



Analyse lexicométrique des tweets sur le #mariagepourtous

Pascal Marchand (pascal.marchand@iut-tlse3.fr)

Pierre Ratinaud (ratinaud@univ-tlse2.fr)



Pourquoi ?

Intérêts théoriques

- dynamique des systèmes de représentations sociales
- étude des "principes générateurs de prises de position" (Doise, 1986) à partir de groupes définis par leurs interactions réelles plutôt que définis a priori par des variables socio-catégorielles
- stabilité des mondes lexicaux (Reinert, 2008)

Et ensuite...

- projection chronologique des liens classes / communautés
- injection des retweets dans l'analyse
- étude des hastags mobilisés
- comparaison avec la presse
- étude des arts
- comparaison avec les débats à l'Assemblée

Pourquoi ?

Intérêts méthodologiques

- pratique pour la mise au point d'outils de visualisation chronologique (twitter est un flux)
- articulation analyse de graphe / lexicométrie

Limites

- Population : jeunes ++, diplômés ++ (Semiocast, 2012 ; Vedel, 2012)
- API (indexation a priori et pas a posteriori + limitations + changements imprévisibles)
- nature des échanges (liens)
- "pauvreté lexicale"

Comment ?

On indexe le flux de twitter à partir de mots clés ou de hashtag
=> **limitation de la thématique**

Outils utilisés :

- yourTwapperKeeper (avant) <https://github.com/540co/yourTwapperKeeper>
- dmi-tcat (maintenant)

Le contexte

Le "mariage pour tous" est une promesse de campagne de François Hollande.
Il s'agit de donner le droit au mariage et à l'adoption aux couples de même sexe.

Quelques dates :

- 07/11/2012 : présentation du texte en conseil des ministres
- 29/01/2013 : début des débats à l'assemblée (110 heures de débats en 10 jours, 5000 amendements)
- 12/02/2013 : vote du texte à l'assemblée (329 pour, 229 contre)
- du 04 au 12/04/2013 : débats au sénat (vote le 12)

Le contexte

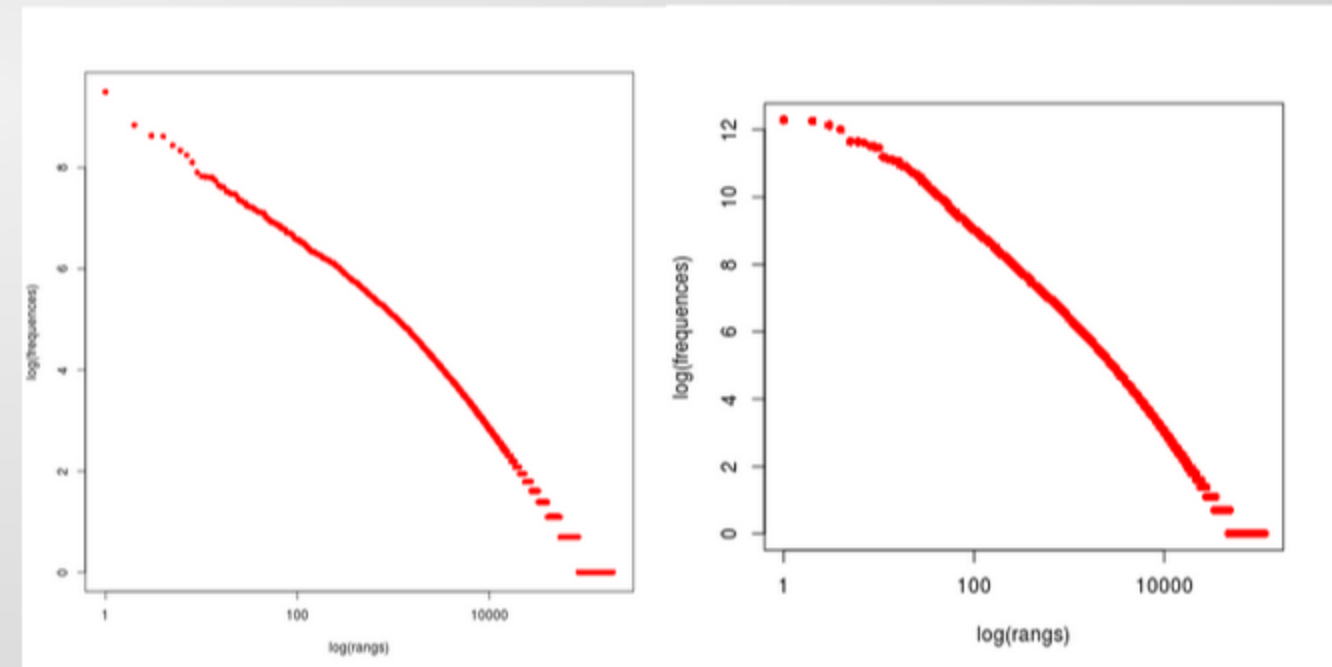
- 15/04/2013 : retour du texte à l'assemblée
- 23/04/2013 : vote du texte à l'assemblée
- 17/05/2013 : validation par le conseil constitutionnel
- fin mai : premier mariage à Montpellier

Entre temps : de nombreuses manifestations contre ("manifpourtous") et pour la loi.

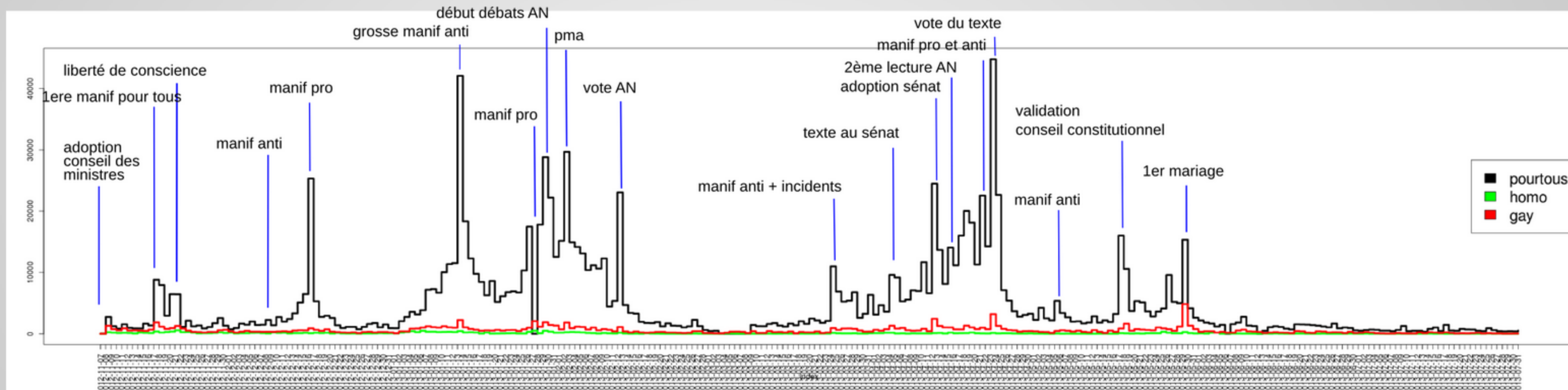
Le corpus

Indexation du mot clé mariagepour tous avec twapperkeeper du 07/11/2012 au 31/07/2013 (266 jours - environ 9 mois)

1 260 092 tweets (dont 61,58% de retweets)
195441 utilisateurs (dont 110532 "hapax" - 56,55%
des utilisateurs - 8,7% des messages)

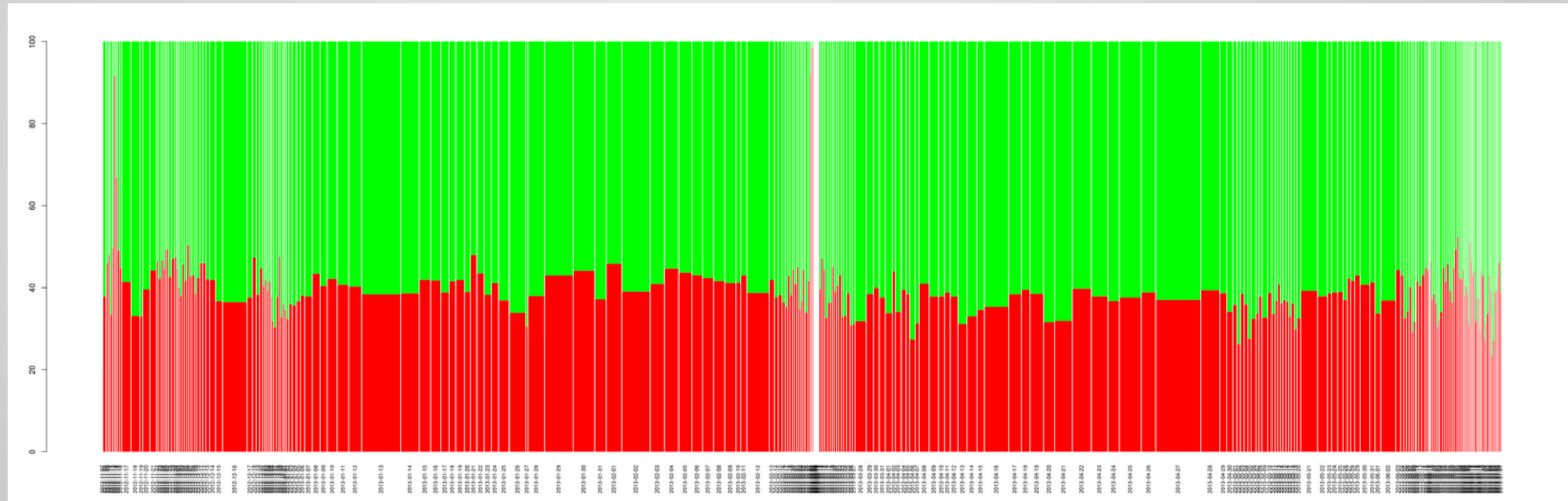


Chronologie et fréquence des tweets



- On conserve uniquement les tweets originaux (= élimination des retweets)
- Nettoyage :
- Elimination des urls
- Elimination du hashtag #mariagepourtous

tweets (en rouge) et retweets (en vert)



Le corpus

- On conserve uniquement les tweets originaux (= élimination des retweets)
- Nettoyage :
- Elimination des urls
- Elimination du hastag #mariagepourtous

Méthode Reinert avec IRaMuTeQ :

70 classes demandées - 8000 formes pleines

On conserve les classes de plus de 10500 tweets

10 classes terminales

88,7% des tweets sont conservés



Le corpus

Le corpus se compose alors de 453 107 tweets.

6 474 386 occurrences (73 380 hapax - 53,6% des formes - 1,13% des occurrences)

Analyse

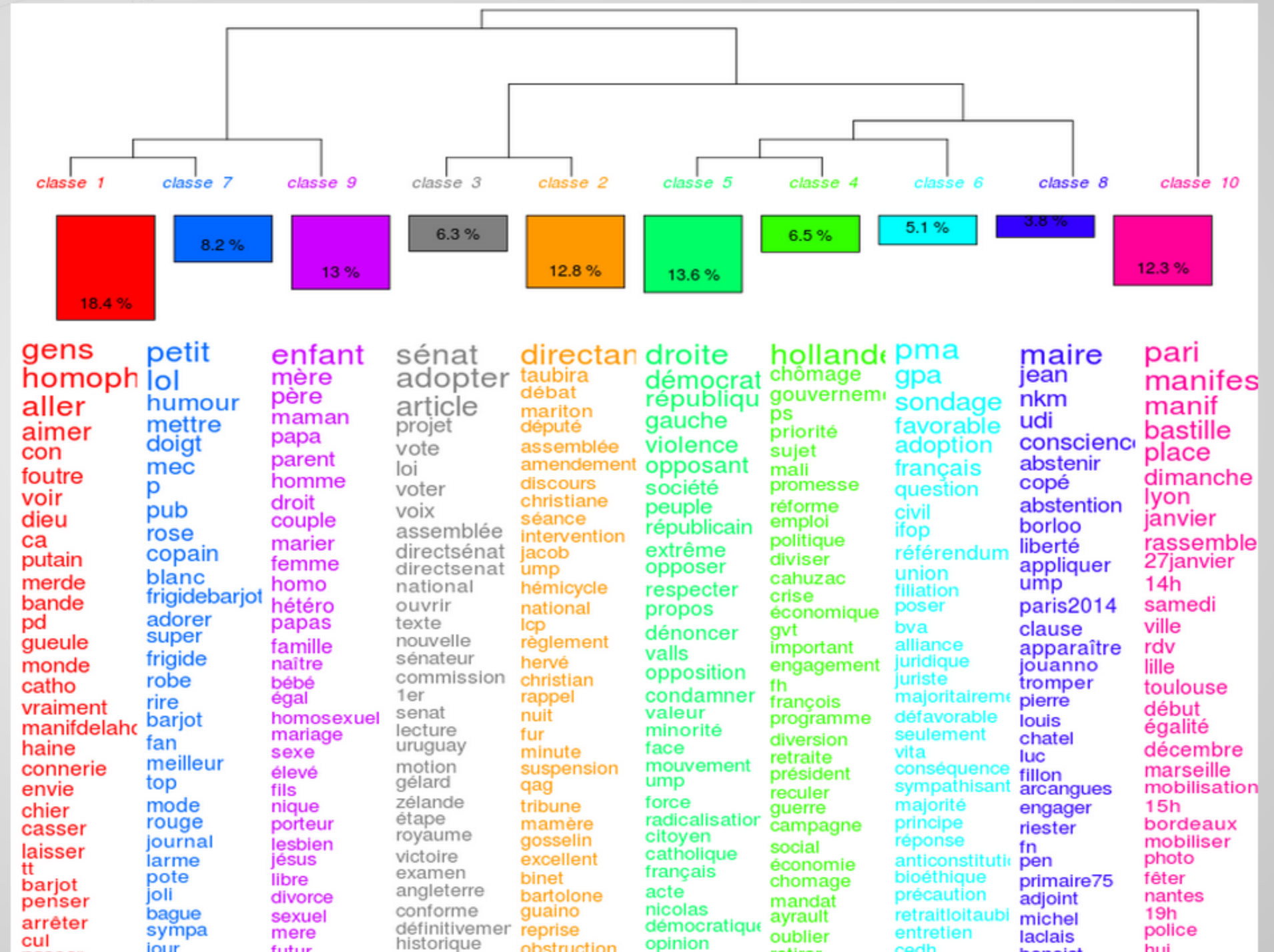
Méthode Reinert avec IRaMuTeQ :

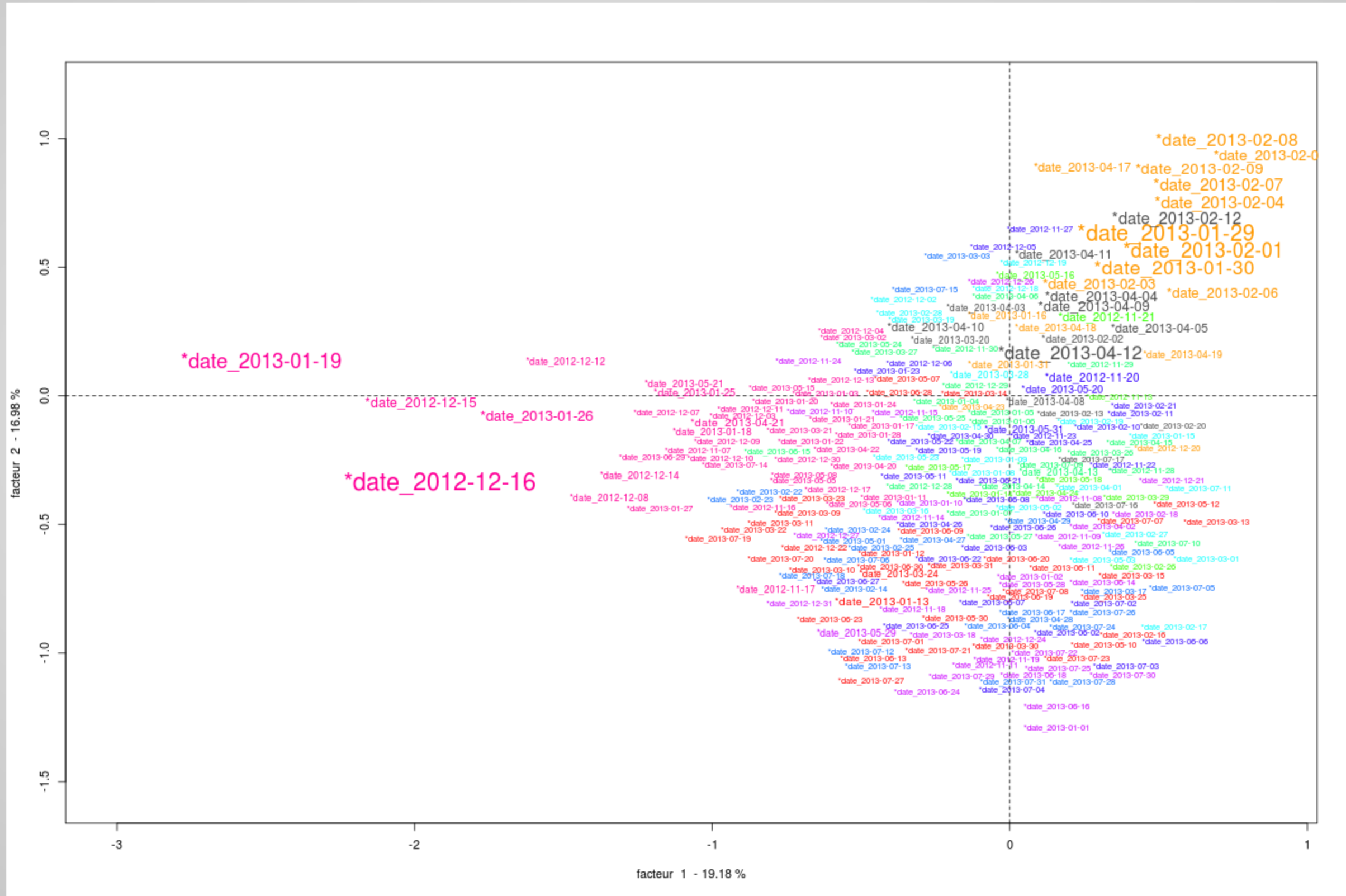
70 classes demandées - 8000 formes pleines

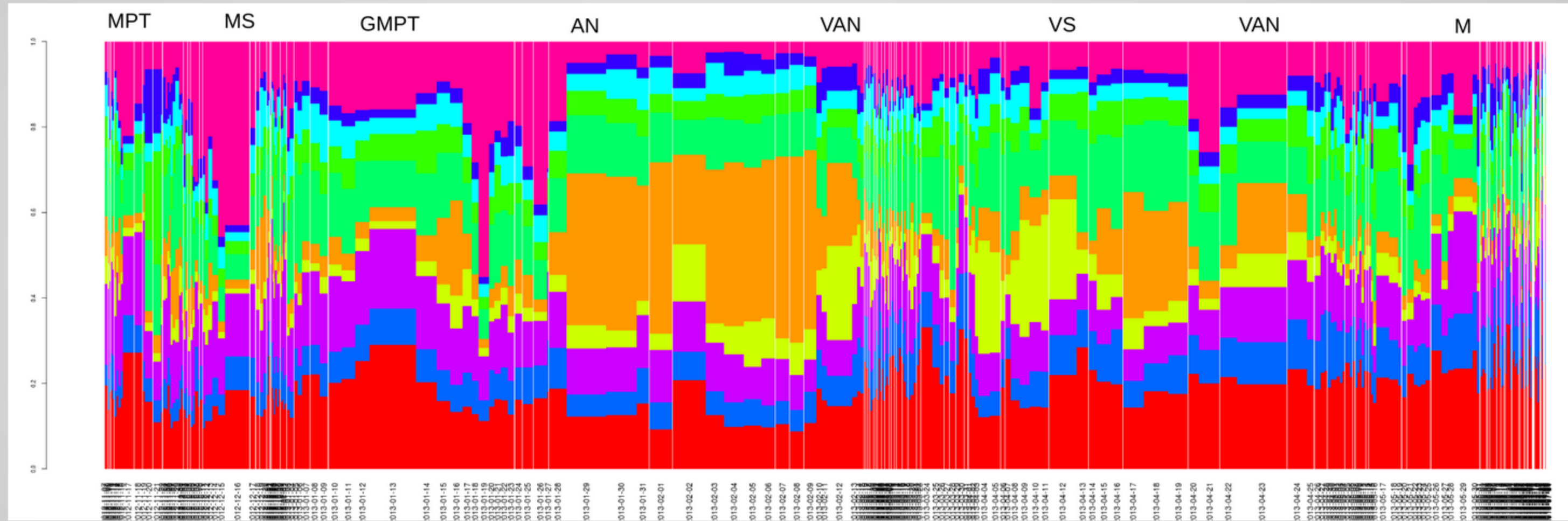
On conserve les classes de plus de 10500 tweets

10 classes terminales

88,7% des tweets sont conservés









Détermination des communautés

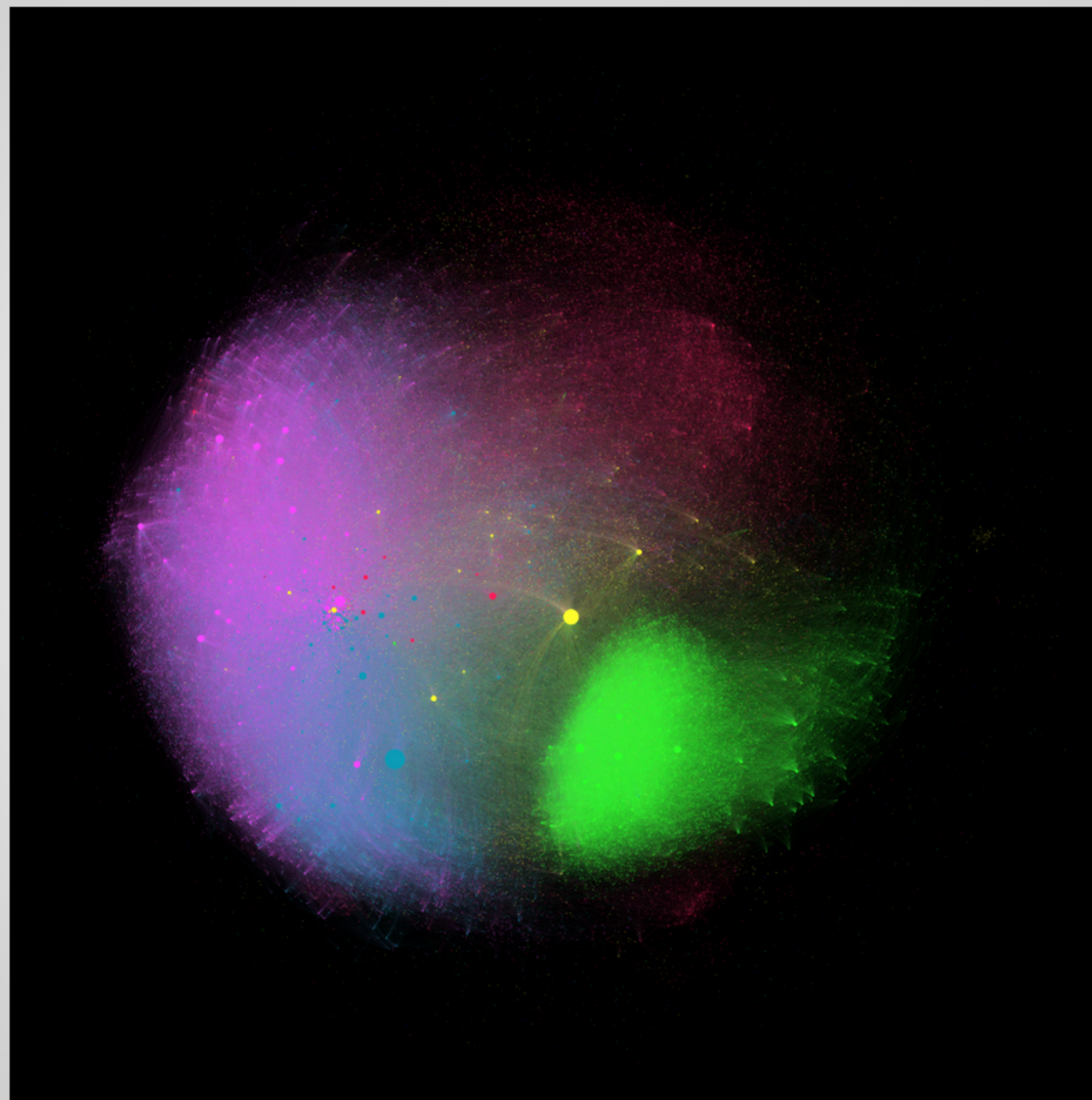
Graphe orienté des retweets des utilisateurs ayant au moins 4 tweets.

Semble efficace pour repérer des communautés d'opinions (Rieder & Smyrnaio, 2012; Smyrnaio & Ratinaud, 2014)

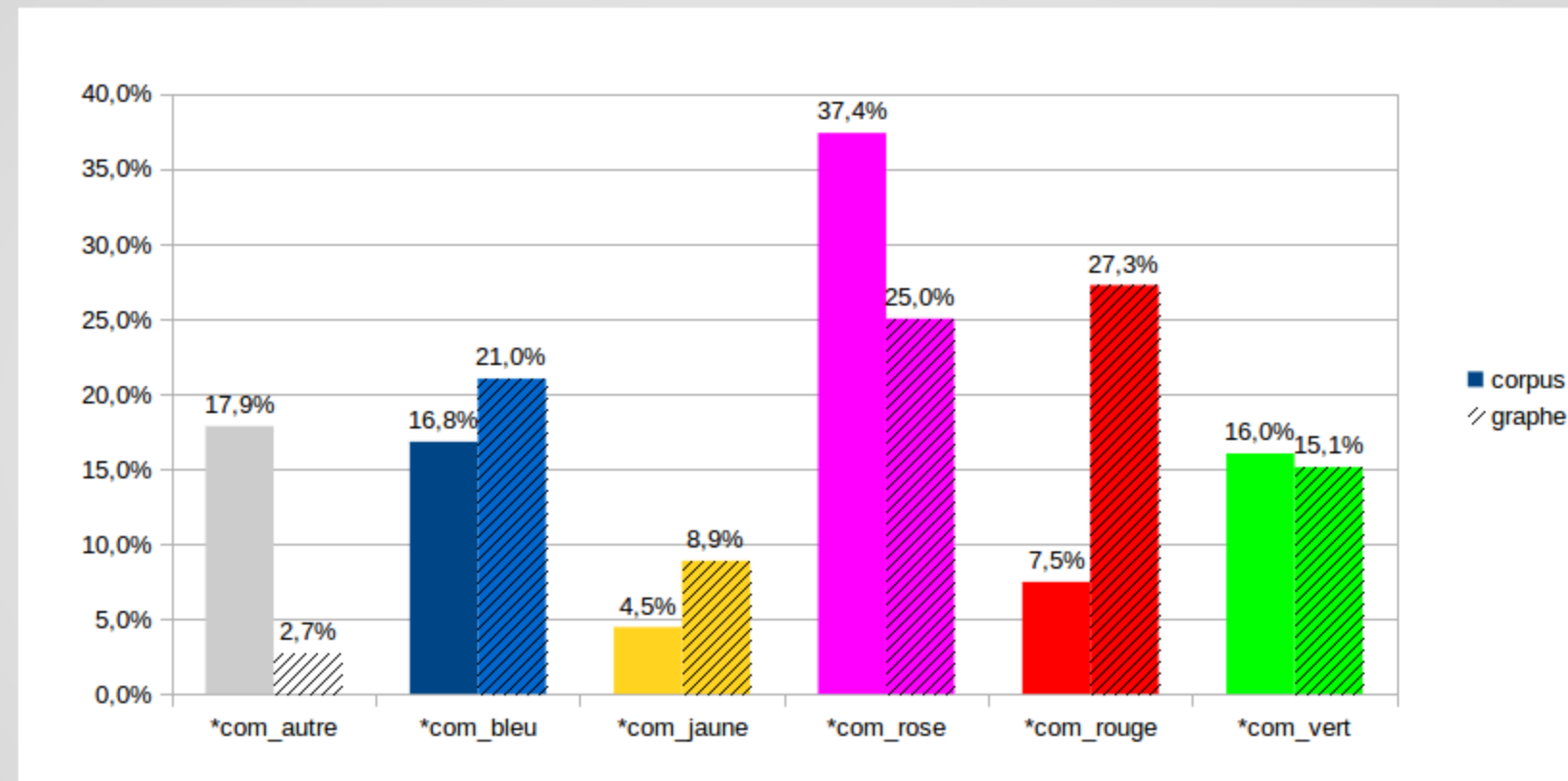
58 587 utilisateurs - 418 844 liens

On applique un algorithme de détection de communautés (Blondel & al, 2008)

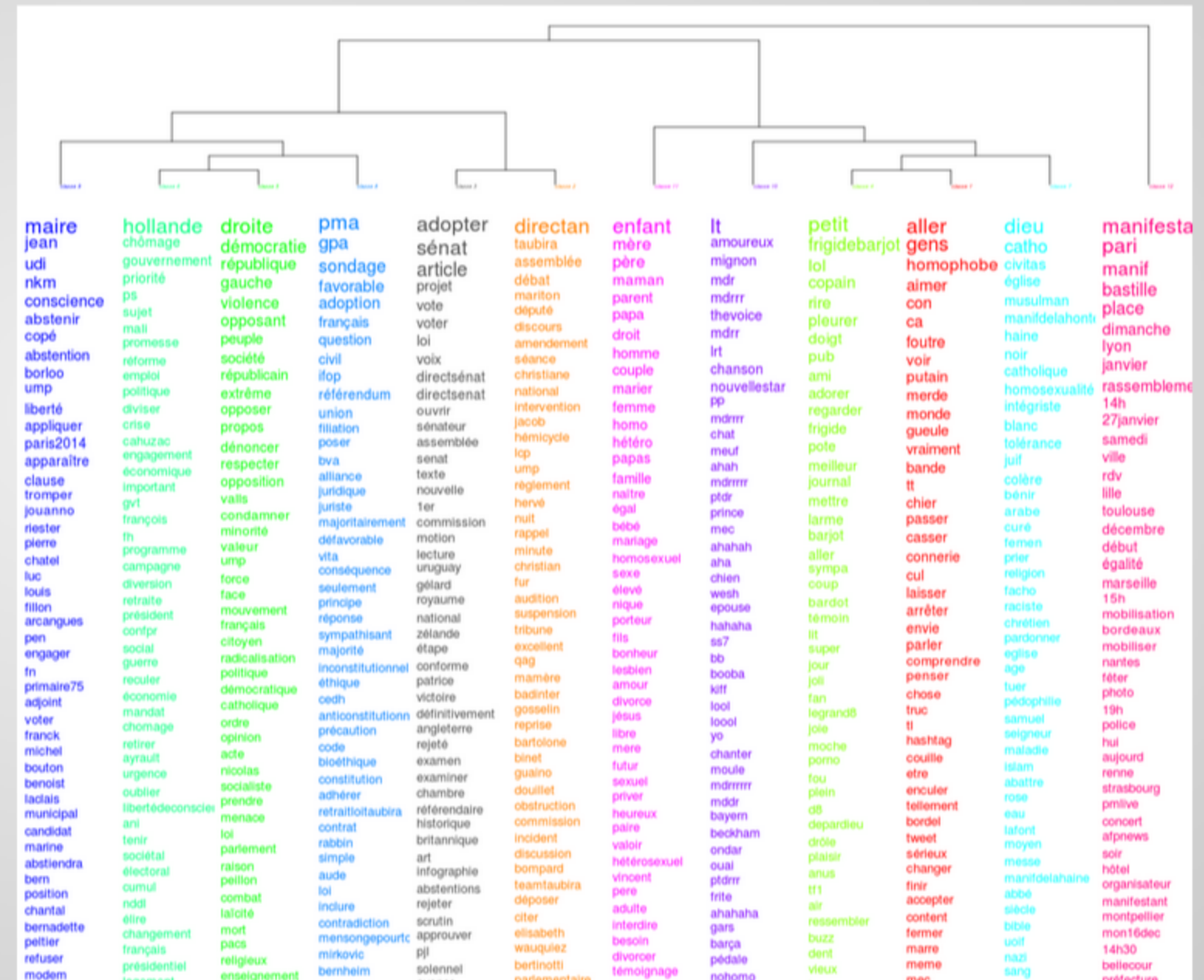
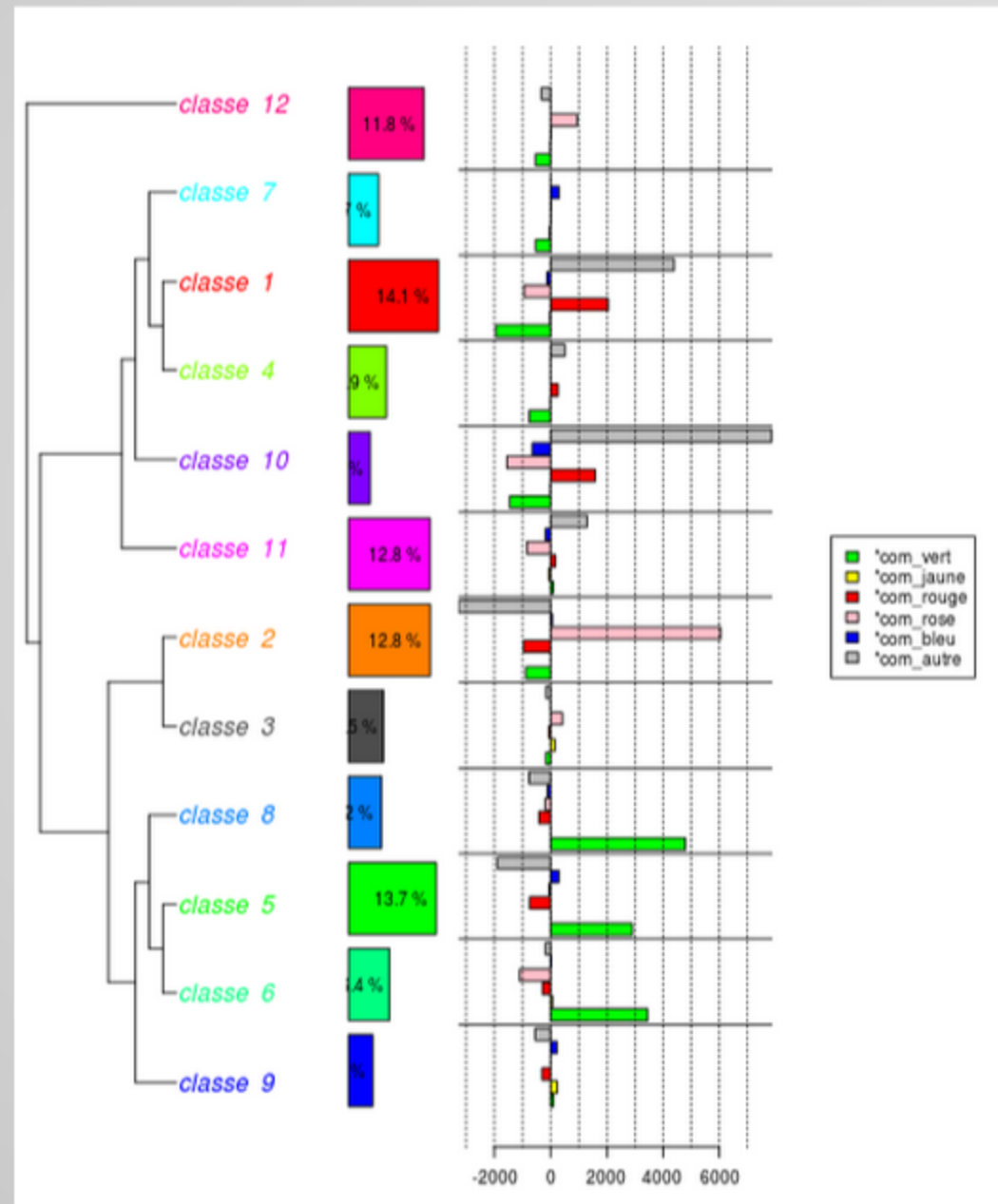
Détermination des communautés



Communautés dans le corpus



Lien entre communautés et classes de discours

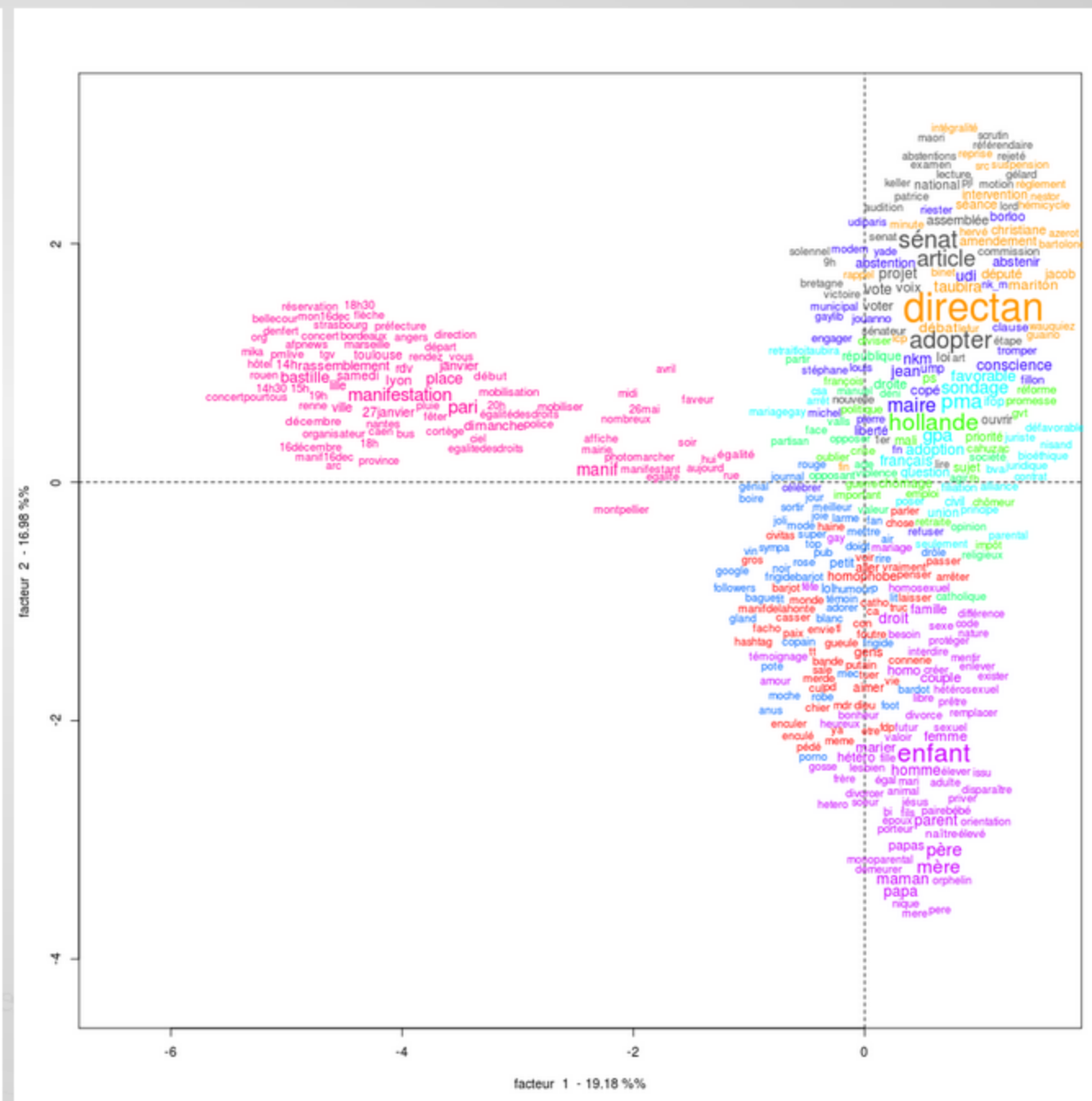
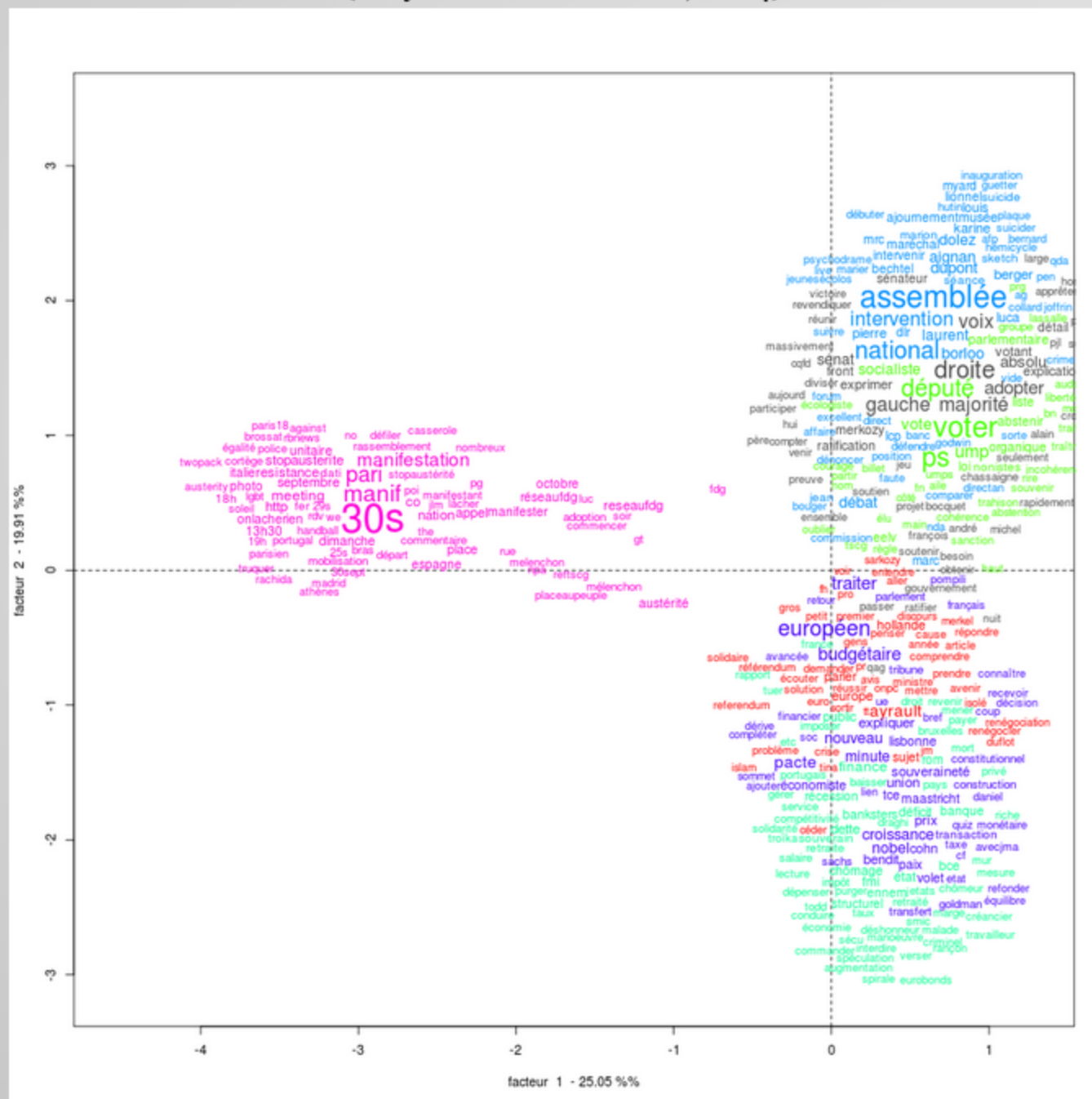


Communautés dans le corpus

Stabilité des mondes lexicaux ? Des structures lexicales ?

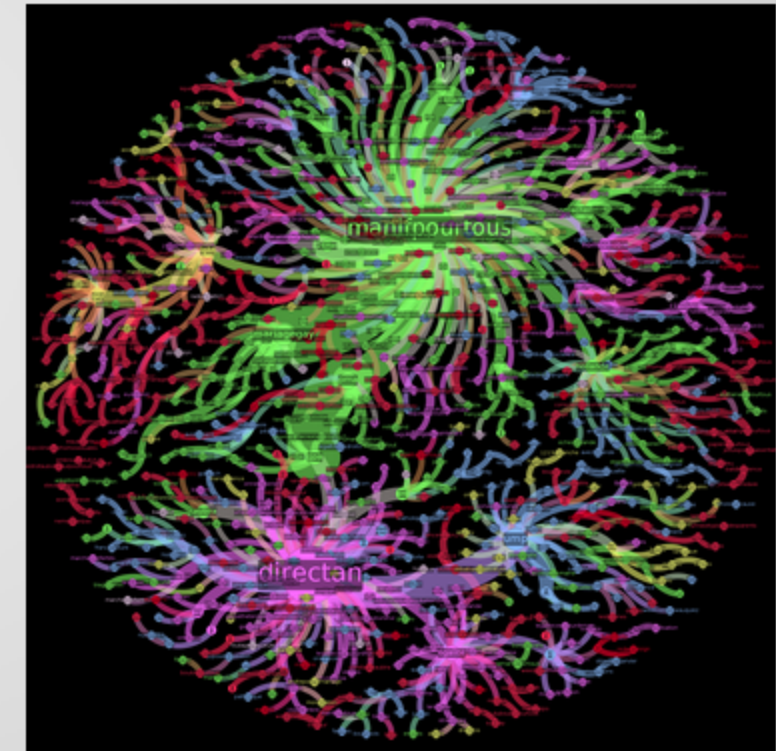
#TSCG - 9312 tweets du 29/09/2012 au 19/10/2012
(Smyrnaiois & Ratinaud, 2014)

#mariagepour tous 453107 tweets du 07/10/2012 au 31/07/2013



Et ensuite...

- projection chronologique des liens classes / communautés
- injection des retweets dans l'analyse
- étude des hastags mobilisés
- comparaison avec la presse
- comparaison avec les débats à l'assemblée et au sénat
- ~~étude des urls~~
- comparaison avec #mariagegay et #mariagehomo





Analyse lexicométrique des tweets sur le #mariagepourtous

Pascal Marchand (pascal.marchand@iut-tlse3.fr)

Pierre Ratinaud (ratinaud@univ-tlse2.fr)

