„Multimedia Information Systems“
at Klagenfurt University

Guest Lecture „Social Network Analysis“

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
About me

**Education:**

- 2002 - 2004
  - PhD. in Knowledge Management, Faculty of Computer Science, TU Graz
- 1997 - 2002
  - M.Sc., Telematik, TU Graz

**Background:**

- July 2007 - present
  - Ass. Prof. (Univ.Ass.), TU Graz, Austria
- 2006 - 2007
  - 15 months Post-Doc, University of Toronto, Canada
- 2002 - 2006
  - Researcher, Know-Center, Austria
Overview

Agenda:

A selection of concepts from Social Network Analysis

- Sociometry, adjacency lists and matrices
- One mode, two mode and affiliation networks
- KNC Plots
- Prominence and Prestige
- Excerpts from Current Research „Social Web“
The Erdös Number

Who was Paul Erdös?

http://www.oakland.edu/enp/

A famous Hungarian Mathematician, 1913-1996
Erdös posed and solved problems in number theory and other areas and founded the field of discrete mathematics.

- 511 co-authors (Erdös number 1)
- ~ 1500 Publications
The Erdös Number

The Erdös Number:
Through how many research collaboration links is an arbitrary scientist connected to Paul Erdös?

What is a research collaboration link?
Per definition: Co-authorship on a scientific paper -> Convenient: Amenable to computational analysis

What is my Erdös Number?
⇒ 5

me -> S. Easterbrook -> A. Finkelstein -> D. Gabbay -> S. Shelah -> P. Erdös
(Work by one of my students, Thomas Noisternig, 2008)
43things.com

- Users
- Listing and Tagging goals

A tripartite graph

- User-Tag-Goal
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?


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

• Relational data is often presented in two-way matrices termed sociomatrices.
Sociometry
[Wassermann and Faust 1994]

![Diagram showing directed lines between six actors: Allison, Drew, Eliot, Keith, Ross, and Sarah. Solid lines, dashed lines, and dotted lines are used to represent different relations.](Images Wasserman/Faust pages 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

<table>
<thead>
<tr>
<th></th>
<th>Allison</th>
<th>Drew</th>
<th>Eliot</th>
<th>Keith</th>
<th>Ross</th>
<th>Sarah</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Friendship at Beginning of Year</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allison</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Drew</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Eliot</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Keith</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ross</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Sarah</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td><strong>Friendship at End of Year</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allison</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Drew</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Eliot</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Keith</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ross</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Sarah</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td><strong>Lines Near</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allison</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Drew</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Eliot</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Keith</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ross</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Sarah</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

Solid lines
Dashed lines
Dotted lines
How can we represent (social) networks?

We will discuss three basic forms:

- Adjacency lists
- Adjacency matrices
- Incident matrices
Adjacency Matrix (or Sociomatrix)

- Complete description of a graph
- The matrix is symmetric for nondirectional graphs
- A row and a column for each node
- Of size m x n (m rows and n columns)
Adjacency matrices

taken from http://courseweb.sp.cs.cmu.edu/~cs111/applications/Ln/lecture18.html
Adjacency lists

taken from http://courseweb.sp.cs.cmu.edu/~cs111/applications/ln/lecture18.html
Incidence Matrix

- (Another) complete description of a graph
- Nodes indexing the rows, lines indexing the columns
- g nodes and L lines, the matrix I is of size g x L
- A "1" indicates that a node $n_i$ is incident with line $l_j$
- Each column has exactly two 1's in it

Table 4.3. Example of an incidence matrix: “lives near” relation for six children

<table>
<thead>
<tr>
<th></th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
<th>$l_4$</th>
<th>$l_5$</th>
<th>$l_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$n_2$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$n_3$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$n_4$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$n_5$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$n_6$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

[Wasserman Faust 1994]
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
Fundamental Concepts in SNA
[Wassermann and Faust 1994]

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

  ![Four possible triadic states in a graph](image)

  **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]

• Social Network
  – Definition: Consists of a finite set or sets of actors and the relation or relations defined on them
  – Focus on relational information, rather than attributes of actors
One and Two Mode Networks

• 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**.
One Mode Networks

- Example: One type of nodes (Person)

Other examples: actors, scientists, students

Taken from:
http://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/bootstrapping_the_foaf_web/
Two Mode Networks

- **Example:**
- **Two types of nodes**

![Diagram of two mode networks]

Type A
- A
- B
- C
- D

Type B
- I
- II
- III
- IV

Examples:
- Type A: actors, scientists, students
- Type B: conferences, courses, movies, articles

Can you give examples of two mode networks?
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

![Bipartite and Complete Bipartite Graphs](image-url)

Fig. 4.15. Bipartite graphs

[Wasserman Faust 1994]
Is this an Affiliation Network? Why/Why not?

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

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.
- Each two mode network can be folded into 2 one mode networks.

Examples:
- Type A: conferences, courses, movies, articles
- Type B: actors, scientists, students

Two mode network 2 One mode networks
Transforming Two Mode Networks into One Mode Networks

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

\[ M_P = M_{PC} \times M_{PC}' \]

C…Children
P…Party

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

<table>
<thead>
<tr>
<th></th>
<th>( n_1 )</th>
<th>( n_2 )</th>
<th>( n_3 )</th>
<th>( n_4 )</th>
<th>( n_5 )</th>
<th>( n_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 )</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>( n_2 )</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( n_3 )</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>( n_4 )</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( n_5 )</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>( n_6 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Fig. 8.6. Event overlap matrix for the three parties

<table>
<thead>
<tr>
<th></th>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1 )</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>( m_2 )</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>( m_3 )</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

[Images taken from Wasserman Faust 1994]
Transforming Two Mode Networks into One Mode Networks

\[ M_P = M_{PC} \times M_{PC}' \]

C...Children
P...Party

<table>
<thead>
<tr>
<th></th>
<th>Allison</th>
<th>Drew</th>
<th>Eliot</th>
<th>Keith</th>
<th>Ross</th>
<th>Sarah</th>
</tr>
</thead>
<tbody>
<tr>
<td>Party 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Party 2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Party 3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{Party 1} & \quad 1 & 0 & 1 \\
\text{Party 2} & \quad 0 & 1 & 1 \\
\text{Party 3} & \quad 1 & 1 & 0 \\
\end{align*}
\]

\[
\begin{align*}
\text{Party 1} & \quad 3 & 2 & 2 \\
\text{Party 2} & \quad 2 & 4 & 2 \\
\text{Party 3} & \quad 2 & 2 & 4 \\
\end{align*}
\]

Output: Weighted regular graph
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

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

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

Connect two users if they share at least $k$ interests in common
Illustration $k=1$

Illustration k=2

Illustration $k=3$

Illustration $k=4$

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
Analysis

Four graphs:

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

All data anonymized, sanitized, downsampled

- Graphs have 100s of thousands to a million users
Examples

- Largest component
- Number of components

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

Live Journal
Users/interests

At k=6, nobody connected

Yahoo Query Logs
webpages/ads

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”

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]

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

**Closeness Centrality**
- Actor Closeness Centrality:  
  - Based on *how close an actor is to all the other actors* in the set of actors  
  - Central nodes are the nodes that have the shortest paths to all other nodes

**Betweenness Centrality**
- Actor Betweenness Centrality:  
  - 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
Centrality and Prestige in Undirected Social Graphs
[Wasserman Faust 1994]

Degree = closeness = betweenness centrality:

\[ n_1 > n_2, n_3, n_4, n_5, n_6, n_7 \]

Betweenness centrality:

\[ n_1 > n_2, n_3 > n_4, n_5 > n_6, n_7 \]

Degree centrality = Betweenness centrality = Closeness centrality:

\[ n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 \]

(a) Star graph

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(b) Circle graph

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(c) Line graph

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 5.1. Three illustrative networks for the study of centrality and prestige
Cliqués, Subgroups
[Wasserman Faust 1994]

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]

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]

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

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.
3 Two-mode networks
  - User-Goal
  - Goal-Tag
  - User-Tag

We have combined information from the
  - User-Goal and
  - Goal-Tag
2-mode networks to construct and study large-scale goal association graphs
(Work by one of my students, Thomas Noisternig, 2008)
Goal Graphs from Search Query Logs

Approach: Treat the set of all queries \{q_{-n} \ldots q_i \ldots q_n\} (n=0) within the n\textsuperscript{th} environment of the explicit intentional query q_i as tags for q_i.

With n= 6, this approach results in tagging “how to have good breast milk” with the following tags (excerpt):

- Breast milk
- [Yellow breast milk]
- [Breast feeding and going back to work]
- [Nestle formula]
- [Free nestle formula]
- [Good start]
- [What fenugreek]

[http://www.verybestbaby.com](http://www.verybestbaby.com)
[http://www.breastfeeding.com](http://www.breastfeeding.com)
Constructing Goal Graphs from Search Query Logs

• Analyzing the tripartite graph of Search
  – Consisting of users, explicit intentional queries and tags

Based on this conceptualization, the following two-mode networks can be folded into one mode networks:

- Intentional Queries – Tags
- Users – Intentional Queries
- Users - Tags
Any questions?

Thank you for your attention.