

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

Éric Leclercq\*, Sergey Kirgizov\*, Maximilien Danisch\*\*

\*LE2I CNRS UMR 6306 - Université de Bourgogne Franche-Comté  
Eric.Leclercq@u-bourgogne.fr, Sergey.krigizov@u-bourgogne.fr

\*\* Télécom ParisTech

**Résumé.** La notion de communauté est largement traitée dans l'analyse des réseaux complexes. Cet article propose de montrer comment une approche pluridisciplinaire revisite la notion de communautés en la connectant à une question de recherche en sciences de l'information et de la communication.

## 1 Introduction et contexte

Le travail que nous décrivons s'inscrit dans le projet international et pluri-disciplinaire TEE 2014 qui vise à étudier la structure de la communication politique sur Twitter durant les élections européennes de 2014, dans 6 pays (France, Allemagne, Italie, Espagne, Belgique, Royaume-Uni). Le corpus généré comporte plus de 50M de tweets sur la période de la campagne des élections. Les données issues des tweets, par les opérateurs (reply, retweet RT, mention @, hashtag # et URL) prennent la forme d'un graphe hétérogène (réseau complexe multi-relationnel) de 50Go environ. Parmi les questions scientifiques abordées par les chercheurs en sciences de la communication (SC), la recherche de communautés (de petite taille), d'utilisateurs échangeant autour de hashtags spécifiques est un enjeu pour la compréhension des formes de communication. À partir des travaux de Goodin et Dryzek (2006), les chercheurs en SC impliqués dans le projet ont proposé une première définition : « 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 ».

Cette définition ne permet pas directement d'établir un lien avec les algorithmes existant. Une analyse exploratoire a été conduite, en appliquant les principaux algorithmes de détection de communautés, mais aucun n'a permis de faire apparaître de manière pertinente les petites communautés recherchées. Seul l'algorithme Walktrap de Pons et Latapy (2005) appliqué sur le réseau multi-relationnel (RT, #, @), associé à une visualisation enrichie avec la connaissance du domaine a mis en évidence des éléments singuliers. En effet, disposant d'une ontologie de domaine constituée par la description des partis politiques, des candidats et de règles logiques, nous avons découvert quelques candidats singuliers, appartenant à un parti mais classés dans une communauté qui rassemble exclusivement des candidats d'un autre parti. À partir de ces premiers éléments et avec l'expérience de l'étude des tweets, la définition des mini-publics a été précisée.

## 2 Modèle et algorithme pour la détection des mini-publics

L'équipe Media-studies de l'Université de Bonn a proposé une définition empirique des mini-publics comportant 10 critères que nous avons partiellement traduit dans un modèle exécutable. Les critères mesurables retenus sont relatifs à la structure du graphe : utilisation intensive de hashtags spécifiques, des opérateurs retweets et reply par un ensemble d'utilisateurs *a priori* non borné. Un mini-public correspond donc à une communauté ego-centrée également désignée par communauté locale à un nœud, Bagrow et Bollt (2005), Danisch et al. (2013). Le critère d'intensité relève à la fois de la temporalité et de la structure. Il est évalué lors de la constitution des jeux de données sous la forme d'un facteur de vieillissement (par exemple proportionnel à la fréquence d'utilisation d'un hashtag par un utilisateur) appliqué sur le poids des liens.

Pour détecter les mini-publics, nous utilisons comme point de départ l'algorithme proposé par Danisch et al. (2013). À partir d'un ensemble de nœuds d'intérêt, une valeur est propagée aux autres nœuds du graphe afin d'en mesurer la proximité. Ainsi, chaque nœud possède une valeur qu'il transmet à ses voisins proportionnellement aux poids des liens. À l'issue d'itérations de ce processus, l'ensemble des valeurs des nœuds doit être stable pour établir un ranking de tous les nœuds du graphe par rapport aux nœuds d'intérêt. L'étude de l'allure de la courbe représentant la proximité en fonction du rang fait apparaître des plateaux qui correspondent à des structures communautaires. La convergence de l'algorithme a été prouvée en utilisant la théorie des chaînes de Markov.

## 3 Résultats, discussion et conclusion

Avec les chercheurs en SC, nous avons déterminé un ensemble d'expériences, dont quelques résultats sont visibles à l'adresse (<http://eric-leclercq.fr/minipublics/>). Nous avons appliqué l'algorithme sur des graphes user/hashtag non dirigés (pondérés), sur des graphes dirigés (user/hashtag/mention/RT) pondérés avec facteur de vieillissement pour des nœuds d'intérêt spécifiques. Les mini-publics obtenus, soumis à l'interprétation des chercheurs en SC, ont validé globalement l'algorithme et montré sa capacité à détecter des mini-publics pertinents, imbriqués dans certains cas, ainsi que les frontières entre plusieurs mini-public. Les travaux actuels se concentrent : 1) sur la modélisation de la notion de conversation de manière à mieux prendre en compte la définition originale 2) sur la modélisation des relations complexes par des hypergraphes.

## Références

- Bagrow, J. et E. Bollt (2005). A local method for detecting communities. *Phys. Rev.*
- Danisch, M., J.-L. Guillaume, et B. Le Grand (2013). Towards multi-ego-centred communities : a node similarity approach. *Int. Journal of Web Based Communities* 9(3), 299–322.
- Goodin, R. et J. S. Dryzek (2006). Deliberative impacts : the macro-political uptake of mini-publics. *Politics & Society* 34, 219–244.
- Pons, P. et M. Latapy (2005). Computing communities in large networks using random walks. In *Computer and Information Sciences-ISCIS 2005*, pp. 284–293. Springer.