Cornell University

Lecture 7 - Graph Features and Metrics

Graph-Based Data Science For Networked Systems

ECE 5260 – ORIE 5735

Anna Scaglione

Cornell Tech

February 18, 2026

Content

Cornell University

Outline

Cornell University

Node features from A and L_0

Cornell University

- In the last lecture we defined some local (nodal) and group properties:
  - Degree distribution
  - Centrality metrics, whose description we will complete today
- Today we also introduce other useful metrics
  - Similarity metrics and assortative networks
  - Clustering coefficient
  - Reciprocity
- Overall these are helpful graph features to understand roles and special relationships in a network and predict what other neighbors can be to complete a network
- In this lecture we also want to talk about **edge metrics** and the 1-Hodge Laplacian.

Node features from A and L_0

Cornell University

- In the last lecture we defined some local (nodal) and group properties:
  - Degree distribution
  - Centrality metrics, whose description we will complete today
- Today we also introduce other useful metrics
  - Similarity metrics and assortative networks
  - Clustering coefficient
  - Reciprocity
- Overall these are helpful graph features to understand roles and special relationships in a network and predict what other neighbors can be to complete a network
- In this lecture we also want to talk about **edge metrics** and the 1-Hodge Laplacian.

Big picture

Cornell University

For the regular Laplacian $\boldsymbol{L}_0$: eigenvectors $\rightarrow$ node embeddings

- small eigenvalues $\rightarrow$ clusters
- leading eigenvector $\rightarrow$ eigenvector centrality

For the 1-Hodge Laplacian $\boldsymbol{L}_1$ we will see that:

- eigenvectors $\rightarrow$ edge embeddings
- kernel $\rightarrow$ topological circulation modes
- nonzero spectrum $\rightarrow$ inconsistency, curvature

So $\boldsymbol{L}_1$ is not about what **node** is important or what **class** it belongs to, but about how **flows** behave.

Overview of the lecture

In today's lecture:

- we look first at an other set of network structural node features (measures and metrics)

Outline

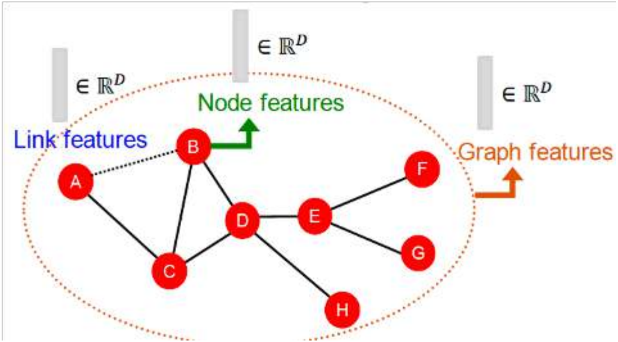
Cornell University

Outline

Network data analysis
Graph analysis

- Part of the data analysis of complex network systems focuses on the network itself, i.e. the (weighted) graph
- The possible inference tasks (classification or prediction) pertain:

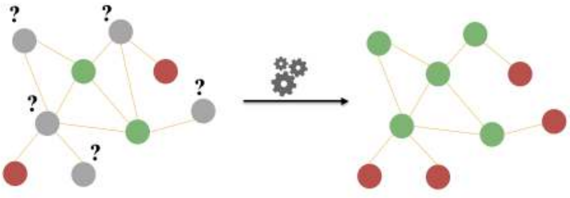
1. individual nodes
2. individual links
3. the entire graph.



Examples of node inference applications

Graph analysis

Cornell University

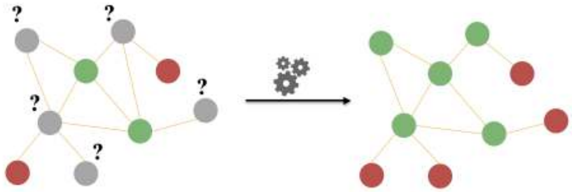


- Social Networks: identifying bots in online social network media. Identifying radicalized terrorists.
- Biological networks: classifying the function of proteins in the interactome
- Knowledge Networks: classifying the topic of a document in a citation network
- Internet: Classification of nodes generated by a Sybil attack (fake nodes in a peer to peer network)

Examples of node inference applications

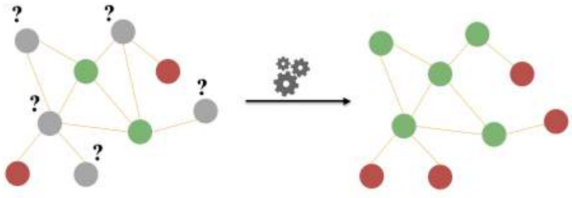
Graph analysis

Cornell University



- Social Networks: identifying bots in online social network media. Identifying radicalized terrorists.
- Biological networks: classifying the function of proteins in the interactome
- Knowledge Networks: classifying the topic of a document in a citation network
- Internet: Classification of nodes generated by a Sybil attack (fake nodes in a peer to peer network)

Examples of node inference applications

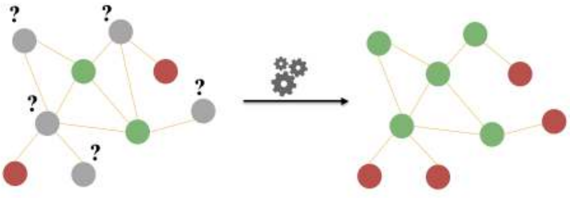


- Social Networks: identifying bots in online social network media. Identifying radicalized terrorists.
- Biological networks: classifying the function of proteins in the interactome
- Knowledge Networks: classifying the topic of a document in a citation network
- Internet: Classification of nodes generated by a Sybil attack (fake nodes in a peer to peer network)

Examples of node inference applications

Cornell University

Graph analysis



- Social Networks: identifying bots in online social network media. Identifying radicalized terrorists.
- Biological networks: classifying the function of proteins in the interactome
- Knowledge Networks: classifying the topic of a document in a citation network
- Internet: Classification of nodes generated by a Sybil attack (fake nodes in a peer to peer network)

The feature extraction step

Feature extraction

Cornell University

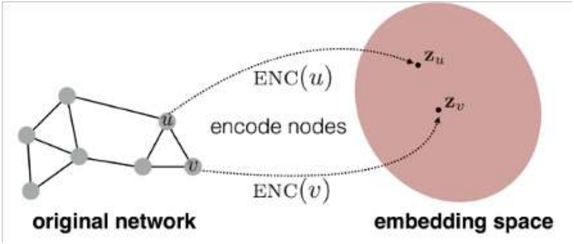
- A preliminary step to perform these tasks is that of extracting features, i.e. finding a function that maps the elements in $\mathcal{V}$, $\mathcal{E}$, the entire graph instance $\mathcal{G}$, possibly also its attributes, onto a point in $\mathbb{R}^d$
- It is desirable that d is small but it has to be such that the embedding is informative.

Node embeddings

Feature extraction

Cornell University

A *shallow embedding* generates a matrix $\boldsymbol{Z} \in \mathbb{R}^{d\times|\mathcal{V}|}$ whose columns are the feature vectors of each node $\boldsymbol{z}_v \forall v \in \mathcal{V}$.



There are several classical nodal features components of $\boldsymbol{z}_u$, that we will discuss

- Node degree
- Different forms Node centrality
- Clustering coefficient
- ...

Outline

Cornell University

2 Graph analysis
- Review
- The ranking of nodes: centrality measures

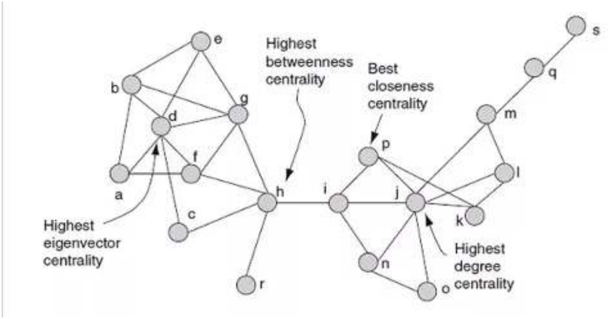
Node centrality measures
graph analysis

Cornell University

Centrality

This measure is used to establish the **most important node in the network**

- They find different nodes, since the form of *importance* is different and has a functional meaning



Outline

Degree centrality
graph analysis

- Degree centrality is one measure. The central node is the one that has maximum degree, i.e.

$$v^{\mathsf{deg}} = \mathop{\mathsf{arg}\max}_{v} d_v.$$

- Is very simple but often very indicative of importance in social networks and of congestion in intrastructures

Degree centrality

graph analysis

- Degree centrality is one measure. The central node is the one that has maximum degree, i.e.

$$v^{\mathsf{deg}} = \mathsf{arg}\max_{v} d_v.$$

- Is very simple but often very indicative of importance in social networks and of congestion in intrastructures

Eigenvector centrality

graph analysis

Cornell University

- Eigenvector Centrality that is defined with a recurrence equation, which captures the fact that an central node must have **central neighbors**:

$$z_u \propto \sum_{v \in \mathcal{N}_u} z_v \quad u \in \mathcal{V} \quad \Rightarrow \quad \lambda \boldsymbol{z} = \boldsymbol{A}\boldsymbol{z}$$

- The eigenvector centrality of a node a node is the corresponding entry of the leading eigenvector, which will assign a positive score to every node

- Let $\boldsymbol{z}_1$ be the leading eigenvector, i.e. $\lambda_1(\boldsymbol{A})\boldsymbol{z}_1 = \boldsymbol{A}\boldsymbol{z}_1$. The most central node is[1]:

$$v^{\text{eig}} = \text{arg}\max_v [\boldsymbol{z}_1]_u$$

- Again, for disconnected networks is more meaningful to have a central node per component.

Eigenvector centrality

graph analysis

Cornell University

- Eigenvector Centrality that is defined with a recurrence equation, which captures the fact that an central node must have **central neighbors**:

$$z_u \propto \sum_{v \in \mathcal{N}_u} z_v \quad u \in \mathcal{V} \quad \Rightarrow \quad \lambda \boldsymbol{z} = \boldsymbol{A}\boldsymbol{z}$$

- The eigenvector centrality of a node a node is the corresponding entry of the leading eigenvector, which will assign a positive score to every node

- Let $\boldsymbol{z}_1$ be the leading eigenvector, i.e. $\lambda_1(\boldsymbol{A})\boldsymbol{z}_1 = \boldsymbol{A}\boldsymbol{z}_1$. The most central node is[1]:

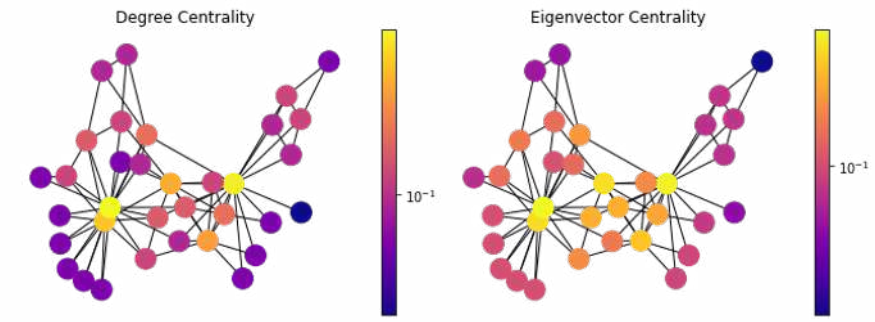
$$v^{\text{eig}} = \text{arg}\max_v [\boldsymbol{z}_1]_u$$

- Again, for disconnected networks is more meaningful to have a central node per component.

Eigenvector centrality example

graph analysis

- The example of the Zachary's Karate Club for both measures illustrates similarties and diferences between the two metrics

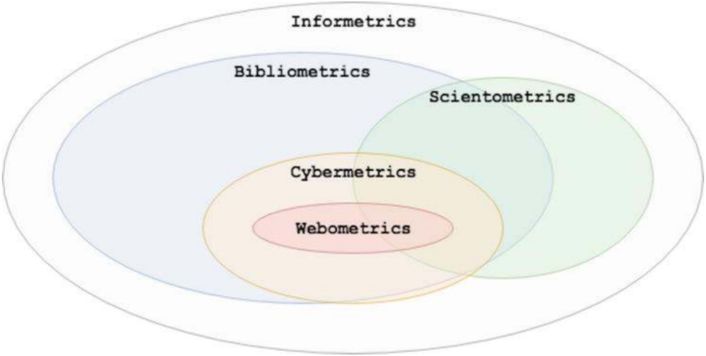


Outline

Cornell University

2 Graph analysis
- Review
- The ranking of nodes: centrality measures
  - Undirected networks
  - Directed networks

Eigenvector centrality for directed networks



- The notion of centrality for directed networks is very important for knowledge networks, in **webometrics**, citations networks etc.

Eigenvector centrality for directed networks

Cornell University

- We applied a similar logic used for the centrality measures introduced but $\boldsymbol{A}$ is not symmetric.
- Between the left and right eigenvectors we said the right eigenvectors are the right choice: *centrality is bestowed by the nodes that point towards you*
- This idea can also be applied to degree centrality (**in degree centrality**)
- We can also define $\lambda_1(\boldsymbol{A})\boldsymbol{z}_1 = \boldsymbol{A}\boldsymbol{z}_1$ to define centrality and $v^{\text{eig}} = \text{argmax}_v[\boldsymbol{z}_1]_u$ is the most central node.

Eigenvector centrality for directed networks

- We applied a similar logic used for the centrality measures introduced but $\boldsymbol{A}$ is not symmetric.
- Between the left and right eigenvectors we said the right eigenvectors are the right choice: *centrality is bestowed by the nodes that point towards you*
- This idea can also be applied to degree centrality (**in degree centrality**)
- We can also define $\lambda_1(\boldsymbol{A})\boldsymbol{z}_1 = \boldsymbol{A}\boldsymbol{z}_1$ to define centrality and $v^{\text{eig}} = \text{arg}\max_v[\boldsymbol{z}_1]_u$ is the most central node.

Eigenvector centrality for directed networks

- Unfortunately this definition can be problematic:
  - if a node has zero in degree has zero centrality score, which may be ok;
  - the problem is that a node that has a large in-degree but all its in-neighbors have zero in-degree also have score zero.
- For instance, a directed acyclic graph, such as an article citation network, hence all nodes get null eigenvector centrality
- This motivates the notion of Katz centrality.

Eigenvector centrality for directed networks

- Unfortunately this definition can be problematic:
  - if a node has zero in degree has zero centrality score, which may be ok;
  - the problem is that a node that has a large in-degree but all its in-neighbors have zero in-degree also have score zero.
- For instance, a directed acyclic graph, such as an article citation network, hence all nodes get null eigenvector centrality
- This motivates the notion of Katz centrality.

Katz centrality
graph analysis

- One solution is to add some *centrality* β to all nodes, equally:

$$z_v = \alpha \sum_{u \in \mathcal{N}_v} z_u + \beta, \quad \Rightarrow \quad \boldsymbol{z} = \alpha \boldsymbol{A} \boldsymbol{z} + \beta \mathbf{1}$$

  where now α is a parameter.

- The score of centrality, named by Katz the author who defined it, is now:

$$\boldsymbol{z}^{\mathsf{Katz}} = \beta(\boldsymbol{I} - \alpha \boldsymbol{A})^{-1}\mathbf{1}.$$

  but we need to choose α and β

- β is sort of a "*self-esteem*" constant, but scales equally everything, does not help the ranking

- A variation of Katz centrality chooses a vector β with different componets based on some side information on the importance of that node

Katz centrality

graph analysis

- One solution is to add some *centrality* β to all nodes, equally:

$$z_v = \alpha \sum_{u \in \mathcal{N}_v} z_u + \beta, \quad \Rightarrow \quad \boldsymbol{z} = \alpha \boldsymbol{A} \boldsymbol{z} + \beta \mathbf{1}$$

  where now α is a parameter.

- The score of centrality, named by Katz the author who defined it, is now:

$$\boldsymbol{z}^{\text{Katz}} = \beta(\boldsymbol{I} - \alpha \boldsymbol{A})^{-1} \mathbf{1}.$$

  but we need to choose α and β

- β is sort of a "*self-esteem*" constant, but scales equally everything, does not help the ranking

- A variation of Katz centrality chooses a vector $\boldsymbol{\beta}$ with different componets based on some side information on the importance of that node

Katz centrality - choice of α

graph analysis

Cornell University

- At $\alpha = 0$ the measure is meaningless $\boldsymbol{z} = \beta\mathbf{1}$
- The values of α that make the matrix not invertible, i.e. such that $\det(\boldsymbol{I} - \alpha\boldsymbol{A}) = 0$, are the eigenvalues of $\boldsymbol{A}$
- The one solution that would weight $\boldsymbol{A}$ the most, while $\det(\boldsymbol{I} - \alpha_1\boldsymbol{A}) \approx 0$.
- $\det(\boldsymbol{I} - \alpha_1\boldsymbol{A})$ is the characteristic equation, again, $\lambda_1(\boldsymbol{A})$
- Values $\alpha \approx \lambda_1(\boldsymbol{A})$ are the popular choice:
  - For undirected or non-acyclic graphs this choice gives results close to the eigenvector centrality
  - For directed networks it does give a different result

Katz centrality - choice of α

graph analysis

Cornell University

- At $\alpha = 0$ the measure is meaningless $\boldsymbol{z} = \beta \mathbf{1}$
- The values of α that make the matrix not invertible, i.e. such that $\det(\boldsymbol{I} - \alpha \boldsymbol{A}) = 0$, are the eigenvalues of $\boldsymbol{A}$
- The one solution that would weight $\boldsymbol{A}$ the most, while $\det(\boldsymbol{I} - \alpha_1 \boldsymbol{A}) \approx 0$.
- $\det(\boldsymbol{I} - \alpha_1 \boldsymbol{A})$ is the characteristic equation, again, $\lambda_1(\boldsymbol{A})$
- Values $\alpha \approx \lambda_1(\boldsymbol{A})$ are the popular choice:
  - For undirected or non-acyclic graphs this choice gives results close to the eigenvector centrality
  - For directed networks it does give a different result

Katz centrality - choice of α

graph analysis

Cornell University

- At $\alpha = 0$ the measure is meaningless $\boldsymbol{z} = \beta \mathbf{1}$
- The values of α that make the matrix not invertible, i.e. such that $\det(\boldsymbol{I} - \alpha \boldsymbol{A}) = 0$, are the eigenvalues of $\boldsymbol{A}$
- The one solution that would weight $\boldsymbol{A}$ the most, while $\det(\boldsymbol{I} - \alpha_1 \boldsymbol{A}) \approx 0$.
- $\det(\boldsymbol{I} - \alpha_1 \boldsymbol{A})$ is the characteristic equation, again, $\lambda_1(\boldsymbol{A})$
- Values $\alpha \approx \lambda_1(\boldsymbol{A})$ are the popular choice:
  - For undirected or non-acyclic graphs this choice gives results close to the eigenvector centrality
  - For directed networks it does give a different result

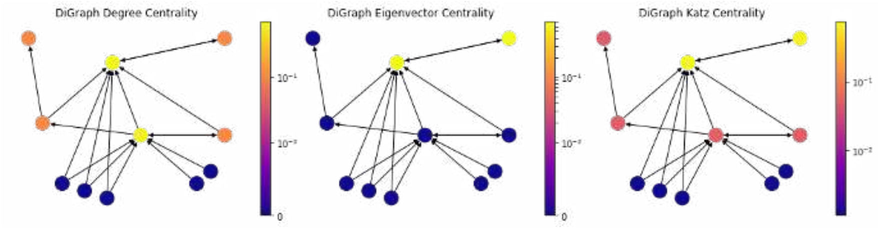
Katz centrality - choice of α

graph analysis

- At $\alpha = 0$ the measure is meaningless $\boldsymbol{z} = \beta \mathbf{1}$
- The values of α that make the matrix not invertible, i.e. such that $\det(\boldsymbol{I} - \alpha \boldsymbol{A}) = 0$, are the eigenvalues of $\boldsymbol{A}$
- The one solution that would weight $\boldsymbol{A}$ the most, while $\det(\boldsymbol{I} - \alpha_1 \boldsymbol{A}) \approx 0$.
- $\det(\boldsymbol{I} - \alpha_1 \boldsymbol{A})$ is the characteristic equation, again, $\lambda_1(\boldsymbol{A})$
- Values $\alpha \approx \lambda_1(\boldsymbol{A})$ are the popular choice:
  - For undirected or non-acyclic graphs this choice gives results close to the eigenvector centrality
  - For directed networks it does give a different result

Comparison with Katz centrality
graph analysis

- The plot below compares the different forms of centrality



- The main drawback of the Katz centrality metric is that a node with high centrality with edges pointing to many others spreads high centrality to its out neighbors.
- **PageRank** centrality fixes the problem.

PageRank centrality

graph analysis

- PageRank centrality normalizes the adjacency matrix by the out degree. Remember that the the outdegree sequence is $\boldsymbol{d}^- = \boldsymbol{A}^\top \mathbf{1}$, because $d_v^- = |\mathcal{N}_v^-| = \sum_{u \in \mathcal{V}} a_{uv}$.

- To fix the case where $d_v^- = 0$, we define $[\boldsymbol{d}^{out}]_v = \max(d_v^-, 1)$

- The normalization $a_{u,v}/d^{+out}$ is obtained by scaling every column of the matrix $\boldsymbol{A}$

$$\boldsymbol{z} = \alpha \boldsymbol{A} \text{diag}^{-1}(\boldsymbol{d}^{out})\boldsymbol{z} + \beta \boldsymbol{I},$$

$$\Rightarrow \quad \boldsymbol{z}^{\text{PageRank}} = \beta(\boldsymbol{I} - \alpha \boldsymbol{A} \text{diag}^{-1}(\boldsymbol{d}^{out}))^{-1}\mathbf{1}$$

- Just like with the Katz centrality, in relative terms β does not matter. We may choose $\beta = 1$.

PageRank centrality
graph analysis

- PageRank centrality normalizes the adjacency matrix by the out degree. Remember that the the outdegree sequence is $\boldsymbol{d}^- = \boldsymbol{A}^\top \mathbf{1}$, because $d_v^- = |\mathcal{N}_v^-| = \sum_{u\in\mathcal{V}} a_{uv}$.
- To fix the case where $d_v^- = 0$, we define $[\boldsymbol{d}^{out}]_v = \max(d_v^-, 1)$
- The normalization $a_{u,v}/d^{+out}$ is obtained by scaling every column of the matrix $\boldsymbol{A}$

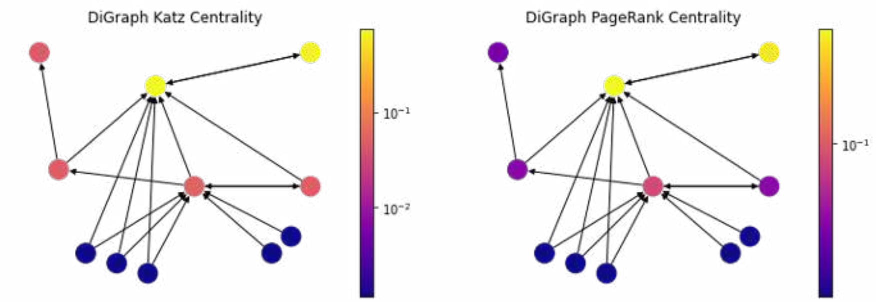
$$\boldsymbol{z} = \alpha \boldsymbol{A}\text{diag}^{-1}(\boldsymbol{d}^{out})\boldsymbol{z} + \beta \boldsymbol{I},$$

$$\Rightarrow \quad \boldsymbol{z}^{\text{PageRank}} = \beta(\boldsymbol{I} - \alpha \boldsymbol{A}\text{diag}^{-1}(\boldsymbol{d}^{out}))^{-1}\mathbf{1}$$

- Just like with the Katz centrality, in relative terms β does not matter. We may choose $\beta = 1$.

PageRank vs Katz centrality
graph analysis

A comparison for the two:



Observations
graph analysis

- The forms of centrality we considered are highlighting well connected nodes, at the receiving end of their neighborhood, or authorities
- Another important measure is what nodes are important to reach them. These nodes are hubs
  - For instance, web pages that are mostly an index that links to other ones may be an important source of information
- The forms of centrality we define next focus on highlighting this notion of centrality that identifies hubs

Observations
graph analysis

- The forms of centrality we considered are highlighting well connected nodes, at the receiving end of their neighborhood, or authorities
- Another important measure is what nodes are important to reach them. These nodes are hubs
  - For instance, web pages that are mostly an index that links to other ones may be an important source of information
- The forms of centrality we define next focus on highlighting this notion of centrality that identifies hubs

Authority and Hub centrality

graph analysis

Cornell University

- One idea is to find a way to simultaneously define hubs an authoritites.

- Let x and y be the authorities and hub centralitites
  - Authority centrality: The authority score is proportional to the hub centralities it connects to
  - Hub centrality: The hub score is proportional to the authority centralities it connects to

- Mathematically:

$$\boldsymbol{x} = \alpha \boldsymbol{A}\boldsymbol{y}, \quad \boldsymbol{y} = \beta \boldsymbol{A}^\top \boldsymbol{x}$$

- Combining the two equations we obtain the eigenvectors equations:

$$\lambda \boldsymbol{x} = \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{x}, \quad \lambda \boldsymbol{y} = \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{y}.$$

Authority and Hub centrality

graph analysis

- One idea is to find a way to simultaneously define hubs an authoritites.
- Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be the authorities and hub centralitites
  - Authority centrality: The authority score is proportional to the hub centralities it connects to
  - Hub centrality: The hub score is proportional to the authority centralities it connects to
- Mathematically:

$$\boldsymbol{x} = \alpha \boldsymbol{A}\boldsymbol{y}, \quad \boldsymbol{y} = \beta \boldsymbol{A}^\top \boldsymbol{x}$$

- Combining the two equations we obtain the eigenvectors equations:

$$\lambda \boldsymbol{x} = \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{x}, \qquad \lambda \boldsymbol{y} = \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{y}.$$

Authority and Hub centrality

graph analysis

- One idea is to find a way to simultaneously define hubs an authoritites.
- Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be the authorities and hub centralitites
  - Authority centrality: The authority score is proportional to the hub centralities it connects to
  - Hub centrality: The hub score is proportional to the authority centralities it connects to
- Mathematically:

$$\boldsymbol{x} = \alpha \boldsymbol{A}\boldsymbol{y}, \quad \boldsymbol{y} = \beta \boldsymbol{A}^\top \boldsymbol{x}$$

- Combining the two equations we obtain the eigenvectors equations:

$$\lambda \boldsymbol{x} = \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{x}, \quad \lambda \boldsymbol{y} = \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{y}.$$

Authority and Hub centrality

graph analysis

Cornell University

- **Fact**: The eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\top$ and $\boldsymbol{A}^\top\boldsymbol{A}$ are the same.
- The eigenvectors are, respectively, the left and right the **singular vectors**, and the equations that yield the **singular value decomposition** of a matrix

$$\boldsymbol{A} = \boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top \quad \text{where} \quad \sigma_n = \sqrt{\lambda_n}$$

- This pair of centrality measures (x_v, y_v) defined by the leading singular vectors for each node $v \in \mathcal{V}$ was defined for hyperlink networks, and its calculation called *hyperlink-induced topic search (HITS)*
- HITS tackles the problem with zero in-degree nodes of Eigenvector Centrality. These zero in-degree nodes become central hubs and contribute to other nodes.

Authority and Hub centrality

graph analysis

Cornell University

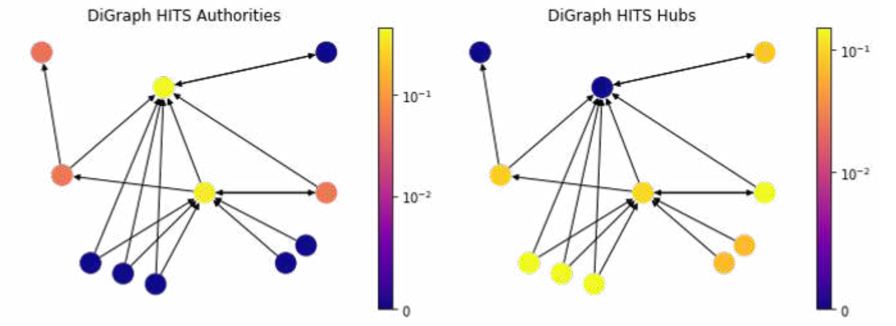
- **Fact**: The eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\top$ and $\boldsymbol{A}^\top\boldsymbol{A}$ are the same.
- The eigenvectors are, respectively, the left and right the **singular vectors**, and the equations that yield the **singular value decomposition** of a matrix

$$\boldsymbol{A} = \boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top \quad \text{where} \quad \sigma_n = \sqrt{\lambda_n}$$

- This pair of centrality measures (x_v, y_v) defined by the leading singular vectors for each node $v \in \mathcal{V}$ was defined for hyperlink networks, and its calculation called *hyperlink-induced topic search (HITS)*
- HITS tackles the problem with zero in-degree nodes of Eigenvector Centrality. These zero in-degree nodes become central hubs and contribute to other nodes.

HITS centrality
graph analysis

HITS centrality example



Outline

Cornell University

Centrality based on distances

graph analysis

- Next we consider two forms of centrality that consider shortest paths in networks
  1. Closeness centrality
  2. Betweeness centrality

Closeness centrality
graph analysis

- Let $dist(v, u)$ be the shortest hop distance between node v and node u
- Closeness centrality measures depends on the average shortest hop distance from node v to every other node in the network

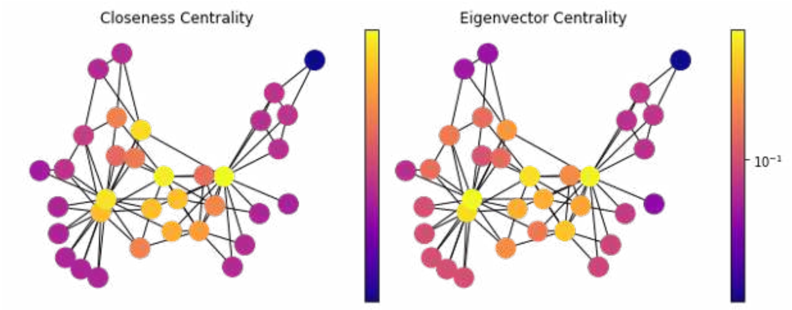
$$\text{Average shortest path length :} \quad \ell_v = \frac{1}{N} \sum_{u \in \mathcal{V}} dist(v, u)$$

which is high for nodes far from everyone.

- Closeness centrality: Is the reciprocal of the average shortest path:

$$z_v^{\text{close}} = \frac{1}{\ell_v}$$

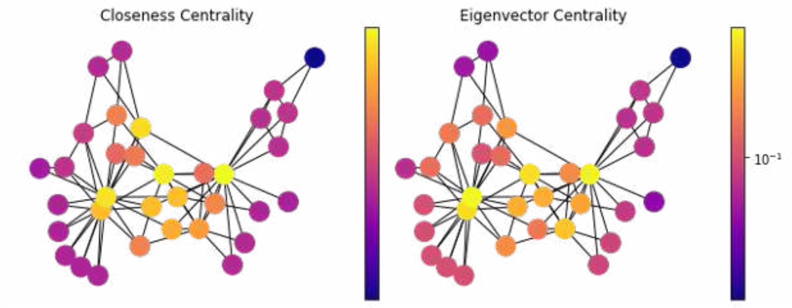
Closeness centrality: example
graph analysis



- Suppose an actor (Christopher Lee) was in many not important movies, where few other stars were cast.
- If you construct a graph that connects actors on the same movie and Christopher Lee has small degree and eigenvector centrality
- Yet Christopher Lee has high closeness centrality

Closeness centrality: example

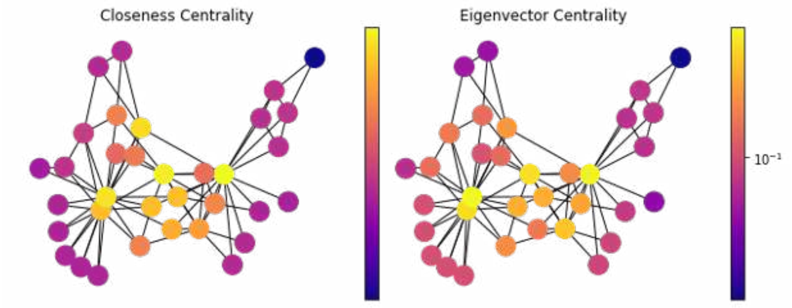
graph analysis



- Suppose an actor (Christopher Lee) was in many not important movies, where few other stars were cast.
- If you construct a graph that connects actors on the same movie and Christopher Lee has small degree and eigenvector centrality
- Yet Christopher Lee has high closeness centrality

Closeness centrality: example
graph analysis

Cornell University



- Suppose an actor (Christopher Lee) was in many not important movies, where few other stars were cast.
- If you construct a graph that connects actors on the same movie and Christopher Lee has small degree and eigenvector centrality
- Yet Christopher Lee has high closeness centrality

Closeness centrality: disconnected networks

Cornell University

- If nodes have more than one component, the distance among nodes in different components is infinite
- There are two logical options
  1. Measure only distances in a certain component.
     - **Problem**: If the components have very different sizes, nodes in small components will have greater closeness centrality
  2. Revise the definition, as follows:

$$z_v^{\text{close-2}} = \frac{1}{N-1} \sum_{u \neq v} \frac{1}{dist(v,u)}$$

     - This definition takes care of it because if v and u are in different components, node u has no contribution to the centrality metric:

$$dist(v,u) = \infty \quad \Rightarrow \quad \frac{1}{dist(v,u)} = 0$$

Closeness centrality: disconnected networks

Cornell University

- If nodes have more than one component, the distance among nodes in different components is infinite
- There are two logical options
  1. Measure only distances in a certain component.
     - **Problem**: If the components have very different sizes, nodes in small components will have greater closeness centrality
  2. Revise the definition, as follows:

$$z_v^{\text{close-2}} = \frac{1}{N-1} \sum_{u \neq v} \frac{1}{dist(v,u)}$$

     - This definition takes care of it because if v and u are in different components, node u has no contribution to the centrality metric:

$$dist(v,u) = \infty \quad \Rightarrow \quad \frac{1}{dist(v,u)} = 0$$

Betweeness centrality

graph analysis

Cornell University

- Measures how frequently a node is in a path between other nodes
- Nodes with high betweeness centrality are crucial in supporting flows, particularly in clustered networks.
- **Betweeness centrality** is defined as the number of times node v appears in the shortest path connecting the network pairs

$$z_v^{\text{betw}} = \sum_{(s,t)\in\mathcal{V}\times\mathcal{V}} n_{s,t}^v$$

$$\text{where} \quad n_{s,t}^v = \begin{cases} 1 & v \text{ is in the path between } (s,t) \\ 0 & \text{else} \end{cases}$$

- This metric works for both directed and undirected graphs

Betweeness centrality

graph analysis

- Measures how frequently a node is in a path between other nodes
- Nodes with high betweeness centrality are crucial in supporting flows, particularly in clustered networks.
- **Betweeness centrality** is defined as the number of times node v appears in the shortest path connecting the network pairs

$$z_v^{\mathsf{betw}} = \sum_{(s,t)\in\mathcal{V}\times\mathcal{V}} n_{s,t}^v$$

$$\text{where} \quad n_{s,t}^v = \begin{cases} 1 & v \text{ is in the path between } (s,t) \\ 0 & \text{else} \end{cases}$$

- This metric works for both directed and undirected graphs

Outline

Cornell University

Homophily

Cornell University

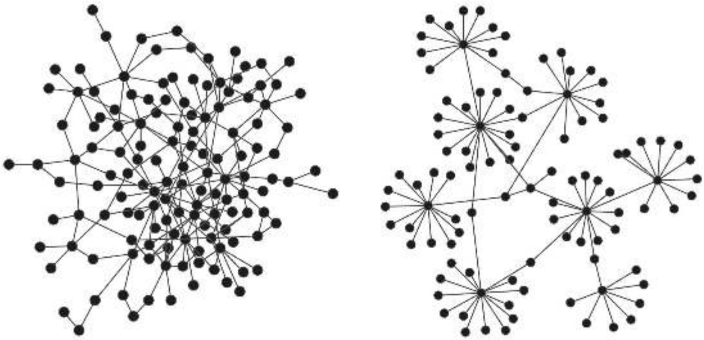
- In addition to ranking the node there are other metrics that are useful to understand trends in a network
- Homophily is a metric that tries to assess how **homogeneous** are the neighborhoods of neighbors with respect to a certain feature
- Next we will focus on the nodes' degrees as their respective features, but other features can be chosen instead

Homophily

- In addition to ranking the node there are other metrics that are useful to understand trends in a network
- Homophily is a metric that tries to assess how **homogeneous** are the neighborhoods of neighbors with respect to a certain feature
- Next we will focus on the nodes' degrees as their respective features, but other features can be chosen instead

Homophily: assortative and disassortative networks

graph analysis



(A) Assortative (B) Disassortative

In an assortative network any node u has neighboring nodes similar with respect to the feature $\phi(u) \approx \phi(v), v \in \mathcal{N}_u$. In the figure the feature $\phi(u) = d_u$. A disassortative network showcases the opposite trend.

Measure of assortativity

- For degree assortativity $\phi(u) = d_u$, which is showcased by the networks in the previous slide, the Pearson correlation coefficient $r \in [-1, 1]$ of **degrees of neighboring nodes** is one of the measures that can be used:

$$r = \frac{\sum_{(u,v)\in\mathcal{E}}(\phi(u) - \overline{\phi}_{\text{in}})(\phi(v) - \overline{\phi}_{\text{out}})}{\sqrt{\sum_{(u,v)\in\mathcal{E}}(\phi(u) - \overline{\phi}_{\text{in}})^2}\sqrt{\sum_{(u,v)\in\mathcal{E}}(\phi(v) - \overline{\phi}_{\text{out}})^2}}$$

$$\overline{\phi}_{\text{in}} = \frac{\sum_{(u,v)\in\mathcal{E}} \phi(u)}{|\mathcal{E}|}, \qquad \overline{\phi}_{\text{out}} = \frac{\sum_{(u,v)\in\mathcal{E}} \phi(v)}{|\mathcal{E}|}$$

- $r > 0$ the network is assortative if $r < 0$ the opposite is true.

Measure of assortativity

- For degree assortativity $\phi(u) = d_u$, which is showcased by the networks in the previous slide, the Pearson correlation coefficient $r \in [-1, 1]$ of **degrees of neighboring nodes** is one of the measures that can be used:

$$r = \frac{\sum_{(u,v)\in\mathcal{E}}(\phi(u) - \overline{\phi}_{\text{in}})(\phi(v) - \overline{\phi}_{\text{out}})}{\sqrt{\sum_{(u,v)\in\mathcal{E}}(\phi(u) - \overline{\phi}_{\text{in}})^2}\sqrt{\sum_{(u,v)\in\mathcal{E}}(\phi(v) - \overline{\phi}_{\text{out}})^2}}$$

$$\overline{\phi}_{\text{in}} = \frac{\sum_{(u,v)\in\mathcal{E}} \phi(u)}{|\mathcal{E}|}, \qquad \overline{\phi}_{\text{out}} = \frac{\sum_{(u,v)\in\mathcal{E}} \phi(v)}{|\mathcal{E}|}$$

- $r > 0$ the network is assortative if $r < 0$ the opposite is true.

Nodes similarity via Cosine similarity

Cornell University

- Cosine similarity: Another measure of similarity applied to undirected networks of neighborhood looks at how **similar** are the neighbors
- Lets us calculate the number of common neighbors:

$$|\mathcal{N}_u \cap \mathcal{N}_v| = n_{u,v} = \sum_{k \in \mathcal{V}} a_{uk} a_{kv} = [\boldsymbol{A}^2]_{u,v}$$

- This is because $\boldsymbol{A} = \boldsymbol{A}^\top$, $\boldsymbol{A}^2 = \boldsymbol{A}^\top \boldsymbol{A}$ and, therefore, $n_{u,v} = [\boldsymbol{A}^2]_{u,v} = [\boldsymbol{A}^\top \boldsymbol{A}]_{u,v}$
- In words $n_{u,v}$ is the projection of the columns of $\boldsymbol{A}$ corresponding to node u and that to node v respectively

Nodes similarity via Cosine similarity

- Cosine similarity: Another measure of similarity applied to undirected networks of neighborhood looks at how **similar** are the neighbors
- Lets us calculate the number of common neighbors:

$$|\mathcal{N}_u \cap \mathcal{N}_v| = n_{u,v} = \sum_{k \in \mathcal{V}} a_{uk} a_{kv} = [\boldsymbol{A}^2]_{u,v}$$

- This is because $\boldsymbol{A} = \boldsymbol{A}^\top$, $\boldsymbol{A}^2 = \boldsymbol{A}^\top \boldsymbol{A}$ and, therefore, $n_{u,v} = [\boldsymbol{A}^2]_{u,v} = [\boldsymbol{A}^\top \boldsymbol{A}]_{u,v}$
- In words $n_{u,v}$ is the projection of the columns of $\boldsymbol{A}$ corresponding to node u and that to node v respectively

Nodes similarity via Cosine similarity

- Cosine similarity: Another measure of similarity applied to undirected networks of neighborhood looks at how **similar** are the neighbors
- Lets us calculate the number of common neighbors:

$$|\mathcal{N}_u \cap \mathcal{N}_v| = n_{u,v} = \sum_{k \in \mathcal{V}} a_{uk} a_{kv} = [\boldsymbol{A}^2]_{u,v}$$

- This is because $\boldsymbol{A} = \boldsymbol{A}^\top$, $\boldsymbol{A}^2 = \boldsymbol{A}^\top \boldsymbol{A}$ and, therefore, $n_{u,v} = [\boldsymbol{A}^2]_{u,v} = [\boldsymbol{A}^\top \boldsymbol{A}]_{u,v}$
- In words $n_{u,v}$ is the projection of the columns of $\boldsymbol{A}$ corresponding to node u and that to node v respectively

Cosine similarity

Cornell University

- For two vectors:

$$\cos(\theta_{xy}) = \frac{\boldsymbol{x}^\top \boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}$$

- This suggests to use a normalized version of $n_{u,v}$:

$$\sigma^{\cos}_{u,v} = \frac{[\boldsymbol{A}^2]_{uv}}{\sqrt{[\boldsymbol{A}^2]_{uu}}\sqrt{[\boldsymbol{A}^2]_{vv}}}$$

Cosine similarity

- For two vectors:

$$\cos(\theta_{xy}) = \frac{\boldsymbol{x}^\top \boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}$$

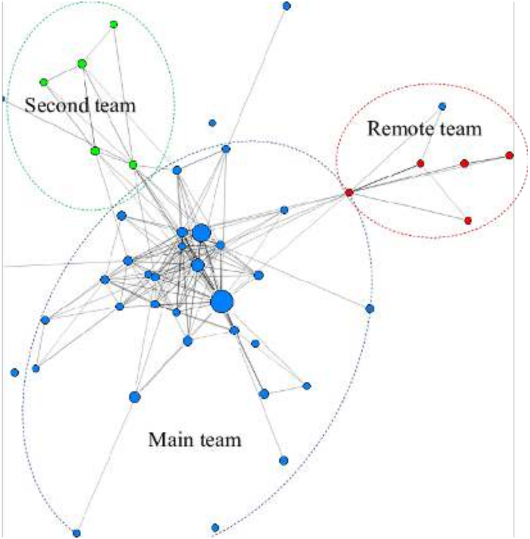
- This suggests to use a normalized version of $n_{u,v}$:

$$\sigma_{u,v}^{\cos} = \frac{[\boldsymbol{A}^2]_{uv}}{\sqrt{[\boldsymbol{A}^2]_{uu}}\sqrt{[\boldsymbol{A}^2]_{vv}}}$$

Cosine similarity example

Cornell University

- The notion of similarity often helps identify nodes, or group of nodes, with different roles or types of interactions
- The example on the left is showcasing how that can work in a social network



Jaccard coefficient

Cornell University

- Another measure of similarity is the Jaccard coefficient

$$\sigma_{uv}^{\text{Jaccard}} = \frac{n_{u,v}}{d_u + d_v - n_{u,v}}$$

- $n_{u,v}$ is like before the number of common neighbors
- $d_u + d_v$ is the number of neighbords of u, plus those of v, i.e. $d_u + d_v = |\mathcal{N}_u \cup \mathcal{N}_v|$
- Therefore the Jaccard coefficient measures how much the two neighborhoods overlap:
  - the higher the overlap $\rightarrow$ the greater is $\sigma_{uv}^{\text{Jaccard}}$,
  - if there u and v have no common neighbors then $\sigma_{uv}^{\text{Jaccard}} = 0$

Jaccard coefficient

Cornell University

- Another measure of similarity is the Jaccard coefficient

$$\sigma_{uv}^{\text{Jaccard}} = \frac{n_{u,v}}{d_u + d_v - n_{u,v}}$$

- $n_{u,v}$ is like before the number of common neighbors
- $d_u + d_v$ is the number of neighbords of u, plus those of v, i.e. $d_u + d_v = |\mathcal{N}_u \cup \mathcal{N}_v|$
- Therefore the Jaccard coefficient measures how much the two neighborhoods overlap:
  - the higher the overlap $\rightarrow$ the greater is $\sigma_{uv}^{\text{Jaccard}}$,
  - if there u and v have no common neighbors then $\sigma_{uv}^{\text{Jaccard}} = 0$

Jaccard coefficient

- Another measure of similarity is the Jaccard coefficient

$$\sigma_{uv}^{\text{Jaccard}} = \frac{n_{u,v}}{d_u + d_v - n_{u,v}}$$

- $n_{u,v}$ is like before the number of common neighbors
- $d_u + d_v$ is the number of neighbords of u, plus those of v, i.e. $d_u + d_v = |\mathcal{N}_u \cup \mathcal{N}_v|$
- Therefore the Jaccard coefficient measures how much the two neighborhoods overlap:
  - the higher the overlap $\rightarrow$ the greater is $\sigma_{uv}^{\text{Jaccard}}$,
  - if there u and v have no common neighbors then $\sigma_{uv}^{\text{Jaccard}} = 0$

Outline

Transitivity

Cornell University

- Transitivity applies to a mathematical relation among elements of a set
  - Examples of transitivity appear in algebra. Equalities and inequalities for example $a = b$ and $b = c$ implies $a = b$, etc.
- For nodes in a graph, this can be applied to predicates regarding pairs of nodes such as one based on connectivity:

  If there is an edge between node u and node v and one between node v and node w, then there is an edge between node u and w.
- In words, the nodes u, v and w form a cycle of length 3.

Transitivity

- Transitivity applies to a mathematical relation among elements of a set
  - Examples of transitivity appear in algebra. Equalities and inequalities for example $a = b$ and $b = c$ implies $a = b$, etc.
- For nodes in a graph, this can be applied to predicates regarding pairs of nodes such as one based on connectivity:

  If there is an edge between node u and node v and one between node v and node w, then there is an edge between node u and w.

- In words, the nodes u, v and w form a cycle of length 3.

Transitivity

Cornell University

- Transitivity applies to a mathematical relation among elements of a set
  - Examples of transitivity appear in algebra. Equalities and inequalities for example $a = b$ and $b = c$ implies $a = b$, etc.
- For nodes in a graph, this can be applied to predicates regarding pairs of nodes such as one based on connectivity:

  If there is an edge between node u and node v and one between node v and node w, then there is an edge between node u and w.
- In words, the nodes u, v and w form a cycle of length 3.

Clustering coefficient: meaning

Cornell University

- Having perfect transitivity is meaningless because that means that every component is a clique.
- It is interesting to measure on average how many nodes form these triplets
- In social networks this measures how more likely it is that a *friend of a friend is a friend.*
- Empirical observations show this is often the case. For instance, actors collaborators networks have clustering coefficients in the order of 0.2.
  - This is high for large networks: if friends where chosen at random, the clustering coefficient should be $O(1/N)$

Clustering coefficient: meaning

Cornell University

- Having perfect transitivity is meaningless because that means that every component is a clique.
- It is interesting to measure on average how many nodes form these triplets
- In social networks this measures how more likely it is that a *friend of a friend is a friend*.
- Empirical observations show this is often the case. For instance, actors collaborators networks have clustering coefficients in the order of 0.2.
  - This is high for large networks: if friends where chosen at random, the clustering coefficient should be $O(1/N)$

Clustering coefficient: meaning

- Having perfect transitivity is meaningless because that means that every component is a clique.
- It is interesting to measure on average how many nodes form these triplets
- In social networks this measures how more likely it is that a *friend of a friend is a friend*.
- Empirical observations show this is often the case. For instance, actors collaborators networks have clustering coefficients in the order of 0.2.
  - This is high for large networks: if friends where chosen at random, the clustering coefficient should be $O(1/N)$

Clustering coefficient: meaning

- Having perfect transitivity is meaningless because that means that every component is a clique.
- It is interesting to measure on average how many nodes form these triplets
- In social networks this measures how more likely it is that a *friend of a friend is a friend*.
- Empirical observations show this is often the case. For instance, actors collaborators networks have clustering coefficients in the order of 0.2.
  - This is high for large networks: if friends where chosen at random, the clustering coefficient should be $O(1/N)$

Clustering coefficient: definition

- The clustering coefficient is defined as follows:

$$C = \frac{\text{(number of closed paths of length two)}}{\text{(number of paths of length two)}}$$

- Equivalently, considering that each triangle contributes 6 paths of length 2, it is:

$$C = \frac{\text{(number of triangles)} \times 6}{\text{(number of paths of length two)}}$$

Clustering coefficient: definition

- The clustering coefficient is defined as follows:

$$C = \frac{\text{(number of closed paths of length two)}}{\text{(number of paths of length two)}}$$

- Equivalently, considering that each triangle contributes 6 paths of length 2, it is:

$$C = \frac{\text{(number of triangles)} \times 6}{\text{(number of paths of length two)}}$$

Local clustering

Cornell University

- We can associate a nodal metric that characterized how clustered is the neighborhood of a node

$$C_v = \frac{(\text{number neighbors in } \mathcal{N}_v \text{ that are connected})}{(\text{number of pairs of neighbors})}$$

where the number of pairs of neighbors $\binom{d_v}{2}$

- The value of C_v is often found to correlate with d_v
- Watts and Strogatz proposed to use an average clustering coefficient as a metric for the type of network:

$$C^{\text{WS}} = \frac{1}{N} \sum_{v \in \mathcal{V}} C_v$$

Local clustering

Cornell University

- We can associate a nodal metric that characterized how clustered is the neighborhood of a node

$$C_v = \frac{(\text{number neighbors in } \mathcal{N}_v \text{ that are connected})}{(\text{number of pairs of neighbors})}$$

where the number of pairs of neighbors $\binom{d_v}{2}$

- The value of C_v is often found to correlate with d_v
- Watts and Strogatz proposed to use an average clustering coefficient as a metric for the type of network:

$$C^{\text{WS}} = \frac{1}{N} \sum_{v \in \mathcal{V}} C_v$$

Local clustering

- We can associate a nodal metric that characterized how clustered is the neighborhood of a node

$$C_v = \frac{\text{(number neighbors in } \mathcal{N}_v \text{ that are connected)}}{\text{(number of pairs of neighbors)}}$$

where the number of pairs of neighbors $\binom{d_v}{2}$

- The value of C_v is often found to correlate with d_v
- Watts and Strogatz proposed to use an average clustering coefficient as a metric for the type of network:

$$C^{\mathsf{WS}} = \frac{1}{N} \sum_{v \in \mathcal{V}} C_v$$

Reciprocity - directed networks

- The clustering coefficient measures the frequency of loops of length three in a network. Why only concentrating on loops of length three?
- For example, in directed network it may be more interesting to look at loops of length two which signal a two way relationship. The study of the Hodge Laplacian is ideal for that.
- If there is a directed edge from node u to node v and one from v to u then we say the edges between u and v are reciprocated or are co-links

Reciprocity - directed networks

Cornell University

- The clustering coefficient measures the frequency of loops of length three in a network. Why only concentrating on loops of length three?
- For example, in directed network it may be more interesting to look at loops of length two which signal a two way relationship. The study of the Hodge Laplacian is ideal for that.
- If there is a directed edge from node u to node v and one from v to u then we say the edges between u and v are reciprocated or are co-links

Reciprocity - directed networks

Cornell University

- The clustering coefficient measures the frequency of loops of length three in a network. Why only concentrating on loops of length three?
- For example, in directed network it may be more interesting to look at loops of length two which signal a two way relationship. The study of the Hodge Laplacian is ideal for that.
- If there is a directed edge from node u to node v and one from v to u then we say the edges between u and v are reciprocated or are co-links

Reciprocity calculation

- The graph Reciprocity r^{rec} is defined as the fraction of edges that are reciprocated.
- Noting that the product of adjacency matrix elements $a_{uv}a_{vu} = 1$ if and only if the edges are reciprocated, the graph reciprocity is:

$$r^{\text{rec}} = \frac{1}{|\mathcal{E}|}\sum_{uv} a_{uv}a_{vu} = \frac{1}{|\mathcal{E}|}\text{Tr}(\boldsymbol{A}^2)$$

where the average is with respect to the total number of (directed) edges in the network.

Reciprocity calculation

- The graph Reciprocity r^{rec} is defined as the fraction of edges that are reciprocated.
- Noting that the product of adjacency matrix elements $a_{uv}a_{vu} = 1$ if and only if the edges are reciprocated, the graph reciprocity is:

$$r^{\text{rec}} = \frac{1}{|\mathcal{E}|}\sum_{uv} a_{uv}a_{vu} = \frac{1}{|\mathcal{E}|}\text{Tr}(\boldsymbol{A}^2)$$

where the average is with respect to the total number of (directed) edges in the network.

Why Stop at Edges?

Cornell University

So far

- Degree is determined by $\boldsymbol{B}_1$.
- Connectivity and diffusion are governed by $\boldsymbol{L}_0 = \boldsymbol{B}_1\boldsymbol{B}_1^\top$.
- Clustering coefficient counts triangles.

Observation

Clustering already uses *2-simplices*. We are no longer in a purely edge-based world.

Natural question

If $\boldsymbol{B}_1$ captures how edges attach to nodes, what operator captures how triangles attach to edges?

Enter $\boldsymbol{B}_2$ and the 1-Hodge Laplacian.

Outline

Cornell University

Beyond the Node Laplacian

Recap: $\boldsymbol{L}_0$

The graph Laplacian $\boldsymbol{L}_0$ acts on node signals and reveals:

- connected components ($\ker \boldsymbol{L}_0$),
- communities (small nonzero eigenvalues),
- node importance (eigenvector centrality).

Key question

What if the data lives on *edges* rather than nodes?

- flows in infrastructure networks,
- traffic and circulation,
- pairwise comparisons or preferences.

The 1-Hodge Laplacian

Definition

The 1-Hodge Laplacian is defined as

$$\boldsymbol{L}_1 = \boldsymbol{B}_1^\top \boldsymbol{B}_1 + \boldsymbol{B}_2 \boldsymbol{B}_2^\top,$$

where:

- $\boldsymbol{B}_1$: node–edge incidence matrix,
- $\boldsymbol{B}_2$: edge–triangle incidence matrix.

Acts on

Edge-valued signals, or flows:

$$\boldsymbol{f} \in \mathbb{R}^{|E|}.$$

Clustering Coefficient and the Upper Laplacian

Cornell University

Clustering coefficient

Clustering measures how many *2-simplices* surround a node v,

$$C_v = \frac{\text{number of triangles incident to } v}{\binom{d_v}{2}}$$

Triangle structure in matrix form

Let $\boldsymbol{B}_2$ be the edge–triangle incidence matrix. Then $\boldsymbol{B}_2\boldsymbol{B}_2^\top \in \mathbb{R}^{|E|\times|E|}$ encodes triangle participation at the edge level:

$$(\boldsymbol{B}_2\boldsymbol{B}_2^\top)_{ee} = \text{number of triangles containing edge } e.$$

Connection to $\boldsymbol{L}_1$

Recall:

$$\boldsymbol{L}_1 = \boldsymbol{B}_1^\top\boldsymbol{B}_1 + \boldsymbol{B}_2\boldsymbol{B}_2^\top.$$

The upper term $\boldsymbol{B}_2\boldsymbol{B}_2^\top$ captures higher-order triangle structure.

Hodge Decomposition of Edge Signals

Cornell University

Hodge decomposition

Any edge flow $\boldsymbol{f}$ can be decomposed as:

$$\boldsymbol{f} = \underbrace{\boldsymbol{B}_1^\top \boldsymbol{\phi}}_{\text{gradient (potential-driven)}} + \underbrace{\boldsymbol{h}}_{\text{harmonic (cycle flow)}} + \underbrace{\boldsymbol{B}_2 \boldsymbol{\psi}}_{\text{curl (local circulation)}} .$$

Interpretation

- Gradient: explainable by node potentials (differences in node values).
- Curl: localized inconsistencies (triangles flows).
- Harmonic: global cycles imposed by topology.

Gradient, Divergence, and Curl on Graphs

Operators and Their Meaning

Object	Operator	Interpretation
Edge flow $\boldsymbol{f}$	$\boldsymbol{B}_1\boldsymbol{f}$	Divergence (net flow at nodes)
Node potential $\boldsymbol{\phi}$	$\boldsymbol{B}_1^\top\boldsymbol{\phi}$	Gradient (potential differences)
Edge flow $\boldsymbol{f}$	$\boldsymbol{B}_2^\top\boldsymbol{f}$	Curl (triangle inconsistency)

Special Cases

- $\boldsymbol{B}_1\boldsymbol{f} = \boldsymbol{0} \Rightarrow$ Divergence-free (pure circulation).
- $\boldsymbol{B}_2^\top\boldsymbol{f} = \boldsymbol{0} \Rightarrow$ Curl-free (no local rotation).
- $\boldsymbol{f} = \boldsymbol{B}_1^\top\boldsymbol{\phi} \Rightarrow$ Gradient flow.
- $\boldsymbol{f} \in \ker \boldsymbol{L}_1 \Rightarrow$ Harmonic flow (divergence- and curl-free).

Topology and $\ker \boldsymbol{L}_1$

Cornell University

Key fact

$$\ker \boldsymbol{L}_1 = \ker \boldsymbol{B}_1 \cap \ker \boldsymbol{B}_2^\top, \quad \text{and} \dim(\ker \boldsymbol{L}_1) = b_1.$$

Consider now $\ker \boldsymbol{L}_1$, i.e. any vector $\boldsymbol{f}$ such that $\boldsymbol{L}_1 \boldsymbol{f} = \boldsymbol{0}$
These flows:

- Are divergence-free $(\boldsymbol{B}_1^\top \boldsymbol{f} = \boldsymbol{0})$
- Are curl-free $\boldsymbol{B}_2 \top \boldsymbol{f} = \boldsymbol{0}$
- Cannot be explained by node potentials

Meaning

- Elements of $\ker \boldsymbol{L}_1$ are *harmonic flows.*
- They represent circulation around topological cycles.
- These modes cannot be removed by local smoothing.

Analogy with Spectral Clustering

Cornell University

L_0 **(nodes)**	L_1 **(edges)**
Clusters	Cycles
Consensus	Circulation
Node embeddings	Edge embeddings
Eigenvector centrality	Edge / cycle importance

Key idea

Small eigenvalues of L_1 reveal *coherent circulation patterns*, just as small eigenvalues of L_0 reveal clusters.

Why L_1 Is Useful

Cornell University

- **Flow denoising**: separate physical flow from noise.
- **Infrastructure analysis**: identify critical loops.
- **Ranking and preferences**: detect global inconsistencies.
- **Anomaly detection**: isolate local vs. global violations.
- **Graph signal processing**: filtering edge-valued data.

Takeaway

L_1 organizes *how things move*, not just how nodes connect.

Spectral Uses of the 1-Hodge Laplacian

Cornell University

Analogy with L_0

- L_0: small eigenvalues reveal **node clusters**.
- L_0: leading eigenvectors define **node centrality**.

What about L_1?

- $\ker L_1$ reveals **topological cycles**.
- Small nonzero eigenvalues reveal **almost-cycles**.
- Eigenvectors of L_1 embed **edges** instead of nodes.

Key idea

L_1 organizes *circulation* the way L_0 organizes *connectivity*.

Spectral Clustering for Edge Signals

Cornell University

Edge centrality eigenproblem

$$\boldsymbol{L}_1 \boldsymbol{f} = \lambda \boldsymbol{f}$$

- Small eigenvalues reveal coherent circulation patterns.
- Edges participating in the same low-frequency mode form **cycle-based communities**.
- Near-zero eigenvalues indicate **almost conserved flows**.

Parallel with L_0

Small eigenvalues of $\boldsymbol{L}_0 \rightarrow$ clusters.
Small eigenvalues of $\boldsymbol{L}_1 \rightarrow$ coherent loops.

Edge Importance and Circulation Centrality

Cornell University

Edge centrality via harmonic modes

Edges that strongly participate in low-frequency eigenvectors of $\boldsymbol{L}_1$ are structurally important for global circulation.

One possible measure:

$$c_e = \sum_{k \in \mathcal{H}} |[\boldsymbol{u}_k]_e|^2,$$

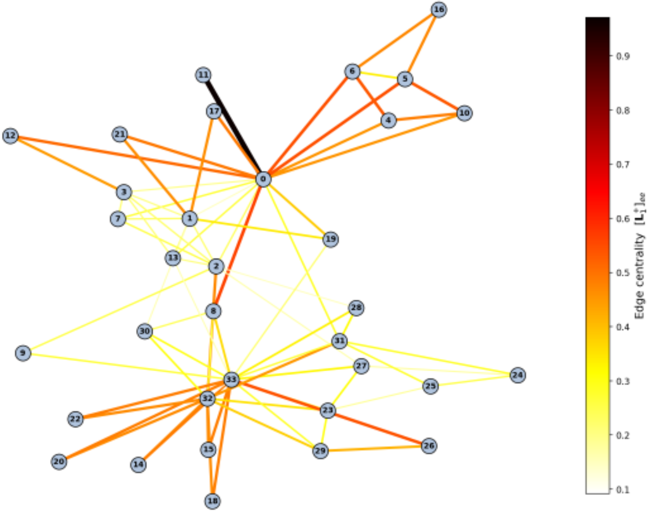
where $\{\boldsymbol{u}_k\}$ are eigenvectors corresponding to small eigenvalues.

Interpretation

- Identifies edges critical to global loops.
- Detects structurally redundant vs. indispensable links.
- Highlights circulation backbones.

Example of Edge Centrality

Zachary's Karate Club — Edge Centrality via Hodge Laplacian $L_1 = B_1^\top B_1 + B_2 B_2^\top$



Edge Centrality via the 1-Hodge Laplacian

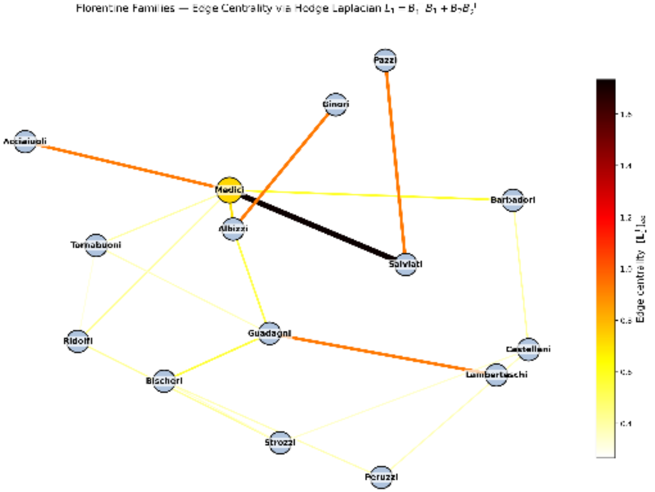
Why is edge $(0, 11)$ maximal?

- Node 11 is a *pendant node* (degree 1).
- Edge $(0, 11)$ belongs to no triangles $\Rightarrow$ no upper-Laplacian contribution $(\boldsymbol{B}_2\boldsymbol{B}_2^\top)$.
- It belongs to no cycles $\Rightarrow$ it is a topological bridge.

Spectral interpretation

The leading eigenvector of $\boldsymbol{L}_1$ (or $\boldsymbol{L}_1^\dagger$) highlights edges that dominate the principal circulation mode. Bridge edges cannot redistribute flow through cycles, so they receive the largest eigenvector weight.

Florentine Families Example



Florentine Families Example

What the Hodge-based edge centrality reveals

- **Top bottleneck: Medici–Salviati** (centrality ≈ 1.73). This edge is the unique conduit to the *Pazzi* family through *Salviati*: removing it disconnects Pazzi from the network (a "bridge of a bridge").
- **Pendant lifelines:** Acciaiuoli–Medici, Salviati–Pazzi, Albizzi–Ginori, Guadagni–Lamberteschi (all ≈ 0.93). Each attaches a degree-1 family with no alternative route.
- **Lowest centrality inside triangles:** Peruzzi–Strozzi (0.27), Ridolfi–Tornabuoni (0.31), Strozzi–Bischeri (0.34). Edges in triangles are redundant: removing one still leaves a 2-step path via the third vertex.

Ranking and Inconsistency Detection

Suppose edges encode pairwise comparisons:

$$f_{ij} = \text{score of } i \text{ vs } j.$$

Hodge decomposition

$$\boldsymbol{f} = \boldsymbol{B}_1^\top \boldsymbol{\phi} + \boldsymbol{B}_2 \boldsymbol{\psi} + \boldsymbol{h}$$

- Gradient component $\rightarrow$ global ranking.
- Curl component $\rightarrow$ local inconsistencies.
- Harmonic component $\rightarrow$ global cyclic contradictions.

Applications

Tournament ranking, preference aggregation, crowd-sourced labeling.

References on L_1 and Hodge Theory

Cornell University

- Jiang, Lim, Yao, Ye (2011). *Statistical Ranking and Combinatorial Hodge Theory.*
- Lim (2020). *Hodge Laplacians on Graphs.* SIAM Review.
- Schaub et al. (2020). *Signal Processing on Higher-Order Networks.*
- Barbarossa & Sardellitti (2020). *Topological Signal Processing over Simplicial Complexes.*

Outline

Cornell University

Conclusions

Cornell University

- We discussed how to associate features to graphs, based on their structure
- We defined local (nodal) and group properties:
  1. Degree distribution
  2. Centrality metrics
  3. Similarity metrics and assortatitive networks
  4. Clustering coefficient
  5. Reciprocity
- We then discussed the use of 1-Hodge Laplacian do extend the feature analysis to network edges
- Next time we will look at models for Random Graphs noting how the different statitical model shape these random graphs measures

Conclusions

Cornell University

- We discussed how to associate features to graphs, based on their structure
- We defined local (nodal) and group properties:
  1. Degree distribution
  2. Centrality metrics
  3. Similarity metrics and assortatitive networks
  4. Clustering coefficient
  5. Reciprocity
- We then discussed the use of 1-Hodge Laplacian do extend the feature analysis to network edges
- Next time we will look at models for Random Graphs noting how the different statitical model shape these random graphs measures

Conclusions

Cornell University

- We discussed how to associate features to graphs, based on their structure
- We defined local (nodal) and group properties:
  1. Degree distribution
  2. Centrality metrics
  3. Similarity metrics and assortatitive networks
  4. Clustering coefficient
  5. Reciprocity
- We then discussed the use of 1-Hodge Laplacian do extend the feature analysis to network edges
- Next time we will look at models for Random Graphs noting how the different statitical model shape these random graphs measures