Class 5:
Affiliation networks

Course: Computational Network Analysis

Prof. Dr. Claudia Müller-Birn

Institute of Computer Science, «Human-Centered Computing»

Feb 26, 2016

Recap

- We discussed the differences of calculating the global clustering coefficient (transitivity) and the Average clustering coefficient.
- We calculated the local clustering exemplary.
- We learnt about models to create scale-free networks and small world networks.
- We talked about implications of these models.

Today's outline

- Examples of affiliation networks
- Basis concepts and terminology
- Presenting affiliation networks
- Converting two-mode networks into one-mode networks
- Analyzing affiliation networks

Motivation

Imagine…

- You as a researcher collect data on:
  - which students in a university belong to which campus organizations,
  - which employees in an organization participate in which electronic discussion forums,
  - which users participate in threads on a Mailinglist
  - which users on Facebook.com participate in specific groups
  - which users use what tags for literature on Bibsonomy.org
  - which users watch what movies on Netflix
  - which customers by what books on Amazon
  - which scientist write which articles together
  - etc.

How can you deal with these data?

One vs. 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

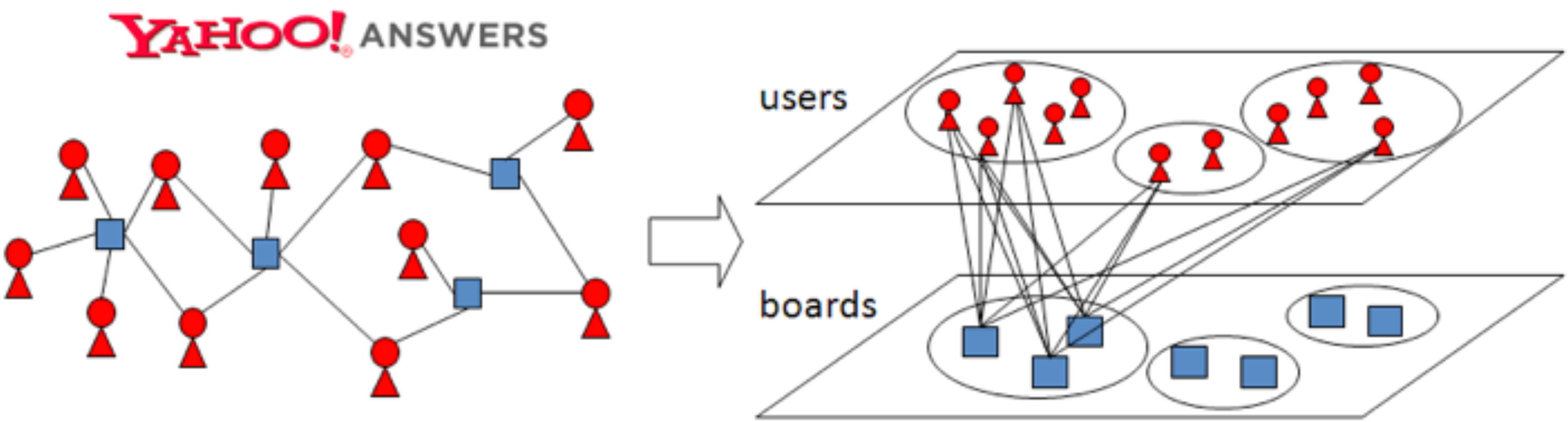
- **Two-mode networks** focus on two sets of actors, or on one set of actors and one set of events

- These kinds of data are often referred to as **affiliations**.

Wasserman, S. & Faust, K. 1997, Social network analysis: methods and applications, Cambridge Univ. Press, Cambridge.

Affiliation 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



Other Examples

	"Actor"	"Events"
Network	**Vertex**	**Group**
Film actors	Actor	Cast of a film
Coauthorship	Author	Authors of an article
Board of directors	Director	Board of a company
Social events	People	Participants of a social event
Recommender system	People	Those who like a book, film, etc.
Keyword index	Keywords	Pages where words appear
Rail connections	Stations	Train routes
Metabolic reactions	Metabolites	Participants in a reaction

Basic Concepts

Terminology - ways and modes

- The ways of a matrix are its dimensions, as in rows and columns (2-way), or rows, columns and levels (3-way)
- Modes of a matrix are the distinct sets of entities pointed to by the ways

	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	1	0	0	1
John	0	1	0	0
Larry	1	0	1	0

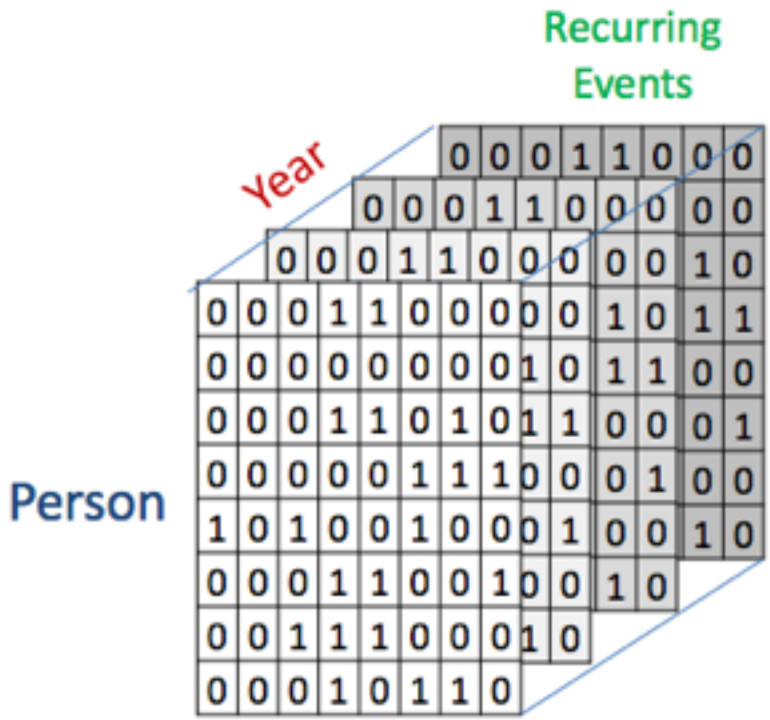
2-way, 1-mode

	Event 1	Event 2	Event 3	Event 4
EVELYN	1	1	1	1
LAURA	1	1	1	0
THERESA	0	1	1	1
BRENDA	1	0	1	1
CHARLO	0	0	1	1
FRANCES	0	0	1	0
ELEANOR	0	0	0	0
PEARL	0	0	0	0
RUTH	0	0	0	0
VERNE	0	0	0	0
MYRNA	0	0	0	0

2-way, 2-mode

Steve Borgatti (2009) The analysis of 2-mode networks. Keynote Conference and Workshop on Two-Mode Social Network Analysis

Terminology - ways and modes



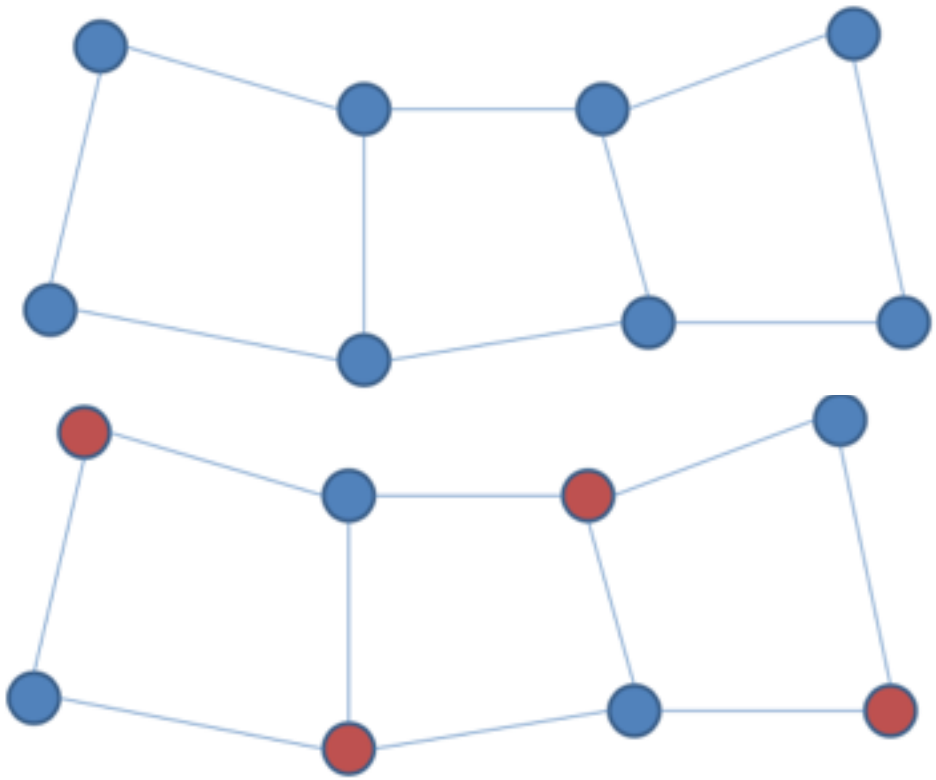
3-way, 3-mode
Longitudinal affiliations data

Steve Borgatti (2009) The analysis of 2-mode networks. Keynote Conference and Workshop on Two-Mode Social Network Analysis

Bipartite graph

- A graph G(V,E) is bipartite if V can be partitioned into V1, V2 such that for all edges (u,v) in E, u belongs to V1 and v belongs to V2
- There are no ties between V1 and V2 (independent sets), and V1+V2 = V
- Bipartite and 2-mode are not interchangeable
- “2-mode network” terminology is misleading

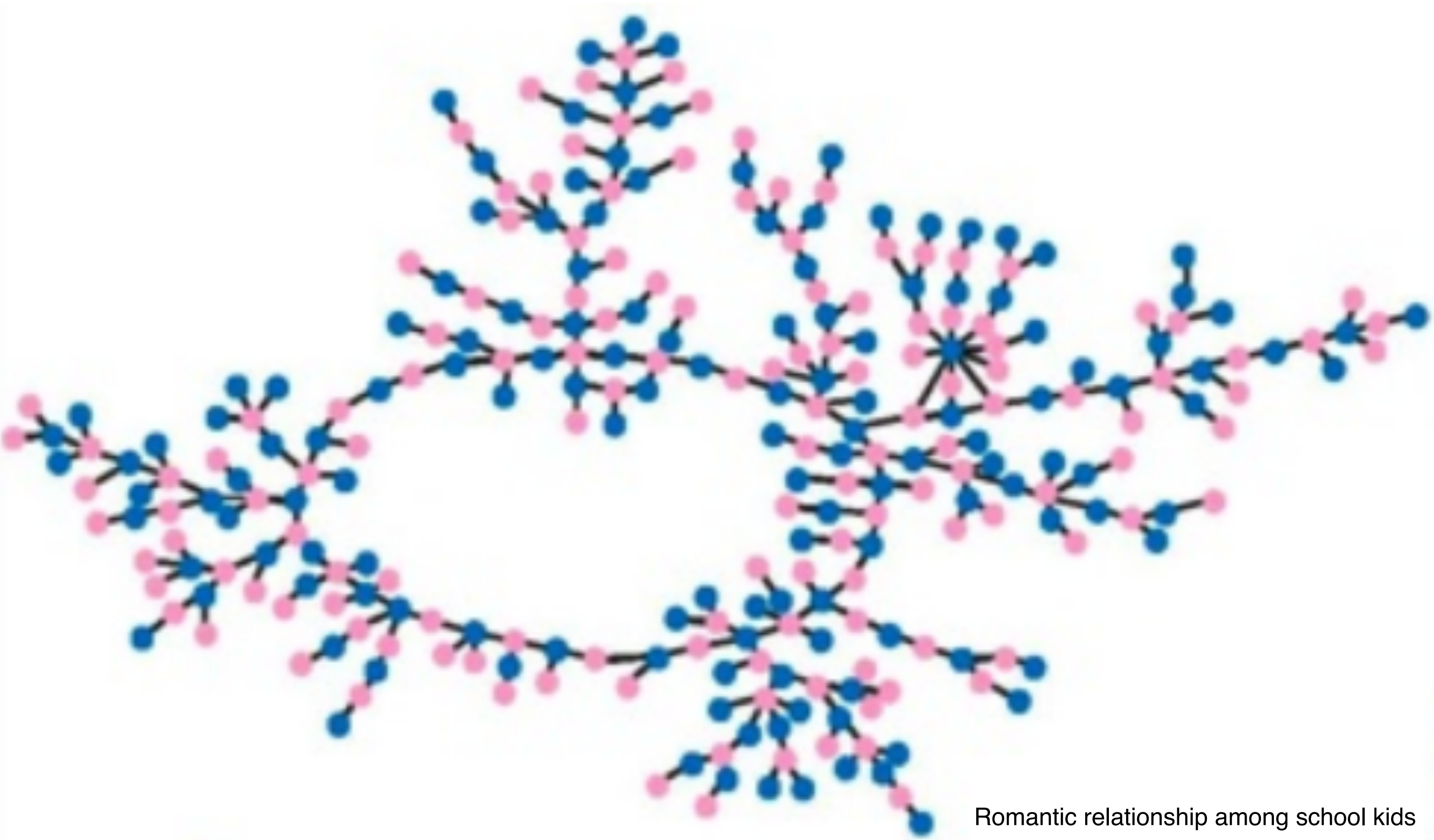
Accidental bipartite-ism



Bipartite does not imply 2-mode

Steve Borgatti (2009) The analysis of 2-mode networks. Keynote Conference and Workshop on Two-Mode Social Network Analysis

Bipartite-ism by choice



Romantic relationship among school kids

two modes are not necessarily needed

Peter Bearman, James Moody, and Katherine Stovel. Chains of affection: The structure of adolescent romantic and sexual networks. American Journal of Sociology, 110(1):44–99, 2004.

Presenting affiliation networks

Affiliation matrix

$$a_{ij} = \begin{cases} 1 \text{ if actor } i \text{ is affiliated with event } j \\ 0 \text{ otherwise.} \end{cases}$$

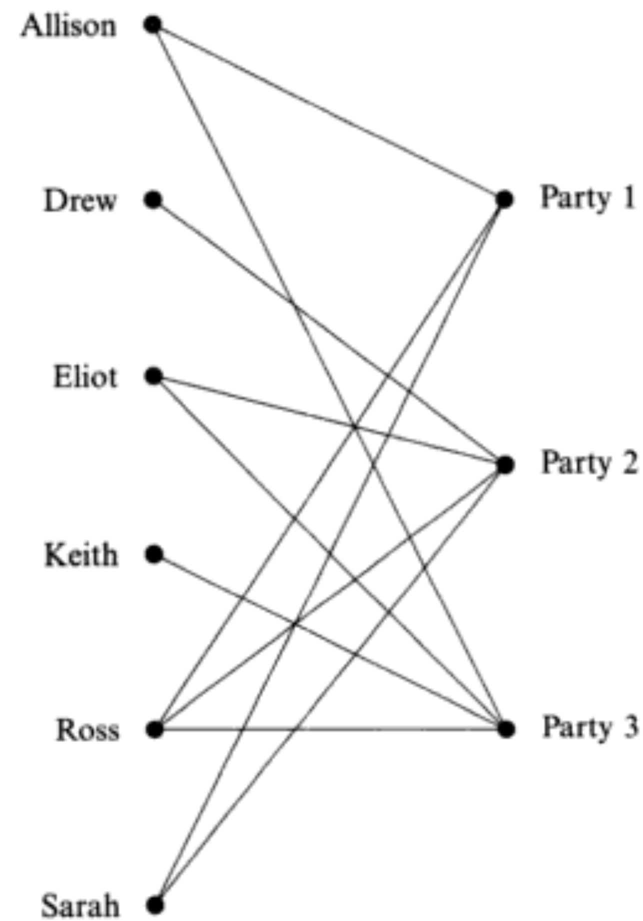
Actor	Event Party 1	Party 2	Party 3
Allison	1	0	1
Drew	0	1	0
Eliot	0	1	1
Keith	0	0	1
Ross	1	1	1
Sarah	1	1	0

- Data are presented in an **affiliation matrix X**
- The actors are the children and the events are the birthday parties they attended
- **Rates of Participation:** the number of foci (e.g. events) with which each person is affiliated
- In our example, the row marginal totals indicate the number of parties a child attended
- **Size of focus:** number of people affiliated with a focus
- Column marginal totals indicate the number of children that attended a party

Wasserman, S, Faust, K. Social Network Analysis. Cambridge University Press. 1997.

Bipartite Network

- Nodes are partitions into two subsets and all lines are between pairs of nodes belonging to different subsets
- As there are g actors and h events, there are g + h nodes
- The lines on the graph represent the relation “is affiliated with” from the perspective of the actor and the relation “has as a member” from the perspective of the event
- No two actors are adjacent and no two events are adjacent



Wasserman, S, Faust, K. Social Network Analysis. Cambridge University Press. 1997.

Advantages and Disadvantages

- Advantages
  - They highlight the connectivity in the network, as well as the indirect chains of connection
  - Data is not lost and we always know which individuals attended which events

- Disadvantage
  - They can be unwieldy when used to depict larger affiliation networks

Wasserman, S, Faust, K. Social Network Analysis. Cambridge University Press. 1997.

Bi-Adjacency matrix

g = 6 children
h = 3 parties
g+h = 9 rows
g+h = 9 cols

	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	-

Wasserman, S, Faust, K. Social Network Analysis. Cambridge University Press. 1997.

Advantages and Disadvantages

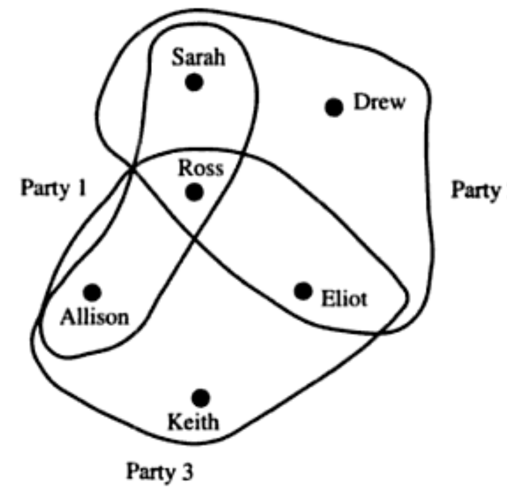
- Advantage
  - It allows the network to be examined from the perspective of an individual actor or an individual event because the actor's affiliations and the event's members are directly listed.

- Disadvantage
  - It can be unwieldy when used to depict large affiliation networks.

Adapted slides by Sai Moturu "Affiliation networks"

Hypergraph

- Affiliation networks can also be described as collections of subsets of entities
- Both actors and events can be viewed as subsets of entities
- Hypergraphs consist of a set of objects, called points and a collection of subsets of objects, called edges



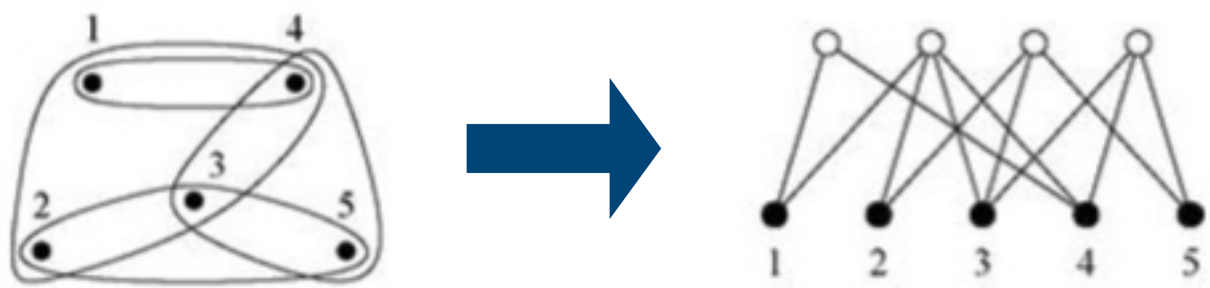
a. Hypergraph: $\mathscr{H}(\mathscr{N},\mathscr{M})$

b. Dual Hypergraph: $\mathscr{H}^*(\mathscr{M},\mathscr{N})$

Adapted slides by Sai Moturu "Affiliation networks"

Advantages and Disadvantages

- Advantage
  - Allows the network to be examined from the perspective of an individual actor or an individual event because the actor's affiliations and the event's members are directly listed.
- Disadvantage
  - It can be unwieldy when used to depict large affiliation networks.
- Hypergraphs have been used to describe urban structures and participation in voluntary organizations.

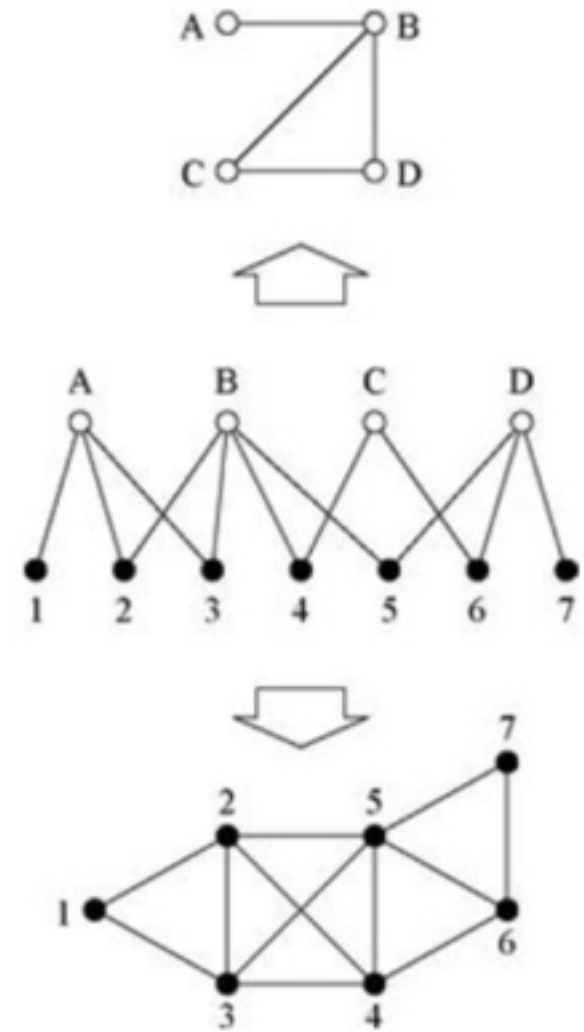


Adapted slides by Sai Moturu "Affiliation networks"

Projections

Two one-mode projections

- Example a network of films and actors
- Case 1: Projection onto the actors alone by constructing the n-vertex network in which the vertices represent actors and two actors are connected by an edge if they have appeared together in a film.
- Case 2: Projection onto the films would be the g-vertex network where the vertices represent films and two films are connected if they share a common actor.



Another example

Zhou, T., Ren, J., Medo, M. c. v., & Zhang, Y.-C. 2007, 'Bipartite network projection and personal recommendation', Phys. Rev. E, 76, 046115.

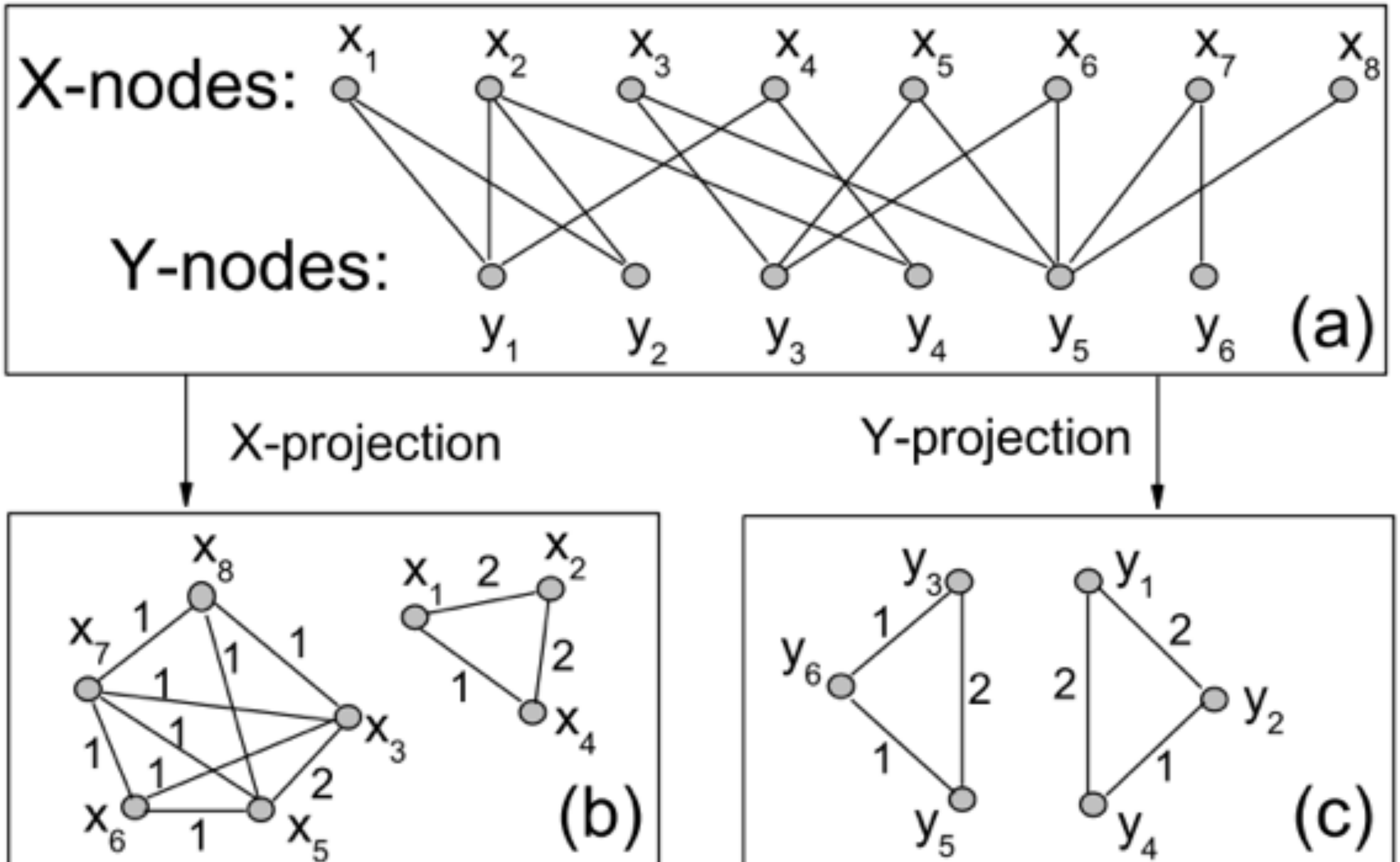


Illustration of a bipartite network (a), as well as its X-projection (b) and Y -projection (c). The edge-weight in (b) and (c) is set as the number of common neighbors in Y and X, respectively.

Analyzing affiliation networks

Deep South - Example of 2-mode network analysis

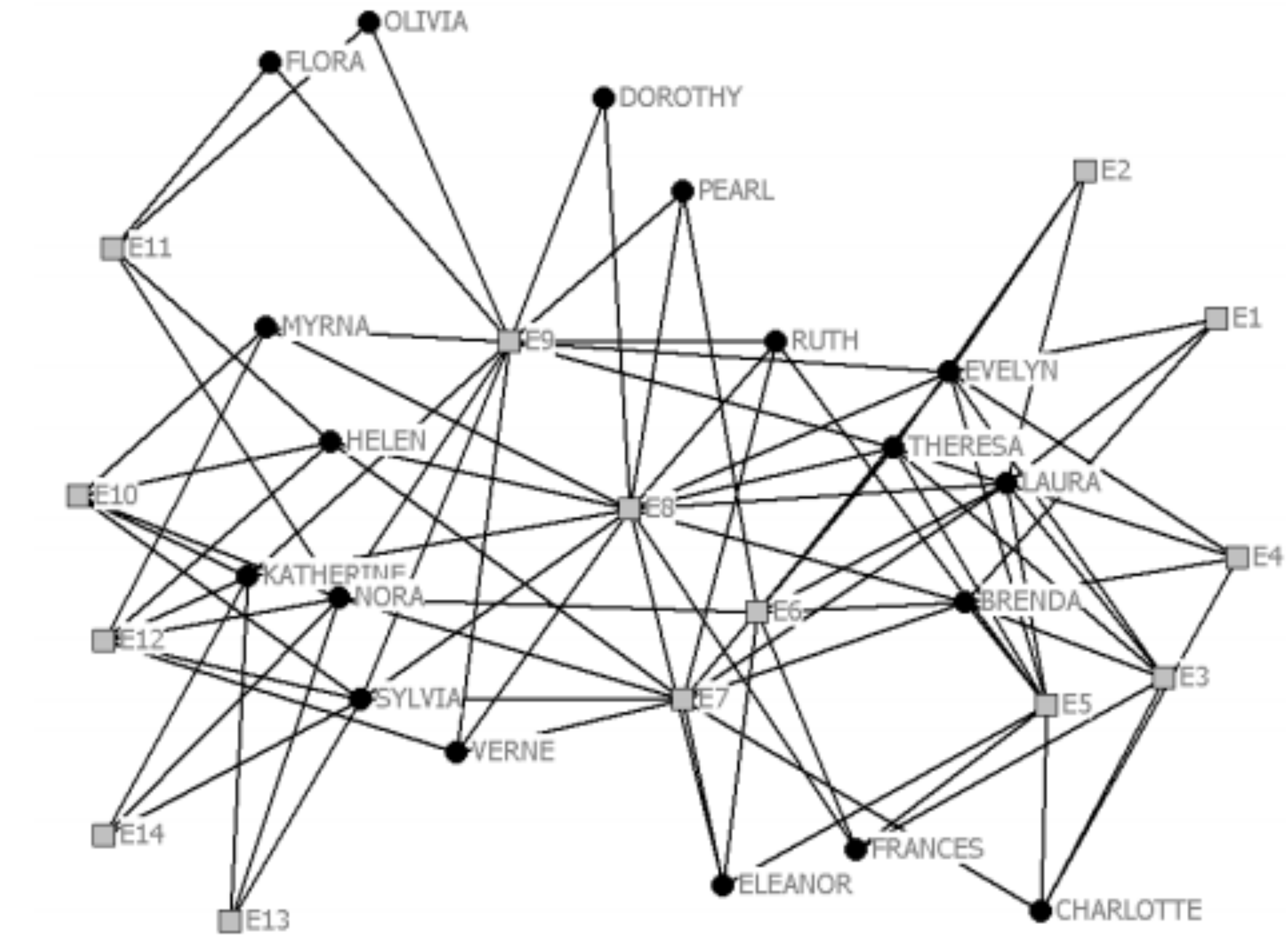
- Goal
  - Examine relation between social class and informal interaction
- Data Collection
  - Spent 9 months in a small town in Mississippi
  - Observing 18 women during 14 informal social events
  - Recording of participation using interviews, observations, guest lists, and newspaper
- Research Questions
  - Is the network connected through social events?
  - Do distinct social groups exist?
  - Which persons are more highly connected than others?

Davis, A., B. Gardner and M. Gardner 1941 *Deep South: A Social Anthropological Study of Caste and Class.* Chicago: University of Chicago Press.

Women-by-events matrix

Names of Participants of Group I	Code Numbers and Dates of Social Events Reported in *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	×	×	×	×	×	×		×	×					
2. Miss Laura Mandeville	×	×	×		×	×	×	×						
3. Miss Theresa Anderson		×	×	×	×	×	×	×	×					
4. Miss Brenda Rogers	×		×	×	×	×	×	×						
5. Miss Charlotte McDowd			×	×	×		×							
6. Miss Frances Anderson			×		×	×		×						
7. Miss Eleanor Nye					×	×	×	×						
8. Miss Pearl Oglethorpe						×		×	×					
9. Miss Ruth DeSand					×		×	×	×					
10. Miss Verne Sanderson							×	×	×			×		
11. Miss Myra Liddell								×	×	×		×		
12. Miss Katherine Rogers								×	×	×		×	×	×
13. Mrs. Sylvia Avondale							×	×	×	×		×	×	×
14. Mrs. Nora Fayette						×	×		×	×	×	×	×	×
15. Mrs. Helen Lloyd							×	×		×	×	×		
16. Mrs. Dorothy Murchison								×	×					
17. Mrs. Olivia Carleton									×		×			
18. Mrs. Flora Price									×		×			

Women-by-events network



Affiliation Matrix

- An affiliation network can be defined by the affiliation matrix X. If there are j number of actors and i number of groups:

Names of Participants of Group I	Code Numbers and Dates of Social Events Reported in *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	×	×	×	×	×	×		×	×					
2. Miss Laura Mandeville	×	×	×		×	×	×	×						
3. Miss Theresa Anderson		×	×	×	×	×	×	×	×					
4. Miss Brenda Rogers	×		×	×	×	×	×	×						
5. Miss Charlotte McDowd			×	×	×		×							
6. Miss Frances Anderson			×		×	×		×						
7. Miss Eleanor Nye					×	×	×	×						
8. Miss Pearl Oglethorpe						×		×	×					
9. Miss Ruth DeSand					×		×	×	×					
10. Miss Verne Sanderson							×	×	×			×		
11. Miss Myra Liddell								×	×	×		×		
12. Miss Katherine Rogers								×	×	×		×	×	×
13. Mrs. Sylvia Avondale							×	×	×	×		×	×	×
14. Mrs. Nora Fayette						×	×		×	×	×	×	×	×
15. Mrs. Helen Lloyd							×	×		×	×	×		
16. Mrs. Dorothy Murchison								×	×					
17. Mrs. Olivia Carleton									×		×			
18. Mrs. Flora Price									×		×			

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0

Bi-Adjacency matrix

	EV	LA	TH	BR	CH	FR	EL	PE	RU	VE	MY	KA	SY	NO	HE	DO	OL	FL	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN																			1	1	1	1	1	1		1	1					
LAURA																			1	1	1		1	1	1	1						
THERESA																				1	1	1	1	1	1	1	1					
BRENDA																			1		1	1	1	1	1	1						
CHARLOTTE																					1	1	1		1							
FRANCES																					1		1	1		1						
ELEANOR																							1	1	1	1						
PEARL																								1		1	1					
RUTH																							1		1	1	1					
VERNE																									1	1	1			1		
MYRNA																										1	1	1		1		
KATHERINE																										1	1	1		1	1	1
SYLVIA																									1	1	1	1		1	1	1
NORA																								1	1		1	1	1	1	1	1
HELEN																									1	1		1	1	1		
DOROTHY																										1	1					
OLIVIA																											1		1			
FLORA																											1		1			
E1	1	1		1																												
E2	1	1	1																													
E3	1	1	1	1	1	1																										
E4	1		1	1	1																											
E5	1	1	1	1	1	1	1		1																							
E6	1	1	1	1		1	1	1						1																		
E7		1	1	1	1		1		1	1			1	1	1																	
E8	1	1	1	1		1	1	1	1	1	1	1	1		1	1																
E9	1		1					1	1	1	1	1	1	1		1	1	1														
E10											1	1	1	1	1																	
E11														1	1		1	1														
E12										1	1	1	1	1	1																	
E13												1	1	1																		
E14												1	1	1																		

Analyzing affiliation networks

Co-Affiliation

Motivation

- In some cases, the purpose of collecting affiliations data is not to understand the pattern of ties between the two sets, but to understand the pattern of ties within one of the sets.

- Given affiliations data, we can construct some kind of tie among members of a node set simply by defining co-affiliation (e.g., attendance at the same events, membership on the same corporate board) as a tie.
  -> *projection*

- For example, we can construct a woman-by-woman matrix S in which sij gives the number of events that woman i and woman j attended together

Women-by-women matrix of overlaps across events

Evelyn attended 8 events

	EVE	LAU	THE	BRE	CHA	FRA	ELE	PEA	RUT	VER	MYR	KAT	SYL	NOR	HEL	DOR	OLI	FLO
EVELYN	8	6	7	6	3	4	3	3	3	2	2	2	2	2	1	2	1	1
LAURA	6	7	6	6	3	4	4	2	3	2	1	1	2	2	2	1	0	0
THERESA	7	6	8	6	4	4	4	3	4	3	2	2	3	3	2	2	1	1
BRENDA	6	6	6	7	4	4	4	2	3	2	1	1	2	2	2	1	0	0
CHARLOTTE	3	3	4	4	4	2	2	0	2	1	0	0	1	1	1	0	0	0
FRANCES	4	4	4	4	2	4	3	2	2	1	1	1	1	1	1	1	0	0
ELEANOR	3	4	4	4	2	3	4	2	3	2	1	1	2	2	2	1	0	0
PEARL	3	2	3	2	0	2	2	3	2	2	2	2	2	2	1	2	1	1
RUTH	3	3	4	3	2	2	3	2	4	3	2	2	3	2	2	2	1	1
VERNE	2	2	3	2	1	1	2	2	3	4	3	3	4	3	3	2	1	1
MYRNA	2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	2	1	1
KATHERINE	2	1	2	1	0	1	1	2	2	3	4	6	6	5	3	2	1	1
SYLVIA	2	2	3	2	1	1	2	2	3	4	4	6	7	6	4	2	1	1
NORA	2	2	3	2	1	1	2	2	2	3	3	5	6	8	4	1	2	2
HELEN	1	2	2	2	1	1	2	1	2	3	3	3	4	4	5	1	1	1
DOROTHY	2	1	2	1	0	1	1	2	2	2	2	2	2	1	1	2	1	1
OLIVIA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2
FLORA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2

Dorothy and Katherine attended 2 events together.

Affiliation data can be a proxy for social relation

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0													
VERNE	0													
MYRNA	0													
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
Size of the events	3	3	6	4	8	8	10	14	12	5	4	6	3	3

How can you consider the size of an event?

Co-affiliations as similarity data

- Data:
  - Woman-by-woman affiliation network

- Procedure:
  - For each pair of women, look at their respective rows in X and count the number of times that they have 1s in the same places
  - In other words, for any pair of women we construct a simple 2-by-2 contingency table that shows the relationship between their pair of rows (as an unnormalized measure of similarity of rows)

a
gives the number of times
that the pair of women co-attended an event

		Woman *j* 1	0	
Woman *i*	1	a	b	a+b
	0	c	d	c+d
		a+c	b+d	n

a+b
gives the total number of events
that woman i attended

n
number of events

a+c
gives the total number of events
that woman j attended

Normalizing co-affiliations

- **How can we bound "a" between 0 and 1 in order to promote comparability across datasets?**
- Simple approach: Divide a by n

$$a^* = \frac{a}{n}$$

- For example, if woman i and woman j attend three events in common, and woman k and woman l do as well, we would likely regard the two pairs as equally close.
- **Problem:** But if i and j each only attended 3 events, whereas k and l each attended 14 events, intuition we would be more likely to conclude that the 100% overlap between i and j signals greater closeness than the 21% overlap between k and l.
- Minimizing approach

$$a_{ij}^* = \frac{a}{Min(a+b, a+c)}$$

The resulting coefficient runs between 0 and 1, where 1 indicates the maximum possible overlap given the number of events attended by i and j.

Normalizing co-affiliations (*cont.*)

- Normalizing “a” by the Jaccard coefficient
- Describes the number of events attended in common as a proportion of events that are “attendable”, as determined by the fact that at least one of the two women attended the event

$$a_{ij}^{*} = \frac{a}{a+b+c}$$

- Additional approaches are Pearson correlation and the Bonacich approach
- All of these normalizations essentially shift the nature of co-affiliation data from frequencies of cooccurrences to tendencies or revealed preferences to co-occur.

When and why is normalization useful?

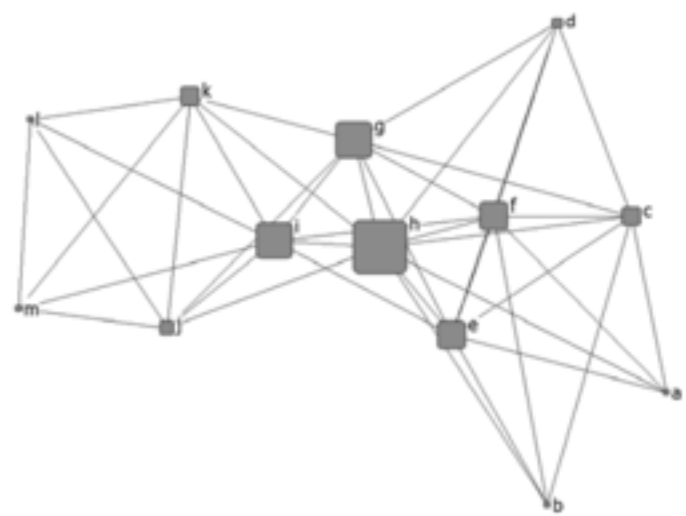
- The normalized measures are the most appropriate, if the reason for studying affiliations is that co-affiliations reveal otherwise unseen relationships between people
- Normalized measures give us the tendency or preference for a pair of women to co-occur while controlling for nuisance variables such as the number of times a woman was observed
- The normalized measures tell us how often two women are co-attending relative to the number of times they could have

Differences between describing co-affiliations

- Data:
  - Group of 13 individuals with their memberships in different social clubs (16 of them)
- Goal:
  - Understanding relationships among the 13 individuals
- Approach:
  - Affiliations data (person-by-social club) are converted into co-affiliations (person-by-person)

Unnormalized co-affiliation matrix

(Co-membership in 2 or more social clubs. Nodes size is based on number of social clubs that each individual is a member of.)



Normalized co-affiliation matrix

Spring Embedding of Jaccard Coefficients. An edge is shown if cij > 0.38.
Nodes size is based on number of social clubs that each individual is a member of.

Normalizing the size of events

- Greater co-affiliation creates more opportunities for social ties to develop, therefore we want to take into account the relative sizes of different events.
- For example, attending an event with 5 people vs. attending an an event where thousands are present
- An obvious approach, then, is to weight events inversely by their size.

- Realisation of the normalized co-affiliations:
  - The quantity n becomes the sum of weights of all events, and the quantity a is the sum of weights of the events that were co-attended by i and j.

Appropriate normalizations by view of data

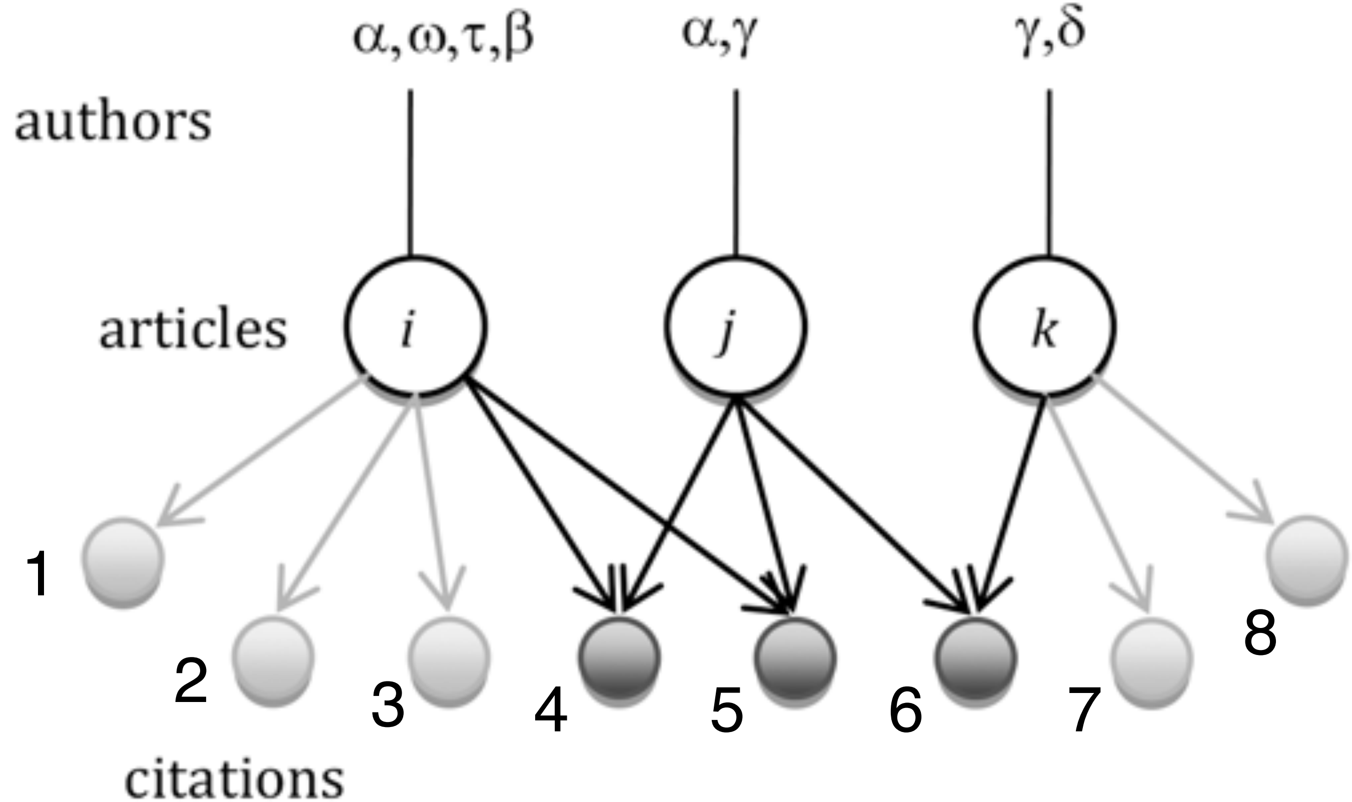
- For convenience, it is assumed that the 2-mode affiliations data are actor-by-event, and that we are interested in constructing the actor-by-actor co-affiliation matrix.
- We refer to the actors/rows as “variables” and the events/columns as “cases”.

Co-Affiliation as Opportunity	Co-Affiliation as Indicator
• No normalization (simple overlap counts) • Case normalization (e.g., weighting inversely by event sizes)	• Variable normalization (e.g., Jaccard or Pearson correlations)

Questions?

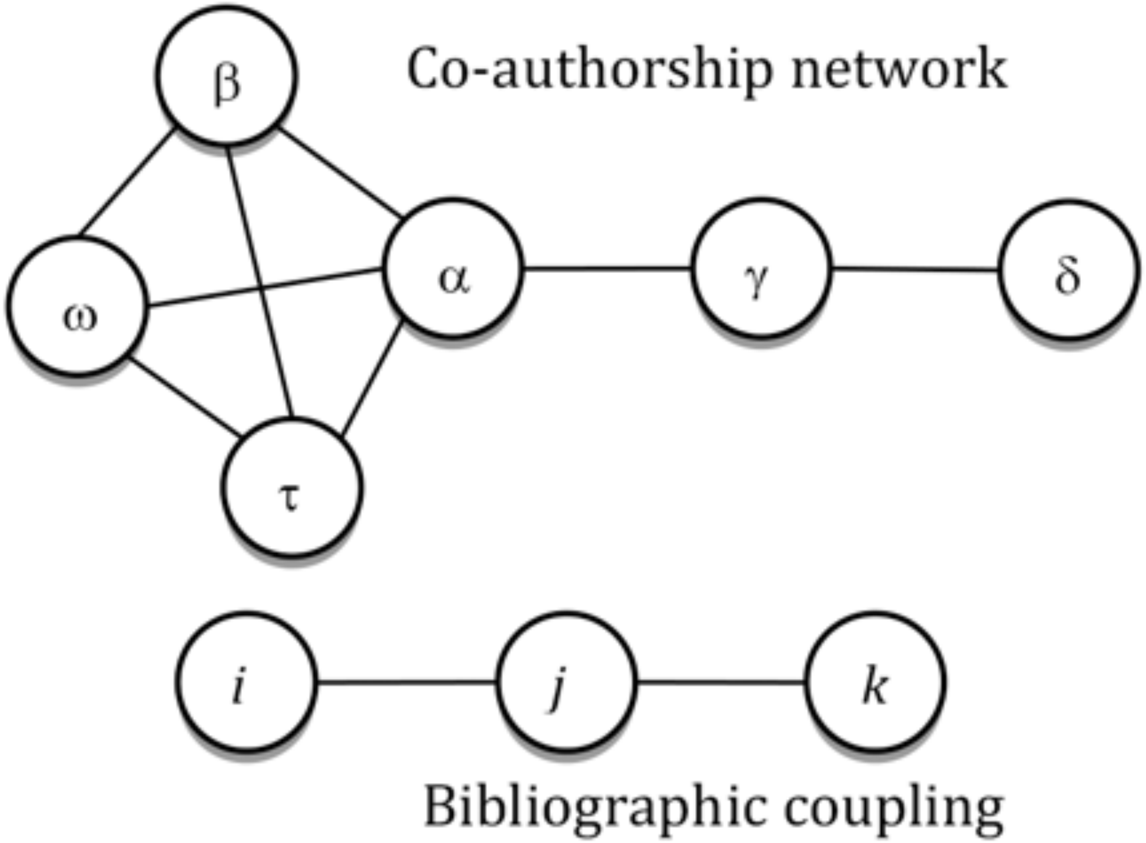
The structure of a scientific literature

Biscaro, C. & Giupponi, C. 2014, 'Co-authorship and bibliographic coupling network effects on citations', PloS one, 9, 6, e99502.



Different views on the same data

Biscaro, C. & Giupponi, C. 2014, 'Co-authorship and bibliographic coupling network effects on citations', PloS one, 9, 6, e99502.



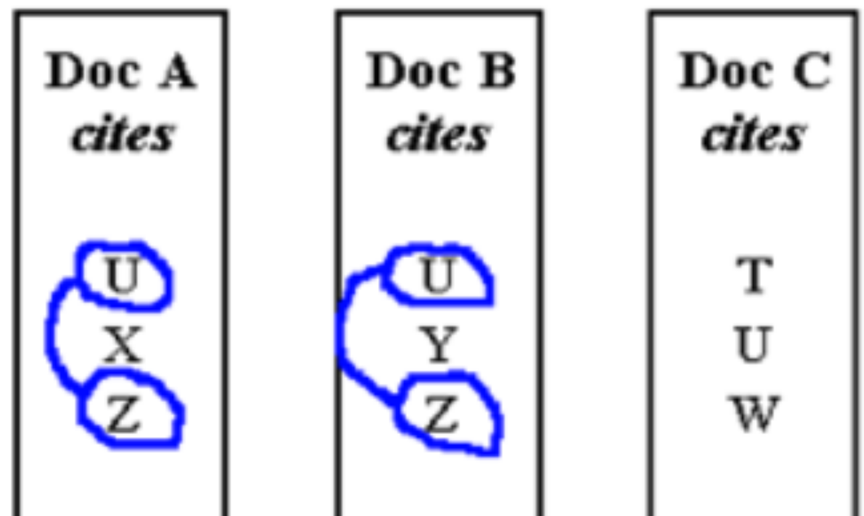
Another perspective includes only the references of these articles.

Cocitation - simple example

- Work A cites Works U, X and Z
- Work B cites Works U, Y and Z
- Work C cites Works T, U and W

- Co-citation link strength (UZ) = 2
- Co-citation link strength (UX) = 1
- Co-citation link strength (UT) = 1
- Co-citation link strength (XZ) = 1 (etc.)

- **Thus, works U and Z are most likely to be about the same subject**



Cocitation (*cont.*)

- Cocitation of two vertices i and j in a directed network is the number of vertices that have outgoing edges pointing to both i and j

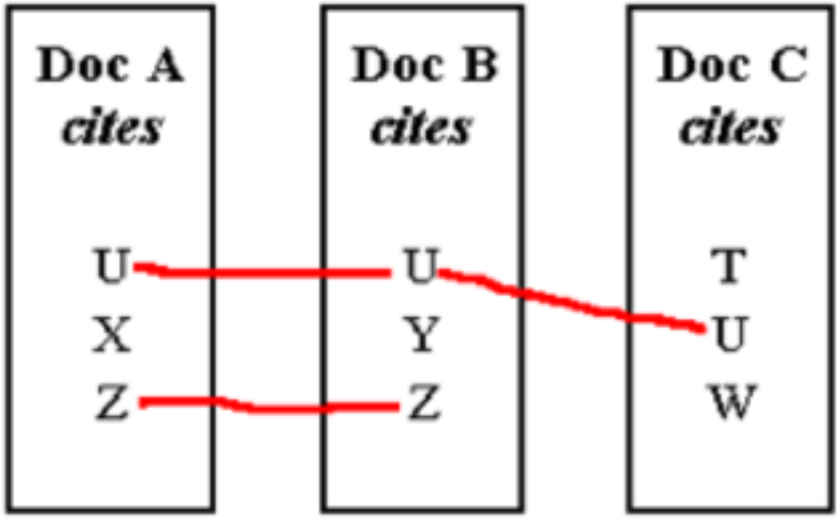
- Citation networks: the cocitation of two papers is the number of other papers that cite both

Bibliographic coupling - simple example

- Work A cites Works U, X and Z
- Work B cites Works U, Y and Z
- Work C cites Works T, U and W

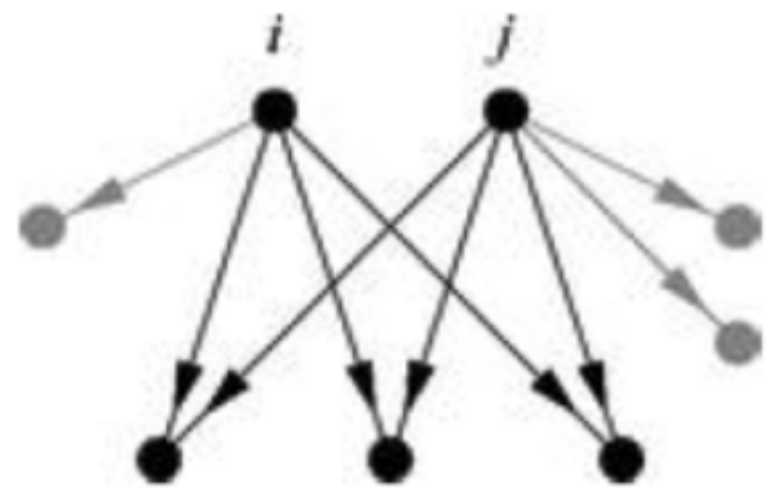
- Bibliographic coupling strength (AB) = 2
- Bibliographic coupling strength (AC) = 1
- Bibliographic coupling strength (BC) = 1

- Thus, Works A and B are most likely to be about the same subject



Bibliographic coupling (cont.)

- The bibliographic coupling of two vertices in a directed network is the number of other vertices to which both point
- Citation network: bibliographic coupling of two papers i and j is the number of other papers that are cited by both i and j



Difference between Bibliographic Coupling and Co-citation

- Cocitation
  - Co-citation coupling is a method used to establish a subject similarity between two documents
  - Two documents are said to be co-cited when they both appear in the reference list of a third document
  - Co-citation focuses on references which frequently come in pair
  - Co-citation is called “**prospective coupling**”

- Bibliographic coupling
  - The term “bibliographic coupling” was coined by M. M. Kessler, he defined a unity of coupling between two papers as an item of reference used by these two papers
  - The two papers are then said to be bibliographic coupled
  - Bibliographic coupling focuses on groups of papers which cite a source document
  - Bibliographic is said to be “**retrospective coupling**”