Manufacturing Consent*

Vivek S. Borkar, *Fellow, IEEE,* Aditya Karnik, *Member, IEEE,* Jayakrishnan Nair, and Sanketh Nalli, *Student Member, IEEE*

***Abstract*—We consider a variant of the gossip algorithm wherein a controller can influence the values at a subset of nodes as well as the averaging weights. This leads to three optimization problems: (i) *Optimal choice of nodes:* a combinatorial optimization problem for which we propose a nonlinear programming relaxation as well as a greedy heuristic; (ii) *Parametric optimization of weights:* a non-convex optimization for which we propose an effective heuristic for a special case; (iii) *Dynamic adjustment of weights:* an optimal control problem. For the dynamic case, we note some empirically observed interesting critical phenomena for the uncontrolled case.**

***Index Terms*—Opinion dynamics, Gossip algorithms, Optimal node placement, Optimal control**

I. Introduction

QUESTIONS such as how individuals form opinions, when does consensus occur, why is there a polarization in opinions, and so on, have interested sociologists for past 4-5 decades. It was recognized early on that the process of opinion formation is a complex process where interpersonal influences play a critical role. Realistic modeling of these influences and their impact on opinion changes turns out to be a major challenge. Beginning with French [4] and followed by DeGroot [5] and Friedkin and Johnsen [6], one line of thinking has matured into what is now referred to as the social influence network theory. Representing opinions by real numbers, it postulates that individuals revise their opinions through weighted averaging of influences on them. Thus opinions evolve from a simple deterministic recursive process as follows.

$$\hat{x}_{n+1}(i) = (1-a)\hat{x}_n(i) + a\sum_j p(i,j)\hat{x}_n(j). \tag{1}$$

Here $x_n(i)$ denotes the opinion of individual i in time period n, and $p(i,j)$, the (i,j)th element of a stochastic matrix P, is the weight individual i assigns to the opinion of individual j. Finally, $0 < a < 1$ is a learning parameter that captures each individual's susceptibility to interpersonal influence. In this model, if the social network is strongly connected (i.e., matrix $P = [[p(i,j)]]_{1\leq i,j\leq N}$ is irreducible), a standard result from Markov chain theory implies that opinions converge to a consensus.

In this paper, we take as a starting point that opinions of individuals / estimates of sensors change as per a recursive process. Specifically, we consider a stochastic variant of (1) wherein $p(i,j)$ represent polling probabilities by agent i. That is, at each time n, agent i randomly polls one of her neighbors $\xi_n(i)$, with neighbor j being picked with probability $p(i,j)$, and updates her opinion as

$$\hat{x}_{n+1}(i) = (1-a)\hat{x}_n(i) + a\hat{x}_n(\xi_n(i)). \tag{2}$$

We may view (2) as a distributed, asynchronous, variant of (1). This is clearly more realistic in decentralized settings such as social networks / sensor networks. Note that (1) describes the evolution of the mean values of the iterates generated via (2).

Our interest is in understanding how a controller can maximize the collective value of the opinions/estimates by forcing some of them to prescribed values and influencing the overall interaction of the rest, under some resource constraints. While there is a substantial body of work on convergence issues in opinion dynamics and gossip algorithms, this work on their control is the first of its kind to our knowledge. Specifically, our contributions are as follows. The foregoing suggests two distinct, albeit related, optimization problems:

1) *The problem of optimal choice of a subset of individuals whose values are to be frozen:* Here, we seek to identify the optimal subset of individuals in order to maximize the rate of convergence to consensus. This turns out to be a hard combinatorial problem. We provide (i) a nonlinear programming relaxation and (ii) a greedy heuristic for a related simpler problem. Via simulations, we demonstrate that the optimal set of individuals can be quite different from the set of top-ranked individuals according to standard centrality measures. Our simulations highlight that a good node selection scheme must capture the (non-trivial) joint influence of sets of individuals on the gossip dynamics of the network.
2) *Optimization of weights with which the agents average their peers' opinions:* We consider two scenarios, both addressing optimization of related objective functions, but over different 'action sets', the first favoring lower complexity over a stronger notion of optimality.
   a) *Parametric optimization of (probability) weights:* Here the tunable parameter(s) affecting the weights

*with apologies to Noam Chomsky

V. S. Borkar was with the Tata Institute of Fundamental Research at the time this work was done and now is with the Dept. of Electrical Engineering, Indian Institute of Technology Bombay (IITB), India. E-mail: borkar.vs@gmail.com. Aditya Karnik was with Global GM R&D India Science Lab, Bangalore, India, and is now with General Electric Global Research Center, Bangalore, India. Jayakrishnan Nair was at the California Institute of Technology, USA, and then at Centrum Wiskunde & Informatica, The Netherlands, when this work was done. He is now with the Dept. of Electrical Engineering, IITB, India. Sanketh Nalli was with the Dept. of Information Technology, National Institute of Technology Karnataka, Surathkal, India, and is now with the Dept. of Computer Science, University of Wisconsin–Madison, USA.

Parts of this work have been presented at the Allerton conference [1], [2] and MIT workshop on social networks [3]. The work of VSB, AK and JN was supported by a grant from Global GM R&D India Science Lab, Bangalore. VSB was also supported by a J. C. Bose Fellowship from the Govt. of India. The work of SN was supported by a summer fellowship from the Indian Science Academies. The work of SN and part of the work of JN was done while visiting TIFR. VSB thanks Prof. A. S. Vasudevamurthy for some useful pointers to the literature.

are to be chosen off-line once for all. This is a non-convex optimization in general, for which we provide an expression of the gradient. We then present a practically relevant special case of the model corresponding to targeted advertising in a social network. For this special case, we establish coordinatewise convexity of the objective, which suggests heuristic coordinate descent methods. Our simulations highlight that it is beneficial to focus advertising effort on a set of influential individuals in the network. As before, the selection of the influential individuals must account for their joint influence on the gossip dynamics of the network.

b) *Dynamic optimization of weights:* Here the weights are to be adapted dynamically with a view to optimize a prescribed performance measure. This makes it an optimal control problem. The parametric optimization of (a) would then correspond to constant controls. For a non-linear variant of (2), we provide a characterization of optimal control via a Hamilton-Jacobi equation and the associated verification theorem. In addition, we record some interesting computational experiments for the uncontrolled scenario as the weight assigned to peer pressure is varied. These exhibit a slew of interesting critical phenomena. We also consider an adversarial situation leading to a differential game.

Note that whereas the opinion dynamics is 'distributed' insofar as the agents update their opinions through local interactions alone as specified by the graph, our controller sits outside of this community (think of a 'social planner') and controls certain aspects of this dynamics with a view to maximize the overall opinion or rating of the community for whatever is being opined about. Thus the control is *centralized* in control theoretic parlance.

A more formal statement of these problems requires setting up some notation, hence is delegated to Section III below after reviewing a small sample of the enormous amount of related literature in Section II. The first problem, that of determining the subset of individuals whose opinions are to be fixed, is discussed in Section IV. In Section V we take up the problem of parametric optimization of opinion values and transition probabilities. A non-linear version of DeGroot's model and its optimal control are discussed in Section VI. We conclude in Section VII.

We conclude this section with some important remarks.

1) There is some empirical work to validate the social influence network theory, e.g., [6], [7] report experimental results for small groups. The linear model ignores the phenomenon of 'confirmation bias' whereby peer opinions are weighed according to their proximity with one's own. To our knowledge the first models which account for this are so called bounded confidence models; see [8]. In these models, the network topology essentially changes with opinions (one only communicates with those who have similar opinions). This could be one way to introduce confirmation bias in our nonlinear model. The linear model, however, does exhibit the 'persuasion bias' [9], [10], also of interest for gossip algorithms for engineering systems. A linear model also fails to capture effects such as herd behavior which involve 'phase transitions'. The nonlinear model we consider later does so.
2) Our analysis presupposes prior knowledge of the model parameters, which is not usually the case. This calls for data driven 'learning' schemes. We briefly mention this fact at the appropriate juncture, but once again this is a nontrivial research direction for the future.

II. Related Work

Various models of opinion dynamics have been proposed in the literature; see, e.g., social impact [11], bounded confidence [8], Bayesian learning [12]. Convergence under repeated averaging has also been addressed in many papers: [13] investigates consensus, [14], [9] investigate whether individuals will learn the true value of a state of nature using (1) if they have noisy versions of it to start with. 'Gossip' algorithms that arise in the context of estimation over a network of sensors and have a similar flavor, have been studied extensively in [15]. Local repeated averaging also appears in other contexts, e.g., see [16] for bird flocking.

While we consider a continuum of opinions, there are situations where the opinions are discrete valued. This has led to interacting particle system models, see, e.g., [17]. Our model can be viewed as a continuum approximation of discrete valued opinions. While it is an interesting problem to justify this rigorously, we do not pursue it here. Cooperative o.d.e.s leading to monotone dynamics play a key role in our analysis. A parallel development for discrete time random models using, e.g., the developments of [18] may be possible. We do not, however, pursue this here. There have also been other models of 'influential nodes', e.g., [19] which considers similar issues in the classical framework of epidemic models.

III. Problem Formulation

Consider a social network with N nodes (agents) $\mathcal{S} = \{1, 2, \cdots, N\}$, with an associated controlled transition probability matrix $P^u = [[p(j|i, u(i))]]_{1\leq i,j\leq N}$, assumed to be irreducible for each fixed u. Here $u(i) \in U_i$ is a tunable parameter for each i (U_i being a compact metric space) and $u := [u(1), \cdots, u(N)]$. At each time instant n, each agent i samples one of the other agents, say j, with probability $p(j|i, u(i))$. Let $\xi_n(i)$ denote the (random) identity of the agent polled. Agent i then replaces her current estimate $\hat{x}_n(i)$ by

$$\hat{x}_{n+1}(i) := (1 - a)\hat{x}_n(i) + a\hat{x}_n(\xi_n(i)), \qquad (2)$$

where $a > 0$ is a small 'learning parameter'. Let $\hat{x}_n = [\hat{x}_n(1), \cdots, \hat{x}_n(N)]^T$. We are interested in the case when a 'controller' fixes the values of $\hat{x}_n(i)$ at some of the nodes (say, $\{m + 1, \cdots, N\}$ for some $m \leq N$) at some fixed values $\nu(i), m < i \leq N$, as well as the control parameters $\{u(i)\}$. The dynamics (2) can then be re-cast as

$$x_{n+1} := (1 - a)x_n + a\left(P_1^u x_n + P_2^u \nu + M_{n+1}\right), \qquad (3)$$

where,

- $x_n \in \mathcal{R}^m$ is the vector of the first m components of $\hat{x}_n$,
- $P_1^u :=$ the submatrix formed by the first m rows and columns of P^u,
- $P_2^u :=$ the submatrix formed by the first m rows and the last $N-m$ columns of P^u,
- $\nu = [\nu(m+1), \cdots, \nu(N)]^T$,
- $M_{n+1} := [\hat{x}_n(\xi_n(1)), \cdots, \hat{x}_n(\xi_n(m))]^T - (P_1^u x_n + P_2^u \nu)$ is a martingale difference sequence.

The idea is that the controller is trying to maximize the collective value of the estimates of the agents under some resource constraints that we shall soon specify. There are three sets of parameters she can optimize over:

1) The subset A of nodes whose values are to be fixed (e.g., influential people in a social network from the standpoint of an advertiser).
2) The actual numbers $\{\nu(i)\}$ at which these are fixed (e.g., budget for sponsored websites, celebrity endorsements, etc.).
3) The transition probabilities which depend on the tunable parameters $u(i)$ (e.g., the amount the advertising effort directed on an individual).[1]

There is some cost associated with any of these choices, making it a legitimate optimization problem that captures a trade-off between cost and payoff. The first of these problems, viz., optimization over subsets A of a fixed cardinality, turns out to be a hard combinatorial optimization problem, so a nonlinear programming relaxation and a greedy heuristic for a simplified version of the problem with provable guarantees are presented. Next we consider the subset A as given and address the second problem, with two different formulations that differ in our handling of the variables $u(i)$. The first considers them as parameters that are chosen once for all, thereby making it a parametric optimization problem. The second allows for dynamic state-dependent choices, i.e., an optimal control problem, which we study for a more general nonlinear variant.

Consider fixed u, ν. Note that (3) can be viewed as a constant stepsize stochastic approximation (see [20, Chapter 9]) which tracks the asymptotic behavior of the o.d.e.

$$\dot{x}(t) = P_1^u x(t) - x(t) + P_2^u \nu \tag{4}$$

in the following sense. Since P_1^u is substochastic with its Perron-Frobenius eigenvalue < 1 under our hypotheses, this o.d.e. has a unique asymptotically stable equilibrium

$$x^* = (I - P_1^u)^{-1} P_2^u \nu. \tag{5}$$

Then

$$\limsup_{n\uparrow\infty} E\left[\|x_n - x^*\|^2\right] = O(a). \tag{6}$$

(See Theorem 3, p. 106, [20].) What (6) says is that the iterates (3) asymptotically concentrate to an $O(a)$ neighborhood around x^* with high probability.[2] In Sections IV and V, we shall refer to x^* loosely as the equilibrium corresponding to our iterates (3).

IV. Choice of Nodes

Consider a scenario where K out of the N nodes, comprising a subset $A \subset \mathcal{S}$, fix their value of the corresponding components of $\hat{x}_n$ for good. Let P_A denote the principal submatrix of P corresponding to nodes in A^c and $\bar{P}_A$ the submatrix with row indices corresponding to A^c and column indices corresponding to A. These are precisely the P_1^u, P_2^u above, tagged by the set A so as to make their dependence on it explicit, and *sans* the control parameter u which does not play any role in the developments of this section. Let I and $\mathbf{1}$ denote respectively the identity matrix and the vector of all 1's (with appropriate dimensions depending on the context).

Note that the equilibrium $x^* = (I - P_A)^{-1}\bar{P}_A\nu$. In the special case $\nu(\cdot) \equiv c$, $x^* = c\mathbf{1}$, i.e., there is asymptotic consensus on the value c. To see this, note that x^* has the well known representation ([21], Ch. IV)

$$x^*(i) = E[\nu(X_\tau)|X_0 = i], i \in A^c, \tag{7}$$

for a Markov chain $\{X_n\}$ with transition matrix P and $\tau := \min\{n \geq 0 : X_n \in A\}$.[3]

The rate of convergence of the iterates x_n to (a neighborhood of) x^* is dictated by the Perron-Frobenius eigenvalue $\lambda(P_A)$ of P_A. Indeed, defining $e_n := x_n - x^*$, (3) can be rewritten as

$$\begin{aligned} e_n &= (aP_A + (1-a)I)e_{n-1} + aM_n \\ &= (aP_A + (1-a)I)^n e_0 + a\sum_{i=1}^{n}(aP_A + (1-a)I)^{n-i}M_i \\ &= T_1 + T_2. \end{aligned}$$

The term T_1 above is $\Theta((a\lambda(P_A) + (1-a))^n)$. Since $\lambda(P_A) < 1$, $a\lambda(P_A) + (1-a) < 1$, implying that T_1 vanishes exponentially fast. The term T_2, which captures the 'persistent' noise, has mean zero, and satisfies $E[\|T_2\|^2] = O(a)$ (see (6)). Thus, the iterates converge 'with high probability' to an $O(a)$ neighborhood of x^* exponentially fast, at rate $a\lambda(P_A)+(1-a)$. Thus, minimizing $\lambda(P_A)$ corresponds to maximizing the rate of convergence of the iterates to (a neighborhood of) the equilibrium x^*. This is the focus of the present section.

Formally, we are interested in the following optimization problem.

$$\min_{A\subset\mathcal{S}, |A|=K} \lambda(P_A) \tag{8}$$

[1] Our model allows for a general functional dependence of the transition probabilities on the control parameters $u(i)$. In Section V, we consider a special case of this dependence that applies to targeted advertising.

[2] This is because in this so called 'constant stepsize' scenario (i.e., a is a constant independent of n, as opposed to a slowly vanishing sequence), the noise perturbation is persistent and any kind of convergence to an equilibrium is untenable. Thus the best we can hope for is a concentration of asymptotic probabilities near the equilibrium. This is plausible if the noise is 'small', which translates into a being small. The aforementioned estimate quantifies this fact. See *ibid.*, Chapter 9 for details.

[3] Suppose that we obtain a reward when the Markov chain $\{X_n\}$ first hits the set A, the reward being $\nu(j)$ if the hitting state is $j \in A$. Then for $i \in A^c$, $r(i) := E[\nu(X_\tau)|X_0 = i]$ is the average reward obtained starting at state i. The Markov property implies that the vector $r = (r(i),\ i \in A^c)$ satisfies the recursion $r = P_A r + \bar{P}_A \nu$, which implies that $r = x^*$.

The minimizing set A will therefore correspond to the K most important nodes to influence from the point of view of rapid opinion dissemination. This is a distinct notion of rating nodes as compared to rating schemes such as Google's PageRank [22] or the 'hub and authority' model of Kleinberg [23], or the various centrality measures proposed in social network research [24].

Our discrete optimization problem (8) is a hard combinatorial problem to solve exactly. In the following section, we consider a continuous relaxation of (8), which is more tractable.

A. Continuous relaxation

We now consider a relaxation of (8) to an optimization problem over reals. Let $P'(\theta) := P\Theta$ where Θ is a diagonal matrix with entries $\theta_1, \cdots, \theta_N \in [0,1]$ on its diagonal. Let $\theta := [\theta_1, \cdots, \theta_N]^T$. We impose the constraint $\sum_i \theta_i = N-K$. Our relaxation of (8) is

$$\min_{\theta \in C_1 \cap C_2} \lambda(\theta), \tag{9}$$

where $\lambda(\theta) = \lambda(P'(\theta))$ denotes the the Perron-Frobenius eigenvalue of $P'(\theta)$, $C_1 := [0,1]^N$, and $C_2 := \{x = [x_1, \cdots, x_N]^T \in \mathcal{R}^N : \sum_i x_i = N-K\}$. Note that a $\{0,1\}$-valued θ will correspond to exactly K zeros and $N-K$ ones, thereby recovering our original formulation. Let $\pi, V \in \mathcal{R}^N$ denote respectively the Perron-Frobenius left- and right- eigenvectors of $P'(\theta)$, i.e., the eigenvectors corresponding to the eigenvalue $\lambda(\theta)$, where we suppress the θ-dependence of π, V for simplicity. That is, for $j \in \mathcal{S}$,

$$\lambda(\theta)V(j) = \sum_k p(j,k)\theta_k V(k), \tag{10}$$

$$\lambda(\theta)\pi(j) = \sum_k p(k,j)\theta_j \pi(k). \tag{11}$$

Differentiating (10) w.r.t. $\theta_i, 1 \le i \le N$, we have

$$\frac{\partial\lambda(\theta)}{\partial\theta_i}V(j)+\lambda(\theta)\frac{\partial V(j)}{\partial\theta_i} = p(j,i)V(i)+\sum_k p(j,k)\theta_k\frac{\partial V(k)}{\partial\theta_i}.$$

Multiplying the above equation by $\pi(j)$, summing over j, and using (11), we get

$$\frac{\partial\lambda(\theta)}{\partial\theta_i} = \frac{V(i)\pi^T p(\cdot,i)}{\pi^T V} = \frac{\lambda(\theta)V(i)\pi(i)}{\theta_i(\pi^T V)},$$

which gives an explicit expression for the gradient of $\lambda(\theta)$ w.r.t. θ. The rightmost expression gives some qualitative insight into parametric dependence of $\lambda(\cdot)$. It suggests that $\lambda(\cdot)$ increases exponentially with respect to the normalized product of left and right eigenvectors at each state i for a given value of θ. This suggests a measure of sensitivity of $\lambda(\cdot)$ vis-a-vis the states. We consider the following projected gradient scheme. Let $\eta > 0$ denote a prescribed step size. Then

$$\theta_{n+1} = \Gamma\left(\theta_n - \eta\frac{\text{diag}(V)P^T\pi}{\pi^T V}\right), \tag{12}$$

where $\Gamma(\cdot)$ is the projection operator onto the set $C_1 \cap C_2$: $\Gamma(x) = \text{argmin}_{y \in C_1 \cap C_2}\|x-y\|$. The projection $\Gamma(x)$ may be computed using the 'successive projections' algorithm due to Boyle, Dykstra and Han [25], [26]. An alternative algorithm for computing $\Gamma(x)$ that exploits the structure of the sets C_1 and C_2 can be found in [27].

The objective function of (9) is not in general convex and hence we can only expect the gradient projection scheme (12) to converge to a local minimum. This suggests resorting to multi-start, simulated annealing, etc. to improve performance. From a candidate solution θ^* to (9), we may obtain a candidate solution to (8) by picking the nodes corresponding to the smallest K components of θ^*. There may be local minima that are not global minima, either in the interior or on the boundary.

B. A greedy algorithm

In view of these difficulties, we consider an alternative formulation. Let $\{X_n\}$ be a Markov chain with transition matrix P. Then $\log(\lambda(P_A))$ has the interpretation of being the asymptotic rate of exit from A^c, i.e., the 'rate of exponential decay' of the tail probability of τ: $\log(\lambda(P_A)) = \lim_{t\uparrow\infty}\frac{\log P(\tau>t)}{t}$ for $\tau := \min\{n \ge 0 : X_n \in A\}$. This suggests looking at a related, more amenable performance measure, the mean exit time $E[\tau]$. Recall that $E[\tau] = \sum_t P(\tau \ge t)$, whereas in light of the above, $P(\tau > t) \approx \lambda(P_A)^t$. Thus we expect the minimization of $\lambda(P_A)$ and $E[\tau]$ to be 'similar'. Assuming uniform initial distribution over A^c, we have the cost criterion

$$E[\tau] = \frac{1}{N-K}\mathbf{1}^T(I-P_A)^{-1}\mathbf{1} = \frac{1}{N-K}\mathbf{1}^T\left(\sum_{n=0}^{\infty}{P_A}^n\right)\mathbf{1}.$$

We now state our optimization objective formally. For $A \subseteq \mathcal{S}$, define $f(A) := \mathbf{1}^T\left(\sum_{n=0}^{\infty}{P_A}^n\right)\mathbf{1}$. Note that $f(\emptyset) = \infty$, and $f(A) < \infty$ for all non-empty $A \subseteq \mathcal{S}$. We seek to solve the following optimization problem:

$$\min_{A \subseteq \mathcal{S}, |A| = K} f(A) \tag{13}$$

Problem (13) is still a hard problem to solve exactly. However, the following result facilitates a greedy heuristic [28].

Theorem 1: f is supermodular, i.e., for $A, B \subseteq \mathcal{S}$,

$$f(A) + f(B) \le f(A \cup B) + f(A \cap B).$$

Proof: For $A \subseteq \mathcal{S}$, define $f^{(n)}(A) = \mathbf{1}^T P_A^n \mathbf{1}$. We will prove that $f^{(n)}(\cdot)$ is supermodular. It is easy to see that this implies supermodularity of f.

Define $g : [0,1]^N \to \mathbf{R}$ as $g(\theta) = \mathbf{1}^T\left(\text{diag}(\theta)\ P\ \text{diag}(\theta)\right)^n\mathbf{1}$. Define

$$\hat{\theta}(A) := (\mathrm{I}_{\{i \notin A\}}, i = 1, 2, \cdots, N),$$

where $\mathrm{I}_{\{z\}}$ equals 1 if z is true and 0 otherwise. Note that $f^{(n)}(A) = g(\hat{\theta}(A))$. It is easy to see that $\frac{\partial^2 g(\theta)}{\partial\theta_i\partial\theta_j} \ge 0$ for all $i \neq j$. This implies that g is supermodular [29, Theorem 10.4], i.e., g satisfies $g(\theta) + g(\tilde{\theta}) \le g(\theta \vee \tilde{\theta}) + g(\theta \wedge \tilde{\theta})$ for all $\theta, \tilde{\theta} \in [0,1]^N$. For $A, B \subseteq \mathcal{S}$,

$$\begin{aligned} f^{(n)}(A) + f^{(n)}(B) &= g(\hat{\theta}(A)) + g(\hat{\theta}(B)) \\ &\le g(\hat{\theta}(A) \vee \hat{\theta}(B)) + g(\hat{\theta}(A) \wedge \hat{\theta}(B)) \\ &= g(\hat{\theta}(A \cap B)) + g(\hat{\theta}(A \cup B)) \\ &= f^{(n)}(A \cap B) + f^{(n)}(A \cup B). \end{aligned}$$

This proves that $f^{(n)}(\cdot)$ is supermodular. ■

Note that the specific choice of initial distribution (uniform in this case) is not crucial here.

1) Greedy Algorithm: Note that the objective function f is monotone non-increasing, i.e., $f(A) \geq f(B)$ whenever $A \subseteq B$. The supermodularity and monotonicity of f (equivalently, the submodularity and monotonicity of $-f$) motivates the following simple greedy heuristic to compute an approximate solution A^G to (13) (see [28]).

1) Set $A_0 = \emptyset$.
2) For $i = 1, 2, \cdots, K$ do

$$\begin{aligned} j_i^* &= \text{argmin}_{j \in A_{i-1}^C} f(A_{i-1} \cup \{j\}), \\ A_i &= A_{i-1} \cup \{j_i^*\}. \end{aligned}$$

3) Set $A^G = A_K$.

The algorithm constructs the set A^G in K stages. In each stage, the node that produces the greatest marginal decrease in the objective function is added to the set.

The greedy algorithm described above involves several evaluations of the function f. For non-empty $A \subset \mathcal{S}$, the computation of $f(A)$ involves the inversion of the typically 'large' matrix $(I - P_A)$. However, these inversions can be greatly simplified using the Sherman-Morrison-Woodbury (SMW) formula (see [30, Section 2.1.3]). In Stage $i \geq 2$ of the algorithm, the SMW formula can be used to compute $(I - P_{A_{i-1} \cup \{j\}})^{-1}$ efficiently from $(I - P_{A_{i-1}})^{-1}$.

2) Suboptimality bound: Since $f(\emptyset) = \infty$, one cannot bound the suboptimality of $f(A^G)$ relative to the optimum value f^* of (13) directly, as in Theorem 4.2 of [28]. We can, however make the following weaker statement. Suppose that we restrict the optimization in (13) to sets A containing a special node i_0; the 'initial node'. Note that the choice of i_0 is also a part of the decision and we assume that it has been arrived at by a suitable heuristic, e.g., one of the standard centrality measures. Let $f_{i_0}^*$ denote the optimum value of this relaxed problem. Consider the following modification to the greedy algorithm above: start with $A_1 = \{i_0\}$, and run the iterations for $i = 2, \cdots, K$ to obtain the set $A^G(i_0)$. Then

$$f(\{i_0\}) - f(A^G(i_0)) \geq \left(1 - \frac{1}{e}\right)(f(\{i_0\}) - f_{i_0}^*),$$

where e denotes the base of the natural logarithm (see Theorem 4.2 of [28]).

C. Experiments

In this section, we evaluate the algorithms proposed in the preceding sections numerically. We illustrate the value of these algorithms by comparing their performance against naive node selection schemes based on two popular centrality notions in the literature.

We consider both synthetic as well as real world datasets. In each case, the network data provides us with an undirected graph, represented by its adjacency matrix $R = [[r(i,j)]]_{1 \leq i,j \leq N}$. From R, we generate the stochastic matrix P by setting $p(i,j) = \frac{r(i,j)}{\deg(i)}$, where $\deg(i)$ denotes the degree of node i. We compare our node selection algorithms with the following simple selection rules: rank the nodes according to either the PageRank algorithm[4] [22] or the HITS algorithm[5] [23], and pick the K highest ranked nodes. Recall that our performance metric is $\lambda(P_A)$, which determines the rate of convergence to consensus, once the subset A of nodes freezes its value.

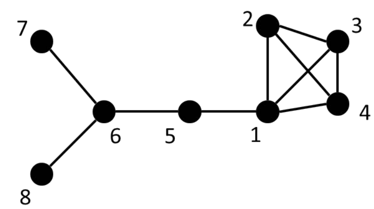

Fig. 1. Toy example with 8 nodes

To provide intuition, we start with the toy network depicted in Fig. 1 with 8 nodes. We set $K = 2$. In this case, the greedy algorithm presented in Section IV-B obtains the optimal solution $\{1, 6\}$, with $\lambda(P_{\{1,6\}}) = 0.873$. On the other hand, both the naive approaches described above pick the subset $\{1, 2\}$, with $\lambda(P_{\{1,2\}}) = 0.974$. Now it is intuitively clear that Node 1 is the most central node in our example. Our naive schemes pick Node 2 as the second node (its ranking under PageRank/HITS is second highest, due in part to its proximity to Node 1). Indeed, it turns out that $\lambda(P_{\{2\}}) < \lambda(P_{\{6\}})$, which suggests that on a stand-alone basis, Node 2 is actually more influential than Node 6. However, the naive schemes fail to take into account the joint influence of groups of nodes. Indeed, the node pair $\{1, 6\}$ is actually more influential than the pair $\{1, 2\}$. Intuitively, this is because the convergence to consensus throughout the network is aided by spreading out the nodes with frozen values in the network.

Next, we now move on to bigger networks. We consider the following three examples.

1) *Zachary Karate Club*: This is a well known network representing friendships between 34 members of a karate club over a period of two years [31]. Here, we take $K = 5$.
2) *Coauthorships in Network Science*: A data set describing a collaboration network of scientists working in network theory and experiments has been prepared by Newman [32]. In this network graph, nodes are scientists and two scientists are connected by an (undirected) edge if they have co-authored a paper. For our experiment, we use the largest connected component of this graph, which contains 379 nodes. Here, we take $K = 20$.
3) *Synthetic network generated via preferential attachment*: We generate a 1000 node network using the well known preferential attachment scheme [33]. Specifically, the graph is generated by adding nodes one at a time, each incoming node attaching itself through an undirected

[4]This corresponds to computing the stationary distribution corresponding to the Markov chain with transition matrix P. The nodes (states) are ranked in decreasing order of their stationary probabilites.

[5]Since the adjacency matrix R in our examples is symmetric, the hub and authority ratings for a node are equal; the vector of hub/authority ratings of the nodes is the Perron-Frobenius eigenvector of R^2.

	Greedy algorithm	*Gradient descent*	*PageRank*	*HITS*
Zachary Karate Club	0.908	0.930	0.930	0.930
Net. Sci. coauthorships	0.989	0.995	0.996	0.999
Preferential attachment network	0.987	-	0.995	0.999

TABLE I
EACH TABLE ENTRY GIVES THE PERRON-FROBENIUS EIGENVALUE OF P_A

edge to an already existing node in the network. With probability 1/2, this neighbor node is picked uniformly at random, and with probability 1/2, the neighbor is picked proportionally to the degree of the existing nodes. For this network, we take $K = 80$.

In Table I, we record the value of $\lambda(P_A)$ obtained for the above networks by selecting the subset A according to the greedy algorithm presented in Section IV-B, the gradient descent algorithm described in Section IV-A[6], and the two naive schemes based on PageRank and HITS. Note that the greedy algorithm produces the lowest value of $\lambda(P_A)$, followed by the projected gradient descent algorithm (except in the 1000 node preferential attachment network; in this case, the gradient descent algorithm takes a prohibitively long time to converge).

In conclusion, we see that our algorithms, particularly the greedy algorithm, consistently outperform naive node selection methods. Moreover, our experiments confirm that our notion of ranking (subsets of) nodes from the point of view of rapid opinion dissemination is very distinct from other popular centrality notions in the literature.

V. PARAMETRIC OPTIMIZATION

We now consider the parametric optimization problem of determining (i) the values at which the opinions of $N - m$ nodes are to be fixed ($\{\nu(i)\}$, $m < i \leq N$) and (ii) parameters $u(i)$, $1 \leq i \leq m$ on which the transition probabilities depend. The payoff – cost structure we consider is the following. At each instant n, we incur a payoff of $\sum_i x_n(i)$, i.e., the net interest level. For each component i, a cost of $g_i(u(i))$ is incurred if the parameter $u(i)$ is chosen, where $g_i \in C(U_i; \mathcal{R})$. In addition, a cost of $h_j(\nu(j)), h_j \in C(\mathcal{R}^+; \mathcal{R}^+)$ is incurred for freezing the opinion of node j. Let $g(u) := [g_1(u(1)), \cdots, g_m(u(m))]^T$ and $h(\nu) := [h_{m+1}(\nu(m+1)), \cdots, h_N(\nu(N))]^T$. This leads to the optimization problem:

$$\begin{aligned}&\text{Maximize } \ \psi(u,\nu) := \mathbf{1}^T x - (\mathbf{1}^T g(u) + \mathbf{1}^T h(\nu)) \\ &\quad \text{over } (u,\nu), \text{ subject to (5).}\end{aligned} \qquad (\dagger)$$

Note that the term in parentheses is the cost associated with parameters u and ν, paid once for all in the beginning when these parameters are frozen, whereas the reward in the first term is a time average.

A. Gradient based schemes

Suppose $u(i)$ take values in some compact subset(s) of a Euclidean space. Assuming continuous differentiability of $g_i(\cdot), p(j|i,\cdot)$, $h_j(\cdot)$, one can use, say, the gradient projection method for maximization over u, ν.

We derive an expression for the gradient for the simple case of $u(i) \in [0,1]$, the general case being similar albeit notationally messy. From (5), we have:

$$\frac{\partial x}{\partial u(i)} = P_1^u \frac{\partial x}{\partial u(i)} + \frac{\partial P_1^u}{\partial u(i)} x + \frac{\partial P_2^u}{\partial u(i)} \nu.$$

Therefore,

$$\begin{aligned}\frac{\partial x}{\partial u(i)} &= (I - P_1^u)^{-1}[0, \cdots, 0, \sum_{j=1}^{m} \frac{\partial p(j|i,u(i))}{\partial u(i)} x(j) \\ &\quad + \sum_{j=m+1}^{N} \frac{\partial p(j|i,u(i))}{\partial u(i)} \nu(j), 0, \cdots, 0]^T,\end{aligned}$$

where on the right hand side the summation is in the ith place. The ith component of the overall gradient then is

$$\frac{\partial \psi}{\partial u(i)} = \sum_{j=1}^{m} \frac{\partial x(j)}{\partial u(i)} - \frac{\partial}{\partial u(i)} g_i(u(i))$$

The gradient with respect to ν is also easy to derive:

$$\frac{\partial \psi}{\partial \nu(j)} = \mathbf{1}^T (I - P_1^u)^{-1} P_2^u e_{(j-m)} - \ h_j'(\nu(j)),$$

where $e_{(j-m)}$ is a column vector with 1 as its $(j-m)$th entry and zero elsewhere.

While this allows us to apply the usual optimization schemes such as projected gradient descent, one is in general not guaranteed to obtain the global optimum. Even in the simple case of $p(\cdot|i,u(i))$ being affine in $u(i)$, $x(j)$ for $1 \leq j \leq m$ will be a rational function of $u(i)$, being the solution of a linear system. Thus typically it will be non-convex with multiple local maxima. One might therefore have to resort to multi-start, simulated annealing, etc.

In the following, we consider a special case of the above abstract model, in which the structure of P^u confers coordinatewise convexity of the objective, allowing for more effective algorithms.

B. A special case of the model

In this section, we consider a special case of our parametric optimization model, which corresponds to targeted advertising in a social network. The purpose of this presentation is twofold. Firstly, we demonstrate a practically relevant application of the abstract model presented above. Secondly, we exploit the special structure of the controlled transition matrix P^u for this case to propose effective algorithms.

Consider a social network with m nodes $\{1, 2, \cdots, m\}$, represented by an undirected, irreducible graph G. Opinion

[6]The step size η was set by trial and error. We ran the algorithm five times, with random initializations of θ; the result corresponding to the best of these runs is reported.

about a product evolves among the nodes of this graph as per a gossip algorithm of the form (3). To define the dynamics of this evolution, we first describe the uncontrolled scenario.

In addition to listening to their neighbors, the nodes also adapt their opinion based on an online product review forum. We model the influence of this review forum by appending a node labeled $m+1$ to our graph G, and adding edges to node $m+1$ from all other nodes. Let us suppose that over the timescale under consideration, this review forum (represented by node $m+1$) projects a constant opinion $\nu(m+1)$ of the product. Node i, $1 \leq i \leq m$, adapts its opinion as per (2), by picking $\xi_n(i)$ uniformly at random from among its neighbors (which includes the product review forum). Note that as per the above dynamics, the opinions of nodes 1 through m would tend to concentrate around the opinion of the review board, i.e., $\nu(m+1)$.

Now, the seller wishes to 'control' the above opinion dynamics using advertisements. We model the advertisements via a Node $m+2$, which projects a fixed, high opinion $\nu(m+2) > \nu(m+1)$ of the product. The control $u(i) \in [0,1]$ for node i operates as follows. At each time step, with probability $u(i)$, $\xi_n(i) = m+2$, i.e., Node i is influenced by the advertisement. With probability $1-u(i)$, $\xi_n(i)$ is picked as per the uncontrolled dynamics.

The above defines a special case of our abstract model presented earlier, with $N = m+2$. The entries of first m rows of the controlled transition matrix P^u in this case are as follows. For $1 \leq i \leq m$,

$$p(j|i,u(i)) = \begin{cases} (1-u(i))\frac{\iota(i,j)}{\deg(i)} & \text{for } 1 \leq j \leq m+1 \\ u(i) & \text{for } j = m+2 \end{cases},$$

where $\iota(i,j)$ equals 1 if there is an edge between i and j in G and zero otherwise, and $\deg(i)$ denotes the degree of node i.

Now, the seller experiences a cost $g_i(u(i))$ for exerting control $u(i)$, and a cost $h(\nu(m+2))$ for creating the advertisement. She seeks to maximize the objective

$$\psi(u,\nu(m+2)) := \mathbf{1}^T x - \sum_{i=1}^{n} g_i(u(i)) - h(\nu(m+2)),$$

where $x = (I-P_1^u)^{-1}P_2^u\nu$ is the equilibrium vector that the opinions in the network would concentrate around. Note that the above formulation allows for targeted advertising, i.e., the seller may tune her advertising effort on a certain node depending on its influence on other nodes in the network. In particular, the seller may selectively focus her advertising efforts on certain 'influential' nodes to maximize her payoff. Indeed, our numerical experiments below seek to demonstrate the benefit of such targeted advertising.

While ψ is in general not concave, the following lemma establishes that ψ is coordinatewise concave, if the functions $g_i(\cdot)$, $1 \leq i \leq m$ and $h(\cdot)$ are convex. This suggests that a local maximum of ψ may be obtained via coordinate descent algorithms that cyclically optimize with respect to each coordinate (see [34, Sec. 2.7]). Each of the coordinate optimizations, being convex programs, can be solved efficiently.

Lemma 2: For $1 \leq i \leq m$, if $g_i(\cdot)$ is convex, then ψ is a concave function of $u(i)$. Also, if $h(\cdot)$ is convex, then ψ is a concave function of $\nu(m+2)$.

Proof: Suppose that $g_i(\cdot)$ is convex. To show that ψ is concave with respect to $u(i)$, it suffices to show that for any j, $x(j)$ is concave with respect to $u(i)$. To show this, we utilize the following interpretation of $x(j)$ as the average terminal reward of an absorbing Markov chain. Specifically, consider a Markov chain $\{X_n\}$ with transition matrix P^u, and absorbing states $m+1$ and $m+2$. Let $\tau := \min\{n \geq 0 \mid X_n \in \{m+1, m+2\}\}$ denote the absorption time. As we have seen before,

$$\begin{aligned} x(j) &= E[\nu(X_\tau) \mid X_0 = j] \\ &= \mathbb{P}(X_\tau = m+1)\,\nu(m+1) \\ &\qquad + (1 - \mathbb{P}(X_\tau = m+1))\nu(m+2) \\ &= \nu(m+2) - (\nu(m+2) - \nu(m+1))\mathbb{P}(X_\tau = m+1)\,. \end{aligned}$$

Thus, to prove that $x(j)$ is concave with respect to $u(i)$, it suffices to prove that $\mathbb{P}(X_\tau = m+1)$ is convex with respect to $u(i)$. We do this as follows.

The 'controlled' dynamics P^u of our Markov chain can be related to the 'uncontrolled' dynamics $\tilde{P} := P^0$ as follows. Under the uncontrolled dynamics $\tilde{P}$, the Markov chain makes a random walk along the nodes $1, \cdots, m$ before being absorbed into the state $m+1$ with probability 1. Under the controlled dynamics P^u, at each time step n, the chain gets absorbed into state $m+2$ with probability $u(X_n)$, and with probability $1-u(X_n)$, makes a transition as per $\tilde{P}$. Thus, $\mathbb{P}(X_\tau = m+1)$ can be calculated by conditioning with respect to paths of the uncontrolled dynamics. Specifically, let $\mathcal{P}$ denote the set of all paths under the uncontrolled dynamics starting at node j and ending at node $m+1$. Then under the controlled dynamics,

$$\mathbb{P}(X_\tau = m+1) = \sum_{p \in \mathcal{P}} \mathbb{P}(p) \prod_{l \in p} (1-u(l)).$$

Note that $\mathbb{P}(p)$ in the above expression is computed according to the uncontrolled dynamics. Now, for any path p, $\prod_{l \in p}(1-u(l))$ is of the form $c(1-u(i))^d$, where c is independent of $u(i)$ and $d \in \mathbb{N}$. It therefore follows easily that $\prod_{l \in p}(1-u(l))$ is convex with respect to $u(i)$. It then follows that $\mathbb{P}(X_\tau = m+1)$ is convex with respect to $u(i)$, since limits of of convex functions are convex.

Finally, concavity with respect to $\nu(m+2)$ is immediate once we note that

$$\mathbf{1}^T x = \mathbf{1}^T (I-P_1^u)^{-1} P_2^u \nu$$

is a linear function of $\nu(m+2)$. ■

A straightforward deduction from the above proof argument is the following. Suppose that all the $u(i)$ are constrained to be equal to $\bar{u}$. This corresponds to a scenario in which the seller is unable to customize the advertising effort on each node in the social network. In this case, assuming that g_i are identical and convex, it follows using an argument similar to that in the above proof that $\psi(\bar{u}\mathbf{1}, \nu(m+2))$ is convex with respect to $\bar{u}$. In our numerical experiments below, we demonstrate the benefit of targeted advertising in a social network by comparing the seller's payoff to that under such 'uniform' advertising.

Numerical experiments: We now present some simple numerics for the model presented in this section. Our objective here is to demonstrate via a toy example that the optimal targeted advertising policy could be much more beneficial to the seller than the best 'uniform' advertising strategy.

We consider once again the network shown in Fig. 1. Thus, in this experiment, $m = 8$. To focus on the contrast between targeted and uniform advertising, we ignore the optimization of $\nu(10)$. We fix $\nu(10) = 10$, $\nu(9) = 5$. The control $u(i)$ is constrained to the set $[0, 0.2]$. Finally, we set $g_i(u) = c \cdot u$ for all $1 \leq i \leq 8$, where c is a positive scalar whose value is varied. We compute a (potentially sub-optimal) targeted advertising strategy via cyclic coordinate descent, and also the optimal uniform advertising strategy. (Note that the latter optimization is convex, based on the discussion above.) Our results are recorded in Table II.

c	Targeted control		Uniform control	
	u	OBJ	$\bar{u}$	OBJ
5	[0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2]	49.74	0.2	49.74
10	[0.2 0.2 0 0 0 0.2 0 0]	44.48	0.12	42.58
15	[0.2 0 0 0 0 0.2 0 0]	41.48	0.02	40.08
20	[0.2 0 0 0 0 0 0 0]	40.46	0	40

TABLE II
COMPARING TARGETED AND UNIFORM ADVERTISING

As expected, when the cost of advertising is sufficiently small, both strategies simply apply the maximum allowed control on all nodes. However, as the cost of control increases, the targeted advertising strategy outperforms the uniform strategy by focusing the advertising effort on a (progressively diminishing) set of 'key' nodes. Interestingly, these sets of 'key' nodes happen to be the same ones picked by our greedy algorithm in our numerical experiments in Section IV.

The above results suggest that a good advertising strategy on a social network would be to focus the advertising effort on a set of influential nodes. Moreover, as was noted before, these influential nodes should not be chosen independently, since it is important to take into account the joint influence of nodes on gossip dynamics in the network. This observation also has algorithmic consequences. Indeed, in a large network, a reasonable strategy might be to short-list a set of candidate influential nodes (perhaps using the algorithms of the preceding section), and then solving the (lower dimensional) optimization of advertising effort over these candidate nodes. While beyond the scope of the present paper, our results motivate larger scale numerical experiments on real world networks with realistic models of advertising costs.

So far, we have considered parametric optimization of the gossip dynamics (3). In these 'static' optimizations, all control parameters are fixed off-line once for all. In the following section, we consider the problem of controlling a non-linear variant of the (3) in the framework of dynamic control.

VI. A NONLINEAR MODEL

Consider a variation of (2) in which an individual agent takes into account what the peers say, but also pursues her own inclination. Specifically, the agent in question, say the ith out of N, holds an opinion $x_k(i) \in \mathcal{R}$ at time instant k and polls a peer j with probability $p(i, j)$, this being the (i, j)th element of an irreducible stochastic matrix P. Let $\xi_k(i)$ denote the (random) identity of the peer who has been polled. She then updates her opinion incrementally according to

$$x_{k+1}(i) = x_k(i) + \gamma[\alpha_i(\textstyle\sum_j I\{\xi_k(i) = j\}x_k(j) - x_k(i)) \\ +(1 - \alpha_i)f_i(x_k(i))]. \tag{14}$$

Here $0 < \alpha_i < 1$ is the weight she attaches to 'peer pressure' while attaching weight $(1 - \alpha_i)$ to her own 'inclination' $f_i(x_k(i))$, where the f_is are bounded Lipschitz. As an example of the latter, consider, e.g., $f_i := \frac{\partial g}{\partial x}$ where g represents a common 'payoff landscape' the agents share. Here $\gamma > 0$ is a stepsize ensuring the incremental nature of the learning process. We consider the case $f_i = f$, $\alpha_i = \alpha$, $\forall i$. Let $F(x_1, \cdots, x_N) := [f(x_1), \cdots, f(x_N)]^T$. Then one can view (14) as a constant stepsize stochastic approximation algorithm with the o.d.e. limit

$$\dot{x}(t) = \alpha(P - I)x(t) + (1 - \alpha)F(x(t)). \tag{15}$$

This is similar to the models of synchronization in natural systems [35]. Using the Hirsch theorem for cooperative o.d.e.s [36][7], one can show that (15) converges for generic initial data. We consider the case where the scalar o.d.e.

$$\dot{z}(t) = f(z(t)) \tag{16}$$

converges to one of finitely many equilibria for any initial condition. If $x^* \in \mathcal{R}$ is one such equilibrium, then $\hat{x}^* := [x^*, \cdots, x^*]^T$ is an equilibrium for (15). We shall call this a *homogeneous* equilibrium of (15). What's more, if x^* is a stable equilibrium for (16), then $\hat{x}^*$ is a stable equilibrium for (15) – this is a special case of the results of [38], [39]. The o.d.e. (15), however, can also have other, 'mixed' equilibria $\hat{x}' := [x'_1, \cdots, x'_N]^T$ where not all x'_i are identical. We investigate the role of α in bringing about consensus or disagreement of opinions, i.e., convergence to homogeneous or mixed equilibria, resp.

Our interest here is in the case of 'opinion manipulation' where some, say agents $\{m + 1, \cdots, N\}$, fix their opinions to some prescribed equilibrium x^* of (16) for all k. Let $\tilde{x}_k$ denote the opinions of the remaining agents. They track the o.d.e.

$$\dot{\tilde{x}}(t) = \alpha(P_1 - I)\tilde{x}(t) + \alpha P_2(x^*\mathbf{1}) + (1 - \alpha)\tilde{F}(\tilde{x}(t)), \tag{17}$$

where $\tilde{F}(x) := [f(x_1), \cdots, f(x_m)]^T$, in the following sense. We assume that P_1 above is irreducible. If B denotes the union of stable equilibria of (17) and B^ϵ an ϵ-neighborhood thereof for some $\epsilon > 0$, then

$$\limsup_{n\uparrow\infty} E\left[\inf_{y\in B^\epsilon} \|x_n - y\|^2\right] = O(\gamma).$$

See Theorem 3, p. 106, and Bullet (ii), p. 109, of [20]. Thus the following result captures the essence of almost sure asymptotic behavior of (14).

[7] Recall that an o.d.e. $\dot{x}(t) = f(x(t))$ is cooperative if $\frac{\partial f_i}{\partial x_j} \geq 0$ for $i \neq j$ ([37], p. 33-34).

Theorem 3: Consider the o.d.e. (17).

(i) A solution $\tilde{x}(t)$ as $t \uparrow \infty$ *cannot* converge to $[x', \cdots, x']^T$ for some equilibrium x' of (16) other than x^*.

(ii) If α is close enough to 1 (i.e., everyone succumbs to peer pressure), $x^*\mathbf{1}$ is the only equilibrium, i.e., a consensus on the desired opinion is obtained.

(iii) Suppose (16) has a continuously differentiable Liapunov function associated with it. Consider $\tilde{x}(0) \in B_R := \{x : \|x\| \leq R\}$. If $\alpha \approx 0$ (i.e., everyone trusts her own judgement better), then for all initial conditions outside a set of small Lebesgue measure, $\tilde{x}(t)$ converges to a small neighborhood of some $\hat{x}' = [x'_1, \cdots, x'_m]^T$, where x'_i are equilibria of (16).

Proof: By (17), it follows that if $\tilde{x}(t)$ converges to $x = [x_1, \cdots, x_m]^T$ where the x_is are equilibria of (16), then $x = x^*(I - P_1)^{-1}P_2\mathbf{1}$. It then follows easily from the discussion of Section IV that $x = x^*\mathbf{1}$, implying (i).

Rewrite the equilibrium condition $\alpha(P_1 - I)x + (1 - \alpha)\tilde{F}(x) + \alpha P_2(x^*\mathbf{1}) = 0$ as $x = G(x)$ for $G(y) := (P_1 y - \frac{1-\alpha}{\alpha}\tilde{F}(y)) + x^* P_2 \mathbf{1}$. Under our hypotheses, $\lambda(P_1) \in (0, 1)$. Let $w = [w_1, \cdots, w_m]^T$ denote the corresponding eigenvector. Then $w_i > 0 \ \forall i$ and P_1 is a strict contraction w.r.t. the norm $\|x\|_{w,\infty} := \max_i |\frac{x_i}{w_i}|$. Then G is a contraction w.r.t. $\| \cdot \|_{w,\infty}$ for $\frac{1-\alpha}{\alpha} < \frac{1-\lambda}{L}$ for $L :=$ the Lipschitz constant for f, so (ii) holds.

Suppose (16) has a continuously differentiable Liapunov function Ψ associated with it. Then $\tilde{\Psi}(x) := \sum_i \Psi(x_i)$ serves as a Liapunov function for the o.d.e.

$$\dot{x}(t) = \tilde{F}(x(t)). \tag{18}$$

If $\alpha \approx 0$, (17) is a regular perturbation of (18). By Theorem 1 of [40], for initial conditions in B_R outside a small neighborhood of unstable equilibria and their stable manifolds, $\tilde{x}(t)$ will converge to a small neighborhood of a stable equilibrium of (18), implying (iii). ■

The situation is more complex if we replace the scalar valued f by a vector valued function. In fact, in this case generic convergence to the set of equilibria is not assured in general unless (16) itself is also cooperative.

A. Experiments

We consider three types of graphs for creating the underlying network of nodes- complete graph, random graphs and planted multi-section [41] graphs. In each case, we pick a particular instance of the above graphs and consider opinion dynamics on it. We do not investigate the impact of the graph type on opinion dynamics. This is an interesting problem in its own right and will be taken up in future research.

The ‘cost function’ assumed in the experiments is shown in Figure 2. It has two minima 0.746058 and 1.751667, with the latter being the global minimum. The results for a 500 node network generated by sampling a random graph (edge probability 0.6) were reported in [3]. Interestingly the results suggested that there is an optimal range of peer pressure (parameter α) in which the population achieves the global optimum. Above this range there is consensus but at a sub-optimal solution, whereas below it there is no consensus.

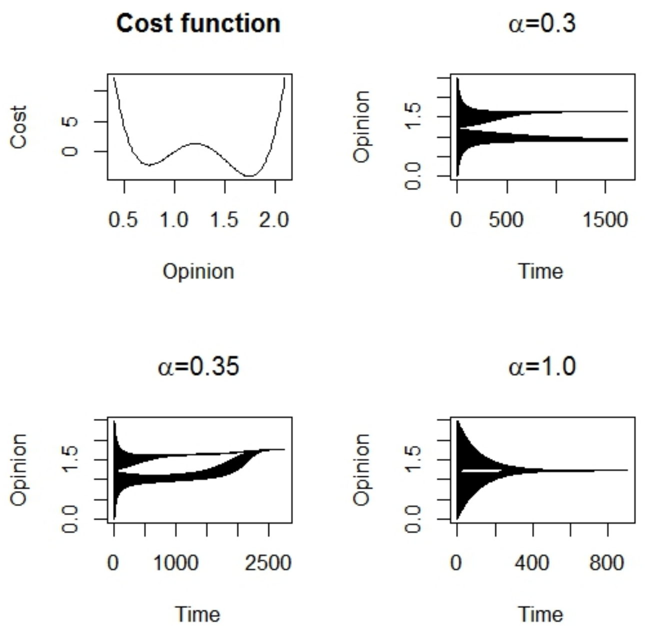

Fig. 2. Convergence for opinions for various values of α (initial condition equals 1).

Here we report results of our experiments on a 500 node planted multi-section graph with two classes (say, A and B), $p = 0.8$ and $q = 0.5$. We consider two initial conditions. In the first, class A nodes have initial opinions less than 1.21 (essentially in the basin of 0.746058) and the remaining have values between 1.21 and 2.5 (thus, in the basin of 1.751667). The initial values are generated uniformly randomly between $[0, 1.21]$ and $[1.21, 2.5]$ for class A and class B respectively. The second choice of initial condition is the same as the first except that 10% nodes of class A have their initial opinions sampled uniformly randomly in $[1.21, 2.5]$, and 10% nodes of class B have their initial opinions generated uniformly randomly in $[0, 1.21]$. The opinions are updated iteratively till the norm of the difference between two successive opinion vectors becomes less than 10^{-4}. We set $\gamma = 0.01$.

Shown in Figure 2 are the opinion trajectories for various values of α when the first initial condition is employed. The following are some qualitative observations.

- At $\alpha = 0$ the population remains divided in terms of opinions.
- It remains so till α exceeds a certain threshold. For the first initial condition, the threshold is approximately 0.32 whereas for the second it is approximately 0.28. Thus, a consensus cannot be reached unless sufficient attention is paid to the peers.
- For values of α greater than the above threshold, there is not only consensus but the population achieves the global optimum 1.751667 (if at all a consensus is reached, it is necessarily at one of the minima).
- At $\alpha = 1$ there is consensus at 1.16. This is the worst case.

Robustness of these results was established by experi-

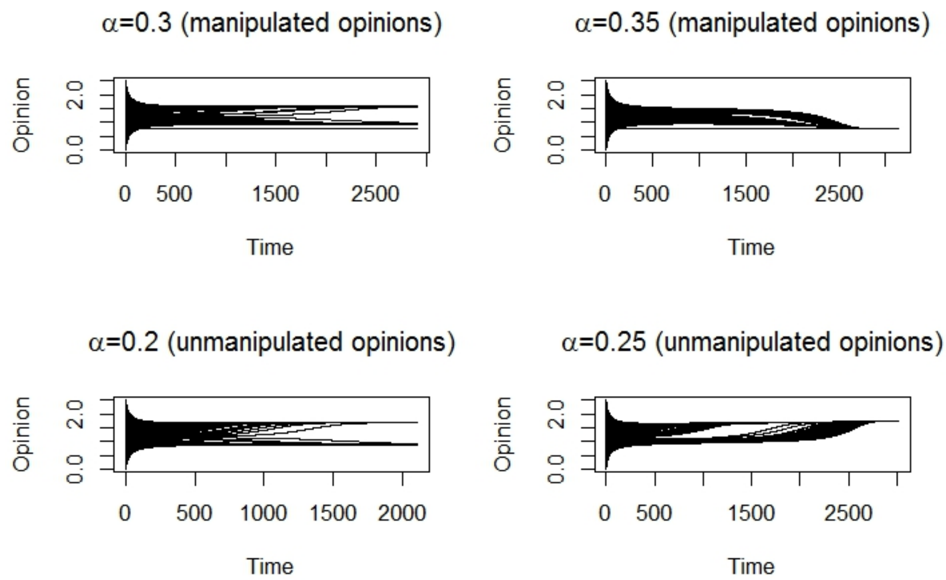

Fig. 3. Convergence for opinions for various values of α in the presence of inflexible agents (opinion manipulation).

menting with other initial conditions (e.g., Uniform $[0, 2.5]$; randomly chosen from $\{0, 1.21, 2.5\}$ with equal probability to extremists and centrists) and two samples of the graph. We omit these results here for lack of space.

For the case of 'opinion manipulation', the set-up is as before: the network an undirected planted multi-section graph of size $n = 500$, $p = 0.8$ and $q = 0.5$, the cost function as before. The initial opinions are generated uniformly randomly in $[0, 2.5]$. Then 10%, i.e., 50 agents are chosen uniformly randomly and their opinions are fixed to 0.746058. The opinions are updated iteratively till the norm of the difference between two successive opinion vectors became less than 10^{-4}. Parameter γ is set to 0.01, with α_i same for all i. Our observations are as follows (see Figure 3 for opinion trajectories for various values of α).

- Till α exceeds 0.31 there is disagreement on opinions. Thus, unless individuals pay sufficient attention to their peers, stubborn individuals cannot force consensus to their opinion.
- For $\alpha > 0.31$ there is consensus at 0.746058, the opinion held by stubborn individuals.
- If opinions are *not* manipulated, the threshold is approximately 0.22. For $\alpha > 0.22$ there is consensus at 1.751667.

Thus if individuals have sufficient pressure from their peers, stubborn individuals can drive the population to agree on a sub-optimal solution.

B. Optimal Control

Consider now a parametrized family $P^u = [[p(j|i, u(i))]]$ as in the preceding sections with $u(i) \in$ a common compact metric space U, and replace (17) correspondingly by

$$\dot{\tilde{x}}(t) = \alpha(P_1^{u(t)} - I)\tilde{x}(t) + \alpha P_2^{u(t)}(x^*\mathbf{1}) + (1-\alpha)\tilde{F}(\tilde{x}(t)), \tag{19}$$

where $u(\cdot) := [u_1(\cdot), \cdots, u_m(\cdot)]^T$ is now a control process, i.e., a function of time. (This is distinct from the preceding sections, where we considered *parametric* optimization where the parameter is chosen once for all and kept fixed.) Assume that the mappings $u \mapsto P_1^u$ and $u \mapsto P_2^u$ are continuous. We first prove a stability result. Let $\check{x}^* := x^*\mathbf{1}$.

Lemma 4: For $R > 0$, $\|\tilde{x}(t)\|$ is uniformly bounded for $\|\tilde{x}(0)\| \leq R$.

Proof: Note that for $\beta(t) := \alpha P_2^{u(t)}(x^*\mathbf{1}) + (1-\alpha)\tilde{F}(\tilde{x}(t))$,

$$\tilde{x}(t) = \Psi(t, 0)\tilde{x}(0) + \int_0^t \Psi(t, s)\beta(s)ds,$$

where Ψ is the transition matrix for the linear system $\dot{z}(t) = \alpha(P_1^{u(t)} - I)z(t)$. View $\alpha(P_1^{u(t)} - I)$ as the transition rate matrix for an absorbing continuous time Markov chain $\{X_t\}$ on $\{1, \cdots, m\}$ with absorbing boundary $\{m+1, \cdots, N\}$. Then $z_j(t) = E[z_{X_t}(0)I\{\tau > t\}|X_0 = j]$, where τ is the time of absorption. Under our irreducibility hypothesis, $P(\tau > t)$ has exponential decay, hence so do $z(\cdot)$ and $\Psi(\cdot, 0)$. Since $\beta(\cdot)$ is uniformly bounded, the claim follows. ∎

Assume that f is continuously differentiable with bounded derivatives. Consider the infinite horizon discounted cost

$$J(u(\cdot), x) := \int_0^\infty e^{-\beta t} \sum_{i=1}^m |\tilde{x}_i(t) - x^*|^2 dt, \tag{20}$$

where $\tilde{x}(0) = x := [x_1, \cdots, x_m]^T$. We assume that $\beta > (1-\alpha)\sup_x |f'(x)|$. Note that there is no cost on the control choice. This will be introduced later. Define the value function

$$V(x) := \inf_{u(\cdot)} J(u(\cdot), x).$$

Lemma 5: The function $x \mapsto V(x)$ is locally Lipschitz, a.e. differentiable, has a minimum at $\check{x}^*$ and satisfies: for all x at which V is differentiable, $\frac{\partial}{\partial x_i}V(x) \geq 0$ if $x_i > x^*$ and < 0 if $x_i < x^*$, $1 \leq i \leq m$.

Proof: Fix $u(\cdot)$ and let $X(x,t), t \geq 0$, denote the solution to (19) with $X(x,0) = x$. Then by standard arguments [42], $x = [x_1, \cdots, x_m]^T \mapsto X(x,t)$ is continuously differentiable and $DX(t) := [[\frac{\partial}{\partial x_j}X_i(x,t)]]$ satisfies the equation of variation

$$\frac{d}{dt}DX(x,t) = (\alpha(P_1^{u(t)} - I) + (1-\alpha)DF(\tilde{x}(t)))DX(x,t),$$

where $DF(x) := diag(\frac{d}{dx_1}f(x_1), \cdots, \frac{d}{dx_m}f(x_m))$. Noting that $DX(x,0) = I$, we get

$$DX(x,t) = I + \int_0^t (1-\alpha)\Psi(t,s)DF(\tilde{x}(s))DX(x,s)ds,$$

where Ψ is the transition matrix for the linear system $\dot{z}(t) = \alpha(P_1^{u(t)} - I)z(t)$. Taking, on both sides of the above equation, the matrix norm induced by the maximum norm over $\mathcal{R}^m$, and then using the Gronwall inequality, we get, for some $c > 0$,

$$\begin{aligned}\|DX(x,t)\|_\infty &\leq c\exp(-(1-\alpha)t\sup_x \|DF(x)\|_\infty) \\ &= c\exp(-(1-\alpha)t\sup_x |f'(x)|).\end{aligned}$$

Now, from Lemma 4, the above bound on $\|DX(x,t)\|_\infty$ and the assumption $\beta > (1-\alpha)\sup_x |f'(x)|$, Lipschitz continuity of $x \mapsto J(u(\cdot),x)$ on compact sets, uniformly in $u(\cdot)$, follows. The Lipschitz continuity of V is immediate from this, from which its a.e. differentiability follows by Rademacher's theorem. Since $V(x) \geq 0$ and $= 0$ only at $\check{x}^*$, $\check{x}^*$ is a minimizer. For the last claim, recall that cooperative o.d.e.s lead to monotone flows [36], [37]. Thus $y \geq x$ implies that $X(y,t) \geq X(x,t)\ \forall\ t \geq 0$ w.r.t. the usual partial order. It follows that $J(u(\cdot),x)$ must be increasing in x_i for $x_i > x^*$ and decreasing otherwise when all other components are kept fixed. Since pointwise infimum of a family of increasing functions is increasing, the same then must hold for V. The third claim follows. ■

Our result below has a simple structure due to the absence of any explicit control cost.

Theorem 6: The optimal control $u^*(\cdot)$ is characterized by: For $1 \leq i \leq m$, if $x_j(t) < x^*$,

$$u_i^*(t) \in \text{Argmax}\left(\sum_{j=1}^m p(j|i,\cdot)x_j(t) + x^* \sum_{j=m+1}^N p(j|i,\cdot)\right), \tag{21}$$

and if $x_j(t) > x^*$,

$$u_i^*(t) \in \text{Argmin}\left(\sum_{j=1}^m p(j|i,\cdot)x_j(t) + x^* \sum_{j=m+1}^N p(j|i,\cdot)\right). \tag{22}$$

Proof: Considering initial conditions in a ball of radius R sufficiently large, Lemma 4 allows us to consider $\tilde{x}(\cdot)$ in a bounded set. Thus we may modify the 'running cost function' $\|x - \check{x}^*\|^2$ outside this ball and suppose that it is bounded Lipschitz. By standard arguments, V is then the unique bounded viscosity solution to the Hamilton-Jacobi equation

$$\begin{aligned}\min_u \Big[&\langle \nabla V(x), \alpha(P_1^u - I)x + \alpha P_2^u(x^*\mathbf{1}) + (1-\alpha)F(x)\rangle \\ &+ \|x - \check{x}^*\|^2 - \beta V(x)\Big] = 0.\end{aligned} \tag{23}$$

Since V is a.e. differentiable, the necessity part of the claim follows from the results of [43] and the sufficiency follows from the results of [44], in view of the last part of Lemma 5 above. (These references deal with finite horizon cost, but then V is also the value function for the 'finite horizon' control problem with cost $E[\int_0^T e^{-\beta t}\|\tilde{x}(t) - x^*\|^2 dt + e^{-\beta T}V(\tilde{x}(T))]$.) ■

This theorem conforms to our intuition that the control would push $\tilde{x}(t)$ towards $\check{x}^*$ as much as possible. This, however, has been possible because there was no cost on the control choice. If we introduce the cost $g(u) := \sum_i g_i(u(i))$ as in the preceding section, the net cost is

$$J(u(\cdot),x) := \int_0^\infty e^{-\beta t}\sum_{i=1}^m (g_i(u_i(t)) + |\tilde{x}_i(t) - x^*|^2)dt, \tag{24}$$

and we have the following extension of Theorem 6:

Theorem 7: The optimal control $u^*(\cdot)$ is characterized by: for $1 \leq i \leq m$,

$$\begin{aligned}u_i^*(t) \in \text{Argmin}\Bigg(&\frac{\partial V}{\partial x_i}(x(t))\sum_{j=1}^m p(j|i,\cdot)x_j(t) \\ &+ x^* \sum_{j=m+1}^N p(j|i,\cdot) + g_i(\cdot)\Bigg),\end{aligned}$$

where V is the unique locally Lipschitz viscosity solution to the Hamilton-Jacobi equation

$$\begin{aligned}\min_u \Big[&\langle \nabla V(x), \alpha(P_1^u - I)x + \alpha P_2^u(x^*\mathbf{1}) + (1-\alpha)F(x)\rangle \\ &+ \|x - \check{x}^*\|^2 + g(u) - \beta V(x)\Big] = 0.\end{aligned}$$

C. Extensions

For $f(\cdot) = [f_1(\cdot), \cdots, f_r(\cdot)]^T \in C_b(\mathcal{R}^r)$, (19) gets replaced by

$$\begin{aligned}\dot{\tilde{x}}_i^k(t) &= \alpha(\sum_{j=1}^m p_k(j|i,u_i(t))\tilde{x}_j^k(t) - \tilde{x}_i^k(t) \\ &+ (\sum_{j=m+1}^N p_k(j|i,u_i(t)))x^*(k)) + (1-\alpha)f_k(\tilde{x}_i^1(t), \cdots, \tilde{x}_i^r(t))\end{aligned}$$

with $u_i(t) := [u_i^1(t), \cdots, u_i^r(t)]$ and the cost

$$J(u(\cdot),x) := \int_0^\infty e^{-\beta t}\sum_{i=1}^m (g_i(u_i(t)) + \|\tilde{x}_i(t) - x^*\|^2)dt,$$

where $\tilde{x}_j(t) = [\tilde{x}_j^1(t), \cdots, \tilde{x}_j^r(t)]^T$ and $x^* = [x^*(1), \cdots, x^*(r)]^T \in \mathcal{R}^r$ is a prescribed stable equilibrium for

$$\dot{x}(t) = f(x(t)).$$

The Hamilton-Jacobi equation is:

$$\min_u \Big[\sum_{i=1}^m\sum_{k=1}^r\Big\{\frac{\partial V}{\partial x_i^k}(x)\Big(\alpha(\sum_{j=1}^m p_k(j|i,u_i^k)x_j^k - x_i^k$$
$$+(\sum_{j=m+1}^N p(j|i,u_i^k))x^*(k)) + \ (1-\alpha)f_k(x_i^1,\cdots,x_i^r)\Big)$$
$$+(g_i(u_i^k)+|x_i^k-x^*(k)|^2)\Big\}\Big] - \beta V(x) = 0.$$

Once again, V is the unique locally bounded viscosity solution of this equation and the optimal $u(t)$ is a.e. given by the minimizer above.

We can also consider an adversarial action effected through another control $v(t) = [[v_i^k(t)]]], v_i^k(t) \in$ a compact set U_i', seeking to *maximize* the cost,[8] whence the dynamics becomes

$$\begin{aligned}\dot{\tilde{x}}_i^k(t) =& \alpha(\sum_{j=1}^m p_k(j|i,u_i^k(t),v_i^k(t))\tilde{x}_j^k(t) - \tilde{x}_k(t)\\ &+(\sum_{j=m+1}^N p_k(j|i,u_i^k(t),v_i^k(t))x^*(k))\\ &+(1-\alpha)f_k(\tilde{x}_i^1(t),\cdots,\tilde{x}_i^r(t)).\end{aligned}$$

We move over to the relaxed (or 'chattering') control framework of [45]. That is, we replace the control spaces U_i by the spaces of probability measures on them with Prohorov topology, and $p(j|i,u,v)$ by $\int\int p(j|i,y,y')u(dy)v(dy')$ correspondingly, $u(\cdot), v(\cdot)$ now being probability measures on U_i, U_i' resp. This amounts to 'mixed strategies' in game theory parlance. The Hamilton-Jacobi equation gets replaced by the Hamilton-Jacobi-Isaacs equation for zero sum differential games:

$$\min_u\max_v h(u,v) = \max_v\min_u h(u,v) = 0,$$

where $h(u,v) =$

$$\Big[\sum_{i=1}^m\sum_{k=1}^r\Big\{\frac{\partial V}{\partial x_i^k}(x)\Big(\alpha\Big[\sum_{j=1}^m\int\int p_k(j|i,y,y')u_i^k(dy)v_i^k(dy')x_j^k$$
$$-x_i^k + (\sum_{j=m+1}^N\int\int p(j|i,y,y')u_i^k(dy)v_i^k(dy'))x^*(k)\Big]$$
$$+(1-\alpha)f_k(x_i^1,\cdots,x_i^r)\Big)$$
$$+(g_i(u_k(i))+|x_i^k-x^*(k)|^2)\Big\}\Big] - \beta V(x)$$

Under our 'relaxed control' formulation, it follows from Theorem 1.10, p. 438, and Theorem 2.6, p. 445, of [46] (see also Proposition 2.9, p. 447 of [46]) that the differential game has an Elliott-Kalton value (see [46], Chapter VIII for background) which coincides with the unique viscosity solution to the above Hamilton-Jacobi-Isaacs equation. (We use the same argument as for Theorem 6 above to consider a bounded payoff function without any loss of generality.) The corresponding verification theorem then is given by Theorem 4.6 of [47].

Note that the controlled o.d.e.s above have been limiting cases of an originally discrete time problem in each case. Thus optimal controls or game strategies have to be correspondingly interpreted as approximations for the discrete time problem. Alternatively, one can write the Bellman equation (for control) or Shapley equation (for games) for the discrete time problem directly. For the scalar case, these are, resp.,

$$V(x) = \min_{u=\{u_i\}}\Big[g(u) + \|x-x^*\|^2 + \ \beta\sum_{\{i'\}}\prod_i p(x_{i'}|x_i,u_i)\times$$
$$V([\cdots, x_i + \gamma\left(\alpha(x_{i'}-x_i) + (1-\alpha)f(x_i)\right),\cdots]^T)\Big],$$

and

$$V(x) = \min_{u=\{u_i\}}\max_{v=\{v_i\}} h(u,v) = \max_{v=\{v_i\}}\min_{u=\{u_i\}} h(u,v),$$

where

$$h(u,v) = \Big[g(u) + \|x-x^*\|^2$$
$$+\beta\sum_{\{i'\}}\prod_i\int\int p(i'|i,y,y')u_i(dy)v_i(dy')\times$$
$$V([\cdots, x_i + \gamma\left(\alpha(x_{i'}-x_i) + (1-\alpha)f(x_i)\right),\cdots]^T)\Big]$$

While this avoids certain technical issues inherent in the continuous time formulation, it is clumsier to work with.

VII. Conclusion

We have considered a model of 'controlled gossip' wherein nodes of a connected network form opinions / learn the average of a quantity by a successive averaging procedure as in the classical gossip algorithms, except that a certain number of nodes has its values frozen to a common value. We consider three optimization problems associated with this model and present algorithms to solve them. Our computational experiments for the uncontrolled scenario demonstrate interesting critical phenomena as weight on the peer pressure is varied.

This work is a first step towards control of opinion dynamics and gossip algorithms, for a simple model that permits analytical treatment to some extent. There are clearly many more theoretical and computational issues involved, which we hope to pursue in future works.

[8] Our game formulation uses a zero sum formulation. In many applications, this may not be accurate. For example, in case of a firm competing for the market share for its product with other firms, it is in general a nonzero sum game. Our formulation is tantamount to a 'worst case' viewpoint clubbing all other firms into a single adversary, which is not accurate. This is because while any market share lost is indeed that gained by the collective lot of competitors, the same does not hold for the costs incurred. Thus the zero sum model is perforce an approximation. It is the 'engineer's license' that is being used, because zero sum models are tractable whereas nonzero sum models are not, except under special structures such as potential games which do not seem realistic here.

References

[1] V. S. Borkar, J. Nair, and S. Nalli, "Manufacturing consent," in *Proc. of 48th Annual Allerton Conference on Communication, Control, and Computing*, 2010, pp. 1550–1555.

[2] V. S. Borkar and A. Karnik, "Controlled gossip," in *Proc. of 49th Annual Allerton Conference on Communication, Control, and Computing*, 2011, pp. 707–711.

[3] ——, "Opinion formation under peer pressure," Interdisciplinary Workshop on Information and Decision in Social Networks, MIT, USA, May 31-June 1, 2011.

[4] J. R. P. French Jr., "A formal theory of social power," *Psychological Review*, vol. 63, pp. 181–194, 1956.

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

[6] N. E. Friedkin and E. C. Johnsen, "Social influence networks and opinion change," *Advances in Group Processes*, vol. 16, pp. 1–29, 1999.

[7] A. G. Chandrasekhar, H. Larreguy, and J. P. Xandri, "Testing models of social learning on networks: Evidence from a framed field experiment," Working Paper, Tech. Rep., 2012.

[8] J. Lorenz, "Continuous opinion dynamics under bounded confidence: A survey," *Intl. J. of Modern Physics C*, vol. 18, no. 12, pp. 1819–1838, 2007.

[9] P. M. DeMarzo, D. Vayanos, and J. Zwiebel, "Persuasion bias, social influence and uni-dimensional opinions," *Quarterly J. of Economics*, vol. 118, no. 3, pp. 909–968, 2003.

[10] L. Corazzini, F. Pavesi, B. Petrovich, and L. Stanca, "Influential listeners: An experiment on persuasion bias in social networks," *European Economic Review*, vol. 56, no. 6, pp. 1276–1288, 2012.

[11] M. Lewenstein, A. Nowak, and B. Latane, "Statistical mechanics of social impact," *Phys. Rev. A*, vol. 45, no. 2, pp. 763–776, 1992.

[12] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dynamic Games and Applications*, vol. 1, pp. 3–49, 2011.

[13] S. Chatterjee and E. Seneta, "Towards consensus: Some convergence theorems on repeated averaging," *J. Applied Probability*, vol. 14, pp. 89–97, 1977.

[14] B. Golub and M. Jackson, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, vol. 2, no. 1, pp. 112–149, 2010.

[15] D. Shah, *Gossip Algorithms*. Now Publishers, 2009.

[16] F. Cucker and S. Smale, "Emergent behavior in flocks," *IEEE Trans. on Auto. Control*, vol. 52, no. 5, pp. 852–862, 2007.

[17] E. Yildiz, D. Acemoglu, A. Ozdaglar, A. Saberi, and A. Scaglione, "Discrete opinion dynamics with stubborn agents," 2011, preprint available at papers.ssrn.com.

[18] I. Chueshov, *Monotone Random Systems - Theory and Applications*. Lecture Notes in Mathematics No. 1779, Springer Verlag, Berlin - Heidelberg, 2002.

[19] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, pp. 888–893, 2010.

[20] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Hindustan Book Agency, New Delhi, and Cambridge Univ. Press, Cambridge, UK, 2008.

[21] ——, *Topics in Controlled Markov Chains*. Longman Scientific and Technical, Harlow, UK, 1991.

[22] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, no. 1-7, pp. 107–117, 1998.

[23] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[24] M. Jackson, *Social and Economic Networks*. Princeton University Press, 2008.

[25] H. Boyle and R. Dykstra, "A method for finding projections onto the intersection of convex sets in Hilbert space," in *Advances in Order Restricted Statistical Inference*, ser. Lecture Notes in Statistics. Springer Verlag, 1985, pp. 28–47.

[26] S.-P. Han, "A successive projection method," *Mathematical Programming*, vol. 40, pp. 1–14, 1988.

[27] N. Maculan, C. Santiago, E. Macambira, and M. Jardim, "An O(n) algorithm for projecting a vector on the intersection of a hyperplane and a box in R^n," *J. of Optimization Theory and Applications*, vol. 117, pp. 553–574, 2003.

[28] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions - I," *Mathematical Programming*, vol. 14, pp. 265–294, 1978.

[29] R. Sundaram, *A First Course in Optimization Theory*. Cambridge University Press, 1996.

[30] G. H. Golub and C. F. Van Loan, *Matrix computations (3rd ed.)*. Johns Hopkins University Press, Baltimore, USA, 1996.

[31] W. Zachary, "An information flow model for conflict and fission in small groups," *J. of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.

[32] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E*, vol. 74, no. 3, 2006.

[33] A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[34] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[35] C. W. Wu, *Synchronization in Complex Networks of Nonlinear Dynamical Systems*. World Scientific, Singapore, 2007.

[36] M. W. Hirsch, "Systems of differential equations that are competitive or cooperative II: convergence almost everywhere," *SIAM J. Math. Anal.*, vol. 16, no. 3, pp. 423–489, 1986.

[37] H. L. Smith, *Monotone Dynamical Systems*. American Math. Soc., Providence, USA, 2008.

[38] K. C. Chang, H. Lee, and Y. Oh, "A generalization of Perron's stability theorems for perturbed linear differential equations," *Kyungpook Math. J.*, vol. 33, no. 2, pp. 163–171, 1993.

[39] T. Taniguchi, "Stability theorems for perturbed linear ordinary differential equations," *J. Math. Anal. and Apppl.*, vol. 149, no. 2, pp. 583—598, 1990.

[40] M. W. Hirsch, "Convergent activation dynamics in continuous time networks," *Neural Networks*, vol. 5, no. 2, pp. 331–349, 1989.

[41] F. McSherry, "Spectral partitioning of random graphs," in *Proc. of 42nd IEEE Symposium on Foundations of Computer Science, Oct. 8-11, 2001*, 2002, pp. 529–537.

[42] V. I. Arnold, *Ordinary Differential Equations*. Springer Verlag, Berlin-Heidelberg, 1992.

[43] E. N. Barron and R. Jensen, "The Pontryagin maximum principle from dynamic programming and viscosity solutions to first-order partial differential equations," *Transactions of the American Math. Society*, vol. 258, no. 2, pp. 635–641, 1987.

[44] X. Y. Zhou, "Verification theorems within the framework of viscosity solutions," *Journal of Optimization Theory and Appl.*, vol. 177, no. 1, pp. 208–225, 1993.

[45] L. C. Young, *Lectures on the Calculus of Variations and Optimal Control Theory*. AMS Chelsea Publishing, Providence, USA, 2000.

[46] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Birkhauser, Boston, 2000.

[47] M. K. Ghosh, K. Suresh Kumar, A. K. Nandakumaran, and K. S. Mallikarjuna Rao, "Differential games of fixed duration in the framework of relaxed strategies," *Differential Equations and Dynamical Systems*, vol. 13, no. 3+4, pp. 251–273, 2005.

Prof. V. S. Borkar got his B.Tech. (Electrical Engg.) from IIT Bombay (1976), M.S. (Systems and Control) from Case Western Reserve Uni. (1977), and Ph.D. (EECS) from Uni. of California, Berkeley (1980). He has held positions at TIFR Centre for Applicable Mathematics and Indian Institute of Science in Bangalore and Tata Institute of Fundamental Research in Mumbai before joining IIT Bombay where he is an Institute Chair Professor. He has held visiting positions at Uni. of Twente, MIT, Uni. of Maryland at College Park, Uni. of California at Berkeley and Uni. of Illinois at Urbana-Champaign. He is a Fellow of science and engineering academies in India and of IEEE, TWAS and AMS. He is the recipient of the S. S. Bhatnagar award given by the Government of India. His research interests are stochastic optimization and applications.

Aditya Karnik received his B.E. (Elec. & Telecom) from the University of Pune, India, and M.E. and Ph.D. (both in ECE) from the Indian Institute of Science, Bangalore, India. He is currently a Senior Scientist in the Modelling & Optimization lab at GE Global research. Prior to joining GE he was with General Motors R&D for more than 5 years. His research interests are in decision sciences and its application to industrial analytics, social networks and communication networks.

Jayakrishnan Nair received his BTech and MTech in Electrical Engg. (EE) from IIT Bombay (2007) and Ph.D. in EE from California Inst. of Tech. (2012). He has held post-doctoral positions at California Inst. of Tech. and Centrum Wiskunde & Informatica. He is currently an Assistant Professor in EE at IIT Bombay. His research focuses on modeling, performance evaluation, and design issues in queueing systems and communication networks.

Sanketh Nalli received his B.Tech. in Information Technology from NIT Karnataka in 2012. He is currently a graduate student in the Dept. of Computer Sciences at University of Wisconsin - Madison. His research focuses on optimizing operating systems and file systems for next generation non-volatile memories.