707.000 Web Science and Web Technology „Social Network Analysis“

How can we analyze social networks?

Markus Strohmaier

Univ. Ass. / Assistant Professor
Knowledge Management Institute
Graz University of Technology, Austria

e-mail: markus.strohmaier@tugraz.at
web: http://www.kmi.tugraz.at/staff/markus

Approximate Course Schedule

	MatLab/Octave Exercises	Project Assignments
March	*Ongoing submission of home assign.!*	
April	Easter holidays	
May		
June		

We are here!
HA1.1 & 1.2 completed
HA1.3 & 1.4 this week

Administrative Issues

- HA1.1 & 1.2 due today
- HA1.3 & 1.4 will be made available this week (Mon & Wed)
- Submission deadline is MAY 3 (for both Has)

- Two lectures this week!
- Next lecture: this Wed APR 21, 12:15-13:45 HS i12
- No lecture next week

- Subsequent lectures:
- MAY 3 (Mon, regular time/Date) and
  MAY 6 (Thu, 9:30 - 11:00 HS i12)

Overview

Today's Agenda: How can we analyze social networks?

A selection of concepts from Social Network Analysis

- Sociometry, adjacency lists and matrices
- Affiliation networks
- KNC Plots
- Prominence
- Cliques, clans and clubs

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

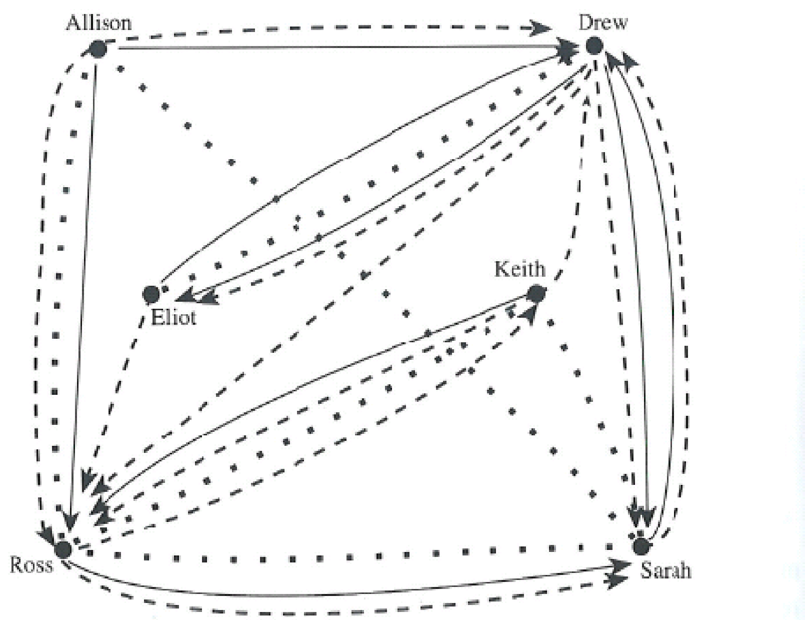


Fig. 3.2. The six actors and the three sets of directed lines — a multivariate directed graph

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

Friendship at Beginning of Year — Solid lines

	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	-

Friendship at End of Year — dashed lines

	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	-

Lives Near — dotted lines

	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	-

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

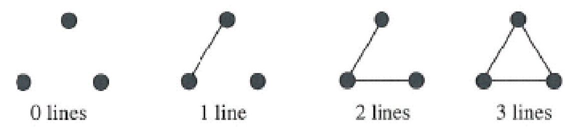
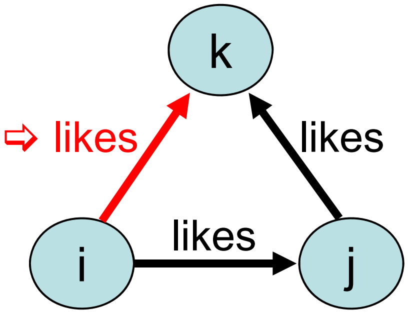
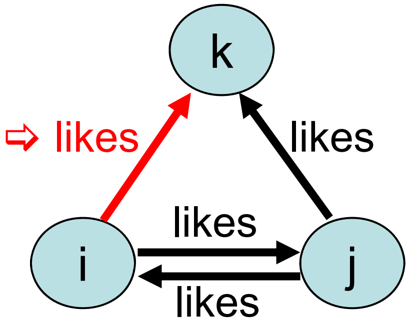


Fig. 4.3. Four possible triadic states in a graph

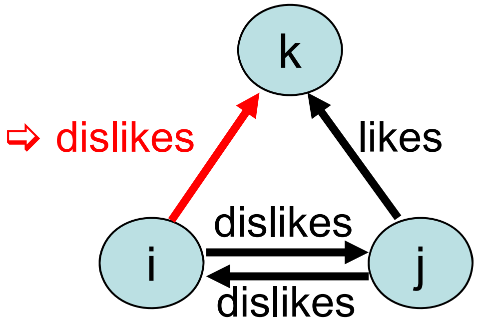
  - **Transitivity**
    - If actor i „likes“ j, and j „likes“ k, then i also „likes“ k
  - **Balance**
    - If actor i and j like each other, they should be similar in their evaluation of some k
    - If actor i and j dislike each other, they should evaluate k differently



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 mode of a network is the **number of sets of entities** on which structural variables are measured
- 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**

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

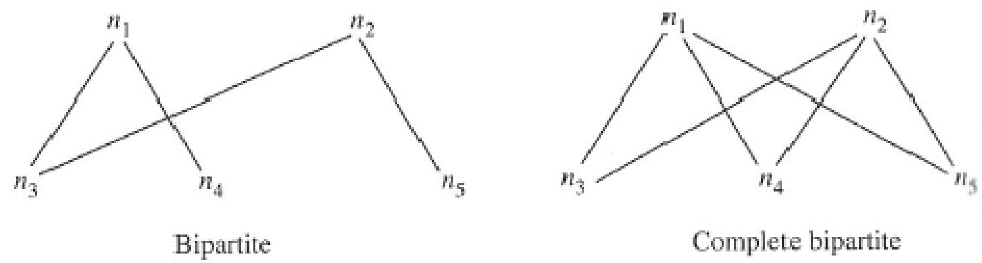
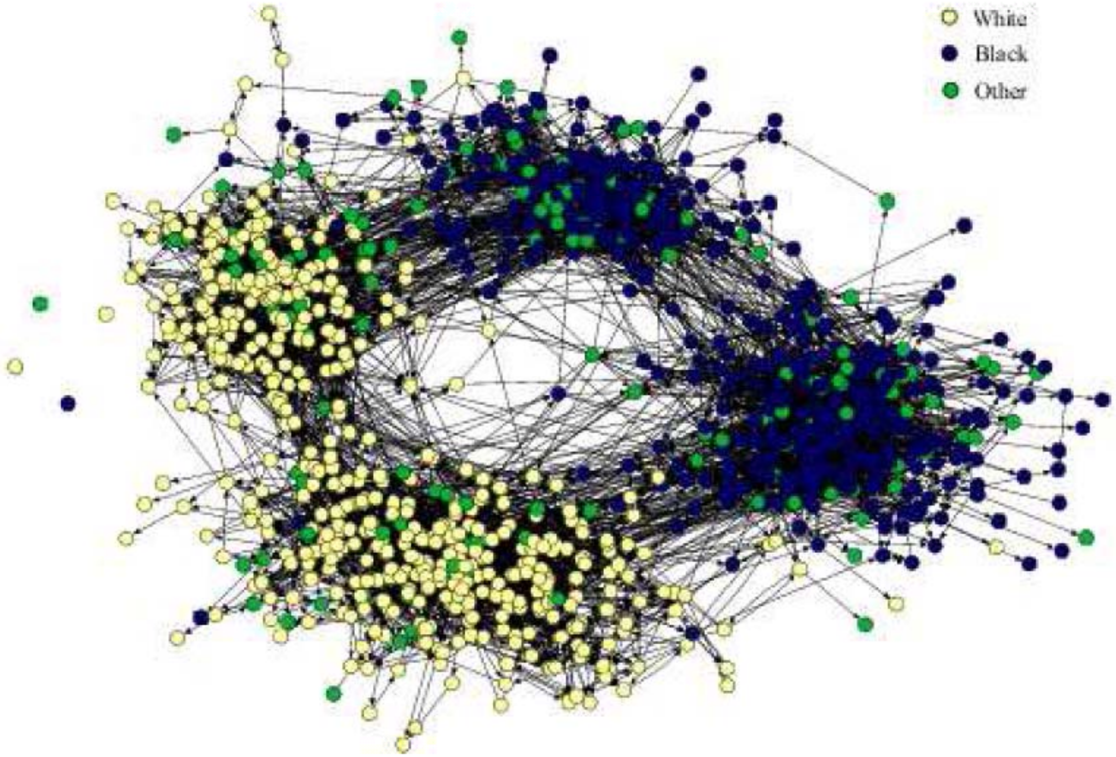


Fig. 4.15. Bipartite graphs

[Wasserman Faust 1994]

Is this an Affiliation Network? Why/Why not?



[Newman 2003]

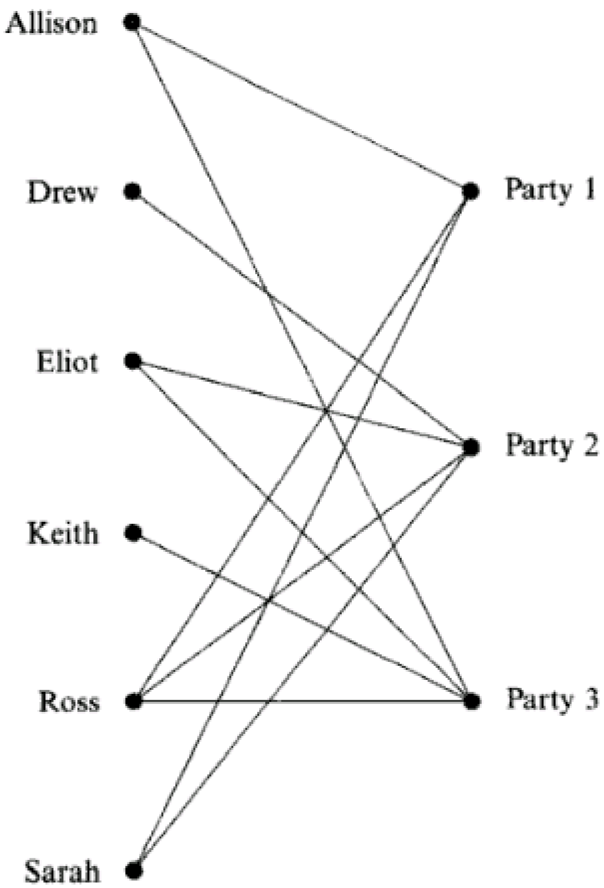
FIG. 8 Friendship network of children in a US school. Friendships are determined by asking the participants, and hence are directed, since A may say that B is their friend but not *vice versa*. Vertices are color coded according to race, as marked, and the split from left to right in the figure is clearly 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]



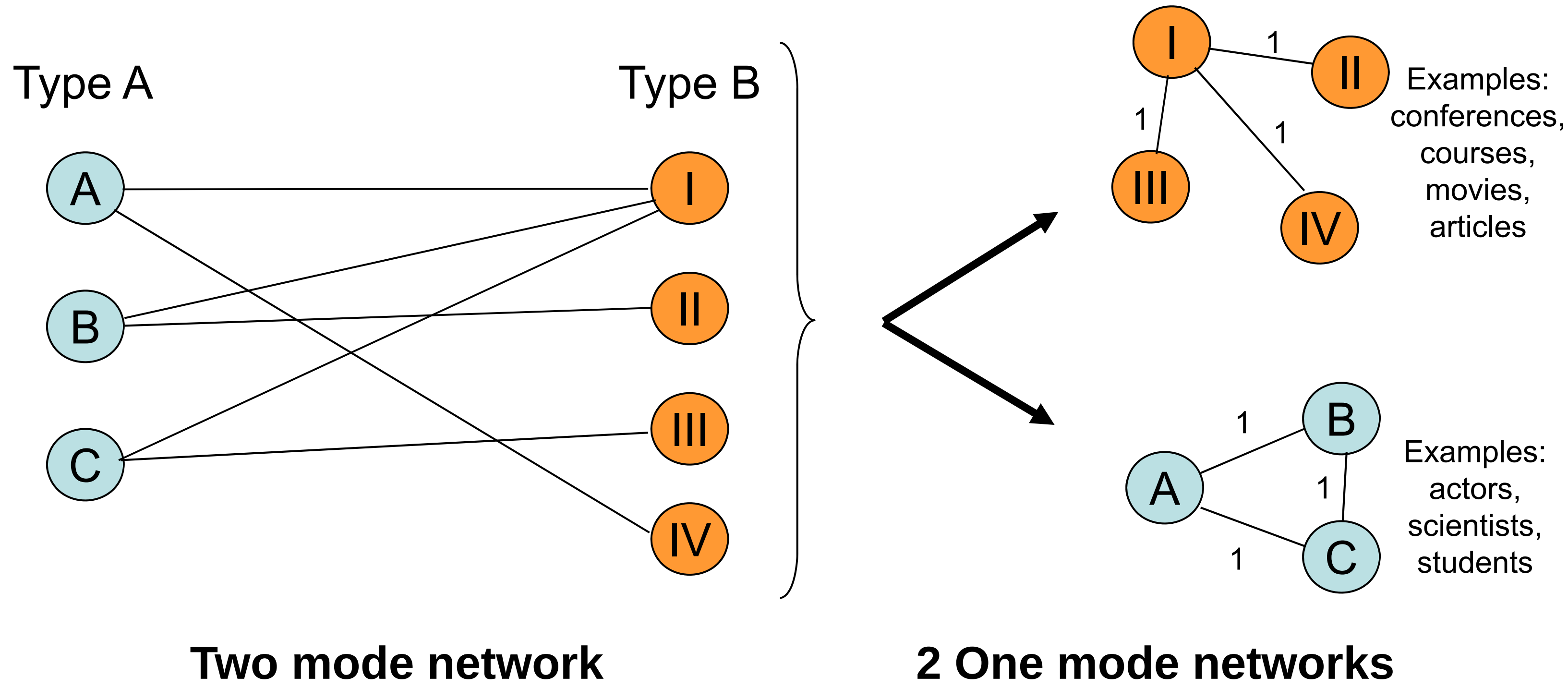
General form: $\begin{pmatrix} 0 & A \\ A' & 0 \end{pmatrix}$

	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



Transforming Two Mode Networks into One Mode Networks

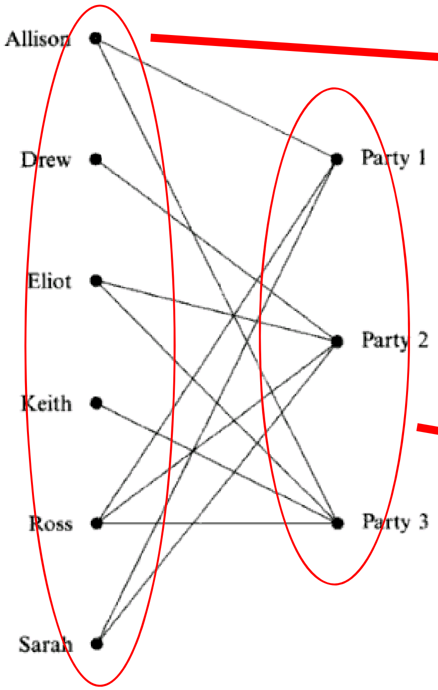
[Wasserman Faust 1994]

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

C…Children

P…Party

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



	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

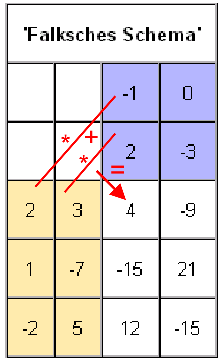
	m_1	m_2	m_3
m_1	3	2	2
m_2	2	4	2
m_3	2	2	4

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]



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

C…Children

P…Party

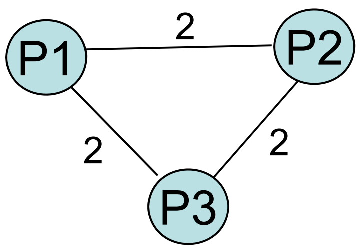
	Allison	Drew	Eliot	Keith	Ross	Sarah
Party 1	1	0	0	0	1	1
Party 2	0	1	1	0	1	1
Party 3	1	0	1	1	1	0

*

	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

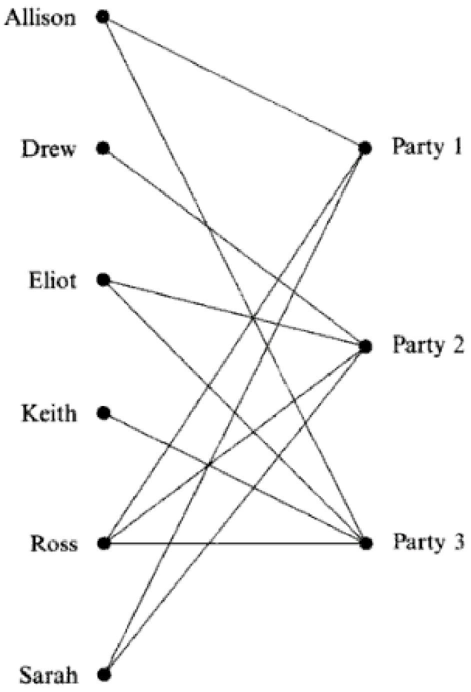


Output: Weighted regular graph

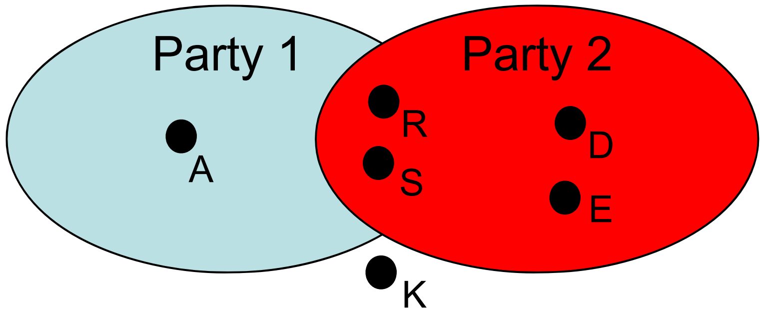
Transforming Two Mode Networks into One Mode Networks

[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 ∩ Y| = 2

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

Binary Approaches (incl. Normalization)

Dice‘s coefficient (D)

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

Jaccard‘s coefficient (J)

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

Cosine coefficient (C)

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

Overlap coefficient (O)

$$\frac{|X \cap Y|}{\min(|X|, |Y|)}$$ =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 ≤ S(X,Y) ≤ 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

$|\vec{x}| = \sqrt{\sum_{i=1}^{n} x_i^2}$

$\vec{x}\cdot\vec{y} = \sum_{i=1}^{n} x_i y_i$

	Binäre Vektoren[1]	Vektoren mit reellen Werten[2]
Raw Measure	$\lvert X \cap Y \rvert$	$\sum_{k=1}^{n}(weight_{xk})(weight_{yk})$
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)

Social Network Theoretic Measures of Similiarity

[Wasserman Faust 1994]

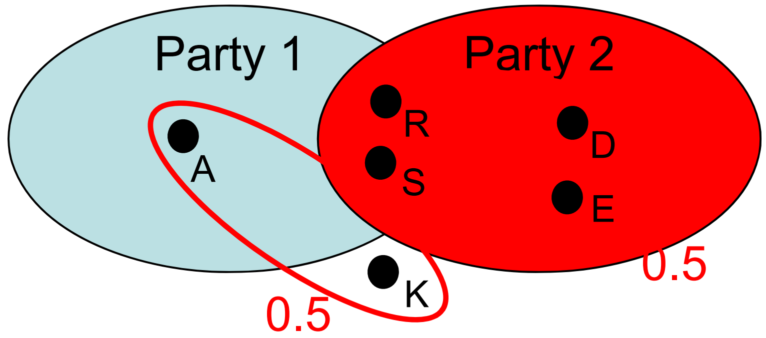
Does attendance of Party 1 have an influence on Party 2 attendance?

Taking Account of Subgroup Size

$$x_{kl}^{\mathcal{MM}} + x_{\bar{k}\bar{l}}^{\mathcal{MM}} + x_{k\bar{l}}^{\mathcal{MM}} + x_{\bar{k}l}^{\mathcal{MM}} = g.$$

	Member of m_l	Not member of m_l
Member of m_k	$x_{kl}^{\mathcal{MM}}$	$x_{k\bar{l}}^{\mathcal{MM}}$
Not member of m_k	$x_{\bar{k}l}^{\mathcal{MM}}$	$x_{\bar{k}\bar{l}}^{\mathcal{MM}}$

Set theoretic interpretation (P1, P2)



Odds ratio: θ

$$\theta_{kl} = \frac{x_{kl}^{\mathcal{MM}} / x_{\bar{k}l}^{\mathcal{MM}}}{x_{k\bar{l}}^{\mathcal{MM}} / x_{\bar{k}\bar{l}}^{\mathcal{MM}}} = \frac{x_{kl}^{\mathcal{MM}} x_{\bar{k}\bar{l}}^{\mathcal{MM}}}{x_{k\bar{l}}^{\mathcal{MM}} x_{\bar{k}l}^{\mathcal{MM}}}.$$

What is $\theta_{P1,P2}$?

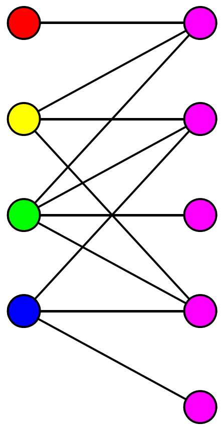
$\theta_{P1,P2} = 2*1 / 2*1 = 1$

- θ is equal to 1, if the odds of being in event P1 to not being in event P1 is the same (p=0.5) for actors in event P2 [D,E,R,S] (p=0.5) as for actors not in event P2 [A,K] (p=0.5)
- If θ is greater than 1, then actors in one event tend to also be in the other, and vice versa.
- If θ is less than 1, then actors in one event tend not to be in the other, and vice versa

Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

The k-neighborhood graph, G_k

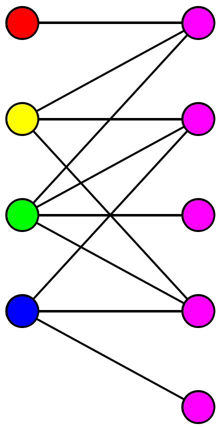
Given bipartite graph B, users on left, interests on right

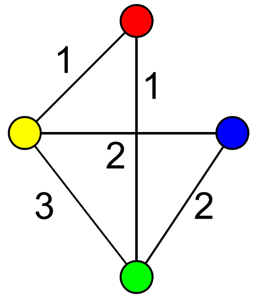


Connect two users if they share at least **k** interests in common

The k-neighborhood graph, G_k

Given bipartite graph B, users on left, interests on right





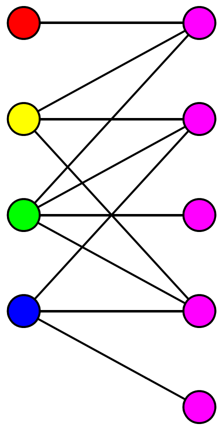
G_1

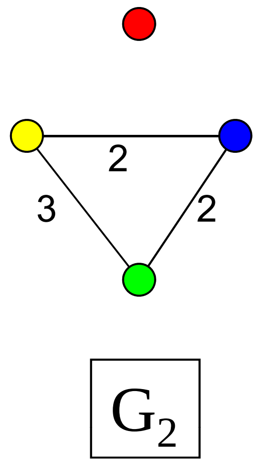
Connect two users if they share at least **k** interests in common

Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

The k-neighborhood graph, G_k

Given bipartite graph B, users on left, interests on right



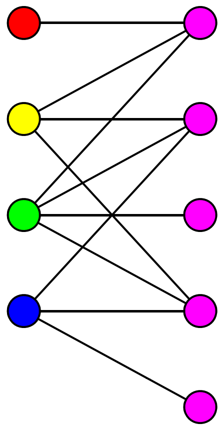


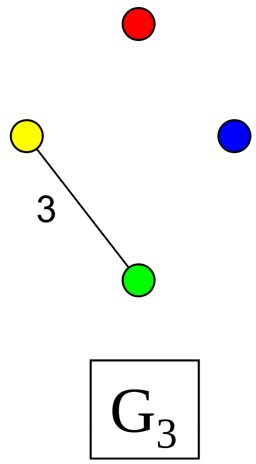
Connect two users if they share at least **k** interests in common

Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

The k-neighborhood graph, G_k

Given bipartite graph B, users on left, interests on right

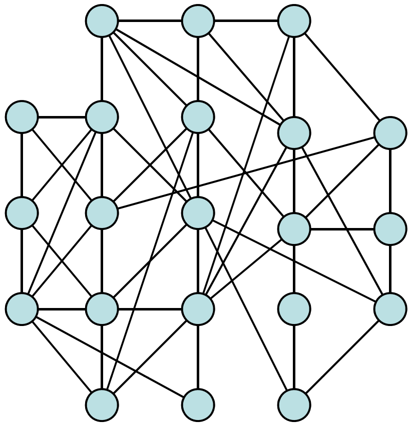




Connect two users if they share at least **k** interests in common

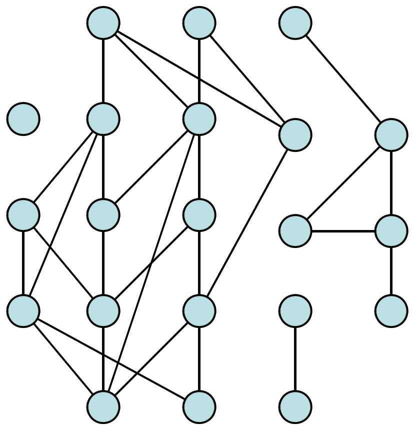
Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

Illustration k=1



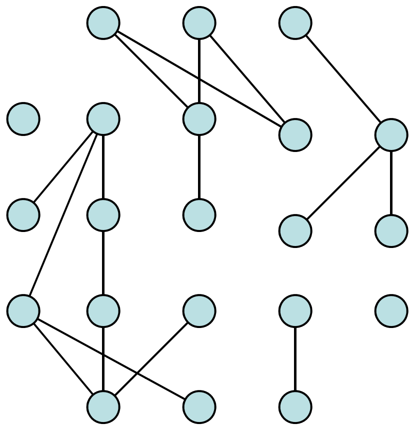
Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

Illustration k=2



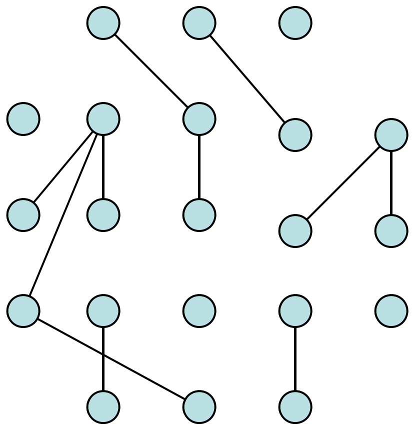
Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

Illustration k=3



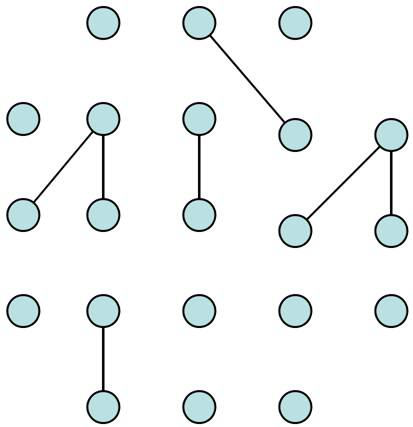
Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

Illustration k=4



Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

Illustration k=5



Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

The KNC-plot

The k-neighbor connectivity plot

- How many connected components does G_k have?
- What is the size of the largest component?

Answers the question:

how many shared interests are meaningful?

- Communities, Cuts

Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

Analysis

Four graphs:

- LiveJournal
  - Blogging site, users can specify interests
- Y! query logs (interests = queries)
  - Queries issued for Yahoo! Search (Try it at www.yahoo.com)
- Content match (users = web pages, interests = ads)
  - Ads shown on web pages
- Flickr photo tags (users = photos, interests = tags)

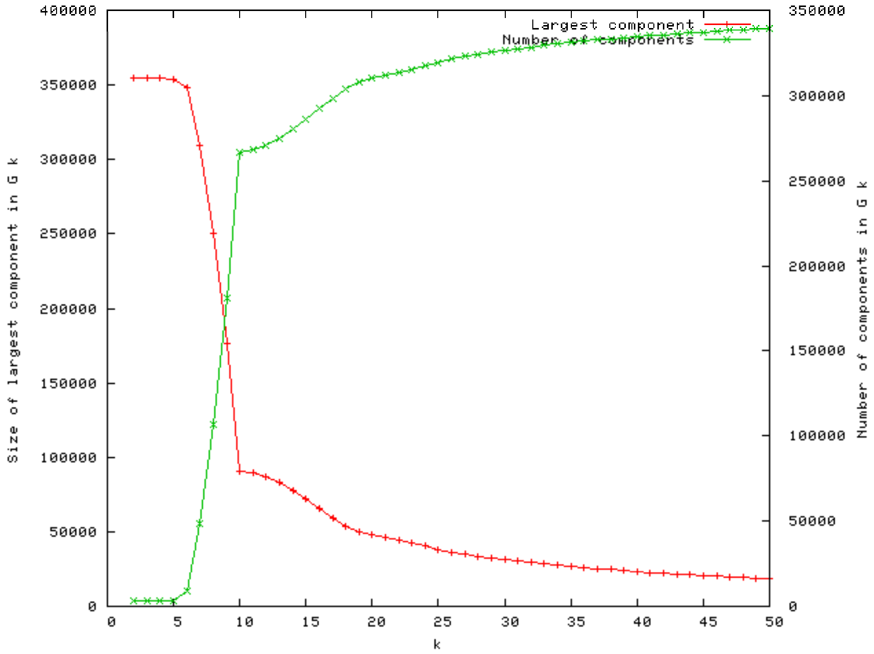
All data anonymized, sanitized, downsampled

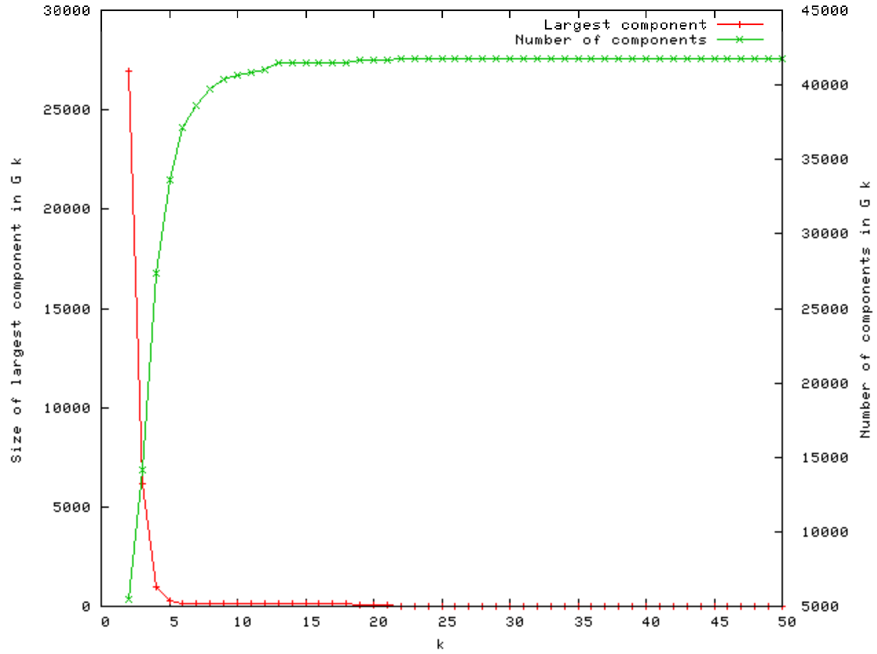
- Graphs have 100s of thousands to a million users

Slides taken from: R. Kumar and A. Tomkins and E. Vee. Connectivity structure of bipartite graphs via the KNC-plot. In Marc Najork and Andrei Z. Broder and Soumen Chakrabarti, editor(s), Proceedings of the Conference on Web Search and Data Mining, WSDM 2008, 129-138, ACM, 2008.

Examples

— Largest component
— Number of components





At k=5, all connected.
At k=6, interesting!

Content match
Web pages = "users"
Ads = "interests"

At k=6, nobody connected

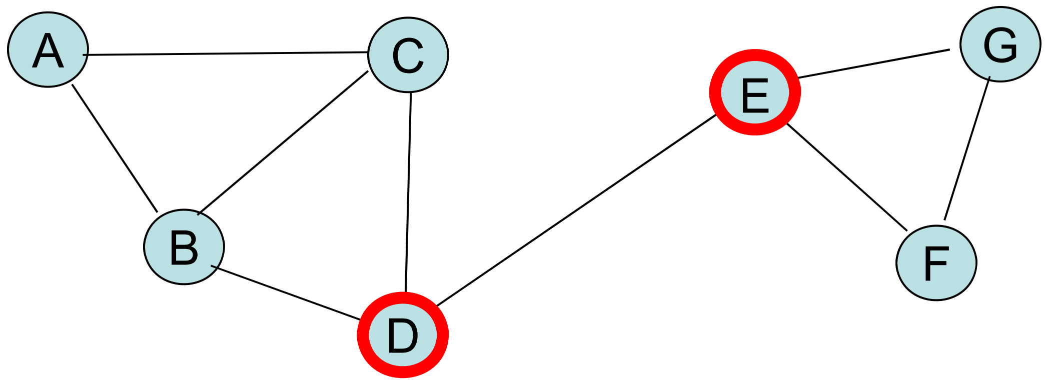
Flickr
Photos = "users"
Tags = "interests"

Cutpoint

A node, n_i, is a cutpoint if the number of components in a graph G that contains n_i is fewer than the number of components in the subgraph that results from deleting n_i from the graph.

Cutpoint or „Articulation point“

Analogous to the concept of bridges, Wasserman p113



Which node(s) represents a cutpoint? Why?

The Web Graph is Flat

Book tip

„Flatland: A romance of many dimensions“
Edwin A. Abbott 1838-1926 (1884)

http://www.geom.uiuc.edu/~banchoff/Flatland/

How can we infer information about the $n^{th}+1$ dimension?

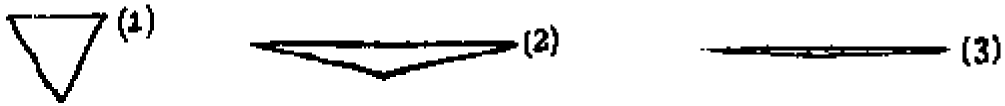
E.g. popularity, trust, prestige, importance, …

http://www.youtube.com/watch?v=BWyTxCsIXE4

http://www.flatlandthefilm.com/

Inhabitants of Flatland

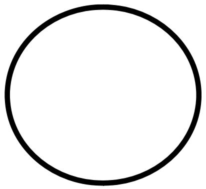
Tradesman



Men (The hero in this novel is **A. Square**)

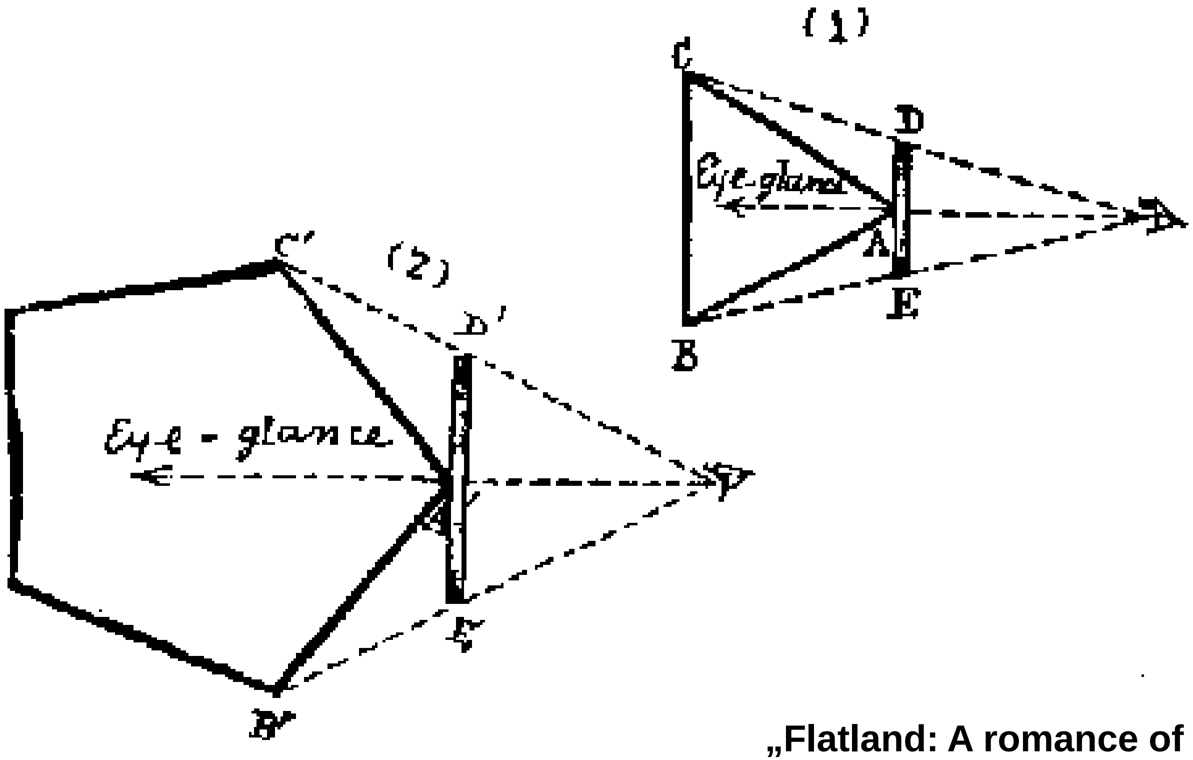
Woman

Priests



Book tip
„Flatland: A romance of many dimensions“
Edwin A. Abbott 1838-1926 (1884)
http://www.geom.uiuc.edu/~banchoff/Flatland/

Recognition by sight



Book tip
„Flatland: A romance of many dimensions“
Edwin A. Abbott 1838-1926 (1884)
http://www.geom.uiuc.edu/~banchoff/Flatland/

What kind of information can we infer from a „flat“ social graph?

Centrality and Prestige [Wasserman Faust 1994]

Which actors are the most important or the most prominent in a given social network?

What kind of measures could we use to answer this (or similar questions)?

What are the implications of directed/undirected social graphs on calculating prominence?

⇨ In directed graphs, we can use Centrality and Prestige

⇨ In undirected graphs, we can only use Centrality

Prominence
[Wasserman Faust 1994]

We will consider an actor to be prominent if the ties of the actor make the actor particularly visible to the other actors in the network.

Actor Centrality [Wasserman Faust 1994]

Prominent actors are those that are extensively involved in relationships with other actors.

This involvement makes them more visible to the others

No focus on directionality -> what is emphasized is that the actor is involved

A *central actor* is one that is involved in many ties. [cf. Degree of nodes]

Actor Prestige [Wasserman Faust 1994]

A prestigious actor is an actor who is the object of extensive ties, thus focusing solely on the actor as a recipient.

[cf. indegree of nodes]

Only quantifiable for directed social graphs.

Also known as *status, rank, popularity*

Different Types of Centrality in Undirected Social Graphs [Wasserman Faust 1994, Scripps et al 2007]

Degree Centrality

- Actor Degree Centrality:
  - *Based on degree only*

$$C_D(n_i) = \sum_j I[(i,j) \in E]$$

Where *I* is a 0=1 indicator function.

Closeness Centrality

$$C_C(n_i) = \left[\sum_{j=1}^{N} d(n_i, n_j)\right]^{-1}$$

d(*u; v*) is the geodesic distance from *u* to *v*.

- Actor Closeness Centrality:
  - Based on how close an actor is to all the other actors in the set of actors
  - Closeness is the reciprocal of the sum of all the geodesic (shortest) distances from a given node to all others
  - Nodes with a small *CC* score are closer to the center of the network while those with higher scores are closer to the edge.

Betweeness Centrality

$$C_B(n_i) = \sum_{j<k} \frac{g_{jk}(n_i)}{g_{jk}}$$

- Actor Betweeness Centrality: where *gjk* is the number of geodesic paths from *j* to *k (j,k all pairs of nodes)* and *gjk*(*ni*) is the number of geodesic paths from *j* to *k* that go through *i*.
  - An actor is central if it lies between other actors on their geodesics
  - The central actor must be between many of the actors via their geodesics

➔All three can be normalized to a value between 0 and 1 by dividing it with its max. value

Centrality and Prestige in Undirected Social Graphs [Wasserman Faust 1994]

Actor = closeness = betweenness centrality:

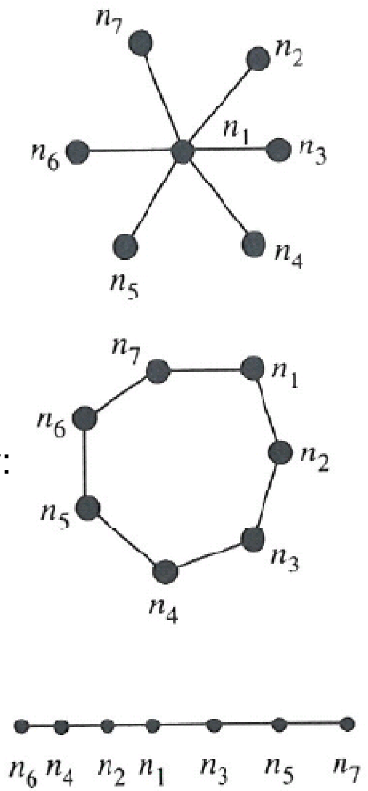
n1>n2,n3,n4,n5,n6 ,n7

Actor centrality = Betweeness centrality = Closeness centrality:

n1=n2=n3=n4=n5=n6 =n7

Betweeness centrality:

n1>n2,n3>n4,n5>n 6,n7



(a) Star graph

0	1	1	1	1	1	1
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0

(b) Circle graph

0	1	0	0	0	0	1
1	0	1	0	0	0	0
0	1	0	1	0	0	0
0	0	1	0	1	0	0
0	0	0	1	0	1	0
0	0	0	0	1	0	1
1	0	0	0	0	1	0

(c) Line graph

0	1	1	0	0	0	0
1	0	0	1	0	0	0
1	0	0	0	1	0	0
0	1	0	0	0	1	0
0	0	1	0	0	0	1
0	0	0	1	0	0	0
0	0	0	0	1	0	0

Fig. 5.1. Three illustrative networks for the study of centrality and prestige

How can we identify groups and subgroups in a social graph?

How can we identify groups and subgroups in a social graph?

Cliques, Subgroups [Wasserman Faust 1994]

What cliques can you identify in the following graph?

Definition of a Clique

- A clique in a graph is a maximal *complete* subgraph of three or more nodes.

Remark:

- Restriction to at least three nodes ensures that dyads are not considered to be cliques
- Definition allows cliques to overlap

Informally:

- A collection of actors in which each actor is adjacent to the other members of the clique

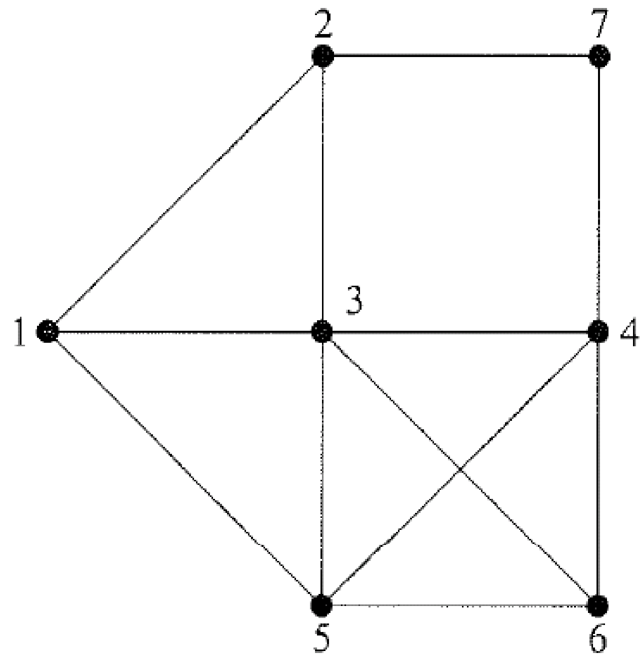


Fig. 7.1. A graph and its cliques

Subgroups [Wasserman Faust 1994]

Cliques are very strict measures

- Absence of a single tie results in the subgroup not being a clique
- Within a clique, all actors are theoretically identical (no internal differentiation)
- Cliques are seldom useful in the analysis of actual social network data because definition is overly strict

⇨ So how can the notion of cliques be extended to make the resulting subgroups more substantively and theoretically interesting?

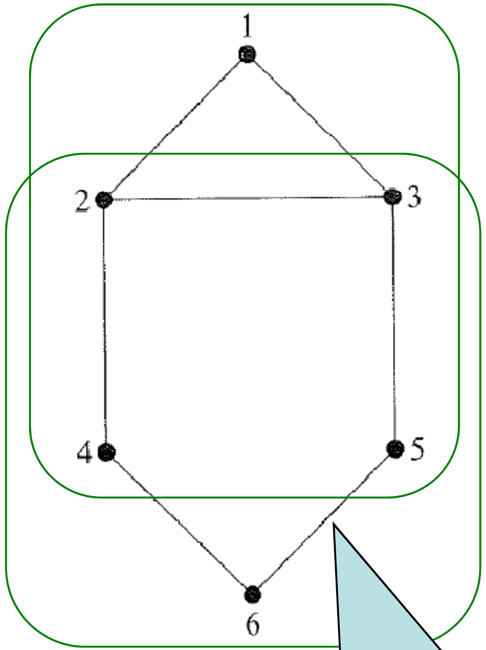
⇨ Subgroups based on reachability and diameter

n cliques
[Wasserman Faust 1994]

Which 2-cliques can you identify in the following graph?

N-cliques require that the **geodesic distances** among members of a subgroup **are small** by defining a **cutoff value n** as the maximum length of geodesics connecting pairs of actors within the cohesive subgroup.

An n-clique is a maximal ~~*complete*~~ subgraph in which the largest geodesic distance between any two nodes is no greater than n.



NOTE: Geodesic distance between 4 and 5 „goes through“ 6, a node which is not part of the 2-clique

Fig. 7.2. Graph illustrating n-cliques, n-clans, and n-clubs

n clans
[Wasserman Faust 1994]

An n-clan **is an n-clique** in which the geodesic distance between all nodes in the subgraph is no greater than n for paths **within** the subgraph.

N-clans in a graph are **those n-cliques** that have diameter less than or equal to n (within the graph).

⇨ All n-clans **are** n-cliques.

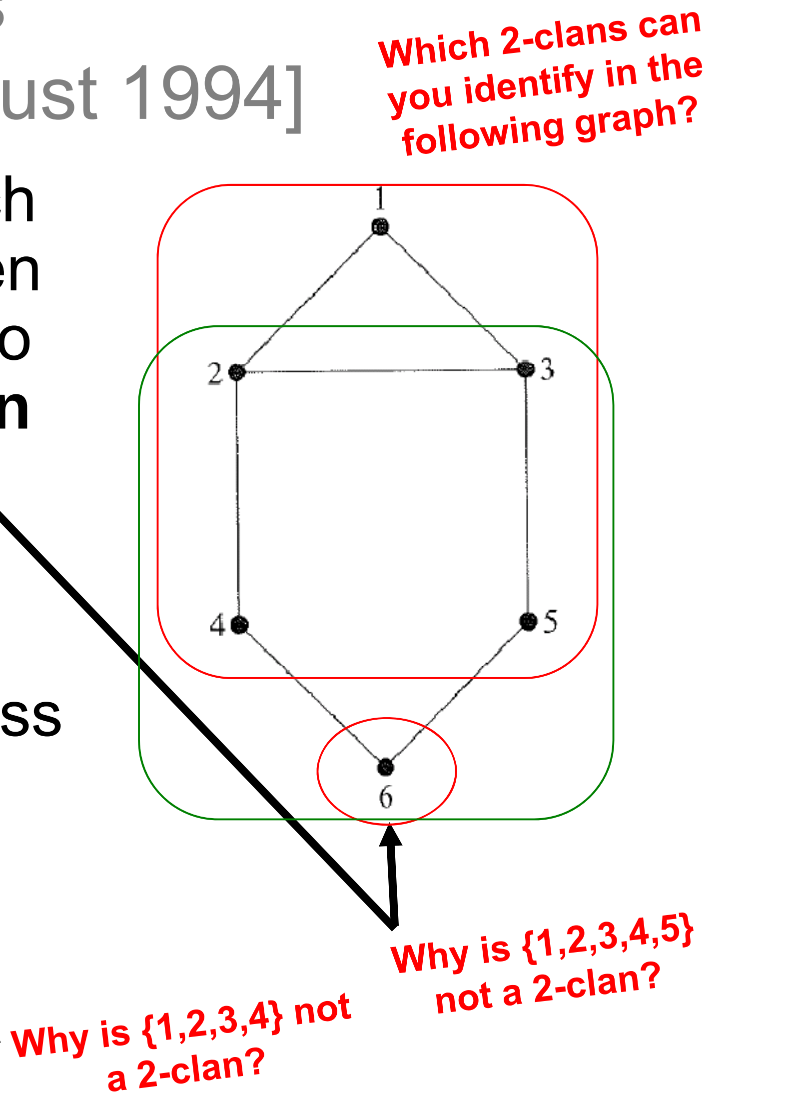


Fig. 7.2. Graph illustrating n-cliques, n-clans, and n-clubs

n clubs
[Wasserman Faust 1994]

Which 2-clubs can you identify in the following graph?

An n-club is defined as a maximal subgraph of diameter n.

No node can be added without increasing the diameter.

A subgraph in which the distance between all nodes **within the subgraph** is less than or equal to n

And no nodes can be added that also have geodesic distance n or less from all members of the subgraph

⇨ All n-clubs are **contained within** n-cliques.
⇨ All n-clans are also n-clubs
⇨ Not all n-clubs are n-clans

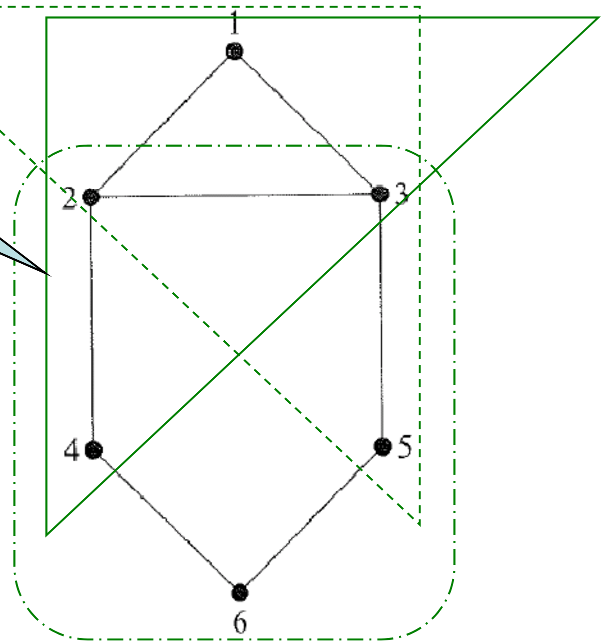


Fig. 7.2. Graph illustrating *n*-cliques, *n*-clans, and *n*-clubs

Subgroups in Co-Affiliation Networks

Borgatti 1997

- The obvious next step would be to try to identify these subgroups in co-affiliation networks.
  - For example, we can search for cliques, n-cliques, n-clans, n-clubs.
- Unfortunately, these methods are not well suited for analysing a bipartite graph.
  - In fact, bipartite graphs contain no cliques
  - In contrast, bipartite graphs contain too many 2-cliques and 2-clans.
  - One of the problems is that, in the bipartite graph, all nodes of the same type are necessarily two links distant.

➔ we need to consider special types of subgraphs which are more appropriate for two-mode data.

Bicliques [Borgatti 1997]

A biclique is a maximal complete bipartite subgraph of a given bipartite graph.

Reasonable to insist on bicliques of the form $K_{m,n}$ where m and n are greater than 2

- Why? Each of the two modes should form (after folding) interesting structures (triads or greater)

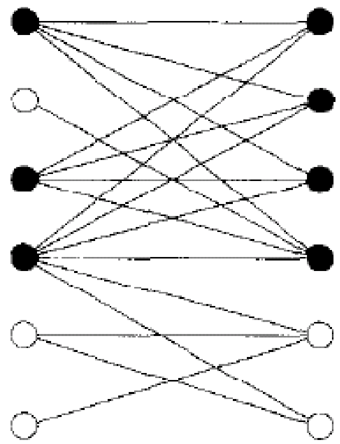
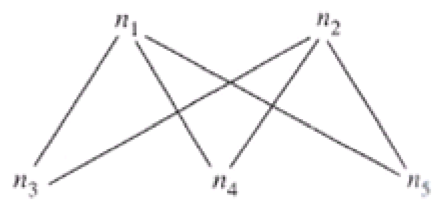


Fig. 10. Dark nodes form a biclique.



Complete bipartite

Wasserman / Faust 1994

Subgroups in Co-Affiliation Networks
Borgatti 1997

- Clearly, we can define extensions of n-cliques, n-clubs and n-clans to n-bicliques, n-biclubs and n-biclans.
- But, the extensions would in many senses be unnatural since n would need to be odd.

- **Next week we will discuss a way to analyze subgroups in affiliation networks: Galois Lattices**

Home Assignment 1.3

- Online Today

- In case of any questions, do not hesitate to post to the newsgroup tu-graz.lv.web-science

Any questions?

See you Wednesday!