Narratives of Foreign Media Ecosystems in Chinese Social Media Discussions of the Russo-Ukrainian War

Hans W. A. Hanley [1]   Yingdan Lu [2]   Jennifer Pan [2]

[1]Stanford University Dept. of Computer Science   [2]Stanford University Dept. of Communication

Overview

- How do ideas and narratives flow across national boundaries and between different media ecosystems?
- We identify narratives about the Russo-Ukraine War shared on Chinese social media.
- We link narratives to those originating in the Russian, Ukrainian, US, and Chinese media ecosystems.
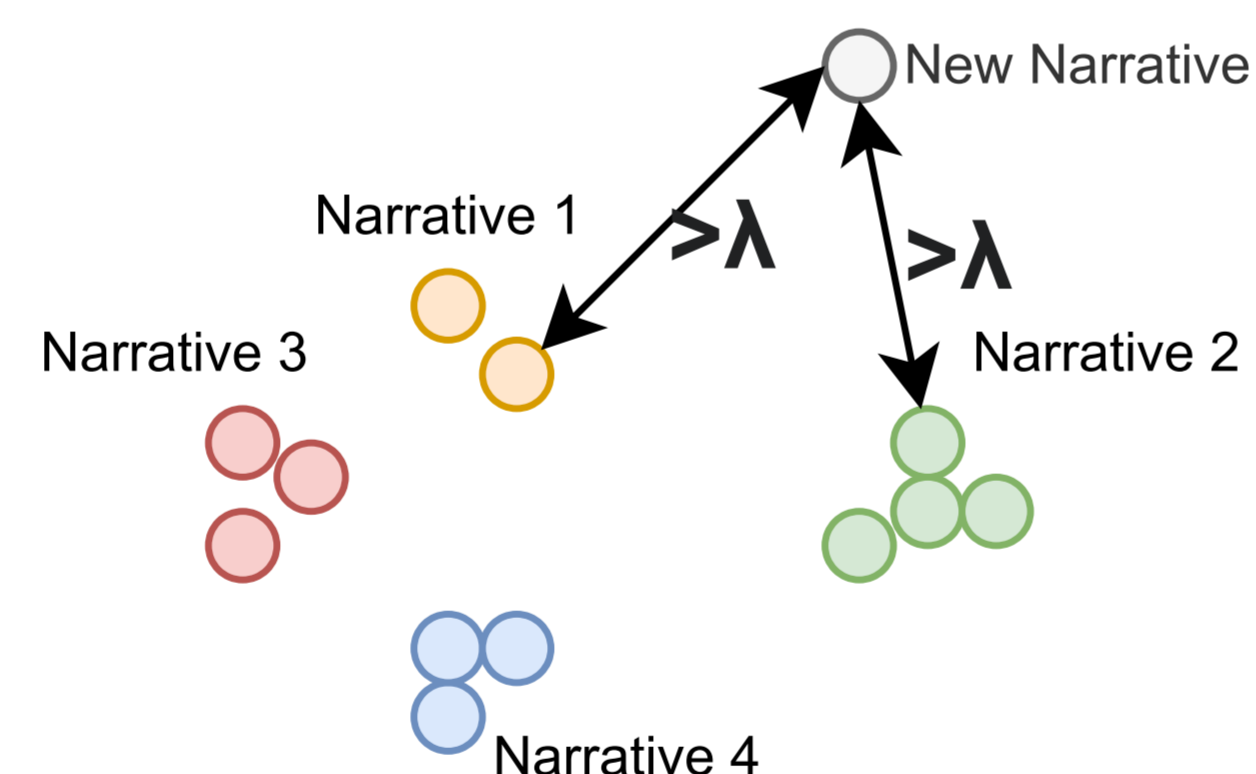
Media Ecosystems Dataset

- Utilize mixture of web crawling techniques
- Subset Alexa top million websites to news websites (from Cloudflare labels) from 4 countries (based on domain registrations)
- 14.4 million articles published Jan. 1 - Apr. 30, 2022 from 3,064 Chinese, 2,844 Ukrainian, 3,954 Russian, 5,839 US news websites

Chinese Social Media Dataset (Weibo)

- Scrape Weibo using keyword-based advanced search ("Ukraine," "Ukrainian" and variants)
- Collect 475,231 posts by 160,792 users published Feb. 1 - Apr. 19, 2022

Identifying Narratives on Chinese Social Media

1. Finetune monolingual (Mandarin) BERT-based model with contrastive learning loss

2. Employ optimized version of DPMeans algorithm to identify narrative clusters

- Embed Weibo posts into a shared vector subspace so posts with similar semantic meaning closer within subspace
- Assess model on 2022 ZH-SemEval Semantic Textual Similarity test dataset, achieving near state-of-the-art Pearson correlation of 0.82
- Non-parametric extension of K-means algorithm does not require *a priori* specification of number of clusters
- When datapoint is λ away from closest cluster, new cluster formed and datapoint assigned to it
- λ = 0.9 set by human validation of cluster coherence
- Cluster embeddings based on cosine similarity
- >70% clusters with texts that refer to singular, identifiable narratives (validation by native-Mandarin speakers)

Finding Narratives in Media Ecosystems Dataset
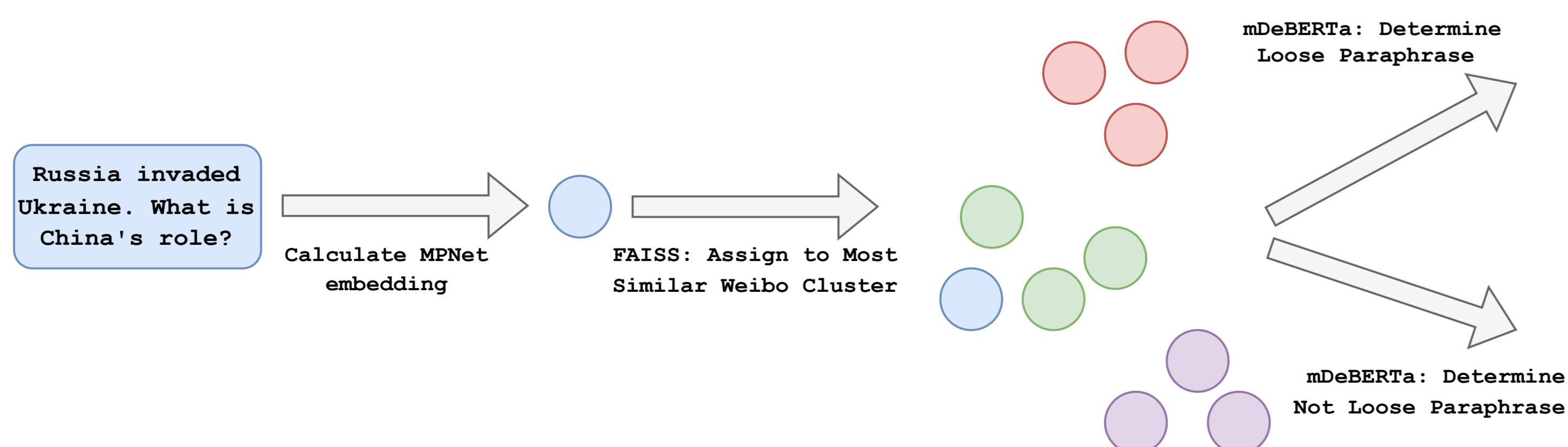
mDeBERTa: trained three models (Russian-Chinese, Ukrainian-Chinese, English-Chinese) to identify paraphrased content between language pairs (~95% accuracy)

Narratives on Chinese Social Media Primarily Sympathetic toward Russia

The situation in Ukraine continues to deteriorate. Ukraine is a pawn used by the US to disrupt Europe and allow capital to flow back and continue competing with China.

Ukrainian President Vladimir Putin said that if Ukraine does not join NATO, Russia will eventually withdraw

Ukraine is really shameless, the United States treats you like a dog, and you treat the United States as your big brother.

After the Ukrainian government announced the registration and evacuation of overseas Chinese, Chinese citizens in Ukraine should not disclose their identities at will.

The situation in Ukraine has sparked heated discussions in the international media.

The Russian army entered Ukraine and just announced that it had seized the biological laboratory set up by the United States in Ukraine.

Russian President Vladimir Putin said that he always believed Russia and Ukraine are the same nation, which clearly signals Russia's future plans.

On Chinese social media, make no mistake, Chinese netizens are saying 'never make fun of the war'.

The press secretary of the Ukrainian president said that the Ukrainian side had rejected the peace talks, sparking heated discussions among netizens.

What happened to the Ukrainian confrontation?

Note: Narrative summaries generated by mT5 multilingual XLSum conditional text generator
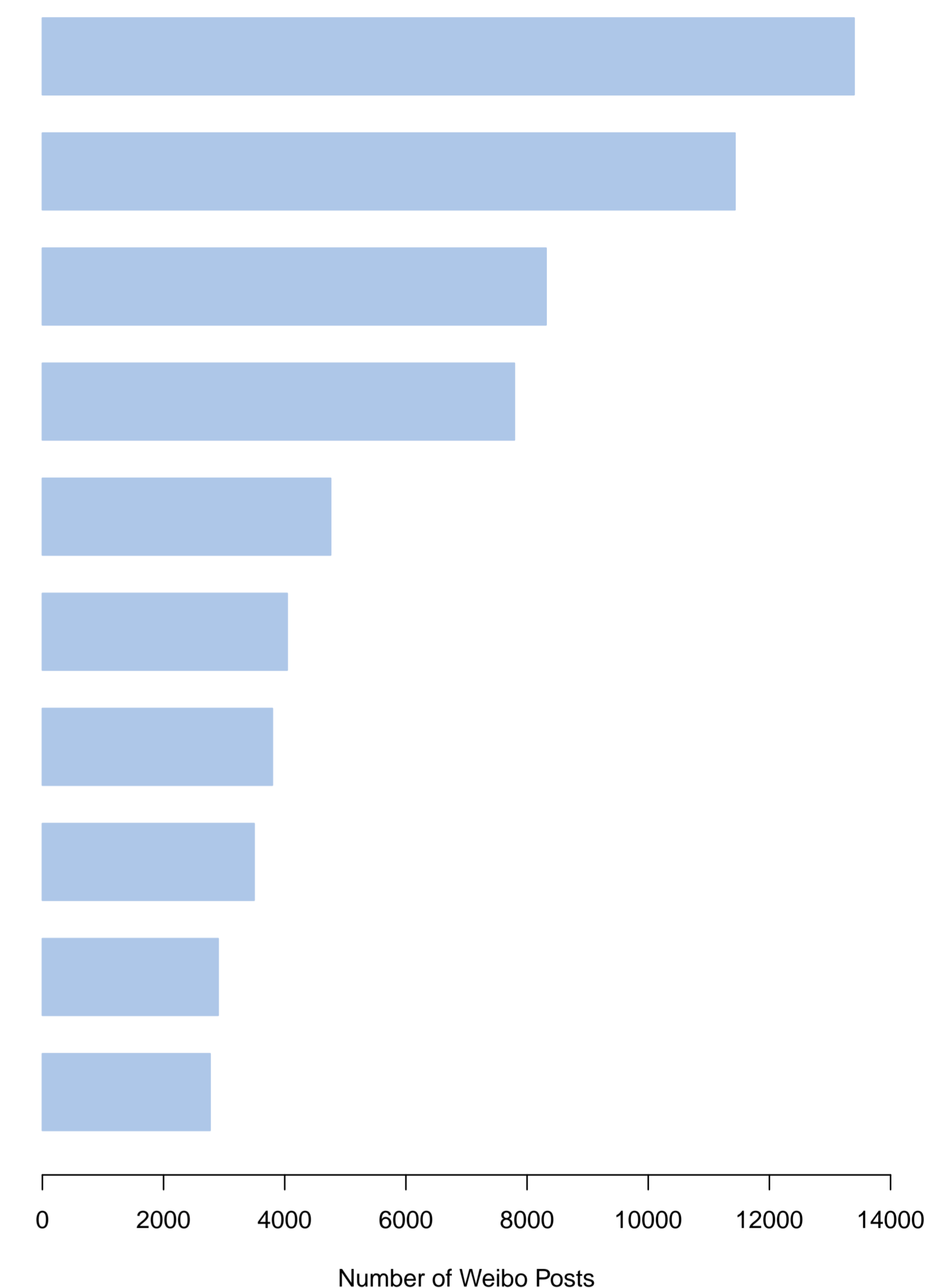
Largest Share of Narratives on Chinese Social Media First Appeared in Russia Media Ecosystem
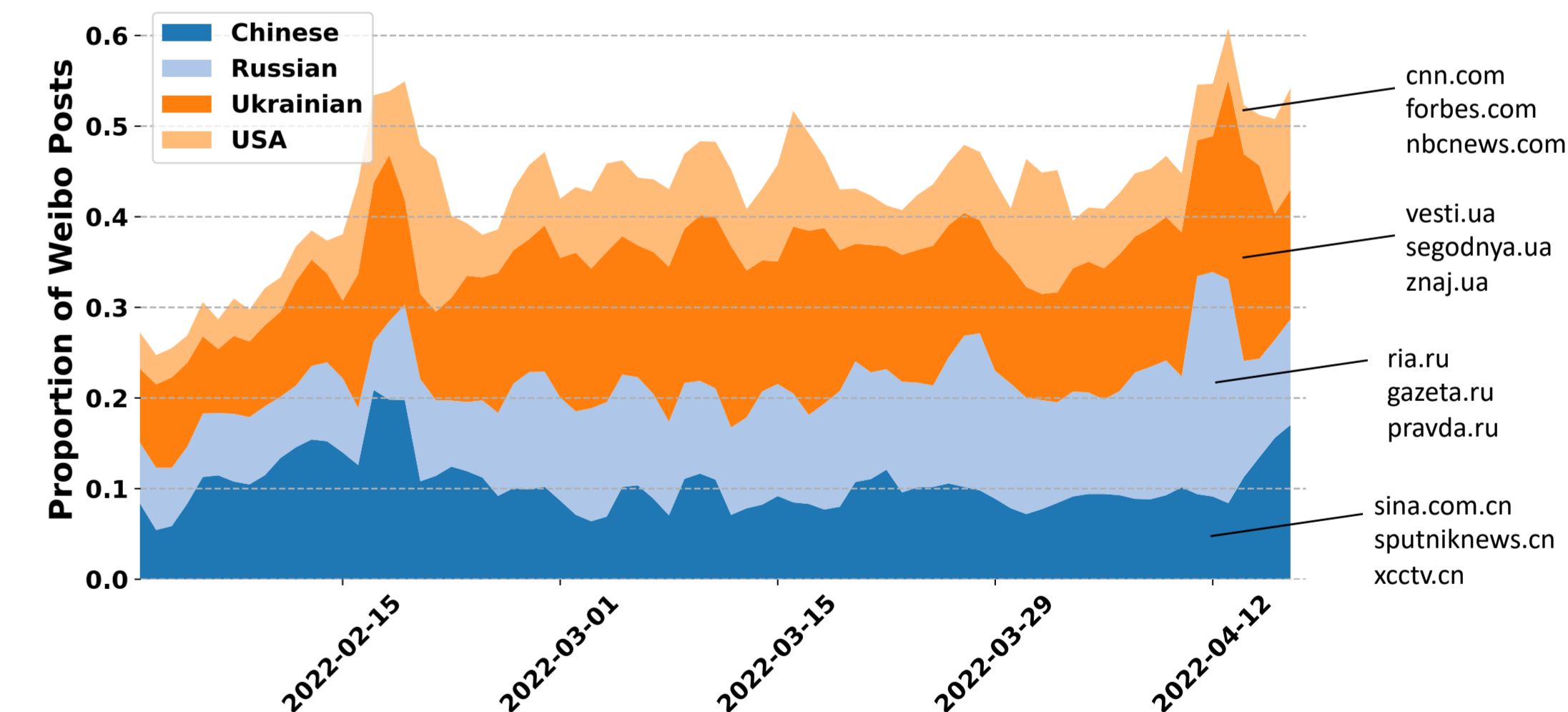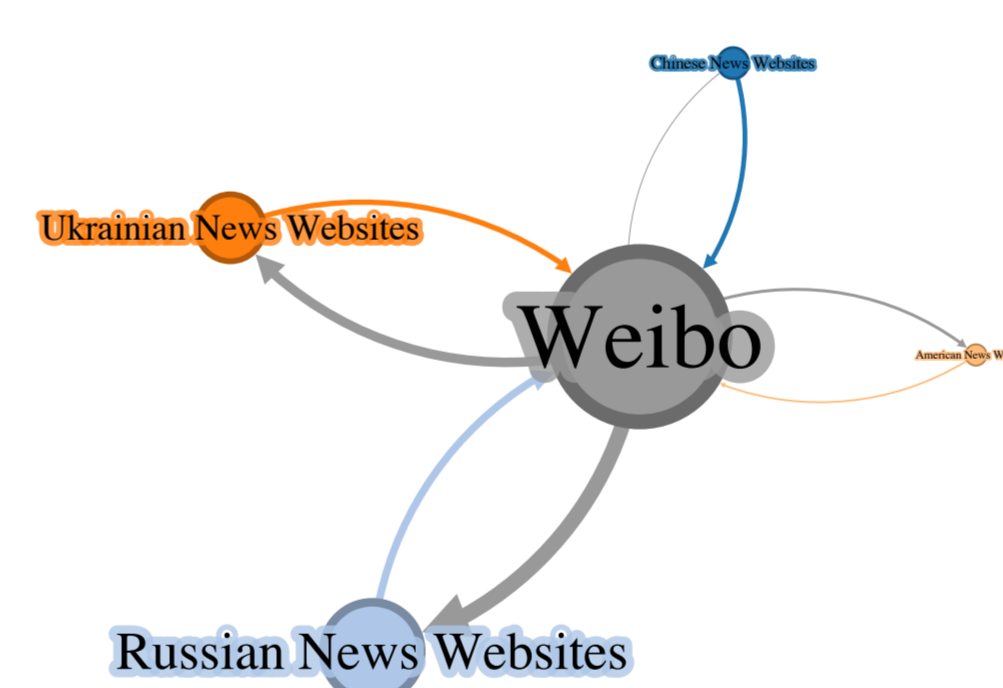
- Arrows represent number of articles / Weibo posts that first appeared on respective platform

- Media ecosystem where Weibo posts originally appeared (by proportion of Weibo posts over time)
- Listed domains are those with the most narratives subsequently found on Weibo

References

Dufour, Sébastien, ,Mohamed Mehdi Kandi, Karim Boutamine, Camille Gosse, Mokhtar Boumedyen Billami, Christophe Bortolaso, and Youssef Miloudi. "BL. Research at SemEval-2022 Task 8: Using various Semantic Information to evaluate document-level Semantic Textual Similarity." *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022).* 2022

Gao, Tianyu, Xingcheng Yao, and Danqi Chen. "SimCSE: Simple Contrastive Learning of Sentence Embeddings." *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing.* 2021.

Hanley, Hans WA, and Zakir Durumeric. "Partial Mobilization: Tracking Multilingual Information Flows Amongst Russian Media Outlets and Telegram." arXiv preprint arXiv:2301.10856. 2023.

He, Pengcheng, Jianfeng Gao , and Weizhu Chen. "DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing." arXiv:2111.09543. 2021.

Song, Kaitao, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. "MPNet: Masked and permuted pre-training for language understanding. " *Adv. in Neural Information Processing Systems.* 2020.