Revised Floyd-Warshall Algorithm to Find Optimal Path in Similarity-Weight Network and Its Application in the Analysis of Global Value Chain

Lizhi Xing¹,², Yujie Li²

¹ College of Economics and Management, Beijing University of Technology, Beijing, China
² School of Online and Continuing Education, Beijing Foreign Studies University, Beijing, China

Email: itwasa@163.com; liyujie@bfsu.edu.cn

Abstract. This paper focuses on measuring the globally and nationally economic system’s connectedness and industrial sector’s function on the Global Value Chain (GVC), as reinforcements to the present studies on international trade. Firstly, we reconsidered the length-related and position-related measures in literatures about vertical specialization from the perspective of econophysics. Secondly, we redefined the inter-country and inter-sector propagating process of intermediate goods and proposed the concept of Strongest Relevance Path Length (SRPL) based on Revised Floyd-Warshall Algorithm (RFWA), which is the basis of new measurement. Thirdly, enlightened by betweenness centrality of node and edge, we introduced SRPL-based index to measure the Value-Added Pivotability of Industrial Sectors and Input-Output Relationships. Fourthly, we also proposed the concepts of Backward Closeness and Forward Closeness of Industrial Sectors based on closeness centrality.

1. Introduction

The rise of Global Value Chain (GVC) has naturally captured the attention of international trade economists eager to bridge the apparent gap between the new characteristics of the international organization of production and the standard methods used to collect, manipulate, and interpret international trade statistics. In particular, a remarkable body of work has devised ingenious empirical methods to disentangle the value-added and intermediate inputs of international trade flows.

According to the definition of value chain proposed by Porter, it represents value-adding process at every stage of production, from R&D, design, to the delivery of the final product and service to consumers. This process may just happen within a country or a region, but it is becoming increasingly international and sophisticated nowadays under the urge of globalization. Thus, trying to measure industrial sectors’ position as well as consequent function on the GVC is not an easy job.

There is no doubt that measurement of length and upstreamness/downstreamness contributes a lot to the studies of vertical specialization. According to our study on the application of complex network theory [1,2,3], network-based algorithms and indices have great potential to shed new light on further understanding of industrial sector’s position and function in consideration of the network-form architecture of GVC. However, only a few scholars adopted this kind of method to analyze countries positions in the international global value networks as far as now, such as Mesa-Arango et al. [4] and Cingolani et al. [5].
2. Construction of the Global Value Chain Network

To establish an industrial complex network, we consider a sector category within a region as a node and the inter-industry IO relationship as a tie, whose weight represents the sale and purchase relationships between producers and consumers. Thus, a graph \( G = (V, E, W) \) containing \( n \) nodes is created, representing sectors within a nation or region, denoted as a node set \( V \). Pairs of nodes are linked by ties reflecting their interdependencies, constituting an asymmetric tie set \( E \). However, in valued graphs, a set \( E \) can be replaced by weight set \( W \), which can be extracted from the region of inter-country inter-industry use and supply in ICIO table. We name the ICIO network model Global Industrial Value Chain Network (GIVCN), since its purpose is to reflect how economic shocks propagate and amplify along GVC, as well as to what extent the industrial impact generates at the inter-country level.

3. Optimal Path Searching Algorithm

We proposed a network-based framework to detect inter-sector relevance. In details, the information of sector’s function and inter-sector relative position on the GVC is already embedded in the topological structure of GIVCN models, so the first thing is to redefine the optimal path to reflect the propagation process of intermediate goods on the premise of considering the properties of economic system.

3.1. The Issue of Path in Similarity-Weight Network

Both Betweenness and Closeness centrality rely on identifying optimal paths, which in the case of binary data can be unproblematically identified as shortest paths. But if the edges are weighted, there are a variety of possibilities in assessing the optimality of a path. In other words, there is no one generalization for weighted networks, and it all depends on what kinds of network process we are studying.

For many researchers studying ICIO networks, the most common style they used is to take the approach of binarizing the weighted edges and running the traditional Breadth-First-Search (BFS) algorithms, in consideration of the computational intension. In addition, someone defined the optimal path as the maximum shortest path length between any two pairs [6], as well as took the minimum shortest path length on the basis of taking the reciprocal of weights. They did so in part because of they wanted to obtain the metric of consecutive paths, however, it is inappropriate.

In our opinion, GVC’s topological structure is very similar with supply chain, which sheds a light on the issue of path in GIVCN models. As we all know, the Bullwhip Effect, also known as the Forrester Effect [7], is a distribution channel phenomenon in which forecasts yield supply chain inefficiencies. It refers to increasing swings in inventory in response to shifts in customer demand as one moves further up the supply chain. For the sake of weakening it, the most direct and efficient way is to reduce the length of transaction process as well as improve the efficiency and effectiveness of each step. Therefore, in this paper, we designed a new sort of BSF algorithm to trace inter-sector crucial connection intermediate goods flowing on in respect of both propagation efficiency and effectiveness.

3.2. Revised Floyd-Warshall Algorithm

In network science, path plays a central role, and many network-based indices are developed around it. The Shortest Path between nodes \( i \) and \( j \), also called Distance or Geodesic Path, is the path with the fewest number of edges linking them, denoted by \( d_{ij} \). If there is no such path between them, it means \( d_{ij} = \infty \). Besides, it is possible to find there are multiple shortest paths of the same length \( d_{ij} \). In general, Floyd-Warshall Algorithm (FWA) as a classical BFS algorithms is used to find all inter-node shortest paths in a recursive way, which means it compares all possible paths through the network between each pair of nodes by incrementally improving an estimate on the shortest path until the estimate is optimal [8]. The core formula of FWA is:

\[
d_{ij}^{(k)} = \min_{i,j,k=1,k \neq i,j} \left\{ d_{ij}^{(k-1)} + d_{ik}^{(k-1)} + d_{kj}^{(k-1)} \right\} (i \neq j)
\]

Eq (1) works by first computing \( d_{ij}^{(k)} \) for all pairs of source-to-sink nodes for \( k = 1 \), then \( k = 2 \), etc. This process continues until \( k = N \), and we have found the shortest path using any intermediate nodes.
However, in weighted networks, particularly those with similarity weights such as IO system, we could use neither the objective function of minimization nor the iterative accumulation of paths. This is because it is meaningless to find the least efficient channel of intermediate goods, and we will get an infinite result inevitably if just taking it as a maximization problem. Hence, we must find one kind of fast and highly effective restraining algorithm to solve this operational research problem about network flow. In fact, it is possible to create a method to reconstruct the actual path between any two endpoint nodes with simple modifications to FWA, just like what scholars have done [9].

Therefore, we divided \( \frac{w_{ik}}{w_{kj}} \) by \( \frac{w_{ik}}{w_{kj}} \) and named the result the \textit{Relevance Path Length (RPL)}.

After transformation, we got
\[
RPL_{ij} = \left( \frac{1}{w_{ik}} + \frac{1}{w_{kj}} + \cdots + \frac{1}{w_{ij}} \right)^{-1} = \left( \sum_{i,j,k,l \in \{1,2,\ldots,N\}} w_{ik}^{-1} w_{kj}^{-1} \right)^{-1}
\]

where \( w_{ij} \) is the weight of an edge on the path connecting nodes \( i \) and \( j \), and the exact number of \( RPL_{ij} \) is up to that of possible paths delivering the inter-sector intermediate goods. From the form, this formula is similar with parallel resistors. As for the self-loops existing in some networks, for instance the ICIO network, \( RPL_{ii} = w_{ii} \) according to the Eq (2).

Then, we considered how to find the optimal one among RPLs, and named it the \textit{Strongest Relevance Path Length (SRPL)}, which is the synthetic gauge of both spreading effectiveness and efficiency of a given path in the meanwhile as far as possible. For this purpose, we put forward the \textit{Revised Floyd-Warshall Algorithm (RFWA)} as an iterative algorithm based on operations research:

\[
SRPL_{ij}^{(k)} = \max_{i,j,k=1,2,\ldots,N} \left\{ w_{ij}^{(k-1)} \times \frac{w_{ik}^{(k-1)} w_{kj}^{(k-1)}}{w_{ik}^{(k-1)} + w_{kj}^{(k-1)}} \right\}
\]

According to Eq (3), we will compare RPLs successively until picking out the largest one. If it is bigger than \( w_{ij}^{(k-1)} \), we keep the record as it is, otherwise just make it equal to \( w_{ij}^{(k-1)} \). When the \( SRPL_{ij}^{(k)} \) happen to be equal to \( w_{ij}^{(k-1)} \), it means the optimal path is just the most direct one between nodes \( i \) and \( j \). In other words, there is no need to take even one more step via another node. Due to the nature of RPLs, the selected SRPL converges when finding the maximum. Besides, the self-loops contain non-negligible topological information, such as self-consumption of industrial sectors in ICIO network. Therefore, we incorporated this sort of edge into the comparison process, which means the source node and sink node of a RPL could be the same one.

3.3. Theoretical Basis of SRPL in GIVCN Model

By Porter’s definition in 1985, an industry value-chain is a physical representation of the various processes involved in producing goods (and services), starting with raw materials and ending with the delivered product. It is based on the notion of value-added at the stage of production level. In fact, Wasilly Leontief’s IO table, published in the 1950s, already provided estimates of the relative importance of each individual link in industry-level value chains. In recent decades, multinational enterprises locate a lot of activities, such as research, development, design, assembly, production of parts, marketing and branding, in different countries around the globe, to enhance efficiency and to optimize profits. It is the theory of comparative advantage behind the globalization, so here comes the concept of GVC. Furthermore, analytical tools stemming from complex network have been proven to be very effective in analysis of GVC, thus we proposed the SRPL in this paper to find the optimal path of intermediate goods propagation based on many theories and models.

The gravity model of international trade in international economics is a model that, in its traditional form, predicts bilateral trade flows based on the economic sizes and distance between two units. The
model was first introduced in economics world by Walter Isard in 1954 [10]. The basic model for trade between two countries \( i \) and \( j \) takes the form of

\[
T_{ij} = \frac{A \times Y_i \times Y_j}{D_{ij}}
\]

(4)

where \( A \) is the constant, \( T_{ij} \) stands for trade flow, \( D_{ij} \) stands for the distance and \( Y_i, Y_j \), as well as \( Y_j, Y_i \), stands for the economic dimensions of the countries that are being measured, often using GDP measurements.

Before defining the RPL and searching SRPL, we noticed that many similar studies didn’t take the length of value chain into account, which makes them overly qualitative or even meaningless. Then, the formula of gravity model shed light on the principle of our BFS algorithm, i.e., an intermediate goods transfer path is directly proportional to the relevant IO relations and inversely proportional to the length of value-added process. Thus, we replaced \( Y_i \times Y_j \) with \( w_i^{(k-1)} \times w_j^{(k-1)} \) to measure the influence from industrial structure, and \( D_{ij} \) with \( w_i^{(k-1)} + w_j^{(k-1)} \) to provide distance information (the denominator of RPLs of two or more steps incorporates the number of relevant steps after employing fractionation, and increasing steps yield more transaction cost as well), explaining the dynamics of intermediate goods transferring from upstream sectors to downstream ones (sometimes they are the same one). As results, SRPL based on RFWA owns two aspects of economical meaning: one is to find all the optimal propagation paths of intermediate goods with both higher IO relationship and lower transaction cost between any pairs of sectors (including self-loops); another is to measure the inter-sector (maybe inner-sector) relationship strength from the standpoint of integral value chain.

4. Derivative Measures

As we all known, comparative advantage is the economic reality describing the work gains from trade for individuals, firms, or nations, which arise from differences in their factor endowments or technological progress. We believe that the GVC exactly stems from comparative advantage of nations all around the world, and SRPL can be used to identify these important value chains from ICIO network. Then, measuring with SRPL-based indicators and forecasting through simulation become feasible. As results, we can propose international trade policies from the angle of econophysics.

4.1. Betweenness Centrality

In similarity-weight network, however, the basis of information propagation differs from Boolean network, thus we proposed Weighted Betweenness Centrality of Node based on RFWA and denote it by \( C_{BFWA}^\text{SRPL} \). Then, we encountered an important issue here, that is, how to treat those nodes locating on both ends of every single SRPL. Can we just ignore them? The answer is no. As the framework has been expended to multi-layer network, in this case, all the industrial sectors processing value-added are between value-added level units and final demand level units. Correspondingly, they as nodes in GIVCN models are neither on the topest of value chain nor on the lowest. Therefore, there is no need to struggle with the function of either source nodes or sink nodes on the SPRLs, just focus on the frequency of appearance. We will assign different meanings or usages to source nodes and sink nodes when running more complex applications. The formula of \( C_{BFWA}^\text{SRPL} \) is:

\[
C_{BFWA}^\text{SRPL} (i) = \sum_{i,j,k \in \{1,2,...,N\}} \text{SRPL}^\text{SRPL} (i)
\]

(5)

where \( \text{SRPL}^\text{SRPL} (i) \) is the number of SRPLs within the scope of the whole network, which are connecting any pairs of others cross a given sector. Of special interest is how to measure the node with self-loop of its being a SRPL. The frequency seems to be 2 because this node appears as both the source node and sink node on this special path, which is, however, irrational. Thus, we set a rule that the number of appearances of the node on the same path is less than or equal to 1.

According to Eq (5), the most significant difference is we do not normalize it by the totality anymore, because we are more concerned with industrial sectors' propagation function on the GVC rather than their weakened heterogeneity reflected by ratio. Above all, we adopted it to measure the
Value-Added Pivotability of Industrial Sectors, with the purpose to evaluate the level of brokerage in the process of turnover of intermediate goods. For example, suppose that a given sector owns high pivotability, many sectors will go through it to reach the others via efficient and effective paths. In principle, this sector has power because it could threaten to stop transferring, making sectors use less efficient and effective paths to reach one another. But this superiority only works if the other sectors cannot easily enter into new trade relations around the world. In other words, not only can we use pivotability to evaluate the current global economic situation, but also offer favorable suggestions for the future.

To find the most important inter-sector economic relationships among numerous couples of sectors in GIVCN models, we introduced Weighted Betweenness Centrality of Edge based on RFWA to measure the Value-Added Pivotability of Input-Output Relationships:

$$C_{E}^{RFWA}(i,j) = \sum_{h,k,l,n \in [1,2,\ldots,N]} SRPL_{hk}^{(N)}(i,j)$$

where $SRPL_{hk}^{(N)}(i,j)$ is the number of SRPLs within the scope of the whole network, which are connecting any pairs of others cross a given couple of sectors $i$ and $j$. Self-loop serving as a SRPL will plus 1 to this indicator of itself.

4.2 Closeness Centrality

GIVCN models took both weight and direction of nodes into account to make maximum use of ICIO data, thus we need to turn to new SRPL-based measures. Firstly, directed closeness centrality should be divided into two sorts: Weighted In-Degree Closeness Centrality and Out-Degree Closeness Centrality based on RFWA, denoted by $C_{C}^{RFWA-IN}$ and $C_{C}^{RFWA-OUT}$, according to nodes’ position on the propagation path, serving as a sink node or source node respectively. Secondly, the shortest paths cannot effectively reflect the propagation efficiency of information in weighted network any longer as mentioned above, and it is therefore replaced with SRPLs while computation. Finally, yet importantly, the relative position of numerator and denominator needs to be changed, because SRPLs reflect the most efficient ways that information passes through in similarity-weight network, and the numerical value of closeness centrality should be proportional to the average value of it. Considering the above-mentioned, new closeness centralities in directed similarity-weight network based on RFWA are introduced:

$$C_{C}^{RFWA-IN}(i) = \frac{\sum_{j=1}^{N} SRPL_{ij}^{(N)}}{N}$$

$$C_{C}^{RFWA-OUT}(i) = \frac{\sum_{j=1}^{N} SRPL_{ji}^{(N)}}{N}$$

By the nature of Eq (7) and Eq (8), the summation of all the backward linkages is just equal to that of forward linkages, so there is a kind of conservation relationship between them in the closed economic system.

From the point of backward linkage, bigger $C_{C}^{RFWA-IN}$ means sectors rely much more on the intermediate goods from their upstream providers in the international division. And from the opposite side, the bigger $C_{C}^{RFWA-OUT}$, the more intermediate goods contributed to the downstream consumers. We defined the $C_{C}^{RFWA-IN}$ and $C_{C}^{RFWA-OUT}$ as Backward Closeness and Forward Closeness of Industrial Sectors, to quantify backward and forward inter-sector closeness degree starting from a given sector respectively. Note that, inner-sector self-consumptions denoted by self-loops in the network are included in both directions.

In sum, our hypothesis is that the relative position of industrial sector on the GVC could be reflected by backward closeness and forward closeness in this paper, and we may figure out what kind of comparative advantage different location will bring to it. Thus, we put forward an analytical framework and will adopt it in empirical studies.
5. Conclusions
The major difference between our framework and the others is that we argue there is no starting point (R&D and design) or ending point (the delivery of final products or services to consumer) in the economic system. Every single sector on the supply side is called the upstream sector only if it directly or indirectly provides intermediate products or services to one consumer at least, while each sector on the demand side is taken as the downstream sector only if it directly or indirectly consumes intermediate products or services from even sole provider.

6 Acknowledgment
The authors acknowledge support from Beijing Municipal Natural Science Foundation (Grant No. 9194024, Grant Recipient: Lizhi Xing), Beijing Municipal Social Science Foundation (Grant No. 17ZGC010, Grant Recipient: Lizhi Xing), Ri-Xin Talents Project of Beijing University of Technology (Grant Recipient: Lizhi Xing), and Technology Plan Key Program of Beijing Municipal Education Commission (Grant No. KZ20181005010, Grant Recipient: Jun Guan).

7. References
[1] Xing L, Ye Q, Guan J. Spreading Effect in Industrial Complex Network Based on Revised Structural Holes Theory [J]. PLoS ONE, 2016, 11(5): e0156270.
[2] Xing L, Dong X, Guan J. Global Industrial Impact Coefficient Based on Random Walk Process and Inter-Country Input-Output Table [J]. Physica A: Statistical Mechanics and its Applications, 2017, 471: 576-591.
[3] Xing L, Guan J, Wu S. Measuring the Impact of Final Demand on Global Production System Based on Markov Process [J]. Physica A: Statistical Mechanics and its Applications, 2018, 502:148-163.
[4] Mesa-Arango R, Narayanan B, Ukkusuri S V. The Impact of International Crises on Maritime Transportation Based Global Value Chains [J]. Networks & Spatial Economics, 2015:1-28.
[5] Cingolani I, Panzarasa P, Tajoli L. Countries’ positions in the international global value networks: Centrality and economic performance [J]. Applied Network Science, 2017, 2(1):2-21.
[6] Tsekeris T. Global value chains: building blocks and network dynamics [J]. Physica A Statistical Mechanics & Its Applications, 2017, 488.
[7] Forrester J W. Industrial dynamics[M]. M.I.T. Press, 1961.
[8] Floyd R W. Algorithm 97: Shortest path [J]. Communications of the Acm, 1962, 5(6):345.
[9] Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths [J]. Social Networks, 2010, 32(3):245-251.
[10] Isard W. Location Theory and Trade Theory: Short-Run Analysis [J]. Quarterly Journal of Economics, 1954, 68(2):305-320.