Audience Analysis for Competing Memes in Social Media

Samuel Carton
University of Michigan
scarton@umich.edu

Souneil Park
Telefonica Research
souneil.park@telefonica.com

Nicole Zeffer
University of Michigan
nzeffer@umich.edu

Eytan Adar
University of Michigan
eadar@umich.edu

Qiaozhu Mei
University of Michigan
qmei@umich.edu

Paul Resnick
University of Michigan
presnick@umich.edu

Abstract

Existing tools for exploratory analysis of information diffusion in social media focus on the message senders who actively diffuse the meme. We develop a tool for *audience analysis*, focusing on the people who are passively exposed to the messages, with a special emphasis on competing memes such as propagations and corrections of a rumor. In such competing meme diffusions, important questions include which meme reached a bigger total audience, the overlap in audiences of the two, and whether exposure to one meme inhibited propagation of the other.

We track audience members' *states of interaction*, such as having been exposed to one meme or another or both. We analyze the marginal impact of each message in terms of the number of people who transition between states as a result of that message. These marginal impacts can be computed efficiently, even for diffusions involving thousands of senders and millions of receivers. The marginal impacts provide the raw material for an interactive tool, RumorLens, that includes a Sankey diagram and a network diagram. We validate the utility of the tool through a case study of nine rumor diffusions. We validate the usability of the tool through a user study, showing that nonexperts are able to use it to answer audience analysis questions.

Introduction

Analyzing the diffusion of competing memes through social media is an important task. A marketing analyst tracking discussion of their product and a competitor's, a journalist tracking positive and negative reactions to an event, and a political analyst tracking the relative popularity of candidates all have reason to do this type of analysis. In this paper we focus on the case of social media rumors, in which the two competing memes are posts that propagate a given rumor and posts that debunk or correct it.

Existing tools for exploratory analysis of meme diffusion focus only on the message senders who actively diffuse the meme. The people who were potentially exposed but did not propagate the meme are not directly represented. This perspective makes certain basic questions hard to answer, such as "how many people were reached by this meme?" It also fails to support inferences about more sophisticated questions such as the overlap in audiences of the two or whether exposure to one inhibited propagation of the other. Throughout this paper we refer to this type of analysis, which requires tracking the people who were potentially exposed to the memes, as *audience analysis*. We contrast this with *propagator analysis*, the conventional sender-focused perspective.

Audience analysis is difficult. The data is much larger in scale than for propagator analysis because for each person propagating a meme, many more are passively exposed to it. It also requires knowing who follows the users who propagated the meme, which in turn requires significantly more effort to retrieve data through rate-limited search APIs. Even when the data has been assembled, there remains the question of how to analyze and visualize it.

Our solution to the last problem is to represent the diffusion of a meme as the mass movement of individuals between states of interaction with that meme. Every post of a meme causes some number of people to make transitions between states, depending on who is a follower of the tweet's author and what state they were in when they saw it. These *marginal impacts* are calculated in a precomputation step, and then displayed in an interactive visualization.

The main contribution of the paper is to demonstrate the feasibility of tools for exploratory recipient-based audience analysis, especially competitive meme diffusions. Subsidiary contributions include:

- Showing that it is possible to efficiently precompute the marginal impact of each propagation in terms of counts of user state transitions.
- Developing an interactive visualization that is easier for people to use for audience analysis than a tabular view of the state transition counts.

Audience analysis of rumors and their corrections is the use case that motivated our development of the tool and hence our name for it, RumorLens. We use this domain to provide concrete examples throughout the paper, though the techniques we discuss are general and could be applied to any set of competing memes.

Related Work

Many visualizations of diffusion use variations of a network diagram, with nodes representing either senders or messages and edges representing follower or reshare relationships. Various techniques are used to simplify the display, including balloon treemaps (Viégas, Wattenberg et al. 2013), node collapsing (Taxidou and Fischer 2014), and adroit use of circular layouts, node sizing and color (Ratkiewicz, Conover et al. 2011, Ren, Zhang et al. 2014).

It is difficult to represent temporal dynamics in network diagrams. Stacked graphs solve this problem by collapsing the activity in each time period into quantities for different categories and showing changes with a horizontal time component. ThemeRiver (Havre, Hetzler et al. 2000) is the earliest approach to apply this technique to large text corpora, and has inspired much subsequent work. Dork, Gruen et al. (2010) are among the first to apply this technique to social media posts, showing the evolution of topics involved in an event on Twitter over time. Later papers add new features or models to this basic approach, such as topic models (Cui, Liu et al. 2011), topic competition and "opinion leaders" (Xu, Wu et al. 2013), sentiment analysis (Wu, Liu et al. 2014), anomaly detection (Zhao, Cao et al. 2014) and topic cooperation (Sun, Wu et al. 2014).

Other approaches have found success as well. Whisper (Cao, Lin et al. 2012) uses radial "floret" diagrams in combination with other elements to surface temporal, geographical and community features of the diffusion of a topic over Twitter in real-time. Vox Civitas (Diakopoulos, Naaman et al. 2010) and TwitInfo (Marcus, Bernstein et al. 2011) both offer dashboards of conventional visualizations of an evolving topic, enhanced in the latter case by an algorithm identifies important sub-events. SocialHelix (Cao, Lu et al. 2014) uses a novel helix visualization to show the divergence of sentiment about a topic on Twitter over time.

These diffusion visualizations have in common that they focus on message senders, showing either individual items or frequencies of categories, topics, or sentiments. Our approach focuses on the experience of message recipients.

RumorLens is most closely related to other systems that use flow diagrams, often in the form of Sankey diagrams, to represent temporal events. Wongsuphasawat and Gotz (2012) introduce the Outflow visualization, which visualizes collections of event sequences as flows between common subsequences whose width are determined by their frequency. Frequence (Perer and Wang 2014) is quite similar, incorporating a frequent event mining stage that allows the technique to be used to explore arbitrary collections of event sequences. von Landesberger, Bremm et al. (2012) combine Sankey diagrams with geographic maps to represent sequences of spatiotemporal events. Ogawa, Ma et al. (2007) and Vehlow, Beck et al. (2014) both use the related parallel sets technique to visualize the evolution of communities, Ogawa in open source software projects, Vehlow in generic social networks. The innovation in RumorLens is to formulate information diffusion as a sequence of temporal exposure events that can be visualized this way. RumorLens also introduces a temporal component that previous systems of this type have not explored.

RumorLens System

RumorLens consists of two stages: a precomputation stage and a visualization stage. The first stage takes a dataset of tweets that have been tagged as propagating or correcting a rumor, as well as the social network of the propagating users, and calculates the marginal impact of each tweet. The second stage visualizes the result of the first.

Precomputation

The precomputation stage performs a one-pass calculation of the marginal impact of each tweet. It can be applied to any set of social media posts that have been labeled into two (or more) classes, and for which the follower list of each author is available. For concreteness, we describe the system in terms of our motivating use case, where the posts are tweets and the two classes of tweets are those that are spreading a given rumor and those that are correcting it.

This stage begins with the set of labeled tweets and a list of follower IDs of each distinct author in that set. From publicly available data, we do not know who actually saw each tweet. In this paper we treat following the author of a tweet as an indicator of potential exposure. Should a more direct measure of exposure become available, the broad procedure we outline would be able to accommodate it.

Maintaining a map of what state each follower is in at any given time, the algorithm iterates through the tweets in chronological order. For each tweet, it determines which, if any, followers of the author of that tweet are prompted to make a state transition by the contents of that tweet. The algorithm has both time and space complexity $O(nd_{AVG})$

where n is the number of tweets and d_{AVG} is the average follower count of propagating users.

For example, if user A has observed only a tweet or tweets of the rumor, then they are in state "Exposed to Rumor". If A then observes a correction tweet from another user they are following, A's state changes to "Exposed to Both". This indicates that A now potentially knows about both types of information, and that any future actions A takes, such as propagating the correction, should be interpreted in this light. As each tweet is processed, the count of each type of transition caused by that tweet is recorded to a database. It is these snapshots of tweets' marginal impact that are visualized by the RumorLens tool.

RumorLens uses a set of exposure states that can answer basic questions about audience and propagator pool size and overlap. It consists of just 7 states: an initial "No exposure" state, and an "Exposed to _____" and "Tweeted _______" state for the rumor, the correction, and for both. We ignore multiple exposures to or propagation of the same meme. We also do not track further exposure after a user has propagated a meme on the basis that since they have already endorsed a position, so further passive exposure is not relevant. This is by no means the only possible set of exposure states to track—we address this topic further in the discussion section.

The precomputation process yields a temporal database with one row for each tweet and one column for each possible state transition (in our case, 14). Each tweet induces a state transition for some of the followers of its author. Each column in a tweet's row contains the number of people that tweet caused to move along that state transition. With a small number of states and thousands of tweets, the total amount of data transferred to the browser is manageable.

Figure 2 shows a raw table view of the information sent to the browser. Each row represents one tweet. The bottom row, for example, represents the 163rd tweet in the dataset; it propagated the rumor and caused 781 people to transition from the "Not exposed" state to the "Exposed to rumor" state and 8 people to transition from the "Exposed to correction" state to the "Exposed to both" state.

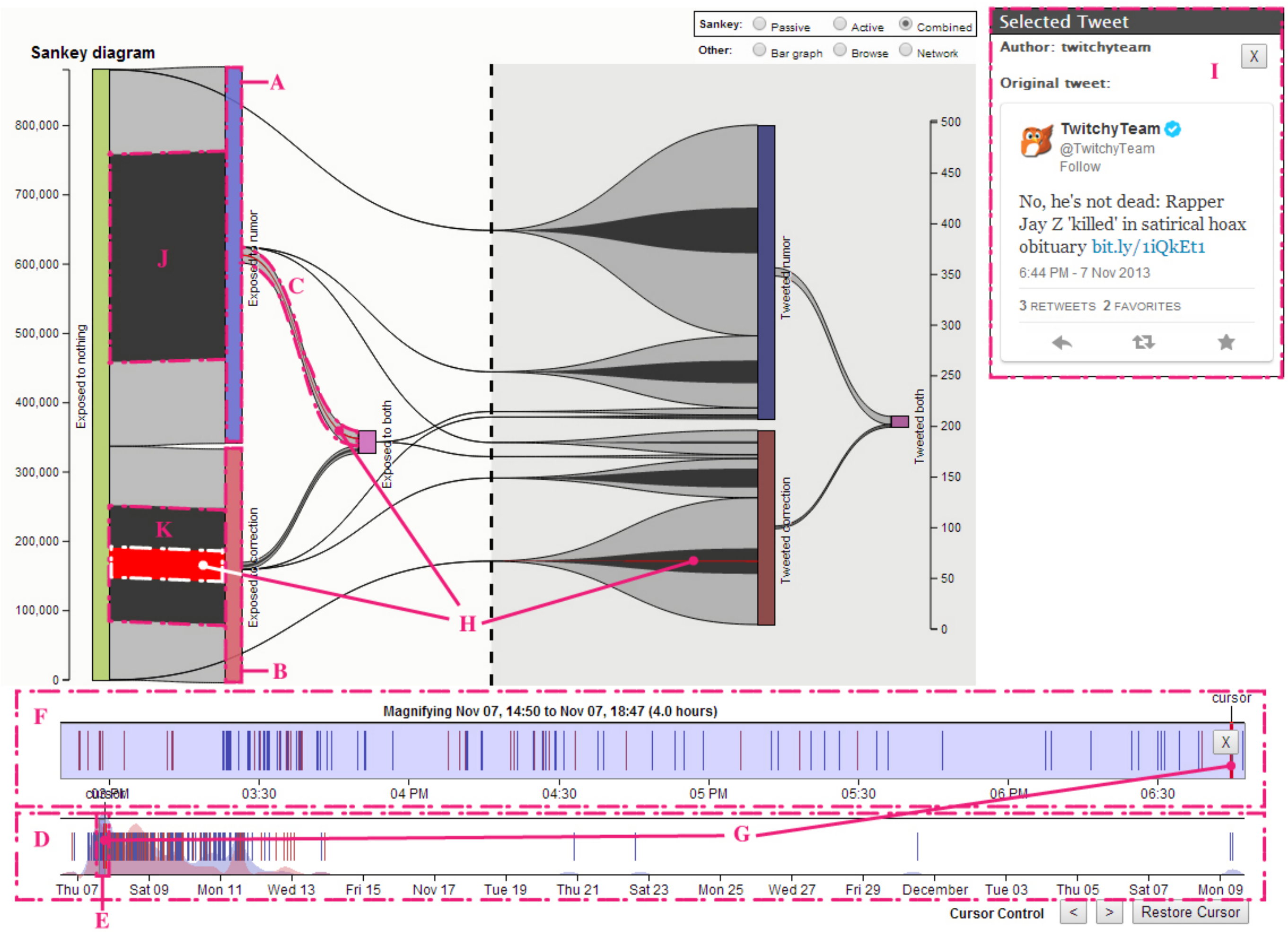


Figure 1: RumorLens interface displaying Sankey diagram

Visual Elements

The RumorLens interface (Figure 1) consists of three coordinated panes, one displaying a timeline of all the tweets (D and F in Figure 1), one for displaying information about a single selected tweet (I in Figure 1), and a central diagram showing information about the aggregate marginal impacts of tweets (the central section of Figure 1). Currently, three alternative visualizations have been implemented for this central piece: a Sankey diagram (Figure 1), a raw table view (Figure 2), and a network diagram (Figure 3).

Tweet timeline

The tweet timeline (D in Figure 1) shows all the tweets as vertical line items spaced out along a labeled timespan, color-coded blue or red for the two competing memes (rumor and correction in our motivating use case).

The analyst can select a window of the timeline (E in Figure 1), shown in a magnified timeline directly above the complete timeline (F). Information about the accumulated marginal impacts of tweets in the selected time-span is displayed on the central diagram, with each diagram displaying this time sub-span information in a different way. The black flows shown on the Sankey diagram are one example of a visual representation of this information.

By dragging the selected time window across the full timeline, the effects shown in the marginal impacts area smoothly and instantly adjust to reflect the newly selected window. This allows the analyst to quickly find periods of time with unusual activity, and to explore the time dynamics of the diffusion process.

On the right side of the tool, there is a panel that presents information related to a single tweet, which can be selected by clicking on a timeline item. Each of the views that can be shown in the marginal impacts pane also has a way of showing the effect of the currently selected tweet. The red flows on the Sankey diagram in Figure 1 are one example of how this information is represented visually.

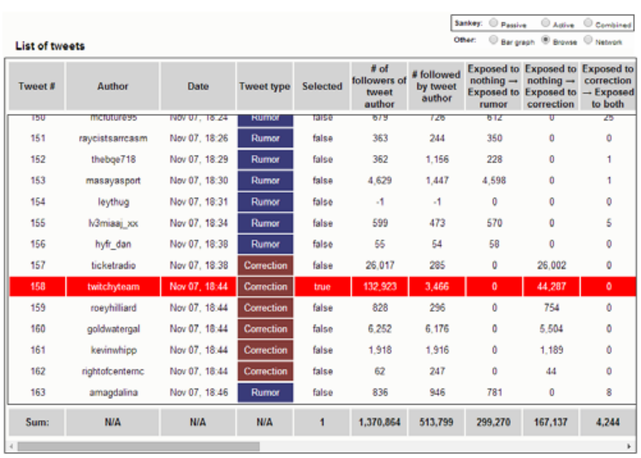
Marginal impact visualization

The most information-rich display-option is the raw table shown in Figure 2. Column footers display the sums of all rows currently contained in the list. Column headers can be used to sort the list by respective columns, allowing easy navigation to extremes of impact and other metrics. When a timespan is highlighted on the timeline, those tweets outside the timespan are removed from the table and the column sums update accordingly. When a tweet is selected from the timeline, its row is highlighted in a bright color.

A more visually accessible option is the Sankey diagram, which represents the aggregate state transitions that occur as a result of the set of tweets that made up the rumor in question. The layout of the nodes conveys information about which transitions are possible (only left-to-right). The size of each node and flow is proportional to the number of tweets it represents, and each flow represents one column sum from the table view. When a timespan is highlighted on the timeline, black sub-flows appear which encode the total transition within that timespan. The visual representation enables comparisons between quantities at a single time or across time as the time slider is moved. When a single tweet is selected, its impact is shown as a set of brightly colored sub-flows overlaid on top of the black sub-flows (e.g., H on top of K in the figure).

The central marginal impacts pane can be switched to a more common representation, a force-directed network layout (Figure 3). Each tweet is represented as a node. Links represent followership: a link exists when the author of a later tweet is a follower of the author of an earlier tweet. Dashed arrows connect tweets by the same author.

Nodes are color-coded blue and red, as in the timeline. As in the London riots rumor visualizations (Proctor, Vis et al. 2011), the size of a node represents the impact of the individual who made the tweet. We use the marginal impacts rather than the raw follower counts: in some cases users with many followers who tweet late may have small impact due to overlapping audiences. (see Table 1).



Tweet #	Author	Date	Tweet type	Selected	# of followers of tweet author	# followed by tweet author	Exposed to nothing → Exposed to rumor	Exposed to nothing → Exposed to correction	Exposed to correction → Exposed to both
150	mcfuture95	Nov 07, 18:24	Rumor	false	679	726	612	0	25
151	raycistsarcasm	Nov 07, 18:26	Rumor	false	363	244	350	0	0
152	thebqe718	Nov 07, 18:29	Rumor	false	362	1,156	228	0	1
153	masayasport	Nov 07, 18:30	Rumor	false	4,629	1,447	4,598	0	1
154	leythug	Nov 07, 18:31	Rumor	false	-1	-1	0	0	0
155	lv3miaaj_xx	Nov 07, 18:34	Rumor	false	599	473	570	0	5
156	hyfr_dan	Nov 07, 18:38	Rumor	false	55	54	58	0	0
157	ticketradio	Nov 07, 18:38	Correction	false	26,017	285	0	26,002	0
158	twitchyteam	Nov 07, 18:44	Correction	true	132,923	3,466	0	44,287	0
159	roeyhilliard	Nov 07, 18:44	Correction	false	828	296	0	754	0
160	goldwatergal	Nov 07, 18:44	Correction	false	6,252	6,176	0	5,504	0
161	kevinwhipp	Nov 07, 18:44	Correction	false	1,918	1,916	0	1,189	0
162	rightofcenternc	Nov 07, 18:44	Correction	false	62	247	0	44	0
163	amagdalina	Nov 07, 18:46	Rumor	false	836	946	781	0	8
Sum:	N/A	N/A	N/A	1	1,370,864	513,799	299,270	167,137	4,244

Figure 2: Raw table

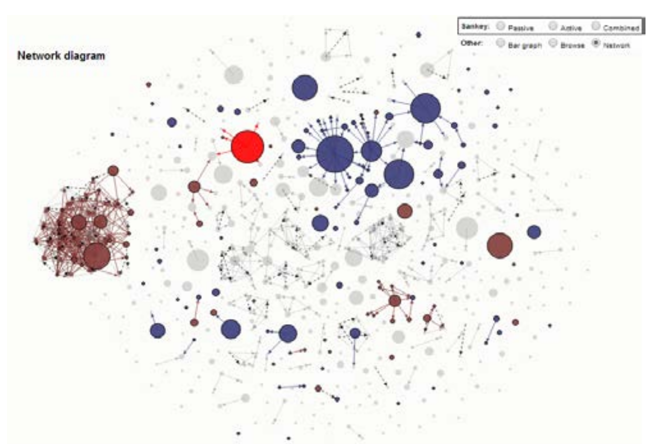


Figure 3: Network diagram

When a timespan is highlighted on the timeline, tweets not in the highlighted span are faded to grey. Because the diagram is force directed, dragging a time window box across the tweet timeline can be used to see the progress of the rumor through the community structure of Twitter. If a single tweet is selected, it is highlighted in a brighter color.

A sample rumor diffusion

To provide a more concrete illustration, we give an example of the type of audience analysis a user could perform with the system[1]. The rumor that we use for demonstration is a 2013 rumor that the rapper Jay-Z had died, first started by a satirical article in a music magazine.

The leftmost rectangles represent states of exposure. From the “Exposed to rumor” rectangle (A in Figure 1), we can see that roughly 550,000 people were exposed to the rumor with no previous exposure to the correction, and roughly 350,000 were exposed to the correction without having first seen the rumor. Hovering over the “Exposed to rumor” (A) and “Exposed to correction” (B) state rectangles would give us the exact numbers associated with these state movements. From the relative size of the “Exposed to rumor” to “Exposed to both” flow (C), we can see that very few people exposed to the rumor were subsequently exposed to the correction, as few people made the transition that would have been induced by a correction tweet.

The black flows overlaying the diagram (e.g. J and K) show the cumulative impact of the subset of tweets in the currently-highlighted timespan of the timeline. These features demonstrate that the 4-hour highlighted timespan accounted for roughly 1/2 of new exposures to the correction (K) and 1/2 of new exposures to the rumor (J). Again, a user could hover over those sub-flows to see the exact numbers associated with that timespan.

The bright red flows (H) represent the impact of a particularly influential correction tweet. The set of transitions spanned by red flows on the diagram indicate that the account had been exposed to the correction before tweeting, but not the rumor, and that it exposed a large number of new people to it. Hovering over that flow would reveal the number of people newly exposed: roughly 45,000.

The state rectangles to the right represent states of propagation activity. The large flow from “No interaction” to “Tweeted rumor” demonstrates that a majority of propagators of this rumor were exposed to it outside Twitter or at least outside of the feed of tweets from people they follow.

Finally, with some careful comparisons of particular flows, we can begin to understand user behavior with respect to this rumor. For instance, the very small size of the “Tweeted both” state and its corresponding flows indicates that users tended to express one opinion and stick with it rather than to vocally change their minds on the issue. Comparing the size of the “Exposed to both” to “Tweeted rumor” flow and the “Exposed to both” to “Tweeted correction” flow shows that while 8 people tweeted the rumor with knowledge of both types of information, only 4 tweeted the correction under those circumstances, indicating that the correction was not found compelling by participants in this rumor’s diffusion.

Case Study

We present a case study of the utility of the tool by using it to analyze nine rumors for which we collected both spreading and correcting tweets. We show that the tool can be used to generate interesting conclusions about the spread of rumors on Twitter. We also provide some evidence that authors’ follower counts are not always a good proxy for the impact that they have in a diffusion.

Data collection

Since we are interested in the interplay between the rumor and its corrections, we selected rumors that ultimately turned out to be false. Some were selected opportunistically, based on stories that came to our attention. Others were selected from the rumors surrounding the Boston Marathon bombing that were written up on Snopes.com and for which we found large numbers of tweets.

For the first rumor, a hoax about Fox news and a Russian meteor, we searched for the keywords “Russia,” “meteor,” “Fox” and “Obama” on the Topsy Twitter search API. All the retrieved tweets were manually labeled. For the other eight rumors we employed the Rec-Req system (Li, Wang et al. 2014) to collect and classify the tweets while requiring fewer human judgments.

We summarize the selected rumors below:

1. *Fox News on the Russian meteor*: This rumor states that Fox News accused President Obama of causing a large meteor to strike Russia in order to spread concern about global warming. 1,210 tweets: 920 rumor; 290 correction.

2. *HIV in Greece*: The World Health Organization (WHO) released a report in September of 2013 that includes a chapter saying that half of new HIV infections in Greece are self-inflicted for the purpose of claiming monthly benefits from the government. 6,939 tweets: 3,941 rumor; 2,996 correction.

3. *JayZ is dead (inside)*: The Rap Insider, a music news publication, ran a story with the title ‘Rapper Jay-Z found dead inside at 43’, a satirical piece about his attitude and his music. 622 tweets: 349 rumor; 265 correction.

Boston Marathon Bombing: A number of rumors spread through social media in the wake of the bombings.

4. *8-year old girl*: One rumor claimed that one of the victims of the bombing was an 8-year-old girl, often claim-

[1] A video demonstration of this sample rumor diffusion analysis is available at http://www.youtube.com/embed/HuvpiNGmFYE

ing that she was a survivor of the Sandy Hook Elementary School shooting who was running in honor of her classmates. 12,010 tweets: 9,426 rumor; 2,584 correction.

5. *Finish line proposal*: This rumor claimed that one victim of the bombing was a woman whose boyfriend was planning to propose to her on the finish line but instead knelt over her as she died. 10,055 tweets: 9,426 rumor; 640 correction.

6. *False victims*: This rumor claimed that some or all of the victims of the bombing were crisis actors sporting fake injuries, implying that the US Government planned the bombing as a "false flag" operation. 1,924 tweets: 1,740 rumor; 184 correction.

7. *Sandy Hook principal*: Another "false flag" rumor states that the late principal of Sandy Hook Elementary School, who died in the shootings there, was present at the Boston Marathon Bombing. 9,109 tweets: 8,452 rumor; 657 correction.

8. *Man on the roof*: An image of a man standing on a rooftop overlooking the marathon route as the bombs were going off sparked rumors that the pictured man was either involved in the attack or under investigation by the Boston Police. 9,523 tweets: 8,611 rumors; 912 corrections.

9. *Facebook early creation*: Some Facebook memorial pages for victims were converted from unrelated pages that had existed prior to the bombing. The creation dates for these pages sparked a rumor that some agency had foreknowledge of the attacks and posted the memorial early. 12,583 tweets: 12,041 rumor; 542 correction

Reach of rumors and corrections

In our datasets, total exposure to the rumor ranges from a minimum of approximately 550,000 for the rumor about Jay-Z's death to a maximum of 7.7 million for that about early Facebook pages.

The Sankey diagram lets us visually compare exposure to the rumor against exposure to the correction. The rumor about the Sandy Hook principal showed the greatest disparity, with a ratio of more than 30 rumor exposures per correction exposure. The HIV rumor had the least disparity between rumor and correction exposure with a ratio of roughly 1.5 rumor exposures per correction exposure, reflecting the fact that the original source of the incorrect information (the WHO) participated in trying to correct it.

Table 1 compares our state-based method of exposure estimation against a naïve technique for estimating the total exposure to each rumor. The "Marginal exposure" column shows our estimate, made by tracking the state transitions of each individual involved in the spread of the rumor. The "Follower counts" column shows the estimate that could be made by simply summing the follower counts of the propagators of the rumor. The table shows that naively summing follower counts results, on our examples, in overestimating total exposure by between 6% and nearly 400%.

Table 1:Comparison of exposure estimates made by summing marginal exposure vs. follower counts

Rumor	Marginal exposure	Follower counts	Ratio
8-year old girl	4,026,105	6,329,859	157%
Facebook early creation	7,765,743	14,588,797	188%
Finish line proposal	4,967,139	7,668,395	154%
False victims	1,280,002	2,530,857	198%
Sandy Hook Principal	5,135,771	9,423,956	184%
Man on the roof	7,363,057	22,599,367	307%
H.I.V in Greece	3,971,845	15,157,016	382%
JayZ is dead (inside)	562,263	788,552	140%
Russian meteor	1,661,729	1,772,952	107%
White House explosion	1,141,498	1,403,032	123%

Speed of diffusion

By leaving the left edge of the sliding window on the far left and dragging its right edge to the right (gradually highlighting the entire timeline) an analyst can study the behavior of a rumor over time. From the data we can see that rumors often spread quite quickly.

For the eight rumors that reached at least one million users, the time it took to do so ranged from 1.2 hours for the Facebook rumor to 10.75 days for the (much smaller) rumor about false victims at the marathon. The time it took to reach 50% of full exposure varied from 3 hours for the Facebook rumor to nearly 8 days for the false victims rumor. The false victims rumor seems to be an outlier in how wide its distribution was over time; no other rumor had a "half-life" of more than a day and a half.

Because corrections appear in response to rumors, one might expect them to lag behind rumors in the speed at which they gain exposure. The data supports this expectation in most cases, but for two rumors it is actually inverted: corrections about the man on the roof peaked 2 hours before the rumor did, while those about the false victims peaked a full 3 days earlier.

Overlap of audiences

By comparing the "No interaction to rumor exposure" flow with the "Rumor exposure to both exposure" flow, we can use RumorLens to examine how well the correction spread to users who had been exposed to the rumor. Conversely, by examining the "Correction exposure to both exposure" flow, we can see how many people had already been "inoculated" with the correction by the time they were exposed to the rumor.

Of people exposed to the rumor, the percentage of people who were later exposed to the correction, or who had already been exposed, ranges from 3% for the finish-line proposal rumor to 39% for the HIV in Greece rumor. HIV is an outlier in this, as the next highest level of such “contested” exposure is 10% for the Sandy Hook student rumor. This relative success may be because the HIV in Greece correction was spread by the same vector that originally spread the rumor, the WHO. Even for that rumor, however, a majority of the people exposed to the rumor were never exposed to a correction tweet.

We can also see what percentage of exposure to the correction was received by users who had been exposed to the rumor: from about 1 in 10 for the rumor about Jay-Z’s death to 1 in 2 for the Facebook rumor.

Effectiveness of corrections

By focusing on the active tweeting states on the right of the diagram, we can explore some of the factors that affect how people spread the rumor and correction, and in particular explore how corrections impact the spread of rumors.

From the “Tweet Rumor” and “Tweet Both” flows we can determine the percentage of users who spread the rumor at least once, and who then change their mind and start spreading the correction: a maximum of about 1 out of 8 people who tweeted the HIV in Greece rumor subsequently tweeted a correction, while fewer than 1 in 1000 did so for the finish-line proposal rumor.

Exposure to the correction may be more or less compelling from rumor to rumor. Among users who are exposed to both rumor and correction and subsequently tweet, a little more than half tweeted the correction in the case of the HIV in Greece rumor. By contrast, fewer than 1 in 10 did so in the case of the Facebook rumor.

How endogenous are rumors to the social graph?

By comparing the flows between the “Start”, “Exposed to Rumor”, and “Tweeted Rumor” states on the active user diagram, we can study what proportion of people who spread the rumor or correction draw their knowledge from sources other than person-to-person hearsay on Twitter, either through other media or through other features of Twitter such as search or following hashtags.

For propagators of the rumor, this percentage varied widely, from more than 4 in 5 for *Jay-Z’s death* to about 1 in 4 for *Russian meteor*. The disparity suggests that different rumors will be more or less responsive to interventions that take place solely on the Twitter social graph.

User Study

A small-scale user study was performed to evaluate the usability of the RumorLens system by nonexperts.

Subjects

31 subjects were recruited from a 100-level undergraduate introduction to information studies course. Subjects ranged from freshman- to junior-year students and represented a broad spectrum of majors, including Computer Science and Economics. 20 males subjects participated; 11 female.

Experiment

We test the usability of the tool by testing the ability of

Table 2: User study questions

	Question
Individual-related	Q1 Identify the user who first tweeted the **rumor**
	Q2 Identify the user who exposed the most people to the **correction** with a single tweet
	Q3 Identify the user who inspired the greatest number of subsequent tweets
	Q4 Identify the user who tweeted the **correction** the largest number of times
	Q5 Identify the user who exposed the greatest number of people to the **correction** who had previously been exposed only to the **rumor**
	Q6 Identify the first user to tweet the **rumor** who had more than 75,000 followers
	Q7 How many people did this user expose to the **rumor**?
Audience-related	Q8 Were more people exposed to the rumor or the **correction**?
	Q9 Roughly speaking, how many people were exposed to the **correction**?
	Q10 How many people were exposed to the **rumor**?
	Q11 What was the distribution of tweet impacts in spreading the **rumor** and **correction**?
	Q12 How many people were exposed to **both the rumor and the correction**?
	Q13 How many people tweeted the **correction**?
	Q14 How many people first tweeted the **rumor**, then changed their minds and tweeted the **correction**?
	Q15 How long did it take for 400,000 people to be exposed to the **rumor**?
	Q16 How long did it take for half of all people to be exposed to the **correction**, who would ever be exposed to it?
	Q17 Were people who had been exposed to both the rumor and the correction more likely to tweet the rumor, or the correction?
	Q18 How many people were exposed to **both the rumor and the correction**?

subjects, with a small amount of training, to effectively use it to analyze the spread of a rumor. To understand the relative strengths of the three visualizations currently offered by RumorLens (Sankey, table and network), each subject was limited to the use of one visualization only.

Each subject session proceeded as follows:

1. The subject was randomly assigned to one of the three visualizations: the Sankey diagram, the network diagram or the raw information table.
2. The subject was lead through a 15-minute tutorial covering the definition of a rumor, the concept of transitioning between states of exposure, and the features of the visualization they were working with.
3. The subject was asked to answer a set of 18 multiple-choice questions (Table 2) about a rumor ("Jay-Z is dead") using the tool, 6 about important individuals in the spread of the rumor, and 11 about the rumor's audience size.

The questions were selected heuristically to represent a range of plausibly interesting questions an analyst might ask about the spread of a rumor, including questions about both audience sizes and notable individuals, questions that required more or less inference, and both precise and approximate questions. All were presented as multiple-choice questions, so that we could easily assess correctness.

Each question came with an option stating: "this question is difficult or impossible to answer with the tool". Subjects were informed that part of their task was to decide, for a given question, whether the visualization was appropriate for that question.

Results

In Table 3 we report the speed and accuracy of subjects in answering questions using the three different visualizations. For each visualization, we report the average across subjects of the percentage of questions they felt they were able to answer, their accuracy when they did so, and their speed in answering questions. We report these statistics for both types of questions, individual and audience-related. To check for statistical significance, we conducted pairwise t-tests comparing each of the other conditions to the Table condition, using an alpha value of 0.025 to adjust for the multiple comparisons.

The results demonstrate that the Sankey diagram is highly effective for the audience-related questions, with all subjects able to answer almost all such questions, with a collective mean accuracy of 85%. The raw table view, though it contains all the same information as the Sankey diagram, was harder to use, with fewer questions answered and more time taken per question. Two time-related questions required subjects to resize or drag the time window in order to find the times when audiences reached certain sizes. Even on these questions, subjects did well with the Sankey diagram: nine of ten subjects answered Q15, all getting it right, and all ten answered Q16, eight getting it right. Using the table, only six of 10 answered Q15, 4 getting it right, and 3 of 10 answered Q16, 2 getting it right.

On questions pertaining to significant individuals, few subjects were able to use the Sankey diagram. Subjects had some success using the raw table, answering 60% of questions at 61% accuracy. Surprisingly, the network diagram was not better than the raw table for these questions.

Discussion

The precomputation process yields a summary of the marginal impact of each tweet as a set of counts of people transitioning between states of interaction with the rumor. Summing counts over a sequence of tweets aggregates those impacts, which can be displayed as column sums in the table view or visually in the Sankey diagram. The case study shows that an analyst who is very familiar with the tools can find interesting insights by visually noticing anomalies. The user study shows that college students with minimal training can quickly answer a variety of specific audience analysis questions.

Table 1 suggests that this type of analysis may be necessary to answer even very simple questions such as how widely a meme spread. Simply counting the size of follower lists may grossly overestimate the total audience reached by a set of tweets: in order to accurately assess the potential reach of a rumor, it is necessary to download the actual follower lists and remove duplicates. Some analyses may be difficult even to estimate without collecting follower lists. Once the follower lists are available, de-duplication and marginal impact calculation can be done in linear time.

Users were less successful at identifying significant individuals in the spread of a meme than in answering questions about audience size, overlap and behavior. This deficiency could be addressed in several ways. The network diagram could include a panel for sizing nodes by any

Table 3: Mean accuracy, percentage of questions answered, and elapsed time per question for each visualization and question type

*significant at $p<0.025$

	Individual-related			Audience-related		
	Answered	Accuracy	Time (s)	Answered	Accuracy	Time (s)
Sankey	0.14*	0.50	41.8*	0.99*	0.85	52.5*
Table	0.64	0.57	66.5	0.76	0.75	72.1
Network	0.53	0.36	63.0	0.40*	0.53*	59.3

available metric of tweet impact. The table view could include a search/filtering interface to allow easier navigation to individual records. On a large screen, all three visualizations could be displayed simultaneously, yoked so that action in any of the three would be reflected in the others.

The precomputed marginal impacts are compact enough to be sent to a browser and cached. They enable display of net impacts in the Sankey diagram, for any time period, and instantaneous update as the time window slider is dragged across the screen. There are tradeoffs involved, however, in providing only the compact precomputed values. They do not enable *all* possible analyses. For example, if an analyst wants to find out the number of people who were exposed to both competing memes *during particular time periods*, the marginal impacts are not sufficient. The aggregated marginal impacts provide a count of how many people reached the state "Exposure to both" during any time period, but does not indicate how many of those people started the time period without any exposure.

We imagine several possible real-world uses for audience analysis tools such as RumorLens. One is for journalists covering specific rumors. They would need to use other tools to retrieve and classify the set of tweets propagating or correcting the story, such as the Rec-Req system we used for the case study in this paper (Li, Wang et al. 2014). Given those tweets, they might check the audience size and whether most of the tweets about the rumor were already corrections, as a way to determine the newsworthiness of the story. They might include audience-analysis statistics in stories they write, and might use the individual analysis tools to identify tweet authors worth interviewing. They might also make the exploratory visualization tools available to readers who wanted to dive in for themselves. Another possible use-case is a brand manager who monitors social media for mentions of the brand or particular products. In that case, the analyst might divide mentions into those that express positive vs. negative sentiments, or track tweets that mention two competing products.

As with any exploratory analysis tool, the value comes not from answering a single question or set of questions. For that, calculation of a single statistic or a tailored visual chart will be better. For each of the 18 questions posed in our experiment, we could have used the precomputed marginal impacts to provide the users with direct answers. Exploratory analysis tools are most useful when the number of possible questions an analyst might ask is large, or when the data themselves may direct the analyst's attention.

That audience analysis benefits from an exploratory tool. The number of states and flows in the Sankey diagram, and thus the number of possible visual comparisons, is limited. It is large enough, however, that the visual layout helps make sense of the many possible comparisons. Moreover, the number of possible analyses grows very large once we introduce time-window questions, such as how long it takes before a correction reaches half its eventual audience or whether there was ever an n-hour period where the correction reached more new people than the rumor did.

That said, repeated use of the exploratory tool may lead to identification of certain questions that should be answered automatically for each new meme diffusion, rather than requiring interaction from the analyst. For example, for any journalist examining the diffusion of a rumor, it may be helpful, prior to engaging in any interactive analysis, to see counts of the number of people exposed to the rumor tweets, to correction tweets, and the percentage of each who were also exposed to the other. However, to identify such emergent needs, it helps for there to be existing available tool, a role which RumorLens fills.

State choice

The essence of our approach is to define a set of states of exposure, compute marginal impacts of each tweet in terms of transitions between those states, and then visualize aggregated marginal impacts across sets of tweets. The utility depends on a designer making a suitable choice for the set of states and creating a good visual for the states and flows.

In principle, every distinct sequence of exposures could define a different sate. That is, exposure to A, then, A again, then B, could be one state and A-B-A a separate state. That would provide the most detailed information to the analyst, but the extra detail would make it harder to notice more general patterns. The designer's choice of states to track and display leads to collapsing some of these exposure sequences and treating them as equivalent. For example, in the RumorLens system described in this paper, we collapse all sequences that include at least one exposure to both a rumor and its correction (and no action of tweeting either one) into the single state "Exposed to both." This makes it easy to notice what fraction of the audience is ever exposed to both, but hides whether audiences for the rumor or correction tended to get multiple exposures.

Consider a couple of alternatives that would afford different analyses. One may define states in terms of proportion of exposures to A or B (e.g., "More B", "Equal", and "More A"). All sequences with more A than B exposures would be treated as equivalent. This would make it easy to see whether, among people who were exposed to both, the amount of exposure tended to favor one or the other.

Another possibility is to define states based on the last exposure. For example, there could be one state if the last exposure was to A and that was the first exposure to A, another state if the last exposure was to A and it was not the first exposure to A. Then, the Sankey diagram would make it easy to see whether, in any given time period, most of the exposures were repeat exposures.

Ultimately, the choice of states to model depends on one's research questions and assumptions about user behavior. The underlying principle of morphing the diffusion into a transition diagram between states of exposure or interaction, however, would remain the same, as would the processes of computing and aggregating marginal impacts.

Conclusion

This paper describes the need for exploratory data analysis tools for information diffusions on social media that focus on the passive audiences rather than just the propagators. We describe such an analysis tool, RumorLens, which solves the data complexity problem by summarizing a diffusion event as the movement of participants between states of interaction with the information being diffused.

Using rumors as a motivating use case, we demonstrate that the precomputation required for audience analysis can be performed efficiently. Through a case study, we show that the RumorLens implementation of audience analysis can be used to pull out interesting facts about particular rumor diffusions. Through a user study, we show that college students without special training can understand the tool and use it to answer audience analysis questions.

Acknowledgments

This work is partially supported by the National Science Foundation under grant numbers IIS-0968489 and SES-1131500. This work is also partially supported by Google.

We thank Yuncheng Shen, Jessica Hullman David Askins, Renee Gross, Michael Betzold, and Ron Dzwonkowski for their helpful critiques of the design of the tool.

References

Cao, N., Y.-R. Lin, X. Sun, D. Lazer, S. Liu and H. Qu. 2012. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics* 18(12): 2649-2658.

Cao, N., L. Lu, Y.-R. Lin, F. Wang and Z. Wen. 2014. SocialHelix: visual analysis of sentiment divergence in social media. *Journal of Visualization*.

Cui, W., S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu and X. Tong. 2011. Textflow: Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics* 17: 2412–2421.

Diakopoulos, N., M. Naaman and F. Kivran-Swaine. 2010. Diamonds in the rough: Social media visual analytics for journalistic inquiry. *VAST*, IEEE.

Dork, M., D. Gruen, C. Williamson and S. Carpendale. 2010. A visual backchannel for large-scale events. *IEEE Transactions on Visualization and Computer Graphics* 16: 1129–1138.

Havre, S., B. Hetzler and L. Nowell. 2000. ThemeRiver: Visualizing theme changes over time. *InfoVis*, IEEE.

Li, C., Y. Wang, P. Resnick and Q. Mei. 2014. ReQ-ReC: High Recall Retrieval with Query Pooling and Interactive Classification. *ACM SIGIR*.

Marcus, A., M. S. Bernstein, O. Badar, D. R. Karger, S. Madden and R. C. Miller. 2011. Twitinfo: aggregating and visualizing microblogs for event exploration. *SIGCHI*, ACM.

Ogawa, M., K.-L. Ma, C. Bird, P. Devanbu and A. Gourley. 2007. Visualizing social interaction in open source software projects. *APVIS*, IEEE.

Perer, A. and F. Wang. 2014. Frequence: interactive mining and visualization of temporal frequent event sequences. *19th international conference on Intelligent User Interfaces*, ACM.

Proctor, R., F. Vis and A. Voss. 2011. How riot rumours spread on Twitter. *The Guardian*.

Ratkiewicz, J., M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini and F. Menczer. 2011. Truthy: mapping the spread of astroturf in microblog streams. *WWW*, ACM.

Ren, D., X. Zhang, Z. Wang, J. Li and X. Yuan. 2014. WeiboEvents: A Crowd Sourcing Weibo Visual Analytic System. *Pacific Visualization Symposium (PacificVis)*, IEEE.

Sun, G., Y. Wu, S. Liu, T.-Q. Peng, J. J. Zhu and R. Liang. 2014. EvoRiver: Visual analysis of topic coopetition on social media.

Taxidou, I. and P. M. Fischer. 2014. RApID: A System for Real-time Analysis of Information Diffusion in Twitter. *CIKM*, ACM.

Vehlow, C., F. Beck, P. Auwärter and D. Weiskopf. 2014. Visualizing the Evolution of Communities in Dynamic Graphs. *Computer Graphics Forum*.

Viégas, F., M. Wattenberg, J. Hebert, G. Borggaard, A. Cichowlas, J. Feinberg, J. Orwant and C. Wren. 2013. Google+ ripples: A native visualization of information flow. *WWW*, International World Wide Web Conferences Steering Committee.

von Landesberger, T., S. Bremm, N. Andrienko, G. Andrienko and M. Tekusova. 2012. Visual analytics methods for categoric spatio-temporal data. *VAST*, IEEE.

Wongsuphasawat, K. and D. Gotz. 2012. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics* 18(12): 2659-2668.

Wu, Y., S. Liu, K. Yan, M. Liu and F. Wu. 2014. OpinionFlow: Visual analysis of opinion diffusion on social media. *IEEE Transactions on Visualization and Computer Graphics*.

Xu, P., Y. Wu, E. Wei, T.-Q. Peng, S. Liu, J. J. Zhu and H. Qu. 2013. Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics* 19: 2012–2021.

Zhao, J., N. Cao, Z. Wen, Y. Song, Y.-R. Lin and C. Collins. 2014. # FluxFlow: Visual Analysis of Anomalous Information Spreading on Social Media. *IEEE Transactions on Visualization and Computer Graphics*.