Interaction-Driven Opinion Dynamics in Online Social Networks

Stacy Patterson
Department of Mechanical Engineering
University of California, Santa Barbara
Santa Barbara, CA 93106
sep@cs.ucsb.edu

Bassam Bamieh
Department of Mechanical Engineering
University of California, Santa Barbara
Santa Barbara, CA 93106
bamieh@engineering.ucsb.edu

ABSTRACT

Online social networks provide a globally available, massive-scale infrastructure for people to exchange information and ideas. A topic of great interest in social networks research is how to model this information exchange and, in particular, how to model and analyze the effects of interpersonal influence on processes such as information diffusion, influence propagation, and opinion formation. Recent empirical studies indicate that, in order to accurately model communication in online social networks, it is important to consider not just relationships between individuals, but also the frequency with which these individuals interact. We study a model of opinion formation in social networks proposed by De Groot and Lehrer and show how this model can be extended to include interaction frequency. We prove that, for the purposes of analysis and design, the opinion formation process with probabilistic interactions can be accurately approximated by a deterministic system where edge weights are adjusted for the probability of interaction. We also present simulations that illustrate the effects of different interaction frequencies on the opinion dynamics using real-world social network graphs.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Sociology*; C.4 [**Performance of Systems**]: Modeling techniques

General Terms

Theory, Performance

Keywords

Distributed consensus, Spectral perturbation analysis

1. INTRODUCTION

Online social networks provide globally available, massive-scale information infrastructures for people to exchange information and ideas. A topic of great interest in social networks research is how to model these exchanges and, in particular, how to model and analyze the effects of interpersonal influence on processes such as information diffusion, influence propagation, and opinion formation. If one can identify the properties of the network that shape these process, then it may be possible to alter these properties to achieve a desired outcome, for example to encourage purchase of a product or support of a cause.

Early work to characterize interpersonal influence in social networks focused on static, structural properties of the network graph such as vertex degree, distance centrality, and betweenness centrality [25]. However, recent investigations of data from real-world, online social networks indicate that the dynamics of user interactions may be equally, if not more important than these static properties in determining how information flows in a social network [14, 26, 24]. It is not sufficient to only consider the network graph derived from relationships such as friendships in Facebook. One must also consider the frequency and timing of interactions between individuals. In fact, the structure of the interaction graph may differ dramatically from that of the relationship graph.

In this paper, we study the process of opinion formation in online social networks and the effects of interaction dynamics on this process. Research on opinion formation in social groups predates the advent of online social networks by decades, and several formal mathematical models of the opinion formation process have been proposed [7, 5, 15, 9, 12]. These models all share the assumption that individuals communicate with each other in a synchronized fashion, and the models do not allow for any variation in the frequency of interaction. We consider one such model which was proposed by De Groot [5] and Lehrer [15], and we show how this model can be extended to include interaction frequency. We then prove that, for the purposes of analysis and design, the opinion formation process with probabilistic interactions can be accurately approximated by a deterministic system with no interaction dynamics, but where edge weights are adjusted for the probability of interaction. The benefit of this result is that any analysis or design that has been done for the model with no interaction dynamics can automatically be extended to a network with probabilistic interactions. We highlight several of these design and analysis results in Section 2. We also present evaluations of the effects of different interaction frequencies on the opinion dynamics using real-world social network graphs. These evaluations also illustrate the accuracy of the proposed de-

1st Workshop on Social Media Analytics (SOMA '10), July 25, 2010, Washington, DC, USA.

terministic model in approximating opinion dynamics with probabilistic interactions.

Related Work

Several models for interaction dynamics in online social networks have been proposed. The most similar to our work is that of Song et al. [22]. In this work, the authors model the diffusion of an innovation in a network using a continuous-time Markov chain, and they adjust the rates of flow between individuals to account for communication delay. They then use this model to rank the influence of users and to determine the correlation between the adoption of an idea at pairs of users. The work by Tang et al. [22] presents a temporal graph model of a social network and proposes global and local (per user) structural metrics for this temporal graph. The authors define several structural properties similar to those studied in static graphs that incorporate communication dynamics.

We also note that the De Groot/Lehrer model of opinion dynamics is very similar to distributed consensus algorithms studied in the context of multi-agent systems and cooperative control. There is a large body of work in this area that may provide insight into the process of opinion formation with different types of interaction dynamics. In particular, this work was inspired by work on consensus in stochastic networks [17, 18, 21, 23, 6, 20].

The remainder of this paper is organized as follows. In Section 2, we describe the model of opinion dynamics which we call classical opinion dynamics, and we highlight important analytical results and applications for this model. Section 3 contains the main contribution of this work. We show how the classical opinion dynamics model can be augmented to include the frequency of interaction between users. We then present analysis of this new model that is analogous to that given for the classical model. We also derive an approximation of the interaction-driven system that can be used to solve network design problems in an efficient manner. Section 4 presents measurements of the efficiency of the opinion formation process in several real world networks for both the interaction-driven model and the approximation of this model. Finally, we conclude in Section 5.

2. CLASSICAL OPINION DYNAMICS

The social network is modeled by a graph $G = (V, E)$ where the vertices represent users or *agents*, with $|V| = n$, and the edges represent relationships between users, with $|E| = m$. The graph may be directed or undirected. A directed graph is used to model networks where relationships are not symmetric, for example, the follower relationship in Twitter. An undirected graph models a network with symmetric relationships such as the friend relationship in Facebook. We say that agent i is a *neighbor* of agent j if $(i, j) \in E$. In this work, we restrict our analysis to undirected graphs, though many of the results presented in this paper hold in both directed and undirected networks. We indicate this fact where applicable.

We consider the model of group opinion formation within a social network that was proposed by De Groot [5] and Lehrer [15]. Each agent i has an initial opinion $x_i(0)$. The opinion is assumed to be a real number, for example a numerical representation of the agent's support for an issue. The agreement process takes place in discrete rounds. In each round, each agent updates his opinion based on information exchanged along edges in the network graph, taking a weighted average of his own opinion and the opinions of his neighbors. Let w_{ij} be the weight that agent i places on the opinion of agent j with the normalization requirement that $\sum_{j=1}^{n} w_{ij} = 1$. We refer to w_{ij} as the weight of edge (i, j). In each round, agent i updates his opinion as follows,

$$x_i(t+1) = w_{i1}x_1(t) + w_{i2}x_2(t) + \ldots + w_{in}x_n(t), \qquad (1)$$

where $w_{ij} > 0$ only if $(i, j) \in E$. We note that the formulation (1) admits the possibility that $w_{ij} = 0$ even if $(i, j) \in E$, meaning i does not place any weight on the opinion of j even though they are neighbors in the network.

The vector $x(t) = [x_1(t)\ x_2(t)\ \ldots\ x_n(t)]^*$ is the *opinion profile* of the network at time t. The opinion profile evolves according to the following linear recursion equation,

$$x(t+1) = Wx(t) \qquad (2)$$

where $W = [w_{ij}]$ is the $n \times n$ matrix of edge weights. An alternate definition of the weight matrix W can be obtained from the Laplacian matrix of the graph G. Let A be the weighted adjacency matrix of the graph, and let the degree of each node i be defined

$$d_i = \sum_{j \in V, j \neq i} w_{ij}.$$

d is the vector of node degrees. The weighted Laplacian matrix of the graph G is

$$L := diag(d) - A.$$

The weight matrix W is equivalent to

$$W = I - L.$$

We say that the opinion formation process described by the recursion (2) *leads to consensus* if eventually, all agents' opinions converge to a single value, or formally, if for every initial profile $x(0)$, there exists a consensus value c such that $\lim_{t\to\infty} x_i(t) = c$ for all $i \in V$.

We note that in the formulation presented above, the communication structure of the opinion formation process is defined by the non-zero entries of W and that this structure remains the same for the duration of the process. Every node communicates with all of its neighbors (for which the entry of W is positive) in every round. In Section 3, we show how the system (2) can be augmented to incorporate the interaction frequency of each pair of nodes. In the remainder of this section, we describe some of the analytical results and applications of the opinion formation model given in (2).

2.1 Social Influence Analysis

The following questions are of particular interest in the study of opinion dynamics and interpersonal influence in social networks [8].

(Q1) *What network topology, edge weights, and initial profile lead to consensus?*

(Q2) *What is the consensus value?*

(Q3) *How efficient is the agreement process?*

We note that the opinion dynamics model (2) is a linear recursion and that the matrix W is stochastic. Therefore, it is possible to draw from various tools in linear analysis and Markov theory to analyze the evolution of the opinion profile $x(t)$ and determine answers to the above questions. Below, we highlight theoretical results that provide answers to these questions.

Consensus Conditions

The following theorem gives a sufficient condition for consensus [5]. This condition applies to both directed and undirected networks.

THEOREM 2.1. *Let the edge weights in the system (2) be such that for every two agents $i, j \in V$, there exists a third agent $k \in V - \{i, j\}$ such $w_{ik} > 0$ and $w_{jk} > 0$. Then, the system leads to consensus.*

We note that if the network graph is strongly connected, then it is always possible to find an assignment of edge weights that leads to consensus [18]. For example, if every edge weight is equal to a value α, $0 < \alpha < \frac{1}{\Delta}$, where Δ is the maximum node degree of the graph, then convergence to consensus is guaranteed.

A necessary and sufficient condition for consensus has also been derived.

THEOREM 2.2. *The system (2) leads to consensus if and only if the matrix W is primitive,* i.e. *if 1 is a simple eigenvalue of W and all other eigenvalues have magnitude less than 1 [10, 18].*

This theorem also holds for both directed and undirected networks.

Consensus Value

In undirected networks, the matrix W is doubly stochastic. It is well-known that if the agents convergence to consensus, the consensus value is the average of the initial opinions, $x_{ave} := \frac{1}{n}\sum_{i\in V} x_i(0)$. In directed networks, the consensus value depends on both the initial opinion profile and the network topology [18].

Agreement Efficiency

Agreement efficiency is a measure of how quickly nodes reach consensus. Before we present results on efficiency, we require an alternate definition of the notion of consensus. Let the *deviation from consensus vector* be defined as

$$\tilde{x}(t) = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^*\right) x(t)$$

where $\mathbf{1}$ is the n-vector of all ones. Each component $\tilde{x}_i(t)$ indicates how far agent i is from the consensus value,

$$\tilde{x}_i(t) = x(t) - x_{ave}.$$

The system (2) leads to consensus if, for every $x(0) \in \mathbb{R}^n$,

$$\lim_{t\to\infty} \|\tilde{x}(t)\| = 0.$$

Agreement efficiency can be defined as the rate at which $\|\tilde{x}(t)\|$ approaches 0, which can be obtained from the following inequality,

$$\|\tilde{x}(t)\| \leq |\lambda_2(W)|^t \|\tilde{x}(0)\|.$$

$\lambda_2(W)$ is the second largest eigenvalue of the weight matrix W (by magnitude). We call $\lambda_2(W)$ the *decay factor* because it is the factor by which $\|x(t)\|$ decreases in each round. This inequality also defines how many rounds are required for agents' opinions to be "close" to each other. This result is formally defined in the following theorem (see [19] for proof).

THEOREM 2.3. *In a system with opinion dynamics of the form (2), the number of rounds required for $\|\tilde{x}(t)\|/\|\tilde{x}(0)\| < \epsilon$ is*

$$\frac{\log \epsilon}{\log(\lambda_2(W))}.$$

The analytical results presented in this section enable the prediction of the outcome of the opinion formation process for a given network. One can predict whether agents will reach consensus and how quickly they will do so. These results also allow for the possibility of *strategic network modification* where the network characteristics are altered in order to achieve a desired outcome [8]. In the next section, we describe some possible approaches for strategic network modification based on the classical opinion dynamics model.

2.2 Strategic Network Modification

Strategic network modification is similar to the problem of system design in cooperative control, and thus we can draw from applications of system design in this area. In this section, we describe two applications that have been proposed in the context of multi-agent systems. We believe that these applications hold value for the the social networks setting as well. Both applications are concerned with increasing the efficiency of the agreement process by minimizing $|\lambda_2(W)|$.

(A1) *Selecting edge weights to optimize efficiency of agreement process [28].*
In this application, the network graph is given, and the goal is to find an assignment of edge weights that results in the fastest convergence to consensus. In the context of social networks, the optimal edge weight assignment can be used as a guide to encourage or discourage information flow between agents so as to bring about the fastest agreement.

The edge weight selection problem corresponds to finding the solution to the following optimization problem,

$$\begin{array}{ll} \text{minimize} & |\lambda_2(W)| \\ \text{subject to} & \lim_{t\to\infty} W^t = \frac{1}{n}\mathbf{1}\mathbf{1}^* \\ & w_{ij} > 0 \text{ only if } (i,j) \in E. \end{array}$$

For undirected graphs, where W is symmetric, this problem can be expressed as a semi-definite program which can be solved efficiently for graphs with on the order of 10^5 edges.

(A2) *Forming relationships to improve efficiency of agreement process [11, 4].*
In this application, we are given a weighted graph $G = (V, E_{base})$ and a set of candidate edges E_{cand}. The goal is to choose k edges from E_{cand}. such that adding these edges to E_{base} results in the graph with the most efficient opinion dynamics. The corresponding optimization problem is

$$\begin{array}{ll} \text{minimize} & |\lambda_2\left(W(E_{base} \cup E)\right)| \\ \text{subject to} & |E| = k \\ & E \subseteq E_{cand}. \end{array}$$

The optimal solution can be found by trying all possible combinations of candidate edges. Clearly, this solution will not scale to the size of large online social networks. However, scalable heuristic-based algorithms have been proposed that have been shown to yield good solutions.

The classical model of opinion dynamics and the theoretical results and applications derived for this model can offer insight into the opinion formation process in online social networks. However, this model depends on the assumption that agents communicate with all of their neighbors in every time period. This assumption does not accurately reflect the interaction dynamics observed in real world social networks. In the next section, we show how the classical model can be modified to included interaction frequency, and we present analogous theoretical results for this augmented interaction-driven model.

3. INTERACTION FREQUENCY

It has been observed that users interact frequently with only a small subset of their neighbors and that communication is infrequent along many edges [26]. It is reasonable to expect that the frequency of interaction will have a large impact on the evolution of opinions in online social networks.

To capture the notion of interaction frequency, we associate a (unique) probability of communication p_{ij} with each edge $(i,j) \in E$. p_{ij} is the probability that agents' i and j will communicate in each round. The evolution of the opinion profile with probabilistic interactions is given by the following recursion,

$$x(t+1) = \left(I - \sum_{(i,j)\in E} \delta_{ij}(t) w_{ij} L_{ij} \right) x(t) \qquad (3)$$

where $\delta_{ij}(t)$ are independent Bernoulli random variables with

$$\delta_{ij}(t) \;:=\; \begin{cases} 1 & \text{with probability } p_{ij} \\ 0 & \text{with probability } 1-p_{ij}. \end{cases}$$

This model has been adapted from the model for multi-agent consensus in stochastic networks [20]. Each L_{ij} is the weighted Laplacian matrix of the graph $G_{ij} = (V, \{(i,j)\})$, the graph that contains the same n vertices as the original graph G and the single edge (i,j). When $\delta_{ij} = 1$, agents' i and j exchange information just as they did in the classical model. When $\delta_{ij} = 0$, agents i and j do not communicate.

We note that

$$W = I - \sum_{(i,j)\in E} L_{ij},$$

where W corresponds to the weight matrix defined in (2). Therefore, if $p_{ij} = 1$ for all $(i,j) \in E$, then the model (3) is equivalent to the classical model given in (2) where all agents communicate with all neighbors in every round. If $p_{ij} = 0$ for all $(i,j) \in E$, then no communication takes place in any round, and all agents keep their initial opinions. We are interested in the opinion dynamics for communication probabilities other than these two extremes, and we would like to analyze the consensus conditions, consensus value, and agreement efficiency as was done for the the classical model.

Consensus Value

In the opinion formation process specified by (3), the information exchange in each round can be represented by a matrix $W(t)$. By definition, each $W(t)$ matrix is symmetric, which implies that if the agents reach consensus, the consensus value will be the average of the initial opinion profile. If instead, we consider the scenario where agents communicate with independent probability in each round, *i.e.* $p_{ij} \neq p_{ji}$, each $W(t)$ is no longer guaranteed to be symmetric. In this case, the consensus value is a weighted average of $x(0)$ that depends on the failure probabilities, edge weights, and network structure (see [6]).

Consensus Conditions and Agreement Efficiency

As in the previous section, we address these questions of consensus conditions and efficiency by studying the deviation from consensus $\tilde{x}(t)$. However, since the system (3) is a stochastic system, we must consider the second order statistics of the deviation vector. To this end, we define the autocorrelation matrix of $\tilde{x}$,

$$M(t) := \mathbf{E}\left\{\tilde{x}(t)\tilde{x}^*(t)\right\} = \mathbf{E}\left\{Qx(t)x^*(t)Q\right\},$$

where $Q := \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^*\right)$. The total variance of the deviation from consensus at time t is given by

$$V(t) = \sum_{i\in V} \tilde{x}_i(t)^2 = \mathbf{tr}\left(M(t)\right)).$$

We say that the system *leads to consensus in mean-square* if

$$\lim_{t\to\infty} V(t) = 0.$$

We measure the efficiency of the agreement process by the rate at which $V(t)$ approaches 0.

In order to study the evolution of $M(t)$, and therefore of $V(t)$, we define the matrix-valued operator $\mathcal{W}$

$$\mathcal{W}(X) = \mathcal{W}_0(X) + \mathcal{W}_1(X) + \mathcal{W}_2(X) \qquad (4)$$

with

$$\begin{aligned} \mathcal{W}_0(X) &= Q\,X\,Q \\ \mathcal{W}_1(X) &= \sum_{(i,j)\in E} p_{ij}\left(-XL_{ij} - L_{ij}X + L_{ij}XL_{ij}\right) \\ \mathcal{W}_2(X) &= \sum_{\substack{(r,s)\in E \\ (r,s)\neq(i,j)}} p_{ij}p_{rs}L_{ij}M(t)L_{rs} \end{aligned}$$

and note that $M(t+1) = \mathcal{W}(M(t))$.

The consensus conditions and agreement efficiency of system with stochastic opinion dynamics (3) are given in the following theorem.

THEOREM 3.1. *The system with second order statistics defined by (4) leads to consensus in mean square if and only if $\rho(\mathcal{W}) < 1$, where $\rho(\mathcal{W})$ denotes the spectral radius of $\mathcal{W}$. The total variance of the deviation from consensus decays, in worst case, as*

$$V(t) \le \rho(\mathcal{W})^t V(0),$$

where this upper bound is tight. We call $\rho(\mathcal{W})$ the decay factor *of the stochastic system.*

PROOF. A proof of the convergence condition and an upper bound on the decay factor is given in [3]. This upper bound is shown be a tight upper bound in [20] for the case where the p_{ij}'s are all equal. A similar argument can be applied to the case where the probabilities of interaction are not identical. □

This theorem implies that the following conditions are sufficient for the system to lead to consensus in mean square.

1. The weight matrix W is such that $\lambda_2(W) < 1$.
2. $p_{ij} > 0$ for all $(i,j) \in E$.

In other words, if $\lambda_2(W) < 1$, as long as there is even the smallest probability of communication along each edge in the network, the agents will reach agreement. We note that condition 1 is a necessary condition for consensus but condition 2 is not. Depending on the network topology, agents may still reach consensus even if communication does not take place along all edges.

We can compute the decay factor for the system described by (4) using the matrix representation of the $\mathcal{W}$ operator

$$\begin{aligned}\mathcal{W} &= \left(Q - \sum_{(i,j)\in E} p_{ij}L_{ij}\right) \otimes \left(Q - \sum_{(i,j)\in E} p_{ij}L_{ij}\right) \\ &+ \sum_{(i,j)\in E} (p_{ij} - p_{ij}^2) L_{ij} \otimes L_{ij}, \end{aligned} \tag{5}$$

where $X \otimes Y$ denotes the Kronecker product of the matrices X and Y. $\rho(W)$ is the absolute value of the largest eigenvalue of this matrix.

This matrix representation can also be used in the network design applications described in the previous section by using $\rho(\mathcal{W})$ in the optimization criteria. However, this approach presents a scalability challenge since for a graph with n nodes, the matrix in (5) is an $n^2 \times n^2$ matrix. Even for moderately sized social networks with on the order of 10,000 nodes, this matrix will be too large for computations with standard software. Therefore we propose a simpler system that approximates the behavior of the stochastic system (3) and can be described by an $n \times n$ matrix for use in analysis and design.

A Deterministic Approximation

We define the linear system

$$z(t+1) = \left(I - \sum_{(i,j)\in E} p_{ij}L_{ij}\right) z(t). \tag{6}$$

where each L_{ij} is the weighted Laplacian with the single edge (i,j) and p_{ij} is the probability of communication over edge (i,j) as defined in the previous subsection. Note that each p_{ij} is a scalar, not a random variable. This system (6) is a deterministic system of the same form as (2), but where each non-zero edge weight w_{ij} is reduced by the factor p_{ij}. Analogous to the total variance of the deviation from consensus for the stochastic system (3), we define

$$\overline{V}(t) = \sum_{i\in V} \tilde{z}_i(t)^2$$

where $\tilde{z}(t) = Qz(t)$. The decay factor of $\overline{V}(t)$ is the spectral radius of the matrix valued operator

$$\overline{\mathcal{W}}(X) = \left(Q - \sum_{(i,j)\in E} p_{ij}L_{ij}\right) X \left(Q - \sum_{(i,j)\in E} p_{ij}L_{ij}\right).$$

Unlike the matrix-valued operator in (4) for the stochastic system, there is a simple relationship between $\rho(\overline{\mathcal{W}})$ and the probability-scaled weight matrix defined in (6),

$$\rho(\overline{\mathcal{W}}) = \left(\lambda_2\left(I - \sum_{(i,j)\in E} p_{ij}L_{ij}\right)\right)^2$$

Therefore, one can perform analysis and design of this system using the $n \times n$ matrix

$$I - \sum_{(i,j)\in E} p_{ij}L_{ij}.$$

rather than an $n^2 \times n^2$ matrix.

We now show that the decay factor of this deterministic system is a close approximation of the decay factor of the system with probabilistic interactions.

THEOREM 3.2. *Consider the opinion formation process over a large, connected, undirected network where the probability of interaction along each edge (i,j) is ϵp_{ij} for some small ϵ. The decay factor of the agreement process given by the stochastic system (3) is identical, up to first order in ϵ, to the decay factor of the deterministic system (6) where the weight of each edge (i,j) is reduced by a factor of ϵp_{ij}.*

PROOF. We first define the probability-scaled Laplacian matrix, where the weight of each edge (i,j) is reduced by a factor of p_{ij},

$$\bar{L} := \sum_{(i,j)\in E} p_{ij}L_{ij}$$

Let the eigenvectors of $\bar{L}$ be v_r, $r = 1 \ldots n$, with $\|v_r\| = 1$. It can be shown that the eigenvalues of $\bar{L}$ are

$$\lambda_r(\bar{L}) = \sum_{(i,j)\in E} p_{ij}w_{ij}(v_r(i) - v_r(j))^2.$$

We note that as long as each $p_{ij} > 0$, $\bar{L}$ has a single eigenvalue equal to 0, with eigenvector $\mathbf{1}$, and all other eigenvalues are in the interval $(0,2)$.

We now use spectral perturbation analysis [13] to find the spectral radius of the $\mathcal{W}$ operator defined in (4). Let $\mathcal{W}(X,\epsilon)$ be the matrix-valued function of a matrix X and a scalar $\epsilon \in \mathbb{R}$ of the form

$$\mathcal{W}(X,\epsilon) = \mathcal{W}_0(X) + \epsilon\mathcal{W}_1(X) + \epsilon^2\mathcal{W}_2(X),$$

where each $\mathcal{W}_i$ is a normal operator. $\mathcal{W}$ is analytic for all $\epsilon \in R$. When ϵ is small, the eigenvalues of $\mathcal{W}$ are analytic in ϵ and are each given by the expansion

$$\lambda_k(\epsilon) = \lambda_k^{(0)} + \epsilon\lambda_k^{(1)} + \epsilon^2\lambda_k^{(2)} + \ldots,$$

where $\lambda_k^{(0)}$ is a eigenvalue of $\mathcal{W}_0$ with eigenmatrix $V_k = v_r v_s^*$. In our case $\mathcal{W}_0$ is the identity transformation. Therefore, for all k, $\lambda_k^{(0)} = 1$. The value of the largest eigenvalue of $\mathcal{W}$, up to first order in ϵ, is then

$$\lambda_1(\epsilon) = 1 - \max_{v_r, v_s \in V} \langle v_r v_s^*, \mathcal{W}_1(v_r v_s^*)\rangle, \tag{7}$$

Note that the inner product on matrices is

$$\langle X, Y\rangle = \mathbf{tr}\,(X^*Y).$$

For the set V we are free to choose any set of vectors that span the space orthogonal to $\mathbf{span(1)}$. Here, we let V be the eigenvectors of $\bar{L}$, excluding $v_r = \mathbf{1}$.

The inner product in (7) simplifies as follows,

$$
\begin{aligned}
&\langle v_r v_s^*, \mathcal{W}_1(v_r v_s^*)\rangle \\
= \;& \mathbf{tr}(v_s^* v_r \sum_{(i,j)\in E} p_{ij} \left(-v_r v_s^* L_{(i,j)} - L_{(i,j)} v_r v_s^* \right.\\
&\left. + L_{(i,j)} v_r v_s^* L_{(i,j)}\right)) \\
= \;& -\sum_{(i,j)\in E} p_{ij} v_r^* L_{ij} v_r - \sum_{(i,j)\in E} p_{ij} v_s^* L_{ij} v_s \\
&+ \sum_{(i,j)\in E} p_{ij} v_r^* L_{ij} v_r v_s^* L_{ij} v_s \\
= \;& -\sum_{(i,j)\in E} p_{ij} w_{ij} (v_r(i) - v_r(j))^2 \\
&- \sum_{(i,j)\in E} p_{ij} w_{ij} (v_s(i) - v_s(j))^2 \\
&+ \sum_{(i,j)\in E} p_{ij} w_{ij}^2 v_r(i) - v_r(j))^2 v_s(i) - v_s(j))^2 \\
= \;& -\lambda_r(\bar{L}) - \lambda_s(\bar{L}) + O\left(\lambda_r(\bar{L})\lambda_s(\bar{L})\right).
\end{aligned}
$$

In the above expansion, the notation $v_r(i)$ refers to the i^{th} component of the r^{th} eigenvector.

The value $\lambda_1(\epsilon)$ is maximized when we choose $v_r = v_s$ with eigenvalue $\underline{\lambda}(\bar{L}) := \lambda_{n-1}(\bar{L})$, the second smallest eigenvalue of $\bar{L}$. Therefore, the spectral radius of $\mathcal{W}$ up to first order in ϵ is

$$\rho(\mathcal{W}) = |1 - 2\epsilon\underline{\lambda}(\bar{L}) + \epsilon O(\underline{\lambda}(\bar{L})^2) + O(\epsilon^2)|.$$

The spectral radius of $\overline{\mathcal{W}}$ is exactly

$$\rho(\overline{\mathcal{W}}) = |1 - 2\epsilon\underline{\lambda}(\bar{L}) + \epsilon^2 \underline{\lambda}\bar{L}^2|.$$

For large networks, $\underline{\lambda}(\bar{L}) >> \underline{\lambda}(\bar{L}^2)$, and therefore the $\underline{\lambda}^2$ term plays a negligible role in the convergence behavior of each system. □

This theorem shows that it is possible to model probabilistic interaction dynamics simply by scaling edge weights by the probability of interaction. The resulting deterministic system can be used in network analysis and in the design problems described in the previous section. These design problems can be solved for the $n \times n$ probability-scaled weight matrix rather than the $n^2 \times n^2$ matrix representation of the stochastic system, and the solutions will be near-optimal for the stochastic system.

4. SIMULATIONS

In this section, we illustrate the effect of the frequency of communication on the efficiency of the opinion formation process using several real-world social networks. Our aim is to highlight the fact that interaction frequency can have a large impact on the opinion formation dynamics. Therefore, it is necessary to incorporate interactions if one wishes to accurately model opinion dynamics in online social networks. We also show the accuracy of efficiency measures obtained from the deterministic system when compared with those of the original system with probabilistic interactions. These results demonstrate that it is not necessary to study the original $n^2 \times n^2$ matrix for the stochastic model of opinion formation with user interactions. Instead, one can study the more compact, deterministic model with modified edge weights and use simulation-free techniques to accurately analyze, predict, and alter interaction-driven opinion dynamics processes

We analyze the opinion dynamics in two online social networks. The first is the General Relativity and Quantum Cosmology collaboration network [16]. The nodes in this network correspond to authors of papers; an edge exists between two nodes if the authors have co-authored at least one paper. The network is determined from papers in e-print arXiv [1] in the period from January 1993 to April 2003. We use the largest connected component of this network, which consists of 4,158 nodes and 13,428 edges. The second network is obtained from a 2009 snapshot of the Facebook Monterey Bay regional network [26]. This network has 6,115 nodes (Facebook users) and 31,374 edges (friendships). As neither of these data sets have edge weights, we generate our own edge weights using the Metropolis-Hastings algorithm (see [2]). The weight on edge (i, j) is

$$w_{ij} = \frac{1}{\max(d_i, d_j)},$$

where d_i and d_j are the degrees of nodes i and j respectively. These edge weights guarantee convergence to consensus for any probability of interaction greater than 0.

We measure the agreement efficiency for three different assignments of probabilities of interaction. In the first scenario, the probability of interaction over each edge is 1, meaning every agent communicates with all of its neighbors in each round. In the second and third scenarios, the probabilities of communication along the edges are generated uniformly at random, from the interval $[0.25, 0.75]$ in the second scenario and the interval $[0.01, 0.1]$ in the third scenario.

We quantify the efficiency of the opinion formation process using two values. The first is the decay factor, which was shown in the previous section to be $\rho(\mathcal{W})$, where $\mathcal{W}$ is the matrix-valued operator that defines the opinion dynamics of the network. The second value is the *consensus time* as defined in [27],

$$\tau := \frac{1}{\log(1/\lambda_2(\mathcal{W}))}.$$

The consensus time is the number of rounds required for the total variance of the deviation from consensus to decrease by a factor of $\frac{1}{e}$.

For the deterministic systems, we find the decay factor by computing the spectral radius of the weight matrix adjusted by the interaction probabilities as defined in (6) and then squaring that value. For the stochastic systems, the associated weight matrices are too large for spectral analysis via readily available software. We therefore find the decay factor through simulations. We select the components of the initial opinion profile $x(0)$ uniformly at random from the interval $[0, 10]$. We then run the opinion formation process with stochastic interactions and plot the logarithm of the variance of the deviation from consensus as a function of time. After some time interval, this plot becomes linear, indicating that the largest eigenvalue dominates the decay. The slope of this line gives us an estimate of $\log \rho(\mathcal{W})$. All computations were done using MATLAB.

The results of our simulations and computations are shown in Figure 1. The first point to note is that, for both networks, the consensus time is dramatically larger when interaction

Table 1: The decay factor and consensus time for two social networks, where interaction probabilities are drawn uniformly at random from the specified intervals. The table presents the values computed for both the stochastic system of the form (3) and the deterministic system of the form (6) that approximates the stochastic system.

Network	Probability of Communication	Stochastic System		Deterministic System	
		Decay Factor	**Consensus Time**	**Decay Factor**	**Consensus Time**
Arxiv GR-QC	$p_{ij} = 1$	n/a	n/a	0.9966	298
Collaboration Network	$p_{ij} \in [.25, 0.75]$	0.9989	883	0.9989	879
(4158 nodes, 13428 edges)	$p_{ij} \in [.01, .1]$	0.999853	6821	0.99987	7774
Facebook	$p_{ij} = 1$	n/a	n/a	0.99097	111
Monterey Bay	$p_{ij} \in [0.25, 0.75]$	0.996855	318	0.997001	330
(6115 nodes, 31374 edges)	$p_{ij} \in [0.01, 0.1]$	0.99977	4230	0.999834	6026

probabilities are low versus when every agent communications in each round. This is evident in both the deterministic and stochastic systems. These results emphasize the empirical and theoretical observations that interaction frequency has a large impact on the efficiency of the agreement process.

We also note that the deterministic system provides accurate estimates for the efficiency measures of the stochastic system. The estimates of consensus times are closer to those of the stochastic system when interaction probabilities are in the interval $[0.25, 0.75]$ than when the probabilities are in the interval $[0.01, 0.1]$. This discrepancy may be due in some part to the fact that computations for the systems with smaller probability values are more sensitive to rounding errors.

Finally, we observe that the the Facebook network exhibits greater efficiency than the collaboration network relative to the number of nodes. The Facebook network also has a greater edge to node ratio than the collaboration network, which is some indication that this network is "more connected". However, the notion of connectivity is not well defined in this context. While it depends in some part on the number of edges, there are other graph attributes that must be considered, including network diameter and node degree distribution. We are currently investigating ways to analytically characterize complex networks like social networks and the effects of these network characteristics on the opinion dynamics.

5. CONCLUSION

In this work, we have extended the model of opinion formation in social networks proposed by De Groot and Lehrer to accommodate frequency of user interactions. We have shown that, for the purposes of analysis and design, the stochastic opinion formation process, where interactions occur with some probability, can be accurately approximated by a deterministic system where edge weights are adjusted for the probability of interaction. We have also presented simulations that demonstrate the effects of different interaction frequencies and the accuracy of the deterministic model that approximates the interaction-driven stochastic system.

While the classical model of consensus is very simple, analysis of this model can provide insight in to how to incorporate interactions into other models of information flow in online social networks. In future work, we plan to extend our mathematical analysis to the opinion dynamics model proposed by Friedkin and Johnsen, [9]. The structure of the Friedkin/Johnsen model is similar enough to the classical model that many of the techniques used in this paper can be applied. Unlike in the classical model, the initial profile plays an important role in determining the final opinions, and one can change the consensus value by altering the initial opinions of a few important individuals. This opens up the possibility of more powerful strategic network modification applications.

6. ACKNOWLEDGEMENTS

This work was funded in part by NSF grant ECCS-0802008.

7. REFERENCES

[1] arXiv.org e-Print Archive, `http://arXiv.org`.

[2] S. Boyd, P. Diaconis, and L. Xiao. Fastest mixing markov chain on a graph. *SIAM Review*, 46(4):667–689, December 2004.

[3] S. Boyd, L. El Ghauoi, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, 1994.

[4] M. C. De Gennaro and A. Jadbabaie. Decentralized control of connectivity for multi-agent systems. In *Proc. 45th IEEE Conference on Decision and Control*, pages 3628–3633, 2006.

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

[6] F. Fagnani and S. Zampieri. Randomized consensus algorithms over large scale networks. *IEEE Journal on Selected Areas in Communications*, 26(4):634–649, 2008.

[7] J. R. P. French. A formal theory of social power. *Psychological Review*, 63:181–194, 1956.

[8] N. E. Friedkin. *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, chapter Social Influence Network Theory: Toward a Science of Strategic Modification of Interpersonal Influence Systems, pages 89–100. The National Academies Press, Washington D.C., 2003.

[9] N. E. Friedkin and E. C. Johnsen. Social influence networks and opinion change. *Advances in Group*

Processes, 16:1–29, 1999.

[10] F. R. Gantmacher. *Applications of the Theory of Matrices*. Interscience, New York, 1959.

[11] A. Ghosh and S. Boyd. Growing well-connected graphs. In *Proc. 45th IEEE Conference on Decision and Control*, pages 6605–6611, 2006.

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

[13] S. L. Jamison. Perturbation of normal operators. *Proceedings of the American Mathematical Society*, 5(1):103–110, 1954.

[14] G. Kossinets, J. Kleinberg, and D. Watts. The structure of information pathways in a social communication network. In *Proc. 14th ACM International Conference on Knowledge Discovery and Data Mining*, pages 435–443, 2008.

[15] K. Lehrer. Social consensus and rational agnoiology. *Synthese*, 31(1):141–160, June 1975.

[16] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1), 2007.

[17] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50:169–182, February 2005.

[18] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. In *Proceedings of the IEEE*, pages 215–233, 2007.

[19] S. Patterson, B. Bamieh, and A. El Abbadi. Convergence analysis of scalable gossip protocols. In *Technical Report 2006-09*, July 2006.

[20] S. Patterson, B. Bamieh, and A. El Abbadi. Convergence rates of distributed average consensus with stochastic link failures. *IEEE Transactions on Automatic Control*, 55(4):880–892, April 2010.

[21] M. Porfiri and D. J. Stilwell. Consensus seeking over random weighted directed graphs. *IEEE Transactions on Automatic Control*, 52(9):1767–1773, September 2007.

[22] X. Song, Y. Chi, K. Hino, and B. L. Tseng. Information flow modeling based on diffusion rate for prediction and ranking. In *Proc. 16th International Conference on World Wide Web*, pages 191–200, 2007.

[23] A. Tahbaz-Salehi and A. Jadbabaie. A necessary and sufficient condition for consensus over random networks. *IEEE Transactions on Automatic Control*, 53(3):791–795, April 2008.

[24] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in Facebook. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, pages 37–42, 2009.

[25] S. Wasserman and K. Faust. *Social Network Analysis*. Cambridge University Press, 1994.

[26] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *Proc. 4th ACM European Conference on Computer Systems*, pages 205–218, 2009.

[27] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78, 2004.

[28] L. Xiao, S. Boyd, and S.-J. Kim. Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, 67(1):33–46, 2007.