

**Vertex centralities in input-output networks reveal the structure of modern economies**Florian Blöchl,<sup>1,\*</sup> Fabian J. Theis,<sup>1,2</sup> Fernando Vega-Redondo,<sup>3,4</sup> and Eric O’N. Fisher<sup>5</sup><sup>1</sup>*Helmholtz Zentrum München, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany*<sup>2</sup>*Institute for Mathematical Sciences, Technische Universität München, Arcisstrasse 21, D-80333 München, Germany*<sup>3</sup>*Department of Economics, European University Institute, Villa San Paolo, Via della Piazzuola 43, I-50133 Firenze, Italy*<sup>4</sup>*Instituto Valenciano de Investigaciones Economicas, E-46020 Valencia, Spain*<sup>5</sup>*Orfalea College of Business, California State Polytechnic University, 1 Grand Avenue, San Luis Obispo, California 93407, USA*

(Received 26 July 2010; revised manuscript received 26 January 2011; published 28 April 2011)

Input-output tables describe the flows of goods and services between the sectors of an economy. These tables can be interpreted as weighted directed networks. At the usual level of aggregation, they contain nodes with strong self-loops and are almost completely connected. We derive two measures of node centrality that are well suited for such networks. Both are based on random walks and have interpretations as the propagation of supply shocks through the economy. Random walk centrality reveals the vertices most immediately affected by a shock. Counting betweenness identifies the nodes where a shock lingers longest. The two measures differ in how they treat self-loops. We apply both to data from a wide set of countries and uncover salient characteristics of the structures of these national economies. We further validate our indices by clustering according to sectors’ centralities. This analysis reveals geographical proximity and similar developmental status.

DOI: [10.1103/PhysRevE.83.046127](https://doi.org/10.1103/PhysRevE.83.046127)

PACS number(s): 89.65.Gh, 89.75.–k, 05.40.Fb

**I. INTRODUCTION**

Within a few weeks of the onset of the financial crisis in 2008, the world economy had plunged into a severe global recession. The volume of international trade contracted sharply, and the world economy did not grow in 2009 for the first time since World War II. Many governments reacted with programs to mitigate the effects of the global downturn on their local economies. The United States spent \$3 billion on the Car Allowance Rebate System. Germany spent an even larger fraction of its national economy (1.5 billion) for a car scrappage program. What effect did these programs have? How did the supply of new cars work its way through the rest of the local economy?

Input-output analysis was designed to explore this kind of effect [1,2]. An *input-output table* is the matrix of the sales of goods and services between the different sectors of an economy. A sector is a fairly coarse level of aggregation; an industry is composed of many firms making an identical product, and a sector is composed of several industries making similar products. “Agriculture” and “Pharmaceuticals” are two typical sectors.

The techniques of input-output analysis have had ready applications in economic planning. It is alleged that Leontief [2] developed aspects of input-output analysis during the Second World War partly as an attempt to help identify strategic weaknesses in the German economy. Ranking the influences of single sectors on national economic activity allows the identification of “key” sectors. For example, there has been much discussion about firms that are “too big to fail,” and there was an implicit understanding that the bailout of General Motors was necessary because of the importance of the automotive sector in the American economy.

To formalize these intuitive ideas, a deeper understanding of the structures of national economies seems to be warranted.

Any national economy is a complex system in which many agents of different sizes interact by buying and selling goods and services. Schweitzer *et al.* [3] suggest that an understanding of these interactions on a systemic level may be achieved by analyzing the underlying complex networks. During the last decade, network analysis has been applied successfully in physics, biology, and the social sciences [4–7].

The literature on economic networks is growing rapidly. Several authors have studied international trade networks. The early work used binary approaches [8,9], but it soon became evident that trade ought to be analyzed as weighted graph [10–12]. Interpreting the gross domestic product (GDP) as a country’s fitness, Garlaschelli and Loffredo [13] proposed a model reproducing the topology of bilateral trade. A gravity model has been used to understand weighted trade networks [10]. Furthermore, innovation networks [14], the “product space” [15], ownership networks [16], and connections between banks [17] have been studied. Recently, Battistion *et al.* [18] analyzed investment stocks, both at the level of firms and aggregated to the level of regions. Lorenz *et al.* [19] developed a general model to understand systemic risk; they investigated cascading failures within systems of many interacting (networked) agents like firms, banks, and funds. Finally, Grassi [20] studied information flow across board members of different firms; she focused on node centralities.

In fact, it is natural to interpret an input-output table as a network. Each sector corresponds to a vertex, and the flow of economic activity from one sector to another constitutes a weighted directed edge. In complex network theory, identifying key sectors and ranking the sectors’ roles in an economy is the task of applying an appropriate measure of node centrality to this input-output graph.

Vertex centrality measures have been studied extensively for quite some time. Freeman introduced the notion of centrality in a graph [21]; he defined the *betweenness centrality* of a node as the average number of shortest paths between pairs of other nodes that pass through it. *Flow betweenness*

\*florian.bloechl@helmholtz-muenchen.de

is based on the maximum capacity of flows between nodes. It also includes contributions from some nongeodesic paths [22]. Another approach, *closeness centrality*, is commonly defined as the inverse of the mean geodesic distance from all nodes to a given one [23]. All these measures require flows in the network to know an ideal route from each source to each target, either order to find a shortest path or to maximize flow. Addressing this potential deficiency, Newman defined *random walk betweenness* [24]. He averages effective visits over all possible random walks in a network.

Three properties of input-output graphs make it hard to apply current centrality measures. First, at the usual level of aggregation, these networks are dense, typically almost completely connected. Thus applying measures based on shortest paths makes little sense. As the topology is nearly trivial, one needs to analyze edge weights. Second, they are directed; for example, in the United States in 2000, \$13.5 billion of rubber and plastic products was used in the production of motor vehicles, but only \$53 million of the output of the motor vehicle sector was used in the production of rubber and plastic products. Third, self-loops play a central role; in the same case, more than 60% of the total output of the cars sector was used as its own input. Some authors, including White and Borgatti [25], have extended centrality concepts to the directed case, but to the best of our knowledge, no one until now has examined node centralities that incorporate self-loops. We derive two measures that are suited for such networks. Both rely on random walks and each has an economic interpretation.

The rest of the paper is structured as follows. The next section provides the basic concepts. Section III derives two centrality measures and shows their relation to economic theory. We contrast our two approaches using a small example. Section IV shows our empirical results using input-output data from a wide range of countries. The proposed measures reveal important aspects of different national economies. Moreover, the consistency of the data allows us to compare nodes' centralities across countries in an intuitive way. Finally, we present some brief conclusions and suggestions for future research. Implementations of the measures, the data, and the results are available at <http://hmgu.de/cmb/ionetworks>.

## II. BASIC CONCEPTS

A graph  $G = (V, E)$  consists of a set of vertices  $V$  and a set of edges  $E \subset V \times V$ . In our case, each edge  $(i, j) \in E$  is directed and assigned a non-negative real weight  $a_{ij}$ . By definition, the graph may contain self-loops. The number of vertices is denoted  $n$ . We consider strongly connected graphs only; for any pair of nodes, there exists a directed path connecting them.

The graph can be represented by its  $n \times n$  adjacency matrix  $A = (a_{ij})$ , where the element  $(i, j)$  represents the weight  $a_{ij}$  of the edge from node  $i$  to node  $j$ . To keep notation simple, we name the vertices by natural numbers, and we can identify them with according indices in the adjacency matrix. Missing edges correspond to zero weights in the adjacency matrix. Then, the out strength of node  $i$  is  $k_i = \sum_{j=1}^n a_{ij}$ . We denote the set of out neighbors of  $i$   $N(i) = \{j | (i, j) \in E\}$ .

### A. Input-output networks

An input-output table  $A$  is an adjacency matrix of a network whose vertices are the sectors of an economy. Its edges quantify the flow of economic activity between sectors. We focus on the table of *intermediate inputs*. It records only sales of goods and services by firms to other firms that are directly consumed or used up as inputs in the production process. It is not a closed system; the row and column sums are not equal. In national accounts, the total value of the gross output of a sector also includes sales for final demand: consumption, investment, government purchases, and net exports. The total value of gross inputs into a sector also includes payments to the factors of production: gross operating surplus, compensation to employees, and indirect business taxes [1].

### B. Random walks

The movement of goods between the sectors of an economy is best modeled as a random walk [26]. In graph theory, a random walker starts out at a given position and repeatedly chooses an edge incident to the current position [27]. These choices are made according to a probability distribution determined by the edge weights. The random walker proceeds for an arbitrarily long time or until a prescribed goal is reached.

An input-output table keeps track of the goods circulating through an economy, consisting of the outputs of a large number of firms in each sector. Hence, each entry is the statistical aggregation of many individual sales. We are interested in the transition probabilities of outputs produced by a sector. These can be obtained by normalizing the input-output matrix by its row sums. Hence in the following we work with the transition matrix

$$M = K^{-1}A, \quad (1)$$

where  $K$  is the diagonal matrix of the out strengths  $k_i$ .

## III. TWO CENTRALITY MEASURES

This section derives two centrality measures that are suited for weighted directed networks with self-loops. First, we explain their economic foundations. Second, we discuss the difference between divisible and indivisible shocks. Third, we define and then relate our measures to others commonly used. Fourth, we give a small example that contrasts them.

### A. Economic intuition

Following the ideas of Fischer Black [28], we design both our centrality measures to quantify the response of sectors to an *economic shock*. Such a shock is a change in an exogenous variable that has repercussions on the endogenous variables under analysis [1]. In input-output accounts, prices, technologies, firms, the distribution of profits, government policy, and vector of final demands are exogenous, and the flows of commodities and corresponding payments between sectors are endogenous. Fischer Black hypothesized that the business cycle might arise because of the propagation of such shocks between the sectors of an economy [28]. Long and

Plosser developed an elegant analysis of the US economy based on this idea [29].

We trace supply shocks as they flow as intermediate inputs through the business sectors of an economy. Their random journeys end at the sector from which the extra output eventually satisfies final demand, which we interpret as the target of some random walk. Consider an extra dollar of production in the car sector—perhaps as a result of a government program—and the target “food products.” The initial output will be sold randomly to another sector, according to the pattern of sales in the input-output table. The original dollar of extra revenue will be paid to capital, labor, or indirect business taxes in “motor vehicles.” The supply shock becomes an input into some sector, and it will increase economic activity there by \$1, akin to the conservation of current in a circuit. The new output, again, will be sold to some sector. Eventually this process will hit the target “food products,” where the extra dollar of output exits the system to satisfy the final demand. Averaging over all initial shocks or over all pairs of shocks and targets, we define a node’s centrality by how quickly or how frequently it is visited during this process. We employ these two different measures, as we will see later that they allow us to emphasize quite different economic aspects.

Every economic transaction consists of a real and a monetary counterpart; thus when keeping track of the flow of goods and services from a source to a destination, at the same time we monitor the flow of a dollar in payments from the destination back to the source.

**B. Divisible and indivisible shocks**

We have intentionally used the metaphor of a single dollar traveling in the network because both our measures rely on the properties of random walks of *indivisible* supply shocks. We impose this restriction for two reasons. First, the economic assumptions inherent in input-output accounting require that the shocks are “not too large.” A large shock would change the very structure of a national economy. Leontief accounting assumes fixed proportions of all inputs into any productive process; input-output tables change only very slowly across time, and they are benchmarked only about once each decade. Thus a small shock corresponds to a change in supply that can be accommodated by the current structure of the national economy. If the shock is sufficiently small, then it is also natural to postulate that its effects are indivisible (or, at least, not freely divisible).

Second, centrality measures, as reviewed in Sec. I, arise from two quite different structural intuitions. First, one may view a node as important in a network if it is close to all other nodes. In this sense, a shock arrives quickly and easily at a central node. Second, a node’s centrality can be based on the degree to which it lies between other nodes; this concept shows how the node mediates between all others. Focusing on how quickly or how frequently a node is visited during the random walk of an indivisible shock allows us to explore two measures that naturally arise from both these graph-theoretic intuitions and apply them readily to our sectoral data on national economies.

Let us now discuss briefly the case of divisible shocks. First, one would have to specify exactly how a shock splits

up for each transaction. A generic approach is to divide the remaining fraction of the shock at each node according to the probabilities inherent in the transition matrix **M**. In this case, studying average properties of walks over source-target pairs makes little sense, since the shock smears out immediately over the whole graph in a densely connected input-output network. Instead, one could analyze the following effect. Suppose a supply shock starts at some sector. Then each transaction splits it up, and its fractional effects accumulate at all sectors of the economy. After some time, these will reach a steady state, since the system conserves the absolute size of the shock. In the long run, this distribution is independent of the initial shock. The proportion of the shock present at any sector in the steady state is therefore also an adequate measure of its centrality. But then the proportion of a divisible shock in a particular sector is the same as the likelihood of finding an indivisible shock there. The frequencies that nodes are visited by an indivisible shock can therefore be understood as a proxy for the steady-state distribution of a divisible one. However, due to the immediate smearing-out over the whole network, divisible shocks do not provide us with a centrality measure arising from the intuition of closeness to other nodes.

**C. Random walk centrality**

Freeman’s closeness centrality [23] is widely used in social network analysis. It is commonly defined as the inverse of the mean geodesic distance from all nodes to a given one. Again, the shortest paths make little sense in densely connected networks like input-output graphs. Moreover, they completely ignore self-loops.

To generalize the concept of closeness, the distance between nodes has to be measured in a different way. We propose using the mean first-passage time (MFPT). This distance is the measure of choice when dealing with random walk processes [27]. The MFPT  $H(s,t)$  from node  $s$  to node  $t$  is the expected number of steps a random walker who starts at  $s$  needs to reach  $t$  for the first time:

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

Here  $P(s \xrightarrow{r} t)$  is the probability that it takes exactly  $r$  steps before the first arrival at  $t$ . Note that  $H(t,t) = 0$  since  $P(t \xrightarrow{r} t) = 0$  for  $r \geq 1$ . The MFPT is not symmetric, even for undirected graphs. This property reflects the fact that it is much more probable to travel from the periphery to the central nodes of a graph than to go the other way around.

We are interested in the first visit of the target node  $t$ . For calculations we can consider an absorbing random walk that, by definition, never leaves node  $t$  once it is reached. It is thus appropriate to modify the transition matrix  $M$  by deleting its  $t$ th row and column. This  $(n - 1) \times (n - 1)$  matrix we denote  $M_{-t}$ .

The element  $(s,i)$  of the matrix  $(M_{-t})^{r-1}$  gives the probability of starting at  $s$  and being at  $i$  in  $r - 1$  steps, without ever having passed through the target node  $t$ . Consider a walk of exactly  $r$  steps from  $s$  that first arrives at  $t$ . Its probability is

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

Plugging this into Eq. (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 over  $r$  is essentially the sum of the geometric series for matrices

$$\sum_{r=1}^{\infty} r (M_{-t})^{r-1} = (I - M_{-t})^{-2}, \quad (3)$$

where  $I$  is the  $n - 1$ -dimensional identity matrix. Making this inversion is the reason for having deleted one row and column from the original transition matrix  $M$ . Lovasz [30] shows that  $(I - M_{-t})$  is invertible as long as there are no absorbing states, whereas  $(I - M)$  is not. So

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

For fast calculation, this can be easily vectorized as  $H(.,t) = (I - M_{-t})^{-2} m_{-t}$ . Here  $H(.,t)$  is the vector of MFPTs for a walk that ends at target  $t$  and  $m_{-t} = (m_{1t}, \dots, m_{t-1,t}, m_{t+1,t}, \dots, m_{nt})'$  is the  $t$ th column of  $M$  with the element  $m_{tt}$  deleted. Further, let  $e$  be an  $n - 1$ -dimensional vector of 1s. Then  $m_{-t} = (I - M_{-t})e$ . Hence

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

This equation allows calculation of the MFPT matrix row by row with basic matrix operations only. Using the Sherman-Morrison formula [31], we can speed up the  $n$  matrix inversions further.

Using the natural analogy with closeness centrality, we define *random walk centrality* as the inverse of the average MFPT to a given node:

$$C_{rw}(i) = \frac{n}{\sum_{j \in V} H(j,i)}. \quad (5)$$

This measure is essentially proposed in Ref. [32]. Random walk centrality incorporates self-loops indirectly because they slow down the traffic between other nodes.

The economic interpretation of this measure is straightforward. Consider a supply shock that occurs with equal probability in any sector. Then a high random walk centrality of a sector means that it is very sensitive to supply conditions anywhere in the economy. Hence, if one could predict sectoral shocks accurately, one would short equity in a central sector and go long on equity in a remote sector during an economic downturn.

#### D. Counting betweenness

Our second approach is inspired by Newman's random walk betweenness [24]. We modify his concept slightly and generalize it to directed networks with self-loops. The proposed measure, denoted as counting betweenness, keeps track of how often a given node is visited on first-passage walks, averaged over all source-target pairs.

For source node  $s$  and target  $t \neq s$ , 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  $(i,j)$  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:

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

Note that a walker never uses an edge  $(i,j)$  if  $j$  is not a neighbor of  $i$ , since the according transition probability is 0. 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 Ref. [24], which 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 visited  $\sum_{j \neq t} (N_{ij}^{st} + N_{ji}^{st})/2$  times. For source  $s$ , target  $t$ , and vertex  $i \neq s,t$ , we define

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

We allow for self-loops, hence a random walker may follow the edge  $(i,i)$ , in which case the vertex  $i$  is visited twice consecutively. Since it is possible that  $i = j \neq t$ , we have to divide by 2 in all cases.

There are two special cases. If  $i = s$ , then the walker visits node  $s$  one extra time when it starts

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

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

$$N^{st}(t) = 1. \quad (7)$$

We define the *counting betweenness* of node  $i$  as the average of this quantity across all source-target pairs:

$$C_c(i) = \frac{\sum_{s \in V} \sum_{t \in (V - \{s\})} N^{st}(i)}{n(n-1)}. \quad (8)$$

Counting betweenness can be used as a microfoundation 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 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 counting betweenness is a drag on the velocity of money in the economy.

#### E. Illustrative examples

Before applying our measures to actual data, we demonstrate their behavior in small artificial examples. Figure 1(b) shows a graph introduced by Newman [24] to illustrate different concepts of centrality. Here, all useful measures should obviously rank nodes of type  $b$  most central. While concepts based on shortest paths do not account

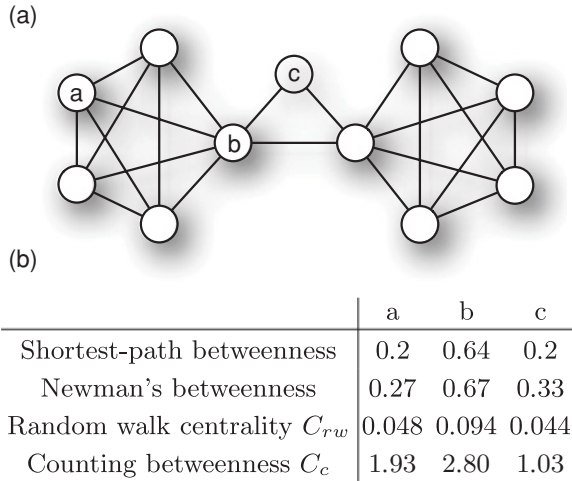


FIG. 1. (a) This network is taken from Ref. [24]. (b) We contrast centrality measures calculated for selected nodes. Even though  $c$  is topologically central, our measures do not rank it highly, in contrast to Newman's betweenness. Instead, they focus on how quickly or how frequently traffic within the network reaches a node. In a graph with two completely connected subcomponents, a slightly remote bridge-like node is not crossed over frequently.

for the topologically central position of node  $c$ , Newman's betweenness gives a high centrality to  $c$ . In contrast, our measures both rank nodes of type  $a$  higher than node  $c$ . A random walk spends a lot of time within the fully connected subgraph on the left and seldom crosses over the bridge-like node  $c$ . The former is why counting betweenness ranks node  $a$  highly, and the latter is why random walk centrality gives it a high ranking.

Figure 2(a) shows a small network that illustrates the differences between our two measures. It emphasizes the role of a self-loop. Depending on the self-loop weight  $a_{44}$  attached to node 4, either node 3 or node 4 has the highest counting betweenness. In contrast, random walk centrality ranks node 3 highest, no matter what the value of  $a_{44}$  is. Counting betweenness strongly emphasizes the importance

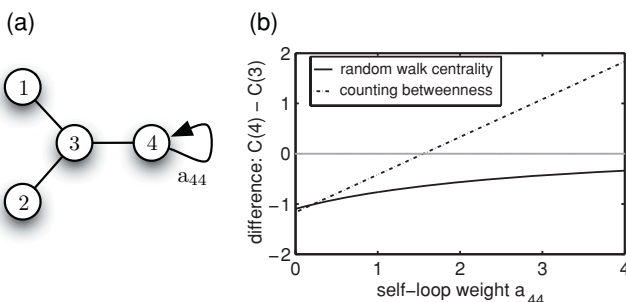


FIG. 2. (a) This small network illustrates the importance of a self-loop. (b) The difference between the centrality of node 4 and that of node 3 as a function of the self-loop weight  $a_{44}$ . All other links have unit weight. Random walk centrality always ranks node 3 highest. Counting betweenness ranks node 4 higher when  $a_{44}$  exceeds a threshold near 1.6. If the self-loop has a large weight, it takes a long time before a random walk leaves node 4 and enters the rest of the network.

of self-loops, which are considered only indirectly by random walk centrality.

IV. CENTRAL SECTORS IN MODERN ECONOMIES

Our data are the input-output accounts from the STAN database at the Organization for Economic Cooperation and Development (OECD), which are available at <http://www.oecd.org/sti/inputoutput/>. They consist of 47 sectors and are benchmarked for 37 countries near the year 2000. Each country's input-output table is one input-output graph. These countries account for more than 85% of the world's GDP.

The data are consistent on three important dimensions. First, they are designed to be consistent across countries. Second, they are consistent with macroeconomic accounts; indeed, they maintain the national income accounting identities. Third, they are consistent across time; so we can compare Germany and the United States to themselves in two different benchmark years. The input-output accounts are reported in local currencies, but we have no need to use exchange rates or GDP deflators because we are only considering the unit-free transition matrices.

Some countries have sectors with no input or 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. Such sectors hinder the matrix inversion in Eq. (3). We therefore assign zero centrality to these nodes and remove them from the adjacency matrix.

A. Results for individual countries

Table I reports each country's most central sector with respect to our two measures. The complete results are available at <http://hmg.u.de/cmb/ionetworks>. It is striking that "wholesale and retail trade" is most frequently the sector with the highest random walk centrality. In many economies, this sector has the highest share of final demand. Still, it is noteworthy that our normalization does not depend on this fact. For example, in Germany in 2000, this sector accounts for 12% of the final demand. "Real estate activities" is the second most important sector, accounting for 9.6% of the final demand, but its random walk centrality is ranked only eighth.

Counting betweenness reveals the importance of Nokia in Finland and the "motor vehicles" sector in several advanced industrialized economies. Textiles play an important role in China, Indonesia, and Turkey, showing the significance of that manufacturing sector in countries with low wages. "Finance and insurance" is most central for Luxembourg. Finally, we note that "public administration, defense, and compulsory social security" is most central in Israel, South Africa, and the United States in 2000.

B. Comparison of different countries

The consistency of the data across countries allows us to immediately compare the centralities of sectors over different countries. We use a clustering technique to visualize our results. A clustering assigns a set of objects into groups according to some measure of similarity. The adjacency

TABLE I. The most central sectors in the economies benchmarked by the OECD.

Country	Random walk centrality	Counting betweenness
Argentina	Food products	Health and social work
Australia	Wholesale and retail trade	Wholesale and retail trade
Austria	Wholesale and retail trade	Wholesale and retail trade
Belgium	Wholesale and retail trade	Motor vehicles
Brazil	Wholesale and retail trade	Food products
Canada	Wholesale and retail trade	Motor vehicles
China	Construction	Textiles
Czech Republic	Wholesale and retail trade	Construction
Denmark	Wholesale and retail trade	Food products
Finland	Wholesale and retail trade	Communication equipment
France	Construction	Motor vehicles
Germany 1995	Wholesale and retail trade	Motor vehicles
Germany 2000	Wholesale and retail trade	Motor vehicles
Great Britain	Wholesale and retail trade	Health and social work
Greece	Wholesale and retail trade	Wholesale and retail trade
Hungary	Wholesale and retail trade	Motor vehicles
Indonesia	Wholesale and retail trade	Textiles
India	Land transport	Food products
Ireland	Construction	Office machinery
Israel	Public administration and defense, social security	Health and social work
Italy	Wholesale and retail trade	Wholesale and retail trade
Japan	Other business activities	Motor vehicles
Korea	Construction	Motor vehicles
Luxembourg	Finance and insurance	Finance and insurance
Netherlands	Wholesale and retail trade	Food products
Norway	Wholesale and retail trade	Food products
New Zealand	Wholesale and retail trade	Food products
Poland	Wholesale and retail trade	Wholesale and retail trade
Portugal	Wholesale and retail trade	Health and social work
Russia	Wholesale and retail trade	Food products
Slovakia	Wholesale and retail trade	Motor vehicles
South Africa	Public administration and defense, social security	Public administration and defense, social security
Spain	Wholesale and retail trade	Construction
Sweden	Other business activities	Motor vehicles
Switzerland	Wholesale and retail trade	Chemicals
Turkey	Food products	Textiles
Taiwan	Wholesale and retail trade	Office machinery
USA 1995	Wholesale and retail trade	Health and social work
USA 2000	Public administration and defense, social security	Public administration and defense, social security

matrices are of dimension  $2209 = 47 \times 47$ , but our focus on centrality reduces each economy to a vector of length 47. Reducing the complex networks to a list of centrality values, we compress y the relevant information dramatically. Moreover, we do not want to attach too much importance to the actual centrality numbers themselves, since we removed sectors without output in some countries. Instead, we are concerned with rankings. Thus, for us two economies are similar if their Spearman rank correlation of centralities across the sectors is high.

An easy and commonly used clustering technique is hierarchical clustering; Hastie [33] gives a good introduction. This iterative algorithm groups economies starting with the closest pair. In Fig. 3(a), Belgium and Spain are the two most similar countries; hence, they are on the lowest-linked branches. Again, by similar, we mean that the Spearman rank correlation of centralities across the sectors is high. We use

complete linkage clustering to complete the dendrogram. This method defines the distance between two sets  $X$  and  $Y$  as the maximum of the distances between any element in  $X$  and any element in  $Y$ . The clustering algorithm proceeds iteratively by identifying nearest neighbors and showing the distance between them using branch heights. When all the initial singletons are linked, the algorithm stops.

Cutting the tree at a predefined threshold gives a clustering at the selected precision. At the threshold 0.65, we find three clusters in Fig. 3(a): (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 3(b) shows a clustering of economies based on the similarity according to counting betweenness. Note that Taiwan is grouped quite differently in the two dendrograms.

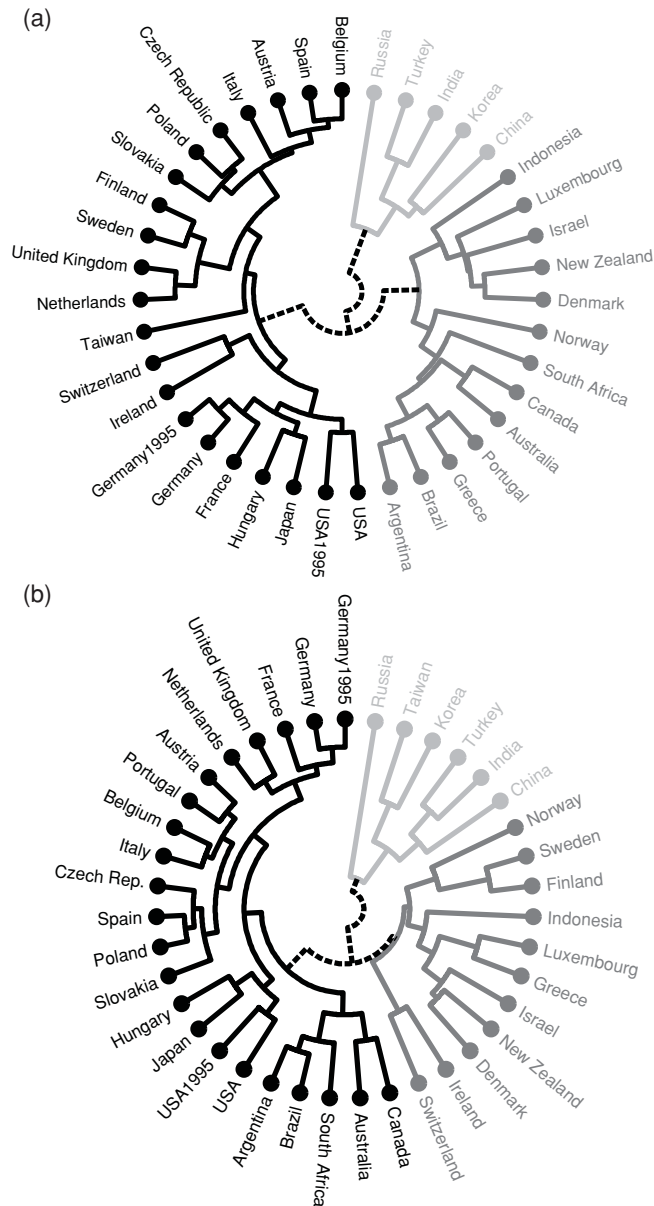


FIG. 3. (a) A hierarchical clustering of countries according to random walk centrality. Gray-scale colors indicate the three important clusters: (1) the industrial countries from Belgium through the United States (black); (2) a mixed group from Argentina through Indonesia, where agriculture and primary products are important (dark gray); and (3) a group of emerging economies from China through Russia (light gray). (b) Clusterings according to counting betweenness. The clusterings according to the two measures are largely stable.

According to random walk centrality, it is in the middle of the advanced industrial economies. But in the clustering according to counting betweenness, it is a close neighbor of Korea, in the “Asian Tigers” subgroup of the emerging economies. An important reason for this difference is that Korea and Taiwan have food products and textiles sectors, both of which have strong self-loops. The clusterings capture the remnants of the historical development process in which both economies were based on manufacturing sectors just one generation ago.

TABLE II. Two advanced economies that are similar in their nodes’ rankings according to random walk centrality.

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	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

It is reassuring that the clusterings are, in large part, stable across the two measures. The groupings are natural; it is appropriate that the American and German economies, each sampled 5 years apart, are most closely related to 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 [2]. These clusterings support his assertion.

C. Two detailed comparisons

Focusing on random walk centrality, we turn briefly to a detailed study of two different pairs of similar economies. Tables II and III examine the details inherent in the sector’s rankings that arise from that measure.

The two nearest neighbors in Fig. 3(a) are Belgium and Spain. Both are advanced economies. Table II reports the 10 most central sectors in each country. There is a remarkable similarity between the flow of intermediate inputs in these economies. The most central sectors in both countries are “retail trade” and “construction.” These sectors are notoriously procyclical, and random walk centrality shows that fact clearly.

India and Turkey are two developing countries that cluster together. This pair is somewhat less similar than Belgium and Spain; in Fig. 3(b), 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

TABLE III. Two emerging economies that are similar in their nodes’ rankings according to random walk centrality.

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 trade	Land transport
9	Chemicals	Travel agencies
10	Production	Machinery and equipment

that the sectoral composition of business cycles is somewhat different in an emerging economy.

## V. CONCLUSION

We have described two vertex centrality measures that are based on random walks. A node's random walk centrality is the inverse of the mean number of steps it takes to reach it, averaged over all starting nodes. Counting betweenness measures the expected number of times that a random walk passes a certain node before it reaches its target, averaged over all pairs of sources and targets. Both measures allow the analysis of weighted directed networks with self-loops. The need for such measures arises from interpreting economic questions within a graph-theoretic framework. We expect that our techniques will be useful for analyzing payment networks and other financial systems. Moreover, any coarsely grained network—such as one describing clubs or teams, not just individuals themselves—will have important self-loops. Our measures will serve well to describe this kind of network architecture. We agree with Estrada *et al.* [34] that there is no best measure of centrality, and we followed their advice and developed two measures that are based on economic theory. We verified our approaches with the application to real complex networks.

We have directed our attention to the flow of economic activity as intermediate inputs before they exited the system

for use in final demand. Our measures identify a central node as a sector that is affected most immediately or most strongly by a random supply shock.

Applying these measures to OECD data has revealed important aspects of different national economies. We have taken full advantage of the consistency of the data across countries and given hierarchical clusterings of the nodes' rankings in these networks. The clusterings were intuitive, grouping countries with similar levels of development.

There is a lot more work to be done in this area. The theory of networks has flourished in the last decade, and consistent international data have also become widely available during this time. These data have a time dimension, and one may begin to study the temporal evolution of economic networks. This may well enable researchers to connect generative models of networks with observations from the real world.

## ACKNOWLEDGMENTS

We thank Carsten Marr and Dominik Wittmann for critical proofreading and the anonymous reviewer for his valuable comments. Eric Fisher thanks the ETH in Zurich and the CES in Munich for the hospitality that allowed this work to be completed. This work was partially supported by the Helmholtz Association (Project CoReNe) and the Federal Ministry of Education and Research (BMBF) through its MedSys initiative (Project SysMBo).

- 
- [1] T. ten Raa, *The Economics of Input-Output Analysis* (Cambridge Books, Cambridge University Press, 2006).
  - [2] W. Leontief, *Input-Output Economics*, 2nd ed. (Oxford University Press, New York, 1986).
  - [3] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, *Science* **325**, 422 (2009).
  - [4] F. Vega-Redondo, *Complex Social Networks* (Cambridge University Press, Cambridge, 2007).
  - [5] A.-L. Barabási and R. Albert, *Rev. Mod. Phys.* **74**, 47 (2002).
  - [6] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
  - [7] S. N. Dorogovtsev and J. F. Mendes, *Evolution of Networks* (Oxford University Press, New York, 2003).
  - [8] D. Garlaschelli and M. I. Loffredo, *Physica A* **355**, 138 (2005).
  - [9] M. A. Serrano and M. Boguñá, *Phys. Rev. E* **68**, 015101 (2003).
  - [10] B. K. G. Mukherjee, J. Saramaki, K. Kaski, and S. S. Manna, *J. Stat. Mech.* (2008) P02002.
  - [11] G. Fagiolo, J. Reyes, and S. Schiavo, *Phys. Rev. E* **79**, 036115 (2009).
  - [12] T. Baskaran, F. Blöchl, T. Brück, and F. J. Theis, *Int. Rev. Econ. Finance* **20**, 135 (2010).
  - [13] D. Garlaschelli and M. I. Loffredo, *Phys. Rev. Lett.* **93**, 188701 (2004).
  - [14] M. König, S. Battiston, M. Napoletano, and F. Schweitzer, *Networks Heterogen. Media* **3**, 201 (2008).
  - [15] C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann, *Science* **317**, 482 (2007).
  - [16] J. B. Glattfelder and S. Battiston, *Phys. Rev. E* **80**, 036104 (2009).
  - [17] G. Iori, G. De Masi, O. V. Precup, G. Gabbi, and G. Caldarelli, *J. Econ. Dyn. Control* **32**, 259 (2008).
  - [18] S. Battiston, J. Rodrigues, and H. Zeytinoglu, *Adv. Complex Syst.* **10**, 29 (2007).
  - [19] J. Lorenz, S. Battiston, and F. Schweitzer, *Eur. Phys. J. B* **71**, 441 (2009).
  - [20] R. Grassi, *Physica A* **389**, 2455 (2010).
  - [21] L. C. Freeman, *Sociometry* **40**, 31 (1977).
  - [22] L. C. Freeman, S. P. Borgatti, and D. R. White, *Social Networks* **13**, 141 (1991).
  - [23] L. C. Freeman, *Social Networks* **1**, 215 (1979).
  - [24] M. E. J. Newman, *Social Networks* **27**, 39 (2005).
  - [25] D. R. White and S. P. Borgatti, *Social Networks* **16**, 335 (1994).
  - [26] S. P. Borgatti, *Social Networks* **27**, 55 (2005).
  - [27] B. Bollobás, *Random Graphs, Cambridge Studies in Advanced Mathematics* (Cambridge University Press, Cambridge, 2001).
  - [28] F. Black, *Business Cycles and Equilibrium* (Basil Blackwell, New York, 1987).
  - [29] J. B. Long and C. I. Plosser, *J. Polit. Econ.* **91**, 39 (1983).
  - [30] L. Lovász, *Combinatorics, Paul Erdős is Eighty* **2**, 1 (1993).
  - [31] G. H. Golub and C. F. Van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, MD, 1996).
  - [32] J. D. Noh and H. Rieger, *Phys. Rev. Lett.* **92**, 118701 (2004).
  - [33] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer, New York, 2001).
  - [34] E. Estrada, D. J. Higham, and N. Hatano, *Physica A* **388**, 764 (2009).