Lecture 07
Network Analysis 2

Andrew Sonta

CIVIL 534: Computational systems thinking for sustainable engineering

09 April 2025

Outline

- Components
- Graph Laplacian and spectral partitioning
- Centrality
  - Degree
  - Spectral
    - Eigenvector, Katz, PageRank
  - Path-based
    - Closeness
    - Betweenness

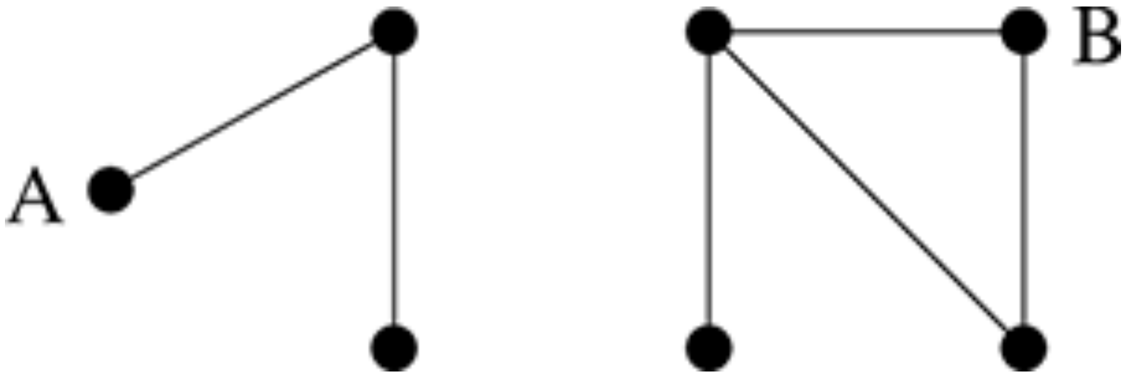
- Corresponding parts of Newman: 6.12; 6.14; 7.1

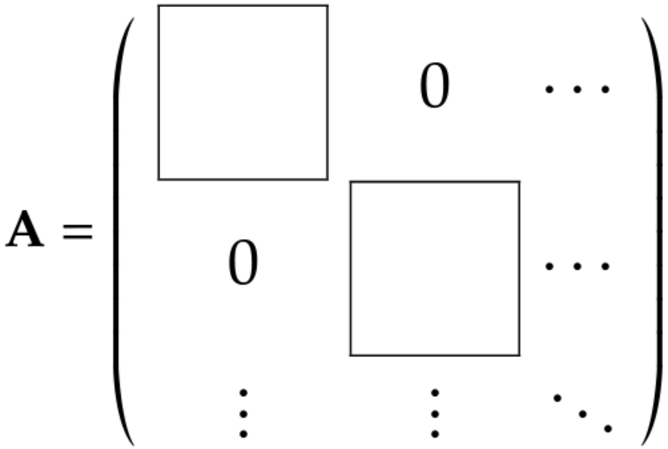
But first…

- Any questions on cut sets, independent paths/connectivity, and min-cut/max-flow theorem?

Components

- Components are parts of the network that are not connected to one another
  - No path from A to B or B to A

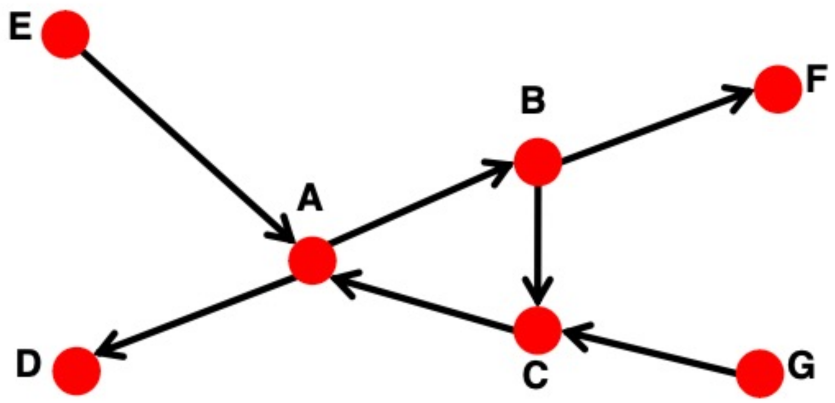




Components can indicate intervention points to increase connectivity

Components on directed graphs

- Extension is not straightforward
- Strongly connected components: there is a path from each node to each other node and vice versa (e.g., A-B and B-A)
- Weakly connected components: connected if we disregard edge directions
  - “In-component”
  - “Out-component”



Graph Laplacian

- The Adjacency matrix is not the only matrix representation of a network
- Graph Laplacian is useful for many different network properties
- Derived from the notion of “diffusion” (something moving along the edges of the network)
- Defined for undirected networks, no direct extension to directed networks
  - Can translate directed network to undirected network if needed

Graph Laplacian - definition

degree of node i

$$L_{ij} = \begin{cases} k_i & \text{if } i = j, \\ -1 & \text{if } i \neq j \text{ and there is an edge between nodes } i \text{ and } j, \\ 0 & \text{otherwise,} \end{cases}$$

Equivalently: $L_{ij} = k_i \delta_{ij} - A_{ij}$

1 iff i=j (Kronecker delta)

Equivalently: L = D – A

$$\mathbf{D} = \begin{pmatrix} k_1 & 0 & 0 & \cdots \\ 0 & k_2 & 0 & \cdots \\ 0 & 0 & k_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

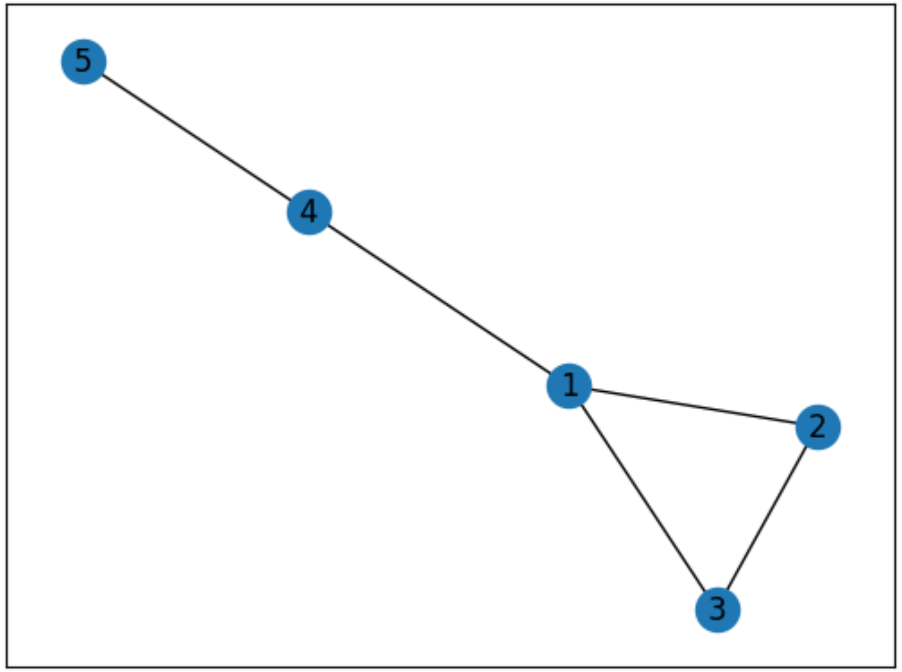
Graph Laplacian - eigenvalues

- Eigenvalues greater than or equal to 0
  - Diffusion intuition: over time we reach equilibrium in the network
- For network with 1 component, there is 1 eigenvalue equal to 0
  - For a network with n components, there are n eigenvalues equal to 0
- The second-smallest eigenvalue/eigenvector pair is a critical value
  - The second-smallest eigenvalue is called the **algebraic connectivity**
    - Measure of how well connected the network is
  - The pair can be used in **spectral partitioning/clustering**

Spectral partitioning

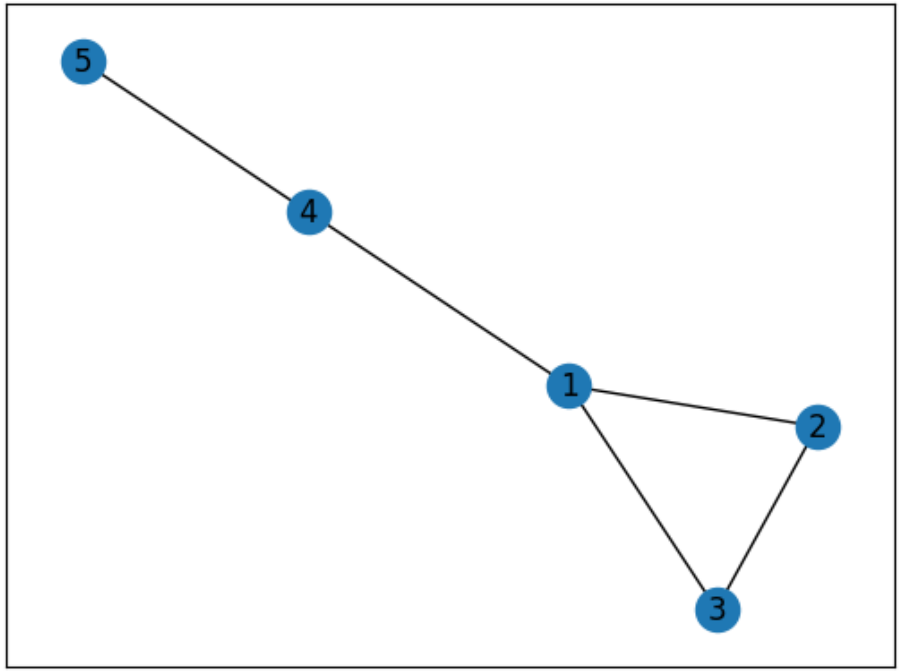
- Given a graph with adjacency matrix A, we want to partition the graph into two subgraphs such that:

$$\min\left(\frac{\#\ edges\ across\ cut}{\#\ node\ pairs\ across\ cut\ that\ could\ support\ edges}\right)$$



Spectral partitioning – Fiedler method

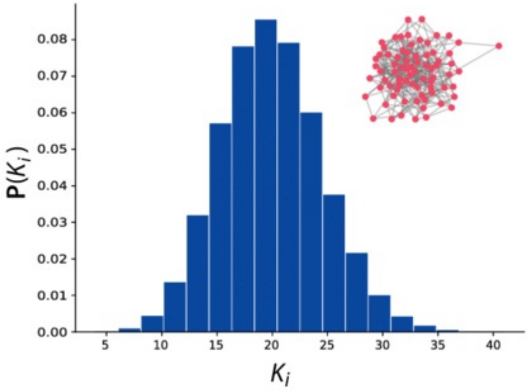
- Second-smallest eigenvalue of L is the algebraic connectivity (less than or equal to vertex connectivity)
- Corresponding eigenvector can be used to separate the graph into 2 communities based on the sign of the vector entry
- Let’s try it!



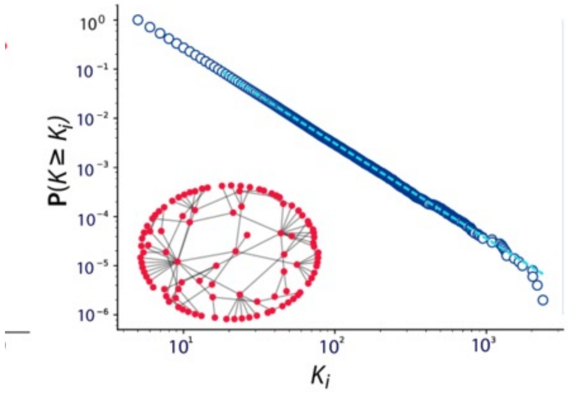
Network metrics

Network metrics – why?

- Network structure in theory tells us everything we need to know about the network
- But in practice, networks are often large and difficult to comprehend
- Metrics and measures are used to distill network information into interpretable values
  - How we interpret each value depends on the way in which the metric/measure was constructed



a) ER



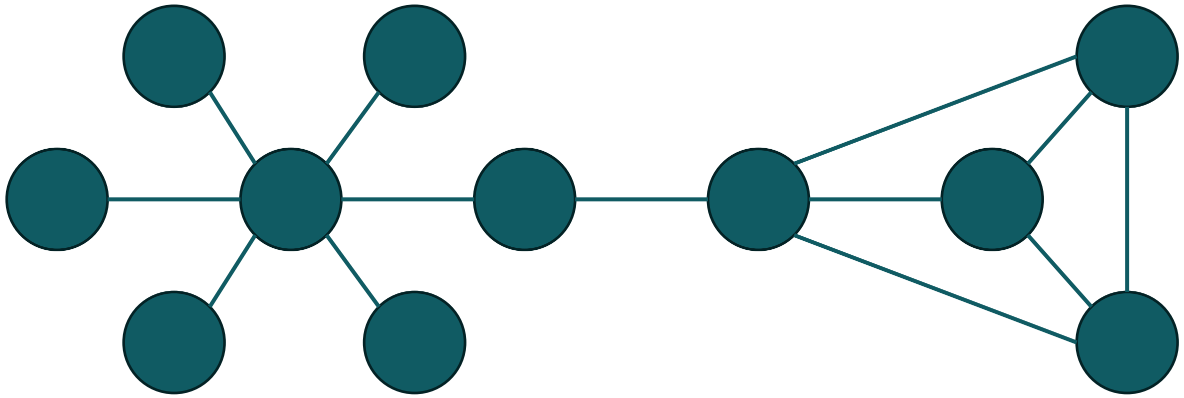
a) BA

Centrality

- Which are the most important or central nodes in a network?
- Many centrality metrics (many ways to define “importance”)

Centrality

- Which are the most important or central nodes in a network?
- Many centrality metrics (many ways to define “importance”)
- Which is the most important node in the following network?



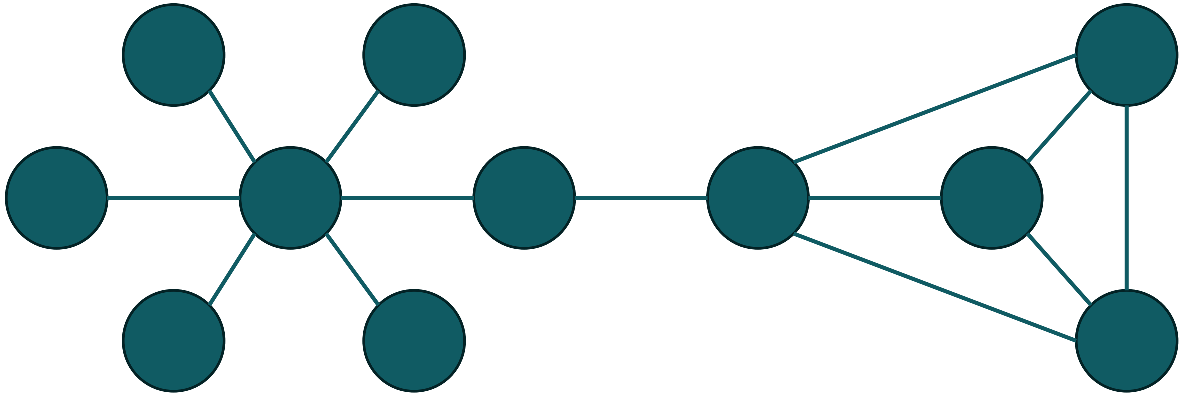
Centrality measures

- Degree centrality
- Spectral centrality measures:
  - Eigenvector centrality
  - Katz
  - PageRank
- Path-based centrality measures:
  - Closeness centrality
  - Betweenness centrality

Degree centrality

- Simplest
- Just the degree of the node
- Undirected: degree
- Directed:
  - In-degree centrality
  - Out-degree centrality

  - **Choose based on application!**



Eigenvector centrality

- Key limitation of degree centrality: it assigns no value to **which nodes** a given node is connected to
- Eigenvector centrality: For a given node's neighbors, add a score proportional to the centrality of each neighbor

$$x_i = \kappa^{-1} \sum_{\substack{\text{nodes } j \text{ that are} \\ \text{neighbors of } i}} x_j$$

$$x_i = \kappa^{-1} \sum_{j=1}^{n} A_{ij} x_j \longrightarrow \mathbf{Ax} = \kappa \mathbf{x}$$

We choose the leading eigenvalue (largest) and associated eigenvector (only eigenvector with all elements non-negative for **A**)

Eigenvector centrality – intuition

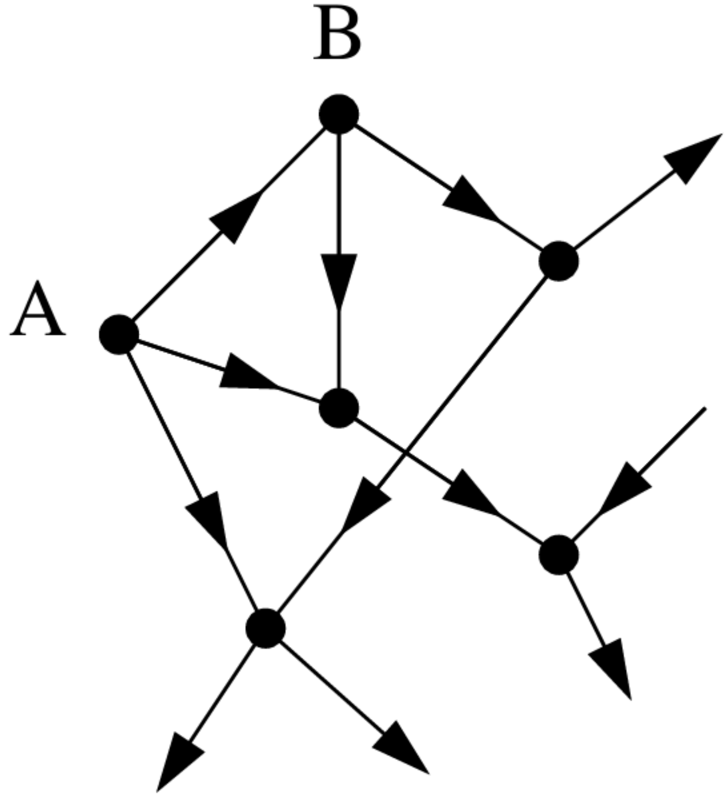
- Each node starts with the same score, and then each node gives away its score to its neighbors (repeat this process)
  - Intuitively: degree counts walks of length 1, eigenvector centrality counts walks of length infinity
- Procedure
  - $\mathbf{c}^{(k)} = \mathbf{A}\mathbf{c}^{(k-1)}$
  - $\mathbf{c}^{(k)} = \mathbf{c}^{(k)} / \|\mathbf{c}^{(k)}\|_2$
  - $k = k + 1$
  - Repeat until convergence

Eigenvector centrality – directed networks

- Non-symmetric adjacency matrix
  - Left and right eigenvectors → which to use?
- Generally, we use the right-eigenvector
  - The rationale is that importance is based more on incoming edges
- The question of extension to directed networks led the development of variants of eigenvector centrality (Katz, PageRank)

Katz centrality

- What happens when a node pointing to another node has zero centrality?
- Katz centrality adds a *free term* to eigenvector centrality
  - All nodes get some centrality for free



$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$

$$\mathbf{x} = \alpha \mathbf{A}\mathbf{x} + \beta \mathbf{1}$$

$$\mathbf{x} = \beta(\mathbf{I} - \alpha\mathbf{A})^{-1}\mathbf{1}$$

$$\mathbf{x} = (\mathbf{I} - \alpha\mathbf{A})^{-1}\mathbf{1}$$

Katz centrality – issues

- If edges are cheap to form (think webpages) then an important node can easily share its centrality with those it points to
- Internet example: Amazon links to millions of pages (e.g., manufacturers)
  - If an important website is very generous with its links, should that count the same as an important link that is sparing with its links?
- Enter PageRank (the first version of Google)
  - Insight: Normalize the score of an incoming edge by the out-degree centrality of the originating node

PageRank centrality

- Extension of Katz centrality in which out-degrees are normalized:

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

$$\mathbf{x} = \alpha \mathbf{A}\mathbf{D}^{-1}\mathbf{x} + \beta\mathbf{1}$$ where $D_{ii} = \max(k_i^{\text{out}}, 1)$

$$\mathbf{x} = \beta(\mathbf{I} - \alpha\mathbf{A}\mathbf{D}^{-1})^{-1}\mathbf{1}$$

$$\mathbf{x} = (\mathbf{I} - \alpha\mathbf{A}\mathbf{D}^{-1})^{-1}\mathbf{1}$$

It is believed that early on, Google used α=0.85

Centrality comparisons

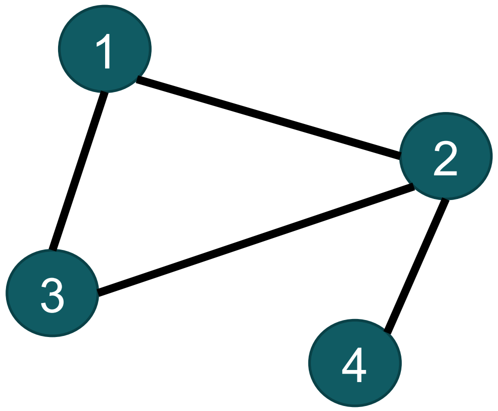
	With constant term	Without constant term
Divide by out-degree	$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A}\mathbf{D}^{-1})^{-1}\mathbf{1}$ PageRank	$\mathbf{x} = \mathbf{A}\mathbf{D}^{-1}\mathbf{x}$ degree centrality
No division	$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A}^{-1})^{-1}\mathbf{1}$ Katz centrality	$\mathbf{x} = \kappa^{-1}\mathbf{A}\mathbf{x}$ eigenvector centrality

Closeness centrality

- Mean distance from a node to the other nodes in the network (using the shortest path between any two nodes)
- Shortest path from *i* to every other node: $\ell_i = \frac{1}{n} \sum_j d_{ij}.$
- Closeness centrality is the inverse: $C_i = \frac{1}{\ell_i} = \frac{n}{\sum_j d_{ij}}$

Closeness centrality

$$C_i = \frac{1}{\ell_i} = \frac{n}{\sum_j d_{ij}}$$



- Advantages:
  - Intuitive
  - Interpretable
- Disadvantages:
  - Can exhibit a small range
  - Requires that the network be strongly connected (d → infinity if not)
    - See "harmonic mean" discussion in Newman pg. 172

Betweenness centrality

- Extent to which a vertex lies on paths between other vertices
- Helps to identify important nodes that control some sort of flow (e.g. messages, people, energy, goods, water)

$$x_i = \sum_{st} \frac{n^i_{st}}{g_{st}}$$

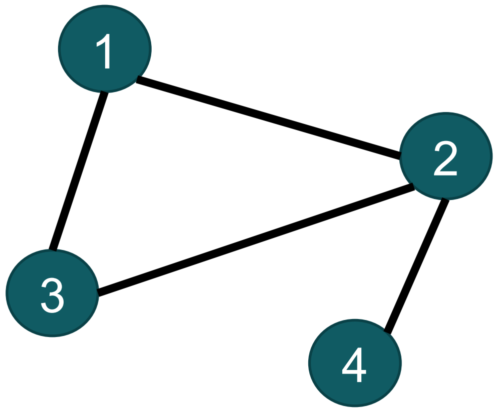
n^i_{st}: number of shortest paths from t → s that pass through *i*

g_{st}: number of shortest paths from t → s

Notes:
- Newman breaks his own convention and defines as s → t
- Newman counts paths from node to itself
- As is often the case in networks, there are variations and the most important thing is to be explicit and consistent

Betweenness centrality – further thoughts

$$x_i = \sum_{st} \frac{n^i_{st}}{g_{st}}$$



- What happens when you remove a node with high betweenness centrality?
- Would you expect betweenness to be correlated with the other centrality measures?

Food for thought…

- Can you think of urban systems examples in which degree, eigvenvector, Katz, PageRank, closeness, and betweenness centralities have useful applications?
- Which node has the highest centrality?