Graph Theory and Complex Networks: An Introduction

Maarten van Steen

VU Amsterdam, Dept. Computer Science
Room R4.20, steen@cs.vu.nl

Chapter 09: Social networks

Version: May 12, 2014

vrije Universiteit amsterdam

Contents

Chapter	Description
01: Introduction	History, background
02: Foundations	Basic terminology and properties of graphs
03: Extensions	Directed & weighted graphs, colorings
04: Network traversal	Walking through graphs (cf. traveling)
05: Trees	Graphs without cycles; routing algorithms
06: Network analysis	Basic metrics for analyzing large graphs
07: Random networks	Introduction modeling real-world networks
08: Computer networks	The Internet & WWW seen as a huge graph
09: Social networks	Communities seen as graphs

Introduction

Observation

Sociologists have always been interested in social structures:

- formation of groups
- influence relationships
- ties of families and friends
- (dis)likings in groups of people

Observation

Graphs form a natural way for modeling social structures

- Sociograms and blockmodeling
- Basic concepts: balance, cohesiveness, affiliation networks
- Equivalence

Example: Workers on strike

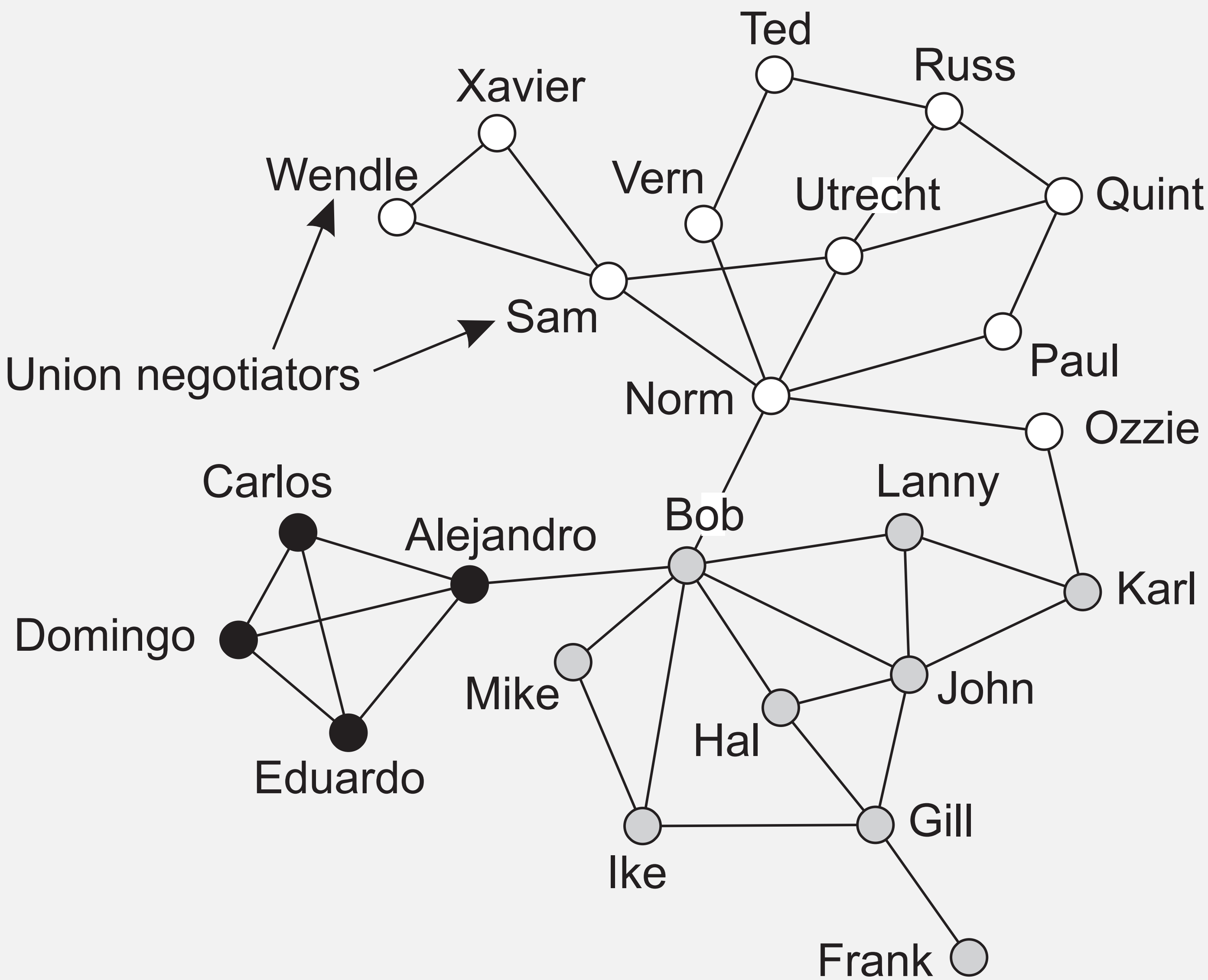
Case

In a small wood-processing firm, management proposed a new compensation package. This led to a strike; management suspected miscommunication. The workers were asked to indicate how often and with whom they discussed the strike.

Model

Graph in which two people were linked if they frequently talked to each other.

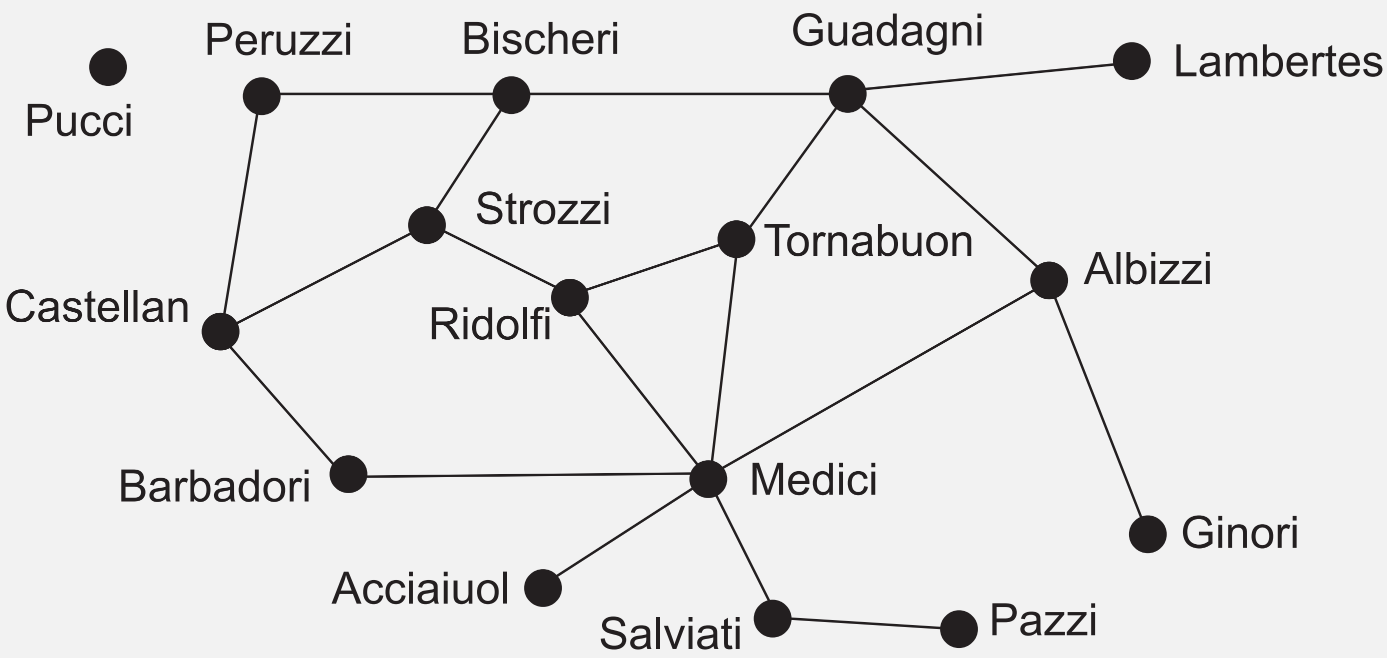
Example: Workers on strike



Example: The influence of the Medici's

Situation

Giovani di Bicci created the Medici Bank and became very rich. His son, Cosimo de' Medici, is the actual founder of the Medici dynasty. Cosimo made sure that the right people got married to each other, resulting in more power.



Example: The influence of the Medici's

Observation

The Strozzi family was richer and had more representatives in the local legislature. Yet the Medici's power surpassed that of the Strozzi's.

Reconsider the **betweenness centrality**:

$$c_B(u) = \sum_{x \neq y \neq u} \frac{|S(x,u,y)|}{|S(x,y)|}$$

with

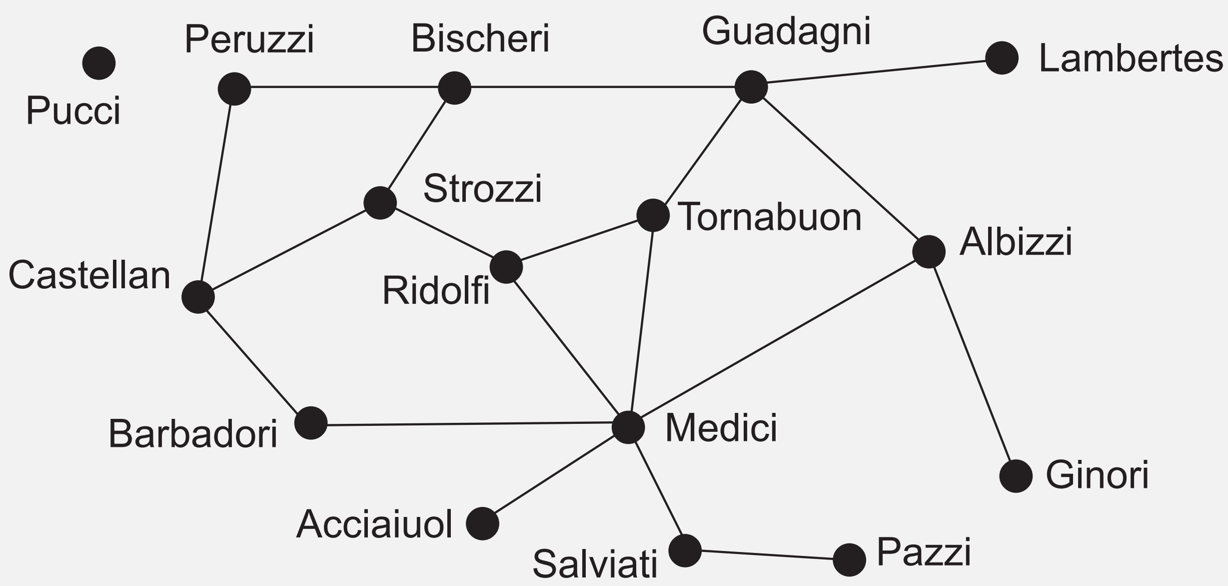
- $S(x,u,y)$ is collection of shortest (x,y) paths containing u
- $S(x,y)$ is set of shortest paths between vertices x and y.

Example: The influence of the Medici's

Normalization

Normalize $c_B(u)$ by the maximum possible pairs of families that u can connect: $\binom{n-1}{2}$

$c_B(\text{Medici}) = 0.522$ whereas $c_B(\text{Strozzi}) = 0.103$



Starters: sociograms

History

Already early in the 1930s Jacob Moreno introduced graph-like representations for social structures and suggested that they could be used for discovering new features.

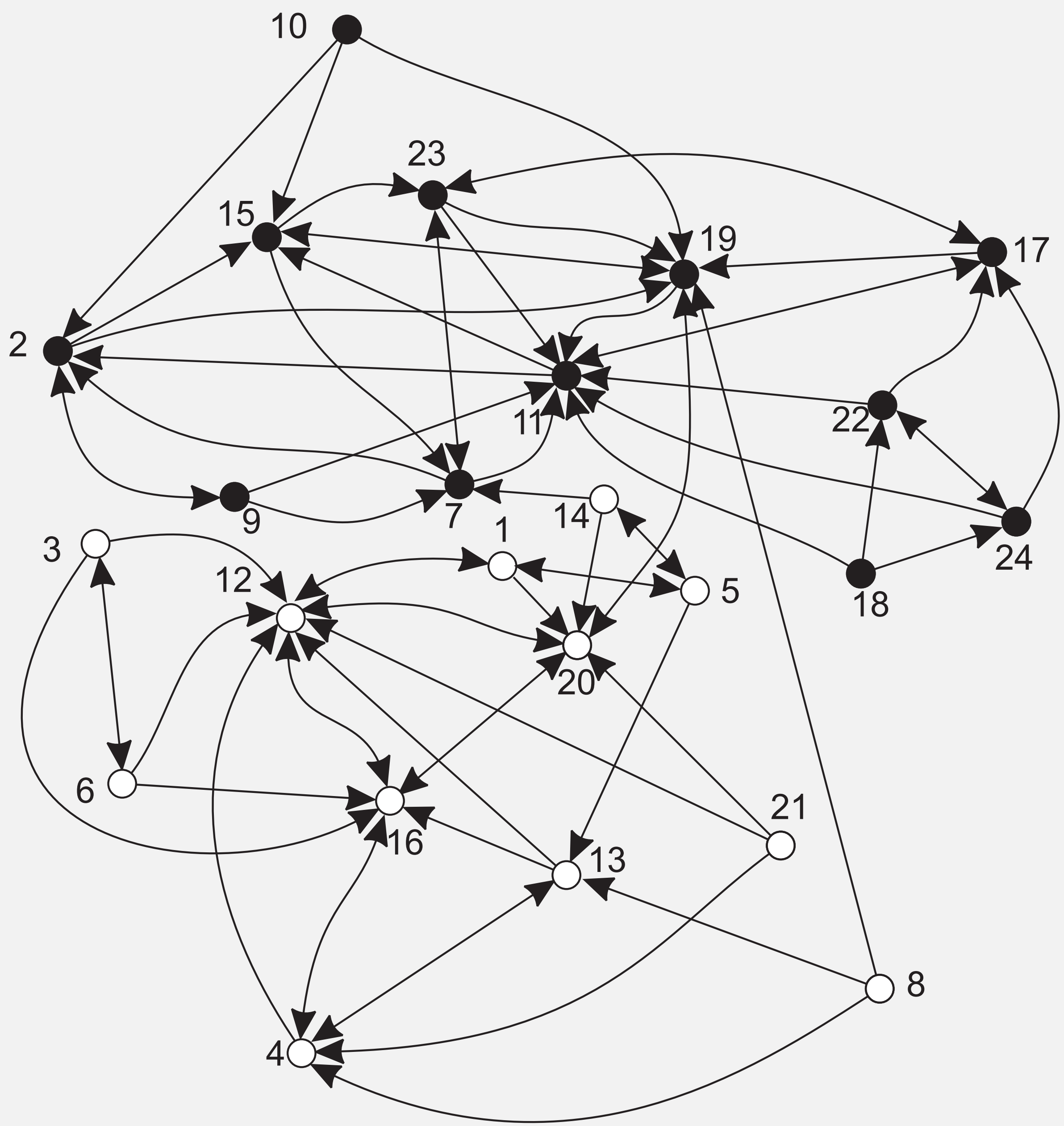
Sociograms in the classroom

In order to get an impression of how a class operates, teachers can ask their pupils to list the three classmates they (dis)like the most.

Example classroom sociogram

Sex	**ID**	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
F	1	■				+				−	−		+								+	−			
M	2	−	■							+						+				+			−	−	
F	3			■			+	−			−		+				+					−			
F	4				■						−		+	+			+			−	−				
F	5	+				■					−			+	+					−	−				
F	6	−		+			■						+			−	+						−		
M	7		+					■			−	+								−		−		+	
F	8				+		−		■					+				−		+					−
M	9		+					+		■		+		−		−									−
M	10		+					−			■	−				+				+		−			
M	11		+								−	■				+		+		−	−				
F	12	+						−			−		■			−	+				+				
F	13				+								+	■			+			−	−	−			
F	14					+	−	+		−					■						+		−		
M	15							+			−				−	■				+				+	−
F	16				+						−		+				■				+		−	−	
M	17				−							+						■		+	−	−		+	
M	18							−				+							■	−		−	+		+
M	19		−									+	−			+				■	+	−			
F	20						−			−			+		−		+			+	■				
F	21	−	−		+								+							−	+	■			
M	22						−			−		+					−	+					■		+
M	23	−						+				+						−		+		−		■	
M	24											+						+			−	−	+	−	■
	+	2	4	1	4	2	1	4	0	1	0	8	8	3	1	4	6	3	0	7	6	0	2	3	2
	−	4	2	0	1	0	4	4	0	4	9	1	1	1	2	3	1	2	0	7	6	10	4	3	3

Classroom example - positive nominations



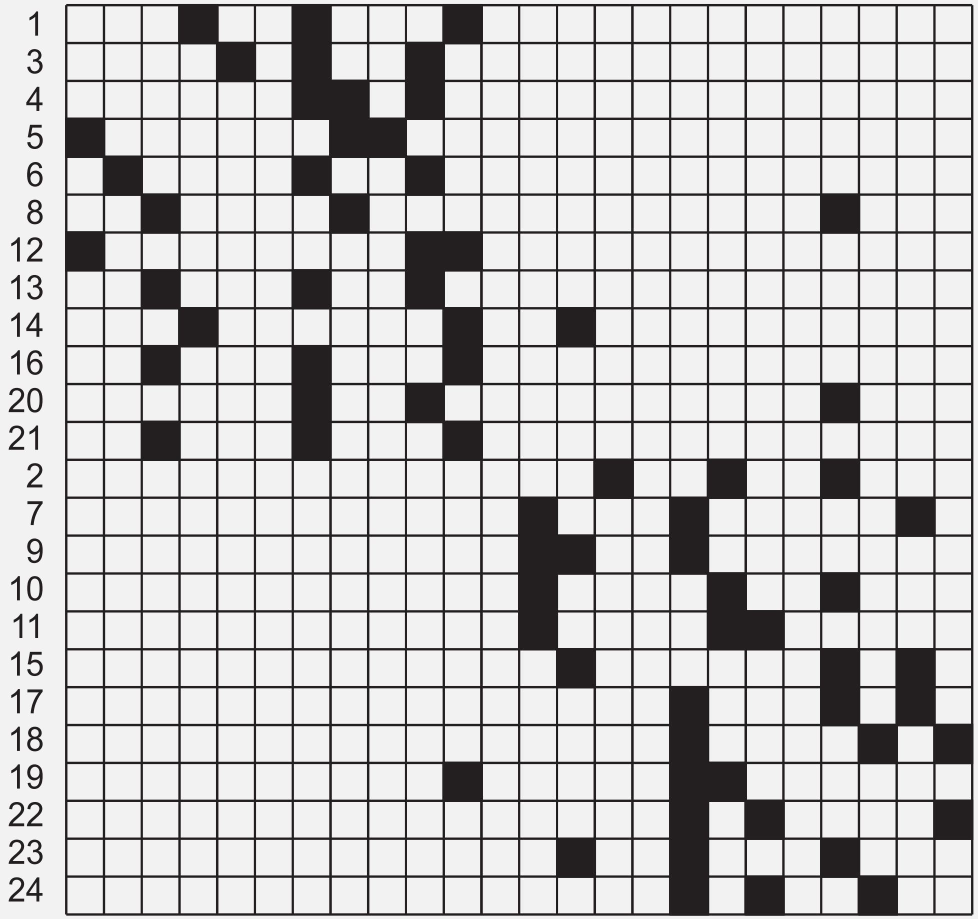
- Clear distinction between boys ("•") and girls ("○")
- Relation between 19 and 20 is important
- There are a few "isolated" children (8 & 10)

Issue

Can we discover these properties **mathematically**?

Blockmodeling

Essence: reorder the rows and columns in the adjacency matrix in order to discover **subgroups**. Can be done automatically (and is then called **clustering**).



Concentrate on SCC (largest Strongly Connected Component)

Eccentricity

Recall: Eccentricity u is maximal minimal distance to other vertices

Child:	1	2	4	5	7	9	11	12
Ecc.:	5	6	6	4	7	7	7	5

Child:	13	14	15	16	17	19	20	23
Ecc.:	6	3	6	5	6	5	4	6

Observations

Child #14 is one of the few nominating a boy *and* a girl. She also seems to be “in the middle.”

Concentrate on SCC

Closeness

Recall: $c_C(u) = \frac{1}{\sum_{v \in V(G)} d(u,v)}$

Child:	1	2	4	5	7	9	11	12
Close:	.23	.21	.18	.25	.18	.18	.18	.22

Child:	13	14	15	16	17	19	20	23
Close:	.18	.30	.21	.21	.21	.25	.25	.21

Observation

The closeness confirms that child #14 is close to **everyone**.

Concentrate on SCC

Betweenness

Child:	1	2	4	5	7	9	11	12
Betw.:	.140	.153	.050	.105	.083	.007	.155	.220

Child:	13	14	15	16	17	19	20	23
Betw.:	.016	.054	.083	.140	.017	.466	.469	.029

Observation

The picture has changed dramatically: child #14 may be close, but her importance should be questioned.

Metrics already discussed

Definition (Vertex centrality)

G is (strongly) connected. The vertex centrality:
$c_E(u) = 1/\max\{d(u,v) | v \in V(G)\}$

Definition (Closeness)

G is (strongly) connected. The closeness: $c_C(u) = 1/\sum_{v \in V(G)} d(u,v)$

Definition (Betweenness)

G is simple and (strongly) connected. $S(x,y)$ is set of shortest paths between x and y. $S(x,u,y) \subseteq S(x,y)$ paths that pass through u.
Betweenness centrality: $c_B(u) = \sum_{x \neq y \neq u} \frac{|S(x,u,y)|}{|S(x,y)|}$.

Prestige

Definition (Degree prestige)

Let D be a directed graph. The degree prestige $p_{deg}(v)$ of a vertex $v \in V(D)$ is defined as its indegree $\delta^{-}(v)$.

Definition (Proximity prestige)

Let D be a directed graph with n vertices. The influence domain $R^{-}(v)$ is the set of vertices from where v can be reached through a directed path, that is, $R^{-}(v) = \{u \in V(D) | \exists (u,v)\text{-path}\}$. The proximity prestige:

$$p_{prox}(v) = \frac{|R^{-}(v)|/(n-1)}{\sum_{u \in R^{-}(v)} d(u,v)/|R^{-}(v)|}$$

Ranked prestige

Definition

Consider a simple directed graph D with vertex set $\{1, 2, \ldots, n\}$ with adjacency matrix $\mathbf{A}$ The ranked prestige of a vertex k is:

$$p_{rank}(k) = \sum_{i=1, i \neq k}^{n} \mathbf{A}[i,k] \cdot p_{rank}(i)$$

Simple example

ID	A	B	C
A	—	0.5	0.4
B	0.1	—	0.6
C	0.9	0.5	—

$$\begin{aligned} p_{rank}(A) &= 0.5 \cdot p_{rank}(B) + 0.4 \cdot p_{rank}(C) \\ p_{rank}(B) &= 0.1 \cdot p_{rank}(A) + 0.6 \cdot p_{rank}(C) \\ p_{rank}(C) &= 0.9 \cdot p_{rank}(A) + 0.5 \cdot p_{rank}(B) \end{aligned}$$

$\mathbf{ID}[i,j]$: how much is i appreciated by j?

Computing ranked presitige

Some simple rewriting

$$
\begin{aligned}
p_{rank}(A) &= 0.5 \cdot p_{rank}(B) + 0.4 \cdot p_{rank}(C) & x &= 0.5 \cdot y + 0.4 \cdot z & (1)\\
p_{rank}(B) &= 0.1 \cdot p_{rank}(A) + 0.6 \cdot p_{rank}(C) & y &= 0.1 \cdot x + 0.6 \cdot z & (2)\\
p_{rank}(C) &= 0.9 \cdot p_{rank}(A) + 0.5 \cdot p_{rank}(B) & z &= 0.9 \cdot x + 0.5 \cdot y & (3)
\end{aligned}
$$

Some simple substitutions

1. Substitute (2) into (3)
2. Substitute (3) into (2)
3. Require that $\sqrt{x^2 + y^2 + z^2} = 1$

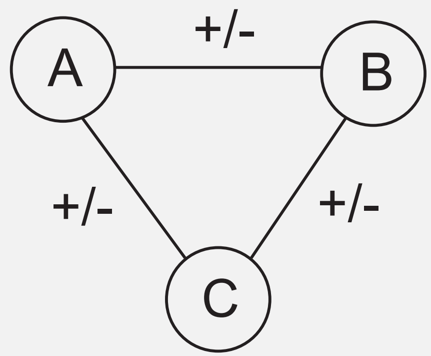
Results

$$x = 0.52 \quad y = 0.48 \quad z = 0.71$$

Structural balance

Basic idea

Consider **triads**: potential relationships between triples of social entities, and label every relationship as positive or negative. We then consider **balanced** triads.



A–B	B–C	A–C	B/I	Description
+	+	+	**B**	Everyone likes each other
+	+	−	**I**	Dislike A–C stresses relation B has with either of them
+	−	+	**I**	Dislike B–C stresses relation A has with either of them
+	−	−	**B**	A and B like each other, and both dislike C
−	+	+	**I**	Dislike A–B stresses relation C has with either of them
−	+	−	**B**	B and C like each other, and both dislike A
−	−	+	**B**	A and C like each other, and both dislike B
−	−	−	**I**	Nobody likes each other

Structural balance: signed graphs

Definition

A **signed graph** is a simple graph G in which each edge e is labeled with either a positive (“$+$”) or negative (“$-$”) sign, $sign(e)$.

Definition

The **product of two signs** s_1 and s_2 is again a sign, denoted as $s_1 \cdot s_2$. It is negative if and only if *exactly one* of s_1 and s_2 is negative. The **sign of a trail** T is the product of the signs of its edges: $sign(T) = \Pi_{e \in E(T)} sign(e)$.

Definition

An undirected signed graph is balanced when all its cycles are positive.

Balanced networks: special characterization

Theorem

An undirected signed complete graph G is balanced if and only if $V(G)$ can be partitioned into two disjoint subsets V_0 and V_1 such that each negative-signed edge is incident to a vertex from V_0 and one from V_1, and each positive-signed edge is incident to vertices from the same set.

More formally

Let $E^-(G)$ be the edges with negative sign, and $E^+(G)$ the ones with positive sign. Then, $E^-(G) = \{\langle x, y\rangle | x \in V_0, y \in V_1\}$ and $E^+(G) = \{\langle x, y\rangle | x, y \in V_0 \text{ or } x, y \in V_1\}$.

Proof: V can be properly partitioned $\Rightarrow$ G is balanced

Every cycle in G contains an even number of edges from $E^-(G)$. All other edges have positive sign. G must be balanced.

Proof

Proof: G is balanced $\Rightarrow$ V can be partitioned

- Let $u \in V(G)$ and let $N^+(u) = \{v \in N(u) | sign(\langle u, v \rangle) = “+"\}$
- Set $V_0 \leftarrow \{u\} \cup N^+(u)$ and $V_1 \leftarrow V(G) \setminus V_0$.
- Consider $v_0, w_0 \in V_0$, other than u. Note: $\langle u, v_0 \rangle$ and $\langle u, w_0 \rangle$ are positive signed $\Rightarrow$ also $\langle v_0, w_0 \rangle$ is positive signed.
- Consider $v_1, w_1 \in V_1$. The triangle with vertices u, v_1, w_1 must be positive; $\langle u, v_1 \rangle$ and $\langle u, w_1 \rangle$ are negative signed $\Rightarrow \langle v_1, w_1 \rangle$ must be positive signed.
- Consider $\langle v_0, v_1 \rangle$, $sign(\langle u, v_0 \rangle)$ is positive, $sign(\langle u, v_1 \rangle)$ negative $\Rightarrow \langle v_0, v_1 \rangle$ must be negative signed.

Balanced networks: path characterization

Theorem

Consider an undirected signed graph G and two distinct vertices $u, v \in V(G)$. G is balanced if and only if all (u, v)-paths have the same sign.

Proof: G is balanced $\Rightarrow$ all (u, v)-paths have the same sign

- Let P and Q be two distinct (u, v)-paths.
- Let $E' = \big(E(P) \cup E(Q)\big) \setminus \big(E(P) \cap E(Q)\big)$.
- $H = G[E']$ consists of edge-disjoint positive-signed cycles.
- For each cycle $C \subseteq H$: $E(C) = E(\hat{P}) \cup E(\hat{Q})$ with $\hat{P}$ a subpath of P and $\hat{Q}$ a subpath of Q.
- $sign(C) = sign(\hat{P}) \cdot sign(\hat{Q})$ is positive $\Rightarrow$ signs of $\hat{P}$ and $\hat{Q}$ must be the same.

Balanced networks: path characterization

Proof: all (u, v)-paths have the same sign $\Rightarrow$ G is balanced

Note:

- u and v have been chosen arbitrarily
- Every cycle C can be constructed as the union of two edge-disjoint paths P and Q

Consequence: for all C: $sign(C) = sign(P) \cdot sign(Q)$ must be positive $\Rightarrow$ G is balanced.

Balanced networks: general characterization

Theorem

An undirected signed graph G is balanced if and only if $V(G)$ can be partitioned into two disjoint subsets V_0 and V_1 such that
$E^-(G) = \{\langle x,y\rangle | x \in V_0, y \in V_1\}$ *and*
$E^+(G) = \{\langle x,y\rangle | x,y \in V_0 \text{ or } x,y \in V_1\}$.

Proof: V can be properly partitioned $\Rightarrow$ G is balanced

- Add $e = \langle u,v\rangle$ to G, with u,v nonadjacent
- u and v in same subset $\Rightarrow$ $sign(e)$ becomes positive, otherwise negative.
- Continue until reaching complete signed graph G^*.
- We know G^* is balanced $\Rightarrow$ G is balanced.

Balanced networks: general characterization

Proof: G is balanced $\Rightarrow$ V can be properly partitioned

- Assume G is connected. Prove by induction on number of edges m.
- Trivially OK for $m = 1$. Assume correct for $m > 1$ edges.
- Consider nonadjacent vertices u and v: all (u, v)-paths have the same sign. Add $e = \langle u, v \rangle$ with $sign(e)$ the same as a (u, v)-path.
- New cycle C will consist of e and a (u, v)-path P from G.
- $sign(C) = sign(e) \cdot sign(P)$, and $sign(e) = sign(P) \Rightarrow C$ must be positive.
- Continue until reaching complete graph G^*, and subsequently partition $V(G^*)$.

Checking for balance

Algorithm (Balanced graphs)

Consider an undirected signed graph G. $N^+(v)$ is the set of vertices adjacent to v through a positive-signed edge. $N^-(v)$ is analogous. Let I be the set of inspected vertices so far.

1. *Select an arbitrary vertex $u \in V(G)$ and set $V_0 \leftarrow \{u\}$ and $V_1 \leftarrow \emptyset$. Set $I \leftarrow \emptyset$.*
2. *Select arbitrary vertex $v \in (V_0 \cup V_1) \backslash I$. Assume $v \in V_i$.*
   - *For all $w \in N^+(v) : V_i \leftarrow V_i \cup \{w\}$.*
   - *For all $w \in N^-(v) : V_{(i+1) \mod 2} \leftarrow V_{(i+1) \mod 2} \cup \{w\}$.*
   - *Also, $I \leftarrow I \cup \{v\}$.*
3. *If $V_0 \cap V_1 \neq \emptyset$ stop: G is not balanced. Otherwise, if $I = V(G)$ stop: G is balanced. Otherwise, repeat the previous step.*

Affiliation networks

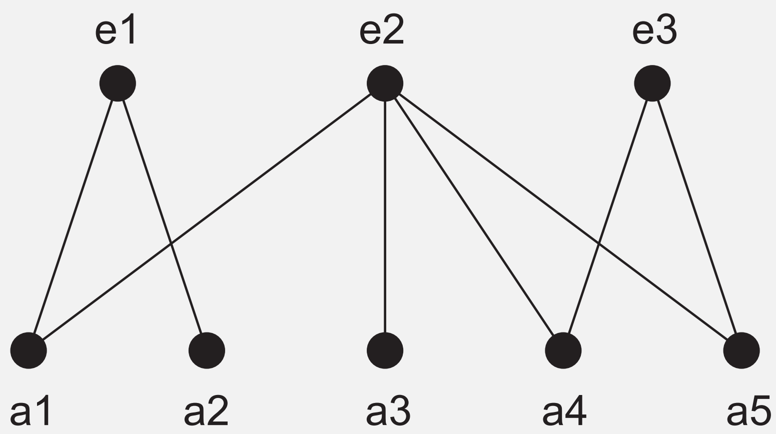
Basic idea

Social structures are assumed to consist of **actors** and **events**. Actors are tied to each other through their participation in an event. Two events are bound through the actors that participate in both events $\Rightarrow$ **two-mode networks**.

Observation

Affiliation networks are naturally represented as **bipartite graphs**: Let V_A represent the actors and V_E the events. Edge $\langle v_a, v_e \rangle$ if actor a participates in event e.

Affiliation networks & adjacency submatrix



	e1	e2	e3
a1	1	1	0
a2	1	0	0
a3	0	1	0
a4	0	1	1
a5	0	1	1

Special tables

Note

$\mathbf{AE}[i,j] = 1$ if and only if actor i participated in event j

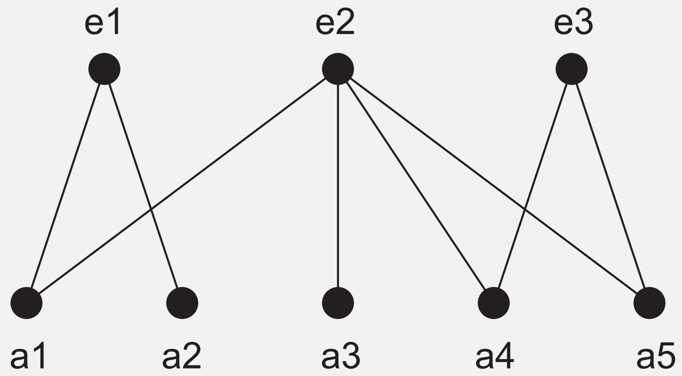
Part 1

$$\mathbf{NE}[i,j] = \sum_{k=1}^{n_E} \mathbf{AE}[i,k] \cdot \mathbf{AE}[j,k]$$

Part 2

$$\mathbf{NA}[i,j] = \sum_{k=1}^{n_A} \mathbf{AE}[k,i] \cdot \mathbf{AE}[k,j]$$

Counting joint participations



NE	a1	a2	a3	a4	a5
a1	2	1	1	1	1
a2	1	1	0	0	0
a3	1	0	1	1	1
a4	1	0	1	2	2
a5	1	0	1	2	2

NA	e1	e2	e3
e1	2	1	0
e2	1	4	2
e3	0	2	2

THE END