

# “Pictures Tell a Thousand Words”: Using Clustering to Explore Trolls’ Visual Strategies

*Keywords: Russian Trolls, Social Media, Clustering, Politics, Misinformation*

## Extended Abstract

The prevalence of state-sponsored information operations has escalated, leveraging trolls to skew public opinion on social media, transforming these platforms into arenas for information warfare. This phenomenon, particularly evidenced by the Russian Internet Research Agency’s (IRA) interference in the 2016 US election (Badawy et al., 2018; Im et al., 2020), underscores the significant challenge of combating online trolls. Despite extensive research into trolls’ behavioral and textual tactics, their use of visual content remains underexplored. Our study aims to bridge this gap by analyzing the image-posting behaviors of IRA-associated trolls, focusing on content themes, temporal patterns, and audience engagement.

We asked the following questions:

**RQ1:** What categories of images were commonly posted by IRA trolls?

**RQ2:** How did the temporal and thematic strategies of IRA trolls’ image posts evolve, particularly around the 2016 U.S. presidential election?

**RQ3:** What is the relationship between image themes and audience engagement (likes, retweets, comments) with IRA trolls’ posts?

To address RQ1, we calculated the frequency of images across these themes. For RQ2, we investigated temporal trends in posting, especially in relation to key political events. RQ3 is explored through regression analysis to understand how different themes influence user engagement metrics. This comprehensive approach aims to shed light on the tactics of IRA trolls in manipulating social media discourse and engaging audiences.

## Data and Approaches

In this study, data on troll accounts and their image postings were extracted from a dataset released by Twitter in 2018, encompassing tweets from August 2011 to April 2018, including engagement metrics. The focus was on images from IRA accounts, resulting in an analysis of 3,667 accounts and over 9 million tweets, with 1,788,068 images analyzed from 3,278 accounts that posted images. A clustering analysis was conducted using the VGG19 network and k-means algorithm to categorize images into 100 themes based on their content. The study further examined the temporal distribution of these image themes around the 2016 U.S. presidential election and their impact on user engagement through a multivariate linear regression model, taking into account various control variables such as the account’s follower count and the nature of the post. This comprehensive approach aimed to shed light on the strategies used by trolls in manipulating public opinion through visual content on social media platforms.

## Results

In the study’s results section, the clustering analysis confirmed the validity of the findings by demonstrating consistency across two separate runs, leading to the identification of 40 distinct

themes from the original 100 clusters. This classification was achieved through a meticulous manual review process, as illustrated in Figures 1 and Figure 2, which display examples of the clustering outcomes. The frequency of images within each theme was quantified to address the first research question, with the findings detailed in Table 1, revealing a significant leaning towards apolitical content—over 70% of the images were non-political, encompassing themes like daily life, city views, and nature scenes.

Temporal pattern analysis aimed to unravel the dynamics of image posting around the 2016 US presidential election. Figures 3 - Figure 8 depict various facets of this analysis, including the monthly, daily, and hourly posting frequencies and the distribution of political versus apolitical images over time. A notable increase in posting activity was observed around key election milestones, with a peak following the first debate and the election results announcement. This trend suggests a strategic deployment of images to influence public opinion at critical junctures. The analysis also revealed a distinct pattern in the types of images posted before versus after the election, indicating a tactical shift from fearmongering to mockery of political figures.

Engagement analysis, detailed through regression models, sought to understand the interaction between image themes and user engagement metrics such as likes, comments, and retweets. In Figure 9 - Figure 11, the differentiation between low-cost (likes and retweets) and high-cost (replies) engagement forms provided insights into how users interact with troll-posted content. Despite the overwhelming presence of apolitical images, political themes, especially those related to military and war, elicited higher levels of high-cost engagement, as shown by the increased tendency of users to reply to these images. This distinction underscores the nuanced impact of trolls' visual content on user engagement and interaction patterns on social media.

In conclusion, the study's results elucidate the sophisticated strategies employed by trolls, particularly in the context of the IRA, to blend into and manipulate social media discourse. Through a combination of apolitical and political imagery, trolls aimed to polarize public opinion and engage users in specific narratives, particularly around contentious topics. The temporal and engagement analyses reveal a calculated adjustment in tactics before and after the election, aimed at maximizing influence and interaction. These findings, supported by visualizations in Figures 1 through 8 and detailed in Table 1, offer a comprehensive overview of the troll's visual strategy and its implications for social media dynamics.

## References

- Badawy, A., Ferrara, E., & Lerman, K. (2018, August). Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. In 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM) (pp. 258-265). IEEE.
- Im, J., Chandrasekharan, E., Sargent, J., Lighthammer, P., Denby, T., Bhargava, A., ... & Gilbert, E. (2020, July). Still out there: Modeling and identifying russian troll accounts on twitter. In Proceedings of the 12th ACM conference on web Science (pp. 1-10).

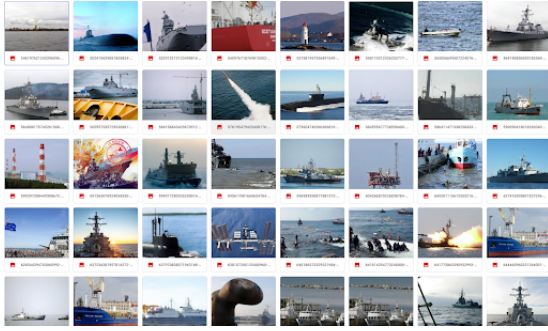


Figure 1: Cluster example of warships



Figure 2: Cluster example of political figures

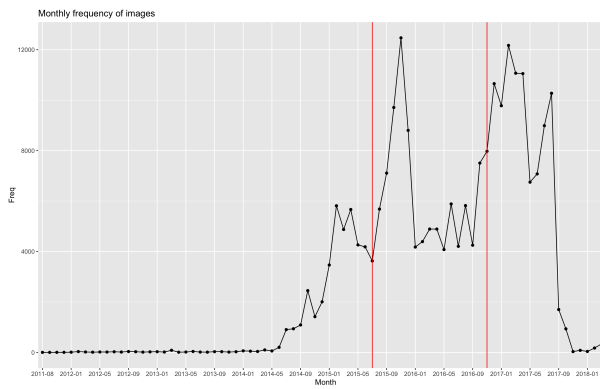


Figure 3: Total monthly frequency of images

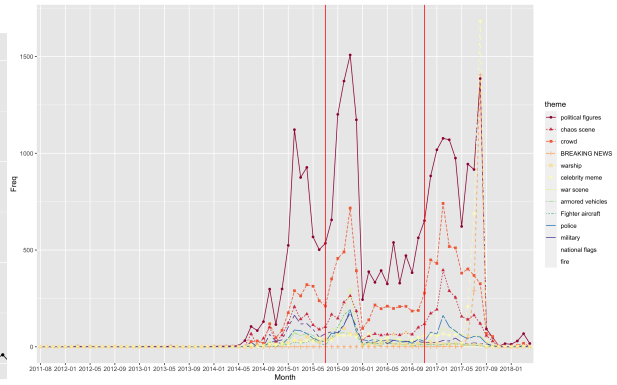


Figure 4: Monthly frequency of images by political theme

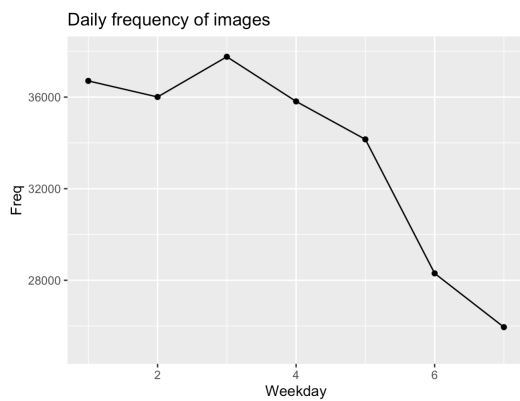


Figure 5: Total daily frequency of images

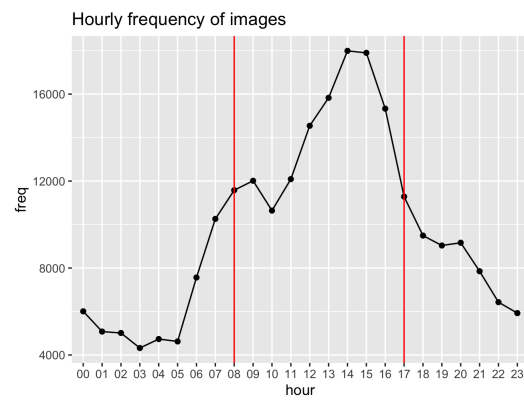


Figure 6: Total hourly frequency of images

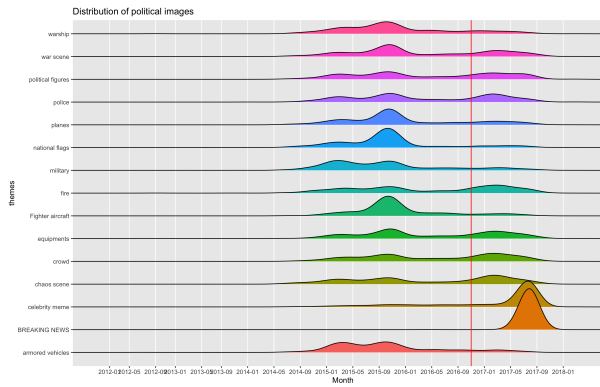


Figure 7: Distribution of political images

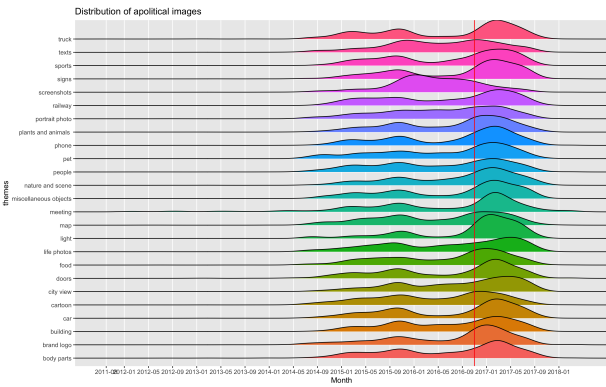


Figure 8: Distribution of apolitical images

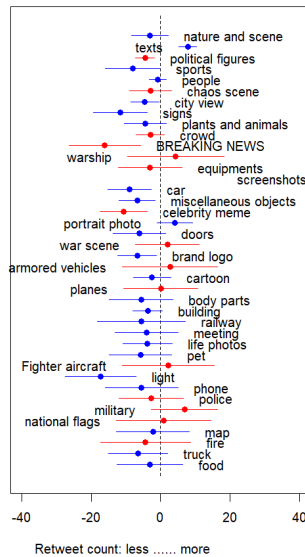


Figure 9: Regression results using # of retweet

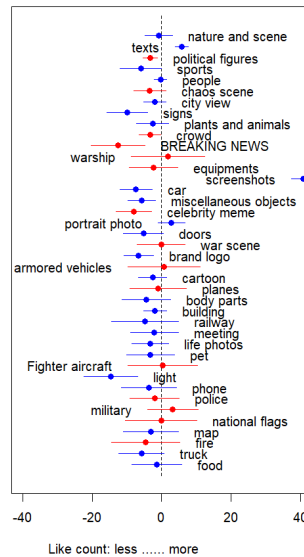


Figure 10: Regression results using # of like

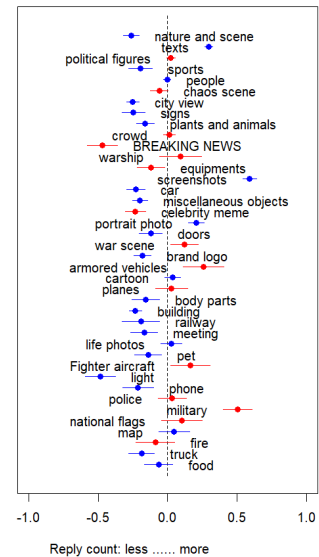


Figure 11: Regression results using # of reply

Theme	Percentage(%)	Category
People	15.66	Apolitical
Texts	14.03	Apolitical
Political Figures	11.03	Political
City Views	4.85	Apolitical
Crowd	4.67	Political
Buildings	4.41	Apolitical
Screenshots	4.01	Apolitical
Portrait Photos	2.99	Apolitical
Miscellaneous Objects	2.90	Apolitical
Nature and Scenes	2.80	Apolitical
Cartoons	2.65	Apolitical
Brand Logos	2.48	Apolitical
Chaotic Scenes	2.13	Political
Plants and Animals	2.10	Apolitical
Cars	2.01	Apolitical
Celebrity Memes	1.70	Political
Life Photos	1.54	Apolitical
Doors	1.31	Apolitical
Signs	1.30	Apolitical
Sports	1.21	Apolitical
Trucks	1.03	Apolitical
Pets	0.93	Apolitical
Meeting	0.92	Apolitical
War Scene	0.91	Political
Body Parts	0.9	Apolitical
Equipments	0.89	Political
Police	0.88	Political
Food	0.85	Apolitical
Military	0.83	Political
Breaking News	0.74	Political
Light	0.72	Apolitical
Phone	0.70	Apolitical
Planes	0.66	Political
Railway	0.47	Political
Fire	0.45	Political
Fighter Aircraft	0.43	Political
National Flags	0.41	Political
Armored Vehicles	0.40	Political
Warship	0.39	Political

Table 1: Image percentages by theme