

---

# DEMO: Using TwitterTrails.com to Investigate Rumor Propagation

**Panagiotis Takis Metaxas**

Computer Science Dept.  
Wellesley College  
Wellesley, MA 02481 USA  
pmetaxas@wellesley.edu

**Samantha Finn**

Computer Science  
Wellesley College  
Wellesley, MA 02481 USA  
sfinn@wellesley.edu

**Eni Mustafaraj**

Computer Science  
Wellesley College  
Wellesley, MA 02481 USA  
emustafa@wellesley.edu

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s). TWITTERTRAILS is supported by NSF CNS-1117693 grant and by the Wellesley Science Trustees Fund. *CSCW'15 Companion*, March 14–18, 2015, Vancouver, BC, Canada. ACM 978-1-4503-2946-0/15/03. <http://dx.doi.org/10.1145/2685553.2702691>

**Abstract**

Social media have become part of modern news reporting, used by journalists to spread information and find sources, or as a news source by individuals. The quest for prominence and recognition on social media sites like Twitter can sometimes eclipse accuracy and lead to the spread of false information. As a way to study and react to this trend, we demo TWITTERTRAILS, an interactive, web-based tool ([twittertrails.com](http://twittertrails.com)) that allows users to investigate the origin and propagation characteristics of a rumor and its refutation, if any, on Twitter. Visualizations of burst activity, propagation timeline, retweet and co-retweeted networks help its users trace the spread of a story. Within minutes TWITTERTRAILS will collect relevant tweets and automatically answer several important questions regarding a rumor: its originator, burst characteristics, propagators and main actors according to the audience. In addition, it will compute and report the rumor's level of visibility and, as an example of the power of crowdsourcing, the audience's skepticism towards it which correlates with the rumor's credibility. We envision TWITTERTRAILS as valuable tool for individual use, and especially for amateur and professional journalists investigating recent and breaking stories.



**Figure 1:** A tweet spreading around 12 noon EST on March 27, 2014, reads (in Spanish) “Picture of the airplane in the sea these moments in Telde, Grand Canary Island”.

## Introduction

The so-called “24 hour news cycle” has led to an increased sensationalism of news stories. Especially with the increase in cable news channels and online news media, the need to catch the attention of the public has led to faster and more hyped up reporting. Many compete to be the first to report a breaking story and present new and exclusive angles. This trend has fed off social media and in turn empowered citizen journalists publishing and transmitting news through websites like Twitter and Facebook. Most of the time the information is true, but the desire to be first and receive more likes and retweets sometimes trumps accuracy and fact checking. In many cases, it may not matter much whether a rumor is true or false, but there are some cases that it matters greatly.

Consider the following scenario, that will serve as a running example in our description: Around noon on March 27, a reporter sees a tweet indicating that an airplane was spotted in the sea near the Canary Islands. For context, this happens just a few weeks after the disappearance of the Malaysian Airlines 370 flight on March 8, which captured the attention of people world wide. Pressing the retweet button is very tempting in this situation, but spreading this information further should not be done automatically. It would be very helpful if the reporter can quickly determine a few facts about this story<sup>1</sup>, including:

- **Originator:** Who “broke” the story first (made it widely known)?

<sup>1</sup>We use the term story to indicate a rumor, true or false, spreading through Twitter.

- **Burst:** When and how did the story break (that is, have the first burst in its propagation)?
- **Timeline:** How is the story propagating over time? Is it still spreading at the time of the investigation?
- **Propagators:** Who has been spreading the story by retweeting, given that retweets often indicate agreement with the message?
- **Negation:** Were there any refutations of the story competing for attention? How widespread were they, compared to the original claim?
- **Main actors:** Who were the main actors in the propagation, according to the Twitter audience?

There is no formal quality control in the realm of citizen reporting. Reliable information can be created by witnesses and spread through social media networks, which could aid journalists when writing a story. But how can journalists or other individuals verify the claims of information they discovered on Twitter? Searching the Internet and social media can be tedious and time consuming, and might require technical information that an individual doesn’t have readily available. In the case of trending stories, massive amounts of data are being created and circulated, and often there will be individuals or bots trying to manipulate this data to promote their agenda.

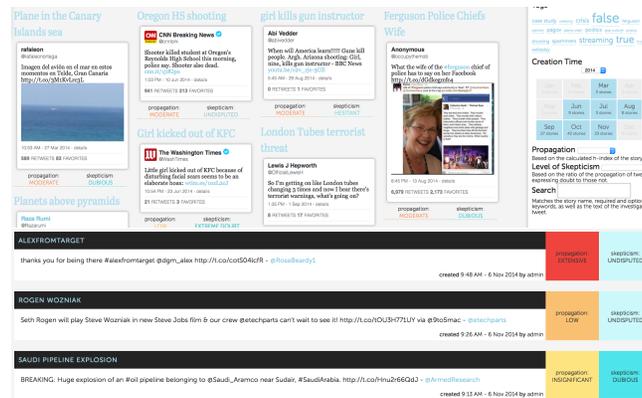
We present and demo [TwitterTrails.com](http://TwitterTrails.com), a new web-based tool for interactive exploration of Twitter information, which helps answer the above-listed questions. Relevant prior work is [1, 2].

- The story was investigated at **12:04 PM on 27 Mar 2014**.
- The first relevant tweet was posted at **1 hour** earlier, at **10:49 AM on 27 Mar 2014**, with a tweet written by **SilviaLuzHernnd**:  
Se estrella un avión en Telde #GRANCANARIA #sucesos @infoemergencias @112canarias - @SilviaLuzHernnd
- It broke on Twitter **4 minutes** later, at **10:53 AM on 27 Mar 2014**, with a tweet written by **rafaleonortega**:  
Imagen del avión en el mar en estos momentos en Telde. Gran Canaria http://t.co/3M1KvLrc3L - @rafaleonortega
- The first 100 breaking tweets were written in **23.6 minutes**
- At the time of inquiry, they had received **5,165 retweets**
- When the story was being investigated, activity was **below average** (1 tpm compared to an average of 52 tpm)
- The incoming tweet rate was **staying fairly constant** (0,000).
- Propagation of this story was **moderate** (42), and in general people were **dubious** (0.742) about the information in the story.
- The main actors who were the most influential spreading and posting relevant information about the story were **laprovincia\_es**, **rafaleonortega**, **aenaerpuertos**, **elmundoes**, and **ActualidadRT**, who wrote the following tweets:

**Figure 2:** The automatically generated summary provides immediate feedback to the user investigating the rumor.

## Overview of TwitterTrails

TWITTERTRAILS (Fig. 3) is an investigative and exploratory tool to analyze the origin and spread of a story on Twitter. While it does not answer directly the question of a story's validity, it provides information that a critically thinking person can use to examine how a Twitter audience reacts to the spreading of the story. TWITTERTRAILS takes as an input from the user a single tweet with information she wishes to investigate, like the one in Figure 1, but allows the user to input keywords from that tweet to collect a set of related tweets. From that set of related tweets it provides visualizations to pinpoint the origin of the investigative tweet: where the information trail started, who initially broke the story. In some cases this may be enough for the user, based on the reputation of the accounts which broke the story on Twitter.



**Figure 3:** The full story view page (top) and condensed view page (bottom) of twittertrails.com.

In cases of more dubious data, or for a more engaged Twitter user or journalist, TWITTERTRAILS provides

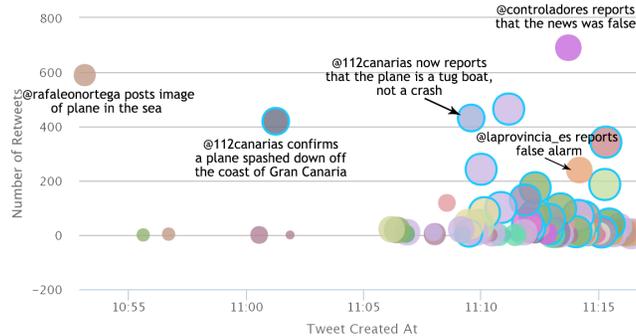
visualizations to trace not only the origin, but the spread of a story. It gives the user tools to answer important questions about the story, as the ones mentioned in the Introduction section. Propagation and Timeline visualizations give the user a meaningful way to browse the data, while network graphs give her an overview of influential users in the data. Moreover, minutes after the launching of an investigation TWITTERTRAILS will give the user a summary of the findings that in most cases may be enough to answer her questions (Fig. 2). If she wants details on how this summary is produced, she can look into each of the sections that the investigation produced.

TWITTERTRAILS is structured around the investigation of a single tweet, which is the first input the user provides (via the url of the tweet). After retrieving the investigative tweet, TWITTERTRAILS provides the Keyword Selection interface, to allow the user to highlight words and phrases from the tweet as keywords, or enter them manually. The system helps the user select the appropriate keywords in a variety of ways. It also suggests new search terms from common words, bigrams and hashtags in the 100 most recent tweets.

The first tool we present to the user is the Propagation Graph (Figure 4): a novel visualization which shows who broke the story on Twitter, and highlights influential and independent content creators. The burstiness algorithm is used to identify the time when the story breaks, and the propagation graph shows the first hundred tweets in the breaking interval.

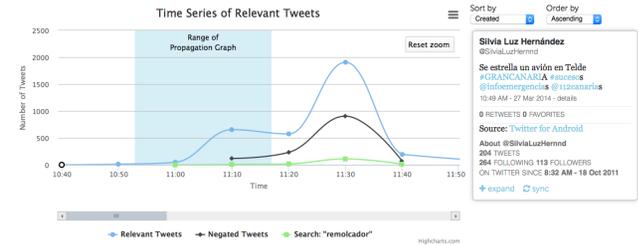
A data point in the Propagation Graph represents a single tweet, and is plotted in several dimensions: the x-axis, which shows *time*; the y-axis, which shows the

number of *retweets* received; and the size of a point, which represents the number of *followers* the tweeter has (scaled logarithmically). Tweets written by *verified* accounts are marked by a bright blue border. We claim that these are key elements in gauging the visibility of the tweet, as well as the degree of credibility other users will assign to the tweet and the amount of trust in the user as a source of information. Since we are trying to track the flow of a story, time is a natural factor to observe. But there are more dimensions that are depicted on the Propagation graph. Tweets with similar language (based on cosine similarity) have the same color, in an attempt to visualize *content independence*.



**Figure 4:** The Propagation Graph from the Plane in the Sea story.

The web interface allows users to view the tweets represented by points on the graph by hovering over or clicking on the points. Studying the Propagation Graph (Fig. 4), we discover some facts about how the “Plane in the Sea” story developed. Despite an admission by the 9-1-1 type service @112canarias about false alarm, news organizations kept propagating the false news for more than an hour.



**Figure 5:** The Timeline visualization from the “Plane in the Sea” story. Selecting a data point brings up a pane with all the tweets sent during this 10-minute interval. Three series are shown in this graph: all the relevant tweets, the negating tweets, and those the user chose to search for containing a particular keyword: remolcador (tug boat). It appears here that the negating tweets have succeeded in affecting the propagation of the rumor.

The Timeline visualization gives an overview of the whole story (Figure 5). The negation tweets are also displayed as a series in this graph, to show when tweets denying the story began to spread. The next two visualizations (not shown here), the Retweet Network and the Co-Retweeted Network, help to answer questions about the main actors who were spreading information. Finally, TWITTERTRAILS produces a summary in the form of a report that refers to its main findings.

## References

- [1] N. Diakopoulos, S. Goldenberg, and I. Essa. Videolyzer: Quality analysis of online info. video for bloggers and journalists. In *ACM CHI*, 2009.
- [2] A. Marcus, M. Bernstein, O. Badar, D. Karger, S. Madden, and R. Miller. Twitinfo: Aggregating and visualizing microblogs for event exploration. In *ACM CHI*, 2011.