Linear Regression with Centrality Measures

Job Market Paper

Yong Cai *
Department of Economics
Northwestern University
yongcai2023@u.northwestern.edu

This version: October 18, 2022.
Newest version here.

Abstract

This paper studies the properties of linear regression on centrality measures when network data is sparse – that is, when there are many more agents than links per agent – and when they are measured with error. We make three contributions in this setting: (1) We show that OLS estimators can become inconsistent under sparsity and characterize the threshold at which this occurs, with and without measurement error. This threshold depends on the centrality measure used. Specifically, regression on eigenvector is less robust to sparsity than on degree and diffusion. (2) We develop distributional theory for OLS estimators under measurement error and sparsity, finding that OLS estimators are subject to asymptotic bias even when they are consistent. Moreover, bias can be large relative to their variances, so that bias correction is necessary for inference. (3) We propose novel bias correction and inference methods for OLS with sparse noisy networks. Simulation evidence suggests that our theory and methods perform well, particularly in settings where the usual OLS estimators and heteroskedasticity-consistent/robust t-tests are deficient. Finally, we demonstrate the utility of our results in an application inspired by De Weerdt and Dercon (2006), in which we consider consumption smoothing and social insurance in Nyakatoke, Tanzania.

Keywords: networks, diffusion centrality, eigenvector centrality.

JEL Classification Codes: C18, C21, C81.

*I am grateful to Ivan Canay, Eric Auerbach and Joel Horowitz for guidance on this project. I have also benefited from the comments and suggestions of Eduardo Campillo-Betancourt, Piotr Dworczak, Grant Goehring and Deborah Kim.

1 Introduction

A large and rapidly growing body of work documents the influence of networks in a wide range of economic outcomes: peer effects drive academic achievement, production networks shape shock propagation in the macroeconomy, social networks affect information- and risk-sharing with important implications for development (see Sacerdote 2011, Carvalho and Tahbaz-Salehi 2019 and Breza et al. 2019 for recent reviews). Many other examples abound.

One particular strand of research has fruitfully explored the relationship between an agent's network position and their economic outcomes. For example, Hochberg et al. (2007) considers the network of venture capital firms and finds that better-networked firms successfully exit a greater proportion of their investments. Meanwhile, Cruz et al. (2017) examines the social networks in the Philippines and shows that more central families are disproportionately represented in political offices. Similarly, Banerjee et al. (2013) studies the problem of diffusing microfinance in India and establishes that seeding information to more central agents led to greater participation in the program.

In these papers, researchers often estimate linear models by ordinary least squares (OLS), using centrality measures as explanatory variables. Centrality measures are node-level statistics that capture notions of importance in a network. Since nodes can be important for many reasons, a variety of centrality measures exist, each capturing a particular aspect of network position. For example, the degree centrality of an agent reflects the number or intensity of their direct links, while eigenvector centrality is designed so that influence of agents is proportional to that of their connections. The correlation between an outcome variable and a particular centrality measure may be revealing about the types of interactions that drive a given economic phenomenon: an outcome that is well-predicted solely by degree is likely to be determined in an extremely local manner, whereas one that is more strongly associated with eigenvector centrality may involve non-linear interactions between agents that are further apart. As such, when researchers estimate these correlations and test their statistical significance, they frequently do so with the goal of drawing conclusions about the economic significance of various centrality measures and the implied mechanisms for outcome determination. Such an exercise is credible only if the OLS estimator is close to the estimand, and if the chosen test statistic (typically the heteroskedasticity-consistent/robust t-statistics) is well described by its asymptotic distribution (standard normal for t-statistics) in finite sample.

However, network data have two features that may threaten the statistical validity of OLS. Firstly, networks may be sparse, with many more agents than links per agent. This could happen because interactions are observed with low frequency, or because the interactions in question are rare. Chandrasekhar (2016) argues that many economic networks are sparse, providing evidence from commonly used social network data (e.g. AddHealth; Karnataka Villages (Banerjee et al. 2013); Harvard social network (Leider et al. 2009)). Sparsity poses a challenge to estimation and inference: if networks are largely empty, there might not be enough variation in centrality measures to identify the parameters of interest. Despite its importance, sparsity has received relatively little attention in the network econometrics literature.

Secondly, the observed network may differ from the true network of interest. Centrality measures are often calculated on data which are obtained by survey or constructed using some proxy for interaction between agents, though subsequent analysis would frequently treat the true network as known. Ignoring measurement error may thus lead to estimates that perform poorly. A growing literature works with networks that are assumed to be measured with error. However, they generally do not consider sparse settings. This is important since sparsity and measurement error are mutually reinforcing: sparser networks contain weaker signals, which are in turn more difficult to pick out from noisy measurements. The upshot is that OLS estimators computed on sparse, noisy networks may have particularly poor properties. Asymptotic theory that ignore these features will provide similarly poor approximations to their finite sample behavior. Consequently, estimation and inference procedures based on these theories may lead to invalid conclusions about the economic significance of centrality measures.

This paper studies the statistical properties of OLS on centrality measures in an asymptotic framework which features both measurement error and sparsity. Our analysis is centered on degree, diffusion and eigenvector centralities, which are among the most popular measures. Our contribution is threefold: (1) We characterize the amount of sparsity at which OLS estimators become inconsistent with and without measurement error, finding that this threshold varies depending on the centrality measure used. Specifically, regression on eigenvector centrality is less robust to sparsity than that on degree and diffusion. This suggests that researchers should be cautious about comparing regressions on different centrality measures, since they may differ in statistical properties in addition to economic significance. (2) We develop distributional theory for OLS estimators under measurement error and sparsity. We restrict ourselves to sparsity ranges under which OLS is consistent, but we find that asymptotic bias can be large even in this case. Furthermore, the bias may be of larger order than variance, in which case bias correction would be necessary for obtaining non-degenerate asymptotic distributions. Additionally, we find that under sparsity, the estimator converges at a slower rate than is reflected by the usual heteroskedasticity-consistent(hc)/robust standard errors, requiring a different estimator. (3) In view of the distributional theory, we propose novel bias-corrected estimators and inference methods for OLS with sparse, noisy networks. We also clarify the settings under which hc/robust t-statistics are appropriate for testing.

Our theoretical results are derived in an asymptotic framework where networks are modeled as realizations of sparse random graphs. As $n \to \infty$, the expected number of links per agent grows much more slowly than n. Because our statistical model captures important features of real world data, we expect our methods to be reliable for estimation and inference with sparse, noisy networks. We provide simulation evidence supporting this view. The utility of our results is also evident from an application inspired by De Weerdt and Dercon (2006), where we conduct a stylized study of consumption smoothing and social insurance in Nyakatoke, Tanzania.

Our choice of asymptotic framework poses technical challenges. Firstly, the eigenvectors and eigenvalues of sparse random graphs are difficult to characterize. We draw on recent advances in random matrix theory (Alt et al. 2021a;b; Benaych-Georges et al. 2019; 2020) to overcome this challenge. Secondly, spectral norms of random matrices concentrate slowly in sparse regimes. Instead, we develop bounds for moments of noisy adjacency matrices by relating them to counts of particular graphs, in the spirit of Wigner (1957) (see Chapter 2 of Tao 2012 more generally). Finally, in order for bias correction to improve mean-squared error, the bias needs to be estimated at a sufficiently fast rate. Because variance is of lower order than bias, a naive plug-in approach does not work for estimating higher order bias terms, although it is sufficient for the first order term. We leverage this fact to recursively construct good estimators for higher order terms.

Related Literature

Our work is most closely related to papers that study linear regression with centrality statistics. To our best knowledge, we are the first to study linear regression with diffusion centrality, though there exist prior work on eigenvector centrality. Le and Li (2020) studies linear regression on multiple eigenvectors of a network assuming the same type of measurement error as this paper. They focus on denser settings than we do and provide inference method only for the null hypothesis that the slope coefficient is 0. We are concerned only with eigenvector centrality, which is the leading eigenvector, but our results cover the sparse case as well as tests of non-zero null hypotheses (more details in Remark 5). Our paper is also related to Cai et al. (2021), which proposes penalized regressions on the leading left and right singular vectors of a network. They consider networks that are as sparse as the ones we study, but their networks are observed with an additive, normally distributed error (more details in Remark 4). Outside of the linear regression setting, Cheng et al. (2021) considers inference on deterministic linear functionals of eigenvectors. They study symmetric matrices with asymmetric noise, proposing novel estimators that leverage asymmetry to improve performance when eigengaps are small. We focus on symmetric matrices with symmetric noise and study the plug-in estimator in which eigenvector is estimated using the noisy adjacency matrix in place of the true matrix.

Our paper also relates to a growing literature that considers sampling and measurement error in networks. Chandrasekhar and Lewis (2016) examines settings in which researchers have access to a panel of networks, but which are constructed using only a partial sample of nodes or edges. Thirkettle (2019) studies a similar missing data problem, but in a cross-sectional setting with only one network. It is concerned with forming bounds on centrality statistics and does not consider subsequent linear regression. Griffith (2022) considers the censoring in network data, which arises when agents are only allowed to list a fixed number of relationships during the sampling process. The above papers study missing data problems under the assumption that the observed network is without error. We assume that the entirety of one network is observed but with error. Lewbel et al. (2021) studies measurement error in peer effects regression, finding that 2SLS with friends-of-friends instruments is valid as long as measurement error is small. They do not discuss centrality regressions.

This paper is also connected to the nascent literature on the statistical properties of sparse networks. A strand of this literature is concerned with network formation models that can give rise to sparsity in the observed data. Dong et al. (2020) and Motalebi et al. (2021) consider modifications to the stochastic block model. A more general model takes the form of inhomogeneous Erdos-Renyi graph, which are generated by a graphon with a sparsity parameter that tends to zero in the limit (see for instacne Bollobás et al. 2007 and Bickel and Chen 2009). Our paper takes this approach. Yet another model for sparse graphs is based on graphex processes, which generalizes graphons by generating vertices through Poisson point processes (see Borgs et al. 2018, Veitch and Roy 2019 and references therein). Our choice of inhomogeneous Erdos-Renyi graphs is motivated by their prevalence in econometrics (Section 3 of De Paula 2017 and Section 6 of Graham 2020a provide many examples), as well as tractability considerations. To our best knowledge, few papers have tackled the challenges that sparse networks pose for regression. Two notable exceptions study network formation models, which take the form of edge-level logistic regressions (Jochmans 2018; Graham 2020b). A separate literature considers estimation of peer effects regressions involving sparse networks using panel data (Manresa 2016; Rose 2016; De Paula et al. 2020). Here, sparsity is an assumption used to justify regularization methods. We consider a node-level regression in a cross-sectional setting with one large network.

More generally, we contribute to the literature on measurement error models, in which economic outcomes are driven by unobserved latent variables, although proxies or noisy measurements of these variables exist. For recent reviews, see Hu (2017) and Schennach (2020). Classical measurement error is an additive noise that is (conditionally) independent of the unobserved regressor and has a long history (e.g. Adcock 1878). When the underlying network is noisily observed, centrality statistics face errors which are non-linear and non-separable. General non-classical measurement error problems have been studied in cross-sectional (e.g. Matzkin 2003; Chesher 2003; Evdokimov and Zeleneev 2022) and panel data settings (e.g. Griliches and Hausman 1986; Evdokimov 2010; Evdokimov and Zeleneev 2020). These papers typically assume that observations can be grouped into units across which the latent variable and measurement error are independent. This excludes our setting, since centrality statistics of any given agent depends on the entire network.

The rest of this paper is organized as follows. Section 2 describes the set-up of our paper. Section 3 presents the theoretical results. Simulation results are contained in Section 4. In Section 5, we apply our results to the social insurance network in Nyakatoke, Tanzania. Section 6 concludes the paper with our recommendations for empirical work. All proofs are contained in Appendix A.

2 Set-Up and Notation

In this section, we introduce notation before describing our econometric model and the asymptotic framework.

We use the following notation. When X is a vector or matrix, X_i and X_{ij} refer the i^{th} and $(i,j)^{\text{th}}$ component of X respectively. Similarly, if X_i or X_{ij} are defined, we use X to denote the full vector or matrix respectively. X' is the transpose of X. When X is a square matrix, $\lambda_j(X)$ denotes the j^{th} eigenvalue of X while $v_j(X)$ denotes the corresponding eigenvector. When $f \in L^2([0,1]^2)$ is a symmetric real function, $\lambda_j(f)$ denotes the j^{th} eigenvalue of the corresponding Hillbert-Schmidt integral operator, $T(g) = \int f(x,y)g(y)dy$, while ϕ_j is the corresponding eigenfunction. For deterministic, monotone sequences x_n and y_n, we write $x_n \succ z_n$ if $x_n/z_n \to \infty$ and $x_n \prec z_n$ if $x_n/z_n \to 0$. $x_n \approx z_n$ indicates that $x_n/z_n \to k$, where $0 < k < \infty$. We write $x_n \succcurlyeq z_n$ to mean $\neg(x_n \prec z_n)$ and similarly for $x_n \preccurlyeq z_n$. Let ι_n be the $n \times 1$ vector of 1's. For two $m \times n$ matrices X and Z, let $X \circ Z$ denote their entrywise (Hadamard) product. Finally, $[n]$ denotes the set of integers from 1 to n.

2.1 Econometric Framework

For simplicity, suppose that there are no covariates besides centrality. Suppose also that the data-generating process yields $\{(Y_i, U_i)\}_{i=1}^n$ which are independent and identically distributed. Y_i is the observed outcome of interest. U_i is an unobserved latent type that will be used to construct the network. In lieu of U_i, we observe either the true adjacency matrix A, or a noisy version, $\hat{A}$. A is an $n \times n$ matrix whose $(i,j)^{\text{th}}$, A_{ij}, records the link intensities between agents i and j. $\hat{A}$ is some estimate of A.

Consider the regression:

$$Y_i = \beta^{(d)} C_i^{(d)} + \varepsilon_i$$

where $C_i^{(d)}$ is a network centrality measure of type d. We do not observe $C_i^{(d)}$, it can be exactly computed using A, or estimated using $\hat{A}$. The parameter of interest is $\beta^{(d)}$. After defining the data-generating process for the latent and observed networks, Assumption 3 will provide conditions allowing us interpret $\beta^{(d)}$ as the slope coefficient in the linear conditional expectation function of Y_i on $C_i^{(d)}$.

In the following, we describe (i) the data-generating process for A and $\hat{A}$ via the U_i's and (ii) the use of A and $\hat{A}$ in computing/estimating centrality statistics for OLS estimation. Throughout our discussion, we motivate econometric framework through the example of consumption smoothing via informal insurance:

Example 1. Suppose we are interested in the relationship between informal insurance and consumption smoothing. This is a question that has been studied by De Weerdt and Dercon (2006); Udry (1994); Kinnan and Townsend (2012) and Bourlès et al. (2021) among many others. Here, we might posit that agents which are more central in the informal insurance network can better smooth consumption. To test this hypothesis, we are interested in the regression where Y_i is variance in i's consumption and $C_i^{(d)}$ is centrality in the informal insurance network. $\beta^{(d)}$ is then the reduction in consumption variance associated with being more central. In the informal insurance network, A_{ij} records the probability that i lends money to j or vice versa in the event of an adverse income shock. However, A is hard to obtain by surveys. Instead, we observe the matrix of actual loans $\hat{A}$, which is a noisy measure of A.

Data-Generating Process for A and $\hat{A}$

Let A be an $n \times n$ symmetric adjacency matrix. We assume that the relationship between two agents in a network is solely determined by their unobserved latent types U_i through the graphon f:

Assumption 1 (Graphon)**.** Suppose $U_i \sim U[0,1]$ and $f : [0,1]^2 \mapsto [0,1]$ is such that:

$$\int_{[0,1]^2} f(u,v)\, du\, dv > 0.$$

Let $p_n \in (0,1]$ and $j > i$, define:

$$A_{ij} = p_n f(U_i, U_j) \ .$$

We set $A_{ji} = A_{ij}$ for $j < i$ and normalize $A_{ii} = 0$ for all $i \in [n]$.

In this model, any two agents have a relationship that is between 0 and 1. We can think of this as a measure of intensity, reflecting factors such as duration of friendship, frequency of interaction, or similarity in personalities. Alternatively, it could be the probability with which a relationship is observed. p_n is a parameter that we will let go to 0 at various rates. As we will explain in Section 2.2, this is a theoretical device that will help us understand the behavior of the OLS estimator when the network is sparse. We restrict our attention to symmetric matrices because eigenvector centrality, one of the most popular network centrality measures, may not be well-defined when the adjacency is not symmetric. We also ignore the trivial case when $f = 0$, in which case the network is always empty. Finally, we note that defining $U_i \sim U[0,1]$ is without loss of generality since we have placed no functional form restrictions on f.

Example 1 (continued)**.** Suppose that $U_i \in [0,1]$ indexes the riskiness of a villager's income as a result of the crops they choose to cultivate. Assumption 1 posits that the relationship between two villagers depends only on their respective income risks. For example, if $f(U_i, U_j) = (U_i - U_j)^2$, then farmers with similar income risks have higher link intensities between them. U_i can also incorporate other observed or unobserved farmer characteristics, such as place of residence, farming skills or gregariousness. Together with the choice of f, the graphon is a rich model of linking behavior.

When A is observed, we say that there is no measurement error. This setting provides a useful benchmark. When A is not observed, we assume that we have access to the noisy version, $\hat{A}$, generated as follows.

Assumption 2 (Measurement Error)**.** The adjacency matrix with measurement error is the $n \times n$ matrix symmetric $\hat{A}$, where for $j > i$,

$$\hat{A}_{ij} \,|\, U \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}\,(A_{ij}) \ .$$

We set $\hat{A}_{ji} = \hat{A}_{ij}$ for $j < i$. $\hat{A}_{ii} = 0$ since $A_{ii} = 0$. Furthermore,

$$\hat{A}_{ij} \perp\!\!\!\perp \varepsilon_i \mid U \ .$$

The form of measurement error we consider randomly rounds A_{ij} into zero or one in proportion to the intensity of the true relationship. Conditional on U, this is an additive error with a mean of 0, and which is independent across agent pairs. Formally, we are assuming a conditionally independent dyad (CID) model for $\hat{A}$. This model is commonly used in econometrics (see for instance Section 3 of De Paula 2017 or Section 6 of Graham 2020a) and is fairly general. By the Aldous-Hoover Theorem (Aldous 1981; Hoover (1979)), all infinitely exchangeable random graphs have such a representation. Roughly, this corresponds to sampling from a large population in which the labels of agents do not matter. Measurement error of this form can arise due to limitations in data collection, or because networks are constructed by proxy. We provide examples of data that fit our measurement error model at the end of this subsection. Importantly, however, it excludes strategic network formation for $\hat{A}$, in which agents' decision to form or report links depends on that of the others.

Finally, we also assume that measurement error is independent of ε_i conditional on U. Together with the CID assumption, measurement error on the network is additive white noise, akin to the type studied in classical measurement error models. In our setting, the key econometric challenge arises because U is unobserved. This is exacerbated by the fact that additive, white noise errors in the network translates into non-linear measurement error in centrality statistics, introducing complications in the analysis.

Remark 1. The challenges that this form of measurement error poses are most severe in cross-sectional data. Intuitively, given long panel data on the networks, we would be able to estimate A_{ij} well using entry-wise averages of the networks over time.

Example 1 (continued)**.** Assumption 2 is reasonable in the context of our leading example. Here, each entry of the unobserved A_{ij} represents the probability of loans. However, $\hat{A}_{ij}$ records actual loans, which are realizations of Bernoulli(A_{ij}). The conditional independence assumption means that conditional on friendship, the decision of i to lend to j is independent of the decision of k to lend to i. This might be the case if the loan amounts are small relative to the income shortfall, so that any agent's decision to lend to i does not significantly reduce their need to borrow. Alternatively, such a condition might be satisfied if borrowing is private, so that friends of i cannot coordinate their lending decisions.

The remainder of this subsection lists examples of network data that could be described by our data-generating process.

Example 2. Carvalho et al. (2021) studies the propagation of shocks through production networks during the Great East Japanese Earthquake of 2011. In the ideal production network, A_{ij} records the value of i's sales to j as a proportion of the value of i's total sales. In turn, A_{ij} depends on U_i and U_j, which might index the quality of a firm's product, with higher quality firms requiring more and higher quality inputs. However, these variables are not observed. Instead, the authors have access to data from a credit reporting agency which includes supplier and customer information for firms. The authors explicitly note two limitations in their data: "First, it only reports a binary measure of interfirm supplier-customer relations... we do not observe a yen measure associated with their transactions. Second, the forms used by [the credit agency] limit the number of suppliers and customers that firms can report to 24 each." Suppose firms only report suppliers from whom they receive delivery during the month in which the forms are filed. Then a supplier that sends fewer inputs are more likely to be omitted in any given month. Abstracting away from concerns about network censoring (see Griffith 2022), the conditional independence assumption would also be satisfied if the delivery schedules for suppliers are independent.

Example 3. Xu (2018) studies how patronage affected the promotion and performance of bureaucrats in the Colonial Office of the British Empire. In the ideal network for measuring patronage, A_{ij} records intensity of the friendship between i and j. Here, U_i might index traits such as gregariousness, polo skills and drinking habits among others. Bureaucrats having more in common with their patrons may be more likely to be recommended for promotion. However, the link intensity between bureaucrats are not observed. Instead, the paper proxies for relationships using indicators for shared ancestry, membership of social groups (such as the aristocracy) or attendance of the same elite school or university. In this context, our data-generating process means that bureaucrats who are closer are more likely to satisfy the above criteria for connection. The conditional independence assumption would be satisfied if the lack of observation are independent across agent pairs, e.g. if some university records were randomly lost.

Centrality Statistics and OLS Estimation

Given our adjacency matrices A and $\hat{A}$, we now define centrality statistics and the OLS estimators that are based on them.

Centrality measures are agent-level measures of importance in a network. Many centrality measures exist, each capturing a different aspect of network position. However, they are all functions of A and can be exactly computed in the no measurement error ase. We focus on three popular measures: degree, diffusion and eigenvector centralities. While they are most intuitive when A is binary, centrality measures should be understood as functions of general weighted (symmetric) adjacency matrices. Except where noted, our definitions are standard (see e.g. Jackson 2010; Bloch et al. 2021).

Definition 1 (Degree Centrality)**.** Degree centrality computed on the $n \times n$ adjacency matrix A is the $n \times 1$ vector:

$$C^{(1)} = A\iota_n \ .$$

Agent i's degree centrality is simply the sum of row i in A. If A is binary, degree centrality is the number of agents with whom i has a relationship.

Definition 2 (Diffusion Centrality)**.** For a given $\delta \in [0, 1]$ and $T \in \mathbf{N}$, diffusion centrality computed on the $n \times n$ adjacency matrix A is the $n \times 1$ vector:

$$C^{(T)} = \left(\sum_{t=1}^{T} \delta^t A^t \right) \iota_n \ .$$

Proposed by Banerjee et al. (2013), diffusion centrality captures the influence of agent i in terms of how many agents they can reach over T periods. Consider again the case of binary A. Then the $(i, j)^{\text{th}}$ of A^t is the number of walks from i to j that are of length t, which can be thought of as the influence of i on j in period t. Diffusion centrality for agent i is simply sum of their influence on all other agents in the network over time up to period T, with a decay of δ per period. Bramoullé and Genicot (2018) provides further discussion on the theoretical foundations of diffusion centrality. In practice, researchers often choose δ to be $1/\lambda_1(\hat{A})$, so that effectively $\delta \to 0$ as $n \to \infty$. An extension of our results to this case is in preparation.

Definition 3 (Eigenvector Centrality)**.** For a given $a_n > 0$, eigenvector centrality computed on the $n \times n$ adjacency matrix A is the $n \times 1$ vector:

$$C^{(\infty)} = a_n v_1(A) \ ,$$

where $v_1(A)$ is the eigenvector corresponding to the eigenvalue of A with the largest absolute value (leading eigenvalue).

Eigenvector centrality is based on the idea that an individual's influence is proportional to the influence of their friends. That is, for some $k > 0$, we seek the following property:

$$C_i = k \sum_{j \neq i} C_j \quad \text{for all } i \in [n] \ . \tag{1}$$

The eigenvectors of A solve the above equations, with k being the corresponding eigenvalue. By the Perron-Frobenius Theorem, the leading eigenvector is the unique eigenvector that be chosen so that every entry is non-negative, motivating its use as a centrality measure. The leading eigenvector of related matrices also emerges as the measure of influence in popular models of social learning (e.g. DeGroot 1974)

The leading eigenvector is well-defined only if the largest eigenvalue of A has multiplicity 1, that is, if $\lambda_1(A) \neq \lambda_2(A)$. To ensure that this occurs with high probability, we introduce the following assumption specific to eigenvector centrality:

Assumption E1. Suppose $\lambda_1(f) \neq \lambda_2(f)$.

Note also that eigenvectors are defined only up to scale: if C satisfies Equation 1, so will $a_n C$ for any $a_n \in \mathbf{R}$. Eigenvector centrality is commonly defined to have length 1 (e.g. Banerjee et al. 2013; Cruz et al. 2017), although researchers sometimes scale eigenvectors so that its standard deviation is 1 (Chandrasekhar 2016; Banerjee et al. 2019). Of the two papers that have considered the statistical properties of regression on eigenvector centrality, Cai et al. (2021) sets the length to $\sqrt{n}$, claiming it to be a normalization. Le and Li (2020) does not fix the length, though their goal is essentially to recover the projection $C^{(d)}\beta^{(d)}$ and not $\beta^{(d)}$ itself. We depart from the literature by leaving a_n as a free parameter. We will analyse the properties of regression on eigenvector centrality, making their dependence on a_n explicit. As we explain in Section 3.1, the choice of a_n is not innocuous and can have implications for estimation and inference.

This paper focuses on the above three centrality measures, which are intimately related (Bloch et al. 2021). When $T = 1$, $C^{(1)} \propto C^{(T)}$. Furthermore, as shown by Banerjee et al. (2019), if $\delta \geq 1/\lambda_1(A)$,

$$\lim_{T \to \infty} C^{(T)} \propto C^{(\infty)} \, .$$

We can thus understand the centrality measures as lying on a line, motivating our notational choice. Notably, we do not discuss betweenness and closeness centralities. These are path-based measures of centrality,which do not have clearly defined counterparts in the graphon. We conjecture that their analysis require a different statistical framework and defer it to future work.

Example 1 (continued)**.** In the context of risk sharing and social insurance, we can interpret

- $C_i^{(1)}$ as the probability-weighted number of friends who will lend to or borrow from i.
- $C_i^{(T)}$ as the probability-weighted number of friends who will lend to or borrow from i directly or through their friends. T is the maximum length of the borrowing chain.

For example if T is 2, i can borrow from friends of friends but not friends of friends of friends. δ is the increased difficulty of borrowing from a person that is one step further, e.g. of borrowing from friends of friends relative to borrowing from a friend directly.

- $C_i^{(\infty)}$ as requiring the borrowing ability of i to be proportional to the borrowing ability of their friends. Implicitly, this means agents can form borrowing chains that are arbitrarily long.

In the no measurement error case, the estimators of interest are:

Definition 4 (OLS Estimators without Measurement Error)**.** Suppose A is observed. For $d \in \{1, T, \infty\}$, define the OLS estimators for $\beta^{(d)}$ to be

$$\tilde{\beta}^{(d)} = \frac{Y' C^{(d)}}{(C^{(d)})' C^{(d)}} \ .$$

When networks are observed with errors, we assume that network centralities are estimated using $\hat{A}$ in place of A:

Definition 5 (Centralities with Measurement Error)**.** Suppose $\hat{A}$ is observed but not A. Define:

$$\begin{aligned}
\hat{C}^{(1)} &= \hat{A}\iota_n \ , \\
\hat{C}^{(T)} &= \left(\sum_{t=1}^{T} \delta^t \hat{A}^t \right) \iota_n \ , \\
\hat{C}^{(\infty)} &= a_n v_1(\hat{A}) \ .
\end{aligned}$$

The corresponding OLS estimators are thus defined:

Definition 6 (OLS Estimators with Measurement Error)**.** Suppose $\hat{A}$ is observed but not A. For $d \in \{1, T, \infty\}$, define the OLS estimators for $\beta^{(d)}$ to be

$$\hat{\beta}^{(d)} = \frac{Y' \hat{C}^{(d)}}{\left(\hat{C}^{(d)}\right)' \hat{C}^{(d)}} \ .$$

Next, define the regression residuals.

Definition 7 (Regression Residuals)**.** For $d \in \{1, T, \infty\}$, define:

$$\tilde{\varepsilon}_i^{(d)} := Y_i - \tilde{\beta}^{(d)} C_i^{(d)} \ , \tag{2}$$

$$\hat{\varepsilon}_i^{(d)} := Y_i - \hat{\beta}^{(d)} \hat{C}_i^{(d)} \ . \tag{3}$$

We conclude this section by stating assumptions on the moments of ε_i conditional on U_i:

Assumption 3. Suppose

(a) $E[\varepsilon_i|U_i] = 0$.

(b) $0 < \underline{\sigma}^2 \leq E[\varepsilon_i^2|U_i] \leq \bar{\sigma}^2 < \infty$.

(c) $E\left[|\varepsilon_i|^3 |U_i\right] \leq \bar{\kappa}_3$.

In the above assumption, (a) justifies linear regression. To see this, write:

$$E\left[\varepsilon_i \,|\, C_i^{(d)}\right] = E\left[E\left[\varepsilon_i \,|\, U, C_i^{(d)}\right] \,\big|\, C_i^{(d)}\right] = E\left[E\left[\varepsilon_i \,|\, U\right] \,\big|\, C_i^{(d)}\right] = E\left[E\left[\varepsilon_i \,|\, U_i\right] \,\big|\, C_i^{(d)}\right] = 0 \ ,$$

where the middle equality follows from the fact that $C_i^{(d)} = g(U_1, ..., U_n)$ for some function g. The subsequent equality follows because $\{(Y_i, U_i)\}_{i=1}^{\infty}$ – or equivalently $\{(\varepsilon_i, U_i)\}_{i=1}^{\infty}$ – are independently and identically distributed. Meanwhile, (b) and (c) control the amount of heterogeneity across different U_i's. (c) implies the upper bound in (b). We introduce $\bar{\sigma}^2$ for notational convenience.

2.2 Sparse Network Asymptotics

To better capture the behavior of estimators when agents in the networks have few relationships with one another, we study their properties under *sparse* network asymptotics. Following Bollobás et al. (2007) and Bickel and Chen (2009), we want to consider settings in which $p_n \to 0$ as $n \to \infty$. p_n is not an empirical quantity. It is a theoretical device to ensure that the sequence of models we consider remains sparse even as $n \to \infty$.

In many settings, a vector or matrix is said to be sparse if many of the entries are 0. In our setting, we say that A and $\hat{A}$ are sparse if their row sums – that is, the actual or observed degrees of agents respectively – are small. Because the entries of $\hat{A}$ are restricted to be binary, having low degrees is the same as having many entries which are 0. We do not restrict the entries of A, so that row sums could be small even if no entry takes value 0, as long as each non-zero entry is small. Sparsity of A should therefore be understood as referring to low intensities of relationship between agents, but which gives rise to observed networks, $\hat{A}$, that are sparse in the conventional sense.

To see how $p_n \to 0$ gives rise to sparsity, suppose for example that $p_n \to c > 0$. Then the network is dense and each agent has total relationships that are roughly of order n in expectation. That is,

$$E\left[C_i^{(1)}\right] \approx n,$$

corresponding to a situation in which each agent is linked to many others. In practice, however, researchers may face sparse networks, in which each agent has few or weak relationships. Choosing $p_n \to 0$ leads to networks that remain sparse as n increases. For example, if we set $p_n = k/n$ for some $k > 0$, then,

$$E\left[C_i^{(1)}\right] \approx 1 \; .$$

That is, each agent has a bounded number of relationships in expectation. A sequence of p_n that goes to 0 more quickly corresponds to data which is more sparse.

To understand the effect of sparsity on OLS estimation, we therefore study how the statistical properties of $\tilde{\beta}^{(d)}$ and $\hat{\beta}^{(d)}$ change as we vary the rate at which $p_n \to 0$. Our goal is to obtain theoretical results that better describe the properties of estimators under sparsity by explicitly modeling its effects. Such an approach has proven fruitful in statistics (e.g. in Bickel and Chen 2009; Bickel et al. 2011 and Wang 2020), but also in econometrics. Notably, Graham (2020b) lets a parameter analogous to p_n here go to 0 at rate $1/n$ to model the fact that individuals at an online market place purchase a bounded number of products in the limit, even though the selection tends to infinity.

A key motivation for our choice of framework is analytical tractability. Our definitions imply that conditional on U, $\hat{A}$ is a sparse inhomogeneous Erdos-Renyi graph, allowing us to borrow results from the random graph literature. However, modeling A as comprising many low intensity links is also reasonable from an economic perspective. In the seminal paper titled “The Strength of Weak Ties”, Granovetter (1973) argues that lower intensity links, which constitute most of any given person’s relationships, are the key drivers of many important social and economic outcomes. For example, in tracing the network of job referrals, the author finds that 83% of recent job changers in a Boston suburb found their new jobs through friends whom they saw fewer than twice a week, and who were only “marginally included in the current network of contacts”. The author further notes: “It is remarkable that people receive crucial information from individuals whose very existence they have forgotten.” A series of empirical work has found evidence in favor of the weak ties theory across diverse applications such as innovation (e.g. Reagans and Zuckerman 2001), economic development (e.g. Eagle et al. 2010) and job referrals (e.g. Rajkumar et al. 2022). This theory lends credence to our econometric model, in which an unobserved network of weak ties not only drives economic effects but also generates a sparse observed network.

As a theoretical device, p_n bears semblance to drifting alternatives in local power analysis (also known as Pitman drift; see Rothenberg 1984). Suppose we want to compare the power of tests for the hypothesis $H_0 : \beta = \beta_0$ against $H_1 : \beta = \beta_1$. Asymptotic analysis with a fixed β_1 is not useful since consistent tests have power that converges to 1 in probability under all alternatives, so that we cannot meaningfully differentiate between these tests. One interpretation of such a failure is that the asymptotic model fails to capture reality: in the limit, $|\beta_1 - \beta_0|$ is large relative to the sampling noise which is of order $1/\sqrt{n}$. In practice, sampling noise can be large relative to the parameter of interest. Local power analysis employs the alternative hypothesis $\beta_1 = \beta_0 + k/\sqrt{n}$. As such, $|\beta_1 - \beta_0| = |k/\sqrt{n}|$ goes to 0 at the same rate as sampling noise. Intuitively, as the sample size gets larger, the testing problem also becomes harder. The upshot is that the testing problem is non-trivial even in the limit, better modeling the finite sample problem.

A similar approach is taken in the weak instruments literature, which is concerned with the instrumental variable regressions in which the relevance condition is barely satisfied. To understand the resulting statistical pathologies, Staiger and Stock (1997) propose to model the strength of the instrument as decaying to 0 at rate $1/\sqrt{n}$, so that strength of the signal in the first stage estimation is on par with sampling uncertainty. This approach has since led to long and productive lines of inquiry (see Andrews et al. 2019 and references therein).

Our drifting parameter p_n serves a similar purpose: by letting $p_n \to 0$, we better capture the statistical properties of estimators when networks are sparse. While we do not focus on any reference level of sparsity, comparing across levels of sparsity will prove instructive.

3 Theoretical Results

In this section, we present our theoretical results about the property of OLS estimators under varying amounts of sparsity. In Section 3.1, we characterize the level of sparsity at which consistency of $\tilde{\beta}^{(d)}$ and $\hat{\beta}^{(d)}$ fails. The upshot is that measurement error renders OLS estimators less robust to sparsity. In particular, eigenvector centrality is less robust to sparsity than degree under measurement error. Section 3.2 presents distributional theory for $\tilde{\beta}^{(d)}$ and $\hat{\beta}^{(d)}$ under regimes of sparsity for under which they are consistent. This leads to tools for bias correction and inference with sparse and noisily measured networks.

3.1 Consistency

This section presents the rates on p_n at which $\tilde{\beta}^{(d)}$ and $\hat{\beta}^{(d)}$ are consistent. We also discuss the role of a_n in ensuring the consistency of $\tilde{\beta}^{(\infty)}$ and $\hat{\beta}^{(\infty)}$.

We first consider the case when the true network A, is observed:

Theorem 1 (Consistency without Measurement Error)**.** Suppose Assumptions 1, 2 and 3 hold. Then,

(a) For $d \in \{1, T\}$, $\tilde{\beta}^{(d)} \xrightarrow{p} \beta^{(d)}$ if and only if $p_n \succ n^{-\frac{3}{2}}$.

(b) Suppose Assumption E1 also holds. Then, $\tilde{\beta}^{(\infty)} \xrightarrow{p} \beta^{(\infty)}$ if and only if $a_n \to \infty$.

As such, we have consistency of OLS for degree and diffusion centralities provided that the network is not *too* sparse. Under extreme sparsity, variation in $C_i^{(d)}$ becomes much smaller than variation in ε_i and it is not possible to learn about $\beta^{(d)}$. In the case of eigenvector centrality, consistency requires conditions on the normalization factor a_n but not on p_n. This is because a_n directly controls the variance of $C^{(\infty)}$, so that it is able to undo the effect of sparsity in the absence of measurement error.

Our result is similar in spirit to Conley and Taber (2011), which studies the properties of difference-in-difference (DiD) estimators when there are few treated unit. In an asymptotic framework that takes the number of treated units to be fixed, the DiD estimator is similarly inconsistent in the limit. More generally, consistency of OLS with i.i.d. data requires $\sqrt{n}\sigma_X \to \infty$, where σ_X is the variance of the regressor. Theorem 1 instantiates this condition for centrality regressions under sparsity.

Interestingly, the choice of a_n matters even when the network is dense. To see why, suppose $f = p_n \cdot 1$ so that $A = p_n \iota_n \iota_n'$. Then $C^{(\infty)}(A) = a_n \iota_n / \sqrt{n}$. Note that it is independent of p_n. We can then write:

$$\tilde{\beta}^{(\infty)} = \frac{\sqrt{n}}{a_n} \cdot \frac{Y' \iota_n}{\iota_n' \iota_n} = \beta^{(\infty)} + \frac{1}{a_n \sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \ .$$

Under our assumptions, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \xrightarrow{d} N(0, \sigma^2)$. Therefore, $a_n \to \infty$ is necessary for the consistency of $\tilde{\beta}^{(\infty)}$.

The above example, together with Theorems 2 and 5 in the next section, makes clear that a_n has important implications for the statistical properties of $\tilde{\beta}^{(\infty)}$ and $\hat{\beta}^{(\infty)}$. We can understand this phenomenon by analogy to OLS with i.i.d. observations, in which we are able to consistently estimate β but not $\sqrt{n}\beta$.

To our knowledge, we are the first to emphasize the importance in choosing a_n appropriately. Cai et al. (2021), which studies eigenvector regressions under a different model for measurement error, sets a_n to $\sqrt{n}$ claiming it to be a normalization. In the formulation of Le and Li (2020), a_n appears only implicitly and they do not prove consistency of $\hat{\beta}^{(\infty)}$. Instead, they show that $\|\hat{\beta}^{(\infty)}\hat{C}^{(\infty)} - \beta^{(\infty)}C(\infty)\|_\infty \xrightarrow{p} 0$. We remark that changing a_n amounts to changing the definition of $\tilde{\beta}^{(\infty)}$. The parameter of interest ultimately depends on the researcher. From the perspective of consistency, however, models with $a_n \to \infty$ are strictly preferable to those with $a_n \preccurlyeq 1$. And as we will see in Theorem 5, particular choices of a_n may be useful for inference.

We next consider the case with measurement error:

Theorem 2 (Consistency with Measurement Error)**.** Suppose Assumptions 1, 2 and 3 hold. Then,

(a) For $d \in \{1, T\}$, $\hat{\beta}^{(d)} \xrightarrow{p} \beta^{(d)}$ if and only $p_n \succ n^{-1}$.

(b) Suppose also that Assumption E1 holds. Then, $\hat{\beta}^{(\infty)} \xrightarrow{p} \beta^{(\infty)}$ if $a_n \to \infty$ and

$$p_n \succ n^{-1}\sqrt{\frac{\log n}{\log\log n}}\,. \tag{4}$$

Suppose p_n satisfies

$$n^{-1}\left(\log\log n\right)^4 \prec p_n \prec n^{-1}\sqrt{\frac{\log n}{\log\log n}}\,. \tag{5}$$

Then $\hat{\beta}^{(\infty)}$ is inconsistent for $\beta^{(\infty)}$.

Theorem 2 gives the rates at which OLS regression on each centrality is consistent under measurement error. We summarize the rates from Theorems 2, together with that from 1, in Figure 1.

Under measurement error, $\hat{\beta}^{(\infty)}$ is consistent under less sparsity than $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$, even when we set $a_n \to \infty$. In other words, $\hat{\beta}^{(\infty)}$ is less robust to sparsity than $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$. This occurs because eigenvectors are poorly behaved under sparsity. The rate in Equation (4) arises because this is the point at which the leading eigenvector of $\hat{A}$ ceases to be informative about the leading eigenvector of A. To be specific, the rate in Equation (4) is the threshold for eigenvector delocalization. Above this threshold, eigenvectors of the noise ($\hat{A} - A$ in our case) are delocalized, meaning that each component is small and roughly of the same order. This ensures that the eigenvectors of $\hat{A}$ is close to that of A. Below this threshold, eigenvector localization occurs (Alt et al. 2021a). That is, the mass of the leading eigenvector concentrates on the agent who happens to have the largest realized degree, which is purely a result of chance. In turn, the leading eigenvector is also noise, rendering $\hat{\beta}^{(\infty)}$ inconsistent. On the other hand, consistency of $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$ only requires the mean of $\hat{A}$ to concentrate to that of A, which occurs as long as $p_n \succ n^{-1}$.

An important implication of our result is that centrality measures may have differing predictive value for outcomes in sparse regimes ,not only because they differ in economic significance, but also because they differ in statistical properties. In particular, suppose diffusion centrality leads to estimates which are significantly different from 0 at some level α, while eigenvector does not. If the underlying networks are sparse, it would be erroneous to conclude that diffusion centrality is structurally meaningful while eigenvector is not, since sparsity might be driving the observed results.

Finally, let us compare the rates in Theorem 2 with those in Theorem 1. As Figure 1 shows, measurement error renders OLS less robust to sparsity. While $\tilde{\beta}^{(1)}$ and $\tilde{\beta}^{(T)}$ are consistent as long as $p_n \succ n^{-3/2}$, $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$ now require that $p_n \succ n^{-1}$. Whereas $\tilde{\beta}^{(\infty)}$ did not require any conditions on p_n for consistency, $\hat{\beta}^{(\infty)}$ does. Moreover, this requirement is more stringent than that on $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$. OLS on eigenvector centrality is therefore more sensitive to measurement error than on degree or diffusion.

Remark 2. To improve the robustness of eigenvector centrality to sparsity, we can consider regularizing $\hat{A}$. Appendix C considers such an approach and finds that consistency with regularized eigenvector obtains when $p_n \succ n^{-1}$.

Remark 3. Theorem 2 does not determine the behavior of $\hat{\beta}^{(\infty)}$ when $p_n \prec n^{-1} (\log \log n)^4$. Up to this threshold, we know by Alt et al. (2021b) that OLS is inconsistent only because the eigenvectors of $\hat{A} - A$ are localized. To our knowledge, recent developments in random matrix theory do not provide any description of eigenvectors below threshold. As such, it is not clear what type of pathologies arises below $p_n \prec n^{-1} (\log \log n)^4$ and how that might affect the behavior of $\hat{\beta}^{(\infty)}$. Description of eigenvalues is more complete: below this point, we know that $\lambda_1(\hat{A})/\lambda_1(A) \to \infty$ (see Alt et al. 2021a; Benaych-Georges et al. 2019; Benaych-Georges et al. 2020). Since the estimated eigenvalues are noise, we conjecture that the estimated eigenvectors are as well. If so, we would not expect $\hat{\beta}^{(\infty)}$ to be consistent.

3.2 Distributional Theory

In this section, we study the asymptotic distributions of $\tilde{\beta}^{(d)}$ and $\hat{\beta}^{(d)}$ under sparsity and measurement error. We focus on regimes of p_n under which each estimator is consistent and find that measurement error still leads to asymptotic bias. Specifically,

$$\hat{\beta}^{(d)} \xrightarrow{p} \beta^{(d)} \quad \text{but} \quad E\left[\lim_{n\to\infty} np_n \left(\hat{\beta}^{(d)} - \beta^{(d)}\right)\right] =: B^{(d)} \neq 0 \; .$$

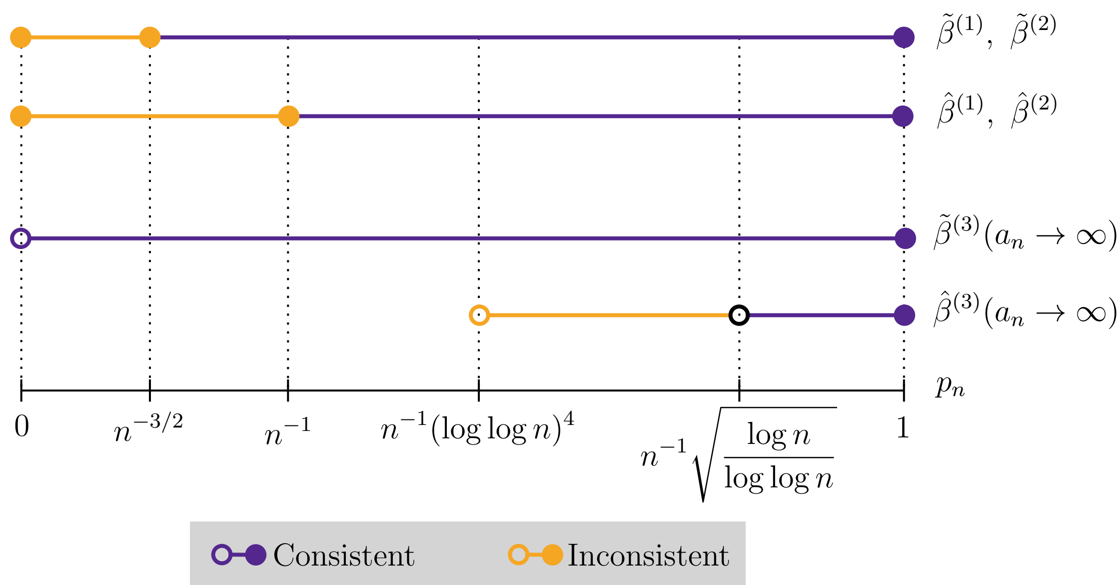

Figure 1: Ranges of consistency for each estimator. When the network is observed with error, regression on eigenvector centrality is less robust to sparsity than on degree or diffusion. When the network is known, much more sparsity can be accommodated.

Furthermore, the bias may be of larger order than the variance of $\hat{\beta}^{(d)}$ in the sense that

$$\frac{\sqrt{\mathrm{Var}(\hat{\beta}^{(d)})}}{B^{(d)}/np_n} \to 0 \ .$$

In this case, it would not be possible to obtain a non-degenerate limiting distribution without bias correction. To see this, write:

$$\frac{\hat{\beta}^{(d)} - \beta^{(d)}}{\sqrt{v_n}} = \underbrace{\frac{\hat{\beta}^{(d)} - \beta^{(d)} - B^{(d)}/np_n}{\sqrt{v_n}}}_{=:\Gamma_1} + \underbrace{\frac{B^{(d)}/np_n}{\sqrt{v_n}}}_{=:\Gamma_2} \ .$$

Suppose we chose $v_n = \mathrm{Var}(\hat{\beta}^{(d)})$. Then $\Gamma_1 \xrightarrow{d} D$, where D is some non-degenerate distribution. However, Γ_2 diverges to $+\infty$ or $-\infty$ depending on the sign of B^d. On the other hand, suppose we chose v_n so that Γ_2 is bounded. Then $\Gamma_1 \xrightarrow{d} 0$ since $\mathrm{Var}(\hat{\beta}^{(d)})/v_n \to 0$. That is, its limit is degenerate. Bias correction is thus necessary for inference.

We propose bias-correction and corresponding inference methods based on $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$. Bias correction is ineffective for $\hat{\beta}^{(\infty)}$. Instead, we propose a data-dependent choice of a_n that leads to convenient properties. Appendix C provides an alternative strategy for regularizing eigenvector centrality.

3.2.1 Centralities without Measurement Error ($\tilde{\beta}^{(1)}, \tilde{\beta}^{(T)}, \tilde{\beta}^{(\infty)}$)

Our first result states that heteroskedasticity-consistent (hc) or robust t-statistics yield vaild inference in the absence of measurement error.

Theorem 3. Suppose Assumptions 1 and 3 hold.

(a) Suppose $p_n \succ n^{-3/2}$. Then, for $d \in \{1, T\}$,

$$\tilde{S}^{(d)} = \frac{\tilde{\beta}^{(d)} - \beta^{(d)}}{\sqrt{\tilde{V}^{(d)}}} \xrightarrow{p} \mathrm{N}(0,1) \ .$$

where $\tilde{V}^{(d)} = \left(\sum_{i=1}^n \left(C_i^{(d)}\right)^2\right)^{-2} \sum_{i=1}^n \left(C_i^{(d)}\right)^2 \left(\tilde{\varepsilon}_i^{(d)}\right)^2$.

(b) Suppose $a_n \to \infty$. Then,

$$\tilde{S}_0^{(\infty)} = \frac{\tilde{\beta}^{(\infty)} - \beta^{(\infty)}}{\sqrt{\tilde{V}^{(\infty)}}} \xrightarrow{p} \mathrm{N}(0,1) \ .$$

where $\tilde{V}^{(\infty)} = \left(\sum_{i=1}^n \left(C_i^{(\infty)}\right)^2\right)^{-2} \sum_{i=1}^n \left(C_i^{(\infty)}\right)^2 \left(\tilde{\varepsilon}_i^{(\infty)}\right)^2$.

In the above, $\tilde{\varepsilon}_i$ is as defined in Equation (2).

Our formulation of the t-statistic highlights that inference on $\beta^{(1)}$ and $\beta^{(T)}$ does not require the sparsity parameter p_n to be specified. This is important since p_n is in general not identified (Bickel et al. 2011) and follows from the convenient fact that the t-statistic is self-normalizing. Intuitively, the sparsity terms in the numerator and the denominator are of the same order, so that they "cancel out". Hansen and Lee (2019) makes a similar observation in the context of cluster-dependent data: although the means of such data converge at a rate that changes based on the dependence structure within each cluster, this rate does not need to be known for estimation and inference, due to the aforementioned self-normalizing property.

We note that $\tilde{V}^{(1)} = O_p\left(n^{-3}p_n^{-2}\right)$, $\tilde{V}^{(T)} = O_p\left(n^{-2T-1}p_n^{-2T}\right)$. These are the rates of convergence for $\tilde{\beta}^{(1)}$ and $\tilde{\beta}^{(T)}$ respectively. In the absence of sparsity (i.e. if $p_n = 1$), the rate of convergence is faster than $n^{-1/2}$. This is because having a network amounts to n^2 observations. Asymptotically, the regressor $C_i^{(d)}$ has much more variation than the regression error ε_i, leading to the higher rate of convergence. Finally, we note that $\tilde{V}^{(\infty)} = O_p(a_n^{-2})$.

In the presence of measurement error, however, the above result does not obtain. The next two subsections presents distributional theory for $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$, and $\hat{\beta}^{(\infty)}$.

3.2.2 Degree and Diffusion Centrality under Measurement Error ($\hat{\beta}^{(1)}$, $\hat{\beta}^{(T)}$)

For $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$, measurement error leads to bias and also slows down the rate of convergence. This is the content of the following theorem:

Theorem 4 (Inference – Degree and Diffusion)**.** Suppose Assumptions 1, 2 and 3 hold and that $p_n \succ n^{-1}$.

(a) Suppose $\beta^{(1)} \neq 0$. Then,

$$\hat{S}^{(1)} := \frac{\hat{\beta}^{(1)} - \beta^{(1)}\left(1 - \hat{B}^{(1)}\right)}{\beta^{(1)}\sqrt{\hat{V}^{(1)}}} \xrightarrow{d} \mathrm{N}(0,1) \ ,$$

where

$$\hat{V}^{(1)} = \frac{1}{2}\left(\sum_{i=1}^{n}\left(\hat{C}_i^{(1)}\right)^2\right)^{-2} \sum_{j\neq i} \hat{A}_{ij}\left(\hat{C}_i^{(1)} + \hat{C}_j^{(1)}\right)^2 \quad , \quad \hat{B}^{(1)} = \left(\sum_{i=1}^{n}\left(\hat{C}_i^{(1)}\right)^2\right)^{-1} \iota_n \hat{A} \iota_n \ .$$

(b) Suppose $\beta^{(T)} \neq 0$. Then,

$$\hat{S}^{(T)} = \frac{\hat{\beta}^{(T)} - \beta^{(T)}\left(1 - \hat{B}^{(T)}\right)}{\beta^{(T)}\sqrt{\hat{V}^{(T)}}} \xrightarrow{d} \mathrm{N}(0,1) \ ,$$

where

$$\hat{V}^{(T)} = \frac{1}{2}\left(\sum_{i=1}^{n}\left(\hat{C}_i^{(T)}\right)^2\right)^{-2} \cdot \delta^{2T} \cdot \iota_n' \left[\hat{A} \circ \left(\sum_{t=1}^{2T}\left(\hat{A}^{2T-t}\iota_n\right)\left(\iota_n'\hat{A}^{t-1}\right)\right)^{\circ 2}\right] \iota_n$$

$$\hat{B}^{(T)} = \left(\sum_{i=1}^{n}\left(\hat{C}_i^{(T)}\right)^2\right)^{-1} \sum_{t=1}^{2T-1} b_T(t,\delta) \cdot \iota_n \hat{A}^t \iota_n \ .$$

Here, $\circ$ denotes the entrywise product. The formula for $b_T(t,\delta)$, up to $T = 10$, can be found in Appendix B.

(c) Suppose for $d \in \{1, T\}$ that $\beta^{(d)} = 0$. Then,

$$\hat{S}_0^{(d)} := \frac{\hat{\beta}^{(d)}}{\sqrt{\hat{V}_0^{(d)}}} \xrightarrow{d} \mathrm{N}(0,1) \ ,$$

where

$$\hat{V}_0^{(d)} = \left(\sum_{i=1}^{n} \left(\hat{C}_i^{(T)} \right)^2 \right)^{-2} \sum_{i=1}^{n} \left(\hat{C}_i^{(d)} \right)^2 \left(\hat{\varepsilon}_i^{(d)} \right)^2 .$$

Here, $\hat{\varepsilon}_i^{(d)}$ is as defined in Equation (3)

Our results are stated in terms of $\hat{B}^{(d)}$ and $\hat{V}^{(d)}$ – estimators for bias and variance – though they should be understood as statements about the true bias and variance of the estimators in combination with statements about estimation feasibility. Note also that results for $\hat{\beta}^{(T)}$ specializes to that for $\hat{\beta}^{(1)}$ when setting $T = \delta = 1$.

When $\beta^{(d)} = 0$ (case (c)), our result asserts that the hc/robust variance estimator is consistent for the variance of $\hat{\beta}^{(d)}$. However, that is no longer the case when then $\beta^{(d)} \neq 0$ (cases (a) and (b)). Here, we find that $\hat{\beta}^{(d)}$ become biased. That is, $\hat{\beta}^{(d)}$ is not centered at $\beta^{(d)}$. The bias of $\hat{\beta}^{(1)}$ comprises only one term. However, bias for $\hat{\beta}^{(T)}$ comprises an exponentially growing number of terms. In addition to eigenvector localization, this provides another intuitive explanation for the poor properties of the eigenvector centrality, since as Banerjee et al. (2019) proves, can be considered the limit of diffusion centrality as $T \to \infty$. Comparing cases (a) and (b) with (c) also shows that the asymptotic distributions for $\hat{\beta}^{(d)}$ are discontinuous in $\beta^{(d)}$ at 0.

Additionally, we see that the asymptotic variance of $\hat{\beta}^{(d)}$ differs from that which is estimated by hc/robust standard error. Note that the difference in asymptotic variance is not the result of bias estimation. In particular, replacing $\hat{B}^{(d)}$ with its limit in probability will not change the asymptotic variance of $\hat{S}^{(d)}$. This stands in contrast to settings such as Regression Discontinuity Design, in which estimation of the asymptotic bias leads to larger asymptotic variance in the relevant test statistic (Calonico et al. 2014). In order to not alter the asymptotic variance, bias must be estimated at a sufficiently fast rate. This is not trivial for $\hat{\beta}^{(T)}$. Bias of the $\hat{\beta}^{(T)}$ comprises terms of the form $\iota_n A^t \iota_n$. However, the naive plug-in estimator $\iota_n \hat{A}^t \iota_n$ does not converge sufficiently fast for $t \geq 2$. Using the fact that $\iota_n \hat{A} \iota_n$ is a good estimator of $\iota_n A \iota_n$, we recursively construct good estimators for $\iota_n \hat{A}^t \iota_n$ when $t \geq 2$, which can then be used to construct $\hat{B}^{(T)}$. The resulting estimator does not have a closed form expression. We provide explicit formula for $T \leq 10$ in Tables 9 and 10.

Hypothesis Testing

Our theory suggests the following test for $d \in \{1, T\}$. Consider testing the hypothesis $H_0 : \beta^{(d)} = \beta_0$ against $H_1 : \beta^{(d)} \neq \beta_0$ by

$$\phi^{(d)} = \begin{cases} \mathbf{1}\left\{\left|\frac{\hat{\beta}^{(d)}}{\sqrt{\hat{V}_0^{(d)}}}\right| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right\} & \text{if } \beta_0 = 0 \\ \mathbf{1}\left\{\left|\frac{\hat{\beta}^{(d)} - \beta_0\left(1 - \hat{B}^{(d)}\right)}{\beta_0\sqrt{\hat{V}^{(d)}}}\right| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right\} & \text{otherwise.} \end{cases}$$

where Φ is the CDF of the standard normal distribution. One-sided tests can be constructed by modifying the rejection rule in the usual way. It is immediate that:

Corollary 1 (Inference for $\beta^{(1)}$ and $\beta^{(T)}$)**.** If $\beta^{(d)} = \beta_0$, $E\left[\phi^{(d)}\right] \to \alpha$.

When $\beta^{(d)} \neq 0$, $\hat{\beta}^{(d)}$ needs to be centered by subtracting $\beta^{(d)}\left(1 - \hat{B}^{(d)}\right)$ instead of $\beta^{(d)}$. We will refer to this form of centering as *bias correction* for $\hat{\beta}^{(d)}$. As we explained at the start of Section 3, bias correction is necessary for $\hat{\beta}^{(d)}$ to attain a non-degenerate limiting distribution when asymptotic bias is of larger order than variance. Indeed, $\hat{B}^{(d)}/\sqrt{\hat{V}^{(d)}} = O_p\left(p_n^{-1/2}\right)$. As such, if $p_n \prec 1$,division by $\sqrt{\hat{V}^{(d)}}$ blows up $\hat{B}^{(d)}$.

In the bias for $\hat{\beta}^{(T)}$, terms with larger t's dominate those with smaller t's. When p_n is dense enough, terms with small t's may actually much smaller than $\sqrt{\hat{V}^{(T)}}$ so that they can be ignored. With only the stipulation that $p_n \succ n^{-1}$ however, a non-degenerate asymptotic distribution can only be achieved when all terms are included.

Confidence Intervals

Because $\hat{V}_0^{(d)}$ estimates variance only when $\beta^{(d)} = 0$, the usual confidence intervals based on $\hat{V}_0^{(d)}$ need not attain nominal coverage. This failure occurs for two countervailing reasons. Firstly, the quantity $\hat{V}_0$ is meant to estimate,

$$\operatorname{Var}\left(\sum_{i=1}^{n} C_i^{(d)} \varepsilon_i\right) =: V_0^{(d)} \; .$$

However, $V_0^{(d)}$ under-estimates variance of $\hat{\beta}^{(d)}$ when $\beta^{(d)} \neq 0$. That is,

$$\operatorname{Var}\left(\frac{\hat{\beta}^{(d)} - E[\beta^{(d)}]}{\sqrt{V_0}}\right) \to \infty \; .$$

On the other hand, the bias in $\hat{\beta}^{(d)}$ means that

$$\hat{V}_0^{(d)} \approx V_0 + \beta^{(d)} \hat{B}^{(d)} \sum_{i=1}^{n} \left(C_i^{(d)} \right)^2 \ .$$

The second term in the above equation can be large, such that $\hat{V}_0^{(d)}$ may exceed $\hat{V}$. This turns out to be the case in our application in Section 5.

To obtain confidence intervals for $\beta^{(d)}$ consider the following:

Definition 8. For $d \in \{1, T\}$ and a given α, let

$$\mathcal{C}_0^{(d)} := \left[\hat{\beta}^{(d)} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{V}_0^{(d)}} \quad , \quad \hat{\beta}^{(d)} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{V}_0^{(d)}} \ \right] \ ,$$

and

$$\mathcal{C}^{(d)} := \left[\frac{\hat{\beta}^{(d)}}{1 - \hat{B}^{(d)} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{V}^{(d)}}} \quad , \quad \frac{\hat{\beta}^{(d)}}{1 - \hat{B}^{(d)} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{V}^{(d)}}} \right] \ .$$

Finally, let $\mathcal{C}_*^{(d)} = \mathcal{C}_0^{(d)} \cup \mathcal{C}^{(d)}$.

The following is immediate:

Corollary 2 (Confidence Interval)**.** $P\left(\beta^{(d)} \in \mathcal{C}_*^{(d)}\right) \to 1 - \alpha$.

We can obtain one-sided confidence intervals by modifying the bounds as usual. More generally, as long as $\mathcal{C}^{(d)}$ is a $1 - \alpha$ confidence interval for $\beta^{(d)} \neq 0$ and $\mathcal{C}_0^{(d)}$ is a $1 - \alpha$ confidence interval when $\beta^{(d)} = 0$, their unions will produce a $1 - \alpha$ confidence interval for $\beta^{(d)}$ unconditionally. In particular, it is always valid to set $\mathcal{C}_0^{(d)} = \{0\}$. This can be useful when it is not important to exclude 0 from the confidence interval. For example, suppose we want one-sided confidence intervals that upper bounds $\beta^{(d)}$. We can then consider using

$$\mathcal{C}_0^{(d)} = \{0\} \quad \text{and} \quad \mathcal{C}^{(d)} = \left(-\infty \quad , \quad \frac{\hat{\beta}^{(d)}}{1 - \hat{B}^{(d)} - \Phi^{-1}(1 - \alpha) \sqrt{\hat{V}^{(d)}}} \right] \ .$$

If $\hat{\beta}^{(d)} > 0$, $\mathcal{C}_*^{(d)} = \mathcal{C}^{(d)}$. For the reasons discussed above, we can also have

$$\mathcal{C}_*^{(d)} \subsetneq \left(-\infty \quad , \quad \hat{\beta} + \Phi^{-1}(1 - \alpha) \sqrt{\hat{V}_0^{(d)}} \ \right] \ .$$

As before, such a situation arises in our application (Section 5).

Bias Correction

Since the bias of the OLS estimators $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$ can be estimated, it is reasonable to consider the following *bias-corrected* estimators:

Definition 9 (Bias-Corrected Estimators)**.** For $d \in \{1, T\}$, define

$$\check{\beta}^{(d)} = \frac{\hat{\beta}^{(d)}}{1 - \hat{B}^{(d)}} \; .$$

Bias-corrected estimators have faster rates of convergence:

Corollary 3. Suppose $p_n \succ n^{-1}$. For $d \in \{1, T\}$, $\check{\beta}^{(d)} - \beta^{(d)} = O_p\left(n^{-2} p_n^{-3/2}\right)$.

For reference, $\hat{\beta}^{(d)} - \beta^{(d)} = O_p\left(n^{-1} p_n^{-1}\right)$.

3.2.3 Eigenvector Centrality under Measurement Error ($\hat{\beta}^{(\infty)}$)

We next consider inference on $\hat{\beta}^{(\infty)}$. Eigenvector centrality can be badly biased under sparsity, which makes inference challenging. However, strategic choices of a_n can overcome many of these issues. We first introduce the following simplifying assumption:

Assumption E2 (Finite Rank)**.** Suppose f has rank $R < \infty$:

$$f(u, v) = \sum_{r=1}^{R} \tilde{\lambda}_r \phi_r(u) \phi_r(v) \quad , \tag{6}$$

where $\|\phi_r\| = 1$ for all $r \in [R]$ and if $r \neq s$,

$$\int_{[0,1]} \phi_r(u) \phi_s(u) du = 0 \; .$$

Furthermore, suppose that

$$\Delta_{\min} = \min_{1 \geq r \geq R-1} \left| \tilde{\lambda}_r - \tilde{\lambda}_{r+1} \right| > 0 \; .$$

In Equation (6), we express f in terms of its eigenfunctions $\{\phi_r\}_{r=1}^{R}$. Assumption E2 implies that the true network has low-dimensional structure and is satisfied by many popular network models, such as the stochastic block model (Holland et al. 1983, also see Example 4 below) and random dot product graphs (Young and Scheinerman 2007). This assumption is also commonly found in the networks literature (e.g. Levin and Levina 2019; Li et al. 2020), and the matrix completion literature more generally (e.g. Candès and Tao 2010; Negahban and Wainwright 2012; Chatterjee 2015; Athey et al. 2021). Importantly, existing papers on inference with eigenvectors (Le and Li 2020; Cai et al. 2021) also make this assumption. Note also that Assumption E2 implies Assumption E1.

Example 4 (Stochastic Block Model)**.** The Stochastic Block Model (SBM) is one of the earliest statistical models of networks. It assumes that individuals fall into groups $g \in \{1, ..., B\}$ and that the true network depends only on group membership. For example, suppose that a classroom has two groups: jocks, nerds. The SBM posits that the strength of the tie between any jock and any nerd are the same. Analogously for that between any two jocks or any two nerds, though all three ties can be of different intensity. Suppose the proportion of each group is π_g and that the link probability is $p_{g,g'} = p_{g,g'}$. Then the graphon is a step-function on $[0,1]^2$ with B^2-steps and rank B. It is visualized in Figure 2.

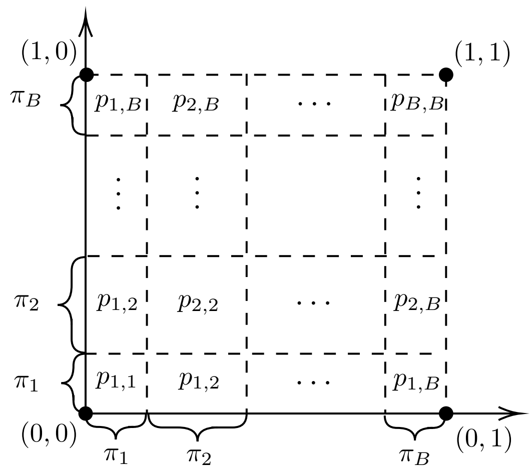

Figure 2: The graphon f of a stochastic block model with B blocks. f is a step-function with B^2 steps and is of rank B.

With the low-rank asumption, we can consider the asymptotic distribution of $\hat{\beta}^{(\infty)}$ in a few cases.

Theorem 5 (Inference – Eigenvector)**.** Suppose Assumptions 1, 2, 3 and E2 hold.

(a) Suppose either:

  (i) $\beta^{(\infty)} = 0$, or,

  (ii) $p_n \succ n^{-1} \log n$ and $a_n \prec (np_n)^{3/2}$, or,

  (iii) For some $\eta > 0$,

$$p_n \succ n^{-1} \left(\frac{\log n}{\log \log n} \right)^{\frac{1}{2}+\eta} . \tag{7}$$

  and $a_n \prec np_n$.

Then,

$$\hat{S}_0^{(\infty)} := \frac{\hat{\beta}^{(\infty)} - \beta^{(\infty)}\left(1 - \hat{B}^{(\infty)}\right)}{\sqrt{\hat{V}_0^{(\infty)}}} \xrightarrow{d} \mathrm{N}\left(0,1\right) \ . \tag{8}$$

where

$$\hat{V}_0^{(\infty)} = \left(\sum_{i=1}^{n}\left(\hat{C}_i^{(T)}\right)^2\right)^{-2}\sum_{i=1}^{n}\left(\hat{C}_i^{(\infty)}\right)^2\left(\hat{\varepsilon}_i^{(\infty)}\right)^2 \quad , \quad \hat{B}^{(\infty)} = \left(\lambda_1(\hat{A})\right)^{-1}$$

In the above, $\hat{\varepsilon}_i^{(\infty)}$ is as defined in Equation (3).

(b) Suppose $p_n \succ \frac{1}{\sqrt{n}}$, $a_n \succ n\sqrt{p_n}$ and $\beta^{(\infty)} \neq 0$. Then,

$$\hat{S}^{(\infty)} := \frac{\hat{\beta}^{(\infty)} - \beta^{(\infty)}\left(1 - \hat{B}^{(\infty)}\right)}{\sqrt{\hat{V}^{(\infty)}}} \xrightarrow{d} \mathrm{N}(0,1)$$

where

$$\hat{V}^{(\infty)} = 2\left(\lambda_1(\hat{A})\sum_{i=1}^{n}\left(\hat{C}_i^{(\infty)}\right)^2\right)^{-2}\sum_{j\neq i}\hat{A}_{ij}\left(\left(\hat{C}_i^{(\infty)}\right)^2 + \left(\hat{C}_j^{(\infty)}\right)^2\right)$$

Note that the statistics above do not require R or p_n to be specified. This is useful since estimating R may be challenging in addition to p_n being unidentified (Bickel et al. 2011).

Our result describes the asymptotic distribution $\hat{\beta}^{(\infty)}$, which depends on $\beta^{(\infty)}$ and a_n. Case (a) gives conditions under which inference with hc/robust t-statistic is appropriate. As with $\beta^{(1)}$ and $\beta^{(\infty)}$, the usual test works if $\beta^{(\infty)} = 0$. However, it also works if $\beta^{(\infty)} \neq 0$ provided that a_n is small. On the other hand, if a_n is large, case (b) suggests that we get behavior that is more in line with that of $\beta^{(1)}$ and $\beta^{(\infty)}$ when target parameters are non-zero. However, to obtain the result in case (b), we require very strong conditions on p_n due to greater difficulty in controlling the behavior of estimated eigenvector, as the discussion following Theorem 4 explains.

When $\beta^{(\infty)} \neq 0$, the differences in case (a) and (b) arise because a_n controls the relative sizes of network measurement error and regression error. The latter dominates if a_n is sufficiently small and has the advantage of being easy to characterize. Hence, in the absence of compelling reasons for choosing a_n to be other values, researchers can consider choosing a_n for statistical convenience. In particular, if a_n is chosen so that case (a) obtains, then usual hc/robust variance estimator based t-statistic and confidence interval have the expected properties. We propose such an a_n below. However, we stress that a smaller a_n also implies a lower rate of convergence. In effect, we are changing the model from one with faster but unknown rate of convergence, to one with a rate that is slower but estimable.

Finally, our result here suggests the use of the bias-corrected estimator, as with degree and diffusion centrality:

$$\breve{\beta}^{(\infty)} = \frac{\hat{\beta}^{(\infty)}}{1 - \hat{B}^{(\infty)}} .$$

Choice of a_n

The following data-dependent choice is convenient:

Corollary 4. Suppose $a_n = \sqrt{\lambda_1(\hat{A})}$ is estimated. If p_n satisfies Equation (7),

$$\frac{\hat{\beta}^{(\infty)} - \beta^{(\infty)}}{\sqrt{\hat{V}^{(\infty)}}} \xrightarrow{d} \mathrm{N}(0,1) .$$

Since $\lambda_1(\hat{A}) \approx \tilde{\lambda}_1 n p_n$, the above choice of a_n satisfies the conditions in case (a)(iii) of Theorem 5, accommodating close to the maximum possible amount of sparsity for $\hat{\beta}^{(\infty)}$. It also obviates the need for bias correction since the bias is always of lower order than the variance at this rate. Furthermore, such an a_n has intuitive appeal since it implies that

$$\hat{C}^{(\infty)} \left(\hat{C}^{(\infty)}\right)' = \lambda_1(\hat{A}) v_1(\hat{A}) \left(v_1(\hat{A})\right)' = \underset{\mathrm{rank}(M)=1}{\arg\min} \left\| M - \hat{A} \right\|_2 .$$

In words, scaling the estimated eigenvectors to $\sqrt{\lambda_1(\hat{A})}$ means that the outer product of $\hat{C}^{(\infty)}$ is the best rank-1 approximation of $\hat{A}$. In proposing eigenvector centrality, Bonacich (1972) in fact cites this property as one of the key motivations, arguing that $\sqrt{\lambda_1(\hat{A})} v_1(\hat{A})$ can be interpreted as the "social interaction potential" of a given agent.

Remark 4. We can compare our results to that of Cai et al. (2021) by setting $a_n = \sqrt{n}$. The condition in Case (a) (ii) of Theorem 5 specializes to $p_n \succ n^{-2/3}$. Cai et al. (2021) can accommodate $p_n \succ n^{-1}$ but they assume measurement error that is additive and i.i.d. Gaussian.

Remark 5. Le and Li (2020) provides methods for testing the hypothesis $\|\beta^{(\infty)} C^{(\infty)}\|^2 = 0$ when $p_n \succ n^{-1/2}$. They accommodate regressions on multiple eigenvectors, but in a setting with only one eigenvector, their result asserts that the t-statistic with the homoskedastic variance estimator can be used to test the hypothesis that $\beta^{(\infty)} = 0$. Theorem 5 does not cover regression on multiple eigenvectors but it accommodates greater sparsity and facilitates tests of $\beta^{(\infty)} = \beta_0$ for $\beta_0 \neq 0$.

4 Simulations

In this section, we present simulation evidence to support our theory. We will consider the unobserved adjacency matrix A defined as

$$A_{ij} = \begin{cases} p_n & \text{if } i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the graphon is $f = 1$. The observed adjacency matrix is $\hat{A}$, where for $i > j$, $\hat{A}_{ij} = \text{Bernoulli}(A_{ij})$. $\hat{A}_{ii} = 0$, $\hat{A}_{ji} = \hat{A}_{ij}$.

Our regression model is:

$$Y_i = \beta C_i^{(d)} + \varepsilon_i$$

where $C_i^{(d)}$ are centrality measures calculated on A. In this simulation, we draw $U_i \stackrel{\text{i.i.d.}}{\sim} U[0,1]$, $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{N}(0,1)$, where $\varepsilon_i \perp\!\!\!\perp U_i$ and $\varepsilon_i \perp\!\!\!\perp \hat{A}_{jk}$ for all $i, j, k \in [n]$. We will set $\beta = 1$. As before, $\tilde{\beta}^{(d)}$ is used when A is observed, and $\hat{\beta}^{(d)}$ is used when only $\hat{A}$ is observed.

We revisit our three sets of results in turn: inconsistency under sparsity, bias correction and normal approximation.

4.1 Inconsistency Under Sparsity

The first regime of interest is $p_n = 1/n$. Theorem 1 asserts that $\tilde{\beta}^{(1)}$ and $\tilde{\beta}^{(T)}$ are consistent. $\tilde{\beta}^{(\infty)}$ is consistent provided that $a_n \to \infty$ for eigenvector centrality.

We start with the last claim, which is supported by Figure 3. For $n = 100$, we see that the choice of scaling clearly affects how well the estimator is able to concentrate around $\beta = 1$. The plots for larger values of n are qualitatively similar. This also hints at the trade-off that made in Theorem 5: we can choose $a_n = \sqrt{\lambda_1(\hat{A})}$ so that the distribution of $\hat{\beta}^{(\infty)}$ is easy to characterize, but this will slow down the rate of convergence. Since this subsection concerns current practice, we will set $a_n = \sqrt{n}$ for its remainder.

We return to the first claim concerning consistency of $\tilde{\beta}^{(d)}$ when $p_n = 1/n$. Figure 4 indeed shows the distribution of $\tilde{\beta}^{(d)}$ for each n. The estimators concentrate around β as n increases, in line with our result. However, Theorem 2 asserts that $\hat{\beta}^{(1)}$, $\hat{\beta}^{(T)}$ and $\hat{\beta}^{(\infty)}$ are all inconsistent when $p_n = 1/n$. Their distributions, presented in Figure 5, concords with our result. Indeed, we see that $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$ are attenuated by constant amount as $n \to \infty$, while $\hat{\beta}^{(\infty)}$ converges in probability to 0.

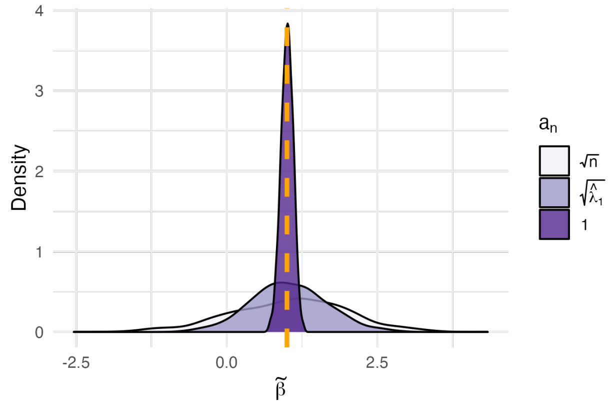

Figure 3: Distribution of $\tilde{\beta}^{(\infty)}$ for $n = 100$, $p_n = 1/n$ under various a_n. $\beta = 1$ (orange dashed line).

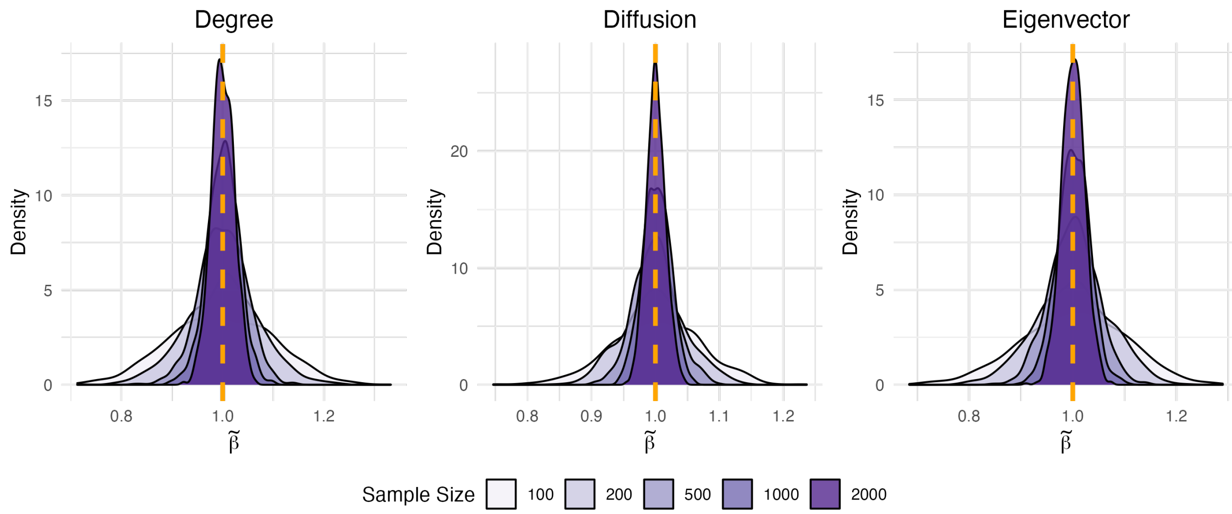

Figure 4: Distribution of $\tilde{\beta}^{(d)}$ for $p_n = 1/n$. For $\tilde{\beta}^{(\infty)}$, $a_n = \sqrt{n}$. $\beta = 1$ (orange dashed line).

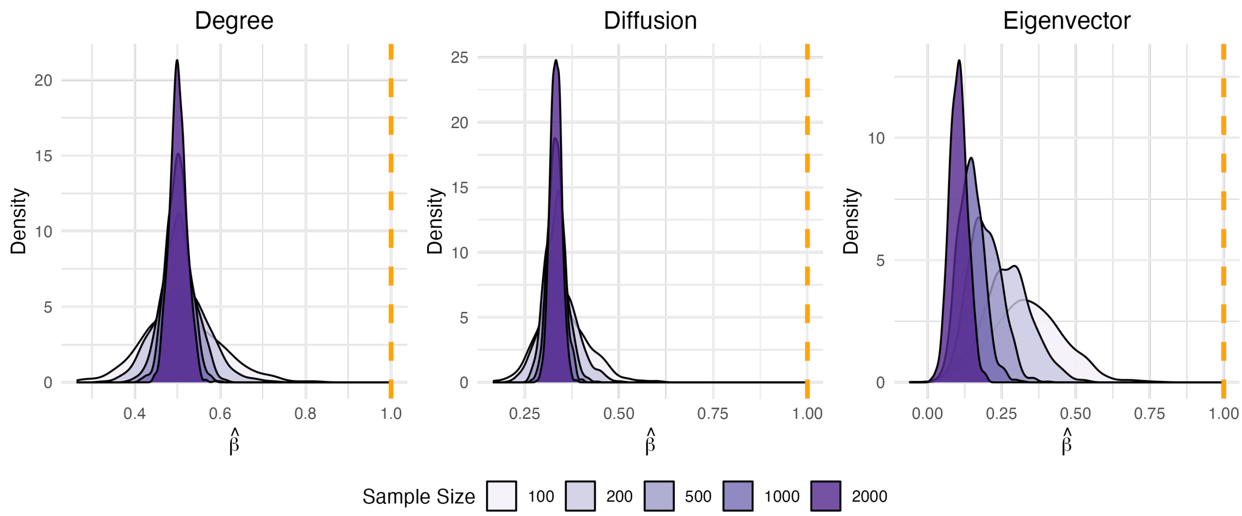

Figure 5: Distribution of $\hat{\beta}^{(d)}$ for $p_n = 1/n$. For $\tilde{\beta}^{(\infty)}$, $a_n = \sqrt{n}$. $\beta = 1$ (orange dashed line).

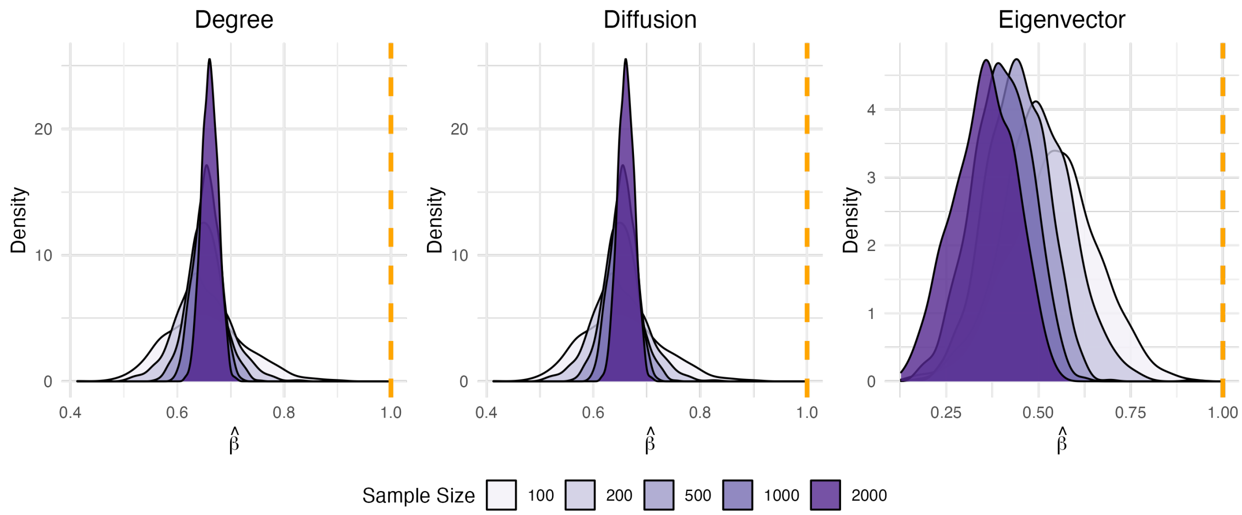

Figure 6: Distribution of $\hat{\beta}^{(d)}$ for $p_n = n^{-1}\sqrt{\log n/\log\log n}$. For $\tilde{\beta}^{(\infty)}$, $a_n = \sqrt{n}$. $\beta = 1$ (orange dashed line).

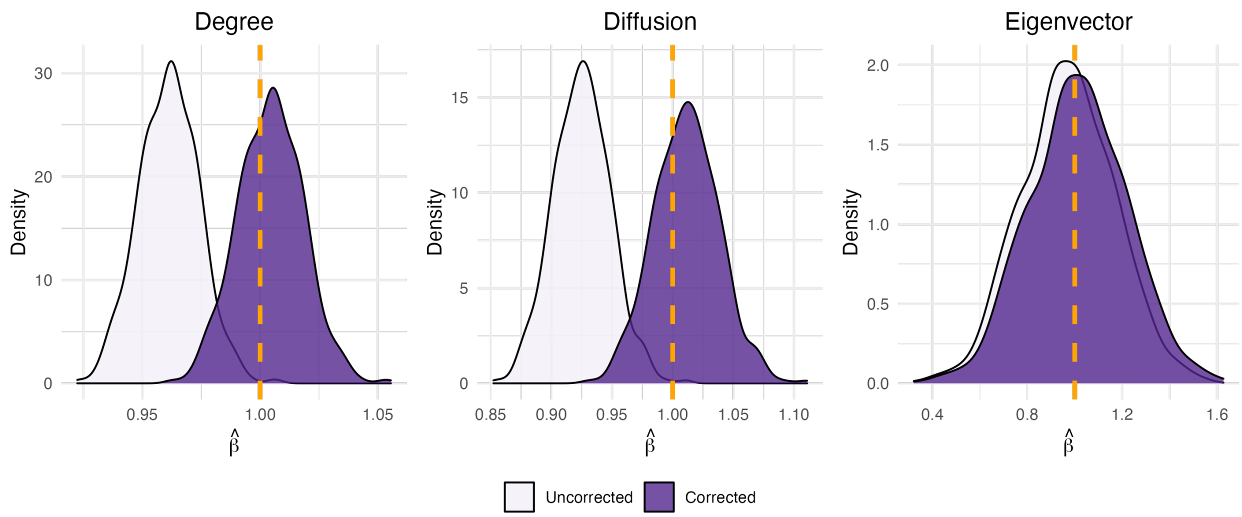

Figure 7: Distributions of $\hat{\beta}^{(d)}$ and their bias corrected versions $\check{\beta}^{(d)}$ for $p_n = 1/\sqrt{n}$. $\beta = 1$ (orange dashed line).

Finally, we consider the case when $p_n = n^{-1}\sqrt{\frac{\log n}{\log\log n}}$. In this regime, $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$ are consistent but $\hat{\beta}^{(\infty)}$ is not. We see suggestive evidence of this in Figure 6, where $\hat{\beta}^{(\infty)}$ is drifting further away from β as n increases. The opposite occurs with $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$. Though the rate of convergence is slow, it is visible.

4.2 Bias Correction

Even in regime dense enough such that $\hat{\beta}^{(1)}$, $\hat{\beta}^{(T)}$ and $\hat{\beta}^{(\infty)}$ are consistent, they can still be subject to biases that affect their rates of convergence. This motivates the bias-corrected

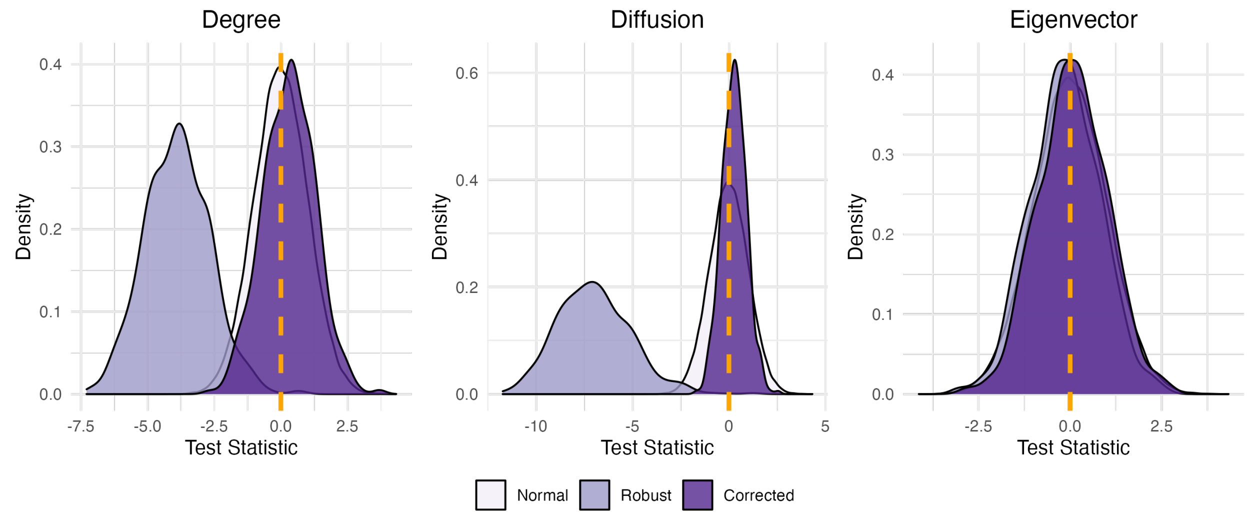

Figure 8: Distribution of the centered and scaled test statistics in Theorems 4 and 5. Robust refers to tests based on t-statistic with robust (heteroskedasticity consistent) standard errors.

estimators in Definition 9. In this subsection, we study the effects of bias correction in the regime $p_n = 1/\sqrt{n}$.

Figure 7 shows the distribution of the estimators when $n = 500$. Here $a_n = \sqrt{\hat{\lambda}_1(\hat{A})}$. We see that bias correction is effective in correctly centering $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$. The same is true for $\hat{\beta}^{(\infty)}$ though to a smaller extent, in line with claims in Corollary 4. Results for other values of n are qualitatively similar.

4.3 Distributional Theory

Finally, we investigate the quality of the normal approximations proposed in Theorems 4 and 5. As before, we consider the regime $p_n = 1/\sqrt{n}$. Figure 8 presents the distribution of test statistics (in purple) which our theorems predict have the standard normal distribution (in gray). We see that the two distributions are indeed close. It is also common for applied researcher to compute the usual t-statistic with heteroskedaticity consistent (robust) standard errors and conduct inference under the assumption that it has a standard normal distribution. For comparison, we include the distribution of the t-statistic (in lavender). Corollary 4 justifies the use of this statistic when $a_n = \sqrt{\hat{\lambda}_1(\hat{A})}$ but our theory for degree and diffusion centralities is based on a different statistic. Indeed, we see that the robust t-statistic can be quite far from the standard normal distribution in both location and dispersion. This suggests that our method would lead to more reliable tests.

We next examine the size and power of tests based on our distributional theory. We consider testing the hypothesis $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ at 5% level of significance.

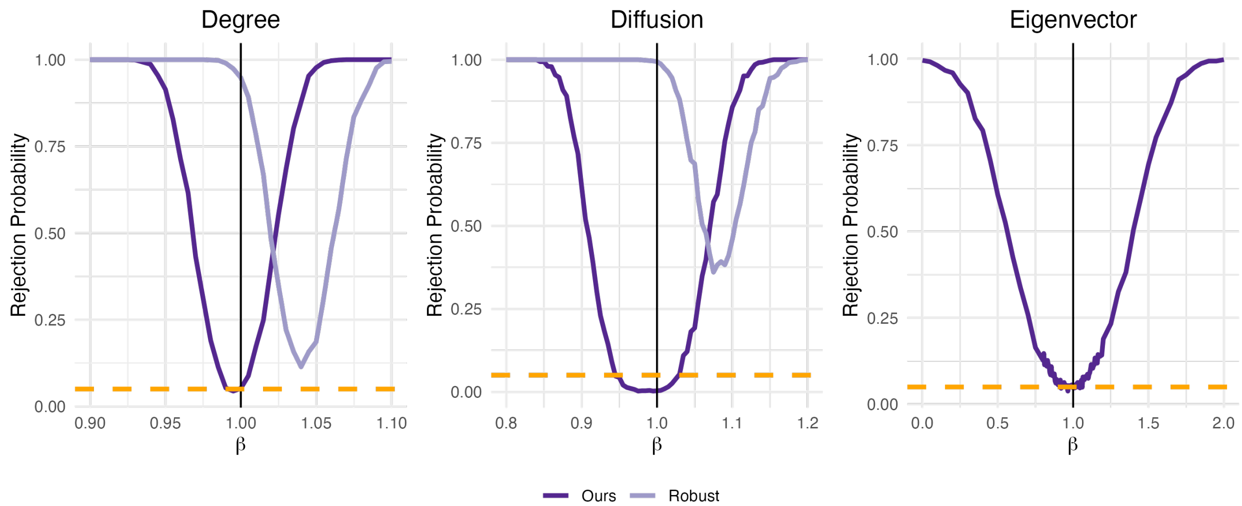

Figure 9: Power of the two-sided test of $H_0 : \beta = 1$ under various alternatives. Test at 5% level of significance (orange dashed line).

Table 1 presents size of the test when $\beta = 1$ is correctly specified. For degree and diffusion centralities, our theory provides test statistics which differ from the robust t-statistic. As we see from the table, the test for degree controls size well. The test for diffusion centrality is more conservative. This is because our estimator $\hat{V}^{(T)}$ is upward-biased in finite sample when there are no other covariates. Tests for degree and diffusion centralities that are based on the robust t-statistic has Type I error over 50% across all sample sizes. For eigenvector centrality, our theory predicts that the robust t-statistic will perform well Indeed, it has size close to 5%. We also consider testing the hypothesis $\beta = 0$. Power for this test is presented in Table 2. For this null hypothesis, our theory suggests the use of the robust t-statistic. Reassuringly, the tests all have power close to 1. To understand how power changes as we vary the alternative hypothesis, we hone in on the case where $n = 500$ and $p_n = 1/\sqrt{n}$. Figure 9 presents the rejection probability of our test under various alternatives. We see that the our tests control size and have good power. Comparatively, tests based on the robust t-statistic have poor size control when $\beta \neq 0$. Furthermore, they can have poor power against particular alternatives owing to the bias. We conclude that our tests have desirable properties and are preferred to the test with robust t-statistic when networks are sparse and observed with noise.

p_n	Statistic		Sample Size				
			100	200	500	1000	2000
0.1	Degree	Ours	0.055	0.052	0.067	0.062	0.065
		Robust	0.573	0.645	0.680	0.662	0.670
	Diffusion	Ours	0.010	0.004	0.007	0.003	0.004
		Robust	0.859	0.883	0.883	0.868	0.897
	Eigenvector		0.054	0.060	0.046	0.052	0.046
$n^{-1/3}$	Degree	Ours	0.066	0.065	0.067	0.058	0.065
		Robust	0.286	0.422	0.556	0.701	0.778
	Diffusion	Ours	0.006	0.001	0.002	0.008	0.003
		Robust	0.621	0.729	0.805	0.888	0.933
	Eigenvector		0.045	0.055	0.050	0.056	0.038
$n^{-1/2}$	Degree	Ours	0.072	0.049	0.051	0.037	0.062
		Robust	0.580	0.761	0.944	0.993	0.999
	Diffusion	Ours	0.015	0.009	0.005	0.001	0.004
		Robust	0.863	0.944	0.993	1.000	1.000
	Eigenvector		0.053	0.049	0.036	0.067	0.056

Table 1: Size of 5% level two-sided tests when $\beta = 1$ is correctly specified. Robust refers to tests based on t-statistic with robust (heteroskedasticity consistent) standard errors.

p_n	Statistic	Sample Size				
		100	200	500	1000	2000
0.1	Degree	1.000	1.000	1.000	1.000	1.000
	Diffusion	1.000	1.000	1.000	1.000	1.000
	Eigenvector	0.880	0.993	1.000	1.000	1.000
$n^{-1/3}$	Degree	1.000	1.000	1.000	1.000	1.000
	Diffusion	1.000	1.000	1.000	1.000	1.000
	Eigenvector	0.994	1.000	1.000	1.000	1.000
$n^{-1/2}$	Degree	1.000	1.000	1.000	1.000	1.000
	Diffusion	1.000	1.000	1.000	1.000	1.000
	Eigenvector	0.860	0.972	0.997	1.000	1.000

Table 2: Power of 5% level two-sided tests of $H_0 : \beta = 0$ when $\beta = 1$. Under this H_0, the our test statistics is the usual t-statistic with robust (heteroskedasticity-consistent) standard errors.

5 Empirical Demonstration

In this section, we demonstrate the relevance of our theoretical findings via an application inspired by De Weerdt and Dercon (2006).[1] In the developing world, social insurance is an important mechanism for smoothing consumption, because of restricted access to formal credit markets (Rosenzweig 1988; Udry 1994; Fafchamps and Lund 2003; Kinnan and Townsend 2012, among many others). De Weerdt and Dercon (2006) examines the case of Nyakatoke, a village with 120 households in rural Tanzania, and find that social insurance helps households to smooth consumption following health shocks. The data they use comprises five rounds of panel data on household consumption, illness among other covariates, collected from February to December 2000. The authors also had access to social network data collected during the first round of the survey, in which households were asked for the identities of those who they depend on or depend on them for help. The authors then regress a household's change in consumption following illness on the mean consumption of their network neighbors, finding evidence of positive co-movements.

Another way to demonstrate the effect of social insurance on consumption smoothing could be to regress variance in consumption on network centrality measures. Specifically, the regression:

$$Y_i = \beta C_i^{(d)} + \varepsilon_i$$

where Y_i is variance in food expenditure and $C_i^{(d)}$ is a centrality measure. The above regression could be preferred to the authors' specification if we are unsure about the covariates that reflect social assistance. For example, it might be a household's stock of savings that co-move with the decision to lend to their friends, rather than their own consumption. We might also be interested in more complex patterns of assistance, which could be summarized in an appropriate centrality measure, but which might not be tractable with covariates.

The above regression requires information on network of social insurance, in which each entry A_{ij} records the probability i lends money to j over the survey period. We can consider obtaining proxies for this network using one of the following:

Unilateral Social (US). $\hat{A}_{ij} = 1$ if either i or j names the other household as a party that they could depend on or which depends on them for help.

Bilateral Social (BS). $\hat{A}_{ij} = 1$ only if both i and j names the other household as a party that they could depend on or which depends on them for help.

[1] The data is obtained from Joachim De Weerdt's website: https://www.uantwerpen.be/en/staff/joachim-deweerdt/public-data-sets/nyakatoke-network/.

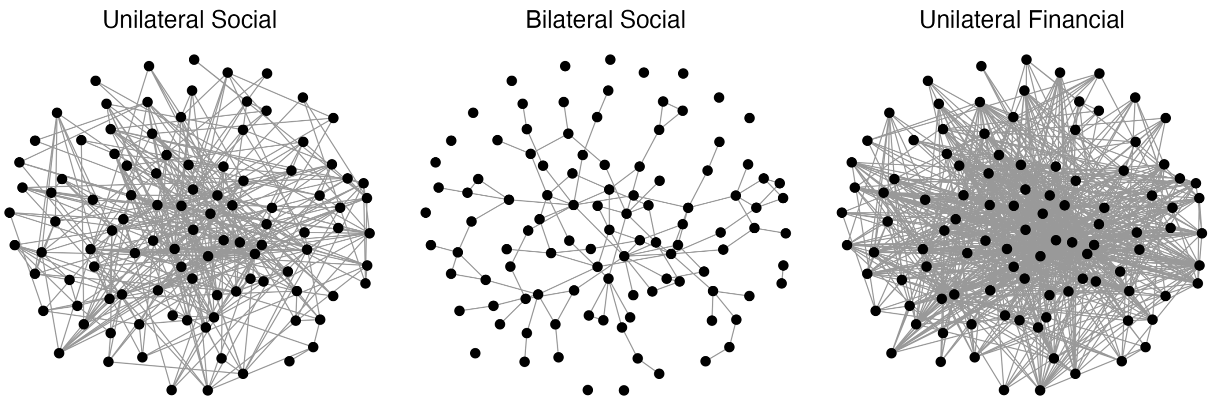

Figure 10: Social and Financial networks in Nyakatoke.

> **Unilateral Financial (UF)**. $\hat{A}_{ij} = 1$ if either i or j lends money to the other at least once over the survey period.

The authors study US and BS. We also consider UF since self-reported loan data is available. The networks are plotted in Figure 10 and the degree distributions are described in Table 3. In a village with $n = 120$ households, mean degree in all three networks are less than 11. We might therefore be concerned that $np_n \prec \sqrt{n}$. By construction, US is much less dense than BS. Due to the availability of five panels, UF is denser than the other two.

$(n = 119)$	Mean	Median	Min	Max
Unilateral Social	8.10	7	0	31
Bilateral Social	2.32	2	0	10
Unilateral Transfer	10.48	8	1	77

Table 3: Degree distributions of various networks in Nyakatoke.

Regression results are presented in Table 4. In this exercise, $a_n = \sqrt{\lambda_1(\hat{A})}$, $\delta = 1/\sqrt{\lambda_1(\hat{A})}$ and $T = 2$. We first note that estimated attenuation factor is the smallest (i.e. furthest from 1) in the sparsest network BS. This is in line with our result that bias is $O_p(n^{-1}p_n^{-1})$. Diffusion centrality is generally estimated to be less attenuated, because δ is small (≈ 0.2). As the last column shows, bias correction can lead to substantially different estimates. Table 4 also presents p-values for tests of the two-sided hypothesis that $\beta^{(d)} = 0$. Centrality statistics on BS appears to be more predictive of the variance in food consumption than that on US and UF. This highlights that researchers should not choose their network proxies on the criteria of sparsity alone. In the case of Nyakatoke, evidence suggests that US reflects "desire to link", rather than actual risk-pooling (Comola and Fafchamps 2014), such that US is a noisier proxy than BS. By the same account, the large number of discrepancies between reporting by borrowers and lenders of the same loan suggests that the loan data is subject to severe mis-reporting (Comola and Fafchamps 2017), rendering it an equally poor proxy. Among the centrality statistics on BS, eigenvector has the least predictive power by far. We reiterate our warning that eigenvector centrality is less robust to sparsity than degree and diffusion, such that the p-values might reflect the poor statistical properties of the measure, rather than its lack of economic significance.

		Estimate	p-value	Atten.	Bias Corr.
Unilateral Social	Degree	-1064	0.67	0.91	-1172
	Diffusion	-4274	0.77	1.00	-4292
	Eigenvector	-12353	0.86	0.91	-13548
Bilateral Social	Degree	-11604	0.06	0.74	-15592
	Diffusion	-23672	0.16	0.94	-25212
	Eigenvector	-10543	0.93	0.78	-13434
Unilateral Financial	Degree	-412	0.70	0.96	-429
	Diffusion	-4559	0.74	1.00	-4561
	Eigenvector	-15040	0.77	0.96	-15699

Table 4: Regression results for various networks. Estimate is $\hat{\beta}^{(d)}$. p-value is for the two-sided test that $H_0 : \beta = 0$. Atten. is the estimated attenuation factor of $\hat{\beta}^{(d)}$ (i.e. $1 - \hat{B}^{(d)}$). Bias Corr. presents the bias corrected estimates, $\check{\beta}^{(d)}$.

Finally, we present one-sided confidence intervals for values of $\beta^{(d)}$ based on our results. These are useful for putting bounds on parameter values. In our example, a lower bound could be intuitively interpreted as the limits to informal risk-sharing, a quantity which could be useful for policymakers deciding whether or not to provide agricultural insurance. We focus on BS since it appears to be the only informative network. Results for degree and diffusion are presented in Table 5. Our confidence intervals leads to tighter lower bound than the those based on the robust t-statistic. Results for eigenvector are omitted since our theory is based on the usual robust t-statistic.

		90%	95%	99%
Degree	Ours	(-18835, ∞)	(-20015, ∞)	(-22680, ∞)
	Robust	(-25050, ∞)	(-28862, ∞)	(-36012, ∞)
Diffusion	Ours	(-34715, ∞)	(-38867, ∞)	(-50112, ∞)
	Robust	(-56938, ∞)	(-66368, ∞)	(-84058, ∞)

Table 5: One-sided confidence intervals for degree and diffusion.

6 Conclusion

In this paper, we studied the properties of linear regression on degree, diffusion and eigenvector centrality when networks are sparse and observed with error. We show that these issues threaten the consistency of OLS estimators and characterize the amount of sparsity at which inconsistency occurs. In doing so, we find that eigenvector centrality is less robust to sparsity than the others and that the statistical properties of the corresponding regression is sensitive to the scaling.

Additionally, we show that an asymptotic bias arises whenever the true slope parameter is not 0 and that the bias can be of larger order than the variance, so that bias correction is necessary to obtain a non-degenerate limiting distribution from the OLS estimator. Finally, we provide estimators for the bias and variance which, together with our central limit theorem, facilitates inference under sparsity and measurement error. We confirm our theoretical results via simulations, which suggest that our approximation result works better for estimation and inference when networks are sparse, particularly when compared to the use of robust standard errors and the associated t-statistics. Finally, we demonstrate the relevance of our theoretical results by studying the social insurance network in Nyakatoke, Tanzania.

In sum, our results suggest that applied researchers view their results with caution when applying OLS to sparse, noisy networks. Specifically, comparing the statistical significance of eigenvector centrality with degree or diffusion may yield misleading conclusions since they differ not only in economic significance but also statistical properties. Provided that the networks are not *too* sparse, the usual t-test is valid for null hypothesis that the slope parameter is 0. However, alternative inference procedures will be necessary for other null hypotheses. Additionally, there may be scope for improving estimation by the use of bias-corrected estimators. Estimation and inference under extreme sparsity remains an open question, though as (Le et al. 2017) show, parametric models such as the Stochastic Block Model may point to a way forward.

References

Adcock, R. J. (1878). A problem in least squares. *The Analyst 5*(1), 53–54.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis 11*(4), 581–598.

Alt, J., R. Ducatez, and A. Knowles (2021a). Extremal eigenvalues of critical Erdős–Rényi graphs. *The Annals of Probability 49*(3), 1347–1401.

Alt, J., R. Ducatez, and A. Knowles (2021b). Poisson statistics and localization at the spectral edge of sparse Erdős–Rényi graphs. *arXiv preprint arXiv:2106.12519*.

Andrews, I., J. H. Stock, and L. Sun (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics 11*(1).

Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association 116*(536), 1716–1730.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science 341*(6144), 1236498.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *The Review of Economic Studies 86*(6), 2453–2490.

Benaych-Georges, F., C. Bordenave, and A. Knowles (2019). Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs. *The Annals of Probability 47*(3), 1653–1676.

Benaych-Georges, F., C. Bordenave, and A. Knowles (2020). Spectral radii of sparse random matrices. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, Volume 56, pp. 2141–2161. Institut Henri Poincaré.

Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences 106*(50), 21068–21073.

Bickel, P. J., A. Chen, and E. Levina (2011). The method of moments and degree distributions for network models. *The Annals of Statistics 39*(5), 2280–2301.

Bloch, F., M. O. Jackson, and P. Tebaldi (2021). Centrality measures in networks. *arXiv preprint arXiv:1608.05845*.

Bollobás, B., S. Janson, and O. Riordan (2007). The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms 31*(1), 3–122.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of mathematical sociology 2*(1), 113–120.

Borgs, C., J. T. Chayes, H. Cohn, and N. Holden (2018). Sparse exchangeable graphs and their limits via graphon processes. *Journal of Machine Learning Research 18*, 1–71.

Bourlès, R., Y. Bramoullé, and E. Perez-Richet (2021). Altruism and risk sharing in networks. *Journal of the European Economic Association 19*(3), 1488–1521.

Bramoullé, Y. and G. Genicot (2018). Diffusion centrality: Foundations and extensions.

Breza, E., A. Chandrasekhar, B. Golub, and A. Parvathaneni (2019). Networks in economic development. *Oxford Review of Economic Policy 35*(4), 678–721.

Cai, J., D. Yang, W. Zhu, H. Shen, and L. Zhao (2021). Network regression and supervised centrality estimation. *Available at SSRN 3963523*.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Candès, E. J. and T. Tao (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory 56*(5), 2053–2080.

Carvalho, V. M., M. Nirei, Y. U. Saito, and A. Tahbaz-Salehi (2021). Supply chain disruptions: Evidence from the great east japan earthquake. *The Quarterly Journal of Economics 136*(2), 1255–1321.

Carvalho, V. M. and A. Tahbaz-Salehi (2019). Production networks: A primer. *Annual Review of Economics 11*, 635–663.

Chandrasekhar, A. (2016). Econometrics of network formation. *The Oxford handbook of the economics of networks*, 303–357.

Chandrasekhar, A. and R. Lewis (2016). Econometrics of sampled networks. *Working Paper*.

Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. *The Annals of Statistics 43*(1), 177–214.

Chen, L. H., L. Goldstein, and Q.-M. Shao (2011). *Normal approximation by Stein's method*, Volume 2. Springer.

Chen, Y., C. Cheng, and J. Fan (2021). Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices. *Annals of statistics 49*(1), 435.

Cheng, C., Y. Wei, and Y. Chen (2021). Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. *IEEE Transactions on Information Theory 67*(11), 7380–7419.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica 71*(5), 1405–1441.

Comola, M. and M. Fafchamps (2014). Testing unilateral and bilateral link formation. *The Economic Journal 124*(579), 954–976.

Comola, M. and M. Fafchamps (2017). The missing transfers: Estimating misreporting in dyadic data. *Economic Development and Cultural Change 65*(3), 549–582.

Conley, T. G. and C. R. Taber (2011). Inference with "difference in differences" with a small number of policy changes. *The Review of Economics and Statistics 93*(1), 113–125.

Cruz, C., J. Labonne, and P. Querubin (2017). Politician family networks and electoral outcomes: Evidence from the philippines. *American Economic Review 107*(10), 3006–37.

De Paula, A. (2017). Econometrics of network models. In *Advances in economics and econometrics: Theory and applications, eleventh world congress*, pp. 268–323. Cambridge University Press Cambridge.

De Paula, A., I. Rasul, and P. Souza (2020). Recovering social networks from panel data: identification, simulations and an application.

De Weerdt, J. and S. Dercon (2006). Risk-sharing networks and insurance against illness. *Journal of development Economics 81*(2), 337–356.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical association 69*(345), 118–121.

Dong, H., N. Chen, and K. Wang (2020). Modeling and change detection for count-weighted multilayer networks. *Technometrics 62*(2), 184–195.

Eagle, N., M. Macy, and R. Claxton (2010). Network diversity and economic development. *Science 328*(5981), 1029–1031.

Eldridge, J., M. Belkin, and Y. Wang (2018). Unperturbed: spectral analysis beyond davis-kahan. In *Algorithmic Learning Theory*, pp. 321–358. PMLR.

Evdokimov, K. (2010). Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Technical report, Citeseer.

Evdokimov, K. and A. Zeleneev (2020). Errors-in-variables in large nonlinear panel and network models. Technical report, Cemmap Working Paper.

Evdokimov, K. S. and A. Zeleneev (2022). Simple estimation of semiparametric models with measurement errors. Technical report, Cemmap Working Paper.

Fafchamps, M. and S. Lund (2003). Risk-sharing networks in rural Philippines. *Journal of development Economics 71*(2), 261–287.

Feige, U. and E. Ofek (2005). Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms 27*(2), 251–275.

Graham, B. S. (2020a). Network data. In *Handbook of Econometrics*, Volume 7, pp. 111–218. Elsevier.

Graham, B. S. (2020b). Sparse network asymptotics for logistic regression.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology 78*(6), 1360–1380.

Griffith, A. (2022). Name your friends, but only five? the importance of censoring in peer effects estimates using social network data. *Journal of Labor Economics 40*(4), 779–805.

Griliches, Z. and J. A. Hausman (1986). Errors in variables in panel data. *Journal of econometrics 31*(1), 93–118.

Hansen, B. E. and S. Lee (2019). Asymptotic theory for clustered samples. *Journal of econometrics 210*(2), 268–290.

Hochberg, Y. V., A. Ljungqvist, and Y. Lu (2007). Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance 62*(1), 251–301.

Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social networks 5*(2), 109–137.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ 2*, 275.

Hu, Y. (2017). The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics. *Journal of econometrics 200*(2), 154–168.

Jackson, M. O. (2010). *Social and economic networks.* Princeton university press.

Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics 36*(4), 705–713.

Kinnan, C. and R. Townsend (2012). Kinship and financial networks, formal financial access, and risk reduction. *American Economic Review 102*(3), 289–93.

Le, C. M., E. Levina, and R. Vershynin (2017). Concentration and regularization of random graphs. *Random Structures & Algorithms 51*(3), 538–561.

Le, C. M. and T. Li (2020). Linear regression and its inference on noisy network-linked data. *arXiv preprint arXiv:2007.00803*.

Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics 43*(1), 215–237.

Leider, S., M. M. Möbius, T. Rosenblat, and Q.-A. Do (2009). Directed altruism and enforced reciprocity in social networks. *The Quarterly Journal of Economics 124*(4), 1815–1851.

Levin, K. and E. Levina (2019). Bootstrapping networks with latent space structure. *arXiv preprint arXiv:1907.10821*.

Lewbel, A., X. Qu, X. Tang, et al. (2021). *Social Networks with Mismeasured Links.* Boston College.

Li, T., E. Levina, and J. Zhu (2020). Network cross-validation by edge sampling. *Biometrika 107*(2), 257–276.

Lovász, L. (2012). *Large networks and graph limits*, Volume 60. American Mathematical Soc.

Manresa, E. (2016). Estimating the structure of social interactions using panel data. *Working Paper*.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica 71*(5), 1339–1375.

Motalebi, N., N. T. Stevens, and S. H. Steiner (2021). Hurdle blockmodels for sparse network modeling. *The American Statistician 75*(4), 383–393.

Negahban, S. and M. J. Wainwright (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research 13*(1), 1665–1697.

Rajkumar, K., G. Saint-Jacques, I. Bojinov, E. Brynjolfsson, and S. Aral (2022). A causal test of the strength of weak ties. *Science 377*(6612), 1304–1310.

Reagans, R. and E. W. Zuckerman (2001). Networks, diversity, and productivity: The social capital of corporate r&d teams. *Organization science 12*(4), 502–517.

Rose, C. (2016). Identification of spillover effects using panel data. Technical report, Working Paper.

Rosenzweig, M. R. (1988). Risk, implicit contracts and the family in rural areas of low-income countries. *The Economic Journal 98*(393), 1148–1170.

Rothenberg, T. J. (1984). Approximating the distributions of econometric estimators and test statistics. *Handbook of econometrics 2*, 881–935.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, Volume 3, pp. 249–277. Elsevier.

Schennach, S. M. (2020). Mismeasured and unobserved variables. In *Handbook of Econometrics*, Volume 7, pp. 487–565. Elsevier.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Tao, T. (2012). *Topics in random matrix theory*, Volume 132. American Mathematical Soc.

Thirkettle, M. (2019). Identification and estimation of network statistics with missing link data. *Working Paper*.

Udry, C. (1994). Risk and insurance in a rural credit market: An empirical investigation in northern nigeria. *The Review of Economic Studies 61*(3), 495–526.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

Veitch, V. and D. M. Roy (2019). Sampling and estimation for (sparse) exchangeable graphs. *The Annals of Statistics 47*(6), 3274–3299.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge university press.

Wang, H. (2020). Logistic regression for massive data with rare events. In *International Conference on Machine Learning*, pp. 9829–9836. PMLR.

Wigner, E. P. (1957). Characteristics vectors of bordered matrices with infinite dimensions ii. *Annals of Mathematics*, 203–207.

Xu, G. (2018). The costs of patronage: Evidence from the british empire. *American Economic Review 108*(11), 3170–98.

Young, S. J. and E. R. Scheinerman (2007). Random dot product graph models for social networks. In *International Workshop on Algorithms and Models for the Web-Graph*, pp. 138–149. Springer.

Appendices

A Proofs

We maintain Assumptions 1, 2 and 3 throughout the entire section. Without loss of generality, let $f(u,v) = f(v,u)$ and the $f(u,u) = 0$. Further define:

$$W = \int_{[0,1]^2} f(u,v)\, dudv \ .$$

We state below a convenient lemma:

Lemma 1 (Concentration in Spectral Norm)**.** Suppose Assumptions 1 and 2 hold. Let $\nu \in (0,1)$. Then with probability at least $1 - \exp\left(-n^2 p_n^2 \sqrt{k \frac{\log n}{\log \log n}}\right)$

$$\left\| A - \hat{A} \right\| \leq k \left(np_n\right)^{(1+\nu)/2} \left(\frac{\log n}{\log \log n}\right)^{(1-\nu)/4}$$

where k is a universal constant. In other words,

$$\left\| A - \hat{A} \right\| = O_p\left(\left(np_n\right)^{(1+\nu)/2} \left(\frac{\log n}{\log \log n}\right)^{(1-\nu)/4} \right) \ . \tag{9}$$

A.1 Proof of Theorem 1

In the setting with no measurement error, we write:

$$\tilde{\beta}^{(d)} = \frac{\sum_{i=1}^n Y_i C_i^{(d)}}{\sum_{i=1}^n \left(C_i^{(d)}\right)^2} = \beta + \frac{\sum_{i=1}^n C_i^{(d)} \varepsilon_i}{\sum_{i=1}^n \left(C_i^{(d)}\right)^2} \ .$$

We first show that OLS is consistent when the lower bounds in the Theorem obtains. Start with degree:

$$\begin{aligned}
\sum_{i=1}^n C_i^{(1)} \varepsilon_i &= \sum_{i=1}^n \sum_{j=1}^n p_n f(U_i, U_j) \varepsilon_i \\
&= \frac{p_n}{2} \binom{n}{2} \cdot \frac{1}{\binom{n}{2}} \sum_{i=1}^n \sum_{j=1}^n f(U_i, U_j)\varepsilon_i + f(U_j, U_i)\varepsilon_j \qquad (10) \\
&= O_p\left(n^{3/2} p_n\right) \ ,
\end{aligned}$$

In the last equality, we use our assumption that $E[\varepsilon_i|U_i] = 0$ and $E[\varepsilon_i^2|U_i] \leq \sigma^2 < \infty$, so that

$$\sqrt{n} \cdot \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} f(U_i, U_j)\varepsilon_i + f(U_j, U_i)\varepsilon_j \xrightarrow{d} \mathrm{N}(0, \gamma)$$

for some $\gamma > 0$ by the standard CLT for U-statistics (e.g. Theorem 12.3 in Van der Vaart 2000). Similarly,

$$\begin{aligned}\sum_{i=1}^{n} \left(C_i^{(1)}\right)^2 &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} p_n f(U_i, U_j)\right)^2 \\ &= p_n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} f(U_i, U_j) f(U_i, U_k) \\ &= p_n^2 \binom{n}{3} \cdot \frac{1}{\binom{n}{3}} \sum_{k=1}^{n} f(U_i, U_j) f(U_i, U_k) = O_p(n^3 p_n^2)\end{aligned} \tag{11}$$

By the LLN for U-statistics,

$$\frac{1}{\binom{n}{3}} \sum_{k=1}^{n} f(U_i, U_j) f(U_i, U_k) \xrightarrow{p} \gamma$$

Next, note that $\gamma > 0$. Let k and B_k be such that $f(u, v) > 1/k$ for all $(u, v) \in B_k$. Our assumption $W > 0$ ensures that there exists k such that $P_k := \int_{[0,1]^2} \mathbf{1}_{B_k} \, du dv > 0$. Then,

$$\begin{aligned}\gamma &= \int_{[0,1]^3} f(u_1, u_2) f(u_1, u_3) \, du_1 du_2 du_3 \\ &\geq \int_{\pi_1(B_k) \times \pi_2(B_k) \times \pi_2(B_k)} k^{-2} \, du_1 du_2 du_3 \geq P_k^2 k^{-2} > 0\end{aligned}$$

where $\pi_j(B_k)$ denotes the projection of B_k onto the j^{th} coordinate. Hence, we have consistency if $n^{3/2} p_n \to \infty$. If $n^{3/2} p_n \approx 1$, the $\tilde{\beta}^{(1)} - \beta^{(1)}$ converges to a normal distribution. If $n^{3/2} p_n \prec 1$, $\tilde{\beta}^{(1)} - \beta^{(1)}$ diverges. Hence, we have consistency if and only if $n^{3/2} p_n \to \infty$

Remark 6. Note that the analysis for normalized degree is similar, except that the numerator is now of order $n^{1/2} p_n$ while the denominator is of order $n p_n^2$. Consistency with the normalized degree thus require $n^{1/2} p_n \to \infty$ as noted in Remark **??**.

Next, consider diffusion centrality. Note that:

$$\sum_{i=1}^{n} C_i^{(T)} \varepsilon_i = \sum_{t=1}^{T} \delta^t \cdot \iota_n' A^t \varepsilon \quad , \quad \sum_{i=1}^{n} \left(C_i^{(T)}\right)^2 = \sum_{t=1}^{T} \delta^{2t} \cdot \iota_n' A^{2t} \iota_n \ .$$

We will identify the dominant terms in the numerator and denominator respectively in each regime of p_n. For $t \geq 2$, write:

$$\left[A^t\right]_{ij} = p_n^t \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_{t-1}=1}^{n} f(U_i, U_{k_1}) f(U_{k_1}, U_{k_2}) \cdots f(U_{k_{t-1}}, U_j) \ .$$

Applying the CLT for U-statistics as before, we have that

$$\iota_n' A^t \varepsilon = p_n^t \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_{t-1}=1}^{n} f(U_i, U_{k_1}) f(U_{k_1}, U_{k_2}) \cdots f(U_{k_{t-1}}, U_j) \varepsilon_j = O_p\left(p_n^t n^{t+1/2}\right) \ .$$

Similarly,

$$\iota_n' A^{2t} \iota_n = p_n^{2t} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_{2t-1}=1}^{n} f(U_i, U_{k_1}) f(U_{k_1}, U_{k_2}) \cdots f(U_{k_{2t-1}}, U_j) = O_p\left(p_n^{2t} n^{2t+1}\right) \ . \tag{12}$$

Next, suppose $np_n \succ 1$. Then the dominant terms in the numerator and denominator are of order $O(p_n^T n^{T+1/2})$ and $O(p_n^{2T} n^{2T+1})$ respectively. As such,

$$\tilde{\beta}^{(T)} - \beta^{(T)} = O_p(p_n^{-T} n^{-T-1/2}) = o_p(1) \ .$$

Suppose instead that $np_n \approx 1$. Then, all terms in the numerator are of the same order. The same is true for the denominator. As before, $\tilde{\beta}^{(T)} - \beta^{(T)} = O_p(n^{-1/2}) = o_p(1)$.

Finally, suppose $np_n \prec 1$. In this regime, diffusion is equivalent to degree to a first order. The dominant terms in the numerator and denominator are of order $O(p_n n^{3/2})$ and $O(p_n^2 n^3)$ respectively. Then, as before, we obtain consistency if and only if $n^{3/2} p_n \to \infty$.

Lastly, consider eigenvector centrality. Given our assumptions, $v_1(A)$ is well-defined with high probability. Next note that by construction, $\sum_{i=1}^n \left(C_i^{(\infty)}\right)^2 = a_n^2$. Furthermore, by our assumptions,

$$E\left[\sum_{i=1}^{n} C_i^{(d)} \varepsilon_i \,\middle|\, U\right] = \sum_{i=1}^{n} C_i^{(d)} E\left[\varepsilon_i \,\middle|\, U\right] = 0$$
$$\mathrm{Var}\left[\sum_{i=1}^{n} C_i^{(d)} \varepsilon_i \,\middle|\, U\right] = \sum_{i=1}^{n} \left(C_i^{(d)}\right)^2 \mathrm{Var}\left[\varepsilon_i \,\middle|\, U\right] \leq a_n^2 \bar{\sigma}^2 \ .$$

As such,

$$\mathrm{Var}\left[\tilde{\beta}^{(\infty)} - \beta^{(\infty)}\right] \leq \frac{\bar{\sigma}^2}{a_n^2} \to 0 \text{ if } a_n \to \infty \ .$$

Thus, $a_n \to \infty$ implies that $\tilde{\beta}^{(\infty)} \xrightarrow{L_2} \beta^{(\infty)}$.

Necessity follows from the counterexample in our main text, reproduced here for completeness. Suppose $f = p_n \cdot 1$ so that $A = p_n \iota_n \iota_n'$. Then $C^{(\infty)}(A) = a_n \iota_n / \sqrt{n}$. Hence,

$$\tilde{\beta}^{(\infty)} = \frac{\sqrt{n}}{a_n} \cdot \frac{Y' \iota_n}{\iota_n' \iota_n} = \beta^{(\infty)} + \frac{1}{a_n \sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \ .$$

Under our assumptions, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \xrightarrow{d} N(0, \mathrm{Var}[\varepsilon_i])$. For $\tilde{\beta}^{(\infty)}$ to be consistent for $\beta^{(\infty)}$, it is therefore necessary for $a_n \to \infty$.

A.2 Proof of Theorem 2

We first write:

$$\hat{\beta}^{(d)} = \beta^{(d)} + \beta^{(d)} \frac{\left(\hat{C}^{(d)}\right)' \left(C^{(d)} - \hat{C}^{(d)}\right)}{\left(\hat{C}^{(d)}\right)' \hat{C}^{(d)}} + \frac{\left(\hat{C}^{(d)}\right)' \varepsilon}{\left(\hat{C}^{(d)}\right)' \hat{C}^{(d)}}$$

For convenience, denote

$$\hat{A}_{ij} = p_n f(U_i, U_j) + \xi_{ij} \quad , \quad E[\xi_{ij} \,|\, U_i, U_j] = 0 \ .$$

Also let $\xi_i = \sum_{j=1}^{n} \xi_{ij} = \sum_{j \neq i} \xi_{ij}$ and $\xi = (\xi_1, ..., \xi_n)'$. Finally, let $\boldsymbol{\xi}$ by the $n \times n$ matrix with $(i,j)^{\text{th}}$ entry ξ_{ij}. Note that $\xi = \boldsymbol{\xi} \iota_n$. By Assumption 2, $\xi_{ij} \perp\!\!\!\perp \varepsilon_k \,|\, U$ for all i, j, k.

A.2.1 Degree

We first show that $np_n \succ 1$ is sufficient for consistency of $\hat{\beta}^{(1)}$. Using our new notation, the numerator is:

$$\left(\hat{C}^{(1)}\right)' \varepsilon = C^{(1)} \varepsilon + \xi' \varepsilon \ .$$

By conditional independence of ξ and ε, $E\left[\xi'\varepsilon\right]=0$

$$
\begin{aligned}
\operatorname{Var}\left[\xi'\varepsilon \,|\, U\right] &= \operatorname{Var}\left[2\sum_{i=1}^{n}\sum_{j>i}\xi_{ij}\varepsilon_i \,\middle|\, U\right] = 4\sum_{i=1}^{n}\operatorname{Var}\left[\varepsilon_i\sum_{j>i}\xi_{ij}\,\middle|\, U\right] \\
&= 4\sum_{i=1}^{n}\left(E\left[\varepsilon_i^2 \,|\, U\right]\sum_{j>i}E\left[\xi_{ij}^2 \,|\, U\right]\right) \\
&\leq 2\bar{\sigma}^2\sum_{i=1}^{n}\sum_{j>i}p_n f(U_i,U_j)\left(1-p_n f(U_i,U_j)\right) \\
&\leq 2\bar{\sigma}^2\sum_{i=1}^{n}\sum_{j>i}p_n f(U_i,U_j)
\end{aligned}
$$

Taking expectations over U, we have that

$$
\operatorname{Var}\left[\xi'\varepsilon\right] \leq 2\bar{\sigma}^2 n^2 p_n \cdot W \quad\Rightarrow\quad \xi'\varepsilon = O_p(n{p_n}^{1/2})
$$

Given Equation (10), $C^{(1)}\varepsilon = O_p(n^{3/2}p_n)$ is thus dominant in the numerator if $np_n \succ 1$.

Next, consider the denominator, which has the form:

$$
\left(\hat{C}^{(1)}\right)'\hat{C}^{(1)} = \left(C^{(1)}\right)'C^{(1)} + 2\left(C^{(1)}\right)'\xi + \xi'\xi \ .
$$

We bound the last term in L_1-norm. Observe that it has conditional expectation:

$$
\begin{aligned}
E\left[\xi'\xi \,|\, U\right] &= E\left[\sum_{i=1}^{n}\left(\sum_{j\neq i}\xi_{ij}\right)^2 \,\middle|\, U\right] = \sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i}E\left[\xi_{ij}\xi_{ik}\,\middle|\, U\right] \\
&= \sum_{i=1}^{n}\sum_{j\neq i}E\left[\xi_{ij}^2 \,\middle|\, U\right] \leq \sum_{i=1}^{n}\sum_{j\neq i}p_n f(U_i,U_j) \ .
\end{aligned}
$$

Taking expectations over U,

$$
E\left[\xi'\xi\right] \leq n^2 p_n \cdot W \quad\Rightarrow\quad \xi'\xi = O_p(n^2 p_n) \ . \tag{13}
$$

Next, consider the middle term, which we will bound in L_2-norm. Write

$$E\left[\left(\left(C^{(1)}\right)'\xi\right)^2 \,\middle|\, U\right] = E\left[\sum_{i=1}^n\sum_{j=1}^n C_i^{(1)}\xi_i C_j^{(1)}\xi_j \,\middle|\, U\right]$$

$$= E\left[\sum_{i=1}^n C_i^{(1)}\xi_i C_i^{(1)}\xi_i \,\middle|\, U\right] + E\left[\sum_{i=1}^n\sum_{j\neq i}^n C_i^{(1)}\xi_i C_j^{(1)}\xi_j \,\middle|\, U\right] .$$

Note that

$$E\left[\sum_{i=1}^n\sum_{j\neq i}^n C_i^{(1)}\xi_i C_j^{(1)}\xi_j \,\middle|\, U\right] = \sum_{i=1}^n\sum_{j\neq i}^n C_i^{(1)}C_j^{(1)}E\left[\xi_i\xi_j \,|\, U\right] = \sum_{i=1}^n\sum_{j\neq i}^n C_i^{(1)}C_j^{(1)}E\left[\xi_{ij}^2 \,|\, U\right]$$

$$\leq \sum_{i=1}^n\sum_{j\neq i}^n C_i^{(1)}C_j^{(1)}p_n f(U_i,U_j)$$

$$= \sum_{i=1}^n\sum_{k=1}^n\sum_{l=1}^n\sum_{j\neq i}^n p_n^3 f(U_i,U_k)f(U_i,U_l)f(U_i,U_j)$$

The second equality above follows form the fact that when $i \neq j$, $E[\xi_{ik}\xi_{jl} \,|\, U] = 0$ unless $k = j$ and $l = i$. Furthermore,

$$E\left[\sum_{i=1}^n C_i^{(1)}\xi_i C_i^{(1)}\xi_i \,\middle|\, U\right] \leq \sum_{i=1}^n \left(C_i^{(1)}\right)^2 E\left[\xi_i^2 \,|\, U\right] = \sum_{i=1}^n \left(C_i^{(1)}\right)^2 E\left[\sum_{j\neq i}\xi_{ij}^2 \,\middle|\, U\right]$$

$$\leq \sum_{i=1}^n \left(C_i^{(1)}\right)^2 \sum_{j\neq i} p_n f(U_i,U_j)$$

$$\leq \sum_{i=1}^n\sum_{k=1}^n\sum_{l=1}^n p_n^2 f(U_i,U_k)f(U_i,U_l)\sum_{j\neq i} p_n f(U_i,U_j)$$

Taking expectation over two displays above,

$$E\left[\left(\left(C^{(1)}\right)'\xi\right)^2\right] = O(n^4p_n^3) \quad \Rightarrow \quad \left(C^{(1)}\right)'\xi = O_p(n^2p_n^{3/2}) \;. \tag{14}$$

By Equation (11), $\left(C^{(1)}\right)'C^{(1)} = O_p(n^3p_n^2)$. Putting the rates we derived together, the denominator is

$$\left(\hat{C}^{(1)}\right)'\hat{C}^{(1)} = O_p(n^3p_n^2) + O_p(n^2p_n^{3/2}) + O_p(n^2p_n) \;. \tag{15}$$

Hence, $np_n \succ 1$ implies that

$$\frac{\left(\hat{C}^{(1)}\right)' \varepsilon}{\left(\hat{C}^{(1)}\right)' \hat{C}^{(1)}} \xrightarrow{p} 0 \ .$$

It remains to note that

$$\left(\hat{C}^{(1)}\right)' \left(C^{(1)} - \hat{C}^{(1)}\right) = \left(C^{(1)}\right)' \xi + \xi' \xi$$

so that by the rates in (13), (14) and (15),

$$\beta^{(d)} \frac{\left(\hat{C}^{(d)}\right)' \left(C^{(d)} - \hat{C}^{(d)}\right)}{\left(\hat{C}^{(d)}\right)' \hat{C}^{(d)}} = O_p(n^{-1}p_n^{-1}) \xrightarrow{p} 0 \ .$$

We can loosely write the above results as

$$\hat{\beta}^{(1)} - \beta^{(1)} \approx \frac{n^2 p_n^{3/2} + n^2 p_n}{n^3 p_n^2 + n^2 p_n^{3/2} + n^2 p_n} + \frac{n^{3/2} p_n + n p_n^{1/2}}{n^3 p_n^2 + n^2 p_n^{3/2} + n^2 p_n}$$

As such, $\hat{\beta}^{(1)}$ is consistent for $\beta^{(1)}$ if $np_n \succ 1$.

Suppose instead that $n^{-2} \prec p_n \prec n^{-1}$. By our rate calculations, we can write

$$\hat{\beta}^{(1)} - \beta^{(1)} = -\beta^{(1)} \cdot \frac{\xi'\xi + o_p(n^2 p_n)}{\xi'\xi + o_p(n^2 p_n)} + o_p(1)$$

In other words, $\hat{\beta}^{(1)} \xrightarrow{p} 0$. Finally, if $p_n \prec n^{-2}$, we $\hat{\beta}^{(1)} - \beta^{(1)}$,

$$\hat{\beta}^{(1)} = \frac{\xi'\varepsilon}{\xi'\xi} + o_p(1) = O_p(n^{-2}p_n^{-1})$$

diverges in probability.

A.2.2 Diffusion Centrality

Diffusion centrality is comprised of terms of the form:

$$\hat{A}^t = (A + \boldsymbol{\xi})^t = \sum_{B \in \tilde{\mathbf{B}}} B$$

Here, $\tilde{\mathbf{B}} = \{A, \boldsymbol{\xi}\}^t$. B is a mixed product of A and $\boldsymbol{\xi}$, and will be the central object of our analysis. For convenience, define:

Definition 10 (Mixed Product of Order t)**.** A mixed product of order t is a term of the form $B = \prod_{j=1}^{t} B_j$ where $B_j \in \{A, \boldsymbol{\xi}\}$. Suppose $B_j = \boldsymbol{\xi}$ for $\tau \geq 0$ number of j's. We will also say that the order of $\boldsymbol{\xi}$ in B is τ. Define $\mathcal{J} = \left\{ j \in [2t+1] \,|\, b_{k_j,k_{j+1}} = \xi_{k_j,k_{j+1}} \right\}$. Then, $\mathcal{J}$ indicate the locations of the $\boldsymbol{\xi}$ in the mixed product B. Let $p = (p_1, ..., p_r)'$ record lengths of the contiguous blocks in $\mathcal{J}$. If $p_1, ..., p_r$ are all even, we say that B is even.

The dependence of $\mathcal{J}$ and p on B is suppressed for convenience.

Example 5. In the above notation,

$$B = A^2 \boldsymbol{\xi}^3 A \boldsymbol{\xi}^2 A \quad \Rightarrow \quad \mathcal{J} = \{(3,4,5)\,,\,(7,8)\,,\,(13,14,15)\,,\,(17,18)\} \quad , \quad p = (3,2,3,2)\ .$$

First note that degree centrality is diffusion centrality with $T = 1$. Since $\hat{\beta}^{(1)}$ is inconsistent for $\beta^{(1)}$ when $np_n \prec 1$, consistency of diffusion centrality also requires that $np_n \succ 1$. We show that $\hat{\beta}^{(T)} \xrightarrow{p} \beta^{(T)}$ when $np_n \succ 1$. Write

$$\hat{C}^{(T)} = \left(\sum_{t=1}^{T} \delta^t \hat{A}^t \right) \iota_n = \left(\sum_{t=1}^{T} \delta^t \left(A + \boldsymbol{\xi} \right)^t \right) \iota_n$$

Expanding the products, we can write $\left(\hat{C}^{(T)}\right)' \hat{C}^{(T)}$ and $\left(\hat{C}^{(T)}\right)' \left(\hat{C}^{(T)} - C^{(T)}\right)$ as sums involving mixed products of A and $\boldsymbol{\xi}$.

We seek to bound $\iota_n' B \iota_n$ in L_2-norm. First note that if $B_j = A$ for all $j \in [t]$, then by Equation (12), $B = O_p(n^{t+1} p_n^t)$. Suppose that $B_j = \xi$ for at least one j. Then,

Lemma 2. Suppose B is a order t mixed product of A and $\boldsymbol{\xi}$. Suppose that the order of $\boldsymbol{\xi}$ in B is $\tau \geq 1$. Then, there exists $\alpha, \beta \in \mathbb{N}$, $\alpha \geq \beta$ such that

$$\iota_n' B \iota_n = O_p\left(n^{t+1-\alpha/2} p_n^{t-\beta/2}\right) \preccurlyeq O_p\left(n^{t+1-\tau/2} p_n^{t-\tau/2}\right)\ .$$

In particular,

$$\iota_n' A^t \iota_n = O_p\left(n^{2t+2} p_n^{2t}\right) \succ n^{2t+2-\alpha} p_n^{2t-\beta}\ .$$

Furthermore, suppose B is not even. Then,

$$\iota_n' B \iota_n = O_p\left(\frac{1}{\sqrt{n}}\right) \cdot O_p\left(n^{t+1-\tau/2} p_n^{t-\tau/2}\right)\ .$$

If B is even, then

$$\iota_n' B \iota_n - E\left[\iota_n' B \iota_n \,|\, U\right] = O_p\left(\frac{1}{\sqrt{n}}\right) \cdot O_p\left(n^{t+1-\tau/2} p_n^{t-\tau/2}\right)\ .$$

Taking expectations over U, we therefore have that

$$\iota_n' A^t \iota_n = O_p(n^{t+1} p_n^t) \succ \iota_n B \iota_n = O_p(n^{t+1-\alpha/2} p_n^{t-\beta/2})$$

under the assumption that $np_n \succ 1$, as long as B contains at least one $\boldsymbol{\xi}$. Now, we return to the nuisance term:

$$\beta^{(T)} \frac{\left(\hat{C}^{(T)}\right)' \left(C^{(T)} - \hat{C}^{(T)}\right)}{\left(\hat{C}^{(T)}\right)' \hat{C}^{(T)}} .$$

By our analysis, the dominant term in the denominator is $\iota_n A^{2T} \iota_n = O_p(n^{2T+2} p_n^{2T})$. Every term in the numerator has strictly smaller order. Hence, we conclude that the nuisance term is $o_p(1)$. It remains to show that

$$\frac{\left(\hat{C}^{(T)}\right)' \varepsilon}{\left(\hat{C}^{(T)}\right)' \hat{C}^{(T)}} \approx \frac{\iota_n A^T \varepsilon}{\iota_n A^{2T} \iota_n} \xrightarrow{p} 0 .$$

Note that the numerator is a U-statistic of order $T+1$. It also has mean 0 by our conditional mean independence assumption. Hence, by the U-statistic LLN, the numerator is of order $o_p(n^{T+1} p_n^T)$, which is again strictly smaller than that of the denominator. Conclude that $\hat{\beta}^{(T)} \xrightarrow{p} \beta^{(T)}$ if $np_n \succ 1$.

A.2.3 Eigenvector Centrality

Inconsistency

We first provide a counterexample under the assumption that p_n satisfies Equation (5). Let $f = 1$, $\beta = 1$ and suppose $\varepsilon_i \perp\!\!\!\perp U_i$ (By Assumption 2, $\varepsilon_i \perp\!\!\!\perp \xi_{jk}$ for all $i, j, k \in [n]$). Theorem 1.7, Remark 1.4 and Remark 1.8 in Alt et al. (2021b) provides the following description of the $v_1(\hat{A})$. Let $i \in [n]$ be a vertex, $B_r(i)$ be the set of vertices which are in the r-neighbourhood of i. Let $S_r(i) = B_r(i) \backslash B_{r-1}(i)$ be the sphere of radius r around i. Let $\mathfrak{u} = \frac{\log n}{np_n \log\log n}$. Then, w.p.a. 1, there exists $\tilde{v}$ such that for any $\eta > 0$,

$$\|\tilde{v} - v_1(\hat{A})\| \leq \frac{1}{\mathfrak{u} \cdot np_n} + \frac{(np_n)^{-1/2+3\eta}}{\sqrt{\mathfrak{u}}} + \frac{1}{np_n} . \tag{16}$$

Furthermore, $\tilde{v}$ has the following structure:

$$\tilde{v} = \sum_{r=0}^{R} u_r s_r(i) \quad , \quad s_r(i) = \frac{\mathbf{1}_{S_r(i)}}{\|\mathbf{1}_{S_r(i)}\|}$$

where $R \prec \frac{np_n}{\log\log n}$ and

$$u_1 = \frac{1}{\sqrt{np_n}} u_0 \quad , \quad u_r \leq \left(\frac{2}{\sqrt{\mathfrak{u}}}\right)^{r-1} u_1$$

and u_0 is defined by the normalization $\|\tilde{v}\| = 1$. The result of Alt et al. (2021b) says that $v_1(\hat{A})$ is well approximated by an eigenvector that is exponentially localized around some vertex i. This vertex is in fact the one with the highest realized degree. Let us calculate a lower bound on u_0.

$$1 = \tilde{v}'\tilde{v} = \sum_{r=1}^{R} u_r^2 \leq u_0^2 + \frac{1}{np_n} u_0^2 \left(\sum_{r=1}^{\infty} \left(\frac{4}{\mathfrak{u}}\right)^{r-1}\right) .$$

The above inequality comes from using upper bounds for u_r and replacing R with ∞. Collecting the u_0's, we find that w.p.a. 1,

$$u_0^2 \geq \frac{1}{1 + \frac{1}{np_n}\frac{1}{1-4/\mathfrak{u}}} .$$

Since $np_n, \mathfrak{u} \to \infty$ when p_n satisfies Equation (5), we have that for n large enough, $u_0 \geq \frac{1}{\sqrt{2}}$ w.p.a. 1. Now, write

$$\begin{aligned}
\hat{\beta}^{(\infty)} &= \frac{a_n \left(v_1(\hat{A})\right)' Y}{a_n^2 \left(v_1(\hat{A})\right)' v_1(\hat{A})} = \frac{1}{a_n} \left(v_1(\hat{A})\right)' Y \\
&= \frac{1}{a_n} \left(v_1(\hat{A})\right)' \left(a_n \frac{\iota_n}{\sqrt{n}} + \varepsilon\right) \\
&= \frac{\left(v_1(\hat{A})\right)' \iota_n}{\sqrt{n}} + \frac{\left(v_1(\hat{A})\right)' \varepsilon}{a_n}
\end{aligned}$$

By independence of ε and $(\boldsymbol{\xi}, U)$, we have that $\mathrm{Var}\left[v_1(\hat{A})'\varepsilon \,\middle|\, U\right] = \|v_1(\hat{A})\|\sigma^2 = \sigma^2$. Hence, σ^2/a_n is a lower bound for the variance of $\hat{\beta}^{(\infty)}$. Hence, $a_n \to \infty$ is necessary for consistency.

Suppose $a_n \to \infty$. We have consistency if and only if

$$\frac{\left(v_1(\hat{A})\right)' \iota_n}{\sqrt{n}} \xrightarrow{p} 1 ,$$

in which case

$$\hat{\beta}^{(\infty)} = \frac{\left(v_1(\hat{A})\right)' \iota_n}{\sqrt{n}} + o_p(1) = \frac{\tilde{v}'\iota_n}{\sqrt{n}} + o_p(1)$$

Notice that the optimization problem:

$$\max_{v \in \mathbb{R}^n} v' \iota_n \quad \text{such that} \quad \|v\| = 1$$

has solution $v = \iota_n / \sqrt{n}$ and optimal value $\sqrt{n}$. We can also consider the constrained optimization problem:

$$\max_{v \in \mathbb{R}^n} v' \iota_n \quad \text{such that} \quad \|v\| = 1 \text{ and } v_1 \geq \frac{1}{\sqrt{2}} \ .$$

This problem has solution $v_1 = \frac{1}{\sqrt{2}}$ and $v_{-1} = \iota_{n-1}/\sqrt{2n}$ and optimal value

$$\gamma^* := \frac{1}{\sqrt{2}} + \frac{n-1}{\sqrt{2n}}$$

The constrained maximization problem corresponds to the best case allocation of $\left(v_1(\hat{A})\right)_{-i}$ that makes $\left(v_1(\hat{A})\right)' \iota$ as close to $\sqrt{n}$ as possible, subject to the requirement that $\left(v_1(\hat{A})\right)_i \geq 1/\sqrt{2}$. As such, w.p.a. 1, we have that

$$\hat{\beta}^{(\infty)} \leq \frac{\gamma^*}{\sqrt{n}} = \frac{1}{\sqrt{2n}} + \frac{n-1}{n\sqrt{2}} \to \frac{1}{\sqrt{2}} \ .$$

Hence, $\hat{\beta}^{(\infty)}$ is bounded away from $\beta^{(\infty)} = 1$ in probability. Conclude that the estimator is inconsistent.

Consistency

We next show that $\hat{\beta}^{(\infty)} \xrightarrow{p} \beta^{(\infty)}$ when $np_n \succ \sqrt{\frac{\log n}{\log \log n}}$. Write

$$\hat{\beta}^{(\infty)} = \beta^{(\infty)} + \beta^{(\infty)} \frac{\left(\hat{C}^{(\infty)}\right)' \left(C^{(\infty)} - \hat{C}^{(\infty)}\right)}{\left(\hat{C}^{(\infty)}\right)' \hat{C}^{(\infty)}} + \frac{\left(\hat{C}^{(\infty)}\right)' \varepsilon}{\left(\hat{C}^{(\infty)}\right)' \hat{C}^{(\infty)}} \tag{17}$$

$$= \beta^{(\infty)} + \beta^{(\infty)} \left(v_1(\hat{A})\right)' \left(v_1(A) - v_1(\hat{A})\right) + \frac{1}{a_n} v_1(\hat{A})' \varepsilon \tag{18}$$

since $\left(\hat{C}^{(\infty)}\right)'\hat{C}^{(\infty)} = a_n^2$ by construction. Therefore, by Lemma 1 and the Davis-Kahan inequality (e.g. Theorem 4.5.5 in Vershynin 2018), for any $\nu \in (0,1)$,

$$\left\|v_1(A) - v_1(\hat{A})\right\| \leq \frac{\|\hat{A} - A\|}{np_n\left(\lambda_1 - \lambda_2\right)} = O_p\left(\left(\sqrt{\frac{\log n}{\log\log n}}\Big/ np_n\right)^{(1-\nu)/2}\right) = o_p(1)$$

where first equality follows from Equation (9) and the second from our assumption on np_n.

Finally, note that

$$E\left[v_1(\hat{A})'\varepsilon\,\middle|\,U\right] \leq \|v_1(\hat{A})\|\bar{\sigma}^2 = \bar{\sigma}^2.$$

Since $a_n \to \infty$, conclude that $\frac{1}{a_n}v_1(\hat{A})'\varepsilon \xrightarrow{p} 0$ and that $\hat{\beta}^{(\infty)} \xrightarrow{p} \beta^{(\infty)}$.

A.3 Proof of Theorem 4

A.3.1 Case (a)

Although case (b) specializes to (a), we will prove (a) separately because

1. The proof for our plug-in estimator for case (a) is also the base case for an induction argument in the proof of case (b)
2. Case (c), by Lemma 3 equivalent to case (a) to a first order.

To prove (a) first recall our analysis in the proof of Theorem 2, which yields:

$$\begin{aligned}\hat{\beta}^{(1)} &= \beta^{(1)} + \beta^{(1)}\frac{\left(\hat{C}^{(1)}\right)'\left(C^{(1)} - \hat{C}^{(1)}\right)}{\left(\hat{C}^{(1)}\right)'\hat{C}^{(1)}} + \frac{\left(\hat{C}^{(1)}\right)'\varepsilon}{\left(\hat{C}^{(1)}\right)'\hat{C}^{(1)}} \\ &= \beta^{(1)} + \beta^{(1)}\frac{\iota_n' A\boldsymbol{\xi}\iota_n + \iota_n'\boldsymbol{\xi}^2\iota_n}{\iota_n A^2\iota_n + o_p\left(\iota_n A^2\iota_n'\right)} + \frac{O_p\left(n^{3/2}p_n\right)}{\iota_n A^2\iota_n + o_p\left(\iota_n' A^2\iota_n\right)}\ . \end{aligned} \tag{19}$$

Recall also that,

$$E\left[\iota_n'\boldsymbol{\xi}\iota_n\,|\,U\right] = \sum_{i=1}^n\sum_{j\neq i} E\left[\xi_{ij}^2\,|\,U\right] = \sum_{i=1}^n\sum_{j\neq i} p_n f(U_i,U_j)\left(1 - p_n f(U_i,U_j)\right)$$

so that the unconditional expectation is

$$E\left[\iota_n'\boldsymbol{\xi}\iota_n\right] = \Omega\left(n^2 p_n\right)\ .$$

To obtain our desired result, we will show that $\iota_n' A\boldsymbol{\xi}\iota_n$ converges to a normal distribution asymptotically once suitable scaled, and that it dominates $\left(\iota_n'\boldsymbol{\xi}^2\iota_n - E\left[\iota_n'\boldsymbol{\xi}\iota_n \,|\, U\right]\right)$. We then show that the population quantities in the CLT can be estimated at a sufficiently fast rate.

First observe that by Assumption 2, $E\left[\iota_n' A\boldsymbol{\xi}\iota_n \,|\, U\right] = 0$. Next, define

$$\begin{aligned} V_*^{(1)}(U) &:= E\left[\left(\iota_n' A\boldsymbol{\xi}\iota_n\right)^2 \,\Big|\, U\right] \\ &= \sum_{j<k} p_n f(U_j, U_k)\left(1 - p_n f(U_j, U_k)\right)\left(\sum_{i\neq j} p_n f(U_i, U_j) + \sum_{i\neq k} p_n f(U_i, U_k)\right)^2 . \end{aligned}$$

Then, by the U-statistics LLN,

$$\frac{1}{n^4 p_n^3} V_*^{(1)}(U) \xrightarrow{p} \int f(U_1,U_2) f(U_1,U_3) f(U_1,U_4)\, dU + \frac{1}{2}\int f(U_1,U_2) f(U_2,U_3) f(U_3,U_4)\, dU > 0 .$$

Next, define the event $\Upsilon^{(1)} := \left\{V^{(1)}(U) > kn^4 p_n^3\right\}$, where k is chosen to be $1/2$ the magnitude of the limit above. We will apply the Berry-Esseen inequality conditional on $U \in \Upsilon^{(1)}$. Note that by Assumption 2, ξ_{ij}'s continue to be independent after conditioning. By Theorem 3.7 in Chen et al. (2011),

$$\sup_{z\in\mathbf{R}} \left| P\left(\frac{\iota_n' A\boldsymbol{\xi}\iota_n}{\sqrt{V_*^{(1)}(U)}} \le z \,\middle|\, U\right) - \Phi(z)\right| \le 10\gamma .$$

Next, we evaluate the third moments of the summands:

$$\begin{aligned} E\left[\left|\xi_{jk}\sum_{i\neq j} p_n f(U_i,U_j)\right|^3 \,\middle|\, U\right] &= \left|\sum_{i\neq j} p_n f(U_i,U_j)\right|^3 E\left[\left|\xi_{jk}\right|^3 \,\big|\, U\right] \\ &\le n^3 p_n^3 \cdot p_n \end{aligned}$$

As such, on $\Upsilon^{(1)}$,

$$\gamma \le \sum_{j<k} \frac{n^3 p_n^4}{\left(kn^4 p_n^3\right)^{3/2}} \approx \frac{n^5 p_n^4}{n^6 p_n^{9/2}} = \frac{1}{np^{1/2}} \to 0$$

where the above bound is independent of U. Furthermore, $P(\Upsilon^{(1)}) \to 1$. Conclude that

$$\frac{\iota_n' A\boldsymbol{\xi}\iota_n}{\sqrt{V_*^{(1)}(U)}} \xrightarrow{d} \mathrm{N}\left(0,1\right) .$$

It remains to show that $\iota_n' \boldsymbol{\xi}^2 \iota_n - E\left[\iota_n' \boldsymbol{\xi}^2 \iota_n \,|\, U\right] = o_p\left(n^2 p_n^{3/2}\right)$. Write

$$\Gamma^{(1)} := \sum_{i_1,\ldots,i_6} E\left[\left(\xi_{i_1,i_2}\xi_{i_2,i_3} - E\left[\xi_{i_1,i_2}\xi_{i_2,i_3} \,|\, U\right]\right)\left(\xi_{i_4,i_5}\xi_{i_5,i_6} - E\left[\xi_{i_4,i_5}\xi_{i_5,i_6} \,|\, U\right]\right) \,\middle|\, U\right] \,.$$

Note that

$$E\left[\xi_{i_1 i_2}\xi_{i_2 i_3}\xi_{i_4 i_5}\xi_{i_5 i_6} \,\middle|\, U\right] = 0 \Rightarrow E\left[\left(\xi_{i_1,i_2}\xi_{i_2,i_3} - E\left[\xi_{i_1,i_2}\xi_{i_2,i_3} \,|\, U\right]\right)\left(\xi_{i_4,i_5}\xi_{i_5,i_6} - E\left[\xi_{i_4,i_5}\xi_{i_5,i_6} \,|\, U\right]\right) \,\middle|\, U\right] = 0\,.$$

This is because for the former to hold, we must have an edge (i_k, i_{k+1}) that is of multiplicity 1, which is sufficient for making the latter conditional expectation 0. Figure 11 shows all possible configurations of indices that will lead to $E\left[\xi_{i_1 i_2}\xi_{i_2 i_3}\xi_{i_4 i_5}\xi_{i_5 i_6} \,\middle|\, U\right] \neq 0$. Table 6 records the frequency of their appearance.

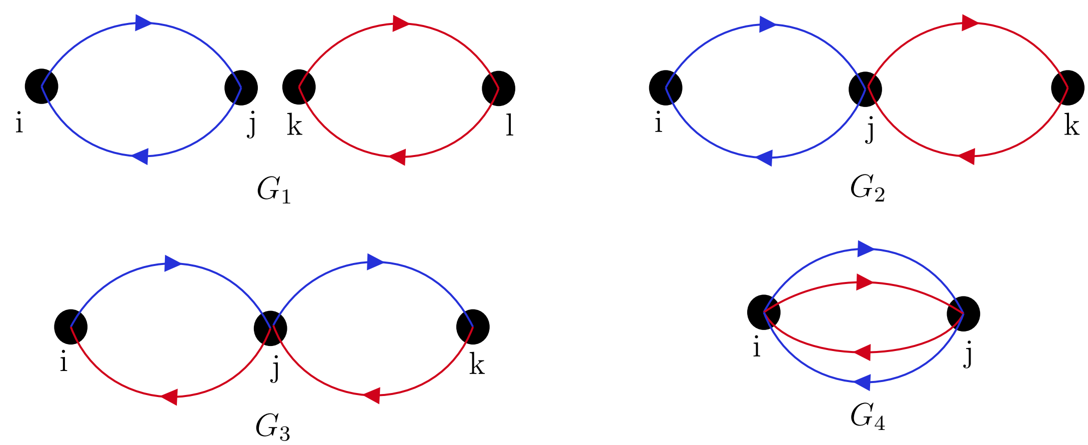

Figure 11: The possible configurations of indices that will lead to $E\left[\xi_{ij}\xi_{jk}\xi_{i'j'}\xi_{j'k'} \,\middle|\, U\right]$ being non-zero. These are the only graphs that can be formed using 2 walks of length 2 and in which each edge has multiplicity at least 2.

Graph	Number of Instances	$E\left[\xi_{ij}\xi_{jk}\xi_{i'j'}\xi_{j'k'} \,\middle|\, U\right]$
G_1	$n(n-1)(n-2)(n-3)$	$p_n^2 f(U_i,U_j)f(U_k,U_l) + O_p(p_n^3)$
G_2	$n(n-1)(n-2)$	$p_n^2 f(U_i,U_j)f(U_k,U_l) + O_p(p_n^3)$
G_3	$n(n-1)(n-2)$	$p_n^2 f(U_i,U_j)f(U_k,U_l) + O_p(p_n^3)$
G_4	$n(n-1)$	$p_n f(U_i,U_j) + O_p(p_n^2)$

Table 6: The number of instances of each graph, as well as the value of their conditional expectations, up to the leading term.

Observe that

$$\begin{aligned} &E\left[\left(\xi_{i_1,i_2}\xi_{i_2,i_3} - E\left[\xi_{i_1,i_2}\xi_{i_2,i_3}\,|\,U\right]\right)\left(\xi_{i_4,i_5}\xi_{i_5,i_6} - E\left[\xi_{i_4,i_5}\xi_{i_5,i_6}\,|\,U\right]\right)\,\middle|\,U\right] \\ &\qquad\neq E\left[\xi_{i_1,i_2}\xi_{i_2,i_3} - E\left[\xi_{i_1,i_2}\xi_{i_2,i_3}\,|\,U\right]\,\middle|\,U\right] E\left[\xi_{i_4,i_5}\xi_{i_5,i_6} - E\left[\xi_{i_4,i_5}\xi_{i_5,i_6}\,|\,U\right]\,\middle|\,U\right] \end{aligned}$$

only if there is an edge that is common to both of the above multiplicands. In particular, G_1 and G_2 will not contribute to $\Gamma^{(1)}$. As such, by Table 6, $\Gamma^{(1)} = O_p\left(n^3 p_n^2\right)$. Conclude that

$$\iota_n' \boldsymbol{\xi}^2 \iota_n - E\left[\iota_n' \boldsymbol{\xi}^2 \iota_n\,|\,U\right] = O_p\left(n^{3/2} p_n\right) = o_p\left(n^2 p_n^{3/2}\right) \ .$$

Using the above results, we can rewrite Equation (19) as

$$\hat{\beta}^{(1)} = \beta^{(1)} + \beta^{(1)} \frac{\iota_n' A \boldsymbol{\xi} \iota_n - E\left[\iota_n' \boldsymbol{\xi}^2 \iota_n\,|\,U\right] + O_p\left(n^{3/2} p_n\right)}{\left(\hat{C}^{(1)}\right)\hat{C}^{(1)}} \ .$$

Consequently,

$$\frac{\hat{\beta}^{(1)} - \beta^{(1)}\left(1 - B^{(1)}\right)}{\beta^{(1)}\sqrt{V^{(1)}}} = \frac{\iota_n' A \boldsymbol{\xi} \iota_n}{\sqrt{V_*^{(1)}}} + \frac{O_p\left(n^{3/2} p_n\right)}{\Omega_p\left(n^2 p_n^{3/2}\right)} \xrightarrow{d} \mathrm{N}(0,1) \ .$$

where

$$\begin{aligned} B^{(1)} &= \left(\left(\hat{C}^{(1)}\right)\hat{C}^{(1)}\right)^{-1} E\left[\iota_n' \boldsymbol{\xi}^2 \iota_n\,|\,U\right] \ , \\ V^{(1)} &= \left(\left(\hat{C}^{(1)}\right)\hat{C}^{(1)}\right)^{-2} V_*^{(1)} \ . \end{aligned}$$

Plug-in Estimation

Finally, we show that $\hat{B}^{(1)}$ and $\hat{V}^{(1)}$ estimate $B^{(1)}$ and $V^{(1)}$ at appropriate rates. Define $\hat{V}_*^{(1)} = \left(\left(\hat{C}^{(1)}\right)\hat{C}^{(1)}\right)^{-2}\hat{V}^{(1)}$. We will show that

$$\frac{\hat{B}^{(1)} - B^{(1)}}{\sqrt{V^{(1)}}} \xrightarrow{p} 0 \quad , \quad \frac{\hat{V}^{(1)}}{V^{(1)}} = \frac{\hat{V}_*^{(1)}}{V_*^{(1)}} \xrightarrow{p} 1 \ .$$

The first statement above is straightforward:

$$\frac{\hat{B}^{(1)} - B^{(1)}}{\sqrt{V^{(1)}}} = \frac{1}{\sqrt{V_*^{(1)}}} \sum_{i \neq j} p_n f(U_i, U_j) + \xi_{ij} - p_n f(U_i, U_j) \left(1 - p_n f(U_i, U_j)\right)$$

$$= O_p\left(\frac{1}{n^2 p_n^{3/2}}\right) \cdot \Bigg(\underbrace{\sum_{i \neq j} \xi_{ij}}_{\substack{=O_p(np_n^{1/2}) \\ \text{by } ((13))}} + \underbrace{\sum_{i \neq j} p_n^2 f^2(U_i, U_j)}_{=O_p(n^2 p_n^2)} \Bigg) = O_p\left(\frac{1}{np_n} + p_n^{1/2}\right) = o_p(1) \ .$$

Next, consider:

$$\begin{aligned}
&2\left(\hat{V}_*^{(1)} - V_*^{(1)}\right) \\
&= \sum_{j \neq k} \hat{A}_{jk} \left(\hat{C}_j^{(1)} + \hat{C}_k^{(1)}\right)^2 - p_n f(U_j, U_k) \left(1 - p_n f(U_j, U_k)\right) \left(\sum_{i \neq j} p_n f(U_i, U_j) + \sum_{i \neq k} p_n f(U_i, U_k)\right)^2 \\
&= \sum_{j \neq k} \hat{A}_{jk} \left(\hat{C}_j^{(1)} + \hat{C}_k^{(1)}\right)^2 - p_n f(U_j, U_k) \left(\sum_{i \neq j} p_n f(U_i, U_j) + \sum_{i \neq k} p_n f(U_i, U_k)\right)^2 + O_p\left(n^4 p_n^4\right) \\
&= \underbrace{2 \sum_{j \neq k} A_{jk} \left(C_j^{(1)} + C_k^{(1)}\right) (\xi_j + \xi_k)}_{=:\Gamma_1^{(1)}} + \underbrace{\sum_{j \neq k} A_{jk} \left(\xi_j + \xi_k\right)^2}_{=:\Gamma_2^{(1)}} \\
&\quad + \underbrace{\sum_{j \neq k} \xi_{jk} \left(C_j^{(1)} + C_k^{(1)}\right)^2}_{=:\Gamma_3^{(1)}} + \underbrace{2 \sum_{j \neq k} \xi_{jk} \left(C_j^{(1)} + C_k^{(1)}\right) (\xi_j + \xi_k)}_{=:\Gamma_4^{(1)}} + \underbrace{\sum_{j \neq k} \xi_{jk} \left(\xi_j + \xi_k\right)^2}_{=:\Gamma_5^{(1)}}
\end{aligned}$$

Recall that $V_*^1 = \Omega_p\left(n^4 p_n^3\right)$. We will show that $\Gamma_a^{(1)} = o_p\left(n^4 p_n^3\right)$ for $a \in [5]$.

$$\begin{aligned}
\Gamma_1^{(1)} &= \sum_{j,k} A_{jk} \left(C_j^{(1)} + C_k^{(1)}\right) \sum_{i \neq j} \xi_{ij} + \sum_{j,k} A_{jk} \left(C_j^{(1)} + C_k^{(1)}\right) \sum_{i \neq k} \xi_{ik} \\
&= 2 \sum_{i,j} \xi_{ij} \sum_{k} A_{jk} \left(C_j^{(1)} + C_k^{(1)}\right) \qquad \text{by symmetry}
\end{aligned}$$

Taking conditional expectations,

$$\begin{aligned}
E\left[\Gamma_1^{(1)}\,|\,U\right] &= E\left[\left(4\sum_{i<j}\xi_{ij}\sum_k A_{jk}\left(C_j^{(1)}+C_k^{(1)}\right)\right)^2 \,\middle|\, U\right] \\
&= 16\sum_{i<j}E[\xi_{ij}^2\,|\,U]\left(\sum_k A_{jk}\left(C_j^{(1)}+C_k^{(1)}\right)\right)^2 \\
&\leq 16n^2p_n\cdot(2np_n)^4 \quad \text{ since } C_j^{(1)}\leq np_n \text{ for all } j\in[n] \\
&= O_p\left(n^6p_n^5\right)
\end{aligned}$$

Hence, $\Gamma_1^{(1)} = O_p\left(n^3p_n^{5/3}\right)$.

Next,

$$\Gamma_2^{(1)} = 2\sum_{j,k}A_{jk}\left(\sum_{i\neq j}\xi_{ij}\right)^2 + \sum_{j,k}A_{jk}\left(\sum_{i\neq j}\xi_{ij}\right)\left(\sum_{i\neq k}\xi_{ik}\right) .$$

First note that

$$\sum_{j,k}A_{jk}\left(\sum_{i\neq j}\xi_{ij}\right)\left(\sum_{i\neq k}\xi_{ik}\right) = \iota_n'\boldsymbol{\xi}A\boldsymbol{\xi}\iota_n = O_p\left(n^{7/2}p_n^{5/2}\right) \qquad \text{by Lemma 2.}$$

Secondly, we have that

$$\begin{aligned}
E\left[\left(\sum_{j,k}A_{jk}\left(\sum_{i\neq j}\xi_{ij}\right)^2\right)^2 \,\middle|\, U\right] &= E\left[\left(\sum_{i,j,l}\xi_{ij}\xi_{jl}\sum_k A_{jk}\right)^2 \,\middle|\, U\right] \\
&\leq n^2p_n^2E\left[\left(\iota_n'\boldsymbol{\xi}^2\iota_n\right)^2 \,\middle|\, U\right] \leq n^2p_n^2\cdot n^4p_n^2
\end{aligned}$$

Noting that the bound above does not depend on U, we have

$$\Gamma_2^{(1)} = O_p\left(n^3p_n^2\right) + O_p\left(n^{7/2}p_n^{5/2}\right) .$$

Now,

$$E\left[\left(\Gamma_3^{(1)}\right)^2\,|\,U\right] = \sum_{j,k}E\left[\xi_{jk}^2\,|\,U\right]\left(C_j^{(1)}+C_k^{(1)}\right)^4 \leq n^2p_n\cdot(2np_n)^4$$

As such, $\Gamma_3^{(1)} = O_p\left(n^3p_n^{5/2}\right)$.

By a similar argument to above, we also have that

$$\Gamma_4^{(1)} = 2\sum_{j,k,l} \xi_{jk}\xi_{jl}\left(C_j^{(1)} + C_k^{(1)}\right) = O_p\left(np_n\right) \cdot O_p\left(\iota_n \boldsymbol{\xi}^2 \iota_n\right) = O_p\left(n^3 p_n^2\right) \ .$$

Finally,

$$\Gamma_5^{(1)} = 2\sum_{j,k} \xi_{jk}\left(\sum_{i\neq j} \xi_{ij}\right)^2 + \sum_{j,k} \xi_{jk} \sum_{i\neq j} \xi_{ij} \sum_{l\neq k} \xi_{kl}$$

First observe that

$$\sum_{j,k} \xi_{jk} \sum_{i\neq j} \xi_{ij} \sum_{l\neq k} \xi_{kl} = \iota_n' \boldsymbol{\xi}^3 \iota_n = O_p\left(n^2 p_n^{3/2}\right) \qquad \text{by Lemma 2.}$$

Now,

$$E\left[\left(\sum_{j,k} \xi_{jk}\left(\sum_{i\neq j} \xi_{ij}\right)^2\right)^2 | \, U\right] = \sum_{i_1,\ldots i_8} E\left[\xi_{i_1 i_2}\xi_{i_1 i_3}\xi_{i_1 i_4} \cdot \xi_{i_5 i_6}\xi_{i_5 i_7}\xi_{i_5 i_8} \,|\, U\right]$$

Relative to Lemma 2, here we are counting the contributions made by two three-pointed stars. The graphs that contribute the above expectation are displayed in Figure 12. Their frequencies and magnitudes are recorded in Table 7. Summing up the contribution of each graph, we have that the above display is $O_p\left(n^4 p_n^2\right)$. Hence, $\Gamma_5^{(1)} = O_p\left(n^2 p_n\right) + O_p\left(n^2 p_n^{3/2}\right) = O_p\left(n^2 p_n\right)$.

Putting all our results together, we have that

$$\frac{\hat{V}_*^1 - V_*^{(1)}}{V_*^{(1)}} = o_p(1) \ ,$$

which together with our central limit therom and result on $\hat{B}^{(1)}$ implies our desired result:

$$\hat{S}^{(1)} := \frac{\hat{\beta}^{(1)} - \beta^{(1)}\left(1 - \hat{B}^{(1)}\right)}{\beta^{(1)}\sqrt{\hat{V}^{(1)}}} \xrightarrow{d} \mathrm{N}(0,1) \ .$$

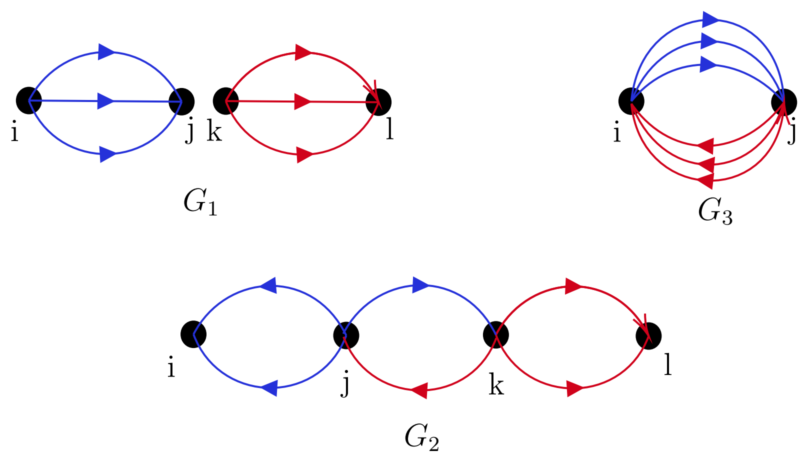

Figure 12: The possible configurations of indices that will lead to $E\left[\xi_{i_1 i_2}\xi_{i_1 i_3}\xi_{i_1 i_4} \cdot \xi_{i_5 i_6}\xi_{i_5 i_7}\xi_{i_5 i_8} \,\middle|\, U\right]$ being non-zero. These are the only graphs that can be formed using 2 3-pointed stars and in which each edge has multiplicity at least 2.

Graph	Number of Instances	$E\left[\xi_{ij}\xi_{jk}\xi_{i'j'}\xi_{j'k'} \,\middle\|\, U\right]$
G_1	$n(n-1)(n-2)(n-3)$	$p_n^2 f(U_i, U_j) f(U_k, U_l) + O_p(p_n^3)$
G_2	$n(n-1)(n-2)(n-3)$	$p_n^3 f(U_i, U_j) f(U_j, U_k) f(U_k, U_l) + O_p(p_n^4)$
G_3	$n(n-1)$	$p_n f(U_i, U_j) + O_p(p_n^2)$

Table 7: The number of instances of each graph, as well as the value of their conditional expectations, up to the leading term. Note that we can consider G_1 with $j = k$, though the contribution of this term is strictly smaller than the contribution of G_1.

A.3.2 Case (b)

As with case (a), our strategy is to remove the bias coming from $\boldsymbol{\xi}^2$ and to obtain a central limit theorem on the leading term of the remainder. Write

$$\hat{\beta}^{(T)} = \beta^{(T)} + \beta^{(T)} \frac{\left(\hat{C}^{(T)}\right)' \left(C^{(T)} - \hat{C}^{(T)}\right)}{\left(\hat{C}^{(T)}\right)' \hat{C}^{(T)}} + \frac{\left(\hat{C}^{(T)}\right)' \varepsilon}{\left(\hat{C}^{(T)}\right)' \hat{C}^{(T)}} \,. \tag{20}$$

As before, $\left(\hat{C}^{(T)}\right)' \left(C^{(T)} - \hat{C}^{(T)}\right)$ comprises mixed products of A and $\boldsymbol{\xi}$, whose order with respect to $\boldsymbol{\xi}$ is at least 1. Let B be such a term. By Lemma 2, if B is even,

$$\iota_n' B \iota_n - E\left[\iota_n' B \iota_n \,|\, U\right] = O_p\left(\frac{n^{t+1} p_n^t}{\sqrt{n}\left(\sqrt{np_n}\right)^{\tau}}\right) \,.$$

Otherwise,

$$\iota_n' B \iota_n = O_p\left(\frac{n^{t+1} p_n^t}{\sqrt{n}\left(\sqrt{np_n}\right)^{\tau}}\right) \,.$$

In other words, once the even terms are centered, the dominant terms in $\left(\hat{C}^{(T)}\right)' \left(C^{(T)} - \hat{C}^{(T)}\right)$ are of order $2T$ overall, and have order 1 with respect to $\boldsymbol{\xi}$. Such terms are dominant provided that they attain the stated upper bounds. There are T of these, taking the form below:

$$\begin{aligned}
&\iota_n' \sum_{t=1}^{T} A^{T+t-1} \boldsymbol{\xi} A^{T-t} \iota_n \\
&= \sum_{j,k} \xi_{jk} \sum_{t=1}^{T} \sum_{i} A_{i_1,i_2} A_{i_2,i_3} \cdots A_{i_{T+t-1},j} \cdot A_{k,i_{T+t+2}} \cdots A_{i_{2T}.i_{2T+1}} \\
&= \sum_{j<k} \xi_{jk} \sum_{t=1}^{T} \Bigg(\sum_{i} A_{i_1,i_2} A_{i_2,i_3} \cdots A_{i_{T+t-1},j} \cdot A_{k,i_{T+t+2}} \cdots A_{i_{2T}.i_{2T+1}} \\
&\qquad\qquad + \sum_{i} A_{i_1,i_2} A_{i_2,i_3} \cdots A_{i_{T+t-1},k} \cdot A_{j,i_{T-t+2}} \cdots A_{i_{2T}.i_{2T+1}} \Bigg) \\
&= \sum_{j<k} \xi_{jk} \sum_{t=1}^{T} \Bigg(\sum_{i} A_{i_1,i_2} A_{i_2,i_3} \cdots A_{i_{T+t-1},j} \cdot A_{k,i_{T+t+2}} \cdots A_{i_{2T}.i_{2T+1}} \\
&\qquad\qquad + \sum_{i} A_{i_{2T+1},i_{2T}} A_{i_{2T},i_{2T-1}} \cdots A_{i_{T+t+2},k} \cdot A_{j,i_{T+t-1}} \cdots A_{i_2.i_1} \Bigg) \text{ by symmetry} \\
&= \sum_{j<k} \xi_{jk} \sum_{t=1}^{2T} \left(\sum_{i} A_{i_1,i_2} A_{i_2,i_3} \cdots A_{i_{t-1},j} \cdot A_{k,i_{t+2}} \cdots A_{i_{2T}.i_{2T+1}} \right) \text{ by change of index.}
\end{aligned}$$

In the above display, summation over i is understood to exclude i_{T+t} and i_{T+t+1}, which have been replaced by j and k. Now define

$$
\begin{aligned}
V_*^{(T)}(U) &:= E\left[\iota_n' \sum_{t=1}^{T} A^{T+t-1}\boldsymbol{\xi} A^{T-t}\iota_n \,\middle|\, U\right] \\
&= \frac{1}{2}\sum_{j,k} A_{jk}(1-A_{jk})\left(\sum_{t=1}^{2T}\sum_{i} A_{i_1,i_2}A_{i_2,i_3}\cdots A_{i_{t-1},j}\cdot A_{k,i_{t+2}}\cdots A_{i_{2T}.i_{2T+1}}\right)^2 .
\end{aligned}
$$

We can get an intuition for the above term by considering binary A, in which case the variance counts the number of ways two paths of length $2T+1$ have at least one overlapping edge. The archetypal motif is displayed in Figure 13.

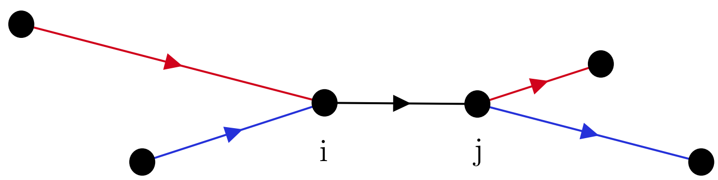

Figure 13: When A is binary, $V_*^{(T)}(U)$ counts motifs like the one displayed. Here, the red path and the blue path have the same length of $2T+1$ and overlap on the edge (i,j).

By the U-statistic LLN,

$$
\begin{aligned}
\frac{1}{n^{4T}p_n^{4T-1}}V_*^{(T)} \xrightarrow{p} \sum_{t=1}^{2T}\sum_{s=1}^{2T}\frac{1}{2}\int & f(U_1,U_2)\cdot[f(U_3,U_4)\cdots f(U_{t+1},U_1)]\cdot[f(U_2,U_{t+2})\cdots f(U_{2T},U_{2T+1})] \\
&\cdot[f(U_{2T+3},U_{2T+4})\cdots f(U_{2T+1+s},U_1)] \\
&\cdot[f(U_2,U_{2T+2+s})\cdots f(U_{4T-s-1},U_{4T-s})]\; dU \; .
\end{aligned} \tag{21}
$$

Notice that

$$
\sqrt{n^{4T}p_n^{4T-1}} = \frac{n^{2T+1}p_n^{2T}}{\sqrt{n}\sqrt{np_n}} \; ,
$$

so that our conjectured leading term in fact strictly dominates all the other terms. Now, let $\mathbf{B}$ be the set of even mixed products in $\left(\hat{C}^{(T)}\right)'\left(C^{(T)}-\hat{C}^{(T)}\right)$. Furthermore, define

$$
\begin{aligned}
B^{(T)} &= \left(\left(\hat{C}^{(T)}\right)'\hat{C}^{(T)}\right)^{-1}\sum_{B\in\mathbf{B}} E\left[\iota_n' B\iota_n \,|\, U\right] \\
V^{(T)} &= \left(\left(\hat{C}^{(T)}\right)'\hat{C}^{(T)}\right)^{-2} V_*^{(T)} \; .
\end{aligned}
$$

Then,

$$\frac{\left(\hat{\beta}^{(T)} - \beta^{(T)} - B^{(T)}\right)}{\beta^{(T)}\sqrt{V^{(T)}}} = \frac{\left(\hat{C}^{(T)}\right)'\left(C^{(T)} - \hat{C}^{(T)}\right) - \sum_{B\in\mathbf{B}} E\left[\iota_n' B \iota_n \,|\, U\right]}{\sqrt{V_*^{(T)}}}$$
$$= \frac{\iota_n' \sum_{t=1}^{T} A^{T+t-1}\boldsymbol{\xi} A^{T-t}\iota_n}{\sqrt{V_*^{(T)}}} + o_p(1) \ .$$

The proof then proceeds similarly to Case (a). Define the event $\Upsilon^{(T)} := \left\{V_*^{(T)} > kn^{4T}p_n^{4T-1}\right\}$, where k is chosen to be 1/2 the magnitude of the limit in Equation (21). Applying the Berry-Eseen inequality of Chen et al. (2011) conditioning on $U \in \Upsilon^{(T)}$ and noting that $P\left(\Upsilon^{(T)}\right) \to 1$, we obtain:

$$\frac{\iota_n' \sum_{t=1}^{T} A^{T+t-1}\boldsymbol{\xi} A^{T-t}\iota_n}{\sqrt{V_*^{(T)}}} \xrightarrow{d} \mathrm{N}(0,1)$$

which yields the desired result.

Plug-in Estimation

We need to estimate the bias for $B \in \mathbf{B}$, as well as $V_*^{(T)}$. We estimate $V_*^{(T)}$, as in case (a), we replace A_{jk} with $\hat{A}_{jk}$, and $(1 - A_{jk})$ with 1. The proof is largely similar to that in Case (a). It is tedious but straightforward since there are no rate requirements on the estimation of $V_*^{(T)}$. We discuss them in turn.

The main challenge in inference for $\beta^{(T)}$ arises because the standard deviation of $\hat{\beta}^{(d)}$ is larger than it's bias. In order for the resulting de-biased inference method to improve mean square error, bias estimation must occur at a sufficiently fast rate.

Our strategy is as follows. Let $B \in \mathbf{B}$ be order t and have order τ with respect to $\boldsymbol{\xi}$. It's block structure is described by $p = (p_1, ..., p_r)$, where each component is even. We first claim is that there exists a function $\tilde{\gamma}(t, A)$ taking the form:

$$\tilde{\gamma}(t, A) = \tilde{\gamma}_1(t) \cdot A + \tilde{\gamma}_2(t) \cdot A^2 + \cdots + \tilde{\gamma}_{t-1}(t) \cdot A^{t-1}$$

such that

$$E\left[\iota_n' B \iota_n \,|\, U\right] - \iota_n' \left(A^{t-\tau} \prod_{j=1}^{r} \tilde{\gamma}(p_j, A)\right)\iota_n = O_p\left(n^{t+1-\tau/2}p_n^{t-\tau/2}\right) \ .$$

In words, the bias of B is "close to" $\tilde{\gamma}$, a polynomial of the unobserved adjacency matrix

A. Then provided that we have good estimators of $\iota_n A^t \iota_n$, we will be able to estimate $E[\iota_n' B \iota_n \,|\, U]$ by substituting them into $\tilde{\gamma}$.

To obtain the $\tilde{\gamma}(t, A)$, write:

$$E[\iota_n' B \iota_n \,|\, U] = \sum_{i_1, \ldots i_{t+1}} E\left[B_{i_1 i_2} \cdots B_{i_t i_{t+1}} \,|\, U\right] \ . \tag{22}$$

We are interested in graphs induced by relationships on the nodes $[n]$ that will lead to non-zero contributions to the above sum. Only nodes corresponding to $\boldsymbol{\xi}$ matters, i.e. i_j s.t $j \in \mathcal{J} \cup (\mathcal{J} + 1)$. Hence, we are interested in graphs formed by overlaying r walks, each of length $p_1, ..., p_r$. For a given graph G, write its order as

$$\sum_{i \in r_G} E\left[B_{i_1 i_2} \cdots B_{i_t i_{t+1}} \,|\, U\right] = O_p(n^{\alpha+1} p_n^{\beta}).$$

where $\alpha, \beta \in \mathbb{N}$ and $\alpha \geq \beta$. By Lemma 2, we know that

$$\iota_n' B \iota_n = O_p\left(n^{t+1-\alpha/2} p_n^{t-\beta/2}\right) \preccurlyeq O_p\left(n^{t+1-\tau/2} p_n^{t-\tau/2}\right) \ .$$

Since $\tau \geq 2$, any graph for which $\alpha > \beta$ satisfies our criteria without bias correction. To achieve the $\frac{1}{\sqrt{n}}$ term in (A.3.2), we only need to deal with graphs for which $\alpha = \beta$. We called these the best case graphs in proof of 2, and they have the same characteristics as before. Namely, every edge must have multiplicity 2 greater than 2, and each edge must involve only one walk, since requiring an edge to be traversed by more than one walk increases α but not β. Thus, it is sufficient to consider the walks separately.

For a given p_j, we need to characterize walks for which $\alpha = \beta$. When $\alpha = \beta$, the order on p_n (i.e. the number of unique edges) is exactly 1 less than the order of n (i.e. the number of nodes). That is, the only graphs that matter are paths.

Let G be a walk with length p_j. Suppose that after removing duplicate edges, G is a path of length s. Then for deterministic vectors x and y,

$$\sum_{i \in r_G} x_{i_1} E\left[\xi_{i_1 i_2} \cdots \xi_{i_t i_{t+1}} \,|\, U\right] y_{i_{s+1}} = (1 + o_p(1)) \cdot \sum_{i} x_{i_1} A_{i_1 i_2} \cdots A_{i_s i_{s+1}} y_{i_{s+1}} \ .$$

The indices on the right hand side are unrestricted. The above assertion arises by the following injective mapping from $r_G \to [n]^s$. By definition, G is a walk of length p_j which traverses $s+1$ unique nodes. Let $j_1, ..., j_s, j_{s+1}$ be the steps at which G reaches a new unique node. Then our injective map is $i \mapsto (i_{j_1}, ..., i_{j_s+1})$.

Example 6. Consider the walk $i_1 \to i_2 \to i_1 \to i_3$, where all the nodes are distinct. Then $j = (1, 2, 4)$. Suppose $i = (5, 10, 5, 4)$. Then $i \mapsto (5, 10, 4)$.

There is a small error term arising from two sources. Firstly, we are only capturing the first order term of the $E\left[\xi_{i_1 i_2} \cdots \xi_{i_t i_{t+1}} \,|\, U\right]$. There are higher order terms whose magnitude are at most of p_n times this term that we omit. Second, there are paths on the right hand side which is not in the range of our injective map. These are in turn i's on which a given node appears more than once. There cannot be more than n^{s-1} such paths. Hence, these paths are at most $O_p\left(\frac{1}{n}\right)$ of the right hand side term.

Noting that $\mathbf{G}$ is finite, the previous display allows to write

$$
\begin{aligned}
E\left[\iota_n \boldsymbol{\xi}^t \iota_n \,|\, U\right] &= \sum_{G \in \mathbf{G}} \sum_{i \in r_G} x_{i_1} E\left[\xi_{i_1 i_2} \cdots \xi_{i_t i_{t+1}} \,|\, U\right] y_{i_{s+1}} && (23) \\
&= \left(1 + O_p\left(p_n \frac{1}{n}\right)\right) \cdot \sum_{G \in \mathbf{G}} x' A^{s(G)} y && (24) \\
&= \left(1 + O_p\left(p_n + \frac{1}{n}\right)\right) \cdot \sum_{s=1}^{t-1} \tilde{\gamma}_s(t) \cdot x' A^{s(G)} y && (25)
\end{aligned}
$$

In the second equation, $s(G)$ is the number of unique edges in G. Note that $s(G) \leq t/2$ since every edge must have multiplicity at least 2. In the last equation, we collected the powers of A and defined $\tilde{\gamma}_s(t)$ to be the number of walks of length t with s unique nodes.

Let us now return to the arbitrary block B. As discussed previously, it is sufficient to consider graphs in which each walk forms a component that is disconnected from all others. On those graphs, each path is independent from the others. Equation (23) therefore allows us to write

$$
\begin{aligned}
E\left[\iota_n' B \iota_n \,|\, U\right] - \iota_n' \left(A^{t-\tau} \prod_{j=1}^{r} \tilde{\gamma}(p_j, A)\right) \iota_n &= O_p\left(p_n + \frac{1}{n}\right) E\left[\iota_n' B \iota_n \,|\, U\right] \\
&= O_p\left(p_n + \frac{1}{n}\right) O_p\left(n^{t+1-\tau/2} p_n^{t-\tau/2}\right) \\
&= O_p\left(\sqrt{p_n} + \frac{1}{\sqrt{n}}\right) O_p\left(\frac{1}{\sqrt{n}}\right) O_p\left(n^{t+1-\tau/2} p_n^{t-\tau/2}\right) \ .
\end{aligned}
$$

The last equality above uses the fact that $np_n \to \infty$, yielding the desired bound. It is difficult to provide closed-form expression for $\tilde{\gamma}_s(t)$ since they are highly combinatorial. However, walks of length t are easy to enumerate on the computer for moderate t. Before proceeding, let us rewrite the $\tilde{\gamma}$ function in a more convenient form. Define $\gamma(B, A)$ such that

$$\left(A^{t-\tau}\prod_{j=1}^{r}\tilde{\gamma}(p_j, A)\right) = \gamma_1(B)A + ... + \gamma_{t-1}(B)A^{t-1} \tag{26}$$

Here, $\gamma_t(B) = 0$ because $\tau \geq 2$, and for each block of $\boldsymbol{\xi}$ in B, we have that every constituent $s(G)$ satisfies $s(G) \leq t/2$.

At this point, we have found a good estimator for $E\left[\iota_n' B \iota_n \,|\, U\right]$ in terms of the unobserved matrix A. In order to estimate $E\left[\iota_n' B \iota_n \,|\, U\right]$ at a good rate, we need good estimators for $\iota_n A^t \iota_n$. Let $\tilde{g}(t)$ be our estimator for $\iota_n A^t \iota_n$. Then we seek:

$$\iota_n' A^t \iota_n - \iota_n' \tilde{g}(t) \iota_n = o_p\left(\frac{1}{\sqrt{n}}\right) O_p\left(n^{t+1} p_n^t\right) \ . \tag{27}$$

Suppose we estimate $\iota_n A^t \iota_n$ with the naive estimator: $\iota_n \hat{A}^t \iota_n$. Our proofs, in particular Lemma 2, yields that the estimator is consistent at the following rate

$$\frac{\iota_n A^t \iota_n}{\iota_n \hat{A}^t \iota_n} = 1 + O_p\left(\frac{1}{n p_n}\right) \ ,$$

so that the error term is too large relative to the variance.

Next write

$$\iota_n' \hat{A}^t \iota_n = \iota_n' A^t \iota_n + \sum_{B \in \mathbf{B}} \iota_n' B \iota_n \ .$$

Suppose for now that we have access to $\tilde{g}(1), ..., \tilde{g}(t-1)$ satisfying Equation (27). We can then consider defining

$$\tilde{g}(t) := \iota_n' \hat{A}^t \iota_n - \sum_{B \in \mathbf{B}} \gamma(B, g)$$

where with some abuse of notation, we define

$$\gamma(B, g) := \gamma_1(B)\tilde{g}(1) + ... + \gamma_{t-1}(B)\tilde{g}(t-1) \ .$$

Since $\tilde{g}(1), ..., \tilde{g}(t-1)$ satisfy Equation (27), and noting that $\mathbf{B}$ is finite,

$$\iota_n' \hat{A}^t \iota_n - \sum_{B \in \mathbf{B}} \gamma(B, g) = \iota_n' A^t \iota_n + O_p\left(\frac{1}{\sqrt{n}}\right) O_p\left(n^{t+1/2} p_n^{t-1/2}\right) \ .$$

which satisfies Equation (27) since $\frac{1}{\sqrt{n p_n}} \to 0$. As such, we can recursively construct $\tilde{g}(t)$ from $\tilde{g}(1), ..., \tilde{g}(t-1)$. However, as our proofs in Case (a) shows, $\tilde{g}(1) = \hat{A}$ is valid. Rewrite

$\tilde{g}$ such that $\tilde{g} = g$ and

$$g(t) = g_1(t)\hat{A} + \cdots + g_t(t)\hat{A}^t \ .$$

The coefficients of $g(t)$ are presented in Table 8.

With $\gamma(\cdot, g)$ in hand, we are able to estimate $E\left[\iota_n B \iota_n \,|\, U\right]$ for arbitrary B. Recall that the bias of $\hat{\beta}^{(T)}$ is

$$B_*^{(T)} = \sum_{B \in \mathbf{B}} \delta^{t(B)} \iota_n' B \iota_n \ .$$

As before, $\mathbf{B}$ is the set of even Bs that are generated. Note that this set is "asymmetric" in that A^t appears as a product from the left but not the right. $t(B)$ is the function giving the order of B. Debiasing by our estimators, we obtain that

$$\sum_{B \in \mathbf{B}} \delta^{t(B)} \iota_n' \left(B - \gamma(B, g)\right) \iota_n = O_p\left(\frac{1}{\sqrt{n}}\right) O_p\left(n^{2T} p_n^{2T-1}\right) \ .$$

This is because B is even and of order at most $2T$. As such, since $\tau \geq 2$,

$$\iota_n' B - \gamma(B, g)\iota_n = O_p\left(\frac{1}{\sqrt{n}}\right) O_p\left(n^{2T+1-\tau/2} p_n^{2T-\tau/2}\right) = O_p\left(\frac{1}{\sqrt{n}}\right) O_p\left(n^{2T} p_n^{2T-1}\right)$$

Now, since $\sqrt{V_*^{(T)}} = \Omega_p\left(n^{2T} p_n^{2T-1/2}\right)$, we conclude that

$$\frac{\hat{B}_*^{(T)} - B_*^{(T)}}{\sqrt{V_*^{(T)}}} = o_p(1) \ .$$

Substituting our computed values of $\tilde{\gamma}_s(t)$ and $g(t)$ yields the formula given in Appendix B.

A.3.3 Case (c)

Suppose $\beta^{(d)} = 0$ for $d \in \{1, T\}$. We can write

$$\hat{\beta}^{(T)} = \frac{\left(\hat{C}^{(T)}\right)' \varepsilon}{\left(\hat{C}^{(T)}\right)' \hat{C}^{(T)}} = \frac{\iota_n \left(\sum_{t=1}^T \delta^t \hat{A}^t\right) \varepsilon}{\iota_n \left(\sum_{t=1}^T \delta^t \hat{A}^t\right)^2 \iota_n}$$

r \ t	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	-1	2	-5	12	-20	-12	295	-1584	5623	-12530	-1806	186702
2		1	-2	4	-8	8	42	-340	1510	-4712	8408	13088	-194318
3			1	-3	7	-14	10	96	-655	2552	-6190	1068	83832
4				1	-4	11	-24	22	142	-1043	4078	-9444	-2150
5					1	-5	16	-39	48	176	-1558	6542	-16554
6						1	-6	22	-60	94	178	-2170	10028
7							1	-7	29	-88	167	122	-2836
8								1	-8	37	-124	275	-26
9									1	-9	46	-169	427
10										1	-10	56	-224
11											1	-11	67
12												1	-12
13													1

r \ t	14	15	16	17	18	19	20
1	-1101323	3938488	-7533897	-13585642	198008994	-999517964	3021609795
2	981200	-3101066	4292162	20354680	-188470026	832916330	-2145039932
3	-530446	2005368	-4310942	-4074647	91205574	-496007668	1614224856
4	151068	-879116	3034670	-4907736	-17574745	186419358	-871382472
5	4548	213314	-1337608	4785512	-8228118	-25081260	283591630
6	-28178	22194	281946	-2019280	7855844	-16309132	-23702626
7	14700	-46038	58866	333648	-2899960	12404253	-30152117
8	-3480	20662	-72062	124800	337020	-3955392	18806973
9	-309	-3984	27916	-108304	232853	246742	-5122509
10	633	-780	-4178	36312	-156828	398862	-1830
11	-290	904	-1503	-3829	45486	-219512	641615
12	79	-368	1252	-2554	-2629	54786	-297780
13	-13	92	-459	1690	-4022	-182	63185
14	1	-14	106	-564	2232	-6010	4010
15		1	-15	121	-684	2893	-8636
16			1	-16	137	-820	3689
17				1	-17	154	-973
18					1	-18	172
19						1	-19
20							1

Table 8: The coefficients $g_r(t)$ for $t \leq 20$.

Let B be a mixed product of order t, and let it have order $\tau \geq 0$ with respect to $\boldsymbol{\xi}$. Then,

$$
\begin{aligned}
&E\left[(\varepsilon' B \iota_n)^2 \mid U, \boldsymbol{\xi}\right] \\
&= \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_{2t+2}=1}^{n} E\left[\varepsilon_{i_1}\varepsilon_{i_{t+2}} \mid U, \boldsymbol{\xi}\right] B_{i_1,i_2} B_{i_2,i_3} \cdots B_{i_t,i_{t+1}} \cdot B_{i_{t+2},i_{t+3}} B_{i_{t+3},i_{t+4}} \cdots B_{i_{2t+1},k_{2t+2}}
\end{aligned}
$$

Now, $E\left[\varepsilon_{i_1}\varepsilon_{i_{t+2}} \mid U, \boldsymbol{\xi}\right] = 0$ unless $i_1 = i_{t+2}$. Hence, we only need to consider sequences i where $i_1 = i_{t+2}$. Under this restriction,

$$
E\left[(\varepsilon' B \iota_n)^2 \mid U, \boldsymbol{\xi}\right] \leq \bar{\sigma}^2 \iota_n' \tilde{B} \iota_n \ .
$$

where $\tilde{B}$ is of order $2t+1$ unconditionally, and of order 2τ with respect to $\boldsymbol{\xi}$. Conclude by Lemma 2 that

$$
\varepsilon' B \iota_n = O_p\left(\sqrt{n^{2t+3/2-\tau} p_n^{2t+1-\tau}}\right) = O_p\left(n^{t+3/4-\tau/2} p_n^{t+1/2-\tau/2}\right) \ . \tag{28}
$$

Next, write

$$
\begin{aligned}
\frac{1}{p_n^T} \iota_n' A^T \varepsilon &= \sum_{i_1,\ldots,i_{T+1}} f(U_{i_1}, U_{i_2}) \cdots f(U_{i_T}, U_{i_{T+1}}) \varepsilon_{i_{T+1}} \\
&= \sum_{i \in I_{T+1}} \sum_{\pi \in \Pi_{T+1}} f(U_{i_{\pi(1)}}, U_{i_{\pi(2)}}) \cdots f(U_{i_{\pi(T)}}, U_{i_{\pi(T+1)}}) \varepsilon_{(i_{\pi(T+1)})} \ .
\end{aligned}
$$

where I_{T+1} comprises all unordered subsets of $T+1$ integers chosen from $[n]$ and Π_{T+1} is the set of permutations on $[T+1]$. We can hence define the following symmetric U-statistic kernel of order $T+1$:

$$
h\left((U_{i_1}, \varepsilon_{i_1}), \cdots, (U_{i_{T+1}}, \varepsilon_{i_{T+1}})\right) = \sum_{\pi \in \Pi_{T+1}} f(U_{i_{\pi(1)}}, U_{i_{\pi(2)}}) \cdots f(U_{i_{\pi(T)}}, U_{i_{\pi(T+1)}}) \varepsilon_{i_{\pi(T+1)}} \ .
$$

Since f is bounded and ε_i has uniformly bounded conditional expectations, $E[h^2] < \infty$. By the U-statistic CLT (Theorem 12.3 in Van der Vaart 2000),

$$
\sqrt{n} \frac{1}{\binom{n}{T+1}} \sum_{i \in I_{T+1}} \sum_{\pi \in \Pi_{T+1}} f(U_{i_{\pi(1)}}, U_{i_{\pi(2)}}) \cdots f(U_{i_{\pi(T)}}, U_{i_{\pi(T+1)}}) \varepsilon_{(i_{\pi(T+1)})} \xrightarrow{d} \mathrm{N}\left(0\,,\, (T+1)^2 \zeta_1\right) \ , \tag{29}
$$

where

$$\begin{aligned}\zeta_1 &= E\left[h\left(\left(U_{i_1},\varepsilon_{i_1}\right),\left(U_{i_2},\varepsilon_{i_2}\right),\cdots,\left(U_{i_{T+1}},\varepsilon_{i_{T+1}}\right)\right)h\left(\left(U_{i_1},\varepsilon_{i_1}\right),\left(U'_{i_2},\varepsilon'_{i_2}\right),\cdots,\left(U'_{i_{T+1}},\varepsilon'_{i_{T+1}}\right)\right)\right] \\ &= E\left[\sum_{\pi\in\Pi_T}\sum_{\pi'\in\Pi_T} f(U_{i_{\pi(1)}},U_{i_{\pi(2)}})\cdots f(U_{i_{\pi(T)}},U_{i_{T+1}})\cdot f(U'_{i_{\pi'(1)}},U'_{i_{\pi'(2)}})\cdots f(U'_{i_{\pi'(T)}},U_{i_{T+1}})\varepsilon^2_{i_{T+1}}\right] \\ &= (T!)^2\, E\left[f(U_1,U_2)\cdots f(U_T,U_{T+1})\cdot f(U_1,U_{T+2})\cdots f(U_{2T},U_{2T+1})\varepsilon_1^2\right] \neq 0 \text{ by assumption.}\end{aligned}$$

As such,

$$\iota_n' A^T\varepsilon = O_p\left(\frac{1}{\sqrt{n}}n^{T+1}p_n^T\right) .$$

Together with the bound in Equation (28), this implies that $\iota_n' A^T\varepsilon$ is the dominant term in the $\left(\hat{C}^{(T)}\right)'\varepsilon$ Next, note that by the U-statistic LLN,

$$\begin{aligned}\frac{1}{n^{2T+1}}\sum_{j=1}^n\left(\iota_n\left(A^T\right)_{\cdot,j}\right)^2\varepsilon_j^2 &= \frac{1}{n^{2T+1}}\sum_{i_1,\dots,i_{2T+1}} f(U_{i_1},U_{i_2})\cdots f(U_{i_T},U_{i_{T+1}})\cdot f(U_{i_1},U_{i_{T+2}})\cdots f(U_{i_{2T}},U_{i_{2T+1}})\varepsilon^2_{i_{2T+1}} \\ &\xrightarrow{p} \frac{1}{(T!)^2}\zeta_1 .\end{aligned}$$

By the usual plug-in arguments, we have that

$$\begin{aligned}\frac{1}{n^{2T+1}}\left(\left(\hat{C}^{(T)}\right)'\hat{C}^{(T)}\right)^2\hat{V}_0^{(T)} &= \frac{1}{n^{2T+1}}\sum_{j=1}^n\left(\hat{C}_j^{(T)}\right)^2\hat{\varepsilon}_j^2 \\ &= \frac{1}{n^{2T+1}}\sum_{j=1}^n\left(C_j^{(T)}\right)^2\varepsilon_j^2 + o_p(1) \quad \text{by the consistency of } \hat{\beta}^{(T)} \\ &= \frac{1}{n^{2T+1}}\sum_{j=1}^n\delta^{2T}\left(\iota_n\left(A^T\right)_{\cdot,j}\right)^2\varepsilon_j^2 + o_p(1) \quad \text{by the dominance of } A^T\end{aligned}$$

As such, the robust/heteroskedasticity consistent t-statistic

$$
\begin{aligned}
\frac{\hat{\beta}^{(T)}}{\sqrt{\hat{V}_0^{(T)}}} &= \frac{\left(\hat{C}^{(T)}\right)'\varepsilon}{\sqrt{\left(\left(\hat{C}^{(T)}\right)'\hat{C}^{(T)}\right)^2 \hat{V}_0^{(T)}}} \\
&= \frac{n^{-(T+1/2)}\delta^T \iota_n' A^T \varepsilon}{\sqrt{\frac{1}{n^{2T+1}}\sum_{j=1}^n \delta^{2T}\left(\iota_n \left(A^T\right)_{\cdot,j}\right)^2 \varepsilon_j^2}} + o_p(1) \\
&= \frac{\frac{\sqrt{n}}{\binom{n}{T+1}}\sum_{i\in I_{T+1}}\sum_{\pi\in\Pi_{T+1}} f(U_{i_{\pi(1)}}, U_{i_{\pi(2)}})\cdots f(U_{i_{\pi(T)}}, U_{i_{\pi(T+1)}})\varepsilon_{(i_{\pi(T+1)})}}{(T+1)!\sqrt{\frac{1}{(T!)^2}\zeta_1}} + o_p(1) \\
&\xrightarrow{d} \mathrm{N}(0,1) \quad \text{by Equation (29)}
\end{aligned}
$$

As such, the robust/heteroskedasticity consistent t-statistic is appropriate for inference under the null hypothesis that $\beta^{(T)} = 0$.

A.4 Proof of Theorem 5

We start by writing

$$
\hat{\beta}^{(\infty)} = \frac{Y'\left(a_n v_1(\hat{A})\right)}{\left(a_n v_1(\hat{A})\right)'\left(a_n v_1(\hat{A})\right)} = \beta^{(\infty)}\left(v_1(A)\right)' v_1(\hat{A}) + \frac{1}{a_n}\varepsilon' v_1(\hat{A}) \ .
$$

The main tool we use to study the above term is the following:

Lemma 3. Suppose Assumption E2 holds and that p_n satisfies Equation (7). Then,

$$
\begin{aligned}
\left(v_1(A)\right)' v_1(\hat{A}) = \left(v_1(A)\right)' v_1(A) + \frac{\left(v_1(A)\right)' \boldsymbol{\xi} v_1(A)}{\lambda_1(A)} + \frac{\left(v_1(A)\right)' \boldsymbol{\xi}^2 v_1(A)}{\left(\lambda_1(A)\right)^2} + o_p\left(\frac{1}{\left(np_n\right)^3}\right) \\
+ \sum_{r=2}^R \frac{\lambda_r(A)}{\lambda_1(A)} \frac{v_r(A)' v_1(\hat{A})}{v_r(A)' v_r(\hat{A})} \cdot O_p\left(\frac{1}{np_n}\right)
\end{aligned}
$$

and

$$
\left(\frac{\varepsilon}{a_n}\right)' v_1(\hat{A}) = \left(\frac{\varepsilon}{a_n}\right)' v_1(A) + o_p\left(\frac{1}{a_n}\right) \ .
$$

Now, by the analogous arguments as in Lemma 2 and proof of Lemma 3,

$$\frac{(v_1(A))' \boldsymbol{\xi} v_1(A)}{\lambda_1(A)} = O_p\left(\frac{1}{\sqrt{n}\sqrt{np_n}}\right) \quad , \quad \frac{(v_1(A))' \boldsymbol{\xi}^2 v_1(A)}{(\lambda_1(A))^2} = O_p\left(\frac{1}{np_n}\right) \quad , \tag{30}$$

$$\frac{(v_1(A))' \boldsymbol{\xi}^2 v_1(A)}{(\lambda_1(A))^2} - \frac{E\left[(v_1(A))' \boldsymbol{\xi}^2 v_1(A) \,|\, U\right]}{(\lambda_1(A))^2} = O_p\left(\frac{1}{\sqrt{n}\,(np_n)}\right) \tag{31}$$

Furthermore, by the Davis-Kahan Inequality,

$$\left|\frac{1}{\lambda_s(A)} v_r(A)' v_s(\hat{A})\right| \leq \left|\frac{1}{\lambda_s(A)} v_r(A)' v_s(A)\right| + \|v_r(A)\| \cdot \left\|\frac{v_s(\hat{A}) - \theta v_s(A)}{\lambda_s(A)}\right\| \ . \tag{32}$$

A.4.1 Case (a)

Let us now consider individual cases in (a). starting with (iii). Suppose we require only that $p_n \succ n^{-1}\left(\frac{\log n}{\log\log n}\right)^{1/2+\eta}$. Then by Lemma 1, we can claim that

$$\left\|\frac{v_s(\hat{A}) - \theta v_s(A)}{\lambda_s(A)}\right\| = o_p(1)$$

but cannot control the rate of convergence. Nonetheless, if $a_n \succ np_n$, Equations (30) and (32), together with Lemma 3 implies that

$$\hat{\beta}^{(\infty)} = \beta^{(\infty)} + \left(\frac{\varepsilon}{a_n}\right)' v_1(A) + o_p\left(\frac{1}{a_n}\right) \ . \tag{33}$$

Note that bias correction is irrelevant in this regime since bias is of smaller order than a_n.

Suppose instead, as in Case (a) (ii) that $p_n \succ n^{-1}\log n$. Then, by Theorem 1.1 in the Supplementary Appendix to Lei and Rinaldo (2015), we can claim that

$$\left\|\frac{v_s(\hat{A}) - \theta v_s(A)}{\lambda_s(A)}\right\| = O_p\left(\frac{1}{\sqrt{np_n}}\right)$$

Then, provided that $a_n \prec (np_n)^{3/2}$,

$$\hat{\beta}^{(\infty)} - \beta^{(\infty)} - \beta^{(\infty)} \frac{E\left[(v_1(A))' \boldsymbol{\xi}^2 v_1(A) \,|\, U\right]}{(\lambda_1(A))^2} = \left(\frac{\varepsilon}{a_n}\right)' v_1(A) + o_p\left(\frac{1}{a_n}\right) \ .$$

This time the bias correction is required.

Finally, consider Case (a) (i), when $\beta^{(\infty)} = 0$. Then it is immediate that Equation

(33) obtains. In all three cases, the asymptotic distribution of the estimator depends on $\left(\frac{\varepsilon}{a_n}\right)' v_1(A)$. Next, notice that

$$\mathrm{Var}\left(\varepsilon' v_1(A) \,|\, U\right) = E\left[\left(\varepsilon' v_1(A)\right)^2 \,|\, U\right] = \sum_{i=1}^{n} \left[v_1(A)\right]_i^2 E\left[\varepsilon_i^2 \,|\, U\right] \ .$$

As such, by Assumption 3,

$$\underline{\sigma}^2 \leq \mathrm{Var}\left(\varepsilon' v_1(A) \,|\, U\right) \leq \bar{\sigma}^2 \ .$$

Now, as in the proof of Lemma 3, let Υ be the event that $|\frac{1}{\sqrt{n}} \sum_{i=1}^{r} \phi_r(U_i)\phi_s(U_i) < 1/R^2$. This happens with probability approaching 1 since R is finite. On this event, $\|v\| = 1$ implies that $|v_r| < 2$ for all r. Furthermore, observe that since $\|f\|_\infty \leq 1$, $\|\phi_r\|_\infty \leq 1$. As such, on Υ, $|v_r(A)\|_\infty \leq 2R/\sqrt{n}$. As such,

$$\sum_{i=1}^{n} \frac{\left[v_1(A)\right]_i^3 E\left[\varepsilon_i^3 \,|\, U\right]}{\mathrm{Var}\left(\varepsilon' v_1(A) \,|\, U\right)} \leq \frac{\bar{\kappa}_3}{\underline{\sigma}^2} \frac{8R^3}{\sqrt{n}} \to 0 \ .$$

Note that the bound on the right-hand side does not depend on U. Putting the above ingredients together, we have that on Υ, the Berry-Esseen Inequality of Chen et al. (2011) yields

$$\sup_{z \in \mathbf{R}} \left| P\left(\frac{\varepsilon' v_1(A)}{\sqrt{\mathrm{Var}\left(\varepsilon' v_1(A) \,|\, U\right)}} \leq z\right) - \Phi(z)\right| \leq 10 \cdot \frac{\bar{\kappa}_3}{\underline{\sigma}^2} \frac{8R^3}{\sqrt{n}} \ .$$

Then since $P(\Upsilon) \to 1$, we have that

$$\frac{\varepsilon' v_1(A)}{\sqrt{\mathrm{Var}\left(\varepsilon' v_1(A) \,|\, U\right)}} \xrightarrow{d} \mathrm{N}(0, 1) \ .$$

Define

$$\begin{aligned} B^{(\infty)} &= \frac{E\left[\left(v_1(A)\right)' \boldsymbol{\xi}^2 v_1(A) \,|\, U\right]}{\left(\lambda_1(A)\right)^2} \\ V_0^{(\infty)} &= \mathrm{Var}\left(\varepsilon' v_1(A) \,|\, U\right) \ , \end{aligned}$$

Then we have that

$$\frac{\hat{\beta}^{(\infty)} - \beta^{(\infty)}\left(1 - B^{(\infty)}\right)}{\sqrt{V_0^{(\infty)}}} = \frac{\varepsilon' v_1(A)}{\sqrt{\mathrm{Var}\left(\varepsilon' v_1(A) \,|\, U\right)}} + o_p(1) \xrightarrow{d} \mathrm{N}(0, 1) \ .$$

The validity of plug-in estimation follows from arguments that are essentially identical to Section A.3.3.

A.4.2 Case (b)

Suppose $a_n \succ n\sqrt{p_n}$. By Lemma 3, we have that

$$\hat{\beta}^{(\infty)} = \beta^{(\infty)} + \beta^{(\infty)}\Bigg[(v_1(A))' \, v_1(A) + \frac{(v_1(A))' \, \boldsymbol{\xi} v_1(A)}{\lambda_1(A)} + \frac{(v_1(A))' \, \boldsymbol{\xi}^2 v_1(A)}{(\lambda_1(A))^2} + o_p\left(\frac{1}{(np_n)^3}\right) + \sum_{r=2}^{R} \frac{\lambda_r(A)}{\lambda_1(A)} \frac{v_r(A)' v_1(\hat{A})}{v_r(A)' v_r(\hat{A})} \cdot O_p\left(\frac{1}{np_n}\right) + o_p\left(\frac{1}{a_n}\right)\Bigg]$$

Furthermore, since $p_n \succ n^{-1}\log n$, the Davis-Kahan Inequality (Theorem 4.5.5 in Vershynin 2018), together with Theorem 1.1 in the Supplymentary Material to Lei and Rinaldo (2015) gives us that

$$\sum_{r=2}^{R} \frac{\lambda_r(A)}{\lambda_1(A)} \frac{v_r(A)' v_1(\hat{A})}{v_r(A)' v_r(\hat{A})} = O_p\left(\frac{1}{\sqrt{np_n}}\right) .$$

As such,

$$\begin{aligned} \hat{\beta}^{(\infty)} - \beta^{(\infty)} \left(1 - B^{(\infty)}\right) &= \beta^{(\infty)} \frac{(v_1(A))' \, \boldsymbol{\xi} v_1(A)}{\lambda_1(A)} + O_p\left(\frac{1}{(np_n)^{3/2}}\right) \\ &= \beta^{(\infty)} \frac{(v_1(A))' \, \boldsymbol{\xi} v_1(A)}{\lambda_1(A)} + O_p\left(\frac{1}{n\sqrt{p_n}}\right) , \end{aligned} \tag{34}$$

since $p_n \succ 1/\sqrt{n}$. Now note that

$$\begin{aligned} &\mathrm{Var}\left(v_1(A)' \boldsymbol{\xi} v_1(A) \,|\, U\right) \\ &= E\left[(v_1(A)' \boldsymbol{\xi} v_1(A))\right] \\ &= E\left[\left(2\sum_{i<j} [v_1(A)]_i \, [v_1(A)]_j \, \xi_{ij}\right)^2 \,\Bigg|\, U\right] \\ &= \sum_{i<j} [v_1(A)]_i^2 \, [v_1(A)]_j^2 \, p_n f(U_i, U_j) \left(1 - p_n f(U_i, U_j)\right) \\ &= 4\left(1 + O_p(p_n)\right) \cdot \sum_{i<j} [v_1(A)]_i^2 \, [v_1(A)]_j^2 \, p_n f(U_i, U_j) \end{aligned}$$

Recall that from the proof of Lemma 3 that

$$A = \sum_{r=1}^{R} \tilde{\lambda} n \left(\frac{\phi_r(U)}{\sqrt{n}}\right) \frac{\phi_r(U)}{\sqrt{n}} \quad \Rightarrow \quad v_1(A) = \sum_{r=1}^{R} \alpha_r \frac{\phi_r(U)}{\sqrt{n}} .$$

so that $|\alpha_r| \leq 2R$ for all $r \in [R]$ w.p.a. 1. We now argue that $\alpha_1 \xrightarrow{p} 1$ and $\alpha_r \to 0$ for $r \geq 2$. Note that we can write

$$\begin{aligned} Av_1(A) &= \left(\sum_{r=1}^{R} \tilde{\lambda}_r n \left(\frac{\phi_r(U)}{\sqrt{n}} \right) \frac{\phi_r(U)}{\sqrt{n}} \right) \left(\sum_{r=1}^{R} \alpha_r \frac{\phi_r(U)}{\sqrt{n}} \right) \\ &= \sum_{r=1}^{R} \tilde{\lambda}_r n \cdot \alpha_r \phi_r \cdot \left(\frac{\phi_r(U)}{\sqrt{n}} \right)' \frac{\phi_r(U)}{\sqrt{n}} + \sum_{r \neq s} \tilde{\lambda}_r \alpha_s \left(\frac{\phi_r(U)}{\sqrt{n}} \right) \frac{\phi_s(U)}{\sqrt{n}} \ . \end{aligned}$$

Consequently,

$$\begin{aligned} (v_1(A))' Av_1(A) = & \sum_{r=1}^{R} \tilde{\lambda}_r n \alpha_r^2 \left(\left(\frac{\phi_r(U)}{\sqrt{n}} \right)' \frac{\phi_r(U)}{\sqrt{n}} \right)^2 \\ &+ \sum_{r=1}^{R} \sum_{s=1}^{R} \tilde{\lambda}_r n \alpha_r \alpha_r \left(\frac{\phi_r(U)}{\sqrt{n}} \right) \frac{\phi_s(U)}{\sqrt{n}} \left(\frac{\phi_r(U)}{\sqrt{n}} \right)' \left(\frac{\phi_r(U)}{\sqrt{n}} \right) \\ &+ \sum_{r=1}^{R} \sum_{s \neq t} \tilde{\lambda}_s n \alpha_t \alpha_r \left(\frac{\phi_r(U)}{\sqrt{n}} \right) \frac{\phi_s(U)}{\sqrt{n}} \left(\frac{\phi_r(U)}{\sqrt{n}} \right) \frac{\phi_t(U)}{\sqrt{n}} \end{aligned}$$

Now,

$$\left(\frac{\phi_r(U)}{\sqrt{n}} \right)' \frac{\phi_r(U)}{\sqrt{n}} = \frac{1}{n} \sum_{i=1}^{n} \phi_r(U_i) \phi_s(U_i) = \begin{cases} 1 + O_p\left(\frac{1}{\sqrt{n}}\right) & \text{if } r = s \\ O_p\left(\frac{1}{\sqrt{n}}\right) & \text{if } r \neq s \end{cases}$$

Since $|\alpha_r| \leq 2R$ w.p.a 1,

$$(v_1(A))' Av_1(A) = \sum_{r=1}^{R} \tilde{\lambda}_r n \alpha_r^2 + o_p(1)$$

Since $|\tilde{\lambda}_1| > |\tilde{\lambda}_r|$ for $r \geq 2$,

$$\frac{(v_1(A))' Av_1(A)}{\tilde{\lambda}_1 n} \xrightarrow{p} 1 \quad \Rightarrow \quad \alpha_1 \xrightarrow{p} 1 \ .$$

Doing a similar expansion for $\|v\|^2$, we arrive at

$$1 = \|v_1(A)\|^2 = \sum_{r=1}^{R} \alpha_r^2 + O_p\left(\frac{1}{\sqrt{n}}\right) \ .$$

Since $\alpha_1 \xrightarrow{p} 1$, $\alpha_r \xrightarrow{p} 0$ for all $r \geq 2$. Since $|\phi_r(U_i)| \leq 1$, the above analysis also implies that

$$|[V_1(A)]_i| = \left| \sum_{r=1}^{R} (\alpha_r + o_p(1)) \frac{\phi_r(U_i)}{\sqrt{n}} \right| \leq \frac{1}{\sqrt{ns}} \sum_{r=1}^{R} |\alpha_r| + o_p(1)$$

where the bound on the right hand side depends on the convergence of α_r and does not vary across i. This yields $\|v_1(A)\|_\infty \leq 2/\sqrt{n}$ w.p.a. 1.

As such,

$$\mathrm{Var}\left(n \cdot v_1(A)' \boldsymbol{\xi} v_1(A) \,|\, U\right) = 4\left(1 + o_p(1)\right) \sum_{i<j} \phi_1(U_i)^2 \phi_1(U_j)^2 p_n f(U_i, U_j) \ .$$

Conclude by the U-statistics LLN that

$$\frac{1}{n^2 p_n} \mathrm{Var}\left(n \cdot v_1(A)' \boldsymbol{\xi} v_1(A) \,|\, U\right) \xrightarrow{p} 2E\left[\phi_1(U_1)^2 \phi_1(U_2)^2 f(U_1, U_2)\right] > 0 \ .$$

As such, if we define $\Upsilon^{(\infty)}$ to be the event on which

$$\mathrm{Var}\left(n \cdot v_1(A)' \boldsymbol{\xi} v_1(A) \,| U\right) > n^2 p_n \cdot E\left[\phi_1(U_1)^2 \phi_1(U_2)^2 f(U_1, U_2)\right]$$

and $\|v_1(A)\|_\infty \leq 2/\sqrt{n}$. By the Berry-Esseen Inequality,

$$\sup_{x \in \mathbf{R}} \left| P\left(\frac{(n \cdot v_1(A))\, \boldsymbol{\xi} v_1(A)}{\sqrt{\mathrm{Var}\left(n \cdot v_1(A)' \boldsymbol{\xi} v_1(A) \,|\, U\right)}} \ \leq \ x \,\middle|\, U, \Upsilon^{(\infty)} \right) - \Phi(x) \right| \leq 10\gamma$$

where

$$\begin{aligned}
\gamma &= \sum_{i<j} \left[\sqrt{n} \cdot v_1(A)\right]_i^3 \left[\sqrt{n} v_1(A)\right]_j^3 \frac{E\left[\xi_{ij}^3 \,|\, U\right]}{\left(\mathrm{Var}\left(n \cdot v_1(A)' \boldsymbol{\xi} v_1(A) \,|\, U\right)\right)^{3/2}} \\
&\leq 64 R^4 \frac{n^2 p_n}{\left(n^2 p_n E\left[\phi_1(U_1)^2 \phi_1(U_2)^2 f(U_1, U_2)\right]\right)^{3/2}} \to 0
\end{aligned}$$

where the last bound follows because we are on the event Υ and does not depend on U otherwise. Substituting this into Equation (34), we have that

$$\frac{\lambda_1(A)\left(\hat{\beta}^{(\infty)} - \beta^{(\infty)}\left(1 - B^{(\infty)}\right)\right)}{\beta^{(\infty)} \sqrt{V^{(\infty)}}} = \frac{(n \cdot v_1(A))\, \boldsymbol{\xi} v_1(A)}{\sqrt{\mathrm{Var}\left(n \cdot v_1(A)' \boldsymbol{\xi} v_1(A) \,|\, U\right)}} + o_p(1) \xrightarrow{d} \mathrm{N}(0, 1) \ ,$$

where the estimate on the error follows because $\lambda_1(A)/\sqrt{V^{(\infty)}} = O_p\left(n^{-1} p_n^{-1/2}\right)$.

The validity of plug-in estimation follows from arguments that are essentially identical to Section A.3.1.

A.5 Proofs of Auxillary Lemmas

A.5.1 Proof of Lemma 1

Suppose $np_n \succcurlyeq \log^2 n$. By Theorem 1.1 in the Supplementary Material to Lei and Rinaldo (2015), noting that the constants in their bounds are uniform in P_{ij} we have that with probability at least $1 - n^{-r}$, where r can be chosen independently of

$$\left\| A - \hat{A} \right\| \leq k(r)\sqrt{np_n}$$

where $k(r)$ is a constant that depends only on r.

Suppose instead that $\sqrt{\frac{\log n}{\log\log n}} \prec np_n \prec \log^2 n$. Our set up satisfies the requirements for Corollary 3.3 in Benaych-Georges et al. (2020). Setting their $\varepsilon^2 = \left(\sqrt{\frac{\log n}{\log\log n}}/(np_n)\right)^{1-\nu}$ and noting that their d is our np_n, we have that with probability at least $1-\exp\left(-(np_n)^{2+\nu}\left(\frac{\log n}{\log\log n}\right)^{(1-\nu)/2} k\right)$

$$\left\| A - \hat{A} \right\| \leq k\,(np_n)^{(1+\nu)/2}\left(\frac{\log n}{\log\log n}\right)^{(1-\nu)/4}$$

where k is a universal constant.

Combining these two inequalities yields the desired results.

A.5.2 Proof of Lemma 2

Let B be a mixed product as in Definition 10. Suppose $B_j = \boldsymbol{\xi}$ for at least one $j \in [t]$ and write:

$$\begin{aligned}(\iota_n' B \iota_n)^2 &= \left(\sum_{i=1}^n \sum_{k_1=1}^n \cdots \sum_{k_{t-1}=1}^n \sum_{j=1}^n b_{i,k_1} b_{k_1,k_2} \cdots b_{k_{t-1},j}\right) \cdot \left(\sum_{i'=1}^n \sum_{k_1'=1}^n \cdots \sum_{k_{t-1}'=1}^n \sum_{j'=1}^n b_{i',k_1'} b_{k_1',k_2'} \cdots b_{k_{t-1}',j'}\right) \\ &= \sum_{k_1=1}^n \sum_{k_2=1}^n \cdots \sum_{k_{2t+2}=1}^n b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_t+1} \cdot b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \cdots b_{k_{2t+1},k_{2t+2}}\end{aligned}$$

In the second line we relabel the indices of summation. Each term in the above sum is a product of $2t$ terms. Note that the term $b_{k_{t+1},k_{t+2}}$ does not exist. Next, we take conditional expectations:

$$E\left[\left(\iota_n' B\iota_n\right)^2 \mid U\right] = \sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\cdots\sum_{k_{2t+2}=1}^{n} E\left[b_{k_1,k_2}b_{k_2,k_3}\cdots b_{k_t,k_t+1}\cdot b_{k_{t+2},k_{t+3}}b_{k_{t+3},k_{t+4}}\cdots b_{k_{2t+1},k_{2t+2}} \mid U\right] .$$

Notice that each summand is non-zero if an only if for each $b_{k_j,k_{j+1}} = \xi_{k_j,k_{j+1}}$, there is some j' such that

$$b_{k_{j'},k_{j'+1}} = \xi_{k_{j'},k_{j'+1}} = \begin{cases} \xi_{k_j,k_{j+1}} & \text{or,} \\ \xi_{k_{j+1},k_j} & \end{cases}$$

In other words, each ξ_{ij} that appears in the summand appears at least twice, either as ξ_{ij} or ξ_{ji}. This property depends on the positions of the A's to the extent that they break up $\boldsymbol{\xi}$: neighbouring ξ_{ij} and ξ_{jk} share an index so that setting $i = k$ is sufficient for the conditional expectation of their product to be non-zero. If ξ_{ij} and ξ_{kl} are separated by at least one A_{pq}, we need to set $k = i$ and $l = j$. The number of restrictions on the indices that are needed for the terms to be non-zero therefore depend on $\mathcal{J}$ and p. In turn, these restrictions determine the order of magnitude of the conditional expectation.

We are interested in relations on $k_1, ..., k_{2t+2}$ which will make

$$E\left[b_{k_1,k_2}\cdots b_{k_t,k_{t+1}}\cdot b_{k_{t+2},k_{t+3}}\cdots b_{k_{2t+1},k_{2t+2}} \mid U\right] \neq 0 .$$

We represent this relation with the multi-graph G on nodes $[n]$ with each ξ_{ij} in the summand corresponding to an edge from i to j. If G is the multi-graph induced by a given relationship, we write that $k_1, ..., k_{2t+2} \in r_G$. Let the contribution of r_G to the our overall sum be:

$$\sum_{(k_1,\ldots,k_{2t+2})\in r_G} E\left[b_{k_1,k_2}\cdots b_{k_t,k_t+1}\cdot b_{k_{t+2},k_{t+3}}\cdots b_{k_{2t+1},k_{2t+2}} \mid U\right] =: S_G$$

For S_G to be non-zero, every edge in G must have multiplicity at least 2. Furthermore, each G is constructed by performing a walk of length p_1, followed by p_2, and so on, until p_r.

The walks relate G to S_G in the following way. Initially, we are given a budget of $n^{2t+2}p_n^{2t}$. The budget on n is the number of times the graph G occurs, corresponding to "degree of freedom". The budget on p_n is the number of unique ξ_{ij} in the term. Given any initial vertex, start the first walk of length p_1. Add one to the multiplicity of each edge taken. In the j^{th} step, incrementing multiplicity from 0 to 1 is free: this corresponds to not restricting k_j and k_{j+1}. Incrementing multiplicity from a to $a + 1$ for $a \geq 1$ costs np_n. This is because such a step corresponds to the restriction $k_j = k_{j'}$ where $k_{j'}$ denotes the other end point of the edge whose multiplicity is being incremented. Furthermore, we "lose" p_n when we restrict $\xi_{k_j,k_{j+1}}$

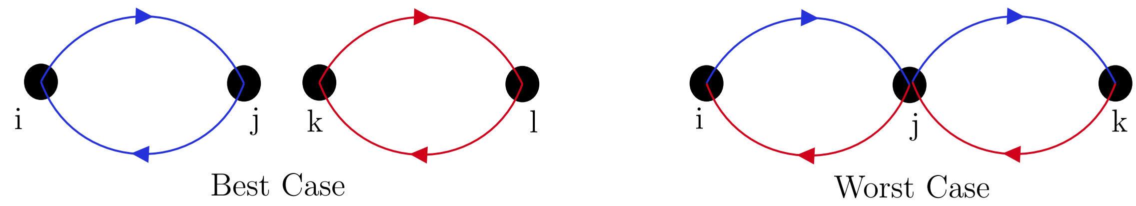

Figure 14: Potential G for $p = (2, 2)$. Red indicates the first walk. Blue indicates the second walk. In the best case (highest order), $S_G = n^4 p_n^2$. In the worst case (lowest order), $S_G = n^3 p_n^2$.

to be equal to an existing edge since there are now fewer unique ξ_{ij}'s. Having completed the first walk, start the second walk. If the first edge of the second walk increments the multiplicity of an edge from 0 to 1, it is free. However, incrementing multiplicity from a to $a+1$ for $a \geq 1$ costs $n^2 p_n$. This is because placing the first edge of a new walk corresponds the two restrictions: $k_j = k_{j'}$, $k_{j+1} = k_{j'+1}$. However, the moments decrease only by p_n since we only lose one unique edge. Continue in the following way until all walks are completed. At the end of the walks, suppose cost is $n^{\alpha} p_n^{\beta}$. If every edge in G has multiplicity at least 2, $S_G = n^{2t+2-\alpha} p_n^{2t-\beta}$ by construction. Otherwise, $S_G = 0$.

Suppose $|\mathcal{J}| = 2\tau$ for some $t \geq 1$. Then at least a edges have multiplicity 2. The minimum cost of such a graph is $n^{\tau} p_n^{\tau}$, so that $\beta \geq \tau$. Note also that each edge costs weakly more n than p_n. As such, $\alpha \geq \beta$ so that $2t + 2 - \alpha - (2t - \beta) \leq 2$. Taking expectations over U, which preserves the order of the terms and then taking square root gives us the order of $\iota_n' B \iota_n$.

Tightened Bounds

Finally, we discuss when the bounds can be tightened by $\frac{1}{\sqrt{n}}$.

Note that given our discussion on costs, we know that the ideal least costly graph have edges of multiplicity exactly 2. Furthermore, all multiplicities of a given edge must belong to the same walk. In particular, the best case is attained only if $p_1, ..., p_r$ are all even. On the other hand, the worst case cost is $n^{2a} p_n^{a}$, which is attained if the second edge are all the initial edges of a new path. For an example, see Figure 14.

Violation of the above "optimality" conditions will result in $\alpha \geq \beta + 1$. This is sufficient to yielding the $1/\sqrt{n}$ improvement. As such, if at least one of $p_1, ..., p_r$ is odd,

$$\iota_n' B \iota_n = O_p\left(\frac{1}{\sqrt{n}}\right) \cdot O_p\left(n^{t+1-\tau/2} p_n^{t-\tau/2}\right) \ .$$

Next, suppose that $p_1, ...p_r$ are all even. Write

$$
\begin{aligned}
&(\iota_n' B \iota_n - E[\iota_n' B \iota_n \,|\, U])^2 \\
&= \sum_{k_1=1}^n \sum_{k_2=1}^n \cdots \sum_{k_{2t+2}=1}^n (b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_t+1} - E[b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_t+1} \,|\, U]) \\
&\qquad\qquad\qquad \cdot \left(b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \cdots b_{k_{2t+1},k_{2t+2}} - E\left[b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \,|\, U\right]\right) \\
&= \sum_G \sum_{k \in r_G} (b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_t+1} - E[b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_t+1} \,|\, U]) \\
&\qquad\qquad \cdot \left(b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \cdots b_{k_{2t+1},k_{2t+2}} - E\left[b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \,|\, U\right]\right)
\end{aligned}
$$

In the above display, we are summing over all G. However, as before, the set of relevant G can be substantially restricted. Define

$$
\begin{aligned}
S_G' := E\Bigg[&\sum_{k \in r_G} (b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_t+1} - E[b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_t+1} \,|\, U]) \\
&\qquad \cdot \left(b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \cdots b_{k_{2t+1},k_{2t+2}} - E\left[b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \,|\, U\right]\right) \,|\, U \Bigg]
\end{aligned}
$$

Note that $S_G = 0 \Rightarrow S_G' = 0$. This is because $S_G = 0$ only if there G has at least one edge with multiplicity exactly 1, which will also set $S_G' = 0$.

We now show that if G is optimal, then $S_G' = 0$. Suppose G attains the optimal rate. Then every edge has multiplicity exactly 2 formed from the same walk. For such a G, $b_{k_i,k_{i+1}}$ for $(i, i+1)$ in the first walk is independent from $b_{k_{i'},k_{i'+1}}$ where $(i', i'+1)$ is in the second walk. As such,

$$
\begin{aligned}
S_G' := \sum_{k \in r_G} &E\left[(b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_t+1} - E[b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_t+1} \,|\, U])\right] \\
&\cdot E\left[\left(b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \cdots b_{k_{2t+1},k_{2t+2}} - E\left[b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \,|\, U\right]\right) \,|\, U\right] = 0.
\end{aligned}
$$

Next, note that by the Cauchy-Schwarz inequality, that for any G', $S_{G'}' \prec S_{G'}$. Let G' be a suboptimal graph. By our study on costs of walks, $S_{G'} \prec \frac{1}{\sqrt{n}} S_G$. Now let $\mathbf{G} = \{G \,|\, S_G \neq 0\}$. Conclude that

$$
E\left[(\iota_n' B \iota_n - E[\iota_n' B \iota_n \,|\, U])^2 \,|\, U\right] = \sum_{G \in \mathbf{G}} S_G' = O_p\left(\frac{1}{\sqrt{n}}\right) O_p(S_G) = O_p\left(\frac{1}{\sqrt{n}}\right) \cdot O_p\left(n^{t+1-\tau/2} p_n^{t-\tau/2}\right) .
$$

A.5.3 Proof of Lemma 3

The proof of this lemma is based on the "Neumann trick" (see for instance, Eldridge et al. 2018, or Theorem 2 of Chen et al. 2021). We use the formulation by Cheng et al. (2021). By their Lemma 1, we have that

$$\frac{\lambda_1(\hat{A})}{\lambda_1(A)\left(v_1(A)'v_1(\hat{A})\right)}w'v_1(\hat{A}) = w'v_1(A) + \underbrace{\frac{w'\boldsymbol{\xi}v_1(A)}{\lambda_1(\hat{A})}}_{=:\Lambda_1} + \frac{w'\boldsymbol{\xi}^2 v_1(A)}{\left(\lambda_1(\hat{A})\right)^2}$$
$$+ \underbrace{\sum_{t=3}^{\infty}\frac{w'\boldsymbol{\xi}^t v_1(A)}{\left(\lambda_1(\hat{A})\right)^t}}_{=:\Lambda_1} + \underbrace{\sum_{r=2}^{R}\frac{\lambda_r(A)}{\lambda_1(A)}\frac{v_r(A)'v_1(\hat{A})}{v_r(A)'v_r(\hat{A})}\left\{\sum_{t=0}^{\infty}\frac{w'\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t}\right\}}_{=:\Lambda_3} . \tag{35}$$

where we used the fact that f is rank R. In the remainder of the proof, we bound Λ_1, Λ_2 and Λ_3 for $w \in \{v_1(A), \varepsilon/a_n\}$.

Bounds for $v_1(A)$

Suppose $w = v_1(A)$. We start by bounding Λ_1. For a given $\nu \in (0,1)$, choose T such that $T(1-\nu) > 4 + 2/\eta$. Then,

$$(np_n)^{-(T(1-\nu)-4)}\left(\frac{\log n}{\log\log n}\right)^{T(1-\nu)/2} \to 0 . \tag{36}$$

This is because the above condition is equivalent to

$$p_n \succ n^{-1}\left(\frac{\log n}{\log\log n}\right)^{\frac{1}{2}+\frac{2}{T(1-\nu)-4}} ,$$

which follows by our choice of T since p_n satisfies Equation (7). Observe that by Weyl's Inequality (e.g. Theorem 4.5.3 in Vershynin 2018),

$$\left\|\lambda_r(\hat{A}) - \lambda_r(A)\right\| \leq \|\boldsymbol{\xi}\| = o_p\left(np_n\right) , \tag{37}$$

the rate estimate follows from Lemma 1. Next, note that $\frac{1}{p_n}A$ is a weighted graph obtained by sampling U on the dense graphon f. As such, by Lemma 10.16 of Lovász (2012), $\frac{\lambda_r(A)}{np_n} = \lambda_r\left(\frac{1}{p_n}A\right)/n \xrightarrow{p} \tilde{\lambda}$. In other words, w.p.a. 1, we have that $\lambda_r(\hat{A}) \geq \tilde{\lambda}np_n/2 > 0$.

Next write

$$
\begin{aligned}
\left|\sum_{t=T+1}^{\infty} \frac{w'\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t}\right| &\leq \sum_{t=T+1}^{\infty} \|w\| \cdot \left(\frac{\|\boldsymbol{\xi}\|}{\tilde{\lambda}_r n p_n/2}\right)^t \cdot \|v_r(A)\| \quad \text{w.p.a. } 1 \\
&= O_p\left(\|w\| \left(\sqrt{\frac{\log n}{\log\log n}} \Big/ (np_n)\right)^{T(1-\nu)/2}\right) \text{ by Lemma 1} \\
&= O_p\left(\frac{\|w\|}{(np_n)^2}\right) \text{ by Equation (36).} \qquad (38)
\end{aligned}
$$

Meanwhile,

$$
E\left[\left|\sum_{t=2}^{T} \frac{w'\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t}\right|\right] \leq E\left[\left|\sum_{t=2}^{T} \frac{w'\boldsymbol{\xi}^t v_r(A)}{\left(\tilde{\lambda}_r n p_n/2\right)^t}\right|\right] \leq \sum_{t=2}^{T} E\left[\frac{\left(w'\boldsymbol{\xi}^t v_r(A)\right)^2}{\left(\tilde{\lambda}_r n p_n/2\right)^{2t}}\right]^{1/2} \qquad (39)
$$

where the last inequality above follows by an application of the triangle and Cauchy-Schwarz inequalities. Since T is finite, it suffices to bound each term individually. The next part is similar to the arguments in the proof Lemma 2.

Next note that if v is an eigenvector of A with eigenvalue λ, it must satisfy:

$$
\begin{aligned}
\lambda v = Mv &= \sum_{r=1}^{R} \tilde{\lambda}_r \frac{\phi_r(U)}{\sqrt{n}} \frac{\phi_r(U)'}{\sqrt{n}} v \\
&= \sum_{r=1}^{R} \tilde{\lambda}_r v_r \frac{\phi_r(U)}{\sqrt{n}} \ .
\end{aligned}
$$

Hence, v is a linear combination of $\phi_r(U)/\sqrt{n}$'s. By convergence of the spectrum, we know that $\limsup \lambda \leq \tilde{\lambda}_1$. Now, let Υ be the event that $|\frac{1}{\sqrt{n}} \sum_{i=1}^{r} \phi_r(U_i)\phi_s(U_i) < 1/R^2$. This happens with probability approaching 1 since R is finite. On this event, $\|v\| = 1$ implies that $|v_r| < 2$ for all r. Furthermore, observe that since $\|f\|_\infty \leq 1$, $\|\phi_r\|_\infty \leq 1$. As such, on Υ, $|v_r(A)\|_\infty \leq 2R/\sqrt{n}$.

Now, for a $U \in \Upsilon$,

$$
\begin{aligned}
&E\left[\left(v_1(A)'\boldsymbol{\xi} v_r(A)\right)^2 \,\middle|\, U\right] \\
&= \frac{1}{n^2}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\cdots\sum_{k_{2t+2}=1}^{n}\Big\{\left[v_1(A)\right]_{k_1}\left[v_1(A)\right]_{k_{t+1}}\left[v_r(A)\right]_{k_{t+1}}\left[v_r(A)\right]_{k_{2t+2}}\cdot \\
&\qquad\qquad E\left[\xi_{k_1,k_2}\xi_{k_2,k_3}\cdots\xi_{k_t,k_t+1}\cdot\xi_{k_{t+2},k_{t+3}}\xi_{k_{t+3},k_{t+4}}\cdots\xi_{k_{2t+1},k_{2t+2}}\,\middle|\,U\right]\Big\} \\
&\le \frac{16R^4}{n^2}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\cdots\sum_{k_{2t+2}=1}^{n}E\left[\xi_{k_1,k_2}\xi_{k_2,k_3}\cdots\xi_{k_t,k_t+1}\cdot\xi_{k_{t+2},k_{t+3}}\xi_{k_{t+3},k_{t+4}}\cdots\xi_{k_{2t+1},k_{2t+2}}\,\middle|\,U\right]
\end{aligned}
$$

where the final inequality follows from our bound on $\|v_r(A)\|_\infty$ and the fact that for all $t \le T$,

$$
E\left[\xi_{ij}^t\,\middle|\,U\right] = p_n f(U_i,U_j)\left(1 - p_n f(U_i,U_j)\right)\cdots\left(1 - t\cdot p_n f(U_i,U_j)\right) \ge 0 \text{ if } p_n \le 1/T\ .
$$

By Lemma 2,

$$
\frac{1}{n^2}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\cdots\sum_{k_{2t+2}=1}^{n}E\left[\xi_{k_1,k_2}\xi_{k_2,k_3}\cdots\xi_{k_t,k_t+1}\cdot\xi_{k_{t+2},k_{t+3}}\xi_{k_{t+3},k_{t+4}}\cdots\xi_{k_{2t+1},k_{2t+2}}\right] = \frac{1}{n^2}O\left(n^{t+2}p_n^t\right)\ .
$$

As such, for $t \ge 2$,

$$
E\left[\frac{\left(v_1(A)'\boldsymbol{\xi}^t v_r(A)\right)^2}{\left(\tilde{\lambda}_r n p_n/2\right)^{2t}}\right]^{1/2} = O\left(\frac{1}{\left(\sqrt{np_n}\right)^t}\right)\ . \tag{40}
$$

Next, suppose $t = 0$. Then $v_1(A)'v_r(A) = 0$ since $r \ne 1$. Suppose $t = 1$.

$$
v_1(A)'\boldsymbol{\xi} v_r(A) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[v_1(A)\right]_i\xi_{ij}\left[v_1(A)\right]_j
$$

As such,

$$
\begin{aligned}
E\left[\left(v_1(A)'\boldsymbol{\xi} v_r(A)\right)^2 \,\middle|\, U\right] &\le \frac{16R^4}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}p_n f(U_i,U_j)\left(1 - p_n f(U_i,U_j)\right) \\
&= O_p\left(p_n\right)\ .
\end{aligned}
$$

Together with the fact that $P(\Upsilon) \to 1$, this yields

$$\frac{v_1(A)'\boldsymbol{\xi} v_r(A)}{\tilde{\lambda}_r n p_n/2} = O_p\left(\frac{1}{n\sqrt{p_n}}\right) \ . \tag{41}$$

Noting that $(v_1(A))' \, v_r(A) = 0$, Equations (38), (40) and (41) yield

$$\sum_{t=0}^{\infty} \frac{w'\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t} = O_p\left(\frac{1}{np_n}\right) \quad , \quad \sum_{t=3}^{\infty} \frac{w'\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t} = o_p\left(\frac{1}{np_n}\right) \tag{42}$$

Substituting Equations and (42) into (35) yields the desired bound for $v_1(A)$.

Bounds for ε/a_n

Next, suppose $w = \varepsilon/a_n$. The proof follows that for $v_1(A)$ up to Equation (39). To proceed, recall that conditional on U, $\boldsymbol{\xi} \perp\!\!\!\perp \varepsilon$. Conditioning again on the evnt Υ,

$$\begin{aligned}
&E\left[\left(\left(\frac{\varepsilon}{a_n}\right)' \boldsymbol{\xi}^t v_r(A)\right)^2 \,\middle|\, U, \boldsymbol{\xi}\right] \\
&= \frac{1}{n \cdot a_n^2} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_{2t+2}=1}^{n} \Big\{ E\left[\varepsilon_{k_1}\varepsilon_{k_{t+2}} \,|\, U, \boldsymbol{\xi}\right] \left[v_r(A)\right]_{k_{t+1}} \left[v_r(A)\right]_{k_{2t+2}} \cdot \\
&\qquad\qquad \xi_{k_1,k_2}\xi_{k_2,k_3} \cdots \xi_{k_t,k_t+1} \cdot \xi_{k_{t+2},k_{t+3}}\xi_{k_{t+3},k_{t+4}} \cdots \xi_{k_{2t+1},k_{2t+2}} \Big\} \\
&\le \frac{4R^2}{n \cdot a_n^2} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_{2t+2}=1}^{n} E\left[\varepsilon_{k_1}\varepsilon_{k_{t+2}} \,|\, U, \boldsymbol{\xi}\right] \xi_{k_1,k_2}\xi_{k_2,k_3} \cdots \xi_{k_t,k_t+1} \cdot \xi_{k_{t+2},k_{t+3}}\xi_{k_{t+3},k_{t+4}} \cdots \xi_{k_{2t+1},k_{2t+2}}
\end{aligned}$$

Recall that conditional on U, $\boldsymbol{\xi} \perp\!\!\!\perp \varepsilon$. If $k_1 \neq k_{t+2}$, we can write:

$$\begin{aligned}
E\left[\varepsilon_{k_1}\varepsilon_{k_{t+2}} \,|\, U, \boldsymbol{\xi}\right] &= E\left[E\left[\varepsilon_{k_1} \,|\, \varepsilon_{k_2}, U, \boldsymbol{\xi}\right] \varepsilon_{k_{t+2}} \,|\, U, \boldsymbol{\xi}\right] \\
&= E\left[E\left[\varepsilon_{k_1} \,|\, U_{k_1}\right] \varepsilon_{k_{t+2}} \,|\, U, \boldsymbol{\xi}\right] \\
&= E\left[\varepsilon_{k_1} \,|\, U_{k_1}\right] E\left[\varepsilon_{k_{t+2}} \,|\, U_{k_{t+2}}\right] = 0 \ .
\end{aligned}$$

Hence, we only need to consider sequences where $k_{t+2} = k_1$, so that

$$E\left[\left(\left(\frac{\varepsilon}{a_n}\right)' \boldsymbol{\xi}^t v_r(A)\right)^2 \big| U, \boldsymbol{\xi}\right]$$
$$\leq \frac{\bar{\sigma}^2}{n \cdot a_n} \sum_{k_1=1}^{n} \cdots \sum_{k_{t+1}=1}^{n} \sum_{k_{t+3}=1}^{n} \cdots \sum_{k_{2t+2}=1}^{n} \xi_{k_1,k_2}\xi_{k_2,k_3} \cdots \xi_{k_t,k_t+1} \cdot \xi_{k_1,k_{t+3}}\xi_{k_{t+3},k_{t+4}} \cdots \xi_{k_{2t+1},k_{2t+2}}$$
$$= \frac{\bar{\sigma}^2}{n \cdot a_n} \iota \boldsymbol{\xi}^{2t+1} \iota$$

Taking expectations over U and $\boldsymbol{\xi}$, we have that

$$E\left[\left(\left(\frac{\varepsilon}{a_n}\right)' \boldsymbol{\xi}^t v_r(A)\right)^2\right] = \frac{\bar{\sigma}^2}{n \cdot a_n^2} O\left(n^{t+1} p_n^{t+1/2}\right)$$

where the rate estimates again follow from Lemma 2. The order on n is smaller than the "best case" by $n^{1/2}$ due to the fact that $2t+1$ is odd, so that at least one edge will not be optimally paired. As such, for $t \geq 1$,

$$E\left[\frac{\left((\varepsilon/a_n)' \boldsymbol{\xi}^t v_r(A)\right)^2}{\left(\tilde{\lambda}_r n p_n/2\right)^{2t}}\right]^{1/2} = O\left(\frac{p_n^{1/4}}{a_n \left(\sqrt{n p_n}\right)^t}\right) = o\left(\frac{1}{a_n}\right) . \tag{43}$$

Note that when $t = 1$, the above bound implies that $\Lambda_1 = o_p\left((a_n)^{-1}\right)$ for $w = \varepsilon/a_n$. If $t = 0$,

$$\left(\frac{\varepsilon}{a_n}\right)' v_r(A) = \frac{1}{a_n \sqrt{n}} \sum_{i=1}^{n} \phi_r(U_i)\varepsilon_i = O_p\left(\frac{1}{a_n}\right) \tag{44}$$

Combining our last two estimates, we have that

$$\sum_{t=0}^{T} \frac{(\varepsilon/a_n)' \boldsymbol{\xi}^t v_r(A)}{\left(\tilde{\lambda}_r n p_n/2\right)^t} = O_p\left(\frac{1}{a_n}\right) \tag{45}$$

Let $\check{\Upsilon}$ be the event that $\lambda_r(\hat{A}) \geq \tilde{\lambda}_r n p_n/2$ and

$$\|\boldsymbol{\xi}\| \leq \sqrt{n p_n} \left(\frac{k \log n}{\log \log n}\right)^{1/4} ,$$

where k is the constant in Lemma 1. Then $P(\check{\Upsilon}) \to 1$ by Lemma 1 and Equation (37). Furthermore,

$$P\left(\left|\sum_{t=T+1}^{\infty}\frac{\varepsilon'\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t}\right| \leq x \,\middle|\, U, \boldsymbol{\xi}, \check{\Upsilon}\right) \leq \frac{1}{x^2}E\left[\left(\sum_{t=T+1}^{\infty}\frac{\varepsilon'\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t}\right)^2 \,\middle|\, U, \boldsymbol{\xi}, \check{\Upsilon}\right]$$

$$\leq \frac{1}{x^2}\bar{\sigma}^2\left\|\sum_{t=T+1}^{\infty}\frac{\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t}\right\|^2 \leq \sum_{t=T+1}^{\infty}\|\boldsymbol{\xi}\|^{2t}$$

$$\leq \frac{\tilde{k}}{np_n} \text{ by Equation (36), on the event } \check{\Upsilon}.$$

The bound on the right hand side does not depend on U and $\boldsymbol{\xi}$ once we condition on $\check{\Upsilon}$. Hence,

$$P\left(\left|\sum_{t=T+1}^{\infty}\frac{\varepsilon'\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t}\right| \leq x\right) \leq P\left(\check{\Upsilon}\right)P\left(\left|\sum_{t=T+1}^{\infty}\frac{\varepsilon'\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t}\right| \leq x \,\middle|\, \check{\Upsilon}\right) + 1 - P\left(\check{\Upsilon}\right)$$

$$\leq \frac{\tilde{k}}{np_n} + 1 - P\left(\check{\Upsilon}\right) \to 0\ .$$

Hence, we conclude that

$$\sum_{t=T+1}^{\infty}\frac{(\varepsilon/a_n)'\,\boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t} = \frac{1}{a_n}o_p(1) = o_p\left(\frac{1}{a_n}\right)\ . \tag{46}$$

Next, note that

$$\begin{aligned}\left|\frac{1}{\lambda_s(A)}v_r(A)'v_s(\hat{A})\right| &\leq \left|\frac{1}{\lambda_s(A)}v_r(A)'v_s(A)\right| + \|v_r(A)\|\cdot\left\|\frac{v_s(\hat{A})-\theta v_s(A)}{2\tilde{\lambda}_s np_n}\right\| + o_p(1)\\ &= \left|\frac{1}{\lambda_s(A)}v_r(A)'v_s(A)\right| + o_p(1)\ ,\end{aligned} \tag{47}$$

where the last equation follows because by the Davis-Kahan Inequality (Theorem 4.5.5 in Vershynin (2018)),

$$\|v_s(\hat{A})-\theta v_s(A)\| \leq \frac{\left\|\hat{A}-A\right\|}{\Delta_{\min}\cdot np_n} = o_p(1) \quad \text{by Lemma 1.}$$

As before, since $v_r(A)'v_1(A) = 0$ for all $r \geq 2$, we can write Λ_3

$$\sum_{r=2}^{R} \underbrace{\frac{\lambda_r(A)}{\lambda_1(A)} \frac{v_r(A)'v_1(\hat{A})}{v_r(A)'v_r(\hat{A})}}_{= o_p(1) \text{ by Eq. (47)}} \left\{ \underbrace{\sum_{t=0}^{T} \frac{(\varepsilon/a_n)' \boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t}}_{= O_p\left(a_n^{-1}\right) \text{ by Eq. (45)}} + \underbrace{\sum_{t=T+1}^{\infty} \frac{(\varepsilon/a_n)' \boldsymbol{\xi}^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t}}_{= o_p\left(a_n^{-1}\right) \text{ by Eq. (46)}} \right\} = o_p\left(\frac{1}{a_n}\right) . \tag{48}$$

Finally, we note that by arguments identical to the above,

$$\sum_{t=1}^{\infty} \frac{(\varepsilon/a_n)' \boldsymbol{\xi}^t v_1(A)}{\left(\lambda_1(\hat{A})\right)^t} = \underbrace{\sum_{t=1}^{T} \frac{(\varepsilon/a_n)' \boldsymbol{\xi}^t v_1(A)}{\left(\lambda_1(\hat{A})\right)^t}}_{= o_p\left(a_n^{-1}\right) \text{ by Eq. (43)}} + \underbrace{\sum_{t=T+1}^{\infty} \frac{w' \boldsymbol{\xi}^t v_1(A)}{\left(\lambda_1(\hat{A})\right)^t}}_{= o_p\left(a_n^{-1}\right) \text{ by Eq. (46)}} = o_p\left(\frac{1}{a_n}\right) . \tag{49}$$

We conclude by remarking that from Equation (47),

$$\frac{\lambda_1(\hat{A})}{\lambda_1(A)\left(v_1(A)'v_1(\hat{A})\right)} = \frac{\lambda_1(A)}{\lambda_1(A)\left(v_1(A)'v_1(A)\right)} + o_p(1) ,$$

so that the LHS of Equation (35) also converges in probability to $w'v_1(\hat{A})$.

B Bias of Diffusion under noise ($\hat{\beta}^{(T)}$)

Tables 9 and 10 presents $b_T(t,\delta)$ used for calculating the bias estimator in Case (b) of Theorem 4. In practice, papers rarely compute $T > 5$. We provide these terms for $T \leq 10$. Functions for computing the bias terms for arbitrary T are available from the author's website.

Each row of Tables 9 and 10 provide the coefficients for δ^s in $b_T(t,\delta)$, for a particular T and t. To obtain the bias formula for a given T, sum across all t's for a given T. For example, when $T = 2$, the correction term is

$$\left(\delta^2 - 3\delta^3 + 3\delta^4\right) \iota_n' \hat{A} \iota_n + \left(3\delta^3 - 2\delta^4\right) \iota_n' \hat{A}^2 \iota_n + \left(2\delta^4\right) \iota_n' \hat{A}^3 \iota_n \ .$$

T	t	δ^2	δ^3	δ^4	δ^5	δ^6	δ^7	δ^8	δ^9	δ^{10}
1	1	1								
	1	1	-3	3						
2	2		3	-2						
	3			2						
	1	1	-3	7	-4	-8				
	2		3	-4		10				
3	3			5	-2	-2				
	4				4	-1				
	5					2				
	1	1	-3	7	-13	-15	91	-182		
	2		3	-4	5	24	-94	160		
	3			5	-7	-1	36	-84		
4	4				8	-6	-2	27		
	5					7	-5			
	6						5	-4		
	7							3		
	1	1	-3	7	-13	-4	161	-500	952	-654
	2		3	-4	5	24	-178	450	-740	314
	3			5	-7	11	57	-222	456	-362
	4				8	-15	6	66	-225	317
5	5					12	-14	4	59	-148
	6						11	-13	6	32
	7							9	-12	6
	8								7	-8
	9									4

Table 9: Coefficients for the bias of diffusion centrality, $b_T(t,\delta)$. Blanks indicate a coefficient of 0.

T	t	δ^2	δ^3	δ^4	δ^5	δ^6	δ^7	δ^8	δ^9	δ^{10}	δ^{11}	δ^{12}	δ^{13}	δ^{14}	δ^{15}	δ^{16}	δ^{17}	δ^{18}	δ^{19}	δ^{20}
	1	1	-3	7	-13	-4	184	-819	1869	-1935	-3737	15981								
	2		3	-4	5	24	-229	770	-1496	954	4740	-14604								
	3			5	-7	11	52	-346	869	-1053	-1526	7728								
	4				8	-15	28	81	-406	915	-674	-1883								
	5					12	-28	22	99	-432	814	-312								
6	6						17	-27	19	94	-359	485								
	7							16	-26	21	67	-209								
	8								14	-25	21	35								
	9									12	-21	15								
	10										9	-13								
	11											5								
	1	1	-3	7	-13	-4	184	-1097	3043	-3843	-6915	49578	-115459	80767						
	2		3	-4	5	24	-229	1088	-2552	2198	9142	-45912	90964	-37634						
	3			5	-7	11	52	-431	1363	-2000	-2751	23692	-59966	51305						
	4				8	-15	28	63	-565	1565	-1556	-5348	25674	-39288						
	5					12	-28	59	102	-692	1717	-1266	-5929	18469						
	6						17	-47	52	130	-739	1625	-718	-4616						
7	7							23	-46	48	127	-669	1263	-272						
	8								22	-45	50	100	-516	758						
	9									20	-44	50	68	-306						
	10										18	-40	44	33						
	11											15	-32	29						
	12												11	-19						
	13													6						
	1	1	-3	7	-13	-4	184	-1097	4525	-7293	-6944	83485	-256225	310955	520469	-2445004				
	2		3	-4	5	24	-229	1088	-3969	4898	11110	-79100	207636	-174066	-652454	2204430				
	3			5	-7	11	52	-431	1963	-3592	-2432	39251	-130367	185191	177779	-1142919				
	4				8	-15	28	63	-696	2366	-3194	-7525	52268	-127455	81923	320735				
	5					12	-28	59	58	-912	2822	-3094	-10153	55373	-107794	30033				
	6						17	-47	110	108	-1096	3070	-2701	-12050	51958	-73965				
	7							23	-73	102	146	-1166	2990	-1896	-11737	35122				
8	8								30	-72	97	145	-1099	2591	-1255	-7492				
	9									29	-71	99	118	-943	2064	-940				
	10										27	-70	99	86	-732	1305				
	11											25	-66	93	51	-426				
	12												22	-58	78	18				
	13													18	-45	49				
	14														13	-26				
	15															7				
	1	1	-3	7	-13	-4	184	-1097	4525	-12637	-1360	110968	-420423	695532	681922	-7068240	17664266	-13712687		
	2		3	-4	5	24	-229	1088	-3969	9368	8424	-109118	350830	-448472	-1023716	6464948	-13721396	6724528		
	3			5	-7	11	52	-431	1963	-6002	474	51064	-209648	389396	179724	-3281313	9018463	-8293240		
	4				8	-15	28	63	-696	3367	-5914	-6669	77446	-236871	246153	850584	-4132503	6454139		
	5					12	-28	59	58	-1079	4123	-6290	-11274	91953	-236476	139244	993350	-2918136		
	6						17	-47	110	18	-1366	4803	-6352	-15294	101721	-231751	105547	672714		
	7							23	-73	188	79	-1614	5172	-5880	-18707	102925	-204469	74412		
	8								30	-107	179	128	-1709	5108	-4789	-19615	88499	-132335		
9	9									38	-106	173	129	-1645	4670	-3937	-15825	53896		
	10										37	-105	175	102	-1486	4121	-3585	-8366		
	11											35	-104	175	70	-1274	3362	-2653		
	12												33	-100	169	35	-968	2058		
	13													30	-92	154	2	-542		
	14														26	-79	125	-16		
	15															21	-60	76		
	16																15	-34		
	17																	8		
	1	1	-3	7	-13	-4	184	-1097	4525	-12637	10890	125858	-597596	1235468	309432	-11044097	36311072	-45046667	-75501514	347810028
	2		3	-4	5	24	-229	1088	-3969	9368	211	-130730	514002	-866532	-979926	10410814	-28739014	23733098	95980642	-310601986
	3			5	-7	11	52	-431	1963	-6002	6504	56591	-292720	663931	-81390	-5051600	18395415	-26721605	-26472010	165085728
	4				8	-15	28	63	-696	3367	-9914	-1678	97369	-363531	535304	1071910	-8050856	19840070	-13232554	-47818679
	5					12	-28	59	58	-1079	5644	-11058	-7738	125398	-398107	394043	1745080	-8688965	16108742	-2368523
	6						17	-47	110	18	-1535	6719	-11878	-13503	148861	-422298	318022	1897993	-7599134	9772214
	7							23	-73	188	-84	-1901	7684	-12237	-18749	165197	-429048	278089	1667351	-4861956
	8								30	-107	301	-11	-2225	8199	-11710	-23857	171609	-407905	236540	1042204
	9									38	-150	291	50	-2347	8154	-10298	-26127	159555	-335522	149307
10	10										47	-149	284	53	-2286	7674	-9217	-22752	124908	-199857
	11											46	-148	286	26	-2124	7102	-8832	-15237	70335
	12												44	-147	286	-6	-1911	6343	-7913	-6786
	13													42	-143	280	-41	-1605	5040	-5264
	14														39	-135	265	-74	-1179	2982
	15															35	-122	236	-92	-637
	16																30	-103	187	-76
	17																	24	-77	111
	18																		17	-43
	19																			9

Table 10: Coefficients for the bias of diffusion centrality, $b_T(t,\delta)$ (continued from Table 9). Blanks indicate a coefficient of 0.

C Eigenvector Regularization

As our analysis in Section 3.1 shows, regression with eigenvector centrality is more sensitive to sparsity under measurement error than degree or diffusion centralities. In this section, we propose a regularization method that makes eigenvector centrality competitive with the alternatives.

Definition 11 (Regularized Eigenvector Centrality)**.** Suppose p_n is known. Let

$$\lambda_i := \min\left\{\frac{2np_n}{\hat{C}_i^{(1)}}\,,\,1\right\}\ .$$

Then, define $\hat{A}_\lambda$ to be the regularized version of $\hat{A}$, where

$$\left(\hat{A}_\lambda\right)_{ij} = \sqrt{\lambda_i\lambda_j}\hat{A}_{ij}\ .$$

Finally, define regularized eigenvector centrality and the corresponding OLS estimator to be:

$$\hat{C}_\lambda^{(\infty)} = a_n v_1\left(\hat{A}_\lambda\right) \quad , \quad \hat{\beta}_\lambda^{(\infty)} = \frac{Y'\hat{C}_\lambda^{(\infty)}}{\left(\hat{C}_\lambda^{(\infty)}\right)'\hat{C}_\lambda^{(\infty)}}\ .$$

Our proposed measure is the principal eigenvector of $\hat{A}_\lambda$, which is in turn a regularized version of the observed adjacency matrix $\hat{A}$. The regularization technique, proposed in Le et al. (2017), re-weights edges so that in $\hat{C}_{\lambda,i}^{(1)} \leq 2np_n$ for all $i \in [n]$. It is well-known that high-degree vertices interfere with concentration of random matrices and that their removal solves the problem (Feige and Ofek 2005). However, such a drastic procedure is not ideal: intuition suggests that high degree vertices are important in a network, forming hubs that connect many individuals. Le et al. (2017) shows that re-weighting the edges of high-degree vertices is sufficient to enforce concentration. In turn, we have consistency of $\hat{\beta}_\lambda^{(\infty)}$ as our next theorem asserts.

Theorem 6 (Consistency with Regularized Eigenvector Centrality)**.** Suppose Assumptions 1 and 2 hold. Suppose further that $E[\varepsilon_i|U_i] = 0$ and $E[\varepsilon_i^2] = \sigma^2 < \infty$ and $\lambda_1(f) > \lambda_2(f)$. Then, $a_n \to \infty$ and $p_n \succ n^{-1}$ implies that $\tilde{\beta}^{(\infty)} \xrightarrow{p} \beta^{(\infty)}$.

Our result shows that $\hat{\beta}_\lambda^{(\infty)}$ is able to accommodate as much sparsity as $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$. As such, when faced with sparse matrices, researchers could benefit from using regularized eigenvector centrality in their regression instead. One difficulty with using the method is that p_n is not known in practice. It is not possible to estimate p_n since the graphon f is unknown. Under a mild assumption, however, the following is possible:

Corollary 5 (Estimated p_n)**.** Suppose $\int_{[0,1]^2} f(u,v)\,dudv \geq M > 0$. Let

$$\rho_n = p_n \int_{[0,1]^2} f(u,v)\,dudv \quad , \quad \hat{\rho}_n = \frac{\iota_n' A \iota_n}{n(n-1)} \ .$$

Next define

$$\hat{\lambda}_i = \min\left\{ \frac{3n\hat{\rho}_n}{M \cdot \hat{C}_i^{(1)}} \, , \, 1 \right\} \ .$$

Using $\hat{\lambda}$ in place of λ in Definition 11 does not change the conclusions of Theorem 6.

C.1 Proof of Theorem 6

As in the Proof of Theorem 2, write

$$\hat{\beta}_{\lambda}^{(\infty)} = \beta^{(\infty)} + \beta^{(\infty)} \left(v_1(\hat{A}_\lambda)\right)' \left(v_1(A) - v_1(\hat{A}_\lambda)\right) + \frac{1}{\sqrt{a_n}} v_1(\hat{A}_\lambda)'\varepsilon \ .$$

By Theorem and Remark 2.2 of Le et al. (2017), with probability at least $1 - n^{-r}$,

$$\left\| A - \hat{A} \right\| \leq k r^{3/2} \sqrt{n p_n}$$

where k is a universal constant. Therefore, by the Davis-Kahan inequality (Theorem 4.5.5 in Vershynin 2018),

$$\left\| v_1(A) - v_1(\hat{A}) \right\| \leq \frac{\|\hat{A} - A\|}{n p_n \left(\lambda_1 - \lambda_2\right)} = O_p\left(\frac{1}{\sqrt{n p_n}}\right) = o_p(1) \ .$$

Again, note that

$$E\left[v_1(\hat{A})'\varepsilon \,\middle|\, U \right] \leq \|v_1(\hat{A})\| \bar{\sigma}^2 = \bar{\sigma}^2.$$

Since $a_n \to \infty$, conclude that $\frac{1}{\sqrt{a_n}} v_1(\hat{A})'\varepsilon \xrightarrow{p} 0$ and that $\hat{\beta}_{\lambda}^{(\infty)} \xrightarrow{p} \beta^{(\infty)}$.

C.2 Proof of Corollary 5

We first note that $\hat{\rho}$ is a good estimator of ρ_n:

Theorem 7 (Theorem 1, Bickel et al. 2011)**.** Under Assumption 1 and 2,

$$\sqrt{n}\left(\frac{\hat{\rho}_n}{\rho_n} - 1\right) \xrightarrow{d} N(0, \sigma^2)$$

for some $\sigma^2 > 0$.

Next, noting that $M/\int f(u,v) \leq 1$

$$P\left(\frac{\hat{\rho}_n}{M} \geq \frac{2Mp_n}{3\int f(u,v)\,,\,dudv}\right) \to 1$$

Setting

$$\hat{\lambda}_i = \min\left\{\frac{3n\hat{\rho}_n}{M \cdot C_i^{(1)}}\,,\,1\right\}$$

ensures that w.p.a. 1,

$$\max_{i\in[n]} C_{\lambda,i}^{(1)} \leq \frac{3n\hat{\rho}_n}{M} \leq \frac{2\int f(u,v)\,dudv}{M} \cdot np_n\ .$$

By Remark 2.1 of Le et al. (2017), the oracle procedure re-weights edges adjacent to fewer than $10/(np_n)$ nodes. Since $\hat{\lambda}_i \geq \lambda_i$, re-weighting using $\hat{\lambda}$ therefore also alters edges adjacent to fewer than $10/(np_n)$ nodes. As such, by Theorem 2.1 of $\|\hat{A}_{\hat{\lambda}} - A\| = O(\sqrt{np_n})$ w.p.a. 1. The proof then proceeds as in that of Theorem 6.