Modeling Opinion Dynamics in Social Networks

Abhimanyu Das
Microsoft Research
Silicon Valley
abhidas@microsoft.com

Sreenivas Gollapudi
Microsoft Research
Silicon Valley
sreenig@microsoft.com

Kamesh Munagala
Dept. of Computer Science
Duke University
kamesh@cs.duke.edu

ABSTRACT

Our opinions and judgments are increasingly shaped by what we read on social media – whether they be tweets and posts in social networks, blog posts, or review boards. These opinions could be about topics such as consumer products, politics, life style, or celebrities. Understanding how users in a network update opinions based on their neighbor's opinions, as well as what global opinion structure is implied when users iteratively update opinions, is important in the context of viral marketing and information dissemination, as well as targeting messages to users in the network.

In this paper, we consider the problem of modeling how users update opinions based on their neighbors' opinions. We perform a set of online user studies based on the celebrated conformity experiments of Asch [1]. Our experiments are carefully crafted to derive quantitative insights into developing a model for opinion updates (as opposed to deriving psychological insights). We show that existing and widely studied theoretical models do not explain the entire gamut of experimental observations we make. This leads us to posit a new, nuanced model that we term the BiasedVoterModel. We present preliminary theoretical and simulation results on the convergence and structure of opinions in the entire network when users iteratively update their respective opinions according to the BiasedVoterModel. We show that consensus and polarization of opinions arise naturally in this model under easy to interpret initial conditions on the network.

1. INTRODUCTION

Opinion formation among individuals and the resulting dynamics it induces in a social network has been widely studied, both from the perspective of analytic modeling as well as experimental psychology. It is known that there are several complementary influence processes that shape opinion formation [6] in society – these include group (or normative) influence, influence due to formation of stereotypes, and so on. One of these processes is termed *informational influence*[1], where users lacking necessary information seek the opinions of their neighbors in order to update their beliefs. Such a process is, for instance, common in opinion formation for fashion, consumer products, and music. Many celebrated theoretical models of opinion formation in networks use informational influence as the underlying premise [4, 17, 5, 16]. Specifically, these works model how an individual user updates her opinion in the face of information learned from her neighbors, and then use the model to characterize the evolution of opinions in the network in terms of convergence time, and the emergence of a consensus or polarization.

However, many of these theoretical models are inspired by processes in physics and biology, for instance, flocking behavior in birds [16], or cellular automata [4, 17, 18]. Though the natural processes inspiring these models are intuitively related to human behavior, it is conceivable that human behavior, even at the level of informational influence, is not as clean as these models make it out to be. In fact, there is scant literature that precisely connects *any* of these quantitative models to empirical human behavior.

We ask: *Do the existing theoretical models adequately explain how users update opinions when faced with opinions of their neighbors? If not, how do we gather empirical data, and can we posit a faithful model from this? And finally, can we understand the dynamics of opinion formation in a network when individuals iteratively update opinions using this model, in terms of convergence to a common opinion, polarization, and convergence time to these states?*

In this paper, we focus on the problem of modeling how users update opinions based on their neighbors' opinions. We begin by a preliminary analysis of opinions expressed by users on Twitter[2] to gain initial understanding into if and how a user's opinions are correlated with her neighborhood opinions. We show that such a correlation indeed exists, even when restricted to sentiments (positive or negative) about a particular topic, such as Hyundai Cars, Tim Cook, or Organic Food. (See Appendix A for details.) Though Twitter is a rich source of information, trying to derive a model of user behavior from such data is not feasible: First, sentiment analyzers we use are noisy and imprecise, producing mainly qualitative data; and secondly and more importantly, there are several competing influence processes

[1]This is different from the information cascade and diffusion models proposed in [14, 20, 13, 11, 25]. See Sections 2.2 and 2.3 for more details.

[2]`http://www.twitter.com`

WSDM'14, February 24–28, 2014, New York, New York, USA.
Copyright 2014 ACM 978-1-4503-2351-2/14/02 ...$15.00.
http://dx.doi.org/10.1145/2556195.2559896.

and group dynamics that operate in such a network, so that teasing out informational influence is a difficult problem.

We therefore perform a carefully designed set of user experiments in order to generate our insights into modeling. These experiments measure the shift in user opinions when exposed to the responses of other users, and are based on the celebrated conformity studies of Asch [1] and subsequent work. While the aim of their experiments was to derive qualitative psychological insights, we design ours carefully toward the end-goal of recovering a coarse *quantitative model* to explain the observations. In our experiments, we recover three distinct types of opinion update processes arising out of *stubbornness, compromise, and biased conformity*. Stubborn users do not change opinion; compromising users update their opinion to be a convex combination of their initial opinion and their neighbors' opinions; and conforming users simply take on the opinions of their neighbors, ignoring their own – the choice of neighbor biased by the closeness of their opinion to the user's own. We show, somewhat surprisingly, that these behaviors are not immutable – a user can take on different behaviors when faced with different amounts and types of information.

The above experiments motivate us to define a new analytic model of user behavior that we term the BIASED-VOTERMODEL. The model is surprisingly simple with only two parameters, one capturing the *stubbornness* in behavior, and the other capturing the *conformity bias* of the user. We study the dynamics that this simple model induces, both analytically as well as via simulations. We show that this model reaches a consensus under reasonable assumptions; however, it can take exponential time to converge to such a consensus, and the states in which it spends most time in transience can be heavily polarized. In this sense, it explains persistent disagreement in society even when users take into account *all* opinions in their neighborhood in their opinion formation process, something that we empirically observe. In contrast, to achieve polarization, existing models either require users to ignore opinions that are less related to their own in their opinion formation process, or require them to always be stubborn regardless of the information they receive – neither of these assumptions is faithful to the experimental data we observe.

2. PRIOR WORK

In this section, we present a brief survey of the commonly used theoretical models of opinion formation. We further survey experimental evidence from psychology and economics literature. In the discussion below, we will use the terms agent, user, node, and vertex interchangeably.

2.1 Theoretical Models

The broad research area of opinion formation is quite classical, and we refer the interested reader to [19] for a survey. The earliest work in this domain comes from the sociology and statistics literature [27, 5, 8].

Averaging Models.

One notable example in this setting is the DEGROOT model [9, 15, 5] which studies in a fixed network how consensus is formed when individual opinions are updated using the average of the neighborhood. Formally, suppose there are n agents located in an undirected graph $G(V, E)$. The Link $(i, j) \in E$ has weight $w_{ij} = w_{ji}$, which captures homophily between agents i and j. Let $N(i)$ denote the set of agents who have an edge to agent i. The opinions of the agents lie on a real line. At time t, let $\vec{z}(t) \in \Re^n$ denote the vector of the opinions of the agents. The update rule in this model is the following:

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

We assume $w_{ii} \geq c \times \sum_{j \in N(i)} w_{ij}$, where c is an absolute constant. We term c the *relative weight* of the initial opinion of i. Typically instances of the DEGROOT model in the literature updated a user's opinion by averaging her opinion with the *mean of her neighboring opinions*. Therefore, for concreteness, we use DEGROOT for the setting where $c \geq 0.25$.

This model can be generalized easily to directed links, and it is easy to show using a contraction mapping argument that this model always reaches a consensus assuming the graph induced by links with $w_{ij} > 0$ is strongly connected.

In reality, consensus is rare in society, and opinions tend to typically be in a state of persistent disagreement, often in highly polarized states. The DEGROOT model was extended by Friedkin and Johnsen [10], to include both disagreement and consensus by associating with each node i an innate opinion s_i in addition to its expressed opinion. In this model, the dynamics becomes:

$$z_i(t+1) = \frac{w_{ii} s_i + \sum_{j \in N(i)} w_{ij} z_j(t)}{w_{ii} + \sum_{j \in N(i)} w_{ij}} \quad (2)$$

We note that averaging models predict persistent disagreement via two means: (1) Agents can be stubborn, *i.e.*, have $w_{ij} = 0$ for all neighbors, and (2) Agents can have different intrinsic opinions s_i.

Flocking Models.

It is folklore that opinion formation has a *conformity bias* [1, 23, 6], *i.e.*, agents assign more weight to opinions that already conform to their beliefs. This phenomenon is a special case of *flocking* behavior (or influence systems [3]).

Several classic opinion formation models combine averaging with flocking. In the popular *bounded confidence* model of Hegselmann and Krause [16] (HK model), opinions lie on a real line. Suppose agent i has opinion z_i. For fixed ϵ, the confidence region of an agent is captured by the set

$$S_i(\vec{z}) = \{j \mid |z_j - z_i| \leq \epsilon\} \quad (3)$$

Each agent i iteratively updates its opinion by averaging the opinions of agents in $S_i(\vec{z})$. In the DW model of Weisbuch et al. [28], two random nodes i and j meet and update their opinions z_i and z_j to the average of the two opinions if these originally satisfied $|z_i - z_j| \leq \epsilon$.

There is also a large body of work (see [3, 24] and references therein) that has focused on characterizing the convergence of bounded confidence dynamics to either *absolute consensus* or some clustering (polarization).

Voter Models.

The voter model was independently proposed by Clifford and Sudbury [4], and by Holley and Liggett [17]. In this model, we have a strongly connected directed graph and opinions are discrete, typically coming from a small set [30]. At each step, a node is selected at random; this node chooses one of its neighbors uniformly at random (including itself) and adopts that opinion as its own. A modification is termed the Label Propagation Algorithm, where the node adopts the majority opinion in its neighborhood (breaking ties randomly). The former algorithm is termed the *voter algorithm*, and can be shown to reach a consensus. Techniques based on coalescing random walks [7] bound the convergence time and steady state distribution of this dynamics.

There are several variants of the voter model known. As a relevant example, Yildiz *et al.* [29] consider the modification where agents can be stubborn. Kempe *et al.* [21] propose a modification where the possible opinions (or types) define a graph, where two types are adjacent if they correspond to similar opinions. They restrict the neighborhood graph on the users to be a *complete graph*. At any step, every user picks a random other user whose type neighbors his own type, and adopts that opinion.

2.2 Asch's Experiments

We now move to experimental evidence to support models for opinion formation. Most of this research comes from social psychology, particularly the work of Asch [1], Deutsch and Gerard [6] and extensive followup research (see [2] for a survey). Asch, in his seminal conformity experiments, observed that a large fraction of his subjects could be swayed to essentially say "white is black" if everyone else in the group said the same. In Asch's experiments, the subject was placed in a room with several stooges who unanimously gave the false (binary) response to a question where the true answer was self-evident. Asch made several observations, the most relevant being the following: (1) The extent of conformity (percent of subjects who conform) depends on the size of the group; and (2) Even when one of the stooges gives the true response, that is sufficient to reduce conformity significantly.

Deutsch and Gerard argued that Asch's observations are the superposition of two types of conformity – *normative*, where subjects conform so as to gain social acceptance into a group, and *informational*, where subjects use the responses of others as evidence of reality, and adjust their beliefs accordingly. They ran similar experiments where the other responses shown to the subject were anonymous and fabricated (though the subject believed these responses were real), and observed similar behavior to Asch's experiments – in effect showing that Asch's observations happen even with purely informational influence and no group pressure, though to a slightly lesser extent. They therefore term Asch's experiments a study of informational influence as opposed to normative influence.

It is also well-known (see [26] for a survey) that there are other mechanisms for opinion formation that apply for medium to long-term interactions. The most widely accepted of these is the notion of stereotyping – subjects form a stereotype of themselves, as well as of others, and use these stereotypes to adopt the opinion of homophilous neighbors. This stereotype could itself be formed based on previous opinions expressed on possibly unrelated topics. Though these mechanisms are important in driving opinion formation, they involve the formation of implicit and explicit groups of users, which will not be the main focus of this paper.

2.3 Diffusion/Influence Models

We briefly mention work on diffusion, influence and information flows in networks that, though not directly relevant to our work, discuss models for the adoption or spread of ideas, rumors or content among online users. Well known models in this domain include Threshold [14] and Cascade models [12] that specify how a node adopts a particular idea or product based on the adoption pattern prevalent in its neighborhood. Subsequently, several papers studied, both theoretically [20] and empirically [11, 13, 25] the phenomenon of diffusion of ideas or content in a social network and the related problem of identifying influential nodes to seed in order to maximize adoption rates. The work in this space is not directly related to our problem since it is mainly concerned with the *dissemination* or the spread of information (in general) through the network while the focus of our work is on the *assimilation* of this information to form an opinion (with possibly a different value from the neighboring values) by a individual user in the network. This often leads to a much finer-grained opinion dynamic model for a single user, which is what we seek to model.

2.4 Our Contribution

In this paper, we focus on informational influence (defined in Section 2.2), which we believe will be the dominant mode of influence in short-term anonymous interactions, for instance on review boards or chat rooms. We analytically model opinion formation in social networks in the presence of informational influence using a model that we term BiasedVoterModel. Our model combines certain aspects of the flocking and the DeGroot models, yet it is different from both. We base our model on several online user experiments that are modeled after the experiments of Deutsch and Gerard – in contrast with their experiments, our experiments do not have an easily discernible ground truth, and the responses are not binary (true/false). We note that binary outcomes can be consistent with several analytical models simultaneously – we therefore make the responses themselves continuous in our experiments, and carefully choose the neighborhood to tease out the modeling desiderata.

We study the response of the subjects not only as a function of the size of the neighborhood, but also as a function of the *distribution* of these opinions, and the *distance* of these opinions from the subject's own. Not only are our observations consistent with prior work on binary opinions (and generalize them), but they also provide us sufficient qualitative and quantitative insights to postulate a model.

3. ONLINE USER STUDIES

We will now present experiments to distinguish all of the above models at a qualitative level. Our online user experiments are modeled after the anonymous experiments of Deutsch and Gerard [6]. The key differences are that we require continuous quantitative responses to questions whose ground truth is uncertain; and further, we ask questions from several different domains to check consistency of the findings. We will show that for these experiments, none of the DeGroot, flocking, or the voter models consistently explains our data, motivating the need for the BiasedVoterModel model.

3.1 Experimental Setup

We used an interacting network of online users recruited from Amazon Mechanical Turk (mturk) and a personal survey website to host the experiments. Subjects were asked to take part in one of two online surveys hosted on the external website. The first survey corresponded to questions about estimating the number of dots in three images. The second survey asked questions about estimating the monthly sales volume of two car models (Hyundai and Ford) and two soda brands (Coke and Pepsi).

Each survey consisted of two sequential phases: In the first phase, for each question on the survey, the user was asked to enter her response to the question. The user did not know apriori there would be a second phase. Then, in the second phase, the user was shown a set of "neighboring opinions" that were synthetically generated (as described later) but were presented to the user as being the most recent answers from previous participants of the survey. The user was then asked to read these neighboring opinions and re-enter her answer to the question. The first answer of the user was used as her initial opinion, while the second answer which was provided by the user after having seen her neighboring opinions was treated as her final opinion.

SURVEY1 (DOTS): This survey consisted of three questions about three images, each containing 10000 5000, and 4000 dots respectively. The user was asked to guess the number of dots in each of the images. We varied the number of neighboring opinions shown for each question: 20 opinions for the first question, 10 for the second question, and 5 for the third question. In all three cases, the neighboring opinion distribution was generated using a (roughly) trimodal distribution with around 20% of the nodes clustered around the neighboring opinion closest to the user's initial opinion, and around 60% clustered around the opinion farthest from the user's initial opinion. We also repeated the entire survey for the case where the neighboring opinions were randomly permuted before being shown to a user, instead of being presented in sorted order.

SURVEY2(CARS AND CANS): This survey asked each user four questions about estimating the sales of car and soda brands (Coke, Pepsi, Ford Focus and Hyundai Elantra) during June 2014. The generation of neighboring opinions for each question was similar to that in the previous survey except for two differences: First, we fixed the number of neighboring opinions to be 20. Second, we used different neighboring opinion distributions: for the Coke and Hyundai questions, we used a Gaussian distribution with mean 0 and standard-deviations of 2000 and 300 respectively. For the Pepsi and Ford questions, we used a truncated Power-Law distribution, with exponent 2. In both cases, we then shifted the set of neighboring opinions (generated using the above distributions) by a constant that depends on the user's initial opinion, as described below.

For all our experiments, the set of neighboring opinions were generated such that they were either all smaller or all larger than the user's initial opinion. Figure 1 gives an illustration of the generation of neighboring opinions based on a user's initial opinion. We use o_i to denote the user's initial opinion, o_e denotes the neighboring opinion closest to o_i, o_c denotes the median neighboring opinion, and o_m denotes the neighboring opinion farthest from o_i. For the dots experiment, we generated the neighboring opinions such that the gap $|o_i - o_e|$ was set to 4000 if o_i was less than 30000 and $o_i - 26000$ otherwise. Similarly, for the cars and soda experiments, the gap $|o_i - o_e|$ was set to 4000 if o_i was less than 60000 and $o_i - 56000$ otherwise.

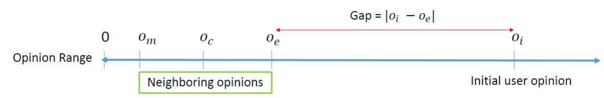


Figure 1: Illustration of user/neighboring opinions for User Experiments

Each survey was answered by around 200 online users. To ensure that the survey response quality was high, we restricted each user to participate in at most one survey.

3.2 Results and Interpretation

As mentioned above, we carefully construct the experiment so that it can qualitatively distinguish between the various models of opinion formation – DEGROOT, Flocking, and VOTER models.

3.2.1 DEGROOT *versus* VOTER *model*

We begin with an experiment which shows that for large number of neighboring opinions, the VOTER model explains the opinion dynamics better compared to the DEGROOT model. We remind the reader of a user following the DEGROOT model [5] averages her opinion with the opinions of her neighbors. As mentioned before we assume for concreteness that such a user gives at least $c \geq 0.25$ relative weight to its own opinion. A user following the VOTER model chooses a neighboring opinion at random and moves to this opinion. Toward this end, we consider a large set of users whose initial opinions are *significantly far* from the set of neighboring opinions. We define these as follows:

DEFINITION 3.1. *Let o_i denote the user's initial opinion. We define o_i to be* significantly far *if $|o_i - o_e| \geq 3|o_c - o_m|$.*

The above definition ensures that we can make the following claim – if many users indeed follow the DEGROOT model, their final opinion should lie in between their initial opinion and the set of the neighboring opinions[3]. Note that this observation precludes the stubborn who do not change their opinions by much. To further make the distinction clear, we restrict our attention to experiments where the neighboring opinions are drawn from a Gaussian distribution with relatively small variance – the Hyundai and Coke experiments. This further ensures that variance in the neighboring opinions does not confuse the analysis. Thus, users whose final opinions lie within the relatively small range of neighboring opinions are likely to follow the VOTER model.

Based on the above, we categorize users based on the following quantity. Let o_f denote the final opinion of the user. Let $s = \frac{|o_i - o_f|}{|o_i - o_e|}$. We define a user to be *stubborn* if $s < 0.1$; it follows the VOTER model if $s > 0.9$, and it follows DEGROOT if $s \in [0.1, 0.9]$. Fig. 2 illustrates this categorization of users. We note here that our results do not qualitatively change much even if we vary the s threshold used for categorizing users into stubborn, DEGROOT and VOTER models slightly; this can be seen by observing in Figures 5(a) and

[3] While the constant 3 in Definition 3.1 is sufficient to ensure that we can distinguish between DEGROOT and VOTER models, our experimental results use even more conservative constants of around 8

5(b) that the distribution of users into the three categories is relatively stable to small perturbations in the s thresholds.

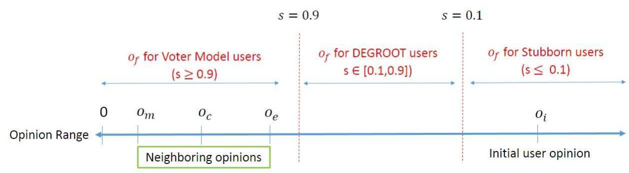


Figure 2: Categorization of user into DEGROOT, Stubborn or VOTER model

We restrict our attention to users whose initial and final opinions are at least 1000 and whose initial opinions satisfy Def. 3.1. In Fig. 3(a), we plot the histograms for the three buckets defined above for both the experiments. We observe that a overwhelming fraction of around 75% of the users indeed adopt their final opinion to be one of (or close to) the neighboring opinions consistent with the VOTER model. Furthermore, we observe that even the number of stubborn users (of around 15%) is larger than the remaining set we surmise to be following the DEGROOT model. This observation follows from the fact that the final opinions of all the DEGROOT user's lie in between their initial opinion and the set of neighboring opinions.

To make the conclusion more concrete, we focus on the Coke experiment. We filter on initial opinions at least $500K$. For such users, the corresponding neighboring opinions lie in a small range $[44K, 56K]$, with mean $50K$. The mean is factor 10 away from the initial opinion, so that a user performing DEGROOT with relative weight at least $c \geq 0.1$ to his own opinion and using 20 neighbors will be clearly separated from users who adopt an opinion close to their neighbors. We plot the histogram in Fig. 3(b). As before, at most 15% of users perform DEGROOT as opposed to more than 65% who follow the VOTER model. To complete the analysis, we expand the user set to include users with initial opinion at least 250K, and split the users with $s > 0.9$ (VOTER model) into three groups: Those for which $s \in (0.9, 1)$ (who adopt an opinion larger than $56K$), those whose final opinions is in $(54K, 56K]$, and those with final opinions at most $54K$. These are the third, fourth, and fifth buckets respectively, while the first two buckets, $s < 0.1$ and $s \in [0.1, 0.9]$ remain the same. We plot this histogram in Fig. 3(c). We note that more than 50% of users move to an opinion at most $56K$, while at most 15% have $s \in (0.1, 1]$, corresponding to an expanded definition of DEGROOT. Even within the former set of users, a vast majority move to an opinion at most $54K$, showing that they use the mean/median/mode of $50K$ in forming their opinion. This shows the predominance of users whose behavior is better modeled by the VOTER model.

3.2.2 *Bias in Opinion Choice*

The above experiment makes a case that the VOTER model is a more accurate model for the Coke and Hyundai experiments than DEGROOT. We now dig deeper into *which* neighboring opinion is adopted more frequently. Is it a an opinion chosen uniformly at random, or is it biased by the user's initial opinion? The Coke experiment (Fig. 3(c)) shows that users move to an easily discernible majority cluster – since the neighboring opinions are drawn from a Gaussian distribution, the users move to the mode of this distribution.

We consider the DOTS experiment where we show 20 neighboring opinions. We filter only users whose initial opinions lie in $[75K, 150K]$ (the upper bound is not critical and is mainly used to filter out users who we believe give arbitrary answers). For these users, the neighboring opinions lie in $[16K, 26K]$. However, the distribution of these opinions is tri-modal: There are 5 distinct opinions in $[25K, 26K]$, the median is $23.5K$, and the mode is $22K$, with 6 neighboring opinions taking this value. When the data is presented in sorted order, the mode is clearly identifiable.

As before, we bucket the users into stubborn ($s < 0.1$), DEGROOT ($s \in [0.1, 0.9)$), and VOTER model ($s \geq 0.9$). Within the third group, we create three sub-groups, one where $s \in [0.9, 1)$ (where the final opinion is larger than $26K$), the second where the final opinion is in $[26K, 25K]$, and the other where the final opinion is smaller than 25K. The histogram of final opinions is shown in Figure 4. We present two histograms, one where the neighboring opinions are shown in sorted order, and the other where they are shown in random order.

We find that around 40% of the users move to an opinion that is close to the extreme cluster around $26K$ – this includes users for which $s \geq 0.9$ and users whose final opinion is in $[25K, 26K]$. This shows that users give more weight to nearby opinions while forming their final opinion. However, note that there is a small but non-trivial fraction of users who perform DEGROOT (around 20%), and that move to an opinion below $25K$ (around 15%).

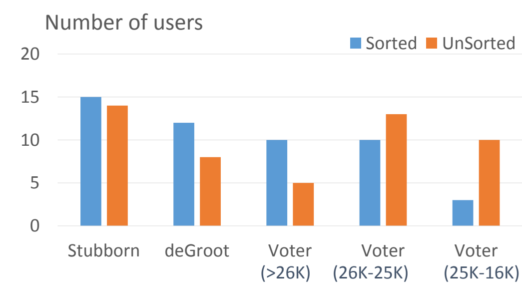
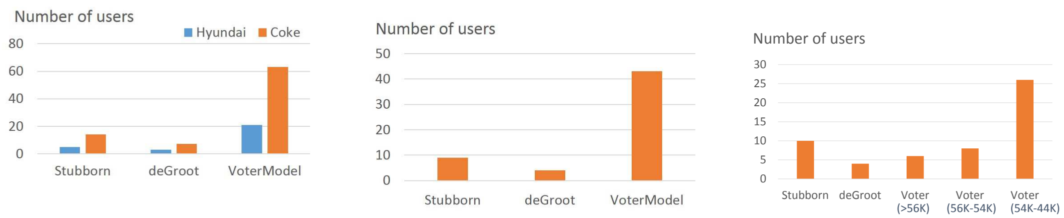


Figure 4: DOTS experiment with 20 neighbors and initial opinions between 75K and 150K

3.2.3 *Effect of Size of Neighborhood*

In the next set of experiments, we measure the effect of the number of neighboring opinions on the final opinion of the user. Toward this end, we considered a set of 20, 10, 5, and 1 neighboring opinions for the DOTS question. As Figures 6(a) and 6(b) shows, there are two effects that are observed. First, stubbornness reduces as users see more neighboring opinions. We observe that the fraction of stubborn nodes decreases from around 35% to 22% as the number of neighboring opinions changes from 1 to 20. This is a natural consequence of a user not seeing enough number of alternate opinions to sway her from her initial opinion, a phenomenon that was documented as early as Asch [1]. Second, we observe more users adopting the DEGROOT model as the number of neighboring opinions decrease. This could either be because users can easily compute averages over a small set of neighboring opinions, or because they are swayed less and adopt an intermediate response. The point we wish

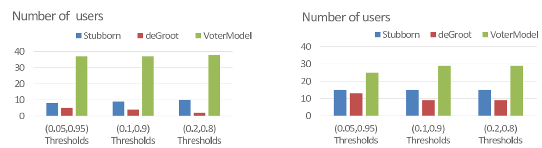


(a) Hyundai and Coke experiments with initial opinions *significantly far*.

(b) Coke experiment with initial opinions at least $500K$, and neighboring opinions in $(44K, 56K)$.

(c) Coke experiment with initial opinions at least $250K$ and finer-grained bucketing of final opinions.

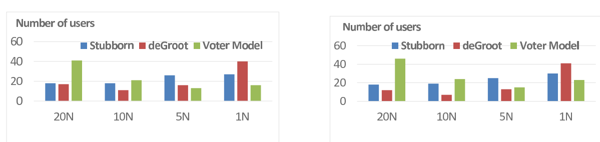
Figure 3: Coke and Hyundai Experiments



(a) Hyundai and Coke experiment.

(b) DOTS experiment.

Figure 5: Distribution of Stubborn, DEGROOT and VOTER users for various s-threshold values used for categorizing users into the three classes



(a) Sorted neighboring opinions.

(b) Unsorted neighboring opinions.

Figure 6: Effect of the number of neighboring opinions. This was measured using the DOTS experiment with significantly far initial opinions

to make is that the pure VOTER model cannot capture the behavior of users when there are few neighboring opinions, since a significant fraction of them move to an opinion in between their initial opinion and the neighboring opinions.

3.2.4 Correlation in Responses

We finally study whether user behavior is consistent across questions – do stubborn users always give stubborn responses, and do conforming users always conform?

To address this, we first consider each user's behavior across all the answered by that user in a survey. We categorize users as *consistently stubborn* if the user is stubborn for all questions, and respectively *consistently* DEGROOT and *consistently* VOTER. We label a user as *mixed* if her behavior changes across questions. Figure 7 summarizes the distribution of these categories in our experiments, for both our Surveys. As seen in the Figure, around 63%(for the DOTS survey) and 55% (for the CARS AND CANS survey) of users exhibit consistent behaviors across the questions answered by them.

Next, we consider a more fine-grained consistency test to understand if user exhibit consistent behavior even within the DEGROOT and VOTER models in terms of selecting the averaging constant (for DEGROOT) or the choice of neighboring opinion to adopt (for VOTER). We consider two questions in the DOTS survey, that use 20 and 10 neighboring opinions respectively. The same user answers both these questions. As before, we restrict only to users whose initial opinions for *both* questions are *significantly far* according to Def. 3.1. We now construct one value q from the responses of the user. We compute the quantity s as before. Let $t = \max(0, \min(s, 1))$. Without loss of generality, assume o_i is larger than all neighboring opinions. Sort the neighboring opinions as $w_1 \geq w_2 \geq \cdots \geq w_k$, so that w_1 is the closest opinion to o_i. Let r denote the index of o_f in this sorted order, where if $o_f > w_1$, the index is 0. Let $r' = 2r/k$. (For instance, if the user reports o_f as the median $w_{k/2}$ of the neighboring opinions, then $r' = 1$.) Let $r'' = \min(r', 1)$. Then, we define $q = t + r''$. Therefore, q is a number between 0 and 2. If $o_f \in (w_1, o_i]$, then $q = s$. If $o_f = w_1$, then $q = 1$; if $o_f = w_{k/2}$, then $q = 2$.

We compute the Pearson's correlation coefficient between the q values for the 20 and 10 neighbors case for the DOTS experiment, restricted to *significantly far* initial responses for both questions. There are 90 such users. The correlation coefficient is 0.57 with p-value less than 0.001. We repeat the above computation for users answering both Hyundai and Ford questions, again restricting to both initial opinions being *significantly far*. We obtain a Pearson correlation coefficient of 0.4 with p-value less than 0.001.

4. DESIDERATA OF A NEW MODEL

We have shown above that by themselves, neither the DEGROOT model, the VOTER model, nor the Flocking model is a good qualitative model for our data; the DEGROOT model cannot explain the Coke and Hyundai experiment where users with large initial opinions move close to a neighboring opinion (nor can it explain Asch's observations); the VOTER Model cannot explain the Dots experiment with few neighboring opinions where users seem to be performing averaging, and the Flocking models cannot explain why users are not stubborn when their initial opinions are very divergent from their neighbors (nor can it explain why stubbornness should decrease with more neighboring opinions). In fact, we observe that users can be largely classified into three basic behavioral types *viz*,

Stubborn Behavior: Users do not change their opinions, regardless of how extreme their opinion is relative to the neighborhood, and how many neighbors they have;

Compromising (deGroot) Behavior: Users choose an opinion that is in between their initial opinion, and the average of the opinions in their neighborhood; and

Biased Conforming Behavior: Users move to an opinion chosen at random from their neighborhood, giving more weight to opinions that are closer by to their initial opinion. We remind the reader that the Voter model corresponds to an unbiased conforming behavior.

Note that the same user can exhibit multiple types of behavior depending on the neighborhood (represented by the *mixed* category in Figure 7). While admitting all these behaviors, we would also like our model to gracefully handle the effects of the neighborhood size as well as a user's bias in opinion adoption we observed in our experiments.

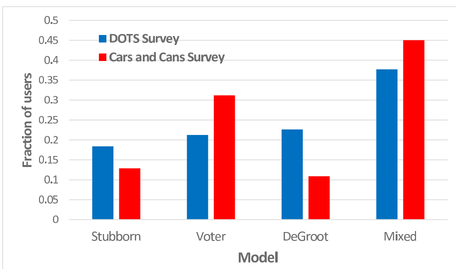


Figure 7: Histogram of users following a certain type of behavior

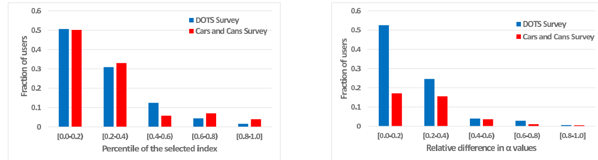
Effect of Size of Neighborhood.

When there are a large number of neighboring opinions, the same user is much more likely to discount his own opinion a lot and simply adopt a neighboring opinion, much like the Voter model. For small neighborhood size, we observe that the same users do indeed adopt average opinions reasonably frequently (much like deGroot). While we did not explicitly vary the number of opinions with time in our user-studies, users, in an online setting, are quite likely to see different number of opinions at different points in time. Therefore, the effect of varying neighboring opinions needs to be carefully modeled, something not addressed in all previous models.

Bias in Opinion Adoption.

Asch's experiments show that conformity with a divergent majority dramatically reduces when there is one neighboring opinion that agrees with the user's own. This is in accordance with the classical Flocking models. However, unlike these models, we do not observe any "confidence interval" beyond which users are stubborn. In fact, we find that users move to the neighboring opinion even when they are arbitrarily far away, the choice of neighbor biased towards the user's own opinion.

We state that our goal of proposing a new model is not to fit our experimental data exactly, since human behavior, even on simple questions, is not likely to admit one simple model. Instead we aim to propose a simple extension to the Voter model that admits the above desiderata. We observe that a geometric model captures the above findings reasonably well. Figure 8(a) illustrates the *bias* in the



(a) Effect of the distance of a neighboring opinion on the likelihood of it getting selected.

(b) The stubbornness of user across multiple questions for a fixed number of neighboring opinions.

Figure 8: Effect of key parameters of a candidate model

choice of neighboring opinions selected by users in our surveys. Clearly, we observe a significant decrease in the probability of choosing a neighbor's opinion further from a user's innate or initial opinion. In other words, as the index of the neighbor's opinion in the presented order increases. On the effect of *stubbornness*, we observe that a user's stubbornness is independent of the distance of neighboring opinions —in fact, it is relatively independent of the specific question in a given survey too. This observation is shown in Figure 8(b) which plots the average of $|s_i - \bar{s_i}|/k$ for all users following the deGroot model. Here $\bar{s_i}$ is the average s of user i across all the k questions she participated in. Note that $s = 0$ denotes a stubborn user while a larger value indicates a user's willingness to conform. In the next section, we will describe our model incorporating both stubbornness and bias parameters.

5. THE BIASED VOTER MODEL

In this section, we propose a model that is more qualitatively consistent with our (and prior) observations that we noted in the previous section. Our model, which we term BiasedVoterModel, uses a combination of the deGroot model and the Voter model. Formally, there is a directed graph $G(V, E)$ on the users. User i has a set of neighbors $N(i)$, which is the set of users i follows, and whose opinions i potentially receives. We assume the opinion dynamics evolve over a sequence of discrete time steps. Let $z_i(t)$ denote the opinions of user i at time step t, and let $S_i(t)$ denote the set of opinions of the neighbors of i at time t. This is precisely the set of opinions $z_j(t)$ for $j \in N(i)$. Each user i then sets $z_i(t+1)$ based on $z_i(t)$ and the set $S_i(t)$, as described by the BiasedVoterModel below;

User i is parametrized by two innate quantities: we term these as the *bias parameter* p_i, and the *stubbornness parameter* α_i. We start with a discrete type space $\mathcal{T} = \{1, 2, \ldots, k\}$ of opinions. The *distance* between two types $a, b \in \mathcal{T}$ is simply $|a - b|$, so that the type of 1 is an extreme opinion, and type of k is the opposite extreme opinion.

Let $z_i(t) = q_0$. For set $S_i(t)$, sort the opinions in increasing order of distance from q_0; denote this multi-set of opinions as $\{q_1, q_2, \ldots, q_d\}$, where d is the degree of user i.

Let $q^* = \text{argmin}_{q \in S}|q - q_0|$ be the "closest" neighboring opinion to i's own. Then:

- With probability p_i, set $z_i(t+1) = q_1$; else with probability p_i, set $z_i(t+1) = q_2$, and so on.
- If the above step did not set $z_i(t+1)$, then:
  - With probability α_i, set $z_i(t+1) = q_0$; and
  - With probability $1 - \alpha_i$ set $z_i(t+1) = q$, where $q \in [q_0, q^*]$ is chosen uniformly at random.

Note that this model is faithful to the experimental results: As $|S|$ increases, the user is more likely to conform to a neighboring opinion, and this opinion is more likely to be closer to the initial opinion. We further note that if all nodes update opinions each step, and if $p_i = 1/d_i$, where d_i is the degree of i in $G(V, E)$, then BiasedVoterModel reduces to the classical Voter model [4, 17] since users choose a neighboring opinion (almost) uniformly at random.

5.1 Dynamics

In this section, we present some analytical results about the dynamics of the BiasedVoterModel, mainly focusing on the nature of stable equilibria, and the rate of convergence towards them. In all the results below, we assume $G(V, E)$ is strongly connected; if this assumption does not hold, the theorems hold in each strongly connected component. The proofs of these observations follow from standard Markov chain analysis techniques [7, 29], and we omit them.

In the results below, by *consensus*, we mean that all users adopt the same opinion. By a *stable equilibrium*, we mean that the opinions of all nodes have converged to some value from which it does not change further.

Theorem 5.1. *In the* BiasedVoterModel, *the following are true:*

- *The frequencies of different opinions converges, though there may be no stable equilibria in the sense that the opinions of a individual user may not converge.*
- *If $\alpha_i, p_i < 1$ for all users i, then the only stable equilibria correspond to consensus.*

The most interesting of the above observations is that in the BiasedVoterModel, if every user compromises a little and there are no stubborn users with $\alpha_i = 1$, then this is enough to make each connected component reach a consensus. However, note that if there is a clique of size k all of whose users have the same opinion, then it takes at least $1/(1-p)^k$ steps for any user in this set to adopt a different opinion. In a sense, it is easy to see that though consensus is the only stable equilibrium, there could be transient states that are polarized, and that are persistent in the sense that the system spends exponential time in these states. Such transience is likely to be pronounced if there are a lot of triadic closures in the graph. On the other hand, if there are stubborn users, consensus is not guaranteed and the opinions of users can oscillate; however the distribution of opinions reaches stability (see Figure 9).

We note that though our model can be viewed as a biased version of the Voter model, the coalescing random walk techniques [7] for analyzing steady state properties in the Voter model do not naturally extend to BiasedVoterModel. Developing new analysis techniques in this domain is an interesting open question. In the interim, we simulate this dynamics to understand the steady state behavior as a function of the stubbornness and bias parameters.

5.2 Simulations

We performed simulations to characterize the effect of stubbornness (parameter α_i) and bias (parameter p_i) exhibited by a node on the dynamics underlying the discrete BiasedVoterModel. Toward this end, we generated graphs having both random and power law degree distributions. The number of nodes in the graph was set to 1000. The maximum degree in the random graph was set to 50 while the corresponding number in the power law graph was set to 200. We assume $k = 10$, so that every node takes on one of 10 opinions in $\{0, 1, 2, \ldots, 9\}$. We define a stubborn node i as one with $\alpha_i = 1$ and $p_i = 0$. We fixed the stubborn nodes to assume one of the two extreme opinions, *viz.*, 0 and 9. The rest of the nodes have the corresponding values drawn uniformly at random. Finally, we note that all results are averaged over 10 runs.

Before we characterize the effect of both the parameters, we make an observation on the convergence of the dynamics. Theorem 5.1 shows that even in the presence of stubbornness and bias, the distribution of opinions in the BiasedVoterModel dynamics always converges, though the rate of convergence can be slower than in the Voter model (where p_i is small). While we observed the voter model converges to a consensus almost always within 3000 iterations on a 1000 node random graph, the corresponding behavior is different in the BiasedVoterModelFirst, there presence of stubborn nodes results in no consensus at convergence. Second, the number of iterations taken to converge could be significantly higher. Figure 9 illustrates the latter for different choices of the fraction of stubborn nodes in the graph where we see that the number of users adopting any of the stubborn opinions (i.e. opinions 0 or 9) stabilizes after around 70,000 iterations.

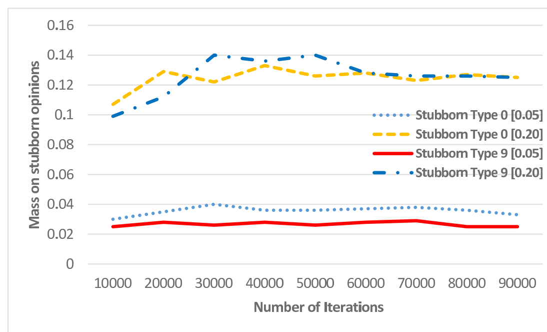


Figure 9: Convergence of the BiasedVoterModel dynamics with different fraction (5% and 20%) of stubborn nodes with one of opinions 0 and 9. The bias parameter p_i is uniformly distributed in [0,1].

Effect of Stubbornness: We varied the fraction of nodes exhibiting stubbornness in the graph and measured the number of nodes adopting any of the stubborn opinions at the 10000th iteration in the dynamic. Figure 10 shows the total number of nodes adopting the stubborn opinions, i.e., 0 or 9 for starting configurations with different fraction of stubborn nodes. Clearly, if all nodes adopts one of the stubborn opinions, we can say that the network is polarized along these opinions. Any decrease in the number of stubborn nodes can only lead to more polarization possibly involving smaller groups and more opinions. Alternately, this can be viewed as a rate of convergence. In this context, we observe that the placement of the stubborn nodes plays a crucial role in the rate of convergence, i.e., the adoption of the stubborn opinions by rest of the graph. Clearly, the slowest rate of convergence is associated with the case when the stubborn nodes are the low degree nodes in the power law graph. On the other hand, when the stubborn users are placed on the high degree nodes in the power law graph, the graph almost immediately converges to the stubborn opinions even when the fraction of the stubborn users is as small as 10% in the graph. Not surprisingly, under uniform seeding of stubborn nodes, the power law graph converges faster compared to the dynamics on a random graph.

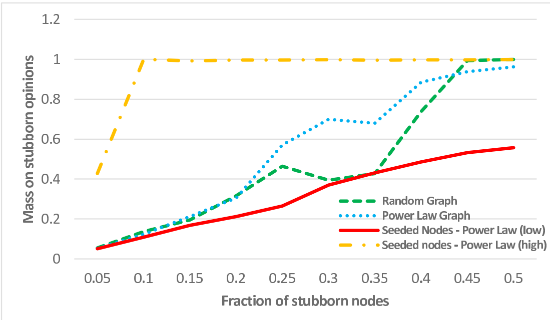
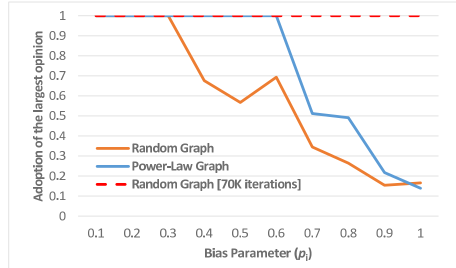


Figure 10: The effect of the placement of the stubborn nodes on the fraction of nodes adopting the stubborn opinions at $t = 10000$. The bias parameter p_i is uniformly distributed in [0,1].

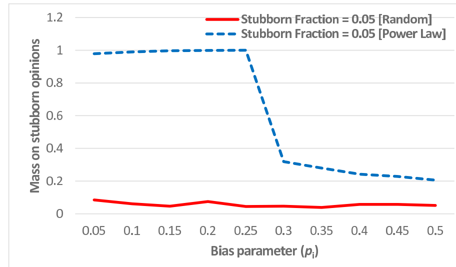
Effect of Bias: Next, we set the bias p_i of every node to be the same and varied the value of this parameter in [0,1]. In the absence of any stubborn nodes, Theorem 5.1 gives us consensus (the dashed line in Figure 11(a) confirms this result). Keeping this in mind, we studied the extent of polarization and consensus in the network at the 10000th iteration. A transient behavior that arises with increase in the bias value is the breakdown of any consensus that existed for small values of p_i. This can be explained by the observation that as the propensity to select the nearest neighbor increases, users tend to quickly converge to smaller groups (having same opinion). Figure 11(a) illustrates this observation. This effect is more pronounced in the power-law graph. On the other hand, when there are stubborn nodes in the graph (see Figure 11(b)), we observe a different behavior. In the case of a random graph, there is hardly any consensus on the stubborn opinions regardless of the value of p_i. In the case of a power-law graph, we note that smaller values of p_i are sufficient to break any consensus compared to the behavior observed in Figure 11(a). Indeed, this is precisely the behavior characteristic of the BIASEDVOTERMODEL wherein the presence of stubborn nodes as well as bias toward particular opinions in the neighborhood can have adverse effect on consensus formation in the network.

6. CONCLUSION

Our work is a preliminary attempt to posit an analytical model for opinion formation and informational influence, based on carefully designed online experiments. We first mention some caveats of our experiments and model, which we plan to resolve in future work. In the user experiments, we have shown neighboring opinions that are chosen in a structured fashion. We have not explored if this has any side effects on opinion formation. Secondly, the questions in this study have a hard to discern ground truth, and are chosen so that users do not have side information apriori. We do not know if this impacts the model for informational influence in a significant way.



(a) No stubborn nodes



(b) The fraction of stubborn nodes is 5%.

Figure 11: Effect of bias parameter p_i on consensus formation after $t = 10000$ iterations.

There are several avenues for future research. The BIASEDVOTERMODEL leads to interesting dynamics that needs to be analytically explored. In the presence of such dynamics, how can messages be targeted to users in the network so that the final set of opinions is close to a desired outcome? We have focused on informational influence and have ignored normative influence as well as stereotyping. How do the presence of these types of influence interact with informational influence?

Acknowledgements. Kamesh Munagala is supported by an award from Cisco, and by NSF via grants CCF-0745761, CCF-1008065, and IIS-0964560.

7. REFERENCES

[1] S. Asch. Opinions and social pressure. *Scientific American*, 193:35–40, 1955.

[2] R. Bond and P. B. Smith. Culture and conformity: A meta-analysis of studies using Asch's line judgment task. *Psychological Bulletin*, 119(1):111–137, 1996.

[3] B. Chazelle. Natural algorithms and influence systems. *Communications of the ACM*, 2012.

[4] P. Clifford and A. W. Sudbury. A model for spatial conflict. *Biometrika*, 60:581–588, 1973.

[5] M. H. DeGroot. Reaching a consensus. *J. American Statistical Association*, 69:118–121, 1974.

[6] M. Deutsch and H. B. Gerard. A study of normative and informational social influences upon individual judgement. *J. Abnormal and Social Psychology*, 51:629–636, 1955.

[7] R. Durrett and H. Kesten. *Random walks, Brownian motion, and interacting particle systems.* Birkhauser Verlag AG, 1991.

[8] J. Fowler and N. Christakis. Cooperative behavior cascades in human social networks. *PNAS*, 107(12):5334–8, 2010.

[9] J. R. P. French. A formal theory of social power. *Psychological Review*, 63:181–94, 1956.

[10] N. E. Friedkin and E. C. Johnsen. Social influence and opinions. *J. Mathematical Sociology*, 15(3-4):193–205, 1990.

[11] S. Goel, D. Watts, and D. Goldstein. The structure of online diffusion networks. In *ACM EC*, pages 623–638, 2012.

[12] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3):211–223, 2001.

[13] M. Gomez Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *Proc. of the 16th ACM SIGKDD*, pages 1019–1028. ACM, 2010.

[14] M. Granovetter. Threshold models of collective behavior. *American journal of sociology*, pages 1420–1443, 1978.

[15] F. Harary. A criterion for unanimity in French's theory of social power. In D. Cartwright, editor, *Studies in Social Power*. Institute for Social Research, Ann Arbor, MI, 1959.

[16] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Artificial Societies and Social Simulations*, 5(3), 2002.

[17] R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The Annals of Probability*, 3(4):643–663, 1975.

[18] P. Holme and M. E. J. Newman. Nonequilibrium phase transitions in the coevolution of networks and opinions. *Phys Rev E Stat Nonlin Soft Matter Phys*, 74:056108, 2006.

[19] M. O. Jackson. *Social and Economic Networks*. Princeton University Press, Princeton, NJ, USA, 2008.

[20] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proc. of the ninth ACM SIGKDD*, pages 137–146. ACM, 2003.

[21] D. Kempe, J. M. Kleinberg, S. Oren, and A. Slivkins. Selection and influence in cultural dynamics. In *ACM EC*, pages 585–586, 2013.

[22] Lexalytics. Salience engine - `http://www.lexalytics.com/technical-info/salience-engine-for-text-analysis`.

[23] C. G. Lord, L. Ross, and M. R. Lepper. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *J. Personality and Social Psychology*, 37(11):2098–2109, 1979.

[24] J. Lorenz. Heterogeneous bounds of confidence: Meet, discuss and find consensus! *Complexity*, 15(4):43–52, 2010.

[25] S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *Proc. of the 18th ACM SIGKDD*, pages 33–41. ACM, 2012.

[26] V. Price, L. Nir, and J. N. Cappella. Normative and informational influences in online political discussions. *Communication Theory*, 16(1):47–74, 2006.

[27] T. C. Schelling. Models of segregation. *American Economic Review*, 59(2):488–493, 1969.

[28] G. Weisbuch, G. Deffuant, F. Amblard, and J.-P. Nadal. Meet, discuss and segregate! *Complexity*, 7(3):55–63, 2002.

[29] E. Yildiz, D. Acemoglu, A. Ozdaglar, A. Saberi, and A. Scaglione. Discrete opinion dynamics with stubborn agents. *ACM EC*, 2013.

[30] M. E. Yildiz, R. Pagliari, A. Ozdaglar, and A. Scaglione. Voting models in random networks. In *Information Theory and Applications Workshop (ITA), 2010*, pages 1–7, 2010.

APPENDIX

A. OPINION CORRELATION ON TWITTER

We present empirical evidence on Twitter to show that for a wide range of topics, the sentiments that a user expresses are correlated with the sentiments in its neighborhood. We note that one cannot conclusively infer evidence of influence from such data, but the very existence of correlation even for sentiments of the *same* topic is motivation to study models of opinion formation and dynamics in social networks.

Dataset: Our dataset comprised of user opinions extracted from a large set of tweets corresponding to one of four topics: "Hyundai", "Tim Cook", "Organic food", and "Electric Cars". We considered all tweets related to these topics (using simple keyword-based classifiers) within a 6-month timeline from 12/1/2012 to 5/31/2013. The total number of tweets in each topic varied between 200000 and 4000000. We then ran each tweet through a commercial sentiment analyzer [22] software to obtain three categorical opinions value: negative, positive, and neutral. For each node (user) and topic, we associated a list of opinions ordered by their timestamps. Using the Twitter follow graph, we computed the induced subgraph over the nodes (across all topics) with around 2 million nodes and 300 million edges.

To unearth statistically-significant evidence of correlations between a user and her neighborhood, we analyze each user's list of opinions expressed in the 6-month window, and compare it with a list of all her neighboring users' opinions expressed in that same window. To alleviate the inherent noise in the opinion mining and sentiment extraction process, we ignore all neutral opinions in our experiments and only consider users that have expressed at least two (non-neutral) opinions. We then perform the following two experiments across each of the 4 topics:

Table 1: Experiment 1

Fraction of +/- opinions in +/- class	Hyundai	Tim Cook	Electric Cars	Organic Food
+ve opinion, +ve class	0.706	0.584	0.430	0.832
+ve opinion, -ve class	0.553	0.550	0.347	0.733
-ve opinion, -ve class	0.331	0.420	0.594	0.206
-ve opinion, +ve class	0.260	0.409	0.529	0.138

Table 2: Experiment 2

Distance between +/- user opinion and +/- nbr opinion	Hyundai	Tim Cook	Electric Cars	Organic Food
+usr, +nbr	884	652	763	600
-usr, +nbr	1149	944	855	788
-usr, -nbr	1007	661	683	1324
+usr, -nbr	1428	928	851	1707

Experiment 1: We categorize users into two classes: a positive (resp. negative) class and a non-positive (resp. non-negative) class, based on whether they have expressed at least one positive (resp. negative) opinion. For each class, we then consider the set of neighboring opinions of users of that class, and measure the overall fraction of positive (resp. negative) opinions in that set. We observe that this fraction is higher for the set of neighboring opinions corresponding to the positive (resp. negative) class, than for the non-positive (resp. non-negative) class. Table 1 illustrates these results. The difference between the two classes is statistically significant at p-value levels of much less than 0.001 using two-sample t-tests.

Experiment 2: Next, we measure the average distance (in time) from a positive opinion of a user to her closest neighboring positive opinion in the past, and compare this to the average distance from a negative opinion of a user to her closest neighboring positive opinion in the past. We observe (in Table 2) that the former distance (in units of seconds) is much smaller than the latter, which suggests that conditioned on her neighborhood expressing a positive opinion, a user is more likely to express a positive opinion than a negative opinion. Similar observations hold for negative opinions as well. As before, we validated that these differences are statistically significant at p-value levels of much less than 0.001 using two-sample t-tests.