



**HAL**  
open science

## Essays in Political Economy

Germain Gauthier

► **To cite this version:**

Germain Gauthier. Essays in Political Economy. Economics and Finance. Institut Polytechnique de Paris, 2023. English. NNT: 2023IPPAX074 . tel-04536950

**HAL Id: tel-04536950**

**<https://theses.hal.science/tel-04536950>**

Submitted on 8 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT  
POLYTECHNIQUE  
DE PARIS

NNT : 2023IPPAX074

Thèse de doctorat



# Essays in Political Economy

Thèse de doctorat de l'Institut Polytechnique de Paris  
préparée à l'École polytechnique

École doctorale n°626 École doctorale de l'Institut Polytechnique de Paris (EDIPP)  
Spécialité de doctorat : Sciences économiques

Thèse présentée et soutenue à Palaiseau, le 30 Août 2023, par

**GERMAIN GAUTHIER**

Composition du Jury :

Julia Cagé Associate Professor, Sciences Po	Présidente
Mirko Draca Associate Professor, University of Warwick	Rapporteur
Ruben Durante Full Professor, Pompeu Fabra University	Rapporteur
Xavier d'Haultfoeuille Professor, ENSAE (CREST)	Examineur
Elliott Ash Assistant Professor, ETH Zürich	Examineur
Alessandro Riboni Full Professor, Ecole polytechnique	Directeur de thèse



# Acknowledgments - Remerciements

Je tiens d'abord à remercier mon directeur de thèse, Alessandro Riboni, pour son soutien lors de ces cinq dernières années. Il a su critiquer mon travail avec bienveillance, de façon à orienter mes recherches sans briser ma motivation. J'ai énormément bénéficié de son oeil de théoricien, m'invitant toujours à mettre en perspective et à prendre du recul sur mes résultats empiriques. Le premier chapitre de ce manuscrit lui doit beaucoup, et j'irais jusqu'à dire qu'il n'aurait probablement pas vu le jour sans ses conseils et encouragements. Il m'a par ailleurs appris que le métier de chercheur ne s'arrête pas à la rédaction d'articles scientifiques, et je lui suis particulièrement reconnaissant de m'avoir guidé dans tous les autres aspects de ce métier. *I pledge to keep an eye on the big picture in Milano, Alessandro!*

Je souhaite aussi remercier Xavier d'Haultfoeuille pour son suivi et ses conseils. Pendant ses années de thèse, il a su approcher mes questions de recherches appliquées avec la rigueur d'un statisticien. Nos discussions d'économétrie ont su satisfaire mon goût pour la certitude d'un résultat mathématique. Paradoxalement, elles m'ont laissé un doute cartésien qui m'assaille dans tous mes projets : « Est-ce que je mesure vraiment ce que je pense mesurer avec cette méthode ? ». Et ce doute ne disparaît généralement qu'après maintes simulations de jeux de données hypothétiques. Merci de m'avoir toujours poussé à développer les outils dont j'avais besoin plutôt qu'à utiliser ceux qu'on mettait à ma disposition.

La recherche n'est pas une entreprise solitaire, et je dois aussi cette thèse à mes formidables co-auteurs. *I would like to thank Elliott Ash for introducing me to the literature on text as data during two wonderful academic visits and supporting me financially during my PhD. Your research has been a true inspiration for my own work.* Pierre Boyer a su stimuler mon goût pour la compétition et la performance et je lui dois beaucoup de ma motivation dans les premières années de ma thèse. Benoit Schmutz m'a toujours impressionné par sa bienveillance et son perfectionnisme, Vincent Rollet par son efficacité et son éthique de travail. J'ai fait mes premiers pas dans le monde de la recherche avec Thomas Delemotte et je garde un très bon souvenir de nos questionnements épistémologiques. Enfin, je dois beaucoup professionnellement et plus encore

à ma co-autrice et compagne Philine Widmer, mais j'y reviendrai.

Je suis très heureux aujourd'hui qu'Alessandro, Xavier, Elliott, Mirko Draca, Ruben Durante, et Julia Cagé aient bien voulu accepter d'être membres de mon Jury de thèse. J'en suis très honoré et je les remercie chaleureusement. *In particular, I would like to warmly thank Mirko Draca and Ruben Durante for accepting to be my referees. Your rigorous lecture and precious feedback are highly appreciated.*

J'ai trouvé un excellent environnement de recherche pour écrire ma thèse au sein du CREST et de l'ETH. J'aimerais remercier le personnel administratif, les nombreux professeurs qui ont pris le temps de discuter mes recherches en séminaire et en privé, ainsi que tous les doctorants qui ont porté ma croix pendant ces années de thèse.

Sur une note plus personnelle, j'aimerais remercier mes amis et ma famille. Mes parents m'ont constamment soutenu et encouragé tout au long de mes années d'études. Je retiens de ma mère sa joie de vivre avec et pour les autres, ainsi que sa passion pour son travail. En t'observant dans la clinique vétérinaire, j'ai toujours su qu'il me fallait trouver une vocation où je ne compterais pas les jours avant les prochaines vacances. Je pense que tu conviendras que je tiens un bon filon avec la recherche en sciences sociales. Après tout, l'Homme est un drôle d'animal ! Mon père et ses projets fantasques n'ont cessé de me montrer que le champ des possibles est plus grand qu'on ne le pense. Je garde de toi cette exhubérance raisonnée et romanesque, et j'espère que nous aurons l'occasion d'écrire encore quelques chapitres ensemble (peut-être au Cape Horn ?). La moue dubitative de mes soeurs, de mes amis d'HEC et de prépa, sur mes recherches m'a toujours poussé à contribuer au débat public plutôt qu'à tomber dans des travers ésotériques. Ma tante, Isabelle, et mon grand-père, André, ont été pour moi des modèles familiaux de chercheurs, et m'ont donné ce goût de la curiosité et de la rigueur intransigeante. Je les en remercie. Enfin, j'aimerais remercier Philine, pour sa contribution à la rédaction de ce manuscrit, pour nos débats autour d'innombrables cafés, mais surtout pour les petites et grandes attentions du quotidien. *Between St.Gallen, Zürich, and Paris, I couldn't have done it without you and I wouldn't have had it any other way. Thank you.* J'ai hâte de vivre les prochaines étapes avec toi et nos filles – Mailin, Solène, et Adaïa – à qui je dédie cette thèse.

# Summary

This PhD dissertation in political economy consists of three independent chapters studying social movements, social media, and political narratives.

Chapter 1 quantifies the impact of the #MeToo movement on sex criminality in the United States. Exploiting the existence of delayed reports to the police, I propose an econometric model to disentangle crime reporting from crime prevalence in crime databases. Solving this long-standing measurement problem allows me to study trends in sex criminality around the #MeToo movement. I find that sexual violence awareness and sex crime reporting had been increasing – and sex crime incidence had been decreasing – for several years before #MeToo in October 2017. Nonetheless, the movement’s sudden mediatization on social media largely reinforced these broader trends.

Chapter 2 argues that social media have ambiguous effects on protest movements. While they facilitate the coordination of protests and lead to persistent online activism, they may also foment radicalization among protesters. To support these claims, the chapter analyzes the Gilets Jaunes protests in France in 2018–2019. It shows that local street protests triggered the creation of large communities of protesters on Facebook. However, these communities progressively became more antagonistic, negative, and ideologically segregated. Moderate discussants left, those who remained radicalized, and Facebook’s recommender algorithm likely contributed its fair share by consistently showcasing radical content.

Chapter 3 develops a new algorithmic approach to measure economic and political narratives from large text corpora. It uses recent neural networks to annotate semantic roles from raw documents, allowing for the construction of *agent-verb-patient* triplets. It then combines phrase embeddings with a clustering algorithm to recover low-dimensional, interpretable narratives (e.g., “taxes kill jobs”). An application to thousands of speeches from the *U.S. Congressional Record* illustrates the method’s potential. An open-source Python package `relatio` is also provided to support further applications.

**Keywords:** social media, social movements, social norms, narratives, natural language processing, measurement of crime, duration models

# Résumé

Cette thèse de doctorat en économie politique est composée de trois chapitres indépendants. Les thèmes principaux abordés sont les mouvements sociaux, leur développement récent sur les réseaux sociaux, ainsi que l'étude quantitative par intelligence artificielle du discours politique.

Le chapitre 1 évalue les effets du mouvement #MeToo sur la criminalité sexuelle aux États-Unis. En exploitant l'existence de plaintes reportées avec un délai à la police, je propose une approche simple pour dissocier la probabilité qu'une victime porte plainte du nombre de crimes sexuels commis dans une série chronologique de crimes sexuels signalés à la police. Résoudre ce problème de mesure de longue date me permet d'étudier l'évolution de la criminalité sexuelle autour du mouvement #MeToo. Je constate que la sensibilisation aux violences sexuelles et la probabilité de porter plainte augmentaient déjà, tandis que les crimes sexuels étaient en baisse depuis plusieurs années avant la médiatisation soudaine de #MeToo sur les réseaux sociaux. Cependant, le mouvement a renforcé ces tendances plus larges.

Le chapitre 2 suggère que les réseaux sociaux ont des effets ambigus sur les mouvements sociaux. Bien qu'ils facilitent la coordination entre les manifestants et permettent de nouvelles formes d'activisme en ligne, ils peuvent également dessiner des trajectoires de radicalisation politique. Pour étayer ces affirmations, le chapitre analyse le mouvement des Gilets Jaunes en France en 2018-2019. Il montre que les manifestations locales ont mené à la création de grandes communautés d'activistes sur Facebook. Ces communautés sont progressivement devenues plus antagonistes, négatives et idéologiquement polarisées. Les plus modérés des interlocuteurs sont partis, tandis que ceux qui sont restés se sont radicalisés. L'algorithme de recommandation de Facebook a probablement contribué à ces trajectoires de radicalisation en présentant systématiquement du contenu radical à ces utilisateurs.

Le chapitre 3 développe une nouvelle approche algorithmique pour mesurer les poncifs économiques et politiques dans de grands corpus de textes. Il utilise des réseaux neuronaux récents pour annoter les rôles sémantiques dans les documents



originaux, permettant la construction de triplets « *agent-verbe-patient* ». Il combine ensuite les plongements lexicaux avec un algorithme de regroupement pour produire des poncifs interprétables de faible dimension (par exemple, « l'impôt tue l'emploi »). La méthode est appliquée à des milliers de discours parlementaires aux États-Unis. Le logiciel libre d'accès *relatio*, écrit en Python, est également fourni pour faciliter d'autres applications.

**Mots clés:** réseaux sociaux, mouvements sociaux, normes sociales, préférences politiques, traitement du langage naturel, mesure de la criminalité, modèles de durée

# Contents

<b>Acknowledgments - Remerciements</b>	<b>i</b>
<b>Summary</b>	<b>iii</b>
<b>Résumé</b>	<b>v</b>
<b>1. Measuring Crime Reporting and Incidence:</b>	
<b>Method and Application to #MeToo</b>	<b>1</b>
1. Introduction . . . . .	3
2. Related Literature . . . . .	6
3. Data and Context . . . . .	9
3.1. Data Sources . . . . .	9
3.2. Descriptive Statistics . . . . .	11
3.3. The Me Too Movement . . . . .	12
4. Conceptual Framework . . . . .	13
4.1. The Canonical Problem . . . . .	13
4.2. Adding Delayed Reports . . . . .	14
4.3. Reported Crimes Around #MeToo . . . . .	15
5. Empirical Framework . . . . .	16
5.1. The Structure of Crime Reports . . . . .	16
5.2. Plaintiff Reporting Hazard . . . . .	17
5.3. Victim Reporting Hazard . . . . .	18
5.4. Modeling Uncertainty . . . . .	19
5.5. Estimation . . . . .	20
5.6. Assumptions . . . . .	23
5.7. Monte Carlo Evidence . . . . .	24
6. Empirical Results . . . . .	25
6.1. Plaintiffs . . . . .	25
6.2. Victims . . . . .	28
6.3. Offenders . . . . .	30
7. Discussion and Mechanisms . . . . .	32

7.1.	Social Norms . . . . .	32
7.2.	Odds of Punishment . . . . .	33
8.	Conclusion . . . . .	33
<b>Appendix</b>		<b>51</b>
A.	Data Sources . . . . .	52
A.1.	City-level Police Records . . . . .	52
A.2.	National Crime Victimization Survey (NCVS) . . . . .	58
A.3.	FBI Datasets . . . . .	58
A.4.	Twitter . . . . .	61
A.5.	Google Trends . . . . .	63
B.	Empirical Issues with Police Data . . . . .	64
C.	Duration Model . . . . .	66
C.1.	Illustration of the Right-truncation Bias . . . . .	66
C.2.	Theoretical Motivation for the Promotion Time Model . . . . .	66
C.3.	Derivation of the Log-Likelihood . . . . .	68
C.4.	Additional Insights from Monte Carlo Simulations . . . . .	72
D.	Robustness and Additional Empirical Results . . . . .	73
<b>2.</b>	<b>Social Media and the Dynamics of Protests</b>	<b>83</b>
1.	Introduction . . . . .	85
2.	Context, Data, and Methods . . . . .	88
2.1.	Brief History . . . . .	88
2.2.	Measuring the Intensity of Mobilization . . . . .	89
2.3.	Textual Analysis of Facebook Discussions . . . . .	91
3.	The Online-Offline Feedback Loop . . . . .	92
3.1.	Empirical Framework . . . . .	94
3.2.	Protests Were Organized Online and Targeted Roundabouts . . . . .	96
3.3.	Roadblocks Spurred Further Online Mobilization . . . . .	97
4.	The Rise of Online Radicalism . . . . .	99
4.1.	Online Discussions Became More Radical Over Time . . . . .	99
4.2.	Facebook's Algorithm Increased Exposure to Radical Content . . . . .	100
4.3.	Radical Discussants Replaced the Moderates . . . . .	102
5.	Conclusion . . . . .	104
<b>Appendix</b>		<b>111</b>
A.	Data Sources . . . . .	112
A.1.	Street Protests . . . . .	112
A.2.	Change.org Petition . . . . .	112
A.3.	Facebook Activity . . . . .	112

A.4.	Tweets of Politicians . . . . .	116
A.5.	Administrative Data . . . . .	116
B.	Supplement for “The Online-Offline Feedback Loop” . . . . .	121
B.1.	Control variables and early mobilization . . . . .	121
B.2.	Details on the roundabout instrument . . . . .	122
B.3.	Robustness of the 2SLS results . . . . .	124
C.	Supplement for “The Rise of Online Radicalism” . . . . .	125
C.1.	Text Pre-processing . . . . .	125
C.2.	Topic Model . . . . .	125
C.3.	Sentiment Analysis . . . . .	126
C.4.	Political Partisanship Model . . . . .	126
<b>3.</b>	<b>Text Semantics Capture Political and Economic Narratives</b>	<b>133</b>
1.	Introduction . . . . .	135
2.	Method: Mining Narratives from Plain Text Corpora . . . . .	136
2.1.	Semantic roles as narrative building blocks . . . . .	137
2.2.	From semantic roles to interpretable actions and entities . . . . .	139
2.3.	Narrative statements and narrative structure . . . . .	140
2.4.	Open-source implementation . . . . .	141
3.	Application to the U.S. Congressional Record . . . . .	142
3.1.	Implementation . . . . .	142
3.2.	Analysis . . . . .	143
4.	Discussion . . . . .	150
4.1.	When should researchers use RELATIO? . . . . .	152
4.2.	Limitations and potential extensions . . . . .	154
	<b>Appendix</b>	<b>161</b>
A.	Python Implementation with RELATIO . . . . .	162
B.	Human Validation . . . . .	165
C.	Inspecting the Quality of Mined Entities . . . . .	169
D.	Summary Statistics . . . . .	173
E.	Examples of Narratives and Sentences . . . . .	174
F.	Comparison to Alternative Text Features . . . . .	187
G.	Additional Results . . . . .	190
H.	Additional Material on Divisive Entities . . . . .	192
I.	Which narratives co-occur together? . . . . .	197
J.	Narratives on Social Media Data . . . . .	200
K.	List of Analyzed Narratives . . . . .	202
L.	List of entities labeled as procedural or noise terms . . . . .	215

M. List of Stopwords . . . . .	217
--------------------------------	-----

# **Chapter 1**

## **Measuring Crime Reporting and Incidence:**

### **Method and Application to #MeToo**

### Abstract

This paper studies the Me Too movement's effects on sex criminality. As many victims do not report to the police, a long-standing empirical challenge with reported crime statistics is that they reflect variations in victim reporting *and* crime incidence. To separate both margins, I develop a duration model that studies the delay between the incident's occurrence and its report to the police. The model accounts for unobserved heterogeneity, never-reporters, and double-truncation in the data. I apply it to the police records of large US cities. Contrary to the widespread view that #MeToo was a watershed moment, I find that sex crime reporting had already been increasing for years before its sudden mediatization in October 2017. Nonetheless, the movement had a positive, persistent impact on victim reporting, particularly for juveniles, racial minorities, and victims of misdemeanors and old crime incidents. The increase in reporting translates into drastically higher probabilities of arrest for sex offenders. Using reported non-sexual crimes as a control group, difference-in-differences estimates suggest the movement also had a sizable deterrent effect.

**Keywords:** crime reporting, crime deterrence, sex crimes, #metoo, survival analysis, double-truncation<sup>1</sup>

---

<sup>1</sup>I am extremely grateful to my advisor Alessandro Riboni for his continuous guidance and encouragement on this project. For helpful discussions and suggestions, I would like to thank Elliott Ash, Christophe Bellego, Christian Belzil, Guillaume Bied, Xavier d'Haultfoeuille, Jean-David Fermanian, Roberto Galbiati, Pauline Rossi, Arne Uhlendorff, Bella Vakulenko-Lagun, Gerard Van den Berg, Philine Widmer, and Yanos Zylberberg. I would also like to thank seminar participants at CREST, ETH Zürich, University of St. Gallen, Bocconi University, University Paris-Nanterre, and HEC Paris. This research is supported by a grant of the French National Research Agency (ANR), "Investissements d'Avenir" (LabEx Ecodec/ANR-11-LABX-0047)".

## 1. Introduction

In October 2017, the Me Too movement led millions of women worldwide to protest against sexual violence. Enthusiastic commentators portrayed the movement as a game-changer in the history of women’s rights.<sup>1</sup> Others, more skeptical, raised concerns about false allegations, backlash effects, and socioeconomic and racial divides.<sup>2</sup> As for most interventions against crime, a major impediment to evaluating the movement’s impact on crime reporting and incidence is that many victims do not come forward. Between 1995 and 2010, U.S. national surveys estimated over 6 million rape and sexual assault victims, of which 60 to 70% did not report the incident to the police (Planty et al. 2013). In turn, a long-standing empirical challenge has been to interpret variations in reported crimes as changes in the number of offenses committed or in victims’ propensity to report (Quêtelet 1831; Levitt 1998).

This paper proposes a general methodology to disentangle victim reporting and crime incidence from police data. I use this novel approach to provide empirical evidence of the Me Too movement’s impact on sex criminality. I find three key results. First, the victim reporting rate had already been increasing for years before #MeToo. Second, the movement had a positive and persistent effect on victim reporting – particularly for the most vulnerable groups of the population. Third, it had a large deterrent effect on sex offenders.

My analysis relies on the police records of New York City, Los Angeles, Cincinnati, and Seattle between 2010 and 2019. I document that more than half of sex crime charges are filed with a delay relative to the incident date. Reports can occur days, months, and sometimes years after the crime. In addition, I show that there was a large increase in reported sex crimes over the decade but that delayed reports primarily drive it. This pattern is consistent with the progressive depletion of a large stock of unreported sex crimes and suggests victims became more likely to report over time.

Building on this intuition, I develop a reduced-form, flexible duration model to analyze delayed reports over time. Victims enter the study on the incident date and exit upon filing a complaint to the police. As some victims never report, I extend the canonical mixed proportional hazards (MPH) model to explicitly account for never-reporters. This modification does not require additional statistical assumptions. However, it requires researcher knowledge about the share of never-reporters at the beginning of the study period. The duration model allows me to reconstruct all subsequent variations

---

<sup>1</sup>For example, see [AP News \(2017\)](#); [Psychology Today \(2017\)](#); [Berkeley Law \(2019\)](#).

<sup>2</sup>For examples, see [Forbes \(2020\)](#); [Harvard Business Review \(2019\)](#); [New York Post \(2020\)](#); [New York Times \(2017a\)](#); [AP News \(2021\)](#); [New York Times \(2017b\)](#).



in this share over time. In turn, I can decompose a time-series of reported crimes into two margins of crime reporting and incidence.

I estimate the model by maximizing the likelihood function integrated over the random effects. However, the model's estimation is complicated by double-truncation in the data. Double-truncation is a non-trivial sample selection scheme that arises because I only observe plaintiffs who report a crime to the police during the study period. Plaintiffs with shorter reporting delays are less likely to enter the study, and thus, left truncation leads to a sample biased towards larger reporting delays. Conversely, plaintiffs with longer reporting delays are less likely to report before the end of the study, and thus, right-truncation leads to a study sample biased towards smaller reporting delays. Without a suitable correction, I show that a naive, out-of-the-box implementation of the MPH model returns severely biased estimates. I thus correct the likelihood to account for double-truncation. The correction weights each observation by the inverse of their sampling probability. The identification assumption is that reporting delays are independent of the date of the incident once conditioning on the history of interventions affecting victim reporting (e.g., #MeToo). In Monte Carlo simulations, the proposed estimator largely outperforms the naive estimator, with no bias in estimates.

I then take the model to the data. First, I focus on *plaintiffs*. I show that abstracting from the subset of victims who did not report to the police provides a lower bound on #MeToo's effects on the hazard of reporting. I find that the hazard of filing a complaint increased by 15 to 20% following the intense mediatization of #MeToo in October 2017. The effect increases over time, suggesting that the movement had a durable impact on reporting norms. Plaintiffs are approximately twice as likely to report past crimes after #MeToo (relative to recent crimes). Thus, encouraging sex crime victims from past decades to reconsider filing a complaint to the police partly allowed older generations to "catch up" with the reporting rate of younger cohorts. Finally, juveniles and racial minorities appear particularly responsive to the intervention. This mitigates concerns of the unequal treatment of age and race in the movement's media coverage (Onwuachi-Willig 2018). These results are robust to various specifications for the baseline hazard, unobserved heterogeneity, incident-level characteristics, and placebo dates.

I further investigate the implications of these findings for the *entire victim population*. In line with national estimates, I assume that 70% of the victims would not have reported incidents for 2010.<sup>3</sup> When accounting for never-reporters, I uncover a large, linear, and positive pre-trend between 2010 and 2017. This contradicts the

---

<sup>3</sup>Absent any future interventions that may affect victim reporting propensity (such as #MeToo).

widespread view that #MeToo was a watershed moment. It is not to say, however, that the movement had no impact. The timing of #MeToo appears as a structural break in the time-series and reinforces pre-existing trends. I estimate that the share of victims who eventually report a sex crime to the police more than doubled over the decade, reaching 75%. #MeToo accounts for approximately 25% of the increase in sex crime reporting over the period. Alternative values for the share of never-reporters in 2010 do not revert these trends but affect their magnitude. I show that my main results are robust to a broad range of reasonable parameter values, ranging from 60 to 80% of never-reporters.<sup>4</sup> Moreover, when focusing on the segment of complaints that lead to arrests, I find qualitatively similar trends, suggesting unfounded allegations do not drive the results.

Next, I reconstruct the time series of sex crime incidence based on estimated reporting rates. Empirical estimates indicate that sex crime incidence decreases by approximately 35% between 2010 and 2019. The decomposition of reported sex crimes thus reveals a substantial increase in sex crime reporting and a substantial decrease in sex crime incidence over time. The two margins partly cancel each other out and are thus less apparent in the time series of reported crimes. Using this newly constructed time-series, I assess the Me Too movement's impact on sex crime incidence. To account for potential confounders, I use reported non-sexual crimes as a plausible control group in a difference-in-differences setup. In the post-treatment period, I find a large and statistically significant deterrent effect of 23% per quarter. I find no effect for placebo dates as well as for non-sexual crimes. My results are also robust to alternative counterfactual models, including an interactive fixed effects model (Xu 2017) and the matrix completion method (Athey et al. 2021). In my baseline specification, the Me Too movement accounts for approximately 12% of the decrease in sex crime incidence over the period.

I consider several channels that may affect victim and offender behaviors. Regarding victims, a plausible explanation is a social norm narrative in which the social cost of reporting a sex crime has decreased. Consistent with this interpretation, I show that the number of tweets surrounding sexual assault has increased over the decade (and so did Google queries). This pattern suggests an increase in sexual violence awareness. Regarding criminal behavior, my results are consistent with a Beckerian model of crime. The reporting rate has increased the probability of arrest for sex offenders, from roughly 15% in 2010 to 37% in 2019. In the data, a one percentage point increase in the probability of arrest is associated with a 0.9 percentage points decrease in sex crimes.

---

<sup>4</sup>This interval encompasses all estimates of the reporting rate of victims of sexual violence by the National Crime Victimization Survey since 2011.

This paper is organized as follows. Section 2 reviews the literature. Section 2 presents the data sources and a brief history of the Me Too movement. Section 4 discusses empirical patterns surrounding #MeToo in light of a simple conceptual framework. Section 5 outlines a general method to infer variations in crime incidence and reporting from police records with delayed reports. Section 6 presents estimates of the Me Too movement's effects on victims and offenders. Section 7 discusses plausible mechanisms. Finally, Section 5 concludes.

## 2. Related Literature

My results inform several distinct strands of literature. First, I contribute to the nascent literature that studies the origins and consequences of the Me Too movement. Public allegations of sexual misconduct have substantially impacted company valuations on financial markets and labor market outcomes (Borelli-Kjaer, Schack and Nielsson 2021; Luo and Zhang 2022; Sophie Calder-Wang and Sweeney 2021; Batut, Coly and Schneider-Strawczynski 2021; Cici et al. 2021; Bernabe 2021; Gertsberg 2022). Closely related to this paper, Levy and Mattsson (2021) estimate that the Me Too movement has led to a 10% increase in sex crime reports in a large sample of OECD countries (including the United States).<sup>5</sup> I study the Me Too movement's impact on sex criminality with a novel methodology. My empirical strategy presents several advantages, including the ability to jointly study variations in crime reporting and incidence at fine-grained scales (e.g., cities). I provide novel insights into the movement's origins and impact in four ways. First, I show that past studies likely underestimated the impact of the Me Too Movement on sex crime reporting. Second – and related to this first point – my results indicate that the movement's deterrent effect is large and statistically significant. Third, I find that the movement had a differential impact along socio-demographic lines and was particularly effective at fostering reporting among racial minorities, juveniles, and past victims of sex crimes. Fourth, I uncover substantial trends in crime reporting and incidence before the Me Too movement. This last result cautions against interpreting social movements as “as-good-as-random” events. If the timing of #MeToo was largely unforeseen, it nonetheless appears in an opportune context of shifting social norms related to sexual violence.

More generally, I contribute to the literature on the under-reporting of gender-based violence (Cheng and Hsiaw 2020). Previous studies have found that the election of female politicians and the number of female officers in the police workforce have increased the reporting of gender-based violence in India (Iyer et al. 2012; Miller and Se-

---

<sup>5</sup>Rotenberg and Cotter (2018) also find similar descriptive evidence for Canada.

gal 2019). Displays of public outrage for highly mediatized affairs, such as allegations of pedophilia in the Catholic Church and particularly gruesome rape cases in India, have also increased the number of victims coming forward (Bottan and Perez-Truglia 2015; Mathur et al. 2019; Sahay 2021; McDougal et al. 2021). The Me Too movement is arguably the largest and most persistent public awareness campaign on sexual violence. I provide empirical evidence that it successfully encouraged victims to file a complaint to the police. Its impact is not limited to victims, however, as the movement is associated with a large decrease in sex crime incidence.

This last finding has implications for the broader literature on crime deterrence. Dating back to Becker (1968), crime deterrence has been extensively studied by scholars (for reviews, see Nagin 2013; Chalfin and McCrary 2017; Doleac 2019). Through the careful study of fines, arrests, and prosecutions, the crime literature analyzes incentives within the criminal justice system but rarely investigates how private attitudes affect enforcement. A smaller literature in political economy, however, has noted that prevailing social norms often co-exist with the criminal justice system and can both facilitate and hinder law enforcement efforts (Hay and Shleifer 1998; Berkowitz, Pistor and Richard 2003; Benabou and Tirole 2011; Young 2015). In particular, the monitoring and punishment of unlawful behaviors depends to a large extent on the willingness of victims to report incidents to law enforcement agencies (Akerlof and Yellen 1994; Dyck, Morse and Zingales 2010; Acemoglu and Jackson 2017). By increasing the probability of arrest and prosecution, my results highlight that changes in reporting norms can successfully enforce socially desirable behaviors when the legal system fails to do so on its own.

This paper also connects to the large amount of literature on the reliability of reported crime statistics. Since the XIX<sup>th</sup> century, scholars refer to the share of crimes that is neither reported to nor recorded by law enforcement agencies as the *dark figure of crime* (Coleman and Moynihan 1996).<sup>6</sup> To this day, unobserved crimes pose a serious empirical challenge for analyzing and interpreting police records.<sup>7</sup> It is particularly problematic when studying the effectiveness of interventions to fight crime. If such interventions affect both crime reporting and incidence, then estimates of treatment effects based on reported crime statistics will be biased (Levitt 1998). Though the dark figure of crime represents a well-documented problem, few approaches attempt to disentangle reporting behaviors from crime rates. As a result, researchers heavily rely on victimization surveys to better monitor victim and offender behaviors. Police

---

<sup>6</sup>The expression *dark figure of crime* is attributed to the Belgian mathematician Adolphe Quételet. Though he raised the issue in the first half of the 19<sup>th</sup> century, it became popular in the 1960s.

<sup>7</sup>For example, around 40% of violent victimizations and 33% of property victimizations were reported to the police in 2019 (Morgan and Thompson 2021).

databases remain, however, the only source of geographically disaggregated data on crime that allows researchers to exploit geographic and time variation in treatment assignment to identify treatment effects. Developing new methods tailored for these records is thus critical to improving our understanding of crime reporting and deterrence. Several contributions in economics have relied on proxy variables to disentangle both margins (e.g., [Aizer 2010](#); [Stephens-Davidowitz 2013](#); [Bellégo and Drouard 2019](#)). Most frameworks implicitly or explicitly assume crimes are reported in short periods to the police or never at all. This is not the case for sex crimes, domestic violence, and harassment complaints, among others. The existence of delayed reports and the empirical issues they raise remain understudied (for notable exceptions, see [Lee and Suen 2019](#); [Klemmer, Neill and Jarvis 2021](#)). I contribute in three ways to this literature. First, I clarify the econometric implications of delayed reports. Second, I propose a solution to monitor variations in victim reporting and crime incidence for crimes reported over long periods. Finally, I provide real-world evidence that underreporting is a serious empirical threat for practitioners and is a first-order concern for the credible impact evaluation of interventions.

On a more technical note, I contribute to the analysis of doubly-truncated data in survival analysis (see [Dörre and Emura 2019](#), for an overview). Beyond police data, double-truncation is a sampling scheme that arises in biostatistics ([LAGAKOS, BARRAJ and Gruttola 1988](#); [Moreira and de Una-Alvarez 2010](#); [Emura and Murotani 2015](#)), engineering ([Ye and Tang 2016](#)), astronomy ([Efron and Petrosian 1999](#)), and economics ([Dörre 2020](#)). Previous research has derived semi-parametric Cox regression models for doubly-truncated data ([Vakulenko-Lagun, Mandel and Betensky 2019](#); [Rennert and Xie 2018](#); [Mandel et al. 2018](#)). These approaches assume *unconditional* independence between the duration and the truncation times. This is often too restrictive an assumption in the presence of time-varying covariates. For instance, in the case of the Me Too movement, it rules out the possibility that the movement would have affected victims' reporting of sex crimes to the police. Furthermore, these models assume homogeneous populations. In practice, unobserved heterogeneity is common, and modeling it is important for valid causal inference ([Abbring and Van den Berg 2003](#)). I thus develop parametric yet very flexible duration models for doubly-truncated data that solve these two methodological shortcomings. First, I relax the unconditional independence assumption to a more realistic *conditional* independence assumption. Controlling for the Me Too movement's viral outbreak on social media thus makes it possible to assess its impact on victim reporting. Second, I extend the model to unobserved heterogeneity.

### 3. Data and Context

#### 3.1 Data Sources

**City-level Police Data** My main data source is detailed incident-level police records for New York City, Los Angeles, Seattle, and Cincinnati, between 2010 and 2019.<sup>8</sup> These cities represent a combined population of approximately 13 million Americans. The records are official administrative data. The data collection is meant to be rigorous and systematic, and is sent later to the FBI for consolidation. Importantly, the records distinguish the incident date from the date of its report to the authorities.<sup>9</sup> This will be crucial for my empirical strategy (see Section 5). For Los Angeles and New York City, I also download offender-level arrest datasets over the same period. This allows me to compute arrest rates per crime category.

I process the data in the following way. I manually classify offenses as *sexual* or *non-sexual*. I exclude sexual offenses related to *pornography, indecency, loitering, sexting, and prostitution*. When possible, I further distinguish sex offenses between *misdemeanors* and *felonies*.<sup>10</sup> For non-sexual offenses, I focus on four broad categories: *burglary, robbery, assault, and murder*. I exclude all other non-sexual offenses from the analysis. For a complete list of offenses considered and excluded, see Appendix A. I also observe the socio-demographic characteristics of plaintiffs and alleged offenders, such as their self-declared race, age, and sex. The sex variable has three groups: *male, female, and unknown*. For the self-declared race, I form four categories: *black, hispanic, white, and other/unknown*. For the age of plaintiffs and offenders, I create a dummy variable *juvenile* that takes values zero for adults (above 18 years old) and one for children (below 18 years old). The age recorded is the plaintiff's (offender's) age upon filing the complaint. Finally, I exclude from the sample all complaints with incoherent dates for the incident or its report to the police.<sup>11</sup>

---

<sup>8</sup>I stop the data collection in December 2019, because lockdowns and other restrictions to fight the COVID pandemic may have affected sex criminality from 2020 onwards.

<sup>9</sup>The Uniform Crime Reporting (UCR) yearly data provides yearly aggregate reports per agency, but no incident-level observations. The National Incident-Based Reporting System (NIBRS) provides incident-level observations, but solely records the incident date and omits the report date to the police. I rely on city-level datasets ([Police Data Initiative 2021](#)). I choose cities that provide (i) incident-level data for the period 2010–2019, (ii) include sex crime reports, and (iii) distinguish the data of the incident and the date of its report to the police. I further exclude Austin and Tucson from the sample, because the distribution of reporting delays in these cities is extremely skewed around the first day and raises data quality concerns. The main results are qualitatively similar when including these two cities in the estimations.

<sup>10</sup>New York City makes a clear distinction between the two categories. Other cities do not.

<sup>11</sup>In some cases, the incident's date is later than the date of its report, or one of the two dates is missing. These represent 1.5% of the raw data.

**The FBI Consolidated Databases** The FBI’s Uniform Crime Reporting (UCR) program provides the general public with a broad range of statistics from local law enforcement agencies. To compare my results to national reported crime statistics, I collect official consolidated crime databases between 1976 and 2019. Reported crimes are harmonized into the Uniform Crime Reporting Summary Reporting System (SRS) (Kaplan 2021). The SRS is a crime report database aggregated by month, agency, and crime category.<sup>12</sup> It is the most comprehensive database on offenses known and clearances by arrest in the United States.<sup>13</sup> The UCR also provides supplementary reports on homicides (Kaplan 2019). This allows me to compute sexual and non-sexual homicide rates.

**The National Crime Victimization Survey** To compare my results to national survey estimates, I collect the National Crime Victimization Survey (NCVS) responses between 2010 and 2019. The Bureau of Justice Statistics (BJS) each year surveys a nationally representative sample of about 240,000 persons in approximately 150,000 households. Respondents are explicitly asked about their experiences of criminal victimization and whether they reported the incidents to the police. Thus, one can compute direct estimates of crime reporting and incidence.

**Google Queries on Sex Criminality** I download from Google the web search interest for the topics “sexual assault” and “Me Too movement” in the United States between 2010 and 2019. The query counts are normalized so that the largest number of queries over the period equals one hundred. I interpret this dataset as a proxy for sexual violence awareness.

**Tweets on Sex Criminality** I download all tweets related to sexual assault on the 1st of each month between 2013 and 2019. I also add tweets from the 15<sup>th</sup> of October 2017 (i.e., the day when #MeToo went viral on social media). I identify tweets on sexual

<sup>12</sup>In 2013, the FBI changed its national definition of rape to include a broader range of sexual offenses. On the FBI’s website, one may read: *“The old definition was ‘The carnal knowledge of a female forcibly and against her will.’ Many agencies interpreted this definition as excluding a long list of sex offenses that are criminal in most jurisdictions, such as offenses involving oral or anal penetration, penetration with objects, and rape of males. The new Summary definition of Rape is: ‘Penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim.’”* This change in definition is unlikely to affect my city-level estimates because I rely on the more detailed city-level classification of offenses to define sexual and non-sexual crimes. In Column 8 of Table 8.2, I show that restricting the sample to 2014 – 2019 does not substantially change my main result.

<sup>13</sup>The National Incident-Based Reporting System (NIBRS) is a more recent data collection effort implemented to improve the overall quality of crime data collected by law enforcement. For each incident, it records the date of its occurrence (but not the date of its report to the police). I do not rely on this database because it has limited geographical coverage relative to the SRS. According to the FBI, in 2017, it covered 33% of the US population.

assault based on a large set of explicit keywords, such as: “rape”, “sexual assault” and “#metoo”. The traffic on Twitter has increased drastically over the years, going from hundreds of daily tweets at its creation in 2006 to 500 million daily tweets in 2021. To make meaningful comparisons over time, I collect the total number of tweets posted on each day of data collection. The complete list of keywords and further details may be found in Appendix A. Once again, I interpret this dataset as a proxy for growing sexual violence awareness.

## 3.2 Descriptive Statistics

Summary statistics on the city-level dataset are provided in Table A.3. Between 2010 and 2019, there were over 2 million crime incidents reported. Among those, approximately 110,000 were sexual crimes. Sexual criminality remains largely gender-specific, with 87% of reports filed by women in the sample (95% when excluding incidents without this information). In terms of declared race, black plaintiffs and Hispanics form the bulk of sex crime reports. Furthermore, children and teenagers are particularly exposed to sexual criminality, with around 43% of plaintiffs declaring being below 18 when filing the complaint. Alleged offenders, on the other hand, are mainly adult males belonging to a racial minority.

A striking feature in the data is the incidence of delayed reporting: approximately 30% of the total number of offenses are not reported on the day of the incident (I refer to these complaints as *delayed* reports as opposed to *direct* ones). But this figure hides substantial heterogeneity across offenses. 42% of sex crimes are direct reports as opposed to 79% for non-sexual assaults and 83% for robberies. The difference is also more sizable for mean reporting delays. The average time to report a sex crime is 197 days, as opposed to less than six days for non-sexual assaults, robberies, and burglaries.<sup>14</sup> The standard deviation in reporting delays is almost sixteen times larger for sex crimes relative to non-sexual assaults, robberies, or burglaries. Figure A.1 presents the sample distribution of delayed sex crime reports. Approximately 40% of plaintiffs report on the day of the incident, 80% within the first month, and 90% within the first year.

Delayed reporting also varies along socio-demographic lines. Longer reporting delays are typically expected for subgroups of the population with higher costs to filing a complaint. Figure A.2 displays boxplots of reporting delays per age group and declared plaintiffs’ race. Juveniles report over larger periods than adults. Hispanic and black plaintiffs also display longer reporting delays than other racial categories.

---

<sup>14</sup>Murders have rather long reporting delays when the police record them with a delay. Those correspond to cold cases.



### 3.3 The Me Too Movement

The Me Too movement is a social movement against all forms of sexual misconduct where people share and publicize allegations of sex crimes. Its explicit goal is to raise awareness of the pervasiveness of sexual violence in society. Social activist Tarana Burke launched the movement on MySpace in 2006. For over ten years, the campaign focused on female minorities (mainly black women) and benefited from limited media coverage. On the 15<sup>th</sup> of October 2017, it was popularized by actress Alyssa Milano in reaction to the Harvey Weinstein affairs. She tweeted: *“If you’ve been sexually harassed or assaulted write ‘me too’ as a reply to this tweet.”* In the following days, the hashtag #MeToo spread virally on social media and was posted millions of times on Twitter, Facebook, and other platforms worldwide.

Following its mass mediatization, mentions of the hashtag #MeToo have dwarfed past references to the Me Too movement (see Figures A.7). However, this large discontinuity in the time series should not oust the social and historical context in which the movement emerged. Several pieces of anecdotal evidence suggest that attitudes toward sex crimes had been changing for over a decade. The movement appeared 11 years before it became popular. The traditional media also brought many affairs to the spotlight during the 2000s: examples include sex crime allegations in the Catholic Church, the Bill Cosby sexual assault cases, and allegations concerning Harvey Weinstein before 2017. On social media, several hashtags denouncing violence against women preceded #metoo but were less viral (e.g., #YesAllWomen, #IAmNotAfraid-ToSpeak, #myHarveyWeinstein and #BeBrave). Given these early signs of changing attitudes towards women, one would expect positive time trends in crime reporting before October 2017. I find empirical support for this hypothesis in Section 6.

Debated on television and in the newspapers, the Me Too movement also raised critiques as it gained momentum. Some pointed to the risk of false allegations.<sup>15</sup> Others claimed the movement failed to recognize the heightened vulnerability that women of color frequently face.<sup>16</sup> Finally, proponents warned against the potential backlash.<sup>17</sup> I address these concerns in my empirical assessment of the movement’s impact on victims and offenders. Overall, in the scope of this study, false allegations are unlikely to be the drivers behind the large increase in sex crime reports, the reporting rate increased more for black, Hispanic, and juvenile victims, and there was no backlash effect in sex crime incidence (see Section 6).

<sup>15</sup>For example, see [Forbes \(2020\)](#) and [New York Post \(2020\)](#).

<sup>16</sup>For example, see [New York Times \(2017a\)](#) and [AP News \(2021\)](#).

<sup>17</sup>For example, see [Harvard Business Review \(2019\)](#).

## 4. Conceptual Framework

*“I do not fear to say that all we possess of statistics of crime and misdemeanors would have no utility at all if we did not tacitly assume that there is a nearly invariable relationship between offenses known and adjudicated and the total unknown sum of offenses committed.”*

– Quêtelet (1831)

This section reviews the empirical issues researchers encounter when studying police data. I focus more specifically on the large share of crimes that are not reported or reported with a delay and their consequences for causal inference.

### 4.1 The Canonical Problem

Consider an analyst disposing of reports recorded by the police between  $\tau_1$  and  $\tau_2$ , respectively the first and last calendar data collection dates. She aims to study the impact of an intervention (e.g., the Me Too movement, an increase in the number of police officers, a harsher institutional penalty) on the number of crimes  $C_t$ . The intervention takes place in period  $t^* \in [\tau_1, \tau_2]$ . For each period  $t$ , she observes  $R_t$ , the number of crimes recorded by the police. As a share of victims does not report the incident to the police, reported crimes  $R_t$  generally do not equate to the total number of crimes  $C_t$ . Let  $r_t$  denote the victim reporting rate. Assuming no delayed reporting, we have

$$R_t = r_t \times C_t. \quad (1.1)$$

It becomes apparent that reported crimes  $R_t$  are a function of two latent variables and that a simple linear regression framework will be subject to an omitted variable bias. Let  $D_t$  be a dummy variable that takes the value one in periods after the intervention (i.e.,  $t \geq t^*$ ) and zero otherwise. We have

$$\log(C_t) = a + bD_t + \varepsilon_t - \log(r_t), \quad (1.2)$$

where  $a$  is an intercept term,  $b$  is the coefficient associated with  $D_t$ , and  $\varepsilon$  is an error term.

If the share of unreported crimes  $r_t$  is correlated to the treatment  $D_t$ , then estimates of  $b$  will be biased. In many applications, researchers explicitly or implicitly assume

that the reporting rate  $r_t$  is orthogonal to  $D_t$  to conduct inference. Though this is a convenient assumption, it is also unlikely to hold in practice. On the contrary, interventions aimed at fighting crime often increase the probability of arrest or the severity of sentencing. Theory suggests that these factors matter for victims and offenders, thus one would expect such interventions to lower crime rates and increase reporting rates.

## 4.2 Adding Delayed Reports

In the absence of delayed reports, reported crimes  $R_t$  may be understood as in Equation 1.1. In practice, however, delayed reports are common in police data, particularly for sex crimes (as documented in Subsection 3.2).

Delayed reports require a careful understanding of the reporting guidelines of law enforcement agencies. There are now two dates to consider: the incident date and the date of its report to the police. Aggregating reported crimes on the date of their report or the incident date will generally lead to different counts (and ultimately estimates). When both dates are available, it is straightforward to show that crime trends are typically sensitive to the choice of the date (see Appendix A for examples based on police data).<sup>18</sup>

One can extend Equation 1.1 to account for delayed reports. Let  $R_{t,\tau_1,\tau_2}$  denote the number of crimes committed in period  $t$  that were reported between  $\tau_1$  and  $\tau_2$ ,  $\tau_0$  the calendar date of the earliest incident in the data,  $Y$  the time to report to the police,  $F$  its associated cumulative distribution function, and  $\chi_t$  the history of interventions.  $R_{t,\tau_1,\tau_2}$  is now

$$\mathbb{E}[R_{t,\tau_1,\tau_2}] = p_t \times C_t, \quad (1.3)$$

where  $p_t = F(\tau_2 - t | \chi_t) - F(\max(\tau_1 - t, 0) | \chi_t)$  is simply the probability of reporting a crime that occurred at date  $t$  within the study period. The omitted variable bias in the classical regression analysis remains (Equation 1.2). Furthermore,  $p_t$  depends on the study period considered (i.e., on  $\tau_1$  and  $\tau_2$ ). Thus, it is also mechanically correlated to  $D_t$ . For instance, the closer  $t$  is to the end of the study period  $\tau_2$ , the smaller the probability of reporting a crime that occurred in period  $t$  before the end of data collection. This implies a spurious decreasing time trend in observed reported crimes  $R_{t,\tau_1,\tau_2}$ . As a result, in the classical regression framework (Equation 1.2), estimates of the marginal effect of  $D_t$  will generally be biased (see Appendix A for an example based on police

<sup>18</sup>In most cases, both dates are not available. For instance, in the United States, the Uniform Crime Reporting (UCR) data focuses on the report date, whereas the National Incident-Based Reporting System (NIBRS) relies on the incident date.

data).<sup>19</sup>

### 4.3 Reported Crimes Around #MeToo

Given never-reporters and delayed reporting, understanding the effects of the Me Too movement’s sudden mass mediatization is not straightforward. City-level and national reported crime statistics indicate that sex crime reports have increased between 2010 and 2019. In Figure A.5, I plot national reported crime trends for sexual and non-sexual crimes over time. The trends are largely similar to those found in my sample of cities (see Panel A of Figure 8.1). Reported sex crimes at the national and city level increase by approximately 40%. Interestingly, the surge starts before #MeToo. Following Equation 1.1, this can be rationalized either as an increase in sex crime reporting or an increase in sex crime incidence, if not variations in both latent variables.

Often, researchers turn to victimization surveys for direct estimates of sex crime incidence and reporting. Thus, in Figure A.4, I present estimates from the NCVS between 2010 and 2019. Survey evidence indicates that increased sex crime incidence mainly drives the rise in sex crime reports (+300% over the decade). The increase is particularly large after #MeToo, suggesting a backlash effect.

At the same time, this narrative appears inconsistent with several pieces of evidence in police records. First, reported non-sexual crimes at the national level decreased by approximately 20% (and they remain stable in the city-level dataset). If we assume that sex crime incidence is positively correlated with non-sexual crimes, then this suggests a decrease in sex crime incidence. Second – and perhaps even more convincingly – homicides related to sex crimes decrease by approximately 70% between 2010 and 2019, whereas homicides unrelated to sex crimes remained stable. Sex crime homicides are very few (less than a hundred cases per year), but arguably do not suffer from under-reporting. Once again, if we assume that sex crime incidence is positively correlated to sex crime homicides, this suggests a decrease in sex crime incidence. Third, in Panel B of Figure 8.1, I distinguish delayed and direct reports for sex crimes. Delayed reports increased twice as much as direct reports over the period. This stylized fact is consistent with the depletion of a large stock of unreported sex crimes being progressively reported to the police. Thus, it also suggests an increase in reporting behaviors rather than an increase in sex crime incidence.

<sup>19</sup>Interestingly, they will be biased even if the intervention  $D_t$  has no effect on crime incidence  $C_t$  or crime reporting  $p_t$ .

## 5. Empirical Framework

Given the prevalence of delayed reporting in police records, the timing of sex crime reporting is an attractive dependent variable to separate sex crime incidence and reporting. Following Equation 1.3, the method proposed in this paper aims at estimating the probability distribution of times to report  $Y$  (which is potentially affected by a series of interventions) to recover an estimate of  $p_t$ . To this end, I develop reduced-form duration models to estimate the reporting hazard a sex crime to the police. In this section, I discuss the models, their estimation, and validate my estimators via Monte Carlo simulations.

### 5.1 The Structure of Crime Reports

To begin, I focus on the data structure of police records. Some of its peculiarities are relevant for survival analysis. Figure 8.2 presents a graphical summary of police data as duration data. The study window goes from  $\tau_1$  to  $\tau_2$  (solid vertical lines), with an intervention  $D_t$  in between represented by a dashed vertical line. Some plaintiffs report before the intervention's implementation and form the control group (non-treated observations). Still, others are affected by the intervention and form the treated group (treated observations). Some plaintiffs report before the start of the study and are unobserved (left-truncated data points). Some plaintiffs have not yet reported a crime to the police by the end of the study but will in the future and are unobserved (right-truncated data points). Finally, some victims may decide never to report and are unobserved. We will call them never-reporters. This graphical depiction raises two empirical challenges to correctly estimate the probability distribution of times to report  $Y$ .

First, the data is doubly-truncated (on the left and the right). Though left-truncation is common in economic applications of survival analysis, right-truncation is a relatively understudied truncation scheme that requires special attention. To account for double-truncation, I provide an analytical correction of the log-likelihood, which I explain in greater detail in Subsection 5.5.

Second, the model needs to account for the share of never-reporters. This implies that the cumulative distribution function of times to report  $Y$  will be improper and have a positive mass as  $y$  tends to infinity. In bio-statistics, such models are referred to as "cure models" (see [Amico and Van Keilegom 2018](#), for a review). I propose a promotion time model to account for this stylized fact, which I explain in greater detail below (see Subsection 5.3).

## 5.2 Plaintiff Reporting Hazard

Let  $(Y, X)$  denote a random vector where  $Y \in \mathbb{R}^+$  is the time to report of a victim and  $X \in \mathbb{R}^d$  contains observed covariates. The standard Mixed Proportional Hazards (MPH) model assumes the conditional reporting hazard at time  $y$  has the form<sup>20</sup>

$$h_{it}^{(p)}(y \mid \gamma_i, x_{ity}) = h_0(y) \exp(\beta' x_{ity}) \gamma_i, \quad (1.4)$$

where  $h_0 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is the baseline hazard function,  $(\beta') \in \mathbb{R}^{d+1}$  is the vector of regression coefficients and  $\gamma_i \in \mathbb{R}$  is an individual-specific, time-invariant, unobserved random effect (i.e., often referred to as unobserved heterogeneity or frailty). When  $\gamma_i = 1$  for all observations, the model boils down to the canonical Cox model (Cox 1972). The core assumptions of this model are (i) proportional hazards and (ii) the multiplicative decomposition between the baseline hazard, the effect of observed covariates, and the effect of individual-specific unobserved heterogeneity (Van den Berg 2001). Note that covariates  $x_{ity}$  may be time-varying. In such cases, their values depend on the calendar incident date  $t$  and the duration  $y$  (e.g., if  $t + y \geq \text{Oct.2017}$ , then  $\text{MeToo}_{ity} = 1$ , otherwise  $\text{MeToo}_{ity} = 0$ ).

In an ideal empirical setting, a researcher would observe all victims, and it would be straightforward to estimate the hazard of reporting a crime to the police. Unfortunately, as many crimes go unreported, fitting the model in Equation 1.4 to police records only provides estimates of the reporting hazard a crime to the police among *plaintiffs* (see Panel A of Figure 8.3). It thus focuses on the duration of reporting delays. There are compelling reasons for focusing on this dependent variable as many studies have highlighted the benefits of early formal disclosure (Klemmer, Neill and Jarvis 2021).<sup>21</sup>

Despite the importance of studying the duration of delays, the duration alone falls short of informing us of the reporting rate in the general population of *victims* (see Panel B of Figure 8.3). At the same time, one would assume that the reporting rate correlates with reporting delays. If plaintiffs report *faster* to the police, then perhaps

<sup>20</sup>I write random variables in uppercase and their realizations in lowercase.

<sup>21</sup>Tavarez (2021) summarizes: “Delayed reporting of sexual violence can adversely impact medical treatment, case outcomes, and the overall mental state of survivors. [...] Survivors reporting and seeking support within the first 24 hours have better treatment outcomes and their medical examinations can provide the best possible forensic evidence to assist prosecution [...] forensic evidence for potential drugging, collected through the survivor’s hair, can only be collected approximately within one month of the rape [...] Therefore, victims who formally disclose sooner (approximately within 72 hours) have been found to experience a broader extent of available options and better case outcomes, such as evidence collection, emergency medical care, and apprehension of the perpetrator.” In addition, longer reporting delays are often associated with repeated, longer periods of abuse. Formal reporting may ensure the safety of victims and prevent future offenses.

victims also report *more*. In what follows, I present a modification of the MPH model that builds on this intuition.

### 5.3 Victim Reporting Hazard

We can explicitly account for never-reporters if we enforce the baseline hazard function as a density function and add an intercept to the regression model of Equation 1.4:

$$h_{it}^{(v)}(y | \gamma_i, x_{ity}) = f_0(y) \exp(\alpha + \beta' x_{ity}) \gamma_i, \quad (1.5)$$

where  $f_0$  is a proper density function<sup>22</sup> and  $(\alpha, \beta') \in \mathbb{R}^{d+1}$  is the vector of regression coefficients. In this model,  $\alpha$  accounts for the share of never-reporters at baseline,  $\beta$  synthesizes covariate effects on the response, and  $f_0$  models the duration influence. The model incorporates the main stylized facts mentioned previously: (i) a share of victims report with a delay, and (ii) a share of victims will never report.<sup>23</sup> In the simplest case of no time-varying covariates, the baseline proportion of victims who will never report a crime to the police is<sup>24</sup>

$$\lim_{y \rightarrow \infty} S_i^{(v)}(y | \gamma_i, x_i) = \exp\left(-\gamma_i \exp(\alpha + \beta' x_i)\right).$$

Like the MPH model, the model's core assumption is proportional hazards. In fact, when the baseline hazard is left unspecified, the two models produce numerically similar estimates for  $\beta$ .<sup>25</sup> The MPH model would, however, require observing the never-reporters as right-censored observations to study the victim reporting propensity. By construction, never-reporters are unobserved in police records. The promotion time model allows me to specify the share of never-reporters at baseline (e.g., in the pre-intervention period) through the plugin parameter  $\alpha$ . Figure 8.3 presents a graphical intuition on the differences between the two models when applied to police records.

In Subsection 5.7, I investigate via simulations the sign of the bias when misspecifying  $\alpha$ . I find that setting a lower value than the true share of never-reporters biases

<sup>22</sup>I systematically distinguish the baseline distribution's cumulative, hazard, density and survival functions (i.e.,  $F_0, h_0, f_0, S_0$ ) from the functions related to the distribution of times to report  $Y$  of victims (i.e.,  $F^{(v)}, h^{(v)}, f^{(v)}, S^{(v)}$ ), and the distribution of times to report  $Y$  of plaintiffs (i.e.,  $F^{(p)}, h^{(p)}, f^{(p)}, S^{(p)}$ ).

<sup>23</sup>The model has a theoretical interpretation as a standard Poisson counting process. See Appendix C.

<sup>24</sup>See Appendix C for the extension to time-varying covariates.

<sup>25</sup>In the MPH regression, the baseline survival function is  $S_{it}^{(p)}(y | \gamma_i, x_i) = \exp(-\gamma_i \exp(\beta' x_i) H_0(y))$ . The baseline cumulative hazard  $H_0$  is unspecified so that it can account for  $\alpha$  and  $F_0$  in Equation 1.5, hence the numerically close estimates.

estimates of covariates on crime reporting toward zero.<sup>26</sup> In the application to the Me Too movement (see Section 6), I set the plugin parameter based on empirical evidence from the National Crime Victimization Survey, which suggests that roughly 70% of sexual crimes are never reported to the police. I also plot crime and reporting trends for a broad range of parameter values as robustness checks (see Appendix D). In practice, the model rescales the observable part of the distribution of times to report  $Y$ . The choice of  $\alpha$  thus changes the magnitude of the effects uncovered but not the underlying trends.

Based on Equation 1.5, it is straightforward to recover an estimate of sex crime incidence. Recall  $C_t$  the total number of crimes committed in period  $t$ ,  $R_{t,\tau_1,\tau_2}$  the total number of reported incidents for date  $t$  by the end of the study period, and  $F^{(v)}$  the cumulative distribution function of times to report in the victim population. Then we have<sup>27</sup>

$$\hat{C}_t = \frac{R_{t,\tau_1,\tau_2}}{\widehat{F^{(v)}}(\tau_2 - t \mid \chi_t) - \widehat{F^{(v)}}(\max(\tau_1 - t, 0) \mid \chi_t)}. \quad (1.6)$$

The resulting estimate  $\hat{C}_t$  accounts both for double-truncation in the data and never-reporters, the two main empirical challenges presented in Section 5.1. It also accounts for variations in the hazard rate related to observable (potentially time-varying) covariates (e.g., a policy intervention).

## 5.4 Modeling Uncertainty

Randomness is modeled through unobserved heterogeneity and the baseline hazard. A common critique of duration models is that their results often depend on too restrictive parametric assumptions. To minimize parametric assumptions, I specify flexible functional forms for the baseline hazard and unobserved heterogeneity.

**Baseline Hazard** A priori, the baseline hazard of delayed reports  $h_0$  in Equations 1.4 and 1.5 is unknown. To infer it from the data, I model the hazard as a piecewise constant function (Yin and Ibrahim 2005). Formally, given a set of time points  $y_1, \dots, y_{m+1} \in (\mathbb{R}^+)^{m+1}$  such that  $y_1 = 0 < y_2 < \dots < y_m < y_{m+1}$ , and marginal hazard

<sup>26</sup>The intuition for this is that ignoring never-reporters compresses differences between the cumulative distribution functions of delayed reports (see Figure 8.3).

<sup>27</sup>In some cases, multiple charges may be filed for one offender. My estimates of sex crime incidence measure the number of sex crimes committed in a period and remain agnostic on the number of offenders involved.



rates  $h_{01}, \dots, h_{0m} \in (\mathbb{R}^+)^m$ , the hazard  $h_0$  may be written

$$h_0(y) = \sum_{l=0}^m h_{0l} I_l(y) \quad \text{with} \quad I_l(t) = \begin{cases} 1 & \text{if } y < y_{l+1} \\ 0 & \text{otherwise.} \end{cases} \quad (1.7)$$

This allows for a flexible yet fully parametric baseline reporting hazard. With a unique time point  $k_1 = 0$ , the model reduces to an exponential cure rate model. With a large number of time points (and a large number of observations), the model resembles semi-parametric models (for which the baseline hazard is left unspecified).<sup>28</sup> Furthermore, the large amount of observations in the data allows me to specify a large number of breaks  $m$ , making the baseline hazard flexible in theory and practice.

**Unobserved Heterogeneity** In my baseline results, I assume unobserved heterogeneity is gamma-distributed with mean one. In a large class of frailty models, the frailty distribution among survivors converges to a gamma distribution under mild regularity assumptions (Abbring and Van Den Berg 2007). This makes it an attractive parametric distribution to account for time-invariant, unobserved determinants of crime reporting.

I also consider an alternative, computationally more intensive approach. I model the population under study as consisting of  $K$  sub-populations with different frailties  $\{\gamma_k\}_{k \in \{1, \dots, K\}}$  and respective shares within the population  $\{s_k\}_{k \in \{1, \dots, K\}}$ . I impose that all parameters are strictly positive and that the sum of the shares is one. I set one frailty effect to one ( $\gamma_1 = 1$ ), so that other frailty effects are relative to this subpopulation. This formulation is a general specification of unobserved heterogeneity, which can account for various distributions. It is well-known that such a parametrization correctly accounts for the bias induced by unobserved heterogeneity, but does not accurately estimate the distribution of unobservables (Heckman and Singer 1984b).

## 5.5 Estimation

**Likelihood** Contrary to common applications in the economics literature, police records present an additional empirical challenge as they raise the issue of double-

---

<sup>28</sup>It may also prove helpful to study the effects of statutes of limitations. A statute of limitations is a time limit for a person to file a charge. If the time limit expires, the individual who wishes to bring forth a claim is barred from doing so. In other terms, the alleged offender may not be prosecuted anymore. A statute of limitations may have two effects on delayed reports. First, it sets a hard threshold for the time to report, beyond which the reporting hazard is null. This is not a problem for the duration model. Second, it could increase the reporting hazard as the time limit draws nearer. Researchers can formally test this by specifying breaks in the baseline hazard  $h$  close to the time limit.

truncation: crime incidents are observed if they are reported within the study period. Some reports may occur after the end of the study (i.e., right-truncation), and others may occur before its start (i.e., left-truncation). Right truncation implies an oversampling of shorter durations. Conversely, left truncation means an oversampling of longer durations. Without an appropriate correction, a naive estimation will lead to biased estimates (Dörre and Emura 2019).

Formally, let  $T$  denote the incident date,  $Y$  the time to report to the police,  $U = \max(0, \tau_1 - T)$  the left-truncation time, and  $V = \tau_2 - T$  the right truncation time. Recall that  $\tau_1$  and  $\tau_2$  are respectively the start and end of the study period. Note also that  $V = U + d$  where  $d = \tau_2 - \tau_1$ .<sup>29</sup> Finally, let  $f$  and  $g$  denote the density functions of  $Y$  and  $U$ . Under double-truncation, I observe  $n$  incidents indexed by  $i$  from the probability distribution  $(T, Y)$  given  $U \leq Y \leq V$ . The density of each data point  $(u_i, y_i, v_i)$  is

$$P(U = u_i, Y = y_i | U \leq Y \leq U + d).$$

In general, when subjects have unequal probabilities of selection, then the observed sample will not be representative of the underlying target population. The associated likelihood is

$$L(\Theta) = \prod_{i=1}^n \frac{f(y_i)g(u_i)}{\int_u \left( \int_u^{u+d} f(y)dy \right) g(u)du}.$$

This likelihood is complex, but under the assumption of independence between  $Y$  and  $U$ , one can decompose it into two, somewhat more tractable conditional likelihoods:

$$L(\Theta) = \prod_{i=1}^n \frac{f(y_i)}{\int_{u_i}^{u_i+d} f(y)dy} \times \prod_{i=1}^n \frac{\left( \int_{u_i}^{u_i+d} f(y)dy \right) g(u_i)}{\int \left( \int_u^{u+d} f(y)dy \right) g(u)du}.$$

I use the first conditional likelihood to make an inference on  $f$ . This first term is relatively intuitive. It is an inverse-probability weighting approach in which observations are weighted by the inverse of their sampling probability. Furthermore, a major advantage of focusing on the first conditional likelihood is that I do not specify the distribution of the truncation time  $U$ . The likelihood to maximize eventually simplifies

<sup>29</sup>In the case of police records,  $V$  is thus entirely determined by  $U$  and  $d$ . This is referred to as *fixed-length* double-truncation, but the results presented below also hold for more general double-truncation schemes.

to

$$L(\Theta) = \prod_{i=1}^n \frac{f(y_i)}{F(u_i + d) - F(u_i)}.$$

Conditioning on observed covariates is straightforward. The covariates can be time-varying (e.g., policy interventions) and thus relax the independence assumption between  $U$  and  $Y$  to a more realistic, conditional independence assumption.<sup>30</sup> If specified, random effects are integrated out. I estimate the models by full maximum likelihood. For further details, see Appendix C.

**Identifiability** The identifiability of proportional hazards models with unobserved heterogeneity has been widely studied (for detailed discussions and proofs, see [Elbers and Ridder 1982](#); [Heckman and Singer 1984a](#); [Abbring and Van den Berg 2003](#)). In general, the baseline hazard and the distribution of unobserved heterogeneity are identified if the mean of the unobserved heterogeneity is finite and there is at least one regressor ([Elbers and Ridder 1982](#)). I thus impose this constraint on the random effects.

A perhaps less well-known issue is the identifiability of proportional hazards models under double-truncation. Let us first consider the case without left-truncation. Then, if the right-truncation time is greater than the maximum victim time to report,  $f$  is (non-parametrically) identifiable. This assumption is referred to as positivity ([Vakulenko-Lagun, Mandel and Betensky 2019](#)). On the other hand, if the maximum truncation time is less than the maximum time to report, then  $f$  is only identifiable up to a constant of proportionality ([Seaman, Presanis and Jackson 2021](#)). In this case, the selection mechanism does not allow us to observe values of  $Y$  greater than the maximum truncation time. Thus, we do not know what proportion of individuals experience a time to report greater than the maximum truncation time. The inclusion of left truncation often allows us to observe the entire distribution of delays. For instance,

<sup>30</sup>Recently, inverse-probability weighting approaches have been proposed for fitting the Cox model to doubly-truncated data ([Mandel et al. 2018](#); [Rennert and Xie 2018](#)), of which right-truncated data is a special case ([Vakulenko-Lagun, Mandel and Betensky 2019](#)). They rely on the non-parametric maximum likelihood estimators of the selection probabilities proposed by [Efron and Petrosian \(1999\)](#) and [Shen \(2010\)](#). The main assumption is that  $Y$ ,  $U$  and  $V$  are *unconditionally* quasi-independent. This would imply that the incident date does not affect the plaintiffs' time to report in the context of crime reports. This runs precisely counter to my research prior: the time to report likely varies with the incident date, as plaintiffs are more or less likely to report a crime to the police over time (e.g., before/after the MeToo movement). I formally test this quasi-independence assumption for each city in my dataset ([Martin and Betensky 2005](#)). I reject the null hypothesis for all cities that survival and truncation times are quasi-independent at all standard significance levels. Thus, these methods are not a good fit in this empirical context. My approach relies on a less demanding and more realistic *conditional* independence assumption. However, it comes at the expense of specifying a parametric baseline hazard. To limit the impact of this parametrization, I specify flexible, piece-wise constant baseline hazards (see Subsection 5.2).

I observe reports filed 30 years after the incident in my data. It appears reasonable to assume this is the maximum time to report and that the entire distribution of times to report of *plaintiffs* can be estimated from the data. However, even in the presence of left-truncation, the entire distribution of times to report of *victims* cannot be estimated from the data. By construction, right-truncation does not allow us to observe the share of never-reporters. This is why the latter, represented by the parameter  $\alpha$ , enters the model as a plugin parameter in Equation 1.5.

## 5.6 Assumptions

The core assumption of the models in Equations 1.4 and 1.5 is proportional hazards. I assume that the effect of covariates does not depend on the time to report  $Y$ . More formally, for all delays  $y$ , I assume that

$$\beta(y) = \beta.$$

In practice, the assumption of proportional hazards can be tested and relaxed on the observable part of the data by allowing for time-dependent covariate effects. The model on plaintiffs (Equation 1.4) can be made fully flexible by interacting all covariate effects with a transform of time to report (e.g., bins of durations or their log). The same can be done for the model on victims (Equation 1.5). We cannot, however, observe the share of never-reporters from police records alone. This implies that one cannot formally test if and how the speed at which plaintiffs report is related to the share of never-reporters.<sup>31</sup>

Though unlikely in reality, it is always possible to construct models in which the two are unrelated. Consider for instance a model that separates the population of victims into two categories: a share  $p$  of victims report to the police (potentially with a delay) and a share  $1 - p$  never do. If we allow covariates to affect both the share of victims who report  $p$  and the shape of the survival function for this category  $S$ , then we have

$$S_{it}^{(v)}(y | x_i) = 1 - p(x_i) + p(x_i)S(y | x_i).$$

In such a model, delayed reporting is unrelated to the share of never-reporters. For example, an extreme case is when covariates do not affect the share of victims who

---

<sup>31</sup>To further test the model's adequacy, criminal victimization surveys could distinguish the incident date from the date of its police report. The share of never-reporters would then be identifiable from the data, and one could formally test for proportional hazards. The data to do this is currently unavailable as the National Crime Victimization Survey asks respondents about their experiences of criminal victimization in the past six months before the interview.

eventually come forward:  $P(x_i) = p$ . Then the distribution of observed times to report is uninformative of the share of never-reporters, and police data with delayed reports is of no help to researchers.

## 5.7 Monte Carlo Evidence

To assess the performance of my estimator, I run a series of Monte Carlo simulations. I benchmark my models and estimators against an out-of-the-box implementation of the MPH model with gamma-distributed frailty. The implementation is from the R package *FrailtyEM* (Balan and Putter 2019). The package handles right-censoring and left-truncation. However, it is not designed (and does not claim) to handle double-truncation. If right-truncation presents a serious empirical issue in our context, this implementation should return biased estimates.

I first simulate datasets from Equation 1.4. The hazard  $h_0$  is modeled as a piecewise constant exponential function with mean  $\lambda_1 = 0.2$  for the first period and mean  $\lambda_2 = 0.01$  for all other periods. Three covariates do not vary over time. Their effects on the hazard are respectively  $-0.5, 0.3$ , and  $0$  (on the log-scale). To capture the double-truncation scheme, observations are exponentially truncated on the right at rate  $\lambda = 0.005$  and on the left at rate  $0.05$ . For 1000 data points, this leaves me with approximately 400 observations per simulation. I simulate 1000 datasets. Results are presented in Panel A of Figure 8.4. Estimates of the MPH model without the analytical correction are severely biased (except for the null effect, which seems correctly estimated). On the other hand, the proposed correction performs well and results in unbiased estimates. This first set of evidence indicates the analytical correction of the likelihood appropriately accounts for double-truncation.

This is reassuring, but police records stem from a complex and messier data-generating process. I thus turn to more demanding and realistic simulations that include the presence of time-varying covariates (e.g., interventions), unobserved heterogeneity, and never-reporters (see Equation 1.5). I simulate time series of crime reports over 200 periods. At each period  $t$ , ten offenses are committed. The hazard  $h_0$  associated with  $F_0$  is modeled as a piecewise constant exponential function with mean  $\lambda_1 = 0.2$  for the first period and mean  $\lambda_2 = 0.01$  for all other periods. This captures the fact that a large share of crimes is reported on the day of the incident. 50% of the victims never report the crime to the police between periods 0 and 75. In period 75, an intervention  $D_1$  permanently decreases the reporting hazard by  $-0.5$  and increases the number of offenses committed to 12. In period 100, an intervention  $D_2$  permanently increases the reporting hazard by  $0.3$  and decreases the number of offenses committed

to 8. In period 125, an intervention does not affect victims and offenders. Unobserved heterogeneity is assumed gamma-distributed with variance 0.3. To capture the double-truncation scheme, I only keep observations for reported incidents between periods 50 and 200, which corresponds to roughly 800 reports out of 4,000 crimes per simulation. I simulate 1,000 datasets. Results are summarized in Panel B of Figure 8.4. Once again, the out-of-the-box MPH regression estimator is severely biased. In a dynamic setting, it does not capture null effects correctly anymore. On the other hand, the proposed estimator for a correctly specified promotion time model presents no bias in estimates (as measured by a t-test).

Given that the share of never-reporters in the first data period is a researcher's input, I also investigate the sensitivity of estimates to model misspecification in another set of simulations. I estimate additional models: a corrected MPH model that assumes all victims eventually report and several corrected promotion time models that assume various shares of never-reporters in the first data period. Results suggest undershooting the share of never-reporters at baseline leads to estimates of covariate effects biased towards zero (see Table C.1). The bias is relatively small in the Monte Carlo Simulations. However, one should not conclude that the plugin parameter  $\alpha$  has no impact on downstream empirical results. A model that assumes 0% of never-reporters largely underestimates the impact of interventions on the share of never-reporters. Conversely, a model that assumes 80% of never-reporters overestimates the impact of interventions. For example, for an intervention effect of 0.3, assuming a share of never-reporters of 50% (80%) implies  $0.5^{\exp(0.3)} \approx 39\%$  ( $0.8^{\exp(0.3)} \approx 74\%$ ) of never-reporters post-intervention.

## 6. Empirical Results

This section studies sex crime incidence and reporting between 2010 and 2019, with a focus on the Me Too movement's impact. I first examine the impact of #MeToo on the plaintiff reporting hazard (see Equation 1.4). I then investigate what these variations imply for the reporting rate of victims (see Equation 1.5). This also allows me to compute estimates of sex crime incidence over the period (see Equation 1.6). Finally, I isolate the effect of #MeToo on sex crime incidence.

### 6.1 Plaintiffs

I first estimate trends in the plaintiff reporting hazard over the period (see Equation 1.4). By abstracting from never-reporters, I focus on the observable part of the distribu-

tion of times to report  $Y$ . From the Monte Carlo simulations, we know that we cannot interpret the resulting hazard ratios as increases in the victim reporting hazard. These estimates are downward-biased if a share of victims never reports to the police. Thus, they may be understood as a lower bound for victims' propensity to report.

**Main Results** The dependent variable is the number of days elapsed between a sex crime being committed and its report to the police. The breaks in the piece-wise constant baseline hazard are set after 1, 30, 90, 180, and 365 days. I assume unobserved heterogeneity is gamma-distributed. My baseline specification is

$$h_{itc}^{(p)}(y) = h_0(y) \exp \left( \delta_c + \sum_{k=\text{Oct.15,2010}}^{\text{Oct.15,2019}} \beta_k \mathbb{1}(t + y \geq k) \right) \gamma_i. \quad (1.8)$$

$\delta_c$  is a city fixed effect that accounts for variations in reporting delays across cities.  $\gamma_i$  is a gamma-distributed unobserved heterogeneity term. For each year, the dummy variables take the value one from the 15<sup>th</sup> of October onwards. The main coefficients of interest are the yearly betas. I interpret  $\beta_k$  as the additional (higher or lower) propensity to report a sex crime to the police in year  $k$  among plaintiffs. Figure 8.5 presents the evolution of the hazard ratio over the decade.<sup>32</sup> Overall, the plaintiff reporting hazard appears flat until October 2017 (i.e., when the Me Too movement went viral on social media). It then increases by approximately 23% in the following two years (significant at all standard significance levels).<sup>33</sup> This finding confirms that #MeToo coincides with a large change in victim reporting behaviors. Overall, the Me Too movement appears to gain momentum, with no apparent decrease by the end of the study period. A closer inspection at the quarterly level confirms this observation (see Figure D.2), suggesting that #MeToo is not simply a one-point-in-time event that briefly increased the saliency of sexual violence in society. On the contrary, the movement coincides with a durable change in reporting norms.

<sup>32</sup>Figure D.1 presents the estimated baseline survival functions per city. For comparison, Figure C.1 also presents the evolution of the hazard ratio for a model that does not account for double-truncation. This naive model has unstable coefficient estimates that are very sensitive to the choice of the end of the study period. In all cases, it estimates that the reporting hazard increased from approximately 70% to 150%. This is, of course, mainly spurious but highlights the risks of failing to account for double-truncation.

<sup>33</sup>This indicates #MeToo is associated with a much larger increase in the reporting hazard than in previous periods. However, given that all these estimates are downward-biased, this should not be interpreted as strong evidence for the absence of pre-trends in the pre-treatment period. As we will see in the next section, once accounting for the share of never-reporters, there are sizable, linear, and positive pre-trends to the movement's mediatization.

**Heterogeneity** I then turn to heterogeneity analysis. To investigate #MeToo’s effects on incident-level characteristics, I focus on approximately 30,000 observations from the New York Police Department. This represents roughly one fourth of the total number of observations. One could worry about selection effects. However, the magnitude and sign of the unconditional #MeToo effect are extremely similar for this subsample than for the overall sample (see Table 8.2). My specification on this restricted sample is:

$$h_{itc}^{(p)}(y) = h_0(y) \exp \left( \delta_c + \zeta X_i + \phi \text{MeToo}_{ity} + \Omega X_i \times \text{MeToo}_{ity} + \Psi \text{Old}_{ity} \times \text{MeToo}_{ity} \right) \gamma_i. \quad (1.9)$$

$X_i$  is a vector of time-invariant incident characteristics. It includes the victim and suspect’s race and sex, a dummy variable for juveniles, and whether the crime is a felony or a misdemeanor. The Me Too movement went viral on social media on the 15<sup>th</sup> of October 2017. Thus,  $\text{MeToo}_{ity}$  is a dummy variable equal to one after this date. Finally,  $\text{Old}_{ity}$  is a dummy variable that takes value one if the incident occurred more than 365 days ago (and 0 otherwise). Figure 8.6 presents estimates of  $\Phi$ ,  $\Omega$ , and  $\Psi$ . The baseline #MeToo effect  $\Phi$  is for white women plaintiffs, filing a complaint against a white suspect for a sexual felony less than a year old. #MeToo has no statistically significant effect on this population, but it has a large and statistically significant effect on old crime incidents (i.e., more than a year old). The intervention appears particularly effective at encouraging victims of past crime incidents to report to the police. However, this is not to say that the movement had no effect on recent crime incidents for all crime types and socio-demographic groups. Misdemeanors are more likely to be reported than felonies at the 10% significance level. Juvenile, Hispanic, and black victims are also more responsive to the intervention. I find no statistically significant effect for male victims, black, Hispanic, and male suspects. Juvenile suspects are far less likely to be reported after #MeToo. This mirrors the increase in juvenile plaintiff reporting, who are more likely to file complaints against adult suspects. Overall, #MeToo had a larger, positive effect on the most vulnerable groups of the victim population, as well as on victims of misdemeanors and past crimes.

**Robustness** I conduct a series of robustness exercises (see Table 8.2). First, I assess the sensitivity of estimates to alternative specifications for unobserved heterogeneity. Columns 1, 2, and 3 present estimates of the #MeToo effect on the plaintiff reporting hazard respectively without unobserved heterogeneity, with a gamma-distributed random effect, and with a discrete random effect (à la Heckman and Singer 1984a). Overall, all specifications indicate a positive impact on the reporting hazard and point



estimates are very close to one another, suggesting the specification of unobserved heterogeneity is unlikely to heavily affect point estimates. Second, I show that my estimates are robust to the inclusion of linear and quadratic time-trends that account for pre-trends (see Columns 4 and 5). Third, I assess the sensitivity of estimates to different breaks in the baseline hazard. Columns 6 to 9 estimate the #MeToo effect on reporting for a broad range of breaks in the baseline hazard. Adding flexibility to the baseline hazard mildly decreases the size of the effect at first, but then point estimates remain remarkably similar past three breaks (and gains in the likelihood become negligible).<sup>34</sup> Fourth, the FBI changed its definition of rape in 2013. City-level police records rely on a different categorization of offenses, which should not affect my results. To remove any doubts, the #MeToo effect remains when I restrict the sample to crimes reported after 2013 (see Column 10). Fifth, I assess the sensitivity of estimates to control variables (see Columns 11 and 12). I find that controlling for all crime incident characteristics barely changes the estimated #MeToo effect. I also run my analysis on non-sexual crimes. The latter are not a particularly relevant control group, because they are generally reported with short delays to the police (see Table A.3). Nonetheless, I interpret resulting point estimates as placebo tests that may capture more general changes in police recording practices. The size of the effects is negligible and suggests that my results are unlikely driven by such dynamics (see Columns 13 to 15). Finally, to relax the proportional hazards assumption, I allow for a differential effect of the yearly dummies on recent crime incidents (of less than 365 days) and older crime incidents (of more than 365 days). Figure D.5 presents the results. I find empirical support for time-dependent effects (as the heterogeneity analysis already indicated). However, in Figure D.6, I show that focusing only on recent crime incidents does not revert the estimated trends in the plaintiff reporting hazard. If anything, it indicates a larger positive pre-trend since 2010.

## 6.2 Victims

Next, I investigate what these trends imply for sex crime reporting in the general population of victims (see Equation 1.5). I estimate the victim reporting hazard between 2010 and 2019. Consistent with estimates of the NCVS, I set  $\alpha$  so that the share of never-reporters in 2010 is approximately equal to 70%. I pool all cities and specify the hazard as follows:

$$h_{itc}^{(v)}(y) = f_0(y) \exp \left( \alpha + \delta_c + \sum_{k=\text{Oct.15,2010}}^{\text{Oct.15,2019}} \beta_k \mathbb{1}(t + y \geq k) \right) \quad (1.10)$$

<sup>34</sup>In practice, many victims report on the first day, so the break on day 1 is the break that matters to estimate reasonable baseline survival functions.

As in Equation 1.8, the main coefficients of interest are the yearly betas. I interpret  $\beta_k$  as the additional (higher or lower) propensity to report a sex crime to the police in year  $k$  among victims.<sup>35</sup> Note that Equation 1.10 also provides us with the total share of victims who will eventually report a sex crime to the police for a given year  $k$  (and in the absence of future events that may shift victim reporting behaviors).

**Main Results** Figure 8.7 decomposes reported sex crimes to the police into estimates of sex crime incidence and reporting (based on Equations 1.6 and 1.10). Accounting for never-reporters uncovers relatively large pre-trends before the Me Too movement. Estimates suggest the share of victims who eventually report increased from 30% in 2010 to 55% before October 2017. The Me Too movement coincides with a reinforcement of these broader trends as the share of reports reaches 75% in 2019. Thus, my estimates suggest that the reporting rate more than doubled during the decade. This also implies a large decrease in sex crime incidence over the same period. These substantial pre-trends partly cancel each other out in the time series of reported crimes and translate into an increase in reported sex crimes (aggregated at the incident date) of approximately 50%.

These results highlight the importance of separating crime incidence and reporting for empirical research on crime. As many have suggested since [Quêtelet \(1831\)](#), reported crime statistics are likely but the tip of the iceberg. Figure 8.7 also questions the context in which successful social movements emerge. Though the precise date on which #MeToo went viral was unforeseen, the movement also appears to take place in the context of a deeper societal change in sex crime reporting and incidence.

**Robustness** I conduct a series of robustness exercises (see Appendix D). First, I present estimated trends under different values for the share of never-reporters in 2010. The model rescales the observable part of the distribution of times to report  $Y$ . The choice of  $\alpha$  thus changes the magnitude of the effects uncovered, but not the underlying trends (see Figure D.4). I show that a broad range of reasonable parameter values for  $\alpha$  – ranging from 60 to 80% of never-reporters at baseline – leaves the main decomposition result qualitatively unchanged (see Figure D.3).<sup>36</sup> Second, given the presence of time-dependent effects, I assess my estimates' sensitivity to this specific

<sup>35</sup>The main difference is that I do not specify a random effect term for unobserved heterogeneity. This is because the promotion time model often fails to identify the presence of unobserved heterogeneity. Yet, it still recovers the simulated treatment effects with seemingly no bias in estimates in the Monte Carlo simulations. Furthermore, we know that a gamma-distributed unobserved heterogeneity barely changes the point estimate for the Me Too movement's impact on the plaintiff reporting hazard. Thus, I do not expect it to largely affect results.

<sup>36</sup>Note that this interval is very large, as it encompasses all estimates of the victim reporting rate by the National Crime Victimization Survey since 2011.

violation of proportional hazards. In Figure D.7, I show that focusing on recent crime incidents does not qualitatively change the estimated trends in sex crime incidence and reporting.

A remaining concern is that my main results assume truthful and founded crime reports. In practice, false allegations of a crime are a rampant concern for the criminal justice system – in particular when it comes to sex offenses. It is notoriously difficult to assess the incidence of false accusations.<sup>37</sup> Recent estimates suggest that baseless rape allegations represented approximately 5% of total rape charges in the United States between 2006 and 2010 (De Zutter, Horselenberg and van Koppen 2017). Though a precise estimate of such allegations is out of reach of researchers, one can still ponder their implications for interpreting my estimates. If the rate of false allegations is positively correlated to #MeToo, then the model will overestimate the movement's effects on the victim reporting rate. As a result, it will also inflate the size of its extrapolated deterrent effect. This is a plausible scenario, particularly if the expected benefits of filing a charge increase for plaintiffs after #MeToo.<sup>38</sup> To assess the impact of unfounded allegations on my estimates, I restrict my sample to sex crimes that resulted in an adult arrest (note that this information is only available for Los Angeles). These reports are more likely to have presented compelling evidence. Estimates from this seriously restricted sample can be understood as highly conservative estimates of sex crime incidence and reporting. Yet they are qualitatively similar to my main results for Los Angeles (see Figure D.8). This suggests unfounded allegations are not driving my results.

### 6.3 Offenders

The careful decomposition of reported sex crimes indicates that sex crime incidence has decreased over the period. I attempt to isolate the contribution of #MeToo to this trend. As #MeToo potentially affected all cities in the United States and worldwide, credible control groups for causal inference are limited. However, the crime literature suggests crime categories are subject to cyclical fluctuations, part of which has been explained by weather conditions, economic downturns, labor market conditions, alcohol consumption, and sports events (Markowitz 2005; Jacob, Lefgren and Moretti 2007). For these reasons, reported non-sexual crimes are a plausible control group. I thus

---

<sup>37</sup>Previous studies have found estimates ranging from 1.5% to 90% of total rape charges filed to the police are false (Rumney 2006). Despite this apparent uncertainty in estimates, a closer look finds that many studies are outdated, lack statistical power, have worked with heterogeneous definitions, and sometimes have questionable methodologies. Maclean (1979) deemed reports false if, for instance, the victim did not appear “dishevelled”, while Stewart (1981) works with a sample of 18 charges.

<sup>38</sup>Through larger financial compensations or higher probabilities of sentencing, for instance.

construct several counterfactuals for quarterly sex crime incidence based on quarterly non-sexual crime reports. For simplicity, my baseline empirical strategy is a difference-in-differences. The specification for crime  $i$ , in quarter  $t$ , in city  $c$  is:

$$\log(\text{Crimes})_{itc} = \beta \text{MeToo}_{it} + \delta_i + \delta_c + \delta_t + \varepsilon_{itc}. \quad (1.11)$$

$\delta_i$ ,  $\delta_t$ , and  $\delta_c$  are respectively crime, time, and city fixed effects.  $\text{MeToo}_{it}$  is a dummy variable that takes value one for sex crimes after October 2017 and thus  $\beta$  is the marginal effect of #MeToo on sex crimes. To construct valid confidence intervals, I sample with replacement observations from the dataset and estimate the duration model, sex crime incidence, and the counterfactuals at each iteration. Sex crime incidence is an extrapolation based on the duration model's output and the bootstrap procedure accounts for this additional source of randomness.

The difference-in-differences estimates may be interpreted causally under the assumption that reported non-sexual crimes and sex crimes would have followed similar trends absent the Me Too movement's sudden mediatization. Despite some noise in the data, the inspection of pre-trends suggests this is a plausible assumption (see Figure 8.8). In addition, given that #MeToo was specifically related to sex crimes, I assume that the reporting rate of non-sexual crimes is uncorrelated to the timing of #MeToo.

**Main Results** Table 8.2 presents the main results. Across specifications, I find a strong, negative, statistically significant decrease in sex crime incidence after #MeToo ( $\approx -23\%$ , see Column 1). I compare my results to estimates to using reported sex crimes as a treated unit (instead of my estimates of sex crime incidence). I use incident dates (see Column 8) or report dates (see Column 9) to aggregate crime reports. When studying incident dates with no correction, one also finds a positive yet statistically insignificant increase in sexual crime reports ( $+20\%$ ). When studying report dates, one finds a large increase in sexual crime reports ( $+32\%$ ). The latter is statistically significant at all standard significance levels. Both estimates have been previously interpreted as an increase in the reporting rate of victims (Levy and Mattsson 2021).

This warrants three remarks. First, in the presence of delayed reports, estimates are sensitive to the date used for aggregate reported crime statistics (i.e., the incident date or of its report). In our empirical context, the size of the effect and its statistical significance vary drastically. Second, one cannot interpret these estimates as an increase in reporting behaviors without making additional, strong, and unverifiable assumptions. Third, if the Me Too movement simultaneously increases victim reporting and decreases sex crime incidence, then both estimates will underestimate the movement's

impact as a whole. According to my estimates, this is clearly the case.

**Robustness** I conduct a series of robustness exercises. First, as presented in Table 8.2, my results are robust to alternative counterfactual models that relax the parallel trends assumption. This includes an interactive fixed effects (IFE) model (Xu 2017) and the matrix completion method (Athey et al. 2021). I find that the IFE model with five factors best fits the pre-treatment period (see Figure D.9). In practice, however, both alternative models barely change point estimates of the movement’s effects on sex crimes (see Columns 2 and 3). Second, I sequentially drop each non-sexual crime used for the counterfactual. In all cases, a strong, negative, statistically significant effect on sex crime incidence remains (see Columns 4 to 7). Third, I replace sex crime incidence as the treated unit with one of the non-sexual crimes used for the counterfactual. For all placebo crimes, I find no statistically significant effect of the Me Too movement (see Columns 10 to 13).

## 7. Discussion and Mechanisms

### 7.1 Social Norms

The recurring public debate over the incidence of false allegations is a reminder that the decision to report a crime to the police is also influenced by social norms and beliefs. Victims may incur social costs for reporting a sex crime to the police in at least two ways. First, libeler narratives can undermine the credibility of the charges and question the motives behind them. In turn, beliefs on the incidence of sex crimes and libelers in one’s society may weigh in a victim’s decision.<sup>39</sup> Second, social conformity concerns may also influence a victim’s decision. If victims care about what other victims do, then a repeated coordination game may easily result in persistent, sub-optimal equilibria.

Unfortunately, there is no systematic empirical evidence of the discourse surrounding sex crimes, and the attitudes of the broader public on the matter.<sup>40</sup> As a proxy for these narratives, I collect and analyze tweets mentioning sex crimes between 2010 and 2019. I present the time-series of tweet counts (in logs) in Figure A.7. The hashtag #MeToo was virtually never used before October 2017 and may lead us to believe the discourse surrounding sex crimes was also very limited (see Panel A). In fact, the

<sup>39</sup>Given the considerable uncertainty on the matter, these beliefs need not be true or unbiased.

<sup>40</sup>To the best of my knowledge, the Views of the Electorate Research (VOTER) survey is the only national survey with explicit questions on sex criminality, but they only ask respondents about sexual harassment (which I do not study in this paper).

topic was already discussed on Twitter its share of the total number of tweets has been increasing since 2010 (see Panel B). Similar observations are made with Google Trends data (see Figure A.8). This suggestive evidence is consistent with the pre-trends in crime reporting uncovered in Section 6.

## 7.2 Odds of Punishment

Figure A.3 compares the evolution of arrest rates for sex crimes to non-sexual crimes between 2010 and 2019 for New York City and Los Angeles.<sup>41</sup> Overall, arrest rates remain relatively stable over the period, with a slight decrease for sexual offenses.

Combining arrest rates with my estimates of sex crime incidence, I also compute the unconditional probability of arrest for committing a sex crime:

$$P_t(\text{Arrest} = 1) = P_t(\text{Arrest} = 1 \mid \text{Report} = 1) \times \hat{F}_t^{(p)}(+\infty \mid \chi_t). \quad (1.12)$$

Figure 8.9 presents the evolution of the probability of arrest for sex offenders over the period. It increases from 16% to 37% between 2010 and 2019. The decrease in sex crime incidence is thus consistent with a Beckerian argument. Between 2010 and 2019, a one percentage point increase in the probability of arrest is associated with a 0.9 percentage point decrease in sex crime incidence. These findings suggest that sex offenders react to the odds of punishment, and thus that increasing the probability of arrest through increased reporting may effectively prevent future offenses.

## 8. Conclusion

In the past two centuries, underreporting has been a major empirical challenge in making sense of reported crime statistics. This paper proposes a methodology to separate sex crime incidence and reporting. My empirical strategy leverages the largely understudied presence of delayed reports in police records. The latter raises new empirical issues but also allows researchers to work with tools from survival analysis to study variations in the reporting hazard over time.

I then study sex crime incidence and reporting surrounding the Me Too movement. I present three key results on its origins and impact. First, #MeToo largely and persistently increased the propensity to report of victims. The effect is larger for juveniles,

---

<sup>41</sup>Seattle and Cincinnati do not provide arrest data for fine-grained categories and only report arrest data for crimes in the Uniform Crime Reporting program.

racial minorities, victims of misdemeanors and past crime incidents. Second, it had a deterrent effect on sexual offenders. Third, the movement appears in a general context of decreasing sex crime rates and increasing sex crime reporting. These substantial trends partly cancel each other out and are less apparent in the time series of reported crimes.

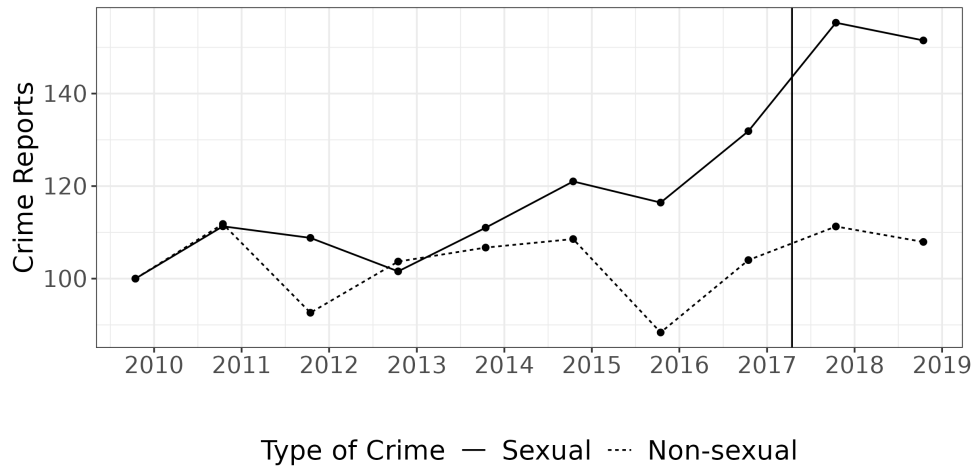
This last finding highlights the importance of disentangling crime incidence and reporting in police data. Though many competing explanations may rationalize my estimates, they are unlikely to be driven by false allegations and changes in the recording guidelines of law enforcement agencies. I present suggestive evidence that rapidly changing social norms increased the reporting rate of victims and, ultimately, the likelihood of arrest for sex offenders.

My results suggest that increasing the reporting rate of victims may significantly deter criminal behavior. They further highlight that social norms can successfully enforce socially desirable behaviors when the legal system fails to do so on its own. The broader legal, political, and economic consequences of the Me Too movement remain unknown and represent an opportunity for future research.

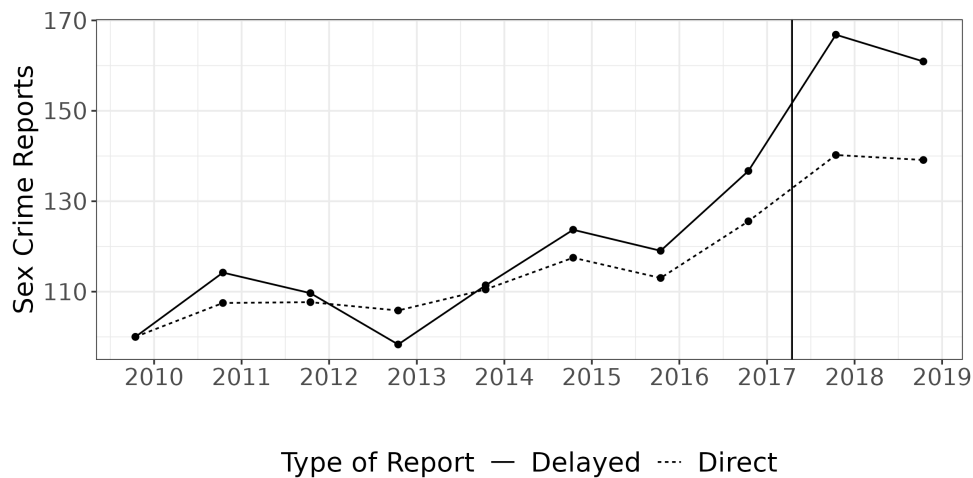
## Figures and Tables

Figure 8.1: Trends in Reported Crimes

### A. Sexual / Non-Sexual Crime Reports



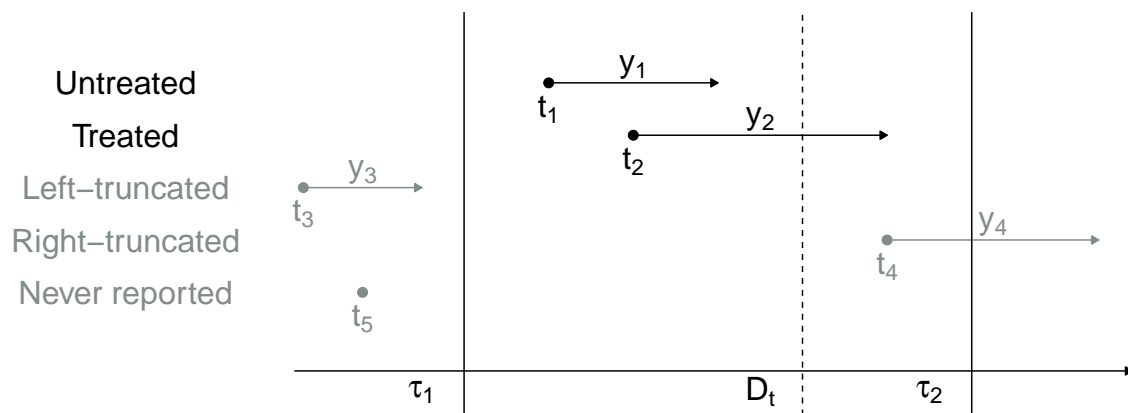
### B. Direct / Delayed Sex Crime Reports



**Notes:** This figure presents trends in reported crimes. Crime reports are aggregated on the report date. Crimes reported on the first day are labeled as a direct report. The vertical solid line is set one period before #MeToo (Oct 2017). Panel A compares sexual crime reports to non-sexual crime reports. Panel B compares direct sexual crime reports to delayed sexual crime reports. A report is labeled as a direct report if it is reported on the day of the incident.

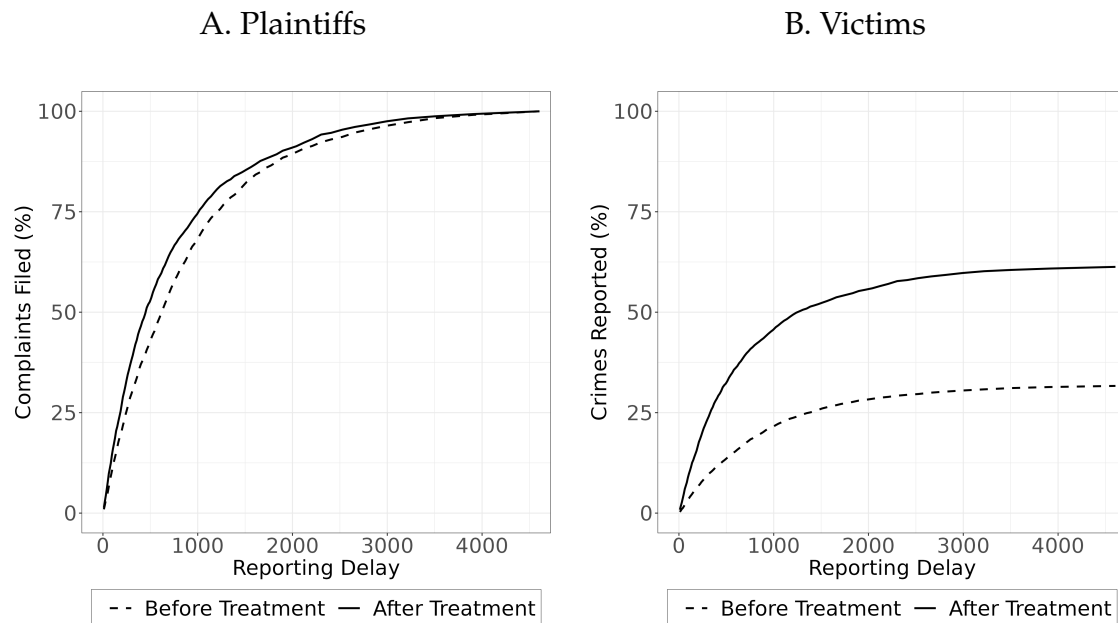


Figure 8.2: The Structure of (Delayed) Crime Reports



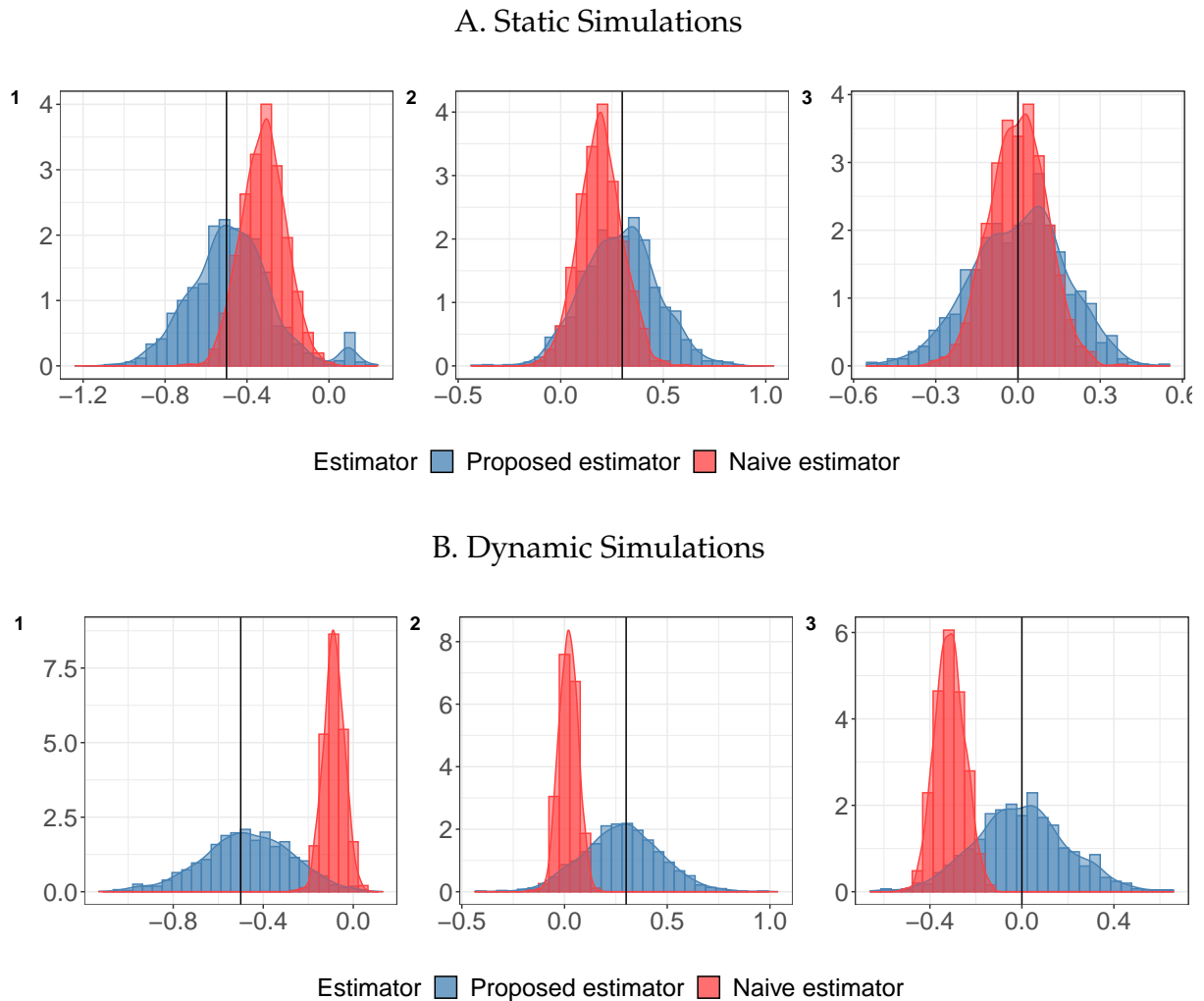
**Notes:** A graphical depiction of time to report data based on police records. The study window goes from  $\tau_1$  to  $\tau_2$  (solid vertical lines), with an intervention  $D_t$  in between represented by a dashed vertical line (e.g. #MeToo). Elements in gray are unobserved. Elements in black are observed. Some plaintiffs will have reported before the intervention's implementation and form the control group (non-treated observations). Others will be potentially affected by the intervention and form the treated group (treated observations). Some plaintiffs have reported before the start of the study and are unobserved (left-truncated data points). Some plaintiffs have not yet reported a crime to the police by the end of the study period but will in the future and are unobserved (right-truncated data points). Finally, a fraction of victims decide to never report and are never observed (i.e., never-reporters). Victims who never report are modeled as never-reporters (but note that this quantity is not identifiable from police records).

Figure 8.3: From Observed Delays to Victim Reporting Rates



**Notes:** These simulated distributions provide graphical intuition on the duration models developed in the paper. A treatment/intervention increases the propensity to report of victims. The plots compare cumulative distribution functions (CDFs) before (dashed line) and after (solid line) the treatment. Panel A plots the CDF of reporting delays among plaintiffs. The researcher observes this in police records. It corresponds to Equation 1.4. Panel B plots the CDF of the entire victim population and corresponds to Equation 1.5. The CDFs do not sum to 100% because a share of victims will never report to the police. Those victims are unobserved. However, suppose one knows the victim reporting rate before the treatment (here 30%). In that case, the observed distribution of reporting delays is sufficient to infer the victim reporting rate after the treatment (here around 60%). This intuition can be generalized to multiple treatments/interventions. For instance, fitting a linear spline over time recovers variations of the reporting rate over time (see the application to #MeToo in Section 6). Note also that ignoring never-reporters (Panel A) tends to underestimate the differences between CDFs in the victim population (Panel B), hence the bias towards zero when setting too low a share of never-reporters at baseline (relative to the 'true' share).

Figure 8.4: Monte Carlo Simulations



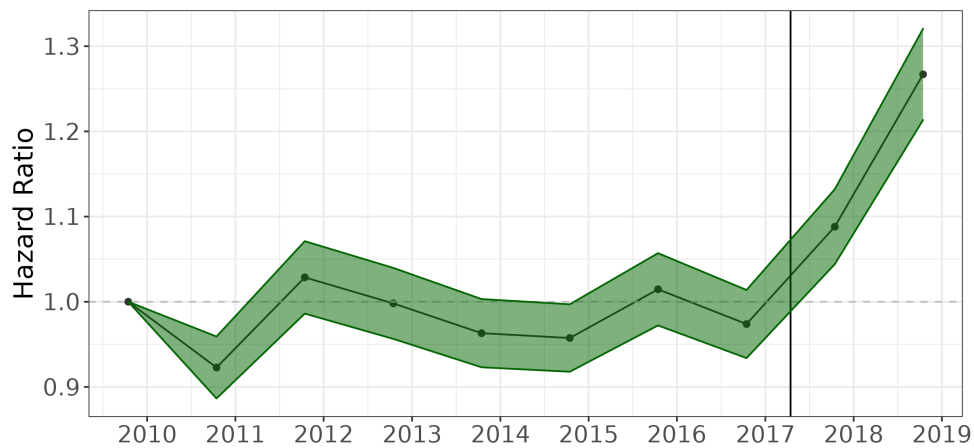
**Notes:** Results of 1000 Monte Carlo simulations. Panel A simulates datasets with time-invariant covariates and double-truncation (based on Equation 1.4). The blue densities present the distribution of estimates with a likelihood that appropriately accounts for double-truncation. The red densities present estimates of the out-of-the-box MPH model as implemented in the R package *frailtyEM* (Balan and Puter 2019). This implementation does not account for double-truncation in the data. Subpanel 1 is for the first intervention with effect -0.5. Subpanel 2 is for the second intervention with effect 0.3. Subpanel 3 is for the third intervention with no effect. Each panel's solid vertical black line is the 'true' parameter value. The results indicate the proposed likelihood correctly accounts for double-truncation in the data by weighting observations by the inverse of their sampling probability. Panel B simulates more realistic datasets of police complaints with time-varying covariates, unobserved heterogeneity, never-reporters, and double-truncation (based on Equation 1.5). In both panels, the densities of the proposed estimator are centered around the true parameter value, whereas the naive estimator is biased.

Table 8.1: #MeToo Effects on Crime Reporting – Lower Bounds and Robustness

Plaintiff Hazard	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Sex Crimes						Assaults						Robberies	Burglaries	
#MeToo: October 15, 2017	0.154 (0.009)	0.169 (0.011)	0.11 (0.01)	0.13 (0.012)	0.10 (0.017)	0.224 (0.012)	0.170 (0.011)	0.166 (0.011)	0.166 (0.011)	0.135 (0.013)	0.196 (0.020)	0.196 (0.019)	0.0018 (0.011)	0.03 (0.011)	0.05 (0.011)
<b>Baseline Hazard</b>															
Day 0+	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Day 1+	X	X		X	X	X	X	X	X	X	X	X	X	X	X
Day 30+	X	X		X	X		X	X	X	X	X	X	X	X	X
Day 90+	X	X		X	X		X	X	X	X	X	X	X	X	X
Day 180+	X	X		X	X		X	X	X	X	X	X	X	X	X
Day 365+	X	X		X	X		X	X	X	X	X	X	X	X	X
<b>Controls</b>															
Crime Category															
Victim Characteristics															
Suspect Characteristics															
City Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<b>Time-Trends</b>															
Linear				X	X										
Quadratic					X										
<b>Unobserved Heterogeneity</b>															
$\tau_1$	2010	2010	2010	2010	2010	2010	2010	2010	2010	2010	2010	2010	2010	2010	2010
$\tau_2$	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019
Cities	All	All	All	All	All	All	All	All	All	All	NYC	NYC	All	All	All
N Observations	111869	111869	111869	111869	111869	111869	111869	111869	111869	72730	32442	32442	50000	50000	50000
Algorithm	BFGS	BFGS	rgenoud	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS
Log-Likelihood	-419895	-416769.5	-421380	-419889	-419886	-420521	-417059	-416848	-416811	-173196	-121129	-119837	-78038	-68458	-108183

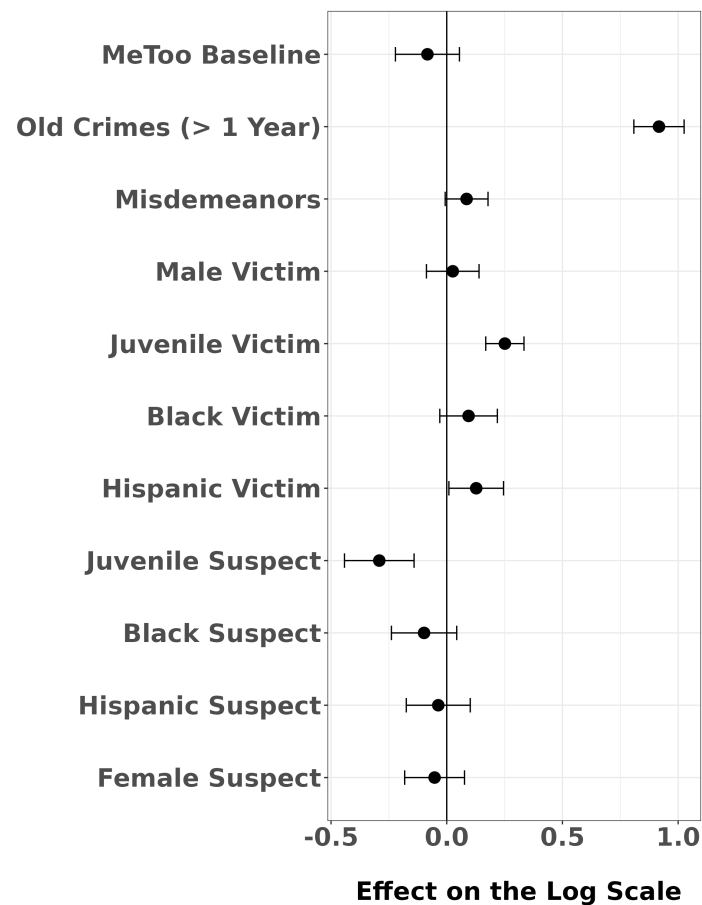
**Notes:** Effect of the Me Too movement's sudden mediatization in Oct. 2017 on the plaintiff reporting hazard (see Equation 1.4) for various specifications and samples. Estimates are presented on the log scale. I focus on plaintiffs and abstract from never-reporters. Thus, the estimates reported in this table are lower bounds for the #MeToo effect on crime reporting. Standard errors are in parentheses. Column 1 presents the baseline estimate. Columns 2 and 3 add a random effect in the estimation to account for time-invariant unobserved heterogeneity. Columns 4 and 5 respectively account for linear and quadratic time-trends in reporting. Columns 6 to 9 sequentially increase the number of breaks in the baseline hazard. Column 10 restricts the study period to 2014-2019. Column 11 presents estimates for New York City. Column 12 controls for incident-level characteristics for New York City. Overall, the effect size is very stable across specifications. Columns 13 to 15 study the hazard of plaintiffs for non-sexual crimes and are interpreted as placebo tests (estimates are based on a random sample of 50,000 incidents). Overall, effect sizes are much smaller than for sex crimes.

Figure 8.5: Yearly Hazard Ratios – Plaintiffs



**Notes:** Yearly estimates of the plaintiff reporting hazard based on Equation 1.8. I focus on plaintiffs and abstract from never-reporters. Thus, the estimates reported in this table are lower bounds for the yearly marginal effects on crime reporting. Breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. The likelihood appropriately corrects for double-truncation. Unobserved heterogeneity is assumed gamma-distributed. 95% confidence intervals (in grey) are computed with a numerical estimation of the hessian combined with the delta method. The vertical solid black line corresponds to the Me Too movement's mediatization.

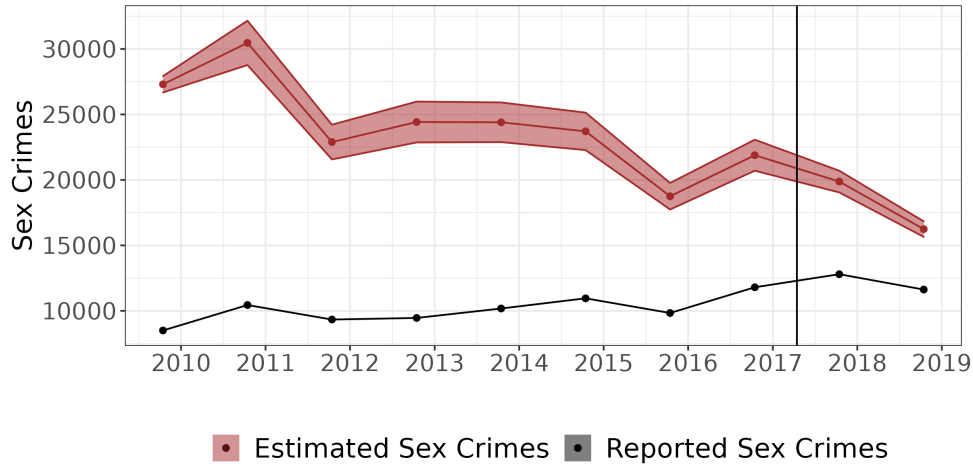
Figure 8.6: #MeToo Effects on Crime Reporting – Heterogeneity Analysis



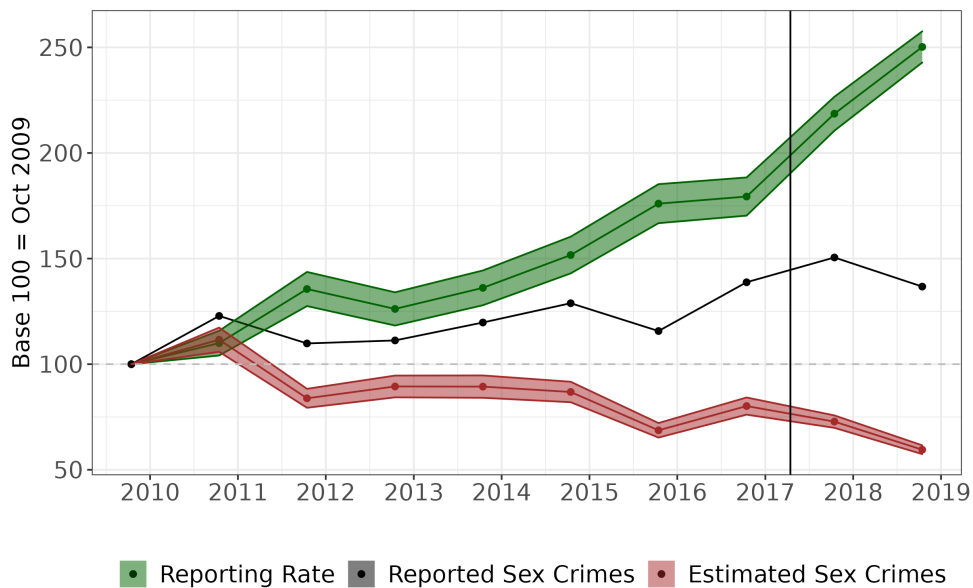
**Notes:** Estimates of the baseline #MeToo effect and additional interaction effects (based on Equation 1.9). The baseline #MeToo effect is for white women plaintiffs, filing a complaint against a white male suspect, for a sexual felony less than a year old. The vertical solid black line corresponds to a null effect. Breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. The likelihood appropriately corrects for double-truncation. Unobserved heterogeneity is assumed gamma-distributed. Estimates are presented on the log scale. 95% confidence intervals.

Figure 8.7: Trends in Sex Crime Reporting and Incidence

## A. Nominal Values



## B. Base 100 = Oct. 15, 2009



**Notes:** Panel A presents estimates of sex crime incidence relative to the observed, reported sex crimes to the police over the period. Panel B decomposes reported sex crimes into an extensive margin (crime reporting) and an intensive margin (crime incidence) based on Equations 1.5 and 1.6. The green line are yearly estimates of the reporting rate of victims of sex crimes. It rescales Figure 8.5 to account for never-reporters. For the baseline specification, I assume 30% of victims would have eventually reported sex crimes committed in 2010 (absent future interventions such as #MeToo). The black line are yearly reported sex crimes. The red line are yearly estimates of sex crime incidence. To obtain these results, breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. The likelihood appropriately corrects for double-truncation. No unobserved heterogeneity. 95% confidence intervals are constructed with a bootstrap procedure and 200 iterations. The vertical solid black line corresponds to the Me Too movement's mediatization.

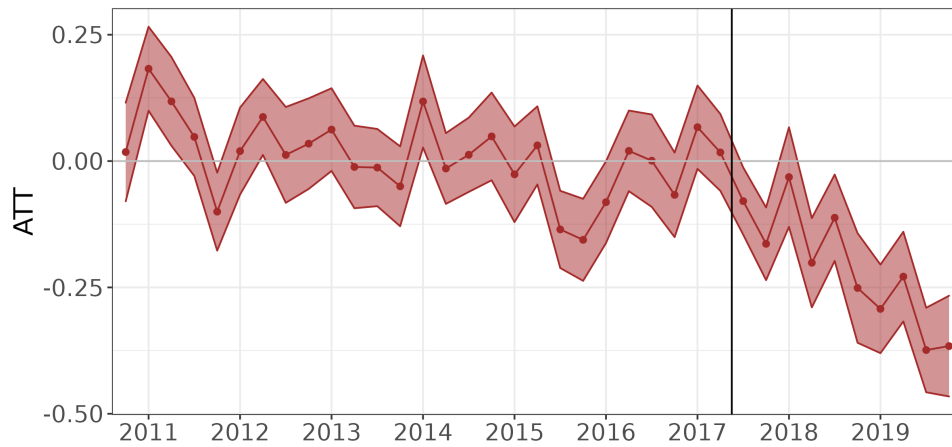
Table 8.2: #MeToo Effect on Sex Crime Incidence – Main Results and Robustness

Treated Group	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Estimated Sex Crimes			Reported Sex Crimes (Incident Date)			Reported Sex Crimes (Report Date)			Murders	Assaults	Robberies	Burglaries
#MeToo: October 15, 2017	-0.23 (0.02)	-0.23 (0.03)	-0.21 (0.015)	-0.21 (0.016)	-0.19 (0.02)	-0.27 (0.02)	-0.26 (0.02)	0.20 (0.105)	0.32 (0.10)	0.11 (0.15)	0.17 (0.12)	-0.16 (0.19)	-0.13 (0.17)
<b>Model</b>	DID	IFE	MC	DID	DID	DID	DID	DID	DID	DID	DID	DID	DID
<b>Fixed Effects</b>													
City Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X	X
Time Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X	X
Crime Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X	X
<b>Control Groups</b>													
Murders	X	X	X	X	X	X	X	X	X	X	X	X	X
Assaults	X	X	X	X	X	X	X	X	X	X	X	X	X
Robberies	X	X	X	X	X	X	X	X	X	X	X	X	X
Burglaries	X	X	X	X	X	X	X	X	X	X	X	X	X
Standard Errors	Bootstrap	Bootstrap	Bootstrap	Bootstrap	Bootstrap	Bootstrap	Bootstrap	Clustered	Clustered	Clustered	Clustered	Clustered	Clustered
N Observations	740	740	740	592	592	592	592	740	740	592	592	592	592

**Notes:** This table presents the Me Too movement’s effects on sex crimes for various specifications. The focus is on the Average Treatment Effect on the Treated (ATT). Dependent variables are on the log scale. The panel data is aggregated at the quarterly level. Standard errors are in parentheses. When the dependent variable involves estimated sex crimes (based on Equation 1.6), I rely on a bootstrap procedure with 200 iterations. Otherwise, standard errors are computed analytically and clustered at the city, crime, and period levels. Column 1 presents the baseline #MeToo effect on sex crime incidence using a difference-in-differences (DID) model. Columns 2 and 3 use alternative counterfactual models that relax the parallel trends assumption. Column 2 relies on the interactive fixed effects (IFE) model (Xu 2017) and Column 3 on the Matrix Completion (MC) method (Athey et al. 2021). Columns 4 to 7 sequentially drop one non-sexual crime from the control group. Column 8 presents the observed effect on reported crimes to the police (aggregated at the incident date). Column 9 does the same exercise but aggregates crime reports at the report date. Columns 10 to 14 sequentially use reported non-sexual crimes as the treated unit and are interpreted as placebo tests.

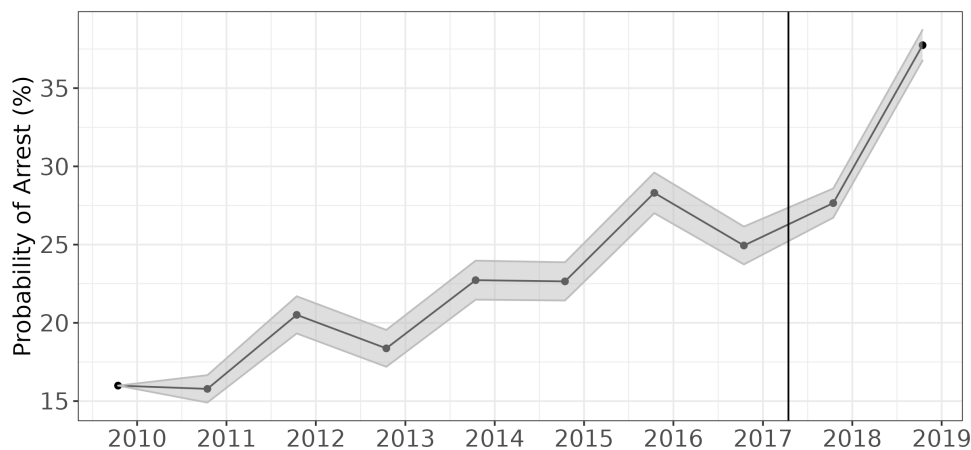


Figure 8.8: #MeToo Effect on Sex Crime Incidence – Event Study



**Notes:** Quarterly estimates of the Average Treatment for the Treated (ATT). The main results are presented in Table 8.2. I present here the results from a standard two-way fixed effects event-study. 95% confidence intervals are constructed with a bootstrap procedure and 200 iterations. The vertical solid black line corresponds to the Me Too movement's mediatization. For other counterfactual models that relax the parallel trends assumption, see Figure D.9.

Figure 8.9: Probability of Arrest for Sex Offenders



**Notes:** Yearly estimates of the probability of arrest for sex offenders based on Equation 1.12 for New York City and Los Angeles. Breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. The likelihood appropriately corrects for double-truncation. No unobserved heterogeneity. 95% confidence intervals (in grey) are constructed with a bootstrap procedure and 200 iterations. The vertical solid black line corresponds to the Me Too movement's mediatization.

# Bibliography

- Abbring, Jaap H, and Gerard J Van den Berg.** 2003. "The nonparametric identification of treatment effects in duration models." *Econometrica*, 71(5): 1491–1517.
- Abbring, Jaap H, and Gerard J Van Den Berg.** 2007. "The unobserved heterogeneity distribution in duration analysis." *Biometrika*, 94(1): 87–99.
- Acemoglu, Daron, and Matthew O Jackson.** 2017. "Social norms and the enforcement of laws." *Journal of the European Economic Association*, 15(2): 245–295.
- Aizer, Anna.** 2010. "The gender wage gap and domestic violence." *American Economic Review*, 100(4): 1847–59.
- Akerlof, George, and Janet L Yellen.** 1994. *Gang behavior, law enforcement, and community values*. Canadian Institute for Advanced Research Washington, DC.
- Amico, Maïlis, and Ingrid Van Keilegom.** 2018. "Cure models in survival analysis." *Annual Review of Statistics and Its Application*, 5: 311–342.
- AP News.** 2017. "The #MeToo Moment: When the Blinders Come Off."
- AP News.** 2021. "Left out of MeToo: New initiative focuses on Black survivors."
- Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi.** 2021. "Matrix completion methods for causal panel data models." *Journal of the American Statistical Association*, 1–41.
- Balan, Theodor Adrian, and Hein Putter.** 2019. "frailtyEM: An R package for estimating semiparametric shared frailty models." *Journal of Statistical Software*, 90: 1–29.
- Batut, Cyprien, Caroline Coly, and Sarah Schneider-Strawczynski.** 2021. "It's a man's world: culture of abuse,# MeToo and worker flows."
- Becker, Gary S.** 1968. "Crime and punishment: An economic approach." In *The economic dimensions of crime*. 13–68. Springer.
- Bellégo, Christophe, and Joeffrey Drouard.** 2019. "Does It Pay to Fight Crime? Evidence From the Pacification of Slums in Rio de Janeiro."
- Benabou, Roland, and Jean Tirole.** 2011. "Laws and norms." National Bureau of Economic Research.
- Berkeley Law.** 2019. "#MeToo—A Watershed Moment."

- Berkowitz, Daniel, Katharina Pistor, and Jean-Francois Richard.** 2003. "Economic development, legality, and the transplant effect." *European economic review*, 47(1): 165–195.
- Bernabe, Angelique.** 2021. "The Impact of News Coverage on Women's Labour Market Decisions: Evidence from the MeToo Movement."
- Borelli-Kjaer, Mads, Laurids Moehl Schack, and Ulf Nielsson.** 2021. "# MeToo: Sexual harassment and company value." *Journal of Corporate Finance*, 67: 101875.
- Bottan, Nicolas L, and Ricardo Perez-Truglia.** 2015. "Losing my religion: The effects of religious scandals on religious participation and charitable giving." *Journal of Public Economics*, 129: 106–119.
- Chalfin, Aaron, and Justin McCrary.** 2017. "Criminal deterrence: A review of the literature." *Journal of Economic Literature*, 55(1): 5–48.
- Cheng, Ing-Haw, and Alice Hsiaw.** 2020. "Reporting sexual misconduct in the # MeToo era." *Tuck School of Business Working Paper*, , (3506936).
- Cici, Gjergji, Mario Hendriock, Stefan Jaspersen, and Alexander Kempf.** 2021. "# MeToo meets the mutual fund industry: Productivity effects of sexual harassment." *Finance Research Letters*, 40: 101687.
- Coleman, Clive, and Jenny Moynihan.** 1996. *Understanding crime data: Haunted by the dark figure*. Vol. 120, Open University Press Buckingham.
- Cox, David R.** 1972. "Regression models and life-tables." *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2): 187–202.
- De Zutter, AWEA, Robert Horselenberg, and Peter J van Koppen.** 2017. "The prevalence of false allegations of rape in the United States from 2006-2010." *Journal of Forensic Psychology*, 2(2): 1–5.
- Doleac, Jennifer L.** 2019. "Encouraging desistance from crime." Working paper.
- Dörre, Achim.** 2020. "Bayesian estimation of a lifetime distribution under double truncation caused by time-restricted data collection." *Statistical Papers*, 61(3): 945–965.
- Dörre, Achim, and Takeshi Emura.** 2019. *Analysis of Doubly Truncated Data: An Introduction*. Springer.
- Dyck, Alexander, Adair Morse, and Luigi Zingales.** 2010. "Who blows the whistle on corporate fraud?" *The journal of finance*, 65(6): 2213–2253.
- Efron, Bradley, and Vahe Petrosian.** 1999. "Nonparametric methods for doubly truncated data." *Journal of the American Statistical Association*, 94(447): 824–834.
- Elbers, Chris, and Geert Ridder.** 1982. "True and spurious duration dependence: The identifiability of the proportional hazard model." *The Review of Economic Studies*, 49(3): 403–409.
- Emura, T, and K Murotani.** 2015. "An algorithm for estimating survival under a copula-based dependent truncation model." *Test*, 24(4): 734–751.

- Forbes.** 2020. "The Dark Side Of #MeToo: What Happens When Men Are Falsely Accused."
- Gertsberg, Marina.** 2022. "The Unintended Consequences of# MeToo-Evidence from Research Collaborations." *Available at SSRN 4105976.*
- Harvard Business Review.** 2019. "The #MeToo Backlash."
- Hay, Jonathan R, and Andrei Shleifer.** 1998. "Private enforcement of public laws: A theory of legal reform." *The American Economic Review*, 88(2): 398–403.
- Heckman, James, and Burton Singer.** 1984a. "The identifiability of the proportional hazard model." *The Review of Economic Studies*, 51(2): 231–241.
- Heckman, James, and Burton Singer.** 1984b. "A method for minimizing the impact of distributional assumptions in econometric models for duration data." *Econometrica: Journal of the Econometric Society*, 271–320.
- Iyer, Lakshmi, Anandi Mani, Prachi Mishra, and Petia Topalova.** 2012. "The power of political voice: women's political representation and crime in India." *American Economic Journal: Applied Economics*, 4(4): 165–93.
- Jacob, Brian, Lars Lefgren, and Enrico Moretti.** 2007. "The dynamics of criminal behavior evidence from weather shocks." *Journal of Human resources*, 42(3): 489–527.
- Kaplan, J.** 2019. "Jacob Kaplan's concatenated files: Uniform Crime Reporting (UCR) program data: Supplementary homicide reports, 1976-2018." *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]*, 07–15.
- Kaplan, J.** 2021. "Jacob kaplan's concatenated files: Uniform crime reporting program data: Offenses known and clearances by arrest (return a), 1960-2020." *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]*. <https://doi.org/10.3886/E100707V17>.
- Klemmer, Konstantin, Daniel B Neill, and Stephen A Jarvis.** 2021. "Understanding spatial patterns in rape reporting delays." *Royal Society open science*, 8(2): 201795.
- LAGAKOS, Stephen W, Leila M BARRAJ, and V de Gruttola.** 1988. "Nonparametric analysis of truncated survival data, with application to AIDS." *Biometrika*, 75(3): 515–523.
- Lee, Frances Xu, and Wing Suen.** 2019. "Credibility of crime allegations." *American Economic Journal-Microeconomics*, forthcoming.
- Levitt, Steven D.** 1998. "The relationship between crime reporting and police: Implications for the use of uniform crime reports." *Journal of Quantitative Criminology*, 14(1): 61–81.
- Levy, Roee, and Martin Mattsson.** 2021. "The effects of social movements: Evidence from# MeToo." *Available at SSRN 3496903.*
- Luo, Hong, and Laurina Zhang.** 2022. "Scandal, social movement, and change: Evidence from# MeToo in Hollywood." *Management Science*, 68(2): 1278–1296.

- Maclean, N.** 1979. "Rape and false accusations of rape." *Police Surgeon*, 15: 29–40.
- Mandel, Micha, Jacobo de Uña-Álvarez, David K Simon, and Rebecca A Betensky.** 2018. "Inverse probability weighted Cox regression for doubly truncated data." *Biometrics*, 74(2): 481–487.
- Markowitz, Sara.** 2005. "Alcohol, drugs and violent crime." *International Review of Law and economics*, 25(1): 20–44.
- Martin, Emily C, and Rebecca A Betensky.** 2005. "Testing quasi-independence of failure and truncation times via conditional Kendall's tau." *Journal of the American Statistical Association*, 100(470): 484–492.
- Mathur, Aparna, Abdul Munasib, Devesh Roy, Akshay Bhatnagar, et al.** 2019. "Sparking the #MeToo revolution in India: The 'Nirbhaya' case in Delhi." American Enterprise Institute.
- McDougal, Lotus, Samuel Krumholz, Nandita Bhan, Prashant Bharadwaj, and Anita Raj.** 2021. "Releasing the tide: how has a shock to the acceptability of gender-based sexual violence affected rape reporting to police in India?" *Journal of interpersonal violence*, 36(11-12): NP5921–NP5943.
- Miller, Amalia R, and Carmit Segal.** 2019. "Do female officers improve law enforcement quality? Effects on crime reporting and domestic violence." *The Review of Economic Studies*, 86(5): 2220–2247.
- Moreira, Carla, and Jacobo de Una-Alvarez.** 2010. "Bootstrapping the NPMLE for doubly truncated data." *Journal of Nonparametric Statistics*, 22(5): 567–583.
- Morgan, Rachel E, and Alexandra Thompson.** 2021. "Criminal victimization, 2020." Washington, DC: National Crime Victimization Survey, Bureau of Justice Statistics. Retrieved Jan, 4: 2022.
- Nagin, Daniel S.** 2013. "Deterrence: A review of the evidence by a criminologist for economists." *Annu. Rev. Econ.*, 5(1): 83–105.
- New York Post.** 2020. "'Being wrongly #MeToo'd has ruined my life'."
- New York Times.** 2017a. "The #MeToo Moment: After Alabama, Black Women Wonder, What's Next?"
- New York Times.** 2017b. "The #MeToo Moment: Blue-Collar Women Ask, 'What About Us?'"
- Onwuachi-Willig, Angela.** 2018. "What about #UsToo: the invisibility of race in the #MeToo movement." *Yale LJF*, 128: 105.
- Planty, Michael, Lynn Langton, Christopher Krebs, Marcus Berzofsky, and Hope Smiley-McDonald.** 2013. *Female victims of sexual violence, 1994-2010*. US Department of Justice, Office of Justice Programs, Bureau of Justice . . . .
- Police Data Initiative.** 2021. "U.S. City-level Police Records."
- Psychology Today.** 2017. "#MeToo: A Watershed Moment."

- Quêtelet, Adolphe.** 1831. *Research on the Propensity for Crime at Different Ages.*
- Rennert, Lior, and Sharon X Xie.** 2018. "Cox regression model with doubly truncated data." *Biometrics*, 74(2): 725–733.
- Rotenberg, Cristine, and Adam Cotter.** 2018. "Police-reported sexual assaults in Canada before and after# MeToo, 2016/2017." *Juristat*, 38(1).
- Rumney, Philip NS.** 2006. "False allegations of rape." *The Cambridge Law Journal*, 65(1): 128–158.
- Sahay, Abhilasha.** 2021. "The Silenced Women."
- Seaman, Shaun R, Anne Presanis, and Christopher Jackson.** 2021. "Estimating a time-to-event distribution from right-truncated data in an epidemic: a review of methods." *Statistical Methods in Medical Research*, 09622802211023955.
- Shen, Pao-sheng.** 2010. "Nonparametric analysis of doubly truncated data." *Annals of the Institute of Statistical Mathematics*, 62(5): 835–853.
- Sophie Calder-Wang, Paul Gompers, and Patrick Sweeney.** 2021. "Venture Capital's "Me Too" Moment." *NBER working paper*.
- Stephens-Davidowitz, Seth.** 2013. "Unreported victims of an economic downturn." *Unpublished paper, Harvard University, Department of Economics, Cambridge, MA.*
- Stewart, CH.** 1981. "A retrospective survey of alleged sexual assault cases." *Police Surgeon*, 28(32).
- Tavarez, Lahiz P.** 2021. "Waiting to Tell: Factors Associated with Delays in Reporting Sexual Violence."
- Vakulenko-Lagun, Bella, Micha Mandel, and Rebecca A Betensky.** 2019. "Inverse probability weighting methods for Cox regression with right-truncated data." *Biometrics*.
- Van den Berg, Gerard J.** 2001. "Duration models: specification, identification and multiple durations." In *Handbook of econometrics*. Vol. 5, 3381–3460. Elsevier.
- Xu, Yiqing.** 2017. "Generalized synthetic control method: Causal inference with interactive fixed effects models." *Political Analysis*, 25(1): 57–76.
- Ye, Zhi-Sheng, and Loon-Ching Tang.** 2016. "Augmenting the unreturned for field data with information on returned failures only." *Technometrics*, 58(4): 513–523.
- Yin, Guosheng, and Joseph G Ibrahim.** 2005. "Cure rate models: a unified approach." *Canadian Journal of Statistics*, 33(4): 559–570.
- Young, H Peyton.** 2015. "The evolution of social norms." 7(1): 359–387.



# Appendix



## A. Data Sources

### A.1 City-level Police Records

City-level crime data was obtained through the *Police Data Initiative*. The main website is <https://www.policedatainitiative.org/datasets/>. The links to the raw datasets may be found below. In addition, I provide a list of all offenses classified as sexual / non-sexual offenses and sex offenses excluded from the analysis.

**New York City** I download the police records from <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i> and the arrest data from <https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u>.

**Los Angeles** I download the police records from <https://data.lacity.org/Public-Safety/Crime-Data-from-2019-to-Present/2nrs-mtv8> and the arrest data from <https://data.lacity.org/Public-Safety/Arrest-Data-from-2010-to-2019/yru6-6re4>.

**Cincinnati** I download the data from <https://data.cincinnati-oh.gov/safety/PDI-Police-Data-Initiative-Crime-Incidents/k59e-2pvf>.

**Seattle** I download the data from <https://data.seattle.gov/Public-Safety/SPD-Crime-Data-2008-Present/tazs-3rd5>.

Table A.1: Classification of Non-Sexual Offenses

Classification	Offense Label
<b>Assault</b>	<p>assault, aggravated driveby shooting, assault, aggravated other domestic violence, assault, aggravated other, assault, aggravated peace officer (non serious inj), assault, aggravated peace officer (serious injury), assault minor injury domestic violence, assault minor injury, assault no injury domestic violence, assault no injury, agg assault, agg assault by public servant, agg assault-fam date violence, agg assault on peace officer, agg assault on public servant, agg assault with motor veh, agg robbery by assault, aggravated assault, aggravated assault weapon or ordnance, aggravated vehicular assault, assault, assault knowingly harm victim, assault recklessly harm victim, assault school personnel, assault 3 &amp; related offenses, assault by contact, assault by contact fam dating, assault by threat, assault by threat fam dating, assault of a pregnant woman, assault offenses, assault on peace officer, assault on public servant, assault w injury fam date viol, assault with deadly weapon on police officer, assault with deadly weapon, aggravated assault, assault with injury, assault, aggravated driveby shooting, assault, aggravated other, assault, aggravated other domestic violence, assault, aggravated peace officer (non serious inj), assault, aggravated peace officer (serious injury), assault minor injury, assault minor injury domestic violence, assault no injury, assault no injury domestic violence, battery simple assault, child abuse (physical) aggravated assault, child abuse (physical) simple assault, crash intoxication assault, expired deadly assault, expired att robbery by assault, felonious assault, felonious assault victim seriously harmed, felonious assault weapon or ordnance, felony assault, intimate partner aggravated assault, intimate partner simple assault, negligent assault, other assault, robbery by assault</p>
<b>Burglary</b>	<p>burglary attempted forcible entry, burglary forcible entry, burglary unlawful entry no force, agg burglary armed w deadly weapon, ordnance, aggravated burglary, aggravated burglary inflict harm, att burglary non residence, att burglary of residence, burglary, burglary from vehicle, burglary from vehicle, attempted, burglary non residence, burglary of coin op machine, burglary of residence, burglary of shed detached garage storage unit, burglary of veh no suspect fu, burglary of vehicle, burglary trespass likely occ struct commit offense, burglary trespass occ struct to commit offense, burglary trespass occ likely occ struct to commit offense, burglary trespass struct to commit offense, burglary, attempted, burglary attempted forcible entry, burglary breaking&amp;entering, burglary forcible entry, burglary unlawful entry no force</p>
<b>Murder</b>	<p>homicide manslaughter, homicide murder, aggravated murder, aggravated murder felony, aggravated murder premeditated, capital murder, crash crim neg homicide, crash intox manslaughter, crash manslaughter, crash murder, crash negligent homicide, crim neg homicide non traffic, criminal homicide, expired att capital murder, expired att murder, homicide offenses, homicide negligent vehicle, homicide negligent,unclassified, homicide manslaughter, homicide murder, invol manslaughter result of misdemeanor, involuntary manslaughter, justified homicide, manslaughter, manslaughter, murder, murder &amp; non negl. manslaughter, negligent homicide, reckless homicide, voluntary manslaughter</p>
<b>Robbery</b>	<p>robbery banks, robbery carjacking, robbery commercial house, robbery convenience store, robbery highway, robbery miscellaneous, robbery residence, robbery service station, agg robbery armed, deadly ordnance, agg robbery armed, deadly weapon, agg robbery inflict attempt serious harm, agg robbery deadly weapon, aggravated robbery, attempted robbery, expired att agg robbery weap, robbery, robbery dangerous weapon, ordnance, robbery use threaten immed use of force, robbery by threat, robbery inflict attempt threat ser phys harm, robbery banks, robbery carjacking, robbery commercial house, robbery convenience store, robbery highway, robbery miscellaneous, robbery residence, robbery service station</p>

**Notes:** This table presents non-sexual offenses used for the empirical analysis. I manually classify them into four broad categories: *assault*, *burglary*, *murder*, and *robbery*.

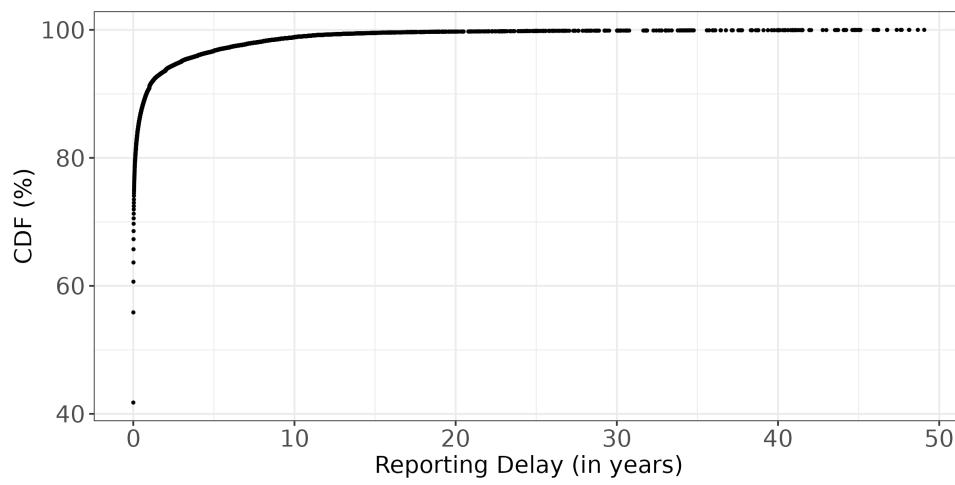
Table A.2: Classification of Sexual Offenses

Status	Offense Label
<b>Excluded</b>	<p>commercialized sex other, commercialized sex pandering, prostitution, vagrancy loitering, agg promotion of prostitution, beastiality, crime against nature sexual aslnt with anim, child pornography, commercialized sex house of ill fame, commercialized sex other, commercialized sex pandering, compelling prostitution, curfew loitering vagrancy violations, expired att agg sex aslnt child, failure to reg as sex offender, fel aslnt sexual conduct w o disclosing hiv knowledge, gross sex imp vict mental physical cond, human trafficking, human trafficking commercial sex acts, human trafficking commercial sex acts, indecency with child exposure, indecent exposure, kidnapping engage in sexual activity, loitering, loitering for drug purposes, loitering in public park, loitering on school prop, loitering deviate sex, loitering gambling (cards, dic, miscellaneous penal law, off. agnst pub ord sensibly &amp;, pornography obscene material, poss promo child pornography, promotion of prostitution, prostitution, prostitution &amp; related offenses, prostitution offenses, public indecency, public indecency appear to be sex act, public indecency engage in sex act, public indecency exposure, purchasing prostitution, sex bat vic minor, off tmp occ discip contr, sex offender registrant out of compliance, sexting depicting a minor, sexting transmit sexual photos, sexual performance by child, vagrancy loitering</p>
<b>Included</b>	<p>sex offenses child molesting, sex offenses exposure, sex offenses lewd &amp; lascivious acts, sex offenses molesting, sex offenses obscene phone calls, sex offenses other (adultry,incest,stat rape,etc), sex offenses peeping tom, sexual assault attempted rape, sexual assault forcible rape, sexual assault other, agg forced sodomy, agg forced sodomy of child, agg rape, agg rape of a child, agg sexual assault child objec, agg sexual assault w object, agg sodomy, assault contact sexual nature, battery with sexual contact, burg of res sexual nature, cont sex abuse of child, expired att agg sexual assault, expired att forced sodomy, expired att rape, expired att rape of a child, expired att sexual assault, expired att sexual aslnt child, expired attaggforcesodomychild, felony sex crimes, forced sodomy, forced sodomy of child, gross sexual imposition, gross sexual imposition &lt; 13 yrs, statutory, gross sexual imposition force, improper contact sex aslnt vict, incest (sexual acts between blood relatives), incest prohibited sex conduct, indecency with a child contact, indecent assault, rape, rape force, threat of, rape substantially impair judgment, rape victim &lt; 13, non forcible, rape victim mental or physical disability, rape of a child, rape, attempted, rape, forcible, sex crimes, sex offenses, sex offenses, consensual, sex offenses child molesting, sex offenses exposure, sex offenses lewd &amp; lascivious acts, sex offenses molesting, sex offenses obscene phone calls, sex offenses other (adultry,incest,stat rape,etc), sex offenses peeping tom, sex,unlawful(inc mutual consent, penetration w frgn obj, sexual assault of child object, sexual assault w object, sexual assault with an object, sexual assault attempted rape, sexual assault forcible rape, sexual assault other, sexual battery, sexual battery mistake for spouse, sexual assault parent or guardian, sexual battery school person of authority, sexual battery victim coerced, sexual coercion, sexual imposition, sexual imposition offensive contact, sexual imposition victim 13, 14, 15, sexual imposition victim impaired, sexual penetration w foreign object, sodomy sexual contact b w penis of one pers to anus oth, statutory rape of child, unlawful sexual conduct with a minor, viol po sexual aslnt victim</p>

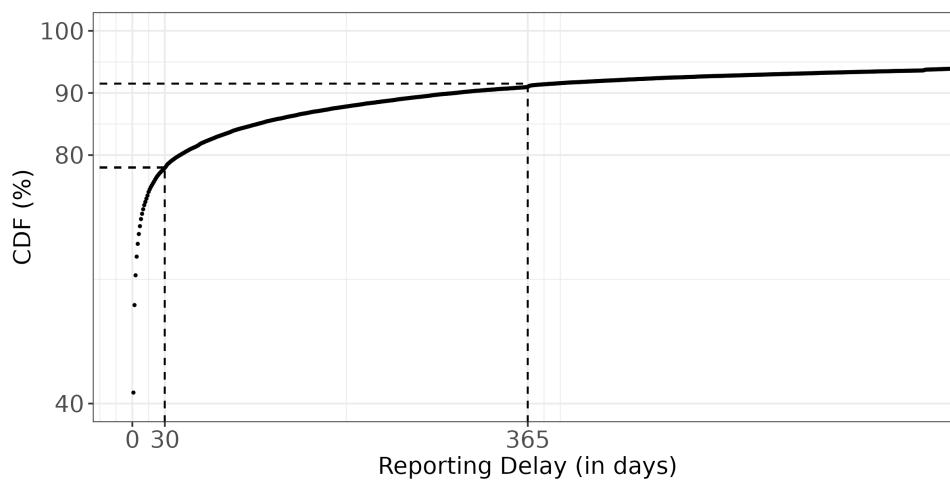
**Notes:** This table presents which sex offenses are used for the empirical analysis. I exclude sex offenses related to *pornography, indecency, loitering, sexting, and prostitution*.

Figure A.1: Distribution of Reporting Delays for Sex Crimes

## A. Entire Distribution



## B. A Closer Look at Shorter Delays



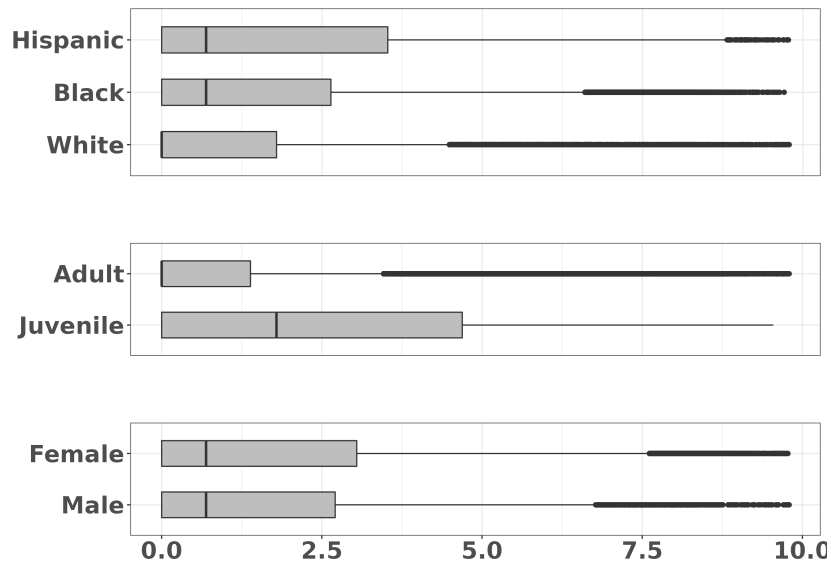
**Notes:** Distribution of observed reporting delays for sex crimes. Approximately 40% of plaintiffs report on the day of the incident, 80% within the first month, and 90% within the first year.

Table A.3: Descriptive Statistics on Police Records

<b>Characteristic</b>	<b>Sex Crime</b>	<b>Murder</b>	<b>Assault</b>	<b>Robbery</b>	<b>Burglary</b>
<b>N</b>	110,591	7,478	1,239,729	295,097	536,312
<b>Report Type</b>					
Delayed	58%	12%	21%	17%	54%
Direct	42%	88%	79%	83%	46%
<b>Time to Report (days)</b>					
Mean	197.19	105.47	4.40	2.68	6.05
Median	2.00	1.00	1.00	1.00	2.00
SD	857.99	948.55	57.12	52.73	62.07
<b>City</b>					
Cincinnati	4.5%	12%	4.4%	9.7%	6.4%
Los Angeles	24%	31%	30%	28%	51%
New York	68%	53%	58%	57%	29%
Seattle	3.8%	3.6%	7.7%	5.2%	14%
<b>Victim Sex</b>					
Female	87%	17%	53%	30%	46%
Male	13%	83%	47%	70%	54%
<b>Victim Age</b>					
Adult	57%	92%	91%	85%	95%
Juvenile	43%	7.5%	9.0%	15%	4.5%
<b>Victim Race</b>					
White	22%	9.9%	16%	23%	42%
Black	40%	67%	48%	40%	37%
Hispanic	38%	23%	36%	37%	20%
<b>Supect Sex</b>					
Female	8.0%	7.8%	25%	6.4%	8.7%
Male	92%	92%	75%	94%	91%
<b>Suspect Age</b>					
Adult	97%	98%	98%	97%	100%
Juvenile	3.4%	2.3%	2.2%	3.2%	0.3%
<b>Suspect Race</b>					
White	14%	7.7%	11%	5.5%	16%
Black	49%	65%	57%	73%	61%
Hispanic	37%	27%	32%	21%	23%

**Notes:** Descriptive statistics for incident-level police records of New York City, Los Angeles, Seattle, and Cincinnati, between 2010 and 2019.

Figure A.2: Reporting Delays per Incident and Victim Characteristics



**Notes:** Boxplots of sex crime report delays per incident characteristics. The darker the gray scale, the longer the delayed reports for each incident characteristic. Reporting delays (in days) are presented on the log scale.

Figure A.3: Arrest Rates

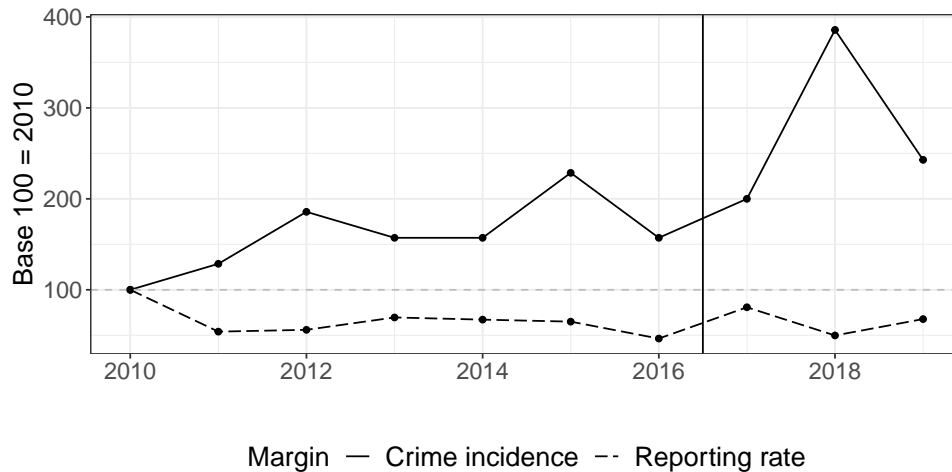


**Notes:** Arrest rates across crime categories between 2010 and 2019 for New York City and Los Angeles.

## A.2 National Crime Victimization Survey (NCVS)

Figure A.4 presents national estimates of sex crime reporting and incidence based on the National Crime Victimization Survey between 2010 and 2019.

Figure A.4: National Survey Estimates of Sex Crime Reporting and Incidence



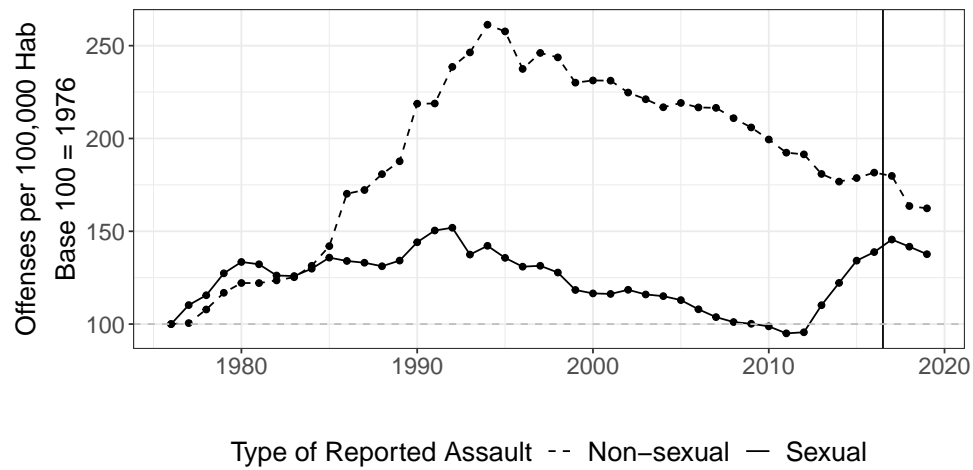
**Notes:** The vertical solid line is set before #MeToo (the viral social media episode took place in Oct 2017).  
**Source:** National Crime Victimization Survey (2010 – 2019)

## A.3 FBI Datasets

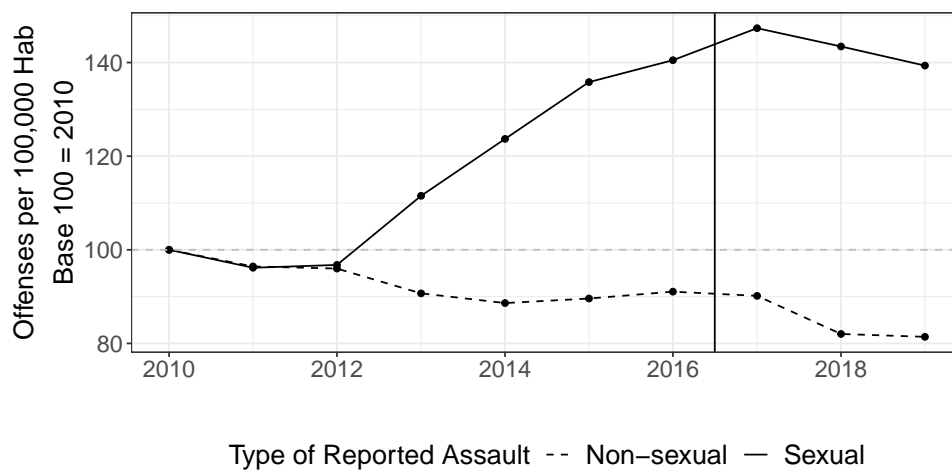
Figure A.5 plots trends in reported sexual and non-sexual assaults. Figure A.6 plots trends in homicides related and unrelated to sexual criminality. These trends are suggestive evidence that the reporting rate of victims of sexual criminality has increased while sex crime incidence has decreased in the U.S. past and more recent history.

Figure A.5: National Trends in Reported Crimes

## A. Long-run (1970-2019)



## B. Short-run (2010-2019)

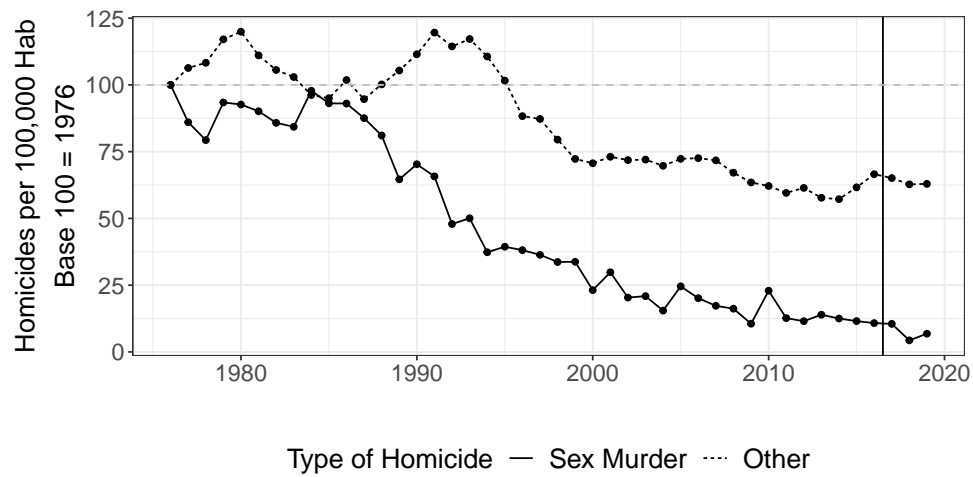


**Notes:** This figure presents national trends in reported crime rates per 100,000 people. Crime reports are aggregated on the report date. Panels A and B distinguish between non-sexual assaults (dashed line) and sexual assaults (solid line). The time-series in Panel A starts in 1970. The time-series in Panel B starts in 2010.

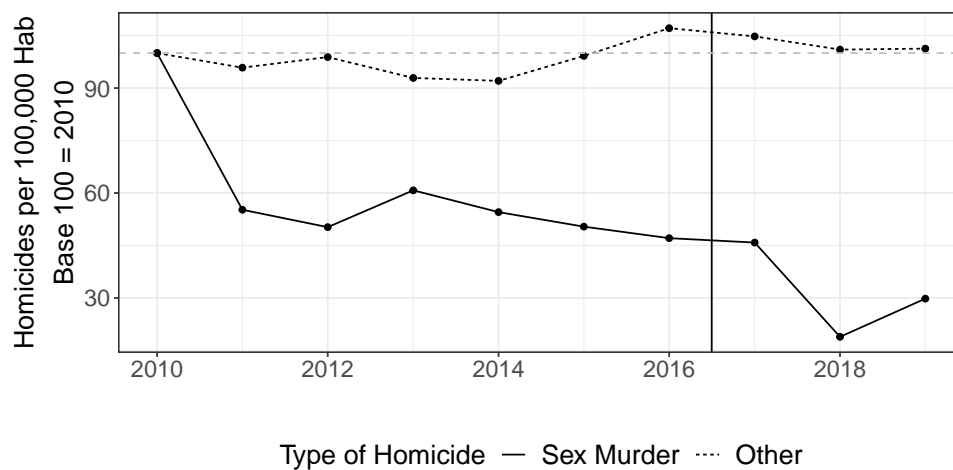


Figure A.6: National Trends in Homicides

## A. Long-run (1970-2019)



## B. Short-run (2010-2019)



**Notes:** This figure presents national trends in homicides per 100,000 people. Crime reports are aggregated on the report date. Panels A and B distinguish between non-sexual homicides (dashed line) and sexual homicides (solid line). The time-series in Panel A starts in 1970. The time-series in Panel B starts in 2010.

## A.4 Twitter

**Keyword Search for Tweets on Sex Crimes** I collect all tweets related to sex crimes for the first day of each month between 2010 and 2019. I also added the 15<sup>th</sup> of October 2017 since the hashtag #MeToo went viral at that time. The exact query sent to the Twitter API is:

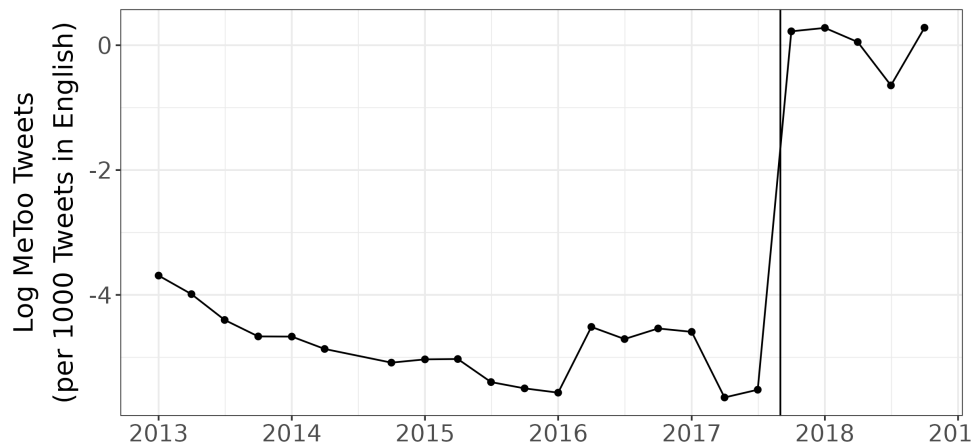
```
'''(sexual assault) OR (sexually assault) OR (sexually assaults)  
OR (sexually assaulted) OR (sexually assaulting) OR (sexual harassment)  
OR (sexually harass) OR (sexually harasses) OR (sexually harassed)  
OR (sexually harassing) OR (sexual abuse) OR (sexually abuse)  
OR (sexually abuses) OR (sexually abused) OR (sexually abusing)  
OR rape OR rapes OR raping OR raped OR #metoo'''
```

**Estimates of Twitter Traffic** To obtain estimates of Twitter traffic, I count the number of English tweets containing the word “a” and/or “the” for the first day of each month between 2010 and 2019. For consistency, I also add the 15<sup>th</sup> of October 2017. The exact query sent to the Twitter API is:

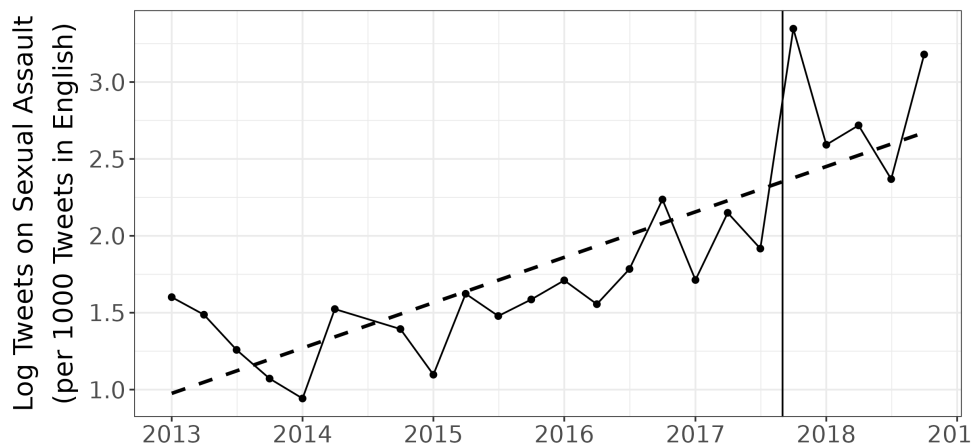
```
"(a OR the) lang:en"
```

Figure A.7: Tweets Related to Sex Crimes

## A. Me Too Tweets



## B. Sex Crime Tweets

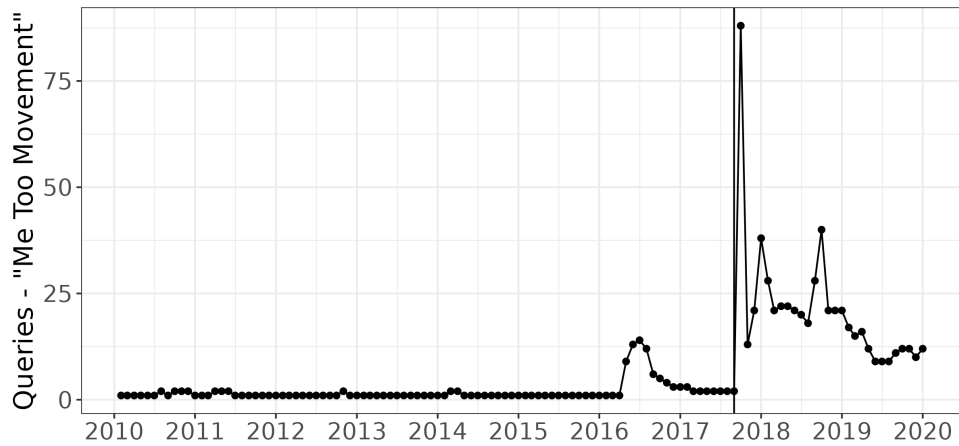


**Notes:** This figure presents trends in the number of tweets with #MeToo (Panel A) and referring to sex crimes more generally (Panel B). The dashed line in Panel B is a linear fit. The structural break in Panel A indicates that the Me Too movement's sudden mediatization brought sex crimes to the forefront of the public debate. However, Panel B nuances this interpretation, as there were clear pre-trends in the number of tweets related to sex crimes before #MeToo was used as a coordination device to combat sexual violence. Before October 2017, the hashtag was marginal on Twitter and rarely referred to sex crimes. The vertical solid line is set one period before #MeToo (Oct 2017).

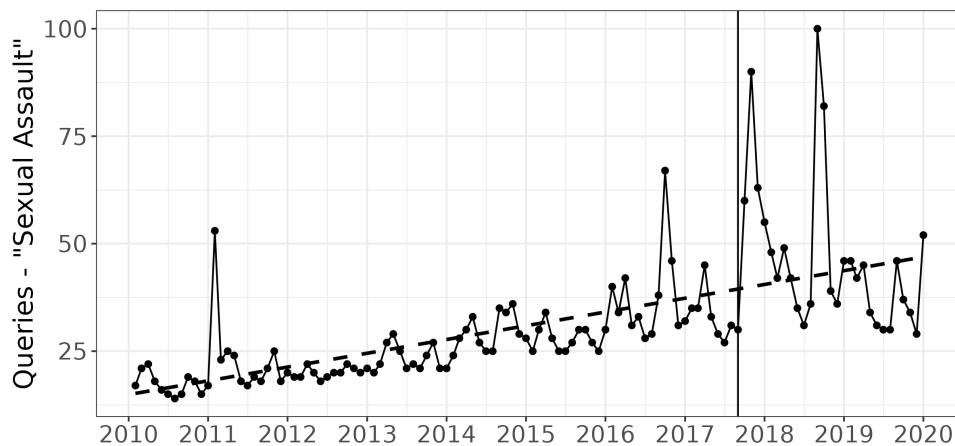
## A.5 Google Trends

Figure A.8: Google Queries Related to Sex Crimes

### A. Queries for Topic “Me Too Movement”



### B. Queries for Topic “Sexual Assault”



**Notes:** This figure presents trends in the number of queries for the topic “Me Too Movement” (Panel A) and for the topic “Sexual Assault” (Panel B). The dashed line in Panel B is a linear fit. The structural break in Panel A indicates that the Me Too movement’s sudden mediatization brought sex crimes to the forefront of the public debate. However, Panel B nuances this interpretation, as there were clear pre-trends in the number of queries related to sex crimes before #MeToo was used as a coordination device to combat sexual violence.

## B. Empirical Issues with Police Data

Delayed reports complicate the analysis of reported crimes  $R_t$ . There are now two dates to consider: the incident date and the date of its report to the police. Aggregating reported crimes on the date of their report or the incident date will generally lead to different counts (and ultimately estimates). When both dates are available, it is straightforward to show that crime trends are typically sensitive to the choice of the date. Figure B.1 plots reported sex crimes aggregated at the incident or the report date. Counts markedly vary across measures, in particular at the beginning and the end of the study period.

This observation can be formalized as follows. First, consider the incident date as the main date for analyzing reported crime statistics. Let  $Y$  denote the time to report to the police,  $F$  its associated cumulative distribution, and  $\chi_t$  the history of interventions.  $R_{t,\tau_1,\tau_2}$  is then

$$R_{t,\tau_1,\tau_2} = \left( F(\tau_2 - t \mid \chi_t) - F(\max(\tau_1 - t, 0) \mid \chi_t) \right) \times C_t.$$

Reported crimes depend on the study period (i.e., on  $\tau_1$  and  $\tau_2$ ). To illustrate this phenomenon, Figure B.2 plots reported sex crimes that occurred in August 2010 depending on the end of the study period  $\tau_2$ . The closer the incident date to the end of the study period, the smaller the number of reported sex crimes for this date. This is an entirely spurious correlation related to the structure of the data (i.e., right-truncation bias in this case).

Next, consider the date of the report as the main date for the analysis of reported crime statistics. Let  $f$  denote the victim population density function of times to report  $Y$ . Then  $R_t$  is

$$R_t = \sum_{j=1}^t f(t-j) \times C_j.$$

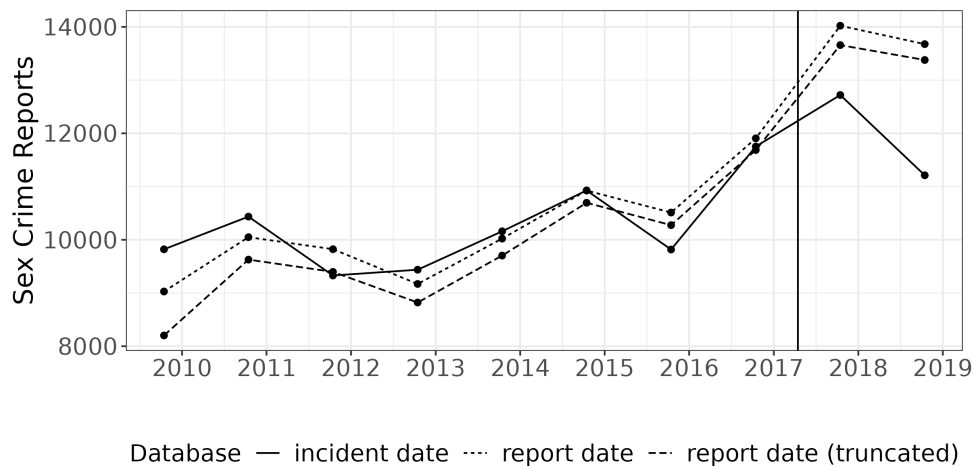
Turning to the linear regression analysis, we have

$$\log \left( \sum_{j=1}^t f(t-j) \cdot C_j \right) = \alpha + \beta_1 D_t + \epsilon_t.$$

Contrary to the incident date, reported crimes aggregated at the report date are not dependent on the study period considered. However, in this last equation, there is a sum in the log transform, and the standard omitted variable framework from Equation

1.1 breaks down. It becomes very hard to measure the effect of the intervention  $D_t$  on crime incidence.

Figure B.1: Sex Crime Reports for Different Police Reporting Guidelines



**Notes:** This figure presents reported sex crime counts for different police reporting guidelines. The solid line aggregates reports by incident date. The dotted line aggregates reports by the report date. The dashed line aggregates reports by the report date, but the data is only records incidents that took place after October 2009. Overall, depending on police recording guidelines, counts may vary substantially.

Figure B.2: Sex Crime Counts for the Same Incident Date But Different Study Periods

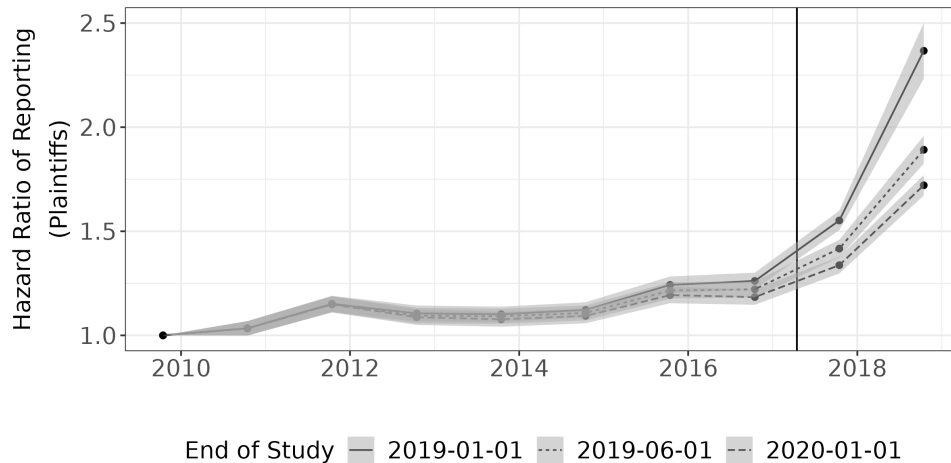


**Notes:** This figure presents reported sex crime counts for the same incident date but different study periods. Each data point corresponds to reported sex crime counts for August 2010 for various end of data collection dates. The closer the incident date to the end of the study period, the more biased downwards the reported sex crime counts. This is because victims often report with a delay.

## C. Duration Model

### C.1 Illustration of the Right-truncation Bias

Figure C.1: Right-truncation leads to a large upward bias.



**Notes:** Yearly estimates of the victim reporting hazard using a naive Cox regression model (i.e., that does not account for right-truncation). Breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. 95% confidence intervals. Without an appropriate correction for right-truncation, estimated hazard ratios are heavily dependent on the end of the study period  $\tau_2$ .

### C.2 Theoretical Motivation for the Promotion Time Model

If a crime is committed, victim  $i$  chooses whether to report the incident in subsequent periods. The psychology literature has highlighted several influential factors in the decision to report (see [Tavarez \(2021\)](#) for a review). Barriers to reporting include internal psychological barriers (e.g., trauma, guilt, and fear), social interactions (e.g., social stigma, relationship with the perpetrator), and the criminal justice system (e.g., negative police interactions, low perceived odds of success in court). In addition, victims are sometimes unaware of resources available and where to report, and it may take time to understand that the situation encountered was in fact rape or sexual assault. Juvenile victims, in particular, lack the level of knowledge needed to recognize and the ability to articulate that a sex crime occurred. All these factors influence both the probability of eventually reporting and reporting delays. To capture internal deliberations of victims in a simple and tractable framework, I assume victim  $j$  is exposed

to  $K_t$  potential decisive arguments to voice out upon a crime being committed.<sup>1</sup>

$$K_t \sim \text{Pois}(\theta_t).$$

These reasons are assumed independent and identically distributed. The time for each argument to trigger a report to the police is drawn from a distribution  $F_0$ .<sup>2</sup> It is then straight-forward to show that the hazard and survival functions of times to report of victims are respectively

$$h_t^{(v)}(y) = \theta_t f_0(y) \quad \text{and} \quad S_t^{(v)}(y) = \exp\left(-\theta_t F_0(y)\right).$$

See below for a formal proof. Note that the survival function has a positive mass as  $y$  tends to infinity, which represents the share of victims who will never report to law enforcement agencies:

$$S_t^{(v)}(+\infty) = \exp(-\theta_t).$$

### Formal Proof

$$\begin{aligned} S_t^v(y) &= P(Y > y) \\ &= P(N = 0) + P(W_1 > y \cap \dots \cap W_N > y \cap N \geq 1) \\ &= \exp(-\theta_t) + \sum_{N=1}^{\infty} (1 - F(y))^N \exp(-\theta_t) \frac{\theta_t^N}{N!} \\ &= \sum_{N=0}^{\infty} (1 - F(y))^N \exp(-\theta_t) \frac{\theta_t^N}{N!} \end{aligned}$$

$$S_t^v(y) = \exp(-\theta_t F(y))$$

<sup>1</sup>An alternative modeling strategy is to formulate the decision process of victims as an optimal stopping problem. It reads as follows. In addition to the costs and benefits of numerous institutional factors (e.g., expected probabilities of success, social pressure), victim  $i$  knows her personal circumstances may change over time. At each period following the incident, she chooses to file a complaint or to postpone the report in the hope of obtaining more favorable circumstances in the future. This is reminiscent of job search models (Mortensen 1986).

<sup>2</sup>Similar models have been used to model duration data in fertility studies and cancer studies (Lambert and Bremhorst 2019).



### C.3 Derivation of the Log-Likelihood

Recall that  $Y$  is the distribution of times to report,  $U$  is the distribution of left-truncation times,  $d$  is the length of the study period (i.e.,  $\tau_2 - \tau_1$ ),  $X$  is a random vector of observed covariates, and  $h_0$ ,  $H_0$  are respectively the baseline hazard and the baseline cumulative hazard of the duration model. In the case of the MPH model, the baseline hazard is a piece-wise constant function. In the case of the promotion time model, the baseline hazard is the density of a distribution of which the hazard is modeled as a piece-wise constant function. Sample observations are indexed by  $i$  and  $\Theta$  is the vector of parameters to estimate. Finally, I refer to random variables with upper-case letters and to their realizations with lower-case letters.

**Time-invariant covariates** I start with the case of time-invariant covariates and no unobserved heterogeneity. I observe a sample of  $n$  realizations  $\{(x_i, y_i)\}_{i \in \{1, \dots, n\}}$  for inference. Under the assumption that  $Y$  and  $U$  are independent conditional on observed covariates  $X$ , we have

$$L(\Theta | x) = \prod_{i=1}^n \frac{P(y_i | x_i)}{P(u_i \leq y_i \leq u_i + d | x_i)},$$

which gives

$$L(\Theta | x) = \prod_{i=1}^n \frac{\exp(\beta' x_i) h_0(y_i) \exp\left(-\exp(\beta' x_i) H_0(y_i)\right)}{\exp\left(-\exp(\beta' x_i) H_0(u_i)\right) - \exp\left(-\exp(\beta' x_i) H_0(u_i + d)\right)},$$

after some factorization, we obtain

$$L(\Theta | x) = \prod_{i=1}^n \frac{\exp(\beta' x_i) h_0(y_i) \exp\left(-\exp(\beta' x_i) (H_0(y_i) - H_0(u_i))\right)}{1 - \exp\left(-\exp(\beta' x_i) (H_0(u_i + d) - H_0(u_i))\right)}.$$

**Time-varying covariates** Next, I extend the model to time-varying covariates. In this case, as soon as one of the covariate changes, the covariate vector needs to be updated. Assume that those variations occur at  $J_i - 1$  occasions  $y_{i1}, \dots, y_{iJ_i-1} \in \mathbb{R}^{J_i-1}$ . Among those, the first  $J_i' + 1$  are observed variations, whereas the remaining are *counterfactual* values for the right-truncation time<sup>3</sup>, such that:  $y_{i0} = u_i \leq y_{i1} \leq \dots \leq y_{iJ_i'} = y_i \leq \dots \leq y_{iJ_i} = v_i \leq \infty$ , yielding the sequence of covariate vectors:

<sup>3</sup>For example, if the victim had not reported to the police in September 2016, she would have been eventually affected by the Me Too movement in October 2017, setting the dummy variable's value to one.

$\chi_i = \{x_{i1}(y_{i0}), \dots, x_{ij_i}(y_{ij_i-1})\} \in \mathbb{R}^{J_i}$ . For  $j = 1, \dots, J_i$ ,  $y_{ij} - y_{ij-1}$  is the time spent by the  $i^{\text{th}}$  subject in his or her  $j^{\text{th}}$  covariate configuration  $x_{ij}(y_{ij-1})$ . Then we obtain

$$L(\Theta | \chi) = \prod_{i=1}^n \frac{\exp(\beta' x_{ij_i'}) h_0(y_{ij_i'}) \exp\left(-\sum_{j=1}^{J_i} \exp(\beta' x_{ij}(y_{ij})) (H_0(y_{ij}) - H_0(y_{ij-1}))\right)}{1 - \exp\left(-\sum_{j=1}^{J_i} \exp(\beta' x_{ij}(y_{ij})) (H_0(y_{ij}) - H_0(y_{ij-1}))\right)}.$$

Note that these last two equations do not require knowing the history of covariates between the incident date and the beginning of follow-up (i.e., between 0 and the left-truncation time). They require, however, knowledge of the covariates until the end of the study period (i.e., for the follow-up time, but also up to the right-truncation time).

**Unobserved Heterogeneity** Next, I extend the model to unobserved heterogeneity. I assume the frailty is a random effect  $\gamma$ . Because the frailty term is unobserved at the individual level, it is necessary to consider the population level and to integrate it out of the likelihood. The likelihood of time-invariant covariates is

$$L(\Theta | \gamma, x) = \prod_{i=1}^n \mathbb{E}_\gamma \left[ P(y_i | u_i \leq y_i \leq u_i + d, x_i) \right].$$

Applying Bayes rule, we have

$$L(\Theta | x) = \prod_{i=1}^n \frac{\mathbb{E}_\gamma \left[ P(y_i | x_i) \right]}{\mathbb{E}_\gamma \left[ P(u_i \leq y_i \leq u_i + d | x_i) \right]}.$$

Replacing expressions with the model parameters gives

$$L(\Theta | \gamma, x) = \prod_{i=1}^n \frac{\mathbb{E}_\gamma \left[ \gamma \exp(\beta' x_i) h_0(y_i) \exp\left(-\gamma \exp(\beta' x_i) H_0(y_i)\right) \right]}{\mathbb{E}_\gamma \left[ \left\{ \exp\left(-\gamma \exp(\beta' x_i) H_0(u_i)\right) - \exp\left(-\gamma \exp(\beta' x_i) H_0(u_i + d)\right) \right\} \right]}.$$

Similar (cumbersome) expressions may be obtained for time-varying covariates. Note that unobserved heterogeneity requires the researcher to know the history of covariates between the incident date and the beginning of follow-up (i.e., between 0 and the left-truncated time). For time-varying covariates, this can be challenging and likely

involves some speculation. For simplicity, I assume there were no interventions that affected victim reporting before the beginning of the study period. As a robustness check, I also estimate the models without left-truncated observations and find qualitatively similar results.

**Parametric Unobserved Heterogeneity** In the case of well-known parametric frailty distributions, the terms can be expressed in terms of the Laplace transform  $\mathcal{L}_\gamma$  and its first derivative  $\mathcal{L}_\gamma^{(1)}$ :

$$L(\Theta \mid \gamma, x) = \prod_{i=1}^n \frac{-\exp(\beta' x_i) h_0(y_i) \mathcal{L}_\gamma^{(1)}(\exp(\beta' x_i) H_0(y_i))}{\left\{ \mathcal{L}_\gamma(\exp(\beta' x_i) H_0(u_i)) - \mathcal{L}_\gamma(\exp(\beta' x_i) H_0(u_i + d)) \right\}}.$$

In the main results, I assume unobserved heterogeneity is gamma distributed with variance  $\Sigma$  (Vaupel, Manton and Stallard 1979; Abbring and Van Den Berg 2007). To ensure that the model is identifiable, I use a parameter restriction for the gamma distribution, such that its mean equals one. For gamma distributions, we know that

$$\mathcal{L}_\gamma(s) = (1 + \Sigma s)^{-\frac{1}{\Sigma}} \text{ and } \mathcal{L}_\gamma^{(1)}(s) = -(1 + \Sigma s)^{-\left(\frac{1}{\Sigma} + 1\right)}.$$

There are theoretical reasons to assume gamma-distributed unobserved heterogeneity. In a large class of frailty models, the frailty distribution among survivors converges to a gamma distribution under mild regularity assumptions (Abbring and Van Den Berg 2007).

**Non-parametric Unobserved Heterogeneity** Nonetheless, in practice, parametric frailties are mainly driven by computational efficiency concerns rather than theoretical justifications. An alternative to parametric distributions is a non-parametric estimation of unobserved heterogeneity introduced by Heckman and Singer (1984). Assume that the population under study consists of  $K$  sub-populations with different frailties  $\{\gamma_k\}_{k \in \{1, \dots, K\}}$  and respective shares within the population  $\{s_k\}_{k \in \{1, \dots, K\}}$ . Further, I impose that all parameters are strictly positive and that the sum of their shares is one. Just like the piece-wise constant function to model baseline hazards, this formulation is a general specification of unobserved heterogeneity, which can account for many

distributions. The likelihood is then

$$L(\Theta | \gamma, x) = \prod_{i=1}^n \frac{\sum_{k=1}^K s_k \left[ \gamma_k \exp(\beta' x_i) h_0(y_i) \exp\left(-\gamma_k \exp(\beta' x_i) H_0(y_i)\right) \right]}{\sum_{k=1}^K s_k \left[ \exp\left(-\gamma_k \exp(\beta' x_i) H_0(u_i)\right) - \exp\left(-\gamma_k \exp(\beta' x_i) H_0(u_i + d)\right) \right]}.$$

**Code Implementation** For models with parametric (or without) unobserved heterogeneity, I use the BFGS algorithm, which builds a picture of the surface to be optimized based on the log-likelihood and its gradient (Nocedal and Wright 1999). For models with non-parametric unobserved heterogeneity, to maximize the odds of finding a global maximum, I rely on a variant of simulated annealing (Bélisle 1992). For simulated annealing and BFGS, I use the *maxLik* package in R (Henningsen and Toomet 2011). As an alternative to simulated annealing, an evolutionary algorithm combined with a derivative-based quasi-Newton method may also be used (Mebane Jr and Sekhon 2011). The evolutionary algorithm is implemented in the R package *rgenoud*.

## C.4 Additional Insights from Monte Carlo Simulations

Table C.1: Estimates for Correct and Misspecified Values for  $\alpha$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<b>Treatment 1</b>	0.443 (0.044)	0.455 (0.044)	0.459 (0.045)	0.466 (0.046)	0.474 (0.048)	0.477 (0.051)	0.112 (0.026)
<b>Treatment 2</b>	-0.131 (0.040)	-0.132 (0.041)	-0.141 (0.042)	-0.149 (0.043)	-0.156 (0.045)	-0.165 (0.046)	-0.039 (0.026)
<b>Treatment 3</b>	-0.008 (0.045)	-0.004 (0.047)	-0.011 (0.048)	-0.010 (0.050)	-0.008 (0.052)	-0.009 (0.054)	-0.299 (0.026)
<b>Baseline hazard</b>							
Day 1	X	X	X	X	X	X	X
Day 2+	X	X	X	X	X	X	X
<b>Never-reporters at baseline (%)</b>	0%	10%	30%	50%	70%	90%	0%
Model	MPH	PT	PT	PT	PT	PT	MPH
Correction for double-truncation	Yes	Yes	Yes	Yes	Yes	Yes	No
Algorithm	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS
N Observations	7059	7059	7059	7059	7059	7059	7059
LogLik	-26026	-26024	-26025	-26025	-26026	-26027	-95716

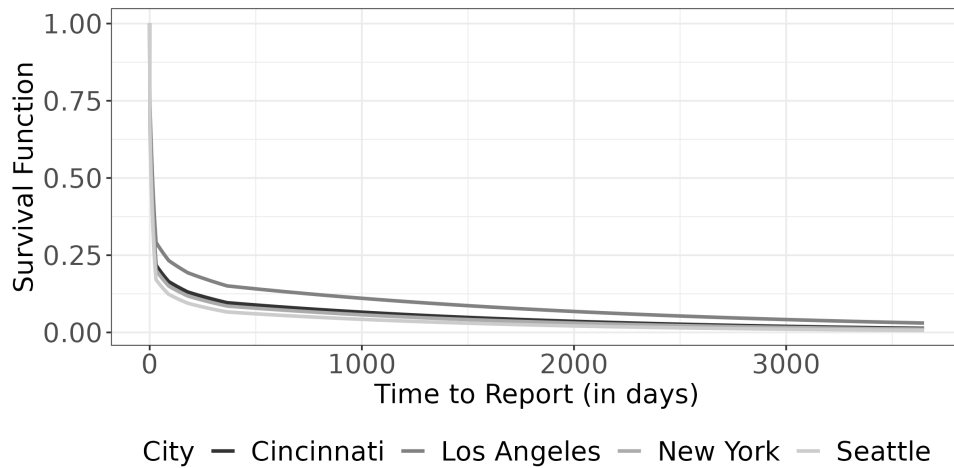
**Notes:** Estimates and standard errors for a simulated police records dataset (based on Equation 1.5). The baseline hazard is 0.2 on the first day, then drops to 0.01. 50% of the victims eventually report to the police at baseline. There is no unobserved heterogeneity. Three interventions/treatments affect the propensity to report of victims by respectively 0.5, -0.2, and 0 on the log-scale. The plugin parameter  $\alpha$  determines the share of never-reporters at baseline (i.e., for the first data period). Misspecifying its value has an impact on estimates. Relying on a model that assumes all victims eventually report (see Equation 1.4) leads to coefficients biased towards zero (see column 1). Relying on the promotion time model, setting too low a share of never-reporters (relative to the ‘true’ value) leads to estimates of treatment effects biased towards zero (see columns 2 to 6). Column 4 is the correctly specified model. In all cases, the resulting bias remains small relative to a naive Mixed Proportional Hazards (MPH) model that does not account for double-truncation in the data (see column 7).

## D. Robustness and Additional Empirical Results

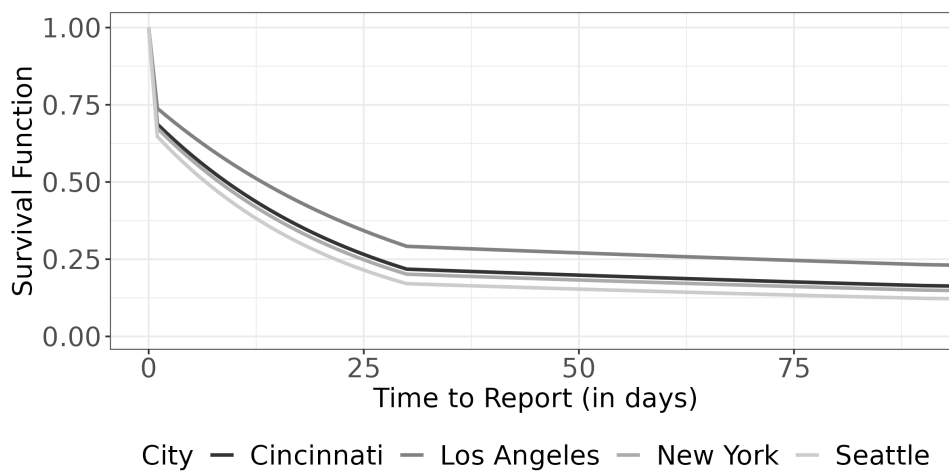
In this appendix, I present additional details and several robustness checks for the main empirical application.

Figure D.1: Estimated Baseline Survival Functions

### A. Over Ten Years



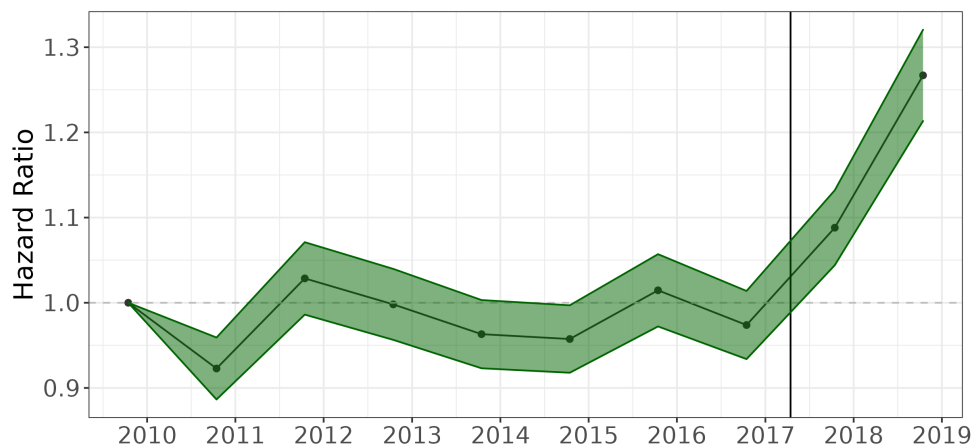
### B. Zoom In on the First 90 Days



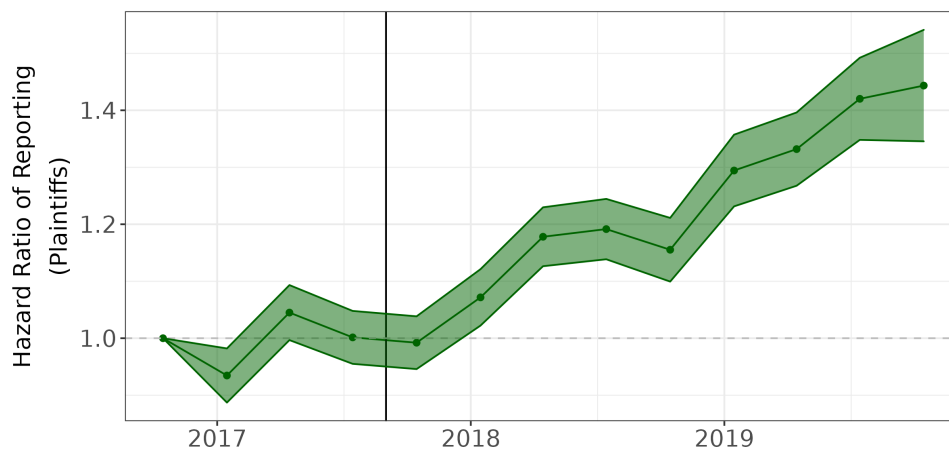
**Notes:** This figure presents estimates of the baseline survival functions implied by Equation 1.4 for all four cities in the sample. Breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. Panel A presents the first ten years. Panel B zooms in on the first 90 days.

Figure D.2: Hazard Ratios Over Time – Plaintiff Reporting

## A. Yearly Hazard Ratios



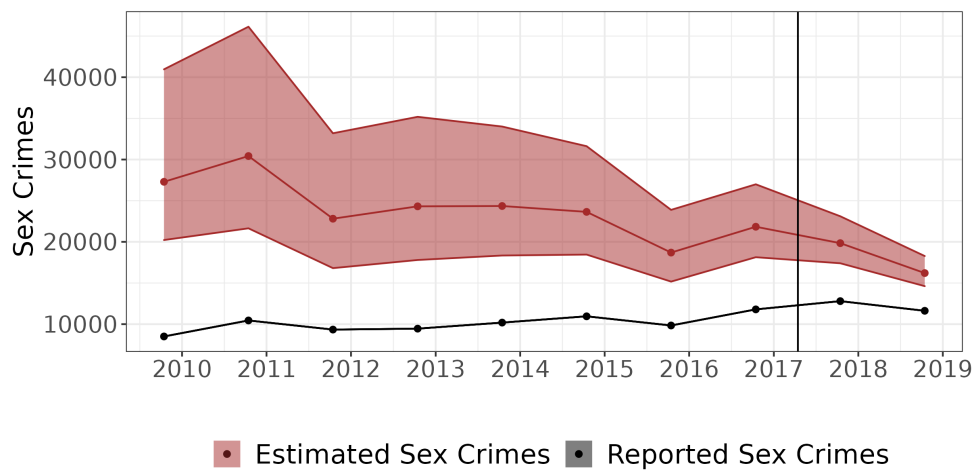
## B. Quarterly Hazard Ratios



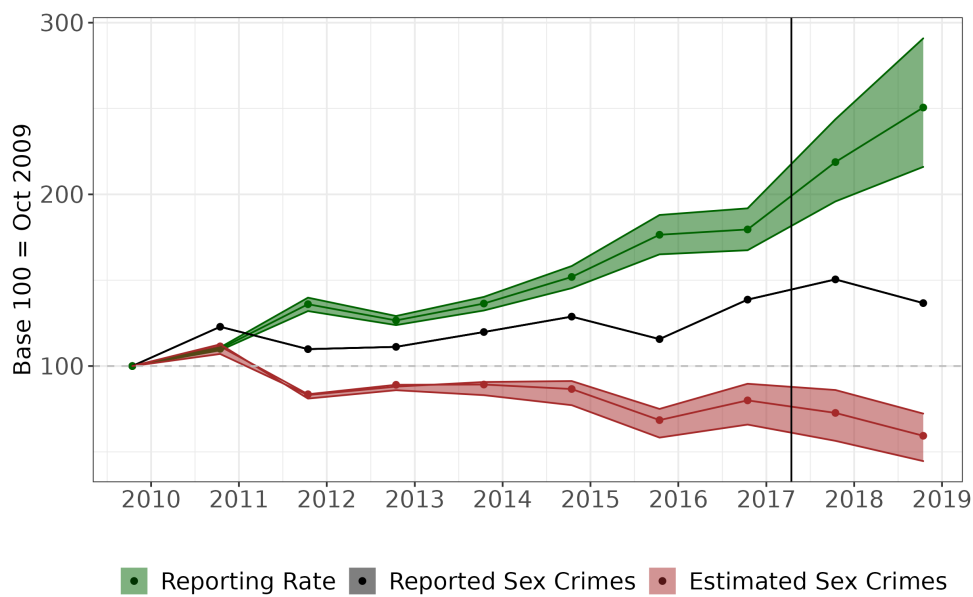
**Notes:** This figure presents estimates for #MeToo’s effect on the plaintiff reporting hazard (see Equation 1.4). Panel A is at the yearly level. Panel B zooms in around #MeToo’s mediatization and is at the quarterly level. The solid vertical line represents the Me Too movement’s intense mediatization on social media (Oct. 2017). Breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. The likelihood appropriately corrects for double-truncation. Unobserved heterogeneity is assumed gamma-distributed. Numerical estimation of the hessian combined with the delta method is used to compute standard errors. 95% confidence intervals.

Figure D.3: Trends in Sex Crime Incidence and Reporting for Different Values of  $\alpha$

A. Nominal Values



B. Base 100 = Oct. 15, 2009

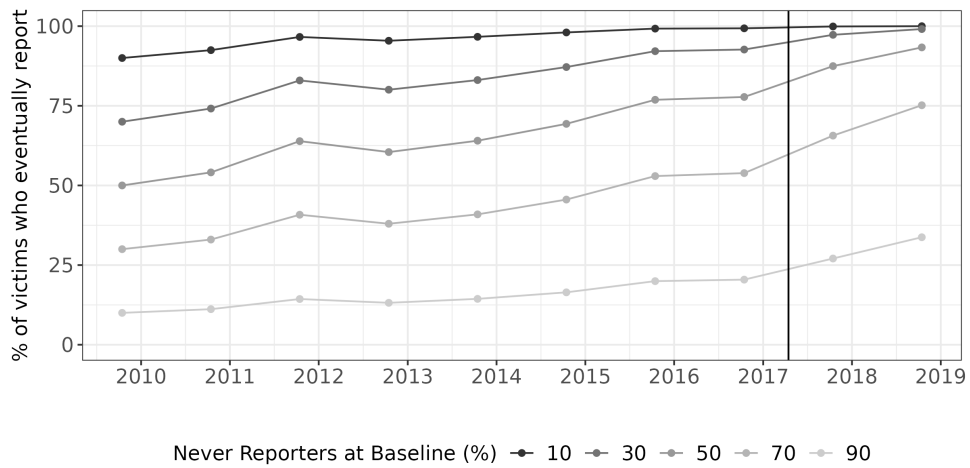


**Notes:** This figure presents estimates of sex crime incidence and victim reporting for different values of  $\alpha$ . The point estimates assume 70% of never-reporters as in the main text. The confidence intervals are built by varying the share of never-reporters at baseline from 60% to 80%. This is in line with the National Crime Victimization Survey’s estimates of the victim reporting rate since 2011.

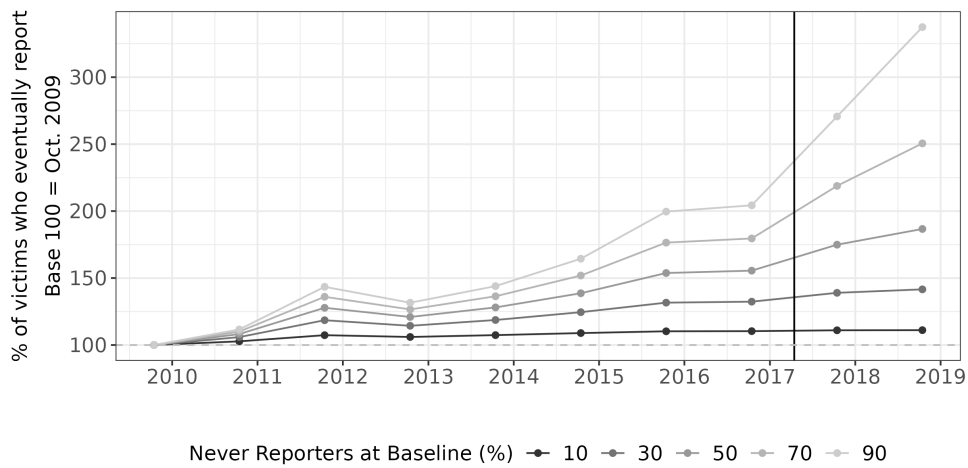


Figure D.4: Share of Never-reporters for Different Values of  $\alpha$

A. Nominal Values

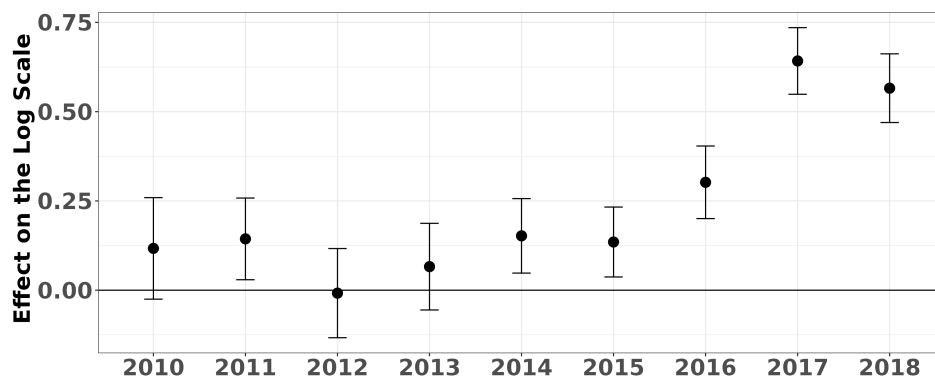


B. Base 100 = Oct. 15, 2009



**Notes:** This figure presents the estimated share of never-reporters for different values of the share of never-reporters in the first period of the study (2010), ranging from 10% to 90%. These trends are to be compared to the share of never-reporters in the main part of the paper (70%). In all cases, I find an increase in the propensity of victims to report a crime to the police. However, depending on  $\alpha$ , the magnitude of the increase varies. The smaller the share, the larger the implied trends in the reporting rate. The interactive fixed effects (IFE) model in column (4) has five factors.

Figure D.5: Test for Time-Dependent Effects



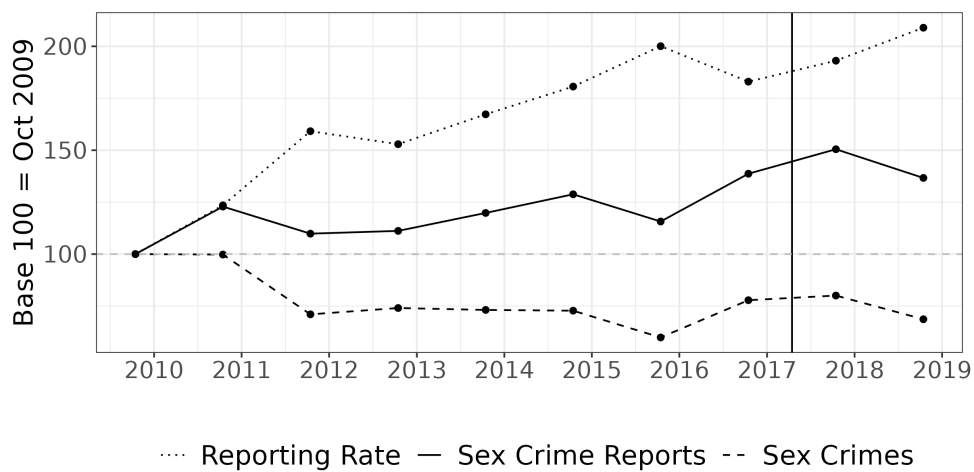
**Notes:** I distinguish yearly hazard ratios for recent sex crime incidents (less than 365 days) and older sex crime incidents (more than 365 days). This figure presents the additional marginal yearly effect on older crime incidents. A statistically significant effect suggests the presence of time-dependent effects and thus a violation of the proportional hazards assumption. Breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. The likelihood appropriately corrects for double-truncation. Unobserved heterogeneity is assumed gamma-distributed. 95% confidence intervals.

Figure D.6: Yearly Plaintiff Hazard Ratios Under Time-Dependent Effects



**Notes:** Yearly estimates of the plaintiff reporting hazard based on Equation 1.8. To account for time-dependent effects, I only keep the effect of year dummies on recent incidents (defined as less than 365 days old). The solid vertical line represents the Me Too movement's intense mediatization on social media (Oct. 2017). Breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. The likelihood appropriately corrects for double-truncation. Unobserved heterogeneity is assumed gamma-distributed. Numerical estimation of the hessian combined with the delta method is used to compute standard errors. 95% confidence intervals.

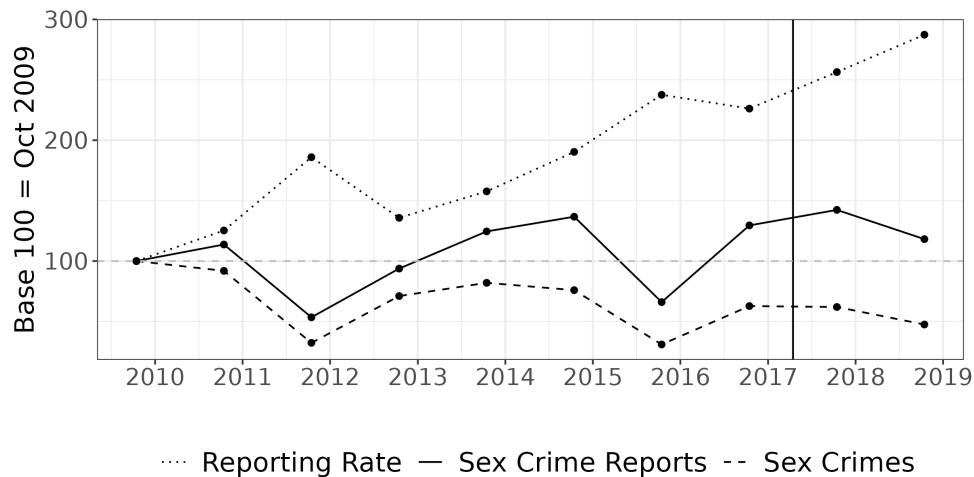
Figure D.7: Trends in Sex Crime Incidence and Reporting Under Time-Dependent Effects



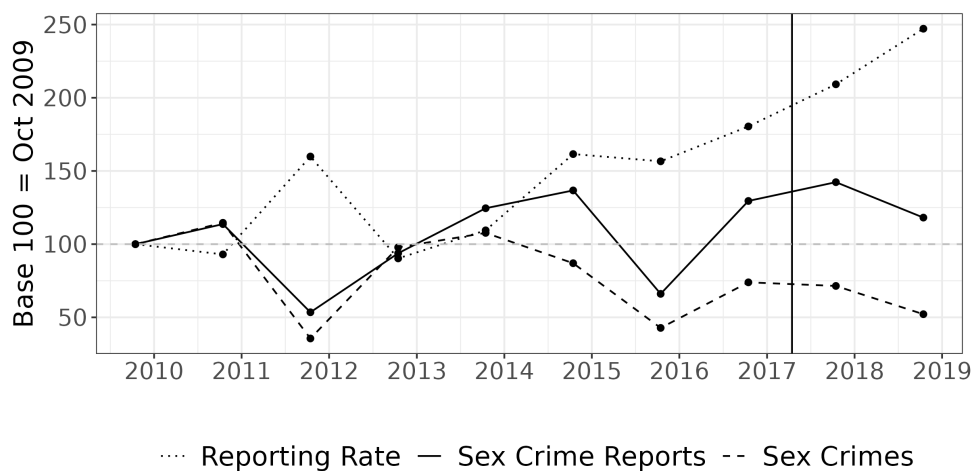
**Notes:** This figure decomposes reported sex crimes into an extensive margin (crime reporting) and an intensive margin (crime incidence) based on Equations 1.5 and 1.6. To account for time-dependent effects, I only keep the effect of year dummies on recent incidents (defined as less than 365 days old). The dotted line are yearly estimates of the reporting rate of victims of sex crimes. The solid line are yearly reported sex crimes. The dashed line are yearly estimates of sex crime incidence. Breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. The likelihood appropriately corrects for double-truncation. Unobserved heterogeneity is assumed gamma-distributed.

Figure D.8: Robustness to Unfounded Allegations

## A. Trends for Los Angeles – Full Sample



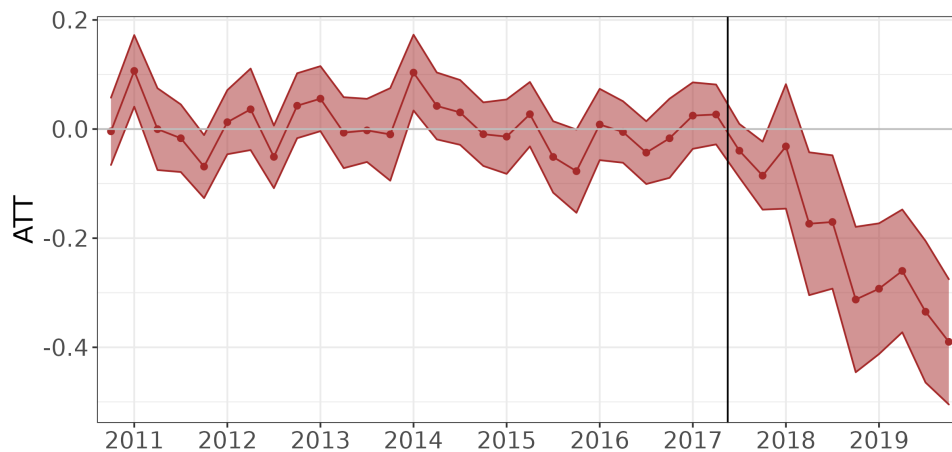
## B. Trends for Los Angeles – Sample Restricted to Arrests



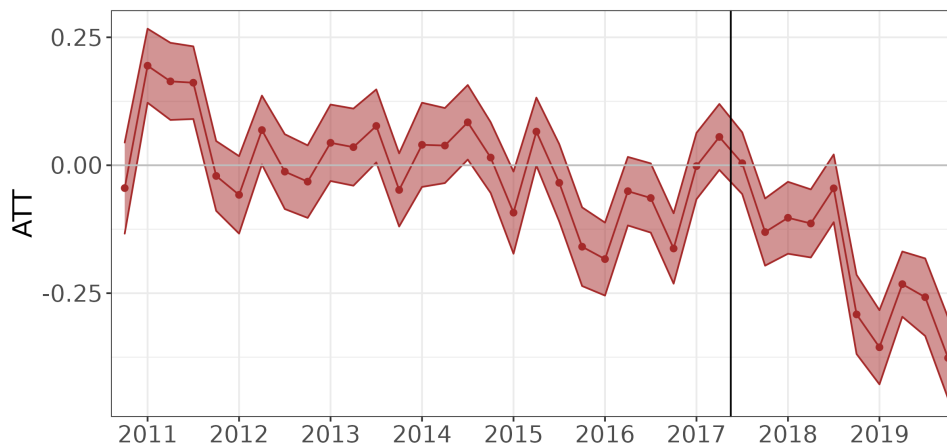
**Notes:** This figure decomposes reported sex crimes into an extensive margin (crime reporting) and an intensive margin (crime incidence) based on Equations 1.5 and 1.6. Panel A estimates trends for the universe of complaints filed at the Los Angeles Police Department. Panel B restricts the sample to sex crime reports that lead to an adult arrest. In doing so, I focus on a subset of reported incidents that are unlikely to be unfounded. I find relatively similar trends as in the main text and Panel A, suggesting false allegations are not driving my results. The dotted line are yearly estimates of the reporting rate of victims of sex crimes. The solid line are yearly reported sex crimes. The dashed line are yearly estimates of sex crime incidence. Breaks in the baseline hazard are set after 1, 30, 90, 180, and 365 days. The likelihood appropriately corrects for double-truncation. Unobserved heterogeneity is assumed gamma-distributed.

Figure D.9: #MeToo Effect on Sex Crime Incidence

## A. Interactive Fixed Effects (IFE)



## B. Matrix Completion



**Notes:** Quarterly estimates of the Average Treatment for the Treated (ATT). The main results are presented in Table 8.2. Panel A uses an interactive fixed effects (IFE) model (Xu 2017). Panel B relies on the Matrix Completion method (Athey et al. 2021). 95% confidence intervals are constructed with a bootstrap procedure and 200 iterations. The vertical solid black line corresponds to the Me Too movement's mediatization. The counterfactual model that maps the best pre-treatment trends in the time series is the interactive fixed effects (IFE) model with two-way fixed effects and five additional factors.

# Bibliography

- Abbring, Jaap H, and Gerard J Van Den Berg.** 2007. "The unobserved heterogeneity distribution in duration analysis." *Biometrika*, 94(1): 87–99.
- Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi.** 2021. "Matrix completion methods for causal panel data models." *Journal of the American Statistical Association*, 1–41.
- Bélisle, Claude JP.** 1992. "Convergence theorems for a class of simulated annealing algorithms on  $\mathbb{R}^d$ ." *Journal of Applied Probability*, 29(4): 885–895.
- Heckman, James, and Burton Singer.** 1984. "A method for minimizing the impact of distributional assumptions in econometric models for duration data." *Econometrica: Journal of the Econometric Society*, 271–320.
- Henningsen, Arne, and Ott Toomet.** 2011. "maxLik: A package for maximum likelihood estimation in R." *Computational Statistics*, 26(3): 443–458.
- Lambert, Philippe, and Vincent Bremhorst.** 2019. "Estimation and identification issues in the promotion time cure model when the same covariates influence long-and short-term survival." *Biometrical Journal*, 61(2): 275–289.
- Mebane Jr, Walter R, and Jasjeet S Sekhon.** 2011. "Genetic optimization using derivatives: the rgenoud package for R." *Journal of Statistical Software*, 42: 1–26.
- Mortensen, Dale T.** 1986. "Job search and labor market analysis." *Handbook of labor economics*, 2: 849–919.
- Nocedal, Jorge, and Stephen J Wright.** 1999. *Numerical optimization*. Springer.
- Tavarez, Lahiz P.** 2021. "Waiting to Tell: Factors Associated with Delays in Reporting Sexual Violence."
- Vaupel, James W, Kenneth G Manton, and Eric Stallard.** 1979. "The impact of heterogeneity in individual frailty on the dynamics of mortality." *Demography*, 16(3): 439–454.

**Xu, Yiqing.** 2017. "Generalized synthetic control method: Causal inference with interactive fixed effects models." *Political Analysis*, 25(1): 57–76.

## Chapter 2

# Social Media and the Dynamics of Protests

**Note:** This chapter is co-authored with Pierre Boyer, Thomas Delemotte, Vincent Rollet, and Benoît Schmutz.



### Abstract

This paper argues that social media have ambiguous effects on protest dynamics. On the one hand, they enable the development of online communities of protesters that keep movements alive. On the other hand, they are fertile ground for political polarization and radicalization. Using data from the 2018-2019 Yellow Vests uprising in France, we show that local street protests triggered the creation of large communities of protesters on Facebook. However, these communities progressively became more antagonistic, negative, and ideologically segregated. While moderate discussants left the discussions, those who remained radicalized. Facebook’s recommender algorithm likely contributed to this pattern by consistently showcasing radical content.

**Keywords:** protests, social media, Yellow Vests movement, NLP techniques, collective action<sup>1</sup>

---

<sup>1</sup>The authors thank Micael Castanheira, Luisa Dörr, Allan Drazen, Elie Gerschel, Julien Grenet, Fanny Henriët, Clément Imbert, Francis Kramarz, James Lo, Nolwenn Loisel, Andrew Lonsdale, Phoebe Mac Donald, Clément Malgouyres, Isabelle Méjean, George Melios, Pierre-Guillaume Méon, Vincent Pons, Audrey Rain, Anasuya Raj, Manon Revel, Ilan Tojerow, Clemence Tricaud, Thomas Van Casteren, Oliver van den Eynde, Lionel Wilner, Ekaterina Zhuravskaya, and many seminar and conference participants at CREST, “Dulbea Workshop on the Political Economy of Mass Demonstrations” at ULB Brussels, Boston University, PSE, Sciences Po, Harvard (Government), Paris 8 and IIPF 2020 for their comments. The authors gratefully acknowledge the Investissements d’Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047), ANR-19-CE41-0011 and ANR-20-CE41-0013 for financial support, the CASD (Centre d’Accès Sécurisé aux Données) and INSEE for the access to French administrative data, and Change.org for sharing their anonymized data.

## 1. Introduction

Since the early 2010s, social media have been instrumental in the emergence of protest movements worldwide (see, for reviews, [Della Porta and Diani 2020](#); [Zhuravskaya, Petrova and Enikolopov 2020](#)). As a result, modern protest movements are now a combination of overlapping online and offline mobilizations. In this paper, we argue that the implications of this hybridization are ambiguous: on the one hand, social media may help protest movements persist over time by nurturing active online communities of protesters; on the other hand, the way social media are used by protesters and structure discussions may be conducive to the radicalization of those communities.

Our analysis is based on the Yellow Vests movement, widely recognized as the most significant social unrest episode in recent French history.<sup>1</sup> Several features of this movement make it particularly well-suited to the study of the impact of social media on protests. First, it was sparked by an online petition and used Facebook to coordinate street protests across the country. Second, it outlived the initial day of protests and stayed active on various online platforms for several months, with hundreds of thousands of petition signatures and lively discussions on dedicated Facebook pages. Third, it turned into a full-blown protest against the government and lost support in the general population. These conflicting patterns suggest that the ease of coordination via social media may come at the cost of the inability to articulate a consistent message. While such a trade-off is not specific to the Yellow Vests,<sup>2</sup> quantitative evidence of it is still missing. We use geolocated data on street protests, Facebook groups, and petition signatures on Change.org, combined with the textual analysis of discussions on Facebook pages, to study the interplay between online and offline mobilizations and its implications for the movement's dynamics.

We start by documenting the existence of a positive feedback loop between mobilization on Facebook and street protests. To that end, we use spatial regressions at the municipality level and focus on the first day of the protests: the 11/17 roadblocks. Those roadblocks had been organized on dedicated Facebook groups all over the country, as confirmed by the strong positive correlation between the presence of an early Facebook group and the occurrence of a roadblock. This correlation aligns with previous research on other settings and is consistent with the history of the movement's

---

<sup>1</sup>For references, see [Algan et al. \(2019\)](#); [Bendali and Rubert \(2020\)](#); [Boyer et al. \(2020\)](#); [Cointet et al. \(2021\)](#); [Sebbah et al. \(2018\)](#).

<sup>2</sup>For example, drawing from a qualitative analysis of the network structure of the Black Lives Matter movement, [Mundt, Ross and Burnett \(2018\)](#) note that “social media use can potentially hinder positive movement impact by making it easier for groups to adopt or appropriate symbols even if they do not share the collective identity or primary focus of the movement.”

early days. However, the reverse direction of the link is novel and more unexpected: those early roadblocks then spurred the creation of a new wave of Facebook groups. To establish causality, we employ an instrumental variable strategy based on the spatial dispersion of roundabouts in French cities. Roundabouts are attractive protest locations because they enable demonstrators to block several roads simultaneously and are easy to set camp on. At the same time, they are widely recognized as architectural fads. Our first main contribution is to show that the presence of a local roadblock led to the creation of 1.2 additional local Facebook groups. We interpret this finding as evidence that protesters sought to continue interacting with each other online after the initial offline contact, resulting in the consolidation of the Yellow Vests' online infrastructure.

Despite the development of active Yellow Vests online communities, the number of protesters in the street quickly subsided. To understand the unraveling of the movement, we turn to the analysis of a corpus of 2.8 million sentences posted on public Facebook pages. Using various Natural Language Processing (NLP) techniques, we show that messages related to organizational concerns and practical demands decreased over time, while those with more antagonistic content, such as insults or mentions of violence, increased. Similarly, the share of messages with a negative sentiment or a higher probability of being written by affiliates of extreme parties also increased. We conduct two quantitative exercises suggesting that the discussions radicalized partly because they were taking place on social media — these correspond to our second main contribution.

First, we show that discussants were disproportionately exposed to radical content. Since our dataset includes information on the ordering of comments displayed to users, we evaluate whether our radicalization measures correlate with the recommendations of Facebook's algorithm. We find that Facebook significantly modified the original comment ordering and that comments associated with our radicalization measures were more likely to appear prominently below Facebook posts. Second, we show that online discussions drove moderate discussants out and, conversely, increasingly attracted radical discussants. Using a de-identified panel of individual discussants and their posts, we can decompose the radicalization process we measure between an extensive margin (changes in the composition of the population of discussants) and an intensive margin (an individual-level increased tendency to post radical messages). Empirically, we find that both margins played an almost equally important role, although the effect of the extensive margin was slightly delayed compared to that of the intensive margin. This result is consistent with the notion that participation in these online communities is volatile and subject to the rise of echo chambers that limit the diversity of participants in favor of the most radically opinionated.

Overall, we view the findings presented in this paper as a cautionary tale on the ambiguous relationship between social media and protest movements. Many studies, relying on various methodologies, have emphasized the importance of digital technologies in large-scale protest movements (see, among others, Loader 2008; Earl and Kimport 2011; Bennett and Segerberg 2012; Castells 2015; Manacorda and Tesei 2020).<sup>3</sup> Social media, in particular, have been shown to facilitate coordination, signaling, and mobilization among protesters.<sup>4</sup> As Qin, Strömberg and Wu (2017), we confirm that the close monitoring of social media may help predict where protests are more likely to occur. Moreover, in line with the intuition of Bastos, Mercea and Charpentier (2015), we show that street protests may revitalize online mobilization, thereby creating a feedback loop that helps protest movements stay active longer than initially planned. As such, social media may be instrumental to the return of local politics documented all over the world (Della Porta and Diani 2020; Le Galès 2021).

However, social media are far from being neutral communication technologies, and they have long been accused of nurturing ideological segregation (Pariser 2011; Ross Arguedas et al. 2022). Even though the real impact of social media on polarization is still an open debate (see, among others, Bail et al. 2018; Fletcher, Cornia and Nielsen 2020), this paper argues that such mechanisms may be particularly relevant in the case of protest movements, which tend to radicalize over time, partly in response to state repression (Della Porta 2018). We show that Yellow Vests online communities were subject to both echo chambers and algorithmic filter bubbles that likely contributed to the radicalization of online discussions, in parallel to high levels of violence in the streets. The ambiguity of social media, as both facilitators of protest movements and instruments of their radicalization, may contribute to explaining why mobilization today can be “very successful in terms of number but tends to be more volatile and intermittent than in the past” (Della Porta 2013).

Our analysis of the Yellow Vests’ discussions also exemplifies that social media allow for real-time surveys of the coalitions behind social movements. Access to detailed information about the political preferences of the protesters, their opinion vis-à-vis policy-makers and their media strategies provides researchers with a unique opportunity to revisit longstanding debates on the nature of protest movements (Lipsky 1968; Meyer 2004). In particular, while a large strand of the literature on protest movements

---

<sup>3</sup>In the US, the Tea Party movement in 2009 is arguably the last large-scale protest movement that relied on the support of traditional media, such as local TV or radio talk shows (Madestam et al. 2013).

<sup>4</sup>Noteworthy examples include Rane and Salem (2012); Acemoglu, Hassan and Tahoun (2018); Clarke and Kocak (2020); Gaffney (2010); Borge-Holthoefer et al. (2011); González-Bailón et al. (2011); Bursztyjn et al. (2021); Enikolopov, Makarin and Petrova (2020); Jost et al. (2018); Fergusson and Molina (2019).

has been devoted to measuring their impact on various outcomes,<sup>5</sup> it is now becoming possible to follow protest dynamics over time and space with an unprecedented level of detail. However, these new monitoring possibilities are not without risks, particularly if online conversations allow authoritarian regimes to identify dissenters (Earl, Maher and Pan 2022).

## 2. Context, Data, and Methods

### 2.1 Brief History

The Yellow Vests movement resulted from chance and the social media ecosystem. In May 2018, a motorist, Priscilla Ludosky, created a petition against gasoline taxes on the Change.org platform. Even though the petition had only garnered a few hundred signatures during its first months, it was mentioned in a local newspaper on October 12, 2018. The wife of a truck driver who had been planning a roadblock of the Paris ring road for November 17<sup>th</sup> read the article and linked the petition on Facebook. Nine days and thousands of local signatures later, a national newspaper published a new article on the petition and the roadblock project, and signatures skyrocketed nationwide. On October 24, the yellow road security jacket, which every car owner is compelled by law to have in her trunk, was proposed as a rallying sign for angry motorists. The organizers of roadblocks heavily relied on Facebook to spread information, and several dedicated websites were created to list relevant local Facebook groups. On November 17<sup>th</sup>, hundreds of thousands of protesters blocked hundreds of roads across France.

The movement resorted to more conventional weekly demonstrations in France's main cities as most roadblocks were quickly evacuated. A climax of violence was reached on December 1<sup>st</sup> in Paris. The following Saturday, police tanks were mobilized, and 2,000 people were arrested. On December 10<sup>th</sup>, as a token of peace, President Macron presented a 10-billion-euro plan that significantly bent the government's budgetary policy. In particular, he pledged a 100 per month increase in the minimum wage and excluded charges and taxes on overtime hours in 2019 and any 2018 end-of-year bonuses paid to employees. He further asked for a compilation of lists of grievances (*Cahiers de doléances*, as took place during the French Revolution in 1789) across the country, which was followed by hundreds of town hall meetings meant to allow everyone to voice their concerns through a "Great National Debate" (*Grand Débat National*).

---

<sup>5</sup>Examples include Giugni (1998); Gillion (2012); Little, Tucker and LaGatta (2015); Frye and Borisova (2019); Ketchley and El-Rayyes (2021); Reny and Newman (2021); Siegel (2009); Meirowitz and Tucker (2013); Wallace, Zepeda-Millán and Jones-Correa (2014); Branton et al. (2015); Mazumder (2018); Larson et al. (2019).

After this response, some roadblocks became permanent campsites, and weekly demonstrations continued for months. However, the number of demonstrators soon became negligible (except in Paris, where some large-scale demonstrations gathering protesters from other parts of France still took place until March 2019). At the same time, the protesters lost public support in the French population and ultimately failed to present a united front for the upcoming elections (the 2019 European Parliament election on May 26<sup>th</sup>). The movement was still active online in 2022 and organized sporadic protests where Yellow Vests were worn as a badge of honor. As such, this simple piece of garment has become a persistent and divisive icon in the French political landscape.

## 2.2 Measuring the Intensity of Mobilization

We gathered data on the mobilization using multiple sources (see Appendix A for details). To understand the movement's roots, we retrieved anonymized data on petition signatories from Change.org. The data includes signatories' ZIP codes, allowing us to geolocate them. By October 16, 2019, the petition had garnered 1,247,816 signatures, including 1,043,337 with a valid French ZIP code. We interpret petition signatures as signaling discontent towards the government and a willingness to protest.

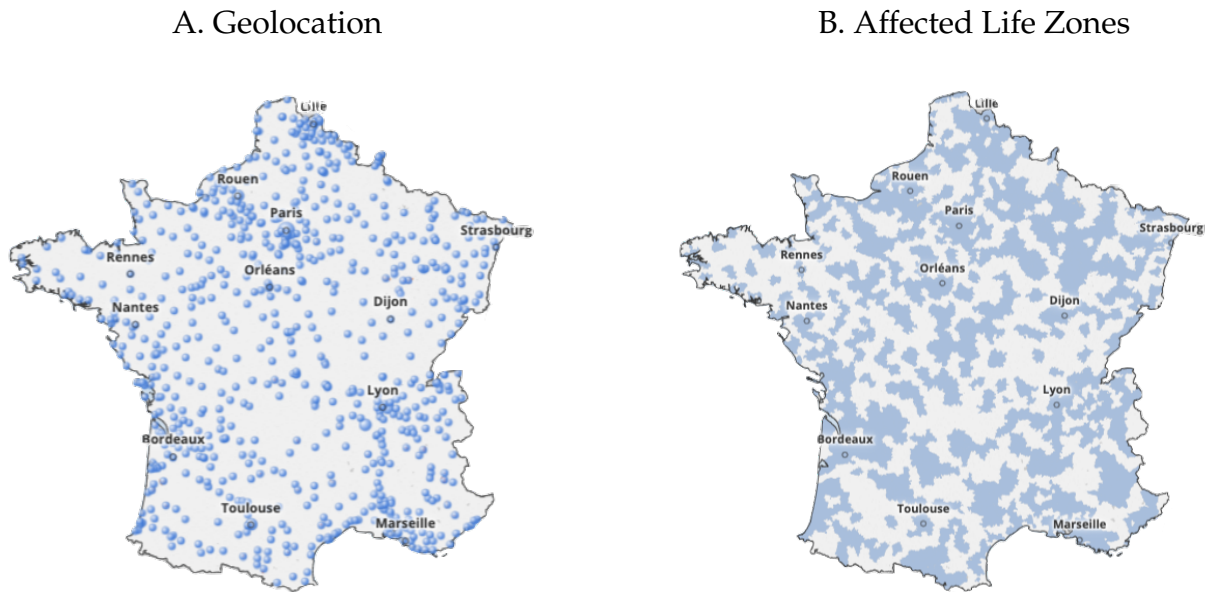
To proxy offline mobilization, we collected a map of planned roadblocks on the evening of November 16<sup>th</sup>. The map was downloaded directly from a website created by the protesters to coordinate demonstrations and roadblocks. This map documented 788 announced roadblocks in metropolitan France (see Panel A in Figure 2.1), which all pointed to precise road infrastructures (e.g., freeway access ramps, parking lots, but primarily roundabouts) and included specific descriptions of the planned events.<sup>6</sup> Many places were chosen for their potential to block traffic and economic activity. To analyze how France was affected by the roadblocks, we use the country's partition in "Life Zones" (hereafter, LZ). They are administrative units defined as the smallest geographical units in which residents can access basic infrastructure and services and conduct a large part of their daily lives. We observe that 551 out of the 1,632 LZs in France experienced a roadblock. They correspond to more than half of the country's population and, as shown in Panel B in Figure 2.1, to a sizable fraction of the French territory.

Finally, to document online mobilization, we searched for all public Facebook groups related to the movement. Using the methodology of [Gaby and Caren \(2012\)](#),

---

<sup>6</sup>Note that these are declarations of intent to demonstrate. However, as the map was created to coordinate the roadblocks, there was little incentive to falsely declare an intent to demonstrate. Note also that the map was updated in real-time.

Figure 2.1: Blocking Half of France at First Try



Notes: Panel A displays the geolocation of the 11/17 roadblocks. Panel B displays the life zones with at least one roadblock on 11/17. These life zones gather 49 million people, 77% of the French mainland population.

we compiled a list of the Facebook groups that were still active one month after 11/17 by performing search requests using a large set of keywords linked to the movement. We recorded each group's name, creation date, number of members, and publications. We identified 3,033 groups with a total of over four million members. Over two-thirds of the groups were associated with a geographical area, and more than 40% of the total members belonged to these localized groups. Moreover, only 20% of the posts emanated from national groups, suggesting that localized groups were the most active ones.

Using a similar method, we also identified 617 Facebook pages and used Netvizz (Rieder 2013) to retrieve their content in March 2019: posts, comments, and interactions (such as likes and shares). This corpus features 120,227 posts, 2.1 million comments, 2.8 million sentences, and 21 million interactions. Netvizz did not provide user identifiers associated with each message. To build a panel of Facebook users, we scraped Facebook a second time in January 2022 and collected additional basic user information. This allows us to study the radicalization of Facebook content for a sample of 120,463 users in Section 4.<sup>7</sup>

<sup>7</sup>To protect users' privacy, all users were de-identified. Approximately 30% of pages had been deleted by January 2022. On the remaining pages, we retrieved 46% of the original posts and 18% of the original comments for this second data retrieval (see Appendix Table A.2). To control for selection bias, we extensively compared both datasets. They are similar in terms of their distribution of political language and in terms of the topics discussed. They also display qualitatively similar trends in our three main outcome variables in Section 4.3, though the second dataset generally displays larger increases in

In Figure 2.2, we depict the daily time series of the number of petition signatories, the number of Facebook group creations, and the number of comments on Facebook pages. The movement culminated in the streets during the first episode of roadblocks. While the petition was mostly signed before 11/17, there were two distinct episodes of group creation: a small one in the weeks before the roadblocks and a large one shortly afterward. This pattern suggests that Facebook groups were used to organize the roadblocks but also served as virtual meeting places that allowed the movement to continue after an initial mobilization in the streets. The evolution of the intensity of the discussions on dedicated Facebook pages corroborates this hypothesis. Discussions gained importance in January 2019 and, contrary to the weekly number of protesters, remained strongly active during the following months.

### 2.3 Textual Analysis of Facebook Discussions

To analyze discussions on Facebook pages, we rely on NLP methods (see, for an overview, [Grimmer and Stewart 2013](#); [Gentzkow, Kelly and Taddy 2019](#)): a topic model, a sentiment analysis, and a political classification of the messages. The technical details of our implementation are provided in Appendix C.

To identify the topics discussed online by the Yellow Vests, we rely on a topic model tailored to analyze short text snippets ([Demszky et al. 2019](#)). In our main specification, we allow for 15 different topics, but qualitatively similar results are found with alternative numbers of clusters. We display the topics obtained in Figure 2.3. We can group topics into different categories, such as protest organization (A and B), socialization (C and G), and online mobilization (D). Other topics reflect the reasons behind the protests and the political goals the Yellow Vests were trying to achieve (E and F). Finally, five topics refer to antagonistic messages (H, I, J, K, and L) and reflect the protesters' anger toward government officials and their policies.

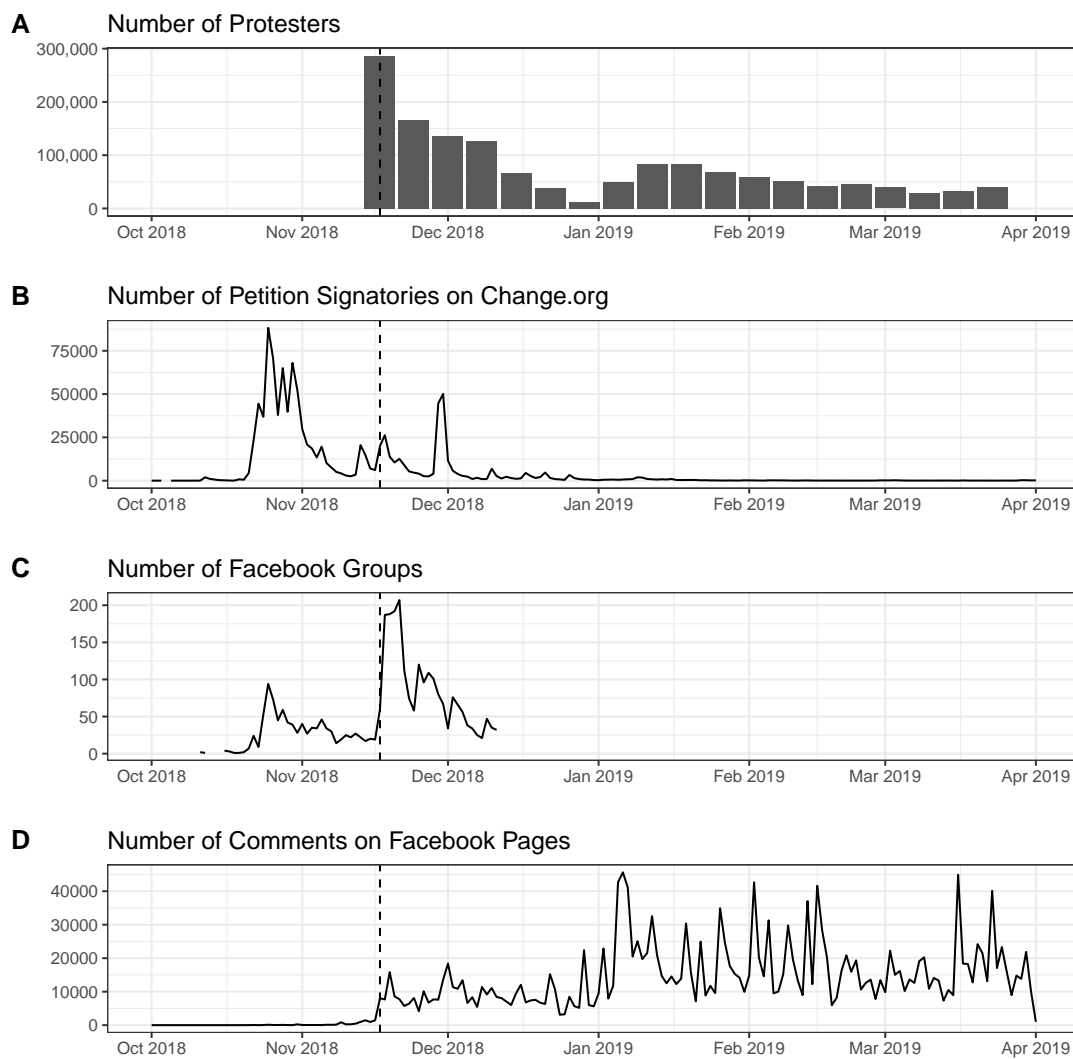
To measure emotional content in messages, we use a dictionary-based approach that assigns to each sentence a sentiment score ranging from -1 (very negative) to 1 (very positive). Figure 2.3 splits the 15 topics between those with an average sentiment of messages above zero and those below zero. The five topics we classify as antagonistic are all associated with negative sentiment. Finally, to understand the political stance of messages, we build a measure of partisanship using a supervised learning model based on tweets from parliamentarians.

---

radical attitudes.



Figure 2.2: Evolution of Online and Offline Mobilizations

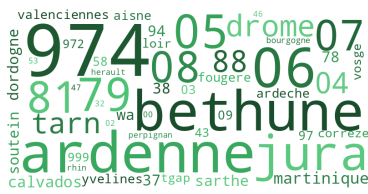


*Notes:* In Panel A, we show the number of demonstrators reported weekly by the Ministry of the Interior. In Panel B, we plot the daily number of petition signatures. In Panel C, we plot the daily number of new Facebook groups created. Finally, in Panel D, we plot the daily number of messages posted on Facebook pages. The vertical dashed line in all panels corresponds to 11/17.

### 3. The Online-Offline Feedback Loop

This section shows that online and offline mobilizations reinforce each other. We estimate spatial regressions where we distinguish between mobilization before and after the 11/17 protests. We first document that the 11/17 protests were organized in localities with higher early online mobilization. This is expected – as most blockades were organized online – and consistent with previous findings showing the facilitating role of social media in organizing protests. However, the reverse direction of the loop (from street protests to further online mobilization) is novel and warrants deeper inves-

Figure 2.3: Results of the Topic Model



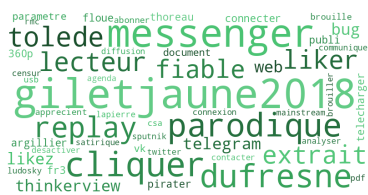
A. Places [2.9%]



B. Towns and hours [4.4%]



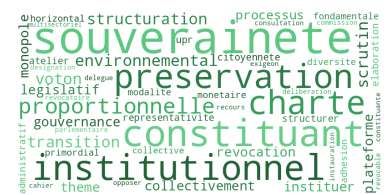
C. Support [3.9%]



D. Diffusion [4.8%]



E. Economic concerns [5.9%]



F. Political institutions [7.6%]



G. Food and objects [6%]



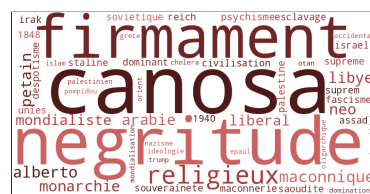
H. Critiques [6%]



I. Insults [4.5%]



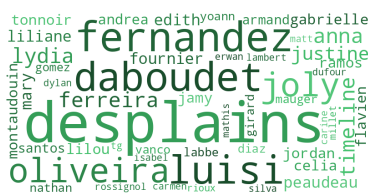
J. Violence [5.9%]



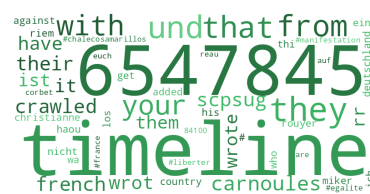
K. Conspiracy [5.9%]



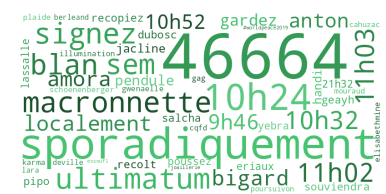
L. Actions [7%]



M. Names [6.6%]



N. Foreign Languages [8.4%]



O. Other [20%]

Notes: This figure shows wordclouds associated with the fifteen topics we identify in our corpus. The size of words is determined by a term frequency-inverse document frequency (TF-IDF) metric, where each document is the entire collection of sentences associated with a given topic. This metric gives higher scores to words that are (i) more frequent in the corpus and (ii) particularly meaningful for each topic. Wordclouds are boxed inside a rectangle when the average sentiment of messages in the topic is negative. Squared brackets indicate the topic frequency (computed as the share of total messages in the corpus).

tigation. While the time series depicted in Figure 2.2 does suggest that new Facebook groups were created in the immediate aftermath of the 11/17 protests, this pattern may simply result from the fact that protesters were simultaneously looking for both online and offline ways of expressing their discontent. Therefore, we use an instrumental variables strategy to establish a causal relationship between offline and online mobilizations.

### 3.1 Empirical Framework

We construct a dataset at the municipality level (indexing observations by  $m$ ).<sup>8</sup> There are more than 35,000 municipalities in France, and their boundaries date back to the French Revolution.<sup>9</sup> They are the lowest government level, allowing us to gather data on various characteristics. We complement those variables with a novel measure of Facebook penetration (see Appendix A.3 for details). Early online mobilization is defined by the vector  $M_m^{\text{pre-11/17}}$ , which includes the signature rate in the municipality before 11/17 and the number of local Facebook groups before 11/17.

The 11/17 protests are measured by  $B_m^{11/17}$ , a dummy variable equal to 1 if there was a roadblock in municipality  $m$  on 11/17. Finally, later mobilization is measured by  $M_m^{\text{post-11/17}}$ , which includes the signature rate in the municipality after 11/17, the number of groups created after 11/17, as well as the number of members and publications observed in those later groups, expressed in per-capita terms.

We document the relationship between  $M_m^{\text{pre-11/17}}$  and  $B_m^{11/17}$  by measuring simple correlations based on the OLS estimation of Equation 2.1:

$$B_m^{11/17} = M_m^{\text{pre-11/17}} \beta + X_m \gamma + \delta_{LZ(m)} + \varepsilon_m \quad (2.1)$$

where  $X_m$  is a large set of economic, geographic, demographic, and political controls (see the full list in Table B.1) and  $\delta_{LZ(m)}$  is a LZ fixed effect to account for fixed unobserved heterogeneity at a higher spatial level.

Conversely, to gauge the impact of  $B_m^{11/17}$  on  $M_m^{\text{post-11/17}}$ , we use 2SLS estimation. For a set of municipal characteristics  $Z_m$  that can serve as instruments of the roadblock probability, we can estimate the following first stage, predicting the probability of a

<sup>8</sup>Some variables were only available at higher geographical levels. When relevant, we apportioned them according to municipal population.

<sup>9</sup>Excluding French overseas territories and Corsica from our sample leaves us with 34,434 municipalities.

roadblock in a municipality:

$$B_m^{11/17} = \alpha_1 + M_m^{\text{pre-11/17}} \beta_1 + X_m \gamma_1 + \delta_1^{\text{LZ}(m)} + Z_m \zeta + \varepsilon_m \quad (2.2)$$

and the following second stage, regressing a measure of online mobilization after 11/17,  $M_m^{\text{post}}$ , on the predicted roadblock probability from Equation 2.2:

$$M_m^{\text{post}} = \alpha_2 + M_m^{\text{pre}} \beta_2 + X_m \gamma_2 + \delta_2^{\text{LZ}(m)} + \eta \widehat{B_m^{11/17}} + \epsilon_m, \quad (2.3)$$

The coefficient  $\eta$  then provides us with the local average treatment effect of an 11/17 roadblock on subsequent online mobilization.

**Easy-to-block locations: Roundabouts.** To instrument the roadblocks, we leverage the presence of roundabouts in each municipality. The rationale for the relevance of this instrument is that calls for demonstrations urged protesters to block roundabouts. By design, they allow to block several roads at a time and possess a central median strip on which it is convenient to set camp.<sup>10</sup> The identifying assumption is that conditional on observable characteristics, the distribution of roundabouts only predicts future online mobilization through its impact on roadblocks. The history of roundabouts makes it likely that the conditional distribution of local roundabout density reflects local idiosyncrasies. Roundabouts are partly a French architectural fad, arguably invented in 1906 by the French urban planner Eugène Hénard. France has over sixty-thousand roundabouts (roughly four times more than the United Kingdom). One-third of French municipalities have at least one. While plausible road safety reasons support their use, they can almost always be replaced with traffic lights.

In support of our exclusion restriction, Appendix Table B.2 explores the determinants of the spatial distribution of roundabouts in France. As one can expect, the distribution of roundabouts in France is closely related to the population distribution: it explains more than 40% of the variance in roundabout density. LZ fixed-effects only have little explanatory power, indicating low levels of spatial auto-correlation in roundabout density. Importantly, other controls only explain a residual fraction of the variation in roundabout density after controlling for the population density. Finally, a map of the prediction error of roundabout density after an OLS regression, including our controls, shows a seemingly random distribution (see Appendix Figure B.1).

<sup>10</sup>There always was a conflict in urban planning between favoring policing (such as during Haussmann's renovation of Paris in the XIX<sup>th</sup> century (Lefebvre 1968)) and favoring social contact. Car-based urban planning is generally viewed as on the policing side (Davis 1992). In that regard, the occupation of roundabouts by the Yellow Vests represents an ironic turn of events.

Assuming the exogeneity of this first instrument, we can leverage a second instrument, which will allow us to test overidentifying restrictions. Indeed, since organizing a roadblock requires significant manpower, protesters had to coordinate to choose roadblock locations. This spatial coordination problem suggests another instrument, which is the mirror image of the first: the density of roundabouts in the other municipalities of the LZ. Because of competition between easy-to-block locations, we expect municipalities surrounded by more roundabouts to be less likely blocked.

### 3.2 Protests Were Organized Online and Targeted Roundabouts

Table 3.1 presents OLS estimates of Equations 2.1 and 2.2. On top of estimates for  $\beta$ , we show estimates for the coefficients on the Facebook penetration rate. Column (1) displays our results without controlling for early mobilization measures. In line with a vast body of evidence on the role of social media in social movements, we find a significant and positive correlation between Facebook penetration and the occurrence of a roadblock. Column (2) shows that the petition signature rate is positively correlated with the occurrence of a roadblock, which suggests that signature rates may be interpreted as a signal for mobilization potential. More importantly, Column (3) shows that, as expected, early mobilization on Facebook is strongly correlated with the occurrence of a roadblock. In addition, when we include the number of Yellow Vests Facebook groups associated with a municipality in the regression, the coefficient associated with the Facebook penetration rate drops by 34%.

Column (4), where we control for both the petition signature rate and the number of groups, shows that the coefficients on the signatures and groups are stable compared to Columns (2) and (3), which suggests that both types of online mobilization are not substitutes for one another. As shown in Column (5), a model that would only control for the existence of those measures would have a predictive power equal to over 60% of that of the model with the full set of municipal covariates but without any Yellow-Vest specific controls.

Finally, Column (6) confirms that roundabouts played an essential role in organizing the protests: increasing the density of roundabouts in a municipality by one standard deviation increases the probability of a roadblock by 1.1 p.p. In addition, an increase of one standard deviation in the density of roundabouts in surrounding municipalities decreases the roadblock probability of a municipality by 8.3 p.p. Both variables are statistically significant at the 99.9% level.

Table 3.1: Predictors of a Roadblock in a Municipality

	(1)	(2)	(3)	(4)	(5)	(6)
Facebook penetration	0.462*** (0.057)	0.476*** (0.057)	0.304*** (0.060)	0.317*** (0.061)		0.301*** (0.059)
Signatures (pre-17/11)		0.007*** (0.001)		0.006*** (0.001)	0.007*** (0.001)	0.006*** (0.001)
Nb. groups (pre-17/11)			0.036*** (0.003)	0.035*** (0.003)	0.056*** (0.001)	0.034*** (0.003)
Roundabouts (municipality)						0.011*** (0.003)
Roundabouts (LZ)						-0.083*** (0.017)
Controls	✓	✓	✓	✓		✓
Life zone FE	✓	✓	✓	✓		✓
Observations	34,449	34,449	34,449	34,449	34,475	34,434
R-squared	0.24	0.24	0.27	0.27	0.15	0.28

*Notes:* This table shows OLS estimates for a linear probability model predicting whether a municipality experienced a roadblock or not, as formalized in Equation 2.1. “Signatures (pre-11/17)” is the municipality’s signature rate of the Change.org petition before 11/17; “Nb. Groups (pre-11/17)” is the apportioned number of Facebook groups (from all geographical levels) created before 11/17. These two variables are standardized. We measure Facebook penetration in a municipality as the number of Facebook users who declare to live or come from that municipality, divided by the municipal population. This variable is standardized and divided by 100. The last column represents the first-stage estimates of Equation 2.3, associated with the second-stage results of Table 3.2. The two instruments we use are the number of roundabouts per squared kilometer in the municipality and the corresponding average for all other municipalities in the LZ. Both variables are standardized. Standard errors are clustered at the LZ level. \*:  $p < 0.1$ , \*\*:  $p < 0.05$ , \*\*\*:  $p < 0.01$ .

### 3.3 Roadblocks Spurred Further Online Mobilization

Table 3.2 presents results for the second stage. The Kleibergen-Paap F-statistic equals 25, suggesting that our instruments are reasonably strong. In addition, the high p-values associated with the Hansen J-statistics indicate that we fail to reject the hypothesis that the overidentifying restrictions are valid. Column (1) shows that even though the bulk of petition signatures occurred before 11/17, having a roadblock increases the post-11/17 signature rate by 1.2 standard deviations. This result suggests that protests helped spread information about the Yellow Vests’ demands at the end of 2018 when public support for the movement was still high. The previous signatory rate is also correlated with subsequent signatory dynamics.

We also find a strong positive impact of roadblocks on subsequent Facebook activ-

Table 3.2: Effects of a Roadblock on Post-11/17 Online Mobilization

	(1)	(2)	(3)	(4)
	Signatures	Groups	Members	Posts
Blockade	1.158*** (0.253)	2.925*** (0.692)	0.214** (0.090)	0.138** (0.065)
Signatures (pre-17/11)	0.599*** (0.022)	-0.001 (0.006)	0.001 (0.008)	-0.002 (0.007)
Groups (pre-17/11)	-0.036*** (0.010)	0.036 (0.036)	-0.001 (0.007)	-0.001 (0.006)
Facebook penetration	-0.440 (0.357)	3.243*** (0.482)	5.912** (2.675)	4.883* (2.753)
Controls	✓	✓	✓	✓
Life zone FE	✓	✓	✓	✓
Observations	34,434	34,434	34,434	34,434
Kleibergen-Paap F-stat	24.8	24.8	24.8	24.8
p-value Hansen	0.664	0.538	0.343	0.310

*Notes:* This table shows 2SLS estimates of the impact of a municipal roadblock on four measures of online mobilization after 11/17: the signature rate of the Change.org petition after 11/17 (column 1), the number of groups created post-11/17 (column 2), the number of members per inhabitant (column 3) and posts per inhabitant (column 4) in these newly created groups. Estimates of the first stage are displayed in column (6) in Table 3.1. All outcome variables are standardized. We cluster standard errors at the LZ level. \*:  $p < 0.1$ , \*\*:  $p < 0.05$ , \*\*\*:  $p < 0.01$ .

ity: a roadblock in a municipality increases the number of new local Facebook groups by 2.9 standard deviations (corresponding to 1.2 additional groups), which translates into an increase in the number of new members per inhabitant by 0.21 standard deviations, and in the number of posts per inhabitant by 0.14 standard deviations. These three measures of later mobilization on Facebook appear uncorrelated with early online mobilization but are positively correlated with the Facebook penetration rate.

Our results are robust to several specification changes (see Appendix Table B.3). In particular, the estimates are stable if we do not include controls (Panel A) or use only one roundabout instrument instead of two (Panels B and C). These three tests are reassuring regarding the validity of the exclusion restriction. Effects are also reasonably similar if we define location fixed effects and the instrument at the commuting zone ( $N = 297$ ) rather than at the LZ ( $N = 1606$ ) level (Panel D) or if we exclude the Paris region, which stands out along many dimensions (Panel E).

Overall, our results suggest that the 11/17 roadblocks contributed to the expansion of online activity. Roadblocks triggered a new wave of popularity for the Change.org petition and led to a consolidation of the online infrastructure of the Yellow Vests

movement on Facebook. While street protests subsided, protesters who had met in the streets relied on Facebook to continue their discussions.

## 4. The Rise of Online Radicalism

In this section, we document the evolution of online discussions among protesters through the textual analysis of the Yellow Vests Facebook pages. Our data allows us to follow the content of Facebook pages between the end of October 2018 and the beginning of April 2019. We first document the evolution of online discussions. We then use the distinction between posts and comments to study the role of Facebook's recommender algorithm in structuring online discussions. Finally, as we observe the messages of individual members over time, we study how the composition of the population of discussants changed and whether some individuals radicalized over the period.

### 4.1 Online Discussions Became More Radical Over Time

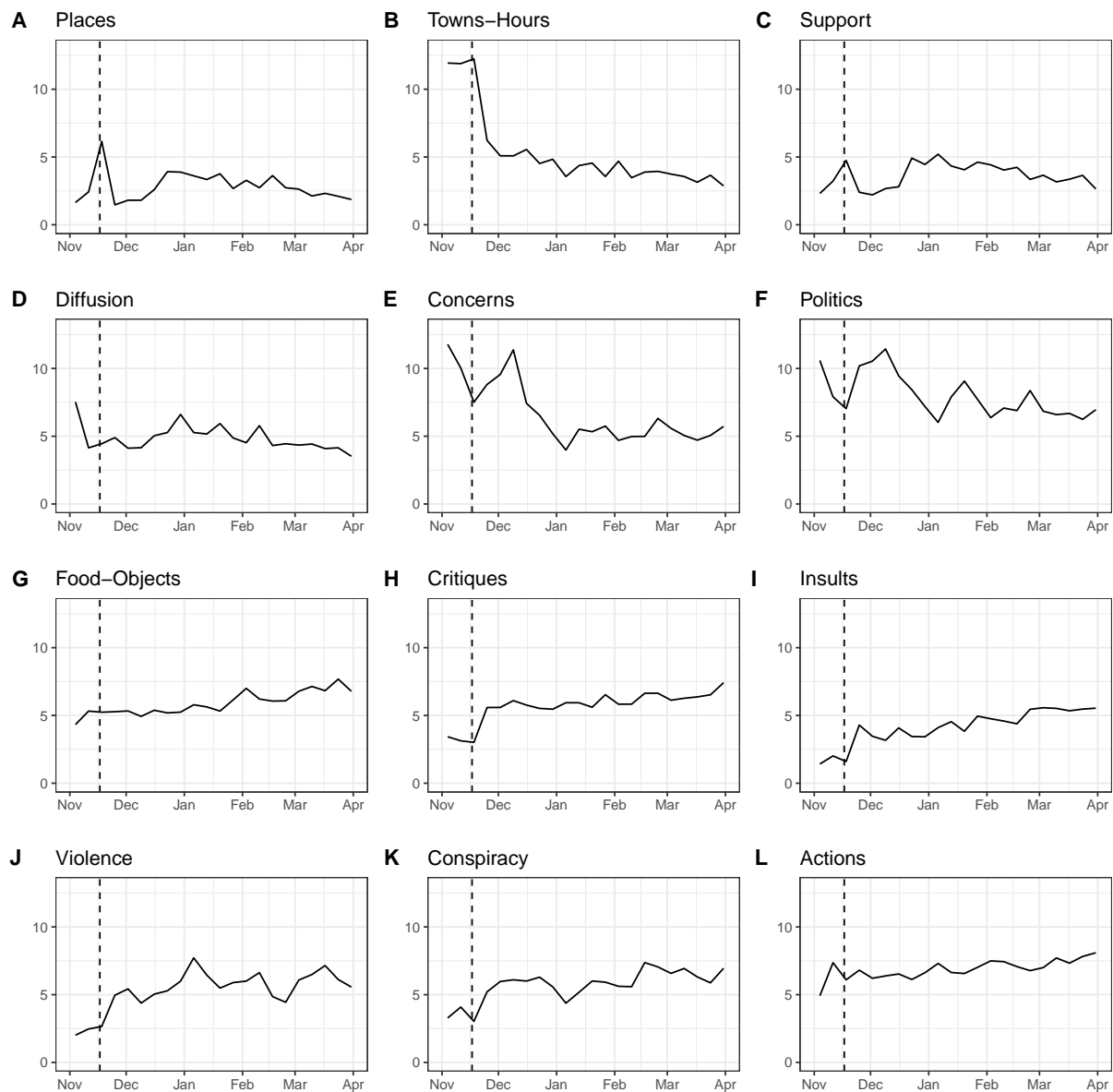
Figure 4.1 shows that the share of messages associated with political or economic concerns declined while messages of violence, conspiracy theories, and insults became more widespread. Overall, the share of messages associated with antagonistic content (topics H to L) increased by 15 p.p. between November 2018 and March 2019. Other classifications of Facebook messages reflect similar trends over the period: the share of messages classified as negative (resp. associated with a far-right or far-left party) increased by 8 p.p. (resp. 6 p.p.).

Altogether, the content of online discussions became more antagonistic, negative, and polarized. We define these three concurrent dynamics as a “rise of online radicalism”. Several non-exclusive reasons may explain this pattern. In particular, one may think of external reasons: on the positive side, the movement succeeded in having the government revert its policy so that economic concerns were less in need to be discussed. On the negative side, some protests were quite violent and met with police repression. In this context, it is no wonder that topics related to violence or insults rose to prominence.

In what follows, we describe two pieces of quantitative evidence suggesting that part of this increase in online radicalism may be due to specific online processes. First, we show that discussants were disproportionately exposed to radical content by Facebook. Second, we show that new, more radical discussants progressively replaced the



Figure 4.1: Topic Shares in Facebook Discussions Over Time



*Notes:* This figure shows weekly shares of the twelve topics of interest shown in Figure 2.3. For all topics, the vertical dashed line corresponds to 11/17. The share of messages associated with violence is below 2.5% in early November and is consistently above 5% after December 10.

more moderate ones.

## 4.2 Facebook's Algorithm Increased Exposure to Radical Content

A first way to assess how radical content spread on the Yellow Vests Facebook pages is to look at the structure of online discussions, which involve an initial post and its associated comments. To assess whether Facebook played an active role in the type of content that the Yellow Vests were exposed to online, we focus on the organization of

Table 4.1: Comments' Rank and Radical Content

	Rank of the Comment (in log)				Comment is Among First Four			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Antagonistic Topic	-0.095*** (0.005)			-0.081*** (0.004)	0.004*** (0.000)			0.003*** (0.000)
Extreme Parties		-0.028*** (0.003)		-0.017*** (0.003)		0.002*** (0.000)		0.002*** (0.000)
Negative Sentiment			-0.085*** (0.005)	-0.065*** (0.004)			0.004*** (0.000)	0.003*** (0.000)
Chronological Order	0.127*** (0.005)	0.126*** (0.005)	0.126*** (0.005)	0.127*** (0.005)	0.014*** (0.003)	0.014*** (0.003)	0.014*** (0.003)	0.014*** (0.003)
Post Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R-Squared	0.81	0.81	0.81	0.81	0.48	0.48	0.48	0.48

*Notes:* This table shows estimates of OLS regressions at the sentence level (N=1,881,976). We restrict the text corpus to comments (and exclude original posts). Some comments are made of several sentences, but results are similar if we restrict the sample to single-sentence comments (61% of the sample). In Columns (1) to (4), the dependent variable is the (log) rank of the comment suggested by Facebook at the time of the scrape. In Columns (5) to (8), the dependent variable is a dummy variable equal to 1 if the comment is among the first four comments suggested by Facebook at the time of the scrape. “Antagonistic Topic” is a dummy variable equal to 1 if the sentence is classified as belonging to an antagonistic topic. “Extreme Parties” is a dummy variable equal to 1 if the sentence is attributed to an extreme party. “Negative Sentiment” is a dummy variable equal to 1 if the sentence is associated with a negative sentiment value. “Chronological Order” is defined as the counterpart of the dependent variable, based on chronological order: the (log) rank of the comment based on chronological order in Columns (1) to (4), and a dummy variable equal to 1 if the comment was among the first four to be posted in Columns (5) to (8). All specifications include a post fixed effect. In all regressions, we cluster standard errors at the post level. \*:  $p < 0.01$ , \*\*:  $p < 0.05$ , \*\*\*:  $p < 0.1$ .

discussions by Facebook’s platform.

While Facebook displays posts chronologically on Facebook pages, it does not deal with their associated comments similarly. Instead, undisclosed algorithms rank comments by what the platform calls “relevance.” Since our dataset contains information on the ordering of comments shown to users at the time of the scrape, we can assess whether our radicalization measures are correlated with the recommendations of Facebook’s algorithm. To that end, we regress the rank of each comment in our text corpus on our measures of radicalism. Since posts vary a lot in their content and the number of comments they generate, we control for post fixed effects.<sup>11</sup> We also control for a measure of the rank of the comment based on the time when the comment was posted.

Results are displayed in Table 4.1. Measures of rank based on Facebook suggestions and time of posting are positively correlated. However, the estimates are quite far from unity, which shows that Facebook strongly alters the original ordering of comments. In

<sup>11</sup>Out of our original corpus of 120,227 posts, we focus on the 35,828 with at least two comments. Estimates are qualitatively similar but slightly higher if we do not include post fixed effects.

particular, comments associated with our radicalization measures are more likely to be found higher on the list. For example, comments associated with antagonistic topics are displayed at a rank between 8 and 10% higher than other comments. The same patterns appear if we focus on the probability of being a “star comment”, which we take as one of the first four comments below the post (10% of our estimation sample). Such comments are likely to appear in users’ newsfeeds without further clicking and are, therefore, much more likely to be salient and read by users. These results show that a chronological order of comments would have provided discussants with less radical content.

### 4.3 Radical Discussants Replaced the Moderates

Filter bubbles that push radical comments on the front page are not the only force driving radicalization. Another specificity of online communities is that participation costs are quite low, and participation may be volatile.<sup>12</sup> As such, online discussions are susceptible to being hijacked by a minority of radical users, who drive others away and attract similarly-minded discussants (Sunstein 2017). To test this mechanism, we decompose the rise in online radicalism between two margins. First, moderate users could have progressively left the movement or been replaced by more radical ones. We refer to this depletion effect as the “extensive margin” of radicalization. Alternatively, active users may have radicalized over time. We refer to such individual changes as the “intensive margin” of radicalization.

To assess whether the trends we observe are more likely to reflect shifts in the intensive or extensive margin of radicalization, we exploit the panel dimension of the data and the fact that we can follow (de-identified) individual Facebook users over time. To identify the intensive margin of radicalization, we can evaluate whether the average user became increasingly likely to post radical messages. To identify the extensive margin, we can evaluate whether the pool of active users becomes increasingly populated with users who (on average) post more radical messages. To this end, we estimate the following fixed-effects equation:

$$Y_s = \delta_{i(s)} + \gamma_{t(s)} + \varepsilon_s, \quad (2.4)$$

where  $Y_s$  is a measure of radicalism of sentence  $s$ ,  $\delta_{i(s)}$  is a fixed effect associated with the user  $i$  who posted sentence  $s$ , and  $\gamma_{t(s)}$  is a fixed effect associated with the month

<sup>12</sup>Measuring participation in online communities is notoriously difficult, because of the presence of non-active participants (Malinen 2015).

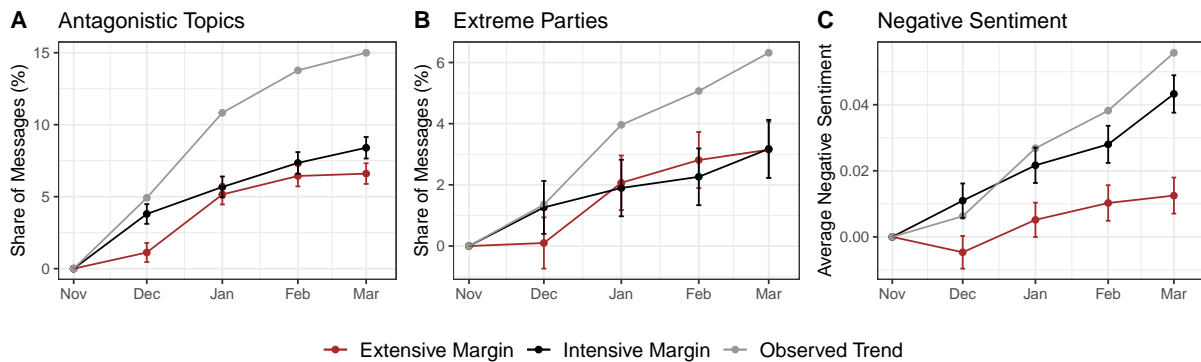
$t$  during which sentence  $s$  was posted. Intuitively,  $\delta_{i(s)}$  measures user  $i$ 's propensity to post radical sentences, and  $\gamma_{t(s)}$  accounts for the additional propensity of users to post radical sentences during month  $t$ .

We can then leverage estimates of user and time fixed effects to decompose the rise of online radicalism into an intensive and extensive margin. Indeed, the average level of radicalism during month  $t$ ,  $\bar{Y}_t$ , can be expressed as:

$$\bar{Y}_t = \underbrace{\hat{\mathbb{E}}_t [\hat{\delta}_i]}_{\text{Extensive margin}} + \underbrace{\hat{\gamma}_t}_{\text{Intensive margin}}, \quad (2.5)$$

where  $\hat{\mathbb{E}}_t [\hat{\delta}_i] = \sum_i s_{i,t} \hat{\delta}_i$  and  $s_{i,t}$  is the share of sentences posted during month  $t$  that originated from user  $i$ . Hence, the first term of expression 2.5 corresponds to the average propensity to post radical sentences for users active during the month  $t$ . An increase of this term over time means that the share of sentences posted by more radical users increases. An increase in the second term of expression 2.5 corresponds to an increase in the propensity of any given user to post a radical sentence at a given time.

Figure 4.2: Extensive and Intensive Margins of Radicalization



*Notes:* This figure decomposes the increase in online radicalism using Equation 2.5. Panel A presents estimates for the probability of posting a sentence associated with an antagonistic topic. Panel B presents estimates for the probability of writing a sentence associated with a politically extreme party (i.e., on the far left or the far right). Panel C presents estimates for negative sentiment. We compute standard errors via bootstrap with 200 iterations and plot confidence intervals at the 95% confidence level.

Figure 4.2 presents a decomposition of our three radicalization measures using Equation 2.5. In Panel A, the outcome variable is a dummy variable that indicates whether a message was associated with an antagonistic topic. In Panel B, the outcome variable is a dummy variable that indicates whether a message was associated with an extreme political party. In Panel C, the outcome variable is the negative sentiment score associated with a sentence, taking values between -1 and 1. For all three dependent variables, our decomposition exercise suggests that both margins contributed to the radicalization of Facebook content. In addition, both margins played a quantita-

tively similar role in two out of three measures, although the impact of the extensive margin seems slightly delayed compared to that of the intensive margin.

We interpret this finding on the role of the extensive margin of radicalization as supporting evidence that participation in online communities is quite volatile and prone to being taken over by more radical discussants over time. However, since we lack information on offline protesters, we cannot tell whether this pattern is specific to online mobilization. In addition, our linear decomposition does not allow us to study the interplay between both margins. For example, the radicalization of other discussants may have led to the departure of moderate discussants, or, conversely, remaining discussants may have faced lower moderation from fellow discussants over time. Further understanding the interplay between online and offline sources of radicalization would require having access to individual newsfeeds, combined with detailed accounts of protesting activity.

## 5. Conclusion

Large protest movements are now a combination of online and offline mobilizations. Many have noted that social media favor the emergence and spread of protests by lowering coordination costs and making it easier to signal discontent. Our study confirms that social media and online protests likely increase the number of street protesters. Moreover, we provide novel evidence that real-life demonstrations may also intensify subsequent online engagement within the same region, prolonging protest movements' lifespan. However, these persistent online communities are subject to radicalization, and we provide several pieces of evidence suggesting that the social media infrastructure itself may drive part of this radicalization process.

Put together, our findings highlight one core tension of hybrid social movements: on the one hand, social media allow online communities to arise from everywhere, even remote locations such as rural areas or suburban fringes; on the other hand, a strong dependency on a leaderless social media infrastructure presents risks of radicalization and may dampen the ability to structure long-lasting, effective political campaigns.

# Bibliography

- Acemoglu, Daron, Tarek A Hassan, and Ahmed Tahoun.** 2018. "The Power of the Street: Evidence From Egypt's Arab Spring." *Review of Financial Studies*, 31(1): 1–42.
- Algan, Yann, Elizabeth Beasley, Daniel Cohen, and Martial Foucault.** 2019. *Les Origines Du Populisme*. La République des Idées. Seuil.
- Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky.** 2018. "Exposure to opposing views on social media can increase political polarization." *Proceedings of the National Academy of Sciences of the United States of America*, 115(37): 9216–9221.
- Bastos, Marco T., Dan Mercea, and Arthur Charpentier.** 2015. "Tents, Tweets, and Events: The Interplay Between Ongoing Protests and Social Media." *Journal of Communication*, 65(2): 320–350.
- Bendali, Zakaria, and Aldo Rubert.** 2020. "Les Sciences Sociales en Gilet Jaune." *Politix*, 4: 177–215.
- Bennett, W Lance, and Alexandra Segerberg.** 2012. "The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics." *Information, Communication & Society*, 15(5): 739–768.
- Borge-Holthoefer, Javier, Alejandro Rivero, Iñigo García, Elisa Cauhé, Alfredo Ferrer, Darío Ferrer, David Francos, David Iniguez, María Pilar Pérez, Gonzalo Ruiz, et al.** 2011. "Structural and Dynamical Patterns on Online Social Networks: The Spanish May 15th Movement as a Case Study." *PloS One*, 6(8): e23883.
- Boyer, Pierre C., Thomas Delemotte, Germain Gauthier, Vincent Rollet, and Benoît Schmutz.** 2020. "The Origins of the Gilets Jaunes Movement." *Revue Économique*, 71(1): 109–138.
- Branton, Regina, Valerie Martinez-Ebers, Tony E Carey Jr, and Tetsuya Matsubayashi.** 2015. "Social Protest and Policy Attitudes: The Case of the 2006 Immigrant Rallies." *American Journal of Political Science*, 59(2): 390–402.

- Bursztyn, Leonardo, Davide Cantoni, David Yang, Noam Yuchtman, and Jane Zhang.** 2021. "Persistent Political Engagement: Social Interactions and the Dynamics of Protest Movements." *American Economic Review: Insights*, 3(2): 233–50.
- Castells, Manuel.** 2015. *Networks of Outrage and Hope: Social Movements in the Internet Age*. Wiley.
- Clarke, Killian, and Korhan Kocak.** 2020. "Launching Revolution: Social Media and the Egyptian Uprising's First Movers." *British Journal of Political Science*, 50(3): 1025–1045.
- Cointet, Jean-Philippe, Pedro Ramaciotti Morales, Dominique Cardon, Caterina Froio, Andrei Mogoutov, Benjamin Ooghe, and Guillaume Plique.** 2021. "De Quelle(s) Couleur(s) Sont Les Gilets Jaunes? Plonger Des Posts Facebook Dans Un Espace Idéologique Latent." *Statistique Et Société*, 9(1-2): 79–107.
- Davis, Mike.** 1992. *City of Quartz: Excavating the Future in Los Angeles*. Vintage Books.
- Della Porta, Donatella.** 2013. *Can Democracy Be Saved?* Polity Press.
- Della Porta, Donatella.** 2018. "Radicalization: A Relational Perspective." *Annual Review of Political Science*, 21: 461–474.
- Della Porta, Donatella, and Mario Diani.** 2020. *Social Movements: An Introduction*. Wiley-Blackwell.
- Demszky, Dorottya, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky.** 2019. "Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings." *arXiv preprint arXiv:1904.01596*.
- Earl, Jennifer, and Katrina Kimport.** 2011. *Digitally Enabled Social Change: Activism in the Internet Age*. MIT Press.
- Earl, Jennifer, Thomas V. Maher, and Jennifer Pan.** 2022. "The digital repression of social movements, protest, and activism: A synthetic review." *Science Advances*, 8(10): eabl8198.
- Enikolopov, Ruben, Alexey Makarin, and Maria Petrova.** 2020. "Social Media and Protest Participation: Evidence From Russia." *Econometrica*, 88(4): 1479–1514.
- Fergusson, Leopoldo, and Carlos Molina.** 2019. "Facebook Causes Protests." *Documento CEDE*, (41).
- Fletcher, Richard, Alessio Cornia, and Rasmus Kleis Nielsen.** 2020. "How Polarized Are Online and Offline News Audiences? A Comparative Analysis of Twelve Countries." *The International Journal of Press/Politics*, 25(2): 169–195.

- Frye, Timothy, and Ekaterina Borisova.** 2019. "Elections, Protest, and Trust in Government: A Natural Experiment From Russia." *Journal of Politics*, 81(3): 820–832.
- Gaby, Sarah, and Neal Caren.** 2012. "Occupy Online: How Cute Old Men and Malcolm X Recruited 400,000 US Users to OWS on Facebook." *Social Movement Studies*, 11(3-4): 367–374.
- Gaffney, Devin.** 2010. "#IranElection: Quantifying Online Activism." CiteSeer.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy.** 2019. "Text as Data." *Journal of Economic Literature*, 57(3): 535–74.
- Gillion, Daniel Q.** 2012. "Protest and Congressional Behavior: Assessing Racial and Ethnic Minority Protests in the District." *Journal of Politics*, 74(4): 950–962.
- Giugni, Marco G.** 1998. "Was It Worth the Effort? The Outcomes and Consequences of Social Movements." *Annual Review of Sociology*, 24(1): 371–393.
- González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno.** 2011. "The Dynamics of Protest Recruitment Through an Online Network." *Scientific Reports*, 1(1): 1–7.
- Grimmer, Justin, and Brandon M Stewart.** 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*, 21(3): 267–297.
- Jost, John T, Pablo Barberá, Richard Bonneau, Melanie Langer, Megan Metzger, Jonathan Nagler, Joanna Sterling, and Joshua A Tucker.** 2018. "How Social Media Facilitates Political Protest: Information, Motivation, and Social Networks." *Political Psychology*, 39: 85–118.
- Ketchley, Neil, and Thoraya El-Rayyes.** 2021. "Unpopular Protest: Mass Mobilization and Attitudes to Democracy in Post-Mubarak Egypt." *Journal of Politics*, 83(1): 291–305.
- Larson, Jennifer M, Jonathan Nagler, Jonathan Ronen, and Joshua A Tucker.** 2019. "Social Networks and Protest Participation: Evidence From 130 Million Twitter Users." *American Journal of Political Science*, 63(3): 690–705.
- Lefebvre, Henri.** 1968. *Le Droit À La Ville*. Anthropos.
- Le Galès, Patrick.** 2021. "The Rise of Local Politics: A Global Review." *Annual Review of Political Science*, 24(1): 345–363.



- Lipsky, Michael.** 1968. "Protest as a Political Resource." *American Political Science Review*, 62(4): 1144–1158.
- Little, Andrew T, Joshua A Tucker, and Tom LaGatta.** 2015. "Elections, Protest, and Alternation of Power." *Journal of Politics*, 77(4): 1142–1156.
- Loader, Brian D.** 2008. "Social Movements and New Media." *Sociology Compass*, 2(6): 1920–1933.
- Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott.** 2013. "Do Political Protests Matter? Evidence From the Tea Party Movement." *Quarterly Journal of Economics*, 128(4): 1633–1685.
- Malinen, Sanna.** 2015. "Understanding user participation in online communities: A systematic literature review of empirical studies." *Computers in Human Behavior*, 46: 228–238.
- Manacorda, Marco, and Andrea Tesei.** 2020. "Liberation Technology: Mobile Phones and Political Mobilization in Africa." *Econometrica*, 88(2): 533–567.
- Mazumder, Soumyajit.** 2018. "The Persistent Effect of US Civil Rights Protests on Political Attitudes." *American Journal of Political Science*, 62(4): 922–935.
- Meirowitz, Adam, and Joshua A Tucker.** 2013. "People Power or a One-Shot Deal? A Dynamic Model of Protest." *American Journal of Political Science*, 57(2): 478–490.
- Meyer, David S.** 2004. "Protest and Political Opportunities." *Annual Review of Sociology*, 30(1): 125–145.
- Mundt, Marcia, Karen Ross, and Charla M Burnett.** 2018. "Scaling Social Movements Through Social Media: The Case of Black Lives Matter." *Social Media + Society*, 4(4): 2056305118807911.
- Pariser, Eli.** 2011. *The Filter Bubble: What the Internet Is Hiding*. Princeton university press.
- Qin, Bei, David Strömberg, and Yanhui Wu.** 2017. "Why Does China Allow Freer Social Media? Protests Versus Surveillance and Propaganda." *Journal of Economic Perspectives*, 31(1): 117–40.
- Rane, Halim, and Sumra Salem.** 2012. "Social Media, Social Movements and the Diffusion of Ideas in the Arab Uprisings." *Journal of International Communication*, 18(1): 97–111.

- Reny, Tyler T, and Benjamin J Newman.** 2021. "The Opinion-Mobilizing Effect of Social Protest Against Police Violence: Evidence From the 2020 George Floyd Protests." *American Political Science Review*, 115(4): 1499–1507.
- Rieder, Bernhard.** 2013. "Studying Facebook via Data Extraction: The Netvizz Application." 346–355, ACM.
- Ross Arguedas, A, Craig Robertson, Richard Fletcher, and Rasmus Nielsen.** 2022. "Echo Chambers, Filter Bubbles, and Polarisation: A Literature Review."
- Sebbah, Brigitte, Natacha Souillard, Laurent Thiong-Kay, and Nikos Smyrniotis.** 2018. "Les Gilets Jaunes, Des Cadrages Médiatiques Aux Paroles Citoyennes." *Rapport De Recherche Préliminaire*.
- Siegel, David A.** 2009. "Social Networks and Collective Action." *American Journal of Political Science*, 53(1): 122–138.
- Sunstein, Cass R.** 2017. # *Republic: Divided Democracy in the Age of Social Media*. Princeton university press.
- Wallace, Sophia J, Chris Zepeda-Millán, and Michael Jones-Correa.** 2014. "Spatial and Temporal Proximity: Examining the Effects of Protests on Political Attitudes." *American Journal of Political Science*, 58(2): 433–448.
- Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov.** 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics*, 12: 415–438.



# Appendix

## A. Data Sources

### A.1 Street Protests

A website ([www.blocage17novembre.fr](http://www.blocage17novembre.fr)) was created to coordinate the mobilization. It provided a map of the organized blockades, updated in real-time. As of November 16, the map documented 788 geolocated blockades. We use this map to document the offline mobilization of the Yellow Vests.

### A.2 Change.org Petition

Change.org generously gave us access to an anonymized list of the signatories of the petition which launched the Yellow Vests movement. Each observation is associated with the date of signature and the ZIP code of the signatory. We restrict the data to signatures in mainland France and with a valid ZIP code. Using the ZIP code, we are able to associate each signature with a municipality, and therefore compute the signature rate in each municipality by dividing the number of signatures in each municipality by its population. In some instances, a ZIP code is associated with several municipalities. In these cases, we allocate signatures associated to this ZIP code across relevant municipalities proportionally to population. In Figure A.2, we map the distribution of signature rates over France.

### A.3 Facebook Activity

The main websites coordinating demonstrations listed local Facebook groups.<sup>1</sup> To document online mobilization, we looked for public Facebook groups and pages related to the movement. Due to the limitations of the Facebook API, we had to look for groups and pages manually, between December 12 and December 15, 2018 for groups and between March 21 and March 23, 2019 for pages. We used Netvizz to retrieve content between April 2 and April 10, 2019. We use a keyword search approach to find Facebook groups and pages, performing requests on Facebook's search engine and manually retrieving results. These searches were performed using temporary sessions in order to minimize bias induced by Facebook's algorithm.

For groups, our aim was to retrieve as many groups linked to the Yellow Vests as possible. To this end, we started by searching for the keywords "gilet jaune" and

---

<sup>1</sup>First [blocage17novembre.fr](http://blocage17novembre.fr), then [gilets-jaunes.com](http://gilets-jaunes.com) and [giletsjaunes-coordination.fr](http://giletsjaunes-coordination.fr).

“hausse carburant”, on their own and associated with the the codes and names of the départements and of the former and current regions, as well as the names of all municipalities with more than 10,000 inhabitants.<sup>2</sup> Then, we performed further searches with the keywords “hausse taxes”, “blocage”, “colere” and “17 novembre”, associated with the names of the French départements, the names of the former and current regions, and the same list of municipalities as before. Finally, we performed searches for the following keywords: “gillet jaune”, “gilets jaune”, “manif 17 novembre”, “manif 24 novembre”, “manif 1 decembre”, “manif 8 decembre”, “macron 17 novembre”, “macron 24 novembre”, “macron 1 decembre”, “macron 8 decembre”, “blocus 17 novembre”, “blocus 24 novembre”, “blocus 1 decembre”, “blocus 8 decembre”, “blocage 17 novembre”, “blocage 24 novembre”, “blocage 1 decembre”, “blocage 8 decembre”.<sup>3</sup>

For pages, as our aim was not to retrieve the universe of active Yellow Vests communities but simply a sample of messages large enough to perform text analysis, we relied on a smaller number of searches, searching for the keywords “gilet jaune” and “blocage hausse carburant” on their own or associated with the codes and names of the départements as well as a list of the largest cities.<sup>4</sup>

## Yellow Vests Groups

For each group, we recorded the name of the group, creation date, number of members, and number of publications. We eventually identified 3,033 groups in total, with over four million members. Over two-thirds of the groups were associated with a geographical area and more than 40% of the total number of members belonged to these localized groups. Moreover, only 20% of the posts emanated from national groups, which suggests that localized groups were the most active type. Table A.1 presents descriptive statistics on the dataset. Figure A.3 displays the spatial distribution of these groups, before (Panel A) and after (Panel B) 11/17.

---

<sup>2</sup>Restricting the keywords used to these large municipalities is necessary as the number of municipalities in France is very high. It might introduce a bias towards groups associated to denser areas. Fortunately, this bias is reduced by a characteristic of Facebook’s algorithm: when searching for groups and pages associated with a municipality on the platform, Facebook also retrieves results associated to nearby municipalities.

<sup>3</sup>We reviewed all the search results manually to only keep the groups clearly associated with the mouvement.

<sup>4</sup>The complete list of further keywords used is the following: paris; marseille; lyon; toulouse; nice; nantes; strasbourg; montpellier; bordeaux; lille; rennes; reims; le havre; saint etienne; toulon; grenoble; dijon; angles; villeurbanne; le mans; nimes; aix en provence; brest; clermont ferrand; limoges; tours.

Table A.1: Characteristics of Facebook groups

Targeted Audience	Groups	Members	Publications
<b>General</b>	502 (63%)	2,372,217	255,131
<b>Region</b>	164 (81%)	244,930	135,857
<b>County</b>	717 (81%)	507,729	320,263
<b>Municipality</b>	1,638 (65%)	983,057	742,036
<b>Total</b>	3,033 (70%)	4,109,325	1,453,878

*Notes:* In the first column of this table, we show the number of Facebook groups for each geographic focus. We infer the group’s targeted audience from its name. In parentheses, we indicate the share of the number of groups created after 11/17. Other columns show the total number of members and the total number of publications (this number is right-censored by Facebook at 10,000 publications per group). The last line (“Total”) includes 12 “foreign” groups, 11 of which were created after 11/17, including 1,392 members and associated with 591 publications.

Table A.2: Comparison Between the Two Data Collections on Facebook Pages

Data Collection	Pages	Posts	Comments	Sentences	Users
First	617	120,242	1,936,921	2,860,427	NA
Second	411	56,062	352,733	706,182	120,463

*Notes:* This table presents simple count metrics to compare the datasets resulting from our two data collections on Facebook pages.

### Yellow Vests Pages

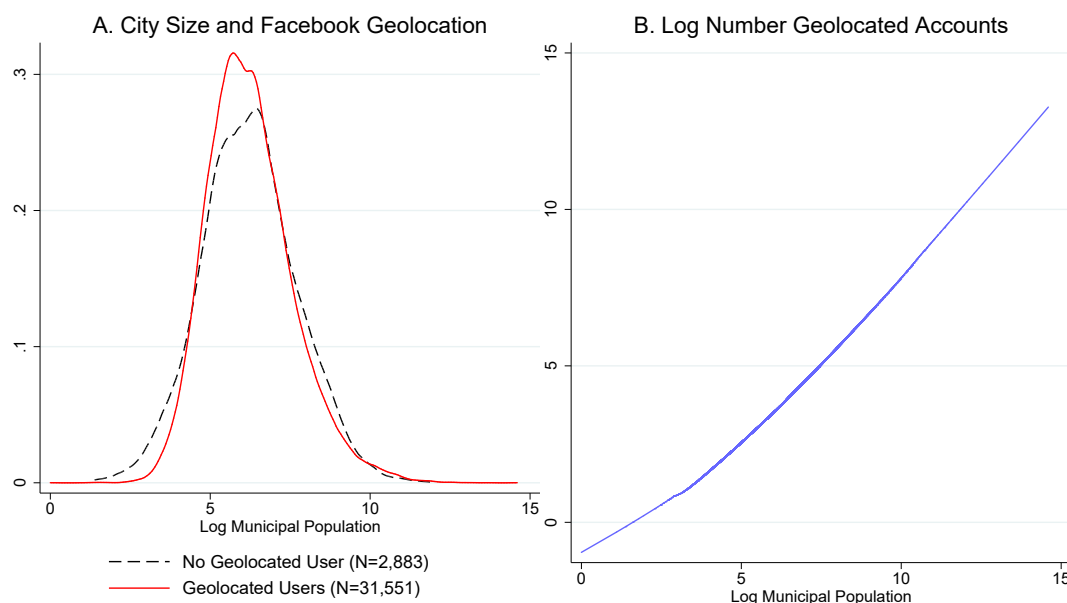
We identified 617 Facebook pages and used Netvizz to retrieve their content (Rieder 2013): posts, comments and interactions (such as likes and shares).<sup>5</sup> This corpus features over 121,000 posts, 2.1 million comments and 21 million interactions. Since Netvizz did not provide user ids associated to scraped content, we scraped Facebook a second time in January 2022 and collect additional basic user information. To protect users’ privacy, all user ids were de-identified. Approximately 30% of pages had been deleted by January 2022. On the remaining pages, we were able to retrieve 46% of the original posts and 18% of the original comments for this second data retrieval (see Table A.2). Both datasets appear similar both in terms of their distribution of political preferences and in terms of the topics discussed. They also display qualitatively similar trends in our three main outcome variables in Section 4.3, though the second dataset generally displays larger increases in radical attitudes.

<sup>5</sup>Netvizz is no longer available since the 21st August, 2019.

## Facebook Penetration Rates

To measure Facebook penetration within France, we leverage one of the largest data leaks in the history of Facebook. In 2019, a massive dataset of Facebook users was made publicly available.<sup>6</sup> For half a billion Facebook accounts (including 30 million French accounts), the dataset contains its creation date, the name of the user, their marital status, self-declared location, and phone number. The hackers compiled a close-to-comprehensive list of Facebook public profiles by searching through Facebook ID numbers in ascending order. Survey data from 2018 shows that about 60% of all French adults have a Facebook account, which is consistent with the number of French accounts leaked as the adult population of the country is slightly above 50 million (see [here](#) for details). It is to this date the largest publicly available dataset of Facebook users with location information.

Figure A.1: Geolocation of Facebook Accounts



*Notes:* Panel A shows the distribution of municipal population for municipalities with at least one geolocated account and municipalities with no geolocated account. Panel B shows a Lowess regression of the (log) number of geolocated municipal accounts and the log of population.

We combine string pattern matching techniques with human supervision to link self-declared locations of users to French administrative data. By doing so, we are able to match 10.8 million accounts to a municipality. Although it is not easy to verify the representativeness of this geolocated subset, we are able to geolocate a positive number of users in over 90% of the municipalities. As shown in Panel A of Figure A.1,

<sup>6</sup>Source: <https://www.bbc.com/news/technology-56772772>



municipalities with no geolocated account are fairly similar in size to the other municipalities. As shown in Panel B of Figure A.1, the log-log relationship between total population and number of geolocated users is almost linear, with a slope of 1. Finally, we define Facebook penetration as the number of accounts per inhabitant. Panel B in Figure A.4 shows the spatial distribution of this variable, which displays much more heterogeneity within regions than between regions.

#### A.4 Tweets of Politicians

We build a dataset of tweets by politicians who belonged to the lower chamber of the French Parliament (the *Assemblée Nationale*) between 2017 and 2022. We consider the five largest French political parties: Rassemblement National (RN), Les Républicains (LR), La République en Marche (LREM), le Parti Socialiste (PS) and La France Insoumise (LFI). Politicians use Twitter to directly speak to their constituents. Thus, tweets are closer to daily social media messages than parliamentary speeches. They provide a natural, labeled dataset to train a machine learning classifier of party affiliation based on written text. We then use our classifier to infer online protesters' political partisanship based on their Facebook messages. The complete list of politicians at the *Assemblée Nationale* is available [here](#). The dataset of French politicians on Twitter is available [here](#). We retrieve the 3200 last tweets of each politician via the Twitter API.

#### A.5 Administrative Data

We construct a wide set of local controls. The set of municipal controls included our regressions may be grouped as follows:

- **Geography** includes the population of the municipality, its density, the distance to the closest city with over 20,000 inhabitants and 100,000 inhabitants, whether the municipality was classified as urban in 2015 and whether it switched from rural to urban between 1999 and 2015.

*Source: Census (RP, complementary exploitation), 2016, INSEE.*

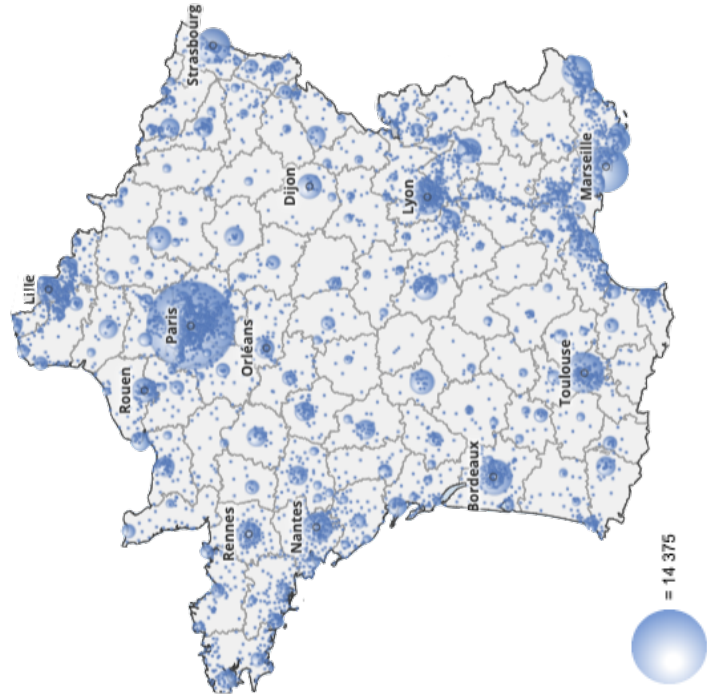
- **Transport** includes the shares of the employed population commuting by car and public transportation, as well as the median commuting distance.

*Source: Census 2016, INSEE. Déclarations Annuelles de Données Sociales (DADS), 2015, INSEE.*

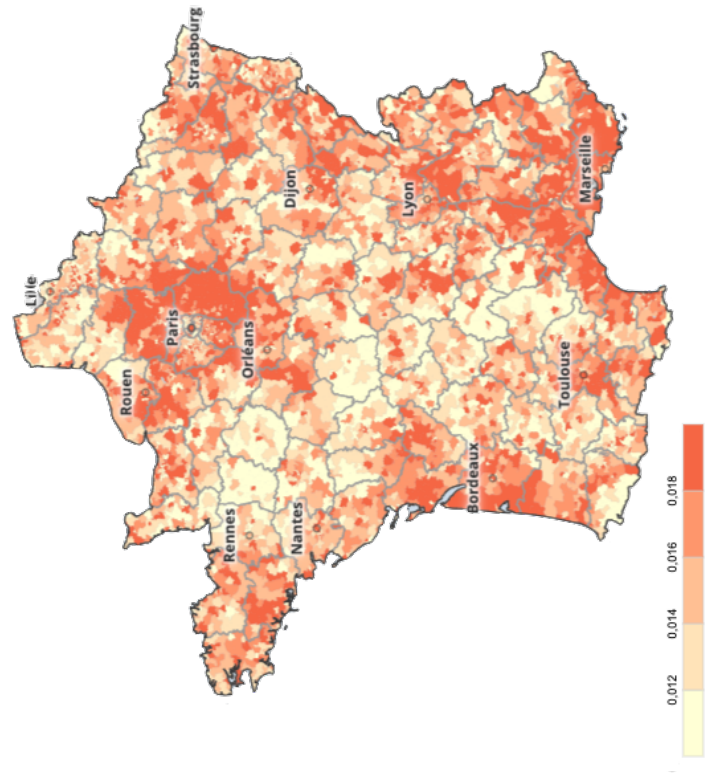
- **Economy** include the local unemployment rate, the fraction of employees with a non-permanent contract, mean income, and population immigrant share.  
*Source: Census 2016, INSEE. DADS, 2015, INSEE.*
- **Occupation** includes the share of the different *catégories socio-professionnelles* defined by INSEE: executive, independent, middle-management, employee, manual worker and agriculture.  
*Source: Census 2016, INSEE.*
- **Age** includes the shares of the population in the following groups: 18-24 y.o.; 25-39 y.o.; 40-64 y.o.; over 65 y.o.  
*Source: Census 2016, INSEE.*
- **Education** includes the shares of the population without the high-school diploma, and with a university degree.  
*Source: Census 2016, INSEE.*
- **Vote** includes the vote share for the five major candidates in the 2017 presidential election (Macron, Le Pen, Fillon, Mélenchon, Hamon), as well as the share of abstention.  
*Source: Ministry of the Interior.*
- **Signature** is the local signature rate of the Change.org petition before 11/17.  
*Source: Change.org.*
- **LZ** is a set of 1,606 dummies for Life Zones.  
*Source: INSEE.*

Figure A.2: Signature Rate of the Change.org Petition by Municipality

A. Absolute value



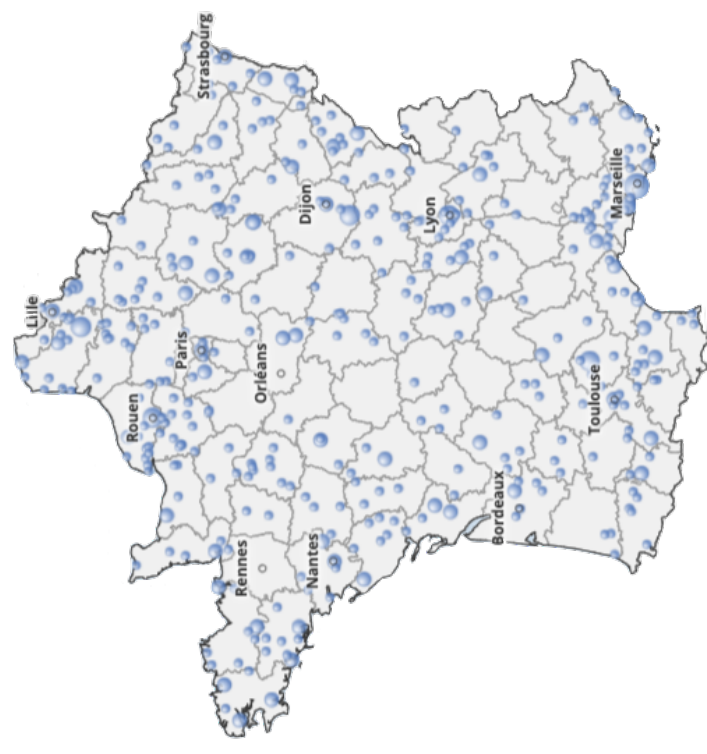
B. Per inhabitant



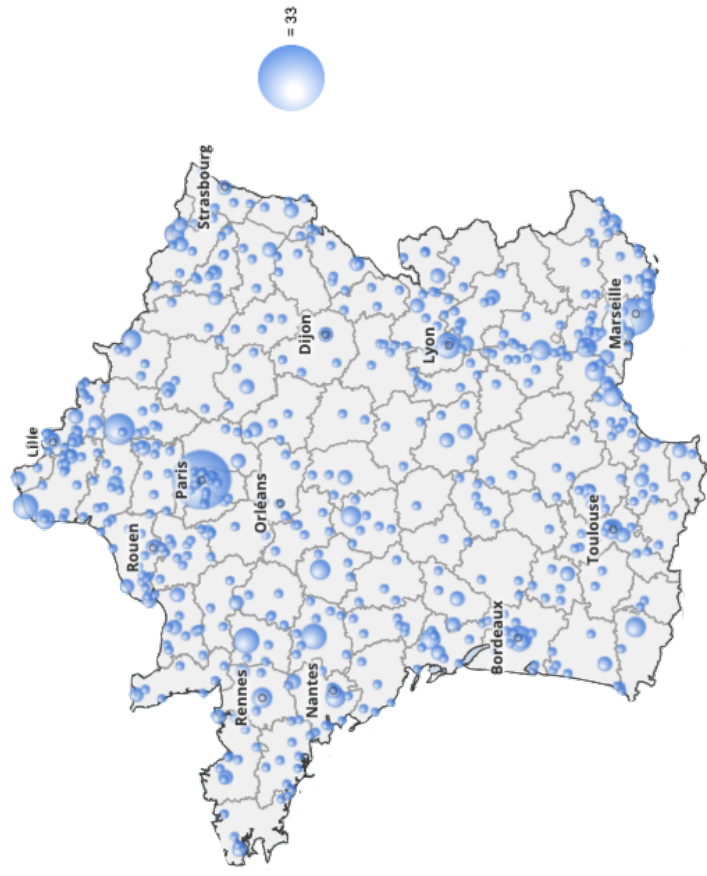
Notes: Figure A displays the number of signature per municipality. Figure B displays the signature rate (signature per inhabitants) by municipalities.

Figure A.3: Number of Local Groups per Municipality.

A. Before 17/11



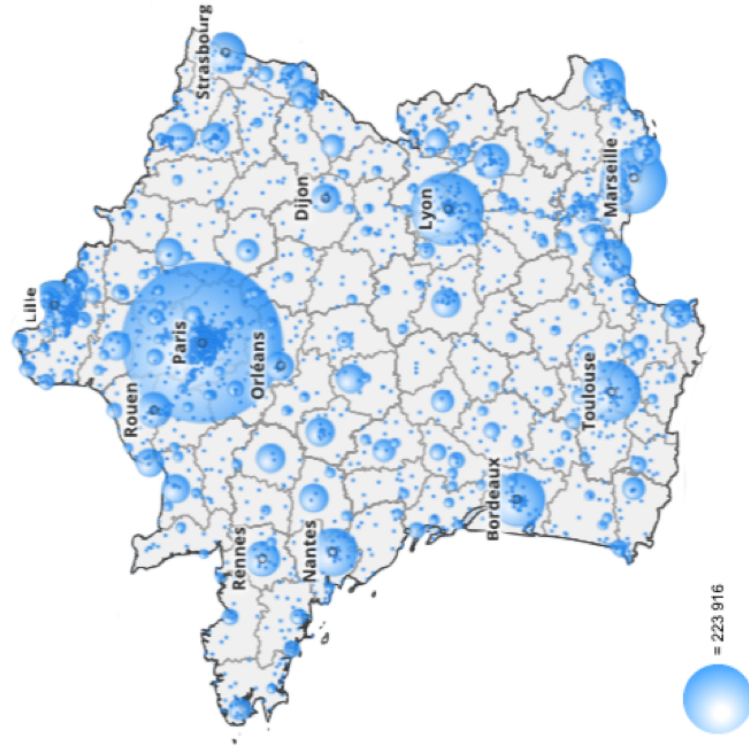
B. After 17/11



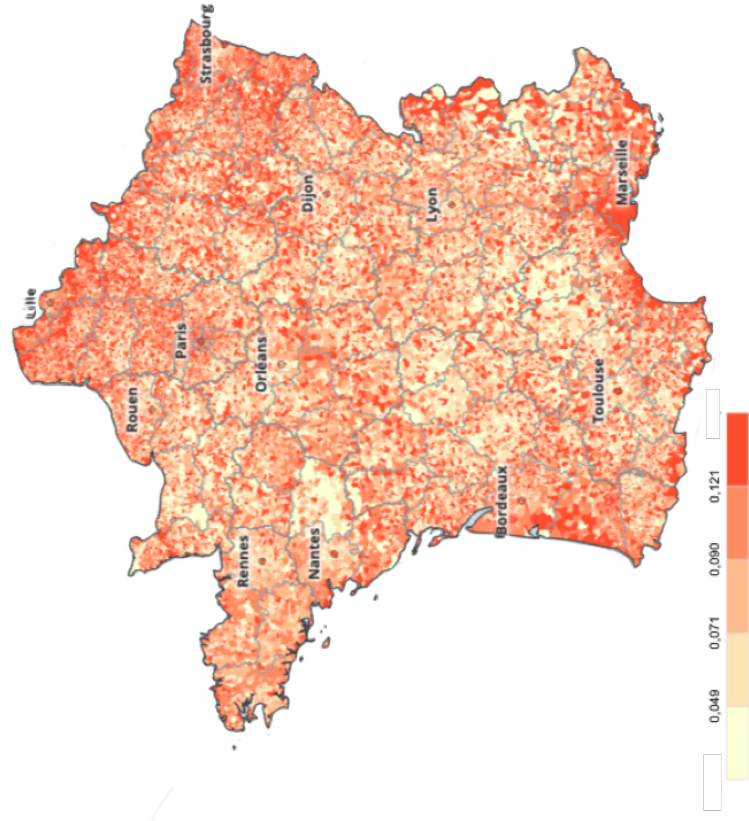
Notes: The two figures display the number of Yellow Vests local group per municipalities. Figure A corresponds to group creations before the First Act of blockades, while Figure B corresponds to group creation after the First Act.

Figure A.4: Spatial Distribution of Facebook Users

A. Absolute value



B. Per inhabitants



Notes: These two maps display the spatial distribution of Facebook users (who declare either to live in or to come from the location) in absolute values at the municipality level. The map on the left-hand side displays the values in proportion of circle radius, while the map on the right-hand side displays the value in terms of color intensity.

## B. Supplement for “The Online-Offline Feedback Loop”

### B.1 Control variables and early mobilization

Table B.1: Variance Decomposition: Yellow Vests Movement (pre-17/11)

	Signatures		Nb. Groups		Blockade	
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Explained variance <math>\hat{y}</math></b>	<b>8.43</b>	<b>37.05</b>	<b>34.61</b>	<b>38.17</b>	<b>17.94</b>	<b>23.48</b>
<b>Fixed effects (LZ)</b>		<b>22.55</b>		<b>1.41</b>		<b>1.00</b>
Density		0.10	0.82	1.35	0.36	0.27
Population	2.90	5.84	21.53	20.75	5.51	8.01
Population squared	0.00	0.00	3.94	7.18	0.63	0.57
Pop. spline 50/75th percentile	0.12	0.26	0.01	0.00	0.00	0.01
Pop. spline 75/100th percentile	2.92	5.90	7.28	5.27	9.18	11.34
<b>Population measures</b>	<b>6.01</b>	<b>12.10</b>	<b>33.59</b>	<b>34.55</b>	<b>15.68</b>	<b>20.19</b>
Dist. to closest mid size city	0.05	0.55	0.02	0.04	0.01	0.04
Dist. to closest large city	0.08	0.17	0.01	0.06	0.01	0.01
Urban municipality	0.10	0.29	0.24	0.53	0.87	0.88
Urbanized since 1999	0.00	0.03	0.01	0.02	0.02	0.02
<b>Geography</b>	<b>0.23</b>	<b>1.04</b>	<b>0.28</b>	<b>0.65</b>	<b>0.91</b>	<b>0.94</b>
Share commuting by car	0.02	0.06	0.04	0.08	0.06	0.07
Share commuting by public transp	0.12	0.11	0.19	0.27	0.14	0.10
Median commuting distance	0.02	0.00	0.00	0.00	0.02	0.02
<b>Commuting</b>	<b>0.17</b>	<b>0.17</b>	<b>0.23</b>	<b>0.35</b>	<b>0.22</b>	<b>0.19</b>
Average wage income	0.24	0.02	0.00	0.00	0.01	0.00
Share in CDI	0.02	0.04	0.01	0.02	0.07	0.07
Unemployment rate	0.09	0.05	0.11	0.22	0.23	0.20
Share retail workers	0.00	0.00	0.00	0.00	0.00	0.01
Share executives	0.04	0.04	0.00	0.01	0.02	0.02
Share intermediate workers	0.10	0.12	0.01	0.03	0.02	0.01
Share clerical workers	0.06	0.07	0.01	0.01	0.02	0.01
Share blue collar	0.03	0.03	0.00	0.00	0.00	0.00
<b>Labor market</b>	<b>0.57</b>	<b>0.36</b>	<b>0.14</b>	<b>0.29</b>	<b>0.37</b>	<b>0.33</b>
Share 18 to 24 y.o.	0.01	0.01	0.13	0.27	0.20	0.23
Share 25 to 39 y.o.	0.04	0.02	0.01	0.02	0.03	0.03
Share 40 to 64 y.o.	0.01	0.01	0.03	0.03	0.04	0.03
Share over 65 y.o.	0.37	0.30	0.00	0.00	0.00	0.01
<b>Age groups</b>	<b>0.43</b>	<b>0.33</b>	<b>0.16</b>	<b>0.33</b>	<b>0.27</b>	<b>0.29</b>
Share with HS degree	0.03	0.00	0.00	0.00	0.01	0.01
Share with college degree	0.10	0.02	0.01	0.02	0.01	0.00
<b>Education</b>	<b>0.13</b>	<b>0.03</b>	<b>0.01</b>	<b>0.02</b>	<b>0.02</b>	<b>0.01</b>
Fillon vote	0.14	0.19	0.00	0.00	0.01	0.00
Hamon vote	0.00	0.00	0.01	0.01	0.01	0.01
Le Pen vote	0.55	0.13	0.00	0.03	0.02	0.02
Macron vote	0.02	0.01	0.00	0.00	0.03	0.03
Far left vote	0.04	0.01	0.01	0.01	0.01	0.01
Abstention	0.00	0.00	0.03	0.12	0.26	0.27
<b>2017 election votes</b>	<b>0.75</b>	<b>0.34</b>	<b>0.05</b>	<b>0.17</b>	<b>0.33</b>	<b>0.34</b>
Share roads with reduced speed	0.01	0.00	0.02	0.05	0.06	0.06
Share of diesel vehicles	0.12	0.09	0.04	0.19	0.05	0.10
<b>Motorists</b>	<b>0.13</b>	<b>0.09</b>	<b>0.06</b>	<b>0.25</b>	<b>0.11</b>	<b>0.16</b>
<b>Facebook penetration</b>	<b>0.01</b>	<b>0.03</b>	<b>0.09</b>	<b>0.16</b>	<b>0.04</b>	<b>0.03</b>

Notes: Following [Shorrocks \(1982\)](#), this table displays the factor contribution of each explanatory variable on the signature rate per inhabitant before 11/17 (Columns (1) and (2)), the number of Facebook groups before 11/17 (Columns (3) and (4)), and a dummy variable for the existence of a blockade on 11/17 (Columns (5) and (6)).

## B.2 Details on the roundabout instrument

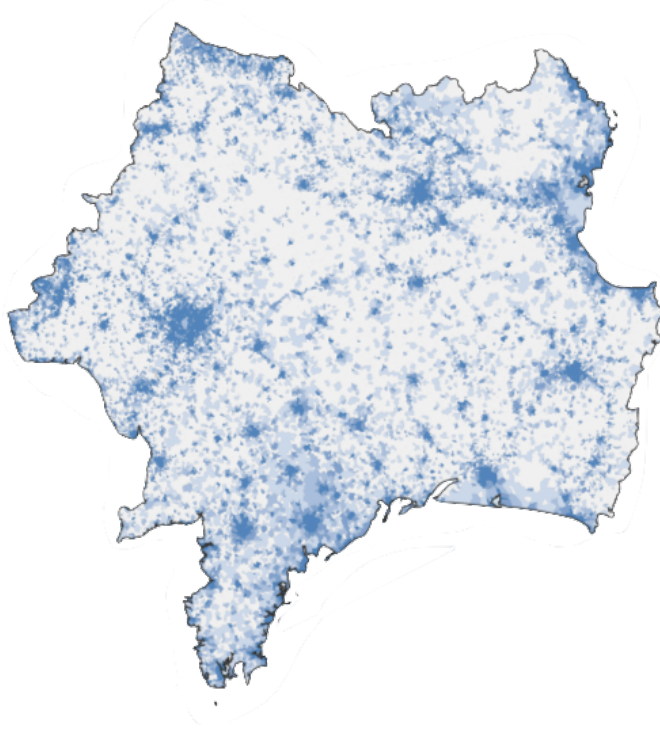
Table B.2: Variance Decomposition: Roundabouts

	Roundabout density	
	(1)	(2)
<b>Explained variance <math>\hat{y}</math></b>	<b>45.33</b>	<b>51.75</b>
<b>Fixed effects (LZ)</b>		<b>3.18</b>
Density	1.69	2.30
Population	18.81	18.41
Population squared	0.04	0.06
Pop. spline 50/75th percentile	0.06	0.04
Pop. spline 75/100th percentile	19.85	19.80
<b>Population measures</b>	<b>40.44</b>	<b>40.61</b>
Dist. to closest mid size city	0.32	0.87
Dist. to closest large city	0.11	0.12
Urban municipality	2.09	3.35
Urbanized since 1999	0.03	0.04
<b>Geography</b>	<b>2.55</b>	<b>4.38</b>
Share commuting by car	0.02	0.06
Share commuting by public transp	0.33	0.54
Median commuting distance	0.04	0.05
<b>Commuting</b>	<b>0.39</b>	<b>0.65</b>
Average wage income	0.04	0.14
Share in CDI	0.25	0.38
Unemployment rate	0.07	0.12
Share retail workers	0.03	0.04
Share executives	0.03	0.04
Share intermediate workers	0.02	0.03
Share clerical workers	0.04	0.05
Share blue collar	0.01	0.01
<b>Labor market</b>	<b>0.50</b>	<b>0.81</b>
Share 18 to 24 y.o.	0.32	0.49
Share 25 to 39 y.o.	0.04	0.08
Share 40 to 64 y.o.	0.09	0.10
Share over 65 y.o.	0.01	0.04
<b>Age groups</b>	<b>0.46</b>	<b>0.72</b>
Share with HS degree	0.00	0.00
Share with college degree	0.05	0.03
<b>Education</b>	<b>0.05</b>	<b>0.03</b>
Fillon vote	0.00	0.00
Hamon vote	0.03	0.03
Le Pen vote	0.07	0.10
Macron vote	0.21	0.18
Far left vote	0.02	0.02
Abstention	0.10	0.22
<b>Motorists</b>	<b>0.11</b>	<b>0.24</b>
Share roads with reduced speed	0.04	0.10
Share of diesel vehicles	0.46	0.72
<b>Motorists</b>	<b>0.50</b>	<b>0.82</b>
<b>Facebook penetration</b>	<b>0.00</b>	<b>0.00</b>

Notes: Following [Shorrocks \(1982\)](#), this table displays the factor contribution of each explanatory variable on the the density of roundabouts in the municipality (Columns (1) and (2)), and on the density of roundabouts in the other municipalities of the LZ (Columns (3) and (4)).

Figure B.1: Roundabout density

A. Roundabouts by squared kilometer



B. Residuals



Notes: Panel A shows the density of roundabouts in mainland France, with darker colors corresponding to higher density. Panel B shows the residual density of roundabouts after controlling for the set of controls described in Section 3. Color intensity corresponds to quantile thresholds.



### B.3 Robustness of the 2SLS results

Table B.3: Impact of Blockades on Post-17/11 Online Mobilization: Alternative Specifications

	(1) Signatures	(2) Groups	(3) Members	(4) Posts
<b>Panel A: Without controls</b>				
Blockade	2.740*** (0.283)	5.874*** (0.490)	0.155*** (0.0513)	0.104*** (0.0385)
Kleibergen-Paap F-stat	30.6	30.6	30.6	30.6
p-value Hansen	0.001	0.022	0.036	0.029
<b>Panel B: Only municipal instrument</b>				
Blockade	1.261*** (0.345)	3.289*** (0.868)	0.287* (0.151)	0.198* (0.110)
Kleibergen-Paap F-stat	14.7	14.7	14.7	14.7
<b>Panel C: Only LZ instrument</b>				
Blockade	1.086*** (0.304)	2.673*** (0.842)	0.163** (0.0740)	0.0971* (0.0566)
Kleibergen-Paap F-stat	37.0	37.0	37.0	37.0
<b>Panel D: Commuting zone instead of LZ</b>				
Blockade	0.632** (0.255)	3.430*** (0.896)	0.289*** (0.0967)	0.186** (0.0745)
Kleibergen-Paap F-stat	12.1	12.1	12.1	12.1
p-value Hansen	0.099	0.868	0.131	0.140
<b>Panel E: Excluding Paris region</b>				
Blockade	0.840*** (0.309)	2.992*** (0.901)	0.395*** (0.144)	0.287** (0.121)
Kleibergen-Paap F-stat	18.9	18.9	18.9	18.9
p-value Hansen	0.656	0.635	0.067	0.065

*Notes:* This table shows estimates corresponding to variations of the regressions of Table 3.2. Panel A shows results for the 2SLS estimation of Table 3.2 when we do not include any municipal control nor LZ fixed effects. In Panel B (resp. C), we show 2SLS results using the roundabout density of the municipality as an instrument (resp., the density of roundabouts in other municipalities of the LZ) only. In Panel D, we control for commuting-zone fixed effects instead of LZ fixed effects and instead of considering the density of roundabouts in other municipalities of the LZ as a second instrument, we use density of roundabouts in other municipalities of the commuting zone. In all panels but Panel E, the number of observations is 34,434. In all regressions, we cluster standard errors at the LZ level (except in Panel E, where we cluster standard errors at the commuting zone level). \*:  $p < 0.01$ , \*\*:  $p < 0.05$ , \*\*\*:  $p < 0.1$ .

## C. Supplement for “The Rise of Online Radicalism”

### C.1 Text Pre-processing

We process all text corpora in the same way. We remove emojis, links, accents, punctuation, social media notifications (e.g., “Yellow Vests changed their profile picture”) and stopwords from the corpus. We also lower-case the text and lemmatize words. We keep hashtags and user mentions, but drop all tokens which occur less than ten times in the Facebook corpus.<sup>7</sup> This leaves us with approximately 40,000 unique tokens in the corpus. Most documents in our corpora are short text snippets (e.g., a phrase or a sentence). Some are longer and span over a multiple sentences (e.g., Facebook posts). To keep all documents comparable, we work with unigrams at the sentence level.

### C.2 Topic Model

The standard approach for topic modeling in the text as data literature is to rely on Latent Dirichlet Allocation (LDA) models. LDA models documents as a distribution over multiple topics. Though this is often a reasonable assumption, it is implausible in the case of short text snippets (such as sentences) which often refer to only one topic (Yan et al. 2013). For this reason, standard topic models are known to perform poorly on such short texts. As an alternative, we build a custom topic model in the spirit of Demszky et al. (2019). First, we produce word embeddings for the corpus and represent each sentence as a vector in the embedding space. We train a Word2Vec model using Gensim’s implementation, with moving windows of eight tokens and ten iterations of training. We build sentence embeddings as the weighted average of the constituent word vectors, where the weights are smoothed inverse term frequencies (to assign higher weights to rare/distinctive words) (Arora, Liang and Ma 2016). The resulting embedding space allows for a low-dimensional representation of text, in which phrases which appear in similar contexts are located close to one another. Second, we group sentence vectors together into a small set of clusters. The goal is to have different clusters for different topics in the text. We rely on the K-Means algorithm. We train the algorithm on 100,000 randomly drawn sentences and predict clusters for the rest of the corpus. We use the ten closest words to the cluster centroids to manually label topics.<sup>8</sup> We choose to work with 15 topics for our main results. However, since the number of

---

<sup>7</sup>The frequency threshold does not influence results, but allows us to remove many uncommon spelling mistakes and other idiosyncrasies related to social media data.

<sup>8</sup>We also considered alternative labeling options, such as term frequency - inverse cluster frequency, which yield similar results.

topics is a hyperparameter in our topic model, we also present resulting topics when specifying 5, 10 and 20 clusters (see Table C.1).

### C.3 Sentiment Analysis

To measure emotional content in Facebook messages, we use a dictionary-based approach that assigns to a sentence a sentiment score ranging from -1 (very negative) to 1 (very positive). For each sentence, the sentiment score is obtained as the average of the sentiment scores of its constituent words. We rely on the [VADER \(Valence Aware Dictionary for Sentiment Reasoning\)](#) library for our main results.

Our measure of sentiment could vary depending on the dictionary used. As a robustness check, we rely on [French TextBlob](#) as an alternative dictionary for word sentiment. We find that the VADER dictionary’s density has larger tails as it tends to classify more sentences to the extremes of the sentiment spectrum. Nonetheless, both measures suggest an increase in average negative sentiment between November 2018 and March 2019. Figure C.1 decomposes the increase in average negative sentiment (as measured by TextBlob) using the method outlined in Section 4.3. Results are qualitatively similar to the main text results.

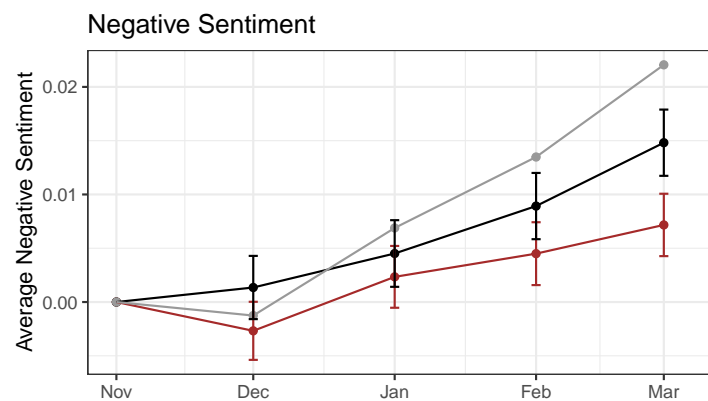


Figure C.1: Margins for Negative Sentiment Using TextBlob

*Notes:* This figure decomposes the increase in average negative sentiment using the method outlined in Section 4.3. We compute sentiment scores based on the TextBlob dictionary. Results are qualitatively similar to the main text results.

### C.4 Political Partisanship Model

Our principal classification method is multinomial logistic regression. Given the large size of the vocabulary, we further penalize the regression with the L1-norm

Table C.1: Results of the Topic Model for Alternative Numbers of Clusters

## Panel A: Results of the Topic Model for 5 clusters

---

**Associated words**

---

04, nimes, arras, nime, 77, narbonne, albi, chambery, 47, orleans  
 pouvoir, etre, consequent, favoriser, necessaire, n, global, politique, specifique, constitue  
 merde, connard, salopard, pourriture, encule, putain, hont, honte, batard, ordure  
 gabin, live, sympa, app, brancher, stp, ramous, cool, stabilisateur, coupure  
 laziah, misfortune, #nousosomesgiletsjaune, dellacherie, exhort, substituons, sansone, pajalo, victory, naeim

---

## Panel B: Results of the Topic Model for 10 clusters

---

**Associated words**

---

etre, n, peuple, meme, politique, faiblesse, nefaste, veritable, gouvernement, destructeur  
 annuel, beneficiaire, compenser, bonus, salaire, taxation, production, exoneration, delocalisation, embauche  
 cr, flic, flics, policier, gazer, projectile, charger, manifestant, matraque, gendarme  
 zappe, zapper, tpm, humoriste, fakenew, interviewe, conversation, cnew, interviewer, bfmtv  
 orlane, magdalena, grilo, correia, gourdon, leal, caudrelier, malaury, macedo, khaye  
 connard, merde, encule, bouffon, conard, pd, salope, enculer, fdp, batard  
 adhesion, charte, valider, definir, modalite, eventuel, prealable, specifique, necessaire, proposer  
 04, nimes, arras, albi, nime, royan, 77, narbonne, chambery, 47  
 courage, courag, bravo, felicitacion, formidable, bisou, bisous, genial, soutien, continuation  
 sansone, dutie, facilitate, soldats, auv, weier, unterstützen, #jiletsjaune, ausbeutung, seem

---

## Panel C: Results of the Topic Model for 20 clusters

---

**Associated words**

---

beneficiaire, compenser, salaire, bonus, annuel, exoneration, plafonner, taxation, embauche, reduction  
 omo, #nousosomesgiletsjaune, laziah, houppette, nooooo, jeoffrey, chab, limitatif, exhort, cageot  
 aller, faire, voir, la, etre, oui, vraiment, merde, savoir, meme  
 englos, royan, sisteron, pontivy, arras, seclin, hendaye, douai, roanne, albi  
 twitter, diffuse, info, publier, fb, diffuser, relater, page, interview, information  
 adhesion, structuration, proposer, proposition, definir, charte, structurer, concertation, revendication, necessaire  
 maud, johanna, gomes, anai, melanie, gregory, rudy, armand, melissa, mathias  
 bisous, courage, felicitacion, courag, bisou, bravo, formidable, soutien, genial, coucou  
 asservissement, domination, peuple, deposseder, destructeur, gouvernance, oppression, politique, veritable, appauvrissement  
 recours, illegal, sanction, infraction, poursuite, condamnation, delit, penal, abusif, commettre  
 41, 52, 58, 47, 38, 61, 69, 37, 46, 82  
 canette, chaussette, bouteille, cendrier, plastique, peintur, toilette, saucisson, scotch, brosse  
 cr, flic, flics, frapper, tabasser, matraquer, policier, gazer, matraque, tabasse  
 mafieux, imposteur, larkin, escroc, acolyte, magouilleur, maffieux, corrompu, dictateur, sbire  
 kassav, akiyo, diritti, sempr, dittaturer, etait, popolo, quando, anch, infami  
 stupide, pathetique, affliger, pitoyable, malsain, stupidite, abject, irrespectueux, insultant, grossier  
 15h, 17h30, 16h30, 10h, 14h00, 11h, gare, 8h30, 18h, 18h30  
 laziah, #nousosomesgiletsjaune, gourdon, misfortune, orlane, grilo, victory, duquesnoy, dellacherie, macedo  
 #jiletsjaune, created, soldats, #assembleenationale, #coletesamarelo, #parisprotest, dutie, unterstützen, #france3, sansone  
 connard, encule, batard, salope, fdp, merde, conard, enculer, pd, salopard

---

*Notes:* This table shows the clusters defined by our the topic model when requesting alternative numbers of topics (5, 10, and 20). For each topic, we report the closest words to the cluster centroid (measured by cosine similarity).

(Lasso) to force some coefficients to zero (Friedman et al. 2001). We consider the five largest French political parties: le Rassemblement National (RN), les Républicains (LR), la République en Marche (LREM), le Parti Socialiste (PS) and la France Insoumise (LFI). We parametrize the probability that a text snippet  $\mathbf{x}$  is from party  $k$  as:

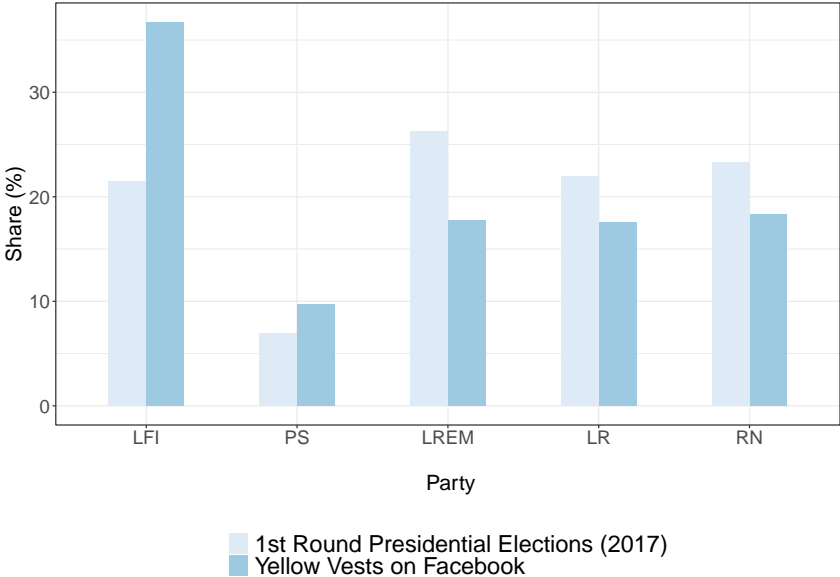
$$P(\text{party} = k | \mathbf{x}) = \frac{\exp(\mathbf{w}_k \cdot \mathbf{x} + b_k)}{\sum_j \exp(\mathbf{w}_j \cdot \mathbf{x} + b_j)}$$

in which  $\mathbf{w}_k$  are specific coefficients to be estimated for party  $k$ . Given the large size of the vocabulary, we further penalize the multinomial logistic regression with the L1-norm (Lasso) to force some coefficients to zero. As some unigrams are not informative of political partisanship, the penalization mitigates over-fitting of the training set by shrinking coefficients.

To validate the model, we shuffle the corpus and split it into 80% training data and 20% test data. We build the classifier in the training set and evaluate its performance in the test set. The model has an accuracy score of 55.5%. A random guess would correctly infer the author’s party 20% of the time. Our model thus assigns the correct party to a text snippet between two and three times more often than a guess at random would. For comparison, Peterson and Spirling (2018) predict party affiliation with an accuracy between 60 and 80% for two parties. In this case, a guess at random would get the label right 50% of the time.

Results are presented in Figure C.2.

Figure C.2: Predicted Political Leaning of the Yellow Vests



Notes: This figure compares the predicted political leaning of the active Yellow Vests users on Facebook (in dark blue) to the scores of each party at the first round of the presidential elections (in light blue). Vote shares at the elections are modified so as to sum up to a hundred (there were other smaller parties that we exclude from the analysis). We assign a political leaning to each Facebook user in our corpus based on the average probability of her sentences being pronounced by a given party according to our classifier.



# Bibliography

- Arora, Sanjeev, Yingyu Liang, and Tengyu Ma.** 2016. "A Simple but Tough-to-beat Baseline for Sentence Embeddings."
- Demszky, Dorottya, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky.** 2019. "Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings." *arXiv preprint arXiv:1904.01596*.
- Friedman, Jerome, Trevor Hastie, Robert Tibshirani, et al.** 2001. *The Elements of Statistical Learning*. Vol. 1, Springer series in statistics New York.
- Peterson, Andrew, and Arthur Spirling.** 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis*, 26(1): 120–128.
- Rieder, Bernhard.** 2013. "Studying Facebook via Data Extraction: The Netvizz Application." 346–355, ACM.
- Shorrocks, Anthony F.** 1982. "Inequality Decomposition by Factor Components." *Econometrica*, 193–211.
- Yan, Xiaohui, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng.** 2013. "A Biterm Topic Model for Short Texts." 1445–1456.





## Chapter 3

# Text Semantics Capture Political and Economic Narratives

**Note:** This chapter is co-authored with Elliott Ash and Philine Widmer.

### Abstract

Social scientists have become increasingly interested in how *narratives* – the stories in fiction, politics, and life – shape beliefs, behavior, and government policies. This paper provides an unsupervised method to quantify latent narrative structures in text documents. Our new software package RELATIO identifies coherent entity groups and maps explicit relations between them in the text. We provide an application to the *United States Congressional Record* to analyze political and economic narratives in recent decades. Our analysis highlights the dynamics, sentiment, polarization, and interconnectedness of narratives in political discourse.

**Keywords:** narratives, memes, natural language processing<sup>1</sup>

---

<sup>1</sup>A special thanks to Andrei Plamada and ETH Scientific IT Services for pivotal contributions to the programmatic implementation RELATIO. We also thank Johannes Binswanger, Jordan Boyd-Graber, Paul Dutronc-Postel, Kfir Eliaz, Malka Guillot, Kai Gehring, Theresa Gessler, Roland Hodler, Philip Resnik, Alessandro Riboni, Michael Roos, Ran Spiegler, and Ekaterina Zhuravskaya for insightful comments. We are grateful for feedback from conference participants at PolMeth U.S., PolMeth Europe, the Swiss Society of Economics and Statistics Conference, the Monash-Warwick-Zurich Text-as-Data Workshop, as well as seminar attendees at CREST, ETH, the Paris School of Economics, and the Universities of Zurich and St.Gallen. This work was supported by the Swiss National Science Foundation. We thank all (early) users of RELATIO for providing valuable feedback, in particular Andrea Sipka and Wenting Song. Our replication materials are available at <https://doi.org/10.7910/DVN/3BRWKK>. The open-source software RELATIO is available at <https://pypi.org/project/relatio/>. Consult the GitHub repository <https://github.com/relatio-nlp/relatio> for additional information and ongoing developments.

## 1. Introduction

If stories are the engines of political identity and collective action, then the best politicians will also be the best storytellers (Patterson and Monroe 1998). Motivated by such ideas, and building on the well-established work in psychology on how narratives shape human perceptions and constructions of social reality (White 1980; Polkinghorne 1988; Bruner 1991; Armstrong 2020), an expanding social-science literature has begun to attend to how narratives drive political, social, and economic outcomes (Branch, McGough and Zhu 2017; Bénabou, Falk and Tirole 2018; Kuhle 2020; Eliaz and Spiegler 2020). Across disciplines, there is a growing recognition that narratives play a fundamental role in both individual and collective decision-making (Akerlof and Shiller 2010; Shiller 2019). Hence, both science and policy would benefit from a better understanding of how narratives form, spread, and influence behavior.

Nonetheless, quantitative analysis of narratives is still largely unexplored (Shiller 2019). A principal impediment to the production of such evidence is the challenge of measuring narratives in written or spoken texts. In particular, such a measure must capture relationships between entities – characters, concepts, organizations, or objects (e.g. Sloman 2005). The existing text-as-data approaches in social science do not account for these relationships (Grimmer and Stewart 2013; Gentzkow, Kelly and Taddy 2019). This paper describes a new method that satisfies this requirement – by identifying *who* does *what* to *whom*, and by mapping the relationships and interactions among entities in a corpus. The accompanying open-source package, RELATIO, allows researchers to measure interpretable narratives from plain text-documents. These narratives can be used as inputs in empirical social science analysis.<sup>1</sup>

The starting point of the narrative mining method is semantic role labeling, a linguistic algorithm that takes in a plain-text sentence and identifies the action, the agent performing that action, and the patient being acted upon. The resulting feature space of agents, actions, and patients is much more informative about narratives than the feature space generated by standard text-as-data methods. Yet, that space is too high-dimensional to be useful for most narrative analyses. Hence, the next part of our narrative mining method is a set of dimensionality reduction procedures. Our entity clustering approach takes the multiple phrase variants referring to the same entity (e.g., “taxes on income” and “income taxation”, or “Former President Reagan” and “Ronald Reagan”) and resolves them into a single entity label.

We demonstrate the usefulness of the method in an application to speeches in the

---

<sup>1</sup>Our replication materials are available at Ash, Gauthier and Widmer (2022). The open-source software RELATIO is available at <https://pypi.org/project/relatio/>.

U.S. Congress (1994-2015). The set of extracted entities is interpretable and includes key actors in the U.S. political economy – Republicans, Democrats, the budget, terrorists, or Medicare, to name just a few. For each entity, our approach highlights what actions (i.e., verbs) connect it to other entities. For instance, we capture that “Republicans” and “Medicare” are related such that “Republicans end Medicare”. “Medicare”, in turn, is also connected to other entities, such as “Medicare provides healthcare”. Overall, the resulting narrative statements are intuitive and close to the original raw text.

The method captures salient narratives around historical events, such as the September 11th attacks and the subsequent U.S. invasion of Iraq. We find that religious invocations such as “God bless America” and “God bless the troops” increased in response to these events. Ranking narratives by relative party usage reveals articulations of partisan values: Democrats are concerned about “Americans losing unemployment benefits”, “budget cuts to Medicare”, and “oil companies making profits”. Republicans want “Americans to keep their money”, decry “government control of healthcare”, and affirm that “Americans rely on oil”.

Besides showing how narratives divide the parties, our approach also demonstrates how narratives connect up with each other to form a broader discourse. We propose a network-based approach to combine several narrative statements in a directed multi-graph, linking up entities with their associated actions. The resulting narrative networks form a visual expression of political worldviews, establishing rich context for qualitative researchers. Further, applying node centrality and graph distance measures to these networks could help illuminate new dimensions of narrative discourse.

The paper concludes with a discussion of how narrative mining fits in with other text-as-data methods. The narrative statements produced by RELATIO are often more informative and interpretable than bag-of-words or n-gram representations of text documents. Finally, we discuss our method’s limitations along with opportunities for improvement and extension.

## 2. Method: Mining Narratives from Plain Text Corpora

According to the Oxford English Dictionary, a narrative is an “account of a series of events, facts, etc., given in order and with the establishing of connections between them”. In a leading framework from social psychology, narratives are similarly defined as sets of relationships between entities that act on each other (Sloman 2005). In human language, such relationships are established through grammatical statements

describing actors, actions, and the acted-upon.

This distinctive relational aspect of narratives is missing from the standard text-as-data-tools in social science. Dictionary methods rely on matching particular words or phrases (Baker, Bloom and Davis 2016; Shiller 2019; Enke 2020). Unsupervised learning methods such as topic models and document embeddings break sentences down into words or phrases and ignore grammatical information (Hansen, McMahon and Prat 2017; Bybee et al. 2020). These previous methods recover information on mentioned entities and mentioned actions, but without connecting them. This section outlines our method for uncovering the recurring relationships established between latent entities and actions in a corpus.

## 2.1 Semantic roles as narrative building blocks

Semantic role labeling (SRL) is a computational-linguistics algorithm that answers basic questions about the meaning of written sentences – in particular, *who* is doing *what* to *whom*.<sup>2</sup> The **agent** (“who”) is the actor in an event – e.g., the grammatical subject of the verb in an active clause. The **verb** (“what”) captures the action in the clause. The **patient** (“whom”) is the entity that is affected by the action – i.e., the object or the target.<sup>3</sup> For example, in the sentence “Millions of Americans lost their unemployment benefits”, “Millions of Americans” is the agent, “lost” the verb, and “their unemployment benefits” the patient.<sup>4</sup>

Not only do semantic roles differentiate actions and entities within a sentence, but they also map the relationships between them. For example, SRL would extract the same directional relation for “Millions of Americans lost their unemployment benefits” as the inverted sentence “Unemployment benefits were lost by millions of Americans.” In some cases, the directions of actions contain pivotal information to understand the narrative.<sup>5</sup> This robustness to word ordering is an important feature of SRL relative to other approaches in natural language processing, such as topic models, which ignore semantic relations in the sentence.

---

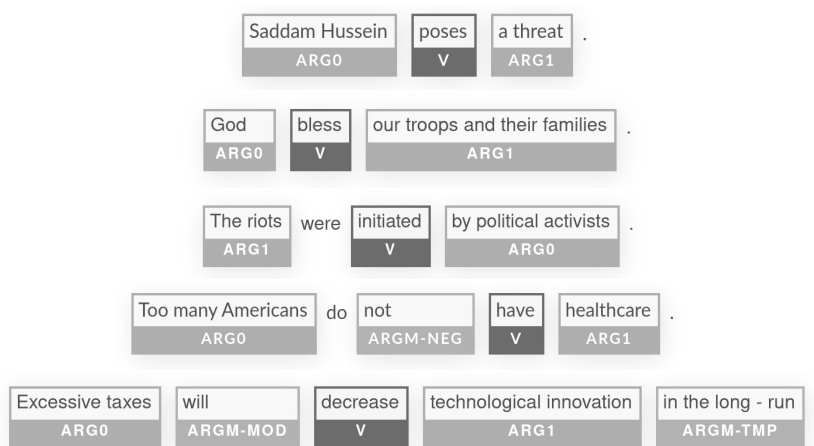
<sup>2</sup>Linguistically, semantic roles are functional components of sentences that are defined based on their relationship with the main verb in a clause. For a detailed background, see Jurafsky and Martin (2020), ch. 20 (and sources cited therein). For information on the particular schema we use, see Bonial et al. (2012).

<sup>3</sup>For simplicity, we use the word “patient” to refer to both linguistic patients and linguistic benefactives – in most sentences, these refer to the direct object and indirect object, respectively.

<sup>4</sup>Any role can be empty (except for the verb). For instance, with “Millions of Americans suffer”, the patient role is empty.

<sup>5</sup>Consider for example: “Terrorists threaten American troops.” and “American troops threaten terrorists”.

Figure 2.1: Examples of Semantic Role Labeling Annotations



**Note:** Examples of semantic role annotations based on allennlp’s programmatic implementation (Gardner et al. 2017). See <https://demo.allennlp.org/semantic-role-labeling> for additional examples. ARG0 refers to the agent, V to the verb, and ARG1 to the patient. The last example shows additional semantic roles, modality (ARGM-MOD), negation (ARGM-NEG), and temporality (ARGM-TMP). Our implementation considers negations (ARGM-NEG). While not further discussed here, it also allows for modal indicators (but not yet for temporality).

We show additional example sentences, with semantic role annotations, in Figure 2.1. These examples come from the state-of-the-art model implemented by AllenNLP (Gardner et al. 2017), used in the empirical application below. The sentences illustrate the flexibility of SRL in capturing variation in sentence content and word order. For example, the agent and patient can both be at the beginning or end of the sentence. Further, while all complete sentences in English have a verb, some sentences lack an agent, patient, or both.<sup>6</sup> A number of additional semantic features, such as modality, negations, and temporal content, are labeled as well.

Our model of language is based on the information contained in semantic roles. Formally, define  $\mathcal{A}_0$ ,  $\mathcal{V}$ , and  $\mathcal{A}_1$  as the sets of phrases respectively annotated as agents, verbs and patients. A sentence can be decomposed into a sequence of semantic roles:

$$\text{AGENT} \xrightarrow{\text{VERB}} \text{PATIENT} \in \mathcal{A}_0 \times \mathcal{V} \times \mathcal{A}_1 \quad (3.1)$$

For example, our previously discussed sentence would be represented as:

$$\text{Millions of Americans} \xrightarrow{\text{lost}} \text{their unemployment benefits}$$

Negation is encoded as a modification of the verb. For example, “did not lose” is replaced with the verb “not-lost”. So, a similar statement with a negation would be

<sup>6</sup>“Americans go” has an agent, but not a patient; “Americans change” has a patient, but not an agent, and “Go” (the imperative) has a verb but neither an explicit agent or patient.

represented as:

Americans  $\xrightarrow{\text{not-lost}}$  their unemployment benefits

## 2.2 From semantic roles to interpretable actions and entities

Let  $\mathcal{S} = \mathcal{A}_0 \times \mathcal{V} \times \mathcal{A}_1$  comprise the space of semantic roles observed in the corpus. These semantic roles capture the relationships between entities that are characteristic of narrative statements. In most real-world corpora, however,  $\mathcal{S}$  is too high-dimensional for further analysis. Thus, the next step is to compress the set of actions and entities to a lower-dimensional yet sufficiently informative representation.

For instance, most people would agree that “Millions of Americans lost their unemployment benefits” and “Many Americans were losing their much-needed unemployment checks” refer to the same underlying narrative. In particular, “much-needed unemployment checks” and “their unemployment benefits” both refer to “unemployment benefits”. Similarly, “Many Americans” and “Millions of Americans” both refer to “Americans”. These examples illustrate that observed agents and patients are drawn from a smaller set of latent entities,  $E$ , such that  $|E| \leq |\mathcal{A}_0 \cup \mathcal{A}_1|$ . Thus, our dimension reduction aims to infer these latent entities and their associated text realizations.

We proceed in two steps for dimension reduction. First, we directly extract coherent entities using named entity recognition (e.g., [Jurafsky and Martin 2020](#), ch. 8). This algorithm automatically extracts references to specific people, organizations, events, and locations. In the second sentence from Figure 2.1, for example, “Saddam Hussein” is identified as a named entity. In practice, we build a vocabulary of named entities as the  $L$  most frequent named entities recognized in the corpus. If a semantic role refers to a named entity from the vocabulary, it is labeled as such. In applications, we have found that it is straight-forward to inspect the list of entities ranked by frequency and set the threshold  $L$  to balance dimensionality and interpretability. Appendix C lists examples of frequent named entities.

The second step of dimension reduction, applied to the remaining agent and patient phrases that do not contain a named entity, consists of semantic clustering. These phrases usually refer to coherently separable entities or groups of entities that we would like to group together – e.g. “unemployment benefits” and “unemployment checks”. Our approach for clustering the phrases associated with such entities begins with a phrase encoder, applied to compress each plain-text entity snippet to a low-dimensional dense vector. As a computationally efficient default, we use a weighted average of the word vectors across each word in the agent or patient segment ([Arora](#),



Liang and Ma 2016). We then apply a clustering algorithm (e.g., K-Means) to the matrix of entity encodings to produce  $K$  clusters. For interpretability, each cluster can be labeled by the most frequent term within the cluster.

Unlike the selection of  $L$  for named entities, the selection of  $K$  for clustered entities is not straightforward. Entities are not clearly demarcated and the preferred level of granularity will depend on the application. For example, should “tax credit” and “tax rebate” be clustered? What about “Republicans” and “conservatives”? Hence, hyperparameter choices for clustering entity embeddings require more care. Automated cluster-quality metrics such as the silhouette score can work as a starting point. But for best results, the phrase clustering output should be produced and inspected for different  $K$  and selected based on the goals of the application.

So far, we have not discussed dimension reduction for verbs. In practice, we find that embedding-based clustering of verbs produces unreliable results, in particular because it frequently assigns antonyms (e.g., “decrease” and “increase”) to the same cluster. In our application, the number of unique verbs is relatively small (60 times smaller than the number of unique agents and patients), so we decide not to dimension-reduce verbs. Hence, we do not make any changes to verbs, except for normalizing the verb tense and adding the “not-” prefix to negated verbs.

## 2.3 Narrative statements and narrative structure

After dimension reduction, we obtain narratives of the form:

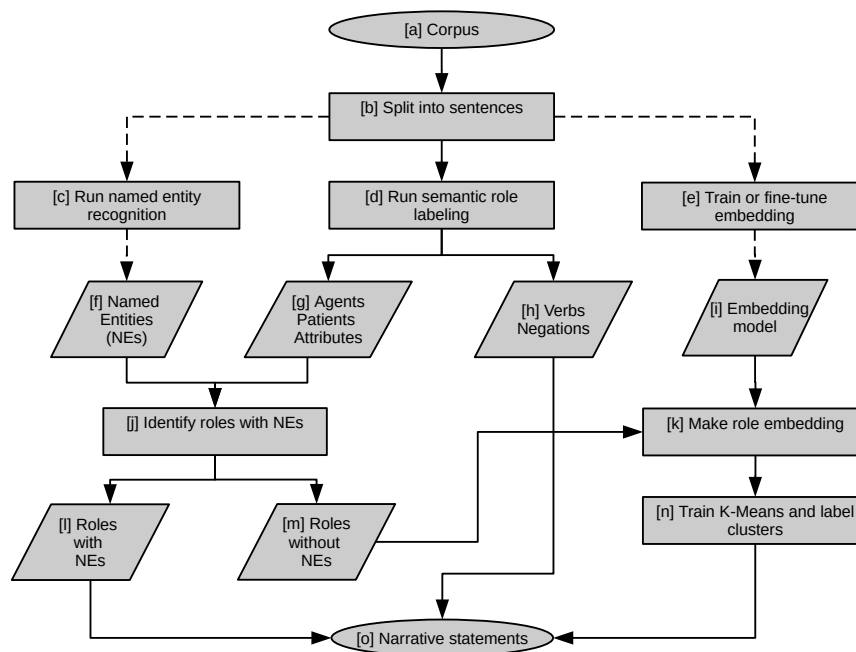
$$\text{AGENT ENTITY} \xrightarrow{\text{(NEGATED) VERB}} \text{PATIENT ENTITY} \in E \times V \times E \quad (3.2)$$

The set  $\mathcal{N} = E \times V \times E$  is the space of all potential narrative statements for a given corpus. Its dimensionality is determined by the number of unique entities and verbs. The set of entities contains named entities and clustered entities.

These narratives can then provide inputs to qualitative or quantitative analysis. In the Congressional Record, for example, we will produce counts by legislator and year for each item in  $\mathcal{N}$ . Such counts can be used for descriptive analysis, as variables in regressions, or as a feature set for machine learning algorithms.

Moving beyond such counts, a key feature of narratives is that they embed entities and relations in a broader, enmeshed structure. Consider for instance: “Taxes fund hospitals and hospitals save lives”. Our method represents this sentence as two sepa-

Figure 2.2: Flowchart for RELATIO



**Note:** Code flowchart for programmatic implementation, open-sourced as the Python RELATIO package ([github.com/relatio-nlp/relatio](https://github.com/relatio-nlp/relatio)). Circles represent the start and the end of the pipeline. Rectangles represent arithmetic operations and data manipulations. Parallelograms represent inputs and outputs.

rate narrative statements:

$$\begin{aligned} \text{taxes} &\xrightarrow{\text{fund}} \text{hospitals} \\ \text{hospitals} &\xrightarrow{\text{save}} \text{lives} \end{aligned}$$

It is easy and intuitive to combine these two simple narrative statements to reveal a more complex narrative chain. This broader narrative has a network structure:

$$\text{taxes} \xrightarrow{\text{fund}} \text{hospitals} \xrightarrow{\text{save}} \text{lives}$$

More generally, a list of simple narrative statements (as captured by Equation 3.2) can be represented as a directed multigraph, in which the edges are actions and the nodes are entities. Formally, let  $n \subset \mathcal{N}$  be a subset of the narrative space. Let  $e \subset E$  be the set of distinct entities in  $n$ . We define a narrative graph as a tuple  $(e, n)$ , in which  $e$  represents the vertices and  $n$  represents the edges of the directed multigraph.

## 2.4 Open-source implementation

Figure 2.2 provides an illustrated summary of our pipeline. The input is a plain-text corpus [a], which is segmented into sentences [b]. Those sentences are the inputs

to named entity recognition [c], semantic role labeling [d], and training (or fine-tuning) a phrase embedding model [e]. The outputs of that first round of operations are a set of named entities [f], annotated semantic roles [g,h], and an embedding model for vectorizing phrases [i]. After tagging named entities [j], the roles containing named entities [l] are finalized for the output. The roles without named entities [m] are vectorized using the phrase embedding model [k], and then piped to the K-means clustering model to produce clustered entities [n]. The final narrative statements [o] are built from the named entities [j], verbs [h], and clustered entities [n].

This system is automated and streamlined as part of the accompanying Python package, RELATIO. A baseline narrative feature set can be produced with only a few lines of code. More advanced users can customize and adapt the inputs, settings, and outputs. For more details, see Appendix A or the package repository, [github.com/relatio-nlp/relatio](https://github.com/relatio-nlp/relatio).

### 3. Application to the U.S. Congressional Record

To demonstrate the workings of our narrative mining method, we apply it to a large corpus of floor speeches in U.S. Congress. The application is designed to illustrate how the method can be used for qualitative and quantitative analysis in social science and digital humanities.

#### 3.1 Implementation

**Data.** Our application uses the *United States Congressional Record* for the years 1994-2015. The *Record* is an exact transcript of speeches given on the floor of the House and the Senate, published in digital format since 1994 by the U.S. Government Publishing Office (GPO). It has been widely used for text-as-data applications in the social sciences (e.g., [Lauderdale and Herzog 2016](#); [Ash, Morelli and Van Weelden 2017](#); [Gentzkow, Shapiro and Taddy 2019](#); [Gennaro and Ash 2021](#)). We link the Congressional speeches to the speakers' metadata (importantly, name and party affiliation).

**Choice of hyperparameters.** We run our pipeline with the default settings on the entire Congressional Record. Appendix A provides some additional material on how the narrative mining system is adapted to the corpus. We limit our analysis to “complete” narratives containing an agent, a verb, and a patient. After some hyperparameter tuning, we select 1000 named entities and 1000 clustered entities (see Appendix B for de-

tails). Clustered entities are first labeled by the most frequent term, with some manual relabeling done after brief inspection.

As the distribution of narrative frequencies is heavily skewed to the left, and since we are interested in recurring narratives, we restrict our attention to narratives pronounced at least 50 times overall in the corpus (i.e., at least twice a year on average across all speakers).<sup>7</sup>

**Resulting narratives.** Table 3.1 lists the ten most frequent narratives, after excluding sentences related to parliamentary procedure. For each narrative, we show two original sentences that the narrative represents. Appendix E shows more examples of sentences of these most frequent narratives in context. The narratives are easily interpretable and semantically close to the original raw text. Thus, the approach satisfies our objective of preserving the important information of *who* does *what* to *whom*.

Appendix D provides summary tabulations on the statements, roles, entities, actions, and narratives extracted from the Congressional Record. The dimensionality reduction is substantial, with the 17.3 million plain-text sentences in the original corpus eventually reduced to 1,638 unique narratives. Besides clustering of entities (see Section 2.2), this reduction is achieved by requiring “complete” narratives (with both an agent and a patient), by the filtering out infrequent narratives (see “Choice of hyperparameters” above), and by dropping narratives containing entities that are procedural or otherwise not related to politics or policy (Appendix L shows a complete list of these entities). The list of narratives, reported in full in Appendix K and readable in a few minutes, provides a concise summary of political discourse in the United States.

## 3.2 Analysis

Now, we use our mined narratives for a descriptive analysis of discourse in U.S. Congress. Starting with Figure 3.1, we first show that narratives capture salient historical events and how those events are framed. Second, we show that narratives contain emotional and partisan resonance. Third, we produce narrative graphs and analyze how political discourse relates entities to one another.

**Narratives reflect key events in U.S. history.** A first task for which our narrative features are useful is to describe changes in discourse over time. To illustrate how

---

<sup>7</sup>This filtering comes at the end and does not influence the construction of our narrative model. Even legislators who speak infrequently show up in our data (see Proksch and Slapin, 2012), unless the narratives pronounced by them are rare overall.

Table 3.1: Most Frequent Non-Procedural Narratives

<b>Narrative</b>	<b>Freq.</b>	<b>Spoken sentence(s)</b>
people lose job	1627	<ul style="list-style-type: none"> <li>– <i>President, people are losing jobs every day.</i></li> <li>– <i>People shouldn't have to lose their jobs to pay for the New York fund.</i></li> </ul>
citizen abide law	1176	<ul style="list-style-type: none"> <li>– <i>I felt kicked around and ignored by the very system the government has in place to protect law-abiding citizens.</i></li> <li>– <i>The vast majority of these private security officers are dedicated, hard-working, law-abiding citizens of this country [...].</i></li> </ul>
american lose job	1168	<ul style="list-style-type: none"> <li>– <i>[...] Americans across this country have lost jobs, unemployment is at a high rate, people are having to make decisions.</i></li> <li>– <i>These funds will go a long way in supporting American workers who have lost their jobs [...].</i></li> </ul>
government run healthcare	1089	<ul style="list-style-type: none"> <li>– <i>I don't understand what they are talking about: "socialized medicine," "Cuban-style, government-run health care."</i></li> <li>– <i>Is that what you are talking about where you all of a sudden shift from people who figure out you can get the government to pay for everything, a government-run health care program?</i></li> </ul>
american have healthcare	1080	<ul style="list-style-type: none"> <li>– <i>[...] [W]e can contain costs and help enable every American to have access to health insurance coverage.</i></li> <li>– <i>[...] Americans already have universal health care because the emergency rooms cannot legally refuse to treat patients.</i></li> </ul>
god bless america	1052	<ul style="list-style-type: none"> <li>– <i>And may God continue to bless America.</i></li> <li>– <i>God bless these heroes, their families and God bless America.</i></li> </ul>
people need help	968	<ul style="list-style-type: none"> <li>– <i>And that is what this welfare reform is all about – to do something about people who are down on their luck and need help.</i></li> <li>– <i>We are supposed to represent the people who need help across this country.</i></li> </ul>
god bless troop	948	<ul style="list-style-type: none"> <li>– <i>In conclusion, God bless our troops, and the President's actions should be based on remembering September the 11th in the global war on terrorism.</i></li> <li>– <i>In conclusion, may God bless our troops, and we will never forget September 11.</i></li> </ul>
worker lose job	889	<ul style="list-style-type: none"> <li>– <i>Democrats want to help more workers who lose their jobs because of trade, especially workers providing services.</i></li> <li>– <i>If we are going to have a real trade package for this country, it has to benefit not just those who win from trade but those who lose from trade as well, including the workers who lose their job through no fault of their own.</i></li> </ul>
small business create job	830	<ul style="list-style-type: none"> <li>– <i>Small businesses create 80 percent of the jobs, so you would think a good piece of the relief would go to small business.</i></li> <li>– <i>Mr. President, small businesses represent more than 99 percent of all employers [...] and create about 75 percent of the new jobs in this country.</i></li> </ul>

dynamic shifts in narratives reflect historical events, we explore the discourse around the September 11th attacks and the subsequent Global War on Terror.<sup>8</sup>

Figure 3.1a shows a time series for a selection of narrative statements pertaining to terrorism and war over the years 1994-2015. First, we see that the narrative on the threat posed by Saddam Hussein spikes in 2002, as the Bush Administration pushed its case for war and for congressional authorization. The supporting narrative of Hussein having or using weapons of mass destruction spikes at the same time (Kull, Ramsay and Lewis 2003), continuing over the subsequent years. Meanwhile, appeals to God blessing America surged in the wake of the September 11th attacks (Klocek and Hassner 2019). Further, starting in 2003 with the Iraq invasion, a second religious narrative of God blessing the troops gained hold and persisted for a decade.

In Appendix I, we show that the statements in Figure 3.1a are part of a broader political story on the American response to 9/11 and the intervention in Iraq. Narratives around Saddam Hussein and Iraq posing a threat are accompanied by a political case for intervention: on people needing help, the nation taking action, and the need to make sacrifices (Appendix Table ??). The constellation of related narratives also includes those used by the anti-war opposition – e.g., that the Bush administration was misleading Americans.

**Popular narratives have emotional resonance.** A commonly discussed feature of narratives is that they appeal to people’s sentiments (Angeletos and La’O 2013; Eliaz and Spiegler 2020). To analyze this dimension in the congressional speeches, we use a sentiment analyzer to score each narrative by the average sentiment of sentences where the narrative appears.<sup>9</sup> We then produce a ranking of narratives by sentiment, both in the positive and negative directions.

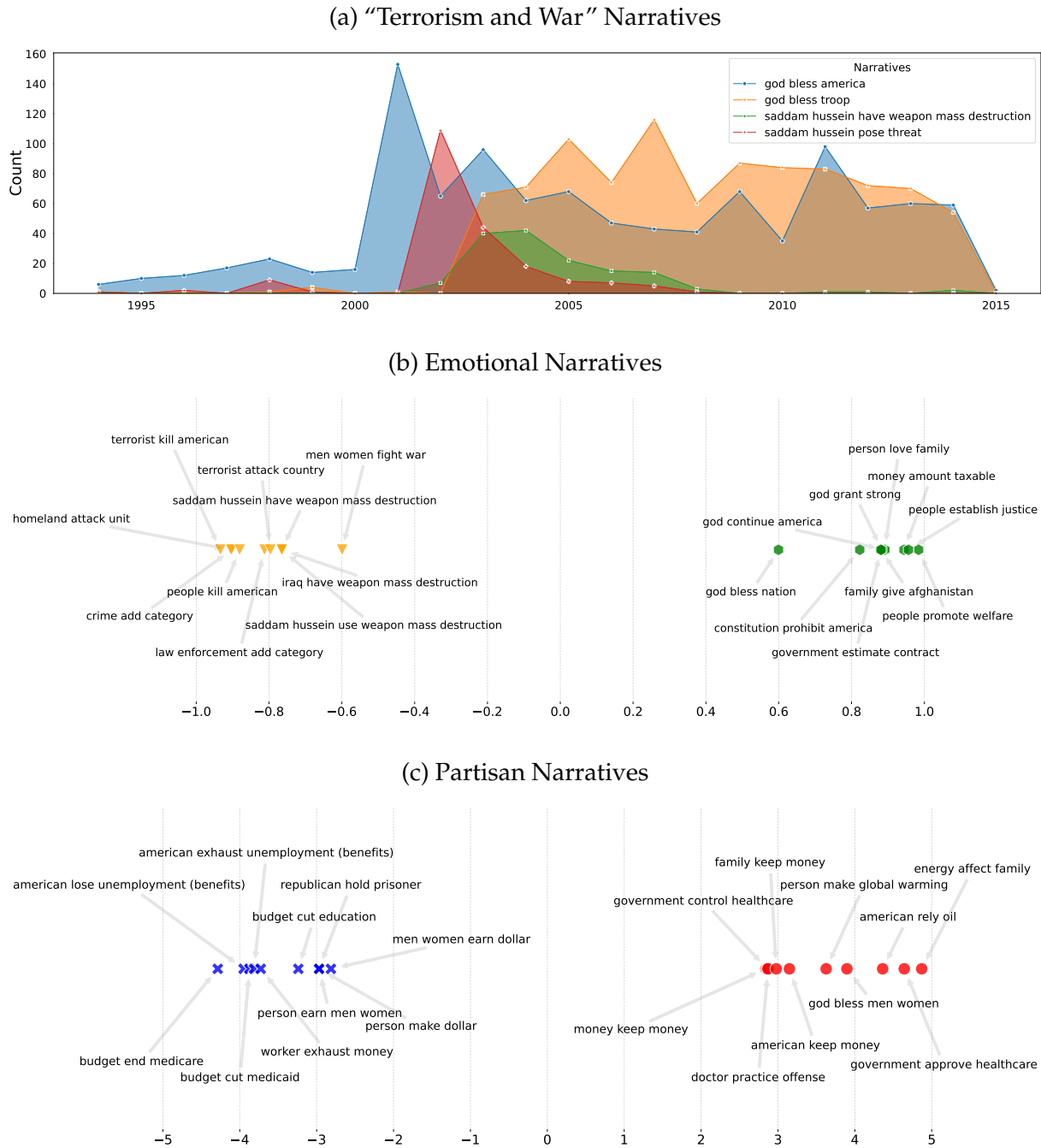
The highest-ranked narratives by positive and negative tone are presented in Figure 3.1b. The most positive narratives include those related to the Constitution and Founding, of the benefits of healthcare, and of small businesses providing jobs. The negative set includes narratives about providing help in times of need and of the threats posed by terrorists.<sup>10</sup> Again, the results provide a qualitative window into the priorities and values held by U.S. Congressmen.

<sup>8</sup>Appendix G shows additional time-series of the most frequent narratives in the US Congress.

<sup>9</sup>We work with Valence Aware Dictionary for Sentiment Reasoning, commonly known as VADER.

<sup>10</sup>“Crime add category” shows up as negatively valent because it includes many sentences related to hate crime. Specifically, “crime” includes “hate crime legislation”, while “category” includes “new categories [to current hate crimes law]”. This narrative was reiterated many times by Senator Gordon Smith, in sentences such as “Each congress, Senator Kennedy and I introduce hate crimes legislation that would add new categories to current hate crimes law, sending a signal that violence of any kind is unacceptable in our society.”

Figure 3.1: History, Sentiment, and Politics in Narrative Discourse



**Note:** This figure shows how narratives may help researchers make sense of speeches in the US Congress. To provide some historical perspective, Panel (a) presents time-series counts of a selection of prevalent narratives in the wake of the 9/11 attacks. The counted narratives are "god bless america" (blue), "god bless troop" (orange), "saddam hussein have weapon mass destruction" (green), and "saddam hussein pose threat" (red). Panels (b) and (c), respectively, plot the 20 most extreme narratives in the U.S. Congressional Record along the sentiment and partisanship dimensions. Panel (b) displays the most positive and negative narratives. The sentiment compound measure, computed using the NLTK VADER package, is averaged over all sentences in which the narrative appears. A high compound sentiment indicates positive sentiment (in green), whereas a low compound sentiment indicates negative sentiment (in orange). Panel (c) displays the most Republican and Democrat narratives. A high log-odds ratio reflects narratives pronounced more often by Republicans relative to Democrats (in red), and vice versa for a low log-odds ratio (in blue).

**Partisan narratives map ideological disagreements.** Scholars have long studied how competing ideologies are reflected in speech, using alternatively n-grams (Ash, Morelli and Van Weelden 2017; Gentzkow, Shapiro and Taddy 2019) or topic models (Quinn et al. 2010). These previous approaches identify those entities and concepts that, when mentioned, tend to signal party affiliation. Our narrative features provide another angle by identifying the *connections* between entities and concepts that, when expressed, signal partisanship.

To explore this dimension, we produce an odds ratio for each narrative as its relative usage by Republicans or Democrats (Monroe, Colaresi and Quinn 2008). Figure 3.1c displays the narratives that are most partisan, with a negative coefficient indicating Democrat-slanted and a positive coefficient indicating Republican-slanted. For example, Democrats are concerned about budget cuts for public programs and Americans losing unemployment benefits, while Republicans care about government interference with healthcare and Christian values.<sup>11</sup> These narratives highlight the mirroring of partisan policy priorities during this recent time period.

To see better how the expressed *connections* between entities signal partisanship, we compute a narrative divisiveness score for each entity, as the average (absolute value) log-odds ratio of the narratives where the entity appears, minus the log-odds ratio of the entity itself. The highest-ranked entities on this score are those agents and patients for which the political parties most differ in their articulated connections to other entities, after adjusting for the partisanship of the entity itself.

The entities with the highest narrative divisiveness are listed in Table 3.2.<sup>12</sup> Panel A shows the top-ten policy-related entities, while Panel B shows the top-ten entities related to identity groups or symbols. The second and third columns, respectively, show the most Democrat-slanted and most Republican-slanted narratives associated with the entity, again as measured by the log-odds ratio.

The list of divisive entities illustrate how the same entities can be used in very different narratives by different parties. On the policy side (Panel A), Democrats lament the profits earned by *oil* companies, while Republicans retort that Americans rely on *oil*. Democrats worry about *budget* cuts to Medicare and Medicaid, while Republi-

<sup>11</sup>The Republican narrative “government approve healthcare” reflects sentences such as “The Virginia court held that the individual mandate requiring every American to purchase government-approved health insurance was unconstitutional.” or “This bill takes away that freedom, requiring every American to purchase a government-approved health plan, pay a tax, or even go to jail.”

<sup>12</sup>We filtered out some infrequent entities that ranked highly in the list due to noise. To be included in this list, the entity must appear in at least six unique narratives, and have at least three Democrat-slanted narratives and three Republican-slanted narratives. Appendix Table H.1 shows a longer, unfiltered list of divisive entities, with the associated scores. For comparison, a list of the least divisive entities are shown in Appendix Table H.2.



Table 3.2: Entities Associated with the Most Divisive Narratives

Entity	Democrat Narratives	Republican Narratives
<b>A. Policy-Related Entities</b>		
oil	oil pay fee oil make profit	oil create job american rely oil
budget	budget end medicare budget cut medicaid	budget raise tax budget increase tax
healthcare	child lose healthcare family afford healthcare	government control healthcare government approve healthcare
interest/rate	reserve raise interest/rate fed raise interest/rate	individual bearing interest/rate capital gain interest/rate
worker	worker exhaust money worker have exhaust	employee hire worker worker support retiree
loan	small business receive loan student take loan	people get loan company hold loan
insurance	insurance deny coverage insurance drop coverage	american buy insurance government run insurance
medicare	budget end medicare republican end medicare	medicare run money doctor see medicare
energy	energy create job nation need energy	american produce energy energy affect family
job	company move job company ship job	american not-do job tax kill job
<b>B. Identity and Symbolic Entities</b>		
men/women	person earn men/women men/women earn dollar	men/women defend nation god bless men women
democrat	republican join democrat democrat balance budget	democrat do nothing democrat raise tax
constitution	constitution prohibit america men women write constitution	constitution give authority constitution give power
american	american lose unemployment (benefits) american exhaust unemployment (benefits)	american keep money american rely oil
community	community need help program provide community	community perform abortion community promote abortion
republican	republican hold prisoner republican refuse action	republican gain power democrat join republican
nation	nation sign treaty nation make progress	men women defend nation nation face fight
people	people lose unemployment (benefits) people face mortgage	people call tax people keep money
family	family afford healthcare family lose healthcare	family keep money energy affect family
person	person earn men/women person make dollar	person love family person make global warming

**Note:** This table shows the set of entities that appear in narratives with a high average log-odds ratio by partisan mentions, adjusted for the log-odds ratio of the entity itself. Panel A includes policy-related entities, while Panel B includes entities are related to groups or ideas. Democrat Narratives and Republican Narratives show the associated narratives with the lowest and highest log odds ratios, respectively.

cans are concerned that the *budget* will increase taxes. Meanwhile, Democrats complain that companies are shipping *jobs* overseas, while Republicans assert that taxes kill *jobs*. On the identity side (Panel B), Democrats attend to *men and women's* earnings, while Republicans celebrate their defense of the nation and invoke God's blessings. For *Americans*: While Democrats emphasize that "Americans lose unemployment benefits", Republicans stress that "Americans rely on oil". Overall, the narratives associated with divisive entities illuminate the key divergences in political priorities between Democrats and Republicans. Identifying such partisan connections would be infeasible with standard text-as-data approaches, such as n-grams.

A lingering question is whether mentions of narratives – that is, connections between entities – are overall more polarized than mentions of singular entities. We test this possibility formally following the approach in [Peterson and Spirling \(2018\)](#). Specifically, we train machine classifiers to predict a speaker's partisanship (Republican or Democrat) in held-out data, using as features either the entities or the narratives pronounced in speeches.<sup>13</sup> The narrative features predict partisanship more accurately: we obtain an out-of-sample accuracy of 81% with narratives, but only 74% for entities.<sup>14</sup> The higher test-set accuracy for narratives suggest that they are more informative about partisanship than mentions of topics or specific entities. The *connections* between entities framed by narratives help politicians tell stories in line with their partisan values.

**Narratives reveal the connected structure of political debates.** So far, we have analyzed narratives in isolation, as two entities connected by one verb. Yet, a defining feature of narratives is that simple statements link up to form more complex stories.<sup>15</sup> As previewed in Section 2.3, we map the narrative space of U.S. congressional speeches as a directed multigraph, with entities composing nodes and verbs composing link between nodes.

The resulting graph of linked entities can be used for a variety of network-based analyses. In particular, centrality measures can be used to determine which entities are pivotal to political narrative structure (e.g. [Hanneman and Riddle 2005](#)). Entities with

---

<sup>13</sup>Our machine learning model is L2-penalized logistic regression. We select the regularization strength using five-fold cross-validated grid search in a 75% training set and evaluate performance in a 25% test set.

<sup>14</sup>To obtain the confidence intervals for these accuracy measures, we employ a five-fold cross-validation in the test set. Averaged across folds, the accuracy is 77% and 73% for narratives and entities, respectively. The corresponding 95%-confidence intervals are [74%, 80%] and [68%, 78%]. The test set contains only 244 speakers; this small sample size across folds likely contributes to the wide confidence intervals.

<sup>15</sup>For example, conspiracy theory narratives have shown to become extremely complex in their proposed connections between disparate entities (e.g., [Wood and Douglas 2015](#); [Tangherlini et al. 2020](#)).

high closeness centrality, for example, are the most connected to the broader narrative network.<sup>16</sup> Out-degree centrality captures influential agents who tend to act on many other entities,<sup>17</sup> while in-degree centrality captures the most receptive patients – for example, policies to be enacted and the beneficiaries of those policies.<sup>18</sup> Such analyses can be further enriched with metadata, notably political party of the speakers.

The narrative graphs are useful to support qualitative descriptive analysis of political corpora. Using an interactive web browser applet, even the full set of 1,638 congressional narratives – 353 entities interwoven by 140 verbs – can be explored efficiently and informatively. Since the entire network is too large to depict as a static figure, we visualize the 100 most frequent narratives here. Figure 3.2 shows this subset of the narrative graph, with thicker arrows indicating higher frequency and colors indicating partisanship – blue for Democrat, and red for Republican. The network summarizes in a single figure many of the stylized facts we have previously discussed. For example, while Democrats lament the problem of “people losing jobs”, Republicans applaud the success of “small businesses creating jobs”. While Democrats want that “Americans have healthcare”, Republicans worry about “government-run healthcare”. Overall, the graph provides a distilled view on the *worldviews* expressed in the U.S. Congress.

Figure 3.2 is just one example of many potentially informative visualizations of the narrative networks. Appendix H provides party-specific network graphs (see Figures H.2 and H.3). Moreover, our GitHub repository provides interactive versions of such network graphs.<sup>19</sup>

## 4. Discussion

To recap, we have designed and implemented a new text-as-data method that provides an intuitive mapping from a corpus of plain-text sentences to a sequence of low-dimensional narrative statements. In this section, we discuss how narrative mining fits in with other text-as-data methods, identify some of its limitations, and propose potential extensions.

---

<sup>16</sup>In our corpus, the top-five entities by closeness centrality are job, family, healthcare, funding, program.

<sup>17</sup>The top-five entities by out-degree centrality are people, Americans, program, government, and person.

<sup>18</sup>The top-five entities by in-degree centrality are job, healthcare, money, family, and services.

<sup>19</sup>See <https://github.com/relatio-nlp/relatio> and the links therein.



## 4.1 When should researchers use RELATIO?

In the application sections above, we focus on results that would be difficult to produce with other text-as-data methods. Mainly, this means looking closely at the narratives themselves. In Figure 3.1a, the depicted narratives capture the dynamics of War on Terror discourse better than plots of specific phrases like “Saddam Hussein” or “God” by themselves. Further, the most emotive and partisan narratives from Table 3.2 and Figure 3.1 are more informative than standard representations (like bag-of-words) because they map explicit relationships, requiring less contextual knowledge to interpret them. Finally, the narrative graphs exemplified by Figure 3.2 have no analogue in other commonly used text methods.

The promise of this mode of quantitative description using narratives is not limited to our particular setting. RELATIO can be used to track the evolution of language in a range of corpora, such as social media and newspapers. In particular, the narrative graphs offer a novel opportunity for data-driven study of worldviews. Appendix J provides an additional application of our method to President Donald Trump’s tweets from 2011 through 2020. Furthermore, since we open-sourced our Python package in August 2021, two working papers have come out using the code for descriptive analysis in different settings, including [Sipka, Hannak and Urman \(2021\)](#) on social media posts about Q-Anon conspiracy theories and [Ottonello and Song \(2022\)](#) on newspaper coverage of banks.

To take a broader view, we now discuss how narrative mining complements existing text-as-data methods. First, the narrative features output by RELATIO can be examined and analyzed the same way that specific word or phrase patterns are used in common dictionary methods. For example, [Baker, Bloom and Davis \(2016\)](#) count articles mentioning both an economy-related word and an uncertainty-related word to measure “economic policy uncertainty.” A RELATIO-based alternative could count narratives containing economy-related entities linked with uncertainty-related attributes or entities. The advantage of the narrative approach over dictionaries is that narratives can more easily specify semantically subtle links, such as distinguishing uncertainty caused by the economy, versus uncertainty caused by policy.

Second, the outputs of RELATIO are similar to approaches using syntactic dependency parsing to identify agents, actions, and objects. In principle, semantic role labeling is designed to out-perform syntax-based methods for this purpose; so, for example, RELATIO could be used to produce measures of relative worker-firm authority in labor union contracts, as an alternative to the dependency-based approach from [Ash et al. \(2020\)](#). A caveat is that SRL is a more complex and error-prone linguistics task than

dependency parsing. The costs of those parsing errors might outweigh the benefits from using SRL, but the methods have not been systematically compared.

Third, topic models such as LDA (Blei, Ng and Jordan 2003) are, like RELATIO, designed to perform informative dimension reduction of documents. Narrative statements could provide an alternative to topics for research questions concerning the prevalence of specific concepts or issues. For example, Catalinac (2016) applies a topic model to parliamentary speeches, manually inspects and labels topics as being related to local pork or national policy, and then examines changes in the pork/policy categories in response to an institutional reform. An alternative approach using narrative mining would be to extract narratives (rather than topics) from those speeches and then label the narratives as being related to pork or policy. Whether topics or narratives are preferred will depend on the use case. Broadly speaking, topic models will identify the prevalence of relatively generic topics in documents, while RELATIO will detect the presence of distinctive arguments or claims.

Besides being useful by themselves, the narrative statements produced by RELATIO can be used as an informative and interpretable feature set in downstream text-as-data applications. For example, document-level counts or frequencies over narrative statements provide an alternative to N-gram frequencies, topic shares, or document embeddings as inputs to text classifiers or text regressions. Similarly, document distance can be computed using cosine similarity between narrative frequency vectors the same way it can be done between N-gram, topic, or embedding vectors. Finally, topic models like LDA can be applied on top of the narrative representation of documents, allowing narrative statements to be topically bundled for further descriptive or empirical analysis.<sup>20</sup>

A systematic comparison of RELATIO features to these existing alternatives for such tasks is a high priority for future work. Our analysis above showing that narrative features are more predictive of partisanship than entity features provides a promising indication in this direction. Holding the quality of the algorithmic outputs constant, narratives might be preferred to other feature sets due to their interpretability and tractability. RELATIO extracts distinctive entity-action tuples which summarize the core claims made in a corpus. This type of interpretable dimension reduction serves the same goals as other text-as-data methods, such as TF-IDF weighting (Grimmer

---

<sup>20</sup>Along those lines, Appendix F presents a topic model on the Congress corpus and compares the words associated with each topic to the narratives associated with each topic. The top narratives tend to provide additional information about topic content which top words fail to capture. Consider the topic we labeled as “economy”. The top (stemmed) words for this topic are “busi”, “small”, “compani”, “capit”, “invest”, “contract”, “administr”, and “job”. The associated narratives describe explicit relationships such as “small business employ workforce”, “individual receive funding”, or “founder start business”.

and Stewart 2013), supervised feature selection (Monroe, Colaresi and Quinn 2008), or filtering on parts of speech to extract noun phrases (Handler et al. 2016). For either supervised or unsupervised learning algorithms, the dimension-reduced inputs will be computationally tractable, while the learned outputs will be informative and interpretable.

## 4.2 Limitations and potential extensions

While our narrative mining system shows some promising preliminary results, there are still a number of limitations and opportunities for improvement.

**SRL quality.** The quality of narrative outputs depends on the quality of the semantic role labeling (SRL) tags. When SRL fails, it produces nonsense data that will mostly be dropped in the pipeline’s filtering steps. The quality of SRL tags correspondingly depends on the quality of text inputs. Hence, messy text with digitization errors, for example due to optical character recognition (OCR), may not produce usable results. A further source of SRL errors is grammatical complexity; SRL performs best on simpler language, such as that in social media posts. The *Congressional Record* – where long, potentially rambling spoken sentences are the norm – is actually a difficult test case for the method.

From the perspective of empirical research design, a relationship between speech complexity and SRL quality is concerning because the performance of our method may differ across subgroups. In our setting, for example, we have evidence that Democrats use somewhat more sophisticated language than Republicans.<sup>21</sup> If SRL errors are higher for these more sophisticated sentences, then they will tend to be selectively dropped from the sample, potentially biasing downstream results.

Our hope is that these issues will diminish as more robust automated SRL models are introduced. For more difficult corpora, such as historical speeches digitized by OCR, it may be fruitful to adapt the approach to work with syntactic dependencies rather than semantic roles.

**Entity extraction.** A persistent practical difficulty is in the extraction of clustered entities. With more clusters, the entities become more coherent and specific, at the expense of higher dimensionality. With fewer clusters, semantically related yet contradictory

---

<sup>21</sup>The number of words per narrative is higher for Democrats, meaning that a narrative entails a marginally higher abstraction. The difference is not statistically significant, however.

phrases – e.g., “tax hike” and “tax cut” – will often be combined into the same entity. The standard cluster quality metrics, such as silhouette scores, provide limited guidance.

As with SRL, this is partly due to limitations with the associated algorithms. It could be that improved or domain-specific named entity recognition, alternative phrase encoding approaches (e.g., [Reimers and Gurevych 2019](#)), or alternative clustering approaches (e.g., [Stambach and Ash 2021](#)), could mitigate these problems. The software package will be continuously updated following such developments.

Even with improved algorithms, however, entity extraction may produce imperfect outputs. For example, say one wants to resolve “american” and “people” to one entity – e.g., for “american lose job” and “people lose job”. This is currently not straightforward: by default, “american” is a named entity while “people” comes from a clustered embedding. A further wrinkle is that the optimal clustering may be context-dependent: for instance, clustering “people” and “worker” together might make sense for the “[...] lose job” narrative, but it may erase useful information in others – e.g., “government support worker” has a different ideological valence from “government support people”. Finally, our entity extraction algorithm cannot directly identify implicit references to entities, such as using the phrase “White House” to refer to the president. In sum, obtaining the entities most relevant to an application with RELATIO is art as much as science.

**Further extensions and limitations.** There are a number of additional extensions to consider. In the SRL step, additional semantic roles could be included, such as temporality or adverbial clause modifiers. The named entity recognizer could be improved to resolve co-references to the same entity. Further experimentation could be done with dimension-reducing verbs, for example by combining an embedding-based and dictionary-based approach (to prevent clustering of antonyms).

Another notable limitation is that the current version only works in the English language. In principle, however, the approach should be applicable in any language where a pre-trained SRL model is available. For small corpora, a pre-trained phrase encoder might also be required. Given the growing plethora of NLP tools across languages, we suspect that these requirements will not pose a significant impediment to diverse applications.





# Bibliography

- Akerlof, George A, and Robert J Shiller.** 2010. *Animal Spirits: How Human Psychology drives the Economy, and why it Matters for Global Capitalism*. Princeton University Press.
- Angeletos, George-Marios, and Jennifer La'O.** 2013. "Sentiments." *Econometrica*, 81(2): 739–779.
- Armstrong, Paul B.** 2020. *Stories and the Brain: The Neuroscience of Narrative*. John Hopkins University Press.
- Arora, Sanjeev, Yingyu Liang, and Tengyu Ma.** 2016. "A Simple but Tough-to-beat Baseline for Sentence Embeddings."
- Ash, Elliott, Germain Gauthier, and Philine Widmer.** 2022. "Replication Data for: "RELATIO: Text Semantics Capture Political and Economic Narratives", <https://doi.org/10.7910/DVN/3BRWKK>, Harvard Dataverse, V1."
- Ash, Elliott, Jeff Jacobs, Bentley MacLeod, Suresh Naidu, and Dominik Stambach.** 2020. "Unsupervised Extraction of Workplace Rights and Duties from Collective Bargaining Agreements." 766–774, IEEE.
- Ash, Elliott, Massimo Morelli, and Richard Van Weelden.** 2017. "Elections and divisiveness: Theory and evidence." *The Journal of Politics*, 79(4): 1268–1285.
- Baker, Scott R, Nicholas Bloom, and Steven J Davis.** 2016. "Measuring Economic Policy Uncertainty." *The Quarterly Journal of Economics*, 131(4): 1593–1636.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan.** 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, 3: 993–1022.
- Bonial, Claire, Jena Hwang, Julia Bonn, Kathryn Conger, Olga Babko-Malaya, and Martha Palmer.** 2012. "English Propbank Annotation Guidelines." *Center for Computational Language and Education Research Institute of Cognitive Science University of Colorado at Boulder*, 48.
- Branch, William, Bruce McGough, and Mei Zhu.** 2017. "Statistical Sunspots."

- Bruner, Jerome.** 1991. "The Narrative Construction of Reality." *Critical Inquiry*, 18(1).
- Bybee, Leland, Bryan T Kelly, Asaf Manela, and Dacheng Xiu.** 2020. "The Structure of Economic News." National Bureau of Economic Research.
- Bénabou, Roland, Armin Falk, and Jean Tirole.** 2018. "Narratives, Imperatives, and Moral Reasoning." National Bureau of Economic Research Working Paper 24798.
- Catalinac, Amy.** 2016. "From Pork to Policy: The Rise of Programmatic Campaigning in Japanese Elections." *Journal of Politics*, 78(1).
- Eliaz, Kfir, and Ran Spiegler.** 2020. "A Model of Competing Narratives." *American Economic Review*, 110(12): 3786–3816.
- Enke, Benjamin.** 2020. "Moral Values and Voting." *Journal of Political Economy*, 128(10): 3679–3729.
- Gardner, Matt, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew Peters, Michael Schmitz, and Luke S. Zettlemoyer.** 2017. "AllenNLP: A Deep Semantic Natural Language Processing Platform."
- Gennaro, Gloria, and Elliott Ash.** 2021. "Emotion and Reason in Political Language." *The Economic Journal*.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy.** 2019. "Text as Data." *Journal of Economic Literature*, 57(3): 535–74.
- Gentzkow, Matthew, Jesse M Shapiro, and Matt Taddy.** 2019. "Measuring group differences in high-dimensional choices: method and application to congressional speech." *Econometrica*, 87(4): 1307–1340.
- Grimmer, Justin, and Brandon M. Stewart.** 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*, 21(3): 267–297.
- Handler, Abram, Matthew Denny, Hanna Wallach, and Brendan O'Connor.** 2016. "Bag of What? Simple Noun Phrase Extraction for Text Analysis." 114–124.
- Hanneman, Robert A, and Mark Riddle.** 2005. "Introduction to social network methods."
- Hansen, Stephen, Michael McMahon, and Andrea Prat.** 2017. "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach." *The Quarterly Journal of Economics*, 133(2): 801–870.

- Jurafsky, Dan, and James H Martin.** 2020. *Speech and Language Processing. Third Edition.*
- Klocek, Jason, and Ron E Hassner.** 2019. "War and Religion: An Overview." *Oxford Research Encyclopedia of Politics.*
- Kuhle, Wolfgang.** 2020. "Thought Viruses and Asset Prices." *Journal of Behavioral Finance*, 1–9.
- Kull, Steven, Clay Ramsay, and Evan Lewis.** 2003. "Misperceptions, the Media, and the Iraq War." *Political Science Quarterly*, 118(4): 569–598.
- Lauderdale, Benjamin E, and Alexander Herzog.** 2016. "Measuring political positions from legislative speech." *Political Analysis*, 374–394.
- Monroe, Burt L, Michael P Colaresi, and Kevin M Quinn.** 2008. "Fightin' words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis*, 16(4): 372–403.
- Ottonello, Pablo, and Wenting Song.** 2022. "Financial Intermediaries and the Macroeconomy: Evidence from a High-Frequency Identification." National Bureau of Economic Research.
- Patterson, Molly, and Kristen Renwick Monroe.** 1998. "Narrative in Political Science." *Annual review of political science*, 1(1): 315–331.
- Peterson, Andrew, and Arthur Spirling.** 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis*, 26(1): 120–128.
- Polkinghorne, Donald E.** 1988. *Narrative Knowing and the Human Sciences.* Suny Press.
- Proksch, Sven-Oliver, and Jonathan B Slapin.** 2012. "Institutional Foundations of Legislative Speech." *American Journal of Political Science*, 56(3): 520–537.
- Quinn, Kevin M, Burt L Monroe, Michael Colaresi, Michael H Crespin, and Dragomir R Radev.** 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science*, 54(1): 209–228.
- Reimers, Nils, and Iryna Gurevych.** 2019. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." *arXiv preprint arXiv:1908.10084.*
- Shiller, Robert J.** 2019. *Narrative Economics: How Stories go Viral and Drive major Economic Events.* Princeton University Press.

- Sipka, Andrea, Aniko Hannak, and Aleksandra Urman.** 2021. "Comparing the Language of QAnon-related content on Parler, Gab, and Twitter." *arXiv preprint arXiv:2111.11118*.
- Sloman, Steven.** 2005. *Causal models: How People Think About the World and its Alternatives*. Oxford University Press.
- Stammbach, Dominik, and Elliott Ash.** 2021. "DocSCAN: Unsupervised Text Classification via Learning from Neighbors." *arXiv preprint arXiv:2105.04024*.
- Tangherlini, Timothy R, Shadi Shahsavari, Behnam Shahbazi, Ehsan Ebrahimzadeh, and Vwani Roychowdhury.** 2020. "An Automated Pipeline for the Discovery of Conspiracy and Conspiracy Theory Narrative Frameworks: Bridgegate, Pizzagate and Storytelling on the Web." *PloS one*, 15(6): e0233879.
- White, Hayden.** 1980. "The Value of Narrativity in the Representation of Reality." *Critical Inquiry*, 7(1): 5–27.
- Wood, Michael J, and Karen M Douglas.** 2015. "Online Communication as a Window to Conspiracist Worldviews." *Frontiers in Psychology*, 6: 836.

# Appendix

## A. Python Implementation with RELATIO

We detail the programmatic implementation of the procedure described in this paper, which is freely distributed as the Python package RELATIO. Our pipeline takes a collection of unstructured text documents as input and returns narrative statements (see Equation 3.2). Figure 2.2 presents a graphical outline of the different steps.

First, we split the raw text into sentences using spaCy (Honnibal et al. 2020). Second, we annotate the semantic roles (SRL). For this task, we rely on a state-of-the-art pre-trained model implemented by Gardner et al. (2017).<sup>1</sup> SRL decomposes every sentence into one or multiple role sequences as defined in Equation 3.1. The current implementation covers agents (ARG0), verbs (V), patients (ARG1), attributes (ARG2), modals (ARGM-MOD) and negations (ARGM-NEG). An example of semantic role labeling annotations is provided in Figure 2.1.

There are hundreds of thousands of unique semantic roles. We seek to regroup them into a smaller set of interpretable groups. An easy step in this direction is to “clean” all roles. We thus lowercase and lemmatize each word. We also drop punctuation, numbers and words from a list of pre-defined stopwords (for the complete list, see M). After these cleaning steps, we drop roles of more than four tokens as they are often uninterpretable.

We do not process the verbs (V) any further, but connect them to their negation (if any). For agents (ARG0), patients (ARG1) and attributes (ARG2), we rely on two methods: Named Entity Recognition (NER) and phrase embedding clustering. We first run spaCy’s Named Entity Recognizer (NER) to identify all named entities mentioned in the corpus. After careful inspection of this list, we keep the 1000 most frequent entities and assign to them their own cluster. Beyond this threshold, named entities are pronounced less than 40 times and are often idiosyncratic. If a role contains one of these named entities, then it is designated as that entity and not processed any further.

For all of the remaining roles that do not contain a frequent named entity, we produce text-embedding clusters as follows. First, we apply a phrase encoder to compress the plain text snippet to a low-dimensional dense vector. There are many encoders fit for this purpose (e.g. Sentence-BERT (Reimers and Gurevych 2019) and its variants). As a computationally efficient default, we use Arora, Liang and Ma’s “simple but tough-to-beat” sentence embeddings, which consist of the weighted average of the GloVe word vectors (Pennington, Socher and Manning 2014) across each word in the role. The weights are chosen to be inversely proportional to the frequency of the

<sup>1</sup>See <https://demo.allennlp.org/semantic-role-labeling>

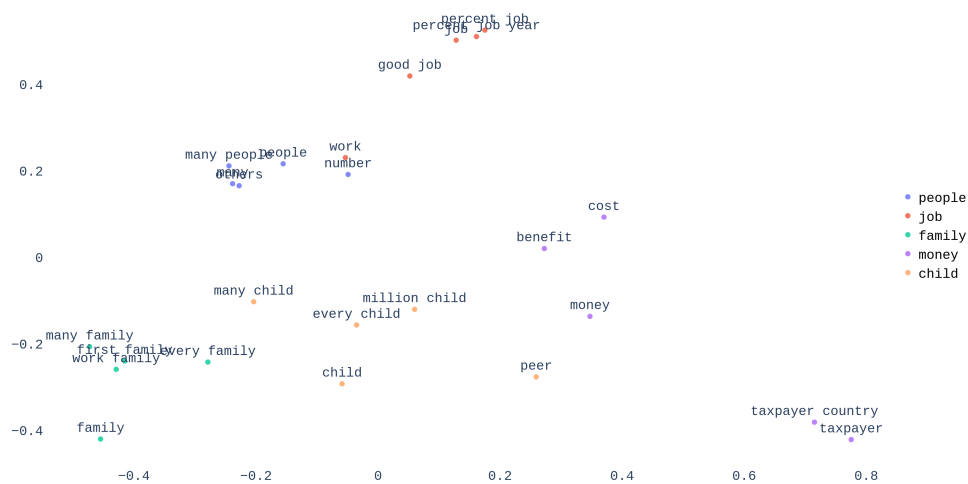
word in the corpus. This means that relatively rare, more distinctive words are up-weighted in the phrase embedding. We use `gensim`'s pre-trained GloVe embeddings with 300 dimensions (Řehůřek and Sojka 2010). The result of this procedure is a vector in the embedding space for each agent, patient and attribute, where linguistically similar phrases (e.g. “higher taxes” or “high tax rates”) have geometrically proximate vector representations. A graphical depiction is presented in Figure A.1.

Exploiting this geometry, we further reduce the dimensionality of our role sets by unsupervised clustering applied to the associated matrix of encoded vectors. As with the phrase encoding step, there are a number of possible approaches for clustering (e.g., DBSCAN, agglomerative clustering). We obtain effective performance with `scikit-learn`'s implementation of the conventional K-means clustering algorithm, which is also computationally lightweight (Pedregosa et al. 2011). For K-means, the only hyperparameter is the number of clusters. Each remaining agent, patient and attribute in the corpus is assigned to a cluster. For interpretability, each cluster is labeled by the most frequent term within the cluster. Given the size of the corpus, we randomly select 50,000 speeches (roughly 5% of the entire corpus) to train the K-Means model, and predict out-of-sample for the remaining speeches. We cluster agents, patients and attributes together. In turn, the same entity can appear as an agent, a patient or an attribute. This joint clustering ensures that roles can be connected easily in a directed multigraph.

Finally, we focus on all narratives that have either an Agent-Verb-Patient or an Agent-Verb-Attribute structure. As the distribution of narrative frequencies is heavily skewed to the left, and since we are interested in recurring narratives, we restrict our attention to narratives pronounced at least 50 times in the corpus (i.e., at least twice a year on average).



Figure A.1: Clustered Embeddings Recover Coherent Entities



**Note:** This figure illustrates the embedding-cluster process. We take the five most frequent entity clusters (produced using *K*-means applied to the phrase embeddings): people, job, family, money, and service. The embedding locations of the most frequent phrases in each cluster are illustrated here in a two-dimensional projection. The embedding space both co-locates and clusters similar phrases, providing dimension reduction while preserving semantically coherent entities. Embeddings are projected to two dimensions using PCA.

## B. Human Validation

### Tuning the number of clusters

The number of mined entities is the core hyperparameter in our analysis. It depends on the number of named entities recognized, as well as the number of clusters we specify for K-Means. To pick the correct number of named entities in our analysis, we manually inspect the most frequent named entities in the corpus. We fix a threshold at 1000 named entities, above which the named entities are pronounced fewer than 40 times and tend to become idiosyncratic. We keep this threshold fixed throughout the rest of the analysis.

Contrary to the number of named entities, the number of clusters to specify for K-Means is not trivial for at least two reasons. First, it is unclear how many unnamed entities are being referred to in the corpus. Second, there is a trade-off between the informativeness of clusters and the magnitude of the dimension reduction. If a small number of clusters is specified, then clusters are likely to become abstract and to encompass a broad variety of phrases. If a large number of clusters is specified, then clusters will be more coherent (and thus potentially more interpretable), but they can also become redundant, with multiple clusters referring to the same underlying entity. This trade-off suggests that we would like to maximize cluster coherence while minimizing cluster redundancy.

We investigate the optimal number of clusters with a simple human validation task. The task is performed by freelancers on Upwork who were based in the U.S. and whose primary language is English. First, we run four clustering scenarios and specify respectively 100, 500, 1000, and 2000 clusters. We then provide annotators with a random sample of semantic roles and their associated cluster labels (from different cluster numbers). Additionally, we add placebo observations, where we show again a random sample of semantic roles, but instead of the predicted cluster label, we combine every role with a random cluster label (drawn from the entire pool of available labels). Annotators are then asked to rank the similarity of the semantic role and the cluster label on a scale of 1 (not similar at all) to 10 (extremely similar/the same). Our random validation sample consists of 4000 unique role and label combinations (500 true and 500 placebo combinations for every clustering scenario). Every unique combination is, on average, annotated by 2.5 freelancers, resulting a total sample of 10000 annotated combinations. Annotators do not see the number of clusters that were used to generate the cluster label, nor whether the semantic role is combined with a predicted cluster label or a placebo label. This procedure jointly evaluates cluster coherence and redun-

dancy (coherence is reflected in how similar the semantic roles are to their predicted label; redundancy is captured by similarities between random combinations of roles and clusters).

## Annotator recruitment

In June 2021, we hired three different freelancers on Upwork to annotate 4000 combinations (applies to two freelancers) or 2000 combinations (applies to one freelancer) of roles and labels. The instructions as communicated in the job post are shown in Figure B.1.

Figure B.1: Wording of the validation job post on Upwork

*“We need freelancers (U.S.-based, fluent in English) to check the similarity between two phrases. In the attached file, please compare the phrase in column 1 and 2. The question is: for every row, how similar are the entities described in column 1 and 2? Please indicate from 0 (not similar at all) to 10 (very similar/the same). Note that you can select values in between 0 and 10 (e.g., 5). This task is about evaluating a computer-based language model from a human perspective. It is about your “hunch” or intuition of how similar the phrases are. Do they refer to a similar entity?”*

Upon posting the job, we sent a sample of 20 combinations to the first five freelancers who expressed their interest in the job. Out of these five, we hired the first three freelancers who submitted the annotated sample (all of these first three had annotated the sample correctly). In Figure B.2, we feature the form presented to the freelancers for annotation.

Figure B.2: Screenshot of Human Annotation File

ARG	ARG-RAW	similarity
water	soil water degradation	
tsa	voter supporter	
care	support treatment	
democrat	sherry life	
bad	product substance regulate	
constitution	constitution high law	
opportunity	many commercial advantage	
labor	nation overall growth	
objection	madison marketing	

## Human validation Results

Table B.1 summarizes the key figures of the human validation across numbers of clusters. The average similarity of clustered roles is 4.49; for the placebos, this figure amounts to 2.57.

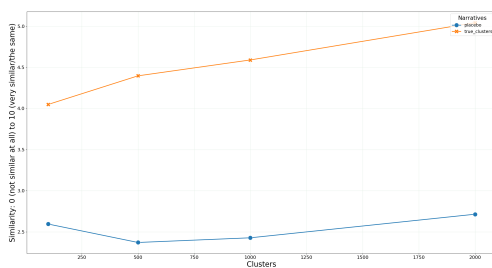
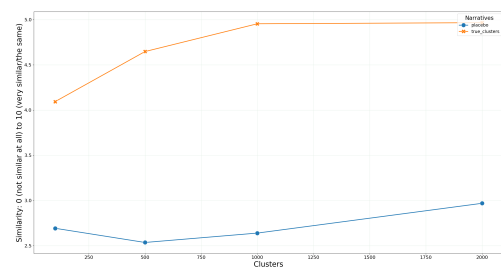
Table B.1: Summary Statistics on the Human Annotations

Total combinations annotated	10000
Unique combinations annotated	4000
Number of annotators	3
Average number of annotators per combination	2.5
Average number of annotations per annotator	3334
Average similarity (true combinations)	4.49
Average similarity (placebo combinations)	2.57
Average inter-annotator similarity standard deviation	2.42

Figure B.3 decomposes the results by number of clusters: it shows the similarity between a random semantic role token and its cluster label for actual (role, cluster) tuples (yellow) and placebo tuples (blue; where roles are combined with a random cluster label), as annotated by human freelancers. Subfigure (1) shows the results for all arguments (that appear in narratives above our usual frequency threshold of 50, see Section 3.1). Here, the similarity increases for the clustering specifications studied here (100, 500, 1000, 2000). However, after 1000 clusters, the placebo similarity also increases.

Next, we seek to understand whether more clusters lead to precision gains for frequent or infrequent entities. Thus, in Subfigure (2), we show the subset of validated arguments that belong to the top 25% of entities (for the respective cluster numbers). Here, increasing the number of clusters seems to yield lower similarity gains after 1000 clusters. These results suggest that higher cluster numbers may be beneficial for less frequent entities. Again, however, the placebo similarity increases more after 1000 clusters. In our application, we focus on 1000 clusters.

Figure B.3: Similarity of semantic roles and cluster labels as reported by humans

All arguments  
(1)Top 25% arguments (by frequency)  
(2)

**Note:** This figure shows the similarity between a random semantic role token and its cluster label for actual (role, cluster) tuples (yellow) and placebo tuples (blue; where roles are combined with a random cluster label), as annotated by human freelancers. Subfigure (1) shows the results for all arguments (that appear in narratives above our usual frequency threshold of 50). Here, the similarity increases for the clustering range studied here. However, after 1000 cluster, the placebo similarity also increases. Subfigure (2) shows the subset of validated arguments that belong to the top 25% of entities (for the respective cluster numbers). Here, increasing the number of clusters seems to yield lower similarity gains after 1000 clusters.

## C. Inspecting the Quality of Mined Entities

Table C.1: Most Frequent Clustered Entities (i)

Label	Most frequent phrases	Frequency
people	people - 182086 — many - 28671 — others - 17168 — many people - 7433 — number - 6196 — have - 5907 — several - 2299 — many others - 1413 — include - 883 — many country - 684	195976
money	money - 30048 — taxpayer - 15388 — benefit - 9898 — cost - 9534 — pay - 2496 — expense - 1298 — amount money - 831 — cash - 744 — taxpayer money - 657 — earn - 529	70503
family	family - 37326 — life - 25338 — work family - 2310 — great - 1981 — many family - 1444 — real problem - 585 — family business - 537 — well life - 524 — great country - 438 — real people - 430	67615
someone	someone - 17689 — anyone - 13401 — everyone - 12217 — everybody - 10819 — somebody - 9924 — anybody - 8159 — nobody - 7671 — whoever - 1030 — someone else - 833 — else - 739	66452
person	person - 30540 — woman - 22444 — man - 10881 — every person - 755 — many woman - 582 — woman child - 545 — first woman - 454 — another person - 404 — man woman - 336 — million woman - 266	58586
job	job - 40808 — work - 16726 — good job - 1357 — job do - 648 — create job - 584 — great job - 548 — work people - 510 — well job - 405 — good pay job - 356 — do job - 353	56871
country	country - 53529 — people country - 2428 — around - 725 — along - 626 — people across country - 388 — across country - 354 — across - 350 — people around world - 311 — around world - 255 — country around world - 229	54044
child	child - 40835 — million child - 1278 — every child - 1224 — peer - 779 — many child - 714 — child family - 593 — child care - 508 — child support - 440 — care need - 439 — family child - 289	43809
business	business - 20863 — industry - 8018 — many company - 719 — management - 719 — enterprise - 691 — foreign company - 628 — financial - 433 — business community - 397 — many business - 386 — management plan - 297	40514
employee	employer - 20943 — employee - 16335 — contractor - 3404 — small employer - 475 — employer employee - 416 — many employer - 393 — private contractor - 334 — percent employer - 232 — many employee - 182 — large employer - 168	39434
law	law - 41489 — public law - 1375 — tax law - 488 — legislative branch - 384 — legal system - 361 — civil action - 327 — legal status - 307 — judicial review - 246 — judicial system - 223 — judicial branch - 172	38472

Table C.2: Most Frequent Clustered Entities (ii)

Label	Most frequent phrases	Frequency
plan	plan - 19094 — proposal - 13111 — initiative - 3198 — propose - 2008 — agenda - 1479 — president plan - 573 — president proposal - 422 — legislative proposal - 312 — be propose - 247 — blueprint - 213	37723
individual	individual - 33657 — account - 3647 — individual family - 643 — many individual - 463 — personal information - 283 — personal responsibility - 235 — individual business - 207 — every individual - 207 — million individual - 179 — family individual - 172	36330
community	community - 13902 — organization - 8763 — local community - 2305 — international community - 2100 — youth - 1386 — membership - 564 — behalf - 487 — many community - 487 — nonprofit organization - 462 — participation - 405	34277
worker	worker - 16944 — million worker - 554 — many worker - 399 — foreign worker - 277 — social worker - 213 — worker right - 133 — every worker - 127 — worker country - 126 — worker family - 126 — percent worker - 123	33601
tax	tax - 15994 — tax cut - 5609 — spending - 4017 — cut - 3774 — tax increase - 1480 — raise - 648 — raise tax - 399 — spending cut - 318 — billion tax cut - 229 — percent cut - 221	32317
company	company - 34769 — consortium - 288 — subsidiary - 222 — affiliate - 196 — ag - 178 — venture - 147 — ge - 129 — foreign subsidiary - 63 — startup - 63 — conglomerate - 62	32205
school	school - 16925 — college - 2924 — local school - 1676 — high school - 1305 — public school - 1178 — college student - 519 — academy - 460 — school system - 417 — many school - 403 — community college - 335	27607
student	student - 18503 — teacher - 6398 — many student - 542 — student family - 407 — million student - 393 — every student - 268 — percent student - 190 — student teacher - 181 — parent teacher - 142 — qualify teacher - 131	26535
fund	fund - 17564 — investment - 3647 — investor - 2817 — foreign investor - 293 — invest - 228 — fund provide - 208 — hedge fund - 192 — foreign investment - 182 — private investment - 175 — mutual fund - 158	26216

**Note:** This table presents the 20 most frequent clustered entities identified in the U.S. Congress. The column “Label” refers to the final label assigned to the entity after manual inspection. The column “Most frequent phrases” lists the ten most frequent phrases which were assigned to these clustered entities, as well as their frequency of occurrence (i.e., phrase - frequency). The column “Frequency” is the total number of mentions of this clustered entity in the corpus.

Table C.3: Most Frequent Named Entities (i)

Label	Most frequent phrases	Frequency
american	american - 49293 — american people - 35168 — million american - 8209 — american public - 3212 — american taxpayer - 3187 — many american - 3118 — american family - 3074 — every american - 3031 — american worker - 2682 — american citizen - 1956	149225
program	program - 44125 — pilot program - 1125 — federal program - 958 — important program - 617 — government program - 580 — medicaid program - 448 — cop program - 403 — education program - 314 — many program - 293 — entitlement program - 280	104383
government	government - 62923 — local government - 5866 — foreign government - 1085 — government official - 553 — government shutdown - 437 — russian government - 383 — tribal government - 357 — government spending - 344 — big government - 303 — branch government - 285	90865
service	service - 14224 — arm service - 3166 — forest service - 2589 — postal service - 2282 — internal revenue service - 896 — wildlife service - 862 — good service - 812 — custom service - 690 — public service - 637 — military service - 607	61409
nation	nation - 34474 — great nation - 918 — nation world - 445 — people nation - 351 — many nation - 280 — entire nation - 277 — nation economy - 272 — rogue nation - 239 — nation child - 231 — foreign nation - 226	57571
republican	republican - 51157 — republican leadership - 6332 — republican majority - 3678 — republican friend - 1246 — republican plan - 1169 — republican side - 614 — republican president - 486 — many republican - 478 — republican proposal - 471 — republican control - 410	56582
administration	administration - 56284 — president administration - 606 — administration official - 484 — previous administration - 476 — administration policy - 268 — administration proposal - 217 — administration budget - 200 — republican administration - 149 — administration president - 125 — last administration - 121	53429
agency	agency - 23605 — federal agency - 6048 — local educational agency - 1605 — government agency - 942 — educational agency - 859 — law enforcement agency - 723 — head agency - 592 — intelligence agency - 530 — public housing agency - 514 — department agency - 389	47452
budget	budget - 28857 — republican budget - 2091 — president budget - 1780 — budget deficit - 663 — democratic budget - 446 — budget plan - 427 — budget agreement - 392 — budget proposal - 345 — budget process - 319 — budget surplus - 298	46434
america	america - 24636 — people america - 971 — rural america - 481 — corporate america - 306 — america senior - 258 — job america - 246 — america family - 209 — america child - 202 — bless america - 197 — everybody america - 191	46053



Table C.4: Most Frequent Named Entities (ii)

Label	Most frequent phrases	Frequency
national	national security - 2623 — national - 1426 — national guard - 1159 — national debt - 1153 — national academy science - 672 — national interest - 563 — national labor relation - 471 — national intelligence director - 422 — director national intelligence - 407 — national science foundation - 358	39427
energy	energy - 5592 — department energy - 2259 — energy policy - 1119 — energy commerce - 652 — high energy - 327 — energy company - 312 — renewable energy - 305 — energy efficiency - 294 — energy independence - 258 — energy crisis - 246	28045
democrat	democrat - 25073 — democrat republican - 2377 — republican democrat - 2183 — democrat majority - 459 — democrat leadership - 418 — president democrat - 406 — republican democrat alike - 297 — many democrat - 280 — democrat republican alike - 260 — liberal democrat - 220	28012
group	group - 8480 — bipartisan group - 464 — special interest group - 458 — work group - 413 — terrorist group - 389 — group people - 256 — environmental group - 252 — interest group - 250 — group health plan - 246 — outside group - 231	25784
project	project - 9652 — demonstration project - 610 — pilot project - 310 — construction project - 241 — program project - 104 — important project - 100 — research project - 100 — military construction project - 98 — project authorize - 93 — water project - 92	23013
war	war - 6299 — war iraq - 831 — world war ii - 819 — war terrorism - 482 — civil war - 476 — cold war - 408 — war terror - 338 — war power - 217 — war drug - 201 — world war - 196	22704
small business	small business - 14213 — small business owner - 1592 — small business administration - 538 — small business concern - 517 — many small business - 439 — small business people - 183 — percent small business - 156 — small business community - 131 — individual small business - 128 — family small business - 124	22379
medicare	medicare - 10868 — medicare beneficiary - 1220 — medicare patient - 808 — medicare system - 309 — medicare trust fund - 215 — medicare benefit - 199 — medicare cut - 150 — medicare recipient - 149 — cut medicare - 138 — traditional medicare - 127	22348
veteran	veteran - 11324 — veteran affair - 1688 — many veteran - 412 — disabled veteran - 377 — nation veteran - 328 — veteran family - 189 — homeless veteran - 188 — million veteran - 181 — veteran benefit - 156 — america veteran - 156	20792
authority	authority - 7485 — local authority - 527 — palestinian authority - 508 — housing authority - 310 — administer authority - 299 — federal authority - 256 — public housing authority - 236 — budget authority - 212 — trade promotion authority - 170 — exist authority - 158	19468

**Note:** This table presents the 20 most frequent named entities identified in the U.S. Congress. The column “Label” refers to the final label assigned to the entity after manual inspection. The column “Most frequent phrases” lists the ten most frequent phrases which were assigned to these named entities, as well as their frequency of occurrence (i.e., phrase - frequency). The column “Frequency” is the total number of mentions of this named entity in the corpus.

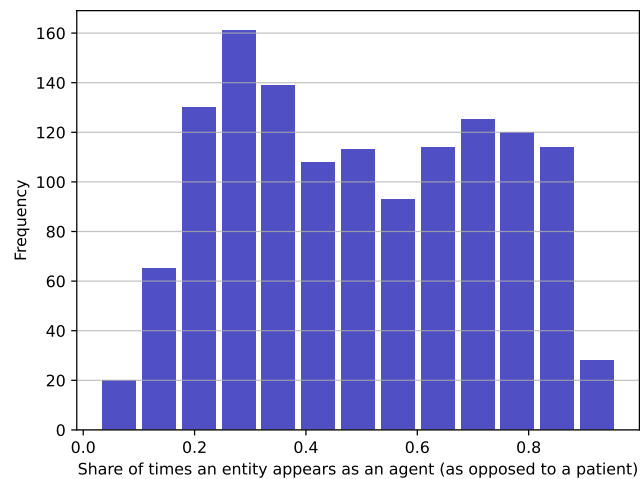
## D. Summary Statistics

Table D.1: Summary Statistics on the Congressional Record Corpus

Variable	All Narratives	Complete Narratives	Frequent Narratives	Relevant Narratives
Speeches	1190969	550561	187379	105514
Sentences	17312358	4196393	279044	160827
Statements	76395962	5774295	346685	180964
Narratives, unique	5963192	3570119	2838	1638
Agents raw	76395962	5774288	346685	180964
Agents raw, unique	1604754	945412	21041	15035
Agents clustered, unique	1884	1879	415	245
Patients raw	76395962	5774293	346685	180964
Patients raw, unique	8322792	1794942	33774	21411
Patients clustered, unique	1884	1877	499	353
Verbs raw	76395962	5774295	346685	180964
Verbs raw, unique	44889	18642	391	276
Verbs cleaned, unique	34780	11807	183	140

**Note:** This table shows descriptive statistics on the U.S. Congressional Record at different steps of the pipeline. “All narratives” are those with at least an agent or a patient. “Complete narratives” are those with both an agent and a patient. Among those, “frequent narratives” were pronounced at least 50 times. Finally, we focus on politically or economically relevant entities (“relevant narratives”) – as opposed to procedural/uninterpretable entities.

Figure D.1: Share of Times An Entity appears as an Agent (as opposed to a Patient)



**Note:** This figure presents, for every entity, the share of times it appears as an agent (as opposed to a patient). For example, if an entity occurred as an agent 8 times and as a patient 2 times, the corresponding share would be 0.8. Most entities appear both as an agent and a patient. However, entities also tend to be “specialized”: Most of them appear as either an agent or a patient more frequently.

## E. Examples of Narratives and Sentences

This appendix shows examples of random sentences and their associated narratives as well as the most frequent narratives and random examples of sentences they appear in. In both cases, we focus on narratives that contain (at least) an agent, a verb, and a patient (i.e., “complete” narratives).

### 1. Sampled sentences and their associated narratives

The following list shows a random sample of sentences, along with the complete narrative(s) we mine for each sentence.

*As the ranking member of the Transportation and Infrastructure Subcommittee during the 111th Congress, he helped us ensure ongoing investment in Federal highway, transit and highway safety programs and the jobs that those investments create.*

**fund create job**

.....

*Many of our Senators during campaigns promise to vote for a constitutional amendment to balance the budget, in the heat of the campaign.*

**constitutional (thing) balance budget**

.....

*Companies doing business online want to reassure consumers and potential customers that their interests will be protected online, not heighten their concern about electronic commerce.*

**company do business**

.....

*The Progressive Caucus Budget raises taxes by almost \$950 billion.*

**budget raise tax**

.....

*Many are asking what happened to the promise Republicans made in 2010 when, in their pledge to America, they wrote—and again I quote—in their pledge to America:*

**republican make promise**

.....

*Faced with soaring gasoline prices, agencies around the country that provide services to the elderly say they are having to cut back on programs such as Meals on Wheels, transportation assistance, and home care, especially in rural areas that depend on volunteers to provide their own gas.*

**agency provide service**

.....

*And we are doing so at a time when our budget, the budget put through by the Republicans who control the Congress, is slashing funds for education, cancer research, health care, other essential needs at home.*

**republican put budget**

.....

*Subject to the direction of the Secretary, a person charged with law enforcement responsibilities by the Secretary who is performing a duty related to enforcement of a law regarding fisheries or other marine resources may make an arrest without a warrant for an offense against the United States committed in his presence, or for a felony cognizable under the laws of the United States, if he has reasonable grounds to believe that the person to be arrested has committed or is committing a felony.*

**person commit (legal) charge**

.....

*We desperately need the largest volunteer effort in our history to mobilize our citizens to act as role models and mentors for young people who have no real family encouragement.*

**people have appreciation**

.....

*What I don't understand is why our Democratic friends want to ruin a good thing that 80 percent of seniors who have access to this prescription drug plan say they like and 90 percent of seniors eligible have signed up for, saving on average \$1,200 a year.*

**senior have prescription drug**

.....

*President, I ask unanimous consent that the Committee on the Judiciary be authorized to meet to conduct a markup on Tuesday, October 5, 2004, at 10 a.m., in SD-226.*

**judiciary conduct markup**

.....

*Recognizing the public health threat caused by antibiotic resistance, Congress took several steps to curb antibiotic overuse in human medicine through amendments to the Public Health Service Act made by section 102 of the Public Health Threats and Emergencies Act , but has not yet addressed antibiotic overuse in agriculture.*

**healthcare make healthcare**

.....

*As the crisis approached, Democrats did nothing.*

**democrat do nothing**

.....

*I join with my friend in asking everyone to support this measure to name this post office after Sergeant Shawn T. Hannon and Master Sergeant Jeffrey J. Rieck, and I align myself with all of the comments that my friend and colleague has made on behalf of these two gentlemen, these two heroes.*

**friend make comment**

.....

*History has shown that, without strict limits on spending, tax increases will simply go to new spending rather than to reduce the deficit.*

**tax reduce deficit**

.....

*Provided further, That funds available under this heading for opportunity scholarships, including from prior-year appropriations acts, may be made available for scholarships to students who received scholarships in the 2009-2010 school year:*

**student receive degree (education)**

.....

*Well, it is important because people who have made America, who have made America strong, ought to in fact be remembered.*

**people make america**

.....  
*President, earlier this week I placed on my Web site– sanders.senate.gov—a letter to Secretary Paulson, and I asked people who shared the sentiments of that letter to sign a petition.*

**people share concern**

**people sign petition**

.....  
*It seeks to solve the problem of soft money , not just with the political parties, but with the special interest groups who run attack ads, who are completely unregulated by the system, and whose contributors are undisclosed.*

**group run ad**

.....  
*I raise these points because when someone stands up and says, “We are the ones who want the taxpayers to keep their money and you on this side of the aisle, you are the ones who want to take it from them.”*

**money keep money**

## **2. Most frequent narratives and examples of sentences they appear in**

The following list shows the most frequent narratives in our corpus (ordered by frequency), along with a random sample of sentences in which they appear.

**people lose job**

*they knew people would lose their jobs.*

*People will lose their jobs, people will lose their savings, people will find that they cannot get the credit necessary to keep their businesses open or to function at a reasonable level.*

*But last week the Associated Press reported that an even greater number of people, 108,000, lost their jobs, as U.S. companies dealt with the battered economy right here at home.*

*This bill allows for clear actions, up-or-down votes on the conference report to prevent Wall Street from melting down like it did 2 years ago and a bill to provide unemployment compensation to people who have lost their jobs who cannot find work in this economy.*

*On the one hand we have people who lose their jobs after working years and years and are temporarily in need of assistance and should have that assistance.*

.....

### **citizen abide law**

*By forcing the IRS, and not the taxpayer, to carry the burden of proof in disputes, we protect legal, law-abiding citizens and end harassing and frivolous claims by maverick agents.*

*A bill to increase and enhance law enforcement resources committed to investigation and prosecution of violent gangs, to deter and punish violent gang crime, to protect law-abiding citizens and communities from violent criminals, to revise and enhance criminal penalties for violent crimes, to expand and improve gang prevention programs, and for other purposes; to the Committee on the Judiciary.*

*Senators from every State of the Union, from every political persuasion, agreed to a version of the PATRIOT Act that would reauthorize the provisions that were set to expire and which provided the Government with the tools to aggressively pursue the war on terror, while protecting the rights of law-abiding citizens.*

*This is a blatant attack on our second amendment rights and on our law-abiding citizens.*

*Furthermore, it is my belief that this basic right should include law-abiding citizens.*

.....

### **american lose job**

*And now this bill that was cooked up in secret between the White House and Congressional Republicans without any input from Congressional Democrats will bring our country further into debt, lead to more hard- working Americans losing their jobs, and put a greater share of the tax receipts in the pockets of the Nation's most privileged.*

*But they are not cheaper because they pay 33 cents an hour; it is just that Americans lost their jobs.*

*How can we stand by as millions of American workers lose their jobs, as thousands of American companies cannot compete fairly, as our country as a whole has wealth drained from it?*

*For all the time that we do not pass it, each month 500,000 Americans will lose their jobs.*

*All these benefits, permanent tax benefits for middle-class families making one-quarter of a million dollars or less, the opportunity to create jobs, the opportunity to take care of a couple million Americans who lost their jobs, the opportunity to bring the private sector back again, the opportunity to give the private sector certainty, none of that is good enough for them.*

.....  
**government run healthcare**

*Secretary Mike Leavitt, who is the former Secretary of the Department of Health and Human Services, said: Advocates for a public health-care plan continue to look for a way to give political cover to moderates while advancing their goal of implementing a government-run health-care system.*

*This is a debate about getting us on that path toward government-run health care.*

*We know that some here in Washington have wanted government-run health care for many years.*

*Now they have a government-run health care full of tax hikes and mandates on small businesses, which the NFIB estimates will cost 1.6 million more jobs lost.*

*We didn't see it materializing into a government-run health care program, as many have alleged here today.*

.....  
**american have healthcare**

*As a nation we have 50 million Americans today who have no health insurance and that number has increased.*

*Speaker, as millions go without health insurance and costs skyrocket, I strongly believe we need commonsense reforms so that all American families can have access to affordable health coverage.*

*The bill would create affordability credits to ensure that all Americans have more affordable health care coverage.*

*What we need to do is subject any health proposal, Democrat or Republican, to one simple test: Does it guarantee that all Americans will have health care coverage?*

*It is time for us to be a part of that representative democracy and forge a bill together that assures all Americans have access to health care reform.*

.....  
**god bless america**

*The President said that we ought to pray that God bless America.*

*When in Namibia, where we were, I suggest you play the Namibian national anthem and leave*



*it to the local mayor to say, as he said to us “God bless America.”*

*Mr. Speaker, God bless all listening, and God bless the United States of America.*

*Mr. Speaker, with that, before I yield back the balance of my time, I will ask God to please bless our men and women in uniform, both in Afghanistan and in Iraq and throughout the world, and I will ask God to please bless America.*

*I pray that God will always continue to bless America.*

.....

**people need help**

*According to the executive director of the Community Action Agency in Kentucky, Kip Bowmar: February of 2008 marked the first time in the program’s history that all 120 Counties in Kentucky ran out of LIHEAP funds, forcing us to close our doors as fuel prices were soaring and people needed help.*

*Look, there are lots of different things, and I will not go through the litany in terms of unemployment provisions and in terms of helping small businesses, because the thing I would like to do is just say thank the gentleman to the gentleman on behalf of New York, on behalf of the liberty zone, on behalf of all those people who need your help.*

*Why then should we be concerned about the increase in numbers of people who are condemned to a career as a temporary worker without benefits or minimum wage workers, people who work every day and still need public help?*

*Those young people need more than help from their families in order to be able to go to college.*

*There are people who need help who don’t have help, and we need to find a way to do that, but it is not this 2,000-page bill.*

.....

**god bless troop**

*In conclusion, God bless our troops, and we will never forget September the 11th and the global war on terrorism.*

*In conclusion, God bless our troops, and we will never forget September 11*

*In conclusion, God bless the memory of Congressman Sonny Montgomery, God bless our troops, and we will never forget September 11.*

*In conclusion, God bless our troops, and we will never forget September the 11th in the global war of terrorism.*

*In conclusion, God bless our troops—and we will never forget September the 11th and the global war on terrorism.*

.....

**worker lose job**

*Workers are losing their jobs, their homes, and their hopes.*

*This would create a system that is more focused on punishing employers than truly helping workers who lose their jobs.*

*There are truckers who are losing their jobs.*

*the impact of not doing something here is huge, on workers losing their jobs.*

*Speaker, Sara Lee is closing their Virginia apparel factory; 42 workers lose their jobs.*

.....

**small business create job**

*The No. 1 issue facing our country is jobs, and the No. 1 goal of Republicans in 2012 is to continue to make it easier for American small business to create jobs.*

*Two out of every three new jobs in the United States is created by a small business, but the shutdown is hitting them hard.*

*Small business owners will be able to spend less time searching for affordable health plans, and more time creating jobs.*

*President, our Nation’s small businesses have created 64 percent of all new jobs in the last 15 years, yet in the last year nearly 85 percent of the jobs lost have come from small businesses.*

*That is why the President said let us focus on doing things that will help these families and equally, if not more importantly, help small businesses create jobs.*

.....

**service connect disability**

*A bill to increase, effective as of December 1, 1996, the rates of disability compensation for veterans with service-connected disabilities and the rates of dependency and indemnity compensation for survivors of such veterans, and for other purposes; to the Committee on Veterans Affairs.*

*I thank the Senator from Maine, Ms. Snowe, the chairman of the Small Business Committee, for making the important point that Federal law requires that a certain percentage of contracts*

*be awarded to firms owned by veterans with service-connected disabilities.*

*As we reported in VA Health Care: Issues Affecting Eligibility Reform , VA uses a complex priority system—based on such factors as the presence and extent of any service-connected disability, the incomes of veterans with nonservice-connected disabilities, and the type and purpose of care needed—to determine which eligible veterans receive care within available resources.*

*Under the law that existed and controlled at that time, only those who had been granted a claim for a service-connected disability or demonstrated financial need could turn to the VA for health care services.*

*Most military retirees who have a service-connected disability are not permitted to collect both their retirement and disability benefits concurrently.*

.....

**constitutional (thing) balance budget**

*I have always wanted to be on the side of those supporting a constitutional amendment to balance the budget.*

*Let me say that I am someone who will support a constitutional amendment to balance the budget.*

*Let me, to the extent that I can, say to the seniors who are listening to those who would like to make you believe that they are really here arguing to save Social Security, suggest to you that what they are really arguing about is that they don't want a constitutional amendment to balance the budget and they have now hit on what I perceive to be a risky gimmick in an effort to frighten seniors and by that approach defeat a constitutional balanced budget amendment.*

*They say if we pass this constitutional amendment to balance the budget, the budget will be balanced.*

*One of the reasons the polls show so many people want a constitutional amendment to balance the budget is two things.*

.....

**people have healthcare**

*But more importantly than that, we have got people who have health care emergencies because of the earthquake, business emergencies, lives literally at stake.*

*The same on health insurance: If you are trying to increase the number of people who have health insurance, you have to do more than what is in this tax package.*

*Some writers have called that group the “contentedly covered,” the people who already have health care coverage in America today.*

*We generally think of most people having private health insurance coverage.*

*People are going to have health insurance they do not now have.*

.....

**public held debt**

*446 511 available to repay debt held by the public.*

*It has brought with it, as Chairman Greenspan noted, a significant increase in revenues to the Federal Government, which are allowing us to pay down even more rapidly than previously thought the debt now held by the public.*

*Change in debt held by the public.*

*The best way to gauge the level of debt held by the public is to compute that debt as a percentage of our national income, our Gross Domestic Product.*

*202 Resolved, That upon the adoption of this resolution it shall be in order to consider in the House the bill to require that the Government prioritize all obligations on the debt held by the public in the event that the debt limit is reached.*

.....

**people pay tax**

*Real jobs in the U.S. economy for people who will be paying taxes, not being paid unemployment compensation for not working.*

*It is the people who are going to pay these taxes.*

*That is why the increase in seniors is going up so dramatically, and the increase in the people working and paying their taxes is going up modestly.*

*We are talking about giving the money back to the people who earn it, and the first place we ought to look is to people who are married who pay more taxes just because they are married.*

*This entire debate is about whether or not people who pay no taxes can get a tax cut.*

.....

**people lose family**

*But I think it is also so important for us to look back, figure out what went wrong, and move*

*forward and fix the problems; not to create a political blame game, not to blame the Clinton administration or the Bush administration, but to fix what went wrong yesterday and make it right for tomorrow, so that we do not have 2,800 people lose their lives again.*

*Now is the time to speak up, not after the people have lost their lives, not after the American people—the American people know what is going to happen here.*

*We have had a number of people—maybe a dozen or so—who have lost their lives prior to this month in Iraq and Afghanistan.*

*Let's recognize, whether they were in the World Trade Center, at the Pentagon, or on airplanes that were commandeered, people were in dangerous situations, and many lost their lives.*

*Yet the U.S. Government has chosen to discriminate against these people who lost their lives at Fort Hood 4 years ago at the hands of a terrorist, who tragically happened to be a member of the uniformed military of the United States, MAJ Nidal Hasan, who has subsequently been convicted of these crimes.*

.....

### **american pay tax**

*Unless there is some extraneous circumstance, I think most Americans voluntarily pay their taxes.*

*I think that Members of Congress, members of the administration need to take their heads out of the sand, need to start dealing with really some of the very tough issues of Medicare, of Social Security, of annual government overspending, and I would just ask an American that pays taxes to spend a few moments thinking about the absurdity of our tax code in this country.*

*However, it is not small to Dave and Joann and to the millions of young Americans who pay more in taxes simply because they have formed the basic unit of society—a family.*

*Why should middle-class Americans, who work hard, pay their taxes, and send their kids to school, not have that kind of security, too?*

*This bill takes away that freedom, requiring every American to purchase a government-approved health plan, pay a tax, or even go to jail.*

.....

### **employee provide healthcare**

*The vast majority of employers who provide health care merely pay for the benefit.*

*We urge the Senate to support your amendment to ensure that employers will not be sued for*

*voluntarily providing health coverage to 172 million workers.*

*As chairman of the Subcommittee on Employer-Employee Relations with jurisdiction over employer-provided health care, the issue of mental health parity falls within the purview of this subcommittee.*

*According to a Gallup poll conducted last September, the vast majority of Americans, 70 percent, are satisfied with their health insurance provided by their employer.*

*I don't think it's right that you mandate an employer to provide health care insurance or*

.....

**people do job**

*We have people doing this work for free in California, not costing any Federal taxpayer dollars, and the amendment that I offered with the gentleman from California that was passed by, I believe, 380 to 0, allows our people to have continued access.*

*If you are giving the money to the people doing the work, they will work wonderfully, and things will happen beautifully.*

*And now you've got people who once had three people doing a job, now there's one.*

*My amendment to the Department of Defense authorization bill is meant to strengthen our efforts to verify if people in the United States are here legally to do their work.*

*The people in Mexico who have the jobs that the people in Ohio used to do do not make enough money to buy the very products they make.*

.....

**men women serve country**

*So I have had the privilege of coming from a State that has sent to Washington, DC, men of great honor, great integrity, men and women who have served the country so well.*

*We are talking about young men and women in our Armed Forces returning home after serving their country in Iraq.*

*Throughout my years of service to Minnesotans in the Senate, Frank provided wise counsel—especially on behalf of the men and women who, like himself, had served their country in time of war.*

*' To the smart young man who wrote me that letter and to all of America's veterans, this bill builds on efforts to meet our country's moral obligations to the men and women who so bravely served our country.*

*I encourage my colleagues to honor the men and women who serve this country and pass this important legislation.*

.....

## F. Comparison to Alternative Text Features

To identify latent topics in political discourse, we run Latent Dirichlet Allocation (LDA) on the U.S. Congressional Record. We then compare our method to topics and n-grams by presenting the ten most representative words as well as the ten most representative narratives associated to each topic share in Tables F.1, F.2, and F.3. We restrict ourselves to narratives that were at least pronounced 50 times in the corpus.

Table F.1: Topics, Words and Narratives (i)

Topic	Top Words	Top Narratives
banking	bank, financi, compani, credit, consum, drug, card, institut, account, market	bank issue card, company hold business, money fund outlay, measure provide authority, company hold loan, asset back military, american want reform, mortgage back military, action have effect, consumer make instruct
communications	page, line, internet, servic, amount, communic, comput, increas, access, broadcast	program serve child, provider offer service, tax benefit american, reconciliation carry recommendation, budget receive recommendation, facility provide service, company provide service, network provide service, provider provide service, citizen have healthcare
crime	crime, law, enforc, victim, crimin, violenc, justic, state, prison, act	law enforcement add category, crime add category, injustice motivate crime, juvenile commit crime, violent crime serve (legal) sentence, program meet regulation , juvenile commit violent crime, offender serve (legal) sentence, child witness violence, offender commit crime
culture	art, smith, museum, delawar, idaho, cultur, music, endow, recognit, artist	american have service, university conduct study, american do job, parent care child, someone seek recognition, program help community, federally fund project, program have (positive/negative) effect, applicant receive grant, people leave country
economy	busi, small, compani, capit, invest, contract, administr, job, creat, econom	person control small business, eligible participate program, program provide fund, small business employ workforce, agency give priority, agency require information , veteran control small business, service control small business, individual receive funding, founder start business
education	educ, school, student, teacher, program, colleg, year, high, state, higher	student use technology, student leave discipline, school reduce class size, school hire student, agency serve school, school use fund, student receive degree (education), agency receives grant, agency require information , student have education
employment and labor	worker, employ, employe, work, labor, job, wage, union, pay, hour	employee pay overtime, employee hire permanent, person earn dollar, men women earn dollar, worker join union, worker form union, person make dollar, market take healthcare, person make workforce, commonwealth perform duty
environment	environment, wast, water, regul, epa, facil, state, clean, protect, administr	individual take opportunity, child have disability, toxic (thing) cause cancer, national preserve social security, program need reform, environmental protection agency do job, supreme declare unconstitutional, republican want repeal (something), money justify money, republican want medicare
family	children, famili, child, parent, welfar, care, young, work, support, kid	girl have baby (boomer), action have effect, family adopt child, good start provide service, parent care child, good start serve child, family leave welfare, child witness violence, child have child, community participate program
farming	food, product, farm, agricultur, farmer, tobacco, produc, price, fda, crop	kid graduate school, F.D.A. regulate tobacco, tobacco cause cancer, child start smoking, food drug administration regulate tobacco, agency require information , nothing limit authority, law provide remedy, F.D.A. have authority, uniform serve country



Table F.2: Topics, Words and Narratives (ii)

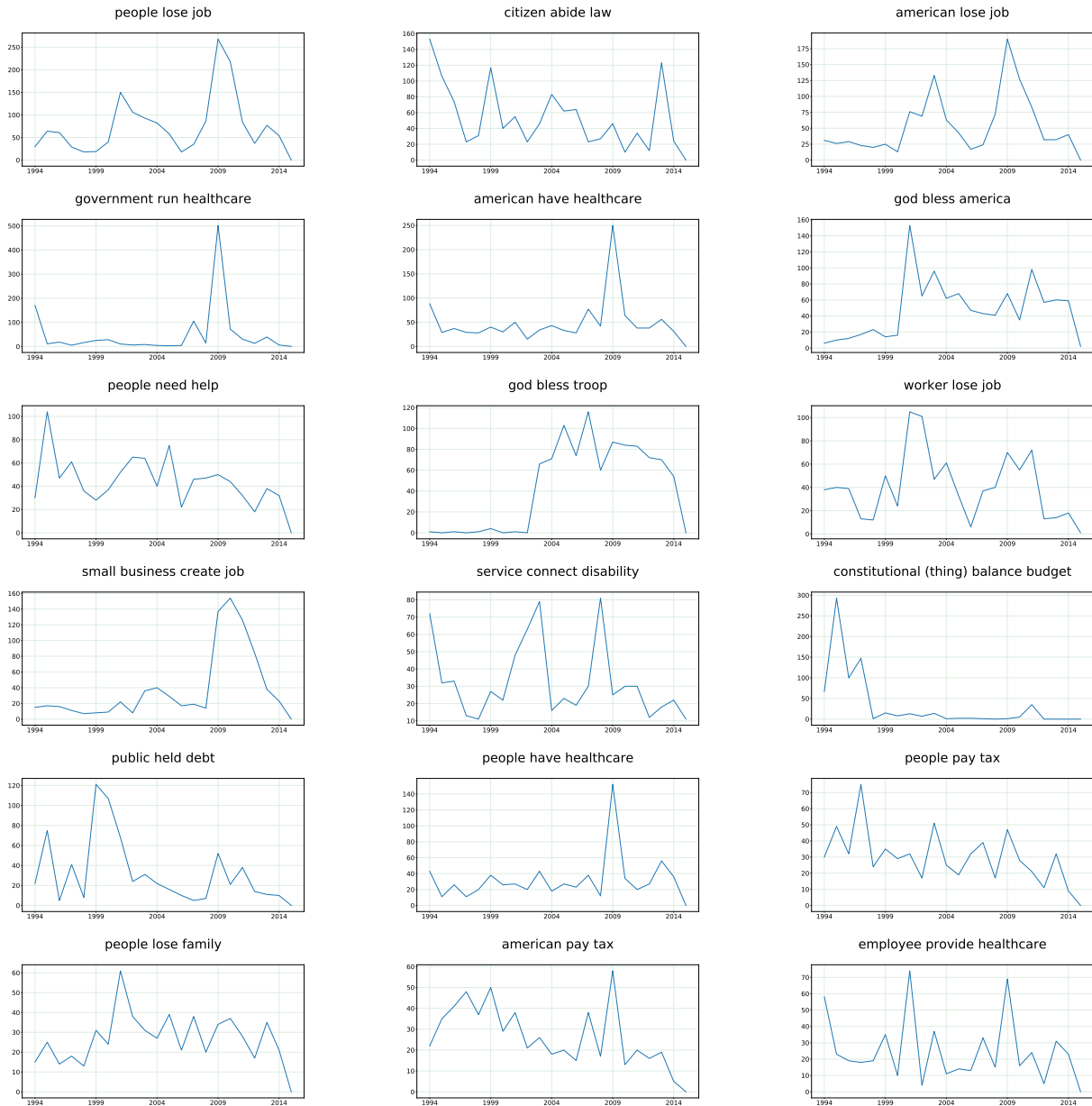
Topic	Top Words	Top Narratives
fiscal policy and taxation	tax, percent, american, incom, famili, pay, job, increas, year, would	measure breach budget, measure would budget, measure provide authority, program require appropriation, budget receive recommendation, reconciliation carry recommendation, measure not-increase deficit, public held debt, fact impose tax, budget add trillion
foreign policy	state, unit, nuclear, treati, weapon, nation, presid, peac, intern, iran	wealthy pay fair, iran obtain nuclear (weapon), iran acquire nuclear (weapon), iran develop nuclear (weapon), iran have nuclear (weapon), iran pose threat, nuclear (weapon) pose threat, bank do business, constitution prohibit america, people help people
gun violence	china, gun, chines, law, weapon, firearm, taiwan, kill, year, state	weapon pierce tank, national provide service, institute conduct study, customer sell gun, veteran need help, soldier pierce tank, gun abide law, crime get gun, witness give witness, american scratch staff
healthcare	hospit, rural, nurs, johnson, home, physician, mitchel, area, facil, lungren	person perform abortion, pregnancy feel pain, healthcare deny coverage, person have corrective action, physician deliver baby (boomer), insurance drop coverage, insurance deny coverage, physician perform abortion, doctor acquire disease, patient get child
housing	loan, hous, home, mortgag, bankruptci, rate, debt, famili, credit, interest	american make fund, homeowner face mortgage, soldier serve country, family face mortgage, law provide protection, federally assist housing, trade create job, people face mortgage, program provide housing, authority provide funding
human rights	unit, govern, state, wherea, right, peopl, human, freedom, intern, religi	program help community, message transmit concurrent, attorney (judge) represent client, people save family, american travel cuba, money amount taxable, burmese contain export, small business provide service, government make payment, business pay tax
immigration	secur, immigr, border, homeland, state, law, illeg, unit, alien, countri	bible fulfil self-employed, american hold market, country participate program, terrorist enter country, service administer program, employee hire illegal immigrant, people cross border, people enter country, employee hire immigrant, immigrant enter country
judiciary	fee, claim, patent, liabil, properti, damag, file, compani, lawsuit, case	party seek remedy, supreme have (legal) statute, association give grade, bush nominate circuit (courts), attorney (judge) follow law, attorney (judge) apply law, bush nominate attorney (judge), law confer power, agreement enters force, attorney (judge) interpret law
legislation	amend, motion, bill, ask, resolut, consent, unanim, read, question, presid	(sick/paid/unpaid) leave be columbia, (sick/paid/unpaid) leave be guam, (sick/paid/unpaid) leave be virgin, money amount taxable, message transmit concurrent, burmese contain export, corporation make charity, natural resource recommend substitute, service recommend substitute, individual make legislative (election)
military	veteran, servic, disabl, militari, care, member, benefit, mental, health, provid	air force request application, force base defense, service connect condition, veteran suffer illness, veteran earn money, veteran suffer diabetes, veteran receive compensation, defense submit service, service connect veteran, defense submit congressional
names	vote, smith, johnson, lewi, brown, taylor, davi, miller, weldon, peterson	application decrease appropriation, law authorize study, brave men women give family, government make payment, corporation make charity, veteran control small business, veteran administer law, nothing limit authority, person commit (legal) charge, congressional budget define money
national pride	american, nation, year, serv, great, honor, mani, today, life, one	people remember courage, god wipe shot, god grant strong, american hold market, word do justice, commonwealth perform duty, african make contribution, church offer prayer, hero give family, god bless family
native americans	indian, tribe, nativ, alaska, tribal, land, mine, trust, reserv, state	decision have effect, budget do job, someone get money, program carry program, employee perform service, healthcare save money, good start provide service, institute conduct study, someone know nothing, raw material meet road

Table F.3: Topics, Words and Narratives (iii)

Topic	Top Words	Top Narratives
natural resources / energy	energi, oil, gas, price, fuel, electr, product, use, produc, natur	someone buy gun, school use fund, ethanol blend gasoline, nation need energy, production shut supply and demand, production use american, oil account america, american rely oil, O.P.E.C. increase production, america need energy
nature	land, nation, area, park, forest, manag, conserv, resourc, state, acr	county mean county, city mean city, authority provide funding, worker earn minimum wage, grant receive grant, agency provide technical assistance, family leave welfare, someone buy gun, nothing affect authority, service do job
no label	presid, veto, congress, aid, reagan, item, bush, earmark, clinton, power	company bring job, story detail fight, energy affect family, people not-want job, american make facility, people send kid, people buy car, country lose job, small business hire people, people try job
political debate	speaker, time, gentleman, chairman, member, hous, would, vote, bill, colleagu	(sick/paid/unpaid) leave be guam, (sick/paid/unpaid) leave be columbia, (sick/paid/unpaid) leave be virgin, someone seek recognition, period offer appropriate, republican refuse action, people have commitment, democratic offer alternative, law authorize appointment, friend make comment
politics	senat, clerk, follow, pro, read, tempor, amend, presid, legis, rule	commonwealth perform duty, judiciary conduct promulgate regulation, air force indicate appointment, self-employed executes budget, commitment have witness, news sign text, party raise soft money, witness limit medicine, group run ad, witness give witness
public services	shall, secretari, fund, provid, program, act, section, state, appropri, avail	law authorize uniform, U.S.C. authorize service, application decrease appropriation, fund subsidize deficit, U.S.C. authorize uniform, fund make funding, congressional budget define money, fund carry funding, person perform abortion, government estimate contract
science	nation, develop, research, technolog, communiti, program, support, scienc, center, improv	commonwealth perform duty, student pursue degree (education), director award grant, agency participate program, (sick/paid/unpaid) leave be columbia, federally fund study, fund pay profit, company participate program, program award grant, program achieve goal
social welfare	drug, medicar, senior, prescript, benefit, cost, plan, program, beneficiari, year	oil account america, production shut supply and demand, medicare operate prescription drug, american rely oil, production use american, senior join H.M.O., senior have coverage, medicare have prescription drug, senior need prescription drug, senior pay premium
sports	univers, team, year, colleg, game, state, nation, wherea, award, first	people have exhaust, parent take child, fund provide service, small business employ workforce, individual perform service, defense prescribed regulation, people sign petition, budget cut education, american obtain healthcare, individual participate program
states	water, project, river, flood, corp, construct, lake, feder, cost, louisiana	water resource development authorize river, person commit offense, university conduct study, law authorize project, project provide water, F.D.A. approve medicine, program have effect, person take healthcare, worker earn minimum wage, agency carry program
trade	trade, u.s, countri, agreement, unit, state, foreign, export, market, american	administration negotiate trade, china join wto, export create job, trade create job, china manipulate dollar, american make life quality, agreement enters force, country have deficit, constitution give authority, bank provide loan
transportation	transport, highway, safeti, airport, system, state, passeng, vehicl, air, rail	grant fund program, amtrak provide service, good start provide service, project meet regulation, railroad provide service, railroad provide transportation, project receive funding, association give grade, carrier provide service, government receive funding
war	iraq, war, troop, terrorist, american, attack, presid, militari, terror, iraqi	iraqi take responsibility, iraqi take power, iraq pose jeopardy, iraq have weapon mass destruction, disability receive education, saddam hussein pose threat, iraq pose threat, saddam hussein use weapon mass destruction, saddam hussein use chemical (weapon), saddam hussein have weapon mass destruction

## G. Additional Results

Figure G.1: Most Frequent Narratives Over Time



**Note:** In this figure, we plot the occurrences of the 18 most frequent narratives in the U.S. Congressional Record over the period 1994 to 2015. We focus on narratives pronounced at least 50 times and remove entities labeled as procedural or noise.

Table G.1: Most Frequent Narratives Per Entity

family	money	someone	agency
people lose family	american keep money	someone have opportunity	agency provide service
american lose family	people keep money	someone pay attention	agency administer program
family lose love	people make money	someone have healthcare	agency furnish information
men women put family	people get money	someone do nothing	agency do job
american give family	government take money	someone have idea	agency receive grant
men women risk family	people have money	someone take healthcare	agency take action
people give family	family keep money	someone pay tax	agency provide funding
men women give family	people earn money	someone do job	agency submit application
god bless family	people put money	someone lose job	agency receive fund
healthcare save family	hard earn money	someone make money	agency participate program
program	american	child	service
program provide service	american lose job	child have healthcare	service connect disability
program provide funding	american have healthcare	child attend school	program provide service
program provide healthcare	american pay tax	child get education	healthcare provide service
agency administer program	american lose healthcare	child receive education	U.S.C. authorize service
program provide grant	american keep money	child have education	agency provide service
program make difference	american lose family	child receive healthcare	government provide service
program create job	american look job	child have opportunity	community provide service
student participate program	american abide law	person have child	service provide service
federally fund program	american not-have healthcare	child need help	company provide service
government run program	american want job	child not-have healthcare	provider provide service
person	budget	law	republican
person have abortion	constitutional (thing) balance budget	citizen abide law	republican take power
person serve country	budget balance budget	american abide law	republican join democrat
person have child	republican balance budget	law respect faith	republican do nothing
person make difference	budget raise tax	people violate law	republican balance budget
person do job	budget increase tax	bush sign law	republican shut government
person earn dollar	budget do nothing	gun abide law	democrat join republican
person make decision	plan balance budget	people break law	republican cut medicare
person have healthcare	budget cut program	law prescribed witness	republican put together
person provide service	government balance budget	veteran administer law	republican cut program
person make disaster	budget reduce deficit	law make appropriation	republican give tax break
employee	healthcare	business	people
employee provide healthcare	government run healthcare	business create job	people lose job
employee sponsor healthcare	american have healthcare	company do business	people need help
employee offer healthcare	people have healthcare	government sponsor business	people have healthcare
employee drop coverage	employee provide healthcare	company hold business	people pay tax
employee provide coverage	american lose healthcare	american do business	people lose family
employee take (paid/unpaid) leave	healthcare provide healthcare	corporation do business	people do job
employee sponsor plan	healthcare provide service	government do business	people take opportunity
employee sponsor coverage	child have healthcare	people do business	people make decision
employee provide service	people lose healthcare	business do business	people look job
employee hire worker	people get healthcare	business hire worker	people have concern
government	job	tax	country
government run healthcare	people lose job	people pay tax	men women serve country
government take healthcare	american lose job	american pay tax	person serve country
government provide service	worker lose job	budget raise tax	veteran serve country
government make decision	small business create job	capital gain tax	problem face country
government take money	people do job	democrat raise tax	people serve country
republican shut government	people look job	budget increase tax	challenge face country
government sponsor business	business create job	tax kill job	people enter country
government do nothing	american look job	family pay tax	country face challenge
government take property	people have job	someone pay tax	american serve country
government run program	people want job	people not-pay tax	country produce oil

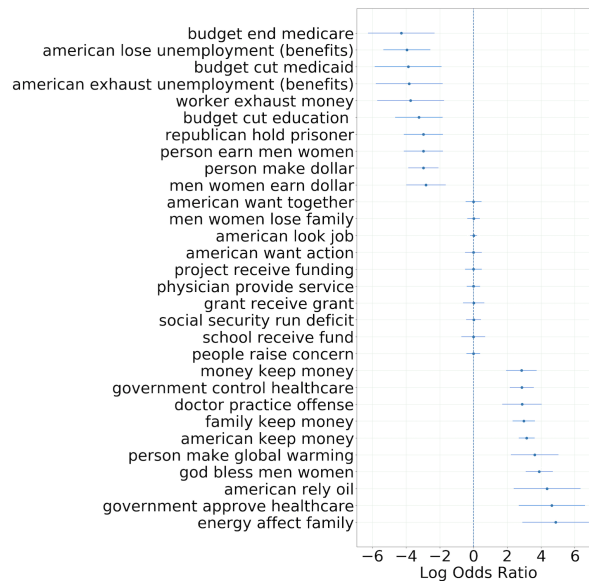
**Note:** In this table, we plot the ten most frequent narratives for the 20 most frequent entities in the U.S. Congressional Record (1994-2015).

## H. Additional Material on Divisive Entities

### Partisan and Neutral Narratives

Figure H.1 shows again the most Republican and Democrat narratives, now along with the most neutral narratives.

Figure H.1: Partisan Narratives in U.S. Congress



**Note:** This figure presents the most partisan narratives in the U.S. Congressional Record, as well as the least partisan narratives (i.e., with log-odds ratio closest to 0). For each narrative, we present the associated log-odds ratio and 95% confidence interval. A high log-odds ratio reflects narratives pronounced more by Republicans (relative to Democrats), and vice versa for a low log-odds ratio. One of the least partisan narratives, “men women lose family”, was pronounced most often in 2001 – suggesting that his narrative captures a reaction to the 09/11 terrorist attacks.

Table H.1: Longer, Unfiltered List of Divisive Entities

Rank	Entity	Score	Log-Odds Ratio		# Narratives		
			Narratives	Entity	Total	Dem	Repub
1	filibuster	1.93	2.10	0.16	2	1	1
2	<b>oil</b>	<b>1.12</b>	<b>1.39</b>	<b>0.27</b>	<b>8</b>	<b>4</b>	<b>4</b>
3	H.M.O.	0.96	1.05	0.10	2	1	1
4	remedy	0.91	0.94	0.02	2	1	1
5	contract	0.87	1.18	0.31	4	2	2
6	mortgage	0.79	1.03	0.24	5	3	2
7	<b>budget</b>	<b>0.76</b>	<b>0.82</b>	<b>0.06</b>	<b>42</b>	<b>17</b>	<b>25</b>
8	men women*	0.75	0.86	0.10	21	10	11
9	workforce	0.75	1.26	0.51	2	1	1
10	economy	0.74	0.78	0.04	4	2	2
11	pregnancy	0.73	1.07	0.34	3	2	1
12	medicaid	0.73	2.43	1.71	2	2	0
13	facility	0.71	0.92	0.21	3	2	1
14	policy	0.70	0.83	0.14	2	1	1
15	gun	0.69	0.85	0.16	6	4	2
16	<b>healthcare</b>	<b>0.68</b>	<b>0.74</b>	<b>0.06</b>	<b>116</b>	<b>76</b>	<b>40</b>
17	environmental protection agency	0.68	1.13	0.46	2	1	1
18	<b>interest/rate</b>	<b>0.65</b>	<b>0.97</b>	<b>0.32</b>	<b>8</b>	<b>5</b>	<b>3</b>
19	<b>loan</b>	<b>0.65</b>	<b>0.72</b>	<b>0.07</b>	<b>12</b>	<b>6</b>	<b>6</b>
20	center	0.64	0.79	0.15	2	1	1
21	democrat*	0.64	1.03	0.40	11	4	7
22	constitution*	0.62	0.63	0.01	8	5	3
23	<b>worker</b>	<b>0.61</b>	<b>1.23</b>	<b>0.62</b>	<b>16</b>	<b>11</b>	<b>5</b>
24	american*	0.60	0.61	0.01	158	86	72
25	plant	0.59	0.97	0.38	3	2	1
26	<b>insurance</b>	<b>0.58</b>	<b>0.67</b>	<b>0.09</b>	<b>16</b>	<b>7</b>	<b>9</b>
27	trade	0.57	0.74	0.17	2	1	1
28	action	0.57	0.61	0.05	14	6	8
29	resource	0.56	0.79	0.23	4	3	1
30	medicine	0.55	0.90	0.35	3	1	2
31	priority	0.54	0.93	0.39	3	2	1
32	community*	0.53	0.60	0.07	14	10	4
33	bank	0.52	0.88	0.35	6	4	2
34	abortion	0.52	0.92	0.39	10	2	8
35	republican*	0.52	1.12	0.60	40	31	9
36	<b>medicare</b>	<b>0.52</b>	<b>0.83</b>	<b>0.31</b>	<b>18</b>	<b>12</b>	<b>6</b>
37	<b>energy</b>	<b>0.52</b>	<b>1.30</b>	<b>0.78</b>	<b>7</b>	<b>3</b>	<b>4</b>
38	<b>job</b>	<b>0.51</b>	<b>0.65</b>	<b>0.14</b>	<b>117</b>	<b>73</b>	<b>44</b>
39	nation*	0.50	0.54	0.04	28	15	13
40	fund	0.49	0.49	0.00	19	8	11
41	plan	0.48	0.53	0.05	20	9	11
42	troop	0.48	0.64	0.16	3	1	2
43	missile	0.48	0.53	0.05	3	1	2
44	weapon mass destruction	0.47	0.64	0.16	3	2	1
45	doctor	0.47	0.79	0.31	16	6	10
46	company	0.46	0.87	0.41	17	11	6
47	people*	0.46	0.49	0.04	240	135	105
48	F.D.A.	0.45	0.68	0.23	8	4	4
49	family*	0.45	0.55	0.10	90	63	27
50	bad (problem)	0.45	0.46	0.01	2	1	1

**Note:** This table shows the full list of most divisive narratives, as measured by the score in column 3, given by average absolute log-odds ratio of the narratives where the entity appears (column 4), minus the absolute log odds ratio of the entity itself (column 5). The other columns give the number of unique narratives (total, Democrat-slanted, and Republican-slanted) where the entity appears. Bold type indicates a policy narrative listed in Table 3.2. Asterisks indicate an identity/symbolic narrative listed in Table 3.2.

Table H.2: List of Entities with Least Divisive Narratives

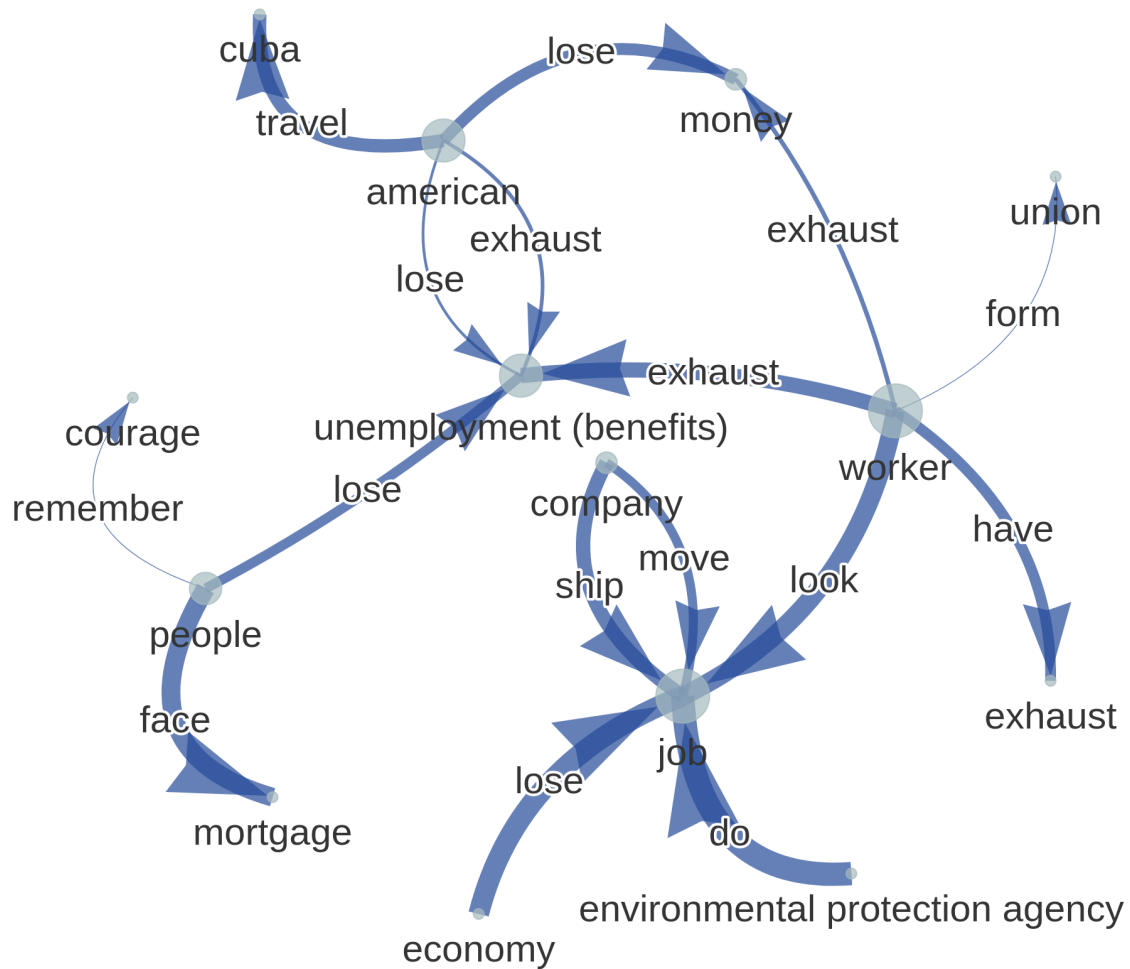
Rank	Entity	Score	Log-Odds Ratio		# Narratives		
			Narratives	Entity	Total	Dem	Repub
416	population	-0.04	0.06	0.10	1	0	1
417	client	-0.04	0.06	0.10	1	0	1
418	(positive/negative) effect	-0.04	0.06	0.10	1	0	1
419	strong	-0.04	0.34	0.38	1	0	1
420	union	-0.04	0.02	0.06	1	0	1
421	natural resource	-0.04	0.02	0.06	1	0	1
422	shot	-0.04	0.06	0.10	1	0	1
423	patient	-0.04	0.23	0.27	7	0	7
424	reform	-0.04	0.81	0.85	2	0	2
425	opinion	-0.05	0.10	0.14	3	0	3
426	truth	-0.05	0.18	0.22	2	0	2
427	buy	-0.05	0.07	0.11	2	0	2
428	county	-0.05	0.17	0.22	3	0	3
429	consequence	-0.06	0.75	0.81	2	0	2
430	railroad	-0.06	0.46	0.52	2	2	0
431	air force	-0.06	0.73	0.79	2	1	1
432	founder	-0.07	0.52	0.59	3	1	2
433	god	-0.08	1.04	1.12	9	1	8
434	border	-0.08	0.66	0.73	2	1	1
435	profit	-0.10	1.10	1.20	8	8	0
436	adult	-0.12	0.37	0.49	4	4	0
437	force	-0.15	0.27	0.42	2	0	2
438	congressional	-0.15	0.75	0.90	2	0	2
439	(sick/paid/unpaid) leave	-0.17	1.32	1.49	2	0	2
440	U.S.C.	-0.18	1.37	1.55	2	0	2
441	freedom	-0.19	0.58	0.77	4	0	4
442	legislative (election)	-0.19	0.58	0.77	2	0	2
443	word	-0.24	0.53	0.77	3	2	1
444	winners and losers	-0.53	0.81	1.33	2	0	2
445	appointment	-0.57	1.44	2.01	2	1	1

**Note:** This table shows the least divisive narratives, as measured by the score in column 3, given by average absolute log-odds ratio of the narratives where the entity appears (column 4), minus the absolute log odds ratio of the entity itself (column 5). The other columns give the number of unique narratives (total, Democrat-slanted, and Republican-slanted) where the entity appears.

## Distinctive Narrative Graphs for Democrats and Republicans

We present narrative graphs for mostly Republican (Democrat) narratives as measured by log-odds ratios.

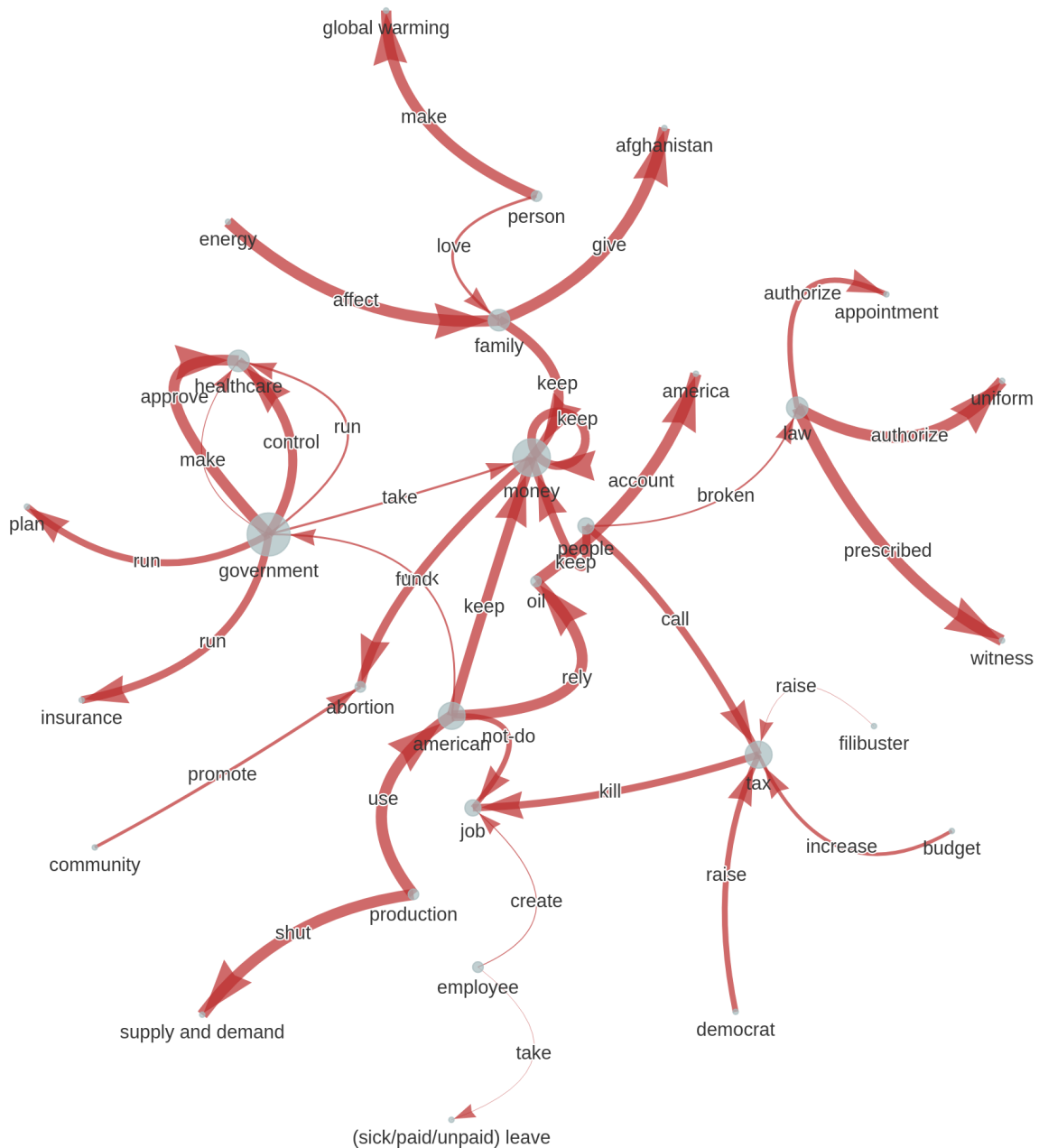
Figure H.2: Partisan Worldviews – Democrats



**Note:** This figure presents the 50 most salient Democrat narratives in the U.S. Congressional Record as measured by a log-odds ratio (i.e., whether a narrative is more pronounced by Democrats relative to Republicans). We represent our narrative tuples in a directed multigraph, in which the nodes are entities and the edges are verbs. The network is pruned so as to plot the largest connected subgraph (for this reason, the graph may display less than 50 narratives). The size of the edges is determined by the log odds ratio. The size of nodes is determined by their degree in the network.



Figure H.3: Partisan Worldviews – Republicans



**Note:** This figure presents the 50 most salient Republican narratives in the U.S. Congressional Record as measured by a log-odds ratio (i.e., whether a narrative is more pronounced by Republicans relative to Democrats). We represent our narrative tuples in a directed multigraph, in which the nodes are entities and the edges are verbs. The network is pruned so as to plot the largest connected subgraph (for this reason, the graph may display less than 50 narratives). The size of the edges is determined by the log odds ratio. The size of nodes is determined by their degree in the network.

## I. Which narratives co-occur together?

In many cases, agent-verb-patient tuples are building blocks of a broader political story. In Figure 3.1a, for example, we show that two prevalent narratives related to the war in Iraq are: “Saddam Hussein has weapons of mass destruction” and “Saddam Hussein poses a threat”. A more comprehensive story might read: “Saddam Hussein has weapons of mass destruction and poses a threat to national security. With a military intervention, the United States can make a difference, but at the cost of American lives.” Similarly, “people lose job” is the most common narrative in the corpus, but to which political stories does it belong to? Can we identify narratives that co-occur together to eventually reconstruct these broader political stories? As a first step in this direction, we build on two common approaches in the text-as-data literature below.

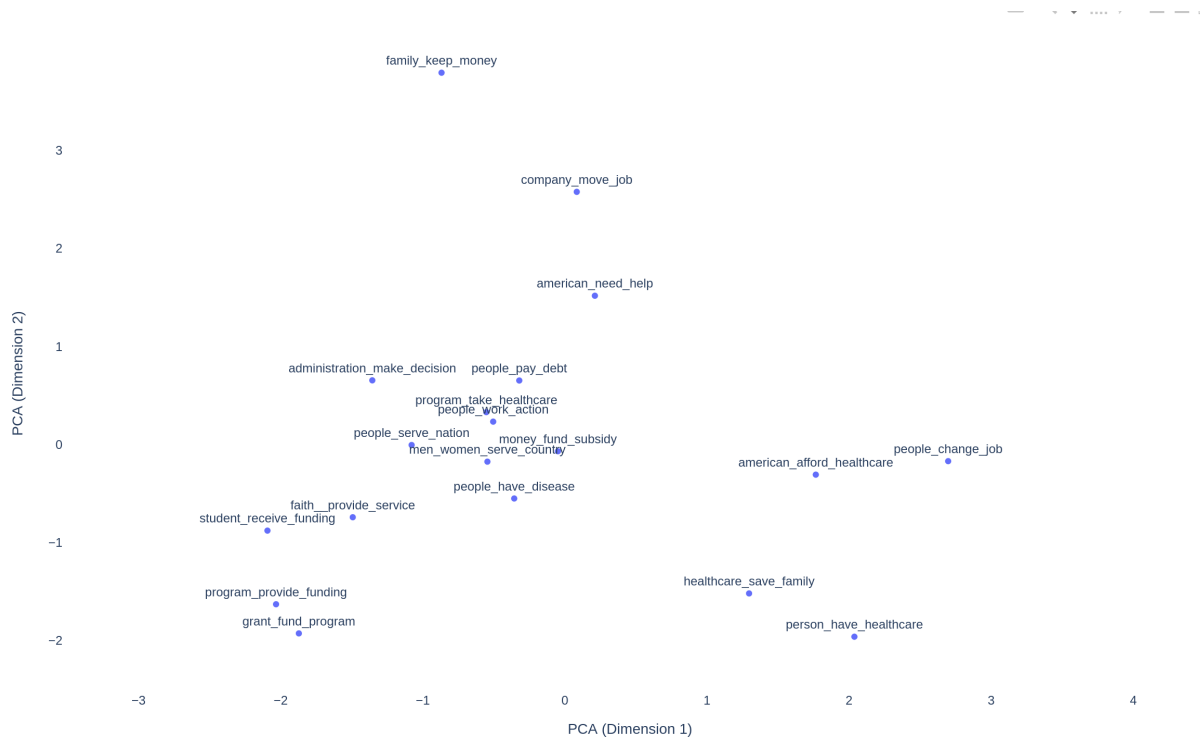
**Pointwise Mutual Information.** A simple approach to recover co-occurring narratives is to compute the pointwise mutual information (PMI) between pairs of narratives. For two narratives  $n_i$  and  $n_j$ , the pointwise mutual information is defined as:

$$PMI(n_i, n_j) = \log \left( \frac{P(n_i \cap n_j)}{P(n_i)P(n_j)} \right) \quad (3.3)$$

If two narratives are independent from one another, then their PMI is equal to zero. The larger the PMI, the more likely two narratives are to co-occur together, adjusting for their overall frequency. To analyze the odds of two narratives co-occurring in the same speech, we compute PMIs for pairs of narratives in the U.S. Congressional Record.

**Narrative Embedding.** A second approach is to train skip-gram embeddings for narratives. As done with word embeddings (Mikolov et al. 2013), this approach locates narratives that appear in similar contexts close to one another in a low-dimensional, dense vectorial space. For the U.S. Congressional Record, we define speeches as documents, narratives as tokens, and then train a Word2Vec model (skip gram with negative sampling) over ten epochs. The window size is set to an arbitrarily large number so that two narratives co-occurring in a speech are always in their respective windows. The dimension of vectors is set to 50. To graphically represent the resulting embedding, Figure I.1 plots a random sample of 20 narratives along the two first dimensions of a Principal Component Analysis (PCA), showing how the geometry represents intuitive topical information. Hence, the relationship between two narratives can be computed as the cosine similarity between their respective vectors.

Figure I.1: Geometry of Embedded Narratives



**Note:** This figure shows 20 random narratives in the embedding, projected to two dimensions using PCA.

**Results.** Table ?? applies these methods to the narrative “Saddam Hussein poses a threat”. We provide the 20 most associated narratives with our PMI metric as well as the cosine similarity of the narrative embedding vectors. Overall, both methods paint a similar picture. Saddam Hussein posed a threat to national security because of (supposedly possessed) nuclear weapons, the United States won the war but American troops had to make a sacrifice. In fact, Saddam Hussein did not possess nuclear weapons. The administration made a mistake and misled Americans.

Table I.1: Co-occurring Narratives – War on Terror

Pointwise Mutual Information	Narrative Embedding
iraq pose threat	iraq pose threat
iraq pose jeopardy	nation love peace
saddam hussein use chemical (weapon)	people love freedom
iraq have weapon mass destruction	american lose job
saddam hussein have weapon mass destruction	god bless troop
saddam hussein invade kuwait	people do job
saddam hussein use weapon mass destruction	saddam hussein have weapon mass destruction
administration mislead american	god bless america
freedom take tree	american lose family
iraqi take responsibility	iraq pose jeopardy
american risk family	men women put family
nation love peace	saddam hussein use chemical (weapon)
men women wear uniform	men women wear uniform
people love freedom	administration make mistake
administration make mistake	person serve country
iraqi take power	terrorist kill american
nation love freedom	economy create job
war torn country	democrat join republican
terrorist pose threat	american do job
military do job	people have problem

**Note:** This table presents narratives that tend to co-occur with “Saddam Hussein poses a threat”. Column (1) lists the 20 narratives with the highest PMI associated to this narrative. Column (2) lists the 20 narratives that are the closest to this narrative in the embedding (as measured by cosine similarity).

## J. Narratives on Social Media Data

Though we illustrate RELATIO’s workings on the U.S. Congressional Record, the package is not designed to be corpus-specific. In this section, we provide a short analysis of a Twitter corpus.<sup>2</sup> We focus on 32,323 tweets from the [Trump Tweet Archive](#). The tweets were written by Donald Trump between August 2011 and October 2020 and amount to 68,616 sentences.

We clean the text the same way as done for the congressional speeches, with the exception of stopwords, taken directly from the spaCy english model “en\_web\_core\_md”. We keep the 100 most frequent named entities and specify 100 additional clusters for the embeddings. We then count the resulting narrative tuples.

Figure J.1 represents Trump’s 50 most frequent narratives in the form of a narrative graph. The former president notably claims that “The Democrats hurt the country”, “steal the election” and “raise taxes”. “China manipulates its currency” and “steals from the people”, while “the establishment and special interests kill the country”.

The corpus is much smaller than the U.S. Congressional Record. It takes just a few minutes to run the full RELATIO pipeline on it. Interested readers will find a step-by-step tutorial to extract narratives from this corpus on our [GitHub repository](#).

---

<sup>2</sup>Researchers not involved in the development of RELATIO have applied it to newspaper articles (Ottonello and Song 2022) as well as social media data from different platforms (Sipka, Hannak and Urman 2021).



## K. List of Analyzed Narratives

Here we list all narratives mined from the Congressional Record corpus. The list includes dimension-reduced narratives appearing at least 50 times in the period 1994-2015. Narratives are sorted by most to least frequent.

people lose job, citizen abide law, american lose job, government run healthcare, american have healthcare, god bless america, people need help, god bless troop, worker lose job, small business create job, service connect disability, constitutional (thing) balance budget, people have healthcare, public held debt, people pay tax, people lose family, american pay tax, employee provide healthcare, people do job, men women serve country, people take opportunity, mortgage back military, american lose healthcare, american keep money, american lose family, program provide service, people make decision, people look job, healthcare provide healthcare, business create job, people have concern, american look job, family lose love, program provide funding, healthcare provide service, people have job, child have healthcare, people lose healthcare, company do business, people have opportunity, people want job, people keep money, american abide law, people get healthcare, U.S.C. authorize service, F.D.A. approve drug, people have insurance, agency provide service, american not-have healthcare, child attend school, period offer appropriate, program provide healthcare, government take healthcare, person have abortion, student attend school, republican take power, people get job, amount bear ratio, people not-have healthcare, defense submit congressional, people love freedom, people take healthcare, american want job, person serve country, american work hard, someone have opportunity, veteran receive healthcare, men women put family, people create job, someone pay attention, someone have healthcare, men women serve nation, people make money, government provide service, veteran serve country, god bless men women, someone do nothing, doctor make decision, republican join democrat, problem face country, agency administer program, agency furnish information , people find job, oil make profit, american give family, people commit crime, doctor take healthcare, law respect faith , child get education , program provide grant, republican do nothing, child receive education , community provide service, budget balance budget, american get healthcare, democrat take power, government make decision, american have concern, employee sponsor healthcare, administration take action, american pay attention, child have education , project create job, challenge face nation, people pay attention, men women risk family, service provide service, people have idea, people serve country, constitutional (thing) require balance budget, american need help, iraq pose threat, agency do job, american do job, program make difference, insurance deny coverage, american find job,

challenge face country, economy create job, coal fire plant, program create job, student participate program, people get money, people have problem, government take money, people need healthcare, people give family, men women give family, republican balance budget, republican shut government, bank make loan, god bless family, law enforcement add category, budget raise tax, government sponsor business, american take opportunity, problem face nation, federally fund program, federally fund study, federally recognize tribe, child receive healthcare, saddam hussein pose threat, healthcare save family, small business provide healthcare, people have money, family make sacrifice, people make choise, capital gain tax, people cross border, democrat raise tax, america have healthcare, child have opportunity, employee offer healthcare, commonwealth perform duty, american make sacrifice, director mean director, god continue america, american have opportunity, men women wear uniform, government do nothing, democrat join republican, company provide service, company hold business, person have child, family need help, budget increase tax, missile propel bomb, someone have idea, agency receive grant, people enter country, american want healthcare, family keep money, doctor provide healthcare, individual not-attain adult, legislative (election) have consequence, child need help, public create job, budget do nothing, country face challenge, government take property, judiciary conduct markup, agency take action, healthcare offer healthcare, family receive funding, administration do nothing, american create job, veteran make sacrifice, american do business, republican cut medicare, child not-have healthcare, people risk family, person make difference, american serve country, someone take healthcare, men women make sacrifice, tax kill job, tide lift ship, provider provide service, plan balance budget, america lead world, people have choise, american receive healthcare, family pay tax, government run program, american have job, people buy healthcare, program award grant, people earn money, people buy insurance, government take action, someone pay tax, people not-pay tax, company make profit, kid get education , people put family, senior have healthcare, corporation pay tax, attorney (judge) make decision, corporation do business, center provide service, person do job, student graduate school, nation face challenge, senior have prescription drug, people violate law, government do everything, american lack healthcare, agreement enters force, school provide education , government do business, government control healthcare, doctor provide service, person earn dollar, raw material meet road, energy create job, people not-have insurance, people express concern, lender make loan, job pay life quality, employee drop coverage, brave men women give family, company create job, employee provide coverage, healthcare offering healthcare, people have family, parent raise child, people buy car, parent send child, employee sponsor plan, person make decision, employee take (sick/paid/unpaid) leave, service connect veteran, program save family, child need healthcare, agency provide funding, people do business, american make choise, iraqi



take responsibility, founder create job, people have child, american lose unemployment (benefits), staff do job, people put money, american need job, nation produce oil, director considers appropriate, saddam hussein have weapon mass destruction, bush sign law, crime add category, person have healthcare, budget cut program, family have healthcare, company move job, business do business, physician provide service, people work job, program receive funding, government do job, people take risk, people lose insurance, stimulus package create job, country produce oil, american know truth, defense waive limit, people not-have job, american make energy, someone do job, hard earn money, people receive money, people need job, mother have child, program save money, program help american, gun abide law, men women serve military, people get education , senior pay premium, (sick/paid/unpaid) leave be guam, doctor perform abortion, american risk family, men women lose family, republican put together, defense notify congressional, individual receive service, terrorist kill american, child have diabetes, people provide service, outlay not-exceed payment, people have coverage, american make decision, american suffer diabetes, plan do nothing, american lose coverage, program provide education , people break law, veteran need healthcare, child review study, agency submit application, senior have choice, person provide service, american not-do job, people make sacrifice, people put together, american have choice, people have skill, individual purchase healthcare, iran develop nuclear (weapon), people need funding, employee sponsor coverage, god wipe shot, person make disaster, american need healthcare, child participate program, law prescribed witness, people participate program, people graduate school, people not-find job, government pick winners and losers, nation owes appreciation, democrat want tax, god grant strong, american make consumer, F.D.A. approve prescription drug, senior get prescription drug, republican cut program, people save money, energy affect family, person give family, government balance budget, fund create job, program help people, american need understand, brave men women risk family, republican give tax break, country support terrorist, people make mistake, person make dollar, agency receive fund, republican cut tax, country face problem, mother raise child, child reach future, family work hard, doctor practice healthcare, family send child, american sent message, program provide food, voter cast ballot, family take healthcare, individual have healthcare, people have income, employee provide service, american travel cuba, someone lose job, program provide money, person lose family, church offer prayer, people lose money, person dedicate family, american need (tax) relief, school hire student, people take money, people need service, american pay income, employee hire worker, administration make decision, people give money, employee drop healthcare, fuel drive economy, budget reduce deficit, grant submit application, people make living, individual participate program, veteran administer law, people make contribution, employee lose job, supreme make decision, story detail fight, company take opportunity,

carrier provide service, people raise concern, constitution require balance budget, individual perform service, someone make money, individual provide service, american have diabetes, budget increase deficit, school participate program, vacancy not-affect power, employee pay minimum wage, people make difference, people drive car, program meet regulation , men women do job, person seek abortion, child receive service, customer sell gun, senior need help, people work minimum wage, people pay money, people get insurance, american have insurance, regulation kill job, family have child, people suffer illness, people need understand, agency participate program, business hire worker, child get healthcare, action have consequence, student need help, veteran receive service, someone get money, drug save family, veteran receive money, person give baby (boomer), agency use fund, child lose parent, citizen pay tax, administration request level (spending), employee hire immigrant, someone seek recognition, republican play politics, small business offer healthcare, government approve healthcare, person have corrective action, american put food, storm cause damage, child witness violence, people do nothing, doctor deliver baby (boomer), american hold market, people call tax, student reduce class size, individual receive money, person have cancer, iran have nuclear (weapon), people work living, medicare provide healthcare, america face challenge, student receive grant, american pay money, american balance budget, life quality pay job, money pick deductible, expenditure not-exceed payment, mother take healthcare, car get gasoline, plan create job, american suffer illness, republican put plan, word have word, grant award grant, iran acquire nuclear (weapon), american want balance budget, budget cut deficit, policy create job, government fund program, car carry passenger, individual receive funding, senior have coverage, republican have idea, illness threaten family, person make contribution, F.D.A. regulate tobacco, american pay gasoline, people have diabetes, economy lose job, someone know nothing, person serve nation, country receive funding, government raise tax, people pay premium, people take grant, attorney (judge) do job, employee provide insurance, neighbor help neighbor, american make contribution, employee participate program, republican have plan, agency issue regulation , mother have diabetes, administration make mistake, american get tax, agency receive funding, firefighter lose family, medicare operate prescription drug, person have pregnancy, people have experience, people work family, pregnancy feel pain, plan meet regulation , people have education , american have government, government borrow money, judiciary recommend substitute, money pay tax, republican propose tax, physician take healthcare, freedom take tree, agency have authority, republican propose budget, attorney (judge) have (legal) statute, federally fund center, word do justice, student take loan, american buy healthcare, people work government, government solve problem, program serve child, people not-have opportunity, family give afghanistan, american want government, money save money, budget cut tax, people make dollar, student receive ed-

ucation , individual attain adult, terrorist pose threat, consumer have choice, grant make grant, brave men women put family, person serve military, american not-find job, program help child, challenge face america, school provide service, employee hire illegal immigrant, family raise child, country need energy, family send kid, american tighten belt, group run ad, program require appropriation, veteran control small business, people start business, company make money, people make fund, someone pay fair, child have future, tax create job, american lose money, friend make comment, F.D.A. approve consumer, government not-create job, law make appropriation, people work hard, people earn minimum wage, senior pay pocket, people need money, people remember courage, child suffer diabetes, someone take opportunity, american get job, people feel life quality, attorney (judge) follow law, appropriation do job, democrat do nothing, corporation make profit, someone make mistake, someone have insurance, hard earn dollar, agency have (legal) statute, bank issue card, government provide healthcare, money keep money, people want government, american make facility, people have illness, veteran get healthcare, employee do job, veteran have service, people make minimum wage, homeowner face mortgage, brave men women serve country, person perform abortion, program provide loan, program improve healthcare, parent make decision, nation face threat, agency receives grant, people lose coverage, people have program, student have education , school need help, american afford healthcare, defense prescribed regulation , republican play game, employee make contribution, disease threaten family, troop do job, american have coverage, crisis face country, constitution give authority, fund subsidize deficit, american take healthcare, american have voice, witness give witness, budget end medicare, people lose unemployment (benefits), person commit crime, people take job, worker join union, government create job, senior need prescription drug, authority provide funding, kid have healthcare, republican want medicare, H.M.O. make decision, program have effect, farmer need help, people love peace, mother have cancer, grant provide funding, people suffer diabetes, director award grant, tax benefit wealthy, people invest money, word give education , republican do job, country face crisis, people abide law, decision affect family, brave men women lose family, someone want job, appropriation make grant, government run deficit, business employ people, agency provide information , budget have deficit, someone have problem, person commit (legal) charge, agency serve school, people change job, american save money, person love family, worker form union, association give grade, community base healthcare, someone make decision, child enter school, men women earn dollar, animal lay milk, veteran suffer diabetes, immigrant enter country, people raise family, child graduate school, people have voice, corporation not-pay tax, child have disability, social security balance budget, american have money, someone get job, administration impose sanction, people filing bankruptcy, nothing create job, american deserve truth, agency provide technical assistance, injury

fire missile, soldier lose family, democrat put together, government provide funding, hard work american, constitution give power, social security run deficit, fund provide service, student take opportunity, grant fund project, law authorize compensation, law enforcement do job, service administer program, father serve war, food drug administration regulate tobacco, problem face american, iraq have weapon mass destruction, employee take opportunity, action have effect, people make future, grant use grant, people want healthcare, attorney (judge) interpret law, community promote abortion, child have family, family lose job, crime get gun, law authorize uniform, worker find job, doctor treat medicare, insurance raise interest/rate, obligation bearing interest/rate, agency enter contract, person get corrective action, people afford healthcare, doctor take medicare, people get help, employee pay premium, nation face crisis, people serve nation, doctor treat patient, money justify money, people have faith , business provide service, small business take opportunity, american earn money, company hold loan, project receive funding, american take job, senior take opportunity, military do job, someone loses job, corrective action save family, american get money, american put family, business hire people, people do best, service recommend substitute, people serve war, republican hold prisoner, attorney (judge) apply law, physician perform abortion, small business hire worker, america create job, american lose faith , american purchase healthcare, asset back military, self-employed executes budget, soldier give family, individual meet regulation , supreme rule unconstitutional, iran pose threat, people have difficulty, fund make funding, money pay money, program provide fund, community receive grant, mother give baby (boomer), bank provide loan, veteran serve nation, parent lose child, democratic offer alternative, physician assist suicide, money fund outlay, administration propose budget, institute conduct study, F.D.A. have authority, fed raise interest/rate, program provide resource, outlay exceed payment, program benefit american, agency share information , employee perform job, family put food, application meet regulation , reserve raise interest/rate, student receive degree (education), threat face nation, company employ people, people want buy, men women defend nation, electronic device save family, american hold love, people lose everything, appropriation fund program, budget reduces deficit, student get education , employee make decision, agency require information , service control small business, employee hire people, person take healthcare, american want together, crime commit crime, american keep healthcare, china join wto, person lose job, defense submit service, people lose love, republican want tax, tax benefit american, kid need help, physician deliver baby (boomer), news sign text, appropriation report appropriation, attorney (judge) enforce law, people do harm, program help poor, employee provide institution, god bless nation, people borrow money, american paid tax, crime use gun, project meet criterion, american deserve (tax) relief, fact impose tax, business make profit, small business hire people, people keep staff, network provide

service, american face problem, american face challenge, people want program, measure breach budget, people want help, government have power, country have deficit, american have enough, administration submit budget, business provide healthcare, child lose healthcare, people solve problem, tax stimulate economy, employee pay overtime, company sell consumer, public provide service, american want action, people have exhaust, employee hire permanent, director determines appropriate, person have baby (boomer), american have faith , worker earn minimum wage, student pursue education , agency receives fund, environmental protection agency regulate emission (pollution), american serve nation, government estimate contract, indian provide service, american send money, challenge face american, people take responsibility, attorney (judge) make determination, physician practice healthcare, justice write opinion, democrat have plan, country need help, project meet regulation , country make progress, nothing limit authority, government not-do nothing, student afford school, american rely oil, production shut supply and demand, company pay tax, worker support retiree, self-employed deduct healthcare, child have special (need/rule), school do job, water resource development authorize river, patient receive healthcare, service do job, people need (tax) relief, people do wrong, healthcare deny coverage, healthcare provide coverage, american keep job, measure would budget, program mean program, country sign treaty, bush nominate attorney (judge), community need help, people paid tax, business take opportunity, people wear clothes, people work money, government receive funding, people have government, democrat balance budget, people keep healthcare, F.D.A. approve facility, american pay fair, republican put budget, people serve military, someone have fact, iran obtain nuclear (weapon), physician make decision, mother pay tax, person make global warming, veteran need help, administration do everything, country participate program, ship carry passenger, person risk family, nation take action, veteran put family, appropriate rely payment, people live family, person earn men women, republican create job, country face threat, kid have opportunity, republican use filibuster, administration request amount, family have income, money fund abortion, threat face country, administration cut program, people lose saving, administration do job, people take action, offender commit crime, coal fire electricity, faith provide service, country have problem, law require budget, parent make choice, policy held debt, people make profit, defense transfer fund, manufacturer collect sale, government make promise, farmer plant crop, small business purchase healthcare, law authorize program, nation face problem, tax have effect, agency make decision, community make decision, grant give priority, budget cut education , people not-afford healthcare, administration threaten veto, immigrant pay tax, program benefit people, people provide healthcare, founder start business, american have housing, american have service, national preserve social security, money fund program, american provide family, federally fund project, patient save family, education award grant,

parent take child, someone have job, interest/rate authorize agency, innocent lose family, ship launch missile, individual pay tax, oil account america, parent have child, law authorize appointment, parent protect child, nothing do nothing, insurance drop coverage, country take opportunity, mother lose job, hero give family, people try help, people have power, legislative (election) call word, community provide funding, people pay debt, budget add trillion, veteran find job, american want reform, government run insurance, child have child, law prohibit practice, people have life quality, person make workforce, force base defense, someone raise staff, law authorize project, people do everything, staff put together, war torn country, plan provide (tax) relief, country lose job, budget reach together, budget report budget, american make ultimate sacrifice, agency award contract, people have opinion, kid graduate school, girl attend school, flood cause damage, senior have medicare, someone need help, men women fight war, american make difference, republican give tax, agency carry program, country have healthcare, physician provide healthcare, american deserve nothing, people receive funding, people make family, employee use (sick/paid/unpaid) leave, terrorist enter country, funding orphan child, juvenile commit violent crime, family adopt child, administration take position, job require skill, news face nation, people have cosponsor, republican end medicare, applicant submit application, program provide help, money fund subsidy, student leave school, story wrench heart, message transmit concurrent, news face country, plan have effect, person terminate pregnancy, patient need healthcare, medicare have prescription drug, people work small business, decision have effect, job pay minimum wage, people put food, child lose family, doctor make healthcare, person choose abortion, worker pay tax, terrorist attack country, american hold accountable, soldier pierce tank, small business pay tax, doctor practice offense, people support family, community participate program, someone get tax, institution provide service, people have confusion, american access healthcare, american have ballot, family make decision, raw material hit road, fund pay profit, people leave welfare, iraq invade kuwait, government have authority, student enter school, U.S.C. authorize uniform, someone commit crime, family balance budget, someone have experience, law provide authority, child eat lunch, medicaid provide healthcare, national conduct study, government issue identification, veteran receive compensation, law provide remedy, city mean city , grant provide service, administration propose tax, american do nothing, military make sacrifice, supreme declare unconstitutional, government take opportunity, america need help, small business create america, senior join H.M.O., family save money, people keep job, firefighter open fire, government issue picture, people move job, air force request application, internal revenue code strike period, mother have baby (boomer), employee sponsor insurance, corporation pay fair, bank lend money, government pay money, kid start smoking, someone do business, program outlive effectiveness, program provide profit, someone have heart,

nation sign treaty, god rest love, company have employee, individual lose job, worker do job, student make difference, government save money, agency conduct analysis, budget cut medicaid, people go job, parent care child, administration provide information , small business get loan, american do best, money outweigh money, physician treat patient, unemployment (benefits) look job, attorney (judge) make law, people get coverage, people balance budget, individual make decision, employee create job, program serve american, medicare cover service, american work money, railroad provide transportation, money get money, program need reform, filibuster raise tax, america need energy, agency make determination, wealthy pay fair, people help people, student receive funding, party reach agreement, program help family, veteran take opportunity, natural resource recommend substitute, people kill people, nuclear (weapon) pose threat, people have common sense, child lack healthcare, program take healthcare, people wear uniform, O.P.E.C. increase production, person have diabetes, market take healthcare, people have disease, senior pay deductible, republican cut education , program have success, police give family, consumer cause harm, child start smoking, american use site, person make choice, oil hit gasoline, republican take action, parent take healthcare, individual lose family, small business pool resource, child get good start, someone paid attention, police do job, child reach adult, government provide information , school use fund, healthcare improve healthcare, people preexist condition, employee not-provide healthcare, person make payment, program provide technical assistance, person commit offense, american deserve healthcare, american receive money, employee offering healthcare, republican want repeal (something), people want tax, people share concern, veteran receive recognition, senior make choice, people want america, senior afford prescription drug, job require education , nation love freedom, people not-get job, family leave welfare, parent send kid, appropriation provide funding, (sick/paid/unpaid) leave be columbia, father have diabetes, family face mortgage, government run plan, law authorize activity, capital gain interest/rate, american put money, veteran have healthcare, insurance provide coverage, america win war, money make difference, program provide essential , veteran participate program, university conduct study, authority make decision, measure not-increase deficit, doctor acquire disease, student use technology, bank provide service, american work government, problem face america, veteran serve war, veteran seek healthcare, person need help, american obtain healthcare, senior depend medicare, people have plan, county mean county, winners and losers appropriate fund, family take opportunity, employee provide money, family make choice, facility provide service, founder endow men women, person have heart, men women serve arm force, people have appreciation, tobacco cause cancer, county lose population, people make america, juvenile commit crime, amtrak provide service, saddam hussein use chemical (weapon), person lay family, individual take opportunity, fund carry funding, saddam hussein use

weapon mass destruction, government have responsibility, community use fund, party raise soft money, nation lead world, small business provide service, agency use information , veteran earn money, bureaucrat make decision, american not-afford healthcare, employee have healthcare, application decrease appropriation, party seek remedy, american feel pain, agency promulgate regulation , oil pay fee, student leave discipline, individual buy healthcare, people have feeling, major issue face country, commitment have witness, budget provide (tax) relief, soldier serve country, grant fund program, saddam hussein invade kuwait, person control small business, country lead world, medicare run money, worker exhaust money, people pay income, country make fund, family receive money, agency receive amount, american make fund, job support family, agency collect information , plan amends plan, disability receive education , african make contribution, judiciary conduct promulgate regulation, someone have diabetes, child need service, provider offer service, employee perform function, father write constitution, money hold bag (trash), person enter contract, medicare manage plan, people promote welfare, person graduate school, university do study, school receive funding, bible fulfil self-employed, student afford education , individual commit crime, medicare provide prescription drug, corporation take opportunity, program have (positive/negative) effect, student pay interest/rate, individual make legislative (election), people save family, family lose healthcare, people get loan, grant provide grant, government receive grant, idea have merit, program do nothing, store collect sale, weapon pierce tank, person receive amount, american lose insurance, company ship job, worker have exhaust, people buy consumer, fund provide funding, someone get healthcare, republican do everything, mother lose mother, mother graduate school, american sell consumer, american work minimum wage, reconciliation carry recommendation, supreme have (legal) statute, prescription drug save family, republican raise tax, school educate child, country head direction, community perform abortion, worker exhaust unemployment (benefits), people kill american, treasury striking-for purpose, offender serve (legal) sentence, railroad provide service, healthcare take healthcare, patient receive patient, people want spend, american work family, doctor see medicare, agency desire grant, american lose concern, republican take money, individual bearing interest/rate, government give money, healthcare work american, republican gain power, government incur money, employee do business, transportation issue certification, business have employee, people have ability, plan cut tax, senior get healthcare, american owe appreciation, people not-pay attention, company participate program, faith base community, service connect condition, program have bipartisan (support), people drive mile , american pay dollar, O.P.E.C. cut production, american have life quality, air force indicate appointment, government have money, people work american, people fight war, people leave country, american support action, law authorize study, attorney (judge) represent client, program reduce disease, administration



enforce law, consumer make instruct, american fight war, project provide water, applicant receive grant, grant receive grant, government make payment, program provide community, baby (boomer) reach adult, people try job, american have idea, someone know someone, child have food, democrat have idea, small business need help, people face mortgage, government protect citizen, people sign petition, republican provide (tax) relief, school make decision, program help nation, people have trouble, money paid amount, federally assist housing, servicemembers lose family, budget receive recommendation, republican propose medicare, senior rely medicare, person obtain abortion, disaster affected area, republican have opportunity, transaction lack abuse, environmental protection agency do job, regulation have effect, budget pay debt, F.D.A. approve plant, people use money, veteran suffer illness, senior not-afford prescription drug, troop make sacrifice, trade create job, people not-have money, constitution vested power, administration make bad (problem), police lose family, patient get child, american rely program, oil reach gasoline, nation make progress, good start serve child, people use drug, american have confusion, men women defend country, people work action, american deserve government, people have courage, people bear staff, program help community, people establish justice, republican make promise, healthcare make healthcare, budget do job, american make life quality, american buy car, program benefit child, injustice motivate crime, people make comment, veteran need service, citizen do extraordinary, corporation make charity, illegal immigrant commit crime, small business afford healthcare, judiciary do job, plan have merit, social security pay money, person receive money, F.D.A. approve medicine, people provide family, student pay loan, people have freedom, national submit intelligence, democrat have power, program provide housing, business pay tax, violent crime serve (legal) sentence, healthcare save money, people pay mortgage, law confer power, people not-want job, medicare pay doctor, people have resource, resource do job, national make recommendation, employee provide notice, program help farmer, people commit suicide, student bring gun, business employ american, measure provide authority, people make judgment, bush threaten veto, child receive food, american want soluton, medicine save family, doctor save family, oil create job, unemployment (benefits) find job, america pay tax, child brought gift, child have succeed, american exhaust unemployment (benefits), someone have concern, girl have baby (boomer), government pay debt, worker change job, tax reduce deficit, water provide water, school reduce class size, family attend school, person perform service, program serve people, company provide healthcare, constitution prohibit america, county provide service, people take loan, people establish constitution, people watch network, employee perform service, good start provide service, government work american, airline lose job, employee provide education , pesticide cause disease, child watch network, someone drive car, victim receive compensation, people have nothing, people have commitment, disability

limit major issue, iraqi take power, iraq pose jeopardy, someone want tax, defendant award punitive damage, employee maintain plan, america have opportunity, parent have choice, appropriation provide fund, judiciary take action, budget take healthcare, eligible participate program, school receive fund, bank do business, worker lose healthcare, refinery make profit, american take grant, money amount taxable, program achieve goal, nation face fight, budget make bad (problem), director issue information , (sick/paid/unpaid) leave be virgin , ethanol blend gasoline, senior receive healthcare, kid not-have healthcare, farmer lose crop, law provide protection, business provide job, republican offer budget, attorney (judge) promulgate regulation , nation love peace, men women write constitution, social security strike period, worker look job, someone serve country, diabetes affect american, program help the needy, plan reduce deficit, american send child, congressional budget define money, export create job, people attend school, china manipulate dollar, program do job, farmer have crop, irs target group, production use american, american provide service, budget set priority, folk do job, citizen have healthcare, nation need energy, administration mislead american, homeland attack unit, business estimate tax, people get tax, lender have lender, child have illness, someone get everything, opinion show american, government not-do job, small business employ workforce, child have cancer, student pursue degree (education), witness limit medicine, small business provide job, crime use weapon, group provide service, disaster cause damage, program carry program, american buy insurance, employee sponsor saving, employee offer coverage, toxic (thing) cause cancer, employee provide employee, project provide service, people face problem, application contain information , men women do nothing, nation make commitment, someone want buy, people stop smoking, american do everything, government make commitment, guard save family, company bring job, administration have plan, people broken law, agency conduct program, self-employed employ individual, director take action, plan save money, plant lose job, program receive grant, employee put money, people loan money, american drive car, car cross border, american produce energy, small business receive loan, katrina hit ocean, burmese contain export, someone have child, federally fund healthcare, republican refuse action, lender participate program, people have disability, budget increase debt, people lose concern, national provide service, government make healthcare, employee have employee, police put family, person make healthcare, student get loan, american earn minimum wage, administration negotiate trade, soldier make sacrifice, american scratch staff, agency give priority, brave men women serve nation, nothing affect authority, program receives funding, people get service, brave men women make sacrifice, men women lose job, child receive funding, family afford healthcare, people send kid, community base community, someone do everything, bush nominate circuit (courts), student earn degree (education), someone buy gun, unemployment (benefits) exhaust money, uniform

make sacrifice, student teach child, person get healthcare, uniform serve country,

## L. List of entities labeled as procedural or noise terms

We present here the entities that we consider to be procedural or not interpretable after dimension reduction. We name them by the most frequent term in their cluster. These entities are excluded in our main analysis. The list below shows all such excluded entities.

absolutely right, accept, accomplish, accurate, adam, address, adjustment, administrative office, advisory, afford, agreeable, akin, alan, allegiance, amaze, anderson, andrew, angel, approach, argument, aristide, arlen, available, ayes, baca, ball, barbara, bartlett, basic, be, beautiful, behind, ben, benjamin, berkley, berman, biggert, billy, bipartisanship, blm, bob, bobby, body, bolton, bonner, bonnie, boswell, bottom, brian, briefing, brook, brownfields, buck, buckley, building, bumper, bureau, burn, burton, byron, cabinet, cafta, candidate, cardoza, carl, carolyn, certain, chamber, chandler, chapter, charles, charlie, cheap, cheney, chenoweth, chris, christopher, chuck, clark, classify annex, clear, close, cloture, coat, code, colin powell, colloquy, columbus, commenters, complicate, compromise, comptroller, conferee, conference, confirm, congressional caucus, congressional gold medal, congressman, consent, conservation, consideration, continue, corp, corrine, costello, counsel, couple, cranston, damato, dan, daniel, danny, date, dave, david, day, de la garza, debate, debbie, demint, dennis, department, designate, desjarlais, determine, dialogue, dianne, diaz balart, dice, dick, dick cheney, different, dirksen office building, disabled, disagree, disappoint, discuss, distinguish, division, do, doable, document, don, double, downstream, draconian, draft, dreier, due, ed, eddie bernice, edward, elizabeth, emerge, enforce, engage, enter, entity, eric, erisa, essence, establish, estimate, estrada, exactly, exceed, excite, executive, exist, extension, extent, extraneous material, extremely important, eye, factor, familiar, fannie, far, far proceeding, federal, filner, final, fine, finish, first, first responder, fit, follow, foot, formal, forward, frank, frank frank, frank gallegly, fred, freddie, fy, ga, gallegly, gary, gate, generalparagraph, george, gibbon, ginny, glosson, go, government accountability office, government printing office, governor, graf, gram, grisham, half, ham, happen, happy, harry, hearing, helm, helpful, henry, herger, herseth, high, hinchey, hj re, holden, hollen, homeownership, hour equally divide, hour workweek, howard, hr, hussein, ii, iii, implement, implication, important, inc, inspire, interior, introduce, ironic, irresponsible, item, jack, james, james madison, jeff, jerry, jesse, jim, jimmy, jo ann, joe, johnny, join, joint, joint chief staff, jon, joseph, journal last day, jr, judicial conference, kansa, keith, kent, kucinich, lahood, larry, late, latourette, led-better, letter, linda, list, lng, local coordinate entity, loretta, lot, lungren, ma, madam, madison, mail, main street, majority, manner, manzullo, mark, martha, martin, mary, max, mccotter, mcdonald, mcmorris, md, medically necessary, mel, michael, miguel

estrada, mike, milosevic, minute, minute remain, mitch, mn, modify, moment, multiple, myrick, na, naive, nancy, nc, nd, net neutrality, newark, newt, nez, nick, nj, nomination, non, norton, np, nt do, numbered, numerous, objection, observation, offer, office, ok, olver, open, oppose, orrin, outside, overall, oversight, overwhelm, owen, page, park, parliamentarian, pas, pat, patrick, patten, patty, paulson, percent, perez, pete, peter, petitioner, petraeus, phil, pitt, plate, point, point order, politically correct, polluter, popular, post, powell, preside officer, president, president tempore, procedure, process, projection, proponent, provision, qualify, quayle, question, quick, randy, rank, rapidly, re, read, ready, recommit, reconsider, redesignate, redesignating, reduction, regular order, rehberg, rehnquist, rein, remain available expend, remark, reno, repeat, report, representative, requisite number word, revolve fund, reyes, richard, rick, riggs, robert, rodney, roger, ron, rothman, rule, rule administration, rulemaking, rumsfeld, russell, russell office building, sam, sander, save, schedule, schmidt, sean, secure, select, select intelligence, select reserve, sen, sense, separate, session, sex offender, shay, shuler, side, sj re, skelton, solomon, somalia, sorry, sotomayor, southern, southwick, specie, speech, spouse, st, stack, standard, starr, statement, station, status quo, stearns, stennis, steve, straight, string, structure, stump, subparagraph, subtitle, sufficient second, sullivan, sunset, supplemental, support, susan, sutton, table, ted, territory, th, thanks, thing, thornton, tim, tim myrick, time, tn, todd, tom, tomorrow, tonight, tradition, tsa, tx, unacceptable, underfunded, underlie, unique, unprecedented, upon, va, valid, ve get, version, view, vitally important, wait, walter, wasserman, way mean, wayne, whatever, whichever, whole, whose, william, willing do, withdraw, woolsey, worthwhile, xiv, xx, xxi, xxii, year, year old, zero, zoe,

## M. List of Stopwords

We remove common words in English, procedural terms specific to the U.S. Congressional Record, numbers, as well as all the names of Congress members and U.S. States. The complete list is below:

i me my myself we us our ours ourselves you your yours yourself yourselves he him his himself she her hers herself it its itself they them their theirs themselves what which who whom this that these those a an the and but if or because as until while of at by for with about against between into through during before after above below to from up down in out on off over under again further then once here there when where why how all any both each few more most other some such no nor not only own same so than too very much actual abraham aderholt adler akaka alexander al-lard allen amash amodei andrews applegate archer arcuri army ashcroft ayotte bac-chus bachmann bachus baesler baird baker baldacci baldwin ballance ballenger barber barca barcia barkley barletta barlow barr barrasso barrett barrow barton bass bate-man baucus bayh bean beatty beauprez becerra begich beilenson bell benishek bennet bennett bentivolio bentley bentsen bera bereuter berry bevill biden bilbray bilirakis bingaman bishop black blackburn blackwell blagojevich bliley blumenauer blumen-thal blunt blute bocchieri boehlert boehner bonamici bond bonilla bonior bono booz-man boren borski boucher boustany boxer boyd boyda bradley brady braley breaux brewster bridenstine bright brooks broun browder brown brownback brownley bryan bryant buchanan bucshon bumpers bunn bunning burgess burns burr burris bustos butterfield buyer byrd byrne callahan calvert camp campbell canady cannon cantor cantwell cao capito capps capuano cardenas cardin carnahan carney carper carson carter cartwright case casey cassidy castle castor castro cazayoux chabot chafee chaf-fetz chambliss chapman chiesa childers chocola christensen chrysler chu cicilline clarke clay clayton cleaver cleland clement clinger clinton clyburn coats coble coburn cochran coffman cohen cole coleman collins combest conaway condit connolly conrad cony-ers cook cooksey cooley coons cooper coppersmith corker cornyn corzine costa cotton courtney coverdell cox coyne craig cramer crane crapo crawford cremeans crenshaw crowley cruz cubin cuellar culberson cummings cunningham dahlkemper daines dan-forth danner darden daschle davis dayton deal deconcini defazio degette delahunt de-laney delauro delay dellums denham dent derrick deutch deutsch dewine dickey din-gell dixon djou dodd doggett dole domenici donnelly dooley doolittle dorgan doyle drake driehaus duckworth duffy duncan dunn durbin durenberger edwards ehlers ehrlich ellison ellmers ellsworth emanuel emerson engel english ensign enyart enzi es-hoo esty etheridge evans everett ewing exon faircloth fallin farenthold farr fattah fawell

fazio feeney feingold feinstein ferguson fields fincher fingerhut fischer fish fitzgerald fitzpatrick flake flanagan fleischmann fleming fletcher flores foley forbes ford fortent-berry fossella foster fowler fox foxx frahm frankel franken franks frelinghuysen frisa frist frost fudge funderburk furse gabbard gallego gallo ganske garamendi garcia gardner garrett gejdenson gekas gephardt geren gerlach gibbons gibbs gibson giffords gilchrest gillibrand gillmor gilman gingrey gingrich glenn glickman gohmert gonzalez goode goodlatte goodling goodwin gordon gorton gosar goss gowdy graham gramm grams grandy granger grassley graves grayson green greene greenwood gregg griffin griffith grijalva grimm grucci gunderson guthrie gutierrez gutknecht hagan hagel hahn hall halverson hamburg hamilton hanabusa hancock hanna hansen hare harkin harman harper harris hart hartzler hastert hastings hatch hatfield hayes hayworth heck hefley heflin hefner heineman heinrich heitkamp heller helms hensarling herrera higgins hill hilleary hilliard himes hinojosa hirono hoagland hobson hochbrueckner hodes hoeffel hoekstra hoeven hoke holding hollings holt honda hooley horn horsford hostettler houghton hoyer hudson huelskamp huffington huffman hughes huizenga hulshof hultgren hunter hurt hutchinson hutchison hutto hyde inglis inhofe inouye inslee isakson israel issa istook jackson janklow jefferson jeffords jeffries jenkins jindal johanns john johnson johnston jones jordan joyce kagen kaine kanjorski kaptur kasich kassebaum kaufman keating keller kelly kempthorne kennedy kennelly kerns kerrey kerry kildee kilmer kilpatrick kilroy kim kind king kingston kinzinger kirk kirkpatrick kleczka klein kline klink klobuchar klug knollenberg kohl kolbe kopetski kosmas kratovil kreidler kuhl kuster kuykendall kyl labrador lafalce lamborn lampson lancaster lance landrieu langevin lankford lantos largent larsen larson latham latta laughlin lautenberg lazio leach Leahy lee lehman levin levy lewis lieberman lightfoot lincoln linder lipinski livingston lloyd lobiondo loeb sack lofgren long longley lott lowenthal lowey lucas luetkemeyer lugar lujan lummi luther lynch machtley mack maffei mahoney majette maloney manchin mann manton marchant marino markey marshall martinez martini mascara massa massie matheson mathews matsui mazzoli mccain mccandless mccarthy mccaskill mccauley mcclintock mccloskey mccollum mcconnell mccrery mcurdy mcdermott mcgovern mchale mchenry mchugh mcinnis mcintosh mcintyre mckeon mckinley mckinney mcleod mcmahon mcmillan mcnerney mcnulty meadows meehan meek meeks melancon menendez meng merkley messer metcalf meyers mfume mica michaud michel mikulski miller mineta minge minnick mitchell moakley molinari mollohan montgomery moore moorhead moran morella moynihan mullin mulvaney murkowski murphy murray murtha musgrave myers nadler napolitano natcher neal nelson nethercutt neugebauer neumann ney nickles noem nolan northup norwood nugent nunes nunn nunnelee nussle nye obama oberstar obey olson ortiz orton osborne ose otter owens oxley packard packwood palazzo pallone pappas parker pascrell pastor paul paulsen paxon payne pearce pease pell pelosi penny

perlmutter perriello perry peters peterson petri phelps pickering pickett pickle pin-  
gree pittenger pitts pocan poe polis pombo pomeroy pompeo porter portman posey  
poshard pressler price pryce pryor putnam quigley quillen quinn radanovich radel ra-  
hall ramstad rangel ravenel redmond reed regula reichert reid renacci renzi reynolds  
ribble rice richardson richmond ridge riegler rigell riley risch rivers robb roberts roby  
rockefeller rogers rodriguez roe roemer rogan rogers rohrbacher rokita rooney ros  
rose roskam ross rostenkowski roth rothfus roukema rowland royce rubio ruiz run-  
yan ruppertsberger rush ryan ryun sabo salazar sali salmon sanchez sanders san-  
dlin sanford sangmeister santorum sarbanes sarpalius sasser sawyer saxton scalise  
scarborough schaefer schaffer schakowsky schatz schauer schenk schiff schneider  
shock schrader schrock schroeder schultz schumer schwartz schwarz schweikert  
scott seastrand sekula sensenbrenner serrano sessions sestak sewell shadegg shaheen  
sharp shaw shays shelby shepherd sherman sherwood shimkus shows shuster sim-  
mons simon simpson sinema sires sisisky skaggs skeen slattery slaughter smith snow-  
barger snowe snyder sodrel solis souder southerland space specter speier spence spratt  
stabenow stenholtz stevens stewart stivers stockman stokes strickland studs stupak  
stutzman sundquist sununu swalwell sweeney swett swift synar takano talent tan-  
credo tanner tate tauscher tauzin taylor tejeda terry tester thomas thompson thorn-  
berry thune thurman thurmond tiaht tiberi tierney tipton titus tonko toomey torkild-  
sen torres torricelli traficant tsongas tucker turner udall unsoeld upton valadao valen-  
tine van vargas veasey vela velazquez vento visclosky vitter voinovich volkmer vu-  
canovich wagner walberg walden walker wallop walorski walsh walz wamp ward  
warner warren waters watson watt watts waxman webb weber webster weiner welch  
weldon weller wellstone wenstrup westmoreland wexler weygand wheat white white-  
house whitfield whitten wicker williams wilson wise wittman wofford wolf wom-  
ack woodall wu wyden wynn yarmuth yoder yoho young zeliff zimmer aye section  
subsection sec act roll yea yes nai no none nay mr mrs senator gentleman gentle-  
men gentlewoman gentlewomen gentlelady gentleladies sir committee speaker chair-  
man ask yield will shall can say let adjourn speak etc thank absent tell per part now  
month today con pro thereabout thereafter thereagainst thereat therebefore therebe-  
fore thereby therefor therefore therefrom therein thereafter thereof thereon thereto  
theretofore thereunder thereunto thereupon therewith therewithal hereabout hereafter  
hereat hereby herein hereinafter hereinbefore hereinto hereof hereon hereto hereto-  
fore hereunto hereunder hereupon herewith whereabouts whereas whereafter whereat  
whereby wherefore wherefrom wherein whereinto whereof whereon whereto where-  
under whereupon wherever wherewith wherewithal bill house member senate amend-  
ment section subsection committee subcommittee line article vote insert clerk speaker  
district state adjourn monday tuesday wednesday thursday friday saturday sunday  
january february march april may june july august september october november de-



cember adjourn board general legislature legislation parliament amend reject appeal  
appear answer appoint ask author calendar chair chairman colleague commission  
congress court file motion floor leader object paragraph unanimous pending revise  
quorum permission secretary resolution pass record print current one two three four  
five six seven eight nine ten eleven twelve thirteen fourteen fifteen sixteen seventeen  
eighteen nineteen twenty thirty forty fifty sixty seventy eighty ninety hundred hun-  
dreds thousand thousands alabama alaska arizona arkansas california colorado con-  
necticut delaware florida georgia hawaii idaho illinois indiana iowa kansas kentucky  
louisiana maine maryland massachussetts michigan minnesota mississippi missouri  
montana nebraska nevada hampshire new jersey mexico york carolina north south  
east west dakota ohio oklahoma oregon pennsylvania island rhode tennessee texas  
utah vermont virginia washington wisconsin wyoming united

# Bibliography

- Arora, Sanjeev, Yingyu Liang, and Tengyu Ma.** 2016. "A Simple but Tough-to-beat Baseline for Sentence Embeddings."
- Gardner, Matt, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew Peters, Michael Schmitz, and Luke S. Zettlemoyer.** 2017. "AllenNLP: A Deep Semantic Natural Language Processing Platform."
- Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adriane Boyd.** 2020. "spaCy: Industrial-strength Natural Language Processing in Python."
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean.** 2013. "Efficient Estimation of Word Representations in Vector Space." *arXiv preprint 1301.3781*.
- Ottonello, Pablo, and Wenting Song.** 2022. "Financial Intermediaries and the Macroeconomy: Evidence from a High-Frequency Identification." National Bureau of Economic Research.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay.** 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, 12: 2825–2830.
- Pennington, Jeffrey, Richard Socher, and Christopher Manning.** 2014. "Glove: Global Vectors for Word Representation." 1532–1543. Doha, Qatar: Association for Computational Linguistics.
- Řehůřek, Radim, and Petr Sojka.** 2010. "Software Framework for Topic Modelling with Large Corpora." 45–50. Valletta, Malta: ELRA. <http://is.muni.cz/publication/884893/en>.
- Reimers, Nils, and Iryna Gurevych.** 2019. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." *arXiv preprint arXiv:1908.10084*.

**Sipka, Andrea, Aniko Hannak, and Aleksandra Urman.** 2021. "Comparing the Language of QAnon-related content on Parler, Gab, and Twitter." *arXiv preprint arXiv:2111.11118*.

**Titre :** Essais en Economie Politique

**Mots clés :** réseaux sociaux, mouvements sociaux, préférences politiques, analyse de texte

**Résumé :** Cette thèse de doctorat en économie politique est composée de trois chapitres indépendants. Les thèmes principaux abordés sont les mouvements sociaux, leur développement récent sur les réseaux sociaux, ainsi que l'étude quantitative par intelligence artificielle du discours politique. Le chapitre 1 évalue les effets du mouvement #MeToo sur la criminalité sexuelle aux États-Unis. En exploitant l'existence de plaintes reportées avec un délai à la police, je propose une approche simple pour dissocier la probabilité qu'une victime porte plainte du nombre de crimes sexuels commis dans une série chronologique de crimes sexuels signalés à la police. Résoudre ce problème de mesure de longue date me permet d'étudier l'évolution de la criminalité sexuelle autour du mouvement #MeToo. Je constate que la sensibilisation aux violences sexuelles et la probabilité de porter plainte augmentaient déjà, tandis que les crimes sexuels étaient en baisse depuis plusieurs années avant la médiatisation soudaine de #MeToo sur les réseaux sociaux. Le mouvement a renforcé ces tendances plus larges. Le chapitre 2 suggère que les réseaux sociaux ont des effets ambigus sur les mouvements sociaux. Bien qu'ils facilitent la coordination entre les manifestants et permettent de nouvelles formes d'activisme en ligne, ils peuvent également dessiner des trajectoires

de radicalisation politique. Pour étayer ces affirmations, le chapitre analyse le mouvement des Gilets Jaunes en France en 2018-2019. Il montre que les manifestations locales ont mené à la création de grandes communautés d'activistes sur Facebook. Cependant, ces communautés sont progressivement devenues plus antagonistes, négatives et idéologiquement polarisées. Les plus modérés des interlocuteurs sont partis, tandis que ceux qui sont restés se sont radicalisés. L'algorithme de recommandation de Facebook a probablement contribué à ces trajectoires de radicalisation en présentant systématiquement du contenu radical à ces utilisateurs. Le chapitre 3 développe une nouvelle approche algorithmique pour mesurer les rôles sémantiques dans les documents originaux, permettant la construction de triplets « *agent-verbe-patient* ». Il combine ensuite les plongements lexicaux avec un algorithme de regroupement pour produire des poncifs interprétables de faible dimension (par exemple, « l'impôt tue l'emploi »). La méthode est appliquée à des milliers de discours parlementaires aux États-Unis. Le logiciel libre d'accès *reLatio*, écrit en Python, est également fourni pour faciliter d'autres applications.

**Title :** Essays in Political Economy

**Keywords :** social media, social movements, political preferences, natural language processing

**Abstract :** This PhD dissertation in political economy consists of three independent chapters studying social movements, social media, and political narratives. Chapter 1 quantifies the impact of the #MeToo movement on sex criminality in the United States. Exploiting the existence of delayed reports to the police, I propose an econometric model to disentangle crime reporting from crime prevalence in crime databases. Solving this long-standing measurement problem allows me to study trends in sex criminality around the #MeToo movement. I find that sexual violence awareness and sex crime reporting had been increasing, and sex crime incidence had been decreasing for several years before #MeToo in October 2017. Nonetheless, the movement's sudden mediatization on social media largely reinforced these broader trends. Chapter 2 argues that social media have ambiguous effects on protest movements. While they facilitate coordination among protesters and lead to persistent online activism, they may also foment radicalization among protesters. To support these claims, the chapter analyzes the Gilets

Jaunes protests in France in 2018–2019. It shows that local street protests triggered the creation of large communities of protesters on Facebook. However, these communities progressively became more antagonistic, negative, and ideologically segregated. Moderate discussants left, those who remained radicalized, and Facebook's recommender algorithm likely contributed its fair share by consistently showcasing radical content. Chapter 3 develops a new algorithmic approach to measure economic and political narratives from large text corpora. It uses recent neural networks to annotate semantic roles from raw documents, allowing for the construction of *agent-verb-patient* triplets. It then combines phrase embeddings with a clustering algorithm to recover low-dimensional, interpretable narratives (e.g., "taxes kill jobs"). An application to thousands of speeches from the *U.S. Congressional Record* illustrates the method's potential. An open-source Python package *reLatio* is also provided to support further applications.