Statistica Sinica Preprint No: SS-2025-0101

Title	Multilayer Network Regression with Eigenvector Centrality and Community Structure
Manuscript ID	SS-2025-0101
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0101
Complete List of Authors	Zhuoye Han, Tiandong Wang and Zhiliang Ying
Corresponding Authors	Tiandong Wang
E-mails	td_wang@fudan.edu.cn
Notice: Accepted author version.	

Statistica Sinica

Multilayer Network Regression with Eigenvector Centrality and Community Structure

Zhuoye Han

School of Mathematical Sciences, Fudan University

Tiandong Wang

Shanghai Center for Mathematical Sciences, Fudan University

Shanghai Academy of Artificial Intelligence for Science

Zhiliang Ying

Department of Statistics, Columbia University

Abstract: In the analysis of complex networks, centrality measures and community structures play pivotal roles. For multilayer networks, a critical challenge lies in effectively integrating information across diverse layers while accounting for the dependence structures both within and between layers. We propose an innovative two-stage regression model for multilayer networks, combining eigenvector centrality and network community structure within fourth-order tensor-like multilayer networks. We develop new community-based centrality measures, integrated into a regression framework. To address the inherent noise in network data, we conduct separate analyses of centrality measures with and without measurement errors and establish consistency for the least squares estimates in the regression model. The proposed methodology is applied to the world input-output dataset, investigating how input-output network data among different countries and industries influence the gross output of each industry.

Key words and phrases: Multilayer Networks, Centrality Measures, Network Regression, Measurement Errors, World Input-Output Database.

1. Introduction

The inherent heterogeneity of real-world systems has led to multilayer networks as a powerful framework for understanding complex network mechanisms. While single-layer network analysis has yielded significant insights across various domains - from social models examining network effects on advertisement (Banerjee et al., 2019) to economic networks studies investigating market structure (Allen et al., 2019) and currency risk premium (Cai et al., 2021; Richmond, 2019) - these approaches typically analyze multilayer data in isolation, neglecting potential inter-layer dependencies. For instance, conventional single-layer analyses of Twitter data treat friendship, reply, and retweet networks separately, disregarding their inherent interdependencies. In contrast, a multilayer network approach conceptualizes these aspects as interconnected layers, enabling a more comprehensive understanding of both intra- and inter-layer dependence structures.

In this paper, we develop new community-based centrality measures for multilayer networks, integrated into a two-stage regression framework. Centrality measures, which quantify node importance, serve as fundamental explanatory variables in single-layer network analysis (Cai et al., 2021; Cai, 2022). Among various proposed measures (Rodrigues, 2019), eigenvector centrality (Bonacich, 1987) has emerged as particularly useful. This concept has been extended to multilayer networks, with our work building upon the tensor-based multilayer network structure introduced by Domenico et al. (2013) and its corresponding eigenvector-like centrality definition presented by De Domenico et al. (2015). Community structure, another fundamental aspect of complex networks, characterizes node similarity through densely connected subgroups. In single-layer networks, community detection has proven invaluable for interpreting complex relational dynamics, with significant methodological advancements (Pothen, 1997; Newman and Girvan, 2004; Palla et al., 2005; Fortunato, 2010). However, existing research has predominantly focused on detection algorithms, with limited integration of community structure with node centrality measures for macroscopic network characterization.

Recent methodological developments in single-layer network analysis provide important foundations for our work. Le and Li (2022) examine linear regression on multiple network eigenvectors, incorporating measurement error and developing inference methods for hypothesis testing. Cai (2022) establish theoretical properties for linear regression with multiple centrality measures in sparse single-layer networks, including consistency and distributional theory for the least squares estimators. Furthermore, Cai et al. (2021) propose the SuperCENT framework, which systematically investigates relationships between monoplex network centralities and response variables of interest.

While these single-layer approaches provide valuable insights, they prove inadequate for multilayer network analysis, as they typically examine layers in isolation, disregarding inter-layer connections. Our work addresses this limitation by developing a tailored analytical framework that explicitly considers layer interactions. By imposing appropriate constraints on community structure, we obtain estimators with superior asymptotic properties for general multilayer networks.

This paper makes three primary contributions: (1) developing a novel method for extracting community-based node centrality features from tensor-form multilayer networks; (2) establishing relationships between these features and macroscopic node characteristics; and (3) deriving asymptotic properties of parameter estimators in our proposed regression model, accommodating both error-free and measurement-error scenarios.

The rest of this paper is organized as follows. Section 2 proposes a multilayer regression framework that incorporates community-based centrality scores into tensorial multilayer networks. Section 3 presents the theoretical properties of the least squares estimators under mild conditions, measurement-error, and error-free scenarios. Section 4 evaluates the finitesample performance of our methodology through simulation studies, while Section 5 demonstrates its practical utility through a real-world application. Proofs of all theorems are provided in the supplementary materials.

2. Methodology

2.1 The multilayer networks framework

In this section, we introduce the framework of multilayer networks. We focus on networks where different layers share a common vertex set, as well as their tensor-based representation. This representation captures both intra- and inter-layer connections, which are essential for comprehensive network analysis. Traditional definition of multilayer networks excluding inter-layer connections is shown below (Tudisco et al., 2018).

Definition 1. An undirected multiplex network $\mathcal{G}$ with L layers is a collection of L graphs:

$$\mathcal{G} = \left\{G^{(\ell)} \equiv A^{(\ell)}\right\}_{\ell\in\mathcal{L}}. \tag{2.1}$$

Each $G^{(\ell)}$ represents the graph of the ℓ-th layer network, where $\mathcal{L} = \{1, \ldots, \mathsf{L}\}$ denotes the set of layers sharing the same node set $\mathcal{V} = \{1, 2, \ldots, \mathsf{N}\}$. For each $\ell \in \mathcal{L}$, $G^{(\ell)}$ is represented by a symmetric adjacency matrix $A^{(\ell)} = \left(A_{ij}^{(\ell)}\right) \in \mathbb{R}^{\mathsf{N}\times\mathsf{N}}$ with nonnegative entries. The entry $A_{ij}^{(\ell)} = \omega_{ij}^{\ell}$

if an edge exists between node i and node j in layer ℓ with weight ω_{ij}^{ℓ}, otherwise $A_{ij}^{(\ell)} = 0$. The network is considered unweighted if all edges have identical weights.

While Definition 1 provides a matrix-based representation, it fails to capture inter-layer connections crucial for multilayer network analysis. These connections have significant practical implications across various domains. In economic input-output networks, nodes represent economic sectors, layers correspond to countries, and inter-layer connections capture international trade relationships. In social networks, nodes represent users, layers represent platforms (e.g., Facebook, Instagram, Twitter), and inter-layer connections reflect cross-platform interactions. Similarly, in transportation networks, nodes represent hubs, layers represent transportation modes (e.g., buses, subways, trains), and inter-layer connections quantify transfer intensities between modes. To adequately model these interactions, we employ a tensor representation (Definition 2) that captures both intra- and inter-layer connections Domenico et al. (2013).

Definition 2. Let $\mathcal{B} = (\mathcal{B}_{i\ell j\ell'}) \in \mathbb{R}^{\mathsf{N}\times\mathsf{L}\times\mathsf{N}\times\mathsf{L}}$ denote the fourth-order adjacency tensor of an undirected weighted multilayer network, where each element is defined as

$$\mathcal{B}_{i\ell j\ell'} = \begin{cases} \omega_{i\ell j\ell'}, \text{ if } (v_i^{\ell}, v_j^{\ell'}) \in E \\ 0, \text{ otherwise}, \end{cases} \tag{2.2}$$

with $i, j \in \mathcal{V}, \ell, \ell' \in \mathcal{L}$, v_i^{ℓ} representing node i in layer ℓ, E denoting the edge set, and $\omega_{i\ell j\ell'}$ representing the edge weight between node i in layer ℓ and node j in layer ℓ'.

For multiplex networks with layers connected solely through shared nodes, we have:

$$\omega_{i\ell j\ell'} = \begin{cases} 1, \text{ if } i = j, \ell \neq \ell' \\ 0, \text{ otherwise}. \end{cases} \tag{2.3}$$

Following Domenico et al. (2013), we construct the supra-adjacency matrix associated with the multilayer adjacency tensor $\mathcal{B}$ as an $\mathsf{NL} \times \mathsf{NL}$ block matrix:

$$B_0 = \begin{bmatrix} A^{(1)} & A^{(1,2)} & \cdots & A^{(1,\mathsf{L})} \\ A^{(2,1)} & A^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & A^{(\mathsf{L}-1,\mathsf{L})} \\ A^{(\mathsf{L},1)} & \cdots & A^{(\mathsf{L},\mathsf{L}-1)} & A^{(\mathsf{L})} \end{bmatrix}, \tag{2.4}$$

where $A^{(\ell)}$ are weighted symmetric adjacency matrices for each layer, and $A^{(\ell,\ell')}$ represent inter-layer adjacency matrices.

2.2 Eigenvector centrality with community structure

In this section, we introduce the construction of eigenvector-like centrality and community-based eigenvector-like centrality for multilayer networks.

2.2.1 Construction of eigenvector-like centrality

Building upon Bonacich (1972)'s definition of eigenvector centrality for single-layer networks, we extend this concept to multilayer networks using the tensor-based framework introduced by De Domenico et al. (2015). For a single-layer network $A = (a_{ij})$, the centrality $c(i)$ of node i satisfies $\lambda c(i) = \sum_{j=1}^{n} a_{ij} c(j)$, $\forall i = 1, \cdots, \mathsf{N}$. This leads to the eigenvector equation $Ac = \lambda c$, where c is the eigenvector corresponding to the largest eigenvalue of A. As A is the adjacency matrix of an undirected (connected) graph, A is positive semi-definite and due to the Perron-Frobenius theorem (Keener, 1993), there exists an eigenvector of the maximal eigenvalue with only nonnegative (positive) entries.

For tensor-based multilayer networks, we define the eigenvector centrality through the matrix $V = (V_{i\alpha}) \in \mathbb{R}^{\mathsf{N} \times \mathsf{L}}$ satisfying

$$\sum_{j=1}^{n} \sum_{\beta=1}^{\mathsf{L}} \mathcal{B}_{i\alpha j\beta} V_{j\beta} = \lambda_1 V_{i\alpha}, \tag{2.5}$$

for $i \in \mathcal{V}$ and $\alpha \in \mathcal{L}$. Here $\mathcal{V}$ denotes the set of vertices and $\mathcal{L}$ denotes the layer set. Then the eigenvector centrality of node i is defined as the i-th row of the matrix V. Here in multilayer networks, the centrality of a node is determined by summing the centralities of its neighbors within the same layer and across different layers, weighted by the strength of both intra-layer and inter-layer connections. Using the supra-adjacency matrix from (2.4), this becomes:

$$B_0 \operatorname{vec}(V) = \lambda_1 \operatorname{vec}(V). \tag{2.6}$$

Here λ_1 in (2.5) and (2.6) is defined as the spectral radius of positive semi-definite matrix B_0. The associated eigenvector is $\operatorname{vec}(V)$, where $\operatorname{vec}(\cdot)$ denotes the standard vectorization operator. By the Perron-Frobenius Theorem, the leading eigenvector is the unique eigenvector that can be chosen so that every entry is non-negative and $\|V\|_F = 1$, motivating its use as a centrality measure. In the following regression step, we consider the covariate $C \in \mathbb{R}^{\mathsf{N}\times\mathsf{L}}$ such that

$$C = a_{\mathsf{N}} V, \tag{2.7}$$

where a_{N} is a scaling parameter related to N. While eigenvector centrality is often normalized to length 1 (Banerjee et al., 2012; Cruz et al., 2017), some studies treat a_{N} as a free parameter (Cai et al., 2021; Cai, 2022). Following Cai (2022), we leave a_{N} as a free parameter to thoroughly examine its impact on regression properties.

Eigenvector-like centrality is valuable in real networks by capturing influence through interconnected relationships. For instance, in economic input-output networks, it reflects sectoral importance across domestic and global links, thereby revealing global economic dynamics. In multilayer social networks, eigenvector-like centrality measures cross-platform influence. Similarly, in multilayer transportation networks, it highlights key hubs via intra- and inter-modal connectivity.

Moreover, to account for real-world measurement noise, we model the observed network B as

$$B = B_0 + E_0, \tag{2.8}$$

where E_0 represents symmetric random noise and B_0 denotes the real undirected network structure. Also, we assume E_0 as a symmetric random matrix since in what follows, we only consider undirected weighted graphs with undirected inter-layer interactions.

2.2.2 Community structure and community-based centrality

Assuming a shared community structure across all layers, we define community assignments $c_i \in \{1, \ldots, \mathsf{R}\}$ for each node $i \in \mathcal{V}$, with R communities of sizes N_r ($\sum_{r=1}^{\mathsf{R}} \mathsf{N}_r = \mathsf{N}$). Following Shen et al. (2022), we construct an $\mathsf{N} \times \mathsf{R}$ community matrix S, where $S_{ij} = 1$ if vertex i belongs to community

j and 0 otherwise. The columns of S are mutually orthogonal, each row sums to 1, and S satisfies $\mathrm{tr}(S^\top S) = \mathsf{N}$ and $S^\top S = \mathrm{diag}\{\mathsf{N}_1, \mathsf{N}_2, \ldots, \mathsf{N}_\mathsf{R}\}$.

Given the community information, we construct the community-based eigenvector centrality matrix $U \in \mathbb{R}^{\mathsf{N}\times\mathsf{L}}$ as

$$U := S(S^\top S)^{-1}(S^\top C), \tag{2.9}$$

where $C \in \mathbb{R}^{\mathsf{N}\times\mathsf{L}}$ is the node-layer centrality matrix as in Equation (2.7). Here $U_{i\ell}$ is the average centrality for the community of node i in layer ℓ.

To address the rank deficiency issue when $\mathsf{L} > \mathsf{R}$, we define the scalar community-based centrality $Z \in \mathbb{R}^{\mathsf{N}\times 1}$ as the average across layers

$$Z := \frac{1}{\mathsf{L}} S(S^\top S)^{-1}(S^\top C)\mathbf{1}_\mathsf{L}, \tag{2.10}$$

Entrywise, we have $Z_i := \frac{1}{\mathsf{L}\mathsf{N}_{c_i}} \sum_{\ell=1}^{\mathsf{L}} \sum_{j\in c_i} C_{j\ell}$, where c_i is the community assignment of node i, and N_{c_i} is the size of that community. This formulation reveals that nodes within the same community share the same community-based centrality in Z, which represents the average centrality across all members of the community and across all layers.

2.3 Multilayer network regression

In this section, we aim to incorporate the multilayer network structure into regression analysis by leveraging eigenvector-like centrality features. Building on the definitions in Section 2.2, we first propose a centrality-based multilayer network regression (C-MNetR) model. However, due to the potential rank deficiency of the centrality matrix C, especially when measurement error is present, a naive plug-in approach may lead to inconsistency in the estimation of coefficients. Hence, we explore two remedies: (i) imposing a baseline constraint, and (ii) incorporating community structure to construct a more robust alternative model, centrality- and community-based multilayer network regression (CC-MNetR).

2.3.1 C-MNetR

Let X denote the covariate matrix of size $\mathsf{N}\times\mathsf{P}$, where P is the number of covariates. The Centrality-based Multilayer Network Regression (C-MNetR) model is specified by

$$y = X\beta_X + C\beta_C + \varepsilon, \tag{2.11}$$

$$\text{where} \qquad B_0\,\mathrm{vec}(V) = \lambda_1\,\mathrm{vec}(V), \quad C = a_{\mathsf{N}}V.$$

Here, β_X and β_C are the regression parameters that can be estimated by the ordinary least squares (OLS) method. When there is no measurement error, C can be uniquely constructed with observations $\{B_0, X, y\}$. The

OLS estimator of $\beta = (\beta_X^\top, \beta_C^\top)^\top$ takes for

$$\hat{\beta}^{(ols)} = (\mathbf{W}_1^\top \mathbf{W}_1)^{-1} \mathbf{W}_1^\top y, \text{ where } \mathbf{W}_1 = (X, C). \tag{2.12}$$

When there is measurement error, C is not directly observed but can be estimated from the observations $\{B, X, y\}$. Specifically, we first compute the top eigenvector of B, denoted as $\text{vec}(\hat{V})$, and subsequently derive $\hat{C} = a_{\mathsf{N}}\hat{V}$. With $\hat{\mathbf{W}}_1 = (X, \hat{C})$, the OLS estimator is

$$\hat{\beta} = (\hat{\mathbf{W}}_1^\top \hat{\mathbf{W}}_1)^{-1} \hat{\mathbf{W}}_1^\top y. \tag{2.13}$$

However, when there is measurement error in the network, the estimated centrality matrix $\hat{C} = a_{\mathsf{N}}\hat{V}$ may not retain the full-column-rank property of C, even if C is theoretically full rank in the noiseless case. This is because the presence of noise in B perturbs the eigenstructure, and $\hat{C}$ is constructed from a potentially misspecified leading eigenvector. As a result, the design matrix $\hat{\mathbf{W}}_1 = (X, \hat{C})$ may be rank-deficient or near-singular, leading to unstable or inconsistent estimation of β_C.

2.3.2 Remedies of C-MNetR

To address the previously discussed limitations of C-MNetR, we propose two complementary strategies: (i) imposing a baseline constraint to interpret coefficients as relative effects across layers, and (ii) aggregating node-layer centrality scores using known community structure. The latter leads to the centrality- and community-based multilayer network regression (CC-MNetR) model.

C-MNetR with baseline constraint. To improve identifiability and interpretability, we set $\beta_{C_1} = 0$, making other coefficients reflect deviations from a baseline layer. This improves numerical stability but does not fully resolve the rank deficiency, especially as the number of layers L grows with N and C becomes increasingly collinear. Thus, the constraint is a practical fix, not a complete solution.

CC-MNetR. Alternatively, we leverage community structure by aggregating centrality features within communities to obtain more stable and interpretable variables. This approach mitigates rank deficiency and is robust to measurement errors. We define the CC-MNetR model as follows:

$$y = X\beta_X + Z\beta_Z + \varepsilon, \tag{2.14}$$

$$\text{where} \quad B_0 \,\mathrm{vec}(V) = \lambda_1 \,\mathrm{vec}(V), \quad C = a_{\mathsf{N}} V,$$

$$Z = \frac{1}{\mathsf{L}} S(S^\top S)^{-1}(S^\top C)\mathbf{1}_{\mathsf{L}}.$$

For CC-MNetR, in the absence of measurement error, the OLS estimator of coefficient $\beta = (\beta_X^\top, \beta_Z^\top)^\top$ is

$$\tilde{\beta}^{(ols)} = (\mathbf{W}_2^\top \mathbf{W}_2)^{-1} \mathbf{W}_2^\top y \tag{2.15}$$

where $\mathbf{W}_2 = (X, Z)$ with observations $\{B_0, S, H, X, y\}$.

When measurement error exists, C can be estimated from $\{B, S, H, X, y\}$ in a similar way. Specifically, we compute the top eigenvector of B, $\text{vec}(\hat{V})$, and

$$\hat{C} = a_{\mathsf{N}} \hat{V}, \quad \hat{U} = S(H \bullet S^\top \hat{C}), \quad \hat{Z} = \frac{1}{\mathsf{L}} \hat{U}\, 1_{\mathsf{L}}\,. \tag{2.16}$$

Let $\tilde{\mathbf{W}}_2 = (X, \hat{Z})$ and

$$\tilde{\beta} = (\tilde{\mathbf{W}}_2^\top \tilde{\mathbf{W}}_2)^{-1} \tilde{\mathbf{W}}_2^\top y. \tag{2.17}$$

The CC-MNetR model leverages community averaging to reduce noise and ensure consistent estimation. Although it assumes a known community matrix S, in practice S is typically estimated (e.g. via spectral clustering (Han et al., 2015)), which may introduce errors. We analyze this impact and provide conditions for estimator consistency when S is estimated; see Section S1.8 of the supplement for details.

Overall, although C-MNetR may face identifiability issues, incorporating baseline constraints or leveraging community structure offers effective remedies. The CC-MNetR model, in particular, improves robustness and enjoys theoretical guarantees, as established in Section 3.

3. Theoretical results

3.1 Model assumptions

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{\mathsf{NL}}$ be eigenvalues of the noiseless network B_0. Define projection matrices $P_X := X(X^\top X)^{-1}X^\top$ and let $\sigma_{\min}((I_{\mathsf{N}} - P_X)V)$ be the smallest singular value of $(I_{\mathsf{N}} - P_X)V$. For a vector $\nu \equiv (\nu_j)_{j=1}^{\mathsf{N}}$, its ℓ_1- and ℓ_2-norms are defined as $\|\nu\|_1 := \sum_{j=1}^{\mathsf{N}} |\nu_j|$ and $\|\nu\|_2 := \left(\sum_{j=1}^{\mathsf{N}} \nu_j^2\right)^{\frac{1}{2}}$, respectively. Additionally, the matrix operator norm of matrix E_0 is defined as $\|E_0\|_2 := \max_{\|u\|_2 \leq 1} \|E_0 u\|_2$. Let C_i denotes the centrality of node i in all L layers with $C = \left[C_1^\top, \cdots, C_{\mathsf{N}}^\top\right]^\top$. We now give crucial model assumptions to establish theoretical properties of the least square estimators.

Assumption 1 (Noise Structure). *The regression noises, ε_i, $1 \leq i \leq \mathsf{N}$, are i.i.d random variables with $\mathbb{E}\left[\varepsilon_i | X, C, E_0\right] = 0$, $Var(\varepsilon_i | X, C, E_0) = \sigma_y^2$, and $\mathbb{E}[\varepsilon_i^4 | X, C, E_0] \leq k_0 < \infty$. The network noise E_0 satisfies $\mathbb{E}\left[\|E_0\|_2^2\right] = O(\mathsf{NL})$.*

We give examples satisfying Assumption 1. Let E_0 be a block diagonal matrix with blocks $E_{01}, E_{02}, \cdots, E_{0\mathsf{L}}$ on the diagonal, where E_{0i},

$i = 1, \cdots, \mathsf{L}$, are symmetric $\mathsf{N} \times \mathsf{N}$ matrices with upper-diagonal entries being i.i.d $\mathcal{N}(0, \sigma_b^2)$. Then from Lemmas 3.7, 3.8, and 3.11 in van Handel (2017), such structure ensures $\mathbb{E}\left[\|E_0\|_2^2\right] = O(\mathsf{NL})$. If instead E_0 is a symmetric matrix with i.i.d. standard Gaussian entries on and above the diagonal, then Lemma 3.11 in van Handel (2017) gives $\mathbb{E}\left[\|E_0\|_2^2\right] = O(\mathsf{NL})$.

Assumption 2. ***(Design Matrix)*** *The design matrix $X \in \mathbb{R}^{\mathsf{N} \times \mathsf{P}}$ satisfies $\mathsf{N} > \mathsf{P} + \mathsf{L}$ for fixed P, and $\frac{\mathsf{L}}{\mathsf{N}} \to 0$ as $\mathsf{N} \to \infty$. For any i, j, k, l (with $i \neq k$ or $j \neq l$), given C and E_0, X_{ij} and X_{kl} are independent. For fixed j, $\{X_{ij}\}$ are i.i.d. with $\mathbb{E}[X_{ij}|C, E_0] = 0$, $\mathbb{E}[X_{ij}^2|C, E_0] < \infty$.*

Note that from Assumption 2, $\frac{1}{\mathsf{N}} X^\top X \xrightarrow{P} V_X$ where V_X is a nonsingular diagonal matrix.

Assumption 3. ***(Identifiability)*** *There exists $l_{\mathsf{N}} > 0$ such that*

$$\sigma_{\min}((I_{\mathsf{N}} - P_X)V) \geq l_{\mathsf{N}} > 0. \tag{3.18}$$

Condition (3.18) in Assumption 3 ensures that $\mathbf{W}_1 = (X, C)$ has full column rank (see Supplement Section S1.1 for details). Multilayer networks lacking inter-layer connectivity violate this assumption, as the top eigenvector of a block-diagonal matrix $B = \operatorname{diag}\{A_1, \ldots, A_{\mathsf{L}}\}$ may replicate a single-layer eigenvector, yielding columns of zeros in C and thus rank deficiency. This condition is more likely to hold when the network exhibits strong and heterogeneous inter-layer connections, allowing the global eigenvector centrality C to capture layer-specific structures and maintain column independence.

Next, we impose constraints on the community structure, which is a key factor that allows us to obtain Theorem 4.

Assumption 4. ***(Centrality and Community Structure)*** *The community sizes $\{\mathsf{N}_i\}_{i=1}^{\mathsf{R}}$ are independent from X, and with probability 1,* $\min_i \frac{\mathsf{N}_i}{\mathsf{N}} > \epsilon$ *for some $\epsilon > 0$. Furthermore,* $\min_{1\le i\le \mathsf{N}} \|C_i\|_1^2 \asymp \frac{a_{\mathsf{N}}^2 \mathsf{L}}{\mathsf{N}}$, *i.e.* $\min_{1\le i\le \mathsf{N}} \|V_i\|_1^2 \asymp \frac{\mathsf{L}}{\mathsf{N}}$ *where $V_i = C_i/a_{\mathsf{N}}$.*

Assumption 4 restricts the minimum sum of centrality values for each node across all layers, excluding nodes with low centrality in all layers. This ensures that the analysis focuses on nodes with significant influence or importance in the network structure, regardless of their specific layer.

Assumption 5. ***(Spectral Gap)*** *Let λ_i be the i-th largest eigenvalue of B_0, and define the spectral gap $\delta := \lambda_1 - \lambda_2$. Assume the spectral gap is sufficiently large so that $\frac{a_{\mathsf{N}}\sqrt{\mathsf{L}}}{\delta} \to 0$, as $\mathsf{N} \to \infty$.*

To ensure Assumption 5 holds, the network matrix B_0 must exhibit a sufficiently large spectral gap δ, reflecting significant differences in connection strengths among the blocks of its partitioned structure. Since different layers share the same community structure, this requires substantial variation in connection strengths across layers. A more detailed discussion on when this condition holds or fails is provided in Supplementary Section S1.9, where we also present an example under which $\delta \asymp \mathsf{NL}$.

In regression analysis, Assumptions 1-4 suffice for consistency in the absence of measurement errors, but Assumption 5 becomes necessary when measurement errors are introduced.

3.2 Centrality-based regression without measurement error

We first consider the case when the true network B_0, is observed. Under the C-MNetR framework, Theorem 1 shows the property of the OLS estimator under mild conditions.

Theorem 1. *Recall the estimator* $\hat{\beta}^{(ols)} := (\hat{\beta}_X^{(ols)\top}, \hat{\beta}_C^{(ols)\top})^\top$ *defined in* (2.12). *Under Assumptions 1, 2 and 3, we have:*

(i) As $\mathsf{N} \to \infty$, $\sqrt{\mathsf{N}}(\hat{\beta}_X^{(ols)} - \beta_X) \xrightarrow{d} \mathcal{N}(0, \sigma_y^2 V_X^{-1})$.

(ii) If $\frac{a_{\mathsf{N}} l_{\mathsf{N}}}{\sqrt{\mathsf{L}}} \to \infty$ *as* $\mathsf{N} \to \infty$, *then* $\hat{\beta}_C^{(ols)} \xrightarrow{L_2} \beta_C$ *and* $\hat{\beta}_C^{(ols)} \xrightarrow{P} \beta_C$.

When $X = 0$ and $\mathsf{L} = 1$, our model reduces to the single-layer centrality regression in Cai (2022), where $P_X = 0$ and (3.18) simplifies to $\sigma_{\min}(V) \geq l_{\mathsf{N}}$. For V normalized with $\sigma_{\min}(V) = 1$, setting $l_{\mathsf{N}} = 1$ recovers their condition $a_{\mathsf{N}} \to \infty$. When X is present, the scaling $a_{\mathsf{N}} l_{\mathsf{N}} \to \infty$ in our framework generalizes this requirement to high-dimensional settings where l_{N} may decay with N.

For CC-MNetR, we consider the community-based centrality Z without measurement error. Theorem 2 gives the properties of $\tilde{\beta}^{(ols)}$ and further establishes the asymptotic normality of $\tilde{\beta}_Z^{(ols)}$ when the network is observed without measurement error.

Theorem 2. *Recall the estimator* $\tilde{\beta}^{(ols)} = (\tilde{\beta}_X^{(ols)\top}, \tilde{\beta}_Z^{(ols)\top})^\top$ *defined in* (2.15), *Under Assumptions 1, 2 and 3, we have the following:*

(i) As $\mathsf{N} \to \infty$, $\sqrt{\mathsf{N}}(\tilde{\beta}_X^{(ols)} - \beta_X) \xrightarrow{d} \mathcal{N}(0, \sigma_y^2 V_X^{-1})$.

(ii) Let $a_{\mathsf{N}} \asymp \sqrt{\mathsf{NL}}$ *and Assumption 4 holds, then* $\tilde{\beta}_Z^{(ols)} \xrightarrow{L_2} \beta_Z$ *and consequently* $\tilde{\beta}_Z^{(ols)} \xrightarrow{P} \beta_Z$. *Moreover, we have* $\sqrt{\frac{Z^\top Z}{\sigma_y^2}} \left(\tilde{\beta}_Z^{(ols)} - \beta_Z\right) \xrightarrow{d} \mathcal{N}(0, 1)$.

A common practice to ensure that elements in C and X are of the same order is to set $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$. This aligns with existing work: Le and Li (2022) used a scaling parameter as $\sqrt{\mathsf{N}}$ in single-layer simulations, while Cai (2022) proposed a data-driven choice $a_{\mathsf{N}} = \sqrt{\lambda_1(\hat{A})}$, which also scales as $\sqrt{\mathsf{N}}$ under their assumptions. Both support $a_{\mathsf{N}} \asymp \sqrt{\mathsf{N}}$ as a natural choice, which generalizes to $\sqrt{\mathsf{NL}}$ in our multilayer settings to reflect the additional layer information.

3.3 Centrality-based regression with measurement error

Real networks often contain noise, which requires consideration of the influence of E_0. The Davis-Kahan theorem (Davis and Kahan, 1970) characterizes the ℓ_2-difference of the top eigenvector. Theorem 3 gives the consistency of the estimators with measurement error for C-MNetR.

Theorem 3. *Let* $\hat{\beta} := (\hat{\beta}_X^\top, \hat{\beta}_C^\top)^\top$ *be defined as in* (2.13)*. When* $a_{\mathsf{N}} \asymp \sqrt{\mathsf{NL}}$ *and Assumptions 1, 2, 3, and 5 hold, we have* $\hat{\beta}_X \xrightarrow{P} \beta_X$*.*

Essentially, the consistency of $\hat{\beta}_X$ is established without requiring the estimated design matrix $\hat{\mathbf{W}}_1 = (X, \hat{C})$ to have full column rank. Our analysis leverages projection-based formulations, which remain well-defined regardless of the rank of $\hat{C}$, allowing the estimator $\hat{\beta}_X$ to be robust to rank-deficiency that may arise from measurement errors in the network layer. Unfortunately, even when Assumption 5 holds and $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$, bias may persist in $\hat{\beta}_C$. This bias is mainly due to the rank deficiency or near rank deficiency of the estimated covariate matrix $\hat{C}$, as later confirmed by our numerical experiments in Table 2.

Note that the presence of measurement noise may lead to unstable or not column full-rank $\hat{C}$, which complicates the estimation and makes

$\hat{\beta}_C$ lack consistency. Therefore, we caution against the use of C-MNetR when reliable community structure information is not available or when measurement errors are significant.

On the other hand, CC-MNetR, which relies on the centrality- and community-based variable Z provides a solution. Theorem 4 guarantees the consistency of the OLS estimators in the presence of measurement errors.

Theorem 4. *Let* $\tilde{\beta} \equiv (\tilde{\beta}_X^\top, \tilde{\beta}_Z^\top)^\top$ *be defined as in* (2.17)*. When* $a_\mathsf{N} \asymp \sqrt{\mathsf{NL}}$ *and Assumptions 1-5 hold,* $\tilde{\beta}_X \xrightarrow{P} \beta_X$ *and* $\tilde{\beta}_Z \xrightarrow{P} \beta_Z$*.*

Theorem 5 shows that even in the presence of measurement error, $\tilde{\beta}_Z$ remains asymptotically normal under the spectral gap condition.

Theorem 5. *Suppose Assumptions 1-4 hold. If* $a_\mathsf{N} \asymp \sqrt{\mathsf{NL}}$ *and* $\frac{a_\mathsf{N}\sqrt{\mathsf{N}}}{\delta} \to 0$ *as* $\mathsf{N} \to \infty$*, then* $\sqrt{\mathsf{N}}(\tilde{\beta}_X - \beta_X) \xrightarrow{d} \mathcal{N}(0, \sigma_y^2 V_X^{-1})$*, and* $\sqrt{\frac{Z^\top Z}{\sigma_y^2}}\left(\tilde{\beta}_Z - \beta_Z\right) \xrightarrow{d} \mathcal{N}(0,1)$*.*

The detailed proofs of Theorem 1-5 are collected in the supplement.

Remark 1 (Standard Error Estimation)**.** The asymptotic normality result in Theorem 5 facilitates inference via standard error estimation. Practically, the standard errors for $\tilde{\beta}_X$ and $\tilde{\beta}_Z$ can be obtained by plug-in methods using sample quantities. Specifically, a consistent standard error estimator for $\tilde{\beta}_X$ is $\widehat{\mathrm{SE}}(\tilde{\beta}_X) = \sqrt{\hat{\sigma}_y^2 (X^\top X)^{-1}}$, where $\hat{\sigma}_y^2$ is the sample residual variance from regressing y on X and the estimated network centrality $\hat{Z}$. Similarly, the standard error of $\tilde{\beta}_Z$ can be estimated by $\widehat{\mathrm{SE}}(\tilde{\beta}_Z) = \sqrt{\frac{\hat{\sigma}_y^2}{Z^\top Z}}$.

Remark 2 (The Role of Layer Number L)**.** The validity of consistency and asymptotic normality results without measurement error in Theorem 2 depends on the signal-to-noise ratio as captured by the spectral gap δ of the noiseless matrix B_0. In particular, when $\delta \asymp \mathsf{NL}$, both consistency and asymptotic normality hold even when the number of layers L is fixed.

In contrast, under measurement error, Theorem 5 requires the condition $\frac{a_{\mathsf{N}}\sqrt{\mathsf{N}}}{\delta} \to 0$ to ensure asymptotic normality. Given that $a_{\mathsf{N}} \asymp \sqrt{\mathsf{NL}}$ and $\delta \asymp \mathsf{NL}$, this condition simplifies to $\frac{1}{\sqrt{\mathsf{L}}} \to 0$, which implies that $\mathsf{L} \to \infty$. That is, even when the spectral gap is sufficiently large, an increasing number of layers is still required to offset the impact of measurement error. This highlights a key distinction between the noiseless and noisy regimes: while a fixed number of layers suffices when the network is observed without error, robustness to noise necessitates layer-wise aggregation to reduce variance.

Remark 3 (Sensitive analysis)**.** We complement Theorem 5 with a sensitivity analysis under increasing network noise levels σ_b. As shown in Section S2.4 of the Supplementary Material, CC-MNetR remains stable under moderate noise and gradually degrades as noise variance increases.

4. Simulation Experiments

In this section, we use simulated data to examine the performance of the proposed C-MNetR and CC-MNetR models, both with and without measurement errors.

4.1 Data Generation without Measurement Error

We simulate a multilayer network with $\mathsf{L} = 2$, where each layer consists of N nodes, partitioned into $\mathsf{R} = 3$ equal-sized communities. The overall network is represented by a supra-adjacency matrix $B_0 \in \mathbb{R}^{2\mathsf{N}\times 2\mathsf{N}}$, constructed from intra- and inter-layer weighted adjacency matrices.

Network construction. For each layer $\ell = 1, 2$, the intra-layer structure is generated from a stochastic block model (SBM) introduced in Holland et al. (1983) with the same community labels across layers. Let $T^{(\ell)} \in \{0,1\}^{\mathsf{N}\times\mathsf{N}}$ denote the binary adjacency matrix of layer ℓ, and $P^{(\ell)} \in [0,1]^{\mathsf{R}\times\mathsf{R}}$ be the community-wise connection probability matrix in layer ℓ. The connection probability matrices for the two layers are

$$P^{(1)} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}, \quad \text{and} \quad P^{(2)} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.9 \end{bmatrix}.$$

Given the community labels $c_1, \ldots, c_{\mathsf{N}} \in \{1, 2, 3\}$, the entries of $T^{(\ell)}$ are independently sampled as

$$T_{ij}^{(\ell)} \mid c_i, c_j \sim \text{Bernoulli}(p_{c_i,c_j}^{(\ell)}), \quad \text{for } i < j,$$

and symmetrized to form an undirected matrix with zero diagonals. Each edge $T_{ij}^{(\ell)} = 1$ is then assigned a weight from $U(1, 2)$, yielding the weighted adjacency matrix $A^{(\ell)}$.

We then simulate the inter-layer connections as

$$A_{ij}^{(1,2)} = T_{ij}^{(1,2)} W_{ij}, \tag{4.19}$$

where $T_{ij}^{(1,2)} \sim \text{Bernoulli}(p_{\text{inter}})$ with $p_{\text{inter}} = 0.2$ and $W_{ij} \sim U(1, 2)$, both independently sampled across all pairs (i, j) and mutually independent. To ensure symmetry of the overall supra-network, we set $A^{(2,1)} = (A^{(1,2)})^\top$. Then the full multilayer network is encoded by the symmetric supra-adjacency matrix $B_0 \in \mathbb{R}^{2\mathsf{N} \times 2\mathsf{N}}$ defined as

$$B_0 = \begin{bmatrix} A^{(1)} & A^{(1,2)} \\ A^{(2,1)} & A^{(2)} \end{bmatrix}.$$

Covariates and centrality measures. The covariate matrix $X \in \mathbb{R}^{\mathsf{N} \times 2}$ has entries independently drawn from $\mathcal{N}(0, 1)$. The centrality measures C

and Z are computed using the eigenvector centrality of B_0 as defined in Equation (2.16). The true coefficients are set as $\beta_X = (1, 2)^\top$, $\beta_C = (1, 2)^\top$, and $\beta_Z = 2$. The response variable is generated as $y = X\beta_X + C\beta_C + \varepsilon$, or $y = X\beta_X + Z\beta_Z + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0, \sigma_y^2 I_{\mathsf{N}})$, $\sigma_y = 1$.

4.2 Performance without measurement error

We now examine the performance of the OLS estimators. Set $\mathsf{N} \in \{100, 200, 500, 1000\}$, $\mathsf{L} = 2$, $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$ and generate $n = 1000$ replications for each choice of N.

N	**C-MNetR**				**CC-MNetR**		
	$\hat{\beta}_{X_1}^{(ols)}$	$\hat{\beta}_{X_2}^{(ols)}$	$\hat{\beta}_{C_1}^{(ols)}$	$\hat{\beta}_{C_2}^{(ols)}$	$\tilde{\beta}_{X_1}^{(ols)}$	$\tilde{\beta}_{X_2}^{(ols)}$	$\tilde{\beta}_{Z}^{(ols)}$
Truth	1	2	1	2	1	2	2
100	1.0009 (0.1033)	1.9970 (0.1032)	0.9660 (1.1546)	2.0335 (1.1734)	1.0010 (0.1024)	1.9971 (0.1026)	1.9988 (0.0984)
200	0.9994 (0.0726)	1.9996 (0.0694)	1.0426 (1.1295)	1.9603 (1.1111)	0.9992 (0.0727)	2.0000 (0.0693)	2.0024 (0.0740)
500	1.0024 (0.0452)	2.0003 (0.0436)	1.0100 (0.9434)	1.9897 (0.9280)	1.0024 (0.0451)	2.0003 (0.0435)	1.9996 (0.0456)
1000	1.0018 (0.0316)	1.9996 (0.0308)	0.9807 (0.7679)	2.0192 (0.7554)	1.0018 (0.0316)	1.9996 (0.0308)	2.0002 (0.0309)

Table 1: Mean and standard deviations of parameter estimates for C-MNetR and CC-MNetR models at different network sizes (N).

Table 1 reports mean and standard deviations of parameter estimates for C-MNetR and CC-MNetR across different sample sizes N. In C-MNetR, estimates of $\hat{\beta}_{X_1}^{(ols)}$ and $\hat{\beta}_{X_2}^{(ols)}$ are consistent and improve with larger N, whereas $\hat{\beta}_{C_1}^{(ols)}$ and $\hat{\beta}_{C_2}^{(ols)}$ show greater variability, suggesting potential rank deficiency in C. On the other hand, CC-MNetR yields stable and consistent estimates for all parameters, $\tilde{\beta}^{(ols)}_{X_1}$, $\tilde{\beta}^{(ols)}_{X_2}$, and $\tilde{\beta}^{(ols)}_{Z}$, with decreasing standard deviations as N increases.

Hence, we conclude that CC-MNetR outperforms C-MNetR in consistency and stability, especially for centrality-related parameters, where C-MNetR shows bias likely due to rank deficiency in C. Boxplots and QQ plots for the noiseless case are provided in Supplement Section S2.1.

4.3 Data Generation with Measurement Error

We consider a setting where measurement error exists. For comparison, we adopt the same sample sizes as in the noiseless case: $\mathsf{N} \in \{100, 200, 500, 1000\}$. Also, suppose $\mathsf{L} = \lfloor \mathsf{N}^{1/4} \rfloor$, allowing it to grow with N, and set $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$. For each configuration, we generate $n = 1000$ replications.

Network construction. The network generation process follows a similar pattern as in Section 4.1. Community labels $c_1, \ldots, c_{\mathsf{N}} \in \{1, 2, 3\}$ are shared across all layers, and edge weights are drawn from $U(1, 2)$. To introduce heterogeneity, the probability matrices of community connections $P^{(\ell)}$ vary by layer. With $\mathsf{L} \leq 5$, we assume

$$P^{(1)} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}, \; P^{(2)} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.9 \end{bmatrix}, \; P^{(3)} = \begin{bmatrix} 0.75 & 0.15 & 0.1 \\ 0.15 & 0.75 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix},$$

$$P^{(4)} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.85 \end{bmatrix}, \quad P^{(5)} = \begin{bmatrix} 0.65 & 0.25 & 0.1 \\ 0.25 & 0.65 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}.$$

For each given N, only the first $\mathsf{L} = \lfloor \mathsf{N}^{1/4} \rfloor$ of these matrices are used to construct the L intra-layer adjacency matrices. The inter-layer adjacency matrices follow the same construction as in the noiseless case described in (4.19), giving a symmetric supra-adjacency matrix $B_0 \in \mathbb{R}^{\mathsf{LN} \times \mathsf{LN}}$.

Measurement error. We include a symmetric perturbation matrix $E_0 \in \mathbb{R}^{\mathsf{LN} \times \mathsf{LN}}$, with upper triangular entries independently drawn from $\mathcal{N}(0, 1)$ and mirrored to ensure symmetry. The observed network is $B = B_0 + E_0$.

Covariates and regression. All other settings, including the generation of covariates, centrality measures, and response variables, follow the procedure in Section 4.1. Under measurement error, we use the estimated centralities $\hat{C}$ and $\hat{Z}$, computed from the noisy network B, as regressors instead of the true C and Z.

4.4 Performance with Measurement Error

Under measurement error, Table 2 shows that $\hat{\beta}$ remains consistent under Assumption 5, while $\hat{\beta}_C$ exhibits persistent bias when $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$, even as

N increases. This remarks on the limitations of C-MNetR in handling measurement error, in contrast with the improved performance of CC-MNetR. On the other hand, Table 3 confirms that $\tilde{\beta}$ from CC-MNetR is consistent and shows more stable asymptotic behavior. Overall, CC-MNetR provides a more reliable approach for multilayer network analysis with community structure. Boxplots and QQ plots for the noisy case are shown in Supplement Section S2.1.

N	$\hat{\beta}_{X_1}$	$\hat{\beta}_{X_2}$	$\hat{\beta}_{C_1}$	$\hat{\beta}_{C_2}$	$\hat{\beta}_{C_3}$	$\hat{\beta}_{C_4}$	$\hat{\beta}_{C_5}$	$\frac{a_{\mathsf{N}}\sqrt{\mathsf{L}}}{\delta}$
Truth	1	2	1	2	3	4	5	-
100	0.9997 (0.1208)	2.0010 (0.1108)	1.5179 (0.6050)	1.9603 (0.6183)	2.5257 (0.5438)	-	-	0.4368
200	1.0002 (0.0738)	1.0004 (0.0722)	1.5425 (0.5832)	1.9754 (0.5512)	2.4974 (0.5201)	-	-	0.2953
500	0.9999 (0.0461)	1.9994 (0.0484)	1.8656 (0.6659)	2.2195 (0.6362)	2.8573 (0.5698)	3.0565 (0.7009)	-	0.1838
1000	1.0003 (0.0329)	2.0016 (0.0327)	2.2347 (0.7718)	2.3905 (0.6648)	3.2160 (0.5963)	3.3280 (0.7237)	3.8200 (0.7455)	0.1242

Table 2: Mean and standard deviation of C-MNetR parameter estimates at different network sizes with measurement error. ($\sigma_b = 1$.)

5. Applications to the World Input-Output Database

We now apply CC-MNetR to the World Input-Output Database (WIOD), which covers 28 EU countries and 15 other major countries or regions from 2000 to 2014 (Timmer et al., 2015). To construct the multilayer network, each node represents an individual industry sector, and each layer corre-

N	$\tilde{\beta}_{X_1}$	$\tilde{\beta}_{X_2}$	$\tilde{\beta}_Z$	$\frac{a_{\mathsf{N}}\sqrt{\mathsf{L}}}{\delta}$
Truth	1	2	2	-
100	1.0008 (0.1088)	2.0021 (0.1021)	2.0088 (0.1037)	0.4368
200	1.0006 (0.0712)	2.0017 (0.0679)	2.0056 (0.0714)	0.2953
500	1.0000 (0.0449)	1.9992 (0.0469)	2.0020 (0.0430)	0.1838
1000	1.0004 (0.0321)	2.0017 (0.0321)	2.0004 (0.0309)	0.1242

Table 3: Mean and standard deviation of C-MNetR parameter estimates at different network sizes with measurement error. ($\sigma_b = 1.$)

sponds to the input-output data for a specific country, resulting in $\mathsf{N} = 56$ and $\mathsf{L} = 43$. Detailed information for each sector is listed in Section S3.3 of the supplement.

Each year's input-output table forms a supra-adjacency matrix B_a. We analyze the symmetrized matrix $B = B_a + B_a^\top$, whose edge weights represent scaled intermediate input-output flows between country-sector nodes. These weights are normalized to lie within the interval $[0, 2]$, making them unitless but interpretable as relative intensities of economic linkages. We set $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$, and the 56 industries are naturally classified into 20 communities based on the International Standard Industrial Classification Revision 4 (ISIC Rev.4). The community information is provided in Section S3.3.

For our analysis, we focus on the input-output tables from 2007 and 2014.

To contextualize the input-output intensities, we see the normalized supra-adjacency matrix shows high edge density. The proportion of nonzero entries is 86.71% overall and 87.20% per-layer in 2014, with similar levels in 2007, indicating extensive intermediate flows across sectors.

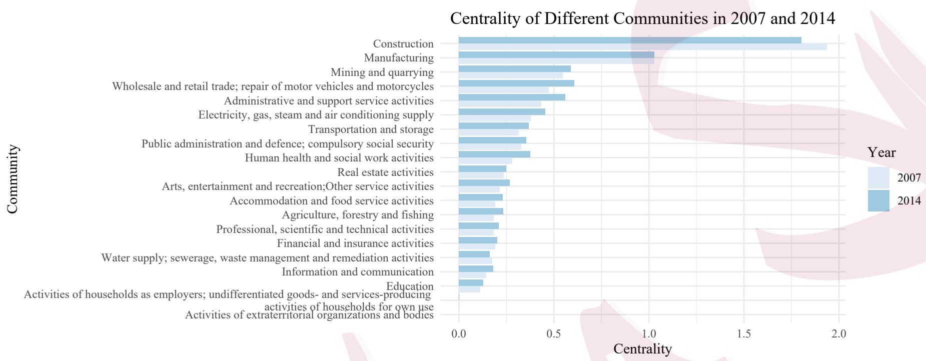

Figure 1: Changes in the centrality of different communities corresponding to industries from 2007 to 2014.

Figure 1 presents community-based centrality measures Z for 2007 and 2014. Communities representing *Construction*, *Manufacturing*, and *Mining and Quarrying* show the highest centrality, consistent with their key roles in GDP and employment (Lean, 2001; Szirmai, 2013; Fugiel et al., 2017). However, communities such as *Extraterritorial Organizations* and *Household Activities* exhibit the lowest centrality scores.

For our analysis, we selected the years 2007 and 2014 to investigate any significant changes before and after the global financial crisis. The centrality score of the *Construction* sector declined from 2007 to 2014, potentially reflecting the impact of the real estate market crash during the financial crisis. Conversely, sectors such as *Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles*, *Human Health and Social Work Activities*, and *Administrative and Support Service Activities* experienced a substantial increase in their centrality scores. This suggests an economic recovery and a shift in the economic structure during the post-crisis period, with increased prominence of sectors like healthcare and services.

We compare our method to a baseline that aggregates intra-layer networks and computes node-level eigenvector centrality, then averages scores within communities. As shown in Section S3.1 of the supplement, our approach, by using cross-country links, uncovers structural patterns missed by the baseline using only domestic (intra-layer) data.

After analyzing the changes in the centrality scores Z, we now apply the CC-MNetR model to the WIOD data using the Socio-Economic Accounts (SEA) dataset for covariates X and response y. The SEA provides industry-level data for each country on various macro-characteristics, including employment, capital stocks, gross output, and value-added at current and constant prices, measured in millions of local currency. The industry classification aligns with the world input-output tables. Since the SEA data is organized at the sector level for each country, we calculate the average sector data across different countries before conducting the regression analysis.

The description of the 10 variables in the SEA dataset is summarized in Supplementary Section S3.2. We choose *Gross Output* (GO) as the response variable y, while the remaining 9 variables are considered potential covariates X. However, since *Intermediate Input* (II) is part of the total flow matrix B and depends on the community-based centrality score Z, we exclude II from our analysis.

A significant issue in model specification is multicollinearity among covariates. Due to strong correlations among EMP (Number of Persons Engaged), LAB (Labor Compensation), EMPE (Number of Employees), and H_EMPE (Total Hours Worked), we retain only EMP to mitigate redundancy. To avoid collinearity with the centrality measure Z, we further exclude Intermediate Input (II), which is partially derived from the same input-output matrix B. This yields a final model using three covariates: Value Added (VA), Number of Persons Engaged (EMP), and Nominal Capital Stock (K). All variables are standardized, and detailed discussions on covariate selection and multicollinearity diagnostics are provided in Section

S3.2 of the supplement.

We compare two regression models to assess the impact of the community-based centrality score Z. The R-squared statistic of the first model $y = \beta X + \varepsilon$ is 0.8152, indicating that 81.52% of the variation in GO (Gross Output) is explained by VA, EMP, and K. Including Z in the CC-MNetR model increases the R-squared statistic to 0.87, showing the importance of Z. Additionally, the F-test results in Table 4 of Supplementary Section S3.4 show that Z is associated with a high F-statistic and a small p-value ($p < 0.001$) for both the years 2007 and 2014, reinforcing its statistical significance in explaining GO.

From Table 4 in S3.4, the estimated regression models for the years 2007 and 2014 are:

$$\mathrm{GO}_{2007} = 0.70Z + 0.89\mathrm{VA} - 0.06\mathrm{EMP} - 0.005\mathrm{K} - 0.38,$$

$$\mathrm{GO}_{2014} = 0.61Z + 0.95\mathrm{VA} - 0.14\mathrm{EMP} - 0.02\mathrm{K} - 0.34.$$

For the year 2007, the coefficient of VA is 0.89, indicating that a unit increase in VA raises GO by 0.89 units. This highlights VA's importance as a key driver of production efficiency and economic growth, as it represents gross output minus intermediate inputs. The negative coefficient for EMP (-0.06) suggests diminishing marginal productivity, where additional labor inputs beyond an optimal point contribute less to GO due to overcrowding or insufficient complementary capital. The positive coefficient for Z (0.70) indicates that sectors with higher centrality scores produce more output, reflecting the importance of network connections in economic activity.

Comparing the models for the years 2007 and 2014, we find that the positive coefficient for Z suggests that as the centrality score of a node increases, so does the Gross Output (y). This aligns with the understanding that sectors with extensive connections to other sectors (i.e., those that are more central in the network) tend to produce more output. Additionally, we observe a decrease in the estimated coefficient for EMP and an increase in its statistical significance. This suggests that the impact of EMP on GO has strengthened over time.

6. Concluding remarks

We propose a multilayer network regression framework using community-based centrality scores as predictors. Theoretical and numerical results show that CC-MNetR effectively captures community influence on network behavior.

Real data analysis (Section 5) aligns with established economic principles. In particular, the centrality measure Z shows a strong, positive association with sectoral flows. Furthermore, stable estimated values for

Z in both years suggest that the significant impact of network centrality measures on economic flows remains unchanged over time.

Supplementary Materials

Section S1 provides technical proofs of main theorems, analyses under unknown community structure, and discussion of key assumptions. Section S2 presents additional simulation results, comparisons with alternative models, and sensitivity analyses. Section S3 collects further details on the real-data application using WIOD, including variable definitions, estimation results, and comparisons of centrality measures.

Acknowledgements

T. Wang gratefully acknowledges Science and Technology Commission of Shanghai Municipality Grant 23JC1400700 and National Natural Science Foundation of China Grant 12301660.

Both Z. Han and T. Wang also thank Shanghai Institute for Mathematics and Interdisciplinary Sciences (SIMIS) for their financial support. This research was partly funded by SIMIS under grant number SIMIS-ID-2024-WE. They are grateful for the resources and facilities provided by SIMIS, which were essential for the completion of this work.

References

Allen, F., J. Cai, X. Gu, J. Qian, L. Zhao, and W. Zhu (2019). Ownership networks and firm growth: What do forty million companies tell us about the chinese economy? *Banking & Insurance eJournal*.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2012). The diffusion of microfinance. *Science 341*.

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

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

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology 92*(5), 1170–1182.

Cai, J., D. Yang, W. Zhu, H. Shen, and L. Zhao (2021). Network regression and supervised centrality estimation. *arXiv preprint arXiv:2111.12921*.

Cai, Y. (2022). Linear regression with centrality measures. *arXiv preprint arXiv:2210.10024*.

Cruz, C., J. Labonne, and P. Querubin (2017). Politician family networks and electoral outcomes: Evidence from the philippines. *American Economic Review 107*(10), 3006–3037.

Davis, C. and W. M. Kahan (1970). The rotation of eigenvectors by a perturbation. iii. *SIAM*

Journal on Numerical Analysis 7(1), 1–46.

De Domenico, M., A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas (2015). Ranking in interconnected multilayer networks reveals versatile nodes. *Nature Communications 6*(1), 6868.

Domenico, M. D., A. Solé-Ribalta, E. Cozzo, M. Kivela, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas (2013). Mathematical formulation of multilayer networks. *Physical Review X 3*, 041022.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports 486*(3-5), 75–174.

Fugiel, A., D. Burchart-Korol, K. Czaplicka-Kolarz, and A. Smoliński (2017). Environmental impact and damage categories caused by air pollution emissions from mining and quarrying sectors of european countries. *Journal of Cleaner Production 143*, 159–168.

Han, Q., K. Xu, and E. Airoldi (2015). Consistent estimation of dynamic and multi-layer block models. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 1511–1520. PMLR.

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

Keener, J. P. (1993). The perron–frobenius theorem and the ranking of football teams. *SIAM Review 35*(1), 80–93.

Le, C. M. and T. Li (2022). Linear regression and its inference on noisy network-linked data.

Journal of the Royal Statistical Society Series B: Statistical Methodology 84(5), 1851–1885.

Lean, C. S. (2001). Empirical tests to discern linkages between construction and other economic sectors in singapore. *Construction Management & Economics 19*(4), 355–363.

Newman, M. E. and M. Girvan (2004). Finding and evaluating community structure in networks. *Physical Review E 69*(2), 026113.

Palla, G., I. Derényi, I. Farkas, and T. Vicsek (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature 435*(7043), 814–818.

Pothen, A. (1997). Graph partitioning algorithms with applications to scientific computing. In D. E. Keyes, A. Sameh, and V. Venkatakrishnan (Eds.), *Parallel Numerical Algorithms*, Dordrecht, pp. 323–368. Springer Netherlands.

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

Rodrigues, F. A. (2019). Network centrality: An introduction. In E. E. N. Macau (Ed.), *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems*, Cham, pp. 177–196. Springer International Publishing.

Shen, C., Q. Wang, and C. E. Priebe (2022). One-hot graph encoder embedding. *IEEE Transactions on Pattern Analysis and Machine Intelligence 45*(6), 7933–7938.

Szirmai, A. (2013). Manufacturing and economic development. *Pathways to industrialization in the twenty-first century: New challenges and emerging paradigms*, 53–75.

Timmer, M. P., E. Dietzenbacher, B. Los, R. Stehrer, and G. J. De Vries (2015). An illustrated user guide to the world input–output database: the case of global automotive production. *Review of International Economics 23*(3), 575–605.

Tudisco, F., F. Arrigo, and A. Gautier (2018). Node and layer eigenvector centralities for multiplex networks. *SIAM Journal on Applied Mathematics 78*(2), 853–876.

van Handel, R. (2017). Structured random matrices. *Convexity and Concentration*, 107–156.

Zhuoye Han, Fudan University

E-mail: zhuoyehan21@m.fudan.edu.cn

Tiandong Wang, Fudan University and Shanghai Academy of Artificial Intelligence for Science

E-mail: td_wang@fudan.edu.cn

Zhiliang Ying, Columbia University

E-mail: zy9@columbia.edu