On the Impact of Social Cost in Opinion Dynamics

Panagiotis Liakos and Katia Papakonstantinopoulou∗
University of Athens, GR15703, Athens, Greece
Email: {p.liakos, katia}@di.uoa.gr

Abstract
We study the formation of opinions in a social context. It has been observed that when in a social environment people often average their opinion with their friends’ opinions, so as to highlight their common beliefs. We analyze a popular social network and verify that social interaction indeed results in influence on opinions among the participants. Moreover we create instances of the network using real data and, based on the game theory framework, we experimentally show that the repeated averaging process results to Nash equilibria which are illustrative of how users really behave.

Introduction
An ever-increasing amount of social activity information is available today, due to the exponential growth of online social networks. The structure of a network and the way the interaction among its users impacts their behavior has received significant interest in the sociology literature for many years. The availability of such rich data now enables us to analyze user behavior and interpret sociological phenomena at a large scale (Anagnostopoulos, Kumar, and Mahdian 2008).

Social influence is one of the ways in which social ties may affect the actions of an individual, and understanding its role in the spread of information and opinion formation is a new and interesting research direction that is extremely important in social network analysis. The existence of social influence has been reported in psychological studies (Kelman 1958), as well as in the context of online social networks (Bond et al. 2012). The latter usually allow users to endorse articles, photos or other items, thus expressing briefly their opinion about them. Each user has an internal opinion, but since she receives a feed informing her about her friends’ endorsements, her expressed (or overall) opinion may well be influenced by her friends’ opinions. This process may lead to a consensus.

The most notable example of studying consensus formation due to information transmission is the DeGroot model (DeGroot 1974). This model considers a network of individuals with an opinion which they update using the average opinion of their friends, eventually reaching a shared opinion. In (Friedkin and Johnsen 1990) the notion of an individual’s internal opinion is added, which, unlike her expressed opinion, is not altered due to social interaction. This captures more accurately the fact that consensus is rarely reached in real world scenarios. The popularity of a specific article, for instance, may vary largely among different communities in a social network. This fact gives rise to the study of the lack of consensus, and the quantification of the social cost that is caused by disagreement (Bindel, Kleinberg, and Oren 2011); the authors here consider a game where user utilities grow with agreement and perform repeated averaging to get the Nash equilibrium of the opinion formation process. The resulting models of opinion dynamics in which consensus is not in general reached allow for testing against real-world datasets to substantiate influence existence.

An approach towards verifying the existence of influence against real-world data based on general models is that of (Anagnostopoulos, Kumar, and Mahdian 2008). However, the proposed models are probabilistic and the correlation the data exhibit is not attributed to influence. Investigating game theoretic models of networks against real data is crucial in understanding whether the behavior they portray depicts an illustration that is close to the real picture.

Our contributions We study the spreading of opinions in social networks, using a variation of the DeGroot model (Friedkin and Johnsen 1990) and the corresponding game detailed in (Bindel, Kleinberg, and Oren 2011). We perform an extensive analysis on a large sample of a popular social network and highlight its properties to indicate its appropriateness for the study of influence. The observations we make verify our intuition regarding the source and presence of social influence. Furthermore, we initialize instances of games using real data and use repeated averaging to calculate their Nash equilibrium. We experimentally show that our model, when properly initialized, is able to mimic the original behavior of users and captures the social cost affecting their activity more accurately than a classification model utilizing the same information. We note that our model does not capture the cases in which a user’s opinion in a social network changes due to reasons other than social influence.
Model

We focus on a setting in which a group of individuals (also called users) are members of a social network, and investigate the impact of social influence on their opinions on some issue. We are concerned with users that perform, within this network, a certain action for the first time. Consider for example a social network in which a user can endorse articles and get informed about her friends’ endorsements. We examine whether this information modifies the users’ opinions on related subjects, e.g., their preferences in articles. We represent the social network as a directed graph. Each node of this graph corresponds to a user, and there is an edge from node $i$ to $j$ iff user $i$ gets informed about the actions of $j$.

In order to identify adjustments in users’ opinions, we have to observe the system for a certain period of time. We keep track of the endorsements of every user (implying their opinion) and the endorsements of their friends, and compare their opinions in the initial state of the system to those in the final state. We assume that if a user endorses an article after some friend of hers has done so, the endorsement results from influence. Therefore, this comparison indicates whether some opinion has changed under social influence.

We model the users’ opinions using the notions of (Friedkin and Johnsen 1990) and the game of (Bindel, Kleinberg, and Oren 2011). Each user $i$ maintains a persistent intrinsic belief $s_i$ and an overall (or expressed) opinion $z_i$; $s_i$ remains constant, while $z_i$ is updated iteratively, during the game, through averaging. In what follows, opinion refers to the overall opinion. We assume that for each user $i$ that endorsed some article prior to all of her friends it is $s_i = 1$, otherwise $s_i = 0$. The real number $z_i$ represents the probability that $i$ endorses the article. At each time step user $i$ updates $z_i$ to minimize the cost of suppressing her internal opinion and disagreeing with her friends in the aforementioned game, using the formula:

$$z_i = s_i + \frac{\sum_{j \in N(i)} w_{ij} z_j}{1 + \sum_{j \in N(i)} w_{ij}}$$

(1)

where $N(i)$ denotes the set of nodes that $i$ follows and $w_{ij}$ expresses the strength of the influence of $j$ on $i$. This averaging process terminates when $z$ converges to the unique Nash equilibrium, which minimizes the social cost, defined to be the sum of all players’ costs. According to our intuition, the influence of $j$ on $i$ regarding a specific article is strong if $i$ generally respects $j$’s opinion and/or $j$ is authoritative on the article under consideration. We therefore define $w_{ij} = a_{ij} b_j$, where $a_{ij}$ expresses how much $j$ influences $i$ in general, and $b_j$ expresses the expertise of user $j$ on the topic of the article. The weights $w_{ij}$ are used to distinguish between influential and weak connections, and are real numbers that may be greater than 1, as a user may value the opinion of a friend higher than her intrinsic belief.

Empirical analysis

We studied the behavior of users in a social network and the impact of social influence on their actions in its context by analyzing a sample of Digg, a social news aggregator, to which we will refer to as the digg dataset in the following (Hogg and Lerman 2012).

Digg allows users to submit links to news stories and vote them up (digg). A user is also able to follow other members and track the stories they recently voted for. The digg dataset consists of the votes of the 3, 553 most popular stories of June 2009, and the directed social graph depicting the followers of each voter. A total of 3, 018, 196 votes from 139, 409 users and 1, 731, 658 follower edges is available.

Information propagation and reproductive ratio

Information can travel through many paths in a social network and identifying word-of-mouth hops that form social cascades is a rather infeasible task. To differentiate the users of digg who endorsed a story due to social influence from the ones that acted freely, we adopt the heuristic used in (Cha, Mislove, and Gummadi 2009) and consider an endorsement to have propagated from user $i$ to user $j$ if $j$ endorsed a story after $i$ did, and $j$ followed $i$ before endorsing the story. If multiple users $i$ satisfy these conditions, we assume that the propagation was caused by all of them. Those users $j$ that endorsed a story having no users $i$ influencing them are considered as seeders.

In epidemiological models, the reproductive ratio, denoted $R_0$, is used to measure the potential for disease spread in a population (Anderson and May 1991). If $R_0 > 1$, an infected individual is expected to infect more than one other individuals and the infection will be able to spread in a population, otherwise the infection will die out. We observe that in more than 92% of the stories of the digg dataset the total ‘infections’ were less than the number of initial seeders, indicating that the reproductive ratio is well below 1.

We also examine the distribution of the total cascades caused by every individual for every story of the dataset, which appears to be heavy-tailed. This verifies our intuition about the presence of authoritative users, although the average transmission probability is quite small. However, it is worth noting that influence varies depending on the story, and even the users that frequently trigger cascades tend to be more effective in certain stories than in others. Therefore a good estimation of $b_j$ can only occur when examining it in the context of a single topic.

Frequent cascade patterns

To further study the complex collective behavior attributed to the interaction of social network users, we mined the frequent cascade patterns occurring in the digg dataset. In particular, we considered a graph with a node $v_{ir}$ for every \{voter $i$, story $r$\} pair and an edge from $v_{ir}$ to $v_{jr}$, if voters $i$ and $j$ are friends and $i$ endorsed a story $r$ before $j$ did. Figure 1 illustrates the 20 most frequent cascade patterns met in such a graph, formed with a sample of 50 stories. We extracted these patterns using GRAMI (Elseidy et al. 2014).
Figure 1: Top-20 cascades that occurred in the digg dataset, ordered by frequency.

We observe that the spread of information for the stories of our dataset exhibits small chain- and tree-like cascades, as was the case with most of the datasets examined in (Leskovec, Singh, and Kleinberg 2006). However, splits are much more infrequent than collisions in our dataset, as opposed to the datasets of (Leskovec, Singh, and Kleinberg 2006); only the smallest possible split was found among the 20 most frequent patterns (ranked 10th).

**Experimental evaluation**

Having verified our intuition regarding the adoption of an opinion due to social influence, we apply (1) on real-world data to examine its fitting performance with respect to our findings. We conduct experiments on the cascade graphs of the digg dataset stories to answer:

- How much more improved is the precision of (1) when distinguishing the crucial aspects of social interaction related to the spread of influence?
- How does (1) perform against a neural network classifier?

**Simulation methodology**

We perform repeated averaging in our model until the opinions converge to the unique Nash equilibrium. To initialize our model we apply the following set-up for each story we examine:

- We consider that every user of digg that endorsed a story before any of the members she follows did so, has a strongly positive intrinsic opinion about it. However, we cannot hypothesize on the intrinsic opinion of users that voted up a story after at least one of the users they follow did so, as their behavior can be attributed to numerous causes. Hence, we consider for user $i$: $s_i = 1$, if $i$ voted a story before any user she follows, and 0 otherwise.
- Regarding the influential strength of $j$ on $i$, $w_{ij}$, we consider two variants:
  
  (i) A straight-forward approach where users are equally authoritative on all stories of digg, i.e., $b_j = 1$ for every $j$. Additionally, users are equally influenced by all the members they follow and are not influenced at all by the rest of the users, i.e., $a_{ij} = 1$ if user $i$ follows user $j$, and 0 otherwise.
  
  (ii) An approach that follows our intuition that the influence of $j$ on $i$ increases with the ratio of votes of $j$ that $i$ followed and builds on the findings reported during our empirical analysis. We specify $a_{ij}$ by how frequently $j$ influences $i$, using information about the total influence of $j$ on $i$ to compute the influence on a certain story:

  \[
  a_{ij} = \frac{\text{# times } i \text{ is influenced by } j}{\text{# votes of } j}
  \]  
  (2)

  Moreover, we quantify $b_j$ by how authoritative user $j$ is for the article under consideration:

  \[
  b_j = \frac{\text{# users influenced by } j \text{ in this story}}{\text{# followers of } j},
  \]  
  (3)

  thus capturing the expertise of each user per story.

Algorithm 1 outlines our approach. We perform repeated averaging with both configurations until we reach convergence to the unique Nash equilibrium of the corresponding games. At the state of convergence, the expressed opinions of the users are given values in $[0, 1]$. Deciding whether a value stands for endorsement or not calls for the use of a threshold. We examine the trade-off between precision and recall by varying the threshold value to obtain the respective curves, where precision is the fraction of users predicted as endorsers that actually voted up, while recall is the fraction of users that voted up that are predicted to do so.

We also use a Neural Network to train a classification model that predict user actions and compare against it. To this end, we perform stratified 10-fold cross-validation using the following independent variables, normalized in the range $[-1, 1]$: (i) $s_i$, (ii) the sum of (2) for every $j$ friend of $i$, and (iii) the sum of (3) for every $j$ friend of $i$.

**Experiments using real-world data**

We performed extensive simulations using our methodology on different stories of the digg dataset, and obtained precision-recall curves for the two configurations of our model. Furthermore, we obtained the respective curves that occur with the use of the Neural Network classifier.

\(^2\)Our implementation is available here: https://bitbucket.org/network-analysis/social-cost

\(^3\)We employ RapidMiner’s Neural Net operator: https://rapidminer.com/
Discussion and open directions

In this work we specified the details which allow for a better formalization of the social effects on a variation of the DeGroot model. To verify our intuition on the causes of social influence, we performed a comprehensive analysis on a real-life popular social network. In addition to this, we initialized several instances of the dataset and applied repeated averaging on them to calculate the state where our model converges. As it is shown in (Bindel, Kleinberg, and Oren 2011), this state is the unique Nash equilibrium of the game defined by the individual cost functions. We presented results comparing this state with the original actions of the network’s members. Our findings show that a properly initialized instance following our model, converges to a Nash equilibrium that closely mimics the original social activity of a real-world dataset. Therefore, we verified that users act in order to minimize their cost as described in this work.

There are still many interesting open directions left for future research. We can proceed with estimating the price of anarchy in such a network. Moreover, we can examine ways to reduce the social cost of these networks.

References