



Brief paper

Community structure recovery and interaction probability estimation for gossip opinion dynamics[☆]Yu Xing^{a,*}, Xingkang He^b, Haitao Fang^{c,d}, Karl Henrik Johansson^a^a Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, and Digital Futures, Stockholm, Sweden^b Ericsson AB, Stockholm, Sweden^c Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, PR China^d School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, PR China

ARTICLE INFO

Article history:

Received 18 February 2021

Received in revised form 26 October 2022

Accepted 12 April 2023

Available online 7 June 2023

Keywords:

Community structure recovery

Opinion dynamics

Gossip models

Stubborn agents

Social networks

Markov chains

ABSTRACT

We study how to jointly recover community structure and estimate interaction probabilities of gossip opinion dynamics. In this process, agents randomly interact pairwise, and there are stubborn agents never changing their states. Such a model illustrates how disagreement and opinion fluctuation arise in a social network. It is assumed that each agent is assigned with one of two community labels, and the agents interact with probabilities depending on their labels. The considered problem is to jointly recover the community labels of the agents and estimate interaction probabilities between the agents, based on a single trajectory of the model. We first study stability and limit theorems of the model, and then propose a joint recovery and estimation algorithm based on a trajectory. It is verified that the community recovery can be achieved in finite time, and the interaction estimator converges almost surely. We derive a sample-complexity result for the recovery, and analyze the estimator's convergence rate. Simulations are presented for illustration of the performance of the proposed algorithm.

© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Networks appear across domains from biology to sociology. Real networks often exhibit community structures, where subsets of nodes have dense connections locally but sparse connections globally (Fortunato & Hric, 2016). Community detection is to partition nodes according to the network topology. There is a growing interest in studying community detection based on state observations of dynamics (Prokhorenkova, Tikhonov, & Litvak, 2022; Schaub, Segarra, & Tsitsiklis, 2020). Lacking topology data makes the problem harder than classic ones. Particularly, it is unclear how to recover communities out of a single trajectory of opinion dynamics (Ravazzi, Dabbene, Lagoa, & Proskurnikov, 2021).

[☆] This work was supported by the Knut and Alice Wallenberg Foundation (Wallenberg Scholar Grant), the Swedish Research Council (Distinguished Professor Grant 2017-01078), the Swedish Foundation for Strategic Research (CLAS Grant RIT17-0046), and the National Natural Science Foundation of China (Grant No. 11931018). The material in this paper was partially presented at the 59th IEEE Conference on Decision and Control, December 14–18, 2020, Jeju Island, Republic of Korea. This paper was recommended for publication in revised form by Associate Editor Julien M. Hendrickx under the direction of Editor Christos G. Cassandras.

* Corresponding author.

E-mail addresses: yuxing2@kth.se (Y. Xing), xingkang0715@gmail.com (X. He), htfang@iss.ac.cn (H. Fang), kallej@kth.se (K.H. Johansson).

1.1. Related work

In this subsection, we first review key community definitions and detection approaches (Fortunato & Hric, 2016), then discuss recovering communities based on state observations, and finally clarify our motivation.

Traditional community detection methods apply classic clustering techniques to node pairs assigned with certain weights (Girvan & Newman, 2002). Newman and Girvan (2004) introduce the concept of modularity to measure the quality of a graph partition. A famous algorithm based on optimizing modularity is the Louvain method (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008), which assigns nodes to one of the communities iteratively to achieve the largest modularity gain. Another approach to community detection is based on statistical inference, which introduces generative network models and considers an observed network as a sample. A canonical model is the stochastic block model (SBM). Abbe (2017) reviews detectability of the SBM and performance of algorithms. Besides optimization and statistical approaches, another method is based on dynamical processes (e.g. Rosvall & Bergstrom, 2008). Morescu and Girard (2010) propose a bounded-confidence model, where agents converge to several clusters corresponding to communities.

Recently, the study of community detection for networked dynamics has emerged. The problem is to recover communities

only based on state observations of a dynamical process. The main difference between this problem and the classic ones, especially the dynamic-based methods, is that the network is not available. Prokhorenkova et al. (2022) and Ramezani, Khodadadi, and Rabiee (2018) apply maximum likelihood methods to cascade data. Prokhorenkova et al. (2022) also propose a two-step procedure, first constructing a network and then clustering agents based on the network. Roddenberry, Schaub, Wai, and Segarra (2020), Schaub et al. (2020) and Wai, Segarra, Ozdaglar, Scaglione, and Jadbabaie (2019) introduce the blind community detection method, using sample covariance matrices of agent states for recovery. Peixoto (2019) investigates simultaneously reconstructing the topology and the community structure for epidemics and the Ising model. Berthet, Rigollet, and Srivastava (2019) study recovery for an Ising blockmodel.

We study how to jointly recover the community structure and estimate the interaction probabilities of gossip opinion dynamics. The problem arises from recent investigation of learning interpersonal influence from dynamics (Ravazzi et al., 2021). Network data is useful for decision making, but directly collecting such data can be hard, due to topic specificity (Cowan & Baldassarri, 2018), consistency issues (Netrapalli & Sanghavi, 2012), and privacy concern (De Montjoye et al., 2018). Learning large-scale networks may be computationally expensive, so recovering communities as a coarse description is a good option. The gossip update rule captures the random nature of individual interactions. It is a fundamental element of many opinion models (Proskurnikov & Tempo, 2017), and has been extensively studied (Boyd, Ghosh, Prabhakar, & Shah, 2006). Stubborn agents, such as media and opinion leaders, play a crucial role in opinion formation (Ramos et al., 2015). Acemoğlu, Como, Fagnani, and Ozdaglar (2013) show that the existence of stubborn agents can explain opinion oscillation. A generalization of stubborn agents is to assume that each agent has some level of stubbornness with respect to its initial belief. This generalization is considered by the Friedkin–Johnsen model and its extensions (Proskurnikov, Tempo, Cao, & Friedkin, 2017; Tian & Wang, 2018).

1.2. Contributions

We consider jointly recovering communities and estimating interaction probabilities for gossip opinion dynamics. Each agent is assigned with one of two community labels, and the agents interact with probabilities depending on their labels. Our contributions are as follows:

1. We study properties of the model by leveraging results on Markov chains and stochastic approximation (SA) (Theorem 1). It is shown that regular-agent states converge in distribution to a unique stationary distribution, and the time average of the agent states converge almost surely. An explicit expression for the mean of the stationary distribution is given (Proposition 2).

2. We develop a joint algorithm (Algorithm 1) to recover the community structure and to estimate the interaction probabilities, based on Polyak averaging and SA techniques. The algorithm is able to recover the communities in finite time, and then able to estimate the interaction probabilities consistently (Theorem 2).

3. We show how to theoretically analyze the developed joint algorithm. A concentration inequality for Markov chains (Lemma 1) is obtained, and it is used in the sample-complexity analysis of the recovery step (Theorem 3). The obtained result shows that the probability of unsuccessful recovery decays exponentially over time. Additionally, we analyze convergence rate of the interaction estimator from an SA argument (Theorem 4).

The obtained results indicate that a Polyak averaging technique can be useful for recovering communities based on a single

trajectory. In addition, we establish a sample-complexity result for successful recovery (recovering all community labels correctly), providing a quantitative dependence of the recovery probability on model parameters. These two points make our paper different from Roddenberry et al. (2020), Schaub et al. (2020) and Wai et al. (2019), which use covariance matrices of samples from several trajectories, and different from Wai, Scaglione, and Leshem (2016), which considers learning a sparse characterization of the network from the gossip model. The considered problem is different from classic system identification (e.g., Sarkar, Rakhlin, & Dahleh, 2021), because stubborn agents normally have fixed states, which does not satisfy input conditions required for system identification, and also because community recovery cannot be obtained directly from parameter estimates. The major differences between this paper and its conference version (Xing, He, Fang, & Johansson, 2020) are that we clarify our assumptions in more detail, characterize the sample complexity and the convergence rate of the algorithm, and add more numerical experiments to illustrate its performance.

1.3. Outline

The rest of the paper is organized as follows. Section 2 formulates the problem. Analysis of the model is given in Section 3, and a joint recovery and estimation algorithm is proposed in Section 4. Section 5 presents convergence results of the algorithm, and Section 6 provides several numerical experiments. Finally, Section 7 concludes the paper. Proofs are given in Xing, He, Fang, and Johansson (2021).

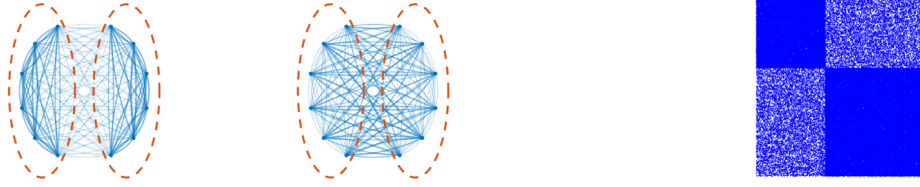
Notation. Denote the n -dimensional Euclidean space, the set of $n \times m$ real matrices, and the set of nonnegative integers by $\mathbb{R}^n$, $\mathbb{R}^{n \times m}$, and $\mathbb{N}$. Denote $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. Let $\mathbf{1}_n$, $\mathbf{e}_i$, I_n , and $\mathbf{0}_{n,m}$ be n -dimensional all-one vector, the unit vector with i th entry being one, the $n \times n$ identity matrix, and the $n \times m$ all-zero matrix. Let $\mathbf{1}_{n_1, n_2} := \mathbf{1}_{n_1} \mathbf{1}_{n_2}^T$. Denote both the Euclidean norm of a vector and the spectral norm of a matrix by $\|\cdot\|$. Denote the diagonal matrix with the elements of a vector x on the main diagonal by $\text{diag}\{x\}$. The i th component of a vector x is denoted by x_i , and the (i, j) -th entry of a matrix A by a_{ij} or $[A]_{ij}$. Denote the spectral radius of A by $\rho(A)$, and the cardinality of a set Ω by $|\Omega|$. $\mathbb{I}_{[\text{property}]}$ is the indicator function. For two sequences $\{a_k\}$ and $\{b_k\}$ with $a_k \in \mathbb{R}^n$ and $0 \neq b_k \in \mathbb{R}$, $k \geq 1$, $a_k = O(b_k)$ means that $\|a_k/b_k\| \leq C$ for all k and some $C > 0$, and $a_k = o(b_k)$ means that $\lim_{k \rightarrow \infty} \|a_k/b_k\| = 0$. An event happens almost surely (a.s.) if it happens with probability one.

2. Problem formulation

This section introduces the considered model and the definition of communities, and formulates the problem.

2.1. Gossip model with stubborn agents

The gossip model is a random process over an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the agent set $\mathcal{V}$, the edge set $\mathcal{E}$, and no self-loops. The agents have two types, regular and stubborn, denoted by $\mathcal{V}_r$ and $\mathcal{V}_s$, respectively ($\mathcal{V} = \mathcal{V}_r \cup \mathcal{V}_s$, $\mathcal{V}_r \cap \mathcal{V}_s = \emptyset$). Each agent i has a state $X_i(t) \in \mathbb{R}$, and the state vector at time $t \in \mathbb{N}$ is $X(t) \in \mathbb{R}^n$. Stubborn agents do not change their states during the process. An interaction probability matrix $W = [w_{ij}] \in \mathbb{R}^{n \times n}$ captures agent interactions, where $w_{ij} = w_{ji} \geq 0$, $w_{ij} > 0 \Leftrightarrow \{i, j\} \in \mathcal{E}$, $i, j \in \mathcal{V}$, and $\mathbf{1}^T W \mathbf{1}/2 = 1$. At time t , edge $\{i, j\}$ is selected with



(a) The left (right) graph demonstrates the case where $w_s > w_d$ ($w_s < w_d$), in which agents within the same community interacting more (less) often than agents between communities. The width of edges is proportional to the number of interactions.

(b) The adjacency matrix of a graph generated from SBM($n, \nu_1, \nu_2, p_s, p_d$) with $n = 5000$, $\nu_1 = 0.4$, $\nu_2 = 0.6$, $p_s = 5 \log n/n$, $p_d = \log n/n$. Dots represent nonzero entries, so the block structure of the matrix is clearly visible.

Fig. 1. Illustration of the interaction model (4) and an adjacency matrix generated from an SBM.

probability w_{ij} independently of previous updates, and agents update as follows, with the averaging weight $q \in [0, 1]$,

$$X_k(t+1) = \begin{cases} qX_k(t) + (1-q)X_l(t), & k \in \mathcal{V}_r \cap \{i, j\}, \\ & l \in \{i, j\} \setminus k, \\ X_k(t), & \text{otherwise.} \end{cases} \quad (1)$$

Define a sequence of independent and identically distributed (i.i.d.) n -dimensional random matrices $\{R(t), t \in \mathbb{N}\}$ such that $\mathbb{P}\{R(t) = R^{ij}\} = w_{ij}$, where $1 \leq i < j \leq n$ and

$$R^{ij} = \begin{cases} I - (1-q)(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T, & \text{if } i, j \in \mathcal{V}_r, \\ I - (1-q)\mathbf{e}_i(\mathbf{e}_j - \mathbf{e}_i)^T, & \text{if } i \in \mathcal{V}_r, j \in \mathcal{V}_s, \\ I - (1-q)\mathbf{e}_j(\mathbf{e}_i - \mathbf{e}_j)^T, & \text{if } i \in \mathcal{V}_s, j \in \mathcal{V}_r, \\ I, & \text{if } i, j \in \mathcal{V}_s. \end{cases}$$

The update rule (1) can then be written as

$$X(t+1) = R(t)X(t). \quad (2)$$

Since stubborn agents never change their states, we rewrite (2) to end up with the following compact form of the gossip model with stubborn agents:

$$X^r(t+1) = A(t)X^r(t) + B(t)X^s(t), \quad (3)$$

where $X^r(t)$ and $X^s(t)$ are the state vectors obtained from stacking the states of regular and stubborn agents, respectively, $X^s(t) \equiv X^s(0)$, and $[A(t) B(t)]$ is the matrix obtained from stacking rows of $R(t)$ corresponding to regular agents. So $\{[A(t) B(t)], t \in \mathbb{N}\}$ is a sequence of i.i.d. random matrices. Assume that the initial vector $X(0)$ is fixed for simplicity. If $X(0)$ is random, we can study the model by conditioning on realizations of $X(0)$.

2.2. Communities

We follow the framework of SBMs (Abbe, 2017) and Ising blockmodels (Berthet et al., 2019), and assume that agents have pre-assigned community labels. We define a community as the set of agents that have the same label.

In particular, we consider the scenario where the network has two disjoint communities, $\mathcal{V}_1$ and $\mathcal{V}_2$. Denote the community label of i by $\mathcal{C}(i)$, so $\mathcal{C}(i) = k$ for $i \in \mathcal{V}_k$, $k = 1, 2$. We call $\mathcal{C}$ the community structure of the network. We further assume that the interaction probability of the agents i and j with $i \neq j$ is

$$w_{ij} = \begin{cases} w_s, & \text{if } \mathcal{C}(i) = \mathcal{C}(j), \\ w_d, & \text{if } \mathcal{C}(i) \neq \mathcal{C}(j), \end{cases} \quad (4)$$

where $w_s, w_d \in (0, 1)$ and $w_s \neq w_d$. Thus agents in the same community (different communities) interact with probability w_s (w_d). Fig. 1(a) illustrates two different interaction models, via an experiment where a gossip model defined by (4) is run for

2000 iterations and the number of interactions between agents is counted. To ease notation, we assume that $\mathcal{V}_1 = \{1, \dots, n_1\}$ and $\mathcal{V}_2 = \{n_1 + 1, \dots, n_1 + n_2\}$ with $n_k := |\mathcal{V}_k|$, $k = 1, 2$, and $n_1 + n_2 = n$. Thus the interaction probability matrix is

$$W = \begin{bmatrix} w_s \mathbf{1}_{n_1, n_1} - \text{diag}\{w_s \mathbf{1}_{n_1}\} & w_d \mathbf{1}_{n_1, n_2} \\ w_d \mathbf{1}_{n_2, n_1} & w_s \mathbf{1}_{n_2, n_2} - \text{diag}\{w_s \mathbf{1}_{n_2}\} \end{bmatrix}. \quad (5)$$

It has a block structure corresponding to the community structure of the network. The following example illustrates how the preceding assumption arises naturally from an SBM. It shows that a graph generated from an SBM defines an interaction probability matrix close to an averaged version with the same structure as (5).

Example 1. Consider an SBM with two communities, commonly studied in community detection (Abbe, 2017). Such an SBM is a random graph, denoted by SBM($n, \nu_1, \nu_2, p_s, p_d$). Here n is the number of agents, $\nu_1 \in (0, 1)$ (resp. $\nu_2 \in (0, 1)$) is the portion of agents with community label 1 (resp. label 2), where $\nu_1 + \nu_2 = 1$ and $\nu_1 n$ and $\nu_2 n$ are integers, and $p_s, p_d \in (0, 1)$ are the link probabilities between agents in the same and in different communities. We assume $\mathcal{C}(i) = 1$, $1 \leq i \leq \nu_1 n$, and $\mathcal{C}(i) = 2$, $\nu_1 n + 1 \leq i \leq n$.

The SBM($n, \nu_1, \nu_2, p_s, p_d$) randomly generates an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$: for $i \neq j$, $\{i, j\} \in \mathcal{E}$ with probability p_s if $\mathcal{C}(i) = \mathcal{C}(j)$ and with probability p_d if $\mathcal{C}(i) \neq \mathcal{C}(j)$, independently of other edges. Here $\mathcal{A} = [a_{ij}]$ is the adjacency matrix. The graph $\mathcal{G}$ defines a gossip model with the interaction matrix $\tilde{W} = \mathcal{A}/\alpha$ and $\alpha = \sum_{i=1}^n \sum_{j=i+1}^n a_{ij} = |\mathcal{E}|$. The inequality $\|\tilde{W} - \mathbb{E}\{\mathcal{A}\}/\mathbb{E}\{\alpha\}\| \leq C/n$ holds with a constant C , except for a probability vanishing as $n \rightarrow \infty$, if $\log n/n = O(\min\{p_s, p_d\})$ (see Xing et al., 2021). This result implies that, if the network of the gossip model is generated from the SBM, then the interaction probability matrix is close to $\mathbb{E}\{\mathcal{A}\}/\mathbb{E}\{\alpha\}$ when n is large. Note that $\mathbb{E}\{\mathcal{A}\}/\mathbb{E}\{\alpha\}$ has exactly the same structure as W in (5) with $n_k = \nu_k n$, $k = 1, 2$, $w_s = p_s/\mathbb{E}\{\alpha\}$, and $w_d = p_d/\mathbb{E}\{\alpha\}$. Fig. 1(b) demonstrates this concentration phenomenon with an obvious two-block structure.

The concentration indicates that behavior of the gossip model over a graph generated from the SBM may not deviate too far from the gossip model over the averaged graph, when n is large. Xing and Johansson (2022) show that the expected stationary states of the two models are close, if $\log n/n = o(\min\{p_s, p_d\})$. This result indicates that the gossip model over the averaged graph can be considered as an approximation of the model over the SBM, and results for the former model can be extended to the latter model.

Remark 1. A general assumption for community labels in the SBM is that each agent gets a label k with probability ν_k independently of each other, $k = 1, 2$. This is essentially equivalent to the label assignment with deterministic node portions when

$n \rightarrow \infty$ (Remark 3 of Abbe, 2017). Note that it is possible to extend the fixed-label assumption considered in Example 1 to the deterministic-portion assumption, by conditioning on each assignment and using the law of total probability. The condition $\log n/n = O(\min\{p_s, p_d\})$ implies that the expected agent degree is at least $O(\log n)$. In this case, the SBM generates connected graphs with high probability. The difference between p_s and p_d has to be large enough to make exact recovery possible (Abbe, 2017). Here we consider the dynamics over the averaged graph, so the detectability only requires $w_s \neq w_d$ (Assumption 1 (ii)). Future work will study detectability in the SBM case.

2.3. Community recovery and interaction estimation

The considered problem is to recover the community structure and to estimate the interaction probabilities based on state observations, as follows.

Problem. Given a trajectory of the gossip model with the interaction matrix (5), develop an algorithm to jointly recover the community structure $\mathcal{C}$ and estimate the interaction probabilities w_s and w_d .

Remark 2. In the problem, we assume that the developed algorithm uses data coming from the gossip model over the averaged graph. A natural question is how this algorithm performs if it uses a trajectory of the gossip model over a graph sampled from an SBM. In Section 6, we illustrate through simulation that the algorithm performs well also in the SBM case. Such performance is guaranteed by that these two processes behave similarly in terms of their stationary states, as explained in Example 1. We use “community recovery” instead of “community detection” to avoid ambiguity, following the terminology of Berthet et al. (2019), because here agent behavior depends directly on the community structure.

Recall $\mathcal{V}_1 = \{1, \dots, n_1\}$ and $\mathcal{V}_2 = \{n_1 + 1, \dots, n_1 + n_2\}$. We further sort the agents as follows: $\mathcal{V}_{r1} = \{1, \dots, n_{r1}\}$, $\mathcal{V}_{s1} = \{n_{r1} + 1, \dots, n_1\}$, $\mathcal{V}_{r2} = \{n_1 + 1, \dots, n_1 + n_{r2}\}$, and $\mathcal{V}_{s2} = \{n_1 + n_{r2} + 1, \dots, n_1 + n_2\}$. Here, $\mathcal{V}_{rk}$ (resp. $\mathcal{V}_{sk}$) is the set of regular (resp. stubborn) agents in the community k , $k = 1, 2$. Denote $n_{rk} := |\mathcal{V}_{rk}|$, $n_{sk} := |\mathcal{V}_{sk}|$, $n_r := |\mathcal{V}_r|$, and $n_s := |\mathcal{V}_s|$. In the considered problem, the total number of agents is known in advance, the network has two communities, and the stubborn-agent states are observable. But difficulty still remains since n_k , n_{rk} , n_{sk} , $k = 1, 2$, and interaction information are unknown. The interaction information cannot be obtained in general situations (e.g., agent states are only observed at some time steps, or observations are corrupted by noise, as discussed in Remark 7).

3. Model analysis

This section studies model behavior, and provides an explicit expression for the mean of the stationary distribution. Assumptions are summarized as follows.

Assumption 1.

(i.1) The agent set $\mathcal{V}$ consists of two communities, $\mathcal{V}_1 = \{1, \dots, n_1\}$ and $\mathcal{V}_2 = \{n_1 + 1, \dots, n_1 + n_2\}$ with $n_1, n_2 > 0$ and $n_1 + n_2 = n$.

(i.2) Both communities have regular agents, namely, $1 \leq n_{r1} \leq n_1$, $1 \leq n_{r2} \leq n_2$.

(ii) The interaction probability matrix W has a block structure (5) with $w_s, w_d > 0$, $w_s \neq w_d$, and

$$(n_1(n_1 - 1) + n_2(n_2 - 1))w_s + 2n_1n_2w_d = 2. \quad (6)$$

(iii) $X(0)$ is deterministic. It holds that $X^r(0) \in \mathcal{S}$ with

$$\mathcal{S} := \{x^r \in \mathbb{R}^{n_r} : x_i^r \in [\underline{s}, \bar{s}], 1 \leq i \leq n_r\}, \quad (7)$$

where $\underline{s} := \min_{1 \leq i \leq n_s} \{x_i^s\}$, $\bar{s} := \max_{1 \leq i \leq n_s} \{x_i^s\}$, $x^s := X^s(0) = [(x^{s1})^T (x^{s2})^T]^T$ is the stubborn state vector, and x^{sk} is the vector for the community k , $k = 1, 2$.

Remark 3. In Assumption 1 (i.1), the order of agents is sorted for convenience, but we do not know to which group each agent belongs, before community recovery. It is necessary to assume $w_s \neq w_d$. Otherwise, W has no block structure. Regular agents are assumed to start from $\mathcal{S}$, which is reasonable and intuitively means that regular states lie between the extreme stubborn states.

Before studying model behavior, we explicitly write the block structures of $\bar{A} := \mathbb{E}\{A(t)\}$ and $\bar{B} := \mathbb{E}\{B(t)\}$ as follows. The block structure of W results in similar updates of agents in the same community.

Proposition 1. Suppose Assumption 1 holds. Then $\bar{A}$ and $\bar{B}$ have the following block structures:

$$\bar{A} = I_{n_r} - (1 - q) \begin{bmatrix} a_1 I_{n_{r1}} - w_s \mathbf{1}_{n_{r1}, n_{r1}} & -w_d \mathbf{1}_{n_{r1}, n_{r2}} \\ -w_d \mathbf{1}_{n_{r2}, n_{r1}} & a_2 I_{n_{r2}} - w_s \mathbf{1}_{n_{r2}, n_{r2}} \end{bmatrix}, \quad (8)$$

$$\bar{B} = (1 - q) \begin{bmatrix} w_s \mathbf{1}_{n_{r1}, n_{s1}} & w_d \mathbf{1}_{n_{r1}, n_{s2}} \\ w_d \mathbf{1}_{n_{r2}, n_{s1}} & w_s \mathbf{1}_{n_{r2}, n_{s2}} \end{bmatrix},$$

where $a_k = w_s n_k + w_d n_{3-k}$, $k = 1, 2$.

Now we provide the stability and limit theorems of the gossip model.

Theorem 1 (Stability and Limit Theorems). Suppose that Assumption 1 holds and there exists at least one stubborn agent in the network (i.e., $n_r < n$). The following results hold for the gossip model with stubborn agents.

(i) The model has a unique stationary distribution π with mean x^r , and $X^r(t)$ converges in distribution to π .

(ii) The expectation of the state vector converges to x^r :

$$x^r = \lim_{t \rightarrow \infty} \mathbb{E}\{X^r(t)\} = (I - \bar{A})^{-1} \bar{B} x^s. \quad (9)$$

(iii) Denote $S^r(t) := \frac{1}{t} \sum_{i=0}^{t-1} X^r(i)$, then

$$\lim_{t \rightarrow \infty} S^r(t) = x^r \quad \text{a.s.} \quad (10)$$

Remark 4. The first two results show that the agent states, although may not converge a.s., converge in distribution to a unique stationary distribution, and their expectations converge to the mean of the stationary distribution. The third result indicates that we can obtain the value of x^r by computing the state time average.

The next proposition shows that x^r also has a block structure, indicating that regular agents in the same community behave similarly on average.

Proposition 2. Under the conditions of Theorem 1, x^r given in (9) has the form

$$x^r = [\chi_1 \mathbf{1}_{n_{r1}}^T, \chi_2 \mathbf{1}_{n_{r2}}^T]^T, \quad (11)$$

where $\chi_k = (\gamma_{kk} \mathbf{1}_{n_{sk}}^T x^{sk} + \gamma_{k,3-k} \mathbf{1}_{n_{s,3-k}}^T x^{s,3-k}) / \delta$, $\delta = w_s^2 n_{s1} n_{s2} + w_s w_d (n_1 n_{s1} + n_2 n_{s2}) + w_d^2 (n_1 n_2 - n_{r1} n_{r2})$, $\gamma_{kk} = w_s^2 n_{s,3-k} + w_s w_d n_k + w_d^2 n_{r,3-k}$, $\gamma_{k,3-k} = w_d (w_s n_{3-k} + w_d n_k)$, and $\mathbf{1}_{n_{sk}}^T x^{sk} := 0$ if $n_{sk} = 0$, $k = 1, 2$.

Remark 5. Appendix B of Xing et al. (2021) studies a multiple-community case, generalizing Proposition 2.

The above proposition means that regular agents in the same community have the same limit, which is a weighted average of stubborn states. Hence it is possible to split regular agents by computing the state time average. However, we are unable to do so if only one community has stubborn agents, or the stubborn states are similar. The following condition rules out these cases.

Assumption 2. Both communities have stubborn agents (i.e., $n_{s1}n_{s2} > 0$), and $\mathbf{x}^s = [(\mathbf{x}^{s1})^T (\mathbf{x}^{s2})^T]^T$ satisfies that $\mathbf{1}_{n_{s1}}^T \mathbf{x}^{s1}/n_{s1} \neq \mathbf{1}_{n_{s2}}^T \mathbf{x}^{s2}/n_{s2}$.

This assumption has a practical meaning: stubborn agents are distributed among communities, and agents from different communities are more likely to have distinct opinions. Under [Assumption 2](#), we have the following result, indicating that the presence of stubborn agents enhances the separation of regular agents.

Proposition 3. Under the conditions of [Theorem 1](#), $\chi_1 \neq \chi_2$ if and only if [Assumption 2](#) holds.

This result shows that [Assumption 2](#) is a necessary and sufficient condition for regular agents from different communities having nonidentical expected stationary states. Note that $\mathbf{1}_{n_{s1}}^T \mathbf{x}^{s1}/n_{s1} \neq \mathbf{1}_{n_{s2}}^T \mathbf{x}^{s2}/n_{s2}$ is generic (i.e., it holds for almost all $\mathbf{x}^s \in \mathbb{R}^{n_s}$).

4. Joint recovery and estimation algorithm

In this section, we design a joint recovery and estimation algorithm ([Algorithm 1](#)) to address the considered problem. We assume the following connections between stubborn and regular agents. The information means that we have prior knowledge about stubborn agents, which may be gathered from other sources in practice.

Assumption 3. For every stubborn agent $i \in \mathcal{V}_s$, it is known for [Algorithm 1](#) that there exists a regular agent $j_i \in \mathcal{V}_r$ such that i and j_i are in the same community (i.e., $\mathcal{C}(i) = \mathcal{C}(j_i)$).

Now we are ready to introduce [Algorithm 1](#), where we denote the estimates at time t of community label $\mathcal{C}(i)$, interaction probabilities w_s and w_d , by $\hat{\mathcal{C}}(i, t)$, $\hat{w}_s(t)$, and $\hat{w}_d(t)$, respectively. We use $S_i^r(t)$ to represent the $(i - n_1 + n_{r1})$ -th entry of $S^r(t)$, $i \in \mathcal{V}_2 = \{n_1 + 1, \dots, n_1 + n_{r2}\}$ for simplicity. Note that both n_1 and n_{r1} are unknown in the algorithm. In the gossip model, agents randomly interact and update states. [Algorithm 1](#) partitions the agents and estimates interaction strength, out of these state observations, without interaction information.

Remark 6. The difficulty of recovery is to find a quantity revealing the community structure. [Algorithm 1](#) exploits the trajectory data by using $S^r(t)$. From [Proposition 2](#) we know that the entries of $S^r(t)$ converge to two distinct values corresponding to the communities. Hence clustering methods (Line 4 of [Algorithm 1](#), or other methods such as k -means) can be used. For estimation of interaction probabilities, the key is to find consistent parameter equations. Here we use the stationary property $\mathbf{x}^r = \mathbf{A}\mathbf{x}^r + \mathbf{B}\mathbf{x}^s$, giving the following equations of $[x \ y]^T$ (see [Xing et al. \(2021\)](#) for the details)

$$\begin{cases} (n_{s1}\chi_1 - \mathbf{1}_{n_{s1}}^T \mathbf{x}^{s1})x + (n_2\chi_1 - n_{r2}\chi_2 - \mathbf{1}_{n_{s2}}^T \mathbf{x}^{s2})y = 0, \\ (n_1(n_1 - 1) + n_2(n_2 - 1))x + 2n_1n_2y = 2. \end{cases}$$

From (9), the system of equations has a unique solution under [Assumptions 1 and 2](#), for fixed n_k , n_{rk} , and n_{sk} . But these quantities are unknown, so we leverage SA techniques to estimate them, as presented in Line 5 of [Algorithm 1](#). Note that the algorithm does not need to know the averaging weight q .

Algorithm 1 (Joint Recovery and Estimation)

Input: $\{X^r(t), t = 0, 1, 2, \dots\}$, $X^s(0)$, step-size parameter a of the interaction estimator with $a > 0$.

Output: $\{\hat{\mathcal{C}}(i, t)\}$, $\hat{w}_s(t)$, $\hat{w}_d(t)$.

- 1: Randomize $\hat{\mathcal{C}}(i, 0)$, $\hat{w}_s(0)$, $\hat{w}_d(0)$, set $S^r(0) = X^r(0)$.
- 2: **for** $t = 1, \dots$ **do**
- 3: Compute

$$\begin{aligned} S^r(t) &= \frac{t}{t+1} S^r(t-1) + \frac{1}{t+1} X^r(t), \\ \bar{S}^r(t) &= \frac{1}{n_r} \mathbf{1}_{n_r}^T S^r(t). \end{aligned}$$

- 4: **Community recovery:**

$$\hat{\mathcal{C}}(i, t) = 2 - \mathbb{I}_{[S_i^r(t) > \bar{S}^r(t)]}, \quad i \in \mathcal{V}_r,$$

$$\hat{\mathcal{C}}(i, t) = \hat{\mathcal{C}}(j_i, t), \quad i \in \mathcal{V}_s,$$

where j_i is defined in [Assumption 3](#).

- 5: **Interaction estimation:**

$$\begin{aligned} \hat{w}_s(t) &= \hat{w}_s(t-1) - \frac{a}{t} \text{sgn}(g(t)) \left(g(t) \hat{w}_s(t-1) + \frac{h_2(t)}{\hat{n}_1(t) \hat{n}_2(t)} \right), \\ \hat{w}_d(t) &= \frac{2 - \hat{w}_s(t)(\hat{n}_1^2(t) + \hat{n}_2^2(t) - \hat{n}_1(t) - \hat{n}_2(t))}{2\hat{n}_1(t)\hat{n}_2(t)}, \end{aligned}$$

where

$$g(t) = h_1(t) - \frac{\hat{n}_1^2(t) + \hat{n}_2^2(t) - \hat{n}_1(t) - \hat{n}_2(t)}{2\hat{n}_1(t)\hat{n}_2(t)} h_2(t),$$

$$h_1(t) = \frac{|\hat{\mathcal{V}}_{s1}(t)|}{|\hat{\mathcal{V}}_{r1}(t)|} \sum_{i \in \hat{\mathcal{V}}_{r1}(t)} S_i^r(t) - \sum_{i \in \hat{\mathcal{V}}_{s1}(t)} X_i^s(0),$$

$$h_2(t) = \frac{\hat{n}_2(t)}{|\hat{\mathcal{V}}_{r1}(t)|} \sum_{i \in \hat{\mathcal{V}}_{r1}(t)} S_i^r(t) - \sum_{i \in \hat{\mathcal{V}}_{r2}(t)} S_i^r(t) - \sum_{i \in \hat{\mathcal{V}}_{s2}(t)} X_i^s(0),$$

$$\hat{n}_k(t) = \sum_{i \in \mathcal{V}} \mathbb{I}_{[\hat{\mathcal{C}}(i, t) = k]},$$

$$\hat{\mathcal{V}}_{rk}(t) = \{i \in \mathcal{V}_r : \hat{\mathcal{C}}(i, t) = k\},$$

$$\hat{\mathcal{V}}_{sk}(t) = \{i \in \mathcal{V}_s : \hat{\mathcal{C}}(i, t) = k\}, \quad k = 1, 2.$$

- 6: **end for**

5. Convergence analysis

This section studies the performance of [Algorithm 1](#). The following result means that communities can be recovered in finite time, and the interaction probability estimates are convergent.

Theorem 2 (Convergence of [Algorithm 1](#)).

Under [Assumptions 1–3](#), the following holds.

(i) The community recovery is achieved in finite time; there exists a positive integer-valued random variable T such that $\hat{\mathcal{C}}(i, t) = \mathcal{C}(i)$, for all $i \in \mathcal{V}$ and $t > T$.

(ii) The interaction estimator converges a.s., namely,

$$\mathbb{P} \left\{ \lim_{t \rightarrow \infty} (\hat{w}_s(t), \hat{w}_d(t)) = (w_s, w_d) \right\} = 1.$$

Remark 7. Since [Algorithm 1](#) uses the property (10), it can also deal with situations where state observations are corrupted. For example, one cannot observe the entire trajectory but can only sample states at some time steps. Ergodic property ensures that the time average of the sampled states still converges, if the sampling process is independent of the update and the number of samples tends to infinity ([Ravazzi, Frasca, Tempo, & Ishii, 2014; Wai et al., 2016](#)). Another situation is that the observations are disturbed by i.i.d. zero-mean noise. The law of large numbers guarantees that the influence of noise vanishes over time.

Now we investigate the sample complexity of the community recovery, and the convergence rate of the interaction estimator.

The following result is useful for studying the sample complexity of the recovery.

Lemma 1. Consider a Markov chain $\{X(t)\}$ taking values on a compact state space $\mathcal{X}$ and having a unique stationary distribution π . For a function $f : \mathcal{X} \rightarrow \mathbb{R}$ and $\alpha := \int_{\mathcal{X}} f(x)\pi(dx)$, denote $g(x) := \sum_{t=0}^{\infty} \mathbb{E}[f(X(t)) - \alpha | X(0) = x]$, and the supremum of $|g|$ on $\mathcal{X}$ by $\|g\|_s := \sup\{|g(x)| : x \in \mathcal{X}\}$. If $\|g\|_s < \infty$, then, for all $\varepsilon > 0$ and $t > 2\|g\|_s/\varepsilon$, it holds for $S_f(t) := \frac{1}{t} \sum_{i=0}^{t-1} f(X(i))$ that

$$\mathbb{P}[|S_f(t) - \alpha| \geq \varepsilon] \leq 2 \exp \left\{ -\frac{(t\varepsilon - 2\|g\|_s)^2}{2t\|g\|_s^2} \right\}.$$

Remark 8. Similar concentration results to Lemma 1 have been obtained in the literature for other models. One class of results leverage Markov chain approaches and normally require stability such as uniform ergodicity (Glynn & Ormoneit, 2002; Paulin, 2015) or explicit bounds of the derivative of the initial measure with respect to the stationary measure (Fan, Jiang, & Sun, 2021). It is hard to derive these properties for Markov chains without continuous distributions (Gibbs & Su, 2002), as in our case. Another line of research studies concentration of Polyak averages, and contains step-size conditions (Mou, Li, Wainwright, Bartlett, & Jordan, 2020), which cannot be applied to our problem either.

Using the preceding lemma, we are able to compute when the differences between entries of $S^r(t)$ and $\mathbf{x}^r$ are small enough, such that agents in different communities have distinct state time averages. As a result, we obtain a sample-complexity result for the community recovery. The next theorem shows that the probability of recovering communities successfully depends on the network, the interaction probabilities, and the stubborn states. The probability tends to one as t goes to infinity.

Theorem 3 (Sample Complexity).

Under the conditions of Theorem 2, for the community recovery step of Algorithm 1, it holds that, for $t > t_0$,

$$\mathbb{P}\{\hat{C}(i, t) = C(i), \forall i \in \mathcal{V}\} \geq 1 - 2n_r \exp \left\{ \frac{-2(t - t_0)^2}{t_0^2 t} \right\},$$

with $t_0 = 4\delta c_{\bar{A}} c_{n_r} c_s / c_w$, where $c_{\bar{A}} = 1/(1 - \rho(\bar{A}))$, $c_{n_r} = n_r^{3/2}(n_r + 1)$, $c_s = \max\{|\underline{s}|, |\bar{s}|\} / |n_{s1} \mathbf{1}_{n_{s2}}^T \mathbf{x}^{s2} - n_{s2} \mathbf{1}_{n_{s1}}^T \mathbf{x}^{s1}|$, $c_w = |w_s^2 - w_d^2|$, δ is given in Proposition 2, and $\underline{s}$ and $\bar{s}$ are given in (7).

Remark 9. This result provides a sample complexity characterization for recovering community from a single trajectory. Multiple-trajectory sample complexity is investigated by Roddenberry et al. (2020), Schaub et al. (2020) and Wai et al. (2019). The parameter δ reflects the combined effect of the cardinality of stubborn and regular agents and the interaction probabilities. $c_{\bar{A}}$ captures the “speed” of information diffusion, and increases with $\rho(\bar{A})$. c_{n_r} depends on the number of regular agents. c_s increases

with the range of the states and decreases with the difference of averaged stubborn states in different communities, and c_w measures the difference between interaction probabilities within and between communities. Smaller δ , $c_{\bar{A}}$, n_r , and $\max\{|\underline{s}|, |\bar{s}|\}$ would make the recovery easier, and so would larger c_w and $|n_{s1} \mathbf{1}_{n_{s2}}^T \mathbf{x}^{s2} - n_{s2} \mathbf{1}_{n_{s1}}^T \mathbf{x}^{s1}|$. For the gossip model over an SBM, Example 1 indicates that the algorithm can recover most community labels, which is illustrated in Section 6.

We have the following result for the convergence rate of the interaction estimator. It shows that the rate also depends on the model parameters, and a large enough step-size a ensures that the rate can achieve $O(1/\sqrt{t})$.

Theorem 4 (Convergence Rate).

Under the conditions of Theorem 2, it holds for $d_0 \in [0, \min\{1/2, a|\eta|\})$ that

$$[\hat{w}_s(t) - w_s, \hat{w}_d(t) - w_d]^T = o(t^{-d_0}), \text{ a.s.,}$$

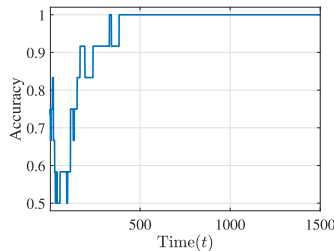
where $a > 0$ is the step-size parameter given in Algorithm 1, $\eta = (w_s n_2 + w_d n_1)(n_{s1} \mathbf{1}_{n_{s2}}^T \mathbf{x}^{s2} - n_{s2} \mathbf{1}_{n_{s1}}^T \mathbf{x}^{s1}) / (\delta n_1 n_2)$, and δ is given in Proposition 2.

Remark 10. In the theorem, η increases with the combined effect of the number of agents and the interaction probabilities (i.e., $(w_s n_2 + w_d n_1)/\delta$) and with the disagreement of stubborn agents, and decreases with the cardinality of each community. When $a \geq 1/(2|\eta|)$, the estimator achieves its optimal rate. Larger η provides a wider selection range. Simulation in Section 6 shows that the algorithm using a trajectory from the gossip model over an SBM can estimate the ratio of the link probabilities.

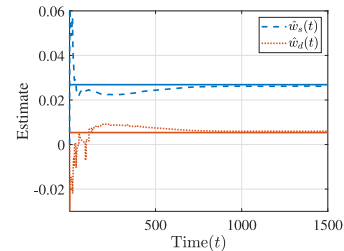
6. Numerical simulation

This section illustrates the performance of Algorithm 1, conducts an algorithm comparison, and applies Algorithm 1 to the SBM case and a real network.

To illustrate the performance of Algorithm 1 under Assumptions 1–3, consider a network consisting of twelve agents. The two communities both have five regular agents and one stubborn agent. Set interaction probabilities be $w_s = 5/186$ and $w_d = 1/186$. The stubborn agent in community 1 (resp. community 2) has state 1 (resp. -1). The initial states of regular agents are drawn from uniform distribution on $(-1, 1)$. The averaging weight is set to be $q = 1/2$ in all experiments. Fig. 2(a) shows that Algorithm 1 recovers the communities in finite time, where the accuracy at time t is defined by $\frac{1}{n}(\max_{\sigma \in S_2} \{\sum_{i=1}^n \mathbb{I}_{[\sigma(\hat{C}(i,t))=C(i)]}\}) \in [0, 1]$. Here $\sigma : \{1, 2\} \rightarrow \{1, 2\}$ is a permutation function (to prevent a reverse distribution of labels), S_2 is the group of permutations on $\{1, 2\}$, $C(i)$ is agent i 's community label, $\hat{C}(i, t)$ is the estimate of agent i 's label at time t , and $n = 12$. Consistency of the interaction estimator with step-size parameter $a = 1$ is demonstrated in Fig. 2(b). These results validate Theorem 2.



(a) Finite-time community recovery. The recovery is achieved after $t = 383$.



(b) Convergence of the interaction estimator. The solid (dashed) lines are true values (estimates).

Fig. 2. Performance of Algorithm 1.

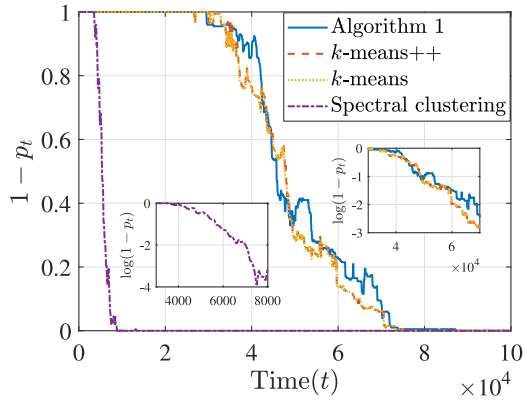
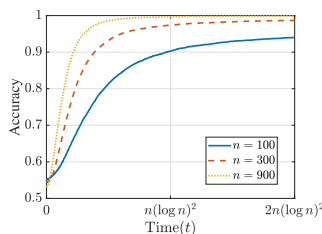


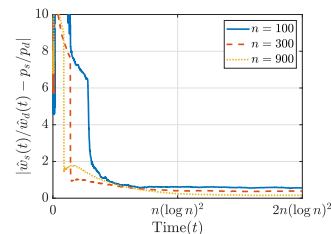
Fig. 3. Performance comparison of four methods.

We now show the sample complexity of the community recovery (Theorem 3) and compare the recovery step with the k -means, k -means++ (Arthur & Vassilvitskii, 2006), and spectral clustering methods (Abbe, 2017). This experiment considers the gossip model under Assumptions 1–3 with $n = 400$, $n_1 = 150$, and $n_{s1} = n_{s2} = 8$. Let $w_s/w_d = 5$ and solve the two parameters from (6). Let stubborn agents in community 1 (resp. community 2) have state 1 (resp. -1), and generate the initial states of other agents from uniform distribution on $(-1, 1)$. By running the algorithms for 200 times, we obtain the relative frequency that the algorithms recover all community labels, defined by $p_t := (\sum_{k=1}^N \max_{\sigma \in S_2} \{\mathbb{I}_{[\sigma(\hat{c}_k(i,t))=C(i), \forall i \in V]}\})/N$, where $N = 200$ and $\hat{c}_k(i, t)$ is the estimate of agent i 's label at time t in the k th run. After computing the time average $S^r(t)$, we use k -means and k -means++ with $k = 2$ instead of Line 4 of Algorithm 1, to recover communities. To implement the spectral clustering method, assume that edge activation is known, and use the activation information to estimate the interaction probability matrix W . Applying spectral clustering to estimates of W obtains community estimates.

Fig. 3 shows that the probabilities of unsuccessful community recovery of all approaches tends to zero exponentially over time. The spectral clustering method performs better than other algorithms, because it directly uses interaction information, but the required time is still of the same order as the other algorithms. The k -means and k -means++ methods perform similarly to each other, and also similarly to Algorithm 1. This observation indicates that the major challenge of the problem is how to use agent states to recover communities without topological information.



(a) Averaged accuracy of community recovery for each SBM.



(b) Median of estimation error of p_s/p_d for each SBM.

Fig. 4. Performance of Algorithm 1 using trajectories of the gossip model over sampled graphs from SBMs with $n = 100, 300, 900$.

We now consider the case where trajectories of the gossip model over graphs sampled from SBMs are given to Algorithm 1. We use three SBMs with size $n = 100, 300, 900$ and with two equal-sized communities ($v_1 = v_2 = 0.5$). Set $n_{r1} = n_{r2} = 0.45n$, and $n_{s1} = n_{s2} = 0.05n$. Let the link probability in the same community be $p_s = (\log n)^2/n$ and the link probability between different communities be $p_d = (\log n)/n$. For each SBM, we generate 20 graph samples. For each graph sample, we run Algorithm 1 for 20 times. Regular states are generated the same as earlier and stubborn agents in community 1 (resp. community 2) have state 1 (resp. -1). Fig. 4(a) shows that Algorithm 1 has high community recovery accuracy, increasing with n . This phenomenon results from the concentration discussed in Example 1. Algorithm 1 outputs $\hat{w}_s(t)$ and $\hat{w}_d(t)$ as estimates of the two distinct non-zero values of $\mathbb{E}\{\mathcal{W}\}/\mathbb{E}\{\alpha\}$. Note that $[cp_s, cp_d]$ defines the same $\mathbb{E}\{\mathcal{W}\}/\mathbb{E}\{\alpha\}$ for all $c > 0$, so we can only estimate the ratio p_s/p_d without knowing the expected number of edges of the SBM. Fig. 4(b) shows that the median of the estimation error for trajectory samples from each SBM is close to zero and decreases with n .

Zachary's karate club network (Zachary, 1977), presented in Fig. 5(a), is used to demonstrate an application of Algorithm 1. An edge represents frequent interaction between the two agents. The strength of interactions between agents is modeled by a weighted adjacency matrix. A conflict between agents 1 and 34 results in a fission of the club. In the experiment, we assume that only the opinions can be observed, instead of interactions between agents. The process is modeled by the gossip model. Agents 1 and 34 are set to be stubborn agents holding different opinions. In addition, one edge in Fig. 5(a) is selected at each time with a probability proportional to interaction strength given in Zachary (1977). The goal is to partition the agents into communities based on only state observations. The network structure departs from our assumptions, but the result shown in Fig. 5(b) indicates that our algorithm can finally recover the community structure as time increases, without topological and interaction information.

7. Conclusion and future work

In this paper, we developed a joint algorithm to recover the community structure and to estimate the interaction probabilities for gossip opinion dynamics. It was proved that the community recovery is achieved in finite time, and the interaction estimator converges almost surely. We analyzed the sample complexity of the recovery and the convergence rate of the estimator. Future work includes to study the case where all regular agents have the same stationary expectation, and to analyze the community detection problem for dynamics over the SBM.



(a) The community structure of Zachary's karate club network. Red squares and green triangles show two communities. (b) Accuracy of community recovery of Algorithm 1 for the gossip model over Zachary's karate club network.

Fig. 5. Numerical experiment over Zachary's karate club network.

Acknowledgments

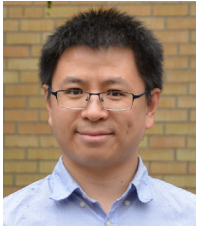
The authors would like to thank the anonymous reviewers for their insightful comments and suggestions.

References

- Abbe, E. (2017). Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(1), 6446–6531.
- Acemoglu, D., Como, G., Fagnani, F., & Ozdaglar, A. (2013). Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, 38(1), 1–27.
- Arthur, D., & Vassilvitskii, S. (2006). *k*-means++: The advantages of careful seeding. *Stanford*.
- Berthet, Q., Rigollet, P., & Srivastava, P. (2019). Exact recovery in the Ising blockmodel. *The Annals of Statistics*, 47(4), 1805–1834.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), Article P10008.
- Boyd, S., Ghosh, A., Prabhakar, B., & Shah, D. (2006). Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6), 2508–2530.
- Cowan, S. K., & Baldassarri, D. (2018). 'It could turn ugly': Selective disclosure of attitudes in political discussion networks. *Social Networks*, 52, 1–17.
- De Montjoye, Y.-A., Gams, S., Blondel, V., Canright, G., De Cordes, N., Dele-taille, S., et al. (2018). On the privacy-conscious use of mobile phone data. *Scientific Data*, 5(1), 1–6.
- Fan, J., Jiang, B., & Sun, Q. (2021). Hoeffding's inequality for general Markov chains and its applications to statistical learning. *Journal of Machine Learning Research*, 22(139), 1–35.
- Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, 659, 1–44.
- Gibbs, A. L., & Su, F. E. (2002). On choosing and bounding probability metrics. *International Statistical Review*, 70(3), 419–435.
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), 7821–7826.
- Glynn, P. W., & Ormoneit, D. (2002). Hoeffding's inequality for uniformly ergodic Markov chains. *Statistics & Probability Letters*, 56(2), 143–146.
- Morarescu, I.-C., & Girard, A. (2010). Opinion dynamics with decaying confidence: Application to community detection in graphs. *IEEE Transactions on Automatic Control*, 56(8), 1862–1873.
- Mou, W., Li, C. J., Wainwright, M. J., Bartlett, P. L., & Jordan, M. I. (2020). On linear stochastic approximation: Fine-grained Polyak-Ruppert and non-asymptotic concentration. In *Conference on learning theory* (pp. 2947–2997).
- Netrapalli, P., & Sanghavi, S. (2012). Learning the graph of epidemic cascades. *ACM SIGMETRICS Performance Evaluation Review*, 40(1), 211–222.
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), Article 026113.
- Paulin, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Probability*, 20, 1–32.
- Peixoto, T. P. (2019). Network reconstruction and community detection from dynamics. *Physical Review Letters*, 123(12), Article 128301.
- Prokhorenkova, L., Tikhonov, A., & Litvak, N. (2022). When less is more: Systematic analysis of cascade-based community detection. *ACM Transactions on Knowledge Discovery from Data*, 16(4), 1–22.
- Proskurnikov, A. V., & Tempo, R. (2017). A tutorial on modeling and analysis of dynamic social networks. Part I. *Annual Reviews in Control*, 43, 65–79.
- Proskurnikov, A. V., Tempo, R., Cao, M., & Friedkin, N. E. (2017). Opinion evolution in time-varying social influence networks with prejudiced agents. *IFAC-PapersOnLine*, 50(1), 11896–11901.
- Ramezani, M., Khodadadi, A., & Rabiee, H. R. (2018). Community detection using diffusion information. *ACM Transactions on Knowledge Discovery from Data*, 12(2), 1–22.
- Ramos, M., Shao, J., Reis, S. D., Anteneodo, C., Andrade, J. S., Havlin, S., et al. (2015). How does public opinion become extreme? *Scientific Reports*, 5(1), 1–14.
- Ravazzi, C., Dabbene, F., Lagoa, C., & Proskurnikov, A. V. (2021). Learning hidden influences in large-scale dynamical social networks: A data-driven sparsity-based approach, in memory of Roberto Tempo. *IEEE Control Systems Magazine*, 41(5), 61–103.
- Ravazzi, C., Frasca, P., Tempo, R., & Ishii, H. (2014). Ergodic randomized algorithms and dynamics over networks. *IEEE Transactions on Control of Network Systems*, 2(1), 78–87.
- Roddenberry, T. M., Schaub, M. T., Wai, H.-T., & Segarra, S. (2020). Exact blind community detection from signals on multiple graphs. *IEEE Transactions on Signal Processing*, 68, 5016–5030.
- Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4), 1118–1123.
- Sarkar, T., Rakhlin, A., & Dahleh, M. A. (2021). Finite time LTI system identification. *Journal of Machine Learning Research*, 22(26), 1–61.
- Schaub, M. T., Segarra, S., & Tsitsiklis, J. N. (2020). Blind identification of stochastic block models from dynamical observations. *SIAM Journal on Mathematics of Data Science*, 2(2), 335–367.
- Tian, Y., & Wang, L. (2018). Opinion dynamics in social networks with stubborn agents: An issue-based perspective. *Automatica*, 96, 213–223.
- Wai, H.-T., Scaglione, A., & Leshem, A. (2016). Active sensing of social networks. *IEEE Transactions on Signal and Information Processing over Networks*, 2(3), 406–419.
- Wai, H.-T., Segarra, S., Ozdaglar, A. E., Scaglione, A., & Jadbabaie, A. (2019). Blind community detection from low-rank excitations of a graph filter. *IEEE Transactions on Signal Processing*, 68, 436–451.
- Xing, Y., He, X., Fang, H., & Johansson, K. H. (2020). Community detection for gossip dynamics with stubborn agents. In *IEEE conference on decision and control* (pp. 4915–4920).
- Xing, Y., He, X., Fang, H., & Johansson, K. H. (2021). Community structure recovery and interaction probability estimation for gossip opinion dynamics. *arXiv preprint arXiv:2102.09683*.
- Xing, Y., & Johansson, K. H. (2022). A concentration phenomenon in a gossip interaction model with two communities. In *European control conference* (pp. 1126–1131).
- Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4), 452–473.



Yu Xing received his B.S. degree in psychology from Peking University in 2014, and his Ph.D. degree in operations research and control theory at Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2020. Since Sep. 2020 he has been a postdoctoral researcher at the Division of Decision and Control Systems, KTH Royal Institute of Technology, Sweden. His research interests include system identification, social opinion dynamics, network inference, and community detection.



Xingkang He was a postdoctoral researcher in the Division of Decision and Control Systems, KTH Royal Institute of Technology, Sweden. He received his Ph.D. degree at the University of Chinese Academy of Sciences in 2018. His research interests include security and privacy of cyber-physical systems, networked control systems, robotics, and machine learning.



Haitao Fang received his B.S. degree in probability and statistics and M.S. degree in applied mathematics in 1990 from the Peking University and 1993 from Tsinghua University, respectively. He received Ph.D. degree in applied mathematics in 1996 from the Peking University. From 1996 to 1998 he was a Postdoc at the Institute of Systems Science, Chinese Academy of Sciences (CAS), and joined the Institute as an Assistant Professor in 1998. He now is with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, CAS as a Full Professor. He currently

serves as Deputy Editor-in Chief of "Journal of Control Theory and Applications" and Associate Editor of "Acta Mathematicae Applicatae Sinica". His research interests include dynamical behavior analysis of networked systems and distributed system estimation, optimization and control.



Karl Henrik Johansson is Professor with the School of Electrical Engineering and Computer Science at KTH Royal Institute of Technology in Sweden and Director of Digital Futures. He received M.Sc. degree in Electrical Engineering and Ph.D. in Automatic Control from Lund University. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests are in networked control systems and cyber-physical systems with applications in transportation, energy, and automation networks. He is President of the European Control

Association and member of the IFAC Council, and has served on the IEEE Control Systems Society Board of Governors and the Swedish Scientific Council for Natural Sciences and Engineering Sciences. He has received several best paper awards and other distinctions from IEEE, IFAC, and ACM. He has been awarded Swedish Research Council Distinguished Professor, Wallenberg Scholar with the Knut and Alice Wallenberg Foundation, Future Research Leader Award from the Swedish Foundation for Strategic Research, the triennial IFAC Young Author Prize, and IEEE Control Systems Society Distinguished Lecturer. He is Fellow of the IEEE and the Royal Swedish Academy of Engineering Sciences.