The Russian invasion of Ukraine through the lens of ex-Yugoslavian Twitter

Bojan Evkoski
bojan.evkoski@ijs.si
Jozef Stefan Institute, and
Jozef Stefan Postgraduate School
Ljubljana, Slovenia

Igor Mozetič
igor.mozetic@ijs.si
Jozef Stefan Institute
Ljubljana, Slovenia

Petra Kralj Novak
petra.kralj.novak@ijs.si
Central European University
Vienna, Austria, and
Jozef Stefan Institute
Ljubljana, Slovenia

Nikola Ljubešić
nikola.ljubesic@ijs.si
Jozef Stefan Institute, and
Faculty of Computer and Information Science,
University of Ljubljana
Ljubljana, Slovenia

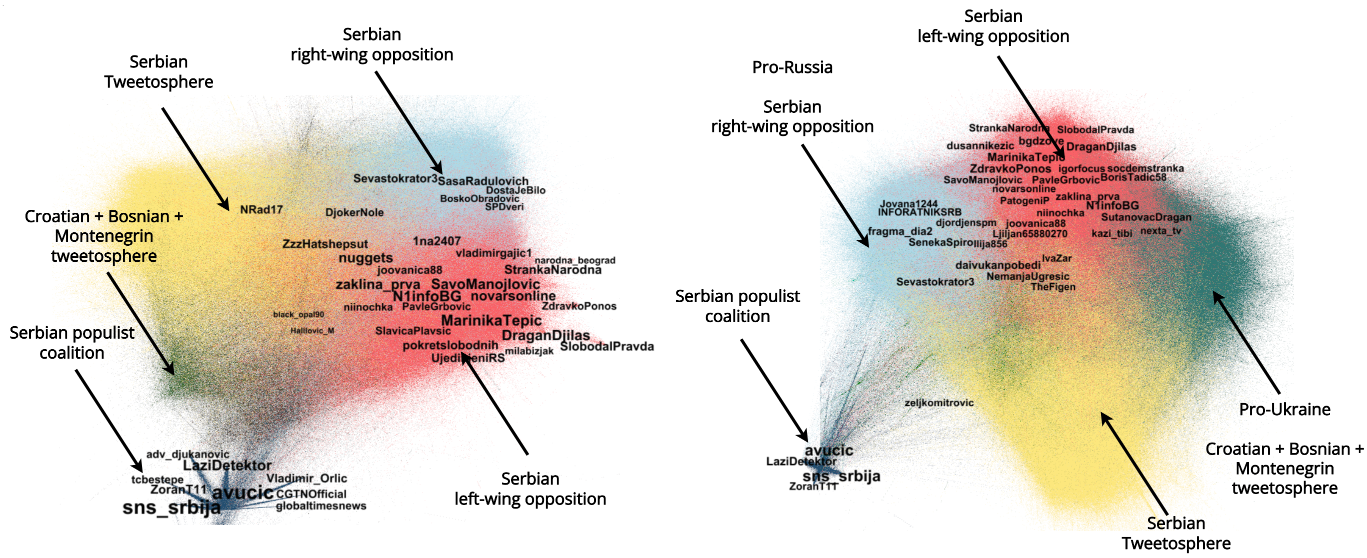

Figure 1: Pre-invasion (left) and invasion (right) ex-Yugoslavian retweet networks. Node colors represent communities. Labeled arrows point to the main communities, with labels inferred from the community users. The in-network labels represent the names of the most retweeted accounts.

ABSTRACT

The Russian invasion of Ukraine marks a dramatic change in international relations globally, as well as at specific, already unstable, regions. The geographical area of interest in this paper is a part of ex-Yugoslavia where the BCMS (Bosnian, Croatian, Montenegrin, Serbian) languages are spoken, official varieties of a pluricentric Serbo-Croatian macro-language [4]. We analyze 12 weeks of Twitter activities in this region, six weeks before the invasion, and six weeks after the start of the invasion. We form retweet networks and detect retweet communities which closely correspond to groups of like-minded Twitter users. The communities are distinctly divided across countries and political orientations. Some communities detected after the start of the Russian invasion also show clear pro-Ukrainian or pro-Russian stance. Such analyses of social media help in understanding the role and effect of this conflict at the regional level.

KEYWORDS

social network analysis, community detection, Twitter

1 INTRODUCTION

The Russian invasion of Ukraine brings about dramatic changes to the world. Analysing the structure and content of the communication on social media, such as Twitter, can give more insight into the causes, developments and consequences of this conflict. The geographical area of interest in our research is a part of ex-Yugoslavia where the BCMS (Bosnian, Croatian, Montenegrin, Serbian) languages are spoken, official varieties of the pluricentric Serbo-Croatian macro-language. This area is strongly politically divided by diverging influences of NATO (Croatia, Montenegro, North Macedonia, Bosniak and Croatian entity in Bosnia and

Information Society 2022, 10–14 October 2022, Ljubljana, Slovenia

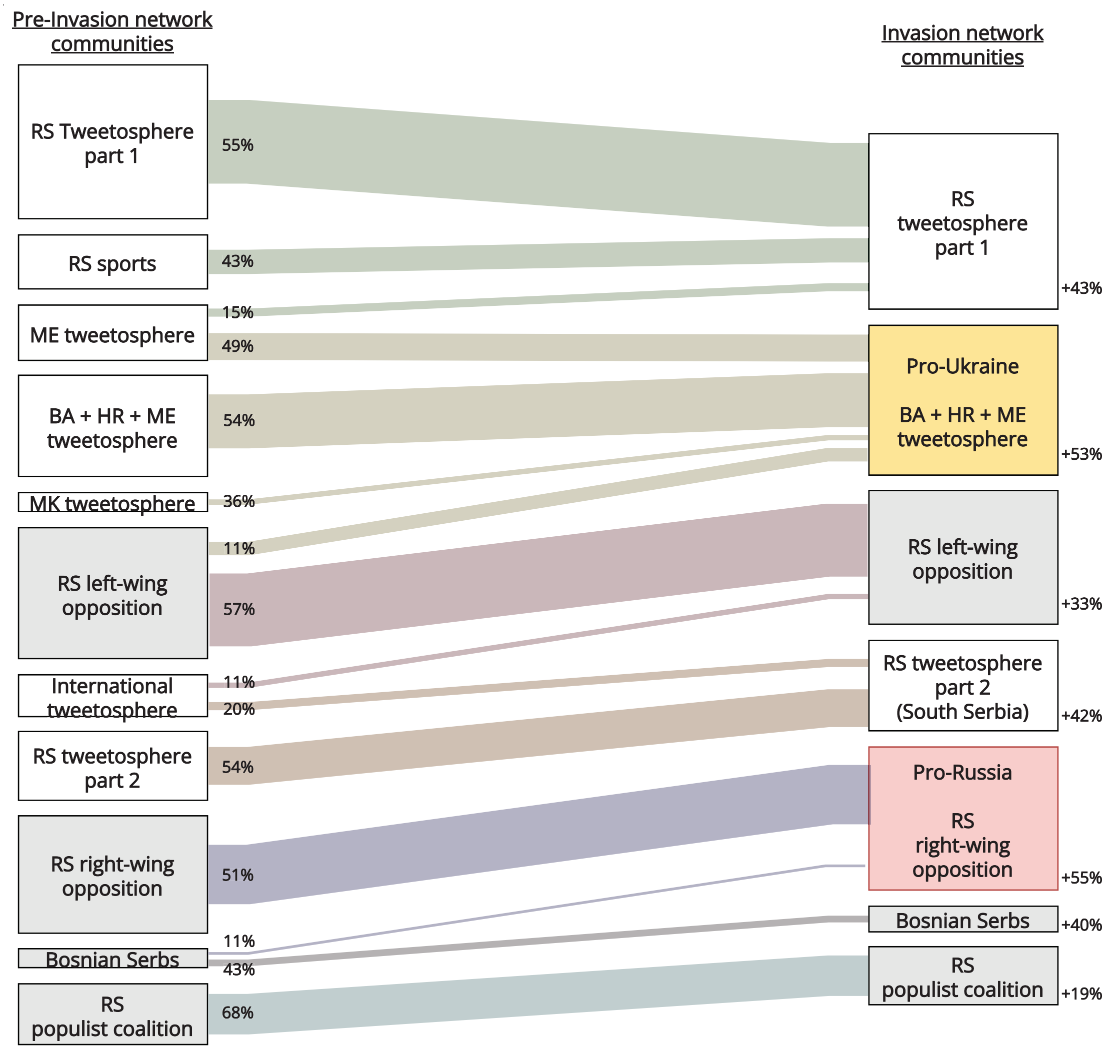

Figure 2: A Sankey diagram showing the transitions of users from the pre-invasion network communities (left) to the invasion network communities (right). Rectangle height is proportional to the community sizes. Percentages near the pre-invasion communities show the portion of users found in the corresponding invasion communities. Percentages on the right-hand side of the invasion communities show the portion of users not previously present in the large communities of the pre-invasion network. Gray rectangles depict the communities tightly related to politics, with the yellow and red denoting the detected pro-Ukraine and pro-Russia leaning communities, respectively.

Herzegovina) and Russia (Serbia, Serbian entity in Bosnia and Herzegovina). While Croatia is full EU member since 2013, Montenegro, North Macedonia and Serbia are EU candidate members, while Bosnia and Herzegovina is a potential candidate. Regarding military alliances, NATO members are Croatia (since 2007), Montenegro (since 2017) and North Macedonia (since 2020), while Serbia does not aspire to join NATO, primarily due to a complex Serbia-NATO relationship caused by the NATO intervention in Yugoslavia in 1999.

To shed light on the impact of the Russian invasion on this brittle and complex geographical and political area, we use social network analysis over available Twitter data, 6-weeks before and 6-weeks during the invasion. We discover a complex landscape of ideology-specific and country-specific communities (see Figure 1), and analyse the transition into evident pro-Ukraine and pro-Russia leanings. We also present a method to measure the similarity of the communities before and during the invasion by analyzing URL and hashtag usage. As the communities show very divergent properties, we echo concerns of the heavy polarization and possible destabilization of this area of the Balkans.

2 RESULTS

The data analysed in this study were collected with the TweetCat tool [3], focused on harvesting tweets of less frequent languages. TweetCat is continuously searching for new users tweeting in the language of interest by querying the Twitter Search API for the most frequent and unique words in that language. Every user identified to tweet in the language of interest is continuously collected from that point onward. This data collection procedure is run for the BCMS set of languages since 2017. During the 12 weeks of our focus, we collected 1.2M tweets and 3.8M retweets from 45,336 users. A rough estimate of the per-country production of tweets via URL usage from country-specific top-level domains (upper part of Table 1) shows for Twitter to be much more popular in Serbia and Montenegro than in Croatia or Bosnia and Herzegovina. This has to be taken into account while analysing the communities of the underlying tweetosphere.

We created **pre-invasion** and **invasion retweet networks** (users as nodes, retweets as edges) from the collected data. We applied community detection (Ensemble Louvain [1]) on the two

Country	Population	URLs		
Serbia (RS)	**7.2M (47.3%)**	**106K (44.2%)**		
Croatia (HR)	3.9M (25.6%)	19.6K (8.1%)		
Bosnia and Herzegovina (BA)	3.5M (23.0%)	14.9K (6.2%)		
Montenegro (ME)	620K (4.1%)	24.7K (10.2%)		
Total	15.2M	242K		
Pre-invasion communities	**Users**	**Tweets**	**Retweets**	**Intra-com. RTs**
RS tweetosphere part 1	**13K (29.0%)**	**125K (24.9%)**	300K (18.9%)	80.3%
RS tweetosphere part 2	2.5K (5.6%)	35.8K (7.1%)	63.2K (4.0%)	62.3%
RS sports	1.6K (3.6%)	12.6K (2.5%)	25.6K (1.6%)	53.8%
ME tweetosphere	1.7K (3.8%)	22.7K (4.5%)	44.6K (2.8%)	74.5%
BA + HR + ME tweetosphere	5.6K (12.4%)	37.8K (7.5%)	59K (3.7%)	75.3%
Macedonian tweetosphere	200 (0.4%)	721 (0.1%)	771 (0.1%)	77.7%
International tweetosphere	934 (2.0%)	8.5K (1.7%)	11.5K (0.7%)	62.3%
RS populist coalition	2.0K (4.8%)	52.4K (10.4%)	396K (24.9%)	**98.7%**
RS left-wing opposition	9.3K (20.6%)	105K (20.9%)	**408K (25.5%)**	80.5%
RS right-wing opposition	7.6K(16.8%)	87.8K (17.4%)	247K (15.5%)	72.1%
Bosnian Serbs	139 (0.3%)	2.2K (0.4%)	3.8K (0.2%)	83.1%
Total	45.3K	502.9K	1590K	
Invasion communities	**Users**	**Tweets**	**Retweets**	**Intra-com. RTs**
RS tweetosphere part 1	**16.9K (29.5%)**	160K (22.4%)	387K (16.8%)	71.1%
RS tweetosphere part 2	4.5K (7.7%)	57.3K (8.1%)	118K (5.1%)	58.1%
Pro-Ukraine BA + HR + ME tweetosphere	12.4K (21.7%)	76.1K (10.6%)	235K (10.2%)	64.7%
Pro-Russia RS right-wing opposition	11.1K (19.4%)	129K (17.9%)	508K (22.1%)	65.1%
RS populist coalition	1.8K (3.1%)	**208K (29.1%)**	450K (19.5%)	**95.6%**
RS left-wing opposition	9.8K (17.2%)	191K (26.7%)	**590K (25.6%)**	72.6%
Bosnian Serbs	356 (0.6%)	5.4K (0.7%)	7.1K (0.3%)	62.3%
Total	57.4K (+26.7%)	717K (+42.8%)	2302K (44.8%)	

Table 1: The first part shows general population of each BCMS country and their respective tweet URL shares (.rs, .hr, .ba and .me). The second part shows the pre-invasion network communities with the number of users, tweets, retweets and intra-community retweets. The third part shows the same statistics for the invasion network communities. Grey rows depict political communities, while yellow and red show the pro-Ukraine and pro-Russia communities, respectively.

networks and analysed the community properties and user transitions [2]. We identified and named the large communities (more than 100 users) by a careful analysis of their most influential users and hashtag/URL usage. Figure 2 depicts the user transitions between the two networks, while Table 1 shows general statistics of each community. We discovered the following peculiarities:

- The BCMS tweetosphere is dominated by Serbian (RS) users and content.
- The political communities are more active compared to the non-political ones.
- RS populist coalition community (led by the Serbian president Aleksandar Vučić) forms a very strong echo chamber, with less than 2% of all users, yet more than 25% of tweets and retweets and more than 95% of intra-community retweets.
- RS populist coalition and left-wing opposition remain neutral on the invasion topic.
- RS right-wing opposition and the Bosnian Serbs show a clear pro-Russia stance.
- Croatian, Bosnian and Montenegrin communities show a clear pro-Ukraine stance.

In order to compare the pre-invasion and invasion communities in terms of content and political leanings, our following goal was to compare the pool of hashtags used and URLs shared by the community users. Therefore, we developed a simple community similarity method. First, we preprocessed the URLs by manually filtering out the ones coming from social media sources like Twitter, Facebook, Youtube etc., as well as URL shorteners. With this, we created a subset in which more than 99% of the URLs were news media, making it ideal for media polarization analysis. Once we extracted the domain of the URLs, we then created sorted lists of the top 50 URL domains and top 50 hashtags for each community, sorted by the usage counts. Finally, in order to calculate the similarities between communities, we used the Rank-biased overlap (RBO) measure for indefinite rankings [5].

We found out that the matchings between the pre-invasion and invasion communities based on highest-user-overlap transitions are also visible through the URL and hashtag similarities (see Figure 3). In fact, for each pre-invasion community, its respective highest-user-overlap invasion community is also the highest RBO pair for both URLs and hashtags. In other words, there is a strong positive correlation between the user transition percentages (Figure 2) and the RBO scores. E.g., 68% of the users from the pre-invasion "RS populist coalition" community transition in the "RS populist coalition" community in the invasion network. Meanwhile, The URL RBO of this pair is 0.64, while the hashtag RBO is 0.43, both as the highest combination for the pre-invasion "RS populist coalition" community, clearly matching it with its invasion transition-based counterpart. This shows that our simple similarity method based on URLs and hashtags can even help in better matching communities in the task of community evolution [6].

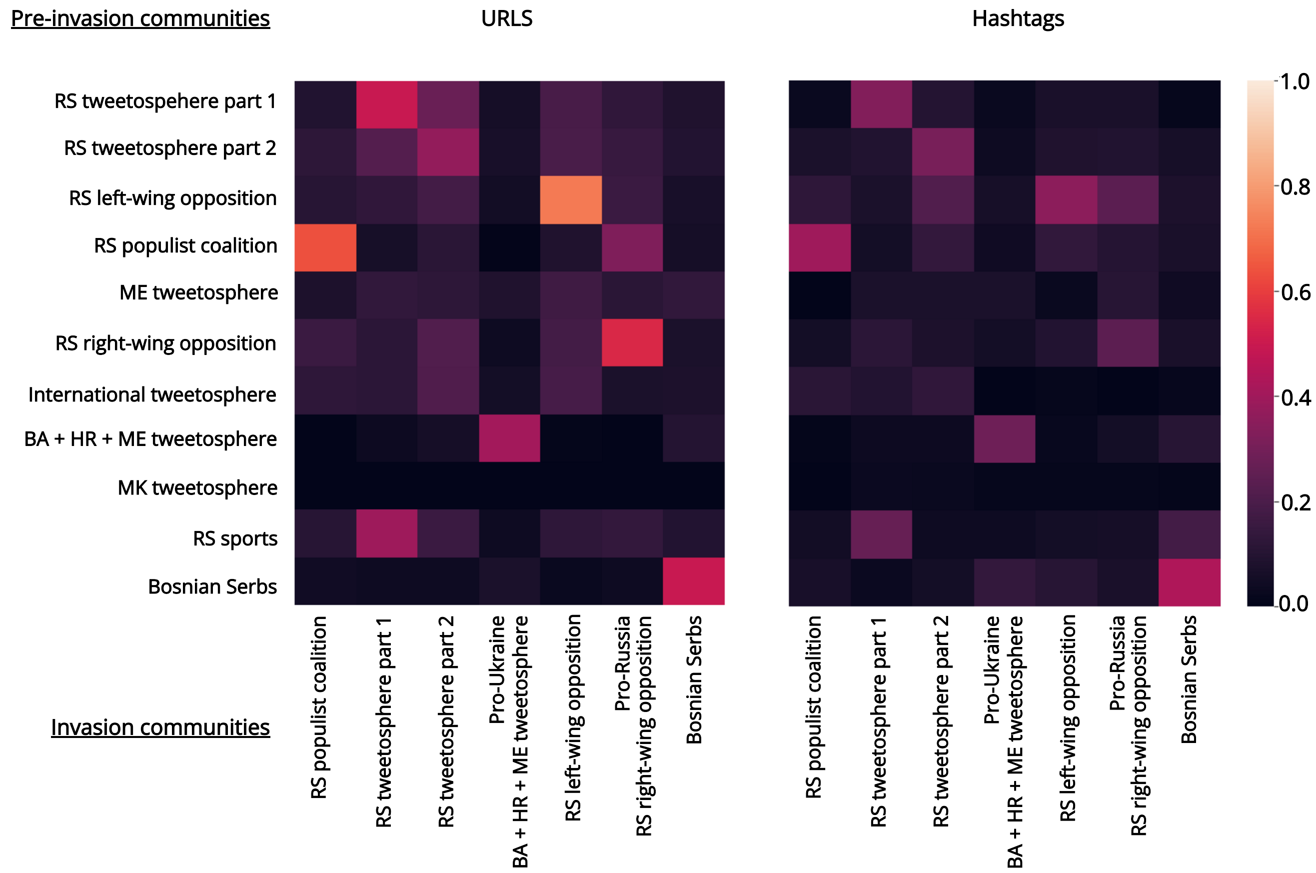

Figure 3: Domain and hashtag community similarities. A heatmap showing the similarities between the pre-invasion and invasion network communities based on the top 50 URLs (left) and hashtags (right). Similarities are calculated using the Rank-biased overlap (RBO) measure for indefinite rankings [5]**.**

3 CONCLUSION

In this work, we investigated the Russian invasion of Ukraine through the lens of Twitter in the ex-Yugoslavian region where Bosnian, Croatian, Montenegrin and Serbian are spoken. We analyzed 12 weeks of Twitter activities in this region, six weeks before the invasion, and six weeks after the start of the invasion. For each period, we created retweet networks and detected retweet communities. We followed the transition of users from the pre-invasion to the invasion period and analyzed these groups of like-minded Twitter users, discovering that they are distinctly divided across countries and political orientations. For the invasion network, we were also able to detect communities which show clear pro-Ukrainian and pro-Russian stance.

Another contribution was a simple method for comparing retweet network communities based on the content of the tweets. The method showed a strong correlation with the most prominent user transitions we formerly discovered.

A continuation of this work is to expand it to a multidisciplinary research, with the aim to meticulously analyze the polarized content between the communities in collaboration with domain experts who are knowledgeable in ex-Yugoslavian politics. Beyond obtaining interesting insights, we also aim to explore two frequent issues in using social media for societal analyses: (1) uptake bias of specific social networks across countries and communities, and (2) entanglement of the main event with other large-scale events.

ACKNOWLEDGMENTS

The authors acknowledge financial support of the Slovenian Research Agency (research core funding no. P2-103 and no. P6-0411).

REFERENCES

[1] B. Evkoski, I. Mozetič, P. Kralj Novak. Community evolution with Ensemble Louvain. In *10th Intl. Conf. on Complex Networks and their Applications*, Book of abstracts, pp. 58–60, Madrid, Spain, 2021.

[2] B. Evkoski, I. Mozetič, N. Ljubešić, and P. Kralj Novak. Community evolution in retweet networks. *PLoS ONE*, 16(9):e0256175, 2021. Non-anonymized version available at https://arxiv.org/abs/2105.06214.

[3] N. Ljubešić, D. Fišer, and T. Erjavec. TweetCaT: a tool for building Twitter corpora of smaller languages. In *Proc. 9th Intl. Conf. on Language Resources and Evaluation*, pp. 2279–2283, ELRA, Reykjavik, Iceland, 2014.

[4] N. Ljubešić, M. Miličević Petrović, and T. Samardžić. Borders and boundaries in Bosnian, Croatian, Montenegrin and Serbian: Twitter data to the rescue. Journal of Linguistic Geography 6:2, DOI 10.1017/jlg.2018.9, pp 100-124, Cambridge University Press, 2018

[5] W. Webber, A. Moffat, Alistair, and J. Zobel. A similarity measure for indefinite rankings. ACM Trans. Information Systems 28(4):20, 2010.

[6] G. Rossetti and R. Caxabet. Community discovery in dynamic networks: a survey. ACM computing surveys (CSUR) 51.2 (2018): 1-37.