2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Compressibility of Voter-Model State Snapshots in the Graph Spectral Basis

Chenyan Zhu and Sandip Roy

***Abstract*— The sparsity or compressibility of network spread states in a graph-spectrum basis is examined, in the context of a stochastic model for influence/spread known as the voter model. In particular, the first- and second- moments are characterized, for the graph-spectrum basis components contained in the voter-model state. These formal characterizations, as well as an asymptotic analysis of the voter model, are used to relate compressibility with the network's graph. An illustrative example as well as a simulation of a larger-scale model are included.**

I. Introduction

During the last 15-20 years, there has been a vibrant effort within the controls community on spread and influence processes in networks [1]–[3]. This effort – which complements a much wider body of work on spread modeling across the sciences [4]– has yielded interesting characterizations of the spatiotemporal dynamics of spread, as well as on the allocation of resources to mitigate or shape spread processes. Controls-engineering approaches hold promise to substantially advance research and practice in the management of spread processes, because they can enable policy design. To date, however, controls-engineering efforts on spread management largely have not been translated to practice (e.g. used by policy organizations like the Center for Disease Control to mitigate spreads), to the best of our knowledge.

One barrier to the practical application of network spread control techniques, and indeed to the development of scientific approaches to managing spread, lies in the inadequacy of models and data for spread. Real-world spread processes exhibit tremendously volatile behaviors at multiple scales, which are governed both by the intrinsic stochastics of the spread processes and complex environmental factors. The simple differential-equation-type models that have been used to design spread controls are ill-equipped to forecast these dynamics (even if they incorporate some stochastics), and data is rarely available at the resolution and scale needed to recover and forecast the processes. As an example, COVID-19 has repeatedly emerged in large-scale outbreaks, with very little advanced forecasting, and inclarity about the process dynamics even after emergence. Similarly, the mechanisms underlying e.g. election/voting processes and misinformation spread are tremendously volatile, and hard to model and forecast. Motivated by this challenge, there has recently been an exciting effort to learn or identify network-theoretic models for spread from data [5]. As an alternative, direct data-driven methods for designing spread controls have also been pursued, as have structural insights into effective controls that are independent of detailed model knowledge [1], [6]. Nevertheless, much remains to be done to develop practical models and controller designs for highly volatile spread processes.

In this study, we approach the modeling of volatile or stochastic spread processes from a complementary perspective, which is grounded in the notion of sparsity or compressibility of the network's state. Specifically, rather than aiming to learn a dynamical model for the spread, we view snapshots of the state at particular times as *scenes* defined on top of the network's graph. Our main hypothesis is that these scenes – although varying stochastically in time, and across different spread events (different outbreaks, different diseases) – are *sparse* in a graph-defined basis. That is, each scene can be approximated well using a (possibly-different) small subset of the basis vectors. Conceptually, sparsity represents the recognition that spread state snapshots vary greatly with time, but have patterns that are closely tied with the network's graph topology.

In a previous CDC paper, we undertook numerical analyses of COVID-19 spread counts and also spread-model-generated data, which suggest that spread snapshots can be sparsified in *graph spectrum* bases [7]. Our main aim in this study is to develop a formal analysis of spread scene snapshots for data generated by a particular stochastic model for spread – specifically, an influence or voter model [8]–[10]. For this model, we statistically characterize the state snapshot components in the graph spectrum basis, so as to give insight into sparsification in this basis. Via the formal analysis and also some examples/simulations, we argue that spread scenes can be sparsifiable or compressible in the graph-spectral basis.

The remainder of the article is organized as follows. The sparsification analysis, and the voter/influence model which is the focus of this study, are introduced in Section II. The main analytical results are developed in Section III. Further interpretations of the sparsity analysis are presented along with simulations in Section IV.

II. Problem Formulation

The goal of our analysis is to understand whether snapshots of network spread processes can be sparsified or compressed in an appropriate basis. Broadly, we consider spread among a network of N nodes, labeled $i = 1, \ldots, n$. Each node i has a spread state $x_i[k] \in R$ associated with it, which evolves along a clocked time axis ($k \in Z$). In general, the nodes and their states may represent different constructs related to a spread process, such as daily infection counts

The authors are with Texas A&M University. This work was conducted at Washington State University and Texas A&M University. Please send correspondence to sandip@tamu.edu .

in different geographic regions, or binary spread/opinion statuses of individuals. The states are assumed to evolve with time via interactions among the nodes defined by a graph Γ, which may be weighted or unweighted, and directed or undirected.

The network spread state $x[k] = \begin{bmatrix} x_1[k] \\ \vdots \\ x_n[k] \end{bmatrix}$ at particular snapshot times k is considered. The hypothesis underlying our work is that these snapshots of the network state, although highly variable, exhibit patterns that are closely tied to the network's graph Γ. In particular, we consider whether the spread state $x[k]$ is sparse or compressible in a basis, which is defined from the spectrum of a matrix which captures the graph structure Γ. To formalize this notion, an $n \times n$ matrix $A = [a_{ij}]$ is defined, which is commensurate with the graph Γ in the following sense: $a_{ij} \neq 0$ for $i \neq j$ if and only if there is an edge from vertex j to vertex i in Γ. The matrix A, for instance, may be the Laplacian matrix associated with the (undirected or directed) graph Γ, or a nonnegative matrix that encapsulates interactions. Then, the graph spectrum matrix V is defined to contain the full spectrum of A, i.e. V is an $n \times n$ matrix whose columns are the eigenvectors and generalized eigenvectors of A.

We are interested in whether network state snapshots $x = x[k]$ are (approximately) sparse in the graph spectrum basis [11]. Specifically, the spread state is said to be K-sparse if a vector s with at most K nonzero entries can be found (i.e. $||s||_0 \leq K$), such that $||x - Vs||_2 \leq \epsilon$ for an accuracy threshold $\epsilon > 0$. It is noted that regularized optimization or regression algorithms, such as the least absolute shrinkage and selection operator (LASSO) may be used to compute the sparse solution [12]. Our primary interest here is in understanding whether spread process snapshots are in fact sparsifiable or compressible in graph spectrum bases.

In a prior article, we experimentally assessed compressibility/ sparsifiability of network spread data, including real COVID-19 infection count data and various model-generated data, and found that the data was often compressible [7]. In this study, our focus is on developing a formal analysis of compressibility, in the context of a canonical stochastic model for spread or influence known as the *voter model*.

The voter model is a discrete-valued Markov stochastic automaton, which tracks binary spread/opinion statuses of individuals or nodes in a network. Formally, within our framework, the voter model's network state $x[k]$ evolves as follows. At an initial time $k = 0$, the state $x_i[k]$ of each node may be arbitrarily set as either 0 or 1. At each subsequent time $k = 1, 2, 3, \ldots$, the state of node i is updated via the following rule:

- A determining or *influencing* node $j = 1, \ldots, n$ is selected with probability a_{ij}, where $a_{ij} \geq 0$ and $\sum_{j=1}^{n} a_{ij} = 1$, independently of all other selections. We note that the influencing node may include the node i itself.
- The time-k state of node i is set of the time-$(k-1)$ state of the influencing node j, i.e. $x_i[k] = x_j[k-1]$.

For the voter model, the influence probabilities can be used to define a weighted digraph that captures the pairwise interactions among nodes. Specifically, the graph Γ is defined to have n vertices corresponding to the n nodes; an edge is drawn from vertex j to i ($i \neq j$) if $a_{ij} > 0$, and is assigned a weight of a_{ij}. For the voter model, the influence matrix $A = [a_{ij}]$ is commensurate with the graph structure; we primarily consider compressibility/sparsity with respect to the graph spectrum matrix V for A. We primarily focus on two broad classes of graphs. First, we consider the case that the graph is strongly connected and aperiodic, which we refer to as a strongly-connected voter model. Second, we consider the case that two nodes are stubborn (i.e. are not influenced by other nodes) but have paths to all other vertices in the network graph. We refer to this case as the stubborn-agents voter model.

The voter model has been used to represent an array of influence and spread processes in networks, including decision-making processes within a group, influences of stubborn/malicious agents on elections, evolution of failure statuses within interconnected systems, and infection processes. In these various contexts, the highly stochastic yet patterned nature of the state has been observed. Our aim here is to give a formal analysis for the pattern.

Our primary goal in this paper is to understand whether, and under what conditions on the influence matrix A, state snapshots of the voter model are compressible in the graph spectral basis. Toward this goal, we undertake a statistical analysis of the state of the influence model in a transformed coordinate, and use this to gain some insight into the sparsity/compressibility of voter model states.

III. Main Results

A statistical analysis of the voter model is undertaken, to characterize the compressibility of state snapshots $x[k]$ in the graph spectrum basis V. As introduced above, compressibility is concerned with whether the state snapshot $x[k]$ can be accurately expressed in the form Vs, where s has only a small number of non-zero entries. Noting that the matrix V is invertible, it is instructive to consider the transformed state $s[k] = V^{-1}x[k] = Wx[k]$, where W is the left-eigenvector matrix of A. To understand why, consider the case that the vector $s[k]$ contains many entries that are near 0. In this case, the state snapshot $x[k]$ is compressible, since $x[k]$ can then be approximated well as $x[k] = V\hat{s}[k]$, where $\hat{s}[k]$ is formed by setting the small entries in $s[k]$ to 0. Thus, to understand compressibility, it is sufficient to evaluate whether or not $s[k] = V^{-1}x[k]$ contains many nonzero entries.

The vector $s[k] = V^{-1}x[k]$ is stochastic, and hence characterizing its entries requires a statistical analysis of $s[k]$. Here, several statistical analyses are pursued. First, for the strongly-connected network, the state snapshot $\mathbf{x}[k]$ in the asymptote of large k is shown to be perfectly 1-compressible, in the sense that only one entry of $s[k]$ is non-zero almost surely. Then, a two-moment analysis of $s[k]$ is undertaken, to characterize the expected sizes (norms) of the entries in $s[k]$ vs time, for both the strongly-connected and stubborn-agent networks. Specifically, the mean $E(s[k])$ is first characterized, to develop conditions under which state snapshots are necessarily not compressible. Then, the correlation matrix $E(s[k]s'[k])$ and covariance matrix $E(s[k]s'[k]) - E(s[k])E(s'[k])$ are analyzed. The second-moment analysis provides insight into compressibility in two ways. First, small values on the diagonal entries of the correlation matrix are indicative of compressibility, as they guarantee that the corresponding entries in $s[k]$ are small with high probability. Second, uncorrelation of the entries in $s[k]$ (i.e. small off-diagonal entries in the covariance matrix) is indicative that the basis V is effective for compression of $x[k]$, as prior work has shown that whitening bases achieve optimal or high compressibility [13].

First, we develop a simple result on the compressibility of state snapshots in the asymptote, when the network graph is strongly connected. The result depends on the recognition that the voter model's node states reach a consensus in the asymptote; this immediately implies that the transformed state vector $s[k]$ has at most one non-zero entry. Here is the result:

Lemma 1: Consider a voter model with a strongly connected network graph. For this model, the transformed state vector $s[k]$ has at most 1 non-zero entry in the asymptote $k \to \infty$, in an almost sure sense.

Proof: For a voter model with a strongly connected and aperiodic network graph, state snapshots $x[k]$ are equal to either $x[k] = \mathbf{0}$ or $x[k] = \mathbf{1}$ almost surely, in the asymptote of large k; this can be readily argued by considering the "master" Markov chain description for the state configuration of the voter model [8]. Next, notice that A has a right eigenvector equal to $\mathbf{1}$, and hence V has one column equal to $\mathbf{1}$. It therefore follows that all rows of $W = V^{-1}$ except one row is orthogonal to the vector $\mathbf{1}$. Therefore, $s[k] = V^{-1}x[k]$ has at most one entry equal to 1, in an almost-sure sense. ■

Next, the mean of the transformed state vector $s[k]$ is characterized, and used to gain some insight into the compressibility of the network state. The following lemma presents a formula for the mean of the transformed state vector:

Lemma 2: For the voter model, the mean of the transformed state vector $s[k]$ is given by

$$E(s[k]) = J^k E(s[0]), \tag{1}$$

where $J = WAV$ is the Jordan matrix of A.

Proof: The mean of the voter model's network state $E(x[k])$ has been characterized in prior work. Briefly, the mean can be characterized via a computation of the node state probabilities $P(x_i[k] = 1)$. By conditioning on the influencing node and using the law of total probability, one immediately finds that $P(x_i[k] = 1) = \sum_{j=1}^{n} a_{ij} P(x_j[k-1] = 1)$. Noting the $E(x_i[k]) = P(x_i[k] = 1)$, it follows that $E(x_i[k]) = \sum_{j=1}^{n} a_{ij} E(x_j[k-1])$. Assembling these equations for $i = 1, \ldots, n$ into a matrix form, we recover $E(x[k]) = AE(x[k-1])$. Finally, substituting for $x[k]$ in terms of $s[k]$ yields $E(s[k]) = WAVE(x[k-1])$. The result in the lemma follows. ■

The analysis of the transformed state vector's mean gives insight into compressibility, in the sense that it identifies entries that are necessarily large and hence cannot be eliminated in a compressed approximation. Specifically, if $E(s_i[k])$ is large (i.e. it is not near zero, or has substantial absolute value), then the random variable $s_i[k]$ is necessarily nonzero with substantial probability. On the other hand, if $E(s_i[k])$ is near zero, then $s_i[k]$ may or may not be nonzero with substantial probability. Hence, large entries in the mean identify state-snapshot components that are incompressible with substantial probability, while small entries are indeterminate with regard to compressibility.

Lemma 2 shows that the entries in the transformed state's mean are directly tied to the eigenvalues of the influence matrix A, and hence the eigenvalues constrain compressibility of the network spread state. It is helpful to characterize the eigenvalues, and hence interpret the compressibility of the network state, for strongly-connected and stubborn-agent networks:

1). For a strongly connected and aperiodic network, A has an eigenvalue $\lambda_1 = 1$ with algebraic multiplicity 1, while the remaining eigenvalues $\lambda_2, \ldots, \lambda_n$ are strictly within the unit circle. Assuming that the initial spread state $x[0]$ and hence initial transformed state $s[0]$ can be arbitrary, compressibility is modulated by powers $|\lambda_1|^k, \ldots, |\lambda_n|^k$ of the eigenvalues. Since $\lambda_1^k = \lambda_1 = 1$, state snapshots have at least one incompressible component. If in addition all or most of the terms $|\lambda_2|^k, \ldots, |\lambda_n|^k$ are large, then each eigenvector component in a state snapshot may not be compressible with substantial probability.

2). For a stubborn-agent network, A has an eigenvalue at 1 with algebraic multiplicity 2 and geometric multiplicity 2; we denote these unity eigenvalues as λ_1 and λ_2. The remaining eigenvalues of A are strictly within the unit circle. Thus, we see that state snapshots have at least two incompressible components. If in addition all or most of the terms $|\lambda_3|^k, \ldots, |\lambda_n|^k$ are large, then each eigenvector component in a state snapshot may not be compressible with substantial probability.

A consequence of the preceding analysis is that compression of state snapshots is difficult for small time horizons k, if all or most eigenvalues of A are all close to the unit circle. This brings forth the question of what conditions on the voter model yield eigenvalues that are close to the unit circle. The following are two types of voter models for which A has eigenvalues near the unit circle, and hence compression in the graph spectrum basis is limited for small k:

1). Voter models where every node is stubborn, in the sense that it has a self-influence probability of at least $1 - \epsilon$ for some small ϵ. In this case, it is easy to check that all eigenvalues λ_i of A are order $\mathcal{O}(\epsilon)$ away from the unit circle, which implies that $|\lambda_i|^k$ is small only if $k > \mathcal{O}(\frac{1}{\epsilon})$.

2). Voter models whose graphs Γ are small perturbations of cycle graphs. If Γ is a cycle graph, all eigenvalues of the influence matrix A lie on the unit circle; thus, upon perturbation, the eigenvalues are necessarily near the unit circle, and again compression may not be possible for any small k. These slow-settling voter models can have state snapshots that are difficult to compress for small time horizons. We note that this analysis holds for both strongly-connected voter models and stubborn-agent models.

Importantly, the first-moment analysis helps to identify voter models whose state snapshots are difficult to compress, but it does not allow characterization of voter models whose snapshots are easy to compress – even if the mean value of $s_i[k]$ is small, it is still possible that state snapshots have a substantial component in this basis direction at each time.

Next, the second moments of the transformed state vector are characterized, as a means to develop a more complete treatment of compressibility. Before presenting the main results on the second moments, we find it convenient to define the notion of a *second-order state matrix*, so as to simplify the formal presentation.

Definition 1: Consider a voter model with n nodes and corresponding state matrix A. An $n^2 \times n^2$ matrix $\hat{A} = [\hat{a}_{ij}]$, called the *second-order state matrix*, is defined as follows:

1). For rows $s = t(n+1) - n$, $t = 1, 2, \cdots$, the entries at columns $q = c(n+1) - n$, $c = 1, 2, \cdots$ are set as $\hat{a}_{sq} = a_{tc}$, while the remaining entries on these rows are set to 0. 2). Any other row i of $\hat{A}$ is set equal to the ith row of $A \otimes A$, i.e. $\hat{A}_i = (A \otimes A)_i$ where the index i refers to the ith row.

The following main theorem gives the second-moment statistical analysis of the transformed state $s[k]$. To simplify the development and presentation of the result, the second moment matrix $E(s[k]s[k]')$ is reshaped into a vector form using a Kronecker product. Specifically, the vector $E(s[k] \otimes s[k])$ is considered as an alternative to the second-moment matrix. The vector contains all second moments (or correlations) of the transformed state. Therefore, we refer to the vector as the second moment vector, and use the notation $E(s_v[k])$ where $s_v[k] = s[k] \otimes s[k]$ for the second-moment vector. Similarly, a second-moment vector for the original state is defined as $E(x_v[k])$, where $x_v[k] = x[k] \otimes x[k]$

Here is the main theorem:

Theorem 1: For the voter model, the second moment vector of the transformed state vector $s[k]$ is given by:

$$E(s_v[k]) = HE(s_v[0]) \tag{2}$$

where $H = (V^{-1} \otimes V^{-1})\hat{A}^k(V \otimes V)$.

The theorem depends on a lemma, given next, which specifies the second-moment vector of the original voter model network state $x[k]$.

Lemma 3: For the voter model, the second moment vector of the voter model state $x[k]$ is given by:

$$E(x_v[k]) = \hat{A}^k E(x_v[0]) \tag{3}$$

The proof of the lemma is given first, followed by the proof of the theorem.

Proof: (Proof of the Lemma). A recursion is developed for the second-moment vector $E(x_v[k])$. To develop the recursion, we develop an expression for each expectation in the time-k second-moment vector in terms of those at time k. A similar analysis has been done in the thesis [8], so we give a brief treatment here.

First, let us consider expectations of the form $E(x_i[k]x_i[k])$. Since $x_i[k]$ is binary, $E(x_i[k]x_i[k]) = E(x_i[k])$. Hence, the expectation is simply the first moment of node state at time k, which can be expressed in terms of time-$(k-1)$ node state expectations as $E(x_i[k]) = \sum_{j=1}^n a_{ij} E(x_j[k-1])$. Therefore, we find that $E(x_i[k]x_i[k]) = \sum_{j=1}^n a_{ij} E(x_j[k-1]x_j[k-1])$.

Next, let us consider expectations of the form $E(x_i[k]x_j[k])$, which are equal to the probability $P(x_i[k] = 1, x_j[k] = 1)$. By conditioning on the influencing nodes for nodes i and j, we obtain that $P(x_i[k] = 1, x_j[k] = 1) = \sum_{p=1}^n \sum_{m=1}^n a_{ip}a_{jm} P(x_p[k-1] = 1, x_m[k-1] = 1)$, which therefore yields that $E(x_i[k]x_j[k]) = \sum_p^n \sum_m^n a_{ip}a_{jm} E(x_p[k-1]x_m[k-1])$. By assembling the expectations on the right side of the equation into a vector, the equation can be rewritten as $E(x_i[k]x_j[k]) = A_i \otimes A_j E(x_v[k-1])$.

Finally, stacking the expressions for each second moment, we readily obtain that $E(x_v[k]) = \hat{A}E(x_v[k-1])$. The lemma result follows. ■

Proof: (Theorem Proof). Substituting $s[k] = V^{-1}x[k]$ into $E(s_v[k])$ produces $E(s_v[k]) = (V^{-1} \otimes V^{-1})E(x_v[k])$. Applying $E(x_v[k]) = \hat{A}^k E(x_v[0])$ yields $E[x_v[k]] = (V^{-1} \otimes V^{-1})\hat{A}^k E(x_v[0])$. With a further substitution $x_v[0] = (V \otimes V)s_v[k]$, the theorem is verified. ■

The second moment analysis is of central importance in understanding sparsity or compressibility of network state snapshots $x[k]$ in the graph spectrum basis. In particular, the second moments are relevant in two senses.

First, the second moments of each transformed state variable ($E(s_i[k]^2)$) are an indication of the expected two-norm of each entry in the transformed state vector, or equivalently the energy of the component of $x[k]$ in the corresponding basis direction. Therefore, if many of these entries in the second-moment vector or second-moment matrix are small, it follows that network state snapshots $x[k]$ are compressible. Alternately, if most entries are large, then $x[k]$ is not easily compressible in this basis. Therefore, the second-moment analysis gives a direct indication of compressibility of the state snapshots. The result in Theorem 1 gives a method for computing the second moments, and hence gauging compressibility/sparsity; we demonstrate this in an example later in the section. The expression also holds promise to give characterizations for compressibility in terms of the network graph, although we have not yet been able to develop a full characterization. Roughly, Theorem 1 suggests that the eigenvalues of A should have a bearing on compressibility, since $H = (V^{-1} \otimes V^{-1})\hat{A}^k(V \otimes V)$ may be approximately viewed as a diagonalizing the matrix $A \otimes A$; this is however a crude approximation, as $\hat{A}$ differs from $A \otimes A$. In the asymptote of large k, the second-moment analysis simplifies further. In particular, let us consider the asymptotic second-moment analysis for the stubborn-agents voter model (since the asymptotics of the strongly connected model have already been characterized). It can be shown that $\hat{A}$ has two eigenvalues at 1 in this case (see [10], whose corresponding eigenvectors determine the steady-state second moment vectors. This simple calculation therefore can be used to gain insight into the compressibility of state snapshots in the asymptote – further details are omitted.

Additionally, the second-moment analysis gives insight into compressibility, with respect to the appropriateness of the graph spectrum basis for compression. Prior work has shown that basis transformations that whiten stochastic vectors, i.e. make the components uncorrelated, tend to optimize compressibility or sparsity [13]. Thus, in our setting, if the matrix H is diagonal or diagonally dominant, the entries in the transformed state vector would be uncorrelated under broad conditions on the initial state. Thus, if this is the case, the transformation is serving to decorrelate the node states, and hence should be well suited for sparsification. Toward understanding whether the transformation is whitening or decorrelating the data, it is also helpful to consider the autocovariance rather than the autocorrelation. The autocovariance can be readily computed based on the first- and second- moment analysis.

Example 1: A stubborn-agents voter model with 4 nodes is analyzed to assess and illustrate the sparsity of the second moment of $s[k]$. The corresponding state matrix is considered shown:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0.2 & 0 & 0.3 \\ 0.4 & 0 & 0 & 0.6 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

The initial state is selected as $x[0] = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$, Figure 1 illustrates the amplitudes of the entries in $E(s[k]s[k]')$ at time $k = 1000$. Each block is colored based on the numerical values of the entries located in the corresponding position in the second-moment matrix.

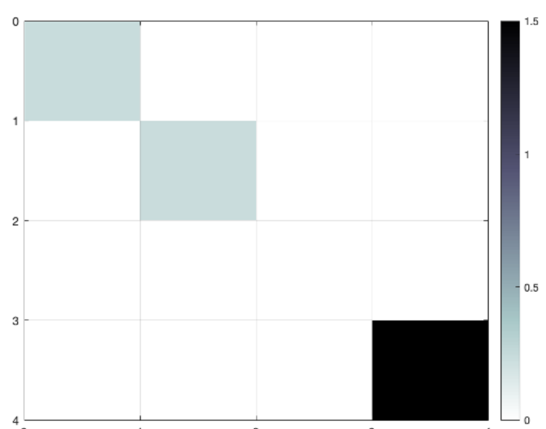


Fig. 1. Visualization of compressibility of a state snapshot at time 1000. The correlation matrix of $s[k]$ only 3 non-zero entries: $E[s[k]s'[k]]_{11} = 0.24$, $E[s[k]s'[k]]_{22} = 0.2344$, and $E[s[k]s'[k]]_{44} = 1.5006$.

As expected, the correlation matrix provided by reshaping the second-moment vector of $s[k]$ presents a high sparsity display. Specifically, zero entries on the off-diagonals numerically support the efficiency of the graph spectrum bases. Also, only one diagonal entry is large, with two others small and one vanishingly so. This structure indicates the sparsity or compressibility of the state in this basis.

Figure 2 visualizes the sparsity of the matrix H. As discussed above, the structure of H is of interest in assessing compressibility/sparsity, as it decides how the initial correlation vector of $s[0]$ is scaled to determine the time k transformed state. As shown in the figure, the matrix indicates the sparsity and diagonal dominance of the H matrix, which also reflects the possibility for sparsity/compressibility.

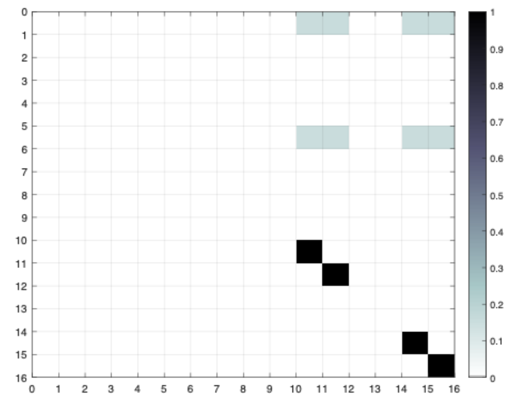


Fig. 2. Visualization of the sparsity of the H matrix. Four diagonal entries with fixed value 1 hold on any snapshot time.

IV. Simulations

Compression of state snapshots in the graph spectrum basis is demonstrated for a stubborn-agents voter model with 200 nodes, whose graph is a line. Specifically, nodes 1 and 200 are stubborn and maintain the states of 0 and 1 respectively, i.e. their self-influencing probabilities are always set to 1. Other nodes will either copy neighbors' states with a probability of 0.4 or maintain the previous state with a probability of 0.2, as shown in Figure 3.

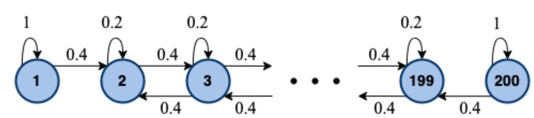


Fig. 3. Visualization of 200 nodes line case graph.

Sparsification of state snapshots is undertaken for this example, and the accuracy of the sparse approximations is illustrated. Specifically, a state snapshot of the 200-node network state at a specific time $k = 500$ (blue line) is compared with a 35-sparse approximation (red line), obtained via the LASSO algorithm (Figure 4). The sparse approximation is seen to effectively approximate the voter model state.

Also, a family of sparse approximations with different sparsity levels is generated using LASSO. The mean-square errors of the approximations are shown as a function of the sparsity level (number of non-zero entries) in Figure 5. As the number of non-zero entries increases, the approximation becomes rapidly more accurate.

The analyses are repeated for a state snapshot at time $k = 1000$. Again, a small number of basis vectors is able

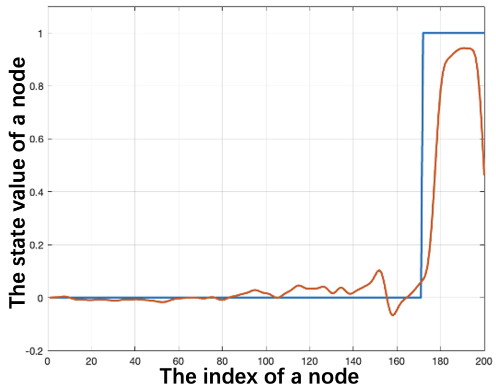


Fig. 4. A state snapshot at a time k=500 and a 35-sparse approximation (35 non-zero entries).

to approximate the state snapshot, although different basis vectors arise in the approximation.

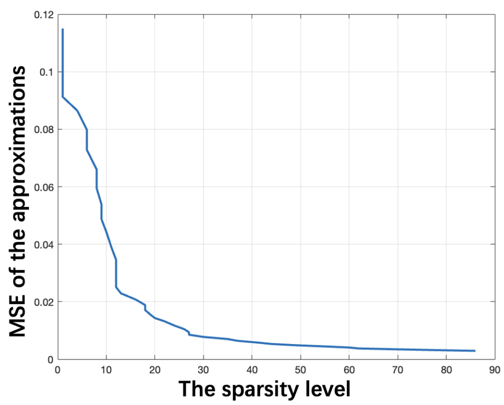


Fig. 5. The mean square error of a sparse approximation is shown as a function of the sparsity level (number of nonzero entries).

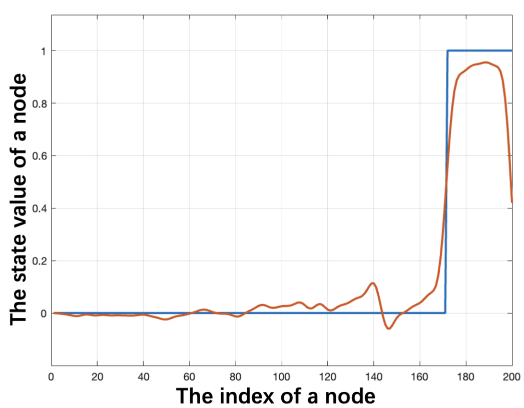
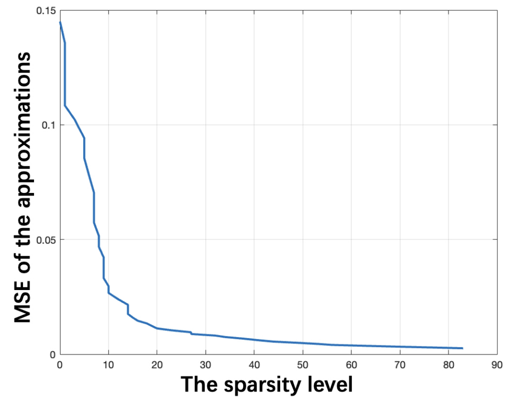


Fig. 6. A state snapshot at a time k=1000 and a 33-sparse approximation (33 non-zero entries).

V. Conclusion

The compressibility of spread processes, modeled using a voter model, is studied using statistical characterizations of the model statistics characterizations. Specifically, the compressibility of the voter model's state snapshot is examined via a first- and second- moment analysis of the state, and also via numerical simulation of the voter model. The analysis suggests that there is a connection between the graph



.

Fig. 7. The mean square error of a sparse approximation at k=1000.

topology and compressibility; formalizing this connection is an important direction of future work.

One practical benefit of compressibility lies in the possibility to estimate or recover spread state snapshots from sparse observations, which would then allow situational awareness about the spread from limited measurements. The sparse recovery of state snapshots, given compressibility, derives from the classical work on compressive sensing [11]. We leave it to future work to formalize the benefits of compressibility from a sparse recovery standpoint, as well as other applications of the compressibility analysis.

References

[1] J. A. Torres, S. Roy, and Y. Wan, "Sparse resource allocation for linear network spread dynamics," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1714–1728, 2016.

[2] P. E. Paré, C. L. Beck, and T. Başar, "Modeling, estimation, and analysis of epidemics over networks: An overview," *Annual Reviews in Control*, vol. 50, pp. 345–360, 2020.

[3] C. Nowzari, V. M. Preciado, and G. J. Pappas, "Analysis and control of epidemics: A survey of spreading processes on complex networks," *IEEE Control Systems Magazine*, vol. 36, no. 1, pp. 26–46, 2016.

[4] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.

[5] P. E. Paré, J. Liu, C. L. Beck, B. E. Kirwan, and T. Başar, "Analysis, estimation, and validation of discrete-time epidemic processes," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 1, pp. 79–93, 2018.

[6] S. Han, V. M. Preciado, C. Nowzari, and G. J. Pappas, "Data-driven network resource allocation for controlling spreading processes," *IEEE Transactions on Network Science and Engineering*, vol. 2, no. 4, pp. 127–138, 2015.

[7] S. Roy and M. Xue, "Compressibility of network opinion and spread states in the laplacian-eigenvector basis," in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 4988–4993.

[8] C. Asavathiratham, "The influence model: A tractable representation for the dynamics of networked markov chains," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.

[9] V. Sood and S. Redner, "Voter model on heterogeneous graphs," *Physical review letters*, vol. 94, no. 17, p. 178701, 2005.

[10] M. Xue and S. Roy, "Averager-copier-voter models for hybrid consensus," *IFAC-PapersOnLine*, vol. 49, no. 4, pp. 1–6, 2016.

[11] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE signal processing magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[12] A. Montanari, Y. Eldar, and G. Kutyniok, "Graphical models concepts in compressed sensing." *Compressed Sensing*, pp. 394–438, 2012.

[13] R. Gribonval, V. Cevher, and M. E. Davies, "Compressible distributions for high-dimensional statistics," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5016–5034, 2012.