Network Regression and Supervised Centrality Estimation

Junhui Cai
University of Notre Dame

Ran Chen
Washington University in St. Louis

Dan Yang
The University of Hong Kong

Wu Zhu
Tsinghua University

Haipeng Shen
The University of Hong Kong

Linda Zhao
University of Pennsylvania

Network Regression and Supervised Centrality Estimation

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 to study the properties of centrality estimation and inference and the subsequent network regression analysis with noisy network observations. Furthermore, we propose a supervised centrality estimation methodology, which aims to simultaneously estimate both centrality and network effect. We showcase the advantages of our method compared with the two-stage method both theoretically and numerically via extensive simulations and a case study in predicting currency risk premiums from the global trade network.

Keywords: Hub centrality, Authority centrality, Network effect, Global trade network, Currency risk premium

1 Introduction

In many disciplines such as economics, finance, and sociology, there has been great interest in studying the network effect, that is, the effect of a network on certain outcomes of interest due to relationships among agents (e.g., individuals, firms, industries, and countries). One popular approach is to bridge the outcome and network via an intermediary or a sufficient statistics – 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 that 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 (Allen et al., 2019, Yang and Zhu, 2020, Liu and Tsyvinski, 2020). Centrality itself has rich implications for studying human capital investment (Jackson et al., 2017), information sharing and advertising (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. 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 to help understand the 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 in strategic industries 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, however, is not directly observable. In practice, researchers often follow a two-stage procedure: in Stage 1, they compute the centrality from a given network adjacency matrix using an algorithm; in Stage 2, the computed centrality is then used as an input in the regression analysis. Such a 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, a network is often observed with noise due to the cost of data collection (Lakhina et al., 2003). There are numerous examples of such noise: 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 (Yang and Zhu, 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, yet it has the following drawbacks. First, 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. Second, 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 follow-up 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. First, to the best of our knowledge, despite the popular adoption of the two-stage procedure, 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.

Second, we are the first to study the properties of the common practice of the two-stage procedure and demonstrate that it can be problematic when the network noise is large. The accuracy of the two-stage centrality estimates in Stage 1 depends on the network noise. When the network noise is large, the centrality estimates are inaccurate, which results in *inaccurate* centrality coefficient estimates with *invalid* ad-hoc inference in Stage 2.

Thirdly, we show theoretically and empirically that the proposed SuperCENT dominates the two-stage procedure universally. Specifically, for (G1), SuperCENT yields a more accurate centrality estimation, especially under large network noise; for (G2), SuperCENT boosts the accuracy of the regression coefficient estimation and 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 procedure to predict the currency risk premium, based on an economic theory that links a country's currency risk premium with its importance within the global trade network (Richmond, 2019). We show that a long-short trading strategy based on SuperCENT centrality estimates yields a return *double* that of 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 strands of literature, including network modeling, network regression with centralities, covariate-assisted network modeling, and network effect modeling. First, the proposed unified framework bridges the gap between research on noisy networks and network regression with centralities. Most of the existing network literature focuses on one of these two aspects. On one hand, in studies involving noisy networks, many empirical works have estimated the true network without incorporating centrality measures (e.g., 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 work, including those mentioned earlier, have 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 has focused mostly on network formation or community detection.

In econometrics, there has been significant effort to model the network effect on an outcome of interest through regression (De Paula, 2017). One popular approach follows the pioneering work of Manski (1993) and his "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 has also been 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 network homophily, while ours concentrates on the nodes' position or importance in the network using the centralities.

It is worth mentioning that there is an extensive body of literature discussing the concept of centrality in networks with negative-weight edges. The foundational work on these networks stems from social balance theory in sociology (Harary, 1953, Cartwright and Harary, 1956). Bonacich and Lloyd (2004) extends the concept of centrality to such networks, providing interpretations grounded in balance theory. Several subsequent studies have built on this foundation (Chiang et al., 2014, Everett and Borgatti, 2014, Singh, 2019, Ma et al., 2019, Gromov, 2025). Empirical research has also explored these networks, including studies on workplace dynamics (Labianca and Brass, 2006) and alliance-enemy networks in wartime (König et al., 2017). Our method is applicable to these networks, and centrality can be interpreted within the framework established by the literature.

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 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://cccfran.github.io/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; 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 an 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 iterating with proper normalization as follows until 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),

$$\begin{cases} \boldsymbol{A} = \boldsymbol{A}_0 + \boldsymbol{E} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top + \boldsymbol{E} = d\boldsymbol{u}\boldsymbol{v}^\top + \sum_{l=2}^{r} d_l \boldsymbol{u}_l \boldsymbol{v}_l^\top + \boldsymbol{E}, & \text{(2a)} \\ \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_x + \boldsymbol{u}\beta_u + \boldsymbol{v}\beta_v + \boldsymbol{\epsilon}, & \text{(2b)} \end{cases}$$

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 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 of interest to be estimated, $(\boldsymbol{u}_2, \ldots, \boldsymbol{u}_r)$ and $(\boldsymbol{v}_2, \ldots, \boldsymbol{v}_r)$ are the non-leading singular vectors orthogonal to $\boldsymbol{u}, \boldsymbol{v}$, 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 centralities, not the estimated ones, that have direct impacts on the response and only $\boldsymbol{u}$ and $\boldsymbol{v}$ are included instead of the entire $\boldsymbol{U}$ and $\boldsymbol{V}$ because it is common practice to consider the network effect via only the centralities.

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 $\boldsymbol{u}, \boldsymbol{v}$; (ii) estimate the regression coefficients $\boldsymbol{\beta}_x, \beta_u, \beta_v$; and (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 that can be rewritten as (2a) (Hoff, 2009, Soufiani and Airoldi, 2012, Fosdick and Hoff, 2015).

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. (Model identifiability) Note that $\boldsymbol{u}, \boldsymbol{v}$ are only identifiable up to a scalar. SVD assumes $\boldsymbol{u}$ and $\boldsymbol{v}$ have unit length. However, we assume $\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = \sqrt{n}$, because 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 $(\boldsymbol{X}, \boldsymbol{u}, \boldsymbol{v})$ 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 provides details on tuning parameter selection and discusses SuperCENT for 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*,

$$\begin{cases} (\hat{d}^{ts}, \hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts}) := \underset{d, \|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = \sqrt{n}}{\arg\min} \|\boldsymbol{A} - d\boldsymbol{u}\boldsymbol{v}^\top\|_F^2, & \text{(3a)} \\ \hat{\boldsymbol{\beta}}^{ts} := ((\hat{\boldsymbol{\beta}}_x^{ts})^\top, \hat{\beta}_u^{ts}, \hat{\beta}_v^{ts})^\top := \underset{\boldsymbol{\beta}_x, \beta_u, \beta_v}{\arg\min} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_x - \hat{\boldsymbol{u}}^{ts}\beta_u - \hat{\boldsymbol{v}}^{ts}\beta_v\|_2^2. & \text{(3b)} \end{cases}$$

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 2. (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 $\operatorname{cov}(\hat{\boldsymbol{\beta}}^{ts}) = \sigma_y^2 (\widehat{\mathbf{W}}^\top \widehat{\mathbf{W}})^{-1}$, 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 account 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 estimators, which depends on the network noise $\boldsymbol{E}$ as well as the singular values and singular vectors of $\boldsymbol{A}_0$, 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 into 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 optimizing the following objective function $\mathcal{L}$ to obtain the SuperCENT estimates,

$$(\hat{d}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}, \widehat{\boldsymbol{\beta}}_x, \hat{\beta}_u, \hat{\beta}_v) := \underset{\substack{\boldsymbol{\beta}_x, \beta_u, \beta_v \\ d, \|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = \sqrt{n}}}{\arg\min} \frac{1}{n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_x - \boldsymbol{u}\beta_u - \boldsymbol{v}\beta_v\|_2^2 + \frac{\lambda}{n^2}\|\boldsymbol{A} - d\boldsymbol{u}\boldsymbol{v}^\top\|_F^2, \quad (4)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix. The above objective function combines the residual sum of squares (3b) and the rank-one approximation error of the observed network (3a). 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 (4), 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., $(\hat{d}^{ts}, \hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts})$. Let $\boldsymbol{P}_{\boldsymbol{A}}$ denotes the projection matrix that projects onto the column space of $\boldsymbol{A}$ and $\|\boldsymbol{A}\|_2$ is the matrix operator norm of matrix $\boldsymbol{A}$. We use $(\hat{d}^{(t)}, \hat{\boldsymbol{u}}^{(t)}, \hat{\boldsymbol{v}}^{(t)}, \widehat{\boldsymbol{\beta}}^{(t)})$ to denote the estimations in the t-th iteration. The complete algorithm with a given tuning parameter λ is shown in Algorithm 1.

Note that although $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ with length $\sqrt{n}$ are only identifiable up to the sign, $\hat{\boldsymbol{u}}\hat{\boldsymbol{v}}^\top$, $\hat{\boldsymbol{u}}\hat{\beta}_u$, $\hat{\boldsymbol{v}}\hat{\beta}_v$ are uniquely identifiable. Without additional structure, the sign is irrelevant since with proper flipping of $(\beta_v, \beta_u, \boldsymbol{u}, \boldsymbol{v})$, the objective function will have the same value and the flipped sequence will remain a valid sequence generated by the algorithm. When additional information is available (e.g., positive entries), one can determine the sign as

Algorithm 1: SuperCENT algorithm with a given tuning parameter.

Result: $\hat{d}$, $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$, and $\widehat{\boldsymbol{\beta}}$.

Input: $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{X} \in \mathbb{R}^{n\times p}$, $\mathbf{y} \in \mathbb{R}^{n}$, tuning penalty parameter λ, tolerance parameter $\rho > 0$, maximum number of iteration T ;

Initiate $(d^{(0)}, \boldsymbol{u}^{(0)}, \boldsymbol{v}^{(0)}) = (\hat{d}^{ts}, \hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts})$, $t = 1$;

while $t \leqslant 2$ *or* $(\max(\|\boldsymbol{P}_{\boldsymbol{u}^{(t-1)}} - \boldsymbol{P}_{\boldsymbol{u}^{(t-2)}}\|_2, \|\boldsymbol{P}_{\boldsymbol{v}^{(t-1)}} - \boldsymbol{P}_{\boldsymbol{v}^{(t-2)}}\|_2) > \rho$ *and* $t < T)$ **do**

1. $\mathbf{W}^{(t-1)} = (\mathbf{X}, \boldsymbol{u}^{(t-1)}, \boldsymbol{v}^{(t-1)})$;
2. $\boldsymbol{\beta}^{(t)} = (\mathbf{W}^{(t-1)\top}\mathbf{W}^{(t-1)})^{-1}\mathbf{W}^{(t-1)\top}\boldsymbol{y}$;
3. $d^{(t)} = \boldsymbol{u}^{(t-1)\top}\mathbf{A}\boldsymbol{v}^{(t-1)}/(\|\boldsymbol{u}^{(t-1)}\|_2^2\|\boldsymbol{v}^{(t-1)}\|_2^2)$;
4. $$\boldsymbol{u}^{(t)} = \left((\beta_u^{(t)})^2 + \frac{1}{n}\lambda(d^{(t-1)})^2\|\boldsymbol{v}^{(t-1)}\|_2^2\right)^{-1} \times \left[\beta_u^{(t)}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_x^{(t)} - \boldsymbol{v}^{(t-1)}\beta_v^{(t)}) + \frac{1}{n}\lambda d^{(t-1)}\mathbf{A}\boldsymbol{v}^{(t-1)}\right];$$
5. $$\boldsymbol{v}^{(t)} = \left((\beta_v^{(t)})^2 + \frac{1}{n}\lambda(d^{(t-1)})^2\|\boldsymbol{u}^{(t)}\|_2^2\right)^{-1} \times \left[\beta_v^{(t)}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_x^{(t)} - \boldsymbol{u}^{(t)}\beta_u^{(t)}) + \frac{1}{n}\lambda d^{(t-1)}\mathbf{A}^\top\boldsymbol{u}^{(t)}\right];$$
6. Normalize $\boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)}$ to have norm $\sqrt{n}$:
$$\boldsymbol{u}^{(t)} = \sqrt{n}\boldsymbol{u}^{(t)}/\|\boldsymbol{u}^{(t)}\|_2, \boldsymbol{v}^{(t)} = \sqrt{n}\boldsymbol{v}^{(t)}/\|\boldsymbol{v}^{(t)}\|_2;$$
7. $t \leftarrow t + 1$;

end

$\mathbf{W}^{(t-1)} = (\mathbf{X}, \boldsymbol{u}^{(t-1)}, \boldsymbol{v}^{(t-1)})$;

$\widehat{\boldsymbol{\beta}} = (\mathbf{W}^{(t-1)\top}\mathbf{W}^{(t-1)})^{-1}\mathbf{W}^{(t-1)\top}\boldsymbol{y}$;

$\hat{d} = \boldsymbol{u}^{(t-1)\top}\mathbf{A}\boldsymbol{v}^{(t-1)}/(\|\boldsymbol{u}^{(t-1)}\|_2^2\|\boldsymbol{v}^{(t-1)}\|_2^2)$;

$\hat{\boldsymbol{u}} = \boldsymbol{u}^{(t-1)}, \hat{\boldsymbol{v}} = \boldsymbol{v}^{(t-1)}$.

follows: identify the entry that has the largest magnitude in $(\hat{\boldsymbol{u}}^\top, \hat{\boldsymbol{v}}^\top)$, make that entry positive, and adjust the signs of the remaining entries in $\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}$ as well as $\hat{\beta}_u, \hat{\beta}_v$ accordingly.

Remark 3 (Convergence of Algorithm 1)**.** SuperCENT will converge to stationary points that have smaller objective values than the objective function value of our initial point. Precisely, $\lim_{t\to\infty} \|\partial\mathcal{L}(\boldsymbol{\beta}^{(t+1)}, d^{(t+1)}, \boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)})\|_2 = 0$ and $\sup_{t\geqslant 1} \mathcal{L}(\boldsymbol{\beta}^{(t+1)}, d^{(t+1)}, \boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)}) \leqslant \mathcal{L}(\boldsymbol{\beta}^{(1)}, d^{(1)}, \boldsymbol{u}^{(0)}, \boldsymbol{v}^{(0)})$, which is all we need for our theoretical results. Details of the proof are deferred to Supplement S3.1.

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 $\boldsymbol{A}_{22}$, as well as the new edges connecting them with the n training observations, $\boldsymbol{A}_{12}$ and $\boldsymbol{A}_{21}$.

Specifically, the original network $\boldsymbol{A}$ is then augmented to $\boldsymbol{A}^{all}$ as follows:

$$\boldsymbol{A}^{all} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{U}_1\boldsymbol{V}_1^\top & \boldsymbol{U}_1\boldsymbol{V}_2^\top \\ \boldsymbol{U}_2\boldsymbol{V}_1^\top & \boldsymbol{U}_2\boldsymbol{V}_2^\top \end{pmatrix} + \begin{pmatrix} \boldsymbol{E}_{11} & \boldsymbol{E}_{12} \\ \boldsymbol{E}_{21} & \boldsymbol{E}_{22} \end{pmatrix}, \tag{5}$$

where $\boldsymbol{A}_{11}$ is the network of the training data, $\boldsymbol{U}_1, \boldsymbol{V}_1$ contain the square roots of the singular values and singular vectors of the training data, and $\boldsymbol{U}_2, \boldsymbol{V}_2$ contain the square roots of the singular values and singular vectors of the testing data and need to be estimated before plugging them into the regression model to make predictions.

To estimate $\boldsymbol{U}_2$ and $\boldsymbol{V}_2$, one can either perform SVD on $\boldsymbol{A}_{22}$ or on $\boldsymbol{A}^{all}$ and reserve only the relevant components of the singular vectors. Another option that improves estimation to $\boldsymbol{U}_2$ and $\boldsymbol{V}_2$ — and thus improves the prediction — is to utilize SuperCENT estimates $\boldsymbol{U}_1$ and $\boldsymbol{V}_1$. To be specific, consider the minimization of the approximation errors related to the two off-diagonal blocks

$$\min_{\boldsymbol{U}_2} \|\boldsymbol{A}_{21} - \boldsymbol{U}_2\boldsymbol{V}_1^\top\|_F^2 \quad \text{and} \quad \min_{\boldsymbol{V}_2} \|\boldsymbol{A}_{12} - \boldsymbol{U}_1\boldsymbol{V}_2^\top\|_F^2. \tag{6}$$

Solving these optimization problems leads to

$$\boldsymbol{U}_2 = \left((\boldsymbol{V}_1^\top \boldsymbol{V}_1)^{-1}\boldsymbol{V}_1\boldsymbol{A}_{21}^\top\right)^\top \quad \text{and} \quad \boldsymbol{V}_2 = \left((\boldsymbol{U}_1^\top \boldsymbol{U}_1)^{-1}\boldsymbol{U}_1\boldsymbol{A}_{12}\right)^\top. \tag{7}$$

Plugging into (7) the estimated $\boldsymbol{U}_1, \boldsymbol{V}_1$ accordingly:

$$\hat{\boldsymbol{U}}_1 = \begin{pmatrix} \hat{d}^{1/2}\hat{\boldsymbol{u}} & \hat{\boldsymbol{U}}_{2:r}\hat{\boldsymbol{D}}_{2:r}^{1/2} \end{pmatrix}, \qquad \hat{\boldsymbol{V}}_1 = \begin{pmatrix} \hat{d}^{1/2}\hat{\boldsymbol{v}} & \hat{\boldsymbol{V}}_{2:r}\hat{\boldsymbol{D}}_{2:r}^{1/2} \end{pmatrix} \tag{8}$$

where $\hat{d}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}$ are SuperCENT estimators of the training network $\boldsymbol{A}_{11}$ and $\hat{\boldsymbol{D}}, \hat{\boldsymbol{U}}, \hat{\boldsymbol{V}}$ are the estimated singular values and singular vectors from plain SVD of $\boldsymbol{A}_{11}$ with appropriate scaling to have length $\sqrt{n}$. Denote the first columns of $\hat{\boldsymbol{U}}_2$ and $\hat{\boldsymbol{V}}_2$ as $\hat{\boldsymbol{u}}^*$ and $\hat{\boldsymbol{v}}^*$ and rescale them by dividing by $\sqrt{\hat{d}}$. We predict the outcome as $\hat{\boldsymbol{y}}^* = \boldsymbol{X}^*\widehat{\boldsymbol{\beta}}_x + \hat{\boldsymbol{u}}^*\hat{\beta}_u + \hat{\boldsymbol{v}}^*\hat{\beta}_v$. The algorithm is formally described in Algorithm S4.

4 Theoretical properties

We investigate the statistical properties of SuperCENT and compare it with the two-stage procedure in this section. We start with introducing notations and assumptions.

Denote SuperCENT estimators from Algorithm 1 with a given tuning parameter λ as $\hat{d}$, $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$, and $\widehat{\boldsymbol{\beta}} = ((\widehat{\boldsymbol{\beta}}_x)^\top, \hat{\beta}_u, \hat{\beta}_v)^\top$ and the two-stage counterparts as $\hat{d}^{ts}$, $\hat{\boldsymbol{u}}^{ts}$, $\hat{\boldsymbol{v}}^{ts}$, and $\hat{\boldsymbol{\beta}}^{ts}$. We further denote $\boldsymbol{A}^\perp = \boldsymbol{U}^\perp \boldsymbol{D}^\perp \boldsymbol{V}^{\perp\top}$ where $\boldsymbol{U}^\perp = (\boldsymbol{u}_2, \ldots, \boldsymbol{u}_r)$, $\boldsymbol{V}^\perp = (\boldsymbol{v}_2, \ldots, \boldsymbol{v}_r)$ and $\boldsymbol{D}^\perp = diag(d_2, \ldots, d_r)$, $\boldsymbol{\Omega} = \begin{pmatrix} \sigma_y^2 \boldsymbol{I}_n & \boldsymbol{0}_{n\times n^2} \\ \boldsymbol{0}_{n^2\times n} & \sigma_a^2 \boldsymbol{I}_{n^2} \end{pmatrix}$, $\boldsymbol{P}_{(\boldsymbol{X}\boldsymbol{u}\boldsymbol{v})}$ as the projection matrix that projects onto the column space of $(\boldsymbol{X}, \boldsymbol{u}, \boldsymbol{v})$, and similarly for $\boldsymbol{P}_{\boldsymbol{X}}$, $\boldsymbol{P}_{\boldsymbol{u}}$ and $\boldsymbol{P}_{\boldsymbol{v}}$. Define $\tilde{\boldsymbol{u}} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{u}$, $\tilde{\boldsymbol{v}} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{v}$, which are the centralities projected onto the orthogonal space of $\boldsymbol{X}$, and $\boldsymbol{C}_{\tilde{\boldsymbol{u}}\tilde{\boldsymbol{v}}} = (\tilde{\boldsymbol{u}},\ \tilde{\boldsymbol{v}})^\top (\tilde{\boldsymbol{u}},\ \tilde{\boldsymbol{v}})$. In the following, the theorems are for $\arg\max_{\mathbf{h}\in\{\hat{\boldsymbol{u}}, -\hat{\boldsymbol{u}}\}} \operatorname{sign}(\mathbf{h}^\top \boldsymbol{u})$ and $\arg\max_{\mathbf{g}\in\{\hat{\boldsymbol{v}}, -\hat{\boldsymbol{v}}\}} \operatorname{sign}(\mathbf{g}^\top \boldsymbol{v})$, and we continue to use $\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}$ to denote them. While these notions are a bit of an abuse of notation, it is reasonable since both the objective function and algorithm are sign-invariant (i.e., proper flipping of signs gives the same value or another valid iteration sequence). The same notation applies in the simulations as well.

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 $n > p+2$, and the dimension p is non-diverging. We further assume $(\boldsymbol{X}, \boldsymbol{u}, \boldsymbol{v})$ is full rank and the condition number of $(\boldsymbol{X}, \boldsymbol{u}, \boldsymbol{v})^\top(\boldsymbol{X}, \boldsymbol{u}, \boldsymbol{v})$ is smaller or equal to $1/\tau^2$ for some positive constant τ.

Assumption 3. The scaled network noise-to-signal ratio $\kappa := \frac{\sigma_a^2}{(d-d_2)^2 n} \to 0$.

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 and SuperCENT, which essentially requires the signal-to-noise ratio (SNR) of the network and the gap between the leading and second singular values of the network to be large enough.

Theorem 1. *Under the unified framework (2) and Assumptions 1, 2 and 3, the SuperCENT estimators converge to the following normal distributions asymptotically,*

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

$$(\hat{\boldsymbol{u}}_i - \boldsymbol{u}_i) \xrightarrow{\mathcal{D}} N\Big(0, \boldsymbol{\Sigma}_{\boldsymbol{u},ii}\Big) \quad \text{and} \quad (\hat{\boldsymbol{v}}_i - \boldsymbol{v}_i) \xrightarrow{\mathcal{D}} N\Big(0, \boldsymbol{\Sigma}_{\boldsymbol{v},ii}\Big), \tag{9}$$

2. *Network effect:*

$$\Big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\Big) \xrightarrow{\mathcal{D}} N\Big(\mathbf{0}_{p+2}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\Big), \tag{10}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{C}_{\boldsymbol{u}}\boldsymbol{\Omega}\boldsymbol{C}_{\boldsymbol{u}}^\top$, $\boldsymbol{\Sigma}_{\boldsymbol{v}} = \boldsymbol{C}_{\boldsymbol{v}}\boldsymbol{\Omega}\boldsymbol{C}_{\boldsymbol{v}}^\top$, *and* $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \boldsymbol{C}_{\boldsymbol{\beta}}\boldsymbol{\Omega}\boldsymbol{C}_{\boldsymbol{\beta}}^\top$ *where* $\boldsymbol{C}_{\boldsymbol{u}} = \begin{pmatrix} \boldsymbol{C}_{11}, & \boldsymbol{C}_{12} \end{pmatrix}$, $\boldsymbol{C}_{\boldsymbol{v}} = \begin{pmatrix} \boldsymbol{C}_{21}, & \boldsymbol{C}_{22} \end{pmatrix}$, *and* $\boldsymbol{C}_{\boldsymbol{\beta}} = \begin{pmatrix} \boldsymbol{C}_{31} & \boldsymbol{C}_{32} \\ \boldsymbol{C}_{41} & \boldsymbol{C}_{42} \\ \boldsymbol{C}_{51} & \boldsymbol{C}_{52} \end{pmatrix}$ *whose specific forms are as follows.*

The matrices related to $\hat{\boldsymbol{u}}$ *and* $\hat{\boldsymbol{v}}$ *are*

$$\begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{pmatrix} = \left[\frac{\lambda d}{n}\begin{pmatrix} dn\boldsymbol{I} & -\boldsymbol{A}^{\perp} \\ (-\boldsymbol{A}^{\perp})^{\top} & dn\boldsymbol{I} \end{pmatrix} + \begin{pmatrix} \beta_u^2 & \beta_u\beta_v \\ \beta_u\beta_v & \beta_v^2 \end{pmatrix} \otimes \Big(\boldsymbol{I} - \boldsymbol{P}_{(\boldsymbol{X}\boldsymbol{u}\boldsymbol{v})}\Big)\right]^{-1} \tag{11}$$
$$\begin{pmatrix} \beta_u\Big(\boldsymbol{I} - \boldsymbol{P}_{(\boldsymbol{X}\boldsymbol{u}\boldsymbol{v})}\Big) & \lambda d\boldsymbol{v}^{\top} \otimes (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{u}})/n \\ \beta_v\Big(\boldsymbol{I} - \boldsymbol{P}_{(\boldsymbol{X}\boldsymbol{u}\boldsymbol{v})}\Big) & \lambda d\left(\boldsymbol{u}^{\top} \otimes (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{v}})/n\right)\boldsymbol{K} \end{pmatrix},$$

the matrices related to $\hat{\beta}_u$ *and* $\hat{\beta}_v$ *are*

$$\begin{pmatrix} \boldsymbol{C}_{31} & \boldsymbol{C}_{32} \\ \boldsymbol{C}_{41} & \boldsymbol{C}_{42} \end{pmatrix} = \boldsymbol{C}_{\tilde{\boldsymbol{u}}\tilde{\boldsymbol{v}}}^{-1}\begin{pmatrix} \tilde{\boldsymbol{u}}^{\top} \\ \tilde{\boldsymbol{v}}^{\top} \end{pmatrix}\begin{pmatrix} -\beta_u\boldsymbol{I}_n & -\beta_v\boldsymbol{I}_n & \boldsymbol{I}_n \end{pmatrix}\begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \\ \boldsymbol{I}_n & \boldsymbol{0}_{n\times n^2} \end{pmatrix}, \tag{12}$$

and the matrices related to $\widehat{\boldsymbol{\beta}}_x$ *are*

$$\begin{pmatrix} \boldsymbol{C}_{51}, & \boldsymbol{C}_{52} \end{pmatrix} = (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\begin{pmatrix} -\beta_u\boldsymbol{I}_n & -\beta_v\boldsymbol{I}_n & -\boldsymbol{u} & -\boldsymbol{v} & \boldsymbol{I}_n \end{pmatrix}\begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \\ \boldsymbol{C}_{31} & \boldsymbol{C}_{32} \\ \boldsymbol{C}_{41} & \boldsymbol{C}_{42} \\ \boldsymbol{I}_n & \boldsymbol{0}_{n\times n^2} \end{pmatrix}. \tag{13}$$

Similarly, we derive the asymptotic distribution of the two-stage estimator in Theorem S1. Comparing the covariance matrices with those of the two-stage in Theorem S1, all $\boldsymbol{\Sigma}_{\boldsymbol{u}}, \boldsymbol{\Sigma}_{\boldsymbol{v}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ involve both σ_a^2 and σ_y^2 due to the simultaneous estimation, while $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{ts}$ and $\boldsymbol{\Sigma}_{\boldsymbol{v}}^{ts}$ of the two-stage only involve σ_a^2 and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{ts}$ involves both. Specifically, $\boldsymbol{C}_{\boldsymbol{u}}$ and $\boldsymbol{C}_{\boldsymbol{v}}$ are functions of $(\sigma_a, \boldsymbol{D}, \boldsymbol{U}, \boldsymbol{V}, \sigma_y, \boldsymbol{X}, \beta_u, \beta_v, \lambda)$, while the two-stage counterparts $\boldsymbol{C}_{\boldsymbol{u}}^{ts}$ and $\boldsymbol{C}_{\boldsymbol{v}}^{ts}$ only involve $(\sigma_a, \boldsymbol{D}, \boldsymbol{U}, \boldsymbol{V})$. Therefore, the difference between SuperCENT and the two-stage estimators of $\boldsymbol{u}$ and $\boldsymbol{v}$ lies in the tuning parameter λ as well as the SNRs of the network and regression. Following Theorem 1, Proposition 1 provides the convergence rates of $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$. We focus on the rank-one scenario where $\boldsymbol{A}_0 = d\boldsymbol{u}\boldsymbol{v}^{\top}$ in model (2a) to provide clearer insights for understanding the difference between SuperCENT and the two-stage estimators.

Proposition 1. *(Convergence rates of* $\hat{\boldsymbol{u}}$*,* $\hat{\boldsymbol{v}}$*) Under the unified framework* (2) *assuming* $\boldsymbol{A}_0$ *to be rank-one and Assumptions 1, 2 and 3, the SuperCENT estimators satisfy the* *following,*

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

$$= \kappa(1+o(1)) - \beta_u^2\delta_{ts,sc}(1+o(1)), \tag{15}$$

$$\frac{1}{n}\mathbb{E}\|\hat{\boldsymbol{v}}-\boldsymbol{v}\|_2^2 = \left(\frac{\sigma_a^2(n-1)}{d^2n^2} - \frac{n-p-2}{n}\beta_v^2\delta_{ts,sc}\right)(1+o(1)) \tag{16}$$

$$= \kappa(1+o(1)) - \beta_v^2\delta_{ts,sc}(1+o(1)), \tag{17}$$

where

$$\delta_{ts,sc} = (\lambda d^2+\beta_u^2+\beta_v^2)^{-2}\left[\frac{2\lambda d^2+\beta_u^2+\beta_v^2}{d^2n}\sigma_a^2 - \sigma_y^2\right]. \tag{18}$$

Remark 4. (The role of $\delta_{ts,sc}$) Comparing Corollaries S1 and 1, the discrepancies between the two-stage and SuperCENT estimators of the centralities are all proportional to $\delta_{ts,sc}$ since $\mathbb{E}\|\hat{\boldsymbol{u}}^{ts}-\boldsymbol{u}\|_2^2/n - \mathbb{E}\|\hat{\boldsymbol{u}}-\boldsymbol{u}\|_2^2/n = \beta_u^2\delta_{ts,sc}$. It can be seen that, whenever $\delta_{ts,sc} > 0$, SuperCENT always outperforms the two-stage.

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. 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 5. (Optimal λ) Minimizing the convergence rates (15) or (17) 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 (18), 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 6. (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 S5; 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 SuperCENT using $\hat{\lambda}_0$ 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 S5, to obtain $\hat{\lambda}_{cv}$ and SuperCENT-$\hat{\lambda}_{cv}$.

Remark 7. (Comparison of the estimation of $\boldsymbol{u}, \boldsymbol{v}$) We further compare the estimation of $\boldsymbol{u}, \boldsymbol{v}$ of two-stage and SuperCENT. Plugging in the optimal λ_0, the rate of $\mathbb{E}\|\hat{\boldsymbol{u}} - \boldsymbol{u}\|_2^2/n$ in (15) becomes

$$\kappa \frac{1 + \kappa \frac{\beta_v^2}{\sigma_y^2}}{1 + \kappa \left(\frac{\beta_u^2}{\sigma_y^2} + \frac{\beta_v^2}{\sigma_y^2} \right)}, \tag{19}$$

which is obviously faster than the rate of $\mathbb{E}\|\hat{\boldsymbol{u}}^{ts} - \boldsymbol{u}\|_2^2/n = \kappa$ in (S26). Given (19), the improvement from SuperCENT boils down to how the regression SNRs regarding $\boldsymbol{u}$ and $\boldsymbol{v}$ compare with the network SNR, i.e., $\kappa\frac{\beta_u^2}{\sigma_y^2}$ and $\kappa\frac{\beta_v^2}{\sigma_y^2}$. One sufficient condition to gain the most improvement for $\hat{\boldsymbol{u}}$ is when $\kappa\frac{\beta_u^2}{\sigma_y^2} \to \infty$ and $\kappa\frac{\beta_v^2}{\sigma_y^2} = O(1)$, meaning that the signal for $\boldsymbol{u}$ is much stronger than the signal of network while that for $\boldsymbol{v}$ is not much stronger than the signal of the network. On the other hand, to gain the most improvement for $\hat{\boldsymbol{v}}$, one sufficient condition is $\kappa\frac{\beta_v^2}{\sigma_y^2} \to \infty$ and $\kappa\frac{\beta_u^2}{\sigma_y^2} = O(1)$. This condition conflicts with the requirement for $\hat{\boldsymbol{u}}$, showing the non-exchangeable roles of $\boldsymbol{u}$ and $\boldsymbol{v}$. Fortunately, when $\kappa\frac{\beta_v^2}{\sigma_y^2} \to \infty$ and $\kappa\frac{\beta_u^2}{\sigma_y^2} \to \infty$, $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ are both improved although on a lesser scale than the first two scenarios. In all the cases above, the network regression model can provide supervision to the centrality estimation; thus 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. We will demonstrate this phenomenon in the simulation.

Remark 8. (Comparison of the estimation of β_u and β_v) The expression of $\hat{\boldsymbol{\beta}}$ and

$\hat{\boldsymbol{\beta}}^{ts}$ for general case is rather complex. To provide clearer patterns and insights, we consider a simplified setting where $\boldsymbol{X} = \mathbf{0}_{n\times p}$. Under the condition when $\hat{\boldsymbol{u}}$ gains the most improvement, as stated in Remark 7, i.e., $\kappa\frac{\beta_u^2}{\sigma_y^2} \to \infty$ and $\kappa\frac{\beta_v^2}{\sigma_y^2} = O(1)$, we have median$(||\hat{\beta}_u| - |\beta_u||) = o($median$(||\hat{\beta}_u^{ts}| - |\beta_u||))$, meaning that the SuperCENT estimator of β_u is more accurate than that of the two-stage. While for β_v, due to the non-exchangeable roles of $\boldsymbol{u}$ and $\boldsymbol{v}$, $\hat{\boldsymbol{v}}$ does not gain enough improvement to boost $\hat{\beta}_v$. Similar to Remark 7, the situation will be reversed between $\hat{\beta}_u$ and $\hat{\beta}_v$ when $\kappa\frac{\beta_v^2}{\sigma_y^2} \to \infty$ and $\kappa\frac{\beta_u^2}{\sigma_y^2} = O(1)$.

Remark 9. (Two-stage CI versus "two-stage-ad-hoc" CI of β_u) Based on Theorem S1, we can construct a valid confidence interval for β_u. Specifically, the asymptotic variance of $\hat{\beta}_u^{ts}$ when $\boldsymbol{A}_0$ is rank-one has the following two-terms: (S28)-(S29) where the first term is the same as the variance in the classical regression results and the second term is due to the randomness nature of $\hat{\boldsymbol{u}}^{ts}, \hat{\boldsymbol{v}}^{ts}$. Compared with the "two-stage-ad-hoc" CI, i.e., the CI that obtained via software directly from the regression in Stage 2, this "two-stage-adhoc" CI uses (S28) alone and is thus invalid unless $\sigma_a = 0$.

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. Additional simulations, including a phase-transition experiment, are deferred to Supplement S5.

5.1 Simulation setup

We generate the network following model (2a). We consider the case of $r = 10$ where the leading singular value $d = 1$ and the non-leading ones as $d_2 = \ldots = d_r = 2^{-1}$. All entries of $\boldsymbol{U}$ are first generated from i.i.d. $N(0,1)$ and $\boldsymbol{V} = 0.5\boldsymbol{U} + \boldsymbol{\epsilon}_{\boldsymbol{V}}$ where $\boldsymbol{\epsilon}_{\boldsymbol{V}}$ are generated from i.i.d. $N(0,1)$. We then apply Gram–Schmidt to ensure orthogonality between columns of $\boldsymbol{U}$ and $\boldsymbol{V}$, and finally rescale each column 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 κ and the regression SNR $(\frac{\beta_u}{\sigma_y}, \frac{\beta_v}{\sigma_y})$ matter. Hence, we fix $n = 2^8$, $d = 1$, $d_2 = \ldots = d_r = 2^{-1}$, 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 controlled by σ_a, we vary $\sigma_a \in 2^{0,2}$. We study the effects of the non-leading singular values $d_2, \ldots, d_r$ in additional simulations in Supplement S5.

For estimation property, we compare the following procedures: 1. **Two-stage**; 2. **SuperCENT-λ_0**, which implements Algorithm 1 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; and 4. **SuperCENT-$\hat{\lambda}_{cv}$** is SuperCENT with tuning parameter $\hat{\lambda}_{cv}$ chosen by cross-validation as in Algorithm S5.

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; and 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

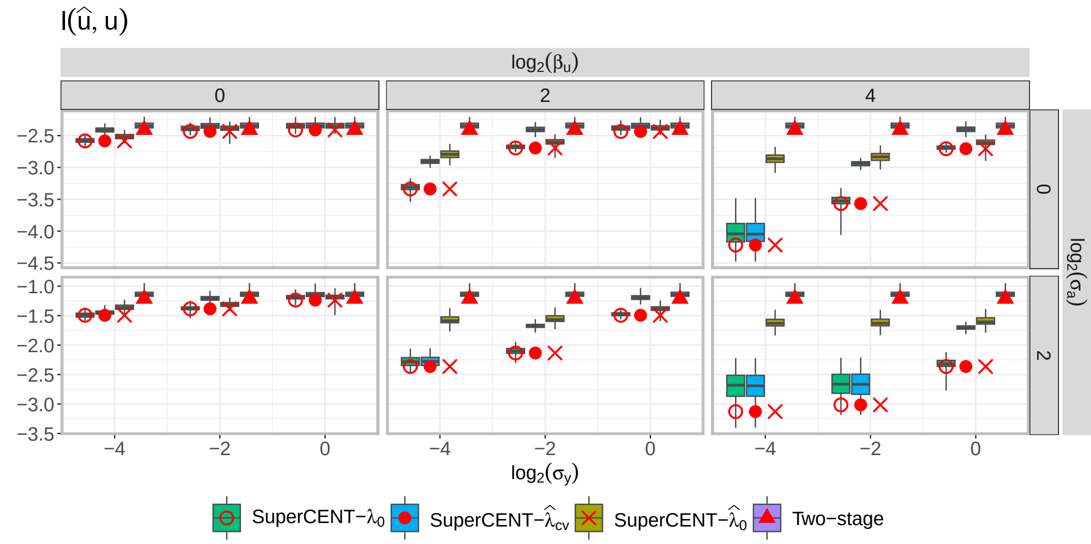
(S28)-(S29) or (S31)-(S32). Note that for the **Two-stage-plugin** and **SuperCENT-$\hat{\lambda}_{cv}$**, $\hat{\sigma}(\hat{\beta}^{ts})$ and $\hat{\sigma}(\hat{\beta}^{\hat{\lambda}_{cv}})$ involve estimation for $\boldsymbol{A}^{\perp} = \boldsymbol{U}^{\perp}\boldsymbol{D}^{\perp}\boldsymbol{V}^{\perp\top}$. For the **Two-stage-plugin**, we plug in the estimate from SVD; for **SuperCENT-$\hat{\lambda}_{cv}$**, we perform SVD on $\boldsymbol{A} - \hat{d}\hat{\boldsymbol{u}}\hat{\boldsymbol{v}}^{\top}$ and then plug in the estimates. The experiments are repeated 500 times.

5.2 Simulation results

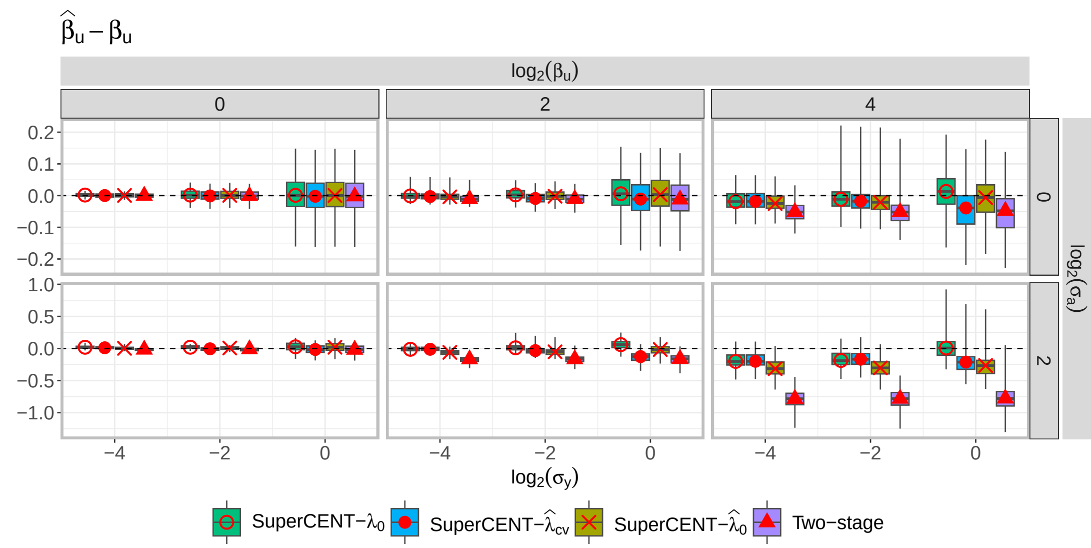
From the perspective of estimation, we compare the following metrics: the estimation accuracy for the centralities, the network, and the regression coefficients. Let $\boldsymbol{P}$ denote the projection matrix. Figure 1 shows the loss $l(\hat{\boldsymbol{u}}, \boldsymbol{u}) = \|\boldsymbol{P}_{\hat{\boldsymbol{u}}} - \boldsymbol{P}_{\boldsymbol{u}}\|_2^2$ and $\hat{\beta}_u - \beta_u$, respectively, across different σ_a, σ_y and β_u with $d = 1$ and $\beta_v = 1$. Losses such 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 $\hat{\beta}_v - \beta_v$ 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 $\hat{\boldsymbol{u}}^{ts}$ and $\hat{\boldsymbol{u}}$ calculated from Theorems S1 and 1 respectively. As expected, 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 4 and 7 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 $\hat{\beta}_u - \beta_u$. With large σ_a or large β_u, the two-stage estimates are inaccurate, while SuperCENT estimates remain accurate. In particular, the two-stage estimates



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

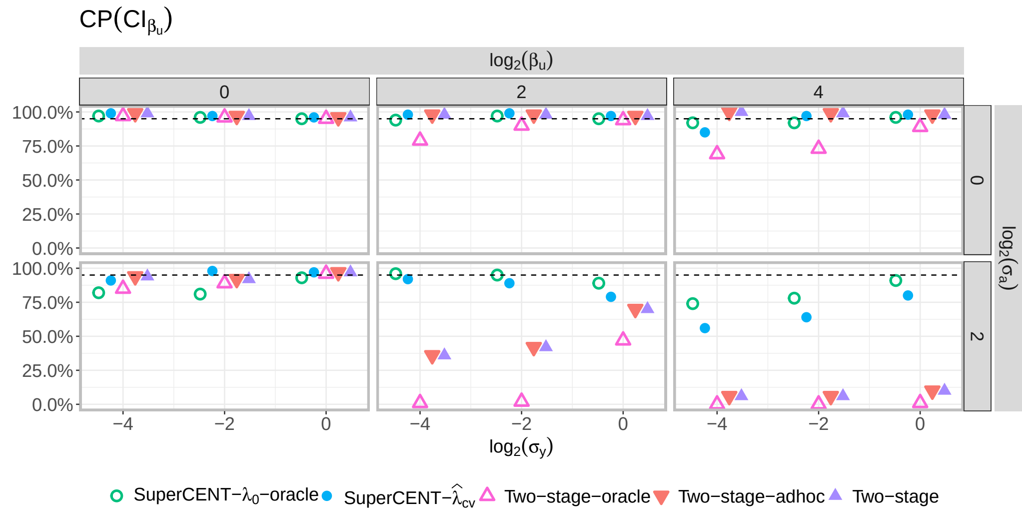


(b) Boxplot of $\hat{\beta}_u - \beta_u$. The dashed lines correspond to 0.

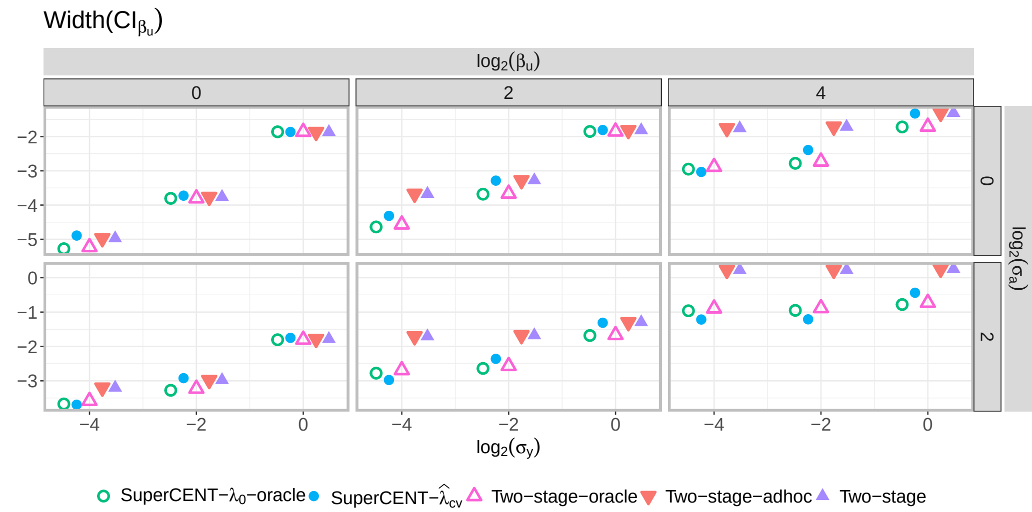
Figure 1. Boxplot of $\log_{10}(l(\hat{\boldsymbol{u}}, \boldsymbol{u}))$ for the four estimators across different σ_a, σ_y and β_u with fixed $d = 1$, $\beta_v = 1$. The super-imposed red symbols show the theoretical rates of the two-stage and SuperCENT calculated from Theorems S1 and 1 respectively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

becomes more inaccurate as σ_a or β_u increases. The three SuperCENT-based methods all outperform the two-stage and SuperCENT-$\hat{\lambda}_{cv}$ and SuperCENT-$\hat{\lambda}_0$ are comparable with the benchmark SuperCENT-λ_0.

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.



(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. 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.

Figure 2a shows how the inaccurate estimation of β_u by the two-stage further affects its confidence interval. Regarding empirical coverage, when β_u is small (leftmost column), all methods are above the nominal level. As β_u increases and σ_a remains small (top right two panels), most methods (except for two-stage-oracle) remain valid, but for different reasons: the two SuperCENT-based methods remain valid due to the accurate estimation of both β_u and the standard error, whereas two-stage and two-stage-ad-hoc remain valid mainly because they over-estimate σ_y^2, and this conservativeness masks the issue of the inaccurate estimation. Two-stage-oracle uses the true σ_y^2 and the issue of the inaccurate estimate cannot be hidden, hence the corresponding intervals undercover. When β_u increases and σ_a gets large too (bottom right panel), the over-estimation of σ_y^2 can no longer hide the issue of inaccurate estimation, causing all two-stage-related methods to become invalid. On the other hand, the empirical coverage of SuperCENT remains 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 becomes more pronounced with larger β_u and σ_a.

6 Global trade network and currency risk premium

In this case study, we demonstrate that SuperCENT can provide a more accurate estimation of the centralities using 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 studied extensively the currency risk premium and remain 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; and (ii) international investors can leverage and profit from a long-short strategy for foreign exchange by taking 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 the 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 five-year moving average: when considering year t, the 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 five-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

[1]The euro was first adopted in 1999. The bilateral trade data is available untill 2013.
[2]The list of country abbreviations is provided in Supplement S6.
[3]We leave the last available year 2013 for the validation purposes.

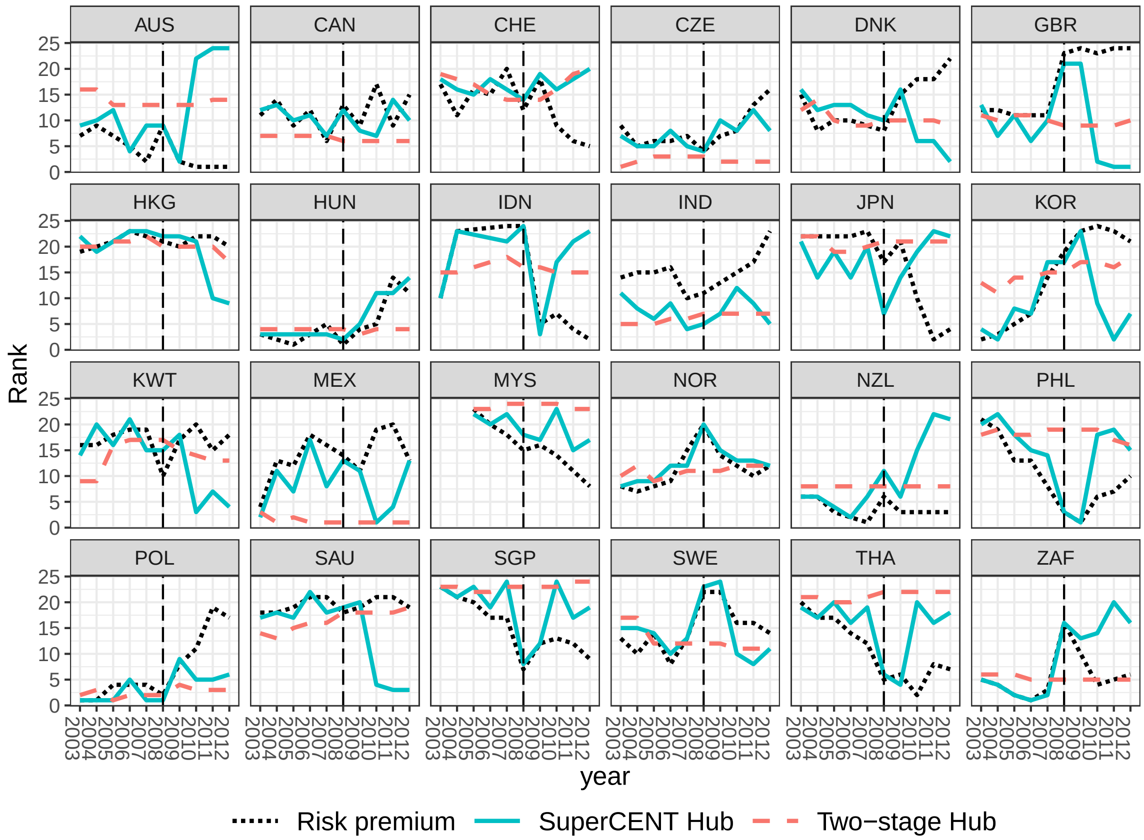


Figure 3. Time series of ranking of risk premium in descending order and ranking of hub centrality estimated by two-stage and SuperCENT in ascending order from 2003 to 2012. The vertical dashed line indicates 2008, the year of the financial crisis.

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 cross-validation. We will refer to SuperCENT-$\hat{\lambda}_{cv}$ as SuperCENT for simplicity and use the superscript sc for all the SuperCENT-$\hat{\lambda}_{cv}$-related estimates. We determine the signs of the centrality estimates by the empirical rule described at the end of Section 3.2. Figure 3

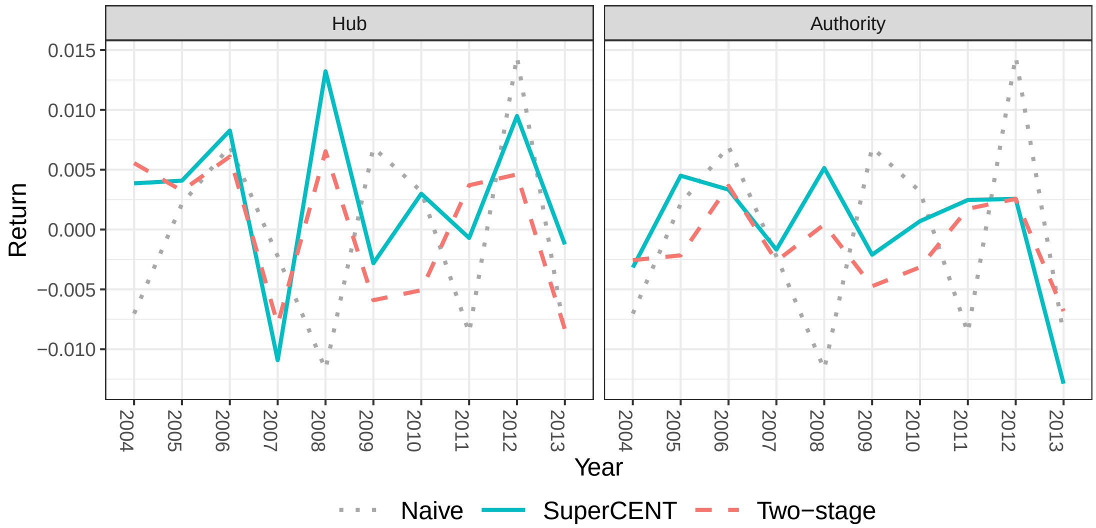


Figure 4. Time series of the next-year return from 2004 to 2013 based on a strategy that takes a long position on the currencies with the lowest 3 centralities and a short position on the currencies with the highest 3 centralities estimated from 2003 to 2012 respectively.

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; while countries like South Africa (ZAF) and New Zealand (NZL) are peripheral. Comparing 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.

Table 1: The 10-year average annualized return.

	Top/Bottom 3		Top/Bottom 4		Top/Bottom 5	
	Hub	Authority	Hub	Authority	Hub	Authority
Naive	-0.13%	-0.13%	-0.86%	-0.86%	-1.86%	-1.86%
SuperCENT CV	3.49%	0.03%	2.17%	0.65%	3.12%	0.28%
Two-stage	0.53%	-1.56%	1.28%	-1.16%	1.50%	-0.65%
SuperCENT Sharpe ratio	0.38	-0.02	0.27	0.07	0.39	0.02
Relative difference of return:						
SuperCENT vs Two-stage	554.5%	102.0%	70.3%	156.5%	108.1%	142.6%

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 based on 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. Similarly, we include a naive long-short strategy based on the return of year $t-1$ as a baseline. Figure 4 shows the year $t+1$ return based on this strategy. The centrality-based portfolios both outperform the naive 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 annualized return and Sharpe ratio (Sharpe, 1994) based on this strategy with the top and bottom 3, 4, and 5 currencies, respectively. The 10-year average return based on SuperCENT centralities doubles that of the two-stage procedure. SuperCENT achieves Sharpe ratios ranging from 0.27 to 0.39, compared to the Sharpe ratios of 0.28 for the Dow Jones, 0.42 for the S&P 500, and 0.39 for the NASDAQ over the sample period (2004-2013).

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 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	−0.0159*	(0.0083)	−0.0159*	(0.0083)	−0.0157***	(0.0039)
Hub	−0.0011	(0.0006)	−0.0011*	(0.0007)	−0.0020***	(0.0001)
Authority	−0.0005	(0.0006)	−0.0005	(0.0006)	−0.0004*	(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 and how much improvement SuperCENT can potentially achieve, we compare the noise-to-signal ratio κ of the trade network and the SNR of the regression. Since both quantities are unknown, we estimate using results from SuperCENT. For the noise-to-signal ratio, $\hat{\kappa}^{sc} = 0.36 \approx 2^{-1.5}$, which is larger than $\kappa = 2^{-8}$ in the simulation where the accuracy of the two-stage estimators is already low. For the SNR of the regression: $(\hat{\beta}_u^{sc}/\hat{\sigma}_y^{sc})^2 = 1.8 \times 10^7 \approx 2^{24}$ and $(\hat{\beta}_v^{sc}/\hat{\sigma}_y^{sc})^2 = 6.1 \times 10^5 \approx 2^{19}$. Compared with the simulation settings 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 significantly outperform the two-stage method in both the estimation and inference of β_u, due to the large value of $(\hat{\beta}_u^{sc}/\hat{\sigma}_y^{sc})^2$ and the fact that $|\hat{\beta}_u^{sc}| \gg |\hat{\beta}_v^{sc}|$ under a relatively large $\hat{\kappa}^{sc}$, while the improvement of β_v is less pronounced.

Table 2 shows the coefficient estimation, the standard error, and the significant level for the two-stage-adhoc, two-stage, and SuperCENT, respectively. The standard errors of two-stage and SuperCENT are based on the trade network being rank-one as we tested using the rank inference by Han et al. (2023) in Supplement S2. For the hub centrality β_u, (i) the estimate from the two-stage methods is −0.0011, while the estimate from SuperCENT is −0.0020, which is consistent with the inaccuracy we observed in the simulation; (ii) the standard errors from the two-stage methods are close to 0.0007, much larger than 0.0001 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; and (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 9; 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 properties of the commonly used two-stage procedure and that it could yield inaccurate centrality estimates and regression coefficient estimates, as well as invalid inference 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 improves over the two-stage in terms of centrality estimation, regression coefficient estimations, and inference. 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.

References

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. 6

Bonacich, P. and Lloyd, P. (2004). Calculating status with negative relations. *Social Networks*, 26(4):331–338. 6

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

Cartwright, D. and Harary, F. (1956). Structural balance: a generalization of heider's theory. *Psychological review*, 63(5):277. 6

Chiang, K.-Y., Hsieh, C.-J., Natarajan, N., Dhillon, I. S., and Tewari, A. (2014). Prediction and clustering in signed networks: a local to global perspective. *The Journal of Machine Learning Research*, 15(1):1177–1213. 6

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

Everett, M. G. and Borgatti, S. P. (2014). Networks containing negative ties. *Social Networks*, 38:111–120. 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

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

Gromov, D. (2025). Social balance-based centrality measure for directed signed networks. *Social Networks*, 80:1–9. 6

Han, X., Yang, Q., and Fan, Y. (2023). Universal rank inference via residual subsampling with application to large networks. *The Annals of Statistics*, 51(3):1109–1133. 30

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

Harary, F. (1953). On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2(2):143–146. 6

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

König, M. D., Rohner, D., Thoenig, M., and Zilibotti, F. (2017). Networks in conflict: Theory and evidence from the great war of africa. *Econometrica*, 85(4):1093–1132. 6

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

Labianca, G. and Brass, D. J. (2006). Exploring the Social Ledger: Negative Relationships and Negative Asymmetry in Social Networks in Organizations. *The Academy of Management Review*, 31(3):596–614. 6

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. 6

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, Y., Zhu, X., and Yu, Q. (2019). Clusters detection based leading eigenvector in signed networks. *Physica A: Statistical Mechanics and its Applications*, 523:1263–1275. 6

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. 6

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

Richmond, R. J. (2019). Trade network centrality and currency risk premia. *The Journal of Finance*, 74(3):1315–1361. 2, 5, 25, 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. 9, 10

Sharpe, W. F. (1994). The sharpe ratio. *Journal of portfolio management*, 21(1):49–58. 29

Singh, R. (2019). On eigenvector structure of weakly balanced networks. *Physica A: Statistical Mechanics and its Applications*, 527:121093. 6

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. 6

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

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

Zhang, Y., Levina, E., and Zhu, J. (2016). Community detection in networks with node features. *Electronic Journal of Statistics*, 10(2):3153–3178. 6

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