Linear Regression with Centrality Measures

Job Market Paper

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

This version: November 22, 2022.
Newest version here.

Abstract

This paper studies the properties of linear regression on centrality measures when net-
work 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 charac-
terize the threshold at which this occurs, with and without measurement error. This
threshold depends on the centrality measure used. Specifically, regression on eigenvec-
tor 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 es-
timators 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 infer-
ence. (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 study the
relationship between consumption smoothing and informal 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 Santiago Camara, 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 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 the 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. First, 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.

Second, 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 focuses 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. First, 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. Second, spectral norms of random matrices concentrate slowly in sparse regimes. To obtain our results, we bound the 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 the variance is of a lower order than the 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 6). 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.
Auerbach (2022) also considers linear regressions with network positions as explanatory variables. However, their approach is nonparametric and results are provided only in the dense case. 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 body of work that considers measurement error in centrality statistics. Early work provided simulation evidence that centrality measures on noisy networks become less accurate as sparsity increases (Costenbader and Valente (2003); Borgatti et al. (2006)). Segarra and Ribeiro (2015) theoretically studies the stability of network statistics under perturbations, finding that degree and eigenvector centralities are stable, while betweenness is not. Avella-Medina et al. (2020) and Dasaratha (2020) consider settings similar to ours with classical measurement error in the network data, but which translates to non-linear measurement error in centrality statistics. These authors provide concentration results for degree and eigenvector centralities among others, but not for diffusion centrality. Additionally, they accommodate less sparsity than us, in part because we are not concerned with estimation of centrality measures, only their use in subsequent regression.

A separate literature has focused on non-classical measurement error in the network data. 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. The author is concerned with forming bounds on centrality statistics and does not consider subsequent linear regression. Griffith (2022) considers 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 more general forms of measurement error in peer effects regression, finding that 2SLS with friends-of-friends instruments is valid as long as measurement error is small. All of the above papers 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 instance 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 or error-in-variable 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 Nyakatokote, Tanzania. Section 6 concludes the paper with our recommendations for empirical work. All proofs are contained in Appendix D.
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 $(i,j)^{th}$ 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 Hilbert-Schmidt integral operator, $T(g) = \int f(x,y)g(y)dy$, while $\phi_j$ 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 $x_n \succeq z_n$ to mean $\neg(x_n \prec z_n)$ and similarly for $x_n \preceq z_n$. Let $\iota_n$ 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 $(i,j)^{th}$, $A_{ij}$, records the link intensities between agents $i$ and $j$. $\hat{A}$ is some estimate of $A$.

Consider the regression:

$$Y_i = \beta^{(d)}C^{(d)}_i + \varepsilon^{(d)}_i$$

where $C^{(d)}_i$ is a network centrality measure of type $d$. While we do not observe $C^{(d)}_i$, 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 true 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 $C^{(d)}_i$.

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 the 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 OLS estimators when the network is sparse. We restrict 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, 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 $j > i$,

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

Set $\hat{A}_{ji} = \hat{A}_{ij}$ for $j < i$. $\hat{A}_{ii} = 0$ since $A_{ii} = 0$. Furthermore, suppose for $d \in \{1, T, \infty\}$ that

$$
\hat{A}_{ij} \perp \perp \varepsilon_{i}^{(d)} \mid U.
$$

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. Importantly, our model excludes strategic network formation for $\hat{A}$, in which agents’ decisions to form or report links depend on that of the others.

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. Nonetheless, the model is a reasonable description of network data. Measurement error of this form can arise due to limitations
in data collection, or because networks are constructed by proxy. Below, we discuss how Example 1 fits our measurement error model. More examples can be found in Appendix B, where we also discuss our econometric framework in the context of the “Weak Ties” theory of social networks (Granovetter 1973).

Finally, we assume that measurement error is independent of \( \varepsilon_i \) 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.

**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.

**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 when \( A \) is observed. 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:

\[
C^{(1)} = A_{tn} .
\]

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 $T \in \mathbb{N}$, 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 $(i, j)^{th}$ 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 $\delta$ per period. Bramoullé and Genicot (2018) provides further discussion on the theoretical foundations of diffusion centrality. In practice, researchers often choose $\delta$ 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:

$$C_i^{(\infty)} = k \sum_{j \neq i} A_{ij} C_j^{(\infty)} \text{ for all } i \in [n] .$$

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 can be chosen so that every entry is non-negative, motivating its use as a centrality measure. The leading eigenvector of related matrices also emerge as measures 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 will make the following assumption when analyzing 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 \( a_n \in \mathbb{R} \). 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 analyze the properties of regression on eigenvector centrality making explicit their dependence on \( a_n \). 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

- \( C^{(1)}_i \) as the probability-weighted number of friends who will lend to or borrow from \( i \).
- \( C^{(T)}_i \) 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. \( \delta \) 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.
- \( C^{(\infty)}_i \) 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.

When \( A \) is observed, we have access to the following infeasible estimators.
**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:

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

$$\hat{C}^{(T)} = \left( \sum_{t=1}^{T} \delta_t \hat{A}^t \right) \iota_n,$$

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

The corresponding OLS estimators are thus defined using the noisy centrality measures.

**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

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

Next, define the regression residuals.

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

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

$$\hat{\varepsilon}_i^{(d)} := Y_i - \hat{\beta}^{(d)} \hat{C}_i^{(d)}.$$ (2)

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

We conclude this section with an assumption on the moments of $\varepsilon_i$ conditional on $U_i$:

**Assumption 3.** Suppose for $d \in \{1, T, \infty\}$ that:

(a) $E \left[ \varepsilon_i^{(d)} | U_i \right] = 0$

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

13
(c) \( E \left[ \left| \varepsilon_i^{(d)} \right|^3 | U_i \right] \leq \bar{\kappa}_3. \)

In the above, (a) justifies linear regression since it implies that
\[
E \left[ \varepsilon_i^{(d)} | C_i^{(d)} \right] = 0.
\]

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 place such a restriction on 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 incorporating it into the asymptotic framework. Using $p_n$ to model sparse networks is standard in statistics literature (see e.g. Bickel and Chen 2009; Bickel et al. 2011, Borgs et al. 2018 Avella-Medina et al. 2020 among many more). Within econometrics, related approaches have been used to study network formation models (Jochmans 2018; Graham 2020b).

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 $\beta_1$ 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. The regimes for $p_n$ cannot be estimated from data. Instead, we provide a rule-of-thumb for gauging the reliability of OLS estimators. Section 3.2 presents distributional theory for $\hat{\beta}(d)$ 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 $\hat{\beta}(d)$ and $\hat{\beta}(d)$ are consistent. We also discuss the role of $a_n$ in ensuring the consistency of $\hat{\beta}(\infty)$ and $\hat{\beta}(\infty)$. Since rates on $p_n$ cannot be estimated, we present rules-of-thumbs for determining the amount of sparsity in practical applications.

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

(b) Suppose Assumption E1 also holds. Then, $\hat{\beta}(\infty) \xrightarrow{p} \beta(\infty)$ 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 $\varepsilon_i$ 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 $C_i^{(\infty)}$, 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 $\sigma_X$ is the variance of the regressor. Theorem 1 instantiate 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 $A = p_n t_n t_n'$. Then $C_i^{(\infty)}(A) = a_n t_n / \sqrt{n}$. Note that it is independent of $p_n$. We can then write:

$$\hat{\beta}(\infty) = \frac{\sqrt{n}}{a_n} \cdot \frac{Y' t_n}{t_n' t_n} = \beta(\infty) + \frac{1}{a_n \sqrt{n}} \sum_{i=1}^{n} \varepsilon_i^{(d)}.$$
Under our assumptions, \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i^{(d)} \xrightarrow{d} N\left(0, \text{Var}\left(\varepsilon_i^{(d)}\right)\right) \). Therefore, \( a_n \to \infty \) is necessary for the consistency of \( \hat{\beta}^{(\infty)} \).

The above example, together with Theorems 2 and 6 in the next section, makes clear that \( a_n \) has important implications for the statistical properties of \( \hat{\beta}^{(\infty)} \) 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 \( \beta \) but not \( \sqrt{n}\beta \).

To our knowledge, we are the first to emphasize the importance in choosing \( a_n \) appropriately. Various \( a_n \)'s are used in applied work and in econometric theory. Applied researchers sometimes set \( a_n = 1 \) (e.g. Cruz et al. 2017; Banerjee et al. 2013). Other times, they divide \( v_1(\hat{A}) \) by its standard deviation, in order to interpret \( \beta^{(\infty)} \) as the effect of a one standard-deviation increase in \( v_1(A) \) (e.g. Chandrasekhar et al. 2018; Banerjee et al. 2019). Corollary 5 shows that \( a_n \approx \sqrt{n} \) under some conditions on \( f \). Cai et al. (2021), which studies eigenvector regressions under a different model for measurement error, sets \( a_n = \sqrt{n} \). 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 \( \|\hat{\beta}^{(\infty)}C^{(\infty)} - \beta^{(\infty)}C^{(\infty)}\|_\infty \xrightarrow{p} 0 \). Of papers which study estimation of centrality statistics, Avella-Medina et al. (2020) sets \( a_n = \sqrt{n} \) while Dasaratha (2020) sets \( a_n = 1 \). We remark that changing \( a_n \) amounts to changing the definition of \( \hat{\beta}'^{(\infty)} \).

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 \( a_n \ll 1 \). And as we will see in Theorem 6, 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 if \( p_n \succ n^{-1} \).

(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

\[
n^{-1} (\log n)^4 \prec p_n \prec n^{-1} \sqrt{\frac{\log n}{\log \log n}}. \tag{5}
\]

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 eigenvalues and eigenvectors are sensitive to noise under sparsity. Our proof uses the homogeneous Erdos-Renyi graph as a counterexample. Suppose \( f = 1 \) so that \( A \) is rank 1 with \( v_1(A) = \iota/\sqrt{n} \). When the lower bound in Equation 5 is satisfied, Alt et al. (2021b) shows that \( \hat{A} \) has an eigenvalue, call it \( \nu \), with a corresponding eigenvector that is approximately \( \iota/\sqrt{n} \). If \( \nu \) is the largest eigenvalue of \( \hat{A} \), \( v_1(\hat{A}) \) is close to \( v_1(A) \). However, when \( p_n \) satisfies Equation 5, \( \nu \) turns out to be much smaller than \( \lambda_1(\hat{A}) \). The result is that \( v_1(\hat{A}) \) is almost orthogonal to \( v_1(A) \). Intuitively, sparsity weakens the signal in \( \hat{A} \), so that its leading eigenvector is pure noise.\(^1\) It then becomes impossible to estimate the leading eigenvector of \( A \) for OLS estimation. 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 \( p_n \sim n^{-1} \).

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 \( \alpha \), 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 \( p_n \sim n^{-3/2} \), \( \hat{\beta}(1) \) and \( \hat{\beta}(T) \) now require that \( p_n \sim n^{-1} \). 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 1. Avella-Medina et al. (2020) and Dasaratha (2020) provide results essentially showing that for \( d \in \{1, \infty\} \), \( \|C^d - C(d)\| \to 0 \) with probability approaching 1 if \( p_n \sim \frac{\log n}{n} \). Our focus is on the OLS estimators \( \tilde{\beta}(d) \) and \( \hat{\beta}(d) \) and we find that thresholds of consistency that are strictly below \( \frac{\log n}{n} \) for all \( d \in \{1, T, \infty\} \).

\(^1\)In fact, \( v_1(\hat{A}) \) exhibits localization. That is, its mass concentrates on the agent who happens to have the largest realized degree, which is purely a result of chance.
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.

**Remark 2.** Theorem 2 does not determine the behavior of $\hat{\beta}^{(\infty)}$ when $p_n < n^{-1} (\log \log n)^4$. Up to this threshold, we know by Alt et al. (2021b) that OLS is inconsistent only because we have descriptions of both eigenvalues and eigenvectors. To our knowledge, recent developments in random matrix theory do not provide any description of eigenvectors below this threshold. Hence, it is not clear what type of pathologies arises below $p_n < n^{-1} (\log \log n)^4$ 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.

**Remark 3.** 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 $p_n > n^{-1}$.

**Rule-of-Thumb for Determining Consistency of $\hat{\beta}^{(d)}$**

Theorems 2 provides consistency results for $\hat{\beta}^{(d)}$ based on the rates at which $p_n \to 0$. It is therefore desirable to have methods for determining if we are in the regime for which OLS is
consistent with measurement error. However, \( p_n \) is a theoretical device and the rate at which it is going to 0 is not a quantity that can be estimated. Instead, we propose a rule-of-thumb for determining the regime of \( p_n \) and the consistency of \( \hat{\beta}^{(d)} \).

**Definition 8** (Connectivity). Let \( \hat{A} \) be a binary network. Two agents \( i \) and \( j \) are connected if \( \hat{A} \) contains a path from \( i \) to \( j \). A collection of agents \( \mathcal{I} \) is connected if every pair in \( \mathcal{I} \times \mathcal{I} \) is connected. We say that \( \hat{A} \) is connected if \( [n] \) is connected.

**Definition 9** (Largest Component). For a binary network \( \hat{A} \), define

\[
L_1 := \max_{\mathcal{I} \subseteq [n], \text{connected}} |\mathcal{I}|.
\]

We say that \( \hat{A} \) has a giant component if \( L_1 \approx n \).

In words, \( \hat{A} \) has a giant component if its largest connected component is a non-vanishing fraction of the total number of nodes.

It is well-known that inhomogeneous random graphs exhibit threshold behavior in connectivity and in the existence of the largest component at the following rates:

**Theorem 3.** Suppose Assumptions 1 and 2 hold. Then as \( n \to \infty \),

(a) (Theorem 3.1, Bollobás et al. 2007). \( p_n \approx n^{-1} \) if and only if \( \hat{A} \) has a unique giant component with probability approaching 1.

(b) (Theorem 1, Devroye and Fraiman 2014). \( p_n \approx n^{-1} \log n \) if and only if \( \hat{A} \) is connected with probability approaching 1.

In words, if we observe that \( \hat{A} \) has a giant component, we can expect that \( p_n \gg n^{-1} \). If we observe that \( \hat{A} \) is connected, then we can expect that \( p_n \gg n^{-1} \log n \gg \sqrt{\frac{\log n}{\log n \log n}} \). Together with Theorem 2, this motivates the following rule-of-thumb:

**Rule of Thumb 1.**

(a) Treat \( \hat{\beta}^{(1)} \) and \( \hat{\beta}^{(T)} \) as consistent only if \( \hat{A} \) has a giant component with \( L_1 > n/2 \).

(b) Treat \( \hat{\beta}^{(\infty)} \) as consistent only if \( \hat{A} \) is connected.

Note that if \( \hat{A} \) is connected, it also has a giant component of size \( n \). Our criteria are therefore nested. The choice of the constant \( 1/2 \) in rule (a) ensures uniqueness of the largest component, but is technically arbitrary. Figure 2 provides graphical illustration.
Figure 2: From left to right, a connected network, a disconnected network with giant component, and a disconnected network without a giant component. Networks are obtained from simulation with $f = 1$, $n = 100$ and $p_n = 1/\sqrt{n}$, $\sqrt{\log n}/n$ and $1/n$ respectively.

Remark 4. Our rule of thumb for the consistency of $\hat{\beta}^{(\infty)}$ essentially requires that $p_n \succ n^{-1}$. This is not necessary. Instead, we could, for example, form a rule-of-thumb based on localization of the leading eigenvector. However it is less appealing to do so since such a criteria is non-nested with our criterion for giant component, and additionally introduces a tuning parameter.

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,

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

Furthermore, the bias may be of larger order than the standard deviation of $\hat{\beta}^{(d)}$. In this case, it would not be possible to obtain a non-degenerate limiting distribution without bias correction.

As such, we propose bias-correction and corresponding inference methods based on $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$. The distribution of $\hat{\beta}^{(\infty)}$ is more tricky to characterize. We will propose a data-dependent choice of $a_n$ that leads to convenient properties. Readers who are only interested in the implementation of inference can refer to the summary in Table 1.
3.2.1 Centralities without Measurement Error \((\hat{\beta}^{(1)}, \hat{\beta}^{(T)}, \hat{\beta}^{(\infty)})\)

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

**Theorem 4.** Suppose Assumptions 1 and 3 hold.

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

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

where \(\tilde{V}^{(d)} = \left(\frac{\sum_{i=1}^{n} (C_{i}^{(d)})^2}{\sum_{i=1}^{n} (\tilde{\varepsilon}_{i}^{(d)})^2}\right)^{-2}\sum_{i=1}^{n} (C_{i}^{(d)})^2 (\tilde{\varepsilon}_{i}^{(d)})^2\).

(b) Suppose \(a_n \rightarrow \infty\). Then,

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

where \(\tilde{V}^{(\infty)} = \left(\frac{\sum_{i=1}^{n} (C_{i}^{(\infty)})^2}{\sum_{i=1}^{n} (\tilde{\varepsilon}_{i}^{(\infty)})^2}\right)^{-2}\sum_{i=1}^{n} (C_{i}^{(\infty)})^2 (\tilde{\varepsilon}_{i}^{(\infty)})^2\).

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 \(\tilde{V}^{(1)} = O_p(n^{-3}p_n^{-2})\), \(\tilde{V}^{(T)} = O_p(n^{-2T-1}p_n^{-2T})\). These are the rates of convergence for \(\hat{\beta}^{(1)}\) and \(\hat{\beta}^{(T)}\) 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 \(\varepsilon_{i}\), 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 5** (Inference – Degree and Diffusion). Suppose Assumptions 1, 2 and 3 hold and that $p_n \gg n^{-1}$.

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

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

where

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

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

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

where

$$\hat{V}^{(T)} = \frac{1}{2} \left( \sum_{i=1}^{n} \left( \hat{C}_i^{(T)} \right)^2 \right)^{-2} \cdot \delta^{4T} \cdot \iota_n' \left[ \hat{A} \circ \left( \sum_{t=1}^{2T} \left( \hat{A}^{t-t_n} \right) \left( \iota_n' \hat{A}^{t-1} \right) \right) \right]^{\odot 2} \iota_n,$$

$$\hat{B}^{(T)} = \left( \sum_{i=1}^{n} \left( \hat{C}_i^{(T)} \right)^2 \right)^{-1} \sum_{t=1}^{2T-1} b_T(t, \delta) \cdot \iota_n \hat{A}_n.$$
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. 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. In fact, \(\hat{V}(d)/\hat{V}_0(d) \to 0\). 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 (appropriately scaled) 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).

Additionally, \(\sqrt{\hat{V}(d)/\hat{B}(d)} = O_p(p_n)\) so that the bias is of larger order than the variance. Bias correction is therefore necessary for obtaining a non-degenerate asymptotic distribution. To see this, write:

\[
\frac{\hat{\beta}(d) - \beta(d)}{\sqrt{\hat{V}_n}} = \frac{\hat{\beta}(d) - \beta(d) - B(d)/np_n}{\sqrt{\hat{V}_n}} + \frac{B(d)/np_n}{\sqrt{\hat{V}_n}} =: \Gamma_1 + \Gamma_2.
\]

Suppose we chose \(v_n = \text{Var}(\hat{\beta}(d))\). Then \(\Gamma_1 \overset{d}{\to} D\), where \(D\) is some non-degenerate distribution. However, \(\Gamma_2\) diverges to \(+\infty\) or \(-\infty\) depending on the sign of \(B(d)\). On the other hand, suppose we chose \(v_n\) so that \(\Gamma_2\) is bounded. Then \(\Gamma_1 \overset{d}{\to} 0\) since \(\text{Var}(\hat{\beta}(d))/v_n \to 0\). That is, its limit is degenerate. Bias correction is thus necessary for inference.

In order for bias correction to improve mean-squared error, 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 $t \geq 2$, even though it works well for $t = 1$. Using this latter fact, 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 in terms of $T$. We provide explicit formulae for $T \leq 10$ in Tables 7 and 8.

Hypothesis Testing

Our theory suggests the following test for $d \in \{1, T\}$.

**Definition 10.** To test the hypothesis $H_0 : \beta(d) = \beta_0$ against $H_1 : \beta(d) \neq \beta_0$ at the significance level of $\alpha$, define

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

(6)

where $\Phi$ 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 $\beta^{(d)} = \beta_0$, $E \left[ \phi^{(d)} \right] \rightarrow \alpha$.

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, $\hat{B}^{(d)}/\sqrt{V^{(d)}} = O_p \left( p_n^{-1/2} \right)$. As such, if $p_n \sim 1$, division by $\sqrt{V^{(d)}}$ 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 $\sqrt{V^{(T)}}$ so that they can be ignored. With only the stipulation that $p_n \sim n^{-1}$ 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,

\[
\text{Var} \left( \sum_{i=1}^{n} C_i \varepsilon_i \right) =: V_0^{(d)} .
\]

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

\[
\text{Var} \left( \frac{\hat{\beta}^{(d)} - E[\beta^{(d)}]}{\sqrt{\hat{V}_0}} \right) \to \infty .
\]

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 11.** For \( d \in \{1, T\} \) and a given \( \alpha \), let

\[
C_0^{(d)} := \left[ \hat{\beta}^{(d)} - \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{V}_0^{(d)}} , \hat{\beta}^{(d)} + \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{V}_0^{(d)}} \right] . \tag{7}
\]

Suppose \( \hat{\beta}^{(d)} \geq 0 \) and let

\[
C^{(d)} := \left[ \frac{\hat{\beta}^{(d)}}{1 - \hat{B}^{(d)} + \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{V}^{(d)}}} , \frac{\hat{\beta}^{(d)}}{1 - \hat{B}^{(d)} + \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{V}^{(d)}}} \right] . \tag{8}
\]

Finally, let \( C^{(d)} = C_0^{(d)} \cup C^{(d)} \).

**Remark 5.** If \( \hat{\beta}^{(d)} < 0 \), the upper bound in the above definition of \( C^{(d)} \) is smaller than the lower bound. In this case,

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

The following is immediate:

**Corollary 2 (Confidence Interval).** \( P \left( \beta^{(d)} \in C^{(d)}_\ast \right) \to 1 - \alpha . \)
We can obtain one-sided confidence intervals by modifying the bounds as usual. More generally, as long as $C(d)$ is a $1-\alpha$ confidence interval for $\beta(d) \neq 0$ and $C_0(d)$ 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 $C_0(d) = \{0\}$. 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

$$C_0(d) = \{0\} \quad \text{and} \quad C(d) = \left(-\infty, \frac{\hat{\beta}(d)}{1 - \hat{B}(d) - \Phi^{-1}(1-\alpha) \sqrt{\hat{V}(d)}}\right].$$

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

$$C_*(d) \subseteq \left(-\infty, \hat{\beta} + \Phi^{-1}(1-\alpha) \sqrt{\hat{V}_0(d)}\right].$$

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

$$\tilde{\beta}(d) = \frac{\hat{\beta}(d)}{1 - \hat{B}(d)}.$$

Bias-corrected estimators have faster rates of convergence:

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

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

**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 $R < \infty$:

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

(9)
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 (9), 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 2 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. Candes 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 2 (Stochastic Block Model).** The Stochastic Block Model (SBM) is one of the earliest statistical models of networks. It assumes that individuals fall into groups $g \in \{1, \ldots, B\}$ 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 $\pi_g$ 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 3.

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

**Theorem 6 (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 $\eta > 0$,

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

(10)
and \( a_n \ll np_n \).

Then,

\[
\hat{S}_0(\infty) := \frac{\hat{\beta}(\infty) - \beta(\infty) \left(1 - \hat{B}(\infty)\right)}{\sqrt{\hat{V}_0(\infty)}} \overset{d}{\to} N(0, 1). \tag{11}
\]

where

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

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

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

\[
\hat{S}(\infty) := \frac{\hat{\beta}(\infty) - \beta(\infty) \left(1 - \hat{B}(\infty)\right)}{\sqrt{\hat{V}(\infty)}} \overset{d}{\to} N(0, 1),
\]

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 5 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:

\[
\hat{\beta}^{(\infty)} = \frac{\hat{\beta}^{(\infty)}}{1 - \hat{B}^{(\infty)}}.
\]

**Choice of \( a_n \)**

The following data-dependent choice is convenient:

**Corollary 4.** Suppose Assumptions 1, 2, 3 and E2 hold. Suppose also that \( a_n = \sqrt{\lambda_1(\hat{A})} \) is estimated. If \( p_n \) satisfies Equation (10),

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

Since \( \lambda_1(\hat{A}) \approx \hat{\lambda}_1np_n \), the above choice of \( a_n \) satisfies the conditions in case (a)(iii) of Theorem 6, 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

\[
\hat{C}^{(\infty)} \left( \hat{C}^{(\infty)} \right)' = \lambda_1(\hat{A})v_1(\hat{A}) \left( v_1(\hat{A}) \right)' = \arg\min_{\text{rank}(M) = 1} \| M - \hat{A} \|_2.
\]
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.

It is also common for applied researchers to scale eigenvector centrality by its standard deviation (e.g. Banerjee et al. 2019; Chandrasekhar et al. 2018). This is typically done so that the regression coefficient can be interpreted as the effect on outcome of a one standard-deviation increase in eigenvector centrality. In effect, this procedure sets $a_n \approx \sqrt{n}$:

**Corollary 5.** Suppose Assumptions 1, 2, 3 and E2 hold. Suppose also that

$$a_n = \left( \frac{1}{n} \sum_{i=1}^{n} \left( [v_1(\hat{A})]_i - \frac{1}{n} \sum_{i=1}^{n} [v_1(\hat{A})]_i \right)^2 \right)^{-1/2}.$$  

Then

$$a_n = (1 + o_p(1)) \cdot \frac{\sqrt{n}}{\sqrt{1 - E[\phi_1(U_i)]^2}}.$$  

As such, Case (a) (ii) of Theorem 6 applies when $p_n \succ n^{-2/3}$.

Note that with this choice of $a_n$, our theorem is able to accommodate less sparsity that if $a_n = \sqrt{\lambda_1(\hat{A})}$.

**Remark 6.** 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 6 does not cover regression on multiple eigenvectors but it accommodates greater sparsity and facilitates tests of $\beta^{(\infty)} = \beta_0$ for $\beta_0 \neq 0$.

**Remark 7.** To compare our results to that of Cai et al. (2021), set $a_n = \sqrt{n}$. The condition in Case (a) (ii) of Theorem 6 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.
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:

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

where $C_i^{(d)}$ are centrality measures calculated on $A$. In this simulation, we draw $U_i \sim U[0, 1]$, $\varepsilon_i^{(d)} \sim \text{N}(0, 1)$, where $\varepsilon_i^{(d)} \perp U_i$ and $\varepsilon_i^{(d)} \perp \hat{A}_{jk}$ 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 4. 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 6: we can choose \( a_n = \sqrt{\lambda_1(A)} \) 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 \( \hat{\beta}^{(d)} \) when \( p_n = 1/n \). Figure 5 indeed shows the distribution of \( \hat{\beta}^{(d)} \) for each \( n \). The estimators concentrate around \( \beta \) 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 6, 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.

Finally, we consider the case when \( p_n = n^{-1} \sqrt{\log \log n} \). 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 7, where \( \hat{\beta}^{(\infty)} \) is drifting further away from \( \beta \) 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 estimators in Definition 12. In this subsection, we study the effects of bias correction in the regime \( p_n = 1/\sqrt{n} \).

Figure 8 shows the distribution of the estimators when \( n = 500 \). Here \( a_n = \sqrt{\lambda_1(A)} \). 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 5 and 6. As before, we consider the regime \( p_n = 1/\sqrt{n} \). Figure 9 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 heteroskedasticity consistent (robust) standard errors and conduct inference under the assumption that it has a standard normal distribu-
Figure 4: Distribution of $\tilde{\beta}(\infty)$ for $n = 100$, $p_n = 1/n$ under various $a_n$. $\beta = 1$ (orange dashed line).

Figure 5: Distribution of $\tilde{\beta}(d)$ 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 $p_n = 1/n$. For $\tilde{\beta}(\infty)$, $a_n = \sqrt{n}$. $\beta = 1$ (orange dashed line).
Figure 7: Distribution of $\hat{\beta}(d)$ for $p_n = n^{-1} \sqrt{\log n / \log \log n}$. For $\beta^{(\infty)}$, $a_n = \sqrt{n}$. $\beta = 1$ (orange dashed line).

Figure 8: Distributions of $\hat{\beta}(d)$ and their bias corrected versions $\tilde{\beta}(d)$ for $p_n = 1/\sqrt{n}$. $\beta = 1$ (orange dashed line).

Figure 9: Distribution of the centered and scaled test statistics in Theorems 5 and 6. Robust refers to tests based on $t$-statistic with robust (heteroskedasticity consistent) standard errors.
tion. 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. Table 2 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 tests control size well. 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 3. 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 10 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.

5 Empirical Demonstration

In this section, we demonstrate the relevance of our theoretical findings via an application inspired by De Weerdt and Dercon (2006). 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. 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: Power of the two-sided test of $H_0 : \beta = 1$ under various alternatives. Test at 5% level of significance (orange dashed line).

<table>
<thead>
<tr>
<th>$p_n$</th>
<th>Statistic</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>0.1</td>
<td>Degree</td>
<td>Ours</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robust</td>
</tr>
<tr>
<td></td>
<td>Diffusion</td>
<td>Ours</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robust</td>
</tr>
<tr>
<td></td>
<td>Eigenvector</td>
<td>Ours</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robust</td>
</tr>
<tr>
<td>$n^{-1/3}$</td>
<td>Degree</td>
<td>Ours</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robust</td>
</tr>
<tr>
<td></td>
<td>Eigenvector</td>
<td>Ours</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robust</td>
</tr>
<tr>
<td>$n^{-1/2}$</td>
<td>Degree</td>
<td>Ours</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robust</td>
</tr>
<tr>
<td></td>
<td>Eigenvector</td>
<td>Ours</td>
</tr>
</tbody>
</table>

Table 2: 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.
<table>
<thead>
<tr>
<th>$p_n$</th>
<th>Statistic</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>0.1</td>
<td>Degree - Robust</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Diffusion - Robust</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Eigenvector</td>
<td>0.845</td>
</tr>
<tr>
<td>$n^{-1/3}$</td>
<td>Degree - Robust</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Diffusion - Robust</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Eigenvector</td>
<td>0.998</td>
</tr>
<tr>
<td>$n^{-1/2}$</td>
<td>Degree - Robust</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Diffusion - Robust</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Eigenvector</td>
<td>0.832</td>
</tr>
</tbody>
</table>

Table 3: 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.

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:

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

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.

**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 11 and the degree distributions are described in Table 4. By construction, BS is much sparser than US. Due to the availability of five panels, UF is denser than the other two. With $n = 120$ households, $\sqrt{n} = 11$. We might therefore be concerned that $np_n \ll \sqrt{n}$ especially for US and BS. Note that US and UF are connected. BS is disconnected but has a clear giant component. Our rule-of-thumb suggests that $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(T)}$ should across all three networks. $\hat{\beta}^{(\infty)}$ should perform well on US and UF, but likely not on BS.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unilateral Social</td>
<td>8.02</td>
<td>7</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>Bilateral Social</td>
<td>2.30</td>
<td>2</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Unilateral Financial</td>
<td>16.53</td>
<td>14</td>
<td>3</td>
<td>79</td>
</tr>
</tbody>
</table>

Table 4: Degree distributions of various networks in Nyakatoke.

Regression results are presented in Table 5. 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 $\delta$ is small ($\approx 0.2$). As the last column shows, bias correction can lead to substantially different estimates. Table 5 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, in line with what is suggested by our rule-of-thumb. 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.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unilateral</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>Degree</td>
<td>-1064</td>
<td>0.67</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>Diffusion</td>
<td>-4274</td>
<td>0.77</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Eigenvector</td>
<td>-12353</td>
<td>0.86</td>
<td>0.91</td>
</tr>
<tr>
<td><strong>Bilateral</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>Degree</td>
<td>-11604</td>
<td>0.06</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>Diffusion</td>
<td>-23672</td>
<td>0.16</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Eigenvector</td>
<td>-10543</td>
<td>0.93</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Unilateral</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial</td>
<td>Degree</td>
<td>-412</td>
<td>0.70</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Diffusion</td>
<td>-4559</td>
<td>0.74</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Eigenvector</td>
<td>-15040</td>
<td>0.77</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table 5: 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, \( \hat{\beta}^{(d)} \).

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 6. Results for eigenvector are omitted since our theory is based on the usual robust \( t \)-statistic. Our confidence intervals leads to tighter lower bound than the those based on the robust \( t \)-statistic. Furthermore, as we increase the desired coverage, our confidence intervals increase much more slowly than the hc/robust confidence intervals. This is because the latter is linear in \( \Phi^{-1}(1 - \alpha/2) \), whereas this term appears in the denominator of \( C^{(d)} \) (as in Definition 8).
Table 6: 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) and Graham (2020b) show, parametric models may point to a way forward.
References


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 distribu-


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


Bourlès, R., Y. Bramoullé, and E. Perez-Richet (2021). Altruism and risk sharing in net-


centrality estimation. *Available at SSRN 3963523*.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence

Candès, E. J. and T. Tao (2010). The power of convex relaxation: Near-optimal matrix

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


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


Appendices

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

Tables 7 and 8 present $b_T(t, \delta)$ used for calculating the bias estimator in Case (b) of Theorem 5. 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 7 and 8 provide the coefficients for $\delta^s$ 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

\[
(\delta^2 - 3\delta^3 + 3\delta^4) t_n' \hat{A}_t n + (3\delta^3 - 2\delta^4) t_n' \hat{A}_t^2 t_n + (2\delta^4) t_n' \hat{A}_t^3 t_n .
\]

<table>
<thead>
<tr>
<th>$T$</th>
<th>$t$</th>
<th>$\delta^2$</th>
<th>$\delta^3$</th>
<th>$\delta^4$</th>
<th>$\delta^5$</th>
<th>$\delta^6$</th>
<th>$\delta^7$</th>
<th>$\delta^8$</th>
<th>$\delta^9$</th>
<th>$\delta^{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>5</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>-3</td>
<td>7</td>
<td>-4</td>
<td>-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>-4</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>-2</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>-3</td>
<td>7</td>
<td>-13</td>
<td>-15</td>
<td>91</td>
<td>-182</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>-4</td>
<td>5</td>
<td>24</td>
<td>-94</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>-7</td>
<td>-1</td>
<td>36</td>
<td>-84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8</td>
<td>-6</td>
<td>-2</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>7</td>
<td>-5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5</td>
<td>-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>-3</td>
<td>7</td>
<td>-13</td>
<td>-4</td>
<td>161</td>
<td>-500</td>
<td>952</td>
<td>-654</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>-4</td>
<td>5</td>
<td>24</td>
<td>-178</td>
<td>450</td>
<td>-740</td>
<td>314</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>-7</td>
<td>11</td>
<td>57</td>
<td>-222</td>
<td>456</td>
<td>-362</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8</td>
<td>-15</td>
<td>6</td>
<td>66</td>
<td>-225</td>
<td>317</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>12</td>
<td>-14</td>
<td>4</td>
<td>59</td>
<td>-148</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>11</td>
<td>-13</td>
<td>6</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>9</td>
<td>-12</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>7</td>
<td>-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Coefficients for the bias of diffusion centrality, $b_T(t, \delta)$. Blanks indicate a coefficient of 0.
Table 8: Coefficients for the bias of diffusion centrality, $b_T(t, \delta)$ (continued from Table 7). Blanks indicate a coefficient of 0.
B Additional Motivation for Econometric Framework

This section discusses the connection of our econometric framework to the “Weak Ties” theory of social networks. We also present other examples of network data that fit our framework.

B.1 “Weak Ties” Theory

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.

B.2 Additional Examples

Example 3. 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 4.** 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.
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 13** (Regularized Eigenvector Centrality). Suppose $p_n$ is known. Let

$$
\lambda_i := \min \left\{ \frac{2np_n}{C_i^{(1)}}, 1 \right\}.
$$

Then, define $\hat{A}_\lambda$ to be the regularized version of $\hat{A}$, where

$$
(\hat{A}_\lambda)_{ij} = \sqrt{\lambda_i \lambda_j} \hat{A}_{ij}.
$$

Finally, define regularized eigenvector centrality and the corresponding OLS estimator to be:

$$
\hat{C}_\lambda^{(\infty)} = a_n v_1 (\hat{A}_\lambda) \quad \text{and} \quad \hat{\beta}_\lambda^{(\infty)} = \frac{Y' \hat{C}_\lambda^{(\infty)}}{(\hat{C}_\lambda^{(\infty)})' \hat{C}_\lambda^{(\infty)}}.
$$

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 7** (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 $p_n > n^{-1}$ implies that $\hat{\beta}_\lambda^{(\infty)} \overset{p}{\to} \beta^{(\infty)}$.

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 6 (Estimated \( p_n \)).** Suppose \( \int_{[0,1]^2} f(u, v) \, du \, dv \geq M > 0 \). Let

\[
\rho_n = p_n \int_{[0,1]^2} f(u, v) \, du \, dv , \quad \hat{\rho}_n = \frac{\iota_n' A \iota_n}{n(n-1)} .
\]

Next define

\[
\hat{\lambda}_i = \min \left\{ \frac{3n \hat{\rho}_n}{M \cdot \tilde{C}_i^{(1)}}, 1 \right\} .
\]

Using \( \hat{\lambda} \) in place of \( \lambda \) in Definition 13 does not change the conclusions of Theorem 7.

**C.1 Proof of Theorem 7**

As in the Proof of Theorem 2, write

\[
\hat{\beta}_\lambda^{(\infty)} = \beta^{(\infty)} + \beta^{(\infty)} \left( v_1(\hat{A}_\lambda) \right)' \left( v_1(A) - v_1(\hat{A}_\lambda) \right) + \frac{1}{\sqrt{a_n}} v_1(\hat{A}_\lambda)' \varepsilon^{(\infty)} .
\]

By Theorem and Remark 2.2 of Le et al. (2017), with probability at least \( 1 - n^{-r} \),

\[
\| A - \hat{A} \| \leq k r^{3/2} \sqrt{\bar{\sigma}^2}
\]

where \( k \) is a universal constant. Therefore, by the Davis-Kahan inequality (Theorem 4.5.5 in Vershynin 2018),

\[
\| v_1(A) - v_1(\hat{A}) \| \leq \frac{\| \hat{A} - A \|}{np_n (\lambda_1 - \lambda_2)} = O_p \left( \frac{1}{\sqrt{np_n}} \right) = o_p(1) .
\]

Again, note that

\[
E \left[ v_1(\hat{A})' \epsilon^{(\infty)} \right] \leq \| v_1(\hat{A}) \|^2 = \bar{\sigma}^2 .
\]

Since \( a_n \to \infty \), conclude that \( \frac{1}{\sqrt{a_n}} v_1(\hat{A})' \epsilon^{(\infty)} \overset{p}{\to} 0 \) and that \( \hat{\beta}_\lambda^{(\infty)} \overset{p}{\to} \beta^{(\infty)} \).

**C.2 Proof of Corollary 6**

We first note that \( \hat{\rho} \) is a good estimator of \( \rho_n \):
Theorem 8 (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 \( M/ \int f(u,v) \leq 1 \)

\[ P \left( \frac{\hat{\rho}_n}{M} \geq \frac{2M \rho_n}{3 \int f(u,v) \, dv} \right) \rightarrow 1 \]

Setting

\[ \hat{\lambda}_i = \min \left\{ \frac{3n \hat{\rho}_n}{M \cdot C_i^{(1)}}, 1 \right\} \]

ensures that w.p.a. 1,

\[ \max_{i \in [n]} C_{\lambda,i}^{(1)} \leq \frac{3n \hat{\rho}_n}{M} \leq \frac{2 \int f(u,v) \, dv}{M} \cdot np_n . \]

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 \( \| \hat{A}_\lambda - A \| = O(\sqrt{np_n}) \) w.p.a. 1. The proof then proceeds as in that of Theorem 7.
D Proofs

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 \( 1 - \exp\left(-n^2 p_n^2 \frac{k}{\log \log n}\right) \)

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

where \( k \) is a universal constant. In other words,

\[
\left\| A - \hat{A} \right\| = O_p \left( (np_n)^{(1+\nu)/2} \left( \frac{\log n}{\log \log n} \right)^{(1-\nu)/4} \right). \tag{12}
\]

D.1 Proof of Theorem 1

In the setting with no measurement error, we write:

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

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

\[
\sum_{i=1}^{n} C_i^{(1)} \varepsilon_i^{(1)} = \sum_{i=1}^{n} \sum_{j=1}^{n} p_n f(U_i, U_j) \varepsilon_i^{(1)} = \frac{p_n}{2} \binom{n}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} f(U_i, U_j) \varepsilon_i^{(1)} + f(U_j, U_i) \varepsilon_j^{(1)} \tag{13}
\]

= \( O_p \left( n^{3/2} p_n \right) \),

In the last equality, we use our assumption that \( E \left[ \varepsilon_i^{(d)} \mid U_i \right] = 0 \) and \( E \left[ \left( \varepsilon_i^{(d)} \right)^2 \mid U_i \right] \leq \sigma^2 < \infty \), so that

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

56
for some $\gamma > 0$ by the standard CLT for U-statistics (e.g. Theorem 12.3 in Van der Vaart 2000). Similarly,

\[
\sum_{i=1}^{n} \left( C_i^{(1)} \right)^2 = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} p_n f(U_i, U_j) \right)^2 \\
= p_n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} f(U_i, U_j) f(U_i, U_k) \\
= p_n^2 \left( \frac{n}{3} \right)^2 \sum_{k=1}^{n} f(U_i, U_j) f(U_i, U_k) = O_p(n^3 p_n^2) 
\]

(14)

By the LLN for U-statistics,

\[
\frac{1}{\left( \frac{n}{3} \right)^2} \sum_{k=1}^{n} f(U_i, U_j) f(U_i, U_k) \xrightarrow{p} \gamma 
\]

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 $P_k := \int_{[0,1]^2} 1_{B_k} dudv > 0$. Then,

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

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 $n^{3/2} p_n \to \infty$

Next, consider diffusion centrality. Note that:

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

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

\[
[A^t]_{ij} = p_n^{t} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_{t-1}=1}^{n} f(U_i, U_{k_1}) f(U_{k_1}, U_{k_2}) \cdots f(U_{k_{t-1}}, U_j) . 
\]
Applying the CLT for U-statistics as before, we have that

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

Similarly,

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

Next, suppose \( np_n > 1 \). 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 \( np_n < 1 \). In this regime, diffusion is equivalent to degree to a first order. The dominant terms in the numerator and denominator are of order \( O(p_n^{3/2}) \) 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,

\[ E \left[ \sum_{i=1}^n C_i^{(\infty)} \varepsilon_i^{(\infty)} \mid U \right] = \sum_{i=1}^n C_i^{(\infty)} E \left[ \varepsilon_i^{(\infty)} \mid U \right] = 0 \]

\[ \text{Var} \left[ \sum_{i=1}^n C_i^{(\infty)} \varepsilon_i^{(\infty)} \mid U \right] = \sum_{i=1}^n \left( C_i^{(d)} \right)^2 \text{Var} \left[ \varepsilon_i^{(\infty)} \mid U \right] \leq a_n^2 \sigma^2. \]

As such,

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

Thus, \( a_n \to \infty \) implies that \( \tilde{\beta}^{(\infty)} \overset{L^2}{\to} \beta^{(\infty)}. \)

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

\[
\tilde{\beta}^{(\infty)} = \frac{\sqrt{n}}{a_n} Y' \tau_n = \beta^{(\infty)} + \frac{1}{a_n \sqrt{n}} \sum_{i=1}^{n} \varepsilon_i^{(\infty)}.
\]

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

### D.2 Proof of Theorem 2

We first write:

\[
\hat{\beta}^{(d)} = \beta^{(d)} + \beta^{(d)} \frac{(\hat{C}^{(d)})' (\hat{C}^{(d)} - \hat{C}^{(d)})}{(\hat{C}^{(d)})' \hat{C}^{(d)}} + \frac{(\hat{C}^{(d)})' \varepsilon^{(d)}}{(\hat{C}^{(d)})' \hat{C}^{(d)}}
\]

For convenience, denote

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

Also let \( \xi_i = \sum_{j=1}^{n} \xi_{ij} = \sum_{j \neq i} \xi_{ij} \) and \( \xi = (\xi_1, ..., \xi_n)' \). Finally, let \( \xi \) by the \( n \times n \) matrix with \( (i, j) \)th entry \( \xi_{ij} \). Note that \( \xi = \xi \tau_n \). By Assumption 2, \( \xi_{ij} \indep \varepsilon_k^{(d)} \mid U \) for all \( i, j, k \) and \( d \in \{1, T, \infty\} \).

#### D.2.1 Degree

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

\[
(\hat{C}^{(1)})' \varepsilon^{(1)} = C^{(1)} \varepsilon^{(1)} + \xi' \varepsilon^{(1)}.
\]
By conditional independence of $\xi$ and $\varepsilon^{(1)}$, $E\left[\xi'\varepsilon^{(1)}\right] = 0$

\[
\text{Var}[\xi'\varepsilon^{(1)} | U] = \text{Var}\left[2 \sum_{i=1}^{n} \sum_{j>i} \xi_{ij} \varepsilon^{(1)}_{i} | U\right] = 4 \sum_{i=1}^{n} \text{Var}\left[\varepsilon^{(1)}_{i} \sum_{j>i} \xi_{ij} | U\right] \\
= 4 \sum_{i=1}^{n} \left(E\left[\left(\varepsilon^{(1)}_{i}\right)^2 | U\right] \sum_{j>i} E\left[\xi_{ij}^2 | U\right]\right) \\
\leq 2\sigma^2 \sum_{i=1}^{n} \sum_{j>i} p_{n}f(U_i, U_j) (1 - p_{n}f(U_i, U_j)) \\
\leq 2\sigma^2 \sum_{i=1}^{n} \sum_{j>i} p_{n}f(U_i, U_j)
\]

Taking expectations over $U$, we have that

\[
\text{Var}[\xi'\varepsilon^{(1)}] \leq 2\sigma^2 n^2 p_{n} \cdot W \Rightarrow \xi'\varepsilon^{(1)} = O_{p}(np_n^{1/2})
\]

Given Equation (13), $C^{(1)}\varepsilon^{(1)} = O_{p}(n^{3/2}p_{n})$ is thus dominant in the numerator if $np_{n} \succ 1$.

Next, consider the denominator, which has the form:

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

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

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

Taking expectations over $U$,

\[
E[\xi' \xi] \leq n^2 p_{n} \cdot W \Rightarrow \xi' \xi = O_{p}(n^{2}p_{n}) \tag{16}
\]
Next, consider the middle term, which we will bound in $L_2$-norm. Write

$$
E \left[ \left( (C^{(1)})' \xi \right)^2 \mid U \right] = E \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} C_{i}^{(1)} \xi_i C_{j}^{(1)} \xi_j \mid U \right] = E \left[ \sum_{i=1}^{n} C_{i}^{(1)} \xi_i \mid U \right] + E \left[ \sum_{i=1}^{n} \sum_{j \neq i} C_{i}^{(1)} C_{j}^{(1)} \xi_i \xi_j \mid U \right] .
$$

Note that

$$
E \left[ \sum_{i=1}^{n} \sum_{j \neq i} C_{i}^{(1)} C_{j}^{(1)} \xi_i \xi_j \mid U \right] = \sum_{i=1}^{n} \sum_{j \neq i} C_{i}^{(1)} C_{j}^{(1)} E \left[ \xi_i \xi_j \mid U \right] = \sum_{i=1}^{n} \sum_{j \neq i} C_{i}^{(1)} C_{j}^{(1)} E \left[ \xi_i^2 \mid U \right]
$$

The second equality above follows form the fact that when $i \neq j$, $E[\xi_i \xi_j \mid U] = 0$ unless $k = j$ and $l = i$. Furthermore,

$$
E \left[ \sum_{i=1}^{n} C_{i}^{(1)} \xi_i C_{i}^{(1)} \xi_i \mid U \right] \leq \sum_{i=1}^{n} \left( C_{i}^{(1)} \right)^2 E \left[ \xi_i^2 \mid U \right] = \sum_{i=1}^{n} \left( C_{i}^{(1)} \right)^2 E \left[ \sum_{j \neq i} \xi_i^2 \mid U \right]
$$

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

Taking expectation over two displays above,

$$
E \left[ \left( (C^{(1)})' \xi \right)^2 \right] = O(n^4 p_n^3) \quad \Rightarrow \quad (C^{(1)})' \xi = O_p(n^2 p_n^{3/2}) . \quad (17)
$$

By Equation (14), $(C^{(1)})' C^{(1)} = O_p(n^3 p_n^2)$. Putting the rates we derived together, the denominator is

$$
\left( \hat{C}^{(1)} \right)' \hat{C}^{(1)} = O_p(n^3 p_n^2) + O_p(n^2 p_n^{3/2}) + O_p(n^2 p_n) . \quad (18)
$$
Hence, $np_n \succ 1$ implies that

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

It remains to note that

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

so that by the rates in (16), (17) and (18),

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

We can loosely write the above results as

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

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

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

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

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

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

diverges in probability.

D.2.2 Diffusion Centrality

Diffusion centrality is comprised of terms of the form:

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

Here, $\tilde{B} = \{A, \xi\}^t$. $B$ is a mixed product of $A$ and $\xi$, and will be the central object of our analysis. For convenience, define:
**Definition 14** (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 $B_j \in \{A, \bm{\xi}\}$. Suppose $B_j = \bm{\xi}$ for $\tau \geq 0$ number of $j$’s. We will also say that the order of $\bm{\xi}$ in $B$ is $\tau$. Define $\mathcal{J} = \{ j \in [2t+1] \mid b_{k_j, k_{j+1}} = \xi_{k_j, k_{j+1}} \}$. Then, $\mathcal{J}$ indicate the locations of the $\bm{\xi}$ in the mixed product $B$. Let $p = (p_1, \ldots, p_r)'$ record lengths of the contiguous blocks in $\mathcal{J}$. If $p_1, \ldots, p_r$ 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,

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

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 $np_n \succeq 1$. We show that $\hat{\beta}^{(T)} \xrightarrow{P} \beta^{(T)}$ when $np_n \succeq 1$. Write

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

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 $\bm{\xi}$.

We seek to bound $t_n B t_n$ in $L_2$-norm. First note that if $B_j = A$ for all $j \in [t]$, then by Equation (15), $B = O_p(n^{t+1}p_n^t)$. Suppose that $B_j = \bm{\xi}$ for at least one $j$. Then,

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

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

In particular,

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

Furthermore, suppose $B$ is not even. Then,

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

If $B$ is even, then

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

63
Taking expectations over $U$, we therefore have that

$$t'_nA^t t_n = O_p(n^{t+1}p_n^{t'}) \succ t_nB t_n = O_p(n^{t+1-\alpha/2}p_n^{t-\beta/2})$$

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

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

By our analysis, the dominant term in the denominator is $t_nA^2 t_n = O_p(n^{2T+2}p_n^{2T})$. 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

$$\frac{\left( \frac{\hat{C}(T)'}{\hat{C}(T)} - \frac{\hat{C}(T)'}{\hat{C}(T)} \right) \frac{\hat{C}(T)'}{\hat{C}(T)}}{t_nA^2 t_n} \to 0.$$

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 $o_p(n^{T+1}p_n^{2T})$, which is again strictly smaller than that of the denominator. Conclude that $\hat{\beta}(T) \overset{p}{\to} \beta(T)$ if $np_n \succ 1$.

### D.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 $\epsilon^{(\infty)}_i \perp \perp U_i$ (By Assumption 2, $\epsilon^{(\infty)}_i \perp \perp \xi_{jk}$ 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 $S_r(i) = B_r(i) \setminus B_{r-1}(i)$ be the sphere of radius $r$ around $i$. Let $u = \log n \cdot np_n \cdot \log \log n$. Then, w.p.a. 1, there exists $\tilde{v}$ such that for any $\eta > 0$,

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

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

$$\tilde{v} = \sum_{r=0}^R u_r s_r(i), \quad s_r(i) = \frac{1_{S_r(i)}}{\|1_{S_r(i)}\|}$$
where $R < \frac{np_n}{\log \log n}$ and

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

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 $u_0$.

$$1 = \tilde{v}'\tilde{v} = \sum_{r=1}^{R} u_r^2 \leq u_0^2 + \frac{1}{np_n} u_0^2 \left( \sum_{r=1}^{\infty} \left( \frac{4}{u_1} \right)^{r-1} \right).$$

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

$$u_0^2 \geq \frac{1}{1 + \frac{1}{np_n} \frac{1}{1-4/u_1}}.$$

Since $np_n, u \to \infty$ 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

$$\hat{\beta}^{(\infty)} = \frac{a_n \left(v_1(\hat{A})\right)' Y}{a_n^2 \left(v_1(\hat{A})\right)' v_1(\hat{A})} = \frac{1}{a_n} \left(v_1(\hat{A})\right)' Y$$

$$= \frac{1}{a_n} \left(v_1(\hat{A})\right)' \left( a_n \frac{\tau_n}{\sqrt{n}} + \varepsilon^{(\infty)} \right)$$

$$= \frac{\left(v_1(\hat{A})\right)' \tau_n}{\sqrt{n}} + \frac{\left(v_1(\hat{A})\right)' \varepsilon^{(\infty)}}{a_n}.$$

By independence of $\varepsilon^{(\infty)}$ and $(\xi, U)$, we have that $\text{Var} \left[ v_1(\hat{A})' \varepsilon^{(\infty)} | U \right] = \|v_1(\hat{A})\| \sigma^2 = \sigma^2$. Hence, $\sigma^2/a_n$ 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

$$\frac{\left(v_1(\hat{A})\right)' \tau_n}{\sqrt{n}} \overset{p}{\to} 1.$$
in which case
\[ \hat{\beta}^{(\infty)} = \left( \frac{v_1(\hat{A})}{\sqrt{n}} \right)' t_n + o_p(1) = \tilde{v}' t_n + o_p(1) \]

Notice that the optimization problem:
\[
\max_{v \in \mathbb{R}^n} v' t_n \text{ such that } \|v\| = 1
\]
has solution \( v = t_n / \sqrt{n} \) and optimal value \( \sqrt{n} \). We can also consider the constrained optimization problem:
\[
\max_{v \in \mathbb{R}^n} v' t_n \text{ such that } \|v\| = 1 \text{ and } v_1 \geq \frac{1}{\sqrt{2}}.
\]
This problem has solution \( v_1 = \frac{1}{\sqrt{2}} \) and \( v_{-1} = t_{n-1} / \sqrt{2n} \) and optimal value
\[ \gamma^* := \frac{1}{\sqrt{2}} + \frac{n-1}{\sqrt{2n}} \]
The constrained maximization problem corresponds to the best case allocation of \( \left( v_1(\hat{A}) \right)' \) that makes \( \left( v_1(\hat{A}) \right)' \) as close to \( \sqrt{n} \) as possible, subject to the requirement that \( \left( v_1(\hat{A}) \right)' \geq 1/\sqrt{2} \). 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 \( \hat{\beta}^{(\infty)} \overset{p}{\rightarrow} \beta^{(\infty)} \) when \( np_n > \sqrt{\frac{\log n}{\log \log n}} \). Write
\[
\hat{\beta}^{(\infty)} = \beta^{(\infty)} + \beta^{(\infty)} \left( \frac{\hat{C}'(\infty)}{\hat{C}(\infty)} - \hat{C}(\infty) \right)' \hat{C}(\infty) + \left( \frac{\hat{C}'(\infty)}{\hat{C}(\infty)} \right)' \hat{C}(\infty) \varepsilon^{(\infty)} \]
\[ = \beta^{(\infty)} + \beta^{(\infty)} \left( v_1(\hat{A}) \right)' (v_1(A) - v_1(\hat{A})) + \frac{1}{a_n} v_1(\hat{A})' \varepsilon^{(\infty)} \]
(20)
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 $\nu \in (0, 1),$

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

where first equality follows from Equation (12) and the second from our assumption on $np_n$.

Finally, note that $E\left[\iota_n' \xi_{t_n} | U\right] = \sum_{i=1}^{n} \sum_{j \neq i} E\left[\xi_{ij}^2 | U\right] = \sum_{i=1}^{n} \sum_{j \neq i} p_n f(U_i, U_j) (1 - p_n f(U_i, U_j))$ so that the unconditional expectation is $E[\iota_n' \xi_{t_n}] = \Omega\left(n^2 p_n\right)$.

To obtain our desired result, we will show that $\iota_n' A \xi_{t_n}$ converges to a normal distribution.
asymptotically once suitable scaled, and that it dominates \((t'_n \xi^2 t_n - E [t'_n \xi t_n | U])\). We then show that the population quantities in the CLT can be estimated at a sufficiently fast rate.

First observe that by Assumption 2, \(E [t'_n A \xi t_n | U] = 0\). Next, define

\[
V^{(1)}_n(U) := E \left[ (t'_n A \xi t_n)^2 \middle| U \right]
\]

\[
= \sum_{j<k} p_n f(U_j, U_k) (1 - p_n f(U_j, U_k)) \left( \sum_{i \neq j} p_n f(U_i, U_j) + \sum_{i \neq k} p_n f(U_i, U_k) \right)^2 .
\]

Then, by the U-statistics LLN,

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

Next, define the event \(\Upsilon^{(1)} := \{ V^{(1)}(U) > k n^4 p_n^3 \}\), 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, \(\xi_{ij}\)'s continue to be independent after conditioning. By Theorem 3.7 in Chen et al. (2011),

\[
\sup_{z \in \mathbb{R}} \left| P \left( \frac{t'_n A \xi t_n}{\sqrt{V^{(1)}_n(U)}} \leq z | U \right) - \Phi(z) \right| \leq 10 \gamma .
\]

Next, we evaluate the third moments of the summands:

\[
E \left[ \left| \xi_{jk} \sum_{i \neq j} p_n f(U_i, U_j) \right|^3 \middle| U \right] = \sum_{i \neq j} p_n f(U_i, U_j) \left| E \left[ |\xi_{jk}|^3 \middle| U \right] \right| \leq n^3 p_n^3 \cdot p_n
\]

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

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

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

\[
\frac{t'_n A \xi t_n}{\sqrt{V^{(1)}_n(U)}} \xrightarrow{d} N(0, 1) .
\]
It remains to show that $l_n \xi^2 l_n - E[l_n \xi^2 l_n \mid U] = o_p \left(n^{2/3}\right)$. Write

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

Note that

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

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 12 shows all possible configurations of indices that will lead to $E[\xi_{i_1,i_2} \xi_{i_2,i_3} \xi_{i_4,i_5} \xi_{i_5,i_6} \mid U] \neq 0$. Table 9 records the frequency of their appearance.

<table>
<thead>
<tr>
<th>Graph</th>
<th>Number of Instances</th>
<th>$E[\xi_{ij}\xi_{jk}\xi_{i'j'}\xi_{j'k'} \mid U]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_1$</td>
<td>$n(n-1)(n-2)(n-3)$</td>
<td>$p_n^4 f(U_i, U_j) f(U_k, U_l) + o_p(p_n^4)$</td>
</tr>
<tr>
<td>$G_2$</td>
<td>$n(n-1)(n-2)$</td>
<td>$p_n^2 f(U_i, U_j) f(U_k, U_l) + o_p(p_n^2)$</td>
</tr>
<tr>
<td>$G_3$</td>
<td>$n(n-1)(n-2)$</td>
<td>$p_n^2 f(U_i, U_j) f(U_k, U_l) + o_p(p_n^2)$</td>
</tr>
<tr>
<td>$G_4$</td>
<td>$n(n-1)$</td>
<td>$p_n f(U_i, U_j) + o_p(p_n^2)$</td>
</tr>
</tbody>
</table>

Table 9: The number of instances of each graph, as well as the value of their conditional expectations, up to the leading term.
Observe that
\[
E \left[ (\xi_{i1,i2}\xi_{i2,i3} - E [\xi_{i1,i2}\xi_{i2,i3} | U]) (\xi_{i4,i5}\xi_{i5,i6} - E [\xi_{i4,i5}\xi_{i5,i6} | U]) | U \right] \\
\neq E \left[ \xi_{i1,i2}\xi_{i2,i3} - E [\xi_{i1,i2}\xi_{i2,i3} | U] \right] E \left[ \xi_{i4,i5}\xi_{i5,i6} - E [\xi_{i4,i5}\xi_{i5,i6} | U] \right] | U
\]
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 9, \(\Gamma^{(1)} = O_p (n^3 p^2_n)\). Conclude that
\[
l'_n \xi^2 t_n - E \left[ l'_n \xi^2 t_n | U \right] = O_p \left( n^{3/2} p_n \right) = o_p \left( n^2 p^3_n \right).
\]
Using the above results, we can rewrite Equation (22) as
\[
\hat{\beta}^{(1)} = \beta^{(1)} + \beta^{(1)} \frac{l'_n A\xi t_n - E \left[ l'_n \xi^2 t_n | U \right] + O_p \left( n^{3/2} p_n \right)}{\hat{C}^{(1)} \hat{C}^{(1)}} .
\]
Consequently,
\[
\frac{\hat{\beta}^{(1)} - \beta^{(1)} (1 - B^{(1)})}{\beta^{(1)} \sqrt{V^{(1)}}} = \frac{l'_n A\xi t_n}{\sqrt{V^{(1)}}} + \frac{O_p \left( n^{3/2} p_n \right)}{\Omega_p \left( n^2 p^{3/2}_n \right)} \rightarrow N(0, 1).
\]
where
\[
B^{(1)} = \left( \left( \hat{C}^{(1)} \right) \hat{C}^{(1)} \right)^{-1} E \left[ l'_n \xi^2 t_n | U \right] ,
\]
\[
V^{(1)} = \left( \left( \hat{C}^{(1)} \hat{C}^{(1)} \right)^{-2} V^{(1)} \right) .
\]
**Plug-in Estimation**

Finally, we show that \(\hat{B}^{(1)}\) and \(\hat{V}^{(1)}\) estimate \(B^{(1)}\) and \(V^{(1)}\) at appropriate rates. Define \(\hat{V}^{(1)} = \left( \left( \hat{C}^{(1)} \hat{C}^{(1)} \right)^{-2} V^{(1)} \right)\). We will show that
\[
\frac{\hat{B}^{(1)} - B^{(1)}}{\sqrt{V^{(1)}}} \xrightarrow{p} 0 ,\quad \frac{\hat{V}^{(1)}}{V^{(1)}} = \frac{\hat{V}^{(1)}}{V^{(1)}} \xrightarrow{p} 1 .
\]
The first statement above is straightforward:

\[
\frac{\hat{B}^{(1)} - B^{(1)}}{\sqrt{V^{(1)}}} = \frac{1}{\sqrt{V^{(1)}}} \sum_{i \neq j} p_n f(U_i, U_j) + \xi_{ij} - p_n f(U_i, U_j) (1 - p_n f(U_i, U_j))
\]

\[
= O_p \left( \frac{1}{n^2 p_n^{1/2}} \right) \cdot \left( \sum_{i \neq j} \xi_{ij} + \sum_{i \neq j} p_n^2 f^2(U_i, U_j) \right) = O_p \left( \frac{1}{n p_n + p_n^{1/2}} \right) = o_p(1).
\]

Next, consider:

\[
2 \left( \hat{V}^{(1)}_* - V^{(1)}_* \right)
\]

\[
= \sum_{j \neq k} \hat{A}_{jk} \left( \hat{C}_{j}^{(1)} + \hat{C}_{k}^{(1)} \right)^2 - p_n f(U_j, U_k) (1 - p_n f(U_j, U_k)) \left( \sum_{i \neq j} p_n f(U_i, U_j) + \sum_{i \neq k} p_n f(U_i, U_k) \right)^2
\]

\[
= \sum_{j \neq k} \hat{A}_{jk} \left( \hat{C}_{j}^{(1)} + \hat{C}_{k}^{(1)} \right)^2 - p_n f(U_j, U_k) \left( \sum_{i \neq j} p_n f(U_i, U_j) + \sum_{i \neq k} p_n f(U_i, U_k) \right)^2 + O_p \left( n^4 p_n^4 \right)
\]

\[
= 2 \sum_{j \neq k} \hat{A}_{jk} \left( C_{j}^{(1)} + C_{k}^{(1)} \right) (\xi_j + \xi_k) + \sum_{j \neq k} \hat{A}_{jk} (\xi_j + \xi_k)^2
\]

\[
= \Gamma_1^{(1)} + \Gamma_2^{(1)}
\]

\[
+ \sum_{j \neq k} \xi_{jk} \left( C_{j}^{(1)} + C_{k}^{(1)} \right)^2 + 2 \sum_{j \neq k} \xi_{jk} \left( C_{j}^{(1)} + C_{k}^{(1)} \right) (\xi_j + \xi_k) + \sum_{j \neq k} \xi_{jk} (\xi_j + \xi_k)^2
\]

\[
= \Gamma_3^{(1)} + \Gamma_4^{(1)} + \Gamma_5^{(1)}
\]

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 \( a \in [5] \).

\[
\Gamma_1^{(1)} = \sum_{j,k} \hat{A}_{jk} \left( C_{j}^{(1)} + C_{k}^{(1)} \right) \sum_{i \neq j} \xi_{ij} + \sum_{j,k} \hat{A}_{jk} \left( C_{j}^{(1)} + C_{k}^{(1)} \right) \sum_{i \neq k} \xi_{ik}
\]

\[
= 2 \sum_{i,j} \xi_{ij} \sum_{k} \hat{A}_{jk} \left( C_{j}^{(1)} + C_{k}^{(1)} \right) \text{ by symmetry}
\]

71
Taking conditional expectations,

\[
E \left[ \Gamma_{1}^{(1)} | U \right] = E \left[ \left( 4 \sum_{i<j} \xi_{ij} \sum_{k} A_{jk} \left( C_{j}^{(1)} + C_{k}^{(1)} \right) \right)^2 | U \right]
\]

\[
= 16 \sum_{i<j} E[\xi_{ij}^2 | U] \left( \sum_{k} A_{jk} \left( C_{j}^{(1)} + C_{k}^{(1)} \right) \right)^2 \leq 16n^2 p_n \cdot (2np_n)^4 \text{ since } C_{j}^{(1)} \leq np_n \text{ for all } j \in [n] \]

\[
= O_p \left( n^6 p_n^5 \right)
\]

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

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

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

Secondly, we have that

\[
E \left[ \left( \sum_{j,k} A_{jk} \left( \sum_{i \neq j} \xi_{ij} \right) \right)^2 | U \right] = E \left[ \left( \sum_{i,j,l} \xi_{ij} \xi_{jl} \sum_{k} A_{jk} \right)^2 | U \right] \leq n^2 p_n^2 E \left[ \left( \iota_n^t \xi^2 \iota_n \right)^2 | U \right] \leq n^2 p_n^2 \cdot n^4 p_n^2
\]

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

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

Now,

\[
E \left[ \left( \Gamma_{3}^{(1)} \right)^2 | U \right] = \sum_{j,k} E \left[ \xi_{jk}^2 | U \right] \left( C_{j}^{(1)} + C_{k}^{(1)} \right)^4 \leq n^2 p_n \cdot (2np_n)^4
\]

As such, \( \Gamma_{3}^{(1)} = O_p \left( n^3 p_n^{5/2} \right) \).
By a similar argument to above, we also have that
\[ \Gamma_4^{(1)} = 2 \sum_{j,k,l} \xi_{jk} \xi_{jl} \left( C_j^{(1)} + C_k^{(1)} \right) = O_p \left(n p_n \right) \cdot O_p \left( \tau_n \xi^2 \right) = O_p \left(n^3 p_n^2 \right). \]

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

First observe that
\[ \sum_{j,k} \xi_{jk} \sum_{i \neq j} \sum_{l \neq k} \xi_{kl} = \tau_n \xi^3 = O_p \left(n^2 p_n^{3/2} \right) \quad \text{by Lemma 2.} \]

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

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 13. Their frequencies and magnitudes are recorded in Table 10. Summing up the contribution of each graph, we have that the above display is \( O_p \left(n^4 p_n^2 \right) \). Hence, \( \Gamma_5^{(1)} = O_p \left(n^2 p_n^2 \right) + O_p \left(n^2 p_n^{3/2} \right) = O_p \left(n^2 p_n \right) \).

Putting all our results together, we have that
\[ \hat{V}_s^{1} - V_s^{(1)} = o_p \left(1\right), \]

which together with our central limit theorem and result on \( \hat{B}^{(1)} \) implies our desired result:
\[ \hat{S}^{(1)} := \frac{\hat{\beta}^{(1)} - \beta^{(1)} \left(1 - \hat{B}^{(1)} \right)}{\beta^{(1)} \sqrt{\hat{V}^{(1)}}} \xrightarrow{d} N(0, 1). \]
Figure 13: The possible configurations of indices that will lead to $E[\xi_{i_1i_2}\xi_{i_3i_4}\xi_{i_5i_6}\xi_{i_7i_8}|U]$ 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.

Table 10: 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 $G_1$. 

| Graph | Number of Instances | $E[\xi_{ij}\xi_{jk}\xi_{ij'}\xi_{j'k'}|U]$ |
|-------|---------------------|------------------------------------------|
| $G_1$ | $n(n-1)(n-2)(n-3)$  | $p_n^2 f(U_i, U_j) f(U_k, U_l) + O_p(p_n^4)$ |
| $G_2$ | $n(n-1)(n-2)(n-3)$  | $p_n^3 f(U_i, U_j) f(U_k, U_l) + O_p(p_n^4)$ |
| $G_3$ | $n(n-1)$            | $p_n f(U_i, U_j) + O_p(p_n^2)$            |
D.3.2 Case (b)

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

$$\hat{\beta}(T) = \beta(T) + \beta(T) \left( \frac{\hat{C}(T)'}{\hat{C}(T)} - \frac{\hat{C}(T)'}{\hat{C}(T)} \frac{\epsilon(T)}{\hat{C}(T)} \right). \quad (23)$$

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

$$\iota'_n B_{\iota_n} - E [\iota'_n B_{\iota_n} | U] = O_p \left( \frac{n^{t+1}p^t}{\sqrt{n} (\sqrt{n}p_n)^t} \right).$$

Otherwise,

$$\iota'_n B_{\iota_n} = O_p \left( \frac{n^{t+1}p^t}{\sqrt{n} (\sqrt{n}p_n)^t} \right).$$

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 $\xi$. Such terms are dominant provided that they attain the stated upper bounds. There are $T$ of these, taking the form below:

$$\iota'_n \sum_{t=1}^T A^{T+t-1} \xi A^{T-t} \iota_n$$

$$= \sum_{j,k} \xi_{jk} \sum_{i} A_{i_1,i_2} \cdots A_{i_{T+t-1,j}} \cdots A_{i_{2T-j2T+1}}$$

$$= \sum_{j<k} \xi_{jk} \sum_{i} A_{i_1,i_2} \cdots A_{i_{T+t-1,j}} \cdots A_{i_{2T-j2T+1}} \cdots$$

$$+ \sum_{i} A_{i_1,i_2} \cdots A_{i_{T+t-1,k}} \cdots A_{i_{2T-j2T+1}}$$

$$= \sum_{j<k} \sum_{i} A_{i_1,i_2} \cdots A_{i_{T+t-1,j}} \cdots A_{i_{2T-j2T+1}} \cdots$$

$$+ \sum_{i} A_{i_1,i_2} \cdots A_{i_{T+t-1,k}} \cdots A_{i_{2T-j2T+1}}$$

by symmetry

$$= \sum_{j<k} \sum_{i} A_{i_1,i_2} \cdots A_{i_{T+t-1,j}} \cdots A_{i_{2T-j2T+1}}$$

by change of index.
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

$$V^*_s(T)(U) := \delta^{4T} E \left[ \left( \sum_{t=1}^{T} A^{T+t-1} \xi A^{T-t} \right)^2 \mid U \right]$$

$$= \frac{1}{2} \delta^{4T} \sum_{j,k} A_{jk} (1 - A_{jk}) \left( \sum_{t=1}^{2T} \sum_{i=1}^{2T} A_{i1,i2} A_{i2,i3} \cdots A_{i_{t-1},j} \cdot A_{k,i_{t+2}} \cdots A_{i_{2T},i_{2T+1}} \right)^2 .$$

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 14.

Figure 14: When $A$ is binary, $V^*_s(T)(U)$ 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,

$$\frac{1}{n^{4T} p_n^{4T-1}} V^*_s(T) \overset{P}{\to} \delta^{4T} \sum_{t=1}^{2T} \sum_{s=1}^{2T} \frac{1}{2} \int f(U_1, U_2) \cdot \left[ f(U_3, U_4) \cdots f(U_{t+1}, U_1) \right] \cdot \left[ f(U_2, U_{t+2}) \cdots f(U_{2T}, U_{2T+1}) \right]$$

$$\cdot \left[ f(U_{2T+3}, U_{2T+4}) \cdots f(U_{2T+s+1}, U_1) \right] \cdot \left[ f(U_2, U_{2T+2+s}) \cdots f(U_{4T-s-1}, U_{4T-s}) \right] \, dU .$$

(24)

Notice that

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

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

$$B^{(T)} = \left( \left( \hat{C}^{(T)} \right)' \hat{C}^{(T)} \right)^{-1} \sum_{B \in \mathcal{B}} E \left[ \nu_n B_{tn} \mid U \right]$$

$$V^{(T)} = \left( \left( \hat{C}^{(T)} \right)' \hat{C}^{(T)} \right)^{-2} V_s^{(T)} .$$
Then,
\[
\frac{\left( \hat{\beta}^{(T)} - \beta^{(T)} - B^{(T)} \right)}{\beta^{(T)} \sqrt{V^{(T)}}} = \frac{\left( \hat{C}^{(T)} \right)' \left( C^{(T)} - \hat{C}^{(T)} \right) - \sum_{B \in \mathcal{B}} \mathbb{E} [ \ell_n' B t_n | U]}{\sqrt{V_s^{(T)}}} = \ell_n' \sum_{t=1}^{T} A^{T+t-1} \xi A^{T-t} t_n \frac{1}{\sqrt{V_s^{(T)}}} + o_p(1) .
\]

The proof then proceeds similarly to Case (a). Define the event \( \Upsilon^{(T)} := \left\{ V_s^{(T)} > k n^{4T} p_n^{4T-1} \right\} \), where \( k \) is chosen to be \( 1/2 \) the magnitude of the limit in Equation (24). 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:
\[
\ell_n' \sum_{t=1}^{T} A^{T+t-1} \xi A^{T-t} t_n \frac{1}{\sqrt{V_s^{(T)}}} \Rightarrow N(0,1)
\]
which yields the desired result.

**Plug-in Estimation**

We need to estimate the bias for \( B \in \mathcal{B} \), as well as \( V^{(T)} \). We estimate \( V_s^{(T)} \), 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 \( V_s^{(T)} \). 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 \( B \in \mathcal{B} \) be order \( t \) and have order \( \tau \) with respect to \( \xi \). It’s block structure is described by \( p = (p_1, \ldots, p_r) \), 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
\[
\mathbb{E} [ \ell_n' B t_n | U] - \ell_n' \left( A^{t-\tau} \prod_{j=1}^{r} \tilde{\gamma}(p_j, A) \right) t_n = O_p \left( n^{t+1-\tau/2} p_n^{t-\tau/2} \right).
\]
In words, the bias of \( B \) is “close to” \( \tilde{\gamma} \), a polynomial of the unobserved adjacency matrix.
Then provided that we have good estimators of \( \pi_n A' \pi_n \), we will be able to estimate \( E [ \pi_n B \pi_n | U ] \) by substituting them into \( \gamma \).

To obtain the \( \gamma(t, A) \), write:

\[
E [ \pi_n B \pi_n | U ] = \sum_{i_1, \ldots, i_{t+1}} E [ B_{i_1 i_2} \cdots B_{i_t i_{t+1}} | U ] .
\] (25)

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 \( \xi \) matters, i.e. \( i_j \) s.t \( j \in J \cup (J + 1) \). Hence, we are interested in graphs formed by overlaying \( r \) walks, each of length \( p_1, \ldots, p_r \). For a given graph \( G \), write its order as

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

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

\[
i_n' B \pi_n = O_p \left( n^{t+1-\alpha/2}p_n^{-\beta/2} \right) \ll O_p \left( n^{t+1-\tau/2}p_n^{-\tau/2} \right).
\]

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 (D.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 \( \alpha \) but not \( \beta \). 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 \( y \),

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

The indices on the right hand side are unrestricted. The above assertion arises by the following injective mapping from \( r_G \rightarrow [n]^s \). By definition, \( G \) is a walk of length \( p_j \) which traverses \( s+1 \) unique nodes. Let \( j_1, \ldots, j_s, j_{s+1} \) be the steps at which \( G \) reaches a new unique node. Then our injective map is \( i \mapsto (i_{j_1}, \ldots, i_{j_{s+1}}) \).
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 \( i \to (5, 10, 4) \).

There is a small error term arising from two sources. Firstly, we are only capturing the first order term of the \( E[\xi_{i_1 i_2} \cdots \xi_{i_t i_{t+1}} | U] \). 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 \( G \) is finite, the previous display allows to write

\[
E \left[ \iota_n \xi_{\iota_n} \right] = \sum_{G \in \mathcal{G}} \sum_{i \in \mathcal{E}_G} x_i E \left[ \xi_{i_1 i_2} \cdots \xi_{i_t i_{t+1}} | U \right] y_{i_{t+1}}
\]

\[
= \left( 1 + O_p \left( \frac{1}{n} \right) \right) \cdot \sum_{G \in \mathcal{G}} x' A^{s(G)} y
\]

\[
= \left( 1 + O_p \left( \frac{1}{n} \right) \right) \cdot \sum_{s=1}^{t-1} \gamma_s(t) \cdot x' A^{s(G)} y
\]

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 \( \gamma_s(t) \) 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 (26) therefore allows us to write

\[
E \left[ \iota_n' B_{\iota_n} | U \right] - \iota'_n \left( A^{t-\tau} \prod_{j=1}^{r} \gamma(p_j, A) \right) \iota_n = O_p \left( p_n + \frac{1}{n} \right) E \left[ \iota_n' B_{\iota_n} | U \right] = O_p \left( p_n + \frac{1}{n} \right) \cdot O_p \left( n^{t+1-\tau/2} p_n^{t-\tau/2} \right)
\]

\[
= O_p \left( \sqrt{p_n} + \frac{1}{\sqrt{n}} \right) O_p \left( \frac{1}{\sqrt{n}} \right) O_p \left( n^{t+1-\tau/2} p_n^{t-\tau/2} \right).
\]

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 \( \gamma_s(t) \) 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

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

(29)

Here, \( \gamma_t(B) = 0 \) because \( \tau \geq 2 \), and for each block of \( \xi \) 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 \( E[\iota_n'B_t \mid U] \) in terms of the unobserved matrix \( A \). In order to estimate \( E[\iota_n'B_t \mid U] \) at a good rate, we need good estimators for \( \iota_n'A_t \). Let \( \tilde{g}(t) \) be our estimator for \( \iota_n'A_t \). Then we seek:

\[
\iota_n'A_t - \iota_n\tilde{g}(t) = o_p\left( \frac{1}{\sqrt{n}} \right) O_p \left( n^{t+1/2}p_n^{t-1/2} \right) .
\]  

(30)

Suppose we estimate \( \iota_n'A_t \) with the naive estimator: \( \iota_n\hat{A}_t \). Our proofs, in particular Lemma 2, yields that the estimator is consistent at the following rate

\[
\frac{\iota_n'A_t}{\iota_n\hat{A}_t} = 1 + O_p\left( \frac{1}{np_n} \right) ,
\]

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

Next write

\[
\iota_n\hat{A}_t = \iota_n'A_t + \sum_{B \in B} \iota_n'B_t .
\]

Suppose for now that we have access to \( \tilde{g}(1), ..., \tilde{g}(t-1) \) satisfying Equation (30). We can then consider defining

\[
\tilde{g}(t) := \iota_n\hat{A}_t - \sum_{B \in B} \gamma(B, g)
\]

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 (30), and noting that \( B \) is finite,

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

which satisfies Equation (30) since \( \frac{1}{\sqrt{np_n}} \to 0 \). 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 11.

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

\[
B_{*}^{(T)} = \sum_{B \in B} \delta^{(B)} t_n' B t_n.
\]

As before, \( B \) is the set of even \( B \)s 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

\[
\sum_{B \in B} \delta^{(B)} t_n' (B - \gamma(B, g)) t_n = O_p \left( \frac{1}{\sqrt{n}} \right) O_p \left( n^{2T - 1} p_n^{2T - 1} \right).
\]

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

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

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 A.

**D.3.3 Case (c)**

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

\[
\hat{\beta}^{(T)} = \frac{(\hat{C}^{(T)})' \varepsilon^{(T)}}{(\hat{C}^{(T)})' \hat{C}^{(T)}} = \frac{t_n \left( \sum_{t=1}^{T} \delta^t \hat{A}^t \right) \varepsilon^{(T)}}{t_n \left( \sum_{t=1}^{T} \delta^t \hat{A}^t \right)^2}.
\]
Table 11: The coefficients $g_r(t)$ for $t \leq 20$.  

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

<table>
<thead>
<tr>
<th>r</th>
<th>t</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>-1101323</td>
<td>3938488</td>
<td>-7533897</td>
<td>-13585642</td>
<td>198008994</td>
<td>-999517964</td>
<td>3021609795</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>981200</td>
<td>-3101066</td>
<td>4292162</td>
<td>20354680</td>
<td>-188470026</td>
<td>832916330</td>
<td>-214503932</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>-530446</td>
<td>2005368</td>
<td>-4310942</td>
<td>-4074647</td>
<td>91205574</td>
<td>-496007668</td>
<td>1614224856</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>151068</td>
<td>-879116</td>
<td>3034670</td>
<td>-4907736</td>
<td>-17574745</td>
<td>186419358</td>
<td>-871382472</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>4548</td>
<td>213314</td>
<td>-1337608</td>
<td>4785512</td>
<td>-8228118</td>
<td>-25081260</td>
<td>283591630</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>-28178</td>
<td>22194</td>
<td>281946</td>
<td>-2019280</td>
<td>7855844</td>
<td>-16309132</td>
<td>-23702626</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>14700</td>
<td>-46038</td>
<td>58866</td>
<td>333648</td>
<td>-2899960</td>
<td>12404253</td>
<td>-30152117</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>-3480</td>
<td>20662</td>
<td>-72062</td>
<td>124800</td>
<td>337020</td>
<td>-3955392</td>
<td>18806973</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>-309</td>
<td>-3984</td>
<td>27916</td>
<td>-108304</td>
<td>232853</td>
<td>246742</td>
<td>-512509</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>633</td>
<td>-780</td>
<td>-4178</td>
<td>36312</td>
<td>-156828</td>
<td>398862</td>
<td>-1830</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>-290</td>
<td>904</td>
<td>-1503</td>
<td>-3829</td>
<td>45486</td>
<td>-219512</td>
<td>641615</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>79</td>
<td>-368</td>
<td>1252</td>
<td>-2554</td>
<td>-2629</td>
<td>54786</td>
<td>-297780</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>-13</td>
<td>92</td>
<td>-459</td>
<td>1690</td>
<td>-4022</td>
<td>-182</td>
<td>63185</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>1</td>
<td>-14</td>
<td>106</td>
<td>-564</td>
<td>2232</td>
<td>-6010</td>
<td>4010</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>1</td>
<td>-15</td>
<td>121</td>
<td>-684</td>
<td>2893</td>
<td>-8636</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>1</td>
<td>-16</td>
<td>137</td>
<td>-820</td>
<td>3689</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>1</td>
<td>-17</td>
<td>154</td>
<td>-973</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>1</td>
<td>-18</td>
<td>172</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>1</td>
<td>-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Let \( B \) be a mixed product of order \( t \), and let it have order \( \tau \geq 0 \) with respect to \( \xi \). Then,

\[
E \left[ \left( (\varepsilon^{(T)})' B_{t_n} \right)^2 \mid U, \xi \right] = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_{2t+2}=1}^{n} E \left[ \varepsilon_{i_1}^{(T)} \varepsilon_{i_{t+2}}^{(T)} \mid U, \xi \right] B_{i_1,i_2} B_{i_2,i_3} \cdots B_{i_{t+2},i_{t+3}} B_{i_{t+3},i_{t+4}} \cdots B_{i_{2t+1},i_{2t+2}}
\]

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

\[
E \left[ \left( (\varepsilon^{(T)})' B_{t_n} \right)^2 \mid U, \xi \right] \leq \sigma^2 \tilde{t}_n \tilde{B}_{t_n}.
\]

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

\[
(\varepsilon^{(T)})' B_{t_n} = O_p \left( \sqrt{n^{2t+3/2-\tau} p_n^{2t+1-\tau}} \right) = O_p \left( n^{t+3/4-\tau/2} p_n^{t+1/2-\tau/2} \right).
\] (31)

Next, write

\[
\frac{1}{p_n^{t+1}} A_{t_n}^{(T)} \varepsilon^{(T)} = \sum_{i_1, \ldots, i_{T+1}} f(U_{i_1}, U_{i_2}) \cdots f(U_{i_T}, U_{i_{T+1}}) \varepsilon_{i_{T+1}}^{(T)}
\]

\[
= \sum_{i \in I_{T+1}} \sum_{\pi \in \Pi_{T+1}} f(U_{\pi(1)}, U_{\pi(2)}) \cdots f(U_{\pi(T)}, U_{\pi(T+1)}) \varepsilon_{\pi(T+1)}^{(T)}
\]

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

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

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

\[
\sqrt{n} \left( \frac{1}{n} \right)^{1/(T+1)} \sum_{i \in I_{T+1}} \sum_{\pi \in \Pi_{T+1}} f(U_{\pi(1)}, U_{\pi(2)}) \cdots f(U_{\pi(T)}, U_{\pi(T+1)}) \varepsilon_{\pi(T+1)}^{(T)} \xrightarrow{d} N \left( 0, (T + 1)^2 \xi_1 \right)
\] (32)
where

$$\zeta_1 = E \left[ h \left( \left( U_1, \varepsilon_1^{(T)} \right), \left( U_2, \varepsilon_2^{(T)} \right), \ldots, \left( U_{T+1}, \varepsilon_{T+1}^{(T)} \right) \right) h \left( \left( U_1, \varepsilon_1^{(T)} \right), \left( U_{T+2}, \varepsilon_{T+2}^{(T)} \right), \ldots, \left( U_{2T+1}, \varepsilon_{2T+1}^{(T)} \right) \right) \right]$$

$$= E \left[ \sum_{\pi \in \Pi_T} \sum_{\pi' \in \Pi_T'} f(U_{\pi(1)}, U_{\pi(2)}) \cdots f(U_{\pi(T)}, U_{T+1}) \cdot f(U_{\pi'(T+1)}, U_{T+2}) \cdots f(U_{\pi'(2T)}, U_{2T+1}) \left( \varepsilon_{T+1}^{(T)} \right)^2 \right]$$

$$\implies (T)^2 E \left[ f(U_1, U_2) \cdots f(U_T, U_{T+1}) \cdot f(U_1, U_{T+2}) \cdots f(U_{2T}, U_{2T+1}) \left( \varepsilon_1^{(T)} \right)^2 \right] \neq 0 \text{ by assumption.}$$

Here, $\Pi_T'$ is the set of permutations on $\{T+1, \ldots, 2T\}$. As such,

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

Together with the bound in Equation (31), this implies that $\iota_n' A^T \varepsilon^{(T)}$ is the dominant term in the $\left( \hat{C}^{(T)} \right)' \varepsilon^{(T)}$. Next, note that by the $U$-statistic LLN,

$$\frac{1}{n^{2T+1}} \sum_{j=1}^n \left( \iota_n \left( A^T \right)_{.,j} \right)^2 \left( \varepsilon_j^{(T)} \right)^2 = \frac{1}{n^{2T+1}} \sum_{i_1, \ldots, i_{2T+1}} f(U_{i_1}, U_{i_2}) \cdots f(U_{i_T}, U_{i_{T+1}}) \cdot f(U_{i_1}, U_{i_{T+2}}) \cdots f(U_{i_{2T}}, U_{i_{2T+1}}) \left( \varepsilon_{i_{2T+1}}^{(T)} \right)^2$$

$$\xrightarrow{p} \frac{1}{(T)^2} \zeta_1.$$

By the usual plug-in arguments, we have that

$$\frac{1}{n^{2T+1}} \left( \left( \hat{C}^{(T)} \right)' \hat{C}^{(T)} \right)^2 \hat{V}_0^{(T)} = \frac{1}{n^{2T+1}} \sum_{j=1}^n \left( \hat{C}_j^{(T)} \right)^2 \left( \varepsilon_j^{(T)} \right)^2$$

$$= \frac{1}{n^{2T+1}} \sum_{j=1}^n \left( C_j^{(T)} \right)^2 \left( \varepsilon_j^{(T)} \right)^2 + o_p(1) \text{ by the consistency of } \hat{\beta}^{(T)}$$

$$= \frac{1}{n^{2T+1}} \sum_{j=1}^n \delta^{2T} \left( \iota_n \left( A^T \right)_{.,j} \right)^2 \left( \varepsilon_j^{(T)} \right)^2 + o_p(1) \text{ by the dominance of } A^T$$
As such, the robust/heteroskedasticity consistent $t$-statistic

$$
\frac{\hat{\beta}(T)}{\sqrt{\hat{V}_0(T)}} = \frac{\left(\hat{C}(T)\right)' \varepsilon(T)}{\left(\hat{C}(T)\right)' \hat{C}(T) \sqrt{\hat{V}(T)}}
$$

$$
= \frac{n^{-(T+1/2)} \delta^T \hat{A}^T \varepsilon(T)}{\sqrt{\frac{1}{n^2 T+1} \sum_{j=1}^{n} \delta_j^2 T \left(\nu_n \left(A^T\right)_j\right)^2 \left(\varepsilon_j(T)\right)^2}} + o_p(1)
$$

$$
= \frac{\sqrt{n}}{(T+1)!} \sum_{i \in I_{T+1}} \sum_{\pi \in \Pi_{T+1}} f(U_{i_1}, U_{i_2}) \cdots f(U_{i_T}, U_{i_{T+1}}) \varepsilon_1^{(T)} + o_p(1)
$$

$$
\Rightarrow N(0, 1) \quad \text{by Equation (32)}
$$

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

### D.4 Proof of Theorem 6

We start by writing

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

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 (10). Then,

$$
(v_1(A))' v_1(\hat{A}) = (v_1(A))' v_1(A) + \frac{(v_1(A))' \xi v_1(A)}{\lambda_1(A)} + \frac{(v_1(A))' \xi^2 v_1(A)}{(\lambda_1(A))^2} + o_p \left(\frac{1}{(np_n)^3}\right)
$$

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

and

$$
\left(\varepsilon^{(\infty)}\right)' v_1(\hat{A}) = \left(\varepsilon^{(\infty)}\right)' v_1(A) + o_p \left(\frac{1}{a_n}\right).
$$
Now, by the analogous arguments as in Lemma 2 and proof of Lemma 3,

\[
\frac{(v_1(A))' \xi v_1(A)}{\lambda_1(A)} = O_p \left( \frac{1}{\sqrt{n} \sqrt{np_n}} \right) , \quad \frac{(v_1(A))' \xi^2 v_1(A)}{(\lambda_1(A))^2} = O_p \left( \frac{1}{np_n} \right) ,
\]

(33)

\[
\frac{(v_1(A))' \xi^2 v_1(A)}{(\lambda_1(A))^2} - E \left[ (v_1(A))' \xi^2 v_1(A) \mid U \right] = O_p \left( \frac{1}{\sqrt{n} (np_n)} \right)
\]

(34)

Furthermore, by the Davis-Kahan Inequality,

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

(35)

**D.4.1 Case (a)**

Let us now consider individual cases in (a). starting with (iii). Suppose we require only that

\[
p_n > n^{-1} \left( \frac{\log n}{\log \log n} \right)^{1/2 + \eta} .
\]

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 \( a_n > np_n \), Equations (33) and (35), together with Lemma 3 implies that

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

(36)

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 \( p_n > n \log n \). 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 \( a_n \prec (np_n)^{3/2} \),

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

This time the bias correction is required.

Finally, consider Case (a) (i), when \( \beta^{(\infty)} = 0 \). Then it is immediate that Equation
(36) obtains. In all three cases, the asymptotic distribution of the estimator depends on $$(\varepsilon^{(\infty)})' v_1(A)/a_n$$. Next, notice that

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

As such, by Assumption 3,

$$\sigma^2 \leq \text{Var} \left( (\varepsilon^{(\infty)})' v_1(A) \big| U \right) \leq \bar{\sigma}^2.$$

Now, as in the proof of Lemma 3, let $\Upsilon$ 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 $\Upsilon$, $\|v_r(A)\|_\infty \leq 2R/\sqrt{n}$. As such,

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

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

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

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

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

Define

$$B^{(\infty)} = E \left[ (v_1(A))' \xi^2 v_1(A) \big| U \right],$$

$$V_0^{(\infty)} = \text{Var} \left( (\varepsilon^{(\infty)})' v_1(A) \big| U \right).$$
Then we have that
\[
\frac{\hat{\beta}(\infty) - \beta(\infty) (1 - B(\infty))}{\sqrt{V_0(\infty)}} = \frac{\varepsilon'v_1(A)}{\sqrt{\text{Var}((\varepsilon(\infty))'v_1(A) | U)}} + o_p(1) \xrightarrow{d} N(0, 1). 
\]

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

D.4.2 Case (b)

Suppose \(a_n \gg n\sqrt{p_n}\). By Lemma 3, we have that
\[
\frac{\hat{\beta}(\infty) - \beta(\infty)}{\sqrt{V_0(\infty)}} = \frac{\varepsilon'v_1(A)}{\sqrt{\text{Var}((\varepsilon(\infty))'v_1(A) | U)}} + o_p(1) \xrightarrow{d} N(0, 1). 
\]

Furthermore, since \(p_n \gg n^{-1} \log n\), the Davis-Kahan Inequality (Theorem 4.5.5 in Vershynin 2018), together with Theorem 1.1 in the Supplementary Material to Lei and Rinaldo (2015) gives us that
\[
\sum_{r=2}^{R} \frac{\lambda_r(A) v_r(A)'v_1(\hat{A})}{\lambda_1(A) v_r(A)'v_r(\hat{A})} = O_p \left( \frac{1}{\sqrt{np_n}} \right).
\]

As such,
\[
\frac{\hat{\beta}(\infty) - \beta(\infty) (1 - B(\infty))}{\sqrt{V_0(\infty)}} = \frac{\varepsilon'v_1(A)}{\lambda_1(A)} + O_p \left( \frac{1}{(np_n)^{3/2}} \right) 
\]
since \( p_n > 1/\sqrt{n} \). Now note that

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

Recall that from the proof of Lemma 3 that

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

so that \(|\alpha_r| \leq 2R\) for all \( r \in [R] \) w.p.a. 1. We now argue that \( \alpha_1 \overset{p}{\to} 1 \) and \( \alpha_r \to 0 \) for \( r \geq 2 \).

Note that we can write

\[
Av_1(A) = \left( \sum_{r=1}^{R} \tilde{\lambda}_r \left( \frac{\phi_r(U)}{\sqrt{n}} \right) \frac{\phi_r(U)}{\sqrt{n}} \right) \left( \sum_{r=1}^{R} \alpha_r \frac{\phi_r(U)}{\sqrt{n}} \right) \\
= \sum_{r=1}^{R} \tilde{\lambda}_r \alpha_r \phi_r \cdot \left( \frac{\phi_r(U)}{\sqrt{n}} \right) \phi_r(U) + \sum_{r \neq s} \tilde{\lambda}_r \alpha_s \left( \frac{\phi_r(U)}{\sqrt{n}} \right) \phi_s(U) \phi_r(U).
\]

Consequently,

\[
(v_1(A))' Av_1(A) = \sum_{r=1}^{R} \tilde{\lambda}_r n \alpha_r^2 \left( \left( \frac{\phi_r(U)}{\sqrt{n}} \right) \phi_r(U) \right)^2 \\
+ \sum_{r=1}^{R} \sum_{s=1}^{R} \tilde{\lambda}_r n \alpha_r \alpha_s \left( \frac{\phi_r(U)}{\sqrt{n}} \right) \phi_s(U) \phi_r(U) \phi_s(U) \\
+ \sum_{r=1}^{R} \sum_{s \neq t} \tilde{\lambda}_s \alpha_t \alpha_r \left( \frac{\phi_r(U)}{\sqrt{n}} \right) \phi_t(U) \phi_r(U) \phi_t(U).
\]

89
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 \(\alpha_1 \xrightarrow{p} 1\),

\[
\alpha_r \xrightarrow{p} 0 \quad \text{for all } r \geq 2 . \tag{38}
\]

Since \(|\phi_r(U_i)| \leq 1\), the above analysis also implies that

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

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

As such,

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

Conclude by the \(U\)-statistics LLN that

\[
\frac{1}{n^2 p_n} \text{Var} \left( n \cdot v_1(A)' \xi v_1(A) \mid U \right) \xrightarrow{p} 2E \left[ \phi_1(U_1)^2 \phi_1(U_2)^2 f(U_1, U_2) \right] > 0 .
\]
As such, if we define \( \Upsilon^{(\infty)} \) to be the event on which
\[
\text{Var} \left( n \cdot v_1(A) \xi_1(A) \mid U \right) > n^2 p_n \cdot E \left[ \phi_1(U_1)^2 \phi_1(U_2)^2 f(U_1, U_2) \right]
\]
and \( \|v_1(A)\|_\infty \leq 2/\sqrt{n} \). By the Berry-Esseen Inequality,
\[
\sup_{x \in \mathbb{R}} \left| P \left( \frac{(n \cdot v_1(A)) \xi_1(A)}{\sqrt{\text{Var} \left( n \cdot v_1(A) \xi_1(A) \mid U \right)}} \leq x \right| - U, \Upsilon^{(\infty)} \right) - \Phi(x) \right| \leq 10\gamma
\]
where
\[
\gamma = \sum_{i<j} \left[ \sqrt{n} \cdot v_1(A) \right]_i^3 \left[ \sqrt{n} v_1(A) \right]_j^3 \frac{E \left[ \xi_{ij}^3 \mid U \right]}{(\text{Var} \left( n \cdot v_1(A) \xi_1(A) \mid U \right))^{3/2}}
\]
\[
\leq 64 R^4 \frac{n^2 p_n}{(n^2 p_n E [\phi_1(U_1)^2 \phi_1(U_2)^2 f(U_1, U_2)])^{3/2}} \rightarrow 0
\]
where the last bound follows because we are on the event \( \Upsilon \) and does not depend on \( U \) otherwise. Substituting this into Equation (37), we have that
\[
\frac{\lambda_1(A) \left( \hat{\beta}^{(\infty)} - \beta^{(\infty)} (1 - B^{(\infty)}) \right)}{\beta^{(\infty)} \sqrt{V^{(\infty)}}} = \frac{(n \cdot v_1(A)) \xi_1(A)}{\sqrt{\text{Var} \left( n \cdot v_1(A) \xi_1(A) \mid U \right)}} + o_p(1) \overset{d}{\rightarrow} N(0, 1),
\]
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 D.3.1.

D.4.3 Proof of Corollary 5

By our analysis of \( v_1(\hat{A}) \) above, we have that
\[
n \cdot a^{-2} = n \left( \frac{1}{n} \sum_{i=1}^{n} [v_1(\hat{A})]_i^2 - \left( \frac{1}{n} \sum_{i=1}^{n} [v_1(\hat{A})]_i \right)^2 \right)
\]
\[
= 1 - \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [v_1(\hat{A})]_i \right)^2 \quad \text{since} \quad \|v_1(\hat{A})\| = 1
\]
\[
= 1 - \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\phi_1(U_i)}{\sqrt{n}} + o_p(1) \right)^2 \quad \text{by Equation (38)}
\]
\[
\overset{p}{\rightarrow} 1 - E [\phi_1(U_1)]^2 .
\]
D.5 Proofs of Auxiliary Lemmas

D.5.1 Proof of Lemma 1

Suppose \( np_n \gtrsim \log^2 n \). 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 \( A - \hat{A} \leq k(r) \sqrt{np_n} \)

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 \( \varepsilon^2 = \left( \frac{\sqrt{\log n}}{\log \log n} / (np_n) \right)^{1-\nu} \) and noting that their \( d \) is our \( np_n \), we have that with probability at least \( 1 - \exp \left( - (np_n)^{2+\nu} \left( \frac{\log n}{\log \log n} \right)^{(1-\nu)/2} k \right) \)

where \( k \) is a universal constant.

Combining these two inequalities yields the desired results.

D.5.2 Proof of Lemma 2

Let \( B \) be a mixed product as in Definition 14. Suppose \( B_j = \xi \) for at least one \( j \in [t] \) and write:

\[
(t_n' B t_n)^2 = \left( \sum_{i=1}^{n} \sum_{k_1=1}^{n} \cdots \sum_{k_{t-1}=1}^{n} \sum_{j=1}^{n} b_{i,k_1,k_2} \cdots b_{k_{t-1},j} \right) \cdot \left( \sum_{i'=1}^{n} \sum_{k'_1=1}^{n} \cdots \sum_{k'_{t-1}=1}^{n} \sum_{j'=1}^{n} b_{i',k'_1,k'_2} \cdots b_{k'_{t-1},j'} \right)
\]

\[
= \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_{2t+2}=1}^{n} b_{k_1,k_2} b_{k_3,k_4} \cdots b_{k_t,k_{t+1}} \cdot b_{k_{t+2},k_{t+3}} b_{k_{t+4},k_{t+5}} \cdots b_{k_{2t+1},k_{2t+2}}
\]

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:

\[
E \left[ (t_n' B t_n)^2 \mid U \right] = \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_{2t+2}=1}^{n} E \left[ b_{k_1,k_2} b_{k_3,k_4} \cdots b_{k_t,k_{t+1}} \cdot b_{k_{t+2},k_{t+3}} b_{k_{t+4},k_{t+5}} \cdots b_{k_{2t+1},k_{2t+2}} \mid U \right].
\]

92
Notice that each summand is non-zero if and 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 $\xi_{ij}$ that appears in the summand appears at least twice, either as $\xi_{ij}$ or $\xi_{ji}$. This property depends on the positions of the $A$’s to the extent that they break up $\xi$: neighbouring $\xi_{ij}$ and $\xi_{jk}$ share an index so that setting $i = k$ is sufficient for the conditional expectation of their product to be non-zero. If $\xi_{ij}$ and $\xi_{kl}$ 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 $J$ 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 $\xi_{ij}$ 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 overall sum be:

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

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 $\xi_{ij}$ 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 $\xi_{k_j,k_{j+1}}$ to be equal to an existing edge since there are now fewer unique $\xi_{ij}$’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 $n^2p_n$. 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 $|\mathcal{J}| = 2t$ 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 $\iota'_n B t_n$.

**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 $p_1, \ldots, p_r$ are all even. On the other hand, the worst case cost is $n^{2t} p_n^\tau$, which is attained if the second edge are all the initial edges of a new path. For an example, see Figure 15.

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 $p_1, \ldots, p_r$ is odd,

$$t'_n B t_n = O_p \left( \frac{1}{\sqrt{n}} \right) \cdot O_p \left( n^{t+1-\tau/2} p_n^{t-\tau/2} \right).$$
Next, suppose that \( p_1, \ldots, p_r \) are all even. Write
\[
(i_n' B_{tn} - E[i_n' B_{tn} \mid U])^2
= \sum_{k_1=1}^n \sum_{k_2=1}^n \cdots \sum_{k_{t+2}=1}^n (b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_{t+1}} - E[b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_{t+1}} \mid U])
\]
\[
\quad \cdot (b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \cdots b_{k_{2t+1},k_{2t+2}} - E[b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \mid U])
\]
\[
= \sum_{G} \sum_{k \in r_G} (b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_{t+1}} - E[b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_{t+1}} \mid U])
\]
\[
\quad \cdot (b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \cdots b_{k_{2t+1},k_{2t+2}} - E[b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \mid U]) \mid U
\]
In the above display, we are summing over all \( G \). However, as before, the set of relevant \( G \) can be substantially restricted. Define
\[
S'_G := E\left[ \sum_{k \in r_G} (b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_{t+1}} - E[b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_{t+1}} \mid U])
\]
\[
\quad \cdot (b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \cdots b_{k_{2t+1},k_{2t+2}} - E[b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \mid U]) \mid U \right]
\]
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,
\[
S'_G := \sum_{k \in r_G} (b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_{t+1}} - E[b_{k_1,k_2} b_{k_2,k_3} \cdots b_{k_t,k_{t+1}} \mid U])
\]
\[
\quad \cdot E\left[ (b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \cdots b_{k_{2t+1},k_{2t+2}} - E[b_{k_{t+2},k_{t+3}} b_{k_{t+3},k_{t+4}} \mid U]) \mid U \right] = 0.
\]
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 \( G = \{ G \mid S_G \neq 0 \} \). Conclude that
\[
E\left[ (i_n' B_{tn} - E[i_n' B_{tn} \mid U])^2 \mid U \right] = \sum_{G \in G} S'_G = O_p\left( \frac{1}{\sqrt{n}} \right) O_p(S_G) = O_p\left( \frac{1}{\sqrt{n}} \right) O_p(n^{t+1-\tau/2} n^{t-\tau/2}).
\]
D.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

\[
\frac{\lambda_1(\hat{A})}{\lambda_1(A)} w'v_1(\hat{A}) = w'v_1(A) + \frac{w' \xi v_1(A)}{\lambda_1(\hat{A})} + \frac{w' \xi^2 v_1(A)}{\lambda_1(A)}^2 \\
+ \sum_{t=3}^{\infty} \frac{w' \xi^t v_1(A)}{\left(\lambda_1(\hat{A})\right)^t} + \sum_{r=2}^{R} \frac{\lambda_r(A) v_r(A')v_1(\hat{A})}{\lambda_1(A) v_r(A) v_r(\hat{A})} \left\{ \sum_{t=0}^{\infty} \frac{w' \xi^t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t} \right\}.
\]

(39)

where we used the fact that \( f \) is rank \( R \). In the remainder of the proof, we bound \( \Lambda_1, \Lambda_2 \) and \( \Lambda_3 \) for \( w \in \{v_1(A), \varepsilon^{(\infty)}/a_n\} \).

Bounds for \( v_1(A) \)

Suppose \( w = v_1(A) \). We start by bounding \( \Lambda_1 \). For a given \( \nu \in (0,1) \), choose \( T \) such that \( T(1-\nu) > 4 + 2/\eta \). Then,

\[
(np_n)^{-(T(1-\nu)-4)} \left( \frac{\log n}{\log \log n} \right)^{T(1-\nu)/2} \rightarrow 0 .
\]

(40)

This is because the above condition is equivalent to

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

which follows by our choice of \( T \) since \( p_n \) satisfies Equation (10). Observe that by Weyl’s Inequality (e.g. Theorem 4.5.3 in Vershynin 2018),

\[
\left\| \lambda_r(\hat{A}) - \lambda_r(A) \right\| \leq \| \xi \| = o_p (np_n) ,
\]

(41)

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), \( \frac{\lambda_r(A)}{np_n} = \lambda_r \left( \frac{1}{p_n} A \right) / n \xrightarrow{p} \hat{\lambda} \). In other words, w.p.a. 1, we have that \( \lambda_r(\hat{A}) \geq \hat{\lambda} np_n/2 > 0 \).
Next write
\[
\left| \sum_{t=T+1}^{\infty} \frac{w^t \xi^t v_r(A)}{\bar{\lambda}_r(\hat{A})} \right| \leq \sum_{t=T+1}^{\infty} \|w\| \cdot \left( \frac{\|\xi\|}{\bar{\lambda}_r np_n/2} \right)^t \cdot \|v_r(A)\| \quad \text{w.p.a. 1}
\]
\[
= O_p \left( \|w\| \left( \sqrt{\frac{\log n}{\log \log n}} \right)^{T(1-\nu)/2} \right) \quad \text{by Lemma 1}
\]
\[
= O_p \left( \frac{\|w\|}{(np_n)^{2}} \right) \quad \text{by Equation (40).} \tag{42}
\]

Meanwhile,
\[
E \left[ \sum_{t=2}^{T} \frac{w^t \xi^t v_r(A)}{\bar{\lambda}_r(\hat{A})} \right] \leq E \left[ \sum_{t=2}^{T} \frac{w^t \xi^t v_r(A)}{(\bar{\lambda}_r np_n/2)^{t}} \right] \leq \sum_{t=2}^{T} E \left[ \frac{(w^t \xi^t v_r(A))^2}{(\bar{\lambda}_r np_n/2)^{2t}} \right]^{1/2} \quad \tag{43}
\]
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 \(\lambda\), it must satisfy:
\[
\lambda v = M v = \sum_{r=1}^{R} \bar{\lambda}_r \frac{\phi_r(U)}{\sqrt{n}} \phi_r(U)^t v
\]
\[
= \sum_{r=1}^{R} \bar{\lambda}_r v_r \frac{\phi_r(U)}{\sqrt{n}} .
\]
Hence, \(v\) is a linear combination of \(\phi_r(U)/\sqrt{n}\)'s. By convergence of the spectrum, we know that \(\limsup \lambda \leq \bar{\lambda}_1\). Now, let \(\Upsilon\) be the event that \(\frac{1}{\sqrt{n}} \sum_{i=1}^{\epsilon} \phi_r(U_i) \phi_s(U_i) < 1/R^2\). 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 \(\Upsilon\), \(\|v_r(A)\|_{\infty} \leq 2R/\sqrt{n}\).
Now, for a $U \in \Upsilon$, 

$$E \left[ (v_1(A)'\xi v_r(A))^2 \mid U \right]$$

$$= \frac{1}{n^2} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_{2t+2}=1}^{n} \left\{ [v_1(A)]_{k_1} [v_1(A)]_{k_{t+1}} [v_r(A)]_{k_{t+1}} [v_r(A)]_{k_{2t+2}} \cdot E \left[ \xi_{k_1,k_2} \xi_{k_2,k_3} \cdots \xi_{k_t,k_{t+1}} \cdot \xi_{k_{t+2},k_{t+3}} \xi_{k_{t+3},k_{t+4}} \cdots \xi_{k_{2t+1},k_{2t+2}} \mid U \right] \right\}$$

$$\leq \frac{16R^4}{n^2} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_{2t+2}=1}^{n} E \left[ \xi_{k_1,k_2} \xi_{k_2,k_3} \cdots \xi_{k_t,k_{t+1}} \cdot \xi_{k_{t+2},k_{t+3}} \xi_{k_{t+3},k_{t+4}} \cdots \xi_{k_{2t+1},k_{2t+2}} \mid U \right]$$

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

$$E \left[ \xi_{t}^2 \mid U \right] = p_n f(U_i,U_j) (1 - p_n f(U_i,U_j)) \cdots (1 - t \cdot p_n f(U_i,U_j)) \geq 0$$ if $p_n \leq 1/T$.

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 $t \geq 2$,

$$E \left[ \left( \frac{(v_1(A)'\xi^t v_r(A))^2}{(\tilde{\lambda}_r n p_n / 2)^{2t}} \right)^{1/2} \right] = O \left( \frac{1}{(\sqrt{n} p_n)^t} \right). \quad (44)$$

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

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

As such,

$$E \left[ (v_1(A)'\xi v_r(A))^2 \mid U \right] \leq \frac{16R^4}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_n f(U_i,U_j) (1 - p_n f(U_i,U_j))$$

$$= O_p \left( p_n \right).$$
Together with the fact that $P(\Upsilon) \to 1$, this yields

$$\frac{v_1(A)^t \xi v_r(A)}{\lambda_r np_n/2} = O_p \left( \frac{1}{n\sqrt{p_n}} \right). \quad (45)$$

Noting that $(v_1(A))^t v_r(A) = 0$, Equations (42), (44) and (45) yield

$$\sum_{t=0}^{\infty} u^t \xi^t v_r(A) = O_p \left( \frac{1}{n p_n} \right), \quad \sum_{t=3}^{\infty} u^t \xi^t v_r(A) = o_p \left( \frac{1}{n p_n} \right) \quad (46)$$

Substituting Equations and (46) into (39) yields the desired bound for $v_1(A)$.

**Bounds for $\varepsilon(\infty)/a_n$**

Next, suppose $w = \varepsilon(\infty)/a_n$. The proof follows that for $v_1(A)$ up to Equation (43). To proceed, recall that conditional on $U, \xi \perp \varepsilon(\infty)$. Conditioning again on the event $\Upsilon$,

$$E \left[ \left( \frac{\varepsilon(\infty)}{a_n} \right)^t \xi^t v_r(A) \right]^2 \mid U, \xi]$$

$$= \frac{1}{n \cdot a_n^2} \sum \sum \ldots \sum \left\{ E \left[ \varepsilon_k^2 \mid U, \xi \right] \varepsilon_k \mid U, \xi \right\} \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots \cdot \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots \cdot \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots \cdot \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots \cdot \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots \cdot \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots \cdot \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots \cdot \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots \cdot \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots \cdot \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots \cdot \xi_k \cdot \xi_{k+1} \cdot \xi_{k+2} \cdot \ldots$$

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

$$E \left[ \varepsilon_{k_1} \mid U, \xi \right] = E \left[ E \left[ \varepsilon_{k_1} \mid U_k \right] \varepsilon_{k_{t+2}} \mid U, \xi \right]$$

$$= E \left[ E \left[ \varepsilon_{k_1} \mid U_k \right] \varepsilon_{k_{t+2}} \mid U, \xi \right]$$

$$= E \left[ \varepsilon_{k_1} \mid U_k \right] E \left[ \varepsilon_{k_{t+2}} \mid U_k \right] = 0.$$
Hence, we only need to consider sequences where \( k_{t+2} = k_1 \), so that

\[
E \left[ \left( \frac{\varepsilon(\infty)}{a_n} \right)' \circ (A) \right]^2 \mid U, \xi \leq \frac{\tilde{\sigma}^2}{n \cdot a_n} \sum_{k_{t+1}=1}^{n} \sum_{k_{t+2}=1}^{n} \xi_{k_1,k_2,k_3,k_4} \cdots \xi_{k_{t+1},k_{t+2}} \xi_{k_{t+1},k_{t+2}}
\]

\[
= \frac{\tilde{\sigma}^2}{n \cdot a_n} \xi^{2t+1}
\]

Taking expectations over \( U \) and \( \xi \), we have that

\[
E \left[ \left( \frac{\varepsilon(\infty)}{a_n} \right)' \circ (A) \right]^2 = \frac{\tilde{\sigma}^2}{n \cdot a_n} O \left( n^{t+1} p_n^{t+1/2} \right)
\]

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 \( t \geq 1 \),

\[
E \left[ \left( \frac{\varepsilon(\infty)}{a_n} \right)' \circ (A) \right]^2 \mid U \leq O \left( \frac{p_n^{1/4} p_n^{t+1/2}}{a_n} \right) = o \left( \frac{1}{a_n} \right).
\]

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

\[
\left( \frac{\varepsilon(\infty)}{a_n} \right)' \circ (A) = \frac{1}{a_n \sqrt{n}} \sum_{i=1}^{n} \phi_i(U_i) \varepsilon_i^{(\infty)} = O_p \left( \frac{1}{a_n} \right)
\]

Combining our last two estimates, we have that

\[
\sum_{t=0}^{T} \left( \frac{\varepsilon(\infty)}{a_n} \right)' \circ (A) / \left( \tilde{\lambda}_r n p_n / 2 \right)^t = O_p \left( \frac{1}{a_n} \right)
\]

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

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

100
where $k$ is the constant in Lemma 1. Then $P(\tilde{\Upsilon}) \to 1$ by Lemma 1 and Equation (41). Furthermore,

$$
P \left( \left\| \sum_{t=1}^{\infty} \frac{(\varepsilon(\infty))^t \xi^t v_r(A)}{\lambda_r(\hat{A})} \right\| \leq x \mid U, \xi, \tilde{\Upsilon} \right) \leq \frac{1}{x^2} \mathbb{E} \left[ \left( \sum_{t=1}^{\infty} \frac{(\varepsilon(\infty))^t \xi^t v_r(A)}{\lambda_r(\hat{A})} \right)^2 \right] \left\| U, \xi, \tilde{\Upsilon} \right\|
$$

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

$$
\leq \frac{\tilde{k}}{n p_n} \text{ by Equation (40), on the event } \tilde{\Upsilon}.
$$

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

$$
P \left( \left\| \sum_{t=1}^{\infty} \frac{(\varepsilon(\infty))^t \xi^t v_r(A)}{\lambda_r(\hat{A})} \right\| \leq x \right) \leq P(\tilde{\Upsilon}) \left( \left\| \sum_{t=1}^{\infty} \frac{(\varepsilon(\infty))^t \xi^t v_r(A)}{\lambda_r(\hat{A})} \right\| \leq x \right) + 1 - P(\tilde{\Upsilon})
$$

$$
\leq \frac{\tilde{k}}{n p_n} + 1 - P(\tilde{\Upsilon}) \to 0.
$$

Hence, we conclude that

$$
\sum_{t=1}^{\infty} \frac{(\varepsilon(\infty)/a_n)^t \xi^t v_r(A)}{\lambda_r(\hat{A})} = \frac{1}{a_n} o_p(1) = o_p \left( \frac{1}{a_n} \right). \quad (50)
$$

Next, note that

$$
\left| \frac{1}{\lambda_s(A)} v_r(A)^t v_s(\hat{A}) \right| \leq \left| \frac{1}{\lambda_s(A)} v_r(A)^t v_s(A) \right| + \| v_r(A) \| \cdot \left| \frac{v_s(\hat{A}) - \theta v_s(A)}{2 \lambda_s n p_n} \right| + o_p(1)
$$

$$
= \left| \frac{1}{\lambda_s(A)} v_r(A)^t v_s(A) \right| + o_p(1), \quad (51)
$$

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

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

101
As before, since \( v_r(A)'v_1(A) = 0 \) for all \( r \geq 2 \), we can write \( \Lambda_3 \)

\[
\sum_{r=2}^{R} \frac{\lambda_r(A)}{\lambda_1(A)} v_r(A)'v_1(\hat{A}) \{ \sum_{t=0}^{T} \frac{(\varepsilon(\infty)/a_n)'\xi_t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t} + \sum_{t=T+1}^{\infty} \frac{(\varepsilon(\infty)/a_n)'\xi_t v_r(A)}{\left(\lambda_r(\hat{A})\right)^t} \} = o_p\left(\frac{1}{a_n}\right) .
\]

(52)

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

\[
\sum_{t=1}^{\infty} \frac{(\varepsilon/a_n)'\xi_t v_1(A)}{\left(\lambda_1(\hat{A})\right)^t} = \sum_{t=1}^{T} \frac{(\varepsilon/a_n)'\xi_t v_1(A)}{\left(\lambda_1(\hat{A})\right)^t} + \sum_{t=T+1}^{\infty} \frac{w'\xi_t v_1(A)}{\left(\lambda_1(\hat{A})\right)^t} = o_p\left(\frac{1}{a_n}\right) .
\]

(53)

We conclude by remarking that from Equation (51),

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

so that the LHS of Equation (39) also converges in probability to \( w'v_1(\hat{A}) \).