Knowledge Graph Representation for Political Information Sources

Tinatin Osmonova, Alexey Tikhonov, Ivan P. Yamshchikov
OSCE Academy, Bishkek
Inworld.AI, Berlin
CAIRO, Technical University of Applied Sciences Würzburg-Schweinfurt (THWS), Würzburg
ivan.yamshchikov@thws.de

Abstract

With the rise of computational social science, many scholars utilize data analysis and natural language processing tools to analyze social media, news articles, and other accessible data sources for examining political and social discourse. Particularly, the study of the emergence of echo-chambers due to the dissemination of specific information has become a topic of interest in mixed methods research areas. In this paper, we analyze data collected from two news portals, Breitbart News (BN) and New York Times (NYT) to prove the hypothesis that the formation of echo-chambers can be partially explained on the level of an individual information consumption rather than a collective topology of individuals' social networks. Our research findings are presented through knowledge graphs, utilizing a dataset spanning 11.5 years gathered from BN and NYT media portals. We demonstrate that the application of knowledge representation techniques to the aforementioned news streams highlights, contrary to common assumptions, shows relative "internal" neutrality of both sources and polarizing attitude towards a small fraction of entities. Additionally, we argue that such characteristics in information sources lead to fundamental disparities in audience worldviews, potentially acting as a catalyst for the formation of echo-chambers.

Keywords: echo chambers, computational social science, knowledge representation

1. Introduction

A knowledge graph, also known as a semantic network, was initially introduced by C. Hoede and F.N. Stokman as a tool for representing the content of medical and sociological texts (Nurdiati and Hoede, 2008). Constructing increasingly larger graphs with the intent of accumulating knowledge was initially deemed to provide a resultant structure capable of operating as an expert system proficient in investigating causes and computing the consequences of certain decisions.

The concept of knowledge graph co-evolved with the rise of computational social science (Conte et al., 2012) and digital data analysis methods (Rogers, 2013). Access to open sources on the Internet has facilitated the measurement of the dynamics of political debates (Neuman et al., 2014). Platforms like Twitter and other microblogging services are widely utilized for studying and modeling social and political discourse (Graham et al., 2016), (Jungherr, 2014) , (Wang et al., 2018). Contemporary researchers even develop a conceptual framework for predicting the morality underlying political tweets(Johnson and Goldwasser, 2018). Moreover, knowledge graphs of fact-checked claims, such as ClaimsKG, have been designed. Such tools facilitate structured queries about truth values, authors, dates, journalistic reviews, and various types of metadata (Tchechmedjiev et al., 2019).

A significant group of studies, advocate usage of graphs for social, political, and business industry data, stating that "graphs greatly increases the clarity of presentation and makes it easier for a reader to understand the data being used"(Kastellec and Leoni, 2007) . Additionally, (Abu-Salih and Beheshti, 2021) explains that knowledge graphs serve as indispensable frameworks that underpin intelligent systems. This is achieved by extracting subtle semantic nuances from textual data sourced from a range of vocabularies and semantic repositories. In the past decade, there has been a notable increase in the examination of political discourse within social content in such a way. The authors discuss in detail the connection between political discussions and the language used in them (Chilton, 2004), (Parker, 2014). Furthermore, the literature examines opinion polarization (Banisch and Olbrich, 2019), attempts to characterize an intuition of the dynamics of the political debate (Yamshchikov and Rezagholi, 2019), and provides techniques for estimating them (Merz et al., 2016), (Subramanian et al., 2017), (Glavaš et al., 2017), (Subramanian et al., 2018) or (Rasov et al., 2020). The extensively employed data sources in studies centered on automated text classification for political discourse analysis involve Manifesto Database (Lehmann et al., 2017) and the proceedings of the European Parliament (Koehn, 2005).

The challenges arising in contemporary studies on observational and discourse analysis are the quality of data (Tweedie et al., 1994) and the credibility of data sources. It is crucial to apply statistical measures and tests to quantify the impact of poor data quality and bias on the results (Abu-Salih and Beheshti, 2021). However, quantifying such effects proves comprehensive in the realm of social sciences due to the numerous indigent properties of

Proceedings of the Second Workshop on Natural Language Processing for Political Sciences (PoliticalNLP 2024), pages 45–54
21 May, 2024.

social datasets (Shah et al., 2015). One significant challenge is associated with the formation of so-called echo-chambers in social structures, which naturally obstruct the propagation of information, reinforcing disparities across various social strata (Goldie et al., 2014), (Colleoni et al., 2014), (Guo et al., 2015) or (Harris and Harrigan, 2015). Addressing the credibility of sources, the phenomenon of fake news draws constant attention from media outlets and researchers. According to (Anderson and Auxier, 2020), 55% of online social network users believe they are accurately informed about recent political updates by the media. Consequently, misleading information and false news have the potential to shape certain beliefs and human behaviors. As a solution, several studies (Allcott and Gentzkow, 2017), (Shu et al., 2017) or (Lazer et al., 2018) analyze and propose methods to enhance the quality of information. Additionally, these studies imply the existence of a certain ground truth that could be universally accepted.

Taking existing knowledge and challenges into account, in this work, we study the issue of news representation from a data analysis perspective. We construct two datasets comprising news articles from "alt-right" and "liberal" news platforms, denoted as Breitbart News (BN) and the New York Times (NYT), spanning 11.5 consecutive years (from 2008 to Fall 2019). We demonstrate that information disparities between these news sources are fundamental regardless of the social structures that encapsulate the readers of the aforementioned outlets. Upon analyzing the findings, we assert that one has to take into consideration these disparities, since they signify fundamental differences in the foundational data that shapes the perspectives, beliefs, and, ultimately, the behavior of readers. Simply put, even if we had no social media information disparities by various news sources could contribute to echo-chamber formation.

2. Data and Methodology

We have parsed two news sites Breitbart News[1] that could be generally associated with the "alt-right" political views and the New York Times[2] associated with "liberal" political views. The choice of these two media platforms was arbitrary to a certain extent. We parsed all news presented on both platforms in the period from 2008 till the fall of 2019. Using the texts of the news as input data we built an information extraction pipeline aimed to reconstruct a form of knowledge graph out of the news texts. To do that we have used state of the art open information extraction (Stanovsky et al., 2018) and named entity recognition (Peters et al., 2017) tools of AllenNLP[3]. The outputs of both models are noisy, so in order to stabilize the resulting signal we came up with the heuristics for substring-matching. We used only ARG0 and ARG1 items of open information extractor and all entities of named entity recognition to extract the most useful objects of the articles. For every entity recognized by both methods, we created a vertex in our knowledge graph. We also applied additional manual 'filtering' of the resulting named entities. The procedure to fix the problems of the different spelling and some artifacts of NER and OIE that crowded the list of entities. Finding longer overlapping substrings with high frequencies we matched longer entities with their shorter "parents". The recognized vertexes were connected with an edge that had an estimate of sentiment and subjectivity calculated with TextBlob[4]. This naive approach yielded a hypergraph of named entities out of both data sources. The weights of the vertexes corresponded to the number of mentions of a given entity. The edges of the graph had three attributes: frequency, polarity, and subjectivity. To facilitate further research of news coverage and political discourse we share the gathered data[5].

3. Do You Know What I Know?

In this chapter, we explore the acquired knowledge graphs. In Section 3.1, we present a bird's-eye view of the graph, including key properties, and delve into the most contrasting entities and topics with varying coverage in two sources. Section 3.2 revisits the graphs, highlighting aspects crucial for differences in political discourse.

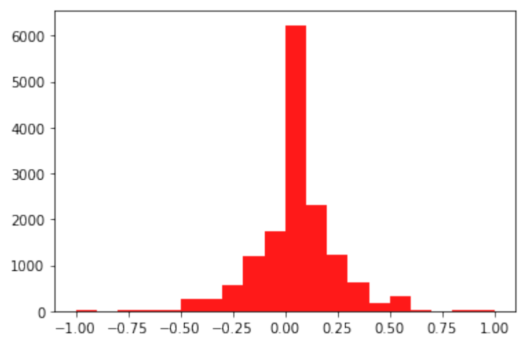

Figure 1: Breitbart News. Distribution of sentiment.

3.1. Bird's-eye View

Figures 5 – 6 show a visualization of two obtained graphs. One can see the divergence of topics:

[1] https://www.breitbart.com/
[2] https://www.nytimes.com/
[3] https://allennlp.org
[4] https://textblob.readthedocs.io/en
[5] https://shorturl.at/ntDOT

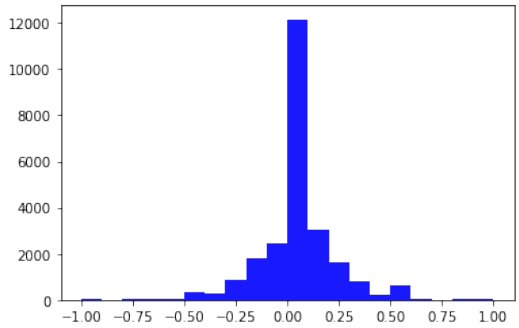

Figure 2: New York Times. Distribution of sentiment.

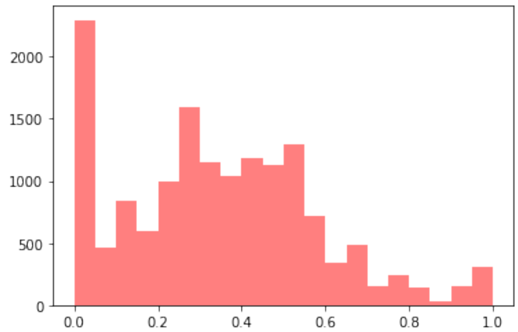

Figure 3: Breitbart News. Distribution of subjectivity.

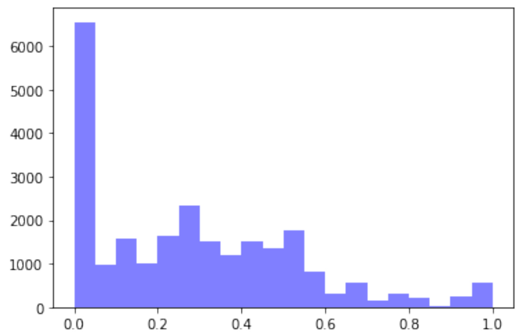

Figure 4: New York Times. Distribution of subjectivity.

Breitbart is more focused around certain personalities, while the New York Times extensively covers foreign affairs. Table 1 shows the first interesting and counter-intuitive result that one can draw when studying obtained graph representations: both media sources are "neutral" on average. Figures 1 – 2 show the distribution of polarity across all edges. The average neutral tone is not a consequence of negatively and positively charged news that balance each other. Distributions in Figures 3 – 4 do not only show that average sentiment across all edges is very close to zero for both graphs, but they also demonstrate and a vast majority of the analyzed relations are presented in a non-polarizing way (at least to the extent to which modern NLP method can distinguish polarity). One can also see the corresponding distributions of subjectivity that are similar for both sources. For the NYT Spearman correlation between polarity and subjectivity is 31%, for Breitbart, it is 23%.

Data	**Radius**	**Diameter**	**Modularity**
BN	6	11	0.43
NYT	7	13	0.53

	Average		
Data	**Path length**	**Polarity**	**Subjectivity**
BN	3.76	0.00	0.12
NYT	3.52	-0.00	0.08.

Table 1: Various parameters of the obtained graph representations. Both sources are neutral on average with Breitbart being just above and NYT just below zero average polarity. Breitbart tends to be more subjective, yet average subjectivity for both sources is at around 10%, with NYT a bit more objective.

Both media sites try to present themselves to the reader as neutral on average and moderately subjective. This stands to reason: an average reader probably neither wants to feel that she wears rose-tinted glasses nor wants to constantly read that the doom is nigh. Majority of the news are neutral, extremely positive and extremely negative news are rare in both sources. At the same time both sources tend to point bias in the coverage "on the other side". Another interesting line of thought that could be developed when regarding Table 1 is the connection between right political actors and propagation of conspiracy theories, see, for example, (Hellinger, 2018). Indeed, the Breitbart graph has smaller modularity and comparable path length. This could imply a lower encapsulation of topics and a higher tendency to connect remote entities. Even a first bird's eye view gives several fundamental insights:

- when assessed formally both right and left media demonstrate qualitatively comparable behavior; they try to cover the news in a relatively neutral tone with a pinch of subjectivity;
- the coverage of various topics differs significantly; the entities that Brietbart constantly covers tend to be people and actors of domestic US politics, whereas NYT pays more attention to institutions and international affairs;
- the overall differences between formally obtained knowledge structures that could proxy right and left world-view are minute, despite our intuition telling us otherwise.

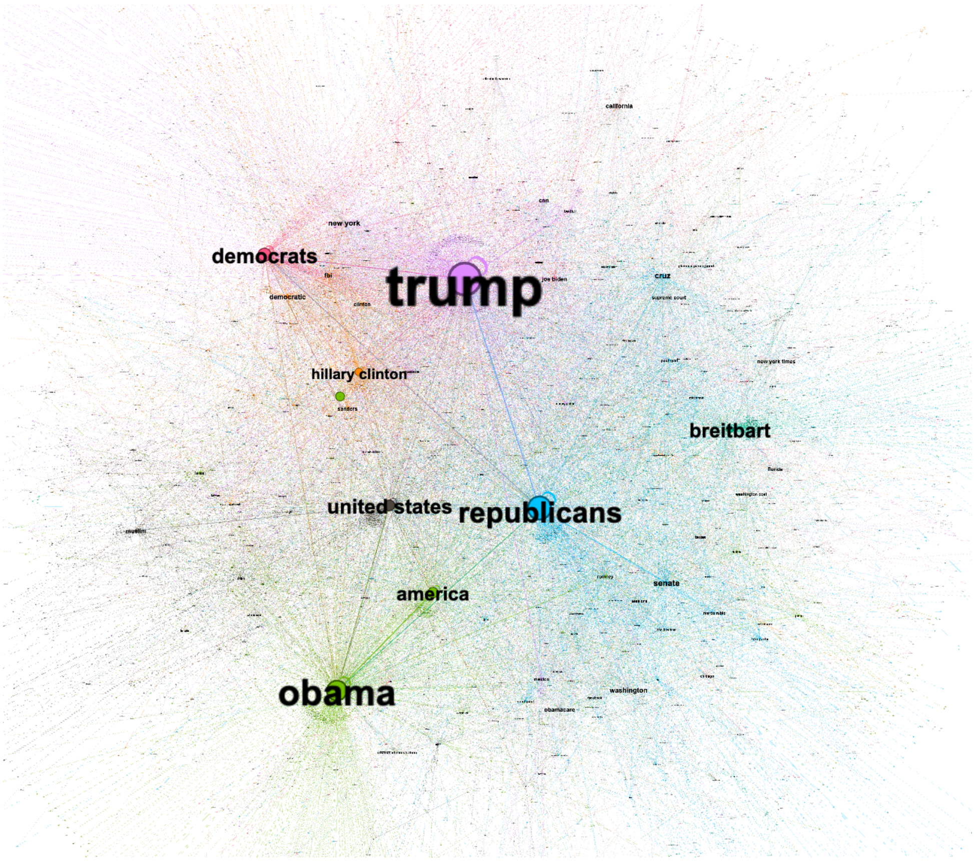

Figure 5: Breitbart News. Overall visualisation of two graphs extracted out of the media sources. The classes found with modularity analysis (Blondel et al., 2008) are highlighted with different colours. Breitbard has smaller number of classes and is centered around US political discourse.

3.2. Politics of Contrasts

Figure 7 shows a joint graph of the most polarized edges. These are the edges between entities for which the polarity in NYT and Breitbart has a different sign. Similarly, in Figure 8 one could see the most contrasting vertexes. These are the entities that have the highest average polarity of the adjacent edge. Effectively these are the representation of the polarizing topics and are covered with different polarity in both news sources.

An interesting difference between the graph of contrasting edges and the graph of contrasting nodes is that the former is mostly populated with domestic political actors, whereas the latter up to a large extent consists of entities connected with foreign affairs. This is interesting. Certain relationships between entities tend to be more polarizing for domestic issues and local politicians, yet when averaged over several such relationships across time the foreign affairs and institutions come forward. This is the same pattern that we saw earlier. One could speculate that contrasting edges highlight certain local events centered around specific politicians. Such events could be highly polarizing yet temporal. At the same time institutions and global affairs might not be as polarizing as a local scandal, yet the position of both sides on them is persistent, so when averaging across adjacent edges one sees Figure 8.

This highlights the fundamental difference between the sources. Though on macro-level both outlets prefer to stick to neutral coverage and refrain from subjectivity when it comes to certain entities and topics they provide different evaluations and tend to be more subjective in these cases. The combination of these two factors is extremely unfortunate since it facilitates social conflict. Indeed, every reader is perfectly convinced that her news source is relevant, objective, and non-biased. This also happens to be true in the vast majority of cases. Yet on a handful of key issues, the media takes a more polarizing and subjective position. Moreover, the local polarizing issues tend to be associated with personalities, while longer, fundamental differences are associated with institutions. This could be attributed to the idea of core political beliefs that could be less polarizing yet may be harder to change in the long run.

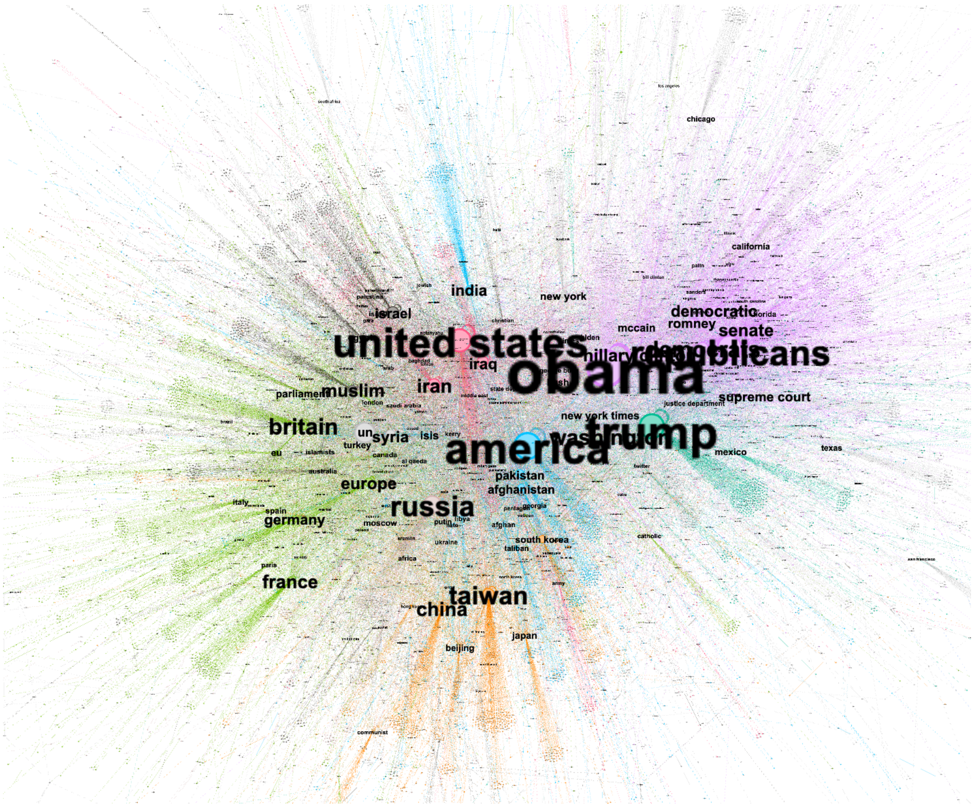

Figure 6: New York Times.Overall visualisation of two graphs extracted out of the media sources. The classes found with modularity analysis (Blondel et al., 2008) are highlighted with different colours. NYT graph has almost twice as many modularity classes and pays way more attention to foreign affairs. This could be partially attributed to the bigger size of the resulting graph, since NYT had more articles published in the studied time period.

4. Discussion

One of the key contributions of this paper is an attempt to demonstrate that an echo-chamber is not exactly a phenomenon based solely on the topology of human social networks. Using modern language processing methods and straight-forward knowledge representation we show that two different media sources paint two different pictures of the political reality. Yet these differences are less obvious than we tend to think and are more subtle. Surprisingly low average polarity and subjectivity for both knowledge structures are extremely intriguing. Assuming there is no ill will on the side of the publisher one can try to explain why the overwhelming amount of news articles try to be non-polarizing and non-subjective and with these attempts reinforce the echo-chambers around them without even trying. Could echo-chambers be a consequence of human psychological trust mechanisms on top of certain social structure formation? In (Levine, 2014) and (Clare and Levine, 2019) the authors discuss the truth-default theory. They demonstrate that when people cognitively process the content of others' communication, they typically do so in a manner characterized by unquestioned, passive acceptance. We could speculate that such behavior naturally transfers to the news sources. High neutrality and low subjectivity reinforce this truth-default. Since the preferred news outlet is often objective and neutral the reader tends to ignore or accept rare polarizing and subjective articles and dismiss the counter-argument of the other side, since in an overwhelming majority of the cases the criticism was not applicable. This might be wild speculation that demands further experimental verification. However, the very idea that echo-chamber formation could be attributed to the personal rather than collective behavior is new to our knowledge.

5. Conclusion

In this paper, we present the graphs of entities that correspond to two major "alt-right" and "liberal" news media and their coverage of the mentioned entities and relations between them. The graphs are obtained without any expert knowledge solely with NLP instruments and methods of knowledge representation. Analyzing obtained graphs we show that despite common intuition they exhibit a lot of structural similarities. We also highlight

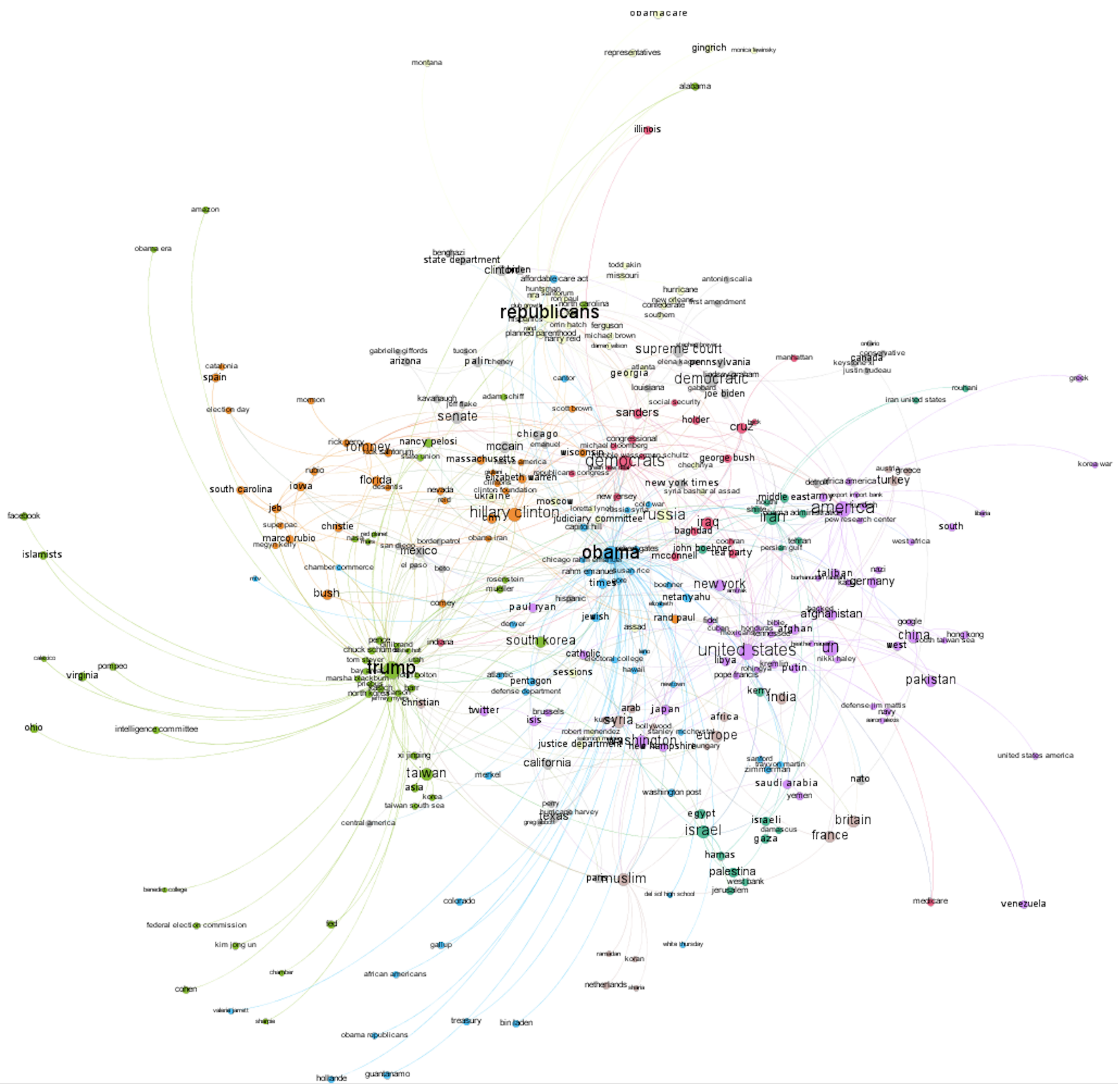

Figure 7: Sub-graph of contrasting edges. These are the edges for which the sign of polarity for BN and NYT is different.

fundamental differences that could be attributed to the formation of echo-chambers and certain biases on the world perception. We suggest that the formation of echo-chambers has more to do with the structure of information consumption and certain core beliefs of the individual rather than social structure that encompasses the aforementioned person.

Limitations

The study covers the period from 2008 to the Fall of 2019, excluding updates beyond 2019. It refrains from a detailed examination of the political aspects and perspectives of Breitbart News and New York Times readers, and it does not develop additional discussions on the global order. Considering recent global crises like wars, economic downturns in specific nations, and the worldwide impact of the COVID-19 pandemic, we anticipate that applying our methodology to recent-year data may produce slightly different findings. Nonetheless, in an effort to encourage transparent research in knowledge representation for social sciences, we provide access to our collected datasets.

Ethics Statement

Our work prioritizes transparency and relies on data collected from open sources. We refrain from making political judgments in our discussion notes to prevent discrimination and minimize potential societal harm.

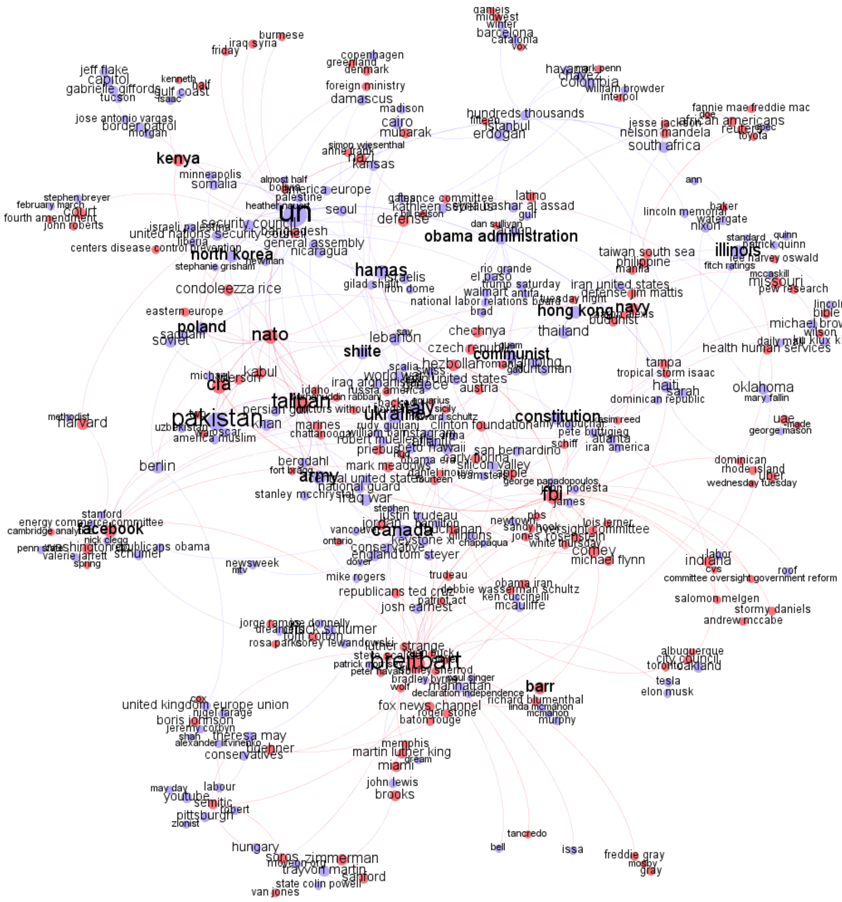

Figure 8: Sub-graph of contrasting vertexes. These are the vertexes for which the average of polarity of the adjacent edges is the highest. Blue nodes are shifted towards NYT, red — towards BN.

6. Bibliographical References

Al-Tawil M. Aljarah I. Faris H. Wongthongtham P. Chan K. Y. Abu-Salih, B. and A. Beheshti. 2021. Relational learning analysis of social politics using knowledge graph embedding. *Data Mining and Knowledge Discovery*, 35(1):1497–1536.

Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36.

American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.

Monica Anderson and Brooke Auxier. 2020. 55 In *Pew Research Center*.

Rie Kubota Ando and Tong Zhang. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853.

Galen Andrew and Jianfeng Gao. 2007. Scalable training of L_1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 33–40.

Sven Banisch and Eckehard Olbrich. 2019. Opinion polarization by learning from social feedback. *The Journal of Mathematical Sociology*, 43(2):76–103.

Francesco Barbieri, Jose Camacho-Collados, Leonardo Neves, and Luis Espinosa-Anke. 2020. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. *arXiv preprint arXiv:2010.12421*.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of* *the Association for Computational Linguistics*, 5:135–146.

E. Borel. 1921. La theorie du jeu et les equations integrales a noyau symetrique. *Comptes rendus hebdomadaires des seances de l'Academie des sciences*, (173):1304–1308.

Romain Campigotto, Patricia Conde-Céspedes, and Jean-Loup Guillaume. 2014. A generalized and adaptive method for community detection.

Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

Wei Chen, Xiao Zhang, Tengjiao Wang, Bishan Yang, and Yi Li. 2017. Opinion-aware knowledge graph for political ideology detection. In *IJCAI*, pages 3647–3653.

P. Chilton. 2004. *Analysing political discourse: Theory and practice*. Routledge.

David D Clare and Timothy R Levine. 2019. Documenting the truth-default: The low frequency of spontaneous unprompted veracity assessments in deception detection. *Human Communication Research*, 45(3):286–308.

Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. 2014. Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. *Journal of communication*, 64(2):317–332.

Rosaria Conte, Nigel Gilbert, Giulia Bonelli, Claudio Cioffi-Revilla, Guillaume Deffuant, Janos Kertesz, Vittorio Loreto, Suzy Moat, J-P Nadal, Anxo Sanchez, et al. 2012. Manifesto of computational social science. *The European Physical Journal Special Topics*, 214(1):325–346.

James W. Cooley and John W. Tukey. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301.

Pieter Delobelle, Murilo Cunha, Eric Cano, Jeroen Peperkamp, and Bettina Berendt. 2019. Computational ad hominem detection. pages 203–209.

Felix Gaisbauer, Armin Pournaki, Sven Banisch, and Eckehard Olbrich. 2023. Grounding force-directed network layouts with latent space models. *Journal of Computational Social Science*, pages 1–33.

Kiran Garimella, Gianmarco Morales, Aristides Gionis, and Michael Mathioudakis. 2016. Quantifying controversy in social media. pages 33–42.

Goran Glavaš, Federico Nanni, and Simone Paolo Ponzetto. 2017. Cross-lingual classification of topics in political texts. In *Proceedings of the Second Workshop on NLP and Computational Social Science*, pages 42–46.

David Goldie, Matthew Linick, Huriya Jabbar, and Christopher Lubienski. 2014. Using bibliometric and social media analyses to explore the "echo chamber" hypothesis. *Educational Policy*, 28(2):281–305.

T. Graham, D. Jackson, and M. Broersma. 2016. New platform, old habits? candidates use of twitter during the 2010 british and dutch general election campaigns. *New Media & Society*, 18(5):765–783.

O. Gross and R. Wagner. 1950. A continuous colonel blotto game. RAND Research Memorandum.

Fangjian Guo, Charles Blundell, Hanna Wallach, and Katherine Heller. 2015. The bayesian echo chamber: Modeling social influence via linguistic accommodation. In *Artificial Intelligence and Statistics*, pages 315–323.

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

Lisa Harris and Paul Harrigan. 2015. Social media in politics: The ultimate voter engagement tool or simply an echo chamber? *Journal of Political Marketing*, 14(3):251–283.

Daniel C Hellinger. 2018. *Conspiracies and conspiracy theories in the age of trump*. Springer.

Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PloS one*, 9:e98679.

Julie Jiang, Xiang Ren, and Emilio Ferrara. 2021. Social media polarization and echo chambers: A case study of covid-19.

Zhijing Jin, Abhinav Lalwani, Tejas Vaidhya, Xiaoyu Shen, Yiwen Ding, Zhiheng Lyu, Mrinmaya Sachan, Rada Mihalcea, and Bernhard Schölkopf. 2022. Logical fallacy detection. page 1.

Kristen Johnson and Dan Goldwasser. 2018. Classification of moral foundations in microblog political discourse. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 720–730.

A. Jungherr. 2014. The logic of political coverage on twitter: Temporal dynamics and content. *Journal of Communication*, 64(2):239–259.

Jonathan P Kastellec and Eduardo L Leoni. 2007. Using graphs instead of tables in political science. *Perspectives on politics*, pages 755–771.

Y. Kim. 2014. Convolutional neural networks for sentence classification. ArXiv:1408.5882.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5, pages 79–86. Citeseer.

D. Kovenock and B. Roberson. 2015. Generalizations of the general lotto and colonel blotto games. CESifo Working Paper 5291.

Werner Krause, Pola Lehmann, Jirka Lewandowski, Theres Matthieß, Nicolas Merz, Sven Regel, and Annika Werner. 2018. Manifesto corpus. *WZB Berlin Social Science Center*.

Onawa P Lacewell and Annika Werner. 2013. Coder training: Key to enhancing coding reliability and estimate validity.

Man Lan, Chew Lim Tan, Jian Su, and Yue Lu. 2008. Supervised and traditional term weighting methods for automatic text categorization. *IEEE transactions on pattern analysis and machine intelligence*, 31(4):721–735.

J.-F. Laslier and N. Picard. 2002. Distributive politics and electoral competition. *Journal of Economic Theory*, (103):106–130.

David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science*, 359(6380):1094–1096.

P. Lehmann, T. Matthieß, N. Merz, S. Regel, and A. Werner. 2017. Manifesto corpus 2017-1. WZB Berlin Social Science Center.

Timothy R Levine. 2014. Truth-default theory (tdt) a theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33(4):378–392.

Steven Loria et al. 2018. textblob documentation. *Release 0.15*, 2(8):269.

Shawn Martin, W. Brown, Richard Klavans, and Kevin Boyack. 2011. Openord: An open-source toolbox for large graph layout. *Proc SPIE*, 7868:786806.

Nicolas Merz, Sven Regel, and Jirka Lewandowski. 2016. The manifesto corpus: A new resource for research on political parties and quantitative text analysis. *Research & Politics*, 3(2):2053168016643346.

Slava Mikhaylov, Michael Laver, and Kenneth R Benoit. 2012. Coder reliability and misclassification in the human coding of party manifestos. *Political Analysis*, 20(1):78–91.

R.B. Myerson. 1993. Incentives to cultivate minorities under alternative electoral systems. *American Political Science Review*, 87:856–869.

Usman Naseem, Imran Razzak, Katarzyna Musial, and Muhammad Imran. 2020. Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Generation Computer Systems*, 113.

R. Neuman, L. Guggenheim, S. Mo Jang, and S.Y. Bae. 2014. The dynamics of public attention: Agenda setting theory meets big data. *Journal of Communication*, 64(2):193–214.

Sri Nurdiati and Cornelis Hoede. 2008. 25 years development of knowledge graph theory: the results and the challenge. *Memorandum*, 1876(2):1–10.

A. Osorio. 2013. The lottery blotto game. *Economics Letters*, 120(2):164–166.

I. Parker. 2014. *Discourse Dynamics (Psychology Revivals): Critical Analysis for Social and Individual Psychology*. Routledge.

Tiago P Peixoto. 2020. Latent poisson models for networks with heterogeneous density. *Physical Review E*, 102(1):012309.

Matthew E. Peters, Waleed Ammar, Chandra Bhagavatula, and Russell Power. 2017. Semi-supervised sequence tagging with bidirectional language models. *ArXiv*, abs/1705.00108.

John Petit, Cong Li, and Khudejah Ali. 2020. Fewer people, more flames: How pre-existing beliefs and volume of negative comments impact online news readers' verbal aggression. *Telematics and Informatics*, 56:101471.

Mohammad Sadegh Rasooli and Joel R. Tetreault. 2015. Yara parser: A fast and accurate dependency parser. *Computing Research Repository*, arXiv:1503.06733. Version 2.

Arsenii Rasov, Ilya Obabkov, Eckehard Olbrich, and Ivan P Yamshchikov. 2020. Text classification for monolingual political manifestos with words out of vocabulary. In *COMPLEXIS*, pages 149–154.

B. Roberson. 2006a. The colonel blotto game. *Economic Theory*, 29(1):1–24.

B. Roberson. 2006b. Pork-barrel politics, discriminatory policies and fiscal federalism. Social Science Research Center Berlin (WZB).

Richard Rogers. 2013. *Digital methods*. MIT press.

Wilbert Samuel Rossi, Jan Polderman, and Paolo Frasca. 2018. The closed loop between opinion formation and personalised recommendations.

Martin Rosvall and Carl T Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the national academy of sciences*, 105(4):1118–1123.

Dhavan V Shah, Joseph N Cappella, and W Russell Neuman. 2015. Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, 659(1):6–13.

Jesse Shore, Jiye Baek, and Chrysanthos Dellarocas. 2018. Network structure and patterns of information diversity on twitter. *MIS Quarterly*, 42.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36.

Gabriel Stanovsky, Julian Michael, Luke Zettlemoyer, and Ido Dagan. 2018. Supervised open information extraction. In *NAACL-HLT*.

Shivashankar Subramanian, Trevor Cohn, and Timothy Baldwin. 2018. Hierarchical structured model for fine-to-coarse manifesto text analysis. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1964–1974.

Shivashankar Subramanian, Trevor Cohn, Timothy Baldwin, and Julian Brooke. 2017. Joint sentence-document model for manifesto text analysis. In *Proceedings of the Australasian Language Technology Association Workshop 2017*, pages 25–33.

Andon Tchechmedjiev, Pavlos Fafalios, Katarina Boland, Malo Gasquet, Matthäus Zloch, Benjamin Zapilko, Stefan Dietze, and Konstantin Todorov. 2019. Claimskg: a knowledge graph of fact-checked claims. In *International Semantic Web Conference*, pages 309–324. Springer.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. *arXiv preprint cs/0607062*.

Richard L Tweedie, Kerrie L Mengersen, and John A Eccleston. 1994. Garbage in, garbage out: can statisticians quantify the effects of poor data? *Chance*, 7(2):20–27.

Suzan Verberne, Eva D'hondt, Antal van den Bosch, and Maarten Marx. 2014. Automatic thematic classification of election manifestos. *Information Processing & Management*, 50(4):554–567.

Zhouxia Wang, Tianshui Chen, Jimmy Ren, Weihao Yu, Hui Cheng, and Liang Lin. 2018. Deep reasoning with knowledge graph for social relationship understanding. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 1021–1028.

Michael D Ward, Katherine Stovel, and Audrey Sacks. 2011. Network analysis and political science. *Annual Review of Political Science*, 14:245–264.

A. Washburn. 2013. Blotto politics. *Operations Research*, 61(3):532–543.

Ivan P Yamshchikov and Sharwin Rezagholi. 2019. Elephants, donkeys, and colonel blotto. In *COMPLEXIS*, pages 113–119.

Cäcilia Zirn, Goran Glavaš, Federico Nanni, Jason Eichorts, and Heiner Stuckenschmidt. 2016. Classifying topics and detecting topic shifts in political manifestos.