Computational Social Science

Course #04199, module 04IN2042

Markus Strohmaier

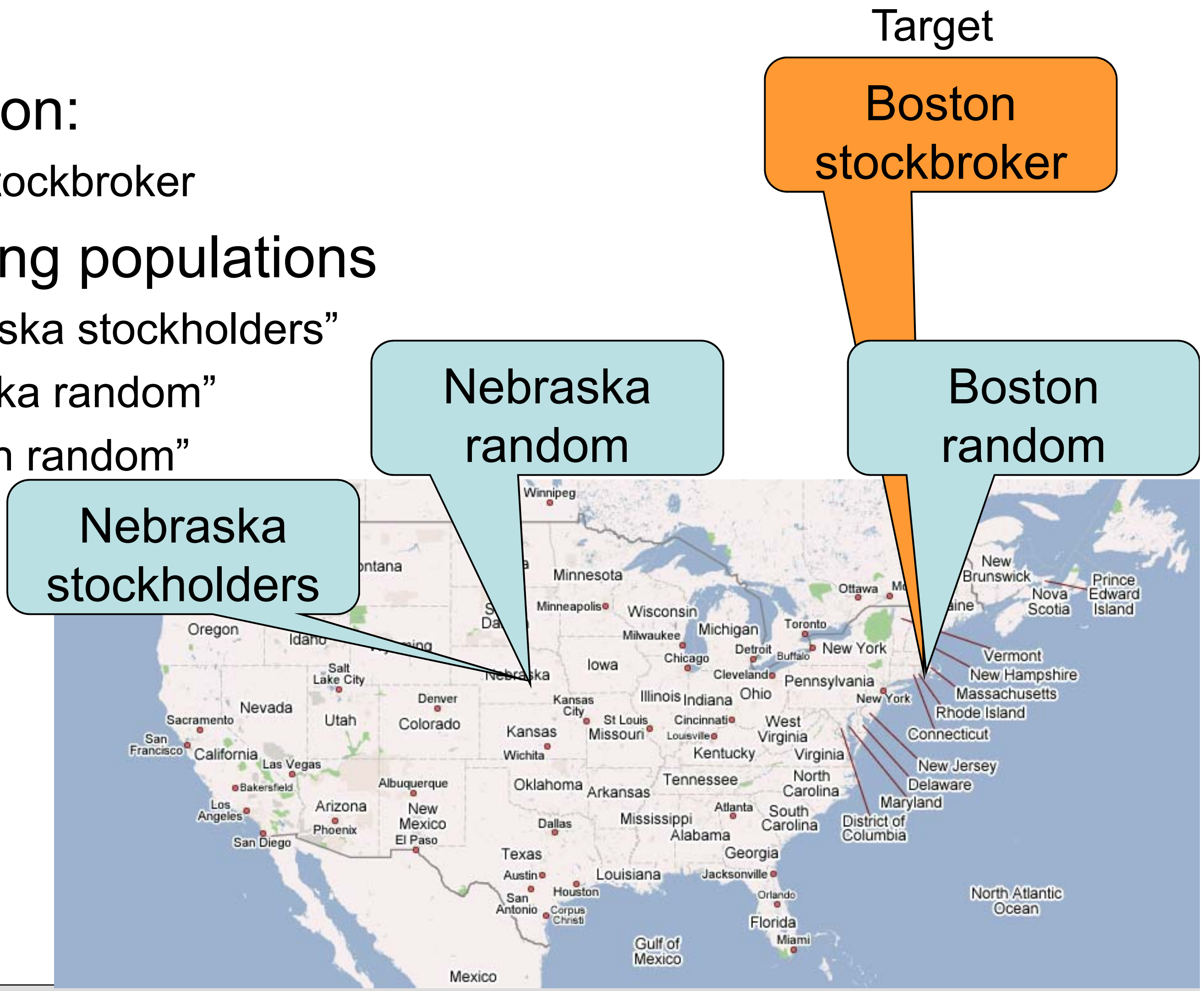
Prof. Dr.

WeST / IWVI U. of Koblenz-Landau, Germany

e-mail: markus.strohmaier@gesis.org
web: http://www.markusstrohmaier.info

A Small World

- Target person:
  - A Boston stockbroker
- Three starting populations
  - 100 “Nebraska stockholders”
  - 96 “Nebraska random”
  - 100 “Boston random”



Formalizing the Small World Problem

[Watts 2003]

Reminder - previous informal definition: SMP exists when every pair of nodes in a graph is connected by a path with an extremely small number of steps.
Does not take searchability into account. Random networks are hard to se

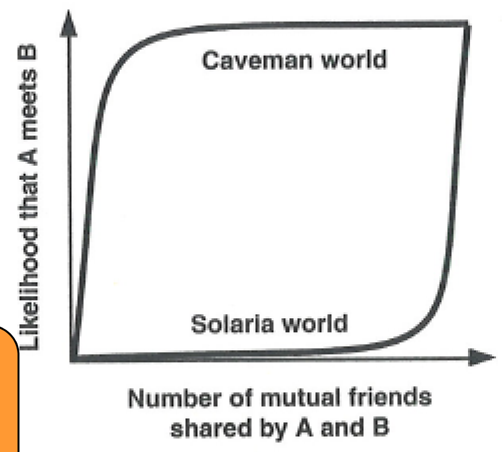


Figure 3.1. Two extreme kinds of interaction rules. In the top curve (caveman world), even a single mutual friend implies that A and B are highly likely to meet. In the bottom curve (Solaria world), all interactions are equally unlikely, regardless of how many friends A and B share.

Under which conditions can these two requirements be reconciled?

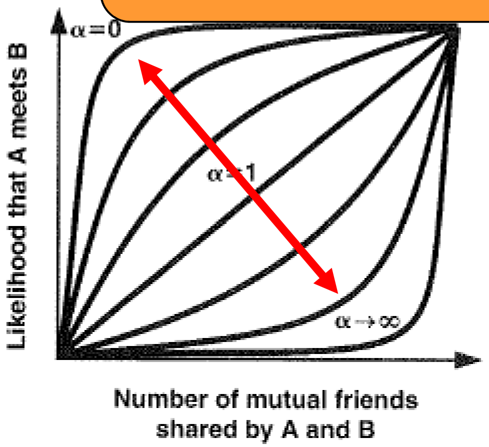


Figure 3.2. Be... two extremes, a whole family of interaction rules exists, each one specified by a particular value of the tuneable parameter alpha (α). When α = 0, we have a caveman world; when α become... we have Solaria.

Search-ability

Two seemingly contradictory requirements for the Small World Phenomenon:

- It should be possible to connect two people chosen at random via chain of only a few intermediaries (as in Solaria world)
- Network should display a large clustering coefficient, so that a node‘s friends will know each other (as in Caveman world)

Today

Agenda:

What are *social* networks?

A selection of concepts from Social Network Analysis

- The strength of weak ties
- Sociometry, adjacency lists and matrices
- Affiliation networks

But …

Isn‘t all of this an over simplification of the world of social systems?

- Ties/relationships vary in intensity
- People who have strong ties tend to share a similiar set of acquaintances
- Ties change over time
- Nodes (people) have different characteristics, and they are *actors*
- *…*

The Strength of Weak Ties [Granovetter 1973]

The strength of an interpersonal tie is a

- (probably linear) combination of
- The amount of time
- The emotional intensity
- The intimacy
- The reciprocal services which characterize the tie

Mark Granovetter,

Stanford University

Can you give examples of strong / weak ties?

The Strength of Weak Ties and Mutual Acquaintances [Granovetter 1973]

Consider:

Two arbitrarily selected individuals A and B and

The set S = C,D,E of all persons with ties to either or both of them

Hypothesis:

The stronger the tie between A and B, the larger the proportion of individuals in S to whom they will both be tied.

Theoretical corroboration:

Stronger ties involve larger time commitments – probability of B meeting with some friend of A (who B does not know yet) is increased

The stronger a tie connecting two individuals, the more similar they are

The Strength of Weak Ties [Granovetter 1973]

The forbidden triad

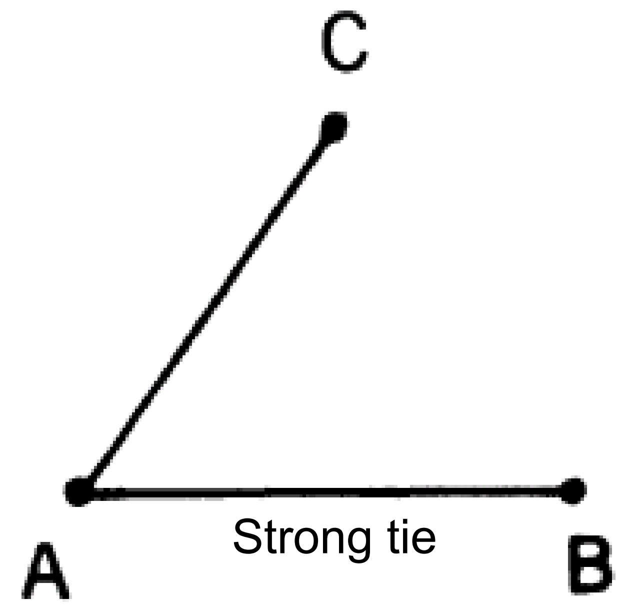


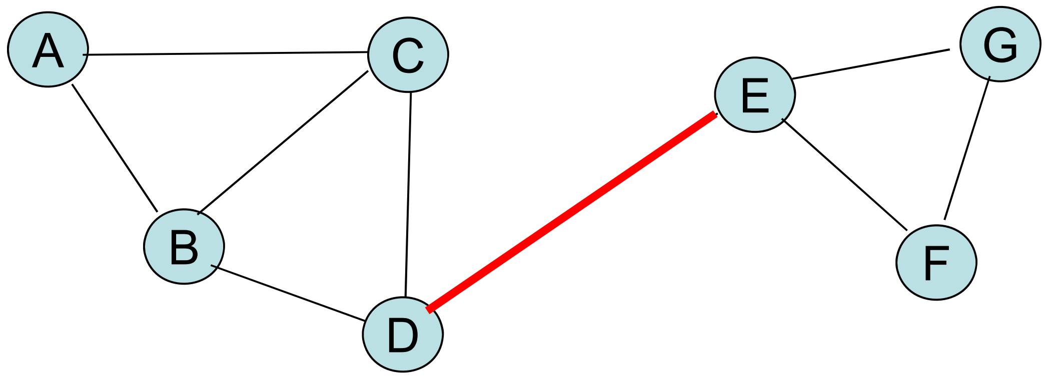
FIG. 1.—Forbidden triad

Why is it called the forbidden triad?

Bridges [Granovetter 1973]

A bridge is a line in a network which provides **the only path** between two points.

In social networks, a bridge between A and B provides the only route along which information or influence can flow from any contact of A to any contact of B



Which edge represents a bridge? Why?

Bridges and Strong Ties [Granovetter 1973]

Example:

1. Imagine the strong tie between A and B
2. Imagine the strong tie between A and C
3. Then, the forbidden triad **implies** that a tie **exists** between C and B (it forbids that a tie between C and B does not exist)

1. From that follows, that A-B is not a bridge (because there is another path A-B that goes through C)

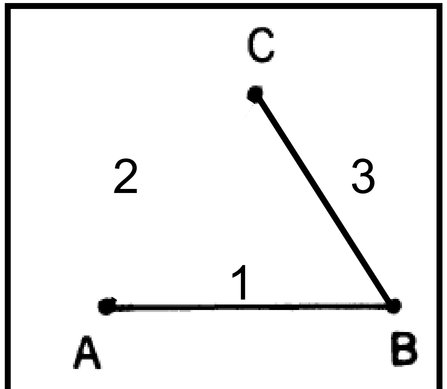


FIG. 1.—Forbidden triad

Why is this interesting?

⇨Strong ties can be a bridge ONLY IF neither party to it has any other strong ties

⇨Highly unlikely in a social network of any size

⇨Weak ties suffer no such restriction, though they are not automatically bridges

⇨But, all bridges are weak ties

In Reality ….
[Granovetter 1973]

it probably happens only rarely, that a specific tie provides the only path between two points

Local bridges: the shortest path between its two points (other than itself)

- Bridges are efficient paths
- Alternatives are more costly
- Local bridges of degree n
- A local bridge is more significant as its degree increases

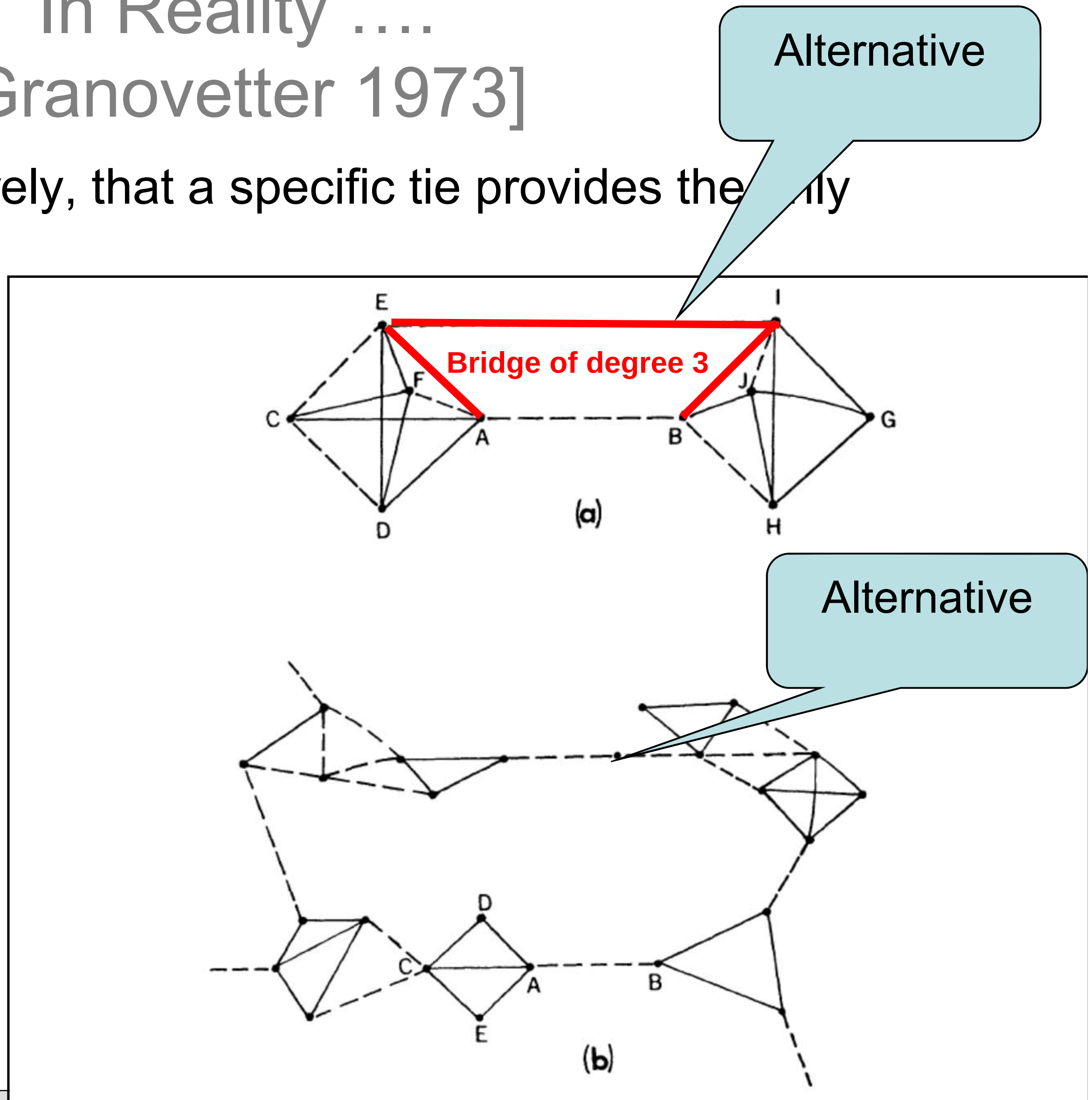


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; — — — = weak tie.

In Reality …

Strong ties can represent *local* bridges BUT
They are weak (i.e. they have a low degree)

Why?

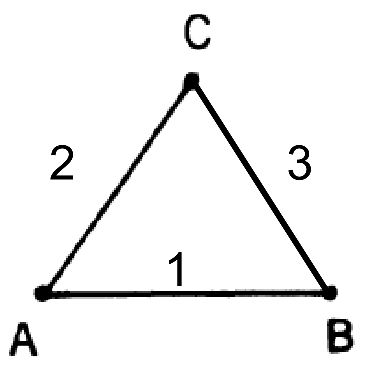


FIG. 1.—Forbidden triad

What's the degree of the local bridge A-B?

Implications of Weak Ties [Granovetter 1973]

- Those weak ties, that are local bridges, create more, and shorter paths.
- The removal of the average weak tie would do more damage to transmission probabilities than would that of the average strong one
- **Paradox**: While *weak ties* have been denounced as generative of alienation, *strong ties*, breeding local cohesion, lead to overall fragmentation

What are sources of weak ties/bridges?

Can you identify some implications for social networks on the web / for search in these networks?

How does this relate to Milgram's experiment?

Completion rates in Milgram's experiment were reported higher for acquaintance than friend relationships [Granovetter 1973]

Implications of Weak Ties [Granovetter 1973]

– Example: Spread of information/rumors in social networks
  - Studies have shown that people rarely act on mass-media information unless it is also transmitted through personal ties [Granovetter 2003, p 1274]
  - Information/rumors moving through strong ties is much more likely to be limited to a few cliques than that going via weak ones, bridges will not be crossed

How does information spread through weak ties?

Sociometry as a precursor of (social) network analysis

[Wasserman Faust 1994]

- Jacob L. Moreno, 1889 - 1974
- Psychiatrist

- born in Bukarest, grew up in Vienna, lived in the US
- Worked for Austrian Government

- Driving research motivation (in the 1930‘s and 1940‘s):
  - Exploring the advantages of picturing interpersonal interactions using sociograms, for sets with many actors

Sociometry
[Wassermann and Faust 1994]

- Sociometry is the study of positive and negative relations, such as liking/disliking and friends/enemies among a set of people.

Can you give an example of web formats that capture such relationships?

FOAF: Friend of a Friend, http://www.foaf-project.org/
XFN: **X**HTML **F**riends **N**etwork, http://gmpg.org/xfn/

- A social network data set consisting of people and measured affective relations between people is often referred to as *sociometric*.

- Relational data is often presented in two-way matrices termed *sociomatrices*.

Sociometry
[Wassermann and Faust 1994]

- Images taken from Wasserman/Faust page 76 & 82

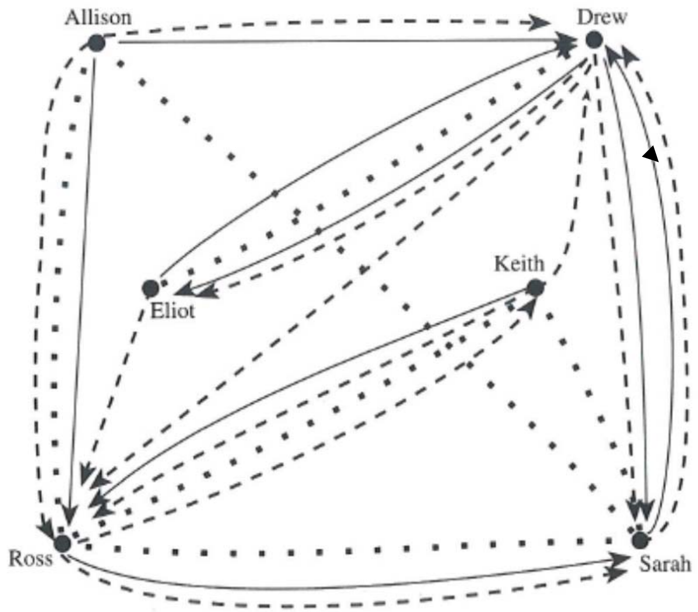


Fig. 3.2. The six actors and the three sets of directed lines — a multivariate directed graph

Table 3.1. *Sociomatrices for the six actors and three relations of Figure 3.2*

Friendship at Beginning of Year

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	0	1
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	0
Ross	0	0	0	0	-	1
Sarah	0	1	0	0	0	-

Solid lines

Friendship at End of Year

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	1	1
Eliot	0	0	-	0	1	0
Keith	0	1	0	-	1	0
Ross	0	0	0	1	-	1
Sarah	0	1	0	0	0	-

dashed lines

Lives Near

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	0	0	0	1	1
Drew	0	-	1	0	0	0
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	1
Ross	1	0	0	1	-	1
Sarah	1	0	0	1	1	-

dotted lines

Fundamental Concepts in SNA
[Wassermann and Faust 1994]

- Actor
  - Social entities
  - Def: Discrete individual, corporate or collective social units
  - Examples: people, departments, agencies
- Relational Tie
  - Social ties
  - Examples: Evaluation of one person by another, transfer of resources, association, behavioral interaction, formal relations, biological relationships
- Dyad
  - Emphasizes on a tie between two actors
  - Def: A dyad consists of two actors and a tie between them
  - An inherent property between two actors (not pertaining to a single one)
  - Analysis focuses on dyadic properties
  - Example: Reciprocity, trust

Which networks would not qualify as *social* networks?

Fundamental Concepts in SNA
[Wassermann and Faust 1994]

- Triad
  - Def: A subgroup of three actors and the possible ties among them

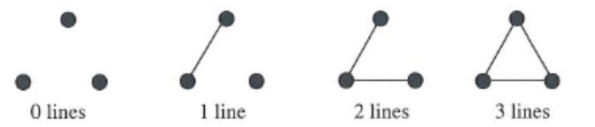
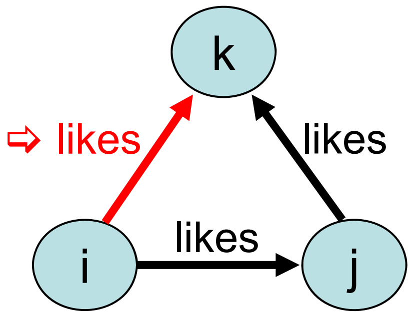
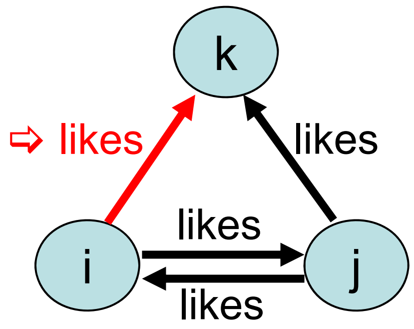


Fig. 4.3. Four possible triadic states in a graph

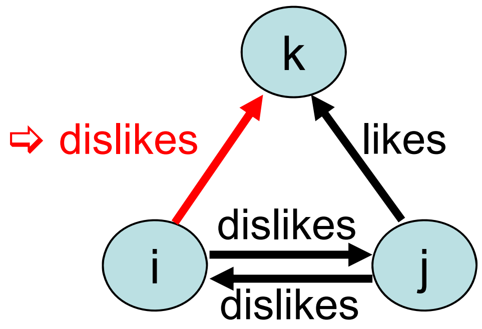
  - **Transitivity**
    - If actor i „likes“ j, and j „likes“ k, then i also „likes“ k
  - **Balance**
    - If actor i and j like each other, they should be similar in their evaluation of some k
    - If actor i and j dislike each other, they should evaluate k differently



Example 1: Transitivity



Example 2: Balance



Example 3: Balance

Fundamental Concepts in SNA

[Wassermann and Faust 1994]

- Definition of a Social Network
  - Consists of a finite set or sets of actors and the relation or relations defined on them
  - Focuses on *relational* information rather than attributes of actors

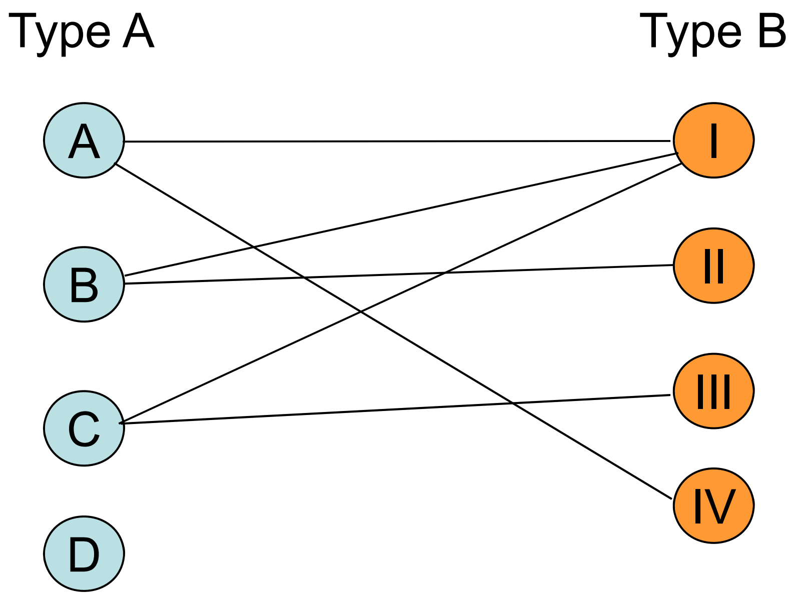
One and Two Mode Networks

[Wasserman Faust 1994]

- The number of modes refers to the **number of distinct kinds** of social entities in a network
- One-mode networks study just a **single set of actors**
- Two mode networks focus on **two sets of actors**, or on **one set of actors** and **one set of events**

Two Mode Networks

- Example:
- Two types of nodes



Examples:
actors,
scientists,
students

Examples:
conferences,
courses,
movies, articles

Can you give examples of two mode networks?

Reminder: Social Networks Examples

Affiliation Networks

- Affiliation networks are two-mode networks
  - Nodes of one type „affiliate“ with nodes of the other type (only!)
- Affiliation networks consist of subsets of actors, rather than simply pairs of actors
- Connections among members of one of the modes are based on linkages established through the second
- Affiliation networks allow to study the dual perspectives of the actors and the events

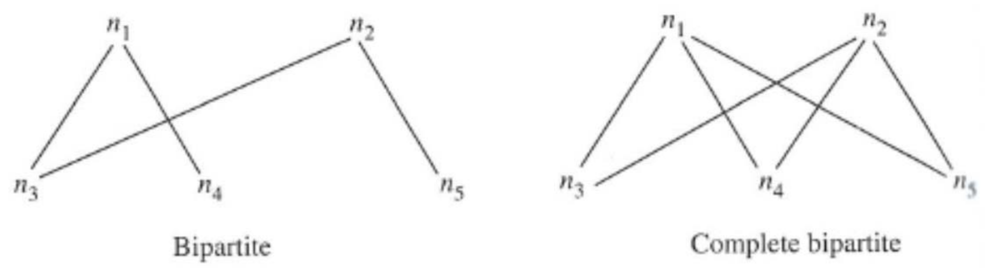
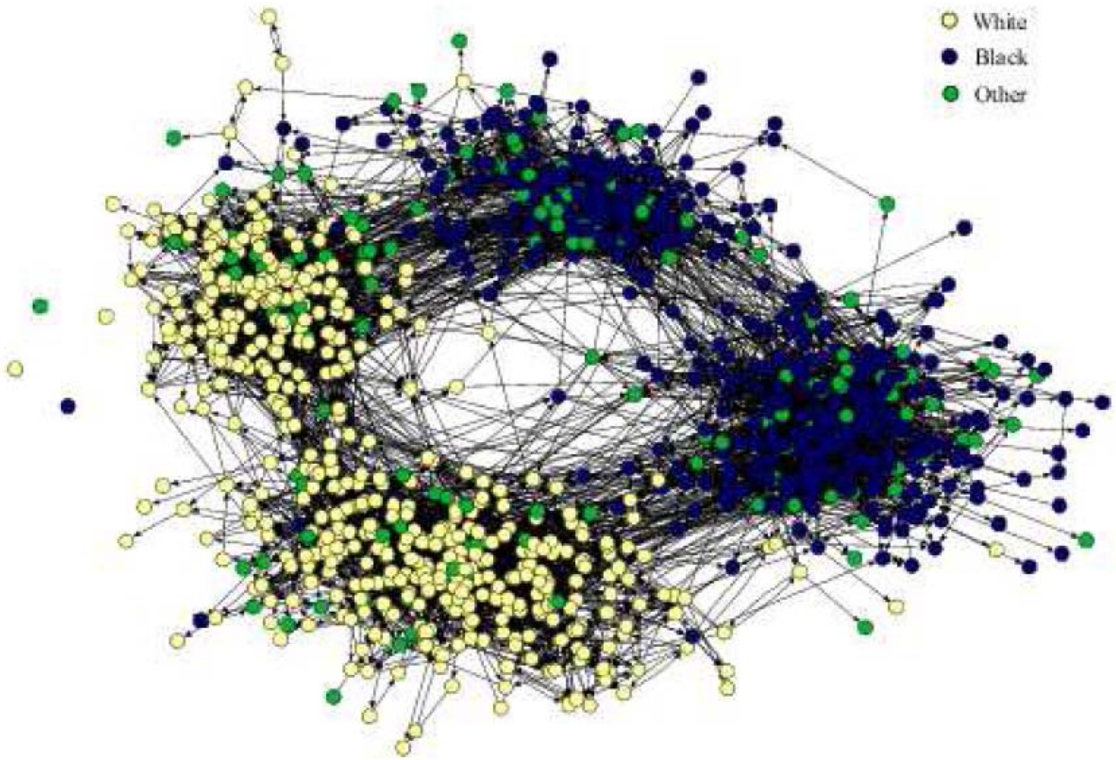


Fig. 4.15. Bipartite graphs

[Wasserman Faust 1994]

Is this an Affiliation Network? Why/Why not?



[Newman 2003]

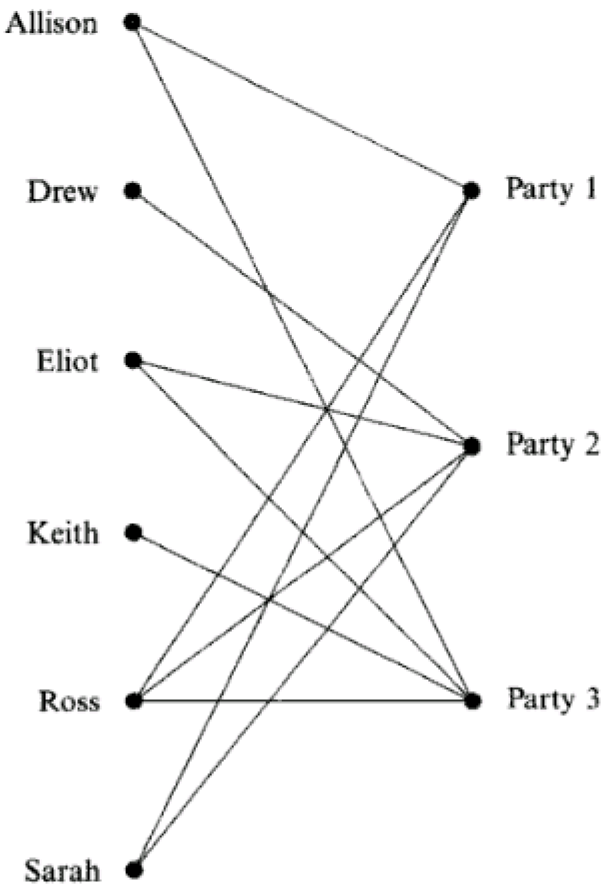
FIG. 8 Friendship network of children in a US school. Friendships are determined by asking the participants, and hence are directed, since A may say that B is their friend but not *vice versa*. Vertices are color coded according to race, as marked, and the split from left to right in the figure is clearly primarily along lines of race. The split from top to bottom is between middle school and high school, i.e., between younger and older children. Picture courtesy of James Moody.

Examples of Affiliation Networks on the Web

- Facebook.com users and groups/networks
- XING.com users and groups
- Del.icio.us users and URLs
- Bibsonomy.org users and literature
- Netflix customers and movies
- Amazon customers and books
- Scientific network of authors and articles
- etc

Representing Affiliation Networks As Two Mode Sociomatrices

[Wasserman Faust 1994]



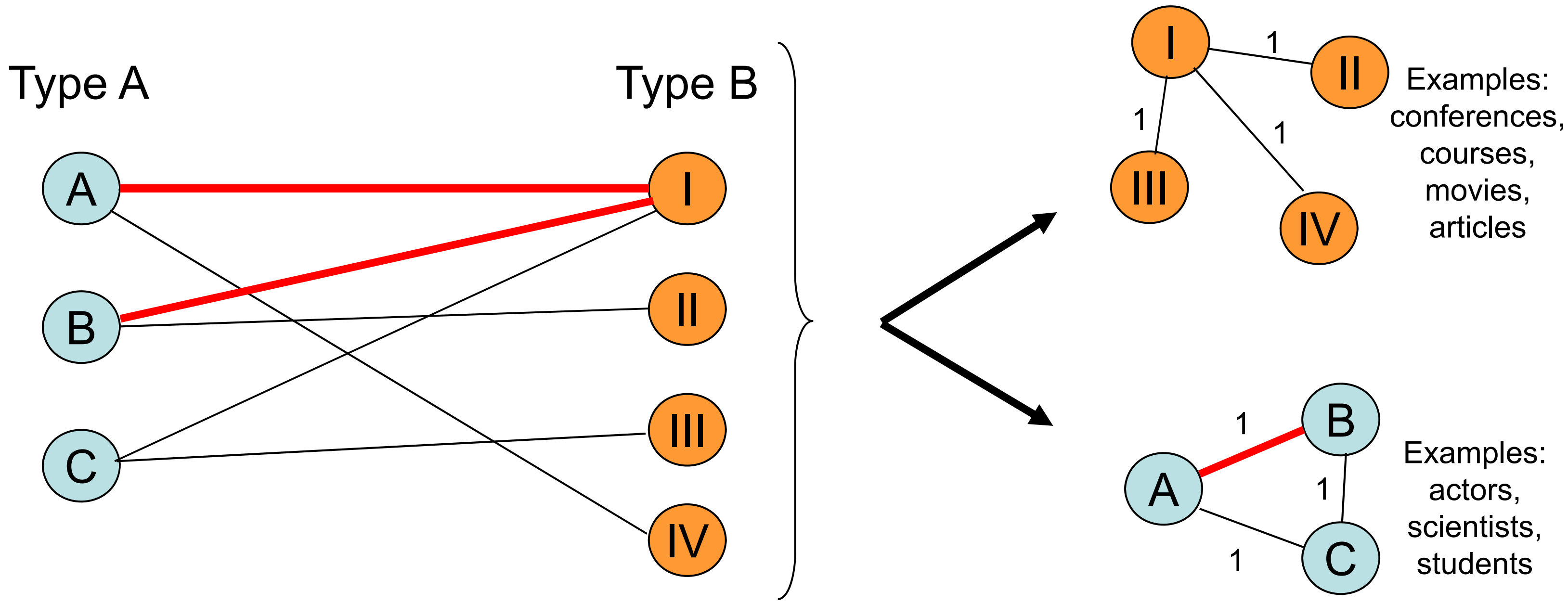
General form: $\begin{pmatrix} 0 & A \\ A' & 0 \end{pmatrix}$

	Allison	Drew	Eliot	Keith	Ross	Sarah	Party 1	Party 2	Party 3
Allison	-	0	0	0	0	0	1	0	1
Drew	0	-	0	0	0	0	0	1	0
Eliot	0	0	-	0	0	0	0	1	1
Keith	0	0	0	-	0	0	0	0	1
Ross	0	0	0	0	-	0	1	1	1
Sarah	0	0	0	0	0	-	1	1	0
Party 1	1	0	0	0	1	1	-	0	0
Party 2	0	1	1	0	1	1	0	-	0
Party 3	1	0	1	1	1	0	0	0	-

Fig. 8.3. Sociomatrix for the bipartite graph of six children and three parties

Two Mode Networks and One Mode Networks

- **Folding** is the process of transforming two mode networks into one mode networks
  - Also referred to as: **T**, ⊥ projections [Latapy et al 2006]
- Each two mode network can be folded into 2 one mode networks



Two mode network **2 One mode networks**

Transforming Two Mode Networks into One Mode Networks

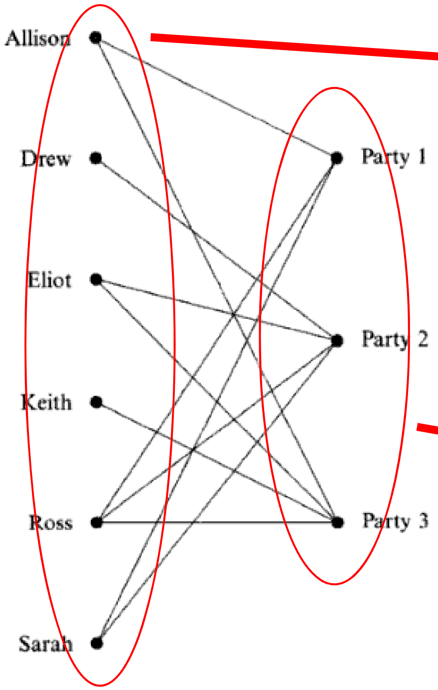
[Wasserman Faust 1994]

•Two one mode (or co-affiliation) networks (folded from the children/party affiliation network)

$\boxed{M_P = M_{PC} * M_{PC}'}$

C…Children

P…Party



	n_1	n_2	n_3	n_4	n_5	n_6
n_1	2	0	1	1	2	1
n_2	0	1	1	0	1	1
n_3	1	1	2	1	2	1
n_4	1	0	1	1	1	0
n_5	2	1	2	1	3	2
n_6	1	1	1	0	2	2

Fig. 8.5. Actor co-membership matrix for the six children

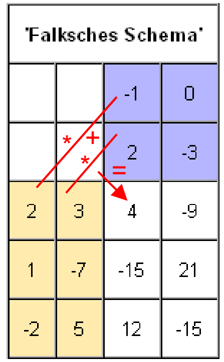
	m_1	m_2	m_3
m_1	3	2	2
m_2	2	4	2
m_3	2	2	4

Fig. 8.6. Event overlap matrix for the three parties

[Images taken from Wasserman Faust 1994]

Transforming Two Mode Networks into One Mode Networks

[Wasserman Faust 1994]



$\boxed{M_P = M_{PC} * M_{PC}'}$

C…Children

P…Party

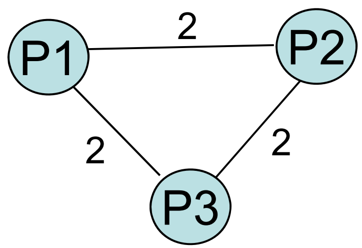
	Allison	Drew	Eliot	Keith	Ross	Sarah
Party 1	1	0	0	0	1	1
Party 2	0	1	1	0	1	1
Party 3	1	0	1	1	1	0

*

	Party 1	Party 2	Party 3
Allison	1	0	1
Drew	0	1	0
Eliot	0	1	1
Keith	0	0	1
Ross	1	1	1
Sarah	1	1	0

=

	Party 1	Party 2	Party 3
Party 1	3	2	2
Party 2	2	4	2
Party 3	2	2	4

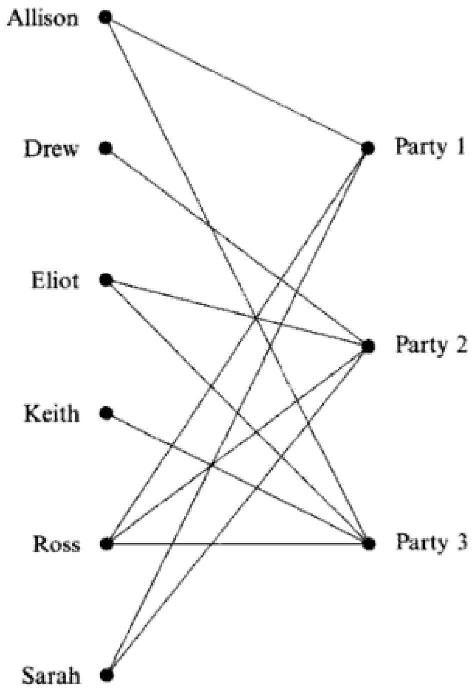


Output: Weighted regular graph

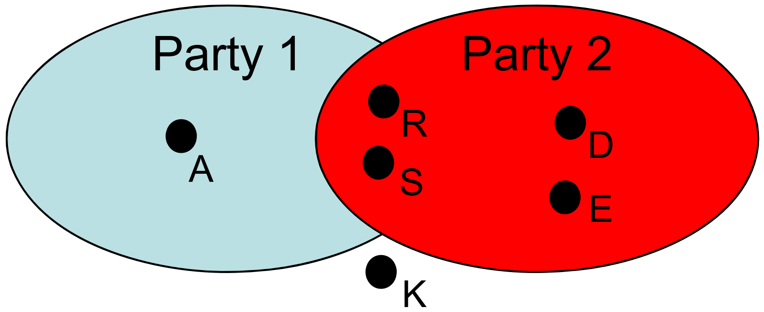
Transforming Two Mode Networks into One Mode Networks

[Wasserman Faust 1994]

Bi-partite representation (entire bipartite graph)



Set theoretic interpretation (P1, P2)



Vector interpretation (P1, P2)

	Party 1	Party 2
Allison	1	0
Drew	0	1
Eliot	0	1
Keith	0	0
Ross	1	1
Sarah	1	1

Set-theoretic/Vector-based Measures of Similiarity

[cf. Manning Schütze 1999, van Rijsbergen 1975]

Vector interpretation (P1, P2)

Party 1	Party 2	
1	0	Allison
0	1	Drew
0	1	Eliot
0	0	Keith
1	1	Ross
1	1	Sarah

counting measure | . | gives the size of the set.

Similiarity between P1 & P2:

Raw measure (or *Simple matching coefficient, result of folding*)

$|X \cap Y| = 2$

(does not take into account sizes of X or Y)

Binary Approaches (incl. Normalization)

Dice's coefficient (D)

$$2\frac{|X \cap Y|}{|X| + |Y|} \quad 2*2/(3+4) = 4/7$$

Jaccard's coefficient (J)

$$\frac{|X \cap Y|}{|X \cup Y|} \quad =2/5$$

Cosine coefficient (C)

$$\frac{|X \cap Y|}{\sqrt{|X| \times |Y|}} \quad 2/(3^{1/2} \times 4^{1/2}) = \sim 0.577$$

Overlap coefficient (O)

$$\frac{|X \cap Y|}{\min(|X|, |Y|)} \quad =2/3$$

All the left (except the raw measure) are normalized similarity measures:

1. For S = D, J, C, O, S(X,Y) = S(Y,X) and S(X; Y) =1 iff X = Y .
2. For S = D, J, C, O, $0 \leq S(X,Y) \leq 1$

[A. Badia and M. Kantardzic. Graph building as a mining activity: finding links in the small. Proceedings of the 3rd International Workshop on Link Discovery, 17--24, ACM Press New York, NY, USA,2005.]

cf. http://www.dcs.gla.ac.uk/Keith/Chapter.3/Ch.3.html

Real-valued Vectors

Manning/Schütze, 2000, 300/301

	Binäre Vektoren[1]	Vektoren mit reellen Werten[2] $\lvert\vec{x}\rvert = \sqrt{\sum_{i=1}^{n} x_i^2}$
Raw Measure	$\lvert X \cap Y \rvert$	$\sum_{k=1}^{n} (weight_{xk})(weight_{yk})$ $\vec{x}\cdot\vec{y} = \sum_{i=1}^{n} x_i y_i$
Dice-Coefficient	$\frac{2\lvert X \cap Y \rvert}{\lvert X \rvert + \lvert Y \rvert}$	$\frac{2\sum_{k=1}^{n}(weight_{xk} \cdot weight_{yk})}{\sum_{k=1}^{n} weight_{xk} + \sum_{k=1}^{n} weight_{yk}}$
Jaccard - Coefficient	$\frac{\lvert X \cap Y \rvert}{\lvert X \cup Y \rvert}$	$\frac{\sum_{k=1}^{n}(weight_{xk} \cdot weight_{yk})}{\sum_{k=1}^{n} weight_{xk} + \sum_{k=1}^{n} weight_{yk} - \sum_{k=1}^{n}(weight_{xk} \cdot weight_{yk})}$
Cosine-Coefficient	$\frac{\lvert X \cap Y \rvert}{\sqrt{\lvert X \rvert \times \lvert Y \rvert}}$	$\frac{\sum_{k=1}^{n} weight_{xk} \cdot weight_{yk}}{\sqrt{\sum_{k=1}^{n} weight_{xk}^{2}} \cdot \sqrt{\sum_{k=1}^{n} weight_{yk}^{2}}}$
Overlap-Coefficient	$\frac{\lvert X \cap Y \rvert}{\min(\lvert X \rvert, \lvert Y \rvert)}$	$\frac{\sum_{k=1}^{n} \min(weight_{xk}, weight_{yk})}{\min(\sum_{k=1}^{n} weight_{xk}, \sum_{k=1}^{n} weight_{yk})}$

[1](Manning/Schütze, 2000, 300/301)
[2](Ferber, 2003)

Home Assignment 1

- Has already been announced last week

- In case of any questions, do not hesitate to post to the newsgroup

Any questions?

See you next week!