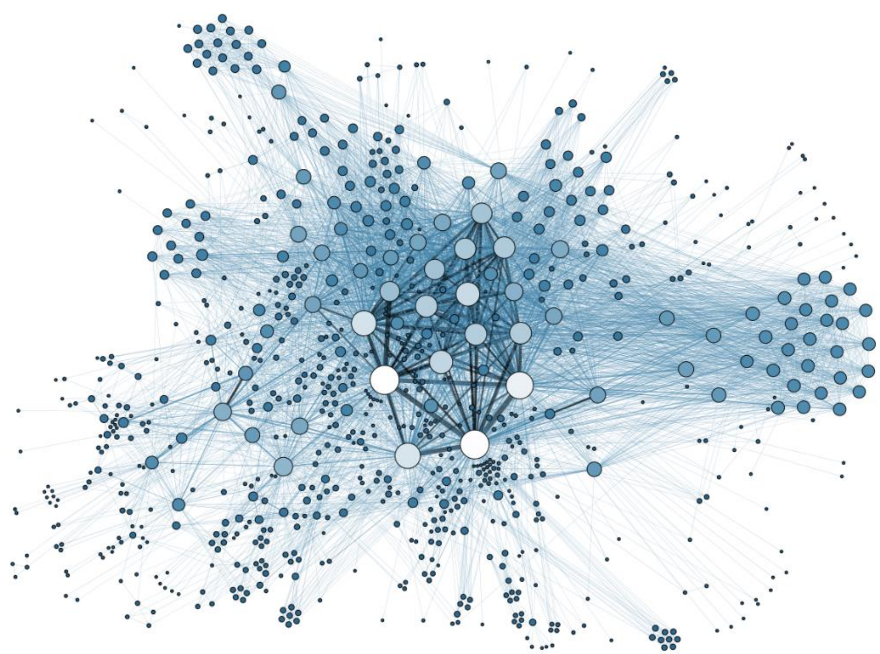
Social Media Analytics

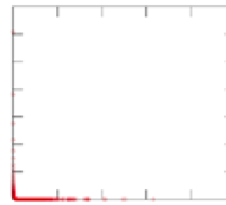
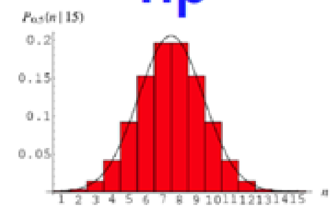
INFS7450

Influence and Homophily



Prof. Hongzhi Yin
School of EECS
The University of Queensland

Recap G_{np}

	MSN	G_{np}	n=180M
Degree distribution:			✗
Avg. path length:	**6.6**	$O(\log n)$ $h \approx 8.2$	✓
Avg. clustering coef.:	*0.11*	$\bar{k} / n$ $C \approx 8 \cdot 10^{-8}$	✗
Largest Conn. Comp.:	*99%*	GCC exists when $\bar{k}$>1. $\bar{k} \approx 14$.	✓

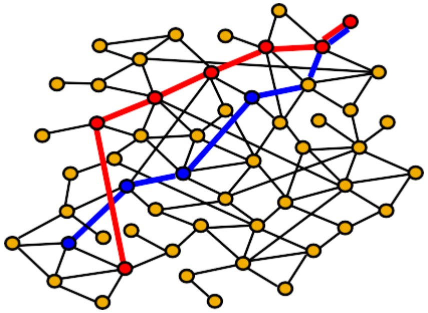
Note: the average degree of the random graph is equal to the average degree in MSN.

Small-World Model

Can we generate graphs with high clustering coefficients and short average paths

The Small-World Experiment

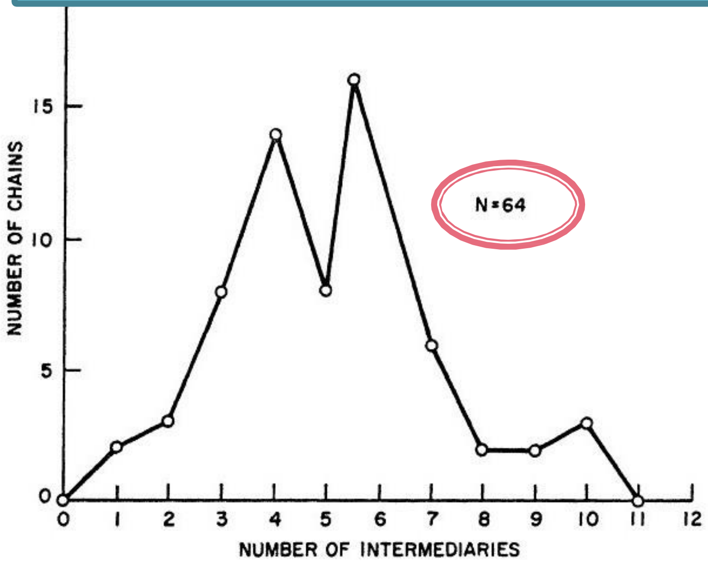
- **What is the typical shortest path length between any two people?**
  - **Experiment on the global friendship network**
    - Can't measure, need to probe explicitly
- **Small-world experiment** [Milgram '67]
  - Picked 296 people in Omaha and Boston
  - Ask them to send a letter to a stock-broker in Boston by passing it through friends
- **How many steps did it take?**



The Small-World Experiment

- **64 chains completed:**
  (i.e., 64 letters reached the target)
  - It took 6.2 steps on the average, thus **“6 degrees of separation”**
- **Further observations:**
  - People who owned stock had shorter paths to the stockbroker than random people: 5.4 vs. 6.7
  - People from the Boston area have even closer paths: 4.4

Milgram’s small world experiment



Clustering Implies Edges Locally

- **MSN network has 7 orders of magnitude larger clustering than the corresponding G_{np}!**
- **Other examples:**

Actor Collaborations (IMDB): N = 225,226 nodes, avg. degree $\bar{k}$ = 61
Electrical power grid: N = 4,941 nodes, $\bar{k}$ = 2.67
Network of neurons: N = 282 nodes, $\bar{k}$ = 14

Network	h_{actual}	h_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power Grid	18.70	12.40	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

h ... Average shortest path length
C ... Average clustering coefficient
"actual" … real network
"random" … random graph with same avg. degree

Regular Lattice

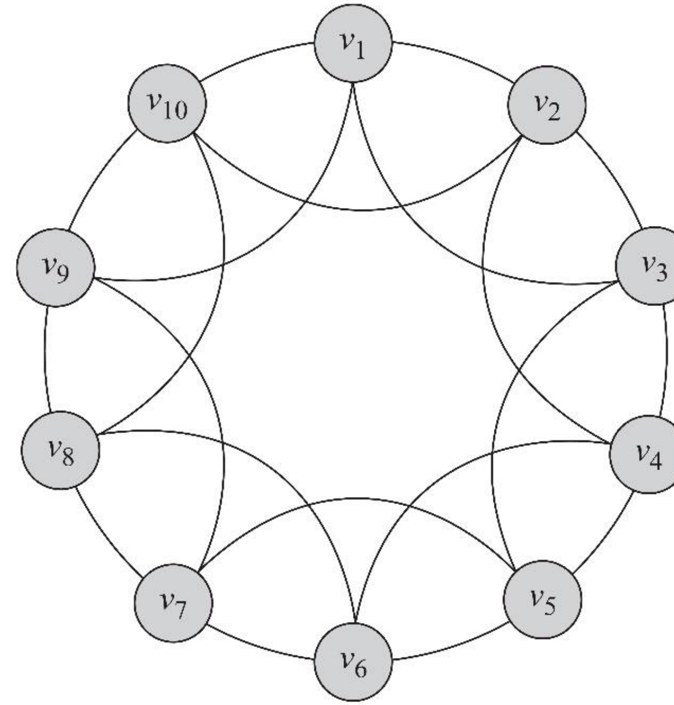
- In real-world interactions, many individuals have **a fixed number** of connections

- In graph theory terms, this assumption is equivalent to embedding users in a regular network

- A regular (ring) lattice is a special case of regular networks where there exists a certain pattern on how ordered nodes are connected to one another

- In a regular lattice of degree c, nodes are connected to their previous $c/2$ and following $c/2$ neighbors

- Formally, for node set V={v_1, v_2, v_3,…, v_n}, an edge exists between node i and j if and only if



$$0 \leq \min(n - |i-j|, |i-j|) \leq c/2$$

Regular Lattice vs. Random Graph

- Regular Lattice:
  - Clustering Coefficient (**high**):

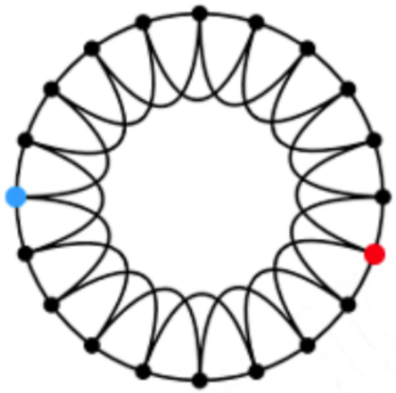
$$\frac{3(c-2)}{4(c-1)} \approx \frac{3}{4}$$

  - Diameter: 0.5 n/0.5 c=$\frac{n}{c}$
  - Average Shortest Path Length (**high**): $n/2c$

- Random Graph:
  - Clustering Coefficient (**low**): p
  - Average Path Length (**ok!**) : $\ln n \,/\, \ln c$

Small World-How

- **Could a network with high clustering also be a small world ($\log n$ dimeter)?**
  - How can we at the same time have **high clustering** and **small diameter?**



High clustering
High diameter

Low clustering
Low diameter

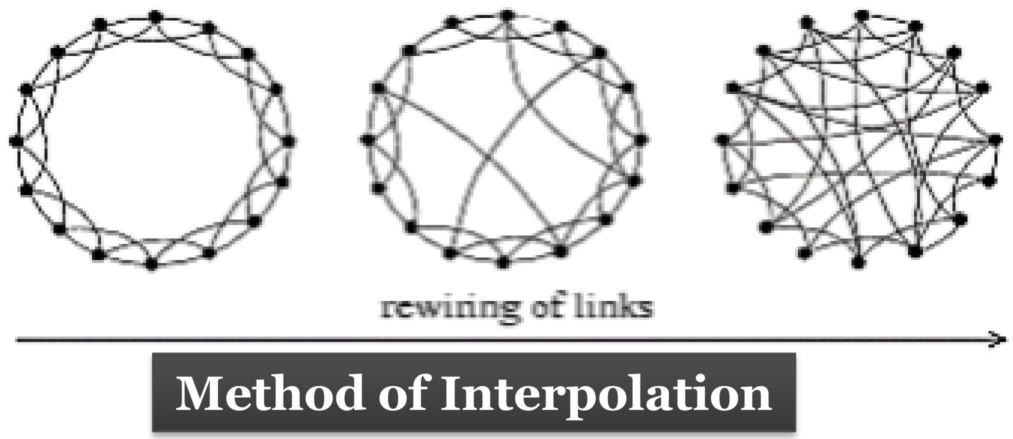
  - Clustering implies edge "locality"
  - Randomness enables "shortcuts"

Small-world Model

- Small-world model
  - A special type of random graph
  - Exhibits small-world properties:
    - Short average path length
    - High clustering coefficient

Generating a Small-World Graph

- The lattice has a **high**, but **fixed**, clustering coefficient
- The lattice has a **high** average path length



- In the small-world model, a parameter $0 \le \beta \le 1$ controls randomness in the model
  – When β is 0, the model is basically a regular lattice
  – When $\beta = 1$, the model becomes a random graph
- The model starts with a regular lattice and starts adding random edges [through **rewiring**]
  – **Rewiring:** take an edge, change one of its end-points randomly

Constructing Small World Networks

Algorithm 4.1 Small-World Generation Algorithm

Require: Number of nodes $|V|$, mean degree c, parameter β

```
return A small-world graph G(V, E)
G = A regular ring lattice with |V| nodes and degree c
for node v_i (starting from v_1), and all edges e(v_i, v_j), i < j do
    v_k = Select a node from V uniformly at random.
    if rewiring e(v_i, v_j) to e(v_i, v_k) does not create self- loops in the graph or
    multiple edges between v_i and v_k then
        rewire e(v_i, v_j) with probability β: E = E−{e(v_i, v_j)}, E = E∪{e(v_i, v_k)};
    end if
end for
Return G(V, E)
```

As in many network generating algorithms

- Disallow self-edges
- Disallow multiple edges

Small-World Summary

- **Could a network with high clustering be at the same time a small world?**
  - Yes! You need only a few random links
- **The Small World Model:**
  - Does not lead to the correct degree distribution
    - Most nodes have similar degrees due to the underlying lattice

How are real social networks formed?

Social Forces

- **Social Forces** connect individuals in different ways
- When individuals get connected, we observe interesting patterns in their connection networks.
  - **Assortativity**, also known as *social similarity*
- Social networks are assortative
  - A high similarity between friends is observed
  - We observe similar behavior, interests, activities, or shared attributes such as age, education background, even income 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.

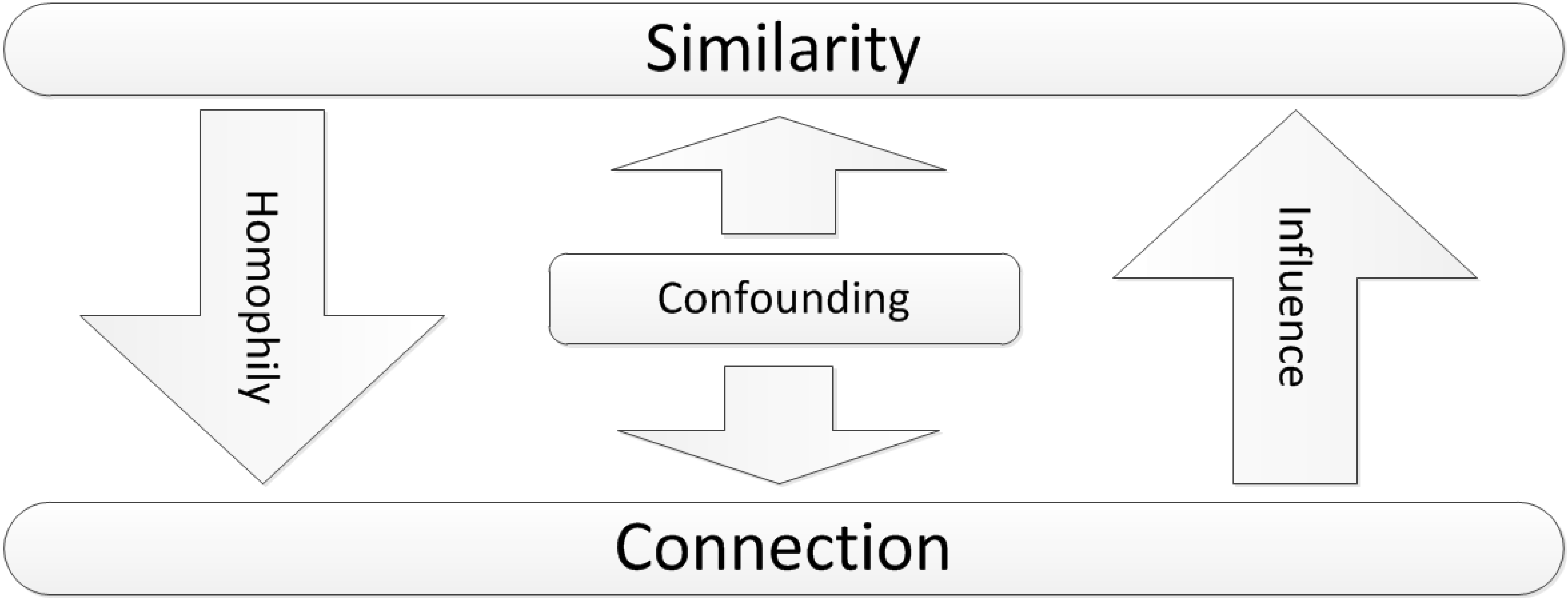
Homophily

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

Confounding Factor

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

Influence, Homophily, and Confounding



Cause and Effect

Source of Assortativity in Networks

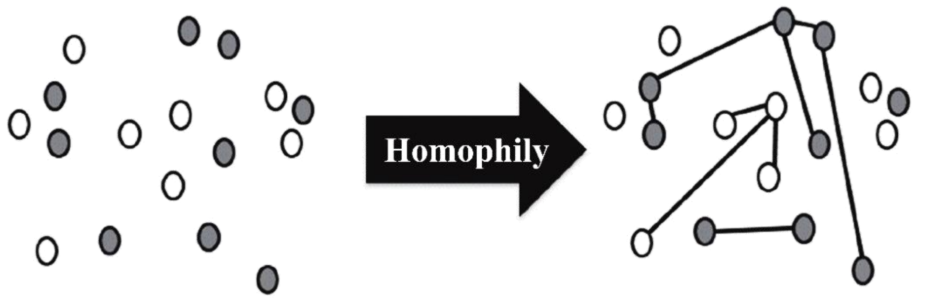
Both influence and homophily lead to social similarity in social networks

Influence

Makes connected nodes similar to each other

Homophily

Selects similar nodes to link together



Assortativity Example

More than 60% of under-16s in Plymouth regularly gather together to smoke

BBC | News | Sport | Weather | Travel | TV | Radio | More... | Search

DEVON

BBC RADIO DEVON
Listen Live   Listen Again

- BBC Local
- Devon
- Things to do
- People & Places
  - Nature & Outdoors
  - History
  - Religion & Ethics
  - Arts & Culture
  - BBC Introducing
- TV & Radio

Local BBC Sites
- News
- Sport
- Weather
- Travel

Neighbouring Sites
- Cornwall
- Dorset
- Somerset

Related BBC Sites
- England

Page last updated at 14:58 GMT, Monday, 14 June 2010 15:58 UK

E-mail this to a friend   Printable version

Patches for Plymouth's young smokers

By Jo Irving
BBC Devon website

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

MORE FROM DEVON
- NEWS
- SPORT
- WEATHER
- TRAVEL

ELSEWHERE ON THE WEB
- Plymouth NHS Trust Stop Smoking Service

Why?

- Smoker friends influence their non-smoker friends (social conformity) **Influence**

- Smokers are easier to become friends **Homophily**

- There are lots of places that people can smoke, and smoking is a type of local culture **Confounding**

Our goal?

1. How can we **measure assortativity**? (This lecture)

2. How can we **measure influence and homophily**? (Next lecture)

3. How can we **model influence and homophily?** (Next lecture)

Measuring Assortativity

Assortativity: An Example

- The friendship network in a US high school in 1994

- Colors represent races,
  - White: 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, gender

- The number of edges between nodes with the same attribute values can be used to measure assortativity of the network
  - 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 v_i's attribute value

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

Kronecker delta function

Assortativity Significance

- **Assortativity significance**
  - The difference between **measured assortativity of a real graph** and **expected assortativity from a special random graph model**
  - 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 (or same) races, then we have a significant finding

How to compute the expected assortativity

- The key is to compute the probability that any pair of nodes with degrees d_i and d_j could be connected.

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

- The link probability can be computed from a configuration model.

How to generate a Configuration model

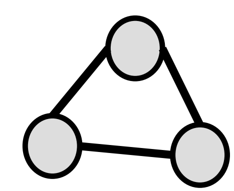
1. Create a list where each node v_i with degree d_i is repeated d_i times
2. Shuffle the list
3. Starting from the first index, join adjacent nodes

Example: Degree sequence $(2,2,2)$

v_1	v_1	v_2	v_2	v_3	v_3

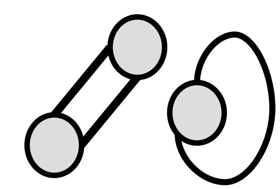
Random Shuffle 1:

v_1	v_2	v_2	v_3	v_3	v_1



Random Shuffle 2:

v_1	v_1	v_2	v_3	v_3	v_2



Properties of the Configuration Model

The probability that node v_i gets connected to node v_j is approximately

$$\frac{d_i d_j}{2m}$$

Proof:

In the shuffled list, for each v_i instance:

- There are d_j instances of v_j that it could be next to
- The probability of being next to v_j is $\frac{d_j}{2m-1}$

- There are d_i instances of v_i ; therefore, the total probability is $(d_i d_j)/(2m-1) \approx (d_i d_j)/2m$

Assortativity Significance

Assortativity

Expected assortativity
(according to configuration model)

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

This is modularity

The maximum happens when all adjacent nodes have the same types and non-adjacent nodes have different types

Matrix Trace

Let $\mathbf{A}$ be a matrix, with

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 3 \\ 11 & 5 & 2 \\ 6 & 12 & -5 \end{pmatrix}$$

Then

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{3} a_{ii} = a_{11} + a_{22} + a_{33} = -1 + 5 + (-5) = -1$$

Trace of a product [edit]

The trace of a square matrix which is the product of two matrices can be rewritten $\boxed{\text{as the sum of entry-wise products of their elements.}}$

$$\operatorname{tr}\left(\mathbf{A}^{\top}\mathbf{B}\right) = \operatorname{tr}\left(\mathbf{A}\mathbf{B}^{\top}\right) = \operatorname{tr}\left(\mathbf{B}^{\top}\mathbf{A}\right) = \operatorname{tr}\left(\mathbf{B}\mathbf{A}^{\top}\right) = \sum_{i,j} A_{ij}B_{ij}.$$

One-hot encoding/vector

id	X
1	a
2	c
3	a
4	b
5	a
6	c
7	c
8	b

One-Hot Encoding

id	X = a	X = b	X = c
1	1	0	0
2	0	0	1
3	1	0	0
4	0	1	0
5	1	0	0
6	0	0	1
7	0	0	1
8	0	1	0

Modularity: Matrix Form

- Let $\Delta \in \mathbb{R}^{n \times k}$ denote the **indicator matrix** and let k denote the number of attribute values; one hot vector

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

- 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,

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

Modularity: Matrix Form

$$B = A - dd^T / 2m$$

$\boldsymbol{d} \in \mathbb{R}^{n \times 1}$ is the degree vector

Modularity can be reformulated as

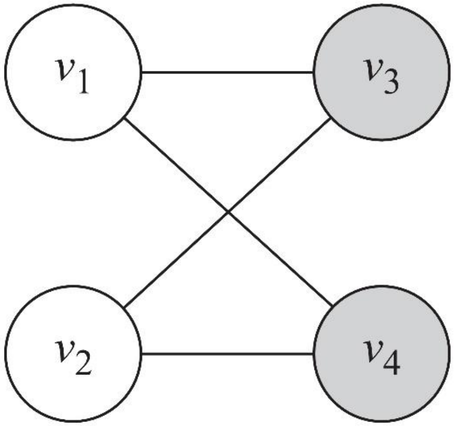
$$Q = \frac{1}{2m} \sum_{ij} \underbrace{\left(A_{ij} - \frac{d_i d_j}{2m} \right)}_{B_{ij}} \underbrace{\delta(\, t(v_i), t(v_j) \,)}_{(\Delta \Delta^T)_{i,j}} = \frac{1}{2m} \mathrm{Tr}(B \Delta \Delta^T)$$

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

Trace of a product [edit]

The trace of a square matrix which is the product of two matrices can be rewritten as the sum of entry-wise products of their elements.

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 (0 vs. 1)

Measuring Assortativity for Ordinal Attributes

- A common measure for analyzing the relationship between ordinal values is **covariance**
- It describes how two variables change together
- In our case, we have a network
  - We are interested in the correlation between attribute values assigned to nodes that are connected via edges
  - The correlation between attribute values of left nodes and right nodes

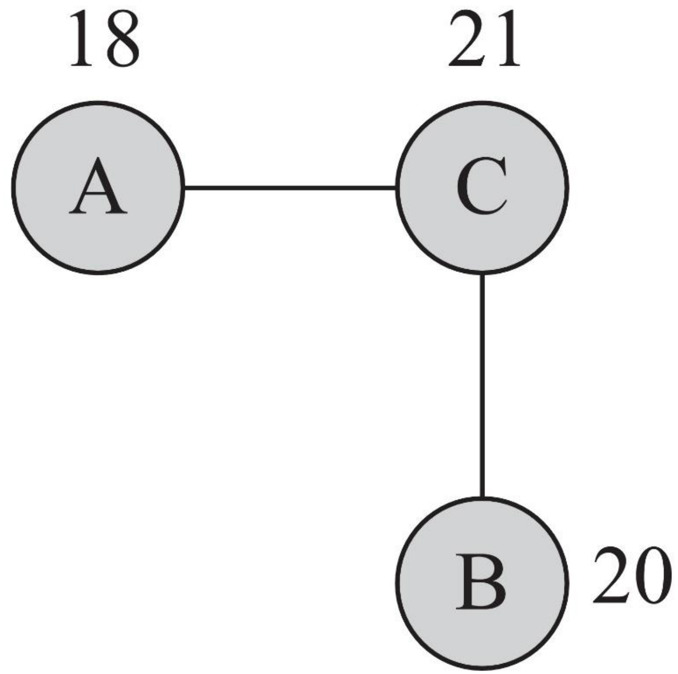
Covariance Variables

- The value assigned to node v_i is x_i
- We construct two variables X_L and X_R
- For any edge (v_i, v_j), we **assume** that x_i is observed from variable X_L and x_j is observed from variable X_R
- X_L represents the ordinal values associated with the left-node (the first node) of the edges
- X_R represents the values associated with the right-node (the second node) of the edges
- We need to compute the covariance between variables X_L and X_R

Covariance Variables: Example

List of edges:

(A, C)
(C, A)
(C, B)
(B, C)



X_L : (18, 21, 21, 20)

X_R : (21, 18, 20, 21)

$$\mathbf{E}(X_L) = \mathbf{E}(X_R)$$

$$\sigma(X_L) = \sigma(X_R)$$

Undirected edges are equivalent to bi-directed edges;

Covariance

For two given column variables X_L and X_R the covariance is

$$\sigma(X_L, X_R) \quad = \quad \mathbf{E}[X_L X_R] - \mathbf{E}[X_L]\mathbf{E}[X_R]$$

$E(X_L)$ is the mean of the variable and $E(X_L X_R)$ is the mean of the multiplication X_L and X_R. The first sum is over edges where i is the index of edges; and the second sum is over nodes where i is the index of nodes.

$$E(X_L) = E(X_R) = \frac{\sum_i (X_L)_i}{2m} = \frac{\sum_i d_i x_i}{2m}$$

$$E(X_L X_R) = \frac{1}{2m} \sum_i (X_L)_i (X_R)_i = \frac{\sum_{ij} A_{ij} x_i x_j}{2m}$$

Covariance

$$\begin{aligned}\sigma(X_L, X_R) &= \mathbf{E}[X_L X_R] - \mathbf{E}[X_L]\mathbf{E}[X_R] \\ &= \frac{\sum_{ij} A_{ij} x_i x_j}{2m} - \frac{\sum_{ij} d_i d_j x_i x_j}{(2m)^2} \\ &= \frac{1}{2m} \boxed{\sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m} \right)} x_i x_j\end{aligned}$$

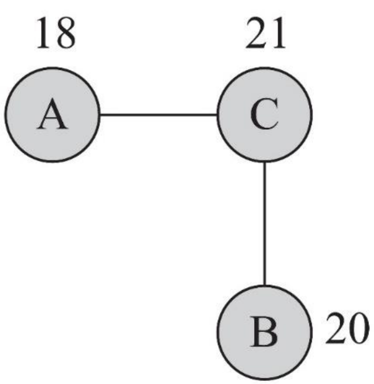
Normalizing Covariance

Pearson correlation $\rho(X,Y)$ is the normalized version of covariance

$$\rho(X_L, X_R) = \frac{\sigma(X_L, X_R)}{\sigma(X_L)\sigma(X_R)}.$$

In our case: $\sigma(X_L) = \sigma(X_R)$ $\sigma^2_X = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - [\mathrm{E}[X]]^2$

Standard deviation **Variance**



$$X_L = \begin{bmatrix} 18 \\ 21 \\ 21 \\ 20 \end{bmatrix} \qquad X_R = \begin{bmatrix} 21 \\ 18 \\ 20 \\ 21 \end{bmatrix}$$

$$\rho(X_L, X_R) = -0.67$$

Node A and Node B are more similar, but they do not have an edge.

References

- R. Zafarani, M. A. Abbasi, and H. Liu, Social Media Mining: An Introduction, Cambridge University Press, 2014.
- http://socialmediamining.info/