# Friendship Patterns

- Transitivity/Reciprocity
- Status/Balance

# I. Transitivity and Reciprocity

# Transitivity

- Mathematic representation:
  - For a transitive relation $R$: $aRb \wedge bRc \rightarrow aRc$

- In a social network:
  - ***Transitivity is when a friend of my friend is my friend***
  - Transitivity in a social network leads to a denser graph, which in turn is closer to a complete graph
  - We can determine how close graphs are to the complete graph by measuring transitivity

# [Global] Clustering Coefficient

- **Clustering coefficient** measures transitivity in **undirected** graphs
  - Count paths of length two and check whether the third edge exists

$$C = \frac{|\text{Closed Paths of Length 2}|}{|\text{Paths of Length 2}|}$$

When counting triangles, since every triangle has **6** closed paths of length 2

$$C = \frac{(\text{Number of Triangles}) \times 6}{|\text{Paths of Length 2}|}$$

# Clustering Coefficient and Triples

Or we can rewrite it as

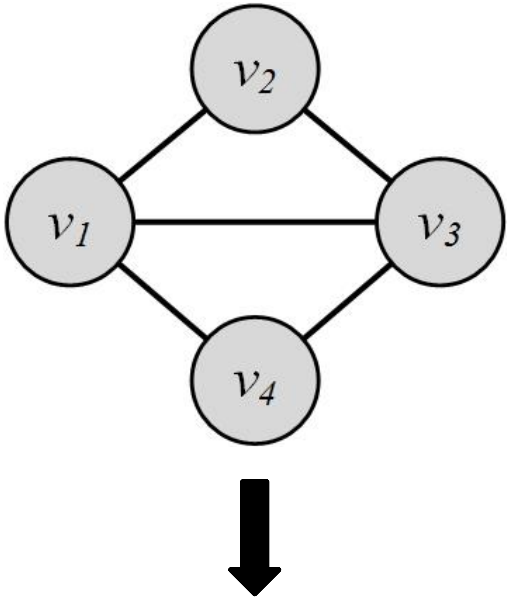
$$C = \frac{(\text{Number of Triangles}) \times 3}{\text{Number of Connected Triples of Nodes}}$$

- **Triple**: an **ordered** set of three nodes,
  - connected by two (open triple) edges or
  - three edges (closed triple)
- A triangle can miss any of its three edges
  - A triangle has **3 Triples**

$v_i v_j v_k$ and $v_j v_k v_i$ are different triples
- The same members
- First missing edge $e(v_k, v_i)$ and second missing $e(v_i, v_j)$

$v_i v_j v_k$ and $v_k v_j v_i$ are the same triple

# [Global] Clustering Coefficient: Example




$$
\begin{aligned}
C &= \frac{(\text{Number of Triangles}) \times 3}{\text{Number of Connected Triples of Nodes}} \\
&= \frac{2 \times 3}{2 \times 3 + \underbrace{2}_{v_2 v_1 v_4,\, v_2 v_3 v_4}} = 0.75.
\end{aligned}
$$

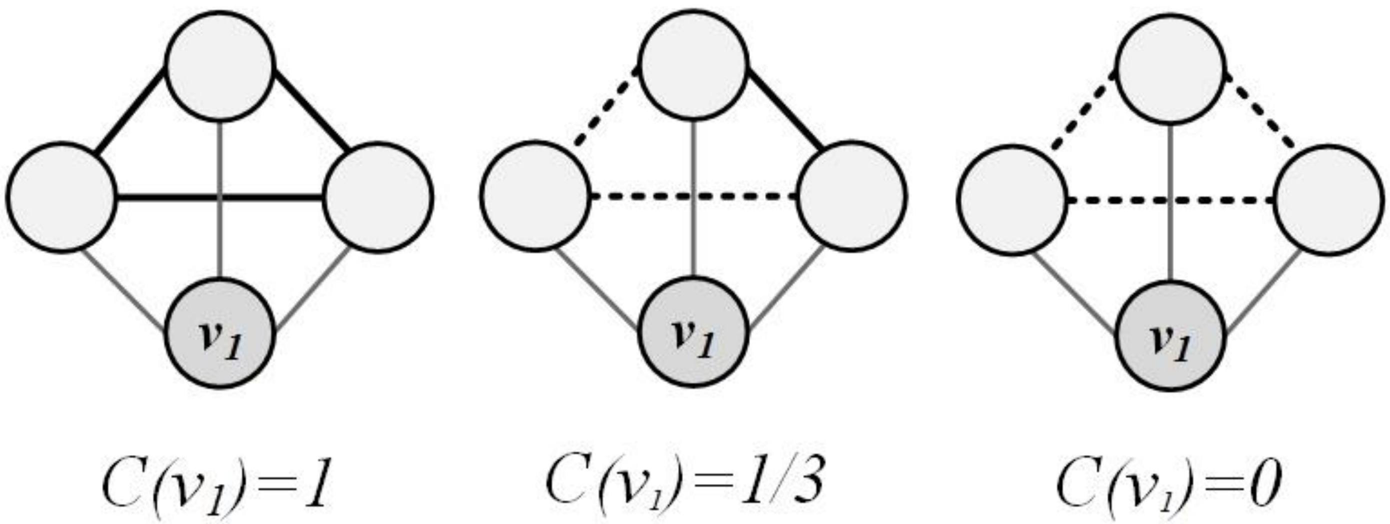
# Local Clustering Coefficient

- Local clustering coefficient measures transitivity at the node level
  - Commonly employed for undirected graphs
  - Computes how strongly neighbors of a node $v$ (nodes adjacent to $v$) are themselves connected

$$C(v_i) = \frac{\text{Number of Pairs of Neighbors of } v_i \text{ That Are Connected}}{\text{Number of Pairs of Neighbors of } v_i}.$$

In an **undirected** graph, the denominator can be rewritten as: $\binom{d_i}{2} = d_i(d_i - 1)/2$

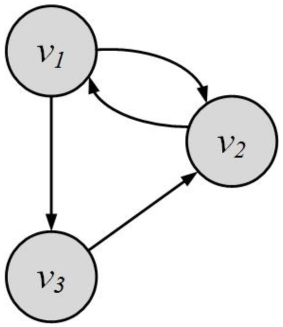
# Local Clustering Coefficient: Example




- Thin lines depict connections to neighbors
- Dashed lines are the missing link among neighbors
- Solid lines indicate connected neighbors
  - When none of neighbors are connected $C = 0$
  - When all neighbors are connected $C = 1$

# Reciprocity

***If you become my friend, I'll be yours***

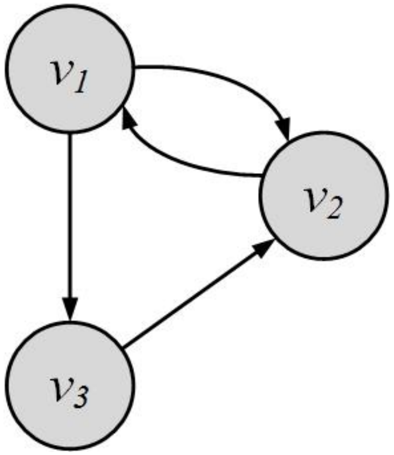



- Reciprocity is simplified version of transitivity
  - It considers closed loops of length 2
- If node $v$ is connected to node $u$,
  - $u$ by connecting to $v$, exhibits reciprocity

$$
\begin{aligned}
R &= \frac{\sum_{i,j,i<j} A_{i,j}A_{j,i}}{|E|/2},\\
&= \frac{2}{|E|}\sum_{i,j,i<j} A_{i,j}A_{j,i},\\
&= \frac{2}{|E|}\times\frac{1}{2}\mathrm{Tr}(A^2),\\
&= \frac{1}{|E|}\mathrm{Tr}(A^2),\\
&= \frac{1}{m}\mathrm{Tr}(A^2).
\end{aligned}
$$

$$\boxed{\mathrm{Tr}(A) = A_{1,1} + A_{2,2} + \cdots + A_{n,n} = \sum_{i=1}^{n} A_{i,i}}$$

# Reciprocity: Example



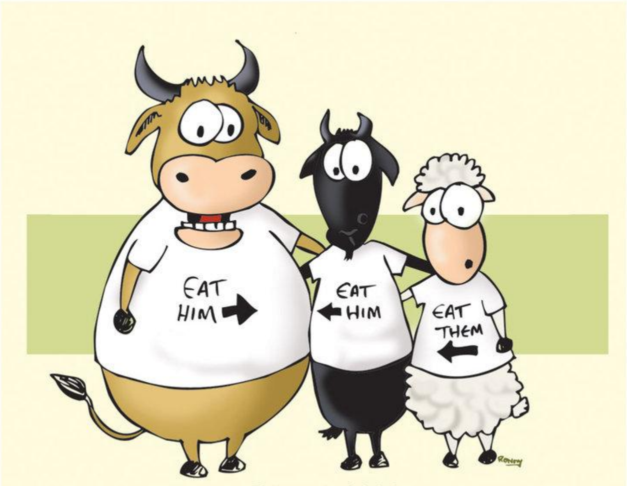

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Reciprocal nodes: $v_1, v_2$

$$R = \frac{1}{m}\mathrm{Tr}(A^2) = \frac{1}{4}\mathrm{Tr}\left(\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}\right) = \frac{2}{4} = \frac{1}{2}.$$

# II. Balance and Status

- Measuring consistency in friendships



# Social Balance Theory

## Social balance theory

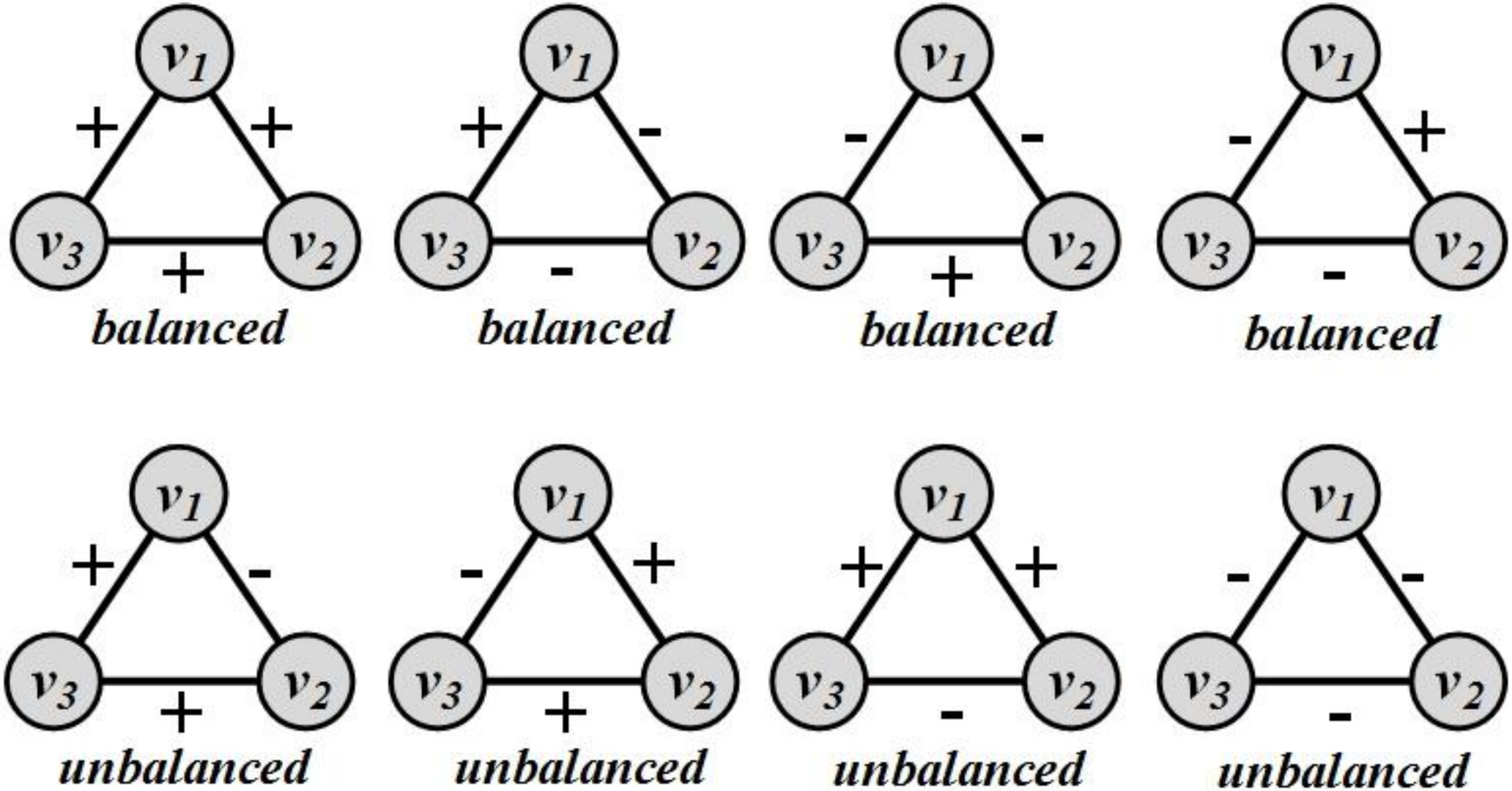
- Consistency in friend/foe relationships among individuals
- Informally, friend/foe relationships are consistent when

$$\begin{aligned}&\textit{The friend of my friend is my friend,}\\&\textit{The friend of my enemy is my enemy,}\\&\textit{The enemy of my enemy is my friend,}\\&\textit{The enemy of my friend is my enemy.}\end{aligned}$$

- In the network
  - Positive edges demonstrate friendships ($w_{ij} = 1$)
  - Negative edges demonstrate being enemies ($w_{ij} = -1$)
- Triangle of nodes $i, j$, and $k$, is balanced, if and only if
  - $w_{ij}$ denotes the value of the edge between nodes $i$ and $j$

$$w_{ij} w_{jk} w_{ki} \geq 0.$$

# Social Balance Theory: Possible Combinations




For any cycle, if the multiplication of edge values become positive, then the cycle is socially balanced

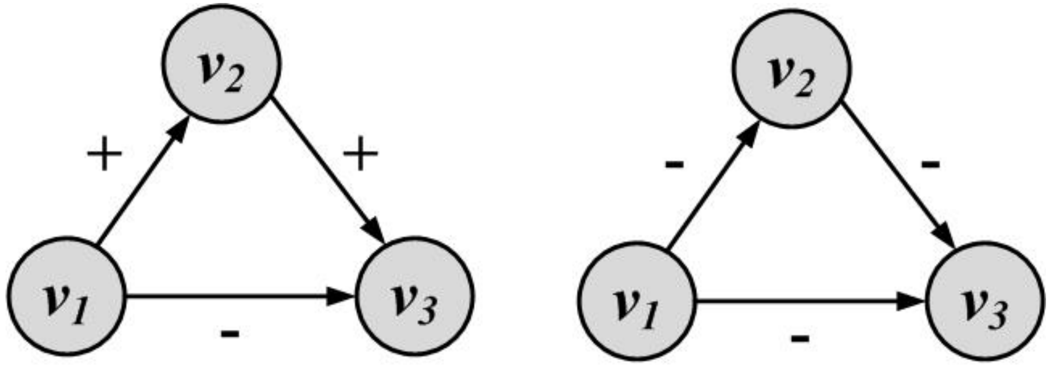
# Social Status Theory

- **Status:** how prestigious an individual is ranked within a society

- **Social status theory:**
  - How consistent individuals are in assigning status to their neighbors
  - Informally,

*If $X$ has a higher status than $Y$ and $Y$ has a higher status than $Z$, then $X$ should have a higher status than $Z$.*

# Social Status Theory: Example




Unstable configuration Stable configuration

- A directed **'+'** edge from node $X$ to node $Y$ shows that $Y$ has a higher status than $X$ and a **'-'** one shows vice versa

# Similarity

How similar are two nodes in a network?

- **Structural Equivalence**
- **Regular Equivalence**

# Structural Equivalence

- **Structural Equivalence:**
  - We look at the neighborhood <u>shared</u> by two nodes;
  - The size of this shared neighborhood defines how similar two nodes are.

- ***Example:***
  - *Two brothers have in common*
    - *sisters, mother, father, grandparents, etc.*
  - *This shows that they are similar,*
  - *Two random male or female individuals do not have much in common and are dissimilar.*

# Structural Equivalence: Definitions

- **Vertex similarity**: $\sigma(v_i, v_j) = |N(v_i) \cap N(v_j)|$

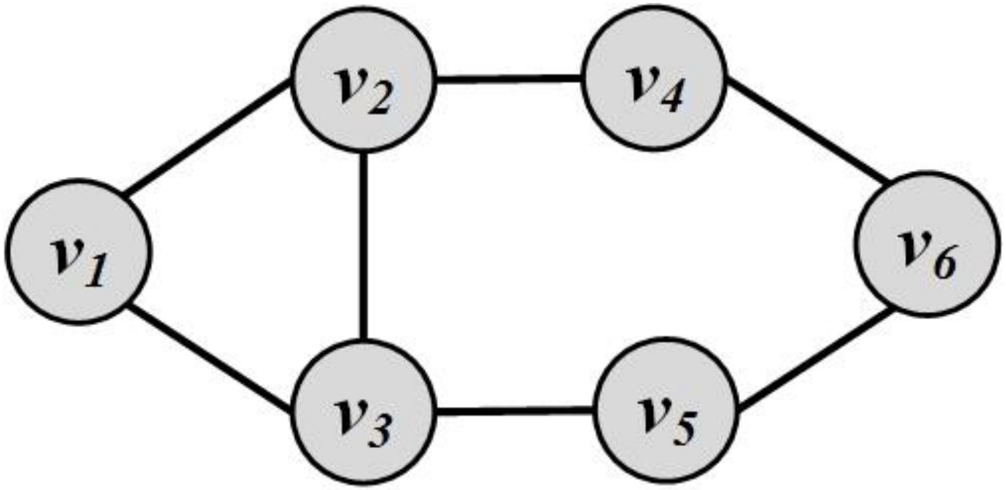
**Normalize?**

**Jaccard Similarity**: $\sigma_{Jaccard}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)|}$

**Cosine Similarity**: $\sigma_{Cosine}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{\sqrt{|N(v_i)||N(v_j)|}}$

- The neighborhood $N(v)$ often excludes the node itself $v$.
  - **What can go wrong?**
    - Connected nodes not sharing a neighbor will be assigned zero similarity
  - **Solution:**
    - We can assume nodes are included in their neighborhoods

# Similarity: Example




$$\sigma_{\text{Jaccard}}(v_2, v_5) = \frac{|\{v_1, v_3, v_4\} \cap \{v_3, v_6\}|}{|\{v_1, v_3, v_4, v_6\}|} = 0.25$$

$$\sigma_{\text{Cosine}}(v_2, v_5) = \frac{|\{v_1, v_3, v_4\} \cap \{v_3, v_6\}|}{\sqrt{|\{v_1, v_3, v_4\}||\{v_3, v_6\}|}} = 0.40.$$

# Similarity Significance

**Measuring Similarity Significance:** compare the calculated similarity value with its expected value where vertices pick their neighbors at random

- For vertices $v_i$ and $v_j$ with degrees $d_i$ and $d_j$ this expectation is $d_i d_j / n$
  - There is a $d_i/n$ chance of becoming $v_i$'s neighbor
  - $v_j$ selects $d_j$ neighbors

- We can rewrite neighborhood overlap as

$$\sigma(v_i, v_j) = |N(v_i) \cap N(v_j)| = \sum_k A_{i,k} A_{j,k}$$

# Normalized Similarity, cont.

$$\bar{A}_i = \frac{1}{n}\sum_k A_{i,k}$$

$$\begin{aligned}
\sigma_{\text{significance}}(v_i, v_j) &= \sum_k A_{i,k}A_{j,k} - \frac{d_i d_j}{n} \\
&= \sum_k A_{i,k}A_{j,k} - n\frac{1}{n}\sum_k A_{i,k}\frac{1}{n}\sum_k A_{j,k} \\
&= \sum_k A_{i,k}A_{j,k} - n\bar{A}_i\bar{A}_j \\
&= \sum_k (A_{i,k}A_{j,k} - \bar{A}_i\bar{A}_j) \\
&= \sum_k (A_{i,k}A_{j,k} - \bar{A}_i\bar{A}_j - \bar{A}_i\bar{A}_j + \bar{A}_i\bar{A}_j) \\
&= \sum_k (A_{i,k}A_{j,k} - A_{i,k}\bar{A}_j - \bar{A}_i A_{j,k} + \bar{A}_i\bar{A}_j) \\
&= \boxed{\sum_k (A_{i,k} - \bar{A}_i)(A_{j,k} - \bar{A}_j)}
\end{aligned}$$

**What is this?**

# Normalized Similarity, cont.

**$n$ times the Covariance between $A_i$ and $A_j$**

$$\frac{1}{n}\sum_k (A_{i,k} - \bar{A}_i)(A_{j,k} - \bar{A}_j)$$

Normalize covariance by the multiplication of Variances.

$$\sqrt{\frac{1}{n}\sum_k (A_{i,k} - \bar{A}_i)^2}\sqrt{\frac{1}{n}\sum_k (A_{j,k} - \bar{A}_j)^2}$$

We get **Pearson correlation coefficient**

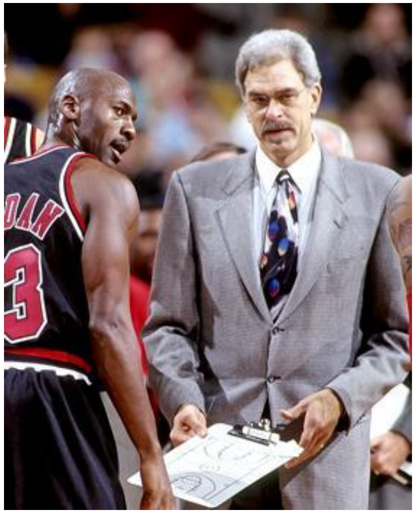
$$\begin{aligned}\sigma_{\text{pearson}}(v_i, v_j) &= \frac{\sigma_{\text{significance}}(v_i, v_j)}{\sqrt{\sum_k (A_{i,k} - \bar{A}_i)^2}\sqrt{\sum_k (A_{j,k} - \bar{A}_j)^2}} \\ &= \frac{\sum_k (A_{i,k} - \bar{A}_i)(A_{j,k} - \bar{A}_j),}{\sqrt{\sum_k (A_{i,k} - \bar{A}_i)^2}\sqrt{\sum_k (A_{j,k} - \bar{A}_j)^2}}\end{aligned}$$

(range of $\sigma \in [-1,1]$ )

# Regular Equivalence

- In regular equivalence,
  - We **do not** look at neighborhoods shared between individuals, but
  - **<u>How neighborhoods themselves are similar</u>**

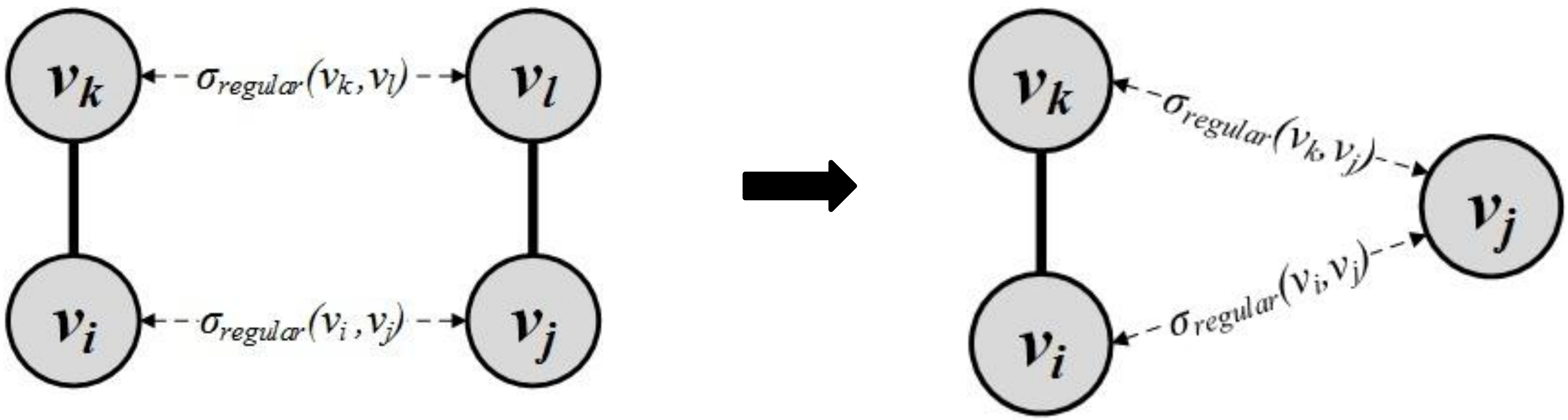
- ***<u>Example</u>*:**
  - ***Athletes are similar not because they know each other in person, but since they know similar individuals, such as coaches, trainers, other players, etc.***



# Regular Equivalence

- $v_i$, $v_j$ are similar when their neighbors $v_k$ and $v_l$ are similar

$$\sigma_{\text{regular}}(v_i, v_j) = \alpha \sum_{k,l} A_{i,k} A_{j,l} \sigma_{\text{regular}}(v_k, v_l)$$




- The equation (left figure) is hard to solve since it is self referential so we relax our definition using the right figure

# Regular Equivalence

- $v_i$ and $v_j$ are similar when $v_j$ is similar to $v_i$'s neighbors $v_k$

$$\sigma_{regular}(v_i, v_j) = \alpha \sum_k A_{i,k} \sigma_{Regular}(v_k, v_j)$$

- In vector format

$$\sigma_{regular} = \alpha A \sigma_{Regular}$$

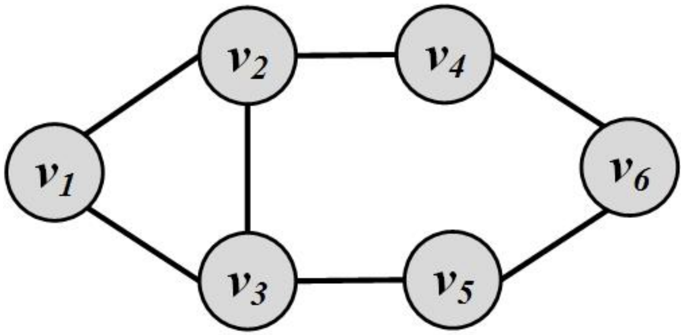
A vertex is highly similar to itself, we guarantee this by adding an identity matrix to the equation

$$\sigma_{regular} = \alpha A \sigma_{Regular} + \mathbf{I}$$

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

**When $\alpha < 1/\lambda_{max}$ the matrix is invertible**

# Regular Equivalence: Example




$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

The largest eigenvalue of $A$ is 2.43

Set $\alpha = 0.3 < 1/2.43$

$$\sigma_{\text{regular}} = (I - 0.3A)^{-1} = \begin{bmatrix} 1.43 & 0.73 & 0.73 & 0.26 & 0.26 & 0.16 \\ 0.73 & 1.63 & 0.80 & 0.56 & 0.32 & 0.26 \\ 0.73 & 0.80 & 1.63 & 0.32 & 0.56 & 0.26 \\ 0.26 & 0.56 & 0.32 & 1.31 & 0.23 & 0.46 \\ 0.26 & 0.32 & 0.56 & 0.23 & 1.31 & 0.46 \\ 0.16 & 0.26 & 0.26 & 0.46 & 0.46 & 1.27 \end{bmatrix}$$

- Any row/column of this matrix shows the similarity to other vertices
- Vertex 1 is most similar (other than itself) to vertices 2 and 3
- Nodes 2 and 3 have the highest similarity (**regular equivalence**)