Social Influence Analysis

Avneesh Singhal, Prakhar Srivastava, Shantam Wadhwa

Abstract –In large social networks, users are influenced by others for various reasons. For example, the colleagues have strong influence on one's work, while the friends have strong influence on one's daily life. Our approach is to differentiate the social influences from different angles (topics) or small-scale networks and to measure the strength of those social influences. In this work, we focus on coming up with an approach for measuring the strength of social influence with a generalized outlook towards the existing social networks.

INTRODUCTION

Social influence occurs when a person's emotions, opinions, or behaviors are affected by others. Social influence takes many forms and can be seen in conformity, socialization, peer pressure, obedience, leadership, persuasion, sales, and marketing. In 1958, Harvard psychologist Herbert Kelman identified three broad varieties of social influence. Social influence is the change in behavior that one person causes in another, intentionally or unintentionally, as a result of the way the changed person perceives themselves in relationship to the influencer, other people and society in general.

Three areas of social influence are conformity, compliance and obedience. Conformity is changing how you behave to be more like others. This plays to belonging and esteem needs as we seek the approval and friendship of others. Conformity can run very deep, as we will even change our beliefs and values to be like those of our peers and admired superiors.

Compliance is where a person does something that they are asked to do by another. They may choose to comply or not to comply, although the thoughts of social reward and punishment may lead them to compliance when they really do not want to comply.

Obedience is different from compliance in that it is obeying an order from someone that you accept as an authority figure. In compliance, you have some choice. In obedience, you believe that you do not have a choice. Many military officers and commercial managers are interested only in obedience.

A. A. Integrated Summary of Referred Papers

PAPER 1:

Through the paper, the author discusses in intricate detail the meaning of Social Influence by a countless number of contents which gets posted on the YouTube everyday. YouTube keeps its competitiveness by maximizing click-through-rate with their cutting-edge video recommendation algorithms. This enables about 20 recommended videos for each video on YouTube.

Each video can be thought as a node and two videos are connected by an edge if one recommends the other, i.e. it creates a large Directed graph. This Paper is all about the characteristics of YouTube videos using diverse techniques such as PageRank, modularity and statistical estimation. By running these analysis algorithm, it is possible to observe a pattern for a popular videos for different categories.

There is a comparison between PageRank and Video feature correlation Analysis. This part analyzed the PageRank score of about 750,000 total Youtube videos by representing each video as a node with a PageRank score. The edges between nodes were formed to represent the relatedness of the two videos. That is, when video A is one of the 20 related videos of video B, then there exists a directed edge from B to A. Edges were not weighted. Using the PageRank score, this project evaluated the influence degree of each video. Thus, if a video has a high PageRank score, that means that video is related to many videos in the graph network, thus have a high influence within the network because it is pointed by many videos in the network.

This mainly examines the correlation between PageRank and multiple different statistical measures for a particular video. Specifically, we looked at the correlation of PageRank with the following six features:

- age (how long ago was the video posted on Youtube)
- number of views (how many people viewed the video)
- number of comments (how many people posted comments for the video)
- length of the video
- rate (the number of likes and dislikes for the video)
- Ratings (the rating scored by viewers out of 5.0 rating system) The purpose of such analysis is to examine what features contribute to what degree in

determining the PageRank score, or the influence level, of the video with in the Youtube network. Then various edge probability Estimation Models are used.

PAPER 2:

This paper introduces a novel, simple and accurate model to continuously rank influential Twitter users in real-time. Their model is based on the information amplification potential of a user, the capacity of the user to increase the audience of a tweet or another username that they find interesting, by retweets or mentions. They incorporate the information amplification using two factors, the first of which indicates the tendency of a user to be retweeted or mentioned, and the second of which is proportional to the size of the audience of the retweets or mentions.

They distinguish between cumulative influence acquired by a user over time, and an important tweet made by an otherwise not-important user, which deserves attention instantaneously, and devise our ranking scheme based on both notions of influence. They show that their methods produce rankings similar to PageRank, which is the basis for several other successful rankings of Twitter users. However, as opposed to PageRank-like algorithms, which take non-trivial time to converge, our method produces rankings in near-real time. Their ranking marginally outperformed PageRank, with 80% of the Top 5 most important users being classified as relevant to the event, whereas, PageRank had 60% of the Top 5 users marked as relevant. However, PageRank produces slightly better rankings, which correlates better with the user-produced rankings, when considering users beyond the top 5.

Many companies are using Twitter to seek information chatter on Twitter relating to their business and also their competitors, showing the immense potential of the platform to retrieve information. Twitter's timeline has been used in the study of event detection and proved to be great diffusor of headline news . The traditional journalism model is changing rapidly as new social media technologies like Twitter appear . Thus, in order to be better informed and receive breaking news about specific topics, it is possible to make an analogy between the action of buying a particular magazine or a newspaper in the traditional journalism model and the action of following a user of your interest on Twitter. Therefore, it is necessary to have a tool that can optimize the search of whom to follow amongst the 140 million currently active accounts, and that can point the best ones which will be relevant to each event or topic.

This work presents a simple and accurate ranking scheme called Information Amplification Rank or *IARank*, which can be used in real-time operations, to rank users as an event unfolds. This rank is based on the notion that the highly influential users in Twitter are those whose usernames are being currently amplified by the Twitter network, via mentions, replies or retweets by other users. Information amplification is measured by two factors. The first factor, which we call *Buzz*, measures the attention the user receives from other users in the network. The second factor, which we term *Structural Advantage*, measures whether the local network structure around a user is better suited to provide information to the network (i.e., the user's tweets are easier to amplify), or seek information from the network (i.e., the username has less potential to spread in the network).

PAPER 3:

Through the paper, the author discusses in intricate detail the meaning of Social Influence in the modern community, and its effects on the social life of the people connected. The author also sets a base for Social Network related approach in the form of different parameters of a graph, for ex. clustering coefficient, in-out degree of nodes etc. and how the measure of these parameters can be used to interpret the behavior of nodes(or users) when put in specific scenarios, for ex. how likely it is that a node befriends or follows someone who is not related to the node's direct circles.

It also differentiates how change of focus from quantity(no. of nodes related) to quality(closeness of the related nodes), could bring unprecedented changes in a user's decisions for what he does on the internet.

PAPER 4:

The existing methods mainly focus on qualitative identification of the existence of influence, but do not provide a quantitative measure of the influential strength.

The approach used in the paper focuses on Topic-based learning which refers to prediction of expertise on basis of the follower network of a person in context to the topic being discussed. Two approaches are used in this paper, Basic TAP, and Distributed TAP. Distributed TAP is based on basic TAP only, but is used for parallel processing of large datasets using multiple systems.

PAPER 5:

The ways to measure the strength of social ties people have cannot yet be broken into solid parameters. This paper proposes a model, focusing on variables which can be used to check the strength of mutual bonds among people, for example broader classification matrix, and autocorrelation improvement. Datasets Examined are undirected (FB and LinkedIn) to redeem maximum possible connections.

PAPER 6:

In this paper, there is an analysis of social network in order to understand the interplay between similarity and social ties. People are similar to their neighbors in a social network for two distinct reasons : Firstly, they grow to resemble their current friends due to social influence. Secondly, they tend to form new links to others who are already like them, this is selection by sociologists. Social Influence can push systems towards uniformity of behavior while Selection can lead to fragmentation. So, they are developing techniques for interaction between Social Influence and Selection using data from online communities where both Social Interaction and changes in behavior over time can be measured.

According to the authors , it is important to isolate the respective effect of Social Influence and Selection as, Social Influence produce network-wide uniformity i.e new behavior spreads along the links. Example- Viral marketing. Whereas Selection tends to drive the network towards smaller clusters of like-minded people. Example- Recommender systems. Developed a framework using data from Wikipedia. Basically, they focused on two central questions: First, Can we characterize how Social Interaction affects interests and vice-versa? Second, Can we characterize the relative degree to which interests and Social Interactions affect what people do?

Wikipedia focuses on two types of recorded actions i.e. editing articles and editing the discussion page associated with particular user. Editing articles are the indicators of interest as each article is on a particular topic whereas Editing the discussion page is the indicator of social ties between the users who are communicating. Considering the pairs of Wikipedia editors who have communicated with one another- after aggregating over all such pairs , they founded that there is a sharp increase in the similarities between two editors before they interact. Therefore, similarity leads to interaction but not vice-versa. Then they compared their model with Holme-Newman Model which is too simple to produce the effects we see in Wikipedia. In order to measure how Selection and Social influence operate , they tracked a time-evolving vector representing each person's activities and they also studied how the vector of two people are changing directly around the moment when they first interact . Mathematical expressions used: cosine metric and Jaccard similarity.

PAPER 7:

There has been a long standing goal of social analysts to characterize the relationship that exists between person's social group and his/her personal behavior. In this paper, the author applied various data mining techniques to study this relation of over 10 million people, by turning to online sources of data. The people who chat with each other are more likely to share interests according to the analysis.

For the characteristics of each person in a network, they turned to two sources: First, is the demographic data of instant messaging(IM) users such as person's age, gender and geographic location. Second is based on personal behavior, for this they turned to web search behavior. For obtaining the relationship between users, conditional probability is used. Two datasets have been used in this paper- MSN messenger instant messaging network and keyword searches made by various users on Microsoft Web search engine along with the information about user's personal characteristics such as zip codes, age and gender. Then there are single-valued attributes and multi-valued attributes. Then this relationship is shown by various histograms.

Then the conditioning is done on personal attributes. The results shows that people who talk to each other on the messenger network are more likely to be similar than a random pair of users, where similarity is measured in terms of matching on attributes such as queries issued, query categories, age, zip and gender.

PAPER 8:

Conformity is a type of social influence involving a change in opinion or behavior in order to fit within a group. This paper is about how the effect of conformity plays a role in changing user's online behavior. Then a Confluence model is proposed in this paper to formalize the effects of social conformity into a probabilistic model.

Confluence can distinguish and quantify the effects of the different types of conformities. There are certain experimental results on four different types of large social networks, i.e., Flicker, Gowalla, Weibo and Co-Author, verify the existence of the conformity phenomena. Action prediction accuracy (AUC) of different methods by considering the effect of conformity in four networks are also given through histograms. The goal is to study how a user's behavior conforms to her peer friends and the communities (groups) that she belongs to.

Three levels of conformities respectively from the aspect of individual, peer relationship, and group are also defined. Then Conformity-aware Factor Graph Model is explained in detail. They simply considered four attribute features, i.e., the number of friends, the number of new friends in the recent three time stamps, the number of total groups that the user joined, and the number of groups the user joined in recent three time stamps. Then there is an algorithm about Distributed Model Learning in order to handle large networks. On all the four data sets, they used the historic users' actions as the training data in different methods and used the learned model to predict users' action in the next time stamp- Prediction performance analysis.

PAPER 9:

In this paper, author characterize the propagation of URLs in the social network of Twitter, a popular microblogging site. We track 15 million URLs exchanged among 2.7 million users over a 300 hour period. Data analysis uncovers several statistical regularities in the user activity, the social graph, the structure of the URL cascades and the communication dynamics. Based on these results we propose a propagation model that predicts which users are likely to mention which URLs. The model correctly accounts for more than half of the URL mentions in our data set, while maintaining a false positive rate lower than 15%.

Microblogging is a relatively new phenomenon in the Web 2.0 world of user generated content. Twitter is one of the most popular microblogging sites today. The goal of the author is to focus on characterizing and modeling the information cascades formed by the individual URL mentions in the Twitter follower graph. The information cascades have been studied before in other Web 2.0 systems, such as Flickr , blogs , Digg and YouTube . Most of the prior work builds information propagation models to either reproduce cascades with statistical properties matching the empirical observations or to predict how far the information will diffuse in the network given its initial spread. They address a different problem: predicting *which users* will tweet *which URLs* given a training set of existing URL mentions.

Accurate prediction of URL mentions is an important enabler of a number of possible applications. First, knowing the tweeting probabilities for each user and URL can be used to generate a ranked list of URLs for each user providing a *personalized recommendation* of URLs that the user is likely to find interesting. For users who follow many other users, this method can be used to prevent information overload by *ranking and filtering* the incoming tweets. Second, aggregating the probabilities per URL quantifies the URL's future potential to diffuse in the social network. This could serve as method for early identification of viral URLs. Third, having an accurate propagation model trained for a specific social network can also help *viral marketing campaigns* to select URL injection points that would maximize the spread of the campaign URL in the network. Finally, a model predicting URL diffusion is not only useful when its predictions are correct, but also when the new data does not match them. Data acquisition : Tweets, URLs, User graphs. Cascade properties: RT-cascades vs F-cascades, Number of subcascades, Subcascade size, Subcascades are shallow.

References

[1] A SURVEY OF MODELS AND ALGORITHMS FOR SOCIAL INFLUENCE ANALYSIS Jimeng Sun Jie Tang

[2] Social Influence Analysis in Large-scale Networks Jie Tang Jimeng Sun Chi Wang and Zi Yang

[3] Modeling Relationship Strength in Online Social Networks .Rongjing Xiang

[4] Jennifer Neville Monica Rogati

[5] Feedback Effects between Similarity and Social Influence in Online Communities David Crandall Dan Cosley Daniel Huttenlocher Jon Kleinberg Siddharth Suri

[6] Yes, There is a Correlation - From Social Networks to Personal Behavior on the Web Parag Singla Matthew Richardson

[7] Confluence: Conformity Influence in Large Social Networks Jie Tang Sen Wu Jimeng

[8] A Survey on Influence Analysis in Social Networks Jessie Yin