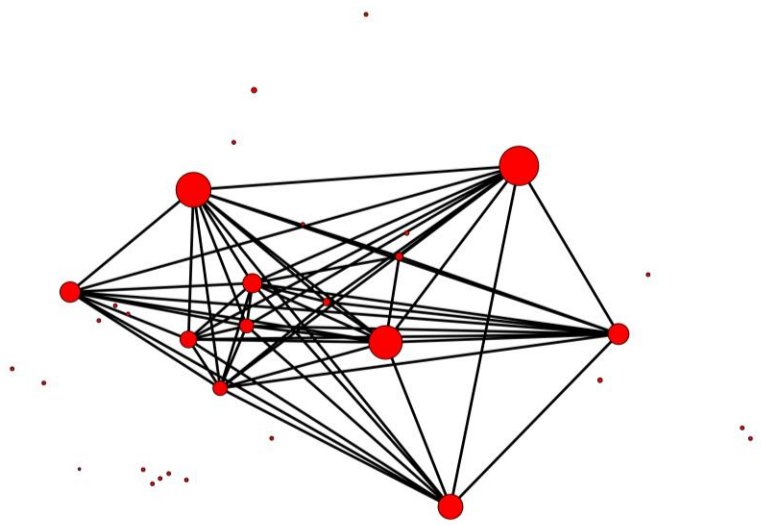


DEFENCE AND SPACE

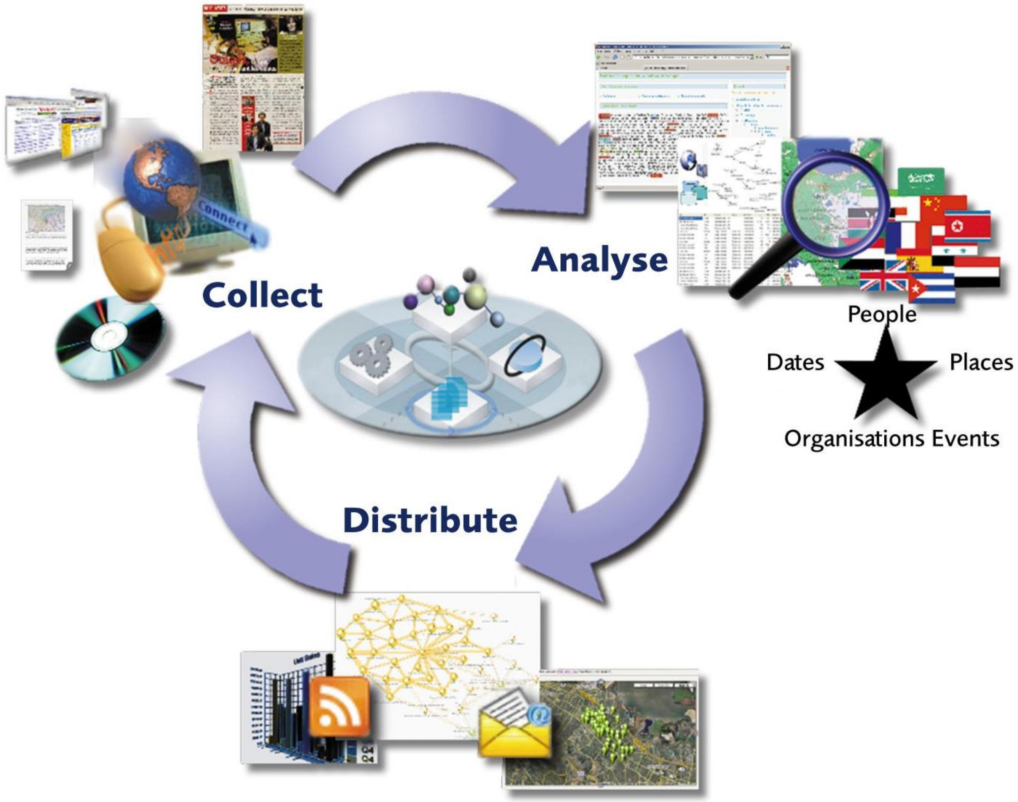
Detection of thematic communities in online social media

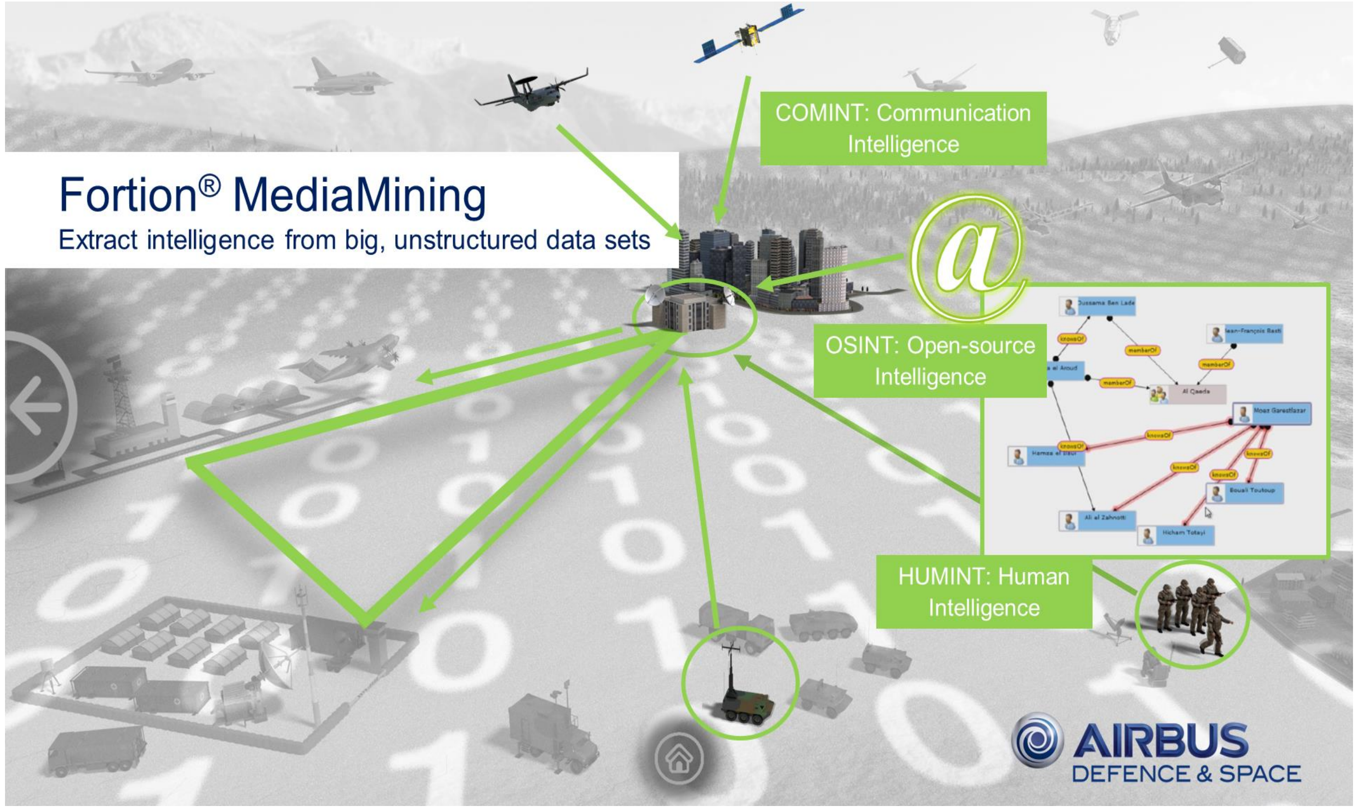
An Airbus & Litis work

Guillaume GADEK, Alexandre PAUCHET, Stéphan BRUNESSAUX, Laurent VERCOUTER,
Khaled KHELIF and Nicolas MALANDAIN

18th October 2018
MARAMI 2018, Avignon

Artificial Intelligence at AIRBUS Defence & Space: Open Source Intelligence Fortion® MediaMining




Fortion® MediaMining
Extract intelligence from big, unstructured data sets
COMINT: Communication Intelligence
OSINT: Open-source Intelligence
HUMINT: Human Intelligence
AIRBUS
DEFENCE & SPACE

Challenge: detect and explore *communities* on social media

Thematic communities:

1) Strong interaction
2) Common centres of interest

Two case studies:

- On (a subset of) Twitter: KevRandTweets

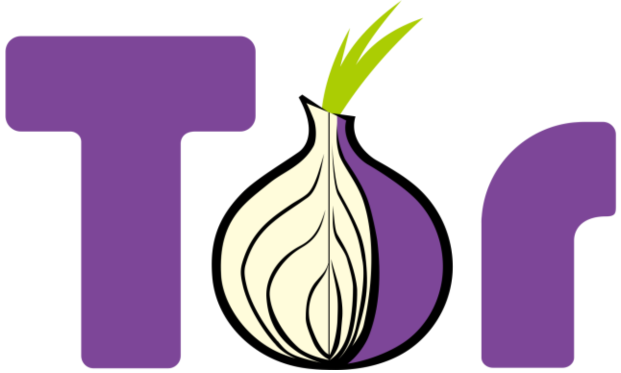
  9,671,711 tweets, December 2016, US politics;
  centered around 5,000 user accounts.

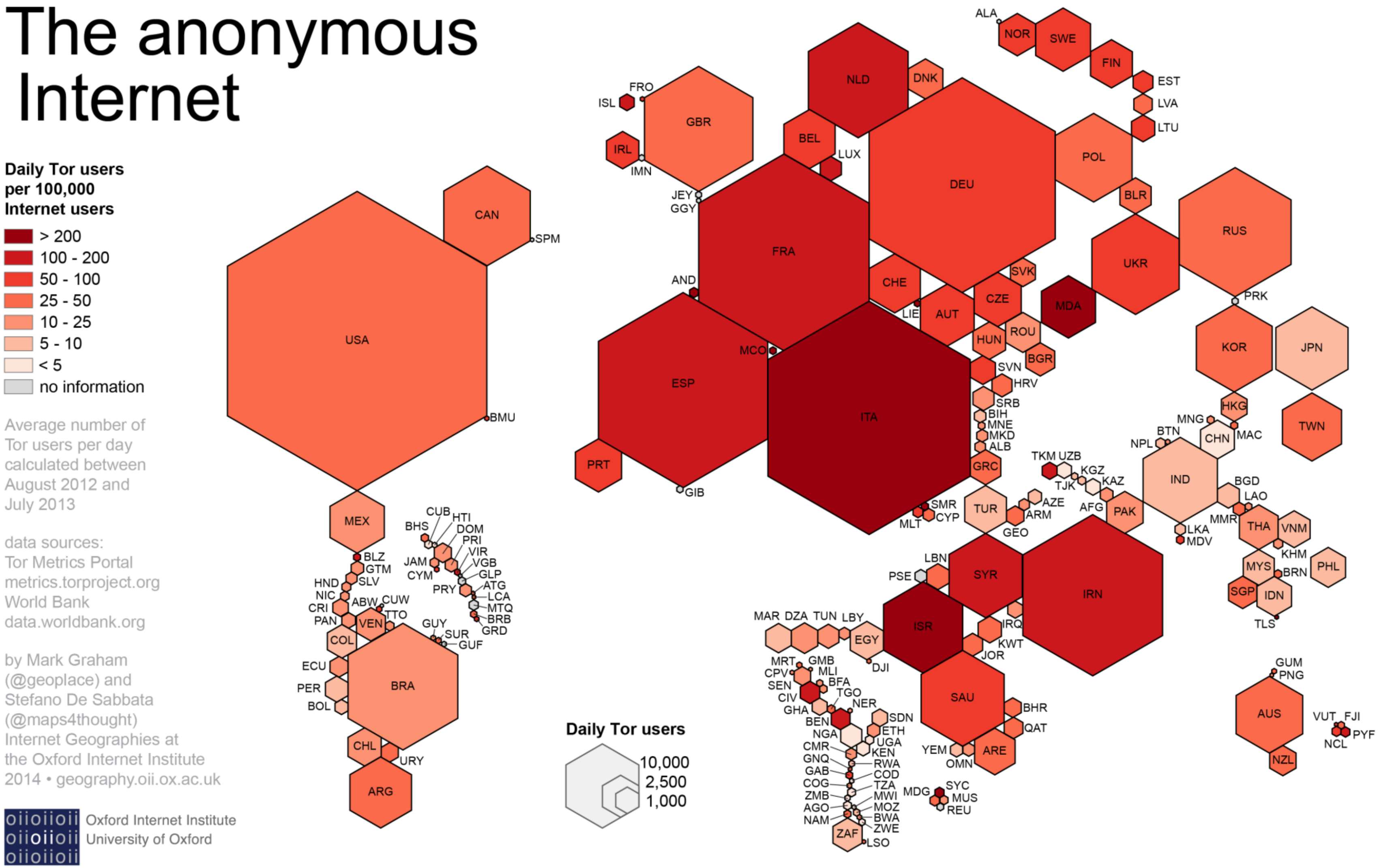
- On Galaxy2

  30,000 posts, 20,000 users, active 2015-2017 on TOR.

Today’s target: Galaxy2, on TOR

TOR : The Onion Router





The anonymous Internet

Daily Tor users per 100,000 Internet users

- > 200
- 100 - 200
- 50 - 100
- 25 - 50
- 10 - 25
- 5 - 10
- < 5
- no information

Average number of Tor users per day calculated between August 2012 and July 2013

data sources:
Tor Metrics Portal
metrics.torproject.org
World Bank
data.worldbank.org

by Mark Graham (@geoplace) and Stefano De Sabbata (@maps4thought)
Internet Geographies at the Oxford Internet Institute
2014 • geography.oii.ox.ac.uk

Oxford Internet Institute
University of Oxford

Daily Tor users

10,000
2,500
1,000

Today’s target: Galaxy2, on TOR

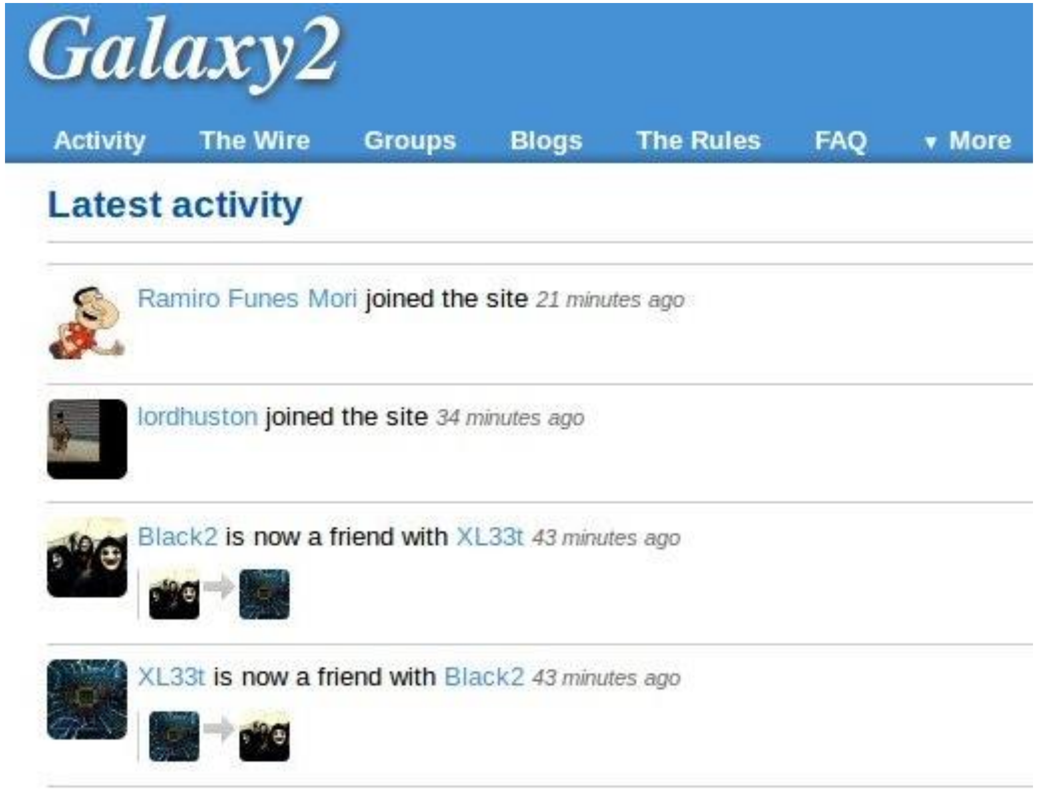
Galaxy2: ‘’the most popular social network on TOR’’

Active in 2015-2016-2017, disrupted since.

Based on the **elgg** open source framework.

Microblogging and friendship features.

About **20,000 user accounts created** in total.



Galaxy2

Latest activity

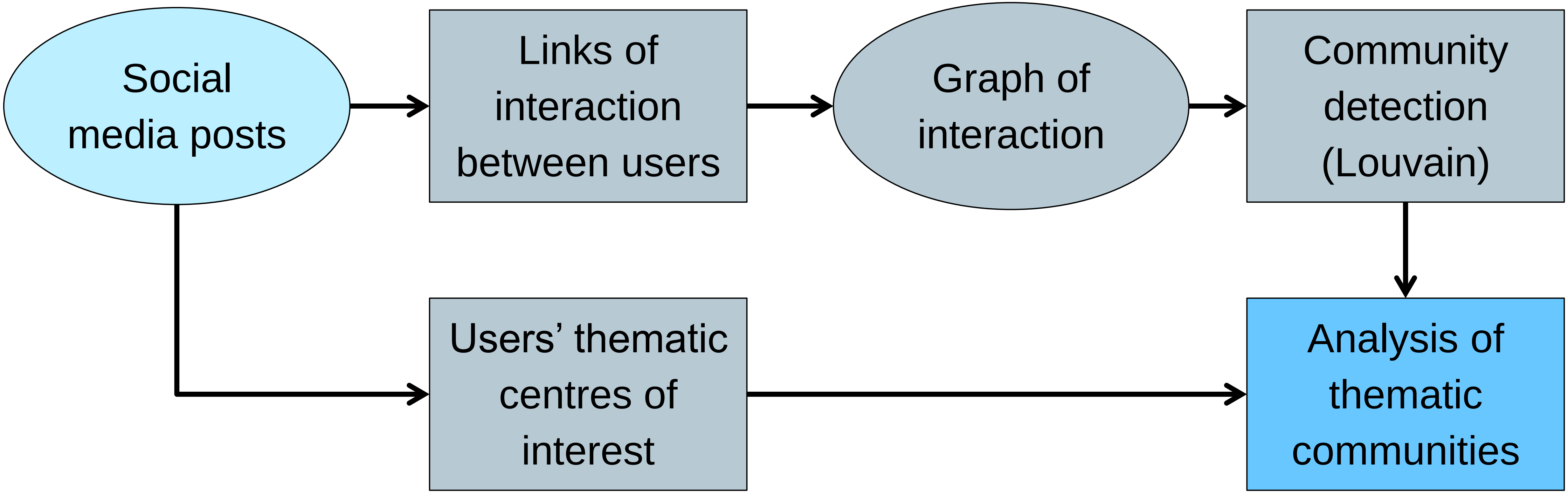
Ramiro Funes Mori joined the site *21 minutes ago*

lordhuston joined the site *34 minutes ago*

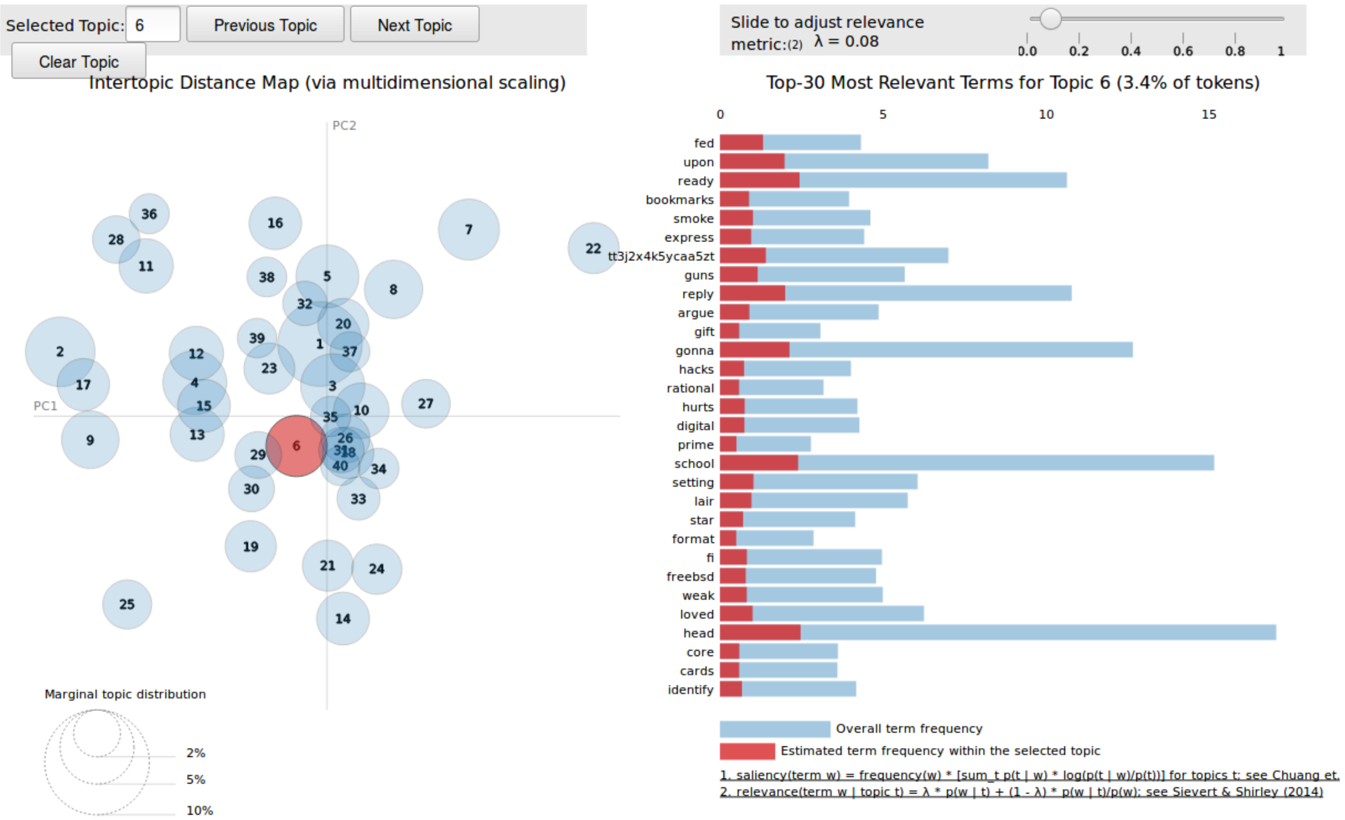
Black2 is now a friend with XL33t *43 minutes ago*

XL33t is now a friend with Black2 *43 minutes ago*

Analysis of social structures: process followed


Social media posts
Links of interaction between users
Graph of interaction
Community detection (Louvain)
Users' thematic centres of interest
Analysis of thematic communities

Textual Analysis: Topic and Sentiment



Selected Topic: 6

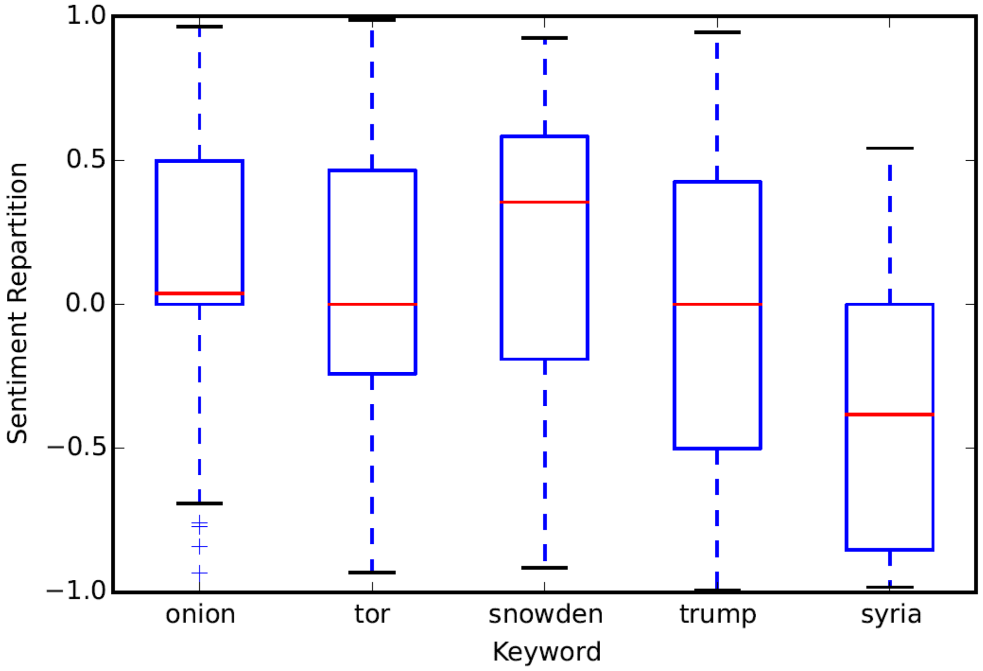
Slide to adjust relevance metric:(2) λ = 0.08

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et.

2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Emergent topic detection and description [C. Sievert 2014]

Coupling Sentiment [C.J. Hutto 2014] with keywords

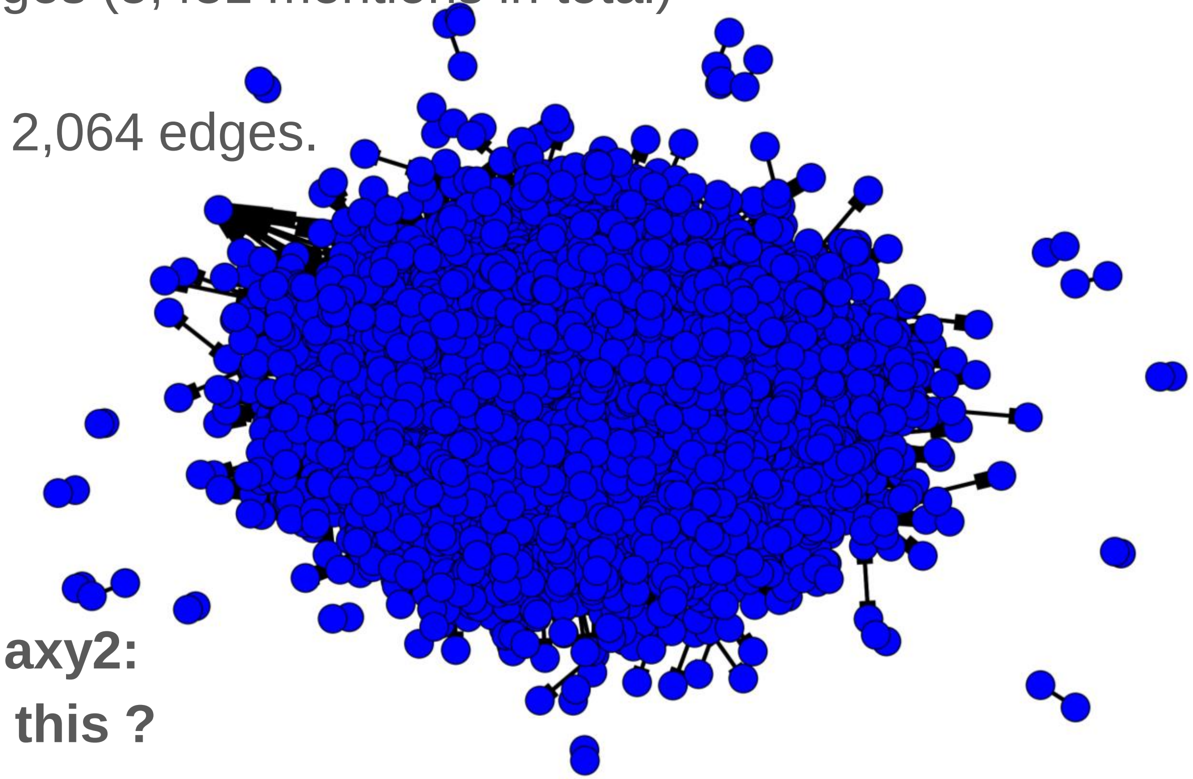


Modelling a social network with graphs

G_F the graph of **Friendship:** 7,356 nodes, 60,860 edges.

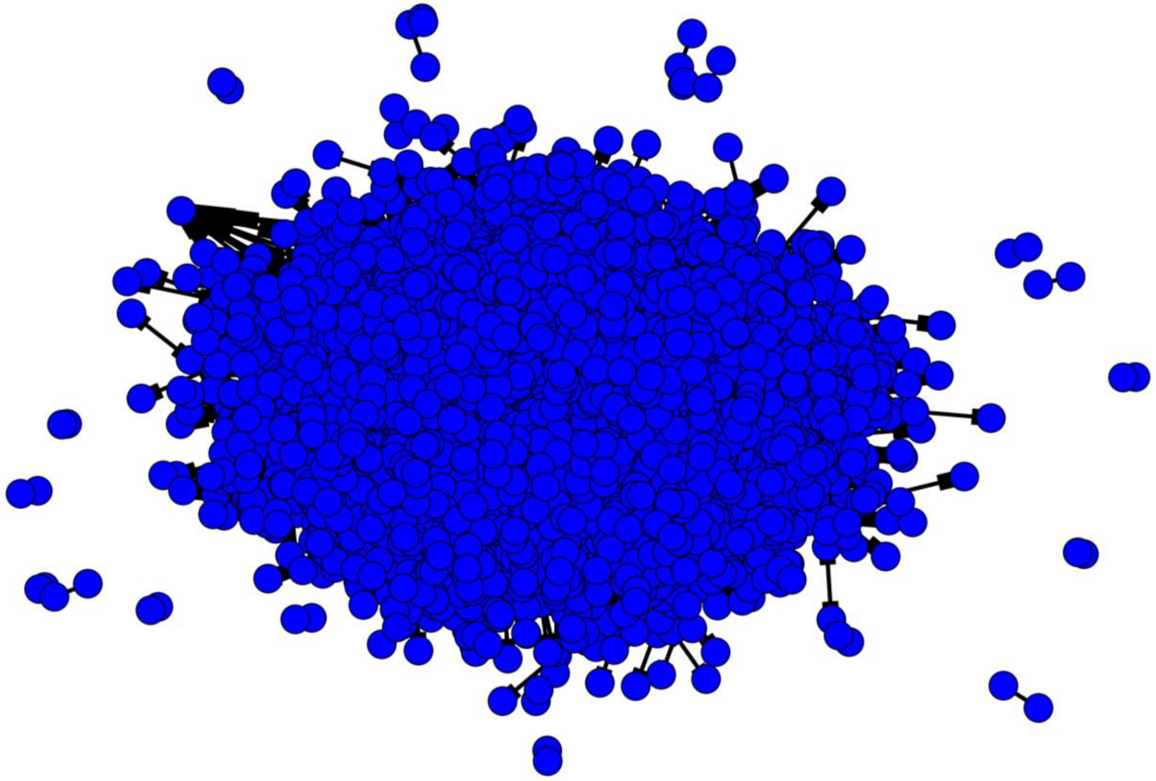
G_I the graph of **Mentions:** 968 nodes, 2,342 edges (5,481 mentions in total)

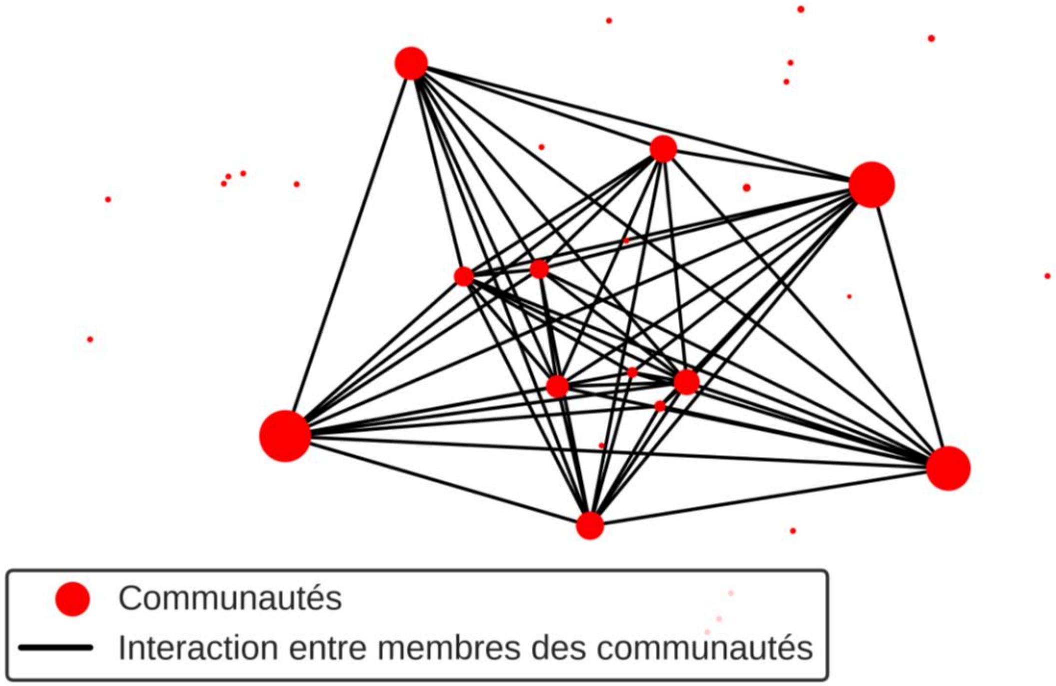
G_Ω the graph of **Objects sharing:** 1,092 nodes, 2,064 edges.



Graph of friendship links on Galaxy2:
How to extract information from this ?

Transforming graphs of users to graphs of groups





Detection and characterisation of communities: measures of cohesion

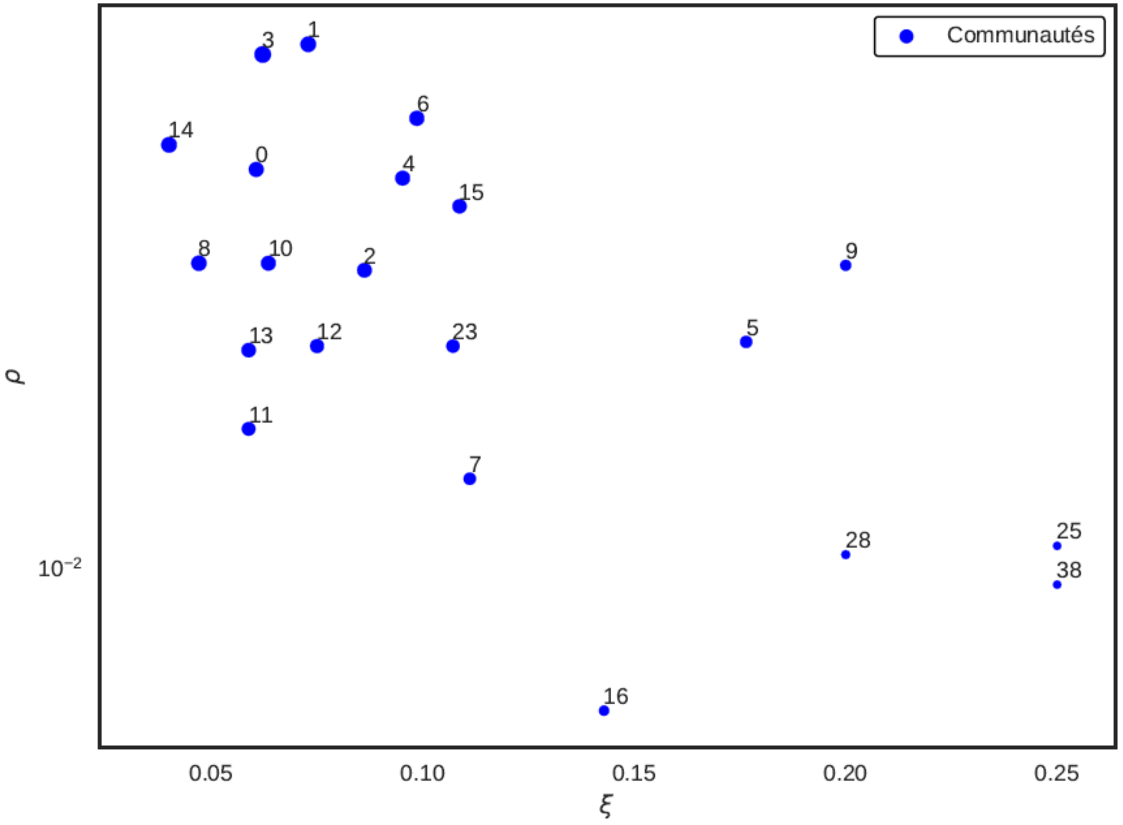
Topological measures [Yang and Leskovec, 2015]**:**

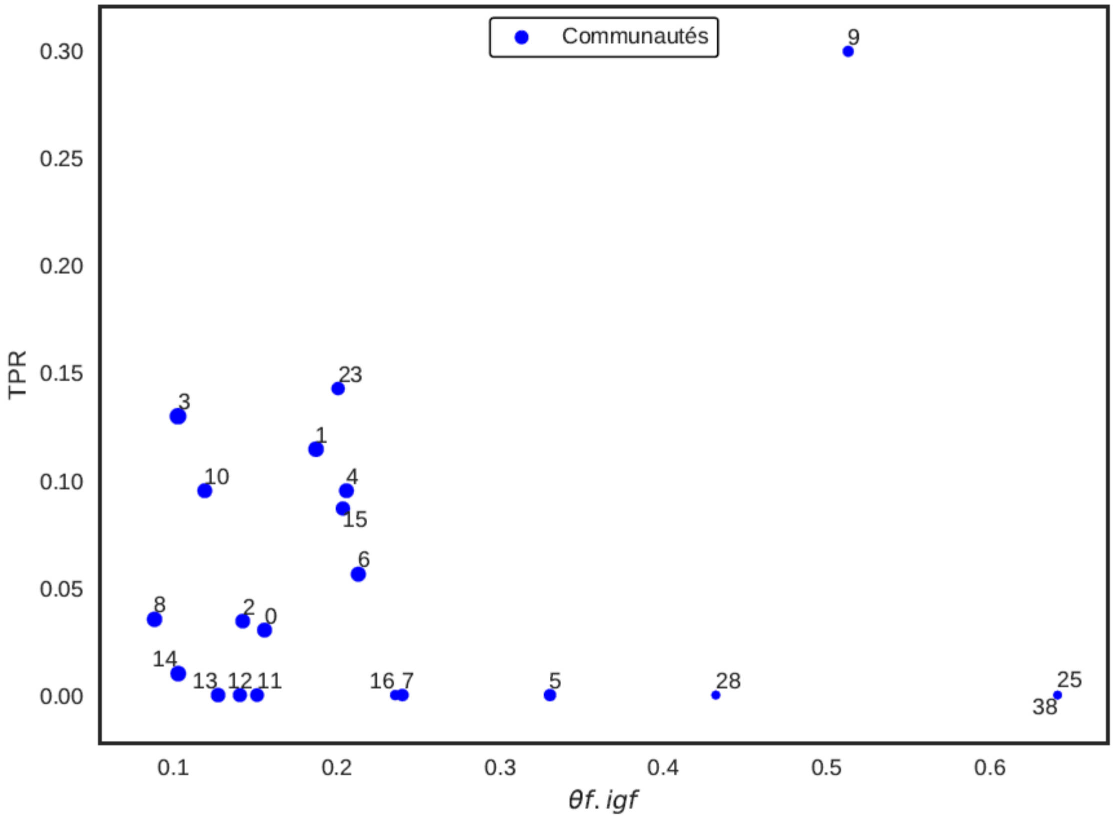
- *Internal density d*: quantity of internal edges
- *Conductance C*: proportion of links toward neighbouring communities
- *Triangles ratio TPR*: proportion of “well-integrated” nodes in the group

Thematic cohesion measures [Gadek, 2017]**:**

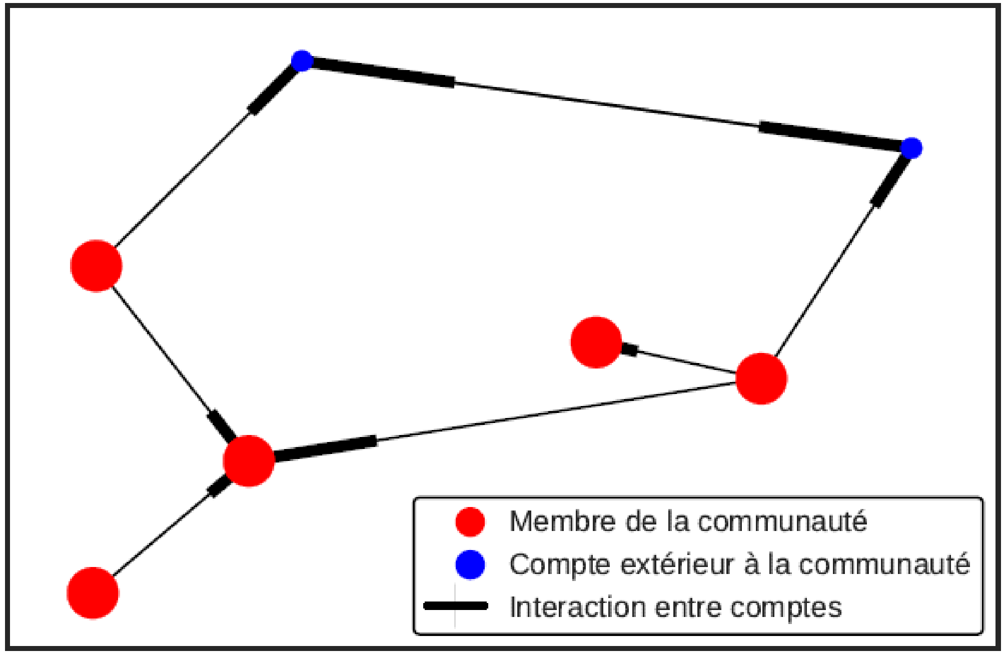
- *Expertise ξ:* strength of a topic in a group
- *Representativeness ρ:* strength of a group on a topic
- *Pertinence θf.igf:* relevance score of a group (similar to *tf.idf*)

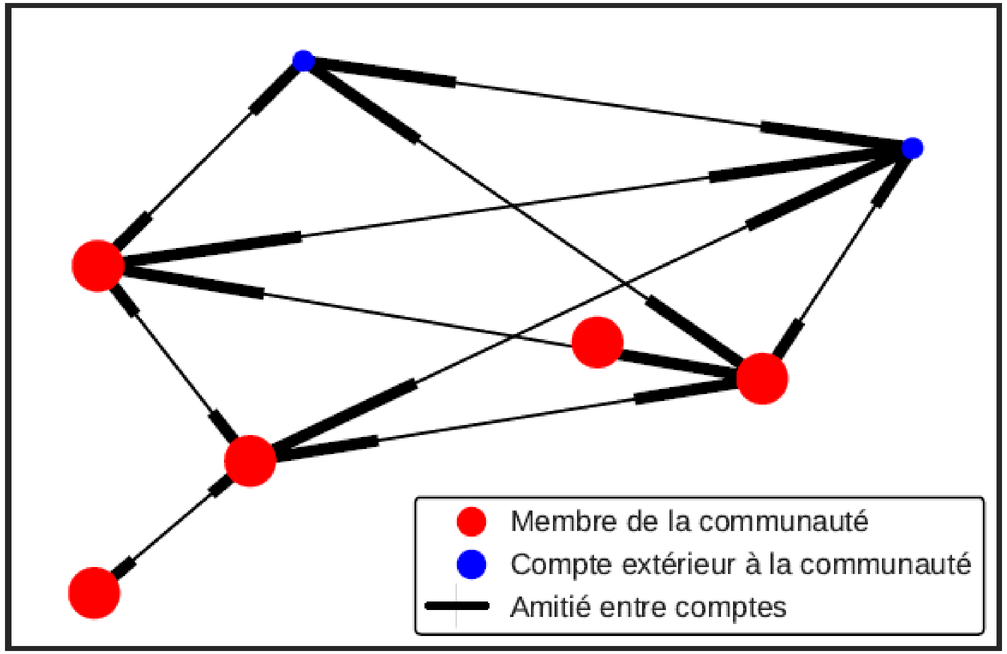
Detection and characterisation of communities





Detection and characterisation of communities


● Membre de la communauté
● Compte extérieur à la communauté
Interaction entre comptes



Community in the graph of interactions…

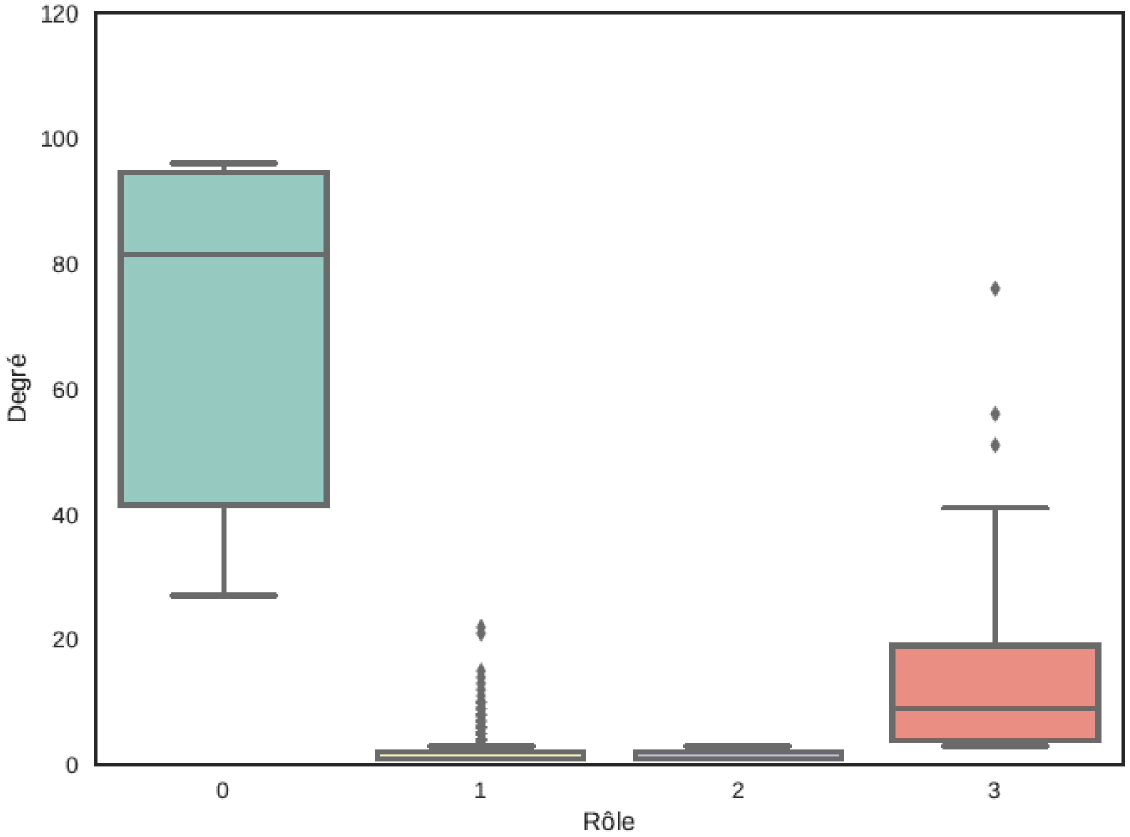
projected in the friendship graph.

Role of users in a network

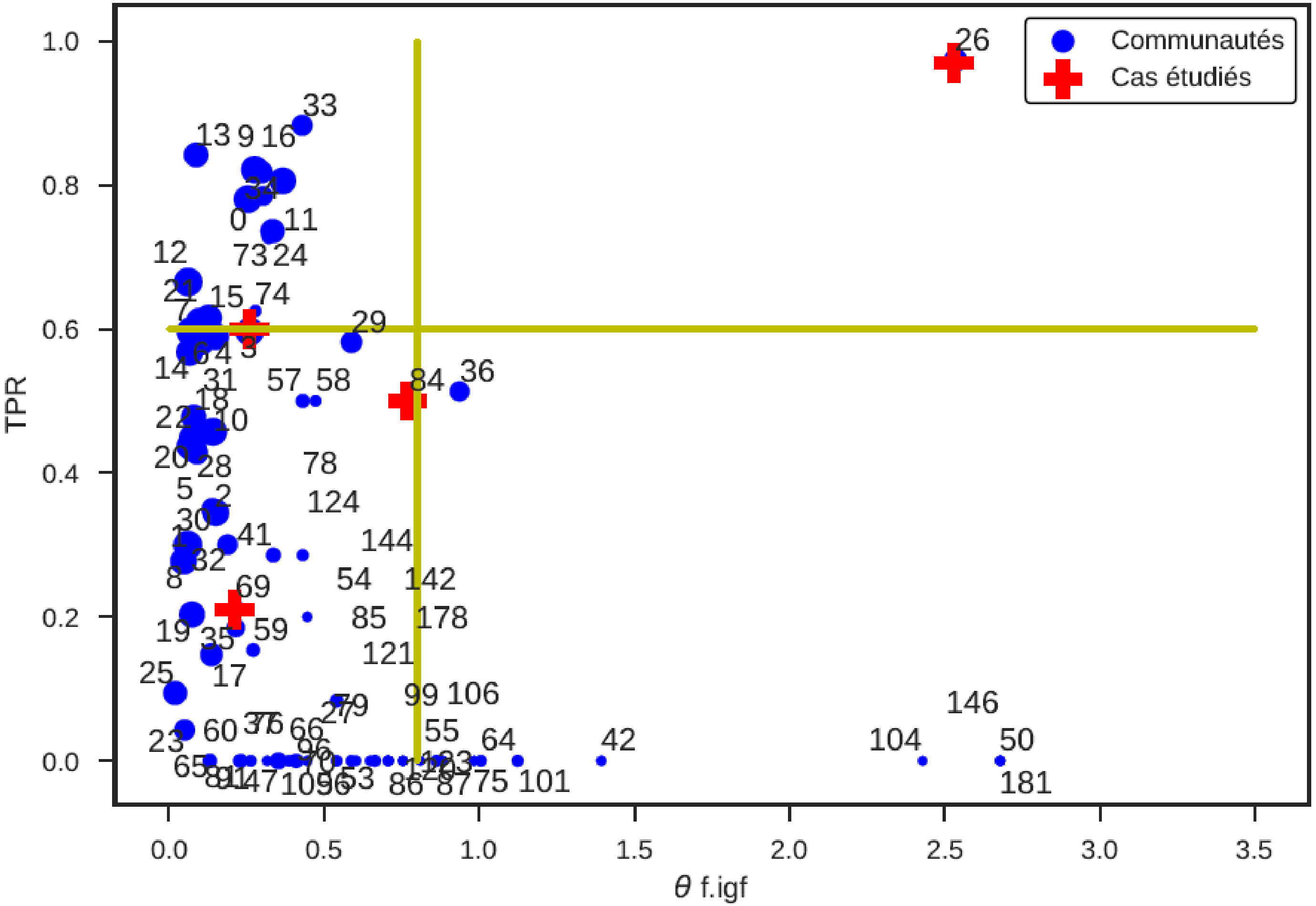
RolX [Henderson et al., 2012]

Unsupervised learning based on topological "scores" of each node's position:
in-/out- degree, centrality, ego-network centralities and betweennesses.

Predetermined number of 4 different roles.
Computed **specifically** on a graph (fig: G_Ω on Galaxy2)
→ Uneasy to interpret and exploit.



Case study on a tweets dataset: KevRandTweets



KevRandTweets:

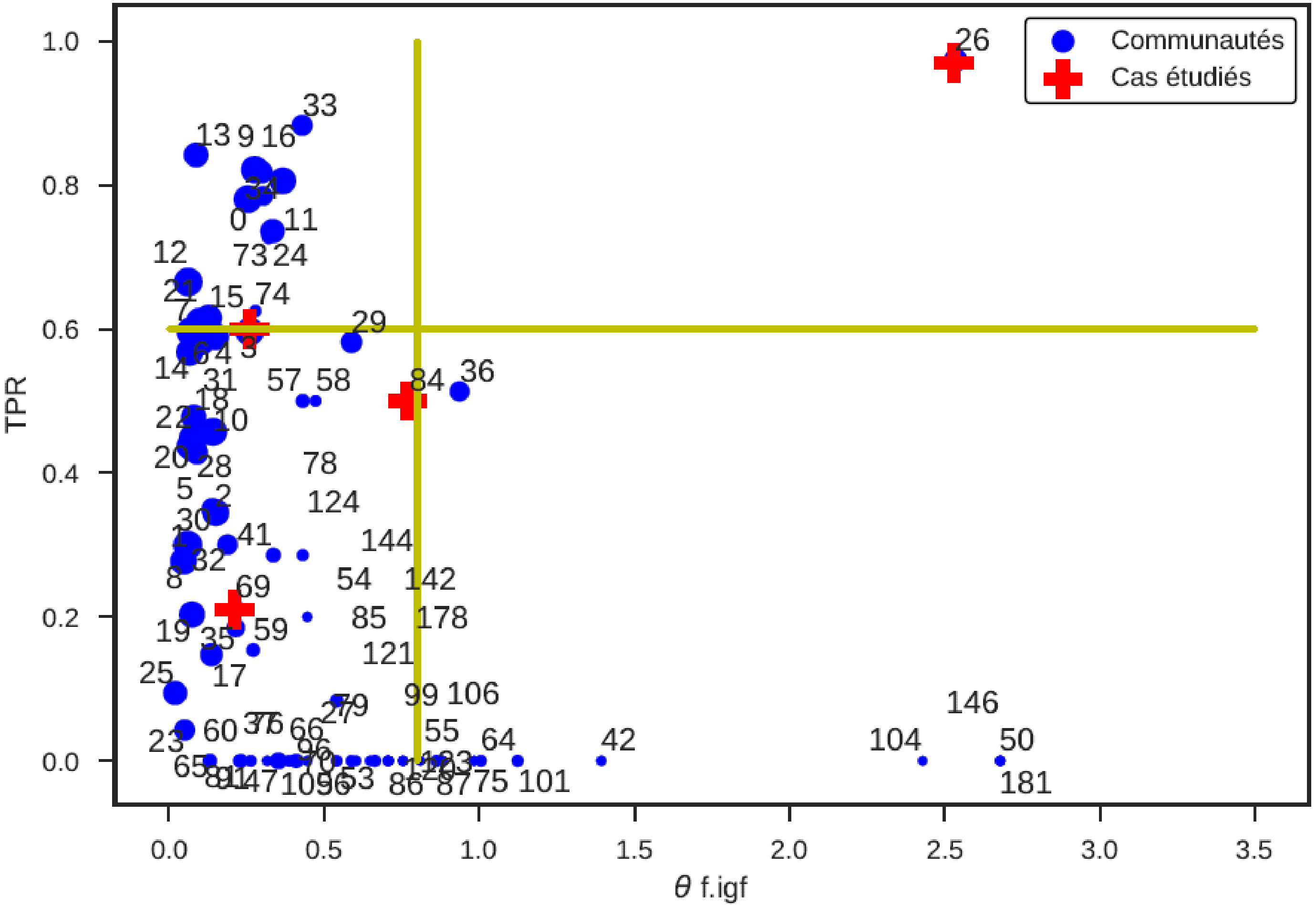
Almost 10 million tweets, containing

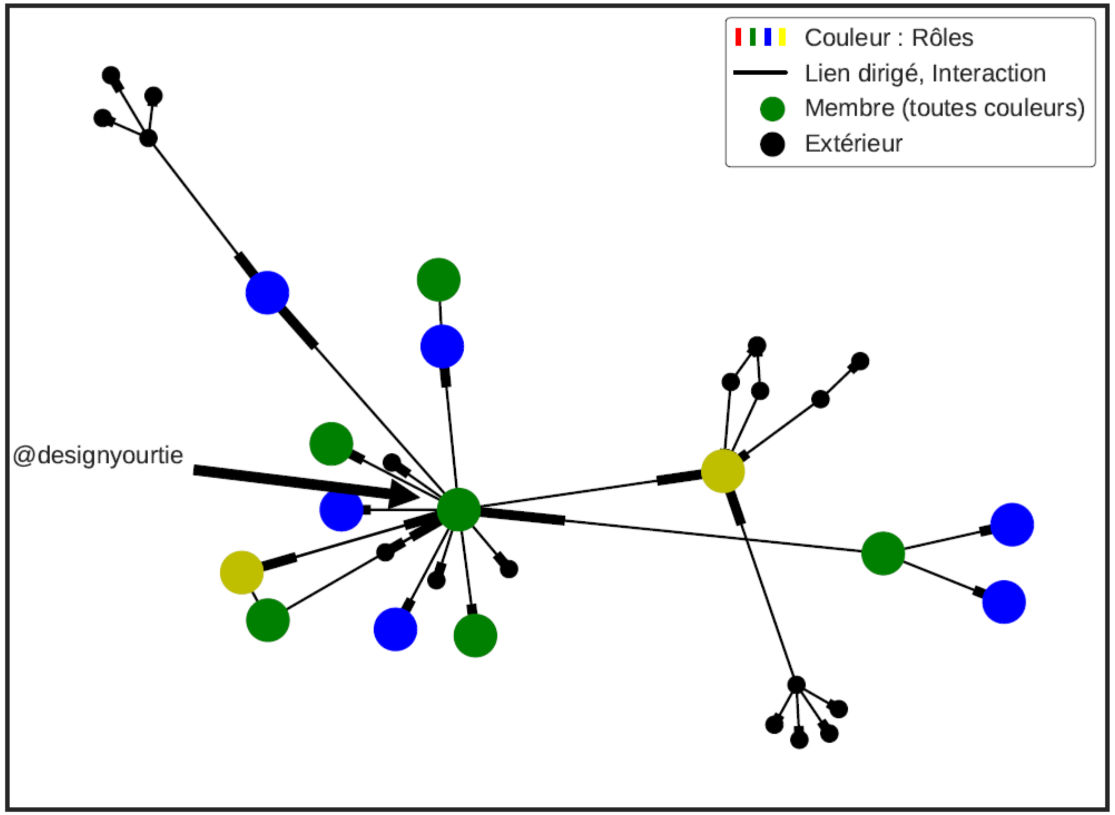
- Every action performed by 5,000 users,
- Every mention of these users

December 2016

US post-election context

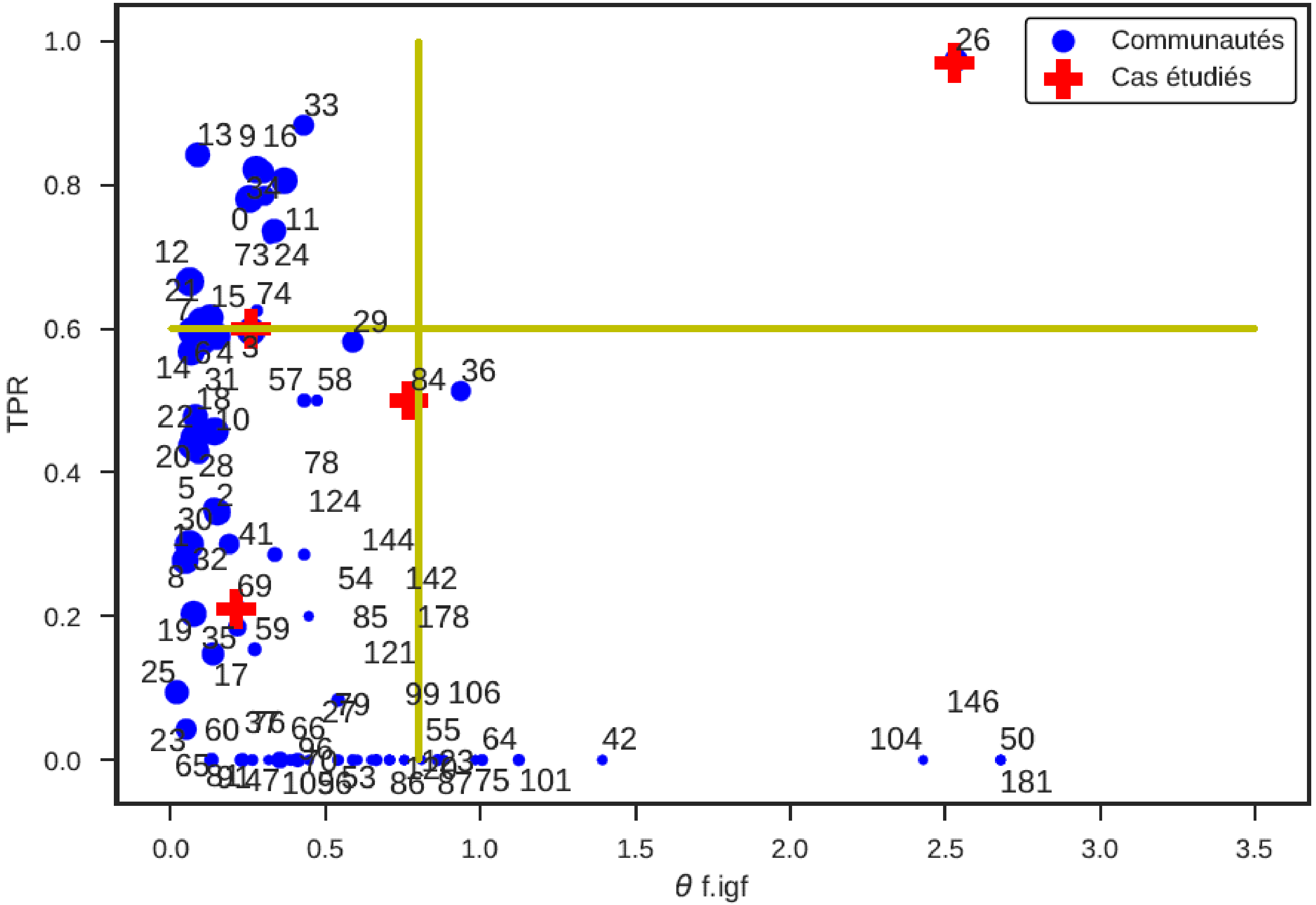
Case study on a tweets dataset: KevRandTweets

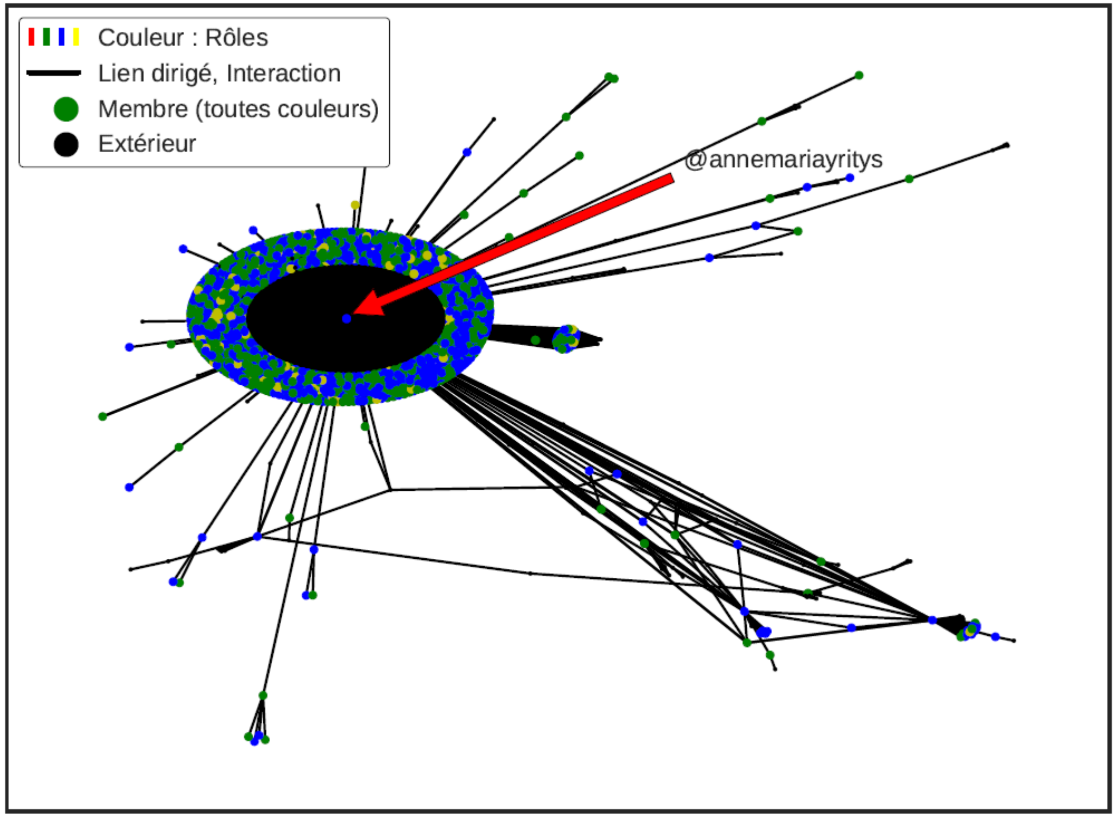




AIRBUS

Case study on a tweets dataset: KevRandTweets





Conclusion: AI uses for Social Media Intelligence

Three levels of analysis

- Textual: sentiment, emotion, topic
- User: role in the network, influence
- Groups: detection, impact, link strength, exploration

Social media : not only Facebook and Twitter

Quickly-evolving domain

- New challenges, and new requests, to be taken in account
- Task-specific modules to benchmark and integrate in a larger solution

General difficulties to obtain and protect the data

- GPDR & privacy
- Proprietary data, access limitations

WebLab – Smart Data Analytics Platform
Extracting valuable information from any source

Context of this work: a PhD thesis (INSA Rouen Normandie)

Thesis title:

Detection of opinions, key-actors and influent communities in online social media

Publications:

— Extracting contextonyms from Twitter for stance detection,
G. Gadek, J. Betsholtz, A. Pauchet, S. Brunessaux, N. Malandain and L. Vercouter, **ICAART**, 2017, Volume 2, 132-141.

— Topical cohesion of communities on Twitter,
G. Gadek, A. Pauchet, N. Malandain, K. Khelif, L. Vercouter and S. Brunessaux, **KES**, 2017, 10p.

— Measures for topical cohesion of user communities on Twitter,
G. Gadek, A. Pauchet, N. Malandain, K. Khelif, L. Vercouter and S. Brunessaux, **WebIntelligence**, 2017, 8p.

— AI techniques to analyse a social network on text, user and group level : application on Galaxy2,
G. Gadek, A. Pauchet, S. Brunessaux, K. Khelif and B. Grilheres, **APIA**, 2018, 9p.

— Topological and topical characterisation of Twitter user communities,
G. Gadek, A. Pauchet, N. Malandain, L. Vercouter, K. Khelif, S. Brunessaux and B. Grilheres, **Data Technologies & Applications Journal**, 2018, 20p.

Thank you

guillaume.gadek@airbus.com

AIRBUS