Inferring Network Controllability from Coarse Measurements

Anonymous authors
Paper under double-blind review

Abstract

The control of large-scale network dynamical systems is important in diverse settings, including data networks, traffic networks, and biological networks. Often our knowledge of network structure is not detailed at the finest scale of the dynamics, due to our inability to observe and characterize at such scales. Rather, our knowledge may be limited to coarse summary information. Our objective herein is to characterize the average controllability of a fine-scale linear time-invariant (LTI) system based on knowledge of only a coarsened approximation of that system. We provide conditions under which the average controllability can be well approximated by that of the (synthesized, reduced-order) coarse-scale system. Not surprisingly, our results show that for the approximation to be good, some inherent fine-scale parametric structure is required as well as a type of coherence–between the coarse observations and underlying structure. The parametric model we assume for the fine-scale network is the stochastic block model (SBM), developed for community detection. We study two complementary methods: a *classical* structure-agnostic model-order reduction method (C-SaMOR) that treats the coarse graph as a surrogate dynamical system, and a learning-based SaMOR (Lb-SaMOR) that infers the latent community representation before computing controllability. We highlight regimes in which each can outperform the other. Finally, we provide simulations to confirm our theoretical results for different network sizes and densities, characterizing how much community structure is retained in the coarse summary. Our work reveals conditions under which coarse measurements support accurate influence estimation for control and intervention in community-structured networked systems.

1 Introduction

This paper studies how to infer the controllability of a fine-scale networked discrete-time linear time-invariant (LTI) system when only coarse measurements of the underlying network are available. In many real-world settings, system identification must rely on low-resolution observations produced by sensing modalities that cannot probe fine-scale structure. Our primary motivation comes from the human brain, though similar challenges arise in data networks, power grids, and traffic networks. While efforts are underway to produce a map of a canonical human brain, our knowledge of the brain as an interconnected, network system is not yet detailed to the level of the individual neuron Betzel & Bassett (2017). And yet, modern neurotechnology increasingly seeks to intervene in and "control" brain dynamics Gu et al. (2015); Karrer et al. (2020). For example, brain implants designed for epilepsy patients attempt to *steer* the brain away from states that correspond to seizures Heck et al. (2014). These applications underscore a fundamental question: how reliably can one reason about controllability when only coarse measurements of a complex network are available?

We ask how coarse network observations can lead to distorted estimates of which node groups are most influential for controlling the fine-scale system. Characterizing this distortion is critical, e.g., in designing next-generation therapeutic neuromodulation implants to improve the accuracy of control strategies when intervening in the brain. More specifically, we aim to estimate a group average controllability vector for an inaccessible fine-scale networked LTI system from coarse summary observations and parametric knowledge

[0]**Abbreviations:** SBM, stochastic block model; MOR, model order reduction; SaMOR, structure-agnostic MOR; C-SaMOR, classical SaMOR; Lb-SaMOR, learning-based SaMOR; LTI, linear time-invariant

of network structure. When a fine-scale system is available, this goal can be approached by first applying some variant of model-order-reduction (MOR) to obtain a lower-dimensional (coarse or coarse-scale) system model.

Our setting differs from traditional MOR, which starts from a known fine-scale system and actively designs a reduced model. In our setting, the coarse measurements are given passively before the latent fine-scale structure is known. This means that the coarse measurements cannot be tuned to align with the fine-scale structure. We therefore term our problem *structure-agnostic model-order reduction* (SaMOR). We study when the controllability of the coarse-scale network is a reliable proxy (see Section 3) for the underlying fine-scale model. The latent statistical structure we use is the stochastic block model (SBM) Schaub et al. (2020); Xing et al. (2020); Abbe (2017). In an SBM, graph nodes are clustered into communities, and the probability of intra-community connections is different from inter-community connections. The SBM provides a community-structured generative model that makes the fine-scale controllability vector estimable from coarse observations. Based on this assumed structure, it is natural that a certain type of synchronization or coherence between coarse observations and underlying structure is required for adequate approximation.

We consider a suitable metric based on the controllability Gramian. Gramian-based metrics can be used to compare the control effort exerted by groups of nodes Pasqualetti et al. (2014); Lindmark (2020); Zhang et al. (2024); Cui et al. (2020); Li et al. (2024). We use this to address the following question. Say there are only certain groups of nodes that we can control and our knowledge is only of the coarse-scale network. Which of the groups is best positioned to control the fine-scale network? Harkening back to our epilepsy-treatment motivations, we would like to identify where an implanted medical device should perturb to be best able to collapse the unstable brain-state oscillations that can lead to seizures O'Leary et al. (2018).

Specifically, we consider the *average controllability* metric defined by the trace of the controllability Gramian of a discrete-time system, where we assume non-negative system matrices Farina & Rinaldi (2000). We characterize the average controllability of a set of systems, each system corresponding to different possible groups of control nodes in the fine-scale network. We concatenate these quantities into an average controllability vector. We present two schemes, each of which estimates this average controllability vector, and characterize the error between the true and estimated vectors. Our main contributions are as follows.

1. *Group average controllability under coarse observations.* We first formulate the object to be estimated from the coarse data. Each coordinate of the average controllability vector corresponds to a group of fine-scale control nodes represented by one coarse node. This formulation allows us to compare the true fine-scale controllability profile with the profiles produced by the two estimators below, using normalized ℓ_1 error metrics that account for coverage size and the inherent shift in average controllability.

2. *C-SaMOR approximation.* In Section 3, we consider an auxiliary LTI system whose state-space matrices are derived from the coarse data. We refer to this system as *a classical structure-agnostic model order reduction (C-SaMOR)*. Each node in the C-SaMOR corresponds to a group of nodes in the fine-scale network (see Fig. 1). In the C-SaMOR, the grouped nodes are both measured and actuated, so our model implicitly couples measurement and actuation. Our objective is to use the average controllability vector of C-SaMOR to approximate that of the fine scale. We derive a tight upper bound on the *approximation-error* between the average controllability of the C-SaMOR and the underlying fine-scale model. The bound goes to zero as (i) the coarse and fine network sizes grow, (ii) the network becomes dense, (iii) the coverage of the fine network by coarse measurements expands, and (iv) the synchronization between the coarse summary data and the underlying community structure increases. We discuss situations in which these four criteria hold. If synchronization is not sufficiently high, the error may not approach zero even as the network size increases.

3. *Learning-based SaMOR estimation.* In Section 4, we learn the average controllability vector of the fine-scale system directly from the coarse data without characterizing an auxiliary SaMOR system. We develop a novel learning-based SaMOR (Lb-SaMOR) algorithm based on a mixed-membership method Mao et al. (2020) for unsupervised learning of the parameters and the community structure of an SBM. We derive a tight upper bound on the estimation error and characterize its convergence

in terms of network parameters. Using this bound, we show that the estimation error approaches zero as long as the community structure parameter estimates are accurate and the coverage of the fine network by coarse measurements is sufficiently rich. Although the Lb-SaMOR's estimate quality is implicitly dependent on synchronization, it demonstrates greater robustness than the C-SaMOR method in the low-synchronization regime.

4. *Simulation study.* In Section 5, we present simulation results to show that estimation errors and their qualitative behaviors are in line with the results in Sections 3 and 4. Importantly, and consistent with theory, simulations show that synchronization has less impact on the estimation error of the Lb-SaMOR method. The Lb-SaMOR approach outperforms when network density and coarse network size are not too small, and the coarse measurements are not fully synchronized with communities.

1.1 Related work

Controllability with Known Network Structure: Our work studies controllability of networked linear time-invariant (LTI) systems when the fine-scale network is not directly observed. Most work on network controllability assumes that the graph is observed at the same resolution at which control is analyzed. Building on the classical theory of controllability for LTI systems Kalman (1960), a central line of research in network control studies energy metrics derived from the controllability Gramian to quantify how easily a network can be driven by external inputs Lindmark (2020); Pasqualetti et al. (2014); Baggio et al. (2022). Among these, average controllability has been widely used to characterize the influence of individual nodes and driver-node sets. Optimization problems involving these metrics have also been studied, including greedy methods for selecting driver nodes or designing network topologies that minimize control energy Cortesi et al. (2014b); Srighakollapu et al. (2021), and analyses of how groups of actuated nodes steer networks to target states with minimum energy Cortesi et al. (2014a); Gu et al. (2015).

In contrast, we consider settings in which only coarse measurements are available. Each coarse measurement/control group induces one entry of the group average controllability vector; estimating this vector lets us compare which groups are most influential. The catch is that this vector must be inferred from coarsened observations because the fine graph is unobserved. Our results provide error bounds relating the controllability of the fine network to that of its coarse representation. This complements recent work on controllability and observability metrics for random linear systems Preciado & Rahimian (2016); Bopardikar (2021).

Model Reduction with Known Fine-Scale Dynamics: Because C-SaMOR uses the coarse graph as a surrogate reduced-order system, it is natural to compare our setting with model order reduction (MOR). MOR is a well-developed area in dynamical systems Moore (1981); Antoulas (2005); Benner et al. (2015). Extensions of MOR to networked systems include methods that simplify individual subsystem dynamics Cheng & Scherpen (2018); Monshizadeh et al. (2013) and methods that reduce the number of nodes through clustering or projection Jongsma et al. (2018); Benner et al. (2021). The accuracy of the reduced model is typically evaluated using system norms such as the H_2 or H_∞ norms Jongsma et al. (2018); Cheng & Scherpen (2021), which compare transfer functions of the original and reduced systems.

These approaches reduce a known system: the full fine-scale model is available, and the reduction is designed deliberately to preserve system behavior. SaMOR addresses the complementary passive setting, where coarse observations arise without prior knowledge of the underlying fine-scale network. Rather than comparing transfer functions, we infer how group-level average controllability statistics change between the fine and coarse networks, using an ℓ_1-based error metric.

Coarsened SBMs for Graph Inference: Our guarantees require statistical structure that connects coarse observations to the unobserved fine graph. Stochastic block models (SBMs) are widely used in machine learning and network science as tractable models for community structure, with applications spanning social, biological, and infrastructure systems Abbe (2017); Funke & Becker (2019); Chen et al. (2025); Henka et al. (2025). Much of this literature assumes that the full network is observable. In large systems, however, only coarse summaries of connectivity may be available; for example, brain measurements from macro-scale electrodes or imaging modalities aggregate signals rather than reveal full connectivity maps Ghoroghchian et al. (2018; 2020).

Motivated by such settings, prior work has studied inference from coarsened network observations. Early work considered recovering graph structure from aggregated measurements Ahn et al. (2012); Dasarathy et al. (2015), while more recent work developed coarsened variants of the SBM for community detection Ghoroghchian et al. (2021). We adopt this coarsened SBM framework as the latent statistical structure that makes fine-scale controllability estimable from coarse measurements.

Standard notations and proofs of our mathematical results are completely developed in the appendix.

2 Preliminaries and Problem Statement

We introduce a class of stable LTI systems that evolve on networks with randomly generated edge weights. We then describe how we use the SBM to generate random *fine* networks with community structure, and how we model the coarse measurements of these networks. We then formulate our problem statement.

2.1 Dynamical systems and controllability

A network is represented by an undirected graph $\mathcal{G} \triangleq (\mathcal{V}, \mathcal{E})$, with node set $\mathcal{V} \triangleq \{1, \ldots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Each edge $(u, v) \in \mathcal{E}$ has a weight $\mathbf{A}_{uv} = \mathbf{A}_{vu} \in \mathbb{R}$. The *weighted symmetric adjacency matrix* of $\mathcal{G}$ is $\mathbf{A} \triangleq [\mathbf{A}_{uv}]$, where $\mathbf{A}_{uv} = 0$ whenever $(u, v) \notin \mathcal{E}$. A random network is an undirected graph whose edge weights $[\mathbf{A}_{uv}]$ are random variables. For a network $\mathcal{G}$ with adjacency matrix $\mathbf{A}$, associate state $x_i[t] \in \mathbb{R}$ to the i-th node at time t. The states evolve according to LTI dynamics as

$$\mathbf{x}[t+1] = \mathbf{A}_{\text{nom}}\mathbf{x}[t] + \mathbf{B}\mathbf{u}[t] \ , \qquad \forall\ t \in \mathbb{N}, \tag{1}$$

where $\mathbf{x}[t] = [x_1[t], \ldots, x_n[t]]^\mathsf{T}$ denotes the network state at time t and $\mathbf{u}[t] \in \mathbb{R}^n$ is the control input. The normalized adjacency matrix is defined as $\mathbf{A}_{\text{nom}} \triangleq \mathbf{A}/(\epsilon + \rho(\mathbf{A}))$ which ensures that the spectral radius $\rho(\mathbf{A}_{\text{nom}}) < 1$ for any $\epsilon > 0$; see Karrer et al. (2020). This normalization guarantees stability of the system. The input matrix $\mathbf{B} = \text{Diag}(\mathbf{b})$, with $\mathbf{b} \in \{0, 1\}^n$, specifies the nodes that receive external inputs. We denote the corresponding *control node set* by $\mathcal{K} = \text{supp}(\mathbf{b}) \subset [n]$. Thus, the control signals are only injected at nodes in $\mathcal{K}$. For example, if $\mathbf{B} = \text{Diag}(\mathbf{1}_{n_1}, \mathbf{0}_{n-n_1})$, then the control node set $\mathcal{K}$ is $\{1, \ldots, n_1\}$.

The T-step controllability Gramian is defined as

$$\boldsymbol{\mathcal{C}}_T(\mathbf{A}_{\text{nom}}, \mathbf{B}) = \sum_{\ell=0}^{T-1} (\mathbf{A}_{\text{nom}})^\ell \mathbf{B}\mathbf{B}^\top (\mathbf{A}_{\text{nom}})^\ell. \tag{2}$$

The rank of $\boldsymbol{\mathcal{C}}_T(\mathbf{A}_{\text{nom}}, \mathbf{B})$ determines whether the system in (1) is controllable. However, this criterion is binary and provides no quantitative insight into how easily the system can be controlled. To address this, several metrics derived from the controllability Gramian have been proposed. A widely used metric in networked systems is the *average controllability* defined as $\text{tr}[\boldsymbol{\mathcal{C}}_T(\mathbf{A}_{\text{nom}}, \mathbf{B})]$ (see Müller & Weber (1972); Pasqualetti et al. (2014); Gu et al. (2015)). Because $\boldsymbol{\mathcal{C}}_T(\mathbf{A}_{\text{nom}}, \mathbf{B}) \succeq 0$, $\text{tr}\left[\boldsymbol{\mathcal{C}}_T(\mathbf{A}_{\text{nom}}, \mathbf{B})\right] \geq 0$. Since the system matrix is normalized so that $\rho(\mathbf{A}_{\text{nom}}) < 1$, the infinite-horizon controllability Gramian $\boldsymbol{\mathcal{C}} = \lim_{T\to\infty} \boldsymbol{\mathcal{C}}_T$ is well defined. In what follows, we work with this infinite-horizon Gramian and write $\boldsymbol{\mathcal{C}}_T$ or $\boldsymbol{\mathcal{C}}$ without explicitly indicating the dependence on $\mathbf{A}_{\text{nom}}$ and $\mathbf{B}$ when the context is clear.

Average controllability can also be interpreted through the nuclear norm of the controllability Gramian. Since $\boldsymbol{\mathcal{C}}_T \succeq 0$, its nuclear norm coincides with its trace, i.e.,

$$\|\boldsymbol{\mathcal{C}}_T\|_* = \text{trace}(\boldsymbol{\mathcal{C}}_T).$$

The nuclear norm is the sum of the singular values of a matrix and is widely used as a convex surrogate for rank in matrix analysis and optimization Recht et al. (2010); Chi et al. (2017). In the present setting, the nuclear norm quantifies the overall magnitude of the controllability Gramian. Larger values correspond to systems that are more easily influenced by control inputs but are also more sensitive to approximation errors in reduced-order representations. This observation motivates our Lb-SaMOR approach (Section 4), which directly estimates average controllability rather than approximating the full Gramian.

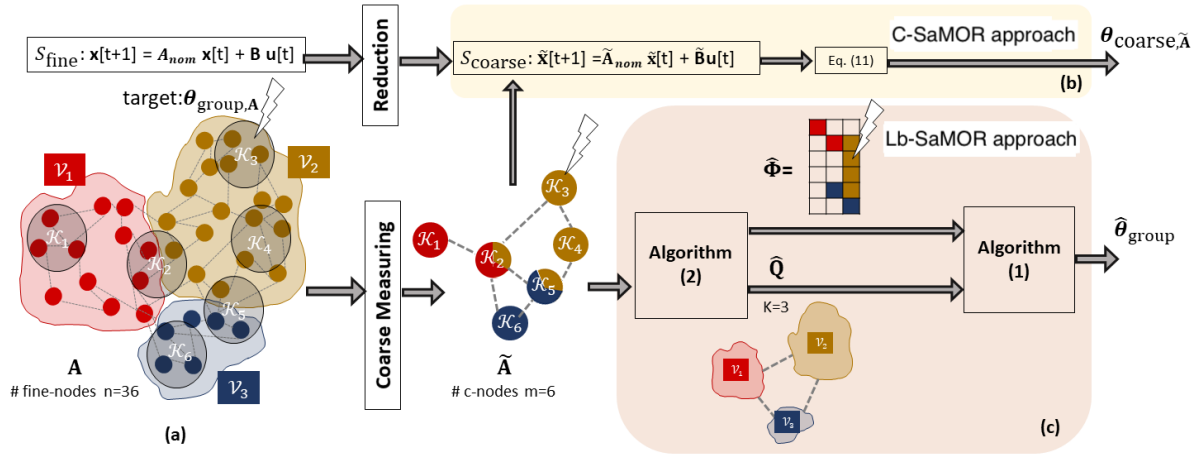


Figure 1: Schematic of C-SaMOR and Lb-SaMOR for estimating $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$. (a) For $\mathcal{G}_{\text{fine}}$ consisting of $n = 36$ nodes, we have $K = 3$ communities ($\mathcal{V}_1$, $\mathcal{V}_2$, $\mathcal{V}_3$) with $m = 6$ coarse c-nodes ($\mathcal{K}_1$,...,$\mathcal{K}_6$), each having coverage size $r = 3$. The control node set is $\mathcal{K}_3$. The lightning bolts show an example of where external input can be exerted. (b) In the C-SaMOR approach, we infer $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ via the reduced-order dynamical system $\mathcal{S}_{\text{coarse}}$. (c) The Lb-SaMOR approach uses the mixed-membership (MM) Algorithm 2 to estimate $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ directly from $\widetilde{\mathbf{A}}$. This avoids the need to consider $\mathcal{S}_{\text{coarse}}$.

Remark 1 (On the interpretation of average controllability)**.** *The inverse of the average controllability is closely related to the control energy required to steer a controllable system from the origin to target states on the unit sphere Müller & Weber (1972); Baggio et al. (2022). Since $\boldsymbol{\mathcal{C}} \succ 0$, it follows that $\text{tr}(\boldsymbol{\mathcal{C}}^{-1}) \geq 1/\text{tr}(\boldsymbol{\mathcal{C}})$. The tightness of this bound is numerically explored in Yan et al. (2012); Karrer et al. (2020). In this work, we focus on the qualitative implication that larger values of $\text{tr}(\boldsymbol{\mathcal{C}})$ correspond to lower levels of average control effort. Thus $\text{tr}(\boldsymbol{\mathcal{C}})$ serves as a tractable proxy for assessing how effectively a given set of control nodes can influence the network in all directions of the state space (see Summers et al. (2015); Ikeda & Kashima (2018); Pirani & Taylor (2020); Srighakollapu et al. (2021); McGowan et al. (2021); Fang et al. (2022); Deng & Gu (2020)).*

2.2 The Stochastic Block Model

The stochastic block model (SBMs) is a probabilistic model that produces random graphs with communities. We use this model to introduce a parametric group structure into the dynamics to allow us to understand the role such community structure plays in the quality of controllability estimation based on coarse measured graphs. Formally, let $\mathcal{G} \triangleq (\mathcal{V}, \mathcal{E})$ be the undirected graph (i.e., the fine graph) with n nodes and random edge weights generated according to the general GSBM$(n, \mathbf{Q}, \mathbf{d})$:

Definition 1. ***(General SBM)*** *Given a natural number K, vector $\mathbf{d} \in [0,1]^K$, and a matrix $\mathbf{Q} \in [0,1]^{K\times K}$, the GSBM$(n, \mathbf{Q}, \mathbf{d})$ is a distribution on a graph $\mathcal{G}$ of size n such that: 1) the $\mathcal{G}$'s node set, denoted by $\mathcal{V}$, $|\mathcal{V}| = n$, is partitioned into K disjoint communities $\mathcal{V} = \cup_{k=1}^{K} \mathcal{V}_k$ where the communities have relative sizes $\mathbf{d} = [\mathbf{d}_1, \dots \mathbf{d}_K]$, $|\mathcal{V}_k| = n\mathbf{d}_k$; and 2) two nodes $u \in \mathcal{V}_k$ and $v \in \mathcal{V}_{k'}$ are joined by an edge with weight $\mathbf{A}_{uv} \in \{0,1\}$ drawn with probability $\mathbf{Q}_{kk'}$ independently from other edges, for all $k, k' \in [K]$.* □

The GSBM$(n, \mathbf{Q}, \mathbf{d})$ generates a random undirected graph with K communities with non-identical intra- and cross-community connection probabilities specified by $\mathbf{Q} \in [0,1]^{K\times K}$. An operational interpretation of the General SBM is provided by the *community membership matrix* $\mathbf{P} \in \mathbb{R}^{K\times n}$:

$$\mathbf{P}_{kv} = \begin{cases} 1 & \text{if} \quad v \in \mathcal{V}_k \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

For any $k, k' \in [K]$, it follows that

$$\mathbf{A}_{uv} \sim \text{Bernoulli}(\mathbf{Q}_{k,k'}), \quad \text{if} \quad \mathbf{P}_{ku} = \mathbf{P}_{k'v} = 1. \tag{4}$$

It is immediate that $\mathbf{P}\mathbf{P}^\mathsf{T} = \text{Diag}\left(|\mathcal{V}_1|, \ldots, |\mathcal{V}_K|\right)$, where $|\mathcal{V}_k|$ is the size of the k-th community. Let $\mathbf{D} \triangleq \frac{1}{n}\mathbf{P}\mathbf{P}^\mathsf{T} = \text{Diag}\left(\mathbf{d}\right)$ be the diagonal matrix of *relative* community sizes.

To facilitate analysis and harness the power of SBM-generated networks, we work with high-dimensional networks. We make some assumptions on the scaling of the parameters of the underlying SBM as $n \to \infty$.

Assumption 1. ***(SBM scaling** Abbe (2017)) There exists a sequence $\rho_n \in (0, 1)$ and a constant $K \times K$ matrix $\mathbf{Q}_\circ \geq 0$ such that*

$$\mathbf{Q} = \rho_n \mathbf{Q}_\circ, \tag{5}$$

where $0 \leq [\mathbf{Q}_\circ]_{kk'} \leq 1$, and $\rho_n \sqrt{n} \to \infty$ as $n \to \infty$.

The parameter ρ_n in (5) regulates the density (or sparsity) of connections in and across communities. For several real-world networks, ρ_n decreases with n Abbe (2017).

Remark 2. *The relative community sizes of the GSBM in Def. 1 are set to $\mathbf{d}$. Another common model is to choose community memberships in an i.i.d. manner from some distribution $\mathbf{p}$. In this model, relative community sizes become random with $\mathbb{E}[\mathbf{d}] = \mathbf{p}$. The analysis of this paper will hold for such a model except that when evoking the expectation of the fine matrix $\mathbf{A}$, an outer expectation should be added to account for the randomness in community sizes.*

2.3 Fine Graph and Coarse Measurements

In this part, we describe the modeling choice and notations made for the coarse measuring process. We assume that the fine graph $\mathcal{G}_{\text{fine}}$ with adjacency matrix $\mathbf{A}$ is drawn from the SBM in Definition 1. Moreover, (1) is the *fine-scale* system dynamics, and we refer to it as $\mathcal{S}_{\text{fine}}$, i.e.

$$\mathcal{S}_{\text{fine}}: \ \mathbf{x}[t+1] = \mathbf{A}_{\text{nom}}\mathbf{x}[t] + \mathbf{B}\mathbf{u}[t] \ , \qquad \forall \ t \in \mathbb{N}. \tag{6}$$

We define the coarse-scale summary of $\mathbf{A}$ as

$$\widetilde{\mathbf{A}} \triangleq \mathbf{W}\mathbf{A}\mathbf{W}^\mathsf{T} \quad \in \mathbb{R}^{m \times m}, \tag{7}$$

where $\mathbf{W} \in \mathbb{R}^{m \times n}$ is the *coarse-measurement matrix* whose rows indicate which parts of the graph are measured. In this paper, we assume (a) the i-th row of the coarse-measurement matrix $\mathbf{W}$ has r_i equal non-zero terms which sum to 1. These non-zero indices are the fine nodes that are measured as a group; and (b) $\mathbf{W}\mathbf{W}^\mathsf{T} = \text{Diag}([1/r_1, \cdots, 1/r_m])$, meaning the sets of fine nodes corresponding to each coarse measurement are disjoint. The coarsened matrix $\widetilde{\mathbf{A}}$ is the starting point of our analysis.

The matrix $\widetilde{\mathbf{A}}$ can be interpreted as the adjacency matrix of an undirected *coarse* graph $\mathcal{G}_{\text{coarse}}$ with m *fictitious* nodes—referred to as the (coarse or) *c-nodes*. This interpretation is helpful as we discuss the network dynamics of $\widetilde{\mathbf{A}}$ in Section 3. The regime where the *total coverage size* $\sum_{i \in [m]} r_i \ll n$ indicates that c-nodes cover the fine graph only sparsely. In other words, there may exist (several) fine nodes that do not contribute to $\widetilde{\mathbf{A}}$ (see Fig. 1). For other coarsening models in the literature see Ghoroghchian et al. (2021); Dasarathy et al. (2015).

2.4 Problem Statement

Our objective is to quantify the average controllability of $\mathcal{G}_{\text{fine}}$ using only $\mathcal{G}_{\text{coarse}}$ (see (7)), and the knowledge of $\mathcal{G}_{\text{fine}}$'s community structure in Definition 1. We do not have access to the way the coarse graph is acquired at the time of decision-making; that is, the coarse-measurement matrix $\mathbf{W}$ is unknown, meaning we do not know which group of fine nodes are mapped to each c-node.

Let $\mathbf{A}_{\text{nom}} = \mathbf{A}/(\epsilon + \rho(\mathbf{A}))$ in (6) be such that $\mathbf{A}$ is drawn from General SBM given by (4). The input matrix $\mathbf{B}_\mathcal{K} \triangleq \mathbf{B}$ is diagonal by definition. Let $\mathcal{K}_i = \text{supp}(\mathbf{w}_i)$, for $i \in [m]$, be the set (hereafter, "group") of nodes

coarsened by the i-th row $\mathbf{w}_i$ of $\mathbf{W}$ in (7). As a result, the control node groups $\{\mathcal{K}_1, \ldots, \mathcal{K}_m\}$ match the measured groups.

Consider the $\mathcal{K}_i$-th group with $\mathbf{B}_{\mathcal{K}_i} = \text{Diag}(\mathbf{w}_i^\mathsf{T})$ and $\mathbf{w}_i$ denoting the coarse measuring vector, where $i \in [m]$. Define the average controllabity metric $\boldsymbol{\theta}^{(i)}_{\text{group},\mathbf{A}} \triangleq \text{tr}[\boldsymbol{\mathcal{C}}(\mathbf{A}_{\text{nom}}, \mathbf{B}_{\mathcal{K}_i})]$ and the m-dimensional group average controllability vector

$$\boldsymbol{\theta}_{\text{group},\mathbf{A}} \triangleq \left[\boldsymbol{\theta}^{(1)}_{\text{group},\mathbf{A}}, \ldots, \boldsymbol{\theta}^{(m)}_{\text{group},\mathbf{A}}\right]^\mathsf{T}. \tag{8}$$

Note that, assuming the fine-scale system is controllable in the classic sense Chen (1999) (i.e., the initial state $\mathbf{x}[0]$ can be steered to any target state $\mathbf{x} \in \mathbb{R}^n$ for some $\mathbf{u}[0], \ldots, \mathbf{u}[T-1]$), $\boldsymbol{\theta}^{(i)}_{\text{group},\mathbf{A}} \geq 0$, due to the positive semi-definiteness of the Gramian $\boldsymbol{\mathcal{C}}(\cdot)$. This vector summarizes the average controllability metrics for all control nodes sets $\{\mathcal{K}_1, \ldots, \mathcal{K}_m\}$. Thus, $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ helps us deduce several quantitative properties that are key to being able to control (6) via $\mathbf{B}_{\mathcal{K}_i}$. For example, $\boldsymbol{\theta}_{\text{group}}$ tells us which $\mathcal{K}_i s$ can drive the network to the desired target state with the least control effort.

Using only the knowledge of $\widetilde{\mathbf{A}}$ in (7), our goal is to estimate the vector $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ in (8). We will develop two approaches. The first, *Structure-agnostic Model Order Reduction (C-SaMOR)*, is based on the traditional *model order reduction* (MOR) approach. Here we rely upon a passively reduced-order auxiliary system $\mathcal{S}_{\text{coarse}}$ in (10), which is governed by $\widetilde{\mathbf{A}}$, to infer $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$. Our second approach is learning-based (Lb-SaMOR). We directly estimate $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ using a clustering-based mixed-membership community-learning algorithm (see Section 4). Here, we bypass the intermediate step of dealing with a reduced-order system. Fig. 1 illustrates the proposed approaches. We use the ℓ_1-norm difference between the true and the estimated controllability vectors to compare the performance of these competing methods (see (12)). Our results in Sections 3 and 4 are probabilistic because $\mathbf{A}$ is a random matrix.

We introduce the expected quantities $\bar{\mathbf{A}} \triangleq \mathbb{E}[\mathbf{A}]$ and $\bar{\widetilde{\mathbf{A}}} \triangleq \mathbb{E}[\widetilde{\mathbf{A}}]$. We later use these to place some mild regulations on $\mathcal{G}_{\text{coarse}}$. From (4) and (7), the expected quantities evaluate to

$$\bar{\mathbf{A}} = \mathbf{P}^\mathsf{T}\mathbf{Q}\mathbf{P} \quad \text{and} \quad \bar{\widetilde{\mathbf{A}}} = \boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^\mathsf{T}, \tag{9}$$

where $\boldsymbol{\Phi} \triangleq \mathbf{W}\mathbf{P}^\mathsf{T} \in \mathbb{R}^{m \times K}$ is the *coarse community membership matrix*, and $\boldsymbol{\Phi}_{ik}$ captures how much the i-th c-node overlaps with the k-th community. To make the analysis less cumbersome, in what follows, we let the coarsening matrix have fixed row support (r). The example below illustrates the definition of the matrices $\mathbf{P}$, $\mathbf{D}$, $\boldsymbol{\Phi}$, and $\mathbf{W}$.

Example 1. *For the network shown in Fig. 2-(a), the fine community membership matrix is*

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \in \{0,1\}^{3\times 8}.$$

The relative fine-scale community sizes are $\mathbf{d} = [\frac{3}{8}, \frac{1}{2}, \frac{1}{8}]$ *and* $\text{Diag}(\mathbf{d}) \in \mathbb{R}^{3\times 3}$. *The coarse-measurement matrix for* $r = 2$ *and the coarse community membership matrix (which tells us what fractions of each coarse node's group of fine nodes belong to each underlying community) are given by*

$$\mathbf{W} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \in \mathbb{R}^{4\times 8}, \quad \textit{and} \quad \boldsymbol{\Phi} = \mathbf{W}\mathbf{P}^\mathsf{T} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \in \mathbb{R}^{4\times 3}.$$

Finally, the rows of $\boldsymbol{\Phi}^\mathsf{T}$ *determine that each community contains what portion of the coarse nodes' groups of fine nodes. Hence,*

$$\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m} = \begin{bmatrix} \frac{5}{16} & \frac{1}{16} & 0 \\ \frac{1}{16} & \frac{6}{16} & \frac{1}{16} \\ 0 & \frac{1}{16} & \frac{1}{16} \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

is a measure of the correlation of (and cross-correlation between) the (coarse membership) representations of communities, i.e., the rows of $\mathbf{\Phi}^\mathsf{T}$. *Each c-node in* $\{1,3\}$ *covers one community; each c-nodes in* $\{2,4\}$ *overlap with two. Matrix* $\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}$ *plays a role in the error bound of Theorem 1 and is discussed in Section 3.* □

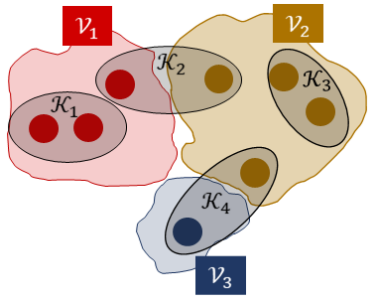


Figure 2: A simple case of a community-structured network $\mathcal{G}_{\text{fine}}$ and its coarse measurements corresponding to Example 1. This graph contains $n = 8$ nodes, $K = 3$ communities ($\mathcal{V}_1$, $\mathcal{V}_2$, $\mathcal{V}_3$) and $m = 4$ coarse c-nodes ($\mathcal{K}_1,\ldots,\mathcal{K}_4$), each having coverage size $r = 2$. Note that c-nodes $1, 3$ are each perfectly synchronized with one community, while c-nodes $2, 4$ each overlap with two communities.

Remark 3. ***(Synchronization of community and coarsening)****: The coarsening operation in Section 2.3 is oblivious to the community structure of* $\mathcal{G}_{\text{fine}}$. *Thus, the* i*-th c-node may contain information about multiple communities if* $|\text{supp}(\mathbf{\Phi}_i)| > 1$ *(the subscript denotes the* i*-th row). Perfect synchronization is the situation in which* $|\text{supp}(\mathbf{\Phi}_i)| = 1$. *This occurs when each c-node covers nodes from only one community.*

3 Approximating Group Average controllability: C-SaMOR Approach

In this section, we define C-SaMOR (classical Structure-agnostic Model Order Reduction) as an auxiliary networked LTI system governed by $\mathcal{G}_{\text{coarse}}$. We then analyze how well the average controllability vector of this reduced system approximates $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ in (8). Formally, consider the coarse system $\mathcal{S}_{\text{coarse}}$,

$$\mathcal{S}_{\text{coarse}}: \ \widetilde{\mathbf{x}}[t+1] = \widetilde{\mathbf{A}}_{\text{nom}}\widetilde{\mathbf{x}}[t] + \widetilde{\mathbf{B}}\widetilde{\mathbf{u}}[t] \ , \tag{10}$$

where $\widetilde{\mathbf{A}}_{\text{nom}} = \widetilde{\mathbf{A}}/(\epsilon + \rho(\widetilde{\mathbf{A}}))$, and $\widetilde{\mathbf{A}}$ is given by (7). Note that $\tilde{\mathbf{x}}[t] \in \mathbb{R}^m$ is not a compressed version of the state $\mathbf{x}[t]$ in $\mathcal{S}_{\text{fine}}$. Rather, $\tilde{\mathbf{x}}[t]$ is a fictitious state controlled by the dynamics of the (scaled) matrix $\widetilde{\mathbf{A}}$. This fictitious state is controlled by the input $\widetilde{\mathbf{B}}\widetilde{\mathbf{u}}[t] \in \mathbb{R}^m$. We refer to this coarse-scale system modeling from the fine-scale system as C-SaMOR because the dimension of the coarse system can be much less than that of the fine system, i.e. $m \ll n$.

Since each group of nodes that receive control input maps to one c-node, $\widetilde{\mathbf{B}}_i = \mathbf{e}_i$ is the coarsened input matrix. Let $\boldsymbol{\theta}^{(i)}_{\text{coarse},\widetilde{\mathbf{A}}} \triangleq \text{tr}[\boldsymbol{\mathcal{C}}(\widetilde{\mathbf{A}}_{\text{nom}}, \widetilde{\mathbf{B}}_i)] \geq 1$ be the average controllability for the i-th node in $\mathcal{G}_{\text{coarse}}$, which is, by definition, lower-bound by 1 (from the definition of trace and the Gramian matrix for positive matrices similar to (2)). Define the coarse average controllability vector:

$$\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}} \triangleq \left[\boldsymbol{\theta}^{(1)}_{\text{coarse},\widetilde{\mathbf{A}}}, \ldots, \boldsymbol{\theta}^{(m)}_{\text{coarse},\widetilde{\mathbf{A}}}\right]^\mathsf{T} \in \mathbb{R}^m. \tag{11}$$

We show in Theorem 1 that the $\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}$ well approximates $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ under some conditions. To quantify the approximation, we define the ℓ_1-error metric:

$$\tilde{\Delta}(\mathbf{A}, \widetilde{\mathbf{A}}) \triangleq \left\| \frac{r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m\|_1} - \frac{\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}} - \mathbf{1}_m}{\|\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}} - \mathbf{1}_m\|_1} \right\|_1 . \tag{12}$$

We prefer $\tilde{\Delta}(\mathbf{A}, \widetilde{\mathbf{A}})$ to the usual metric $\|\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}\|_1$ for the following reasons. First, the factor of r in the first term of $\tilde{\Delta}(\mathbf{A}, \widetilde{\mathbf{A}})$ in (12) accounts for the additional scaling by $1/r$ in the fine system input matrix $\mathbf{B}_{\mathcal{K}_i} = \text{Diag}(\mathbf{w}_i^\mathsf{T})$ (see Lemma 3 in the appendix). Second, the shift by -1 discounts the inherent $+1$ shift in the average controllability definition (see (29) in the appendix). Third, the denominators in (12) ensure that the error is scale-invariant. In other words, the error is unaffected if all elements of either of the controllability vectors are multiplied by a common scaling. Moreover, we chose $\|.\|_1$ over other vector

norms (like $\|.\|_{\max}$) as it aggregates the errors when choosing a singular or any combinations of c-nodes to stimulate sequentially.

The matrix below appears in Theorems 1 and contributes to upper bounding the estimation errors

$$\boldsymbol{\Upsilon} \triangleq \mathbf{D}^{-\frac{1}{2}} \left(\left[\mathbf{I} - (\iota \mathbf{D}^{\frac{1}{2}} \mathbf{Q}_{\circ} \mathbf{D}^{\frac{1}{2}})^2 \right]^{-1} - \mathbf{I} \right) \mathbf{D}^{-\frac{1}{2}} \in \mathbb{R}^{K \times K}, \tag{13}$$

where $\iota = 1/(\frac{\epsilon}{\rho_n n} + \rho(\mathbf{Q}_{\circ}\mathbf{D}))$ is the normalization factor $1/(\epsilon + \rho(\bar{\mathbf{A}}))$ in (6) after substituting $\bar{\mathbf{A}}$ with $\mathbf{A}$, $\rho(\cdot)$ is the spectral radius, and ρ_n and $\mathbf{Q}_{\circ}$ are defined in Assumption 1. We set the normalized matrix $\mathbf{Z} = \iota \mathbf{D}^{\frac{1}{2}} \mathbf{Q}_{\circ} \mathbf{D}^{\frac{1}{2}}$. Hence, the term $\left[\mathbf{I} - (\iota \mathbf{D}^{\frac{1}{2}} \mathbf{Q}_{\circ} \mathbf{D}^{\frac{1}{2}})^2 \right]^{-1}$ in (13) becomes the resolvent of $\mathbf{Z}$, i.e. $(z\mathbf{I} - \mathbf{Z}^2)^{-1}$, evaluated at $z = 1$. Also, note that matrix $\mathbf{D}^{\frac{1}{2}} \mathbf{Q}_{\circ} \mathbf{D}^{\frac{1}{2}}$ is the intra- and cross-community probability matrix that is re-weighted according to the relative community sizes. Recall that the resolvent captures the direct and indirect effects (via the sum of the powers) of $\mathbf{D}^{\frac{1}{2}} \mathbf{Q}_{\circ} \mathbf{D}^{\frac{1}{2}}$. As a result, matrix $\boldsymbol{\Upsilon}$ in (13) encapsulates the size-weighted community-connection probability matrix's direct and indirect impacts. For more information on the significance and the properties of resolvents see Dunford & Schwartz (1988). Later, in Lemma 2, we see how $[\boldsymbol{\Upsilon}]_{kk}$ moves in tandem (i.e. increase and decrease together) with both $\boldsymbol{\theta}^{(i)}_{\text{group},\bar{\mathbf{A}}}$ and $\boldsymbol{\theta}^{(i)}_{\text{coarse},\widetilde{\bar{\mathbf{A}}}}$ for i-th c-nodes that are perfectly synchronized with community k. The aforementioned definitions coupled with Lemma 2 in Section 4, contribute to the proof of the next Theorem.

Theorem 1. *(ℓ_1 **bound on** $\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}$): Let $\tilde{\Delta}(\mathbf{A}, \widetilde{\mathbf{A}})$ be defined as in (12) for $\mathbf{A} \in \{0,1\}^{n \times n}$ and $\widetilde{\mathbf{A}} \in \mathbb{R}^{m \times m}$, and $0 < \delta < 1$ be a constant. Then, under Assumption 1, there exists $n_0, m_0 \in \mathbb{N}$ such that for $n > n_0, m > m_0$, we have*

$$\tilde{\Delta}(\mathbf{A}, \widetilde{\mathbf{A}}) = \mathcal{O}\left(\frac{\eta + \tilde{\eta}}{\rho_n} + \sqrt{\frac{n}{mr}} \sqrt{\frac{\log(\frac{1}{\delta})}{2}} + \frac{\|\mathbf{L}_{\boldsymbol{\Phi},\boldsymbol{\Upsilon}}\|_1}{m} + \left\| \mathbf{D} - \frac{\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Phi}}{m} \right\|_{\max} \right)$$

with probability at least $1 - 3\delta$. Here $\eta = \frac{\sqrt{\log(\frac{n^2}{\delta})/(2n)}}{\|\mathbf{Q}\mathbf{D}\mathbf{Q}\|_{\min}}$; $\tilde{\eta} = \frac{\sqrt{\log(\frac{m^2}{\delta})/(2m)}}{\|\mathbf{Q}\frac{\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Phi}}{m}\mathbf{Q}\|_{\min}}$; $\mathbf{L}_{\boldsymbol{\Phi},\boldsymbol{\Upsilon}} = \boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon}) - \text{diag}(\boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^{\mathsf{T}})$; $\mathbf{D} = (1/n)\mathbf{P}\mathbf{P}^{\mathsf{T}}$; $\boldsymbol{\Upsilon}$ is defined in (13); $\boldsymbol{\Phi}$ is defined in (9) and $\mathbf{Q}$ in Def. 1.

Theorem 1 indicates that $\tilde{\Delta}(\mathbf{A}, \widetilde{\mathbf{A}})$ approaches zero as: (i) the total coverage proportion, i.e. mr/n, increases; (ii) η and $\tilde{\eta}$ –which guarantee that the bound on $\tilde{\Delta}(\mathbf{A}, \widetilde{\mathbf{A}})$ exists with large probability–become negligible; and (iii) $\frac{1}{m}\|\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon}) - \text{diag}(\boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^{\mathsf{T}})\|_1$ and $\|\mathbf{D} - \frac{\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Phi}}{m}\|_{\max}$ decrease. The latter two terms have to do with *synchronization (or overlap)*, i.e., the extent to which coarse measurements are synchronized with (overlap) communities (see Remark 3). The first term $\frac{1}{m}\|\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon}) - \text{diag}(\boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^{\mathsf{T}})\|_1$ is equal to zero if perfect synchronization holds. The relationship between the second term, $\|\mathbf{D} - \frac{\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Phi}}{m}\|_{\max}$, and community overlap is less straightforward. To understand the role of this term, note that matrix $\frac{\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Phi}}{m}$ (cf., Example 1) quantifies a measure of the correlation between (coarse membership) representations of communities. This matrix reveals two natural tradeoffs in the coarsening process. The first is *overlap*: the off-diagonal elements, $[\frac{\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Phi}}{m}]_{kk'} = \sum_{i \in [m]} \boldsymbol{\Phi}_{ik}\boldsymbol{\Phi}_{ik'}/m$ for $k \neq k'$, show the extent of synchronization between coarse measurements and communities (c.f., Remark 3 and Example 1). The second is *balanced-ness*: the diagonal elements, $[\frac{\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Phi}}{m}]_{kk} = \sum_{i \in [m]} \boldsymbol{\Phi}_{ik}^2/m$ for all $k \in [K]$, reflect the relative community sizes *after* coarse measuring. Separately considering diagonal and off-diagonal elements, helps demystify the role of the error term $\|\mathbf{D} - \frac{\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Phi}}{m}\|_{\max}$ in Theorem 1:

$$\left| [\mathbf{D} - \frac{\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Phi}}{m}]_{k,k'} \right| = \begin{cases} \sum_{i \in [m]} \boldsymbol{\Phi}_{ik}\boldsymbol{\Phi}_{ik'}/m & \text{if } k \neq k' \\ |\mathbf{d}_k - \sum_{i \in [m]} \boldsymbol{\Phi}_{ik}^2/m| & \text{if } k = k'. \end{cases} \tag{14}$$

From (14) we infer that $\|\mathbf{D} - \frac{\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Phi}}{m}\|_{\max}$ approaches zero when c-nodes overlap with fewer communities, *and* c-nodes are more evenly spread over communities. The latter means that larger communities get more coarse measurements, and vice versa. Finally, we note that not all degrees of coarseness are acceptable in

Theorem 1. Hence, for each set of parameters in a problem, we can sweep over different values of m_0 to find the acceptable ranges.

Proof sketch of Theorem 1: Using the triangle inequality on matrix norms, we upper bound the left-hand side by a sum of three norms, corresponding to the ℓ_1 relative controllability error at (i) the fine scale (between $\mathbf{A}$ and $\bar{\mathbf{A}}$, see Theorem 2) (ii) at the coarse scale (between $\widetilde{\mathbf{A}}$ and $\bar{\widetilde{\mathbf{A}}}$)), and (iii) average matrices at coarser versus finer scales (between $\bar{\mathbf{A}}$ and $\bar{\widetilde{\mathbf{A}}}$). We then find upper bounds separately on each term using Lemma 1. We find closed-form values for the controllability values corresponding to $\bar{\mathbf{A}}$ and $\bar{\widetilde{\mathbf{A}}}$, as well as probabilistic gaps between the stochastic matrices and their expected values (i.e., an upper bound on the differences that hold with a certain probability). The complete proof is in Appendix Section (B.3). □

In the following, we elaborate on two key components that contribute significantly to the proof of Theorem 1. Define the error $\Delta(\mathbf{A}, \bar{\mathbf{A}})$ similar to (12):

$$\Delta(\mathbf{A}, \bar{\mathbf{A}}) \triangleq \left\| \frac{r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m\|_1} - \frac{r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}} - \mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}} - \mathbf{1}_m\|_1} \right\|_1. \tag{15}$$

The following theorem bounds the difference between the group controllability of the true and expected $\mathbf{A}$ in (15). As hinted in Section 4, this difference, when small, will allow the introduction of the candidate estimator in (21). The estimator is only a function of $\boldsymbol{\Phi}, \mathbf{Q}$, and $\mathbf{D}$, which are retrievable from the coarse-scale matrix $\widetilde{\mathbf{A}}$.

Theorem 2. *(ℓ_1 **error bound between** $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ and $\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}}$): Let $\Delta(\mathbf{A}, \bar{\mathbf{A}})$ be as in (15) for $\mathbf{A} \in \{0,1\}^{n\times n}$ and $\widetilde{\mathbf{A}} \in \mathbb{R}^{m\times m}$, and $0 < \delta_\circ, \delta < 1$ be two constants. Then, under Assumption 1, there exists $n_0 \in \mathbb{N}$ such that for $n > n_0$, we have*

$$\Delta(\mathbf{A}, \bar{\mathbf{A}}) \leq \frac{2\left(\frac{\epsilon}{\rho_n n} + \rho(\mathbf{Q}_\circ \mathbf{D})\right)^2}{\|\mathbf{Q}_\circ \mathbf{D} \mathbf{Q}_\circ\|_{\min}} \left[1 + \sqrt{\frac{n}{mr}} \sqrt{\frac{\log\left(\frac{1}{\delta_\circ}\right)}{2}}\right] \alpha_n = \mathcal{O}\left(\frac{\eta}{\rho_n} + \sqrt{\frac{n}{mr}} \sqrt{\frac{\log\left(\frac{1}{\delta_\circ}\right)}{2}}\right), \tag{16}$$

holds with probability at least $(1-\delta_\circ)(1-\delta)$, where η is defined in (19); n and m are the dimensions of $\mathbf{A}$ and $\widetilde{\mathbf{A}}$; $\mathbf{Q}$ is defined in Def. 1; and $\mathbf{D} = (1/n)\mathbf{P}\mathbf{P}^\mathsf{T}$.

Theorem 2 bounds the relative controllability at the fine scale. The final step in the proof of Theorem 1 is provided by Lemma 1, which relates the relative controllability between true and averaged systems, both in the coarse and fine scales.

Lemma 1. *(**Error between the Gramians of random and expected LTI systems**): Let $\boldsymbol{\theta}_{\text{fine},\mathbf{A}}, \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}$, $\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}$, and $\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}$ denote the quantities defined in 11. Under Assumption 1, the following holds with probability at least $1-\delta$:*

$$\alpha_n \triangleq \left\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\right\|_1 = \mathcal{O}(\tfrac{\eta}{\rho_n}), \tag{17}$$

and with probability at least $1-\tilde{\delta}$:

$$\tilde{\alpha}_m \triangleq \left\|\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}} - \boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}\right\|_1 = \mathcal{O}(\tfrac{\tilde{\eta}}{\rho_n}), \tag{18}$$

where

$$\eta = \frac{\sqrt{\log(\frac{n^2}{\delta})/(2n)}}{\|\mathbf{Q}\mathbf{D}\mathbf{Q}\|_{\min}}, \tilde{\eta} = \frac{\sqrt{\log(\frac{m^2}{\delta})/(2m)}}{\|\mathbf{Q}\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m}\mathbf{Q}\|_{\min}} \tag{19}$$

At a high level, Lemma 1 is a concentration result for the Gramians of $\mathcal{S}_{\text{fine}}$ (or $\mathcal{S}_{\text{coarse}}$) and $\overline{\mathcal{S}}_{\text{fine}}$ (or $\overline{\mathcal{S}}_{\text{coarse}}$). Note that $\boldsymbol{\mathcal{C}}(\mathbf{Z}, \mathbf{I}) = (\mathbf{I} - \mathbf{Z}^2)^{-1}$, where $\mathbf{Z}$ can take $\mathbf{A}_{\text{nom}}$, $\bar{\mathbf{A}}_{\text{nom}}$, $\widetilde{\mathbf{A}}_{\text{nom}}$, or $\bar{\widetilde{\mathbf{A}}}_{\text{nom}}$. Lemma 1 is effectively bounding the difference of resolvents $(z\mathbf{I} - \mathbf{Z}^2)^{-1}$ evaluated at $z = 1$. However, the rate at which the difference

goes to zero is different for the fine- and coarse systems. For $\mathbf{A} = \widetilde{\mathbf{A}}$, we have $n = m$ and $\frac{\mathbf{\Phi}^\top \mathbf{\Phi}}{m} = \mathbf{D}$. Hence, α_n and $\widetilde{\alpha}_n$ coincide.

In closing, we highlight that the C-SaMOR estimate $\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}$ will well approximate $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ when (i) the graph density ρ_n and number of fine- and c-nodes are sufficiently large, when (ii) coarse measurements are sufficiently synchronized with the communities, and when (iii) coarse measurements are distributed across communities in a balanced way. This way, Theorem 1 helps assess the reliability of our control strategy from the network properties of the coarsely approximated system, properties that we can evaluate. The required conditions will be (roughly) satisfied in practice when communities and coarse measurements are spatially (geographically) localized Sporns & Betzel (2016); Galhotra et al. (2018) and aligned. In many setting, nearby nodes tend to be connected and form communities (e.g., social networks of friends and neighbours, regions with distinct functional and structural features in the brain). Furthermore, coarse measurements are often also local (e.g., parcellation of brain imaging data, electrophysiological recording of the brain, aggregation of geographically adjacent nodes). Hence in many settling local communicies and local measures align spatially and so will (approximately) satisfy the needed conditions. Finally, we note that often coarse measurements are uniformly distributed across a network. In such a setting larger communities are typically measured more frequently.

4 Approximating Group Average Controllability: Lb-SaMOR Approach

We present Lb-SaMOR, a learning-based approach to estimate $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$. Unlike the C-SaMOR approach that relies on $\mathcal{S}_{\text{coarse}}$, we directly estimate elements in $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ using a mixed-membership (MM) community-learning algorithm. Specifically we use the mixed-membership algorithm Mao et al. (2020), which is numerically efficient and has a strong theoretical base.

The lemma below is instrumental in the proof of both of our primary results, Theorems 1 and 3. This lemma provides motivation for our candidate estimator in (21). In essence, this lemma characterizes the group-average controllability vector associated with the expected matrix $\bar{\mathbf{A}}$ in (9).

Lemma 2. ***(Group Controllability of Expected Dynamics):*** *Define $\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}}$ by replacing $\mathbf{A}$ in (8) with the expected matrix $\bar{\mathbf{A}}$ given by (9). Let $\mathbf{\Phi}$ be as in (9), $\mathbf{P}$ be in (3), and r be the coverage size of $\mathbf{W}$ in (7). Under Assumption 1, we have*

$$\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}} = \frac{1}{r}\left(\mathbf{1}_m + \frac{1}{n}\mathbf{\Phi}\text{diag}(\mathbf{\Upsilon})\right), \tag{20}$$

where n and m are the dimensions of $\mathbf{A}$ and $\widetilde{\mathbf{A}}$; moreover $\mathbf{\Upsilon}$ is defined in (13).

Lemma 2 gives a formula to compute the group average controllability vector associated with the expected matrix $\bar{\mathbf{A}}$. Recall that $\mathbf{\Phi}_{i,k} = |\{v : v \in \mathcal{V}_k \cap \text{supp}(\mathbf{w}_i)\}|/|\text{supp}(\mathbf{w}_i)|$ ($k \in [K]$, $i \in [m]$) is the fraction of the i-th c-node's overlap with community k. From Lemma 2, it follows that $\boldsymbol{\theta}^{(i)}_{\text{group},\bar{\mathbf{A}}} \propto \sum_{k\in[K]} \mathbf{\Phi}_{i,k}\mathbf{\Upsilon}_{kk}$. In view of this observation, we infer that the c-nodes that have the most overlap with communities of largest $\mathbf{\Upsilon}_{kk}$ are the most controllable. Moreover, Theorem 2 (see the Appendix) shows that $\Delta(\mathbf{A}, \bar{\mathbf{A}})$ can be arbitrarily small with high probability. With Theorem 2 and Lemma 2, we propose our candidate estimator

$$\hat{\boldsymbol{\theta}}_{\text{group}} \triangleq \mathbf{1}_m + \hat{\mathbf{\Phi}}\text{diag}(\hat{\mathbf{\Upsilon}}), \tag{21}$$

where

$$\hat{\mathbf{\Upsilon}} \triangleq \hat{\mathbf{D}}^{-\frac{1}{2}}\left(\left[\mathbf{I} - (\hat{\imath}\hat{\mathbf{D}}^{\frac{1}{2}}\hat{\mathbf{Q}}_\circ\hat{\mathbf{D}}^{\frac{1}{2}})^2\right]^{-1} - \mathbf{I}\right)\hat{\mathbf{D}}^{-\frac{1}{2}}, \text{ and } \hat{\imath} = \epsilon/n + \rho(\hat{\mathbf{Q}}\hat{\mathbf{D}}). \tag{22}$$

The estimator in (21) is similar to the true group controllability vector in (20), but there are two important differences. First, the true terms $\mathbf{\Phi}, \mathbf{Q}$, and $\mathbf{D}$ are replaced with their estimated ("hatted") counterparts. Second, the constant coefficients $1/r$ and $1/n$ in (20) are omitted since they affect neither the error metric (see (23)—wherein $1/r$ is compensated and $1/n$ cancels out in both the numerator and denominator)—nor

the order of nodes when assessed based on their controllability value. The hatted quantities $\hat{\mathbf{\Phi}}$ and $\hat{\mathbf{Q}}$ are obtained from Algorithm 2, which takes as input $\widetilde{\mathbf{A}}$ and the number of communities K[1]. We obtain $\hat{\mathbf{D}}$ from Algorithm 1. Importantly, $\hat{\boldsymbol{\theta}}_{\text{group}}$ in (21) is obtained from coarsened matrix $\widetilde{\mathbf{A}}$ and not from the fine scale matrix $\mathbf{A}$. The following assumption is required by Mao et al. (2020) for the recovery of $\hat{\mathbf{\Phi}}, \hat{\mathbf{Q}}$ using their proposed Algorithm 2.

Assumption 2. ***(Existence of a perfectly-synchronized c-node):*** *For all $k \in [K]$, there exist at least one coarse node $i \in [m]$ such that $\mathbf{\Phi}_{ik} = 1$.*

The assumption is that, $\mathcal{G}_{\text{coarse}}$ has at least one perfectly-synchronized (or equivalently "pure") c-node per community $k \in [K]$. In real-world networks communities, coarse measurements are often spatially localized (see Section 3 for a more detailed explanation). Hence, we expect it is reasonable to assume that there is at least one coarse measurement per community that does not overlap with multiple communities. The next theorem bounds the error of our proposed Lb-SaMOR estimation in (21) using the following metric:

$$\widehat{\Delta}(\widetilde{\mathbf{A}}) = \left\| \frac{\hat{\boldsymbol{\theta}}_{\text{group}} - \mathbf{1}_m}{\|\hat{\boldsymbol{\theta}}_{\text{group}} - \mathbf{1}_m\|_1} - \frac{r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m\|_1} \right\|_1 . \tag{23}$$

Theorem 3. *(ℓ_1 **error bound on $\hat{\boldsymbol{\theta}}_{\text{group}} - \boldsymbol{\theta}_{\text{group},\mathbf{A}}$**): Let $\widehat{\Delta}(\widetilde{\mathbf{A}})$ be defined as in (23) for $\mathbf{A} \in \{0,1\}^{n\times n}$ and $\widetilde{\mathbf{A}} \in \mathbb{R}^{m\times m}$, and $0 < \delta < 1$ be a constant. Then, under Assumptions 1,2, there exists $n_0 \in \mathbb{N}$ such that for $n > n_0$, we have*

$$\widehat{\Delta}(\widetilde{\mathbf{A}}) = \mathcal{O}\left(\frac{\eta}{\rho_n} + \sqrt{\frac{n}{mr}}\sqrt{\frac{\log(\frac{1}{\delta})}{2}} + \frac{\|\mathbf{L}_{\mathbf{\Phi}}\|_{1,1}}{m} + \|\mathbf{L}_{\mathbf{Q}}\|_{\max} + \|\mathbf{L}_{\mathbf{D}}\|_{\max} \right)$$

holds with probability at least $1-2\delta$, where $\mathbf{L}_{\mathbf{\Phi}} = \hat{\mathbf{\Phi}} - \mathbf{\Phi}$, $\mathbf{L}_{\mathbf{Q}} = \hat{\mathbf{Q}} - \mathbf{Q}$, and $\mathbf{L}_{\mathbf{D}} = \hat{\mathbf{D}} - \mathbf{D}$; $\eta = \frac{\sqrt{\log(\frac{n^2}{\delta})/(2n)}}{\|\mathbf{QDQ}\|_{\min}}$; n and m are the dimensions of $\mathbf{A}$ and $\widetilde{\mathbf{A}}$; $\mathbf{\Phi}$ is defined in (9) and $\mathbf{Q}$ in Def. 1; and $\mathbf{D} = (1/n)\mathbf{P}\mathbf{P}^{\mathsf{T}}$.

Theorem 3 suggests that for sufficiently large n (smaller η which yields high probabilistic guarantees for the bound on $\widehat{\Delta}(\widetilde{\mathbf{A}})$) and large total coverage proportion $\frac{mr}{n}$, the estimate $\hat{\boldsymbol{\theta}}_{\text{group}}$ approximates $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ to arbitrary precision under the following conditions. First, the graph is dense enough (sufficiently large ρ_n). Second, the coarse community membership matrix is well estimated (smaller $\|\mathbf{L}_{\mathbf{\Phi}}\|_{1,1}/m$). Third, cross-community probability estimation does not suffer from high error (smaller $\|\mathbf{L}_{\mathbf{Q}}\|_{\max}$), and the relative community sizes estimated from the coarse graph are close to the ones in the fine graph (smaller $\|\mathbf{L}_{\mathbf{D}}\|_{\max}$). The authors of Mao et al. (2020) show that $\|\mathbf{L}_{\mathbf{\Phi}}\|_{1,1}/m$ and $\|\mathbf{L}_{\mathbf{Q}}\|_{\max}$ in Theorem 3 approach zero under some conditions as $m \to \infty$. While it is unclear whether these theoretical conditions hold in our setting (given our coarsening process), our simulations show that $\widehat{\Delta}(\widetilde{\mathbf{A}})$ decreases with an increase in m. The study and characterization of the behavior of $\|\mathbf{L}_{\mathbf{\Phi}}\|_{1,1}/m$ and $\|\mathbf{L}_{\mathbf{Q}}\|_{\max}$ with respect to graph scaling (i.e. ρ_n) is left for future work.

Proof sketch of Theorem 3: By an application of the triangle inequality for matrix norms, we upper bound $\widehat{\Delta}(\widetilde{\mathbf{A}})$ in (23) as

$$\widehat{\Delta}(\widetilde{\mathbf{A}}) \leq \underbrace{\left\| \frac{r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}} - \mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}} - \mathbf{1}_m\|_1} - \frac{r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m\|_1} \right\|_1}_{=\Delta(\mathbf{A},\bar{\mathbf{A}})} + \underbrace{\left\| \frac{\hat{\boldsymbol{\theta}}_{\text{group}} - \mathbf{1}_m}{\|\hat{\boldsymbol{\theta}}_{\text{group}} - \mathbf{1}_m\|_1} - \frac{r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}} - \mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}} - \mathbf{1}_m\|_1} \right\|_1}_{=\widehat{\Delta}(\bar{\mathbf{A}},\bar{\mathbf{A}})} . \tag{24}$$

The first term on the right-hand side of the inequality in (24) has already been bounded in (16). We then bound the second term in (24) by substituting $\hat{\boldsymbol{\theta}}_{\text{group}}$ with the output of Algorithm 1, and $\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}}$ with

[1] Algorithm 2, a mixed-membership algorithm, is adapted from Mao et al. (2020). It is a type of spectral clustering method that first performs eigen decomposition of $\widetilde{\mathbf{A}}$ to find the overlapping membership ($\mathbf{\Phi}_{ik}$) of the fine nodes. A pruning step is included (see steps 4 and 5 in Algorithm 2) that improves algorithm performance.

(20), and $\boldsymbol{\Upsilon}$ defined in (13) to get:

$$\widehat{\Delta}(\widetilde{\mathbf{A}}, \bar{\mathbf{A}}) = \left\| \frac{\hat{\boldsymbol{\Phi}}\text{diag}(\hat{\boldsymbol{\Upsilon}})}{\|\hat{\boldsymbol{\Phi}}\text{diag}(\hat{\boldsymbol{\Upsilon}})\|_1} - \frac{\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})}{\|\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})\|_1} \right\|_1 . \tag{25}$$

We successively apply the triangle inequality to achieve an upper bound on (25) as a function of the norm errors between $\boldsymbol{\Phi}, \mathbf{Q}, \mathbf{D}$ and their estimates. □

The result of Theorem 3 is important because one can directly infer the most influential control node groups, $\mathcal{K}_i$ (i.e., the one with the highest average controllability) via the most influential i-th c-nodes, and vice versa. In contrast to the C-SaMOR error characterization in Theorem 1, the Lb-SaMOR's error does not suffer explicitly from non-synchronization of coarse measurements and communities, although the error is implicitly influenced by the estimation quality of $\boldsymbol{\Phi}, \mathbf{Q}, \mathbf{D}$. Thus, the Lb-SaMOR estimation of group controllability is more robust than that of the C-SaMOR when community overlap is present. This hypothesis is validated via simulation in Section 5. Finally, the Lb-SaMOR approach improves the accuracy of control designs of many real-world scenarios wherein coarse measurements are taken blindly with respect to the underlying community organization.

Algorithm 1 Direct Inference of the Group Average Controllability

Require: estimates $\hat{\boldsymbol{\Phi}}$ and $\hat{\mathbf{Q}}$ from Algorithm 2, and the number of communities K
compute $\hat{\mathbf{D}} = \text{Diag}(\frac{\hat{\boldsymbol{\Phi}}^\mathsf{T}\mathbf{1}_m}{\mathbf{1}_K\hat{\boldsymbol{\Phi}}^\mathsf{T}\mathbf{1}_m})$ and $\hat{\imath} = \frac{\epsilon}{n} + \rho(\hat{\mathbf{Q}}\hat{\mathbf{D}})$
return $\hat{\boldsymbol{\theta}}_{\text{group}} = \mathbf{1}_m + \hat{\boldsymbol{\Phi}}\text{diag}\left(\hat{\mathbf{D}}^{-\frac{1}{2}}\left(\left[\mathbf{I} - (\hat{\imath}\hat{\mathbf{D}}^{\frac{1}{2}}\hat{\mathbf{Q}}_\circ\hat{\mathbf{D}}^{\frac{1}{2}})^2\right]^{-1} - \mathbf{I}\right)\hat{\mathbf{D}}^{-\frac{1}{2}})\right)$

Algorithm 2 Mixed-Membership Community Estimation Algorithm Mao et al. (2020)

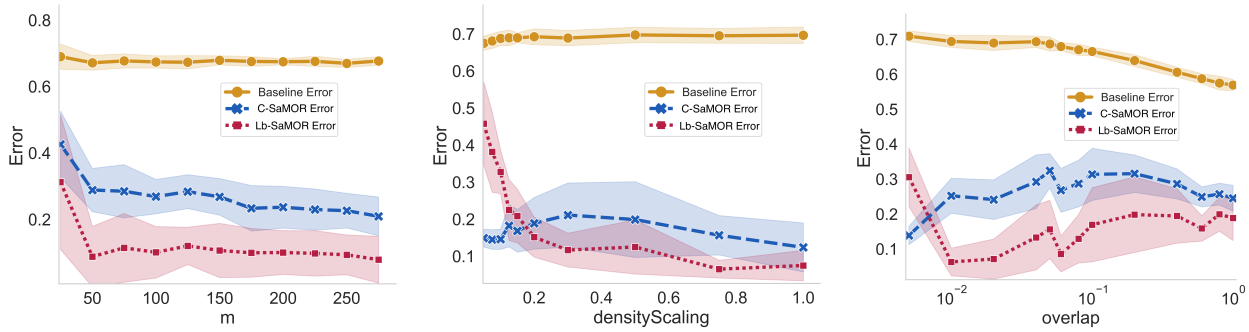
Require: Coarse adjacency matrix $\widetilde{\mathbf{A}}$, number of communities K
compute the highest K spectral-decomposition of $\widetilde{\mathbf{A}}$ as $\hat{V}\hat{\Lambda}\hat{V}^\mathsf{T}$ and set $\mathcal{S}_{\text{pruned}} = \text{Prune}(\hat{V})$
set $X = \hat{V}([m]\backslash\mathcal{S}_{\text{pruned}}, :)$ and compute $\mathcal{S}_{\text{pure}} = \text{Successive-Projection-Algorithm}(X^\mathsf{T})$
set $X_{\text{pure}} = X(\mathcal{S}_{\text{pure}}, :)$ and compute un-normalized $\hat{\boldsymbol{\Phi}}^{\text{un-nom}} = \hat{V}X_{\text{pure}}^{-1}$
$\hat{\boldsymbol{\Phi}}_{ik}^{\text{un-nom}} \leftarrow 0$ if $\hat{\boldsymbol{\Phi}}_{ik}^{\text{un-nom}} < e^{-12}, \forall i \in [m], k \in [K]$
return $\hat{\boldsymbol{\Phi}} = \text{Diag}^{-1}(\hat{\boldsymbol{\Phi}}^{\text{un-nom}}\mathbf{1}_K)\hat{\boldsymbol{\Phi}}^{\text{un-nom}}$ and $\hat{\mathbf{Q}} = X_{\text{pure}}\hat{\Lambda}X_{\text{pure}}^\mathsf{T}$

5 Simulations

We validate our theoretical results[2] in Theorems 1 and 3, by plotting the errors $\tilde{\Delta}(\mathbf{A}, \widetilde{\mathbf{A}})$ and $\widehat{\Delta}(\widetilde{\mathbf{A}})$ and show that these errors are comparable to the bounds we obtained in Theorems 1 and 3 We generate $\mathbf{A} \sim \text{GSBM}(n, \mathbf{Q}, \mathbf{d})$ and then determine $\widetilde{\mathbf{A}} = \mathbf{WAW}^\mathsf{T}$. We obtain the support of each row of $\mathbf{W}$ independently, by the three-step process:

1. We sample an i.i.d vector $\varrho \in [K]^m$ from the distribution specified the probability vector $[\mathbf{d}_k]_{k=1}^K$. The i-th element $\varrho_i \in [K]$ represents the community with which the i-th c-node has the most overlap.
2. We generate a set of random vectors π from the distribution $\text{Dir}(\beta, \cdots, 1, \cdots, \beta) \in [0, 1]^K$, where 1 is in the ϱ_i-th position and Dir(.) is the Dirichlet multivariate probability distribution parameterized by $0 < \beta < 1$ Lin (2016). Specifically, $\text{Dir}(\beta, \cdots, 1, \cdots, \beta)$ outputs a vector of size K whose ϱ_i-th component is the largest, whereas the strength of the other components is proportionate to β, and the sum of the vector's elements equals 1.

[2]The results in this paper are reproducible. The theoretical findings are annotated and step-by-step elaborated in the appendix. The data generation process and the parameter values used for numerical simulations are fully explained. In addition, the Python code from which the simulation figures are generated is available upon request. The code is also on GitHub and the repository will be made public upon submission acceptance.



(a) Error versus number of coarse nodes (m).

(b) Error versus graph density scaling parameter (ρ_n).

(c) Error versus overlap extent parameter (β).

Figure 3: Error estimate based on C-SaMOR and Lb-SaMOR approaches, where C-SaMOR Error $= \tilde{\Delta}(\mathbf{A}, \widetilde{\mathbf{A}})$ and Lb-SaMOR Error $= \widehat{\Delta}(\widetilde{\mathbf{A}})$. The shaded region represents one standard deviation computed over 20 independent realizations.

3. For each community k where $\pi_k > 0$, we randomly chose $\lfloor r \cdot \pi_k \rfloor$ fine nodes from $\mathcal{V}_k^{\text{non-sel}}$ (i.e., non-selected fine indices in community k) and set these to be the support of $\mathbf{w}_i$. The chosen indices are removed from $\mathcal{V}_k^{\text{non-sel}}$.

This process continues until the support of all $\mathbf{w}_i$s are selected. The first step ensures that the community balancedness criterion illustrated in Section 3 is maintained, i.e., the coverage of each community after coarse-measuring remains (in expectation) proportionate to the fine-scale community sizes. The second step controls the level of community overlap (or synchronization); the smaller β is chosen, the more coarse nodes get synced with communities, and vice versa (c.f., Remark 3).

We set the number of fine nodes $n = 5000$, the overlap parameter $\beta = 0.05$, and the number of communities $K = 4$. For $\mathbf{Q} = \rho_n \mathbf{Q}_\circ \in \mathbb{R}^{K \times K}$, we set $[\mathbf{Q}_\circ]_{kk} = p = 0.5$, $[\mathbf{Q}_\circ]_{kk'} = q = 0.1$ (for $k \neq k'$). If not specified, the number of c-nodes $m = 100$, the density scaling parameter $\rho_n = 0.1$, and the coverage size per c-node $r = 10$.

Fig. 3a-3c illustrate the qualitative behavior of the errors with respect to changes in the degree of (non-) synchronization in coarse nodes (i.e., β), m, and ρ_n. To compare these errors fairly, we also consider a *baseline* random vector $\boldsymbol{\mu} \in [1, 2]^m$ whose elements are generated (drawn) independently from the uniform distribution on the interval $[1, 2]$, independently of $\widetilde{\mathbf{A}}$. We then define the *baseline error* as $\left\| \frac{\boldsymbol{\mu} - \mathbf{1}_m}{\|\boldsymbol{\mu} - \mathbf{1}_m\|_1} - \frac{r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m\|_1} \right\|_1$. We make the following observations.

First, both the Lb-SaMOR and C-SaMOR errors are consistently better than the random baseline. That said, the Lb-SaMOR approach has consistently smaller errors than the C-SaMOR approach for large parametric regimes. Fig. 3a shows that all errors monotonically decrease as m increases. Fig. 3b shows that errors decrease as ρ_n increases. This is expected because larger ρ_n results in more distant intra- and cross-community edge densities. This makes both community representation extraction and controllability estimation easier.

Finally, Fig. 3c demonstrates the higher tolerance of the Lb-SaMOR approach to situations wherein coarse measurements are less synchronized (larger β), in comparison to the C-SaMOR method. From Fig. 3c, we observe that in the perfectly synchronized regime (the left of the x-axis), C-SaMOR performs very well. As the overlap increases up to the point around $\beta \simeq 0.1$, measurements become less synchronized with communities, and the estimation errors of both schemes increase. After this point, the overlap becomes so high that the controllability values become closer and closer to one another, which results in a slight decrease in the C-SaMOR estimation error. These results are consistent with our bounds in Theorems 1 and 3.

6 Conclusions

In this paper, we introduced two frameworks that exploit the power of community-based representations to infer the average controllability of fine graphs from coarse summary data. One approach is learning-based (Lb-SaMOR), the other is the classical model order reduction (C-SaMOR) approach. We compare the per-

formance of the two. For both methods, we derived high probability error bounds on the difference between the controllability estimate and ground truth and validated the theory with numerical simulations. Our results highlight the impact of fine- and coarse-network sizes, graph density, and measurement synchronization with communities in the estimation accuracy. Interestingly, for the Lb-SaMOR approach, we show that the estimation error decreases with network size despite the synchronization bias, which is not the case with the C-SaMOR approach. In terms of future work, we plan to use our theory to study the role of coarsening, community structures, and synchronization on the controllability of brain networks. Extending the proposed tools and techniques to controllability metrics other than average controllability is also left for future work.

Broader Impact Statement

This work is primarily methodological. It develops tools for estimating quantities related to controllability in networked systems when only coarse measurements are available. One potential area of positive impact is neuroscience, where brain networks are commonly modeled using structural or functional connectivity data at resolutions coarser than the underlying neural dynamics. Improved estimation of controllability metrics may help researchers assess when coarse network measurements preserve information relevant to stimulation, intervention, or disease-related changes. Similar benefits may extend to other domains that rely on large-scale network models, such as systems biology, infrastructure networks, and multi-agent systems.

However, controllability metrics should not be interpreted as direct prescriptions for intervention. Our results rely on modeling assumptions, including stochastic block structure and specific coarsening mechanisms, and provide statistical guarantees rather than deterministic conclusions about any particular real system. Misuse could arise if such metrics are treated as sufficient evidence for clinical, operational, or policy decisions without domain-specific validation. Responsible use requires domain expertise, careful model checking, validation against the system of interest, and expert interpretation when translating controllability estimates into practice.

References

Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Graph sketches: Sparsification, spanners, and subgraphs. pp. 5–14. Proc. of the 31st ACM SIGMOD-SIGACT-SIGAI symposium on Principles of Database Syst., 2012.

Athanasios C. Antoulas. *Approximation of Large-Scale Dynamical Systems*. Society for Industrial and Applied Mathematics, 2005.

Giacomo Baggio, Fabio Pasqualetti, and Sandro Zampieri. Energy-aware controllability of complex networks. *Annual Review of Control, Robotics, and Autonomous Systems*, 5(1), 2022.

Rémi Bardenet and Odalric-Ambrym Maillard. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 2015.

Peter Benner, Serkan Gugercin, and Karen Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.

Peter Benner, Sara Grundel, and Petar Mlinaric. *Clustering-Based Model Order Reduction for Nonlinear Network Systems*, pp. 75–96. Springer Int. Publishing, Cham, 2021. doi: 10.1007/978-3-030-72983-7_4. URL https://doi.org/10.1007/978-3-030-72983-7_4.

Richard F Betzel and Danielle S Bassett. Multi-scale brain networks. *Neuroimage*, 160:73–83, 2017.

Shaunak D Bopardikar. A randomized approach to sensor placement with observability assurance. *Automatica*, 123:109340, 2021.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

Chi-Tsong Chen. Linear system theory and design. 1999.

Yuzhou Chen, Xiao Guo, and Shujie Ma. Stochastic block model-aware topological neural networks for graph link prediction. *Transactions on Machine Learning Research*, 2025.

Xiaodong Cheng and Jacquelien M.A. Scherpen. Robust synchronization preserving model reduction of Lur'e networks. In *European Control Conference*, pp. 2254–2259. European Control Conference, 2018. doi: 10.23919/ECC.2018.8550058.

Xiaodong Cheng and JMA Scherpen. Model reduction methods for complex network systems. *Annual Review of Control, Robotics, and Autonomous Syst.*, 4:425–453, 2021.

Chong-Yung Chi, Wei-Chiang Li, and Chia-Hsiang Lin. *Convex optimization for signal processing and communications: From fundamentals to applications*. CRC press, 2017.

Fabrizio L. Cortesi, Tyler H. Summers, and John Lygeros. Submodularity of energy related controllability metrics. In 53rd *IEEE Conference on Decision and Control*, pp. 2883–2888. 53rd IEEE Conference on Decision and Control (CDC), 2014a. doi: 10.1109/CDC.2014.7039832.

Fabrizio L Cortesi, Tyler H Summers, and John Lygeros. Submodularity of energy related controllability metrics. In 53rd *IEEE Conference on Decision and Control*, pp. 2883–288. 53rd IEEE Conference on Decision and Control (CDC), 2014b.

Yulong Cui, Shibo He, Mincheng Wu, Chengwei Zhou, and Jiming Chen. Improving the controllability of complex networks by temporal segmentation. *IEEE Trans. on Network Sci. and Eng.*, 7(4):2765–2774, 2020.

Gautam Dasarathy, Parikshit Shah, Badri Narayan Bhaskar, and Robert D Nowak. Sketching sparse matrices, covariances, and graphs via tensor products. *IEEE Trans. on Inf. Theory*, 61(3):1373–1388, 2015.

Shikuang Deng and Shi Gu. Controllability analysis of functional brain networks. *arXiv preprint arXiv:2003.08278*, 2020.

Nelson Dunford and Jacob T Schwartz. *Linear operators, part 1: General theory*, volume 10. John Wiley & Sons, 1988.

Feng Fang, Beata Godlewska, Raymond Y Cho, Sean I Savitz, Sudhakar Selvaraj, and Yingchun Zhang. Effects of escitalopram therapy on functional brain controllability in major depressive disorder. *Journal of Affective Disorders*, 2022.

Lorenzo Farina and Sergio Rinaldi. *Positive linear systems: Theory and applications*, volume 50. John Wiley & Sons, 2000.

Thorben Funke and Till Becker. Stochastic block models: A comparison of variants and inference methods. *PloS one*, 14(4):e0215296, 2019.

Sainyam Galhotra, Arya Mazumdar, Soumyabrata Pal, and Barna Saha. The geometric block model. In *Proc. of the AAAI Conference on Artificial Intelligence*, volume 32. Proc. of the AAAI Conference on Artificial Intelligence, 2018.

Nafiseh Ghoroghchian, Stark C Draper, and Roman Genov. A hierarchical graph signal processing approach to inference from spatiotemporal signals. In *2018 29th Biennial Symp. on Commun.*, pp. 1–5. IEEE 29th Biennial Symp. on Commun., 2018.

Nafiseh Ghoroghchian, David M Groppe, Roman Genov, Taufik A Valiante, and Stark C Draper. Node-centric graph learning from data for brain state identification. *IEEE Trans. on Signal and Inf. Process. over Networks*, 6:120–132, 2020.

Nafiseh Ghoroghchian, Gautam Dasarathy, and Stark Draper. Graph community detection from coarse measurements: Recovery conditions for the coarsened weighted stochastic block model. In *Int. Conference on Artificial Intelligence and Statistics*, pp. 3619–3627. PMLR, 2021.

Shi Gu, Fabio Pasqualetti, Matthew Cieslak, Qawi K Telesford, B Yu Alfred, Ari E Kahn, John D Medaglia, Jean M Vettel, Michael B Miller, Scott T Grafton, et al. Controllability of structural brain networks. *Nature Communications*, 6(1):1–10, 2015.

Christianne N Heck, David King-Stephens, Andrew D Massey, Dileep R Nair, Barbara C Jobst, Gregory L Barkley, Vicenta Salanova, Andrew J Cole, Michael C Smith, Ryder P Gwinn, et al. Two-year seizure reduction in adults with medically intractable partial onset epilepsy treated with responsive neurostimulation: Final results of the rns system pivotal trial. *Epilepsia*, 55(3):432–441, 2014.

Noureddine Henka, Mohamad Assaad, and Sami Tazi. Mixture degree-corrected stochastic block model for multi-group community detection in multiplex graphs. *Transactions on Machine Learning Research*, 2025.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

Takuya Ikeda and Kenji Kashima. Sparsity-constrained controllability maximization with application to time-varying control node selection. *IEEE Control Syst. Lett.*, 2(3):321–326, 2018.

H.-J. Jongsma, P. Mlinarić, S. Grundel, P. Benner, and H. L. Trentelman. Model reduction of linear multi-agent systems by clustering with H_2 and H_∞ error bounds. *Mathematics of Control, Signals, and Syst.*, (6), 2018.

R.E. Kalman. On the general theory of control systems. *IFAC Proceedings Volumes*, 1(1):491–502, 1960. ISSN 1474-6670. 1st Int. IFAC Congress on Automatic and Remote Control, Moscow, USSR, 1960.

Teresa M Karrer, Jason Z Kim, Jennifer Stiso, Ari E Kahn, Fabio Pasqualetti, Ute Habel, and Danielle S Bassett. A practical guide to methodological considerations in the controllability of structural brain networks. *Journal of Neural Eng.*, 17(2):026031, 2020.

Fritz Keinert. Course notes: Applied linear algebra, Fall 2019. https://orion.math.iastate.edu/keinert/math507/notes/chapter5.pdf.

Renjie Li, Chengxia Sun, Miao Dong, Meijuan Wang, Qing Gao, and Xian Liu. The controllability analysis of brain networks during rhythmic propagation. *IEEE Trans. on Network Sci. and Eng.*, 2024.

Jiayu Lin. On the Dirichlet distribution. *Department of Mathematics and Statistics, Queens University*, 2016.

G. Lindmark. *Controllability of Complex Networks at Minimum Cost*. PhD thesis, Linköping University Electronic Press, 2020. (visited on 04/24/2022).

Xueyu Mao, Purnamrita Sarkar, and Deepayan Chakrabarti. Estimating mixed memberships with sharp eigenvector deviations. *Journal of the American Statistical Association*, (just-accepted):1–24, 2020.

Amanda L McGowan, Linden Parkes, Xiasong He, Ovidia Stanoi, Yoona Kang, Silicia Lomax, Mia Jovanova, Peter J Mucha, Kevin N Ochsner, Emily B Falk, et al. Controllability of structural brain networks and the waxing and waning of negative affect in daily life. *Biological Psychiatry Global Open Science*, 2021.

Nima Monshizadeh, Harry L. Trentelman, and M. Kanat Camlibel. Stability and synchronization preserving model reduction of multi-agent systems. *Syst. & Control Lett.*, 62(1):1–10, 2013. ISSN 0167-6911.

B. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. on Automatic Control*, 26(1):17–32, 1981.

PC Müller and HI Weber. Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems. *Automatica*, 8(3):237–246, 1972.

Gerard O'Leary, David M Groppe, Taufik A Valiante, Naveen Verma, and Roman Genov. NURIP: Neural interface processor for brain-state classification and programmable-waveform neurostimulation. *IEEE Journal of Solid-State Circuits*, 53(11):3150–3162, 2018.

Fabio Pasqualetti, Sandro Zampieri, and Francesco Bullo. Controllability metrics, limitations and algorithms for complex networks. *IEEE Trans. on Control of Network Syst.*, 1(1):40–52, 2014.

Mohammad Pirani and Joshua A Taylor. Controllability of AC power networks with dc lines. *IEEE Trans. on Power Syst.*, 36(2):1649–1651, 2020.

Victor M Preciado and M Amin Rahimian. Controllability gramian spectra of random networks. In *2016 American Control Conference*, pp. 3874–3879. IEEE American Control Conference, 2016.

Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

Michael T Schaub, Santiago Segarra, and John N Tsitsiklis. Blind identification of stochastic block models from dynamical observations. *SIAM Journal on Mathematics of Data Science*, 2(2):335–367, 2020.

Olaf Sporns and Richard F Betzel. Modular brain networks. *Annual review of psychology*, 67:613–640, 2016.

Manikya Valli Srighakollapu, Rachel Kalpana Kalaimani, and Ramkrishna Pasumarthy. Optimizing network topology for average controllability. *Syst. & Control Lett.*, 158:105061, 2021.

Tyler H Summers, Fabrizio L Cortesi, and John Lygeros. On submodularity and controllability in complex dynamical networks. *IEEE Trans. on Control of Network Syst.*, 3(1):91–101, 2015.

Eugene Evgenevich Tyrtyshnikov. *A brief introduction to numerical analysis*. Springer Science & Business Media, 1997.

Yu Xing, Xingkang He, Haitao Fang, and Karl H Johansson. Community detection for gossip dynamics with stubborn agents. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pp. 4915–4920. IEEE, 2020.

Gang Yan, Jie Ren, Ying-Cheng Lai, Choy-Heng Lai, and Baowen Li. Controlling complex networks: How much energy is needed? *Physical Review Lett.*, 108(21):218703, 2012.

Yu Zhang, Jie Ding, and Xiang Li. Network controllability robustness learning via spatial graph neural networks. *IEEE Trans. on Network Sci. and Eng.*, 2024.

A Appendix

This section contains proofs for the main results in Sections 3 and 4.

Notation: We denote vectors and matrices using bold faced small and upper case letters, e.g., $\mathbf{z}$ and $\mathbf{Z}$. For $\mathbf{Z} \in \mathbb{R}^{n\times n}$ wehere $\mathbf{Z}_{uv}$ is the etnry in the uth row and vth column, define $\|\mathbf{Z}\|_\infty = \max_{1\leq u\leq n} \sum_{v=1}^n |\mathbf{Z}_{uv}|$; $\|\mathbf{Z}\|_{\max} = \max_{u,v}|\mathbf{Z}_{uv}|$; $\|\mathbf{Z}\|_{1,1} = \sum_{u=1}^n \sum_{v=1}^n |\mathbf{Z}_{uv}|$; and $\|\mathbf{Z}\|_2 = \sqrt{\lambda_{\max}(\mathbf{Z}^\mathsf{T}\mathbf{Z})}$, where $\lambda_{\max}$ is the maximum eigenvalue. Denote the spectral radius of $\mathbf{Z}$ by $\rho(\mathbf{Z}) = \max_i\{|\lambda_i(\mathbf{Z})|\}$, where λ_i is the i-th eigenvalue; $\mathrm{diag}(\mathbf{Z}) = [\mathbf{Z}_{11}, \ldots, \mathbf{Z}_{nn}]^\mathsf{T} \in \mathbb{R}^n$; and $\mathrm{Diag}(\mathbf{Z})$ sets the off-diagonal entries of $\mathbf{Z}$ to zero. The matrix inequality $\mathbf{Z}_1 \leq \mathbf{Z}_2$ implies element-wise inequality. Denote the length-n all-ones and all-zeros vectors by $\mathbf{1}_n$ and $\mathbf{0}_n$. For $\mathbf{z} \in \mathbb{R}^n$, define $\mathrm{supp}(\mathbf{z})$ as the set of indices i for which $\mathbf{z}_i \neq 0$ and $\|\mathbf{z}\|_1 = \sum_{u=1}^n |\mathbf{z}_u|$. For a positive integer n, denote $[n] \triangleq \{1, \ldots, n\}$. The cardinality of a set $\mathcal{V}$ is denoted by $|\mathcal{V}|$. Vector $\mathbf{e}_i$ is the "standard" basis vector containing all 0s except for its i-th element which is 1. The Hadamard (element-wise) product is denoted by $\odot$. We write $f(n) = \mathcal{O}(h(n))$ iff there exist positive reals c_0 and n_0 such that $|f(n)| \leq c_0 h(n)$ for all $n \geq n_0$. Conversely, $f(n) = \Omega(h(n))$ iff there exist $c_0, n_0 > 0$ such that $|f(n)| \geq c_0 h(n)$ for all $n \geq n_0$. $[a, b]$ shows the interval (set of real numbers) between scalar values a and b.

We begin with a few auxiliary results. In what follows, $\mathbf{1}_{K,K}$ is $K \times K$ dimensional all-one matrix.

Useful Matrix Norm Equivalence Horn & Johnson (2012); Keinert (Fall 2019): For two real, square, and symmetric matrices $\mathbf{Z}, \mathbf{Y} \in \mathbb{R}^{n\times n}$, the following hold:

1. If $\mathbf{Z}, \mathbf{Y}$ are invertible: $\mathbf{Z}^{-1} - \mathbf{Y}^{-1} = \mathbf{Z}^{-1}(\mathbf{Y} - \mathbf{Z})\mathbf{Y}^{-1}$.
2. $\mathrm{tr}(\mathbf{ZY}) \leq \|\mathbf{Z}\|_2 \mathrm{tr}(\mathbf{Y})$.
3. For all norms: if $\|\mathbf{Z}\| < 1 \Rightarrow \|(\mathbf{I} - \mathbf{Z})^{-1}\| < \frac{1}{1-\|\mathbf{Z}\|}$.
4. $\|\mathbf{ZY}\|_{\max} \leq \|\mathbf{Z}\|_\infty \|\mathbf{Y}\|_{\max}$.
5. $\|\mathbf{Z}\|_\infty \leq n\|\mathbf{Z}\|_{\max}$.
6. $\|\mathbf{Z}\|_2 \leq n\|\mathbf{Z}\|_{\max}$.
7. $\|\mathbf{ZY}\|_{\max} \leq n\|\mathbf{Z}\|_{\max}\|\mathbf{Y}\|_{\max}$.

Note that for $\mathbf{Z} = \mathbf{Z}^\mathsf{T}$, the matrix 2-norm and the spectral radius coincide ($\rho(\mathbf{Z}) = \|\mathbf{Z}\|_2$) and are used interchangeably.

Proposition 4. ***(Hoeffding's Inequality) Boucheron et al. (2013)****: Let $\{\mathbf{z}_\ell\}_{\ell=1}^{|\mathbf{z}|}$ be bounded independent random variables. Then the following holds with probability at least $1 - \delta$, for some $\delta > 0$:*

$$\begin{aligned}\left|\sum_{\ell=1}^{|\mathbf{z}|} \mathbf{z}_\ell - \mathbb{E}\left[\sum_{\ell=1}^{|\mathbf{z}|} \mathbf{z}_\ell\right]\right| &\leq \sqrt{\frac{|\mathbf{z}|}{2}\log\left(\frac{1}{\delta}\right)}\,(\max(\mathbf{z}) - \min(\mathbf{z})) \\ &\leq \sqrt{\frac{|\mathbf{z}|}{2}\log\left(\frac{1}{\delta}\right)}\frac{\max(\mathbf{z})-\min(\mathbf{z})}{\mathbb{E}\left[\sum_{\ell=1}^{|\mathbf{z}|} \mathbf{z}_\ell\right]}\mathbb{E}\left[\sum_{\ell=1}^{|\mathbf{z}|} \mathbf{z}_\ell\right] \\ &\leq \sqrt{\frac{|\mathbf{z}|}{2}\log\left(\frac{1}{\delta}\right)}\frac{\max(\mathbf{z})-\min(\mathbf{z})}{\min\left(\mathbb{E}\left[\sum_{\ell=1}^{|\mathbf{z}|} \mathbf{z}_\ell\right]\right)}\mathbb{E}\left[\sum_{\ell=1}^{|\mathbf{z}|} \mathbf{z}_\ell\right],\end{aligned} \tag{26}$$

where $\max(\mathbf{z})$ and $\min(\mathbf{z})$ respectively denote the maximum and minimum possible values of $\{\mathbf{z}_\ell\}_{\ell=1}^{|\mathbf{z}|}$.

Proposition 5. ***(Monotone Matrix Norms Tyrtyshnikov (1997))****: Suppose the element-wise inequality $\mathbf{Z} \geq \mathbf{Y}$ holds. Then, $\|\mathbf{Z}\|_p \geq \|\mathbf{Y}\|_p$, for a positive integer p.*

Proposition 6. ***(An upper bound on the ℓ_1 norm of the difference between two ℓ_1-normalized vectors)****: Let $\mathbf{z}, \mathbf{y} \in \mathbf{R}^m$ be such that $\mathbf{z} \geq \mathbf{1}_m$ and $\mathbf{y} \geq \mathbf{1}_m$, where $\mathbf{1}_m$ is the all-ones vector. Then,*

$$\left\|\frac{\mathbf{z} - \mathbf{1}_m}{\|\mathbf{z}-\mathbf{1}_m\|_1} - \frac{\mathbf{y} - \mathbf{1}_m}{\|\mathbf{y} - \mathbf{1}_m\|_1}\right\|_1 \leq 2\frac{\|\mathbf{z}-\mathbf{y}\|_1}{\max\left(\|\mathbf{z}-\mathbf{1}_m\|_1, \|\mathbf{y}-\mathbf{1}_m\|_1\right)}.$$

Proof. Using the triangle inequality, we have:

$$\begin{aligned}\left\|\frac{\mathbf{z}-1}{\|\mathbf{z}-\mathbf{1}_m\|_1}-\frac{\mathbf{y}-1}{\|\mathbf{y}-\mathbf{1}_m\|_1}\right\|_1 &= \sum_{i=1}^{m}\left|\frac{\mathbf{z}_i-1-(\mathbf{y}_i-1)+(\mathbf{y}_i-1)}{\|\mathbf{z}-\mathbf{1}_m\|_1}-\frac{\mathbf{y}_i-1}{\|\mathbf{y}-\mathbf{1}_m\|_1}\right| \\ &= \sum_{i=1}^{m}\left|\frac{\mathbf{z}_i-\mathbf{y}_i}{\|\mathbf{z}-\mathbf{1}_m\|_1}-(\mathbf{y}_i-1)\left(\frac{1}{\|\mathbf{y}-\mathbf{1}_m\|_1}-\frac{1}{\|\mathbf{z}-\mathbf{1}_m\|_1}\right)\right| \\ &= \sum_{i=1}^{m}\left|\frac{\mathbf{z}_i-\mathbf{y}_i}{\|\mathbf{z}-\mathbf{1}_m\|_1}-(\mathbf{y}_i-1)\left(\frac{\sum_{j=1}^{m}(\mathbf{z}_j-\mathbf{y}_j)}{\|\mathbf{y}-\mathbf{1}_m\|_1\,\|\mathbf{z}-\mathbf{1}_m\|_1}\right)\right| \\ &\leq \frac{\sum_{i=1}^{m}|\mathbf{z}_i-\mathbf{y}_i|}{\|\mathbf{z}-\mathbf{1}_m\|_1}\left(1+\frac{\sum_{i=1}^{m}|\mathbf{y}_i-1|}{\|\mathbf{y}-\mathbf{1}_m\|_1}\right)=2\frac{\|\mathbf{z}-\mathbf{y}\|_1}{\|\mathbf{z}-\mathbf{1}_m\|_1}.\end{aligned} \tag{27}$$

By exchanging the roles of $\mathbf{z}_i$ and $\mathbf{y}_i$, we find a similar bound that combined to (27) gives the inequality in the statement of Lemma. □

B Lemma: Relationship Between Fine- and Group- Average Controllability

Let $\mathbf{Z}$ be an $n\times n$ system matrix that could be either $\mathbf{A}$ or $\bar{\mathbf{A}}$ (the expected quantity). Recall the controllability Gramian of the LTI system in (1) has been defined in (2). We define the n-dimensional *fine* average controllability vector $\boldsymbol{\theta}_{\text{fine},\mathbf{Z}}$ similar to (8) as

$$\boldsymbol{\theta}_{\text{fine},\mathbf{Z}}^{\mathsf{T}} \triangleq \left[\text{tr}[\boldsymbol{\mathcal{C}}(\mathbf{Z}_{\text{nom}},\mathbf{e}_1)] \quad \dots \quad \text{tr}[\boldsymbol{\mathcal{C}}(\mathbf{Z}_{\text{nom}},\mathbf{e}_n)]\right], \tag{28}$$

where $\mathbf{e}_i$ is the i-th canonical basis vector in $\mathbb{R}^n$. By definition in (2), the following holds:

$$\begin{aligned}\boldsymbol{\theta}_{\text{fine},\mathbf{Z}}^{(i)} &= \text{tr}\left[\boldsymbol{\mathcal{C}}(\mathbf{Z}_{\text{nom}},\mathbf{e}_i)\right]=\textstyle\sum_{\tau=0}^{\infty}\text{tr}(\mathbf{Z}_{\text{nom}}^{\tau}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\mathbf{Z}_{\text{nom}}^{\tau}) \\ &= \textstyle\sum_{\tau=0}^{\infty}[\text{diag}(\mathbf{Z}_{\text{nom}}^{2\tau})]_i = 1+[\text{diag}(\mathbf{Z}_{\text{nom}}^{2})]_i+\cdots=[\text{diag}\left((\mathbf{I}-\mathbf{Z}_{\text{nom}}^{2})^{-1}\right)]_i.\end{aligned} \tag{29}$$

The series converges to the final equality because $\rho(\mathbf{Z})\leq 1$. Putting (29) into a vector form gives:

$$\boldsymbol{\theta}_{\text{fine},\mathbf{Z}}=\text{diag}\left(\boldsymbol{\mathcal{C}}(\mathbf{Z}_{\text{nom}},\mathbf{I}_{n\times n})\right)=\text{diag}\left((\mathbf{I}-\mathbf{Z}_{\text{nom}}^{2})^{-1}\right). \tag{30}$$

The following lemma states the group average controllability vector in (8) is a linear mapping of the fine average controllability vector defined in (30).

Lemma 3. *With the notation defined in (30) for* $\mathbf{Z}$ *consider the average controllability vector in (8). Then* $\boldsymbol{\theta}_{\text{group},\mathbf{Z}}=\frac{1}{r}\mathbf{W}\boldsymbol{\theta}_{\text{fine},\mathbf{Z}}$.

Proof. We begin by simplifying the group average controllability using its definition in (8) and the definition of Gramian in (2), for a general matrix $\mathbf{Z}$:

$$\begin{aligned}\boldsymbol{\theta}_{\text{group},\mathbf{Z}}^{(i)} &= \text{tr}\left[\boldsymbol{\mathcal{C}}(\mathbf{Z}_{\text{nom}},\text{diag}(\mathbf{w}_i^{\mathsf{T}}))\right] \\ &= \text{tr}\left[\sum_{\tau=0}^{\infty}\mathbf{Z}_{\text{nom}}^{\tau}\text{diag}(\mathbf{w}_i^{\mathsf{T}})\text{diag}(\mathbf{w}_i^{\mathsf{T}})^{\mathsf{T}}\mathbf{Z}_{\text{nom}}^{\tau}\right] \\ &= \text{tr}\left[\sum_{\tau=0}^{\infty}\mathbf{Z}_{\text{nom}}^{\tau}\text{diag}((\mathbf{w}_i\odot\mathbf{w}_i)^{\mathsf{T}})\mathbf{Z}_{\text{nom}}^{\tau}\right] \\ &\overset{\text{(a)}}{=} \frac{1}{r^2}\text{tr}\left[\sum_{\tau=0}^{\infty}\mathbf{Z}_{\text{nom}}^{\tau}(\sum_{v\in\text{supp}(\mathbf{w}_i)}\mathbf{e}_v\mathbf{e}_v^{\mathsf{T}})\mathbf{Z}_{\text{nom}}^{\tau}\right] \\ &= \frac{1}{r^2}\sum_{v\in\text{supp}(\mathbf{w}_i)}\underbrace{\text{tr}\left[\sum_{\tau=0}^{\infty}\mathbf{Z}_{\text{nom}}^{\tau}\mathbf{e}_v\mathbf{e}_v^{\mathsf{T}}\mathbf{Z}_{\text{nom}}^{\tau}\right]}_{\boldsymbol{\theta}_{\text{fine},\mathbf{Z}}^{(v)}} \\ &= (\mathbf{w}_i\odot\mathbf{w}_i)^{\mathsf{T}}\boldsymbol{\theta}_{\text{fine},\mathbf{Z}},\end{aligned} \tag{31}$$

where step (a) is due to the assumption of $\text{supp}(\mathbf{w}_i) = r$ at the end of Section 2.4, for all $i \in [m]$; $\odot$ denotes the Hadamard product; and $\boldsymbol{\theta}_{\text{fine},\mathbf{Z}}$ has been defined in (30) in which $\mathbf{A}$ is substituted with $\mathbf{Z}$. Putting (31) in vector form concludes the proof. □

Lemma 3 states that the i-th element in $\boldsymbol{\theta}_{\text{group},\mathbf{Z}}$ is the weighted average of a group of elements in $\boldsymbol{\theta}_{\text{fine},\mathbf{Z}}$; this group corresponds to the fine node indices that map to the i-th c-node.

Lemma 4. ***(spectral radius of $\bar{\mathbf{A}}$):*** *Let $\mathbf{D} = \frac{1}{n}\mathbf{P}\mathbf{P}^\mathsf{T}$ with ρ_n be as in (5). Then the spectral radius of $\bar{\mathbf{A}} = \mathbf{P}^\mathsf{T}\mathbf{Q}\mathbf{P}$ in (9) is given as*

$$\rho(\bar{\mathbf{A}}) = n\rho(\mathbf{Q}\mathbf{D}) = n\rho_n\rho(\mathbf{Q}_\circ\mathbf{D}).$$

Proof. We recall that the eigenvalues of $\mathbf{AB}$ and $\mathbf{BA}$ coincide for any two square matrices $\mathbf{A}$ and $\mathbf{B}$. Thus, $\rho(\bar{\mathbf{A}}) = \rho(\mathbf{Q}\mathbf{P}\mathbf{P}^\mathsf{T}) = n\rho(\mathbf{Q}\mathbf{D}) = n\rho_n\rho(\mathbf{Q}_\circ\mathbf{D})$, where for the last equality we use the relation in (5). □

The result below gives an equivalent expression for the spectral radius of the expected coarse-scale adjacency matrix $\bar{\widetilde{\mathbf{A}}}$.

Lemma 5. ***(spectral radius of $\bar{\widetilde{\mathbf{A}}}$):*** *The spectral radius of $\bar{\widetilde{\mathbf{A}}}$ can be expanded as*

$$\rho(\bar{\widetilde{\mathbf{A}}}) = \rho(\boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^\mathsf{T}) = m\rho\left(\mathbf{Q}\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m}\right). \tag{32}$$

Proof. Similar to Lemma 4. Details are omitted. □

B.1 Proof of Lemma 1: Error between the Gramians of random and expected LTI systems

Let $\overline{\mathcal{S}}_{\text{fine}}$ denote the expected dynamics of the LTI system in (6); specifically, $\mathbf{A}$ is replaced with the expected quantity $\bar{\mathbf{A}}$, while the rest of the parameters remain unchanged. We define $\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}$ similar to $\boldsymbol{\theta}_{\text{fine},\mathbf{A}}$ in (30), only by replacing $\mathbf{A}$ with $\bar{\mathbf{A}}$. Similarly, $\overline{\mathcal{S}}_{\text{coarse}}$ refers to the expected dynamics of (10) when $\widetilde{\mathbf{A}}$ is replaced with $\bar{\widetilde{\mathbf{A}}}$, without altering the rest of the parameters. $\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}$ is then defined similar to $\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}$ (c.f., (11)), as the controllability vector of $\overline{\mathcal{S}}_{\text{coarse}}$, by replacing $\widetilde{\mathbf{A}}$ with $\bar{\widetilde{\mathbf{A}}}$. The following result provides an error bound on the respective differences of the fine- and coarse- average controllability vectors. The proof of this result demonstrates how this error bound is equivalent to bounding the trace of the difference between the Gramians of $\mathcal{S}_{\text{fine}}$ and $\overline{\mathcal{S}}_{\text{fine}}$, as well as between the Gramians of $\mathcal{S}_{\text{coarse}}$ and $\overline{\mathcal{S}}_{\text{coarse}}$.

Proof. We begin by deriving an upper bound for α_n in (17). Consider,

$$\begin{aligned}
\alpha_n &\overset{(a)}{=} \left\|\text{diag}\left[\boldsymbol{\mathcal{C}}(\mathbf{A}_{\text{nom}}, \mathbf{I}_{n\times n}) - \boldsymbol{\mathcal{C}}(\bar{\mathbf{A}}_{\text{nom}}, \mathbf{I}_{n\times n})\right]\right\|_1 \\
&\overset{(b)}{=} \text{tr}\left(\left|\boldsymbol{\mathcal{C}}(\mathbf{A}_{\text{nom}}, \mathbf{I}_{n\times n}) - \boldsymbol{\mathcal{C}}(\bar{\mathbf{A}}_{\text{nom}}, \mathbf{I}_{n\times n})\right|\right) \\
&\overset{(c)}{=} \text{tr}\left(\left|(\mathbf{I} - \mathbf{A}_{\text{nom}}^2)^{-1} - (\mathbf{I} - \bar{\mathbf{A}}_{\text{nom}}^2)^{-1}\right|\right) \\
&\overset{(d)}{=} \text{tr}\left((\mathbf{I}-\mathbf{A}_{\text{nom}}^2)^{-1}|\mathbf{A}_{\text{nom}}^2 - \bar{\mathbf{A}}_{\text{nom}}^2|(\mathbf{I}-\bar{\mathbf{A}}_{\text{nom}}^2)^{-1}\right) \\
&\overset{(e)}{\leq} \left\|(\mathbf{I}-\mathbf{A}_{\text{nom}}^2)^{-1}\right\|_2 \text{tr}\left(\left|\mathbf{A}_{\text{nom}}^2 - \bar{\mathbf{A}}_{\text{nom}}^2\right|\right) \left\|(\mathbf{I}-\bar{\mathbf{A}}_{\text{nom}}^2)^{-1}\right\|_2 \\
&\overset{(f)}{\leq} \frac{1}{1-\|\mathbf{A}_{\text{nom}}^2\|_2}\frac{1}{1-\|\bar{\mathbf{A}}_{\text{nom}}^2\|_2}\text{tr}\left(\left|\mathbf{A}_{\text{nom}}^2 - \bar{\mathbf{A}}_{\text{nom}}^2\right|\right),
\end{aligned} \tag{33}$$

where step (a) follows from (30); step (b) follows from the equivalence of the of trace of the absolute value of a matrix and the ℓ_1 norm of its diagonal entries; steps (d)-(f) follow respectively from items 1-3 in the matrix

equivalence list at the beginning of the appendix; and finally, step (c) follows by expanding the Gramian matrix in (2) using the Neumann series:

$$\begin{aligned}\boldsymbol{\mathcal{C}}(\mathbf{A}_{\text{nom}}, \mathbf{I}_{n\times n}) &\triangleq \lim_{T\to\infty} \boldsymbol{\mathcal{C}}_T(\mathbf{A}_{\text{nom}}, \mathbf{I}_{n\times n}) \\ &= \lim_{T\to\infty} \textstyle\sum_{t=0}^{T-1} (\mathbf{A}_{\text{nom}})^{2t} = (\mathbf{I} - \mathbf{A}_{\text{nom}}^2)^{-1},\end{aligned} \tag{34}$$

and similarly,

$$\boldsymbol{\mathcal{C}}(\bar{\mathbf{A}}_{\text{nom}}, \mathbf{I}_{n\times n}) \triangleq \lim_{T\to\infty} \boldsymbol{\mathcal{C}}_T(\bar{\mathbf{A}}_{\text{nom}}, \mathbf{I}_{n\times n}) = (\mathbf{I} - \bar{\mathbf{A}}_{\text{nom}}^2)^{-1}.$$

The limit exits because $\mathbf{A}$ and $\bar{\mathbf{A}}$ are symmetric matrices and that $\|\mathbf{A}_{\text{nom}}\|_2 = \frac{\rho(\mathbf{A})}{\epsilon+\rho(\mathbf{A})} < 1$ and $\|\bar{\mathbf{A}}_{\text{nom}}\|_2 = \frac{\rho(\bar{\mathbf{A}})}{\epsilon+\rho(\bar{\mathbf{A}})} < 1$.

We now find an upper bound for $\text{tr}(|\mathbf{A}_{\text{nom}}^2 - \bar{\mathbf{A}}_{\text{nom}}^2|)$. Using the Hoeffding's inequality in Proposition 4, the following holds for $i, j \in [n]$ and $i \neq j$, with probability at least $1 - \delta$:

$$\begin{aligned}\left|[\mathbf{A}^2]_{ij} - \mathbb{E}[\mathbf{A}^2]_{ij}\right| &= \left|\textstyle\sum_{\ell\in[n]} \mathbf{A}_{i\ell}\mathbf{A}_{j\ell} - \mathbb{E}[\sum_{\ell\in[n]} \mathbf{A}_{i\ell}\mathbf{A}_{j\ell}]\right| \\ &\leq \sqrt{\frac{n\log(\frac{1}{\delta})}{2}} \frac{\max(\mathbf{A}_{i\ell}\mathbf{A}_{j\ell}) - \min(\mathbf{A}_{i\ell})}{\min([\bar{\mathbf{A}}^2]_{ij})} [\bar{\mathbf{A}}^2]_{ij} \\ &\overset{(a)}{\leq} \sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{2n}} \frac{1}{\|\mathbf{QDQ}\|_{\min}} [\bar{\mathbf{A}}^2]_{ij},\end{aligned} \tag{35}$$

where step (a) follows by the definition of $\mathbf{D} = (1/n)\mathbf{P}\mathbf{P}^\mathsf{T}$ for $\mathbf{P}$ in (3), coupled with the fact that

$$\bar{\mathbf{A}}^2 = \mathbf{PQP}^\mathsf{T}\mathbf{PQP}^\mathsf{T} = n\mathbf{PQDQP}^\mathsf{T}. \tag{36}$$

Similarly for the diagonal elements, the following inequality holds with probability at least $1 - \delta$:

$$\begin{aligned}\left|[\mathbf{A}^2]_{ii} - \mathbb{E}[\mathbf{A}^2]_{ii}\right| &= \left|\sum_{\ell\in[n]} \mathbf{A}_{i\ell}^2 - \mathbb{E}[\mathbf{A}_{i\ell}^2]\right| = \left|\sum_{\ell\in[n]} \mathbf{A}_{i\ell} - \mathbb{E}[\mathbf{A}_{i\ell}]\right| \\ &\overset{(a)}{\leq} \sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{2n}} \frac{1}{\|\mathbf{Qd}\|_{\min}} \mathbb{E}[\mathbf{A}^2]_{ii},\end{aligned} \tag{37}$$

where step (a) is due to Hoeffding's inequality and the identity

$$\mathbb{E}\left[\mathbf{A}^2\right]_{ii} = \mathbf{P}_i\mathbf{QP}\mathbf{1}_n = \mathbf{P}_i\mathbf{Qd} \tag{38}$$

($\mathbf{P}_i$ is the i-th column of $\mathbf{P}$). We will bound the difference between $\mathbb{E}[\mathbf{A}^2]$ and $\bar{\mathbf{A}}_{\text{nom}}^2$ shortly as we simplify $\text{tr}(|\mathbf{A}_{\text{nom}}^2 - \bar{\mathbf{A}}_{\text{nom}}^2|)$. By putting the element-by-element bounds in (35) and (37) into a matrix, the following holds with probability at least $1 - \delta$:

$$\begin{aligned}\mathbf{A}^2 &\leq (1+\eta)\mathbb{E}[\mathbf{A}^2] \\ &\overset{(a)}{\leq} (1+\eta)(\bar{\mathbf{A}}^2 - \text{Diag}[n\mathbf{P}^\mathsf{T}\mathbf{QDQP}] + n\text{Diag}[\mathbf{P}^\mathsf{T}\mathbf{Qd}]) \\ &= (1+\eta)n\left(\mathbf{P}^\mathsf{T}\mathbf{QDQP} + \text{Diag}[\mathbf{P}^\mathsf{T}(\mathbf{Q}\odot(\mathbf{1}_{K,K} - \mathbf{Q})\mathbf{d})]\right),\end{aligned} \tag{39}$$

where step (a) follows from the fact that the off-diagonal elements of $\mathbb{E}[\mathbf{A}^2]$ (as a sum of $\sum_{\ell\in[n]} A_{i,\ell}A_{j,\ell}, \forall i \neq j$ is composed of independent random variables and is equal to the off-diagonal elements of $\bar{\mathbf{A}}^2$. The rest adds and subtracts the diagonal elements from (38) and (36) to adjust those components. Also, η is defined as

$$\frac{1}{\sqrt{2n}} \max\left(\frac{\sqrt{\log(\frac{n^2-n}{\delta})}}{\|\mathbf{QDQ}\|_{\min}}, \frac{\rho_n\sqrt{\log(\frac{n}{\delta})}}{\|\mathbf{Qd}\|_{\min}}\right) \leq \sqrt{\frac{\log\left(\frac{n^2}{\delta}\right)}{2n}} \frac{1}{\|\mathbf{QDQ}\|_{\min}} \triangleq \eta.$$

Similarly, for the lower bound we get:

$$\mathbf{A}^2 \geq (1-\eta)n\big(\mathbf{P}^\mathsf{T}\mathbf{Q}\mathbf{D}\mathbf{Q}\mathbf{P}+\mathrm{Diag}\big[\mathbf{P}^\mathsf{T}(\mathbf{Q}\odot(\mathbf{1}_{K,K}-\mathbf{Q})\mathbf{d})\big]\big). \tag{40}$$

From (39) and (40), the following upper bound is obtained for the $\mathrm{tr}\left(\left|\mathbf{A}^2_{\mathrm{nom}} - \bar{\mathbf{A}}^2_{\mathrm{nom}}\right|\right)$ term:

$$\begin{aligned}
&\mathrm{tr}\left(\left|\mathbf{A}^2_{\mathrm{nom}}-\bar{\mathbf{A}}^2_{\mathrm{nom}}\right|\right) \qquad (41)\\
&\overset{(a)}{=} \mathrm{tr}\left(\left|\frac{\mathbf{A}^2-\bar{\mathbf{A}}^2}{(\epsilon+\rho(\bar{\mathbf{A}}))^2}+\left(\frac{\mathbf{A}^2}{\epsilon+\rho(\mathbf{A})}-\frac{\mathbf{A}^2}{\epsilon+\rho(\bar{\mathbf{A}})}\right)\right|\right)\\
&\overset{(b)}{\leq} \mathrm{tr}\left(\left|\frac{\eta\bar{\mathbf{A}}^2+\eta n\mathrm{Diag}[\mathbf{P}^\mathsf{T}\mathbf{Q}\odot(\mathbf{1}-\mathbf{Q})\mathbf{d}]}{(\epsilon+\rho(\bar{\mathbf{A}}))^2}+\left(\frac{\mathbf{A}^2}{(\epsilon+\rho(\mathbf{A}))^2}-\frac{\mathbf{A}^2}{(\epsilon+\rho(\bar{\mathbf{A}}))^2}\right)\right|\right)\\
&\overset{(c)}{\leq} \eta\frac{\mathrm{tr}\left(\bar{\mathbf{A}}^2\right)}{(\epsilon+\rho(\bar{\mathbf{A}}))^2}+\eta n\frac{\mathrm{tr}\left(\mathrm{Diag}[\mathbf{P}^\mathsf{T}\mathbf{Q}\odot(\mathbf{1}-\mathbf{Q})\mathbf{d}]\right)}{(\epsilon+\rho(\bar{\mathbf{A}}))^2}+\left|\frac{1}{(\epsilon+\rho(\mathbf{A}))^2}-\frac{1}{(\epsilon+\rho(\bar{\mathbf{A}}))^2}\right|\mathrm{tr}\left(\mathbf{A}^2\right)\\
&\overset{(d)}{\leq} \eta\frac{n^2\mathrm{tr}\left((\mathbf{D}\mathbf{Q})^2\right)}{(\epsilon+n\rho(\mathbf{Q}\mathbf{D}))^2}+\eta\frac{n^2\|\mathbf{Q}\odot(\mathbf{1}-\mathbf{Q})\mathbf{d}\|_{\max}}{(\epsilon+n\rho(\mathbf{Q}\mathbf{D}))^2}+\left|1-\frac{\rho(\mathbf{A})}{\rho(\bar{\mathbf{A}})}\right| \qquad (42)\\
&\qquad\times\left(1+\frac{\rho(\mathbf{A})+2\epsilon}{\rho(\bar{\mathbf{A}})}\right)\frac{(1+\eta)n^2\left[\mathrm{tr}\left((\mathbf{D}\mathbf{Q})^2\right)+\|\mathbf{Q}\mathbf{d}\|_{\max}\right]}{(1-\eta)\left(n^2\rho^2(\mathbf{Q}\mathbf{D})+n\|\mathbf{Q}\odot(\mathbf{1}_{K,K}-\mathbf{Q})\mathbf{d}\|_{\max}\right)}\\
&\overset{(e)}{\leq} \eta\left[\frac{1/\rho_n+\mathrm{tr}\left((\mathbf{D}\mathbf{Q}_\circ)^2\right)}{\rho^2(\mathbf{Q}_\circ\mathbf{D})}+\frac{1}{2}(1+4\eta)\frac{\left[\mathrm{tr}\left((\mathbf{D}\mathbf{Q}_\circ)^2\right)+\frac{1}{\rho_n}\right]}{\rho^2(\mathbf{Q}_\circ\mathbf{D})}\left(1+\sqrt{1+\eta}\sqrt{1+\frac{1}{n\rho_n\rho^2(\mathbf{Q}_\circ\mathbf{D})}}+\frac{2\epsilon}{n\rho_n\rho(\mathbf{Q}_\circ\mathbf{D})}\right)\right],
\end{aligned}$$

where step (a) follows from the normalization factors in (1) and an additive-subtractive term; step (b) from (39); step (c) from the triangle inequality and the linearity of trace; and step (d) is because of: first, (65) and (66) which are derived at the end of this proof; second the definition of $\mathbf{P}$ in (3); third, the inequality

$$\begin{aligned}
\left|\frac{1}{(\epsilon+\rho(\mathbf{A}))^2}-\frac{1}{(\epsilon+\rho(\bar{\mathbf{A}}))^2}\right| &= \left|\frac{(\epsilon+\rho(\bar{\mathbf{A}}))^2-(\epsilon+\rho(\mathbf{A}))^2}{(\epsilon+\rho(\mathbf{A}))^2(\epsilon+\rho(\bar{\mathbf{A}}))^2}\right|\\
&= \left|\frac{(\rho(\bar{\mathbf{A}})+\rho(\mathbf{A})+2\epsilon)(\rho(\bar{\mathbf{A}})-\rho(\mathbf{A}))}{(\epsilon+\rho(\mathbf{A}))^2(\epsilon+\rho(\bar{\mathbf{A}}))^2}\right|\\
&\leq \frac{\left[1+\frac{\rho(\mathbf{A})+2\epsilon}{\rho(\bar{\mathbf{A}})}\right]\left|1-\frac{\rho(\mathbf{A})}{\rho(\bar{\mathbf{A}})}\right|}{(\epsilon+\rho(\mathbf{A}))^2};
\end{aligned}$$

and finally the equality $\mathrm{tr}\left(\bar{\mathbf{A}}^2\right) = n^2\mathrm{tr}\left((\mathbf{D}\mathbf{Q})^2\right)$. For the last equality, (9) and the invariance of trace under cyclic permutation were used; finally, (e) follows from (67), (68) (derived at the end of this proof), and two easy-to-validate inequalities $\sqrt{1-\eta} \geq 1-\eta/2$ and $\frac{1+\eta}{1-\eta} \leq 1+4\eta, \forall\eta < 1/2$. From (41) and the normalization factor in (6) for $\mathbf{A}$ and its expectation, we further simplify (33) as

$$\begin{aligned}
\alpha_n &\leq \frac{\eta\left[\frac{1/\rho_n+c_1}{c_\circ^2}+(\frac{1}{2}+2\eta)\frac{\left[c_1+\frac{1}{\rho_n}\right]}{c_\circ^2}\left(1+\sqrt{1+\eta}\sqrt{1+\frac{1/\rho_n}{nc_\circ^2}}+\frac{2\epsilon/\rho_n}{nc_\circ}\right)\right]}{\left[1-\left(\frac{c_\circ}{\epsilon/\rho_n+c_\circ}\right)^2\right]\left[1-\left(\frac{\rho(\mathbf{A})}{\epsilon+\rho(\mathbf{A})}\right)^2\right]} \qquad (43)\\
&\overset{(a)}{=} \mathcal{O}\left(\frac{\eta}{\rho_n}\right).
\end{aligned}$$

In (43), $c_\circ = \rho(\mathbf{Q}_\circ\mathbf{D}), c_1 = \mathrm{tr}\left((\mathbf{D}\mathbf{Q}_\circ)^2\right)$, and step (a) holds in the asymptotic scenario when $\mathbf{D}, \mathbf{Q}$ are constant and $\sqrt{n}\rho_n$ is very large and hence, η becomes very small.

The proof of the inequality in (18) bears some similarity to the upper bound derivations for α_n in (33)-(43).

Following similar lines as those in (33), the upper bound on $\tilde{\alpha}_m$ (18) can be simplified to

$$\tilde{\alpha}_m \leq \frac{1}{1-\|\widetilde{\mathbf{A}}^2_{\mathrm{nom}}\|_2}\frac{1}{1-\|\bar{\widetilde{\mathbf{A}}}^2_{\mathrm{nom}}\|_2}\mathrm{tr}(|\widetilde{\mathbf{A}}^2_{\mathrm{nom}}-\bar{\widetilde{\mathbf{A}}}^2_{\mathrm{nom}}|). \tag{44}$$

To get (44), one should note that (2) and the Neumann series yield

$$\begin{aligned}\mathcal{C}(\widetilde{\mathbf{A}}_{\text{nom}}, \mathbf{I}_{m\times m}) &\triangleq \lim_{T\to\infty} \mathcal{C}_T(\widetilde{\mathbf{A}}_{\text{nom}}, \mathbf{I}_{m\times m}) \\ &= \lim_{T\to\infty} \sum_{t=0}^{T-1} \widetilde{\mathbf{A}}_{\text{nom}}^t \left(\widetilde{\mathbf{A}}_{\text{nom}}^{\mathsf{T}}\right)^t \\ &= \lim_{T\to\infty} \sum_{t=0}^{T-1} \widetilde{\mathbf{A}}_{\text{nom}}^{2t} = (\mathbf{I} - \widetilde{\mathbf{A}}_{\text{nom}}^2)^{-1},\end{aligned} \tag{45}$$

since $\|\widetilde{\mathbf{A}}_{\text{nom}}\|_2 = \frac{\rho(\widetilde{\mathbf{A}})}{\epsilon+\rho(\widetilde{\mathbf{A}})} < 1$. Next, we simplify the term $\text{tr}(|\widetilde{\mathbf{A}}_{\text{nom}}^2 - \bar{\widetilde{\mathbf{A}}}_{\text{nom}}^2|)$ in (44) using similar concentration bounds as those in (41). From Proposition 4, Lemma 5, and the definition of $\bar{\widetilde{\mathbf{A}}}$ in (9), the following holds for $i, j \in [m]$ and $i \neq j$ with probability at least $1 - \tilde{\delta}$:

$$\begin{aligned}|[\widetilde{\mathbf{A}}^2]_{ij} - \mathbb{E}[\widetilde{\mathbf{A}}^2]_{ij}| &= |\sum_{\ell\in[m]} \widetilde{\mathbf{A}}_{i\ell}\widetilde{\mathbf{A}}_{j\ell} - \underbrace{\mathbb{E}[\sum_{\ell\in[m]} \widetilde{\mathbf{A}}_{i\ell}\widetilde{\mathbf{A}}_{j\ell}]}_{[\bar{\widetilde{\mathbf{A}}}^2]_{ij}}| \\ &\leq \sqrt{\frac{m\log\left(\frac{1}{\tilde{\delta}}\right)}{2}} \frac{\max(\widetilde{\mathbf{A}}_{i\ell}\widetilde{\mathbf{A}}_{j\ell}) - \min(\widetilde{\mathbf{A}}_{i\ell}\widetilde{\mathbf{A}}_{j\ell})}{\min(\mathbf{\Phi}_i^{\mathsf{T}}\mathbf{Q}\mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}_j)} \mathbb{E}[\widetilde{\mathbf{A}}^2]_{ij} \\ &\leq \sqrt{\frac{\log\left(\frac{1}{\tilde{\delta}}\right)}{2m}} \frac{1}{\left\|\mathbf{Q}\frac{\mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}}{m}\mathbf{Q}\right\|_{\min}} \mathbb{E}[\widetilde{\mathbf{A}}^2]_{ij}.\end{aligned} \tag{46}$$

Similarly, for the diagonal elements, we have:

$$\left|[\widetilde{\mathbf{A}}^2]_{ii} - \mathbb{E}[\widetilde{\mathbf{A}}_{ii}^2]\right| = \left|\sum_{\ell\in[m]} \widetilde{\mathbf{A}}_{i\ell}^2 - \mathbb{E}[\widetilde{\mathbf{A}}_{i\ell}^2]\right| \leq \sqrt{\frac{m\log\left(\frac{1}{\tilde{\delta}}\right)}{2}} \frac{\max(\widetilde{\mathbf{A}}_{i\ell}^2) - \min(\widetilde{\mathbf{A}}_{i\ell}^2)}{\min(\mathbb{E}[\sum_{\ell\in[m]} \widetilde{\mathbf{A}}_{i\ell}^2])} \mathbb{E}[\sum_{\ell\in[m]} \widetilde{\mathbf{A}}_{i\ell}^2]. \tag{47}$$

Next, we simplify the term $\mathbb{E}[\widetilde{\mathbf{A}}_{i\ell}^2]$ in (47). From Section 2.2, we can verify that

$$r^2\widetilde{\mathbf{A}}_{ij} \sim \text{Poisson-Binomial}(\{\underbrace{\mathbf{Q}_{kk'}\cdots\mathbf{Q}_{kk'}}_{r^2\mathbf{\Phi}_{ik}\mathbf{\Phi}_{jk'}} : k, k' \in [K]\}), \tag{48}$$

since $r^2\widetilde{\mathbf{A}}_{ij}$ is a sum of independent (but not necessarily identically) Bernoulli random variables, parameterized by $\mathbf{Q}_{kk'}$s for $r^2\mathbf{\Phi}_{ik}\mathbf{\Phi}_{jk'}$ of the variables contributing to the sum. We can denote (48) concisely by:

$$r^2\widetilde{\mathbf{A}}_{ij} \sim \text{Poisson-Binomial}(\{\text{vec}(\mathbf{Q})\}^{\odot\text{vec}(r^2\mathbf{\Phi}_i\mathbf{\Phi}_j^{\mathsf{T}})}), \tag{49}$$

where $\text{vec}(\mathbf{Z})$ outputs a vector containing all the elements in its input matrix $\mathbf{Z}$. We use $\{\text{vec}(\mathbf{Z})\}^{\odot\text{vec}(\mathbf{Y})}$ to denote a vector that contains all the elements of $\mathbf{Z}$, where $\mathbf{Z}_{ij}$ has $\mathbf{Y}_{ij}$ occurrences. From (49), and using the variance definition of Poisson-Binomial, the term $\mathbb{E}[\widetilde{\mathbf{A}}_{i\ell}^2]$ in (47) can be represented as

$$\begin{aligned}\mathbb{E}[\widetilde{\mathbf{A}}_{i\ell}^2] &= \mathbb{E}[\widetilde{\mathbf{A}}_{i\ell}]^2 + \text{var}(\widetilde{\mathbf{A}}_{i\ell}) = (\mathbf{\Phi}_i^{\mathsf{T}}\mathbf{Q}\mathbf{\Phi}_\ell)^2 + \frac{1}{r^4}(\mathbf{1} - \{\text{vec}(\mathbf{Q})\}^{\odot\text{vec}(r^2\mathbf{\Phi}_i\mathbf{\Phi}_\ell^{\mathsf{T}})})^{\mathsf{T}}\{\text{vec}(\mathbf{Q})\}^{\odot\text{vec}(r^2\mathbf{\Phi}_i\mathbf{\Phi}_\ell^{\mathsf{T}})} \\ &= (\mathbf{\Phi}_i^{\mathsf{T}}\mathbf{Q}\mathbf{\Phi}_\ell)^2 + \frac{1}{r^4}\text{vec}(r^2\mathbf{\Phi}_i\mathbf{\Phi}_\ell^{\mathsf{T}})^{\mathsf{T}}[\text{vec}(\mathbf{Q}) \odot (\mathbf{1}_{K^2} - \text{vec}(\mathbf{Q}))] \\ &= \rho_n^2(\mathbf{\Phi}_i^{\mathsf{T}}\mathbf{Q}_\circ\mathbf{\Phi}_\ell)^2 + \frac{\rho_n}{r^2}\text{vec}(\mathbf{\Phi}_i\mathbf{\Phi}_\ell^{\mathsf{T}})^{\mathsf{T}}\text{vec}(\mathbf{Q}_\circ \odot (\mathbf{1}_{K,K} - \mathbf{Q})).\end{aligned} \tag{50}$$

We then sum over the ℓ index of (50) to obtain the $\mathbb{E}[\widetilde{\mathbf{A}}_{ii}^2]$ term in (47):

$$\begin{aligned}\mathbb{E}[\widetilde{\mathbf{A}}_{ii}^2] = \mathbb{E}[\sum_{\ell\in[m]}\widetilde{\mathbf{A}}_{i\ell}^2] =& \mathbf{\Phi}_i^\mathsf{T}\mathbf{Q}\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}_i + \frac{1}{r^2}\mathrm{vec}(\mathbf{\Phi}_i\overbrace{\sum_{\ell\in[m]}\mathbf{\Phi}_\ell^\mathsf{T}}^{\mathbf{1}_m^\mathsf{T}\mathbf{\Phi}})^\mathsf{T}\mathrm{vec}(\mathbf{Q}\odot(\mathbf{1}_{K,K}-\mathbf{Q})) \\ =& \mathbf{\Phi}_i^\mathsf{T}\mathbf{Q}\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}_i + \frac{1}{r^2}\mathrm{vec}(\mathbf{\Phi}_i\mathbf{1}_m^\mathsf{T}\mathbf{\Phi})^\mathsf{T}\mathrm{vec}(\mathbf{Q}\odot(\mathbf{1}_{K,K}-\mathbf{Q})) \\ \geq& m\|\mathbf{Q}\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\mathbf{Q}\|_{\min}.\end{aligned} \tag{51}$$

Similarly an upper bound for (51) is obtained:

$$\begin{aligned}\mathbb{E}[\widetilde{\mathbf{A}}_{ii}^2] = \mathbb{E}[\sum_{\ell\in[m]}\widetilde{\mathbf{A}}_{i\ell}^2] &= m\rho_n^2\mathbf{\Phi}_i^\mathsf{T}\mathbf{Q}_\circ\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\mathbf{Q}_\circ\mathbf{\Phi}_i + \frac{\rho_n}{4r^2}\mathrm{vec}(\mathbf{\Phi}_i\mathbf{1}^\mathsf{T}\mathbf{\Phi})^\mathsf{T}\mathbf{1} \\ &\leq m\rho_n^2\|\mathbf{Q}_\circ\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\mathbf{Q}_\circ\|_{\max} + \frac{\rho_n}{4r^2}\underbrace{\sum_{\ell\in[m],k,k'\in[K]}\mathbf{\Phi}_{ik}\mathbf{\Phi}_{\ell k'}}_{=m}.\end{aligned} \tag{52}$$

From (51), and the fact $0 \leq \widetilde{\mathbf{A}}_{i,\ell} \leq 1$ by definition in (7), the upper bound on (47) can be further simplified as

$$|[\widetilde{\mathbf{A}}^2]_{ii} - \mathbb{E}[\widetilde{\mathbf{A}}_{ii}^2]| \leq \sqrt{\frac{\log\left(\frac{1}{\tilde{\delta}}\right)}{2m}}\frac{1}{\|\mathbf{Q}\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\mathbf{Q}\|_{\min}}\mathbb{E}[\sum_{\ell\in[m]}\widetilde{\mathbf{A}}_{i\ell}^2]. \tag{53}$$

By merging (46) and (53) and putting $[\widetilde{\mathbf{A}}^2]_{ij}$'s in a matrix form, we get

$$\begin{aligned}\widetilde{\mathbf{A}}^2 &\leq (1+\tilde{\eta})\mathbb{E}[\widetilde{\mathbf{A}}^2] \\ &\leq (1+\tilde{\eta})\left(\bar{\widetilde{\mathbf{A}}}^2 + \frac{m\rho_n}{4r^2}\mathbf{I}\right) \\ &= (1+\tilde{\eta})\left((\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^\mathsf{T})^2 + \frac{m\rho_n}{4r^2}\mathbf{I}\right) \\ &= (1+\tilde{\eta})m\left(\mathbf{\Phi}\mathbf{Q}\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\mathbf{Q}\mathbf{\Phi}^\mathsf{T} + \frac{\rho_n}{4r^2}\mathbf{I}\right),\end{aligned} \tag{54}$$

and

$$\widetilde{\mathbf{A}}^2 \geq (1-\tilde{\eta})m\mathbf{\Phi}\mathbf{Q}\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\mathbf{Q}\mathbf{\Phi}^\mathsf{T}, \tag{55}$$

which both hold with probability at least $1-\tilde{\delta}$ for $\tilde{\eta}$ in 19. Similar to the α_n derivations, further norm inequalities are obtained from (54) and (55), which will be then used to simplify the upper bound on $\tilde{\alpha}_m$ in (44). Using Proposition 5), $\rho(\widetilde{\mathbf{A}})^2$ can be bounded from below by

$$\rho(\widetilde{\mathbf{A}})^2 = \rho(\widetilde{\mathbf{A}}^2) \geq (1-\tilde{\eta})m^2\rho_n^2\rho^2\left(\mathbf{Q}_\circ\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\right), \tag{56}$$

and bounded from above by

$$\rho(\widetilde{\mathbf{A}})^2 \leq (1+\tilde{\eta})m[m\rho_n^2\rho^2\left(\mathbf{Q}_\circ\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\right) + \frac{\rho_n}{4r^2}]. \tag{57}$$

Equation (56), coupled with Lemma 5, gives

$$\begin{aligned}\frac{\rho(\widetilde{\mathbf{A}})}{\rho(\bar{\widetilde{\mathbf{A}}})} &\leq \frac{\sqrt{(1+\tilde{\eta})m[m\rho_n^2\rho^2\left(\mathbf{Q}_\circ\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\right)+\frac{\rho_n}{4r^2}]}}{m\rho\left(\mathbf{Q}\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\right)} \\ &\leq \sqrt{1+\tilde{\eta}}\sqrt{1+\frac{1}{4r^2m\rho_n\rho^2\left(\mathbf{Q}_\circ\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\right)}}.\end{aligned} \tag{58}$$

Similarly, with the help of (57), we obtain

$$\frac{\rho(\widetilde{\mathbf{A}})}{\rho(\bar{\widetilde{\mathbf{A}}})} \geq \sqrt{1-\tilde{\eta}}\sqrt{1+\frac{1}{4r^2m\rho_n\rho^2\left(\mathbf{Q}_\circ\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\right)}} \geq \sqrt{1-\tilde{\eta}}. \tag{59}$$

In addition, (9) and the invariance of trace under cyclic permutation, yield the following inequality

$$\mathrm{tr}\left(\bar{\widetilde{\mathbf{A}}}^2\right) = m^2\mathrm{tr}\left((\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\mathbf{Q})^2\right). \tag{60}$$

Combining (60) and (54), gives

$$\begin{aligned}\mathrm{tr}\left(\widetilde{\mathbf{A}}^2\right) &\leq \mathrm{tr}\left[(1+\tilde{\eta})m\left(\mathbf{\Phi}\mathbf{Q}\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\mathbf{Q}\mathbf{\Phi}^\mathsf{T}+\frac{\rho_n}{4r^2}\mathbf{I}\right)\right] \\ &\leq (1+\tilde{\eta})m^2\left[\mathrm{tr}\left((\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\mathbf{Q})^2\right)+\frac{\rho_n}{4r^2}\right].\end{aligned} \tag{61}$$

Similar to (43), we bound the term $\mathrm{tr}(|\widetilde{\mathbf{A}}^2_{\mathrm{nom}} - \bar{\widetilde{\mathbf{A}}}^2_{\mathrm{nom}}|)$ to finalize the upper bound simplification in (44) (inner steps are removed due to redundancy):

$$\begin{aligned}\tilde{\alpha}_m &\leq \frac{\tilde{\eta}\left[\frac{1/\rho_n+\tilde{c}_1}{\tilde{c}_\circ^2}+(\frac{1}{2}+2\tilde{\eta})\frac{\left[\tilde{c}_1+\frac{1}{\rho_n}\right]}{\tilde{c}_\circ^2}\left(1+\sqrt{1+\tilde{\eta}}\sqrt{1+\frac{1/\rho_n}{m\tilde{c}_\circ^2}}+\frac{2\epsilon/\rho_n}{m\tilde{c}_\circ}\right)\right]}{\left[1-\left(\frac{\tilde{c}_\circ}{\epsilon/\rho_n+\tilde{c}_\circ}\right)^2\right]\left[1-\left(\frac{\rho(\widetilde{\mathbf{A}})}{\epsilon+\rho(\widetilde{\mathbf{A}})}\right)^2\right]} \\ &= \mathcal{O}(\frac{\tilde{\eta}}{\rho_n}),\end{aligned} \tag{62}$$

where $\tilde{c}_\circ = \rho\left(\mathbf{Q}_\circ\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\right), \tilde{c}_1 = \mathrm{tr}\left((\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}\mathbf{Q}_\circ)^2\right)$. The last scaling function is similarly found as that of (43), for the asymptotic scenario when $\sqrt{m}\rho_n$ is very large and hence $\tilde{\eta}$ becomes very small.

Bounds on $\rho(\mathbf{A}), \mathrm{tr}\left(\mathbf{A}^2\right)$, and $\frac{\rho(\mathbf{A})}{\rho(\mathbf{A})}$: From (39), (40), and Proposition 5, we obtain several inequalities:

$$\begin{aligned}\rho^2(\mathbf{A}) &= \rho(\mathbf{A}^2) \\ &\geq (1-\eta)\left(\rho(\bar{\mathbf{A}})^2 n\rho(\mathrm{Diag}[\mathbf{P}^\mathsf{T}(\mathbf{Q}\odot(\mathbf{1}_{K,K}-\mathbf{Q})\mathbf{d})])\right) \\ &= (1-\eta)\left(n^2\rho^2(\mathbf{QD})+n\|\mathbf{Q}\odot(\mathbf{1}_{K,K}-\mathbf{Q})\mathbf{d}\|_{\max}\right),\end{aligned} \tag{63}$$

$$\rho^2(\mathbf{A}) \leq (1+\eta)n\left(n\rho^2(\mathbf{QD})+\|\mathbf{Q}\odot(\mathbf{1}_{K,K}-\mathbf{Q})\mathbf{d}\|_{\max}\right) \tag{64}$$

$$\leq (1+\eta)n^2\rho_n^2\left(\rho^2(\mathbf{Q}_\circ\mathbf{D})+\frac{1}{n\rho_n}\right), \tag{65}$$

Moreover, from (39) and the invariance of trace under cyclic permutation, we obtain

$$\begin{aligned}\mathrm{tr}(\mathbf{A}^2) &\leq \mathrm{tr}\left[(1+\eta)n\left(\mathbf{P}^\mathsf{T}\mathbf{QDQP}+\mathrm{Diag}[\mathbf{P}^\mathsf{T}(\mathbf{Q}\odot(\mathbf{1}_{K,K}-\mathbf{Q})\mathbf{d})]\right)\right] \\ &\leq (1+\eta)n^2\left[\mathrm{tr}\left((\mathbf{DQ})^2\right)+\|\mathbf{Qd}\|_{\max}\right]\end{aligned} \tag{66}$$

Moreover, Lemma 4 and (63) give

$$\frac{\rho(\mathbf{A})}{\rho(\bar{\mathbf{A}})} \leq \frac{\sqrt{(1+\eta)n^2\left(\rho^2(\mathbf{QD})+\frac{1}{n\rho_n}\right)}}{n\rho(\mathbf{QD})} \leq \sqrt{1+\eta}\sqrt{1+\frac{1}{n\rho_n\rho^2(\mathbf{Q}_\circ\mathbf{D})}}. \tag{67}$$

Similarly, by combining Lemma 4 and (65), we have

$$\frac{\rho(\mathbf{A})}{\rho(\bar{\mathbf{A}})} \geq \sqrt{(1-\eta)+(1-\eta)\frac{\|\mathbf{Q}_\circ \odot (\mathbf{1}_{K,K}-\mathbf{Q})\mathbf{d}\|_{\max}}{n\rho_n\rho^2(\mathbf{Q}_\circ\mathbf{D})}} \geq \sqrt{1-\eta}. \tag{68}$$

B.2 Proof of Lemma 2

We find the group controllability of $\overline{\mathcal{S}}_{\text{fine}}$, obtained by replacing $\mathbf{A}$ with $\bar{\mathbf{A}}$ in (6). To do so, we simplify $\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}$ and substitute the result in Lemma 3. By replacing $\mathbf{Z}$ in (28) with $\bar{\mathbf{A}}$, we have

$$\begin{aligned}
\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}} &= \text{diag}\left(\textstyle\sum_{\tau=0}^{\infty} \bar{\mathbf{A}}_{\text{nom}}^{2\tau}\right) \overset{(a)}{=} \text{diag}\left(\textstyle\sum_{\tau=0}^{\infty} (\frac{\iota}{n}\mathbf{P}^\mathsf{T}\mathbf{Q}_\circ\mathbf{P})^{2\tau}\right) \\
&= \text{diag}\left(\mathbf{I}+(\frac{\iota}{n})^2\mathbf{P}^\mathsf{T}\mathbf{Q}_\circ\mathbf{P}\mathbf{P}^\mathsf{T}\mathbf{Q}_\circ\mathbf{P}+\cdots\right) \\
&= \mathbf{1}_n+\frac{1}{n}\text{diag}\left(\mathbf{P}^\mathsf{T}\mathbf{D}^{-\frac{1}{2}}\left[(\iota\mathbf{D}^{\frac{1}{2}}\mathbf{Q}_\circ\mathbf{D}^{\frac{1}{2}})^2+\cdots\right]\mathbf{D}^{-\frac{1}{2}}\mathbf{P}\right) \\
&= \mathbf{1}_n+\frac{1}{n}\text{diag}\left(\mathbf{P}^\mathsf{T}\boldsymbol{\Upsilon}\mathbf{P}\right) \overset{(b)}{=} \mathbf{1}_n+\frac{1}{n}\mathbf{P}^\mathsf{T}\text{diag}(\boldsymbol{\Upsilon}),
\end{aligned} \tag{69}$$

where $\iota \triangleq \frac{1}{\frac{\epsilon}{\rho_n n}+\rho(\mathbf{Q}_\circ\mathbf{D})} = \mathcal{O}(1)$ since $\frac{\epsilon}{n\rho_n} \xrightarrow{n\to\infty} 0$, and $\boldsymbol{\Upsilon}$ is given by

$$\boldsymbol{\Upsilon} = \mathbf{D}^{-\frac{1}{2}}\left(\left[\mathbf{I}-(\iota\mathbf{D}^{\frac{1}{2}}\mathbf{Q}_\circ\mathbf{D}^{\frac{1}{2}})^2\right]^{-1}-\mathbf{I}\right)\mathbf{D}^{-\frac{1}{2}} = \iota^2\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ(\mathbf{I}-(\iota\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ)^2)^{-1}; \tag{70}$$

step (a) in (69) follows from the definition of $\bar{\mathbf{A}}$ in (9) and Lemma 4; step (b) of (69) is due to the special structure of $\mathbf{P}$ in (3), where for an arbitrary matrix $\mathbf{Z}$ of appropriate size the following equality holds:

$$[\text{diag}(\mathbf{P}^\mathsf{T}\mathbf{Z}\mathbf{P})]_i = \sum_{k,k'\in[K]} \mathbf{P}_{ki}\mathbf{Z}_{k,k'}\mathbf{P}_{k'i} = \sum_{k\in[K]} \mathbf{P}_{ki}^2\mathbf{Z}_{k,k} \overset{(c)}{=} \mathbf{P}_i\text{diag}(\mathbf{Z}),$$

in which step (c) is true because $\mathbf{P}$ is binary. Combining Lemma 3 and (69), with $\boldsymbol{\Phi}$ defined in (9), yields:

$$\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}} = \frac{1}{r}\mathbf{W}\left(\mathbf{1}_n+\frac{1}{n}\mathbf{P}^\mathsf{T}\text{diag}(\boldsymbol{\Upsilon})\right) = \frac{1}{r}\left(\mathbf{1}_m+\frac{1}{n}\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})\right).$$

B.3 Proof of Theorem 1

Let $i \in [m]$, and recall that $\mathbf{B}_i = \text{diag}(\mathbf{w}_i^\mathsf{T})$ in (6) and $\widetilde{\mathbf{B}}_i = r\mathbf{W}\text{diag}(\mathbf{w}_i^\mathsf{T})$ in (10). Consider the following bound on $\tilde{\Delta}(\mathbf{A},\widetilde{\mathbf{A}})$ in (12), using the triangle inequality for matrix norms:

$$\begin{aligned}
\tilde{\Delta}(\mathbf{A},\widetilde{\mathbf{A}}) \leq\ & \underbrace{\left\|\frac{r\boldsymbol{\theta}_{\text{group},\mathbf{A}}-\mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\mathbf{A}}-\mathbf{1}_m\|_1}-\frac{r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}}-\mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}}-\mathbf{1}_m\|_1}\right\|_1}_{=\Delta(\mathbf{A},\bar{\mathbf{A}})} \\
&+\underbrace{\left\|\frac{\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}-\mathbf{1}_m}{\|\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}-\mathbf{1}_m\|_1}-\frac{\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}-\mathbf{1}_m}{\|\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}-\mathbf{1}_m\|_1}\right\|_1}_{=\tilde{\tilde{\Delta}}(\widetilde{\mathbf{A}},\bar{\widetilde{\mathbf{A}}})} \\
&+\underbrace{\left\|\frac{r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}}-\mathbf{1}_m}{\|r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}}-\mathbf{1}_m\|_1}-\frac{\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}-\mathbf{1}_m}{\|\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}-\mathbf{1}_m\|_1}\right\|_1}_{=\tilde{\Delta}(\bar{\mathbf{A}},\bar{\widetilde{\mathbf{A}}})},
\end{aligned} \tag{71}$$

where $\boldsymbol{\theta}_{\text{group},\mathbf{A}}$ and $\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}$ are in (8) and (11). We define $\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}}$ and $\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}$ by replacing $\mathbf{A}$ with $\bar{\mathbf{A}}$ in (8), and $\widetilde{\mathbf{A}}$ with $\bar{\widetilde{\mathbf{A}}}$ in (11).

An upper bound on $\Delta(\mathbf{A},\bar{\mathbf{A}})$ has been provided in Theorem 2. We now derive an upper bound on $\tilde{\bar{\Delta}}(\widetilde{\mathbf{A}},\bar{\widetilde{\mathbf{A}}})$ with the help of Proposition 6. Replacing $\mathbf{Z}$ with $\bar{\widetilde{\mathbf{A}}}$ in (30) yields $\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}} = \text{diag}\left((\mathbf{I}-\bar{\widetilde{\mathbf{A}}}^2_{\text{nom}})^{-1}\right)$. Hence

$$\begin{aligned}\left\|\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}-\mathbf{1}\right\|_1 &= \text{tr}\left[(\mathbf{I}-\bar{\widetilde{\mathbf{A}}}^2_{\text{nom}})^{-1}-\mathbf{I}\right] = \text{tr}\left[\bar{\widetilde{\mathbf{A}}}^2_{\text{nom}}(\mathbf{I}-\bar{\widetilde{\mathbf{A}}}^2_{\text{nom}})^{-1}\right] \\ &\geq \text{tr}\left[\bar{\widetilde{\mathbf{A}}}^2_{\text{nom}}\right] \overset{(a)}{\geq} \frac{\text{tr}\left[(\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m}\mathbf{Q}_\circ)^2\right]}{\left(\frac{\epsilon}{m\rho_n}+\rho\left(\mathbf{Q}_\circ\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m}\right)\right)^2},\end{aligned} \tag{72}$$

where for step (a) (9) the normalization factor in (10), and the invariance of trace and spectral radius under cyclic permutation are used. Replacing $\mathbf{z}$ and $\mathbf{y}$ in Proposition 6 with $\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}$ and $\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}$ yields the following bound on $\tilde{\bar{\Delta}}(\widetilde{\mathbf{A}},\bar{\widetilde{\mathbf{A}}})$ in (71):

$$\begin{aligned}\tilde{\bar{\Delta}}(\widetilde{\mathbf{A}},\bar{\widetilde{\mathbf{A}}}) &\leq \frac{2\|\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}-\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}\|_1}{\max\left(\|\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}-\mathbf{1}_m\|_1,\|\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}-\mathbf{1}_m\|_1\right)} \\ &\overset{(a)}{\leq} \frac{2\tilde{\alpha}_m}{\|\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}-\mathbf{1}_m\|_1} \overset{(b)}{\leq} 2\frac{\left(\frac{\epsilon}{m\rho_n}+\rho\left(\mathbf{Q}_\circ\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m}\right)\right)^2}{\text{tr}((\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m}\mathbf{Q}_\circ)^2)}\tilde{\alpha}_m \overset{(c)}{=} \mathcal{O}\left(\frac{\tilde{\eta}}{\rho_n}\right),\end{aligned} \tag{73}$$

with probability at least $1-\tilde{\delta}$, where in step (a), $\left\|\boldsymbol{\theta}_{\text{coarse},\widetilde{\mathbf{A}}}-\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}\right\|_1$ is substituted with $\tilde{\alpha}_m$ as defined in (18); in (b), we replace $\|\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}-\mathbf{1}_m\|_1$ by its upper bound in (72); and in (c), the scaling function in (18) is used, assuming n,m are sufficiently large.

Next, we find an upper bound for the remaining error term $\tilde{\Delta}(\bar{\mathbf{A}},\bar{\widetilde{\mathbf{A}}})$ in (71). The proof is similar to the derivations of (16) and (73); hence, we provide high-level details. We first derive a closed-form formula for $\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}}$:

$$\begin{aligned}\boldsymbol{\theta}_{\text{coarse},\bar{\widetilde{\mathbf{A}}}} &= \text{diag}\left(\sum_{\tau=0}^{\infty}\bar{\widetilde{\mathbf{A}}}^{2\tau}_{\text{nom}}\right) = \text{diag}\left(\sum_{\tau=0}^{\infty}(\frac{\tilde{\iota}}{m}\boldsymbol{\Phi}\mathbf{Q}_\circ\boldsymbol{\Phi}^\mathsf{T})^{2\tau}\right) \\ &= \mathbf{1}_m + \frac{1}{m}\text{diag}\left(\boldsymbol{\Phi}\tilde{\boldsymbol{\Upsilon}}\boldsymbol{\Phi}^\mathsf{T}\right),\end{aligned} \tag{74}$$

where $tilde\boldsymbol{\Upsilon} \triangleq \tilde{\iota}^2\mathbf{Q}_\circ\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m}\mathbf{Q}_\circ(\mathbf{I}-(\tilde{\iota}\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m}\mathbf{Q}_\circ)^2)^{-1}$ and $\tilde{\iota} \triangleq 1/\left(\frac{\epsilon}{\rho_n m}+\rho\left(\mathbf{Q}\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m}\right)\right)$. From (20) and (74), the upper bound on $\tilde{\Delta}(\bar{\mathbf{A}},\bar{\widetilde{\mathbf{A}}})$ in (71) is simplified as

$$\begin{aligned}\tilde{\Delta}(\bar{\mathbf{A}},\bar{\widetilde{\mathbf{A}}}) &= \left\|\frac{\frac{1}{n}\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})}{\|\frac{1}{n}\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})\|_1} - \frac{\frac{1}{m}\text{diag}(\boldsymbol{\Phi}\tilde{\boldsymbol{\Upsilon}}\boldsymbol{\Phi}^\mathsf{T})}{\|\frac{1}{m}\text{diag}(\boldsymbol{\Phi}\tilde{\boldsymbol{\Upsilon}}\boldsymbol{\Phi}^\mathsf{T})\|_1}\right\|_1 \\ &\overset{(a)}{\leq} 2\frac{\|\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})-\text{diag}(\boldsymbol{\Phi}\tilde{\boldsymbol{\Upsilon}}\boldsymbol{\Phi}^\mathsf{T})\|_1}{\max\left(\|\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})\|_1,\|\text{diag}(\boldsymbol{\Phi}\tilde{\boldsymbol{\Upsilon}}\boldsymbol{\Phi}^\mathsf{T})\|_1\right)} \\ &\leq 2\frac{\|\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})-\text{diag}(\boldsymbol{\Phi}(\tilde{\boldsymbol{\Upsilon}}-\boldsymbol{\Upsilon}+\boldsymbol{\Upsilon})\boldsymbol{\Phi}^\mathsf{T})\|_1}{m\|\boldsymbol{\Upsilon}\|_{\min}} \\ &\leq 2\frac{\|\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})-\text{diag}(\boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^\mathsf{T})\|_1+m\|\tilde{\boldsymbol{\Upsilon}}-\boldsymbol{\Upsilon}\|_{\max}}{m\|\boldsymbol{\Upsilon}\|_{\min}} \\ &\overset{(b)}{=} \mathcal{O}\left(\frac{\|\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})-\text{diag}(\boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^\mathsf{T})\|_1}{m}+\|\mathbf{D}-\frac{\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}}{m}\|_{\max}\right),\end{aligned} \tag{75}$$

where step (a) follows from similar logic as in Proposition 6; (b) is similar to a later elaborate derivation in Section B.5 which results in (94), except that $\hat{\mathbf{Q}}$ in (94) is replaced with $\mathbf{Q}$ (i.e. $\mathbf{L}_{\mathbf{Q}} = 0$) and $\hat{\mathbf{D}}$ in (94) is substituted by $\frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}$ (i.e. $\mathbf{L}_{\mathbf{D}} = \mathbf{D} - \frac{\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}}{m}$). The statement of the theorem follows with probability at least $1 - \delta_\circ - \delta - \tilde{\delta}$ (using union bound on the probability of joint events) from (71), (16), (73), and (75). To simplify the notations, we set $\delta_\circ = \tilde{\delta} = \delta$. The proof is now complete. □

B.4 Proof of Theorem 2: ℓ_1 error between the group controllability of the system with true and expected $\mathbf{A}$

Proof. We start by providing an upper bound on $\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\|_{\max}$ which will be shortly used to bound (16). Using the triangle inequality on matrix norm, we have:

$$\begin{aligned}\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\|_{\max} &\leq \|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \mathbf{1}_m\|_{\max} + \|\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}} - \mathbf{1}_m\|_{\max} \\ &\overset{(a)}{\leq} \|(\mathbf{I} - \mathbf{A}_{\text{nom}}^2)^{-1} - \mathbf{I}\|_{\max} + \|\frac{1}{n}\mathbf{P}^\mathsf{T}\text{diag}(\mathbf{\Upsilon})\|_{\max} \\ &\overset{(b)}{\leq} \sum_{\ell=1}^{\infty} \frac{\|(\mathbf{A}^2)^\ell\|_{\max}}{(\epsilon + \rho(\mathbf{A}))^{2\ell}} + \frac{1}{n}\|\text{diag}(\mathbf{\Upsilon})\|_{\max} \\ &\overset{(c)}{\leq} \sum_{\ell=1}^{\infty} (1+\eta)^\ell n^\ell \frac{f(\ell)}{\rho^{2\ell}(\mathbf{A})} + \frac{1}{n}\|\mathbf{\Upsilon}\|_{\max},\end{aligned} \tag{76}$$

where step (a) follows from (30) and (69); step (b) is because $\mathbf{P}$ is binary; and finally step (c) is due to (39) and defining $f(\ell) \triangleq \|\left(\mathbf{P}^\mathsf{T}\mathbf{Q}\mathbf{D}\mathbf{Q}\mathbf{P} + \text{Diag}[\mathbf{P}^\mathsf{T}(\mathbf{Q} \odot (\mathbf{1}_{K,K} - \mathbf{Q})\mathbf{d})]\right)^\ell\|_{\max}$. We, now, have $f(1) = \|\mathbf{Q}\mathbf{D}\mathbf{Q}\|_{\max} + \|\mathbf{Q}\mathbf{d}\|_{\max}$, and for all $\ell > 1$

$$\begin{aligned}f(\ell) &\overset{(a)}{\leq} \|\mathbf{P}^\mathsf{T}\mathbf{Q}\mathbf{D}\mathbf{Q}\mathbf{P} + \text{Diag}[\mathbf{P}^\mathsf{T}(\mathbf{Q} \odot (\mathbf{1}_{K,K} - \mathbf{Q})\mathbf{d})]\|_\infty \|\left(\mathbf{P}^\mathsf{T}\mathbf{Q}\mathbf{D}\mathbf{Q}\mathbf{P} + \text{Diag}[\mathbf{P}^\mathsf{T}(\mathbf{Q} \odot (\mathbf{1}_{K,K} - \mathbf{Q})\mathbf{d})]\right)^{\ell-1}\|_{\max} \\ &\overset{(b)}{\leq} (\|\mathbf{P}^\mathsf{T}\mathbf{Q}\mathbf{D}\mathbf{Q}\mathbf{P}\|_\infty + \|\mathbf{Q} \odot (\mathbf{1}_{K,K} - \mathbf{Q})\mathbf{d})\|_{\max}) f(\ell - 1) \\ &\overset{(c)}{\leq} (n\|\mathbf{Q}\mathbf{D}\mathbf{Q}\|_{\max} + \|\mathbf{Q}\mathbf{d}\|_{\max}) f(\ell - 1),\end{aligned} \tag{77}$$

where step (a) is due to item 4 in the matrix norm equivalence list at the beginning of the appendix; step (b) follows from the triangle inequality on matrix norms matrix, $\mathbf{P}$ being binary, and the definition of $f(\ell)$ prior to (77); and finally step (c) is because of item 5 of the list at the start of the appendix. Using (77), we upper bound the summation term in (76):

$$\begin{aligned}\sum_{\ell=1}^{\infty} (1+\eta)^\ell n^\ell \frac{f(\ell)}{\rho^{2\ell}(\mathbf{A})} &\overset{(a)}{\leq} \sum_{\ell=1}^{\infty} \left(\frac{1+\eta}{1-\eta}\right)^\ell n^\ell \frac{(n\|\mathbf{Q}\mathbf{D}\mathbf{Q}\|_{\max} + \|\mathbf{Q}\mathbf{d}\|_{\max})^\ell}{(n^2\rho^2(\mathbf{Q}\mathbf{D}) + n\|\mathbf{Q} \odot (\mathbf{1}_{K,K} - \mathbf{Q})\mathbf{d}\|_{\max})^\ell} \\ &= \sum_{\ell=1}^{\infty} \left(\frac{1+\eta}{1-\eta} \frac{\|\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ\|_{\max} + \frac{1}{n\rho_n}\|\mathbf{Q}_\circ\mathbf{d}\|_{\max}}{\rho^2(\mathbf{Q}_\circ\mathbf{D}) + \frac{1}{n\rho_n}\|\mathbf{Q}_\circ \odot (\mathbf{1}_{K,K} - \mathbf{Q})\mathbf{d}\|_{\max}}\right)^\ell,\end{aligned} \tag{78}$$

Using (78), we can further simplify (76) as

$$\begin{aligned}\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\|_{\max} &\leq \frac{\frac{1+\eta}{1-\eta}\frac{\|\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ\|_{\max} + \frac{1}{n\rho_n}\|\mathbf{Q}_\circ\mathbf{d}\|_{\max}}{\rho^2(\mathbf{Q}_\circ\mathbf{D}) + \frac{1}{n\rho_n}\|\mathbf{Q}_\circ \odot (\mathbf{1}_{K,K} - \mathbf{Q})\mathbf{d}\|_{\max}}}{1 - \frac{1+\eta}{1-\eta}\frac{\|\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ\|_{\max} + \frac{1}{n\rho_n}\|\mathbf{Q}_\circ\mathbf{d}\|_{\max}}{\rho^2(\mathbf{Q}_\circ\mathbf{D}) + \frac{1}{n\rho_n}\|\mathbf{Q}_\circ \odot (\mathbf{1}_{K,K} - \mathbf{Q})\mathbf{d}\|_{\max}}} + \frac{1}{n}\|\mathbf{\Upsilon}\|_{\max} \\ &\overset{(a)}{=} \mathcal{O}(1).\end{aligned} \tag{79}$$

(a) in (79) is because η is very small when n is very large, and $\|\mathbf{\Upsilon}\|_{\max}$ is a constant (c.f. (70)):

$$\begin{aligned}\|\mathbf{\Upsilon}\|_{\max} &= \left\|\mathbf{D}^{-\frac{1}{2}}\left(\left[\mathbf{I} - (\iota\mathbf{D}^{\frac{1}{2}}\mathbf{Q}_\circ\mathbf{D}^{\frac{1}{2}})^2\right]^{-1} - \mathbf{I}\right)\mathbf{D}^{-\frac{1}{2}}\right\|_{\max} \\ &= \mathcal{O}(1).\end{aligned} \tag{80}$$

We return to bounding $\Delta(\mathbf{A}, \bar{\mathbf{A}})$. Using (15) and Lemma 3, we obtain:

$$\begin{aligned}
\Delta(\mathbf{A}, \bar{\mathbf{A}}) &= \sum_{i=1}^{m} \left| \frac{r\boldsymbol{\theta}^{(i)}_{\text{group},\mathbf{A}} - 1}{\|r\boldsymbol{\theta}_{\text{group},\mathbf{A}} - \mathbf{1}_m\|_1} - \frac{r\boldsymbol{\theta}^{(i)}_{\text{group},\bar{\mathbf{A}}} - 1}{\|r\boldsymbol{\theta}_{\text{group},\bar{\mathbf{A}}} - \mathbf{1}_m\|_1} \right| \\
&= \sum_{i=1}^{m} \left| \frac{\mathbf{w}_i\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - 1}{\|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \mathbf{1}_m\|} - \frac{\mathbf{w}_i\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}} - 1}{\|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}} - \mathbf{1}_m\|_1} \right| \\
&\overset{(a)}{\leq} 2\frac{\|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \mathbf{W}\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\|_1}{\max\left(\|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - 1\|_1, \|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}} - 1\|_1\right)} \\
&\overset{(b)}{\leq} 2\frac{\frac{1}{r}\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\|_{1,(rm)}}{\max\left(\|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - 1\|_1, \|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}} - 1\|_1\right)} \\
&\overset{(c)}{\leq} 2\frac{\frac{1}{r}\left[\frac{mr}{n}\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\|_1 + \sqrt{\frac{mr\log\left(\frac{1}{\delta_\circ}\right)}{2}}\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\|_{\max}\right]}{\max\left(\|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - 1\|_1, \|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}} - 1\|_1\right)} \\
&= 2\frac{\left[\frac{m}{n}\alpha_n + \sqrt{\frac{m\log\left(\frac{1}{\delta_\circ}\right)}{2r}}\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\|_{\max}\right]}{\max\left(\|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - 1\|_1, \|\mathbf{W}\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}} - 1\|_1\right)} \\
&\overset{(d)}{\leq} 2\frac{\left[\frac{m}{n}\alpha_n + \sqrt{\frac{m\log\left(\frac{1}{\delta_\circ}\right)}{2r}}\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\|_{\max}\right]}{\frac{m}{n}\frac{1}{\frac{\epsilon}{\rho_n n} + \rho(\mathbf{Q}_\circ\mathbf{D})}\|\text{diag}(\boldsymbol{\Upsilon})\|_{\min}} \\
&\overset{(e)}{\leq} 2\frac{\left(\frac{\epsilon}{\rho_n n} + \rho(\mathbf{Q}_\circ\mathbf{D})\right)^2}{\|\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ\|_{\min}} \\
&\qquad \times \left[\alpha_n + \frac{1+\eta}{1-\eta}\sqrt{\frac{n}{mr}}\sqrt{\frac{\log\left(\frac{1}{\delta_\circ}\right)}{2}}\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\|_{\max}\right] \\
&\overset{(f)}{=} \mathcal{O}\left(\frac{\eta}{\rho_n} + \sqrt{\frac{n}{mr}}\sqrt{\frac{\log\left(\frac{1}{\delta_\circ}\right)}{2}}\right)
\end{aligned} \tag{81}$$

that holds with probability at least $(1-\delta)(1-\delta_\circ) \geq 1-\delta_\circ-\delta$ (c.f. probability of joint events), for sufficiently large n, i.e. there exists $n_0 : n \geq n_0$. (a) follows from Proposition 6 where $\mathbf{z}$ is replaced with $\mathbf{W}\boldsymbol{\theta}_{\text{fine},\mathbf{A}}$ and $\mathbf{y}$ with $\mathbf{W}\boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}$; (c) uses (79); (d) is from (20) which proves:

$$\begin{aligned}
\|r\boldsymbol{\theta}^{(i)}_{\text{group},\bar{\mathbf{A}}} - 1\|_1 &= \left\|\frac{1}{n}\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Upsilon})\right\|_1 \\
&\geq \frac{1}{n}\|\boldsymbol{\Phi}\mathbf{1}_K\|_1\|\boldsymbol{\Upsilon}\|_{\min} \geq \frac{m}{n}\|\boldsymbol{\Upsilon}\|_{\min} \\
&\overset{(g)}{=} \Omega\left(\frac{m}{n}\right),
\end{aligned} \tag{82}$$

where (g) in (82) follows from

$$\|\boldsymbol{\Upsilon}\|_{\min} = \left\|\mathbf{D}^{-\frac{1}{2}}\left(\left[\mathbf{I} - (\iota\mathbf{D}^{\frac{1}{2}}\mathbf{Q}_\circ\mathbf{D}^{\frac{1}{2}})^2\right]^{-1} - \mathbf{I}\right)\mathbf{D}^{-\frac{1}{2}}\right\|_{\min} = \Omega(1). \tag{83}$$

(e) uses the definition of $\boldsymbol{\Upsilon}$ in (13); (f) due to the definition of α_n in (17) and an easy-to-verify inequality $\frac{1+\eta}{1-\eta} \leq 1+4\eta$ for $\eta < 1/2$ that is satisfied for sufficiently large n; and finally, (b) is the result of the properties of the coarse-measurement matrix in (7), the following notation definition

$$\left\|\boldsymbol{\theta}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}_{\text{fine},\bar{\mathbf{A}}}\right\|_{1,(rm)} \triangleq \sum_{i \in \cup_{j\in[m]}\mathcal{K}_j} \left|\boldsymbol{\theta}^{(i)}_{\text{fine},\mathbf{A}} - \boldsymbol{\theta}^{(i)}_{\text{fine},\bar{\mathbf{A}}}\right|$$

for $\mathcal{K}_i = \text{supp}(\mathbf{w}_i)$ defined in Section 2.4, and invoking (85) which is explained below.

Sum of Random Samples from a set of Data Points: Using a Hoeffding's bound equivalence for the *sum of random samples of a set of data points* $\{\mathbf{z}_i\}_{i\in\mathcal{I}}$ indexed by $\mathcal{I}$, we have (Bardenet & Maillard, 2015, Proposition 1.2)

$$\begin{aligned}\mathbb{P}\{|\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}|\mathbf{z}_i| - \frac{1}{|\mathbf{z}|}\sum_{i\in[|\mathbf{z}|]}|\mathbf{z}_i|| \leq \eta_\circ\} &\geq 1 - \exp\left(-\frac{2|\mathcal{I}|\eta_\circ^2}{(\max(|\mathbf{z}|) - \min(|\mathbf{z}|))^2}\right) \\ &\geq 1 - \exp\left(-\frac{2|\mathcal{I}|\eta_\circ^2}{\max^2|\mathbf{z}|}\right)\end{aligned} \tag{84}$$

which means with probability at least $1 - \delta_\circ$

$$\begin{aligned}\|\mathbf{z}\|_{1,(rm)} &\leq \frac{|\mathcal{I}|}{|\mathbf{z}|}\|\mathbf{z}\|_1 + \sqrt{\frac{|\mathcal{I}|\log\left(\frac{1}{\delta_\circ}\right)}{2}}\max|\mathbf{z}| \\ &\leq \left[\frac{|\mathcal{I}|}{|\mathbf{z}|} + \sqrt{\frac{|\mathcal{I}|\log\left(\frac{1}{\delta_\circ}\right)}{2}}\frac{\|\mathbf{z}\|_{\max}}{\|\mathbf{z}\|_1}\right]\|\mathbf{z}\|_1.\end{aligned} \tag{85}$$

To simplify the notations, we set $\delta_\circ = \delta$. This concludes the proof. □

B.5 Proof of Theorem 3

Continuing the proof sketch started in the main text, in the following, we bound the second term in (24). Let $\mathbf{L}_{\mathbf{\Phi}} = \hat{\mathbf{\Phi}} - \mathbf{\Phi}$, $\mathbf{L}_{\mathbf{Q}} = \hat{\mathbf{Q}} - \mathbf{Q}$, and $\mathbf{L}_{\mathbf{D}} = \hat{\mathbf{D}} - \mathbf{D}$ be error matrices of appropriate sizes which we substitute into (25). We successively apply the triangle inequality, and we use Proposition 6 to obtain

$$\begin{aligned}\widehat{\Delta}(\widetilde{\mathbf{A}}, \bar{\mathbf{A}}) &\leq 2\frac{\||\hat{\mathbf{\Phi}}\text{diag}(\hat{\mathbf{\Upsilon}}) - \mathbf{\Phi}\text{diag}(\mathbf{\Upsilon})\|_1}{\max\left(\|\hat{\mathbf{\Phi}}\text{diag}(\hat{\mathbf{\Upsilon}})\|_1, \|\mathbf{\Phi}\text{diag}(\mathbf{\Upsilon})\|_1\right)} \\ &\leq 2\frac{\||(\mathbf{\Phi} + \mathbf{L}_{\mathbf{\Phi}})\text{diag}(\hat{\mathbf{\Upsilon}}) - \mathbf{\Phi}\text{diag}(\mathbf{\Upsilon})\|_1}{\|\mathbf{\Phi}\text{diag}(\mathbf{\Upsilon})\|_1} \\ &\leq 2\frac{\|\mathbf{\Phi}\text{diag}(\hat{\mathbf{\Upsilon}} - \mathbf{\Upsilon})\|_1 + \|\mathbf{L}_{\mathbf{\Phi}}\text{diag}(\hat{\mathbf{\Upsilon}})\|_1}{m\|\text{diag}(\mathbf{\Upsilon})\|_{\min}} \\ &\leq 2\frac{\|\hat{\mathbf{\Upsilon}} - \mathbf{\Upsilon}\|_{\max} + \|\hat{\mathbf{\Upsilon}}\|_{\max}\|\mathbf{L}_{\mathbf{\Phi}}\|_{1,1}/m}{\|\mathbf{\Upsilon}\|_{\min}}.\end{aligned} \tag{86}$$

We simplify $\|\hat{\mathbf{\Upsilon}} - \mathbf{\Upsilon}\|_{\max}$ in (86) using the two error matrices, $\mathbf{L}_{\mathbf{Q}}$ and $\mathbf{L}_{\mathbf{D}}$, introduced in the statement of the theorem:

$$\begin{aligned}(\mathbf{Q}_\circ + \mathbf{L}_{\mathbf{Q}})(\mathbf{D} + \mathbf{L}_{\mathbf{D}})(\mathbf{Q}_\circ + \mathbf{L}_{\mathbf{Q}}) &= \mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ + \mathbf{\Gamma}_1, \\ (\mathbf{D} + \mathbf{L}_{\mathbf{D}})(\mathbf{Q}_\circ + \mathbf{L}_{\mathbf{Q}}) &= \mathbf{D}\mathbf{Q}_\circ + \mathbf{\Gamma}_2, \\ (\mathbf{D}^{\frac{1}{2}} + \mathbf{L}_{\mathbf{D}}^{\frac{1}{2}})(\mathbf{Q} + \mathbf{L}_{\mathbf{Q}})(\mathbf{D}^{\frac{1}{2}} + \mathbf{L}_{\mathbf{D}}^{\frac{1}{2}}) &= \mathbf{D}^{\frac{1}{2}}\mathbf{Q}\mathbf{D}^{\frac{1}{2}} + \mathbf{\Gamma}_3, \\ (\mathbf{D}\mathbf{Q}_\circ + \mathbf{\Gamma}_2)^{2\ell} &= (\mathbf{D}\mathbf{Q}_\circ)^{2\ell} + \mathbf{\Gamma}_4.\end{aligned} \tag{87}$$

Let ℓ be a positive integer. Define the following:

$$\begin{aligned}
\mathbf{\Gamma}_1 &\triangleq \mathbf{Q}_\circ \mathbf{D}\mathbf{L}_\mathbf{Q} + \mathbf{Q}_\circ \mathbf{L}_\mathbf{D}\mathbf{Q}_\circ + \mathbf{Q}_\circ \mathbf{L}_\mathbf{D}\mathbf{L}_\mathbf{Q} + \mathbf{L}_\mathbf{Q}\mathbf{D}\mathbf{Q}_\circ \\
&\quad + \mathbf{L}_\mathbf{Q}\mathbf{D}\mathbf{L}_\mathbf{Q} + \mathbf{L}_\mathbf{Q}\mathbf{L}_\mathbf{D}\mathbf{Q}_\circ + \mathbf{L}_\mathbf{Q}\mathbf{L}_\mathbf{D}\mathbf{L}_\mathbf{Q}, \\
\mathbf{\Gamma}_2 &\triangleq \mathbf{D}\mathbf{L}_\mathbf{Q} + \mathbf{L}_\mathbf{D}\mathbf{Q}_\circ + \mathbf{L}_\mathbf{D}\mathbf{L}_\mathbf{Q} \\
\mathbf{\Gamma}_3 &\triangleq \mathbf{D}^{\frac{1}{2}}\mathbf{L}_\mathbf{Q}\mathbf{D}^{\frac{1}{2}} + \mathbf{D}^{\frac{1}{2}}\mathbf{Q}\mathbf{L}_\mathbf{D}^{\frac{1}{2}} + \mathbf{D}^{\frac{1}{2}}\mathbf{L}_\mathbf{Q}\mathbf{L}_\mathbf{D}^{\frac{1}{2}} \\
&\quad + \mathbf{L}_\mathbf{D}^{\frac{1}{2}}(\mathbf{Q}\mathbf{D}^{\frac{1}{2}} + \mathbf{L}_\mathbf{Q}\mathbf{D}^{\frac{1}{2}} + \mathbf{Q}\mathbf{L}_\mathbf{D}^{\frac{1}{2}} + \mathbf{L}_\mathbf{Q}\mathbf{L}_\mathbf{D}^{\frac{1}{2}}) \\
\mathbf{\Gamma}_4 &\triangleq (\mathbf{D}\mathbf{Q}_\circ)^{2\ell-1}\mathbf{\Gamma}_2 + \mathbf{\Gamma}_2(\mathbf{D}\mathbf{Q}_\circ)^{2\ell-1} + (\mathbf{D}\mathbf{Q}_\circ)^{2\ell-2}\mathbf{\Gamma}_2\mathbf{D}\mathbf{Q}_\circ \\
&\quad + (\mathbf{D}\mathbf{Q}_\circ)^{2\ell-2}\mathbf{\Gamma}_2^2 + \mathbf{\Gamma}_2(\mathbf{D}\mathbf{Q}_\circ)^{2\ell-1} + \mathbf{\Gamma}_2(\mathbf{D}\mathbf{Q}_\circ)^{2\ell-2}\mathbf{\Gamma}_2 + \cdots.
\end{aligned} \tag{88}$$

For any norm, $\|\mathbf{\Gamma}_3\| = \mathcal{O}(\|\mathbf{L}_\mathbf{Q}\| + \|\mathbf{L}_\mathbf{D}\|), \|\mathbf{\Gamma}_4\| = \mathcal{O}(\|\mathbf{\Gamma}_2\|)$ when $\|\mathbf{L}_\mathbf{D}\|, \|\mathbf{L}_\mathbf{Q}\| \ll 1$. From (87), we also have

$$|\rho(\mathbf{Q}\mathbf{D} + \mathbf{\Gamma}_3) - \rho(\mathbf{Q}\mathbf{D})| \overset{(a)}{\leq} \|\mathbf{\Gamma}_3\|_2 \overset{(b)}{\leq} K\|\mathbf{\Gamma}_3\|_{\max}, \tag{89}$$

where step (a) is due to Proposition 5, and step (b) because $\|\mathbf{Z}\|_2 \leq n\|\mathbf{Z}\|_{\max}$. Using the above identities, we simplify (86) as

$$\begin{aligned}
\|\hat{\mathbf{\Upsilon}} - \mathbf{\Upsilon}\|_{\max} &= \|\hat{\iota}^2\hat{\mathbf{Q}}\hat{\mathbf{D}}\hat{\mathbf{Q}}(\mathbf{I} - (\hat{\iota}\hat{\mathbf{D}}\hat{\mathbf{Q}})^2)^{-1} - \iota^2\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ(1 - (\iota\mathbf{D}\mathbf{Q}_\circ)^2)^{-1}\|_{\max} \\
&= \hat{\iota}^2\|(\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ + \mathbf{\Gamma}_1)(\mathbf{I} - (\hat{\iota}\hat{\mathbf{D}}\hat{\mathbf{Q}})^2)^{-1} - (\frac{\iota}{\hat{\iota}})^2\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ(1 - (\iota\mathbf{D}\mathbf{Q}_\circ)^2)^{-1}\|_{\max} \\
&= \hat{\iota}^2\|(\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ)[(\mathbf{I} - (\hat{\iota}\hat{\mathbf{D}}\hat{\mathbf{Q}})^2)^{-1} - (\mathbf{I} - (\iota\mathbf{D}\mathbf{Q}_\circ)^2)^{-1}] \\
&\quad + \mathbf{\Gamma}_1(\mathbf{I} - (\hat{\iota}\hat{\mathbf{D}}\hat{\mathbf{Q}})^2)^{-1} - \frac{\iota^2 - \hat{\iota}^2}{\hat{\iota}^2}\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ(\mathbf{I} - (\iota\mathbf{D}\mathbf{Q}_\circ)^2)^{-1}\|_{\max} \\
&\overset{(c)}{\leq} \Big[\|\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ\|_{\max}\|\mathbf{\Gamma}_5\|_{\max} + \|\mathbf{\Gamma}_1\|_{\max}\|(1 - (\iota\mathbf{D}\mathbf{Q}_\circ)^2)^{-1}\|_{\max} \\
&\quad + \left(2\rho(\mathbf{Q}\mathbf{D}) + \rho(\mathbf{\Gamma}_3) + \frac{2\epsilon}{n}\right)\frac{\|\mathbf{\Gamma}_3\|_2}{\left(\frac{\epsilon}{n} + \rho(\mathbf{Q}\mathbf{D})\right)^2} \\
&\quad \times \|\mathbf{Q}_\circ\mathbf{D}\mathbf{Q}_\circ\|_{\max}\|(\mathbf{I} - (\iota\mathbf{D}\mathbf{Q}_\circ)^2)^{-1}\|_{\max}\Big] K\hat{\iota}^2 \\
&\overset{(d)}{=} \mathcal{O}(\|\mathbf{\Gamma}_5\|_{\max} + \|\mathbf{\Gamma}_1\|_{\max} + \|\mathbf{\Gamma}_3\|_{\max}),
\end{aligned} \tag{90}$$

where $\mathbf{\Gamma}_5 \triangleq (\mathbf{I} - (\hat{\iota}\hat{\mathbf{D}}\hat{\mathbf{Q}})^2)^{-1} - (\mathbf{I} - (\iota\mathbf{D}\mathbf{Q}_\circ)^2)^{-1}$; (c) follows from the triangle inequality of matrix norms, multiple use of item 7 in the list at the beginning of the appendix, (13), (22), and inequality (a) in (89); and (d) is a consequence of inequality (b) in (89), $\mathbf{Q}_\circ, \mathbf{D}$ being constant values, and $\mathbf{\Gamma}_1, \mathbf{\Gamma}_3, \mathbf{\Gamma}_5 \ll 1$.

We simplify (90). First, we upper bound $\|\mathbf{\Gamma}_5\|_{\max}$ in (90) using Neumann series, (87), and the triangle inequality:

$$\begin{aligned}
\|\mathbf{\Gamma}_5\|_{\max} &= \left\|\textstyle\sum_{\ell\geq 1}[(\hat{\iota}\hat{\mathbf{D}}\hat{\mathbf{Q}})^{2\ell} - (\iota\mathbf{D}\mathbf{Q}_\circ)^{2\ell}]\right\|_{\max} \\
&= \left\|\textstyle\sum_{\ell\geq 1}\hat{\iota}^{2\ell}\mathbf{\Gamma}_4 - (\iota^{2\ell} - \hat{\iota}^{2\ell})(\mathbf{D}\mathbf{Q}_\circ)^{2\ell}\right\|_{\max} \\
&\leq \left\|\hat{\iota}^2(1 - \hat{\iota}^2)^{-1}\mathbf{\Gamma}_4\right\|_{\max} \\
&\quad + \left\|(\mathbf{D}\mathbf{Q}_\circ)^2[1 - (\mathbf{D}\mathbf{Q}_\circ)^2]^{-1}\right\|_{\max}\max_{\ell\geq 1}|\iota^{2\ell} - \hat{\iota}^{2\ell}| \\
&\overset{(a)}{=} \mathcal{O}\left(\|\mathbf{\Gamma}_4\|_{\max} + |\iota - \hat{\iota}|\right) \overset{(b)}{=} \mathcal{O}(\|\mathbf{\Gamma}_2\|_{\max} + \|\mathbf{\Gamma}_3\|_{\max}),
\end{aligned} \tag{91}$$

where (a) holds because $\iota, \hat{\iota}, \mathbf{D}, \mathbf{Q}_\circ = \mathcal{O}(1)$; (b) follows from (88) and (89) and the fact that

$$|\iota - \hat{\iota}| = \frac{|\rho(\mathbf{Q}\mathbf{D} + \mathbf{\Gamma}_3) - \rho(\mathbf{Q}\mathbf{D})|}{(\epsilon/(\rho_n n) + \rho(\mathbf{Q}_\circ\mathbf{D}))\left(\epsilon/(\rho_n n) + \rho(\hat{\mathbf{Q}}\hat{\mathbf{D}})\right)},$$

Similar to (91), $\|\mathbf{\Gamma}_1\|_{\max}$ and $\|\mathbf{\Gamma}_2\|_{\max}$ can be upper bounded as

$$\begin{aligned}\|\mathbf{\Gamma}_1\|_{\max} \leq & K^2(\|\mathbf{Q}_\circ\mathbf{D}\|_{\max}\|\mathbf{L_Q}\|_{\max} + \|\mathbf{Q}_\circ\|_{\max}\|\mathbf{L_D}\|_{\max}\|\mathbf{Q}\|_{\max} \\ & + \|\mathbf{Q}_\circ\|_{\max}\|\mathbf{L_D}\|_{\max}\|\mathbf{L_Q}\|_{\max} + \|\mathbf{L_Q}\|_{\max}\|\mathbf{D}\mathbf{Q}_\circ\|_{\max} \\ & + \|\mathbf{L_Q}\|_{\max}\|\mathbf{D}\|_{\max}\|\mathbf{L_Q}\|_{\max} + \|\mathbf{L_Q}\|_{\max}\|\mathbf{L_D}\|_{\max}\|\mathbf{Q}_\circ\|_{\max} + \|\mathbf{L_Q}\|_{\max}\|\mathbf{L_D}\|_{\max}\|\mathbf{L_Q}\|_{\max}) \\ \leq & K^2(\|\mathbf{L_Q}\|_{\max} + \|\mathbf{L_D}\|_{\max} + \|\mathbf{L_D}\|_{\max}\|\mathbf{L_Q}\|_{\max} + \|\mathbf{L_Q}\|_{\max} \\ & + \|\mathbf{L_Q}\|_{\max}^2 + \|\mathbf{L_Q}\|_{\max}\|\mathbf{L_D}\|_{\max} + \|\mathbf{L_Q}\|_{\max}^2\|\mathbf{L_D}\|_{\max}) = \mathcal{O}(\|\mathbf{L_Q}\|_{\max} + \|\mathbf{L_D}\|_{\max}),\end{aligned} \tag{92}$$

and

$$\begin{aligned}\|\mathbf{\Gamma}_2\|_{\max} &= \|\mathbf{D}\mathbf{L_Q} + \mathbf{L_D}\mathbf{Q}_\circ + \mathbf{L_D}\mathbf{L_Q}\|_{\max} \\ &\leq K(\|\mathbf{D}\|_{\max}\|\mathbf{L_Q}\|_{\max} + \|\mathbf{L_D}\|_{\max}\|\mathbf{Q}_\circ\|_{\max} + \|\mathbf{L_D}\|_{\max}\|\mathbf{L_Q}\|_{\max}) \\ &\overset{(b)}{\leq} K(\|\mathbf{L_Q}\|_{\max} + \|\mathbf{L_D}\|_{\max} + \|\mathbf{L_D}\|_{\max}\|\mathbf{L_Q}\|_{\max}) \\ &= \mathcal{O}(\|\mathbf{L_Q}\|_{\max} + \|\mathbf{L_D}\|_{\max}),\end{aligned} \tag{93}$$

where (b) in (93) is by the definitions of $\mathbf{D}$ and $\mathbf{Q}_\circ$ in Assumption 1, and the explanation below (4); and the scaling behaviour in the last lines of both in (92) and (93) follows when $\|\mathbf{L_Q}\|_{\max}, \|\mathbf{L_D}\|_{\max} \ll 1$. From (91), (92), and (93), $\|\hat{\mathbf{\Upsilon}} - \mathbf{\Upsilon}\|_{\max}$ of (90) becomes:

$$\|\hat{\mathbf{\Upsilon}} - \mathbf{\Upsilon}\|_{\max} = \mathcal{O}(\|\mathbf{L_Q}\|_{\max} + \|\mathbf{L_D}\|_{\max}), \tag{94}$$

From (94) and (86), we assert that

$$\widehat{\Delta}(\widetilde{\mathbf{A}}, \bar{\mathbf{A}}) = \mathcal{O}\left(\frac{\|\mathbf{L_\Phi}\|_{1,1}}{m} + \|\mathbf{L_Q}\|_{\max} + \|\mathbf{L_D}\|_{\max}\right), \tag{95}$$

where we used $\|\mathbf{\Upsilon}\|_{\min}{=}\Omega(1)$ and $\|\hat{\mathbf{\Upsilon}}\|_{\max}{=}\mathcal{O}(1)$ given in (13) and (22). The proof concludes by invoking (95) and (16) in (24). □