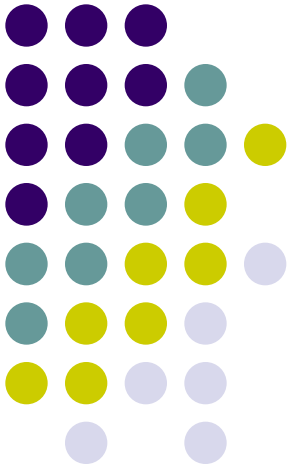
Social Media Analytics

Homophily vs. Social Influence

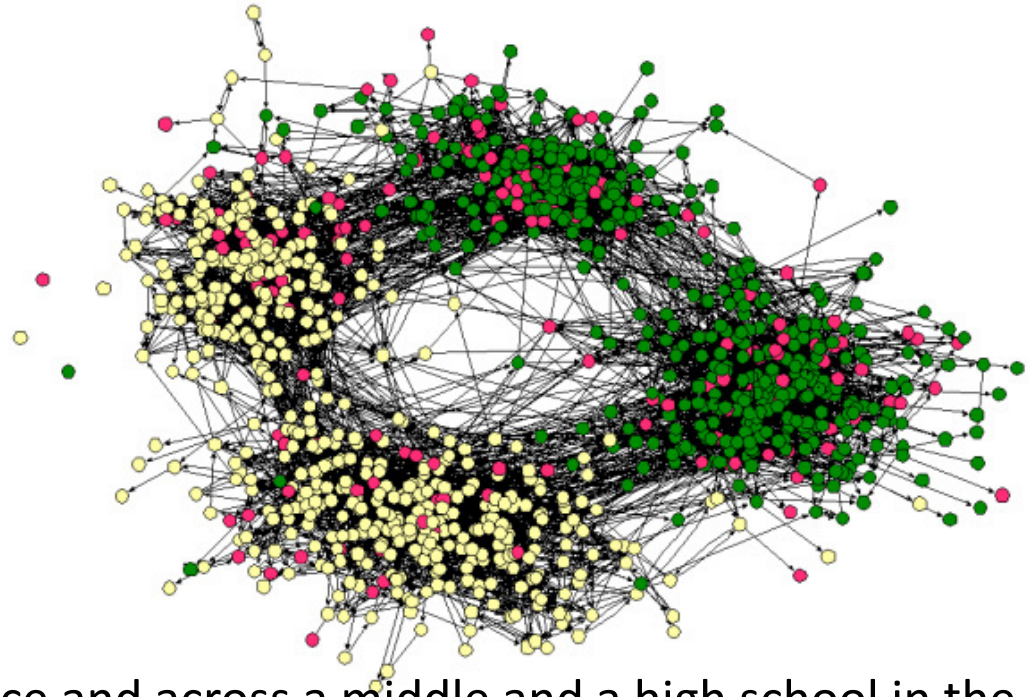
MSBA, S2022, Feb 21

Dr. Anitesh Barua
David Bruton Jr. Centennial Chair Professor of Business
Distinguished Fellow, INFORMS Information Systems Society
Stevens Piper Foundation Professor
University of Texas Distinguished Teaching Professor
Associate Director, Center for Research in e-Commerce
McCombs School of Business, University of Texas at Austin
Email: aniteshb@gmail.com



Homophily (Similarity)

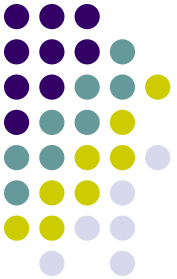
- "Birds of a feather flock together"
  - Your friends/contacts vs. a random sample of people
  - Social networks tend to connect people who are similar to each other



Friendships by race and across a middle and a high school in the same school district

Source: Easley & Kleinberg

Distinguishing Between Social Influence and Homophily

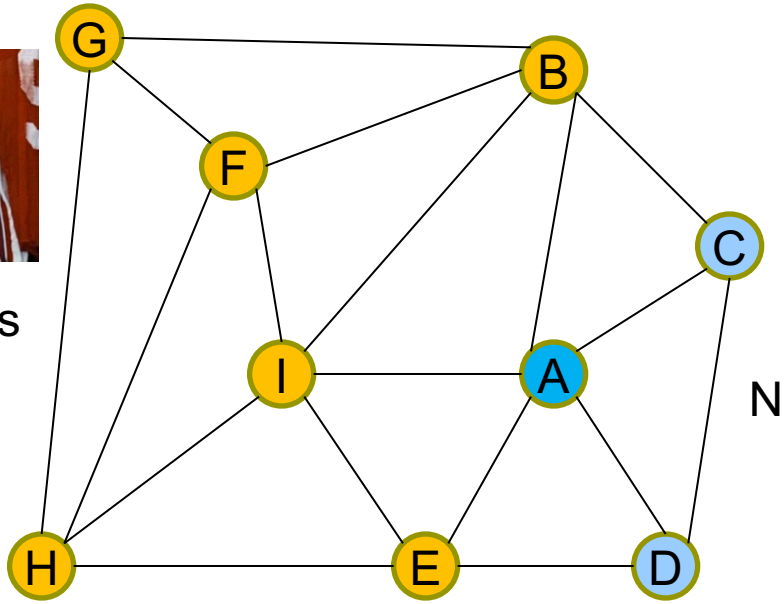


- Can opinions, attitudes & purchases be attributed to social influence?
- Or is it due to homophily?
- E.g., did I buy something because
  - you influenced me?
  - we are just similar?
- What difference would it make to a company's strategy?

Detecting Homophily for Static Attributes

- Have to know what attribute(s) may be relevant
- E.g., gender, interest, educational background, etc.

Jocks



Nerds

- Does this network exhibit homophily?
- What measure can we use?

Source: Easley & Kleinberg

A Little Theoretical Detour

A network with a set of nodes (V) & randomly assigned edges (E^r):

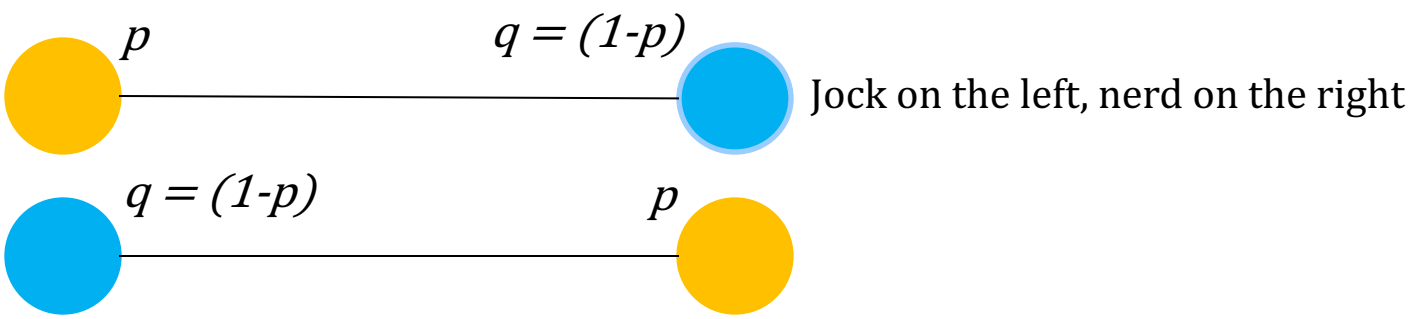
$$R = (V, E^r)$$

Each node is assigned an attribute: say, type = jock with probability p, and type = nerd with probability $q = 1\text{-}p$

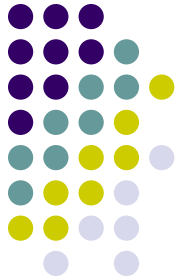
Consider any edge $(i,j) \in E^r$ of this random network R.

Let the random variable $X_{ij} = 1$ if it is a "cross-edge", and $X_{ij} = 0$ otherwise. Then X_{ij} is a Bernoulli random variable such that

$P(X_{ij} = 1) = 2pq$



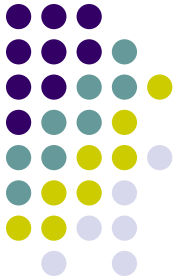
Dynamic Attributes: How Can We Distinguish Between Homophily & Influence?



- Need multiple snapshots in time
- Homophily: Due to **similar** attributes in time *t*, some people may choose to become friends in *t*+1
  - E.g., high achievers in a class may form links
- But some people may become friends in *t*+1 even though their attributes were different in *t*
- Check which effect is stronger

Source: LaFond & Neville

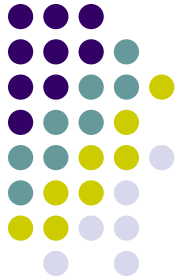
Test for Homophily



- Homophily exists if
- *p*(Becoming friends in *t*+1 where attributes were same in *t*) > *p*(Becoming friends in *t*+1 where attributes were different in *t*)
- *p*(Dissolving friendships in *t*+1 where attributes were same in t) < *p*(Dissolving friendships in *t*+1 where attributes were different in *t*)

Source: LaFond & Neville

Detecting Social Influence



- Some **friends** at *t* with different attributes may become similar in *t*+1 (due to social influence)
  - E.g., some buy a product their friends have
  - Some change their beliefs & attitudes
- But people who are not friends and have different attributes at *t* can also become similar at *t*+1 due to "other" factors
- Which effect is stronger?
  - I.e., is *p*(Attributes becoming same in *t*+1 where the individuals were friends in *t*) > *p*(Attributes becoming same in *t*+1 where the individuals were not friends in *t*)?
  - Is *p*(Attributes becoming different in *t*+1 where the individuals were friends in *t*) < *p*(Attributes becoming different in *t*+1 where the individuals were not friends in *t*)?

Source: LaFond & Neville

Distinguishing Between Homophily & Social Influence: The Case of Dynamic Attributes

Time: *t* *t*+1

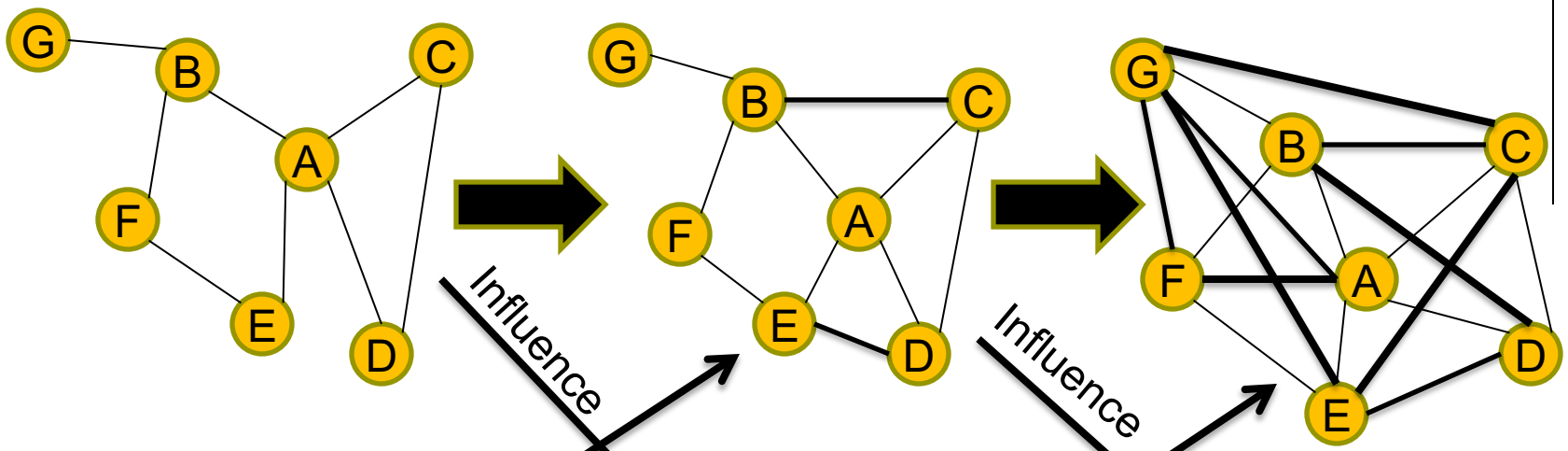
	Subscription
A	Yes
B	No
C	No
D	Yes
E	No
F	Yes
G	Yes

Attribute (e.g., subscription to a music service at time *t*

A	Yes
B	Yes
C	Yes
D	Yes
E	No
F	Yes
G	Yes

Attribute at time t+1

Evidence of Homophily Versus Social Influence



Time: *t*

t+1

t+2

Influence

Homophily

Influence

Homophily

A	Yes
B	No
C	No
D	Yes
E	No
F	Yes
G	Yes

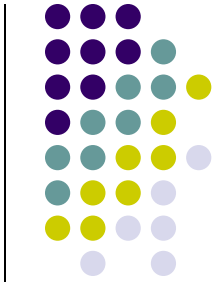
Attribute at time *t*

A	Yes
B	Yes
C	Yes
D	Yes
E	No
F	Yes
G	Yes

Attribute at time *t*+1

Attribute at time *t*+2

Testing Significance Levels



X^i is an atttribute of node v_i

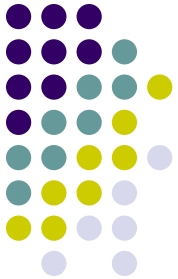
P_R is a set of related nodes (friends) in a network

	$X^i = X^j = x$	$\neg (X^i = X^j = x)$
$(v_i, v_j) \in P_R$	a	b
$(v_i, v_j) \notin P_R$	c	d

$$Relational\ autocorrelation\ C(X,G) = \chi^2 = \frac{N.(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Source: LaFond & Neville

Chi-Square (for Contingency Tables)



	Improved Outcome	Didn't improve	Total
Treatment	36	14	50
No Treatment	30	25	55
Total	66	39	105

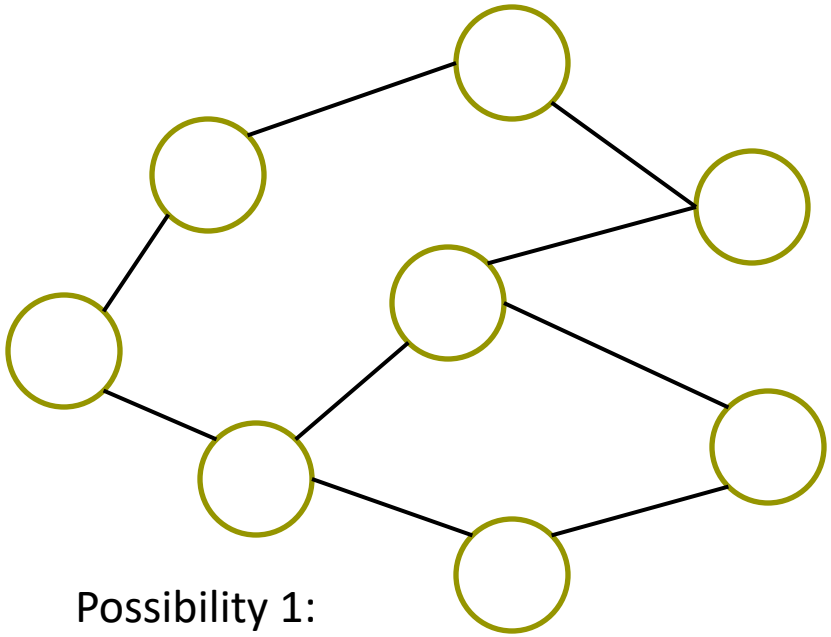
$$\chi^2 = ?$$

$$\chi^2 = \frac{105*(36*25-30*14)^2}{50*55*66*39} = 3.42$$

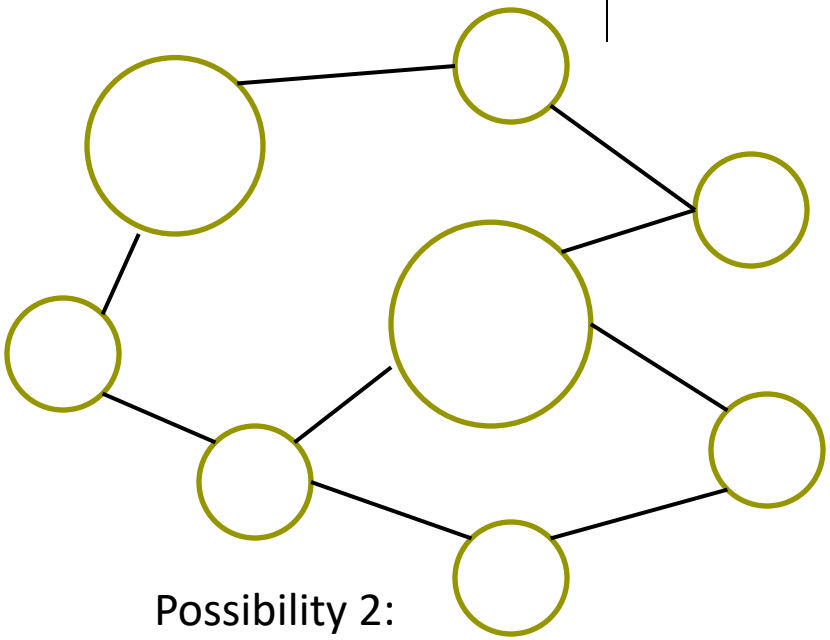
Significance Tests

- Homophily: $C(X_t, G_{t+1}) > C(X_t, G_t)$
- Social influence: $C(X_{t+1}, G_t) > C(X_t, G_t)$

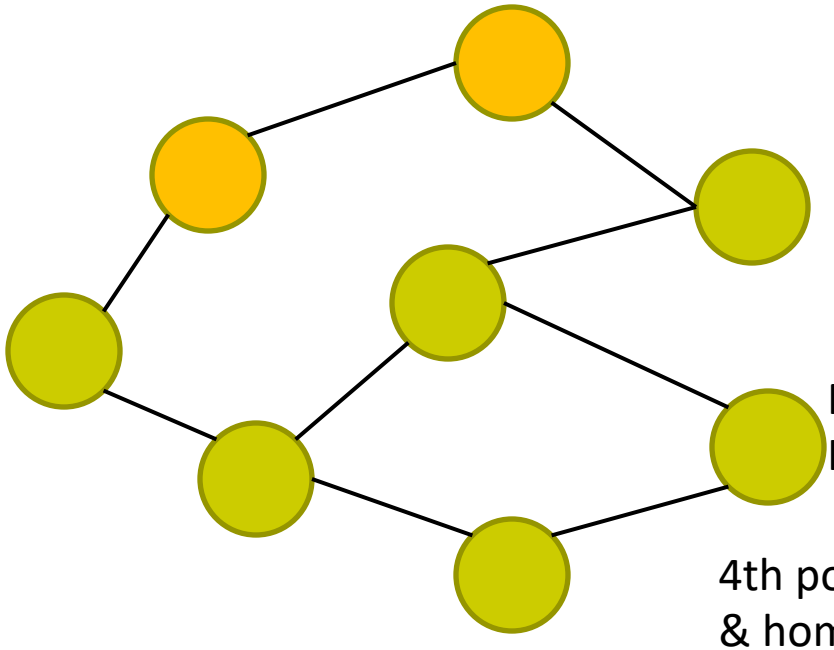
Not all Networks are Created Equal



Possibility 1:
No social influence, no homophily



Possibility 2:
Social influence but no homophily



Possibility 3:
Homophily but no social influence

4th possibility (not shown): Both influence & homophily