

Spotting trendsetters: Inference for network games

Randall Berry and Vijay G. Subramanian EECS Department

Northwestern University, Evanston, IL 60208

Email: rberry@eecs.northwestern.edu, v-subramanian@northwestern.edu

Abstract—Network games provide a basic framework for studying the diffusion of new ideas or behaviors through a population. In these models, agents decide to adopt a new idea based on optimizing pay-off that depends on the adoption decisions of their neighbors in an underlying network. Assuming such a model, we consider the problem of inferring early adopters or first movers given a snap shot of the adoption state at a given time. We present some results on solving this problem in the low temperature regime. We conclude with a discussion on reducing the complexity of such inference problems for large networks.

I. INTRODUCTION

Increased digital connectivity has led to an ever increasing confluence of social, economic and information networks [16]. This in turn has led to much work on understanding network structures and the processes they enable. One fundamental example is the *diffusion* [5], [6] of new ideas or behaviors through a population, where an individual's decision to adopt a new idea or behavior is influenced by her neighbors in an underlying social network. One of the main theoretical frameworks for analyzing diffusion processes is via a *network game*, i.e., a model in which rational agents make decisions to maximize a pay-off that depends on the underlying network structure [3], [4], [7], [8], [13], [15], [18]; these are also referred to as graphical games [9]. This line of work has focused on analyzing equilibrium properties of such games but has largely not considered inference questions in such settings. In this paper, we focus on inferring the time history of a given diffusion process given one or more snap shots of the process. Based on the logistic function trend of technology adoption [5], the adopters are classified into the following different types based on when they take-up the products: first-movers/early adopters, early majority, late majority and laggards. Thus, an important special case of the inference problem is to determine the agents which started the diffusion, i.e. the *first movers* or *early adopters*.

In a basic diffusion model, there is one *status quo* action A and one new action B . For example, A could be an existing smart phone operating system and B could be a newly introduced operating system, which is not fully compatible with A . Agents decide to adopt the new action or stay with the *status quo* based in part on the decisions of their “peers” (e.g., co-workers, friends, etc.), which are identified as one-hop neighbors in an underlying social network. For the smart phone example, this peer influence can model the fact that B might enable certain applications to be used only with other users of B . Alternatively, this might simply be the result of getting favorable reviews of the product from one's peers.

In a network game model, agents decide to adopt B or stay with A by choosing the action that maximizes their own benefit. The influence of an agent's neighbors is captured by assuming that the benefit an agent receives from an action is a function of the number of neighbors who have chosen that action. When one agent in such a network makes a new decision, this can lead her neighbors to also change their decision; the diffusion process is simply the history of these changes. In many cases, the entire time-history of a diffusion may not be available, e.g. due to technical limitations, privacy considerations or simply that one was not recording data about the given diffusion until some time after it started. For such cases, we seek to infer the most likely history of diffusion given one or more snap shots of the process.

A prominent alternative to game theoretic modeling of diffusion processes is to employ *epidemic models* [1], [2], in which the adoption of the new action B is viewed as the spread of an infection, with uninfected nodes becoming infected with a probability that depends on the number of its neighbors that are infected. For an epidemic model, a problem related to identifying innovators or early adopters is the *influence maximization problem* [10], [12]. This problem does not focus on identifying early adopters for a given partial history of a diffusion, but rather seeks to identify a set of users, who if they were first movers would yield the largest expected spread of the new innovation. Closer to our work, the problem of estimating the source of an infection in an epidemic model has been studied in [17], [19], [20]. The main difference of these models from ours is that they do not model any strategic choices that agents may make, and so the resulting diffusion dynamics can be quite different [3], [7], [15]. For random graphs, best-response dynamics have been analyzed in [26]. In [21], [22], the complexity of the seed selection problem or the starting set problem is analyzed, while in [27], the same problem is considered but with irreversible or monotone dynamics.

II. COORDINATION GAMES & DIFFUSION

In this paper we present an analysis for a simple network coordination game. However, the analysis easily extends to more general network games. For concreteness, we describe this in terms of a technology adoption scenario, in which a set of n users are choosing between the *status quo* technology A and a new technology B . The users are represented as nodes in an undirected graph $\mathcal{G} = (V, E)$, which models an underlying social network. An example of such a network for 8 nodes is shown in Figure 1, which loosely captures a

key property of realistic network graphs: there are a few nodes with a high degree whereas most nodes have low degrees.

Each user is playing a game in which her strategy set $\mathcal{A} = \{A, B\}$ is simply the choice of technology. The users' pay-offs are given by the sum of per-link pay-off functions, where each link's pay-off function corresponds to a *coordination game* between the end-points of the link, i.e., a game in which the two players receive a better pay-off when they both either adopt the new technology or stay with the *status quo*. More precisely, let $u(a_i, a_j)$ be the link pay-off function for any two neighbors i and j , where a_i represents the action of agent i . If we normalize each player's link payoffs for adopting different technology to be zero, then this implies that $u(A, B) = u(B, A) = 0$. Let $v_A = U(A, A)$ and $v_B = U(B, B)$ be the player's payoffs when they both adopt technology A or B , respectively. The coordination game assumption requires that both v_A and v_B be greater than zero. Additionally, we assume that the new technology is superior given that both users adopt it, so that $v_B > v_A$. Given these link payoffs, the total payoff for each agent i is then given by $U_i(a_i, a_{-i}) = \sum_{j \in N(i)} u(a_i, a_j)$, where a_{-i} denotes the actions of all the players other than i and $N(i)$ denotes the set of neighbors to i in the social network. For example, in Figure 1, node 1's pay-off will depend on her own actions and those of nodes 4, 6 and 8.

In a diffusion scenario all nodes are initially playing the *status quo* A , and at some point one or more nodes adopt B . Subsequently, at random times each node has the opportunity to change her strategy based on the current actions of their neighbors. For example, nodes may employ a best response update, in which they maximize their payoff assuming the strategies of all other users are fixed. Returning to Figure 1, if $v_A = 1$ and $v_B = 1.5$, then node 1's best response will be to adopt B , when at least two of her neighbors have adopted B ; otherwise node 1's best response will be to stay with the *status quo* A . Note that a node's best response only depends v_A and v_B through their ratio or equivalently through the parameter $h = \frac{v_B - v_A}{v_B + v_A} \in (0, 1)$, which will be a convenient equivalent parameterization of the game. Note also that the best response is equivalent to choosing state B only when the number of neighbors who are in state B exceed a node-dependent threshold. Assuming best response updates, the "starting set" problem considered in [21], [22] is closely related to inferring first movers. This problem seeks to identify the smallest number of nodes that need to have adopted B in order for the whole network to adopt B following a sequence of best responses, and is shown to be NP-hard and also hard to approximate.

Instead of a pure best response, we assume a *noisy best response* [8], which allows agents to choose non-optimal actions with some probability. One reason for choosing such models is that they account for imperfect decisions made by people in practice. A second reason is to model certain learning behaviors of agents. A third reason is simply that such models have nice mathematical properties and can approximate true best response dynamics in a limiting sense.

The specific dynamics we consider are noisy best-response

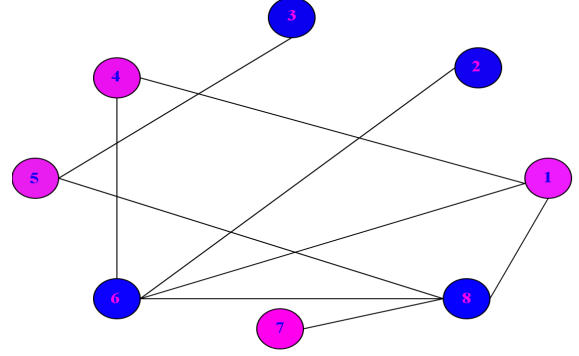


Fig. 1: A social network among 8 users.

dynamics following a logistic response function. This is parameterized by a non-negative number β defined as follows. Let the state of the system at time t be given by the vector of actions chosen by the different nodes, $X^t = (a_1^t, a_2^t, \dots, a_n^t)$. Given that node i can update her action at time t , the probability that node i chooses action $\hat{a}_i \in \mathcal{A}$ is given by

$$p^\beta(X_{-i}^t, \hat{a}_i) = \frac{e^{\beta U_i(\hat{a}_i, X_{-i}^t)}}{\sum_{a \in \mathcal{A}} e^{\beta U_i(a, X_{-i}^t)}}$$

where X_{-i}^t represents the vector of the actions of all nodes other than i ; we will denote the transition probability by $p^\beta(X^t, X^{t+1})$ for a transition from state X^t to X^{t+1} where we allow the two states to be the same. Thus, the dynamics are such that when $\beta = 0$, a node picks an action at random, but when $\beta \uparrow \infty$, the preferred action is the one that maximizes the payoff, i.e., one gets best-response dynamics. Note that under these dynamics it is possible for a node to reverse its adoption decision at a given time t , i.e., it could adopt B and then at a latter time switch back to A .

Each user is assumed to have a timer whose duration is randomly chosen independent of every other timer and exponentially distributed with parameter 1. When a user's timer expires, then the user changes its action/state and immediately restarts the timer with an independent and identically distributed value; note that this implies that only one user will change her state at any given time. With the above specification we model the dynamics as a Markov process with state-space $\mathcal{X} = \mathcal{A}^V$. This formulation is well accepted [3], [4], [7], [8], [13], [15], [18] for analyzing games over networks; the specific dynamics are called *Glauber dynamics* in the Markov-Chain-Monte-Carlo literature [14]. Note that this asynchronous update model introduces randomness even when $\beta = \infty$, i.e., when each user chooses her action based on the best-response calculation.

III. INFERRING HISTORY OF DIFFUSION

Next we state our inference problem more precisely and discuss some results. Suppose that prior to time 0, the system is in the *ground* or initial state corresponding to all users adopting the *status quo*; we denote this by $X^{0^-} = \mathbf{A}$, where \mathbf{A} is the "all A " state, and \mathbf{X}^t denotes the overall state (i.e.,

current vector of actions) at time t . Suppose further that some user changes its state at time 0 and at time t we find the system in a state z in which several users have adopted B . The problem is to identify the most likely history of adoption leading to x , and in particular to identify the most likely user who changed its state at time 0. Denote by e_i the state whose i^{th} component is B while the rest are A . Then the most likely first mover is given by the maximum likelihood estimate, $\hat{i} \in \arg \max_{i \in V} P_{e_i}^{t,\beta}(z)$, where $P_{e_i}^{t,\beta}(z)$ is the probability that the Markov process starting out in state e_i at time 0 is in state z at time t .

This formulation is also applicable in other contexts that correspond to transitions from one state in the configuration space to another. For example, consider if the system being in state x corresponds to a certain state of the market. Then a transition to a state with a smaller number of B 's corresponds to a reduction in market share for the company selling product B . Finding out the first customer to change her/his state corresponds to locating the most likely customer to leave, which is of considerable interest, e.g., for reducing churn for wireless service providers.

Since the Markov process can also be described via its generator matrix Q^β and since the state space is finite, it follows that $P_{e_i}^{t,\beta}(x)$ is the (e_i, x) entry of $\exp(Q^\beta t)$, i.e., a specific term of the matrix exponential obtained from the generator. The accurate calculation of the matrix exponential is a hard problem computationally; for our particular case the complexity is $O(|\mathcal{X}|^4)$ which is exponential in n . Additionally, with the underlying Markov process being ergodic, it follows that irrespective of the starting state y , $P_y^{t,\beta}(x)$ converges to the stationary probability of being in state x as $t \rightarrow \infty$, which we denote by $\pi^\beta(x)$. Therefore the maximum likelihood estimation procedure must be carried out with high fidelity. This makes a compelling case for an alternate solution to this problem. Our approach is to consider the asymptotic regime of large β while assuming that the time t of observation stays fixed.¹

The underlying Markov model for the diffusion is identical to models of Glauber dynamics studied in the statistical physics literature [11], [14], [15]. From the analysis of these models, it follows that the stationary probability $\pi^\beta(x)$ is proportional to $\exp(-\beta H(x))$, where

$$H(x) = - \sum_{(i,j) \in E} x_i x_j - \sum_{i \in V} h d_i x_i$$

and d_i is the degree of node i . Furthermore, transitions only occur between states $x, y \in \mathcal{X}$ that differ in one-coordinate/user with probability $p^\beta(x, y) = \exp(-\beta G(x, y) + o(\beta))$, where $G(x, y) = \max(H(x), H(y)) - H(x)$. Noting that the probability of being in state x at time t when starting out from state e_i at time 0 can be obtained by adding together the probabilities of all paths in state-space (with repeats) that start at e_i and end at x , we define the *length* of a path w in configuration

space to be sum of the weights of the edges traversed, i.e., $W(w) = \sum_{k=1}^{|w|-1} G(w_k, w_{k+1})$ where w_t is the node at head of the edge used at step t and $|w|$ is exactly one more than the total number of edges traversed. A path from state x to y that reaches y only at the last step is also denoted using the following notation: $w : x \rightarrow y$. Then a key initial result that we can prove is the following.

Theorem 1: For a given diffusion process on a social network defined by logistic best response dynamics with parameter β ,

$$\lim_{\beta \rightarrow \infty} \frac{-\log P_y^{t,\beta}(x)}{\beta} = \inf_{w: x \rightarrow y} W(w)$$

for every $x, y \in \mathcal{X}$, where w is an allowed path in configuration space from x to y .

See Appendix A. The exponential decay of the likelihood probability in β indicated in the theorem can be used to determine the behavior of the maximum likelihood estimator. Let $ML^{t,\beta}(z)$ be the set of maximum likelihood estimates of the first movers given the state of the process at time t being z , i.e.,

$$ML^{t,\beta}(z) := \arg \max_{i \in V} P_{e_i}^{t,\beta}(z),$$

then Theorem 1 yields the following important corollary.

Corollary 1: For every $t > 0$, there exists a large enough but finite β^* such that for all $\beta > \beta^*$ we have

$$\arg \max_{i \in V} P_{e_i}^{t,\beta}(z) = ML^{t,\beta}(z) \subseteq \arg \min_{i \in V} \inf_{w: e_i \rightarrow z} W(w)$$

See Appendix B. In other words, for large but finite β , we can obtain the maximum likelihood estimate by solving a much simpler shortest path problem in configuration space where the states are connected via a directed graph with weights $G(x, y)$. The solution to the shortest path problem also yields the most likely paths for the state to be x at time t . When there are multiple solutions to the shortest path problem, not all solutions necessarily have the maximum likelihood. Separating such solutions will involve a more refined analysis of the constant terms in $P_{e_i}^{t,\beta}(z)$, which is for future work.

Note that we have to solve a shortest-path problem on a directed graph with weights $G(a, b)$ on the edges. Thus, the estimate for the set of early adopters is

$$\hat{i} \in \arg \min_i \inf_{w: e_i \rightarrow z} W(w).$$

In the reversible setting, which holds owing to Glauber dynamics, one can obtain another estimate (that will coincide) by the following property. From reversibility we get that

$$\pi^\beta(x) P_x^{t,\beta}(y) = \pi^\beta(y) P_y^{t,\beta}(x)$$

so that

$$P_x^{t,\beta}(y) = \pi^\beta(y) P_y^{t,\beta}(x) / \pi^\beta(x).$$

Thus, in our setting with the Glauber dynamics, we get

$$\lim_{\beta \rightarrow \infty} \frac{-\log P_x^{t,\beta}(y)}{\beta} = \lim_{\beta \rightarrow \infty} \frac{-\log P_y^{t,\beta}(x)}{\beta} + H(x) - H(y)$$

¹The same asymptotic regime has been considered in the physics literature [11] for analysis of metastability.

where we further have that

$$\lim_{\beta \rightarrow \infty} \frac{-\log P_y^{t,\beta}(x)}{\beta} = \inf_{w:y \rightarrow x} W(w).$$

Therefore, the estimate for the set of early adopters is also given by

$$\hat{i} \in \arg \min_i H(e_i) + \inf_{w:z \rightarrow e_i} W(w),$$

i.e., one considers the minimum cost path from z to each e_i and amends its cost with a termination cost which is the cost of state e_i . We can solve this problem using the Dijkstra algorithm [24] with complexity $O(|\mathcal{X}| \log(|\mathcal{X}|))$. Note that this is less complex than computing the matrix exponential but is still of exponential complexity in n . Notice also that the shortest path estimator does not depend on t , the time when the state is observed, whereas using the matrix exponential does not have this failing. Thus, our next goal is to develop an estimator that explicitly depends on t but satisfies the following two requirements: for large enough β it agrees with the shortest path estimator and has complexity that is $O(|\mathcal{X}| \log(|\mathcal{X}|))$.

A. Approximation by hitting times

We will provide an alternate proof for our main result by analyzing hitting times. The alternate proof provides further intuition as to why the shortest path calculation yields an asymptotically accurate estimate. In addition, we will use this to achieve our goal of a time-dependent asymptotically optimal estimator that has lower complexity than using the matrix exponential.

For two distinct states x, y define τ_x^y to be the first time the Markov process hits y when it starts in state x , i.e.,

$$\tau_x^y = \inf\{u \geq 0 : X^u = y\}.$$

With the above definition, we can write down the following integral equation

$$P_{e_i}^{t,\beta}(z) = \int_0^t P_z^{t-u,\beta}(z) dP_{e_i}^{u,\beta}(\tau_{e_i}^z = u),$$

i.e., in $(u, u + du)$ the Markov chain hits state s for the first time and then it returns to state s in $t - u$ time. An upper bound follows immediately since $P_z^{u,\beta}(z) \leq 1$, namely,

$$P_{e_i}^{t,\beta}(z) \leq \int_0^t dP_{e_i}^{u,\beta}(\tau_{e_i}^z = u) = P^\beta(\tau_{e_i}^z \leq t).$$

Now we use the proof of [14, Prop. 10.18] and reversibility of the Markov chain to prove that for all $x \in \mathcal{X}$, $P_x^{t+u,\beta}(x) \leq P_x^{t,\beta}(x)$ for all $t \geq 0$ and $u \geq 0$; reversibility is key in the proof, see Lemma 1 in Appendix C. The lower bound this yields is

$$P_{e_i}^{t,\beta}(z) \geq P_z^{t,\beta}(z) P^\beta(\tau_{e_i}^z \leq t).$$

The intuition from the upper and lower bounds is that if one can find an $x \in \{e_1, \dots, e_n\}$ such that the random variable τ_x^z is made as small as possible (in a stochastic ordering sense), then we can well approximate the maximum likelihood estimate \hat{i} whenever $P_z^{t,\beta}(z)$ is close to 1. We note

that minimizing the random variable τ_x^z is also non-trivial.⁴ Thus, we consider the behaviour of $P_x^\beta(\tau_x^z \leq t)$ as $\beta \uparrow \infty$. Note that t should be much smaller than the mixing time of the Markov chain for the maximum likelihood problem to have a non-trivial solution, otherwise $P_x^{t,\beta}(z)$ will be close to $\pi^\beta(z)$ for any feasible $x \in \mathcal{X}$; in general the mixing time is an increasing function of β . We can, thus, ignoring this *caveat* for the case of finite t and use the fact that $P_z^t(z) \geq e^{-nt}$ (none of the Poisson clocks tick) to show that

$$\lim_{\beta \rightarrow \infty} \frac{-\log P_{e_i}^{t,\beta}(z)}{\beta} = \lim_{\beta \rightarrow \infty} \frac{-\log P^\beta(\tau_{e_i}^z \leq t)}{\beta},$$

assuming that both limits exist. We can prove that the asymptote of $P^\beta(\tau_x^z \leq t)$ is exactly the solution to the shortest-path problem.

Proposition 1: For a given diffusion process on a social network defined by logistic best response dynamics with parameter β ,

$$\lim_{\beta \rightarrow \infty} \frac{-\log P^\beta(\tau_{e_i}^z \leq t)}{\beta} = \inf_{w:x \rightarrow y} W(w)$$

for every $x, y \in \mathcal{X}$, where w is an allowed path in configuration space from x to y .

See Appendix C. Having established this result, we will use the intuition of stochastically minimizing the hitting time to develop an estimator that incorporates the time of measurement: given t and e_i the goal will be to maximize $P^\beta(\tau_{e_i}^z \leq t)$.

B. Approximation by max-product algorithm

The inference problem that we are interested in specifies that there is a transition at time 0 and at time t the state of the Markov chain is z . We are, however, not given the states through which the Markov chain reaches z starting out at \mathbf{A} at time 0. If we are also given the number of transitions, say k , that it takes to reach state z , then we have a hidden Markov model problem. We can then use the backward Chapman-Kolmogorov equations, i.e., the sum-product algorithm [25], to determine the distribution of the state at time 0 (i.e., at the first transition) and then find the maximizer to obtain the maximum likelihood estimate of the first adopters. Following the intuition from the previous results for large enough β that using the most likely path to z also leads to the maximum likelihood estimator and this also coincides with stochastically minimizing the hitting time to z , we will, instead, find the most likely sequence of states to reach state z and use that to construct the estimate. In more detail, given k for each $i \in \{1, 2, \dots, n\}$ we find

$$w_{e_i \rightarrow z}^*(k) \in \arg \max_{|w|=k, w:e_i \rightarrow z} \prod_{t=0}^{k-1} p^\beta(w_t, w_{t+1}),$$

which can be obtained using the max-product algorithm [25], i.e., the Viterbi algorithm; denote the maximum by $p^\beta(w_{e_i \rightarrow z}^*(k))$. Then we determine the number of steps that yield the maximum probability to reach z from e_i , i.e.,

$$k^*(i) \in \arg \max_{k \in \{0, 1, 2, \dots\}} \frac{t^k}{\Gamma(k+1)} p^\beta(w_{e_i \rightarrow z}^*(k)).$$

It is easy to see that we only need to consider a finite number of terms in the above so that the maximum exists and this can be computed: determine the probability of one path from e_i to s , say $\tilde{p}(e_i \rightarrow s)$, then there exists a \tilde{k} such that $\frac{t^k}{\Gamma(k+1)}$ is less than $\tilde{p}(e_i \rightarrow z)$ for all $k \geq \tilde{k}$; the summability of $\{\frac{t^k}{\Gamma(k+1)}\}_{k=\{0,1,2,\dots\}}$ guarantees this. Then the estimator for the early adopters is

$$\hat{i} \in \arg \max \left\{ i \in \{1, \dots, n\} : \frac{t^{k^*(i)}}{\Gamma(k^*(i)+1)} p^\beta(w_{e_i \rightarrow z}^*(k^*(i))) \right\}.$$

The complexity of running the Viterbi algorithm [25] is linear in the size of the state space, i.e., it is $O(|\mathcal{X}|)$. As mentioned earlier the Viterbi algorithm has to be run a finite number of times that depends on t and β but not the size of the state space. The final step of taking the maximum involves evaluating the probabilities for each e_i with $i \in \{1, 2, \dots, n\}$ which results in the final complexity being $O(|\mathcal{X}| \log(|\mathcal{X}|))$. We also have the following result

Proposition 2: For every t , there exists a β^* such that for all $\beta \geq \beta^*$ we have

$$\arg \max \left\{ i \in \{1, \dots, n\} : \frac{t^{k^*(i)}}{\Gamma(k^*(i)+1)} p^\beta(w_{e_i \rightarrow z}^*(k^*(i))) \right\} \\ = \arg \min_i \inf_{w: e_i \rightarrow z} W(w)$$

We skip the proof as it is carried out in a similar manner to the proof of Corollary 1.

C. Example

We illustrate these results for the small example in Figure 1, for which the maximum likelihood estimate can be calculated exactly. We consider the case of $x = (BAABBABA)$ and $h = \{0.55\}$ with the maximum likelihood estimates shown in Figure 2. When h assumes the intermediate value, then the maximum likelihood estimate is eventually 8, which coincides with the shortest path estimate. An interesting feature in this example is at the time of sampling, node 8 is in state A, i.e., even though node 8 is the likely first mover to B, she changes her state in the future. Such complex behaviour makes the problem hard to analyze. In Figure 2 we also point out that node 1 is the maximum likelihood estimate of the early adopter for intermediate β . Interestingly, in this case the dynamics are monotone, i.e., no node switches back to A once it has adopted B.

IV. CONCLUSIONS

For network games with noisy best response dynamics with the level of noise being inversely proportional to parameter $\beta \geq 0$, we formulated an inference problem that seeks to determine early adopters when a given system state is observed. Modeling the dynamics as a Markov chain, we can solve the inference problem by calculating a matrix exponential. Known algorithms for directly solving this maximum likelihood problem via calculating the matrix exponential function have complexity $O(|\mathcal{X}|^4)$, which is of exponential complexity in the number of players, n . Our results show that

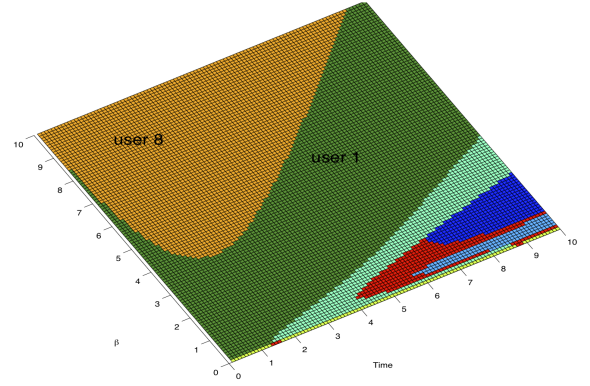


Fig. 2: Maximum likelihood estimate for state $x = (BAABBABA)$.

in the regime of large β , one can obtain the answer exactly by solving a shortest path problem. We then showed using reversibility of the underlying Markov chain that the shortest path problem can be solved using the Dijkstra algorithm with complexity $O(|\mathcal{X}| \log(|\mathcal{X}|))$. Using this idea we then developed an asymptotically optimal algorithm based on the max-product algorithm. While this yields a considerable reduction in complexity, it is still of exponential complexity in n . Thus, in future work we will look at much faster algorithms to approximately solve the shortest path problem so that it will scale to large data sets. However, based on the results on the starting set problem [21], [22], we believe that an exact solution is computationally intractable. Additionally, one can consider similar problems when the social network and the payoffs are unknown and have to be estimated from data.

APPENDIX I PROOF OF THEOREM 1

Note that we can look at a discrete-time version of the continuous-time Markov chain by using uniformization using a Poisson process at rate n , i.e., we sample the continuous-time Markov chain at instances of a Poisson process (of fast-enough rate). Using this random time-change representation we can then write the following decomposition for the probability of being in state y at time t when starting from state x (assume that $x \neq y$, namely,

$$P_x^{t,\beta}(y) = \sum_{w: x \rightarrow y} e^{-nt} \frac{(nt)^{|w|}}{\Gamma(|w|+1)} \prod_{t=0}^{|w|-1} \frac{1}{n} p^\beta(w_t, w_{t+1})$$

where w is a path from x to y , $|w|$ the path-length and $p^\beta(x, y)$ are transition probabilities for the discrete-time chain obtained using uniformization, i.e., the probabilistic rule imposed once a state is picked. When we assume that the transition probabilities are given by $p(a, b) = \exp(-\beta G(a, b) + o(\beta))$ with $\beta \in [0, \infty)$, we get

$$P_x^{t,\beta}(y) = \sum_{w: x \rightarrow y} e^{-nt} \frac{t^{|w|}}{\Gamma(|w|+1)} e^{-\beta \sum_{t=0}^{|w|-1} G(w_t, w_{t+1}) + |w| o(\beta)}.$$

For later use, for a path $w : x \rightarrow y$ let $W(w) = \sum_{t=0}^{|w|-1} V(w_t, w_{t+1})$ be the path cost. We are interested in the case of $\beta \rightarrow \infty$, and would like to prove that we can approximate $P_x^{t,\beta}(y) \approx C \exp(-\beta I)$ for some C and I . The term I is the exponential decay rate of $P_x^{t,\beta}(y)$ in β and is given by

$$\lim_{\beta \rightarrow \infty} \frac{-\log P_x^{t,\beta}(y)}{\beta} = I \stackrel{(\text{Conjecture})}{=} \inf_{w=x \rightarrow y} \sum_{t=0}^{|w|-1} G(w_t, w_{t+1}).$$

One part of the bound is obvious since

$$P_x^{t,\beta}(y) \geq e^{-nt} \frac{t^{|w|}}{\Gamma(|w|+1)} e^{-\beta \sum_{t=0}^{|w|-1} G(w_t, w_{t+1}) + |w|o(\beta)}$$

for every path w from x to y . This then yields

$$\limsup_{\beta \rightarrow \infty} \frac{-\log P_x^{t,\beta}(y)}{\beta} \leq \inf_{w=x \rightarrow y} \sum_{t=0}^{|w|-1} G(w_t, w_{t+1}).$$

Let $W^* = \inf_{w=x \rightarrow y} \sum_{t=0}^{|w|-1} G(w_t, w_{t+1})$. It is obvious that $W^* \in [0, \infty)$ and also that there exists at least one path that has this cost. In fact, for every path $w : x \rightarrow y$ we can find a sub-path within that never revisits x by going back to the last time the discrete-time chain visits x before time $|w|$. Furthermore, for paths that never revisit x , one can determine a loop-free path by ignoring some transitions. Since we are always ignoring transitions, i.e., certain probabilities, the probability of a path w is upper-bounded by the probability of the loop-free path, denote this set by \mathcal{W} : starting with the state visited right after x (say z), find the last time z is visited and ignore all transitions in between, and then repeat this procedure, then the finite length of each path ensures that this procedure terminates. Note that there are only a finite number of loop-free paths from x to y owing to finiteness of the state-space. Next for path-length k , there can be at the most $2^{n(k-1)}$ paths from x to y , as fixing the end-points at x and y we have $k-1$ degrees of freedom. Using these definitions we can write down the following (crude) upper bound

$$\begin{aligned} P_x^{t,\beta}(y) &\leq \sum_{k=0}^{\infty} e^{-nt} \frac{2^{n(k-1)} t^k}{\Gamma(k+1)} \sum_{w \in \mathcal{W}} e^{-\beta W(w) + |w|o(\beta)} \\ &= 2^{-n} e^{(2^n - n)t} \sum_{w \in \mathcal{W}} e^{-\beta W(w) + |w|o(\beta)}. \end{aligned}$$

Now the result follows by the principle of the largest term [23] since \mathcal{W} is a finite set.

APPENDIX II

PROOF OF COROLLARY 1

Let A be the adjacency matrix of the graph implied by $p^\beta(\cdot, \cdot)$ where we include the self-transitions as well. Then the number of paths from x to y of length k are given by $A^k(x, y)$. We will use this fact to refine our estimates above. We can write the following when $x \neq y$

$$P_x^{t,\beta}(y) = \sum_{k=1}^{\infty} e^{-nt} \frac{t^k}{\Gamma(k+1)} \sum_{w:x \rightarrow y, |w|=k} e^{-\beta V(w) + ko(\beta)}.$$

Since the lower bound is the same as before, we get a better estimate for the upper bound. From the proof, we can upper bound the probability of each path from x to y by the probability of a path that only visits nodes only once. Since there are only finitely many such paths, we can further upper bound the probability by considering the largest probability; it is easy to see that this would scale as $e^{-\beta \inf_{w:x \rightarrow y} W(w) + o(\beta)}$. Thus, we have

$$P_x^{t,\beta}(y) \leq e^{-\beta \inf_{w:x \rightarrow y} W(w) + o(\beta)} \sum_{k=1}^{\infty} e^{-nt} \frac{t^k}{\Gamma(k+1)} A^k(x, y).$$

Denote by $Q^t = \exp(At)e^{-nt}$ where $\exp(\cdot)$ the matrix exponential function, then

$$P_x^{t,\beta}(y) \leq e^{-\beta \inf_{w:x \rightarrow y} W(w) + o(\beta)} Q^t(x, y).$$

The lower bound then implies that

$$P_x^{t,\beta}(y) \geq e^{-\beta \inf_{w:x \rightarrow y} W(w) + o(\beta)} e^{-nt} \frac{t^{k^*}}{\Gamma(k^*+1)} A^{k^*}(x, y)$$

where k^* is the length of the shortest path that takes the least number of steps. The conclusion then holds by contradiction. Comparing the upper and lower bounds for different β , for large enough β the values corresponding to the shortest path estimate are greater than any others so that the maximum likelihood estimate has to belong to the shortest path estimate set. Note that we are not claiming that the maximum likelihood estimate converges as $\beta \rightarrow \infty$, which is a much harder result to establish.

APPENDIX III

PROOF OF PROPOSITION 1

We first start by proving the following lemma.

Lemma 1: For a finite-state, irreducible and reversible Markov chain for all $x \in \mathcal{X}$, $P_x^{t+u,\beta}(x) \leq P_x^{t,\beta}(x)$ for all $t \geq 0$ and $u \geq 0$.

By the Chapman-Kolmogorov equations we can write

$$P_x^{t+s,\beta}(x) = \sum_{y,v \in \mathcal{X}} P_x^{t/2,\beta}(y) P_y^{s,\beta}(v) P_v^{t/2,\beta}(x).$$

Multiplying on both sides by $\pi^\beta(x)$ we get using reversibility that

$$\begin{aligned} &\pi(x) P_x^{t+s,\beta}(x) \\ &= \sum_{y,v \in \mathcal{X}} \pi^\beta(x) P_x^{t/2,\beta}(y) P_y^{s,\beta}(v) P_v^{t/2,\beta}(x) \\ &= \sum_{y,v \in \mathcal{X}} P_y^{t/2,\beta}(x) \pi^\beta(y) P_y^{s,\beta}(v) P_v^{t/2,\beta}(x) \\ &= \sum_{y,v \in \mathcal{X}} P_y^{t/2,\beta}(x) P_v^{s,\beta}(y) \pi^\beta(v) P_v^{t/2,\beta}(x) \\ &= \sum_{y,v \in \mathcal{X}} P_y^{t/2,\beta}(x) \sqrt{\pi^\beta(y) P_y^{s,\beta}(v)} \times \\ &\quad \sqrt{P_v^{s,\beta}(y) \pi^\beta(v)} P_v^{t/2,\beta}(x) \end{aligned}$$

where once again by reversibility $\pi^\beta(y) P_y^{s,\beta}(v) = P_v^{s,\beta}(y) \pi^\beta(v)$. Defining

$\varphi(y, v) := P_y^{t/2, \beta}(x) \sqrt{\pi^\beta(y) P_y^{s, \beta}(v)} \geq 0$, we note that the right side above is nothing but $\sum_{y, v \in \mathcal{X}} \varphi(y, v) \varphi(v, y)$ which by the Cauchy-Bunyakovsky-Schwarz inequality is upper-bounded by $\sum_{y, v \in \mathcal{X}} \varphi(y, v)^2$. This then yields, again using reversibility, that

$$\begin{aligned} & \pi(x) P_x^{t+s, \beta}(x) \\ & \leq \sum_{y, v \in \mathcal{X}} \varphi(y, v)^2 \\ & = \sum_{y, v \in \mathcal{X}} P_y^{t/2, \beta}(x) P_y^{t/2, \beta}(x) \pi^\beta(y) P_y^{s, \beta}(v) \\ & = \sum_{y \in \mathcal{X}} P_y^{t/2, \beta}(x) P_y^{t/2, \beta}(x) \pi^\beta(y) \\ & = \sum_{y \in \mathcal{X}} \pi^\beta(x) P_x^{t/2, \beta}(y) P_y^{t/2, \beta}(x) = \pi^\beta(x) P_x^{t, \beta}(x) \end{aligned}$$

which proves our result since $\pi^\beta(x) > 0$ for all $x \in \mathcal{X}$. A few remarks are in order: i) in the discrete-time setting of [14, Prop. 10.18], the result translates to $\max(P_x^{2k+1}(x), P_x^{2k+2}(x)) \leq P_x^{2k}(x)$ for $k = 0, 1, 2, \dots$ and using this one can prove that $P_x^k(z) \geq P(\tau_x^z \leq k) P_z^{\lceil k/2 \rceil}(z)$; ii) one can also prove the result using the eigen-decomposition of reversible transition matrices as suggested in [14, Exer. 12.6]; and (iii) the ordering obtained is exactly like the lazy chain ordering of [14, Prop. 10.18] since for finite chains one can take h small enough that $P_x^h(x) \geq 1/2$ for all $x \in \mathcal{X}$.

Let A be a subset of \mathcal{X} . We denote by $p_A^\beta(\cdot, \cdot)$ the transition probability matrix restricted to $\mathcal{X} \setminus A$. Denote $p^\beta(x, A) := \sum_{y \in A} p^\beta(x, y)$ and for $x, v \in \mathcal{X} \setminus A$, let $p_A^{\beta, k}(x \rightarrow v)$ be the probability of going from state x to state v in k steps while never visiting A ; note that this is the (x, v) entry of the k -fold product of $p_A^\beta(\cdot, \cdot)$. Then $dP(\tau_x^A = t)/dt$ (i.e., the density of the hitting time random variable) is given by

$$\begin{aligned} & \frac{dP(\tau_x^A = t)}{dt} \\ & = \sum_{k=1}^{\infty} e^{-nt} n \frac{(nt)^{k-1}}{\Gamma(k)} \sum_{v \in \mathcal{X} \setminus A} \frac{1}{n^{k-1}} p_A^{\beta, k-1}(x, v) \frac{1}{n} p^\beta(v, A) \\ & = \sum_{k=1}^{\infty} e^{-nt} \frac{t^{k-1}}{\Gamma(k)} \sum_{v \in \mathcal{X} \setminus A} p_A^{\beta, k-1}(x, v) p^\beta(v, A). \end{aligned}$$

Therefore, we have the following

$$\begin{aligned} & P(\tau_x^A \leq t) \\ & = \sum_{k=1}^{\infty} e^{-nt} \frac{1}{n^k} \left(\sum_{l=k}^{\infty} \frac{(nt)^l}{\Gamma(l+1)} \right) \sum_{v \in \mathcal{X} \setminus A} p_A^{\beta, k-1}(x, v) p^\beta(v, A) \\ & = \sum_{v \in \mathcal{X} \setminus A} \sum_{k=1}^{\infty} e^{-nt} \frac{1}{n^k} \left(\sum_{l=k}^{\infty} \frac{(nt)^l}{\Gamma(l+1)} \right) p_A^{\beta, k-1}(x, v) p^\beta(v, A). \end{aligned}$$

Using the same steps as in proof of theorem 1 we have the following

$$\begin{aligned} & \lim_{\beta \uparrow \infty} \frac{-\log(P(\tau_x^A \leq t))}{\beta} \\ & = \inf_{v \in \mathcal{X} \setminus A} \inf_{w: x \rightarrow v, w \in (\mathcal{X} \setminus A)^{|w|}} V(w) + V(v, A) \\ & = \inf_{w: x \rightarrow A} V(w). \end{aligned}$$

Thus, the exponent is the value of the shortest path to A when starting out in x . Therefore, the estimate will be the e_i that minimizes this exponent over all initial states when $A = \{z\}$.

ACKNOWLEDGMENT

This research was supported in part by NSF under grant IIS-1219071.

REFERENCES

- [1] M. Granovetter, "Threshold models of collective behaviour," *American Journal of Sociology*, 1978, 83:1420–1443.
- [2] M. Morris, "Epidemiology and social networks: Modeling structure diffusion," *Sociol. Methods Res.*, 1993, 22:99–126.
- [3] G. Ellison, "Learning, local interaction, and coordination," *Econometrica*, 1993, 61:1047–1071.
- [4] M. Kandori, H. Mailath and F. Rob, "Learning, mutation, and long run equilibria in games," *Econometrica*, 1993, 61:29–56.
- [5] E. M. Rogers, "Diffusion of Innovations," *Free Press, New York*, 1995, 4th ed.
- [6] T. W. Valente, "Network Models of the Diffusion of Innovations," *Hampton, Cresskill, NJ*, 1995.
- [7] L. Blume, "The statistical mechanics of best-response strategy revision," *Games and Economic Behavior*, Elsevier, 1995, 11(2), pp. 111–145.
- [8] D. Fudenberg and D. Levine, "The Theory of Learning in Games," *MIT Press, Cambridge, MA*, 1998.
- [9] M. Kearns, M. Littman and S. Singh, "Graphical models for game theory," In *Proc. of the 17th Conf. on Uncertainty in AI*, 2001.
- [10] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. 7th ACM SIGKDD*, 2001, pp. 57–66.
- [11] A. Bovier and F. Manzo, "Metastability in Glauber dynamics in the low-temperature limit: Beyond exponential asymptotics," *J. Stat. Physics*, 107:757–779, 2002.
- [12] D. Kempe, J. Kleinberg and É. Tardos, "Maximizing the spread of influence in a social network," in *Proc. 9th ACM SIGKDD*, 2003, pp. 137–146.
- [13] M. O. Jackson and L. Yariv, "Diffusion of Behavior and Equilibrium Properties in Network Games," *American Economic Review*, 2007, 97(2), pp. 92–98.
- [14] D. A. Levin, Y. Peres, and E. L. Wilmer, "Markov chains and mixing times," *AMS, Providence, RI*, 2009.
- [15] A. Montanari and A. Saberi, "The spread of innovations in social networks," in *PNAS USA*, 2010, 107(47), 20196–20201.
- [16] D. Easley and J. Kleinberg, "Networks, Crowds, and Markets: Reasoning About a Highly Connected World," *Cambridge University Press*, 2010.
- [17] D. Shah and T. Zaman, "Detecting Sources of Computer Viruses in Networks: Theory and Experiment," *ACM Sigmetrics*, 2010.
- [18] H. P. Young, "The Dynamics of Social Innovation," forthcoming in *PNAS USA*, 2011.
- [19] D. Shah and T. Zaman, "Rumors in a Network: Who's the Culprit?," *IEEE Trans IT*, Aug. 2011, 57:5163–5181.
- [20] D. Shah and T. Zaman, "Finding Rumor Sources on Random Graphs," 2011, *eprint*, arXiv 1110.6230.
- [21] N. Chen, "On the Approximability of Influence in Social Networks," *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2008.
- [22] N. Chen, "On the Approximability of Influence in Social Networks," *SIAM Journal on Discrete Mathematics*, V.23(3), 1400–1415, 2009.
- [23] A. Dembo and O. Zeitouni, "Large Deviations Techniques and Applications," 2nd ed., Applications of Mathematics, Vol. 38., *Springer, New York*, 1998.

- [24] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, Introduction to Algorithms (Second ed.), *MIT Press and McGraw-Hill*, 2001.
- [25] C. M. Bishop, Pattern Recognition and Machine Learning, *Springer*, 2006.
- [26] H. Ohta and S. Sasa, "A universal form of slow dynamics in zero-temperature random-field Ising model," *Europhys. Lett.*, 90, 27008-1-5, 2010.
- [27] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta and Riccardo Zecchina, "The Spread Optimization Problem," *preprint*, arXiv:1203.1426.