

A Topological Centrality Measure for Directed Networks

Fenghuan He*

December 24, 2021

Abstract

Given a directed network G , we are interested in studying the qualitative features of G which govern how perturbations propagate across G . Various classical centrality measures have been already developed and proven useful to capture qualitative features and behaviors for undirected networks. In this paper, we use topological data analysis (TDA) to adapt measures of centrality to capture both directedness and non-local propagating behaviors in networks. We introduce a new metric for computing centrality in directed weighted networks, namely the *quasi-centrality* measure. We compute these metrics on trade networks to illustrate that our measure successfully captures propagating effects in the network and can also be used to identify sources of shocks that can disrupt the topology of directed networks. Moreover, we introduce a method that gives a hierarchical representation of the topological influences of nodes in a directed network.

Key Words: Complex Networks, Centrality, Directed Networks, Trade Networks, Topological Data Analysis, Persistent Homology, Hierarchical Clustering

Contents

1	Introduction	2
2	Preliminary	3
2.1	Networks	3
2.2	Existing centrality measures	4
2.3	Algebraic topology and homology	5
2.4	Persistent homology	6
3	Quasi-centrality measure	8
4	Quasi-centrality applications	11
4.1	2007 Asia machinery production network	13
4.2	2011 Asia machinery production network	15
4.3	Summary	18
5	Clustering nodes by topological influences	19
5.1	Node hierarchy using TDA	19
5.2	Hierarchical clustering of the 2007 Asia machinery production network	21
5.3	Hierarchical clustering for the 2011 Asia machinery production network	24
5.4	Summary	27

*Commonwealth School, Boston, MA. Email: lhe@commschool.org

1 Introduction

Networks are a useful abstraction for many real-world systems, representing interactions between objects within a system. Network analysis examines relationships among entities, such as persons, organizations, or documents [Wri15] and has been extensively adopted for modeling systems in various domains with a long history of applications [New03] [Cos+11], including neuroscience [Si+21], biology [Lei+20], and social networks [OR02]. Networks exist in various forms: weighted or unweighted, directed or undirected. However, many complex systems are more accurately modeled by directed weighted networks, where the relation between two different entities in the system is asymmetric and exists in a range of intensities. For example, citation networks, immigration networks, and trade networks are all accurately modeled by directed weighted networks.

Directed networks are mathematically represented as a directed graph where vertices represent the objects or entities in the system, and edges encode the interaction between individual objects. Previous studies have implemented a wide range of measures for studying networks [Rid05], including node centrality [Fic+21], clustering coefficients [Mas+18] and path lengths between nodes [MG16]. In this paper, we focus on analyzing node centrality in directed weighted networks.

Node centrality measures the influence of a node in the entire network by assigning rankings and numbers to nodes within the network based on their network position. Centrality measures enable us to detect various real-world phenomena, including the identification of banks that are too-connected-to-fail [Gof17] and decisions regarding human capital or education [Hab+15]. Different centrality measures have been developed to capture different behaviors. For example, betweenness centrality has been widely used in social networks [Lee+21], eigenvector centrality has proved to be useful in temporal networks [Tay+16], optimal percolation centrality has been widely used in large networks [LSV20], and K-core centrality has been widely used in dynamic networks [Liu+20].

In this paper, we are interested in studying the propagating properties of a network since influences of individuals in many real-world complex systems have the potential to propagate over the entire system. This phenomena can be meaningfully illustrated on the global trade network, where globalization of trade and intertwined economies around the world can cause economic perturbations originated in a single country propagate elsewhere. Moreover, since world economies are exhibiting increasing levels of local heterogeneity and global interdependency [SBV07], it is crucial to take a topological standpoint on defining the centrality measure for understanding propagating effects in trade networks. Thus, we define a new centrality measure using persistent homology, namely the *quasi-centrality* (Definition 3.4) to determine the ranks of the nodes in directed weighted networks based on this property.

Persistent homology is a central tool used in topological data analysis (TDA), which is a burgeoning field in math and data analysis where concepts from algebraic topology are used to simplify, summarize, and compare complex data sets. TDA has found successful applications in neuroscience [Dab20], biological models [TZH15] machine learning [Che+15], and other related fields in statistics, math, physics, and biology. However, while many applications of TDA have focused on simplifying and identifying the intrinsic shapes of complex data sets, it has not been extensively applied to analyze networks. Hence the significance of our work is shown by utilizing TDA concepts in examining network properties.

The quasi-centrality measure (Definition 3.4) is defined based on the idea that if a node plays a crucial role in the connectivity of the network, we would expect perturbations originating from this node to propagate significantly through the network. Roughly, we measure a node's role in the overall connectivity of a directed network by computing the difference between the connectivity

of the network before and after deleting this node. In particular, we use the “size” of homology groups as a proxy to determine the connectivity of a directed network (Remark 3.5).

We show that the quasi-centrality is optimal for measuring the influence of perturbations originated from individual nodes by (1) computing the quasi-centralities on the star-shape network in Example 3.8 and comparing with existing centrality measures, and (2) analyzing the importance of country-industry pairs in the Asia machinery production network (Section 4) by computing the quasi-centralities and comparing with existing centrality measures.

Moreover, given a directed weighted network G , we introduce a method (Definition 5.6) that extracts a hierarchical representation of nodes in G based on their topological impacts. This method combines the bottleneck distance (Definition 5.5), a tool used in TDA, with hierarchical clustering, a method used in cluster analysis that determines similarity between objects.

We present our findings as follows: in Section 2, we provide background knowledge in networks and existing centrality measures, algebraic topology, and TDA (specifically persistent homology). In Section 3, we define quasi-centrality (Definition 3.4). In Section 4, we illustrate the practicality of quasi-centrality by computing it on the Asia machinery production network. In Section 5, we present our method that determines the hierarchy of nodes in a given directed weighted network, and analyze the hierarchical dendrograms extracted from the Asia machinery production networks. And finally in Section 6, we summarize our results and discuss future directions.

Data and implementations: Our data sets and software are available on <https://github.com/lindahe8989/A-topological-centrality-measure-for-directed-networks>.

For plotting persistence diagrams and hierarchical dendrograms, we partially used <https://github.com/fmemoli/PersNet> for reference.

2 Preliminary

In this section, we provide background knowledge in algebraic topology and persistent homology, as well as related work in network analysis. We begin by introducing networks.

2.1 Networks

Definition 2.1. A *network* is a pair (X, w_X) where X is a finite set and $w_X : X \times X \rightarrow \mathbb{R}$ is called the weight function. An *undirected network* is a network such that $w_X(x_1, x_2) = w_X(x_2, x_1)$ for all $x_1, x_2 \in X$, and a *directed network* is a network such that $w_X(x_1, x_2) \neq w_X(x_2, x_1)$ for some $x_1, x_2 \in X$.

Our definition of quasi-centrality (Definition 3.4) is restricted to dissimilarity networks, introduced below:

Definition 2.2. A *dissimilarity network* is a network $G = (X, w_X)$ where $w_X : X \times X \rightarrow \mathbb{R}$ is called a *dissimilarity function*, i.e. a map such that $w_X(x, x') = 0$ if and only if $x = x'$ for all $x, x' \in X$.

Remark 2.3. A dissimilarity network can be mathematically represented as a directed graph without self-loops.

Given a directed network $G = (X, m_X)$, the dissimilarity function $m_X(x_1, x_2)$ describes the intensity of interaction between the two nodes. We can associate a distance metric w_X to G such that if two nodes interact frequently, i.e., $m_X(x_1, x_2) \gg 0$, then they are close to each other in this distance metric, i.e., $w_X(x_1, x_2) \approx 0$. We adapt the definition of *effective distance* [BH13]¹,

¹Successful applications include [Ian+17], [BH13], etc.

which has been proved to have successfully interpreted the relationship between distance and the volume of interaction between two nodes into our distance metric m_X as follows:

Definition 2.4. Let $G = (X, w_X)$ be a (dissimilarity) network. Define $\gamma(G)$ to be (X, m_X) where $m_X : X \times X \rightarrow \mathbb{R}$ is given by:

$$m(x_i, x_j) = \begin{cases} 1 - \log \frac{w(x_i, x_j)}{\sum_{k \neq j} w(x_k, x_j)} \geq 1 & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases}$$

Hence in $\gamma(G)$, two vertices x_i, x_j with large volume of interaction $w(x_i, x_j)$ will correspond to a closer distance $m(x_i, x_j)$.

Remark 2.5. The notion of effective distance will be used heavily in our definition of quasi-centrality (Definition 3.4).

Remark 2.6. For simplicity in our calculation, when $w(x_i, x_j) = 0$, we choose a large number for $m(x_i, x_j)$, i.e., we let the distance between nodes x_i, x_j in the new metric m be very far apart so that it is almost indistinguishable when no relation exists between the two nodes. In Example 3.8, we let $m(x_i, x_j) = 24.026$.

2.2 Existing centrality measures

Given a network $G = (X, w_X)$, we can study the qualitative features of the network. One such interesting feature is to detect the importance of individual nodes in the network. In particular, *node centrality* measures have been extensively applied on empirical networks to detect ranks and influence of individual nodes. Here we summarize existing measures of node centrality for directed networks. In Section 5 we will use these as a comparison with quasi-centrality (Definition 3.4).

In the following definitions, we let $G = (X, w_X)$ be a directed network and A_{ij} be its adjacency matrix.

- *Degree centrality:* for a directed network G , the *in-degree centrality* and *out-degree centrality* for a node $i \in X$ are defined to be the number of edges going in or out of i .
- *Katz centrality:* for a directed network G , the the Katz centrality x_i for node $i \in X$ is defined to be $x_i = \alpha \sum_j A_{ij} x_j + \beta$, where α and β are constants and A_{ij} is an element of the adjacency matrix.
- *Pagerank centrality:* for a directed network G , the Pagerank centrality x_i for node $i \in X$ is defined to be $x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta$, where α and β are constants, A_{ij} is an element of the adjacency matrix, and k_j^{out} is the out-degree of the node i .
- *HITS hubs and authorities centrality:* The HITS hubs and the HITS authorities centrality measures are defined such that one measure depends on the centrality measure determined by the other measure.

Let α and β be non-negative constants. For a directed network G , the HITS-authority centrality x_i for node $i \in X$ is defined to be the sum of the hub centralities y_j which point to the node i : $x_i = \alpha \sum_j A_{ij} y_j$, and the HITS-hubs centrality y_i for node $i \in X$ is defined to be the sum of the authority centrality x_j which are pointed by the node i : $y_i = \beta \sum_j A_{ji} x_j$.

We note that all of the above centrality measures are based on the adjacency matrix A_{ij} , showing that the centrality measure for a node i is completely dependent on the direct neighbors of the node, while not taking the propagating influence of the node on the entire network into account. The goal of our research is to provide a centrality measure that takes propagating influences into account, and we do so by using algebraic topology and persistent homology, which we will provide background knowledge in the following section.

2.3 Algebraic topology and homology

We begin by introducing simplicial complexes, which are the foundational objects of study in algebraic topology and homology.

Definition 2.7. Given a finite set T , an (abstract) simplicial complex X is a subset of $\text{Pow}(T)$, the power set on T , such that

- singletons $\{x\} \subseteq T$ belong to X , and
- whenever $\sigma \in \text{Pow}(T)$ belongs to X , then any subset $\tau \subseteq \sigma$ belongs to X as well. We refer to a proper subset τ of σ as a *face* of σ .

The elements of T are called the *vertices* of X , the two-element subsets of T that belong to X are called the *edges*, and in general the k -element subsets of T are called the $(k-1)$ -simplices in X .

Given a simplex $\sigma = \{x_0, \dots, x_n\}$, an *orientation* of σ is a choice of an equivalence class of the orderings of its vertices. We say that two orderings are *equivalent* if they differ by an even permutation. For $n > 0$, there are thus two equivalence classes of orderings. A choice of one of the classes will be called an orientation of σ . If $\sigma = \{v_0, \dots, v_n\}$, then σ with a choice of an orientation will be denoted by $\sigma = [v_0 : \dots : v_n]$.

We now define homology for simplicial complexes. Let X be a simplicial complex, and let X_n denote the set of oriented n -simplices in X , then for any dimension $n \in \mathbb{Z}_+$, the *vector space of n -chains* of X over the field \mathbb{K} , denoted by $C_n(X, \mathbb{K})$, is the free \mathbb{K} -vector space with basis the set of oriented n -simplices:

$$C_n(X, \mathbb{K}) = \left\{ \sum \alpha_i \sigma_i : \sigma_i \in X_n, \alpha_i \in \mathbb{K} \right\}.$$

Definition 2.8. Given any n -chain vector space $C_n(X, \mathbb{K})$, where $n \in \mathbb{Z}_+$, we define the boundary map $\partial_n : C_n(X, \mathbb{K}) \rightarrow C_{n-1}(X, \mathbb{K})$ as the linear transformation specified on the generators as

$$\partial_n([v_0, \dots, v_n]) = \sum_{i=1}^n (-1)^i [v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_n].$$

Informally, ∂_n maps each $(n+1)$ -simplex to its boundary comprised of its faces. It is easy to check that the boundary of a boundary is empty, and hence the linear maps ∂_n satisfy the property that composing any two consecutive boundary maps yields the zero map:

Proposition 2.9. [Mun84] For all $n \in \mathbb{Z}_+$, $\partial_{n-1} \circ \partial_n = 0$.

We axiomatize the properties of $C_n(X; \mathbb{K})$ in the definition of a chain complex. Namely, a *chain complex of vector spaces* A_\bullet is a collection of vector spaces $\{A_n\}$ such that the linear transformations $\partial_n : A_n \rightarrow A_{n-1}$ satisfy $\partial_{n-1} \circ \partial_n = 0$ for all $n \in \mathbb{Z}_+$. From the chain complex A_\bullet , one may define the following two subspaces:

$$k\text{-cycles} : Z_k(A_\bullet) := \ker(\partial_k) = \{a \in A_k : \partial_k a = 0\},$$

$$k\text{-boundaries} : B_k(A_\bullet) := \text{im}(\partial_k) = \{a \in A_k : a = \partial_{k+1} b\}.$$

Remark 2.10. By Proposition 2.9, the image of ∂_{n+1} is contained in the kernel of ∂_n , so we can take the quotient of $\ker(\partial_n)$ by the $\text{im}(\partial_{n+1})$.

Definition 2.11. For a chain complex A_\bullet , the n th *homology group* $H_n(A_\bullet)$ is defined as the quotient

$$H_n(A_\bullet) = \ker(\partial_n) / \text{im}(\partial_{n+1}).$$

Roughly, we can think of homology groups as the subspace of homotopy classes of cycles in $Z_k(A_\bullet)$ that are not the boundaries of elements of $B_{k+1}(A_\bullet)$. The quotient vector space $H_n(A_\bullet)$ is called the k -th homology of the chain complex A_\bullet .

Definition 2.12. The dimension of $H_n(A_\bullet)$ over \mathbb{K} is called the n -th Betti number of A_\bullet , denoted by $\beta_n(A_\bullet)$.

Example 2.13. Suppose X is the simplicial complex homeomorphic to the boundary of a triangle with vertices v_0, v_1 , and v_2 . Then $C_0(X, \mathbb{K}) = \mathbb{K}^3$, generated by the vertices $[v_1], [v_2], [v_3]$, and $C_1(X, \mathbb{K}) = \mathbb{K}^3$, generated by the edges $[v_0, v_1], [v_1, v_2], [v_0, v_2]$, and $C_n(X, \mathbb{K}) = 0$ for all $n > 1$. Then

$$H_0(X) = \frac{Z_0(X)}{B_0(K)} = \frac{\langle [v_0], [v_1], [v_2] \rangle}{\langle [v_0] - [v_1], [v_0] - [v_2] \rangle} = \mathbb{K},$$

$$H_1(X) = \frac{Z_1(X)}{B_1(K)} = \frac{\langle [v_0, v_1] - [v_0, v_2] + [v_1, v_2] \rangle}{\langle 0 \rangle} = \mathbb{K}.$$

All other homology groups are trivial because $C_n(X; \mathbb{K}) = 0$ for $n \geq 2$. The *geometric realization*² $|X|$ of this simplicial complex X is homotopic³ to S^1 , so this example computes the homology of S^1 .

Note that the dimension $H_0(S^1)$ corresponds to the number of path components of S^1 , and the dimension of $H_1(S^1)$ corresponds to the number of 1-dimensional cycles in S^1 . In general, if we have a n -dimensional sphere S^n , then $H_n(S^n) = \mathbb{K}$ and $H_0(S^n) = \mathbb{K}$ and all other homology groups are trivial.

In fact, the homology groups have a very nice interpretation: for all simplicial complexes X , $H_0(X)$ is the free \mathbb{K} -vector space on the set of 0 -dimensional cycles of X modulo boundaries. i.e., two vertices are equivalent if there exists a sequence of edges between them, which can be visualized as *path components* in $|X|$; $H_1(X)$ is the free \mathbb{K} -vector space on the set of 1 -dimensional cycles of X modulo boundaries, which can be visualized as *1-dimensional holes* in $|X|$; more generally, $H_k(X)$ is the free \mathbb{K} -vector space on the set of k -dimensional cycles of X modulo boundaries, and can be visualized as *k-dimensional holes* in $|X|$.

Fact 2.14. Let X and Y be topological spaces⁴, then a continuous map $f : X \rightarrow Y$ induces a \mathbb{K} -linear map $H(f) : H_*(X; \mathbb{K}) \rightarrow H_*(Y; \mathbb{K})$.

Remark 2.15. Let X and Y be topological spaces. If $f, g : X \rightarrow Y$ are homotopic, i.e., they can be continuously deformed into each other, then they induce the same \mathbb{K} -linear homomorphisms on the respective homology groups.

Remark 2.16. If two topological spaces X, Y are homotopy equivalent⁵, then X and Y have isomorphic homology groups.

2.4 Persistent homology

Computing the homology of finite simplicial complexes can be simplified to computing linear algebra. However, most empirical data sets do not exist in the form of simplicial complexes. Here we introduce the method in which we build simplicial complexes from directed networks.

²See Section 1.8 in [RB20]: vaguely, the geometric realization $|X|$ of a simplicial complex X is the operation that interprets each algebraic n -simplex as a standard topological n -simplex gluing together in a “nice” way.

³See Section 1.8 in [RB20]: vaguely, two topological spaces X and Y are homotopic if one can be continuously deformed into the other.

⁴In fact, we can define homology for “sufficiently nice” topological spaces, not just abstract simplicial complexes.

⁵Two topological spaces X and Y are homotopy equivalent if there exists a pair of continuous maps $f : X \rightarrow Y$ and $g : Y \rightarrow X$ such that $f \circ g$ is homotopic to id_Y and $g \circ f$ is homotopic to id_X .

Given a directed network G , one can induce a sequence of simplicial complexes $\mathcal{D}_{\delta,G}$ for a sequence of values of parameter $\delta \in \mathbb{R}_+$ such that $\mathcal{D}_{\delta=m,G} \subseteq \mathcal{D}_{\delta=n,G}$ for all $m \leq n$, where $\mathcal{D}_{\delta,G}$ is called the Dowker sink complex (Definition 2.17), introduced first in [CM18]:

Definition 2.17. Given a network $G = (X, w_X)$ and $\delta \in \mathbb{R}$, define $R_{\delta,G} \subseteq X \times X$ as:

$$R_{\delta,G} := \{(x, x') : w_X(x, x') \leq \delta\}.$$

Using $R_{\delta,G}$, we build a simplicial complex $\mathcal{D}_{\delta,G}$ called the *Dowker sink complex* as:

$$\mathcal{D}_{\delta,G} := \{\sigma \in \text{Pow}(X) : \text{there exists } p \in X \text{ s.t. } (x_i, p) \in R_{\delta,G} \text{ for each } x_i \in \sigma\}.$$

The node p is called the *sink* for the simplex σ , and one can check that $\mathcal{D}_{\delta,G} \subseteq \mathcal{D}_{\delta',G}$ for all $\delta \leq \delta'$.

Remark 2.18. Let $G = (X, w_X)$ be a dissimilarity network, then $R_{\delta=0,G}$ is precisely the set of $|X|$ vertices, i.e., 0-simplices.

We note that the Dowker sink complex and the *Vietoris-Rips complex* [Ott+17], one of the most prevalent tools used in TDA, are analogous. While the Vietoris-Rips complex induces a filtration of simplicial complexes from a point cloud data set, the Dowker sink complex induces a filtration of simplicial complexes from a directed network.

Definition 2.19. A *filtration of simplicial complexes* is a set of simplicial complexes X_1, X_2, \dots, X_n such that each simplicial complex is contained in its successive simplicial complex:

$$X = X_1 \subseteq X_2 \subseteq \dots \subseteq X_n.$$

X is called a *filtered simplicial complex*.

Example 2.20. Let $G = (X, w_X)$ be a directed network in a finite metric space, and let the simplicial complexes $\mathcal{D}_{\delta_1,G}, \mathcal{D}_{\delta_2,G}, \dots, \mathcal{D}_{\delta_n,G}$ be the Dowker sink complexes at $\delta_1 \leq \delta_2 \leq \dots \leq \delta_n$ respectively, then the sequence

$$\mathcal{D}_{\delta_1,G} \subseteq \mathcal{D}_{\delta_2,G} \subseteq \mathcal{D}_{\delta_3,G} \subseteq \dots \subseteq \mathcal{D}_{\delta_n,G} = \mathcal{D}_G$$

is a filtration of simplicial complexes, and \mathcal{D}_G is a filtered simplicial complex.

One can compute the homology groups of the simplicial complex $\mathcal{D}_{\delta,G}$ for a given value of δ . As the value of δ increases, one can record the birth (the value of δ at which a homological feature appears) and death (the value of δ at which a homological feature disappears) values for all homological features (path components, 1-cycles, etc.) that appear in \mathcal{D}_G .

Remark 2.21. The definitions of birth, death values, and a persistence barcode of a homological feature in a filtered simplicial complex are presented in Section 4 of [Ott+17]. We will omit these definitions since they are not necessary for the work presented in this paper.

Thus, given a filtered simplicial complex X , we can associate X a union of *persistence barcodes*, each representing the birth and death values of a homological feature of X . Persistence barcodes record changes in the n -dimensional homology group H_n for some $n \in \mathbb{Z}_{\geq 0}$ when one simplicial complex X_i is included in its successor X_{i+1} in the filtration. In particular, each persistence barcode $[b, d)$ represents the lifetime of a generator of a homology group H_n , where b is the birth value of the feature and d is the death value of the feature in the filtered simplicial complex X . The value $d - b$ is called the *persistence* of the feature.

Definition 2.22. A *persistence diagram* \mathbf{D} is a union of a finite set of points above the diagonal $D = \{(x, y) \in \mathbb{R}^2 | x = y\}$ in $\mathbb{R}_{\geq 0}^2$ and the entire diagonal D .

Remark 2.23. Each point $(x, y) \in \mathbf{D}$ represents the birth and death δ values of a persistent barcode $[b, d)$ with $x = b$ and $y = d$.

Remark 2.24. Roughly, a persistence diagram inherits the same data as the persistence barcodes. However, the persistence diagram \mathbf{D} contains the diagonal D , which will be useful for comparing the similarity of data sets. Specifically, a matching between two persistence diagrams map each point of a persistence diagram to another point of the other persistence diagram or to the diagonal, For more concrete details, see Section 5.1.

Generally, given any dissimilarity network $G(X, w_X)$, we denote $\text{Dgm}_n(G)$ as the n -dimensional persistence diagram for the filtered simplicial complex \mathcal{D}_G . Moreover, we denote

$$\mathbf{P}_n(G) := \{[b_1, d_1), [b_2, d_2), \dots, [b_k, d_k)\}$$

to be the set of n -dimensional persistence barcodes for the filtered simplicial complex \mathcal{D}_G .

3 Quasi-centrality measure

In this section, we present our definition of *quasi-centrality* for a directed dissimilarity network based on persistent homology and the Dowker sink complex, which is given in Definition 3.4.

Definition 3.1. Let $G = (X, w_X)$ be a (dissimilarity) network and let $x \in X$, define $f(G, x) = (X \setminus \{x\}, w_X)$ to be the sub-network induced by deleting x and all edges incident to x in G .

Remark 3.2. We note that for a trade network $G = (X, w_X)$, where $w_X(x_1, x_2)$ represents the trade volume between individual countries or industries $x_1, x_2 \in X$, the induced network $f(G, x)$ models an embargo trade barrier⁶ on x .

Definition 3.3. Let $G = (X, w_X)$ be a (dissimilarity) network and $\gamma(G) = (X, m_X)$ be defined as in Definition 2.4. Given, $x \in X$, define $\mu(x)$ to be the minimum distance between x and any other node $x' \in X$ with respect to the metric m_X .

Definition 3.4. Let $G = (X, w_X)$ be a (dissimilarity) network and let $x \in X$. We define the *quasi-centrality* $C(x)$ as follows:

$$C(x) = \sum_{b \in \mathbf{P}_0(f(\gamma(G), x))} l(b) - \sum_{b \in \mathbf{P}_0(\gamma(G))} l(b) + \mu(x)$$

where b represents a persistence interval in \mathbf{P}_0 and $l(b)$ is the persistence of the interval, i.e., the length of the interval, and $\gamma(G)$ is given in Definition 2.4 and $\mu(x)$ is given in Definition 3.3.

Remark 3.5. We first give an intuitive explanation of what $C(x)$ measures before proving that $C(x)$ is nonnegative for all $x \in X$. We note that $\gamma(G)$ not only can be treated as a topological space where $\mathcal{D}_{\gamma(G)}$ encompasses the data of its homological features, but it also can be treated as a metric space since it comprises the data of the pairwise distances between nodes in the network.

Roughly speaking, we think of the sum $\sum_{b \in \mathbf{P}_0(\gamma(G))} l(b)$ as a measure of how disconnected $\gamma(G)$ is, i.e., if \sum is large, then $\gamma(G)$ is considered to be highly disconnected and vice versa. In particular, since the dimension of the 0-th homology group of a simplicial complex X measures the number of path components (disconnected components) in $|X|$, we consider $\sum_{b \in \mathbf{P}_0(\gamma(G))} l(b)$ to be the “size” of the 0-th homology group of $\gamma(G)$, and $\gamma(G)$ is considered to be highly disconnected if the sum \sum is large and vice versa. Similarly, the sum $\sum_{b \in \mathbf{P}_0(f(\gamma(G), x))} l(b)$ as a measure of

⁶For our purposes, an embargo on x refers to the situation in which all other countries/nodes refuse to trade with x .

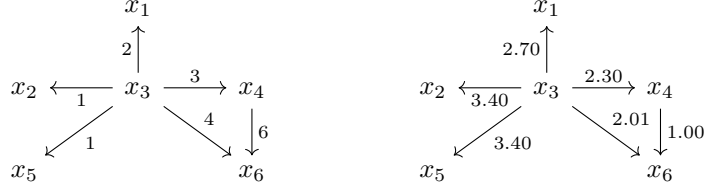


Figure 1: Left: $G = (X, w_X)$. Right: $\gamma(G) = (X, m_X)$.

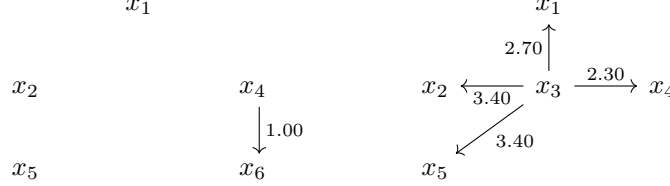


Figure 2: Left: $f(\gamma(G), x_3)$. Right: $f(\gamma(G), x_6)$.

how disconnected $f(\gamma(G), x)$ is. We think of the difference as measuring of how much node x contributes to the overall connectivity of the network.

Intuitively, a network should become ‘more disconnected’ after deleting a node, which is made precise by the following theorem.

Theorem 3.6. *For a (dissimilarity) network $G = (X, w_X)$, $C(x)$ is nonnegative for all $x \in X$.*

Proof. Recall that each $[0, d_j] \in \mathbf{P}_0(\gamma(G))$ represents the lifetime of a path component in $\gamma(G)$, hence by definition one connected component α joins another connected component β at $\delta = d_j$. Therefore, at least two 0-simplices $[x_1] \in \alpha$ and $[x_2] \in \beta$ belong to a 1-simplex $\pm[x_1 : x_2]$ at $\delta = d_j$. By definition of the Dowker sink complex, there exists $p \in X$ such that $m(x_1, p) \leq d_j$ and $m(x_2, p) \leq d_j$.

If $x \in \{x_1, x_2, p\}$, then the 0-th persistence barcode in $\mathbf{P}_0(f(\gamma(G), x))$ representing the lifetime of path component α or β will have a death resolution δ greater than or equal to d_j . If $x \notin \{x_1, x_2, p\}$, then the 0-th persistence barcode in $\mathbf{P}_0(f(\gamma(G), x))$ representing the lifetime of path component α or β is still equal to d_j .

Hence all persistence barcodes representing the lifetime of connected components have death values in $\mathbf{P}_0(\gamma(G))$ greater than or equal to the death values of corresponding components in $\mathbf{P}_0(f(\gamma(G), x))$. Therefore

$$\sum_{b \in \mathbf{P}_0(f(\gamma(G), x))} l(b) + \mu(x) \geq \sum_{b \in \mathbf{P}_0(\gamma(G))} l(b),$$

and our proof is complete. \square

Remark 3.7. As the value of δ increases, $C(x)$ measures how often node x bridges between two disconnected components. We think of the node x as serving as a δ -sink not only for a pair of nodes in a neighborhood of x , but rather the pair of connected components the nodes belong to. In this way, the quasi-centrality $C(x)$ takes into account not only the local neighborhood of node x , but also propagating effects in said neighborhood.

Example 3.8. In this example, we compute the quasi-centrality measure on nodes in an example directed weighted network G as shown in Figure 1 (left), and compare the result with existing centrality measures.

The illustrations of G and $\gamma(G)$ are shown in Figure 1 (left and right respectively). In the illustration of $\gamma(G)$, we leave out the edge e_{ij} when $w(x_i, x_j) = 0$. However, in our calculation of $C(x)$, we let the $m(x_i, x_j)$ to be a very big number when $w(x_i, x_j) = 0$ so that nodes x_i, x_j in m_X are very far apart that it is almost indistinguishable when no relation exists between the two nodes. In this example, we let $m(x_i, x_j) = 24.026$ to make computations more convenient.

Figure 2 illustrates the networks for $f(\gamma(G), x_3)$ and $f(\gamma(G), x_6)$, and Figure 3 illustrates the Dowker persistence barcodes for $f(\gamma(G), x_3)$, $f(\gamma(G), x_6)$, and $\gamma(G)$.

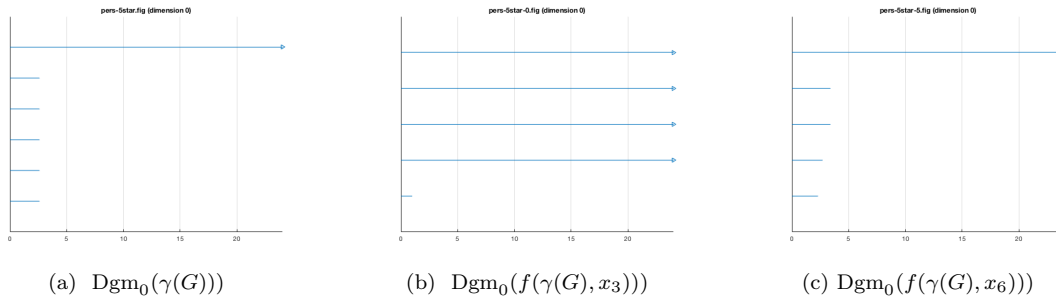


Figure 3

We illustrate the computation of $C(x_3)$ as an example. Since $\mu(x_3) = 2.01$,

$$\begin{aligned}
 C(x_3) &= \sum_{j=1}^{n-1} d'_j - \sum_{j=1}^n d_j + 2.01 \\
 &= (1 + 24.026 \times 4) - (1 + 2.01 + 2.70 + 3.40 + 3.40 + 24.026) + 2.01 \\
 &= 65.978.
 \end{aligned}$$

Similarly, we have $C(x_1) = C(x_2) = C(x_4) = C(x_5) = 0$ and $C(x_6) = 0.29$. This shows that in G , node x_3 has the most propagating influence in the entire network, followed by node x_6 .

Figure 4 illustrates the values obtained by other existing centrality measures. We have included two measures for the Pagerank centrality: Pagerank 1 outputs the centrality values for nodes in the original network G and Pagerank 2 outputs the centrality measures for nodes in the network obtained by switching the direction of edges from G .

We included Pagerank 2 for the following reason: since all existing centrality measures are based on the term $\sum A_{ij}x_j$, i.e., the centrality value x_i for node i only sums up the contribution values x_j of nodes j that points toward node i , by computing Pagerank 2 and comparing to the

Node	Quasi-centrality	Katz	Pagerank 1	Pagerank 2	HITS-hubs	HITS-authorities
x_1	0	0.38	0.15	0.50	0.00	0.11
x_2	0	0.38	0.13	0.09	0.00	0.06
x_3	65.978	0.23	0.12	0.09	0.47	0.00
x_4	0	0.38	0.16	0.09	0.53	0.17
x_5	0	0.38	0.13	0.14	0.00	0.06
x_6	0.29	0.61	0.31	0.09	0.00	0.60

Figure 4: Centrality values for G obtained from the quasi-centrality, Katz centrality, HITS-hubs and authorities centralities, and Pagerank centrality.

results obtained by the quasi-centrality, we essentially eliminates the possibility that the differences in the centrality values is solely due to the fact of the direction of the edges.

We note that x_3 has low centrality values for all existing centrality measures, including both Pagerank centrality measures. On the contrary, we note that x_3 has the high quasi-centrality value, which is what we would expect since node x_3 holds the center position in the directed dissimilarity network G . Therefore in this case, the quasi-centrality better measures the propagating influences of nodes in this example.

We expect the reason for all existing centrality measures having resulted in a low centrality value for x_3 for two reasons: firstly, all existing centrality measures depend on the adjacency matrix A_{ij} and contain the term $\sum A_{ij}x_j$, which sums up the contribution values of nodes j that links with node i . However, it is possible that the nodes j are not central in the network and hence does not contribute to the centrality value of node i , which can be shown in this example. However, it is still possible that node i holds the center position in the network even when its neighbors are not important in the network (such as in this example), in which all existing centrality measure fail to accurately measure its importance. Secondly, the adjacency matrix term $\sum A_{ij}x_j$ only sums up the contribution values of nodes j that has an ingoing edge to node i while ignoring the nodes that have an outgoing edge from node i , meaning that the centrality values are only taking into account of propagating influences on one side while ignoring the other.

Generally, when measuring the propagating influence of a node, both incoming edges and outgoing edges should positively impact the centrality value. For example, when measuring the influence of perturbations originated from a country or industry in the empirical trade production network, if a particular country or industry is both an important exporter and importer, it will have a greater influence on the entire economy compared to a country or industry that is only an important importer. The quasi-centrality in this case would give a more thorough measure of a node’s propagating influence as it takes into account of the nodes’ impact in both directions, whereas other centrality measures might only focus on ingoing edges or outgoing edges.

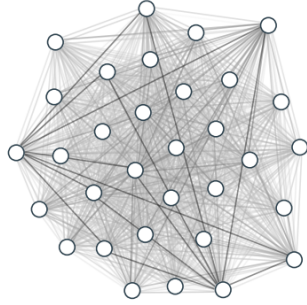
4 Quasi-centrality applications

In this section, we apply the quasi-centrality measure to an empirical network describing trade among countries and industries—the Asia machinery production network in 2007 and 2011. We show that the quasi-centrality measure gives a more refined evaluation of the propagating influences of nodes in the trade network than existing centrality measures.

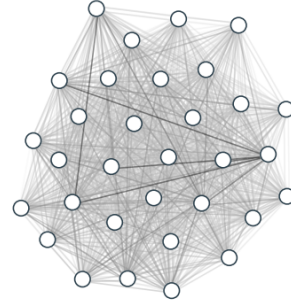
The trade network exhibits many properties common to complex networks, including high clustering coefficient and degree-degree correlation between different vertices [SBn03]. Moreover, research suggests that random microeconomic shocks have the ability to propagate through a network with high inter-connectivity and cause “cascade effects” in the entire network and economy [Eco]—a prime example is the Asiatic crisis. Therefore, it is important to take into account of the propagating effects of the trade network when evaluating node centrality. We later show that the quasi-centrality accurately predicts the influences of nodes in the trade network as it takes into account of such propagating effects.

We examine the machinery production network in Asia during 2007 and 2011 (network illustrations are shown in Figure 5), before and after the 2008-2009 financial crisis. Our network consists of both different intermediate products and different countries as nodes, as either counterpart plays a crucial role in supply and production chains⁷, which are central for our understanding of the entire network.

⁷A supply/production chain is a system of intermediate transactions, suppliers, distributors, consisting of different countries and intermediate products.



(a) 2007 Asia machinery production network



(b) 2011 Asia machinery production network

Figure 5

Our data consists of 8×4 pairs of countries and industries, obtained from the OECD Inter-Country Input-Output (ICIO) Tables [OECD21], which provides transport values between country-industry pairs. The eight countries and four industries are labeled in the following table:

Country Label	Country Name
KOR	North Korea
JPN	Japan
MYS	Malaysia
SGP	Singapore
TWN	Chinese Taipei
CN	China
VNM	Vietnam
THA	Thailand

Industry Label	Industry Name
C29	Machinery and equipment
C30	Computer, electronic and optical equipment
C31	Electrical machinery and apparatus
C34	Motor vehicles, trailers and semi-trailers

For brevity, we abbreviate the 32 nodes by [label of industry]-[label of the country]. For example, the machinery and equipment industry in Korea is referred to as KOR-C29.

Remark 4.1. In our data obtained from the OECD Inter-Country Input-Output (ICIO) Tables, imports are valued at basic prices of the country of origin, i.e., the domestic and international distribution included in goods imports in c.i.f. purchasers’ prices are re-allocated to trade, transport and insurance sectors of foreign and domestic industries.

Remark 4.2. We exclude taxes paid and subsidies received in foreign countries in our network.

Remark 4.3. For the rest of Section 4, we will use “node” and “country-industry pair” interchangeably.

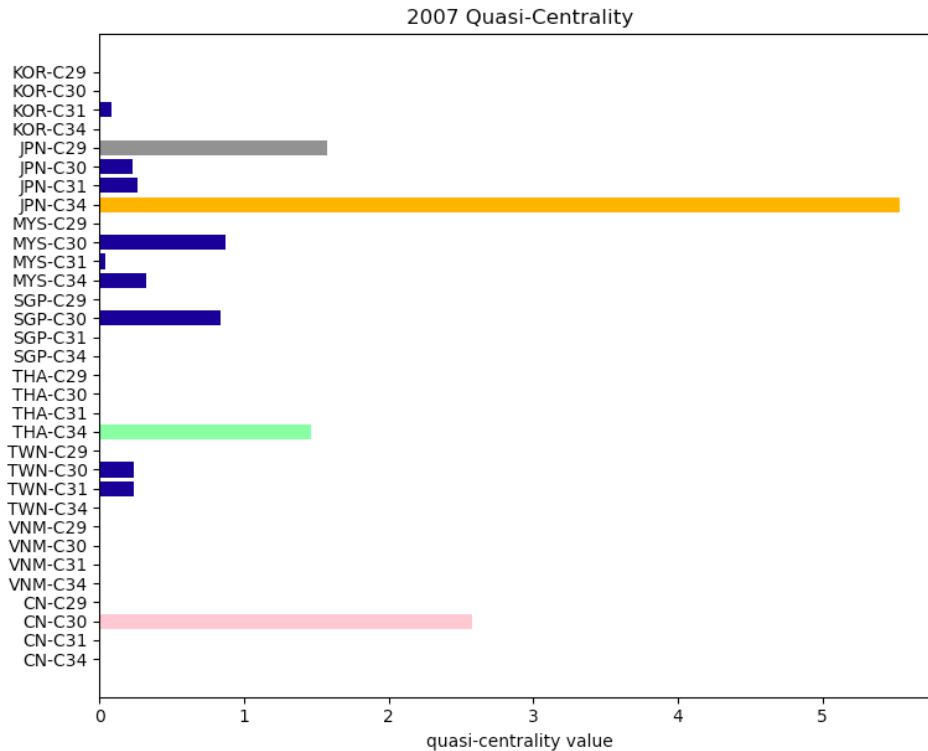


Figure 6: Quasi-centralities for the 2007 Asia machinery production network.

4.1 2007 Asia machinery production network

We first examine nodes' propagating influences in the 2007 Asia machinery production network. We compute the quasi-centralities (Figure 6) for nodes in the network and compare with existing centrality measures (Figure 7) and later demonstrate that quasi-centrality successfully captures propagating effects and more accurately evaluates nodes' propagating influences in the network.

According to Figure 6, JPN-C34 (marked in orange) has the highest quasi-centrality value, implying that it has the highest propagating influence over the network, followed by CN-C30 (marked in pink), JPN-C29 (marked in grey), THA-C34 (marked in green), etc. These values reflect that China and Japan were the two largest contributors to the machinery production industry in Asia in 2007 [OK16].

We note that THA-C34 also has a relatively high quasi-centrality value. Although Thailand only accounted for 1.12% of global exports of vehicle parts in 2007, while Japan, Korea, and China accounted for 10.2 %, 4.05 %, and 3.98 % ([OEC21]) respectively, Thailand holds an important network position locally in Asia due to its high inter-correlation with China, the leading player in the network. According to [OEC15], between 2000 to 2010, Thailand's exports in motor vehicles parts, electronics and electrical appliances increased by > 40 percent, and growth in trade with China (both imports and exports) also grew substantially. By 2010, China became Thailand's lead export destination.

Since quasi-centrality is designed to capture how much a node contributes to the overall connectivity of the network, meaning that higher intensity of links between nodes will lead to a higher connectivity, we may also conclude the following: if a node a is highly influential in the network, and if another node b exhibits strong linkage with node a , then node b also plays a relatively im-

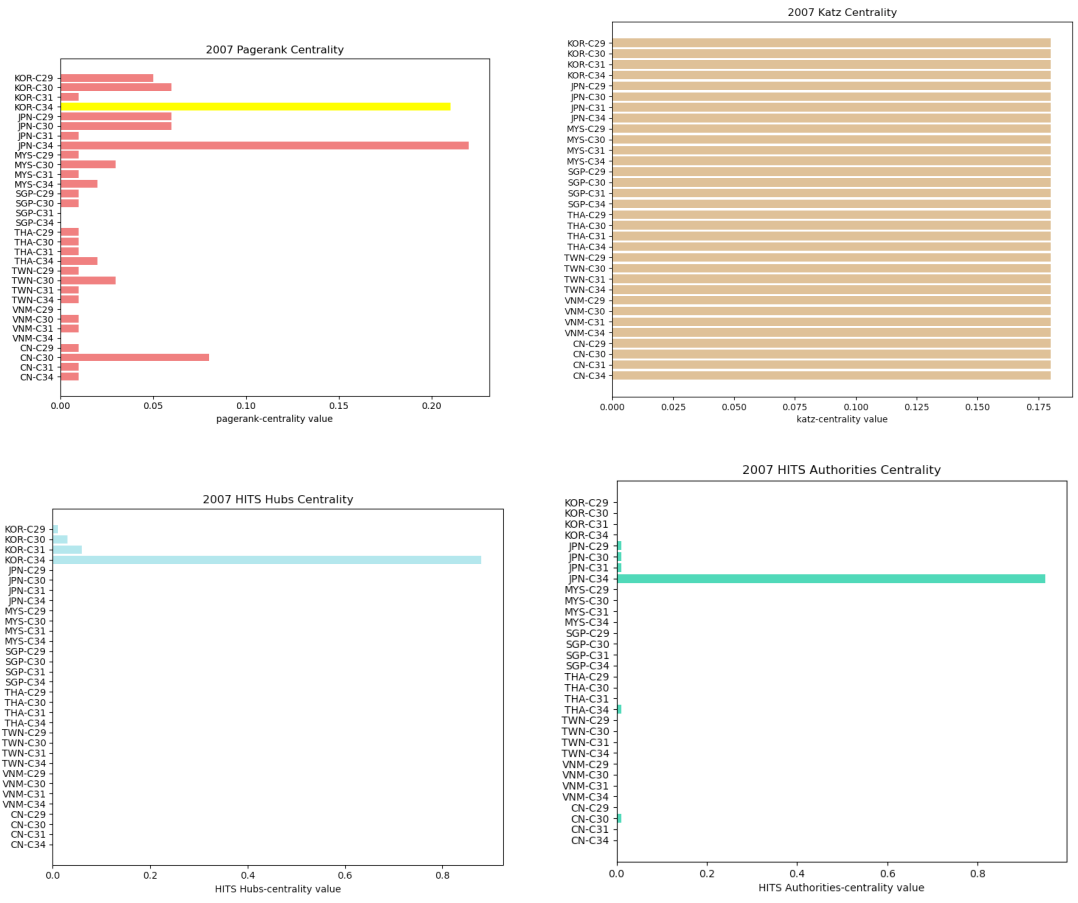


Figure 7: Existing centralities (Katz, Pagerank, HITS-hubs, HITS-authority) for the 2007 Asia machinery production network.

portant role in the network because perturbations originated from b will have a high influence on a , and will then be highly influential to the entire network. Since Thailand is highly interconnected with China, one of the main players in the network, we may infer that perturbations originating from Thailand have a high potential to propagate over the entire network.

Comparison with existing centrality measures. We observe that the relative importance of THA-C34 according to the quasi-centrality is not reflected in other existing centrality measures. We hypothesize that this is due the fact that the existing centrality measures are dependent on the adjacency matrix A_{ij} , implying that the centrality value for THA-C34 is based on its local neighborhood, while the quasi-centrality evaluates the THA-C34’s contribution to the connectivity of the entire network, emphasizing its interconnectedness with CN-C30.

According to Figure 7, the Katz centrality returns similar centrality values for all nodes in the network, failing to capture subtleties in the influences of nodes. Since the trade network exhibits high clustering properties, i.e., every country-industry pair is linked with almost every other country-industry pair, the Katz centrality is unable to capture the relative differences in the influence of nodes. We may conclude that this is due to the fact that Katz centrality ignores the importance of the intensity of interaction between edge links, i.e. it does not take edge weights into account.

Moreover, both HITS-hubs and HITS-authorities centralities are not refined since their output values fail to evaluate the relative scale of the the influences of nodes in the network: they are only able detect a single node that is highly influential, while ignoring the influences of other nodes. This is not helpful for detecting potential sources of shocks in supply chains because we are only able to detect one single chain instead of many chains.

Of the existing centrality measures, the Pagerank centrality gives similar relative information of the influence of nodes compared to the gradation given by the quasi-centrality. However, we note that the Pagerank centrality outputs a high centrality value for KOR-C34 (marked in yellow), while the quasi-centrality measure for KOR-C34 is zero. According to [OK16], even though Korea has high import value for machinery final products, its export value is relatively low—in particular, KOR-C34 has a many more incoming edges than outgoing edges. Since the Pagerank centrality x_i for a node i is $\alpha \sum_j A_{ij}x_j + \beta$, which only sums up the contribution value of nodes j that has an edge that points toward node i , KOR-C34 has a high Pagerank centrality value as it has a large number of incoming edges, i.e., imports.

However, a country-industry pair’s propagating influence in the entire trade network should depend on both its export and import values, as a country-industry pair with both high export and import should have a larger propagating influence than a country-industry pair with only high import values. Thus in this case, quasi-centrality offers a better assessment of a node’s influence than Pagerank does. While the Pagerank centrality only remembers one direction at a time, meaning that it captures some propagating influence but not all, the quasi-centrality captures both directions and gives a more comprehensive evaluation of the propagating influence of a node.

4.2 2011 Asia machinery production network

Now we proceed to analyze the propagating influences of nodes in the 2011 Asia machinery production network, following the 2007-2008 financial crises. The objective is to identify changes in power dynamics of nodes in the network compared to 2007.

According to Figure 8, CN-C30 (marked in pink) has the highest quasi-centrality value, followed by JPN-C34 (marked in orange) and SGP-C30 (marked in red). It is worth noting the change in the quasi-centrality values for JPN-C34, CN-C30, and SGP-C30 from 2007 to 2011. While the quasi-centrality value for CN-C30 and SGP-C30 have significantly increased, the value for JPN-C34 has significantly decreased.

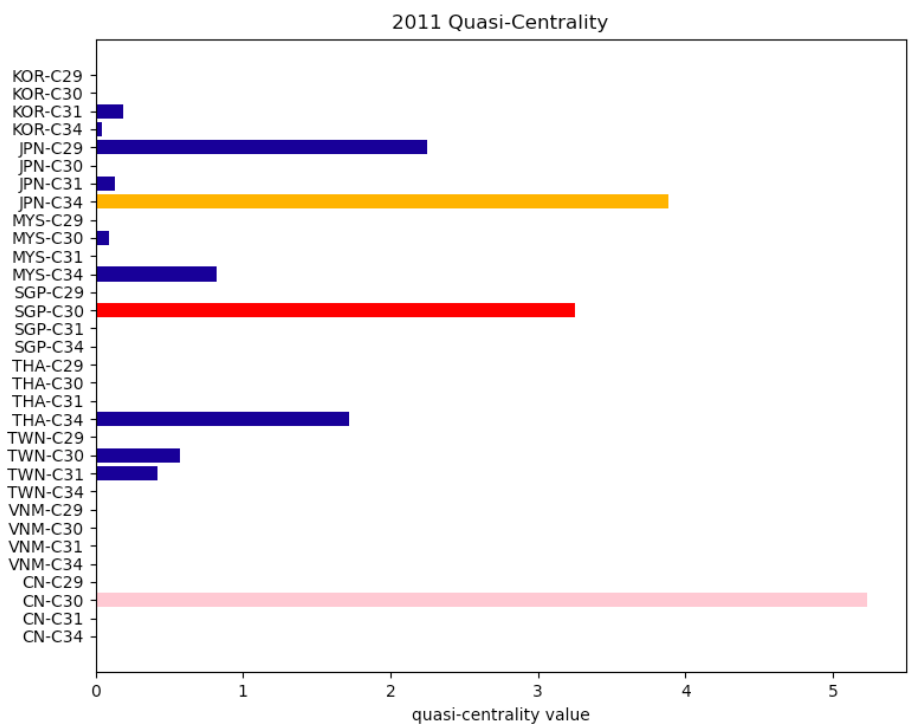


Figure 8: Quasi-centralities for the 2011 Asia machinery production network.

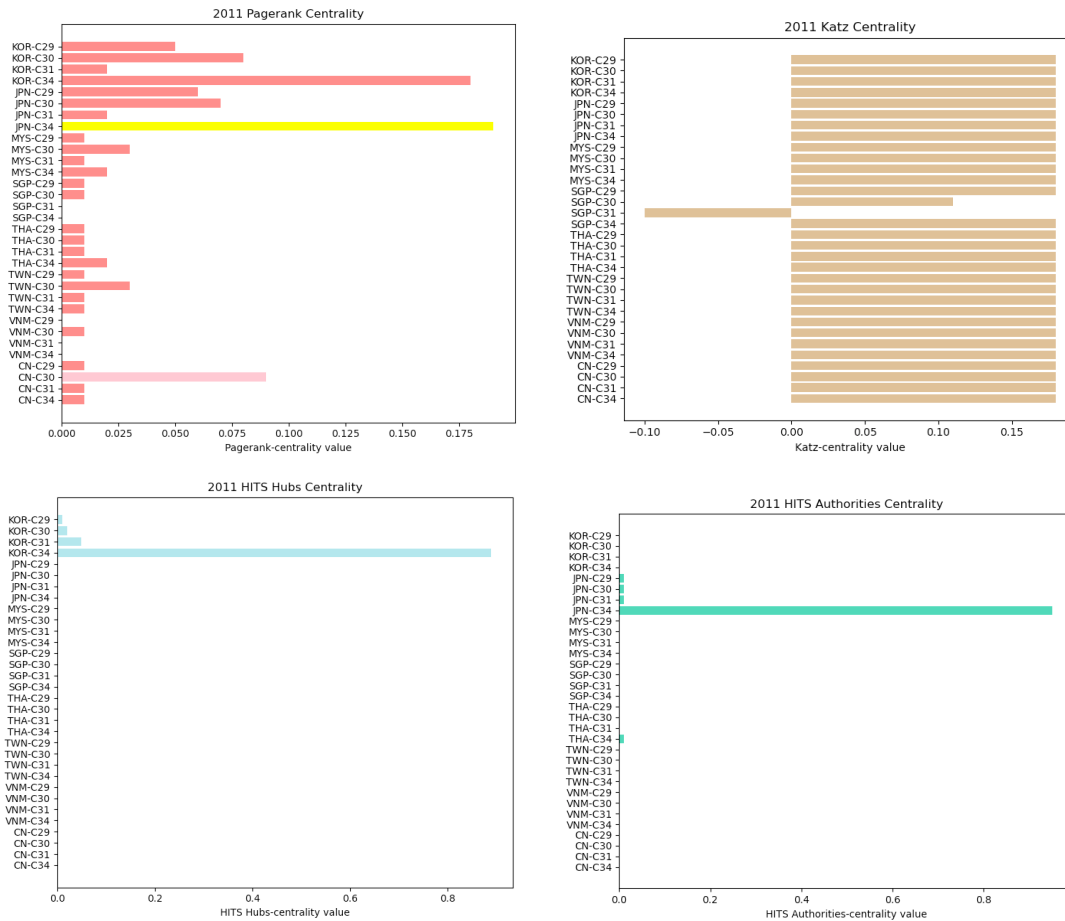


Figure 9: Existing centralities (Katz, Pagerank, HITS-hubs, HITS-authority) for the 2011 Asia machinery production network.

According to [OK16], China became a dominant player in the global machinery production network in terms of *both* export value and the diversity of industry-destination pairs following the 2008-2009 financial crisis. Since the quasi-centrality measures how a node contributes to the overall connectivity of the network, accounting for *all* factors of import, export values, and industry-destination pairs, China’s growth of impact in the network is substantially illustrated by its increase in the quasi-centrality value.

According to [OK16], Japan’s machinery production growth, on the other hand, stagnated because of two main reasons after the financial crisis of 2008-2009. First, transportation links between Japan and multiple countries weakened, specifically with Korea. Second, there was a noticeable decrease in the number of product–destination pairs in Japan’s exports of machinery. Both of these factors have led to Japan’s stagnation in economic growth, and the stagnation is distinctively illustrated by the change of Japan’s quasi-centrality value relative to others as computed in Figure 8.

In addition to the significant power dynamic changes of Japan and China in the network, we also note that the quasi-centrality value for SGP-C30 has significantly increased compared to 2007. According to [Hoo22], the Singapore electrical machinery and apparatus industry became more *central* in numerous supply chains (see footnote 7) after the financial crisis, thereby playing an important role as an intermediary between an abundance of production, supply, distribution, and post-sales activities of goods and services.

Comparison with existing centrality measures. We note that China’s dominance in the network and Japan’s stagnation of economic growth are not reflected in the existing centrality measures. Figure 9 shows that the HITS-hubs and authorities centrality measures are not refined in that they continue to only detect the node with the highest influence in the network (KOR-C34 and JPN-C34 respectively), and that the Katz centrality continues to output the same centrality measure for almost all the nodes.

Moreover, the Pagerank centrality continues to measure JPN-C34 (marked in yellow) as being the most influential node in the network. We hypothesize that this is due to the fact that since the Pagerank centrality only sums up the contribution values of nodes having an incoming to JPN-C34, the decrease in product–destination pairs for JPN-C34 of *exports* would not affect its Pagerank centrality. Furthermore, CN-C30 (marked in pink) also has a relatively low Pagerank centrality value even though it is the dominant player in the network in 2011. We hypothesize that this is again due to the fact that the Pagerank centrality only sums up the contribution values of nodes having an incoming edge to CN-C30, but CN-C30 exports far more than imports. Specifically, China’s raw export value of machines globally was \$894 billion dollars in 2011, while its raw import value of machines globally was only \$328 billion dollars.

Moreover, the increase of Singapore electrical machinery industry’s influence in the network is substantially reflected by its quasi-centrality value, while not reflected in the existing centrality measures, again showing how the quasi-centrality is effective in measuring a node’s propagating influence in the network.

4.3 Summary

We computed the quasi-centralities of nodes in the 2007 and 2011 Asia machinery production network and compared the result with the existing centrality measures. Based on empirical trade statistics, we showed that quasi-centrality is optimal for detecting propagating influences of a node in the network as it takes into account both the import and export values of a country and industry pair, as well as the node’s number of industry-destination pairs and its role in intermediary transactions.

Specifically, we showed that empirical facts support the information obtained by computing

the quasi-centrality measure: a set of countries and industries such that perturbations originating from this set of nodes can propagate significantly through the network.

Potential applications Trade is often adversely affected by perturbations in a single country, including natural disasters such as a hurricane, earthquake or flood, and political turmoil or an armed conflict [KPD17]. Understanding which collections of nodes in the network might potentially cause significant propagating impacts can aid in identifying bottlenecks, thereby making the trade network more resilient. We propose that using the quasi-centrality measure can ultimately serve to protect world economies.

5 Clustering nodes by topological influences

In section 5.1, we introduce a method (Definition 5.6) that combines hierarchical clustering and persistent homology to construct a hierarchy of nodes based on their impact in the overall topology of a directed weighted network. In sections 5.2 and 5.3, we apply our method to the 2007 and 2011 Asia machinery production network.

5.1 Node hierarchy using TDA

Before introducing our method in Definition 5.6, we recall how hierarchical clustering works.

Given a set of objects and a specified weight function [Kul02] that computes the similarity between two objects in the set, hierarchical clustering is a method for detecting community structures within this set of objects by arranging the objects into a hierarchy of groups according to the weight function. For a given value of parameter t , hierarchical clustering groups objects that are “at most t far apart from each other” according to the weight function.

Definition 5.1. Given a finite set X , a *partition* P of X is a collection $P = \{B_1, \dots, B_k\}$ of subsets of X such that:

1. For all $i \neq j$, $B_i \cap B_j = \emptyset$, and
2. The union $\cup_{i=1}^k B_i = X$.

Each B_i is referred to as a *block* of P . We denote the set of all partitions of X by $\text{Part}(X)$.

Because the partition depends on t , we would like a family of t -partitions, hence a hierarchical dendrogram is defined to be a function that outputs a partition given an input value of t :

Definition 5.2. Let X be a finite set. A *dendrogram over X* is a weight function $D_X : \mathbb{R}_+ \rightarrow \text{Part}(X)$ such that:

1. For $t' \geq t$, every block of $D_X(t)$ is contained in a block of $D_X(t')$.
2. There exists $t_F \in \mathbb{R}_+$ such that for all $t \geq t_F$, $D_X(t) = \{X\}$.
3. The weight function $D_X(0)$ consists of singletons $\{x\}$ for all $x \in X$.
4. For all $t \in \mathbb{R}_+$, there exists $\epsilon > 0$ such that $D_X(t) = D_X(t')$ for all $t' \in [t, t + \epsilon]$.

As the value of parameter t increases, the partition $D_X(t)$ gets “coarser”, i.e., objects that are within t similar to each other belong to the same block in $D_X(t)$.

A dendrogram can be thought of as a (nested) family of partitions of X into *similarity classes*, with the variable t specifying the desired fineness and granularity of the partition $D_X(t)$. Elements that belong to the same block in $D_X(t)$ for small values of $t \in \mathbb{R}_+$ are more similar according to the weight function D_X , and vice versa.

Remark 5.3. We will not be going into detail of the computer algorithm that we use to construct our hierarchical dendrogram since the exact computer algorithm will not be necessary for explaining

our work. We will use *single-linkage clustering*⁸ as our clustering algorithm, which is also used in Definition 5.6.

Now we explain how we can extract a hierarchical dendrogram from a dissimilarity directed network $G = (X, w_X)$ based on nodes' impact in the topology of the network (Definition 5.6). We first introduce the set of objects included in our hierarchy:

Definition 5.4. Given a dissimilarity network $G = (X, w_X)$, let $\gamma(G)$ be as defined in Definition 2.4 and assume that $|X| = n$, the number of nodes in G . For each node $x \in X$, let $f(\gamma(G), x)$ be as defined in Definition 3.1. Define the set of objects denoted by S_G in the hierarchical dendrogram associated to G as:

$$S_G := \{\text{Dgm}(f(\gamma(G), x)) \mid x \in \gamma(G)\} \cup \{\text{Dgm}(\gamma(G))\}.$$

We are interested in using dendrograms to express the hierarchy of nodes in a given directed network based on nodes' topological influences in the network. The main idea is that the bottleneck distance between the Dowker persistence diagram associated to G and the diagram associated to the induced network $G_x = (X \setminus \{x\}, w_X)$ is small if x does not play an important role in the topology of G , and vice versa. We first introduce the set of objects S_G in the hierarchical dendrogram associated to G :

Recall that given a directed weighted network, the Dowker persistence diagram summarizes the homological features in the Dowker complex associated to the network. We may then conclude that given two directed weighted networks, if the two Dowker persistence diagrams associated to the networks are similar, then the networks have similar topology.

We measure the similarity between two persistence diagrams using the bottleneck distance (Definition 5.5), which describes the cost of the optimal *matching*⁹ between points of the two diagrams [Aga20], introduced below:

Definition 5.5. Given two persistence diagrams D_1 and D_2 , the *bottleneck distance* of D_1 and D_2 denoted by $d_{B\infty}(D_1, D_2)$ is defined as

$$d_{B\infty}(D_1, D_2) = \inf_{\eta: D_1 \rightarrow D_2 \cup \Delta} \sup_{x \in D_1} \|x - \eta(x)\|_\infty$$

where η ranges over all embeddings $\eta: D_1 \hookrightarrow D_2 \cup \Delta$, where Δ denotes the diagonal and $\|(x, y)\|_\infty$ is the usual L_∞ norm.

If the bottleneck distance between two Dowker persistence diagrams is close to zero, i.e., if $d_{B\infty}(\text{Dgm}_1, \text{Dgm}_2) \approx 0$, then we may conclude that two corresponding directed weighted networks have similar topological structure, and vice versa.

Using S_G , we may obtain the hierarchical dendrogram associated to the dissimilarity directed network G as follows:

Definition 5.6. Given a dissimilarity network $G = (X, w_X)$ with the bottleneck distance, the *hierarchical dendrogram* \mathcal{H}_G associated to G is the function

$$\begin{aligned} \mathcal{H}_G : \mathbb{R}_+ &\rightarrow \text{Part}(S_G) \\ t &\mapsto \text{partition of } S_G \text{ (} S_G, \text{ bottleneck distance)}. \end{aligned}$$

And we implement the single-linkage clustering (see footnote 8) algorithm to extract the hierarchical dendrogram from the function \mathcal{H}_G .

For a given object $x \in X$ and parameter value $t = t'$, we will denote the block containing x in $\mathcal{H}_G(t')$ by $\mathcal{B}_{\mathcal{H}_G}(x, t')$.

⁸You can read more concrete details about the algorithm in Section II of [VSB19].

⁹A matching η between persistence diagrams D_1 and D_2 pairs each point in D_1 with a point in D_2 or a point on the diagonal line, and pairs each point in D_2 with a point in D_1 or a point on the diagonal.

5.2 Hierarchical clustering of the 2007 Asia machinery production network

In this section, we examine the 2007 Asia machinery production network by analyzing the hierarchical dendrogram associated to the network obtained by the method we introduced in the previous section (Definition 5.6).

Figure 10 shows the hierarchical dendrogram associated to the 2007 Asia machinery production network, where the y -axis shows the 33 *nodes*¹⁰ in the dendrogram, corresponding to the elements in S_G , where each of $\text{Dgm}(f(\gamma(G), x))$ is labeled by the node x deleted, and $\text{Dgm}(\gamma(G))$ is labeled by STANDARD, and the values along the x -axis corresponds to the values of the parameter t in the weight function \mathcal{H}_G . The vertical line at $t = t'$ corresponds to the blocks of the partition $\mathcal{H}_G(t')$.

We consider the smallest value of t for which $\text{Dgm}(f(\gamma(G), x))$ belongs to the cluster containing STANDARD to be its topological impact on the network. In other words, we expect nodes in the dendrogram belonging to the block containing the STANDARD node at smaller values of t in the partition $\mathcal{H}_G(t)$ to have a smaller impact in the topology of the network, i.e., if given two nodes a, b and $t < t'$, and that $a \in \mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, t)$ and $b \in \mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, t')$ then b has a greater topological influence in the network topology.

According to Figure 10, we see that $\text{THA-C30}, \text{VNM-C34} \in \mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, t = 0)$, implying that the country-industry pairs THA-C30 and VNM-C34 are insignificant in the topological structure of the network. We consider these nodes as “peripheral,” i.e. deleting them doesn’t really change the network topology.

On the contrary, when $t < 0.48$, $\text{JPN-C34} \notin \mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, t)$ while $\mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, 0.42) = \{S_G \setminus \{\text{JPN-C34}\}\}$, thus we conclude that deleting JPN-C34 has the most significant impact in the topological structure of the network. Indeed, we see that the 1-dimensional barcodes of $\text{Dgm}(f(\gamma(G), \text{JPN-C34}))$ (Figure 11(b)) differs significantly from the 1-dimensional barcodes of $\text{Dgm}(\gamma(G))$ (Figure 11(a)): while all 1-dimensional barcodes in $\text{Dgm}(\gamma(G))$ have death parameters $\delta \leq 4.38$, there exists a 1-dimensional barcode in $\text{Dgm}(f(\gamma(G), \text{JPN-C34}))$ with death parameter $\delta = 5.64$, indicating that JPN-C34 has a significant impact in the one-dimensional homological features of $\gamma(G)$, reflecting its importance in the network topology.

We note that from the result illustrated in Section 4.1, JPN-C34 also has the highest quasi-centrality value in the 2007 Asia machinery production network. Since the quasi-centrality measures the propagating influence of JPN-C34 and the network hierarchy reflects JPN-C34’s impact in the overall topological structure of the network, we conclude that perturbations originating from JPN-C34 not only effectively propagate throughout the entire network, but also changes the network topology significantly.

Similarly, when $t < 0.33$, $\text{CN-C31} \notin \mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, t)$, implying that CN-C31 also has a significant impact in the network topology. Indeed, the 1-dimensional barcodes of $\text{Dgm}(f(\gamma(G), \text{CN-C31}))$ (Figure 11(d)) also differ significantly from the 1-dimensional barcodes of $\text{Dgm}(\gamma(G))$ (Figure 11(a)). Hence similar to JPN-C34, we conclude that perturbations originating from CN-C31 not only effectively propagate throughout the entire network, but also changes the network topology significantly.

On the contrary, although VNM-C31 has zero quasi-centrality value from the result in Section 4.1, when $t < 0.43$, $\text{VNM-C31} \notin \mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, t)$, implying that VNM-C31 has a significant impact in the network topology. Indeed, the 1-dimensional barcodes of $\text{Dgm}(f(\gamma(G), \text{VNM-C31}))$ (Figure 11(c)) exhibit considerable differences compared to the one-dimensional barcodes of $\text{Dgm}(\gamma(G))$ (Figure 11(a)). However, the 0-dimensional persistence diagrams do not reflect significant differences. Since the quasi-centrality is measured based on how much VNM-C31 contributes to the overall connectivity of the network and is measured in terms of connected components (0-dimensional

¹⁰For brevity, we will refer to the objects in the hierarchy to be the nodes in the dendrogram.

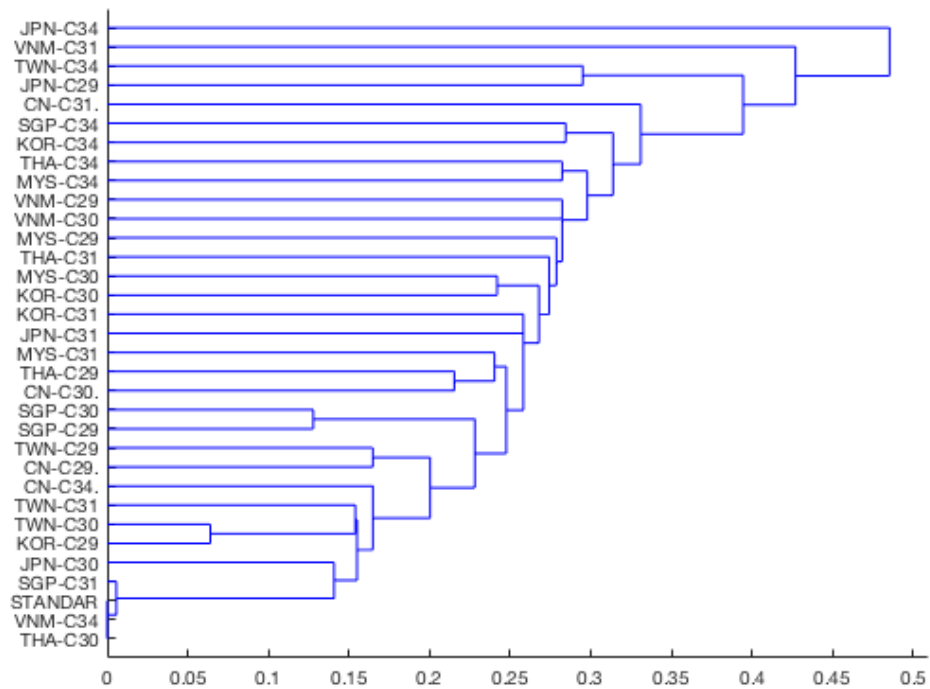


Figure 10: Hierarchical dendrogram associated to the 2007 Asia machinery production network.

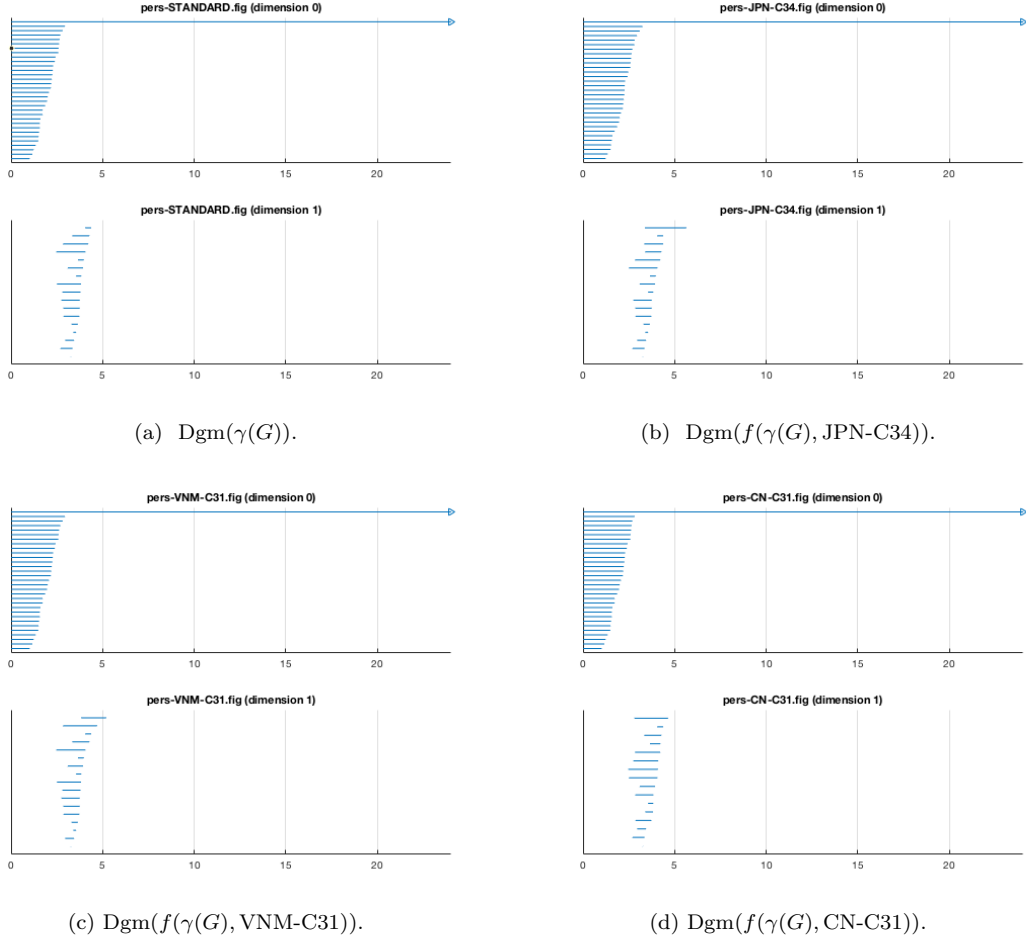


Figure 11: Dowker persistence diagrams for sub-networks obtained from the 2007 Asia machinery production network.

homological features), we conclude that although perturbations originating from VNM-C31 cannot propagate effectively throughout the entire network, it can potentially cause changes in the network topology.

In fact, the hierarchical dendrogram produced using the method in Definition 5.6 not only illustrates the hierarchy of the nodes based on their topological impact in the network, but also allows us to identify which collections of nodes have *similar* impacts in the network. In particular, the nodes that first belong to $\mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, t)$ at approximately the same value of t are considered to have similar topological influence in the network.

For example, it is clear from Figure 10 that the nodes THA-C31, MYS-C29, VNM-C30, VNM-C29 all first join $\mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, t)$ at *close* values of t ($0.26 < t < 0.28$), implying that this set of nodes have similar topological influence in the network. Indeed, a close resemblance between the Dowker persistence diagrams corresponding to this set of nodes can be observed in Figure 12.

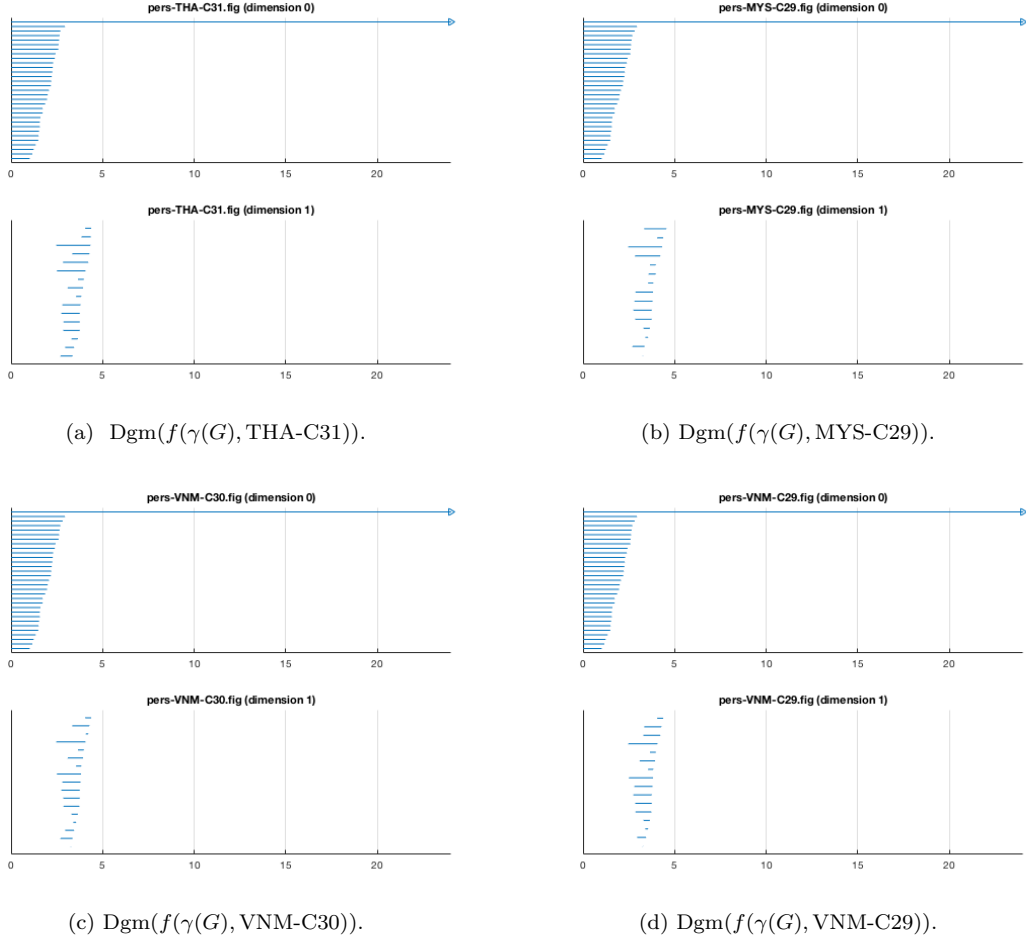


Figure 12: Dowker persistence diagrams for sub-networks obtained from the 2007 Asia machinery production network.

5.3 Hierarchical clustering for the 2011 Asia machinery production network

Now we analyze the hierarchical dendrogram associated to the 2011 Asia machinery production network obtained from Definition 5.6. We hope that comparing the 2011 network hierarchical dendrogram (Figure 13) with the 2007 network hierarchical dendrogram (Figure 10) will afford insights into how the country-industry pairs' influence in the topological structure of the network changed in the wake of the financial crisis.

In contrast to the 2007 hierarchical dendrogram where $\text{JPN-C34} \notin \mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, t)$ when $t < 0.48$ while $\mathcal{B}_{\mathcal{H}_G}(\text{STANDARD}, 0.42) = \{S_G \setminus \{\text{JPN-C34}\}\}$, the 2011 hierarchical dendrogram (Figure 13) reflects that the absolute power of the node JPN-C34 in the network has significantly weakened; it now shares the influential role with another node: VNM-C31. The position change of JPN-C34 in the hierarchy further reflects that over the course of 2007 to 2011, the Japanese industry has stagnated [OK16] both in terms of its topological impacts in the structure of the network (reflected by the hierarchy) and its propagating influence through the entire network

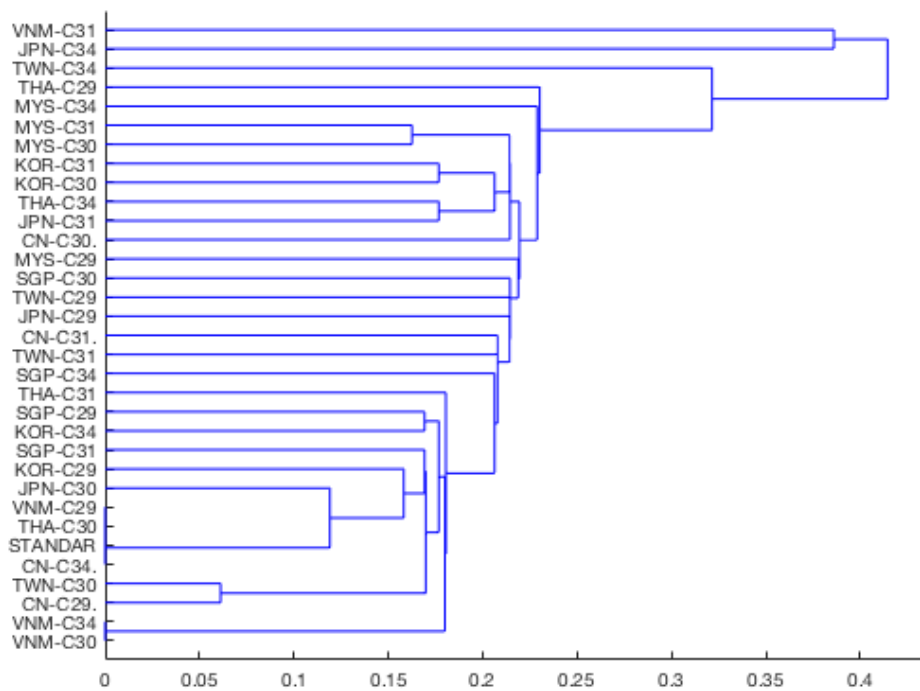


Figure 13: Hierarchical dendrogram associated to the 2011 Asia machinery production network.

(reflected by the quasi-centrality).

The Dowker persistence diagram of the original network $\gamma(G)$ is illustrated in Figure 14(a) and the Dowker persistence diagrams obtained by deleting the node JPN-C34 (VNM-C31, resp.) from $\gamma(G)$ are illustrated in Figure 14(b) ((c), resp.) respectively. We see that both JPN-C34 and VNM-C31 had caused big changes in the network topology in terms of the 1-cycles. Indeed, because JPN-C34 and VNM-C31 are "well-connected" to other nodes, deleting these nodes have a high potential disrupt the 1-dimensional homological features.

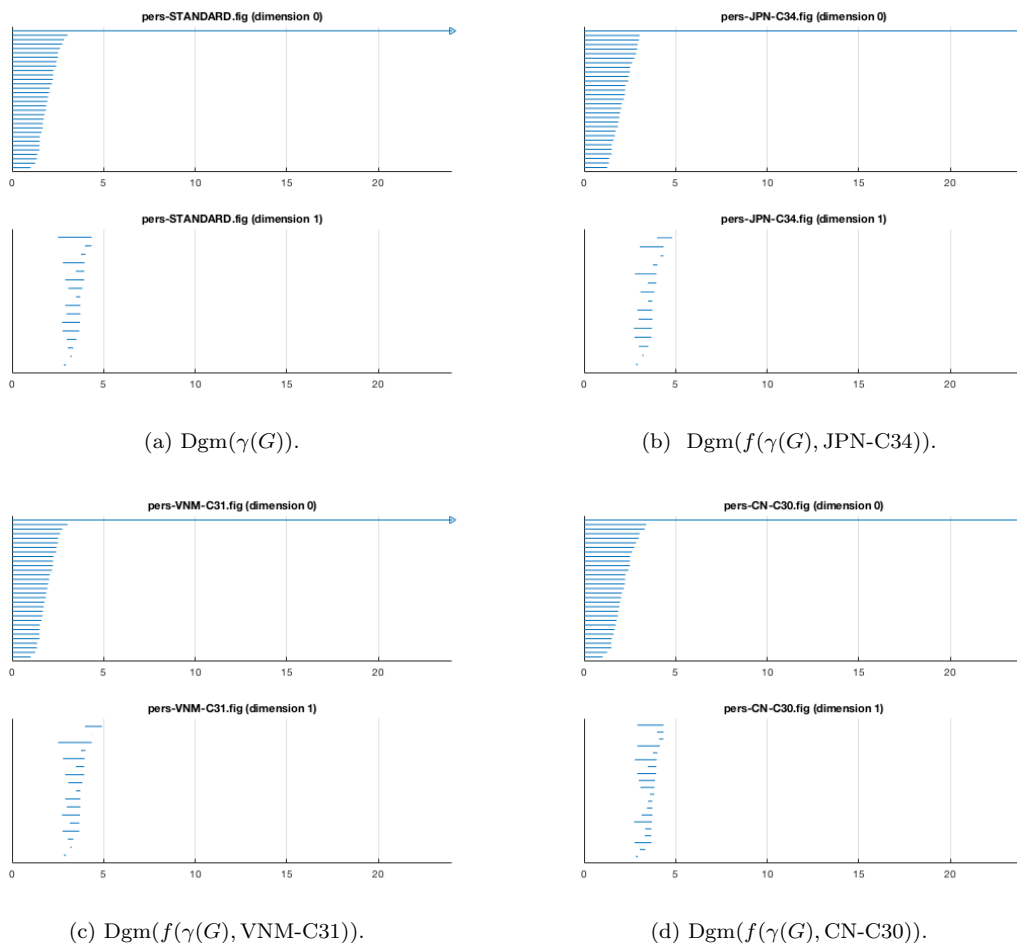


Figure 14: Dowker persistence diagrams for sub-networks obtained from the 2011 Asia machinery production network.

Although CN-C30 had the largest quasi-centrality value as computed in Section 4.2, and that the Dowker persistence diagram of $f(\gamma(G), \text{CN-C30})$ illustrated in 14(c) also reflects significant changes in the 0-dimensional barcodes compared to the Dowker persistence diagram of $\gamma(G)$, its impact on the 1-dimensional homological features is notably less than JPN-C34 and VNM-C31.

We note that the hierarchical clustering method incorporates higher dimensional topology in the network while the quasi-centrality gives a direct measure of the extent of propagating influence of a node in terms of changes in the 0-dimensional homological features.

5.4 Summary

We used the bottleneck distance between persistence diagrams obtained from the Dowker sink complex, to cluster nodes by their impact on the topology of the network, with a high emphasis on higher order homological features. We showed that our hierarchical clustering dendrogram reflects both the extent to which each node changes the topology of the network and which collections of persistence diagrams have similar topological effects.

We note the key distinction between computing quasi-centrality on nodes and hierarchical clustering: while the quasi-centrality gives a direct measure of the node’s propagating influence by recording how frequently the node serves as a bridge between path components, i.e., 0-dimensional homological features, the hierarchical clustering method directly compares the Dowker persistence diagrams, which includes higher dimensional homological features. While the deletion of a node can potentially have a smaller impact on the 0-dimensional persistence barcodes, it might disrupt 1-cycles or higher dimensional cycles, illustrating its impact in the overall network topology.

6 Discussion

Summary In this paper, we apply persistent homology to develop a network centrality measure *quasi-centrality* that captures propagating influences in directed networks. We computed the quasi-centrality on the Asia machinery production network and demonstrated that our centrality measure accurately reflects the propagating influences of perturbations originating from individual nodes.

Furthermore, we introduced a method that allows us to express the hierarchy of nodes based on their impact in the topology of a directed network. By incorporating hierarchical clustering and the bottleneck distance, we are able to not only detect which nodes play a significant role in the topology of the network, but also determine which nodes change the topology of the network in a similar fashion.

Future directions We expect that many other properties in networks (directed or undirected) admit characterizations using tools in TDA: one direction for future research is to define other measures in network analysis using TDA, such as density, robustness, efficiency, connectivity, etc. We expect that using TDA to investigate network properties allows us to bring a different perspective to examine networks.

Another open question is how to relate higher dimensional homological features in the simplicial complexes that arises from networks to real-world phenomena, such as trade flow, embargo, value and supply chains, patterns of trade, biological cycles, etc.

Another direction for future research is to apply the quasi-centrality measure on other directed networks. We expect that the quasi-centrality can not only be useful in assessing the influence of individual components within industries or trade networks, but also in other highly clustered complex directed networks, such as biological networks, air flight networks, etc.

We believe that many network structures such as trade flows, supply chains, and biological cycles can be better understood from a topological point of view by employing tools in TDA. Just as TDA has been successfully applied to detect intrinsic shapes in complex data sets, we believe that TDA can also find interesting structures in network topology and detect structural properties in networks, allowing us to better understand these networks, and gain more insights into the complex systems they model.

Limitations We note that although the quasi-centrality measure and the hierarchical clustering function we defined in this paper can be useful for detecting nodes’ propagating influences and topological impacts in a given directed network, computations and dendrograms generated are highly computationally costly as the number of nodes in the network exceeds 70-80 nodes.

7 Acknowledgements

I would like to sincerely thank my mentor Lucy Yang of Harvard University for providing thorough guidance and invaluable advice. I would like to thank the MIT-PRIMES program for providing me this invaluable opportunity and resources to work on this project. I would like to thank Professor Memoli of the Ohio State University and Dr. Chowdhury of Stanford University for their generous help and discussions. I would also like to thank Dr. Tanya Khovanova and Dr. Kent Vashaw of MIT for carefully proofreading the paper and offering helpful suggestions.

References

- [Aga20] Sarit Agami. *Comparison of Persistence Diagrams*. 2020. arXiv: [2003.01352](https://arxiv.org/abs/2003.01352) [stat.AP].
- [BH13] D. Brockmann and D. Helbing. “The Hidden Geometry of Complex, Network-Driven Contagion Phenomena”. In: *Science* 342.6164 (2013), 1337–1342. DOI: [10.1126/science.1245200](https://doi.org/10.1126/science.1245200).
- [Che+15] Sofya Chepushtanova et al. “Persistence Images: An Alternative Persistent Homology Representation”. In: *CoRR* abs/1507.06217 (2015). arXiv: [1507.06217](https://arxiv.org/abs/1507.06217). URL: <http://arxiv.org/abs/1507.06217>.
- [CM18] Samir Chowdhury and Facundo Mémoli. “A functorial Dowker theorem and persistent homology of asymmetric networks”. In: *Journal of Applied and Computational Topology* 2.1-2 (2018), 115–175. DOI: [10.1007/s41468-018-0020-6](https://doi.org/10.1007/s41468-018-0020-6).
- [Cos+11] Luciano da Fontoura Costa et al. “Analyzing and modeling real-world phenomena with complex networks: a survey of applications”. In: *Advances in Physics* 60.3 (2011), 329–412. ISSN: 1460-6976. DOI: [10.1080/00018732.2011.572452](https://doi.org/10.1080/00018732.2011.572452). URL: <http://dx.doi.org/10.1080/00018732.2011.572452>.
- [Dab20] Yuri Dabaghian. “From Topological Analyses to Functional Modeling: The Case of Hippocampus”. In: *Frontiers Comput. Neurosci.* 14 (2020), p. 593166. DOI: [10.3389/fncom.2020.593166](https://doi.org/10.3389/fncom.2020.593166). URL: <https://doi.org/10.3389/fncom.2020.593166>.
- [Eco] “The Network Origins of Aggregate Fluctuations”. In: *Econometrica* 80.5 (2012), 1977–2016. DOI: [10.3982/ecta9623](https://doi.org/10.3982/ecta9623).
- [Fic+21] Annamaria Ficara et al. *Correlation analysis of node and edge centrality measures in artificial complex networks*. 2021. arXiv: [2103.05427](https://arxiv.org/abs/2103.05427) [cs.SI].
- [Gof17] Michael Gofman. “Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions”. In: *Journal of Financial Economics* 124.1 (2017), pp. 113–146. ISSN: 0304-405X. DOI: <https://doi.org/10.1016/j.jfineco.2016.12.009>. URL: <https://www.sciencedirect.com/science/article/pii/S0304405X16302471>.
- [Hah+15] Youjin Hahn et al. *Teams, Organization and Education Outcomes: Evidence from a field experiment in Bangladesh*. CEPR Discussion Papers 10631. C.E.P.R. Discussion Papers, 2015. URL: <https://EconPapers.repec.org/RePEc:cpr:ceprdp:10631>.
- [Hoo22] Hian Teck Hoon. *The Singapore economy dynamism and inclusion*. Routledge, 2022.
- [Ian+17] Flavio Iannelli et al. “Effective distances for epidemics spreading on complex networks”. In: *Physical Review E* 95.1 (2017). ISSN: 2470-0053. DOI: [10.1103/physreve.95.012313](https://doi.org/10.1103/physreve.95.012313). URL: <http://dx.doi.org/10.1103/PhysRevE.95.012313>.
- [KPD17] Yevgeniya Korniyenko, Magali Pinat, and Brian Dew. “Assessing the Fragility of Global Trade: The Impact of Localized Supply Shocks Using Network Analysis”. In: *IMF Working Papers* 17.30 (2017), p. 1. DOI: [10.5089/9781475578515.001](https://doi.org/10.5089/9781475578515.001).

- [Kul02] V. V. Kulish. *Hierarchical methods*. Kluwer Academic Publishers, 2002.
- [Lee+21] Jongshin Lee et al. “Betweenness centrality of teams in social networks”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31.6 (2021), p. 061108. ISSN: 1089-7682. DOI: [10.1063/5.0056683](https://doi.org/10.1063/5.0056683). URL: <http://dx.doi.org/10.1063/5.0056683>.
- [Lei+20] Ian Leifer et al. “Circuits with broken fibration symmetries perform core logic computations in biological networks”. In: *PLOS Computational Biology* 16.6 (2020). Ed. by MercedesEditor Pascual, e1007776. ISSN: 1553-7358. DOI: [10.1371/journal.pcbi.1007776](https://doi.org/10.1371/journal.pcbi.1007776). URL: <http://dx.doi.org/10.1371/journal.pcbi.1007776>.
- [Liu+20] Jingxin Liu et al. “K-Core based Temporal Graph Convolutional Network for Dynamic Graphs”. In: *IEEE Transactions on Knowledge and Data Engineering* (2020), 1–1. ISSN: 2326-3865. DOI: [10.1109/tkde.2020.3033829](https://doi.org/10.1109/tkde.2020.3033829). URL: <http://dx.doi.org/10.1109/TKDE.2020.3033829>.
- [LSV20] Alane M. de Lima, Murilo V. G. da Silva, and André L. Vignatti. *Estimating the Percolation Centrality of Large Networks through Pseudo-dimension Theory*. 2020. arXiv: [1910.00494](https://arxiv.org/abs/1910.00494) [cs.DS].
- [Mas+18] Naoki Masuda et al. “Clustering Coefficients for Correlation Networks”. In: *Frontiers in Neuroinformatics* 12 (2018). ISSN: 1662-5196. DOI: [10.3389/fninf.2018.00007](https://doi.org/10.3389/fninf.2018.00007). URL: <http://dx.doi.org/10.3389/fninf.2018.00007>.
- [MG16] Sergey Melnik and James P. Gleeson. *Simple and accurate analytical calculation of shortest path lengths*. 2016. arXiv: [1604.05521](https://arxiv.org/abs/1604.05521) [physics.soc-ph].
- [Mun84] James R. Munkres. *Elements of Algebraic Topology*. Addison Wesley Publishing Company, 1984. ISBN: 0201045869. URL: <http://www.worldcat.org/isbn/0201045869>.
- [New03] M. E. J. Newman. “The Structure and Function of Complex Networks”. In: *SIAM Review* 45.2 (2003), pp. 167–256. DOI: [10.1137/S003614450342480](https://doi.org/10.1137/S003614450342480). eprint: <https://doi.org/10.1137/S003614450342480>. URL: <https://doi.org/10.1137/S003614450342480>.
- [OEC15] OECD. *Green Growth in Bangkok, Thailand*. 2015, p. 148. DOI: <https://doi.org/https://doi.org/10.1787/9789264237087-en>. URL: <https://www.oecd-ilibrary.org/content/publication/9789264237087-en>.
- [OEC21] OEC. *Life expectancy*. data retrieved from OEC-The Observatory of Economic Complexity, <https://oec.world/>. 2021.
- [OECD21] OECD. *OECD Inter-Country Input-Output (ICIO) Tables*. data retrieved from OECD Inter-Country Input-Output (ICIO) Tables, <https://www.oecd.org/sti/ind/inter-country-input-output-tables.htm>. 2021.
- [OK16] Ayako Obashi and Fukunari Kimura. “The Role of China, Japan, and Korea in Machinery Production Networks”. In: *International Economic Journal* 30.2 (2016), pp. 169–190. DOI: [10.1080/10168737.2016.1148398](https://doi.org/10.1080/10168737.2016.1148398). eprint: <https://doi.org/10.1080/10168737.2016.1148398>. URL: <https://doi.org/10.1080/10168737.2016.1148398>.
- [OR02] Evelien Otte and Ronald Rousseau. “Social network analysis: a powerful strategy, also for the information sciences”. In: *Journal of Information Science* 28.6 (2002), pp. 441–453. DOI: [10.1177/016555150202800601](https://doi.org/10.1177/016555150202800601). eprint: <https://doi.org/10.1177/016555150202800601>. URL: <https://doi.org/10.1177/016555150202800601>.
- [Ott+17] Nina Otter et al. “A roadmap for the computation of persistent homology”. In: *EPJ Data Science* 6.1 (2017). DOI: [10.1140/epjds/s13688-017-0109-5](https://doi.org/10.1140/epjds/s13688-017-0109-5).

- [RB20] Rabadan Raul and Andrew J Blumberg. *Topological data analysis for genomics and evolution: topology in biology*. Cambridge: Cambridge university press, 2020.
- [Rid05] Christine Rider. “Social Economics”. In: *Encyclopedia of Social Measurement* (2005), 501–508. DOI: [10.1016/b0-12-369398-5/00552-1](https://doi.org/10.1016/b0-12-369398-5/00552-1).
- [SBn03] Ma Ángeles Serrano and Marián Boguñá. “Topology of the world trade web”. In: *Phys. Rev. E* 68 (1 2003), p. 015101. DOI: [10.1103/PhysRevE.68.015101](https://doi.org/10.1103/PhysRevE.68.015101). URL: <https://link.aps.org/doi/10.1103/PhysRevE.68.015101>.
- [SBV07] M. Ángeles Serrano, Marián Boguñá, and Alessandro Vespignani. “Patterns of dominant flows in the world trade web”. In: *Journal of Economic Interaction and Coordination* 2.2 (2007), 111–124. ISSN: 1860-7128. DOI: [10.1007/s11403-007-0026-y](https://doi.org/10.1007/s11403-007-0026-y). URL: <http://dx.doi.org/10.1007/s11403-007-0026-y>.
- [Tay+16] Dane Taylor et al. *Eigenvector-Based Centrality Measures for Temporal Networks*. 2016. arXiv: [1507.01266](https://arxiv.org/abs/1507.01266) [[physics.soc-ph](https://arxiv.org/archive/physics)].
- [TZH15] Chad M. Topaz, Lori Ziegelmeier, and Tom Halverson. “Topological Data Analysis of Biological Aggregation Models”. In: *PLOS ONE* 10.5 (2015). Ed. by BardEditor Ermentrout, e0126383. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0126383](https://doi.org/10.1371/journal.pone.0126383). URL: <http://dx.doi.org/10.1371/journal.pone.0126383>.
- [VSB19] Vijaya, Shweta Sharma, and Neha Batra. “Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering”. In: *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*. 2019, pp. 568–573. DOI: [10.1109/COMITCon.2019.8862232](https://doi.org/10.1109/COMITCon.2019.8862232).
- [Wri15] James D. Wright. *International encyclopedia of the social behavioral sciences*. Elsevier, 2015.
- [Şi+21] Berfin Şimşek et al. *Geometry of the Loss Landscape in Overparameterized Neural Networks: Symmetries and Invariances*. 2021. arXiv: [2105.12221](https://arxiv.org/abs/2105.12221) [[cs.LG](https://arxiv.org/archive/cs)].