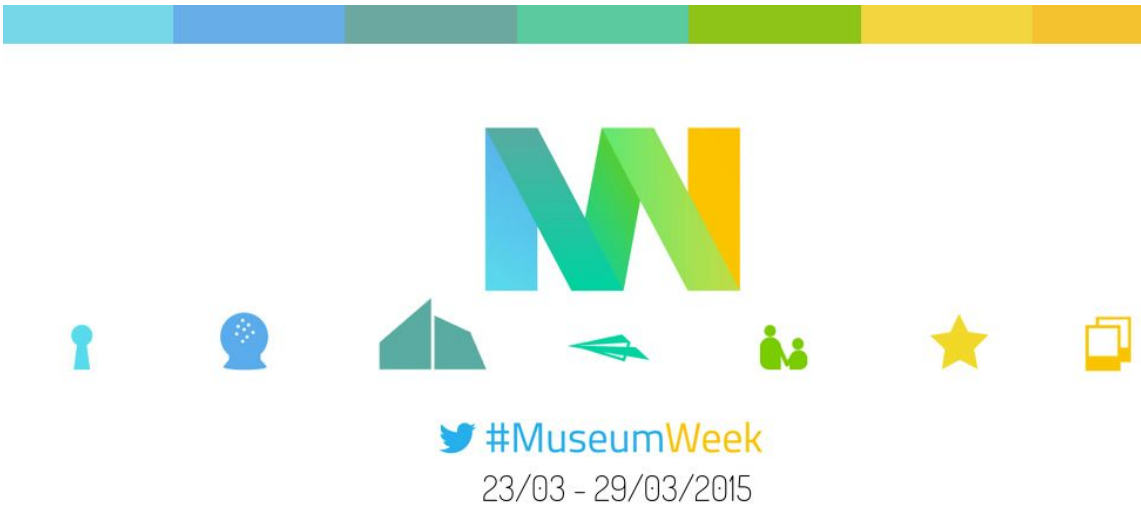
Analyses quantitatives et catégorielles des tweets émis dans le cadre de l'évènement #MuseumWeek2015

Antoine Courtin, Nicolas Foucault



Analyses quantitatives et catégorielles des tweets émis dans le cadre de l'évènement #MuseumWeek2015

-

Labex *Les passés dans le présent*

Septembre 2015

Version

0.2.

Objet du livrable

Résultat de l'étude sur l'opération *#MuseumWeek* de 2015 (2ème édition), suite au devis contractualisé entre le Labex *Les passés dans le présent* et le Ministère de la Culture et de la Communication.

Projet associé

Comm-Num - Médiation culturelle et circulation des connaissances à l'ère numérique : communication institutionnelle, pratiques éditoriales, pratiques informationnelles et dispositifs socio-techniques.

Projet du Labex piloté par le laboratoire MoDyCo, UMR 7114 (Université Paris Ouest Nanterre La Défense-CNRS) ; www.modyco.fr

Date

Septembre 2015.

Coordinatrice

Brigitte JUANALS, Maître de conférences habilitée à diriger des recherches en Sciences de l'information et de la communication.

Auteurs

Antoine Courtin, ingénieur d'étude, Labex *Les passés dans le présent.*

Nicolas Foucault, ingénieur de recherche, Labex *Les passés dans le présent.*

Sommaire

Préambule **3**
Organisation du document 3
Les données 3
Cadre d'étude **4**
Le projet Comm-Num 4
Objet d'étude : la #MuseumWeek 4
1 Étude **5**
1.1 Problématique et analyses 5
1.2 Un contexte d'étude double 6
1.2.1 Contexte international 6
1.2.2 Contexte national : le cas français 7
1.2.3 Quelques chiffres au niveau national 7
1.3 Le traitement des données 9
1.3.1 Collecte 10
1.3.2 Prétraitement 12
1.3.3 Échantillonnage 13
2 Résultats **14**
2.1 Analyses quantitatives 14
2.1.1 Tweets, langues et origines des participants 14
2.1.2 Identification des institutions françaises les plus actives 20
2.2 Analyses catégorielles 22
2.2.1 Méthodologie de classification des tweets 23
2.2.2 Classification des tweets et des institutions françaises 25
2.3 Analyse factorielle : analyse en composantes principales 32
3 Conclusion **35**
Annexes **36**
A1 Informations demandées à Twitter pour collecter et préparer les données 37
A2 Exemple de données collectées durant la #MuseumWeek 42
A3 Données supplémentaires et hashtags non officiels 44
A4 Pays ayant une institution inscrite à la #MuseumWeek 45
A5 Pays d'origine des participants à la #MuseumWeek 46
A6 Erreurs de classification langagière commises par Twitter 48
A7 Autres résultats 51
A8 Devis proposé au Ministère de la Culture et de la Communication 58

Préambule

Ce rapport a été rédigé conjointement par Antoine Courtin, ingénieur d'étude et Nicolas Foucault, ingénieur de recherche au Labex *les passés dans le présent* et supervisé par Jean-Luc Minel, professeurs des universités et directeur du laboratoire MoDyCo. Les analyses, résultats et conclusion faites à travers ce rapport relèvent donc de leur entière responsabilité et n'engagent en rien les institutions citées ou partenaires du Labex.

Nous tenons à remercier chaleureusement toutes celles et ceux qui nous ont permis de réaliser ce travail et en particulier les nombreux professionnels des musées avec qui nous avons eu la chance de collaborer lors des réunions et ateliers réalisés en concertation avec le Ministère de la Culture et de la Communication (MCC).

Organisation du document

Ce rapport est découpé en quatre parties de la manière suivante :

(i) Introduction ;
(ii) Étude ;
(iii) Résultats ;
(iv) Conclusion.

L'introduction présente la cadre et le focus de l'étude que nous avons menée cette année à propos de la *#MuseumWeek*.

La seconde partie est construite en deux temps. Dans un premier temps, nous présentons la problématique, les analyses et le double contexte (international et national) de l'étude. Dans un second temps, nous présentons en détail les traitements que nous avons réalisés sur les données d'étude collectées durant la *#MuseumWeek*.

La troisième partie présente les résultats des analyses associées[1].

La dernière partie présente les premières conclusions que nous avons pu tirer de ce travail.

Les données

Une partie des données collectés durant la semaine de la *#MuseumWeek*[2] est disponible via Internet sur un espace de stockage du MoDyCo réservé au MCC. Les visuels produits dans ce rapport sont accessibles par le même biais au format png et/ou svg par souci d'exploitation.

[1] Ces résultats sont produits en adéquation avec les lots de prestations définis dans le devis proposé au Ministère courant mai 2015. Le devis est joint en annexe de ce livrable en section A8.

[2] Ces données ne sont pas fournies à l'état brut, mais prétraité (cf. section 1.3). Elles correspondent aux corpus de type fr-fr$_1$ et fr-fr$_2$ décrits en section 2.1.1 et référencés au tableau 3 de la page 18.

Cadre d'étude

Dans cette partie, nous présentons le cadre et l'objet d'étude de ce rapport.

Le projet Comm-Num

Le projet Comm-Num du Labex *Les passés dans le présent* porte à la fois sur l'analyse des dynamiques d'articulation entre la médiation culturelle et scientifique et sur les techniques numériques innovantes. Ce projet comporte 2 axes de recherche pensés en interaction :

- Le premier axe, « Actualité, politique et pratiques de la publicisation des connaissances par les institutions culturelles et scientifiques à l'ère numérique », traite des politiques de communication et des pratiques éditoriales de ces institutions. Elles sont matérialisées dans la construction d'un *écosystème informationnel numérique* et les mutations des formes de médiation culturelle et scientifique.
- Le deuxième axe, « Humanités numériques et transformations dans la production et la circulation des connaissances scientifiques et culturelles », traite des actions d'une dizaine d'institutions patrimoniales sur les réseaux sociaux. Cet axe se fonde sur des techniques appliquées en Traitement Automatique des Langues.

Objet d'étude : la *#MuseumWeek*

La *#MuseumWeek* est un événement culturel qui prend naissance dans les sphères numériques et se matérialise dans des lieux culturels tels que les musées.

Cet évènement d'un nouveau genre a été impulsée par une douzaine de Community Managers d'institutions culturelles françaises et Twitter France en 2014. Sa seconde édition a eu lieu cette année en 2015. Comme l'an passé, elle s'est déroulée sur une durée de sept jours au cours de la semaine du 23 au 29 mars. Cette année, la *#MuseumWeek* a bénéficié d'un rayonnement international, ce qui n'était pas le cas l'an passé.

L'étude que nous présentons dans ce rapport porte sur l'édition 2015 de la *#MuseumWeek*[3]. Chaque jour, un hashtag "thématique" différent[4] permettait aux établissements participants de valoriser leurs collections, leurs activités ou leur programmation sur la plate-forme Twitter, tout en incitant les publics à partager leurs propres expériences et contenus sur ce réseau social.

[3] http://museumweek2015.org/fr
[4] http://museumweek2015.org/fr/programme

1 Étude

Cette partie se décompose en 3 volets. À la section 1.1, nous présentons la problématique, les analyses et les objectifs de notre étude. À la section 1.2, nous présentons son contexte. Enfin, à la section 1.3, nous présentons les traitements de données que nous avons réalisés.

1.1 Problématique et analyses

Notre problématique d'étude est la compréhension des aspects communicationnels de la *#MuseumWeek*. Les analyses menées pour y répondre se décomposent comme suit :

1. Des analyses manuelles dans la presse et les médias (classiques et numériques) à propos des aspects de communication événementielle de l'opération et des discours qui l'accompagne. Ces analyses font l'objet d'un second rapport[5].
2. Des analyses automatiques du contenu des tweets de la *#MuseumWeek*. Ces analyses font l'objet du présent rapport. Nous avons opté pour des approches développées en Traitement Automatique des Langues[6] (ou TAL) afin de les réaliser.

Les analyses menées dans notre étude résultent donc d'une approche hybride qui se situe à la frontière entre la linguistique et l'informatique. Ces analyses portent aussi bien sur les tweets émis par les institutions inscrites à la *#MuseumWeek* que sur ceux émis par les autres acteurs (institutionnels ou non) qui y ont participé sans être inscrits.

Nos analyses sont de deux types :

(i) Analyses quantitative ;
(ii) Analyses catégorielles.

Le premier type d'analyse permet d'observer globalement la manière dont s'est déroulée la *#MuseumWeek*. Dans ce volet, les analyses sont basées sur l'exploitation des corpus collectés durant l'opération, on observe des comportements de masse. **Le second type d'analyse** permet d'appréhender plus précisément le comportement des institutions. Dans ce volet, les analyses sont basées sur la classification automatique supervisée des tweets à partir des indices linguistiques identifiés dans les tweets. Dans les deux cas, nous avons décidé d'accorder une place importante à la représentation visuelle des résultats afin de faciliter leur compréhension et leur exploitation[7].

[5] Clara Licht, Marion Rampini, *L'évènement culturel #MuseumWeek*, rapport Labex, juin 2015.
[6] https://fr.wikipedia.org/wiki/Traitement_automatique_du_langage_naturel
[7] Nous rappelons que les visuels produits dans ce rapport sont à disposition du MCC sur un espace de stockage du MoDyCo réservé à cet effet (cf. préambule).

1.2 Un contexte d'étude double

Cette année, la *#MuseumWeek* a pris un essor remarquable en France comme à l'étranger. Dans cette section, nous offrons quelques précisions dans ces deux contextes.

1.2.1 Contexte international

Au niveau international, on compte 2 917 institutions inscrites à la *#MuseumWeek* (sur une période de 58 jours) dont 2 825 avec un compte Twitter (contre 630 l'an dernier), depuis 1 104 villes et 71 pays différents[8]. Les figures 1 et 2 illustrent respectivement le nombre d'institutions inscrites par jour à la *#MuseumWeek* et leur localisation sur la planisphère.

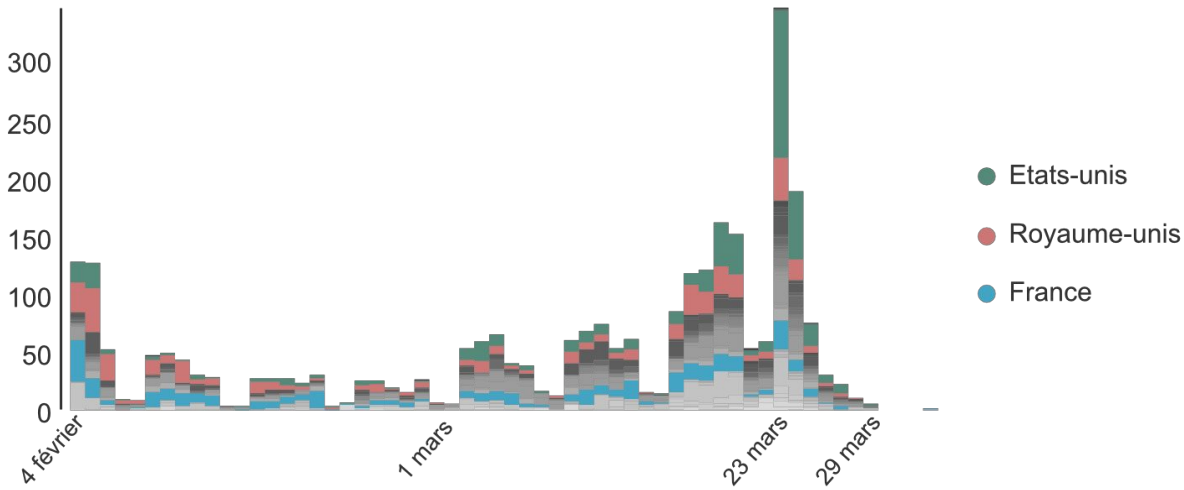


Figure 1 | Nombre d'institutions culturelles inscrites à la *#MuseumWeek* (2015) pour toute la durée de l'opération (58 jours). 4 février : début de l'opération. 23-29 mars : semaine de la *#MuseumWeek*.

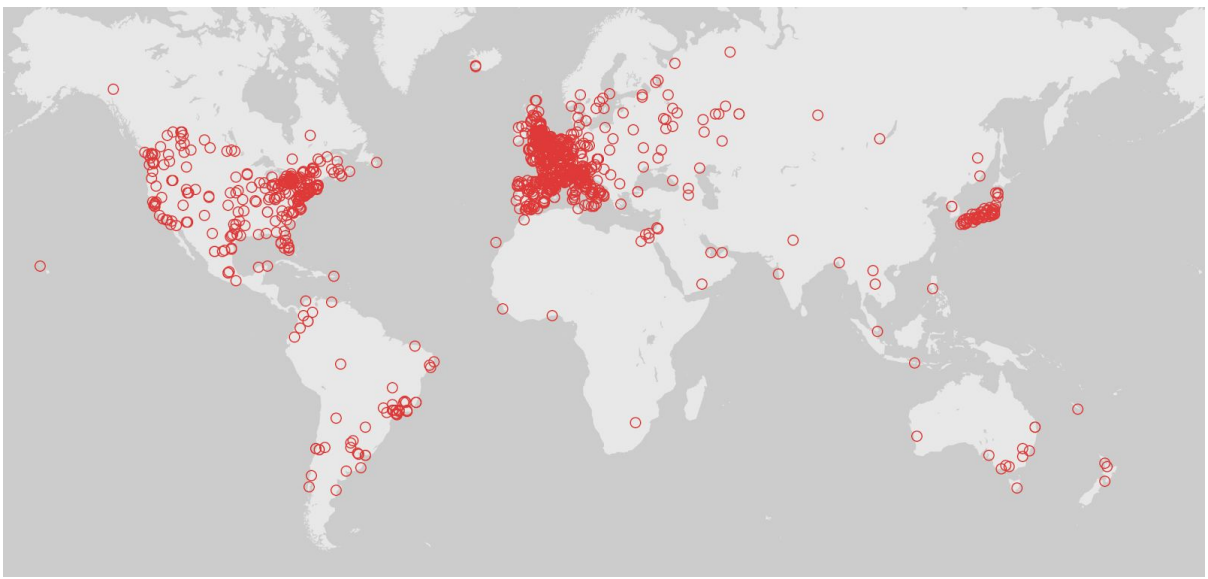


Figure 2 | Localisation des institutions culturelles inscrites à la *#MuseumWeek* (2015), obtenue à partir des informations de géolocalisation du centre d'émission des tweets par Twitter.

[8] La liste complète de ces pays est donnée en annexe de ce rapport à la section A4.

1.2.2 Contexte national : le cas français

Il faut cependant relativiser l'ampleur de cette nouvelle édition de la *#MuseumWeek*. En effet, si on note par exemple en France 184 inscrits l'année dernière contre 358 cette année (soit une hausse de 94% d'inscrits, presque 2 fois plus donc), il faut préciser que les contraintes d'admissibilités ont été élargies cette année. Ainsi, certains inscrits ne sont pas des institutions culturelles au sens propre (p. ex. *AhAhAh*, *Collectif Arts croisés*, *Sciences Aco* et *Mom'Art*), mais plutôt des *acteurs culturels* ; en l'occurrence des collectifs, associations, centres ou communautés qui évoluent dans l'univers du culturel et du numérique. À noter aussi, la forte présence d'autres types d'institutions culturelles, comme par exemple les services d'archives[9]. Ainsi, l'opération *#MuseumWeek* accroît son rayonnement culturel cette année par l'intégration d'institutions non muséales à l'ensemble de ses participants.

1.2.3 Quelques chiffres au niveau national

Type d'institutions participantes

Aux 358 institutions françaises inscrites à la *#MuseumWeek* cette année, correspondent 107 types différents d'institutions culturelles. Les 10 types les plus fréquents sont indiqués dans le tableau 1, accompagnés du nombre d'institutions couvertes à chaque fois.

Type d'institution	#inscrits
Musée, Monument historique	26
Centre de culture scientifique, technique, industrielle	18
Monument historique	12
Musée, Centre de culture scientifique, technique, industrielle	7
Archives	6
Centre d'interprétation	6
Musée, Centre de culture scientifique, technique, industrielle, Monument historique	4
Orchestre	4
Théâtre	4
Bibliothèque	3
Total	**180**

Tableau 1 | Nombre d'institutions françaises inscrites à la *#MuseumWeek* (2015) pour chacun des 10 types d'institution les plus fréquemment inscrites à l'événement en France.

[9] On doit préciser ici que 8 de ces institutions peuvent être assimilés à des services d'archives, bien que seulement 6 soient identifier comme tel dans le formulaire d'inscription.

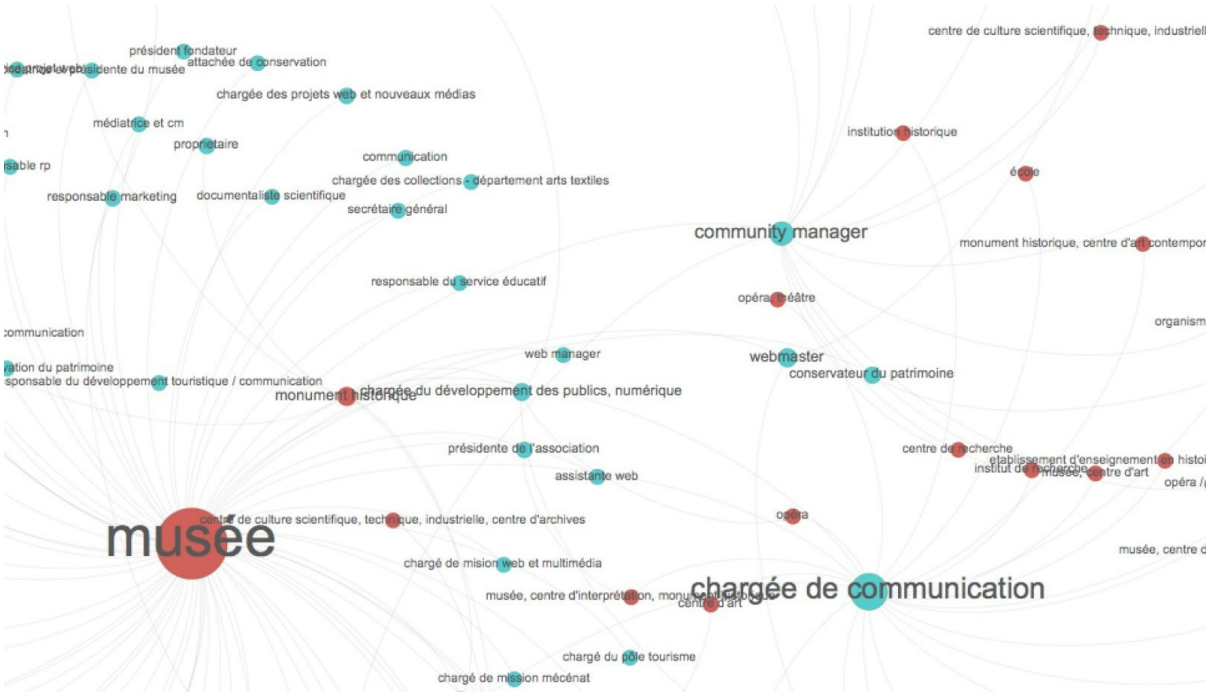


Figure 3 | Extrait des différents types d'institutions culturelles françaises ayant participées à la *#MuseumWeek* (2015) en lien avec les attributions majeures de leur représentants.

Représentants institutionnels et attributions professionnelles

La figure 3 donne un extrait des différents types d'institutions culturelles françaises ayant participées à la *#MuseumWeek*, en lien avec les attributions majeures de leur représentant (c.-à-d. celui qui a inscrit l'institution qu'il représente à la *#MuseumWeek*).

Aux 358 représentants des institutions françaises inscrites à la *#MuseumWeek* cette année, correspondent 140 attributions professionnelles différentes dont les sept plus fréquentes sont[10] :

- Responsable (de) ;
- Assistante (de) ;
- Chargé (de) ;
- Attaché (de) ;
- Chef ;
- Coordinateur ;
- Directeur.

[10] Nous avons obtenus ces différentes attributions après regroupement des divers intitulés de poste trouvés dans les formulaires d'inscription de la *#MuseumWeek* par le biais du logiciel **OpenRefine** (http://openrefine.org).

1.3 Le traitement des données

Le traitement des données est une tâche centrale de notre étude[11]. Elle a notamment fait l'objet de nombreuses discussions lors des réunions de co-organisation de la cellule d'évaluation de la *#MuseumWeek* au MCC sous la direction de M^me^ Jacqueline Eidelman.

Comme l'an passé, nous avons réalisé nous-mêmes les traitements de données de la *#MuseumWeek* car Twitter a refusé cette année de s'investir dans cette tâche, ne serait-ce qu'en nous offrant un accès privilégié aux données de la *#MuseumWeek* (nos échanges à ce propos sont disponibles en annexe de ce rapport à la section A1).

La figure 4 illustre globalement la chaîne de traitements que nous avons appliquées aux données de la *#MuseumWeek* en indiquant les formats d'entrée/sorties de chaque étape de traitement et le type des mécanismes de pilotage des traitements associé à chacune d'elle.

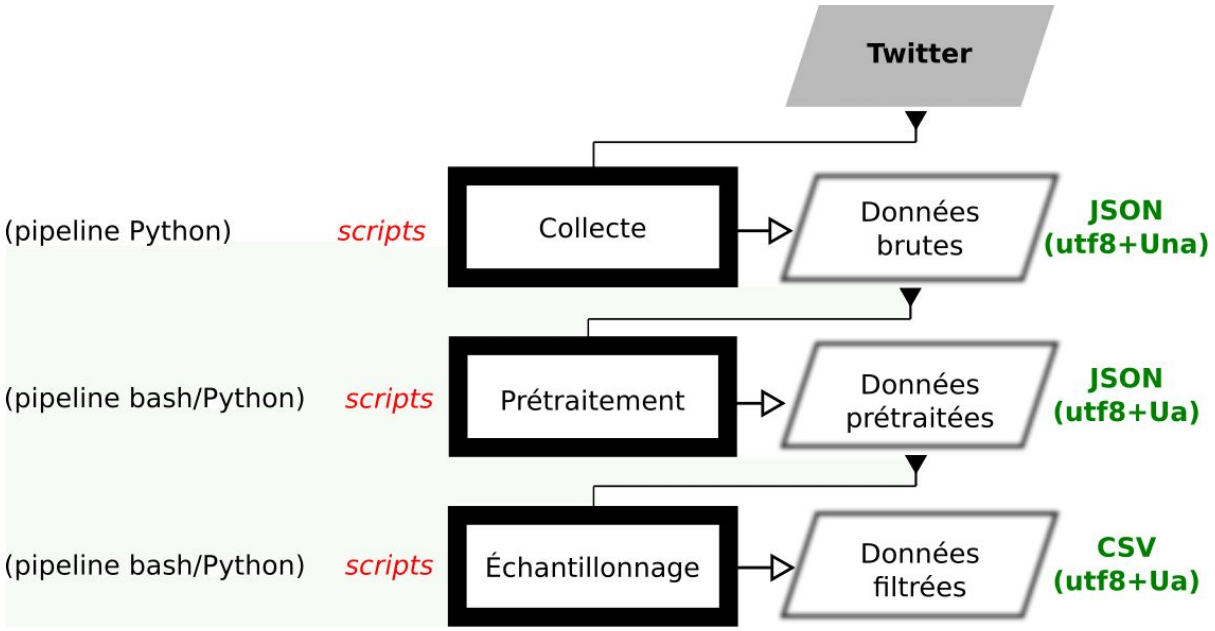


Figure 4 | Chaîne de traitement globale appliquée sur les données de la *#MuseumWeek* (2015). Au centre (boîte noires) : étapes de traitement. À droite en gris : entrées/sorties et en vert : formats d'entrées/sorties et encodage (Una : Unicode avec représentation non ascii, Ua : Unicode avec représentation ascii). Texte à gauche (rouge/noir) : type des mécanismes de pilotage des traitements.

Le découpage de cette section suit les étapes de traitements indiquées à la figure 4 : la section 1.3.1 présente la collecte des données, la section 1.3.2 présente leur prétraitement et la section 1.3.3 leur échantillonnage en sous corpus.

[11] Nous rappelons qu'une partie des données de la *#MuseumWeek* est à disposition du MCC sur un espace de stockage du MoDyCo réservé à cet effet (cf. préambule).

1.3.1 Collecte

La collecte est une étape qui a consisté à obtenir les données d'étude de la *#MuseumWeek*.

Nous présentons dans cette section l'approche globale que nous avons suivi, les contrôles de données que nous avons réalisés, les stratégies spécifiques de collecte appliquées ainsi qu'une première présentation des corpus obtenus.

Approche

Nous avons utilisés les API publiques de Twitter[12] sans utiliser de services tierces pour collecter les données de la *#MuseumWeek*. Ceci nous a permis d'une part de contrôler la fiabilité des données collectées et d'autre part de ne pas fournir nos clés de compte Twitter à un service tierce comme c'est souvent le cas. En revanche, en utilisant un accès public (c.-à-d. non payant) aux API de Twitter, nous avons été confrontés aux nombreuses restrictions imposées par la société Twitter dans ce cas[13].

Contrôle

Nous avons réalisés de nombreux tests des procédures de collecte proposées par Twitter afin d'éviter tout problème au moment de leur utilisation lors de la *#MuseumWeek*. Cependant, malgré ces précautions, nous avons dû revoir nos stratégies de collecte de tweets au cours de l'opération. Ainsi, nous sommes passés d'une stratégie de collecte des tweets par *streaming*[14] à une stratégie de collecte de tweets par *searching*[15, 16].

Si l'équipe d'organisation de la *#MuseumWeek* chez Twitter France n'a pas participé au traitement des données de l'événement dans notre étude, elle nous a cependant permis de consulter les données archivées dans la *TimeCapsule*[17]. Nous avons ainsi pu constater que les données issues de notre collecte sont valides au ses où elles correspondent substantiellement à celles qui se trouvent archivées dans la *TimeCapsule*[18].

---

[12] Il n'y a pas d'autre moyen que d'utiliser les API de Twitter pour collecter des tweets.
[13] Pour un tour d'horizon de ces limitations, se référer à : https://dev.twitter.com/rest/public/rate-limits.
[14] En *streaming*, on récupère les tweets au moment même de leur envoi via l'API *stream* de Twitter.
[15] En *searching*, on récupère les tweets jusqu'à 7 jours après leur envoi via l'API *search* de Twitter.
[16] Lors d'événements mondiaux comme la *#MuseumWeek*, le flux de tweets est tel que Twitter ne peut pas répondre aux requêtes en temps réel. Ainsi, en streaming, certains tweets sont manqués.
[17] http://bit.ly/1e4RE5m
[18] Le corpus de la *TimeCapsule* (voir note précédente), comme le notre, compte environ 180K tweets (sans doublon, retweets exclus), émis par les participants de la *#MuseumWeek* durant l'opération (23-29 mars ; institutions ou individus). Les tweets de la *TimeCapsule* et ceux de notre corpus contiennent au moins un des huit hashtags thématiques de l'événement et sont structurés au format json. À l'instar de notre corpus, certains noms de métadonnées des tweets ou métadonnées elle-mêmes de la *TimeCapsule* diffèrent de ceux habituellement utilisés par Twitter.

Stratégies

Nous avons réalisée **2 collectes de données** pour notre étude :

1. Une collecte des tweets émis par les participants à la *#MuseumWeek* (institutions ou individus). Les informations collectées englobent l'ensemble des métadonnées de tweet définies par Twitter[19] (un exemple est donné en annexe à la section A2.1) ;
2. Une collecte des comptes des institutions inscrites à la *#MuseumWeek*. Les informations collectées englobent l'ensemble des métadonnées biographiques de comptes définies par Twitter[20] (un exemple est donné en annexe à la section A2.2).

La première collecte a duré environ 5 mois (4 février - 30 juin), tandis que la seconde a duré 4 mois (4 mars - 30 juin).

Voici les stratégies que nous avons appliqué dans chaque cas :

- **Stratégies de collecte des tweets** :
    - Entre le 4 février et le 23 mars (c.-à-d. entre l'annonce officielle et le début de l'opération), nous avons mis en place une collecte des tweets par *streaming* pour chacun des 8 hashtags thématiques de l'opération (7 hashtags journaliers et 1 hashtag général)[21] à l'aide de la librairie Twarc[22] ;
    - Du 23 au 29 mars, semaine de la *#MuseumWeek,* et jusqu'au 30 juin, nous avons mis en place une collecte des tweets par *searching*. Celle-ci a été faite pour les mêmes 8 hashtags que précédemment, toujours à l'aide de Twarc.
- **Stratégie de collecte des biographies de comptes** :
    - Entre le 4 mars et le 30 juin, nous avons mis en place la collecte journalière des biographies de comptes Twitter des institutions inscrites à la *#MuseumWeek,* à l'aide de la librairie *Twython*[23].

[19] Listing des métadonnées associées à un tweet : https://dev.twitter.com/overview/api/tweets.
[20] Listing des métadonnées associées à un compte Twitter : https://dev.twitter.com/overview/api/users.
[21] Les 7 hashtags journaliers sont les suivants : **#secretsMW**, **#souvenirsMW**, **#architectureMW**, **#inspirationMW**, **#familyMW**, **#favMW**, **#poseMW**. Le hashtag général est le suivant : **#MuseumWeek**.
[22] **Twarc** est une librairie écrite dans le langage de programmation Python qui permet de communiquer facilement avec les API de Twitter pour la collecte, le post-traitement et l'analyse de tweets ; https://github.com/edsu/twarc.
[23] **Twython** est une librairie écrite dans le langage de programmation Python qui offre des fonctionnalités similaires à celles de Twarc, sans les aspects de post-traitement et d'analyse de tweets proposé par Twarc ; https://github.com/ryanmcgrath/twython. Nous avons opté pour une collecte des biographies de comptes via Twython plutôt que Twarc, car nous disposions déjà d'un tel système de collecte développé au MoDyCo par Antoine Courtin à l'occasion de la *#MuseumWeek* de 2014.

Corpus

Les 2 collectes que nous avons mené ont abouti à la création de **10 corpus de données**[24] :

- 9 corpus de tweets :
    - 8 corpus thématiques (un corpus par hashtag de la *#MuseumWeek*) ;
    - 1 corpus global (ALL) qui réuni les 8 corpus thématiques précédents.
- 1 corpus de biographies.

Les corpus collectés sont conséquents et possèdent une couverture linguistiques étendue. Par exemples, à l'état brut (c.-à-d. sans pré-traitement), le corpus global de tweets contient presque 2 millions de tweets pour une taille de 10 GB et compte 41 langues différentes tandis que le corpus de biographies représente un volume de 1 GB[25, 26]. Nous projetons d'exploiter le corpus de biographies très prochainement.

Le nombre de tweets par corpus collecté est indiqué dans le tableau 2 en fin de section.

1.3.2 Prétraitement

Le prétraitement est une étape qui consiste à préparer les données de collecte pour leur échantillonnage. Les étapes intermédiaires de pré-traitements de données que nous avons appliqués sur les corpus thématiques de tweets collectés sont les suivants :

1. Suppression des données non valides (c.-à-d. mal formées ou incomplètes) ;
2. Suppression des données en doublon ;
3. Résolution des problèmes d'encodage (transcodage des caractères HTML[27] inclus) ;
4. Conversion du format de collecte (json) vers le format d'analyse (csv) ;
5. Normalisation textuelle des données en français standard[28] avec correction des mentions et des hashtags non valides présents dans les tweets.

Le nombre de tweets par corpus prétraités est indiqué dans le tableau 2 en fin de section.

---

[24] Nous disposons par ailleurs d'un 11ème corpus de 1 704 tweets japonais que nous projetons d'utiliser à terme pour consolider les données de la *#MuseumWeek*. Nous donnons plus de détail à propos de ce corpus en annexe (voir section A3).

[25] Ces biographies, correspondent à des versions journalières des 2 825 biographies d'institutions inscrites à la *#MuseumWeek*, collectées entre le 4 mars et le 30 juin, à compter du jour d'inscription des institutions à la *#MuseumWeek*.

[26] Les informations biographiques collectées ne sont pas forcément exhaustives ni homogènes entre institutions, tout dépend du statut privé/public que les institutions ont choisi d'attribuer sur la période de collecte à la fois à leur compte Twitter (statut public durant l'événement pour 100% des institutions françaises inscrites à la *#MuseumWeek*) et à chaque donnée biographique (contrôle impossible à mettre en place à moins de poser la question directement à chacune des institutions inscrites).

[27] Toutes nos données sont encodées en UTF-8 ; https://fr.wikipedia.org/wiki/UTF-8.

[28] Ce travail inclut notamment la conversion vers leur forme complète des abréviations et de certains termes ou symboles utilisés pour gagner de la place dans les tweets (p. ex. ***pquoi*** ou ***collec°*** sont réécrit en ***pourquoi*** et ***collection*** respectivement et **+** est réécrit ***plus*** selon le contexte).

1.3.3 Échantillonnage

L'échantillonnage est une étape qui consiste à filtrer les données de prétraitement afin de pouvoir les analyser.

Les données peuvent être filtrées sur la base de différents critères (p. ex. la langue des tweets, leur date d'émission ou bien encore leur auteur). Tout dépend des analyses à entreprendre.

Le critère de filtrage initial que nous avons utilisé est :

- **La date d'envoi des tweets** ; la fourchette de temps choisie court sur la semaine du 23 au 29 mars 2015 (c.-à-d. la semaine de la *#MuseumWeek*).

Nous reviendrons ultérieurement sur les critères de filtrage supplémentaires que nous avons utilisé dans notre étude.

Le nombre de tweets par corpus échantillonné sur la date d'envoi des tweets, comme spécifié plus haut, est indiqué dans le tableau 2.

Corpus	*Collecte*	*Prétraitement*	*Échantillonage*
			23 - 29 mars
secretsMW	147 672	84 715	81 955
souvenirsMW	172 010	70 797	70 344
architectureMW	202 182	101 549	100 086
inspirationMW	173 784	70 874	69 313
familyMW	155 201	56 222	53 894
favMW	158 442	58 380	55 568
poseMW	151 459	48 549	43 326
MuseumWeek	606 477	435 548	371 179
ALL	**1 767 227**	**660 312**	**589 107**

Tableau 2 | Nombre de tweets par corpus et par étape de traitement[29].

[29] Ici, nous n'avons considéré qu'une partie des données collectées, à savoir les données collectées entre le 4 février et le 3 avril. Ainsi, le nombre total de données de collecte indiqué dans le tableau 2 pour cette période (c.-à-d. 1 757 227 tweets) est bien inférieur au nombre total de données de collectes dont nous disposons, collectées sur l'intégralité de la période de collecte des tweets (4 février - 30 juin).

2 Résultats

Cette partie se décompose en deux sections : la section 2.1 présente les résultats des analyses quantitatives menées sur les corpus de tweets que nous avons collecté durant la semaine de la *#MuseumWeek* alors que la section 2.2 présente les résultats de l'analyse catégorielle que nous avons mené sur ces mêmes corpus.

2.1 Analyses quantitatives

Dans cette section, nous présentons des résultats attenants à la nationalité des participants de la *#MuseumWeek* ainsi qu'aux langues qu'ils parlent ou qu'ils ont utilisé pour tweeter avec un regard spécifique sur la Francophonie, la France et le français.

2.1.1 Tweets, langues et origines des participants

Les résultats présentés dans cette section sont issus de l'analyse des tweets émis durant la *#MuseumWeek* et plus précisément, sur ceux émis durant la semaine de la *#MuseumWeek*. Nous les avons obtenus en utilisant tantôt la nationalité des participants tantôt leur langue ou celle des tweets comme critères de filtrage supplémentaires pour échantillonner les données (cf. section 1.3.1 et section 1.3.3 pour plus de détail à propos de l'échantillonnage des données et des corpus utilisés dans cette section).

Origines des participants

On compte pas moins de 182 nationalités représentées par l'ensemble des participants de la *#MuseumWeek* (institutions et individus confondus). La différence entre l'ensemble des pays touchés par la *#MuseumWeek* (182 pays au total) et l'ensemble des pays qui comptent au moins une institution inscrite à la *#MuseumWeek* (71 pays au total) est de 111 (avec entre autre le Népal, le Ghana, le Soudan et l'Irak). Cette différence montre clairement l'intérêt porté à la *#MuseumWeek* dans le monde entier et pas seulement dans des pays concernés activement par l'événement (au sens où ils possèdent des institutions inscrites).

Nous fournissons **à titre indicatif** le classement des 10 pays ayant émis le plus de tweets pendant la *#MuseumWeek* à la figure 5. En effet, nous avons pu réaliser au cours de nos analyses de corpus que les attributs de métadonnées des tweets qui pouvaient permettre d'identifier la nationalité des participants, à savoir les attributs `geo`, `coordinates`, `country` et `location`, étaient ou obsolètes (`geo`) ou non représentatifs (`coordinates` et `country`) ou encore difficilement exploitables (`location`). Nous expliquons cela en détail en annexe à la section A5.

Les résultats de la figure 5 ont été obtenu sur la base de l'attribut location. Cet attribut possède une valeur dans 70,1% des cas (c.-à-d. pour 414 525 tweets sur l'ensemble des données). Nous avons pu inférer le pays d'origine des auteurs des tweets dans 62.7% des cas (c.-à-d. pour 259 146 tweets, soit pour 43.9% des données). La méthode d'inférence que nous avons utilisé est décrite en annexe à la section A5.2. Il serait nécessaire d'enrichir cette méthode afin d'améliorer ses performances et donc les résultats de la figure 5 qui en découlent.

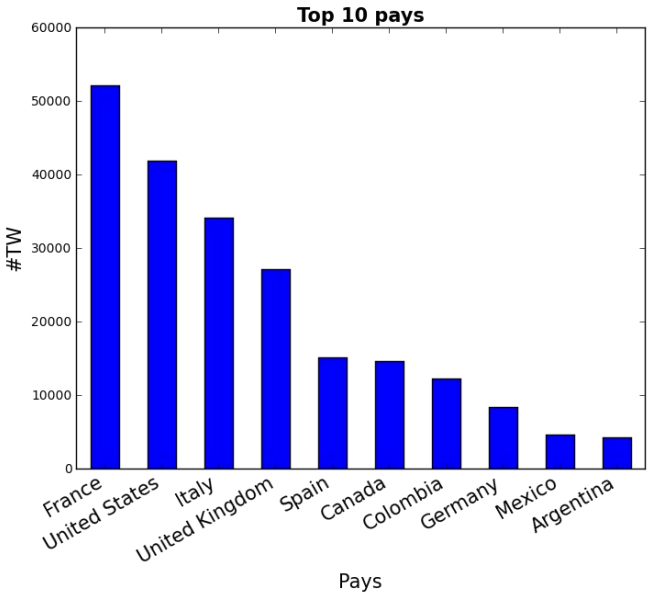


Figure 5 | Classement des 10 pays ayant émis le plus de tweets pendant la *#MuseumWeek* (2015).

Notamment, on voit à la figure 5 que la France serait le pays qui a soumis le plus de tweets durant la *#MuseumWeek* devant les États-Unis et le Royaume-Uni qui sont pourtant les 2 pays à avoir le plus grand nombre d'institutions inscrites à cette opération devant la France (avec respectivement 520, 442 et 358 institutions inscrites)[30].

De même, on peut par exemple s'interroger sur l'absence de la Russie et du Japon dans ce classement alors que ces 2 pays font partis des 10 pays avec le plus d'institutions inscrites à la *#MuseumWeek* (respectivement 127 et 74 institutions), loin devant la Colombie et l'Argentine qui apparaissent dans la seconde moitié de ce classement (avec respectivement 22 et 35 institutions inscrites).

Les résultats présentés au paragraphe suivant pourront être recoupés avec les résultats présentés à la figure 5 afin d'illustrer les propos que nous avons avancé jusqu'à maintenant.

[30] On rappelle que la liste des pays avec au moins un musée inscrit à la *#MuseumWeek*, accompagnés du nombre de musées inscrits pour chacun, se trouve en annexe à la section A4.

Langues des participants versus langues des tweets

Contrairement aux attributs que nous avons mentionnés précédemment pour identifier l'origine des participant, l'attribut (lang) Twitter qui encode le langage utilisé par les participants à l'aide d'un identifiant BCP 47[31], est présent pour chacun des tweets que nous avons collecté, et ce, toujours avec une valeur exploitable. Il en va de même pour l'attribut Twitter qui encode la langue dans laquelle un tweet est écrit. Il est important de préciser que les identifiants BCP 47 trouvés dans nos données codent presque toujours pour une langue générique plutôt qu'une langue spécifique (p. ex. *ar* pour *arabe* versus *ar-DZ* pour *algérien*).

Nous avons extrait séparément les valeurs BCP 47 associés aux attributs de langue des tweets et de leur auteurs (ou participants) et nous avons établi les classements de la figure 6. Le premier classement correspond aux Top_{10} des langues les plus utilisées par les participants de la *#MuseumWeek* (c.-à-d. la langue des tweets). Le second classement correspond à celui des langues les plus parlées par les participants (leur langue d'origine).

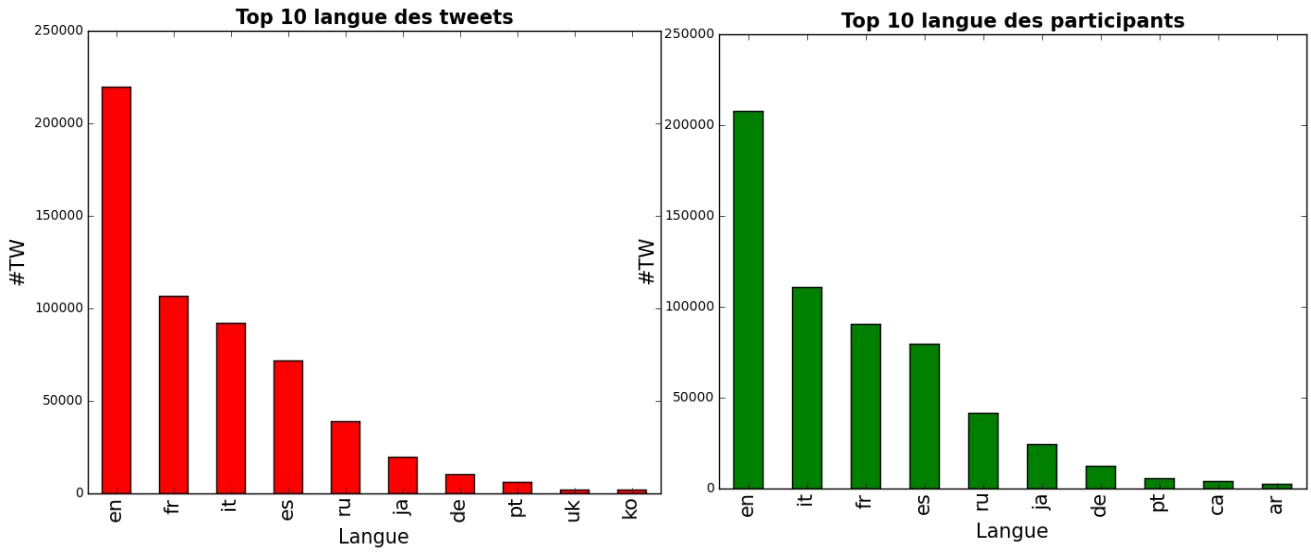


Figure 6 | Nombre de tweets émis durant la *#MuseumWeek* (#TW) selon la langue des tweets (à gauche) et celle des participants (à droite)[32].

On peut remarquer à la figure 6 que les 2 distributions représentées se ressemblent fortement. La majorité des tweets émis durant la *#MuseumWeek* sont donc écrits dans la langue que parlent les participants, l'anglais étant sans surprise la langue dominante dans un cas comme dans l'autre (avec respectivement 219 771 et 207 533 tweets).

[31] https://tools.ietf.org/html/bcp47

[32] La signification de chaque codes BCP 47 indiqués dans les 2 classements est : **en** : anglais, **fr** : français, **it** : italien, **es** : espagnol, **ru** : russe, **ja** ; japonais, **de** : allemand, **pt** : portugais, **uk** : ukrainien, **ko** : coréen, **ca** : catalan, **ar** : arabe.

En regardant plus attentivement les distributions de la figure 6, on se rend compte que certaines langues sont plus utilisées que d'autres par les participants à l'international. C'est notamment le cas du français qui à l'inverse de l'italien se place en 3ème position dans le Top_{10} des langues des participants avec moins de 100K tweets (c.-à-d. 96 575 tweets) alors qu'il se trouve 2ème dans l'autre Top_{10} avec plus de 100K tweets (c.-à-d. 106 583 tweets).

Par ailleurs, on peu noter que si l'arabe (toute distinction confondu) est présent dans le classement des langues les plus parlés par les participants (avec 2 426 tweets), aucune langue à dominante arabe n'apparaît dans le classement des langues les plus utilisées dans les tweets. A contrario, certaines langues absentes du premier classement apparaissent dans le second comme l'ukrainien et le coréen (avec respectivement 1 742 et 1 675 tweets).

De la même façon que pour l'origine des participants, on peut s'interroger sur le fait que le chinois n'apparaissent dans aucun des 2 classements présentés à la figure 6. On peut se demander si d'autres réseaux sociaux que Twitter, plus prisés en Chine (p. ex. Weibo[33]), ne leur aurait pas servi à échanger à propos de la *#MuseumWeek*.

La Francophonie, la France et le français

Nous avons vu qu'avec 106 583 tweets, le français est la 2ème langue la plus utilisée après l'anglais dans le corpus de tweets de la *#MuseumWeek*. Cette quantité de tweets représente environ 19% des tweets collectés durant la *#MuseumWeek* (37% pour l'anglais).

Ce total englobe l'ensemble des tweets français envoyés par des participants de toute nationalité durant la *#MuseumWeek*. Ce total correspond à celui trouvé à la ligne ALL, dans la colonne **xx-fr** (Monde+tweets en français) du tableau 3.

Le tableau 3 présente également le détail d'autres comptes attenants au français et/ou pour des participants francophones (colonnes **F-xx**, **f-xx** et **f-fr**). Notamment, nous avons défini une Francophonie au sens large ou Francophonie *étendue* (F-xx) et une Francophonie au sens strict ou Francophonie *stricte* (f-xx et f-fr). La Francophonie étendue inclus les pays francophones officiels[34] ainsi que des pays africains et arabes où l'on parle couramment français mais qui ne sont pas officiellement reconnus comme étant francophones. La Francophonie stricte inclus seulement les pays francophones officiels ; le tableau 4 donne le détail des pays francophones couverts par nos données avec, pour chaque pays concerné, le nombre d'institutions inscrites.

[33] https://fr.wikipedia.org/wiki/Sina_Weibo
[34] https://fr.wikipedia.org/wiki/Liste_des_pays_ayant_le_fran%C3%A7ais_pour_langue_officielle

Corpus	Échantillonnage (Sous-corpus)						
	Date	**xx-fr**	**F-xx**	**f-xx**	**f-fr**	**fr-fr$_1$**	**fr-fr$_2$**
	23 - 29 mars	Monde + Tweets en français métadonnées Twitter uniquement	Francophonie (étendue) + Tweets toute langue métadonnées Twitter uniquement	Francophonie (stricte) + Tweets toute langue métadonnées Twitter uniquement	Francophonie (stricte) + Tweets en français métadonnées Twitter uniquement	France + Tweets en français métadonnées Twitter uniquement	France + Tweets en français *institutions françaises* métadonnées Twitter et inscriptions
secretsMW	81 955	14 414	13 118	12 359	10 781	10 749	2 359
souvenirsMW	70 344	14 499	12 662	12 504	10 942	10 919	2 975
architectureMW	100 086	16 380	13 599	13 282	11 407	11 379	2 526
inspirationMW	69 313	13 020	10 934	10 774	8 867	8 859	2 453
familyMW	53 894	10 838	8 841	8 767	7 711	7 706	2 305
favMW	55 568	13 483	10 879	10 754	9 011	8 966	1 808
poseMW	43 326	7 767	7 334	7 213	5 380	5 346	1 275
MuseumWeek	371 179	68 576	60 739	59 293	50 227	50 211	11 375
ALL	**589 107**	106 583	93 007	90 575	76 466	76 291	**17 650**

Tableau 3 | Nombre de tweets par corpus de collecte en fonction de différents critères d'échantillonnage. **xx-fr** : tout participant, tweets en français. **F-xx** : participants francophones (Francophonie étendue), tweets toute langue. **f-xx** : participants francophones (Francophonie stricte), tweets toute langue. **f-fr** : participants francophones (Francophonie stricte), tweets en français. **fr-fr$_1$** : participants français (d'après les métadonnées Twitter), tweets en français. **fr-fr$_2$** : institutions françaises uniquement (basé sur les données d'inscription de la *#MuseumWeek*), tweets en français.

Francophonie étendue	#inscrits	Francophonie étendue	#inscrits
France	358	Bénin	1
Québec (Canada)	42 (218)	Maroc	1
Belgique	22	Tunisie	1
Luxembourg	7	**TOTAL (7)**	**432**

Tableau 4 | Liste des pays francophones (Francophonie étendue) de la *#MuseumWeek* (2015).

On peut voir au tableau 3 que le nombre de tweets émis par des participants francophones ne fait pas grande différence entre Francophonie étendue et stricte, quelque soit la langue des tweets (F-xx : 93 007 tweets versus f-xx : 90 575 tweets). Ceci s'explique par le nombre très restreint de pays africains et arabes de la Francophonie étendue qui soient présents dans les données (seulement le Bénin, le Maroc et la Tunisie comme indiqué au tableau 4).

Il en va de même entre le nombre total de tweets francophones écrits en français (f-fr : 76 466 tweets) et le nombre total de tweets en français émis par des français (**fr-fr$_1$** : 76 291 tweets). Cela suggère à première vue que la majorité des tweets écrits en français ont été émis par des français. Or, on a constaté que la plupart des institutions québécoises inscrites [35] se considèrent françaises plutôt que canadiennes ; le code langue indiqué dans leur biographique est le français (*fr*) et non le canadien (*fr-CA* ou *en-CA*). Le nombre d'institutions québécoises inscrites n'étant pas négligeable (42 comme indiqué au tableau 4), nous avons pris soin de retirer leur tweets de l'ensemble des tweets français émis par des institutions françaises d'après les métadonnées de Twitter. Nous avons aussi retiré les tweets français émis par des acteurs non institutionnels (c.-à-d. émis par des particuliers et non par des institutions) pour se faire une idée plus claire du taux de participation général des institutions françaises à la *#MuseumWeek*. Le résultat obtenu est indiqué dans la dernière colonne du tableau 3 (**fr-fr$_2$**). Le taux de participation à la *#MuseumWeek* des institutions françaises s'élève au final à 17 650 tweets soit 23% des tweets francophones (f-fr) et 3% des tweets monde (23 -29 mars)[36].

Remarque

Les critères de filtrage que nous avons utilisé pour obtenir les résultats de la colonne fr-fr$_2$ du tableau 3 ne s'appuient ni sur la langue des tweets, ni sur celles relatives à la biographie Twitter des institutions comme c'est le cas pour le reste des résultats de ce tableau. Dans ce cas, nous nous sommes appuyés sur le nom des institutions mentionnés dans les formulaires d'inscriptions à la *#MuseumWeek*. D'une part, cela nous a permis d'identifier avec certitude les tweets émis par des institutions françaises tout en évitant d'intégrer ceux émis par des institutions étrangères, par exemple des institutions québécoises. D'autre part, cela nous a permis de préserver les tweets d'institutions françaises que nous aurions écarté autrement, en raison de certaines erreurs de classification commises par Twitter pour déterminer la langue des institutions (cf. section A6 en annexe pour plus de détail).

[35] https://twitter.com/MCCQuebec/lists/museumweek-qu-bec/members

[36] Ces taux s'élèvent respectivement à 25% et 3,3%, si on considère l'ensemble des tweets émis par des institutions françaises, quelque soit la langue employée (19 541 tweets au total).

2.1.2 Identification des institutions françaises les plus actives

Dans cette section, nous avons cherché à identifier sur la base des tweets du corpus fr-fr$_2$ quelles institutions parmi les 358 institutions françaises inscrites à la *#MuseumWeek* étaient les plus actives.

Activité en terme de nombre de tweets

La figure 7 illustre la participation, en nombre de tweets émis par les institutions. On voit que la participation des institutions est plutôt homogène comme le confirme les histogrammes de la figure, même si on peut constater quelques singularités puisque l'émetteur le plus productif (Palais de la découverte ; 779 tweets) a émis presque deux fois plus de tweets que la seconde institution la plus productive (Archives Nationales de France ; 491 tweets).

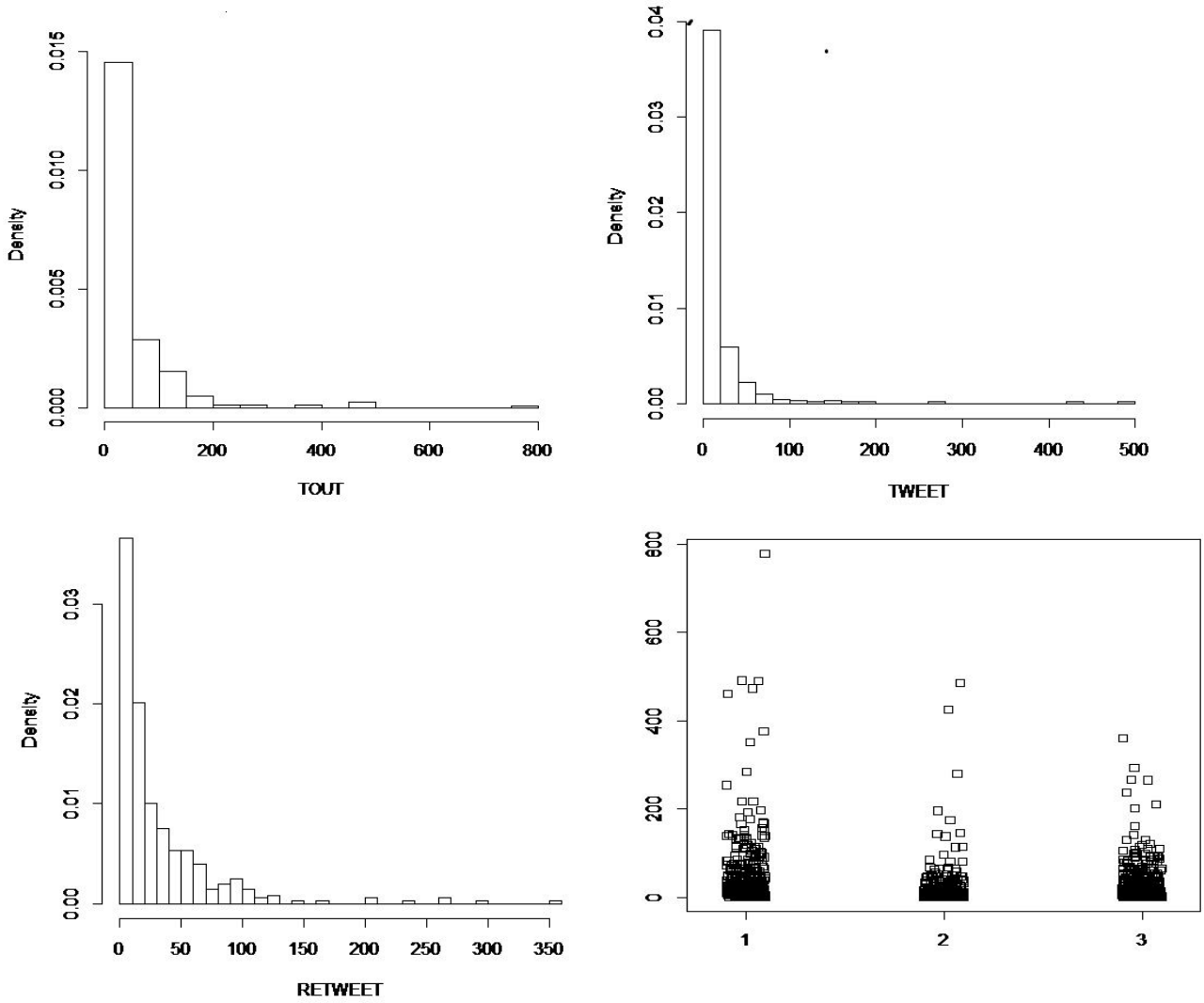


Figure 7 | Production de tweets (graphique en bas à droite) et histogrammes relatifs. 1 : dispersion tweets initiaux et retweets émis par institution (histogramme relatif : TOUT). 2 : dispersion tweets initiaux (histogramme relatif : TWEET). 3 : dispersion retweets (histogramme relatif : RETWEET).

Le tableau 5 (quantiles) et la figure 8 (boîte à moustaches) s'associent aux résultats de la figure précédente pour illustrer la répartition des différents types de tweets (TOUT, TWEET et RETWEET) et se faire une idée générale de la dispersion des distributions associées. Le tableau 7 donné plus loin détaille le nombre de tweets du corpus fr-fr$_2$ par type de tweets.

Quantiles	*0%*	*25%*	*50%*	*75%*	*100 %*
TOUT	0	8	24	58	779
TWEETS	0	1	5	15	486
RETWEET	0	6	16	41	360

Tableau 5 | Quantiles relatifs aux dispersions de tweets de la figure 7.

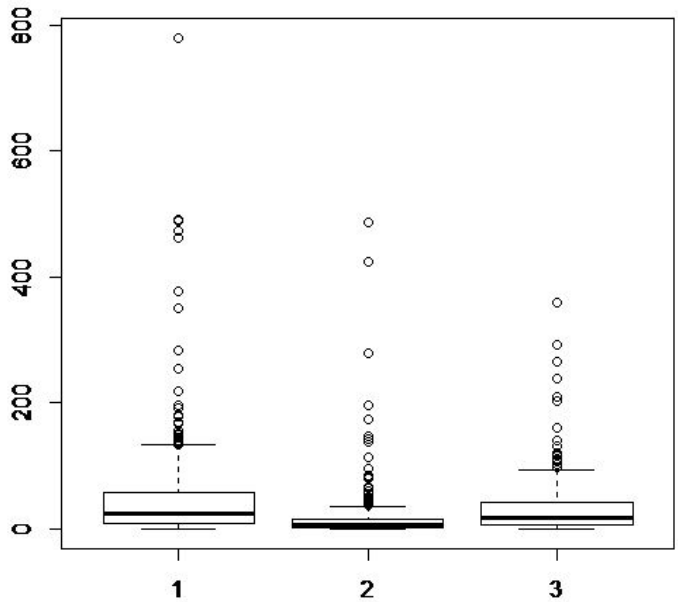


Figure 8 | Boîtes à moustache relatives aux dispersions de tweets de la figure 7.

Institutions meneuses versus institutions relayeuses

Comme pour la *#MuseumWeek* 2014, nous avons calculé pour chaque institution inscrite à la *#MuseumWeek* cette année, 2 indicateurs de participation (IP) :

- **IPL** = N_TW / NT_TW ;
- **IPS** = N_RT / NT_RT, avec :
    - N_TW : nombre de tweets initiaux émis par l'institution ;
    - N_RT : nombre de retweets émis par l'institution ;
    - NT_TW : nombre de tweets initiaux total (toutes institutions confondues) ;
    - NT_RT : nombre de retweets total (toutes institutions confondues).

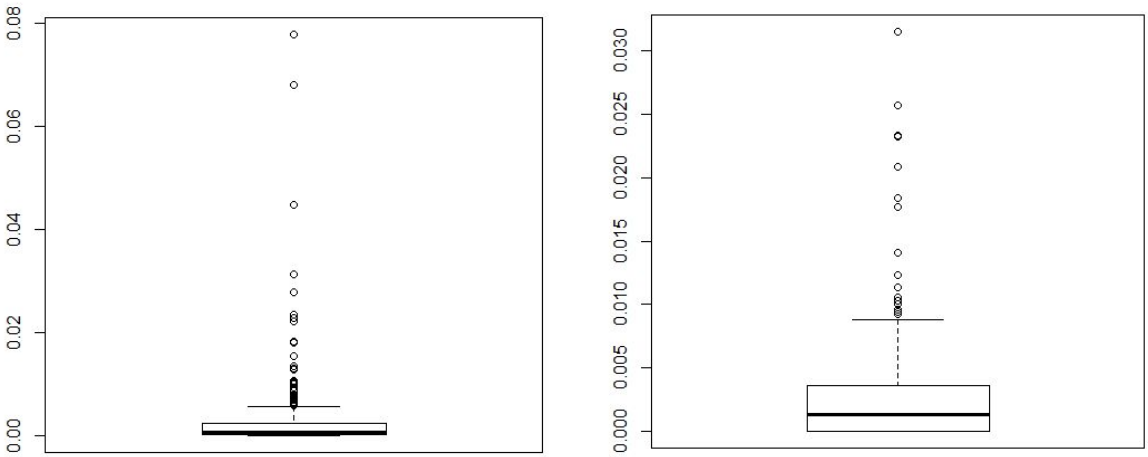


Figure 9 | Boîtes à moustache relative aux indicateurs IPL (à gauche) et IPS (à droite).

Le premier indicateur, **IPL** (indicateur de participation en tant que *lanceur*), permet de définir dans quelle mesure une institution joue un rôle de meneur via les tweets initiaux (c.-à-d. si elle produit du contenu original). Le second indicateur, **IPS** (indicateur de participation en tant que *suiveur*), permet de définir dans quelle mesure une institution joue un rôle de relayeur par l'entremise des retweets (c.-à-d. si elle produit du contenu de reprise).

La moyenne des indicateurs IPL et IPS est de 0,00297, suggérant une parité meneur/suiveur entre les différentes institutions. Néanmoins, comme on le voit d'après la figure 9, si la majorité des meneurs comme des relayeurs suggère une similitude dans le comportement des institutions (les valeurs IPS et IPL sont resserrées), la plus grande disparité des valeurs IPS, et ce, malgré un taux d'outliers plus faible, suggère que le comportement des suiveurs est moins homogène que celui des meneurs.

2.2 Analyses catégorielles

Nous présentons dans cette section les résultats d'analyses catégorielles. Ces analyses sont complémentaires des analyses quantitatives qui précèdent au sens où elles offrent une compréhension plus fine du comportement des institutions. Ces analyses portent exclusivement sur les tweets du corpus **fr-fr$_2$**, c'est-à-dire les tweets émis en français par les 358 institutions françaises ayant participées à la *#MuseumWeek*.

Cette section s'articule en 2 parties. La section 2.2.1 traite de la méthodologie de classification que nous avons suivie pour catégoriser les tweets du corpus fr-fr$_2$. La section 2.2.2 présente les résultats de classification de ces tweets et les résultats de classification des institutions associées, globalement et par institution.

2.2.1 Méthodologie de classification des tweets

Le comportement des institutions est définie à partir de la classification automatiquement de leur tweets en 4 catégories prédéfinies. Cette classification est réalisée à l'aide d'un classifieur, un modèle statistique obtenu par apprentissage (ou *machine learning*) sur la base d'exemples de référence qui, dans notre cas, prennent la forme de tweets annotés à la main par des experts du domaine culturel en fonction des catégories prédéfinies.

Les approches par apprentissage sont communément utilisées en TAL dans des tâches de classification de document textuel (p. ex. pages web). Elles ont récemment été utilisé avec succès dans des tâches de classification de tweets. Par exemple en analyse de sentiment, le but étant de définir automatiquement la valence des tweets (positif, négatif ou neutre) et/ou leur intensité (p. ex. fortement négatif versus faiblement négatif).

Notre modèle permet de catégoriser les (re)tweets événementiels ou institutionnel attenants au domaine de la culture comme ceux de la *#MuseumWeek*, émis par une institution ou un individu. Plus précisément, notre modèle classe les tweets en catégories de type communicationnel sur la base du contenu textuel des tweets. Nous utilisons des traits linguistiques (majoritairement lexicaux, mais aussi les marques de ponctuation), des traits spécifiques aux tweets (p. ex. présence/absence de hashtags dans les tweets) ainsi que certaines métadonnées comme l'identité des auteurs des tweets.

Nous donnons au tableau 6, la taxonomie de tweets que nous avons mis en place spécifiquement pour la *#MuseumWeek* l'an dernier avec un spécialiste en communication de l'information et les avis de community managers de musées parisiens. Celle-ci a été affinée et réutilisée cette année pour classer les tweets du corpus fr-fr$_2$. Elle nous a aussi servie cette année au cours d'un atelier ministériel donné dans le cadre de la CMMin[37] du MCC au Centre Georges Pompidou pour faire annoter différents corpus de tweets à 16 des community managers présents. Les institutions culturelles représentés provenaient de différents horizons (Paris, région, musées, archives, théâtres, public et privé). Les corpus ainsi annotés nous ont servi d'exemples pour l'apprentissage du classifieur.

La présentation détaillée du classifieur, de ces données et traits d'apprentissage ainsi que son évaluation devrait faire l'objet d'une publication de recherche à la rentrée dans le cadre des 8ème Journées internationales de Linguistique de Corpus[38]. Le résumé de sélection soumis pour cette publication est fourni dans l'archive qui accompagne ce livrable.

[37] http://cmmin.fr
[38] http://jlc2015.sciencesconf.org.

Catégorie principale	Sous catégorie	**Définition**	*Exemple*
1 Encourager la participation (IRL[39]/on-line)	**Encourager la contribution on-line** (a)	Inviter les twittos à participer à une opération sur Twitter (emploi de hashtag, quizz, etc.)	*Attention ! Plus qu'1h pour répondre à notre question #OùestSaintLouis et remporter le lot spécial #jeu #concours*
	Encourager la contribution IRL (b)	Inviter les twittos à venir sur place pour participer à une activité, exposition, etc.	*Demain, dernier jour pour profiter de l'exposition #Matisse au @CentrePompidou ! http://t.co/sWeFkPiR #art #Paris*
2 Partager son expérience, avis, sentiment	**Partager une expérience** (a)	Partager son avis, une émotion (coup de coeur/gueule)	*Bravo pour avoir déniché cette oeuvre superbe ;) RT @axellemaemi: Joueur de luth, musée du Moyen - Age. Cluny http://t.co/BorpourquoiFfIB1*
	Remercier (b)	Remercier/exprimer sa gratitude envers un ou plusieurs émetteurs	*Merci à tous les participants au #betatest de #Blinkster @CentrePompidou qui sont restés tard pour nous donner leur avis. http://ow.ly/6Maqq*
	Commenter un tweet (c)	Ajouter un contenu à un tweet	*Participez aux Impromptus, atelier en famille à 15h ! MT @Babetchou : après-midi au @centrepompidou (photo @AnnSlls) http://t.co/mfbxyq0lR3*
3 Interagir avec sa communauté	**Répondre à un compte** (a)	Répondre à une question	*@sebtik non malheureusement il ne fait pas partie des œuvres exposées actuellement*
	Interpeller un compte (b)	Mentionner un compte en l'interpellant, en lui faisant un clin d'oeil, en lui posant une question, etc.	*Tiens, tiens, les #lolcats du Moyen Age sont de retour, cette fois @britishlibrary http://t.co/OkMtzeSN cc @v_septembre*
4 Promouvoir-informer sur l'institution (activités ou collections)	**Informer sur les collections** (a)	Diffuser une information concernant les oeuvres du musée	*La "Roue de bicyclette" de Marcel #Duchamp et quoi d'autre? Les œuvres qui seront exposées au #CPmobile à #Libourne: http://t.co/tp9g13Ad*
	Promouvoir le musée émetteur du tweet (b)	Informer et promouvoir les activités et informations pratiques du musée	*Nous ouvrons l'accès à l'expo #Dali les dimanches dès 9h30 du matin pour les visiteurs munis de billets et adhérents http://t.co/daYh8D79*
	Promouvoir un autre musée (c)	Informer et promouvoir les activités et informations pratiques des autres musées	*#ArtCetEte Une selection d'exposition à visiter à l'étranger cet été [ESP] http://t.co/ncbsGkVh vía @TerraOcioES @museoguggenheim*

Tableau 6 | Typologie utilisée pour la classification des tweets en catégorie communicationnelle.

[39] In Real Life.

2.2.2 Classification des tweets et des institutions françaises

Dans cette section, nous présentons les résultats de classification des tweets du corpus fr-fr_2 et des 358 institutions françaises associées, globalement et par institution. Concernant les institutions, il est à noter que nous utiliserons de manière interchangeable les termes *classification* et *comportement*.

Classification globale

On présente au tableau 7 le nombre de tweets et de retweets émis par les 358 institutions françaises inscrites à la *#MuseumWeek* en fonction des deux types de tweets suivants :

- Les tweets initiaux (ou proto-tweets) ;
- Les retweets.

On voit d'après les résultats de ce tableau que les institutions ont majoritairement émis des tweets (TWEET : 65% versus RETWEET : 35%) et que le nombre moyen de tweets émis par institution est presque 2 fois supérieur à celui des retweets (TWEET : 32 versus RETWEET : 17). Ceci nous permet de quantifier les analyses faites précédemment en section 2.1.2 à propos de l'activité des institutions sur Twitter au cours de la *#MuseumWeek* où nous avions déjà pu observer que le nombre de tweets initiaux émis par les institutions françaises semblait plus important que celui des retweets.

fr-fr_2	#tweet	%TOUT	#tweet moyen par institution
TOUT	**17 650**	100	49
RETWEET	6 232	35	17
TWEET	11 418	65	32

Tableau 7 | Nombre de tweet (#tweet) et pourcentage associé (%TOUT) tout confondu (TOUT), seulement pour les retweets (RETWEET) et pour les tweets initiaux (TWEET) du corpus fr-fr_2 (c.-à-d. les tweets français émis par les 358 institutions françaises durant la *#MuseumWeek* en 2015).

On présente au tableau suivant (le tableau 8) les résultats de classification des tweets du corpus fr-fr_2 en lien avec les 2 catégories de tweets considérés précédemment. On peut voir d'après ce tableau que seulement 2% des tweets n'ont pas réussi à être classé selon notre approche ce qui représente un total de 459 tweets sur les 17 650 catégorisés, les tweets étant légèrement plus difficile que les retweets à classer (113 retweets non classés versus 346 proto-tweets non classés). Cela montre que notre classifieur a une bonne couverture dans son domaine.

fr-fr$_2$	TOUT		RETWEET (RT)		TWEET (TW)	
	#tweet	%TOUT	**#tweet**	%RT	**#tweet**	%TW
Promouvoir	4 867	28	436	7	4 431	39
Interagir	655	4	341	5	314	3
Encourager	2 227	13	227	4	2 000	17
Partager	9 442	53	5 115	82	4 327	38
Autre	459	2	113	2	346	3
Total	**17 650**	100	**6 232**	100	**11 418**	100

Tableau 8 | Détail des résultats présentés au tableau 7 par catégorie de classification des tweets. La catégorie *Autre* correspond aux tweets pour lesquels notre méthode de classification est en échec.

On voit également dans ce tableau que la distribution des tweets émis par les institutions n'est pas homogène, tout type de tweet considéré : la catégorie dominante dans le corpus est la catégorie Partager son expérience, un avis, un sentiment (que nous renommerons *Partager* dans la suite de cette section afin d'alléger la lecture) qui comptabilise la moitié des données classées (53% exactement). La seconde catégorie dominante est la catégorie Promouvoir-informer sur l'institution à propos de ces activités ou de ses collections (désormais *Promouvoir*) qui comptabilise 28% des données classés. Enfin, viennent les catégories Encourager la participation IRL ou on-line et Interagir avec sa communauté (désormais *Encourager* et *Interagir*) qui comptabilisent 13% et 4% des tweets.

De ces résultats, on pourrait conclure que les institutions ont été plus enclines à partager leurs propres expériences ainsi que promouvoir et/ou informer les publics à propos de leur établissement et/ou d'autres établissements plutôt que de les encourager à participer et interagir avec elles dans le cadre de la *#MuseumWeek*. Néanmoins, si on considère le type de tweets émis, on s'aperçoit alors que dans la majorité (soit 9942 tweets), classés dans la catégorie *Partager*, sont des retweets (82% d'entre eux). Or, les retweets issus de cette catégorie sont, la plupart du temps, des tweets mettant en avant l'expérience d'un individu plutôt que celle d'une institution (p. ex. son appréciation à propos de la visite d'un musée). Il s'avère donc que le contenu réellement produit par les institutions durant la *#MuseumWeek* se reflète plus du côté des proto-tweets que des retweets. Sous cet angle, on constate que les catégories de tweets dominantes dans le corpus fr-fr$_2$ sont : (i) *Promouvoir* et (ii) *Partager* (avec presque 40% de couverture chacune). Viennent ensuite, les catégories *Encourager* (couverture : 17%) et *Interagir* (couverture : 3%).

Une telle répartition pourrait s'expliquer par le caractère relativement *story telling*[40] et *do it yourself*[41] de la *#MuseumWeek* cette année qui ne va pas dans le sens de l'échange et de l'interaction comme c'était le cas l'an dernier avec des thématiques telles que *QuizzMW* et *QuestionMW* où les publics étaient amener à engager la conversation avec les community manager des musées participants.

Dans la suite de cette section, nous avons délaisser l'analyse des retweets, cantonnés à une seule catégorie de classification, au profit de l'analyse approfondie des proto-tweets émis par les institutions françaises (soit 11 418 tweets), dans l'idée de mieux cerner leur comportement durant la *#MuseumWeek*.

Comportements généraux des institutions

La figure 10 donnée plus bas illustre le comportement global des 358 institutions françaises ayant participé à la *#MuseumWeek* obtenu par classification de leur tweets (retweets exclus) à travers les 4 catégories principales définies dans notre taxonomie de tweets. Nous avons regroupé ces institutions, de façon arbitraire, en 5 groupes, relativement au nombre de tweets qu'elles ont émis. Le groupes sont les suivants :

1. Les institutions ayant émis plus de 100 tweets;
2. Celles ayant émis entre 50 et 100 tweets ;
3. Celles ayant émis entre 25 et 50 tweets ;
4. Celles ayant émis entre 10 et 25 tweets ;
5. Celles ayant émis moins de 10 tweets (celle n'ayant émis aucun tweet incluses).

La répartition des institutions dans ces groupes et leur couverture en terme de tweets est :

Groupe	#institution	Couverture (%)
1 > 100 tweets	19	29
2 [50 ; 100] tweets	54	33
3 [25 ; 50 [tweets	63	20
4 [10 ; 25 [tweets	103	14
5 [0 ; 10 [tweets	119	4

[40] J1 (#secretsMW) : découverte du quotidien des musées par le grand public avec des anecdotes et secrets racontées par les community manager des institutions participantes. J2 (#souvenirsMW) : partage de souvenirs de visites par le grand public ; J3 (#architectureMW) : découverte de l'histoire des musées ; J6 (#favMW) : incitation des musées à partager les photos des visiteurs, leur coup de coeur.

[41] Avec les deux journées consacrées à la créativité numérique et, entre autres, aux selfies (J4 : #inspirationMW et J7 : #poseMW) ; on rappelle que la définition des hashtags thématiques officiels de la *#MuseumWeek* se trouve sur Internet en suivant ce lien : http://museumweek2015.org/fr/programme.

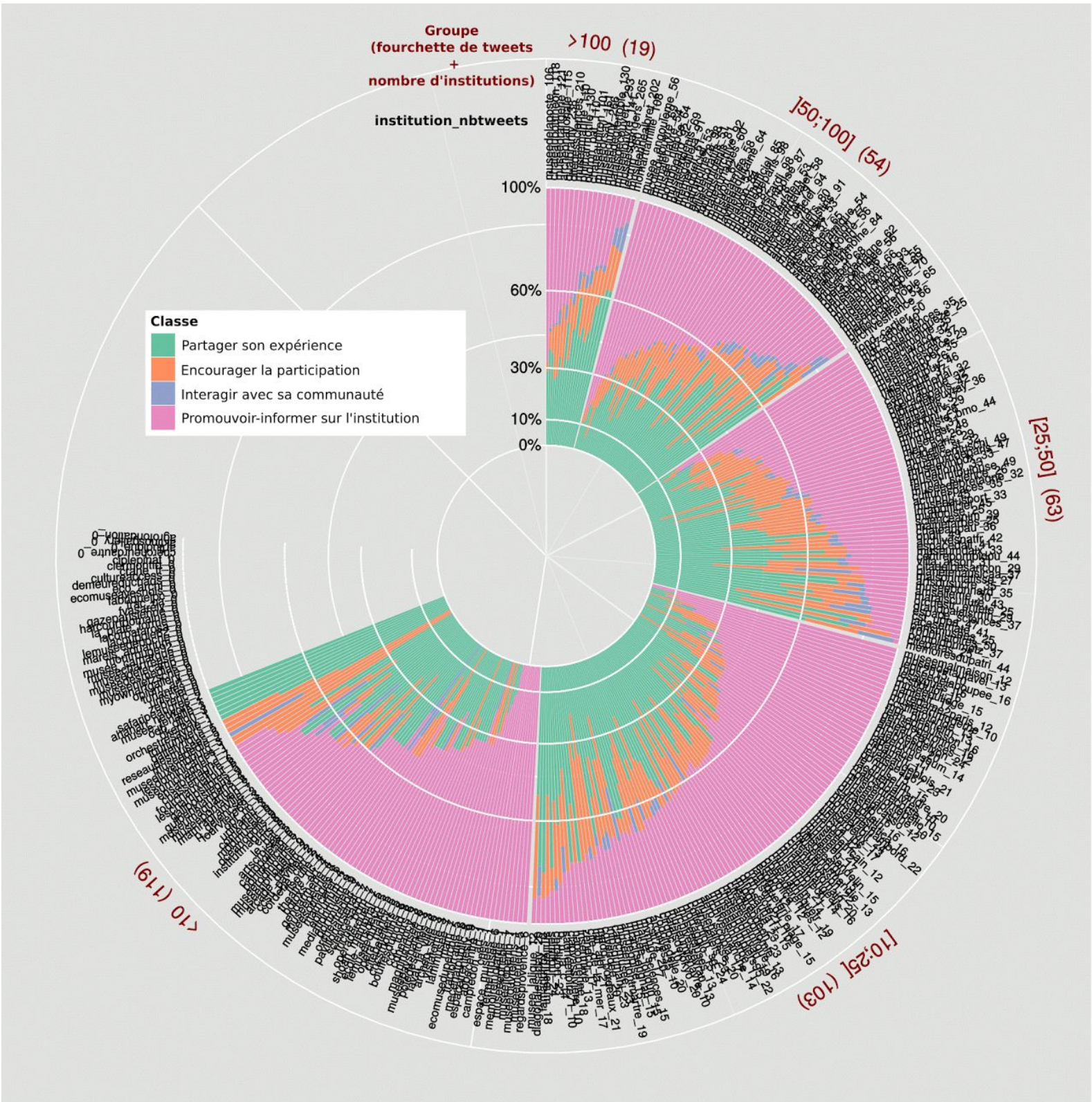


Figure 10 | Classification globale des 358 institutions françaises inscrites à la *#MuseumWeek* (2015), basée sur le contenu communicationnel des tweets émis par chacune d'entre elle durant l'événement (retweets exclus).

La figure 10 nous renseigne sur le comportement global des institutions : **plus une institution participe, moins elle centre sa communication sur elle-même** (catégorie *Promouvoir*) et plus elle joue le jeu de la *#MuseumWeek*. Ce constat est validé par les chiffres indiqué aux tableau 9 qui présente le nombre moyen de tweets émis par l'ensemble des 358 institutions françaises ayant participées à la *#MuseumWeek* par catégorie de tweet en fonction des différents groupes d'institution. Il est plus clairement visible au niveau de la figure A7.1 donnée en annexe (cf. section A7) qui présente la même visualisation que la figure 10 ci-dessus mais uniquement pour les 3 premiers groupes. À titre de comparaison, nous donnons à la figure A7.2 (section A7 également) la classification globale des 103 institutions françaises qui ont participé à la *#MuseumWeek* l'année dernière.

On voit également d'après la figure 10 que seules les institutions ayant émis plus de 25 tweets (c.-à-d. les institutions des 3 premiers groupes) interagissent avec leur communauté (moins de 4% des tweets seulement). On s'aperçoit aussi qu'un certains nombre d'institutions ce sont inscrites à la MuseumWeek, mais n'ont pas participé (cf. groupe 5 sur la figure). Ces dernières sont listées à la fin de l'annexe de ce rapport (en section A7 toujours). Nous avons dénombré 32 institutions dans ce cas précis.

Groupe	*Promouvoir*	*Interagir*	*Encourager*	*Partager*
1 > 100 tweets	35,27	3,75	15,86	41,93
2 [50 ; 100] tweets	42,22	2,24	17,64	35,27
3 [25 ; 50 [tweets	35,66	2,68	19,85	38,01
4 [10 ; 25 [tweets	45,00	1,62	16,19	34,72
5 [0 ; 10 [tweets	48,74	2,79	11,74	33,62

Tableau 9 | Pourcentage moyens de tweets émis par l'ensemble des 358 institutions françaises ayant participées à la *#MuseumWeek* (2015) par catégorie et par groupe de classification[42].

Nous avons illustré à la figure 11 les résultats indiqué dans le tableau 9 sous la forme de représentation en étoile (ou *représentation radar*) en prenant le même jeu de couleur que celui choisi pour la représentation radiale de la figure 10. Nous avons sorti du groupe 5 les 32 institutions qui n'ont émis aucun tweet. Les catégories ont été abréviées dans la légende de la figure (**Pro.** : *Promouvoir*, **Int.** : *Interagir*, **Enc.** : *Encourager* et **Par.** : *Partager*).

[42] La somme des lignes indiquées dans ce tableau est inférieur à 100% puisque les valeurs de colonnes ne sont des valeurs exactes mais des moyennes.

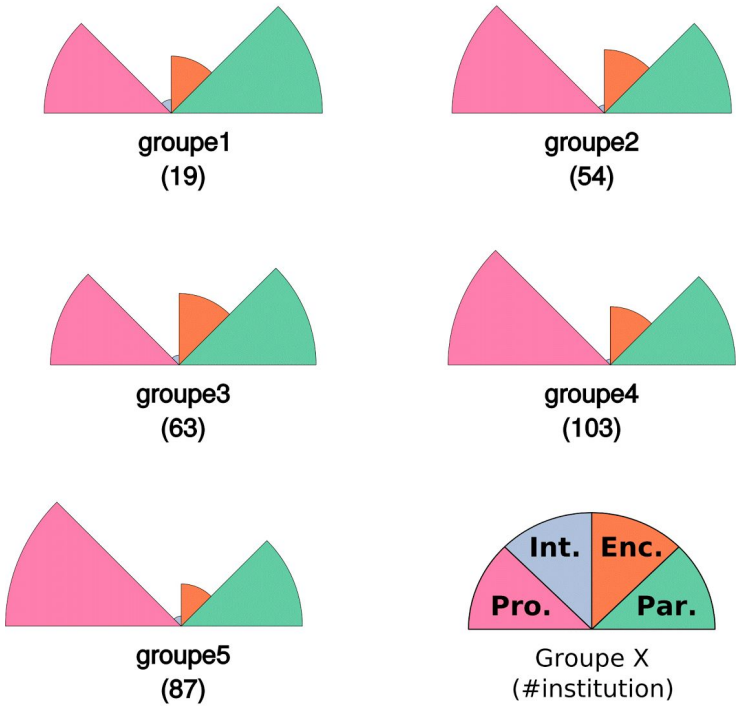


Figure 11 | Comportements communicationnels moyens des 358 institutions françaises ayant participées à la *#MuseumWeek* (2015).

Le constat général qui ressort de cette section comme l'illustre aussi la figure 11 est que le comportement des différentes institutions est globalement le même : les patterns communicationnel obtenus sont relativement proches. Mais il ne faut pas oublier que ces derniers correspondent à des comportements moyens et que, dans la réalité, les institutions qui ont participé à la *#MuseumWeek* ne ce sont pas toutes comportées à l'identique, ainsi que l'attestent les patterns communicationnels individuels des 19 institutions du groupe 1 illustrés à la figure 12 en fin de section.

Comportements individuels des 19 institutions du groupe 1 (>100 tweets)

Figure 12 | Comportement communicationnel des 19 institutions françaises ayant émis le plus de tweets durant la *#MuseumWeek* (2015), basée sur le contenu des tweets émis par chacune d'elle durant l'événement (retweets exclus).

Les comportement individuels des institutions des autres groupes communicationnels (cf. tableau 9 et figure 11) sont donnés sous la même forme qu'à la figure 12 en annexe de ce livrable (cf. section A7, figures A7.3 à A7.6 précisément).

2.3 Analyse factorielle : analyse en composantes principales

À titre exploratoire, nous avons réalisé une Analyse en Composantes Principales (ou ACP) sur un échantillon composé des 20 institutions ayant émis le plus de tweets (retweets inclus) au cours de la *#MuseumWeek* 2015. Cet échantillon représente environ 33% du volume total de tweets émis par l'ensemble des institutions françaises (corpus fr-fr$_2$).

Voici les 7 variables que nous avons choisi pour réaliser l'ACP[43] :

- Nombre de tweets (émis par les institutions) ;
- Nombre de retweets (ibidem) ;
- Nombre de tweets non catégorisés lors de la classification ;
- Nombre de tweets catégorisés **Partager son expérience** ;
- Nombre de tweets catégorisés **Encourager la participation** ;
- Nombre de tweets catégorisés **Interagir avec sa communauté** ;
- Nombre de tweets catégorisés **Promouvoir-informer sur l'institution**.

	TW	RWT	NON_CATEGO	PARTAGER	ENCOURAGER	INTERAGIR	PROMOUVOIR
PALAIS_DECOUV	486	293	5	379	16	26	60
ARCHIV_FR	425	66	14	364	12	25	10
CITE_SCIENCES	279	210	4	213	18	13	31
LECMM	113	360	0	101	3	2	7
MUS_ANGERS	196	265	5	151	5	13	22
MqB	138	238	1	111	4	6	16
MUS_CLUNY	85	266	2	67	1	7	8
MUS_LOUVRE	143	141	0	123	0	1	19
MOMART_FAM	146	108	3	122	9	6	6
MEMOIR_PATR	174	44	3	156	4	8	3
MUS_J_ALBRET	16	202	0	12	1	2	1
MUS_GRENOBLE	67	130	0	55	2	6	4
MUS_ORSAY	31	161	1	24	1	1	4
CHATE_OIRON	80	101	0	70	4	5	1
CHAT_BOUTEON	59	118	2	44	5	4	4
MUS_LAPOSTE	63	106	0	56	1	6	0
PHILARMONIE	37	130	1	28	2	3	3
MUSM_GRENOBLE	96	70	3	80	3	7	3
MUS_ORANGERIE	35	121	0	28	2	0	5
MUS_BALYON	67	84	1	55	4	6	1

Figure 13 | Données de la *#MuseumWeek* (2015) utilisées pour l'ACP (TW : tweets, RWT : retweets).

Les données d'ACP sont présentées à la figure 13. Le résultat de l'ACP est présenté à travers les figures 14, 15 et 16 ci-après.

On constate à la figure 14 que les 2 premières dimensions de l'ACP couvrent près de 89% des données analysées, elles sont donc très significatives. Les valeurs obtenus selon ses dimensions pour chacune des variables de l'ACP se trouvent en fin de section à la figure 16.

[43] Les catégorie utilisées ici sont celles définies aux section précédentes, nommées différemment.

Cette figure (ou *carte de facteur*) indique le rôle des 7 variables d'analyse en fonction des 2 dimensions calculées. On peut voir que les variables ne sont pas négativement corrélées et sont mêmes très proches pour *Encourager*, *Partager* et *Interagir*.

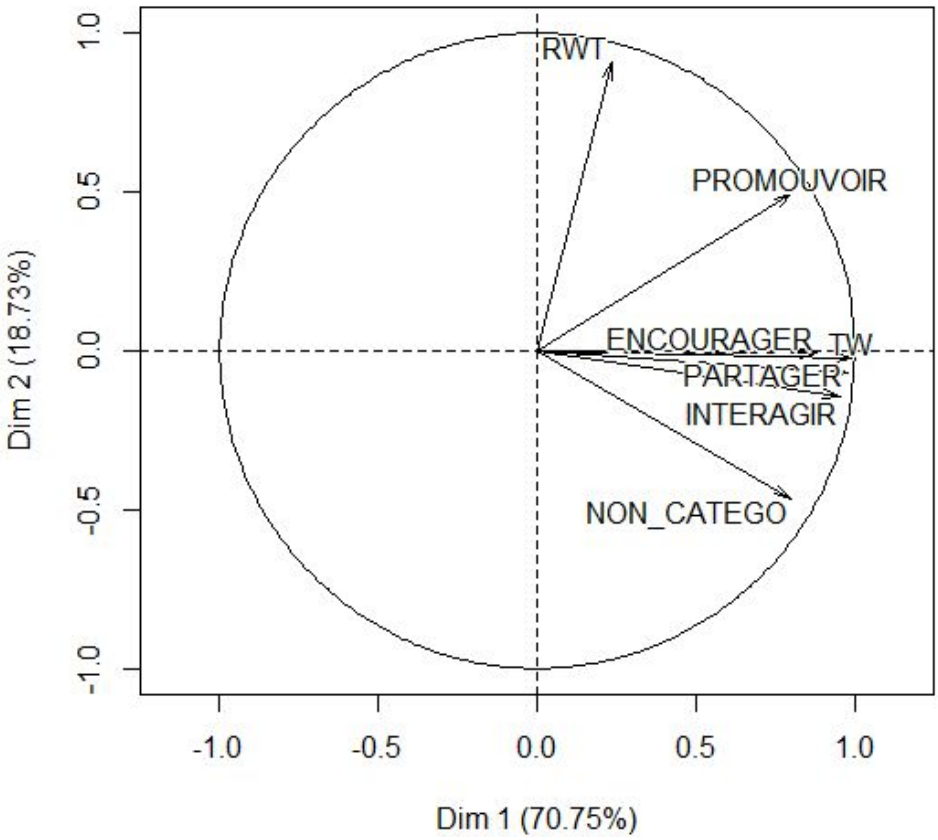


Figure 14 | Carte de facteur (ACP).

D'après la figure 15, trois institutions, les Archives Nationales de France (ARCHIV_FR), le Palais de la Découverte (PALAIS_DECOUV) et le CMN (LECMM), se détachent des 17 autres.

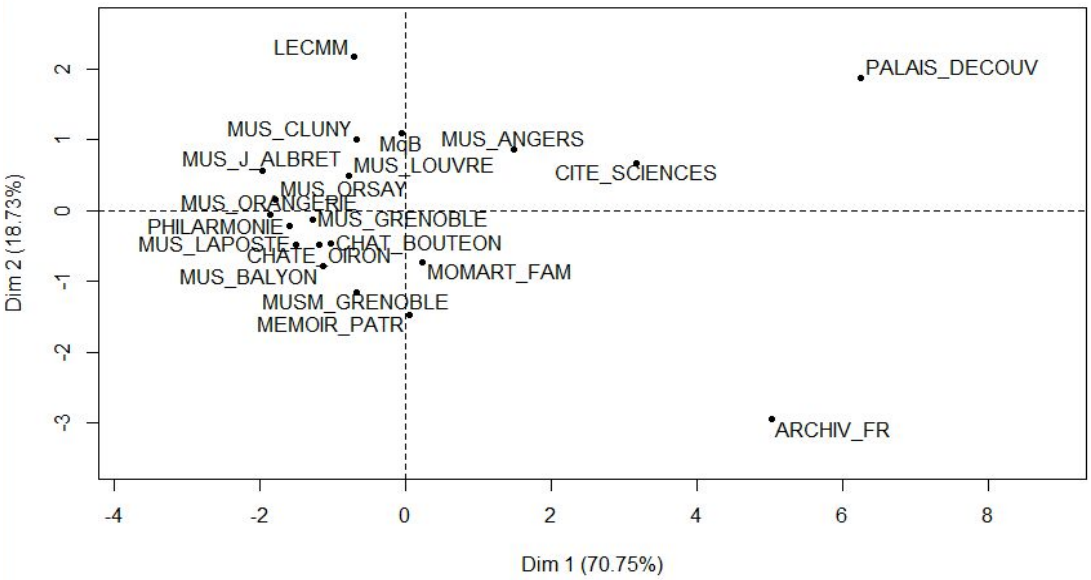


Figure 15 | Clustering factoriel (ACP).

	Dim.1	Dim.2
TW	19.745265	0.036473179
RWT	1.086882	62.992916411
NON_CATEGO	12.896294	16.445185096
PARTAGER	19.297718	0.379013246
ENCOURAGER	15.672369	0.004356541
INTERAGIR	18.358834	1.585151058
PROMOUVOIR	12.942639	18.556904470

Figure 16 | Poids des différentes variables (tweets, retweets, etc.) utilisées pour calculer l'ACP, en fonction des dimensions 1 et 2 (Dim.1 et Dim.2 respectivement).

Au final, et sous réserve d'une analyse plus approfondie, on constate d'après l'analyses factorielle que nous avons mené que les pratiques de communication des institutions françaises les plus actives pendant la *#MuseumWeek* 2015 sont, contrairement à l'an passé, très proches (la majorité des tweets ayant principalement invités les twetto à partager leur expérience) malgré des différences institutionnelles marquées dans certains cas (p. ex. Archives Nationales de France, le Palais de la Découverte et le CMN)[44].

[44] Le comportement communicationnel de ces 3 institutions est visible à la figure 12 donnée à la fin de la section précédente. On voit également que dans ce cas, le comportement de ces 3 institutions est très différents, le comportement du Palais de la Découverte se situant à la frontière entre celui du CMN et des Archives Nationale de France, de la même manière que l'indique le clustering factoriel présentée à la figure 15 dans cette section.

3 Conclusion

Le premier constat que l'on puisse faire tient à la très grande différence, en terme de participation des institutions, entre les deux éditions de la *#MuseumWeek* (2014 versus 2015). Nous avons montré à la section 1.2 le double contexte, international et national, de l'évènement en 2015.

Le deuxième constat que l'on puisse faire tient au manque de fiabilité des métadonnées fournies par les API de Twitter et notamment celles relatives à la langue des tweets et au pays d'origine des institutions (attributs `langue` et `country` notamment). Ceci nous a obligé à effectuer de nombreux pré-traitements supplémentaires avant de pouvoir réaliser des analyses quantitatives rigoureuses. On peut regretter à ce propos le manque de coopération de la par de Twitter France, co-organisateur de l'évènement avec le MCC.

Le troisième constat que l'on puisse faire concerne la grande disparité des pratiques des community managers en charge de la politique de communication sur Twitter pour couvrir l'évènement, notamment en nombre de tweets émis, et surtout dans l'usage des retweets. Ces différences de pratiques qui reflètent une conception de l'auctorialité très différente d'un community manager à l'autre étaient déjà présentes en 2014.

Le quatrième constat que l'on puisse faire concerne la grande disparité qu'il y a entre les 20 institutions les plus productrices de tweets dont le contenu relève essentiellement de la catégorie *Partager* (cf. section 2.3) et le reste des 358 institutions participantes dont le contenu des tweets relève majoritairement de la catégorie *Promouvoir* (cf. section 2.2).

Enfin, le dernier constat que l'on puisse faire, issu des résultats d'analyse factorielle (ACP) donnés à la section 2.3, révèle trois institutions singulières quand à la volumétrie des tweets émis et de leurs contenus. Ces dernières sont : les Archives Nationales de France, le Palais de la Découverte et le CMN.

Annexes

A1 Informations demandées à Twitter pour collecter et préparer les données

Dans cet annexe, nous présentons les différentes demandes/remarques (9 au total) que nous avons faites à Twitter France pour nous aider dans la collecte et le pré-traitement des données de la *#MuseumWeek* 2015.

Chacune demande/remarque est accompagnée des échanges associés entre le MoDyCo et Twitter. Les différentes demandes/remarques sont séparées les unes des autres dans le texte à l'aide d'un bandeau jaune comme on peut le voir pour les 2 premières demandes sur cette page.

- **DEMANDE 1**
  - MoDyCo :
    - **Peut-on récupérer toutes les données au format cvs et json ?**
  - Twitter :
    - Les données sont au format JSON.
- **DEMANDE 2**
  - MoDyCo :
    - **Peut-on récupérer 1 corpus par jour ou par hashtag (si les hashtags FR et internationaux sont différents) avec en plus 1 colonne avec la provenance (le code pays) du tweet ?**
  - Twitter :
    - Nous ne prévoyons pas de faire ce type d'extraction. Vous pouvez utiliser nos API pour traquer les Tweets comportant les 8 hashtags : https://dev.twitter.com/streaming/overview. En cas de problème, on pourra vous mettre en relation avec l'un de nos developer advocates.
  - MoDyCo :
    - Ok mais d'un point de vu général, nous risquons d'être confronter à ces limites (http://goo.gl/uLXxZd).
  - Twitter :
    - Il faut utiliser la streaming API : vous pouvez accéder à 1% du firehose, soit 5M Tweets/day! (rappel : la MuseumWeek en 2014 = 260 000 Tweets).
  - MoDyCo :
    - Nous parlions en général (voir également demande 3).

- **DEMANDE 3**
  - MoDyCo :
    - **Pourrions-nous récupérer la liste de tous les comptes participants à l'opération avec pour chacun de ces comptes, la biographie et les 4 données chiffrées suivantes (à partir du 23 mars) : le nombre de tweets, d'abonnés, de followers pour chaque compte ainsi que son pays ?**
  - Twitter :
    - Sous réserve d'avoir la liste des comptes en question à partir du formulaire du blog avant le 23 mars. Oui pour le nombre de Tweets, abonnements, followers et PEUT ETRE pour le pays.
  - MoDyCo :
    - Il s'agit d'avoir ces infos non pas seulement pour les institutions mais pour tous les comptes ayant utilisé au moins 1 fois 1 des #. De plus, dans votre liste de ce que vous acceptez, vous ne parlez pas des biographies.
  - Twitter :
    - Impossible à fournir de notre côté.
  - MoDyCo :
    - On peut récupérer la bio et ces infos, c'est juste que cela prend énormement de temps avec les limitations de l'API (cf remarque ci-dessus). Pour prendre un cas concret, récupérer 9000 bio nous prend plus de 2 jours.
  - Twitter :
    - Rectification de notre spécialiste API : en connaissant les user IDs dont on cherche à connaître les informations détaillées, la meilleure façon de procéder est d'utiliser la REST API via GET users/lookup qui permet "d'hydrater" 100 users à la fois sur chaque requête. Autrement dit, récupérer 9000 bios ne prend plus alors que 90 requêtes, soit moins de deux minutes.
- **DEMANDE 4**
  - MoDyCo :
    - **Serait-il possible de récupérer, pour toutes les institutions participantes (récupérées grâce au formulaire) depuis l'annonce de l'événement officiel sur Twitter jusqu'à 1 mois après l'opération le nombre de followers journalier de chaque comptes participant à la MuseumWeek ?**
  - Twitter :
    - Non pour TOUTES les institutions. Uniquement pour les principales avec une liste restreinte à déterminer car c'est un processus manuel de notre côté. La livraison des données se fera en une fois à la clôture de la période d'étude.
  - MoDyCo :

        - Avec l'accord de l'équipe "Evaluation", la liste pourrait être restreinte aux institutions françaises même si nous nous étonnons que l'opération soit manuelle.
    - Twitter :
        - Nous vous confirmons, c'est un processus qui n'est pas adapté pour un grand nombre de comptes.
    - MoDyCo :
        - Pourtant, il nous semble que c'est quelque chose de faisable avec votre API, non ?
    - Twitter :
        - Rectification de notre spécialiste API : à partir du moment où vous avez une liste des usernames, il est possible d'obtenir cette information en effectuant un GET users/lookup chaque jour avec la REST API.
- **DEMANDE 5**
    - MoDyCo :
        - **Pour chaque tweet, pourrait-on, en plus des données habituelles, connaître le matériel et le logiciel utilisés (p. ex. appli web, Android, iOS, tweetdeck) pour l'envoi des tweets ?**
    - Twitter :
        - Il est possible de récupérer l'app utilisée dans le champ "source" d'un Tweet, par exemple "Twitter for iPhone", mais pas possible de voir par exemple "iPhone 6 iOS 8".
    - MoDyCo :
        - On peut récupérer cette information par l'API ? De ce que nous savons, cette info est présente seulement dans l'archive d'un compte (donc à faire par le propriétaire du compte).
    - Twitter :
        - Nous vous confirmons, le champ source est dans l'API https://twittercommunity.com/t/how-to-get-the-users-device-information-of-a-tweet-via-streaming-api/15012.
    - MoDyCo :
        - Merci pour la confirmation.
- **DEMANDE 6**
    - MoDyCo :
        - **Pourrait-on récupérer les URLs originales présentent dans les tweets (plutôt que leur forme raccourcies) ?**
    - Twitter :

        - Non. Vous devez pou cela utiliser l'API pour filtrer les Tweets et récupérer les URLs. Précision : l'objet "entities" expose déjà toutes les URLs étendues dans les Tweets, à la fois en REST et Streaming.

- **DEMANDE 7**
    - MoDyCo :
        - **Pour le corpus de tweets, avoir une pré-segmentation par conversation autour d'un tweet (grâce à l'id_tweet et l'id_conversation).**
    - Twitter :
        - Nous n'avons pas d'API conversation pour lister toutes les réponses à un tweet.
    - MoDyCo :
        - Nous sommes surpris que vous de puissiez pas le faire étant donné que vous listez le nombre de réponse avec les id_conversation. À vérifier de notre côté.
    - Twitter :
        - Nous vous confirmons que ce n'est pas disponible.
- **DEMANDE 8**
    - MoDyCo :
        - **Pourrions-nous récupérer les analytics proposés par la plateforme officielle de Twitter ? Avec en plus les informations suivantes (pour chaque tweet) :**
            - Le nombre d'impressions (nombre de fois que des utilisateurs ont vu le tweet sur Twitter) ;
            - Le taux d'engagement (et donc tous les détails des interactions, à savoir : clics sur les médias intégrés, ouverture des détails, retweets, favoris, clic sur profil utilisateur, clic sur le lien, clics sur le hashtag, réponse, partager par e-mail, prise de nouveaux followers).
    - Twitter :
        - Il n'y a pas d'API pour analytics. Il est possible d'exporter ces analytics en CSV pour chaque compte. Cependant, 2 réserves :
            - Les statistiques ne sont activées qu'à partir du moment où le musée a accédé une première fois à analytics.twitter.com : c'est un point à aborder impérativement lors du webinar notamment ;
            - Ces données ne sont en principe pas partageables avec un tiers. Il est obligatoire que les musées concernés vous fournissent eux-mêmes leur archive personnel de tweets au format CSV.
    - MoDyCo :
        - C'est un point très important effectivement. D'autant, que sauf erreur de notre part (en tout cas, c'était comme ça il y a quelques mois à l'ouverture de

ces données), la plateforme analytics est liée à la plateforme “publicité” et il donc est nécessaire d’indiquer une carte-bleue, même si ces fonctions sont gratuites. De plus, les filtres temporels sont peu pratiques étant donné qu’il existe soit “les 7 derniers jours” soit “les 28 derniers jours”. Nous n’arrivons pas à trouver dans les CGU où il est précisé que les données ne sont pas partageables. Ne faut-il pas alors mettre dans le formulaire une case à cocher pour indiquer que les institutions sont d’accord pour participer à l’étude et accepte que Twitter partage ses données là à notre équipe d’évaluation ? Ou faut-il demander à toutes les institutions, peut-être via le ministère, de partager sur un ftp partagé et sécurisé ces données exportées ?

- Twitter :
  - Non, analytics.twitter est accessible depuis n’importe quel compte, mais pour son propre compte seulement. À vous de voir pour que les institutions vous envoient leur CSV à la fin de la MuseumWeek.
- MoDyCo :
  - Impossible à mettre en place dans le temps imparti...

- **REMARQUE**
  - MoDyCo :
    - Sur notre compte Twitter (dev), nous n’avons plus la possibilité de créer d’application. Nous dépassons la limite (message d’erreur). En scripant Twarc pour utiliser l’API Stream de Twitter, il nous semble que l’API nous limite à 2 streams en simultané avec le même jeu d’API (secret/acces token). Donc pour prévoir les 8 hashtags (de la MuseumWeek) à streamer en simultané, nous avons du créer plusieurs jeux d’accès. Cependant, nous restons confrontés au même problème !
  - Twitter :
    - Il faut maintenir une seule connexion à la Streaming API à tout instant. Sur une seule connexion Streaming, il est possible de filtrer en temps réel jusqu’à 400 mots clés avec le paramètre ‘track’. Plus d’infos : https://dev.twitter.com/streaming/overview/connecting. Cela devrait donc également résoudre le problème d’une nombre d’apps puisqu’en l’occurrence une seule suffira pour tout le projet.

A2 Exemple de données collectées durant la *#MuseumWeek*

A2.1 Tweet

Nous donnons dans cette section à la figure A2.1 un extrait de tweets collecté à l'aide du hashtag #architectureMW pendant la *#MuseumWeek* (le mardi 24 mars). On notera la présence de ce hashtag dans l'extrait au niveau de l'attribut de métadonnée : **"text"**.


```
{ "created_at": "Tue Mar 24 23:00:37 +0000 2015",
  "id": 580504519974498300,
  "id_str": "580504519974498300",
  "text": "De quel roi ce porc-épic est-il l'animal-symbole? #AnnedeBretagne #MuseumWeek
  #architectureMW http://t.co/zkM5WoqBm8,
  …
  "user": {
        "id": 293534469,
        "id_str": "293534469",
        "name": "Château de Nantes",
        "screen_name": "ChateauNantes",
        "location": "Nantes",
        "url": "http://t.co/AINzCLd5AX",
        "description": "Bienvenue au Château des ducs de Bretagne - Musée d'histoire urbaine de la
        ville de Nantes / participez à la #MuseumWeek",
        "protected": false,
        "followers_count": 3907,
        "friends_count": 316,
        "listed_count": 151,
        "created_at": "Thu May 05 14:28:44 +0000 2011",
        "favourites_count": 438,
        "geo_enabled": true,
        "verified": false,
        "statuses_count": 2741,
        "lang": "fr",
        … },
  "geo": null,
  "coordinates": null,
  "place": null,
  "contributors": null,
  "retweet_count": 1,
  "favorite_count": 0,
  "entities": {
         "hashtags": [
              { "text": "AnnedeBretagne", "indices": [50,65] }, …
              { "text": "architectureMW", "indices": [101, 109] } ],
         … },
  "favorited": false,
  "retweeted": false,
  "lang": "fr"
}
```


Figure A2.1 | Extrait de tweet (id 580504519974498300) émis collecté durant la *#MuseumWeek* (2015) ; tweet du Château de Nantes. Les attributs de métadonnées primaires sont indiqués en gras, les autres non. Les valeurs associées aux attributs (primaires ou non) sont indiquées en gris.

A2.2 Biographie

Nous donnons dans cette section à la figure A2.2 un extrait de biographie du compte Twitter du Château de Nantes collecté après la *#MuseumWeek* (le samedi 4 avril).


```
{ "created_at": "Thu May 05 14:28:44 +0000 2011",
  "id": 293534469,
  "id_str": "293534469",
  "name": "Château de Nantes",
  "screen_name": "ChateauNantes",
  "location": "Nantes",
  "url": "http://t.co/AINzCLd5AX",
  "description": "Bienvenue au Château des ducs de Bretagne - Musée d'histoire urbaine de la ville de Nantes / LT #expoLaboureur",
  "protected": false,
  "followers_count": 3907,
  "friends_count": 317,
  "listed_count": 157,
  "favourites_count": 452,
  "geo_enabled": true,
  "verified": false,
  "statuses_count": 2813,
  "status": {
        "contributors": null,
        "truncated": false,
        "text": "RT @LucasLechat: Au pied du plus haut clocheton du @ChateauNantes http://t.co/LH4bU50wvd",
        "created_at": "Sat Apr 04 15:05:23 +0000 2015"
        "geo": null,
        "coordinates": null,
        "place": null,
        "contributors": null,
        "retweet_count": 1,
        "favorite_count": 2,
        "entities": {
              "hashtags": [
                    { "text": "AnnedeBretagne", "indices": [50,65] },
                     …
                    { "text": "architectureMW", "indices": [101, 109] } ],
               … },
        "favorited": false,
        "retweeted": false,
        "lang": "fr" }
…
}
```


Figure A2.2 | Extrait de biographie (id 293534469) collectée durant la *#MuseumWeek* (2015) ; biographie du Château de Nantes. Les attributs de métadonnées primaires sont indiqués en gras, les autres non. Les valeurs associées aux attributs (primaires ou non) sont indiquées en gris.

A3 Données supplémentaires et hashtags non officiels

En plus des 10 corpus de tweets (9 thématiques et 1 global) que nous avons collecté durant la *#MuseumWeek*, nous disposons d'un 11ème corpus de 1 704 tweets, spécifique au Japonais. En effet, nous nous sommes aperçu le lendemain du démarrage de la *#MuseumWeek* que des versions japonaises[45] non officielles des hashtags de l'opération étaient parfois utilisées à la place des hashtags officiels pour tweeter au Japon. Nous projetons de consolider les données de la *#MuseumWeek* avec ce corpus, mais ne l'avons pas fait dans notre étude.

Par ailleurs, nous nous demandons dans quelle mesure cela ne serait pas aussi le cas dans d'autres pays et notamment, dans des pays dont la langue utilise un autre alphabet que l'alphabet latin tel que l'alphabet cyrillique ou arabe comme en Russie et dans les pays du Maghreb respectivement.

[45] http://bit.ly/1eM83f9

A4 Pays avec au moins une institution inscrite à la *#MuseumWeek*

Pays	#musées	Pays	#musées	Pays	#musées
United States	**520**	Turkey	9	Lithuania	2
United Kingdom	**442**	Belarus	8	Iceland	2
France	**358**	South Korea	7	Greece	2
Italy	**256**	Portugal	7	Georgia	2
Spain	**220**	Luxembourg	7	Armenia	2
Canada	**217**	South Africa	5	Yemen	1
Japan	**127**	Poland	5	Tunisia	1
Germany	92	Finland	5	Sierra Leone	1
Russia	74	Croatia	5	República	1
Mexico	72	Slovenia	4	Dominicana	1
Brasil	67	Serbia	4	Peru	1
Australia	39	Puerto Rico	4	Paraguay	1
Argentina	35	Norway	4	Mongolia	1
Netherlands	24	New Zealand	4	Monaco	1
Colombia	22	Israel	4	Malta	1
Belgium	22	India	4	Laos	1
Sweden	17	Venezuela	3	Indonesia	1
Switzerland	16	Uruguay	2	Hungary	1
Chile	16	United Arab Emirates	2	Guatemala	1
Austria	13	Slovakia	2	Egypt	1
Ireland	12	Singapore	2	Ecuador	1
Denmark	11	Qatar	2	Bénin	1
Ukraine	10	Philippines	2	Bangladesh	1
Latvia	10	Maroc	2	Bahrain	1

Tableau A4 | Pays avec au moins 1 institution inscrite à la *#MuseumWeek* (2015) et leurs inscrits[46].

[46] Résultats obtenus d'après la liste officielle d'inscriptions à la *#MuseumWeek* (http://bit.ly/1K2M8OL).

A5 Pays d'origine des participants à la *#MuseumWeek*

Cette section s'articule en 2 temps. Dans un premier temps, nous présentons les différents attributs de métadonnées des tweets qui peuvent être utilisés pour identifier : le lieu d'envoi d'un tweet, les lieux attenants à celui-ci et le pays d'origine (ou nationalité) de son auteur. Dans un second temps, nous présentons la méthode que nous avons mis en place afin d'inférer automatiquement, à partir des attributs précédents, le pays d'origine des participants à la *#MuseumWeek*.

A5.1 Métadonnées des tweets et pays d'émission

Les attributs de métadonnées associés à la localisation du lieu depuis lequel a été envoyé un tweet sont :

1. **geo** : information sur le lieu d'émission du tweet ;
2. **coordinates** : coordonnées (longitude/latitude) du point d'émission du tweet ;
3. **place** : coordonnées des lieux connus associés au point d'émission du tweet (sans pour autant forcément coïncider avec ce dernier) et informations sur le pays où sont localisés les lieux associés dont un attribut **country**.

Le premier type d'information est obsolète, tous les tweets récents ne le précise donc pas. Le deuxième type d'information est disponible uniquement quand l'émetteur du tweet a donné son accord pour sa géolocalisation, ce qui reste très rare dans les données de la *#MuseumWeek* (moins de 3% des participants). Le dernier type d'information, comme le précédent, est relativement rare dans les données de la *#MuseumWeek* (environ 3% des tweets seulement).

Aucune de ces métadonnées n'est donc suffisamment représentatifs des données de la *#MuseumWeek* pour s'assurer d'obtenir des statistiques significatives ni sur le lieu d'émission des tweets ni sur la nationalité de leurs auteurs.

A5.2 Méthode d'inférence mise en place dans notre étude

L'alternative que nous avons trouvé pour déterminer le pays d'origine des participants à la *#MuseumWeek* (institutions et individus) consiste à utilisé l'attribut de métadonnée **location**, situé au niveau de la section utilisateur des tweets (c.-à-d. l'attribut **user** indiqué à la figure A2.1 de la section A2.1 donnée en annexe). Cet attribut est toujours présent dans les métadonnées, avec une valeur non nulle dans 70,4% des cas (soit 414 525 tweets).

Comme indiqué au niveau de l'API de Twitter[47], l'attribut location n'est pas nécessairement un lieu. Nos analyses manuelles ont montrées que cet attribut n'est pas toujours rempli par les utilisateurs sur leur compte Twitter, mais que dans tous les cas, il l'est en bien plus grande quantité que les attributs que nous avons mentionnées précédemment (c.-à-d. pour 70,1% des données). Par ailleurs, quand cet attribut a une valeur, il contient le plus souvent des informations relatives à un lieu telles que : une adresse, un acronyme de province (p. ex. *TX* pour Texas), un nom de ville, voire un nom de pays, plus ou moins bien orthographié.

Nous avons donc tenter de déterminé autant que faire se peut les noms de pays associés aux valeurs de l'attribut location. Pour cela, nous avons utilisé la librairie Python *geonamescache*[48] (version 0.18), en présupposant que les informations (ou token) trouvées dans cet attribut devaient référer au pays d'origine des utilisateurs Twitter (et donc des participants à la *#MuseumWeek*).

Focus sur notre méthode d'inférence

Pour notre étude, nous avons en particulier utilisé les dictionnaires de pays et de villes reconnus dans *geonamescache* et les avons combiné pour obtenir automatiquement un nom de pays à partir d'un token ville quand aucun token pays n'est présent dans l'attribut location.

Il s'avère que la solution mise en place n'est pas encore satisfaisante puisqu'elle ne fonctionne que dans 62,7% des cas (soit pour 259 146 des 413 227 tweets pour lesquels ont pourraient inférer le pays d'origine des participants à la *#MuseumWeek*).

La solution n'est pas optimale car nous n'avons exploité qu'une partie des ressources de la librairie *geonamescache* sans même les combiner à d'autres ressources disponible via d'autres librairies Python qui pourraient s'avérer utile pour améliorer les performances de notre méthode d'inférence. En particulier, nous pensons aux libraires *guess-language-spirit* et *pycountry*[49] qui fonctionnent sur la base d'une représentation ISO commune avec la librairie *geonamescache. guess-language-spirit* pourrait nous permettre d'inférer la nationalité d'un utilisateur à partir des tokens de l'attribut location en identifiant avec précision la langue utilisée parmi 60 langues différentes[50]. *pycountry* nous permettrait de faire la même chose que *geonamescache* mais avec une couverture plus large des noms de pays (3 166 au total versus 252 seulement avec *geonamescache*).

[47] https://dev.twitter.com/overview/api/tweets
[48] https://pypi.python.org/pypi/geonamescache
[49] https://pypi.python.org/pypi/guess_language-spirit
[50] *guess-language-spirit* est par exemple capable de distinguer anglais-américain d'anglais-britannique.

A6 Erreurs de classification langagière commises par Twitter

Cette section traite des erreurs de classification commises par Twitter pour identifier correctement la langue des institutions participantes à la *#MuseumWeek*. Nous avons déceler ce genre d'erreurs à la création des corpus de tweets relatifs aux institutions françaises inscrites à la *#MuseumWeek* (c.-à-d. pour la création des sous corpus de type fr-fr_2 mentionnés dans le tableau 3 de la section 2.1.1 à la page 18).

A6.1 Attribut en défaut

L'attribut *secondaire* lang renseigné par Twitter pour spécifier la langue d'un utilisateur au niveau d'un tweet est parfois erroné (voir à la figure A2.1 à la section A2.1 pour repérer cet attribut dans la structure des métadonnées associés aux tweets collectés).

A6.2 Exemple

Nous donnons à la figure A6.2 un exemple de tweet erroné. Cet exemple correspond à un tweet émis par le Centre des Monuments Nationaux. On voit dans cet exemple que le champ n°3 (c.-à-d. la valeur de l'attribut secondaire lang) n'est pas fr (pour français), comme attendu pour le CMN, mais en-gb (pour anglais britannique) !

```
1 leCMN,
2 506082474,
3 en-gb,
4 580822526701543426,
5 fr,
6 Ne serait-ce pas la @FondationLV que l'on distingue depuis le toit de
l'#Arcdetriomphe?
  #MuseumWeek #ArchitectureMW http://t.co/UuWUGRmIN8
```

Figure A6.2 | Exemple d'erreur de classification langagière commise par Twitter. Donnée prétraitée (format csv, un champ par ligne) extraite du corpus global de tweets (ALL ; type fr-fr_2). **Champ 1** : émetteur du tweet. **Champ 2** : identifiant du compte associé. **Champ 3** : langue du compte associé. **Champ 4** : identifiant du tweet. **Champ 5** : langue du tweet. **Champ 6** : tweet (contenu textuel).

A6.3 Taux d'erreurs

Nous n'avons pas évalué le taux exact d'erreurs de classification faites par Twitter pour l'ensemble des données de la *#MuseumWeek*. Nous nous sommes contenté de l'estimer à partir du nombre d'erreurs total commis par Twitter sur l'ensemble des 17 650 tweets de langue française émis par les 358 institutions françaises inscrites à la *#MuseumWeek* (c.-à-d. l'ensemble des tweets contenu dans le sous corpus global de type fr-fr_2 : ALL).

Les taux d'erreurs calculés nous ont permis :

1. De garantir la validité des statistiques de corpus fournies dans le tableau 3 de la page 18 et en particulier, celles relatives aux corpus de type fr-fr_2 ;
2. De garantir la validité des résultats d'analyses catégorielles présentés à la section 2.2 de ce rapport qui dépendent des corpus de type fr-fr_2.

Taux d'erreur relatif aux tweets émis par des institutions françaises

Nous donnons aux tableau A6.3 différentes statistiques établies lors de l'analyses des erreurs de classification langagière commises par Twitter sur les 17 650 tweets français émis par les institutions françaises inscrites à la *#MuseumWeek*.

Cas d'analyse	Inscription MW : France	$lang_{user}$	#institution	#tweet	@
1	x	en-gb en	2	473 6	leCMN ComedieFr (C-F)
2	x	fr en	2	5 4	ArtLudique AktinosGallery
3	x	-	22	0	...[51]
4	x	und[52]	1	1	EcomuseAvesnois

Tableau A6.3 | Informations relevées durant l'analyse des erreurs de classification langagière commises par Twitter sur le corpus global de type fr-fr_2 : ALL (cas d'analyse 1). Autres cas d'analyse : musées français inscrits à la *#MuseumWeek* n'ayant émis aucun tweets en français (2), aucun tweets du tout (3) et seulement des tweets français mais non reconnus par Twitter (4). @ : mention institutionnelle sur Twitter (donnée en minuscule uniquement et sans le symbole *arroba*).

D'après les résultats présentés au tableau X, moins de 1% des institutions françaises couvertes par nos analyses sont concernées par les erreurs de classification langagière de Twitter (c.-à-d. 2/358 institutions analysées), le nombre total de tweets correspondants s'élevant à tweets 479 (CMN : 473 tweets ; C-F : 6 tweets) soit 2,7% exactement du corpus global de type fr-fr_2 : ALL. Par souci d'exactitude, ces données devraient être ajoutées aux corpus de données francophones et françaises associés aux statistiques du tableau 3 page 18 (c.-à-d. les corpus de type F-xx, f-xx, f-fr et fr-fr_1 respectivement)[53].

[51] agrofondation, chercheurdautre, ciplepinat, crane_fr, cultureaccess, demeureduchaos, eco_archi, frac_ca, fvasarely, lacorbatarosa, lacoupole62, la_plate_forme, lemuseedepoche, mareis_aquarium, musee_camargue, museechartreuse, museenietzsche, myownartgallery, opium_art, tunantes, unisciel, zebreric.

[52] Und = Undefined (pour indéterminé).

[53] En effet cela a été fait pour les corpus de type fr-fr_2.

Taux d'erreur relatif à l'ensemble des tweets de collecte

Le taux d'erreurs de classification langagière commis par Twitter sur l'ensemble des données collectées est estimé à 15 987 tweets, c'est-à-dire près de 3% des données (2,7% exactement).

Cela semble raisonnable compte tenu de la quantité de données générés par Twitter durant la *#MuseumWeek* et nous assure que les résultats produits dans notre étude à partir de ces données, bien qu'erronées pour certaines, sont significatifs.

A7 Autres résultats

A7.1 Classification : institutions françaises de 2015 (groupes 1-3)

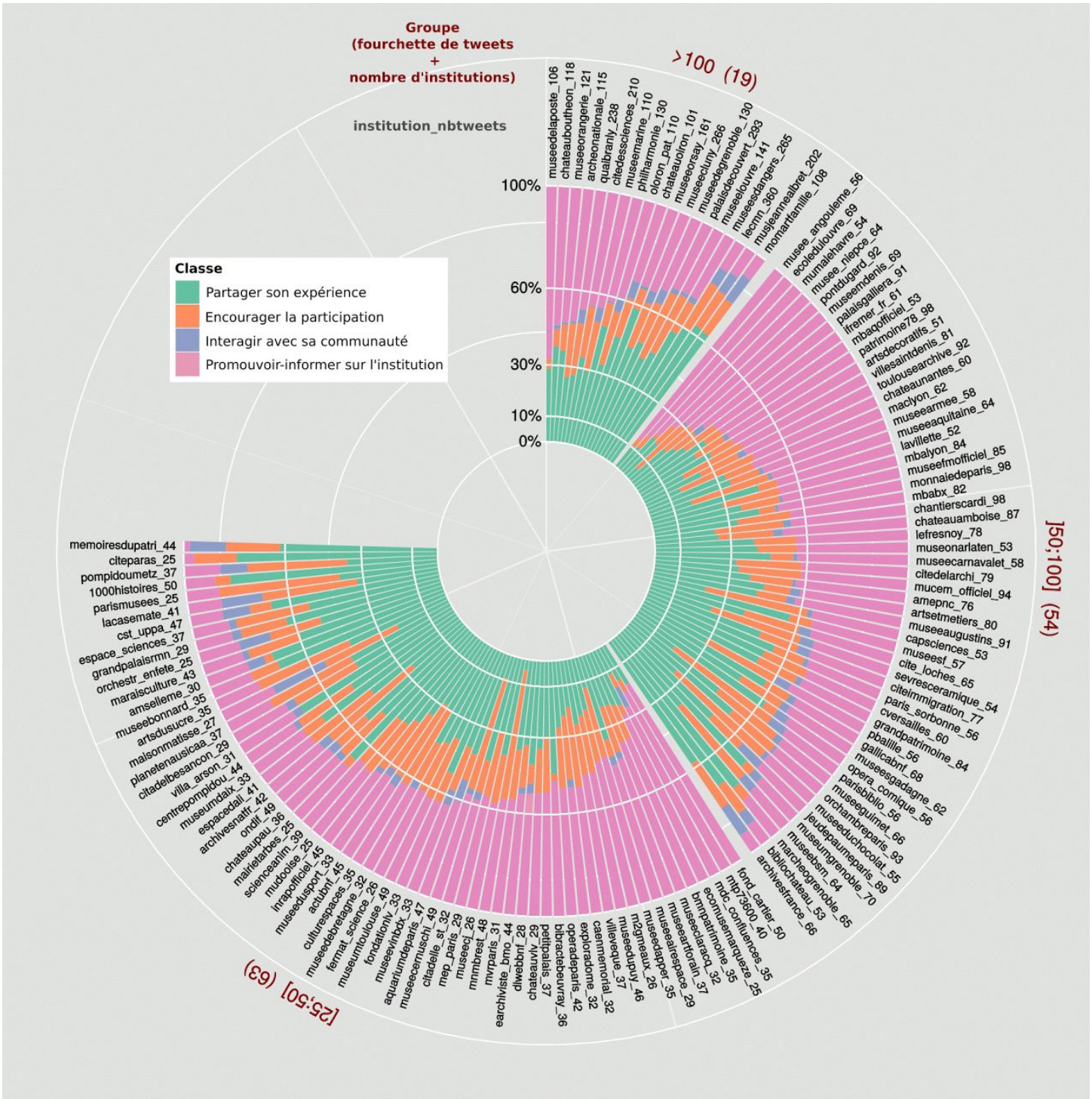


Figure A7.1 | Classification globale des 137 institutions françaises ayant émis le plus de tweets durant la *#MuseumWeek* (2015), basé sur le contenu communicationnel des tweets (retweets exclus).

A7.2 Classification : musées français à la *#MuseumWeek* de 2014

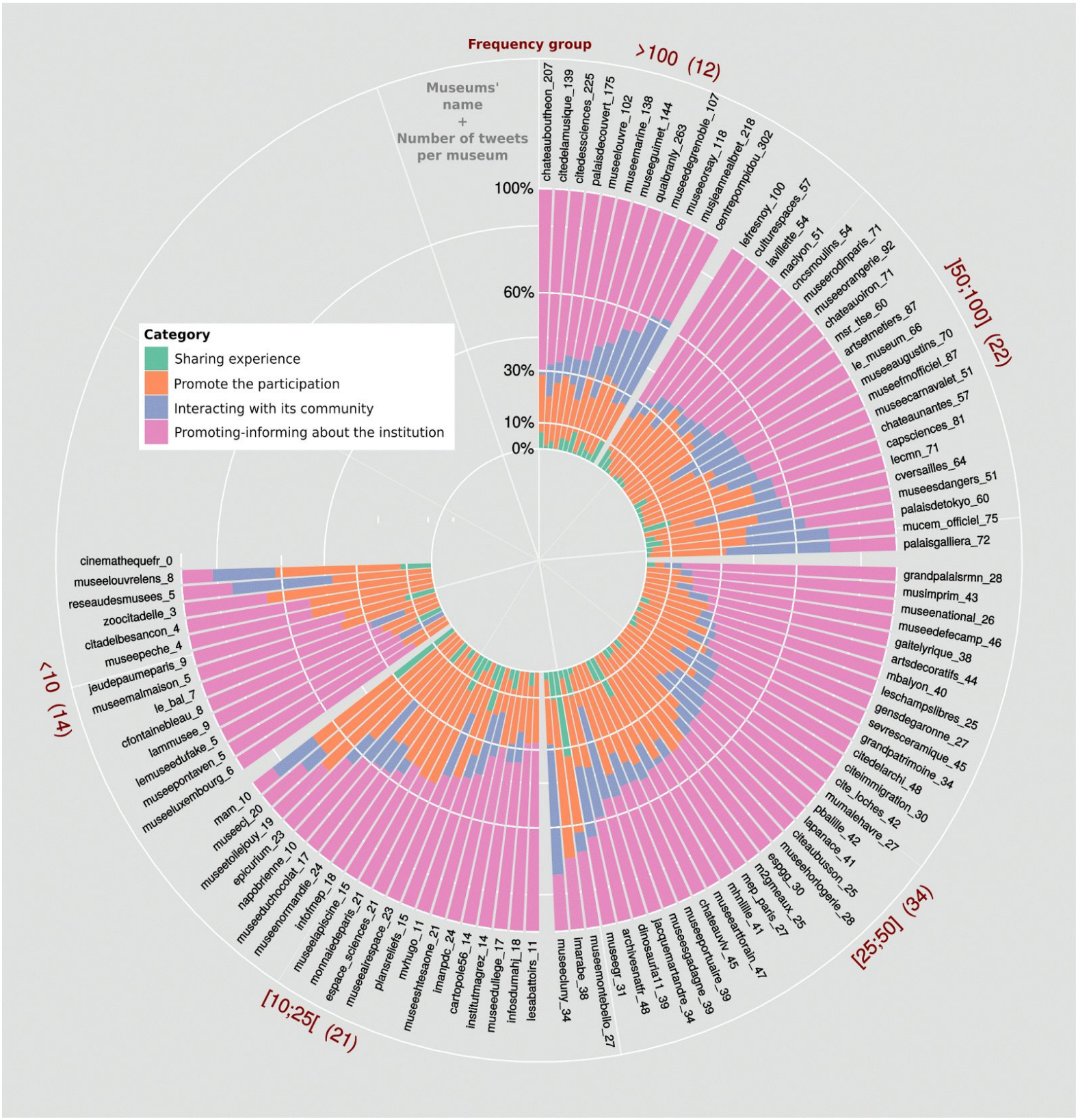


Figure A7.2 | Classification globale des 103 musées français inscrits à la *#MuseumWeek* l'an passée (2014), basée sur le contenu communicationnel de leur tweets (retweets exclus).

A7.3 Comportement communicationnel : institutions du groupe 2

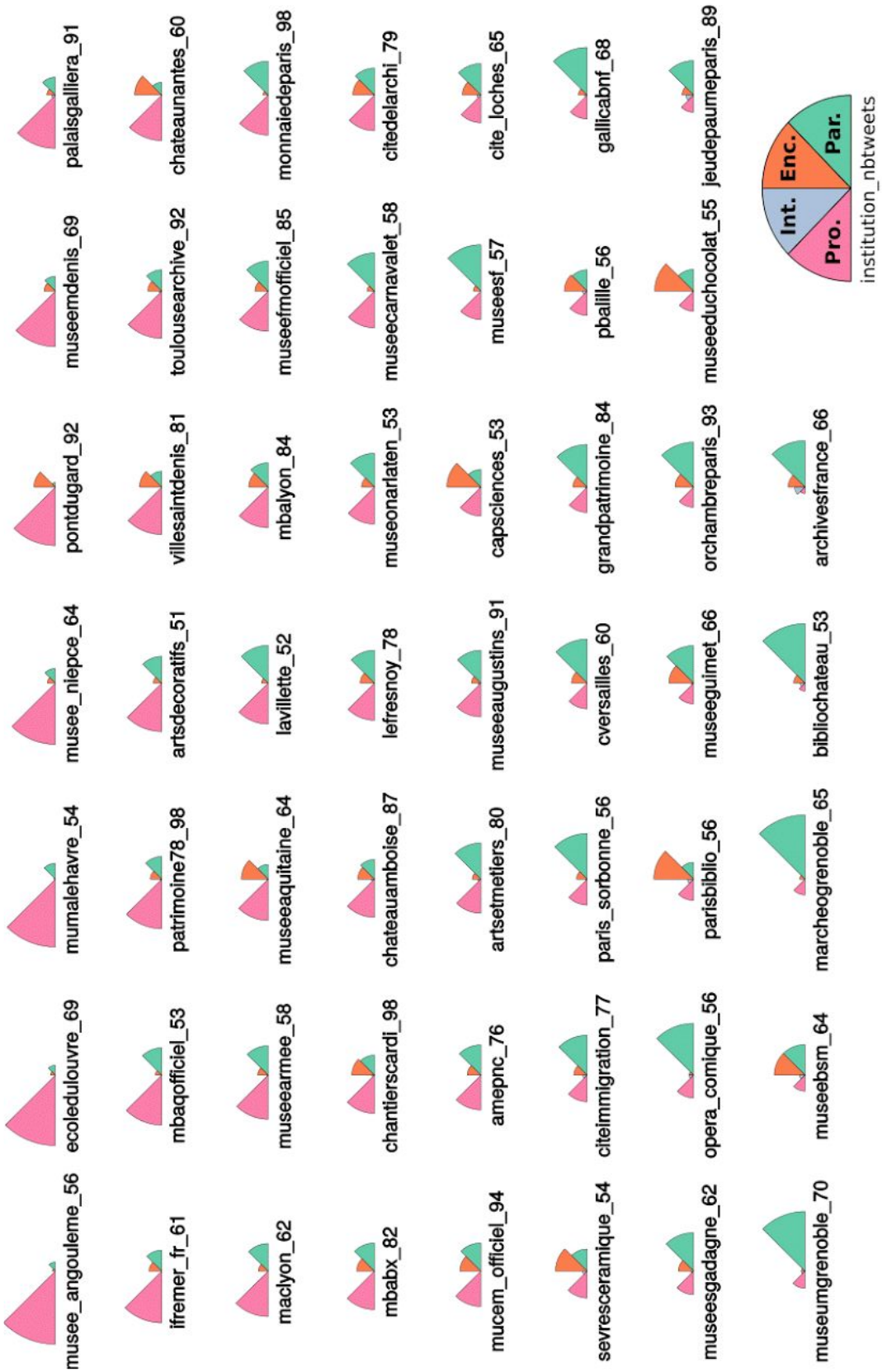


Figure A7.3 | Comportement communicationnel des 54 institutions françaises ayant émis entre 50 et 100 tweets durant la *#MuseumWeek* (2015) ; retweets exclus.

A7.4 Comportement communicationnel : institutions du groupe 3

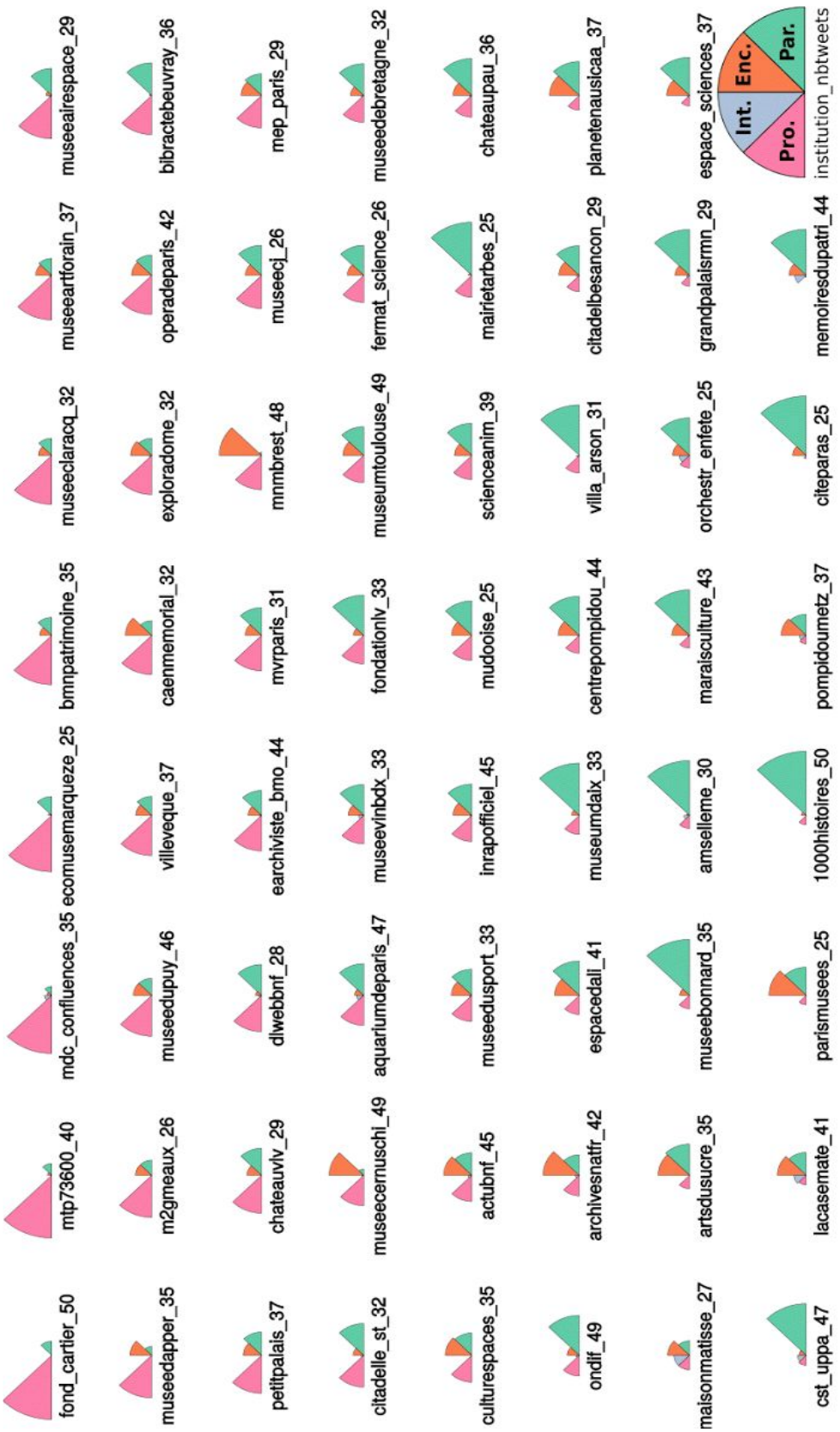


Figure A7.4 | Comportement communicationnel des 63 institutions françaises ayant émis entre 25 et 50 tweets durant la *#MuseumWeek* (2015) ; retweets exclus.

A7.5 Comportement communicationnel : institutions du groupe 4

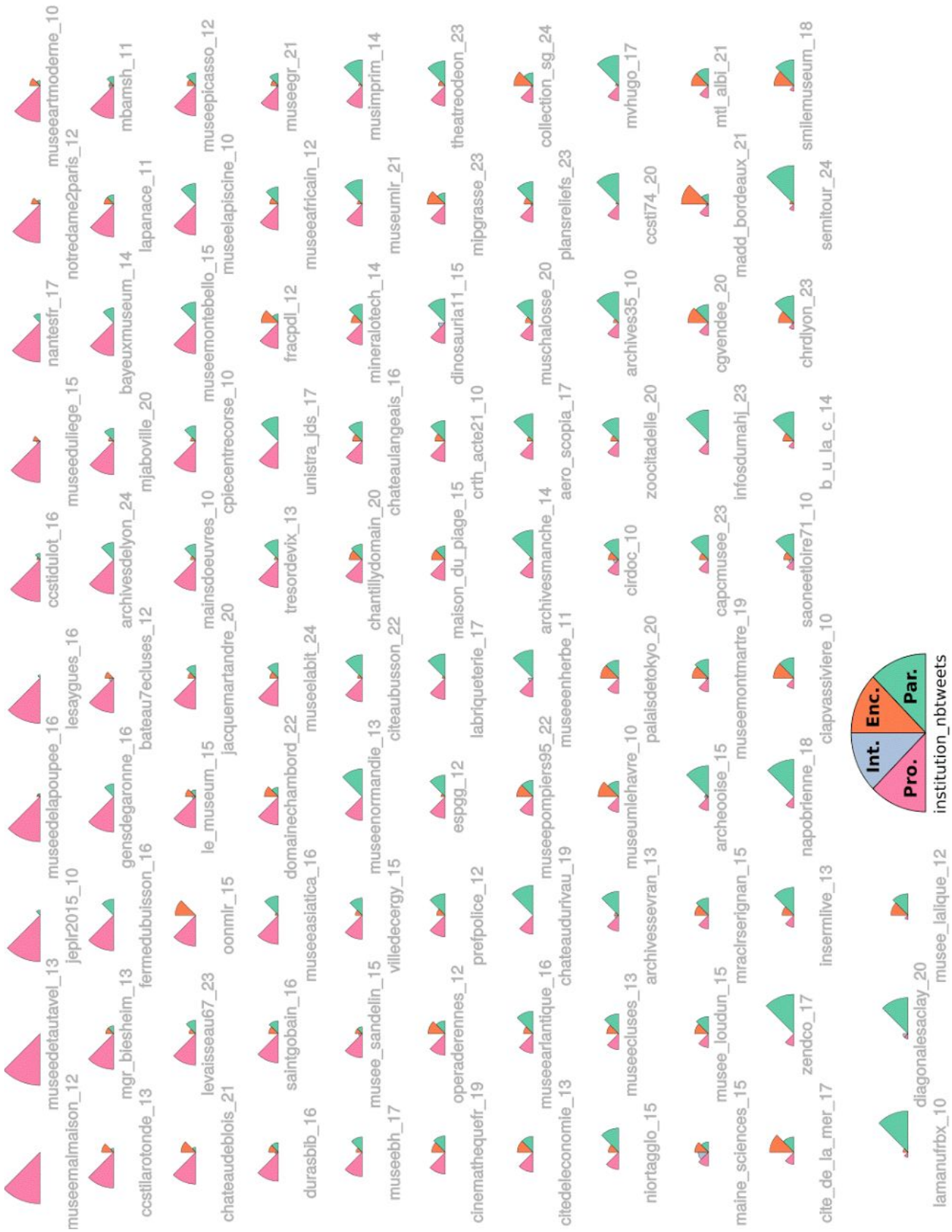


Figure A7.5 | Comportement communicationnel des 103 institutions françaises ayant émis entre 10 et 25 tweets durant la *#MuseumWeek* (2015) ; retweets exclus.

A7.6 Comportement communicationnel : institutions du groupe 5

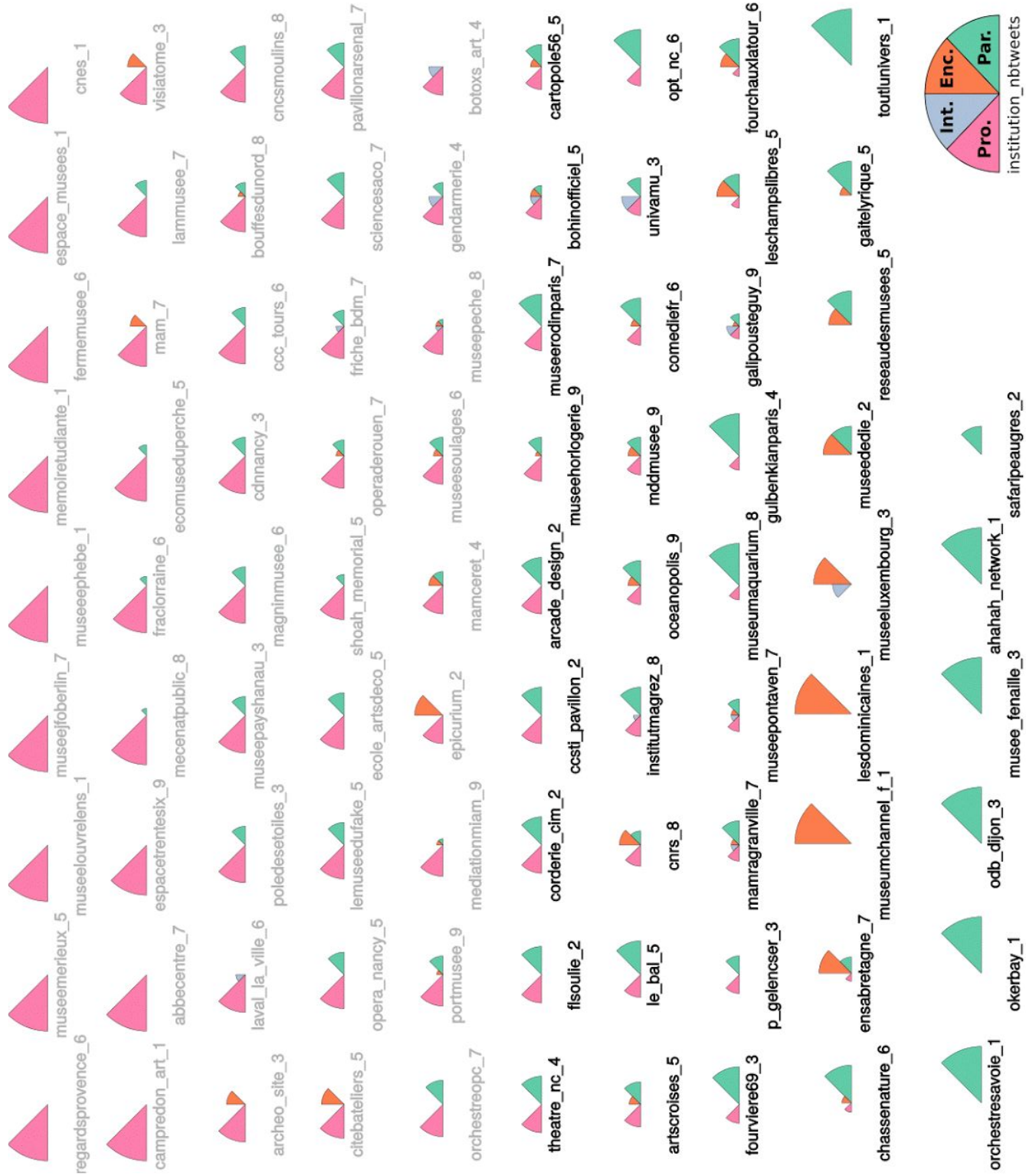


Figure A7.6 | Comportement communicationnel des 87 institutions françaises ayant émis moins de 10 tweets (mais au moins 1) durant la *#MuseumWeek* (2015) ; retweets exclus.

A7.7 Comportement communicationnel : institutions du groupe 6

On donne à la figure A7.7 la liste des institutions françaises inscrites à la *#MuseumWeek*, mais qui n'ont émis aucun tweet en français (ou reconnu comme tel par Twitter) durant l'événement.

1. agrofondation
2. aktinosgallery
3. artludique
4. chercheurdautre
5. ciplepinat
6. clermontfd*
7. crane_fr
8. cultureaccess
9. demeureduchaos
10. eco_archi
11. ecomuseavesnois
12. fabriquejpc*
13. frac_ca
14. fvasarely
15. gazeparisienne*
16. harcourtdomaine*
17. lacorbatarosa
18. lacoupole62
19. la_plate_forme
20. lemuseedepoche
21. mareis_aquarium
22. montrouge92
23. musee_camargue
24. museechartreuse
25. museedefecamp*
26. museedesmedias*
27. museenietzsche
28. myownartgallery
29. opium_art
30. tunantes
31. unisciel
32. zebreric

Figure A7.7 | Liste des comptes Twitter (sans *arroba* et tout en minuscule) relatifs aux 32 institutions françaises n'ayant émis aucun tweet durant la *#MuseumWeek* (2015) ; retweets exclus. Les 6 institutions ayant émis des retweets (en français) sont marquées d'un astérix.

A8 Devis proposé au Ministère de la Culture et de la Communication

Dans cet annexe, nous présentons le devis de prestation tel qu'il a été fourni au ministère et dont l'étude visée fait l'objet de ce livrable.

À la différence de l'édition 2014 où la sphère francophone avait des hashtags dédiés qui permettaient d'identifier facilement les corpus sur lesquels travailler, l'édition 2015 de la #MuseumWeek a promu une utilisation commune des hashtags à l'internationale en choisissant des hashtags à consonance/graphie anglo-saxonne.
Ce choix stratégique des hashtags combiné à la popularité de l'évènement (c.-à-d. de la masse de données collectées) sans oublier les restrictions imposées par l'API publique de Twitter, demandent pour cette édition 2015 de revoir certains des développements mis en place pour traiter les données de l'édition précédente.

Le contenu de cette première prestation de 20 jours facturée 5 000 euros inclus l'extension des traitements réalisées sur les données de l'édition 2014 de la #MuseumWeek dans le but de pouvoir les appliquées aux données de l'édition 2015. Autrement dit, la majeure partie des prestations de ce lot consiste à compléter et/ou adapter certains modules de la chaîne de traitement de données actuelle mais pas seulement ; certains des lots proposés pour cette prestation impliquent la création de modules de traitements supplémentaires et notamment de traitements type visualisation de données.
Par ailleurs, les différents traitements de données et analyses statistiques sont cumulatifs et non restrictifs au sens où les paramètres choisis pour une analyse peuvent être croisés avec ceux d'autres analyses afin d'aboutir à des résultats plus précis. Par exemple, identifier les 10 comptes non institutionnels les plus actifs parmi l'ensemble des comptes non institutionnels avant le début de l'opération (c.-à-d. entre le 4 février et le 23 mars).

Code prestation	Description	Type de traitement				Nombre de jours
		PT	AQ1	AQ2	V	
Lot1_01	Nettoyage et consolidation des données (p. ex. intégration des tweets avec hashtags non officiels ou erronés).	X	X		X	**4**
Lot1_02	Découpage + regroupement des données en fonction de la langue des tweets et de la nationalité des comptes (p. ex. réunion des pays francophones versus réunion des tweets écrits en français quelque soit la nationalité des auteurs).	X	X		X	**5**
Lot1_03	Analyse catégorielle des tweets (c.-à-d. classification automatique des tweets en catégorie communicationnelle basée sur leur contenu textuel) ; analyse globale, par institution et par individu (c.-à-d. les non institutionnels).	X	X	X		**4**
Lot1_04	Réalisation de visuels pour chaque institution française engagée dans la MW ; basée sur la classification automatique des tweets.				X	**3**
Lot1_05	Identification des comptes actifs ; basée sur le calcul d'indicateurs relatifs aux nombres (globaux/spécifiques) de (re)tweets émis durant la *#MuseumWeek*.		X		X	**2**
Lot1_06	Évolution du nombre de tweets en fonction du temps (sur une fenêtre de +1/-1 semaine autour de l'événement) ; évolution global par hashtag.	X	X		X	**2**
Total						20

Tableau A8 | Description et détail des traitements associés à chaque lot de prestation proposés dans le devis n°1 de subvention du Ministère de la Culture et de la Communication. PT : pré-traitement ; inclus le *pré-traitement* des données collectées pour la *#MuseumWeek* ainsi que le *pré-traitement* des données de visualisation tels que des graphiques et des tableaux de statistiques, lorsque la case *visualisation* située le long de la même ligne est cochée. AQ1 : analyse quantitative. AQ2 : analyse qualitative. V : visualisation.