Twitter Trendfinder: Visualization and analysis of Twitter trends

Milad Gholami milad621@gmail.com
Peixin Gao gaopei@umd.edu
Chris Imbriano imbriano@umd.edu
Mahfuza Sharmin msharmin@umiacs.umd.edu
Jin Sun jinsun@cs.umd.edu

Abstract

As the rapid growing of social networks, people start to realize the influence of the social networks like Twitter in terms of information diffusion over the networks, and more and more research on social network analysis has been towards the information diffusion models and influence patterns of social networks, which helps the growth of social networks and the implementation of commercial models over social networks. Previous work on influence pattern in social networks was mainly focusing on building mathematical models using tons of training data. However, many details and key characters of the information diffusion process may be eliminated via the statistical approach of analysis. In this project, we investigate and analyze the influence pattern behavior via visualization of tweet-retweet interaction and evolution of different trends in Twitter, in order to find out valuable informational details and undiscovered insights, meanwhile verify the results reached by previous work using statistical approach, analysis, and find valuable insights from the visualization.

Background

In literature there has been a lot of study under the basic social network analysis on Twitter. Among these research topics on Twitter developed recently, there are several topics that are still open to research.

Social Network Analysis: Users and Tweets

In [1] the authors analyzed characteristics of users, messages communicated among them, and networking behaviors or activities. Some of their interesting findings are, high buzz is associated with large number of retweets, retweets are correlated with follower counts etc. In [2], the authors proposed a graph based method to find a user who propagates retweets of matching interest. And they visualized that graph (overlay graph) to check the frequency and content of tweets for validation purpose. Group behavior and growth patterns based on twitter usage has
been done in [3].

The characteristics of tweets and their flowing pattern and dissemination has also been a concern for study. In [7] authors applied machine learning (PCA and GLM) to analyze the features of tweets. In another study temporal changes of retweets has been conducted [8]. Another large scale study on deleted tweets has been conducted in [10] to answer the connection of those tweets with users, comments, geotagging information etc. In [9] authors conducted an interesting study analyzing how twitter users can differ from average users with respect to movie ratings and trying to answer whether or not Twitter sentiment is a good predictor for box office (movie) success.

Influential Accounts within Social Networks

Hao et al [12] proposed method to discover influential users by modeling user relations as graph and using community detection algorithm to classify users with regard to their influence. [13] ranks users on twitter based on their information amplification potential. The potential is defined with two factors: one indicates the tendency of a user to be retweeted or mentioned, second is proportional to the size of the audience, i.e. the number of followers. The ranking algorithm is similar to that of PageRank.

Li et al [14] address this problem from game theory approach. They defined standards of influential user by set of parameters such as message quality, message generation rate, following tactics and following rate. This paper reveals the strategy to become an influential user. Game theory model is used to simulate the evolution of the network and spread of the influence. Having also built a graph to represent user interactions in social network, [15] takes a different perspective in modeling the information flow and identifies influential users using random walks. The intuition is that influential mobile users may be visited more frequently by random walks initialized from different small group of users.

Trend Detection and Monitoring

Trends are topics that involve a significant amount of users of discussion. One interesting question is when does one topic become a trend, and how? Mathioudakis et al [11] proposed method to detect trend over twitter stream in real time. Trend is found by detecting and grouping ‘bursty’ keywords. This is closely related to our task that to retrieve the moment when a topic becoming a trend over time.

Social Media vs Traditional Media

Social networks, especially ones like Twitter, not only connect people together, but also play as media to disseminate information like news and messages, for which they are sometimes also called social media. In this project we also want to investigate the role of Twitter as a social
media and its difference between traditional media like newspaper and television, according to the information diffusion models that have been developed and the results that we will find via our experiments.

H Kwak et al conducted a quantitative study on the entire Twittersphere and information diffusion on it, where three ranking systems are used to identify influential users. They found a non-power-law follower distribution, a short effective diameter, and low reciprocity, which all mark a deviation from known characteristics of human social networks.

There is also a comparison between social media and traditional media made by A Hausman. And the social media is labeled as two-way conversation, open system, transparent, one-on-one marketing, about "you", brand and user-generated content, free platform, actors: users/ influencers, community decision-making, unstructured communication, real time creation, bottom-up strategy, informal language and active involvement. On the contrary, traditional media is remarked as one-way conversation, closed system, about "me", polished content, paid platform, actors: celebrities, economic decision-making, controlled communication, pre-produced/ scheduled, top-down strategy, formal language and passive involvement. These differences make the two parties have different perspectives and advantages over different types or information.

Problem Description and Motivation

In this project, we want to build up a visualization tool to visualize the tweet-retweet pattern of different trends in Twitter, and use the tool to explore in discovering the user(s) who are responsible for initiating the trend. This are useful to identify influential persons on a certain topic, which topics spread fast etc. We think this is useful because current research on the topic of information distribution and influence diffusion is mainly about global analysis which may fail in discovering a lot of detail information. People will not be able to find the evolution patterns of a specific trend from macroscope analysis. In our project, we want to find out details and insights that hasn’t been discovered in previous work. Information and insights from the visualization can be useful in many cases, for example advertising in social networks and analysis on the relation between information diffusion patterns and different topics.

Data Processing

We collected original tweets data using Twitter Stream API via a Ruby wrapper. The experiment is done with a dataset of 100 tweets about topic “#ObamaCare”. The raw tweets data is further converted to json format for visualization purpose.

The original tweet data has many fields and we select a subset of fields:

dict{'tweets': [{'id': 1, 'text': 'RT @LadyVeteran23: ObamaCare: Illegal Voter Mine for Dem Party? Bill specifies low income ppl R auto registered to vote.'}]}
'id' is the identifier we assigned for each tweet. 'text' contains the content of the tweet. 'created_at' shows the time this tweet was created. 'retweeted_id' is the id of the tweet being retweeted by this tweet. This field is critical since we need it to trace back to the very tweet which started the current trend.

After import the json file to our D3 visualization engine, we are able to draw the graph such that each node is an individual tweet and each link represents the relation between tweets (retweeted/reply).

**System Overview**

Twitter Trendfinder visualization follows the Visual Information Seeking Mantra: overview first, zoom and filter, then details-on-demand. First by default we will show the overview picture of tweets' occurrence over time. Then the slider control enables user to zoom in a specific time period for closer investigation. Finally, by clicking on the tweet node, detailed information is presented in the information panel.

Tweets are represented as nodes in trees. Here nodes include independent tweet, reply and retweet. We show edges from one tweet to another for following two cases: a) if tweet 2 is retweeted from tweet 1, there is an edge from tweet 1 to tweet 2. b) if tweet 2 is a reply of tweet 1, we add an edge from tweet 1 to tweet 2. The tree is drawn using force layout. The x-coordinate of the tweet is set by the time it was created. The y-coordinate is automatically calculated by algorithm used in force layout.

In the tree view we have added some user interactions, for example, user can select nodes. On selection of a node (representative of a tweet/reply/retweet) user can see the details information related to that tweet. For example, what is the tweet about, exactly when the tweet has been made, whether this is a reply or a retweet etc. Also if the user mouseovers on the node some basic information regarding the tweet is shown. Additionally to see the trend more clearly we highlight the whole subtree under a node. [Fig-1]
If there are lots of tweets on a specific topic the visualization looks cluttered. For better arrangement of the nodes we implemented force layout for the visualization. Even with the force layout the graph can be difficult to inspect with large number of tweets. That’s why we have added another component, range slider. Range slider helps to see the dense graph closely. As we move the slider towards left (right) the nodes/tweets in the left (right) becomes invisible.

For our graph visualization we add a control panel for filtering tweets. There are two types of filtering in our implementation. We can filter in/out tweets which are independent tweets (no one retweeted and no one replied) and/or tweets which are either retweeted or replied by other tweets. The 1st group is called ‘Singleton’. So, in our control panel we have option either to show or hide ‘Singleton’. The other filter option is to show which tweets are reply and which are retweets.

**Evaluation**

The purpose of our evaluation is to gauge the usability and ease of navigation of the Twitter Trendfinder web site for specific target audiences, namely researchers at Universities, IT company engineers, and the general public.

During December 2013, we will test the Twitter Trendfinder web site with 10-15 individuals pulled from the defined audience groups. Our goals is to determine what is or is not working successfully on the new web site from the users’ perspective.
After a brief introduction and explanation of the tool, we allowed the participants to have a freeform exploration of the data until they are ready to take the test. We first ask them to finish seven tasks. For each task, they have 2 minutes to finish it. We will employ a task-based think-aloud protocol, in which we will ask users to communicate their thought processes verbally while they work. We will ask them to vocalize what path they take to find information, what questions they have, and what surprises or confuses them as they go through the application. Then we check the correctness of the answers, in order to determine the usability of our design.

After finishing all the tasks, the participants will be asked to fill in a feedback form of six questions. We also encourage them to express any comment that they might have. Test results, their comments and our observations (mistakes they made, unreasonable learning curves, bugs, confusing interface elements, missing items, etc.) will be documented in handwritten and typed notes.

The tasks we designed for the usability test are:

- Given the a certain date, find the tweets about the trend and what’s the amount?
- On what date the trend started to blow up?
- What’s the peak (maximum amount of tweets per day) about the topics
- Who is the person that started the topic?
- After you click a certain tweet, find its children, tell if they are retweets or replies?
- Find all the singletons on a certain date.

And we also have the follow-up questions asking about their impression of the website, followed by an open-ended general discussion period where participants can share their thoughts on any aspect of the web site or testing with us.

**Conclusion and Future Directions**

Based on the feedback from the usability test, we will try to improve the usability via modifying the design. Meanwhile, we will add more functions of visualization into the webpage to offer more information to the users.

**Reference**


