Physica A 581 (2021) 126196

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Regional and sectoral structures of the Chinese economy: A network perspective from multi-regional input–output tables

Tao Wang [a], Shiying Xiao [a], Jun Yan [b], Panpan Zhang [c,*]

[a] School of Statistics, Shanxi University of Finance and Economics, Taiyuan, 030006, Shanxi, China
[b] Department of Statistics, University of Connecticut, Storrs, 06269, CT, USA
[c] Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, 19104, PA, USA

ARTICLE INFO

Article history:
Received 23 February 2021
Received in revised form 31 May 2021
Available online 22 June 2021

MSC:
91D03
05C82

Keywords:
Backbone structure
Dynamical analysis
Input–output table
Network analysis
Power law
Region-sector economy

ABSTRACT

A multi-regional input–output table (MRIOT) containing the transactions among the region-sectors in an economy defines a weighted and directed network. Using network analysis tools, we analyze the regional and sectoral structure of the Chinese economy with the province-sector MRIOTs of China in 2007 and 2012. Global analyses are done with network topology measures. Growth-driving province-sector clusters are identified with community detection methods. Influential province-sectors are ranked by weighted PageRank scores. The results have revealed a few interesting and telling insights. The level of inter-province activities increased with the rapid growth of the national economy, but not as fast as that of intra-province economic activities. Regional community structures were deeply associated with geographical factors. The community heterogeneity across the regions was high and the regional fragmentation increased during the study period. Quantified metrics assessing the relative importance of the province-sectors in the national economy echo the national and regional economic development policies to a certain extent.

1. Introduction

The rapid growth of the Chinese economy over the last three decades has drastically elevated its global importance. The annual growth rate of gross domestic product (GDP) during 1990–2010 was 10.4% [1]. The growth in China was a driving force for the recovery of the world from the financial crisis in 2008 [2]. As the Chinese economy matured, the GDP growth slowed down to 6.74% over 2015–2019, but it was still much higher than the average of global economy, 2.82%, during the same period [3]. In 2019, China contributed 16.34% to the global GDP, second only to the United States and almost tripling the contribution by Japan which ranked third [4]. China has become an integrated part of the global economy. As a top trader, China accounted for 10.14% of the global imports [5] and 10.61% of the global exports in 2019 [6]. Behind the growth of the Chinese economy, there have been dramatic structural changes such as urbanization and industrialization [7,8]. The regional and sectoral structures of the Chinese economy are heavily affected by internal government policies such as the Great Western Development Strategy [9] or external factors such as the World Trade Organization accession in 2001 [10] and the 2008 global financial crisis [11].

Given its size and impact, the Chinese economy is central to important regional and sectoral structure issues in economic development theory and practice. The disparities in sectoral structure and economic growth at the province

* Corresponding author.
E-mail address: panpan.zhang@pennmedicine.upenn.edu (P. Zhang).

https://doi.org/10.1016/j.physa.2021.126196

level in China are high and have been increasing [e.g.,12–14]. Liberalized and globalized industries are mostly aggregated in the coastal regions while low technology, resource-based, and protected industries are widely dispersed in the inland regions [e.g.,15]. Emerging industries are more likely to enter the regions that are globalized, economically liberalized and fiscally healthy [16]. On one hand, the distribution of value-added across regions has been flattened due to the expansion of inter-regional trade [17]. On the other hand, growing inter-regional competition and local protection have jointly led to various inter-regional trade barriers and severe regional fragmentation [18,19]. Understanding the regional and sectoral structures of Chinese economy is critical for economic development and resource efficiency not only in China but also the entire world.

Multi-regional input–output tables (MRIOTs) are the most prevalent tool for studying the inter-dependencies among the sectors from different regions in an economy. An MRIOT records the transactions among the sectors within multiple regions [e.g.,20,21]. A few MRIOTs at the global level have been available with bilateral trade information for a large number of countries annually for several decades [22]. Within China, however, MRIOTs are not as available. Two versions of MRIOTs have been constructed, both based on the raw inflow and outflow tables released by each province every One was constructed for 2002 and 2007 [23] based on a dual-constraint gravity model with maximum entropy estimation [24,25]. The other was constructed for 2007 and 2012 [26,27] with a newer gravity model [28,29]. Although both versions were estimated based on the same published inter-regional flows, the raw data of 2002 or earlier were less accurate due to the higher aggregation level. The MRIOTs of China and their extensions have been used in the analyses of the impact of government infrastructural plans [30], provincial and sector-level material footprints [31,32], and cross-country sectoral price comparison [33], among others.

An MRIOT inherently defines a weighted, directed network which facilitates analyses with statistical methods for networks. The world input–output tables (WIOTs) [34] are MRIOTs at the global level. Traditional tools for input–output table analysis include multipliers, linkages, and structural paths [35,36], which measure the impact from each region-sector in the table. Network analytic tools of MRIOTs construct a network using the intermediate use matrix across all the region-sectors (see Table 1), which enables not only measures at the region-sector level such as centrality, but also natural investigation on local clustering, community detection, and backbone extraction as well as global network features such as assortativity and clustering coefficient [37,38]. Although the network tools are of limited use for the final use section in a MRIOT due to lack of details on data aggregated by region-sectors, the network perspective based on the intermediate use section is of great value in structural and regional economic analyses [39–41]. For the Chinese MRIOTs, in part due to their limited availability, few network analysis work has been done to explore the regional and sectoral structures of the Chinese economy. One exception is Sun et al. [42], who used the old version of MRIOTs for 2007 [23] to investigate the key sectors in single provincial input–output networks and inter-provincial input–output networks in China.

Our contributions are two-fold. First, to the best of our knowledge, this is the first comprehensive (from global, regional and local perspectives) network analysis of the most recent MRIOTs of China to study regional and sectoral structure of the Chinese economy. Through the analysis, we have evidently observed a clear pattern of increased regional fragmentation from 2007 to 2012. Some of the research outcomes have not been reported in the literature, and may not be straightforwardly uncovered via traditional input–output analysis tools. Our second contribution is the application of several novel network measures specifically developed for weighted and directed networks to analyze the Chinese MRIOTs. As mentioned, the networks associated with MRIOTs are both weighted and directed. If (edge) weight and/or direction are disregarded in the computations, the features of the networks such as assortativity and centrality are not precisely characterized. The adopted network measures help correct the misleading results and inaccurate inference driven from the classical unweighted versions.

The rest of the manuscript is organized as follows. In Section 2, we give a brief introduction of the compiled MRIOTs in China, and demonstrate the network analysis setup. The specific network analysis methods are presented in Section 3, followed by the applications to the MRIOTs in Section 4. Finally, we address some concluding remarks and follow-up discussions in Section 5.

2. MRIOTs of China

The MRIOTs of China that we used were those constructed with the gravity method, which are available for the year of 2007 [26], 2010 [43], and 2012 [27]. The databases were jointly developed by the Institute of Geographic Sciences and the Natural Resources Research of the Chinese Academy of Sciences, and the National Bureau of Statistics of China. The entries of inter-province-sector economic transactions were obtained by applying the gravity model [28,29] to the input–output tables reported by all the participating provinces. Table 1 shows the fundamental structure of MRIOT. We focus on the 2007 and 2012 tables in the present study because the compilation of the 2010 table was based on the 2007 table in addition to the input–output tables of the 17 provinces rather than direct data collection and investigation [44], which might cause measurement errors and bias. An MRIOT consists of four parts: (I) intermediate flow matrix, (II) final use, (III) imports, and (IV) value added. The intermediate flow matrix records the economic exchanges among the sectors from different provinces, reflecting their intricate economic relations (e.g., supply and demand) as well as their interdependence and mutual constraints.

The data in the MRIOTs were pre-processed to prepare for the analyses. The 2007 table covered 30 provincial units with each containing 30 sectors. The 2012 table, however, covered 31 provincial units due to the debut of Tibet and 42

Table 1
Fundamental structure of an MRIOT.

			Intermediate use							Final use	Total output
			region 01			...	region 30				
			sector 01	...	sector 30	...	sector 01	...	sector 30		
	region 01	sector 01									
		⋮									
		sector 30									
Intermediate input	⋮	⋮				I				II	
	region 30	sector 01									
		⋮									
		sector 30									
Imports						III					
Value added						IV					
Total input											

Table 2
Description of the sectors in the MRIOTs.

Code	Sector	Code	Sector
01	Agriculture, forestry, animal husbandry and fishery	16	General and specialist machinery
02	Coal mining and processing	17	Transport equipment
03	Petroleum and gas extracting	18	Electrical equipment
04	Metals mining/processing	19	Electronic equipment
05	Nonmetal mining/processing	20	Instrument and meter
06	Food processing and tobaccos	21	Other manufacturing
07	Textiles	22	Electricity and heat production and supply
08	Clothing, leather, fur, etc.	23	Gas and water production and supply
09	Wood processing and furnishing	24	Construction
10	Paper making, printing, stationery, etc.	25	Transport and storage
11	Petroleum refining, coking, etc.	26	Wholesale and retail
12	Chemical industry	27	Hotel and restaurant
13	Nonmetal products	28	Leasing and commercial services
14	Metallurgy	29	Scientific research
15	Metal products	30	Other services

sectors that were further divided from the 30 sectors in the 2007 table. For the purpose of comparison over time, we only included the 30 provinces that appeared in both tables and aggregated the 42 sectors in 2012 to the 30 sectors in 2007. Table 2 lists the codes with detailed descriptions of the 30 sectors. The monetary units for both tables were set to 10,000 Chinese Yuan (CNY). To adjust for inflation, we converted the entries in the 2012 table to 2007 CNY price by using the GDP price deflator [45].

We constructed the multi-regional input–output networks (MRIONs) based on the MRIOTs. In an MRION, each vertex represents a sector within a province; each directed edge represents an existing transaction from the source province-sector to the target one, with weight representing the multiplier of 10,000 CNY in transaction volume. Therefore, the MRIONs are weighted and directed. The number of vertices in each MRION is 900. The link densities are respectively 0.7685 in 2007 and 0.9212 in 2012, suggesting that the vertices in the MRIONs are densely connected. The top two panels of Fig. 1 show the chord plots of MRIONs that are aggregated according to sectors with self-loops removed. Each of the outer arcs with a distinct color represents a sector, with arc length representing the sum of the inflows and outflows. A chord from one arc to another represents the transaction from the corresponding sector to the other. Its width is proportional to the volume of the transaction, while its color remains the same as the color of the source sector. For both years (2007 and 2012), the main suppliers are "metallurgy" (14), "chemical industry" (12), "other services" (30), and "agriculture, forestry, animal husbandry and fishery" (01). They supply a large portion of the intermediate products or services that are needed by many other sectors. The most notable receivers are "construction" (24) and "other services" (30). These two sectors may have strong pulling effects on the whole economy. Especially for "construction" (24), the proportion of inflows in its inter-sectoral transaction exceeds 90%. One notable change from 2007 to 2012 is the share of the transactions associated with sector "scientific research" (29), which is quadrupled from 0.24% to 0.97%.

In addition, we provide the chord plots of the MRIONs aggregated by provinces (with self-loops removed as well) for 2007 and 2012, shown in the bottom two panels in Fig. 1. The arcs and chords are defined analogously as the top ones. The main suppliers in 2007 were Hebei (03), Guangdong (19) and Jiangsu (10), but Hebei (03) was replaced with

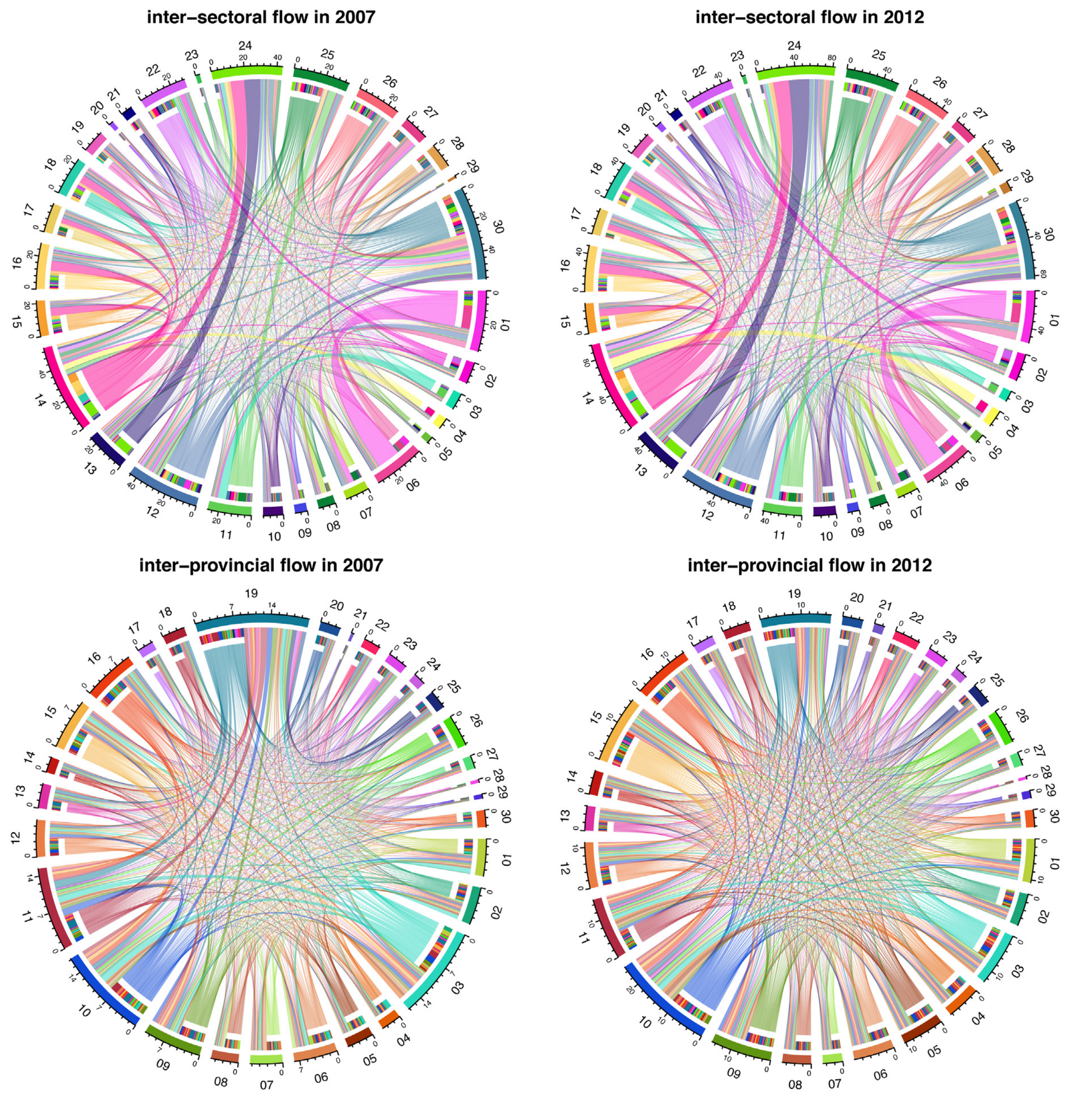

Fig. 1. Intermediate flows (across multiple regions) in the MRIONs from 2007 and 2012. The top two panels are for sectors (aggregated with respect to provinces), where the codes are referred to Table 2. The bottom two panels are for provinces (aggregated with respect to sectors), where the codes are referred to Appendix A. The bandwidth of chord (connecting the arcs) is proportional to the size of economic flow. Long arcs indicate large outputs. The unit of economic flow is 100 billion CNY.

Shandong (15) in 2012, indicating that the production capacity of Shandong (15) became strong (from 2007 to 2012). For both years, the most notable receivers were Jiangsu (10), Zhejiang (11) and Guangdong (19). These three provinces have promoted a large number of inter-provincial trade exchanges with the others across the nation. In fact, the majority of the provinces presenting high inter-provincial trade amount were from the coastal region. In spite of the substantial drops in the proportions of inter-provincial trade in Guangdong (19) and Zhejiang (11) from 2007 to 2012, they were still top 5 over the nation and remained the driving forces contributing to the multi-regional economy in China. More quantitative assessment calls for detailed network analyses.

3. Methods

Our approach to investigating the MRIOTs of China are network-based analytics. Let $G(V, E)$ denote a directed network with vertex set V and edge set E. By convention, each vertex $i \in V$ represents a data point. Given a pair of vertices $i, j \in V$, if there is a directed edge from i to j, then we have $e_{ij} \in E$. Vice versa, the existence of an edge $e_{ij} \in E$ suggests a (directed) link from i to j. One of the most popular ways of displaying a network structure is adjacency matrix. For a network with $n = |V|$ vertices (i.e., the cardinality of set V is the number of elements in V), its adjacency matrix is denoted by $\boldsymbol{A} := (a_{ij})_{n\times n}$ with $a_{ij} = 1$ if $e_{ij} \in E$ and $a_{ij} = 0$ otherwise. For a weighted and directed network, the weighted counterpart is denoted by $\boldsymbol{W} := (w_{ij})_{n\times n}$, where w_{ij} represents the weight of edge e_{ij}. The weighted adjacency matrix $\boldsymbol{W}$ is equivalent to $\boldsymbol{A}$ if $w_{ij} = 1$ for all $e_{ij} \in E$.

3.1. Degree and strength distributions

In a directed network, the degree of a vertex i (denoted d_i) is comprised of in-degree (denoted by $d_i^{(\text{in})}$) and out-degree (denoted by $d_i^{(\text{out})}$), which are, respectively, the number of edges pointing into and emanating out of vertex i. To account for edge weight, we define the in-strength and out-strength of vertex i as $s_i^{(\text{in})} := \sum_{j\in V} w_{ji}$ and $s_i^{(\text{out})} := \sum_{j\in V} w_{ij}$, respectively. The strength of vertex i, denoted s_i, is the sum of its in-strength and out-strength. In traditional network analyses, degrees and strengths are used to show the importance of the vertices in a network [46].

The degree distribution is the probability distribution $\pi(\cdot)$ of the vertex degrees over the entire network; that is, $\pi(k)$ is the probability of a vertex with degree $k \in \{0, 1, 2, \ldots\}$. The degree distribution plays an important role in theoretical and applied network analyses. In a completely random network [47], the degree distribution is Poisson, whereas the tail of the degree distribution of a scale-free network [48] follows a power law. Pennock et al. [49] pointed out that most real networks fall between these two extreme classes. It is evident that the degree distributions of economic networks are likely to exhibit power-law patterns [50,51]. The goodness-of-fit of power-law tails can be tested via the Kolmogorov–Smirnov statistic [52] with p-values obtained from bootstrapping.

In a weighted network, the strength distribution, which is based on the vertex strengths, usually better captures the network structure than the degree distribution. While the degree distribution is always discrete, the strength distribution can be either discrete or continuous, depending on weight characteristic. As the MRIONs are weighted and directed, we conducted analogous analyses on the strength, in-strength and out-strength distributions and made comparisons with their degree counterparts.

3.2. Assortativity

Assortativity (or assortative mixing) refers to the tendency that the vertices in a network are connected according to a pair of (vertex-specific) features [53]. It is a measure of homophily among the vertices based on the two given features. A commonly used assortativity measures is the degree–degree correlation [53,54], which is analogous to Pearson correlation coefficient. Its value is between -1 and 1, with a positive (negative) value indicating that high-degree vertices are more likely to be connected with high-degree (low-degree) vertices. This measure is amenable to directed networks [55,56]. See Noldus and Van Mieghem [57] for a comprehensive survey.

Since the MRIONs are not only directed but also weighted, we adopted a class of assortativity measures proposed by Yuan et al. [58] to incorporate edge weight and direction. Let $(\alpha, \beta) \in \{\text{in}, \text{out}\}$ index strength type. The assortativity based on α- towards β-type strength is

$$\rho_{\alpha,\beta}(G) = \frac{\sum_{i,j\in V} w_{ij}\left[\left(s_i^{(\alpha)} - \bar{s}_{\text{sou}}^{(\alpha)}\right)\left(s_j^{(\beta)} - \bar{s}_{\text{tar}}^{(\beta)}\right)\right]}{W\sigma_{\text{sou}}^{(\alpha)}\sigma_{\text{tar}}^{(\beta)}}, \tag{1}$$

where $W := \sum_{i,j\in V} w_{ij}$ is the total weight, $s_i^{(\alpha)}$ is the α-type strength of source vertex i, $s_j^{(\beta)}$ is the β-type strength of target vertex j,

$$\bar{s}_{\text{sou}}^{(\alpha)} = \frac{\sum_{i,j\in V} w_{ij}s_i^{(\alpha)}}{W} \quad \text{and} \quad \bar{s}_{\text{tar}}^{(\beta)} = \frac{\sum_{i,j\in V} w_{ij}s_j^{(\beta)}}{W}$$

are, respectively, the weighted mean of the α-type strength of the source vertices and β-type strength of the target vertices, and

$$\sigma_{\text{sou}}^{(\alpha)} = \sqrt{\frac{\sum_{i,k\in V} w_{ik}\left(s_i^{(\alpha)} - \bar{s}_{\text{sou}}^{(\alpha)}\right)^2}{W}} \quad \text{and} \quad \sigma_{\text{tar}}^{(\beta)} = \sqrt{\frac{\sum_{k,j\in V} w_{kj}\left(s_j^{(\beta)} - \bar{s}_{\text{tar}}^{(\beta)}\right)^2}{W}}$$

are the associated weighted standard deviations. A positive (negative) $\rho_{\alpha,\beta}(G)$ suggests assortative-mixing (disassortative-mixing), and zero assortativity indicates no obvious pattern of assortative- or disassortative-mixing.

For a network like MRION, the weighted adjacency matrix can be decomposed into one comprised of diagonal blocks only and the other comprised of off-diagonal blocks. The former contains the information of economic transactions within each province (called intra-province), while the latter records the exchanges across multiple provinces (called inter-province). The proposed assortativity measure can be applied to the decomposed adjacency matrices to investigate the correlation structures at the intra- and inter-province levels.

3.3. Clustering coefficient

Clustering coefficient is a measure quantifying the tendency that the vertices in a network are clustered together, usually characterized by a simple measure like connection density [59]. In the literature, clustering coefficient is also known as transitivity coefficient [60]. Classical clustering coefficient was proposed for undirected and unweighted networks [61], and later were extended for analyzing weighted and directed networks [62–67].

In the present study, we adopted the weighted and directed clustering coefficients developed by Clemente and Grassi [67]. The local clustering coefficient of vertex i in an unweighted and undirected network $G(V, E)$ is the ratio of the number of links connecting the neighbors of i (i.e., $\{j \in V : e_{ij} \in E\}$) to the maximum possible value. When edges have weights, the weighted adjacency matrix $\boldsymbol{W}$ plays an important role. Self-loops are removed prior to the computation since they do not practically contribute to the network clustering property. The local clustering coefficient of i can be precisely expressed via matrix notations:

$$C_i^{\text{tot}} = \frac{\left[\left(\boldsymbol{W}+\boldsymbol{W}^\top\right)\left(\boldsymbol{A}+\boldsymbol{A}^\top\right)^2\right]_{ii}}{2\left[s_i(d_i-1)-(\boldsymbol{AW}+\boldsymbol{WA})_{ii}\right]}, \tag{2}$$

where $\boldsymbol{A}^\top$ is the transpose of $\boldsymbol{A}$, and we use $(\boldsymbol{A})_{ii}$ to denote the ith element on the diagonal of square matrix $\boldsymbol{A}$.

The superscript distinguishes it from the four kinds of distinct local clustering coefficients induced from four types of directed triangles [66,67]. Namely, they are in-, out-, mid- and cyc-clustering coefficients. When computing a specific local clustering coefficient, the denominator needs to be updated to the number of corresponding triplets. For instance, the local in-clustering coefficient of i is the number of triangles such that the neighbors (say, j and k) both link towards i alongside with an edge (in either direction) connecting j and k out of the number of triplets with both j and k generating directed edges to i (disregarding whether or not j and k are connected). All of the other local clustering coefficients are defined analogously. The local out-clustering coefficient of i is the proportion of triangles which have two edges from i pointing to j and k and an edge linking j and k in either direction. The local mid-clustering coefficient of i considers the proportion of the triangles in which i is a middleman: neighbor j (or k) either has a directed link to neighbor k (or j) or forms a directed path $j \to i \to k$ (or $k \to i \to j$). The local cyc-clustering coefficient of i only counts the triangles of which the directed edges form a cycle. See Fig. B.8 in Appendix B for graphical illustrations and the formulae therein for practical computations. Accordingly, there are five kinds of global clustering coefficients (including that does not account for edge direction, c.f. Eq. (2)) on the network base, obtained by averaging the associated local clustering coefficients over all the vertices.

Similar to assortativity, any kind of clustering coefficients introduced in this section can be applied to the decomposed adjacency matrices of MRION to uncover the clustering properties at the intra- and intro-province levels.

3.4. Community detection

Community detection aims to group the entities with similar characteristics from a network in the same community. The entities in the same community are expected to be densely linked, while those from different communities are loosely linked. There are two major classes of community detection methods, model-based [68,69] and metric-based [70,71] methods. In this study, we used a metric-based method for community detection in MRIONs. Interested readers are referred to Goldenberg et al. [72] for a comprehensive survey for community detection techniques.

Specifically, we exploited the modularity maximization algorithm proposed by Newman [73]. An objective function called modularity is defined to measure the quality of all clustering strategies, and then the one with maximal modularity is selected. The underlying principle of modularity maximization is that the number of links among the vertices within a community is significantly more than expected number of random allocation (based on the Erdös–Rényi model which is generally used as the null model), while the counterpart across different communities is significantly less. Newman's algorithm is built upon recursive bi-partitioning. For a weighted and undirected network $G(V, E)$, its modularity matrix $\boldsymbol{B} := (b_{ij})_{n\times n}$ is defined as

$$b_{ij} := w_{ij} - \frac{s_i s_j}{W},$$

where W is identical to that defined in Section 3.2. The term $s_i s_j/W$ is interpreted as the expected weight of the edges connecting i and j if all the edges are randomly placed among the vertices in the network. Let $\boldsymbol{c} := (c_i)_{i=1}^n$ denote a clustering strategy. A bi-partitioning algorithm admits two clusters, so c_i's take value 1 or -1 representing distinct membership. Conditional on $\boldsymbol{c}$, the modularity score is defined as

$$Q = \frac{1}{W}\sum_{i,j} b_{ij} I(c_i = c_j),$$

where $I(\cdot)$ is the indicator function. The expression of Q can be regarded as a reward-penalty system. Given $c_i = c_j$, the value of Q increases if $b_{ij} > 0$, but decreases if $b_{ij} < 0$. Besides, the larger b_{ij} is (given $c_i = c_j$), the more reward is granted. Subsequent bi-partitioning continues within each resulting community until no more partition in any existing community leads to an increase in modularity score.

For large networks, parsimonious algorithms [71,74,75] are needed to solve the optimization problem. We adopted the greedy algorithm developed by Clauset et al. [74] in the present study.

3.5. Centrality

The centrality of each vertex measures its relative importance in a network. Vertices with high centrality scores altogether form the main frame of the network. There are various ways of defining centrality depending on practical needs and interpretations, such as degree centrality [63], closeness and betweenness [76], eigenvector centrality [77], and PageRank [PR,78]. Measures like degree centrality are too simple to capture the vertex characteristics in complex networks. PR is extended from eigenvector centrality which undergoes the limitation of being inapplicable to directed acyclic networks. See Das et al. [79] for a concise survey of centrality measures and their applications to social network analysis.

Classical centrality measures and their variants that have been used to analyze input–output networks include Rasmussen [80], Katz [81], Laumas [82], Dietzenbacher [83], but they all undergo the limitation of making full use of information from weighted, directed networks for precise characterizations. In this article, we adopted a class of extended PR measures [84] for weighted and directed networks. Specifically, the weighted PR centrality of vertex $i \in V$ is given be

$$P_i = \gamma \sum_{j\in V} \left(\theta \frac{w_{ji}}{s_j^{(\text{out})}} + (1-\theta)\frac{a_{ji}}{d_j^{(\text{out})}} \right) P_j + \frac{(1-\gamma)\lambda_i}{\sum_{i\in V}\lambda_i}, \qquad i = 1, \ldots, n, \tag{3}$$

where $\theta \in [0, 1]$ is a tuning parameter indicating the proportion of edge weight (versus edge number) accounted in weighted PR, γ is a damping factor that prevents the algorithm from getting stuck in sinking vertices (those without outgoing edges), and λ_i is a prior measure (which can be dependent or independent of network structure) of the relative importance of vertex i. When there is no information available for γ, it takes value 0.85 as suggested by Page et al. [85]. In spite of the prior information specified by λ_i's, Eq. (3) suggests that a vertex receives a high PR score (with $\theta = 1$) if (i) it receives a large number of incoming edges from the others in the network; (ii) the weights of the incoming edges are large;

When there is no prior information about λ_i's, they can be set the same, in which case the second term on the right hand side of Eq. (3) is simplified to $(1-\gamma)/n$. A standard method to solve Eq. (3) is power iteration, but the convergence of this algorithm may be slow for large-scale networks. A remedy is to utilize the stochastic process theory and convert the problem to finding the stable distribution of an underlying Markov Chain [86]. The investigations of the crucial properties of the proposed PR measure can be found in Zhang et al. [84].

3.6. Backbone

The backbone of a network is the fundamental but essential structure of a network [38]. Non-essential links, which act like noise in a large network, can be removed without affecting the backbone. Extracting the backbone of a massive and dense network like MRION is critical, as hundreds of edges with minimal weights would overwhelm the analysis. Proper removal of non-essential edges helps succinctly characterize a complex network system, and meanwhile enhance computation speed. There have been a few promising backbone extraction methods, such as the disparity filter method [87,88], the locally adaptive network sparsification algorithm [89], and two classes of node-based filtering approaches [90]. We used the disparity filter method which has been applied to the analysis of WIOTs by Xu and Liang [38].

The rational of the disparity filter method is as follows. Consider the normalized weights of $d \in \mathbb{Z}^+$ edges of a vertex. Under the null hypothesis that the normalized weights are generated from a uniform random assignment, they can be regarded as obtained by dividing the unit interval by $(d - 1)$ randomly placed points. The lengths of the subintervals, which represent the normalized weights, have density function

$$p(x; d) = (d-1)(1-x)^{d-2}, \quad x \in (0, 1).$$

An overly large normalized weight relative to this distribution means that the corresponding edge is unlikely to be from the uniform random assignment, which supports the corresponding edge to be part of the backbone. This idea can be formulated as obtaining the p-value of each normalized weight. Define $\tilde{w}_{ij} = w_{ij}/s_i$. The p-value of $\tilde{w}_{ij}$ is

$$\delta_{ij}(\tilde{w}_{ij}; d_i) = \int_{\tilde{w}_{ij}}^{1} p(x; d_i)\mathrm{d}x.$$

The backbone with level $\delta \in (0, 1)$ is obtained by retaining only those edges whose p-values are less than δ.

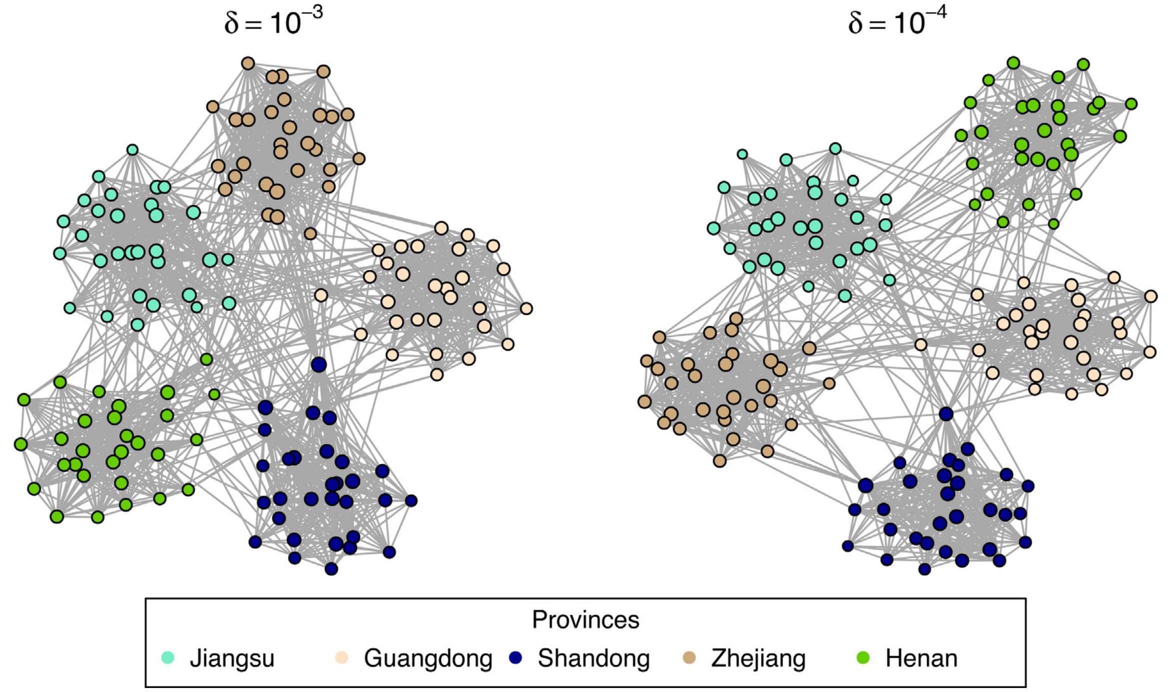

Fig. 2. Examples of the sub-MRIONs comprised of the sectors from the provinces with top 5 regional GDP in 2012. The self-loops are not presented. The significance levels are respectively 10^{-3} (left panel) and 10^{-4} (right panel).

For a directed network, normalized out- and in-strength of each edge can be defined similarly as $\tilde{w}_{ij}^{(\text{out})} = w_{ij}/s_i^{(\text{out})}$ and $\tilde{w}_{ij}^{(\text{in})} = w_{ij}/s_j^{(\text{in})}$ for all $i, j \in V$. The corresponding p-values are $\delta_{ij}(\tilde{w}_{ij}^{(\text{out})}; d_i^{(\text{out})})$ and $\delta_{ij}(\tilde{w}_{ij}^{(\text{in})}; d_j^{(\text{in})})$, respectively. For backbone with level $\delta \in (0, 1)$, edge e_{ij} is preserved if at least one of the two p-values is less than δ.

In some rare cases of $d_i^{(\text{out})} = 1$, $d_i^{(\text{in})} = 1$ or $d_i^{(\text{out})} = d_i^{(\text{in})} = 1$, special treatments may be needed, depending on the specific features of the networks as well as the practical interpretation of network heterogeneity. These rare cases do not occur in our study. In Fig. 2, we present the sub-networks of the MRION in 2012 consisting of the sectors only from top 5 regional GDP provinces with significance level $\delta \in \{10^{-3}, 10^{-4}\}$. Though the sub-networks remain dense, they relatively better reflect the basic structure of the MRION, and furthermore, suggest province-based market fragmentation as well as community structure, which are consistent with some of the results shown in Section 4.

4. Results

We apply the methods in Section 3 to the 2007 and 2012 MRIONs of China, and present the corresponding results in this section. The interpretations of the analysis results are given from both statistical and economic perspectives.

4.1. Degree and strength distributions

Fig. 3 shows the histograms for the in-, out- and total-degree distributions of the MRIONs in 2007 and 2012. These degree distributions appear to share two features. First, there is a strictly positive probability of degree zero, which corresponds to province-sectors with no edges (i.e., isolated vertices). For example, the sectors "coal mining and processing" (02), "metals mining/processing" (04) and "nonmetal mining/processing" (05) in Shanghai are singletons with neither inbound nor outbound links. Second, the nonzero degrees are close to the maximum degree and skewed to the left, more significantly reflected in out-degrees than in-degrees. This is a result of heavily connected province-sectors. From 2007 to 2012, all three degree distributions shifted to the right with more left skewness, suggesting the increase in the number of links among the province-sectors during this period in China.

More information about the magnitude of the economic transactions is provided in the strength distributions on the log scale shown in Fig. 4. The strength distributions are mixtures of a point mass at zero and a positive continuous distribution. Such distributions are often used to model zero-inflated non-negative continuous data through the mechanism of two-part models or hurdle models [91]. The mass at zero is inherited from the zero degrees from the degree distributions. The positive strengths on the log scale are skewed to the left. On the original scale, however, the positive strengths are skewed to the right. Similar to the degree distributions, the strength distributions all shifted to the right more significantly from

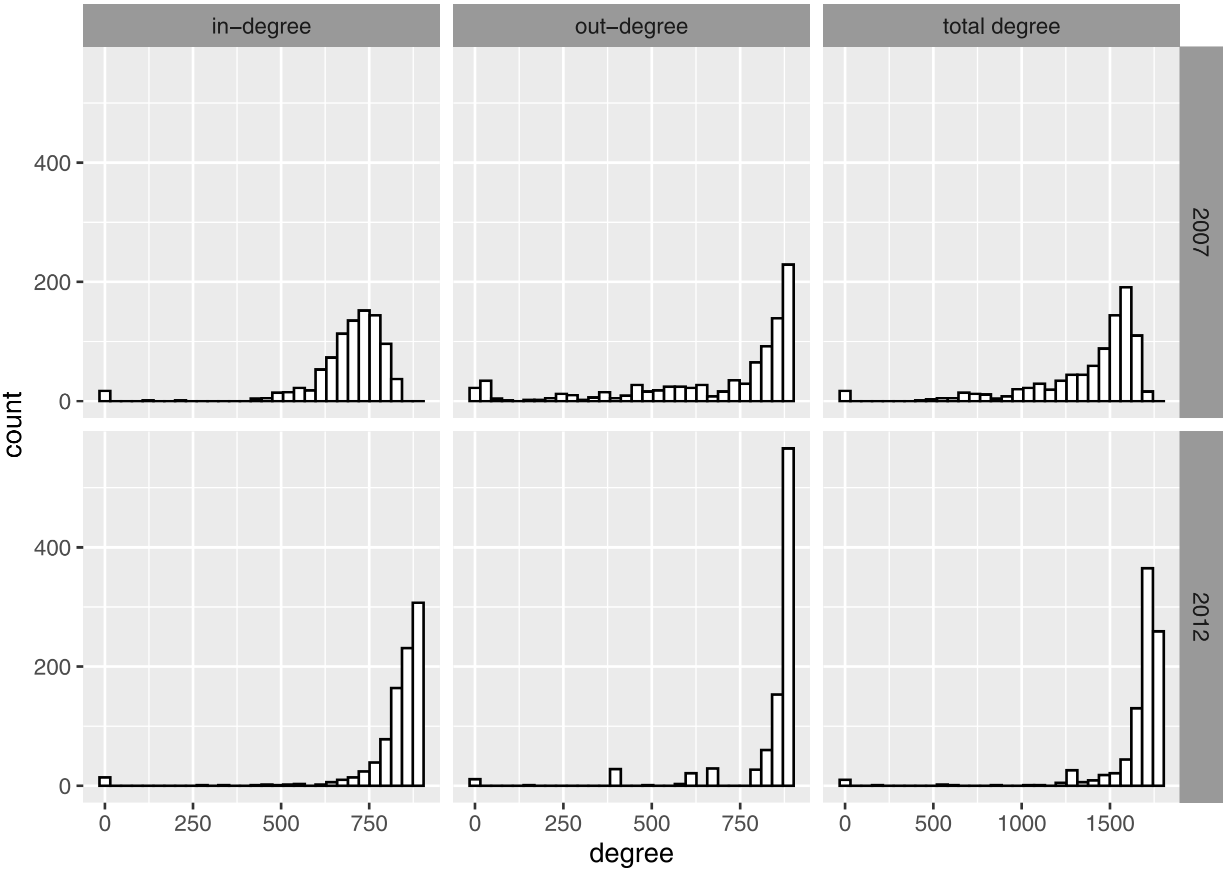

Fig. 3. In-, out- and total-degree distributions of the MRIONs in 2007 and 2012.

Table 3
Estimated parameters for power-law tails for the in-, out- and total-strength distributions of the MRIONs in 2007 and 2012 and p-values of the goodness-of-fit tests; the unit of threshold is 1 million CNY.

	2007			2012		
	In-strength	Out-strength	Total-strength	In-strength	Out-strength	Total-strength
Threshold	9.11	20.91	20.99	20.83	24.41	46.96
Exponent	2.65	3.56	2.84	3.19	3.29	3.33
p-value	0.00	0.27	0.01	0.18	0.46	0.42

2007 to 2012 with most of the quartiles more than doubled, reflecting an expansion of economic transactions among the province-sectors during this period.

Strength distribution tail is of great importance, especially in extreme value theory, as it characterizes the features of the distribution far way from the mean, indicating the relative probability of the occurrence of some "unusual" events, i.e., extensively large strength sectors. Specifically, we are interested in a particular class of heavy tail distributions—power laws, which, as mentioned in Section 3.1, have been observed in a variety of economic networks. Fig. 5 shows the empirical survival curves of the three strength distributions in 2007 and 2012 with both axis on the log scale. Empirical distributions of such shapes appear to be typical for MRIONs [e.g.,38]. The tails of the distributions show plausible linear patterns (on the log scale) which are the characteristics of power laws. To verify, we conducted a goodness-of-fit test for power law tails [52]. Table 3 summarizes the estimated thresholds and exponent parameters as well as the p-values of the power law tails beyond the estimated thresholds obtained from bootstrapping. The power law provides adequate fit to all three strength distributions in 2012 and the out-strength distribution in 2007; it is rejected at significance level 0.01 for the in-strength and total-strength distribution in 2007. This is consistent with the lower panels of Fig. 5, which show the empirical conditional survival curves beyond the estimated thresholds. For the out-strength, the smaller exponent parameter estimate in 2012 than in 2007 indicates heavier tails in the magnitude of the extremely large transactions in 2012 than in 2007.

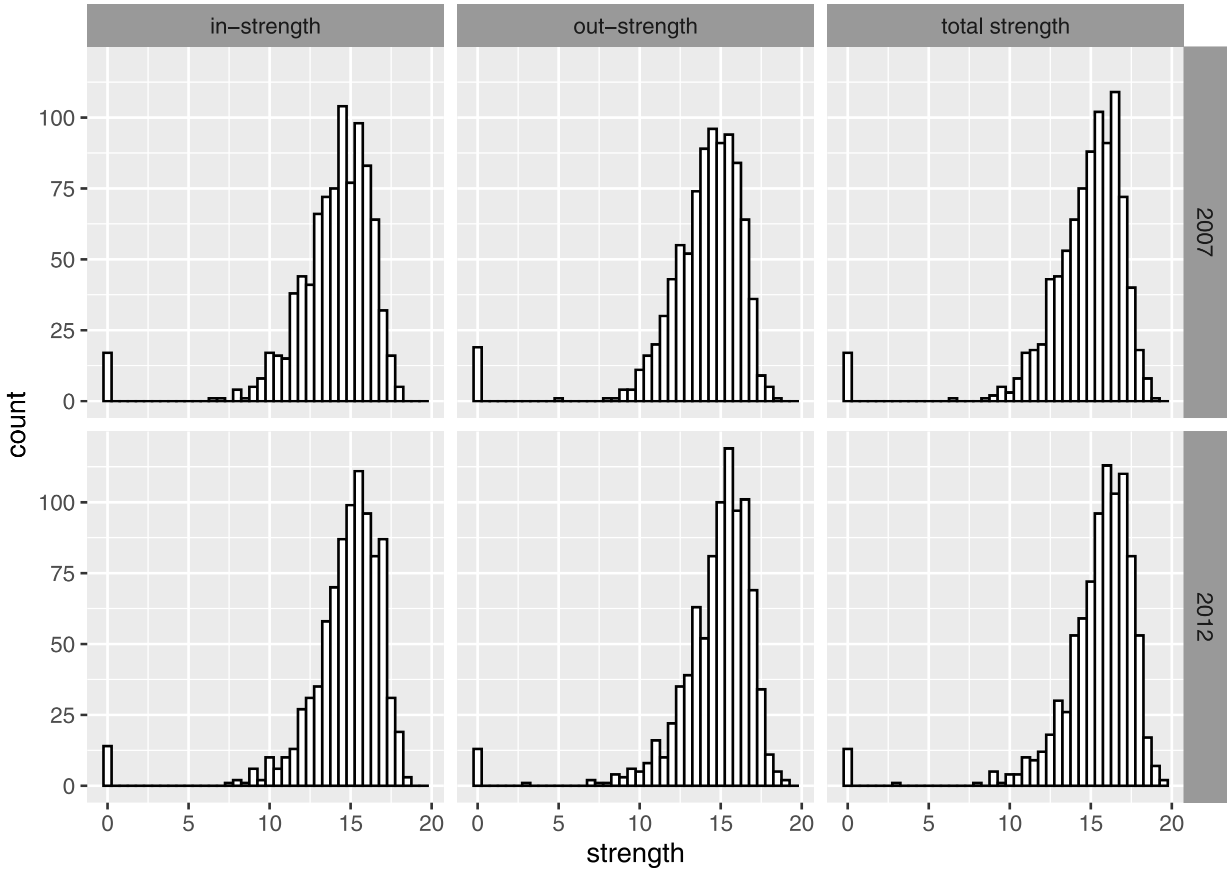

Fig. 4. In-, out- and total-strength distributions (after logarithmic transformation) of the MRIONs in 2007 and 2012.

Table 4
Five kinds of assortativity coefficients (including "total" that does not account for edge direction) of the (national, intra-province and inter-province) MRIONs in 2007 and 2012, where "UW" and "W" respectively represents the unweighted and weighted versions of the assortativity measures.

Type	2007						2012					
	National		Intra-prov.		Inter-prov.		National		Intra-prov.		Inter-prov.	
	UW	W	UW	W	UW	W	UW	W	UW	W	UW	W
in-in	−0.010	0.501	0.258	0.610	−0.033	0.023	−0.003	0.573	0.209	0.649	−0.008	−0.022
in-out	−0.001	0.447	0.124	0.549	−0.006	−0.024	−0.001	0.516	0.109	0.594	−0.002	−0.034
out-in	−0.123	0.493	0.007	0.599	−0.136	0.117	−0.110	0.563	0.010	0.635	−0.111	0.130
out-out	−0.024	0.474	0.070	0.572	−0.028	0.047	0.015	0.536	0.095	0.618	0.016	−0.001
total	−0.070	0.418	0.140	0.544	−0.080	0.046	−0.030	0.457	0.158	0.561	−0.031	0.028

4.2. Assortativity

Table 4 presents a collection of assortativity coefficients for the MRIONs of 2007 and 2012. For each year, the assortativity coefficients were computed for both directed (four types) links and undirected links; intra-province links, inter-province links, and nationwide links; unweighted links and weighted links.

Our first observation is that the unweighted assortativity coefficients [53,56] are not informative for characterizing the MRIONs. The unweighted versions are similar to the weighted versions in magnitude only for inter-province links, which are all close to zero. However, they are mostly different in signs. For nationwide links and intra-province links, the weighted and unweighted assortativity coefficients of all five types, including four directed and one undirected, are notably different. The maximum magnitude of the unweighted version is only 0.123 (out-in in 2007), which is much lower than the magnitudes of the weighted versions in the range of 0.4–0.6. The two versions have completely different signs for all five types of nationwide links in 2007. The unweighted versions suggest that there is a negligible pattern of disassortative mixing, while the weighted versions suggest assortative mixing. The weighted versions are more consistent with intuition.

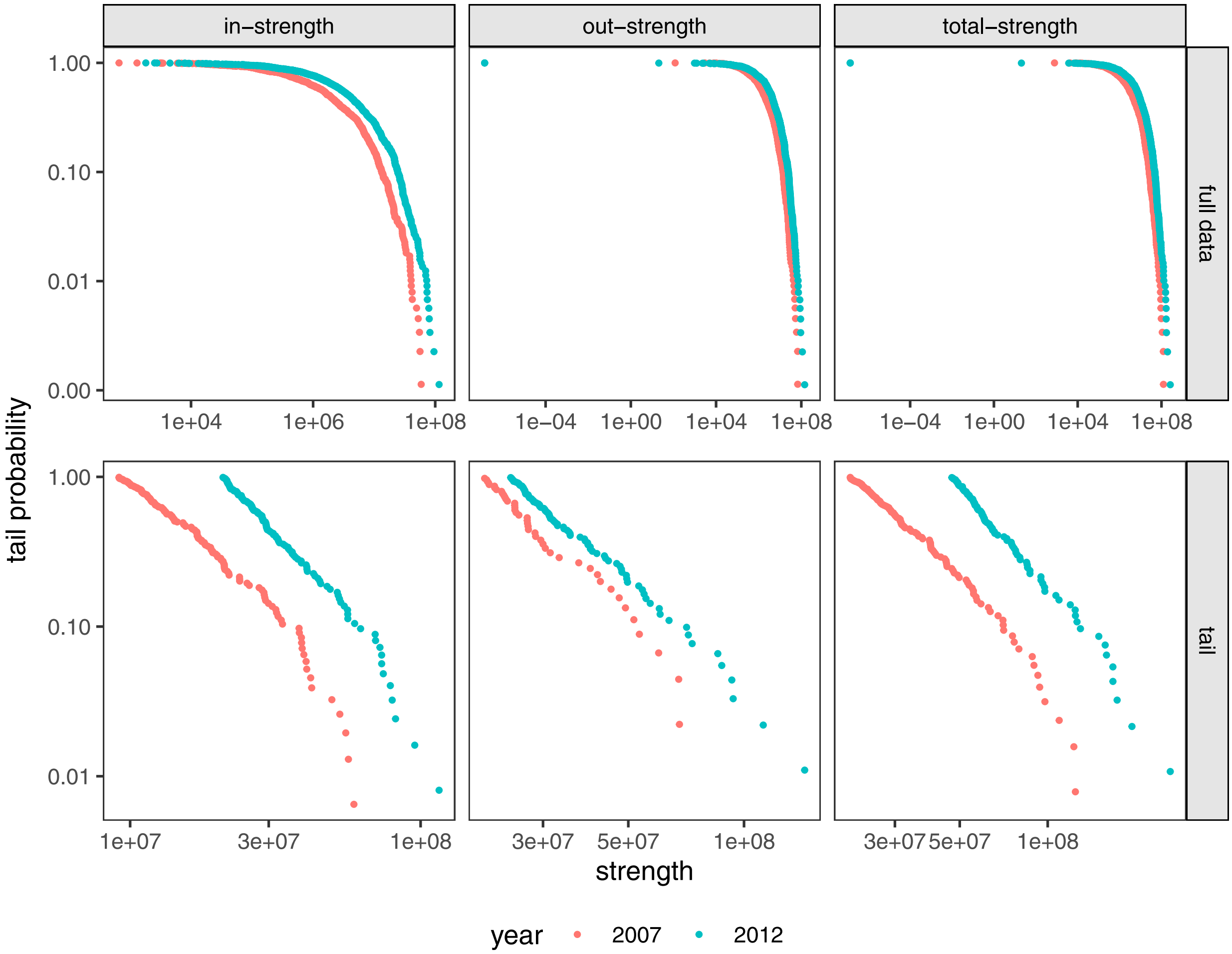

Fig. 5. Tail distributions of in-, out- and total-strengths of the MRIONs in 2007 and 2012 (top three panels). Their correspondingly zoomed tails after the estimated thresholds are given in the bottom three panels.

Some existing analyses of the WIOTs without weight also reported "close to zero" assortativity coefficients [e.g.,39], which need to be revisited by using the weighted assortativity measures [58].

Based on the results from the weighted definitions, the assortativity coefficients for nationwide links of all kinds have magnitude from 0.4 to 0.6. These are moderately strong assortative mixing. Take the out-in assortativity for nationwide links as an example. The coefficient of 0.493 in 2007 and 0.563 in 2012 suggest that province-sectors with large inputs are likely to take high transaction volumes from the others with high outputs in the network. Decomposing the weighted adjacency matrix helps separate the contributions respectively from intra-province and inter-province links. All five assortativity coefficients for intra-province links are much greater than those for inter-province links, which are close to zero. That is, economic transactions among the sectors from the same province are extensively closely connected, with high transaction volumes; in contrast, sectors from multiple regions are relatively loosely connected, and the majority of the transaction volumes represented by existing link weights are extremely small. The differences between intra- and inter-province assortativity coefficients support the well-known theory of regional fragmentation [19].

The assortativity coefficients of all types for nationwide links and intra-province links increased from 2007 to 2012, while those for inter-province links remained close to zero. The increases in nationwide links are therefore attributed to the increase in intra-province links, suggesting an increase in the degree of provincial segmentation. One possible explanation, for example, is that the economic stimulus plan after the 2008 financial crisis stimulated the construction industry most, which propagated to upstream metal and non-metal mining/processing sectors within each province. Despite the increased inter-provincial transactions, there is no clear assortative pattern among these transactions in contrast to the intra-provincial transactions.

The assortativity coefficient provides a platform to demonstrate the effectiveness of backbone. Fig. 6 shows the national assortativity coefficients for the backbones of the MRIONs in 2007 and 2012 for a sequence of significance levels

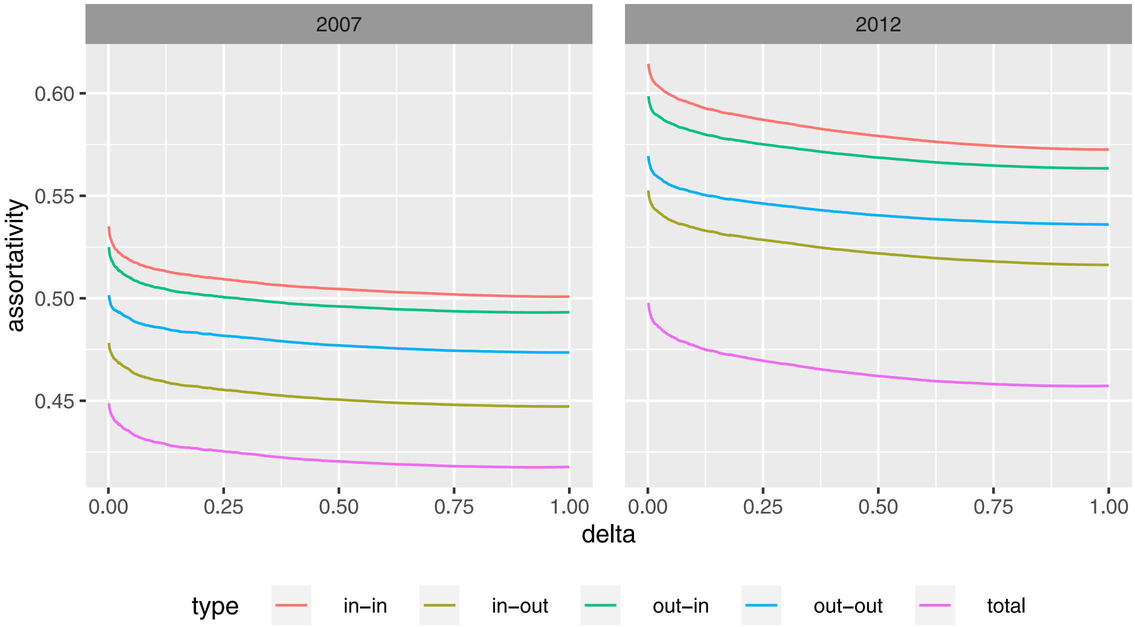

Fig. 6. Five kinds of assortativity coefficients (nationwide) at a sequence of significance levels, $\delta = \{0.001, 0.002, \ldots, 0.999\}$ in 2007 and 2012.

Table 5
Total, cycle-, middleman-, in- and out-clustering coefficients of the MRIONs at national, intra-province and inter-province levels in 2007 and 2012.

Type	2007			2012		
	National	Intra-prov.	Inter-prov.	National	Intra-prov.	Inter-prov.
Total	0.874	0.942	0.855	0.968	0.969	0.937
Cycle	0.828	0.933	0.815	0.960	0.962	0.933
Middleman	0.927	0.952	0.888	0.975	0.975	0.942
In	0.914	0.949	0.881	0.968	0.971	0.935
Out	0.843	0.939	0.819	0.966	0.967	0.939

$\delta = \{0.001, 0.002, \ldots, 0.999\}$. For each year, the value of each assortativity coefficient increases, but only slightly, with the decrease of δ due to the removal of the non-essential edges. The removed edges are supposed to impose limited impact on the overall structure of the network. As a result, the magnitude of change in each assortativity coefficient is small. Furthermore, for each δ, all kinds of the assortativity coefficients in 2012 are greater than their counterparts in 2007, which is consistent with the results from Table 4. Therefore, backbone is a parsimonious and powerful tool for uncovering the fundamental and essential properties of a network, especially for the large-scale networks that are likely to cause computational expensiveness.

4.3. Clustering coefficients

Clustering coefficients were computed with and without edge direction for three link types, nationwide, intra-province, and inter-province, in the 2007 and 2012 MRIONs; see Table 5. All the clustering coefficients have large values (close to 1), providing stronger evidence than simple link densities for immense connectivity of the MRIONs. The larger values in 2012 suggest a higher tendency that the province-sectors would cluster together in terms of forming triangles. Decomposing the nationwide links to intra-province and inter-province links reveals that the nationwide increase from 2007 to 2012 was mainly due to the increase in inter-province components. For example, the nationwide cycle-clustering coefficient increased from 0.828 to 0.960; the intra-province coefficient 0.933 in 2007 was quite high, leaving not much room to increase; the inter-province coefficient increased from 0.815 in 2007 to 0.933 in 2012. The emergence of more transactions across the inter-province sectors may be attributed to the government's strategies and policies such as the Great Western Development Strategy. The increase in inter-province transaction is not in contradiction to its small proportion in the overall magnitude, so it did not affect the manifestation of regional fragmentation in China.

Among the four types of directed clustering coefficients, the cycle- and out-clustering coefficients have increased more notably than the others. In a cyclic-triangle connection, each province-sector is an upstream as well as a downstream

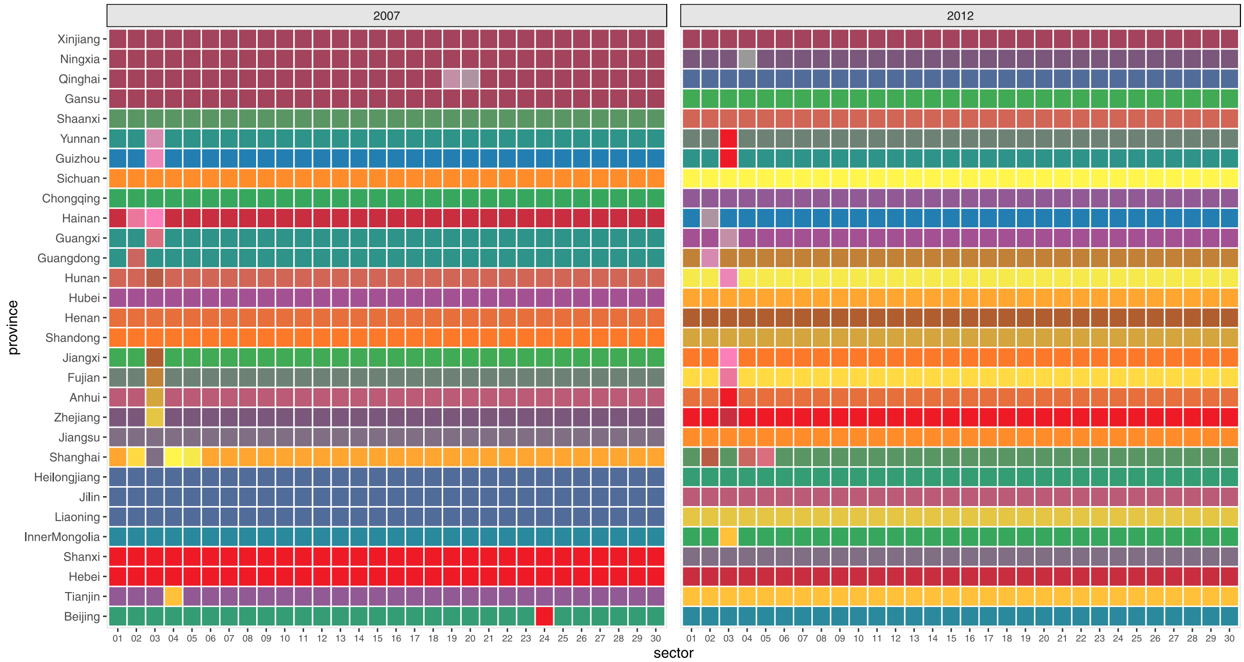

Fig. 7. Communities of the MRIONs in 2007 (left panel) and 2012 (right panel).

of its neighbors. A higher value of cycle-clustering coefficient indicates a higher proportion of triangular (supply and demand) chains formed by the province-sectors. In an out-triangle connection, a province-sector is always the upstream to its neighbors. A higher value of out-clustering coefficient suggests an increased proportion transactions among the downstream sectors.

4.4. Community detection

The community detection results from modularity maximization are visualized in two side-by-side heat maps in Fig. 7 respectively for 2007 and 2012. For each heat map, province-sectors in the same community have the same color. Between the two years, however, the colors are not comparable because these colors are nominal within each community detection task. There were 39 and 40 communities in 2007 and 2012, respectively. A common feature is that most sectors from the same province belong to the same community. This is expected, as intra-province economic ties are naturally tighter than inter-province economic ties for geographic, historical, and administrative reasons. An interesting discovery is that heavy industry sectors, such as "coal mining and processing" (02), "petroleum and gas extracting" (03), and "metals mining/processing" (04) usually form singletons independent from province-based communities. For example, Shanghai as a manufacturing and business center usually inquires a high demand of raw materials like coal, which heavily relies on the supplies from other provinces. Consequently, "coal mining and processing" (02) of Shanghai forms a singleton instead of falling into the same community formed by the most of the other sectors from Shanghai.

From 2007 to 2012, the community structure has shown a notable change. Sectors from the provinces in the same geographic region tended to stay in the same community in 2007. For example, three northeastern provinces Heilongjiang, Jilin, and Liaoning were in one community; four northwestern provinces Xinjiang, Ningxia, Qinghai, and Gansu belonged to another community; two central north provinces Shanxi and Hebei were placed in the same community. In 2012, however, this pattern was no longer observed, each province appearing to be a community of its own. A closer look at the data reveals that the growth rate of inter-province trade is much smaller than that of intra-province trade. In 2007, there were 9.6 trillion CNY of inter-province trade and 35.5 trillion CNY intra-province trade in China. In 2012, the inter-province trade increased to 12.8 trillion CNY (an increase of 30%), while the intra-province trade increased to 68.4 trillion CNY (an increase of 90%). In some provinces, Jilin for instance, there has been almost no change in inter-province trade from 2007 to 2012, whereas the intra-province trade has been tripled.

The community detection results again echos the regional fragmentation of the Chinese economy at the province level [19]. In the early 2000s, under the effect of regional integration strategies such as the Great Western Development and Northeast Revitalization, the transactions among the sectors in certain regions mushroomed, forming multi-province communities. In response to the 2008 global financial crisis, the central government advocated a four trillion Chinese economic stimulus plan, which sustained the economic growth while the world economy slowed down [92]. The majority

Table 6
The top 10 province-sectors in weighted PR scores ($\gamma = 0.85$) from the MRIONs in 2007 and 2012 with and without TVA (total value-added) as prior information.

Rank	$\theta = 0$		$\theta = 1$			
	No prior		No prior		Using TVA as prior	
	2007	2012	2007	2012	2007	2012
1	Beijing 15	Hainan 21	Beijing 30	Yunnan 24	Guangdong 30	Yunnan 24
2	Beijing 10	Hainan 12	Guangdong 30	Zhejiang 24	Beijing 30	Guangdong 30
3	Beijing 14	Hainan 06	Yunnan 24	Guangdong 24	Guangdong 19	Guangdong 24
4	Beijing 16	Hainan 17	Guangdong 19	Jiangsu 17	Jiangsu 30	Zhejiang 24
5	Beijing 12	Hainan 14	Guangdong 18	Hebei 24	Shandong 06	Jiangsu 30
6	Beijing 13	Hainan 16	Guangdong 17	Shaanxi 24	Zhejiang 30	Jiangsu 17
7	Beijing 17	Hainan 09	Jiangsu 30	Shanghai 24	Guangdong 18	Shandong 06
8	Jiangsu 17	Hainan 13	Guangdong 08	Qinghai 24	Shanghai 30	Jiangsu 12
9	Jiangsu 18	Tianjin 12	Shandong 06	Sichuan 24	Jiangsu 12	Shandong 12
10	Jiangsu 22	Hebei 06	Shanghai 30	Xinjiang 24	Shandong 30	Guangdong 19

of these funds were reallocated from the budget of provincial and local governments, which supported infrastructure projects and housing developments; some assisted local governments to lend to state-owned companies for developing estates [93]. Therefore, by 2012, the extent of the regional fragmentation has expanded despite the original goals of implementing the regional integration strategies in the early 2000s. The extent of the regional fragmentation in the Chinese economy after 2012 is of great interest when the 2017 MRIOT becomes available.

4.5. Centrality

The PR scores were used to rank the relative importance of province-sectors in the Chinese economy. The dampling factor was set to $\gamma = 0.85$ as recommended by Brin and Page [78]. For different values of $\theta = \{0, 1\}$, we computed the PR scores for the 900 province-sectors in 2007 and 2012 with or without accounting for prior information. Specifically, the total value added (TVA) of each province-sector was adopted as node-specific prior information, as it indicates the added value contributed by each sector to the national economy. The TVAs of province-sectors in each year are recorded in quadrant IV of the MRIOT; see Table 1. We present the province-sectors with top 10 PR scores in Table 6.

Without considering the weight ($\theta = 0$), most of the top 10 province-sectors are from Beijing in 2007 and from Hainan in 2012. Beijing's top ranking in 2007 may be explained by its special function as the nation's capital with an advantage in access to the resources nationwide. The preparation for the 2008 Olympic Games in urban infrastructure, ecological environment, electronic technology and other aspects had a huge pull effect on the economic development of Beijing, especially in the manufacturing sectors. Nonetheless, the sizes of the sectors in Beijing are smaller than those in eastern coastal provinces such as Guangdong or Jiangsu. The top ranking of Beijing without weight is, therefore, not consistent with broad perceptions. The top ranking of seven sectors in Hainan in 2012 is even more puzzling. As an island with a small population, Hainan is known to be relatively less developed in comparison to many other provinces in China. The unweighted PR scores are unsatisfactory in measuring the province-sector centrality in the context of MRIONs.

With PR scores fully based on weights instead of counts ($\theta = 1$), the top 10 province-sectors have changed substantially. Province-wise, most of the top 10 province-sectors in 2007 are from the eastern coast (Guangdong, Jiangsu, Shandong, and Shanghai). These provinces are more developed than others, and consistently make large contribution to the national GDP. In 2012, however, some sectors from much less developed provinces (Shaanxi, Qinghai, and Xinjiang) joined those from traditional developed provinces in the top 10. This may be a result of these less developed provinces in northwest China benefiting from the effectiveness of the China Western Development policy. Sector-wise, the sector of "other services" (30) appeared most often in the top 10 in 2007, but "construction" (24) became dominant in 2012. Note that "other services" cover some essential services like financial services and information technology, both of which are critical to modern economic development. It is evident that these services have provided unprecedented support to the growth of many other sectors in China in the early 2000s. The 2008 global financial crisis stroke many such services. Much of the four-trillion CNY stimulus program funded projects like railway, highway, bridge, and aviation construction. Further, one of the aims of the China Western Development program was to strengthen the infrastructure construction in the participating provinces.

The utilization of TVA as prior information led to noticeable changes in both lists of top 10 PR province-sectors. The changed results are more consistent with intuition. For 2007, Guangdong became less dominant than otherwise, albeit still with the highest frequency in the top 10. Most of the provinces came from the eastern/southern coast (except for Beijing). Different from the diversity in province, "other services" (30) appeared to be most influential sector-wise, as it occupied six positions of the top ten. The impact of the TVA prior is more notable in 2012 than in 2007. Sectors from western provinces like Shaanxi, Qinghai and Xinjiang were gone, while more sectors from the coastal provinces emerged in the top 10 list. The "construction" (24) sector is less dominant, but remaining most frequent in the top 10. Other leading province-sectors are "other services" (30) from Guangdong and Jiangsu and "chemical industry" (12) from Jiangsu and

Table 7
The province-sectors with top 10 PR scores ($\gamma = 0.85$, $\theta = 1$) of the backbones of MRIONs in 2007 and 2012 at significance level $\delta = \{10^{-3}, 10^{-4}\}$. TVA is used as prior information. The blue colored province-sectors with * do not appear in the corresponding columns for $\delta = 1$ (original networks).

Rank	$\delta = 1$		$\delta = 10^{-3}$		$\delta = 10^{-4}$	
	2007	2012	2007	2012	2007	2012
1	Guangdong 30	Yunnan 24	Guangdong 30	Guangdong 24	Guangdong 30	Guangdong 24
2	Beijing 30	Guangdong 30	Beijing 30	Guangdong 30	Beijing 30	Guangdong 30
3	Guangdong 19	Guangdong 24	Guangdong 19	Zhejiang 24	Guangdong 19	Jiangsu 17
4	Jiangsu 30	Zhejiang 24	Shandong 06	Jiangsu 17	Shandong 06	Zhejiang 24
5	Shandong 06	Jiangsu 30	Jiangsu 30	Yunnan 24	Jiangsu 30	Yunnan 24
6	Zhejiang 30	Jiangsu 17	Guangdong 18	Jiangsu 30	Guangdong 08*	Jiangsu 30
7	Guangdong 18	Shandong 06	Guangdong 08*	Shandong 06	Guangdong 18	Shandong 06
8	Shanghai 30	Jiangsu 12	Guangdong 17*	Jiangsu 12	Zhejiang 30	Jiangsu 12
9	Jiangsu 12	Shandong 12	Zhejiang 30	Guangdong 19	Guangdong 17*	Shandong 12
10	Shandong 30	Guangdong 19	Shandong 30	Shandong 12	Shandong 30	Guangdong 19

Shandong. The updated results with the TVA prior seem more reasonable because TVA contains the information about self-loops which were otherwise discarded but are useful in assessing centrality.

It is worth special attention that “construction” (24) of Yunnan is top 1 in 2012 in spite of the inclusion of prior information. Albeit a developing inland province, Yunnan is one of the largest tourism provinces in China. The fast development of tourism in Yunan has boosted the development of infrastructure construction such as transportation facilities and hotel accommodations. Located in the geographical center of Asia, connecting southeast Asia with China and inland with coastal regions, Yunnan has been crucial to the China Western Development. With favorable domestic policies and economic cooperation with Southeast Asian countries, Yunnan’s GDP grew with a rate consistently higher than the national average during this period, for which the construction sector played an important pulling role [94]. In fact, construction is a long-known key sector nationwide of at the province level. For example, the random walk closeness centrality [42] indicated that the most essential sectors in almost two thirds of the provinces were “construction” (24) in 2007. For the MRIONs, however, our results appear to be more consistent with intuition than those obtained with centrality measures not accounting for edge weight or TVA information [42].

The PR score provides another opportunity to demonstrate the effectiveness of backbone. Table 7 summarizes the province-sectors with top 10 PR scores in the backbones of MRIONs with significance level $\delta = \{10^{-3}, 10^{-4}\}$. No drastic change in the lists for both 2007 and 2012 are observed. Only two new province-sectors, “clothing, leather, fur” (08) and “transport equipment” (17) from Guangdong, ranked among the top 10 in both of the backbones but not in the original 2007 MRION. In fact, they were ranked respectively at 13 and 14 in the original MRION, with small difference in the magnitude of PR score from the bottom of top 10. For 2012, the province-sectors in the top 10 list remained the same for both backbones, in spite of some changes in order. The traditional strong sectors “construction” (24) and ”other services” (30) in Guangdong, “transport equipment” (17) in Jiangsu and “construction” (24) in Zhejiang surpassed “construction” (24) in Yunnan. This was because “construction” (24) in Yunnan was more connected by insignificant edges in the MRION of 2012, but these edges were filtered out in the backbones. Between the two lists for the backbones at different significance levels, we only observe small dispersion in province-sector orders. For instance, in the two lists of 2012, the province-sectors at rank 3 and 4 and those at rank 9 and 10 were respectively switched, with the rest remaining identical. This comparison, again, shows that backbone is capable of detecting the central province-sectors in the MRIONs.

5. Discussions

MRIOTs provide a natural arena for network analyses to study regional and sectoral structure of an economy. Our study is the first of such kind to analyze the most recent MRIONs of China for the year of 2007 and 2012. All three types of strength distributions (in, out, and total) were skewed to the left, where the degree of skewness of each increased over time. For each MRION, the positive assortativity coefficients suggested assortative mixing across the province-sectors, especially intra-province-sectors. As indicated by close-to-one clustering coefficients, the province-sectors tended to cluster together, and the tendency increased in inter-province transactions. Province-based community structures were detected. There were communities containing multiple provinces in 2007 but none in 2012, suggesting increased regional fragmentation. The most essential province-sectors in the Chinese economy were identified through a new class of weighted PR measures. Province-wise, Guangdong, Jiangsu and many eastern coastal provinces contained most sectors in the top list. Sector-wise, “construction” (24) and “other services” (30) appeared to be dominant.

Our study suggests a few methodological caveats for network analyses of MRIONs, some of which may be applicable beyond MRIONs. First, it is critical to account for edge weight when summarizing MRIONs as these networks are weighted; otherwise, misleading inference may be made. For example, unweighted assortativity coefficients suggest almost no pattern of assortative (or disassortative) mixing in the MRIONs, whereas the weighted counterparts indicate a moderately positive assortative mixing. While no weight is accounted, the classical PR algorithm have produced a top 10 list in 2012 containing 8 sectors form Hainan, which is not consistent with the fact. Rather, the weighted PR algorithm (with or without

Table A.8
Codes of the provinces in the MRIONs.

Code	Province	Code	Province
01	Beijing	16	Henan
02	Tianjin	17	Hubei
03	Hebei	18	Hunan
04	Shanxi	19	Guangdong
05	Inner Mongolia	20	Guangxi
06	Liaoning	21	Hainan
07	Jilin	22	Chongqing
08	Heilongjiang	23	Sichuan
09	Shanghai	24	Guizhou
10	Jiangsu	25	Yunnan
11	Zhejiang	26	Shaanxi
12	Anhui	27	Gansu
13	Fujian	28	Qinghai
14	Jiangxi	29	Ningxia
15	Shandong	30	Xinjiang

using TVA as prior information) has provided a much more reasonable result. Second, precise interpretation of the network measures is crucial. For instance, clustering coefficient is a relative measure. A large local clustering coefficient of a vertex only tells how likely its two neighbors are connected; it says nothing about how many neighbors it has. For instance, the province-sector of the highest local clustering coefficient in 2012 is given to "construction" (24) from Inner Mongolia according to the computation. This specific province-sector receives a high value since it has fewer neighbors compared to the rest in the nation, leading to a higher proportion conversely. Third, when ranking the vertices in a network, we recommend to make full use of the possible vertex-specific auxiliary information from the MRIOTs such as final use, import, or value-added, which helps lead to more accurate results. Finally, when the final use section is of an MRIOT is of specific interest, network analysis based on the intermediate use section my be of limited help. A final use network could be constructed by aggregating the region-sectors over regions at the cost of some information loss, and its significance merits further study.

CRediT authorship contribution statement

Tao Wang: Conceptualization, Resources, Writing - original draft. **Shiying Xiao:** Software, Validation, Data curation, Writing - original draft, Visualization. **Jun Yan:** Conceptualization, Writing - review & editing, Supervision. **Panpan Zhang:** Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Tao Wang and Shiying Xiao were supported by the National Bureau of Statistics of China (2018lz33).

Appendix A. Province codes

Table A.8 summarizes the codes and the names of the 30 provinces used in the chord graph in Fig. 1.

Appendix B. Four types of local clustering coefficients

In practice, the formulations of the four kinds of local clustering coefficients are respectively given by

$$C_i^{\text{in}} = \frac{\left[\boldsymbol{W}^\top \left(\boldsymbol{A} + \boldsymbol{A}^\top\right) \boldsymbol{A}\right]_{ii}}{2s_i^{(\text{in})}\left(d_i^{(\text{in})} - 1\right)}, \tag{B.1}$$

$$C_i^{\text{out}} = \frac{\left[\boldsymbol{W} \left(\boldsymbol{A} + \boldsymbol{A}^\top\right) \boldsymbol{A}^\top\right]_{ii}}{2s_i^{(\text{out})}\left(d_i^{(\text{out})} - 1\right)}, \tag{B.2}$$

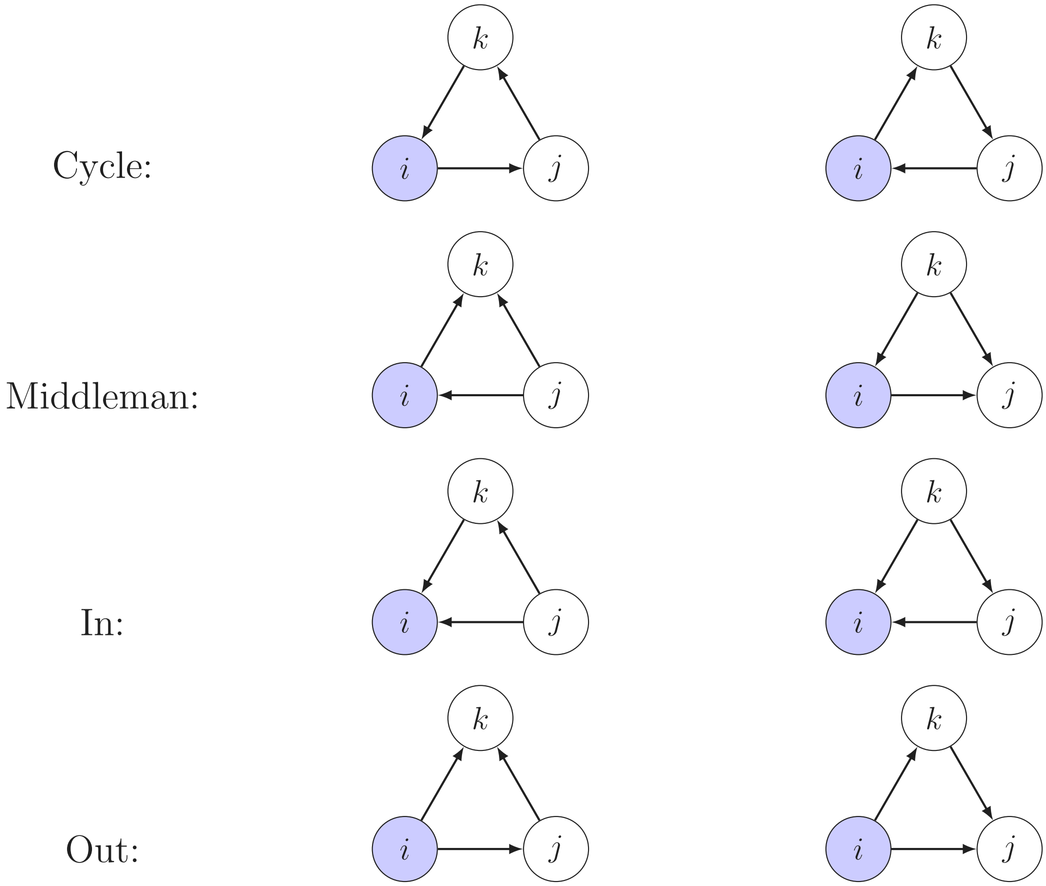

Fig. B.8. Four types of triangles proposed in Fagiolo [66], and reused in Clemente and Grassi [67].

$$C_i^{\text{mid}} = \frac{\left(\boldsymbol{W}^\top \boldsymbol{A}\boldsymbol{A}^\top + \boldsymbol{W}\boldsymbol{A}^\top \boldsymbol{A}\right)_{ii}}{\left(s_i^{(\text{in})} d_i^{(\text{out})} + s_i^{(\text{out})} d_i^{(\text{in})}\right) - (\boldsymbol{A}\boldsymbol{W} + \boldsymbol{W}\boldsymbol{A})_{ii}}, \tag{B.3}$$

$$C_i^{\text{cyc}} = \frac{\left[\boldsymbol{W}\boldsymbol{A}^2 + \boldsymbol{W}^\top \left(\boldsymbol{A}^\top\right)^2\right]_{ii}}{\left(s_i^{(\text{in})} d_i^{(\text{out})} + s_i^{(\text{out})} d_i^{(\text{in})}\right) - (\boldsymbol{A}\boldsymbol{W} + \boldsymbol{W}\boldsymbol{A})_{ii}}. \tag{B.4}$$

References

[1] International Monetory Fund, GDP, current prices [data file], 2020, Available online: https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD.
[2] J.Y. Lin, China and the global economy, China Econ. J. 4 (1) (2011) 1–14.
[3] The World Bank, GDP Growth (annual %) [data file], 2020, Available online: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG.
[4] The World Bank, GDP (current US$) [Data file], 2020, Available online: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.
[5] The World Bank, Imports of goods and services (current US$) [Data file], 2020, Available online: https://data.worldbank.org/indicator/NE.IMP.GNFS.CD.
[6] The World Bank, Exports of goods and services (current US$) [Data file], 2020, Available online: https://data.worldbank.org/indicator/NE.EXP.GNFS.CD.
[7] S. Fan, X. Zhang, S. Robinson, Structural change and economic growth in China, Rev. Dev. Econ. 7 (3) (2003) 360–377.
[8] S. Chen, G.H. Jefferson, J. Zhang, Structural change, productivity growth and industrial transformation in China, China Econ. Rev. 22 (1) (2011) 133–150.
[9] J. Jia, G. Ma, C. Qin, L. Wang, Place-based policies, state-led industrialisation, and regional development: Evidence from China's great western development programme, Eur. Econ. Rev. 123 (2020) 103398.
[10] G.C. Chow, Impact of joining the WTO on China's economic, legal and political institutions, Pac. Econ. Rev. 8 (2) (2003) 105–115.
[11] C. Yuan, S. Liu, N. Xie, The impact on Chinese economic growth and energy consumption of the global financial crisis: An input–output analysis, Energy 35 (4) (2010) 1805–1812.
[12] S. Fan, R. Kanbur, X. Zhang, China's regional disparities: Experience and policy, Rev. Dev. Finance 1 (1) (2011) 47–56.
[13] H. Li, K.E. Haynes, Economic structure and regional disparity in China: Beyond the Kuznets transition, Int. Reg. Sci. Rev. 34 (2) (2011) 157–190.

[14] B.-S. Lee, J. Peng, G. Li, J. He, Regional economic disparity, financial disparity, and national economic growth: Evidence from China, Rev. Dev. Econ. 16 (2) (2012) 342–358.
[15] C. He, J. Wang, Regional and sectoral differences in the spatial restructuring of Chinese manufacturing industries during the post-WTO period, Geol. J. 77 (3) (2012) 361–381.
[16] C. He, Y. Yan, D. Rigby, Regional industrial evolution in China, Pap. Reg. Sci. 97 (2) (2018) 173–198.
[17] B. Meng, Y. Fang, J. Guo, Y. Zhang, Measuring China's domestic production networks through trade in value-added perspectives, Econ. Syst. Res. 29 (1) (2017) 48–65.
[18] A. Young, The Razor's edge: Distortions and incremental reform in the People's Republic of China, Q. J. Econ. 115 (4) (2000) 1091–1135.
[19] S. Poncet, A fragmented China: Measure and determinants of Chinese domestic market disintegration, Rev. Int. Econ. 13 (3) (2005) 409–430.
[20] L.N. Moses, The stability of interregional trading patterns and input-output analysis, Am. Econ. Rev. 45 (5) (1955) 803–826.
[21] W. Leontief, A. Strout, Multiregional input-output analysis, in: T. Barna (Ed.), Structural Interdependence and Economic Development: Proceedings of an International Conference on Input-Output Techniques, Geneva, September 1961, Palgrave Macmillan UK, London, UK, 1963, pp. 119–150.
[22] A. Tukker, E. Dietzenbacher, Global multiregional input-output framework: An introduction and outlook, Econ. Syst. Res. 25 (1) (2013) 1–19.
[23] Y. Zhang, S. Qi, China Multi-Regional Input-Output Models, China Statistics Press, Beijing, ISBN: 9787503765049, 2012.
[24] T.R. Anderson, Intermetropolitan migration: A comparison of the hypotheses of Zipf and Stouffer, Am. Sociol. Rev. 20 (3) (1955) 287–291.
[25] A.G. Wilson, A statistical theory of spatial distribution models, Transp. Res. 1 (3) (1967) 253–269.
[26] W. Liu, J. Chen, Z. Tang, H. Liu, D. Han, F. Li, Theory and Practice of Compiling China 30-Province Inter-Regional Input-Output Table of 2007, China Statistics Press, Beijing, China, ISBN: 9787503764714, 2012.
[27] W. Liu, Z. Tang, M. Han, The 2012 China Multi-Regional Input-Output Table of 31 Provincial Units, China Statistics Press, Beijing, China, ISBN: 9787503784521, 2018.
[28] J.H. Bergstrand, The gravity equation in international trade: Some microeconomic foundations and empirical evidence, Rev. Econ. Stat. 67 (3) (1985) 474–481.
[29] A.L.M. Sargento, Empirical examination of the gravity model in two different contexts: Estimation and explanation, Rev. Reg. Res. 27 (2) (2007) 107–127.
[30] J. Ji, Z. Zou, Y. Tian, Energy and economic impacts of China's 2016 economic investment plan for transport infrastructure construction: An input-output path analysis, J. Cleaner Prod. 238 (2019) 117761.
[31] S. Liang, Y. Wang, T. Zhang, Z. Yang, Structural analysis of material flows in China based on physical and monetary input-output models, J. Cleaner Prod. 158 (2017) 209–217.
[32] M. Jiang, P. Behrens, T. Wang, Z. Tang, Y. Yu, D. Chen, L. Liu, Z. Ren, W. Zhou, S. Zhu, C. He, A. Tukker, B. Zhu, Provincial and sector-level material footprints in China, Proc. Natl. Acad. Sci. USA 116 (52) (2019) 26484–26490.
[33] K. Fujikawa, C. Milana, Input-output decomposition analysis of sectoral price gaps between Japan and China, Econ. Syst. Res. 14 (1) (2002) 59–79.
[34] M.P. Timmer, E. Dietzenbacher, B. Los, R. Stehrer, G.J. de Vries, An illustrated user guide to the world input–output database: The case of global automotive production, Rev. Int. Econ. 23 (3) (2015) 575–605.
[35] J. Defourny, E. Thorbecke, Structural path analysis and multiplier decomposition within a social accounting matrix framework, Econ. J. 94 (373) (1984) 111–136.
[36] E.J. Feser, E.M. Bergman, National industry cluster templates: A framework for applied regional cluster analysis, Reg. Stud. 34 (1) (2000) 1–19.
[37] A. Leonidov, E. Serebryannikova, Dynamical topology of highly aggregated input–output networks, Physica A 518 (2019) 234–252.
[38] M. Xu, S. Liang, Input-output networks offer new insights of economic structure, Physica A 527 (2019) 121178.
[39] F. Cerina, Z. Zhu, A. Chessa, M. Riccaboni, World input-output network, PLoS One 10 (7) (2015) 1–21.
[40] R.M. del Río-Chanona, J. Grujić, H. Jeldtoft Jensen, Trends of the world input and output network of global trade, PLoS One 12 (1) (2017) 1–14.
[41] J. Amador, S. Cabral, Networks of value-added trade, World Econ. 40 (7) (2017) 1291–1313.
[42] X. Sun, H. An, X. Liu, Network analysis of chinese provincial economies, Physica A 492 (2018) 1168–1180.
[43] W. Liu, Z. Tang, J. Chen, P. Yang, The 2010 China Multi-Regional Input-Output Table of 30 Provincial Units, China Statistics Press, Beijing, China, ISBN: 9787503771316, 2014.
[44] Z. Mi, J. Meng, H. Zheng, Y. Shan, D. Guan, A multi-regional input-output table mapping China's economic outputs and interdependencies in 2012, Sci. Data 5 (1) (2018) 180155.
[45] The World Bank, Inflation, GDP deflator (annual %) - China [Data file], 2020, Available online: https://data.worldbank.org/indicator/NY.GDP.DEFL.KD.ZG?locations=CN.
[46] M.E.J. Newman, Networks: An Introduction, Oxford University Press, New York, NY, USA, ISBN: 9780199206650, 2010.
[47] P. Erdös, A. Rényi, On random graphs. I, Publ. Math. Debrecen 6 (1959) 290–297.
[48] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[49] D.M. Pennock, G.W. Flake, S. Lawrence, E.J. Glover, C.L. Giles, Winners don't take all: Characterizing the competition for links on the web, Proc. Natl. Acad. Sci. USA 99 (8) (2002) 5207–5211.
[50] X. Gabaix, Zipf's law for cities: An explanation, Q. J. Econ. 114 (3) (1999) 739–767.
[51] L. Kaplow, Pareto principle and competing principles, in: S.N. Durlauf, L.E. Blume (Eds.), The New Palgrave Dictionary of Economics, Palgrave Macmillan, London, UK, 2008, pp. 4807–4812.
[52] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703.
[53] M.E.J. Newman, Assortative mixing in networks, Phys. Rev. Lett. 89 (20) (2002) 208701.
[54] R. van der Hofstad, N. Litvak, Degree-degree dependencies in random graphs with heavy-tailed degrees, Internet Math. 10 (3–4) (2014) 287–334.
[55] M.E.J. Newman, Mixing patterns in networks, Phys. Rev. E 67 (2) (2003) 026126.
[56] J.G. Foster, D.V. Foster, P. Grassberger, M. Paczuski, Edge direction and the structure of networks, Proc. Natl. Acad. Sci. USA 107 (24) (2010) 10815–10820.
[57] R. Noldus, P. Van Mieghem, Assortativity in complex networks, J. Complex Netw. 3 (4) (2015) 507–542.
[58] Y. Yuan, J. Yan, P. Zhang, Assortativity coefficients for weighted and directed networks, J. Complex Netw. 9 (2) (2021) cnab017.
[59] T. Opsahl, P. Panzarasa, Clustering in weighted networks, Social Networks 31 (2) (2009) 155–163.
[60] M.E.J. Newman, D.J. Watts, S.H. Strogatz, Random graph models of social networks, Proc. Natl. Acad. Sci. USA 99 (Suppl. 1) (2002) 2566–2572.
[61] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.
[62] P. Grindrod, Range-dependent random graphs and their application to modeling large small-world proteome datasets, Phys. Rev. E 66 (2002) 066702.
[63] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proc. Natl. Acad. Sci. USA 101 (11) (2004) 3747–3752.
[64] J.-P. Onnela, J. Saramäki, J. Kertész, K. Kaski, Intensity and coherence of motifs in weighted complex networks, Phys. Rev. E 71 (2005) 065103.
[65] B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis, Stat. Appl. Genet. Mol. Biol. 4 (1) (2005) 17.

[66] G. Fagiolo, Clustering in complex directed networks, Phys. Rev. E 76 (2007) 026107.
[67] G.P. Clemente, R. Grassi, Directed clustering in weighted networks: A new perspective, Chaos Solitons Fractals 107 (2018) 26–38.
[68] T.A.B. Snijders, K. Nowicki, Estimation and prediction for stochastic blockmodels for graphs with latent block structure, J. Classification 14 (1) (1997) 75–100.
[69] M.S. Handcock, A.E. Raftery, J.M. Tantrum, Model-based clustering for social networks, J. Roy. Statist. Soc. Ser. A 170 (2007) 301–354.
[70] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99 (12) (2002) 7821–7826.
[71] G. Ouyang, D.K. Dey, P. Zhang, Clique-based method for social network clustering, J. Classification 37 (1) (2020) 254–274.
[72] A. Goldenberg, A.X. Zheng, S.E. Fienberg, E.M. Airoldi, A survey of statistical network models, Found. Trends Mach. Learn. 2 (2) (2010) 129–233.
[73] M.E.J. Newman, Modularity and community structure in networks, Proc. Natl. Acad. Sci. USA 103 (23) (2006) 8577–8582.
[74] A. Clauset, M.E.J. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E 70 (6) (2004) 066111.
[75] A.Y. Ng, M.I. Jordan, Y. Weiss, On spectral clustering: Analysis and an algorithm, in: T.G. Dietterich, S. Becker, Z. Ghahramani (Eds.), Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, The MIT Press, Cambridge, MA, USA, 2001, pp. 849–856.
[76] M.E.J. Newman, Scientific collaboration netowrks. II. Shortest paths, weighted networks, and centrality, Phys. Rev. E 64 (1) (2001) 016132.
[77] P. Bonacich, Power and centrality: A family of measures, Am. J. Sociol. 92 (5) (1987) 1170–1182.
[78] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Comput. Netw. ISDN Syst. 30 (1) (1998) 107–117.
[79] K. Das, S. Samanta, M. Pal, Study on centrality measures in social networks: A survey, Soc. Netw. Anal. Min. 8 (2018) 13.
[80] P. Rasmussen, Studies in Intersectoral Relations (Ph.D. thesis), University of Copenhagen, Amsterdam, Netherlands, 1956.
[81] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1) (1953) 39–43.
[82] P.S. Laumas, The weighting problem in testing the linkage hypothesis, Q. J. Econ. 90 (2) (1976) 308–312.
[83] E. Dietzenbacher, The measurement of interindustry linkages: Key sectors in the Netherlands, Econ. Model. 9 (4) (1992) 419–437.
[84] P. Zhang, T. Wang, J. Yan, Pagerank centrality and algorithms for weighted, directed networks with applications to world input-output tables, 2021, arXiv preprint, arXiv:2104.02764.
[85] L. Page, S. Brin, R. Motwani, T. Winograd, The pagerank citation ranking: Bringing order to the web, in: P.H. Enslow, A. Ellis (Eds.), Proceedings of the 7th International World Wide Web Conference, Elsevier, 1998, pp. 161–172.
[86] P. Berkhin, A survey on pagerank computing, Internet Math. 2 (1) (2005) 73–120.
[87] M.Á. Serrano, M. Boguñá, A. Vespignani, Extracting the multiscale backbone of complex weighted networks, Proc. Natl. Acad. Sci. USA 106 (16) (2009) 6483–6488.
[88] X. Zhang, J. Zhu, Skeleton of weighted social network, Physica A 392 (6) (2013) 1547–1556.
[89] N.J. Foti, J.M. Hughes, D.N. Rockmore, Nonparametric sparsification of complex multiscale networks, PLoS One 6 (2) (2011) 1–10.
[90] Z. Ghalmane, C. Cherifi, H. Cherifi, M. El Hassouni, Extracting backbones in weighted modular complex networks, Sci. Rep. 10 (1) (2020) 15539.
[91] L. Liu, Y.-C.T. Shih, R.L. Strawderman, D. Zhang, B.A. Johnson, H. Chai, Statistical analysis of zero-inflated nonnegative continuous data: A review, Statist. Sci. 34 (2) (2019) 253–279.
[92] M. Ouyang, Y. Peng, The treatment-effect estimation: A case study of the 2008 economic stimulus package of China, J. Econometrics 188 (2) (2015) 545–557.
[93] D. Chen, S. Khan, X. Yu, Z. Zhang, Government intervention and investment comovement: Chinese evidence, J. Bus. Finance Account. 40 (3–4) (2013) 564–587.
[94] X. Su, Multi-scalar regionalization, network connections and the development of yunnan province, China, Reg. Stud. 48 (1) (2014) 91–104.