BLOGS, FAKE NEWS, AND INFORMATION ACTIVITIES

Chapter · November 2017

CITATIONS
2

READS
2,948

2 authors:

Nitin Agarwal
University of Arkansas at Little Rock
253 PUBLICATIONS **2,831** CITATIONS
SEE PROFILE

Kiran Kumar Bandeli
Walmart Inc.
10 PUBLICATIONS **56** CITATIONS
SEE PROFILE

2

BLOGS, FAKE NEWS, AND INFORMATION ACTIVITIES

Nitin Agarwal, Kiran Kumar Bandeli

Blogs provide fertile ground for framing narratives. This chapter demonstrates that aside from the blog post itself, reader comments can make the narrative more persuasive. However, the absence of a social network structure for blogs inhibits the dissemination of these narratives. Social media platforms such as Twitter and Facebook are used as vehicles to disseminate the content using cross-media and mixed-media tactics. The link between blogs and social media platforms is vital for understanding contemporary disinformation campaigns.

INTRODUCTION

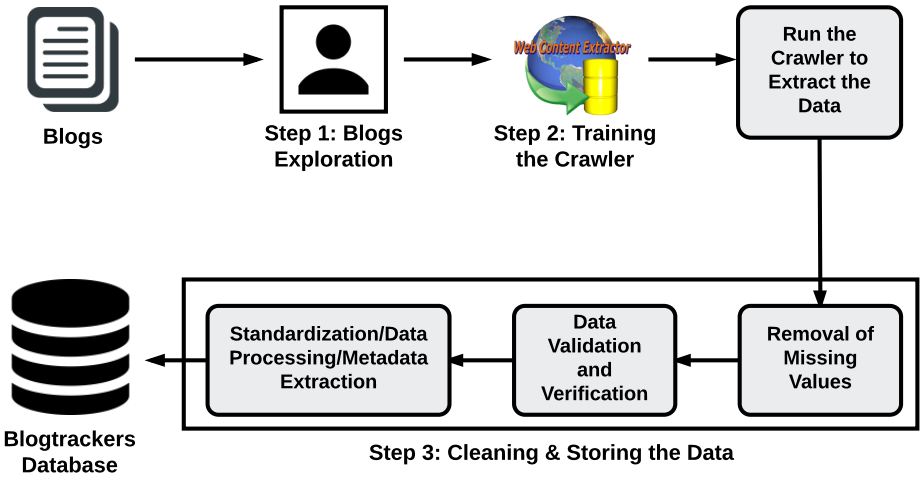
Blogs have ushered in an era of citizen journalism that has irreversibly changed the way we consume information, partly supplanting traditional journalism. Blogs have endowed citizens with the power and freedom to express their opinions or frame narratives for a greater audience; readers' comments on blogs afford greater inclusiveness and dialog. Blogs cater to the needs of the public to receive information in manageable chunks, tailored to their individual preferences. They can provide intimate details and live accounts featuring compelling, on-the-ground-style coverage of an event. Together, these two capabilities—news chunking and first-person reporting—can create the capability to orchestrate highly biased, partial, and distorted information, i.e. an information campaign.

Blogs alone are not effective in conducting information campaigns. Blogs provide fertile ground for framing narratives, but the absence of a social network structure inhibits dissemination. Various social media platforms, such as Twitter, Facebook, and VK, are then used as vehicles to disseminate the content. Nine out of ten bloggers have Facebook accounts. 78% of bloggers use Twitter to promote their content. This percentage is higher, i.e. almost 90% for professional and full-time bloggers.[137] In addition to bloggers promoting their content, studies have widely reported the exploitation of computer programs,[138] also known as social bots, to massively amplify content dissemination via Twitter. The ability to embed YouTube videos, SoundCloud files, and Internet-based memes in blogs has led to unprecedented convenience in framing narratives, disseminating them widely, and driving online traffic to generate a rich conversation around a chosen topic. In addition to content promotion, prolific media integration helps boost search rankings artificially—a technique known as *link farming*, which is a well-known strategy for search engine optimization. Gaming search engines by using prolific linking to blogs has become part of modern information activity. By further examining the information flows within the media networks, we attempt to understand the sources of mis-/disinformation and their reach; if we can detect how far and how quickly the mis-/disinformation can travel, we can also understand the extent to which information is being manipulated. This chapter will present an in-depth examination of the social media networks using a social-network-analysis-based methodology to identify the prominent information brokers and leading coordinators of disinformation campaigns. A methodological section will describe how the data is fetched from different sources, and the approach we propose for studying information flows. The analysis and findings below provide a deep dive into the research questions we set out to answer in this study.

METHODOLOGY

For the purposes of our analysis, we examined several blogs and identified common attributes among them, such as *title, date and time of posting, author/*

THE DATA COLLECTION PROCESS FOR BLOGS



blogger, blog post content, comments, and *permalinks*. We collected and indexed all blog content from four different blog datasets into our Blogtrackers database. The database can be accessed at http://blogtrackers.host.ualr.edu/. The dataset consists of 372 blog sites, 7576 bloggers, and 196940 blog posts riddled with false and misleading information. To crawl these blogs from different sources, we setup crawler(s) for each blog to extract all the required attributes. There are three main steps in crawling data from a blog: (1) exploring the blog site, (2) crawling the blog site, and (3) cleaning and storing the data in a database for analysis and retrieval. Figure above represents the data crawling process for the blogs.[139]

For this study, data was collected from four diverse sources. The descriptions associated with the attributes in these four types of datasets are as follows:

- *Fake news dataset from* kaggle.com. This dataset has 244 blogs, 2236 bloggers, 12,999 posts, and 20 attributes. Some of the key attributes in this dataset are: domain name, site_url, author, post title, text, published date, language, comments, replies_count, shares, and likes. The dataset is available at https://www.kaggle.com/mrisdal/fake-news.
- *Dr. Melissa Zimdars' compiled list of fake news blogs.* Dr. Melissa Zimdars, a professor from Merrimack College (http://bit.ly/2wTMlUb), compiled blogs featuring fake news. These blog sites are available at http://bit.ly/2ezvFbV. This dataset has 37 blogs, 971 bloggers, 96,056 posts, and 79 attributes. The key attributes are: blog name,

blogger, blog post title, blog post, posting date, location, and language.

- *Blogs containing disinformation regarding the Baltic States.* This dataset has 21 blogs, 728 bloggers, 16,667 posts, and 79 attributes. The key attributes are: blog name, blogger, blog post title, blog post, posting date, location, and language.
- *Blogs containing disinformation regarding NATO exercises/ activities.* This dataset contains blogs collected by the Blogtrackers tool that posted mis-/disinformation during various exercises conducted by NATO, such as the Trident Juncture 2015, Brilliant Jump 2016, and Anakonda 2016. This dataset has 70 blogs, 3641 bloggers, 71,218 posts, and 79 attributes. The key attributes are: blog name, blogger, blog post title, blog post, posting date, location, and language.

The characteristics of these four datasets are presented in Table below. Next we present the research methodology used to analyse these blogs in order to examine the spread of disinformation.

In this study, we plan to answer the following research questions:

- What are the typical characteristics of mis-/ disinformation-riddled blogs?
- Can we track the origins of the content, such as memes, images, etc., appearing in these blogs?
- What strategies are common in disseminating the content (e.g. mixed-media and cross-media)? And, can we identify the other media sites that are predominantly used to disseminate the original blog posts?
- How do antagonistic narratives travel?

Dataset	Number of Blogs	Bloggers	Number of Posts	Attributes
Fake news from Kaggle.com	244	2236	12,999	20
Prof. Melissa Zimdars' compiled fake news blogs	37	971	96,056	79
Blogs containing disinformation regarding the Baltic States	21	728	16,667	79
Blogs containing disinformation regarding NATO exercises/ activities	70	3641	71218	79

TYPICAL CHARACTERISTICS OF DISINFORMATION-RIDDLED BLOGS

What are the typical characteristics of mis-/disinformation-riddled blogs? Based on our observations and the work of other experts, we provide a set of heuristics to identify blogs that are potentially riddled with mis-/disinformation.[140] These heuristics are:

1. **Pay attention to the 'contact us' section of the page to validate and verify site authors.** The contact information sections of these blogs do not provide real contact information for the author. For instance, one such real-looking contact URL is http://abcnews.com.co/.
2. **Do not read just the headline; instead, skim the body content** to familiarize yourself with the details of the story. For example, the headline '*Obama Signs Executive Order Declaring Investigation into Election Results; Revote Planned for Dec. 19th – ABC News*' is a false story with a catchy headline. But, reading through the content will enable the reader better to evaluate the story.
3. **Pay close attention to the URLs, sources, images, and editorial standards of the writing.** For instance, the URL bloomberg.ma is used to imitate the well-known site bloomberg.com.
4. **Always crosscheck the story with fact-checking websites,** such as snopes.com, factcheck.org, mediabiasfactcheck.com, or politifact.com for the credibility of the story. For example, a blog post titled '*The Amish In America Commit Their Vote to Donald Trump; Mathematically Guaranteeing Him a Presidential Victory – ABC News*' is a fake story reported by the well-known fact checking website snopes.com.
5. **Search for the post in well-known search engines**, such as Google, Bing, Yahoo, etc., to see if the same post or content is repeated on other sites using mix/cross media approaches to disseminate the narrative. For instance, the blog post '*Obama Signs Executive Order Declaring Investigation into Election Results; Revote Planned for Dec. 19th – ABC News*' has been shared on many websites, indicating the use of a mixed-media approach.
6. **Check if the article has been previously published and if it is being reused to affect perceptions about an event.** For example, a blog post title 'Muslims BUSTED: They Stole Millions in Govt Benefits' published in 2016, contained an image that was reused from the year 2013.
7. **Check if the post is disturbing or controversial.**

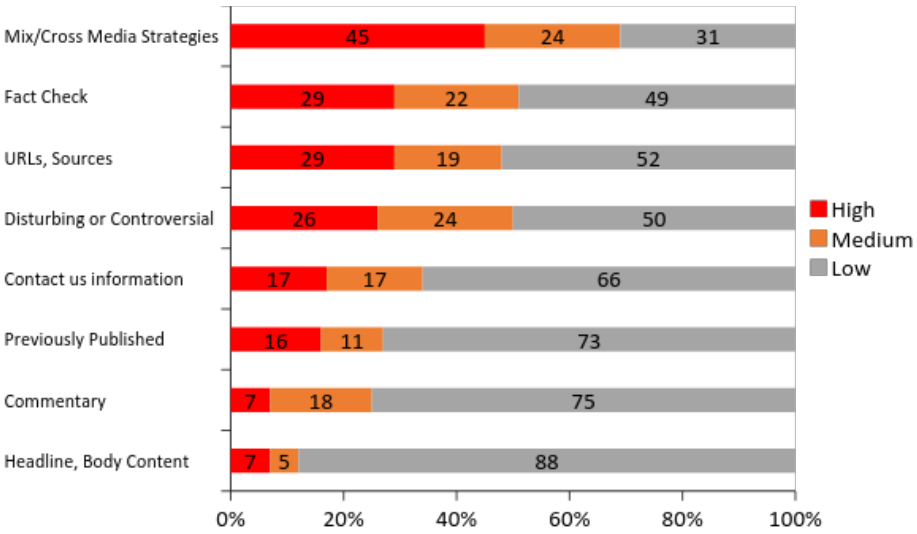
Fake stories usually appear under sensational headlines. For instance, the blog post titled '*EU NATO Commit Adultery, Prince Charles Saudi Trade & More*' presents disturbing information. Disinformation narratives are often embedded in such stories.

8. **Check if the post has any 'likes', 'replies', or 'comments'.** This will indicate how interested readers are in a given story, and whether they agree or disagree. The *sentiment* can be used to infer this. For example, a blog post titled '*NASA Confirms—Super Human Abilities Gained*' has a lot of comments, many of which were debunking the story.

To evaluate the efficacy of these eight criteria, we conducted a survey. We randomly selected 96 blogs featuring mis-/disinformation and asked survey participants to rate (low, medium, high) how effective each of the eight criteria was in assessing whether the blog site contained misleading or false information.

After collecting the survey data, we constructed a stacked bar for each of the criteria where the X-axis represents values (0%–100%) indicating participant

EFFECTIVENESS OF EACH OF THE 8 CRITERIA IN IDENTIFYING BLOGS CONTAINING MISINFORMATION OR DISINFORMATION.



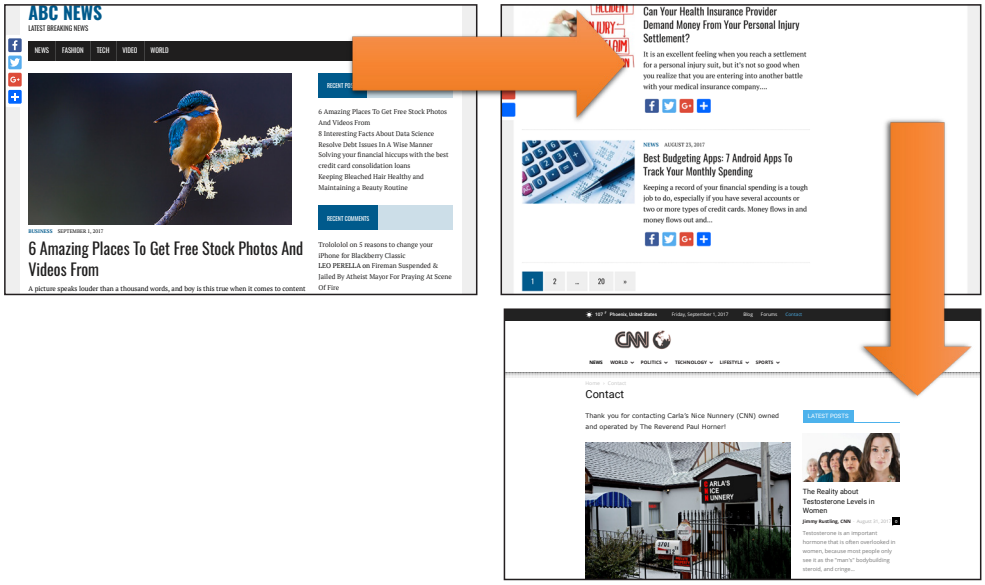
** The criteria are sorted in decreasing order of effectiveness. The smaller the gray bar the more effective the criterion is. Numbers on the colored bars indicate the number of blog sites identified as containing misleading or false information with a confidence of High, Medium, and Low.*

confidence in the 96 blogs rated low, medium, or high, and the Y-axis denotes the eight criteria. Looking at Figure on page 36, it is clear that the best criterion is the use of mix/cross media strategies by the blog site in disseminating the content. This can be used as the superlative feature for assessing the mis-/disinformation contained in any blog post. The next best feature is fact-checking websites.

Next, we present some empirical observations vis-à-vis mis-/disinformation heuristics on the fake news dataset collected from kaggle.com. Incidentally, most of the posts had very few comments or none, which might imply that the stories were mainly disseminated but not discussed much on these sites. We also found that during the US elections many posts were primarily intended to reach large numbers, draw their attention, and direct them to non-factual stories with the intention of influencing readers. For example, 96% (12,468 of 12,999) of the posts had zero 'likes' and 94% (12,304 of 12,999) of the posts had zero 'replies'. These posts were primarily intended to be disseminated to reach more people and mislead. We also observed that the majority of the stories originated from a set of domains that are usually reported as containing false information by snopes.com.

We further examined the website structure disseminating these false stories. We found, in many cases, that the 'contact us' page does not provide any real contact information or redirects readers to another website, usually a social media site, e.g. Facebook or Twitter. The example below illustrates how a

ILLUSTRATION: THE'CONTACT US' PAGE REDIRECTS TO ANOTHER WEBSITE

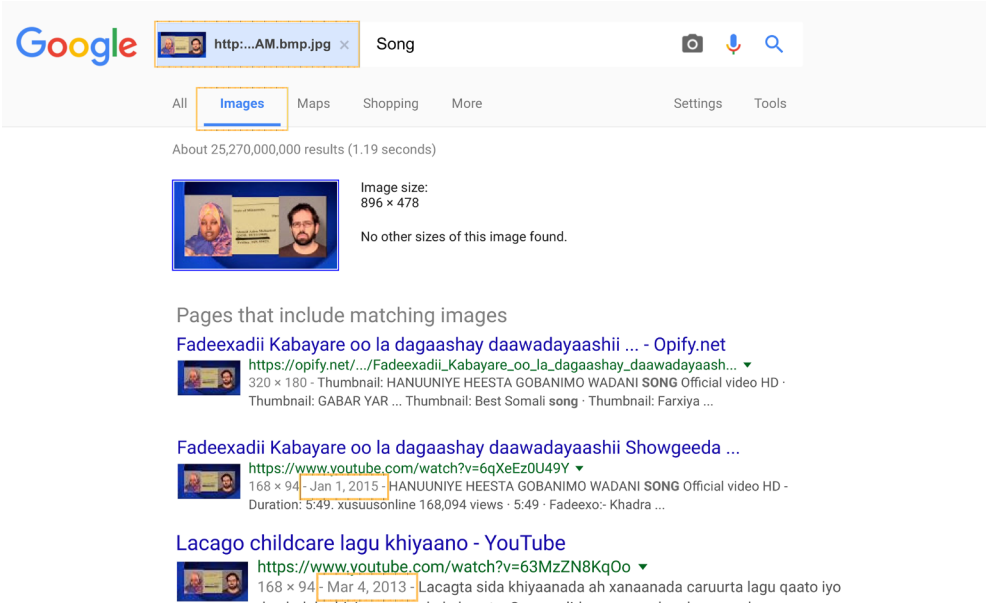


site redirects to another website when readers look for contact information for the author. The example provided here refers to the site name – ABC NEWS with the URL http://abcnews.com.co/. The contact information link is present at the bottom of the page for this site. If the reader clicks on 'contact us', he is redirected to another site named CNN with the URL http://cnn.com.de/contact/. The http://cnn.com.de/ website closely mimics the CNN News website (http://www.cnn.com/), even using the CNN logo, website structure, etc. However, cnn.com.de is riddled with false stories and conspiracy theories. When posted on Facebook, an article from cnn.com.de would bear the CNN logo and appear as if the article were actually published by the genuine CNN.com. This deception tactic is highly effective in disseminating disinformation originating on blogs via other social media channels.

TRACKING THE ORIGINS OF MISLEADING BLOG CONTENT

Can the origins of misleading content, such as memes, images, etc., which appear on these blogs, be tracked? We began our analysis with a 'reverse image search' (i.e. searching for the URL of a given image on Google Images to identify other sources that have used the image) and found that the images were not unique for each article and not relevant to the context they are used for. The same image was reused with different narratives, as shown below. Images lend credibility to a narrative

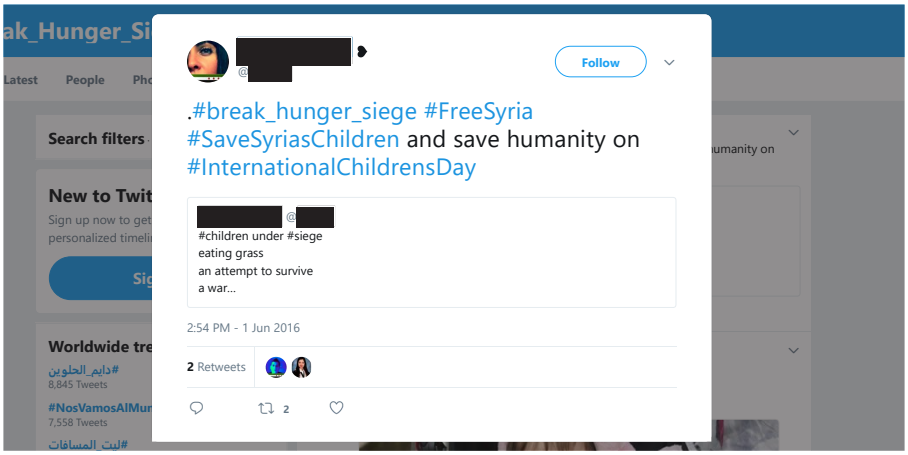
REVERSE IMAGE SEARCH SHOWS THE USE OF ONE IMAGE WITH DIFFERENT NARRATIVES



and are more effective than text alone for fabricating perceptions. The use of images and videos in framing narratives is effective because multiple modalities are exploited to influence thinking.[141] We also observed a pattern in which a post shared on Twitter was actually linked to a blog post using hashtags and links. This pattern is common across various social media channels, i.e. the origin of the content is generated on a blog and later disseminated

BLOG POST USES HASHTAGS AND LINKS TO REFER TO TWITTER

TWEET USES HASHTAGS/LINKS



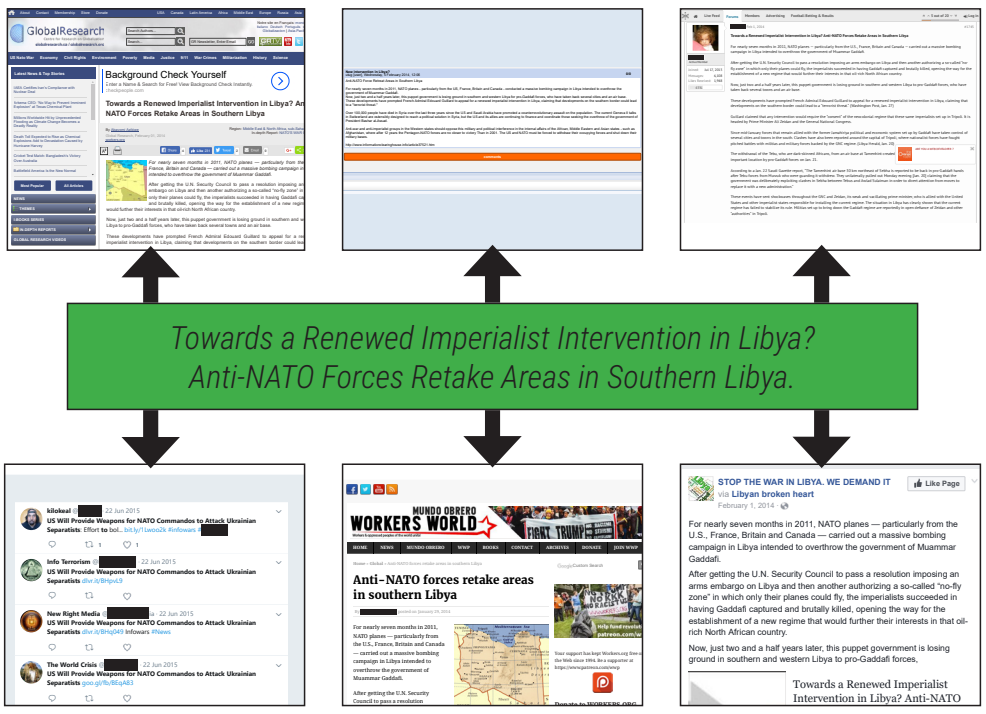
through social media channels. Figures below depict this pattern. Initially the content is generated on blog posts where the use of hashtags and links serve as the vehicles connecting to other social media channels, in this case to Twitter.

MIXED-MEDIA VS. CROSS-MEDIA APPROACHES

A mixed-media information dissemination campaign uses multiple social media channels to perpetuate a narrative. More precisely, the information campaign can be observed on multiple social media sites through the use of text, images, and audio and video content. Although the content may not be strictly identical on the various social media channels where it appears, it clearly pertains to a single information campaign.

A cross-media information dissemination campaign is characterized by a central channel around which the campaign is built. More precisely, the information is

MIXED-MEDIA STRATEGY FOR DISSEMINATING MISINFORMATION OR DISINFORMATION ON DIFFERENT WEBSITES.



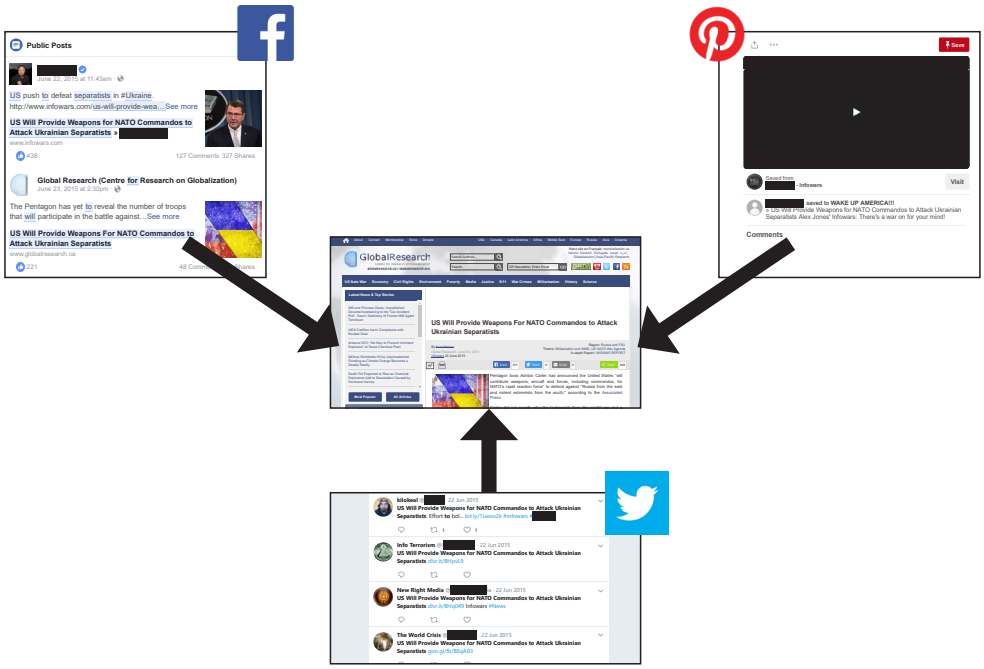
The image above shows the mixed-media dissemination campaign for 'Towards a Renewed Imperialist Intervention in Libya? Anti-NATO Forces Retake Areas in Southern Libya'

hosted on a website (e.g. text on a blog or video on a YouTube channel) and is widely distributed through other social media channels that provide established social network structures, such as Twitter, Facebook, etc.

First we investigated the use of the mixed-media approach in disseminating stories. In this study, we encountered cases where an article was shared on different sites as shown on page 40. For instance, a story titled *'Towards a Renewed Imperialist Intervention in Libya? Anti-NATO Forces Retake Areas in Southern Libya'* was disseminated on multiple sites, i.e. facebook.com, oroom.org, twitter.com, globalresearch.ca, hotnews.ro, and workers.org.[142]

Next, we examined the cross-media information dissemination approach. This tactic was observed to a good effect in our dataset. There were many sites that shared links to specific social media channels such as Twitter, Facebook, and Reddit sites. For instance, a blog site named 'globalresearch.ca' had a post entitled *'US Will Provide Weapons For NATO Commandos to Attack Ukrainian*

CROSS MEDIA INFORMATION DISSEMINATION STRATEGY FOR DISSEMINATING MISINFORMATION OR DISINFORMATION ON SOCIAL MEDIA



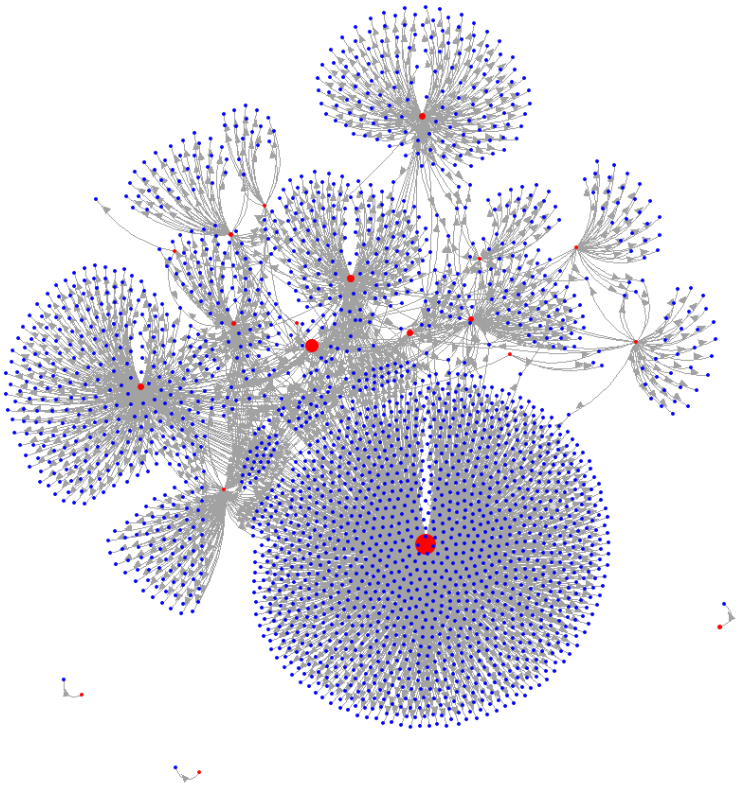
Separatists' with the link – http://bit.ly/2ewVTg7. This post was shared on Twitter (http://bit.ly/2xEQxnU), Pinterest (http://bit.ly/2x02sQ0), and Facebook (http://bit.ly/2wrIhZD) as depicted below. This clearly indicates a cross-media pattern.

TRACKING HOW AN ANTAGONISTIC NARRATIVE TRAVELS

To analyze how a narrative travels, we examined the 'likes' and 'comments' features available on blogs. A higher number of retweets, shares and comments at blog level show that posts have been circulated widely, demonstrating that media integration strategies do help in disseminating the narratives. Readers can like the content and comment on the post. Note that the 'like' feature on the blogs embeds various social plugins from Twitter, Facebook, Reddit, etc. These social plugins allow readers to like the page simultaneously on the different social media platforms, thereby disseminating the content on a variety of platforms simultaneously. For instance, a blog site, 21stcenturywire.com, published a blog post on September 18, 2016 entitled *'Syria: No "Dusty Boy" Outrage for 7 yr old Haider, Sniped by NATO Terrorists in Idlib Village of Foua'*. This blog post received 65 comments in which the audience presented their views. Moreover, the article was shared on other social media channels such as Twitter, where it got 19 retweets, 5 likes, and 2 replies. The same post on Facebook got 6 reactions, 3 comments, and 2 shares. Also, many groups posted this article to disseminate to an intended audience. The same blog, i.e. 21stcenturywire.com, published another blog post on September 27, 2016 entitled *'EU NATO Commit Adultery, Prince Charles Saudi Trade & More'* that again presented factually incorrect information. As we did with the previous example, we tracked how this post was disseminated through different social media channels. This blog post, however, received no comments. The article was shared on Twitter, but it got only 1 retweet, 1 like, and no replies. The same post was also shared on Facebook, where it received 27 reactions, 1 comment, and 11 shares. But all the shares were coming from the same group, 21stcenturywire.com. No other Facebook group posted this article. Since not many individuals or groups showed interest in spreading this information, it is clear that this article did not get any traction on blogs and not much on other social media platforms.

Next, we analyzed the effects network of blogs have on content dissemination. Unlike social media platforms, blogs do not have a social network structure, i.e. there is no follow-follower relation among blogs. However, it is still possible to observe the information flow network in blogs based on who links to whom. More specifically, we examined the hyperlinks in the blogs to extract the blog network. We used this approach to extract the network of the blogs containing disinformation regarding Baltic States. We used specific software to visualize the network[143], as depicted in on page 43. The network

NETWORK* OF BLOGS AND SHARED HYPERLINKS

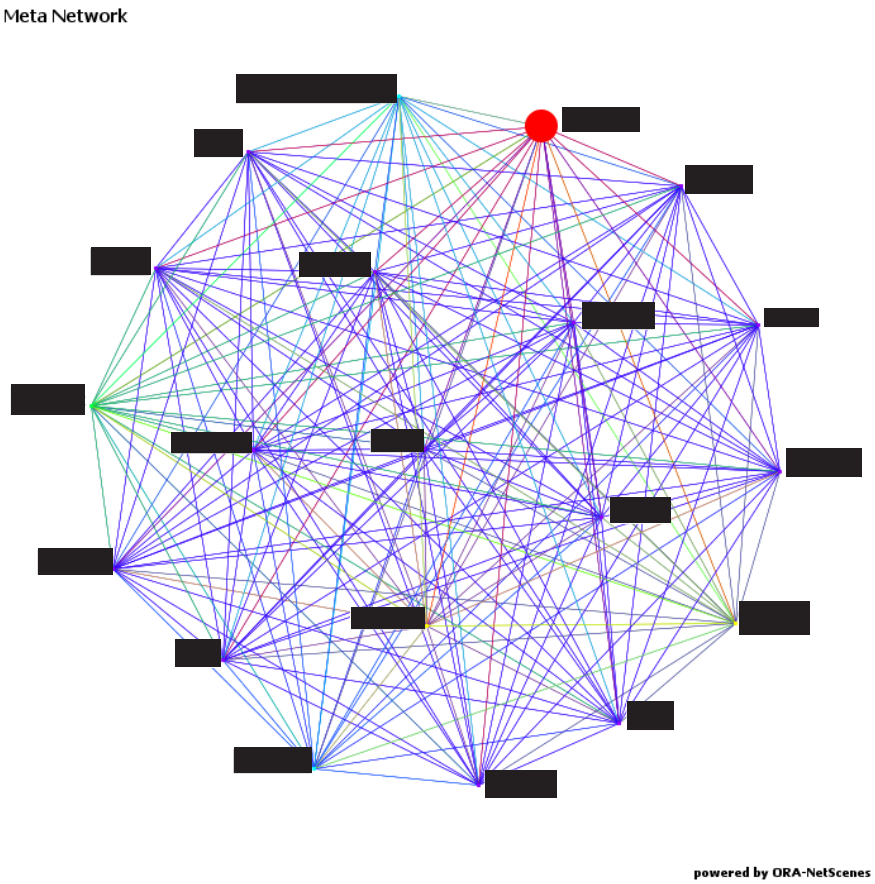


** The network contains 21 blogs (red nodes) and 2321 hyperlinks (blue nodes). Size of a node is proportional to the number of shared hyperlinks (i.e. out-degree centrality). Edge thickness is proportional to the number of times a blogs shared a hyperlink.*

contains 21 blogs (red nodes) and 2321 hyperlinks (blue nodes). Further analysis of the blog network helps in identifying 5 blogs out of 21 that were the most resourceful (having the most hyperlinks), as well as the most exclusive in resources (i.e. they shared hyperlinks that no other blogs shared). 10 out of 2321 hyperlinks were the most shared and most exclusively shared, i.e., these hyperlinks were shared by only a few blogs. Most of these top ten shared hyperlinks have a domain suffix from the Baltic nations, i.e. '.ee' for Estonia, '.lv' for Latvia, and '.lt' for Lithuania.

The exclusivity of resource sharing by a few blogs hints at information campaign coordination. To dig deeper, we construct a blog network based on the commonly shared hyperlinks. The blog network thus identified is depicted on the next page. The network is fully connected, i.e. every blog connects to every other blog. This suggests that every blog in this set shared the same hyperlinks. This confirms our conjecture that there is intensive campaign coordination among these blogs. Further investigation is required to know if these blogs belong to

A NETWORK* OF BLOGS BASED ON COMMONLY SHARED HYPERLINKS

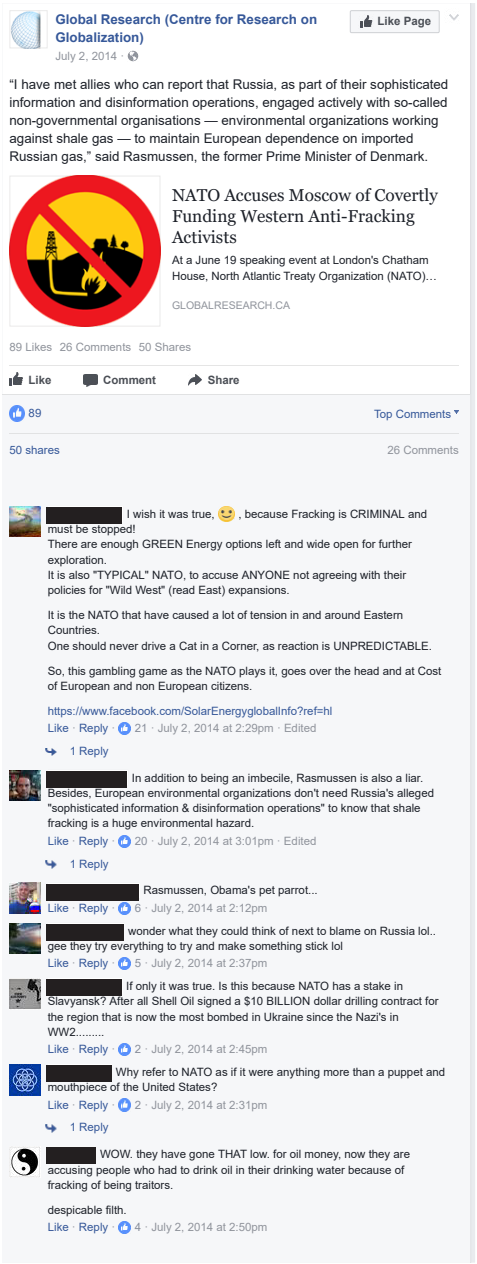


** The network is fully connected, i.e. a clique, where every blog is connected with every other blog. This depicts massively coordinated information campaign.*[144]

or are controlled by the same individual or a group. [144]

Next, we analyzed the role of blogs in providing a persuasive dimension to the narrative. We examined how 'exemplified accounts'[145] in the user comments to a story may influence audience perceptions.[146] We provide an example where commentary lends a persuasive dimension to the blog post.

On page 45 it is possible to see how exemplified accounts in users' comments for a post may influence the audience perceptions.[147] After reading through the comments, we can actually observe that some of the commenters' accounts help in developing a persuasive discourse.



An example illustrating exemplified accounts in comments may shape the discourse to form a persuasive dimension

We observe that user comments actually augment the narrative presented in the blog post. We can see a lot of users commenting about the post to further strengthen the narrative. At the same time, we can see patterns such as linking this content to other websites or pages (such as Facebook fan pages), sharing to other channels (50 shares) to further raise discussions.

CONCLUSIONS

Blogs are becoming virtual town halls that are shaping the public perceptions and narratives of regional events. Narratives are first framed on the blogs, then they are disseminated through other social media channels. The key findings include the identification of massively coordinated information campaigns among blogs by applying social network analysis concepts—and demonstrating that commentary on blogs lends a persuasive dimension to the discourse.

In our research, we highlighted the role that blogs can have in weaponizing narratives and conducting disinformation campaigns, suggesting that action be taken towards developing countermeasures. The major contributions of this chapter include: assessment of guidelines for detecting blogs containing misinformation or disinformation; tracking the origins of the content on blogs such as memes, images, videos, etc.; evaluating mixed-media and cross- media narrative dissemination strategies; tracking how the narratives originating in blogs travel

in the social media ecosystem; and analyzing campaign coordination from blog networks. We studied four different blog datasets consisting of 372 blog sites, 7576 bloggers, and 196,940 blog posts riddled with misleading or false information. Social network analysis of the blog network revealed most resourceful blogs and blogs that were most exclusive in sharing resources. Furthermore, a massive misinformation coordination campaign was discovered.

Acknowledgements

This research is funded in part by the U.S. National Science Foundation (IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2605, N00014-17-1-2675), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059) and the Jerry L. Maulden/Entergy Fund at the University of Arkansas at Little Rock. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.