Identifying key sectors in the regional economy: a network analysis approach using input–output data

Fernando DePaolis[1*], Phil Murphy[2] and M. Clara De Paolis Kaluza[3]

*Correspondence:
fdepaolis@middlebury.edu

[1] Center for the Blue Economy,
Middlebury Institute
of International Studies,
Monterey, CA, USA
[2] META Lab, Middlebury
Institute of International Studies,
Monterey, CA, USA
[3] Khoury College of Computer
Science, Northeastern University,
Boston, MA, USA

Abstract

By applying network analysis techniques to large input–output system, we identify key sectors in the local/regional economy. We overcome the limitations of traditional measures of centrality by using random-walk based measures, as an extension of Blöchl et al. (Phys Rev E 83(4):046127, 2011). These are more appropriate to analyze very dense networks, i.e. those in which most nodes are connected to all other nodes. These measures also allow for the presence of recursive ties (loops), since these are common in economic systems (depending to the level of aggregation, most firms buy from and sell to other firms in the same industrial sector). The centrality measures we present are well suited for capturing sectoral effects missing from the usual *output* and *employment* multipliers. We also develop and make available an R implementation for computing the newly developed measures.

Keywords: Input–output system, Regional economies, Multiplier effects

JEL Classification: C67, D85, P25

Introduction

Wassily Leontief won the Nobel prize in economics in 1973 for developing the Input-Output [I–O] modeling system, in the years before and during WWII. The original purpose of the I–O approach was to identify flows among economic sectors of a region or a country. Those flows represent the exchanges that take place among sectors. In market economies, this functions as a fair representation of sectors buying from and selling to one another. Over time it was clear that the identification of key sectors in the economy was emerging as a new and exciting application as shown in the early contributions of Rasmussen (1956), Laumas (1975), Schultz (1977) and Hewings (1982). In those early times, the implication was that any industrial policy intended to protect or stimulate specific sectors would start with the proper identification of their importance to the system. Today this application still remains relevant, even though I–O systems themselves have been superseded by more sophisticated and complex economic modeling approaches, such as general equilibrium (Jorgenson 2016). Local and regional policy makers and planners can certainly refer to the

DePaolis *et al. Applied Network Science*   (2022) 7:86

Page 2 of 19

sectors thus identified to allocate infrastructure funds, grant tax exemptions, employment support programs, and implement other policy actions geared toward strengthening the economic system and ensuring its resilience.

The identification of the most relevant sectors is not, however, an easy task. Is the most important sector the one that produces the highest output? or the one with the highest employment? or the one that buys the most from local suppliers? It is quite obvious that the choice of measurement will greatly affect the results; different sectors are "key" under different assumptions, and for different purposes.

The I–O approach provides a rather straight forward process to describing local economies. This was discovered very early by economic geographers, planners and others who began using it to, first, describe, and later analyze the relationships among local sectors and the emergence or presence of clusters or other groupings that were meaningful in the performance of the local/regional economy as a system (Hubbell 1965; Streit 1969; Roepke et al. 1974 and Campbell 1975).

In recent years, there has a been a growing interest in revisiting those early attempts, but this time with the aid of network analysis techniques. For the most part those efforts have focused on national economies, either at the country level (Giuliani 2013), the regional block level (Guo and Planting 2000; García Muñiz et al. 2008; Montresor and Marzetti 2009; Aroche Reyes and Marquez Mendoza 2012; García Muñiz 2013) or at the global scale (Blöchl et al. 2011). On the other hand, there are relatively few examples of this type of analysis at the sub-national level.

Therefore, there remains a clear and strong motivation to find alternative ways for the identification of key sectors in the local and regional economy, or at the very least the systematic analysis of the connections among sectors (Reid et al. 2008). One such approach is social network analysis (SNA). Though the roots of SNA, as a field, stretch back to the 1930s with Jacob Moreno's *sociometry* (Moreno 1934), the field experienced its greatest initial growth and expansion beginning with advances in computing power and ubiquity in the 1970s (Freeman 2004). Although the vast majority of early efforts in SNA applied to networks that are interpersonal and social in nature, the field has since bloomed and has become well-developed, with extensions into a variety of disciplines, such as biology, genetics, physics, and economics. At the core of the network analysis approach is a set of measures that provide a variety of conceptualizations of how one may operationalize the concept of relative prominence of each of the nodes that constitute the network. At its simplest, prominence may be measured as a function of *the frequency* in which a node is connected to others. Somewhat more nuanced conceptualizations of such measures, however, will consider the structure of the *pathways* that the ties create and *relative distances* between nodes in the network when traveling such paths. Frequency, path, and distance considerations have been incorporated into dozens of *centrality* measures (degree, closeness, betweenness, eigenvector, reach, and flow, just to mention some of the most commonly used measures). The concept of prominence within a network, as operationalized through centrality measures, remains fairly fluid and context-dependent. Even the earliest treatments on the topic, Freeman (1978) asserted that there was no unanimity on what centrality is or on the proper method(s) for its measurement. Despite attempts

DePaolis *et al. Applied Network Science*   (2022) 7:86

Page 3 of 19

to evaluate and prioritize various centrality measures (e.g., Harrison et al. 2016; Meng et al. 2017), context and interpretability remain the most valid determinants of their selection and application.

As we see it, we contribute to the advancement of the field by setting multiple objectives. First, we want to consolidate definitions and interpretations from existing frameworks. As it is common with methods that extend into other fields, multiple terms are used to refer to the same concept, and vice-versa. We are not aware of similar efforts. Second, we test whether the newly developed network metrics are applicable to I–O systems at the subnational level (state, and even municipal). This is not a technical issue, but one of interpretation. As we show, the same metrics have different values at the national, state and local level, hinting at the different role that specific industrial sectors play at different level of geographic aggregation. We have not identified other contributions in the literature that analyze subnational geographies in similar fashion. Finally, we implement all the metrics discussed in the paper in R. The code is publicly available and—with relatively little pre-processing—a user can compute all metrics for any network.[1]

Theoretical considerations

As discussed above, I–O systems can be represented by networks in which the nodes (also referred to as *vertices*) are economic sectors or industries, and the links connecting them (also conceptialized as *ties*) represent the flows among those industries. More precisely, I–O systems are *very dense, valued, directed* networks. In a network, density refers to the proportion of links that actually exist as a share of all possible links. I–O networks are *very dense* because–especially at high levels of aggregation–most, if not all, nodes (industries) will be connected to almost all other nodes.That is, network density (*d*) approaches 1 (in symbols, $d \rightarrow 1$).[2] They are *valued* networks because the links do not only represent the presence of a connection, but such a connection has a specific magnitude. Finally, they are also *directed* networks because I–O systems represent bi-directional flows between economic sectors. That is, each pair of nodes is connected by *two* links, one for each of the directions in which transactions may take place, typically with differing values. At greater levels of sectoral disaggregation, a more granular classification captures narrower and narrower definitions of industries and commodities. This results in more differentiation in the flows between sectors, with one direction potentially overshadowing the other by orders of magnitude (Lovász 2009; Miller and Blair 2009). Ties with similar values in both directions are extremely rare in I–O systems.

If the objective is to determine how important a sector is in the economic network, one may consider using vertex centrality measures such as those introduced by Freeman (Freeman 1977, 1978; Freeman et al. 1991). Common vertex centrality measures are of ambiguous applicability in the case of I–O networks. The difficulties associated with applying such vertex centrality measures to I–O systems become apparent when considering some common features that describe such networks. Of particular relevance

---

[1] NOTE: In the field of network analytics it is very common to find multiple definitions and implementations for the same metric, or one with very slight variations. Whenever possible, we refer to the earliest definition of a measure or concept.

[2] One reason for high density is high levels of sectoral aggreation. But even highly disaggregated systems, with hundreds of nodes, can have, for example $d > 0.7$, where 70% of all possible ties are present.

are the values of the ties between vertices, loops representing recursive trade within an industry, and the overall density of I–O networks.[3]

The simplest of the measures Freeman defined, *degree*, is calculated as the number of ties that are incident upon a node.[4] As such, the *degree* of a node describes how often each industry participates in the production function of others, and which sectors are part of its own production function. Given the high density found in I–O networks, however, the number of links that are incident upon any given node is not likely to vary greatly throughout the network, making *degree* a relatively poor measure of a given industry's relative prominence in a local economy.[5] In addition, because *degree* measures only direct access to others, it fails to capture the larger systemic effects that are distributed throughout the wider network.

Path-based measures were introduced to take into account a node's place within the larger network. Two measures that were introduced to take the entire network into account were *closeness* and *betweenness* Freeman (1978). *Closeness* provides a measure of the inverse distance between a node and all other nodes reachable from it. More specifically, *closeness* centrality for a given node *i* is calculated as

$$C_i^{CLO} = \left[\sum_{j=1}^{n} d_{(i,j)} \right]^{-1} \tag{1}$$

where d_{ij} is the distance of the shortest path (i.e., geodesic distance, or, simply, geodesic) between node *i* and any other node *j*. In this manner, closeness provides a measure of a node's strategic positioning within an network in terms of the speed or efficiency with which the flows within a network will pass through a particular node. Larger measures of closeness may, for example, indicate a node that will be able to access information or materials more frequently or quickly than others with lower values.

Another means of conceptualizing prominence of a node in terms of flows through a network is *betweenness*, which measures a node's potential for being able to capture, enable, or impede the passage of informaiton or materials in a network. As such, it is calculated as

$$C_i^{BET} = \frac{g_{jk}(i)}{g_{jk}} \tag{2}$$

where g_{jk} is the number of shortest paths (i.e., geodesics) between node *j* and node *k*, and $g_{jk}(i)$ is the number of those paths that include *i*. In this manner, *betweenness* measures the degree to which a particular node *i* commands strategic junctures within the network.

---

[3] Loops may disappear (or be drastically reduced) at higher levels sectoral aggregation.

[4] In directed networks, the degree measure can be separated into *in-degree* and *out-degree* to account for ties coming to or going from a node, respectively.

[5] When running the analysis at the 2-digic NAICS level (North American Industrial Classification System level which contains 20 sectors), we found degree values averaging around 19, meaning the almost every sector was connected to every other sectors.

DePaolis *et al. Applied Network Science*   (2022) 7:86

Page 5 of 19

A major weakness of both *betweenness* and *closeness*, as they are commonly used, is that neither measure was conceived to take the value of the ties into account. For I–O networks, this is a major shortcoming, as the actual magnitudes of intersectoral flows are a critical consideration. A solution has been proposed by Dijkstra (1959), Brandes (2001) and Newman (2001) and further modified by Opsahl et al. (2010) to seek the path with the least cumulative impedance, as opposed to the one with the fewest steps. The idea that drives this modification is that ties with lower values transmit less, and may be considered to impede flow more than ties with greater tie values. The resulting implementation is a weighted distance measure given by

$$d_{(i,j)}^{w\alpha} = min\left(\frac{1}{(w_{ih})^{\alpha}} + \cdots + \frac{1}{(w_{hj})^{\alpha}} \right) \tag{3}$$

where the inverse of the tie values (i.e., weights) w_{ij} is summed for each of the paths between node i and node j, and the path of least resistance (i.e., the one with the lowest value) is selected. The α coefficient functions as a tuning parameter that is used to either emphasize or deemphasize whether the number of steps should be taken into account. Setting $\alpha = 0$ produces the same measure as if the ties were of binary values; setting $\alpha = 1$ sums the inverse tie values; setting $0 < \alpha < 1$ favors fewer steps; and setting $\alpha > 1$ favors stronger tie weights in calculating shortest path distances.

Opsahl et al. (2010) employed the weighted geodesic measure shown in (3) in both *closeness* (4) and *betweenness* (5) in a manner that is fairly straightforward.

$$C_i^{WCLO} = \left[\sum_{j=1}^{n} d_{(i,j)}^{w\alpha} \right]^{-1} \tag{4}$$

$$C_i^{WBET} = \frac{g_{jk}^{w\alpha}(i)}{g_{jk}^{w\alpha}} \tag{5}$$

In each case, the weighted geodesic (i.e., shortest path) distance measure has been substituted for the binary form. In the case of *weighted closeness*, depending on the α setting, the relative distance in terms of summed inverse tie values is substituted for a count of the number of steps in each path when selecting the lowest value. For, *weighted betweenness*, on the other hand, $g_{jk}^{w\alpha}$ is a count of the number of geodesics occurring between node j and node k (Opsahl 2015).

Weighted path-based centrality measures hold the potential to reveal the relative prominence of nodes in a valued network. Each is well suited for use with valued networks, though there are some shortcomings for each that should be noted in regard to their potential for application in I–O networks. The characteristics of I–O networks that make them a challenge for both standard and weighted network metrics include the values of ties in I–O networks, the recursive loops present in aggregated networks, and their density merit consideration in modeling the flow of resources.

If one considers the weighted distance equation given in Eq. (3) to modify closeness (4) and betweenness (5), it should quickly become apparent that such a weighting metric will function in a manner similar to the measures designed for dichotomous ties (Eqs.

(1) and (2)) only when the tie values are limited to a relatively narrow range of integer values. Given that I–O networks are expected to take a theoretically unlimited range of positive continuous values, it becomes increasingly likely that the measure will produce one unique shortest path for any given node pair. Such a solution would emphasize the prominence of nodes that are situated in some of the most proiminent production sectors in the region being evaluated. Although prominence within key sectors will produce useful information, the lack of alternate shortest paths between nodes holds the potential to mask the relative importance of other nodes that may hold secondary importance–something that is also important from a planning and disaster mitigation standpoint. The ability to tune the measure using the α setting helps to reduce this tendency, but also requires a more standardized approach to tuning that has not yet been evaluated in I–O networks at this point.

An additional challenge to the anlaysis of I–O networks occurs when considering the recursive loops that are used to best model flows within a system that has been aggregated to create a set of nodes that would normally trade amongst themselves into a metanode that represents an entire industrial sector. Depending on the level of aggregation, the presence of loops within I–O networks can be substantial. However, both the classic binary and the weighted assessments of shortest paths through a network will logically always ignore such loops, as they would not normally constitute a "shorter" path through the network. This is not a realistic representation of the actual behavior of flows within an I–O network.

The final consideration that should be important to those modeling I–O networks is the high density of ties within the network. Using path-based measures that treat ties as binary tends to make it look as though all nodes are relatively "close" to one another since, on average, the paths throughout exceptionally dense networks will be of roughly the same length. This tendency is somewhat reduced by using weighted ties. When employing weighted measures, both closeness and betweenness will produce measures of prominence that are relative to the total network. They do not provide any special consideration for the more immediate neighborhood (e.g., manufacturing sector) around each node.

One way of avoiding most of those limitations is adopting measures of centrality based on random walks. Newman (2005) explored and further developed the notion of a measure of centrality based on random walks to overcome the need for a pre-determined, known flow from each source (*s*) to each target (*t*), and stated that "the random-walk betweenness of a vertex *i* is equal to the number of times that a random walk starting at *s* and ending at *t* passes through *i* along the way, averaged over all *s* and *t*."

Experts in I–O modeling, not yet familiar with network metrics, perceive the notion of random walk as one that does not correctly represent the relationships in the economic system. Their complaints are not without merit because, after all, industrial sectors do not consume a random collection of inputs in the hopes of producing a very specific output. Those input are indeed very well defined by a sector's production function. In the same vein, one could conceptualize random walks as representing a way of exhausting all possible combinations of inputs that produce all outputs in the economy. Because the computation of any random walk-based measure involves the "value" of the flow between two given sectors, only the ties that are present will influence the magnitude of

DePaolis *et al. Applied Network Science*   (2022) 7:86

Page 7 of 19

Table 1 Equivalency of measurements

García Muñiz et al. (2008)	Blöchl et al. (2011)
Immediate effects	Random walk centrality (analogous to closeness centrality)
Mediative effects	Counting betweenness (analogous to betweenness centrality)

the metric. Therefore, all random walked-based measures of an I–O system will reflect the true underlying production functions that link those sectors.

Network metrics that use random walks could identify those sectors that participate in all other sectors' production functions more often and with a greater insidence. This is one of the most significant differences between simple measures of centrality (which would identify the "presence" of the tie, in an on–off fashion) and random walk-based measures of centrality, which also include an apt representation of the intensity of the tie.

Methods

Following Friedkin (1991) we present a means of measuring *immediate effects* and *mediative effects*, in addition to the already defined *total effects*. Lee (2006) provides a more extended discussion about the equivalency between total, immediate and mediative effects and eigenvector, closeness and betweenness centrality, respectively. In this section we discuss approaches that provide a better definition of mediative and immediate effects within a network, as presented in García Muñiz et al. (2008) and Blöchl et al. (2011). Table 1 presents our own attempt to establish the equivalence between the measures proposed by each author.

What follows is a discussion of two alternative conceptualizations and corresponding definitions for *closeness* and *betweenness* centrality. In each case, these metrics provide a measurement that more closely mimics flows within I–O networks than the more widely shared (Freeman 1978) versions. We have elected to implement the definitions set by Blöchl et al. (2011). The metrics proposed in García Muñiz et al. (2008) are the inverse of those considered in Blöchl et al. (2011). This does not alter the outcome in terms of– for instance–sectoral rankings because the information transmitted by those metrics is equivalent.

One question remains. Is it appropriate to add the *immediate* and *mediative* effects to obtain a sort of *total* effects? Friedkin (1991) refers to total effects in the context of social networks but we have not found any work in which such definition is applied to I–O networks. Our cursory examination shows that immediate and mediative effects could show opposite directions, producing ambiguous interpretation of total effects. This is perhaps a computational confirmation of our intuition, which is that immediate and mediative effects correspond to two substantially different processes, and their combination is–at least–questionable. This is certainly an area for further investigation.

Random walk centrality for immediate effects

Sectors that have effects transmitted over long sequences of economic relations have lower economic impacts than those sectors that have a large number of direct linkages.

The measure *immediate effects* is the reciprocal of the mean length of the sequences of relations from the j*th* sector to all others.

Consider a weighted network, either directed or undirected, with n nodes denoted by j = 1,..., n; and a random walk process on this network with a transition matrix M.[6] The m_{jk} element of M describes the probability of the random walker that has reached node i, proceeds directly to node j. These probabilities are defined by Eq. (6).

$$M_{ij} = \frac{a_{ij}}{\sum_{j=1}^{n} a_{ij}} \tag{6}$$

where a_{ij} is the (i, j)*th* element of the weighting matrix A of the network. When there is no tie between two nodes, the corresponding element of the A matrix is zero. In the case of I–O system as we implement it, A is a matrix of regional absorption coefficients. Equation (7) shows the random walk closeness centrality of a node i as a function of the inverse of the average mean first passage time to that node.

$$C_i^{RWC} = \frac{n}{\sum_{j=1}^{n} H_{ij}} \tag{7}$$

The *mean first passage time* (MFPT) from node i to node j is the expected number of steps it takes for the process to reach node j from node i for the first time (Eq. (8)). Random walk centrality is the inverse of MFPT to a given sector. MFPT is the starting point in the computation of a random walk-based measure. Noh and Rieger (2004) define it as the expected number of steps a random walker who starts at source i needs to reach target j for the first time, H_{ij}.

$$H_{ij} = \sum_{j=1}^{\infty} r P_{ijr} \tag{8}$$

where P_{ijr} denotes the probability that it takes exactly r steps to reach j from i for the first time. To calculate these probabilities of reaching a node for the first time in r steps, it is useful to regard the target node as an absorbing one, and introduce a transformation of M (in Eq. (6) by deleting its j*th* row and column and denoting it by M_{-j}.

As the probability of a process starting at i and being in k after r-1 steps is simply given by the (i, k)*th* element of M_{-j}^{r-1}, $P(i, j, r)$ can be expressed as:

$$P_{ijr} = \sum_{k \neq j} ((M_{-j}^{r-1}))_{ik} m_{kj} \tag{9}$$

where m_{jk} is a column of M with the element m_{kk} deleted. Substituting Eq. (9) into equation(8), and vectorizing for computational convenience yields:

$$H(\bullet, j) = (I - M_{-j})^{-1} e \tag{10}$$

---

where $H(\bullet,j)$ is the vector of first passage times for a walk ending at node j, e is an n-1 dimensional vector of ones and I is the identity matrix. The result is then used in Eq. (7). The calculation of random walk betweenness centrality is very computationally intensive; for that reason we conducted most of our calculation at the 86-sector level of aggregation[7] rather at the 536-sector aggregation[8], which proved to be excessively challenging.

Counting betweenness for mediative effects

Mediative effects capture the prominence of sectors as instruments of transmission of total effects. The assumption is that it measures the involvement of a sector in the paths connecting other sectors, as if they were intermediaries in the transactions among other sectors. The more paths a sector participates in, the more its relevance as a connector or conduit in the overall economy. *Counting Betweenness* generalizes Newman's random walk for directed networks with loops. It measures how often a sector (a node in the network) is visited on first-passage walks, averaged over all source-target pairs, N_{ij}^{st}, and it is shown in Eq. (11).

$$C_i^{CB} = \frac{\sum_{j=1}^{n} \sum_{k \neq j}^{n} N^{jk}(i)}{n(n-1)} \tag{11}$$

where $N^{jk}(i)$ is a measure of the frequency with which a random walker reaches node i while going from j to k.[9]

Data and application to two cases

In the U.S., the commercial product IMPLAN® is regarded as an industry standard in terms of its widespread use as a tool for regional economic analysis. IMPLAN captures the inter-sector relationships with the parameter "gross absorption," which is the total amount of each commodity that is needed for production in one sector.[10] We decided to use "regional absorption," which is the amount of the required inputs (gross absorption) purchased locally.[11] This, we think, captures the relationships at the local level better than gross absorption, and is a better starting point in the identification of key sectors in the local economy. A sector's relative importance in the local economy increases as it uses more locally sourced inputs.

- *Gross Absorption* describes the total amount of each commodity that is needed in the production of a sector's output.
- *Regional Absorption* describes the amount of the required inputs that can be obtained locally (the total amount of production requirements that can be sourced within the model's geography).

---

[7] Equivalent to the 3-digit level of NAICS.

[8] Full specification of the IMPAN dataset.

[9] For a detail discussion of the calculations, see Blöchl et al. (2011). They provide the corresponding Matlab code, but it did not work as stated by the authors.

[10] IMPLAN granted us special permission to use internal components of their software for the development of our research, for which we are grateful.

[11] The relationship between gross and regional absorption is capture by "regional purchase coefficients".

We developed two different settings for the application of our algorithms. In the first case, we examine the economic structure of Monterey County, California. Here we show the differences between the more commonly used multipliers (total output and employment multipliers) and the centrality measures defined above. For the second application, we compare a metro area (Wayne County, Michigan, where the Detroit metropolitan region is located), a state (Michigan), and the whole country (United States). In this case, we show how the same centrality measures vary with the geographic scale and offer some interpretations.

At the time of the analysis, the IMPLAN system provided data at the local level, counties or zipcodes, using a 536-sector aggregation schema. Sectoral aggregations at 2- and 3-digit NAICS were also available. We produced datasets at all three levels: 2-digit NAICS (20 sectors); 3-digit NAICS (86 sectors); and native IMPLAN format (536 sectors). At the 86-sector level, the base matrix is a grid which cells are 86 by 86 matrices, that is, its full dimension is 7396 by 7396, or almost 55 million cells; at the 536-sector level, the full dimension is over 82 billion cells. These values are a measure of the computational demands under each aggregation schema.[12] The results presented below correspond to the 86-sector, 3-digit NAICS structure because we think that is the best sectoral map for this application. There are enough sectors to distinguish between subactivities that would be masked under the 20-sector schema, but not so many sectors that would result in the processing of huge matrices. The 3-digit NAICS aggregation has the additional advantage of providing a realistic description of almost any local or regional economy, with relatively very few missing sectors.

Results and interpretation

When applying our algorithm to the datasets in Blöchl et al. (2011), we replicated their results for a variety of countries. Although, they do not report the actual values of their metrics, we are confident our algorithm performs as expected because the high concordance between the rankings they report and our own computations. The R code is available on RPubs.[13]

For Monterey County, "Appendix 1" shows both sets of results, that is the traditional multipliers and the network metrics developed in our research. It shows the top-10 sectors, ranked by each of the measures (Output multiplier, Employment multiplier, Random Walk Centrality and Counting Betweenness). "Appendix 2" shows the ranking of all 86 sectors for the same four measures.

For the metropolitan area of Detroit, the state of Michigan, and the United States, "Appendix 3", "Appendix 4" and "Appendix 5" show the top ten sectors for Random Walk Centrality and Counting Betweenness, and the complete ranks in all three measures for all 86 sectors, respectively.

Both measures, Random Walk Centrality (RWC) and Counting Betweenness (CBET), are dynamic in nature as they capture the impacts on each sector as effects propagate throughout the local economy.[14] Random walk centrality measures to what extent a

---

[12] IMPLAN has since shifted to a browser-based implementation and the data structure might have changed.

[13] Contact the corresponding author for additional details.

[14] In this context, I–O modelers can simulate *"shocks"* to the systems by, for instance, increasing or decreasing the value of output in a sector

sector will be impacted earlier or later during the shock transmission process. Counting betweenness is a measure of impedance in the flow of a shock. A shock will reach a sector with a higher RWC before a sector with lower value of RWC. A shock will transit faster–or will reach more sectors from–an industry with a larger value of CBET. The commonly used output and employment multipliers cannot capture these dynamic effects. In fact, if the objective is to identify systemic weakness in the local economy, the multipliers could be misleading. Compared to a game of falling dominoes, the multipliers would be equivalent to the size of a single domino, while RWC and CBET would show how fast and in which direction the flow of falling dominoes will move, once the initial wave is set in motion (i.e. a shock).[15]

In the case of Monterey County, the sectors at the top of the RWC and CBET rankings are quite different from those for the output and employment multipliers. *Professional and Scientific Services* and *Management of Companies* are first and second for RWC and first and eight for CBET, while they are 32*nd* and 12*th*, and 41*st* and 50*th* for output and employment multipliers, respectively. One interpretation is that RWC and CBET for those sectors show that, although their output and employment levels are not very high, they do participate in many (or most) other sectors' production functions. This is one way of assigning preeminence to a sector in the context of the local economy.

When comparing the results for the Detroit metro region, the state of Michigan and the United States, the rankings of RWC and CBET show some interesting effects. For example, sector #9, Utilities, has RWC ranking values 8, 10 and 13 respectively for each geography. That would imply that at the smaller geography (Detroit metro), the "Utilities" sector plays a more important role than in the larger geographies (although it is quite important in all of them). The rankings in CBET show a similar effect. Conversely, for sector #20, Chemical Manufacturing, the effects seem to be opposite. The ranking is much lower for the Detroit metro area (70) than for the state and the nation (37 and 12, respectively). This means one of two things: (1) There is some production for the sector in the Detroit metro region but its output is consumed mostly outside the region, and therefore its local effect is lower; or (2) There is some local production but other local sectors utilize inputs from outside the region (this could be the result of the sectoral aggregation being too coarse to capture sub-sector effects). On the other end of the geographic scale, at the national level the Chemical Manufacturing sector provides inputs to all local production. That becomes meaningless when considering that "local" in this case refers to the whole of the United States.

The reader can derive similar inferences for all other sectors shown in "Appendix 5". Regardless of the remaining ambiguity, the interpretation confirms that both RWC and CBET measures performed as expected in identifying those sectors that are more relevant as contributors of locally sourced production. This attribute is not directly captured by output multipliers, which makes the identification of key sectors in the local economy more difficult.

---

[15] The description of the process gives the appearance of dynamics, but in reality we are just computing the aggregate effects as if it were happening instantaneously. In our metrics, there is no explicit modeling of time.

Closing remarks

Following the work of Blöchl et al. (2011) and García Muñiz et al. (2008), we implemented random walk-based measures of centrality in I–O networks to identify key sectors in the local and regional economy, using IMPLAN data and our own R code. The brief examination of the literature confirmed that the need for identifying key sectors remains and that improved network analysis techniques are apt instruments for the task. Thus we completed the three original objectives of our research (1) implement RWC and CBET in R; (2) test the implementation on subnational regional datasets; and (3) the presenting a preliminary consolidated version of network metrics from the existing literature.

Future explorations include the analysis of other applications of these network metrics, such as identification of industrial clusters.We posit that the measures developed in this paper could uncover previously unidentified clusters and analyze their evolution over time; both could be instrumental in informing future policy actions.

Appendix 1: Top 10 sectors by selected indicators for Monterey County

Order	Output multiplier	Employment multiplier	Random walk centrality	Counting betweenness
1	Securities other financial	Social assistance	Professional, scientific, tech svcs	Professional, scientific, tech svcs
2	Funds, trusts, other finan	Misc retailers	Management of companies	Real estate
3	Electronics, appliances stores	Personal, laundry svcs	Real estate	Admin support svcs
4	Broadcasting	Private households	Admin support svcs	Wholesale Trade
5	Food products	Admin support svcs	Wholesale Trade	Construction
6	Personal, laundry svcs	Educational svcs	Construction	Insurance carriers, related
7	Social assistance	Nursing, residential care	Government, non NAICs	Monetary authorities
8	Sightseeing transportation	Sports, hobby, book, music stores	Monetary authorities	Management of companies
9	Religious, grantmaking, similar orgs	Food svcs, drinking places	Securities, other financial	Securities, other financial
10	Telecommu-nications	Ag, Forestry Svcs	Food svcs, drinking places	Government, non NAICs

Appendix 2: Complete rankings for Monterey County

Sector ID	Sector description	Output multiplier	Employment multiplier	Random walk centrality	Counting betweenness centrality
1	Crop Farming	66	36	38	37
2	Livestock	20	66	49	48
3	Forestry & Logging	35	29	78	79

Sector ID	Sector description	Output multiplier	Employment multiplier	Random walk centrality	Counting betweenness centrality
4	Fishing- Hunting & Trapping	70	19	79	72
5	Ag & Forestry Svcs	55	10	81	49
6	Oil & gas extraction	72	68	35	34
7	Mining	83	73	71	71
8	Mining services	58	53	37	31
9	Utilities	85	85	27	20
10	Construction	23	44	6	5
11	Food products	5	62	26	23
12	Beverage & Tobacco	36	70	64	64
13	Textile Mills	78	58	74	75
14	Textile Products	42	43	70	70
15	Leather & Allied	15	31	80	80
16	Wood Products	59	46	65	65
17	Paper Manufacturing	71	74	75	76
18	Printing & Related	47	37	56	57
19	Petroleum & coal prod	27	83	59	59
20	Chemical Manufacturing	79	84	46	44
21	Plastics & rubber prod	81	69	76	77
22	Nonmetal mineral prod	69	65	72	73
23	Primary metal mfg	61	61	69	69
24	Fabricated metal prod	74	55	60	60
25	Machinery Mfg	75	76	30	28
26	Computer & oth electron	65	71	36	35
27	Electircal eqpt & appliances	77	67	77	78
28	Transportation eqpmt	80	80	53	51
29	Furniture & related prod	60	47	58	58
30	Miscellaneous mfg	51	51	68	68
31	Wholesale Trade	54	60	5	4
32	Motor veh & parts dealers	67	40	33	36
33	Furniture & home furnishings	37	27	61	61
34	Electronics & appliances stores	3	11	66	66
35	Bldg materials & garden dealers	33	25	45	46
36	food & beverage stores	50	15	51	52
37	Health & personal care stores	26	20	39	39
38	Gasoline stations	34	34	43	43
39	Clothing & accessories stores	49	22	21	24
40	Sports- hobby- book & music stores	38	8	52	53
41	General merch stores	48	16	31	32
42	Misc retailers	17	2	41	41
43	Non-store retailers	68	35	15	16
44	Air transportation	41	64	32	33
45	Rail Transportation	29	59	62	62
46	Water transportation	86	86	86	86
47	Truck transportation	28	39	19	22
48	Transit & ground passengers	40	17	20	30
49	Pipeline transportation	45	63	73	74
50	Sightseeing transportation	8	32	18	15
51	Couriers & messengers	18	14	24	25
52	Warehousing & storage	30	33	28	27

Sector ID	Sector description	Output multiplier	Employment multiplier	Random walk centrality	Counting betweenness centrality
53	Publishing industries	76	72	57	56
54	Motion picture & sound recording	82	77	34	26
55	Broadcasting	4	56	47	42
56	Internet publishing and broadcasting	62	82	12	13
57	Telecommunications	10	52	48	47
58	Internet & data process svcs	64	81	63	63
59	Other information services	43	75	50	50
60	Monetary authorities	16	45	8	7
61	Credit inmediation & related	19	48	14	14
62	Securities & other financial	1	13	9	9
63	Insurance carriers & related	57	54	13	6
64	Funds- trusts & other finan	2	18	55	55
65	Real estate	84	78	3	2
66	Rental & leasing svcs	63	49	17	21
67	Lessor of nonfinance intang assets	56	79	16	18
68	Professional- scientific & tech svcs	32	41	1	1
69	Management of companies	12	50	2	8
70	Admin support svcs	39	5	4	3
71	Waste mgmt & remediation svcs	53	57	22	17
72	Educational svcs	21	6	54	54
73	Ambulatory health care	25	30	67	67
74	Hospitals	22	38	82	81
75	Nursing & residential care	13	7	84	83
76	Social assistance	7	1	84	83
77	Performing arts & spectator sports	14	21	23	19
78	Museums & similar	11	26	25	83
79	Amusement- gambling & recreation	31	23	44	45
80	Accommodations	46	24	29	29
81	Food svcs & drinking places	24	9	10	11
82	Repair & maintenance	52	28	11	12
83	Personal & laundry svcs	6	3	42	40
84	Religious- grantmaking- & similar orgs	9	12	40	38
85	Private households	44	4	86	86
86	Government & non NAICs	73	42	7	10

Appendix 3: Top 10 sectors by *random walk centrality* for Detroit (Wayne County), State of Michigan, and the United States

Order	Detroit	Michigan	United States
1	Professional- scientific and tech svcs	Professional- scientific and tech svcs	Professional- scientific and tech svcs
2	Management of companies	Management of companies	Management of companies
3	Real estate	Real estate	Real estate
4	Admin support svcs	Admin support svcs	Admin support svcs
5	Construction	Construction	Wholesale Trade
6	Wholesale Trade	Wholesale Trade	Construction
7	Petroleum and coal prod	Insurance carriers and related	Petroleum and coal prod

Order	Detroit	Michigan	United States
8	Utilities	Monetary authorities	Securities and other financial
9	Monetary authorities	Securities and other financial	Monetary authorities
10	Insurance carriers and related	Utilities	Oil and gas extraction

Appendix 4: Top 10 sectors by *counting betweenness* for Detroit (Wayne County), State of Michigan, and the United States

Order	Detroit	Michigan	United States
1	Professional- scientific and tech svcs	Professional- scientific and tech svcs	Professional- scientific and tech svcs
2	Utilities	Insurance carriers and related	Insurance carriers and related
3	Real estate	Real estate	Real estate
4	Admin support svcs	Admin support svcs	Admin support svcs
5	Insurance carriers and related	Utilities	Utilities
6	Construction	Construction	Chemical Manufacturing
7	Wholesale Trade	Wholesale Trade	Wholesale Trade
8	Petroleum and coal prod	Monetary authorities	Securities and other financial
9	Management of companies	Securities and other financial	Management of companies
10	Monetary authorities	Management of companies	Monetary authorities

Appendix 5: Complete rankings for Detroit (Wayne County), State of Michigan, and the United States

Sector ID	Sector description	Random walk centrality			Counting betweenness		
		Detroit	Michigan	United States	Detroit	Michigan	United States
1	Crop Farming	11	61	55	70	59	54
2	Livestock	78	69	58	76	64	55
3	Forestry and Logging	83	65	66	81	56	58
4	Fishing- Hunting and Trapping	82	68	69	77	79	77
5	Ag and Forestry Svcs	80	67	67	78	66	65
6	Oil and gas extraction	31	23	10	28	24	14
7	Mining	79	46	29	79	46	25
8	Mining services	63	35	14	60	33	13
9	Utilities	8	10	13	2	5	5
10	Construction	5	5	6	6	6	11
11	Food products	39	48	41	31	45	35
12	Beverage and Tobacco	71	63	62	68	62	61
13	Textile Mills	75	78	73	73	76	70
14	Textile Products	74	81	79	72	80	78
15	Leather and Allied	81	82	82	80	81	81
16	Wood Products	62	44	49	57	42	42
17	Paper Manufacturing	77	53	35	75	49	32
18	Printing and Related	57	54	45	55	54	44

Sector ID	Sector description	Random walk centrality			Counting betweenness		
		Detroit	Michigan	United States	Detroit	Michigan	United States
19	Petroleum and coal prod	7	20	7	8	20	12
20	Chemical Manufacturing	70	37	12	67	34	6
21	Plastics and rubber prod	33	41	31	29	40	31
22	Nonmetal mineral prod	76	34	42	74	32	41
23	Primary metal mfg	47	49	23	43	47	19
24	Fabricated metal prod	45	29	18	42	30	17
25	Machinery Mfg	18	55	21	15	55	22
26	Computer and oth electron	40	73	25	34	70	18
27	Electircal eqpt and appliances	35	64	51	32	65	52
28	Transportation eqpmt	23	33	34	17	29	28
29	Furniture and related prod	50	74	64	48	72	62
30	Miscellaneous mfg	66	75	65	62	73	64
31	Wholesale Trade	6	6	5	7	7	7
32	Motor veh and parts dealers	36	38	57	36	38	57
33	Furniture and home furnishings	69	76	78	66	74	76
34	Electronics and appliances stores	73	79	81	71	78	80
35	Bldg materials and garden dealers	52	56	74	49	57	72
36	food and beverage stores	61	71	76	59	69	74
37	Health and personal care stores	44	45	63	41	48	66
38	Gasoline stations	55	58	72	53	58	71
39	Clothing and accessories stores	25	30	50	25	31	51
40	Sports- hobby- book and music stores	60	66	77	58	67	75
41	General merch stores	34	36	56	33	35	56
42	Misc retailers	46	51	71	46	51	69
43	Non-store retailers	27	26	43	27	27	43
44	Air transportation	42	25	33	38	25	37
45	Rail Transportation	54	62	46	51	63	47
46	Water transportation	68	72	70	65	71	68
47	Truck transportation	16	17	27	16	17	30
48	Transit and ground passengers	17	18	28	39	44	46
49	Pipeline transportation	41	39	48	37	39	48
50	Sightseeing transportation	49	24	30	47	22	24
51	Couriers and messengers	24	22	32	24	19	33
52	Warehousing and storage	22	28	38	21	28	39
53	Publishing industries	59	42	52	56	37	50

Sector ID	Sector description	Random walk centrality			Counting betweenness		
		Detroit	Michigan	United States	Detroit	Michigan	United States
54	Motion picture and sound recording	51	47	47	44	41	38
55	Broadcasting	58	57	54	54	52	49
56	Internet publishing and broadcasting	21	15	16	19	11	15
57	Telecommunications	43	19	24	40	21	27
58	Internet and data process svcs	67	52	61	64	53	63
59	Other information services	56	43	40	52	43	40
60	Monetary authorities	9	8	9	10	8	10
61	Credit inmediation and related	19	14	19	18	15	20
62	Securities and other financial	14	9	8	11	9	8
63	Insurance carriers and related	10	7	11	5	2	2
64	Funds- trusts and other finan	64	59	68	61	60	67
65	Real estate	3	3	3	3	3	3
66	Rental and leasing svcs	20	21	26	20	23	29
67	Lessor of nonfinance intang assets	26	16	22	26	16	26
68	Professional- scientific and tech svcs	1	1	1	1	1	1
69	Management of companies	2	2	2	9	10	9
70	Admin support svcs	4	4	4	4	4	4
71	Waste mgmt and remediation svcs	29	27	36	23	18	36
72	Educational svcs	65	70	75	63	68	73
73	Ambulatory health care	72	80	80	69	77	79
74	Hospitals	84	83	83	82	82	82
75	Nursing and residential care	84	83	83	83	83	83
76	Social assistance	84	83	83	83	83	83
77	Performing arts and spectator sports	28	31	37	22	26	34
78	Museums and similar	30	32	39	83	83	83
79	Amusement- gambling and recreation	53	60	60	50	61	60
80	Accommodations	32	77	44	30	75	45
81	Food svcs and drinking places	13	12	17	13	13	21
82	Repair and maintenance	15	13	20	14	14	23
83	Personal and laundry svcs	48	50	59	45	50	59
84	Religious- grantmaking- and similar orgs	37	40	53	35	36	53
85	Private households	38	86	86	86	86	86

Author contributions
DePaolis and Murphy contributed equally. De Paolis Kaluza contributed to the code development. All authors read and approved the final manuscript.

Availability of data and materials
The R code is available on RPubs (https://rpubs.com/RStudio_knight/368268). Contact the corresponding author for sample datasets and additional details.

Declarations

Competing interests
The authors declare that there are no competing interests or conflicts related to this research.

References

Aroche Reyes F, Marquez Mendoza MA (2012) An economic network in North America. MPRA Paper 61391, University Library of Munich, Germany

Blöchl F, Theis FJ, Vega-Redondo F, Fisher E (2011) Vertex centralities in input–output networks reveal the structure of modern economies. Phys Rev E 83(4):046127

Blum A, Hopcroft J, Kannan R (2015) Foundations of data science

Brandes U (2001) A faster algorithm for betweenness centrality. J Math Sociol 25:163–177

Campbell J (1975) Application of graph theoretic analysis to inter-industry relationships. Reg Sci Urban Econ 5:91–106

Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1:269–271

Freeman L (1977) A set of measures of centrality based on betweenness. Sociometry 40(1):35–41

Freeman L (1978/79) Centrality in social networks conceptual clarification. Soc Netw 1(3):215–239

Freeman L, Borgatti S, White D (1991) Centrality in valued graphs: a measure of betweenness based on network flow. Soc Netw 13:141–154

Freeman LC (2004) The development of social network analysis: a study of the sociology of science. Empirical Press, Vancouver

Friedkin N (1991) Theoretical foundations for centrality measures. Am J Sociol 96(6):1478–1504

García Muñiz A (2013) Modelling linkages versus leakages networks: the case of Spain. Reg Sect Econ Stud 13(1):43–54

García Muñiz A, Morillas Raya A, Ramos Carvajal C (2008) Key sectors: a new proposal from network theory. Reg Stud 42(7):1013–1030

Giuliani E (2013) Network dynamics in regional clusters: evidence from Chile. Res Policy 42(8):1406–1419

Guo J, Planting M (2000) Using input–output analysis to measure the U.S. economic structural change over a 24 year period. BEA Papers 0004, Bureau of Economic Analysis

Harrison K, Ventresca M, Ombuki-Berman B (2016) A meta-analysis of centrality measures for comparing and generating complex network models. J Comput Sci 17:205–215

Hewings GJD (1982) The empirical identification of key sectors in an economy: a regional perspective. Dev Econ 20(2):173–195

Hubbell C (1965) An input–output approach to clique identification. Sociometry 28(4):377–399

Jorgenson D (2016) Econometric general equilibrium modeling. J Policy Model 38(3):436–447

Laumas P (1975) Key sectors in some underdeveloped countries. Kyklos 28(1):62–79

Lee C-Y (2006) Correlations among centrality measures in complex networks. ArXiv Physics e-prints

Lovász L (2009) Very large graphs. Curr Dev Math 2008:67–128

Meng F, Gu Y, Fu S, Wang M, Guo Y (2017) Comparison of different centrality measures to find influential nodes in complex networks. In: Wang G, Atiquzzaman M, Yan Z, Choo K (eds) Security, privacy, and anonymity in computation, communication, and storage. Springer, pp 415–423

Miller R, Blair P (2009) Input–output analysis, foundations and extensions. Cambridge University Press, Cambridge

Montresor S, Marzetti G (2009) Applying social network analysis to input–output based innovation matrices: an illustrative application to six OECD technological systems for middle 1990s. Econ Syst Res 21(2):129–149

Moreno J (1934) Who shall survive? Nervous and Mental Disease Publishing Company, Washington

Newman MEJ (2001) Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys Rev E 64:016132

Newman M (2005) A measure of betweenness centrality based on random walks. Soc Netw 27(1):39–54

Noh J, Rieger H (2004) Random walks on complex networks. Phys Rev Lett 92:118701

Opsahl T (2015) tnet: software for analysis of weighted, two-mode, and longitudinal networks

Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: generalizing degree and shortest paths. Soc Netw 32:245–251

Rasmussen PN (1956) Studies in inter-sectoral relations. E. Harck, Copenhagen

Reid N, Smith B, Carroll M (2008) Cluster regions. Econ Dev Q 22(4):345–352

Roepke H, Adams D, Wiseman R (1974) A new approach to the identification of industrial complexes using input–output data. J Reg Sci 14(1):15–29

Schulman LS (2016) Transition matrix from a random walk. ArXiv e-prints arXiv:1605.04282
Schultz S (1977) Approaches to identifying key sectors empirically by means of input–output analysis. J Dev Stud 14(1):77–96
Streit M (1969) Spatial associations and economic linkages between industries. J Reg Sci 9(2):177–187

Publisher's Note