

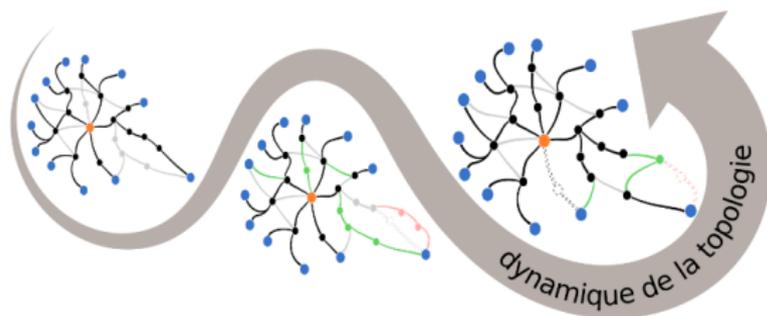
Densité temporelle des réseaux complexes: détection et visualisation de l'évolution de la structure ego-communautaire

Sergey Kirgizov

UBFC

23 Mars 2017

Thèse : Analyse empirique et modélisation de la dynamique de la topologie de l'internet



complexnetworks.fr

We are interested in all aspects of real-world networks and their models, from internet measurements to random graphs, from social network analysis to spreading phenomena, and from graph algorithms to biological networks.



UPMC
SORBONNE UNIVERSITÉS

Postdoc et ATERs



Thèmes :

- Analyse et modélisation des réseaux complexes
- Théorie des graphes
- Combinatoire énumérative

Publications : analyse des réseaux sociaux, dynamique de la topologie de l'Internet, bioinformatique, apprentissage par renforcement, informatique théorique.

1. Nos projets : TEE 2014 et PEPS 2015-2016
2. Structures communautaires
3. Densité temporelle des réseaux complexes & Évolution de la structure communautaire
4. Conclusion & Discussion

“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

Graph-streaming pour l'étude de la dynamique des sphères **mediatiques** et politiques

Participants :

Le2i CombNet :

Benoit Darties (porteur), Olivier Togni, Nicolas Gastineau

Le2i SISI : Eric Leclercq, Sergey Kirgizov

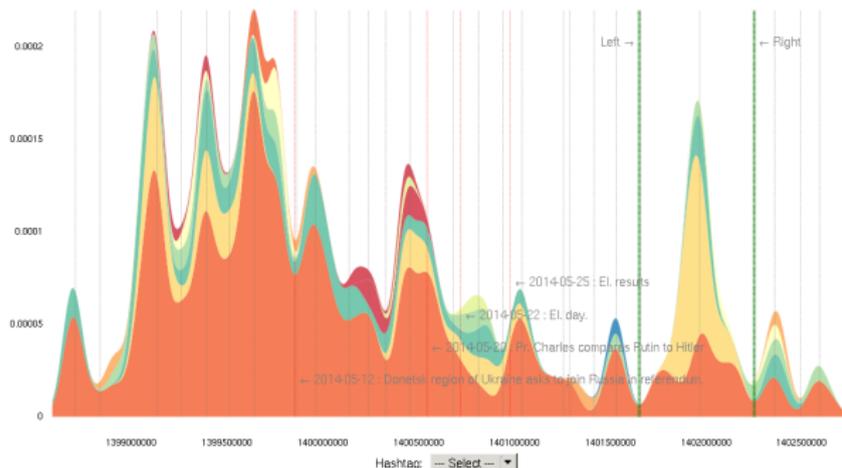
LIRIS GOAL : Hamamache Kheddouci, Hamida Seba Lagraa

CIMEOS 3S et TIL : Gilles Brachote, Alexander Frame et Tatiana Kondrashova

Observatoire de la dynamique de Twitter

Densité temporelle des hashtags + caractérisation des périodes sélectionnées

Time density



Top 10 hashtags

50 #hashtag1
60 #HASHTAG3
60 #hashtag2
...

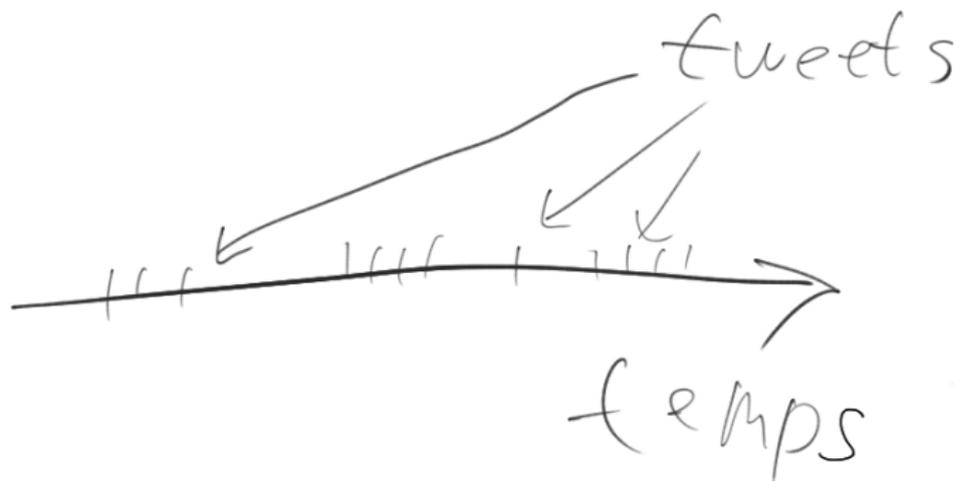
Top 10 users

500 Ya
400 Ty
300 On
200 Ona
...

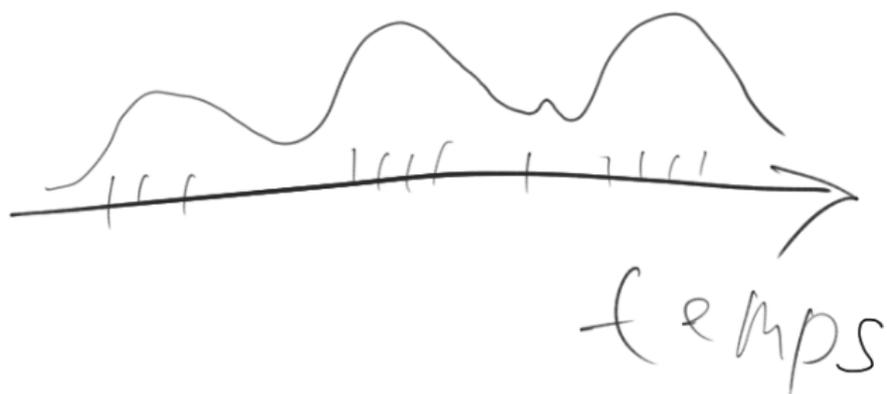
Tweets

1398777777: User1: bla-bla #hashtag1 #hashtag2
1398777778: User1: bla-bla #HASHTAG3
1398777788: User2: bla-bla #hashtag2
...

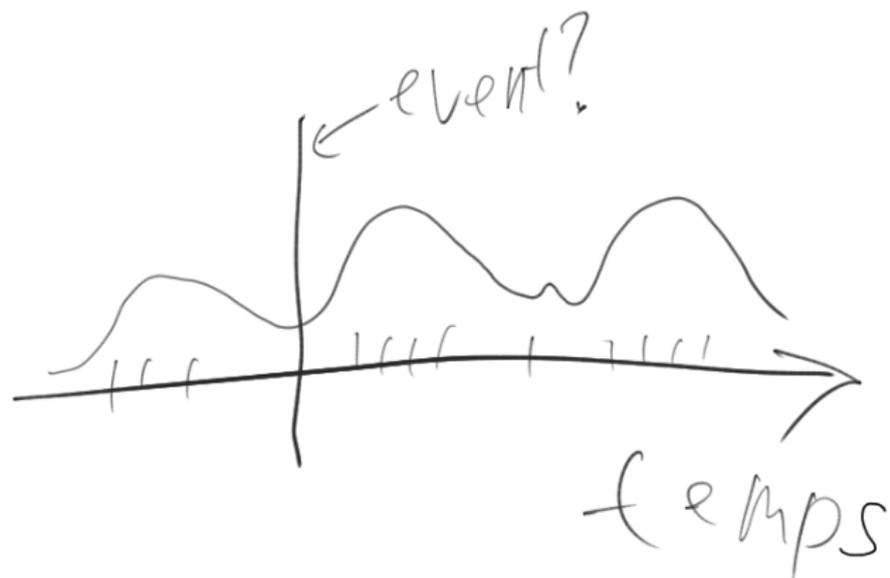
Densité temporelle des hashtags



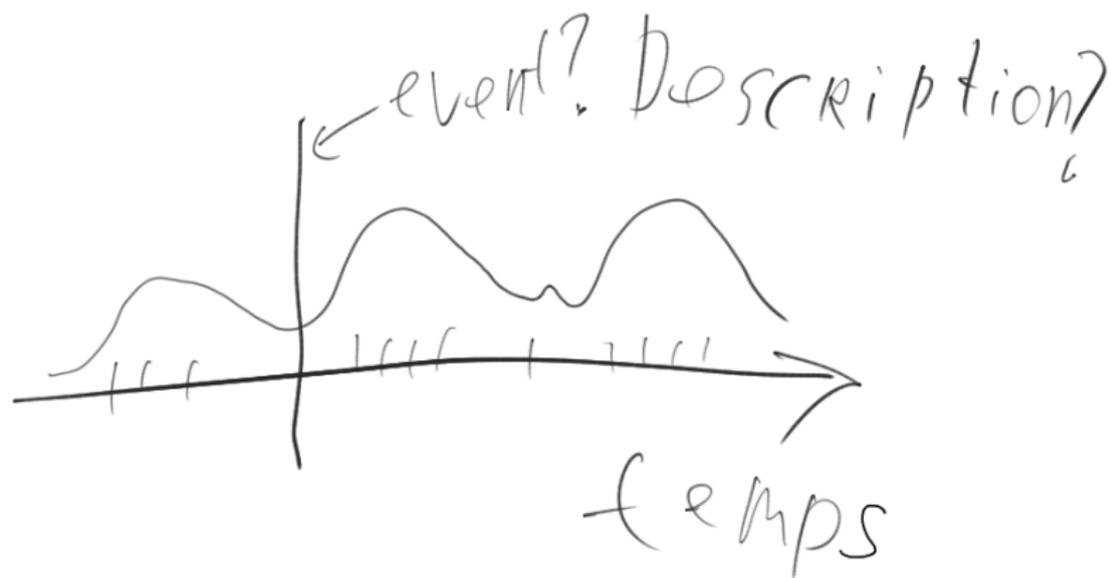
Parkzen - Rozenblatt



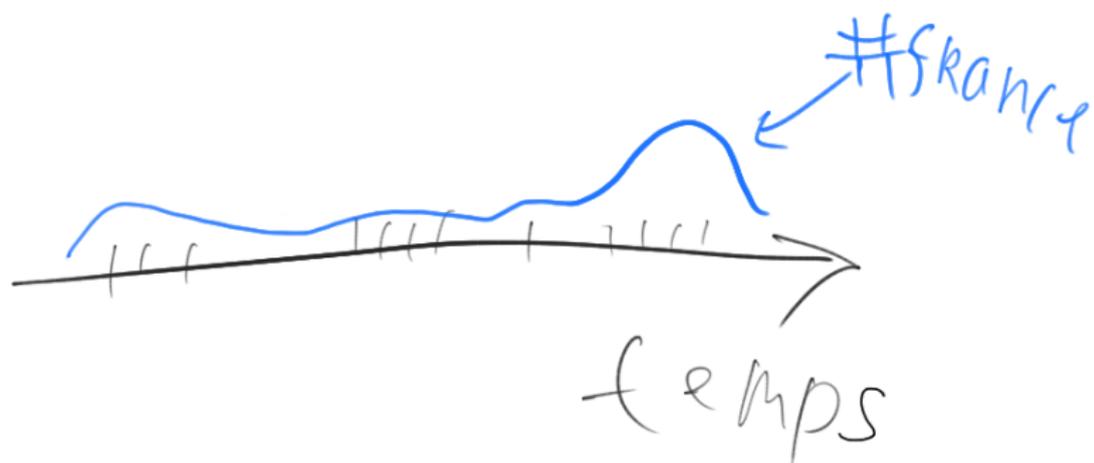
Parkzen - Rozenblatt



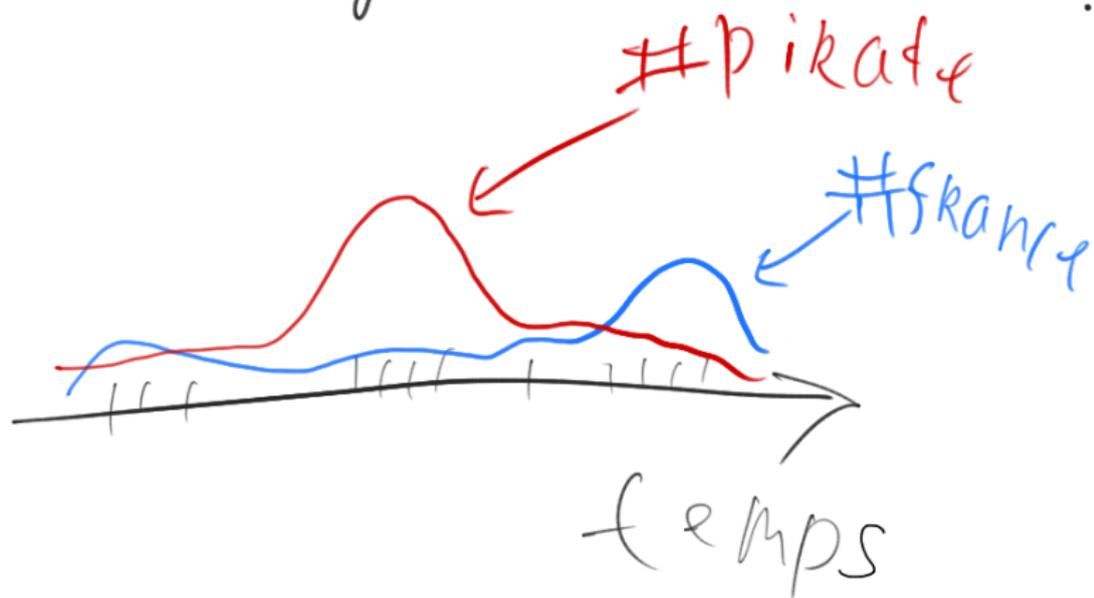
Parkzen - Rozenblatt

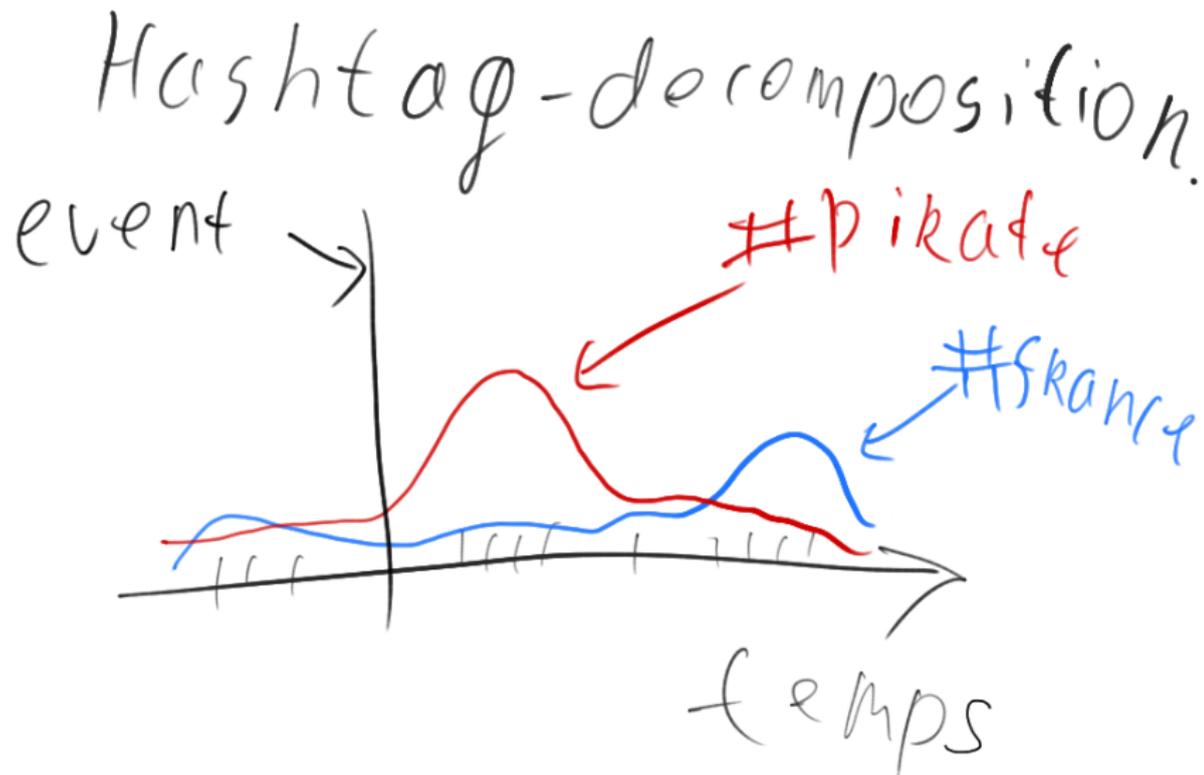


Hashtag-décomposition.



Hashtag-décomposition.

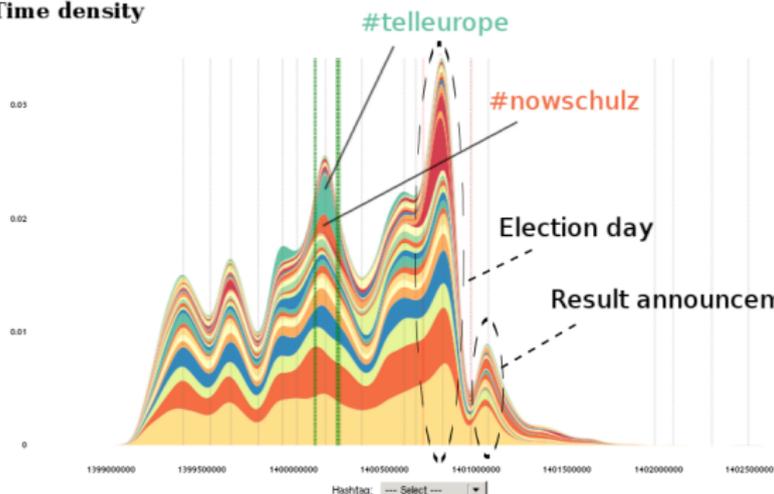




Observatoire de la dynamique de Twitter

Densité temporelle des hashtags + caractérisation des périodes sélectionnées

Time density



Time window:

Left border:



Right border:



MAKE ME HAPPY!

Top 10 hashtags

count	hashtag
1	570 ep2014
2	443 telleurope
3	438 ee2014
4	256 notreurope
5	237 europeennes2014
6	200 eelv
7	190 nowschulz
8	108 europe
9	91 tafta
10	85 ump

Top 10 users

count	from_user_name	from_user
1	121 JS Herpin	@jsherpin
2	108 Elodie Massé	@masselodie
3	81 Yannick Jadot	@yjadot
4	79 Béatrice DELGENDRE	@bdelgendre
5	74 Guillaume Balas	@BalasGuillaume
6	53 Raquel Garrido	@RaquelGarridoPG
7	47 Aleksander GLOGOWSKI	@Aleks_Paris
8	44 Sandrine Belier	@sandrinebelier
9	39 Dolores BAUDELLOT	@Yodado
10	38 Franck Proust	@frankproust

Publications et talks sur ce sujet

- 1** Un observatoire pour la modélisation et l'analyse des réseaux multirelationnels : une application à l'étude du discours politique sur twitter Ian Basaille, SK, Éric Leclercq, Marinette Savonnet, Nadine Cullot, Thierry Grison, and Elisabeth Gavignet *Revue Document Numérique*, 2017+. à paraître
- 2** Towards a Twitter Observatory : A multi-paradigm framework for collecting, storing and analysing tweets. Ian Basaille, SK, Éric Leclercq, Marinette Savonnet, et Nadine Cullot, *RCIS 2016*
- 3** A la recherche des mini-publics : un problème de communautés, de singularités et de sémantique. Eric Leclercq, SK et Maximilien Danisch, *journée "Données Participatives et Sociales"*, EGC 2016, Reims, 2016
- 4** (Re)constuire la temporalité d'un événement médiatique sur Twitter : une étude contrastive. Tatiana Kondrashova, Alex Frame et SK, *XXe congrès de la SFSIC : Temps, temporalités et information-communicatio*, 2016 Metz
- 5** Twitter in mediatized society : the dynamics of news circulation through politicians' tweets. Alexander Frame et Tatiana Kondrashova, *6th International "Language In The Media" Conference, September 2015, Hamburg*
- 6** SNFreezer : a Platform for Harvesting and Storing Tweets in a Big Data Context. *chapitre du livre*
- 7** Approche multi-paradigmes pour l'analyse et la caractérisation des réseaux sociaux complexes Éric Leclercq, *Séminaire Traitement de l'information multimodale et "Big Data" Direction Générale de l'Armement, Arcueil, October 2015*
- 8** A web application for event detection and exploratory data analysis for Twitter data, SK, Eric Leclercq, Marinette Savonnet, Alexander Frame, Ian Basaille-Gahite *Twitter at the European Elections 2014 : International Perspectives on a Political Communication Tool*, Dijon, 2015

Étude de l'évolution de communautés dans un graphe dynamique

1. Nos projets : TEE 2014 et PEPS 2015-2016
2. Structures communautaires
3. Densité temporelle des réseaux complexes & Évolution de la structure communautaire
4. Conclusion & Discussion

L'immense article de survey
de Santo Fortunato
"Community detection in graphs"
contient 500 références et 103
pages

HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.



SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.

- 1 Centralité spectrale (eigenvector centrality)
- 2 Mincut
- 3 **Pagerank**
- 4 Shi–Malik : normalised relaxed mincut
- 5 Pons–Latapy : Walktrap
- 6 Danisch–Guillaume–Le Grand : Carryover opinion
- 7 Méthodes basées sur la propagation de la chaleur
- 8 HITS de Kleinberg
- 9 ...

V : ensemble des nœuds

E : ensemble des liens

G : graphe

A : matrice d'adjacence du graphe G , éventuellement pondéré

Pagerank personnalisé

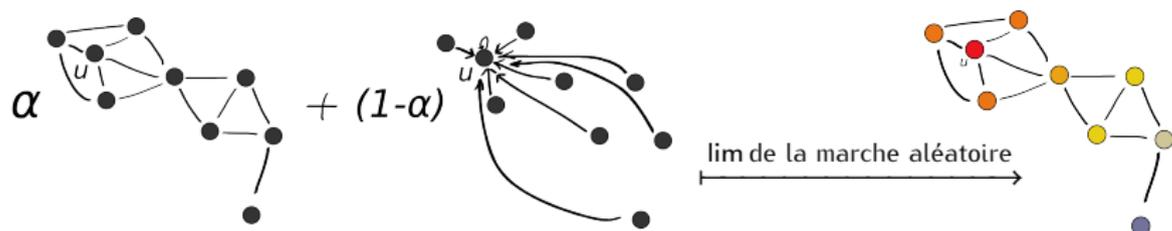
$M = (D^{-1}A)^T$: matrice des transitions

x_0 : distribution initiale de probabilité définie sur les nœuds

α : un paramètre de téléportation

$Y = (1 - \alpha)x_0\mathbf{1}^T$: matrice de téléportation

$\hat{M} = \alpha M + (1 - \alpha)Y$: matrice de google



Pagerank personnalisé est un vecteur r t.q. $r = \hat{M}r$

The anatomy of a large-scale hypertextual Web search engine

Brin et Page, 1998

Pagerank personnalisé, trois interprétations

Soit M une matrice symétrique, stochastique, irréductible et apériodique :

Convergence de la chaîne de Markov,
Marche aléatoire sur graphe

$$r = \lim_{n \rightarrow \infty} M^n x_0$$

Interprétation spectrale

$$r = v_{max}$$

vecteur propre associé à la plus grande valeur propre λ_{max} de M .

Optimisation

$$\operatorname{argmax}_{\|x\|=1} \sum M_{ij} x_i x_j = v'_{max}, \quad v_{max} = \frac{v'_{max}}{\|v'_{max}\|_1}$$

Si la matrice M est non symétrique, c'est un peu plus compliqué, car l'interprétation spectrale ne marche pas directement. Pour montrer la relation entre les coupures minimales et les vecteur propres les gens font une certaine symétrisation.

Voir par exemple,

The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume

1990 Lovász et Simonovits

Local partitioning for directed graphs using pagerank, 2007

Reid Andersen, Fan Chung et Kevin Lang

Réseaux complexes du monde réel



Pagerank personnalisé trouve de bonnes communautés (même si elles se chevauchent) dans les réseaux complexes du monde réel (DBLP, Youtube, Amazon)

Community membership identification from small seed sets

Kloumann et Kleinberg, SIGKDD, 2014

Overlapping Community Detection Using Neighborhood-Inflated Seed Expansion

Whang, Gleich, Dhillon, 2015¹⁸

1. Nos projets : TEE 2014 et PEPS 2015-2016
2. Structures communautaires
3. Densité temporelle des réseaux complexes & Évolution de la structure communautaire
4. Conclusion & Discussion

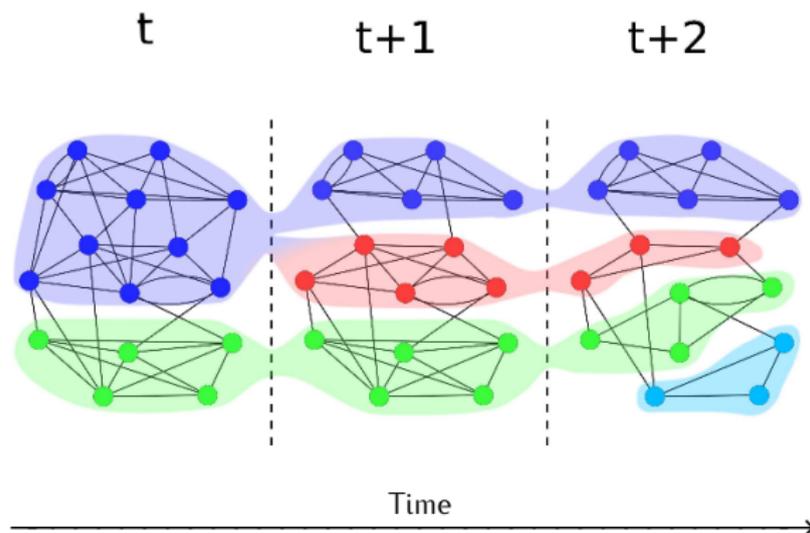
Graphes dynamiques

- Snapshots [Hopcroft et al., 2004, Leskovec et al., 2005],
- Time-varying graphs [Casteigts et al., 2012, Wehmuth et al., 2013]
- Flot de liens [Viard et al., 2016]

Dans ces modèles, les changements de la structure sont des changements discrets, c'est-à-dire la fonction de présence des liens est de type $\text{Temps} \times V^2 \rightarrow \{0, 1\}$.

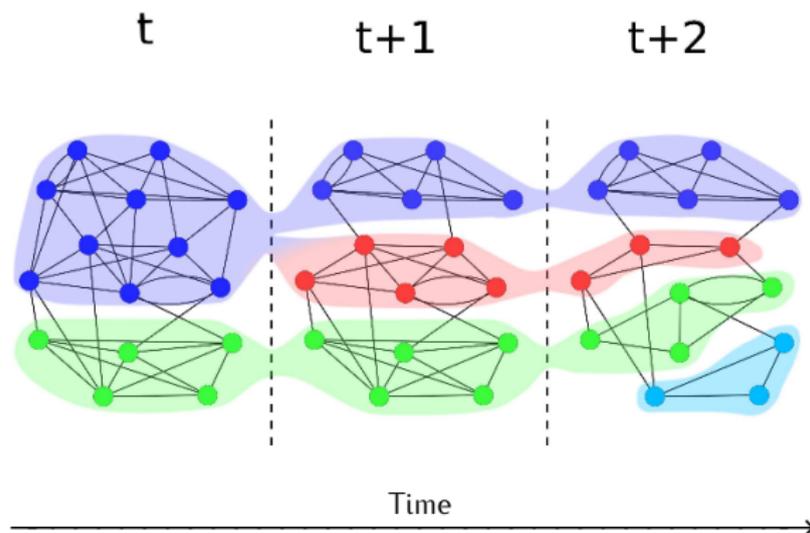
Evolution de communautés classiques (snapshots)

Detection/visualisation basées sur le snapshots



Evolution de communautés classiques (snapshots)

Detection/visualisation basées sur le snapshots

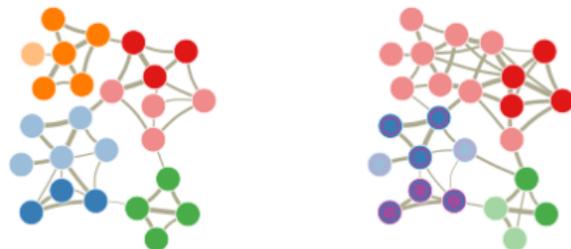


Cette visualisation ne convient pas aux grands graphes / grand nombre des shapshots

Evolution de communautés classiques (snapshots)

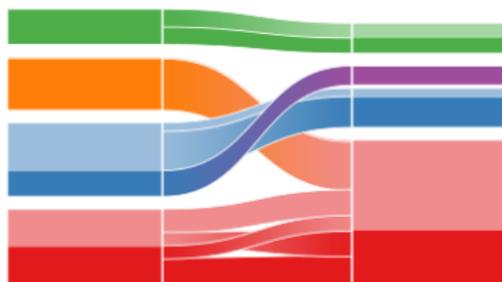
Mapping change in large networks

Rosvall, Bergstrom, 2010



Mapping
change

Change over time

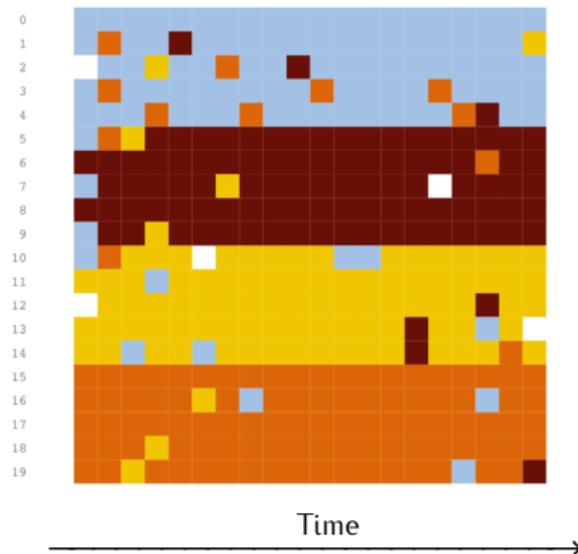


Time 1

Time 2

Une autre méthode de visualisation

Les lignes correspondent aux nœuds, couleurs aux différentes communautés.

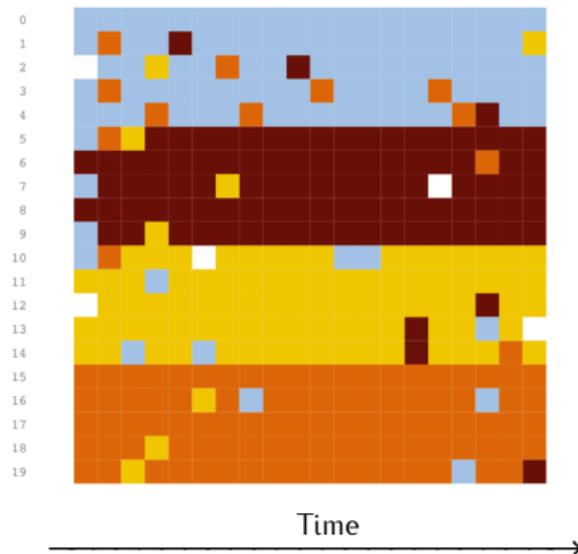


Intrinsically Dynamic Network Communities

by Mitra, Tabourier, Roth 2011
22

Une autre méthode de visualisation

Les lignes correspondent aux nœuds, couleurs aux différentes communautés.



En général, les structures des réseaux complexes évoluent plus doucement

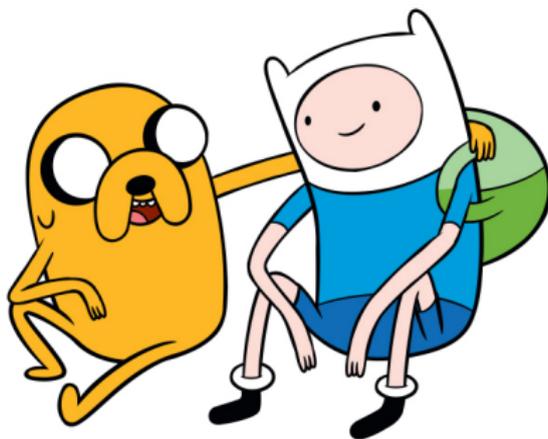


Détecter les communautés dans les réseaux dynamiques et visualiser leur évolution !



Ce n'est pas si facile

Détecter les communautés dans les réseaux dynamiques et visualiser leur évolution !



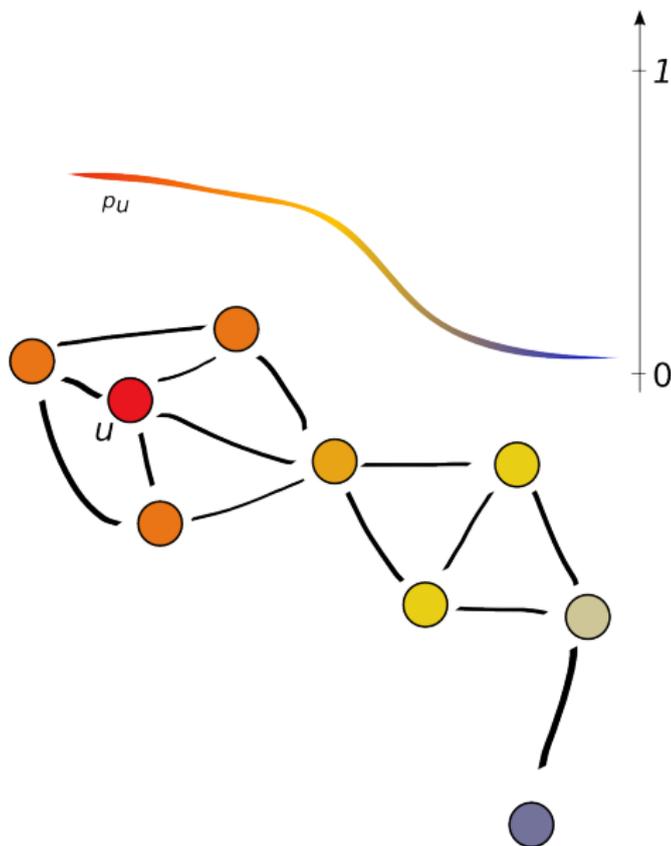
Au lieu de cela, nous considérons
l'évolution de la structure
ego-communautaire

Définition d'une structure ego-communautaire

Le nœud u est un centre de la communauté.

Fonction $p_u : V \rightarrow [0, 1]$ donne la probabilité que le nœud v soit dans la communauté du nœud u .

Le rouge signifie une grande probabilité, le bleu signifie une petite.



Notre approche

- 1 Lisser les données d'entrée discrets
- 2 Couper les données lissées en tranches de temps
- 3 Faire du pagerank pour toutes tranches de temps
- 4 Visualiser

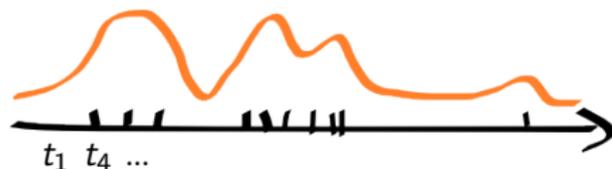
Un nœud interagit avec un autre nœud à l'instant t .

a	b	t_1
c	b	t_2
d	c	t_3
a	b	t_4
d	b	t_5
	...	

Flot de liens entre les nœuds a et b

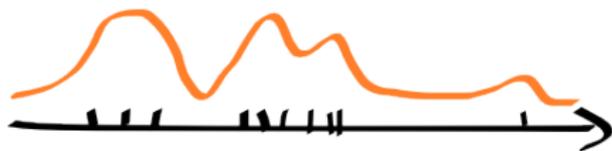
a	b	t_1
c	b	t_2
d	c	t_3
a	b	t_4
d	b	t_5
	...	

se transforme en



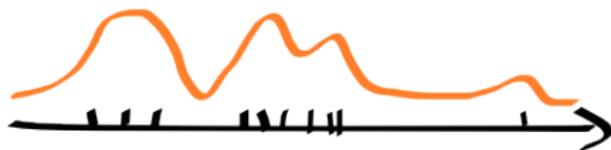
(par Parzen–Rosenblatt méthode)

Densité temporelle entre a et b



Lisser les données d'entrée

Densité temporelle entre a et b

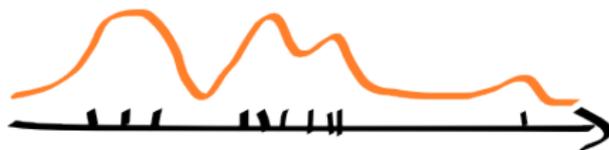


Densité temporelle entre c et b



Lisser les données d'entrée

Densité temporelle entre a et b



Densité temporelle entre c et b



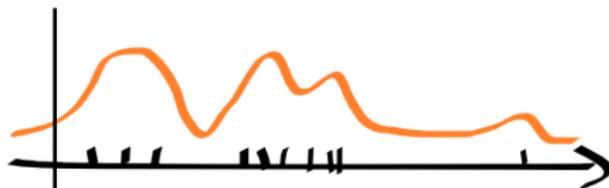
Densité temporelle entre a et c



...

Couper & pagerank

Densité temporelle entre a et b



Densité temporelle entre c et b



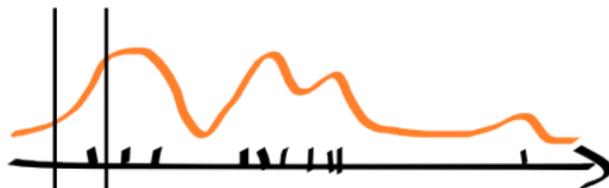
Densité temporelle entre a et c



...

Couper & pagerank

Densité temporelle entre a et b



Densité temporelle entre c et b



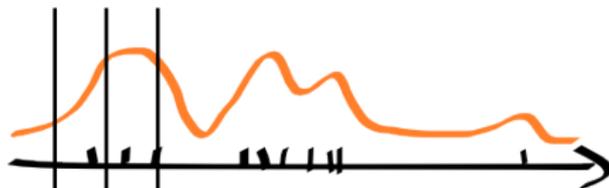
Densité temporelle entre a et c



...

Couper & pagerank

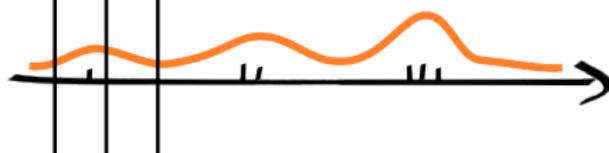
Densité temporelle entre a et b



Densité temporelle entre c et b



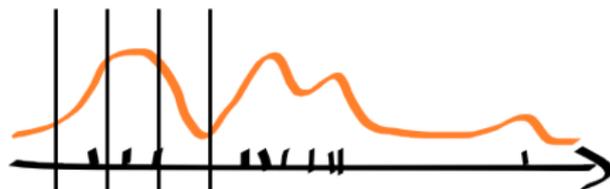
Densité temporelle entre a et c



...

Couper & pagerank

Densité temporelle entre a et b



Densité temporelle entre c et b



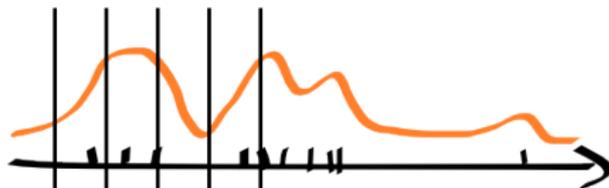
Densité temporelle entre a et c



...

Couper & pagerank

Densité temporelle entre a et b



Densité temporelle entre c et b



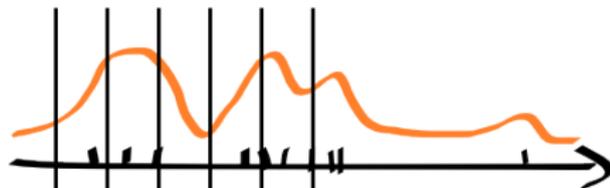
Densité temporelle entre a et c



...

Couper & pagerank

Densité temporelle entre a et b



Densité temporelle entre c et b



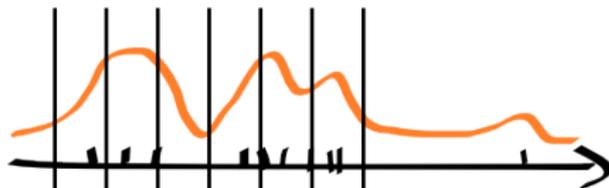
Densité temporelle entre a et c



...

Couper & pagerank

Densité temporelle entre a et b



Densité temporelle entre c et b



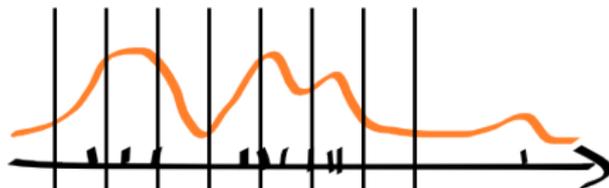
Densité temporelle entre a et c



...

Couper & pagerank

Densité temporelle entre a et b



Densité temporelle entre c et b



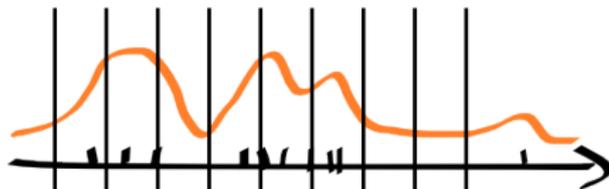
Densité temporelle entre a et c



...

Couper & pagerank

Densité temporelle entre a et b



Densité temporelle entre c et b



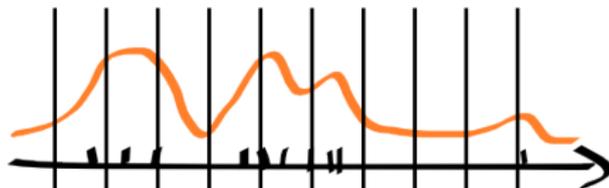
Densité temporelle entre a et c



...

Couper & pagerank

Densité temporelle entre a et b



Densité temporelle entre c et b



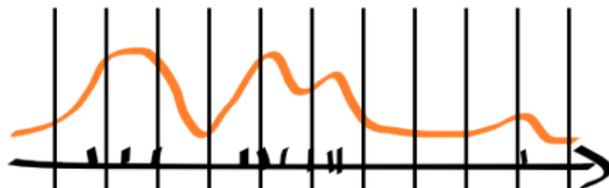
Densité temporelle entre a et c



...

Couper & pagerank

Densité temporelle entre a et b



Densité temporelle entre c et b



Densité temporelle entre a et c



...

1. Couper les données lissées en tranches de temps
2. Pour chaque timeslice effectuer un pagerank personnalisé en utilisant des valeurs lissées comme des poids des liens.

Évolution de la structure communautaire

Maintenant, la matrice d'adjacence dépend du temps. Et donc, la structure ego-communautaire aussi.

$p_{u,v}(t)$: probabilité que le nœud v soit dans la communauté du nœud u à l'instant t .

$p_{u,v}(t) =$ 

Théorème

$A(t)$ est lisse $\Rightarrow p_{u,v}(t)$ est lisse.

Schéma de la preuve

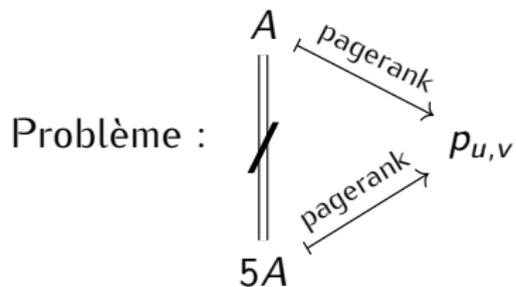
Pagerank : $A(t) \mapsto \hat{M}(t) \mapsto p_{u,v}(t)$

\hat{M} est irréductible et apériodique

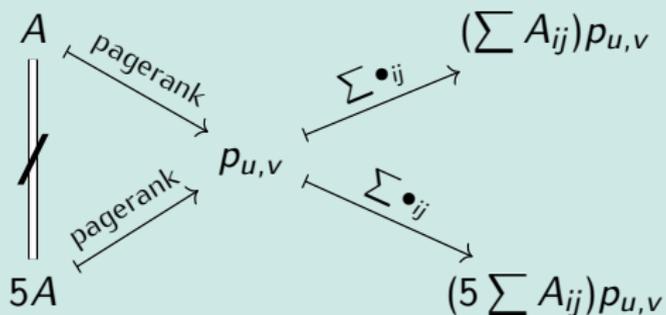
Théorème 3.2 de **A Note on Perturbations of Stochastic Matrices**

Huppert et Willems, 2000

Dénormalisation



Solution : dénormalisation !



SocioPatterns

[ABOUT](#) | [GALLERY](#) | [PUBLICATIONS](#)

DATASET: Primary school temporal network data

Release data: Sep 30, 2015

This data set contains the temporal network of contacts between the children and teachers used in the study published in BMC Infectious Diseases 2014, 14:695. The file contains a tab-separated list representing the active contacts during 20-second intervals of the data collection. Each line has the form "t i j Ci Cj", where i and j are the anonymous IDs of the persons in contact, Ci and Cj are their classes, and the interval during which this contact was active is [t – 20s, t]. If multiple contacts are active in a given interval, you will see multiple lines starting with the same value of t. Time is measured in seconds.

Terms and conditions

The data are distributed to the public under a [Creative Commons Attribution-NonCommercial-ShareAlike license](#). When this data is used in published research or for visualization purposes, please cite the following papers:

242 nœuds, 125 773 liens

Visualisation de l'évolution de la communauté

Nœud u c'est un élève de la classe "4A"

v et w représentent autres élèves.



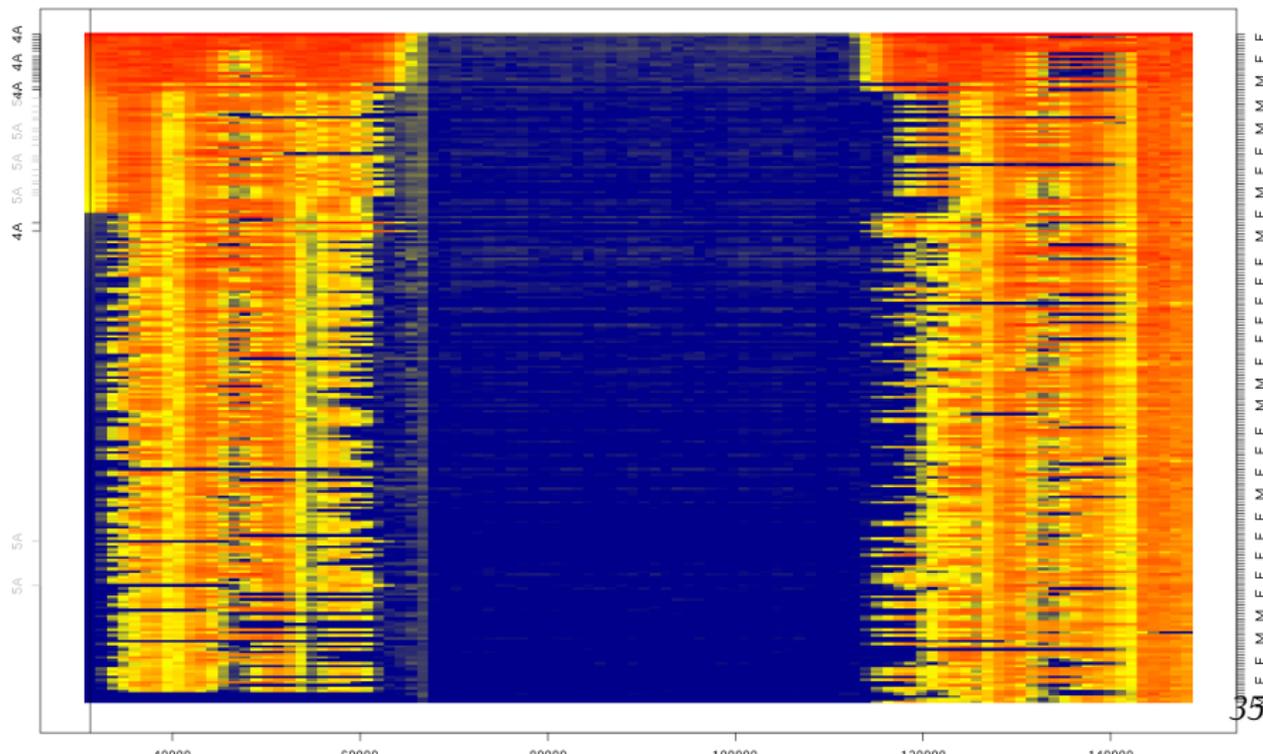
Il faut choisir l'instant t et faire le tri !

La visualisation dépend de ce tri.

Visualisation de l'évolution de la communauté

les lignes sont les élèves,

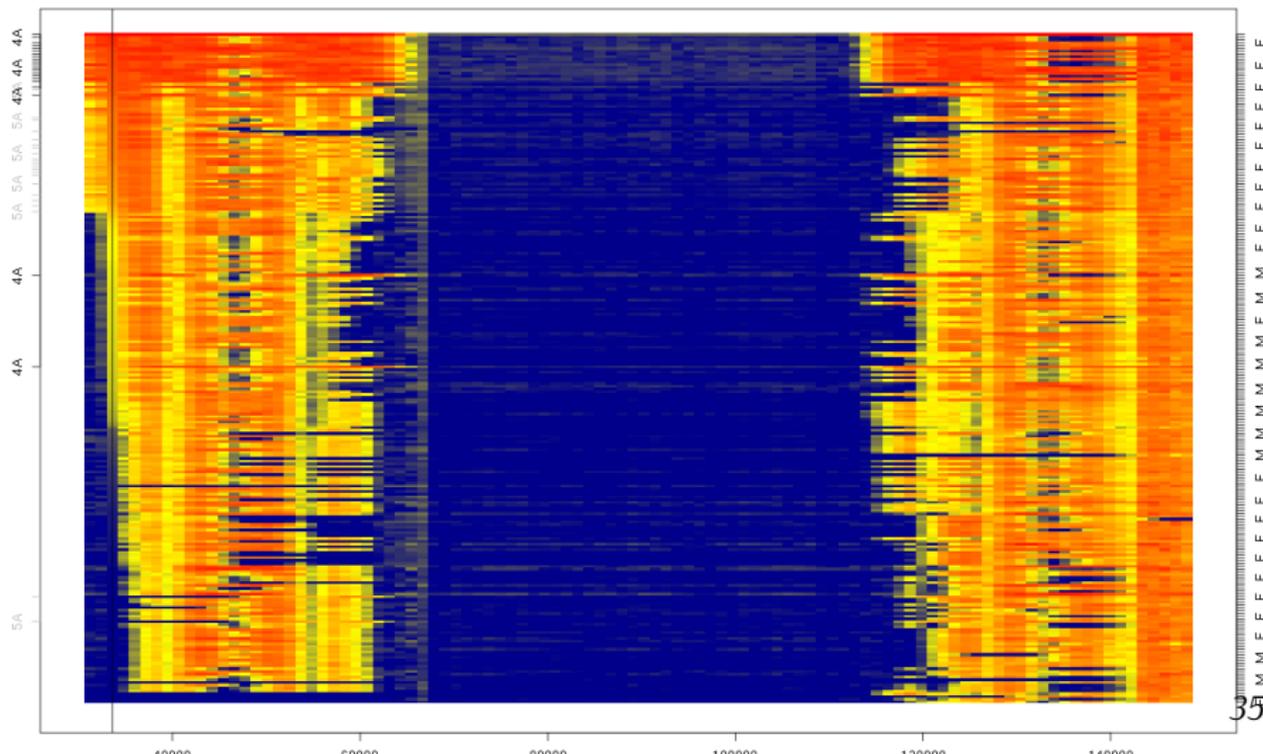
les colonnes sont les structures ego-communautaires $p_{u,t}(v)$



Visualisation de l'évolution de la communauté

les lignes sont les élèves,

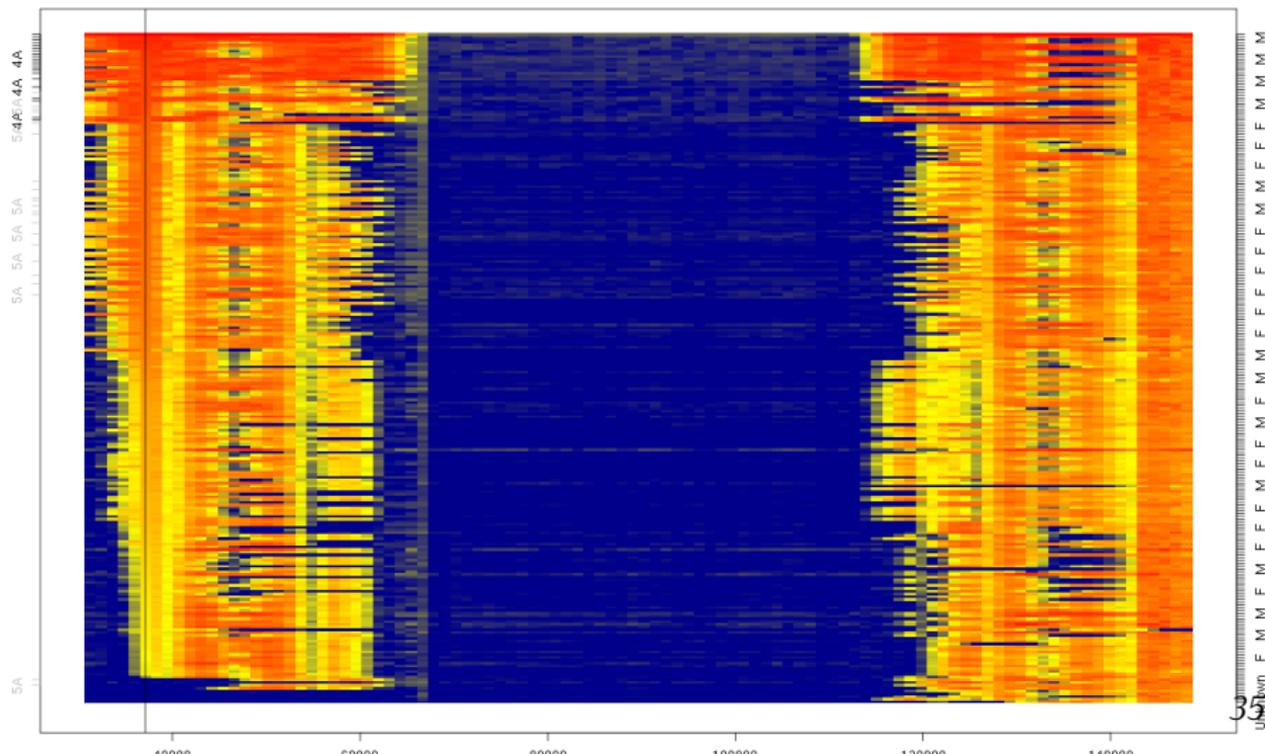
les colonnes sont les structures ego-communautaires $p_{u,t}(v)$



Visualisation de l'évolution de la communauté

les lignes sont les élèves,

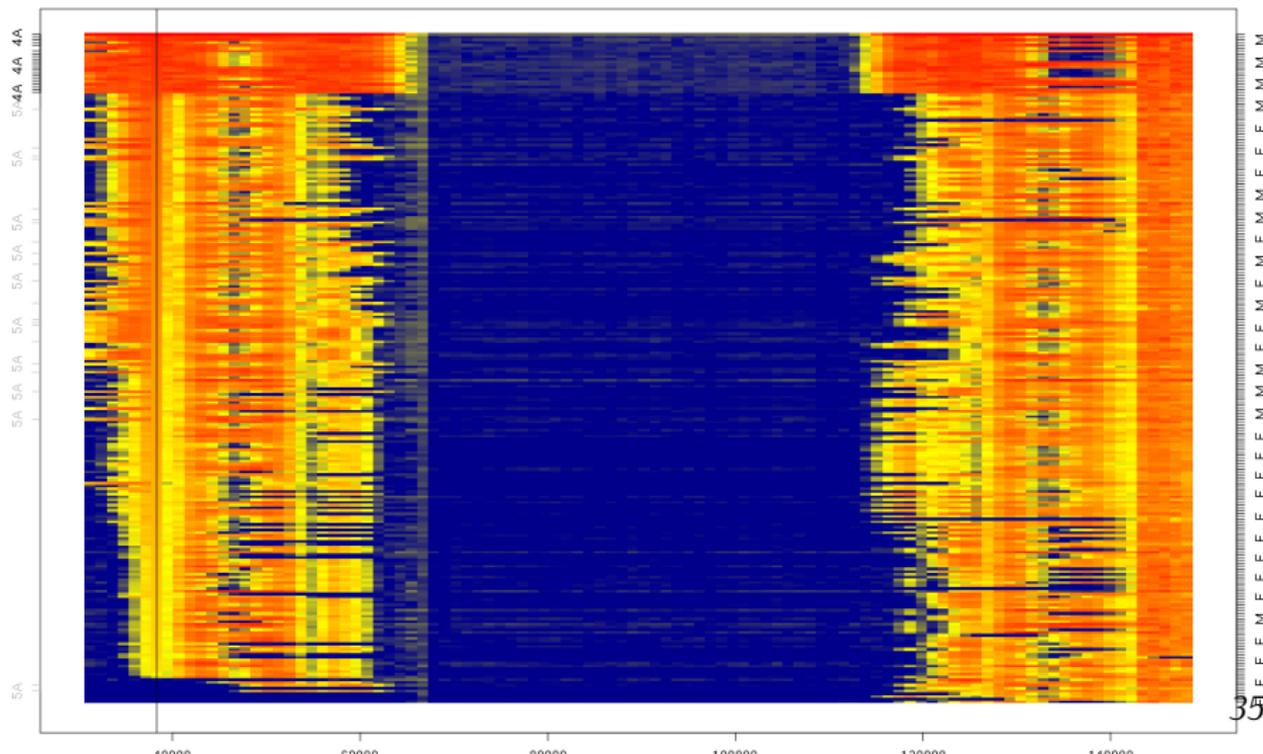
les colonnes sont les structures ego-communautaires $p_{u,t}(v)$



Visualisation de l'évolution de la communauté

les lignes sont les élèves,

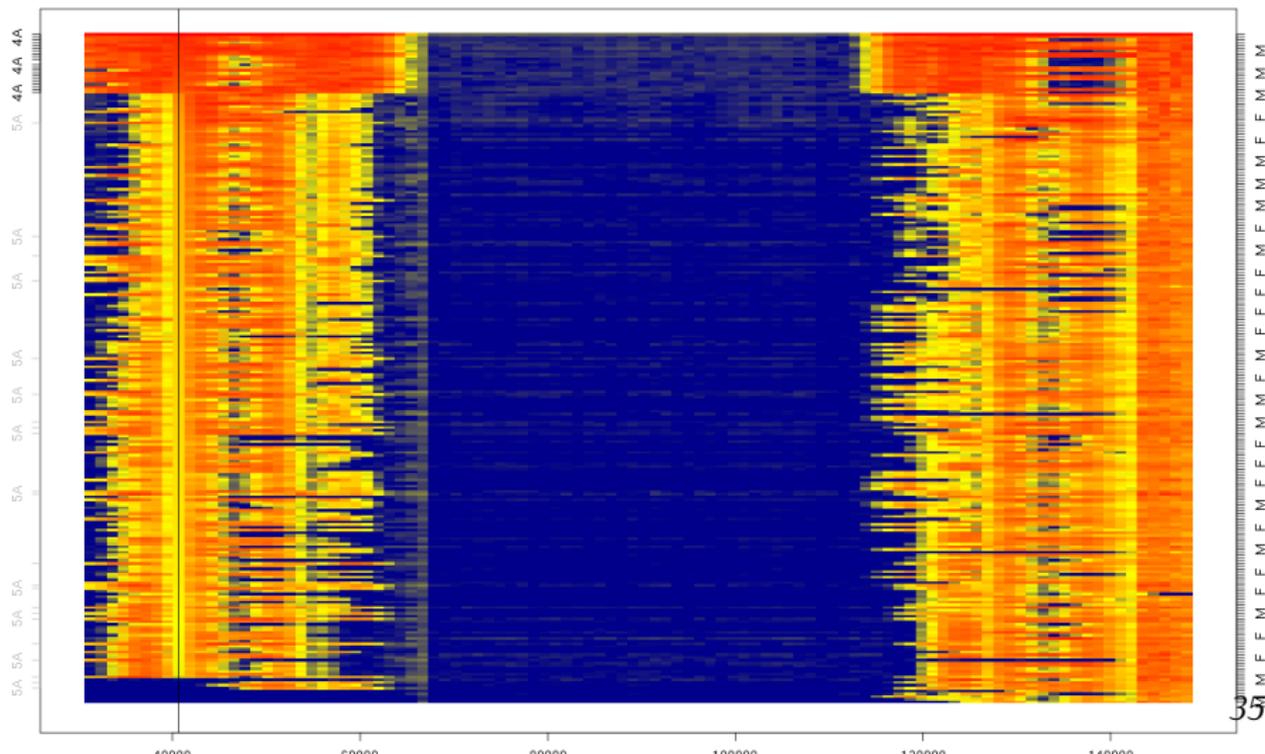
les colonnes sont les structures ego-communautaires $p_{u,t}(v)$



Visualisation de l'évolution de la communauté

les lignes sont les élèves,

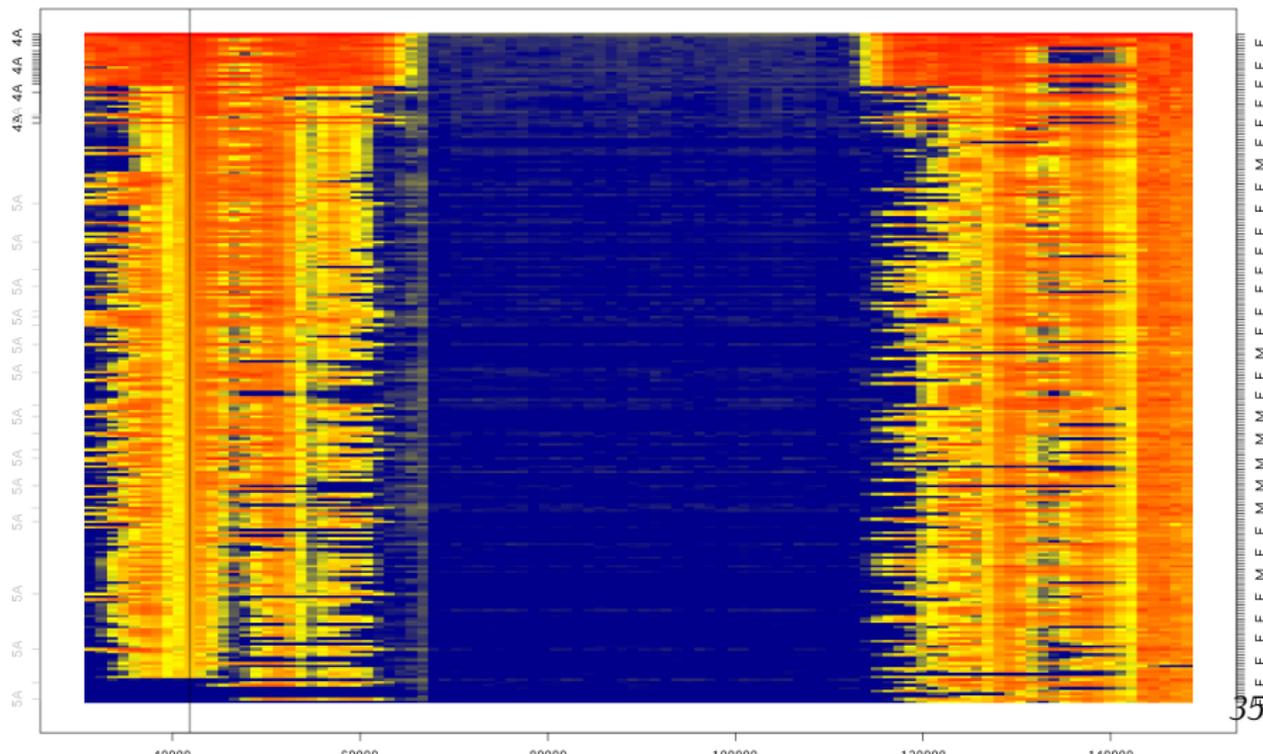
les colonnes sont les structures ego-communautaires $p_{u,t}(v)$



Visualisation de l'évolution de la communauté

les lignes sont les élèves,

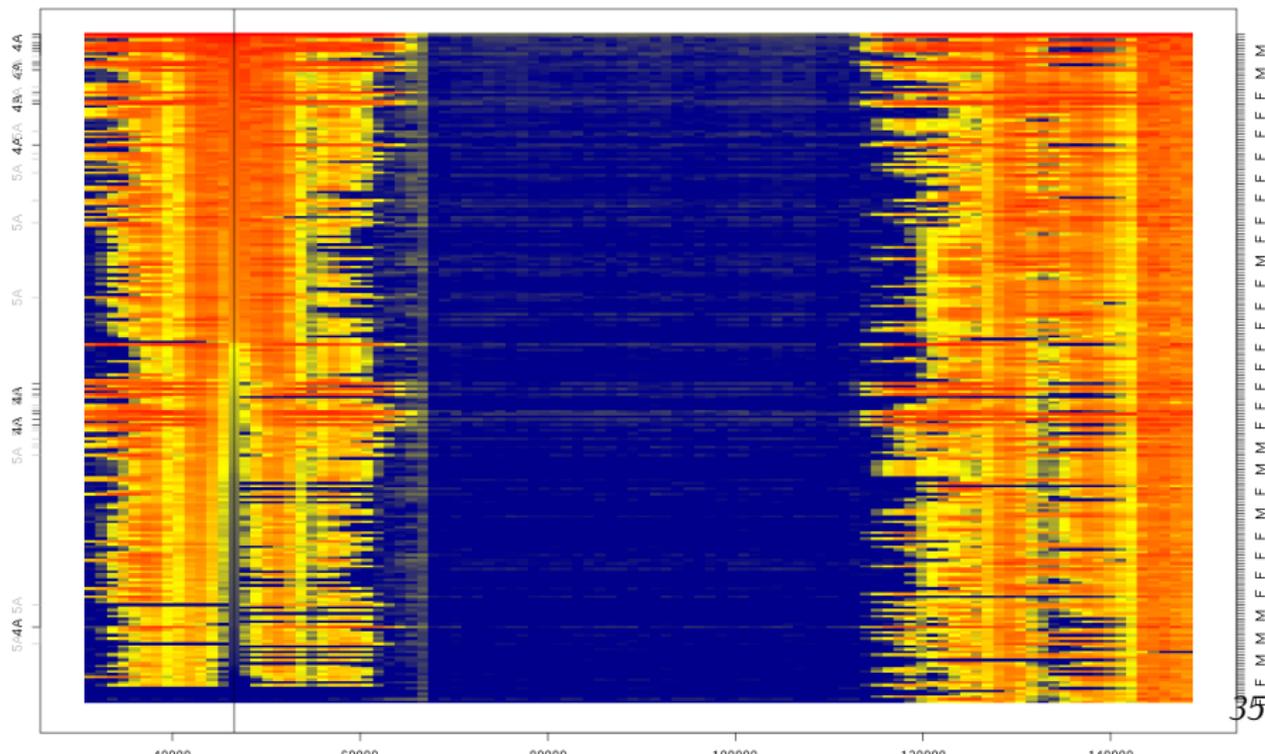
les colonnes sont les structures ego-communautaires $p_{u,t}(v)$



Visualisation de l'évolution de la communauté

les lignes sont les élèves,

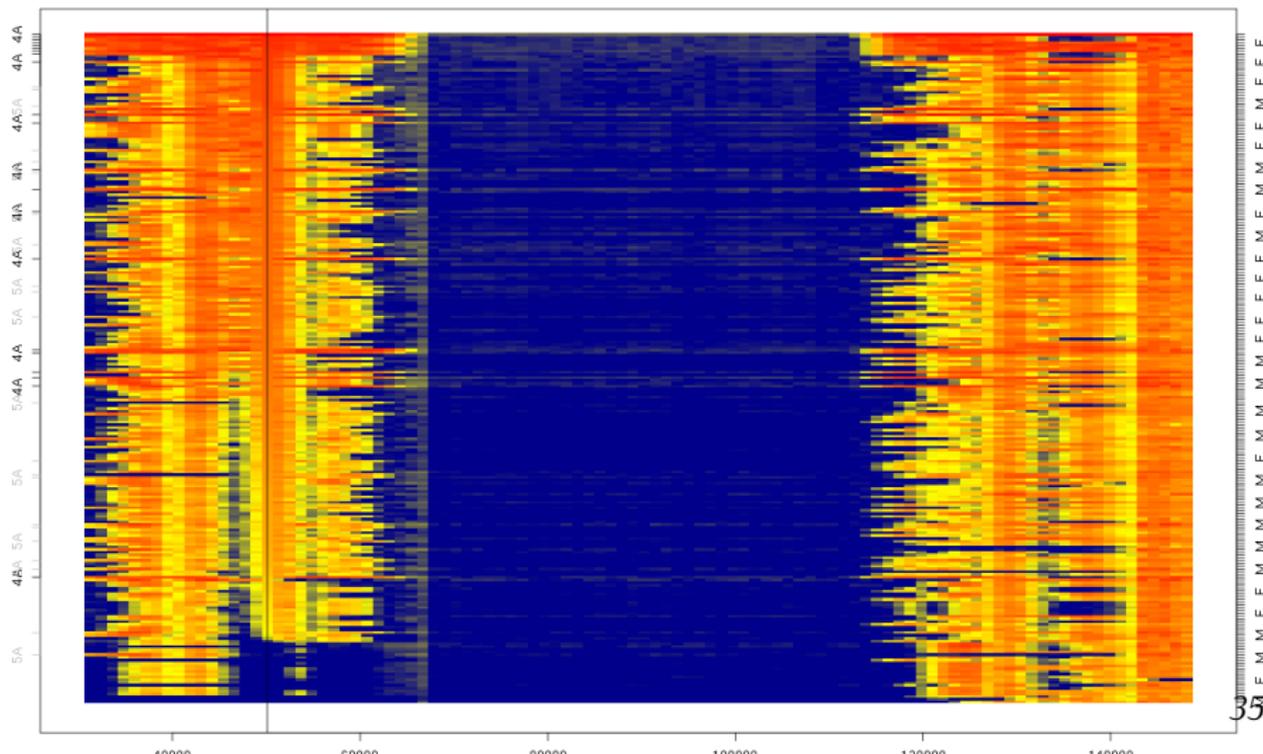
les colonnes sont les structures ego-communautaires $p_{u,t}(v)$



Visualisation de l'évolution de la communauté

les lignes sont les élèves,

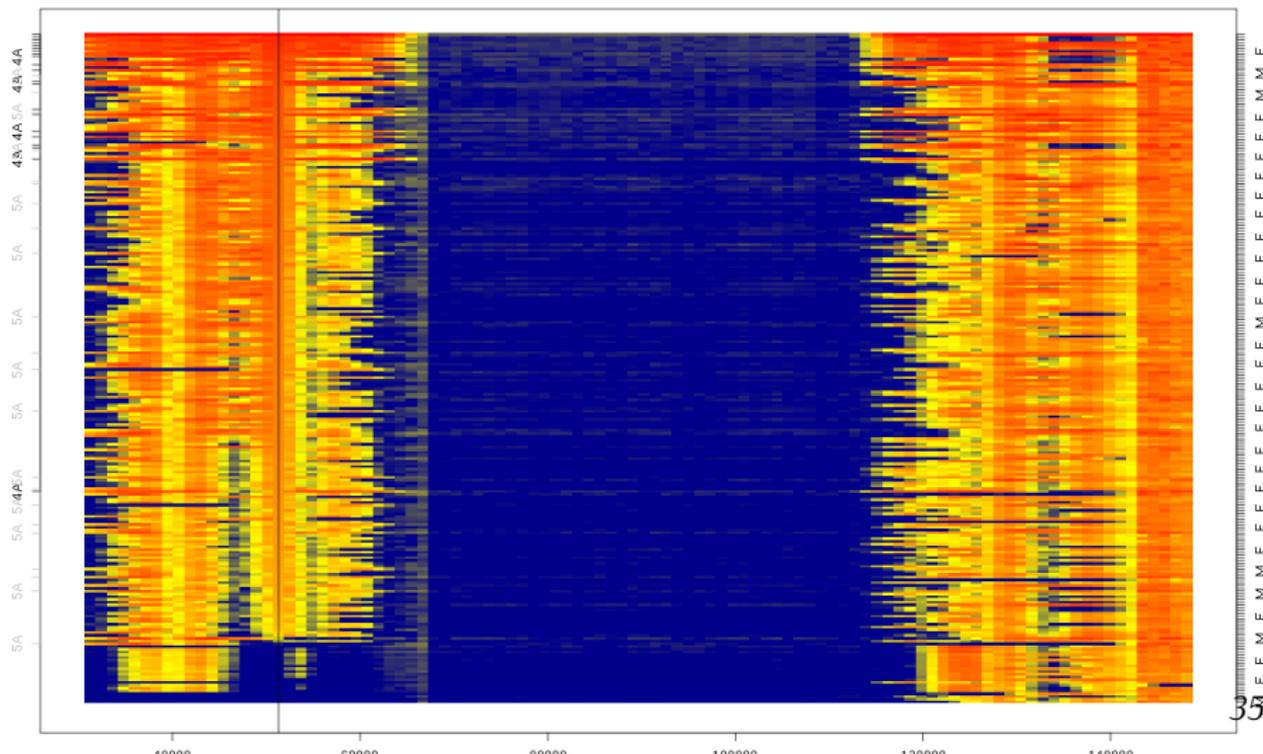
les colonnes sont les structures ego-communautaires $p_{u,t}(v)$



Visualisation de l'évolution de la communauté

les lignes sont les élèves,

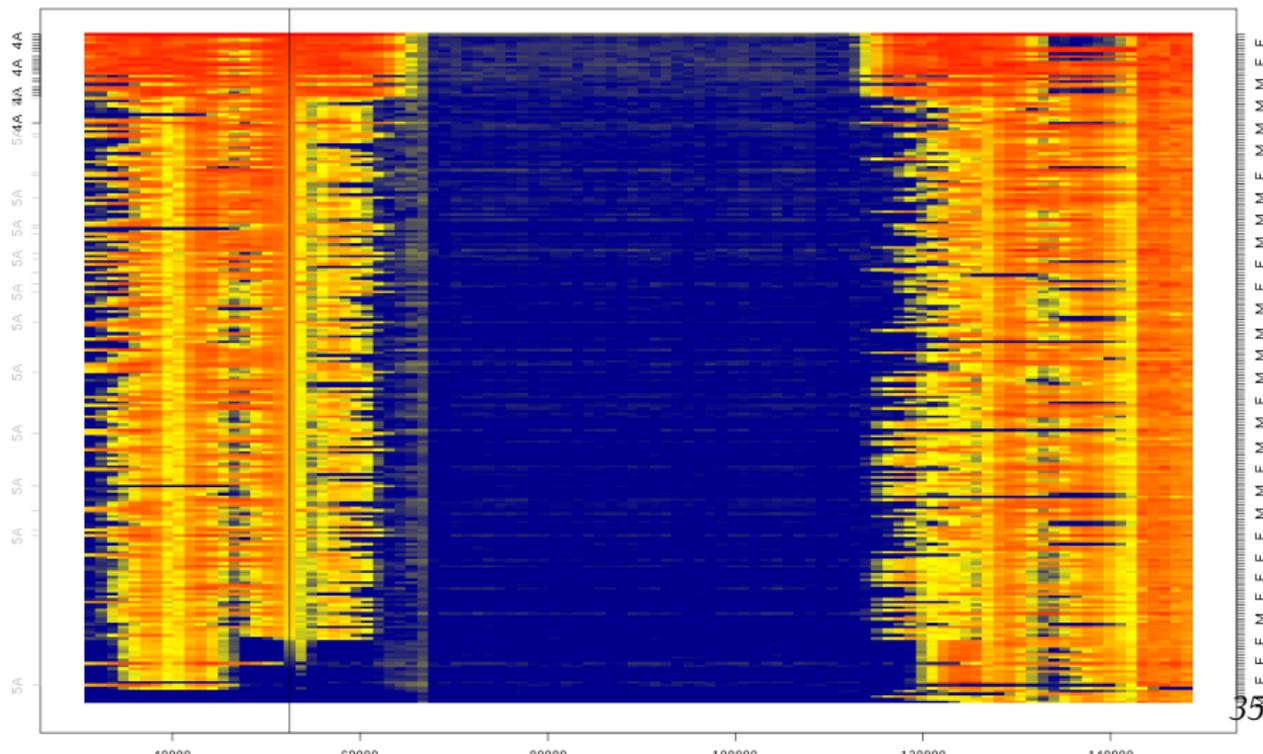
les colonnes sont les structures ego-communautaires $p_{u,t}(v)$



Visualisation de l'évolution de la communauté

les lignes sont les élèves,

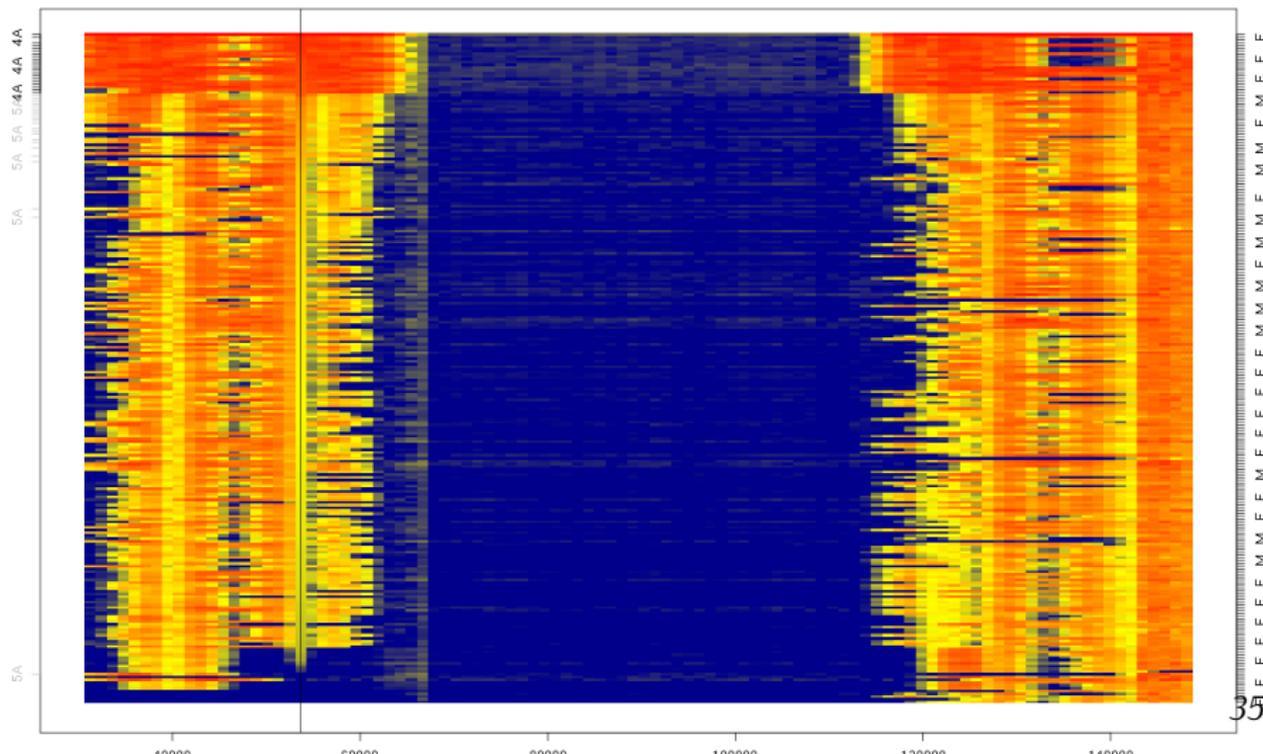
les colonnes sont les structures ego-communautaires $p_{u,t}(v)$



Visualisation de l'évolution de la communauté

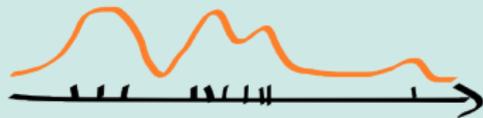
les lignes sont les élèves,

les colonnes sont les structures ego-communautaires $p_{u,t}(v)$

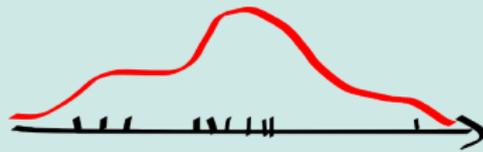


Paramètres

Lissage de la densité temporelle

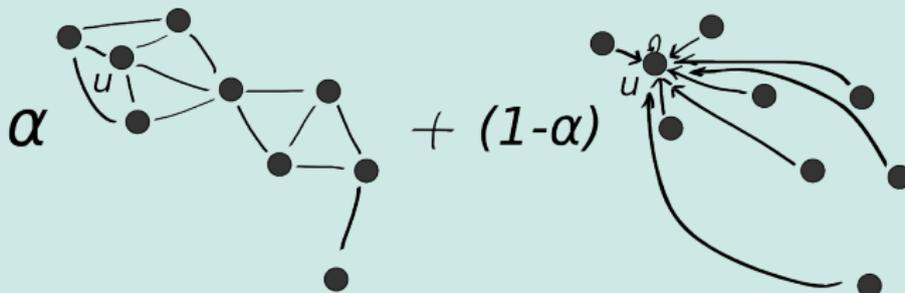


événements locaux



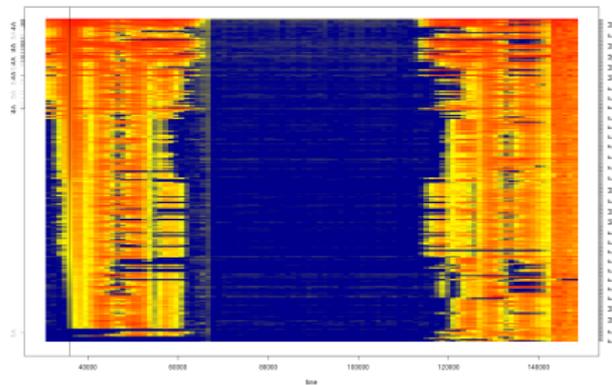
événements globaux

α du pagerank (paramètre de téléportation)

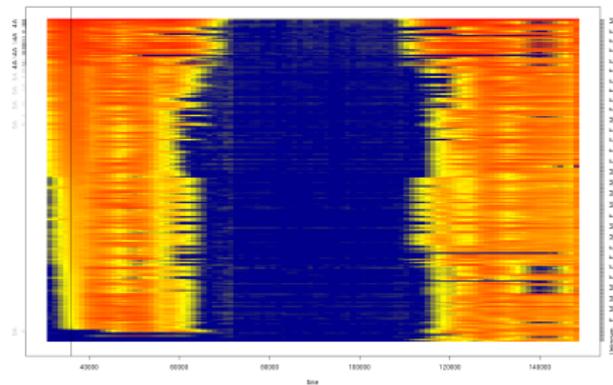


Plus α est petite, plus vite la marche aléatoire revient son origine

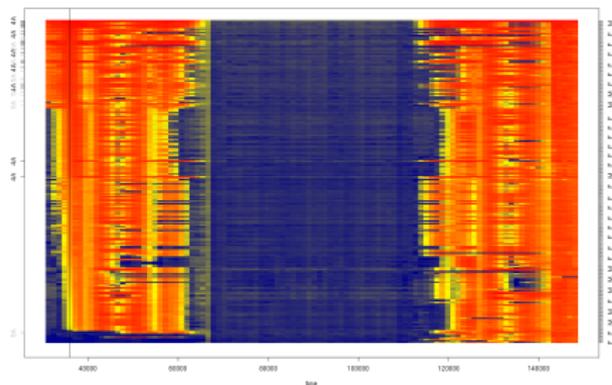
Visualisation



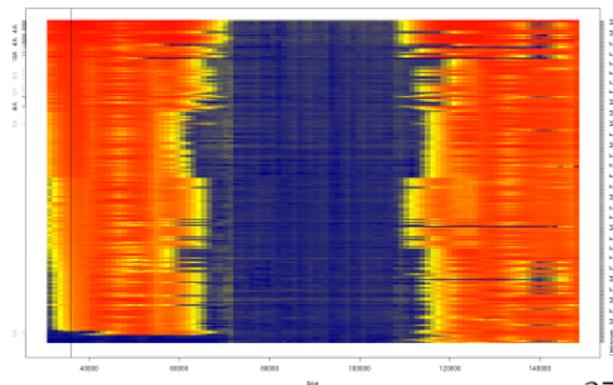
lissage 600, $\alpha = 0.2$



lissage 1200, $\alpha = 0.2$



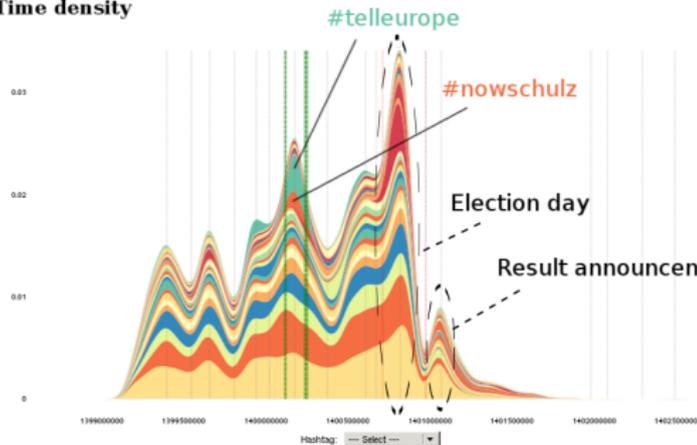
lissage 600, $\alpha = 0.8$



lissage 1200, $\alpha = 0.8$

European Elections 2014 in Twitter

Time density



Time window:

Left border:



Right border:



Top 10 hashtags

	count	hashtag
1	570	ep2014
2	443	telleurope
3	438	ee2014
4	256	notreeurope
5	237	europeennes2014
6	200	eelv
7	190	nowschulz
8	108	europe
9	91	tafta
10	85	ump

Top 10 users

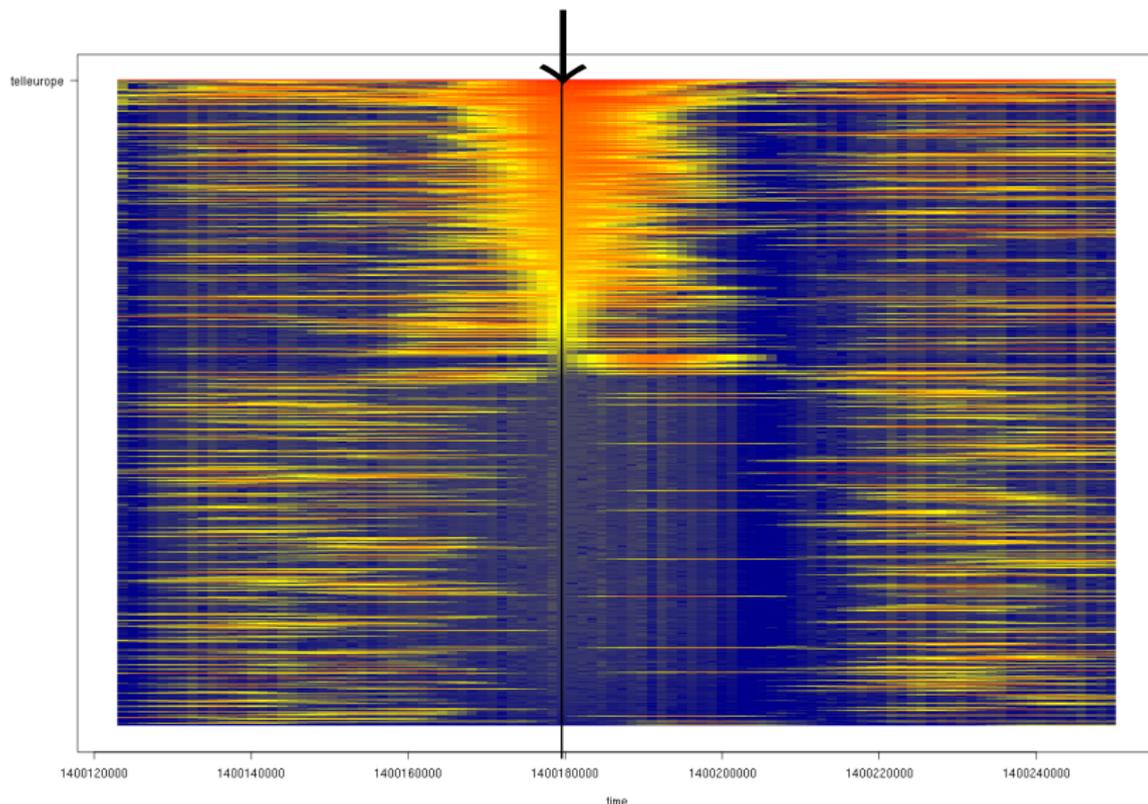
	count	from_user_name	from_user
1	121	JS Herpin	@jsherpin
2	108	Elodie Massé	@masselodie
3	81	Yannick Jadot	@yjadot
4	79	Béatrice DELGENDRE	@bdelgendre
5	74	Guillaume Balas	@BalasGuillaume
6	53	Raquel Garrido	@RaquelGarridoPG
7	47	Aleksander GLOGOWSKI	@Aleks_Paris
8	44	Sandrine Bélier	@sandrinebelier
9	39	Dolores BAUDELOT	@Ybdado
10	38	Franck Proust	@franckproust

Towards a Twitter Observatory : A multi-paradigm framework for collecting, storing and analysing tweets

Ian Basaille, Kirgizov, Éric Leclercq, Marinette Savonnet, and Nadine Cullot
RCIS 2016

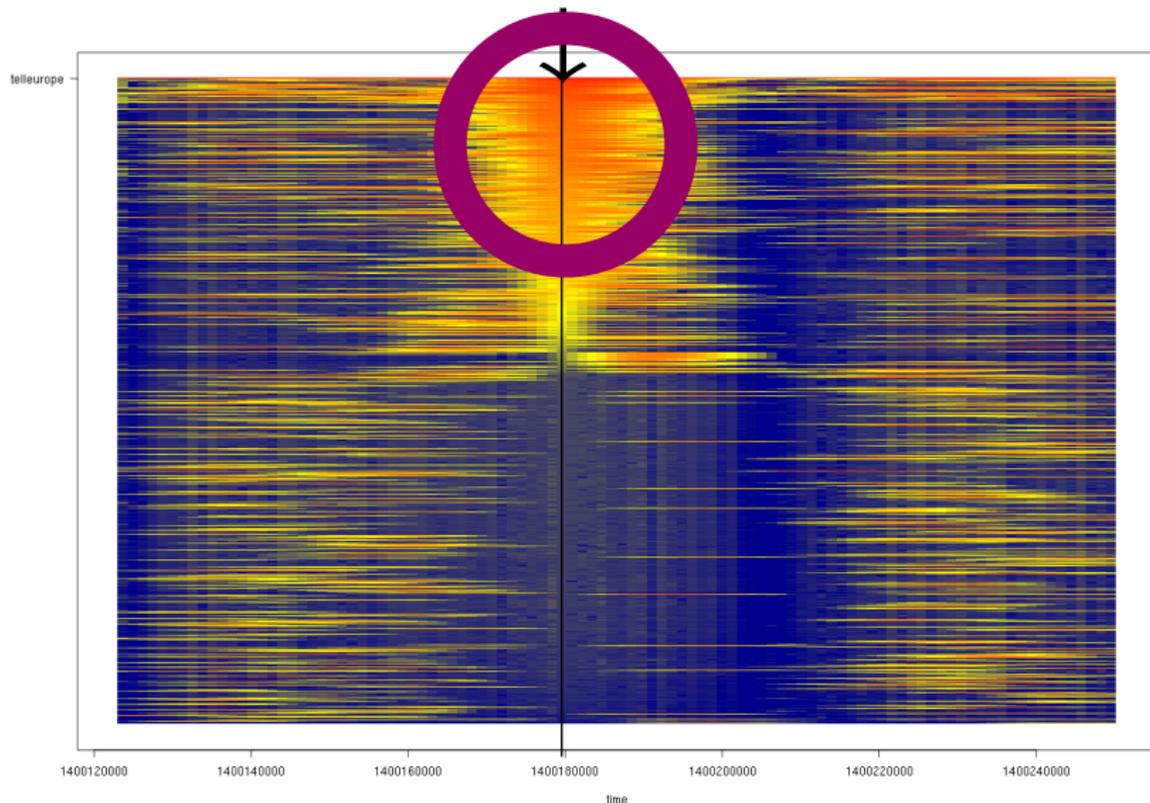
European Elections 2014 in Twitter

Dynamic community structure around #telleurope
Sort by values at this timestamp.



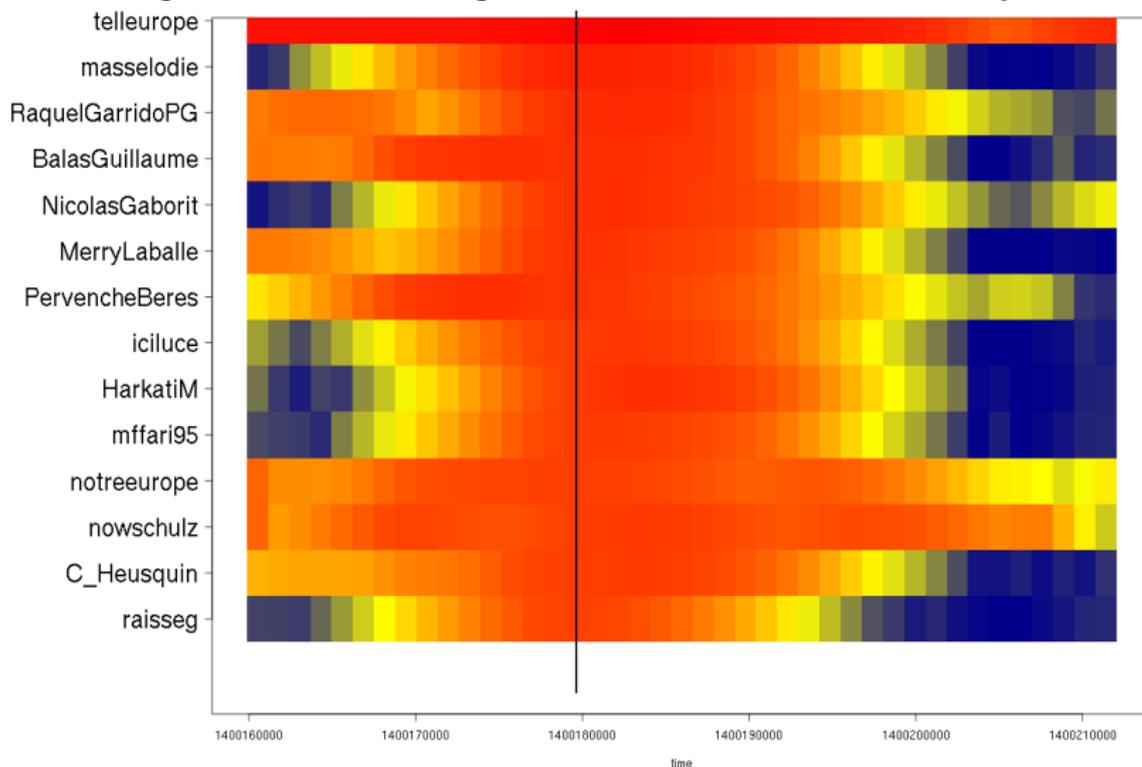
European Elections 2014 in Twitter

Dynamic community structure around #telleurope
Sort by values at this timestamp.



European Elections 2014 in Twitter

Dynamic community structure around #telleurope



1. Nos projets : TEE 2014 et PEPS 2015-2016
2. Structures communautaires
3. Densité temporelle des réseaux complexes & Évolution de la structure communautaire
4. Conclusion & Discussion

Densité temporelle pour

- ♣ visualiser un grand nombre de liens à la fois
- ♠ localiser et décrire les événements
- ★ étudier l'évolution de la structure (ego-)communautaire

Future

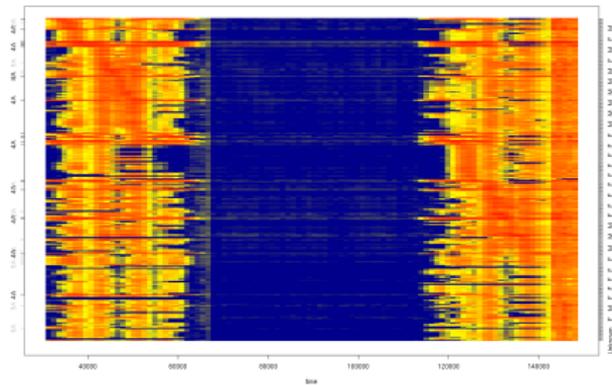
- ★ Utiliser d'autres algorithmes (HITS, par exemple)
- ♣ Traiter de données en ligne
- ♦ Proposer de meilleurs tris pour la visualisation
- ★ Modéliser la dynamique des structures communautaires en utilisant des approches de la combinatoire.
- ★ Appliquer aux autres réseaux

Merci de votre attention.

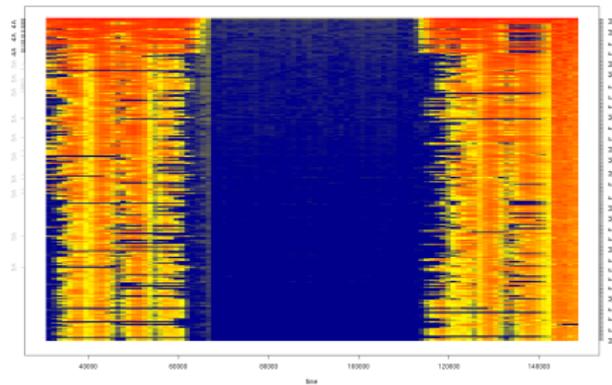
Questions ?

<http://kirgizov.link>

<https://github.com/kerzol/ego-evolution>



Tri par timestamp de max de
chaque ligne



Tri par somme de lignes

Algorithme et sa complexité

k : nombre des tranches du temps (j'ai choisi 100 pour la visualisation)

n : nombre de nœuds

m : nombre total de liens

μ : max nombre de liens entre deux nœuds

Algorithme

- 1 Lisser les fonctions de présence de chaque lien (Binned FFT), faire k tranches du temps — $O(m(\mu + k \log k))$
- 2 Pour chaque tranche faire du pagerank $\approx k \cdot O(m \log m)$ (ça dépend de la deuxième valeur propre)
- 3 Dénormalisation

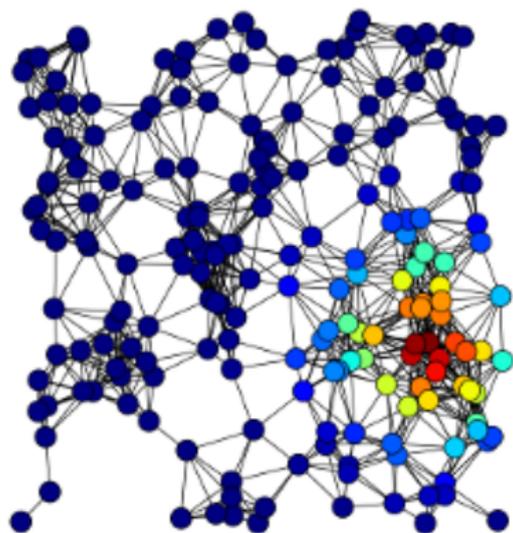
au totale... $\approx O(km \log(km) + m\mu)$

Fast computation of kernel estimators

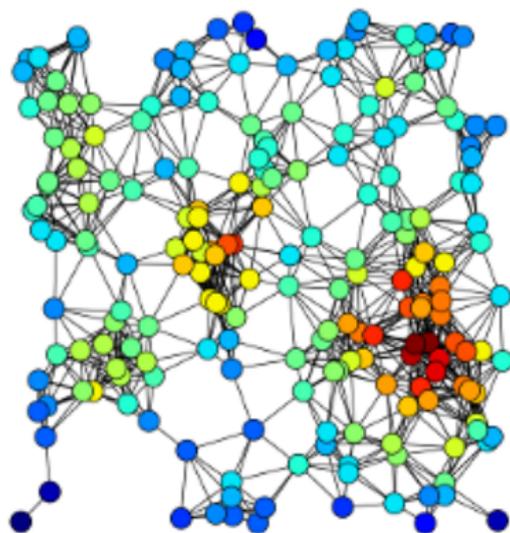
Raykar, Duraiswami, et Zhao, 2010

Using pagerank to locally partition a graph

Andersen, Chung, Lang, 2007⁴³



Centralité spectrale

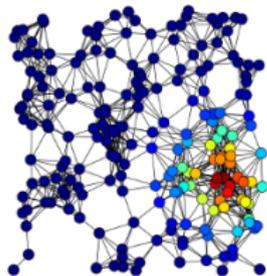


Degrés

Tapiocozzo@Wikipedia, CC BY-SA 4.0

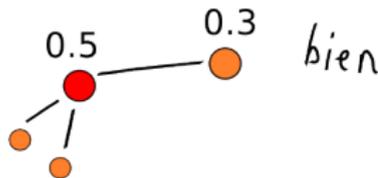
Centralité spectrale, deux interprétations

Soit A une matrice symétrique, par exemple une matrice d'adjacence du graphe non orienté.



Optimisation

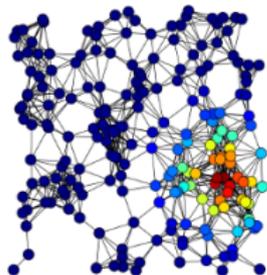
$$\max \sum A_{ij} x_i x_j$$
$$\|x\| = 1$$



On met plus de poids sur les nœuds adjacents !

Centralité spectrale, deux interprétations

Soit A une matrice symétrique, par exemple une matrice d'adjacence du graphe non orienté.



Optimisation

$$\max \sum A_{ij}x_i x_j$$
$$\|x\| = 1$$

Interprétation spectrale, théorème min-max de Courant-Fischer

$$\sum A_{ij}x_i x_j = xAx^T$$

$$\max_{\|x\|=1} x^T Ax = \lambda_{max}$$

$$\operatorname{argmax}_{\|x\|=1} xAx^T = v_{max}$$

v_{max} est le vecteur propre associé à $\lambda_{max}(A)$,
C'est mieux quand ils sont uniques.

Une version de mincut, deux interprétations

Soit A une matrice symétrique, par exemple une matrice d'adjacence du graphe non orienté.

Une version de mincut

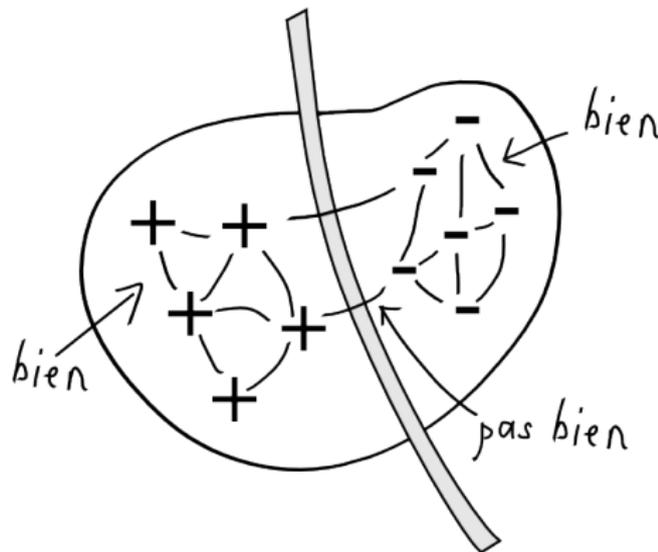
$$\max \sum A_{ij} x_i x_j$$

$$\|x\| = 1, \sum x_i = 0$$

$$S = \{i \in V : x_i \leq 0\}$$

$$\bar{S} = \{i \in V : x_i > 0\}$$

$$G = S + \bar{S} + \text{cut}(S, \bar{S})$$



Une version de mincut, deux interprétations

Soit A une matrice symétrique, par exemple une matrice d'adjacence du graphe non orienté.

Une version de mincut

$$\max \sum A_{ij} x_i x_j$$

$$\|x\| = 1, \sum x_i = 0$$

$$S = \{i \in V : x_i \leq 0\}$$

$$\bar{S} = \{i \in V : x_i > 0\}$$

$$G = S + \bar{S} + \text{cut}(S, \bar{S})$$

$$\sum x_i = 0 \iff x \perp \mathbf{1}$$

Une version de mincut, deux interprétations

Soit A une matrice symétrique, par exemple une matrice d'adjacence du graphe non orienté.

Une version de mincut

$$\max \sum A_{ij} x_i x_j$$

$$\|x\| = 1, \sum x_i = 0$$

$$S = \{i \in V : x_i \leq 0\}$$

$$\bar{S} = \{i \in V : x_i > 0\}$$

$$G = S + \bar{S} + \text{cut}(S, \bar{S})$$

$$\sum x_i = 0 \iff x \perp \mathbf{1}$$

On transforme un peu la matrice de telle sorte que $v_{\max} = \mathbf{1}$

Une version de mincut, deux interprétations

Soit A une matrice symétrique, par exemple une matrice d'adjacence du graphe non orienté.

Une version de mincut

$$\max \sum A_{ij} x_i x_j$$

$$\|x\| = 1, \sum x_i = 0$$

$$S = \{i \in V : x_i \leq 0\}$$

$$\bar{S} = \{i \in V : x_i > 0\}$$

$$G = S + \bar{S} + \text{cut}(S, \bar{S})$$

$$\sum A_{ij} x_i x_j = xAx^T$$

Courant-Fischer

$$\max_{\substack{\|x\|=1 \\ x \perp \mathbf{1}}} x^T Ax = \lambda_2$$

$$\operatorname{argmax}_{\substack{\|x\|=1 \\ x \perp \mathbf{1}}} xAx^T = v_2$$

v_2 est le vecteur propre associé à la deuxième plus grande valeur propre de A .

Inégalité(s) de Cheeger

Définitions

δ_v : degré du nœud v

D : matrice diagonale des degrés

$M = D^{-1}A$: matrice des transitions

$$G = S + \bar{S} + \text{cut}(S, \bar{S}), \quad \text{Vol}(S) = \sum_{v \in S} \delta_v$$

Conductance

$$\Phi(G) = \min_{\emptyset \subset S \subset V} \frac{|\text{cut}(S, \bar{S})|}{\text{Vol}(S)\text{Vol}(\bar{S})} \cdot 2|E|$$

Inégalité(s) de Cheeger

$$\frac{\Phi^2}{16} \leq 1 - \lambda_2 \leq \Phi$$

λ_2 deuxième plus grande valeur propre de M

Articles, livres

- Four Cheeger-type Inequalities for Graph Partitioning Algorithms**
Fan Chung, 2007
- Normalized cuts and image segmentation**
Jianbo Shi and Jitendra Malik, 2000
- Eigenvalues of graphs**
László Lovász, 2007
- Spectres de graphes**
Yves Colin de Verdière, 1998

Algorithme HITS (Hyperlink-Induced Topic Search)

A une matrice d'adjacence du graphe orienté.
Elle n'est pas toujours symétrique.

Mais AA^T et $A^T A$ sont toujours symétriques !

HITS de Kleinberg

$v_{max}(AA^T)$: scores de hub (hub \approx degré sortant élevé)

$v_{max}(A^T A)$: scores de autorité (autorité \approx degré entrant élevé)

Authoritative Sources in a Hyperlinked Environment

1999 Kleinberg

Je remercie les personnes suivantes pour les discussions et l'inspiration :

Eric Leclercq, Maximilien Danisch, Benjamin Gras,
Armen Petrossian, Nicolas Gastineau, Jean-luc Baril,
Tiphaine Viard, Clémence Magnien, Jean-Loup
Guillaume et Emmanuel Orsini