SIR model:

- In addition to the **I** and **S** states, a <u>recovery state **R**</u> is present

- Individuals get infected and some recover

- Once hosts recover (or are removed) they can no longer get infected and are not susceptible

$$I \; + \; S \; + \; R \; = \; N$$

$$\frac{dS}{dt} \; = \; -\beta IS,$$

$$\frac{dI}{dt} \; = \; \beta IS - \gamma I,$$

$$\frac{dR}{dt} \; = \; \gamma I.$$

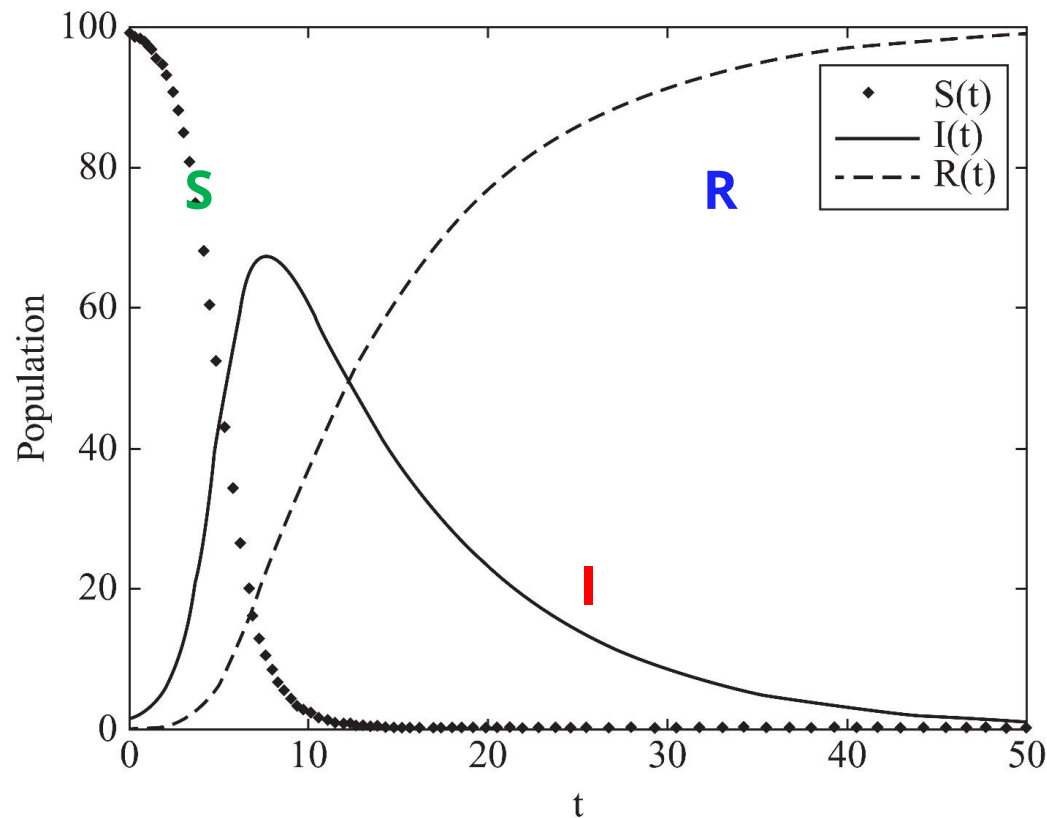**γ** defines the recovering probability of an infected individual at a time stamp

$$\frac{dS}{dt} = -\beta IS,$$
$$\frac{dI}{dt} = \beta IS - \gamma I,$$
$$\frac{dR}{dt} = \gamma I.$$

➡️ $$\frac{dS}{dR} = -\frac{\beta}{\gamma}S$$

➡️ $$\log \frac{S_0}{S} = \frac{\beta}{\gamma}R \qquad (R_0 = 0)$$

$$\frac{dR}{dt} = \gamma(N - S - R)$$

$$\frac{dR}{dt} = \gamma(N - S_0 e^{-\frac{\beta}{\gamma}R} - R)$$

➡️ $$t = \frac{1}{\gamma}\int_0^R \frac{dx}{N - S_0 e^{-\frac{\beta}{\gamma}x} - x}$$

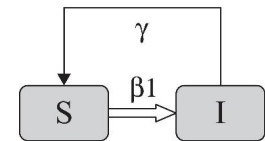There is no closed form solution for this integration and only numerical approximation is possible.

SIR Model: Example

SIR model simulated with $S_0 = 99$, $I_0 = 1$, $R_0 = 0$, $\beta = 0.01$, and $\gamma = 0.1$

- The **SIS** model is the same as the **SI** model with the addition of infected nodes recovering and becoming susceptible again

$$\frac{dS}{dt} \;=\; \gamma I - \beta IS, \qquad \frac{dI}{dt} \;=\; \beta IS - \gamma I$$

$$\frac{dI}{dt} \;=\; \beta I(N - I) - \gamma I = I(\beta N - \gamma) - \beta I^2$$

$$\frac{dI}{dt} = \beta I(N - I) - \gamma I = I(\beta N - \gamma) - \beta I^2$$
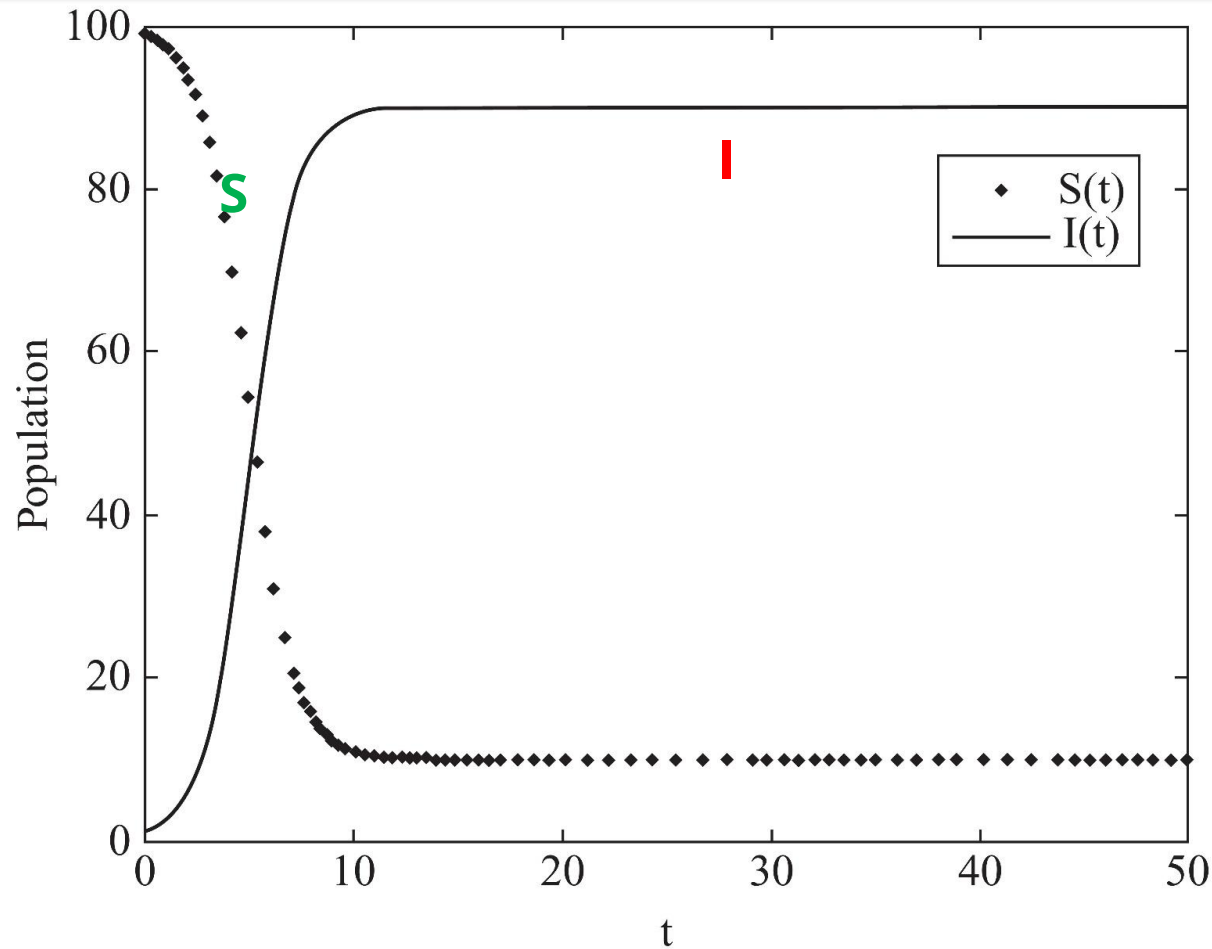
Case 1: When $\beta N \leq \gamma$ (or when $N \leq \frac{\gamma}{\beta}$):

- The first term will be at most zero or negative
- The whole term becomes negative
- In the limit, I(t) will decrease exponentially to zero

Case 2: When $\beta N > \gamma$ (or when $N > \frac{\gamma}{\beta}$):

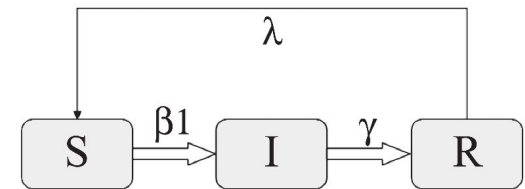- We will have a logistic growth function like the **SI** model

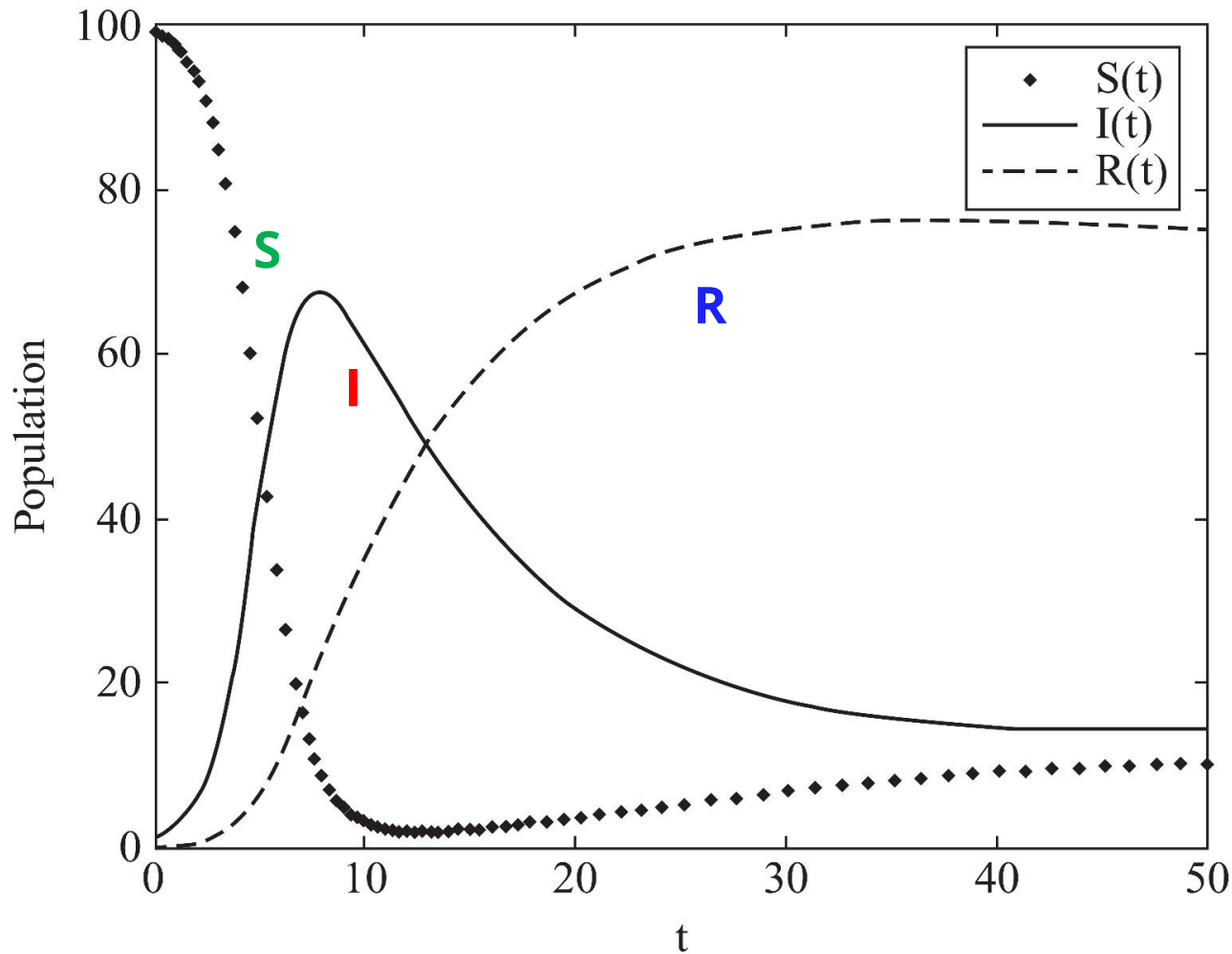**SIS** model simulated with $S_0 = 99$, $I_0 = 1$, $\beta = 0.01$, and $\gamma = 0.1$

The individuals who have recovered will lose immunity after a certain period of time and will become susceptible again

$$\frac{dS}{dt} = \lambda R - \beta IS,$$

$$\frac{dI}{dt} = \beta IS - \gamma I,$$

$$\frac{dR}{dt} = \gamma I - \lambda R.$$

Like the SIR, model this model has no closed form solution, so numerical integration can be used

SIRS model simulated with $S_0 = 99$, $I_0 = 1$, $R_0 = 1$, $\beta = 0.01$, $\lambda = 0.02$, and $\gamma = 0.1$

- Suppose that we have a susceptible society and want to prevent more spread by vaccinating the most vulnerable individuals

- What is the best way to vaccinate a population?

Randomly pick some nodes and ask them who is the most popular from their point of view, then vaccinate those individuals!

Epidemic Intervention: *Mad Cow Disease*

- Jan. 2001
  - First case observed in UK
- Feb. 2001
  - 43 farms infected
- Sep. 2001
  - 9,000 farms infected
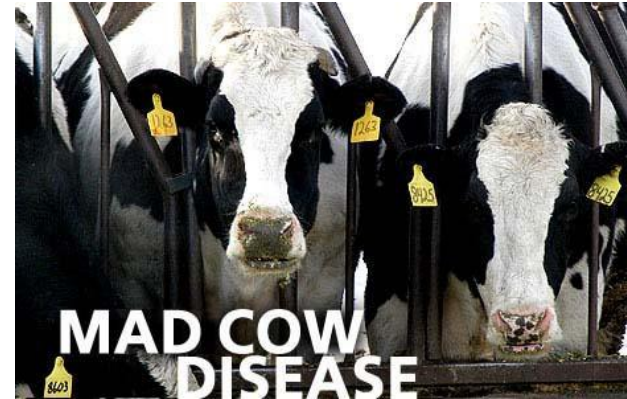
In the mad cow disease case, we have weak ties
- Animals being bought and sold
- Soil from tourists, etc.

How to stop the disease:
- Banned movement (make contagion spread more difficult, or quarantine)
- Killed millions of animals (remove weak ties)

CS 579: Online Social Network Analysis

Influence and Homophily

Kai Shu

Spring 2023

Social Forces

- **Social Forces** connect individuals in different ways

- When individuals get connected, we observe distinguishable patterns in their connectivity networks.
  - **Assortativity**, also known as *social similarity*

- In networks with assortativity:
  - Similar nodes are connected to one another more often than dissimilar nodes.
- Social networks are *assortative*
  - A high similarity between friends is observed
  - We observe similar behavior, interests, activities, or shared attributes such as language among friends

Why are connected people similar?

Influence

- The process by which a user (i.e., influential) affects another user
- The influenced user becomes more similar to the influential figure.
  - **Example:** If most of our friends/family members switch to a cellphone company, we might switch [i.e., become influenced] too.
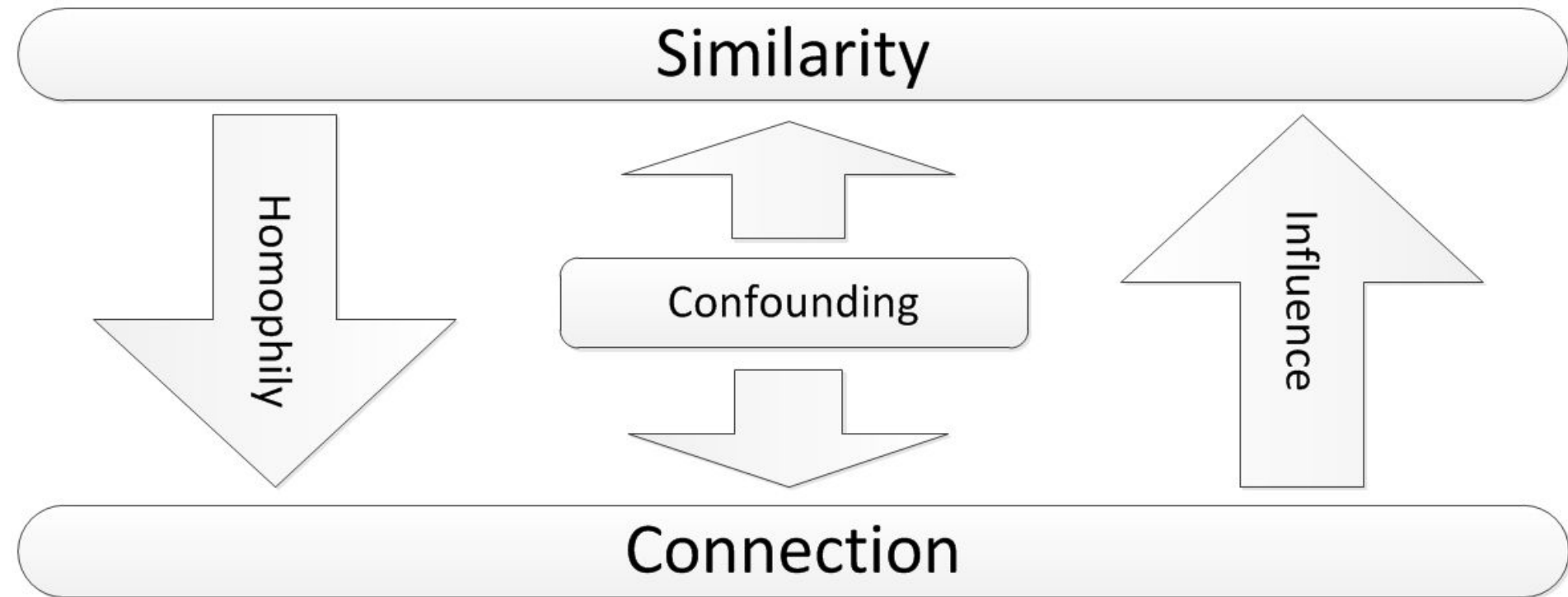
Homophily

- Similar individuals becoming friends due to their high similarity
  - **Example:** Two musicians are more likely to become friends.

Birds of a Feather Flock Together

Confounding

- The environment's effect on making individuals similar
  - **Example:** Two individuals living in the same town are more likely to become friends than two random individuals

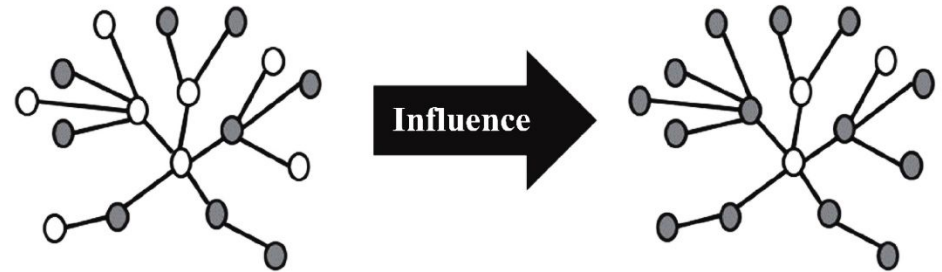# Influence, Homophily, and Confounding

Source of Assortativity in Networks

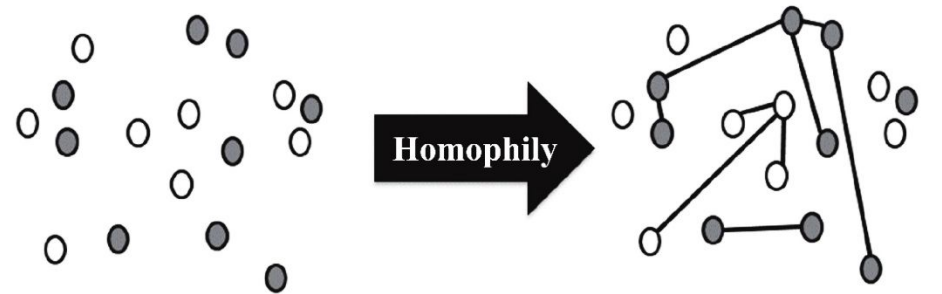Both influence and Homophily generate similarity in social networks

Influence
> Makes connected nodes similar to each other

Homophily
> Selects similar nodes and links them together

Assortativity Example

The city's draft tobacco control strategy says more than 60% of under-16s in Plymouth smoke regularly

More than 60% of Plymouth's under-16s smoke

Why?

- Smoker friends influence their non-smoker friends

Influence

- Smokers become friends
  - Can this explain smoking behavior?

Homophily

- There are lots of places that people can smoke

Confounding

Our goal

1. How can we **measure assortativity**?

2. How can we **measure influence** or **homophily**?

3. How can we **model influence** or **homophily**?

4. How can we **distinguish between the two**?

Measuring Assortativity

Assortativity: An Example

- The friendship network in a US high school in 1994

- Colors represent races,
    - : Whites
    - **Grey**: Blacks
    - **Light Grey**: Hispanics
    - **Black**: others

- High assortativity between individuals of the same race

Measuring Assortativity for Nominal Attributes

- Assume **nominal** attributes are assigned to nodes
  – Example: race

- Edges between nodes of the same type can be used to measure assortativity of the network
  – Same type = nodes that share some attribute value(s)
  – Node attributes could be nationality, race, sex, etc.

$$\frac{1}{m} \sum_{(v_i,v_j) \in E} \delta(t(v_i), t(v_j)) = \frac{1}{2m} \sum_{ij} A_{ij} \delta(t(v_i), t(v_j))$$

$t(v_i)$ denotes type of vertex v_i

$$\delta(x, y) = \begin{cases} 0, & \text{if } x \neq y \\ 1, & \text{if } x = y \end{cases}$$

Kronecker delta function

Assortativity **Significance**

- **Assortativity significance**
  - The difference between **<u>measured assortativity</u>** and **<u>expected assortativity</u>**
  - The higher this difference, the more significant the assortativity observed

Example
  - In a school, 50% of the population is **White** and the other 50% is **Hispanic**.
  - We expect 50% of the connections to be between members of different races.
  - If all connections are between members of different races, then we have a significant finding

Assortativity **Significance**

Assortativity

Expected assortativity
(according to configuration model)

$$
Q = \boxed{\frac{1}{2m} \sum_{ij} A_{ij} \delta(t(v_i), t(v_j))} - \boxed{\frac{1}{2m} \sum_{ij} \frac{d_i d_j}{2m} \delta(t(v_i), t(v_j))}
$$

$$
= \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \delta(t(v_i), t(v_j)).
$$

This is **modularity**

The maximum happens when all vertices of the same type are connected to one another

$$Q_{\text{normalized}} = \frac{Q}{Q_{\text{max}}}$$

$$= \frac{\frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{d_i d_j}{2m}\right)\delta(t(v_i), t(v_j))}{\max\left[\frac{1}{2m}\sum_{ij}A_{ij}\delta(t(v_i), t(v_j)) - \frac{1}{2m}\sum_{ij}\frac{d_i d_j}{2m}\delta(t(v_i), t(v_j))\right]}$$

$$= \frac{\frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{d_i d_j}{2m}\right)\delta(t(v_i), t(v_j))}{\frac{1}{2m}2m - \frac{1}{2m}\sum_{ij}\frac{d_i d_j}{2m}\delta(t(v_i), t(v_j))}$$

$$= \frac{\sum_{ij}\left(A_{ij} - \frac{d_i d_j}{2m}\right)\delta(t(v_i), t(v_j))}{2m - \sum_{ij}\frac{d_i d_j}{2m}\delta(t(v_i), t(v_j))}$$

- Let $\Delta \in \mathbb{R}^{n \times k}$ denote the **indicator matrix** and let k denote the number of types

$$\Delta_{x,k} = \begin{cases} 1, & \text{if } t(x) = k; \\ 0, & \text{if } t(x) \neq k \end{cases}$$

- The **Kronecker delta** function can be reformulated using the indicator matrix

$$\delta(t(v_i), t(v_j)) = \sum_k \Delta_{v_i,k} \Delta_{v_j,k}$$

- Therefore,

$$(\Delta\Delta^T)_{i,j} = \delta(t(v_i), t(v_j))$$

Let Modularity matrix be

$$B = A - dd^T/2m$$
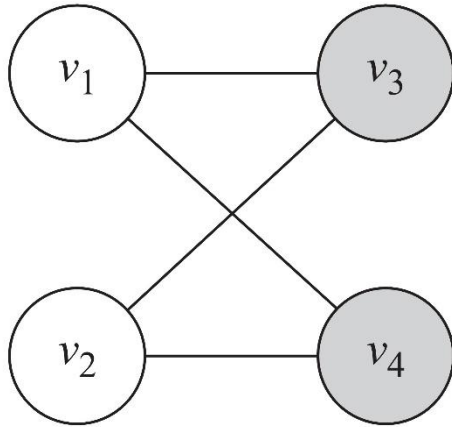
$d \in \mathbb{R}^{n \times 1}$ is the degree vector

Modularity can be reformulated as

$$Q = \frac{1}{2m} \sum_{ij} \underbrace{(\, A_{ij} - \frac{d_i d_j}{2m}\,)}_{B_{ij}} \underbrace{\delta(\, t(v_i), t(v_j)\,)}_{(\Delta \Delta^T)_{i,j}} = \frac{1}{2m} \mathrm{Tr}(B \Delta \Delta^T)$$

$$= \frac{1}{2m} \mathrm{Tr}(\Delta^T B \Delta)$$

Modularity Example

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}, \quad \Delta = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}, m = 4$$

$$B = A - \mathbf{d}\mathbf{d}^T/2m = \begin{bmatrix} -0.5 & -0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 & -0.5 \\ 0.5 & 0.5 & -0.5 & -0.5 \end{bmatrix}$$

$$Q = \frac{1}{2m}\mathrm{Tr}(\Delta^T B \Delta) = -0.5$$

The number of edges between nodes of the **same color** is less than the **expected** number of edges between them