



Research Article

Search engine manipulation to spread pro-Kremlin propaganda

The Kremlin's use of bots and trolls to manipulate the recommendation algorithms of social media platforms is well-documented by many journalists and researchers. However pro-Kremlin manipulation of search engine algorithms has rarely been explored. We examine pro-Kremlin attempts to manipulate search engine results by comparing backlink and keyphrase networks of US, European, and Russian think tanks, as well as Kremlin-linked "pseudo" think tanks that target Western audiences. Our evidence suggests that pro-Kremlin pseudo-think tanks are being artificially boosted and co-amplified by a network of low-quality websites that generate millions of backlinks to these target websites. We find that Google's search algorithm appears to be penalizing Russian and pseudo-think tank domains.

Authors: Evan M. Williams (1), Kathleen M. Carley (1)

Affiliation: (1) School of Computer Science, Carnegie Mellon University, USA

How to cite: Williams, E. M., & Carley, K. M. (2023). Search engine manipulation to spread pro-Kremlin propaganda. *Harvard Kennedy School (HKS) Misinformation Review*, 4(1).

Received: August 10th, 2022. Accepted: January 13th, 2022. Published: February 16th, 2023.

Research questions

- Are there "link scheme" attempts to manipulate search traffic by using low-quality websites generated to create large amounts of links to pro-Kremlin think tank and pseudo-think tank domains?
- What qualitative characteristics do we observe in these amplifier websites, and what keyphrases do they target?
- Are other think tanks co-amplified by the same link scheme websites?

Essay summary

- We used network analysis and data analysis to explore search engine optimization (SEO) efforts around Kremlin-linked think tank and pseudo-think tank websites.
- We found that pro-Kremlin pseudo-think tank websites are heavily co-amplified by low-quality link scheme domains and that these sites rank highly for many conspiratorial keyphrases.
- We found that the network of websites amplifying pseudo-think tanks has strong connections to various US, European, and Russian think tanks.

¹ A publication of the Shorenstein Center on Media, Politics and Public Policy at Harvard University, John F. Kennedy School of Government.

- If search engine algorithms are impacted by this attempted manipulation, more users will unknowingly enter pro-Kremlin media ecosystems through search engines.
- Google's search algorithm appears to be penalizing Russian and pseudo-think tanks; we made no assessments for other search engines.

Implications

In 2014, Vitaly Bespalov was hired as a writer by a secretive organization, later revealed to be the Kremlin-linked Internet Research Agency (IRA), a troll farm that was indicted by a US federal grand jury for online US election interference in 2018 (United States of America v. Internet Research Agency LLC et al). On an average day, Bespalov recounts being tasked with rewriting articles on Ukraine over and over, each time keeping roughly 70 percent of the original text. Bespalov states he was asked to change words like "terrorist" to "militia" or to write "national guard" instead of "Ukrainian army," and he was instructed to never criticize Russia in these articles (Poppins & Cobiella, 2017). The goal of this operation, according to Bespalov, was to get the articles to the top of search engine results (Poppins & Cobiella, 2017). These kinds of activities present an important and understudied component of Kremlin-linked attempts to spread online misinformation and propaganda.

In recent years, the Kremlin's influence operations on social media platforms have been the subject of widespread public attention and research. In 2017, an estimated 127 million individuals were exposed to Russian disinformation on Facebook alone (Isaac & Wakabayashi, 2017). In 2021, Facebook reported in its *Threat Report* on influence operations that Russia remains the biggest driver of disinformation. In that same report, Facebook² details the removal of over five Kremlin-backed networks targeting a number of countries and regions. Facebook claims it was able to link these attacks back to the Internet Research Agency (IRA) and other non-IRA Russian military intelligence organizations (Facebook, 2021). To make these information operation attacks successful, IRA *trolls* (fake personas on social media platforms that seek to disrupt governments, individuals, or institutions) rely on a wide range of techniques to manipulate the recommendation algorithms of social media platforms (Carley, 2020). By strategically amplifying certain media, hashtags, and accounts, the IRA can greatly extend the reach of such messages (Benigni et al., 2017). Similarly, using what is known about search engine ranking algorithms, bad actors can attempt to artificially inflate website search result rankings.

Search engine rankings represent an important line of inquiry in misinformation research, as the rankings can substantially impact both what information users consume and what information they believe. High-ranking pages—for example, the pages that appear at the top of a Google search result—are far more likely to be seen. In a 2013 analysis of 300 million search engine clicks, 92% were on the first page of search results, and 51% of those were the first or second result (Chitika Insights, 2014). Additionally, users were found to be 140% more likely to click the last result on the first page than the first result on the second page (Chitika Insights, 2014). More recent analyses by Backlinko, an SEO firm, and Ignite Visibility, a digital marketing firm, both found the click-through rate of the first result was ten times higher than that of the tenth result (Dean, 2022; Lincoln, 2020). It has been demonstrated that search engine rankings can have an impact on the political beliefs and voting patterns of users. Three laboratory experiments with double-blind control group designs found that relatively minor changes in search engine rankings could influence decisions of undecided voters by 20% (Epstein & Robertson, 2015). While the magnitude of this

² Now Meta.

effect has been contested,³ if the manipulation of search engines is successful, more individuals can be exposed to artificially amplified content.

This increased exposure can have substantial impacts on information environments and has become increasingly concerning in recent years with the rise of what Marwick and Partin call *populist expertise*, or the rejection of experts and traditional information vectors in favor of alternative “home-grown” knowledges (Marwick & Partin, 2022). On QAnon message boards, users implore other users to “do the research” in order to understand the conspiracies in Q’s messaging. More recently, the phrase “do your own research” has become synonymous with the vaccine-hesitancy movement that seeks to lead users down anti-vax rabbit holes (Ballantyne & Dunning, 2022). A 2021 search engine audit study found that Google was highly effective at suppressing material that promotes misinformation given conspiracy keyphrases, whereas Bing, Yahoo, Yandex, and DuckDuckGo were far less effective (Urman et al., 2022). Consequently, right-wing extremists and conspiracy theorists, angry at what they view as Google’s suppression of information—reliable or otherwise—have begun urging followers to use DuckDuckGo and other search engines that are less effective at suppressing misinformation (Thompson, 2022). As susceptible users are being told to seek information in less-reliable information ecosystems and increasingly rejecting expert advice, understanding the environment in which users “do the research” is important.

While the Kremlin’s influence operations on social media platforms have received widespread attention, there is currently very little research on any state-backed attempts to manipulate search engine recommendation rankings. Several audit studies have been conducted where researchers query terms on multiple search engines and compare the rankings of relevant results. In one audit, Kravets and Toepfl (2021) found that Yandex returned far fewer websites than Google with information on anti-regime Russian protests. An additional cross-country audit found Google to be effective at suppressing pro-Kremlin content in five countries (Toepfl et al., 2021). Additionally, one audit compared Google and Yandex rankings of content related to Alexei Navalny and found that Google promoted opposition sources more frequently than Yandex (Makhortykh et al., 2022). Search engine audit studies are useful but are constrained by the set of keyphrases chosen by the researchers. Addressing some of these limitations, Samantha Bradshaw (2017) used proprietary data to examine search optimization strategies of junk news sites. Hrckova et al. (2021) used the SEO tool, Ahrefs, to explore linking patterns of partisan news sites in central Europe. We have adopted a similar methodology to that of Bradshaw (2017) and Hrckova et al. (2021) in this study.

We have examined the networks of backlinks and *keyphrases*—search terms for which websites rank highly—of US, European, Russian, and what we refer to as pro-Kremlin “*pseudo*” *think tanks*—organizations that blur the lines between news, think tank content, misinformation, and propaganda. Details on these networks can be found in the Methods section of this paper. Successful search manipulation (see the Appendix for an overview) can result in websites ranking highly on search engines for specific keyphrases. Conspiracy theories have a distinct advantage in this space, as conspiracy keyphrases can be highly specific and have very low competition (Golebiewski & Boyd, 2019). For example, Katehon is a pseudo-think tank with over 99% of its links coming from low-quality domains.⁴ Despite its weak overall profile, as of September 24, 2022, a highly conspiratorial article produced by Katehon appears

³ A statistical critique by Katarina A. Zweig contested the results, arguing that the authors may have overestimated effect size with their choices of statistical tests. Zweig estimates that real effects were likely 2-4%, which is still an important effect, particularly in settings where elections are close (Zweig, 2017).

⁴ Over 99% of backlinks to Katehon come from websites with a normalized (0–100) page rank score (calculated by Ahrefs) of 20 or lower.

in the top three results on Google, DuckDuckGo, Bing, Yahoo, and Yandex when searching for the keyphrase “Rothschild criminal.”⁵

Our principal recommendation is that further exploration is needed. While Google appears to be penalizing the Russian and pseudo-think tanks in this report, we do not have enough data to determine how widespread these manipulation behaviors are or how Google’s penalization algorithm generalizes to other contexts. If Google is penalizing based solely on curated lists—where the links between these think tanks and the Kremlin have been publicly documented by academic and government sources—then newly-created sites or sites not otherwise identified as misinformation may not be penalized. If, in contrast, Google’s algorithm is penalizing these domains as a result of their backlink profiles, research would need to be done to ensure smaller “innocent” sites are not also being penalized. Finally, we note that even if Google perfectly penalizes these sites, these sites can still attract new users through other search engines, social media, or through direct links from other domains. Search engine companies have monetary interests in stopping search manipulation, which makes countering search manipulation an area where the goals of companies, regulators, and other stakeholders are aligned (Dipayan & Scott, 2018). This study provides a methodological framework to help researchers, journalists, and other stakeholders further explore and expose state-aligned search manipulation attempts in other country and topic contexts.

Findings

Finding 1: Pro-Kremlin websites are heavily amplified by domains seemingly built for generating backlinks.

We find a highly imbalanced backlink volume across networks. Global Research, a Kremlin-aligned pseudo-think tank, receives 22.1 million backlinks, more than all US, European, and Russian think tanks combined. We visualized the 15 think tanks with the largest backlink volume in the left panel of Figure 1. In total, American think tanks received 14.1 million links, European think tanks received 4.6 million links, Russian think tanks received 1 million links, and pseudo-think tanks received 30.8 million backlinks. Although the pseudo-think tanks received more links than Russian and US think tanks combined, most of these links were from low-quality domains. Ahrefs (see Methods section) calculates PageRank scores (see Appendix) for every domain in its database normalized from 0–100, where 100 signifies the most authoritative. Websites that linked only to pseudo-think tanks in our network had a mean PageRank score of 17.04, whereas the mean domain rank for all other backlinking websites was 35.59.

⁵ The article promotes longstanding antisemitic conspiracies that a global cabal of Jewish people is secretly running the world and is headed by the Rothschild family. This term is defined in ADL’s Glossary of Extremism (<https://extremismterms.adl.org/glossary/rothschilds>).

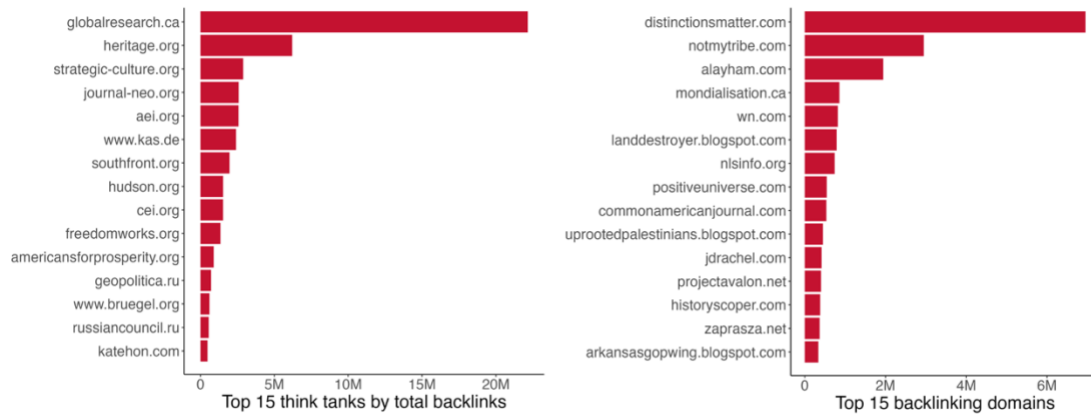


Figure 1. Left: Top 15 think tanks by backlink volume. Right: Top 15 backlinking websites.

Distinctionsmatter, the website with the largest volume of outbound links (all to pseudo-think tank domains), bears many of the indicators of a link scheme website. There is no “about” or contact information for the author, there are no ads on the page, and the page generates millions of links to unreliable news sites. Distinctionsmatter posted 6.6 million links to Global Research—more than any other site. Its second most linked-to site (with 2.2 million links as of 22 September, 2022) is Zero Hedge, a news outlet accused of spreading Russian propaganda by US intelligence agencies (Walsh, 2022).

The site that generates the second most backlinks, notmytribe.com, has the same general tone as many IRA bots identified in Twitter information operations. IRA trolls have been known to taunt users by sarcastically declaring that the accounts were Russian bots. In that same vein, the navigation banner on notmytribe.com’s landing page has “News,” “Culture,” “Work,” and “Disinfo” drop-down sections, shown in Figure 2. The syntax of the drop-down labels suggests non-native English or machine translation. These include “Info Virus,” “On War Machine,” “AntiGlobalization,” and “Nighttime Gardening.” The purpose of the website is almost certainly to boost ranking of other sites, as it generates 2.9 million backlinks to Global Research. We inspected each of the 15 top backlinking domains to see if the domains 1) spread pro-Kremlin propaganda, 2) linked disproportionately to pro-Kremlin domains, or 3) linked rarely to non-pro-Kremlin domains. We found 13 of 15 top backlinking domains, with the exceptions of wn.com and nlsinfo.org, met these criteria. By pro-Kremlin propaganda, we mean content that consistently aligns with the geopolitical goals of the Kremlin, for example, using misinformation to attack Ukraine and Western institutions critical of Russia while also never criticizing the Russian government. These sites appear to have been built for the purpose of either spreading pro-Kremlin propaganda or increasing the search rankings of pro-Kremlin websites by using automation to generate hundreds of thousands or millions of links to these domains.



Figure 2. Notmytribe.com’s site navigation bar contains a “Disinfo” dropdown with conspiratorial subsections.

Finding 2: Keyphrases of pseudo-think tanks exhibit high internal overlap and appear to target conspiracy theorists.

Many of the top keyphrases shared between think tank groups are names of people, but many are very specific and conspiratorial, particularly within the pseudo-think tank network. These keyphrases appear to be targeting conspiratorial “data voids” where search results are sparse and low-quality websites can rank highly (Golebiewski & Boyd, 2019). Among the pseudo-think tank keyphrases are phrases like “Is Zelensky a drug addict,” “Zelensky on cocaine,” “neutron bomb in Yemen,” and “subcortical vascular dementia Hillary Clinton.” The keyphrases network in Figure 3 visualizes the keyphrases used by two or more think tanks. The Russian think tanks are characterized largely by Russian-language keyphrases and only have one overlapping keyphrase with another think tank group.⁶ However, a handful of keyphrases cross think tank groups (see the red nodes in Figure 3). Several keyphrases are shared between pseudo-think tanks and hudson.org, including “climate change money trail,” three variations of “Mike Pompeo speech,” and six keyphrases about former CIA director John Brennan voting communist.⁷ EU and pseudo-think tanks share two common keyphrases: “great prophet 17” and “Europe’s reaction to Donald Trump.” We found, on average, Russian and pseudo-think tanks rank much lower for keyphrases than US and EU think tanks. Averaging Google positions⁸ over the top 1,000 keyphrases for each network, we find significant disparity between think tank groups. The average positions ranking over keyphrases are as follows: US (7), Europe (11), Russia (35), and pseudo-think tanks (34). In other words, on average, the best terms for US and European think tanks rank higher on Google than those of pseudo and Russian think tanks. This suggests Google is penalizing the pseudo-think tank domains despite, or perhaps as a result of, the ongoing link scheme manipulation attempts. In contrast, US and European think tanks do not appear to be penalized.

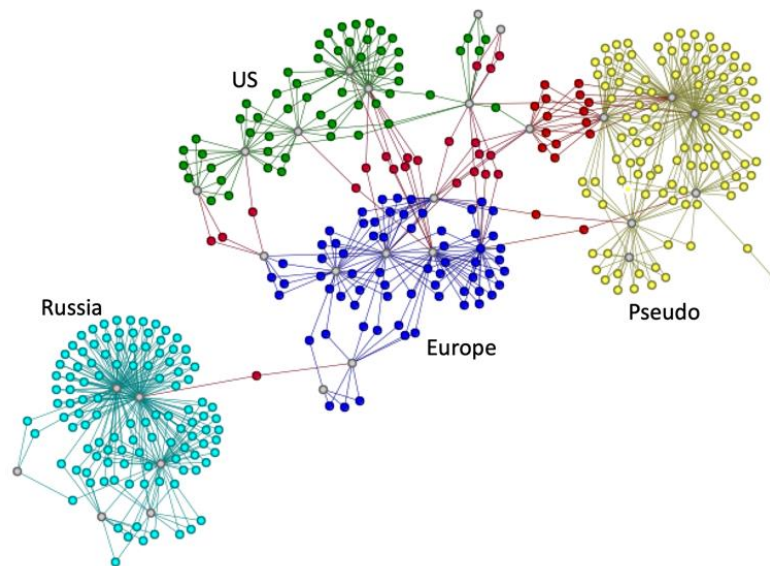


Figure 3. Keyphrase network visualization. Grey nodes are think tanks. Blue nodes are EU keyphrases, teal are Russian keyphrases, green are US keyphrases, yellow are pseudo-think tank keyphrases, and red are keyphrases shared across different think tank groups.

⁶ Russiacouncil.ru (Russia) and ifri.org (Europe) share “csto.”

⁷ “Did John Brennan vote communist,” “John Brennan communist,” “John Brennan communist vote,” “John Brennan voted for a communist,” “John Brennan voted communist,” “John Brennan voted for communist.”

⁸ Extracted via Ahrefs.

Finding 3: Many pseudo-think tanks are strongly amplified by the same websites.

We constructed a co-amplification network for inbound links using the pairwise co-amplification weights between each website, and we visualized a filtered version of this co-amplification network in Figure 4. Edges in this network reflect the strength at which think tanks are linked to by the same domains (details in the Methods section). We only visualized edges with edge weights larger than 15k.⁹ The most substantial co-amplification happens within the pseudo-think tank network. The sites with the highest overall co-amplification scores are Global Research (4.4M), Strategic Culture Foundation (4M), Heritage Foundation (3.8M), New Eastern Outlook (3.2M), and American Enterprise Institute (2.7M). Each of these think tanks receives high volumes of links from domains that heavily link to other think tank websites.

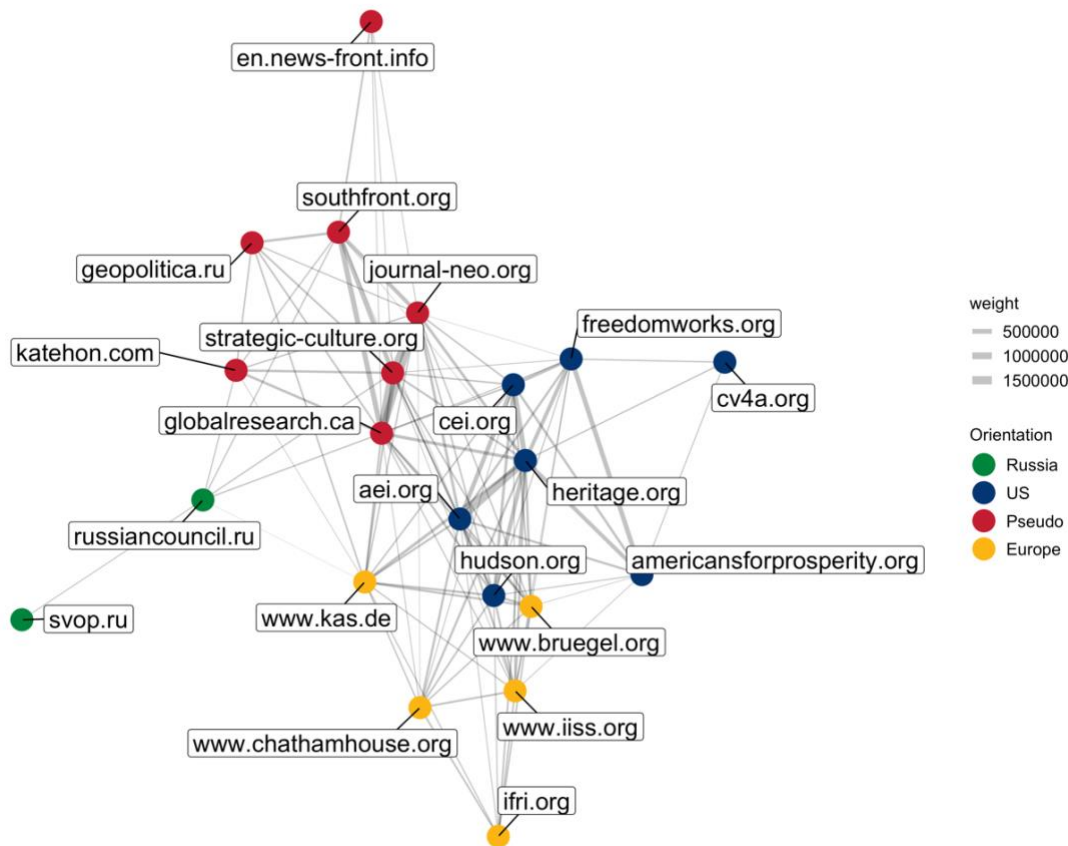


Figure 4. Filtered co-amplification network. Each edge depicted indicates at least 15k links from the same set of referring domains. Green nodes are Russian, yellow nodes are European, blue nodes are US, and red nodes are pseudo-think tanks.

The strongest edge weights for these sites are between think tanks of the same group, but there is notable co-amplification across groups, particularly between pseudo-think tanks and American right-wing think tanks. The Heritage Foundation has an edge weight of 163k with New Eastern Outlook and an edge weight of 213k with Global Research. The Heritage Foundation, American Enterprise Institute, and New Eastern Outlook all received almost exactly 102.2k backlinks from historyscoper-islamwatch.blogspot.com, an anti-Islamic blog that heavily links to a variety of right-wing sources and contains millions of external links to what appear to be conspiracy sites. oldephartte.blogspot.com is a more standard SEO blog that links

⁹ This number is arbitrarily chosen, but it both highlights the strongest pairwise ties and improves readability.

heavily to Global Research, Strategic Culture Foundation, and New Eastern Outlet while also often linking to the Heritage Foundation's website.

Methods

Data collection

To choose the initial domains of interest, we took advantage of the fact that the Kremlin does not invest solely in hard-power operations to spread its influence. It also invests in numerous soft-power initiatives, one of which is Kremlin-linked think tanks and pro-Kremlin pseudo-think tanks. We examined the websites of eight of the Kremlin-linked think tanks, identified by the Institute of Modern Russia, with primarily domestic Russian audiences (Smagliy, 2018). These think tanks primarily target Russian civilians and academics, and as a result, most of these think tanks publish primarily in Russian. We contrasted these with eight influential Western European Think Tanks and eight US conservative think tanks, which we hypothesized could provide a baseline without strong evidence of search manipulation. For Europe, we used the top eight Western European think tanks ranked in the University of Pennsylvania's *2020 Global Go To Think Tank Index Report* (McGann, 2021). To identify US conservative think tanks, we use the eight think tanks that supplied the largest number of staff, cabinet, and political appointees in the Trump administration (Kravitz et al., 2019). We contrasted these think tank networks with a network of the seven Kremlin-backed proxy outlets identified in the Global Engagement Center's *Pillars of Russia's Disinformation and Propaganda Ecosystem* report (U.S. Department of State: Global Engagement Center, 2020). We call these *pseudo-think tanks*, as each of these proxy outlets blurs the lines between news, think tanks, misinformation, and propaganda. The purpose of these pseudo-think tanks is to spread Russian state propaganda to Western audiences. All think tanks and their network assignments can be found in Table 1.

Table 1. All think tank domains and their corresponding networks.

Russia	Europe	US	Pseudo
cskp.ru	bruegel.org	freedomworks.org	strategic-culture.org
rethinkingrussia.ru	realinstitutoelcano.org	cei.org	globalresearch.ca
russiancouncil.ru	clingendael.org	heritage.org	journal-neo.org
svop.ru	ifri.org	cv4a.org	geopolitica.ru-en
foreignpolicy.ru	chathamhouse.org	hudson.org	southfront.org
iiseps.org	dc.fes.de	americaneconomicfreedomalliance.com	en.news-front.info
doc-research.org	kas.de	americansforprosperity.org	katehon.com-en
eurasian-strategies.ru	iiss.org	aei.org	

All data for this project was collected using the platform Ahrefs.¹⁰ Ahrefs advertises a 12-trillion link database and the second most active commercial web crawler after Google.¹¹ For each of the 31 think tanks we identified, we used the platform to pull three different networks. A table of these think tank domains and their alignments can be found in the Appendix in Table 1. For each of the 31 think tanks, we pulled 1) a network of websites with the largest number of backlinks to the source website and 2) a network of the top keyphrases, as identified by Ahrefs, for which the source website ranks. For each target domain, we pulled the 1,000 top backlinking websites by backlink volume and the 1,000 keyphrases with the highest Google positions for each domain. If a website had fewer than 1,000 results for any query, all results were returned. In the backlink networks, source nodes are the think tanks, target nodes are domains, and edge weights are the number of backlinks between every nonzero domain pair. In the keyphrase network, source nodes are think tank websites, target nodes are keyphrases, and edges reflect a connection. These queries resulted in two groups of four, or eight separate networks. Summary statistics for each of these networks are reported in Table 2 (backlink network) and Table 3 (keyphrase network). Nodes are the number of websites or websites and keyphrases in a group (under 8,000 means not all target domains had at least 1,000 backlinking domains or keyphrases). Edges represent unweighted ties between unique domains or keyphrases. The total weighted degree is the total volume of backlinks in the network. Data were collected on September 22, 2022.

Table 2. Backlink network statistics for each think tank group.

	Russia	Europe	US	Pseudo
Nodes	3,247	5,362	4,729	4,427
Edges	4,170	7,422	7,071	6,995
Total Weighted Degree	1,017,762	4,631,448	14,130,634	30,863,844
Clustering Coef.	0.006	0.002	0.12	0.2
Density	0.0008	0.0005	0.0006	0.0007

Table 3. Keyphrase network statistics for each think tank group.

	Russia	Europe	US	Pseudo
Nodes	3,184	7,938	6,963	3,834
Edges	3,293	8,000	7,011	3,954
Total Weighted Degree	3,293	8,000	7,011	3,954
Clustering	0	0	0	0
Density	0.0006	0.0003	0.0003	0.0005

Co-amplification

To measure the co-amplification of websites, we needed to quantify the shared backlink overlap between domains. A more formal mathematical description of our overlap metric can be found in the Appendix. Intuitively, if website i links strongly to both websites u and v , the overlap score for u and v with respect to i should be high. Conversely, if website i links strongly to u but not to v or links weakly to both sites, the

¹⁰ <https://ahrefs.com/>

¹¹ <https://ahrefs.com/robot>

overlap score between u and v with respect to i should be low. We, therefore, take the minimum number of times website i links to u or v . This is repeated for each websites i linking to u and v . The sum of those minimums is our overlap metric. More formally, $\mathcal{O}_{u,v} = \sum_{i=1}^d \min(u_i, v_i)$. This overlap score is high when sites are strongly co-amplified by the same set of URLs and low otherwise.

Bibliography

- Ballantyne, N., & Dunning, D. (2022). Skeptics say, ‘do your own research.’ It’s not that simple. *The New York Times*. <https://www.nytimes.com/2022/01/03/opinion/dyor-do-your-own-research.html>
- Benigni, M., Kenny, J., & Carley, K. M. (2017). Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter. *PLOS ONE*, 12(12), e0181405. <https://doi.org/10.1371/journal.pone.0181405>
- Carley, K. M. (2020). Social cybersecurity: An emerging science. *Computational and Mathematical Organizational Theory*, 26(4), 365–381. <https://doi.org/10.1007/s10588-020-09322-9>
- Chitika Insights (2013). *The value of Google result positioning*. <https://research.chitika.com/wp-content/uploads/2022/02/chitikainsights-valueofgoogleresultspositioning.pdf>
- Dean, B. (2022). We analyzed 4 million Google search results: Here’s what we learned about organic click through rate. Backlinko. <https://backlinko.com/google-ctr-stats>
- Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, 112(33), E4512–E4521. <https://doi.org/10.1073/pnas.1419828112>
- Facebook. (2021). *Threat report: The state of influence operations 2017-2020*. <https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf>
- Ghosh, D., & Scott, B. (2018). *Digital deceit: The technologies behind precision propaganda on the Internet*. New America. <https://d1y8sb8igg2f8e.cloudfront.net/documents/digital-deceit-final-v3.pdf>
- Golebiewski, M., & Boyd, D. (2019). *Data voids: Where missing data can easily be exploited*. Data & Society Research Institute. <https://datasociety.net/library/data-voids/>
- Hrckova, A., Moro, R., Srba, I., & Bielikova, M. (2021). Quantitative and qualitative analysis of linking patterns of mainstream and partisan online news media in Central Europe. *Online Information Review*, 46(5), 954–973. <https://doi.org/10.1108/OIR-10-2020-0441>
- Isaac, M., & Wakabayashi, D. (2017, October 30). Russian influence reached 126 million through Facebook alone. *The New York Times*. <https://www.nytimes.com/2017/10/30/technology/facebook-google-russia.html>
- Kravitz, D., & Shaw, A., & Perlman, C., & Mierjeski, A. (2019). *Trump town: Track White House staff, cabinet members and political appointees across the government*. ProPublica. <https://projects.propublica.org/trump-town/>
- Kravets, D., & Toepfl, F. (2021). Gauging reference and source bias over time: How Russia’s partially state-controlled search engine Yandex mediated an anti-regime protest event. *Information, Communication & Society*, 25(15), 2207–2223. <https://doi.org/10.1080/1369118X.2021.1933563>
- Lincoln, J. E. (2020). Google click-through rates (CTR) by ranking position [2020]. <https://ignitevisibility.com/google-ctr-by-ranking-position/>
- Makhortykh, M., Urman, A., & Wijermars, M. (2022). A story of (non) compliance, bias, and conspiracies: How Google and Yandex represented Smart Voting during the 2021 parliamentary elections in Russia. *Harvard Kennedy School (HKS) Misinformation Review*, 3(2). <https://doi.org/10.37016/mr-2020-94>

- Marwick, A. E., & Partin, W. C. (2022). Constructing alternative facts: Populist expertise and the QAnon conspiracy. *New Media & Society*. <https://doi.org/10.1177/14614448221090201>
- McGann, J. G. (2021). *2020 global go to think tank index report*. Think Tanks and Global Societies Program, University of Pennsylvania. https://repository.upenn.edu/think_tanks/18/
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab. <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>
- Popken, B., & Cobiella, K. (2017, November 16). *Russian troll describes work in the infamous misinformation factory*. NBC News. <https://www.nbcnews.com/news/all/russian-troll-describes-work-infamous-misinformation-factory-n821486>
- Smagily, K. (2018). *Hybrid analytica: Pro-Kremlin expert propaganda in Moscow, Europe and the US: A case study on think tanks and universities*. Institute of Modern Russia. <https://www.underminers.info/publications/hybridanalytica>
- Thompson, S. A. (2022, February 23). Fed up with Google, conspiracy theorists turn to DuckDuckGo. *The New York Times*. <https://www.nytimes.com/2022/02/23/technology/duckduckgo-conspiracy-theories.html>
- Toepfl, F., Kravets, D., Ryzhova, A., & Beseler, A. (2022). Who are the plotters behind the pandemic? Comparing Covid-19 conspiracy theories in Google search results across five key target countries of Russia's foreign communication. *Information, Communication & Society*. <https://doi.org/10.1080/1369118X.2022.2065213>
- Tripodi, F. B. (2022). *The propagandists' playbook: How conservative elites manipulate search and threaten democracy*. Yale University Press.
- United States of America v. Internet Research Agency LLC et al. (2018). Case 1:18-cr-00032-DLF. United States District Court for the District of Columbia. <https://www.justice.gov/file/1035477/download>
- Urman, A., Makhortykh, M., Ulloa, R., & Kulshrestha, J. (2022). Where the earth is flat and 9/11 is an inside job: A comparative algorithm audit of conspiratorial information in web search results. *Telematics and Informatics*, 72, 101860. <https://doi.org/10.1016/j.tele.2022.101860>
- U.S. Department of State: Global Engagement Center (2020). *Pillars of Russia's disinformation and propaganda ecosystem*. https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia's-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf
- Walsh, J. (2022). U.S. reportedly accuses financial blog Zero Hedge of publishing Russian propaganda. *Forbes*. <https://www.forbes.com/sites/joewalsh/2022/02/15/us-reportedly-accuses-financial-blog-zero-hedge-of-publishing-russian-propaganda/?sh=65f9fd08d2c4>
- Zweig, K. A. (2017). *Watching the watchers: Epstein and Robertson's "search engine manipulation effect."* Algorithm Watch. <https://algorithmwatch.org/en/watching-the-watchers-epstein-and-robertsons-search-engine-manipulation-effect/>

Funding

The research for this paper was supported in part by the ARMY Scalable Technologies for Social Cybersecurity, Office of Naval Research's 1) MURI: Near Real Time Assessment of Emergent Complex Systems of Confederates, 2) MURI: Persuasion, Identity & Morality in Social-Cyber Environments, 3) Scalable Tools for Social Media Assessment, and 4) Group Polarization in Social Media under grants N000141712675, N000142112749, N00014-21-1-2229, and N00014-18-1-2106. It was also supported by the Center for Informed Democracy and Social-cybersecurity (IDeaS) and the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research or the US Government.

Competing interests

The authors declare no conflicts of interest.

Ethics

Institutional review was not necessary as it analyzes only relations between websites.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

Data availability

All materials needed to replicate this study are available via the Harvard Dataverse: <https://doi.org/10.7910/DVN/OKNLOW>. The data provider, Ahrefs (ahrefs.com), must be cited if data are used.

Appendix: Details on search engine manipulation and co-amplification

What is search engine manipulation?

Search engines can rank websites using a variety of algorithms that are not made public, but by far the most well-known is PageRank, proposed by Google co-founder Larry Page (Page et al., 1999). Google's algorithm has become far more complex since PageRank was initially proposed, but the original PageRank ranking factors, backlink quantity, and quality, remain important. PageRank determines the ranks of websites by both the quantity and quality of their inbound links. Although many other attributes are weighted by search recommendation algorithms, link quality, and quantity remain important attributes. While it may be difficult to obtain a backlink from high-quality domains, link quantity is much easier to game.

Manipulating rankings through the creation of many new backlinks can involve paying third-party services to post a target website across third-party websites, creating websites to amplify a target website, hacking webpages and injecting invisible URLs, and many other maneuvers. Google calls these kinds of algorithmic manipulations “link schemes”¹² and explicitly forbids them in its webmaster guidelines.¹² When link quantity is manipulated, a website's ranking can be boosted for various keyphrases. This can result in the website appearing more frequently in the first page of search results, thereby increasing site traffic. *Search engine optimization* (SEO) is the broad class of actions one might take to increase the visibility of a website on search engines. We focus on the “link scheme” subset of SEO forbidden by webmaster guidelines (commonly referred to as “black-hat” SEO), which we refer to as *search engine manipulation* (SEM). In this work, our main interest is the prevalence of link schemes created to boost pro-Kremlin domains. For more details, we refer the reader to (Tripodi, 2022).

We also explore *keyphrases*—search terms for which websites rank highly. Conspiracy theories have a distinct advantage in this space, as keyphrases can be highly specific and have very low competition (Golebiewski & Boyd, 2019). For example, if a user in a vaccine hesitancy group or forum sees a post falsely accusing Pfizer of a specific criminal action, a conspiracy website could reiterate the claim and rank highly for its keyphrases, as there's often little keyphrase competition for emergent conspiratorial stories.

Co-amplification

To measure the co-amplification of websites, we needed to quantify the shared backlink overlap between domains. We define *backlink co-amplification* as the number of websites that are linked at high volumes by the same domains. For an unweighted adjacency matrix $A \in \mathbb{R}^{m \times n}$ overlap $\mathcal{O}$ can be calculated using $\mathcal{O} = A^T A$. However, when adjacency matrices are weighted, matrix multiplication can cause the metric to lose interpretability. Intuitively, if a domain was linked to by i 10 times and by j 100 times, our co-amplification score should not exceed 10. For a vector of weights, it does not make sense for overlap to exceed the total sum of inbound or outbound links present in either of the two domains. Formally, we want to constrain overlap for domain $u \in \mathbb{R}^d$ and domain $v \in \mathbb{R}^d$ so that $\mathcal{O}_{u,v} \leq \sum_{i=1}^d u_i \wedge \mathcal{O}_{u,v} \leq \sum_{i=1}^d v_i$. To satisfy this constraint, we define a co-amplification score as the overlap of domains u and v as the sum of the minimum pair-wise overlap: $\mathcal{O}_{u,v} = \sum_{i=1}^d \min(u_i, v_i)$. The diagonal of the matrix was zeroed out to remove self-links. Intuitively, if two domains are heavily linked to by the same websites, this score will be high. If not, the score will be low.

¹² <https://developers.google.com/search/docs/advanced/guidelines/link-schemes>