

Détection non supervisée de position dans les médias sociaux : une approche générique

Maïa Sutter*, Antoine Gourru*
Christine Largeron* Amine Trabelsi**

* Laboratoire Hubert Curien, UMR CNRS 5516, Saint-Etienne, France
maia.d.sutter@gmail.com, antoine.gourru@univ-st-etienne.fr,
christine.largeron@univ-st-etienne.fr

** Université de Sherbrooke, Department of Computer Science, Canada
amine.trabelsi@usherbrooke.ca

Résumé. La tâche de détection de position (stance detection) vise à déterminer la position d'un texte ou de son auteur, comme étant pour ou contre un sujet cible donné. Pour résoudre cette tâche sans annotation coûteuse de données, l'approche non supervisée est privilégiée. Néanmoins, les modèles actuels ont été conçus pour des types de réseaux spécifiques, soit homophiliques, soit hétérophiliques. Dans cet article, nous étudions tout d'abord la capacité de généralisation des modèles existants récents à ces deux types de réseaux très différents. Nous proposons ensuite un modèle basé sur des plongements de texte propagés avec un réseau neuronal graphique (GNN), applicable aussi bien aux réseaux hétérophiliques qu'homophiliques. Nos expériences confirment que ce modèle surpasse, en moyenne, les autres méthodes pour les deux types de réseaux. En outre, elles montrent que la combinaison d'informations textuelles et de réseaux est plus performante que l'utilisation du texte seul, et que la taille du modèle de langage n'a qu'un impact limité sur les résultats.

Cet article est un résumé en français de l'article "Unsupervised Stance Detection for Social media Discussions : A Generic Baseline", paru à EACL 2024 ¹

1 Introduction

La détection de position (stance detection) consiste à déterminer la position d'un texte ou de son auteur par rapport à une cible donnée, sous la forme d'une étiquette prenant la valeur "pour" ou "contre", ou éventuellement "neutre", "inconnu" ou "ni l'un ni l'autre". La cible peut être une entité, un thème ou un autre sujet, tel qu'une revendication ou un événement. La plupart des travaux dans ce domaine ont été réalisés suivant une approche de classement supervisé comme (Largeron et al. (2021)). Seule une gamme limitée de méthodes, par exemple, (Trabelsi et Zaiane (2018)), a été proposée récemment en suivant une approche non supervisée qui s'avère plus appropriée pour les sujets où l'annotation est coûteuse ou lorsque les positions sont très

1. <https://aclanthology.org/2024.eacl-long.107.pdf>

Détection non supervisée de position dans les médias sociaux

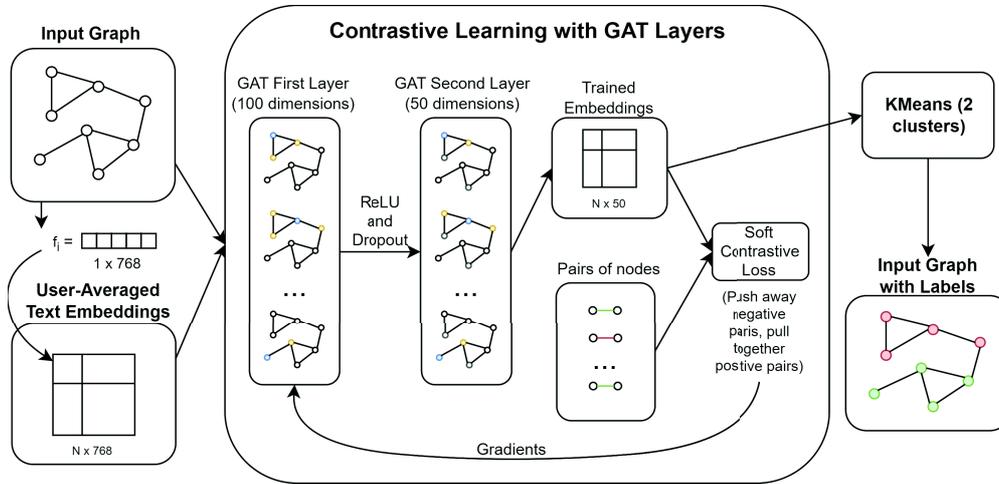


FIG. 1 – Architecture de GUSD.

évolutives notamment sur des sujets d'actualité. Par ailleurs, la plupart des méthodes, qu'elles soient supervisées ou non, exploitent un seul type d'information, soit le contenu textuel comme Ghosh et al. (2018) soit le réseau d'interaction des utilisateurs modélisé par un graphe, comme Li et al. (2022).

De plus, les méthodes existantes ont le fort désavantage d'être conçues pour un type spécifique de source de données et peuvent difficilement être généralisées à d'autres types. En effet, des plateformes telles que X (anciennement Twitter) tendent davantage vers l'homophilie (Khanam et al. (2023)) car les utilisateurs ont tendance à interagir avec d'autres utilisateurs qui partagent leurs opinions (McPherson et al. (2001)), tandis que d'autres, telles que les forums de débat, sont plutôt hétérophiliques (Pick et al. (2022)). De ce fait, les méthodes actuelles fournissent de bons résultats pour l'un ou l'autre de ces types de réseaux, mais beaucoup moins bons sur des données ne correspondant pas au type de réseau pour lequel elles ont été conçues.

À notre connaissance, nous sommes les premiers à proposer une méthode non supervisée de détection de position qui exploite à la fois les informations sémantiques via les plongements de texte et les informations sur le réseau à l'aide d'un réseau neuronal de graphes (GNN) pour traiter les deux types de réseaux (homophilique et hétérophilique). Cette méthode, GUSD, est entraînée dans un contexte contrastif contrôlable. Comme les approches non supervisées sont rares, nous avons choisi trois méthodes récentes pouvant être considérées comme références par rapport à GUSD. Chacune est conçue pour des réseaux homophiliques ou hétérophiliques, exploite d'une manière ou d'une autre la structure des graphes ou des données textuelles, et ont leur code disponible. Ces méthodes sont InfoVGAE (Li et al. (2022)) et STEM (Pick et al. (2022)), qui n'utilisent pas l'information textuelle, et la méthode de Darwish et al. (2020), qui n'utilise également pas directement l'information textuelle mais la fréquence d'utilisation des hashtags. Nos expérimentations montrent qu'en moyenne, la méthode GUSD est plus performante que ces méthodes de référence. Elles confirment également que la combinaison d'informations textuelles et de réseaux est plus avantageuse que l'utilisation du texte seul, et que la taille du modèle linguistique n'a qu'un impact limité sur les résultats.

Dataset	Source	Nodes	Edges	Textes/utilisateur	Dyadicité	Hétérophilie
Euro16	Twitter	343	654	5.87	1.96	0.11
ConRef	Twitter	178	208	13.40	1.66	0.30
TIMME-Pure	Twitter	389	4544	46.76	1.91	0.02
TIMME-All	Twitter	942	14558	93.62	1.80	0.10
CD-All	CreateDebate	247	724	15.11	0.58	1.41
CD-Abortion	CreateDebate	104	245	14.85	0.55	1.43
CD-GayRights	CreateDebate	90	207	10.87	0.54	1.62
CD-Marijuana	CreateDebate	43	60	7.47	0.63	1.37
CD-Obama	CreateDebate	59	109	10.83	0.53	1.53

TAB. 1 – Caractéristiques des jeux de données. Homophily : Diadicité > 1 ou Hétérophily < 1 - Heterophily : Diadicité < 1 ou Hétérophily > 1

2 Un modèle générique pour la détection non supervisée de la posture (GUSD)

Cette section présente notre modèle, GUSD (Generic model for Unsupervised Stance Detection), dont l’architecture est décrite dans la Figure 1. Dans ce cadre, V représente l’ensemble d’utilisateurs, les interactions entre eux sont modélisées par un graphe $G = (V, E)$ non orienté où il existe une arête pondérée $(v_i, v_j) \in E$ entre v_i et v_j s’ils ont interagi d’une manière ou d’une autre (retweet, mention, réponse, etc.) dans le réseau et le poids de l’arête est le nombre d’interactions entre eux. En outre, chaque utilisateur est associé à un ensemble de documents qu’il a publiés en ligne. Soit F une matrice de taille $|V| \times d$ où chaque ligne f_i correspond à la représentation vectorisée de la production textuelle de l’utilisateur i obtenue via un modèle de langage pré-entraîné, un encoder Transformer BERT Devlin et al. (2019) dans nos expériences. Plus précisément, il s’agit de la moyenne de la représentation des tokens [CLS] de chaque document produit par l’utilisateur.

Nous nous appuyons sur les réseaux d’attention graphique (GAT) (Veličković et al. (2018)) pour construire des représentations informatives des utilisateurs qui intègrent à la fois le graphe et ces représentations textuelles. Nous rappelons que le calcul du plongement $z_i^l \in \mathbb{R}^r$ du nœud v_i dans la couche l est défini comme suit :

$$z_i^l = \sigma \left(\sum_{j=1}^{|V|} \alpha_{ij} W_l z_j^{l-1} \right) \quad (1)$$

avec σ une fonction d’activation, W_l une matrice de poids apprise, et partagée entre tous les nœuds, et α_{ij} un coefficient d’attention appris pour le nœud v_j par rapport au nœud v_i qui capture l’importance des voisins et une représentation initiale z_j^0 égale à f_j .

Comme aucune annotation de la position des utilisateurs n’est fournie dans ce cadre non supervisé, nous utilisons un objectif auto-supervisé comme ((Hamilton et al., 2017)) pour apprendre la représentation des nœuds. La tâche d’apprentissage est la reconstruction du graphe au moyen d’une approche contrastive. L’objectif est de maximiser la probabilité des arêtes observées et de minimiser la probabilité pour un ensemble d’exemples négatifs (un sous-ensemble de nœuds non connectés du graphe). Nous utilisons une perte contrastive dite douce telle que présentée dans Oh et al. (2019) :

$$\mathcal{L}_{softcon} = \begin{cases} -\log p(m|z_1, z_2) & \text{if } \hat{m} = 1 \\ -\log(1 - p(m|z_1, z_2)) & \text{if } \hat{m} = 0 \end{cases} \quad (2)$$

où $p(m|z_1, z_2) := \sigma(-a|z_1 - z_2|_2 + b)$ est la probabilité qu’une paire de nœuds ait une arête entre eux, avec z les plongements finaux (la dernière représentation de la couche GAT), $a > 0$ et $b \in \mathbb{R}$ fournissent un seuil entraînable pour la distance, et \hat{m} est la fonction indicatrice qui vaut 1 pour les paires positives et 0 pour les paires négatives.

Notons que cette stratégie auto-supervisée est tout à fait standard (Oh et al. (2019); Hamilton et al. (2017)), et n’est donc pas bien adaptée pour traiter à la fois les graphes homophiliques et hétérophiliques. Pour contourner ce problème, nous proposons deux versions de cet objectif d’apprentissage. L’une s’adapte à un réseau homophilique, tandis que la seconde peut traiter un réseau hétérophilique.

Pour la version homophilique, les exemples positifs sont des nœuds connectés et les exemples négatifs sont tirés parmi des paires de nœuds non connectés. Plus précisément, pour une paire positive, c’est-à-dire une arête entre les nœuds u et v , nous tirons aléatoirement un nœud w qui n’est pas connecté à u , et nous ajoutons la paire (u, w) à l’ensemble des paires négatives. Pour la version hétérophilique, nous construisons une matrice d’adjacence alternative A' à partir de la matrice d’adjacence originale A de G . Avec $c(\cdot)$, la fonction qui transforme les valeurs positives non nulles en 1 et les valeurs négatives en 0, nous calculons $A' = c(AA^T) - c(A)$. Dans A' , il y a une arête entre deux utilisateurs qui ont interagi avec au moins un utilisateur commun (par exemple, qui ont débattu avec la même personne dans le cadre d’un débat). Il supprime également toutes les interactions directes du graphe initial. Ce processus transforme le graphe de débat/hétérophilique en un graphe homophilique. Les valeurs positives de A' constituent les exemples positifs, tandis que pour les exemples négatifs, nous utilisons les valeurs positives de $c(A)$ pour donner un signal fort des utilisateurs qui devraient être éloignés dans l’espace latent.

Après avoir entraîné le modèle dans ce cadre auto-supervisé, nous réalisons un partitionnement à l’aide de K-means sur les représentations z , fournissant des groupes d’utilisateurs, représentant leur proximité d’opinion et donc leur position.

3 Expériences

Le premier objectif de nos expériences est de comparer GUSD, avec comme encodeur de texte BERT-large, avec l’état de l’art sur divers jeux de données ayant des caractéristiques différentes et de montrer sa capacité à s’adapter à différents types de réseaux (homophilique et hétérophilique, en particulier).

3.1 Protocole

Le tableau 1 présente les caractéristiques des neuf jeux de données utilisés : nombre de nœuds et d’arêtes dans le graphe, nombre moyen de texte par utilisateur et indicateurs d’homophilie. **Euro16** (Li et al. (2022)), **ConRef** (Lai et al. (2018)) et **TIMME** (Xiao et al. (2020)) sont plutôt homophiliques tandis que **CreateDebate** (CD) (Hasan et Ng (2014)) et ses 5 sous-ensembles peuvent être considérés comme hétérophiliques d’après les scores d’hétérophilie et de dyadicité

Dataset	GUSD		InfoVGAE	
	Acc.	F1	Acc.	F1
Euro16	87.75 ± 12.25	87.63 ± 12.59	92.9 ± 1.31	92.69 ± 1.38
ConRef	68.31 ± 6.01	70.5 ± 5.48	53.11 ± 3.11	56.94 ± 2.87
TIMME-Pure	97.89 ± 0.63	97.88 ± 0.63	70.49 ± 1.76	67.69 ± 2.3
TIMME-All	97.26 ± 0.31	97.26 ± 0.3	-	-
CD-All	76.11 ± 1.02	76.23 ± 1.02	47.3 ± 2.08	41.08 ± 2.89
CD-Abortion	61.34 ± 3.71	61.33 ± 3.78	47.09 ± 2.38	40.49 ± 3.86
CD-GayRights	81.33 ± 2.15	81.99 ± 2.01	43.23 ± 15.6	40.22 ± 16.81
CD-Marijuana	69.76 ± 2.94	70.83 ± 2.82	48.8 ± 14.12	46.64 ± 15.04
CD-Obama	78.31 ± 1.97	78.3 ± 1.96	62.77 ± 2.49	59.13 ± 2.54
Mean score	79.78 ± 12.68	80.22 ± 12.35	58.21 ± 16.72	55.61 ± 18.1
Dataset	STEM		Darwish et al.	
	Acc.	F1	Acc.	F1
Euro16	58.89 ± 0.94	59.2 ± 0.99	63.06 ± 3.83	63.18 ± 4.01
ConRef	54.16 ± 0.89	57.77 ± 0.81	94.78 ± 0.27	94.67 ± 0.27
TIMME-Pure	54.32 ± 21.12	53.34 ± 21.31	97.92 ± 0.08	97.92 ± 0.08
TIMME-All	61.21 ± 18.94	56.4 ± 23.19	94.52 ± 0.2	94.53 ± 0.2
CD-All	82.67 ± 1.01	82.73 ± 1.0	56.6 ± 0.46	43.86 ± 0.64
CD-Abortion	75.96 ± 0	75.97 ± 0	54.04 ± 0.88	43.99 ± 0.79
CD-GayRights	85.56 ± 0	85.99 ± 0	-	-
CD-Marijuana	73.02 ± 1.2	73.83 ± 1.15	-	-
CD-Obama	81.19 ± 2.46	81.25 ± 2.38	-	-
Mean score	69.66 ± 12.59	69.61 ± 12.85	76.82 ± 20.97	73.03 ± 25.85

TAB. 2 – Précision moyenne et scores F1 pondérés et écart-type (s.d.) avec score moyen sur tous les jeux de données indiqués en bas.

moyenne pondérée à partir des deux étiquettes (pour et contre) (Park et Barabási (2007)), dont des valeurs supérieures à 1 indiquent respectivement l’hétérophilie ou la dyadicité (c’est-à-dire l’homophilie) et des valeurs inférieures à 1 indiquent l’inverse.

Comparaison avec la vérité terrain Nous comparons notre méthode à InfoVGAE (Li et al. (2022)), STEM (Pick et al. (2022)), et la méthode de Darwish et al. (2020) selon le taux de bien classés et la F1 mesure. Comme un utilisateur publie potentiellement plusieurs messages avec des positions différentes, nous déterminons leur position globale de vérité terrain sur le sujet en sélectionnant leur position la plus fréquente, similairement à la littérature (Li et al. (2022); Darwish et al. (2020)).

Paramètres des compétiteurs et configuration de GUSD²

Compte tenu des contraintes de pages, concernant le paramétrage des méthodes concurrentes, nous renvoyons le lecteur intéressé à la version complète de cet article Sutter et al. (2024).

Pour GUSD, nous utilisons un GAT à deux couches avec des hyperparamètres standard : dimensions cachées de 100 et 50, fonction d’activation ReLU sur la première couche, ainsi qu’une couche de drop-out de 20 % entre les deux couches du GAT.

La phase d’apprentissage auto-supervisé contrastif est exécutée pendant 10 000 itérations avec un arrêt précoce basé sur le score de reconstruction du graphe sur un ensemble de validation, après quoi les plongements générés sont passés par l’algorithme K-means pour générer les deux clusters de positions (plus précisément, nous sélectionnons, l’époque qui minimise l’inertie de la sortie de K-means parmi les 100 dernières étapes).

2. <https://github.com/anongusd/GUSD>

Nous avons utilisé un ensemble d'hyperparamètres standard pour l'architecture GNN, qui est commun à tous les ensembles de données, de sorte qu'il n'y a pas de réglage des hyperparamètres (nombre de couches, taux d'apprentissage, etc.). Deuxièmement, nous avons utilisé un arrêt précoce et un ensemble de tests composé de paires de nœuds, liés ou non, où les arêtes ont été cachées pendant l'étape d'apprentissage, comme cela a été fait dans la littérature. Concrètement, en considérant l'apprentissage auto-supervisé du GNN, nous avons utilisé 5 % des liens (qui ont été cachés dans l'ensemble d'apprentissage) comme ensemble d'évaluation pour calculer le critère de convergence : lorsque la précision de la prédiction des liens cesse d'augmenter sur cet ensemble, nous arrêtons la phase d'optimisation.

3.2 Resultats

Le tableau 2 détaille le taux de bien classés en classification et le score F1 pondéré de détection de position pour chaque modèle sur chaque ensemble de données, moyennés sur 10 exécutions. En outre, la moyenne globale et l'écart-type de ces mesures, calculés sur les neuf ensembles de données, sont indiqués ensuite.

Comme le montrent les résultats, les méthodes de l'état de l'art sont meilleurs dans des contextes spécifiques, mais en dehors de ces contextes, leurs performances sont souvent comparables à celles du hasard. InfoVGAE a du mal avec les réseaux hétérophiles, car il utilise des convolutions de graphe sur le graphe bipartite et agrège donc jusqu'aux voisins à deux sauts, c'est-à-dire les utilisateurs qui ont partagé le même contenu. Il obtient de meilleurs résultats sur les réseaux homophiles avec un pourcentage plus élevé de retweets en raison de son utilisation d'un graphe bipartite et de convolutions de graphe. En effet, un graphe contenant de nombreux textes uniques (c'est-à-dire des textes postés par un seul utilisateur sans retweets, comme dans le cas de CreateDebate) produira un graphe moins connecté et réduira l'efficacité du modèle. STEM, quant à lui, a été conçu pour les réseaux d'interaction hétérophiles, comme le confirme sa faible performance sur les jeux de données issus de ce type de média social. La méthode de Darwish et al. fonctionne mieux sur les données Twitter homophiles qui permettent de construire des vecteurs de fréquence plus informatifs, ce qui signifie que la moyenne des tweets, des hashtags ou des mentions par utilisateur est relativement élevée, comme dans les ensembles de données ConRef et TIMME. Cela permet aux vecteurs UMAP de faible dimension d'être plus efficaces dans la classification des utilisateurs.

Bien que GUSD ne surpasse pas tous les modèles sur tous les jeux de données, ce qui n'était pas son objectif, il surpasse ces modèles en moyenne sur tous ces jeux. Parce qu'il peut être exécuté sur des réseaux d'interactions homophiles et hétérophiles, il est applicable pour différents types de réseaux issus de médias sociaux variés, alors que ce n'est pas le cas des autres méthodes du fait de leurs hypothèses sous-jacentes. En outre, contrairement aux autres modèles, GUSD fonctionne sans problème sur tous les jeux de données et est moins sensible au bruit supplémentaire que contient le jeu TIMME-All par rapport à TIMME-Pure, qui n'est filtré que sur les données relatives aux hommes politiques.

4 Conclusion

Bien qu'il existe des modèles non supervisés pour la détection des positions, ils peinent à se généraliser aux types de réseaux qui ne respectent pas les hypothèses sur lesquelles les méthodes

sont basées. Ainsi, aucun d’entre eux ne peut servir de référence pour des jeux de données variés présentant des caractéristiques opposées d’homophilie ou d’hétérophilie. Pour palier cette limite, nous avons conduit une analyse comparative des modèles récents sur différents jeux de données et proposé un nouveau modèle générique. Contrairement aux travaux existants, il exploite à la fois le contenu du texte et la structure du graphe en utilisant des plongements de texte propagés via des réseaux neuronaux de graphes, ce qui le rend plus généralisable à ces différents types de réseaux. GUSD est en moyenne plus performant que les autres modèles non supervisés et résiste aux changements dans le nombre de paramètres du modèle de langage utilisé pour construire les plongements de texte.

Remerciements

Ce travail a été partiellement soutenu par la MANUTECH-Sleight Graduate School grâce à une subvention publique de Saint-Etienne Métropole, par le projet CODANA du CNRS IEA, et par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) sous la subvention n° RGPIN-2022-04789.

Références

- Darwish, K., P. Stefanov, M. Aupetit, et P. Nakov (2020). Unsupervised User Stance Detection on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media 14*, 141–152.
- Devlin, J., M.-W. Chang, K. Lee, et K. Toutanova (2019). BERT : Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186. Association for Computational Linguistics, doi: 10.18653/v1/N19-1423.
- Ghosh, S., K. Anand, S. Rajanala, A. B. Reddy, et M. Singh (2018). Unsupervised stance classification in online debates. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pp. 30–36. ACM.
- Hamilton, W., Z. Ying, et J. Leskovec (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems 30*.
- Hasan, K. S. et V. Ng (2014). Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 751–762.
- Khanam, K. Z., G. Srivastava, et V. Mago (2023). The homophily principle in social network analysis : A survey. *Multimedia Tools and Applications 82*(6), 8811–8854.
- Lai, M., V. Patti, G. Ruffo, et P. Rosso (2018). Stance evolution and twitter interactions in an italian political debate. In *International Conference on Applications of Natural Language to Data Bases*.
- Largerone, C., A. Mardale, et M.-A. Rizoiu (2021). Linking the dynamics of user stance to the structure of online discussions. In P. H. Abreu, P. P. Rodrigues, A. Fernández, et J. Gama

- (Eds.), *Advances in Intelligent Data Analysis XIX*, pp. 275–286. Springer International Publishing.
- Li, J., H. Shao, D. Sun, R. Wang, Y. Yan, J. Li, S. Liu, H. Tong, et T. Abdelzaher (2022). Unsupervised Belief Representation Learning with Information-Theoretic Variational Graph Auto-Encoders. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1728–1738.
- McPherson, M., L. Smith-Lovin, et J. M. Cook (2001). Birds of a feather : Homophily in social networks. *Review of Sociology* 27, 415–444.
- Oh, S. J., A. C. Gallagher, K. P. Murphy, F. Schroff, J. Pan, et J. Roth (2019). Modeling uncertainty with hedged instance embeddings. In *International Conference on Learning Representations*.
- Park, J. et A.-L. Barabási (2007). Distribution of node characteristics in complex networks. *Proceedings of the National Academy of Sciences* 104(46), 17916–17920.
- Pick, R. K., V. Kozhukhov, D. Vilenchik, et O. Tsur (2022). STEM : Unsupervised Structural EMbedding for Stance Detection. *Proceedings of the AAAI Conference on Artificial Intelligence* 36(10), 11174–11182.
- Sutter, M., A. Gourru, A. Trabelsi, et C. Largeron (2024). Unsupervised stance detection for social media discussions : A generic baseline. In Y. Graham et M. Purver (Eds.), *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2024*, pp. 1782–1792.
- Trabelsi, A. et O. Zaiane (2018). Unsupervised model for topic viewpoint discovery in online debates leveraging author interactions. *Proceedings of the International AAAI Conference on Web and Social Media* 12(1), doi: 10.1609/icwsm.v12i1.15021.
- Veličković, P., G. Cucurull, A. Casanova, A. Romero, P. Liò, et Y. Bengio (2018). Graph attention networks. In *International Conference on Learning Representations*.
- Xiao, Z., W. Song, H. Xu, Z. Ren, et Y. Sun (2020). TIMME : Twitter Ideology-detection via Multi-task Multi-relational Embedding. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2258–2268. ACM.

Summary

With the ever-growing use of social média to express opinions on the national and international stage, unsupervised methods of stance detection are increasingly important to handle the task without costly annotation of data. The current unsupervised state-of-the-art models are designed for specific network types, either homophilic or heterophilic, and they fail to generalize to both. In this paper, we first analyze the generalization ability of recent baselines to these two very different network types. Then, we conduct extensive experiments with a baseline model based on text embeddings propagated with a graph neural network that generalizes well to heterophilic and homophilic networks. We show that it outperforms, on average, other state-of-the-art methods across the two network types. Additionally, we show that combining textual and network information outperforms using text only, and that the language model size has only a limited impact on the model performance.