

À la recherche des *mini-publics* : un problème de communautés, de singularités et de sémantique

Éric Leclercq¹ Sergey Kirgizov¹ Maximilien Danisch²

¹Équipe Systèmes d'Information et Systèmes d'Images, LE2I CNRS UMR 6306, Université de Bourgogne Franche-Comté

²Télécom Paristech, Département Informatique et Réseaux

19 Janvier 2016

“Twitter aux élections européennes : une étude contrastive internationale des utilisations de Twitter par les candidats aux élections au Parlement Européen en mai 2014”

≈ 45 chercheurs (majoritairement politologues, sociologues, chercheurs en communication)

10 laboratoires de recherche

6 pays européens (France, Allemagne, Belgique, Italie, Espagne et Angleterre)

50M de tweets pour un volume total de 50Go

“Mini-Publics are defined as groups small enough to be genuinely deliberative. Mini publics serve as smaller circles of (better) informed groups, which engage in important information exchange processes and discourses”

– chercheurs en sciences de la communication

Mini-Publics == communautés ?

Quels critères, quelles caractéristiques ?

Mini-Publics de Twitter, 10 caractéristiques

- 1 **Hashtag-based (mostly)**
- 2 Spontaneity (e.g. sudden events)
- 3 Current topic (e.g. major breaking news)
- 4 User-initiated (non-supervised, bottom-up, unannounced)
- 5 Rapid growth (virality)
- 6 High intensity
- 7 Undetermined participants/actors
- 8 Structural openness
 - Dimension 1 : On Twitter, each individual is able to create and use whatever hashtags he feels is appropriate (no approval from administrators is needed)
 - Dimension 2 : Principle of open participation : every user can contribute join a Twitter discussion (e.g., marked by a certain hashtag) and help spreading the word
- 9 Short duration
- 10 Volatility

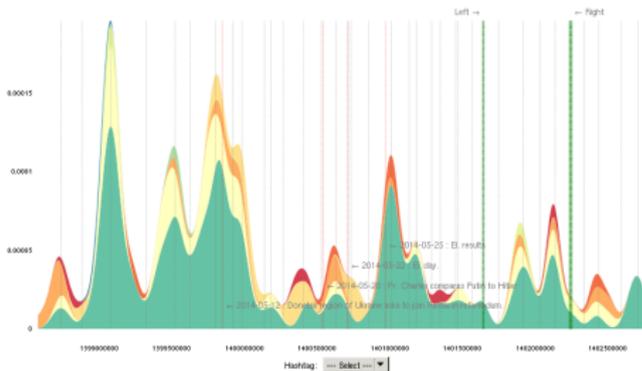
Densité temporelle des hashtags

Tweets Observatory: tweets with **Crimea's hashtags** emitted during #EE2014

Tweets emitted by [french](#) / [german](#) / [english](#) / [russian](#) candidates. [All tweets](#)

You can move Left and right borders of the time window.
You may click the button.
You may wait for results.
(Last steps in [unix-manch](#) format).

Time density



Time window:

Left border:

Right border:

MAKE ME HAPPY!

Tweets

timestamp0	id	from_user_name	from_user	text
1	1401838294			
2	1401879572			
3	1401881487			
4	1401919735			
5	1401921191			
6	1401924247			

Top 10 hashtags

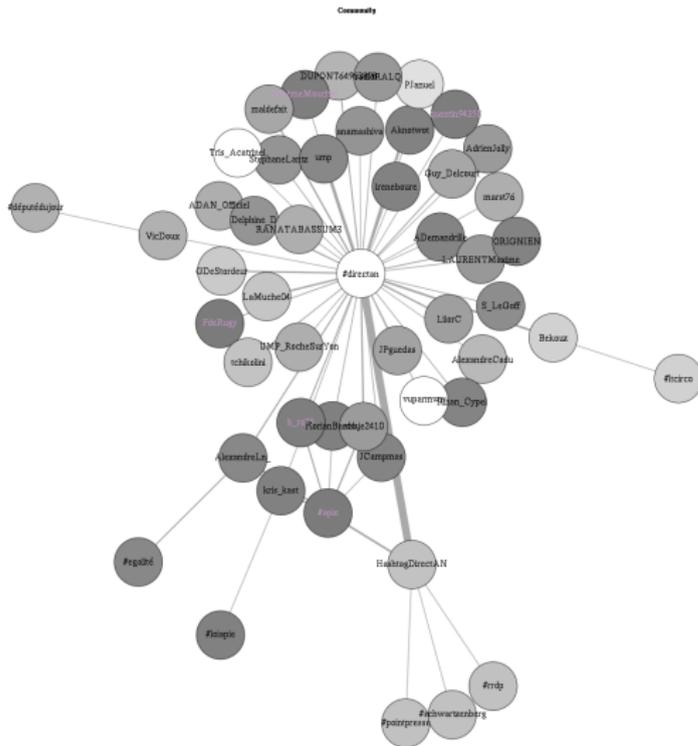
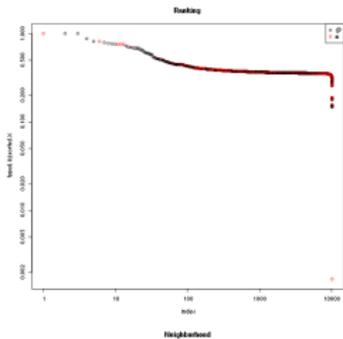
count	hashtag
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Top 10 users

count	from_user_name	from_user
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

- Les Mini-publics ressemblent à de petites communautés
- Les algorithmes existants les plus connus :
 - ne reposent pas une définition de communauté spécifique, c'est-à-dire basée sur le domaine, mais plutôt une définition abstraite
 - généralement ne détectent pas de petites communautés
 - ne considèrent pas les communautés chevauchantes
- **Point de départ** : Algorithme “l'opinion propagée” pour la détection de communautés egocentrées
[Danisch, Guillaume, Legrand, 2012]
- **Idées principales** :
 - mesurer la proximité des nœuds d'intérêt à tous les autres nœuds du graphe “hashtags – utilisateurs” avec les liens pondérées par l'intensité d'utilisation de hashtags.
 - proximité \implies classement

Nœud d'intérêt : un hashtag (#directan)



L'algorithme utilise des liens
entre les nœuds pour
"propager" l'information virtuel
d'un nœud à ses voisins

Algorithme "l'opinion propagée"

Matrice d'adjacence A .

Matrice de transition $M_{ij} = A_{ij} / \sum_j A_{ij}$.

Vecteur x qui contient les opinions de nœuds.

1 Initialisation $x_0 = (1, 0, 0, 0, \dots)$

2 Moyennage $x \mapsto xM$

3 Mise à l'échelle $x \xrightarrow{g} \frac{x - \min(x)}{1 - \min(x)}$

4 Réinitialisation $x \xrightarrow{f} x + x_0 - x \cdot x_0$

$$x_t = f\left(g(x_{t-1}M)\right)$$

Algorithme itératif : $x_0 \mapsto x_1 \mapsto x_2 \mapsto \dots$ convergence ! ?

Malheureusement, cet
algorithme ne converge pas
pour les graphes bipartis (e.g.
un graphe de Twitter
“hashtags – utilisateurs”)

Heureusement, on sait
comment résoudre ce
problème !

Les propriétés théoriques : convergence

Conjecture

L'algorithme converge si la chaîne de Markov associée à la matrice de transition M est non-périodique.

Confirmé expérimentalement. Il y a une idée de la preuve formelle.

Pour assurer la convergence, on ajoute des boucles :

$$A \mapsto A + \begin{pmatrix} \alpha & & & \\ & \alpha & & \\ & & \dots & \\ & & & \alpha \end{pmatrix}$$

Les propriétés théoriques : vitesse de convergence

Les valeurs propres de la matrice M jouent un rôle principal dans l'évaluation de la vitesse de convergence.

Afin d'évaluer l'impact du poids de boucles α sur les valeurs propres, et donc sur la vitesse de convergence, on utilise les inégalités de Weyl.

[“Honeycombs and Sums of Hermitian Matrices”, Knutson, Tao, 2000]

Informellement,

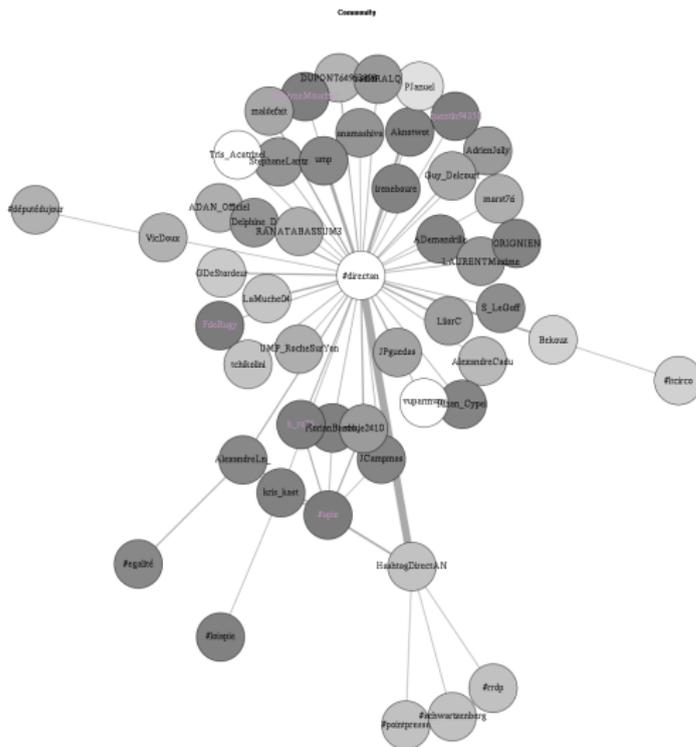
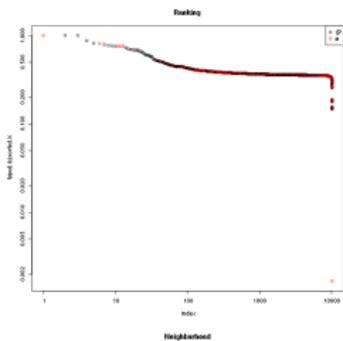
- plus α est grande, plus rapidement ça va converger.
- mais... en augmentant la valeur de α , nous nous éloignons du graphe original.

Pour assurer la convergence, on ajoute des boucles :

$$A \mapsto A + \begin{pmatrix} \alpha & & & \\ & \alpha & & \\ & & \dots & \\ & & & \alpha \end{pmatrix}$$

Exemples

Nœud d'intérêt : un hashtag (#directan)



Résultats

- ♣ Une certaine formalisation du concept de mini-publics
- ★ Algorithme adapté pour la détection de mini-publics
- ♦ Solution pour le problème de la convergence

Travaux futurs

- ♣ Comparaison avec d'autres algorithmes (par exemple, pagerank enraciné)
- ♠ Comparaison avec la vérité du terrain (?)

Merci de votre attention.

Questions ?