Which sectors of a modern economy are most central?

Florian Blöchl
HelmholtzZentrum Munich

Eric O'N. Fisher
California Polytechnic State University

Fabian Theis
HelmholtzZentrum Munich *

February 18, 2010

Abstract

We analyze input-output matrices for a wide set of countries as weighted directed networks. These graphs contain only 47 nodes, but they are almost fully connected and many have nodes with strong self-loops. We apply two measures: random walk centrality and one based on count-betweenness. Our findings are intuitive. For example, in Luxembourg the most central sector is "Finance and Insurance"and the analog in Germany is "Wholesale and Retail Trade" or "Motor Vehicles", according to the measure. Rankings of sectoral centrality vary by country. Some sectors are often highly central, while others never are. Hierarchical clustering reveals geographical proximity and similar development status.

1 Introduction

It is natural to think of an input-output matrix as a network. Each sector is a vertex, and the flow of economic activity from one sector to another constitutes an edge. Studying the network properties of these matrices poses four practical problems. First, at the usual level of aggregation, these networks are dense; they are typically completed connected. Second, they are directed; for example, in the United States in 2003, \$11.3 billion of rubber and plastic products were used in the production of motor vehicles, but only \$97 million of the output of the motor vehicle industry was used in the production of rubber and plastic products. Third, these networks have self-loops; in the same case, more than thirty percent of total industry output was used as its own input. Fourth, they are weighted; for example, the output of the sector transit and ground transportation was only about six percent of that of motor vehicles in the United States in 2003.

In this paper we develop two measures of betweenness that are suited for these networks. We apply our measures to a wide array of input output tables for the OECD countries. These data are

*The authors' emails are florian.bloechl@helmholtz-muenchen.de, efisher@calpoly.edu, and fabian.theis@helmholtz-muenchen.de. Fisher thanks the ETH at Zurich for the hospitality that allowed this work to be completed. All the data and Matlab programs are available upon request.

consistent in two ways. First, they are derive from macroeconomic accounts; the total of value added across sectors is equal to national income. Second; they are consistent across countries. The level of aggregation and the definition of sectors allows us to compare networks across countries in an appealing and intuitive way.

Each edge is the local currency value of a sector's output that is used as an input into another sector, including perhaps itself. An industry's outputs need not be closely related to its inputs, and each row of the table is as a directed flow of economic activity. For example, motor vehicle industry may be severely affected by bottlenecks in the production of rubber and plastic products, but the converse is not true.

We are not the first to emphasize that an input output matrix is a directed network. Gancho Ganchev, Lothar Krempel, and Margarita Shivergeva [6] give a nice visual representation of the structure of the Bulgarian economy. McNerney [12] also investigates applications of network analysis to national input-output tables, although he empahsizes measures of flow, not centrality.

Freeman [3] introduced the notion of centrality in a network; he defined the centrality of a node as the average number of shortest links between pairs of other nodes that pass through it. His definition is not adequate for an economic network in which edges may have different capacities. Freeman, Borgatti, and White [5] describe a measure of flow for weighted networks that is based upon a maximum capacity of flows between nodes. Their measure ignores the possibility of parallel processing, whereby information might flow between nodes through many different channels. Addressing this deficiency forthrightly, Newman [13] defined random-walk betweenness. Our measures build upon his important work, and they are easy to calculate.

It is alleged that Leontief [9] developed aspects of input-output accounting during the Second World War partly as an attempt to help identify strategic weaknesses in the German economy. The techniques of input-output accounting have ready applications in economic planning. One of the most important reasons for collecting and constructing economic data at a disaggregated level is to identify the influence of sectors on national economic activity. Fischer Black (1987) hypothesized the business cycle might arise because of the propagation of shocks between the sectors of an economy, and Long and Plosser [10] developed an elegant analysis of the United States economy based on this idea. Our analysis is an attempt to quantify which sectors are most central in this process.

The first step in our analysis is to normalize the row sums to unity.[1] Then a row shows the shares of output that goes into each sector; it is also the probability that a given dollar's worth of output will flow into any one sector. Hence, we do not distinguish between economies that have very different vectors of aggregate output, as long as commodities flow the same way in every sector. This has the advantage of allowing us to compare economies of very different sizes, and it is somewhat akin to defining countries as having identical technologies when their unit input requirements are identical. For us, two economies are identical when their normalized output flows are the same for every sector.

Consider a unit of agricultural output. In equilibrium, a farmer will be indifferent between selling to the manufacturing sector or the construction sector since the marginal revenues are identical. Hence the output will flow randomly to any one of several sectors. Also, payments received will arrive from any random business using the output as an intermediate input; after all, a dollar is green no matter its provenance. Thus centrality measures based upon the random flow of goods between sectors–and the corresponding random flows of payments between businesses–will be quite apt. We

[1] Indeed, the matrices as raw data are not comparable for two reasons. First, each table is defined in a local currency. Second, there is an enormous difference between th volume of economic activity in American, which accounts for about 24% of world GDP, and the Slovak Republic, which contributes 0.2% to world output.

develop and implement two of them.

Our first centrality measure is based upon the concept of Mean First Passage Time, a measure of the distance between a source and target sector. Imagine a supply shock to any one sector. The incremental output will flow randomly into all the sectors of the economy. We compute the expected time it takes first to reach any particular target sector, where it is used for final demand. We argue that a sector is most central if on average any random supply shock first flows through it. For example, in the United States economy in 2000, the sector called “Public admin. & defence; compulsory social security”has the lowest Mean First Passage Time when one averages across all possible supply shocks. We consider it the most central sector in the American economy; in essence, the government “purchases”as intermediate inputs a broad array of the outputs (including compulsory social insurance) from many different sectors. Thus it will feel the effects of supply shocks fairly early.

Our second measure is counting centrality. Again, imagine a supply shock falling on any one sector. The extra value added will eventually leak out of the system of intermediate inputs as a good or service used for final demand. But first the incremental output will flow randomly through the economy, causing secondary effects everywhere before it leaks out to satisfy the final demand for consumption, investment, government purchases, or net exports. We keep track of how often it is expected to visit any node, and we average these numbers of visits across all possible pairs of supply shocks and sectoral outflow for final demand For example, in the German economy in 2000, the sector called “Motor vehicles”has the highest counting centrality.

The rest of this paper is structured as follows. The second section gives some preliminary definitions. In the third section, we develop our two centrality measures and then contrast using a simple network. The fourth section shows our results; to our knowledge, it is the first use in economics of hierarchical clustering to identify similarities among countries. The fifth section presents some brief conclusions and suggestions for future research.

2 Definitions

We describe two measures of centrality that are designed to highlight aspects of the input output matrices. Both are based on the concept of random walks in graphs. These measures are highly correlated, but each has a slightly different focus.

Let $G = (V, E)$ be a connected, weighted, and directed graph, consisting of a set of vertices V and a set of edges $E \subset V \times V$. Each edge $(i, j) \in E$ is assigned a non-negative real weight a_{ij}. Our graph may contain self-loops but one an only one edge connects the ordered pair (i, j). The number of vertices and edges is denoted by n and m respectively.

The graph can be represented by its $n \times n$ adjacency matrix $A = (a_{ij})$, where the $(i, j) - th$ element represents the weight of edge $i \to j$. To keep notation simple, we name the vertices by natural numbers, and we can identify them with the according positions in the adjacency matrix. The out-degree of node i is $k(i) = \sum_{j=1}^{n} a_{ij}$ and the set of out-neighbors of i by $N(i) = \{j \mid (i, j) \in E\}$, so $k(i) = \sum_{j \in N(i)} a_{ij}$. This yields the total weight of the graph $k(G) = \sum_{i \in V} k(i)$.

Any real economic transaction has its monetary counterpart. Thus we can model the movement of goods between sectors or the corresponding flow of payments. The weight a_{ij} in an input-output matrix corresponds to the value of goods produced in sector i sold to sector j. Hence it is the nominal value of a flow of commodities or services from i to j and also the corresponding flow of monetary payments from j to i.

We model the movement of goods by random walks; see [1] for more details. In graph theory, a random walker starts out at a given position with an intended destination. He or she repeatedly chooses an edge incident to the current position, and these choices are made according to a probability distribution determined by the edge weights. The random walker proceeds until the goal is reached. In an input-output table, a random walk keeps track of a dollar circulating through the economy, with the transition probabilities given by the flow of goods and services between sectors. Because of the dual nature of all economic transactions, we are keeping track of the flow of the value of goods and services from the source to the destination and also the flow of a dollar from the destination back to the source.

The input-output matrices give us the sales of a large number firms in each sector. Hence by normalizing an input-output matrix by its row sums, we get the transition probabilities for sales of output by sector. We work with the Markov matrix

$$M = K^{-1}A, \tag{1}$$

where K is the diagonal matrix of the out-degrees $k(i)$ defined above.[2] It is entirely possible for a dollar to become stuck in a sector if it makes sales only to itself and records no other transactions with the rest of the economy, including payments to the factors of production.

The input-output matrix A is not a closed system. In particular, its row sums are not equal to its column sums. The table records only sales by firms to other firms of goods and services used as intermediate inputs in the production process. In national accounts, the total value of the gross output of a sector is a row sum that includes sales for final demand, broken into consumption, investment, government purchases, and net exports. The total value of gross inputs into a sector is a column sum that includes payments to the factors of production called, gross operating surplus, compensation to employees, and indirect business taxes.

3 Two Centrality Measures

In this section, we define two centrality measures that describe the nodes in input-output matrices well. We also give a simple example that contrasts them.

3.1 Random Walk Centrality

In social network analysis, closeness centrality, introduced by Freeman [4], is a widely used measure. It is usually defined as the inverse of the mean geodesic distance from all nodes to a given one. For an input-output network, this measure makes little sense; no dollar knows how to travel along a shortest path between sectors. It can take an arbitrarily long route, and it may even pass over the same link more than once. In fact, a dollar could easily cycle for a long time between sectors i to j before eventually moving on to k. Indeed, all real economies are so densely connected that one can get from any sector to any other in at most two transactions. Since there is an ineluctable element of

[2] In our empirical implementation, there are countries with sectors recording no output. These arise because of data limitations in the local national accounts. The most serious case is the Russian Federation, where the OECD records output in only 22 sectors. In essence, such a sector splits the economy into two disconnected components. Our empirical work is based upon matrices where none of the row sums is zero. Then we assign zero centrality to a sector with no output.

randomness in how a dollar flows around the economy, we could have labeled our measure random walk closeness. But we wish to pay homage to [14]

Hence, we need to measure distance between nodes in a different way. We propose using the Mean First Passage Time (MFPT) as a metric when dealing with random walk processes [1]. The MFPT from node s to t is the expected number of steps a random walker starting at node s needs to reach node t for the first time:

$$H(s,t) := \sum_{r=1}^{\infty} r \cdot Pr(s \xrightarrow{r} t) \ . \tag{2}$$

Here $Pr(s \xrightarrow{r} t)$ is the probability that it needs exactly r steps before the first arrival. [3] As we are interested in the first visit of the target node, we consider an absorbing random walk, means we never leave node t after we went there. It is appropriate to modify the Markov matrix M by deleting its $t-th$ row and column, resulting in a $(n-1) \times (n-1)$ matrix that we denote by M_{-t}.

The (s,i) element of the matrix

$$((M_{-t})^{r-1})_{si}. \tag{3}$$

gives the probability of starting at s and being at i in $r-1$ steps, without ever having passed through t. Consider a walk of exactly r steps from s that first arrives at t. Its probability is:

$$Pr(s \xrightarrow{r} t) = \sum_{i \neq t} ((M_{-t})^{r-1})_{si} m_{it} \ .$$

Plugging this into equation (2), we find

$$H(s,t) = \sum_{r=1}^{\infty} r \sum_{i \neq t} ((M_{-t})^{r-1})_{si} m_{it} \ .$$

The infinite sum $\sum_{r=1}^{\infty} r(M_{-t})^{r-1} = (I - M_{-t})^{-2}$, where I is the $n-1$ dimensional identity matrix. Being able to make this inversion is the reason for deleting one row and column from the original transition matrix M. Lovasz [11] shows that $(I - M_{-t})$ is invertible as long as there are no absorbing states, whereas $(I - M)$ is not since M is a Markov matrix. So

$$H(s,t) = \sum_{i \neq t} \left((I - M_{-t})^{-2} \right)_{si} m_{it} \ .$$

This can be easily vectorized:

$$H(.,t) = (I - M_{-t})^{-2} \, m_{-t} \ .$$

where $H(.,t)$ is the vector of mean first hitting times for a walk that ends at target t and $m_{-t} = (m_{1t}, ..., m_{t-1,t}, m_{t+1,t}, ..., m_{nt})'$ is the $t-th$ column of M with the element m_{tt} deleted. Further, let e be an $n-1$ dimensional vector of ones. Then $m_{-t} = (I - M_{-t})e$. Hence

$$H(.,t) = (I - M_{-t})^{-1} \, e \ . \tag{4}$$

[3] By convention, $H(t,t) = 0$ since $Pr(t \xrightarrow{r} t) = 0$ for $r \geq 1$.

This equation allows calculation of the MFHT matrix row-by-row with basic matrix operations only. Using Sherman-Morrison formula [7], we can speed up the n matrix inversions further.

In principle, the MFHT is not symmetric, even for undirected graphs. This property reflects the fact that it is much easier to travel from the periphery to the center than it is to go the other way around. Using the natural analogy with closeness centrality, we define random walk centrality as the inverse of the average mean first hitting time to a given node:

$$\mathcal{C}_1(i) = \frac{n}{\sum_{j \in V} H(j, i)} \,. \tag{5}$$

This measure is similar to the one proposed in [14]. Consider a supply shock that occurs with equal probability in any sector. Then random walk centrality is the inverse of the expected number of steps it will take for this shock to be felt in sector i. If this is a low number, then sector i is very sensitive to supply conditions anywhere in the economy. Hence a sector that depends on a wide array of inputs will tend to have a high random walk centrality.

3.2 Counting Centrality

Our second approach is inspired by Newman's random walk betweenness [13]. We modify this concept slightly and generalize it to directed networks with self-loops. Also, this measure is a generalization of betweenness centrality [3]: Betweenness centrality measures how often a certain node lies on a shortest path if one averages over all possible pairs of source and target. A fast algorithm for computing it is developed in [2]. Betweenness centrality is inadequate for networks based on input output matrices since the shortest path concept is not useful in this case; also, betweenness centrality does not allow for self-loops. We build on the concept of random walk betweenness and define a measure that we call counting betweenness. It counts how often a given node is visited on first passage walks, averaged over all pairs of source and target.

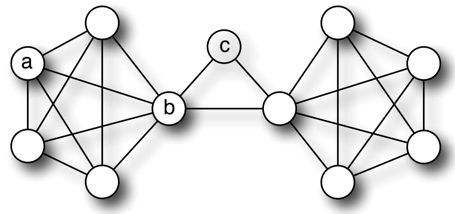
For a source node s and target node $t \neq s$, the probability of being at node $i \neq t$ after r steps is $((M_{-t})^r)_{si}$. Then the probability of going from i to j is m_{ij}. So the probability that a walker uses the edge $i \to j$ immediately after r steps is $((M_{-t})^r)_{sj} m_{ij}$. Summing over r we can calculate how often the walker is expected to use this edge:

$$\begin{aligned} \sum_r ((M_{-t})^r)_{si} m_{ij} &= \quad m_{ij} \textstyle\sum_r ((M_{-t})^r)_{si} \\ &= \quad m_{ij} \, ((I - M_{-t})^{-1})_{si} \\ &=: \quad N_{ij}^{st} \end{aligned}$$

Notice that a walker never uses an edge $i \to j$ if j is not a neighbor of i. The total number of times we go from i to j and back to i is $N_{ij}^{st} + N_{ji}^{st}$. Here we differ from [13], who excludes walks that oscillate and thus counts only the net number of visits. On any walk from s to t, we enter node $i \neq s, t$ as often as we leave it. Hence, on a path from s to t, vertex i is visted $\sum_{j \neq t} (N_{ij}^{st} + N_{ji}^{st})/2$ times. For source s and target t and vertex $i \neq s, t$, we define:

$$N^{st}(i) = \sum_{j \neq t} (N_{ij}^{st} + N_{ji}^{st})/2 \,. \tag{6}$$

A



B

	a	b	c
Shortest-path betweenness	0.2	0.64	0.2
Random Walk betweenness	0.27	0.67	0.33
Counting betweenness	1.93	2.80	1.03
Random Walk Centrality	0.048	0.094	0.044

Figure 1: **A** A toy network, adapted from [13]. **B** Different centrality measures calculated for selected nodes.

Since self loops are completely normal in an input-output network, a random walk can follow the edge $i \to i$, in which case the vertex i is visited twice consecutively. Since it is possible that $i = j \neq t$, we must take care to divide by 2 in all cases.

There are two special cases. If $i = s$, then the walker visits vertex s one extra time

$$N^{st}(s) = \sum_{j \neq t} (N^{st}_{sj} + N^{st}_{js})/2 + 1.$$

Also, if $i = t$, then the walker is absorbed by vertex t the first time it arrives there and

$$N^{st}(t) = 1.$$

The counting betweenness of node i is the average of this quantity across all source-target pairs:

$$\mathcal{C}_2(i) = \frac{\sum_{s \in V} \sum_{t \in (V - \{s\})} N^{st}(i)}{n(n-1)} . \tag{7}$$

Counting betweenness can be used as a micro-foundation for the velocity of money. Consider a dollar of final demand that is spent with equal probability on the output of any sector, and assume that all transactions for intermediate goods must be paid for with cash, not credit. Then the counting betweenness of sector i is the expected number of periods that this dollar will spend there. If it is a high number, then that sector requires many transactions before the money is eventually returned to the household sector as a payment to some factor of production. If each transaction takes a fixed amount of time, then a sector with a high betweenness is a drag on the velocity of money in the economy.

3.3 Toy examples

Before applying our measures to the Input-Output graphs, we study their behavior on small toy examples. Figure 1 shows a graph introduced by Newman [13] to illustrate different concepts of centrality measures. Here, all useful measures should rank nodes b as the most central ones. However, while concepts based on shortest paths do not account for the topologically central position of node a, the Random Walk Betweenness does. Calculating our measures from the last section, we find that in contrast both rank nodes of type a higher than node c. A random walker causes a large amount of traffic within the strongly connected subgraph which dominates the lower traffic over the bridge c. Thus, as we do not average out traffic in opposite directions, this leads to a large counting betweenness of the nodes a.

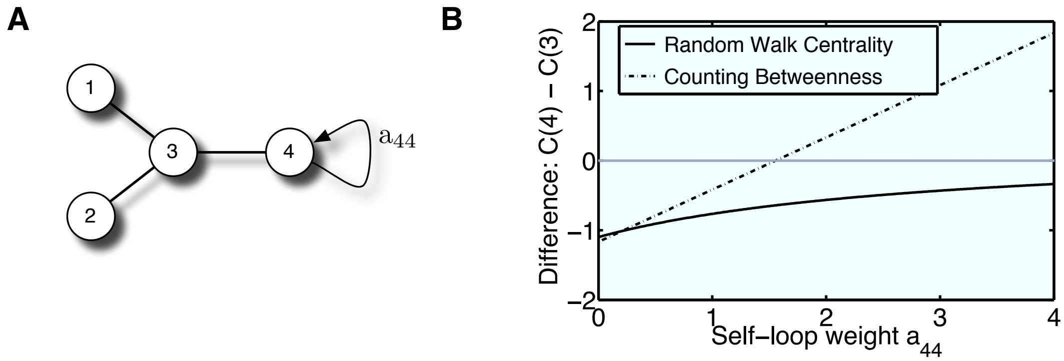


Figure 2: **A** A small network with a self-loop. **B** Centrality measure in

In Figure 2 we plot a small network illustrating the differences between our two centrality approaches regarding the role of self-loops. Depending on the self-loop weight a_{44} either node 3 or 4 has the highest counting betweenness in this network. In contrast, independently of a_{44} node 3 is always most central with respect to random walk centrality.

4 Central Sectors in Modern Economies

Table 1 presents the most central sectors in each economy. Our countries are Argentina, Australia, Austria, Belgium, Brazil, Canada, China, the Czech Republic, Denmark, Finland, France, Germany in 1995 and in 2000, Great Britain, Greece, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Korea, Luxembourg, the Netherlands, New Zealand, Norway, Poland, Portugal, the Russian Federation, Slovakia, South Africa, Spain, Sweden, Switzerland, Turkey, Taiwan, and the United States in 1995 and in 2000. These countries account for more than 85% of world gross domestic product.

It is striking that “Wholesale and retail trade”is most frequently the sector with highest centrality. In many economies, this sector has the highest share of final demand. Still, it is noteworthy that our normalization does not depend upon this fact. For example, in Germany in 2000, this sector accounts for 12% of final demand, but out normalization makes this sector’s entries sum to unity, just like any others. Real estate activities is the second most important sector accounting for 9.6% of final demand, but its random walk centrality is ranked only eighth. One can tentatively conclude that high random walk centrality is actually based upon a rich pattern of output linkages, not on the sector’s absolute importance in the economy.

Counting centrality captures sectors with hig average betweenness and also important self-loops. Focussing on counting centrality reveals the importance of Nokia in Finland and the motor vehicle sector in several advanced industrialized economies. Textiles play an important role in China, Indonesia, and Turkey, showing the importance of that manufacturing sector in a countries with low wages. Finally, it is worth noting that public administration, defence, and compulsory social security is most central in Israel, South Africa, and the United States.

Table 1: Most Central Sectors

Country	Random Walk Centrality	Counting Centrality
arg1997	Food products	Health and social work
aus199899	Wholesale and retail trade	Wholesale and retail trade
aut2000	Wholesale and retail trade	Wholesale and retail trade
bel2000	Wholesale and retail trade	Motor vehicles
bra2000	Wholesale and retail trade	Food products
can2000	Wholesale and retail trade	Motor vehicles
che2001	Wholesale and retail trade	Chemicals excluding pharmaceuticals
chn2000	Construction	Textiles
cze2000	Wholesale and retail trade	Construction
deu1995	Wholesale and retail trade	Motor vehicles
deu2000	Wholesale and retail trade	Motor vehicles
dnk2000	Wholesale and retail trade	Food products
esp2000	Wholesale and retail trade	Construction
fin2000	Wholesale and retail trade	Communication equipment
fra2000	Construction	Motor vehicles
gbr2000	Wholesale and retail trade	Health and social work
grc1999	Wholesale and retail trade	Wholesale and retail trade
hun2000	Wholesale and retail trade	Motor vehicles
idn2000	Wholesale and retail trade	Textiles
ind199899	Land transport	Food products
irl2000	Construction	Office machinery
isr1995	Defence and social security	Health and social work
ita2000	Wholesale and retail trade	Wholesale and retail trade
jpn2000	Other Business Activities	Motor vehicles
kor2000	Construction	Motor vehicles
lux2000	Finance and insurance	Finance and insurance
nld2000	Wholesale and retail trade	Food products
nor2000	Wholesale and retail trade	Food products
nzl200203	Wholesale and retail trade	Food products
pol2000	Wholesale and retail trade	Wholesale and retail trade
prt2000	Wholesale and retail trade	Health and social work
rus2000	Wholesale and retail trade	Food products
svk2000	Wholesale and retail trade	Motor vehicles
swe2000	Other Business Activities	Motor vehicles
tur1998	Food products	Textiles
twn2001	Wholesale and retail trade	Office machinery
usa1995	Wholesale and retail trade	Health and social work
usa2000	Defence and social security	Defence and social security
zaf2000	Defence and social security	Defence and social security

It is impressive to visualize the structure in complicated sets of data by using clustering techniques. A clustering assigns a set of objects into groupings according to a measure of similarity. Our normalized data are of dimension $2209 = 47 * 47$, but our focus on centrality reduces each economy to an element in a 47-dimensional space. Reducing the complex networks to a list of centrality val-

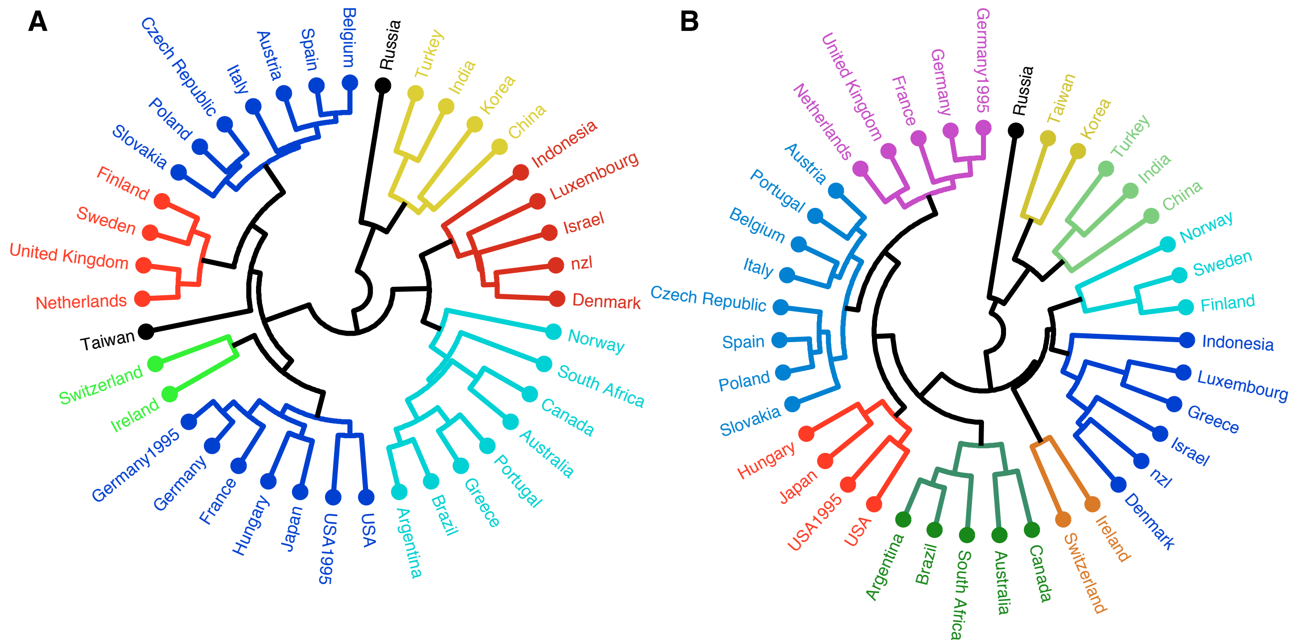


Figure 3: Dendrograms of clusterings according to **A** Random Walk Centrality and **B** Counting Betweenness

ues, we can dramatically compress the relevant information. Even more, we do not want to attach too much importance to the actual centrality numbers themselves. Instead, we are concerned with their rankings. Thus, for us two economies are similar if their Spearman rank correlation of centrality across sectors is high. This also captures the fact that we had to remove sectors without input or output to keep our measures well-defined.

Perhaps the easiest and most commonly used clustering method is hierarchical clustering, for a detailed treatment see e.g. [8]. This iterative algorithm groups economies starting with the most similar ones. Our distance measure is Spearman rank correlation. Figure 2 shows that Belgium and Spain are the two most similar networks; hence, they are the closest two networks. We use complete linkage clustering to draw the rest of the dendrogram with respect to a ranking according to a betweenness measure: Let A and B be two sets; then distance between them is $d(A,B) = max\{d(x,y) : x \in A, y \in B\}$. The clustering algorithm proceeds iteratively by identifying nearest neighbors and showing their distance using branch heights in the dendrogram. When all the initial singletons are linked, the algorithm stops. Cutting the tree at a predefined threshold gives a clustering at the selected precision. For example, at the threshold 0.65, there are three clear clusters in Figure 2: (1) a group of advanced industrial economies ranging from Belgium through the United States; (2) a mixed group of countries where agriculture may be important; and (3) a group of rapidly emerging economies ranging from China through Russia.

Figure 3b shows a clustering based upon the similarity of networks according to counting centrality. Taiwan is grouped quite differently in the two clusterings. According to random walk centrality, it is in the middle of the advanced industrial economies. But in the clustering according to count centrality, it is a close neighbor of Korea, in the "Asian Tigers"sub-group of the emerging economies.

An important reason for this different grouping is that Korea and Taiwan have food products and textiles industries that both have large self-loops. This clustering captures the remnants of the historical development process in which both economies were based on manufacturing sectors just one generation ago.

It is reassuring that the clusterings are stable across the two measures. The groupings are natural; it is appropriate that the American and German economies, each sampled five years apart, are most closely related their former selves. Leontief argued that the stability of input-output relations across time was a good empirical justification for using a fixed-coefficients technology in his original work. These clusterings support his assertion.

Focusing on Random Walk Centrality, we turn briefly to a detailed study of two different pairs of similar economies. Tables 2 and 3 look into the details inherent in the clusterings that arise from our the Random Walk Centrality. The two nearest neighbors are Belgium and Spain. We report the ranks of the ten most central sectors in each country. There are forty-seven sectors in each case, but reporting all would overwhelm the reader. The list of ten sectors shows the level of disaggregation of our data. The main conclusion drawn from Table 2 is that there is a remarkable similarity between the flow of intermediate inputs in each of these economies. Retail trade and construction are notoriously pro-cyclical, and Random Walk Centrality shows that fact clearly.

Table 2: Two Similar Advanced Economies

Rank	Sector in Belgium	Sector in Spain
1	Wholesale and retail trade	Wholesale and retail trade
2	Construction	Construction
3	Other Business Activities	Hotels and restaurants
4	Food products	Other Business Activities
5	Chemicals excluding pharmaceuticals	Food products
6	Hotels and restaurants	Real estate activities
7	Travel agencies	Travel agencies
8	Motor vehicles	Other social services
9	Agriculture	Motor vehicles
10	Health and social work	Agriculture

India and Turkey cluster together, but they are somewhat less similar than Belgium and Spain; in Fig. 2, the length of the branch that brings them together is twice as high as that for Belgium and Spain. Food products, construction, and hotels and restaurants all have high centrality rankings. These rankings seem to indicate that the sectoral composition of business cycles is somewhat different in an emerging economy.

Table 3: Two Similar Emerging Economies

Rank	Sector in India	Sector in Turkey
1	Land transport	Food products
2	Food products	Wholesale and retail trade
3	Agriculture	Construction
4	Construction	Hotels and restaurants
5	Hotels and restaurants	Agriculture
6	Textiles	Finance and insurance
7	Health and social work	Textiles
8	Wholesale and retail trades	Land transport
9	Chemicals excluding pharmaceuticals	Travel agencies
10	Electricity	Machinery, nec

5 Conclusion

We developed two centrality measures suited for measuring the flow of economic activity between sectors in an economy. A node's random walk centrality is the inverse of its mean first hitting time, averaged over all pairs of source and target. In an input-output table a central sector will bear the brunt of a supply shock very quickly. A node's counting centrality measures expected number of times that a commodity passes though before it exits the system as a good for final demand. Again, we average over all pairs of source and target. The main difference between the two measures is that counting centrality captures the effects of self-loops, an important part of economic networks.

We have concentrated on the flow of economic activity across the sectors of the economy. Our normalization was to divide the input-output table by its row sums. Hence we treated sectors that had large volumes of output of intermediate goods in the same way as those that had small volumes. We directed our attention to the flow of economic activities as intermediate outputs before they exited the system for use in final demand.

Taking full advantage of the the consistency of the data across countries, we have given the first hierarchical clusterings of these economic networks. The clusterings were intuitive. They revealed level of development and similarity of the same economy across time. To the best of our knowledge, we have given the first hierarchical clusterings of the production structures of different economies. We anticipate that others will build on this aspect of our work.

Using random walk centrality, we interpret a central node as a sector that it is most immediately affected by a random supply shock. Hence, if one could predict sectoral shocks accurately, one would short equity in a central sector and go long equity in a remote sector during an economic downturn. Using counting centrality, we interpret a central sector as one that is both central and has a strong self-loop. This measure of centrality is related to the velocity of money, and it does not ignore that firms in the same sector actually buy and sell from each other.

There is a lot more work to be done in this area. The theory of networks has flourished in the last two decades, and consistent international data has also become widely available during this time. In the social sciences, network theory arose from theoretical sociology, but it obviously has ready applications to economic data, both real and financial. we expect that our techniques will be useful for analyzing payment networks and other financial systems. Also, our measure of counting betweenness is useful for any network where self-loops are important. If the nodes of a social network describe aggregates such as clubs or teams–not just individuals themselves–then self-loops become an important part of its architecture.

References

[1] B. Bollobás. *Random graphs*. Cambridge Studies in Advanced Mathematics, 2001.

[2] Ulrik Brandes. A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.

[3] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40:31–41, 1977.

[4] L. C. Freeman. Centrality in social networks: Conceptual clarification ı. *Social Networks*, 1(4):215–239, 1979.

[5] L. C. Freeman, S. P. Borgatti, and D. R White. Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks*, 13(2):141–154, 1991.

[6] Lothar Krempel Gancho Ganchev and Margarita Shivergeva. How to view structural change: The case of economic transition in bulgaria. Research Project at the Max-Planck Institute for the Study of Societies: Visualizing Economic Tansition in Bulgaria, July 2001.

[7] G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins University Press, 1996.

[8] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.

[9] Wassily Leontief. *Input-Output Economics (2nd ed.)*. Oxford University Press, New York, 1986.

[10] John B. Long and Charles I. Plosser. Real business cycles. *Journal of Political Economy*, 91:39–69, 1983).

[11] L. Lovász. Random walks on graphs: a survey. *Combinatorics*, 2(80):1–46, 1993.

[12] James McNerney. Network properties of economic-input output networks. International Institute for Applied Systems Analysis Interim Repor IR-09-003, January 2009.

[13] M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, 2005.

[14] Jae Dong Noh and Heiko Rieger. Random walks on complex networks. *Phys. Rev. Lett.*, 92(11):118701, 2004.