Network Regression and Supervised Centrality Estimation

Junhui Cai
University of Notre Dame

Ran Chen
Massachusetts Institute of Technology

Dan Yang
The University of Hong Kong

Wu Zhu
Tsinghua University

Haipeng Shen
The University of Hong Kong

Linda Zhao
University of Pennsylvania

Abstract

The centrality in a network is often used to measure nodes' importance and model network effects on a certain outcome. Empirical studies widely adopt a two-stage procedure, which first estimates the centrality from the observed noisy network and then infers the network effect from the estimated centrality, even though it lacks theoretical understanding. We propose a unified modeling framework, under which we first prove the shortcomings of the two-stage procedure, including the inconsistency of the centrality estimation and the invalidity of the network effect inference. Furthermore, we propose a supervised centrality estimation methodology, which aims to simultaneously estimate both centrality and network effect. The advantages in both regards are proved theoretically and demonstrated numerically via extensive simulations and a case study in predicting currency risk premiums from the global trade network.

Keywords: Hub centrality, Authority centrality, Measurement error, Global trade network, Currency risk premium

1 Introduction

In many disciplines such as economics, finance, and sociology, there have been great interests in studying the network effect, that is, the effect of a network on certain outcomes of interest due to relationships among agents (e.g., individual persons, firms, industries, and countries). One popular approach is to bridge the outcome and network via an intermediary – the centrality of the network.

As a low-rank summary of a network, centrality is a common metric to measure agents' importance in the network, which in turn induces a wide range of agent behaviors and consequently shapes certain outcomes of them. A strong motivation for centrality is that many real-world networks exhibit a low-rank structure, i.e., the leading singular value dominates the rest in magnitude (Zhu and Yang, 2020, Liu and Tsyvinski, 2020). Centrality itself has rich implications on studying human capital investment (Jackson et al., 2017), information sharing and advertisement (Banerjee et al., 2019, Breza and Chandrasekhar, 2019), firms' investment decision-making (Allen et al., 2019), the identification of banks that are too-connected-to-fail (Gofman, 2017), and stock returns (Ahern, 2013, Richmond, 2019), among many others.

To be specific, researchers often regress the outcome of interest on the network centrality to study the network effect. This approach has been implemented in many fields including portfolio management, finance, and social networks, among others. In portfolio management, Hochberg et al. (2007), Ahern (2013) and Richmond (2019) demonstrated that, for a trade network of firms or countries, a strategy that shorts portfolios with high centralities and longs those with low centralities yields a significant excess return, and regressing risk metrics on the centrality of the financial institutions helps avoid amplification of severe adversarial shocks to the central institutions in the network. Liu (2019) examined the effect of centrality in the production network on the government's investment to illustrate the effectiveness of industrial policies. For social networks, Ozsoylev et al. (2014) and Rossi et al. (2018) regressed the excess returns of investment managers on the centrality of their social networks to study trading behaviors; Kornienko and Granger (2018) and Mojzisch et al. (2021) studied the network effect on mental health by regressing the stress level on the network centrality.

Network centrality is not directly observable however. In practice, researchers often, in Stage 1, compute the centrality from a given network adjacency matrix via some algorithm and then, in Stage 2, feed the computed centrality into the regression. Such practice will be referred to as the *two-stage* procedure throughout.

The validity of the two-stage procedure, however, hinges upon one critical assumption that the centrality is computed from a *noiseless* observed adjacency matrix in Stage 1 so that it is accurate. In reality, network is often observed with noise due to the cost of data collection (Lakhina et al., 2003). For example, the friendship network on Facebook or Twitter is far from a perfect measure of real-life social connections; using self-reported friendships to measure social ties suffers from subjective biases (Banerjee et al., 2013); using patent citations to measure the knowledge flow between companies neglects the communication among workers or executives (Zhu and Yang, 2020). Overlooking noise in networks has demonstrable consequences for network analysis (Borgatti et al., 2006, Frantz et al., 2009, Wang et al., 2012, Martin and Niemeyer, 2019, Candelaria and Ura, 2022).

Given a *noisy* observed network, one has two goals in understanding the network effect:

(G1) Estimate centrality accurately from the observed noisy network;

(G2) Estimate and conduct valid inference of the network effect through the centrality.

The two-stage procedure attempts to achieve these two goals in a sequential manner. It has the following drawbacks: Stage 1 only uses the information from the noisy network to estimate centrality without incorporating the auxiliary information from the regression on the centrality, which can result in inaccurate estimation of the centrality due to large observational errors in the network; Stage 2 is contingent upon Stage 1 – regressing the outcome on the inaccurately estimated centrality exacerbates an inaccurate estimation of the regression coefficients, thereby invalidating the followup statistical inference.

To remedy the shortcomings of the two-stage procedure, we first propose a *unified* framework that fuses two models to achieve the two goals: one network generation model based on the centralities for (G1) and one network regression model for the dependency of the outcome on the centralities for (G2). We then propose a novel *supervised network centrality estimation* (SuperCENT) methodology that accomplishes both (G1) and (G2) *simultaneously*, instead of sequentially.

SuperCENT exploits information from the two models – the network regression model contains auxiliary information on the centrality in addition to the network, and thus provides *supervision* to the centrality estimation. The supervision effect improves the centrality estimation, which in turn benefits the network regression. Therefore, the centrality estimation and the network regression complement and empower each other. Under the unified framework, we derive the theoretical convergence rates and asymptotic distributions of the centralities and regression coefficients estimators, for both the two-stage and SuperCENT methods, which can be used to construct confidence intervals.

We summarize our contributions as follows. Firstly, to the best of our knowledge, we are the first to provide a unified framework to study properties of centrality estimation and inference, and the subsequent network regression analysis when the observed network is noisy.

Secondly, we are the first to demonstrate that the common practice of two-stage can be problematic. For centrality estimation (G1), the two-stage centrality estimates in Stage 1 are inconsistent when the network noise is large. This finding of inconsistency extends the phase transition phenomenon of the singular vectors (Shabalin and Nobel, 2013) and the eigenvectors (Shen et al., 2016) to the network centrality problem. For the network regression (G2), the centrality coefficient estimates are *biased* and *inconsistent*, and the ad-hoc inference can be either *conservative* or *invalid* depending on the network noise level.

Thirdly, we show theoretically and empirically that the proposed SuperCENT dominates the two-stage universally. Specifically, for (G1), SuperCENT yields a more accurate centrality estimation; for (G2), SuperCENT mitigates the coefficient estimation bias, and thus boosts the estimation accuracy under large network noise. Furthermore, SuperCENT provides confidence intervals that are *valid* and *narrower* than the ad-hoc two-stage confidence intervals.

Lastly, we apply both SuperCENT and the two-stage to predict the currency risk premium, based on an economic theory on the relationship between a country's currency risk premium and its importance within the global trade network (Richmond, 2019). We show that a long-short trading strategy based on the SuperCENT centrality estimates yields a return *three times* as high as that by the two-stage procedure. Furthermore, SuperCENT can verify the economic theory via a rigorous statistical test while the two-stage fails.

Our paper contributes to several lines of literature on network modeling, network regression with centralities, covariate-assisted network modeling, network effect modeling, and measurement error. Firstly, the proposed unified framework unites the literature on the noisy network and network regression with centralities. Most existing network literature focuses on only one of the two aspects. On one hand, in the presence of noisy networks, many empirical studies estimated the true network without involving centrality (Lakhina et al., 2003, Handcock and Gile, 2010, Banerjee et al., 2013, Le et al., 2018, Rohe, 2019, Breza et al., 2020). On the other hand, numerous research, including those aforementioned, focused on the network regression model with centralities while ignoring the estimation error of the centralities inherited from the noise of the network.

Our unified framework also relates to the line of research on networks with covariates supervision (Zhang et al., 2016, Li et al., 2016, Binkiewicz et al., 2017, Yan et al., 2019, Ma et al., 2020). One major difference is that SuperCENT uses both the covariates and the response to supervise the estimation, instead of only the covariates. In addition, the existing literature focused mostly on network formation or community detection.

In econometrics, there have been significant efforts to model the network effect on an outcome of interest through regression (De Paula, 2017). One popular approach follows the pioneer work of Manski (1993), the "reflection model" (Lee, 2007, Bramoullé et al., 2009, Lee et al., 2010, Hsieh and Lee, 2016, Zhu et al., 2017). This approach models the network effect through the observed adjacency matrix itself, not through the centralities like ours. There is also a recent surge of literature in network recovery based on the reflection model (De Paula et al., 2019, Battaglini et al., 2021). This literature focuses on the issue of identifiability of the network effect, while our work attends to both estimation and inference of the network effect. Another popular approach assumes that the outcome depends on individual fixed effects, and casts the role of the network through the Laplacian matrix, such that connected nodes share similar individual fixed effects (Li et al., 2019, Le and Li, 2020). This approach emphasizes the network homophily, while ours concentrates on the nodes' position or importance in the network using the centralities.

Lastly, our methodology further contributes to the measurement error literature. Most measurement error literature concerns a regression setup where the covariates are directly observed with errors, which leads to bias in the coefficient estimation (Garber and Klepper, 1980, Pischke, 2007, Abel, 2017). We extend it to the network regression problem. Specifically, the two-stage procedure resembles the measurement error problem: the estimated centralities that are used as covariates in the regression of Stage 2 contain estimation error instead of measurement error. Nevertheless, the derivation of the two-stage bias is not a trivial extension of the classical results because it involves the analysis of the asymptotic joint distributions of the two-stage centrality estimators. Furthermore, SuperCENT corrects the coefficient estimation bias induced by the estimation errors and provides valid inference for the regression coefficients under less restrictive assumptions.

The rest of this article is organized as follows. Section 2 provides the background and formally introduces the unified framework. Descriptions of the two-stage procedure and SuperCENT are given in Section 3. Theoretical properties are studied in Section 4 and the simulation study is shown in Section 5. Section 6 presents the case study of the relationship between currency risk premiums and the global trade network centralities. Section 7 concludes with a summary and future work. The supplementary materials contain additional background information on network and centralities, detailed descriptions of the algorithms for rank-one, multi-rank network models as well as undirected networks, more simulation results, additional information of the case study, some concrete mathematical expressions, and the proofs. We developed an R package, `SuperCENT`, that implements the methods (`https://jh-cai.com/SuperCENT`).

2 A unified framework

2.1 Set-up and background of network

We observe a sample of n observations $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)$ where $y_i \in \mathbb{R}$ is the response and $\mathbf{x}_i \in \mathbb{R}^{p-1}$ is the vector of $p-1$ covariates for the i-th observation as in the multivariate regression setting. Let $\boldsymbol{y} \in \mathbb{R}^n$ denote the column vector of outcome and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ denote the design matrix including the intercept, which is assumed to be fixed.

In a network, the nodes are agents and the edges represent relationships between the agents. The edges can be directed or undirected depending on whether the relationships are reciprocal. This article focuses on directed networks, and the Supplement provides the results for undirected ones. A weighted directed network with n nodes can be represented by an *asymmetric* adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ where a_{ij}'s represent the weighted edges.

Researchers have used multiple versions of network centrality. We refer to Chapter 2 of Jackson (2010) for a comprehensive introduction to centrality. We focus on the *hub* and *authority centralities* (Kleinberg, 1999), which extend the well-known eigenvector centrality associated with the undirected network to the directed network.

For directed networks, there is a distinction between the giver and the recipient, such as the citee-citor in citation networks or web-page networks, the exporter-importer in trade networks, and the investor-investee in investment networks. The hub and authority centralities take into account the different roles of the giver and the recipient, and thus measure the importance of nodes from these two different perspectives. The concept of "hubs and authorities" originated from web searching. Intuitively, the hub centrality of a web page depends on the total level of authority centrality of the web pages it links to, while the authority centrality of a web page depends on the total level of hub centrality of the web pages it receives links from. Supplement S1 provides a toy example to further illuminate this intuition.

Let u_i denote the hub centrality and v_i denote the authority centrality for node i, and let $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^\top$, $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)^\top$. Their relationship hence satisfies $\boldsymbol{u} = \mathbf{A}\boldsymbol{v}$, $\boldsymbol{v} = \mathbf{A}^\top\boldsymbol{u}$. Given $\mathbf{A}$, to calculate the centralities, Kleinberg (1999) proposes to iterate with proper normalization as follows till convergence, for $k = 1, 2, 3, \ldots$,

$$\boldsymbol{u}^{(k)} \leftarrow \mathbf{A}\boldsymbol{v}^{(k-1)}, \quad \boldsymbol{v}^{(k)} \leftarrow \mathbf{A}^\top\boldsymbol{u}^{(k)}. \tag{1}$$

This iterative algorithm is also well known as the power method to compute the singular value decomposition (SVD) of $\boldsymbol{A}$ (Van Loan and Golub, 1996). Therefore, the hub and authority centralities are the leading left and right singular vectors of $\boldsymbol{A}$ respectively. It is worth mentioning that such definition of centrality and the algorithm essentially assume that the adjacency matrix $\boldsymbol{A}$ is noiseless.

2.2 A unified framework

We propose the following unified modelling framework that encapsulates (G1)-(G2),

$$\begin{cases} \boldsymbol{A} = \boldsymbol{A}_0 + \boldsymbol{E} = d\boldsymbol{u}\boldsymbol{v}^\top + \boldsymbol{E}, & \text{(2a)} \\ \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_x + \boldsymbol{u}\beta_u + \boldsymbol{v}\beta_v + \boldsymbol{\epsilon}. & \text{(2b)} \end{cases}$$

The intuitions of the unified framework are as follows. The hub and authority centralities are calculated as the leading left and right singular vectors of the observed adjacency matrix. As such, it is natural to consider the generative model (2a) for the observed adjacency matrix, where $\boldsymbol{A}_0$ is the true adjacency matrix, the true centralities $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ are the parameters to be estimated, and $\boldsymbol{E}$ is the additive noise of mean zero. Then, (2b) naturally models the relationship between the centralities and the response variable. Here, $\boldsymbol{\beta}_x \in \mathbb{R}^p$ is the vector of the regression coefficients, $\beta_u, \beta_v \in \mathbb{R}$ are the coefficients of the hub and authority centralities, and the regression error $\boldsymbol{\epsilon}$ has mean zero. Note that in (2b) it is the true centralites, not the estimated ones, that have direct impacts on the response.

Under the unified framework (2) with observed data $\{\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{y}\}$, our original two goals (G1)-(G2) become concrete: (i) estimate the true centralities $\boldsymbol{u}, \boldsymbol{v}$ and the true network $\boldsymbol{A}_0$; (ii) estimate the regression coefficients $\boldsymbol{\beta}_x, \beta_u, \beta_v$; (iii) construct *valid* confidence intervals (CIs) for the centralities and the regression coefficients.

The low-rank mean plus noise model (2a) has been commonly adopted for matrix estimation or denoising (Shabalin and Nobel, 2013, Yang et al., 2016, Cai and Zhang, 2018), matrix completion (Candes and Plan, 2010), and network community detection with slight modifications (Rohe et al., 2011, Zhao et al., 2012, Lei and Rinaldo, 2015, Le et al., 2016, Gao and Ma, 2021). There is a strand of literature on latent variables network models which can be rewritten as (2a) (Hoff, 2009, Soufiani and Airoldi, 2012, Fosdick and Hoff, 2015).

Model (2a) assumes a rank-one instead of multi-rank structure for $\boldsymbol{A}_0$ for multiple reasons. Firstly, since the centralities are our focus and defined to be the leading pair of singular vectors in the literature, the rank-one structure is a reasonable approximation. Secondly, it is commonly observed that real networks' first singular value dominates the latter ones, a phenomenon that we demonstrate with four real networks in Supplement S2, namely, the global trade, innovation, production, and equity-holding networks. Thirdly, the rank-one structure simplifies the theoretical analysis, which offers many insights for understanding the two-stage and SuperCENT estimators, as well as for extension to multi-rank. Lastly, Remark 1 extends the rank-one model to a multi-rank version, and Supplements

S3.2 and S5.4 contain the corresponding estimation procedure, the simulation results, and the theoretical analysis. As a prelude, the same SuperCENT methodology (Section 3.2) remains valid for multi-rank models and the simulation results are qualitatively similar.

The unified framework unites our estimation goals and provides a theoretical framework to study the behaviors of the two-stage procedure and motivates our new methodology. Under Model (2a) and some extra assumptions on the noise, Shabalin and Nobel (2013) proves that if the noise-to-signal ratio is large, the leading singular vector of $\boldsymbol{A}$ and that of $\boldsymbol{A}_0$ converge to orthogonal as n goes to infinity. This implies that the naive estimation of the centralities by implementing SVD on the observed network will fail in the presence of large noise, which invalidates the common practice of two-stage. Furthermore, unifying the two models motivates our supervised network centrality estimation (SuperCENT) methodology, which we will describe formally in the next section. We name it the "supervised" centrality estimation because $(\boldsymbol{X}, \boldsymbol{y})$ in the regression (2b) can be thought of as the supervisors that offer additional supervision to the centrality estimation. It is expected that if the centralities indeed have strong predictive power (that is, the centrality regression coefficients β_u, β_v are large compared with the regression noise level), the estimation of the centralities will be better when considering both (2a) and (2b) instead of only (2a). With the improved estimation of the centralities, SuperCENT can further improve the estimation and inference of the regression model.

Remark 1. (Multi-rank unified framework) The unified framework can be extended to a multi-rank version, by substituting the rank-one network model (2a) with a multi-rank network model with non-diverging rank $r \leqslant n$, i.e.,

$$\boldsymbol{A} = \boldsymbol{A}_0 + \boldsymbol{E} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top + \boldsymbol{E} = d\boldsymbol{u}\boldsymbol{v}^\top + \sum_{l=2}^{r} d_l \boldsymbol{u}_l \boldsymbol{v}_l^\top + \boldsymbol{E}, \tag{3}$$

where $\boldsymbol{D}$ is a diagonal matrix of dimension $r \times r$ with the singular values $d > d_2 \geqslant \ldots \geqslant d_r \geqslant 0$ as the diagonal entries, and $\boldsymbol{U} = (\boldsymbol{u}, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r)$ and $\boldsymbol{V} = (\boldsymbol{v}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r)$ are two matrices of size $n \times r$ with orthogonal columns of length $\sqrt{n}$. The hub and authority centralities,

$\boldsymbol{u}$ and $\boldsymbol{v}$, still correspond to the leading left and right singular vector respectively. The regression model (2b) remains unchanged and only includes $\boldsymbol{u}$ and $\boldsymbol{v}$ instead of the entire $\boldsymbol{U}$ and $\boldsymbol{V}$ because it is common practice to consider the network effect via only the centralities.

Remark 2. (Model identifiability) Note that $\boldsymbol{u}, \boldsymbol{v}$ are only identifiable up to a scalar. SVD assumes $\boldsymbol{u}, \boldsymbol{v}$ have unit length. However, we assume $\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = \sqrt{n}$, in view of the fact that the network can grow and consequently the centralities should roughly be on the same scale with the network. This prevents the centrality regression coefficients from exploding as the network grows. We further assume $n > p + 2$ and $\boldsymbol{X}$ is full rank.

3 Methodology

Sections 3.1-3.2 formally introduce the two-stage procedure and SuperCENT respectively. Section 3.3 is devoted to the prediction problem when new nodes are added to the network together with their covariates. Supplement S3 contains tuning parameter selection and SuperCENT for multi-rank networks and undirected networks with eigenvector centrality.

3.1 The two-stage procedure

As mentioned in the introduction, given the unified framework (2) and the observed data $\{\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{y}\}$, a natural and ad-hoc procedure is the two-stage estimator, which can serve as a benchmark. In view of (2a), the first stage is to perform SVD on the observed adjacency matrix $\boldsymbol{A}$ and take its leading left and right singular vectors and rescale them to have length $\sqrt{n}$, denoted as $\hat{\boldsymbol{u}}^{ts}$ and $\hat{\boldsymbol{v}}^{ts}$, as the estimates for the centralities $\boldsymbol{u}$ and $\boldsymbol{v}$ respectively. The superscript ts stands for **t**wo-**s**tage. In view of (2b), given the estimates $\hat{\boldsymbol{u}}^{ts}$ and $\hat{\boldsymbol{v}}^{ts}$, the second stage performs the ordinary least square (OLS) regression of $\boldsymbol{y}$ on $\boldsymbol{X}$ and $\hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts}$, treating $\hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts}$ as fixed covariates.

Hence, the two-stage procedure solves the following two optimizations *sequentially*,

$$
\begin{cases}
(\hat{d}^{ts}, \hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts}) := \underset{d, \|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = \sqrt{n}}{\arg\min} \|\boldsymbol{A} - d\boldsymbol{u}\boldsymbol{v}^\top\|_F^2, & \text{(4a)} \\
\hat{\boldsymbol{\beta}}^{ts} := ((\hat{\boldsymbol{\beta}}_x^{ts})^\top, \hat{\beta}_u^{ts}, \hat{\beta}_v^{ts})^\top := \underset{\boldsymbol{\beta}_x, \beta_u, \beta_v}{\arg\min} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_x - \hat{\boldsymbol{u}}^{ts}\beta_u - \hat{\boldsymbol{v}}^{ts}\beta_v\|_2^2. & \text{(4b)}
\end{cases}
$$

It follows that $\hat{\boldsymbol{\beta}}^{ts} = (\widehat{\mathbf{W}}^\top\widehat{\mathbf{W}})^{-1}\widehat{\mathbf{W}}^\top\mathbf{y}$, where $\widehat{\mathbf{W}} = (\mathbf{X}, \hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts})$.

Remark 3. (Two-stage "ad-hoc" confidence interval (CI)) Besides the estimation of the unknown parameters, valid inference is necessary to evaluate the network effect. Empirical studies usually construct CIs of the regression coefficients from the second stage regression by assuming that $\hat{\boldsymbol{u}}^{ts}$ and $\hat{\boldsymbol{v}}^{ts}$ are fixed and noiseless. This assumption simplifies the inferential statement, because it follows that $\text{cov}(\hat{\boldsymbol{\beta}}^{ts}) = \sigma_y^2(\widehat{\mathbf{W}}^\top\widehat{\mathbf{W}})^{-1}$, where $\widehat{\mathbf{W}} = (\mathbf{X}, \hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts})$. However, the observed network $\boldsymbol{A}$ is one realization from $\boldsymbol{A}_0 + \boldsymbol{E}$ as in Model (2a), which makes its singular vectors $\hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts}$ random. If one ignores this randomness, the inference becomes invalid. We refer to such "ad-hoc" CI as the "two-stage-adhoc" method. To correct for the randomness of the estimated singular vectors $\hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts}$ and obtain valid inference, Section 4 derives the asymptotic distribution of the two-stage estimator and discusses the theoretical property of the naive two-stage-adhoc CI. Section 5 shows that the two-stage-adhoc CI is either conservative or invalid, depending on the network noise level.

3.2 SuperCENT methodology

In the two-stage procedure, the estimation of the regression model in Step 2 depends on the centrality estimation in Step 1. The more accurate the centrality estimates are, the better we are able to make inference in the regression model. On the other hand, the centralities are incorporated in the regression model as regressors, so $(\boldsymbol{X}, \boldsymbol{y})$ can supervise centrality estimation and thus boost the estimation accuracy.

Motivated by the above intuition, we propose to optimize the following objective function to obtain the SuperCENT estimates,

$$(\hat{d}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}, \widehat{\boldsymbol{\beta}}_x, \hat{\beta}_u, \hat{\beta}_v) := \mathop{\arg\min}_{\substack{\boldsymbol{\beta}_x, \beta_u, \beta_v \\ d, \|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = \sqrt{n}}} \frac{1}{n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_x - \boldsymbol{u}\beta_u - \boldsymbol{v}\beta_v\|_2^2 + \frac{\lambda}{n^2}\|\boldsymbol{A} - d\boldsymbol{u}\boldsymbol{v}^\top\|_F^2, \quad (5)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix. The above objective function combines the residual sum of squares (4b) and the rank-one approximation error of the observed network (4a). The connection between the two terms is the centralities. The trade-off between them can be tuned through a proper selection of the hyper-parameter λ.

To solve (5), we use a block gradient descent algorithm by updating $(\hat{d}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}, \widehat{\boldsymbol{\beta}})$ iteratively until convergence, where $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_x^\top, \hat{\beta}_u, \hat{\beta}_v)^\top$. The initialization can be the first stage of the two-stage procedure, i.e., $(\hat{d}^{ts}, \hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts})$. The complete algorithm with a given tuning parameter λ is shown in Algorithm S1 of Supplement S3.1.

Note that although $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ with length $\sqrt{n}$ are only identifiable up to the sign, $\hat{\boldsymbol{u}}\hat{\boldsymbol{v}}^\top$, $\hat{\boldsymbol{u}}\hat{\beta}_u$ and $\hat{\boldsymbol{v}}\hat{\beta}_v$ are uniquely identifiable. One can determine the sign of all the parameters as follows: find the entry that has the largest magnitude in $(\hat{\boldsymbol{u}}^\top, \hat{\boldsymbol{v}}^\top)$, and then make that entry positive and determine the signs of the rest in $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ accordingly.

On a separate note, under the multi-rank unified framework presented in Remark 1, the same SuperCENT methodology applies without any modification (see Section S3.2). Simulation studies show similar results as the rank-one setting (see Section S5.4).

3.3 Prediction

Once the model is fitted with training data, it can be used for prediction. Suppose there are n^* new observations, which includes the new covariates $\boldsymbol{X}^*$, the new network among themselves $\boldsymbol{A}^*$, as well as the new edges connecting them with the n training observations. The original network $\boldsymbol{A}$ is then augmented to $\boldsymbol{A}^{all}$ as follows:

$$\boldsymbol{A}^{all} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}^* \end{pmatrix} = d \begin{pmatrix} \boldsymbol{u}\boldsymbol{v}^\top & \boldsymbol{u}\boldsymbol{v}^{*\top} \\ \boldsymbol{u}^*\boldsymbol{v}^\top & \boldsymbol{u}^*\boldsymbol{v}^{*\top} \end{pmatrix} + \begin{pmatrix} \boldsymbol{E}_{11} & \boldsymbol{E}_{12} \\ \boldsymbol{E}_{21} & \boldsymbol{E}_{22} \end{pmatrix}. \tag{6}$$

The above expression is obtained from assuming (2a) for $\boldsymbol{A}^{all}$, i.e.

$$\boldsymbol{A}^{all} = d\boldsymbol{u}^{all}\boldsymbol{v}^{all\top} + \boldsymbol{E}^{all}, \tag{7}$$

where $\boldsymbol{u}^{all} = (\boldsymbol{u}^\top, \boldsymbol{u}^{*\top})^\top$ and $\boldsymbol{v}^{all} = (\boldsymbol{v}^\top, \boldsymbol{v}^{*\top})^\top$. Given the regression equation $\hat{\boldsymbol{y}}^* = \boldsymbol{X}^*\widehat{\boldsymbol{\beta}}_x + \boldsymbol{u}^*\hat{\beta}_u + \boldsymbol{v}^*\hat{\beta}_v$, to predict $\boldsymbol{y}^*$, $\boldsymbol{u}^*$ and $\boldsymbol{v}^*$ needs to be estimated. One can either perform SVD on $\boldsymbol{A}^*$ or SVD on $\boldsymbol{A}^{all}$ and reserve only the relevant components of the singular vectors. The latter approach is more accurate and is formally described in

Algorithm S3 of the Supplement.

Since $\hat{\boldsymbol{u}}^{all}$ and $\hat{\boldsymbol{v}}^{all}$ are only identifiable up to sign, we determine their signs in Step 2 of Algorithm S3 such that the angles between the training proportions and the SuperCENT estimates are less than 90 degrees. Recall that for identifiability, $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$ are of length $\sqrt{n}$, and $\hat{\beta}_u$, $\hat{\beta}_v$ are of the corresponding scale. In prediction, we need to scale $\hat{\boldsymbol{u}}^*$ and $\hat{\boldsymbol{v}}^*$ accordingly so that $\hat{\beta}_u\hat{\boldsymbol{u}}^* + \hat{\beta}_v\hat{\boldsymbol{v}}^*$ is on par with $\hat{\beta}_u\hat{\boldsymbol{u}} + \hat{\beta}_v\hat{\boldsymbol{v}}$ (See Step 3 of Algorithm S3).

4 Theoretical properties

We investigate the statistical properties of the two-stage procedure in Section 4.1 and SuperCENT in Section 4.2. We start with introducing notations and assumptions. Let

$$\delta_{ts,sc} = (\lambda d^2 + \beta_u^2 + \beta_v^2)^{-2}\left[\frac{2\lambda d^2 + \beta_u^2 + \beta_v^2}{d^2 n}\sigma_a^2 - \sigma_y^2\right], \tag{8}$$

which depends on the network signal and noise strengths d, σ_a, the regression signal and noise levels $\beta_u, \beta_v, \sigma_y$, the sample size n, and the tuning parameter λ. As we will show later, $\delta_{ts,sc}$ is a crucial quantity that measures the discrepancy between the two-stage (ts) and SuperCENT (sc).

The properties of two-stage and SuperCENT will be studied under a different subset of the following assumptions respectively.

Assumption 1. The network noise $\boldsymbol{E}$ and regression noise $\boldsymbol{\epsilon}$ have independent normal entries with mean 0 and variance σ_a^2 and σ_y^2 respectively, and they are independent.

Assumption 2. The fixed design matrix in the regression $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ satisfies $n > p + 2$, $\boldsymbol{X}^\top\boldsymbol{X}$ is invertible, and the dimension p is non-diverging.

Assumption 3. The scaled network noise-to-signal ratio $\kappa := \frac{\sigma_a^2}{d^2 n} \to 0$.

Assumption 4. (i) The scaled signal-and-noise relationship for the hub centrality from both network and regression $(\kappa - \beta_u^2\delta_{ts,sc}) \to 0$; (ii) The scaled signal-and-noise relationship for the authority centrality from both network and regression $(\kappa - \beta_v^2\delta_{ts,sc}) \to 0$.

In Assumption 1, the independence is assumed for simplicity. If the network noises e_{ij}'s or the regression noises ϵ_i's are dependent with known covariance, the theorems and the corollaries below still hold with slight modifications by simply plugging their covariance matrices into appropriate places; if they are dependent with unknown covariance, extra assumptions on the covariance structure need to be made and new methodologies and theories should be developed. Assumption 2 simply states that the regression is in the conventional low-dimensional fixed-design regime.

Assumption 3 is required for the consistency of the two-stage, which essentially requires the signal-to-noise ratio (SNR) of the network to be large enough. Otherwise, the two-stage centrality estimation is inconsistent and the regression coefficient estimation is biased. Assumption 4 is required for the consistency of SuperCENT and it is less restrictive in general than Assumption 3, which will be further explained in Remark 6. Therefore, to achieve consistency, SuperCENT requires *weaker* network SNR than the two-stage.

Throughout this section, we focus on the unified framework with the rank-one network model (2), whose properties are more comprehensible. The same proof strategy can be applied under the multi-rank unified framework.

4.1 Theoretical properties of the two-stage procedure

In this section, we summarize the theoretical properties of the two-stage estimator. Additional theorems and interpretations are given in Supplement S4.1.

Under Assumptions 1-3, the two-stage estimators are consistent, with asymptotic distributions in Theorem S1 and convergence rates in Corollaries S1-S2. They have the following implications on centrality estimation and inference of the regression: 1. For centrality estimation, Corollary S1 states that the convergence rate of two-stage is essentially the network noise-to-signal ratio κ. Since real networks tend to have diverging noise level or shrinking signal strength as the network grows, the two-stage estimators will be thus inconsistent; 2. For regression inference, Theorem S1 and Corollary S2 show that the asymptotic covariance of $\widehat{\boldsymbol{\beta}}^{ts}$ involves both σ_y^2 and σ_a^2, not just σ_y^2, which signifies that the two-stage-adhoc CI is invalid as it assumes error-less $(\hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts})$.

Corollary 1 states that, when Assumption 3 is violated, the two-stage estimator is inconsistent in Stage 1, leading to estimation biases for the regression coefficients in Stage 2.

Corollary 1. *(Bias of $\hat{\beta}_u^{ts}$, $\hat{\beta}_v^{ts}$ when two-stage is inconsistent) Let $\rho = \mathrm{cor}(\boldsymbol{u}, \boldsymbol{v})$. If $\boldsymbol{\beta}_x = \mathbf{0}$ or $\mathrm{cov}(\boldsymbol{X}, (\boldsymbol{u}\boldsymbol{v})) = \mathbf{0}_{p\times 2}$, then under the unified framework (2) and Assumptions 1-2,*

$$\mathrm{plim}\ \hat{\beta}_u^{ts} = \frac{(1+\kappa-\rho^2)\beta_u + \kappa\rho\beta_v}{(1+\kappa)^2-\rho^2} \quad \text{and} \quad \mathrm{plim}\ \hat{\beta}_v^{ts} = \frac{(1+\kappa-\rho^2)\beta_v + \kappa\rho\beta_u}{(1+\kappa)^2-\rho^2}. \tag{9}$$

Corollary 1 explicates the directions of biases in $\hat{\beta}_u^{ts}$ and $\hat{\beta}_v^{ts}$ and has consequences on the inference, resembling the measurement error problem which we review in Remark S6. The following remarks discuss the a few special cases of the corollary and the two-stage ad-hoc confidence interval (CI) of β_u. Remark S5 discusses the conditions in the corollary.

Remark 4. (Directions of the biases of $\hat{\beta}_u^{ts}$ and $\hat{\beta}_v^{ts}$) For Corollary 1 there are some special cases when Assumption 3 is violated, i.e., $\kappa \nrightarrow 0$. (i) When the two true centralities are uncorrelated, $\rho = 0$, $\mathrm{plim}\ \hat{\beta}_u^{ts} = \frac{1}{1+\kappa}\beta_u$ and $\mathrm{plim}\ \hat{\beta}_v^{ts} = \frac{1}{1+\kappa}\beta_v$. The OLS estimate is biased towards zero, and the degree of bias depends on the attenuation factor, $\frac{1}{1+\kappa}$, similar to the classical measurement error results. The classical results, however, are derived under simpler assumptions; hence they are not directly applicable under our unified framework. Note that if $\kappa \to 0$, $\mathrm{plim}\ \hat{\beta}_u^{ts} = \beta_u$ and $\mathrm{plim}\ \hat{\beta}_v^{ts} = \beta_v$; but the traditional estimate of σ_y^2, $\frac{RSS}{n-1}$, overestimates σ_y^2. (ii) When $\rho \neq 0$, if $|\beta_u| \gg |\beta_v|$, then $\mathrm{plim}\ \hat{\beta}_u^{ts} \approx \frac{(1+\kappa-\rho^2)\beta_u}{(1+\kappa)^2-\rho^2}$, which is equivalent to $\mathrm{plim}\ \hat{\beta}_u^{ts} - \beta_u \approx -\frac{(1+\kappa)\kappa}{(1+\kappa)^2-\rho^2}\beta_u$. This implies that $\hat{\beta}_u^{ts}$ has attenuation bias. As for $\hat{\beta}_v^{ts}$, $\mathrm{plim}\ \hat{\beta}_v^{ts} - \beta_v \approx \frac{\rho\kappa}{(1+\kappa)^2-\rho^2}\beta_u$, which implies that $\hat{\beta}_v^{ts}$ is biased away from zero if $\mathrm{sign}(\beta_u) = \mathrm{sign}(\beta_v)$. (iii) When β_u and β_v have similar size, the directions of biases depend on the relationships of β_u, β_v, ρ and κ. For $\hat{\beta}_u^{ts}$, the asymptotic bias is $\mathrm{plim}\ \hat{\beta}_u^{ts} - \beta_u = \frac{-\kappa[(1+\kappa)\beta_u - \rho\beta_v]}{(1+\kappa)^2-\rho^2}$. Because $\kappa > 0$ and $0 < \rho < 1$, the denominator is always larger than 0; thus, the direction of the bias depends on the sign of $(1+\kappa)\beta_u - \rho\beta_v$: when $\beta_u > \frac{\rho}{1+\kappa}\beta_v$, $\mathrm{plim}\ \hat{\beta}_u^{ts} - \beta_u < 0$; when $\beta_u < \frac{\rho}{1+\kappa}\beta_v$, $\mathrm{plim}\ \hat{\beta}_u^{ts} - \beta_u > 0$. Similar conclusions can be drawn for $\hat{\beta}_v^{ts}$.

Remark 5. (Comments on the two-stage "ad-hoc" CI of β_u) Remarks 4 and S6 have a few implications on the CI for β_u when the two-stage estimator is consistent ($\kappa \to 0$). (i) When all the quantities in the asymptotic variance of $\hat{\beta}_u^{ts}$ (S28)-(S29) are known, both terms are needed to construct valid inference. Using (S28) alone, i.e., the "two-stage-adhoc" method obtained via software directly from the regression in Stage 2, yields invalid inference unless $\sigma_a = 0$ or $\tilde{\boldsymbol{u}} \perp \tilde{\boldsymbol{v}}$ as discussed in Remark S3. (ii) When the quantities are unknown and need to be estimated, the inference based on the two estimated terms of (S28)-(S29) will be unnecessarily wide, thereby even more conservative. This is because the inference based on the first estimated term (S28) alone is already conservative and $\hat{\sigma}_y^2$ further over-estimates σ_y^2 in (S29). On the other hand, when $\kappa \nrightarrow 0$, the two-stage inference is invalid.

4.2 Theoretical properties of SuperCENT

Denote SuperCENT estimators from Algorithm S1 with a given tuning parameter λ as $\hat{d}$, $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$, and $\widehat{\boldsymbol{\beta}} = ((\widehat{\boldsymbol{\beta}}_x)^\top, \hat{\beta}_u, \hat{\beta}_v)^\top$. Let $\widehat{\boldsymbol{A}} = \hat{d}\hat{\boldsymbol{u}}\hat{\boldsymbol{v}}^\top$ be the SuperCENT estimate of $\boldsymbol{A}_0$.

Theorem 1. *Under the unified framework (2) and Assumptions 1, 2 and 4, the SuperCENT estimators converge to the following normal distributions asymptotically,*

1. *Centralities: for each $i = 1, \ldots, n$*

$$(\hat{\boldsymbol{u}}_i - \boldsymbol{u}_i) \xrightarrow{\mathcal{D}} N\Big(0, \boldsymbol{\Sigma}_{\boldsymbol{u},ii}\Big) \quad \text{and} \quad (\hat{\boldsymbol{v}}_i - \boldsymbol{v}_i) \xrightarrow{\mathcal{D}} N\Big(0, \boldsymbol{\Sigma}_{\boldsymbol{v},ii}\Big); \tag{10}$$

2. *Network: for each $i, j = 1, \ldots, n$*

$$\Big(\widehat{\boldsymbol{A}}_{ij} - \boldsymbol{A}_{0,ij}\Big) \xrightarrow{\mathcal{D}} N\Big(0, \boldsymbol{\Sigma}_{\boldsymbol{A}_{0,i+n(j-1),i+n(j-1)}}\Big); \tag{11}$$

3. *Network effect:*

$$\Big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\Big) \xrightarrow{\mathcal{D}} N\Big(\mathbf{0}_{p+2}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\Big), \tag{12}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{u}}, \boldsymbol{\Sigma}_{\boldsymbol{v}}, \boldsymbol{\Sigma}_{\boldsymbol{A}_0}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ are functions of $(\sigma_a, d, \boldsymbol{u}, \boldsymbol{v}, \sigma_y, \boldsymbol{X}, \beta_u, \beta_v, \lambda)$, whose specific forms are given in Supplement S4.3.

Comparing the covariance matrices with those of the two-stage in Theorem S1, all $\boldsymbol{\Sigma}_{\boldsymbol{u}}, \boldsymbol{\Sigma}_{\boldsymbol{v}}, \boldsymbol{\Sigma}_{\boldsymbol{A}_0}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ involve both σ_a^2 and σ_y^2 due to the simultaneous estimation, while

$\boldsymbol{\Sigma}^{ts}_{\boldsymbol{u}}, \boldsymbol{\Sigma}^{ts}_{\boldsymbol{v}}$ and $\boldsymbol{\Sigma}^{ts}_{\boldsymbol{A}_0}$ of the two-stage only involve σ_a^2 and $\boldsymbol{\Sigma}^{ts}_{\boldsymbol{\beta}}$ involves both. Following Theorem 1, Corollaries 2-3 provide the convergence rates of the network-related quantities ($\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$, $\widehat{\boldsymbol{A}}$) and the regression coefficient estimates ($\widehat{\boldsymbol{\beta}}$) respectively.

Corollary 2. *(Convergence rates of $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$ and $\widehat{\boldsymbol{A}}$) Under the unified framework (2) and Assumptions 1, 2 and 4, the SuperCENT estimators satisfy the following,*

$$\frac{1}{n}\mathbb{E}\|\hat{\boldsymbol{u}} - \boldsymbol{u}\|_2^2 = \left(\frac{\sigma_a^2(n-1)}{d^2n^2} - \frac{n-p-2}{n}\beta_u^2\delta_{ts,sc}\right)(1+o(1)) = O\left(\kappa - \beta_u^2\delta_{ts,sc}\right), \tag{13}$$

$$\frac{1}{n}\mathbb{E}\|\hat{\boldsymbol{v}} - \boldsymbol{v}\|_2^2 = \left(\frac{\sigma_a^2(n-1)}{d^2n^2} - \frac{n-p-2}{n}\beta_v^2\delta_{ts,sc}\right)(1+o(1)) = O\left(\kappa - \beta_v^2\delta_{ts,sc}\right), \tag{14}$$

$$\mathbb{E}\frac{\|\widehat{\boldsymbol{A}} - \boldsymbol{A}_0\|_F^2}{\|\boldsymbol{A}_0\|_F^2} = \left(\frac{\sigma_a^2(2n-1)}{d^2n^2} - \frac{n-p-2}{n}\left(\beta_u^2+\beta_v^2\right)\delta_{ts,sc}\right)(1+o(1)) = O\left(\kappa - \left(\beta_u^2+\beta_v^2\right)\delta_{ts,sc}\right) \tag{15}$$

Corollary 3. *(Asymptotic property of $\widehat{\boldsymbol{\beta}}$) Under the unified framework (2) and Assumptions 1, 2 and 4, the SuperCENT estimators satisfy the following,*

$$\mathbb{E}(\hat{\beta}_u - \beta_u)^2 = \mathbb{E}(\hat{\beta}_u^{ts} - \beta_u)^2 = O\left(\frac{\sigma_y^2}{n} + \frac{\sigma_a^2(\beta_u^2+\beta_v^2)}{d^2n^2}\right),$$

$$\mathbb{E}(\hat{\beta}_v - \beta_v)^2 = \mathbb{E}(\hat{\beta}_v^{ts} - \beta_v)^2 = O\left(\frac{\sigma_y^2}{n} + \frac{\sigma_a^2(\beta_u^2+\beta_v^2)}{d^2n^2}\right),$$

$$Cov\left(\widehat{\boldsymbol{\beta}}_x - \boldsymbol{\beta}_x\right) = Cov\left(\widehat{\boldsymbol{\beta}}_x^{ts} - \boldsymbol{\beta}_x\right).$$

The two-stage estimator $\hat{\boldsymbol{\beta}}^{ts}$ achieves the same rate under Assumptions 1, 2 and 3.

Remark 6. (The role of $\delta_{ts,sc}$) Comparing Corollaries S1 and 2, the discrepancies between the two-stage and SuperCENT estimators of the centralities and the true network are all proportional to $\delta_{ts,sc}$.The consistency of the two-stage requires Assumption 3, i.e., $\kappa \to 0$, whilst the consistency of SuperCENT requires Assumption 4, i.e., $(\kappa - \beta_u^2\delta_{ts,sc}) \to 0$ and $(\kappa - \beta_v^2\delta_{ts,sc}) \to 0$. It can be seen that, whenever $\delta_{ts,sc} > 0$, SuperCENT always outperforms the two-stage. When both Assumptions 3-4 hold, two-stage and SuperCENT are both consistent, but SuperCENT converges faster; when Assumption 3 is violated and Assumption 4 holds, SuperCENT is consistent while the two-stage is not.

The positiveness of $\delta_{ts,sc}$ requires $\frac{2\lambda d^2+\beta_u^2+\beta_v^2}{d^2n}\sigma_a^2 - \sigma_y^2 > 0$, which depends on the interplay of $(\sigma_a, d, \sigma_y, \beta_u, \beta_v, n, \lambda)$. Specifically, $\delta_{ts,sc}$ is positive, when the signal of the regression

β_u, β_v is large, the regression noise σ_y is small, the signal of the network d is small, or the network noise σ_a is large. This exactly verifies our intuition: when the regression SNR is high, we gain information from the regression to assist centrality estimation; and the advantage is more pronounced when the network SNR is weak, which is exactly when the two-stage becomes inconsistent while SuperCENT remains consistent. Moreover, $\delta_{ts,sc}$ involves a tuning parameter λ, and is positive when λ is large enough. This is especially true when λ takes the optimal value $\lambda_0 = n\sigma_y^2/\sigma_a^2$ given in the remark below.

Remark 7. (Optimal λ) Minimizing the convergence rates (13), (14), or (15) with respect to λ leads to the optimal tuning parameter $\lambda_0 = \frac{n\sigma_y^2}{\sigma_a^2}$. With the optimal λ_0, SuperCENT achieves its best performance and obtains the most improvement over the two-stage. Plugging the optimal λ_0 into (8), we obtain the discrepancy $\delta_{ts,sc} = \frac{\frac{\kappa^2}{\sigma_y^2}}{1+\kappa\left(\frac{\beta_u^2}{\sigma_y^2}+\frac{\beta_v^2}{\sigma_y^2}\right)}$, which is always positive. This implies that as long as the tuning parameter is properly selected, SuperCENT will always be superior over the two-stage.

Remark 8. (SuperCENT-$\hat{\lambda}_0$ and SuperCENT-$\hat{\lambda}_{cv}$) The benefit of the optimal value λ_0 is two-fold: 1) to benchmark the cross-validation (CV) procedure in Algorithm SS4; 2) to provide a candidate for the tuning parameter λ by plugging in the two-stage estimates of σ_y^2 and σ_a^2, i.e., $\hat{\lambda}_0 = n(\hat{\sigma}_y^{ts})^2/(\hat{\sigma}_a^{ts})^2$, instead of the time-consuming cross-validation. We refer to the SuperCENT with $\hat{\lambda}_0$ as SuperCENT-$\hat{\lambda}_0$. Furthermore, $\hat{\lambda}_0$ can be used as a guide to lay out the cross-validation grid points in Algorithm SS4, to obtain $\hat{\lambda}_{cv}$ and SuperCENT-$\hat{\lambda}_{cv}$.

Remark 9. (Comparison of the estimation of $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{A}_0$ when two-stage is *inconsistent*) For the two-stage procedure, $\hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts}, \widehat{\boldsymbol{A}}^{ts}$ is consistent if and only if $\kappa = \frac{\sigma_a^2}{d^2 n} \to 0$, which implies the network SNR has to be large enough for the two-stage to be consistent. When $\kappa = O(1)$, the two-stage procedure is inconsistent. Can the SuperCENT estimates remain consistent under this regime? The answer is positive.

Plugging in the optimal λ_0, the rate of $\mathbb{E}\|\hat{\boldsymbol{u}} - \boldsymbol{u}\|_2^2/n$ in (13) becomes

$$\kappa \frac{1+\kappa\frac{\beta_v^2}{\sigma_y^2}}{1+\kappa\left(\frac{\beta_u^2}{\sigma_y^2}+\frac{\beta_v^2}{\sigma_y^2}\right)}, \tag{16}$$

which is obviously faster than the rate of $\mathbb{E}\|\hat{\boldsymbol{u}}^{ts}-\boldsymbol{u}\|_2^2/n\kappa$ in (S26). We want the above rate to converge to 0 when $\kappa = O(1)$. Given (16), the convergence of $\hat{\boldsymbol{u}}$ boils down to the SNR of $\boldsymbol{u}$ and $\boldsymbol{v}$ in the network regression model, i.e., $\frac{\beta_u^2}{\sigma_y^2}$ and $\frac{\beta_v^2}{\sigma_y^2}$. One sufficient condition for convergence is when $\frac{\beta_u^2}{\sigma_y^2} \to \infty$ and $\frac{\beta_v^2}{\sigma_y^2} = O(1)$, which means that, to guarantee convergence of $\hat{\boldsymbol{u}}$, the signal for $\boldsymbol{u}$ has to be stronger than the signal for $\boldsymbol{v}$ and the noise σ_y.

If we want to guarantee the convergence of $\hat{\boldsymbol{v}}$ under this regime, one sufficient condition is $\frac{\beta_v^2}{\sigma_y^2} \to \infty$ and $\frac{\beta_u^2}{\sigma_y^2} = O(1)$. This conflicts with the requirement of the convergence of $\hat{\boldsymbol{u}}$. Fortunately, the rates of both $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ are faster than those of $\hat{\boldsymbol{u}}^{ts}$ and $\hat{\boldsymbol{v}}^{ts}$, so SuperCENT always improves the estimation: when $\beta_u^2/\sigma_y^2 \gg \beta_v^2/\sigma_y^2 = O(1)$ or $\beta_v^2/\sigma_y^2 \gg \beta_u^2/\sigma_y^2 = O(1)$, one of $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ will be consistent. In other words, no matter $\kappa \to 0$ holds or not, $\hat{\boldsymbol{u}}$ is consistent under Assumption 4(i) and $\hat{\boldsymbol{v}}$ is consistent under Assumption 4(ii). We will demonstrate this phenomenon in the simulation.

For the estimation of $\boldsymbol{A}_0$ using the optimal λ_0, the rate of $\mathbb{E}\|\widehat{\boldsymbol{A}}-\boldsymbol{A}_0\|_F^2/\|\boldsymbol{A}_0\|_F^2$ in (15) becomes $\kappa\left[1+\kappa\left(\frac{\beta_u^2}{\sigma_y^2}+\frac{\beta_v^2}{\sigma_y^2}\right)\right]^{-1}$, which is much faster than the rate of $\mathbb{E}\|\widehat{\boldsymbol{A}}^{ts}-\boldsymbol{A}_0\|_F^2/\|\boldsymbol{A}_0\|_F^2 = \kappa$ in (S27). Better yet, to ensure $\widehat{\boldsymbol{A}}$ is consistent, we only require either $\frac{\beta_u^2}{\sigma_y^2} \to \infty$ or $\frac{\beta_v^2}{\sigma_y^2} \to \infty$. This means that, as long as one of $\hat{\boldsymbol{u}}$ or $\hat{\boldsymbol{v}}$ is consistent, $\widehat{\boldsymbol{A}}$ is also consistent, while two-stage $\widehat{\boldsymbol{A}}^{ts}$ is only consistent when both $\hat{\boldsymbol{u}}^{ts}$ and $\hat{\boldsymbol{v}}^{ts}$ are consistent.

Remark 10. (Comparison of $\widehat{\boldsymbol{\beta}}^{ts}$ and $\widehat{\boldsymbol{\beta}}$) Note that Corollary 3 states that the asymptotic variances and covariances of the two-stage and SuperCENT estimators of $\beta_u, \beta_v, \boldsymbol{\beta}_x$ are the same. But the property for the two-stage holds under Assumption 3, while the property for the SuperCENT holds under Assumption 4. When both Assumptions 3-4 hold, i.e., both the two-stage and SuperCENT are consistent, from the perspective of regression coefficient estimation, SuperCENT and the two-stage are similar and the supervision effect of SuperCENT only takes place for the estimation of $\boldsymbol{u}, \boldsymbol{v}$ and $\boldsymbol{A}_0$. However, if Assumption

3 does not hold but Assumption 4 does (with a large regression SNR and a proper tuning parameter), the two-stage regression coefficient estimation is biased as shown in Corollary 1, but the SuperCENT regression coefficient estimates are still consistent and unbiased with the covariances shown in Corollary 3.

5 Simulation

In this section, we investigate the empirical performances, including the *estimation* and *inference properties* of the two-stage and SuperCENT estimators under various settings. Section 5.1 describes the simulation setups and Section 5.2 shows the results under the inconsistent regimes of the two-stage. Additional simulations, including the consistent regime of the two-stage, a phase-transition experiment, and robustness check under the multi-rank network model, are deferred to Supplement S5. Messages under multi-rank model remain the same as those under the rank-one model.

5.1 Simulation setup

We generate the network following (2a). The elements of $\boldsymbol{u}$ are first generated from i.i.d. $N(0,1)$ and $\boldsymbol{v} = 0.5\boldsymbol{u} + \boldsymbol{\epsilon_v}$ where $\boldsymbol{\epsilon_v}$ are generated from i.i.d. $N(0,1)$. $\boldsymbol{u}$ and $\boldsymbol{v}$ are then re-scaled to have length $\sqrt{n}$. For the regression model (2b), the regression coefficients are $\boldsymbol{\beta}_x = (1,3,5)^\top$, the design matrix $\mathbf{X}$ consists of a column of 1's and $p-1$ columns whose entries follow $N(0,1)$ independently.

For the properties of the estimators and inference, only the network SNR d/σ_a and the regression SNR $(\beta_u/\sigma_y, \beta_v/\sigma_y)$ matter. Hence, we fix $n = 2^8$, $d = 1$, and $\beta_v = 1$ and vary σ_a, σ_y, and β_u. To study the effect of the regression SNR, we consider $\sigma_y \in 2^{-4,-2,0}$ and $\beta_u \in 2^{0,2,4}$, while ensuring $\frac{\beta_u^2}{\sigma_y^2} \geqslant \frac{\beta_v^2}{\sigma_y^2}$. As the network SNR is solely controlled by σ_a, we vary σ_a to cover both regimes when the two-stage estimator is consistent or inconsistent. Specifically, under the *consistent regime of the two-stage*, i.e., when the network noise-to-signal ratio $\kappa = \frac{\sigma_a^2}{d^2 n} \to 0$, we consdier $\sigma_a \in 2^{-4,-2}$ to keep $\kappa < 2^{-12}$; under the *inconsistent regime of the two-stage*, i.e., $\kappa = O(1)$, we consider $\sigma_a \in 2^{0,2}$ so that $\kappa \in 2^{-8,-4} = O(1)$.

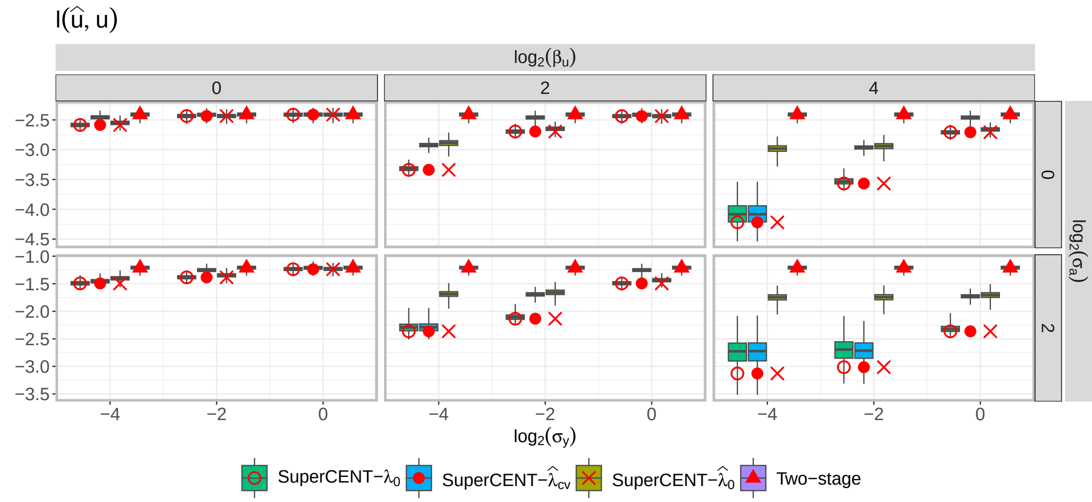
The simulation results under the inconsistent regime are presented below in this section, while those under the consistent regime are given in Supplement S5.1.

For estimation property, we compare the following procedures: 1. **Two-stage**; 2. **SuperCENT-λ_0**, which implements Algorithm S1 with oracle $\lambda_0 = n\sigma_y^2/\sigma_a^2$ using the true σ_y, σ_a and serves as the benchmark; 3. **SuperCENT-$\hat{\lambda}_0$** is SuperCENT with estimated tuning parameter $\hat{\lambda}_0 = n(\hat{\sigma}_y^{ts})^2/(\hat{\sigma}_a^{ts})^2$, where $(\hat{\sigma}_y^{ts})^2 = \frac{1}{n-p-2}\|\hat{\boldsymbol{y}}^{ts} - \boldsymbol{y}\|_2^2$ and $(\hat{\sigma}_a^{ts})^2 = \frac{1}{n^2}\|\widehat{\boldsymbol{A}}^{ts} - \boldsymbol{A}_0\|_F^2$ are estimated from the two-stage procedure; 4. **SuperCENT-$\hat{\lambda}_{cv}$** is SuperCENT with tuning parameter $\hat{\lambda}_{cv}$ chosen by cross-validation as in Algorithm SS4.

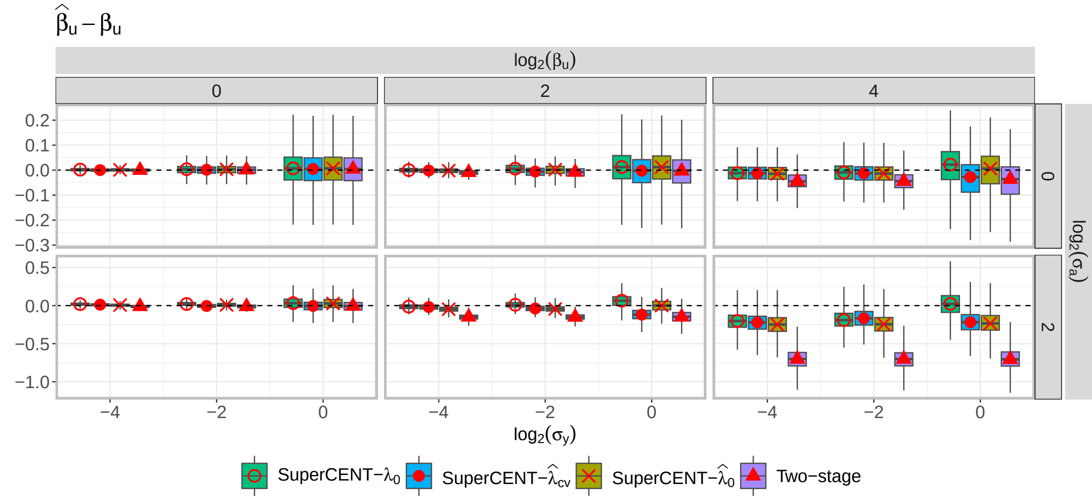
For inference property, we consider the following procedures to construct the confidence intervals (CIs) for the regression coefficient: 1. **Two-stage-adhoc**: $\hat{\beta}^{ts} \pm z_{1-\alpha/2}\hat{\sigma}^{OLS}(\hat{\beta}^{ts})$, where $z_{1-\alpha/2}$ denote the $(1-\alpha/2)$-quantile of the standard normal distribution, $\hat{\beta}^{ts}$ is the two-stage estimate of β and $\hat{\sigma}^{OLS}(\hat{\beta}^{ts})$ is the standard error from OLS, assuming $\hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts}$ are fixed predictors; 2. **Two-stage-oracle**: $\hat{\beta}^{ts} \pm z_{1-\alpha/2}\sigma(\hat{\beta}^{ts})$, where $\sigma(\hat{\beta}^{ts})$ is the standard error of $\hat{\beta}^{ts}$, whose mathematical expressions are given in (S28)-(S29) or (S31)-(S32) with the true parameters plugged in; 3. **Two-stage-plugin**: $\hat{\beta}^{ts} \pm z_{1-\alpha/2}\hat{\sigma}(\hat{\beta}^{ts})$, where $\hat{\sigma}(\hat{\beta}^{ts})$ is the standard error of $\hat{\beta}^{ts}$ by plugging all the two-stage estimators into (S28)-(S29) or (S31)-(S32); 4. **SuperCENT-λ_0-oracle**: $\hat{\beta}^{\lambda_0} \pm z_{1-\alpha/2}\sigma(\hat{\beta}^{\lambda_0})$, where $\hat{\beta}^{\lambda_0}$ is the estimate of β by SuperCENT-λ_0 and $\sigma(\hat{\beta}^{\lambda_0})$ follows (S28)-(S29) or (S31)-(S32), with the true parameters plugged in; 5. **SuperCENT-$\hat{\lambda}_{cv}$**: $\hat{\beta}^{\hat{\lambda}_{cv}} \pm z_{1-\alpha/2}\hat{\sigma}(\hat{\beta}^{\hat{\lambda}_{cv}})$, where $\hat{\beta}^{\hat{\lambda}_{cv}}$ is the estimate of β by SuperCENT-$\hat{\lambda}_{cv}$ and $\hat{\sigma}(\hat{\beta}^{\hat{\lambda}_{cv}})$ is obtained by plugging the SuperCENT-$\hat{\lambda}_{cv}$ estimates into (S28)-(S29) or (S31)-(S32). The experiments are repeated 500 times.

5.2 Simulation results under the inconsistent regime of two-stage

From the perspective of estimation, we compare the following metrics: the estimation accuracy for the centralities, the network, and the regression coefficients. Let $\boldsymbol{P}$ denote the projection matrix. Figure 1 shows the loss $l(\hat{\boldsymbol{u}}, \boldsymbol{u}) = \|\boldsymbol{P}_{\hat{\boldsymbol{u}}} - \boldsymbol{P}_{\boldsymbol{u}}\|_2^2$ and the bias $\hat{\beta}_u - \beta_u$, respectively, across different σ_a, σ_y and β_u with $d = 1$ and $\beta_v = 1$. Losses such



(a) Boxplot of $\log_{10}(l(\hat{\boldsymbol{u}}, \boldsymbol{u}))$.



(b) Boxplot of the bias $\hat{\beta}_u - \beta_u$. The dashed lines correspond to no bias.

Figure 1. Inconsistent regime of two-stage: Boxplot of $\log_{10}(l(\hat{\boldsymbol{u}}, \boldsymbol{u}))$ for the four estimators across different σ_a, σ_y and β_u with fixed $d = 1$, $\beta_v = 1$. The super-imposed red symbols show the theoretical rates of the two-stage in Corollary S1 and SuperCENT in Corollary 2 in Figure 1a and the median of the bias in Figure 1b respectively.

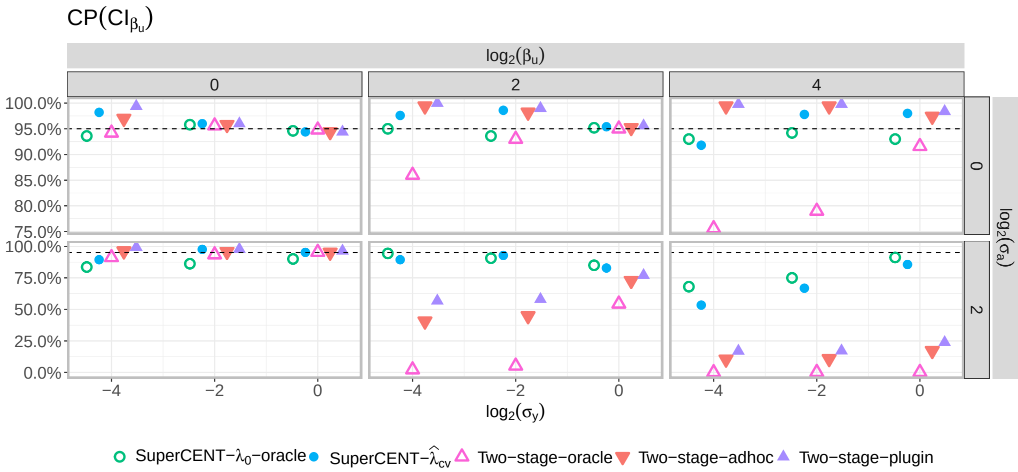
as $l(\hat{\boldsymbol{v}}, \boldsymbol{v}) = \|\boldsymbol{P}_{\hat{\boldsymbol{v}}} - \boldsymbol{P}_{\boldsymbol{v}}\|_2^2$, $l(\widehat{\boldsymbol{A}}, \boldsymbol{A}_0) = \|\widehat{\boldsymbol{A}} - \boldsymbol{A}_0\|_F^2 / \|\boldsymbol{A}_0\|_F^2$, $l(\hat{\beta}_u, \beta_u) = (\hat{\beta}_u - \beta_u)^2/\beta_u^2$, $l(\hat{\beta}_v, \beta_v) = (\hat{\beta}_v - \beta_v)^2/\beta_v^2$, and the bias $\hat{\beta}_v - \beta_v$ are given in Supplement S5.2.

Figure 1a shows the boxplot of $\log_{10}(l(\hat{\boldsymbol{u}}, \boldsymbol{u}))$. The rows correspond to $\log_2(\sigma_a)$ and the columns correspond to $\log_2(\beta_u)$. For each panel, the x-axis is $\log_2(\sigma_y)$ and the y-axis is $\log_{10}(l(\hat{\boldsymbol{u}}, \boldsymbol{u}))$. The super-imposed red symbols show the theoretical rates of $\hat{\boldsymbol{u}}^{ts}$ in Corollary S1 and that of $\hat{\boldsymbol{u}}$ in Corollary 2. As expected, the three SuperCENT-based methods estimate $\boldsymbol{u}$ much more accurately than the two-stage procedure. In particular, the supervision effect of $(\boldsymbol{X}, \boldsymbol{y})$ is more pronounced when the noise of the outcome regression (σ_y) is small, or when the signal of the outcome regression (β_u) is large, or when the network noise-to-signal $(\frac{\sigma_a}{d} = \sigma_a)$ is large. The numerical comparison validates Remarks 6 and 9 on the theoretical comparison of the estimators. Comparing the three SuperCENT-based methods, SuperCENT-$\hat{\lambda}_{cv}$ and SuperCENT-$\hat{\lambda}_0$ are sometimes worse than the benchmark SuperCENT-λ_0, but still better than the two-stage. SuperCENT-$\hat{\lambda}_0$ is typically comparable to or worse than SuperCENT-$\hat{\lambda}_{cv}$, because SuperCENT-$\hat{\lambda}_0$ fails to locate the optimal λ_0 due to inaccurate estimate of σ_a and σ_y from the two-stage procedure.

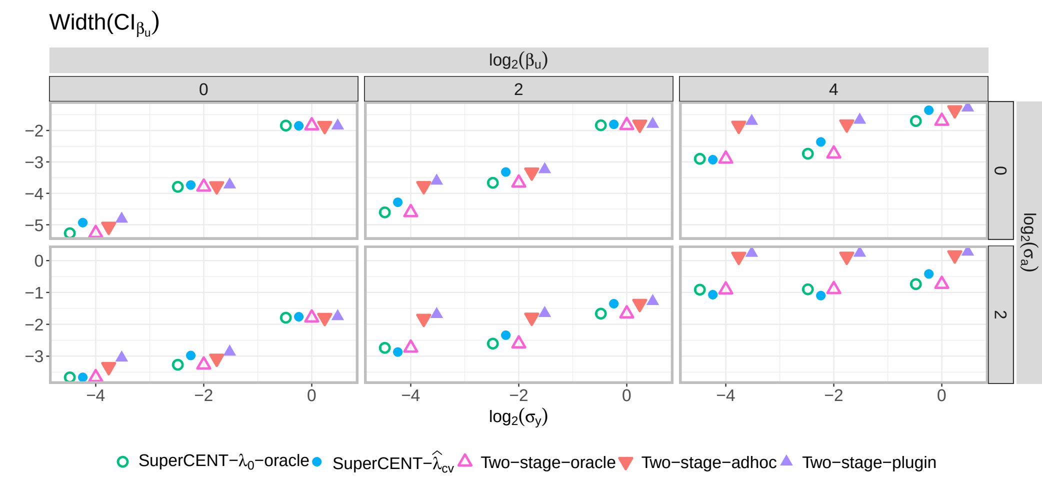
Figure 1b shows the estimation bias $\hat{\beta}_u - \beta_u$. With large σ_a or large β_u, the two-stage estimates suffer from sever attenuation bias, while SuperCENT can alleviate the bias. The attenuation bias of the two-stage can be explained by Corollary 1 and Remark 4 as follows. In this regime where $\kappa = 2^{-8,-4} \nrightarrow 0$, $\boldsymbol{u}$ and $\boldsymbol{v}$ are correlated with $\rho = \frac{2^{-1}}{\sqrt{1.25}}$, and $\beta_u \in 2^{2,4} > 1 \cdot 2^{-1.2} = \beta_v \frac{\rho}{(1+\kappa)}$, then plim $\hat{\beta}_u^{ts} - \beta_u < 0$. Hence, $\hat{\beta}_u^{ts}$ has an attenuation bias and the bias becomes larger as β_u increases.

From the perspective of inference property, Figure 2 shows the empirical coverage probability (CP) and the average width of the 95% confidence interval for β_u respectively. The CP and width for the centralities, the network, and β_v are given in the Supplement.

Figure 2a shows that the bias in the estimation of β_u by the two-stage further affects its confidence interval. For the empirical coverage, when β_u is small (leftmost column), all the methods are close to the nominal level. When β_u increases and σ_a remains small (top right two panels), all the methods (except two-stage-oracle) remain valid, though for different reasons: the two SuperCENT based methods remain valid because there is no estimation bias and the estimation of the standard error is accurate; while two-stage and two-stage-adhoc remain valid mainly because they over-estimate σ_y^2, and this conservativeness hides the issue of bias. Two-stage-oracle uses the true σ_y^2 and the issue of bias can not hide, hence the corresponding intervals undercover. When β_u increases and σ_a gets large too



(a) Empirical coverage of CI_{β_u}. The dashed lines show the nominal confidence level 0.95.



(b) $\log_{10}$ of the width of CI_{β_u}.

Figure 2. Inconsistent regime of two-stage: Empirical coverage and $\log_{10}$ of the width of CI_{β_u} across different σ_a, σ_y and β_u with $d = 1$ and $\beta_v = 1$. SuperCENT variants are labelled as circles (○ •) and the two-stage variants are labelled as triangles (△ ▼ ▲). The hollow ones are for oracles and the solid ones are for non-oracles.

(bottom right two panels), over-estimation of σ_y^2 can no longer conceal the issue of bias and all two-stage related methods become invalid. Again, SuperCENT can mitigate the bias and the CP is closer to the nominal level.

As for the width of CI_{β_u}, Figure 2b shows that the confidence intervals by the SuperCENT based methods have better coverage and are narrower than those by the two-stage methods. The improvement in width is more significant with larger β_u, σ_a.

6 Global trade network and currency risk premium

In this case study, we demonstrate that SuperCENT can provide more accurate estimation of the centralities from the global trade network. This has a profound and lucrative implication on portfolio management because the centrality is closely related to currency risk premium, i.e., the excess return from holding foreign currency compared to the US dollar. We further show the advantage of SuperCENT over the two-stage in the inference of regression coefficients, and thus strengthens a related economic theory.

In international finance literature, economists have been studying extensively the currency risk premium and puzzled by its driving forces. One recent theory, developed by Richmond (2019) using a general equilibrium, shows that countries' positions in the trade network can explain the difference in currency premiums and countries that are central in the trade network exhibit lower currency risk premiums. This theory has two implications: (i) the regression coefficients for the centralities should be negative; (ii) international investors can leverage and profit from a long-short strategy for foreign exchange – take a long position in currencies of countries with low centralities and a short position in currencies of countries with high centralities. Therefore, if the centralities can be estimated accurately, one can yield a significant investment return based on the strategy.

Motivated by Richmond (2019), we investigate how the global trade network drives the currency risk premium by regressing the currency risk premium on the centrality of the international trade network. To be specific, we consider a triplet of $\{\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{y}\}$, where $\boldsymbol{A}$ is the country-level trade network, $\boldsymbol{y}$ is the currency risk premium, and $\boldsymbol{X}$ is the share of world's GDP. Since all these quantities are not directly available, we compute them following Richmond (2019). It is worth mentioning that the trade linkage in $\boldsymbol{A}$ is defined as the trade volume normalized by the pair-wise total GDP, which represents the relative trade (export/import) intensity between two countries. We use a 5-year moving average:

when considering year t, average is taken from year $t-4$ to year t. More details are provided in Supplement S6.1. We focus on the period between 1999 and 2013 and include the 24 countries/regions whose exchange rates are available during this period.[1,2] In Figure 3, the dotted line shows the time series plot of the rank of the 5-year moving average of risk premium from 2003 to 2012 for the 24 countries/regions.[3] In each year, we rank the 24 countries/regions' risk premiums from the largest to the smallest as the 1st to 24th. We show a circular plot to visualize the average trade volume (2003-2012) in Figure S29.

Centrality estimation. Since neither the two-stage nor SuperCENT is applicable for panel data, we will repeat the analysis for each year from 2003 to 2012. Besides the network and the response variable, we also include the GDP share as a predictor, which is defined as the percentage of country/region GDP among the total GDP of all available countries in the sample for that year. In summary, the unified framework is, for each t,

$$\begin{aligned} a_{ijt} &= d \cdot \text{Hub}_{it} \times \text{Authority}_{jt} + e_{ijt}, \\ y_{it} &= \alpha + \beta_{ut} \cdot \text{Hub}_{it} + \beta_{vt} \cdot \text{Authority}_{it} + \beta_{xt} \cdot \text{GDP share}_{it} + \epsilon_{it}. \end{aligned}$$

In Sections 4 and 5, we have demonstrated that the two-stage is problematic under large network noise. In this case study, the observational error of the network comes from two sources: GDPs and the trade volumes, because each entry of the observed network a_{ijt} is defined as the trade volume normalized by their GDPs. The accounting of GDP has been a challenge in macroeconomics (Landefeld et al., 2008). For the trade volume, measurement errors are mostly due to (i) underground or illegal import and export; (ii) excluding service trade; (iii) trade cost like transportation or taxes (Lipsey, 2009). Consequently, the observed trade network can be very noisy and the two-stage will perform badly.

On the other hand, SuperCENT can significantly improve over the two-stage when the network noise is large. In what follows, we focus on SuperCENT-$\hat{\lambda}_{cv}$ using 10-fold

[1] Euro was first adopted in 1999. The bilateral trade data is available till 2013.
[2] The list of country acronyms is provided in Supplement S6.
[3] We leave the last available year 2013 for the validation purpose.

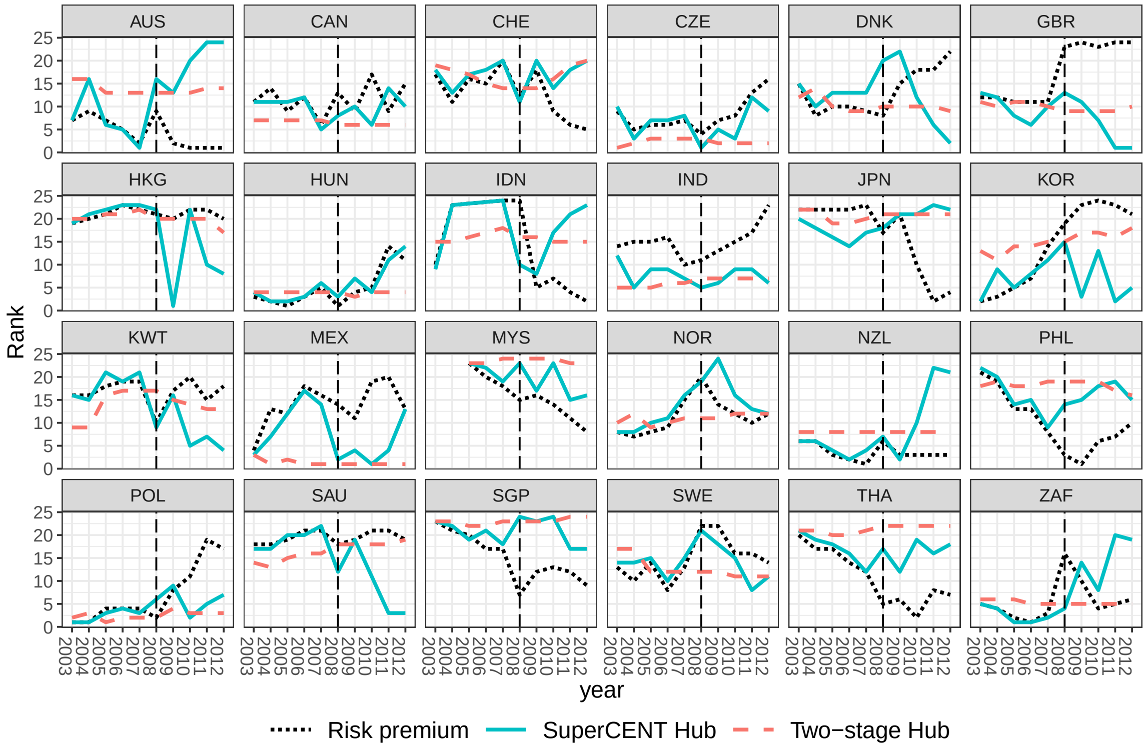


Figure 3. Time series of ranking of risk premium in descending order and ranking of hub centrality estimated by two-stage and SuperCENT in ascending order from 2003 to 2012. The vertical dashed line indicates 2008, the year of the financial crisis.

cross-validation. We will refer SuperCENT-$\hat{\lambda}_{cv}$ to SuperCENT for simplicity and use the superscript sc for all the SuperCENT-$\hat{\lambda}_{cv}$ related estimates. Figure 3 shows the time series plots of the ranking of the hub centrality estimated by two-stage and SuperCENT for the 24 countries/regions, together with the ranking of the currency risk premium. Figure S28 is for the authority centrality. We rank the centrality in ascending order and the risk premium in descending order. Based on the negative relationship between centralities and risk premium established in Richmond (2019), the closer the trends of rankings between centralities and risk premium are, the better the centralities capture the time variation in the risk premium. The centrality estimated by the two-stage procedure is relatively more stable over time compared to SuperCENT. This is because SuperCENT incorporates information of both the GDP share and the currency risk premium, which is more volatile than the trade network itself. Asian trade hubs such as Hong Kong (HKG) and Singapore (SGP) are the most central ones; while countries like South Africa (ZAF) and New Zealand

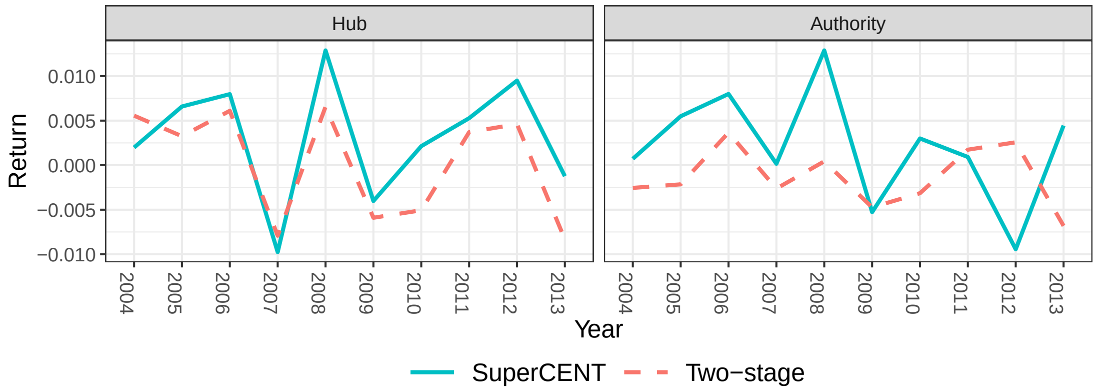


Figure 4. Time series of the next-year return from 2004 to 2013 based on a strategy that takes a long position on the currencies with the lowest 3 centralities and a short position on the currencies with the highest 3 centralities estimated from 2003 to 2012 respectively.

(NZL) are peripheral. Comparing with the ranking of risk premium, the time variation is not reflected in the centrality estimated by the two-stage procedure, while it is well captured by SuperCENT. For the 2008 financial crisis, the SuperCENT centralities fluctuate together with the risk premium while the two-stage centralities mostly remain unchanged.

To emphasize the importance of accurate centrality estimation for portfolio management, we examine whether a long-short strategy based on SuperCENT's estimated centrality can significantly boost investment performance over two-stage. For either two-stage or SuperCENT, we take a long position on the currencies with the lowest 3 centralities (bottom 10%) and a short position on the currencies with the highest 3 centralities (top 10%). We obtain a return based on the estimated centrality of the period between year $t-4$ and t. Figure 4 shows the year $t+1$ return based on this strategy. The return based on the centrality estimated by SuperCENT is much higher than that of the two-stage procedure. Table 1 shows the 10-year average return based on this strategy with the top and bottom 3, 4, and 5 currencies, respectively. The 10-year average return based on the SuperCENT centralities increased more than twice from that of the two-stage procedure.

Inference of regression. We further demonstrate the superiority of SuperCENT in inference. Again since our method is not directly applicable to longitudinal data, we take the 10-year average of trade volume and GDP to construct a 10-year trade network and

Table 1: The 10-year average return

	Top/Bottom 3		Top/Bottom 4		Top/Bottom 5	
	Hub	Authority	Hub	Authority	Hub	Authority
SuperCENT	0.0031	0.0021	0.0036	0.0019	0.0033	0.0014
Two-stage	0.0003	-0.0014	0.0008	-0.001	0.001	-0.0006
Relative difference	1 136%	253%	338%	285%	237%	320%

Table 2. The summary table of the regression comparing three methods in terms of coefficient estimation, standard error (in parenthesis) and the significant level (by asterisks).

	Two-stage-adhoc		Two-stage		SuperCENT-$\hat{\lambda}_{cv}$	
GDP share $\boldsymbol{\beta}_x$	-0.0159^{*}	(0.0083)	-0.0159^{*}	(0.0083)	-0.0162^{***}	(0.0037)
Hub β_u	-0.0011	(0.0006)	-0.0011^{*}	(0.0006)	-0.0021^{***}	(0.0002)
Authority β_v	-0.0005	(0.0006)	-0.0005	(0.0006)	-0.0003	(0.0003)
Note:						*p<0.1; **p<0.05; ***p<0.01

GDP share. Similarly, we take the 10-year average of risk premium as the response.

To better understand the behavior of the two-stage and SuperCENT estimators, it is crucial to know which regime the trade network belongs to. However, the true noise-to-signal ratio κ of the trade network is unknown, so we estimate it using SuperCENT: $\hat{\kappa}^{sc} = 0.154 \approx 2^{-3}$, which falls in the inconsistent regime of the two-stage. Note that in the simulation study, when $\kappa = 2^{-8}$, two-stage already shows inconsistency.

To further comprehend the behavior of SuperCENT and gauge how much improvement it can potentially achieve in the inconsistent regime, we estimate the SNR of the regression: $(\hat{\beta}_u^{sc}/\widehat{\sigma}_y^{sc})^2 = 7.6\times 10^6 \approx 2^{23}$ and $(\hat{\beta}_v^{sc}/\widehat{\sigma}_y^{sc})^2 = 1.8\times 10^5 \approx 2^{17}$. Compared with the simulation settings in the inconsistent regime where $\kappa = 2^{-4}$, $\beta_u^2/\sigma_y^2 \leqslant 2^{16}$ and $\beta_v^2/\sigma_y^2 \leqslant 2^8$, we expect SuperCENT to improve greatly over two-stage for both the estimation and inference of β_u, due to a large $(\hat{\beta}_u^{sc}/\widehat{\sigma}_y^{sc})^2$ and $\hat{\beta}_u^{sc} \gg \hat{\beta}_v^{sc}$ under a relatively large $\hat{\kappa}^{sc}$.

Table 2 shows the coefficient estimation, the standard error, and the significant level for the two-stage-adhoc, two-stage, and SuperCENT, respectively. For the hub centrality β_u, (i) the estimate from the two-stage methods is -0.0011, while the estimate from SuperCENT is -0.0021, which demonstrates the severe bias of two-stage in the inconsistent regime and the bias is towards zero because $|\hat{\beta}_u^{sc}| = 0.0021 \gg 0.0003 = |\hat{\beta}_v^{sc}|$;[4] (ii) the standard errors from the two-stage methods are close to 0.0006, much larger than 0.0002 from SuperCENT, which reinforces the problem of overestimation of σ_y^2 in two-stage; (iii) the above two facts combined make the confidence intervals by two-stage-adhoc and two-stage unnecessarily wide, yet still invalid: consequently the hub centrality β_u is barely significant at level 0.1 using two-stage and is insignificant using two-stage-adhoc; (iv) the two facts in (i) and (ii) also lead to a valid but narrower confidence interval for SuperCENT, making the hub centrality a significant factor at level 0.01 for the currency risk premium; (v) conclusions drawn from the two-stage-adhoc and two-stage methods contradict the theory in Richmond (2019), while SuperCENT supports the theory. Other regression coefficients' significance can be also explained by Remark 4; the details are given in Supplement S6.2.

7 Conclusion and discussion

Motivated by the rising use of centrality in empirical literature, we examined centrality estimation and inference on a noisy network (G1) as well as network effect through the centralities in the subsequent network regression (G2). We proposed a unified framework that incorporates the network generation model and the network regression model to achieve both goals. Under the unified framework, we showed that the commonly used two-stage procedure could yield inaccurate centrality estimates, biased regression coefficient estimates, and invalid inference, especially when the noise-to-signal ratio of the network is large. We proposed SuperCENT which incorporates the two models and simultaneously estimates the centralities and the effects of the centralities on the outcome. We further derived the convergence rate and the distribution of the SuperCENT estimator and provided valid confidence intervals for all the parameters of interest. We showed that SuperCENT dominates the two-stage universally and the SuperCENT estimates remain consistent under

[4]Specifically, $\hat{\beta}_u^{sc} = -0.0021 < -0.0003 \times \frac{0.673}{1+0.154} \approx -0.0002 = \hat{\beta}_v^{sc} \frac{\rho^{sc}}{1+\kappa^{sc}}$, then plim $\hat{\beta}_u^{ts} - \beta_u > 0$ as in Remark 4, and therefore the two-stage estimate is biased towards zero.

less restrictive assumptions than those required by the two-stage. The theoretical results are corroborated with extensive simulations and a real case study in predicting currency risk premiums from the global trade network.

The unified framework and SuperCENT methodology can be extended in multiple directions. One can consider a generalized linear model for the outcome model and extend SuperCENT to generalized SuperCENT. In the case when only a subset of covariates and outcomes are observed, semi-supervised SuperCENT can be developed. In the case when the network is partially observed, we can perform matrix completion with supervision. SuperCENT can also be extended to a longitudinal model with additional assumptions by using techniques from tensor decomposition as well as functional data analysis to obtain centralities that are smooth over time. For ultra high-dimensional problems, sparsity can be imposed on centralities due to the existence of abundant peripheral nodes.

Acknowledgement

Shen's research is supported in part by Hong Kong CRF C7162-20GF, the Ministry of Science and Technology Major Project of China 2017YFC1310903, University of Hong Kong (HKU) Stanley Ho Alumni Challenge Fund, and HKU BRC Grant. Yang's research is supported in part by NSF grant IIS-1741390, Hong Kong GRF 17301620 and CRF C7162-20GF. Zhao's research is supported in part by Wharton Global Initiative Fund. Zhu's research is supported in part by Tsinghua University Initiative Scientific Research Program and Tsinghua University School of Economics and Management Research Grant.

References

Abel, A. B. (2017). Classical measurement error with several regressors. Technical report, University of Pennsylvania. 6

Ahern, K. R. (2013). Network centrality and cross section of stock returns. *SSRN 2197370*. 2

Allen, F., Cai, J., Gu, X., Qian, J., Zhao, L., and Zhu, W. (2019). Ownership network and firm growth: What do five million companies tell about chinese economy. *SSRN 3465126*. 2

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144). 3, 5

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *The Review of Economic Studies*, 86(6):2453–2490. 2

Battaglini, M., Patacchini, E., and Rainone, E. (2021). Endogenous social interactions with unobserved networks. *The Review of Economic Studies*, 89(4):1694–1747. 6

Binkiewicz, N., Vogelstein, J. T., and Rohe, K. (2017). Covariate-assisted spectral clustering. *Biometrika*, 104(2):361–377. 5

Borgatti, S. P., Carley, K. M., and Krackhardt, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social Networks*, 28(2):124–136. 3

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55. 6

Breza, E. and Chandrasekhar, A. G. (2019). Social networks, reputation, and commitment: evidence from a savings monitors experiment. *Econometrica*, 87(1):175–216. 2

Breza, E., Chandrasekhar, A. G., McCormick, T. H., and Pan, M. (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, 110(8):2454–84. 5

Cai, T. T. and Zhang, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60–89. 9

Candelaria, L. E. and Ura, T. (2022). Identification and inference of network formation games with misclassified links. *Journal of Econometrics*. 3

Candes, E. J. and Plan, Y. (2010). Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936. 9

De Paula, A. (2017). Econometrics of network models. In *Advances in Economics and Econometrics: Theory and Applications: Eleventh World Congress*, volume 1, pages 268–323. Cambridge University Press, Cambridge. 6

De Paula, Á., Rasul, I., and Souza, P. (2019). Identifying network ties from panel data: theory and an application to tax competition. *arXiv preprint arXiv:1910.07452*. 6

Fosdick, B. K. and Hoff, P. D. (2015). Testing and modeling dependencies between a network and nodal attributes. *Journal of the American Statistical Association*, 110(511):1047–1056. 9

Frantz, T. L., Cataldo, M., and Carley, K. M. (2009). Robustness of centrality measures under uncertainty: Examining the role of network topology. *Computational and Mathematical Organization Theory*, 15(4):303–328. 3

Gao, C. and Ma, Z. (2021). Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing. *Statistical Science*, 36(1):16–33. 9

Garber, S. and Klepper, S. (1980). Extending the classical normal errors-in-variables model. *Econometrica: Journal of the Econometric Society*, pages 1541–1546. 6

Gofman, M. (2017). Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *Journal of Financial Economics*, 124(1):113–146. 2

Handcock, M. S. and Gile, K. J. (2010). Modeling social networks from sampled data. *The Annals of Applied Statistics*, 4(1):5. 5

Hochberg, Y. V., Ljungqvist, A., and Lu, Y. (2007). Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance*, 62(1):251–301. 2

Hoff, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and mathematical organization theory*, 15(4):261–272. 9

Hsieh, C.-S. and Lee, L. F. (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, 31(2):301–319. 6

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton University Press. 7

Jackson, M. O., Rogers, B. W., and Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of Economic Literature*, 55(1):49–95. 2

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632. 7, 8

Kornienko, O. and Granger, D. A. (2018). Peer networks, psychobiology of stress response, and adolescent development. *Oxford handbook of evolution, biology, and society*, pages 327–348. 3

Lakhina, A., Byers, J. W., Crovella, M., and Xie, P. (2003). Sampling biases in IP topology measurements. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 1, pages 332–341. IEEE. 3, 5

Landefeld, J. S., Seskin, E. P., and Fraumeni, B. M. (2008). Taking the pulse of the economy: Measuring gdp. *Journal of Economic Perspectives*, 22(2):193–216. 27

Le, C. M., Levin, K., and Levina, E. (2018). Estimating a network from multiple noisy realizations. *Electronic Journal of Statistics*, 12(2):4697–4740. 5

Le, C. M., Levina, E., and Vershynin, R. (2016). Optimization via low-rank approximation for community detection in networks. *The Annals of Statistics*, 44(1):373–400. 9

Le, C. M. and Li, T. (2020). Linear regression and its inference on noisy network-linked data. *arXiv preprint arXiv:2007.00803*. 6

Lee, L.-F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140(2):333–374. 6

Lee, L.-f., Liu, X., and Lin, X. (2010). Specification and estimation of social interaction models with network structures. *The Econometrics Journal*, 13(2):145–176. 6

Lei, J. and Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237. 9

Li, G., Yang, D., Nobel, A. B., and Shen, H. (2016). Supervised singular value decomposition and its asymptotic properties. *Journal of Multivariate Analysis*, 146:7–17. 5

Li, T., Levina, E., Zhu, J., et al. (2019). Prediction models for network-linked data. *The Annals of Applied Statistics*, 13(1):132–164. 6

Lipsey, R. E. (2009). *Measuring International Trade in Services*. University of Chicago Press. 27

Liu, E. (2019). Industrial policies in production networks. *The Quarterly Journal of Economics*, 134(4):1883–1948. 2

Liu, E. and Tsyvinski, A. (2020). Dynamical structure and spectral properties of input-output networks. Technical report, National Bureau of Economic Research. 2

Ma, Z., Ma, Z., and Yuan, H. (2020). Universal latent space model fitting for large networks with edge covariates. *Journal of Machine Learning Research*, 21:4–1. 5

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542. 6

Martin, C. and Niemeyer, P. (2019). Influence of measurement errors on networks: Estimating the robustness of centrality measures. *Network Science*, 7(2):180–195. 3

Mojzisch, A., Frisch, J. U., Doehne, M., Reder, M., and Häusser, J. A. (2021). Interactive effects of social network centrality and social identification on stress. *British Journal of Psychology*, 112(1):144–162. 3

Ozsoylev, H. N., Walden, J., Yavuz, M. D., and Bildik, R. (2014). Investor networks in the stock market. *The Review of Financial Studies*, 27(5):1323–1366. 3

Pischke, S. (2007). Lecture notes on measurement error. Technical report, London School of Economics, London. 6

Richmond, R. J. (2019). Trade network centrality and currency risk premia. *The Journal of Finance*, 74(3):1315–1361. 2, 5, 26, 28, 31

Rohe, K. (2019). A critical threshold for design effects in network sampling. *The Annals of Statistics*, 47(1):556–582. 5

Rohe, K., Chatterjee, S., and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915. 9

Rossi, A. G., Blake, D., Timmermann, A., Tonks, I., and Wermers, R. (2018). Network centrality and delegated investment performance. *Journal of Financial Economics*, 128(1):183–206. 3

Shabalin, A. A. and Nobel, A. B. (2013). Reconstruction of a low-rank matrix in the presence of gaussian noise. *Journal of Multivariate Analysis*, 118:67–76. 4, 9, 10

Shen, D., Shen, H., and Marron, J. (2016). A general framework for consistency of principal component analysis. *The Journal of Machine Learning Research*, 17(1):5218–5251. 4

Soufiani, H. A. and Airoldi, E. (2012). Graphlet decomposition of a weighted network. In *Artificial Intelligence and Statistics*, pages 54–63. PMLR. 9

Van Loan, C. F. and Golub, G. (1996). *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)*. The Johns Hopkins University Press. 8

Wang, D. J., Shi, X., McFarland, D. A., and Leskovec, J. (2012). Measurement error in network data: A re-classification. *Social Networks*, 34(4):396–409. 3

Yan, T., Jiang, B., Fienberg, S. E., and Leng, C. (2019). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association*, 114(526):857–868. 5

Yang, D., Ma, Z., and Buja, A. (2016). Rate optimal denoising of simultaneously sparse and low rank matrices. *The Journal of Machine Learning Research*, 17(1):3163–3189. 9

Zhang, Y., Levina, E., and Zhu, J. (2016). Community detection in networks with node features. *Electronic Journal of Statistics*, 10(2):3153–3178. 5

Zhao, Y., Levina, E., and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266–2292. 9

Zhu, W. and Yang, Y. (2020). Networks and business cycles. *Available at SSRN*. 2, 3

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017). Network vector autoregression. *The Annals of Statistics*, 45(3):1096–1123. 6