



HAL
open science

Fouille de données hétérogènes pour la caractérisation de fake news : analyse empirique croisée de la presse en ligne et Twitter

Matthieu Bachelot, Inna Lyubareva, Thomas Epalle, Romain Billot,
Raphaël-David Lasserri

► To cite this version:

Matthieu Bachelot, Inna Lyubareva, Thomas Epalle, Romain Billot, Raphaël-David Lasserri. Fouille de données hétérogènes pour la caractérisation de fake news : analyse empirique croisée de la presse en ligne et Twitter. SCAI : Infox sur Seine, François Bancilhon, Mar 2023, Paris, France. hal-04221835

HAL Id: hal-04221835

<https://hal.science/hal-04221835>

Submitted on 28 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fouille de données hétérogènes pour la caractérisation de fake news : analyse empirique croisée de la presse en ligne et Twitter

Matthieu Bachelot^{1,2,3}, Inna Lyubareva¹, Thomas Epalle³, Romain Billot²,
Raphaël-David Lasserri³

¹ IMT Atlantique, LEGO, France

² IMT Atlantique, DECIDE, France

{prenom.nom}@imt-atlantique.fr

³ Magic LEMP, France

{nom}@magic-lemp.com

Mots-clés : *informations fallacieuses, analyse des réseaux sociaux, graphes, propagation, fouille de données, visualisation*

1 Introduction

Les réseaux sociaux jouent un rôle significatif dans la transmission de l'information, notamment de par leur facilité d'accès sur plusieurs supports et leur ergonomie. Cependant, l'abondance et la diversité des points de vues permises par le contexte numérique ne sont pas toujours synonymes d'une meilleure qualité de l'information en ligne. Parmi les problèmes récents, la prolifération des rumeurs ou fausses informations (fake news) constitue un enjeu majeur pour l'ensemble des acteurs économiques du secteur, mais aussi plus largement en matière de sérénité du débat public voire de sécurité nationale. Le problème de la détection des informations fallacieuses a été étudié depuis longtemps [6], mais les approches se basant uniquement sur les contenus deviennent insuffisantes dans le contexte actuel [5]. Des techniques utilisant d'autres leviers ont été développées pour essayer de détecter ces fake news de manière précoce, en y joignant par exemple des profilages d'utilisateurs [3] ou bien des analyses temporelles [4]. Cependant ces méthodes sont encore perfectibles et surtout se fondent trop souvent sur un nombre réduit de jeu de données de référence (Weibo, Twitter15, Twitter16 ...). Le corollaire est qu'il n'y a que très peu de travaux de recherches dans la détection d'informations fallacieuses sur un jeu de données en français, quel que soit le support (liste d'articles, base de données de tweets). Notre travail vise ainsi à combler ces lacunes et à proposer une des premières analyses empiriques en langue française, en croisant plusieurs sources de données.

2 Méthodologie

2.1 Données

L'originalité de notre approche réside dans la proposition d'une méthodologie de modélisation d'informations fallacieuses qui utilise plusieurs support de propagations : d'un côté, des articles de presse issus des sites web des principaux médias français, et de l'autre la base de données du réseau social Twitter. Nous avons sélectionné plusieurs événements (ie. un ensemble cohérent d'articles et de tweets autour d'un même sujet) pouvant s'apparenter à des rumeurs ou fake news ainsi que d'autres correspondant à des événements normaux (vraies informations) dans une base de données contenant plus d'un millions d'articles et un demi million de tweet récoltés depuis Novembre 2021. Cette approche par événement est d'autant plus intéressante que les résultats de l'analyse seront plus facilement généralisables.

2.2 Modélisation et analyses

Nous avons modélisé les événements par deux graphes, un par support étudié (Twitter et articles). Pour les données Twitter, le graphe a été reconstitué à partir des retweets et des réponses. L'analyse de ces graphes, notamment des composantes connexes a mis en lumière leurs différences de taille (en terme de nombre de sommets). Aussi nous avons fait le choix de nous intéresser à la plus grande composante connexe pour chaque événement dans le cadre des graphes Twitter.

Pour les articles, les arêtes du graphes s'appuient sur la notion de proximité sémantique des contenus. Après avoir sélectionné un échantillon d'articles traitant de l'évènement, nous avons testé plusieurs méthodes d'embedding des contenus dans l'optique de calculer leurs similarité deux à deux dans l'espace latent. Ces calculs de similarité ont permis de générer une matrice d'adjacence pour pouvoir créer le graphe de propagation de l'information dans la presse écrite. Une fois les graphes obtenus nous avons divisé notre étude en trois axes : un sémantique, un structurel et un dernier temporel.

L'étude sémantique : Nous avons complété l'analyse des plongements sémantiques afin de valider notre méthode et de visualiser l'évolution du traitement de l'information au cours du temps. Pour ce faire, nous avons comparé l'évolution de la distance des plongements sémantiques des articles et des tweets dans l'espace latent de plusieurs manières différentes (par rapport à un article ou tweet de référence et sur une fenêtre glissante de taille variable).

L'étude structurelle : Elle porte sur l'étude des marqueurs structurels des graphes de propagations. Nous avons voulu confirmer ou non les résultats produits par différentes études descriptives sur des données anglophones [1] récentes ainsi que sur d'autres travaux de détections d'informations fallacieuses [2]. Nous avons conduit plusieurs tests à différents niveaux de granularité (égo, communauté et le graphe dans son ensemble) afin avoir une compréhension plus fine de l'organisation des différents acteurs de la propagation de l'information. De plus, nous avons testé différents types de graphes (notamment types de pondérations) pour voir si certaines configurations permettent de mieux différencier la propagation des vraies des fausses informations.

L'étude temporelle : Nous avons enfin réalisé une étude temporelle exhaustive afin d'étudier plusieurs aspects relatifs à la propagation de l'information. Cette analyse temporelle se divise en trois parties : une analyse des délais de propagation, une analyse des sous-graphes représentant les différentes phases de la propagation de l'information, et enfin l'analyse de la propagation au stade précoce. Pour générer les sous-graphes, nous avons analysé les tendances globales des tweets (respectivement des articles) pour déterminer le nombre de périodes d'intérêt. Nous avons alors reconstruit les graphes en isolant les sommets dans ces périodes, et nous avons analysé leurs structure. Enfin en ce qui concerne la détection précoce, nous avons regardé la dynamique de propagation entre le début de l'évènement (défini à la main) jusqu'à un certain seuil (12H, 24H 72H ...) et nous avons étudiées les différences entre les distributions des délais ainsi que les différentes structure de graphes.

3 Premiers résultats

L'analyse sémantique a permis de mettre en évidence une évolution du traitement de l'information au cours du temps entre les vraies et les fausses informations. Dans le cadre de l'analyse des distances sémantiques au cours du temps, les événements traitant de fausses informations ont une distance systématiquement plus faible (en moyenne et en écart type) que les événements traitant des vraies informations, ce qui indiquerait que les rumeurs utilisent la même rhétorique au cours du temps, alors que les vrai informations suivent le cours de l'actualité.

Dans le cadre de l'étude sur les différences structurelles, nous avons dégagé des tendances distinctes entre les événements vrai et fake qui confirment certaines tendances décrites dans

l'état de l'art. Ces tendances en questions portent sur des mesures de *centralité* dans les graphes de propagations complets, autrement dit une mesure de l'importance d'un sommet dans ces graphes. Nous avons répété ces calculs avec différentes modélisations (graphes pondérés en fonction des délais de propagation par exemple), ce qui a permis de trouver des indicateurs discriminants entre vrai et fausses informations. De plus, nous avons notamment mis en évidence l'importance de certains utilisateurs Twitter (nommés *leader* ou *super spreader* dans la littérature anglophone). Ces *leaders* ont des caractéristiques différentes en fonction du type d'information propagées (vrai ou fausse information) .

L'analyse temporelle a permis de visualiser deux dynamiques bien particulières entre vraies et fausses informations. Bien que l'analyse des délais de propagation n'a pas donné de différences significatives entre vraies et fausses information, le découpage des graphes en sous graphes temporels a été plus fructueuse. La conclusion principale est que la dynamique de propagation de l'information dans la presse écrite et dans le réseau sociale Twitter est "découplée" dans le cadre des fausses informations.

De manière générale les premiers résultats empiriques présentés dans ce travail sont prometteurs et permettent de décliner un certain nombre de propriétés de graphe inhérentes aux informations fallacieuses, en se fondant sur les métriques classiques de l'analyse des réseaux sociaux et des analyses statistiques uni- et multidimensionnelles.

Références

- [1] A. Bodaghi and J. Oliveira. The theater of fake news spreading, who plays which role? A study on real graphs of spreading on Twitter. *Expert Systems with Applications*, 189 :116110, Mar. 2022.
- [2] M. Davoudi, M. R. Moosavi, and M. H. Sadreddini. DSS : A hybrid deep model for fake news detection using propagation tree and stance network. *Expert Systems with Applications*, 198 :116635, July 2022.
- [3] Y.-J. Lu and C.-T. Li. GCAN : Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media, Apr. 2020. Number : arXiv :2004.11648 arXiv :2004.11648 [cs, stat].
- [4] N. Ruchansky, S. Seo, and Y. Liu. CSI : A Hybrid Deep Model for Fake News Detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 797–806, Singapore Singapore, Nov. 2017. ACM.
- [5] K. Shu, S. Wang, and H. Liu. Beyond News Contents : The Role of Social Context for Fake News Detection. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 312–320, Melbourne VIC Australia, Jan. 2019. ACM.
- [6] Tiantian Qin, J. Burgoon, J. Blair, and J. Nunamaker. Modality Effects in Deception Detection and Applications in Automatic-Deception-Detection. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, pages 23b–23b, Big Island, HI, USA, 2005. IEEE.