From Authority-Respect to Grassroots-Dissent: Degree-Weighted Learning and Convergence Speed

Chen Cheng[*], Xiao Han[†], Xin Tong[‡], Yusheng Wu[‡], and Yiqing Xing[§]

Abstract

Opinions are influenced by neighbors, with varying degrees of emphasis based on their connections. Some may value more connected neighbors' views due to authority-respect, while others might lean towards grassroots perspectives.The emergence of ChatGPT could signify a new "opinion leader" whose views people put a lot of weight on. This study introduces a degree-weighted DeGroot learning model to examine the effects of such belief updates on learning outcomes, especially the speed of belief convergence. We find that the effect of authority-respect is non-monotonic: greater respect for authority does not guarantee faster convergence. The convergence speed, influenced by increased authority-respect or grassroots-dissent, hinges on the unity of elite and grassroots factions. This research sheds light on the growing skepticism towards public figures and the ensuing dissonance in public debate.

Main

In the interconnected web of social networks, people's beliefs and opinions are intrinsically shaped by those around them[1, 2]. Not all voices carry the same weight: mainstream media, celebrities, and traditional authorities have historically held sway[3]. The digital age ushers in another layer of complexity with tools like ChatGPT, which blend the realms of technological advancement with societal influence, potentially ushering in a new echelon of opinion leaders[4]. Yet, the shifting sands of public trust, particularly post-COVID, hint at an evolving dynamic where traditional experts face increasing skepticism.[5] These developments necessitate

[*]Johns Hopkins University, Carey Business School
[†]University of Science and Technology of China
[‡]University of Southern California, Marshall School of Business
[§]Peking University, NSD

a theoretical framework that analyzes the implications of both “elite-respect” and “grassroots-dissent”[6], which is essential for understanding the outcomes of social learning amidst these evolving trends.

As beliefs evolve, the credence given to others’ views is often tied to their perceived influence and connectivity within their social network [7]. This phenomenon raises crucial questions: How does this weighting impact collective learning outcomes? Do beliefs converge faster when individuals respect authority more or when there is a prevalent grassroots-dissent? These questions are especially pertinent given the growing skepticism towards public figures and in a world where authority and grassroots voices clash [8].

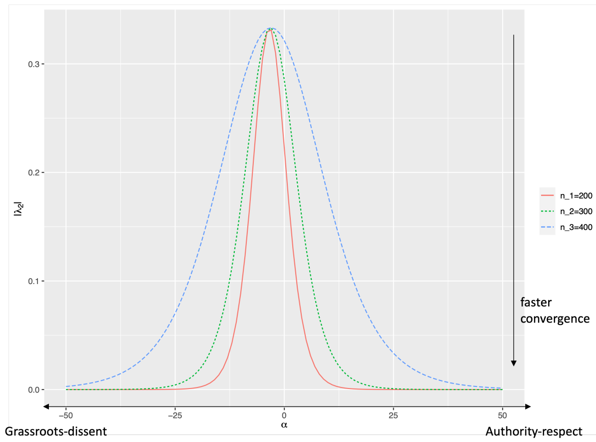
To navigate the intricate terrain of social influence, we introduce a degree-weighted learning heuristic to the classical DeGroot learning model[1, 9, 10, 11], in which agents iteratively refine their beliefs by considering a weighted average of their neighbors’ current opinions. The uniqueness of our approach lies in how the influence of a neighbor’s opinion is determined. We introduce a mechanism where the impact of each neighbor’s view changes based on their network prominence. This means that we can capture cases in which the opinions of more prominent neighbors hold greater sway, reflecting a respect for elite. Conversely, less influence is given to the views of prominent neighbors in other cases, indicating a tendency towards grassroots dissent. Hence, our model captures the full range of dynamics from a preference for established elite to a leaning towards grassroots perspectives.

Our findings reveal non-monotonic patterns in increasing authority-respect: a heightened deference to authority does not necessarily lead to quicker consensus. The speed at which beliefs converge, whether influenced by authority reverence or grassroots-dissent, pivots on the unity within elite and grassroots factions[12].

Specifically, in the baseline case with one elite and one grassroots group, the slowest convergence occurs when the influences of the elite and the grassroots are relatively balanced.

Amplifying either authority-respect or grassroots-dissent accelerates convergence (Figure 1).

Figure 1: One grassroots group and one elite group



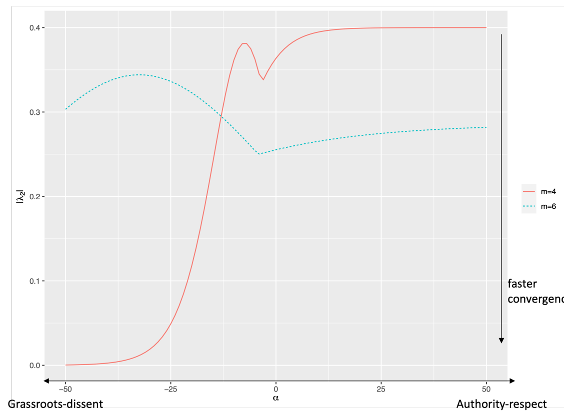
Notes: Graph of convergence speed versus the extent of degree-dependence. This graph shows that when there is one elite group and one grassroots group, either increasing authority-respect (by increasing α from a threshold to ∞) or enhancing grassroots-dissent (by decreasing α from a threshold to $-\infty$) can speed up convergence. Specifically, convergence speed is inversely related to the magnitude of the second largest eigenvalue of the learning matrix $\mathbf{T}^*$, denoted by $|\lambda_2(\mathbf{T}^*)|$, and the index α measures the degree-dependence of belief updating: a larger α indicates a greater influence of an individual's degree in the updating process. When $\alpha = 0$, individuals weigh their neighbors' opinions equally, regardless of their degrees. In the graph, the weighting function is defined as $\phi(\alpha, d) = d^\alpha$, where d is the degree of an individual. The size of the first group, n_1, is set to 200, 300, and 400, and the total population size n is 1000. The within-group linking probability p is 0.4, and the between-group linking probability q is 0.2.

When there is one grassroots group and multiple elite groups (Figure 2a), we witness phase transitions in the dominant tension hindering belief convergence. In Regime I, where elite groups are more influential, the predominant tension lies in *intra-group* conflicts among elites (e.g., elite polarization [12]). These conflicts within elite groups are the main factor slowing down convergence, with the presence of the grassroots group acting as a mediator to reconcile these elite viewpoints. This results in a slower convergence as elite influence becomes more pronounced. In Regime II, characterized by less dominant elite influence, the tension shifts to *inter-group* conflicts between the grassroots and elite groups. Slow convergence is most evident when both groups have similar levels of influence, and convergence speeds up as the grassroots group gains more influence.

Figure 2: One grassroots group and multiple elite groups

(a) 3 elite groups.

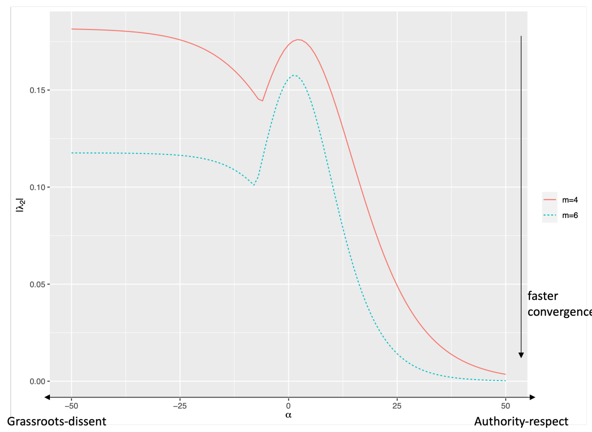
(b) 3 or 5 elite groups.



Notes: Graph of $|\lambda_2(\mathbf{T}^*)|$ versus α, with $\phi(\alpha, d) = d^\alpha$, with $\phi(\alpha, d) = d^\alpha$, $n_1 = 200$, $(m-1)n_2 = 1200$, $m = 4, 6$, $p = 0.3$, $q = 0.1$. When there is one grassroots group but multiple elite groups, increasing authority-respect (e.g., when $\alpha \to \infty$) can lead to slower convergence. As a counterpoint to the scenario with one elite group and multiple grassroots groups, when α is positive and large, the initial beliefs that lead to the slowest convergence are those in which elite groups hold differing views from one another. In this case, the grassroots group acts as an information mediator between the conflicting elite groups. Thus, lowering the weight placed on elite groups (decreasing α) results in a lower λ_2, signifying faster convergence. Conversely, when α is negative and large in magnitude, the initial beliefs that result in the slowest convergence are those where elite groups share identical views amongst themselves, yet these views differ from those of the grassroots group.

A phase transition is also observed when there is one elite group but multiple grassroots groups (Figure 3). When the grassroots groups are relatively more influential, it is the intra-group conflicts among these groups that matters more. Slow convergence is noted when different grassroots groups (for example, multiple minority cultural or ethnic groups) hold conflicting views. As an authority group and/or ChatGPT emerge and gain influence, convergence occurs more rapidly. This trend continues until the influence of the grassroots and the authority becomes roughly equal, marking a phase transition. Beyond this point, it is the inter-group conflict that predominantly slows convergence. In this later regime, convergence is slow when the influences of the two populations are relatively balanced and accelerates when the authority group becomes dominant.

Figure 3: Multiple grassroots groups and one elite group



Notes: Graph of $|\lambda_2(\mathbf{T}^*)|$ versus α, with $\phi(\alpha, d) = d^\alpha$, $m = 4, 6$, $n_1 = 400$, $(m-1)n_2 = 600$, $p = 0.5$, $q = 0.3$. When there is one elite group but multiple grassroots groups, increasing grassroots-dissent (e.g., as α approaches $-\infty$) can lead to slower convergence. When α is negative and large in magnitude, the initial beliefs that lead to the slowest convergence are those in which grassroots groups hold different views from each other. Conversely, when α is positive and large, the initial beliefs that result in the slowest convergence are those where grassroots groups share the same views with each other, but these views differ from those of the elite group.

This refined model bridges the traditional social learning models with contemporary societal nuances. Existing literature on social learning demonstrates convergence under very mild conditions (aperiodicity and strongly connectedness) [1, 13, 9]. Other studies, including this paper, are interested in how fast beliefs converge. Unlike most investigations that uncover the impacts of network structures[14, 10], our contributions stand as to emphasize the effects of a learning heuristic—authority-respect and grassroots-dissent—on convergence speed.

Results

To investigate the influence of authority-respect and grassroots-dissent, captured by degree-weighted learning heuristics, on belief convergence rates, we structure our exposition through a series of methodical steps. Initially, we integrate a degree-weighted learning heuristic into the established DeGroot learning framework. Subsequently, we delineate the metric for convergence speed within this augmented model. In the penultimate phase, we employ the stochastic block model to introduce a variable degree of heterogeneity among the agents. Finally, we demonstrate how varying degrees of deference to authority can significantly alter the rate at which convergence is achieved.

Our framework is built upon the classical DeGroot learning model [1], where an agent's belief—such as the probability of global warming or the safety of a COVID vaccine—is formed through a linear aggregation of the neighboring agents' beliefs from the previous period, as follows:

$$\mathbf{b}(t) = \mathbf{T}^t \mathbf{b}_0 \,,$$

where $\mathbf{b}(t)$ is the belief vector of n vertices at time $t = 0, 1, 2, \ldots$, $\mathbf{b}_0 \in \mathbb{R}^n$ with $\|\mathbf{b}_0\|_2 = 1$ is the initial belief vector, and $\mathbf{T}$, the learning matrix, is an $n \times n$ row-stochastic matrix. In this matrix, each element T_{ij} represents the weight that vertex i assigns to the belief of vertex j.

We generalize [10]'s setup which assumes equal-weighting of one's neighbors: given an undirected and unweighted network, captured by a $n \times n$ adjacency matrix $\mathbf{A}$, the learning matrix $\mathbf{T}$ is defined by the relation $T_{ij} = A_{ij}/d_i(\mathbf{A})$, where $d_i(\mathbf{A}) = \sum_j A_{ij}$ represents i's degree—that is, the number of connections or friends agent i possesses.

We introduce a new degree-weighted learning matrix $\mathbf{T}(\mathbf{A})$, as follows:

$$T_{ij}(\mathbf{A}) = \frac{A_{ij}\phi(\alpha, d_j(\mathbf{A}))}{\sum\limits_k A_{ik}\phi(\alpha, d_k(\mathbf{A}))} \,, \tag{1}$$

where $\phi : \mathbb{R}^2 \to \mathbb{R}$ is a function that modulates the influence of the degree of vertices, with a canonical example being $\phi(\alpha, d) = d^\alpha$. The index, $\alpha \in \mathbb{R}$, captures the dependence of the weight on one's degree: a larger α corresponds to larger weights to neighbors with higher degrees, or more authority-respect;[1] and a smaller α corresponds to more grassroots-dissent.[2] So that both dynamics are captured in the same framework.

To explore the degree-weighted learning heuristic, we adopt a stochastic block model, $\mathbf{A}(\mathbf{P}, \mathbf{n})$ ([15, 16, 17]), to generate the adjacency matrices. Besides its prevalence in network analysis ([18, 19, 10, 20]), the model is a natural choice to produce networks where vertices have heterogeneous degrees and to differentiate the elites and the grassroots.

Elite-Grassroots model. Particularly, we introduce a case referred to as the Elite-Grassroots model, consisting of n agents distributed into $m \geq 2$ groups, with each group comprising n_i agents for $i \in \{1, \cdots, m\}$. The linking probability matrix $\mathbf{P}$ is structured as follows:

$$\mathbf{P} = \begin{bmatrix} p & q & \dots & q \\ q & p & \dots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \dots & p \end{bmatrix}, \tag{2}$$

where p denotes the within-group linking probability, and q the between-group linking probability. The ratio of p to q serves as a measure of homophily ([10]). We make the following simplifying assumption regarding group sizes:

$$n_1 \neq n_2 = \ldots = n_m. \tag{A1}$$

Let d_k^* denote the expected degree of a vertex in group $k \in \{1, \ldots, m\}$, then $d_1^* = n_1 p + (m-1) n_2 q$ and $d_k^* = n_1 q + n_2 p + (m-2) n_2 q, k = 2, \ldots, m$.

[1] This corresponds to a common premise that vertices with higher degrees, such as celebrities, experts, or opinion leaders, are deemed more influential.

[2] This reflects a trend where, aided by technological advancements, the opinions of the broader community have become increasingly accessible and influential.

We term the group with the larger expected degree the *elite*, and the group with the smaller expected degree the *grassroots*. This simple structure captures the following three cases:

1. one grassroots group and one elite group (Figure 1), which captures the baseline patterns for authority-respect and grassroots-dissent;

2. one grassroots group and multiple elite groups (Figure 2), which captures the case where elites may represent different campaigns and/or hold conflicting views[12];

3. one elite group and multiple grassroots groups (Figure 3), which corresponds to the scenario with multiple (cultural, ethnic, etc.) groups whose initial beliefs may be divergent.

Now we present the formal results. Our first result confirms, in our context, the inverse relationship between convergence speed and $|\lambda_2|$ documented in the literature (e.g. [21, 22]).

Lemma 1 (Convergence Speed Inversely Related to $|\lambda_2|$)**.** *Given* $\mathbf{T}^\infty = \lim_{t\to\infty} \mathbf{T}^t$*,*[3] *at time* t*, the distance between the current belief and the limiting belief is*

$$\max_{\|\mathbf{b}_0\|_2=1} \|(\mathbf{T}^t - \mathbf{T}^\infty)\mathbf{b}_0\|_2 = |\lambda_2(\mathbf{T})|^t \,, \tag{3}$$

where $\lambda_2(\mathbf{T})$ *is the second largest eigenvalue in magnitude of* $\mathbf{T}$ *and* $\|\cdot\|_2$ *is the Euclidean norm. Therefore, convergence speed decreases in* $|\lambda_2|$.

With the largest eigenvalue of a row-stochastic matrix being 1, the expression on the right side of equation (3) diminishes to zero as t increases, and thus $|\lambda_2(\mathbf{T})|$ primarily determines the convergence speed.[4]

[3] We assume convergence, i.e., $\lim_{t\to\infty} \mathbf{T}^t$ exists, which only requires mild conditions (aperiodicity and strongly connectivity, see, e.g. [1] and [13]). Assumption 1 in Section ensures convergence with high probability.

[4] For an intuitive explanation of why the magnitude of the second largest eigenvalue correlates with convergence speed, interested readers may consult [22].

Results for "Expectation"–Replacing $\mathbf{A}$ with $\mathbb{E}[\mathbf{A}]$. With the foundational concepts established, our task is to determine the second largest eigenvalue (in magnitude) of $\mathbf{T}$. In a stochastic block model, $\mathbf{T}$ is inherently a random variable. Here, we present our results concerning the expectations of $\mathbf{T}$. In Section , we complete the analysis by demonstrating that the difference between a random network and its expected value becomes negligibly small for a sufficiently large n.

Recall that $\mathbf{A}(\mathbf{P}, \mathbf{n})$ denotes the random matrix generated by the stochastic block model. Now, define $\mathbf{R} = \mathbb{E}[\mathbf{A}(\mathbf{P}, \mathbf{n})]$, where $\mathbf{R}$ is the expected adjacency matrix, with $R_{ij} = P_{kl}$ when vertex i is in group k and vertex j is in group l. Let $\mathbf{T}^*$ denote the linear updating mechanism, which we define using $\mathbf{R}$ as follows:

$$T^*_{ij} = \frac{R_{ij}\, \phi(\alpha, d_j(\mathbf{R}))}{\sum\limits_j R_{ij}\, \phi(\alpha, d_j(\mathbf{R}))}\,, \tag{4}$$

where we define the expected degree $d_i(\mathbf{R})$ as follows: $d_i(\mathbf{R}) = \mathbb{E}[d_i(\mathbf{A})] = \sum_j R_{ij}$ for $i = 1, \ldots, n$. In the definition of $\mathbf{T}^*$, note that we replace all instances of $\mathbf{A}$ with $\mathbf{R}$ in the expression of $\mathbf{T}$ in (1).

We are now ready to fully characterize $|\lambda_2|$ and the impact of α on it.

Theorem 1 (Non-monotonic Impact of Degree-Dependence, α)**.** *Let $\mathcal{D}_\alpha$ be the region such that $|\lambda_2(\mathbf{T}^*)|$ is decreasing in α for $\alpha \in \mathcal{D}_\alpha$ and increasing for $\alpha \notin \mathcal{D}_\alpha$. Then $\mathcal{D}_\alpha$ is characterized as follows:*

$$\mathcal{D}_\alpha = \begin{cases} \left[g^{-1}\left(\frac{n_1}{n_2}\right), \infty\right),^5 & \textit{if } m = 2 \\ (-\infty, g^{-1}(n_1/n_2)] \cup \left[g^{-1}\left(\frac{n_1}{n_2}\left(\frac{p}{(m-1)(p+(m-2)q)}\right)^{1/2}\right), \infty\right), & \textit{if } m \geq 3, d_1^* > d_2^* \\ \left(g^{-1}\left(\frac{n_1}{n_2}\left(\frac{p}{(m-1)(p+(m-2)q)}\right)^{1/2}\right), g^{-1}(n_1/n_2)\right], & \textit{if } m \geq 3, d_1^* < d_2^* \end{cases} \tag{5}$$

In other words, when there is only one elite group and one grassroots group (Case 1, Figure 1), an increase in α beyond $g^{-1}(\frac{n_1}{n_2})$ as well as a decrease in α below this threshold can both lead to faster belief convergence. In the scenario with one grassroots group and multiple elite groups (Case 2, Figure 2), a lower α facilitates faster convergence, while a higher α results in slower convergence. Conversely, with one elite group and multiple grassroots groups (Case 3, Figure 3), a higher α accelerates belief convergence, whereas a lower α decelerates. In addition, Figure 5 compares Cases 1 and 2 and Figure 4 Cases 2 and 3.

Recall that the above patterns are presented based on the expected matrix, $\mathbf{R}$. In Section , we complete the analysis by demonstrating that the difference between a random network and its expected value becomes negligibly small for a sufficiently large n.

Discussion

Understanding regime changes using the worst-case initial beliefs.

Our results reveal that simply amplifying either authority-respect or grassroots-dissent does not necessarily lead to faster convergence. To understand why, it is crucial to examine the regime changes in each scenario. This involves investigating the worst-case beliefs – the initial beliefs that result in the slowest convergence – which highlight the primary tensions impeding convergence in each case.

In the scenario with one elite and one grassroots group (Case 1, Figure 1), the worst-case occurs when these groups hold divergent initial beliefs (e.g., 1's for one group and 0's for the other). As α approaches $-\infty$, agents predominantly consider grassroots opinions, leading to grassroots dominance; conversely, as α approaches $+\infty$ with greater trust in authorities, the elite group becomes dominant. As α shifts from $-\infty$ to $g^{-1}(\frac{n_1}{n_2})$, society transitions from a single dominant viewpoint to dual viewpoints – one from the elite and one from the grassroots – which slows down belief convergence as α increases in this range. Further increase in α eventually reverts the society back to a single dominant viewpoint, thereby accelerating convergence.

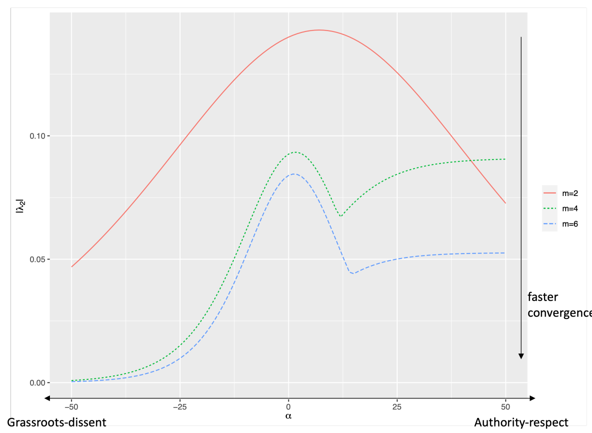
In the case of one grassroots group and multiple elite groups (Case 2, Figure 2a), different dynamics emerge. In regime I, where α is large enough for elites to be more influential, the worst-case involves conflicting views within different elite groups. These intra-group conflicts significantly slow down convergence, with the grassroots acting as mediators that harmonize these conflicts, resulting in decreased convergence speed as α increases. In regime II, with a smaller α, inter-group conflicts between the grassroots and elite groups drive slower convergence. Here, convergence is slowest when both populations are similarly influential and speeds up as the grassroots become more dominant.

Understanding the dynamics in Case 2 simplifies the comprehension of Case 3. When grassroots groups are more influential (when α is negative but large in maganitude), slow convergence is observed when different grassroots groups (e.g., those from distinct cultural or ethnic backgrounds) hold conflicting views. The emergence and increasing influence of an elite group or ChatGPT (a positve and large α) enhance social learning effectiveness. This continues until the influences of the grassroots and the elite roughly equalize, marking a phase transition where the main tension becomes the one between these two populations. In this final regime, convergence is slower when the influence of both populations is balanced and speeds up as the elite group gains dominance.

Takeaway. Our study sheds light on the influence of skepticism towards experts and the discord in public discourse. In society's complex discourse, our views are shaped by various influences, from authorities to grassroots movements and opinion leaders. The weight we give to each voice can significantly affect the persistence of societal disagreements. The key takeaway is that a pronounced respect for authority does not uniformly speed up consensus. The rate of belief convergence, influenced by deference to authority or grassroots-dissent, depends on the unity within the elite and grassroots factions.

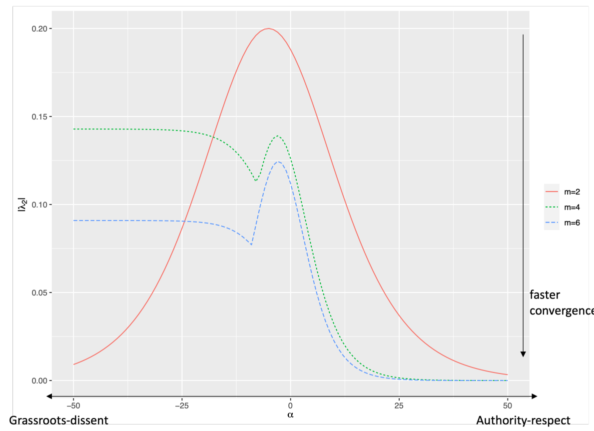
Limitations and Future Work. Our study employs an agent's network of connections, or "degree," as a surrogate for attributes like influence, fame, and popularity. This degree-dependent method in our updating matrix captures essential aspects of social learning. However, there are potential alternative dimensions to consider. For instance, during information updates, the weighting matrix might be based on the precision of an agent's signal rather than their neighbors' degree. While this presents a compelling avenue of exploration, a thorough analysis of such dimensions is reserved for future research.

Figure 4: One grassroots group + one vs. multiple elite groups



Notes: Graph of $|\lambda_2(\mathbf{T}^*)|$ vs. α, with $\phi(\alpha, d) = d^\alpha$, $n_1 = 800$, $(m-1)n_2 = 600$, $p = 0.3$, $q = 0.4$ and $m = 2, 4, 6$. This graph shows that when there is one grassroots group but multiple elite groups ($m = 4, 6$, green-dotted or blue-dashed), unlike the case with one elite group ($m = 2$, red-solid), increasing authority-respect (e.g., when $\alpha \to \infty$) can lead to slower convergence.

Figure 5: One vs. multiple grassroots groups + one elite group



Notes: Graph of $|\lambda_2(\mathbf{T}^*)|$ vs. α, with $\phi(\alpha, d) = d^{\alpha}$, $n_1 = 500$, $n_2 = 300$, $p = 0.3$, $q = 0.2$ and $m = 2, 4, 6$. This graph shows that when there is one elite group but multiple grassroots groups ($m = 4, 6$, green-dotted or blue-dashed), unlike the case with one grassroots group ($m = 2$, red-solid), increasing grassroots-dissent (e.g., when $\alpha \to -\infty$) can cause slower convergence.

Materials and Methods

We showed our results using an example of the degree-dependent function $\phi(d) = d^{\alpha}$. All results hold for general ϕ as long as it satisfies the following properties:

Property 1. *$\phi(\alpha, d) \in \mathcal{C}^2(\mathbb{R} \times [0, \infty))$ is nonnegative and $\phi(0, d) \equiv 1$ for all $d \in [0, \infty)$. In addition, ϕ is monotonically increasing in α for $\alpha \in \mathbb{R}$, monotonically increasing in d for $\alpha \in (0, \infty)$ and monotonically decreasing in d for $\alpha \in (-\infty, 0)$.*

Property 2. *For any two degrees $d_1 > d_2$, the ratio $\phi(\alpha, d_2)/\phi(\alpha, d_1)$ is strictly decreasing in α:*

$$\frac{d}{d\alpha}\left(\frac{\phi(\alpha, d_2)}{\phi(\alpha, d_1)}\right) < 0\,. \tag{6}$$

In addition, $\lim\limits_{\alpha \to \infty} \frac{\phi(\alpha,d_2)}{\phi(\alpha,d_1)} = \lim\limits_{\alpha \to -\infty} \frac{\phi(\alpha,d_1)}{\phi(\alpha,d_2)} = 0$.

Property 3. *ϕ satisfies*

$$\limsup_{d\to\infty} \frac{\partial\phi(\alpha,d)/\partial d}{\phi(\alpha,d)/d} < \infty\,, \tag{7}$$

and

$$\limsup_{d\to\infty} \frac{\partial^2\phi(\alpha,d)/\partial d^2}{(\partial\phi(\alpha,d)/\partial d)/d} < \infty\,. \tag{8}$$

Properties 1 and 2 are the very basic requirements for ϕ in order for our discussion to be meaningful. Property 3 consists of two technical conditions needed for the concentration result that we derive later in Lemma 2 for random networks. Roughly speaking, Property 3 says that we do not want a tiny change in the degree d to cause a huge increase in the ϕ.

Fully characterizing λ_2

Theorem 1 relies on the following result which fully characterizes the second largest eigenvalue in magnitude of learning matrix $\mathbf{T}^*$.

Proposition 1 (λ_2 in Elite-Grassroots Model)**.** *Recall that $d_1^* = n_1p + (m-1)n_2q$ and $d_2^* = n_1q + n_2p + (m-2)n_2q$ be the expected degrees and let $g : \mathbb{R} \to \mathbb{R}$ be the function:*

$$g(\alpha) = \frac{\phi(\alpha, d_2^*)}{\phi(\alpha, d_1^*)}, \tag{9}$$

where ϕ satisfy Properties 1 - 3. Then under assumptions of linking probabilities and group sizes in (2) and (A1) we have:

When $m = 2$ (Case (1)), the second largest eigenvalue (in magnitude) is:

$$|\lambda_2(\mathbf{T}^*)| = \left|\frac{n_1p\,\phi(\alpha,d_1^*)}{n_1p\,\phi(\alpha,d_1^*) + n_2q\,\phi(\alpha,d_2^*)} - \frac{n_1q\,\phi(\alpha,d_1^*)}{n_1q\,\phi(\alpha,d_1^*) + n_2p\phi(\alpha,d_2^*)}\right| \tag{10}$$

When $m \geq 3$ and $d_1^* > d_2^*$ (Case (2)), we have:[6]

$$|\lambda_2(\mathbf{T}^*)| = \begin{cases} \left|\frac{n_1p\,\phi(\alpha,d_1^*)}{n_1p\,\phi(\alpha,d_1^*)+(m-1)n_2q\,\phi(\alpha,d_2^*)} - \frac{n_1q\,\phi(\alpha,d_1^*)}{n_1q\,\phi(\alpha,d_1^*)+(n_2p+(m-2)n_2q)\phi(\alpha,d_2^*)}\right|, & \alpha \geq g^{-1}(n_1/n_2) \\ \left|\frac{n_2(p-q)\,\phi(\alpha,d_2^*)}{n_1q\,\phi(\alpha,d_1^*)+(n_2p+(m-2)n_2q)\phi(\alpha,d_2^*)}\right|, & \alpha < g^{-1}(n_1/n_2) \end{cases} \tag{11}$$

[6]Note that the existence of the inverse g is guaranteed: property 2 of ϕ implies that $\frac{\phi(\alpha,d_2^*)}{\phi(\alpha,d_1^*)}$ is strictly monotone and thus invertible.

When $m \geq 3$ and $d_1^* < d_2^*$ (Case (3)), the second largest (in magnitude) eigenvalue is:

$$|\lambda_2(\mathbf{T}^*)| = \begin{cases} \Big|\frac{n_1 p\,\phi(\alpha,d_1^*)}{n_1 p\,\phi(\alpha,d_1^*)+(m-1)n_2 q\,\phi(\alpha,d_2^*)} \\ \quad -\frac{n_1 q\,\phi(\alpha,d_1^*)}{n_1 q\,\phi(\alpha,d_1^*)+(n_2 p+(m-2)n_2 q)\phi(\alpha,d_2^*)}\Big|, & \alpha \leq g^{-1}(n_1/n_2) \\ \left|\frac{n_2(p-q)\,\phi(\alpha,d_2^*)}{n_1 q\,\phi(\alpha,d_1^*)+(n_2 p+(m-2)n_2 q)\phi(\alpha,d_2^*)}\right|, & \alpha > g^{-1}(n_1/n_2) \end{cases} \tag{12}$$

Worst initial beliefs

It can be shown that the second eigenvector of $\mathbf{T}^*$ takes the following form:

$$\mathbf{v}_2 = (\underbrace{v_{21},\ldots,v_{21}}_{n_1},\underbrace{v_{22},\ldots,v_{22}}_{n_2},\ldots,\underbrace{v_{2m},\ldots,v_{2m}}_{n_2})^\top.$$

Note here that since agents within the same group are identical ex ante, their beliefs quickly converge after one period of updating. As a result, in expectation, agents in the same group hold identical views. Only beliefs across groups could differ.

Then we have for case 1, i.e., $m = 2$:

$$v_{21} = -\frac{b}{en_2 + n_1 b^2/e},\ v_{22} = \frac{1}{n_2 + n_1 b^2/e^2}\,.$$

In other words, when we have only two groups of agents, for each α, the initial beliefs that lead to slowest convergence is the one that elite group holds a different view than the grassroots group.

For case 2, when $m \geq 3$ and there is one elite group and multiple grassroots groups: if $\alpha > g^{-1}(\frac{n_1}{n_2})$, the worst initial beliefs are that elite group agents share the same initial beliefs, and *all* the grassroots groups share the same beliefs. Whereas when $\alpha < g^{-1}(\frac{n_1}{n_2})$, the worst initial beliefs are that elite group agents share the same belief, but the grassroots groups hold *different* opinions from each other (i.e., grassroots groups 2 to m holds different initial views).

Moving to case 3 in which there is one grassroots group and multiple elite groups, if $\alpha > g^{-1}(\frac{n_1}{n_2})$, the worst initial beliefs are such that different elite groups hold different views, and grassroots agents hold the same view; and when $\alpha < g^{-1}(\frac{n_1}{n_2})$, the worst initial beliefs are such different elite groups holds the same view, but the view is different from grassroots view. More formally, the following proposition summarizes the above results about worst initial beliefs.

Proposition 2. *Denote the second eigenvector of* $\mathbf{T}^*$ *by*

$$\mathbf{v}_2 = (\underbrace{v_{21}, \ldots, v_{21}}_{n_1}, \underbrace{v_{22}, \ldots, v_{22}}_{n_2}, \ldots, \underbrace{v_{2m}, \ldots, v_{2m}}_{n_2})^\top.$$

For $d_1^* > d_2^*$*, when* $\alpha > g^{-1}(\frac{n_1}{n_2})$*, we have*

$$v_{21} = -\frac{(m-1)b}{e(m-1)n_2 + n_1(m-1)^2b^2/e},\; v_{22} = \ldots = v_{2m} = \frac{1}{(m-1)n_2 + n_1(m-1)^2b^2/e^2}\,.$$

When $\alpha < g^{-1}(\frac{n_1}{n_2})$*, let* $\mathbf{v}' = (\underbrace{v_1', \ldots, v_1'}_{n_1}, \underbrace{v_2', \ldots, v_2'}_{n_2}, \ldots, \underbrace{v_m', \ldots, v_m'}_{n_2})^\top$*, where*

$$v_1' = -\frac{(m-1)b}{e(m-1)n_2 + n_1(m-1)^2b^2/e},\; v_2' = \ldots = v_m' = \frac{1}{(m-1)n_2 + n_1(m-1)^2b^2/e^2}\,.$$

then $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{-1/2}\mathbf{v}_2$ *is any vector orthogonal to* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{v}'$ *and* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{1}$*, where* $\mathbf{D}_{1,\mathbf{R}}$ *and* $\mathbf{D}_{2,\mathbf{R}}$ *are diagonal matrices defined by:*

$$(\mathbf{D}_{1,\mathbf{R}})_{ii} = \sum_j R_{ij}\,\phi(\alpha, d_j(\mathbf{R}))\,,$$
$$(\mathbf{D}_{2,\mathbf{R}})_{ii} = \phi(\alpha, d_i(\mathbf{R}))\,,$$

in which $\mathbf{T}^* = \mathbf{D}_{1,\mathbf{R}}^{-1}\mathbf{R}\mathbf{D}_{2,\mathbf{R}}$ *, and* $\mathbf{R} = \mathbb{E}\mathbf{A}$*.*

When $\alpha = g^{-1}(\frac{n_1}{n_2})$*,* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{-1/2}\mathbf{v}_2$ *is any vector orthogonal to* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{1}$*.*

In contrast if $d_1^* < d_2^*$*, when* $\alpha < g^{-1}(\frac{n_1}{n_2})$*, we have*

$$v_{21} = -\frac{(m-1)b}{e(m-1)n_2 + n_1(m-1)^2b^2/e},\; v_{22} = \ldots = v_{2m} = \frac{1}{(m-1)n_2 + n_1(m-1)^2b^2/e^2}\,.$$

When $\alpha > g^{-1}(\frac{n_1}{n_2})$*, then* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{-1/2}\mathbf{v}_2$ *is any vector orthogonal to* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{v}'$ *and* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{1}$*.*

When $\alpha = g^{-1}(\frac{n_1}{n_2})$*,* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{-1/2}\mathbf{v}_2$ *is any vector orthogonal to* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{1}$*.*

For $m = 2$*, the second eigenvector is*

$$v_{21} = -\frac{b}{en_2 + n_1b^2/e},\; v_{22} = \frac{1}{n_2 + n_1b^2/e^2}\,.$$

Results for random learning matrix

Section summarizes our findings about convergence speed in the case of "expectation." To see the full picture, we need to show that the result for the random case is arbitrarily "close" to the "expectation" when n is sufficiently large. To address this issue, we first make a few assumptions:

Assumption 1 (Density). *Let* $\tau_n = \min_i \frac{d_i(\boldsymbol{R})}{n}$, $\tilde{\tau}_n = \max_i \frac{d_i(\boldsymbol{R})}{n}$, *where* $\mathbf{R} = \mathbb{E}\mathbf{A}$. *Assume that*

$$\lim_{n\to\infty} \frac{\tau_n}{\sqrt{\frac{\tilde{\tau}_n \log n}{n}}} = \infty \,. \tag{13}$$

Assumption 2 (No Vanishing Groups). *For all* $k = 1, \ldots, m$*:*

$$\liminf_n \frac{n_k}{n} > 0 \,. \tag{14}$$

Assumption 3 (Comparable Densities). [7]

$$\limsup_n \frac{p_n}{q_n} < \infty \,. \tag{15}$$

We note first that with the minimum density assumed (Assumption 1), the network is connected with high probability. Restricting to this high probability event, the directed graph $\mathcal{G}(\mathbf{T})$ corresponding to the matrix $\mathbf{T}$ is strongly connected. In addition, $\mathbf{T}$ is aperiodic with high probability. With the known fact that strongly connectivity and aperiodicity together implies convergence (see [13]), we see that the limit exists with high probability. We characterize what limit beliefs look like in the following proposition.

Proposition 3 (Consensus). *The limit* $\mathbf{T}^\infty$ *as* $t \to \infty$ *is given by:*

$$\mathbf{T}^\infty = \begin{bmatrix} T_1 & T_2 & \ldots & T_n \\ T_1 & T_2 & \ldots & T_n \\ \vdots & \vdots & \ddots & \vdots \\ T_1 & T_2 & \ldots & T_n \end{bmatrix} ,$$

[7] Here, we consider increasing n, so p, q in (2) are with subscript n.

where

$$T_j = \frac{\sum_i A_{ij}\, \phi(\alpha, d_j(\mathbf{A}))\phi(\alpha, d_i(\mathbf{A}))}{\sum_{i,j} A_{ij}\, \phi(\alpha, d_j(\mathbf{A}))\phi(\alpha, d_i(\mathbf{A}))},$$

for $j = 1, \ldots, n$. *This means that the limiting beliefs* $\mathbf{b}(\infty) = \mathbf{T}^{\infty}\mathbf{b}_0$ *would shift towards the beliefs of high degree neighbors as* α *increases.*[8]

With Assumptions 1 - 3, we achieve our main technical Lemma:

Lemma 2. *Suppose* $\mathbf{A}(\mathbf{P}, \mathbf{n})$ *is generated by the stochastic block model with* $\mathbf{P}$ *and* $\mathbf{n}$*. In addition, suppose Assumptions 1, 2, and 3 hold. Then, there exists a positive constant*[9] $\tilde{C}$*, independent of* n*, such that for all* $n > 0$*,*

$$\mathbb{P}\left\{ |\lambda_2(\mathbf{T}) - \lambda_2(\mathbf{T}^*)| \geq \tilde{C}\frac{\sqrt{\tilde{\tau}_n \log n}}{\tau_n \sqrt{n}} \right\} \leq \frac{16}{n^2}.$$

Note for this lemma, we do not require the structure assumed in elite-grassroots model (i.e., we do not need Assumption A1). By Lemma 2, we see that the error gets arbitrarily small when n goes to infinity. The remaining part is that although the monotonicity in the case "expectation" does not happen in the case of random networks, we can still find out how much increase in α could give a decrease in the eigenvalue that exceeds the error created by randomness. Given an α_0, suppose α rises from α_0 to α_1, then how large do α_1 need to be to see an increase in the the convergence speed with high probability? We answer this question by applying Lemma 2 in the following theorem.

Theorem 2. *Given a constant* α_0*. Let* α_1 *be bounded away from* ∞*. If* $\alpha_0, \alpha_1 \in \mathcal{D}_\alpha$ *as defined in Theorem 1 and*

$$\alpha_1 - \alpha_0 > \hat{C}\frac{\sqrt{\tilde{\tau}_n \log n}}{\tau_n \sqrt{n}} \left[\frac{\partial\phi/\partial\alpha(\alpha_1, d_1)}{\phi(\alpha_1, d_1)} - \frac{\partial\phi/\partial\alpha(\alpha_1, d_2)}{\phi(\alpha_1, d_2)} \right]^{-1}, \tag{16}$$

[8] Of course, if all degrees $d_j(\mathbf{A})$ are equal for all $j = 1, \ldots, n$, then $\mathbf{b}(\infty)$ is independent of α, but this happens with negligible probability if $\mathbf{A}$ is generated randomly as in stochastic block model.

[9] The constant $\tilde{C}$ may depend on α and ϕ.

where $\hat{C} > 0$ is a constant[10] *independent of* n *, then there exists an integer* $n_0 \in \mathbb{N}$*, such that for all* $n > n_0$*,*

$$\mathbb{P}\left\{|\lambda_2(\mathbf{T}(\alpha_0))| - |\lambda_2(\mathbf{T}(\alpha_1))| > 0\right\} \geq 1 - \frac{16}{n^2}\,.$$

Remark 1. *The condition in (16) may seem complicated but it is not restrictive at all: the right hand side of the inequality approaches zero as* n *goes to* ∞*. This condition is meant to describe the minimum size of* $\alpha_1 - \alpha_0$*.*

This theorem gives the size of increase from α_0 to α_1 such that the change in the magnitude of the eigenvalue would be greater than the error, whose size is given by Lemma 2. Therefore, Theorem 2 tells us that if α increases from α_0 to α_1 that satisfy the conditions in Theorem 2, then we will see a decrease in $|\lambda_2(\mathbf{T})|$ and thus an increase in convergence speed with a probability tending to one as n goes to ∞.

Extension: Perturbation of Adjacency Matrix

We can further examine what happens if we allow for perturbation of the network structure. A perturbation term is added to each entry of the adjacency matrix $\mathbf{A}$ and the impact of this on Theorem 2 is investigated.[11] More specifically, for $i, j = 1, 2, \ldots, n$, let ϵ_{ij} be independent and identically distributed random variables supported on $[0, 1]$. Then, let the perturbed adjancency matrix $\tilde{\mathbf{A}}$ be defined entrywise by:

$$\tilde{A}_{ij} = (1-\delta)A_{ij} + \delta\epsilon_{ij}\,,$$

for $\delta \in [0, 1]$. Replacing $\mathbf{A}$ by $\tilde{\mathbf{A}}$ in (1), we define the perturbed weight matrix $\tilde{\mathbf{T}}$ by:

$$\tilde{T}_{ij} = \frac{\tilde{A}_{ij}\phi(\alpha, d_j(\tilde{\mathbf{A}}))}{\sum_j \tilde{A}_{ij}\phi(\alpha, d_j(\tilde{\mathbf{A}}))}\,.$$

[10] $\hat{C}$ depends on $\limsup_n p_n/q_n$, $\limsup_n n_2/n_1$, $\limsup_n \phi(\alpha_1, d_2)/\phi(\alpha_1, d_1)$ and the constant $\tilde{C}$ in Lemma 2.

[11] Ideally, we would add the perturbation to the matrix $\mathbf{T}$ but the structure of the row-stochastic matrix can easily be broken in this way. Therefore, we pass the perturbation to the adjancency matrix $\mathbf{A}$ for theoretical simplicity.

Similarly, let $\tilde{\mathbf{R}} = \mathbb{E}\tilde{\mathbf{A}}$ and define the deterministic matrix $\tilde{\mathbf{T}}^*$ by:

$$\tilde{T}^*_{ij} = \frac{\tilde{R}_{ij}\phi(\alpha, d_j(\tilde{\mathbf{R}}))}{\sum_j \tilde{R}_{ij}\phi(\alpha, d_j(\tilde{\mathbf{R}}))} .$$

Then, we have the following Corollary:

Proposition 4. *Theorem 1 and Theorem 2 hold for the perturbed random matrix* $\tilde{\mathbf{T}}$.

It is clear that the extra perturbation term does not change the structure assumed in Theorem 1. The concentration result of Lemma 2 also holds since the perturbation ϵ_{ij} is bounded so that the concentration inequalities as our main tools for proving of Lemma 2 are unaffected. The result of Theorem 2 follows.

References

[1] DeGroot, M. H. Reaching a consensus. *Journal of the American Statistical Association* **69**, 118–121 (1974).

[2] Golub, B. & Sadler, E. Learning in social networks. In Bramoullé, A. G., Yann & Rogers, B. W. (eds.) *The Oxford Handbook of the Economics of Networks* (Oxford Academic, 2016).

[3] Katz, E., Lazarsfeld, P. F. & Roper, E. Personal influence: The part played by people in the flow of mass communications. In Breiger R, P. P., Carley K (ed.) *Dynamic Social Network Modeling and Analysis*, 162—-173 (National Academies Press, Washington DC, 2003).

[4] Anderson, J. & Rainie, L. The future of truth and misinformation online. Tech. Rep., Pew Research Center (2017).

[5] Kennedy, B., Tyson, A. & Funk, C. Americans' trust in scientists, other groups declines. Tech. Rep., Pew Research Center (2022).

[6] Burt, R. S. Social contagion and innovation: Cohesion versus structural equivalence. *American journal of Sociology* **92**, 1287–1335 (1987).

[7] Centola, D. *How behavior spreads: The science of complex contagions*, vol. 3 (Princeton University Press Princeton, NJ, 2018).

[8] O'Connor, C. & Weatherall, J. O. *The misinformation age: How false beliefs spread* (Yale University Press, 2019).

[9] Golub, B. & Jackson, M. Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics* **2**, 112–49 (2010).

[10] Golub, B. & Jackson, M. How homophily affects the speed of learning and best response dynamics. *The Quarterly Journal of Economics* **127** (2012).

[11] Banerjee, A., Breza, E., Chandrasekhar, A. G. & Mobius, M. Naive learning with uninformed agents. *American Economic Review* **111**, 3540–3574 (2021).

[12] Banda, K. K. & Cluverius, J. Elite polarization, party extremity, and affective polarization. *Electoral Studies* **56**, 90–101 (2018).

[13] Meyer, C. *Matrix Analysis and Applied Linear Algebra* (SIAM, 2000).

[14] Acemoglu, D., Ozdaglar, A. & ParandehGheibi, A. Spread of (mis) information in social networks. *Games and Economic Behavior* **70**, 194–227 (2010).

[15] Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: First steps. *Social Networks* **5**, 109 – 137 (1983).

[16] Wang, Y. J. & Wong, G. Y. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association* **82**, 8–19 (1987).

[17] Abbe, E. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research* **18**, 6446–6531 (2017).

[18] Decelle, A., Krzakala, F., Moore, C. & Zdeborová, L. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E* **84**, 066106 (2011).

[19] Rohe, K., Chatterjee, S. & Yu, B. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics* **39**, 1878–1915 (2011).

[20] Abbe, E., Bandeira, A. S. & Hall, G. Exact recovery in the stochastic block model. *IEEE Transactions on information theory* **62**, 471–487 (2015).

[21] Levin, D. A. & Peres, Y. *Markov chains and mixing times*, vol. 107 (American Mathematical Soc., 2017).

[22] Jackson, M. O. *et al.* *Social and economic networks*, vol. 3 (Princeton university press Princeton, 2008).

[23] Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, 2018).

Acknowledgments

The authors express their sincere gratitude to Ben Golub, Matt Jackson, Steven Heilman, David Miller, and participants of the NSF Network Conference 2022, CCERSI 2023, CMGTA 2023 and SNAB 2023 for their invaluable suggestions and expert advice. Additionally, Chen Cheng wishes to acknowledge the financial support received through the Discovery Award from Johns Hopkins University.

Supplementary materials

Omitted Proofs

We will use the following notations in our proofs. By (1), our updating matrix $\mathbf{T}$ is defined as:

$$\mathbf{T} = \mathbf{D}_{1,\mathbf{A}}^{-1}\mathbf{A}\mathbf{D}_{2,\mathbf{A}}\,, \tag{17}$$

where $\mathbf{D}_{1,\mathbf{A}}$ and $\mathbf{D}_{2,\mathbf{A}}$ are diagonal matrices with diagonal entries:

$$(\mathbf{D}_{1,\mathbf{A}})_{ii} = \sum_j A_{ij}\,\phi(\alpha, d_j(\mathbf{A}))\,,$$

$$(\mathbf{D}_{2,\mathbf{A}})_{ii} = \phi(\alpha, d_i(\mathbf{A}))\,.$$

Recall that in section 2, we have restricted $\mathcal{G}(\mathbf{T})$ to be strongly connected, which is equivalent to say $\mathbf{T}$ is irreducible. In addition, we study only the case in which $\lim_t \mathbf{T}^t$ converges.

Throughout the proofs in the appendix, there are constants in equalities and inequalities. The capital letter C will denote all such constants (possibly different) that are positive and are independent of n.

Proof of Lemma 1

Preliminaries of Proof of Lemma 1

Lemma 1 is a generalization of the result in [10]. For the proof, we proceed by finding an upper bound and a lower bound and show that they are indeed the same. The main part of the proof is the same as that of [10] but some details are adjusted to work for our matrix $\mathbf{T}$ and the norm

$\|\cdot\|_2$. The key component of the proof is to apply the spectral theorem to get a decomposition of the matrix $\mathbf{T}$. We first show that such decomposition exists by the following Lemma:

Lemma 3. *The matrix* $\mathbf{T}$ *in (17) is diagonalizable.*

Proof. First, note that

$$(\mathbf{D}_{1,\mathbf{A}}\mathbf{D}_{2,\mathbf{A}})^{1/2}\,\mathbf{T}\,(\mathbf{D}_{1,\mathbf{A}}\mathbf{D}_{2,\mathbf{A}})^{-1/2} = \mathbf{D}_{1,\mathbf{A}}^{-1/2}\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{T}\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}\,. \tag{18}$$

The equality holds because the diagonal matrices $\mathbf{D}_{1,\mathbf{A}}$ and $\mathbf{D}_{2,\mathbf{A}}$ commute. Equation (18) says $\mathbf{T}$ is similar to a symmetric matrix. Then, since any real symmetric matrix is diagonalizable, $\mathbf{T}$ is also diagonalizable. □

Lemma 3 allows us to apply spectral theorem for diagonalizable matrices [13] to $\mathbf{T}$. Let the spectral decomposition of $\mathbf{T}$ be:

$$\mathbf{T} = \sum_{i=1}^{n} \lambda_i \mathbf{U}_i\,, \tag{19}$$

where λ_i are the eigenvalues of $\mathbf{T}$ in decreasing order (in magnitude) and $\mathbf{U}_i$ are the orthogonal projections onto the eigenspace of $\mathbf{T}$ associated with λ_i. The next lemma, commonly referred to as the Perron-Fronbenius Theorem, gives an important property of the eigenvalues of the matrix $\mathbf{T}$ that is applied in the proof of Lemma 1.

Lemma 4. *For a nonnegative irreducible stochastic matrix* $\mathbf{T}$*, its spectral radius* $\rho(\mathbf{T}) = 1$ *is a simple eigenvalue. In addition, the limit* $\mathbf{T}^\infty$ *exists and takes the form:*

$$\mathbf{T}^\infty = \lambda_1 \mathbf{U}_\mathbf{1} = \mathbf{v}\mathbf{w}^\top\,,$$

where $\mathbf{v}$ *and* $\mathbf{w}$ *are left and right eigenvectors of* $\mathbf{T}$ *corresponding to* λ_1*, normalized so that* $\mathbf{w}^\top\mathbf{v} = 1$.

The proof is given in [13].

Proof of Lemma 1

Upper Bound

In this part, we want to achieve an upper bound of the distance $\left\|\left(\mathbf{T}^t - \mathbf{T}^\infty\right)\mathbf{b}\right\|_2$ by applying the spectral decomposition of the matrix $\mathbf{T}$ in (19). Note that

$$\mathbf{T}^t - \mathbf{T}^\infty = \sum_{i=2}^{n} \lambda_i^t \mathbf{U}_i \,. \tag{20}$$

The sum from 2 to n is justified by Lemma 4. The largest eigenvalue of $\mathbf{T}$ in magnitude is 1 and all other eigenvalues have magnitude less than 1. Then, the $i = 1$ part in the sum $\mathbf{T}^t = \sum_{i=1}^{n} \lambda_i^t \mathbf{U_i}$ cancels with $\mathbf{T}^\infty$, since $\mathbf{U}_1 = (\mathbf{v}\mathbf{w}^\top)^t = \mathbf{v}(\mathbf{w}^\top\mathbf{v})^{t-1}\mathbf{w}^\top = \mathbf{v}\mathbf{w}^\top = \mathbf{T}^\infty$ for $t \in \mathbb{N}$. Applying (20), we have

$$\begin{aligned}
\left\|\left(\mathbf{T}^t - \mathbf{T}^\infty\right)\mathbf{b}\right\|_2^2 &= \left\|\sum_{i=2}^{n} \lambda_i^t \mathbf{U}_i \mathbf{b}\right\|_2^2 && (21)\\
&= \sum_{i=2}^{n} |\lambda_i|^{2t} \left\|\mathbf{U}_i \mathbf{b}\right\|_2^2 && (\mathbf{U}_i\mathbf{U}_j = 0 \text{ for } i \neq j)\\
&\leq |\lambda_2|^{2t} \sum_{i=2}^{n} \left\|\mathbf{U}_i \mathbf{b}\right\|_2^2\\
&= |\lambda_2|^{2t} \left\|\sum_{i=2}^{n} \mathbf{U}_i \mathbf{b}\right\|_2^2 . && (\mathbf{U}_i\mathbf{U}_j = 0 \text{ for } i \neq j)
\end{aligned}$$

View $\mathbf{U} = \sum_{i=2}^{n} \mathbf{U}_i$ as a single orthogonal projection and it has the property that: $\mathbf{U} = \mathbf{U}^\top = \mathbf{U}^2$. Then, for any vector $\mathbf{x}$,

$$\|\mathbf{U}\mathbf{x}\|_2^2 = \langle \mathbf{U}\mathbf{x}, \mathbf{U}\mathbf{x} \rangle = \langle \mathbf{U}^\top\mathbf{U}\mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{U}\mathbf{x}, \mathbf{x} \rangle \leq \|\mathbf{U}\mathbf{x}\|_2 \|\mathbf{x}\|_2 \,.$$

The last inequality is the Cauchy-Schwarz inequality. By (21),

$$\left\|\left(\mathbf{T}^t - \mathbf{T}^\infty\right)\mathbf{b}\right\|_2^2 \leq |\lambda_2|^{2t} \|\mathbf{b}\|_2^2 = |\lambda_2|^{2t} \,.$$

The upper bound is obtained by taking square root on both sides.

Lower Bound

Here, we obtain a lower bound by considering a specific $\mathbf{b}$. Let $\mathbf{b}$ be the eigenvector of $\mathbf{T}$ corresponding to its second largest eigenvalue λ_2 in magnitude, with $\|\mathbf{b}\|_2^2 = 1$. Note $(\mathbf{T}^t - \mathbf{T}^\infty)\,\mathbf{b} = \lambda_2^t\mathbf{b}$. This is true because $\mathbf{U}_i\mathbf{b} = 0$ for all $i \neq 2$. Then,

$$\begin{aligned}\left\|\left(\mathbf{T}^t - \mathbf{T}^\infty\right)\mathbf{b}\right\|_2^2 &= \left\|\lambda_2^t\mathbf{b}\right\|_2^2 \\ &= |\lambda_2|^{2t}\,.\end{aligned}$$

By taking square root on both sides, we see that the lower bound is the same as the upper bound.

■

Proof of Proposition 3

We first show a simple lemma that is used in the proof of Proposition 3.

Lemma 5. *Consider similar $n \times n$ matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ such that $\mathbf{M}_2 = \mathbf{P}\mathbf{M}_1\mathbf{P}^{-1}$, where $\mathbf{P}$ is some invertible $n \times n$ matrix. If $\mathbf{v}$ is an eigenvector of $\mathbf{M}_1$ corresponding to eigenvalue λ, then $\mathbf{P}\mathbf{v}$ is an eigenvector of $\mathbf{M}_2$ corresponding to the same eigenvalue λ.*

Proof. Note that

$$\mathbf{M}_2(\mathbf{P}\mathbf{v}) = \left(\mathbf{P}\mathbf{M}_1\mathbf{P}^{-1}\right)(\mathbf{P}\mathbf{v}) = \mathbf{P}\mathbf{M}_1\mathbf{v} = \lambda\mathbf{P}\mathbf{v}\,.$$

By the definition of eigenvectors, $\mathbf{P}\mathbf{v}$ is an eigenvector of $\mathbf{M}_2$ corresponding to the eigenvalue λ. □

Proof of Proposition 3

We want to find the limit:

$$\mathbf{T}^\infty = \lim_{t\to\infty}\mathbf{T}^t = \lim_{t\to\infty}\mathbf{D}_{1,\mathbf{A}}^{-1/2}\mathbf{D}_{2,\mathbf{A}}^{-1/2}\left(\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}\mathbf{A}\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}\right)^t\mathbf{D}_{1,\mathbf{A}}^{1/2}\mathbf{D}_{2,\mathbf{A}}^{1/2}\,. \qquad (22)$$

Given that the limit converges,

$$\lim_{t\to\infty}\left(\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}\mathbf{A}\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}\right)^t = \lambda_1\mathbf{v}_1\mathbf{v}_1^\top\,, \tag{23}$$

where $\lambda_1 = 1$ is the largest eigenvalue in magnitude and $\mathbf{v}_1$ is the corresponding unit eigenvector of the matrix $\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}\mathbf{A}\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}$. The limit being in this form is a result of the Perron-Frobenius Theorem, which is the same as what we apply in equality (21) in the proof of Lemma 1. Note that since the matrix in (23) is symmetric, the left and right eigenvectors are the same. To find $\mathbf{v}_1$, apply Lemma 5: it's easy to see that $\mathbf{D}_{1,\mathbf{A}}^{-1}\mathbf{A}\mathbf{D}_{2,\mathbf{A}}$ has eigenvector $\mathbf{e} = (1,\ldots,1)^\top$, corresponding to the eigenvalue 1. Then, $\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}\mathbf{A}\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}$ has eigenvector $\mathbf{w}_1 = \mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{1/2}\mathbf{e}$. To make it into a unit vector, we divide it by its magnitude $\|\mathbf{w}_1\|_2$, which has the expression:

$$\|\mathbf{w}_1\|_2 = \left\|\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{1/2}\mathbf{e}\right\|_2 = \left(\sum_{i,j} A_{ij}\,\phi(\alpha, d_j(\mathbf{A}))\phi(\alpha, d_i(\mathbf{A}))\right)^{1/2}.$$

Since $\mathbf{v}_1 = \mathbf{w}_1/\|\mathbf{w}_1\|_2$, (22) becomes:

$$\begin{aligned}\mathbf{T}^\infty =& \mathbf{D}_{1,\mathbf{A}}^{-1/2}\mathbf{D}_{2,\mathbf{A}}^{-1/2}\left(\|\mathbf{w}_1\|_2^{-1}\,\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{1/2}\mathbf{e}\right)\left(\|\mathbf{w}_1\|_2^{-1}\,\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{1/2}\mathbf{e}\right)^\top\mathbf{D}_{1,\mathbf{A}}^{1/2}\mathbf{D}_{2,\mathbf{A}}^{1/2}\\ =& \|\mathbf{v}_1\|_2^{-2}\,\mathbf{E}_n\mathbf{D}_{1,\mathbf{A}}\mathbf{D}_{2,\mathbf{A}}\,,\end{aligned}$$

where $\mathbf{E}_n$ is an $n \times n$ matrix with all entries equal to 1. This leads us to the result of the limit:

$$\mathbf{T}_{ij}^\infty = \frac{\sum_i A_{ij}\,\phi(\alpha, d_j(\mathbf{A}))\phi(\alpha, d_i(\mathbf{A}))}{\sum_{i,j} A_{ij}\,\phi(\alpha, d_j(\mathbf{A}))\phi(\alpha, d_i(\mathbf{A}))}\,,$$

for each $i, j = 1, \ldots, n$. ■

Proof of Proposition 1

In this section, our results are for the nonrandom matrix $\mathbf{T}^*$ defined by:

$$\mathbf{T}^* = \mathbf{D}_{1,\mathbf{R}}^{-1}\mathbf{R}\mathbf{D}_{2,\mathbf{R}}\,, \tag{24}$$

where $\mathbf{R} = \mathbb{E}\mathbf{A}$ and $\mathbf{D}_{1,\mathbf{R}}$, $\mathbf{D}_{2,\mathbf{R}}$ are diagonal matrices defined by:

$$(\mathbf{D}_{1,\mathbf{R}})_{ii} = \sum_j R_{ij}\, \phi(\alpha, d_j(\mathbf{R}))\,,$$

$$(\mathbf{D}_{2,\mathbf{R}})_{ii} = \phi(\alpha, d_i(\mathbf{R}))\,.$$

Before the proof of Proposition 1, we first show a Lemma:

Lemma 6. *Let the $m \times m$ matrix $\mathbf{F}_{\mathbf{T}^*}$ be defined as:*

$$(\mathbf{F}_{\mathbf{T}^*})_{kl} = \frac{n_l P_{kl}\, \phi(\alpha, \sum_h n_h P_{lh})}{\sum_l n_l P_{kl}\, \phi(\alpha, \sum_h n_h P_{lh})}\,. \tag{25}$$

Then, $\mathbf{F}_{\mathbf{T}^*}$ has the same eigenvalues as $\mathbf{T}^*$.

Proof. First, note $\mathbf{R}$ is a matrix with m^2 blocks. Within each block, the entries are identical. So is $\mathbf{T}^*$. From the definition of $\mathbf{T}^*$, we see that if vertex i belongs to group k and vertex j belongs to group l,

$$T^*_{ij} = \frac{R_{ij}\, \phi(\alpha, d_j(\mathbf{R}))}{\sum_j R_{ij}\, \phi(\alpha, d_j(\mathbf{R}))} = \frac{P_{kl}\, \phi(\alpha, \sum_h n_h P_{lh})}{\sum_l n_l P_{kl}\, \phi(\alpha, \sum_h n_h P_{lh})}\,.$$

After suitable rearrangement of the vertices, this matrix $\mathbf{T}^*$ is in the following block form:

$$\begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \dots & \mathbf{B}_{1m} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \dots & \mathbf{B}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{m1} & \mathbf{B}_{m2} & \dots & \mathbf{B}_{mm} \end{bmatrix},$$

where each $\mathbf{B}_{kl}$ is a block matrix and within each block the entries are identical. Denote the entries in block $\mathbf{B}_{kl}$ by B_{kl}, which takes the value of T^*_{ij} in (25), if vertex i is in group k and vertex j is in group l. Consider eigenvectors in the form of $\mathbf{v} = (v_1, v_1, \dots, v_2, v_2, \dots, v_m, v_m \dots)^\top$. $\mathbf{T}^*\mathbf{v} = \lambda\mathbf{v}$ implies:

$$n_1 B_{11} v_1 + n_2 B_{12} v_2 + \ldots + n_m B_{1m} v_m = \lambda v_1\,,$$

$$\vdots$$

$$n_1 B_{m1} v_1 + n_2 B_{m2} v_2 + \ldots + n_m B_{mm} v_m = \lambda v_m \,,$$

which completes the proof. □

Proof of Proposition 1

Proof. In this proof, we focus on the case that $d_1^* > d_2^*$, the proof of $d_1^* < d_2^*$ is almost the same and thus we omit it. By applying Lemma 6, we are able to reduce the $n \times n$ matrix $\mathbf{T}^*$ to an $m \times m$ matrix $\mathbf{F}_{\mathbf{T}^*}$. By our assumptions of $\mathbf{n}$ and $\mathbf{P}$ in (A1) and (2), we see there are two expected degrees, denoted as d_1^* and d_2^*. For vertices in group 1 (the group with size n_1), $d_1^* = n_1 p + (m-1) n_2 q$ and for vertices in the rest groups, $d_2^* = n_1 q + n_2 p + (m-2) n_2 q$. Then, the matrix $\mathbf{F}_{\mathbf{T}^*}$ is in the following form:

$$\mathbf{F}_{\mathbf{T}^*} = \begin{bmatrix} a & b & b & b & \ldots & b \\ e & c & d & d & \ldots & d \\ e & d & c & d & \ldots & d \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ e & d & d & \ldots & c & d \\ e & d & d & \ldots & d & c \end{bmatrix} , \tag{26}$$

where

$$\begin{aligned}
a &= \frac{n_1 p\, \phi(\alpha, d_1^*)}{n_1 p\, \phi(\alpha, d_1^*) + (m-1) n_2 q\, \phi(\alpha, d_2^*)} \,, \\
b &= \frac{n_2 q\, \phi(\alpha, d_2^*)}{n_1 p\, \phi(\alpha, d_1^*) + (m-1) n_2 q\, \phi(\alpha, d_2^*)} \,, \\
c &= \frac{n_2 p\, \phi(\alpha, d_2^*)}{n_1 q\, \phi(\alpha, d_1^*) + n_2 p\, \phi(\alpha, d_2^*) + (m-2) n_2 q\, \phi(\alpha, d_2^*)} \,, \\
d &= \frac{n_2 q\, \phi(\alpha, d_2^*)}{n_1 q\, \phi(\alpha, d_1^*) + n_2 p\, \phi(\alpha, d_2^*) + (m-2) n_2 q\, \phi(\alpha, d_2^*)} \,, \\
e &= \frac{n_1 q\, \phi(\alpha, d_1^*)}{n_1 q\, \phi(\alpha, d_1^*) + n_2 p\, \phi(\alpha, d_2^*) + (m-2) n_2 q\, \phi(\alpha, d_2^*)} \,.
\end{aligned}$$

We perform row operations on the matrix $\mathbf{F}_{\mathbf{T}^*} - \lambda \mathbf{I_m}$ before finding the zeros of the characteristic polynomial:

$$\det(\mathbf{F}_{\mathbf{T}^*} - \lambda \mathbf{I_m}) = 0 \,. \tag{27}$$

Subtract last row from rows $2, 3, \ldots, m-1$ and equation (27) becomes:

$$\det \begin{bmatrix} a-\lambda & b & b & b & \dots & b \\ 0 & c-d-\lambda & 0 & 0 & \dots & -(c-d-\lambda) \\ 0 & 0 & c-d-\lambda & 0 & \dots & -(c-d-\lambda) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & c-d-\lambda & -(c-d-\lambda) \\ e & d & d & \dots & d & c-\lambda \end{bmatrix} = 0\,.$$

We see now one eigenvalue is $\lambda = c-d$, with algebraic multiplicity $m-2$.

Suppose $\lambda \neq c-d$. Multiply rows $2, 3, \ldots, m-1$ by $1/(c-d-\lambda)$, we have:

$$\det \begin{bmatrix} a-\lambda & b & b & b & \dots & b \\ 0 & 1 & 0 & 0 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \\ e & d & d & \dots & d & c-\lambda \end{bmatrix} = 0\,.$$

Then, subtract multiplies of rows $2, 3, \ldots, m-1$ from the first and last row, we have:

$$\det \begin{bmatrix} a-\lambda & 0 & 0 & 0 & \dots & (m-1)b \\ 0 & 1 & 0 & 0 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \\ e & 0 & 0 & \dots & 0 & c-\lambda+(m-2)d \end{bmatrix} = 0\,. \tag{28}$$

Computing the determinant (28), we see the eigenvalues different from $c-d$ should satisfy the equation:

$$(a-\lambda)(c-\lambda+(m-2)d) - e(m-1)b = 0\,.$$

Note whether m is even or odd does not change the equation. By solving the quadratic equation, we see that the other two eigenvalues are $\lambda = 1$ and $\lambda = a-e$. To get the second largest eigenvalue, we compare $a-e$ and $c-d$, both being positive. So, in the computation below, $|\lambda_2| = \lambda_2$ and the absolute value is omitted. We then compare the two candidates $a-e$ and

$c - d$. Denoting the denominators of a and c as D_1 and D_2 respectively, we have

$$
\begin{aligned}
& D_1 D_2 (a - e - c + d) \\
= & \, n_1 p \, \phi(\alpha, d_1^*)[n_1 q \, \phi(\alpha, d_1^*) + n_2 p \, \phi(\alpha, d_2^*) + (m-2) n_2 q \, \phi(\alpha, d_2^*)] \\
& - (n_2(p-q)\phi(\alpha, d_2^*) + n_1 q \, \phi(\alpha, d_1^*))[n_1 p \, \phi(\alpha, d_1^*) + (m-1) n_2 q \, \phi(\alpha, d_2^*)] \\
= & \, (m-1) n_2 q \, \phi(\alpha, d_2^*)(p-q)(n_1 \, \phi(\alpha, d_1^*) - n_2 \, \phi(\alpha, d_2^*)) \, .
\end{aligned} \tag{29}
$$

This proves (11) in Proposition 1, since $n_1 \phi(\alpha, d_1^*) \geq n_2 \, \phi(\alpha, d_2^*)$ is equivalent to $\alpha \geq g^{-1}(n_1/n_2)$. We are left with the special case of $m = 2$. We see that if $m = 2$, there is no eigenvalue being $c - d$. This completes the computations in Proposition 1. □

Proof of Theorem 1

In this proof we focus on the case that $d_1^* > d_2^*$, the proof of $d_1^* < d_2^*$ is almost the same and thus we omit it. For $m \geq 3$, to show the monotonicity we first consider $\alpha \geq g^{-1}(n_1/n_2)$, for which the second largest eigenvalue is:

$$
\begin{aligned}
\lambda_2 &= a - e \\
&= \frac{n_1 p \, \phi(\alpha, d_1^*)}{n_1 p \, \phi(\alpha, d_1^*) + (m-1) n_2 q \, \phi(\alpha, d_2^*)} - \frac{n_1 q \, \phi(\alpha, d_1^*)}{n_1 q \, \phi(\alpha, d_1^*) + (n_2 p + (m-2) n_2 q)\phi(\alpha, d_2^*)} \, ,
\end{aligned} \tag{30}
$$

From this expression, we see that the limit is 0 as α goes to ∞. We rewrite the above expression of λ_2 as:

$$
\lambda_2 = \frac{1}{1 + C_1 g(\alpha)} - \frac{1}{1 + C_2 g(\alpha)} \, , \tag{31}
$$

where $C_1 = \frac{(m-1)n_2 q}{n_1 p}$, $C_2 = \frac{n_2 p + (m-2) n_2 q}{n_1 q}$ and $g(\alpha, d_1^*, d_2^*) = \frac{\phi(\alpha, d_2^*)}{\phi(\alpha, d_1^*)}$. Since d_1^* and d_2^* are fixed here, we omit them in the notation below and write $g(\alpha)$.

Note that $0 < C_1 < C_2$. Then, the derivative with respect to α is:

$$\begin{aligned} \frac{\partial \lambda_2}{\partial \alpha} &= -\frac{C_1 dg/d\alpha}{(1+C_1 g(\alpha))^2} + \frac{C_2 dg/d\alpha}{(1+C_2 g(\alpha))^2} \\ &= \frac{dg}{d\alpha}\left[\frac{(C_2-C_1)(1-C_1C_2g^2(\alpha))}{(1+C_1g(\alpha))^2(1+C_2g(\alpha))^2}\right] . \end{aligned} \tag{32}$$

Property 2 in (6) implies that $\frac{dg}{d\alpha} < 0$. Thus, the $\frac{\partial \lambda_2}{\partial \alpha} \leq 0$ for $1 - C_1C_2g^2(\alpha) \geq 0$, which is equivalent to:

$$\alpha \geq g^{-1}\left(\frac{1}{\sqrt{C_1C_2}}\right) = g^{-1}\left(\frac{n_1}{n_2}\left(\frac{p}{(m-1)(p+(m-2)q)}\right)^{1/2}\right) . \tag{33}$$

Similarly, for $\alpha < g^{-1}(n_1/n_2)$,

$$\lambda_2 = c - d = \frac{n_2(p-q)g(\alpha)}{n_1q + n_2(p+(m-2)q)g(\alpha)} . \tag{34}$$

Then, taking the derivative and simplify, we get

$$\frac{\partial \lambda_2}{\partial \alpha} = \frac{\partial g}{\partial \alpha}\frac{n_1n_2(p-q)q}{(n_1q + n_2(p+(m-2)q)g(\alpha))^2} < 0 , \tag{35}$$

since $\frac{\partial g}{\partial \alpha} < 0$. For $m = 2$, noticing that $\lambda_2 = a - e$ and hence the proof is similar to (30)–(33). ■

Proposition 5. *Denote the second eigenvector of* $\mathbf{T}^*$ *by*

$$\mathbf{v}_2 = (\underbrace{v_{21}, \ldots, v_{21}}_{n_1}, \underbrace{v_{22}, \ldots, v_{22}}_{n_2}, \ldots, \underbrace{v_{2m}, \ldots, v_{2m}}_{n_2})^\top .$$

For $d_1^* > d_2^*$, *when* $\alpha > g^{-1}(\frac{n_1}{n_2})$, *we have*

$$v_{21} = -\frac{(m-1)b}{e(m-1)n_2 + n_1(m-1)^2b^2/e},\ v_{22} = \ldots = v_{2m} = \frac{1}{(m-1)n_2 + n_1(m-1)^2b^2/e^2} .$$

When $\alpha < g^{-1}(\frac{n_1}{n_2})$, *let* $\mathbf{v}' = (\underbrace{v'_1, \ldots, v'_1}_{n_1}, \underbrace{v'_2, \ldots, v'_2}_{n_2}, \ldots, \underbrace{v'_m, \ldots, v'_m}_{n_2})^\top$, *where*

$$v'_1 = -\frac{(m-1)b}{e(m-1)n_2 + n_1(m-1)^2b^2/e},\ v'_2 = \ldots = v'_m = \frac{1}{(m-1)n_2 + n_1(m-1)^2b^2/e^2} .$$

then $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{-1/2}\mathbf{v}_2$ *is any vector orthogonal to* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{v}'$ *and* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{1}$*, where* $\mathbf{D}_{1,\mathbf{R}}$ *and* $\mathbf{D}_{2,\mathbf{R}}$ *are defined below* (24)*.*

When $\alpha = g^{-1}(\frac{n_1}{n_2})$*,* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{-1/2}\mathbf{v}_2$ *is any vector orthogonal to* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{1}$*.*

In contrast if $d_1^* < d_2^*$*, when* $\alpha < g^{-1}(\frac{n_1}{n_2})$*, we have*

$$v_{21} = -\frac{(m-1)b}{e(m-1)n_2 + n_1(m-1)^2b^2/e},\ v_{22} = \ldots = v_{2m} = \frac{1}{(m-1)n_2 + n_1(m-1)^2b^2/e^2}\,.$$

When $\alpha > g^{-1}(\frac{n_1}{n_2})$*, then* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{-1/2}\mathbf{v}_2$ *is any vector orthogonal to* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{v}'$ *and* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{1}$*.*

When $\alpha = g^{-1}(\frac{n_1}{n_2})$*,* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{-1/2}\mathbf{v}_2$ *is any vector orthogonal to* $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{1}$*.*

For $m = 2$*, the second eigenvector is*

$$v_{21} = -\frac{b}{en_2 + n_1b^2/e},\ v_{22} = \frac{1}{n_2 + n_1b^2/e^2}\,.$$

Proof. Similar to the previous proof, we focus on the case that $d_1^* > d_2^*$. By Proposition 1, we conclude that the second eigenvalue is $a - e$ or $c - d$ depending on α. Denote the second eigenvector by $\mathbf{v}_2 = (\underbrace{v_{21}, \ldots, v_{21}}_{n_1}, \underbrace{v_{22}, \ldots, v_{22}}_{n_2}, \ldots, \underbrace{v_{2m}, \ldots, v_{2m}}_{n_2})^\top$. When $\alpha > g^{-1}(\frac{n_1}{n_2})$, the second eigenvalue is $a - e$, combining (26) with Lemma 6, we have the following equations for $\mathbf{v}_2$

$$av_{21} + bv_{22} + \ldots + bv_{2m} = (a-e)v_{21}\,, \tag{36}$$

$$\vdots$$

$$ev_{21} + cv_{2i} + d\sum_{j\neq i, j>1} v_{2j} = (a-e)v_{2i}\,, \tag{37}$$

$$\vdots$$

$$ev_{21} + dv_{22} + \ldots + cv_{2m} = (a-e)v_{2m}\,. \tag{38}$$

By (36) we have

$$-ev_{21} = b\sum_{j=2}^{m} v_{2j}\,. \tag{39}$$

Substituting this into (37), we have

$$(a+b-e-c)v_{2i} + (b-d)\sum_{j\neq i, j>1} v_{2j} = 0,\ i = 2,\ldots,m. \tag{40}$$

Therefore, the vector $(v_{22},\ldots,v_{2m})^\top$ is the eigenvector of the matrix $diag((a+d-e-c),\ldots,(a+d-e-c)) + (b-d)\mathbf{1}\mathbf{1}^\top$ corresponding to the zero eigenvalue. Noticing that $a+d-e-c = -(m-1)(b-d)$, we have

$$v_{22} = \ldots = v_{2m}\,.$$

Combining this with (39), we have

$$v_{21} = -\frac{(m-1)b}{e(m-1)n_2 + n_1(m-1)^2b^2/e},\ v_{22} = \ldots = v_{2m} = \frac{1}{(m-1)n_2 + n_1(m-1)^2b^2/e^2}\,.$$

Similarly, when $\alpha < g^{-1}(\frac{n_1}{n_2})$, recalling the definitions of $\mathbf{D}_{1,\mathbf{R}}$ and $\mathbf{D}_{2,\mathbf{R}}$ below (24) and notice that they are diagonal matrices. We imply that $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{-1/2}\mathbf{v}_2$ is any vector orthogonal to $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{v}'$ and $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{1}$.

When $\alpha = g^{-1}(\frac{n_1}{n_2})$, $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{-1/2}\mathbf{v}_2$ is any vector orthogonal to $(\mathbf{D}_{1,\mathbf{R}}\mathbf{D}_{2,\mathbf{R}})^{1/2}\mathbf{1}$.

For $m = 2$, there is no eigenvalue being $c-d$ and thus the second eigenvector is

$$v_{21} = -\frac{b}{en_2 + n_1b^2/e},\ v_{22} = \frac{1}{n_2 + n_1b^2/e^2}\,.$$

□

Proof of Lemma 2

Preliminaries of Proof of Lemma 2

Recall that $\mathbf{R} = \mathbb{E}\mathbf{A}$. It is easy to see from the definition of matrices in (17) and (24) that to compare the eigenvalues $\lambda_2(\mathbf{T})$ and $\lambda_2(\mathbf{T}^*)$, it is equivalent to compare the eigenvalues of

$\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}\mathbf{A}\mathbf{D}_{2,\mathbf{A}}^{1/2}\mathbf{D}_{1,\mathbf{A}}^{-1/2}$ and $\mathbf{D}_{2,\mathbf{R}}^{1/2}\mathbf{D}_{1,\mathbf{R}}^{-1/2}\mathbf{R}\mathbf{D}_{2,\mathbf{R}}^{1/2}\mathbf{D}_{1,\mathbf{R}}^{-1/2}$ because they are similar to $\mathbf{T}$ and $\mathbf{T}^*$ correspondingly. For simplicity, denote

$$\mathbf{D}_{\mathbf{A}} = \mathbf{D}_{2,\mathbf{A}}^{-1}\mathbf{D}_{1,\mathbf{A}}\,,$$

and

$$\mathbf{D}_{\mathbf{R}} = \mathbf{D}_{2,\mathbf{R}}^{-1}\mathbf{D}_{1,\mathbf{R}}\,.$$

Then, by Weyl's inequality,

$$|\lambda_2(\mathbf{T}) - \lambda_2(\mathbf{T}^*)| \leq \left\|\mathbf{D}_{\mathbf{A}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathbf{A}}^{-1/2} - \mathbf{D}_{\mathbf{R}}^{-1/2}\mathbf{R}\mathbf{D}_{\mathbf{R}}^{-1/2}\right\|\,, \tag{41}$$

where $\|\cdot\|$ is the spectral norm. Therefore, it is sufficient to bound the spectral norm on the right and this is done in the proof of Lemma 2.

Due to the assumptions (14) and (15) made in Lemma 2, the expected degrees satisfy:

$$1 < \liminf_n \frac{\min_i d_i(\mathbf{R})}{\max_i d_i(\mathbf{R})(\mathbf{R})} \leq \limsup_n \frac{\max_i d_i(\mathbf{R})(\mathbf{R})}{\min_i d_i(\mathbf{R})} < \infty\,.$$

which implies that there exists a positive constant C such that for all n,

$$\min_i d_i(\mathbf{R}) < \max_i d_i(\mathbf{R})(\mathbf{R}) < C\min_i d_i(\mathbf{R})\,. \tag{42}$$

In addition,

$$\min_i (\mathbf{D}_{\mathbf{R}})_{ii} = \min_i \frac{\sum_j R_{ij}\,\phi(\alpha, d_j(\mathbf{R}))}{\phi(\alpha, d_i(\mathbf{R}))} = \min_i \sum_j R_{ij}\frac{\phi(\alpha, d_j(\mathbf{R}))}{\phi(\alpha, d_i(\mathbf{R}))} \geq \min_i d_i(\mathbf{R})\,. \tag{43}$$

Before the proof of Lemma 2, we first list 2 propositions that are used in Lemma 2. The proofs of these 2 propositions are left at the end of the section.

Proposition 6.

$$\|\mathbf{A} - \mathbf{R}\| \leq 3\sqrt{\max_i d_i(\mathbf{R})\log n}\,,$$

with probability at least $1 - \frac{2}{n^2}$.

Proposition 7. *With the assumptions (13)-(15) of Lemma 2,*

$$\left\|\mathbf{D}_{\mathbf{A}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\right\| \leq C\sqrt{\max_i d_i(\mathbf{R})\log n}\left(\min_i d_i(\mathbf{R})\right)^{-3/2},$$

with probability at least $1-\frac{14}{n^2}$.

Proof of Lemma 2

Proof. First, note that the spectral norm in (41) can be split as:

$$\begin{aligned}
&\left\|\mathbf{D}_{\mathbf{A}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathbf{A}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\mathbf{R}\mathbf{D}_{\mathbf{R}}^{-1/2}\right\| \\
\leq &\left\|\mathbf{D}_{\mathbf{A}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathbf{A}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathbf{A}}^{-1/2}\right\| + \left\|\mathbf{D}_{\mathbf{R}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathbf{A}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathbf{R}}^{-1/2}\right\| \\
&+\left\|\mathbf{D}_{\mathbf{R}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathbf{R}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\mathbf{R}\mathbf{D}_{\mathbf{R}}^{-1/2}\right\| \\
\leq &\left\|\mathbf{D}_{\mathbf{A}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\right\|\cdot\|\mathbf{A}\|\cdot\left\|\mathbf{D}_{\mathbf{A}}^{-1/2}\right\| + \left\|\mathbf{D}_{\mathbf{A}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\right\|\cdot\|\mathbf{A}\|\cdot\left\|\mathbf{D}_{\mathbf{R}}^{-1/2}\right\| \\
&+\|\mathbf{A}-\mathbf{R}\|\cdot\left\|\mathbf{D}_{\mathbf{R}}^{-1}\right\|.
\end{aligned}$$

By triangular inequality,

$$\|\mathbf{A}\| \leq \|\mathbf{A}-\mathbf{R}\| + \|\mathbf{R}\|.$$

The spectral norm is bounded above by the Frobenius norm:

$$\|\mathbf{R}\| \leq \|\mathbf{R}\|_F = \sqrt{\sum_{i,j} R_{ij}^2} \leq Cn\max_{i,j} P_{ij} \leq C\min_j d_j(\mathbf{R}),$$

for all $n>0$, where $\max_{i,j} P_{ij}$ is the largest entry of the matrix $\mathbf{P}$ in the stochastic block model.

Then applying Proposition 6 together with assumption 1 in (13), we get

$$\|\mathbf{A}\| \leq C\min_j d_j(\mathbf{R}),$$

with probability at least $1-\frac{2}{n^2}$. Similarly, by Proposition 7,

$$\left\|\mathbf{D}_{\mathbf{A}}^{-1/2}\right\| \leq C\left(\min_j d_j(\mathbf{R})\right)^{-1/2},$$

with probability at least $1-\frac{14}{n^2}$, for n large enough. Conditioning on the event

$$\left\{\|\mathbf{A}-\mathbf{R}\| \leq 3\sqrt{\max_i d_i(\mathbf{R})\log n}\right\} \bigcap \left\{\left\|\mathbf{D}_{\mathbf{A}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\right\| \leq C\sqrt{n\log n}\left(\min_i d_i(\mathbf{R})\right)^{-3/2}\right\},$$

which takes place with probability at least $1-\frac{16}{n^2}$, we get

$$\left\|\mathbf{D}_{\mathbf{A}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathbf{A}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\mathbf{R}\mathbf{D}_{\mathbf{R}}^{-1/2}\right\| \leq C\frac{\sqrt{\max_i d_i(\mathbf{R})\log n}}{\min_i d_i(\mathbf{R})} = C\frac{\sqrt{\tilde{\tau}_n \log n}}{\tau_n\sqrt{n}}, \tag{44}$$

which finishes the proof. □

Proof of Theorem 2

Recall that in Proposition 1 and Theorem 1, we computed $|\lambda_2(\mathbf{T}^*(\alpha))|$ and $\frac{\partial|\lambda_2(\mathbf{T}^*(\alpha))|}{\partial\alpha}$. Since $|\lambda_2(\mathbf{T}^*(\alpha))|$ is positive and monotonically decreasing on $\mathcal{D}_\alpha$, for $\alpha_1 > \alpha_0$,

$$|\lambda_2(\mathbf{T}^*(\alpha_0))| - |\lambda_2(\mathbf{T}^*(\alpha_1))| = \lambda_2(\mathbf{T}^*(\alpha_0)) - \lambda_2(\mathbf{T}^*(\alpha_1)). \tag{45}$$

By mean value theorem,

$$\lambda_2(\mathbf{T}^*(\alpha_0)) - \lambda_2(\mathbf{T}^*(\alpha_1)) \geq (\alpha_0-\alpha_1)\left.\frac{\partial\lambda_2(\mathbf{T}^*(\alpha))}{\partial\alpha}\right|_{\alpha=\alpha_1}. \tag{46}$$

We want to find the size of the derivative. Let

$$C_1 = \frac{(m-1)n_2 q}{n_1 p},$$

$$C_2 = \frac{n_2 p + (m-2)n_2 q}{n_1 q}.$$

Then, for $\alpha \geq g^{-1}(n_1/n_2)$, using the expression (32) in the proof of Theorem 1, we have

$$\left.\frac{\partial\lambda_2(\mathbf{T}^*(\alpha))}{\partial\alpha}\right|_{\alpha=\alpha_1} = -C_3\left(\frac{n_2}{n_1},\frac{q}{p},\frac{\phi(\alpha_1,d_2)}{\phi(\alpha_1,d_1)}\right)\left[\frac{\partial\phi/\partial\alpha(\alpha_1,d_1)}{\phi(\alpha_1,d_1)} - \frac{\partial\phi/\partial\alpha(\alpha_1,d_2)}{\phi(\alpha_1,d_2)}\right], \tag{47}$$

where C_3 is some positive constant which depends on the fractions $\frac{n_2}{n_1}$, $\frac{q}{p}$ and $\frac{\phi(\alpha_1,d_2)}{\phi(\alpha_1,d_1)}$ and it does not depend on n when n is large. By assumption,

$$\alpha_1-\alpha_0 > \hat{C}\left(\frac{n_2}{n_1},\frac{q}{p},\frac{\phi(\alpha_1,d_2)}{\phi(\alpha_1,d_1)},\tilde{C}\right)\frac{\sqrt{\tilde{\tau}_n\log n}}{\tau_n\sqrt{n}}\left[\frac{\partial\phi/\partial\alpha(\alpha_1,d_1)}{\phi(\alpha_1,d_1)} - \frac{\partial\phi/\partial\alpha(\alpha_1,d_2)}{\phi(\alpha_1,d_2)}\right]^{-1},$$

where $\tilde{C}$ is the constant in Lemma 2. Combining with (45) – (47), we have

$$|\lambda_2(\mathbf{T}^*(\alpha_0))| - |\lambda_2(\mathbf{T}^*(\alpha_1))| > \hat{C}C_3\frac{\sqrt{\tilde{\tau}_n \log n}}{\tau_n\sqrt{n}}\,, \tag{48}$$

for $\alpha \geq g^{-1}(n_1/n_2)$.

On the other hand, for $\alpha < g^{-1}(n_1/n_2)$, using expression (35) in the proof of Theorem 1, we have equation (47), for a positive constant C_4 possibly different from C_3. Similar to (48), we then get

$$|\lambda_2(\mathbf{T}^*(\alpha_0))| - |\lambda_2(\mathbf{T}^*(\alpha_1))| > \hat{C}C_4\frac{\sqrt{\tilde{\tau}_n \log n}}{\tau_n\sqrt{n}}\,, \tag{49}$$

for $\alpha < g^{-1}(n_1/n_2)$.

We then notice that

$$\begin{aligned} |\lambda_2(\mathbf{T}(\alpha_0))| - |\lambda_2(\mathbf{T}(\alpha_1))| =& |\lambda_2(\mathbf{T}^*(\alpha_0))| - |\lambda_2(\mathbf{T}^*(\alpha_1))| + (|\lambda_2(\mathbf{T}(\alpha_0))| - |\lambda_2(\mathbf{T}^*(\alpha_0))|) \\ &+ (|\lambda_2(\mathbf{T}^*(\alpha_1))| - |\lambda_2(\mathbf{T}(\alpha_1))|) \\ >& |\lambda_2(\mathbf{T}^*(\alpha_0))| - |\lambda_2(\mathbf{T}^*(\alpha_1))| + C_5\frac{\sqrt{\tilde{\tau}_n \log n}}{\tau_n\sqrt{n}}\,, \end{aligned}$$

with probability at least $1 - \frac{16}{n^2}$ for some constant C_5, by Lemma 2. Letting $\hat{C} = \frac{(1-C_5)}{\min(C_3, C_4)}$, we see that the result of Theorem 2 follows. ■

Proof of Proposition 6

To bound $\|\mathbf{A} - \mathbf{R}\|$, rewrite it as:

$$\mathbf{A} - \mathbf{R} = \mathbf{Y} + \sum_{1\leq i<j\leq n} \mathbf{X}^{i,j}\,,$$

where $\mathbf{X}^{i,j} = (\mathbf{A} - \mathbf{R})(\mathbf{e}_i\mathbf{e}_j^\top + \mathbf{e}_j\mathbf{e}_i^\top)$, $\mathbf{Y} = \text{diag}(A_{11} - R_{11}, \ldots, A_{nn} - R_{nn})$ and $\mathbf{e}_i$ is the standard basis of $\mathbb{R}^n$. Note $\mathbf{X}^{i,j}$ are independent with mean zero and they are independent of $\mathbf{Y}$, which also has mean zero. Then, the matrix Bernstein inequality, [23] Theorem 5.4.1 implies:

$$\mathbb{P}\left\{\left\|\mathbf{Y} + \sum_{1\leq i<j\leq n} \mathbf{X}^{i,j}\right\| \geq t\right\} \leq 2n\exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right)\,, \tag{50}$$

where

$$\|\mathbf{Y}\|, \left\|\mathbf{X}^{i,j}\right\| \leq K = 1$$

and

$$\sigma^2 = \left\|\mathbb{E}\mathbf{Y}^2 + \sum_{1\leq i<j\leq n} \mathbb{E}(\mathbf{X}^{i,j})^2\right\|$$

Note $(\mathbf{X}^{i,j})^2 = (A_{ij} - R_{ij})^2(\mathbf{e}_i\mathbf{e}_i^\top + \mathbf{e}_j\mathbf{e}_j^\top)$ so that

$$\mathbb{E}(\mathbf{X}^{i,j})^2 = (R_{ij} - R_{ij}^2)(\mathbf{e}_i\mathbf{e}_i^\top + \mathbf{e}_j\mathbf{e}_j^\top)\,.$$

And

$$\mathbb{E}\mathbf{Y}^2 = \sum_{i=1}^{n}(R_{ii} - R_{ii}^2)\mathbf{e}_i\mathbf{e}_i^\top\,.$$

Since each A_{ij} is a Bernoulli random variable, $R_{ij} \in [0,1]$. Therefore, $\sum_{j=1}^{n}(R_{ij} - R_{ij}^2) \leq \max_i d_i(\mathbf{R})$, we see $\sigma^2 \leq \max_i d_i(\mathbf{R})$. Let $t = 3\sqrt{\max_i d_i(\mathbf{R}) \log n}$. By (50) we get:

$$\begin{aligned}\mathbb{P}\left\{\left\|\sum_{1\leq i<j\leq n} \mathbf{Y} + \mathbf{X}^{i,j}\right\| \geq 3\sqrt{\max_i d_i(\mathbf{R}) \log n}\right\} &\leq 2n \exp\left(-\frac{\frac{9\max_i d_i(\mathbf{R})\log n}{2}}{\max_i d_i(\mathbf{R}) + \frac{3\sqrt{\max_i d_i(\mathbf{R})\log n}}{3}}\right) \\ &\leq 2n\exp(-3\log n) \\ &= \frac{2}{n^2}\,.\end{aligned}$$

■

Proof of Proposition 7

The proof of Proposition 7 can be broken into proofs of several lemmas. We present the statement and proofs of these lemmas below.

Lemma 7. *Let $K > 1$ be a constant independent of n. Then, for $\alpha \geq 0$,*

$$\phi(\alpha, \min_i d_i(\mathbf{R})) \leq \phi(\alpha, K\min_i d_i(\mathbf{R})) \leq C_1\phi(\alpha, \min_i d_i(\mathbf{R}))\,, \tag{51}$$

and

$$C_2 \frac{\partial \phi}{\partial d}(\alpha, \min_i d_i(\mathbf{R})) \leq \frac{\partial \phi}{\partial d}(\alpha, K \min_i d_i(\mathbf{R})) \leq C_3 \frac{\partial \phi}{\partial d}(\alpha, \min_i d_i(\mathbf{R}))\,, \tag{52}$$

where C_1, C_2, C_3 are some positive constants independent of n. On the other hand, for $\alpha < 0$,

$$C_1 \phi(\alpha, \min_i d_i(\mathbf{R})) \leq \phi(\alpha, K \min_i d_i(\mathbf{R})) \leq \phi(\alpha, \min_i d_i(\mathbf{R}))\,, \tag{53}$$

and

$$C_3 \frac{\partial \phi}{\partial d}(\alpha, \min_i d_i(\mathbf{R})) \leq \frac{\partial \phi}{\partial d}(\alpha, K \min_i d_i(\mathbf{R})) \leq C_2 \frac{\partial \phi}{\partial d}(\alpha, \min_i d_i(\mathbf{R}))\,, \tag{54}$$

Proof. We first prove (51). $\phi(\alpha, \min_i d_i(\mathbf{R})) \leq \phi(\alpha, K \min_i d_i(\mathbf{R}))$ is by the monotonicity in property 1 of ϕ for $\alpha \geq 0$. For the other side,

$$\begin{aligned} \log \frac{\phi(\alpha, K \min_i d_i(\mathbf{R}))}{\phi(\alpha, \min_i d_i(\mathbf{R}))} &= \log \phi(\alpha, K \min_i d_i(\mathbf{R})) - \log \phi(\alpha, \min_i d_i(\mathbf{R})) \\ &= (K-1) \min_i d_i(\mathbf{R}) \frac{\partial}{\partial d} \log \phi(\alpha, d_0)\,, \end{aligned}$$

for some $d_0 \in [\min_i d_i(\mathbf{R}), K \min_i d_i(\mathbf{R})]$. Then, by property 3 in (7),

$$\begin{aligned} (K-1) \min_i d_i(\mathbf{R}) \frac{\partial}{\partial d} \log \phi(\alpha, d_0) &= (K-1) \min_i d_i(\mathbf{R}) \frac{\partial \phi / \partial d(\alpha, d_0)}{\phi(\alpha, d_0)} \\ &\leq (K-1) K d_0 \frac{\partial \phi / \partial d(\alpha, d_0)}{\phi(\alpha, d_0)} \leq C\,. \end{aligned}$$

Then,

$$\frac{\phi(\alpha, K \min_i d_i(\mathbf{R}))}{\phi(\alpha, \min_i d_i(\mathbf{R}))} = \exp\left(\log \frac{\phi(\alpha, K \min_i d_i(\mathbf{R}))}{\phi(\alpha, \min_i d_i(\mathbf{R}))} \right) \leq e^C\,.$$

The proof of (52) is the same using property 3 in (8). Then, (53) and (54) are implied by (51) and (52), respectively, by reversing all the inequalities, since $\partial \phi / \partial d \leq 0$ for $\alpha < 0$. □

Lemma 8. *For a fixed j,*

$$|d_j(\mathbf{A}) - d_j(\mathbf{R})| \leq \sqrt{8 \max_i d_i(\mathbf{R}) \log n}\,, \tag{55}$$

with probability at least $1 - \frac{2}{n^4}$. *In addition,*

$$|\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_j(\mathbf{R}))| \leq \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,1}) \right| , \tag{56}$$

for some $c_{j,1} \in \left[d_j(\mathbf{R}) - \sqrt{8 \max_i d_i(\mathbf{R}) \log n}, d_j(\mathbf{R}) + \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \right]$ *for each* $j = 1, \ldots, n$, *with probability at least* $1 - \frac{2}{n^4}$.

Proof. Recall $d_j(\mathbf{A}) = \sum_{i=1}^{j} A_{ij}$ and $d_j(\mathbf{R}) = \sum_{i=1}^{j} R_{ij}$. Then, by Berstein inequality,

$$\begin{aligned}
\mathbb{P}\{|d_j(\mathbf{A}) - d_j(\mathbf{R})| \geq t\} &= \mathbb{P}\left\{ \left| \sum_{i=1}^{n} (A_{ij} - R_{ij}) \right| \geq t \right\} \\
&\leq \mathbb{P}\left\{ \sum_{i=1}^{n} (A_{ij} - R_{ij}) \geq t \right\} + \mathbb{P}\left\{ \sum_{i=1}^{n} (R_{ij} - A_{ij}) \geq t \right\} \\
&\leq 2 \exp\left(-\frac{t^2}{2(\max_i d_i(\mathbf{R}) + t/3)} \right) .
\end{aligned}$$

Let $t = \sqrt{8 \max_i d_i(\mathbf{R}) \log n}$, we get:

$$\mathbb{P}\{|d_j(\mathbf{A}) - d_j(\mathbf{R})| \geq \sqrt{8 \max_i d_i(\mathbf{R}) \log n}\} \leq 2e^{-4 \log n} = \frac{2}{n^4} , \tag{57}$$

which is (55). For (56), notice that by mean value theorem,

$$|\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_j(\mathbf{R}))| = |d_j(\mathbf{A}) - d_j(\mathbf{R})| \left| \frac{\partial \phi}{\partial d}(\alpha, c_j) \right| ,$$

where $c_j \in [d_j(\mathbf{R}) - |d_j(\mathbf{A}) - d_j(\mathbf{R})|, d_j(\mathbf{R}) + |d_j(\mathbf{A}) - d_j(\mathbf{R})|\,]$. Then by the previous bound (57), we have

$$|\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_j(\mathbf{R}))| \leq \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,1}) \right| ,$$

for $c_{j,1} \in \left[d_j(\mathbf{R}) - \sqrt{8 \max_i d_i(\mathbf{R}) \log n}, d_j(\mathbf{R}) + \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \right]$ for each $j = 1, \ldots, n$, with probability at least $1 - \frac{2}{n^4}$. $\square$

Lemma 9. *With the assumptions (13)-(15) of Lemma 2,*

$$\|\mathbf{D_A} - \mathbf{D_R}\| \leq C \sqrt{\max_i d_i(\mathbf{R}) \log n} ,$$

with probability at least $1 - \frac{14}{n^2}$ *and C is some fixed positive constant independent of n.*

Proof. Note both $\mathbf{D_A}$ and $\mathbf{D_R}$ are diagonal matrices. We first achieve an entrywise bound. For any $i = 1, \ldots, n$, we have

$$\begin{aligned}
&\left|\frac{\sum_j A_{ij}\,\phi(\alpha, d_j(\mathbf{A}))}{\phi(\alpha, d_i(\mathbf{A}))} - \frac{\sum_j R_{ij}\,\phi(\alpha, d_j(\mathbf{R}))}{\phi(\alpha, d_i(\mathbf{R}))}\right| \\
\leq &\left|\frac{\sum_j A_{ij}\,\phi(\alpha, d_j(\mathbf{A}))}{\phi(\alpha, d_i(\mathbf{A}))} - \frac{\sum_j R_{ij}\,\phi(\alpha, d_j(\mathbf{R}))}{\phi(\alpha, d_i(\mathbf{A}))}\right| + \left|\frac{\sum_j R_{ij}\,\phi(\alpha, d_j(\mathbf{R}))}{\phi(\alpha, d_i(\mathbf{A}))} - \frac{\sum_j R_{ij}\,\phi(\alpha, d_j(\mathbf{R}))}{\phi(\alpha, d_i(\mathbf{R}))}\right| \\
= &\textcircled{1} + \textcircled{2}. \qquad (58)
\end{aligned}$$

Define $d_{j,-i}(\mathbf{A}) = d_j(\mathbf{A}) - A_{ij}$. Note $d_{j,-i}(\mathbf{A})$ is independent of A_{ij}. Then, the numerator of the first term $\textcircled{1}$ in (58) can be split as:

$$\begin{aligned}
&\left|\sum_j A_{ij}\,\phi(\alpha, d_j(\mathbf{A})) - \sum_j R_{ij}\,\phi(\alpha, d_j(\mathbf{A}))\right| \\
\leq &\left|\sum_j A_{ij}\left(\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{A}))\right)\right| + \left|\sum_j (A_{ij} - R_{ij})\phi(\alpha, d_{j,-i}(\mathbf{A}))\right| \\
&+ \left|\sum_j R_{ij}\left[\phi(\alpha, d_{j,-i}(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{R}))\right]\right| + \left|\sum_j R_{ij}\left[\phi(\alpha, d_j(\mathbf{R})) - \phi(\alpha, d_{j,-i}(\mathbf{R}))\right]\right| \\
= &\,\mathbf{G} + \mathbf{H} + \mathbf{J} + \mathbf{K}. \qquad (59)
\end{aligned}$$

Bound for G in (59):

By mean value theorem and Lemma 8,

$$\begin{aligned}
&\left|\{\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{A}))\} - \{\phi(\alpha, d_j(\mathbf{R})) - \phi(\alpha, d_{j,-i}(\mathbf{R}))\}\right| \\
\leq &\left|\{\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_j(\mathbf{R}))\} - \{\phi(\alpha, d_{j,-i}(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{R}))\}\right| \\
\leq &\,2\max\left(\left|\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_j(\mathbf{R}))\right|, \left|\phi(\alpha, d_{j,-i}(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{R}))\right|\right) \\
\leq &\,2\sqrt{8\max_i d_i(\mathbf{R})\log n}\left|\frac{\partial\phi}{\partial d}(\alpha, c_{j,2})\right|,
\end{aligned}$$

for some $c_{j,2} \in \left[d_j(\mathbf{R}) - \sqrt{8\max_i d_i(\mathbf{R})\log n} - 1, d_j(\mathbf{R}) + \sqrt{8\max_i d_i(\mathbf{R})\log n}\right]$ for each

$j = 1, \dots, n$, with probability at least $1 - \frac{2}{n^4}$. Then, by triangular inequality,

$$|\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{A}))| \leq |\phi(\alpha, d_j(\mathbf{R})) - \phi(\alpha, d_{j,-i}(\mathbf{R}))| + 2\sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,2})\right|$$
$$\leq \left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,3})\right| + 2\sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,2})\right| ,$$

for some $c_{j,3} \in [d_{j,-i}(\mathbf{R}), d_j(\mathbf{R})]$ for each $j = 1, \dots, n$, with probability at least $1 - \frac{2}{n^4}$. Note the bound obtained is not optimal but is more than enough for the proof of the Lemma. Consider the event:

$$\Lambda = \bigcap_j \left\{ |\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{A}))| \leq \left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,3})\right| + 2\sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,2})\right| \right\} .$$

We see that by union bound[12], $\mathbb{P}\{\Lambda\} \geq 1 - \frac{2}{n^3}$. Denote $\mathbb{P}_S\{T\} = \mathbb{P}\{S \cap T\}$ for events S and T.

Let $t = 2\left(\log n \sum_j \left(\left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,3})\right| + 2\sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,2})\right|\right)^2\right)^{1/2}$. Then, by Hoeffding's inequality, we have

$$\begin{aligned}
&\mathbb{P}\left\{\left|\sum_j (A_{ij} - R_{ij})\left(\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{A}))\right)\right| \geq t\right\} \\
&\leq \mathbb{P}_\Lambda\left\{\left|\sum_j (A_{ij} - R_{ij})\left(\phi(\alpha, d_j(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{A}))\right)\right| \geq t\right\} + \mathbb{P}\{\Lambda^C\} \\
&\leq 2\exp\left(-\frac{4 \log n \sum_j \left(\left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,3})\right| + 2\sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,2})\right|\right)^2}{\sum_j \left(\left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,3})\right| + 2\sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left|\frac{\partial \phi}{\partial d}(\alpha, c_{j,2})\right|\right)^2}\right) + \frac{2}{n^3} \\
&\leq 2\exp(-4 \log n) + \frac{2}{n^3} \\
&\leq \frac{4}{n^3} . \qquad (60)
\end{aligned}$$

Bound for H in (59):

[12] Consider events $E_1, \dots, E_n$ and $\mathbb{P}(E_j) \geq 1 - c_j$ for all $j = 1, \dots, n$. Then, $\mathbb{P}(\cap_j E_j) = \mathbb{P}((\cup_j E_j^C)^C) = 1 - \mathbb{P}(\cup_j E_j^C) \geq 1 - \sum_j \mathbb{P}(E_j^C) \geq 1 - \sum_j c_j$.

By mean value theorem and Lemma 8,

$$
\begin{aligned}
|\phi(\alpha, d_{j,-i}(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{R}))| &\leq \sqrt{8 \max_i d_i(\mathbf{R}) \log(n-1)} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,4}) \right| \\
&\leq \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,4}) \right| ,
\end{aligned}
$$

for some $c_{j,4} \in \left[d_j(\mathbf{R}) - \sqrt{8 \max_i d_i(\mathbf{R}) \log n} - 1, d_j(\mathbf{R}) + \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \right]$ for each $j = 1, \ldots, n$, with probability at least $1 - \frac{2}{n^4}$. Consider the event:

$$
\Gamma = \bigcap_j \left\{ \phi(\alpha, d_{j,-i}(\mathbf{A})) \leq \phi(\alpha, d_j(\mathbf{R})) + \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,4}) \right| \right\} .
$$

With the exact same proof in Lemma 8 and union bound, $\mathbb{P}\{\Gamma\} \geq 1 - \frac{2}{n^3}$.

Let $t = 2 \left(\log n \sum_j \left(\phi(\alpha, d_j(\mathbf{R})) + \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,4}) \right| \right)^2 \right)^{1/2}$.

Then, by Hoeffding's inequality, we have

$$
\begin{aligned}
&\mathbb{P} \left\{ \left| \sum_j (A_{ij} - R_{ij}) \phi(\alpha, d_{j,-i}(\mathbf{A})) \right| \geq t \right\} \\
&\leq \mathbb{P}_\Gamma \left\{ \left| \sum_j (A_{ij} - R_{ij}) \phi(\alpha, d_{j,-i}(\mathbf{A})) \right| \geq t \right\} + \mathbb{P}\{\Gamma^C\} \\
&\leq 2 \exp \left(- \frac{4 \log n \sum_j \left(\phi(\alpha, d_j(\mathbf{R})) + \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,4}) \right| \right)^2}{\sum_j \left(\phi(\alpha, d_j(\mathbf{R})) + \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,4}) \right| \right)^2} \right) + \frac{2}{n^3} \\
&\leq 2 \exp(-4 \log n) + \frac{2}{n^3} \\
&= \frac{4}{n^3} .
\end{aligned}
\tag{61}
$$

Bound for J in (59):

Under the event Γ,

$$
\left| \sum_j R_{ij} \left[\phi(\alpha, d_{j,-i}(\mathbf{A})) - \phi(\alpha, d_{j,-i}(\mathbf{R})) \right] \right| \leq \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \sum_j R_{ij} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,4}) \right| .
\tag{62}
$$

Thus, the inequality holds with probability at least $1 - \frac{2}{n^3}$.

Bound for K in (59):

By mean value theorem,

$$\left| \sum_j R_{ij} \left[\phi(\alpha, d_j(\mathbf{R})) - \phi(\alpha, d_{j,-i}(\mathbf{R})) \right] \right| \leq \sum_j R_{ij} R_{ij} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,5}) \right| \leq \sum_j R_{ij} \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,5}) \right| , \tag{63}$$

for some $c_{j,5} \in [d_{j,-i}(\mathbf{R}), d_j(\mathbf{R})]$ for each $j = 1, \ldots, n$.

Bound for ① in (58):

The assumptions of Lemma 2 are applied here. Note that assumptions 2 and 3 would imply that $d_j(\mathbf{R})$ for all j are with the same order. This is also true for $\phi(\alpha, d_j(\mathbf{R}))$ and $\frac{\partial \phi}{\partial d}(\alpha, d_j(\mathbf{R}))$ for all j, by Lemma 7. Hence, by assumption 1 of Lemma 2 together with the property 3 (7) of the function ϕ, we combine the 4 bounds (60) – (63) to get a bound for numerator of ①:

$$\left| \sum_j A_{ij} \, \phi(\alpha, d_j(\mathbf{A})) - \sum_j R_{ij} \, \phi(\alpha, d_j(\mathbf{A})) \right| \leq C \sqrt{\max_i d_i(\mathbf{R}) \log n} \, \phi(\alpha, d_i(\mathbf{R})) \, , \tag{64}$$

with probability at least $1 - \frac{10}{n^3}$ and C is a fixed constant independent of n. By Lemma 8 and the triangular inequality,

$$\begin{aligned} \frac{1}{\phi(\alpha, d_i(\mathbf{A}))} &\leq \frac{1}{\phi(\alpha, d_i(\mathbf{R})) - |\phi(\alpha, d_i(\mathbf{A})) - \phi(\alpha, d_i(\mathbf{R}))|} \\ &\leq \frac{1}{\phi(\alpha, d_i(\mathbf{R})) - \sqrt{8 \max_i d_i(\mathbf{R}) \log n} \, \left| \frac{\partial \phi}{\partial d}(\alpha, c_{j,1}) \right|} \, , \end{aligned}$$

with probability at least $1 - \frac{2}{n^4}$. Therefore,

$$① = \left| \frac{\sum_j A_{ij} \, \phi(\alpha, d_j(\mathbf{A}))}{\phi(\alpha, d_i(\mathbf{A}))} - \frac{\sum_j R_{ij} \, \phi(\alpha, d_j(\mathbf{R}))}{\phi(\alpha, d_i(\mathbf{A}))} \right| \leq C \sqrt{\max_i d_i(\mathbf{R}) \log n} \, ,$$

with probability at least $1 - \frac{12}{n^3}$.

Bound for ② in (58):

By Lemma 8 and the bound for $\frac{1}{\phi(\alpha, d_i(\mathbf{A}))}$,

$$
\begin{aligned}
② &= \left| \frac{\sum_j R_{ij}\, \phi(\alpha, d_j(\mathbf{R}))}{\phi(\alpha, d_i(\mathbf{A}))} - \frac{\sum_j R_{ij}\, \phi(\alpha, d_j(\mathbf{R}))}{\phi(\alpha, d_i(\mathbf{R}))} \right| \\
&= \sum_j R_{ij}\, \phi(\alpha, d_j(\mathbf{R})) \frac{|\phi(\alpha, d_i(\mathbf{A})) - \phi(\alpha, d_i(\mathbf{R}))|}{\phi(\alpha, d_i(\mathbf{A}))\phi(\alpha, d_i(\mathbf{R}))} \\
&\leq \sum_j R_{ij}\, \phi(\alpha, d_j(\mathbf{R})) \frac{\sqrt{8 \max_i d_i(\mathbf{R}) \log n}\ \left|\frac{\partial \phi}{\partial d}(\alpha, c_{i,0})\right|}{\phi(\alpha, d_i(\mathbf{R})) \left(\phi(\alpha, d_i(\mathbf{R})) - \sqrt{8 \max_i d_i(\mathbf{R}) \log n}\ \left|\frac{\partial \phi}{\partial d}(\alpha, c_{i,0})\right|\right)} \\
&\leq C d_i(\mathbf{R}) \frac{\sqrt{4 \max_i d_i(\mathbf{R}) \log n}\ \left|\frac{\partial \phi}{\partial d}(\alpha, c_{i,0})\right|}{\phi(\alpha, d_i(\mathbf{R}))} \\
&\leq C \sqrt{\max_i d_i(\mathbf{R}) \log n}\,, \qquad (65)
\end{aligned}
$$

with probability at least $1 - \frac{2}{n^4}$.

Bound for ① + ② in (58):

Combining (64) and (65), we get the entrywise bound. For any i,

$$
\left| \frac{\sum_j A_{ij}\, \phi(\alpha, d_j(\mathbf{A}))}{\phi(\alpha, d_i(\mathbf{A}))} - \frac{\sum_j R_{ij}\, \phi(\alpha, d_j(\mathbf{R}))}{\phi(\alpha, d_j(\mathbf{R}))} \right| \leq C \sqrt{\max_i d_i(\mathbf{R}) \log n}\,,
$$

with probability at least $1 - \frac{14}{n^3}$. Finally, by union bound,

$$
\|\mathbf{D_A} - \mathbf{D_R}\| \leq C \sqrt{\max_i d_i(\mathbf{R}) \log n}\,,
$$

with probability at least $1 - \frac{14}{n^2}$. □

Proof of Proposition 7

Rewrite the left hand side as:

$$\begin{aligned}\left\|\mathbf{D}_{\mathbf{A}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\right\| &= \left\|\left(\mathbf{D}_{\mathbf{A}}^{-1}-\mathbf{D}_{\mathbf{R}}^{-1}\right)\left(\mathbf{D}_{\mathbf{A}}^{-1/2}+\mathbf{D}_{\mathbf{R}}^{-1/2}\right)^{-1}\right\| \\ &= \left\|\left(\mathbf{D}_{\mathbf{A}}-\mathbf{D}_{\mathbf{R}}\right)\left(\mathbf{D}_{\mathbf{A}}\mathbf{D}_{\mathbf{R}}\right)^{-1}\left(\mathbf{D}_{\mathbf{A}}^{-1/2}+\mathbf{D}_{\mathbf{R}}^{-1/2}\right)^{-1}\right\| \\ &\leq \left\|\mathbf{D}_{\mathbf{A}}-\mathbf{D}_{\mathbf{R}}\right\|\cdot\left\|\mathbf{D}_{\mathbf{A}}^{-1}\right\|\cdot\left\|\mathbf{D}_{\mathbf{R}}^{-1}\right\|\cdot\left\|\left(\mathbf{D}_{\mathbf{A}}^{-1/2}+\mathbf{D}_{\mathbf{R}}^{-1/2}\right)^{-1}\right\|\,.\end{aligned}$$

Condition on the event $\|\mathbf{D}_{\mathbf{A}}-\mathbf{D}_{\mathbf{R}}\|\leq C\sqrt{n\log n}\,$,

$$\begin{aligned}\left\|\left(\mathbf{D}_{\mathbf{A}}^{-1/2}+\mathbf{D}_{\mathbf{R}}^{-1/2}\right)^{-1}\right\| &= \max_i\left((\mathbf{D}_{\mathbf{A}})_{ii}^{-1/2}+(\mathbf{D}_{\mathbf{R}})_{ii}^{-1/2}\right)^{-1} \\ &= \left(\min_i(\mathbf{D}_{\mathbf{A}})_{ii}^{-1/2}+(\mathbf{D}_{\mathbf{R}})_{ii}^{-1/2}\right)^{-1} \\ &\leq \left(\min_i(\mathbf{D}_{\mathbf{R}})_{ii}^{-1/2}+((\mathbf{D}_{\mathbf{R}})_{ii}+|(\mathbf{D}_{\mathbf{A}})_{ii}-(\mathbf{D}_{\mathbf{R}})_{ii}|)^{-1/2}\right)^{-1} \\ &\leq C\left(\min_i(\mathbf{D}_{\mathbf{R}})_{ii}^{-1/2}\right)^{-1} \\ &\leq C\max_i\sqrt{d_i(\mathbf{R})} \\ &\leq C\min_i\sqrt{d_i(\mathbf{R})}\,.\end{aligned}$$

Similarly,

$$\left\|\mathbf{D}_{\mathbf{A}}^{-1}\right\| = \left(\min_i(\mathbf{D}_{\mathbf{A}})_{ii}\right)^{-1}\leq C\left(\min_i d_i(\mathbf{R})\right)^{-1}\,.$$

Combining the results and applying Lemma 9, we have

$$\left\|\mathbf{D}_{\mathbf{A}}^{-1/2}-\mathbf{D}_{\mathbf{R}}^{-1/2}\right\|\leq C\sqrt{\max_i d_i(\mathbf{R})\log n}\left(\min_i d_i(\mathbf{R})\right)^{-3/2}\,,$$

with probability at least $1-\frac{14}{n^2}$. ■

Proof of Proposition 4

The fact that Theorem 1 and Theorem 2 holds for $\tilde{\mathbf{T}}$ follows trivially from the proofs of Theorem 1, Lemma 2 and Theorem 2. ■