

# WEAPONIZED INFORMATION OUTBREAK:

A Case Study on COVID-19, Bioweapon Myths, and the Asian Conspiracy Meme

PRESENTED BY  
**RUTGERS**  
Miller Center for Community  
Protection and Resilience

**Savvas Zannettou**, *Author*

The Network Contagion Research Institute  
Max Planck Institute for Informatics

**Jason Baumgartner**, *Author*

The Network Contagion Research Institute

**Joel Finkelstein**, *Corresponding Author*

The Network Contagion Research Institute  
The James Madison Program in American Ideals and Institutions, Princeton University  
joel@ncri.io

**Alex Goldenberg**, *Corresponding Author*

The Network Contagion Research Institute  
alex@ncri.io

**John Farmer**, *Contributing Editor*

Former New Jersey State Attorney General and Chief Counsel, 9/11 Commission  
Director, Miller Center for Community Protection and Resilience  
Rutgers, the State University of New Jersey

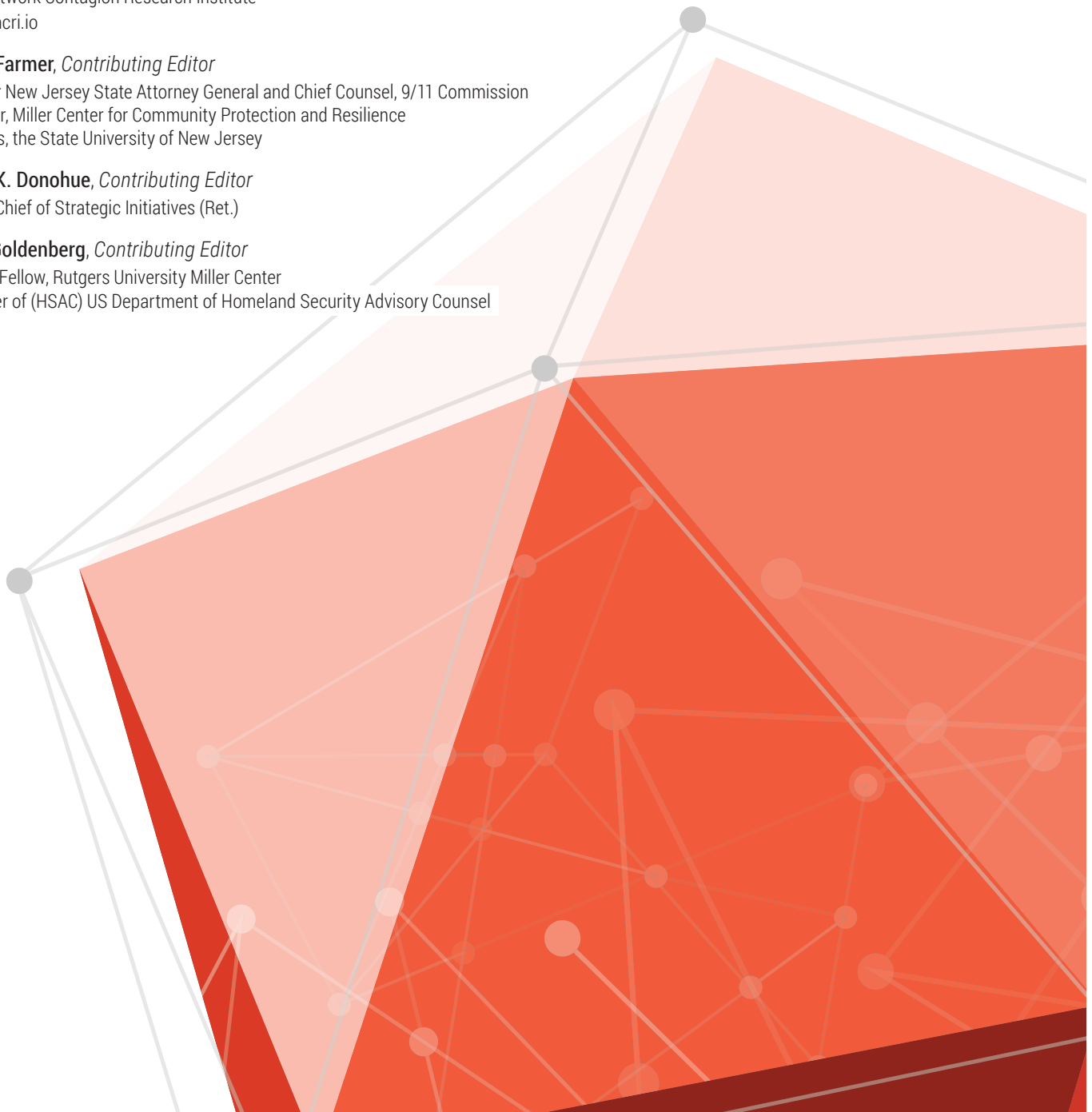
**John K. Donohue**, *Contributing Editor*

NYPD Chief of Strategic Initiatives (Ret.)

**Paul Goldenberg**, *Contributing Editor*

Senior Fellow, Rutgers University Miller Center  
Member of (HSAC) US Department of Homeland Security Advisory Council

POWERED BY  
**NETWORK  
CONTAGION**  
RESEARCH INSTITUTE



## INTRODUCTION

Violent, ethnic hate is exacerbated instantaneously by outbreaks of weaponized information during episodes of viral pandemics. Even as intelligence and law enforcement chart new waves of attacks against Asian citizens<sup>1</sup>, they remain challenged to understand how social media empowers the self-organization of these massive waves of violence from weaponized information.

As paranoia now rises around COVID-19, several major challenges prevent the protection of these vulnerable outgroups from ethnic hate, especially on social media. Objectively identifying which groups attract hostility in a timely fashion, cataloging cryptic, hateful language and specific conspiracies as they evolve and determining which communities and actors source hostility—all pose significant challenges to law enforcement and intelligence. In the face of these challenges it is becoming increasingly critical to illuminate how hate intersects conspiracies and weaponized information within communities because such research promises to be instrumental in countering both. However, absent clear indicators and contextual cues, traditional methods of open-source investigation are unequipped to respond to acute episodes of rapid and targeted hate at such a massive scale.

Using innovative methodology developed by NCRI affiliates and other colleagues in a recent report<sup>2</sup> to chart flows of targeted hate, we perform a case study on the spread of a “bioweapon conspiracy” being weaponized against Asian citizens to opportunistically promote violent sentiment around the COVID-19 pandemic on social media. The NCRI documents an outbreak of Sinophobic hate emerging with COVID-19 on 4chan, an influential and extremist Web community, which uses memes—inside jokes or coded images—to promulgate violence and conspiracies. We employ innovative methods to detect surges in Sinophobic hate in close to real time as it emerges from Web communities. We explore the case of the growth of a conspiracy which depicts China engineering COVID-19 as bioweapon—to illustrate how specific conspiracies become weaponized and spread for ethnic hate—and how this hatred capitalizes on COVID-19 itself. NCRI deploys topic networks<sup>3</sup> to expose an entire catalog of viral conspiracies and codewords around the bioweapon motif eliciting targeted ethnic hate toward Asians. Utilizing theme-subtraction methods, we more fully illuminate the relationship between the ethnic hate and this cryptic disinformation about bioweapons and unveil a more principled means to resolve anti-Asian coded conspiracies more objectively and in aggregate. Finally, we perform a preliminary investigation on Twitter and Reddit and find evidence suggesting contagion may be growing for the conspiracy in both communities.

The findings and methods in this case study are meant to showcase to platforms, law enforcement, intelligence, and civil-society organizations how the most innovative general methods to intersect hate on the social-cyber domain can be deployed for specific needs—as the need to detect, predict and intersect emerging threats from outbreaks of weaponized information expands in tandem with COVID-19. In these circumstances the NCRI’s massive data ingestion and analytic service, *Contextus*, now seeks to operate in close to real time to capture important underlying dynamics in outbreaks of weaponized information with flexibility across time, across topics, and across identities.

---

<sup>1</sup> <https://www.nbcnews.com/news/asian-america/asian-americans-report-nearly-500-racist-acts-over-last-week-n1169821>

<sup>2</sup> <https://arxiv.org/abs/2004.04046>

<sup>3</sup> Finkelstein, J., Zannettou, S., Bradlyn, B., & Blackburn, J. (2018). A quantitative approach to understanding online antisemitism. *arXiv preprint arXiv:1809.01644*.

## CONTEXTUS DETECTS SURGES OF ANTI-ASIAN, TARGETED HATE ON 4CHAN AND REVEALS COVID-19 AS THE UNDERLYING CONTEXT

We begin by detecting and documenting measurable increases in Sinophobia and anti-Asian sentiment over a large Web community, which we tie directly to the virus itself. Using *Contextus*, we monitor hostility against specific identities by measuring how identity terms become targeted by hateful associations in aggregate on Web communities. In figure 1, we describe large, sustained surges in Sinophobic terms on 4chan. In addition to this however, we also deploy targeted-hate detection to illuminate the underlying context of hostility as evident in figure 2. On over millions of comments, we witness the term “chink” becoming contextually closer in meaning to the term “virus” as the COVID-19 outbreak spreads, suggesting that the context for this increase in hate is the virus itself.

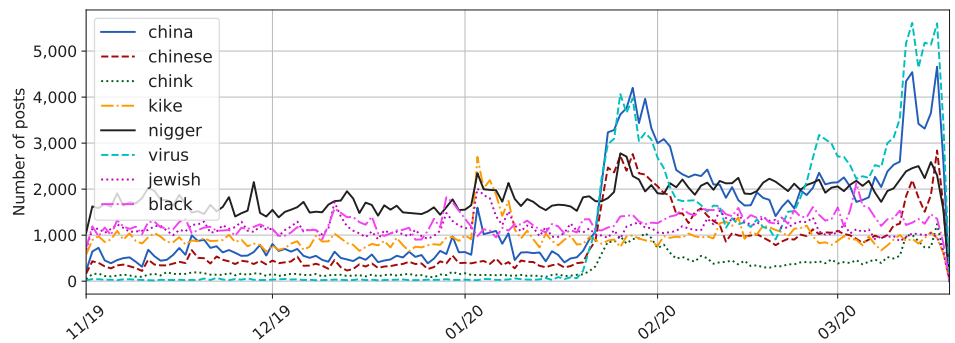


Fig 1. Timeline of identity terms and slurs over the past 120 days along with the word “virus”. Since January, “chink,” “china,” and “chinese” show sharp relative increases in use in tandem with “virus.” By contrast, terms and slurs for Jews and for African Americans remain stable, suggesting targeted hate increasing toward Asian communities. Terminal drops are artifacts.

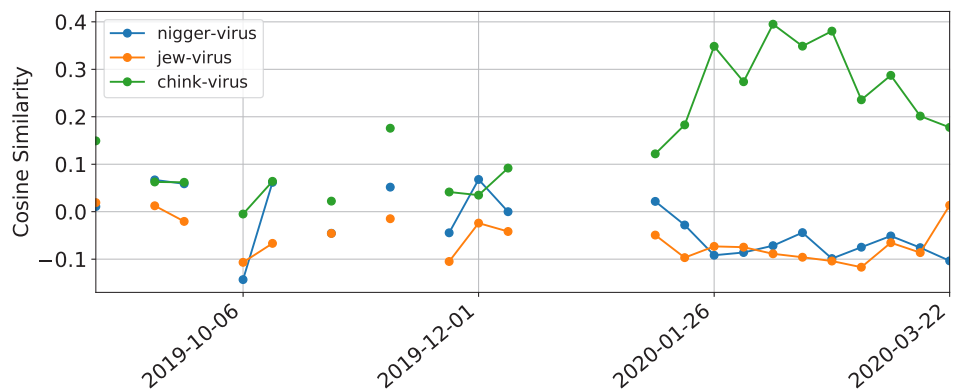
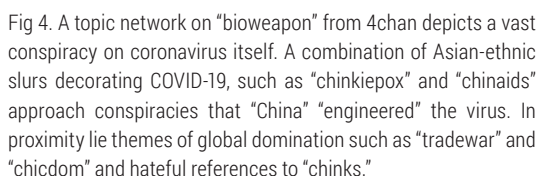
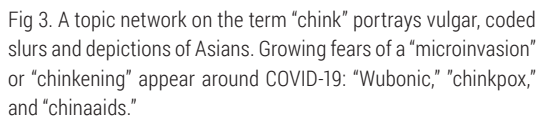


Fig 2. Contextual proximity of identity terms and slurs over the past 120 days to the word “virus.” Since January, “chink” shows sharp increases in relatedness to the virus. By contrast terms for Jews and a slur for African Americans show a slight decrease in relatedness. This technique allows the NCRI to reveal how flows of targeted hate evolve topically with COVID-19 using evidence-based approaches. Gaps in data reflect weeks with insufficient samples.

By contrast, terms related to Jews and derogatory terms for African Americans actually seem to show flat or decreasing relatedness to topics around the virus. At least on 4chan, these findings do not support recent reports which emphasize fears and dangers of ethnic blame around the virus towards other ethnicities. While these fears are rational, and likely to be at play in pockets of social media, this work recommends data-driven and evidence-based approaches as an important supplement to more qualitative investigations of individual anecdotes in resolving concerns around targeted hate. In the current volatility of the information ecosystem, qualitative reports, taken out of context, may serve to aid disinformation campaigns because they can overgeneralize threats.



We next deploy topic networks on *Contextus* to resolve the specific makeup of an Asian conspiracy theory and hateful code words which this method reveals. In figure 3, we use the seed term “chink” to examine an entire spectrum of anti-Asian propaganda and slurs and notably, an entire cluster of Corona-related material appears to have evolved recently in the community. We discover conspiracies connecting the racial slur, “chink,” to the virus. These include a “chinobyl” conspiracy, the accidental release of a designed bioweapon, from a “lab” in “wuhan” as well as “microinvasions.” The topic network thus contains elements of militarization and conspiracy around the virus. Indeed, according to our open-source intelligence investigation, the theme of the virus being militarized as a bioweapon comprises the most widespread of the Sinophobic conspiracy theories that we encounter.

## NCRI'S *CONTEXTUS*-THEME SUBTRACTION METHOD UNVEILS CODE WORDS AND DISINFORMATION IN AGGREGATE ON SOCIAL MEDIA

The NCRI thus deployed a theme-subtraction method on *Contextus* to create a more principled filter on Sinophobic hate as it becomes weaponized as information during the COVID-19 epidemic. *Contextus* uses embedding methods such as Word2Vec, and these methods help us evaluate language using statistical relationships between words. These embeddings permit the NCRI to use math on language in ingenuitive ways on any given corpus.

The most well-understood illustration of this technique is taking a word such as “king” and subtracting the word “man” from “king.” This mathematical operation would return gender neutral terms such as “ruler” or “monarch” as an output. We thus subtract the vector value of the word “chink” from the topic network of the word “bioweapon” and examine the resulting topic network. We hypothesize that subtracting the hateful slur will remove all of the Sinophobic contexts from the topic network. Remarkably, when we subtract the word “chink,” from the word “bioweapon” as evident in figure 5, the coronavirus itself disappears altogether in the topic network of “bioweapons.” In addition, all of the conspiracies around the coronavirus—its weaponization, use for invasion and multitude of Asian ethnic slurs—evaporate completely. This suggests that the association of coronavirus to conspiracy is contextually nested in Sinophobic hate on 4chan.



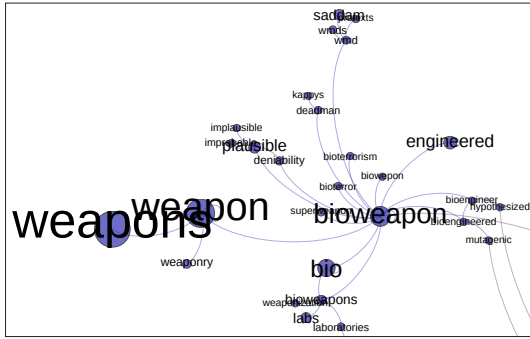


Fig 5. When we filter “chink” from the “bioweapon” topic network it eliminates all signatures of weaponized information, ethnic hostility and any mention of COVID-19 from the network, leaving a background of entirely mundane conversation as filtrate.

In addition, once we subtract “chink” from the “bioweapon” topic network, the remaining themes in the network read as factual descriptors of bioweapons and the process leaves no themes of weaponized information, conspiracy, or hateful content. Theme filtering represents an important capacity not just to identify the makeup and context of hate with more evidence-based methodology, but to share context and understanding as a public good to converge on common norms.

## THEMES OF ANTI-ASIAN BIOWEAPON CONSPIRACIES ARE INTACT AND SURFACING ON TWITTER AND REDDIT

Finally, we run a preliminary investigation on NCRI’s technology platform, *Pushshift*, using Twitter’s verified account feed and Reddit to analyze the prevalence of the term “bioweapon” among users (as seen in figure 6). Our initial results show a sharp uptick in the term corresponding with the emergence of COVID-19. Furthermore, a simple word-cloud analysis of “bioweapon” depicts that themes of the entire Asian conspiracy meme appear on these networks. While additional investigation is needed, weaponized information surfacing on mainstream platforms bears considerable risk for indoctrinating or radicalizing users who are more public facing and, ultimately, drawing them into contagions of ignorance and hate.

As Sinophobic, weaponized information leaves more extremist sites and traffics into mainstream communities, the risk of Asian conspiracy memes mainstreaming on a national scale becomes a concerning and increasingly likely outcome

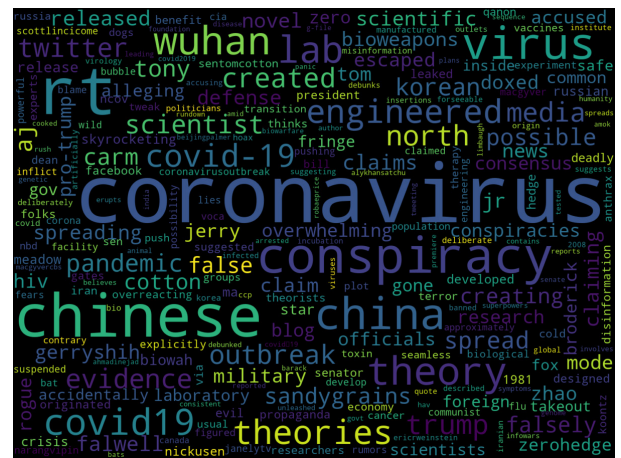
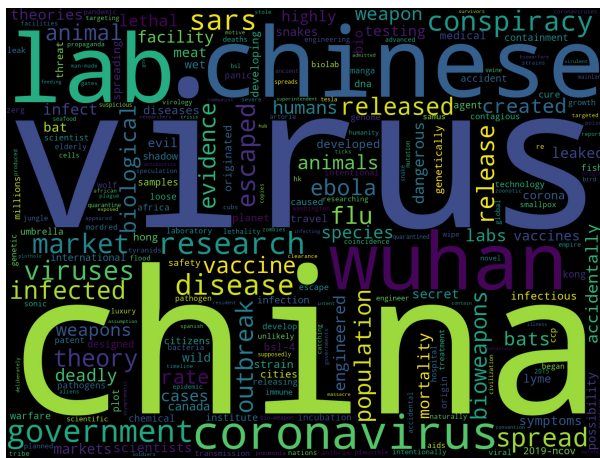
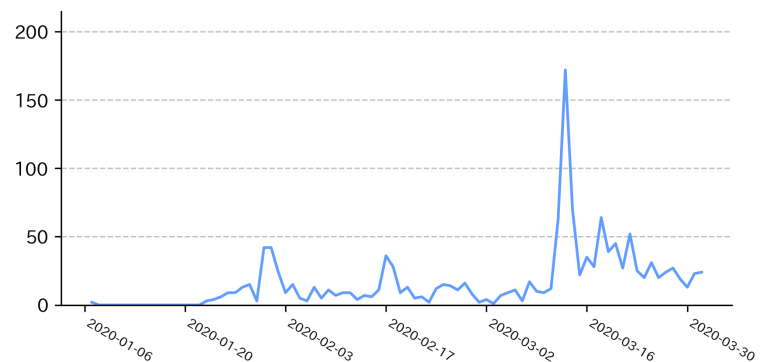
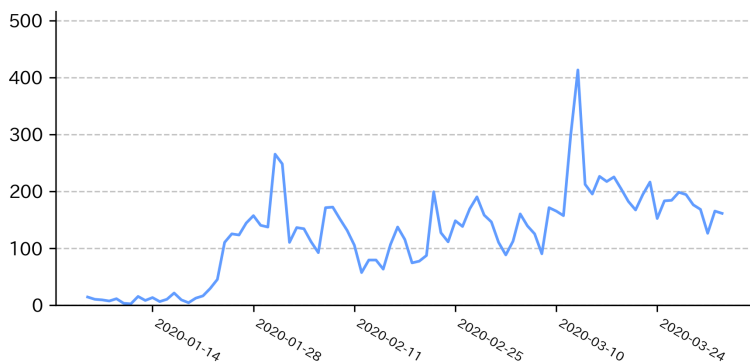


Fig 6. A timeline and word-cloud from Reddit (L) and Twitter’s verified feed (R) on the term “bioweapon” shows elements of the Asian conspiracy meme intact on both communities and surfacing in tandem with COVID-19.



Image of a sinophobic modified “Happy Merchant” meme. The “Happy Merchant” meme is one of the most popular memes on subcultural online forums such as 4chan. The meme traditionally depicts a drawing of a Jewish man with stereotyped facial features, but is often modified.



The Honkler, a cartoonish and racist figure, here unrolls the term “chink” while wearing a protective bio suite. 4chan is one of the most influential communities the NCRI studies for spreading cartoonish and cryptic memes such as this one.

## KEY FINDINGS


1. In conjunction with research premiering newly published methodology for intersecting targeted hate, the NCRI performs a case study on an outbreak of a Sinophobic and hateful bioweapons conspiracy emerging with COVID-19 on one of the most influential and extremist Web communities, 4chan; a trend highly likely to be reflected on other extreme Web communities.
2. Through *Contextus*, NCRI uses data-driven and evidence-based approaches to persistently and in close to real time, document acute increases in both the vitriol and magnitude of ethnic hate.
3. The NCRI can use topic networks to catalog violent conspiracies around COVID-19 and objectively profile and enumerate the most pronounced items of weaponized information.
4. Using theme subtraction, the NCRI publicly illuminates the relationship between ethnic hate and cryptic disinformation.
5. Using *Pushshift*, NCRI observes features of the Asian conspiracy meme surfacing on Twitter and Reddit, suggesting the growing presence of the conspiracy in the mainstream.

## CONCLUSIONS AND RECOMMENDATIONS

As economic conditions suffer and psychological factors such as isolation and anxiety grow increasingly in quarantine, the forthcoming outlet for organizing political hostility may increasingly become social media. As a conjoined threat, outbreaks of hate and disinformation on social media comprise unparalleled dangers to society in the face of actual viral pandemics, such as COVID-19. NCRI’s research indicates that hateful communities may serve as sources of spread for disinformation and propaganda during politically volatile events for purposes of hate. The hate and threats of ethnic violence which emerge from the cyber domain are highly complex and have moved beyond the playbook of civil society and law enforcement at every turn. Furthermore the disinformation poses unique dangers in its own right, not merely in the form of violence, but in the form of contagious ignorance about COVID-19. Outbreaks of weaponized information serve to attack public trust and undermine democratic institutions at a key moment of global vulnerability. The danger of weaponized information is that it may cultivate a communication and security environment that stands to nurture transmission of the COVID-19 virus itself.

In the face of these threats, the NCRI herein submits a series of strategic recommendations, as matter of public service, to incisively face down outbreaks of weaponized information in order to help restore public trust:

1. We recommend a public initiative to monitor the weaponized information outbreak using these methods in order to create transparency at scale and provide responsible, sober and timely information in the form of public service announcements, wherever massive disinformation campaigns emerge.
2. We recommend that methods to detect targeted flows of ethnic hate be deployed in collaboration with Web platforms to intersect these outbreaks where and when they occur, and to create a public facing climate station to track, expose, and combat weaponized information.

**Anonymous** ID:lwvzIOAI Mon 23 Mar 2020 23:51:02 No.249867662   
 >>249859732  
 Chink virus.  
 I'm gonna go out and beat an asian now because of this thread.

Screenshot from 4chan /pol/.



Jose L. Gomez, accused of stabbing Asian family over coronavirus, was charged with three counts of attempted capital murder and one count of aggravated assault with a deadly weapon.



Emerging threats of anti-Asian violence appeared on Instagram feeds on April 1, 2020. Calls for mass violence from social media comprise imminent threats that must be tracked and detected in the weaponized information landscape.

3. We recommend consensus building and democratic outreach within and between social media platforms to create better integration across communities and converge on shared norms around weaponized information. Methods such as theme subtraction can provide a much-needed objective avenue to clarify ambiguity in code words in aggregate. As bad actors seize on obfuscated code words with the goal of agitating hate, their true meaning can now be more plainly revealed. Application of this method will afford the well-intended a more objective method by which to distance themselves from bad actors who hijack their messaging in order to misinform or to incite hate. Theme filtering can best be used to share context to avoid stigma and demonization and help heal the bitter social rifts where weaponized information, like a virus, can proliferate.
4. The way platforms currently perform censorship are neither democratic nor transparent. Furor around disinformation can threaten to encourage excess in this practice. The heavy use of censorship by platforms amid the current outbreak of disinformation is understandable, but excesses in censorship carry risks that need to be better understood as well. We recommend utilizing the tools of NCRI to help us move from excesses in the practices of censorship and add tools to help us move towards public conciliation and shared norms in the face of hate.

## ACKNOWLEDGMENTS

Sponsorship by:

The Miller Center for Community Security at Rutgers University



Special thanks to:

Pushshift, for contributing data to this report



**THE NETWORK CONTAGION RESEARCH INSTITUTE (NCRI)** is a neutral and independent third party whose mission it is to track, expose, and combat misinformation, deception, manipulation, and hate across social media channels.

Acting as a public benefit corporation, NCRI is a not-for-profit organization that seeks to explore safe ways to audit, reveal challenges, devise solutions, and create transparency in partnerships with social media platforms, public safety organizations, and government agencies.