On The Limiting Behavior of Parameter Dependent Network Centrality Measures

Michele Benzi and Christine Klymko

Presented by: Himanshu Kharkwal

Goals

- a broad class of walk-based, parameterized node centrality measures are considered for network analysis.
- measures are expressed in terms of functions of the adjacency matrix and generalize various well-known centrality indices, including Katz and subgraph centrality.
- To show that the parameter can be “tuned” to interpolate between degree and eigenvector centrality, which appear as limiting cases.
- To explain certain correlations often observed between the rankings obtained using different centrality measures, and provide some guidance for the tuning of parameters

Motivation

One of the most basic questions about network structure is the identification of the "important" nodes in a network and accurate rankings are becoming increasingly more important in many fields, such as study of

- human interactions, e.g., via social networks
- Target marketing
- Government infrastructure spending, e.g., power grids and rail routes
- Study of protein–protein interactions, the basis for diseases such as Alzheimer's and cancer.

Introduction

- computational measures of node importance, called centrality measures, are used to rank the nodes in a network
- centrality measures often provide rankings that are highly correlated, at least when attention is restricted to the most highly ranked nodes
- considering centrality measures based on functions of the adjacency matrix, in addition to degree and eigenvector centrality

Background and Definitions

- A **walk** of length k in G is a list of nodes $i_1,i_2,...i_k,i_{k+1}$ such that for all $1 \leq l \leq k$, there is a edge between i_l and i_{l+1}.
- A **closed walk** is a walk where $i_1 = i_{k+1}$
- A graph is **simple** if it has no loops, no multiple edges, and unweighted edges.
- An undirected graph is **connected** if there exists a path between every pair of nodes

Properties of Adjacency matrix

If G is a simple, undirected graph

- A is binary and symmetric with zeros along the main diagonal
- the eigenvalues of A will be real.

If G is connected, then

- $\lambda 1 > \lambda 2 ... \geq \lambda n$ by the Perron-Frobenius theorem.
- A can be decomposed into $A = Q \Lambda Q^T$ where $\Lambda = diag(\lambda 1, \lambda 2, ..., \lambda n)$, $Q=[q1,q2,...,qn]$ is orthogonal
- The dominant eigenvector,q1, can be chosen to have positive entries when G is connected: we write this $q1>0$.

Common centrality measures

- Degree centrality
- Eigenvector centrality
- Exponential Subgraph centrality
- Resolvent Subgraph centrality
- Total Communicability
- Katz centrality

Some other famous centrality measures

- Betweenness centrality
- Closeness centrality

Degree centrality

- simple centrality measure that counts how many neighbors a node has.
- If the network is directed:
  - in-degree is the number of incoming links
  - out-degree is the number of outgoing links
- $d_i = [A\ 1]_i$
- Degree of each vertex is sum of rows/columns of the adjacency matrix A.

Eigenvector centrality

- A node is important if it is linked to by other important nodes.
- It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

- Where **M(v)** is a set of the neighbors of **v** and **λ** is a constant. With a small rearrangement this can be rewritten in vector notation as the eigenvector equation: **Ax = λx**

Exponential Subgraph centrality

- subgraph centrality measure of a node in a network as the weighted sum of the numbers of closed walks of different lengths that start and end at that node.
- Specifically, if A is the adjacency matrix of a network, the numbers of closed walks of length k are the diagonal entries of the matrix A_k.
- Decreasing weights with path length ensures the convergence of the series while guaranteeing that short-range interactions are given more weight than long-range ones

$$SC_i(\boldsymbol{\beta}) = [e^{\boldsymbol{\beta} A}]_{ii}$$

$$e^{\beta A} = I + \beta A + \frac{(\beta A)^2}{2!} + \cdots + \frac{(\beta A)^k}{k!} + \cdots = \sum_{k=0}^{\infty} \frac{(\beta A)^k}{k!}.$$

Total Communicability

- Total communicability is closely related to subgraph centrality.
- This measure also counts the number of walks starting at node i, scaling walks of length k by **β**/k!
- However, rather than just counting closed walks, total communicability counts all walks between node i and every node in the network
- Represents the row sums of the exponential subgraph centrality

$$TC = [e^{\beta A} \mathbf{1}]_i$$

Exponential vs Resolvent based centralities

Exponential : Penalize long walks by α_k = 1/k!,so the ordinal node rankings are obtained from the centrality vector

$$C_{\mathbf{exp}}(A) = e^{A}\mathbf{1}.$$

$$e^{\beta A} = I + \beta A + \frac{(\beta A)^2}{2!} + \cdots + \frac{(\beta A)^k}{k!} + \cdots = \sum_{k=0}^{\infty} \frac{(\beta A)^k}{k!}.$$

Resolvent : Penalize long walks by $\alpha_k = \alpha^k$,so for $\alpha \in (0,1/\lambda_1)$the ordinal node rankings are obtained from the centrality vector C_{res}, where λ_1 is the largest eigenvalue of A and I is the identity matrix.

$$C_{\mathbf{res}}(A) = (I - \alpha A)^{-1}\mathbf{1},$$

$$(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \cdots + \alpha^k A^k + \cdots = \sum_{k=0}^{\infty} \alpha^k A^k.$$

Resolvent Subgraph centrality

- The resolvent subgraph centrality of node i, $[(I-\alpha A)^{-1}]_{ii}$, counts the total number of closed walks in the network which are centered at node i, weighing walks of length k by α_k
- penalize longer walks which can be seen by considering the Neumann series expansion of $(I-\alpha A)^{-1}$, valid for $0 < \alpha < 1/\lambda_1$

$$RC_i(\boldsymbol{\alpha}) = [(I-\alpha A)^{-1}]_{ii}$$

$$(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \cdots + \alpha^k A^k + \cdots = \sum_{k=0}^{\infty} \alpha^k A^k.$$

Katz centrality

- Katz centrality of node i counts all walks beginning at node i, penalizing the contribution of walks of length k by α_k

$$K_i(\boldsymbol{\alpha}) = [(I-\alpha A)^{-1} 1]_i$$

The bounds on $\alpha (0 < \alpha < 1/\lambda_1)$ ensure that the matrix $I-\alpha A$ is invertible and that the power series converges to its inverse. The bounds on α also force $[(I-\alpha A)^{-1}]_{ii}$ to be nonnegative.

Governing Conditions for Admissible Matrix functions

- function should be defined by a power series with real coefficients, such that f(A) has real entries for any real A

  $$f(z) = \sum_{k=0}^{\infty} c_k z^k, \quad c_k \geq 0 \quad \text{for} \quad k \geq 0\,.$$

  further required that $c_k > 0$ for all k= 1,2,...,n−1, so as to guarantee that $[f(A)]_{ij} > 0$ for all i!=j

- Introducing a (scaling) parameter t, and considering parameterized matrix function f(tA) for A for values of t such that the power series is convergent

  $$f(tA) = c_0 I + c_1 tA + c_2 t^2 A^2 + \cdots = \sum_{k=0}^{\infty} c_k t^k A^k$$

Limiting Behavior of Parametrized Centrality Measures

Limits: $$\alpha \to 0+, \quad \alpha \to \frac{1}{\lambda_1}-, \quad \beta \to 0+, \quad \beta \to \infty\,.$$

- Seen:
  - the actual scores may vary by orders of magnitude
  - rankings are quite stable, in the sense that they do not appear to change much for different choices of α and β
- Expectation:
  - Same behavior when using parameterized centrality measures based on analytic functions f

Undirected Case

G = (V,E) be a connected, undirected, unweighted network with adjacency matrix A, assumed to be primitive

- As t→0+, the rankings produced by both SC(t) and TC(t) converge to those produced by d = (di), the vector of degree centralities
- As t→t∗− the rankings produced by both SC(t) and TC(t) converge to those produced by eigenvector centrality, i.e., by the entries of q1, the dominant eigenvector of A

Method	Limiting ranking scheme: degree	Limiting ranking scheme: eigenvector
$RC(\alpha)$, $K(\alpha)$	$\alpha \to 0+$	$\alpha \to \frac{1}{\lambda_1}-$
$EC(\beta)$, $TC(\beta)$	$\beta \to 0+$	$\beta \to \infty$

PageRank

$G = (V,E)$

$H = A^T D^{-1}$, where A is the adjacency matrix and D is a special diagonal matrix

$S = H + 1/n[\mathbf{1}\, a^T]$, where $a_i = 1$ for indices which have zero out degree, else 0

To obtain an irreducible matrix, we take $\alpha \in (0,1)$ and construct the “Google matrix”

$P = \alpha S + (1-\alpha) v \mathbf{1}^T$, where v is an arbitrary probability distribution vector

Normalized eigenvector of P, p is called the pagerank vector and it is used to rank nodes in the digraph

Possible reformulation of the pagerank problem can be given by the linear system of equations

$$(I - \alpha H)x = v, \qquad p = x/(x^T \mathbf{1})$$

For each $\alpha \in (0,1)$, the coefficient matrix is nonsingular, hence it is invertible with a nonnegative inverse $(I - \alpha H)^{-1}$, which is quite similar to Katz centrality.

Using this equivalence, we can easily describe the limiting behavior of PageRank for $\alpha \to 0+$, showing that the rankings from $p(\alpha)$ coincide with those from the row sums of H

Discussion

- the degree centrality of node i measures the local influence of i and the eigenvector centrality measures the global influence of i.
- When the centrality measures associated with an analytic function $f \in P$,
  - walks of all lengths are included
  - a weight c_k is assigned to the walks of length k, where $c_k \to 0$ as $k \to \infty$. Hence, both local and global influence are now taken into account, but with longer walks being penalized more heavily than shorter ones.

- Parameter t permits further tuning of the weights,
  - As t is decreased, the weights corresponding to larger k decay faster and shorter walks become more important
  - As t is increased, walks of “infinite” length dominate and the centrality rankings converge to those of eigenvector centrality.
- Parameterized centrality measures are likely to be most useful when both local and global influence need to be considered in the ranking of nodes in a network

For Quantitative assessment, estimate how fast the limiting rankings given by degree and eigenvector centrality are approached for $t \to 0+$ and $t \to t_*-$

- when the spectral gap is large, the rankings obtained using parameterized centrality will converge to those given by eigenvector centrality more quickly as t increases than in the case when the spectral gap is small
- we can expect the degree centrality limit to be attained more rapidly, for $t \to 0+$, for networks with low clustering coefficient than for networks with high clustering coefficient

Questions!