Information Diffusion in Social Networks - A survey

IOANNIS KALLIANTZIS, Aristotle University of Thessaloniki, Greece
THEOKTISTI MARINOPOULOU, Aristotle University of Thessaloniki, Greece
ADAMANTIOS MINAS, Aristotle University of Thessaloniki, Greece
KYRIAKI PANTELIDOU, Aristotle University of Thessaloniki, Greece

ABSTRACT Social networks are constructs where people can connect with each other and share information. Since their creation, social networks have been used thoroughly and have now become mainstream, in the sense that most people maintain a profile in at least one social network. While the network grows larger, so does the information it contains, thus the information that arrives to each user. The spreading of information in a social network is called Information Diffusion, and it is a topic of interest for social network researchers, especially with the rise of Fake News. In this survey, we studied the existing literature and we present algorithms that model the Information Diffusion in a network. Such models can be used to predict whether a piece information will pass down from a user to another, which can find multiple applications, for example in the spreading of Fake News.

ACM Reference Format:
Ioannis Kalliantzis, Theoktisti Marinopoulou, Adamantios Minas, and Kyriaki Pantelidou. 2018. Information Diffusion in Social Networks - A survey. 1, 1 (June 2018), 21 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Authors' addresses: Ioannis Kalliantzis, Aristotle University of Thessaloniki, Greece, ioankall@csd.auth.gr; Theoktisti Marinopoulou, Aristotle University of Thessaloniki, Greece, tmarinop@csd.auth.gr; Adamantios Minas, Aristotle University of Thessaloniki, Greece, adamantcm@csd.auth.gr; Kyriaki Pantelidou, Aristotle University of Thessaloniki, Greece, kyrpanpan@csd.auth.gr.

XXXX-XXXX/2018/6-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In this survey we study the information diffusion on Online Social Media from a perspective of trying to identify true and false news based on their diffusion patterns.

Information Diffusion is defined as "the process by which a piece of information is communicated through certain channels over time among the members of a social system". Described a piece of information as any new idea, news, article or media object considered new to an individual.

The study of information diffusion originates from the research of computer virus (epidemic) spreading over network. Information Diffusion in Social Networks describes the anatomy and intricacies of the network under consideration.

Knowing what information is being produced and where, along with to where, how and why it is being transferred, will help in identifying the hidden functionalities of Social Networks and to false news detection.

Critical or malicious information can spread uncontrollably over a network. The management and prediction of information diffusion in Social Networks is not possible until there is a model that can capture and analyse the hidden mechanism underlying diffusion.

The traditional analysis tools may show when and to where information propagated, but it fails to explain how and why did it propagate. The parameters like diffusion rate and who influenced whom plays a decisive role in understanding the dynamics of Social Networks.

Researchers have in recent years developed a variety of techniques and models to capture Information Diffusion in Social Networks, analyse it, extract knowledge from it and predict it. The aim of this survey is to study such models and create a knowledge base for future research.

In the section 2 of this paper we are going to analyze such models, in the section 3 we will see the connection between information diffusion and fake news, in section 4 we will acknowledge the sentiment factor in the information diffusion, in section 5 we will study the diffusion of media on social medias and in section 6 we will introduce a model that predicts which node will influence which other nodes in the network.

2 MODELLING APPROACHES

2.1 Cascades

Modelling Information Diffusion refers to the process of creating a model that can describe how information spreads between users. Such models have a wide variety of uses - for example, one could predict if a certain piece of information would spread to a specific user, or the amount of users it will spread to.

Most approaches embed in their logic a graph that describes the network. The nodes usually represent users, while the edges represent a relationship between the two users. Such a relationship can either be a following/follows relationship as in Twitter, or just representing that there exists a means of communication between the two users.

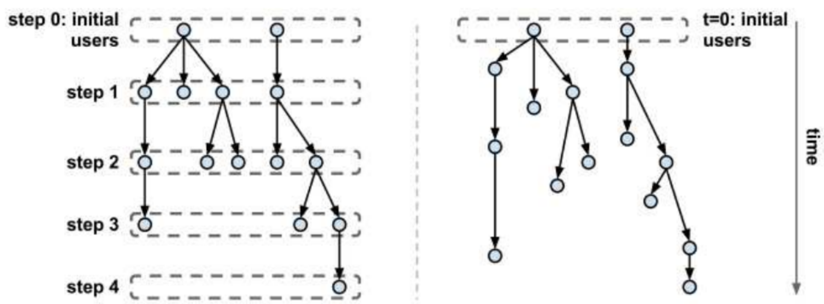
Two widely used Information Diffusion models are the Independent Cascade (IC) and Linear Threshold (LT) models. Both models are based on a directed graph, where each edge is associated with a diffusion probability in the case of IC or a weight for LT. They work in similar ways - both iteratively mark nodes as active or not based on the edge data, until all information has propagated through the network. A node is marked as active in an iteration of the algorithm if the information has successfully spread to it, or rather if a previously active node successfully influenced it. If the information did not spread, it is marked as inactive.

IC and LT differ on how they mark nodes as active or not, or in other words, how they model the propagation of information. IC uses a diffusion probability per edge - when an active node tries to influence an inactive one, it does so with this probability. Thus, it follows a sender-centric approach. On LT, an inactive node is activated if the total edge weights of all his active neighbours exceed a threshold T. This second approach is a receiver-centric one. The probabilities or edge weights of both models need to be estimated - such problems are known in machine learning as Maximum-Likelihood problems, where the task is to find such parameters that maximize a likelihood function.

Both IC and LT work synchronously, in a discrete time axis. On each time step, a piece of information is propagated to the neighbouring nodes, with a chance of activating them or not. This however is not the way social networks operate - in the real world, this process is asynchronous. Saito et al. proposed 2 extensions to both models, called AsIC and AsLT [6], meaning Asynchronous IC and Asynchronous LT respectively, which address this issue by introducing a time-delay parameter on each edge.

In AsIC, each edge (u, v) is also assigned a time-delay parameter $r_{u,v}$ along with its activation probability $k_{u,v}$, as discussed above. Starting from an initial active set of nodes S, the process unfolds in continuous time t. If a node u is activated on time t, it tries once to activate its inactive children. If it succeeds, they will become active on time $t + \delta$. For each child v, a delay-time δ is chosen from an exponential distribution with parameter $r_{u,v}$. If u still has any inactive children in time $t + \delta$, it attempts to activate them with probability $k_{u,v}$ for each child v, and if it succeeds, the child becomes active in time $t + \delta$.

Fig. 1. IC model (left) versus AsIC model (right)



AsLT employs the continuous time in a similar way to AsIC, with a slight difference - each time-delay parameter for each node u is only node dependent, so it is properly written as r_u. This further highlights the fact that the LT model is receiver-centric.

It is evident that time plays an important role in information diffusion, since it affects the users directly. In [2], Guille&Hacid present a model based on AsIC, that aims to capture the temporal dynamics taking place during diffusion. The purpose of this model is to predict diffusion. Since it is based on AsIC, it is a sender-centric model. The dataset used comprised of 467 million Twitter posts from 20 million users, over a 7 month period (June 1 2009 - December 31 2009). From this dataset they extracted multiple subsets with the following strategy: starting with an initial user (seed), they gathered his followers at a distance of at most 2 hops, and then connected the nodes based on their relationships (following or not). For each such subset, an AsIC cascade was created.

To estimate the cascade's parameters, they used machine learning on a number of features gathered per user. They distinguish between three types of features - social, semantic and temporal - social features concern the network's properties and relations, semantic features refer to the content of each tweet (Twitter post) and temporal features are derived from the activity of each user per time slot. More specifically, they employ the following features:

Social

(1) Activity - users activity, average number of tweets per hour
(2) Social homogeneity - similarity of sets of users that user1 and user2 talk to

(3) Directed tweets to user - ratio of tweets that are targeted to the user
(4) Boolean - true if user2 is mentioned in tweets of user1
(5) Mention rate - represents user popularity based on mentions

Semantic

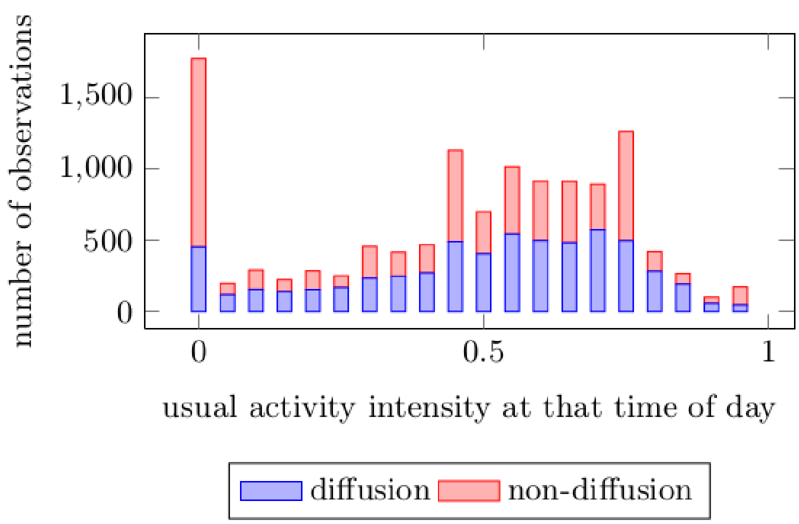
(1) Boolean - true if the user used a keyword present in his past exchanges in his current post

Temporal

(1) 6-Dimensional vector of user's activity (fraction of tweets) - each day is broken down into 6 blocks of 4 hours each, each position on the vector represents the user's activity in each block

Throughout their paper, Guille&Hacid highlight the importance of the temporal features in information diffusion. The intuition behind this comes from the fact that the active hours of a user greatly affect whether he will spread a post, since they explicitly affect if he will see it. Figure 2 shows how the user's activity affects whether diffusion happens or not - it is evident when the user is more active, more diffusions happen.

Fig. 2. Information diffusion distribution over user's activity

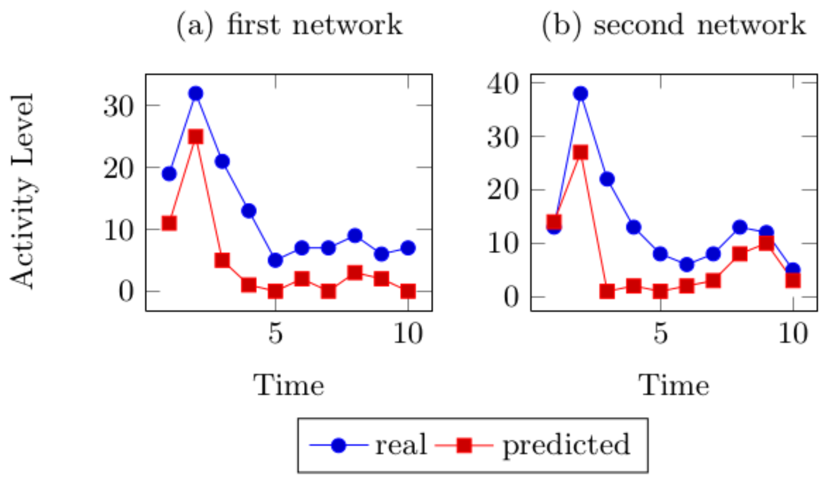


The aforementioned features are later used to train algorithms to estimate the cascade's parameters. The classification task is defined as follows: given a vector V as described above, the goal is to estimate the probability $P(Y|V)$, where $Y = diffusion, non - diffusion$. The authors experimented with Decision Trees (C4.5), Linear and Multilayer Perceptrons, as well as Bayesian Logistic Regression, which they finally settle to. Figure 3 shows the performance of the model for two different networks - in the first network the cascade was constructed with n=11 starting users, while the second network with n=14. Both networks were tasked with modelling the diffusion of a certain topic, which was the acquisition of a startup by Google.

2.2 Epidemiological Models

Other approaches use epidemiological models. Such models describe they way a disease spreads between neighbouring nodes, resembling the way information spreads on a social network. Jin et al. [4] present the use of epidemiological modelling for information diffusion of news and rumors on Twitter. They use the SEIZ model, which splits the user into four categories:

Fig. 3. Predictions for different amounts of starting users



(1) Susceptible (S)
(2) Exposed (E)
(3) Infected (I)
(4) Skeptic (Z)

Susceptible are the users who have not yet received the piece of information under study. Exposed are the users who have received the information, but have not tweeted yet about it. Infected are the users who have tweeted this piece of information, while Skeptic are users who have received the information but choose not to tweet about it. Time is also important in this model as well, since exposed users introduce a delay between receiving the information and tweeting about it.

In the SEIZ model, users are moved between categories with the following rules:

(1) With rate b, susceptible users may become skeptic with probability l, or they may become exposed with probability $(1 - l)$
(2) A susceptible user will automatically believe a news story or rumor with probability p, or he will be moved to the exposed category with probability $(1 - p)$.
(3) Exposed users can turn into infected by i) coming in contact with another infected user, with contact rate ρ, ii) or by himself, with rate ϵ.

The parameters of the model are described on table 1, as they appear on the original paper:

The model is described using the 4 Ordinary Differential Equations shown in figure 4. Parameter estimation is done using these equations, and the optimal parameters are the ones that minimize $|I(t) - tweets(t)|$, which is the absolute difference of the number of infected users and the number of tweets at time t.

Table 1. Parameters of the SEIZ model for Information Diffusion

Parameter	Definition
β	S-I contact rate
b	S-Z contact rate
ρ	E-I contact rate
ϵ	Incubation rate
$\frac{1}{\epsilon}$	Average Incubation Time
bl	Effective rate of S -> Z
β_ρ	Effective rate of S -> I
$b(1-l)$	Effective rate of S -> E via contact with Z
$\beta(1-p)$	Effective rate of S -> E via contact with I
l	S->Z Probability given contact with skeptics
$1-l$	S->E Probability given contact with skeptics
p	S->I Probability given contact with adopters
$1-p$	S->E Probability given contact with adopters

Fig. 4. Ordinary Differential Equations used to describe the SEIZ model

$$\frac{d[S]}{dt} = -\beta S \frac{I}{N} - bS \frac{Z}{N}$$

$$\frac{d[E]}{dt} = (1-p)\beta S \frac{I}{N} + (1-l)bS \frac{Z}{N} - \rho E \frac{I}{N} - \epsilon E$$

$$\frac{d[I]}{dt} = p\beta S \frac{I}{N} + \rho E \frac{I}{N} + \epsilon E$$

$$\frac{d[Z]}{dt} = lbS \frac{Z}{N}$$

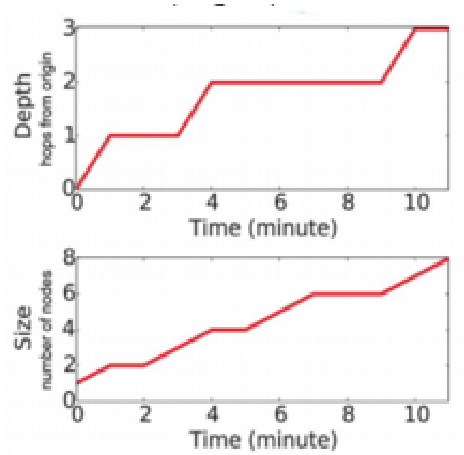
The dataset the authors used was a set of gathered tweets about specific topics, both news, like the Boston Marathon Bombings and rumors, like the Doomsday of December 21 2012. They demonstrated that epidemiological model like SEIZ can be successfully used to model Information Diffusion, for both news and rumors.

3 INFORMATION DIFFUSION AND FAKE NEWS

A rumor cascade begins on Twitter when a user makes an assertion about a topic in a tweet, which could include written text, photos, or links to articles online. Others then propagate the rumor by retweeting it. A rumorâĂŹs diffusion process can be characterized as having one or more cascades, which we define as instances of a rumor spreading pattern that exhibit an unbroken retweet chain with a common, singular origin. For example, an individual could start a rumor cascade by tweeting a story or claim with an assertion in it, and another individual could independently start a second cascade of the same rumor (pertaining to the same story or claim) that is completely independent of the first cascade, except that it pertains to the same story or claim. If they remain independent, they represent two cascades of the same rumor. Cascades can be as small as size one (meaning no one retweeted the original tweet). The number of cascades that make up a rumor is equal to the number of times the story or claim was independently tweeted by a user (not retweeted). So, if a rumor âĂIJAâĂİ is tweeted by 10 people separately, but not retweeted, it would have 10 cascades, each of size one. Conversely, if a second rumor âĂIJBâĂİ is independently tweeted by two people and each of those two tweets is retweeted 100 times, the rumor would consist of two cascades, each of size 100.

Here it has been investigated the differential diffusion of true, false, and mixed news stories using a comprehensive data set of all of the fact-checked rumor cascades that spread on Twitter from its inception in 2006 to 2017. The data include 126,000 rumor cascades spread by 3 million people more than 4.5 million times. It is sampled all rumor cascades investigated by six independent fact-checking organizations (snopes.com, politifact.com, factcheck.org, truthorfiction.com, hoax-slayer.com, and urbanlegends. about.com) by parsing the title, body, and verdict (true, false, or mixed) of each rumor investigation reported on their websites and automatically collecting the cascades corresponding to those rumors on Twitter. It has been cataloged the diffusion of the rumor cascades by collecting all English-language replies to tweets that contained a link to any of the aforementioned websites from 2006 to 2017 and used optical character recognition to extract text from images where needed. For each reply tweet, it has been extracted the original tweet being replied to and all the retweets of the original tweet. Each retweet cascade represents a rumor propagating on Twitter that has been verified as true or false by the fact-checking organizations. It has been quantified the cascadesâĂŹ depth (the number of retweet hops from the origin tweet over time, where a hop is a retweet by a new unique user) and size (the number of users involved in the cascade over time). As a rumor is retweeted, the depth and size of the cascade increase. (Fig.5).

Fig. 5. Cascade's depth and size



A greater fraction of false rumors experienced between 1 and 1000 cascades, whereas a greater fraction of true rumors experienced more than 1000 cascades (Fig. 6B); this was also true for rumors based on political news (Fig. 6D). The total number of false rumors peaked at the end of both 2013 and 2015 and again at the end of 2016, corresponding to the last U.S. presidential election. (Fig. 6C and E).

Politics was the largest rumor category in our data, with 45,000 cascades, followed by urban legends, business, terrorism, science, entertainment, and natural disasters (Fig.6F).

When we analyzed the diffusion dynamics of true and false rumors, we found that falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information. A significantly

Fig. 6. Number of Cascades and Political Cascades

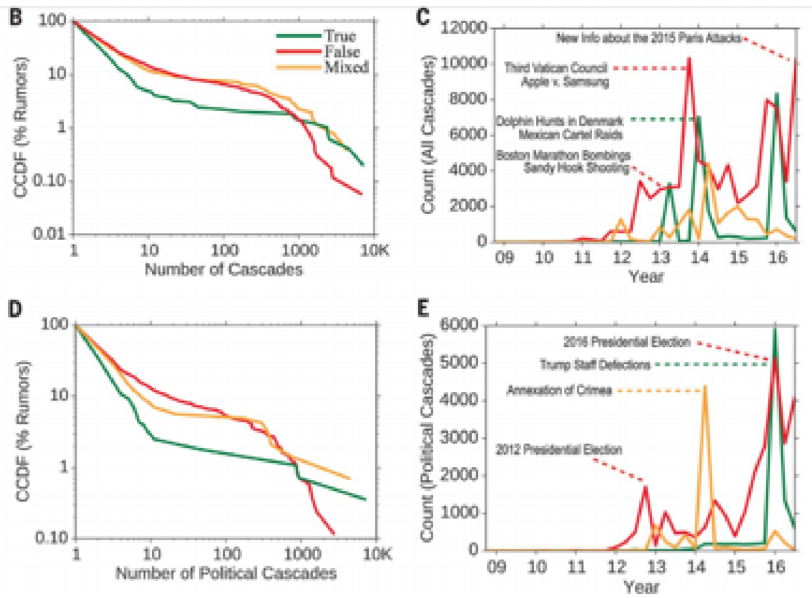
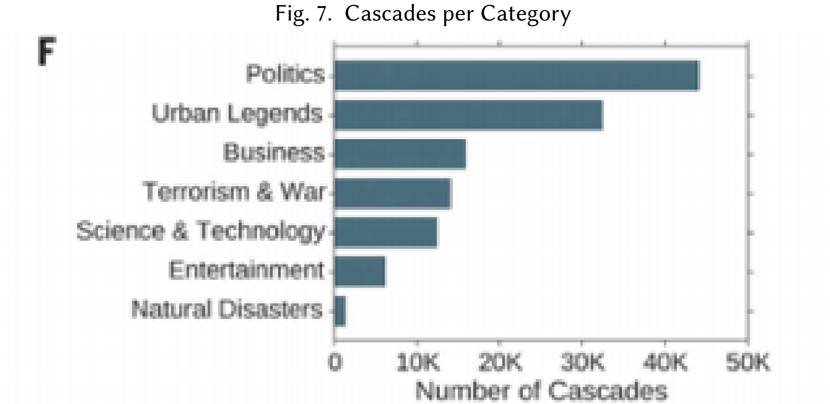


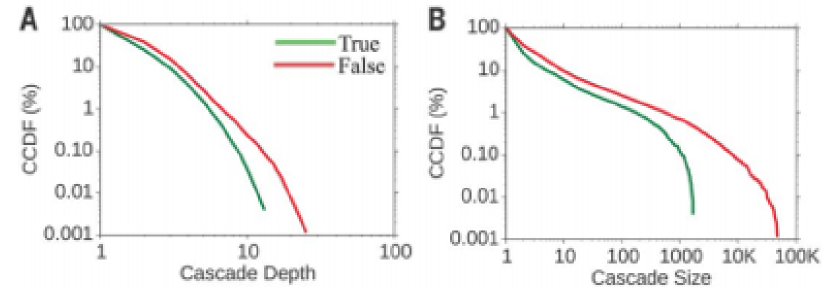
Fig. 7. Cascades per Category



greater fraction of false cascades than true cascades exceeded a depth of 10, and the top 0.01% of false cascades diffused eight hops deeper into the Twittersphere than the truth, diffusing to depths greater than 19 hops from the origin tweet (Fig. 7A). Falsehood also reached far more people than the truth. Whereas the truth rarely diffused to more than 1000 people, the top 1% of false-news cascades routinely diffused to between 1000 and 100,000 people (Fig. 7B).

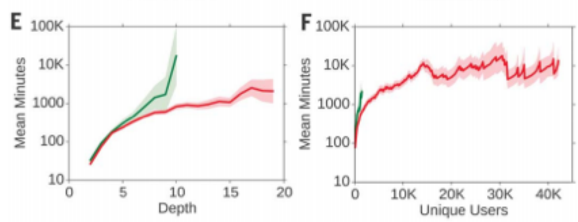
It took the truth about six times as long as falsehood to reach 1500 people (Fig. 8B) and 20 times as long as falsehood to reach a cascade depth of 10 (Fig.8A). As the truth never diffused beyond a depth of 10, we saw that falsehood reached a depth of 19 nearly 10 times faster than the truth reached a depth of 10 (Fig. 9E). False

Fig. 8. True and False cascade depth and size



political news traveled deeper (Fig. 10A) and more broadly (Fig. 10C), reached more people (Fig.10B) than any other category of false information.

Fig. 9. Mean Minutes for false and real news to be spread



One might suspect that structural elements of the network or individual characteristics of the users involved in the cascades explain why falsity travels with greater velocity than the truth. Perhaps those who spread falsity âĂIJfollowedâĂİ more people, had more followers, tweeted more often, were more often âĂIJverifiedâĂİ users, or had been on Twitter longer. But when we compared users involved in true and false rumor cascades, we found that the opposite was true in every case. Users who spread false news had significantly fewer followers, followed significantly fewer people, were significantly less active on Twitter, were verified significantly less often, and had been on Twitter for significantly less time (Fig. 11A). Falsehood diffused farther and faster than the truth despite these differences, not because of them.

Fig. 10. Diffusion of false political news

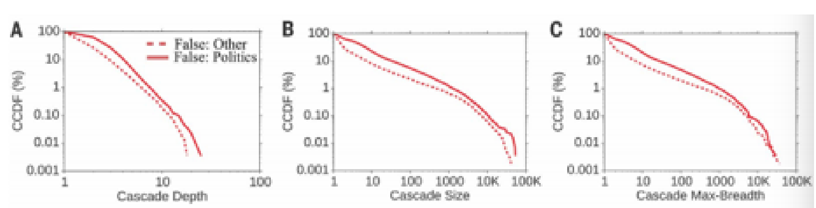
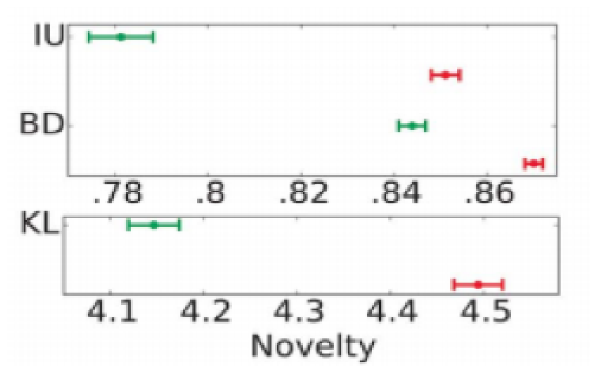


Fig. 11. Comparison of true and false users

	median		mean		mean (log)		stdv (log)		ks-test
	false	true	false	true	false	true	false	true	
followers	410	466	2234	5240	2.62	2.68	0.69	0.88	D=0.104, p∼0.0
followees	383	509	1002	1707	2.59	2.72	0.85	0.96	D=0.136, p∼0.0
verified	0	0	0.002	0.006	nd	nd	nd	nd	D=0.005, p<0.001
engagement	9.52	9.54	19.70	24.65	0.91	0.90	0.65	0.76	D=0.054, p∼0.0
account age	982	1214	1072	1269	2.90	2.97	0.39	0.42	D=0.125, p∼0.0

One alternative explanation emerges from information theory and Bayesian decision theory. Novelty attracts human attention, contributes to productive decision-making, and encourages information sharing because novelty updates our understanding of the world. When information is novel, it is not only surprising, but also more valuable, both from an information theoretic perspective and from a social perspective. We therefore tested whether falsity was more novel than the truth and whether Twitter users were more likely to retweet information that was more novel. We found that false rumors were significantly more novel than the truth across all novelty metrics, displaying significantly higher information uniqueness (Fig. 12).

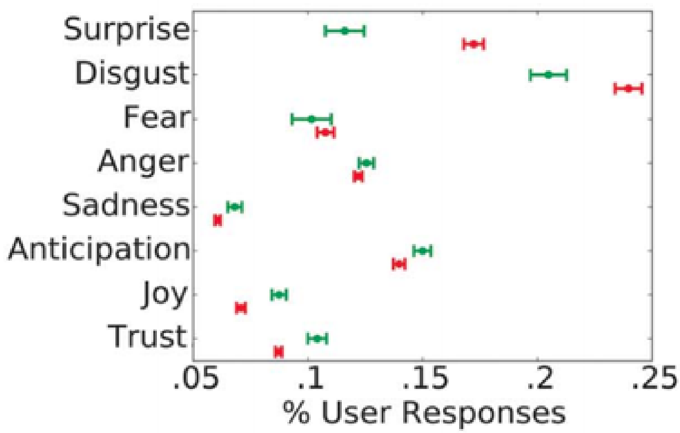
Fig. 12. Novelty of False and Real news



Although false rumors were measurably more novel than true rumors, users may not have perceived them as such. We therefore assessed usersâĂŹ perceptions of the information contained in true and false rumors by comparing the emotional content of replies to true and false rumors. It has been categorized the emotion in the replies by using the leading lexicon curated by the National Research Council Canada (NRC), which provides a comprehensive list of 140,000 English words and their associations with eight emotions based on PlutchikâĂŹs work on basic emotionâĂŤanger, fear, anticipation, trust, surprise, sadness, joy, and disgust âĂŤand a list of 32,000 Twitter hashtags and their weighted associations with the same emotions. It has been removed stop words and URLs from the reply tweets and calculated the fraction of words in the tweets that related to each of the eight emotions, creating a vector of emotion weights for each reply that summed to one across the emotions. It has been found that false rumors inspired replies expressing greater surprise (corroborating the novelty hypothesis, and greater disgust, whereas the truth inspired replies that expressed greater sadness, anticipation, joy, and trust (Fig. 13)

The emotions expressed in reply to falsehoods may illuminate additional factors, beyond novelty, that inspire people to share false news. Although we cannot claim that novelty causes retweets or that novelty is the only reason why false news is retweeted more often, we do find that false news is more novel and that novel information is more likely to be retweeted

Fig. 13. Emotions for real and false news



4 THE EFFECT OF SENTIMENT ON INFORMATION DIFFUSION

4.1 Sentiment analysis

To attach a sentiment score to the tweets in the dataset, it has been used SentiStrength, a promising sentiment analysis algorithm. The algorithm assigns to each tweet t a positive $S^+(t)$ and negative $S^-(t)$ sentiment score, both ranging between 1 (neutral) and 5 (strongly positive/negative). Starting from the sentiment scores, we capture the polarity of each tweet t with one single measure, the polarity score S(t), defined as the difference between positive and negative sentiment scores:

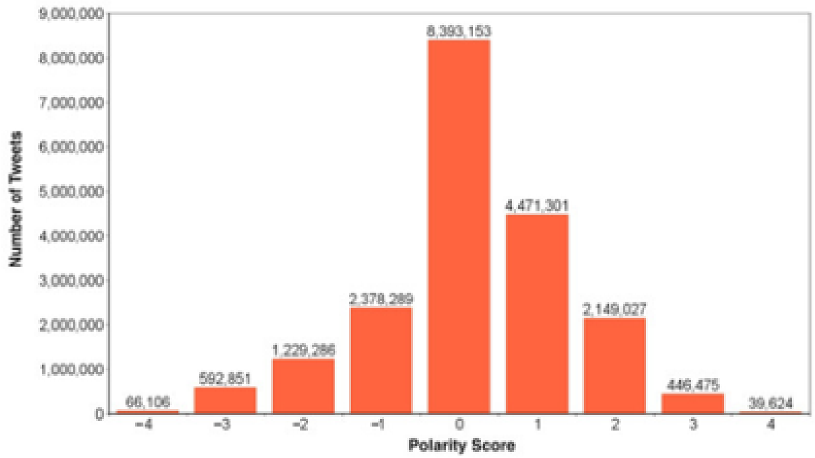
$$S(t) = S^+(t) - S^-(t) \tag{1}$$

The above-defined score ranges between −4 and +4. The former score indicates an extremely negative tweet, and occurs when $S^+(t) = 1$ and $S^-(t) = 5$. Vice-versa, the latter identifies an extremely positive tweet labeled with $S^+(t) = 5$ and $S^-(t) = 1$. In the case $S^+(t) = S^-(t) - positive$ and negative sentiment scores for a tweet t are the same - the polarity $S(t) = 0$ of tweet t is considered as neutral. There is a focus on the polarity score (rather than the two dimensions of sentiment separately) because previous studies highlighted the fact that measuring the overall sentiment is easier and more accurate than trying to capture the intensity of sentiment.

4.2 Dataset

The dataset adopted in this study contains a sample of all public tweets produced during September 2014 and it has been extracted all tweets in English that do not contain URLs or media content (photos, videos, etc.) produced in that month. This dataset comprises of 19,766,112 tweets. produced by 8,130,481 distinct users. All tweets are processed by SentiStrength and attached with sentiment scores (positive and negative) and with the polarity score calculated as described before. It has been identified three classes of tweetsâĂŹ sentiment: negative (polarity score $S \leqslant -1$), neutral ($S = 0$), and positive ($S \geqslant 1$). Negative, neutral, and positive tweets account for, respectively, 21.59%, 42.46% and 35.95% of the total. The distribution of polarity scores is captured by Fig 14.

Fig. 14. Distribution of polarity scores



4.3 Results

Here it has been studied the relation between content sentiment and information diffusion. Figure 15 shows the effect of content sentiment on the information diffusion dynamics and on content popularity. It has been measured three aspects of information diffusion, as function of tweets polarity scores: Fig.15A shows the average number of retweets collected by the original posts as function of the polarity expressed therein; similarly, Fig. 15B shows the average number of times the original tweet has been favorited; Fig. 15C illustrates the speed of information diffusion, as reflected by the average number of seconds that occur between the original tweet and the first retweet. Note that a large fraction of tweets are never retweeted (79.01% in our dataset) or favorited (87.68%):

Two important considerations emerge from the analysis of Fig.15: (i) positive tweets spread broader than neutral ones, and collect more favorites, but interestingly negative posts do not spread any more or less than neutral ones, neither get more or less favorited. This suggests the hypothesis of observing the presence of positivity bias (Garcia, Garas and Schweitzer, 2012) (or Pollyanna hypothesis (Boucher and Osgood, 1969)), that is the tendency of individuals to favor positive rather than neutral or negative items, and choose what information to favor or rebroadcast further accordingly to this bias. (ii) Negative content spread much faster than positive ones, albeit not significantly faster than neutral ones. This suggests that positive tweets require more time to be rebroadcasted, while negative or neutral posts generally achieve their first retweet twice as fast.

To investigate how sentiment correlates with content popularity, we now only consider active and exclusive discussions occurred on Twitter in September 2014. Each topic of discussion is here identified by its most common hashtag. Active discussions are defined as those with more than 200 tweets and exclusive ones are defined as those whose hashtag never appeared in the previous (August 2014) and the next (October 2014) month. Inspired by previous studies that aimed at finding how many types of different conversations occur on Twitter (Kwak et al., 2010; Lehmann et al., 2012), we characterize our discussions according to three features: the proportion pb of

Fig. 15. The effect of content sentiment

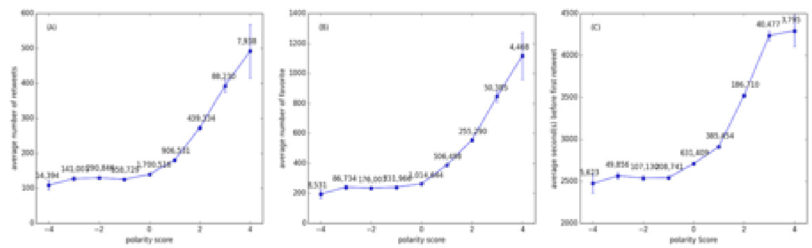
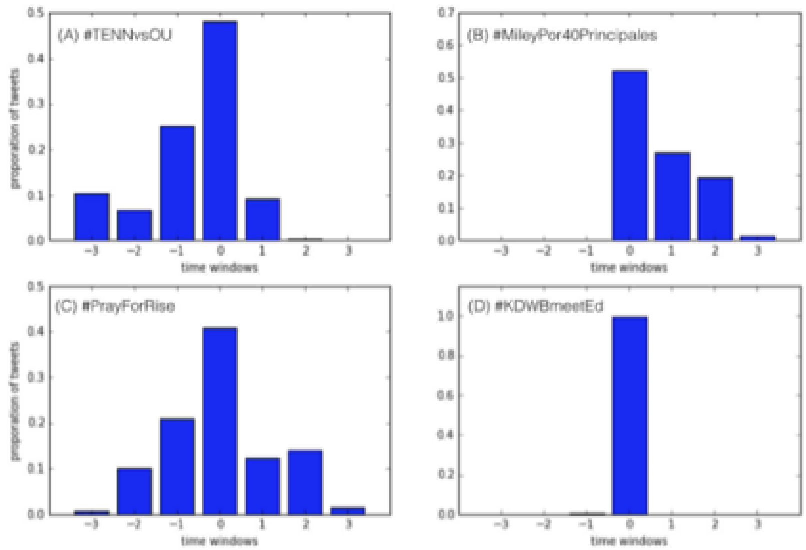


Figure 7(A,B,C): The effect of content sentiment

tweets produced within the conversation before its peak, the proportion pd of tweets produced during the peak, and finally the proportion pa of tweets produced after the peak. The peak of popularity of the conversation is simply the day which exhibits the maximum number of tweets with that given hashtag. The outcome of this process determines that the optimal number of components is four.i) anticipatory discussions, (ii) unexpected events, (iii) symmetric discussions , and (iv) transient events. Anticipatory conversations exhibit most of the activity before and during the peak. These discussions build up over time registering an anticipatory behavior of the audience, and quickly fade out after the peak. The complementary behavior is exhibited by discussions around unexpected events: the peak is reached suddenly as a reaction to some exogenous event, and the discussion quickly decays afterwards. Symmetric discussions are characterized by a balanced number of tweets produced before, during, and after the peak time. Finally, transient discussions are typically bursty but short events that gather a lot of attention, yet immediately phase away afterwards. According to this classification, out of 1,522 active and exclusive conversations (hashtags) observed in September 2014, we obtained 64 hashtags of class A (anticipatory), 156 of class B (unexpected), 56 of class C (symmetric), and 1,246 of class D (transient), respectively.

Figure 16 shows the evolution of sentiment for the four classes of Twitter conversations: it can be useful to remind the average proportions of neutral (42.46%), positive (35.95%), and negative (21.59%) sentiments in our dataset, to compare them against the distributions for popular discussions. (A) For anticipatory events, the amount of positive sentiment grows steadily until the peak time, while the negative sentiment is somewhat constant throughout the entire anticipatory phase. Notably, the amount of negative content is much below the dataset average, fluctuating between 9% and 12% (almost half of the dataset average), while the positive content is well above average, ranging between 40% and 44%. This suggests that, in general, anticipatory popular conversations are emotionally positive. (B) The class of unexpected events intuitively carries more negative sentiment, that stays constant throughout the entire discussion period to levels of the dataset average. (C) Symmetric popular discussions are characterized by a steadily decreasing negative emotions, that goes from about 23% (above datasetâĂŹs average) at the inception of the discussions, to around 12% toward the end of the conversations. Complementary behavior happens for positive emotions, that start around 35% (equal to the dataset average) and steadily grow up to 45% toward the end. This suggests that in symmetric conversations there is a general shift of emotions toward positiveness over time. (D) Finally, transient events, due to their short-lived lengths, represent more the average discussions, although they exhibit lower levels of negative sentiments (around 15%) and higher levels of positive ones (around 40%) with respect to the datasetâĂŹs averages.

Fig. 16. Example of four types of Twitter conversation



5 FAKE IMAGES DIFFUSION ON TWITTER

Nowadays, on social media users do not only share text but also numerous media files like audio, images, videos, GIFs etc. Fake news are not only spread through the text that a user is sharing but also through the media mentioned above. For this reason, we study and present a paper [3] that examined the fake photos' diffusion on Twitter during Hurricane Sandy.

Hurricane Sandy caused mass destruction and turmoil in and around USA from October 22nd to October 31st, 2012. Social media such as Twitter and Facebook were widely used by people to keep abreast about latest updates of the storm. Additionally, social media was also widely exploited by malicious entities during Sandy, to spread rumors and fake pictures in real-time. Such fake images and news became extremely viral and caused panic and chaos among the people affected by the hurricane. Hence, it is an ideal event, to analyze the spread and impact of fake and incorrect information on social media.

5.1 Dataset

For data collection from Twitter the authors of this work had a 24 âĹŮ 7 setup, which has been functional for about last 20 months. They collected data from Twitter using the Streaming API. This API enables researchers to extract tweets in real-time, based on certain query parameters like words in the tweet, time of posting of tweet, etc. They queried the Twitter Trends API after every hour for the current trending topics, and collected tweets corresponding to these topics as query search words for the Streaming API.

Hurricane SandyâĂŹs impact lasted from Oct. 20th to Nov. 1st, 2012, hence from all the tweets collected during this period, they filtered out tweets containing the words âĂŸsandyâĂŹ and âĂŸhurricaneâĂŹ. About 1.8 million tweets by 1.2 million unique users on Hurricane Sandy from Oct. 20th to Nov. 1st, 2012 was filtered out.

Using certain online resources (articles, tweets and blogs) they identified certain URLs that belonged to fake pictures of Hurricane Sandy. One of the prominent data sources used by them was the list of fake and real images made public by the Guardian news media company. The list provided by Gaurdian, classified the top image URLs shared during the hurricane as fake or real image URLs, which they used to form the dataset.

5.2 Methodology for tweets analysis

First they performed temporal analysis on the fake images tweets by analyzing how many such tweets were shared per hour on Twitter. Also, they analyzed the sudden peaks (from x1 hour to x1+1) in the graph more closely by constructing a retweet graph for the sudden peak in the temporal analysis, to find out what changes in the network topology lead to the viral spread of these images. This led to certain useful insights, about the nature and spread of fake image URLs on Twitter, which are summarized in the next subsection

Next, they analyzed what role the social network graph of a user on Twitter plays in propagation of fake URLs. The explicit social network of a user on Twitter, is that of his follower graph. We wanted to analyze what percentage of information diffusion takes place via this follower network graph of a user. The details of the algorithm used to compute are summarized in the following algorithm (Algorithm 1).

Fig. 17. Algorithm 1

Algorithm 1 Compute_Overlap

```
Create_Graph_Retweets()
Create_Graph_Followers()
for each edge in the retweet network do
    num_retweet_edges++
    Insert edge into hashmap, H[1..n]
end for
for each edge in the follower network do
    Insert each edge in hashmap, H[1..n]
    if collision then
        intersections++
    end if
end for
%overlap = (intersections/num_retweet_edges) * 100
```

In the function, Create Graph Followers, they crawled the follower network of all the unique users that had tweeted the fake images, using the REST API of Twitter. The network created had 10,779,122 edges and 10,215 nodes. In Create Graph Retweets, they created a retweet network, where an edge between two nodes exists if one user had retweeted the otherâĂŹs tweet. A hashmap, H[1..n], is created to compute the overlap between the follower and retweets graphs.

5.3 Results

They found that out of the 10,350 tweets identified by them, containing fake images URLs, about 86% were retweets. That is, only about 14% people posted original tweeted that contained such URLs. From the temporal analysis, they plotted the per hour tweeting activity of the fake images URLs. From Figure 18 we see that the fake URLs spread spikes at, 12 hours after the introduction of the URLs in the Twitter network. Then analyzed the spread of these picture URLs one hour before and after the spike by constructing the reply and retweet graph for the tweets sharing these fake picture URLs on October 29th, at 21 hours and 22 hours, as shown in Figure 20. We see that there are only a few users with very high degree, that is, only a few users results in majority of the retweets. They confirmed this statistically, Figure 19 (CDF) shows that top 30 users (0.3% of the users) resulted in 90% of retweets of the fake images. Combining results from both the graphs, we conclude that though the fake

URLs were present in the Twitter network for almost 12 hours before they became viral, also the sudden spike in their propagation via retweets happened only because of a few users.

Fig. 18. Details of data collected for the fake images URL sharing. Temporal distribution of tweets, hour wise, starting from the first hour that a fake image tweet was posted.

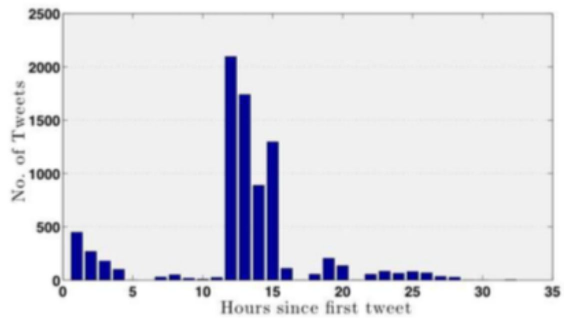
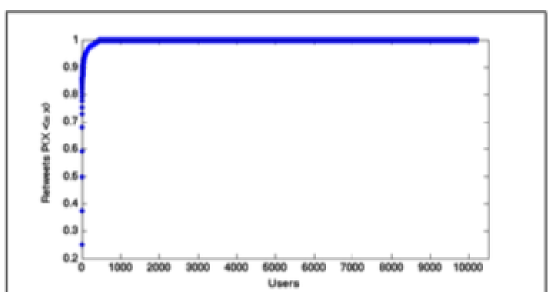
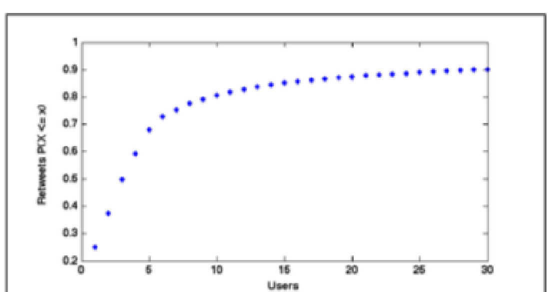


Fig. 19. CDF of retweets of the fake image tweets by the users. It shows that top 30 users (0.3% of the users) resulted in 90% of retweets of the fake images

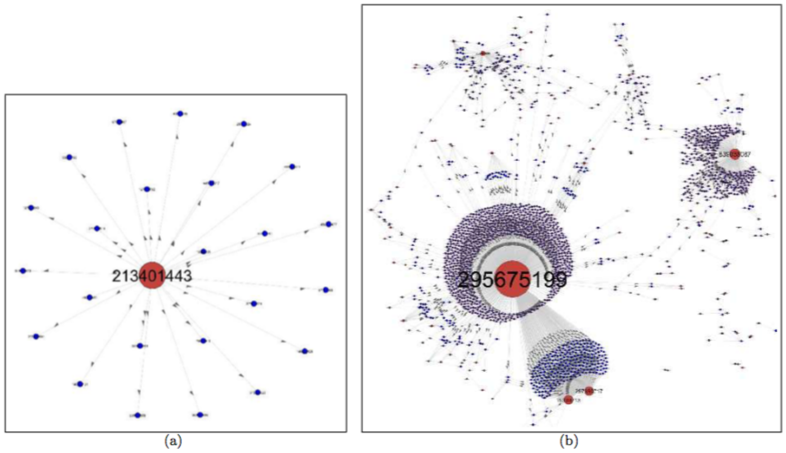


(a) All users



(b) Top 30 users

Fig. 20. Spread of fake pictures URLs (retweet and reply graph), the number on the node is user profile ID on Twitter. The figure shows that the fake images became viral very fast, within an hour there was a tremendous growth in the number of people tweeting them. (a) Oct. 29, 2100 hours (b) Oct. 29, 2200 hours.



6 MODELING INFORMATION DIFFUSION IN IMPLICIT NETWORKS

The diffusion of information has been an active research area recently. However modeling the diffusion in social networks has proven to be a challenging task, due to the difficulty in obtaining large scale diffusion data. Diffusion and cascading behavior models need to make several assumptions: (a) complete network data is available, (b) contagion can only spread over the edges of the underlying network, (c) the structure of the network itself is sufficient to explain the observed behavior. However, in many scenarios, the network over which diffusion takes place is in fact implicit or even unknown. Commonly, they only observe when nodes got âĂIJinfectedâĂİ but not who infected them.

6.1 Modeling diffusion

They address the above issues by developing a model of diffusion where no explicit knowledge of the network is necessary. Rather than predicting which node in the network will infect which other nodes, they focus on modeling the global influence a node has on the rate of diffusion through the (implicit) network.

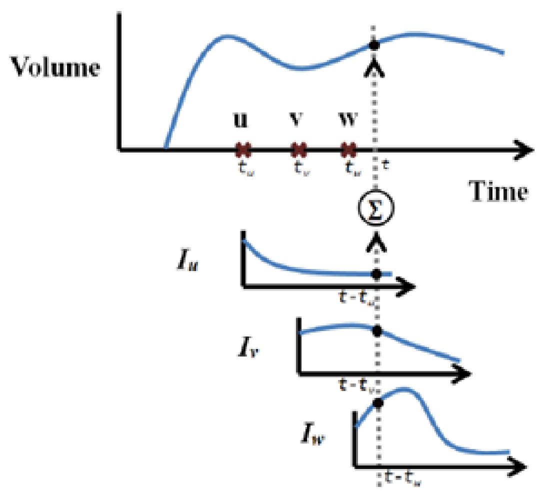
They formulate the Linear Influence Model (LIM) by starting with the assumption that the number of newly infected nodes depends on which other nodes got infected in the past. They then model the number of newly infected nodes as a function of the times when other nodes got infected in the past. In this model, each node has an influence function associated with it. Then the number of newly infected nodes at time t is a function of influences of nodes that got infected before time t. Node influence functions can be efficiently estimated by formulating a regression task where the goal is to learn an influence function Iu(t) for each node u such that the overall number of newly infected nodes at time t is the sum of influences of previously infected nodes. They model influence functions in a non-parametric way and show that they can be estimated using a simple least squares procedure. Even though their model is widely applicable, they restrict their discussion to the setting of information diffusion in online media.

About model formulation, Consider a set of nodes that participate in a diffusion process. As the information diffuses, nodes become "infected" when they adopt (mention) the information. They consider the setting where they observe only the time tu when a particular node u mentioned the information and do not require the knowledge of the network. They define the volume, V(t), as the number of nodes that mention the information at time t. They aim to model the volume over time as a function of which other nodes have mentioned the information beforehand. Each node u has a particular non-negative influence function Iu(l) associated with it. One can simply think of Iu(l) as the number of followup mentions l time units after node u adopted the information. Or equivalently, after node u mentions the information, this triggers an additional Iu(1) mentions in the next time step, Iu(2) mentions after two time steps, and so on. Now, they aim to model the relation between the volume V(t), and the influence functions of nodes u that mention the information at times tu (tu < t). They simply assume that the volume V(t) is the sum of properly aligned influence functions of nodes u:

$$V(t+1) = \sum_{u2A(t)} Iu(t{-}tu) \tag{2}$$

where A(t) denotes the set of already active (infected, influenced) nodes u that got activated prior to time t (tu âĽď t).

Fig. 21. Illustration of LIM model



The curve on the top of Figure 21 represents the volume V(t) over time, and tu, tv, and tw denote the times when nodes, u, v and w, got infected. After the nodes got infected, they each influence additional Iu(tâĹŠtu), Iv(tâĹŠtv) and Iw(tâĹŠtw) infections at time t. So the volume V(t) at time t is the sum of the influences of the three nodes.

A natural question then is how to model the individual influence functions Iu(l). The use of a non-parametric approach is the answer. This way they do not make any assumptions about the shape of the influence functions and let the model estimation procedure find the most appropriate shapes. They achieve this by considering the time to increase in discrete intervals (e.g., one hour). Then they can represent an influence function Iu(l) as a non-negative vector of length L, where lth value represents the value of Iu(l). Setting the length of vector Iu to L simply means that the influence of a node drops to zero after L time units.

6.2 Extensions

Accounting for novelty: So far, they have assumed that a node has the same influence regardless of how early or late in the diffusion they appear. This means that the influence of a node is same even if it mentions the information very early or very late. However, nodes are more likely to adopt novel and recent information while ignoring old and obsolete information. In order to account for this effect of recency and novelty they introduce a multiplicative factor Îś(t) that models how much more/less influential a node is at the time when it mentions the information. They refer to this model as a-LIM:

$$V(t+1) = a(t) \sum_{u2A(t)} Iu(t{-}tu) \tag{3}$$

Accounting for imitation: Another aspect of information diffusion and adoption is the effect of imitation, where nodes imitate one another because the information is popular and everyone talks about it. They refer to the contribution of the imitation as the latent volume in a sense that this volume is caused not by influence, but by other factors. They model the latent volume with an additive factor b(t) and refer to the model as the B-LIM model:

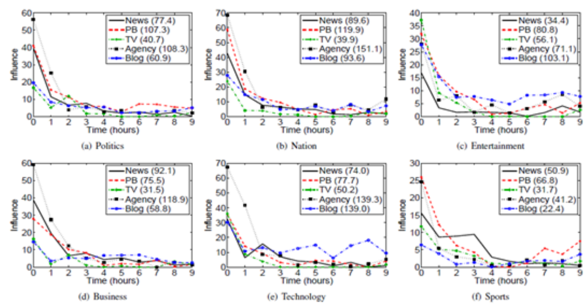
$$V(t+1) = b(t) + \sum_{u2A(t)} Iu(t{-}tu) \tag{4}$$

6.3 Experiments

Describing the dataset, first, they consider modeling the diffusion of short textual phrases over the online media space. With Memetracker methodology, they extract 343 million short textual phrases from a set of 172 million news articles and blog posts collected from more than 1 million online sources between September 1 2008 to August 31 2009. They choose 1,000 phrases with highest volume in a 5 day window around their peak volume. For each phrase, they track which websites mention it during 5 days around its peak volume.Second, they analyze the diffusion of hashtags on Twitter. They collect a stream of 580 million Twitter posts (40-50% of all posts) between June 2009 and February 2010. They identify 6 million different hashtags, and then discard hashtags that do not experience a significant peak in their volume. They then select 1,000 highest total volume hashtags during the 5 days around their peak volume. As Twitter users adopt at most 1% of the hashtags, we mitigate this data sparsity issue by grouping users into groups of 100 users. They consider 100 groups and model each group as a node. They then model the collective behavior of each group by aggregating all the mentions within the group. **Influence of textual phrases on Memetracker dataset:** For their experiment, they examine how the influence of various types of nodes changes depending on the topic of the information and the type of a node. Memetracker dataset consists of a wide range of media sites from traditional mass media such as newspapers, nationwide TV stations and press agencies, to modern online independent news sites, professional and personal blogs. They categorize textual phrases into six different topics. For each topic, they then estimate the influence functions of various types of sites Different types of media also have different types of textual phrases. For the purpose of the experiment, they identify five types of media: Newspapers (New York Times, USA Today), Professional blogs (Salon, Huffingtonpost), TV stations (ABC, CBS), News agencies (AP, Reuters) and (personal) Blogs. In total they select 22 sites, and group them in the above five groups. Now, they estimate the influence functions of 22 media sites on each of the six topics, by fitting LIM with the phrases in the topic.

Figure 6(a) shows the amount of influence of users grouped based on their total volume. All groups tend to have similar form of total influence. The group with the third largest volume has the most total influence, while the highest volume group has the lowest. Similarly, Figure 6(b) shows the influence functions of users grouped

Fig. 22



based on their total number of followers. Surprisingly, they find that the Twitter users with the intermediate number of followers have much higher influence than the highest in-degree nodes.

The results on Twitter nicely align with the experiments on Memetracker data. As the Memetracker experiments focused on online media and adoption of short textual phrases, they find that the mainstream media holds the most influential position in the dissemination of news content. On the other hand, hashtags on Twitter are a very different type of contagions. Hashtags are not news but rather socially contagious tags that are adopted in a distributed manner without a central supervision. Therefore, the diffusion of hashtags is mostly governed by the Twitter social/information network. This way Twitter users with âĂIJtoo highâĂİ number of followers, which usually correspond to celebrities and organizations, may be very influential in propagating the âĂIJinformationâĂİ contagions such as news, but not in diffusing more âĂIJsocialâĂİ contagions such as hashtags.

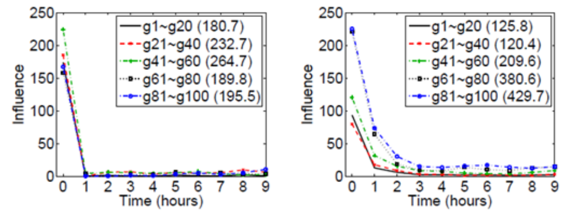
Notice that in general influence functions tend to decay rapidly over time. While the decay is particularly pronounced for business and politics, for entertainment or sports the influence seems to last somewhat longer.Similarly, the influence of bloggers tends to be lower at start, but tends to last longer (in particular for entertainment and technology). Politics, business, technology and the nation tend to be dominated by news agencies. Professional blogs are the second in terms of total influence in politics and national news, newspapers are the second in business, and personal blogs are in technology. In entertainment and sports, the situation is somewhat reverse. For entertainment it is the personal blogs that are the most influential, while for sports it is the professional blogs followed by the newspapers.

Influence of users on Twitter: They consider a set of 10,000 Twitter users, and aggregate them into 100 groups of 100 users. They consider two different types of grouping. First, they order users by the amount of their activity (hashtag volume) and second they order them based on the number of their followers. They fit B-LIM and examine the relation between the hashtag volume and the influence they have on the adoption of hashtags across the whole Twitter network.

Figure 23(a) shows the amount of influence of users grouped based on their total volume. All groups tend to have similar form of total influence. The group with the third largest volume has the most total influence, while the highest volume group has the lowest. Similarly, Figure 23(b) shows the influence functions of users grouped based on their total number of followers. Surprisingly, they find that the Twitter users with the intermediate number of followers have much higher influence than the highest in-degree nodes.

The results on Twitter nicely align with the experiments on Memetracker data. As the Memetracker experiments focused on online media and adoption of short textual phrases, they find that the mainstream media holds the most influential position in the dissemination of news content. On the other hand, hashtags on Twitter are a very different type of contagions. Hashtags are not news but rather socially contagious tags that are adopted in a

Fig. 23. Amount of influence of users grouped based on their total volume



(a) Average influence functions of each group based on volumes

(b) Average influence functions of each group based on follower counts

distributed manner without a central supervision. Therefore, the diffusion of hashtags is mostly governed by the Twitter social/information network.

This way Twitter users with "too high" number of followers, which usually correspond to celebrities and organizations, may be very influential in propagating the "information" contagions such as news, but not in diffusing more "social" contagions such as hashtags.

7 CONCLUSIONS

REFERENCES

[1] Z. Yang E. Ferrara. 2015. Quantifying the effect of sentiment on diffusion in social media.

[2] Adrien Guille and Hakim Hacid. 2012. A predictive model for the temporal dynamics of information diffusion in online social networks. In *WWW*.

[3] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking Sandy: Characterizing and Identifying Fake Images on Twitter during Hurricane Sandy. In *Conference Paper - May 2013*. https://doi.org/10.1145/2487788.2488033

[4] Fang Jin, Edward Dougherty, Parang Saraf, Yang Cao, and Naren Ramakrishnan. 2013. Epidemiological Modeling of News and Rumors on Twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis (SNAKDD '13)*. ACM, New York, NY, USA, Article 8, 9 pages. https://doi.org/10.1145/2501025.2501027

[5] S. Aral S. Vosoughi, D. Roy. [n. d.]. The spread of true and false news online.

[6] Kazumi Saito, Masahiro Kimura, Kouzou Ohara, and Hiroshi Motoda. 2010. Selecting Information Diffusion Models over Social Networks for Behavioral Analysis. In *ECML/PKDD*.

[7] Jaewon Yang and Jure Leskovec. [n. d.]. Modeling Information Diffusion in Implicit Networks.