



POWERED BY

ISD | Institute
for Strategic
Dialogue
CASM
technology

Information Warfare and Wikipedia

Carl Miller, Melanie Smith, Oliver Marsh,
Kata Balint, Chris Inskip, Francesca Visser

Contents

Executive Summary	5
Introduction	6
I. The Evolution of Information Operations	8
II. Threats to Wikipedia	9
III. Responses	13
Counter-Vandalism and Investigations Communities	13
IV. Case Study: The Article on the Russo-Ukrainian War	15
‘Not here to build an encyclopaedia’	15
Data Collection: 86 Banned Editors of the Article on the Russo-Ukrainian War	15
Identifying pages that have been targeted	15
State-Affiliated Sourcing	17
Detecting Possible Coordination through URLs	20
Linking state-affiliated domains to Wikipedia pages	22
Limitations and Future Work	23

Acknowledgements

The report authors would like to thank the expert interviewees who provided invaluable insights and helped inform our research. Notably, Stephen Harrison, a journalist for Slate; Piotr Konieczny, Wikipedia researcher and active member of Wikipedia for fifteen years; Anton Protsiuk, Programmes Co-ordinator for Wikimedia Ukraine and a volunteer editor and admin since 2012; Nazar Tokar, an administrator on the Ukrainian Wikipedia and Patrick Hilsman, a journalist and researcher of the Syrian conflict.

Executive Summary

Joining air, sea, land, space and cyber, information is increasingly seen as a theatre of war. In this report, ISD and CASM Technology set out to examine the ways in which Wikipedia may be vulnerable to the forms of systematic manipulation that have been exposed on Facebook, Twitter, YouTube, Reddit and a number of other information spaces. The report combines a literature review on publicly available research and information around Wikipedia, expert interviews and a case study.

For the case study, the English-language Wikipedia page for the Russo-Ukrainian war was chosen, where accounts that edited the page and have subsequently been blocked from editing were examined. Their editing behaviour on other Wikipedia pages was mapped to understand the scale and overlap of contributions. This network mapping has seemed to identify a particular strategy used by bad actors of dividing edits on similar pages across a number of accounts in order to evade detection. Researchers then tested an approach of filtering edits by blocked editors based on whether they add references to state-media affiliated or sponsored sites, and found that a number of edits exhibited narratives consistent with Kremlin-sponsored information warfare. Based on this, researchers were able to identify a number of other Wikipedia pages where blocked editors introduced state-affiliated domains, which helps spotlight various regions of Wikipedia that might be investigated more closely.

Introduction

Information environments are being reshaped around the world by the idea that they are theatres of war. 'Information warfare' conceives of information not as a tool or an act, but as a contested space that war occurs within. Whether for geopolitical advantage, electoral success, the promotion of certain ideas, or for economic gain, within this idea of warfare the value of information is purely instrumental, a way of achieving the ulterior ends of changes in attitudes, beliefs, identity and behaviour.

One of the tools of information warfare is disinformation, and huge amounts of academic research and journalistic investigation have exposed disinformation in its various forms across social media platforms. But joining disinformation are a suite of other techniques that can be less obvious: platform manipulation, inorganic amplification, the exploitation of false identity, malicious automation, coordinated inauthentic behaviour and a list of others that continues to grow.

Campaigns involving these techniques happen for a wide number of different reasons, from influencing elections and court cases to terrorist attacks, diplomatic summits, global conversations around climate action and, of course, in support of military action occurring in its more conventional forms. And whilst attribution of these campaigns has always been difficult, it is likely that the perpetrators are just as varied, ranging from military and cyber-operations groups to extremist mobilisations, cyber-criminals, conspiracy theory 'investigators', illicit advertising consultancies, and others.

This report is a short contribution to the discussions around the threats posed by information warfare. Whilst a great deal of attention has been placed on detecting illicit campaigns on mainstream social media platforms such as Twitter and Facebook, much less has been spent on detecting information warfare elsewhere. For this paper, we look at arguably the most epistemically consequential of all the possible targets of information warfare: Wikipedia.

Wikipedia has a unique position in the global information ecosystem. It is the largest source of truth that humankind has ever created, visited by over a billion people every month and (in a 2014 poll) trusted more than BBC News.¹ It is the most common destination following any Google search, its content populates Google's knowledge panels,² and is the source tech giants themselves are turning to for help confronting disinformation on their own platforms.³ Wikipedia is also a crucial destination for people to visit to access credible information about events as they happen, and pages on these events can see continuous editing to keep them updated as new information becomes available.

Wikipedia has been famously resilient to vandalism. All edits are open, vandalism can be rolled back quickly, pages can be locked and protected, and the site is patrolled by a combination of bots and editors. However, the question that remains is: how vulnerable is Wikipedia to information warfare which might use subtler methodologies and be executed over longer lengths of time?

This report combines a literature review on publicly available research and information around Wikipedia, expert interviews, and case study research to discuss three related areas:

- A look at the likely threat vectors to Wikipedia from actors wishing to undermine its information integrity in coordinated ways;
- A summary of the known responses to, and mitigations against, these threats;
- A case study that scopes approaches to detect possible influence operations on Wikipedia at scale.

This study is indicative rather than exhaustive. The interviews do not try to be representative of Wikipedia's editor-base and should be read as the views of a small number of expert individuals who have first-hand experience of many of the issues engaged with in this report. Likewise, the desk-based research and case studies are not representative of

all Wikipedia, but are indicative of areas we believe deserve more research, scrutiny and visibility.

We, the authors also recognise that Wikipedia is not just an encyclopaedia but the community that protects and expands it. None of the authors are Wikipedia editors; indeed we spend our time working to detect information operations occurring in online venues that are very different from Wikipedia. We hope that the perspective we offer here may be helpful precisely because it is an external one that crosses information domains, but we do not claim any deep expertise or experience in Wikipedia itself.

I. The Evolution of Information Operations

Over the last few years, the general tradecraft in information warfare has developed quickly, as far as we can track it. So before turning to the question of Wikipedia we first foreground some of the emerging techniques and strategies seen in contemporary campaigns identified on Facebook,⁴ Twitter, and YouTube,⁵ Weibo, Wechat, and TikTok.⁶

Manufacturing Credibility

This can include the creation of fake experts to validate a specific narrative,⁷ the involvement of specialised companies,⁸ or (see below) the manipulation of legitimate research.⁹ Recent investigations have also identified sophisticated strategies which have been linked to individuals connected to states and state-owned companies and infrastructures. A November 2021 report published by Meta¹⁰ uncovered a network of over 500 accounts on Facebook and Instagram behind a fake Swiss expert, who alleged that the United States was putting pressure on World Health Organization scientists studying the origin of COVID-19 to blame the virus on China. These stories are then amplified by state media networks and social media networks that are a mix of inauthentic accounts and - in this case - authentic accounts that belonged to employees of Chinese state infrastructure companies. The 2020 PeaceData campaign targeted freelance journalists, offering commissions for them to produce stories that were then amplified by a network of conspiracist sites.¹¹

Collaborations between State and Private Sector Actors

Information operations are not monolithic and can involve collaborations between actors, sectors, specialisms and geographies. Another investigation looked into Chinese government public procurement databases and revealed how the Chinese government has been contracting out social media influence work to private internet companies. The investigation found that government agencies have been hiring professional companies to publish positive stories about the government, local culture, and other aspects of life in China.¹²

Distorted Research

Information operations increasingly invest more effort and skill in the distortion of evidence to conform to desired narratives. A distorted version of legitimate research on the dispersal of radioactive material released into the Pacific Ocean after the Fukushima disaster, for instance, was spread and disseminated by various pro-China sources after being initially manipulated. An edited version of the original video was propagated online to criticise Japan after they announced a plan to release contaminated wastewater from the Fukushima nuclear plant in the next two years.¹³

The Amplification and Promotion of Western Voices

This often occurs within campaigns when some Western voices align with the aims of the influence operation. This can involve using diplomatic and state media outlets to amplify interventions from specific Western voices, and to actually conduct and translate interviews. Investigations have provided evidence that a number of these Western influencers have accepted financial support from autocratic state entities for producing content, but have denied that such support has influenced their material.¹⁴

Hacking or hijacking legitimate accounts.

Information operations increasingly tap into more classical cyber-offensive operations, including attacks on the social media accounts of elected political figures and the acquisition or purchase of longer-term compromised accounts.¹⁵

II. Threats to Wikipedia

Alongside broader trends in information operations, next we look at the known platform-specific threats to Wikipedia identified through expert interview and literature review.

Undisclosed Paid Editing

One of the main issue areas identified by our interviewees is undisclosed paid editing. Especially within the industry of 'reputation protection' or 'reputation management', companies offer services to confront and correct damaging information across a number of platforms, including Wikipedia. A 'necessary evil' as one interviewee put it, Wikipedia grudgingly allows paid editing even though the platform was not created for this purpose. It is prohibited however when the editing otherwise violates Wikipedia's policies and principles, misuses powers of office (see below) or is, in some cases, undisclosed.

Experienced editors can be directly contracted to perform paid edits, but they risk losing their reputation within the community. On the other hand, there are also larger companies that specialise in larger-scale editing projects. One interviewee described this phenomenon as a big problem on Ukrainian Wikipedia, mentioning a specific company offering paid editing services, which runs a sophisticated operation to circumvent the mechanisms in place that are meant to block undisclosed paid editing behaviour. This company, the interviewee alleged, also blackmails editors, and sometimes when an article is nominated for deletion pays editors to withdraw the deletion. He argues that there are no known state-backed operations carried out by this company, but they work with whoever will pay for their services.

Another interviewee stated that it is also common for paid editors to pay low-quality news sites to publish an article, which then can be used as a source to justify the claims made on Wikipedia. There are also instances when these paid editors get administrator or patroller rights, however, in the experience of the interviewees, most of them

lost their rights after it was revealed that they were paid editors.

Adversarial Editing

An associated phenomenon is the concerted and coordinated attempt to edit Wikipedia for ideological and political reasons. These might be to celebrate or promote a particular linguistic group or to 'get the truth out' about a particular conflict or controversy. Such editing becomes adversarial when these identities, ideologies or interests collide with others, resulting in continuous revisions and counter-revisions across a page or number of pages over time.

Whether independent, paid-for, or state-backed, bad actors can use various tactics to improve (i) the speed and scale at which they can make adversarial edits and (ii) the probability that edits will survive without another editor removing them. Examples taken from external research¹⁶ include:

Tactic	Description
Tag teams	A set of users coordinating to introduce fake content by exploiting the weakness of communities and systems.
Sock-puppets	Multiple online identities used for purposes of deception.
Click farms	Where a large group of low-paid workers are hired to perform microtasks to deceive online systems.
Deep fakes	Creation of fake videos or images through AI.
Data voids	Exploitation of missing data to manipulate research results.
Edits in languages other than English	Manipulation of content in languages other than English with a scarce editors or a homogeneous editing community that would make the Wikipedia project more vulnerable to political biases.
Hoax articles or facts	Creation of a deliberately fabricated falsehood. Hoaxes on Wikipedia include hoax articles or hoax facts that can be embedded in an otherwise accurate article.

State Editing

There are few known instances of illicit behaviour on Wikipedia clearly attributed to a state. Perhaps the clearest attributions are edits made from known Government IP addresses, and a number of bots on Twitter monitor this activity, highlighting incidents when they occur.¹⁷ These edits do not imply any sort of coordinated or concerted campaign, and IP addresses can be easily spoofed or obscured.

The clearest indication of a more concerted and strategic attempt to change Wikipedia by a state came in September 2021, when the Wikimedia Foundation banned seven editors linked to a mainland China group for editing Wikipedia articles with the objective of promoting “the aims of China”.¹⁸ Wikimedia notably called this behaviour an “infiltration” which “threatened the very foundations of Wikipedia”.

Unreliable Sources

Sourcing on Wikipedia is extremely important. Wikipedia articles must not contain original research; all claims, facts, ideas, allegations require support from external sources recognised as reliable by Wikipedia’s community. They include peer-reviewed journals, books, and mainstream news. Most interviewees agreed biased sources on the platform are a concern. Wikipedia has built lists of banned sources in order to mitigate this risk.¹⁹ These include Russian and Chinese state media as well as Western sources viewed by the community to be of dubious quality or strong political bias.

These source lists, interviewees said, differed across different language-versions of Wikipedia, and sources from one list were not necessarily banned on another, whilst editors may still advised not to use them as references. Interviewees reported that biased references can therefore still persist, either because of the lack of experience of an editor or by deliberate violation. Sputnik, RT, TASS Russian News Agency, and China Daily sometimes appear as sources on Wikipedia pages. One interviewee noted that when an editor consistently uses dubious

sources with either bad journalistic standards or skewed, ideologically-driven motivations, it is a good indicator of paid editing.

Gaming Wikipedia’s Governance, Administration, Policies, and Enforcement

Wikipedia’s governance practices, as explored below, provide an intricate and sophisticated defence against vandalism and platform misuse. However, these practices can themselves be a target of attack. At least in theory, the nature of this threat is not vandalism but entryism: states (or other bad actors) mounting organised, long-term, strategic attempts to infiltrate the community of administrators and Wikipedia office-holders (bureaucrats, check-users, SPI clerks and others, see below) who enact Wikipedia’s policies.

It is conceivable that states might support actors to supply a large number of legitimate and constructive edits to Wikipedia with absolutely no visible link to each other. Over time, these actors would earn trust from the community. They might then engage in the consensus and poll-based processes to gain positions of authority within the community and become active in discussions around Wikipedia’s policies; including those around legitimate sourcing, contentious political topics, even the identification of sock-puppets and sanctions. This threat vector, different from vandalism, would allow states to change Wikipedia in ways that the system allows, and indeed to change what the system allows.

Perhaps the most explicit example of an entryist attack on Wikipedia’s community is the one already mentioned relating to Chinese state-linked editors, prompting Wikimedia’s intervention to ban seven users and the removal of administrator powers from a further 12 in September 2021. In that case, Wikimedia described a security risk that “relates to information about infiltration of Wikimedia systems, including positions with access to personally identifiable information and elected bodies of influence ... [and] people deliberately seeking to ingratiate themselves with their communities in

order to obtain access and advance an agenda contrary to open knowledge goals.”²⁰

Attacks against the Wikipedia Editor Community

Wikipedia’s main defense against all the threats mentioned above is to ensure there are enough willing, able, and good-faith volunteers to detect and respond to them. A number of interviewees who have been involved in (in their view) protecting key pages from systematic attempts to change their content, report however that this undertaking is not particularly creatively fulfilling, and may be protracted and frustrating.²¹ Members can burn out and either become less active on the platform or leave the community altogether.

Attacks on pages are not confined to the edits themselves, but also attacks on the editors who work to reverse them. This can involve:

- Leaving aggressive messages on the talk pages of other editors;
- Tarnishing the reputation of editors amongst other Wikipedians;
- Leaking (and/or threatening to leak) personal information;
- Finding and harassing family members.

The Wikimedia Foundation put in place policies and set up a Trust and Safety team in order to mitigate these harms.²² However, this was an area of particular concern (and personal experience) for interviewees. Interviewees argued that policies in place aimed at countering bullying do not work. One interviewee argued that “if you accuse someone of being a state agent, you’re more likely to be sanctioned than they are.” They also suggested that “good-faith actors are more vulnerable to harassment than anonymous trolls,” as the latter can create armies of sock accounts, which are used to attack good-faith actors with the aim of driving them away from Wikipedia.

Non-English Wikipedias and ‘Community Capture’

Official Wikipedias have been created in 325 languages, and 314 are active.²³ As one interviewee argued, these Wikipedias can be divided into three groups in this regard:

1. The English language Wikipedia, which is the largest and has the highest readership. It is the primary battlefield, as it can have significant effects in shaping public opinion on certain issues;
2. Wikipedias in less problematic regions or languages, where there is most likely no major malicious editing, as the states are less likely to engage in such behaviour;
3. Non-English Wikipedias in more problematic contexts where they could be wide-scale attempts to control the narrative.

Some Wikipedias, an interviewee argued, were vulnerable to ‘community capture’. Due to the scarcity of editors or the homogeneity of the editing community, a Wikipedia that has been ‘captured’ in this way becomes expressive of the majority views of a particular ethno-linguistic or national group, regardless of whether these views are recognised outside of that group. An example of this is the Croatian-language version of Wikipedia, which was subject to an investigation by Wikimedia.²⁴ The investigation found that for over a decade, a group of ideologically driven editors and administrators had distorted content in line with the narratives of political organisations and groups of the Croatian radical right, whitewashing crimes committed by Croatia’s Nazi-allied Ustashe regime during the Second World War and promoting a fascist worldview.²⁵

One of our interviewees believes that the Chinese and Russian regimes are mostly interested in the English language Wikipedia as well as their own languages, but that they could be carrying out coordinated editing activity across other regions as well, including Ukraine, Estonia, Latvia, and Lithuania (in the case of Russia). Pro-Russian undisclosed

paid editing also occurs not only on the Russian and Ukrainian Wikipedias, but other language Wikipedias as well. He thinks paid editors were trying to block renaming the article on Kyiv, the capital city of Ukraine, so that the article remains with the incorrect spelling. Another example of pro-Russian paid editing was the editing of an article on timezones to include a map that depicted Crimea as being part of Russia. The interviewee believes there is likely paid editing occurring around topics related to the history of Russia and Ukraine.

III. Responses

In order to fight vandalism and manipulation, amongst a great many other reasons, Wikipedia has developed a rich and copious network of governing bodies, policies, guidelines, resolution and arbitration processes and communal, consensus-seeking spaces for public discussion and debate.²⁶

The governance structure is intricate. The Wikimedia Foundation owns Wikipedia and has a governance structure similar to many charities, with a board of trustees, and a body of staff and contractors split across a number of functions. However, Wikimedia is not involved in either Wikipedia's policies or how they are implemented and enforced. This is left to the community, a self-organising, self-governing collection (in its loosest sense) of 43 million registered users across the world.²⁷

Counter-Vandalism and Investigations Communities

There are a number of roles and bodies in Wikipedia that are more specialised in responding to systematic, coordinated and bad faith editing.

The **Counter-Vandalism Unit** is a Wikipedia project aiming to identify and undo vandalism on the platform. Since the Unit is open for anyone to join, it has numerous volunteer members, who are regularly

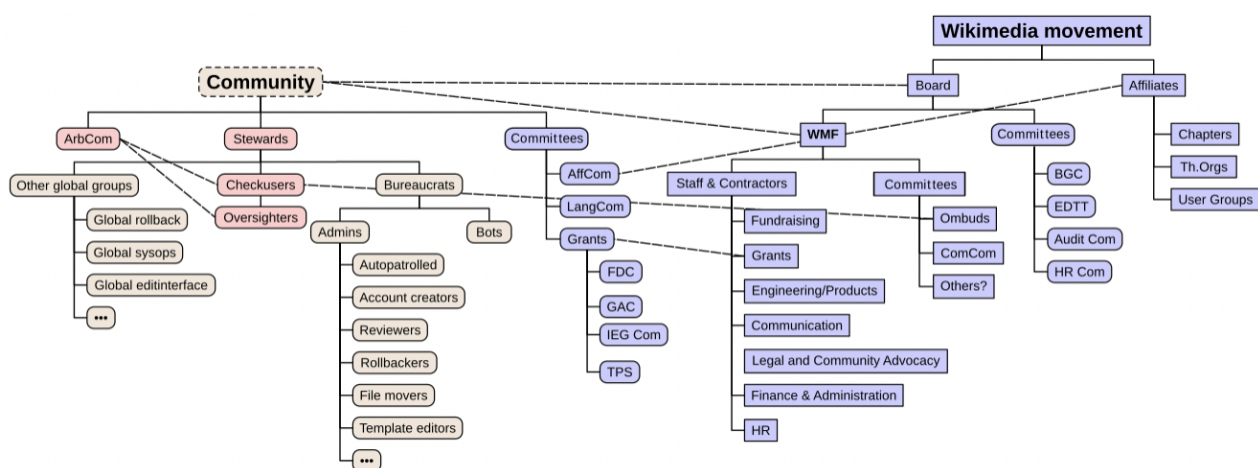
informed about software updates as well as policy changes.²⁹ The Unit also organises a **Counter-Vandalism Academy** to train new or inexperienced users in identifying manipulation of content.³⁰

Similarly to the Counter-Vandalism Unit, though with a much smaller membership, the **Counter-vandalism Network** also works to prevent and clean up vandalism, specifically on Wikimedia Foundation wikis and Fandom wikis. The Network develops tools to monitor changes in real-time, assists in the removal of vandalism, provides advice and offers semi-automated coordination for members working around these issues, such as bots.³¹

Sock Puppet Investigations. Wikipedia permits users to own multiple accounts but using multiple accounts in certain ways is understood as 'abusive sock puppetry'. This includes behaviours that are disruptive, abusive or misleading, such as multiple voting in polls to circumvent policies or deceptively influence debates.³²

Any editor can file for the opening of a sock puppet investigation. They are required to gather evidence and to present this to the community. If the evidence is seen to have merit, an investigation is opened,

The administrative structure of Wikipedia²⁸



conducted first by an administrator or clerk, who is an editor typically specialised in sock puppet investigations. They look for behavioural evidence which, if found, escalates the case to another kind of account: a CheckUser.

The **CheckUsers** team is a small group of 52 users, including volunteers and staff of the Wikimedia Foundation, who have access to the CheckUser tool.^{33, 34} The tool helps determine the IP addresses as well as other restricted technical information of specific user accounts in a set of limited circumstances.

IV. Case Study: The Article on the Russo-Ukrainian War

‘Not here to build an encyclopaedia’

Any detection regime must distinguish between behaviour likely to be illicit from behaviour which is not, and ultimately this is often a question of motivation. However, any number of different underlying motivations can manifest as the same surface-level behaviour online. Ideologically-driven editing can appear on Wikipedia as indistinguishable from commercially-driven edits and those done on behalf of states. Whether something is done for personal gain or genuinely held political conviction, by pride in any kind of identity, or for propagandistic reasons - or indeed any mixture of these - may not always even be clear to the actors themselves.

Discerning motivation often requires more attribution and contextualisation than scaled research methods allow. However, Wikipedia itself has a useful umbrella concept to distinguish between good-faith and bad-faith behaviour on the platform which it calls “here to build an encyclopaedia”. Editors are “here to build an encyclopaedia” when they show genuine interest in improving articles, respect for Wikipedia’s policies and a desire to personally improve, take feedback and learn lessons as they go. Editors are “not here to build an encyclopaedia” when they treat editing as a battleground, try to game the platforms and its policies and essentially have any long-term agenda or interest which is inconsistent with Wikipedia’s overall mission of amassing encyclopaedic knowledge.

Our case study aims to explore the behavioural patterns of editors already identified as “not here to build an encyclopaedia”. Starting with a single ‘seed page’ - the English-language Wikipedia page for the Russo-Ukrainian war - we intend to identify the editors on this page that have been banned, describe the edits they have made, look at where else they have edited on Wikipedia and whether any of this behaviour might lead us to new detections.

Data Collection: 86 Banned Editors of the Article on the Russo-Ukrainian War

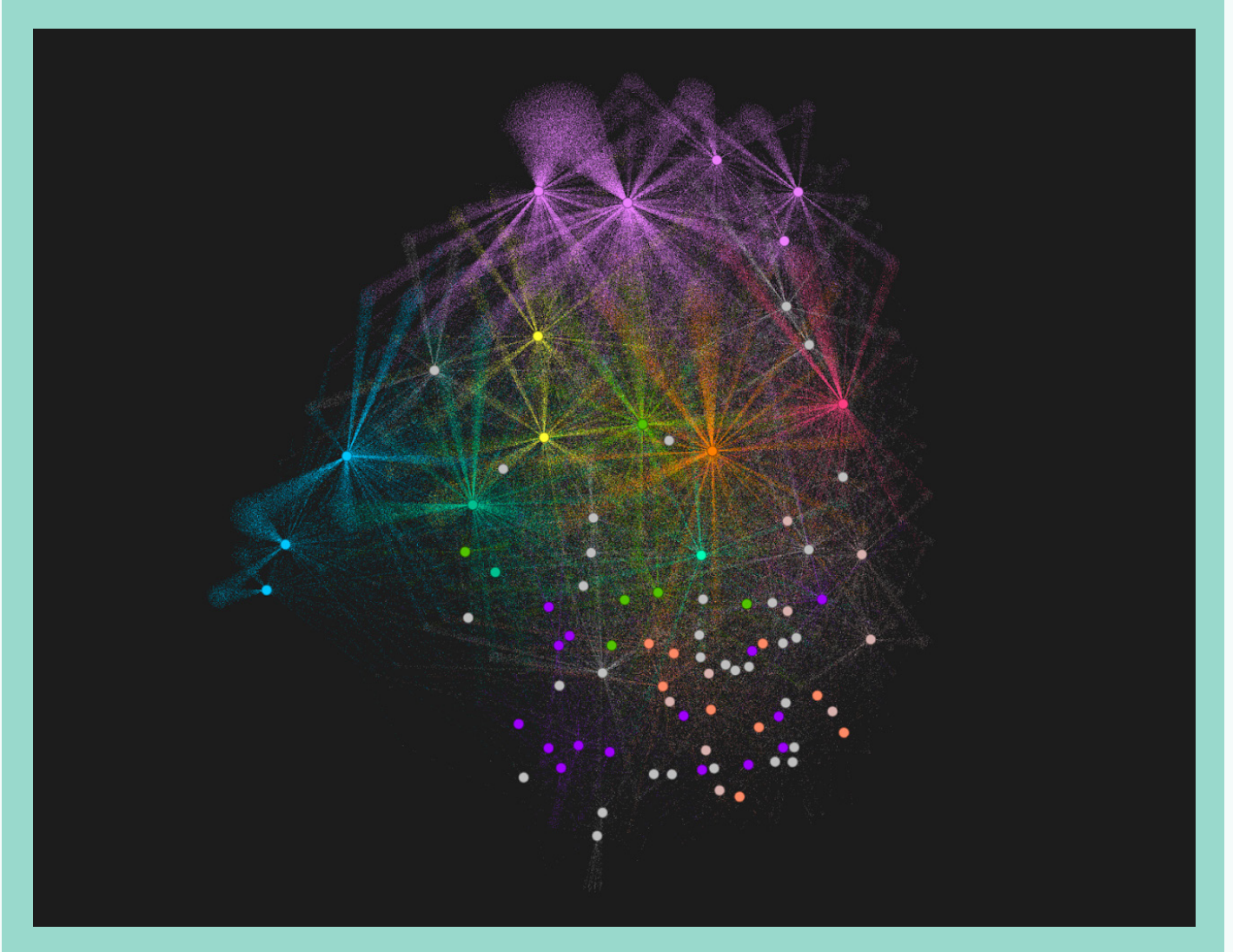
To begin the analysis, we focused on a single page: the English-language Wikipedia page for the Russo-Ukrainian war. As of 1st March 2022, there were 7,905 edits by 1,766 users to this page, dating back to its creation in early 2014.

89 editors who have made changes to this page have been blocked from further editing activity through the result of Wikipedia’s processes and policies. A variety of reasons have been given for these blocks: violating Wikipedia’s rules through vandalism, abuse and personal attacks; copyright violations; compromised accounts; and using multiple accounts (sock-puppets). Three of these accounts are self-classified as bots that were blocked due to malfunctioning. This study focuses on the remaining 86. 41 of these had been blocked due to suspected sock puppet activity, for using multiple accounts or otherwise as the result of a CheckUser investigation (see above), and a number of others for abuse to other editors, the persistent addition of unsourced claims, vandalism or because the account was compromised.

Identifying pages that have been targeted

Our analysis begins at the broadest possible level, by mapping the other pages that the 86 blocked accounts also edited.

In total, the 86 blocked editors have made 794,771 revisions made across 332,990 pages; so to visualise co-editing behaviour we created a network map showing blocked editors as large nodes, pages as small nodes, and edits between them as links. We filtered to only include pages that have been edited by at least two of the blocked accounts, creating a network of 86 blocked editors, 22,335 edited pages, and 53,552 connecting edits. The colours are automatically added by the network visualisation tool to help distinguish clusters of similar editors.



The map shows a number of highly connected editors in the centre and upper areas, indicating that they made edits to a large number of different pages. The lower region of the map shows editors that are less connected, making edits to fewer pages. Within this network, there are small, tight clusters of pages only edited by a small number of editors. Again, this does not imply that such editing is coordinated - indeed, it would be understandable for individual editors, at least, to focus on certain themes, regions or areas in which they are expert or interested. However, it does provide insight into the scale and overlap of contributions. One cluster shows a group of pages largely associated with Judaism and Poland.



Another cluster shows distinct editing around airports and commercial aviation, where there is a high degree of overlap by three separate editors.



A third cluster of pages is related to Iraq, Libya and the so-called Islamic state and involves accounts that are now known to be sock-puppets.



This network mapping may identify a particular strategy used by bad actors of splitting their edit histories between a number of accounts to evade detection. This was described by an administrator during a sock-puppet investigation:

“Accounts made similar edits to JA21 (diff1, diff2) and Khaby Lame (diff3, diff4) a few hours after each other. Their editing histories also show similar patterns, both contributing mostly to articles on airports, flight crashes and politics, but hardly ever the same ones. Example: Sarah Carvalho recently edited iPhone 12 (diff 5), while EricSDA made similar contributions to iPhone 13 (diff6) and iPhone 13 Pro (diff7). Possibly an attempt to avoid scrutiny by splitting editing history. — Ætoms [talk] 13:53, 27 November 2021 (UTC)”

State-Affiliated Sourcing

Next, we move from understanding broad editor- and community-level patterns to the question of how to analyse the Wikipedia edits themselves at scale. An edit is more complex to study than a Tweet or a Facebook post, because each act can involve not only the addition of content but also its relocation or deletion, often in combination.

One of the threats to Wikipedia (as identified through the interviews) is the use of suspicious or tendentious sources. We first, therefore, tested an approach of filtering edits by blocked editors based on whether they add references to state-media affiliated or sponsored sites.

The 86 blocked editors are responsible for conducting 681 edits on the Russo-Ukrainian war, or 8.61% of the total. Across these edits, links containing 347 separate domains were added at least once. Each edit was manually evaluated by analysts to determine whether the domain that featured was a state-affiliated source, using the Alliance for Securing Democracy’s³⁵ collection of state-sponsored or affiliated news and information channels and the Media Bias / Fact Check website³⁶. When a domain was not present in either of these collections, analysts searched for other sources to verify the affiliation of the domain or examined the content of the source to understand the ways in which events related to the Russo-Ukrainian war were presented.

22 edits containing 37 domains were considered by analysts to be state-sponsored or affiliated. These were:

rostec.ru	mk.ru
libertyunyielding.com	kremlin.ru
gazeta.ru	news.kremlin.ru
e1.ru	ria.ru
kommersant.com	en.c-inform.info
ukrinform.ua	wnd.com
zavtra.ru	rt.com
ura.ru	lenta.ru
sudact.ru	sputniknews.com
pln-pskov.ru	southfront.org
top.rbc.ru	112.ua
islamnews.ru	msk.kp.ru
russian.rt.com	news.xinhuanet.com
vedomosti.ru	informationclearinghouse.info
theblaze.com	eng.kremlin.ru
en.kremlin.ru	presstv.com
volgograd.kp.ru	Voiceofrussia.com
itar-tass.comtass.ru	zik.ua

Of course, there are a number of reasons why any editor might add a link on Wikipedia. The team therefore manually assessed the edits containing these links, and found that 16 of these 22 edits were contentious, exhibiting narratives consistent with Kremlin-sponsored information warfare, with the following themes:

Casting doubt on objectivity

Altering language to minimise objectivity of pro-Western accounts and maximise objectivity of pro-Kremlin accounts.

Previous

```
[[Marc Weller (professor)|Marc Weller]], a
[[Cambridge University]] professor of
[[international law]], analyzed some legal
aspects of Russia's actions, concluding
that they are for the most part
contraventions of established law. He has
opined as follows:
```

Revised

```
[[Marc Weller (professor)|Marc Weller]], a
[[Cambridge University]] professor of
[[international law]], assumed that Russia
militarily intervened in Crimea and on this
basis analyzed some legal aspects of "the
intervention" concluding that they are for
the most part contraventions of established
law. He has opined as follows:
```

```
Russia contested that it has undertaken a
[[humanitarian intervention]] to protect
ethnic Russians in Crimea from threats to
their safety.<ref name="bbc-2014-analysis-
legal">{{cite news
```

```
Russia consistently states that it will but
has not yet undertaken military
intervention to protect people living in
the Ukraine and in particular ethnic
Russians living there.
```

```
<!-- No bolding per WP:BOLDTITLE --
>Following the events of the [[2014
Ukrainian Revolution]], a [[2014 Crimean
crisis|secession crisis]] began on
[[Ukraine]]'s Russian-leaning [[Crimean
Peninsula]]. In late February 2014,
unidentified armed forces began to take
over the [[Crimean Peninsula]] in
[[Ukraine]]. Ukraine and Western
commentators identified the gunmen as
Russian Special Forces<ref>{{cite news
|url=http://gulfnnews.com/news/world/other-
world/russian-special-forces-on-crimea-
frontline-experts-1.1299168 |title=Russian
special forces on Crimea frontline: experts
|date=March 4, 2014 |accessdate=March 4,
2014 |publisher=Gulf News }}</ref> and
allied paramilitaries.
```

```
<!-- No bolding per WP:BOLDTITLE --
>Following the events of the [[2014
Ukrainian Revolution]], a [[2014 Crimean
crisis|secession crisis]] began on
[[Ukraine]]'s Russian-leaning [[Crimean
Peninsula]]. In late February 2014,
unidentified armed forces began to take
over the [[Crimean Peninsula]]. Ukraine and
Western commentators branded the gunmen as
Russian Special Forces<ref>{{cite news
|url=http://gulfnnews.com/news/world/other-
world/russian-special-forces-on-crimea-
frontline-experts-1.1299168 |title=Russian
special forces on Crimea frontline: experts
|date=March 4, 2014 |accessdate=March 4,
2014 |publisher=Gulf News }}</ref> and
allied paramilitaries.
```

Alternative reporting

Adding material which supports Kremlin descriptions of ongoing situations (or removing material which does the opposite).

Previous

Revised

```
On January 29, the chief of Ukraine's
General Military Staff [[Viktor Muzhenko]]
acknowledged that Russian regular army
units were not involved in combat action in
the troubled Donetsk and Luhansk regions.
<ref>
[http://sputniknews.com/europe/20150129/1017
514425.html]</ref>
```

```
In September 2014 Russia's defence minister
called for the deployment of reinforcements
to the Crimea and southern Russia, citing
the worsening crisis in Ukraine and buildup
of foreign forces nearby. [[Sergei Shoigu]]
said the "situation in Ukraine has sharply
worsened and the foreign military presence
has increased very close to our border."
<ref>{{cite
news|url=http://www.breitbart.com/Big-
Peace/2014/09/16/Russia-to-deploy-
reinforcements-to-Crimea-area--
minister|agency=AFP|title=russia to deploy
reinforcements|date=16 September
2014|accessdate=16September2014}}</ref>
```

Historical narratives

Introducing topics which sway historical background toward pro-Russia narratives.

Previous

```
Russia does not recognize the newly installed
interim government in Ukraine, instead
recognizing ousted-President [[Viktor
Yanukovych]],{{efn|Walker (2014) "Putin also
insisted that ousted Ukrainian leader, Viktor
Yanukovych, was the legitimate leader of
Ukraine and that the "so-called" acting
president had no authority and the new
government in Kiev illegal."<ref name="walker-
the-guardian-2014-descend" />}}
```

Revised

```
Russia does not recognize the newly installed
interim government in Ukraine, instead
recognizing ousted-President [[Viktor
Yanukovych]], as the last democratically
elected and as the current legitimate
President.{{efn|Walker (2014) "Putin also
insisted that ousted Ukrainian leader, Viktor
Yanukovych, was the legitimate leader of
Ukraine and that the "so-called" acting
president had no authority and the new
government in Kiev illegal."<ref name="walker-
the-guardian-2014-descend" />}} Russia and
others accused the United States and the EU of
funding and directing the
[[2014 Ukrainian revolution|Crimean
revolution]].
<ref>http://www.presstv.com/detail/2014/02/06/3
49524/us-accused-of-funding-ukraine-
rioters/</ref><ref>http://rt.com/op-edge/us-eu-
interference-in-ukraine-402/</ref>
```

Adding Kremlin quotations and press releases explicitly into the page to increase the salience of pro-Russian arguments and viewpoints.

Previous

```
|accessdate=5 March 2014 }}</ref> The
Ukrainian military reaction has included a
[[mobilization]] of [[Armed Forces of
Ukraine|Ukraine's armed forces and
reserves]]. Western media reported that as
of 3 March, Russia had stated that its
troops would stay until the political
situation was "normalised"<ref
name="autogenerated2">{{Cite
news|url=http://www.bbc.com/news/world-
europe-26414600|title=Ukraine crisis:
Russia vows troops will stay|date=3 March
2014|accessdate=3 March 2014|publisher=BBC}}
</ref>
<ref>http://online.wsj.com/news/articles/SB1
0001424052702304360704579417151023498892</re
f> however Russia never stated that its
troops were involved in the conflict.
```

```
In mid-April, the US<ref name="StateDept"
/> accused Russia of responsibility for the
[[2014 pro-Russian unrest in
Ukraine|seizure of buildings in eastern
Ukraine]].
```

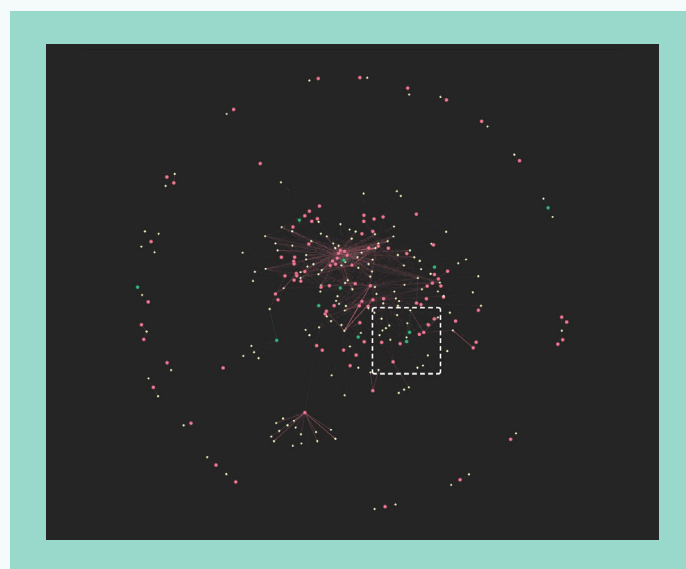
Revised

```
|accessdate=5 March 2014 }}</ref> Russia
accuses the United States and the EU of
funding and directing the Ukrainian
revolution.
<ref>http://www.presstv.com/detail/2014/02/0
6/349524/us-accused-of-funding-ukraine-
rioters/</ref><ref>http://rt.com/op-edge/us-
eu-interference-in-ukraine-402/</ref>
<ref>http://intellihub.com/india-sides-
russia-ukraine-crisis/</ref>
<ref>http://pando.com/2014/02/28/pierre-
omidyar-co-funded-ukraine-revolution-groups-
with-us-government-documents-show/</ref>
<ref>http://voiceofrussia.com/2014_03_06/US-
NATO-CIA-supporting-nazis-in-Ukraine-
project-2569/</ref> The Ukrainian military
reaction has included a [[mobilization]] of
[[Armed Forces of Ukraine|Ukraine's armed
```

```
In mid-April, the US<ref name="StateDept"
/> accused Russia of responsibility for the
[[2014 pro-Russian unrest in
Ukraine|seizure of buildings in eastern
Ukraine]], with the Russian Foreign
Ministry sharply criticizing the accusation
stating "The US Department of State
feverishly collects all speculations
distributed by the current Kiev authorities
in order to substantiate accusations
against Russia of instigating and all but
organizing disorders in Ukraine's
southeast". To one of the points of
evidence raised by the US, that the
militants are carrying Kalashnikov rifles,
it responded "As if the American foreign
policy department does not know that the
rifles are in service with the Ukrainian
army and law enforcement structures".
<ref>http://en.itar-
tass.com/russia/728304</ref>
```

Detecting Possible Coordination through URLs

The analysis shown above can be extended to show patterns where a number of editors consistently add the same links. To do so, we created another network map, but rather than showing edits, each edge captures the action by an editor of adding a state-sponsored or affiliated URL³⁷ to the page on the Russo-Ukrainian war. The links are represented as URL nodes in yellow, the 86 blocked editors in green and unblocked editors in red. Editors that are near each other on this map are likely to have added a similar profile of state-affiliated links to the page.



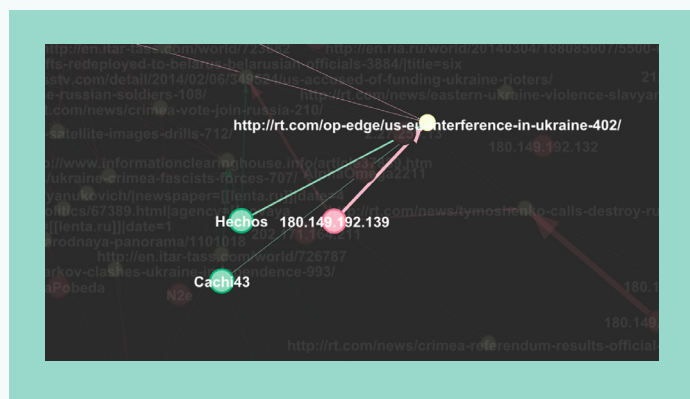
In the figures below, we again see how two known sock-puppet accounts are situated in close proximity within the network, here adding seven URLs introduced to the seed page.



Through this view, there is potential to identify additional editors that have added the same links as the blocked editors and query their edits. In this instance, we see an IP user introduce the same URL on the page as two known sock-puppets.



Querying the edits made by this IP user suggests edits mirroring almost verbatim the edit of the known sock-puppet user.



Known sock-puppet editor.³⁸

Russia does not recognize the newly installed interim government in Ukraine, instead recognizing ousted-President [[Viktor Yanukovich]],{{efn|Walker (2014) "Putin also insisted that ousted Ukrainian leader, Viktor Yanukovich, was the legitimate leader of Ukraine and that the

Russia does not recognize the newly installed interim government in Ukraine, instead recognizing ousted-President [[Viktor Yanukovich]], as the last democratically elected and as the current legitimate President.{{efn|Walker (2014) "Putin also insisted that ousted Ukrainian

Unblocked IP-editor.³⁹

summit-ukraine-thursday-175852481.html
[title=EU leaders to hold summit on Ukraine on Thursday – Yahoo News |publisher=Yahoo! News |date=3 March 2014]]</ref>

summit-ukraine-thursday-175852481.html
[title=EU leaders to hold summit on Ukraine on Thursday – Yahoo News |publisher=Yahoo! News |date=3 March 2014]]</ref> Russia and others accused the United States and the EU of funding and directing the revolution.
<ref>http://www.presstv.com/detail/2014/02/06/349524/us-accused-of-funding-ukraine-rioters/</ref><ref>http://rt.com/op-edge/us-eu-interference-in-ukraine-402/</ref>

Linking state-affiliated domains to Wikipedia pages

Last, we can begin to tie the above analyses together by identifying all the state-affiliated domains added by the 86 blocked editors across all the pages that they have edited and evaluating which pages have seen the greatest intensity of activity. Of the 794,771 edits made across 332,990 pages, 2,421 were found to introduce links of state-affiliated domains; these edits are made across 667 pages.

The table below shows the pages where relevant domains have been introduced at least 10 times. This procedure appears to provide a relatively focused subset of pages likely to be subjected to similar edit behaviour to the seed page, and can potentially serve as a high precision proxy for building out the network/ecosystem of pages under threat.

Of course, this does not necessarily identify coordination or strategic intent, but much like the research carried out above, it can spotlight various regions of Wikipedia that might be investigated more closely.

Page	Edits made containing state media
List_of_traffic_collisions_(2010–2014)	243
List_of_clashes_in_the_North_Caucasus_in_2010	211
List_of_clashes_in_the_North_Caucasus_in_2012	134
List_of_clashes_in_the_North_Caucasus	128
2014_Crimean_crisis	70
War_in_Donbas	43
List_of_clashes_in_the_North_Caucasus_in_2011	41
Russo-Ukrainian_War	39
2010_Pakistan_floods	24
2021_Formula_One_World_Championship	24
Republic_of_Crimea_(country)	23
Daniel_Ricciardo	19
Political_status_of_Crimea	18
2014_Winter_Olympics	16
Revolution_of_Dignity	16
Fan_Bingbing	16
List_of_earthquakes_in_2014	15
Spetsnaz	14
2014_pro-Russian_unrest_in_Ukraine	14
Luhansk_People's_Republic	13
Talk:List_of_sovereign_states	12
List_of_clashes_in_the_North_Caucasus_in_2016	12
RT_(TV_network)	12
List_of_terrorist_incidents_in_May_2016	12
Fake_news_website	11
Rape_during_the_Bosnian_War	10
October_2013_Volgograd_bus_bombing	10
Annexation_of_Crimea_by_the_Russian_Federation	10
AK-47	10
List_of_terrorist_incidents_in_June_2016	10

Limitations and Future Work

This short report could only focus on a small number of methodologies, and there are a number of limitations, unanswered questions and extension opportunities of our research we have identified.

- **Move from description to detection.** Our research did not set out to identify unknown suspicious activity, but rather to understand and describe activity that was already known to be suspicious. The next stage would be to deploy some of the methodologies described above on unknown data, in order to assess whether they are capable of identifying patterns that might result in new detections.
- **Comparison between blocked and unblocked behaviours.** Our case studies were dedicated to describing the behaviour of blocked editors. However, in order to discern patterns in a systematic way, these behaviours should be compared to non-suspicious baseline behaviours of unblocked accounts in order to empirically identify the areas that discriminate suspicious behaviours the greatest.
- **Deploy natural language processing technology to classify edits.** Most of our research is focused on editor- and community-level editing patterns, and comparatively little on the edits themselves beyond manual appraisal. However, a great deal of capability exists within the field of natural language processing to analyse edits at scale in order to understand possible patterns.

Endnotes

- 1 <https://yougov.co.uk/topics/politics/articles-reports/2014/08/09/more-british-people-trust-wikipedia-trust-news>
- 2 Google's knowledge panels are information boxes that appear on Google when you search for entities (people, places, organizations, things). They are meant to give a quick snapshot of information on a topic based on Google's understanding of available content on the web.
- 3 <https://www.wired.com/story/youtube-will-link-directly-to-wikipedia-to-fight-conspiracies/>
- 4 <https://demtech.oii.ox.ac.uk/research/posts/chinas-public-diplomacy-operations-understanding-engagement-and-inauthentic-amplification-of-chinese-diplomats-on-facebook-and-twitter/#continue>
- 5 Spamouflage Survives: CCP-aligned Disinformation Campaign Spreads on Facebook, Twitter, and YouTube (substack.com)
- 6 <https://freedomhouse.org/report/china-media-bulletin/2021/beijings-global-content-manipulation-plenum-propaganda-sports>
- 7 <https://www.bbc.co.uk/news/world-asia-china-59456548>
- 8 <https://chinatalk.substack.com/p/selling-chinas-story-how-the-chinese?s=r>
- 9 https://www.isdglobal.org/digital_dispatches/how-distorted-research-found-legitimacy-thanks-to-superspreaders-chinese-officials-and-state-media/
- 10 <https://about.fb.com/wp-content/uploads/2021/12/November-2021-CIB-Report.pdf>
- 11 <https://www.theguardian.com/technology/2020/sep/04/russia-media-disinformation-fake-news-peacedata>
- 12 <https://chinatalk.substack.com/p/selling-chinas-story-how-the-chinese?s=r>
- 13 https://www.isdglobal.org/digital_dispatches/how-distorted-research-found-legitimacy-thanks-to-superspreaders-chinese-officials-and-state-media/
- 14 <https://www.nytimes.com/interactive/2021/12/13/technology/china-propaganda-YouTube-influencers.html?referring-source=articleShare>
- 15 https://www.isdglobal.org/digital_dispatches/twitter-accounts-of-european-politicians-hijacked-to-spread-ccp-propaganda/
- 16 Diego Sáez-Trumper, (2019), Online Disinformation and the Role of Wikipedia. ArXiv, abs/1910.12596. <https://arxiv.org/pdf/1910.12596.pdf>
- 17 <https://www.wired.co.uk/article/russia-edits-mh17-wikipedia-article>
- 18 <https://www.bbc.co.uk/news/technology-58559412>
- 19 https://en.wikipedia.org/wiki/Wikipedia:Deprecated_sources
- 20 <https://lists.wikimedia.org/hyperkitty/list/wikimedia-l@lists.wikimedia.org/message/6ANVSSZWOGH27OXAIN2XMJ2X-7NWRVURF/>
- 21 Piotr Konieczny, (2018), Volunteer Retention, Burnout and Dropout in Online Voluntary Organizations: Stress, Conflict and Retirement of Wikipedians, in Patrick G. Coy (ed.) Research in Social Movements, Conflicts and Change (Research in Social Movements, Conflicts and Change, Volume 42) Emerald Publishing Limited, pp.199 - 219
- 22 https://meta.wikimedia.org/wiki/Trust_and_Safety
- 23 https://en.wikipedia.org/wiki/List_of_Wikipedias
- 24 https://meta.wikimedia.org/wiki/Croatian_Wikipedia_Disinformation_Assessment-2021
- 25 <https://www.fastcompany.com/90666412/non-english-wikipedia-misinformation>
- 26 <https://en.wikipedia.org/wiki/Wikipedia:Administration>
- 27 <https://en.wikipedia.org/wiki/Wikipedia:Wikipedians>
- 28 <https://en.wikipedia.org/wiki/Wikipedia:Administration>
- 29 https://en.wikipedia.org/wiki/Wikipedia:Counter-Vandalism_Unit
- 30 https://en.wikipedia.org/wiki/Wikipedia:Counter-Vandalism_Unit/Academy
- 31 https://meta.wikimedia.org/wiki/Countervandalism_Network
- 32 https://en.wikipedia.org/wiki/Wikipedia:Sockpuppet_investigations/SPI/Guidance#Defending_yourself_against_claims
- 33 <https://en.wikipedia.org/wiki/Wikipedia:CheckUser>
- 34 https://en.wikipedia.org/wiki/Wikipedia:Wikipedians#Number_of_editors
- 35 <https://securingdemocracy.gmfus.org/hamilton-dashboard/>
- 36 <https://mediabiasfactcheck.com/>
- 37 Where a domain is a broader site which can act as an umbrella for a collection of pages - for example rt.com - a URL on Wikipedia points to a specific page.
- 38 <https://en.wikipedia.org/w/index.php?diff=603376885>
- 39 https://en.wikipedia.org/w/index.php?title=Russo-Ukrainian_War&diff=prev&oldid=602034769

