Chapter 1

Introduction

Bryan S. Graham[a] **and Áureo de Paula**[b]
[a]*Department of Economics, University of California - Berkeley, Berkeley, CA, United States,*
[b]*University College London, CeMMAP, IFS and CEPR, London, United Kingdom*

Contents

In this chapter we provide the foundational vocabulary for discussing, describing and summarizing network data like that shown in Fig. 1.1. This figure depicts buyer–supplier relationships among publicly traded firms in the United States. Each dot (*node* or *vertex*) in the figure corresponds to a firm. If a firm, say, United Technologies Corporation, supplies inputs to another firm, say Boeing Corporation, then there exists a *directed edge* (also referred to as *link* or *tie*) •—• from United Technologies to Boeing. The supplying firm (left node) is called the *tail* of the edge, while the buying firm (right node) is its *head*. The set of all such supplier–buyer relationships forms the graph $G(\mathcal{V}, \mathcal{E})$, a directed network or *digraph* defined on $N = |\mathcal{V}|$ vertices or agents (here publicly traded firms). The set $\mathcal{V} = \{1, \ldots, N\}$ includes all agents (firms) in the network and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ the set of all directed links (supplier–buyer relationships) among them.[1],[2] The number of nodes N is sometimes referred to as the *order* of the digraph and $|\mathcal{E}|$, its *size*.[3]

[1] Here $\mathcal{U} \times \mathcal{V}$ denotes the Cartesian product of the set $\mathcal{U}$ and $\mathcal{V}$ (i.e., $\mathcal{U} \times \mathcal{V} = \{(u, v) : u \in \mathcal{U}, v \in \mathcal{V}\}$).

[2] The buyer-supplier network depicted in Fig. 1.1 was constructed using information on large customers disclosed by firms when filing with the United States Securities and Exchange Commission (SEC). Statement of Financial Accounting Standards (SFAS) regulation 131, in effect since 1998, requires firms to report sales to customers which account for 10 percent or more of all firm sales in a given year. Prior to 1998, SFAS regulation 14 imposed similar requirements. Firm-reported large customers are included in the Compustat – Capital IQ customer segments file. Cohen and Frazzini (2008) and Atalay et al. (2011) also construct Supplier–Buyer networks from Compustat data.

[3] In what follows, depending on the context, we will refer synonymously to (di)graphs or (directed) networks.

The Econometric Analysis of Network Data. https://doi.org/10.1016/B978-0-12-811771-2.00007-9

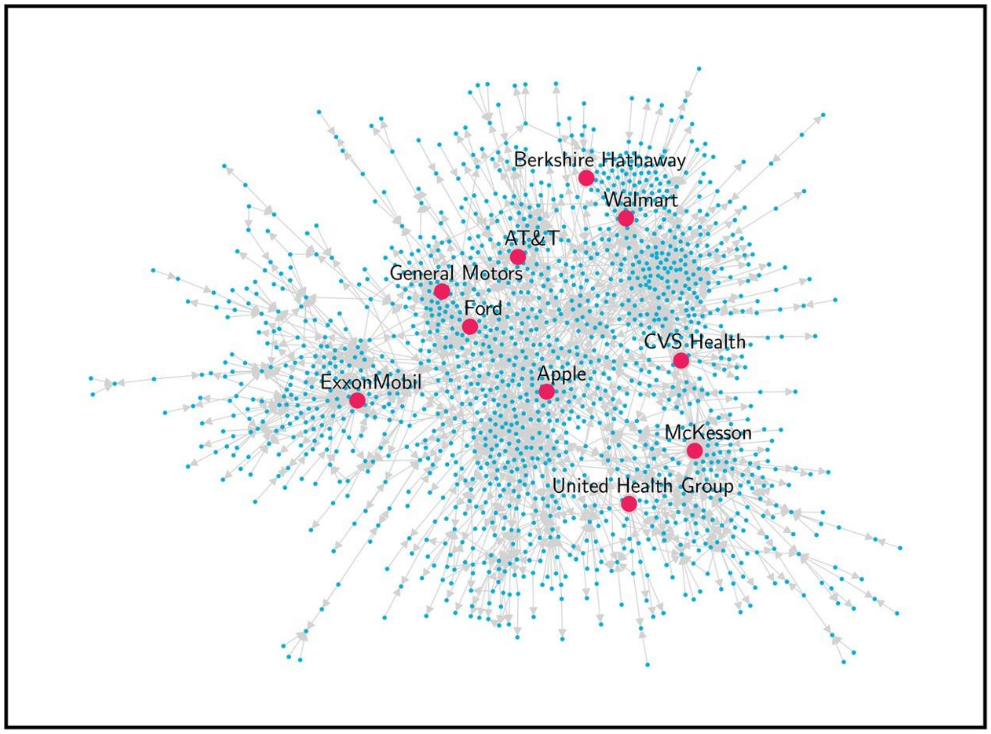

FIGURE 1.1 US buyer–supplier production network, 2015. *Sources:* Compustat – Capital IQ and authors' calculations.

If i directs an edges to j, and j likewise directs and edge back to i, we say the link is *reciprocated*. In some settings links are automatically reciprocated, or naturally "directionless" (e.g., partnerships), in which case the network is undirected. This is the case, for instance, in Fafchamps and Lund (2003) which focuses on (reciprocal) risk-sharing relationships in the rural Phillipines. Links in such an undirected graph are represented as unordered pairs of nodes instead of ordered pairs as described previously. In what follows we will present results for both directed and undirected networks depending on a combination of our immediate pedagogical goals, the illustrating application, and the state of the literature. While analogs of methods and algorithms available for directed networks are typically available for undirected ones, and vice versa, this is not always the case.

Returning to the digraph discussed earlier, the US supplier–buyer network is extraordinarily complex. Its structure may have implications for regulation and industrial policy, the diffusion of technology, and even macroeconomic policymaking (e.g., Carvalho, 2014; Acemoglu et al., 2016). In order to study this network, and others like it, we first need to know how to summarize its essential features. In non-network settings, empirical research often begins by tabulating a variety of summary statistics (e.g., means, medians, standard deviations, correlations). How might a researcher similarly summarize a dataset of relationships among agents? We outline some answers to this question in what follows.

While Fig. 1.1 is interesting to look at, it is not especially useful for statistical analysis. For this purpose it is convenient to represent $G(\mathcal{V}, \mathcal{E})$ by its $N \times N$ *adjacency matrix* $\mathbf{D} = \left[D_{ij}\right]$ where

$$D_{ij} = \begin{cases} 1, & (i, j) \in \mathcal{E}(G), \\ 0, & \text{otherwise.} \end{cases} \tag{1.1}$$

Here $D_{ij} = 1$ if agent i "sends" or "directs" a link to agent j (and zero otherwise), while $D_{ji} = 1$ if agent j directs a link to i. While the adjacency matrix for an undirected network will be symmetric as links are reciprocal, the adjacency matrix for a digraph need not (and typically will not) be symmetric. Self-links, or loops, are ruled-out here, though they may be allowed in different contexts, such that $D_{ii} = 0$ for all $i = 1, \ldots, N$. Econometric analysis of network data typically involves operations on the adjacency matrix as they allow one to focus on algebraic operations rather than graph-theoretic, combinatorial manipulations.[4] These matrices can also encode the strength of any links between a pair of nodes if this is available, like the traffic flow (edge) from one city (node) to another. A network with unweighted edges is typically referred to as a *simple graph* in the graph theory literature.

Once we define the basic objects of interest as graphs or adjacency matrices representing those, we can expand our discussion on probabilistic processes leading to observed social and economic networks. In other words, one can postulate a statistical model on the examined networks. Letting $\mathcal{G}$ be a particular set of graphs or networks, we can define a probability distribution over that particular set of graphs taken as a sample space. These probability models can be and usually are indexed by features or parameters related to the graphs in $\mathcal{G}$, like the number of vertices and/or other features. A collection of such models provides the basis for a statistical model. One of the early models, for example, imposes a uniform probability on the class of graphs with a given number of nodes, N, and a particular number of edges, $|\mathcal{E}|$, for $g \in \mathcal{G}$ (see Erdös and Rényi (1959) and Erdös and Rényi (1960)). Another basic, canonical random-graph model is one in which the edges between any two nodes follow an independent Bernoulli distribution with equal probability, say p. For a large enough number of nodes and sufficiently small probability of link formation p, the degree distribution approaches a Poisson distribution, and the model is consequently known as the Poisson random-graph model. This class of models appears in Gilbert (1959) and Erdös and Rényi (1960) and has since been studied extensively. While they

[4] There are other matrix representations for a network. For example, the *incidence* matrix will list vertices as rows and edges as columns. In digraphs, a node-edge entry in the incidence matrix is 1 if the node is the tail of the edge and -1 if it is the head. For undirected networks, the (unoriented) incidence matrix entry is 1 if the node in the row is part of the edge in that column. Such representations are related and informative about features of the network (e.g., via its eigenvalues and eigenvectors) and might be adequate for different purposes.

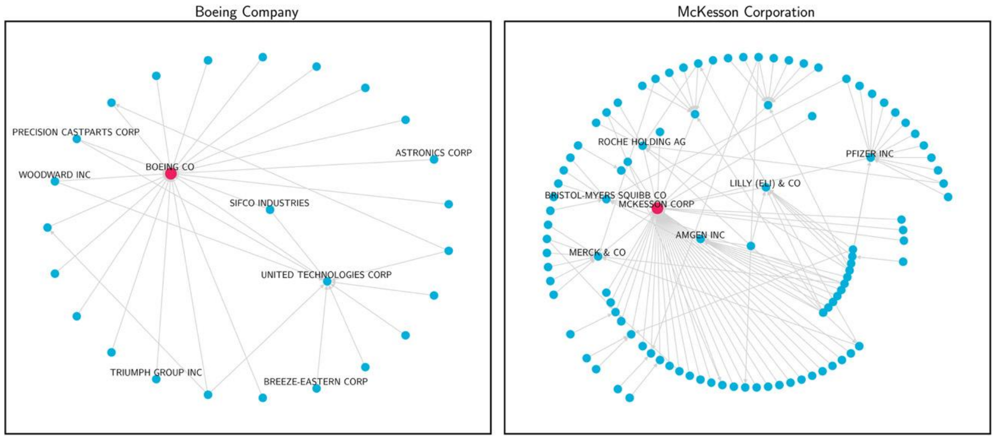

FIGURE 1.2 Boeing and McKesson supply chains, 2015. *Sources:* Compustat – Capital IQ and authors' calculations.

fail to reproduce important dependencies observed in social and economic networks accurately, they form important antecedents for the ensuing discussion in this volume.

Paths, distance and diameter

Imagine Fig. 1.1 is a map showing one way roads (edges) between cities (nodes). Under this analogy reciprocated links correspond to two way roads. If an individual can legally travel from city i to j along a sequence of one way roads (edges), we say there is a *walk* from i to j. When the walk does not repeat any cities (nodes) along the way, it is called a *path* and when it does not repeat any edges, it is called a *trail*. If our traveler traverses k edges on her trip, then we say the walk is of *length* k. Walks are directed in a digraph: it may be possible to go from i to j, but not back from j to i. If a walk runs from i to j, but not from j to i, we say i and j are *weakly connected*. If a walk runs in both directions, the two agents are *strongly connected*. In this case, a walk from city i to j and back that does not repeat any cities in between is a trail, in fact a *cycle*, but not a path. The shortest walk from i to j equals the *distance* from i to j.

The left-hand panel of Fig. 1.2 shows Boeing's supply chain. Inspecting this figure we can see that there is a length 1 path from Precision Castparts Corporation to Boeing. There is also a length 2 path which runs through United Technologies Corporation. Precision Castparts is both a direct and indirect supplier to Boeing. The distance from Precision Castparts to Boeing is one. The distance from Breeze Eastern Corporation to Boeing is two. Note that there is no directed path from Boeing to Breeze Eastern; the distance from Boeing to Breeze Eastern is infinite.

We say a directed network is *weakly connected* if for any two agents, there is a directed path connecting them. The network is *strongly connected* if there is a directed path from both i to j and j to i for all pairs of agents i and j.

Most real-world directed networks are not strongly connected, but many are weakly connected or, more precisely, contain a large *giant component* that is weakly connected. Fig. 1.1 actually does not show the full US buyer-supplier network, instead it just shows its largest weakly connected component (i.e., the maximum subset of nodes such that there is a directed path between all nodes in the subset). This weakly connected component includes over 80 percent all publicly traded firms in the United States. This indicates the substantial level of interconnectedness across the supply-chains of large firms in the United States economy. Such interconnectedness implies that shocks to just a few firms may affect the macroeconomy. Carvalho et al. (2016) show how the Great East Japan Earthquake of 2011, while directly impacting only a small fraction of Japanese firms, ultimately disrupted production in large portions of the Japanese and, to a lesser extent, global economies.

It turns out that we can count the number of k-length walks connecting two agents in a network, by inspecting powers of the adjacency matrix. Consider first the square of the adjacency matrix:

$$\mathbf{D}^2 = \begin{pmatrix} \sum_j D_{1j}D_{j1} & \sum_j D_{1j}D_{j2} & \cdots & \sum_j D_{1j}D_{jN} \\ \sum_j D_{2j}D_{j1} & \sum_j D_{2j}D_{j2} & & \sum_j D_{2j}D_{jN} \\ \vdots & & \ddots & \\ \sum_j D_{Nj}D_{j1} & \sum_j D_{Nj}D_{j2} & \cdots & \sum_j D_{Nj}D_{jN} \end{pmatrix}. \tag{1.2}$$

The ijth element of (1.2) coincides with the number of length two walks from agents i to j. If i links to k, and k links to j, then there exists a length two walk from i to j. The ijth element of (1.2) is a summation over all such length two walks. The diagonal elements of (1.2) equal the number of reciprocated ties to which agent i is party. Observe that reciprocated links are equivalent to length two walks from an agent back to herself.

Calculating $\mathbf{D}^3$ yields

$$\mathbf{D}^3 = \begin{pmatrix} \sum_{j,k} D_{1j}D_{jk}D_{k1} & \sum_{j,k} D_{1j}D_{jk}D_{k2} & \cdots & \sum_{j,k} D_{1j}D_{jk}D_{kN} \\ \sum_{j,k} D_{2j}D_{jk}D_{k1} & \sum_{j,k} D_{2j}D_{jk}D_{k2} & & \\ \vdots & & \ddots & \\ \sum_{j,k} D_{Nj}D_{jk}D_{k1} & \sum_{j,k} D_{Nj}D_{jk}D_{k2} & \cdots & \sum_{j,k} D_{Nj}D_{jk}D_{kN} \end{pmatrix}$$

whose ijth element gives the number of walks of length 3 from i to j. Note these walks may pass through a single agent twice. For example a length three path from i to j may involve walking from i to k, then back to i (via a reciprocated link), and then finally to j.

Proceeding inductively it is easy to show that the ijth element of $\mathbf{D}^k$ gives the number of walks of length k from agent i to agent j.

Theorem 1.1. *For a digraph G with adjacency matrix $\mathbf{D}$ and k a positive integer, the number of k-length walks from agents i to j coincides with the ijth element of $\mathbf{D}^k$.*

Proof. Let $D_{ij}^{(k)}$ denote the ijth element of $\mathbf{D}^k$. Begin by observing that $\mathbf{D}^0 = I_N$, correctly implying that the only zero length walks in the network are those from each agent to herself. Under the maintained hypothesis, $D_{ij}^{(k)}$ equals the number of k-length paths from i to j. The number of $k+1$ length paths from i to j then equals

$$\sum_{k=1}^{N} D_{ik}^{(k)} D_{kj},$$

which equals the ijth element of $\mathbf{D}^{k+1}$. The claim follows by induction. □

We can also use powers of the adjacency matrix to calculate shortest path distances or "degrees of separation". Specifically,

$$M_{ij} = \min_{k \in \{1,2,3,\ldots\}} \left\{ k : D_{ij}^{(k)} > 0 \right\} \tag{1.3}$$

equals the distance from i to j (if it is finite). For modestly-sized networks M_{ij} can be calculated by taking successive powers of the adjacency matrix. If the network is strongly connected, we can compute the *average distance* as

$$\overline{M} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} M_{ij}. \tag{1.4}$$

Since few directed networks are strongly connected, (1.4) is rarely finite. Consequently it can be insightful to first convert a directed network to an undirected one and then compute average distance as

$$\overline{M} = \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j<i} M_{ij}.$$

If the undirected network is not connected, then the average can be taken across dyads within its largest connected component.

The *diameter* of a network is the largest distance between any two agents in it. It will be finite if the network consists of a single strongly connected component (in which case all agents are "reachable" starting from any other agent) and infinite in weakly connected networks, or in those consisting of multiple strongly connected components (in which case there are no paths connecting some pairs of agents). As with average distance, it can sometimes be fruitful to first convert a directed network to an undirected one prior to computing is diameter.

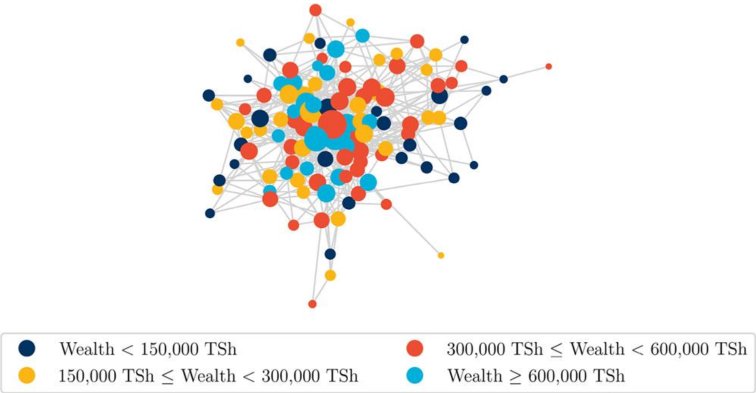
Wealth < 150,000 TSh
150,000 TSh ≤ Wealth < 300,000 TSh
300,000 TSh ≤ Wealth < 600,000 TSh
Wealth ≥ 600,000 TSh

FIGURE 1.3 Nyakatoke risk-sharing network. *Sources:* De Weerdt (2004) and authors' calculations.

TABLE 1.1 Frequency of degrees of separation in the Nyakatoke network.

	1	2	3	4	5
Count	490	2666	3298	557	10
Frequency	0.0698	0.3797	0.4697	0.0793	0.0014

Source: De Weerdt (2004) and authors' calculations.

An illustration of these concepts is provided by the Nyakatoke risk-sharing network first studied by De Weerdt (2004). This network is depicted in Fig. 1.3, which plots risk-sharing links between households in the village of Nyakatoke, Tanzania. Households in Nyaktoke were asked about other individuals in the village they could "personally rely on for help". The network in Fig. 1.3 was constructed by placing an undirected edge between two households if a member in one reports being able to rely on help from a member in another, the opposite, or both.

The Nyakatoke network consists of a single giant component. Table 1.1 tabulates the frequency of shortest path lengths across all $\binom{119}{2} = 7,021$ dyads in the Nyakatoke network. The Nyakatoke network is, in many ways, prototypical of other small and medium-sized social and economic networks. First it is relatively *sparse*: only 490 out of 7,021 dyads in the Nyakatoke are directly connected (less than seven percent).[5] While only a small fraction of all possible links are present, shortest path lengths between any two nodes are small: over 85 percent of dyads are less that three degrees apart. The maximum distance between any two households, corresponding to the diameter of the network, is also small, equaling five.

The conjunction of sparseness and low diameter is common in social and economic networks and sometimes called the "small world phenomenon." This

[5] In statistical models of network formation investigated later, sparsity typically refers to the number of links, being $O_p(N^2)$, when N is allowed to grow.

phrase was popularized by the social psychologist Stanley Milgram (1967) who argued, on the basis of computer simulations and real-world data collected through a series of postal experiments in the 1960s, that any two individuals in the United States are often connected through a short chain of acquaintances (e.g., "six degrees of separation").

Network sparseness and low diameter make the statistical analysis of network data challenging. Intuitively these two properties imply that there is little data and (perhaps) appreciable dependence across observations. Much of modern statistical analysis involves understanding what can be learned by averaging many independent pieces of data. Network statistical analysis often requires assessing what can be learned from small amounts of dependent data.

Measuring homophily

A well-documented feature of many real-world social and economic networks is homophily: the tendency of agents to form links with others similar to themselves (e.g., McPherson et al., 2001; Pin and Rogers, 2016). Many types of social relationships occur more frequently between individuals with similar socio-demographic attributes (i.e., race, gender, social class; cf., Marsden, 1987). Homophily also extends beyond social links to economic ones. For example, Bengtsson and Hsu (2015) present evidence that co-ethnicity of investors and company founders is an important predictor of venture capital flows in the United States.[6]

The presence and magnitude of homophily and degree heterogeneity has implications for how information diffuses, the spread of epidemics, as well as the speed and precision of social learning (e.g., Pastor-Satorras and Vespignani, 2001; Jackson and Rogers, 2007; Golub and Jackson, 2012; Jackson and López-Pintado, 2013).[7]

In this section we consider the measurement of homophily in practice. For simplicity we focus on the *undirected* network case.

A measure of homophily captures the extent to which observed agent attributes X_i and X_j are more similar (in value) across agents who are linked ($D_{ij} = 1$) relative to those who are not ($D_{ij} = 0$); or relative to some benchmark model (e.g., a null model where agents match completely at random). In the statistical physics literature homophily is typically measured by what Newman (2010) calls the *modularity* of a network; this measure is now widely used in other fields as well. In the case of a binary attribute, network modularity is closely related to standard (and decades old) measures of residential segregation. As in the literature on the measurement of segregation, statistical measures of homophily are often presented as denizens of the sample data alone. That

[6] Another empirically robust example is assortative-matching by race among high schoolers in the United States (e.g. Currarini et al., 2009).

[7] Apicella et al. (2012) even study the relationship between homophily and the emergence of cooperation in hunter-gatherer societies.

is, without the context of a clear generative or population model (cf., Graham, 2018). The lack of such a generative model makes the interpretation and analysis of homophily measures difficult though connections with statistically-based models where ties form based on communities have been established (see Newman (2016)).

In this section we introduce some notation and use it to provide a simple probabilistic interpretation of network modularity. Our approach is guided (albeit rather indirectly) by graphon representations of probability distributions for exchangeable random graphs (e.g., Diaconis and Janson, 2008; Lovász, 2012; Bickel and Chen, 2009). Let $X_i \in \mathbb{X} \subset \mathbb{R}^1$ be some scalar-valued agent attribute and imagine that the link probability between i and j is guided by such attributes. Adapting the sample-based definition given by Newman (2010, p. 779), we define the *assortativity coefficient* or *normalized modularity* as

$$\rho_{\mathrm{AC}} = \frac{\mathbb{E}\left[X_i X_j \middle| D_{ij} = 1\right] - \mathbb{E}\left[X_i \middle| D_{ij} = 1\right] \mathbb{E}\left[X_j \middle| D_{ij} = 1\right]}{\mathbb{E}\left[X_i^2 \middle| D_{ij} = 1\right] - \mathbb{E}\left[X_i \middle| D_{ij} = 1\right]^2}. \tag{1.5}$$

Eq. (1.5) is reminiscent of the definition of correlation between two random variables Goldberger (1991, p. 66). In fact (1.5), as we will demonstrate shortly, has such an interpretation, but, in the absence of additional structure, it is difficult to make much sense of the expected values present in (1.5).

We begin by establishing notation for the conditional probability of the event $D_{ij} = 1$ given that $X_i = x$ and $X_j = y$:

$$\omega(x, y) = \Pr\left(D_{ij} = 1 \middle| X_i = x, X_j = y \right). \tag{1.6}$$

Integrating (1.6) over x and y gives, in a small abuse of notation, the marginal link probability

$$\rho = \int \omega(x, y) f_X(x) f_X(y) \,\mathrm{d}x\mathrm{d}y. \tag{1.7}$$

Finally, Bayes' law, together with (1.6) and (1.7), gives

$$f_{X_i, X_j | D_{ij}}\left(x, y \middle| D_{ij} = 1 \right) = \frac{\omega(x, y)\, \rho}{f_X(x) f_X(y)}, \tag{1.8}$$

which illustrates how linking behavior determines the conditional distribution of covariates across linked dyads and hence homophily. The elements in the numerator of (1.8) are features of the network formation process, while those entering the denominator are features of the population of agents. Both are familiar objects. The distribution (1.8) can be used to understand the expectations appearing in (1.5) above.

In the Nyakatoke network the assortativity coefficient takes a value of 0.073 for the logarithm of land and livestock wealth (converted into Tanzanian shillings) and 0.094 for age of household head in years.

Measuring agent centrality

A natural question to ask, when viewing Fig. 1.1, is: which firms are most important or *central* in the US economy? It turns out that this is a classic question in network analysis, with a long history across several disciplines (e.g., Wasserman and Faust, 1994). Here we review a handful of centrality measures that economists undertaking network analysis have found especially useful.

Acemoglu et al. (2012) study how firm-level production shocks may cascade through the economy via supplier connections (cf., Carvalho, 2014). They argue that local shocks to certain 'key', or central, firms may have sizable aggregate effects. Ballester et al. (2006) develop a model of criminal behavior where a specific measure of centrality identifies those criminals in a network whose apprehension would lead to the greatest reductions in criminal activity. Kim et al. (2015) use various centrality measures to target a peer-spread public health intervention in Honduras (cf., Banerjee et al., 2013). In the wake of the 2007 to 2009 financial crisis, regulators have been interested in identifying financial institutions which are 'too connected to fail' (e.g., Battiston et al., 2012; Denbee et al., 2014). Measures of agent centrality feature in all of these research projects.

Degree centrality

In a digraph, the *indegree* of agent i coincides with the number of arcs directed toward her, while her *outdegree* equals the number of arcs she directs toward other agents. Arithmetically, the indegree of agent i equals $D_{+i} = \sum_j D_{ji}$, while her outdegree equals $D_{i+} = \sum_j D_{ij}$ (here the '+' denotes summation over the replaced index). For an undirected network, the degree of a vertex is simply the number of edges incident with that node. The indegree *sequence* of the network, $\mathbf{D}_{+\bullet}$, equals the vector of column sums of the adjacency matrix (i.e., $\mathbf{D}_{+\bullet} = \mathbf{D}'\iota_N$). The *outdegree sequence*, $\mathbf{D}_{\bullet+}$, equals the vector of row sums of the adjacency matrix (i.e., $\mathbf{D}_{\bullet+} = \mathbf{D}\iota_N$). When the network is undirected, the *degree sequence* is given by the vector of column or row sums of the adjacency matrix, which are equal by symmetry of that matrix.

In sociology an agent's indegree is often called degree prestige (e.g., Wasserman and Faust, 1994, p. 202). Outdegree sequences are less well-studied, but can be important in economics. For example, the outdegree may be informative about agents who are well positioned to disperse information quickly (e.g., by sending 'news' to agents to which they have directed ties). In the context of production networks, the outdegree may help to identify critical input suppliers; firms whose output is used by many different downstream firms (e.g., Acemoglu et al., 2012).

Table 1.2 lists those firms with the largest number of suppliers according to the Compustat production network dataset (i.e., an indegree ranking). The list is populated by a mix of large retailers (Walmart, Home Depot and Target), healthcare and pharmaceutical firms (McKesson, Cardinal Health and AmerisourceBergen), as well automakers (Ford, General Motors), an energy

TABLE 1.2 US firms with the most suppliers, 2015.

Firm	Number of suppliers
Walmart Stores Inc.	115
Royal Dutch Shell plc	48
McKesson Corp.	41
Cardinal Health Inc.	40
Home Depot Inc.	37
AmerisourceBergen Corp.	35
Ford Motor Co.	31
General Motors Co.	28
Target Corp.	26
AT&T Inc.	22
Chevron Corp.	22

Notes: List of top ten firms by indegree (number of suppliers) in the US economy in 2015. Note there is a tie for 10th place.
Sources: Compustat – Capital IQ and authors' calculations.

conglomerate (Shell), and a communications company (AT&T). As detailed by Carvalho (2014), the high indegree of the "Big Three" automakers, as well as the overlap among their suppliers, was used as an argument for the 2009 government rescue of General Motors and Chrysler.

Unfortunately the Compustat data is not helpful for ranking firms by outdegree, since most firms list (at most) their ten largest customers when filing with the SEC. This reporting rule artificially truncates firm outdegree at ten (cf., Atalay et al., 2011).

Refinements of degree centrality

While degree-based centrality measures are simple to understand and compute, they have well-known limitations. Consider two firms, both with ten upstream suppliers. For one of those firms, each of its suppliers, itself has 10 suppliers further upstream, while the other firm's suppliers do not (e.g., they are just raw materials suppliers). Indegree centrality ranks these two firms identically, while intuition would suggest that the former firm is more central since its suppliers have higher indegree. Bonacich (1972), building on earlier work by Katz (1953), introduced a measure of centrality, called *eigenvector centrality*, designed to ameliorate this limitation of degree centrality.[8] In directed networks there are two variants of Bonacich's (1972) measure, respectively generalizing indegree and outdegree centrality. In what follows we first introduce various generaliza-

[8] The eigenvector centrality also appears in the independent work by Gould (1967) and is alternatively known as Gould's accessibility index in geography.

tions of indegree centrality before subsequently discussing how these measures may be adapted to generalize outdegree centrality.

The eigenvector centrality of an agent is recursively defined as a linear combination of the centralities of those who direct links toward her:

$$c_i^{\mathrm{EC}}(\mathbf{D}) = \sum_j c_j^{\mathrm{EC}}(\mathbf{D}) D_{ji},$$

or, in matrix form, with $\mathbf{c}^{\mathrm{EC}}(\mathbf{D}) = \left(c_1^{\mathrm{EC}}(\mathbf{D}), \ldots, c_N^{\mathrm{EC}}(\mathbf{D})\right)$,

$$\mathbf{c}^{\mathrm{EC}}(\mathbf{D}) = \mathbf{c}^{\mathrm{EC}}(\mathbf{D})\mathbf{D}. \tag{1.9}$$

Like other centrality measures, this is a self-referential measure: a vertex is central if it is connected to more central vertices. Inspection of (1.9) indicates that $\mathbf{c}^{\mathrm{EC}}(\mathbf{D})$ is a row (or left) eigenvector of $\mathbf{D}$ associated with an eigenvalue 1. Therefore (1.9) only has a non-zero solution if 1 is an eigenvalue of the adjacency matrix. To ensure a non-zero solution Bonacich (1972) suggests replacing (1.9) with $\mathbf{c}^{\mathrm{EC}}(\mathbf{D}, \phi) = \phi\mathbf{c}^{\mathrm{EC}}(\mathbf{D}, \phi)\mathbf{D}$, where ϕ equals the inverse of the largest eigenvalue of $\mathbf{D}$.[9] An alternative approach, first suggested by Katz (1953), is to row-normalize the adjacency matrix. Define the *row-normalized* adjacency matrix as

$$\mathbf{G} = \mathrm{diag}\left\{\max(1, D_{1+}), \ldots, \max(1, D_{N+})\right\}^{-1} \times \mathbf{D}. \tag{1.10}$$

Observe that the ith row of (1.10) sums to either zero (if agent i has an outdegree of zero) or one (if agent i has positive outdegree). If all agents have positive outdegree, then $\mathbf{G}$ will be a row-stochastic matrix. Replacing $\mathbf{D}$ with its row-normalized counterpart $\mathbf{G}$ in (1.9) yields

$$\mathbf{c}^{\mathrm{K}}(\mathbf{D}) = \mathbf{c}^{\mathrm{K}}(\mathbf{D})\mathbf{G}. \tag{1.11}$$

If $\mathbf{G}$ is row-stochastic, then $\mathbf{c}^{\mathrm{K}}(\mathbf{D})$ corresponds to a stationary vector of a Markov chain with transition matrix $\mathbf{G}$. From the theory of Markov chains we know that if the matrix $\mathbf{G}$ is irreducible, then this stationary vector is unique. It turns out that irreducibility holds if, and only if, the network is strongly connected. Unfortunately, as noted earlier, few real-work social and economic networks are strongly connected (at least when edges are directed). This includes the production network depicted in Fig. 1.1. Indeed, not only does strong connectivity typically fail, but many directed networks have "dangling nodes"

[9] This gives $\mathbf{c}^{\mathrm{EC}}(\mathbf{D}, \phi)$ as the solution to $\mathbf{c}^{\mathrm{EC}}(\mathbf{D}, \phi)\left[\frac{1}{\phi} I_N - \mathbf{D}\right] = 0$, which corresponds the left eigenvector associated with the largest eigenvalue of $\mathbf{D}$. If the adjacency matrix is nonnegative and corresponds to a strongly connected network, a linear algebra result known as the Perron–Frobenius theorem guarantees that there is a dominant real eigenvalue corresponding to the one (up to normalization) eigenvector that can be taken to have positive entries. Its entries correspond to the eigenvector centrality.

(agents with zero indegree); ${c_i}^{\mathrm{K}}(\mathbf{G})$ will equal zero for such agents. This will also be the case for all agents with incoming links solely from dangling nodes and so on.

PageRank

The problem of dangling nodes, as well as the failure of strong connectivity, motivated Sergey Brin and Lawrence Page, at the time graduate students in computer science at Stanford University, to develop the PageRank centrality measure, now used by Google to rank web-search results (Brin and Page, 1998; Page et al., 1999). Brin and Page made two modifications to the basic Katz (1953) measure. First, they regularized the (row-normalized) adjacency matrix so that all rows, including those associated with dangling nodes, sum to one. Specifically, they introduced what is now called the *Google Matrix* $\mathbf{H} = \left[H_{ij}\right]$ with elements

$$H_{ij} = \begin{cases} \phi G_{ij} + \frac{(1-\phi)}{N} & \text{if } D_{i+} > 0, \\ \frac{1}{N} & \text{otherwise.} \end{cases} \tag{1.12}$$

Observe that $\mathbf{H}$ is both row-stochastic and irreducible.

Second, as first suggested by Katz (1953) and Bonacich (1987), they endow each agent with a small amount of exogenous centrality:

$$\mathbf{c}^{\mathrm{PR}}(\mathbf{D}, \phi) = \phi \mathbf{c}^{\mathrm{PR}}(\mathbf{D}, \phi)\,\mathbf{H} + \left(\frac{1-\phi}{N}\right)\iota_N'. \tag{1.13}$$

Here ι_N denotes a $N \times 1$ vector of ones. A typical value for ϕ, at least in web search, is 0.85.[10] For $|\phi| < 1$ the matrix $I_N - \phi\mathbf{H}$ is strictly (row) diagonally dominant (I_N is the $N \times N$ identity matrix). By the Levy–Desplanques theorem (e.g., Horn and Johnson, 2013) it is therefore non-singular. Non-singularity of $(I_N - \phi\mathbf{H})$ allows us to solve for the PageRank vector as

$$\mathbf{c}^{\mathrm{PR}}(\mathbf{D}, \phi) = \left(\frac{1-\phi}{N}\right)\iota_N' \left(I_N - \phi\mathbf{H}\right)^{-1}. \tag{1.14}$$

To motivate the PageRank measure we can appeal to a random web surfer stochastic process. Imagine an individual surfing the web. With probability ϕ she moves to another page by choosing one of the outgoing links at her current location, each with equal probability. With probability $1-\phi$ she instead chooses a page at random from the set of all pages. If the current page corresponds to a dangling node, she just chooses a page at random. Given the above process,

[10] This value is related to the magnitude of the second eigenvalue of the Google Matrix, the size of which determines the speed with which (1.13) may be iteratively solved for $\mathbf{c}^{\mathrm{PR}}(\mathbf{D}, \phi)$. In modestly sized networks it is generally possible to set ϕ much closer to one.

TABLE 1.3 Central buying firms in the US economy, 2015.

Firm	Buyer's PageRank
Walmart Stores Inc.	0.0272
CVS Health Corp.	0.0198
Royal Dutch Shell plc	0.0124
AmerisourceBergen Corp.	0.0094
McKesson Corp.	0.0086
Cardinal Health Inc.	0.0081
Home Depot Inc.	0.0060
HP Inc.	0.0056
Express Scripts Holding Co.	0.0050
BP Plc.	0.0047
Apple Inc.	0.0047
Boeing Co.	0.0047

Notes: List of ten most central firms in the US economy in 2015 according to PageRank ($\alpha = 0.95$). Note there is a three-way tie for 10th place.
Sources: Compustat – Capital IQ and authors' calculations.

${c_i}^{\mathrm{PR}}(\mathbf{D}, \phi)$ corresponds to the frequency with which our surfer visits page i in equilibrium.[11]

Table 1.3 lists the top ten firms in the US economy according to the PageRank index. The list largely overlaps with the indegree ranking reported in Table 1.2, but there are also important differences. Specifically the aircraft manufacturer Boeing, and the computer companies Hewlett-Packard and Apple, enter the top 10; displacing the car manufacturers Ford and General Motors.

Some insight into why the relative rankings according to indegree and PageRank of, for example, Ford and Boeing, differ is provided by examining their respective supply chains. The *ancestors* of node i consist of all nodes with a directed path from themselves to i. In the context of a supply chain ancestors of a firm include its direct suppliers, its supplier's suppliers and so on. Figs. 1.2 (left panel) and 1.4 (right panel) display the subgraphs induced by, respectively, Boeing and all its direct and indirect suppliers and those induced by Ford and all its direct and indirect suppliers. Ford's supply chain takes a traditional "vertical" form, with many firms delivering intermediate parts to Ford for final assembly into cars. Boeing's supply chain is more complicated; United Technologies di-

[11] If we use the series expansion

$$(I_N - \phi\mathbf{H})^{-1} = \sum_{k=0}^{\infty} \phi^k \mathbf{H}^k,$$

as well as the fact that $\mathbf{H}\iota_N = \iota_N$ (and hence that $\mathbf{H}^k \iota_N = \iota_N$ for $k \geq 1$) it is easy to verify that $\sum_{i=1}^{N} {c_i}^{\mathrm{PR}} = \left(\frac{1-\phi}{N}\right) \iota_N' \left(I_N - \phi_0 \mathbf{H}\right)^{-1} \iota_N = 1$. Hence $\mathbf{c}^{\mathrm{PR}}(\mathbf{D}, \phi)$ is a valid probability distribution.

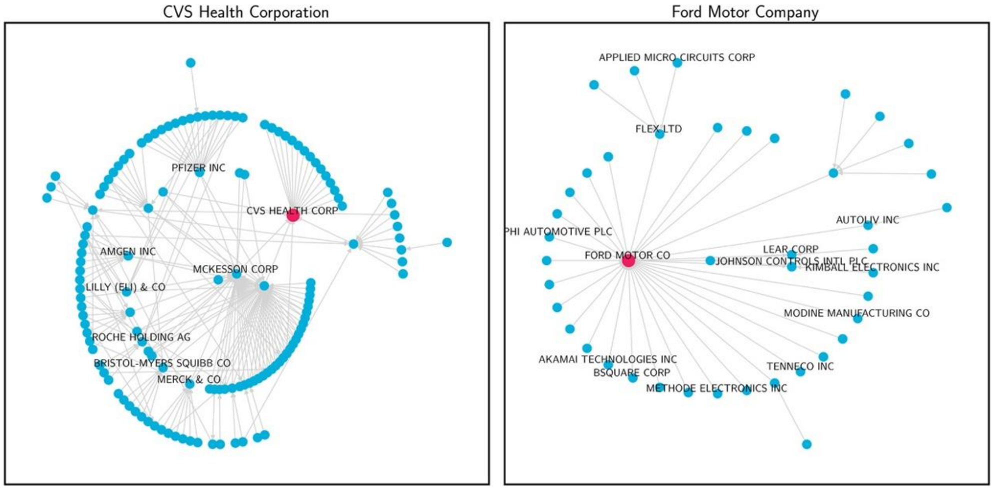

FIGURE 1.4 Supply chain structure and PageRank. *Notes:* The left-hand figure displays the induced subdigraph associated with CVS and all its ancestor nodes (i.e., the CVS supply chain). The right-hand figure displays the corresponding Ford Motor Company supply chain. *Sources:* Compustat - Capital IQ and authors' calculations.

rectly supplies Boeing, while also being a large buyer of intermediate inputs. Furthermore the two firms share three suppliers in common. These features of Boeing's supply chain architecture drive its higher PageRank customer ranking compared to Ford.

PageRank and the social multiplier

The concept of a social multiplier has been a key theme of empirical work on social interactions in economics since the publication of Manski (1993). It features in, for example, Brock and Durlauf (2001), Glaeser and Scheinkman (2001, 2003), Graham (2008), Graham et al. (2010), Angrist (2014) and Galeotti et al. (2017). In the presence of social multiplier effects, the full impact of an intervention exceeds the initial impact due to feedback effects across agents. When interactions occur on a non-trivial network, the magnitude of any multiplier effect will also depend upon exactly which agent is initially acted upon by the policy-maker. This is the intuition behind *social multiplier centrality*. In turns out that PageRank centrality shares a close connection with *social multiplier centrality*.

There are several ways to make the connection between PageRank and the social multiplier; the easiest involves introducing a simple quadratic complementarity game of the type recently surveyed by Jackson and Zenou (2015). Let Y_i be some continuously-valued action chosen by agent $i = 1, \ldots, N$. Let $\mathbf{Y}$ be the $N \times 1$ vector of all agents' actions. Let, as before, ι_N be an $N \times 1$ vector of ones, and $\mathbf{G}$ be the row-normalized network adjacency matrix. Initially assume that this matrix is row-stochastic and irreducible (i.e., that the network is strongly connected).

Observe that

$$\mathbf{G}_i\mathbf{y} = \sum_{j \neq i} G_{ij} y_j \overset{def}{\equiv} \bar{y}_{n(i)}$$

equals the average action of player i's direct peers (i.e. those players to whom she has directed a link) under the (perhaps hypothetical) action profile $\mathbf{y}$. Here $\mathbf{G}_i$ denotes the ith row of $\mathbf{G}$, $n(i)$ the index set $\left\{j \ : \ D_{ij} = 1, \ \ j \neq i\right\}$, and $\bar{y}_{n(i)}$ the average of Y_j over these indices.

Following Blume et al. (2015), among others, assume that the utility agent i receives from action profile $\mathbf{y}$ given the network structure is

$$\begin{aligned} u_i\left(\mathbf{y};\mathbf{D}\right) &= \left(\alpha_0 + U_i\right) y_i - \frac{1}{2} y_i^2 + \beta_0 \bar{y}_{n(i)} y_i \\ &= \left(\alpha_0 + U_i\right) y_i - \frac{1}{2} y_i^2 + \beta_0 \mathbf{G}_i \mathbf{y} y_i \end{aligned} \tag{1.15}$$

with $0 < |\beta_0| < 1$ and $\mathbb{E}\left[U_i\right] = 0$. Here U_i captures heterogeneity in agents' preferences for action.

The marginal utility associated with an increase in y_i is increasing in the average action of one's peers, $\bar{y}_{n(i)}$. Specifically,

$$\frac{\partial^2 u_i\left(\mathbf{y},\mathbf{D}\right)}{\partial y_i \partial \bar{y}_{n(i)}} = \beta_0.$$

That is, if $\beta_0 > 0$, own- and peer-action are complements. In the terminology of Manski (1993), the magnitude of β_0 is an index for the strength of any *endogenous social interactions*.

Assume that the observed action $\mathbf{Y}$ corresponds to a Nash equilibrium where no agent can increase her utility by changing her action given the actions of all other agents in the network. Agents observe $\mathbf{D}$, the network structure, and $\mathbf{U}$, the $N \times 1$ vector of individual-level heterogeneity terms.

The first order condition for optimal behavior associated with (1.15) generates the following best response function:

$$y_i = \alpha_0 + \beta_0 \bar{y}_{n(i)} + U_i \tag{1.16}$$

for $i = 1, \ldots, N$. Eq. (1.16) is a special case of what is called the linear-in-means model of social interactions (e.g., Brock and Durlauf, 2001). An agent's best reply varies with the average action of those to whom she is directly connected, y, as well as the unobserved own attribute, U_i (which shifts the marginal utility of action across agents).

Eq. (1.16) defines an $N \times 1$ system of simultaneous equations. Since observed actions correspond to equilibrium values, the econometrician observes actions which satisfy (1.16). Writing the system defined by (1.16) in matrix form gives

$$\mathbf{Y} = \alpha_0 \iota_N + \beta_0 \mathbf{G}\mathbf{Y} + \mathbf{U}. \tag{1.17}$$

For $|\beta_0| < 1$, solving (1.17) for the equilibrium action vector, $\mathbf{Y}$, as a function of $\mathbf{D}$ and $\mathbf{U}$ alone, yields the reduced form

$$\mathbf{Y} = \alpha_0 \left(I_N - \beta_0 \mathbf{G}\right)^{-1} \iota_N + \left(I_N - \beta_0 \mathbf{G}\right)^{-1} \mathbf{U} \tag{1.18}$$

or, using a series expansion (see footnote 11),

$$\mathbf{Y} = \frac{\alpha_0}{1-\beta_0} \iota_N + \left[\sum_{k=0}^{\infty} \beta_0^k \mathbf{G}^k \right] \mathbf{U}. \tag{1.19}$$

Eq. (1.19) provides some insight into what various researchers have called the social multiplier. Consider a policy which increases the ith agent's value of U_i by Δ. We can conceptualize the full effect of this increase on the network's distribution of outcomes as occurring in "waves". In the initial wave only agent i's outcome increases. The change in the entire action vector is therefore

$$\triangle \mathbf{e}_i,$$

where $\mathbf{e}_i$ is an N-vector with a one in its ith element and zeros elsewhere.

In the second wave all of agent i's peers experience outcome increases. This is because their best reply actions change in response to the increase in agent i's action in the initial wave. The action vector in wave two therefore changes by

$$\triangle \beta_0 \mathbf{G} \mathbf{e}_i.$$

In the third wave the outcomes of agent i's friends' friends change (this may include a direct feedback effect back onto agent i if some of her links are reciprocated). In wave three we get a further change in the action vector of

$$\Delta \beta_0^2 \mathbf{G}^2 \mathbf{e}_i.$$

In the kth wave we have a change in the action vector of

$$\triangle \beta_0^{k-1} \mathbf{G}^{k-1} \mathbf{e}_i.$$

Observing the pattern of geometric decay we see that the "long-run" or full effect of a Δ change in U_i on the entire distribution of outcomes is given by

$$\triangle \left(I_N - \beta_0 \mathbf{G}\right)^{-1} \mathbf{e}_i. \tag{1.20}$$

Eq. (1.20) indicates the effect of perturbing U_i by $\triangle$ on the equilibrium action vector coincides with the ith column of the matrix $\triangle \left(I_N - \beta_0 \mathbf{G}\right)^{-1}$. The total effect on aggregate action is therefore given by the sum of the ith column of this matrix. Hence the row vector

$$\mathbf{c}^{\text{SM}}(\mathbf{D}, \beta) = \iota_N' (I_N - \beta \mathbf{G})^{-1} \tag{1.21}$$

equals a *social multiplier centrality* measure for each agent in the network. By construction this measure is greater than or equal to one for $\beta_0 \geq 0$. If $c_i^{\text{SM}}(\mathbf{D}, \beta) = 2$, then the effect of intervening to increase U_i by Δ on the aggregate action $\sum_{i=1}^{N} Y_i$ is twice the initial direct effect of $\triangle$. Averaging over all agents we get

$$\frac{1}{N}\sum_{i=1}^{N} c_i^{\text{SM}}(\mathbf{D}, \beta) = \frac{1}{1-\beta}$$

(again see footnote 11); this is the form of the social multiplier in the linear-in-means model as first formulated by Manski (1993) (cf., Glaeser and Scheinkman, 2001, 2003). In the presence of non-trivial network structure, the full effect of an intervention will, unlike in the model of Manski (1993), vary heterogeneously across agents. If we multiply the elements of (1.21) by $(1-\beta)/N$ we recover the PageRank centrality measure of Brin and Page (1998).

Our analysis assumes that $\mathbf{G}$ is irreducible. In cases where it is not, replacing $\mathbf{G}$ in (1.21) with the Google matrix (1.12), with ϕ set equal to a value 'close to one', yields a centrality measure with approximately the same interpretation. In this case ϕ is a regularization parameter.

Katz–Bonacich centrality

Closely related to both PageRank and social multiplier centrality, but older than either, is the Katz–Bonacich centrality measure (Bonacich, 1987; Bonacich and Lloyd, 2001). This measure also often arises in the context of quadratic complementarity games played on networks (Ballester et al., 2006; Calvó-Armengol et al., 2009; Jackson and Zenou, 2015). Katz–Bonacich centrality is increasing in the indegree of an agent, the indegree of those agents who direct link to her and so on, with weights discounted according to the degree of separation:

$$\begin{aligned}\mathbf{c}^{\text{KB}}(\mathbf{D}, \phi) &= \phi \iota_N' \mathbf{D} + \phi^2 \iota_N' \mathbf{D}^2 + \phi^3 \iota_N' \mathbf{D}^3 + \cdots \\ &= \left(\phi \iota_N' \mathbf{D}\right)\left(I_N + \phi \mathbf{D} + \phi^2 \mathbf{D}^2 + \cdots\right) \\ &= \left(\phi \iota_N' \mathbf{D}\right)\left[\sum_{k=0}^{\infty} \phi^k \mathbf{D}^k\right].\end{aligned}$$

For $\phi < 1/\lambda_1$, with λ_1 the maximum eigenvalue of the adjacency matrix, the sequence in brackets converges so that the vector of Katz–Bonacich centralities equals

$$\mathbf{c}^{\mathrm{KB}}(\mathbf{D},\phi)=\left(\phi\iota_N'\mathbf{D}\right)(I_N-\phi\mathbf{D})^{-1}. \tag{1.22}$$

Outdegree-based centrality measures

PageRank, social multiplier and Katz–Bonacich centrality, as introduced above, are all prestige-type measures: central nodes have links directed toward them by other central nodes and so on. In settings where the process of information diffusion or shock propagation is of central interest, outdegree-based measures may be of greater interest. For an empirical example in economics consider the work of Acemoglu et al. (2012) and Carvalho (2014). These papers study the macro effects of output shocks on downstream firms. They argue that shocks to firms that supply many firms (or supply a firm that itself supplies many firms) may have large aggregate effects.

For concreteness consider the Buyer–Supplier network depicted in Fig. 1.1. Replacing $\mathbf{D}$ with $\mathbf{D}'$ in (1.14) yields an outdegree-based variant of PageRank. While $\mathbf{c}^{\mathrm{PR}}(\mathbf{D},\phi)$ will tend to rank large downstream firms with many suppliers as central, $\mathbf{c}^{\mathrm{PR}}\left(\mathbf{D}',\phi\right)$ will instead rank key upstream firms as central (i.e., suppliers with high outdegree, or suppliers-of-suppliers with high outdegree and so on).

A stochastic process interpretation of $\mathbf{c}^{\mathrm{PR}}\left(\mathbf{D}',\phi\right)$ may be helpful. Consider an input purchaser traversing our buyer-supplier network. During each period she makes, with probability ϕ, an intermediate input purchase from one of the suppliers (predecessors) of the current firm; choosing one supplier at random. With probability $1-\phi$ she makes a purchase completely at random from the set of *all* firms. She then moves *upstream* to the *selling* firm from which she made a purchase and repeats the purchasing process. If, during this process, she ends up at a firm with no suppliers (e.g., a raw materials company) she simply makes a purchase at random from the set of all firms. In equilibrium $c_i^{\mathrm{PR}}\left(\mathbf{D}',\phi\right)$ equals the fraction, out of all her intermediate input purchases, that come from firm i (i.e., where firm i is the selling or supplying firm). Hence if $c_i^{\mathrm{PR}}\left(\mathbf{D}',\phi\right)$ is large we might reasonably call firm i an ‘important’ or a *central intermediate input supplier*.

The analogous stochastic process for $\mathbf{c}^{\mathrm{PR}}(\mathbf{D},\phi)$, PageRank as initially introduced, involves a hypothetical saleswoman. During each period she sells her current firm’s output, with probability ϕ, to one of its buyers (successors). With probability $1-\phi$ she sells to a random firm chosen from the set of all firms. She then moves *downstream* to the *buying* firm which made the purchase from her and repeats the sales process. If, during this process, she ends up at a firm with no buyers (e.g., a large retail firm that sells only to final consumers like Walmart) she makes a sale at random to a firm chosen from the set of all firms. In equilibrium $c_i^{\mathrm{PR}}(\mathbf{D},\phi)$ equals the fraction of all intermediate input sales made to, or purchases made by, firm i. Hence if $c_i^{\mathrm{PR}}(\mathbf{D},\phi)$ is large we might reasonably call firm i an ‘important’ or a *central intermediate input buyer*.

The work of Acemoglu et al. (2012) and Carvalho (2014) focuses on the macroeconomic implications of productivity shocks to intermediate goods producers. In that context, $c_i^{\mathrm{PR}}\left(\mathbf{D}', \phi\right)$, measures *supplier centrality*. Acemoglu et al. (2016) additionally explore the macroeconomic implications of firm-specific demand shocks. In that case, $c_i^{\mathrm{PR}}\left(\mathbf{D}, \phi\right)$, *buyer centrality*, plays a central role.

References

Acemoglu, D., Akcigit, U., Kerr, W., 2016. Networks and the macroeconomy: an empirical exploration. NBER Macroeconomics Annual 31 (1), 273–335.

Acemoglu, D., Carvalho, V., Ozdaglar, A., Tahbaz-Salehi, A., 2012. The network origins of aggregate fluctuations. Econometrica 80 (5), 1977–2016.

Angrist, J., 2014. The perils of peer effects. Labour Economics 30, 98–108.

Apicella, C.L., Marlowe, F.W., Fowler, J.H., Christakis, N.A., 2012. Social networks and cooperation in hunter-gatherers. Nature 481 (7382), 497–501.

Atalay, E., Hortaçsu, A., Roberts, J., Syverson, C., 2011. Network structure of production. Proceedings of the National Academy of Sciences 108 (13), 5199–5202.

Ballester, C., Calvó-Armengol, A., Zenou, Y., 2006. Who's who in networks. wanted: The key player. Econometrica 74 (5), 1403–1417.

Banerjee, A., Chandrasekhar, A.G., Dulfo, E., Jackson, M.O., 2013. The diffusion of microfinance. Science 341 (6144), 363–370.

Battiston, S., Puliga, M., Kaushik, R., Tasca, P., Caldarelli, G., 2012. Debtrank: Too central to fail? financial networks, the fed and systemic risk. Scientific Reports 2 (541).

Bengtsson, O., Hsu, D.H., 2015. Ethnic matching in the u.s. venture capital market. Journal of Business Venturing 30 (2), 338–354.

Bickel, P.J., Chen, A., 2009. A nonparametric view of network models and Newman-Girvan and other modularities. Proceedings of the National Academy of Sciences 106 (50), 21068–21073.

Blume, L.E., Brock, W.A., Durlauf, S.N., Jayaraman, R., 2015. Linear social interaction models. Journal of Political Economy 123 (2), 444–496.

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

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

Bonacich, P., Lloyd, P., 2001. Eigenvector-like measures of centrality for asymmetric relations. Social Networks 23 (3), 191–201.

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. Computer Networks 30 (1–7), 107–117.

Brock, W.A., Durlauf, S.N., 2001. Handbook of Econometrics. North-Holland, Amsterdam, pp. 3297–3380. volume 5, chapter Interactions-based models.

Calvó-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. The Review of Economic Studies 76 (4), 1239–1267.

Carvalho, V., 2014. From micro to macro via production networks. Journal of Economic Perspectives 28 (4), 23–48.

Carvalho, V.M., Nirei, M., Saito, Y.K., Tahbaz-Salehi, A., 2016. Supply chain disruptions: evidence from the great east japan earthquake. Cambridge University.

Cohen, L., Frazzini, A., 2008. Economic links and predictable returns. Journal of Finance 63 (4), 1977–2011.

Currarini, S., Jackson, M., Pin, P., 2009. An economic model of friendship: homophily, minorities and segregation. Econometrica 70 (4), 1003–1045.

De Weerdt, J., 2004. Insurance Against Poverty, chapter Risk-sharing and endogenous network formation. Oxford University Press, Oxford, pp. 197–216.

Denbee, E., Julliard, C., Li, Y., Yuan, K., 2014. Network risk and key players: A structural analysis of interbank liquidity. LSE Working Paper.

Diaconis, P., Janson, S., 2008. Graph limits and exchangeable random graphs. Rendiconti di Matematica 28 (1), 33–61.

Erdös, P., Rényi, A., 1959. On random graphs. Publicationes Mathematicae Debrecen 6, 290–297.

Erdös, P., Rényi, A., 1960. On the evolution of random graphs. Publications of the Mathematical Institute of the Hungarian Academy of Sciences 86 (5), 17–61.

Fafchamps, M., Lund, S., 2003. Risk sharing networks in rural Philippines. Journal of Development Economics 71 (2), 261–287.

Galeotti, A., Golub, B., Goyal, S., 2017. Targeting Interventions in Networks. Technical Report arXiv:1710.06026.

Gilbert, E., 1959. Random graphs. Annals of Mathematical Statistics 30 (4), 1141–1144.

Glaeser, E.L., Scheinkman, J.A., 2001. Social Dynamics, chapter Measuring social interactions. The MIT Press, Cambridge, MA, pp. 83–132.

Glaeser, E.L., Scheinkman, J.A., 2003. Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, volume 1, chapter Non-market interactions. Cambridge University Press, Cambridge, pp. 339–369.

Goldberger, A.S., 1991. A Course in Econometrics. Harvard University Press, Cambridge, MA.

Golub, B., Jackson, M.O., 2012. How homophily affects the speed of learning and best-response dynamics. Quarterly Journal of Economics 127 (3), 1287–1338.

Gould, P., 1967. On the geographical interpretation of eigenvalues. Transactions of the Institute of British Geographers 42, 53–83.

Graham, B., 2008. Identifying social interactions through conditional variance restrictions. Econometrica 76 (3), 643–660.

Graham, B., 2018. Identifying and estimating neighborhood effects. Journal of Economic Literature 56 (2), 450–500.

Graham, B.S., Imbens, G.W., Ridder, G., 2010. Measuring the effects of segregation in the presence of social spillovers: a nonparametric approach. Working Paper 16499. NBER.

Horn, R.A., Johnson, C.R., 2013. Matrix Analysis, 2nd edition. Cambridge University Press, Cambridge.

Jackson, M.O., López-Pintado, D., 2013. Diffusion and contagion in networks with heterogeneous agents and homophily. Network Science 1 (1), 49–67.

Jackson, M.O., Rogers, B.W., 2007. Relating network structure to diffusion properties through stochastic dominance. B.E. Journal of Theoretical Economics 7 (1), 6 (Advances).

Jackson, M.O., Zenou, Y., 2015. Handbook of Game Theory, vol. 4. Amsterdam. chapter Games on networks, (pp. 95–163). North-Holland.

Katz, L., 1953. A new status index derived from sociometric analysis. Psychometrica 18 (1), 39–43.

Kim, D.A., Hwong, A.R., Stafford, D., Hughes, D.A., O'Malley, A.J., Fowler, J.H., Christakis, N.A., 2015. Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. Lancet 386 (9989), 145–153.

Lovász, L., 2012. Large Networks and Graph Limits. American Mathematical Society Colloquium Publications., vol. 60. American Mathematical Society.

Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. Review of Economic Studies 60 (3), 531–542.

Marsden, P.V., 1987. Core discussion networks of Americans. American Sociological Review 52 (1), 122–131.

McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: homophily in social networks. Annual Review of Sociology 27 (1), 415–444.

Milgram, S., 1967. The small-world problem. Psychology Today 1 (1), 61–67.

Newman, M.E.J., 2010. Networks: An Introduction. Oxford University Press, Oxford.

Newman, M.E.J., 2016. Community detection in networks: Modularity optimization and maximum likelihood are equivalent. Physical Review E 94 (5), 052315.

Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The PageRank citation ranking: bringing order to the web. Technical report. Stanford University.

Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale-free networks. Physical Review Letters 86 (14), 3200–3203.

Pin, P., Rogers, B., 2016. The Oxford Handbook on the Economics of Networks. Oxford University Press, Oxford. chapter Stochastic network formation and homophily, (pp. 138–166).

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge.