

Structural Characterization and Graph-based Detection of Fake News in Portuguese

Roney Lira de Sales Santos¹, Thiago Alexandre Salgueiro Pardo¹

¹Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematical and Computer Sciences, University of São Paulo
São Carlos, Brazil

roneysantos@usp.br, taspardo@icmc.usp.br

Abstract. *The production of fake news is a problem nowadays. With social networks, fake news spread in easier and less costly ways, having the power to reach a large number of people in a short time. In this paper, we investigate graph-based approaches for fake news characterization and detection, taking into account widely used measures of graphs and complex networks. Our results show that some network measures are useful for structurally characterizing fake and true news and that machine learning-based solutions over this kind of feature produce promising results.*

Resumo. *A produção de notícias falsas é um problema dos dias atuais. Com as redes sociais, as notícias falsas se espalham de forma mais fácil e barata, podendo chegar a um grande número de pessoas em um curto espaço de tempo. Neste artigo, investigamos abordagens baseadas em grafos para caracterização e detecção de notícias falsas, levando em consideração medidas amplamente utilizadas de grafos e redes complexas. Nossos resultados mostram que algumas medidas de rede são úteis para caracterizar estruturalmente notícias falsas e verdadeiras e que soluções baseadas em aprendizado de máquina sobre esse tipo de atributo produzem resultados promissores.*

1. Introduction

Nowadays, fake news detection has become an important research topic, and the need to always evaluate the veracity of digital content has been raised by the constant spread of fake news [Figueira and Oliveira 2017]. According to [Rubin et al. 2015], in the Natural Language Processing (NLP) area, two main approaches may be used for automatically identifying fake news: linguistic and network approaches.

In linguistic approaches, there is an attempt to find linguistic clues that can differentiate fake news from true news, and these clues generally work well when analyzing text features such as lexical, morpho-syntactic, syntactic, semantic, psycho-linguistic [Pérez-Rosas and Mihalcea 2015, Pérez-Rosas et al. 2017] and readability measures [Santos et al. 2020]. In the case of network approaches, there are fact-checking models, which aim at relating the elements of an statement (e.g., by generic and domain-specific relations as “is a”, “member of” and “is married to”) to subsidize content verification [Ciampaglia et al. 2015]. In a related research line, graph-based approaches to detect fake news are present in the literature with considerable results, using knowledge

graphs [Ciampaglia et al. 2015, Pan et al. 2018], biclique identification, graph-based feature vector learning and label spreading [Gangireddy et al. 2020] and graph-based modeling of online communities [Chandra et al. 2020], among others. This type of approach is important to detect false information, as it allows examining beyond linguistic characteristics of the writing and determining the structural behavior of fake news.

This paper has as contribution the investigation of two graph-based approaches to deal with fake news for the Portuguese language: the SentiElection, an ensemble of global centrality measures over reference graphs, proposed by [Vilarinho and Ruiz 2018] for the task of sentiment analysis; and a complex network-based approach, in which the news are converted into complex networks and several of the most used network measures in the literature are applied, to then classify the news as fake or true. To the best of our knowledge, this is the first time that such methods are explored for fake news detection.

Trying to uncover the structural characteristics of fake and true news in Portuguese, we show that some complex network measures are interesting to distinguish them, as clustering coefficient, eigenvector, katz and pagerank values. In machine learning-based detection approaches, the network measures are used as features and some of them produce promising results, as katz and density, with K-Means and SVM techniques producing the best results. The adaptation of SentiElection to the task in hands shows that the accuracy of fake news detection increases as the reference graphs grow, although its results do not overcome the ones produced by the complex network measures.

The remainder of this paper is organized as follows. Section 2 briefly introduces the main related work in the area. Section 3 details the graph-based approaches that we explore in this paper. Experiments and results are presented in Section 4 and some final remarks are made in Section 5.

2. Related Work

Graph-based (or network-based) approaches generally address and try to solve the task of automatic fact-checking. [Thorne and Vlachos 2018] formally describe fact-checking as the task of evaluating whether the statements made, in both written and spoken language, are true, which is usually done by trained professionals. For automatic checking, the authors suggest that the input to fact-checking approaches should be triples of the form (subject, predicate, object), which facilitate fact-checking in structured (and semi-structured) databases. Such a triple is one of the main patterns we have today for the construction of graphs, where the subject and object are nodes of the graph and the predicate is the relationship (edge) between the nodes.

Approaches using knowledge graphs, with a triple being the main element, can provide a rich set of structured information related to world knowledge stored in a machine-readable format that supports the task of checking facts, being widely used in the literature [Ciampaglia et al. 2015, Shi and Weninger 2016, Pan et al. 2018], including the case of Portuguese language [Santos and Pardo 2020], which had 74% average accuracy when simple statements were checked (i.e., *Brasília is the capital of Brazil*).

Complex networks are often used in detecting the spread of fake news [Lind et al. 2007, Alassad et al. 2019]. In [Paluch et al. 2018], the authors tried to find a method that executes the task of identification of the propagation source in reasonable

time on large complex networks and delivers high quality localization results at the same time. [Zhou and Zafarani 2019] work with a pattern called Denser-Network to identify fake news spreaders, since the authors hypothesize that fake and true news spreaders show different graph properties, reaching 90% accuracy in this approach. However, to the best of our knowledge, there are no works that use measures extracted from a complex network to directly characterize and classify news text as fake or true.

In the next section we introduce the techniques that we explore in this paper.

3. Graph and Complex Network-based Approaches

3.1. SentiElection Approach (SEA)

The SentiElection approach is the method proposed by [Vilarinho and Ruiz 2018], which is a voting system over three measures (katz, eigenvector and pagerank) computed for a text in relation to reference graphs. A text of interest is classified according to the reference graph that it best fits (as indicated by the taken measures in the voting strategy).

In this work, SEA behaves as a classifier in detecting fake news. The approach is made with two graphs formed by news words: one graph with fake news (G-FAKE) and other with true news (G-TRUE). The graphs are formed with a variation of the number of edges that each word has, called the word frame¹.

True and fake news from the Fake.Br Corpus² [Monteiro et al. 2018, Silva et al. 2020] are used as input to the classifier. Of the 3,600 fake news items present in the corpus, a certain number of news items were randomly selected to form the fake news graph. The same process was used for the construction of the true news graph. Details of the experiments are shown in Section 4.1.

3.2. Complex Network Measures-based Classifier (CNMC)

The Complex Network Measures-based Classifier is a classifier that works with measures widely used in Complex Networks approaches, namely: betweenness, closeness, eigenvector, katz, pagerank, hubs and authorities, cluster coefficient (and its average), correlation, transitivity and density. The measures are briefly explained³ in Table 1.

The mentioned measures were applied in all news from Fake.Br Corpus. Firstly, the graph of each news item was generated, following the same procedures that were used to generate the graphs of the SEA, but with one change: the word frame is 1, i.e., each word is linked only to its adjacent words.

With the generated graphs, the second part of the process was to compute the complex network measures. Such measures are extracted from the implementations in the NetworkX framework [Hagberg et al. 2008], using the default implementations of each measure, without additional configurations, except for the centrality measures katz, pagerank and hubs and authorities, where a setting of maximum interaction of the algorithm is

¹The word frame represents the distance between the analyzed word and the words in sequence. If the distance is 1, the edge in the graph would be from word n to $n+1$; if the distance is 2, the edge in the graph would be from word n to $n+1$ and to $n+2$, where n is the position of the word in the sentence.

²Details about the dataset, such as how and when the data was collected, the sources and news annotation, can be found in [Monteiro et al. 2018] and [Silva et al. 2020].

³According to NetworkX documentation available at <https://networkx.org/>. Further details about the measures can be found in [Morais and Prati 2013] and [Comin et al. 2020].

Type	Metric	Explanations
Centrality	Betweenness	Quantifies the participation of a node v in paths of minimum length
	Closeness	Measures the proximity of node v to all nodes in the network
	Eingenvector	Brings out the centrality of a node v based on the centrality of their neighbours
	Katz centrality	Computes the centrality for a node based on the centrality of its neighbors, on local and global influence
	Pagerank centrality	Computes a ranking of the nodes in the graph G based on the structure of the incoming links
	Hubs	Estimates the node value based on outgoing links
	Authorities	Estimates the node value based on incoming links
Clustering	Cluster coefficient and Cluster coefficient average	Quantifies the fraction of possible triangles through that node that exist. A triangle is a set of three nodes, where each node has a relationship to all other nodes
	Correlation	Computes the similarity of connections in the graph with respect to the node degree
	Transitivity	Computes the fraction of all possible number of "triads" (two edges with a shared vertex)
Other	Density	Measures of how is a graph in which the number of edges is close or far to the max number of edges

Table 1. Complex Network measures

determined (1,000 epochs). This configuration was necessary because interactions below this value did not result in reliable results.

Besides attempting to structurally characterizing the news, the measures served as input features to various machine learning techniques, both symbolic and statistical. The analysis were divided into verifying the results with the combined features and with individual features. Details of the cited experiments are reported in Section 4.2.

4. Experiments and Results

To test the approaches reported in this paper, some experiments were performed. In all the experiments, the truncated versions of the fake and true news in the Fake.Br Corpus were used (in order to guarantee similar text size and more reliable experiment results).

4.1. The SEA

For the analyses of the SEA experiments, graphs of different sizes were built to observe the accuracy of the classifier when news are analyzed by the amount of information present. Graphs were formed with 100, 200, 300, 400, 500, 600, 700 and 800 news, both fake and true, totaling 16 graphs with all news being randomly selected for each set. The test was done with 1,000 news, also randomly selected, being 500 fake news and 500 true news. It was necessary to keep such balance for more faithful results.

The pipeline used in the experiment had three steps: 1) the extraction of news tokens; 2) the construction of the graph⁴; and 3) the analysis of the SEA classification results. From the selected news set, each news item is tokenized with the NLTK tokenizer [Bird et al. 2009] and is tagged by NLPNet [Fonseca et al. 2015]. If the token is an adverb, adjective, noun (common or proper) or a verb, the token is selected to form the graph; if the token is a stopword, it is discarded. Using only content words ensures more informative links between words. The words are then linked to form the graph, using word frame = 3, since [Vilarinho and Ruiz 2018] demonstrated that this was the best configuration in their experiments. The graphs were formed using the NetworkX framework. Finally, for the test, each graph formed by the union of the *reference graph* + *news graph* is submitted to the centrality calculation of the eigenvector, katz and pagerank measures, which work as a voting system, indicating the category of the news (true or fake) according to the graph that produces the best results.

Analyzing the 1,000 news in the 16 G-FAKE and G-TRUE graphs, we synthesize the results in Figure 1. The evaluation measures were the geometric mean (as in [Vilarinho and Ruiz 2018] - blue line) and accuracy (orange line).

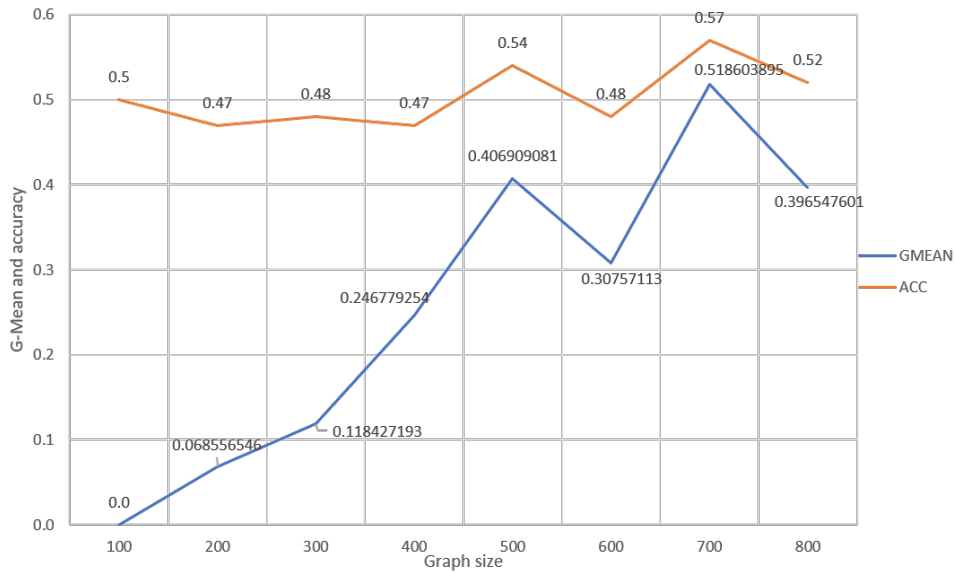


Figure 1. Results of the SEA experiment

The performance of the algorithm increases proportionally to the size of the test graph, with an outlier for the graph of size 600, understandable by the randomness of the news selection process. The best performance was when the graph reached size 700, with a drop in performance when it reached size 800.

4.2. The CNMC

Unlike the SEA, the news graphs for the CNMC experiment were unique, i.e., a single graph for each news was formed for the analysis. The selected word frame was 1, with each word being linked with its immediate neighbors.

⁴Details about graph construction can be found in [Vilarinho and Ruiz 2018].

Experiments were carried out with graphs with and without stopwords⁵. It is noteworthy that the stopwords were extracted from the NLTK, similar to what we use in the SEA, however, here we use the full return of NLTK stopwords, without distinguishing part-of-speech classes. For each of the types of graphs (fake or true), the 12 measures of complex networks cited before were extracted by the NetworkX framework.

It was possible to make a first comparative analysis between the complex network measures from fake and true news. Table 2 shows the comparison between the individual metrics in each of the news types and graph composition (with or without stopwords). The numbers represented in Table 2 take into account when **metric(true) > metric(fake)** happened and the corresponding percentage. It is important to mention that 3,600 comparisons were made between news, since Fake.Br Corpus is an aligned dataset, with each fake news having a corresponding true one.

Measure	with stopwords		without stopwords	
	metric(true)>metric(fake)	%	metric(true)>metric(fake)	%
Betweenness	1,756	0.48	1,592	0.44
Closeness	1,664	0.46	1,446	0.40
Eigenvector	1,645	0.45	1,379	0.38
Katz	1,660	0.46	1,400	0.38
Page Rank	1,705	0.47	1,362	0.37
Hubs	1,836	0.51	1,882	0.52
Authorities	1,751	0.48	1,685	0.46
Cluster Coeff Avg	1,758	0.48	1,444	0.40
Cluster Coeff	585	0.16	837	0.23
Correlation	1,630	0.45	1,957	0.54
Transitivity	1,549	0.43	1,443	0.40
Density	1,828	0,51	1,673	0.46

Table 2. Comparison based in absolute values of the metrics

It is possible to see that the biggest difference occurred in the cluster coefficient, where only 16% of the true news with stopwords and 23% of the true news without stopwords had this measure with a value greater than the fake news. The other measures reached very close values, being the difference a little more relevant in the centrality measures eigenvector, katz and pagerank when the graphs did not contain stopwords.

The next experiment was to use the measures extracted from the graphs of each news to create classifiers. Nine Machine Learning (ML) techniques were selected: Decision Tree (DT), Naive Bayes (NB), Random Forest (RF), K-Nearest Neighbors (KNN), Support Vectors Machine (SVM), Multi-Layer Perceptron (MLP), OneRule (OneR), One Class (OC) and K-Means (KM). The implementations followed the pattern of Scikit-Learn framework [Pedregosa et al. 2011], except for the OneRule classifier, which implementations of the MLxtend library [Raschka 2018] were used. Standard implementations of Scikit-Learn and MLxtend were used in all approaches except for: MLP, where training

⁵Stopwords may be a sensible issue in complex network approaches, as they may significantly change the network topology and the corresponding measures.

was limited to 1,000 training epochs; OneClass, where training used the kernel coefficient = auto, i.e., using $1/n$ features; and K-Means, where the number of clusters was 2.

Two sets of experiments were performed: (i) classifiers with individual features and (ii) classifiers with combined features. In (i), each classifier received as input the individual values of each metric of each news item. For example, an SVM model was created that received as input the values of the betweenness metric for each news item. The model would then draw a cutting line where values above or below it would be fake or true. This experiment was interesting to verify if some metric by itself could be discriminating in news classification, based in what we found in the Table 2. In (ii), all news features were combined as input to one of the machine learning models.

Table 3 contains the experiments with the best results for each measure in (i). Each classifier received as input the individual measures of the news with and without stopwords. Cross-validation with $k = 5$ was the evaluation method and the accuracy results show the average of the executions of the ML approaches (10 executions).

Measure	with stopwords		without stopwords	
	model(s)	acc	model(s)	acc
Betweenness	DT, RF, SVM, MLP	0.51	NB, SVM, MLP	0.54
Closeness	NB, SVM, MLP, KM	0.52	NB, MLP	0.56
Eigenvector	NB, SVM, KM	0.53	MLP	0.56
Katz	KM	0.63	SVM, MLP	0.55
Pagerank	DT, RF, SVM	0.52	SVM, MLP	0.56
Hubs	OC	0.58	DT, RF	0.53
Authorities	KM	0.67	SVM	0.53
Cluster Coeff Avg	NB	0.51	DT, RF	0.53
Cluster Coeff	DT, KNN, MLP	0.51	DT, KNN	0.53
Correlation	NB, SVM, MLP	0.53	SVM, MLP, KM	0.53
Transitivity	KM	0.61	KM	0.53
Density	KM	0.69	DT, RF	0.52

Table 3. Results of the experiments with individual measures on the ML approaches

The best results come from graphs with stopwords. The density measure reached the best accuracy, with 69%, followed by authorities with 67% of accuracy and the katz measure with 63%. It is also interesting to note that the best results came from the K-Means approach. The graphs without stopwords had results close to random ($\approx 50\%$), with the best accuracies coming from closeness and pagerank, reaching 56%.

Experiment (ii) took into account all measures as input features to the ML approaches. The results achieved are presented in Table 4. Again, the evaluation method used was the cross-validation with $k = 5$ and the accuracy results show the average of the executions of the MLs approaches (10 executions).

SVM obtained the best accuracy in this experiment, reaching 60% with graphs without stopwords, followed by MLP with 59%, also with the same setup. K-Means, which was the ML approach that had the best results when using the individual features, reached only 54% accuracy, below what was achieved in the previous experiment.

ML approach	with stopwords	without stopwords
Decision Tree	0.51	0.55
Naive Bayes	0.54	0.58
Randon Forest	0.54	0.58
KNN	0.53	0.57
SVM	0.55	0.60
MLP	0.55	0.59
OneR	0.50	0.51
One Class	0.24	0.24
K-Means	0.48	0.54

Table 4. Results of the experiments with combined measures on the ML approaches

Overall, the achieved results are encouraging, showing that true and fake news do have distinguishing structural characteristics. The use of these characteristics may support the identification of fake content in different ways, possibly helping overcoming the limitations of some linguistic-based approaches that check the veracity of news by evaluating their writing style and linguistic choices. Bringing additional structural knowledge to the situation may allow to deal with more difficult cases, for instance, statements with half-truths

The use of complex network measures has not been much explored for the detection of fake news. This paper was a first attempt in this endeavour. The addition of new measures present in the literature may be important to increase the accuracy of the classification and remains for future work, as we comment in the following section.

5. Final Remarks

This work focused on structural approaches for characterizing and detecting fake news for Portuguese. We adapted the SentiElection method [Vilarinho and Ruiz 2018] and explored measures of complex networks, showing that fake and true news do have distinguishing structural characteristics. The classification results did not overcome the ones achieved in linguistic-based approaches (see, e.g., [Silva et al. 2020]), but we believe that they may help improving such approaches.

Future work includes performing deeper analyses of network measures, looking for other structural patterns, and tackling more challenging phenomena, as half-truth and post-truth statements. It may also be interesting to perform cross-lingual tests to check whether the learned patterns keep the same across different natural languages.

The interested reader may find more information about this work at the web portal of the POeTiSA project⁶.

Acknowledgments

The authors are grateful to CAPES, USP Research Office (PRP #668), and the Center for Artificial Intelligence (C4AI) of the University of São Paulo, sponsored by IBM and FAPESP (grant #2019/07665-4).

⁶<https://sites.google.com/icmc.usp.br/poetisa>

References

- Alassad, M., Hussain, M. N., and Agarwal, N. (2019). Finding fake news key spreaders in complex social networks by using bi-level decomposition optimization method. In Agarwal, N., Sakalauskas, L., and Weber, G.-W., editors, *Modeling and Simulation of Social-Behavioral Phenomena in Creative Societies*, pages 41–54.
- Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*. O’Reilly Media, Inc., 1st edition.
- Chandra, S., Mishra, P., Yannakoudakis, H., Nimishakavi, M., Saeidi, M., and Shutova, E. (2020). Graph-based modeling of online communities for fake news detection.
- Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., and Flammini, A. (2015). Computational fact checking from knowledge networks. *PloS one*, 10(6):e0128193.
- Comin, C. H., Peron, T., Silva, F. N., Amancio, D. R., Rodrigues, F. A., and da F. Costa, L. (2020). Complex systems: Features, similarity and connectivity. *Physics Reports*, 861:1–41.
- Figueira, A. and Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science*, 121:817–825.
- Fonseca, E. R., Rosa, J. a. L. G., and Aluísio, S. M. (2015). Evaluating word embeddings and a revised corpus for part-of-speech tagging in portuguese. *Journal of the Brazilian Computer Society*, 21(1):2.
- Gangireddy, S. C. R., P, D., Long, C., and Chakraborty, T. (2020). Unsupervised fake news detection: A graph-based approach. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, page 75–83.
- Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx. In *Proceedings of the 7th Python in Science Conference*, pages 11 – 15.
- Lind, P. G., da Silva, L. R., Andrade, J. S., and Herrmann, H. J. (2007). Spreading gossip in social networks. *Phys. Rev. E*, 76:036117.
- Monteiro, R. A., Santos, R. L. S., Pardo, T. A. S., de Almeida, T. A., Ruiz, E. E. S., and Vale, O. A. (2018). Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In *Computational Processing of the Portuguese Language*, pages 324–334.
- Morais, G. and Prati, R. C. (2013). Complex network measures for data set characterization. In *2013 Brazilian Conference on Intelligent Systems*, pages 12–18.
- Paluch, R., Lu, X., Suchecki, K., Szymański, B. K., and Hołyst, J. A. (2018). Fast and accurate detection of spread source in large complex networks. *Scientific reports*, 8(1):1–10.
- Pan, J. Z., Pavlova, S., Li, C., Li, N., Li, Y., and Liu, J. (2018). Content based fake news detection using knowledge graphs. In *The Semantic Web – ISWC 2018*, pages 669–683.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau,

- D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pérez-Rosas, V., Kleinberg, B., Lefevre, A., and Mihalcea, R. (2017). Automatic detection of fake news. *CoRR*, abs/1708.07104.
- Pérez-Rosas, V. and Mihalcea, R. (2015). Experiments in open domain deception detection. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1120–1125.
- Raschka, S. (2018). Mlxtend: Providing machine learning and data science utilities and extensions to python’s scientific computing stack. *The Journal of Open Source Software*, 3(24).
- Rubin, V. L., Conroy, N. J., and Chen, Y. (2015). Towards news verification: Deception detection methods for news discourse. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS48) Symposium on Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium*, pages 5–8.
- Santos, R., Pedro, G., Leal, S., Vale, O., Pardo, T., Bontcheva, K., and Scarton, C. (2020). Measuring the impact of readability features in fake news detection. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1404–1413.
- Santos, R. L. d. S. and Pardo, T. A. S. (2020). Fact-checking for portuguese: Knowledge graph and google search-based methods. In *Computational Processing of the Portuguese Language*, pages 195–205.
- Shi, B. and Wenginger, T. (2016). Fact checking in heterogeneous information networks. In *Proceedings of the 25th International Conference Companion on World Wide Web*, page 101–102.
- Silva, R. M., Santos, R. L., Almeida, T. A., and Pardo, T. A. (2020). Towards automatically filtering fake news in portuguese. *Expert Systems with Applications*, 146:113–199.
- Thorne, J. and Vlachos, A. (2018). Automated fact checking: Task formulations, methods and future directions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3346–3359.
- Vilarinho, G. and Ruiz, E. (2018). Global centrality measures in word graphs for twitter sentiment analysis. In *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 55–60.
- Zhou, X. and Zafarani, R. (2019). Network-based fake news detection: A pattern-driven approach. *SIGKDD Explor. Newsl.*, 21:48–60.