2022 European Control Conference (ECC)
July 12-15, 2022. London, United Kingdom

# A Concentration Phenomenon in a Gossip Interaction Model with Two Communities

Yu Xing and Karl H. Johansson

*Abstract*— **We study a concentration phenomenon in a gossip model that evolves over a stochastic block model (SBM) with two communities. We study the conditional mean of the stationary distribution of the gossip model over the SBM, and show that it is close to the mean of the stationary distribution of the gossip model over an averaged graph, with high probability. As a consequence, regular (non-stubborn) agents in the same community of the gossip model over the SBM have stationary states with similar expectations. The results show that it is possible to use the gossip model over the averaged graph to approximate and analyze the gossip model over the SBM, and establish a correspondence between agent states and community structure of a network. We present numerical simulations to illustrate the results.**

## I. Introduction

A crucial problem in community detection is whether community structure of a network corresponds to labels that come from non-topological information [1]. For large-scale networks, research finds that the correspondence hardly holds [2]. To deal with the disagreement of community structure and non-topological labels, most existing papers focus on developing approaches that integrate both information. The papers [3], [4] utilize Bayesian methods for the integration, which compute the correlation between community structure and non-topological labels, and use the labels when the correlation is high. The authors in [5] further introduce statistical techniques to exploit non-topological labels when the labels correlate only weakly with community structure. However, it should be noted that many realistic networks have labels that are possibly generated from underlying dynamics. For example, in Zachary's karate club network [6], the labels of agents come from a group fission, which could be a consequence of an opinion formation process. Ignorant of complex dynamics evolving over a network, we may fail to predict non-topological labels, even when we have the complete knowledge of the network. Thus, there is a need to investigate behavior of complex dynamics over networks with community structure. This type of results may deepen our understanding of how community structure links to labels with non-topological information and how to make predictions from the community structure.

A related problem in the field of opinion dynamics is how to theoretically characterize large-scale opinion landscapes [7], [8]. Many papers use statistical indices such as the variance of agent states, and local and global disagreement, to characterize the separation of agent states [9], [10], [11], but these indices can provide only vague descriptions. The paper [12] is one of the earliest researches endeavoring to establish more precise characterizations of agent states (i.e., an entry-wise description of the agent-state vector). The authors study a gossip model with stubborn agents, and show that, if the network is highly fluid and the size of the network is large, almost all regular (non-stubborn) agents have stationary states with similar expectations. The paper [13] studies the DeGroot model (a deterministic counterpart of the gossip model) with stubborn agents over a weighted complete graph, and provides conditions for agent states gathering around two extreme points. However, there is still a need to establish a precise characterization of agent states for this type of models over general graphs.

In this paper, we study a concentration phenomenon in the gossip model that evolves over a stochastic block model (SBM) with two communities. In the literature, concentration means that a random variable that smoothly depends on the equal influence of many independent variables is close to a constant with high probability [14]. Here we borrow the term to refer to the phenomenon that an agent-state vector depending on the SBM is close to a constant vector with high probability. Our contributions are summarized as follows.

We show that the conditional mean of the stationary distribution of the gossip model over the SBM is close to that of the gossip model over an averaged graph with high probability (Theorem 1), by using concentration inequalities of random variables and random matrices. As a consequence of this result, we show that regular agents in the same community of the gossip model over the SBM have stationary states with similar expectations (Corollary 1).

The studied model captures random opinion evolution over a network with community structure. The obtained results provide a possible explanation of how community structure influences group dynamics, which assign agents with non-topological labels. Also, the analysis shows that it is possible to use the gossip model over an averaged graph to approximate and study the gossip model over the SBM. Since it is much easier to analyze the former model, this finding provides insight into studying precise characterization of agent states for complex dynamics. The discovered correspondence between community structure and agent states can also inspire the design of online community-detection algorithm based on state observations [15].

This work was supported by the Knut & Alice Wallenberg Foundation and the Swedish Research Council. The authors are with Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, and with Digital Futures, Stockholm, Sweden. Email: {yuxing2,kallej}@kth.se

### A. Outline

The rest of the paper is organized as follows. In Section II, we introduce the definitions and properties of the SBM and the gossip model, and in Section III we formulate the problem. Section IV presents our main results. We provide numerical experiments in Section V and conclude the paper in Section VI.

### B. Notation

Denote the $n$-dimensional Euclidean space by $\mathbb{R}^n$, the set of $n \times m$ real matrices by $\mathbb{R}^{n\times m}$, the set of nonnegative integers by $\mathbb{N}$, and $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. Let $\mathbf{1}_n$ be the all-one vector with dimension $n$, $e_1, \ldots, e_n$ be the canonical basis of $\mathbb{R}^n$, $I_n$ be the $n\times n$ identity matrix (we omit $n$ if there is no confusion). Denote both the Euclidean norm of a vector and the spectral norm of a matrix by $\|\cdot\|$. For a vector $x \in \mathbb{R}^n$, denote its $i$-th entry by $x_i$, and for a matrix $A \in \mathbb{R}^{n\times n}$, denote its $(i,j)$-th entry by $a_{ij}$ or $[A]_{ij}$. The cardinality of a set $\Omega$ is denoted by $|\Omega|$. For two sequences of real numbers $\{a_k\}$ and $\{b_k\}$ with $b_k \neq 0$, $k \in \mathbb{N}^+$, we denote $a_k \sim b_k$ if $\lim_{k\to\infty} a_k/b_k = 1$, $a_k = O(b_k)$ if $|a_k/b_k| \leq C$ for all $k \in \mathbb{N}^+$ and some positive constant $C$, $a_k = o(b_k)$ if $\lim_{k\to\infty} |a_k/b_k| = 0$, and $a_k = \omega(b_k)$ if $b_k = o(a_k)$. We call an event happens almost surely (a.s.) if it happens with probability one, and call an event depending on a parameter $n \in \mathbb{N}$ happens with high probability, if the probability that this event happens tends to one as $n \to \infty$. Denote the expectation of a random vector $X$ by $\mathbb{E}\{X\}$. Denote an undirected graph by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, where $\mathcal{V}$ is the agent set, $\mathcal{E}$ is the edge set, and $A = [a_{ij}]$ is the adjacency matrix (i.e., $a_{ij} = a_{ji} = 1$ if $\{i,j\} \in \mathcal{E}$, and $a_{ij} = a_{ji} = 0$ otherwise).

## II. Preliminaries

In this section, we introduce the definitions of the SBM and the gossip model.

### A. Stochastic Block Model

The SBM is a canonical model studied in the statistical framework for community detection [16]. It is a random graph model with community labels. In this paper we consider a simplified two-community version of the SBM:

*Definition 1 (Stochastic block model):* A stochastic block model is a random graph model with parameters $n$, $l_s$, and $l_d$, denoted by $\mathrm{SBM}(n, \boldsymbol{l})$, where $\boldsymbol{l} = [l_s\ l_d]^T$. Here $n$ is an even number, and $l_s$ and $l_d$ are parameters in $(0,1)$ depending on $n$. $\mathrm{SBM}(n,\boldsymbol{l})$ generates an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ without self-loops, where $|\mathcal{V}| = n$, according to the following rule: in the first step, the model assigns half of the agents with community label 1, and the other half with label 2 (we denote the community label of agent $i$ by $\mathcal{C}_i$, $1 \leq i \leq n$); in the second step, for all $i, j \in \mathcal{V}$ with $i \neq j$, the model adds edge $\{i,j\}$ to $\mathcal{E}$ with probability $p_{ij}$ independently of other edges, where $p_{ij} = l_s$ if $\mathcal{C}_i = \mathcal{C}_j$ and $p_{ij} = l_d$ if $\mathcal{C}_i \neq \mathcal{C}_j$. ■

### B. Gossip Model with Stubborn Agents

The gossip model with stubborn agents is a random process evolving over a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ with $|\mathcal{V}| = n \geq 2$.

The underlying graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ of the gossip model is undirected and has no self-loops. In addition, $\mathcal{V}$ contains two types of agents, regular and stubborn, denoted by $\mathcal{V}_r$ and $\mathcal{V}_s$, respectively ($\mathcal{V} = \mathcal{V}_r \cup \mathcal{V}_s$ and $\mathcal{V}_r \cap \mathcal{V}_s = \emptyset$). Each agent $i$ in the graph possesses a state $Z_i(t) \in \mathbb{R}$, $t \in \mathbb{N}$. Stacking all agent states, we denote the state vector at time $t$ by $Z(t) \in \mathbb{R}^n$. We assume that the model starts with a deterministic initial vector $Z(0)$ for simplicity.

The random interaction of the gossip model is captured by an interaction probability matrix $W = [w_{ij}] \in \mathbb{R}^{n\times n}$ satisfying that $w_{ij} = w_{ji} = a_{ij}/\alpha$, where $\alpha = \sum_{i=1}^n \sum_{j=i+1}^n a_{ij}$. Hence $\mathbf{1}^T W \mathbf{1}/2 = 1$.

At time $t$, edge $\{i,j\}$ is selected with probability $w_{ij}$ independently of previous updates, and agents update as follows,

$$Z_k(t+1) = \begin{cases} \frac{1}{2}(Z_i(t) + Z_j(t)), & \text{if } k \in \mathcal{V}_r \cap \{i,j\}, \\ Z_k(t), & \text{otherwise.} \end{cases} \quad (1)$$

The averaging weight is assumed to be $1/2$, but general weights can be considered. For $1 \leq i < j \leq n$, define

$$V^{ij} = \begin{cases} I - \frac{1}{2}(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T, & \text{if } i,j \in \mathcal{V}_r, \\ I - \frac{1}{2}\mathbf{e}_i(\mathbf{e}_i - \mathbf{e}_j)^T, & \text{if } i \in \mathcal{V}_r, j \in \mathcal{V}_s, \\ I - \frac{1}{2}\mathbf{e}_j(\mathbf{e}_j - \mathbf{e}_i)^T, & \text{if } i \in \mathcal{V}_s, j \in \mathcal{V}_r, \\ I, & \text{if } i,j \in \mathcal{V}_s, \end{cases}$$

and a sequence of independent and identically distributed (i.i.d.) $n$-dimensional random matrices $\{V(t),\ t \in \mathbb{N}\}$ such that $\mathbb{P}\{V(t) = V^{ij}\} = w_{ij}$, $1 \leq i < j \leq n$. The compact form of update rule (1) is

$$Z(t+1) = V(t)Z(t). \quad (2)$$

Since stubborn agents never change their states during the process, we rewrite (2) and obtain the following compact form of the gossip model:

$$X(t+1) = Q(t)X(t) + R(t)\mathbf{z}^s, \quad (3)$$

where $X(t)$ and $\mathbf{z}^s$ are the state vectors obtained by stacking the states of regular and stubborn agents, respectively, and $[Q(t)\ R(t)]$ is the matrix obtained by stacking rows of $V(t)$ corresponding to regular agents.

Denote $\bar{Q} := \mathbb{E}\{Q(t)\}$ and $\bar{R} := \mathbb{E}\{R(t)\}$. We here present some basic properties of the gossip model for completeness:

*Proposition 1:* (Stability and limit theorems) Suppose that $\mathcal{G}$ is connected and there exists at least one stubborn agent in the network. The following results hold for the gossip model (3).
(i) The model has a unique stationary distribution $\pi$ with mean $\mathbf{x}$, and $X(t)$ converges in distribution to $\pi$, as $t \to \infty$.
(ii) The expectation of the regular-agent state vector converges to the mean of the stationary distribution $\pi$, namely,

$$\mathbf{x} = \lim_{t\to\infty} \mathbb{E}\{X(t)\} = (I - \bar{Q})^{-1}\bar{R}\mathbf{z}^s. \quad (4)$$

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

$$\lim_{t\to\infty} S(t) = \mathbf{x} \quad \text{a.s.} \tag{5}$$

Proofs of the preceding results can be found in [12], and [15] provides an analysis for the case where the graph $\mathcal{G}$ is a weighted complete graph. In short, the results show that, although agent states may not reach a consensus or converge to a fixed value (instead, they may fluctuate a.s. [12]), they converge in distribution to a stationary distribution. Also, if we compute the time average of the agent states, then it converges to the same limit $\mathbf{x}$. This vector $\mathbf{x}$ can be considered as a characterization of the final positions of regular agents on average.

## III. Problem Formulation

In this paper we study the gossip model evolving over an SBM. We define it in Section III-A, and in Section III-B we define the gossip model over an averaged graph. In Section III-C we formulate the considered problem.

### A. Gossip Model over SBM

In this paper we study the gossip model evolving over an SBM. To include stubborn agents in the SBM, we assume that there is a portion $s_0$ of agents being stubborn:

*Definition 2 (SBM with stubborn agents):* A stochastic block model with a portion $s_0$ of agents being stubborn, denoted by $\mathrm{SBM}(n, \boldsymbol{l}, s_0)$, is a random graph model with parameters $n$, $\boldsymbol{l} = [l_s\ l_d]^T$, and $s_0 \in (0,1)$. Here $n$ is an even number, $l_s$ and $l_d$ are parameters in $(0,1)$ depending on $n$, and we assume that $s_0 n/2$ is an integer. The model constructs a graph with stubborn agents according to the following rule: in the first step, a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ is generated from $\mathrm{SBM}(n, \boldsymbol{l})$; in the second step, $s_0 n/2$ agents in each community are labeled as stubborn. ■

Without loss of generality, we sort the agents in any graph generated from $\mathrm{SBM}(n, \boldsymbol{l}, s_0)$ as follows: agents $1, \ldots, r_0 n/2$ (resp. $1+r_0 n/2, \ldots, r_0 n$) are regular agents in community 1 (resp. community 2), and agents $r_0 n + 1, \ldots, r_0 n + s_0 n/2$ (resp. $1 + r_0 n + s_0 n/2, \ldots, n$) are stubborn agents in community 1 (resp. community 2), where $r_0$ is the portion of regular agents and satisfies that $r_0 + s_0 = 1$.

We define the gossip model over $\mathrm{SBM}(n, \boldsymbol{l}, s_0)$ as follows.

*Definition 3 (Gossip model over SBM):* The gossip model over $\mathrm{SBM}(n, \boldsymbol{l}, s_0)$ (the gossip model over the SBM, for short) is the gossip model (3) that evolves over a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ generated from $\mathrm{SBM}(n, \boldsymbol{l}, s_0)$. ■

*Remark 1:* The gossip model over the SBM consists of two random components: the random generation of a graph from $\mathrm{SBM}(n, \boldsymbol{l}, s_0)$ and random interactions between agents in the gossip model (3). We use $(\Omega, \mathcal{F}, \mathbb{P})$ to denote the entire probability space defined by the two components, and use $\mathbb{P}\{\cdot|\mathcal{G}\} = \mathbb{P}_{\mathcal{G}}\{\cdot\}$ and $\mathbb{E}\{\cdot|\mathcal{G}\} = \mathbb{E}_{\mathcal{G}}\{\cdot\}$ to represent the conditional probability and expectation with respect to a graph $\mathcal{G}$ generated from $\mathrm{SBM}(n, \boldsymbol{l}, s_0)$. ■

Since the underlying graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ of the gossip model over the SBM is random, we now use $\bar{Q}$ and $\bar{R}$ to represent the conditional expectation of $Q(t)$ and $R(t)$ with respect to $\mathcal{G}$, respectively; that is, $\bar{Q} := \mathbb{E}_{\mathcal{G}}\{Q(t)\}$ and $\bar{R} := \mathbb{E}_{\mathcal{G}}\{R(t)\}$. From (3), it follows that $\mathbb{P}_{\mathcal{G}}\{[Q(t)]_{ii} = 1/2\} = \sum_{j=1}^{n} a_{ij}/\alpha = 1 - \mathbb{P}_{\mathcal{G}}\{[Q(t)]_{ii} = 1\}$, $1 \le i \le r_0 n$, where $\alpha = \sum_{i=1}^{n}\sum_{j=i+1}^{n} a_{ij}$. Similarly, $\mathbb{P}_{\mathcal{G}}\{[Q(t)]_{ij} = 1/2\} = a_{ij}/\alpha = 1 - \mathbb{P}_{\mathcal{G}}\{[Q(t)]_{ij} = 0\}$ for $1 \le i \ne j \le r_0 n$, and $\mathbb{P}_{\mathcal{G}}\{[R(t)]_{ij} = 1/2\} = a_{i,r_0 n+j}/\alpha = 1 - \mathbb{P}_{\mathcal{G}}\{[R(t)]_{ij} = 0\}$ for $1 \le i \le r_0 n$ and $1 \le j \le s_0 n$. Hence,

$$\bar{Q} = \mathbb{E}_{\mathcal{G}}\{Q(t)\} = I_{r_0 n} - \frac{1}{2\alpha}\bar{M}, \tag{6}$$

$$\bar{R} = \mathbb{E}_{\mathcal{G}}\{R(t)\} = \frac{1}{2\alpha}\bar{U}, \tag{7}$$

where

$$\bar{M} := \begin{bmatrix} \sum_{j=1}^{n} a_{1j} & -a_{12} & \cdots & -a_{1,r_0 n} \\ -a_{21} & \sum_{j=1}^{n} a_{2j} & \cdots & -a_{2,r_0 n} \\ \vdots & & \ddots & \vdots \\ -a_{r_0 n,1} & \cdots & -a_{r_0 n, r_0 n-1} & \sum_{j=1}^{n} a_{r_0 n, j} \end{bmatrix},$$

$$\bar{U} := \begin{bmatrix} a_{1,r_0 n+1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{r_0 n, r_0 n+1} & \cdots & a_{r_0 n, n} \end{bmatrix}.$$

From the preceding discussion, we see that $\bar{Q}$ and $\bar{R}$ are random matrices depending on $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, which is generated by $\mathrm{SBM}(n, \boldsymbol{l}, s_0)$. If $(I - \bar{Q})^{-1}$ exists, then we use

$$\mathbf{x}^{\mathcal{G},n} = (I - \bar{Q})^{-1}\bar{R}\mathbf{z}^s \tag{8}$$

to represent the mean of the stationary distribution of the gossip model over the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ (it is a conditional expectation with respect to $\mathcal{G}$). Otherwise we define $\mathbf{x}^{\mathcal{G},n} = +\infty$. Note that $\mathbf{x}^{\mathcal{G},n}$ is a random vector depending on $\mathcal{G}$, and the superscript $n$ emphasizes the dependence of the considered vector on the network size.

### B. Gossip Model over Averaged Graph

In the study of SBMs, the expectation of the adjacency matrix $A$ of a graph $\mathcal{G}$, generated from an SBM, plays a crucial role (see e.g., Section 4.5 of [17]). Note that by averaging all possible graphs generated from $\mathrm{SBM}(n, \boldsymbol{l}, s_0)$, we can obtain a weighted complete graph with weighted adjacency matrix $\mathbb{E}\{A\}$. We define the gossip model over the averaged graph as follows.

*Definition 4 (Gossip model over averaged graph):* The gossip model over the averaged graph is the gossip model (3) that evolves over the complete graph with weighted adjacency matrix $\mathbb{E}\{A\}$, which is obtained from averaging all graphs generated from $\mathrm{SBM}(n, \boldsymbol{l}, s_0)$. ■

In the gossip model over the averaged graph, the interaction probability matrix is $\mathcal{W} := \mathbb{E}\{A\}/\mathbb{E}\{\alpha\}$, which is defined by $\mathbb{E}\{A\}$, the expectation of the adjacency matrix $A$ of the SBM. So $\mathcal{W}$ can be considered as an approximation of the interaction probability matrix $W$ of the gossip model over the SBM.

Define $\bar{\mathcal{Q}} := I - \mathbb{E}\{\bar{M}\}/(2\mathbb{E}\{\alpha\})$ and $\bar{\mathcal{R}} := \mathbb{E}\{U\}/(2\mathbb{E}\{\alpha\})$, which are the counterparts of $\bar{Q}$ in (6) and $\bar{R}$ in (7), respectively. It is not hard to see that the matrices $\mathcal{W}$, $\bar{\mathcal{Q}}$, and $\bar{\mathcal{R}}$ all have block structure corresponding to the community structure, and it is shown in [15] that $(I - \bar{\mathcal{Q}})^{-1}$ exists and the conclusions of Proposition 1 hold for the gossip model over the averaged graph. Thus we use

$$\mathbf{x}^{*,n} := (I - \bar{\mathcal{Q}})^{-1}\bar{\mathcal{R}}\mathbf{z}^s \tag{9}$$

to represent the mean of the stationary distribution of the gossip model over the averaged graph. It is verified in Proposition 2 of [15] that $\mathbf{x}^{*,n} = [\chi_1 \mathbf{1}_{r_0 n/2}\ \chi_2 \mathbf{1}_{r_0 n/2}]$ for some constants $\chi_1$ and $\chi_2$ (these constants are weighted averages of the stubborn-agent states). In other words, agents in the same community have stationary states with the same expectation, in the gossip model over the averaged graph.

### C. Concentration Problem

In this paper, we study the relationship between $\mathbf{x}^{\mathcal{G},n}$ and $\mathbf{x}^{*,n}$, given by the gossip model over the SBM and the gossip model over the averaged graph, respectively. More precisely, we are interested in whether there exists a concentration phenomenon for $\mathbf{x}^{\mathcal{G},n}$. We state the considered problem as follows:

> **Problem.** Provide a high-probability bound for the difference of $\mathbf{x}^{\mathcal{G},n}$ and $\mathbf{x}^{*,n}$, $\|\mathbf{x}^{\mathcal{G},n} - \mathbf{x}^{*,n}\|$. ■

The problem is related to whether we can use the gossip model over the averaged graph to approximate the gossip model over a network with community structure. Since it could be simpler to analyze the gossip model over the averaged graph, the approximation is useful for studying behavior of the gossip model over the considered network.

## IV. Main Results

In this section we study the problem introduced in Section III-C. We show that the mean of the stationary distribution of the gossip model over the SBM is close to that of the gossip model over the averaged graph.

Our main result given in the following shows the difference of $\mathbf{x}^{\mathcal{G},n}$ and $\mathbf{x}^{*,n}$ is an infinitesimal of $\|\mathbf{z}^s\|$ with high probability.

*Theorem 1:* Suppose that the gossip model over the SBM and the gossip model over the averaged graph start with the same stubborn-agent state vector $\mathbf{z}^s$. If $l_s = \omega((\log n)/n)$ and $l_d = \omega((\log n)/n)$, then

$$\mathbb{P}\left\{\|\mathbf{x}^{\mathcal{G},n} - \mathbf{x}^{*,n}\| = O\left(\sqrt{\frac{\log n}{n(l_s + l_d)}}\right)\|\mathbf{z}^s\|\right\} \geq 1 - n^{-c},$$

where $\mathbf{x}^{\mathcal{G},n}$ and $\mathbf{x}^{*,n}$ are given in (8) and (9), respectively, and $c$ is a positive constant. ■

*Proof:* Due to page limit, we provide a proof sketch. Note that

$$\begin{aligned}
&\|\mathbf{x}^{\mathcal{G},n} - \mathbf{x}^{*,n}\| \\
&= \|(I - \bar{Q})^{-1}\bar{R}\mathbf{z}^s - (I - \bar{\mathcal{Q}})^{-1}\bar{\mathcal{R}}\mathbf{z}^s\| \\
&= \left\|\left[\left(\frac{n\bar{M}}{\alpha}\right)^{-1}\frac{n\bar{U}}{\alpha} - \left(\frac{n\mathbb{E}\{\bar{M}\}}{\mathbb{E}\{\alpha\}}\right)^{-1}\frac{n\mathbb{E}\{\bar{U}\}}{\mathbb{E}\{\alpha\}}\right]\mathbf{z}^s\right\|,
\end{aligned}$$

so it suffices to bound

$$\begin{aligned}
&\left\|\left(\frac{n\bar{M}}{\alpha}\right)^{-1}\frac{n\bar{U}}{\alpha} - \left(\frac{n\mathbb{E}\{\bar{M}\}}{\mathbb{E}\{\alpha\}}\right)^{-1}\frac{n\mathbb{E}\{\bar{U}\}}{\mathbb{E}\{\alpha\}}\right\| \\
&\leq \left\|\left(\frac{n\bar{M}}{\alpha}\right)^{-1}\right\|\left\|\frac{n\bar{U}}{\alpha} - \frac{n\mathbb{E}\{\bar{U}\}}{\mathbb{E}\{\alpha\}}\right\| \\
&\quad + \left\|\left(\frac{n\bar{M}}{\alpha}\right)^{-1} - \left(\frac{n\mathbb{E}\{\bar{M}\}}{\mathbb{E}\{\alpha\}}\right)^{-1}\right\|\left\|\frac{n\mathbb{E}\{\bar{U}\}}{\mathbb{E}\{\alpha\}}\right\|.
\end{aligned} \tag{10}$$

To bound the first term, we need analyze $\|n\bar{M}/\alpha - n\mathbb{E}\{\bar{M}\}/\mathbb{E}\{\alpha\}\|$), which can be decomposed as

$$\begin{aligned}
&\left\|\frac{n}{\alpha}\bar{M} - \frac{n}{\mathbb{E}\{\alpha\}}\mathbb{E}\{\bar{M}\}\right\| \\
&\leq \left\|\frac{n}{\alpha}\bar{M}\left(1 - \frac{\alpha}{\mathbb{E}\{\alpha\}}\right)\right\| + \left\|\frac{n}{\mathbb{E}\{\alpha\}}(\bar{M} - \mathbb{E}\{\bar{M}\})\right\| \\
&=: (I) + (II).
\end{aligned}$$

Utilizing the Chernoff inequality [17] and the Gershgorin circle theorem, we can show that $(I) \leq C/\sqrt{n}$ with high probability for some positive constant $C$. The matrix Bernstein inequality [17] yields that

$$(II) = O\left(\sqrt{\frac{\log n}{n(l_s + l_d)}}\right),$$

with probability at least $1 - n^{-c}$, where $c$ is a positive constant. Therefore, with high probability,

$$\left\|\frac{n}{\alpha}\bar{M} - \frac{n}{\mathbb{E}\{\alpha\}}\mathbb{E}\{\bar{M}\}\right\| = O\left(\sqrt{\frac{\log n}{n(l_s + l_d)}}\right).$$

Note that, if $(n\bar{M}/\alpha)^{-1}$ and $(n\mathbb{E}\{\bar{M}\}/\mathbb{E}\{\alpha\})^{-1}$ exist, then it follows from (5.8.1) of [18] that

$$\begin{aligned}
&\left\|\left(\frac{n\bar{M}}{\alpha}\right)^{-1} - \left(\frac{n\mathbb{E}\{\bar{M}\}}{\mathbb{E}\{\alpha\}}\right)^{-1}\right\| \\
&\leq \left\|\left(\frac{n\bar{M}}{\alpha}\right)^{-1}\right\|\left\|\left(\frac{n\mathbb{E}\{\bar{M}\}}{\mathbb{E}\{\alpha\}}\right)^{-1}\right\|\left\|\frac{n}{\alpha}\bar{M} - \frac{n}{\mathbb{E}\{\alpha\}}\mathbb{E}\{\bar{M}\}\right\|.
\end{aligned}$$

Utilizing the Chernoff inequality and the Gershgorin circle theorem, we know that both $\|(n\bar{M}/\alpha)^{-1}\|$ and $\|(n\mathbb{E}\{\bar{M}\}/\mathbb{E}\{\alpha\})^{-1}\|$ can be upper bounded by some positive constants. In this way, we obtain a high-probability bound for the first term of (10).

Similarly, it is able to use the Chernoff inequality and the Matrix Bernstein inequality for rectangular matrices [17] to analyze the second term of (10). The conclusion follows from combining the bounds of the two terms. ■

*Remark 2:* If $l_s = a_s(\log n)/n$ and $l_d = a_d(\log n)/n$ (we refer to it as the critical case), then when $(a_s + a_d)/2 > 1$, SBM$(n, \boldsymbol{l})$ is connected with high probability (i.e., it generates a connected graph with high probability) [16], [19]. So our assumption in Theorem 1 requires the graph generated from SBM$(n, \boldsymbol{l})$ is slightly denser than a random graph with expected degree of order $O(\log n)$. It can be shown that $\|\mathbf{x}^{\mathcal{G},n} - \mathbf{x}^{*,n}\| = O(1)\|\mathbf{z}^s\|$ in the critical case. In

Defintion 2 we assume that the stubborn agents constitute a linear fraction of both communities, but in practice there could be only a few stubborn agents in a social network. One way to capture this situation is to set $s_0$ given in Definition 2 to be small enough. Another way is to assume that the number of stubborn agents is $o(n)$. We will study this case in the future. ■

*Remark 3:* Note that $\sqrt{(\log n)/[n(l_s+l_d)]} = o(1)$ from the assumptions of Theorem 1, so the inequality in the theorem reveals a concentration phenomenon in the gossip model over the SBM. It shows that $\mathbf{x}^{\mathcal{G},n}$ lies in a neighborhood of its expected version, $\mathbf{x}^{*,n}$. If the entries of $\mathbf{z}^s$ is uniformly bounded, then $\|\mathbf{z}^s\| = O(\sqrt{n})$, and hence $\|\mathbf{x}^{\mathcal{G},n} - \mathbf{x}^{*,n}\| = o(\sqrt{n})$. If $\mathbf{x}^{\mathcal{G},n}$ and $\mathbf{x}^{*,n}$ have uniformly bounded entries, then a trivial bound for their difference is $\|\mathbf{x}^{\mathcal{G},n} - \mathbf{x}^{*,n}\| = O(\sqrt{n})$. Therefore, Theorem 1 indicates that most entries of $\mathbf{x}^{\mathcal{G},n}$ must be not too far away from those of $\mathbf{x}^{*,n}$. In particular, the difference between $\mathbf{1}_n^T\mathbf{x}^{\mathcal{G},n}/n$ and $\mathbf{1}_n^T\mathbf{x}^{*,n}/n$ is $o(1)$. That is, the mean of $\mathbf{x}^{\mathcal{G},n}$ is close to that of $\mathbf{x}^{*,n}$. ■

With the help of Theorem 1, it is able to approximate the entries of $\mathbf{x}^{\mathcal{G},n}$ by using $\mathbf{x}^{*,n}$, as the following corollary shows.

*Corollary 1:* For all $\varepsilon > 0$, denote $\mathcal{V}^{\varepsilon,n} := \{i \in \mathcal{V} : |\mathbf{x}_i^{\mathcal{G},n} - \mathbf{x}_i^{*,n}| > \varepsilon\}$. Under the same conditions of Theorem 1, if all entries of $\mathbf{z}^s$ are uniformly bounded by some constant $C$ not depending on $n$, then it holds for all $\varepsilon > 0$ that

$$\mathbb{P}\left\{|\mathcal{V}^{\varepsilon,n}| = O\left(\sqrt{\frac{n\log n}{(l_s+l_d)}}\right)\right\} \geq 1 - n^{-c},$$

where $c$ is a positive constant. ■

*Remark 4:* The corollary indicates that for all agents but a small part of them with size of order $o(n)$, $\mathbf{x}_i^{\mathcal{G},n}$ is very close to its expected version $\mathbf{x}_i^{*,n}$ with high probability. Since, from [15], $\mathbf{x}_i^{*,n} = \chi_k$ for all $1 \leq i \leq r_0 n$ and $k = 1, 2$ such that $\mathcal{C}_i = k$, this result establishes a correspondence between stationary agent states and community labels. The result also shows that it is possible to use the gossip model over the averaged graph, which is much easier to analyze, to study behavior of the gossip model over the SBM. ■

## V. Numerical Simulation

In this section, we present numerical experiments to illustrate the obtained results. The first numerical experiment demonstrates the concentration phenomenon in the gossip model over the SBM: the mean of the stationary distribution of the gossip model over the SBM, $\mathbf{x}^{\mathcal{G},n}$, is close to that of the gossip model over the averaged graph, $\mathbf{x}^{*,n}$, and their difference can be upper-bounded by $o(1)\|\mathbf{z}^s\|$, where $n$ is the number of agents and $\mathbf{z}^s$ is the stubborn-agent state vector. This experiment also shows that most entries of $\mathbf{x}^{\mathcal{G},n}$ is close to $\mathbf{x}^{*,n}$ for large enough $n$. The second experiment illustrates that a similar concentration phenomenon also appears in the critical case, in which the parameters $l_s$ and $l_d$ of the SBM are of order $O((\log n)/n)$.

In both numerical experiments, the portion of stubborn agents, $s_0$, in SBM$(n, \boldsymbol{l}, s_0)$ is set to be 0.2, and the states of stubborn agents in community 1 (resp. community 2) are generated independently from uniform distribution $(0.9, 1)$ (resp. from uniform distribution $(0, 0.1)$).

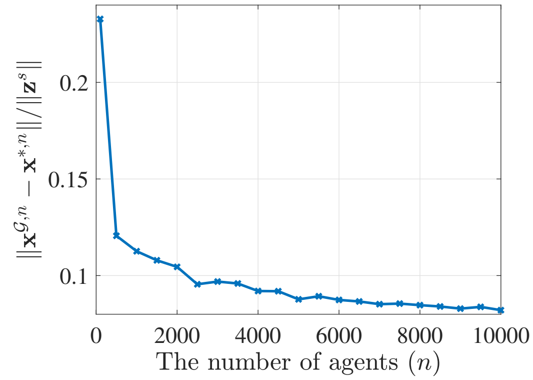



Fig. 1. The ratio $\|\mathbf{x}^{\mathcal{G},n} - \mathbf{x}^{*,n}\|/\|\mathbf{z}^s\|$ decreases as $n$ increases.

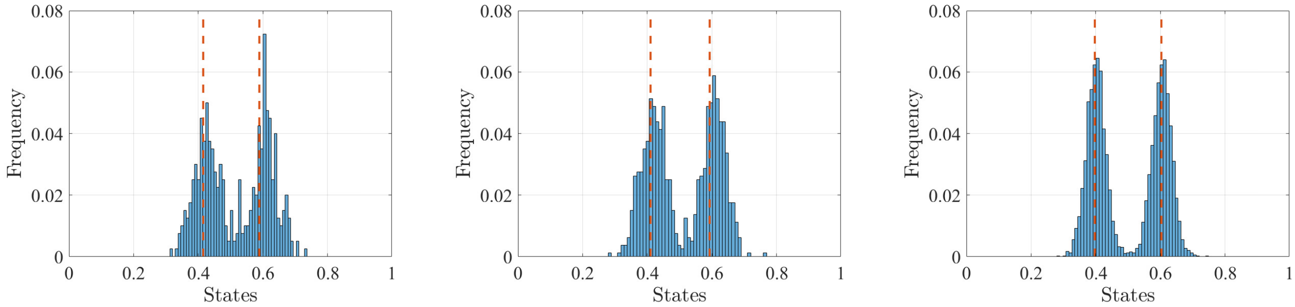
In the first experiment, we set $l_s = (\log n)^2/n$ and $l_d = (\log n)(\log\log n)/n$. For a given $n$, we generate a graph $\mathcal{G}$ from SBM$(n, \boldsymbol{l}, s_0)$, and compute $\|\mathbf{x}^{\mathcal{G},n} - \mathbf{x}^{*,n}\|/\|\mathbf{z}^s\|$. Fig. 1 shows that this ratio decreases as $n$ increases. This observation validates the concentration result given in Theorem 1, and implies that $\mathbf{x}^{\mathcal{G},n}$ lies in a neighborhood of its expected version $\mathbf{x}^{*,n}$ for large $n$.

We further draw the histogram of $\mathbf{x}^{\mathcal{G},n}$ for $n = 500$, 1000, and 5000 in Fig. 2. This figure shows that, as $n$ grows, the entries of $\mathbf{x}^{\mathcal{G},n}$ become closer to the corresponding entries of $\mathbf{x}^{*,n}$, as Corollary 1 states. As discussed in Remark 4, $\mathbf{x}^{*,n}$ has two distinct values, $\chi_1$ and $\chi_2$. Thus, for regular agent $i$, $\mathbf{x}_i^{\mathcal{G},n}$ is close to one of these two values that corresponds to agent $i$'s community. The results indicate that the correspondence between agent states and community labels emerges for large networks, and we can use the gossip model over the averaged graph to characterize the positions of agent states in the gossip model over the SBM.

In the second experiment, we set $l_s = 60(\log n)/n$ and $l_d = 20(\log n)/n$, which is an example of the critical case discussed in Remark 2. We draw the histogram of $\mathbf{x}^{\mathcal{G},n}$ for $n = 100$, 500, and 1000 in Fig. 3. The result also shows that more and more entries of $\mathbf{x}^{\mathcal{G},n}$ gather around the entries of $\mathbf{x}^{*,n}$ as $n$ grows. This observation indicates that the concentration phenomenon still exists in the critical case.

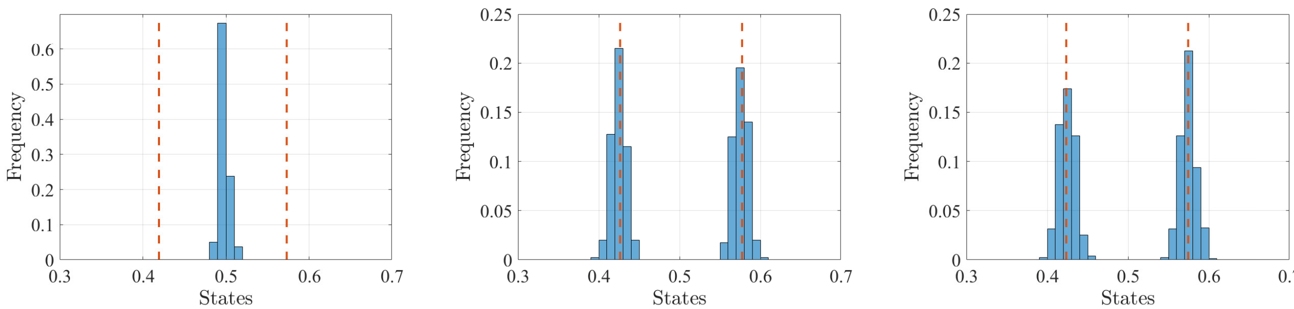
## VI. Conclusion

In this paper we studied a concentration phenomenon in a gossip model with stubborn agents over an SBM. We obtained a concentration result for the conditional mean of the stationary distribution of the gossip model over the SBM. It is shown that this conditional mean is close to the mean of the stationary distribution of the gossip model over an averaged graph. As a result, most entries of the two vectors are close. This result establishes a correspondence between the stationary agent states of the gossip model over the SBM and the community labels of the agents, and indicates that it is possible to use the gossip model over the averaged graph




(a) $n = 500$ (b) $n = 1000$ (c) $n = 5000$

Fig. 2. Concentration phenomenon for $l_s = (\log n)^2/n$ and $l_d = (\log n)(\log \log n)/n$. The subfigures show the histogram of $\mathbf{x}^{\mathcal{G},n}$ with $n = 500$, 1000, and 5000, respectively. The red dotted lines represent the two distinct values of $\mathbf{x}^{*,n}$.




(a) $n = 100$ (b) $n = 500$ (c) $n = 1000$

Fig. 3. Concentration phenomenon for $l_s = 60(\log n)/n$ and $l_d = 20(\log n)/n$. The subfigures show the histogram of $\mathbf{x}^{\mathcal{G},n}$ with $n = 100$, 500, and 1000, respectively. The red dotted lines represent the two distinct values of $\mathbf{x}^{*,n}$.

to approximate behavior of the gossip model over the SBM. Future work includes to study the critical case of Theorem 1 and to investigate the gossip model over general SBMs.

## References


[1] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics Reports*, vol. 659, pp. 1–44, 2016.

[2] D. Hric, R. K. Darst, and S. Fortunato, "Community detection in networks: Structural communities versus ground truth," *Physical Review E*, vol. 90, no. 6, p. 062805, 2014.

[3] M. E. Newman and A. Clauset, "Structure and inference in annotated networks," *Nature Communications*, vol. 7, no. 1, pp. 1–11, 2016.

[4] D. Hric, T. P. Peixoto, and S. Fortunato, "Network structure, metadata, and the prediction of missing nodes and annotations," *Physical Review X*, vol. 6, no. 3, p. 031038, 2016.

[5] L. Peel, D. B. Larremore, and A. Clauset, "The ground truth about metadata and community detection in networks," *Science Advances*, vol. 3, no. 5, p. e1602548, 2017.

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

[7] A. Flache, M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, and J. Lorenz, "Models of social influence: Towards the next frontiers," *Journal of Artificial Societies and Social Simulation*, vol. 20, no. 4, 2017.

[8] A. V. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part I," *Annual Reviews in Control*, vol. 43, pp. 65–79, 2017.

[9] A. Matakos, E. Terzi, and P. Tsaparas, "Measuring and moderating opinion polarization in social networks," *Data Mining and Knowledge Discovery*, vol. 31, no. 5, pp. 1480–1505, 2017.

[10] C. Musco, C. Musco, and C. E. Tsourakakis, "Minimizing polarization and disagreement in social networks," in *Proceedings of the 2018 World Wide Web Conference*, pp. 369–378, 2018.

[11] J. Gaitonde, J. Kleinberg, and E. Tardos, "Adversarial perturbations of opinion dynamics in networks," in *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 471–472, 2020.

[12] D. Acemoğlu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Mathematics of Operations Research*, vol. 38, no. 1, pp. 1–27, 2013.

[13] G. Como and F. Fagnani, "From local averaging to emergent global behaviors: The fundamental role of network interconnections," *Systems & Control Letters*, vol. 95, pp. 70–76, 2016.

[14] M. Talagrand, "A new look at independence," *The Annals of Probability*, pp. 1–34, 1996.

[15] Y. Xing, X. He, H. Fang, and K. H. Johansson, "Detecting communities in a gossip model with stubborn agents," *arXiv preprint arXiv:2102.09683*, 2021.

[16] E. Abbe, "Community detection and stochastic block models: Recent developments," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6446–6531, 2017.

[17] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 2012.

[19] E. Abbe, A. S. Bandeira, and G. Hall, "Exact recovery in the stochastic block model," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 471–487, 2015.