



#ww1. The Great War on Twitter

Frédéric Clavert, Université de Lausanne, DH2016
@inactinique / frederic.clavert@unil.ch




A corpus of (yet) two million tweets

- April 2014-April 2016
- 2 096 968 tweets
 - Few noise but #11Nov #Somme #Verdun
 - 730 111 without retweets
- 542 570 accounts
 - Of all sorts
 - Individuals, memory-based institutions, media, research projects, bots...
- 124 424 hashtags
 - 54 566 just once in the database
 - 107 047 ten times or less in the database
- Harvesting still on-going
 - With the 1916 battles commemoration (Verdun / Somme): 2.8 million.

Technical infrastructure

- LAMP server
- PHP script: 140dev
 - Connects to Twitter Public Streaming API
 - Collect tweets based on keywords (#ww1, #1gm, #ewk and many others)
 - Parse JSON into MySQL
- Export to CSV for data analyses
 - No «real time» analysis

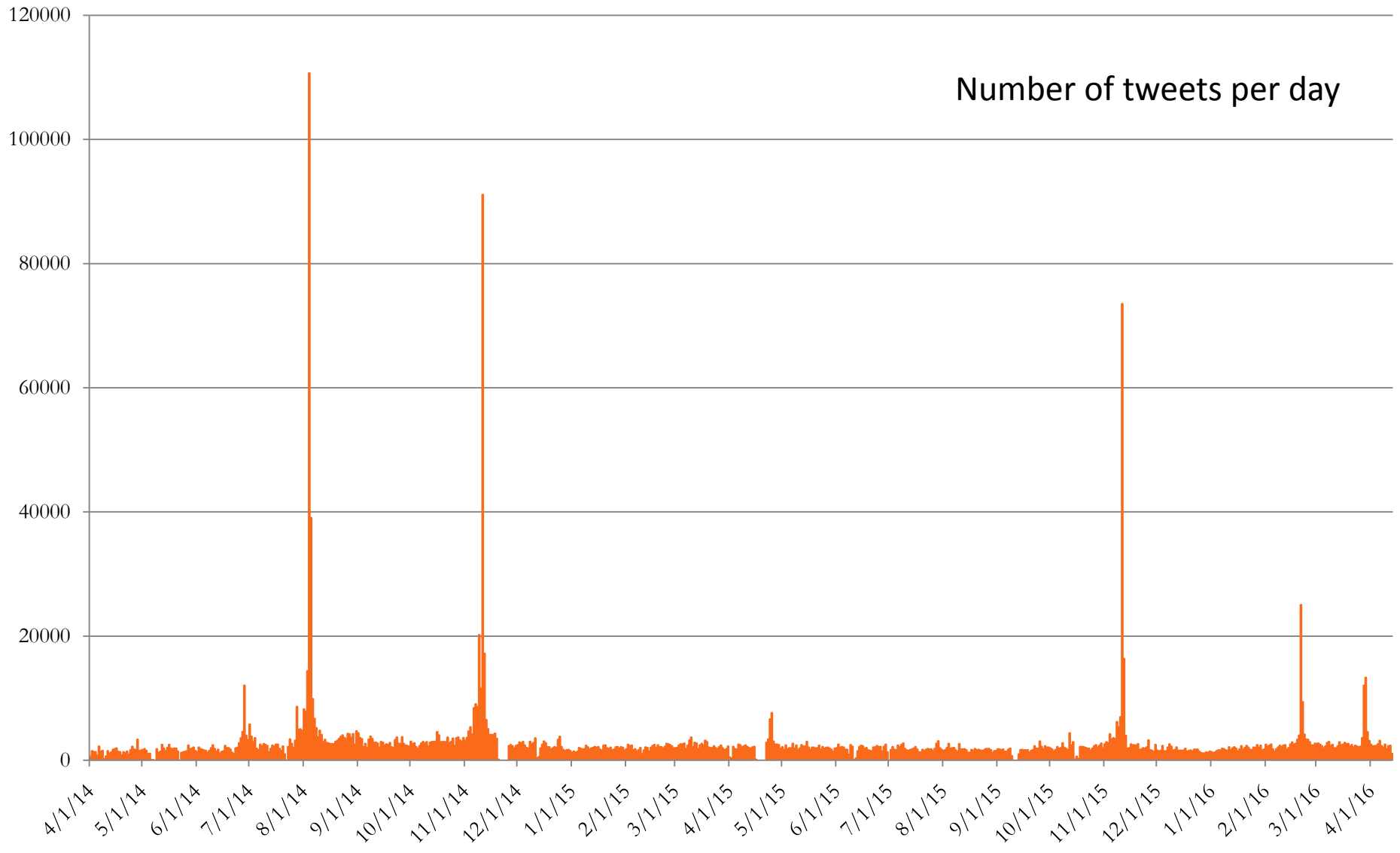
A painting of a man with reddish-brown hair, seen from behind, wearing a dark suit. He stands on a rocky outcrop, looking out over a vast, hazy landscape with rolling hills and distant mountains under a pale sky. The style is impressionistic, with soft, blended colors.

Historian facing a sea of data

How to read
two millions tweets?

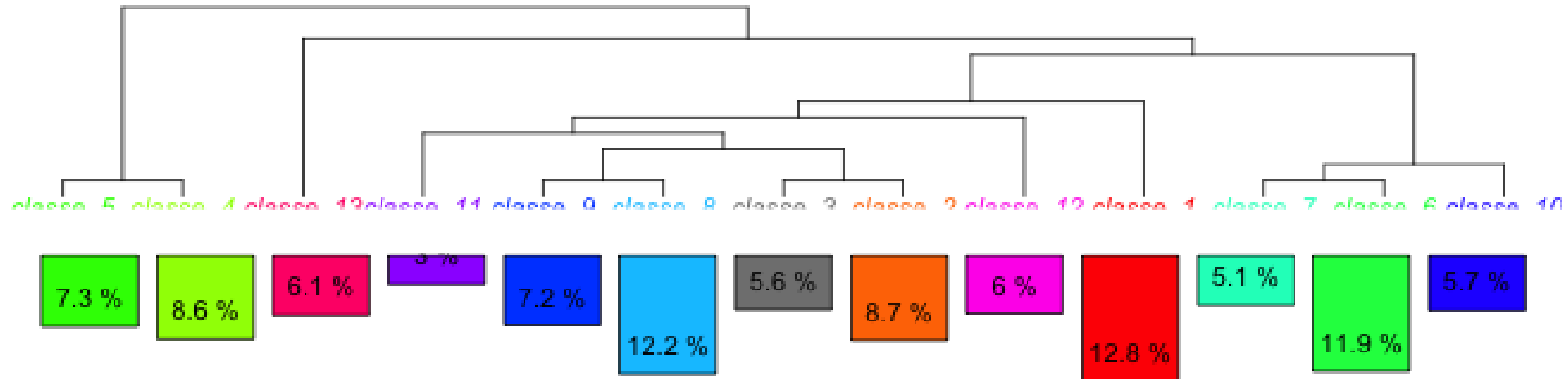
A close-up, dark, and somewhat abstract image of a rocky outcrop or cliff edge, with a dark, moody sky in the background.

Basic statistics



Clustering

- Text as statistics (text-mining)
- Software: Iramuteq
 - Groups similar tweets in clusters
 - Similarity = co-occurrences of words
 - Tree diagram (dendrogram)
 - Words that are shown are the most typical words of each cluster



2e
cl
rire
joseph
sdt
ric
classe
rac
jean
marius
hôpital
louis
caporal
beauséjour
emile
françois
4e
rg
béziers
bc
ra
auguste
16e
mpf
paul
3e
hurlus
dec
canonnier
bcp
st
hopital

tuer
rire
ennemi
naître
louis
dcd
blessure
pierre
marier
décéder
meuse
jean
henri
suite
362e
belgique
coray
douaumont
haumont
neuville
bonneau
alain
vaast
valoir
avril
septembre
caporal
eugène
disparaître
jules
soldat
bois

livre
indexés
ficher
commun
indexatic
matricule
poilu
registre
partir
indexées
indexer
collaborativ
défi
mam
indexations
carte
homme
transcrire
maj
mdh
compléter
memoire
participant
inscrire
désormais
entièrement
intégraleme
indexons
icaunais
etape
mémoire

mondi
premie
guerre
1ère
lecteur
dessiner
inventer
au_dela
londres
jardin
tropical
peugeot
infographie
moooc
bande
fois
allier
lorrain
fin
conflit
invitation
évacuation
quentin
1ere
accepter
webdocumen
conséquences
be
chaîne
canada
apocalypse
commémores
video

lire
lettre
dernier
article
raconter
dossier
nouveau
blog
revue
récit
lecture
numéro
recherche
destin
page
témoignage
courrier
épisode
breton
écouter
historien
roman
correspond
relire
nouvel
sujet
publier
cœur
femme
consacrer

grand
guerre
découvrir
exposition
archive
gt
sita
photo
expo
journal
histoire
sélection
image
musée
colloque
retrouver
retour
front
document
source
veille
collecte
collection
dessin
travers
fonds
art
ressource
quotidien
civil
jeu
présenter

soir
novem
dimanch
conféren
prochain
samedi
mercredi
ouvrir
visite
18h
mardi
rdv
rendez_voi
porte
guider
programme
table
soirée
ouvert
manquer
11h
lundi
salle
19h
ronde
cinéma
h
18h30
20h

beau
élève
matin
école
travail
militaire
ville
inviter
maire
cérémoni
projet
année
participer
reportage
cimetière
chanter
lycée
rue
marseillais
voir
public
pari
parisien
mission
remise
musique
hôtel
collège
conseil

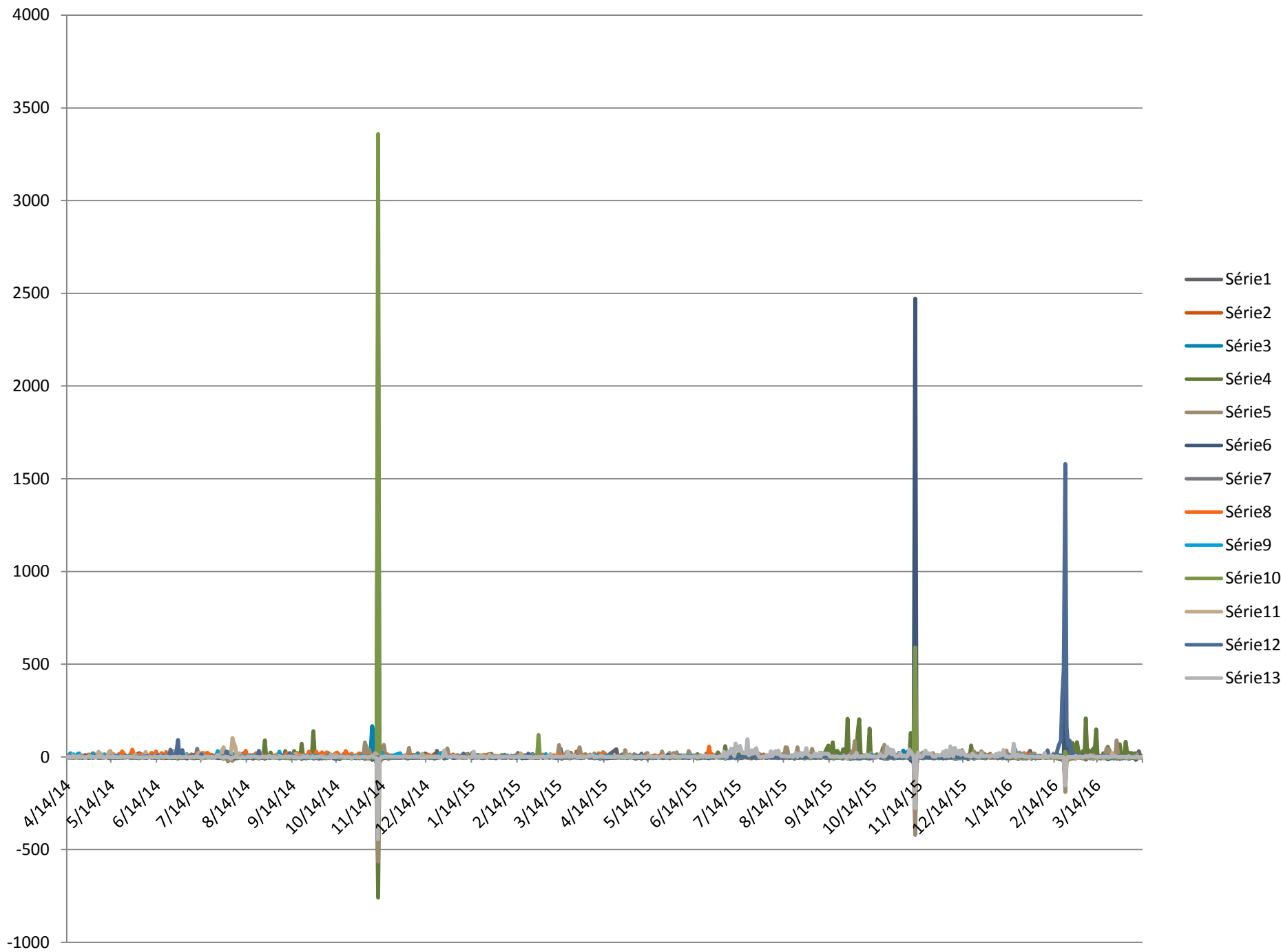
bataill
verdun
champ
enfer
début
direct
centenaire
commémori
débuter
elysées
siècle
symbole
suivre
mémorial
célébration
émission
australien
chiffre
réouverture
mythe
krumeich
ossuaire
rouvrir
élysées
gerd
spécial
live
xxe
jaurés
fresque
joe
démarrer
sacco

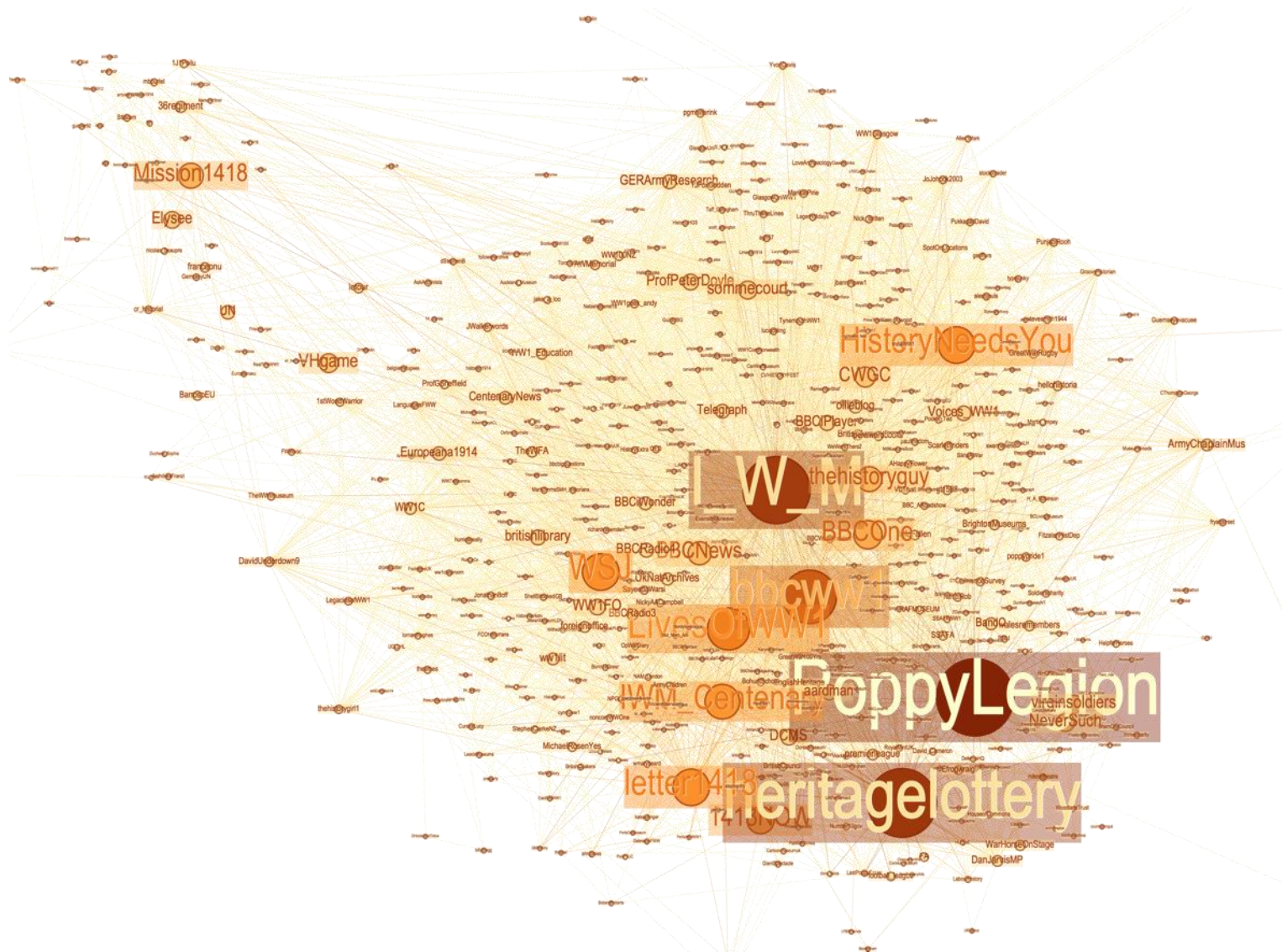
allema
obus
tranchée
batterie
nuit
tirer
canon
attaque
coup
bombardem
calme
artillerie
tir
heure
offensive
pluie
7e
troupe
arriver
blessé
neige
bombarder
infanterie
officier
côté
feu
village
repos
occuper
général
carnet
près de

homm
rendre
france
mort
victime
bel
tomber
industriel
combattant
soldat
unir
anatole
poilu
déclarer
royaume
héros
faber
joueur
allemagne
héraldais
patrie
champion
tour
vainqueur
stade
ardenne
driant
mourir
loi
rugbyman
mention
député
descendant

pensé
oublier
battre
souvenir
liberté
paix
pays
france
honneur
respect
sacrifice
patrie
défendre
vie
combattre
héros
père
donner
sacrifier
honorer
sauver
mort
peuple
merci
tomber
français
rester
vivre
génération
arrière
ninine

férié
nabilla
tt
gens
chose
con
penser
parler
important
vraiment
hollande
monde
fautre
hashtag
putain
aller
préférer
jour
fête
honte
main
pauvre
poignée
passer
gars
respecter
ya
mec
cool
faim
hnsse





As a conclusion

- Understanding the missing parts
 - Where are the German Twitter accounts?
 - Where are prisoners, women, colonial troops, soldiers who did not fall on the battlefield?
 - Where are other fronts than the West-European one?