SOCIAL NETWORK ANALYTICS

Homophily

Prakash C O
Department of Computer Science and Engineering

SOCIAL NETWORK ANALYTICS

Homophily

Prakash C O
Department of Computer Science and Engineering

Homophily

➢ **Homophily:** (i.e., love of the same)

- **The tendency of individuals to associate and bond with similar others.**
- **People interact more often with people who are "like them" than with people who are dissimilar.**

➢ **What leads to Homophily?**

- Beliefs, Values, Attitudes, Abilities, Aspirations, Education, Social status, Age, Race, Sex, Gender, Occupation, Religion etc.

Idioms: 'birds of a feather flock together' = 'Similar people find each other'

Homophily

- **Homophily:**
  - It refers to **the degree to which pairs of individuals who interact are similar with respect to certain attributes**, such as beliefs, values, education, social status, age, race, gender, occupation, religion etc.
  - Individuals in **homophilic relationships share common characteristics that make communication and relationship formation easier.**
    Example: Two musicians are more likely to become friends.
  - The homophily makes networks to contain cliques, and near-cliques, meaning subnets which have connections between almost any two nodes within them. Hence, homophily creates high clustering in networks.

Homophily

➢ **Homophily in Social Networks:**

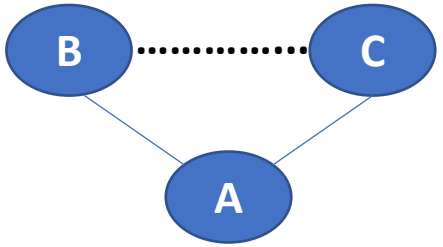
- **One of the most basic notions governing the structure of social networks is homophily** — the principle that we tend to be similar to our friends.
  The pervasive fact is that **links in a social network tend to connect people who are similar to one another**.

- Homophily in social relations may lead to the creation of clusters that have been observed in social networking services.

➢ Homophily is a key topic in network science as **it can determine the speed of the diffusion of information and ideas.**

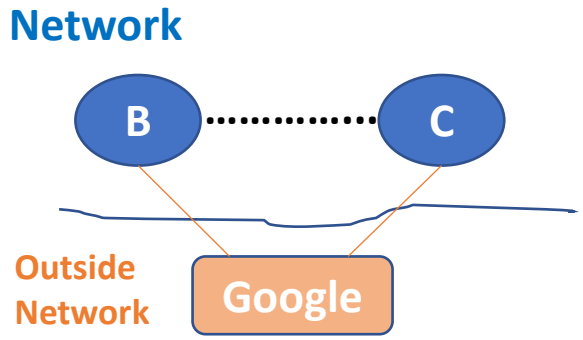
Homophily

- Homophily provides us with a first, fundamental illustration of **how a network's surrounding contexts can drive the formation of its links.**
- Consider the basic contrast between
  1. **a friendship that forms because two people are introduced through a common friend**
  2. a friendship that forms because two people attend the same school or work for the same company.
- **In the first case**, (i.e., with Triadic closure-intrinsic mechanism)
  - **a new link is added for reasons that are *intrinsic* to the network itself; we need not look beyond the network to understand where the link came from.**



Homophily

- Homophily provides us with a first, fundamental illustration of **how a network's surrounding contexts can drive the formation of its links.**
- Consider the basic contrast between
  1. a friendship that forms because two people are introduced through a common friend
  2. **a friendship that forms because two people attend the same school or work for the same company.**
- **In the second case**, (i.e., with Homophily - contextual mechanism)
  - the new link arises for an equally natural reason, but one that makes sense only when we look at the *contextual factors* beyond the network — at some of the social environments (in this case schools and companies) to which the nodes belong.



Homophily

- When we look at a network, such contexts capture some of the dominant features of its overall structure.
- Figure 1, for example, depicts the **social network of middle school and high school students**; in this image, students of different races are drawn as differently-colored circles.
- **Two dominant divisions within the network** are apparent.
  - **One division is based on race** (from left to right in the figure);
  - **the other, based on age and school attendance, separates students in the middle school from those in the high school** (from top to bottom in the figure).
- There are many other structural details in this network, but the effects of these two contexts stand out when the network is viewed at a global level.

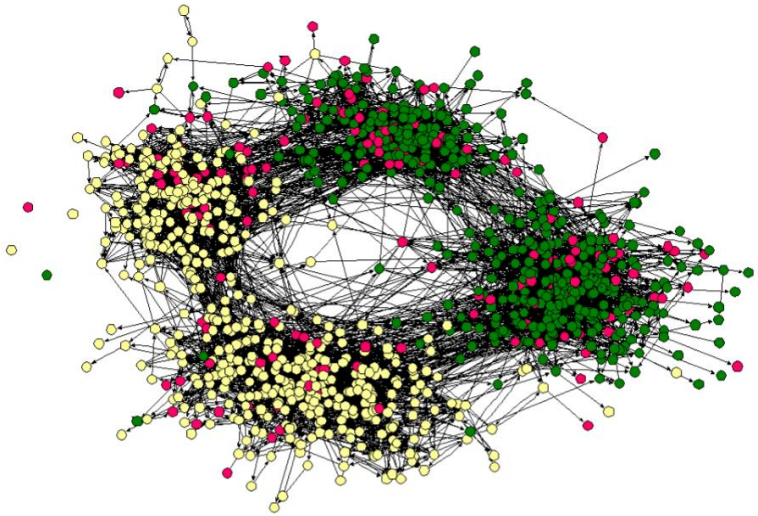


Figure 1

SOCIAL NETWORK ANALYTICS

Homophily

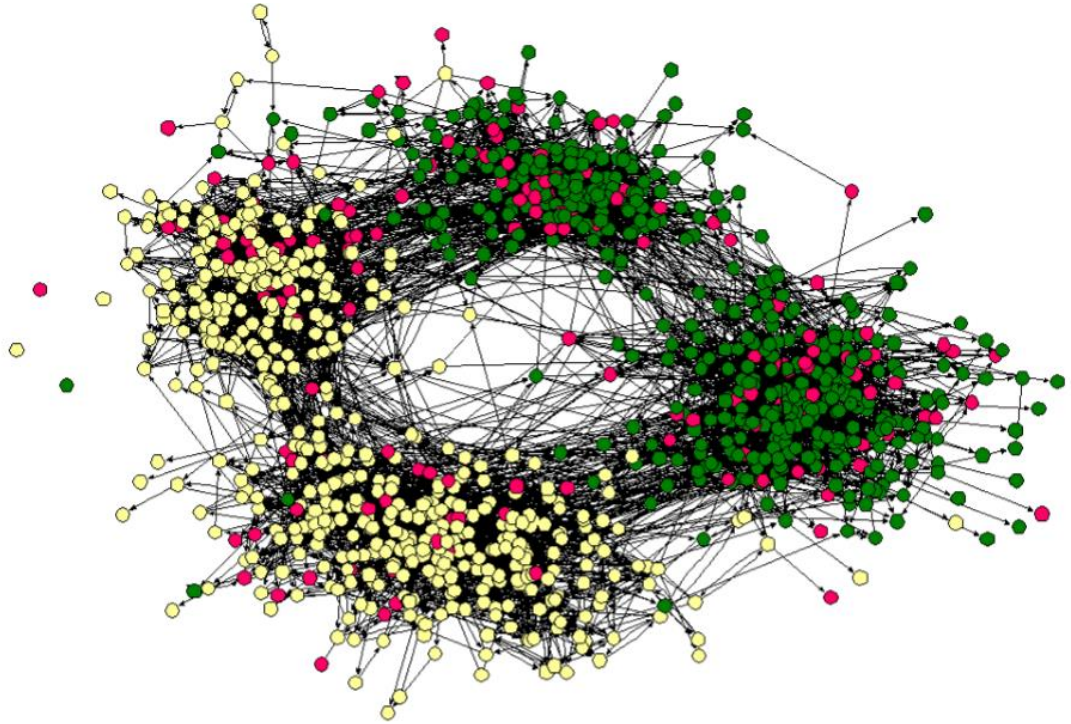


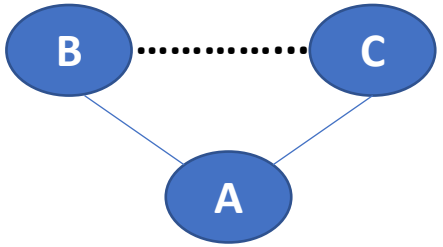
Figure 1: Homophily can produce a division of a social network into densely-connected, homogeneous parts that are weakly connected to each other. In this social network from a town's middle school and high school, two such divisions in the network are apparent: one based on race (with students of different races drawn as differently colored circles), and the other based on friendships in the middle and high schools respectively.

SOCIAL NETWORK ANALYTICS

Homophily

- **In social networks, there are strong interactions between *intrinsic* and *contextual* effects on the formation of any single link; they are both operating concurrently in the same network.**

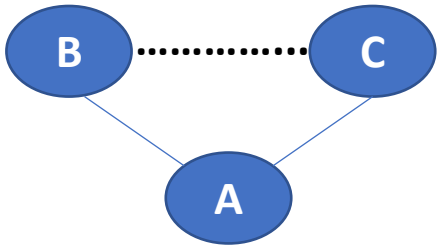
- **For example, the principle of triadic closure** - that triangles in the network tend to "close" as links form between friends of friends - is supported by a range of mechanisms that range from the intrinsic to the contextual.



  - **The triadic closure is motivated by hypothesizing intrinsic mechanisms:** When individuals B and C have a common friend A, then there are increased opportunities and sources of trust on which to base their interactions, and A will also have incentives to facilitate their friendship.
  - **Social contexts also provide natural bases for triadic closure**: Since we know that A-B and A-C friendships already exist, the principle of homophily suggests that B and C are each likely to be similar to A in a number of dimensions, and hence quite possibly similar to each other as well. As a result, based purely on this similarity, there is an elevated chance that a B-C friendship will form; and this is true even if neither of them is aware that the other one knows A.

Homophily

- As we take into account more and more of the factors that drive the formation of links in a social network, it inevitably becomes difficult to attribute any individual link to a single factor.



- Ultimately, one expects most links to in fact arise from a combination of several factors

1. partly due to the effect of other nodes in the network, and
2. partly due to the surrounding contexts.

Homophily

Measuring Homophily

- **Is there a simple test we can apply to a network in order to estimate whether it exhibits homophily according to some characteristic?**

- Let's suppose in particular that we have the friendship network (6-boys and 3-girls) of an elementary-school classroom, and we suspect that it exhibits homophily by gender:
  - boys tend to be friends with boys, and
  - girls tend to be friends with girls.

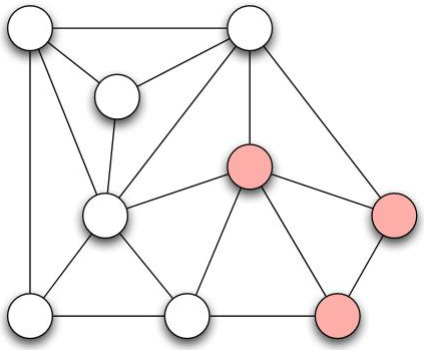


Figure 2: friendship network of a (hypothetical) classroom: shaded nodes are girls and the six unshaded nodes are boys

Homophily

Measuring Homophily

- Suppose we have a network in which
  - a **p fraction of all individuals are male,** and
  - a **q fraction of all individuals are female**.
- Consider a given edge in this network.
  - If we independently assign each node the gender male with probability p and the gender female with probability q, then
    - **If both ends of the edge is male,** so this happens with probability p^2,
    - **Is both ends of the edge is female,** so this happens with probability q^2,
    - If the first end of the edge is male and the second end is female, so this happens with probability **pq**.

      If the first end of the edge is female and the second end is male, so this happens with probability **qp**.

      then we have **a cross-gender edge**, with probability **2pq**.

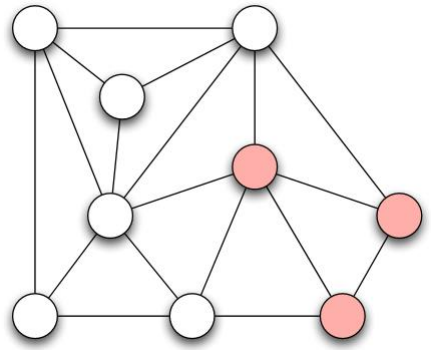


Figure 2: friendship network of a (hypothetical) classroom: shaded nodes are girls and the six unshaded nodes are boys
P=6/9 and q=3/9

Homophily

Measuring Homophily

- The test for homophily according to gender is as follows:

  Homophily Test: ***If the fraction of cross-gender edges is significantly less-than 2pq, then there is evidence for homophily.***

- In this example(Figure 2), Since p = 6/9 = 2/3 and q = 3/9 = 1/3, then
  - The probability of a p-p link = 2/3 * 2/3 = 4/9
  - The probability of a p-q link = 2/3 * 1/3 = 2/9
  - The probability of a q-p link = 1/3 * 2/3 = 2/9
  - The probability of a q-q link = 1/3 * 1/3 = 1/9

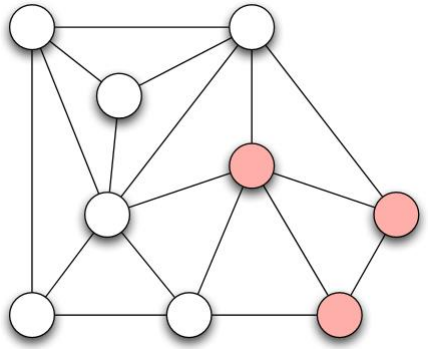


Figure 2:

$$(p+q)^2 = p^2 + 2pq + q^2 = [4/9 + ((2/9)+(2/9)) + 1/9\ ] = 1$$

Homophily

Measuring Homophily

- The test for homophily according to gender is as follows:
  Homophily Test: ***If the fraction of cross-gender edges is significantly less-than 2pq, then there is evidence for homophily.***
- In Figure-2, 5 of the 18 edges in the graph are cross-gender. Since p = 6/9 = 2/3 and q = 3/9 = 1/3 in this example, we should be comparing the fraction of cross-gender edges to the quantity 2pq =2*(2/3)*(1/3) = 4/9 = 8/18.

  5/18 is significantly less than 8/18

- In other words, with no homophily, one should expect to see 8 cross-gender edges rather than 5, and so this example shows some evidence of homophily.

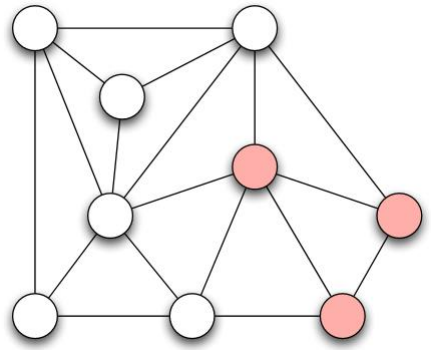


Figure 2:

Measuring Homophily

There are a few points to note on ***Homophily Test.***

- First, the number of cross-gender edges in a random assignment of genders will deviate some amount from its expected value of 2pq, and so **to perform the test in practice one needs a working definition of "significantly less than**."
  Standard measures of statistical significance (quantifying the significance of a deviation below a mean) can be used for this purpose.
- Second, it's also easily possible for a network to have a fraction of cross-gender edges that is significantly more than 2pq. In such a case, we say that the network exhibits **inverse homophily**.

Homophily

Quiz

- Consider the network of a school where 1/4 fraction of the students are boys and 3/4 fraction are girls. For an evidence of Homophily in this network, what should be the fraction of cross gender edges?

1) Significantly greater than 3/8.
2) Significantly lesser than 3/8.
3) Significantly greater than 1/3.
4) Significantly lesser than 1/3.

Homophily

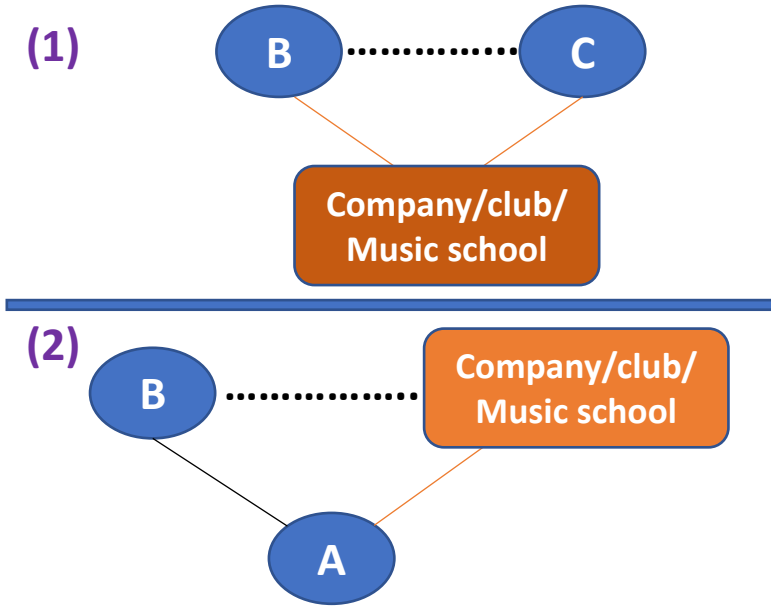
Mechanisms Underlying Homophily

- There are two mechanisms by which ties among similar people are preferentially formed: (**Homophily has 2 mechanisms for link creation**)

1. **Selection:** the tendency of people to form friendships with others with similar characteristics(that are immutable such as race or ethnicity, occupation, …).

2. **Social influence:** people may modify their behaviors, activities, interests, beliefs, and opinions (that are more mutable) to bring them more closely into alignment with the behaviors of their friends.

- **With selection,** the individual characteristics drive the formation of links, while **with social influence,** the existing links in the network serve to shape people's (mutable) characteristics.



Mechanisms Underlying Homophily

The Interplay of Selection and Social Influence

- Most of the times, both mechanisms apply and interact with each other.
- Studies show that teenage friends are similar to each other in their behaviors, and both selection and social influence apply:
  - teenagers seek social circles of people like them and
  - peer pressure causes conform to behavioral patterns within these circles

Homophily

Quiz

- Dynamics of friendships formation and behaviour of people in a network is:
  1. Impacted by neither - selection and social influence.
  2. Impacted by both, selection as well as social influence.
  3. Impacted by selection but not social influence.
  4. Impacted by social influence but not selection.

Homophily

Important questions about homophily

- An observation of homophily is often not an endpoint in itself, but rather the starting point for deeper questions, such as

1. Why the homophily is present?
2. How its underlying mechanisms will affect the further evolution of the network?
3. How these mechanisms interact with possible outside attempts to influence the behavior of people in the network.

Homophily

Assignment – Paper reading

BIRDS OF A FEATHER: Homophily in Social Networks - Miller McPherson, Lynn Smith-Lovin, and James M Cook

SOCIAL NETWORK ANALYTICS

Affiliation Networks

Prakash C O

Department of Computer Science and Engineering

SOCIAL NETWORK ANALYTICS

Affiliation Networks

Prakash C O
Department of Computer Science and Engineering

Affiliation Networks

- Studied so far:
  - Homophily groups together similar nodes
  - Selection and social influence determine the formation of links in a network
  - Similarity of nodes based on characteristics
- How to model these characteristics?
  - They represent surrounding contexts of networks
  - They exist “outside” the network
  - How to put these contexts into the network itself?

SOCIAL NETWORK ANALYTICS

Affiliation Networks

- Activities such as
  - being part of a particular company, organization, or neigborhood;
  - frequenting a particular place;
  - pursuing a particular hobby or interest

  when shared between two people, tend to increase the likelihood that they will interact and hence form a link in the social network.

- **Adopting terminology due to Scott Feld, we'll refer to such activities as foci (i.e., “focal points" of social interaction).**
- **Foci** - constitutes “social, psychological, legal, or physical entities around which joint activities are organized (e.g. workplaces, voluntary organizations, hangouts, etc.)"

Affiliation Networks

- In constructing affiliation networks, as a first step, we can represent the participation of a set of people in a set of foci using a graph as follows.
  1. **a node for each person,** and
  2. **a node for each focus,** and
  3. **connect person A to focus X by an edge if A participates in X.**
- **Example:** Figure below shows two people and two foci. The graph indicates that Anna participates in both of the foci, while Daniel participates in only one.

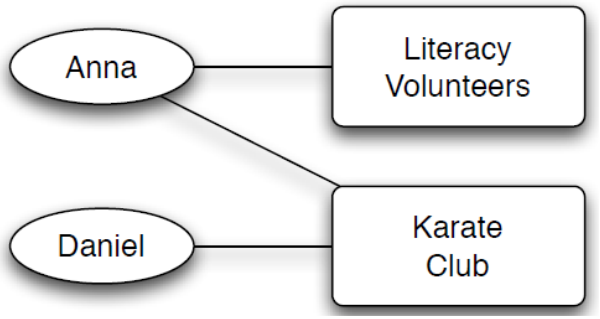
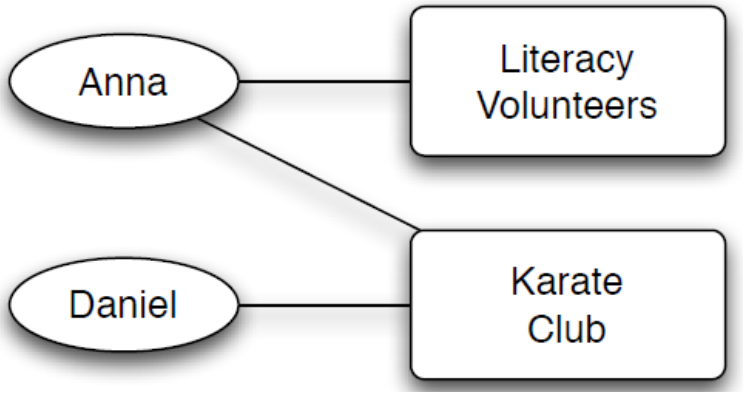


Figure 4.3: An affiliation network is a bipartite graph that shows which individuals are affiliated with which groups or activities.

Affiliation Networks



- **The representation of participation of a set of people(actors) in a set of foci(activities/events) is referred to as an affiliation network, since it represents the affiliation of people with foci.**
- More generally, **affiliation networks are examples of a class of graphs called bipartite graphs(Sociomatrix).**

Affiliation Networks

Affiliation Networks are Relational

- They show how people(actors) and activities(events) are related
- They show how activities(events) create ties among people(actors)
- They show how people(actors) create ties among activities(events)

Affiliation Networks

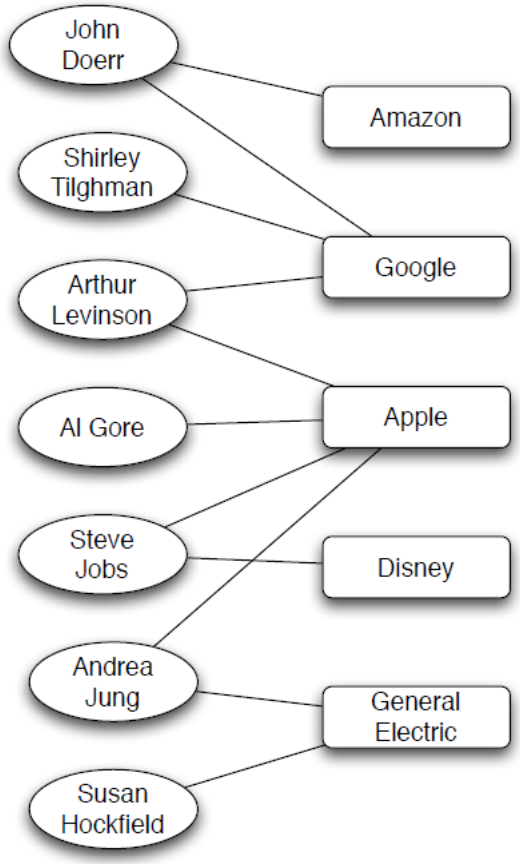


Figure : One type of affiliation network that has been widely studied is **the memberships of people on corporate boards of directors**. A very small portion of this network (as of mid-2009) is shown here. The structural pattern of memberships can reveal subtleties in the interactions among both the board members and the companies.

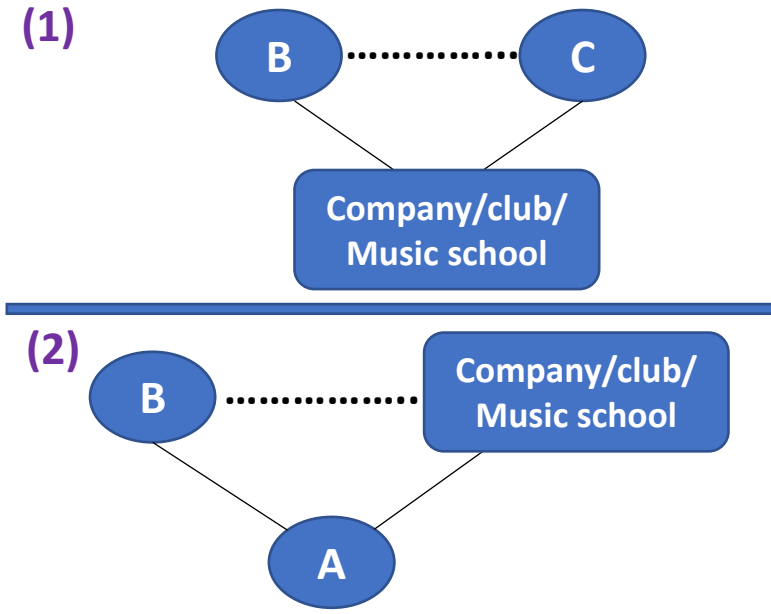
SOCIAL NETWORK ANALYTICS

Social-Affiliation Networks

- Both social networks and affiliation networks change over time:
  - new friendship links are formed, and
  - people become associated with new foci.

These changes represent a kind of co-evolution that reflects the interplay between selection and social influence:

1. if two people participate in a shared focus, this provides them with an opportunity to become friends; and
2. if two people are friends, they can influence each other's choice of foci.



SOCIAL NETWORK ANALYTICS

Social-Affiliation Networks

➢ **Social-Affiliation Networks has two distinct kinds of edges.**

1. **The first kind of edge functions as an edge in a social network:** **it connects two people and indicates friendship** (or alternatively some other social relation, like professional collaboration).
2. **The second kind of edge functions as an edge in an affiliation network**: **it connects a person to a focus** and indicates the participation of the person in the focus.

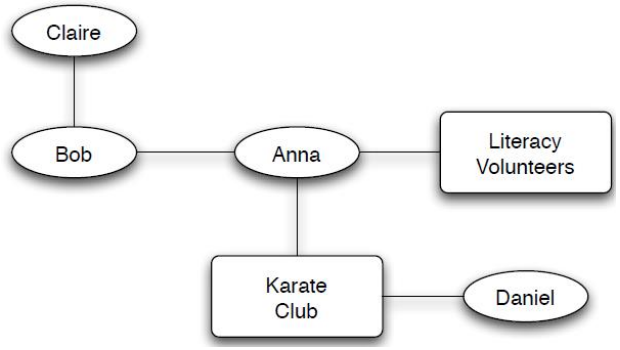


Figure : A social-affiliation network shows both the friendships between people and their affiliation with different social foci.

➢ **Social-affiliation network**, reflects the fact that it simultaneously contains

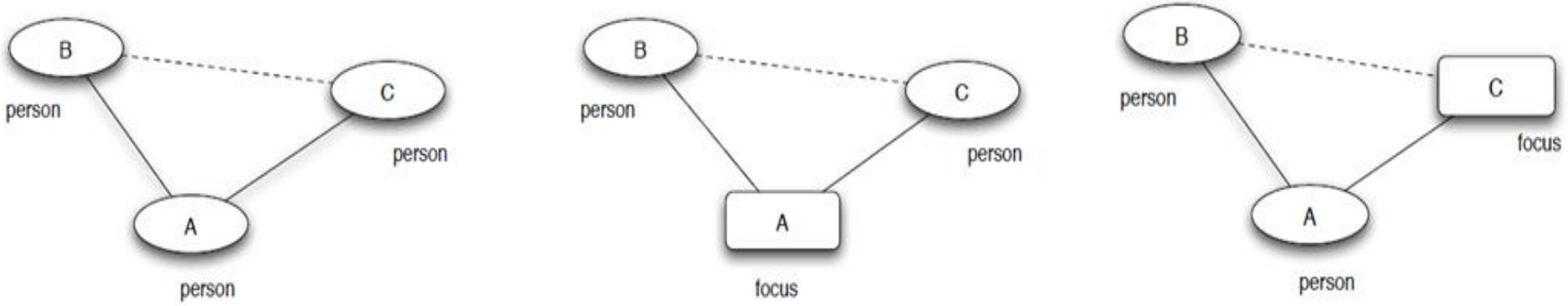
- a **social network** on the people and
- an **affiliation network** on the people and foci.

SOCIAL NETWORK ANALYTICS

Social-Affiliation Networks

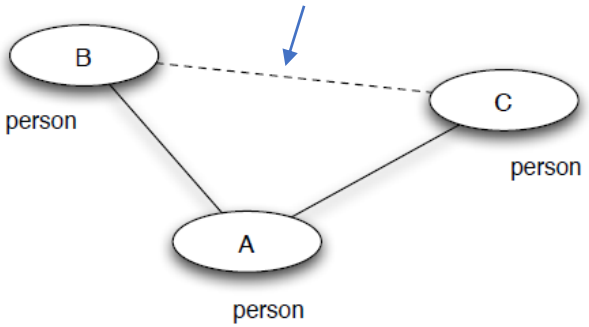
- A range of **different mechanisms for link formation** can all be viewed as types of **closure processes**, in that **they involve "closing" the third edge of a triangle in the network.**

- Suppose we have two nodes B and C with a common neighbor A in the network, and suppose that an edge forms between B and C.
  There are several interpretations for what this corresponds to, depending on whether A, B, and C are people or foci.

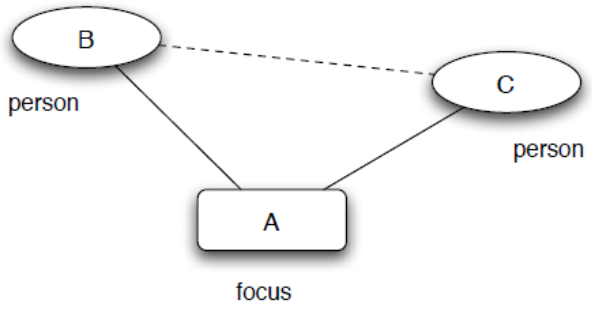


Social-Affiliation Networks

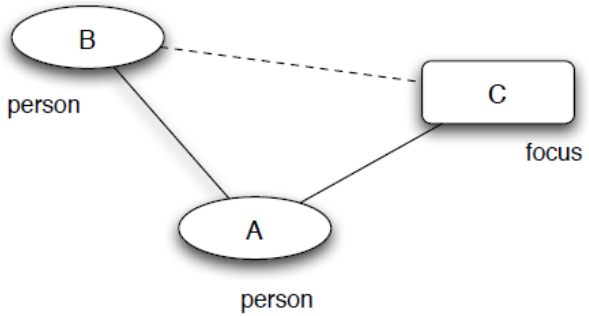
1) If A, B, and C each represent a person, then the formation of the link between B and C is **triadic closure**. (See Figure (a).)



(a) *Triadic closure*



(b) *Focal closure*

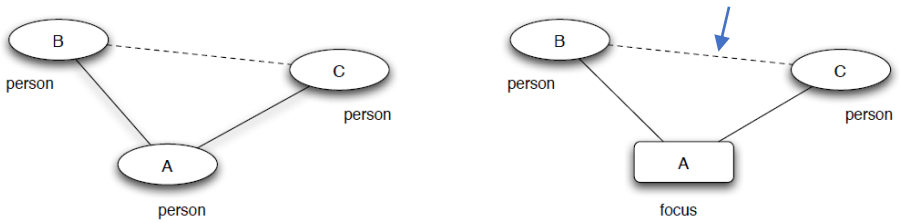


(c) *Membership closure*

Social-Affiliation Networks

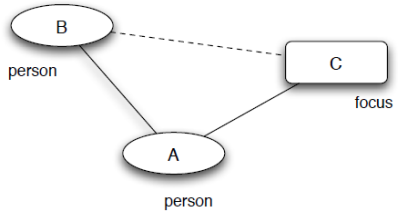
2) If B and C represent people, but A represents a focus, then this is something different: **it is the tendency of two people to form a link when they have a focus in common**. (See Figure (b).)

This is an aspect of the more general principle of selection, forming links to others who share characteristics with you. To emphasize the analogy with triadic closure, this process has been called **focal closure**.



(a) *Triadic closure*

(b) *Focal closure*

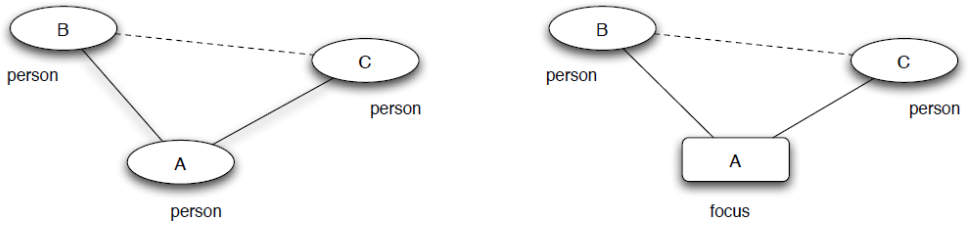


(c) *Membership closure*

Social-Affiliation Networks

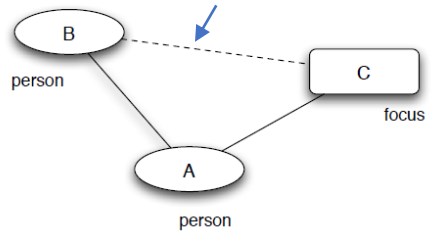
3) If A and B are people, and C is a focus, then we have the formation of a new affiliation: B takes part in a focus that her friend A is already involved in. (See Figure (c).)

This is a kind of social influence, in which B's behavior comes into closer alignment with that of her friend A. Continuing the analogy with triadic closure, we will refer to this kind of link formation as **membership closure**.



(a) *Triadic closure*

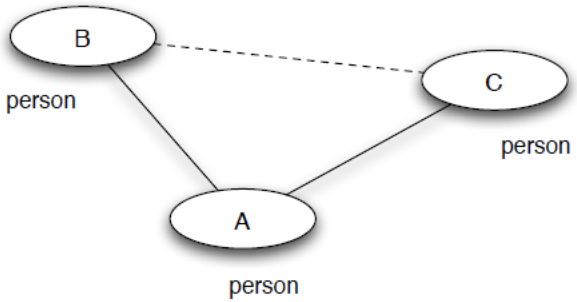
(b) *Focal closure*



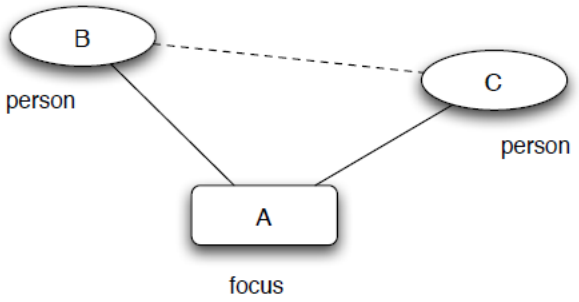
(c) *Membership closure*

Social-Affiliation Networks

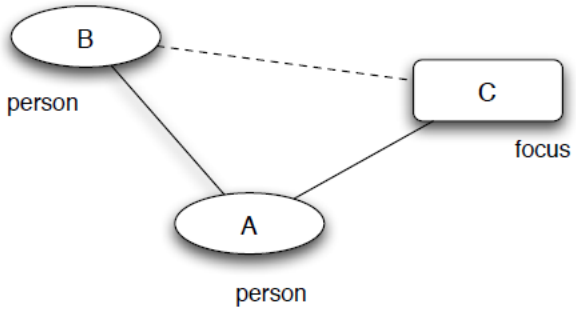
- Figure: Each of triadic closure, focal closure, and membership closure corresponds to the closing of a triangle in a social-affiliation network.



(a) *Triadic closure*



(b) *Focal closure*



(c) *Membership closure*

SOCIAL NETWORK ANALYTICS

Social-Affiliation Networks

- Figure below shows all three kinds of closure processes at work:
  1. **triadic closure** leads to a new link between Anna and Claire;
  2. **focal closure** leads to a new link between Anna and Daniel; and
  3. **membership closure** leads to Bob's affiliation with the karate club.
- Oversimplifying the mechanisms at work, they can be summarized in the following succinct way:
  1. Bob introduces Anna to Claire (two people with a friend in common)
  2. Karate introduces Anna to Daniel (two people with a focus in common)
  3. Anna introduces Bob to Karate (a person joining a focus that a friend is already involved in)

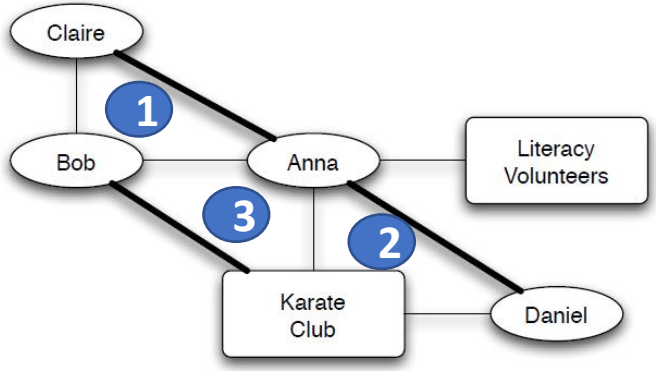


Figure: In a social-affiliation network containing both people and foci, edges can form under the effect of several different kinds of closure processes.

Quiz

- If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future. The above principle is referred as

  1. Triadic closure
  2. Foci closure
  3. Membership closure
  4. None of the above

Quiz

- If two people in a social network have a foci in common, then there is an increased likelihood that they will become friends themselves at some point in the future. The above principle is referred as

1. Triadic closure
2. Foci closure
3. Membership closure
4. None of the above

Quiz

- Consider the following two fictional cases:

  Case 1: Akshay and Abhishek become friends as they have n common friends.

  Case 2: Virender and Mahinder become friends as they have n common social foci. Assuming n as a large number, choose the correct option from the following:

  a) Case 1 and Case 2 are equally probable.

  b) Case 2 is more probable than Case 1.

  c) Case 1 is more probable than Case 2.

  d) None of the above

Social-Affiliation Networks

Assignment: Paper Reading

- Analyzing Affiliation Networks - Stephen P. Borgatti and Daniel S. Halgin
- Affiliation Networks- Silvio Lattanzi, D. Sivakumar Google Inc.

References

1. “Networks, Crowds, and Markets: Reasoning About a Highly Connected World”, D Easley and J Kleinberg, Cambridge University Press, 2010.
2. Wikipedia – Current Literature

THANK YOU

Prakash C O

Department of Computer Science and Engineering

coprakasha@pes.edu

+91 98 8059 1946