Diffusion and Consensus in a Weakly Coupled Network of Networks

Yuhao Yi, Anirban Das, Bassam Bamieh, *Fellow, IEEE*, Zhongzhi Zhang, *Member, IEEE*, and Stacy Patterson, *Member, IEEE*

***Abstract*—We study diffusion and consensus dynamics in a Network of Networks model. In this model, there is a collection of sub-networks, connected to one another using a small number of links. We consider a setting where the links between networks have small weights, or are used less frequently than links within each sub-network. Using spectral perturbation theory, we analyze the diffusion rate and convergence rate of the investigated systems. Our analysis shows that the first order approximation of the diffusion and convergence rates is independent of the topologies of the individual graphs; the rates depend only on the number of nodes in each graph and the topology of the connecting edges. The second order analysis shows a relationship between the diffusion and convergence rates and the information centrality of the connecting nodes within each sub-network. We further highlight these theoretical results through numerical examples.**

***Index Terms*—Distributed systems, gossip protocols, diffusion, randomized consensus, perturbation analysis, network of networks.**

I. INTRODUCTION

DIFFUSION and consensus dynamics play a fundamental role in the coordination of many complex networks, from networks of autonomous vehicles [1], to power grids [2], to social networks [3], and beyond. As such, significant research effort has been devoted to development of analytical characterizations of the performance of diffusion processes and consensus algorithms based on the network topology and the node interactions.

The vast majority of this work has considered a single, isolated network model. However, many complex networks can be more accurately represented by a set of interacting networks. For example, in vehicular ad-hoc networks, the network topology often consists of clusters of sub-networks, made up of co-located vehicles, that periodically communicate with one another [4]. A large-scale power network can also be viewed as a composition of subsystems which represent "areas" of the network, where the dynamics in each subsystem operates at a faster time-scale than the network-wide dynamics [5]–[7]. Another example can be found in social networks, where people are often clustered into communities; interaction within communities is frequent, and interaction across communities less so [8]. These examples motivate the *Network of Networks* (NoN) model, where multiple individual networks, or *subgraphs*, are connected using a set of links to form a connected network.

To differentiate an NoN model from a general graph, additional restrictions on the connecting graph or the subgraphs are often imposed to characterize patterns of real NoNs in different applications [9]–[12]. For example, to model a networked system where multiple signal channels work in parallel, an intensely studied model (multiplex networks) [13]–[15] assumes that the number of nodes in each subgraph is the same, and the links between any two subgraphs are given by an identity map. And, the graph Cartesian product generates NoNs with replicated subgraphs connected by inter-layer identity maps [16]. However, in the aforementioned examples of vehicular ad-hoc networks, power networks, and social networks, nodes in different subgraphs cannot be viewed as copies of the same set of nodes. In such NoNs, subgraphs may have different number of nodes and topologies. Further, the inter-network connections are often sparse [17] and weakly coupled, and are often established between gateway nodes.

We study an NoN model where inter-network edges have small weights, and each subgraph has only one node connecting to other subgraphs. Using this model, we can explicitly characterize the relationship between the subgraph topologies, the connecting graph topology, and the selection of the connecting nodes. We analyze the diffusion rate of the considered NoN model and the convergence rate of consensus algorithms in such a model using spectral perturbation theory-based methods. For the diffusion process in an NoN, we study a system in which all weights between subgraphs are multiplied by a small parameter ϵ. This setting captures diffusion processes in many complex network systems, for example, social networks with weak inter-community links. RC-circuits and their analogs are well characterized by these diffusion dynamics [5]. The linearization of the electromechanical swing dynamics of power systems can also be viewed as an extension of a diffusion process in an NoN [7], [18].

In a consensus network, we consider a setting where the

This work was supported in part by NSF under Grant CNS-1553340 and Grant CNS-1816307, in part by NSF Grant ECCS-1932777 and Grant CMI-1763064, and in part by National Natural Science Foundation of China under Grant 61872093 and Grant U20B2051.

Yuhao Yi is with the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden, and he is also affiliated with Digital Futures (email: yuhaoy@kth.se). This work was done when he was with the Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York, 12180 USA.

Anirban Das and Stacy Patterson are with the Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York, 12180 USA. (email: dasa2@rpi.edu; sep@cs.rpi.edu).

Bassam Bamieh is with the Department of Mechanical Engineering, UC Santa Barbara, Santa Barbara, CA 93106 USA (email: bamieh@ucsb.edu).

Zhongzhi Zhang is with the Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai, 200433 China (email: zhangzz@fudan.edu.cn).

links between subgraphs may be costly to use, and so they are used sparingly in the consensus algorithm. We model this setting using a stochastic system where links that connect subgraphs are active in each iteration with some small probability p. This setting applies to architectures like vehicle networks and the Internet of Things, where nearby nodes can communicate using free local communication, e.g., Bluetooth, but where distant nodes must communicate using potentially costly cellular or satellite communication.

We show that the diffusion rate is directly related to the convergence rate of the expected system of the stochastic consensus network. Our analytical results show that up to first order in ϵ, the diffusion rate depends on the generalized Laplacian matrix of the connecting graph, which is determined by the number of nodes in each subgraph and the topology of the interconnecting links. The rate does not depend on the topologies of the individual subgraphs nor on which nodes are used to connect the subgraphs to one another. The first-order approximations are accurate when ϵ is sufficiently small. The second order perturbation analysis, however, shows that choosing nodes with the largest information centrality [19] as connecting nodes maximizes the diffusion rate up to second order in ϵ. We also study the mean square convergence rate of the consensus network, which includes an additional variance term. The analysis gives similar results. In addition, we conduct experiments to show that our analysis accurately captures the behavior of the studied dynamics for small values of ϵ or p.

Related work: Several previous papers have studied the diffusion process in various NoN models. [13] provides an upper bound for the diffusion rate of a NoN where each layer of subgraph has the same number of nodes and the inter-network links between any two adjacent layers of networks are restricted to be an identity map. [14] studies the same model as [13] using perturbation theory. [20] studies optimal weights for inter-layer links in the case where intra-layer network may be directed. [15] also studies same model as [13] and derives relationships between λ_2 of the supra-Laplacian and topological properties of the subgraphs. We note that all these works are based on the homogeneous one-to-one inter-layer connection assumption made in [13], which limits the application of these models. In addition, [16] studies diffusion in Cartesian product of graphs as a model of NoN and gives some analysis based on numerical experiments. In contrast to these models, subgraphs can be arbitrary in our model, and the subgraphs are loosely coupled, both in the number and strength of the links between them.

As for discrete-time consensus dynamics, there has been a significant amount of work devoted to the analysis of distributed consensus algorithms in time-varying networks and stochastic networks, e.g., [21]–[26]. In this work, we employ a model similar to that studied in [27]–[29], which all study the convergence rate of the mean-square deviation from consensus in a stochastic network. [27] presents bounds based on the spectrum of the expected weight matrix, whereas [28] and [29] give analytical expressions for the convergence rate itself. None of these works considered an NoN model.

The NoN consensus model was introduced in [12] and [30], where they measure network performance by analyzing its robustness against random node failures. More recent work [31] considers an NoN model with noisy consensus dynamics and proposes methods to identify the optimal interconnection topology. And, in [32], the authors consider a similar NoN model, but with slightly different dynamics. They show that interconnection between the nodes of subgraphs with the highest degree maximizes the robustness of the NoN. Conditions for controllability of NoN models have also been studied recently [33], [34]. While these works focus on robustness and controllability of an NoN, our work in contrast, focuses on the rate at which nodes reach consensus, and in particular, how this rate relates to the topologies of the interconnecting network and the subgraphs.

A clustering based order reduction method has been used to model dynamics in power networks [5]–[7]. In these works the dynamics is decomposed into fast local motions and slow network-wide motion, where the slow motion is characterized by variables obtained by collapsing the subgraphs. A major difference of our work is that our aim is to study the roles of the connecting graph, the subgraphs, and the positions of the connecting nodes, from a network design point of view, rather than to reduce the order of the system. We directly analyze the spectrum of the Laplacian matrix of the network. This enables us to track every eigenvalue of the system up to its second order approximation, which reveals the impact of all the aforementioned factors. Our analysis directly implies a heuristic algorithm for choosing the optimal connecting nodes.

We note that the problem of designing network structure to improve the algebraic connectivity in a general undirected graph has been proven to be **NP**-hard [35]. As for a variation of such a problem in directed graphs, only special families of graphs have been studied [36].

A preliminary version of this work appeared in [37]. This conference paper presented first-order perturbation analysis only. Further, this analysis was restricted to consensus algorithms, In this paper, we study both diffusion and consensus dynamics, and more significantly, we include second-order perturbation analysis.

Outline: Section II describes our system model and the problem formulation. Section III provides background on spectral perturbation analysis. In Section IV, we present analysis of the diffusion and convergence rates in an NoN, including its first- and second-order behaviors. In Section V, we present our analysis of the mean square convergence of consensus algorithms for a special case of stochastic dynamics. Section VI gives numerical evaluations that highlight key results of our theoretical analysis, followed by the conclusion and discussion in Section VII.

II. System Model

A. Diffusion Dynamics in Network of Networks

We consider a system of D disjoint graphs $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i, w_i)$, $i = 1, \ldots, D$. Each graph $\mathcal{G}_i$ is weighted, undirected, and connected. We call these graphs the *subgraphs* of the NoN. The set $\mathcal{V}_i$ denotes the node set of $\mathcal{G}_i$, with $|\mathcal{V}_i| = N_i$, and $\mathcal{E}_i$ is the set of links. We let $N = \sum_{i=1}^{D} N_i$ be the total number of nodes in the network. An edge between node $r \in \mathcal{V}_i$ and $s \in \mathcal{V}_i$ is denoted by $e(r,s)$, and $\mathcal{N}_i(j)$ denotes the neighbor set of node j in subgraph $\mathcal{G}_i$. The function $w_i : \mathcal{E}_i \mapsto \mathbb{R}^+$ defines a non-negative weight $w_i(r,s)$ for each edge $e(r,s) \in \mathcal{E}_i$. Let $\mathbf{L}_i$ be the weighted Laplacian matrix of subgraph $\mathcal{G}_i$:

$$\mathbf{L}_i(r,s) = \begin{cases} \sum_{k \in \mathcal{N}_i(r)} w_i(r,s) & \text{for } r = s \\ -w_i(r,s) & \text{otherwise.} \end{cases}$$

Further, we define $\mathbf{L}_{sub}$ to be the $N \times N$ block diagonal matrix with blocks $\mathbf{L}_i$, $i = 1 \dots D$.

We construct an NoN by connecting the D subgraphs with a small number of edges. The set $\mathcal{V}$ is the NoN vertex set, $\mathcal{V} = \bigcup_{i=1}^{D} \mathcal{V}_i$, with $|\mathcal{V}| = N$. Without loss of generality, we identify the nodes in $\mathcal{V}$ as $1, 2, \dots, N$. The NoN edge set $\mathcal{E}$ consists of all edges in $\mathcal{E}_1 \cup \ldots \cup \mathcal{E}_D$, as well as a set of undirected *connecting edges* $\mathcal{E}_{con} = \{e(r,s) \mid r \in \mathcal{G}_i, s \in \mathcal{G}_j, i \neq j\}$. We call the nodes $u \in \mathcal{V}$ that are adjacent to some edge in $\mathcal{E}_{con}$ *connecting nodes*, and we denote the set of connecting nodes by $\mathcal{V}_{con}$. We assume that there is only one connecting node s_i in each subgraph $\mathcal{G}_i$. The *connecting graph* is defined as $\mathcal{G}_{con} = (\mathcal{V}, \mathcal{E}_{con}, w_{con})$, where $w_{con} : \mathcal{E}_{con} \mapsto \mathbb{R}^+$ is a function that defines a non-negative weight $w_{con}(r,s)$ for each edge $e(r,s) \in \mathcal{E}_{con}$. The weighted Laplacian matrix of the connecting graph is denoted by an $N \times N$ matrix $\mathbf{L}_{con}$. For a matrix $\mathbf{Q}$, we use the symbol $\widehat{\mathbf{Q}}$ to denote the principle submatrix of $\mathbf{Q}$ whose rows and columns correspond to vertices in $\mathcal{V}_{con}$. For example, $\widehat{\mathbf{L}}_{con}$ is the $D \times D$ weighted Laplacian of the graph $\widehat{\mathcal{G}}_{con} = (\mathcal{V}_{con}, \mathcal{E}_{con}, w_{con})$. We assume $\widehat{\mathcal{G}}_{con}$ is connected, and therefore, $\widehat{\mathbf{L}}_{con}$ has a single zero eigenvalue.

With these definitions, the NoN is thus formally defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$, where $w(r,s) = w_i(r,s)$ for $r, s \in \mathcal{V}_i$ and $w(r,s) = w_{con}(r,s)$ for $r \in \mathcal{V}_i, s \in \mathcal{V}_j, i \neq j$. We further define the *strength* of a node r as $\Delta_r = \sum_{s \in \mathcal{N}(r)} w(r,s)$, where $\mathcal{N}(r)$ denotes the neighbor set of node r in graph $\mathcal{G}$.

We study diffusion dynamics in this NoN where there is weak coupling between subgraphs. This weak coupling is enforced both by limiting the number of connecting nodes in each subgraph to one and by selecting a small inter-subgraph diffusion coefficient. For each subgraph $\mathcal{G}_i$, every node $r \in \mathcal{V}_i$ has a scalar-valued state denoted by x_r. Its dynamics is:

$$\dot{x}_r = \sum_{s \in \mathcal{N}_i(r)} w(r,s)(x_s - x_r) + \epsilon \sum_{e(r,u) \in \mathcal{E}_{con}} w(r,s)(x_u - x_r),$$

where ϵ is the diffusion coefficient between subgraphs. Let $\mathbf{x}_i$ denote the vector of node states for graph $\mathcal{G}_i$, and let $\mathbf{x}$ denote the states of all nodes in the system, i.e., $\mathbf{x} = [\mathbf{x}_1^T \ \mathbf{x}_2^T \ \dots \ \mathbf{x}_D^T]^T$. The dynamics of the entire NoN is:

$$\dot{\mathbf{x}} = -(\mathbf{L}_{sub} + \epsilon \mathbf{L}_{con})\mathbf{x}. \tag{1}$$

The matrix $\mathbf{L} = \mathbf{L}_{sub} + \epsilon \mathbf{L}_{con}$ is called the *supra-Laplacian* of the NoN.

We investigate the smallest non-zero eigenvalue of the Laplacian matrix $\mathbf{L}$, which decides the rate of diffusion in (1). It is also called the spectral gap of $\mathbf{L}$.

Definition II.1. *The spectral gap of* $\mathbf{L}$ *is defined as the smallest non-zero eigenvalue of* $\mathbf{L}$*, denoted as* $\alpha(\mathbf{L})$.

The spectral gap determines the slowest speed that the diffusion process (1) converges to its steady state from any initial state and therefore is also referred to as the *diffusion rate*. Since $\mathbf{L}$ is positive semi-definite and has eigenvalue zero with multiplicity 1 for any connected graph $\mathcal{G}$, we know that $\alpha(\mathbf{L}) > 0$. In particular, we study how the spectral gap $\alpha(\mathbf{L})$ is related to matrix $\mathbf{L}_{sub}$ and matrix $\mathbf{L}_{con}$. We recall that $\mathbf{L}_{con}$ is decided by the set of connecting nodes $\mathcal{V}_{con}$ and the structure of the connecting graph, characterized by $\widehat{\mathbf{L}}_{con}$. Further, we show how are analysis can be used to select connecting nodes within the subgraphs that maximize the spectral gap.

B. Connection to Consensus in Stochastic Networks

There is a close relationship between $\alpha(\mathbf{L})$ the convergence rate of discrete-time consensus dynamics in stochastic networks. Through this relationship, we identify an alternate interpretation of $\alpha(\mathbf{L})$. We consider a consensus network where links within each subgraph are always active, e.g., due to the proximity of agents within the subgraph to one another. Since subgraphs may be separated spatially, communication between subgraphs may be infrequent and/or lossy. We model this by activating the connecting edges in $\mathcal{E}_{con}$ each time step ℓ with some small probability p. One can define the dynamics as a consensus network with stochastic communication links. For a node $r \in \mathcal{V}_i$,

$$\begin{aligned} x_r(\ell+1) = x_r(\ell) - &\sum_{v \in \mathcal{N}_i(r)} w(r,v)\big(x_r(\ell) - x_v(\ell)\big) \\ &- \beta \sum_{e(r,s) \in \mathcal{E}_{con}} \delta_{rs}(\ell) w(r,s)(x_r(\ell) - x_s(\ell)). \end{aligned}$$

We assume that for all $r \in \mathcal{G}, r \in \mathcal{G}_i, \sum_{v \in \mathcal{N}_i(r)} w(r,v) < 1$. In addition, we assume $\beta \leq \frac{1}{2\Delta}$, where $\Delta = \max(\Delta_i)$ is the maximal node strength of $\mathcal{G}$.

$$\delta_{rs}(\ell) = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

where $\delta_{rs}(\ell)$ are Bernoulli random variables that are not necessarily mutually independent. We note that all $\delta_{rs}(\ell)$ are independent of $\mathbf{x}(\ell)$.

1) Convergence Rate of Expected System: Let $\mathbf{A}$ be the block diagonal matrix $\mathbf{A} = \mathbf{I} - \mathbf{L}_{sub}$. We also define an $N \times N$ matrix $\mathbf{B}_{rs} = \beta \cdot w(r,s) \cdot \mathbf{b}_{rs}\mathbf{b}_{rs}^T$, where $\mathbf{b}_{rs}$ is a binary N-vector with the r^{th} element equal to 1, the s^{th} element equal to -1, and the remaining elements equal to 0. The dynamics of the stochastic NoN can then be written as

$$\mathbf{x}(\ell+1) = \mathbf{A}\mathbf{x}(\ell) - \sum_{e(r,s) \in \mathcal{E}_{con}} \delta_{rs}(\ell)\mathbf{B}_{rs}\mathbf{x}(\ell). \tag{2}$$

We further let $\bar{\mathbf{x}}(\ell) = \mathbf{E}[\mathbf{x}(\ell)]$ and $\mathbf{B} = \sum_{e(r,s) \in \mathcal{E}_{con}} \mathbf{B}_{rs}$. By taking expectation of both sides of (2), we obtain

$$\bar{\mathbf{x}}(\ell+1) = \overline{\mathbf{A}}\bar{\mathbf{x}}(\ell), \tag{3}$$

where $\overline{\mathbf{A}} = \mathbf{A} - p\mathbf{B}$ is the *expected weight matrix*. The equality follows from the fact that $\delta_{rs}(\ell)$ is independent of $\mathbf{x}(\ell)$.

Definition II.2. *The* convergence rate of the expected system *of (3), denoted* $\rho_{ess}(\overline{\mathbf{A}})$*, is defined as the second largest* *eigenvalue of* $\overline{\mathbf{A}}$*, also called the* essential spectral radius *of* $\overline{\mathbf{A}}$.

Given the condition $\sum_{v\in\mathcal{N}_i(r)} w(r,v) < 1$, the matrix $\mathbf{A}_i := \mathbf{I} - \mathbf{L}_i$ has 1 as a simple eigenvalue with eigenvector $\mathbf{1}$ for all subgraph $\mathcal{G}_i$, then matrix $\mathbf{A}$ has eigenvalue 1 with multiplicity D. Since $\mathcal{G}_{con}$ is connected, the matrix $\overline{\mathbf{A}}$ has eigenvalue 1 with multiplicity 1, and its corresponding eigenvector is $\mathbf{1}$. Under the assumption $\beta \leq \frac{1}{2\Delta}$, the convergence rate of the expected system (3) is characterized by the second largest eigenvalue of $\overline{\mathbf{A}}$ [38].

Next, noting that $\mathbf{A} - p\mathbf{B} = \mathbf{I} - (\mathbf{L}_{sub} + p\beta\mathbf{L}_{con})$, we state a simple relationship between $\alpha(\mathbf{L})$ and $\rho_{ess}(\overline{\mathbf{A}})$.

Proposition II.3. *The spectral gap* $\alpha(\mathbf{L})$*, where* $\mathbf{L} = \mathbf{L}_{sub} + \epsilon\mathbf{L}_{con}$*, and the essential spectral radius* $\rho_{ess}(\overline{\mathbf{A}})$*, where* $\overline{\mathbf{A}} = \mathbf{A} - p\mathbf{B}$*, as given by Definitions II.1 and II.2, respectively, satisfy*

$$\rho_{ess}(\mathbf{A} - p\mathbf{B}) = 1 - \alpha(\mathbf{L}_{sub} + p\beta\mathbf{L}_{con}). \tag{4}$$

2) Mean-Square Convergence Rate: We also study the *mean square convergence rate* of the stochastic NoN in (3). Let $\tilde{\mathbf{x}}(\ell) = \mathbf{P}\mathbf{x}(\ell)$ be the deviation from average vector, where $\mathbf{P}$ is the projection matrix, $\mathbf{P} = (\mathbf{I}_N - \frac{1}{N}\mathbf{1}\mathbf{1}^T)$. If $\lim_{t\to\infty} \mathbf{E}\left[\|\tilde{\mathbf{x}}(\ell)\|_2\right] = 0$, we say the system *converges in mean square*.

We start by investigating the case where all edges in $\mathcal{G}_{con}$ are activated together with some probability p in each time step t. We include the discussion of the i.i.d. case in a technical report [39].

Assumption II.4. *All edges in* $\mathcal{G}_{con}$ *are online or offline with probability* p *and* $1-p$ *at time step* ℓ*, decided by a Bernoulli random variable* $\delta(\ell)$*.*

We define the autocorrelation matrix of $\tilde{\mathbf{x}}(\ell)$ by $\mathbf{\Sigma}(\ell) = \mathbf{E}\left[\tilde{\mathbf{x}}(\ell)\tilde{\mathbf{x}}(\ell)^T\right]$ and note that $\mathbf{\Sigma}(\ell) = \mathbf{E}\left[\mathbf{P}\mathbf{x}(\ell)\mathbf{x}(\ell)^T\mathbf{P}\right]$. Using a similar method to that in [40], it can be shown that $\mathbf{\Sigma}(\ell)$ satisfies the matrix recursion

$$\mathbf{\Sigma}(\ell+1) = (\mathbf{P}\bar{\mathbf{A}}\mathbf{P})\mathbf{\Sigma}(\ell)(\mathbf{P}\bar{\mathbf{A}}\mathbf{P}) + \sigma^2\mathbf{B}\mathbf{\Sigma}(\ell)\mathbf{B}. \tag{5}$$

where the zero-mean random variable $\mu(\ell)$ is defined as $\mu(\ell) = \delta(\ell) - p$, and $\sigma^2 = \mathbf{var}\left[\mu(\ell)\right]$. The errors $\mathbb{E}[\tilde{x}_r(\ell)^2]$ are given by the diagonal of $\mathbf{\Sigma}(\ell)$, and we are interested in how they evolve. We define the matrix-valued operator

$$\mathcal{A}(X) = (\mathbf{P}\bar{\mathbf{A}}\mathbf{P})X(\mathbf{P}\bar{\mathbf{A}}\mathbf{P}) + \sigma^2\mathbf{B}X\mathbf{B}, \tag{6}$$

and note that $\mathbf{\Sigma}(\ell+1) = \mathcal{A}(\mathbf{\Sigma}(\ell))$. The rate of decay of the entries of $\mathbf{\Sigma}(\ell)$ is given by the spectral radius of $\mathcal{A}$, denoted by $\rho(\mathcal{A})$ [40].

Definition II.5. *The* mean square convergence rate *of the system (2), under Assumption II.4, is defined as* $\rho(\mathcal{A})$*.*

III. Background on Spectral Perturbation Theory

Our analytical approach is based on spectral perturbation analysis [41], [42], especially the analysis where repeated eigenvalues are considered [42]. Here, we provide a brief overview of this material.

Let $\mathcal{M}(\epsilon, X)$ be a symmetric vector-valued (or matrix-valued operator) of a real parameter ϵ and a variable X of the form

$$\mathcal{M}(\epsilon, X) = \mathcal{M}_0(X) + \epsilon\mathcal{M}_1(X) + \epsilon^2\mathcal{M}_2(X) \tag{7}$$

and let $(\gamma(\epsilon), W(\epsilon))$ be an eigenvalue-eigenvector (or eigenvalue-eigenmatrix) pair of $\mathcal{M}(\epsilon, .)$, as a function of ϵ

$$\mathcal{M}(\epsilon, W(\epsilon)) = \gamma(\epsilon)W(\epsilon).$$

According to spectral perturbation theory, the functions γ and W are well-defined and analytic for small values of ϵ. The power series expansion of γ is

$$\gamma(p) = \lambda(\mathcal{M}_0) + C^{(1)}\epsilon + C^{(2)}\epsilon^2 + \cdots \tag{8}$$

where $\lambda(\mathcal{M}_0)$ is an eigenvalue of the operator $\mathcal{M}_0$, and $C^{(1)}$ and $C^{(2)}$ are coefficients for the first and second order correction terms.

Let eigenvalue $\lambda(\mathcal{M}_0)$ have multiplicity K, and let $\mathbf{W}_i$, $i = 1 \ldots K$, be K orthonormal eigenvectors (or eigenmatrices) of $\mathcal{M}_0$ that form a basis for the eigensubspace of $\lambda(\mathcal{M}_0)$. We form the $K \times K$ matrix $\mathbf{F} = [f_{i,j}]$, with entries defined by

$$f_{ij} = \frac{\langle \mathbf{W}_i, \mathcal{M}_1(\mathbf{W}_j)\rangle}{\langle \mathbf{W}_i, \mathbf{W}_i\rangle}. \tag{9}$$

When $\mathcal{M}$ is a vector-valued operator, the inner product is the standard vector inner product (for $\mathcal{M}$ a matrix-valued operator, the matrix inner product is $\langle \mathbf{X}, \mathbf{Y}\rangle := \mathbf{tr}\left(\mathbf{X}^*\mathbf{Y}\right)$). Let $\nu_1, \nu_2, \ldots, \nu_K$ be the eigenvalues of $\mathbf{F}$, with repetition. Then, the K first-order perturbation constants are $C_i^{(1)} = \nu_i$, for $i = 1 \ldots K$.

We also study the second order perturbation terms $C^{(2)}$. According to [41], [42], for an eigenvalue $\lambda(\mathcal{M}_0)$ with multiplicity $K > 1$, when $\mathbf{F}$ is diagonal, the second order terms $C_i^{(2)}$, $i = 1 \ldots K$, are

$$C_i^{(2)} = \sum_{\lambda_m(\mathcal{M}_0)\neq\lambda(\mathcal{M}_0)} \frac{\langle \mathbf{W}_i, \mathcal{M}_1(\mathbf{W}_m)\rangle^2}{\lambda(\mathcal{M}_0) - \lambda_m(\mathcal{M}_0)} \tag{10}$$

where $\mathbf{W}_i$ is the i^{th} eigenvector (or eigenmatrix) of $\mathcal{M}_0$ with eigenvalue λ, for $i = 1 \ldots K$, and $(\lambda_m(\mathcal{M}_0), \mathbf{W}_m)$ is an eigenpair of $\mathcal{M}_0$ with $\lambda_m(\mathcal{M}_0) \neq \lambda(\mathcal{M}_0)$.

IV. Analysis

In this section, we use spectral perturbation analysis to study $\alpha(\mathbf{L})$ and $\rho_{ess}(\overline{\mathbf{A}})$.

A. The Spectral Gap in Diffusion Dynamics

We first study the convergence of system (1), assuming the diffusion coefficient ϵ between subgraphs is small. The dynamics in (1) can be expressed using a vector-valued operator of the form given by (7) as $\dot{\mathbf{x}} = \mathcal{M}(\epsilon, \mathbf{x})$, where $\mathcal{M}_o(\mathbf{x}) = \mathbf{L}_{sub}\mathbf{x}$, $\mathcal{M}_1(\mathbf{x}) = \mathbf{L}_{con}\mathbf{x}$, and $\mathcal{M}_2(\mathbf{x}) = 0$.

We note that $\mathbf{L}_{sub}$ is the Laplacian matrix of a graph with D connected components (the subgraphs). Thus, it has an eigenvalue of 0 with multiplicity D. However, since $\widehat{G}_{con}$ is connected, $\mathbf{L}$ has an eigenvalue of 0 with multiplicity 1. The smallest $D-1$ nonzero eigenvalues of $\mathbf{L}$ correspond to the perturbed 0 eigenvalue of $\mathbf{L}_{sub}$. Therefore we study the perturbations to the 0 eigenvalue of $\mathbf{L}_{sub}$.

We begin by defining the generalized Laplacian matrix of the connecting graph $\widehat{\mathcal{G}}_{con}$ [43].

Definition IV.1. *Let* $\mathbf{r} = [N_1\ N_2\ \ldots\ N_D]^T$*, and let* $\mathbf{R}$ *be the* $D \times D$ *diagonal matrix with diagonal entries* $\mathbf{r}$*. The* generalized Laplacian matrix *of* $\widehat{\mathcal{G}}_{con}$ *is* $\widehat{\mathbf{M}} = \mathbf{R}^{-\frac{1}{2}}\widehat{\mathbf{L}}_{con}\mathbf{R}^{-\frac{1}{2}}$.

Note that $\widehat{\mathbf{M}}$ is symmetric positive semidefinite. It has an eigenvalue of 0 with eigenvector $\mathbf{r}^{1/2}$, and if $\widehat{\mathcal{G}}_{con}$ is connected, its second smallest eigenvalue $\lambda_2(\widehat{\mathbf{M}})$ is greater than 0. We now give a relationship between this eigenvalue and the spectral gap.

Theorem IV.2. *The spectral gap of the matrix* $\mathbf{L} = \mathbf{L}_{sub} + \epsilon\mathbf{L}_{con}$*, up to first order in* ϵ*, is*

$$\alpha(\mathbf{L}) = \epsilon\lambda_2(\widehat{\mathbf{M}})\,,$$

in which $\lambda_2(\widehat{\mathbf{M}})$ *is the smallest nonzero eigenvalue of* $\widehat{\mathbf{M}}$*.*

Proof: We determine the perturbation coefficients by forming the matrix $\mathbf{F}$ in (9). To do so, we must find an orthonormal set of eigenvectors for D zero eigenvalues of $\mathbf{L}_{sub}$, denoted as $\{\mathbf{v}_1, \ldots, \mathbf{v}_D\}$.

Let $\mathbf{u}_1, \ldots, \mathbf{u}_D$ be orthonormal eigenvectors of $\widehat{\mathbf{M}}$, and let $\lambda_1(\widehat{\mathbf{M}}) \leq \lambda_2(\widehat{\mathbf{M}}) \leq \ldots \leq \lambda_D(\widehat{\mathbf{M}})$ be the corresponding eigenvalues. We define the eigenvectors $\mathbf{v}_i$, $i = 1 \ldots D$, to be $\mathbf{v}_i = [\theta_i^{(1)}\mathbf{1}_{N_1}^T\ \theta_i^{(2)}\mathbf{1}_{N_2}^T\ \ldots\ \theta_i^{(D)}\mathbf{1}_{N_D}^T]^T$, with

$$\theta_i^{(j)} = \frac{1}{\sqrt{N_j}}u_{ij} \tag{11}$$

where u_{ij} denotes the j^{th} component of the eigenvector $\mathbf{u}_i$. We observe that the eigenvectors $\mathbf{v}_i$, $i = 1 \ldots D$, are orthonormal. We now find the entries of the $D \times D$ matrix $\mathbf{F}$ defined by (9). For f_{ij}, we have

$$\begin{aligned} f_{ij} = \langle \mathbf{v}_i, \mathbf{L}_{con}\mathbf{v}_j\rangle &= \mathbf{u}_j^T\mathbf{R}^{-\frac{1}{2}}\widehat{\mathbf{L}}_{con}\mathbf{R}^{-\frac{1}{2}}\mathbf{u}_i \\ &= \mathbf{u}_j^T\widehat{\mathbf{M}}\mathbf{u}_i = \lambda_i(\widehat{\mathbf{M}})\mathbf{u}_i^T\mathbf{u}_j. \end{aligned} \tag{12}$$

The equalities follow by the definition of $\mathbf{v}_i$, $\widehat{\mathbf{M}}$, and $\mathbf{u}_i$. If $i \neq j$, then because $\mathbf{u}_i$ and $\mathbf{u}_j$ are orthonormal, $f_{ij} = 0$. Thus $\mathbf{F}$ is a diagonal matrix, and its eigenvalues are

$$C_i^{(1)} = \lambda_i(\widehat{\mathbf{M}}),\ i = 1 \ldots D. \tag{13}$$

This completes the proof. ∎

Theorem IV.2 shows that the diffusion rate, up to first order in ϵ, is decided by an expression that depends on the smallest nonzero eigenvalue of $\widehat{\mathbf{M}}$. We note that $\widehat{\mathbf{M}}$ depends on the topology and edge weights of the connecting graph, as well as the number of vertices in each subgraph. However, $\widehat{\mathbf{M}}$ does not depend on the topology or edge weights of the subgraphs. Further, it does not depend on the choice of connecting node in each subgraph. An intuition for this result is that the connecting link is a bottleneck in the diffusion process. The diffusion rate within each graph is much faster than the diffusion rate across the connecting link. The role of the connecting link is to transfer information between the two graphs, and the amount of information that needs to be exchanged is proportional to the sizes of the graphs. It has been shown that $\lambda_i(\widehat{\mathbf{M}})$ also determines the convergence rate of load balancing diffusion algorithms in heterogeneous systems [43]. Following this analogy, we can view just the edges in $\mathcal{G}_{con}$ as executing a load balancing algorithm. The role of the connecting graph is to transfer load (i.e., node state) between the subgraphs, and the load that needs to be transferred out of each subgraph to balance the system is be proportional to the number of nodes in that subgraph.

Then we study the diffusion rate of (1) upto second order of ϵ. We note that it is decided by the spectral gap of $\mathbf{L}$.

Theorem IV.3. *The spectral gap of the matrix* $\mathbf{L} = \mathbf{L}_{sub} + \epsilon\mathbf{L}_{con}$*, up to second order in* ϵ*, is*

$$\alpha(\mathbf{L}) = \epsilon\lambda_2(\widehat{\mathbf{M}}) - \epsilon^2((\lambda_2(\widehat{\mathbf{M}}))^2(\mathbf{u}_2^*\widehat{\mathcal{S}}\mathbf{u}_2)\,, \tag{14}$$

where $\lambda_2(\widehat{\mathbf{M}})$ *is the smallest nonzero eigenvalue of* $\widehat{\mathbf{M}}$*, and* $\mathbf{u}_2$ *is its corresponding eigenvector. The* $D \times D$ *diagonal matrix* $\widehat{\mathcal{S}}$ *has diagonal entries* $\widehat{\mathcal{S}}(k,k) := N_k \cdot \mathbf{L}_k^\dagger(s_k, s_k)$*.* $\mathbf{L}_k^\dagger$ *is the Moore-Penrose inverse of* $\mathbf{L}_k$*,* s_k *is the connecting node in graph* $\mathcal{G}_k$*, and* $\mathbf{L}_k^\dagger(s_k, s_k)$ *is the diagonal entry of* $\mathbf{L}_k^\dagger$ *that corresponds to node* s_k*.*

Proof: In order to study second order perturbation coefficients using (10), we need to find all N eigenvectors of the matrix $\mathbf{L}_{sub}$.

We recall that the eigenvectors of $\mathbf{L}_{sub}$ corresponding to zero eigenvalues are defined as $\mathbf{v}_i = [\theta_i^{(1)}\mathbf{1}_{N_1}^T\ \theta_i^{(2)}\mathbf{1}_{N_2}^T\ \ldots\ \theta_i^{(D)}\mathbf{1}_{N_D}^T]^T$, where $\theta_i^{(1)}$ is defined by (11), for $i = 1 \ldots D$.

We define the remaining eigenvectors of $\mathbf{L}_{sub}$ as follows. Consider the Laplacian matrix $\mathbf{L}_i$ for subgraph i, and let $\mathbf{p}_{i_\psi}$, $\psi = 1 \ldots N_i$, be a set of N_i orthonormal eigenvectors of $\mathbf{L}_i$. Since $\mathcal{G}_i$ is connected, its 0 eigenvalue has multiplicity 1. We let $\mathbf{p}_{i_\psi}$, $\psi = 2 \ldots N_i$, be the eigenvectors associated with nonzero eigenvalues. Then we define the remaining $\mathbf{v}_m$, $m = (D+1) \ldots N$, to be $\mathbf{v}_m = [\mathbf{0}_{N_1}^T\ \ldots \mathbf{0}_{N_{k-1}}^T\ \mathbf{p}_m^T\ \mathbf{0}_{N_{k+1}}^T\ \ldots\ \mathbf{0}_{N_D}^T]^T$, where $\mathbf{p}_m \in \{\mathbf{p}_{i_\psi} : i \in [D] \text{ and } \psi \in \{2, \ldots, N_i\}\}$.

By applying (10) we attain

$$\begin{aligned} C_i^{(2)} &= \sum_{\lambda_m(\mathbf{L}_{sub}) \neq 0} \frac{\mathbf{v}_i^*\mathbf{L}_{con}\mathbf{v}_m\mathbf{v}_m^*\mathbf{L}_{con}\mathbf{v}_i}{0 - \lambda_m(\mathbf{L}_{sub})} \\ &= \sum_{\lambda_m(\mathbf{L}_{sub}) \neq 0} \frac{\widehat{\mathbf{v}}_i^*\widehat{\mathbf{L}}_{con}\widehat{\mathbf{v}}_m\widehat{\mathbf{v}}_m^*\widehat{\mathbf{L}}_{con}\widehat{\mathbf{v}}_i}{0 - \lambda_m(\mathbf{L}_{sub})} \end{aligned}$$

We recall that s_k is the vertex index of the connecting node in subgraph $\mathcal{G}_k$. Then

$$\begin{aligned} C_i^{(2)} &= \sum_{k=1}^{D} \sum_{\substack{m: \\ supp(\mathbf{v}_m) \subset \mathcal{V}_k}} \frac{\widehat{\mathbf{v}}_i^*\widehat{\mathbf{L}}_{con}(p_{m,s_k}^2\mathbf{E}_k)\widehat{\mathbf{L}}_{con}\widehat{\mathbf{v}}_i}{-\lambda_m(\mathbf{L}_k)} \\ &= \sum_{k=1}^{D} \widehat{\mathbf{v}}_i^*\widehat{\mathbf{L}}_{con}\mathbf{R}^{-\frac{1}{2}} \left(\sum_{\substack{m: \\ supp(\mathbf{v}_m) \subset \mathcal{V}_k}} \frac{p_{m,s_k}^2\mathbf{R}^{\frac{1}{2}}\mathbf{E}_k\mathbf{R}^{\frac{1}{2}}}{-\lambda_m(\mathbf{L}_k)} \right) \mathbf{R}^{-\frac{1}{2}}\widehat{\mathbf{L}}_{con}\widehat{\mathbf{v}}_i \\ &= \sum_{k=1}^{D} \widehat{\mathbf{v}}_i^*\widehat{\mathbf{L}}_{con}\mathbf{R}^{-\frac{1}{2}} \left(\sum_{\substack{m: \\ supp(\mathbf{v}_m) \subset \mathcal{V}_k}} \frac{r_{kk} \cdot p_{m,s_k}^2\mathbf{E}_k}{-\lambda_m(\mathbf{L}_k)} \right) \mathbf{R}^{-\frac{1}{2}}\widehat{\mathbf{L}}_{con}\widehat{\mathbf{v}}_i\,, \end{aligned}$$

where $\mathbf{E}_k$ is a $D \times D$ matrix with only one non-zero entry $\mathbf{E}_{k,k} = 1$. p_{m,s_k} is the entry of $\mathbf{p}_m$ associated with the connecting node s_k. We can further derive

$$\begin{aligned} C_i^{(2)} &= -\mathbf{u}_i^* \mathbf{R}^{-\frac{1}{2}} \widehat{\mathbf{L}}_{con} \mathbf{R}^{-\frac{1}{2}} \widehat{\mathcal{S}} \mathbf{R}^{-\frac{1}{2}} \widehat{\mathbf{L}}_{con} \mathbf{R}^{-\frac{1}{2}} \mathbf{u}_i \\ &= -\mathbf{u}_i^* \widehat{\mathbf{M}} \widehat{\mathcal{S}} \widehat{\mathbf{M}} \mathbf{u}_i = -(\lambda_i(\widehat{\mathbf{M}}))^2 (\mathbf{u}_i^* \widehat{\mathcal{S}} \mathbf{u}_i)\,, \end{aligned} \tag{15}$$

where the $D \times D$ diagonal matrix $\widehat{\mathcal{S}}$ has its entries $\widehat{\mathcal{S}}(k,k) := r_{kk} \cdot \mathbf{L}_k^{\dagger}(s_k, s_k)$. From (8) we attain the result. ■

We further obtain the following corollary for all the eigenvalues of $\mathbf{L}$ up to first order and second order in ϵ.

Corollary IV.4. *For any nonzero eigenvalue $\lambda_i(\mathbf{L})$, $i = 2, \ldots, D$ in the studied network of networks system (2), the first order approximation of $\lambda_i(\mathbf{L})$ is independent of the choices of connecting nodes, the second order approximation of $\lambda_i(\mathbf{L})$ is maximized when each connecting node is chosen as the one with maximum information centrality in each subgraph.*

Proof: By Theorem IV.2, the first order approximation of $\lambda_i(\mathbf{L})$ does not depend on the choice of the connecting nodes.

Then we take into account the second order perturbation terms given by (15). We note that once the structure and the weight function of the connecting graph are fixed, $\lambda_i(\widehat{\mathbf{M}})$ and $\mathbf{u}_i$ are determined for all i. As long as the choice of connecting nodes is concerned, $C_i^{(2)}$ is maximized when $\widehat{\mathcal{S}}$ is minimized in the Loewner order. This is achieved when the diagonal entries $\widehat{\mathcal{S}}(k,k)$ are all minimized simultaneously. This is then achieved when each bridge node is chosen as the node with maximum information centrality [19] in that subgraph, because $r_{k,k} = N_k$ is the same for any choice in that subgraph. ■

Corollary IV.4 shows that the second-order perturbation terms are affected by the choice of connecting node in each subgraph. The second-order approximations of all eigenvalues are maximized simultaneously when each connecting node is chosen as the node with maximum information centrality in the subgraph. This result naturally decouples the optimization problem into independent subproblems, which can be addressed in parallel.

B. Analytical Examples

1) Analysis for $D = 2$: For an NoN consisting of two subgraphs $\mathcal{G}_1$ and $\mathcal{G}_2$, the backbone graph $\mathcal{G}_{con}$ is an edge.

Corollary IV.5. *In an NoN where two subgraphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are connected by a single edge, the spectral gap $\alpha(\mathbf{L})$, up to first order in ϵ, is $\alpha(\mathbf{L}) = \epsilon N/(N_1 N_2)$.*

Proof: The generalized Laplacian matrix $\widehat{\mathbf{M}}$ is given by $\widehat{\mathbf{M}} = [1/\sqrt{N_1}\ (-1/\sqrt{N_2})]^\top [1/\sqrt{N_1}\ (-1/\sqrt{N_2})]$. $\widehat{\mathbf{M}}$ has two eigenvalues, $\lambda_1(\widehat{\mathbf{M}}) = 0$ and $\lambda_2(\widehat{\mathbf{M}}) = \frac{1}{N_1} + \frac{1}{N_2}$. Their corresponding eigenvectors are $\mathbf{u}_1 = \frac{1}{\sqrt{N}}[\sqrt{N_1}\ \sqrt{N_2}]^T$ and $\mathbf{u}_2 = \frac{1}{\sqrt{N}}[\sqrt{N_2}\ (-\sqrt{N_1})]^T$. Applying the definition for $F_i^{(1)}$ in (13), we obtain the result in Corollary IV.5. ■

This theorem shows that the first order approximation of $\alpha(\mathbf{L})$ depends on the number of nodes in each subgraph. The first order approximation does not depend on the structures of the subgraphs or the choice of bridge node within each subgraph, as we have observed in Theorem IV.2. We can also observe from Corollary IV.5 that when $N_1 = N_2 = \frac{N}{2}$, the first order approximation of $\alpha(\mathbf{L})$ is minimized.

2) Analysis for $D > 2$ with Equally Sized Graphs: We next consider the case where $N_1 = N_2 = \ldots = N_D = \frac{N}{D}$, i.e., all subgraphs have the same number of nodes.

Corollary IV.6. *Consider a composite system consisting of D subgraphs $\mathcal{G}_1, \ldots, \mathcal{G}_D$, each with $\frac{N}{D}$ nodes, and a backbone graph $\mathcal{G}_{con}$. The spectral gap $\alpha(\mathbf{L})$, up to first order in ϵ, is $\alpha(\mathbf{L}) = \epsilon\left(\frac{D}{N}\right)\lambda_2(\widehat{\mathbf{L}}_{con})$, where $\lambda_2(\widehat{\mathbf{L}})_{con}$ is the second smallest eigenvalue of $\widehat{\mathbf{L}}_{con}$.*

Proof: Given $N_1 = N_2 = \ldots = N_D = \frac{N}{D}$, we attain $\widehat{\mathbf{M}} = \frac{D}{N}\widehat{\mathbf{L}}_{con}$. Then we obtain the result in Corollary IV.6 by applying Theorem IV.2. ■

As with the case where $D = 2$, up to the first order approximation, the convergence factor is independent of the topology of the subgraphs, and it is independent of the choice of connecting nodes. The diffusion rate depends on $\lambda_2(\widehat{\mathbf{L}}_{con})$, also called the *algebraic connectivity* of the backbone graph. If $\mathcal{G}_{con}$ is not connected, then $\lambda_2(\widehat{\mathbf{L}}_{con}) = 0$, meaning, as expected, the system does not converge. The diffusion rate increases as the algebraic connectivity of $\mathcal{G}_{con}$ increases.

C. Convergence Rate of the Expected Consensus Network

Next we study the convergence rate of the the expected consensus network (3). By using the analytic results we developed in IV-A, as well as the connection between the spectral gap of $\mathbf{L}$ and the essential spectral radius of $\overline{\mathbf{A}}$, we obtain the following corollary.

Corollary IV.7. *The essential spectral radius of the expected weight matrix $\overline{\mathbf{A}}$, up to first order in p, is*

$$\rho_{ess}(\overline{\mathbf{A}}) = 1 - p\beta\lambda_2(\widehat{\mathbf{M}});$$

the essential spectral radius, upto second order in p, is

$$\rho_{ess}(\overline{\mathbf{A}}) = 1 - p\beta\lambda_2(\widehat{\mathbf{M}}) + p^2\beta^2((\lambda_2(\widehat{\mathbf{M}}))^2(\mathbf{u}_2^*\widehat{\mathcal{S}}\mathbf{u}_2).$$

We omit the proof of Corollary IV.7 because the results follow straightforwardly from Proposition II.3, Theorem IV.2, and Theorem IV.3.

According to Proposition II.3 and Corollary IV.4, we conclude that the first order approximation of $\rho_{ess}(\overline{\mathbf{A}})$ is independent of the choices of connecting nodes; the second order approximation shows that choosing nodes with maximum information centrality as the connecting node in each subgraph leads to the fastest convergence rate for the expected system.

V. Analysis of Mean Square Convergence Rate

We now use spectral perturbation analysis to study the mean square convergence rate of an NoN in which all edges in $\mathcal{E}_{con}$ are activated together with small probability p.

A. Mean Square Perturbation

We write the operator in (6) as a matrix-valued operator $\mathcal{A}(X,p)$ of both a matrix X and the small probability p in the form (7), with

$$\mathcal{A}_0(X) = \tilde{\mathbf{A}} X \tilde{\mathbf{A}} \tag{16}$$

$$\mathcal{A}_1(X) = -\mathbf{B} X \tilde{\mathbf{A}} - \tilde{\mathbf{A}} X \mathbf{B} + \mathbf{B} X \mathbf{B} \tag{17}$$

$$\mathcal{A}_2(X) = \mathbf{B} X \mathbf{B} - \mathbf{B} X \mathbf{B} = \mathbf{0}\,, \tag{18}$$

where $\tilde{\mathbf{A}} = \mathbf{PAP}$. Recall that $\mathbf{A} = \mathbf{I} - \mathbf{L}_{sub}$. Given the assumption that for all $r \in \mathcal{G}, r \in \mathcal{G}_i, \sum_{v\in\mathcal{N}_i(r)} w(r,v) < 1$, then for each subgraph $\mathcal{G}_i$, $\mathbf{L}_i$ has a single 0 eigenvalue. Then the matrix $\mathbf{L}_{sub}$ has eigenvalue 0 with multiplicity D, it follows that $\tilde{\mathbf{A}} = \mathbf{P} - \mathbf{L}$ has eigenvalue 1 with multiplicity $D-1$. Therefore, the operator $\mathcal{A}_0$ has an eigenvalue of 1 with multiplicity $(D-1)^2$. When the system is perturbed by $p\mathcal{A}_1$, these 1 eigenvalues are perturbed. The perturbed eigenvalue with largest magnitude is $\rho(\mathcal{A})$.

For any pair of eigenvectors $\mathbf{w}_i$ and $\mathbf{w}_j$ of the matrix $\tilde{\mathbf{A}}$, $\mathbf{W}_{ij} := \mathbf{w}_i \mathbf{w}_j^*$ is an eigenmatrix of $\mathcal{A}_0$ with eigenvalue $\lambda_{ij}(\mathcal{A}_0) = \lambda_i(\tilde{\mathbf{A}})\lambda_j(\tilde{\mathbf{A}})$. Because $\tilde{\mathbf{A}} = \mathbf{P} - \mathbf{L}$ is symmetric, its left and right eigenvectors satisfy $\mathbf{w}_i^*\mathbf{w}_i = 1$ for $i \in [N]$ and $\mathbf{w}_i^*\mathbf{w}_j = 0$ for any $i,j \in [N]$, $i \neq j$.

Lemma V.1. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an NoN with the dynamics as defined in (2). There exists a set of vectors $\{\mathbf{w}_i : i = 2,\ldots,D\}$ and an induced set of matrices $\{\mathbf{W}_{ij} = \mathbf{w}_i\mathbf{w}_j^* : i,j \in \{2,\ldots,D\}\}$ such that*

$$\mathcal{A}_0(\mathbf{W}_{ij}) = \mathbf{W}_{ij}, \qquad \forall i,j \in \{2,\ldots,D\}\,, \tag{19}$$

$$\mathbf{w}_i^*\mathbf{w}_i = 1, \qquad \forall i, \{2,\ldots,D\}\,, \tag{20}$$

$$\mathbf{w}_i^*\mathbf{w}_j = 0, \qquad \forall i,j \in \{2,\ldots,D\}, i \neq j \tag{21}$$

$$\mathbf{w}_i^*\mathbf{1} = 0, \qquad \forall i \in \{2,\ldots,D\}\,, \tag{22}$$

$$\mathbf{w}_i^*\mathbf{B}\mathbf{w}_j = 0, \qquad \forall i,j \in \{2,\ldots,D\}, i \neq j \tag{23}$$

The mean square convergence rate of system (2) satisfying Assumption II.4, up to first order in p, is $\rho(\mathcal{A}) = \max_{ij}\left(1 + p f_{ij}^{(1)}\right)$, in which

$$f_{ij}^{(1)} = -\mathbf{w}_i^*\mathbf{B}\mathbf{w}_i - \mathbf{w}_j^*\mathbf{B}\mathbf{w}_j + (\mathbf{w}_i^*\mathbf{B}\mathbf{w}_i)\left(\mathbf{w}_j^*\mathbf{B}\mathbf{w}_j\right)\,. \tag{24}$$

Proof: Let $\mathbf{M}_{ij} = \mathbf{m}_i\mathbf{m}_j^*$, $i,j \in \{2\ldots D\}$ be any set of (mutual) orthonormal eigenmatrices of $\mathcal{A}_0$ associated with eigenvalue 1. The vectors $\mathbf{m}_i$, $i = 2\ldots D$ are eigenvectors of $\tilde{\mathbf{A}}$ such that $\tilde{\mathbf{A}}\mathbf{m}_i = \mathbf{m}_i$; further, they are mutually orthonormal and are all orthogonal to the vector $\mathbf{1}$.

We define a matrix $\mathbf{H}$ whose entries are defined as $h_{ij} = \mathbf{m}_i^*\mathbf{B}\mathbf{m}_j$. Let $\mathbf{U}$ be the matrix whose columns are $\mathbf{m}_i$, $i \in \{2\ldots D\}$. Then it is clear that $\mathbf{H} = \mathbf{U}^*\mathbf{B}\mathbf{U}$. Let $\mathbf{H} = \mathbf{S}\boldsymbol{\Lambda}\mathbf{S}^*$ be the spectral decomposition of $\mathbf{H}$. $\mathbf{S}$ is an unitary matrix, $\mathbf{s}_i$ is the ith column of $\mathbf{S}$. Therefore $\mathbf{B} = \mathbf{U}\mathbf{S}\boldsymbol{\Lambda}\mathbf{S}^*\mathbf{U}^*$. We define $\mathbf{w}_i := \mathbf{U}\mathbf{s}_i$, for all $i \in \{2\ldots D\}$. It is easy to verify that the vectors in $\{\mathbf{w}_i : i = 2,\ldots,D\}$ satisfy the properties (19)-(23) stated in the lemma. We note that by (20) and (21), $\langle \mathbf{W}_{ij}, \mathbf{W}_{ij}\rangle = 1$ for all $i,j \in \{2\ldots D\}$; $\langle \mathbf{W}_{ij}, \mathbf{W}_{pq}\rangle = 0$ for all $i \neq p$ or $j \neq q$. Therefore, we consider the entries of the $(D-1)^2 \times (D-1)^2$ matrix $\mathbf{F}$:

$$\begin{aligned} f_{ij,pq} &= \langle \mathbf{w}_i\mathbf{w}_j^*, \mathcal{A}_1(\mathbf{w}_p\mathbf{w}_q^*)\rangle \\ &= \mathbf{tr}\left(\mathbf{w}_j\mathbf{w}_i^*\left(-\mathbf{B}\mathbf{w}_p\mathbf{w}_q^*\tilde{\mathbf{A}} - \tilde{\mathbf{A}}\mathbf{w}_p\mathbf{w}_q^*\mathbf{B} + \mathbf{B}\mathbf{w}_p\mathbf{w}_q^*\mathbf{B}\right)\right) \\ &= -\mathbf{tr}\left(\mathbf{w}_j\mathbf{w}_i^*\mathbf{B}\mathbf{w}_p\mathbf{w}_q^*\right) - \mathbf{tr}\left(\mathbf{w}_j\mathbf{w}_i^*\mathbf{w}_p\mathbf{w}_q^*\mathbf{B}\right) \\ &\quad + \mathbf{tr}\left(\mathbf{w}_j\mathbf{w}_i^*\mathbf{B}\mathbf{w}_p\mathbf{w}_q^*\mathbf{B}\right) \end{aligned} \tag{25}$$

where the last equality holds since $\tilde{\mathbf{A}}\mathbf{w}_p = \mathbf{w}_p$ and similarly, $\mathbf{w}_q^*\tilde{\mathbf{A}} = \mathbf{w}_q^*$. the expression can further be written as

$$\begin{aligned} f_{ij,pq} = &- \mathbf{w}_i^*\mathbf{B}\mathbf{w}_p\mathbf{w}_q^*\mathbf{w}_j - \mathbf{w}_i^*\mathbf{w}_p\mathbf{w}_q^*\mathbf{B}\mathbf{w}_j \\ &+ \left(\mathbf{w}_i^*\mathbf{B}\mathbf{w}_p\right)\left(\mathbf{w}_j^*\mathbf{B}\mathbf{w}_q\right). \end{aligned}$$

If $i = p$ and $j = q$, then noting that $\mathbf{w}_i^*\mathbf{w}_p = 1$ and $\mathbf{w}_j^*\mathbf{w}_q = 1$, it follows that

$$f_{ij}^{(1)} := f_{ij,ij} = -\mathbf{w}_i^*\mathbf{B}\mathbf{w}_i - \mathbf{w}_j^*\mathbf{B}\mathbf{w}_j + (\mathbf{w}_i^*\mathbf{B}\mathbf{w}_i)\left(\mathbf{w}_j^*\mathbf{B}\mathbf{w}_j\right).$$

Furthermore, since $\mathbf{w}_i^*\mathbf{B}\mathbf{w}_j = 0$ for any $i \neq j$, all off diagonal entries are zeros. ∎

We next use this lemma to characterize the convergence factor in two classes of NoNs.

B. Analysis for Special Cases

We give results for the mean square convergence rate for the two cases which we have discussed in Section IV.

Corollary V.2. *For an NoN consisting of two subgraphs $\mathcal{G}_1$ and $\mathcal{G}_2$, with the dynamics (2) satisfying Assumption II.4, the mean square convergence rate, up to first order in p, is*

$$\rho(\mathcal{A}) = 1 - 2p\beta N/(N_1N_2) + p\beta^2\left(N/(N_1N_2)\right)^2\,. \tag{26}$$

Proof: We define the vector $\mathbf{w}_2 = [\theta^{(1)}\mathbf{1}_{N_1}^T \;\; \theta^{(2)}\mathbf{1}_{N_2}^T]^T$, where $\theta^{(1)} = \sqrt{\frac{N_2}{N\cdot N_1}}$ and $\theta^{(2)} = -\sqrt{\frac{N_1}{N\cdot N_2}}$. It is easily observed that $\mathbf{w}_2$ is an eigenvector of $\tilde{\mathbf{A}}$ with eigenvalue 1, and $\mathbf{w}_2$ is orthogonal to $\mathbf{1}$. When $D = 2$, the matrix $\mathbf{F}$ consists of a single element. Applying the definition for $f_{22}^{(1)}$ in (24),

$$\begin{aligned} f_{22}^{(1)} &= -2\beta(\theta^{(1)} - \theta^{(2)})^2 + \beta^2(\theta^{(1)} - \theta^{(2)})^4 \\ &= -2\beta\left(N/(N_1N_2)\right) + \beta^2\left(N/(N_1N_2)\right)^2. \end{aligned}$$ ∎

From (26) we observe that given N, the magnitude of $\rho(\mathcal{A})$ is maximized when the graphs are of the same size, i.e., $N_1 = N_2$. It is minimized when $N_1 = 1$, $N_2 = N-1$ or $N_2 = 1$, $N_1 = N-1$. This means that the speed of convergence is slower between balanced subgraphs. By comparing (26) to Corollary IV.5 we note that for two subgraphs, both $\rho_{ess}(\overline{\mathbf{A}})$ and $\rho(\mathcal{A})$ are determined by the strength (activation probability) of the connecting graph and the number of nodes in both subgraphs.

Corollary V.3. *For an NoN consisting of D subgraphs $\mathcal{G}_1,\ldots,\mathcal{G}_D$, each with $\frac{N}{D}$ nodes, with the system dynamics (2) satisfying Assumption II.4, the mean square convergence factor, up to first order in p, is*

$$\rho(\mathcal{A}) = 1 - p\left(2\beta\left(D/N\right)\lambda_2(\widehat{\mathbf{L}}_{con}) - \beta^2\left(D/N\right)^2\left(\lambda_2(\widehat{\mathbf{L}}_{con})\right)^2\right).$$

where $\lambda_2(\mathbf{L}_{con})$ is the second smallest eigenvalue of $\mathbf{L}_{con}$.

Proof: We obtain this result by defining the $D-1$ eigenvectors of $\tilde{\mathbf{A}}$ with eigenvalue 1 as follows. Let $\mathbf{u}_1, \ldots, \mathbf{u}_D$ be an orthonormal set of eigenvectors of the $D \times D$ matrix $\widehat{\mathbf{L}}_{con}$ with eigenvalues $0 = \lambda_1(\widehat{\mathbf{L}}_{con}) \leq \ldots \leq \lambda_D(\widehat{\mathbf{L}}_{con})$. Let $\mathbf{u}_1 = (1/\sqrt{D})\mathbf{1}$, and thus $\mathbf{L}_{con}\mathbf{u}_0 = 0$. The i^{th} eigenvector of $\tilde{\mathbf{A}}$, $i = 2 \ldots D$, is $\mathbf{w}_i = [\theta_i^{(1)}\mathbf{1}_{N_1}^T \;\; \theta_i^{(2)}\mathbf{1}_{N_2}^T \;\; \ldots \;\; \theta_i^{(D)}\mathbf{1}_{N_D}^T]^T$ with $\theta_i^{(j)} = (\sqrt{D/N})u_{ij}$, where u_{ij} denotes the j^{th} component of the eigenvector $\mathbf{u}_i$, $j = 1 \ldots D$. Therefore, the first order perturbation term of the eigenvalue corresponds to eigenmatrix $\mathbf{W}_{ij} = \mathbf{w}_i\mathbf{w}_j^*$ is $f_{ij}^{(1)} = -\beta\left(\frac{D}{N}\right)(\lambda_i(\widehat{\mathbf{L}}_{con}) + \lambda_j(\widehat{\mathbf{L}}_{con})) + \beta^2\left(\frac{D}{N}\right)^2\lambda_i(\widehat{\mathbf{L}}_{con})\lambda_j(\widehat{\mathbf{L}}_{con})$. By Lemma V.1,

$$\rho(\mathcal{A}) = \max_{i,j\in\{2,\ldots,D\}} 1 - p\Big(2\beta\,(D/N)\Big(\lambda_i(\widehat{\mathbf{L}}_{con}) + \lambda_j(\widehat{\mathbf{L}}_{con})\Big) - \beta^2\,(D/N)^2\,\lambda_i(\widehat{\mathbf{L}}_{con})\lambda_j(\widehat{\mathbf{L}}_{con})\Big). \quad (27)$$

The maximum node degree of any node $v \in \mathcal{V}_{con}$ is $D-1$; thus, the eigenvalues of $\widehat{\mathbf{L}}_{con}$ are in the interval $[0, 2\Delta]$ [44]. Since $\beta < \frac{1}{2\Delta}$, we have $\beta\lambda_j(\widehat{\mathbf{L}}_{con}) \in [0,1)$ for $j = 2 \ldots D$. Further we attain that $2\beta(\frac{D}{N}) - \beta^2(\frac{D}{N})^2\lambda_j(\widehat{\mathbf{L}}_{con}) > 0$ for $j = 2 \ldots D$. Thus, the right hand side of expression (27) is maximized when $\lambda_i(\widehat{\mathbf{L}}_{con})$ is minimized. The same analysis holds for $\lambda_j(\widehat{\mathbf{L}}_{con})$. So the right hand side of expression (27) is maximized when both $\lambda_i(\widehat{\mathbf{L}}_{con})$ and $\lambda_j(\widehat{\mathbf{L}}_{con})$ are equal to $\lambda_2(\widehat{\mathbf{L}}_{con})$, which proves the theorem. ■

We observe from Corollary V.3 and Corollary IV.6 that for subgraphs with the same number of nodes, both $\rho_{ess}(\overline{\mathbf{A}})$ and $\rho(\mathcal{A})$ are determined by the coupling strength, the algebraic connectivity of the connecting graph, as well as the number of nodes in each subgraph.

We note that the second-order perturbation analysis similar to Corollary IV.4 can also be applied to the analysis of mean-square convergence rate of (3) satisfying Assumption II.4. We leave the related discussion to a technical report [39].

VI. Numerical Results

We give some numerical examples to support our analytic results. Edges are weighted 1 in these examples unless otherwise specified. All experiments were done in MATLAB.

First, we investigate the spectral gap of the supra-Laplacian matrix in the diffusion dynamics. In Fig. 1, we compare the spectral gap estimated by first order perturbation analysis (labeled 'SPA') and second order perturbation analysis (labeled 'SPA2') to the spectral gap directly computed using $\mathbf{L}$ (labeled 'Exact') for various ϵ. Each figure shows plots for different numbers of subgraphs, $D = 2$, $D = 4$, and $D = 8$, as the sizes of the subgraphs increase. Each subgraph is an Erdős Rényi (ER) graph with connecting probability 0.6. In each NoN, all subgraphs have the same number of nodes. The connecting graph $\mathcal{G}_{con}$ is a complete graph, and the connecting node is chosen uniformly at random in each subgraph.

As expected, the spectral gap decreases as the sizes of the individual subgraphs increase. Also, in general, we see the trend that when ϵ is held constant, with larger values of D, the spectral gap is higher. We explore this phenomenon further in subsequent experiments. We observe that the spectral gap generated by first- and second-order perturbation analysis closely approximates the exact diffusion rate for $\epsilon = 0.001$ to $\epsilon = 0.01$. This is in accordance with spectral perturbation theory. The result given by SPA diverges from the exact diffusion rate for a larger value $\epsilon = 0.1$. However, SPA2 still gives good approximation for the spectral gap when $\epsilon = 0.1$.

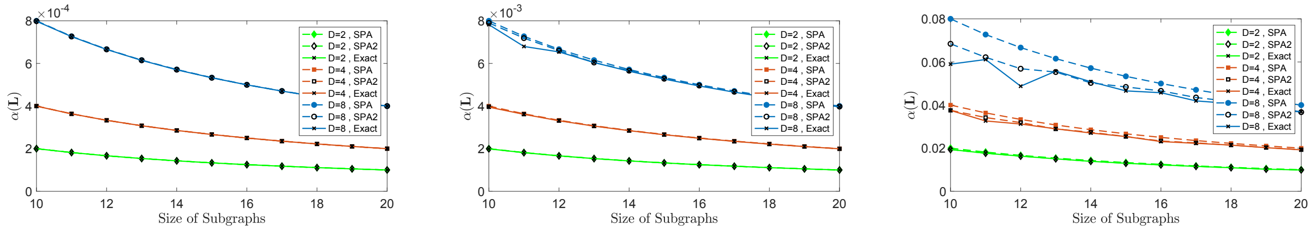
In Fig. 2, we show results using the same network scenarios as in Fig. 1, with the exception that the connecting graphs $\mathcal{G}_{con}$ are path graphs. To make the experiment homogeneous, the connecting nodes are selected as end nodes of each path graph. Again, we note the spectral gap decreases as the size of individual subgraphs increase for all ϵ. The results of SPA and SPA2 closely approximate the exact spectral gap for $\epsilon = 0.001$. The result of SPA2 still well approximates the spectral gap for $\epsilon = 0.01$, though with less accuracy than in Fig. 1. Both SPA and SPA2 fail to approximate the spectral gap for $\epsilon = 0.1$. Thus, we observe that the accuracy of the spectral perturbation analysis depends on the network topology. For each topology, there is some threshold for which, when ϵ is smaller than this threshold, the approximations are accurate. However, this threshold varies for different topologies.

We also note that, in comparing Fig. 1 and Fig. 2, it can be observed that the diffusion rate given by SPA coincide for networks of the same size. This conforms with our analysis that the first-order approximation of convergence factor of the NoN obtained from spectral perturbation analysis only depends on the sizes of the subgraphs and not on their individual topologies.

In Fig. 3 we study the dependency of the spectral gap on the topology of $\mathcal{G}_{con}$ as the number of subgraphs varies. Each subgraph is an ER graph with edge probability 0.6. All subgraphs have 10 nodes. We let $\epsilon = 0.01$, and we compute the diffusion rates when $\mathcal{G}_{con}$ is a complete graph or a ring.

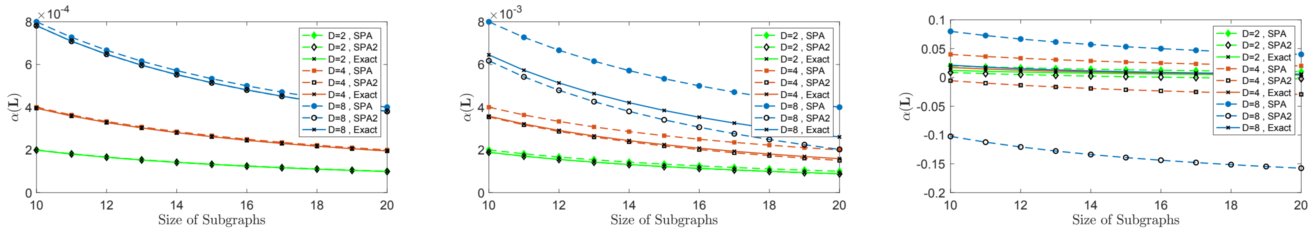
We observe that, when $\mathcal{G}_{con}$ is complete, the spectral gap of $\mathbf{L}$, in Exact, SPA , and SPA2, increases with the increase in the number of subgraphs. To better understand this phenomenon, let us assume all subgraphs are of the same size Φ. For $\mathcal{G}_{con}$ a complete graph, $\mathbf{L}_{con}$ has one eigenvalue of 0 and $D-1$ eigenvalues equal to D. For the SPA diffusion rate given in Theorem IV.2, we know that up to first order in ϵ, $\rho(\mathcal{A}) = \epsilon\left(\frac{D}{N}\right)\lambda_2(\widehat{\mathbf{L}}_{con}) = \epsilon\left(\frac{D}{\Phi}\right)$. Since Φ and ϵ are held constant, with the increase in D, the diffusion rate increases. We also note that the diffusion rate when $\mathcal{G}_{con}$ is a ring graph is smaller than the diffusion rate when $\mathcal{G}_{con}$ is a complete graph. This can be explained in part by the fact that the algebraic connectivity of a ring graph decreases as its number of nodes increases.

In Fig. 4 and Fig. 5 we show the exact diffusion rates (given by $\alpha(\mathbf{L})$) and the mean square convergence rates (given by $\rho(\mathcal{A})$) of systems with different connecting nodes. In both examples we have two ER subgraphs connected by a single edge. The probability that two nodes in the same subgraph are connected is set to 0.2. And both subgraphs are connected. For the diffusion dynamics, we set $\epsilon = 0.1$. For the consensus dynamics, we let $p = 0.1$ and $\beta = \frac{1}{21}$, and $w = \frac{1}{21}$ for all edges in both subgraphs. In the proposed heuristic, we choose the bridge nodes as the ones with maximum information centrality in each subgraph. We compare the results with the true optimum given by brute-force search, as well as the result of a random choice. The results show that our strategy hits



(a) $\epsilon = 0.001$ (b) $\epsilon = 0.01$ (c) $\epsilon = 0.1$

Fig. 1: Spectral gap of the supra-Laplacian matrix, Exact and predicated by perturbation analysis (SPA and SPA2), for composite graphs as the sizes of the individual graphs increase, for various ϵ. The individual graphs are ER graphs with connecting probability 0.6, and the connecting graph $\mathcal{G}_{con}$ is a complete graph.



(a) $\epsilon = 0.001$ (b) $\epsilon = 0.01$ (c) $\epsilon = 0.1$

Fig. 2: Spectral gap of the supra-Laplacian matrix, Exact and evaluated by perturbation analysis (SPA and SPA2), for composite graphs as the sizes of the individual graphs increase, for various values of ϵ. The individual graphs are path graphs, and the connecting graph $\mathcal{G}_{con}$ is a complete graph.

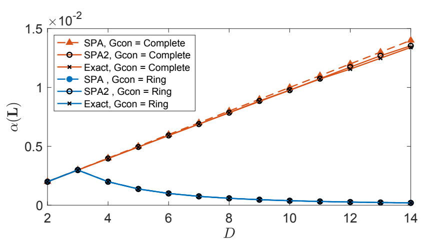


Fig. 3: Spectral gap for Exact, SPA, and SPA2, with increasing NoN sizes for ring and complete $\mathcal{G}_{con}$ topologies. Subgraphs graphs are ER graphs each with 10 nodes. ϵ is set to 0.01.

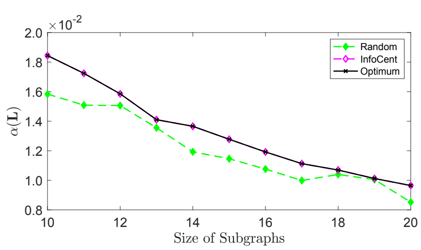


Fig. 4: Diffusion rates for the system consists of two subgraphs connected by an edge with bridge nodes selected by different strategy. Subgraphs are ER graphs with connecting probability 0.2. The diffusion coefficient ϵ is set to 0.1.

optimal solutions in all occasions, and evidently outperforms the random strategy. We have shown in Theorem IV.4 that the second-order approximation of spectral radius of the supra-Laplacian is maximized when connecting nodes are chosen as the ones with largest information centrality. In Fig. 4, we show that by using this result we actually obtain an optimal

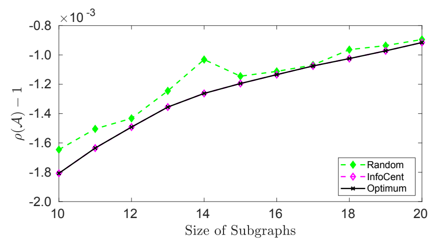


Fig. 5: Mean square convergence rates for the system consists of two subgraphs connected by an edge with bridge nodes selected by different strategy. Subgraphs are ER graphs with connecting probability 0.2. The activation probability of edges in $\mathcal{E}_{con}$ is $p = 0.1$. β takes the value of $1/21$.

connecting node in each subgraphs. In Fig. 5, similar results are observed. Therefore, empirically speaking, this approach can also be used as a heuristic to find connecting nodes that lead to a good mean square convergence rate.

Finally, we apply our algorithm to an NoN with subnetworks obtained from the University of Illinois' Power System Test Cases [45]. We choose five networks from the data set and use them as subgraphs, which are connected by a path graph or a complete graph using only one connecting node for each subgraph. For the path connecting graph, the subgraphs are connected in the order G_9, G_{24}, G_{39}, G_{30}, G_{14}, where the subscripts indicate the sizes of the corresponding subgraphs.

Table I shows the actual diffusion rate of the system in which connecting nodes are chosen as the nodes with optimal information centrality or as random nodes, for the two different connecting graphs. The results show that our algorithm picks better connecting nodes than random choices. We note that the

TABLE I: Diffusion Rates for Different Connecting Graphs.

Connecting Gaph	InfoCent	Random
Path	0.2311	0.2253
Complete Graph	1.2609	1.0594

improvement is more significant when the connecting graph is a complete graph.

VII. Conclusion and Discussion

We have investigated the rate of diffusion in an NoN model, as well as the convergence rate in a consensus NoN with a stochastically switching connecting graph. We showed that the first-order perturbation term is determined by the spectral gap of the generalized Laplacian matrix of the connecting network. In addition, using second-order perturbation analysis, we showed the connection between information centrality and the optimal connecting nodes in subgraphs. Finally, we presented numerical results to substantiate our analysis.

We note that the application of algebraic connectivity is not limited to diffusion and consensus processes. For example, it also determines the mixing time of a Markov chain. In addition, the whole spectrum of the graph can be approximated using the perturbation theory approach. Therefore, properties related to robustness, such as the $\mathcal{H}_2$ norm of the system, can be analyzed. Further, local behaviors of nonlinear systems can be linearized and approximated using this approach.

In future work, we plan to extend our analysis to more general models of NoNs. One can apply similar analysis to a model with multiple bridge nodes in each subgraph. By defining the condensed graph as a graph generated from the connecting graph by collapsing all nodes in the same subgraph into a single node, one can show that the first order perturbation term is determined by the spectral gap of the generalized Laplacian matrix of the condensed graph. The second order perturbation analysis will require more analytical expressions to uncover the roles of the subgraphs and the connecting nodes in such a model.

References

[1] W. Ren and R. W. Beard, *Distributed consensus in multi-vehicle cooperative control*. Springer, 2008.

[2] S. Kar and G. Hug, "Distributed robust economic dispatch in power systems: A consensus + innovations approach," in *2012 IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–8.

[3] R. Hegselmann, U. Krause *et al.*, "Opinion dynamics and bounded confidence models, analysis, and simulation," *J. Artif. Soc. Soc. Simul.*, vol. 5, no. 3, 2002.

[4] F. Li and Y. Wang, "Routing in vehicular ad hoc networks: A survey," *IEEE Veh. Tech. Mag.*, vol. 2, no. 2, pp. 12–22, June 2007.

[5] J. H. Chow, G. Peponides, P. Kokotovic, B. Avramovic, and J. Winkelman, *Time-scale modeling of dynamic networks with applications to power systems*. Springer, 1982, vol. 46.

[6] J. Chow and P. Kokotovic, "Time scale modeling of sparse dynamic networks," *IEEE Trans. Autom. Control*, vol. 30, no. 8, pp. 714–722, 1985.

[7] D. Romeres, F. Dörfler, and F. Bullo, "Novel results on slow coherency in consensus and power networks," in *Euro. Control Conf.* IEEE, 2013, pp. 742–747.

[8] M. Granovetter, "The strength of weak ties: A network theory revisited," *Sociological theory*, pp. 201–233, 1983.

[9] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *science*, vol. 328, no. 5980, pp. 876–878, 2010.

[10] R. Parshani, S. V. Buldyrev, and S. Havlin, "Interdependent networks: Reducing the coupling strength leads to a change from a first to second order percolation transition," *Phys. Rev. Lett.*, vol. 105, no. 4, p. 048701, 2010.

[11] J. Gao, D. Li, and S. Havlin, "From a single network to a network of networks," *National Science Review*, vol. 1, no. 3, pp. 346–356, 2014.

[12] J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, "Robustness of a network of networks," *Phys. Rev. Lett.*, vol. 107, no. 19, p. 195701, 2011.

[13] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C. J. Perez-Vicente, Y. Moreno, and A. Arenas, "Diffusion dynamics on multiplex networks," *Physical review letters*, vol. 110, no. 2, p. 028701, 2013.

[14] A. Sole-Ribalta, M. De Domenico, N. E. Kouvaris, A. Diaz-Guilera, S. Gomez, and A. Arenas, "Spectral properties of the laplacian of multiplex networks," *Physical Review E*, vol. 88, no. 3, p. 032807, 2013.

[15] G. Cencetti and F. Battiston, "Diffusive behavior of multiplex networks," *New Journal of Physics*, vol. 21, no. 3, p. 035006, mar 2019.

[16] H. Shao, Y. Xi, M. Mesbahi, D. Li, Y. Xu, and Z. Gan, "Relative tempo of consensus dynamics on multiplex networks," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 5184 – 5189, 2017, 20th IFAC World Congress.

[17] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[18] F. Dorfler and F. Bullo, "Kron reduction of graphs with applications to electrical networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 1, pp. 150–163, 2013.

[19] K. Stephenson and M. Zelen, "Rethinking centrality: Methods and examples," *Social networks*, vol. 11, no. 1, pp. 1–37, 1989.

[20] A. Tejedor, A. Longjas, E. Foufoula-Georgiou, T. T. Georgiou, and Y. Moreno, "Diffusion dynamics and optimal coupling in multiplex networks with directed layers," *Physical Review X*, vol. 8, no. 3, p. 031071, 2018.

[21] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988–1001, June 2003.

[22] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Fourth Int. Sym. Information Processing in Sensor Networks*, April 2005, pp. 63–70.

[23] H. K. Mousavi, C. Somarakis, M. Bahavarnia, and N. Motee, "Performance bounds and optimal design of randomly switching linear consensus networks," in *Proc. American Control Conf.*, May 2017, pp. 4347–4352.

[24] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Trans. Autom. Control*, vol. 50, no. 2, pp. 169–182, Feb 2005.

[25] J. Zhou and Q. Wang, "Convergence speed in distributed consensus over dynamically switching random networks," *Automatica*, vol. 45, no. 6, pp. 1455–1461, 2009.

[26] N. Abaid, I. Igel, and M. Porfiri, "On the consensus protocol of conspecific agents," *Linear Algebra Appl.*, vol. 437, no. 1, pp. 221–235, 2012.

[27] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3315–3326, July 2008.

[28] F. Fagnani and S. Zampieri, "Average consensus with packet drop communication," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 102–133, 2009.

[29] S. Patterson, B. Bamieh, and A. El Abbadi, "Convergence rates of distributed average consensus with stochastic link failures," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 880–892, April 2010.

[30] T. P. Peixoto and S. Bornholdt, "Evolution of robust network topologies: Emergence of central backbones," *Phys. Rev. Lett.*, vol. 109, no. 11, p. 118703, 2012.

[31] E. Mackin and S. Patterson, "Optimizing the coherence of a network of networks," *IEEE Trans. Control of Netw. Syst.*, pp. 1–1, 2020.

[32] R. Santini, A. Gasparri, F. Pasqualetti, and S. Panzieri, "Network composition for optimal disturbance rejection," in *Proc. American Control Conf.*, July 2016, pp. 3764–3769.

[33] Y. Hao, Q. Wang, Z. Duan, and G. Chen, "Controllability of kronecker product networks," *Automatica*, vol. 110, p. 108597, 2019.

[34] Y. Hao, Z. Duan, G. Chen, and F. Wu, "New controllability conditions for networked, identical LTI systems," *IEEE Trans. Autom. Control*, vol. 64, no. 10, pp. 4223–4228, 2019.

[35] D. Mosk-Aoyama, "Maximum algebraic connectivity augmentation is np-hard," *Oper. Res. Lett.*, vol. 36, no. 6, pp. 677–679, 2008.

[36] Y. Hao, Q. Wang, Z. Duan, and G. Chen, "The role of reverse edges on consensus performance of chain networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.

[37] A. Das, Y. Yi, S. Patterson, B. Bamieh, and Z. Zhang, "Convergence rate of consensus in a network of networks," in *Proc. 57th IEEE Conf. Decision and Control*, 2018, pp. 459–465.

[38] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65 – 78, 2004.

[39] Y. Yi, A. Das, S. Patterson, B. Bamieh, and Z. Zhang, "Diffusion and consensus in a weakly coupled network of networks," *arXiv preprint arXiv:2002.06109*, 2020.

[40] S. Patterson and B. Bamieh, "Convergence rates of consensus algorithms in stochastic networks," in *Proc. 29th IEEE Conf. Decision and Control*. IEEE, 2010, pp. 6608–6613.

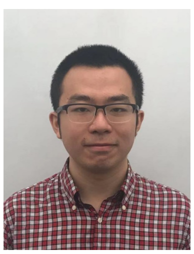
[41] H. Baumgärtel, *Analytic Perturbation Theory for Matrices and Operators*. Birkhäuser, 1985.

[42] B. Bamieh, "A tutorial on matrix perturbation theory (using compact matrix notation)," *arXiv preprint arXiv:2002.05001*, 2020.

[43] T. Rotaru and H.-H. Nägeli, "Dynamic load balancing by diffusion in heterogeneous systems," *J. Parallel Distrib. Comput.*, vol. 64, no. 4, pp. 481–497, 2004.

[44] R. Merris, "Laplacian matrices of graphs: a survey," *Linear Algebra Appl.*, vol. 197, pp. 143–176, 1994.

[45] Information Trust Institute, University of Illinois at Urbana-Champaign, "Power system test cases," accessed 2020-02-08. [Online]. Available: Available at https://icseg.iti.illinois.edu/power-cases/.



Yuhao Yi is a Postdoctoral Researcher in the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science at KTH Royal Institute of Technology. He received the B.Sc. degree and the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2013 and 2019, respectively. From 2016-2017, he was a Visiting Scholar with the Rensselaer Polytechnic Institute, Troy, NY, USA, where he was a Postdoctoral Researcher from 2019-2020. His research interests include network science and distributed control. His current research interests include analyzing and optimizing the performance of large-scale networked systems.

Anirban Das is a Ph.D. candidate in the department of Computer Science at Rensselaer Polytechnic Institute, Troy, USA. He has received a M.S. degree in computer science from the same university in 2020. He received his Bachelor of Technology degree in electronics and communication engineering in 2014 from WBUT, Kolkata, India. His research interests lie in distributed systems, distributed machine learning and resource allocation in cloud computing.

Bassam Bamieh (F'08) received the B.Sc. degree in electrical engineering and physics from Valparaiso University, Valparaiso, IN, USA, in 1983, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from Rice University, Houston, TX, USA, in 1986 and 1992, respectively. From 1991 to 1998 he was an Assistant Professor with the Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, after which he joined the University of California at Santa Barbara (UCSB) where he is currently a Professor of Mechanical Engineering. His research interests include robust and optimal control, distributed and networked control and dynamical systems, shear flow transition and turbulence, and the use of feedback in thermoacoustic energy conversion devices. He is a past recipient of the IEEE Control Systems Society G. S. Axelby Outstanding Paper Award (twice), the AACC Hugo Schuck Best Paper Award, and the National Science Foundation CAREER Award. He is Distinguished Lecturer of the IEEE Control Systems Society, a Fellow of the IEEE, and a Fellow of the International Federation of Automatic Control (IFAC).

Zhongzhi Zhang (M'19) received the B.Sc. degree in applied mathematics from Anhui University, Hefei, China, in 1997 and the Ph.D. degree in management science and engineering from Dalian University of Technology, Dalian, China, in 2006.

From 2006 to 2008, he was a Post-Doctoral Research Fellow with Fudan University, Shanghai, China, where he is currently a Full Professor with the School of Computer Science. He has published over 140 papers in international journals or conferences. He has over 3100 ISI Web of Science citations with an H-index of 33 according to the Clarivate. He was one of the most cited Chinese researchers (Elsevier) in 2019. His current research interests include network science, graph data mining, computational social science, spectral graph theory, and random walks.

Dr. Zhang was a recipient of the Excellent Doctoral Dissertation Award of Liaoning Province, China, in 2007, the Excellent Post-Doctor Award of Fudan University in 2008, the Shanghai Natural Science Award (third class) in 2013, and the Wilkes Award for the best paper published in The Computer Journal in 2019. He is a member of the IEEE.

Stacy Patterson is an Associate Professor in the Department of Computer Science at Rensselaer Polytechnic Institute. She received the MS and PhD in computer science from UC Santa Barbara in 2003 and 2009, respectively. From 2009-2011, she was a postdoctoral scholar at the Center for Control, Dynamical Systems and Computation at UC Santa Barbara. From 2011-2013, she was a postdoctoral fellow in the Department of Electrical Engineering at Technion - Israel Institute of Technology. Dr. Patterson is the recipient of a Viterbi postdoctoral fellowship, the IEEE CSS Axelby Outstanding Paper Award, and an NSF CAREER award. Her research interests include distributed systems, machine learning, and the Internet of Things.