

Evaluation de l'influence sur Twitter: Application au projet "Twitter aux Elections Européennes 2014"

Lobna Azaza¹, Sergey Kirgizov¹, Eric Leclercq¹, Marinette Savonnet¹, Alex Frame²

1. LE2I (UMR 6306), Université de Bourgogne, 21000 Dijon, France

Lobna.Azaza@u-bourgogne.fr

Sergey.Kirgizov@u-bourgogne.fr

Eric.Leclercq@u-bourgogne.fr

Marinette.Savonnet@u-bourgogne.fr

2. Centre Interlangues Text, Image, Langage (EA 4182), Université de Bourgogne, 21000 Dijon, France

Alexander.Frame@u-bourgogne.fr

Mots clés: Web politique, Influence sur Twitter, Marqueurs d'influence, Dynamique des réseaux

1. Introduction

Le Web a eu les impacts les plus importants sur la vie humaine tout au long de ces dernières années. Les utilisateurs dans les réseaux sociaux sont de plus en plus actifs et génèrent une grande quantité de nouvelles données. L'un des caractéristiques distinctives des réseaux sociaux en ligne est la diffusion de l'information à travers les liens sociaux [6]. En effet, les liens entre les utilisateurs déterminent la circulation de l'information et donc indiquent l'influence d'un utilisateur sur d'autres, un concept qui est essentiel en sociologie.

Le Web et les réseaux sociaux ont pris tout leur sens en stratégie de communication en politique en étant désormais au cœur de la société démocratique [5]. Avec une efficacité sans précédent, ils permettent à des groupes d'internautes de se réunir et rendent plus transparente l'action des gouvernements et des responsables politiques, ce qui favorise la participation des citoyens aux processus démocratiques. En particulier, ils peuvent reconnecter les citoyens avec leurs institutions démocratiques (parlements et partis politiques), grâce à de nouvelles formes de relations, plus interactives. Ainsi, l'influence qu'un candidat peut avoir sur les internautes, et la façon avec laquelle les candidats communiquent sur les réseaux sociaux sont devenues un enjeu majeur.

Dans ce travail, nous proposons une approche pour mesurer l'influence des candidats sur Twitter et analysons son évolution à travers le temps. La mesure de cette influence est principalement basée sur les différentes interactions entre les internautes [4].

2. TEE 2014

Les travaux de recherche menés se déroulent dans le cadre du projet TEE 2014 dont l'intitulé exact est : « Twitter aux élections européennes : Une étude contrastive internationale des utilisations de Twitter par les candidats aux élections au Parlement Européen en mai 2014 ». Ce projet international, mené par la Maison des Sciences de l'Homme (MSH) de Dijon, réunit près de 45 chercheurs (majoritairement politologues, sociologues, chercheurs en communication), de 10 laboratoires de recherche répartis sur 6 pays européens (la France, l'Allemagne, la Belgique, l'Italie, l'Espagne et l'Angleterre). L'objectif global de ce projet est d'observer et d'analyser la communication des politiques sur Twitter durant la période électorale de mai 2014 dans différents pays d'Europe Occidentale.

La collecte des Tweets durant la période électorale a permis de constituer un corpus qui sera ensuite analysé. Trois types d'informations (généralisées sous le terme "source") sont pris en paramètre dans cette collecte : des comptes utilisateurs ; des hashtags et des mots ou phrases. L'objectif de la collecte est de récupérer les tweets mentionnant les utilisateurs désignés, ceux contenant un certain hashtag, un mot ou une phrase, ou encore les tweets envoyés par les utilisateurs particuliers. Ces

différentes sources ont été choisies par les politologues, et on retrouve parmi elles les noms des principaux candidats, leurs comptes Twitter, et les hashtags relatifs à ces candidats, leurs partis, ou plus généralement les élections. L'infrastructure déployée pour la collecte permet de disposer d'un très grand nombre de tweets (37 millions) recueillis pendant 50 jours consécutifs, et de traiter massivement ces données.

Dans nos travaux, nous analysons la diffusion de l'information à grande échelle à travers les données que nous avons collectées de Twitter. En outre, nous recueillons des informations sur ces tweets tel que les tweets retweetés et les utilisateurs mentionnés dans les tweets.

Nous analysons les données collectées pour répondre aux questions suivantes :

- Comment mesurer l'influence ? Qui sont les candidats les plus influents ?
- Comment cette influence évolue à travers le temps ?
- Comment l'information se propage sur Twitter ?

Dans cet article, nous aborderons les deux premières questions.

3. Influence sur Twitter : un aperçu

L'influence sur Twitter est définie comme le potentiel d'une action d'un utilisateur à lancer une nouvelle action par un autre utilisateur [1]. Le terme utilisateur est défini par la plate-forme Twitter. Le terme action mérite de plus amples explications. Comprendre le terme action en ce qui concerne l'influence sur Twitter dépend de la structure fondamentale des idées dans l'environnement et la façon dont ces idées se propagent [3]. Ainsi, le nombre de followers d'un utilisateur peut être utilisé pour mesurer l'influence [4], c'est-à-dire le nombre d'abonnés ou de personnes qui suivent un utilisateur et ses tweets. Dans [6], les auteurs utilisent le nombre de commentaires et le nombre de favoris pour mesurer l'influence. Les travaux décrits dans [7] visent à mesurer l'influence selon trois critères différents : nombre de followers, retweets et mentions. Mais ils n'offrent pas de combinaison des critères proposés. Dans [2], les auteurs mesurent justement l'influence en combinant deux critères : similarité des sujets et structure des liens entre les utilisateurs, mais ils ignorent d'autres critères qui sont importants tels que le retweet et la mention.

Les travaux existants proposent de nombreuses méthodes pour mesurer l'influence sur les réseaux sociaux. Cependant, aucun d'entre eux présente une approche globale qui combine les différents critères qui peuvent marquer l'influence. De plus, ils existent des critères importants comme les réponses et l'utilisation des hashtags, qui marquent bien l'influence mais qui n'ont jamais été utilisés dans le but de la mesurer.

4. Approche proposée

Nous proposons dans cet article une approche qui repose sur un modèle et une méthode de mesure d'influence des candidats sur Twitter.

4.1. Modélisation

Les unités fondamentales sur Twitter sont l'utilisateur et le tweet. Une action sur Twitter comprend toutes les interactions d'un utilisateur avec ce tweet. Tandis que les actions sur Twitter comprennent à la fois les interactions reconnues par la plate-forme ainsi que des comportements émergents inattendus qui deviennent largement utilisés par les utilisateurs. Twitter reconnaît trois actions qui peuvent se produire : la réponse, le retweet et le favoris. Ils sont classés comme actions parce qu'ils sont appliqués par un utilisateur pour un autre tweet. Bien que les actions ont des buts différents, les trois sont destinées à propager le contenu à d'autres utilisateurs (quoique de manière différente). Si une réponse, un retweet, ou un favoris existe par rapport à un tweet donné, ces trois actions sont la preuve de l'influence qui s'est produite. Nous utilisons aussi le nombre de followers comme critère pour mesurer l'influence qui est une action reconnue par la plateforme Twitter et qui se produit entre un utilisateur et un autre. Littéralement, les actions sont des marqueurs d'influence.

D'autres actions qui apparaissent fréquemment sur Twitter, non reconnues par le système mais assez populaires pour être adoptées par les utilisateurs : la mention, le hashtag et le lien URL.

Similaire à la réponse, le retweet, et le favoris, ces actions sont classées comme marqueurs d'influence parce qu'elles sont aussi appliquées par un utilisateur pour un contenu de tweet.

Les différentes actions qui s'appliquent entre les deux éléments fondamentaux de Twitter (utilisateurs U et tweets T) se font à travers différents opérateurs représentés dans le *Tableau 1* :

Opérateur	Action
@ : U × U × T	mention
RT : U × U × T	retweet
# : U × T × H	hashtag
URLs : U × T × URLs	lien URLs
Resp : U × T × U × T	réponse
Abonné : U × U	follower
Favoris : U × U × T	favoris

Tableau 1 : Modèle opérateur- action

4.2. Vers l'élaboration d'une mesure multicritères d'influence

Dans cet article, en utilisant des données collectées de Twitter sur les élections européennes 2014 (37 millions de tweets), nous comparons les différents marqueurs d'influence définis précédemment puis nous proposons une nouvelle mesure multicritères de l'influence en se basant sur la combinaison des différents marqueurs. A travers ces mesures, nous étudions l'influence des utilisateurs et son évolution temporelle.

Le *Tableau 2* issu d'une expérimentation sur le corpus français présente le classement des dix premiers candidats les plus influents selon les différents marqueurs (abonnées, retweets, mentions, réponse) et montre qu'en fonction du marqueur, la position du candidat n'est pas la même. D'où l'importance de la définition d'une mesure multicritères d'influence prenant en compte contexte.

Classement	Abonnées	Retweet	Mention	Réponse
1	Marine Le Pen	Marine Le Pen	Marine Le Pen	Christine Boutin
2	Jean-Luc Melenchon	Florian Philippot	Christine Boutin	Marine Le Pen
3	Eva Joly	Jean-Luc Melenchon	Jean-Luc Melenchon	Florian Philippot
4	Christine Boutin	Aymeric Chauprade	Florian Philippot	Jean-Luc Melenchon
5	Nicolas Dupont-Aignan	François Asselineau	Nicolas Dupont-Aignan	Louis de Gouyon Matignon
6	Vincent Peillon	Corinne Morel-Darleux	José Bove	Nicolas Dupont-Aignan
7	Corinne Lepage	Nicolas Dupont-Aignan	Aymeric Chauprade	Jean-Sébastien Herpin
8	Philippe Poutou	Louis Aliot	Raquel Garrido	Julien Rochedy
9	Chantal Jouanno	Denis Payre	Jérôme Lavrilleux	Geoffroy Didier
10	José Bove	Yannick Jadot	Marielle de Sarnez	Louis Aliot

Tableau 2 : Classement des dix premiers candidats les plus influents

L'approche proposée combine différents critères de mesure d'influence ou marqueurs d'influence : le nombre de followers, les retweets, les mentions, les réponses, les favoris, les hashtags, et les URLs. Nous utilisons des modèles théoriques de représentation et de raisonnement sur les informations incertaines. Plus particulièrement, il s'agit d'utiliser les modèles basés sur les fonctions de croyance, en vue de tenir compte de toute forme d'incertitude dans les étapes de calcul de l'influence.

Nous représentons l'influence qui peut être exercée d'un utilisateur sur un autre et la propagation d'un tweet dans le réseau à travers des graphes de diffusion. La *Figure 1* illustre un exemple d'un graphe de diffusion. Les nœuds représentent les utilisateurs et les liens entre ces nœuds sont les différentes formes d'influence qui existent entre eux. Le graphe est orienté pour montrer le sens de propagation d'un tweet. Il contient différents niveaux. L'influence directe est représentée dans un premier niveau. Cette influence est celle produite entre un utilisateur et ses followers directs. Alors que l'influence indirecte désigne le cas où le tweet est diffusé jusqu'à atteindre les followers de ses followers directs. L'influence indirecte est représentée sur des niveaux secondaires. Ce graphe permet de connaître les nœuds à partir desquels l'information est diffusée avec un taux de diffusion assez important par rapport aux autres nœuds.

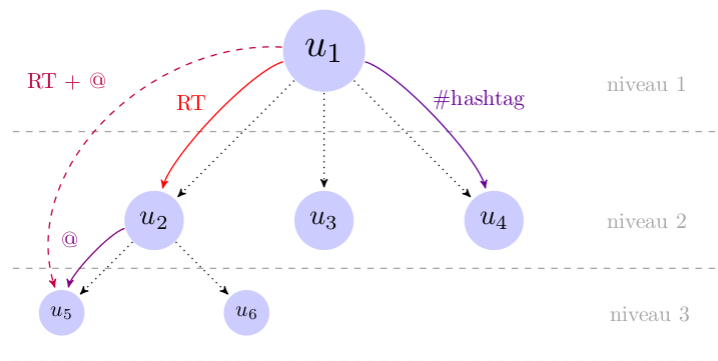


Figure 1. Exemple d'un graphe de diffusion

Nous étudions l'évolution de l'influence à travers le temps. Les candidats les plus influents ne sont pas les mêmes avant et après les évènements (par exemple avant /après le jour des élections). L'utilisation d'algorithmes de détection d'évènements comme changepoint [8] mettent en évidence des évènements importants.

Les expériences effectuées sur le corpus français nous ont permis de faire plusieurs observations intéressantes. Premièrement, les candidats les plus influents selon le critère nombre de followers ne sont pas nécessairement les candidats les plus influents en termes de retweets ou mentions. Deuxièmement, l'influence n'est pas acquise spontanément ou accidentellement, mais à travers des efforts concertés comme la publication massive des tweets par les candidats pour avoir plus de propagation de leurs tweets et atteindre un maximum d'internautes. Troisièmement, l'influence des utilisateurs n'est pas stable et évolue à travers le temps.

5. Conclusion

Dans ce travail, nous exploitons un grand volume de données collectées de Twitter dans le contexte des élections européennes 2014, environ 37 millions de tweets collectés pendant 50 jours. Nous proposons une mesure multicritères de l'influence sur Twitter tout en tenant compte de toute forme d'incertitude observée dans les différentes étapes de calcul. Nous étudions aussi l'évolution de l'influence à travers le temps et représentons la propagation de l'information à travers des graphes de diffusion. Les travaux effectués nous ont conduit à tirer des observations intéressantes.

Références

- [1] Alex Leavitt, Evan Burchard, David Fisher, et Sam Gilbert. The Influentials: New Approaches for Analyzing Influence on Twitter. *Analyzing Influence on Twitter* (2009).
- [2] Jianshu Weng, Ee-Peng Lim, Jing Jiang, et Qi He. 2010. TwitterRank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining (WSDM '10)*. ACM, New York, NY, USA, 261-270.
- [3] Jimene Sun et Jie Tang. A survey of models and algorithms for social influence analysis. *Social network data analysis*. pp 177-214. (2011).
- [4] Maximilien Danisch, Nicolas Dugué, et Anthony Perez. On the importance of considering social capitalism when measuring influence on Twitter. *2014 International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC 2014)*.
- [5] McClurg, Scott D. Social Networks and Political Participation: The Role of Social Interaction in Explaining Political Participation. *Political Research Quarterly*. pp 449-464.
- [6] Meeyoung Cha, Alan Mislove, et Krishna P. Gummadi. 2009. A measurement-driven analysis of information propagation in the flickr social network. *Proceedings of the 18th international conference on World wide web (WWW '09)*. ACM, New York, NY, USA, pp 721-730.
- [7] Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, et Krishna P. Gummadi. Measuring user influence in Twitter: The million follower fallacy. *ICWSM '10: Proceedings of international AAAI Conference on Weblogs and Social* (2010).
- [8] Rebecca Killick et Idris A. Eckley. changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software* (2014).