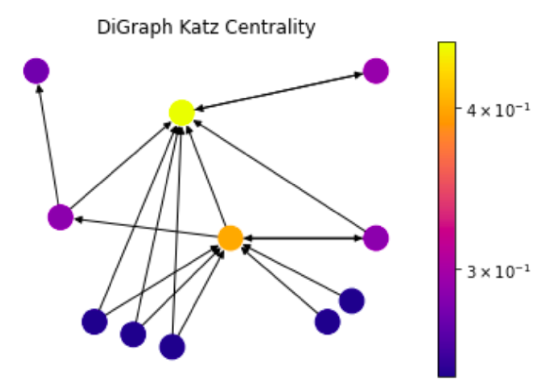
A Topological Centrality Measure for Directed Networks

Fenghuan (Linda) He lhe@commschool.org[1]
Mentor: Lucy Yang lyang@math.harvard.edu[12]
[1]MIT-PRIMES (Commonwealth School) [2]Harvard University

Background and Motivation

- **Networks** model complex systems as (directed) graphs
- **Node centrality** measures output a nodes' importance in a given network:
  - Betweenness centrality in social networks (J.Lee, 2021)
  - Eigenvector centrality in temporal networks (D.Taylor, 2016)

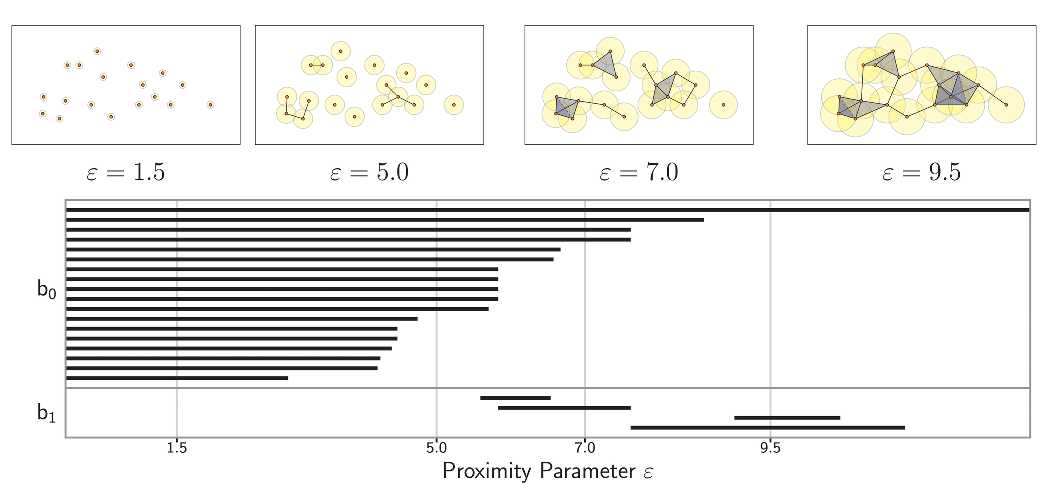


$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{\text{out}}} + \beta$$

Goal
Defining a topological centrality measure that captures propagating effects and directedness.

Topological Data Analysis (TDA)

Central Idea
Determining the intrinsic robust features of complex data sets.



Suppose data is embedded in an underlying topological space:

- Algebraic topology measures qualitative features in simplicial complexes.
- Persistent homology measures the "robustness" of homological features.

TDA in Networks

Definition
A **network** G is a pair (X, w_X) where X is a finite set and $w_X : X \times X \to \mathbb{R}$ is called the weight function.

Definition (F.Memoli and S.Chowdhury, 2016)

Given a network $G = (X, w_X)$ and $\delta \in \mathbb{R}$, let $R_{\delta,G} \subseteq X \times X$ be defined 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 Complex**:

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

The node p is called a sink for the simplex σ.

We denote $\mathsf{P}_n(G)$ as the set of n-dimensional persistence barcodes corresponding to the Dowker sink complex $\mathcal{D}_{\delta,G}$.

Quasi-centrality Measure

Goal
We measure the extent in which a node contributes to the connectivity of a network G.

Effective Distance

Let $G = (X, w_X)$ be a network, define $\gamma(G)$ to be (X, m_X) where $m_X : X \times X \to \mathbb{R}$ is given by:

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

Main Definition (Quasi-centrality Measure)

Given a network $G = (X, w_X)$, let $f(G, x) = (X \setminus \{x\}, w_X)$. Moreover, given $x \in X$, let $\mu(x)$ to be the minimum distance between x and any other node $x' \in X$ with respect to the metric m_X. We define the **quasi-centrality** $C(x)$ for $x \in X$ as follows:

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

where b represents a persistence interval in P_0 and $l(b)$ is the persistence of the interval.

Theorem

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

idea
As the value of δ increases, $C(x)$ measures how often node x bridges between two disconnected components.

Proof.

- $\gamma(G)$ can be treated as both a topological space with underlying homological features and as a metric space.
- The sum $\sum_{b \in \mathsf{P}_0(\gamma(G))} l(b)$ as a measure of how **disconnected** $\gamma(G)$ is.
- The 0-th homology group $\mathsf{P}_0(\gamma(G))$ measures the number of **disconnected path components** in $\mathcal{D}_{\gamma(G)}$, hence $\sum_{b \in \mathsf{P}_0(\gamma(G))} l(b)$ can be used as a proxy of the **size** of the 0-th homology group.
- Similarly, the sum $\sum_{b \in \mathsf{P}_0(f(\gamma(G),x))} l(b)$ as a measure of 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.

Applications

Trade Networks:

- Interdependency between local and global communities.
- Economic perturbations originated in a single country can propagate elsewhere.

Previous Studies:

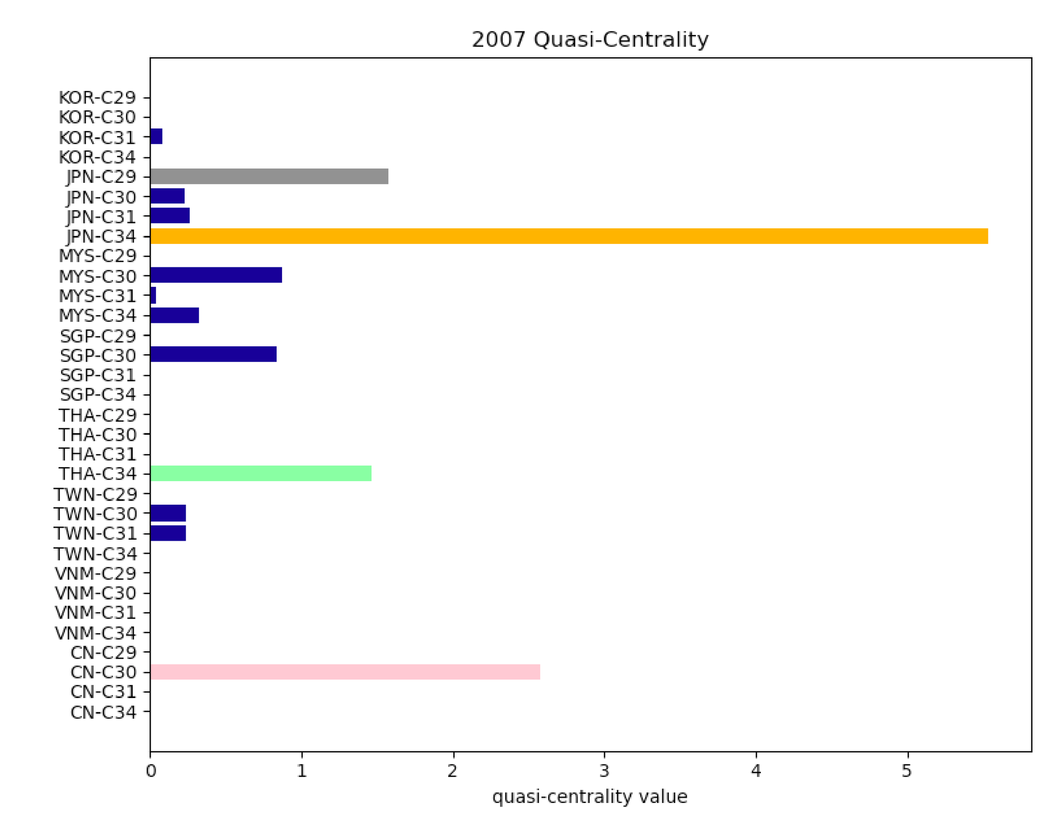
- Assessing the Fragility of Global Trade (Y.Korniyenko, 2017).
- International Trade Network: Country centrality and COVID-19 pandemic (R.Antonietti, 2021).

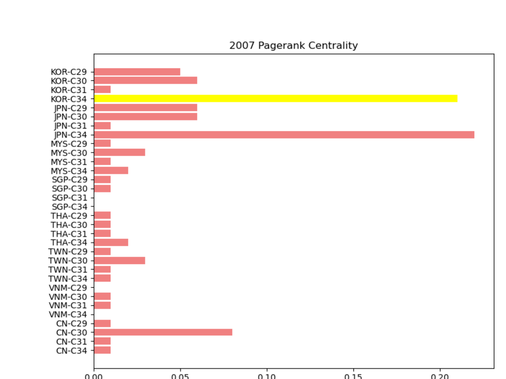
Goal
Assess the influence of a node in a trade network by computing the quasi-centrality measure.

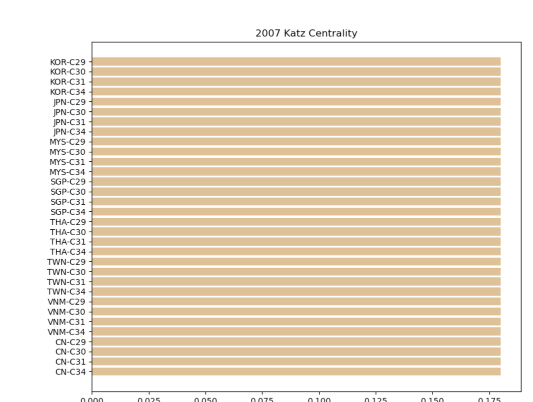
Data:

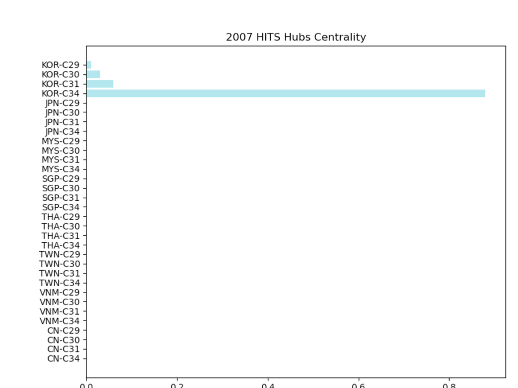
- OECD Inter-Country Input-Output (ICIO) Tables
- Machinery production network in Asia during 2007 (financial crises)
- Industries: machinery equipment, computer and electronics, electrical machinery, auto machinery

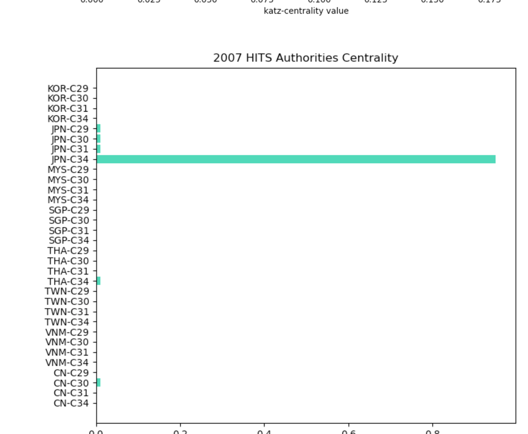
Selected Results











Hierarchical Clustering

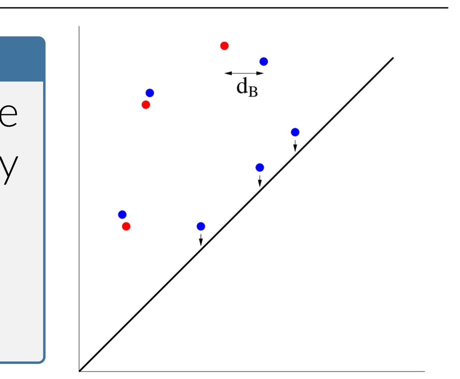
Goal
Determine the hierarchy of nodes in terms of their impact in the topology of the network.

Hierarchical clustering arranges the objects of a set into a hierarchy of groups using a specified weight function.

Bottleneck Distance

Definition
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

$$d_{B_\infty}(D_1, D_2) = \inf_{\eta: D_1 \to D_2} \sup_{x \in D_1} \|x - \eta(x)\|_\infty$$



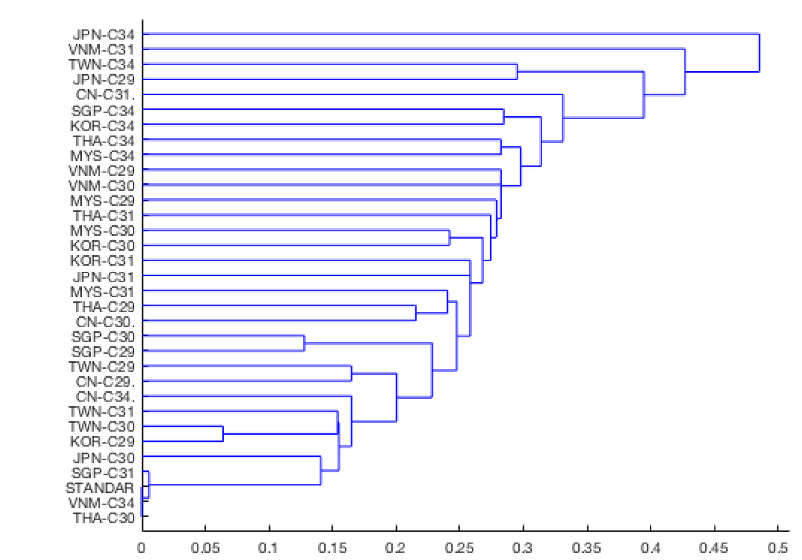
Definition

Define S_G to be the following set:

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

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

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



Future Work

- Many properties in networks admit characterizations using tools in TDA. (Density, robustness, efficiency, etc.)
- Relate higher dimensional homological features in networks to real-world phenomena.
- Apply the quasi-centrality measure on other directed networks.

Acknowledgement and references

This work is conducted as part of the MIT-PRIMES program. I am grateful for my mentor Lucy Yang of Harvard University, my advisor Prof. Memoli of the Ohio State University, and the MIT-PRIMES organizers for offering this research opportunity.

[1] Samir Chowdhury and Facundo Mémoli. A functorial dowker theorem and persistent homology of asymmetric networks. *Journal of Applied and Computational Topology*, 2(1-2):115–175, 2018.

[2] Yevgeniya Korniyenko, Magali Pinat, and Brian Dew. Assessing the fragility of global trade: The impact of localized supply shocks using network analysis. *IMF Working Papers*, 17(30):1, 2017.

[3] Jongshin Lee, Yongsun Lee, Soo Min Oh, and B. Kahng. Betweenness centrality of teams in social networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(6):061108, Jun 2021.

[4] Dane Taylor, Sean A. Myers, Aaron Clauset, Mason A. Porter, and Peter J. Mucha. Eigenvector-based centrality measures for temporal networks, 2016.