(Tarik Haiga/Unsplash)

HUMANS

Researchers Developed AI That Tells Apart True Conspiracies From Conspiracy Theories

TIMOTHY R. TANGHERLINI, THE CONVERSATION 7 JANUARY 2021

The audio on the otherwise shaky body camera footage is unusually clear. As police officers search a handcuffed man who moments before had fired a shot inside a pizza parlor, an officer asks him why he was there.

The man says to investigate a pedophile ring. Incredulous, the officer asks again. Another officer chimes in, "Pizzagate. He's talking about Pizzagate."

In that brief, chilling interaction in 2016, it becomes clear that conspiracy theories, long relegated to the fringes of society, had moved into the real world in a very dangerous way.

Conspiracy theories, which have the potential to cause significant harm, have found a welcome home on social media, where forums free from moderation allow like-minded individuals to converse. There they can develop their theories and propose actions to counteract the threats they "uncover."

But how can you tell if an emerging narrative on social media is an unfounded conspiracy theory? It turns out that it's possible to distinguish between conspiracy theories and true conspiracies by using machine learning tools to graph the elements and connections of a narrative. These tools could form the basis of an early warning system to alert authorities to online narratives that pose a threat in the real world.

The culture analytics group at the University of California, which I and Vwani Roychowdhury lead, has developed an automated

approach to determining when conversations on social media reflect the telltale signs of conspiracy theorizing.

We have applied these methods successfully to the study of Pizzagate, the COVID-19 pandemic and anti-vaccination movements. We're currently using these methods to study QAnon.

Collaboratively constructed, fast to form

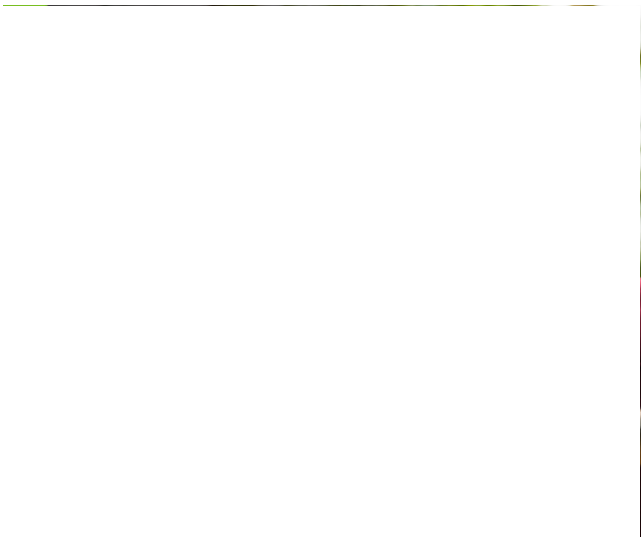
Actual conspiracies are deliberately hidden, real-life actions of people working together for their own malign purposes. In contrast, conspiracy theories are collaboratively constructed and develop in the open.

Conspiracy theories are deliberately complex and reflect an all-encompassing worldview. Instead of trying to explain one thing, a conspiracy theory tries to explain everything, discovering connections across domains of human interaction that are otherwise hidden – mostly because they do not exist.

While the popular image of the conspiracy theorist is of a lone wolf piecing together puzzling connections with photographs and red string, that image no longer applies in the age of social media. Conspiracy theorizing has moved online and is now the end-product

of a collective storytelling. The participants work out the parameters of a narrative framework: the people, places and things of a story and their relationships.

The online nature of conspiracy theorizing provides an opportunity for researchers to trace the development of these theories from their origins as a series of often disjointed rumors and story pieces to a comprehensive narrative. For our work, Pizzagate presented the perfect subject.



Pizzagate began to develop in late October 2016 during the runup to the presidential election. Within a month, it was fully formed, with a complete cast of characters drawn from a series of otherwise unlinked domains: Democratic politics, the private lives of the Podesta brothers, casual family dining and satanic pedophilic trafficking.

The connecting narrative thread among these otherwise disparate domains was the fanciful interpretation of the leaked emails of the Democratic National Committee dumped by WikiLeaks in the final week of October 2016.

AI narrative analysis

We developed a model – a set of machine learning tools – that can identify narratives based on sets of people, places and things and their relationships. Machine learning algorithms process large amounts of data to determine the categories of things in the data and then identify which categories particular things belong to.

We analyzed 17,498 posts from April 2016 through February 2018 on the Reddit and 4chan forums where Pizzagate was discussed. The model treats each post as a fragment of a hidden story and sets about to uncover the narrative. The software identifies the people, places and things in the posts and determines which are major elements, which are minor elements and how they're all connected.

The model determines the main layers of the narrative – in the case of Pizzagate, Democratic politics, the Podesta brothers, casual dining, satanism and WikiLeaks – and how the layers come together to form the narrative as a whole.

To ensure that our methods produced accurate output, we compared the narrative framework graph produced by our model with illustrations published in *The New York Times*. Our graph aligned with those illustrations, and also offered finer levels of detail about the people, places and things and their relationships.

Sturdy truth, fragile fiction

To see if we could distinguish between a conspiracy theory and an actual conspiracy, we examined Bridgegate, a political payback operation launched by staff members of Republican Gov. Chris Christie's administration against the Democratic mayor of Fort Lee, New Jersey.

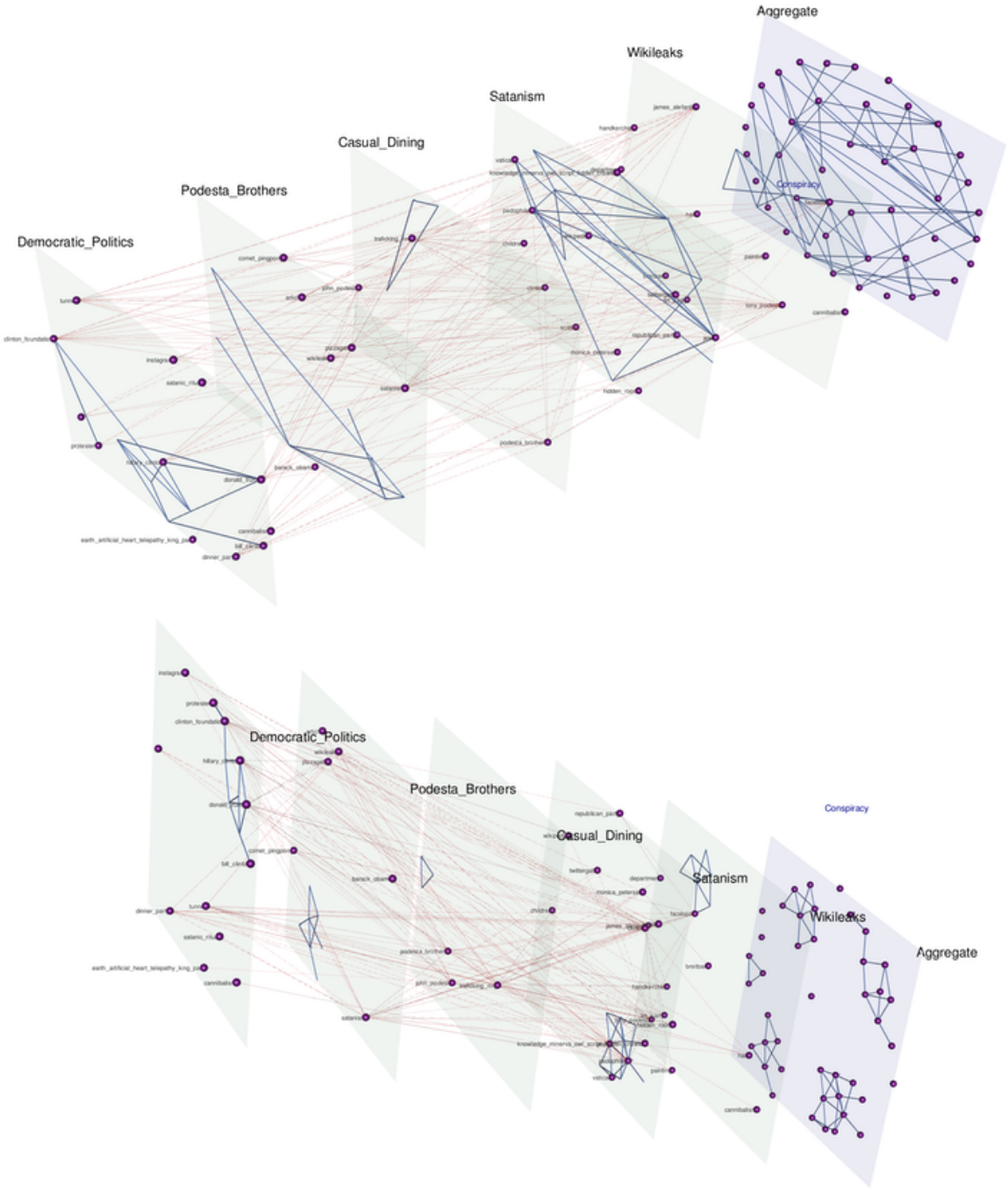
As we compared the results of our machine learning system using the two separate collections, two distinguishing features of a conspiracy theory's narrative framework stood out.

First, while the narrative graph for Bridgegate took from 2013 to 2020 to develop, Pizzagate's graph was fully formed and stable within a month. Second, Bridgegate's graph survived having elements removed, implying that New Jersey politics would continue as a single, connected network even if key figures and relationships from the scandal were deleted.

The Pizzagate graph, in contrast, was easily fractured into smaller subgraphs. When we removed the people, places, things and relationships that came directly from the interpretations of the WikiLeaks emails, the graph fell apart into what in reality were the unconnected domains of politics, casual dining, the private lives of the Podestas and the odd world of satanism.

In the illustration below, the green planes are the major layers of the narrative, the dots are the major elements of the narrative, the blue lines are connections among elements within a layer and the red lines are connections among elements across the layers.

The purple plane shows all the layers combined, showing how the dots are all connected. Removing the WikiLeaks plane yields a purple plane with dots connected only in small groups.



(Tangherlini et al., CC BY 4.0)

Above: The layers of the Pizzagate conspiracy theory combine to form a narrative, top right. Remove one layer, the fanciful interpretations of emails released by WikiLeaks, and the whole story falls apart, bottom right.

Early warning system?

There are clear ethical challenges that our work raises. Our methods, for instance, could be used to generate additional posts to a conspiracy theory discussion that fit the narrative framework at the root of the discussion. Similarly, given any set of domains, someone could use the tool to develop an entirely new conspiracy theory.

However, this weaponization of storytelling is already occurring without automatic methods, as our study of social media forums makes clear. There is a role for the research community to help others understand how that weaponization occurs and to develop tools for people and organizations who protect public safety and democratic institutions.

Developing an early warning system that tracks the emergence and alignment of conspiracy theory narratives could alert researchers – and authorities – to real-world actions people might take based on these narratives.

Perhaps with such a system in place, the arresting officer in the Pizzagate case would not have been baffled by the gunman's response when asked why he'd shown up at a pizza parlor armed with an AR-15 rifle.

Timothy R. Tangherlini, Professor of Danish Literature and Culture, University of California, Berkeley.

This article is republished from The Conversation under a Creative Commons license. Read the original article.

Learn More

What Insulin Option is Trending in Diabetes Care?
ReachMD

Learn why memory care is person-centered care.
Brought to you by Brookdale Senior Living

Insulin is Expensive, Explore These Savings & Discounts
ReachMD

What to look for in a memory care community.
Brought to you by Brookdale Senior Living

Powered by TREND MD

Think Testosterone is the Answer? Mike Tyson Says "Think again"
Mike Tyson

How Victoria Principal Looks At 71 Is Heartbreaking
Wordsa

This "1" Knee Pain Reliever Has Knee Surgeons Speechless
AmRelieve
Learn More

Sponsored Links

How Thousands of People Keeping Toenail Fungus Off (Try It At Home)
Fungus Eliminator
Learn More

If You Have Foot Neuropathy Pa
AmRelieve

How Victoria Principal Looks At 71 Is Heartbreaking
Wordsa | Sponsored

Read Next Story ›