

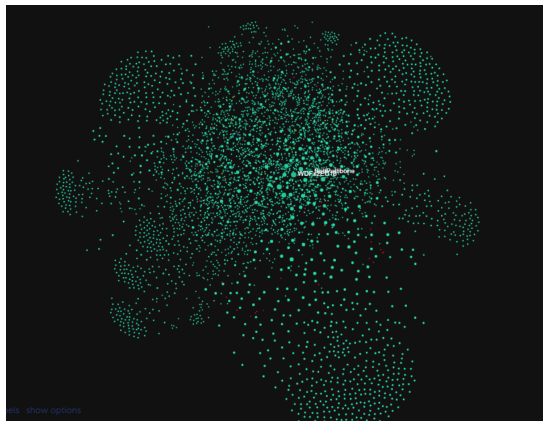


Investigating the infamous #Pizzagate conspiracy theory

Panagiotis Metaxas and Samantha Finn

Highlights

- Describes the TwitterTrails tool that enables the investigation of rumors on Twitter
- Using TwitterTrails, analyzes Pizzagate, an infamous conspiracy theory that gained prominence during the 2016 US Presidential election
- Reveals several previously unknown facts about the story, including the role that a Turkish journalist played in internationalizing the rumor, visualizes the echo chamber of those who promoted it, and describes their identities and connections



The echo-chamber formed by the highly visible actors promoting the #pizzagate conspiracy theory on Twitter reveals a dense group of accounts who do not question the obvious lies they propagate.

Abstract

In the 2016 US Presidential elections, we saw the rise of the so-called “fake news” phenomenon –a term used to describe lies, formatted as news articles, aiming to confuse and/or anger the public. One of the better-known instances of “fake news” is the infamous “Pizzagate” conspiracy theory that received significant attention in the media. In this paper

we use the TwitterTrails tool to investigate it. TwitterTrails (twittertrails.com) is an interactive, Web-based investigative tool that allows its users to examine on Twitter the origin and propagation characteristics of a rumor and its refutation. We designed TwitterTrails as a tool for journalists (amateur and professional) investigating recent and breaking stories. However, with little training, it can be used by anyone who wants to find answers to some basic facts about the spreading of rumors online.

Results Summary: Searching the Web, one finds dozens of news articles and analyses related to the debunked Pizzagate conspiracy theory. There are several facts that previous articles missed, however, and our TwitterTrails investigation reveals some of them. In particular, it points to an earlier starting date on Twitter than the one previously reported. It visualizes the identities and the close relationship between the main actors of the online echo chamber promoting the Pizzagate conspiracy theory, and it reveals how a bilingual troll misled a Turkish journalist who played a significant role in internationalizing the conspiracy theory.

Introduction

Online social media, such as Twitter and Facebook, have become part of modern news reporting, and it is used daily to inform its users. It is also used to inform journalists who, in turn, can have a greater influence on the public, including people who may not even be on social media. It's not really that "the internet is full of lies" [1], as some people casually claim; if it were, the internet would be eventually ignored. Instead, the vast majority of people use the internet dozens of times every day. The negative reaction that some people have towards the internet probably has more to do with characteristics of human behavior [2]. People respond emotionally and selectively to those recognizable lies that bother them.

The so-called "24-hour news cycle" has led to increased sensationalism in news stories [3]. At a time of increased competition among cable news channels and online news media, the need for journalists to catch the attention of the public has led to faster and more hyped-up reporting [4]. Many compete to be the first to report a breaking story and present new and exclusive angles. This trend feeds off social media and empowers citizen journalists [5] who publish and transmit news through websites like Twitter and Facebook. However, a journalist's desire for fame and recognition on social media can sometimes suppress critical thinking in favor of reporting speed. At the same time, it is not clear anymore who qualifies as a journalist.

It is not just citizen journalists and so-called traditional media sources that compete for our attention. Masquerading as news organizations, propagandists seize the opportunity to misinform. They produce "fake news," that is, falsehoods formatted and circulated online in such a way that a reader might mistake them for legitimate news articles from reputable sources. (We should mention that there is some confusion around the term "fake news," as it means different things to different people. Some people use it to denote opinions with which

they disagree as well as reporting errors and incorrect predictions. None of these are “fake news.”)

“Fake news” has been around for a long time, but social media technology makes it possible to produce and consume it today on a massive scale [6]. Such articles appear on a variety of unknown websites that adopt familiar-sounding names, such as abcNews.com.co, [7] designed to be confused with the known abcNews.com. Many of these sites profit by gaining clicks that lead to advertisements on social media sites. In order to be successful in attracting user attention, these websites present a sensational made-up story of a political nature, religious nature, or anything that would have strong emotional appeal to a subset of the public. Typically, “fake news” stories are planted on social media sites using provocative titles and images. “Clickbait” titles, as they are called, attract the attention of unsuspecting and confused social media users who click on links thinking they are visiting a legitimate news site. These users are drawn in by the emotional appeal, and “fake news” providers get a share of advertising money from each click.

Of course, not every “fake news” creator’s purpose is simply financial gain. Many have political motivations, and they would be better characterized as propagandists of national or international origins. Their made-up stories are a form of propaganda, aiming to trick readers into behaving in ways beneficial to the “fake news” provider, the propagandist [8]. The benefit may be political (e.g., persuading readers to vote as the propagandist wants), financial (e.g., persuading readers to click on the advertisements and bring revenue to the propagandist), religious (e.g., persuading readers that a particular religion is good or bad), entertaining (e.g., persuading readers to spread a joke and show how gullible people are), etc. While millions of such stories exist, the vast majority of them do not succeed in getting widespread attention. Those that are successful, however, spread for a variety of reasons [9]. One of the main reasons behind successful spreading is that most readers are familiar with the historically valid model of trusting professional news sources, and cannot tell the difference [10] between professional reporting and “fake news.” The former is typically edited and printed by some authoritative source to the Web, while the latter looks authoritative but could be anyone’s opinion.

While we can often identify and avoid being tricked by “fake news,” the unfortunate fact is that any of us could fall for one of these lies, especially if it is presented in a way that matches our biases and prior beliefs [11]. In order to recognize “fake news,” diversity of information sources and informants is key. It can be easier to recognize “fake news” if a group of diverse people, with a broad variety of different biases (and therefore tendencies to believe or be skeptical about different stories), engages with the information. On the other hand, members of a homogeneous group (an “echo chamber”) can be easily fooled when presented with lies that conform to their common biases.

Unfortunately, people tend to form echo chambers in social media and in society. We find comfort and safety with others who are similar to ourselves. We do not like to believe lies.

However, when we are presented with evidence that our beliefs are incorrect, we try hard to avoid challenging our belief system [2]. This is when we are most susceptible to lies: when they are presented in a way that conforms to our prior beliefs.

Background

As a way of studying and reacting to this trend, we are using TwitterTrails [12] (twittertrails.com), an interactive, web-based investigative tool that uses visualization and allows users to investigate the origin and propagation characteristics of a rumor and its refutation, on Twitter. In this section, we review related work focusing on tools that use similar techniques and data sources as TwitterTrails, and on tools visualizing data for similar purposes.

On any particular day, there may be hundreds of thousands of tweets commenting on a topic, and it is impossible for someone to make sense of the comments by reading them. Big data necessitates sense-making tools, especially in the form of interactive visualizations, to allow humans to process and interpret the data. Pirolli and Card [26] study how information should be organized for intelligence analysis, and to this end introduce a "sense-making loop" describing the process in which a tool like TwitterTrails can help its user analyze information. They postulate that visualizations should be used as a sort of external memory for a user, to improve a user's memory and processing capabilities. TwitterTrails follows a bottom-up data process similar to the one described in [26], in which the tool gathers and refines data with human input, and then creates visualizations to help the user filter and consume the data in a meaningful way in order to formulate theories about the data.

Similar to TwitterTrails are tools that focus on timelines to visualize the spread and propagation of a story or real-time event, often focusing on summarization of the data. Fisher et. al. track the frequency of terms in blog data to track the evolution of news stories [27]. A meme-tracking tool developed by Leskovec et al. maps the rise and fall of memes in the blogosphere and the news media [28]. Swan and Jensen [29] create "overview timelines" by extracting nouns and named entities and charting the frequency of these features over time.

The tools described above take data from news media, blogs, and internet searches. TwitterTrails focuses on the spread of data on social media, generated by both official news media and individuals reporting and spreading stories. Social media websites generate massive amounts of data every day. Twitter's rise to prominence in both daily and professional life has led to the creation of many tools to make sense of its data. TwitInfo is a visual interface to assist users in summarizing event specific data, such as a sports match [30]. Eddi mines a single user's stream, which can be overwhelmed by hundreds of tweets daily, and utilizes topical analysis to enable users to browse their stream [31]. It aggregates topical data from Twitter's Streaming API and automatically identifies and labels peaks in the data. It uses a timeline interface similar to TwitterTrails to allow users to browse through data but adds sentiment and geolocation to give more information about the data. Vox Civitas, a

graphical tool created by Diakopoulos et al., has a motivation similar to TwitterTrails: to assist users, specifically journalists, in extracting interesting and meaningful information from social media streams [32]. Their tool mines query specific data from media events on Twitter and visualizes both sentiment and topics over time. When journalists evaluated Vox Civitas, one reaction was that they would use it to track sources of data. TwitterTrails focuses on this goal, using a similar time series graph to allow users to navigate through Twitter data.

Monitoring and evaluating the propagation of a rumor has recently gotten a lot of attention. For example, RumorLens [33] analyzes the spread of rumors on Twitter and prompts user feedback to classify results as propagating, debunking, or unrelated to the original rumor. It will then use a text classifier to garner more widespread results. Rumor Cascades on Facebook have also been studied by a Facebook team, [34] focusing on tracking the way that rumors propagate on Facebook, mainly those that have been verified in snopes.com.

One of the earliest systems that focused on studying patterns of information propagation in online social networks like Twitter is Truthy [35]. Now part of a broader project called Osome, Truthy is based on the concept of memes that spread in the network. Such memes are detected and followed over time to capture their diffusion patterns. Through visualizations of propagation patterns and other metrics (e.g., sentiment analysis), Truthy can enable a user to come to a certain conclusion on her own. Finally, we note that Shane Greenup maintains an extensive list of projects working to address online misinformation [25].

Methods

In this paper, we use TwitterTrails [12] (twittertrails.com) to investigate the origin and propagation characteristics of a rumor, and its refutation, on Twitter. Within minutes after launching, TwitterTrails collects relevant tweets and automatically answers several important questions regarding a rumor such as its originator, spreading pattern, groups of propagators and main actors. In addition, TwitterTrails computes and reports the rumor's level of *visibility* and the audience's *skepticism* towards the rumor, which correlates with the rumor's *credibility*.

We designed TwitterTrails to be a valuable tool for journalists (amateur and professional) investigating recent and breaking stories. Further, its growing collection of investigated rumors can answer questions regarding the amount and success of misinformation on Twitter [13]. Not every piece of “fake news” has the same importance. Believing the “miraculous” characteristics of some beauty product may cost us a bit of money, but it is much less important than believing the “miraculous” medicinal effects of a substance that, supposedly, can cure a lethal disease. The latter can lead to the death of someone who might otherwise be treated. Believing as true a “gaffe” that some member of the British royal family supposedly committed can give us a laugh, but believing that a major politician is involved in a despicable child-abusing scandal can hurt the way that democracy should work. In any case, it would be useful if we could examine how people on social media react to a rumor

they heard. Doing so could give us a broader perspective and help us identify opinions of people who may not share the same biases with us.

TwitterTrails has been used to investigate over a thousand stories [14] since 2014. In the remainder of this section, we describe how TwitterTrails works through a case study. In the next section we describe our findings from investigating the infamous Pizzagate conspiracy theory.

Reading a TwitterTrails “story” page through an investigation example

Consider the following scenario, which will serve as a running example in this section: On September 18, 2015, while the refugee crisis appears in the news on a daily basis, a TwitterTrails user sees a tweet: “Desperate Refugee gently waves ISIS flag at German police as a friendly help & thanks for making me welcome???” The tweet is accompanied by a photo of someone apparently using an ISIS flag, ready to hit a man in uniform with POLIZEI written on its back. The whole scene is one of confrontation and anger.

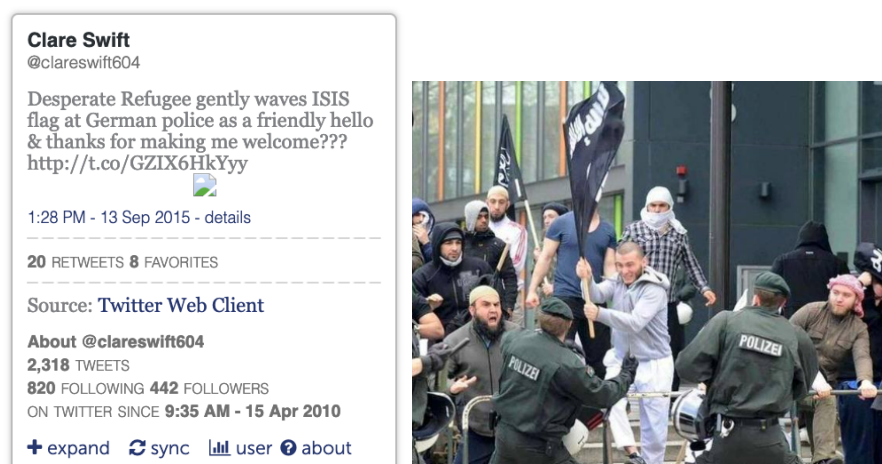


Figure 1. Tweet that started the TwitterTrails “story” used in the investigation example.

Our user notices that the tweet was sent on Sept. 13, 2015, so it is already 5 days old. “Is it true?,” our user would like to know. The sender of that tweet is unknown to our user. It could be a witness, a so-called “citizen reporter.” Reliable information can be provided by witnesses and spread through social media networks, which could aid journalists when writing a story about this confrontation. But how can journalists or other individuals verify the claims of information they discover on Twitter? There is no formal quality control in the realm of the “citizen reporter” —anyone could claim this title. Searching the Internet and social media can be tedious and time-consuming and might require technical skills that an individual doesn’t have readily available. It is also not clear what keywords our user should use in this case to verify the validity of this event.

Our user launches a TwitterTrails investigation that will produce a “story,” an automatically created Web page that our user can examine to evaluate the claim of the tweet. To launch the TwitterTrails investigation our user gives as a starting point the tweet at hand, along with relevant keywords. In this case, a relevant keyword is the string “REFUGEE ISIS FLAG,” which in connection to the starting tweet should collect *all relevant tweets* that have been sent 7 days prior to the time of the investigation using the Search API [15]. If some of the tweets collected reference older tweets, these older tweets are also retrieved (as is the case with the story we examine below).

The initial keyword selection is the only manual portion of launching a story using TwitterTrails. After that, the Web page with the story will be created automatically. Finding the right keywords that produce the most relevant tweets is not a trivial task, and this is the reason we opened this step only to trained journalists and not to the general public. However, anyone can request an investigation by completing the request form at <http://bit.ly/TTrequest> or emailing the administrators. In our case, the selection of “refugee isis flag” was determined to be the right string. Using fewer than these three words in a search string (e.g., “refugee isis”) would have retrieved less relevant tweets that could dilute the accuracy of the search. Using more words in the search string, such as “German” or “police,” would have missed relevant tweets. In any case, the trained user who launches the story has a set of tools to measure the number of relevant tweets, precision, accuracy, and current frequency, as well as to experiment with various search strings.

Within minutes, the Web page with the story collects the relevant 10,000 tweets and analyzes them. (The reader is welcome to follow the description by examining the story online at <http://bit.ly/2Hywb81>.) The story reads as follows:

Welcome to TwitterTrails, a system to investigate the spread and validity of stories on Twitter. TwitterTrails gathers data about news stories, rumors, events, and memes on Twitter, to present in useful and meaningful visualizations that can help users answer questions about how the story spread. Scroll down for the visualizations, or click on “overview” on the top left of the page to view data about this story. For more information about the specific visualizations or the TwitterTrails system, please read our blog or follow us on Twitter @tweet_trails.

This page, created automatically by TwitterTrails at 11:47 AM on 18 Sep 2015, investigates a story based on the following tweet:

Desperate Refugee gently waves ISIS flag at German police as a friendly hello & thanks for making me welcome??? <http://t.co/GZIX6HkYyy> - @clareswift604

Data collected were tweets posted for about the week prior to the start of the investigation. During that time, propagation of this story was insignificant, and in general people were extremely doubtful of the information presented.

The last reference to insignificant propagation and doubt is part of the analysis that TwitterTrails performs [16]. In our case, TwitterTrails indicates that the tweet has not widely spread and may not be a true claim. But what evidence is there about the claim? TwitterTrails was designed so that one could quickly answer the following questions:

- Who originated the claim (i.e., wrote the first relevant tweet) and when?
- Is the story still spreading at the time of the investigation?
- Who “broke the story,” that is, when did the claim become more widely known? This step identifies an early tweet that attracted a relative spike of attention in terms of retweets. (It does not refer to the tweet with the overall highest number of retweets.)
- Who are the “main actors,” that is, who are the accounts that are more visible to the audience? Main actors are measured in terms of retweets in the current investigation, not the overall number of followers an account may have.
- The main actors are automatically grouped [37] into “communities” based on the way the audience sees them. How do the community members describe themselves?
- What pictures were used in the tweets?
- How much has the story been spreading? If it is related to a claim, is there any doubt in the audience [38] about the claim?

The idea behind TwitterTrails is that by reviewing the answers to these questions, our user will have a reasonable idea as to whether this claim is to be trusted or not. Let’s follow the description on the TwitterTrails story page to see how these questions are answered. Each question is treated in a **section** in the story page, and the section has some visualization element, preceded by an explanation of the visualization. We present each section below. The first section is about the Propagation graph.

WHO BROKE THE STORY AND WHEN?

The Propagation Graph highlights the tweets which were influential in “breaking” the story on Twitter, and highlights independent content creators.

Each circle on the Propagation graph represents a tweet, and hovering over or clicking on the circle will display the tweet to the right of the graph.

Tweets are plotted on x-axis of the graph based on the time they were posted, and on the y-axis by the number of retweets they have received (at the time of data collection).

Metaxas P, Finn Investigating the infamous #Pizzagate conspiracy theory. *Technology Science*. 2019121802. December 18, 2019. <https://techscience.org/a/2019121802>

Circles are sized based on the number of followers the user who posted the tweet has. Circles are drawn by default as gray. Circles with other colors represent tweets with nearly identical texts.

Additionally, circles with a bright blue border indicate tweets written by verified accounts.

Propagation Graph



Figure 2. The Propagation Graph of the investigation example.

According to the explanation of this TwitterTrails section, in order to see which tweet is credited with “breaking the story,” our user should hover over the circles to reveal which are the actual tweets they represent. The most prominent tweet is in the upper right corner of the propagation graph and reads: “You’ve probably seen a picture of a refugee holding an #ISIS flag. It’s a complete lie.” The sender of this tweet with Twitter handle @7pillars apparently claims that the picture is a lie. The tweet has only received 50 retweets, although the account has over 60,000 followers. The color of that tweet’s circle (purple) appears in many instances of the Propagation graph, indicating that many of the tweets at that time repeat roughly the same text as the breaking account. Note that the Propagation graph’s x-axis starts at 4:30 AM ET and ends at 7:00 AM ET on Sept. 15, 2015, a relatively long period of time compared to many of the 600 stories we have investigated in TwitterTrails, indicating a slow propagation. We can click on the name of tweet senders to learn more about them. The time that the breaking tweet was sent (6:39 AM) is not unreasonable for someone in New York, but the early

tweets in the propagation graph indicate the topic was already discussed in Europe in mid-morning. In fact, almost all of the tweets in the graph are by senders in the UK and EU.

So, our user realizes that the claim under investigation did not receive much attention, even 2 days after it started spreading, and no verified account had been tweeting just before the story was breaking. (Verified accounts often belong to media organizations and to well-known individuals. They should not be considered as belonging to reliable entities by default. However, people often are not aware of this fact.)

But how was the claim started and by whom? The second section with the Time Series of Relevant Tweets histogram has a few answers for our user's questions.

WHO AND WHEN THE STORY ORIGINATED? IS IT STILL SPREADING?

The Time Series shows the activity over time of relevant data collected.

Time is on the x-axis and the number of tweets generated is on the y-axis. Each point represents a ten minute time span.

Selecting a point on the time series will display on the right a list of the tweets at that time span. These tweets are sorted by the number of retweets they have received (highest on top), and can be re-sorted using the drop down menus. If there are more than 50 tweets in the time span, links to navigate the tweets 50 at a time are provided.

You can zoom in on the graph by clicking and dragging your mouse over a period of time.

Clicking on Manage Series on the bottom right of the display will open a panel which you can use to add new time series to the graph by checking the box on the left.

The Search field takes a search term and will display all tweets contain (the exact) search term when you check the box on the left of Search.

The shape of the Time Series indicates how the story was spreading and whether it was still spreading at the time of the investigation.

Time Series of Relevant Tweets %

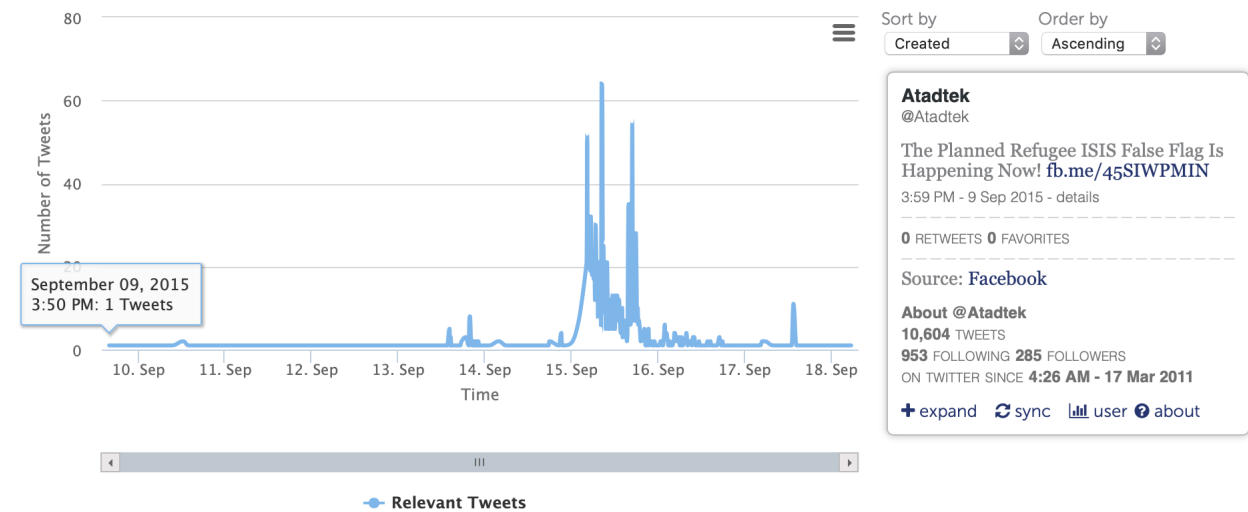


Figure 3. Time Series of Relevant Tweets of the investigation example.

The first relevant tweet the TwitterTrails investigation found was sent on Sept. 9, 2015, almost 10 days *before* our user's investigation. It appears that this tweet may not be directly relevant to the investigation and was only collected because it accidentally contains the target keywords "refugee," "ISIS" and "flag." It focuses on the term "false flag." Readers familiar with the term "false flag" [18] will recognize a codeword used by right-wing extremist groups. Did any such groups have a role in the development of the story we investigate? Our user would like to know how many tweets contain the term "false flag" and investigates by entering the keywords "false flag" in the search button under "Manage Series." It turns out that only 16 tweets have this codeword. Were they relevant? Let's continue our investigation to find out.

Time Series of Relevant Tweets %

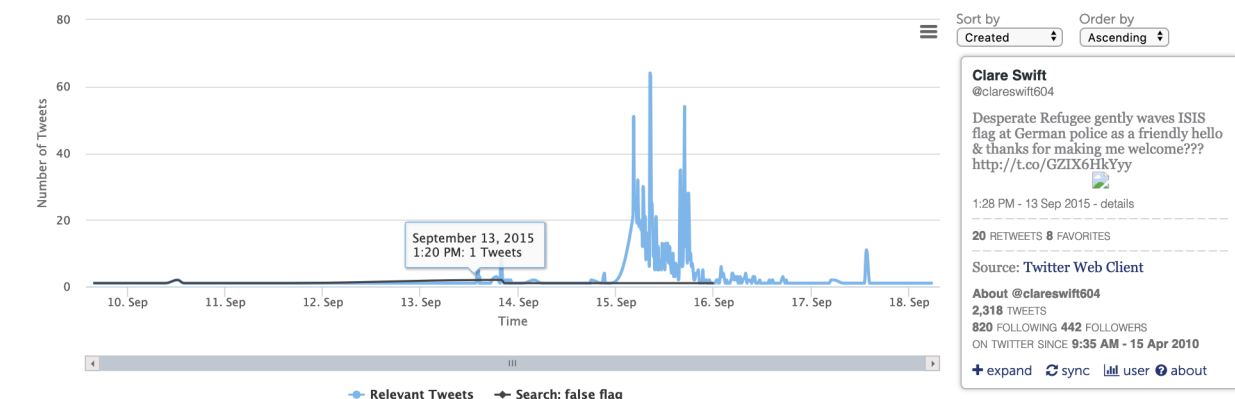


Figure 4. Displaying the trend of specific keywords chosen by the user in the example.

Ignoring the “false flag” tweets, our user can check which is the first tweet in the investigation that *does not* have the codeword and identify the originator of the rumor. It turns out that the first tweet was the one sent by the account with handle @clareswift604 that actually prompted this investigation on Sept. 13, 2015, and it got only 20 retweets. Was there any relationship between the “false flag” initiator and the originator of the rumor?

Realizing that a rumor has not spread much can be very useful to journalists. Often, some journalists see an outrageous claim and will go to great lengths to debunk it. Unfortunately, this may work against their intentions and make the claim more widely known, especially if they try to use sarcasm that can be misunderstood.

Recall that the tweet that broke the story wrote that the claim was a lie. “Are there other tweets that also use the word “lie” in their tweets?” our user might wonder. TwitterTrails allows us to search for any word in the collection. Clicking on “Manage Series” reveals a menu of histograms to include in the Time Series. Our user chooses to search for “lie”:

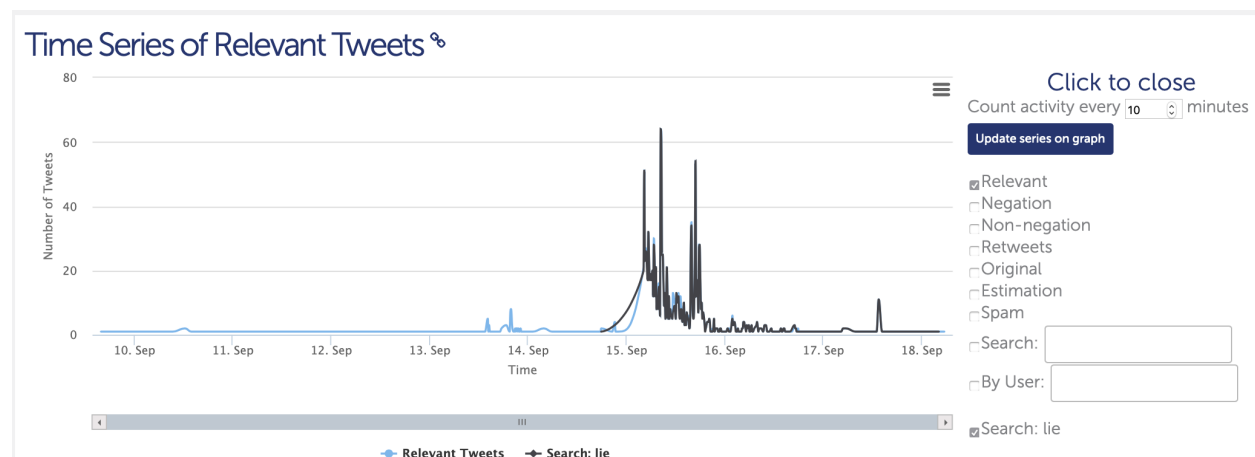


Figure 5. The Manage Series menu gives the user options to further an investigation.

Indeed, it turns out that most tweets in the collection did contain the word “lie.” It seems that the audience did not “buy” the claim. In fact, it is because so many tweets in the collection included the negative-sentiment word “lie” that TwitterTrails declared that “people were extremely doubtful of the information presented” in the first section of the story.

Next, TwitterTrails provides an answer as to who are the most visible members of the Twitter community that discuss this topic. Typical social network analyses look at the **retweet** network, which shows the retweeting activity in the tweet collection, and we briefly describe it here. The nodes of the retweet network represent accounts, and the arcs represent retweets. A node X represents an account that retweeted the account of another node Y. An arc from X to Y represents such a retweet. The more times X retweets Y, the thicker the arc, and the closer X and Y are drawn. Drawing of the retweet graph is done using a force-directed algorithm [19].

Retweet networks are typically very large networks. Even in stories without much spreading, like the one we are describing here, the retweet network can be large and provides information about node popularity, which is not very useful in our investigation. That's why by default, TwitterTrails does not show the retweet network. However, we show it here for the curious reader and for comparison with the co-retweeted network presented in the next section.

For our story, the retweet network contains 1253 nodes. There were altogether 1203 retweets in the retweeted network, and the only popular node, the account for *The Independent* newspaper, received 866 retweets by 853 users. In fact, it sent 6 almost identical tweets trying to promote its message. This story, therefore, is an example of a rumor that became known more because of the eagerness of journalists to debunk it than because of the success of the rumor spreaders!

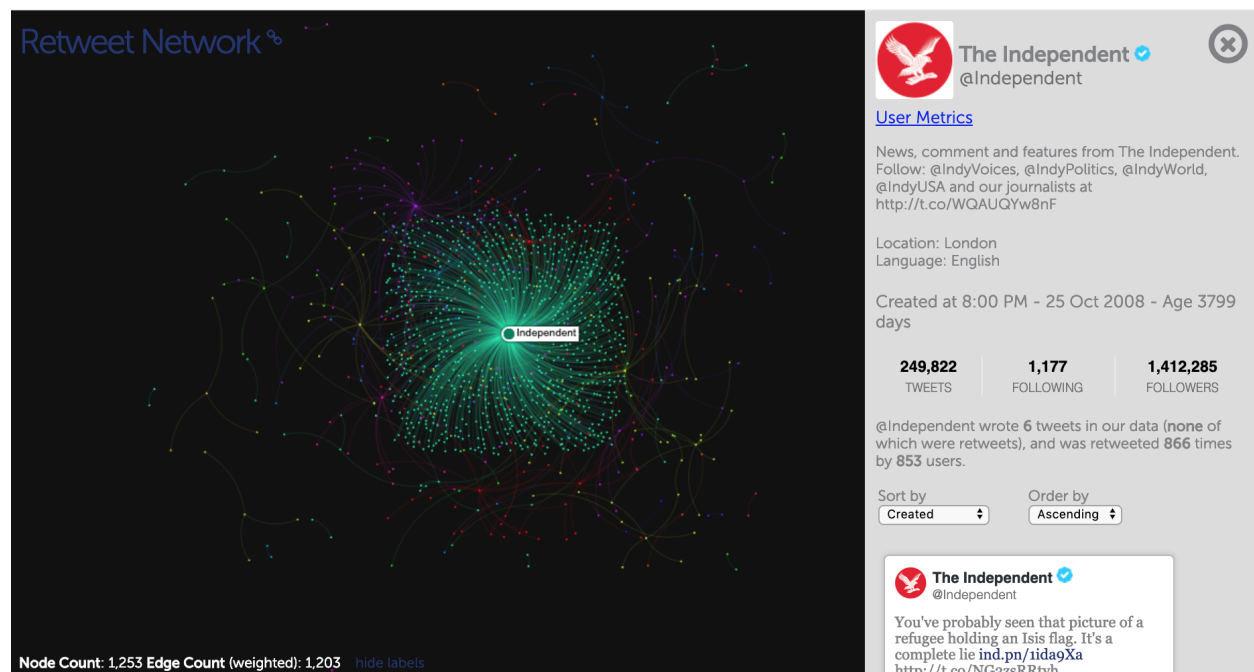


Figure 6. The Retweet Graph of the investigation example shows the most popular tweets but has limited application on discovering the groups participating in the online exchange.

Many systems that analyze Twitter activity stop at the retweet network. While the retweet network is useful in pointing out accounts with popular tweets, it cannot reveal other important information. In particular, it cannot reveal whether there are any groups of accounts that may be presenting different versions of the story. This is accomplished by the next network TwitterTrails uses, the **co-retweeted** network.

WHO ARE THE MAIN ACTORS OF THE INVESTIGATION?

The co-retweeted network shows the clusters, communities that participate in this investigation, and highlights influential accounts in the retweet network. It is generated by connecting and clustering accounts based on mutual retweeting by other users. (That is, if User A and User B in the co-retweeted network are connected by an edge, it means at least one other user (part of the “audience”) has retweeted both User A and User B. The more members of the audience are retweeting both User A and User B, the stronger the edge among them, and the closer they appear in the cluster.)

Clusters look like clouds and are forming automatically (based on the force-directed algorithm) and indicate the strongest agreement regarding the topic being investigated.

Communities are often parts of the cluster clouds and are colored automatically (based on the Louvain algorithm) and indicate similarity between users in the community: they have stronger connections within their community than outside of it. There could be several communities within a cluster.

Hovering over a point will display the name of the influential account it represents, and clicking on it will bring up information about that account on the right of the graph. The user information also contains the tweets written/retweeted by that user in the dataset.

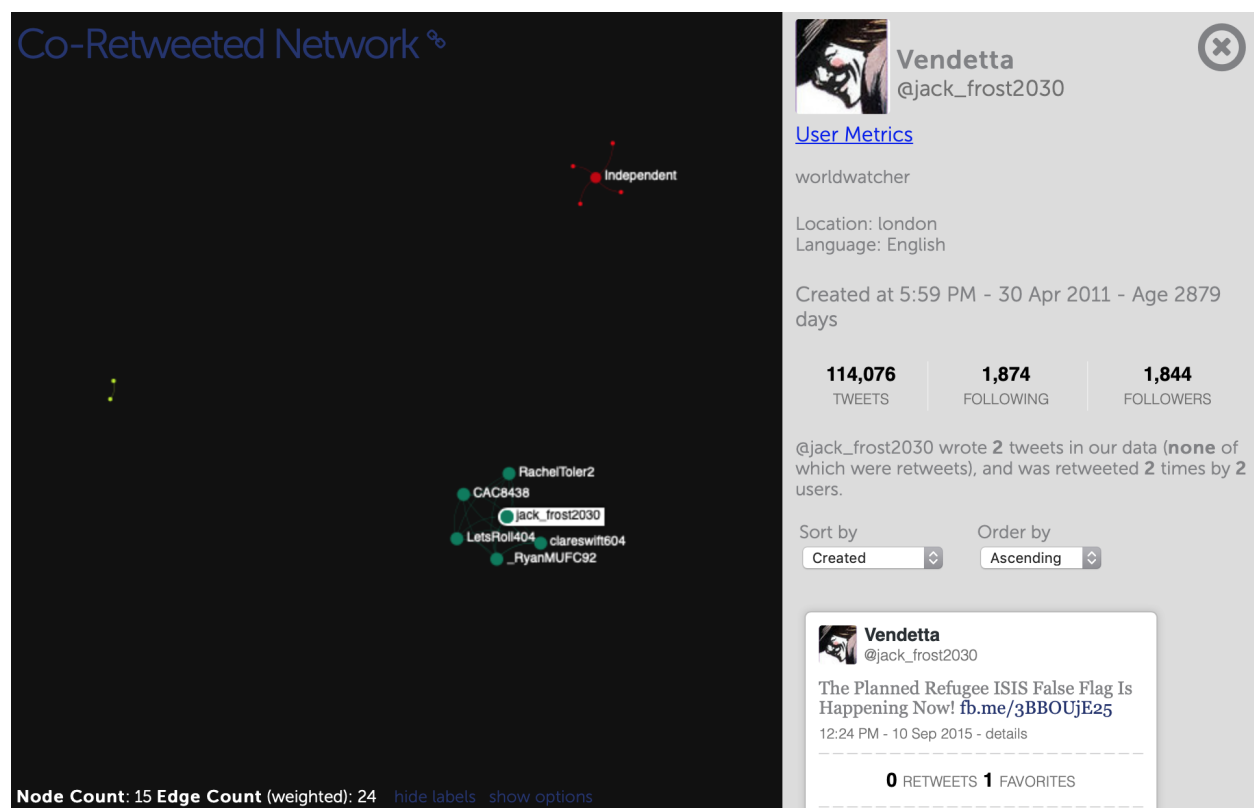


Figure 7. The Co-Retweeted Network of the investigation example shows two separate groups discussing the rumor.

The co-retweeted network is an important visualization because it can (a) reveal the major actors in the story and (b) show whether there is any polarization between them. Nodes represent major actors that have been retweeted by others. The accounts doing the retweeting of major actors are not represented in the network, but their co-retweeting actions are represented as undirected edges. An edge between nodes A and B means that a third account C has retweeted both A and B, that is, A and B are both visible to C. It also means that, according to C, A's and B's retweeted messages are in agreement because, by and large, retweeting means endorsement [20]. The co-retweeted network is drawn using Gephi's Force Atlas 2 [21] algorithm, a force-directed layout, which means that, the more accounts retweet nodes A and B, the closer A and B are drawn.

In our story the co-retweeted network reveals that there are not many visible actors and that the actors are polarized. They appear in two larger groups. One group has 5 nodes, and its central dominating node is *The Independent* newspaper's account. The remaining 4 are repeating the Independent's tweet, one of them corresponding to the account that broke the story, his tweet propagating faster than the newspaper's first tweet!

The other group contains 6 accounts that were spreading a "false flag" conspiracy theory and promoting the false claim. Both the "false flag" initiator and the rumor originator are in the same groups. The central node is a conspiracy theorist who spread many of the "fake news" stories that the now-banned Russian IRA trolls [22] also spread. Including the 16 accounts that promoted the "false flag" conspiracy, one finds 5 now-deleted accounts, one left-wing, three environmentally concerned, one that serves soft porn, and 8 right-wing. We would guess that almost all are fake accounts.

The co-retweeted network reflects the polarization easily. But what can we tell about these communities? The next TwitterTrails section, "Network Statistics," gives us word clouds of the terms the community members use in order to describe themselves. The 6 members of the "Blue group" (looks green in the image) we saw above, often use words such as "christian," "people" and "god" in their profiles. The group that *The Independent* participates in ("Red group") and the two small ones have no common words.

HOW ARE COMMUNITY MEMBERS DESCRIBING THEMSELVES?

There are a total of 4 communities of similar users in the Co-Retweeted Network. The largest community has 6 users in it, and the smallest has 2. Nodes in the co-retweeted graph are colored based on their community.

Each community is also represented by a word cloud in the colored rectangle below. The more often community members use a word in their profile, the larger that word appears in the word cloud.

To view aggregation statistics about any of the communities, you can either click on a node in the graph, or select a community from the panel below.

Co-Retweeted Network Statistics

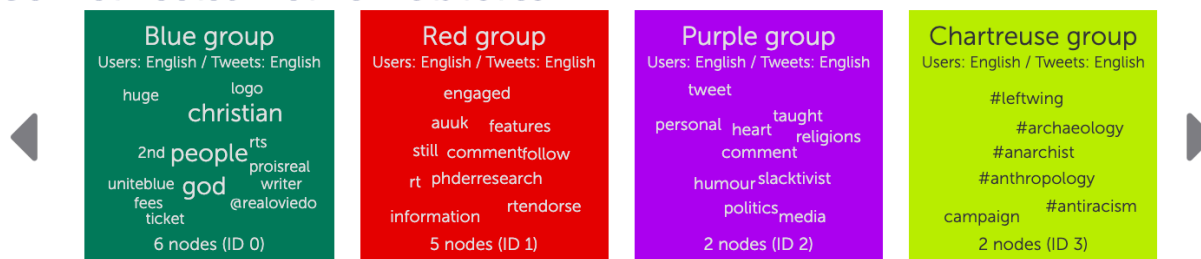


Figure 8. The reader of the investigation example can get a sense of the characteristics of the groups involved in the online exchange by looking at some of the statistics of the Co-Retweeted network groups. This figure shows the more common keywords used in the profiles of the group members and the group sizes.

While the co-retweeted network gives us some insight in this example, it is in large networks that this visualization is particularly useful, because large networks may display a higher degree of polarization. The visualization can also reveal an echo chamber, as we will see in the Pizzagate story in the next section. In the current example of the “ISIS flag” fake picture, there is a small set of communities, and one can examine the prominent members easily. In fact, in doing so we see that, as of the time of this writing, two of the 6 accounts in the Blue group were deleted by Twitter.

Finally, the story presents a grid of photos used in the collection. Almost all of the pictures are copies of the same fake photo. There is also a missing picture. When TwitterTrails can no longer find a photo because it was deleted along with the owning account, there is an empty spot in the photo grid.

WHAT PICTURES WERE USED IN THE TWEETS?

Research has shown that a picture is a powerful way to promote a message because it has strong emotional impact on people. This section shows the most retweeted images are displayed in this investigation. Hover your cursor over on an image to see the tweet in which that image was posted.

In Pictures

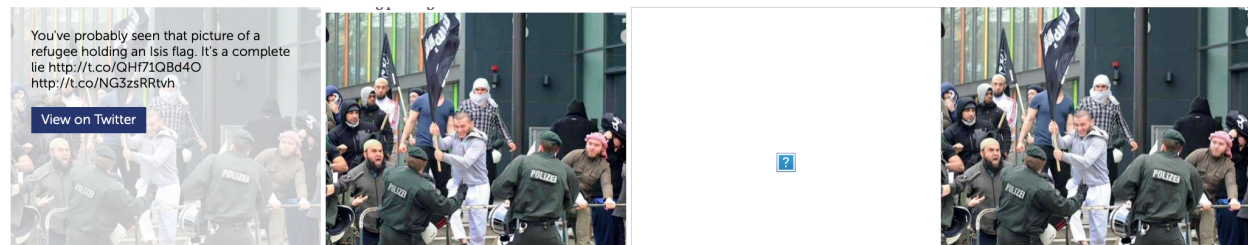


Figure 9. Some of the pictures that were included by the participants of the investigation example. A missing picture indicates that the account that posted it was deleted by Twitter.

The story Web page starts with the title of the story and two semi-circles visually indicating the spread and skepticism of the claim. In our case, the story had insignificant spread, and extremely high skepticism. In fact, most of the tweets in the story were from those who were fact-checking the photo and describing it at a lie.

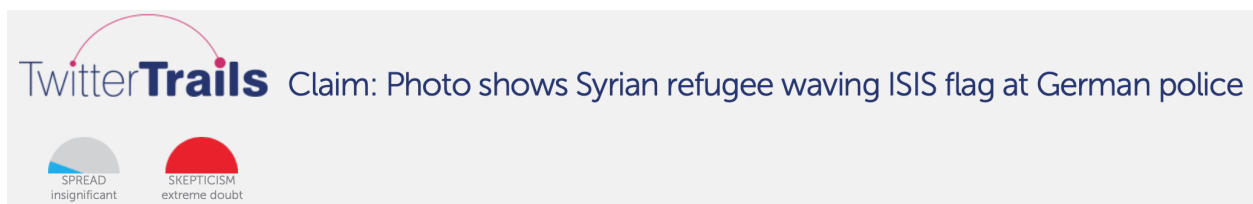


Figure 10. The title of the TwitterTrails “story” of the investigation example contains a measurement of the spread that the story had along with a measurement of the skepticism on its validity based on the words that the collected tweets contained. In our example it appears that the rumor had very low spreading and that most of the online participants were expressing doubt on its validity.

We have now seen all the components of a TwitterTrails story using a small example. We are ready to explore the Pizzagate conspiracy theory that is the main topic of this paper.

Results

TwitterTrails investigation of #Pizzagate

The previous section described in some detail TwitterTrails, a system that investigates a rumor by retrieving all recent tweets related to keywords of the rumor. The example in the previous section was a story that did not spread widely. The Pizzagate story we describe in this section, however, had a much greater spread, causing investigation by many news media [23] and even the US Congress. A few days before the 2016 Presidential elections, a rumor spread on Reddit and then major social media that claimed that Hilary Clinton, John Podesta

Metaxas P, Finn Investigating the infamous #Pizzagate conspiracy theory. *Technology Science*. 2019;12:1802. December 18, 2019. <https://techscience.org/a/2019121802>

and other well-known Democrats were involved in a pedophile ring operating out of the basement of a pizza store, the Comet Ping Pong restaurant in Washington, DC. The evidence supposedly came from hacked emails that WikiLeaks had revealed earlier that year, and it spread despite the fact that some people visited the business, even recording live, and found no children and no basement in the store. The harassment of the store owner and other nearby businesses continued until a self-appointed savior took upon himself to walk into the store on December 4, 2016, discharging his gun and threatening the employees.

Searching the Web one finds hundreds of thousands of news articles and analyses related to this discredited conspiracy theory [39], and it is a reasonable question to ask whether TwitterTrails can add anything to what we already know. In fact, as we demonstrate in this section, TwitterTrails can offer facts and insights that were not previously known.

TwitterTrails' investigation of #PizzaGate can be found in <http://bit.ly/TTpizzagate>, an online interactive page. Interested readers are encouraged to visit the page and interact with it on their own. Be aware that the data set is large and may take several minutes to load on a computer.

We started the investigation on Dec. 2, 2016, prompted by a tweet from an account that tries to spread conspiracy theories that the stock market is about to crash. Their tweet was a bit different: "Keep telling yourself it's 'fake news.' Nobody wants to believe people are killing and torturing children. Not in America! #PIZZAGATE - @ZVixGuy."

Examining the TwitterTrails story reveals several points worth noting. Here are a few of them:

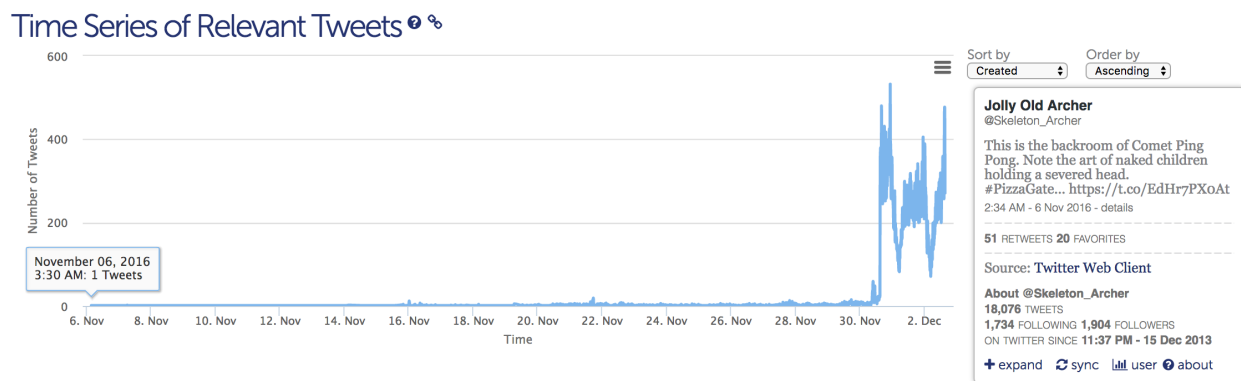


Figure 11. The Time Series of the #Pizzagate investigation shows that the hashtag started on Twitter two days before the 2016 elections.

The timeline reveals that #Pizzagate was a rumor that had not gotten attention on Twitter for much of November 2016, until groups formed an echo chamber to discuss it.

The rumor on Twitter started two days before the 2016 elections

It has been widely reported [40] that the hashtag #PizzaGate appeared on Twitter on Nov. 7, 2016, a day before the US Presidential elections. In news reports, there is no mention of who first used the hashtag. TwitterTrails data reveals that the hashtag appeared earlier (Nov. 6, 2016 at 3:30 AM ET) and was created by a trolling account, which is currently suspended on Twitter, that had not tweeted since Feb. 13, 2017. Before it was suspended, the account often tweeted pro-Nazi content, violating Twitter rules. This may be the reason for the suspension.

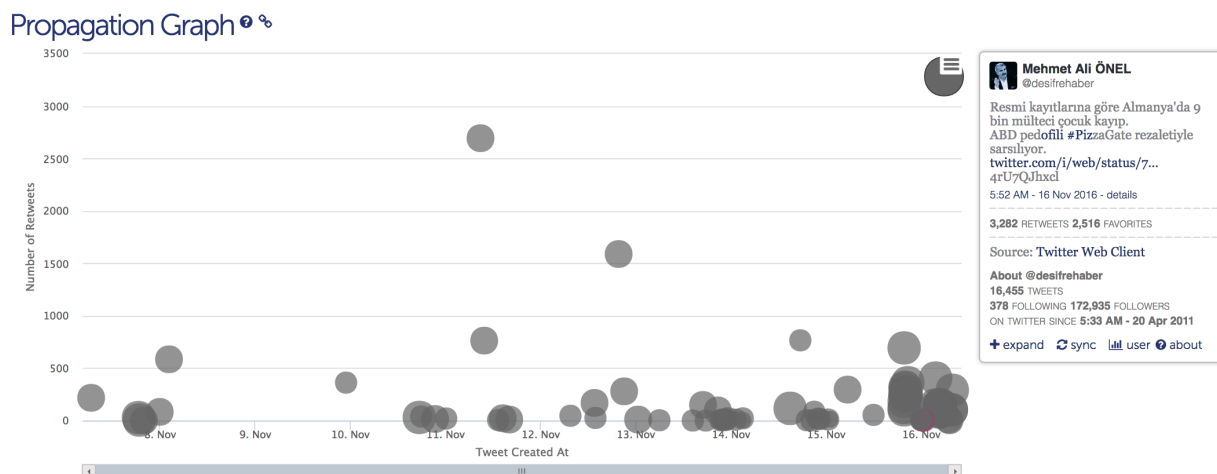


Figure 12. The Propagation Graph of the #Pizzagate investigation shows that the tweet that made the rumor more known belonged to a journalist in Turkey.

The propagation graph shows who “broke” the story on Twitter and when. In the upper right corner the tweet that received the most retweets early on appears as a (partially covered) gray circle. Moving the cursor over the circle shows the actual tweet on the right. Under this tweet is a barrage of tweets sent a few hours earlier by a troll, “informing” and provoking the Turkish journalist.

A Turkish journalist “broke” the story outside the US led by a troll

It has been reported that Turkish journalists promoted #PizzaGate, [41] but no direct evidence has been given on how large was that role. Our data identify a pro-Erdogan Turkish journalist, Mehmet Ali Önel, who played a major role in the internationalization of the rumor earlier than the news articles reported. At that time, the Erdogan government was criticized for a proposed law that would enable rapists to marry their victims in order to avoid prosecution [42]. The journalist tries to counter by claiming that such atrocities are happening in Germany and the US: “According to official records, 9,000 refugee children are missing in Germany. US pedophilia #PizzaGate shaken. So where are these kids?”

TwitterTrails points out who prompted the Turkish journalist about the rumor. A bilingual troll bombarded the Turkish Twittersphere with at least 118 tweets about conspiracy theories

Metaxas P, Finn Investigating the infamous #Pizzagate conspiracy theory. *Technology Science*. 2019121802. December 18, 2019. <https://techscience.org/a/2019121802>

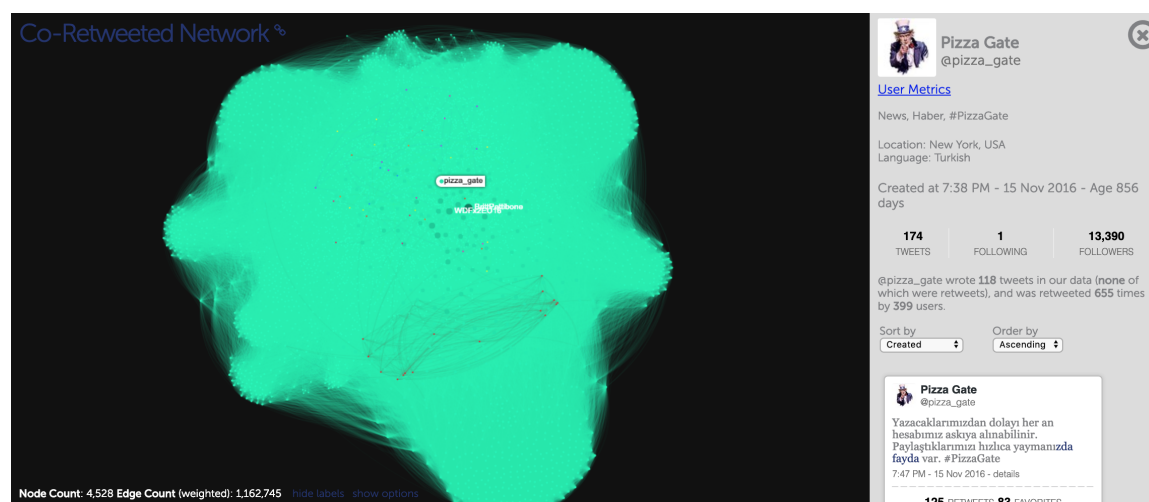
a few hours before the journalist picked it up. This trolling account was created just minutes before tweeting for the first time and within a two-week period acquired more than 13,000 followers. This troll was clearly part of a network of propagandists and bots and tweeted exclusively on Pizzagate.



Figure 13. Our investigation revealed that a bilingual troll account was very active and very influential in promoting the false rumor in Turkish shortly before it was tweeted by journalists.

An echo chamber of conspiracy theorists

One might wonder whether there was any skepticism during the spreading of the rumor. The answer is no, because the rumor spread in a dense echo chamber, creating a perfect environment for growing the conspiracy theory. The Turkish troll is positioned close to the center of the echo chamber, along with many of the Russian accounts shown to spread propaganda during the 2016 US elections, such as AmelieBaldwin, TheFoundingSon, DorotheBell, PatriotBlake, DonnaBRivera, CooknCook, RealRobert1987, OneMightyFish, and March_for_Trump.



Metaxas P, Finn Investigating the infamous #Pizzagate conspiracy theory. *Technology Science*. 2019;12:1802. December 18, 2019. <https://techscience.org/a/2019121802>

Figure 14. The Co-Retweeted Network shows a dense echo-chamber of accounts repeating the false rumor to each other At the center of this echo-chamber are the accounts that were instrumental in making the false rumor widely known.

The co-retweeted network of the echo chamber formed by the group discussing #Pizzagate. Typically, co-retweeted networks of stories that are of a political nature show at least two polarized groups. For example, in US politics one typically sees two groups, one liberal and the other conservative. In this story, however, there is a single network representing an echo chamber wherein claims are accepted without strong doubt.

Co-Retweeted Network Statistics

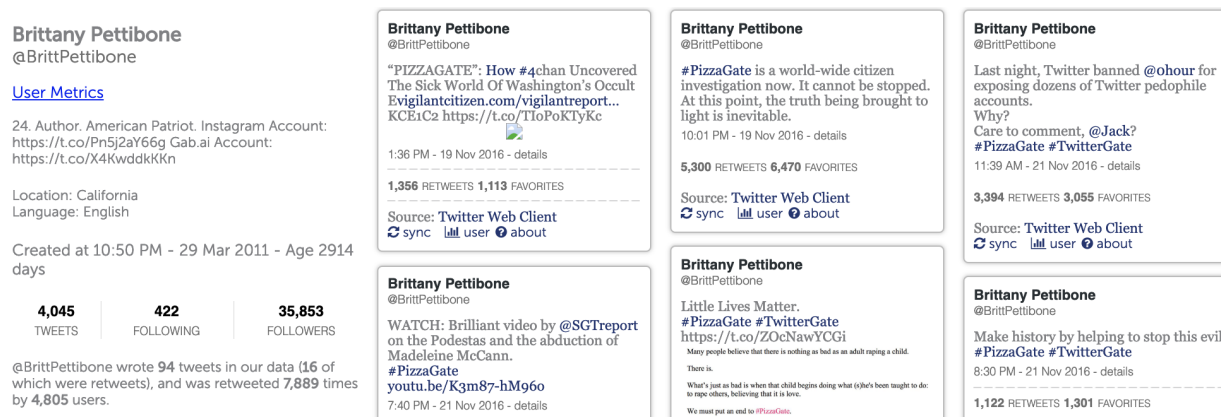
There are a total of 17 communities of similar users in the Co-Retweeted Network. The largest community has 4474 users in it, and the smallest has 2, with an average of 266 users, and median of 2 users. Nodes are colored based on their community. To view aggregation statistics about any of the communities, you can either click on a node in the graph, or select a community from the panel below.



Figure 15. The Co-Retweeted network statistics shows the most common words that the accounts in the echo-chamber were using in describing themselves. The more common words are “#maga”, “trump”, “truth”, and “love”.

There was one major group discussing #Pizzagate, effectively forming a dense echo chamber. The keywords in the word cloud are those that appear more often in the profiles of the participants in the group, effectively describing the group members. In the large group the keywords are trump, #maga, truth, love, and god.

TwitterTrails’ co-retweeted graph can easily show what the echo chamber looked like, what were the main keywords in the profiles of the group that populated the echo chamber, and who were the main actors in its spreading. It turns out that Brittany Pettibone, an aspiring teenage writer, played a major role in spreading the rumor in the US [36].



Metaxas P, Finn Investigating the infamous #Pizzagate conspiracy theory. *Technology Science*. 2019121802. December 18, 2019. <https://techscience.org/a/2019121802>

Figure 16. One of the more influential accounts in promoting the false rumor was describing herself as “24. Author. American Patriot” along with her other social media accounts.

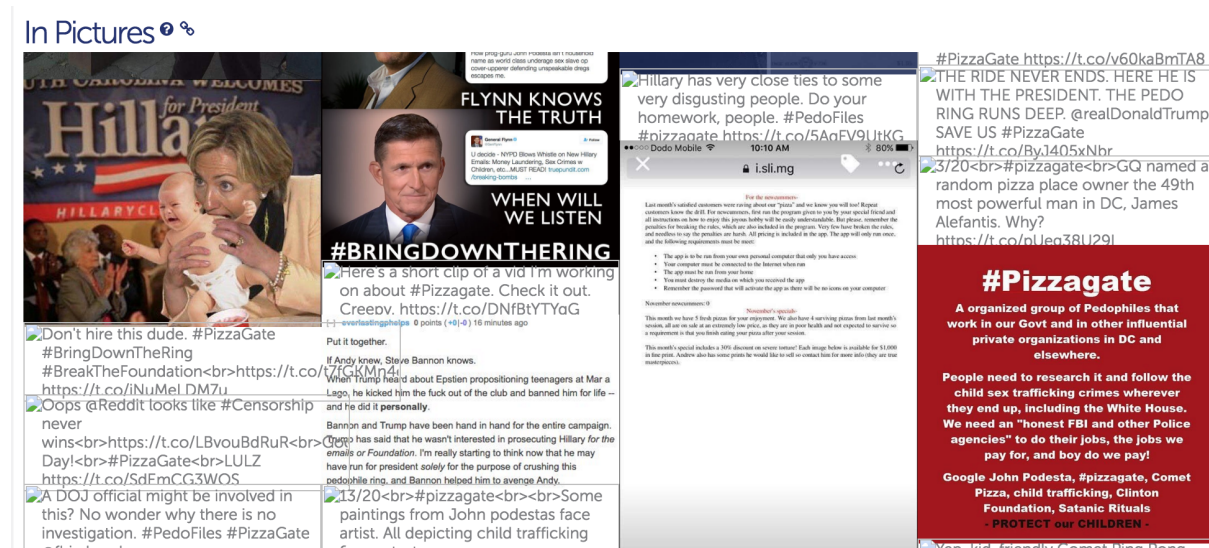


Figure 17. The pictures that the members of the echo-chamber promoting the false #Pizzagate rumor were very graphic, mainly photoshopped images aiming to anger, disgust and to spread conspiracy theories. Most of them belonged to accounts that are deleted at the time of this writing.

There are many more discoveries that one can make by interacting with the TwitterTrails page <http://bit.ly/TTpizzagate>, such as the pictures posted by the major actors. Given the limited space for this publication, we will leave further investigation to the interested reader.

An investigation starts with data collection. Using appropriate keywords that should guarantee as great precision and recall as possible, we collect all relevant tweets. For our case study, we performed a keyword search on Twitter for a single keyword, “#Pizzagate,” on Dec. 2, 2016. After that step, TwitterTrails is fully automatic. Throughout the process, we also collect relevant data, such as the pictures shared with these tweets and the URLs of any websites mentioned.

Discussion

Even though #Pizzagate has been extensively examined, TwitterTrails easily points out facts and insights previously missed. This is but one of the over 500 publicly visible investigations we conducted using our system. Many others conducted by journalists are visible only to them. We welcome and work with journalists or researchers who want to make use of our

system. If interested, please contact us and follow us on Twitter (Trails Research @tweet_trails).

In addition to the findings we describe above, this investigation reveals the technique that propagandists use to create a conspiracy theory on Twitter: Find a community that is emotionally charged on some issue; create fake accounts that become members of the community; then launch the conspiracy theory and sit back and watch while the rest of the community promotes the conspiracy without skepticism [24].

TwitterTrails was designed and implemented with the goal of providing a vital service to users who want to engage with Twitter as a source of reliable information, either for their own consumption, or as a source for journalism, both professional and amateur.

TwitterTrails makes it easy to investigate a suspicious story. By inputting a single tweet to the system, and selecting keywords relevant to the story being investigated, the system will gather a dataset of tweets through which the user can trace the story's origin. The Tweet Propagation visualization focuses on the moment the story first broke on Twitter, while the Timeline visualization shows how it spread. Both allow users to meaningfully and easily sort through hundreds to thousands of tweets. The visualizations highlight both tweets and time periods most interesting to the story. The two network visualizations, a Retweet network and a Co-Retweeted network, allow users to study accounts on Twitter that were both influential propagators of information and sources that other users put trust in.

Our system leads us to conjecture that true and false rumors have different footprints in terms of how they propagate and invoke skepticism by their audience. False rumors are more likely to be negated if exposed to a larger audience.

The most pressing areas of future study for TwitterTrails are to design and implement a user evaluation of the tool, and to further improve its functionality and usefulness.

We also plan on evaluating and improving our algorithms to detect when a story breaks on Twitter, and filtering for relevant tweets. We also hope to pursue more methods of customizing TwitterTrails for users in ways specific to the story they are investigating. This may include creating more visualizations, so users can select those more appropriate to their investigation, and creating more meaningful ways in which these visualizations can interact with each other.

We close this section with a note and invitation that our system is open, our collection is not based on proprietary data, our methods are simple and easily implementable, and interested researchers can replicate and verify our work.

References

1. New Scientist, "Web of lies: Is the internet making a world without truth?" Published: 17 February 2016. Available at: <https://www.newscientist.com/article/2077405-web-of-lies-is-the-internet-making-a-world-without-truth/>
2. P.T. Metaxas, "Propaganda and the limits of human intellect," arXiv:1806.09541, 2018 - arxiv.org. Available at: <http://bit.ly/2OzSB8f>
3. Neil Postman, "Amusing Ourselves to Death: Public Discourse in the Age of Show Business," Penguin Books 1985.
4. Jeff Sorensen, "24 Hour News Killed Journalism." Huffington Post Blog, Aug.20, 2012. Available at: <http://huff.to/1tRlttE>
5. P.T. Metaxas, Eni Mustafaraj, Samantha Finn and Andres Monroy-Hernandez, "Hiding in Plain Sight: A Tale of Trust and Mistrust Inside a Community of Citizen Reporters". Proceedings of the ICWSM Conference, Dublin, Ireland, June 2012. Available at <http://bit.ly/JbaAeW>
6. Public Policy Practicum, Stanford Law School, "Fake news and misinformation" 2017. Available at <https://law.stanford.edu/wp-content/uploads/2017/10/Fake-News-Misinformation-FINAL-PDF.pdf>
7. Wikipedia entry: <https://en.wikipedia.org/wiki/ABCnews.com.co>
8. P.T. Metaxas and J. Destefano, "Webspam, Propaganda and Trust, Proceedings of the WWW/AIRWeb2005, May 10, 2005, Chiba, Japan. Available at <http://bit.ly/WjNKEM>
9. Pennycook, Gordon and Rand, David G., "Who Falls for Fake News? The Roles of Bullshit Receptivity, Overclaiming, Familiarity, and Analytic Thinking" (March 22, 2019). Available at: <https://ssrn.com/abstract=3023545>
10. C. Pollalis et al. "Classroom Activity for Critical Analysis of News Propagation Online". Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. Available at: <https://dl.acm.org/citation.cfm?id=3170427.3174355>
11. B. Southwell and E. Thorson, "The Prevalence, Consequence, and Remedy of Misinformation in Mass Media Systems," Journal of Communication, 2015. Available at: <https://onlinelibrary.wiley.com/doi/10.1111/jcom.12168>
12. P. T. Metaxas, S. Finn, and E. Mustafaraj, "Using twittertrails.com to investigate rumor propagation," in Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work; Social Computing, ser. CSCW'15 Companion. New York, NY, USA: ACM, 2015, pp. 69–72. [Online]. Available: <http://bit.ly/2uGeoEc>

13. S. Finn, P. T. Metaxas, and E. Mustafaraj, "Spread and skepticism: Metrics of propagation on twitter," in Proceedings of the ACM Web Science Conference, ser. WebSci '15. New York, NY, USA: ACM, 2015, pp. 39:1–39:2. [Online]. Available at: <http://bit.ly/2n7o2Ml>
14. TwitterTrails Blog. Available at: <https://blogs.wellesley.edu/twittertrails/2014/09/17/hello-world/>
15. Twitter Standard Search API. Available at: <https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.html>
16. P. T. Metaxas, S. Finn, and E. Mustafaraj, "Using twittertrails.com to investigate rumor propagation," in Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing, ser. CSCW'15 Companion. New York, NY, USA: ACM, 2015, pp. 69–72. [Online]. Available: <http://bit.ly/2uGeoEc>
17. Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment*, 10.
18. Merrill Perlman, "How the term 'false flag' migrated to the right." *Columbia Journalism Review*, October 29, 2018. Available at: https://www.cjr.org/language_corner/false-flags.php
19. Fruchterman, T. M. J. and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21:1129– 1164.
20. P.T. Metaxas et. al., "What do Retweets indicate? Results from User Survey and Meta-Review of Research," Proceedings of the 9th AAAI ICWSM, 2015. Available at: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10555/10467>
21. Gephi (2010). ForceAtlas2, the new version of our home-brew layout. Available at: <http://bit.ly/1deeWht>
22. Oliver Roeder, "Why We're Sharing 3 Million Russian Troll Tweets" *FiveThirtyEight* Blog, Jul. 31, 2018. Available at: <https://fivethirtyeight.com/features/why-were-sharing-3-million-russian-troll-tweets/>
23. Cecilia Kang. "Fake News Onslaught Targets Pizzeria as Nest of Child-Trafficking," *NY Times*, Nov. 21, 2016. Available at: <https://www.nytimes.com/2016/11/21/technology/fact-check-this-pizzeria-is-not-a-child-trafficking-site.html>

24. E. Mustafaraj and P. T. Metaxas, "The "fake news" spreading plague: Was it preventable?" in Proceedings of the 2017 ACM on Web Science Conference, ser. WebSci '17. New York, NY, USA: ACM, 2017, pp. 235–239. [Online]. Available at: <http://bit.ly/2sehUCv>
25. Shane Greenup. Catalogue of all projects working to solve Misinformation and Disinformation. On the MisInfoCon Web site, accessed 26 Aug. 2019. Available at: <http://bit.ly/2zoQD5a>
26. Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In Proc. of the International Conference on Intelligence Analysis, 2005
27. Fisher, Danyel et al. "Narratives: A visualization to track narrative events as they develop." 2008 IEEE Symposium on Visual Analytics Science and Technology (2008): 115-122.
28. Jure Leskovec, Lars Backstrom, and Jon Kleinberg. 2009. Meme-tracking and the dynamics of the news cycle. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '09). ACM, New York, NY, USA, 497-506.
29. R. Swan, D. Jensen. Timemines: Constructing timelines with statistical models of word usage. KDD-2000 Workshop on Text Mining, 73-80
30. Marcus, Adam et al. "Twitinfo: aggregating and visualizing microblogs for event exploration." CHI (2011).
31. Bernstein, Michael S. et al. "Eddi: interactive topic-based browsing of social status streams." UIST (2010).
32. Diakopoulos et. al. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In IEEE Symposium on Visual Analytics Science and Technology, 2010.
33. Paul Resnick, et.al. Rumorlens: A system for analyzing the impact of rumors and corrections in social media. In Amer. Journalism Review, 2014.
34. A. Friggeri, et.al. Rumor Cascades. In Intl. Conf. on Weblogs and Social Media, 2014.
35. J. Ratkiewicz, et. al. Truthy: mapping the spread of astroturf in microblog streams. In WWW, pages 249--252. ACM, 2011.

36. “Why 3 anti-Islam activists were refused entry to the UK,” BBC Trending, 14 March 2018, on the BBC site, accessed on 29 Sept. 2019. Available at: <https://www.bbc.com/news/blogs-trending-43393035>
37. We are using Gephi’s Louvain modularity algorithm [17] implementation for this step.
38. We are using a skepticism [13] score that is inspired from by the h-index in academic citations for this purpose.
39. See, e.g., related articles at The New York Times or Snopes.
40. See, e.g., “Pizzagate: From rumor, to hashtag, to gunfire in D.C.” by Marc Fisher, John Woodrow Cox and Peter Hermann, Dec. 6, 2016, *The Washington Post*, <http://wapo.st/2h9EXgi>
41. See, e.g., “The saga of ‘Pizzagate’: The fake story that shows how conspiracy theories spread,” BBC Trending, Dec. 2, 2016: <http://bbc.in/2v5wLmX>
42. The law was abandoned a week later: See "[Turkey scraps plans for law that would have pardoned rapists who marry underage victims.](#)" The Independent (22 November 2016).

Authors

Panagiotis (Takis) Metaxas is a Professor and Chair of Computer Science at Wellesley College. He is the founding Director of the Media Arts and Sciences program at Wellesley and affiliated faculty at the Center for Research on Computation and Society at Harvard University and at the Centre of Technology and Global Affairs at the University of Oxford. He has been studying online misinformation since 2002 and has received four “best paper” awards for his work. With the help of several NSF awards he has been designing tools to counter computational propaganda. For more information, go to <http://bit.ly/pmetaxas>

Samantha Finn is in Library and Technology Services at Wellesley College. A graduate of Wellesley, Samantha has been instrumental in developing TwitterTrails and updating and maintaining it over the years. Her work has been recognized with a “best paper” award.

Acknowledgments

This research was partially supported by NSF grant CNS-1117693 and by the Wellesley Science Trustees Fund. The authors would like to thank the students who have worked in the implementation of the system, especially Prof. Eni Mustafaraj, Laura Zeng, Lindsey Tang, Susan Tang, Megan O’Keefe, and Christina Pollalis.

Metaxas P, Finn Investigating the infamous #Pizzagate conspiracy theory. *Technology Science*. 2019121802. December 18, 2019. <https://techscience.org/a/2019121802>

Referring Editor: Latanya Sweeney

Citation

Metaxas P, Finn S. Investigating the infamous #Pizzagate conspiracy theory. *Technology Science*. 2019121802. December 18, 2019. <https://techscience.org/a/2019121802>

Data

Available at <http://twittertrails.com/>