Social Media Analytics

Network Measures

— SL03 —

Mourad Khayati

mourad.khayati@unifr.ch

Table of contents — SL03

1. Centrality
   Definition & Application
   Measures

2. Similarity

Social Influence/1

What can we do if we can rank all the nodes in a graph (e.g., Facebook, LinkedIn, Twitter)?

- Find **celebrities** or influential people in a social network
- Find **gatekeepers** who connect communities (headhunters love to find them on LinkedIn)
- etc.

Social Influence/2

Klout:

Influence is the ability to drive action. When you share something on social media or in real life and people respond, that's influence. The more influential you are, the higher is your Klout Score.

Most Influential Celebrities on Twitter (Klout)

Sort By: Name - Fans - Fans Today - Talking About - Followers - Followers Today - Lists - Klout - Total

View:

Rank	Name	Facebook (Fans | Talking About)	Twitter (Followers | Klout)	Total
#1 tweet this	Barack Obama, Politician, Like, Follow, Follow on Foursquare	55,276,944; 1,465,248 Talking About	98,596,510; 98.6	153,873,454
#2 tweet this	Donald J. Trump, Executive, Like, Follow	23,019,477; 1,490,872 Talking About	45,997,920; 94.4	69,017,397
#3 tweet this	Britney Spears, Musician, Like, Follow	39,014,869; 176,665 Talking About	56,472,273; 94.4	95,487,142
#4 tweet this	The Beatles, Musician, Like, Follow	41,612,126; 116,691 Talking About	3,614,465; 93.9	45,226,591
#5 tweet this	Eminem, Musician, Like, Follow	89,765,790; 156,086 Talking About	22,277,142; 93.6	112,042,932

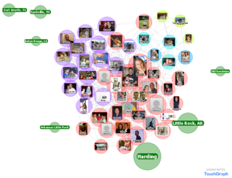
RANKING OF TOP FINTECH INFLUENCERS IN DACH (GERMANY, SWITZERLAND, AUSTRIA)

Centrality: Definition

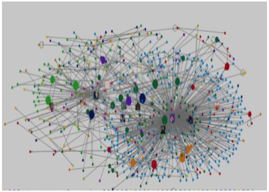
- The centrality measures are used to define the central node which:
  - is important and/or powerful
  - has an influential/advantageous position on the network
- There are a vast number of different centrality measures that have been proposed over the years.
- We will study various measures which have been associated to the concept of centrality.

CENTRALITY SOCIAL QUERIES/1

- The centrality measures allow to answer social network-based queries such as:



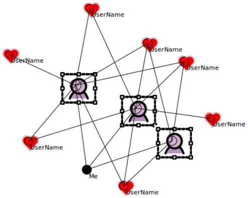
Who is the most central person among my Facebook friends?



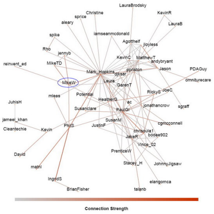
What interaction patterns are common on my Twitter network?

Centrality Social Queries/2

- The centrality measures allow to answer social network-based queries such as:



Who are the like-minded users among my Instagram friends?



Who is is the most efficient spreader(s) on my Linkedin network?

Centrality Measures: Overview

The most common centrality measures are:

- Based on degree:
  - Degree centrality
- Based on geodesics:
  - Closeness centrality
  - Betweenness centrality
- Recursive:
  - Eigenvector centrality
  - PageRank centrality

Node Degree Centrality: Definition

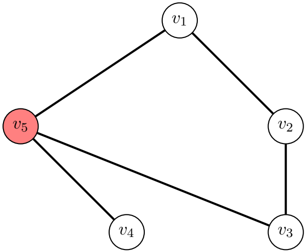
Idea:
A central actor is the one with many connections.

Definition:

- The degree of centrality $d(v_i)$ of a node v_i is the number of ties that v_i has in the network.
- In social media networks, $d(v_i)$ represents the number of friends person v_i has.

Node Degree Centrality: Example

- $d(v_1)$= 2
- $d(v_2)$= 2
- $d(v_3)$= 2
- $d(v_4)$= 1
- $d(v_5)$= 3



Node Degree Centrality: Computation

- In directed graphs, the degree centrality value is computed using the in-degree, the out-degree, or the combination.
- $k(v_i, v_j)$ is the function that returns 1 if there exists a connection from v_j to v_i, 0 otherwise.
- The centrality value can be:
  - in-degree centrality: $d^i(v_i) = \sum_{j, j \neq i} k(v_i, v_j)$
  - out-degree centrality: $d^o(v_i) = \sum_{j, j \neq i} k(v_j, v_i)$
  - combined centrality: $d^c(v_i) = d^i(v_i) + d^o(v_i)$

Normalized Degree Centrality

- In order to compare centrality values across networks, the degree centrality needs to be normalized.
- The degree centrality can be normalized by:
  - the maximum possible degree: $\tilde{d}(v_i) = \frac{d(v_i)}{n-1}$
  - the maximum degree: $d^{max}(v_i) = \frac{d(v_i)}{max_{j,j\in\{1,n\}} d(v_j)}$
  - the degree sum: $d^{sum}(v_i) = \frac{d(v_i)}{\sum_{j,j\in\{1,n\}} d(v_j)} = \frac{d(v_i)}{2\times|E|} = \frac{d(v_i)}{2\times m}$

Graph Degree Centrality

- The centrality of a graph is a degree vector V that contains the degree centrality of all nodes.
- Let A be the adjacency matrix of a undirected graph and let $U \in \mathbb{R}^n$ be the ones vector (all ones). Then,

$$V = A \cdot U$$

Graph Degree Centrality

- The centrality of a graph is a degree vector V that contains the degree centrality of all nodes.
- Let A be the adjacency matrix of a undirected graph and let $U \in \mathbb{R}^n$ be the ones vector (all ones). Then,

$$V = A \cdot U$$

- Example:

$$\underbrace{\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array}\begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array}}_{\text{A}} \cdot \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{U} = \underbrace{\begin{bmatrix} 2 \\ 2 \\ 2 \\ 1 \\ 3 \end{bmatrix}}_{V}$$

DEGREE CENTRALITY APPLICATION

The degree centrality makes sense when the number of contacts is important:

- Number of supporters
- Number of confidants
- Number of trading partners
- Immediate influence

Closeness Centrality: Definition

Idea:
A central node is the one that is close, on average, to other nodes.

Definition:

- The closeness centrality emphasizes the distance of a node to other nodes on a network.
- In social networks, closeness can be seen as a measure of how long it will take to spread information from a node v_i to all other nodes sequentially.

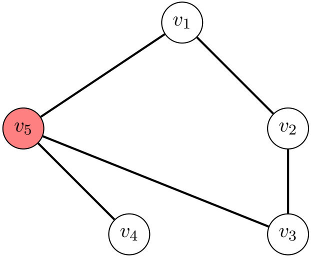
Closeness Centrality: Formula

- The geodesic distance $d(v_i, v_j)$ is the minimal path length from v_i to v_j.
- The closeness centrality of node v_i is defined as:

$$c^c(v_i) = \frac{1}{\frac{1}{n-1}\sum_{j,j\neq i} d(v_i, v_j)}$$

Closeness Centrality: Example

- $c^c(v_1) = \frac{1}{((1+1+2+2)/4)} = 0.66$
- $c^c(v_2) = \frac{1}{((1+1+2+3)/4)} = 0.57$
- $c^c(v_3) = \frac{1}{((1+1+2+2)/4)} = 0.66$
- $c^c(v_4) = \frac{1}{((1+2+2+3)/4)} = 0.5$
- $c^c(v_5) = \frac{1}{((1+1+1+2)/4)} = 0.8$



Closeness Centrality Application

Closeness centrality is often applied whenever direct access is important:

- Access to information
- Opinion formation
- Spread or rumors
- Adoption of new technology

Betweenness Centrality: Definition

Idea:
A central actor is the one that acts as a bridge, broker or gatekeeper.

Definition:

- Betweenness centrality measures the number of times a node acts as a bridge along the shortest path between two other nodes.
- In this context, a node that has a high probability to occur on a randomly chosen shortest path between two randomly chosen nodes, will have a high betweenness value.

Betweenness Centrality: Computation/1

- The betweenness centrality of a node v_i is computed as follows:

$$c^b(v_i) = \sum_{i \neq j \neq k} \frac{\sigma_{v_j v_k}(v_i)}{\sigma_{v_j v_k}}$$

  - $\sigma_{v_j v_k}$ is the total number of shortest paths from node v_j to node v_k.
  - $\sigma_{v_j v_k}(v_i)$ is the number of those paths that pass through v_i.

Betweenness Centrality: Computation/2

The betweenness of a node v_i in a graph $\mathcal{G} = (V, E)$ is computed as follows:

- For each pair of nodes (v_j, v_k), compute the shortest paths between them.

Betweenness Centrality: Computation/2

The betweenness of a node v_i in a graph $\mathcal{G} = (V, E)$ is computed as follows:

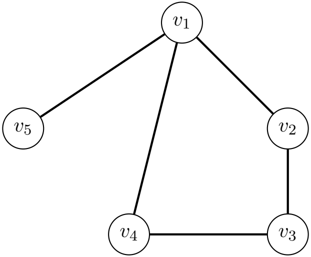
- For each pair of nodes (v_j, v_k), compute the shortest paths between them.
- For each pair of nodes (v_j, v_k), determine the fraction of shortest paths that pass through the node in question (here, node v_i).

Betweenness Centrality: Computation/2

The betweenness of a node v_i in a graph $\mathcal{G} = (V, E)$ is computed as follows:

- For each pair of nodes (v_j, v_k), compute the shortest paths between them.
- For each pair of nodes (v_j, v_k), determine the fraction of shortest paths that pass through the node in question (here, node v_i).
- Sum up this fraction over all pairs of vertices (v_j, v_k).

Betweenness Centrality: Example



$$
\begin{aligned}
c^b(v_1) &= \underbrace{0/1}_{j=v_2,k=v_3} + \underbrace{1/2}_{j=v_2,k=v_4} + \underbrace{1/1}_{j=v_2,k=v_5} + \underbrace{0/1}_{j=v_3,k=v_4} + \underbrace{2/2}_{j=v_3,k=v_5} + \underbrace{1/1}_{j=v_4,k=v_5} \\
&= 3.5
\end{aligned}
$$

$$
\begin{aligned}
c^b(v_2) &= \underbrace{1/2}_{j=v_1,k=v_3} + \underbrace{0/1}_{j=v_1,k=v_4} + \underbrace{0/1}_{j=v_1,k=v_5} + \underbrace{0/1}_{j=v_3,k=v_4} + \underbrace{1/2}_{j=v_3,k=v_5} + \underbrace{0/1}_{j=v_4,k=v_5} \\
&= 1
\end{aligned}
$$

Betweenness Centrality Application

Betweenness centrality make sense when you gain from bridging between different groups

- Brokering between groups
- Control of information
- Innovation
- Collaboration

Betweenness Centrality Limits

Lower limit:

- $0 \leq c^b(v_i)$
- $c^b(v_i) = 0$ if v_i lies on none of the shortest paths

Upper limit:

- $$c^b(v_i) \leq \binom{n-1}{2} \qquad \text{binomial}$$
  $$\leq \frac{(n-1)(n-2)}{2}$$
- $c^b(v_i) = \frac{(n-1)(n-2)}{2}$: if v_i lies on all shortest paths

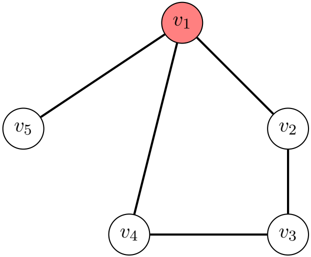
Normalized Betweenness Centrality

- The betweenness centrality needs to be normalized to be comparable across networks.
- The normalized betweenness centrality is obtained by dividing the betweenness centrality by the maximum betweenness value.
- Formally,

$$\tilde{c}^b(v_i) = \frac{2 \times c^b(v_i)}{(n-1)(n-2)}$$

Normalized Betweenness Centrality: Example

- $n = 5$
- $\tilde{c}^b(v_1) = \frac{2 \times c^b(v_1)}{4 \times 3} = 0.58$
- $\tilde{c}^b(v_2) = \frac{2 \times c^b(v_2)}{4 \times 3} = 0.16$
- $\tilde{c}^b(v_3) = \frac{2 \times c^b(v_3)}{4 \times 3} = 0.08$
- $\tilde{c}^b(v_4) = \frac{2 \times c^b(v_4)}{4 \times 3} = 0.16$
- $\tilde{c}^b(v_5) = \frac{2 \times c^b(v_5)}{4 \times 3} = 0$



Eigenvector Centrality: Definition/1

Idea:
A central actor is the one that is connected to other central actors (having most information flowing through).

Eigenvector Centrality: Definition/1

Idea:
A central actor is the one that is connected to other central actors (having most information flowing through).

Definition:

- Eigenvector centrality is used in undirected graphs.
- Let A be the adjacency matrix of a graph $\mathcal{G}$ and $a_{ij} \in A$, then the eigenvector centrality of a node v_i is:

$$c^e(v_i) = \frac{1}{\lambda} \sum_{j, j \neq i} (a_{ij} \times c^e(v_j))$$

  where $\lambda \neq 0$ is a constant
- A node's eigenvector centrality is proportional to the centrality of its neighbors.

Eigenvector Centrality: Definition/2

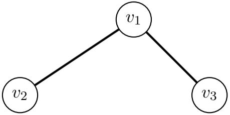
- The eigenvector centrality is expressed in matrix form as follows:

$$AC^e = \lambda C^e \tag{1}$$

where $C^e = \{c^e(v_1), c^e(v_2), \ldots, c^e(v_n)\}$

- C^e is an eigenvector of the adjacency matrix A and λ is its corresponding eigenvalue.
- The formula (1) can be solved using basic linear algebra.

EIGENVECTOR CENTRALITY: EXAMPLE 1



$$A = \begin{array}{c c} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} v_1 \\ v_2 \\ v_3 \end{array} & \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

Let $C^e = [u_1, u_2, u_3]^T$ and I the identity matrix. We have

$$AC^e = \lambda C^e \Rightarrow$$
$$(A - \lambda I)C^e = 0 \Rightarrow$$
$$det(A - \lambda I) = 0$$

In order to compute λ and C^e we need to solve $det(A - \lambda I) = 0$.

Eigenvector Centrality: Example 1 (cont.)

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}; \lambda I = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \Rightarrow A - \lambda I = \begin{bmatrix} -\lambda & 1 & 1 \\ 1 & -\lambda & 0 \\ 1 & 0 & -\lambda \end{bmatrix}$$

$$\begin{aligned} det(A - \lambda I) &= \begin{vmatrix} -\lambda & 1 & 1 \\ 1 & -\lambda & 0 \\ 1 & 0 & -\lambda \end{vmatrix} \\ &= -\lambda \begin{vmatrix} -\lambda & 0 \\ 0 & -\lambda \end{vmatrix} - 1 \begin{vmatrix} 1 & 0 \\ 1 & -\lambda \end{vmatrix} + 1 \begin{vmatrix} 1 & -\lambda \\ 1 & 0 \end{vmatrix} \\ &= -\lambda(-\lambda \times -\lambda) - 1(-\lambda - 0) + 1(0 - (-\lambda)) \\ &= -\lambda^3 + 2\lambda = \lambda(2 - \lambda^2) = 0 \Rightarrow \lambda = (0, -\sqrt{2}, \sqrt{2}) \end{aligned}$$

We select the largest $\lambda = \sqrt{2}$ to find the corresponding eigenvector.

Eigenvector Centrality: Example 1 (cont.)

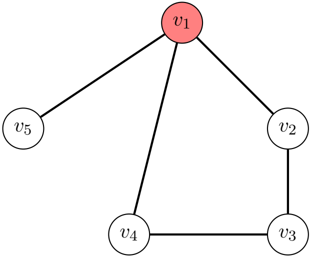
$$A - \lambda I = 0 \Rightarrow$$

$$\begin{bmatrix} -\sqrt{2} & 1 & 1 \\ 1 & -\sqrt{2} & 0 \\ 1 & 0 & -\sqrt{2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow$$

$$\begin{cases} -\sqrt{2}u_1 + u_2 + u_3 = 0 \\ u_1 - \sqrt{2}u_2 = 0 \\ u_1 - \sqrt{2}u_3 = 0 \end{cases} \Rightarrow \begin{cases} u_1 = t \\ u_2 = \frac{t}{\sqrt{2}} \\ u_3 = \frac{t}{\sqrt{2}} \end{cases}$$; we set $t = \dfrac{1}{\sqrt{2}}$ and get

$$\begin{cases} u_1 = \frac{1}{\sqrt{2}} \\ u_2 = \frac{1}{2} \\ u_3 = \frac{1}{2} \end{cases} \Rightarrow C^e = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

Eigenvector Centrality: Example 2



$$A = \begin{array}{c c} & \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array} & \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$\lambda = (-2.136, \underbrace{2.136}_{\lambda_{max}}, -0.662, 0.662, 0) \Rightarrow C^e = \begin{bmatrix} 0.557 \\ 0.465 \\ 0.435 \\ 0.465 \\ 0.261 \end{bmatrix}$$

Eigenvector Centrality Application

Eigenvector centrality is often applied when you want to extend the influence beyond the direct connections into the wider network:

- long-term influence over the whole network
- understand human social networks
- understand propagation in networks (e.g., malware propagation)
- large graphs

PageRank Centrality

Google describing PageRank:

Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important".

PageRank Centrality: Definition/1

Idea:
A central node is the one that is either linked from other important and link parsimonious nodes or highly linked with others.

PageRank Centrality: Definition/1

Idea:
A central node is the one that is either linked from other important and link parsimonious nodes or highly linked with others.

Definition:

- Let A be the adjacency matrix of a graph $\mathcal{G}$ and $a_{ij} \in A$, then the Google PageRank centrality of a node v_i is:

$$c^p(v_i) = \alpha \sum_{j,j \neq i} \left(\frac{a_{ij}}{d^o(v_j)} \times c^p(v_j)\right) + \beta$$

where α and β are constants, and $d^o(v_j)$ is the out-degree of node v_j.

PageRank Centrality: Definition/2

The PageRank centrality is expressed in matrix form as follows:

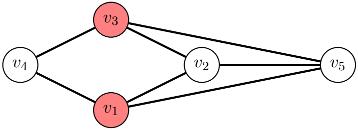
$$C^p = \beta(I - \alpha A^T D^{-1})^{-1} \times \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

where

- D^{-1} is a diagonal matrix with the i-th diagonal element equal to $\frac{1}{d^o(v_i)}$
- $0 < \alpha < \frac{1}{\lambda}$; $\lambda = 1$ in undirected graphs

PageRank Centrality: Example

We assume $\alpha = 0.95$ and $\beta = 0.1$



$$A = \begin{array}{c c} & \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array} & \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

$$C^p = \beta(I - \alpha A^T D^{-1})^{-1} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2.14 \\ 2.13 \\ 2.14 \\ 1.45 \\ 2.13 \end{bmatrix}$$

PAGERANK CENTRALITY APPLICATION

PageRank centrality is often applied when you want to take link direction and weight into account:

- understand citations (e.g. patent citations, academic citations)
- visualize network activity
- identify leadership groups on social networks
- large graphs

Centrality Measures Comparison

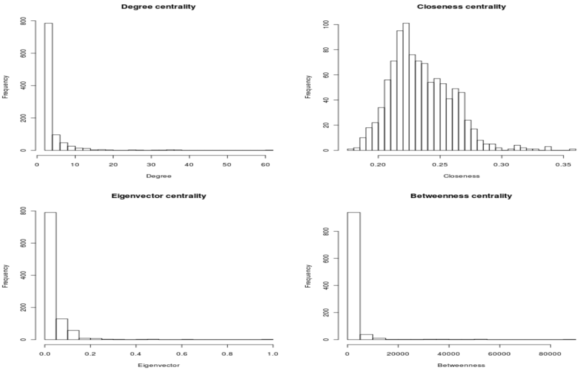
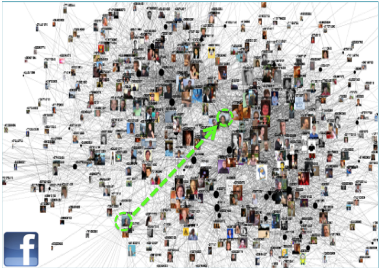


Table of contents — SL03

1. Centrality

2. Similarity
   Introduction
   Measures

Similarity on Social Networks



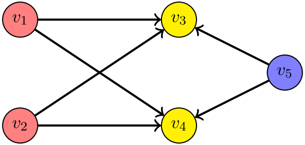
- For each network, how much similar are the persons in green?
- How can we quantify the similarity?
- Which persons on each network are most similar to one another?

SIMILARITY VS. CENTRALITY

- The equivalence is used to measure the similarity: two nodes (or other more elaborate structures) are similar if they fall in the same "equivalence class".
- The structural equivalence is often applied to quantify the equivalence.

Structural Equivalence: Definition

- Two nodes are structurally equivalent if they share many of the same network neighbors.
- For example, two students attending the Social Media Analytics course have in common teachers, colleagues and could be considered similar. Two random individuals will have less in common and can be considered dissimilar.
- The structurally equivalent nodes have been colored with the same color.

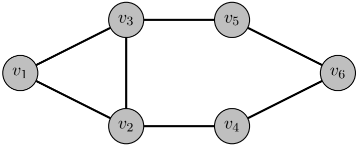


Structural Equivalence: Measures

- Given $N(v_i)$ that returns the neighbors of v_i, the structural equivalence can be computed using the following similarity measures:
  - Vertex similarity
  - Jaccard similarity
  - Cosine similarity
  - Adamic-Adar similarity
- The definition of $N(v_i)$ excludes the node itself (v_i).

Vertex Similarity

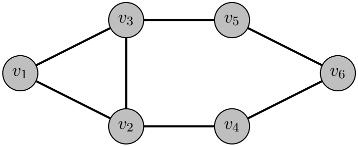
- **Idea**: This measure computes the number of neighbors that two nodes v_i and v_j have in common.
- **Interpretation**: Two FB persons are similar when they have lots of shared friends.
- **Formula**: $\forall i \neq j : \sigma(v_i, v_j) = \left|N(v_i) \cap N(v_j)\right|$
- **Example**:



  - $\sigma(v_3, v_6) = \left|\{v_1, v_2, v_5\} \cap \{v_4, v_5\}\right| = 1$

JACCARD SIMILARITY

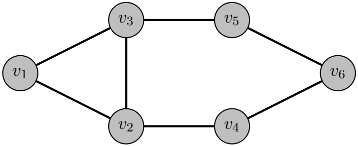
- **Idea**: This measure computes the probability that both v_i and v_j have a neighbor v_k, for a randomly selected neighbor v_k that either v_i or v_j has.
- **Interpretation**: Two FB persons are similar when they have lots of shared friends, and are close to each other.
- **Formula**: $\forall i \neq j : \sigma_{Jac}(v_i, v_j) = \frac{\left|N(v_i) \cap N(v_j)\right|}{\left|N(v_i) \cup N(v_j)\right|}$
- **Example**:



  - $\sigma_{Jac}(v_3, v_6) = \frac{\left|\{v_1, v_2, v_5\} \cap \{v_4, v_5\}\right|}{\left|\{v_1, v_2, v_4, v_5\}\right|} = \frac{1}{4} = 0.25$

Cosine Similarity

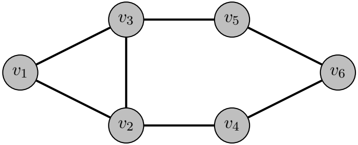
- **Idea**: This measure returns how well connected are two nodes v_i and v_j.
- **Interpretation**: Two FB friends are similar when they have lots of shared friends and few non shared ones.
- **Formula**: $\forall i \neq j : \sigma_{Cos}(v_i, v_j) = \frac{\left|N(v_i) \cap N(v_j)\right|}{\sqrt{\left|N(v_i)\right| \cdot \left|N(v_j)\right|}}$
- **Example**:



- $\sigma_{Cos}(v_3, v_6) = \frac{\left|\{v_1, v_2, v_5\} \cap \{v_4, v_5\}\right|}{\sqrt{\left|\{v_1, v_2, v_5\}\right| \left|\{v_4, v_5\}\right|}} = \frac{1}{2.44} = 0.4$

Adamic-Adar Similarity

- **idea**: This measure refines the simple counting of vertex similarity by weighting less-connected neighbors more heavily.
- **Interpretation**: Two FB persons who are friends with a famous person are probably less similar than those who are friends with a less famous person.
- **Formula**: $\forall i \neq j : \sigma_{AA}(v_i, v_j) = \sum_{z \in N(v_i) \cap N(v_j)} \frac{1}{\log \left| N(z) \right|}$
- **Example**:

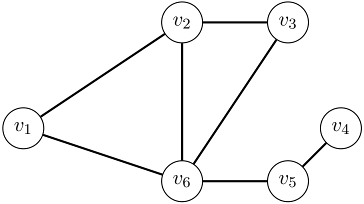


  - $\sigma_{AA}(v_3, v_6) = \frac{1}{\log \left| N(v_5) \right|} = \frac{1}{\log 2} = 3.32$

Exercise: Similarity measures

In the graph below, compute the similarity matrix using each of the following similarity measures:

- Vertex similarity
- Jaccard similarity
- Cosine similarity
- Adamic-Adar similarity



Exercise: Vertex Similarity matrix

EXERCISE: JACCARD SIMILARITY MATRIX

Exercise: Cosine Similarity matrix

Exercise: Adamic-Adar Similarity matrix