Modeling and Visualizing Information Propagation in a Micro-blogging Platform

Chien-Tung Ho, Cheng-Te Li, and Shou-De Lin
Graduate Institute of Networking and Multimedia
National Taiwan University
R302, CSIE Building, No.1, Sec.4, Roosevolt Rd., Taipei 106, Taiwan (R.O.C.)
TEL: +886-2-33664888 Ext:302
ho_tongong@yahoo.com.tw, {d98944005, sdlin}@csie.ntu.edu.tw

Abstract—**Micro-blogging is a type of social networking service that has become ubiquitous in Web 2.0 era. Micro-blogs allows bloggers to exchange information, discuss ideas, and share experiences with friends or even strangers with similar interests. In this paper, we try to identify ways to measure how information is propagated in micro-blogs. More specifically, we consider the following issues. (1) How to quantify a person's capability to disseminate ideas via a micro-blog. (2) How to measure the extent of propagation of a concept in a micro-blog. (3) How to demonstrate and visualize information propagation in a micro-blog. We propose methods to effectively measure each user's ability to disseminate information via micro-blogs. The design of the measure considers three factors: (a) the number of people influenced; (b) the speed of propagation; and (c) the geographic distance of the propagation. We also provide an online demonstration micro-blog system that allows the users to explore the information propagation. The system shows the propagation paths and social graphs, influence scores, timelines, and geographical information among people for the user-given terms.**

Keywords-information propagation, microblogging

I. INTRODUCTION

A micro-blog differs from a traditional blog in that its messages are typically shorter in terms of their individual size and their aggregated size. The concept of micro-blogging becomes popular with the emergence of Tumblr [1] and Twitter [2] in 2006. Since then, several online micro-blogging services have been established, e.g., Plurk, Squeelr, Jaiku, Beeing, identi.ca, and Emote.in. These services generated tremendous amounts of micro-blog data, which has become an important resource for social network researchers. Although each micro-blog service has its own interface and functions, all such services share one important feature, namely, the embedded timeline (either explicit or implicit) that records the message posting/replying time stamp of each user. The update cycles of other online data sources, such as web pages and blogs, are measured in days, weeks or even months. In contrast, micro-blogs are updated every few minutes or even seconds, so they can be regarded as real-time services. Therefore, micro-blog data is a perfect resource for studying the dynamic nature of information, in particular how information is propagated and distributed in a large social network.

The idea of social-based information propagation can be traced back to the concept of "viral marketing" [3], which considers that, in the marketplace, ideas are spread by word of mouth. In his book "Media Virus: Hidden Agendas in Popular Culture", Douglas Rushkoff [4] argued that ideas and inference could be propagated even more powerfully via the Internet than by traditional channels. The argument is even more pertinent now, thanks to the emergence of social network services. Many companies and politicians now use Facebook or other social network services to promote themselves and increase their influence.

In recent years, a number of micro-blogging services have provided APIs that enable developers to extract the content of messages and obtain information about the social connections among users. As a result, some interesting analyses of micro-blogging have been published. For example, Sun et al. [5] used data provided by Facebook to determine the correlations between different communities; and Kwak et al. [6] used data from Twitter to construct the relationships among people for advanced analysis of social networks.

Most researchers have focused on the qualitative study or mathematical modeling of information propagation in social networks. By contrast, in this paper, our objective is to determine how to measure or quantify the propagation of information in micro-blogging platforms. We also design a real-world online demo system that allows users to analyze and visualize the propagation of certain concepts of interest. We consider the following issues.

1. How to quantify a person's ability to disseminate certain idea via a micro-blog.
2. How to measure the extent of a concept's propagation in a micro-blog.
3. How to visualize information propagation in a micro-blog.

The above issues are important because, from a research perspective, academics in different disciplines (e.g., sociology and mass communications) are interested in how information is spread among people, and are urgently in need of a platform or tool to perform experiments. Micro-blog services provide a perfect environment for such studies. Being able to measure and even visualize the level of information propagation in a social network could enhance social science research, e.g. by

978-0-7695-4375-8/11 $26.00 © 2011 IEEE
DOI 10.1109/ASONAM.2011.37

learning how information influences the evolution of a society or by studying how a burst of popularity can arise for a particular product. From the perspective of applications, knowing how to identify individuals with the ability to propagate certain ideas could benefit the recommendation of products in social networks. For example, companies could use influential people as seed candidates to make marketing and advertising campaigns more effective.

Our work contributes in three ways:

1. We propose an information propagation model to measure the ability of users to propagate ideas via micro-blogs. Our measure focuses on three factors: the number of people influenced, the speed of propagation, and the geographic range of propagation. We also define a loose criterion and a rigid criterion to bound the level of quantification.
2. We propose a method for measuring the level of spread of a query term in a micro-blog. The method allows us to quantify the influence of query terms in micro-blogs and rank their popularity.
3. We provide a visualization framework with an online search-based service that implements the proposed methods for demonstration purposes. Given a concept, our system finds the top micro-bloggers that have disseminated the concept in the network, and then displays different kinds of propagation values for users. To the best of our knowledge, this is the first online system that allows users to investigate and visualize information-propagation via micro-blogs. The system is available at http://mslab.csie.ntu.edu.tw/plurpagation/index.php.

The remainder of this paper is organized as follows. In Section 2, we present our methodology. In Section 3, we describe the demonstration system together with the case study with some snapshots of the visualization system applied on the Plurk micro-blog service. In Section 4, we review related works. Section 5 contains some concluding remarks.

II. METHODOLOGY

Ideas are shared and propagated frequently in micro-blogs. In this paper, a person who posts messages about a topic is called the *topic propagator* and people who receive the information are called *topic receivers*. If a receiver then disseminates the information to his friends, he becomes the next propagator and we say that the information has been propagated to 2^{nd} degree neighbors (usually denoted as L2 neighbors in the literature). For example, on Twitter, one of the most popular micro-blog services, the function "Retweet" can be regarded as information propagation tool because a blogger simply re-posts an original message from someone else's page to his own page. Facebook and Plurk provide similar functions. It is common for a person to post a friend's message on his own board in order to share it with the other friends. Based on the above discussion, we propose some simple yet intuitive methods to measure information propagation in micro-blogging platforms.

A. Model Overview

The system framework is shown in Figure 1. First, the user inputs the query term and the time period of interest. Given the input constraints, the system identifies a set of corresponding posts and their replies, as well as the time stamps on them. Then, based on the above information, two kinds of inference trees are constructed for each blogger: a rigid lower-bound influence (LBI) tree and a loose upper-bound influence (UBI) tree. For each tree, it is possible to produce three kinds of propagation values: (1) the number of individuals influenced by the propagated information, (2) the speed of the propagation, and (3) the geographic distance of the propagated information. After accumulating the propagation scores of all the users, we can determine the level of propagation of the query.

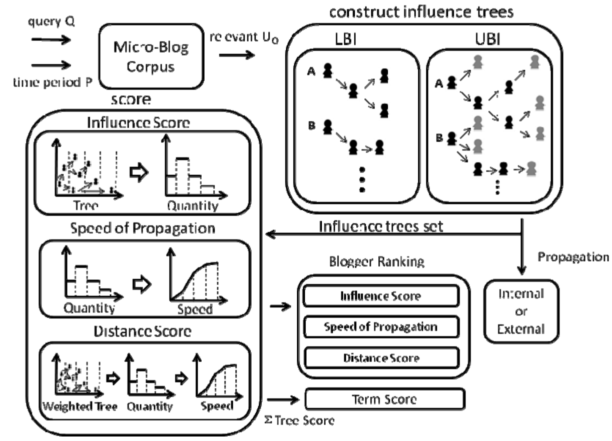

Figure 1. System Overview.

B. Tree Construction

For a given user query Q, we construct a set of influence trees to model the propagation among users in a micro-blog platform. Before discussing the tree construction, we define some key terms.

Definition 1. Micro-blog Corpus: A micro-blog corpus C is comprised of three kinds of entities: users, posts, replies of posts. The corpus C contains a set of users U. Each user has a set of posts and each post might contain a set of reply messages. All posts and replies are time stamped.

Definition 2. Relevant Micro-Bloggers: Given a query Q and a specific time period P, we define relevant micro-bloggers with respect to Q as a set of bloggers $UQ=\{u1,...,un\}$ that satisfy the following requirements: (1) ui has a set of posts $Ai=<ai,1, ai,2, ..., ai,m>$ $(k\geq1)$ containing the query term Q; and (2) the time stamp of each $ai,k \in Ai$ is within the specified period P. We also associate each $ui \in UQ$ with a time stamp of the first message that ui posts about Q.

To demonstrate the propagation of a concept Q from a user A to another user B in a micro-blog, we design two models: a rigid-propagation relationship model and a loose-propagation relationship model.

Rigid-propagation relationship model: If X posts a message that contains Q and an individual Y replies to the message and also posts another message relevant to Q, then we say the concept Q has been propagated from X to Y. Consequently, we assume there is a rigid-propagation relationship between X and Y. If two individuals X and Y satisfy the above conditions, we assume there is a rigid-propagation relationship between them. We consider that message propagation involves two actions: receiving and distributing. If Y replies to X, we consider that the message has been received by Y. If Y subsequently posts a similar message, then we say the message has been distributed by Y.

Loose-propagation relationship model: If X posts a message containing Q and Y replies to the message, we consider that Q has been propagated from X to Y; and we assume there is a loose-propagation relationship between X and Y. It is less rigid than the first criterion for propagation because it does not have to satisfy the 're-post' condition stated in the previous model. Here we still assume that a propagation event can be decomposed into receiving and distributing actions. In some micro-blog systems, people can view the replies to posts; therefore, Y's reply to X's post can be viewed by Y's friends. To a certain extent, this can be regarded as the propagating the information.

Based on above models, we define two kinds of influence tree for information propagation, as shown in Figure 2.

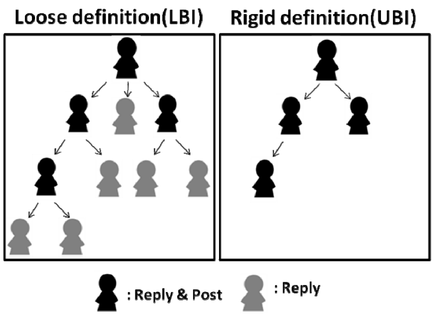

Figure 2: Two kinds of influence trees. The tree with the loose definition can be considered as the wider propagation.

Definition 3. UBI tree (Upper-bound Influence Tree): An upper-bound influence tree of a person u_x is a tree structure rooted at u_x. Each edge in the tree represents a loose-propagation relationship between the parent node and the child node.

Definition 4. LBI tree (Lower-bound Influence Tree): A lower-bound influence tree of a person u_x is a tree structure rooted at u_x. Each edge in the tree represents a rigid-propagation relationship between the parent node and the child node.

Each tree represents a propagation pattern from the root node. Clearly, the LBI tree is a subgraph of the UBI tree. We hypothesize that the true propagation pattern is bounded by these two trees. That is, the true propagation pattern of the root is a *subgraph of the UBI tree* and *a supergraph of the LBI tree*. The LBI and UBI trees can be constructed using BFS-like search method on the micro-blog data, as shown in Algorithm 1.

Algorithm 1. Influence Tree Set Construction

Input: a micro-blog corpus C; a query Q; a specified influence relationship LBI or UBI; a time period P.
Output: an influence tree set $F=\{T_1,...,T_n\}$.

Based on query Q and time period P, retrieve the relevant user U_Q from the corpus C.
for each $u_i \in U_Q$ **do**
  $T_i = T_i \cup u_i$.
  enqueue(Queue, u_i).
  while size(Queue) $\neq 0$ **do**
    x = dequeue(Queue).
    for each $y \in \{U_Q \setminus T_i\}$ **do**
      if x and y confirm influence relationship **do**
        $Ti = Ti \cup y$.
        Construct a link between x and y in T_i.
        enqueue(Queue, y).

C. Use Influence Trees to Quantify Propagation

To measure the degree of propagation, we consider three factors given an influence trees: (1) the number of people influenced, (2) the speed of propagation, and (3) the geographic distance of propagation.

Scale of Propagation. To quantify the scale of the propagation, we can simply count the total number of people (i.e., nodes) in the corresponding influence tree (Figure 3). The higher the number, the greater will be the scale of the propagation, as shown by the two-dimensional diagram in Figure 3. The horizontal axis represents the sequence of time stamps {t1,t2,…,tm}, and the vertical axis represents the number of people influenced during the specified time period. Note that each person should have two scores, one from the LBI tree and the other from the UBI tree. The LBI tree score is subject to tighter constraint; therefore, it represents the lower-bound value of the propagation. The UBI tree score represents the upper-bound value of the propagation.

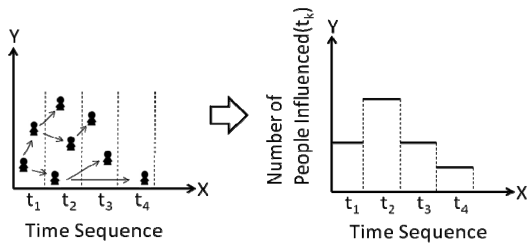

Figure 3. The distribution of people influenced during a specific period of time

Speed of Propagation. Besides the amount of propagation, we are interested in the speed of propagation. A person who is capable of affecting a large number of people in a short time is considered as a strong candidate for disseminating information. Based on the constructed influence trees and the time stamp of each entity, we propose a method for estimating the propagation speed of a message sent by the person at the root node. Figure 4 illustrates the above concept. The left-hand figure, which is similar to the one in Figure 3, captures the propagation status over time. We aggregate the number of individuals influenced over time to produce the graph on the right hand side of Figure 4. Then, we use the area under curve to represent the speed of propagation. Intuitively, the higher the propagation speed, the larger will be the number of people influenced in a short time. This would cause the curve to reach a higher level earlier and therefore increase the area under the curve. Similarly, each root node (person) should have LBI and UBI values to measure the speed of propagation.

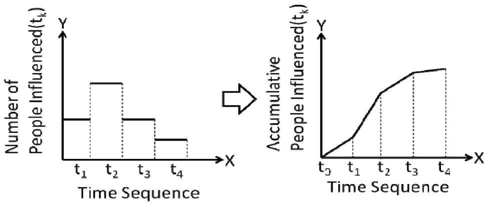

Figure 4. The distribution of accumulative people influenced during the time

Distance of Propagation, Geographic information enables us to observe whether a message is disseminated globally or locally. In this section, we propose a method for measuring the propagation in terms of the geographic distance. An intuitive way to determine the propagation distance is to calculate the total distance from the root user to the people he/she influences. This can be achieved by a micro-blogging service as long as it provides location information, such as city and country of the users. We can then pinpoint a location by its longitude and latitude coordinates and calculate the geographic distance between two geocodes. We use an x-y diagram, as shown in Figure 5, to illustrate the geographical spread. The vertical axis represents the total distance from the root user to the people influenced in a specific time period. We use the LBI and UBI scores to represent the constraints on the propagation measurement.

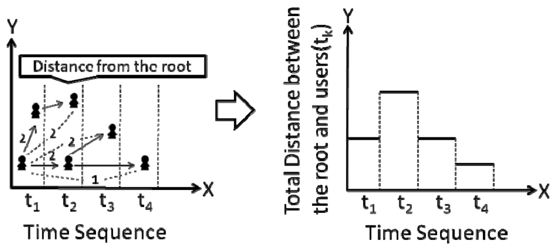

Figure 5: The distribution of total distance

D. Measuring the propagation of a query term

Clearly, the different kind of message can be propagated differently in a micro-blog. Some topics (e.g., information about a natural disaster) are likely to be propagated more rapidly than others. In the following, we propose a method for measuring the level of propagation of a concept or term in a micro-blog. First, we define the concept seed users.

Definition 5. (Seed Users): Given a query Q and a time period P, we define seed users as a set of users US={u1,...,us} that satisfies the following requirements: (1) u_i is a relevant user who posted a message about Q; and (2) u_i is not the *topic receiver* in a loose or rigid propagation relationship.

Seed users play the role of message sources, i.e., they initiate the propagation of messages. They are the topic providers in the micro-blog. To quantify the level propagation of a concept, we sum the propagation scores of all seed users (i.e., topic providers). Formula 1 below lists the three components of the measurement. Each component can contain an upper-bound value and a lower-bound value for a term. A concept is regarded as well-propagated if a large number of seed people share an idea with many other people, or propagate it over a long distance.

Formula 1. Six kinds of Term Scores

	LBI	UBI
Propagation Score	$\sum_{i=1 \cap i \in Seed}^{n} \mathrm{LBI}(\mathrm{PScore}_i)$	$\sum_{i=1 \cap i \in Seed}^{n} \mathrm{UBI}(\mathrm{PScore}_i)$
Propagation Speed	$\sum_{i=1 \cap i \in Seed}^{n} \mathrm{LBI}(\mathrm{PSpeed}_i)$	$\sum_{i=1 \cap i \in Seed}^{n} \mathrm{UBI}(\mathrm{PSpeed}_i)$
Distance Score	$\sum_{i=1 \cap i \in Seed}^{n} \mathrm{LBI}(\mathrm{DScore}_i)$	$\sum_{i=1 \cap i \in Seed}^{n} \mathrm{UBI}(\mathrm{DScore}_i)$

III. SYSTEM DEMO

There is no benchmark for measuring information propagation in a micro-blog service. Thus, instead of trying to prove the correctness of our model, we implement an online system to demonstrate its usage. We also propose some ways to display global and local information in the system.

A. Data Description

We implemented our system on the Plurk micro-blog system, a popular micro-log service in Asia. Unlike other social network services, Plurk implements a special feature called a dragable dynamic time-line to display messages. According to the analysis in Lai et al. [7], more than 5 million users formed a giant graph in Plurk in 2009. In Dec 2009, Plurk introduced an API that allows developers to search its content using any given query.

B. System Architecture

Our system serves as a search platform for users to observe and analyze the propagation of information in a micro-blog. The system can also be regarded as a visualization tool for information-propagation. In the system, we display the local propagation chart as well as some global statistics to facilitate further analysis.

There are three stages in the system, as shown in Figure 6. First, a crawler collects Plurk data using the provided API. Our system utilizes a topic term and the indicated time period to crawl the related posts and replies as well as the relevant users of Plurk. We crawl four kinds of content: 1) relevant posts; 2) the replies to each post; 3) the time stamps of the posts and the replies; and 4) the geographic information provided by the relevant users. Second, we use the crawled data to construct propagation trees for each relevant user. Then, we use the proposed measures to quantify the propagation capability of each relevant user. Finally, we display the information propagated by top-ranked users. We show the top five micro-bloggers for each measure. For each micro-blogger, users can click the icon to see the associated propagation paths and other relevant information, as will be shown later.

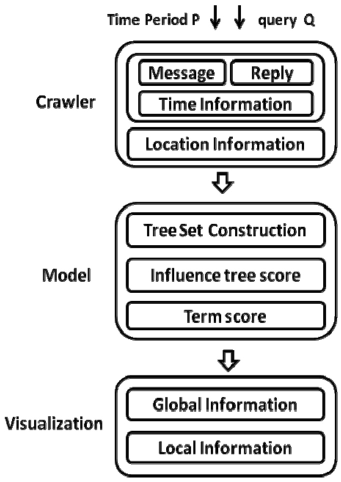

Figure 6. The system architecture.

Case Study

We use "FIFA 2010" as the topic (query term) to demonstrate how our system works. The demonstration results can be viewed at the following link: http://tinyurl.com/4kzsab9. FIFA2010 began on June 11, 2010. It was one of the most popular sports events in 2010. We also set the time period as June 11 to June 13, 2010. In the following, we demonstrate information propagation from the global and local perspectives.

(1) Global Information for Propagation

Given a query term (topic), the system displays four types of global information, as shown in Figure 9: (1) Histogram of Repliers, (2) Daily-post Distribution, (3) Web Content and (4) Influence Table. We describe each type in detail below.

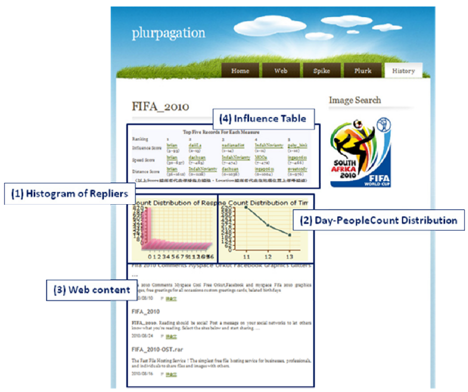

Figure 9: The system's global information page

Histogram of Repliers. We gather the statistical information about the number of people that replied to each message in the given time period, and then use an open source "open flash chart" [8] to display the distribution. Figure 10 shows our example of 'FIFA 2010'. The horizontal axis represents the number of people that replied to each message, and the vertical axis indicates the frequency of each message. In this example, we can conclude that the distribution follows a power law distribution.

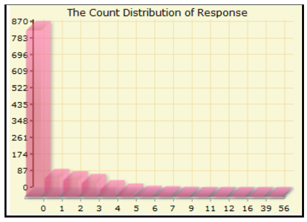

Figure 10: The visualization image of Reply-PeopleCount Distribution

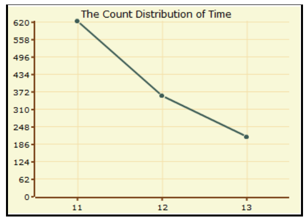

Figure 11: The visualization image of Day-PeopleCount Distribution

Daily-post Distribution. In the Daily-post Distribution diagram, we generate the statistics of the number of messages post with respect to time. The result are shown in Figure 11, where the horizontal axis represents the specified time period (e.g. June 11-13 in this example) and the vertical axis indicates the total number of messages posted about the topic each day.

Web Information. We exploit the Yahoo Search API for a given query, and display the top three related news items as well as the top image search results. The system can also serve as a search engine that provides some information about the term itself.

Influence Table. For each query term, we generate an influence table to show the top 5 users given each propagation measure (i.e. the number of people influenced, the speed of propagation, and the geographical distance). We use the LBI and UBI scores as the lower/upper bounds in Figure 12. The system takes the average score of the two bounds to rank the influence of each user, as shown in Figure 13. Each user has a hyperlink to display the local propagation information, which we discuss in the next section.

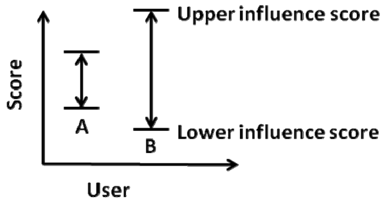

Figure 12: Boundary in the LBI and UBI methods

Top Five Records For Each Measure

Ranking	1	2	3	4	5
Influence Score	briian (5~93)	daiiiLa (2~13)	nadianadiot (1~14)	IndahNovianty (1~11)	peby_bink (1~10)
Speed Score	briian (30~637)	dachuan (7~485)	IndahNovianty (7~474)	MOOa (7~472)	inga90611 (7~466)
Distance Score	briian (56~1612)	IndahNovianty (0~1128)	dachuan (0~1036)	inga90611 (0~1004)	sweetcody (0~976)

(以上Score越高者代表傳播能力越強，Location越高者代表在地理位置上傳得越遠)

Figure 13: The influence table

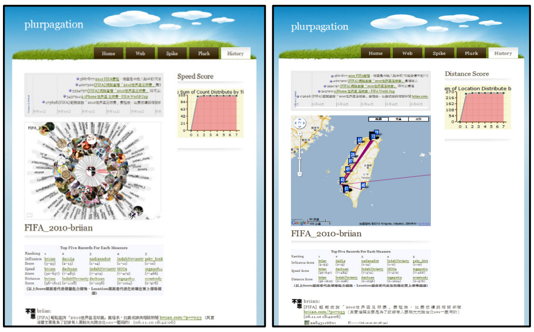

Figure 15: Two pages of local information

(2) Local Information for Propagation

For each top-ranked micro-blogger, our system provides five kinds of local information: (1) a timeline, which displays the content of the messages and replies as well as the temporal relationship between them; (2) the structure of the influence tree as a visualization of the propagation; (3) a map to show the geographic propagation; (4) an aggregated spread count over time to indicate the speed of propagation; and (5) the content of the messages and replies in table format. Our system displays two kinds of pages, as shown in Figure 15. Here we use the top blogger "briian" for demonstration.

Timeline. We use the tool "Simile Widgets" [9] to visualize information about the timeline of the related messages, which is root from "briian", as shown in Figure 16 (http://tinyurl.com/4o2ggfn). The timeline shows how quickly a message is spread within given time period.

Figure 16: The timeline of influence tree with "briian" as root

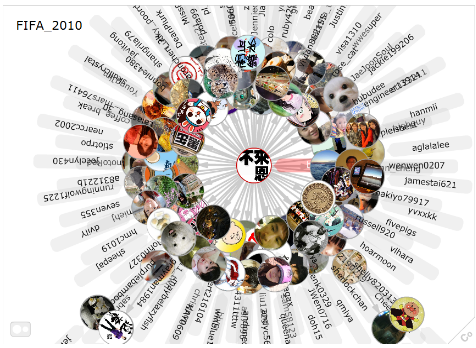

Figure 17. The influence tree with "briian" as the root

Graph Gear. The UBI influence tree rooted in "briian" is shown in Figure 17. We use the open source tool "GraphGear" [10] to visualize the tree. The images in the circles are the pictures of the bloggers in Plurk. The red circle indicates the root user. By following the directions of the edges, it is possible to identify the paths of information propagation. The display allows the users to determine whether the high influence of a person is caused by the long propagation paths, or simply the consequence of its high degree.

Dynamic Map Marking for Geographical Information. We use Google Map to mark the geographic propagation location (i.e., the latitude and longitude coordinates of each user (node) in order to visualize the geographic influence of a UBI influence tree. The tree rooted in "briian" is shown in Figure 18 (http://tinyurl.com/4h5pa43). The marks on the map appear in order of the posting time. Through the links, we can visualize the geographic propagation path. This feature also allows us to determine whether the query topic is propagated locally or globally.

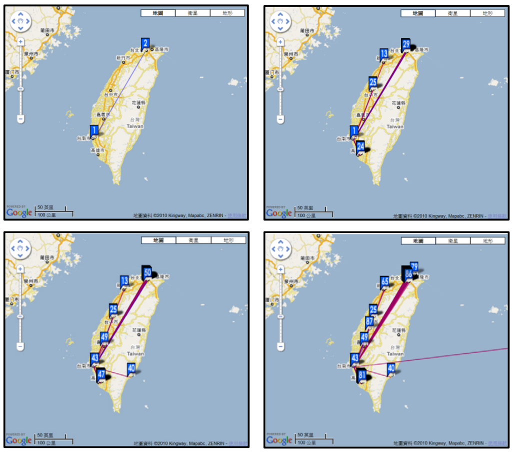

Figure 18. Dynamic visualization of propagation on Google Map

Aggregated spread count over time. The aggregated spread count chart represents an accumulation of propagated person-count or total distance over time in an influenced tree, which is similar to the chart in Figure 4. The feature allows users to determine the propagation speed of the root user.

Message. Finally the system displays the content of message posts and replies, as shown in Figure 19. Here, we find that the bloggers "Cheese0831" and "andonedemon" have posted similar information.

Figure 19. The system displays the content of message posts and replies

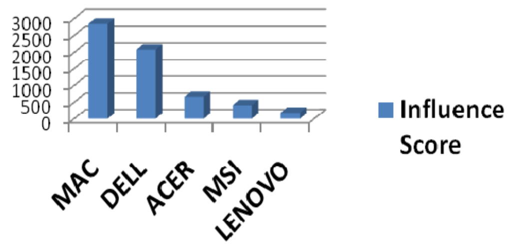

Figure 20: Term Scores based on the five topics

C. Rank on Term Score

In this section, we evaluate how the results of our model compare to those generated by humans in terms of ranking the level of propagation of terms. We choose five well-known brands of 3C products, namely MAC, DELL, ACER, MSI and LENOVO, as query topics to generate term scores. The results are shown in Figure 20. In Plurk, MAC has the highest level of propagation followed by Dell. We then find 5 humans to rank these five products based on how they believe the level of propagation of each idea in general. The results show that the Kendall tau rank correlation among our system's results and the human results is 0.60, which is considered fairly significant in statistics. The results show that our measures match the impressions of the human participants to a certain extent.

IV. Related Work

A. Model-based Information Diffusion

Richardson and Domingos [11] proposed a probabilistic method for extracting information from a knowledge-sharing network and put forward a hypothesis about the most effective individuals for viral marketing. Subsequently, Kempe et al. [12] proposed a model to maximize the influence of a social network. First, they showed that finding the most influential people is an NP-hard problem. Then, they proposed two models, the Linear Threshold Model and the Independent Cascade Model, and used them to simulate information propagation in a social network. Meanwhile, Gruhl et al. [13] developed a propagation model for blogs based on the theory of infectious diseases; while Song et al. [14] proposed a method to predict the target flow of information and how long it takes for a user to obtain new information. The major difference between the above works and our approach is that they focus on developing models to simulate, predict, or explain information propagation, rather than measuring and quantifying information propagation in a real-world micro-blog environment.

B. Information Propagation on Real Data

With the increasing availability of social network data in recent years, researchers have applied different models to analyze the data. Cha et al. [15] exploited Flickr data to construct the relationships between photos and the photographers. They also tried to determine how widely information can be spread and what role word of mouth plays in such a network. Sun et al. [5] investigated the propagation phenomenon of Facebook's News Feed, and created a social network based on users and fans of Facebook for analysis. Their objective was to observe the relationships between different kinds of propagation communities and determine if several short diffusion chains tend to merge together. They exploited zero-inflated negative binomial regressions to model the phenomenon. Kwak et al. [6] used the relationships between the follower and following in Twitter to construct a social network for advanced analysis. Their results indicate that there exists a gap in influence inferred from the number of followers and that from the popularity of one's tweets. Goyal et al. [16] propose a parameterized model to learn and predict the time required by a user to perform an action. They used Flickr data to verify the model's accuracy. Sakaki et al. [17] proposed a real-time event detection system using Twitter data. They regard Twitter users as event sensors and use the messages posted by the sensor users to train the features set by their model and analyze the location and temporal information for some events. The authors apply their system to detect earthquakes and typhoons. The above works use micro-blog data for certain kinds of analysis, but they do not consider the issues related to measuring the scale and speed of propagation. Furthermore, to the best of our knowledge, there is no other free-access, online system that allows users to visualize and analyze the propagation of information.

V. Conclusion

Some people believe that online advertising could provide a profitable business model for social network services. Thus, being able to quantify and measure the propagation of information would facilitate the expansion of online advertising services in social networks. In this work, we propose a model for estimating information propagation. Instead of producing an exact quantification score, our model provides upper-bound and lower-bound values using the UBI and LBI tree scores. Besides proposing a simple yet intuitive way to measure information propagation by counting the number of people influenced, we present novel ways to measure the propagation speed and the geographical distances. We have also implemented an online system, called Plurpagation, trying to visualize information propagation in Plurk. The system displays global information about concept propagation as well as local, dynamic information that allows users to gain more insights into propagation patterns. Our system and model are for general purposes, so can easily be applied to other micro-blog services, such as Twitter. We believe that our system will provide researchers in other areas, such as social science, with an alternative way to collect data and conduct social-network research.

In the future, we will exploit certain NLP techniques, such as opinion analysis, to analyze the content of messages and provide a deeper analysis of the propagation. Moreover, the current system is not very efficient because the query API is slow. To resolve the problem, we will investigate cloud-computing based approaches to collect information through the MapReduce framework in order to speed up the process.

References

[1] Tumblr.com. Micro-blogging web site: http://www.tumblr.com/

[2] Twitter.com. Micro-blogging web site: http://twitter.com/

[3] "Viral Marketing": http://en.wikipedia.org/wiki/Viral_marketing

[4] Rushkoff, D.(1994) Media virus! : hidden agendas in popular culture. Ballantine Books

[5] Sun, E., Rosenn, I., Marlow, C. and Lento, T. 2009. Gesundheit! Modeling Contagion through Facebook News Feed. In Proc. of the 3rd Int'l AAAI Conf. on Weblogs and Social Media

[6] Kwak, H., Lee, C., Park, H., and Moon, S. 2010. What is Twitter, a social network or a news media?. In Proceedings of the 19th international Conference on World Wide Web. WWW '10.

[7] Hung-Che Lai, Chien-Wen Chen, P.S. Liu, and Shou-De Lin 2009. Exploiting Cloud Computing for Social Network Analysis – Exemplified in Plurk Network Analysis, TAAI '09.

[8] Open Flash Chart; http://teethgrinder.co.uk/open-flash-chart/

[9] Simile-widgets; http://www.simile-widgets.org/timeline/

[10] GraphGear; http://www.creativesynthesis.net/recycling/graphgeardemo/

[11] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In Proc. of the 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining

[12] D. Kempe, J. M. Kleinberg, and ´E. Tardos. Maximizing the spread of influence through a social network. In Proc. of the 9th ACM Int. Conf. on Knowledge Discovery and Data Mining

[13] Gruhl, D., Guha, R., Liben-Nowell, D., and Tomkins, A. Information diffusion through blogspace. In Proceedings of the 13th international Conference on World Wide Web. WWW '04.

[14] Xiaodan Song, Yun Chi, Koji Hino, Belle L. Tseng: Information flow modeling based on diffusion rate for prediction and ranking. WWW '07

[15] Cha, M., Mislove, A., and Gummadi, K. P. 2009. A measurement-driven analysis of information propagation in the flickr social network. In Proceedings of the 18th international Conference on World Wide Web. WWW '09.

[16] Goyal, A., Bonchi, F., and Lakshmanan, L. V. 2010. Learning influence probabilities in social networks. In Proceedings of the Third ACM international Conference on Web Search and Data Mining. WSDM '10. ACM

[17] Sakaki, T., Okazaki, M., and Matsuo, Y. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international Conference on World Wide Web. WWW '10. ACM

[18] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. S. Glance. Cost-effective outbreak detection in networks. In Proc. of the 13th ACM Int. Conf. on Knowledge Discovery and Data Mining KDD'07.