Fake News and Brazilian politics – temporal investigation based on semantic annotations and graph analysis

Luiz Gomes-Jr[1], **Gabriel Frizzo**[1]

[1]Universidade Tecnológica Federal do Paraná (UTFPR) – Curitiba – Brazil

gomesjr@dainf.ct.utfpr.edu.br, gfrizzo@alunos.utfpr.edu.br

Abstract. *The widespread use of misleading news articles has been threatening democratic processes such as elections and referendums. Understanding how fake news address social entities (e.g. personalities and institutions) and how they react to social events are important factors in the fight against the trend. This paper employs a new approach based on semantic annotations and graph analysis to study fake news articles about Brazilian politics in a two year time span. We demonstrate how graph analysis can be used to track topic evolution and cluster related entities. A preliminary result also indicates that fake news tend to be influenced by public interest (and not the other way around).*

1. Introduction

Fake news has become a major source of concern in modern societies. The problem derives from the combination of cheap means of content publication provided by the Web with the easy and manipulable means of content promotion provided by social media. This scenario threatens democratic processes such as elections and referendums, and is characterized by a complex interaction between economic incentives and human vulnerabilities [Allcott and Gentzkow 2017].

The serious social threats and the difficulties in tackling the problem make it one of the most urgent research topics currently. Understanding the evolution of fake news over time and their interaction with social events can provide insights on the reasoning behind the publication of unreliable content. This understanding can then provide the basis for better algorithms for classification of non-credible sources and for better legal actions against such sources.

This paper investigates the topics and entities targeted by fake news in the Brazilian press in a two years time span. We employ semantic annotations and graph analysis models to assess, for the entities mentioned in the published news: the clusters formed (which can represent the latent topics, Section 4.1), the relative importance of entities through time (Section 4.2), and the interplay between the calculated metrics and social interest in the entities (Section 4.3). We discuss preliminary results of the techniques in Section 5.

The goal of this work-in-progress paper is twofold: (i) to provide a glimpse into the topics and entities discussed in the Brazilian fake news, and (ii) to demonstrate how semantic annotations and graph analysis can be employed for temporal analysis of text (not restricted to fake news analysis).

2. Related work

The term fake news has gained importance in recent years, mainly due to the widespread use of misinformation in democratic elections and referendums. Fake news can be defined as distorted signals that do not correlate with the truth [Allcott and Gentzkow 2017]. One important aspect of the fake news phenomenon is the mechanisms behind their dispersion. Allcott and Gentzkow [Allcott and Gentzkow 2017] analyse the influence of Social Media on the dissemination of fake news in the 2016 presidential election in the United States. The authors discuss the economics of fake news, how social media is a source of traffic for fake news websites, how the news favored the candidates and how education plays a role in people's vulnerability to false content.

Given the discussed importance of social media in the dissemination of fake content, it is important to understand the dynamics of the origins and spreading of misinformation. Jang et al. [Jang et al. 2018] analysed several tweets related to fake news stories in the 2016 US election employing a network science approach. The authors created phylogenetic trees for the tweets associated with each of the fake news topics. The analysis of the trees suggested interesting trends, like the tweets being promoted majoritarily by ordinary users and the tendency of fake news tweets mentioning non-credible sources.

Therefore, it is important to be able to detect non-credible sources. Linguistic patterns can provide valuable information about the likelihood of a content being fake. Monteiro et al. [Monteiro et al. 2018] produced the first reference corpus of fake news for the Portuguese language. The authors manually added fake news from online sources and aligned each article with a reliable story on the same topic. With the aligned corpus the authors were able to determine linguistic differences between fake and true content, such as the higher number of misspellings in fake news. The authors also set up a classification task and run several experiments to identify the best features to classify news as fake or true.

In this paper we use the corpus provided by Monteiro et al. but focus on a time analysis of the evolution of topics and entities cited in the fake news. Unlike Jang et al. [Jang et al. 2018], our focus is on the content produced by non-credible sources and not on its spread. We aim at understanding the interplay between fake news and the society in terms of political events and popular interest.

3. Data collection and cleaning

The corpus used in the analysis is the Fake.Br Corpus [Monteiro et al. 2018]. Only news classified as fake by the authors are used here. Each news article was annotated with DBPedia[1] entities using the SpotLight tool [Mendes et al. 2011]. For each article, a graph was formed with all entities mentioned as vertices. Edges in the graph connect all entities (complete graph), forming the article's co-citation graph. The graphs for the individual articles were used to build two distinct datasets: (i) a global graph for the entire time span of the corpus, and (ii) a series of graphs representing time partitions in the corpus.

For the global graph, the individual co-citation graphs were merged, with the number of occurrences of a given edge being used as its weight. Edges that occurred less than 20 times were omitted from the analysis. This dataset was used to identify clusters of

[1] https://dbpedia.org/

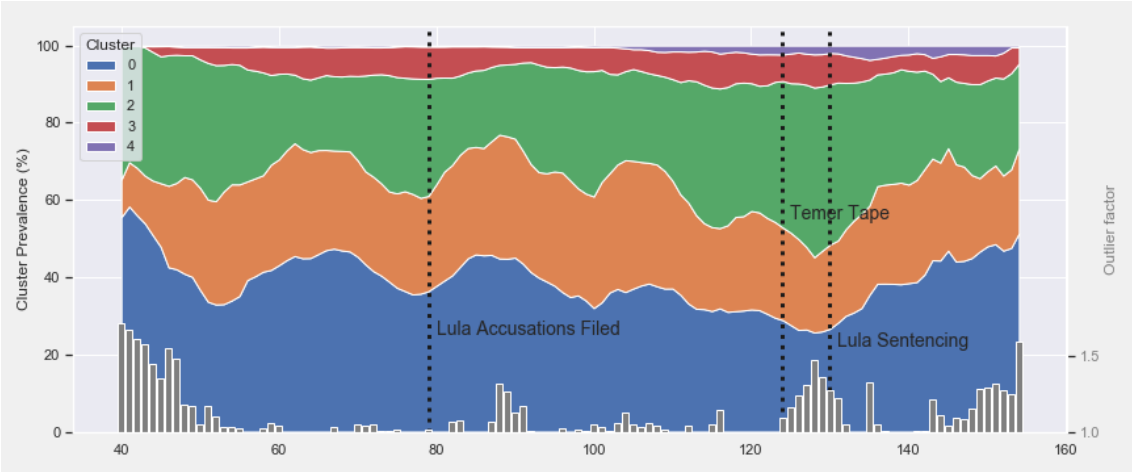

Figure 1. Prevalence of each cluster, outlier factors and events for reference

entities (Section 4.1). As for the time series of graphs, we defined time windows of 14 days with an overlap of 7 days (i.e. one 14 days window starts every week). Each window is considered a partition, which are numbered from 0 (starting on 2015-01-10) to 198 (ending on 2018-11-04). Only partitions between 38 and 155 are kept because the others have many gaps and much fewer collected articles. For each partition we merged the graphs of the respective articles and calculated several centrality and clustering metrics. This dataset was used to assess the evolution of graph measurements and other statistics over time (Section 4.2).

4. Data analysis

4.1. Cluster analysis

To define clusters of entities correlated in the graph, we applied the modularity algorithm (see [da F. Costa et al. 2007]) over the complete graph (all partitions merged – 154 nodes, 665 edges). The algorithm detected 5 distinct clusters, as shown below. The latent topic in each cluster seems to be (0): PT (workers party) politicians, (1): Lava Jato (car wash) investigations, (2): President Rousseff's Impeachment, (3): (residual cluster) Supreme Court, (4): (residual cluster) Bolsonaro and others.

Figure 1 shows the evolution of the prevalence of clusters through time. Each color represents a cluster with its width proportional to the percentage of entities of that cluster appearing in news in a given partition window. The time series for each clusters was smoothed to reduce noise and emphasize the trends. It can be seen, for example, how Lula's cluster (cluster 0, drawn in blue) becomes more prevalent after accusations were filed against him and also after he was considered guilty in the process.

To automatically detect important events in the period covered by the news, we employed outlier detection techniques. The algorithms are applied to single out partitions whose metrics deviate significantly from what would be expected.

The initial technique used is the Local Outlier Factor (LOF). It is applied to the data measuring the prevalence of each cluster (modularity class) in each partition. LOF is a multidimensional technique which in this case identifies isolated partitions in the 5-

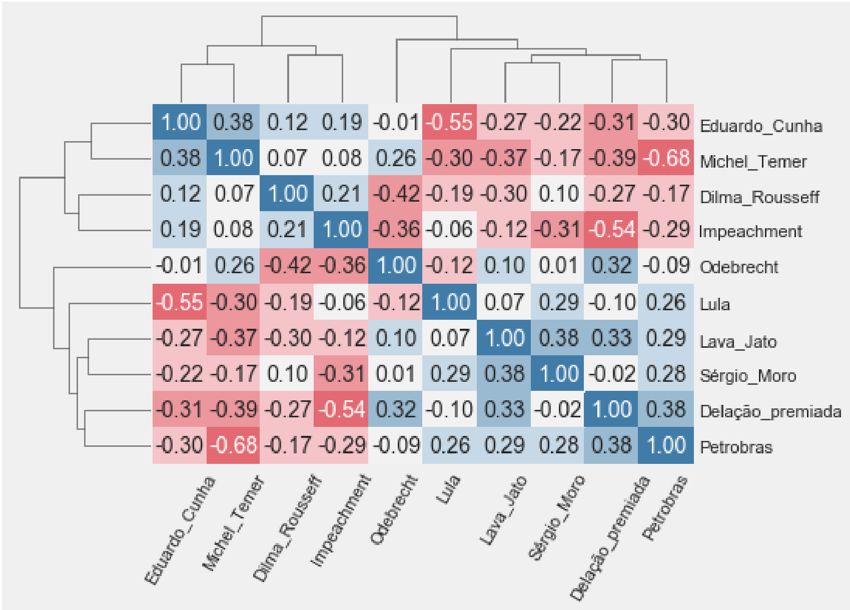

Figure 2. Correlations between top entities

dimensional space of cluster metrics. Intuitively, the algorithm will identify important shifts in discourse in the context of the news.

Figure 1 shows the Local Outlier Factor (bars in grey) representing how unexpected a given partition and its mix of clusters are. In the Figure, we overlaid three major political events in the period. There seems to be a correlation between the events and the distribution of the clusters, but this investigation is part of our ongoing efforts.

4.2. Time analysis of entities

The main focus of this paper is on understanding how the citation of individual entities vary with time. To assess the patterns, we need measurements of the importance or relevance of each entity in each time window. For each time window, we constructed its graph and calculated centrality measurements for the vertices (i.e. entities). The metrics used were *eigenvector centrality*, *closeness centrality*, and *betweenness centrality* (see [da F. Costa et al. 2007] for definitions).

The centrality measurements allow us to analyze the correlation of entities for the entire time period (all time windows). We can also cluster the entities based on spatial similarities of their vectors. Figure 2 shows the correlation matrix for top entities end their eigenvector centrality with dendograms displaying possible groupings. In the figure, it can be seen clusters of positive and negative correlations grouping entities related to the impeachment process and also with the car wash investigation. The eigenvector centrality metric demonstrated a clearly better separation among all metrics, but further analysis is needed to substantiate its superiority in this type of task.

4.3. Fake News vs. popular interest

One central aspect in understanding the mechanisms behind fake news is determining causal relationships between fake news cycles and popular interest. The basic question

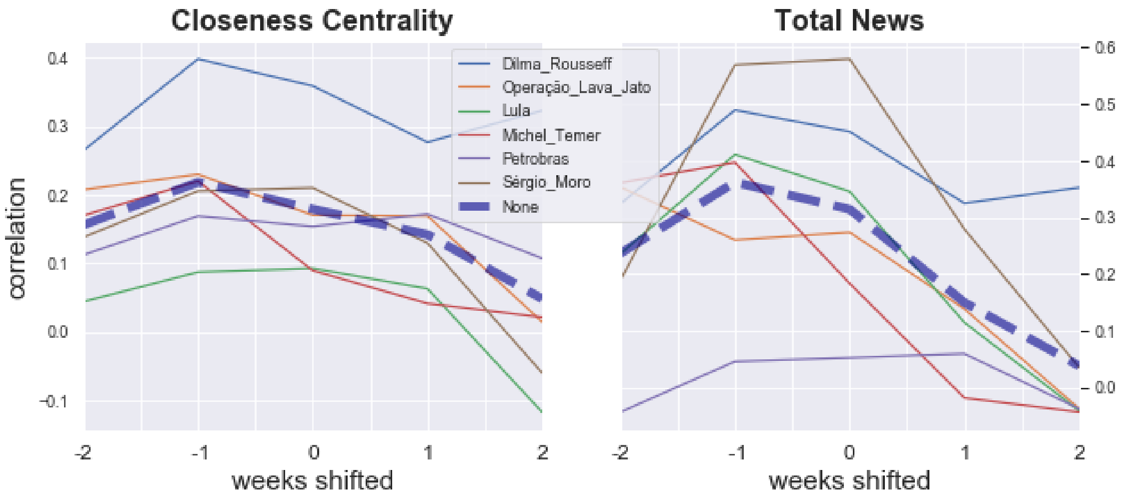

Figure 3. Correlation between metrics and Google Trends

here is whether fake news influence popular demand of topics or whether it feeds on the demand to promote their hidden agendas.

To make an initial assessment of this relationship, we first selected five top entities in our dataset, focusing on persons and entities that were part of the popular discussion in the period covered by the fake news corpus. The entities chosen were *Dilma Rousseff* (Brazil's impeached president), *Operação Lava Jato* (the codenamed 'car wash' investigation), Lula (ex-president and central theme in the investigations), Michel Temer (Rousseff's vice-president), Petrobras (Brazilian state-run oil company), Sérgio Moro (judge behind the car wash prosecutions).

For each of the chosen entities, we obtained their Google Trends[2] counts (representing search demands) for the weeks covered in the fake news corpus. We then analyzed the correlations between the Google Trends and the time partition measurements in our dataset. The correlation between total number of citation in fake news and Google Trends count reaches 0.58 for Sérgio Moro, which we consider high, since we believe that only the more polemical news would generate a web search response. The best centrality metric in terms of correlation was closeness centrality, reaching 0.36 for Dilma Rousseff.

The most interesting patterns appear when we time-shift the Google Trends counts. The idea is to correlate the entity measurements with their popular interest weeks before or ahead the time window of the fake news. Figure 3 (left) shows the variation of correlation for closeness centrality compared against the Google Trends for two weeks before, one week before, the same week, one week after, and two weeks after. The correlations tend to peak for the Google Trends of the week before (-1 in the x axis), indicating that the fake news tend to follow popular interest – at least in general terms. The dashed line shows the aggregated mean of the correlations, highlighting the aforementioned trend.

Figure 3 (right) shows the trends considering total mentions of entities in the fake news articles. The correlation is higher than that for closeness centrality, albeit more irregular (the Petrobras entity now has almost no correlation). Although more tests are

[2]https://trends.google.com/trends/

necessary, this result indicates that the graph metrics like closeness centrality might be more reliable indicators of entity relevance.

5. Discussion and Conclusion

This paper presented a graph-based approach to analyse the contents and evolution of fake news about Brazilian politics. The preliminary results presented here suggest that graph analysis can capture several aspects of the discourse space.

By using graph-based clustering we detected the main latent topics in the corpus. The goal was to match the evolution through time of the topic mix with real world events. As presented here, we are now starting to apply outlier detection algorithms to identify important periods. We are currently testing other outlier algorithms and developing a better methodology to match the outliers with real-world events. We expect that in the future this analysis will provide insights on the interplay between fake news and real events.

The analysis of the graph measurements for the entities over time also showed promising results. Using traditional clustering algorithms over the entity vectors containing the measured values, we were able to find clusters of related entities that correspond to real world political developments. We plan to also apply outlier detection over individual entities in the future.

Finally, in the most important preliminary result in this paper, we assessed how the graph measurements for some top entities correlate with popular interest (represented by Google Trends counts). The analysis shows a clear tendency of fake news on a given topic to vary based on recent popular interest. This suggest that fake news, at this moment, may be more of a reactive mechanism than a driver of the discourse. Our hypothesis is that real news would not show this marked trend, correlating more closely with popular interest, but we still do not have an adequate corpus to verify. If the hypothesis is verified, this information could be used to automatically identify unreliable sources of information.

References

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36.

da F. Costa, L., Rodrigues, F. A., Travieso, G., and Boas, P. R. V. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242.

Jang, S. M., Geng, T., Li, J.-Y. Q., Xia, R., Huang, C.-T., Kim, H., and Tang, J. (2018). A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis. *Computers in Human Behavior*, 84:103–113.

Mendes, P. N., Jakob, M., García-Silva, A., and Bizer, C. (2011). Dbpedia spotlight: Shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems*. ACM.

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. Springer International Publishing.