Computational Network Analysis of Metajournalistic Discourse: Analyzing Journalism Coverage of Russia's Invasion of Ukraine

Iuliia Alieva[1,2 [0000-0001-6270-8985]] and Kathleen M Carley[1 [0000-0002-6356-0238]]

[1] Carnegie Mellon University, Pittsburgh PA, USA
[2] University of Stuttgart, Germany
ialieva@cmu.edu

Abstract. This study aims to integrate theoretical frameworks from media sociology, focusing on metajournalistic discourse, to examine discussions surrounding war journalism coverage during Russia's full-scale invasion of Ukraine. The study employs computational methods and network science through a mixed-method approach. This analysis explores war journalism reporting practices related to the conflict by integrating social media analytics, Twitter (X) data collection, data analysis, information processing, and an investigation of the actors and narratives.

Keywords: Social Network Analysis, Metajournalistic Discourse, Journalism, Conflict, Mixed-Method Analysis.

1 Introduction

1.1 Metajournalistic discourse

Traditionally, researchers investigating discussions surrounding journalism issues employ a metajournalistic discourse theoretical framework when analyzing conversations about journalism practices [1, 2]. Metajournalistic discourse illustrates how various actors engage in processes of defining journalism, setting its boundaries, and making judgments about its legitimacy. It also explains how meanings surrounding journalism are formed [1]. This approach allows journalism to be situated as a cultural practice by analyzing various actors both inside and outside the field, as well as examining conversations where meanings about journalism are constructed, reiterated, and challenged [1]. Among other issues, it includes cultural and rhetorical conversations about journalism's mission, questions regarding journalistic ethics, and the societal role of the profession. The theory of metajournalistic discourse connects three components: actors, sites/audiences, and topics. These components shape processes of definition-making, boundary work, and legitimation within the field of journalism through discussions between internal practitioners and external actors. Based on theoretical premises, the discourse demonstrates how journalism varies across time and space, encompassing dif-

ferent types of practices, various normative understandings that legitimize these practices, and the conditions in which news practices emerge, such as regulatory regimes, political and media institutions, and cultural norms that shape news forms. It also shows that journalism is contextual, developing as a result of social relationships. Additionally, it illustrates how understandings of journalism practices are formed and how professional meanings are developed. This is central to journalism as a gatekeeping institution, as the profession lacks firm boundary markers [3] and represents the primary framework through which journalists construct their professional norms and ideals [4].

In studies of metajournalistic discourse, scholars have primarily focused on journalists as research subjects, downplaying the role of other actors in shaping journalistic values and norms [3]. Previously, metajournalistic discourse methodology has included qualitative analysis of professional trade journals and journalists' reflections and opinions, often overlooking the social and narrative connections, as well as how audiences, influencers, and various societal groups participate in discussions that influence journalistic practices. To enhance metajournalistic discourse analysis, we propose an approach that links the investigation of various actors, the sites where they speak, the audiences they address, and the topics they discuss, for a more comprehensive understanding of narratives, boundaries, and legitimization in journalism. To achieve this, we propose a mixed-method social network analysis to identify key groups and communities, influential actors, their roles in the discussions, and the topics and narratives they initiate.

Social network analysis (SNA) can be used to examine complex social discourse by combining quantitative and qualitative approaches, allowing for the investigation of the form and content of relationships, opinions, and discussions [5, 6]. The theory of social networks integrates traditional structural analysis with a strong emphasis on meanings within networks. Networks are viewed as social structures with three dimensions: the structure of social relationships, individual actors and their connections, and the meaning associated with these networks [7]. SNA has the potential to enhance metajournalistic discourse by offering a mixed-method approach to understanding actors, connections, and narratives.

Using SNA, this study aims to integrate the theoretical frameworks of media sociology, with a focus on metajournalistic discourse, to examine discussions surrounding journalism reporting during Russia's full-scale invasion of Ukraine. It employs computational methods and network science through a mixed-method approach. This analysis explores war journalism reporting practices related to the conflict in Ukraine, encompassing social media analytics, Twitter data collection, data analysis, and information processing, as well as the investigation of actors and narratives.

1.2   Research Questions

For this study, we propose a mixed-methods approach that incorporates computational network analysis to identify metajournalistic discourse. In addition to existing qualitative discourse analysis previously used for this purpose, this methodology acknowledges the influence of external actors and societal attitudes on journalistic conversa-

tions. Network analysis helps identify the main actors and groups, influencers, and various discussions surrounding the phenomena, situating metajournalistic discourse within a broader context. Additionally, Leiden clustering and sphere of influence (ego network) analysis are implemented to identify influential communities and their networks. Natural language processing with BERTopic is employed to identify discussion topics. We also incorporate qualitative analysis to investigate actors and narratives in the discourse for better research triangulation. To guide our proposed analysis, we formulated the following research questions:

- Which actors and communities are present in the discourse surrounding journalistic coverage of Russia's invasion of Ukraine?
- How do actors and communities articulate conversations about journalism and war coverage to influence journalistic practices?
- What narratives are identified and promoted by the actors and communities within this discourse?

2   Methodology

This study employs a mixed-method approach that combines computational methods, including network science methodology, community clustering, sphere of influence (ego network) analysis, and natural language processing, alongside qualitative discourse analysis. Through network analysis, community clustering, and sphere of influence analysis, we identify influential actors and connected clusters using network metrics, with Twitter (X) serving as an aggregator for discussions. The Twitter (X) dataset, collected using relevant keywords, focuses on war journalism during Russia's invasion of Ukraine. The methodology encompasses network science metrics for influencer identification, sphere of influence (ego network) analysis for examining groups of journalism/media accounts, Leiden clustering for community detection, the BERTopic method for natural language processing to identify main topics and narratives, as well as qualitative discourse analysis (see Figure 1 for an overview of the methodology approach).
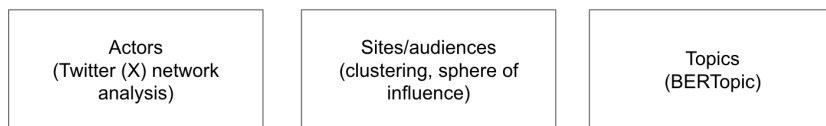
Actors (Twitter (X) network analysis)	Sites/audiences (clustering, sphere of influence)	Topics (BERTopic)

Fig. 1. Metajournalistic discourse methodology with computational network analysis.

We collected tweets using the Twitter (X) Academic Access API version 2 with the Python package twarc, employing keywords that combined terms such as "journalism," "conflict," "Ukraine," and others. The collection period spanned from three months before the invasion (November 24, 2021) to one year after the invasion (March 16, 2023). As a result, we gathered 1,527,846 tweets from 571,980 actors. We converted the raw Twitter data into a meta-network consisting of user-to-user communication networks, user-to-tweet networks, and user-to-various tweet artifacts (hashtags, URLs) using ORA software to conduct network analysis [8].

ORA was also used to identify influencers in the communication network. Influencers are accounts, whose tweets wield substantial influence within the social network due to their significant follower count and strategic network position. The narratives they propagate have the power to sway the opinions of other users in the network. ORA software identifies Twitter influencers based on computed network metrics from the Twitter data and communication networks between users in the dataset. This study focuses on two specific categories: super spreaders (identified using a combination of network analysis metrics such as out-degree centrality, page rank centrality, and k-core) and super friends (determined through a combination of total degree centrality and k-core). The super spreaders list includes users who consistently produce content that is widely and efficiently circulated throughout the network. The super friends list encompasses users engaged in regular two-way communication, thereby contributing to the development of extensive and resilient communication networks.

Additionally, we implement sphere of influence (ego network) analysis for the list of media and journalism accounts. The sphere of influence illustrates the direct relationships between an account and its neighbors, where the neighbors are the nodes within a specified path length.

To identify network communities participating in the conversations, we use the Leiden clustering method. The Leiden clustering algorithm involves network partitioning and node movement, ensuring the formation of well-connected communities. The Leiden algorithm has been shown to be more efficient than others, such as Louvain; it is also faster and uncovers better partitions [9]. After identifying the communities, qualitative methods were employed to compare content and user characteristics between groups. We conducted influencer analysis on the largest Leiden communities, pinpointing the primary attitudes expressed by influencers within each group. Due to the substantial number of communities and actors, we opted to concentrate on the most influential users, as their content enjoys widespread dissemination within each community.

BERTopic modeling is employed to identify major topics for the groups using the textual content of tweets. BERTopic has demonstrated superiority over other topic modeling strategies, such as LDA, particularly for short-text documents [10]. The BERTopic pipeline involves using an embedding model, typically BERT-based, to generate vector representations of the text. This is followed by a dimensionality reduction step and a clustering step to group similar documents [11].

3   Results

Through social network analysis, we identified several communities and actors participating in discussions about journalistic coverage of Russia's invasion of Ukraine. This analysis revealed various groups of journalists and influencers, including conservative and alt-right communities. The largest group comprises conservative and alt-right influencers and opinion leaders, Western reporters who support Russian pro-government narratives, low-credibility sites with anti-West sentiments, and those with sympathetic views toward authoritarian regimes. The second-largest group consists of pro-Ukraine actors, including media figures and politicians. Additionally, we identified

other actors and communities, such as American liberal and conservative media and journalists, British media and journalists, Western war correspondents and analysts, Indian media and journalists, as well as international media organizations, reporters, and mainstream media outlets.

We identified topics in the discussion using BERTopic and conducted a qualitative investigation. For the largest group, we found multiple critical narratives regarding Western journalism and media, including accusations of biased coverage, neglect of the people in the Donbas region of Eastern Ukraine, and the crimes of the Ukrainian government. Additionally, the actors referenced the Julian Assange case to highlight perceived hypocrisy in Western coverage (see Figure 2 for examples of topics).
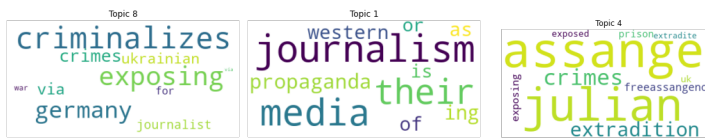

Fig. 2. Topics for the largest group.

The second largest group presents narratives more closely related to war news and discourse, addressing the deaths and dangers faced by journalists working in conflict zones. This group also covers speeches and announcements from the Ukrainian government, as well as anti-war protest activities, exemplified by Russian journalist Marina Ovsyannikova, who worked for Russian state television and protested against the war during a main news program (see Figure 3 for examples).

Fig. 3. Topics for the second largest group.

As further steps in the analysis, we plan to investigate groups, topics and narratives for the general understanding of the metajournalistic discourse and discussions in more details.

4   Conclusion

Social network analysis (SNA) helped us identify multiple communities and discussions related to the journalism coverage of Russia's invasion of Ukraine. We observed that the main discussions focus on different sides of the conflict. We identified actors, influencers, communities, and topics using mixed-method analysis; however, there is a

need to decipher and interpret these topics, taking into account the biases of the communities and groups and how they fit into the conversation around journalistic practices.

For future steps, we plan to implement additional qualitative methods, such as interviews and coverage analysis, and potentially identify disinformation and influence campaigns through discourse analysis.

References

1. Carlson, M.: Metajournalistic discourse and the meanings of journalism: Definitional control, boundary work, and legitimation. Communication theory 26(4), 349-368 (2016).
2. Perreault, G., Perreault, M. F., & Maares, P.: Metajournalistic discourse as a stabilizer within the journalistic field: Journalistic practice in the COVID-19 pandemic. Journalism Practice 16(2-3), 365-383 (2022).
3. Farkas, J.: Fake News in Metajournalistic Discourse. Journalism Studies 24(4), 423–441. (2023).
4. Vos, T. P., & Thomas, R. J.: The Discursive (Re)construction of Journalism's Gatekeeping Role. Journalism Practice, 13(4), 396–412. (2018).
5. Edwards, G., & Crossley, N.: Measures and Meanings: Exploring the Ego-Net of Helen Kirkpatrick Watts, Militant Suffragette. Methodological Innovations Online, 4(1), 37-61. (2009).
6. Schipper, D. & Spekkink, W.: Balancing the Quantitative and Qualitative Aspects of Social Network Analysis to Study Complex Social Systems. Complexity, Governance & Networks 1 (1): 5–22. (2015).
7. Fuhse J., & Mützel, S.: Tackling connections, structure, and meaning in networks: quantitative and qualitative methods in sociological network research. Quality & Quantity: International Journal of Methodology, Springer, 45(5), 1067-1089. (2011).
8. Carley, K.M.: ORA: A Toolkit for Dynamic Network Analysis and Visualization. In Reda Alhajj and Jon Rokne (Eds.) Encyclopedia of Social Network Analysis and Mining, Springer. (2014).
9. Traag, V. A., Waltman, L., & Van Eck, N. J.: From Louvain to Leiden: guaranteeing well-connected communities. Scientific reports 9(1), 5233 (2019).
10. de Groot, M., Aliannejadi, M., & Haas, M. R.: Experiments on generalizability of BERTopic on multi-domain short text. *arXiv preprint arXiv:2212.08459* (2022).
11. Grootendorst, M.: BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).