Artificial Dissimilarity: Multi-Modal Content Similarities in Online Disinformation Campaigns

Cody Buntain, New Jersey Institute of Technology
Vishakh Padmakumar, New York University
Richard Bonneau, New York University
Jonathan Nagler, New York University
Joshua A. Tucker, New York University

1. INTRODUCTION

Disinformation campaigns are growing in frequency globally, with evidence of state-sponsored social media manipulation appearing across 70 countries in 2019 [Bradshaw and Howard 2019]. Studies of these disinformation campaigns reveal composite efforts that touch multiple online spaces, across Twitter, Facebook, YouTube, Reddit, fake websites, and podcasts [US Department of Justice 2018; Yin et al. 2018; Gadde and Roth 2018] and multiple modalities, from text to images to videos. These multifaceted strategies demonstrate the need for new approaches that look at online disinformation across multiple modalities and across the *broader information ecosystem*.

This abstract describes a study into a disinformation campaign's similarities in messaging across content types and social platforms. We compare information sharing behaviors of the Russian IRA troll efforts, looking for similarities in the text, hyperlinks, images, and YouTube channels they share. These different modalities can expose important factors of coordination, as evidence shows the Russian IRA has leveraged many more than just text as part of their campaign, with a heavy reliance on links to news and YouTube shares [Yin et al. 2018; Zannettou et al. 2020].

To evaluate these coordination dimensions, we leverage data published by Twitter's electoral integrity program [Gadde and Roth 2018] and by Reddit's 2017 transparency report [Huffman 2018]. Twitter's dataset contains 9,041,308 tweets by 3,667 distinct Russian IRA accounts, and Reddit's listing contains 12,718 posts by 937 accounts. For purposes of comparison, we also have constructed 5,000-account samples of random accounts and politically engaged accounts from Twitter and Reddit, resulting in approximately 17 million tweets and 350 thousand Reddit submissions.

Our findings identify some systematic consistencies in the IRA's messaging: First, the IRA exhibits a surprising dissimilarity in the text and images they share, as they appear significantly less similar compared to randomly selected accounts. This result suggests active attempts to appear uncoordinated, which manifests as inorganic dissimilarity, and may be unique to malevolent disinformation campaigns. While this dissimilarity results holds in the Russian IRA's link and YouTube channel sharing in Twitter, the IRA is significantly *more similar* to each other in terms of *news sharing*, consistent with a political orientation, as news domains have been shown to capture ideological information [Messing et al. 2017].

Taken together, these results can inform future work into characterizing malevolent information campaigns, as legitimate campaigns may similarly coordinate around a political message but may be less likely to exhibit the same dissimilarity shown by the Russian IRA.

2. RESULTS

Below, we present results for comparing Russian IRA accounts' sharing behaviors to our comparison groups across several modalities: text, links, images, and YouTube channels. In each case, we compare Russian IRA behavior to the two comparison groups, consisting of politically engaged users (i.e., users who followed at least five politicians) and randomly selected users. These comparisons quantify how similar accounts in the same group are to each other and an analysis of Russian IRA accounts compared to both comparisons groups. These latter comparisons give us insight into whether the Russian IRA behaves more like politically engaged accounts or random accounts. In all cases, we use Euclidean distance and use bootstrapping to construct the sample distribution.

Similarity in Text Figures 1a and 1a present the mean pairwise distances in terms of text sharing in Twitter and Reddit, respectively. In both platforms, we see this distance among Russian IRA accounts (the grey curve) is significantly higher compared to the other groups. More interestingly though, this mean pairwise distance among IRA accounts is significantly higher than what we see from two random accounts (the yellow curve), suggesting the IRA is artificially dissimilar, which would be consistent with an effort to appear uncoordinated. As one might expect for a politically motivated group, however, the Russian IRA is significantly more similar to political accounts (the green curve) than they are to random accounts (the red curve).

Similarity in Image Sharing Consistent with text sharing, Figures 1c and 1d present a space in which Russian IRA accounts are also significantly more divergent from each other than one would expect from a group of randomly selected accounts. As with text sharing, image sharing by politically engaged accounts is the most similar, and the IRA again shares content more similar to political accounts than to random accounts (also as one would expect from a politically motivated organization).

Similarity in Links Figures 1e and 1f show how these accounts compare in terms of the top 150 most popular web domains each group shares. For Twitter, as with text and image sharing, we see that within-group similarity for IRA accounts is much less similar to within-group similarity among the political or random groups, again suggesting the IRA's efforts to appear dissimilar manifest in link sharing. Unlike with text/image sharing, however, across-group similarity for IRA accounts show the IRA is more similar to each other than they are to political or random accounts, consistent with a coordinated sharing effort, in that they are pushing links to similar domains one might not see shared by other groups. Surprisingly, Reddit exhibits a wholly different structure, with the political group being the most divergent, and the Russian IRA being the most similar in the links they share.

While the above comparisons are based on the most popular domains shared, we also evaluate web domain sharing specific to news (Figures 1g and 1h). Here, we see accounts in the Russian IRA group are *much more* similar to each other than accounts from either political or random populations, suggesting the Russian IRA consistently shares links to similar news domains.

Similarity in YouTube Sharing As with message text, Figures 1i and 1j show YouTube channels IRA accounts share on Twitter are conspicuously dissimilar compared to political and random groups. This result does not hold in Reddit, however, with IRA YouTube sharing in Reddit falling between political and random groups (more consistent with an absence of active attempts to appear uncoordinated). In both platforms, however, Russian IRA accounts are more similar to the random group than the political group in the YouTube channels they share.

3. CONCLUSION

The comparisons we present show the IRA's messaging exhibit systematic patterns across several modalities *and* platforms. Across both Twitter and Reddit, political news sharing among the IRA show

evidence of coordination around a consistent set of information sources, whereas IRA text and image sharing are significantly dissimilar, which may be evidence of inauthentic manipulation. Future work will explore these sharing behaviors across multiple disinformation campaigns.
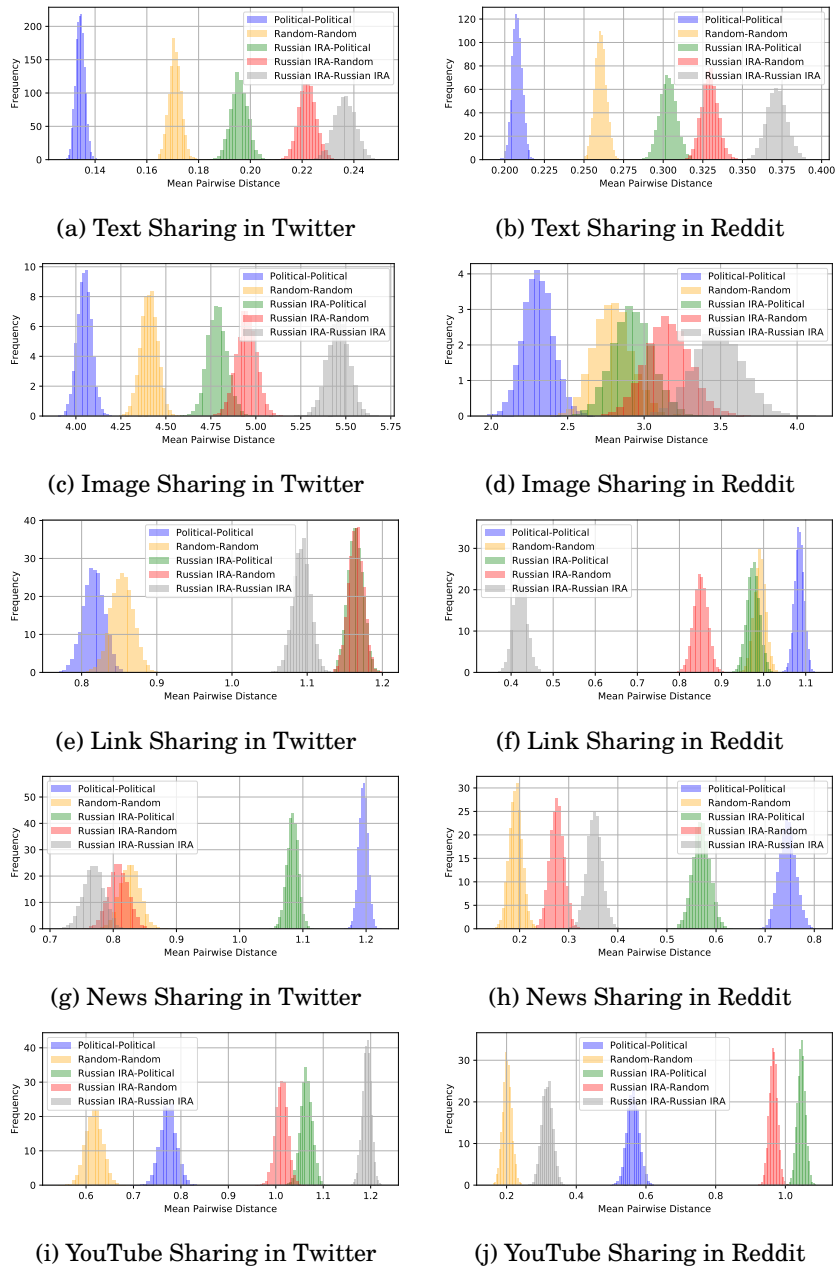

(a) Text Sharing in Twitter

(b) Text Sharing in Reddit

(c) Image Sharing in Twitter

(d) Image Sharing in Reddit

(e) Link Sharing in Twitter

(f) Link Sharing in Reddit

(g) News Sharing in Twitter

(h) News Sharing in Reddit

(i) YouTube Sharing in Twitter

(j) YouTube Sharing in Reddit

Fig. 1: Bootstrapped distributions of mean distances among groups across different contexts.

REFERENCES

Samantha Bradshaw and Philip N Howard. 2019. The Global Disinformation Order 2019 Global Inventory of Organised Social Media Manipulation. (2019), 25. https://comprop.oii.ox.ac.uk/research/cybertroops2019/

Vijaya Gadde and Yoel Roth. 2018. Enabling further research of information operations on Twitter. (2018). https://blog.twitter.com/en{_}us/

Steve Huffman. 2018. *Reddit's 2017 transparency report and suspect account findings : announcements*. Technical Report. Reddit. https://www.reddit.com/r/announcements/comments/8bb85p/reddits

Solomon Messing, Van Kessel Kessel, and Adam Hughes. 2017. *Sharing the News in a Polarized Congress*. Technical Report December. Pew Research Center.

US Department of Justice. 2018. *United States of America v. Internet Research Agency LLC*. Technical Report. US District Court for the District of Columbia, Washington, DC. https://www.justice.gov/file/1035477/download

Leon Yin, Franziska Roscher, Richard Bonneau, Jonathan Nagler, and Joshua A Tucker. 2018. *Your Friendly Neighborhood Troll : The Internet Research Agency ' s Use of Local and Fake News in the 2016 US Presidential Campaign*. Technical Report. Social Media and Political Participation Lab, NYU, New York, New York, USA.

Savvas Zannettou, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, Jeremy Blackburn, Tristan Caulfield, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Jeremy Blackburn. 2020. Characterizing the Use of Images by State-Sponsored Troll Accounts on Twitter. *ICWSM* (2020). http://arxiv.org/abs/1901.05997