CSI 445/660 – Part 3

(Networks and their Surrounding Contexts)

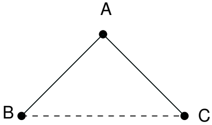
Ref: Chapter 4 of [Easley & Kleinberg].

External Factors Affecting Network Evolution

Homophily:

- A basic principle: “We tend to be similar to our friends”.
- Governs the structure of social networks.
- Has a long history:
  - Socrates: “People love those who are like themselves”.
  - Plato: “Similarity begets friendship”.
  - Well known proverb: “Birds of a feather flock together”.
- Provides an illustration of how the **surrounding context** drives the formation of networks.

Triadic Closure



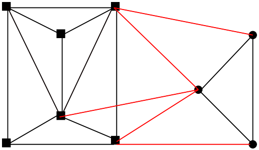
- Having a common friend is one reason for triadic closure.
- Homophily provides another reason.

- Suppose B and C are majors in the same department.
- They may become friends even though there is no common friend. (This is an effect of the surrounding context).

Measuring Homophily:

- A characteristic must be specified.
- **Examples:** Age, gender, ethnicity.
- How can we check whether a given network exhibits homophily with respect to a specified characteristic?

Measuring Homophily (continued)



- Friendship network of some children in an elementary school.
- Circles denote girls and squares denote boys.

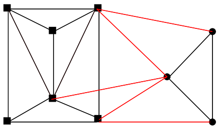
- We want to check whether this network exhibits **gender homophily**.
- **Extreme case of homophily:** The network does not have any "cross-gender edge" (i.e., an edge joining a boy and a girl). This is not typical.
- One can develop a numerical measure of homophily with respect to a characteristic.
- This will be illustrated using a characteristic (namely, gender) which has two possible values.

Description of the Method: See Handout 3.1.

Measuring Homophily (continued)

Homophily Test: Consider a network H with N_B boys and N_G girls. Let $p = N_B/(N_B + N_G)$ and $q = 1 - p = N_G/(N_B + N_G)$. If the fraction of cross edges in H is significantly below $2pq$, then there is evidence for gender homophily.

Example:



- Here, $N_B = 6$ and $N_G = 3$.
- Total number of edges $= 18$.
- No. of cross edges $= 5$.
- So, fraction of cross edges $= 5/18$.

- $p \ = \ N_B/(N_B + N_G) \ = \ 6/9 = 2/3$.
- $q \ = \ 1 - p = 1/3$.
- $2pq \ = \ 4/9 = 8/18$.
- Since the actual fraction of cross edges (5/18) is less than the fraction $2pq$, we conclude that the network exhibits some degree of homophily.

Mechanisms Underlying Homophily

- Homophily is observed behavior.
- Sociologists want to understand the mechanisms that lead to homophily.
- Two known mechanisms are **selection** and **socialization**.

Selection:

- Applies to **immutable** characteristics (such as ethnicity or race).
- People "select" friends with similar characteristics.

Socialization or Social Influence:

- Applies to **mutable** characteristics (e.g. behaviors, interests, beliefs, opinions).
- People may modify their characteristics to align with the behaviors of their friends.

Selection and Socialization

- Socialization may be viewed as the **reverse** of selection.
- **Reason:**
  - With selection, individual characteristics drive the formation of links.
  - With socialization, links in a network shape people's (mutable) characteristics.
- In general, there is also some interplay between the two mechanisms.

Longitudinal Studies to Understand Link Formation:

- From a single snapshot of a network, it is generally difficult to determine the reason for the formation of links.
- **Longitudinal** studies, where links and behaviors are tracked over a period of time, are needed.

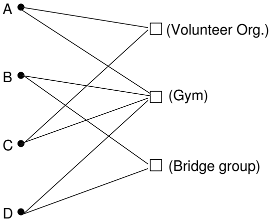
A Famous Longitudinal Study: Summary

- Published in 2007 by Nicholas Christakis (Yale University) and James Fowler (UC San Diego).
- Longitudinal study (part of Framingham Heart Study) over a 32 year period (1971 to 2003) involving 12,067 people.
- **Focus:** Obesity status.
- **Observation:** Normal weight and overweight people formed clusters in the network consistent with homophily.
- The main cause of homophily was social influence; changes in the obesity status of one's friends had a significant effect on the person.
- The authors go on to suggest that obesity is a form of **contagion** that spreads through a social network. **(This suggestion has been questioned by other researchers.)**

Affiliation Networks

- **So far:** Surrounding context not part of the network.
- The idea of **affiliation networks** allows the surrounding context to be part of the network.
- Introduce activities or **focal points** as nodes in the network, leading to a more general form of the network.
- Examples of focal points: Hobbies, interests.

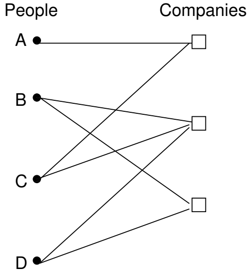
Affiliation Network Example:



- Dark circles: People.
- Squares: Focal points.

Affiliation Networks (continued)

An Affiliation Network from Previous Discussion:

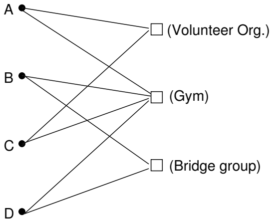


Note: An edge between a person x and a company y indicates that x serves (or served) on the Board of Directors for y.

Focus: Formation of edges between people due to focal points.

Affiliation Networks (continued)

Some Graph Theoretic Definitions:



- An example of a **bipartite** graph.
- There are two sets of nodes.
- Each edge joins a node from one set to a node in the other set. **(No edge joins a pair of nodes in the same set.)**

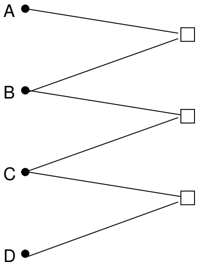
Observation: Each cycle in a bipartite graph must contain an **even** number of nodes and edges.

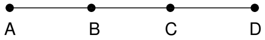
Projected Networks of Affiliation Networks

Projected Network:

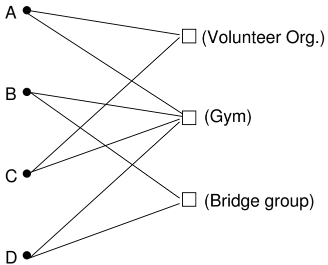
- Network on the nodes representing people.
- There is an edge between two people if they both have edges to at least one **common** focal point.

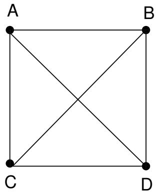
Example 1: An affiliation network and the corresponding projected network.





Example 2: Another affiliation network and the corresponding projected network.

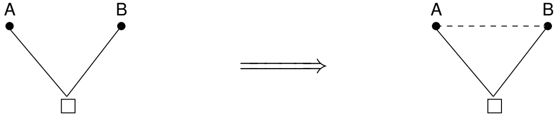




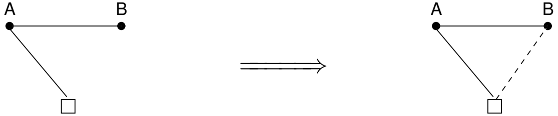
Link Formation in Affiliation Networks

Three forms of Closure Processes:

- Triadic closure: Due to a common friend or homophily.
- **Focal closure:** A new edge may form because of a common focal point (effect of homophily).



- **Membership closure:** A new edge may form between a person and a focal point (also an effect of homophily).



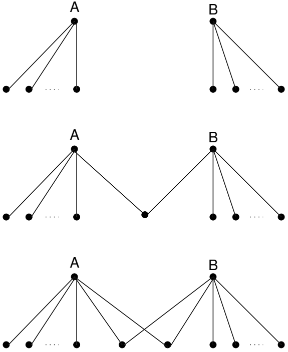
Link Formation in ... (continued)

Question: Can we study the link formation issue in a quantitative fashion?

Illustration – Study of Triadic Closure:

- Study done by Kossinets and Watts [2006].
- **Caveat:** Study uses online data; conclusions from the study may not be applicable to settings based on human interactions.
- **Basic questions:**
  1. How does the likelihood of the formation of a link increase when two people have one friend in common (compared to when they they have no common friend)?
  2. How does the likelihood increase when two people have two or more friends in common?

Study by Kossinets & Watts (continued)



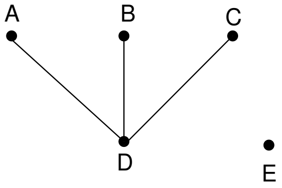
- No common friend.
- One common friend.
- Two common friends.

Note: We would expect the likelihood to increase as the number of common friends increases.

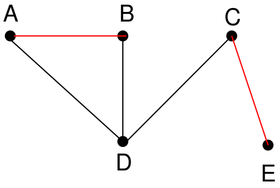
Study by Kossinets & Watts (continued)

Description of Methodology: See Handout 3.2.

Example to Illustrate the Methodology:



Network N_1

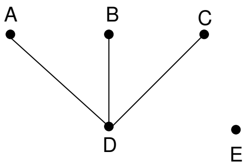


Network N_2

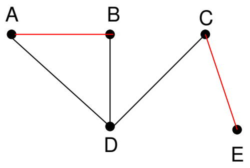
Note:

- In N_1, pairs of nodes have no common neighbor or one common neighbor.
- So, according to the methodology, we must construct the sets S_0 and S_1.

Illustrative Example (continued)



Network N_1



Network N_2

S_0 = Set of pairs (x, y) such that x and y have **no** common neighbor in N_1 and the edge $\{x, y\}$ is not in N_1

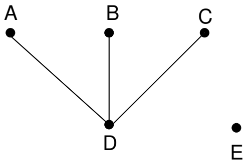
= {(A,E), (B,E), (C,E), (D,E)}

Q_0 = Subset of S_0 such that for each pair (x, y) in Q_0 the edge $\{x, y\}$ is in N_2

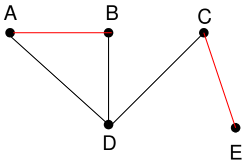
= {(C,E)}

Thus, $|S_0| = 4$, $|Q_0| = 1$ and $T(0) = |Q_0|/|S_0| = 1/4$.

Illustrative Example (continued)



Network N_1



Network N_2

S_1 = Set of pairs (x, y) such that x and y have **one** common neighbor in N_1 and the edge $\{x, y\}$ is not in N_1

= {(A,B), (A,C), (B,C)}

Q_1 = Subset of S_1 such that for each pair (x, y) in Q_1 the edge $\{x, y\}$ is in N_2

= {(A,B)}

Thus, $|S_1| = 3$, $|Q_1| = 1$ and $T(1) = |Q_1|/|S_1| = 1/3$.

Study by Kossinets & Watts

Details About the Data Set:

- Data from email communication between students at a US university. No. of students $\approx$ 22,600.
- Observation period: One year.
- Each student is a node; the edge $\{x, y\}$ is added when they exchanged email.
- By considering multiple pairs of snapshots of the network, they constructed an average value of $T(k)$ for each value of k.

Results:

- $T(0)$ (the likelihood of link formation with no common friends) is close to 0.
- Probability of link formation increases with the number of common friends (k).
- Having two common friends increases the likelihood by a factor of more than 2 compared to having one common friend.

Study by Kossinets & Watts: Triadic Closure

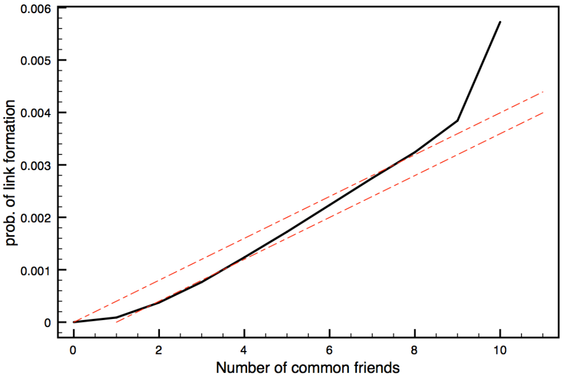


Figure 4.9: Quantifying the effects of triadic closure in an e-mail dataset [259]. The curve determined from the data is shown in the solid black line; the dotted curves show a comparison to probabilities computed according to two simple baseline models in which common friends provide independent probabilities of link formation.

Comparison with a Baseline Model

Assumption: There is a (small) value p such that for each pair of people x and y, each common friend causes the the link $\{x, y\}$ **independently** with probability p.

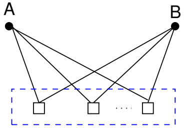
Model Derivation:

- Suppose x and y have $k \geq 1$ friends in common.
- The probability that they **don't** form a link is $(1-p)^k$.
- So, the probability $T_b(k)$ that they **do** form a link is given by $T_b(k) \; = \; 1-(1-p)^k$.

Notes:

- The plot in Slide 3-21 shows the actual curve sandwiched between $T_b(k)$ and $T_b(k-1)$.
- The value of p is chosen so that the model provides a good alignment with the actual curve.
- The plot suggests that the baseline model is reasonable for low values of k.

Focal Closure – Brief Discussion



- **Goal:** To understand how the likelihood of the link {A, B} depends on the number of common foci.

- Kossinets & Watts used classes as the foci.
- They computed the empirical estimates of the probability values using the methodology discussed in Handout 3.2. (The results are shown in Slide 3-24.)
- When the number of common courses is small, the likelihood of link formation increases.
- A subsequent increase in the number of common courses has a "diminishing returns" effect.

Study by Kossinets & Watts: Focal Closure

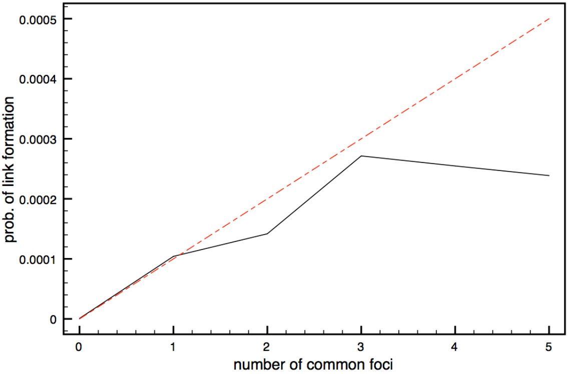


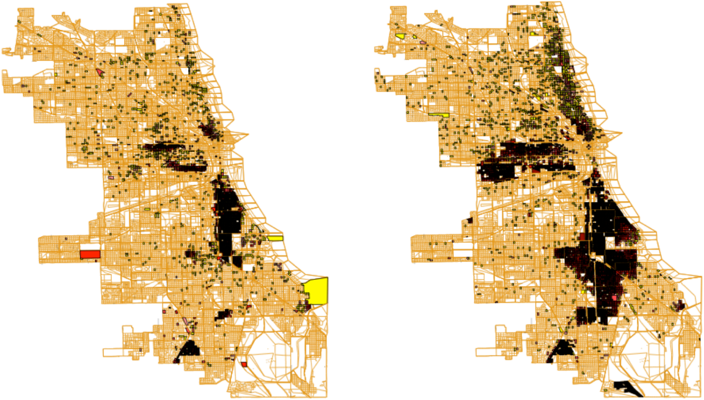
Figure 4.10: Quantifying the effects of focal closure in an e-mail dataset [259]. Again, the curve determined from the data is shown in the solid black line, while the dotted curve provides a comparison to a simple baseline.

Schelling's Model of Spatial Segregation

- Thomas Schelling (1921 –)
- Professor Emeritus, University of MD & New England Complex Systems Institute.
- Winner of the 2005 Nobel Prize in Economics (for contributions to Game Theory).

- **An easily seen effect of homophily:** Racially homogeneous neighborhoods (see Slide 3-26).
- In 1972, Schelling suggested a spatial model to explain this.
  - The model shows that global patterns of segregation can arise due to homophily operating at a local level.
  - These mechanisms operate even when no single individual wants segregation.

Schelling's Model ... (continued)

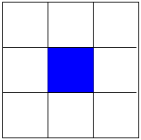


(a) *Chicago, 1940* (b) *Chicago, 1960*

Figure 4.14: The tendency of people to live in racially homogeneous neighborhoods produces spatial patterns of segregation that are apparent both in everyday life and when superimposed on a map — as here, in these maps of Chicago from 1940 and 1960 [302]. In blocks colored yellow and orange the percentage of African-Americans is below 25, while in blocks colored brown and black the percentage is above 75.

Details of the Model (Game):

- Grid representation for a city, with each cell representing a section of the city.
- There are two types of people (**agents**), denoted by `X` and `O`.
- This agent classification is based on an **immutable** characteristic.
- Some grid cells have agents while others are empty.



- Each cell may have up to 8 neighbors.
- Boundary cells have fewer neighbors.

Schelling's Model ... (continued)

Constraint: Each agent wants to have **at least** t other agents of the same type as its neighbors.

- The parameter t is called the **threshold**. (For example, t may be chosen as 3.)
- An agent with **less than** t neighbors of the same type is "unsatisfied" and wants to move to another cell where the threshold is satisfied.

Dynamics of Movement:

- Movement happens in **rounds**.
- Actions carried out in each each round:
  - Unsatisfied agents are considered in some (predetermined) order.

Schelling's Model ... (continued)

- Actions in each round (continued):
  - Each unsatisfied agent is moved to a cell where the threshold is satisfied. (If there are many possible cells, one is chosen randomly. If there is no such cell, the agent may be moved to a random cell or left where it is.)
  - Such movements may cause other agents to become "unsatisfied".
- Rounds are repeated until all agents are satisfied. (The game may never end.)

An Example of a Configuration:

X1*	X2*				
X3	O1*		O2		
X4	X5	O3	O4	O5*	
X6*	O6			X7	X8
	O7	O8	X9*	X10	X11
		O9	O10	O11*	

Notes:

- Agents are numbered so that their movements can be readily followed.
- Threshold value $t = 3$.
- An asterisk is used to indicate an unsatisfied agent.

Schelling's Model ... (continued)

A Subsequent Configuration:

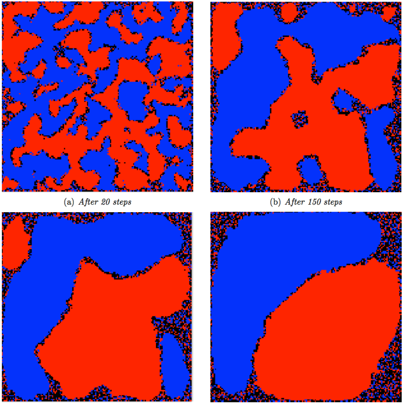
X3	X6	O1	O2		
X4	X5	O3	O4		
	O6	X2	X1	X7	X8
O11	O7	O8	X9	X10	X11
	O5	O9	O10*		

Notes:

- Small examples are useful in understanding the model.
- Several applets are available on the web to try larger examples.

Schelling's Model ... (continued)

Results from an Applet: (Threshold = 4)



(a) *After 20 steps* (b) *After 150 steps*

(c) *After 350 steps* (d) *After 800 steps*

Figure 4.19: Four intermediate points in a simulation of the Schelling model with a threshold t of 4, on a 150-by-150 grid with 10,000 agents of each type. As the rounds of movement progress, large homogeneous regions on the grid grow at the expense of smaller, narrower regions.

Observations Regarding Schelling's Model

- Segregation takes place (at the global level) even though no agent is actively seeking it. (Each agent is willing to have a minority of neighbors of its type.)
- Segregation is **not built into** the model; there are patterns where there is not much segregation. (However, empirical evidence suggests that such patterns are hard to reach from random initial configurations.)
- In real life, segregation effect is amplified by genuine desire on the part of a small fraction of people who want to avoid other types of people.
- Schelling's work also suggests that immutable characteristics (e.g. race, ethnicity) may be highly correlated with certain mutable characteristics (e.g. decision about where to live).