Feature Extraction Methods on Graphs without Learning (Node-Level)

ACMS 80770: Deep Learning with Graphs

Instructor: Navid Shervani-Tabar

Department of Applied and Comp Math and Stats

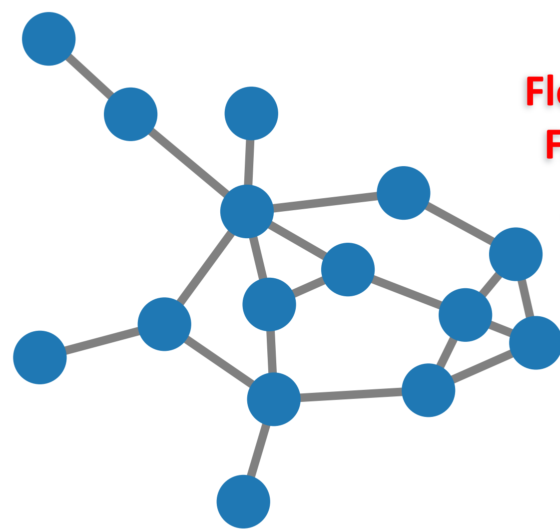
UNIVERSITY OF NOTRE DAME

Feature Extraction

- The structure of a graph is fully described by its nodes and their connections.
- While a small graph may be qualitatively analyzed by eye inspection, large scale study of graphs needs mathematical measures to quantify their structure.
- The extracted features should characterize a graph on different levels for different tasks.
- These features are inputs for standard machine learning algorithms.
- We look at measures of extracting node and edge-level features and statistics.

Centrality

❖ Consider the study of nodes in the following social graph.



Florentine Families graph

❖ Centrality measures how central and important a node is.

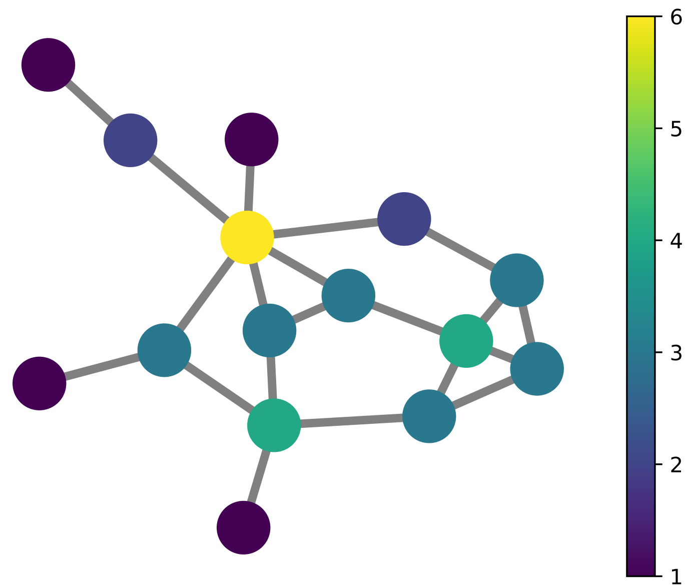
❖ Various definitions of importance lead to various centrality measures.

Degree Centrality

- **Degree centrality:** measures the number of a node's neighbors.

$$d_i = \sum_{v_j \in V} A_{ij}$$

- Number of friends in a social network (undirected).
- Number of citations in a citation graph (directed)



Eigenvector Centrality

- But not all neighbors are as important.
- One may also consider the importance of the neighbors.
- **Eigenvector centrality:** Centrality of a node v_i is proportional to the sum of the centrality of its neighbors.

$$e_i = \frac{1}{\lambda} \sum_{v_j \in N(v_i)} e_j$$

Centrality

$$= \frac{1}{\lambda} \sum_{v_j \in V} \mathbf{A}_{ij} e_j$$

- Writing in matrix notation yields eigendecomposition of **A**.

$$\lambda \boldsymbol{e} = \boldsymbol{A} \boldsymbol{e}$$

Eigenvector Centrality

- But not all neighbors are as important.
- One may also consider the importance of the neighbors.
- **Eigenvector centrality:** Centrality of a node v_i is proportional to the sum of the centrality of its neighbors.

$$e_i = \frac{1}{\lambda} \sum_{v_j \in N(v_i)} e_j$$
$$= \frac{1}{\lambda} \sum_{v_j \in V} \mathbf{A}_{ij} e_j$$

Centrality

- Writing in matrix notation yields eigendecomposition of **A**.

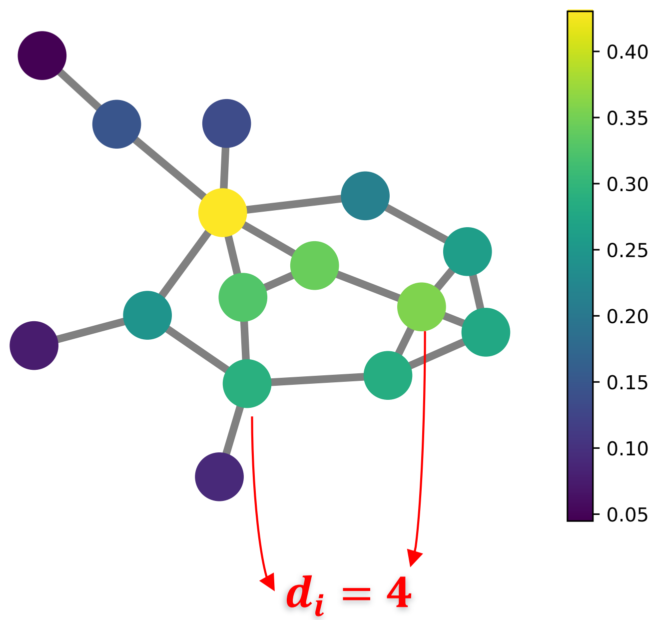
$$\lambda \boldsymbol{e} = \boldsymbol{A} \boldsymbol{e}$$

Leading eigenvalue

- To keep the centrality of nodes non-negative, we select the leading eigenvector as the solution (Perron-Forbenius theorem).

Eigenvector Centrality

➢ Florentine Families graph



❖ Eigenvector centrality is more suitable for undirected graphs.

Eigenvector Centrality

- For a directed graph, the adjacency matrix is (generally) asymmetric and hence has two sets of right and left eigenvalues.
- Incoming edges is usually a better indicator of importance of the node.
- Web, citation
- We look at the rows of the adjacency matrix for the incoming edges

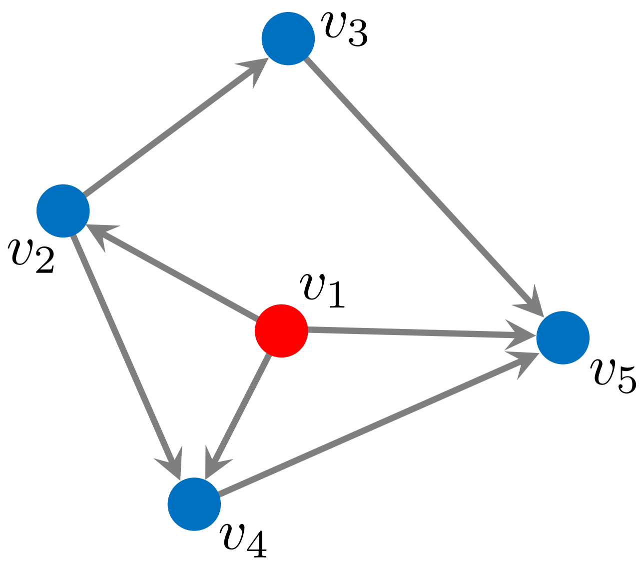
$$e_i = \frac{1}{\lambda} \sum_{v_j \in V} \mathbf{A}_{ij} e_j$$

- As a result, to value incoming edges, one may select right leading eigenvector as the centrality measure.

$$\lambda \boldsymbol{e} = \boldsymbol{A}\boldsymbol{e}$$

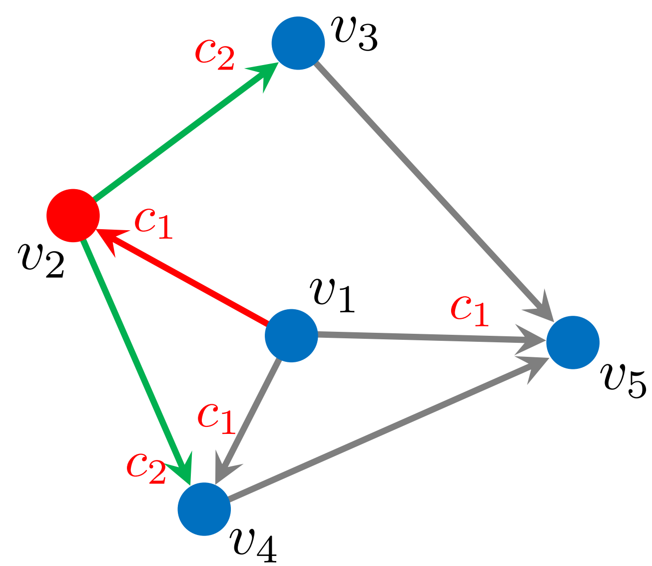
Eigenvector Centrality

- Consider v_1 which has no incoming edges.
- It may seem reasonable to set centrality of v_1 to zero.



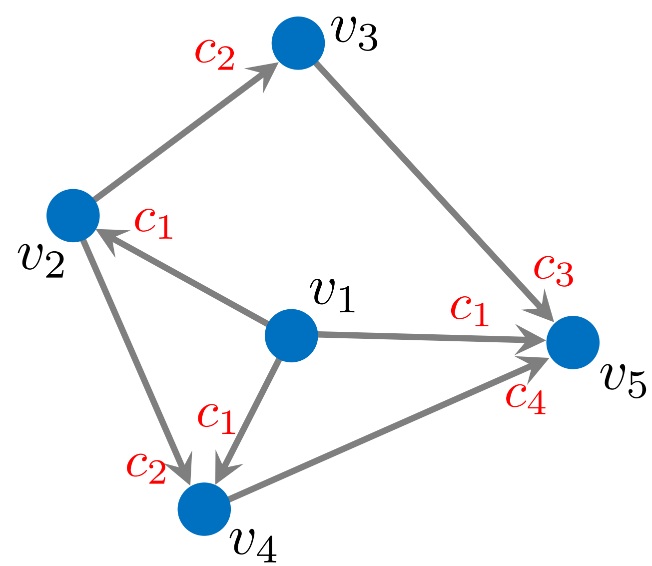
Eigenvector Centrality

- Consider v_1 which has no incoming edges.
- It may seem reasonable to set centrality of v_1 to zero.
- Consider v_2, which has one incident edge.



Eigenvector Centrality

- Consider v_1 which has no incoming edges.
- It may seem reasonable to set centrality of v_1 to zero.
- Consider v_2, which has one incident edge.
- But what if the node connected to the incoming node had no importance?



- Acyclic citation network.

Katz Centrality

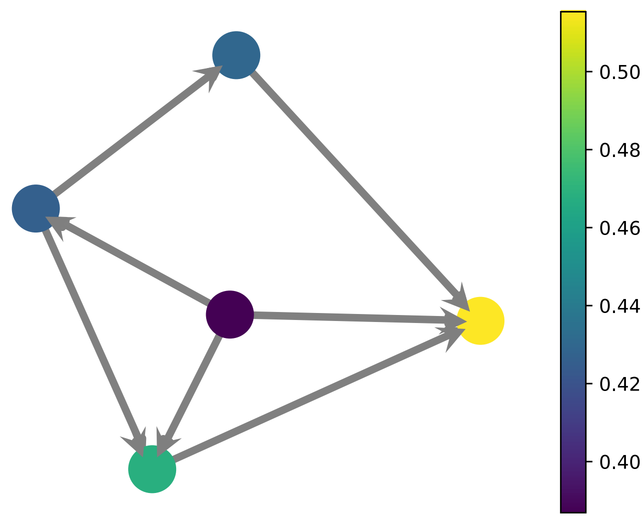
- **Katz centrality:** Adds a small centrality to every node, regardless of their in-degree.

$$c_i = \alpha \sum_{v_j \in V} \mathbf{A}_{ij} c_j + \beta$$

- In matrix notation, $\boldsymbol{c} = \alpha \boldsymbol{A} \boldsymbol{c} + \beta \vec{\mathbf{1}}$. Rearranging yields

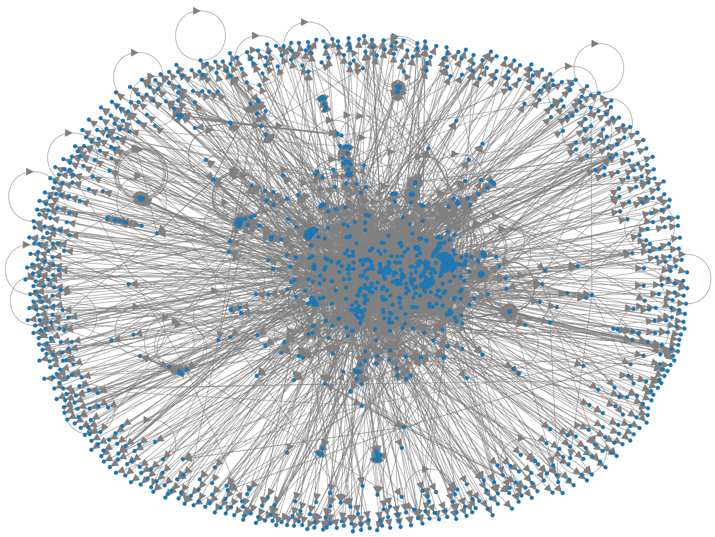
$$\boldsymbol{c} = \beta (\boldsymbol{I} - \alpha \boldsymbol{A})^{-1} \vec{\mathbf{1}}$$

- α balances between eigenvector centrality term and the constant term.



PageRank

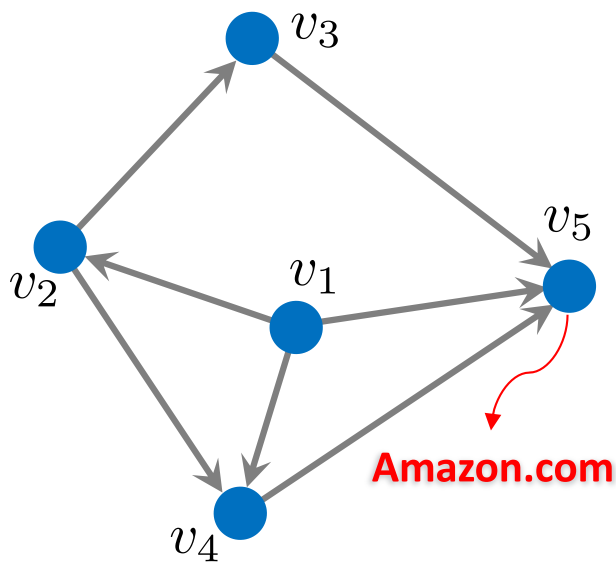
- Consider WWW.
- We would like to rank pages based on their importance for a search engine.
- We model web as a graph, node represent webpage and edges represent hyperlinks.
- A centrality measure can be used to rank pages (nodes) in this graph.



nd.edu graph

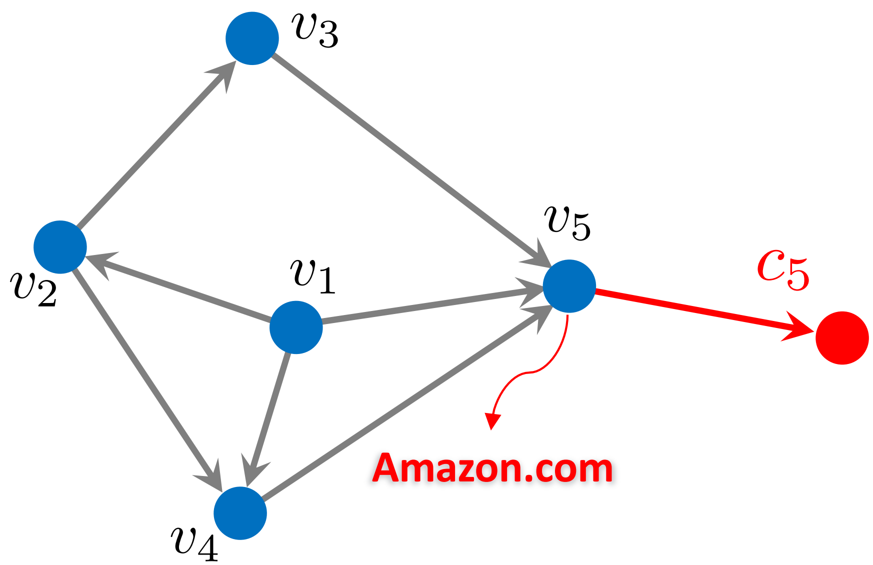
PageRank

- Let's consider using the Katz centrality.
- A website like Amazon has a high importance and centrality.



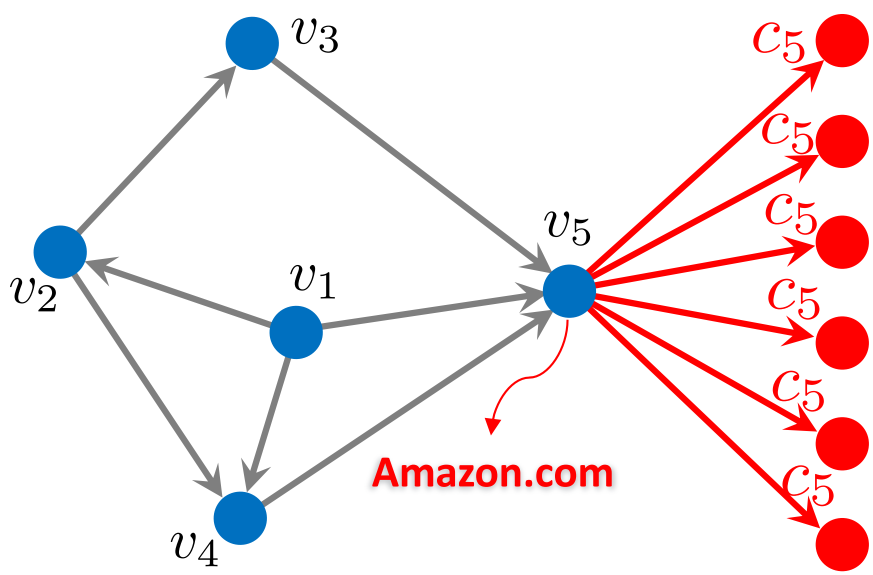
PageRank

- Let's consider using the Katz centrality.
- A website like Amazon has a high importance and centrality.
- Webpages linked by amazon would also have high centrality.



PageRank

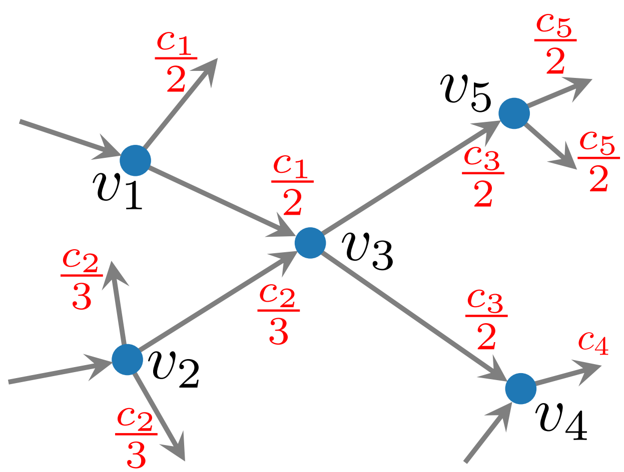
- Let's consider using the Katz centrality.
- A website like Amazon has a high importance and centrality.
- Webpages linked by amazon would also have high centrality.
- But many unimportant pages are linked by Amazon, which all receive high centrality.



PageRank

- It is reasonable to reduce the outgoing centrality of node based on the number of its outgoing edges.
- Hence, each node passes its centrality divided by its out-degree through each edge.

$$
\begin{aligned}
c_i &= \alpha \sum_{v_j \in N^{(in)}(v_i)} \frac{c_j}{d_j^{(\text{out})}} \\
&= \alpha \sum_{v_j \in V} A_{ij} \frac{c_j}{d_j^{(\text{out})}}
\end{aligned}
$$

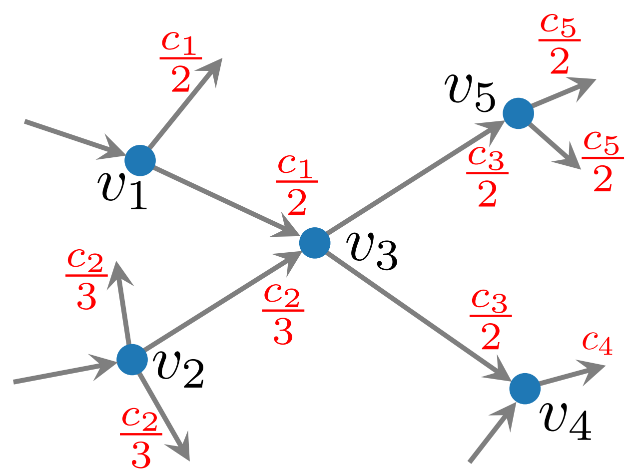


- For nodes without outgoing edges, we set $d_j = 1$.

PageRank

- It is reasonable to reduce the outgoing centrality of node based on the number of its outgoing edges.
- Hence, each node passes its centrality divided by its out-degree through each edge.

$$c_i = \alpha \sum_{v_j \in N^{(in)}(v_i)} \frac{c_j}{d_j^{(\mathrm{out})}} + \beta$$
$$= \alpha \sum_{v_j \in V} A_{ij} \frac{c_j}{d_j^{(\mathrm{out})}} + \beta$$



- For nodes without outgoing edges, we set $d_j = 1$.
- Similar to Katz, we add β for webpages with no in-links.

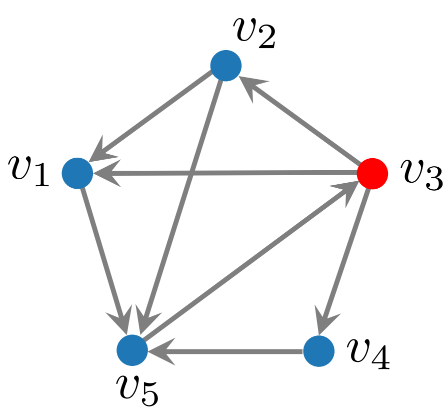
PageRank

- In matrix notation (with $\beta = 1$),

$$\mathbf{c} = (I - \alpha \underbrace{\mathbf{A}\mathbf{D}^{-1}}) ^{-1} \vec{\mathbf{1}}$$

Rank vector

Stochastic Adjacency Matrix



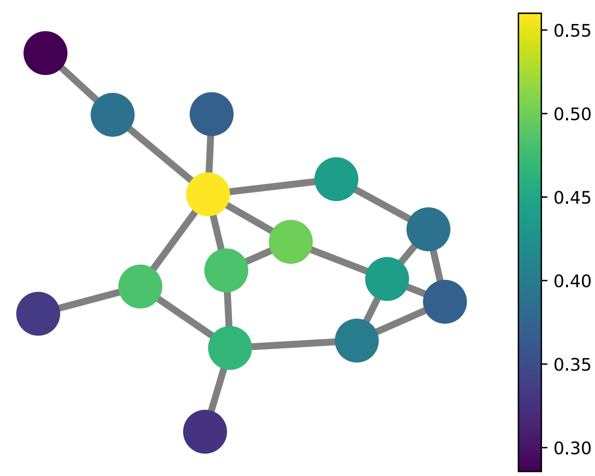
	$1/2$	$1/3$		
		$1/3$		
				1
		$1/3$		
1	$1/2$		1	

- Each column in the stochastic adjacency matrix sums to 1.
- The stochastic adjacency matrix is a left stochastic matrix.
  - Google

Closeness Centrality

- Apart from matrix concepts, we may use shortest path to define centrality.
- **Closeness centrality:** Measures the mean distance from a node to other nodes.
- If l_{ij} denotes the shortest distance between nodes v_i and v_j , the closeness centrality is computed as

$$c_i = \frac{1}{\frac{1}{|V|}\sum_{v_j \in V} l_{ij}} = \frac{|V|}{\sum_{v_j \in V} l_{ij}}$$

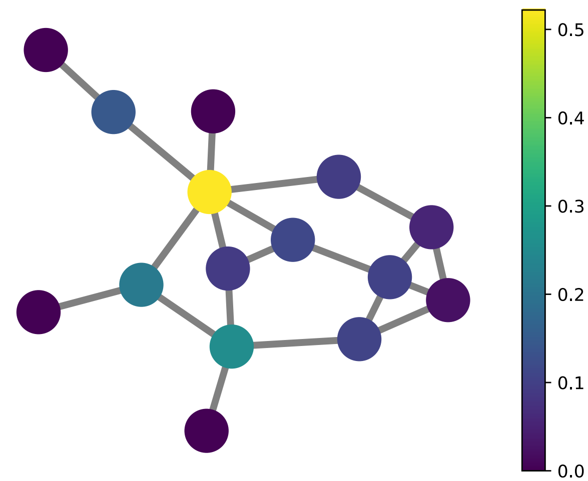


➢ Faster spread of someone's opinions through a community.

Betweenness Centrality

- **Betweenness centrality:** Measures how often a node lies on the path between two nodes in the graph.

$$c_k = \sum_{v_i \in V} \sum_{v_j \in V} n_{ij}^k$$



where

$$n_{ij}^k = \begin{cases} 1 & \text{if } v_k \text{ on the shortest} \\ & \text{path between } v_i, v_j \\ 0 & \text{otherwise.} \end{cases}$$

- A measure of nodes influence on the flow of information in the graph.
  - Cities and postal services.

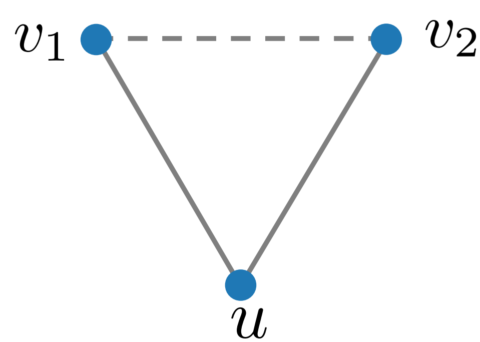
Clustering Coefficient

- **Clustering coefficient:** Computes the ratio of connected neighbors v_1 and v_2 of node u to the total pair of nodes in its neighborhood.

Connected pair of neighbors

$$c_u = \frac{|(v_1, v_2) \in \mathcal{E} : v_1, v_2 \in \mathcal{N}(u)|}{\frac{1}{2} d_u (d_u - 1)}$$

Total pair of neighbors



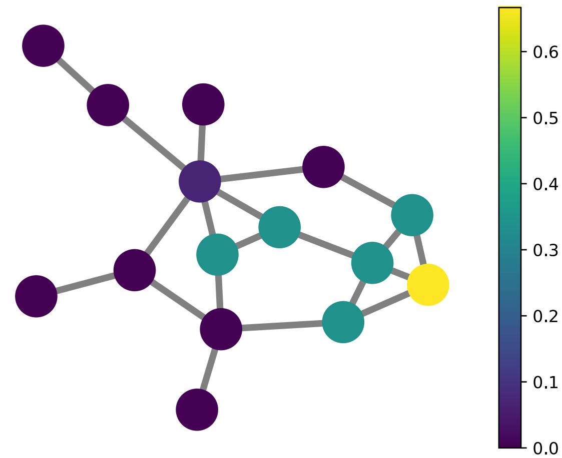
Clustering Coefficient

- **Clustering coefficient:** Computes the ratio of connected neighbors v_1 and v_2 of node u to the total pair of nodes in its neighborhood.

Connected pair of neighbors

$$c_u = \frac{|(v_1, v_2) \in \mathcal{E} : v_1, v_2 \in \mathcal{N}(u)|}{\frac{1}{2} d_u (d_u - 1)}$$

Total pair of neighbors



- If ego graph of node u is a clique, then $c_u = 1$.

Clustering Coefficient

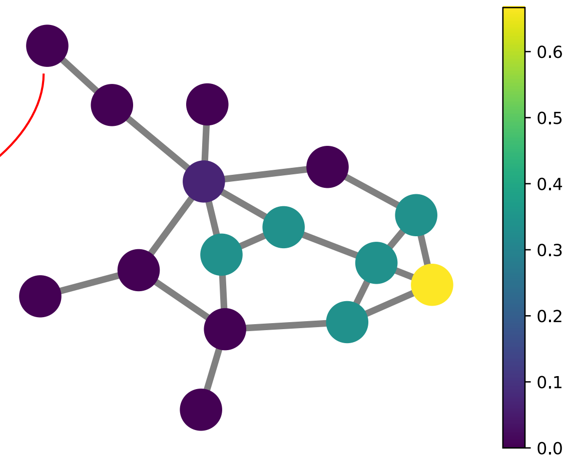
- **Clustering coefficient:** Computes the ratio of connected neighbors v_1 and v_2 of node u to the total pair of nodes in its neighborhood.

$$c_u = \frac{|(v_1, v_2) \in \mathcal{E} : v_1, v_2 \in \mathcal{N}(u)|}{\frac{1}{2} d_u (d_u - 1)}$$

Connected pair of neighbors

Total pair of neighbors

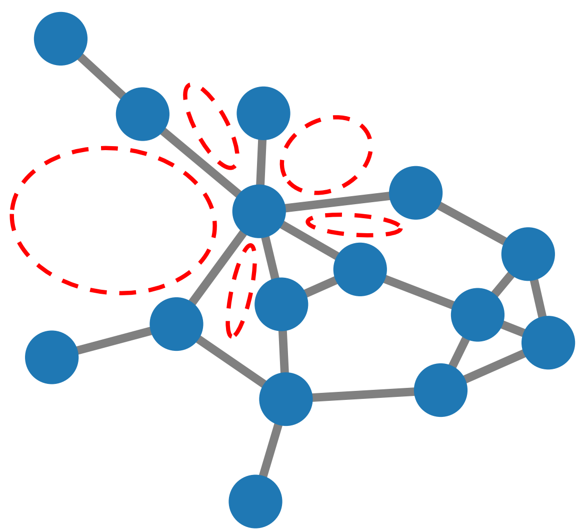
$c_u = 0$



- If ego graph of node u is a clique, then $c_u = 1$.
- By convention, clustering coefficient is zero for disconnected nodes and nodes with 1 neighbor.

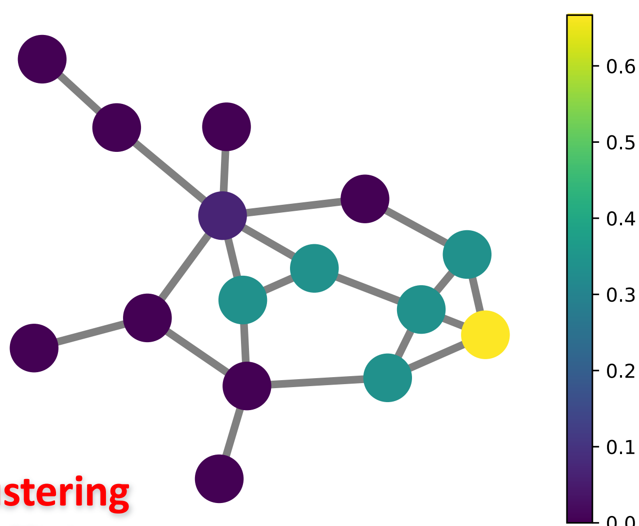
Structural Hole

- **Structural hole:** Disconnected neighbors introduce structural holes in the graph.
- Structural holes can have both advantages and disadvantages:
  - High traffic flow in a poorly connected neighborhood.
  - Added value for transit hubs.

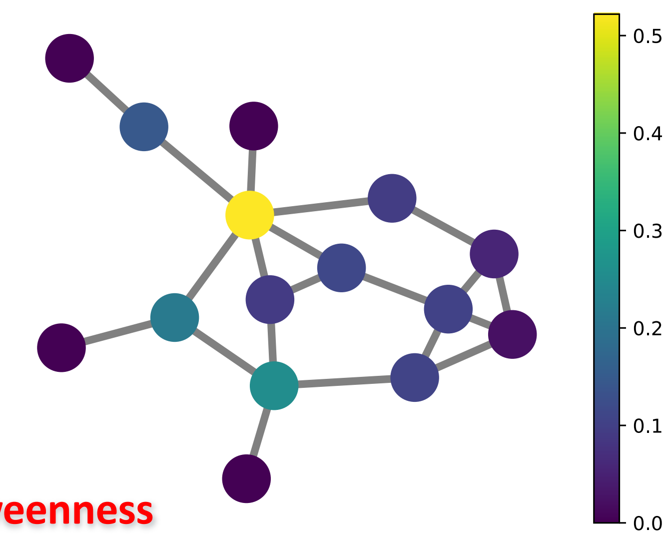


Clustering Coefficient

- Like betweenness, clustering coefficient reflects control of the node over the flow of information.
- Betweenness centrality shows nodes centrality in a global perspective, while clustering coefficient shows a local measure.
- Betweenness centrality is computationally expensive.



Clustering Coefficient

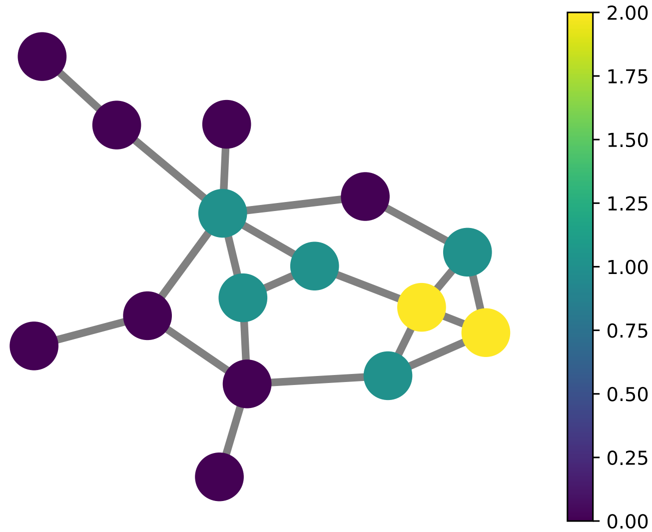


Betweenness Centrality

Motif Count

❖ **Motif count**: Counts number if arbitrary structures of given size within a node's neighborhood

➢ Triangles



Summary

- Node-Level Features:
  - Node Centrality:
    - Degree centrality
    - Eigenvector centrality
    - Katz centrality
    - PageRank
    - Closeness Centrality
    - Betweenness Centrality
  - Clustering Coefficient
  - Structural Holes
  - Motif Count