# Social Network Analysis

## Basic Concepts, Methods & Theory

**University of Cologne**
**Johannes Putzke**

# Agenda

- Introduction
- Basic Concepts
- Mathematical Notation
- Network Statistics

# Textbooks

- Hanneman & Riddle (2005) *Introduction to Social Network Methods,* available at *http*://faculty.ucr.edu/~hanneman/nettext/
- Wasserman & Faust (1994): *Social Network Analysis – Methods and Applications*, Cambridge: Cambridge University Press.

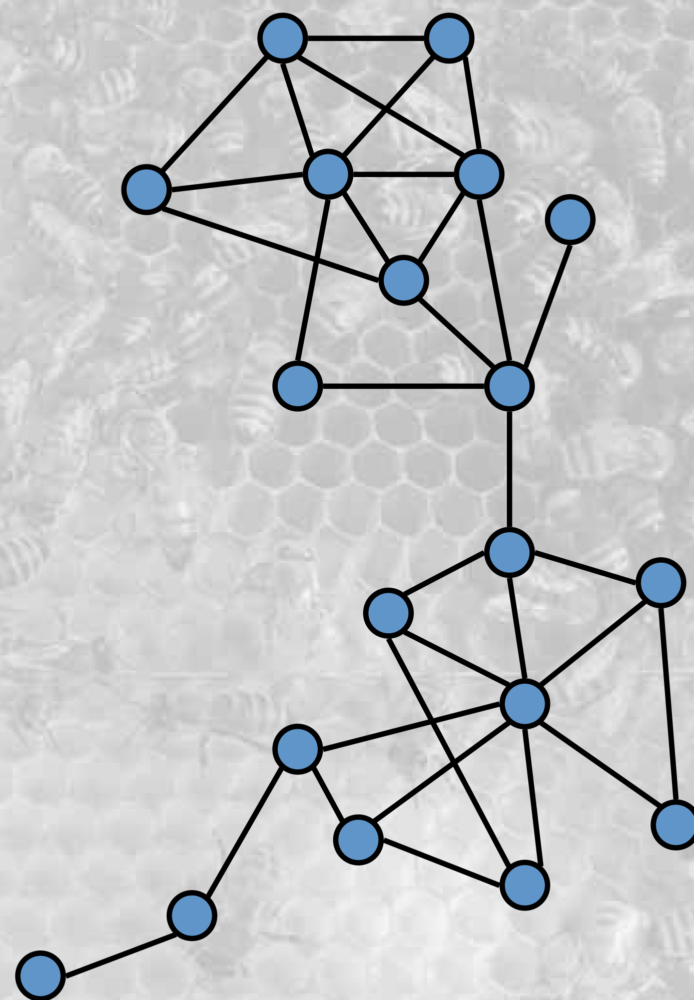
# Introduction

# Basic Concepts

What is a network?

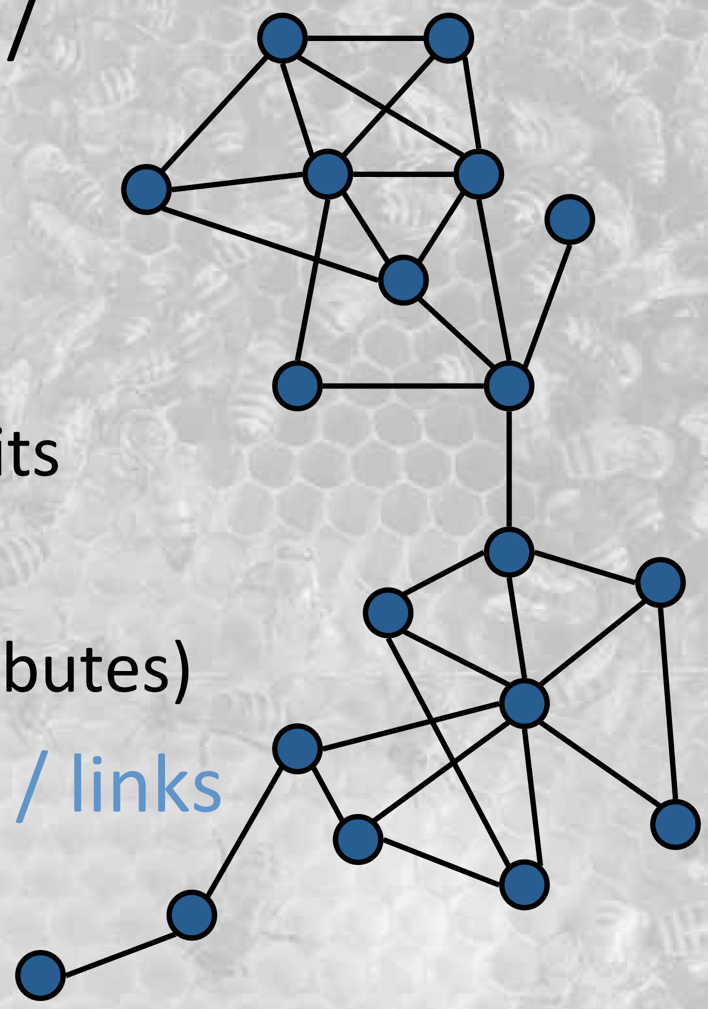
# What is a Network?

- Actors / nodes / vertices / points
- Ties / edges / arcs / lines / links



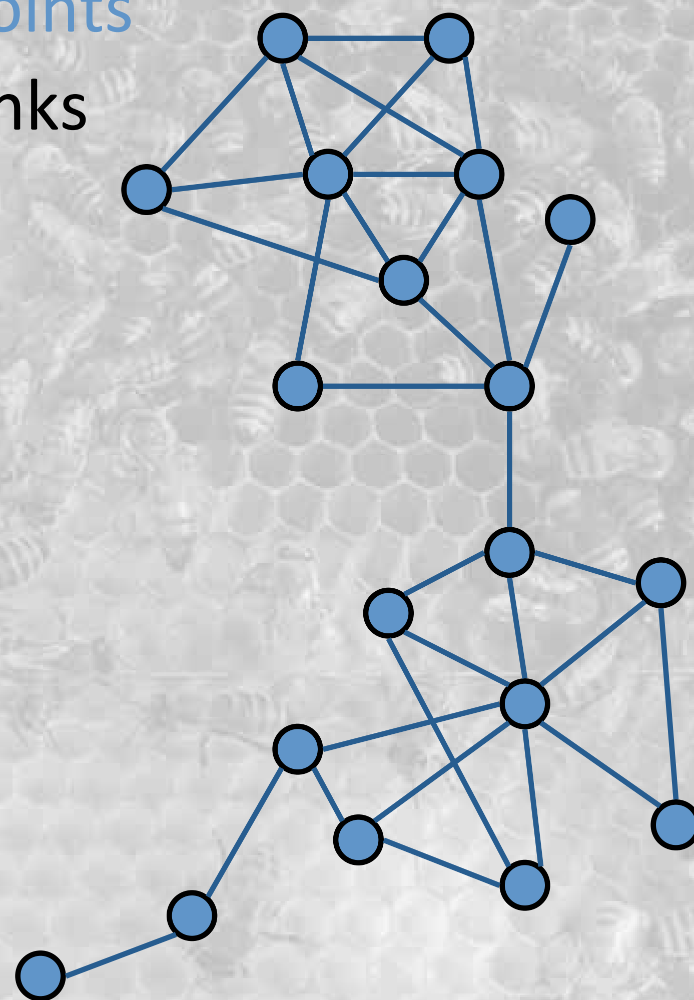
# What is a Network?

- Actors / nodes / vertices / points
  - Computers / Telephones
  - Persons / Employees
  - Companies / Business Units
  - Articles / Books
  - Can have properties (attributes)
- Ties / edges / arcs / lines / links

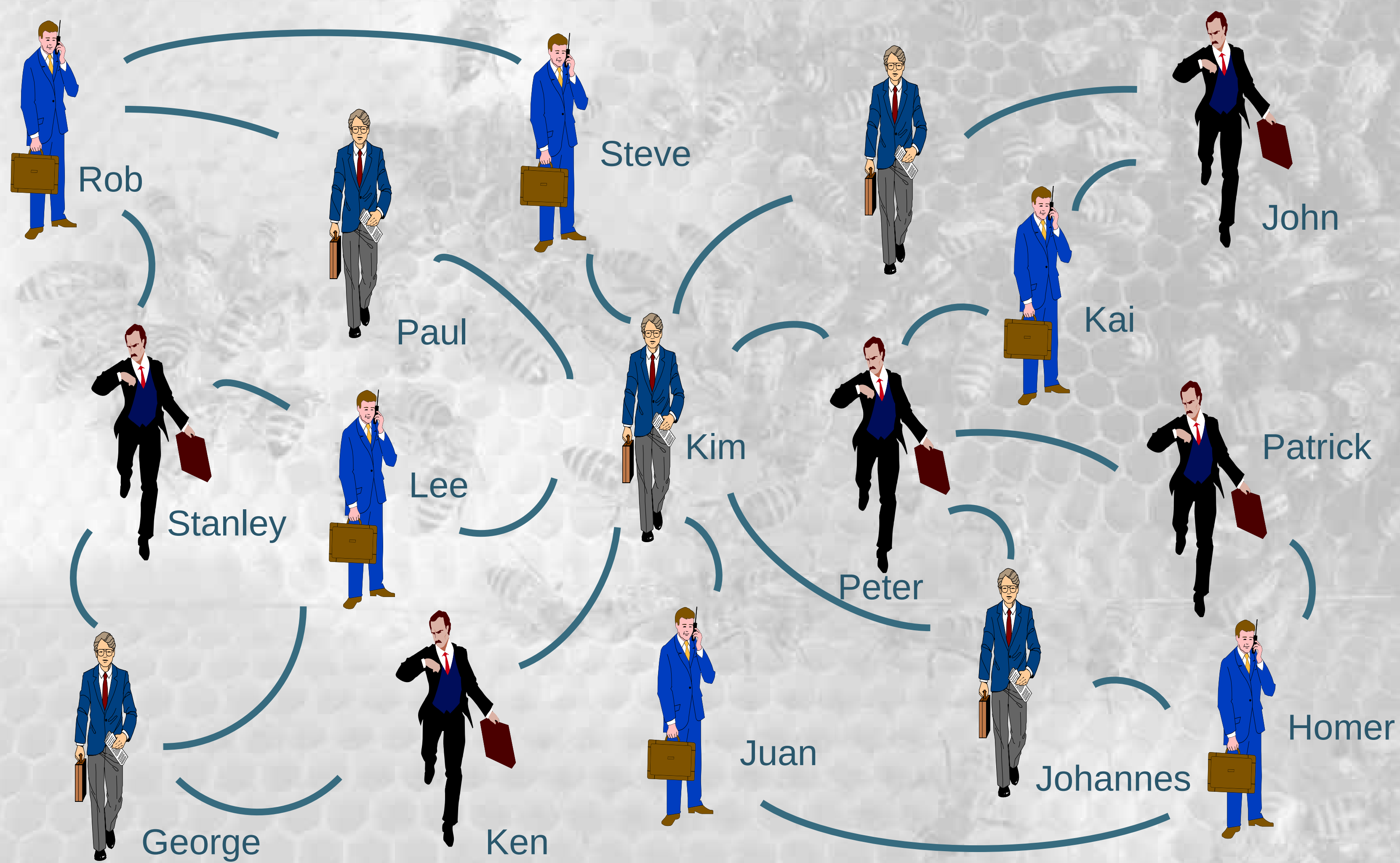


# What is a Network?

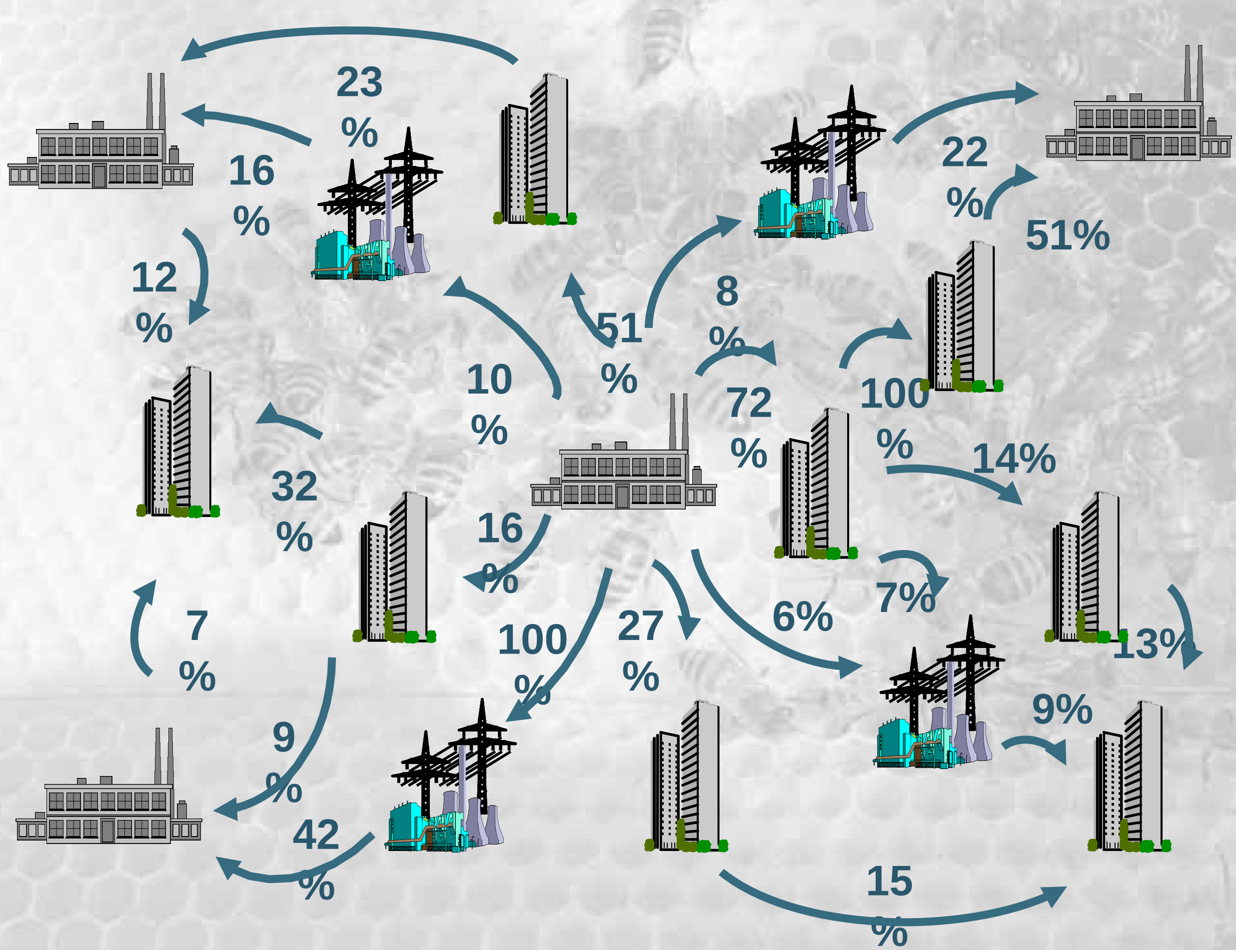
- Actors / nodes / vertices / points
- Ties / edges / arcs / lines / links
  - connect pair of actors
  - types of social relations
    - friendship
    - acquaintance
    - kinship
    - advice
    - hindrance
    - sex
  - allow different kind of flows
    - messages
    - money
    - diseases



# What is a Social Network? - Relations among People



# What is a Network? - Relations among Institutions




- as institutions
  - owned by, have partnership / joint venture
  - purchases from, sells to
  - competes with, supports
- through stakeholders
  - board interlocks
  - Previously worked for

# Why study social networks?

# Example 2) Homophily Theory

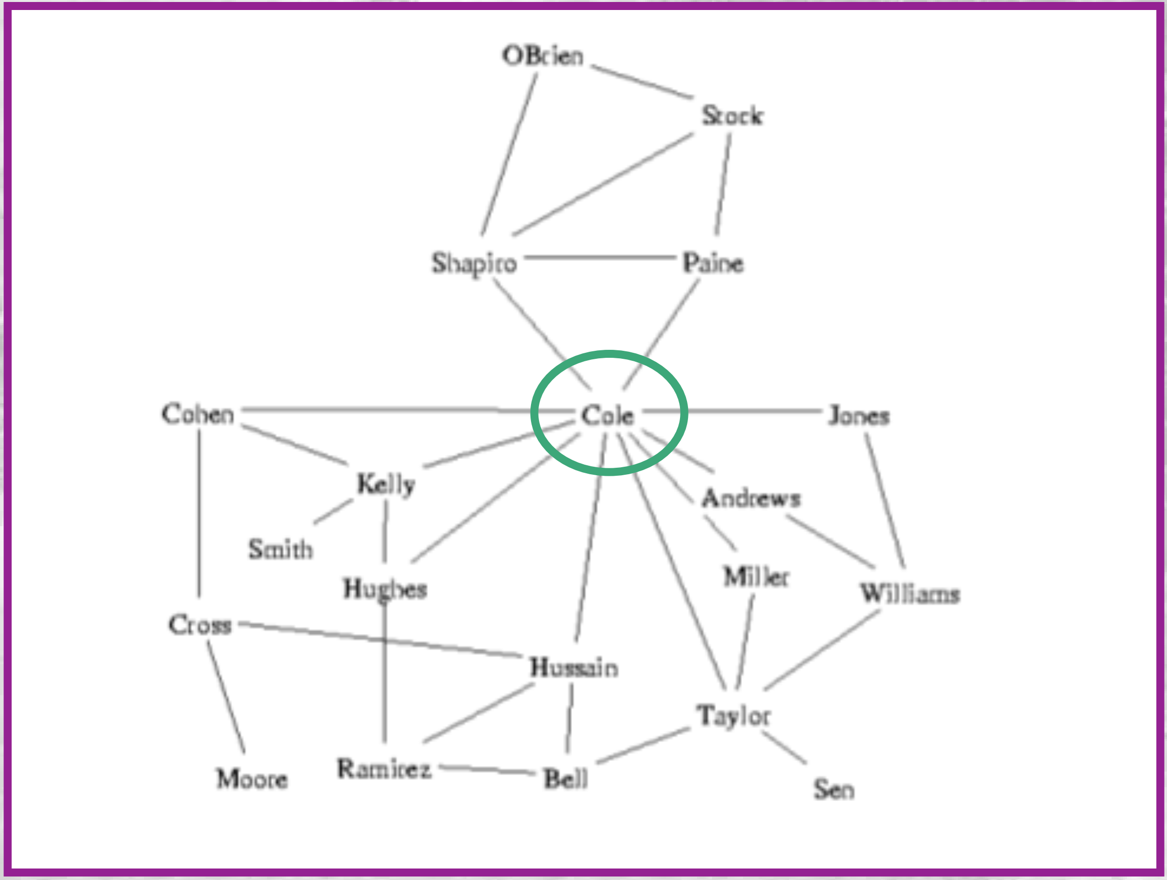
| | Male | Female |
|---|---|---|
| Male | 123 | 68 |
| Female | 95 | 164 |

- Birds of a feather flock together
- See McPherson, Smith-Lovin & Cook (2001)

| | 0-13 | 14-29 | 30-44 | 45-65 | >65 |
|---|---|---|---|---|---|
| 0-13 | 212 | 63 | 117 | 72 | 91 |
| 14-29 | 83 | 372 | 75 | 67 | 84 |
| 30-44 | 105 | 98 | 321 | 214 | 117 |
| 45-65 | 62 | 72 | 232 | 412 | 148 |
| >65 | 90 | 77 | 124 | 153 | 366 |

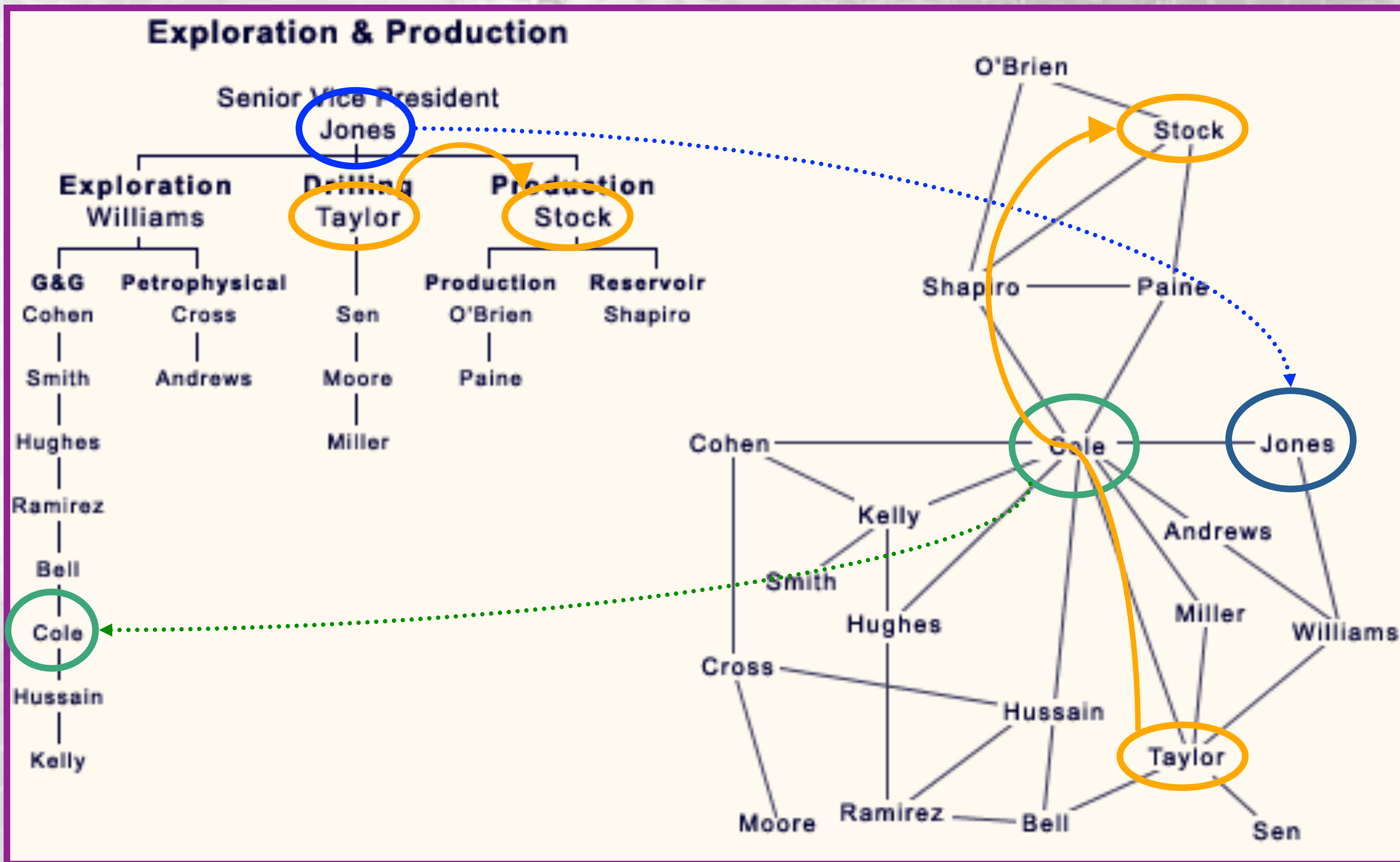
- age / gender → network

# Managerial Relevance – Social Network…



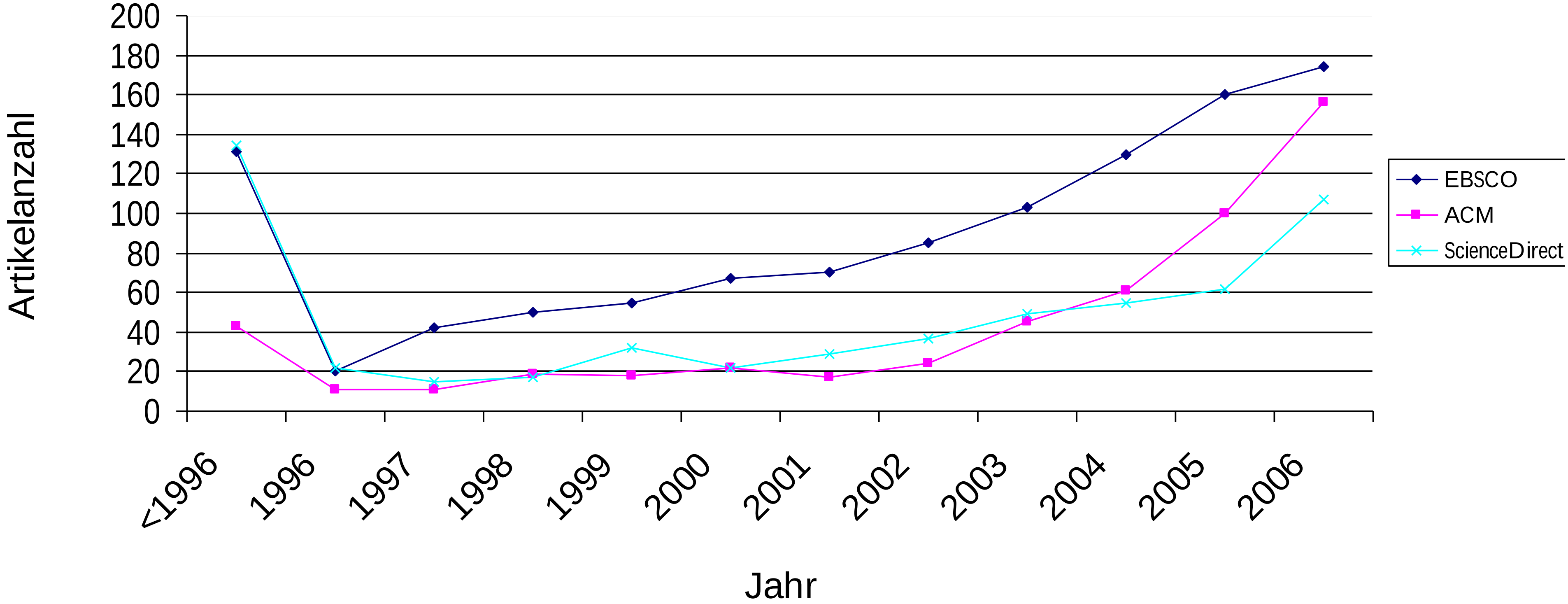

Source: http://www.robcross.org/sna.htm

# …vs. Organigram




Source: http://www.robcross.org/sna.htm

# SNA – A Recent Trend in IS Research



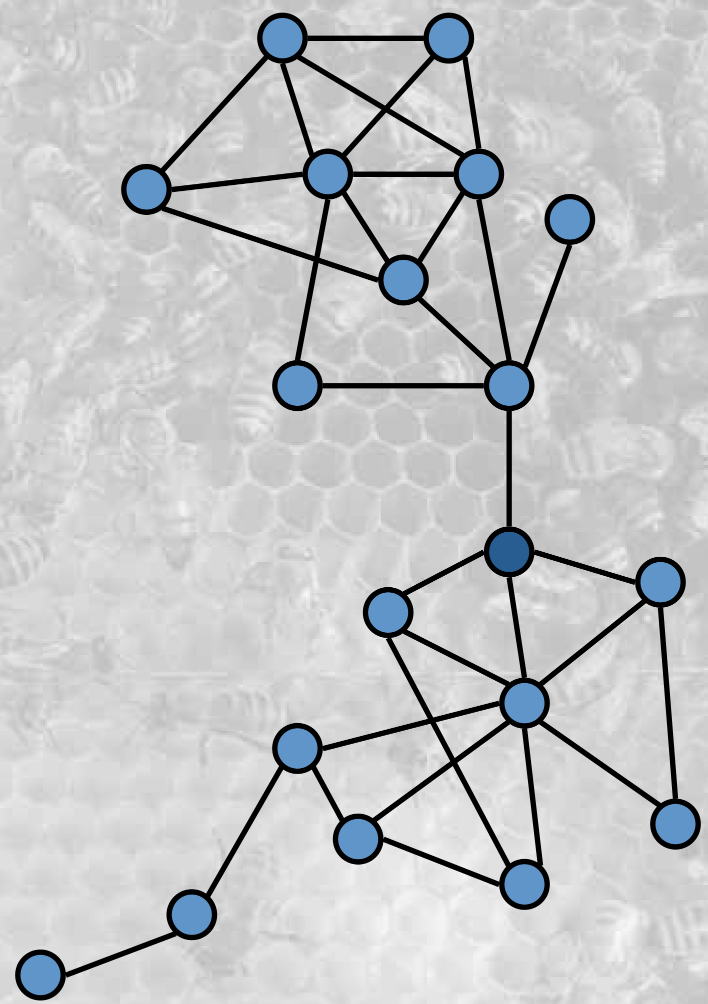
# How to analyze Social Networks?

# Example: Centrality Measures

- Who is the most prominent?
  - *Who knows the most actors?*
    *(Degree Centrality)*
  - Who has the shortest distance to the other actors?
  - Who controls knowledge flows?
  - ...

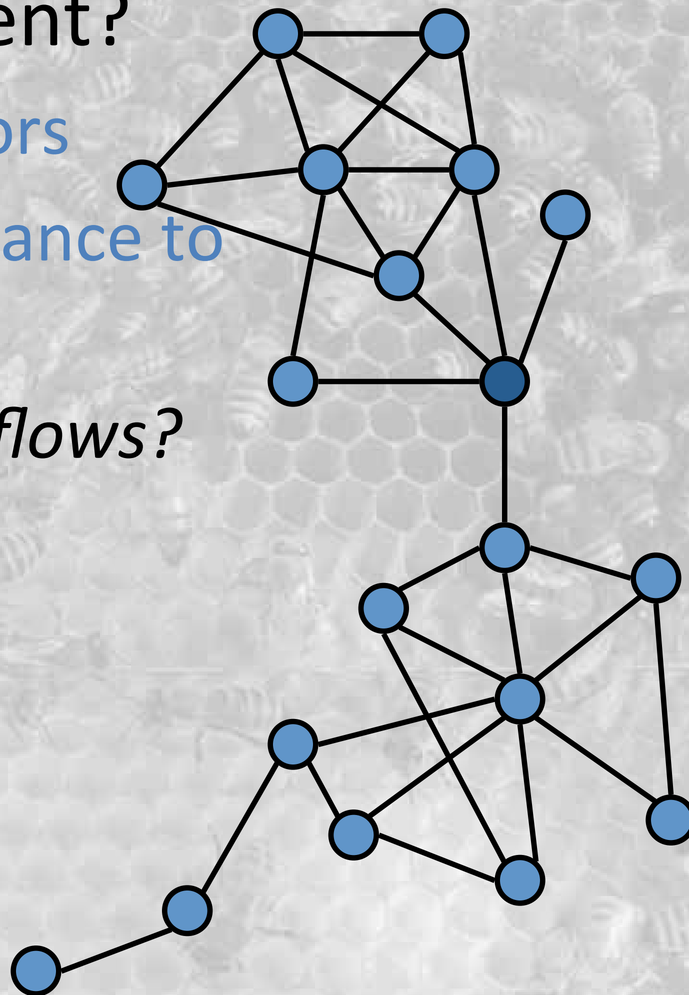
# Example: Centrality Measures

- Who is the most prominent?
  - Who knows the most actors?
  - *Who has the shortest distance to the other actors? (Closesness Centrality)*
  - Who controls knowledge flows?
  - …



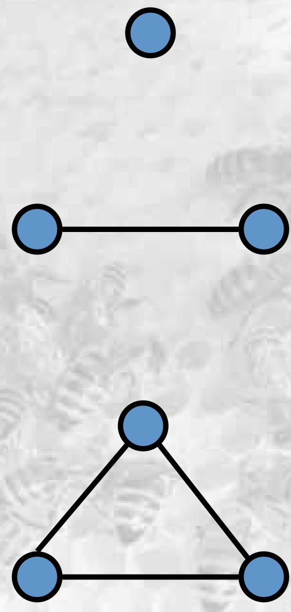
# Example: Centrality Measures

- Who is the most prominent?
  - Who knows the most actors
  - Who has the shortest distance to the other actors?
  - *Who controls knowledge flows? (Betweenness Centrality)*
  - …



# Basic Concepts

# Dyads, Triads and Relations

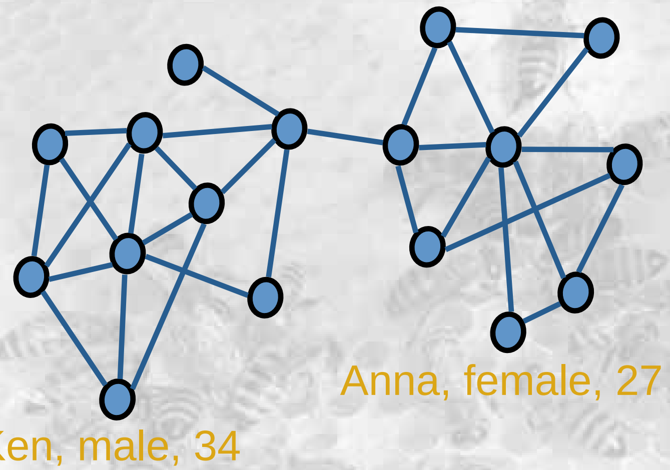


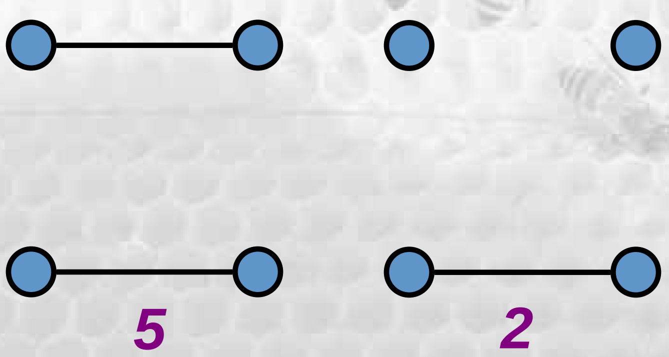
friendship

kinship

- *actor*
- *dyad*
- *triad*
- *relation:*
  - collection of specific ties among members of a group

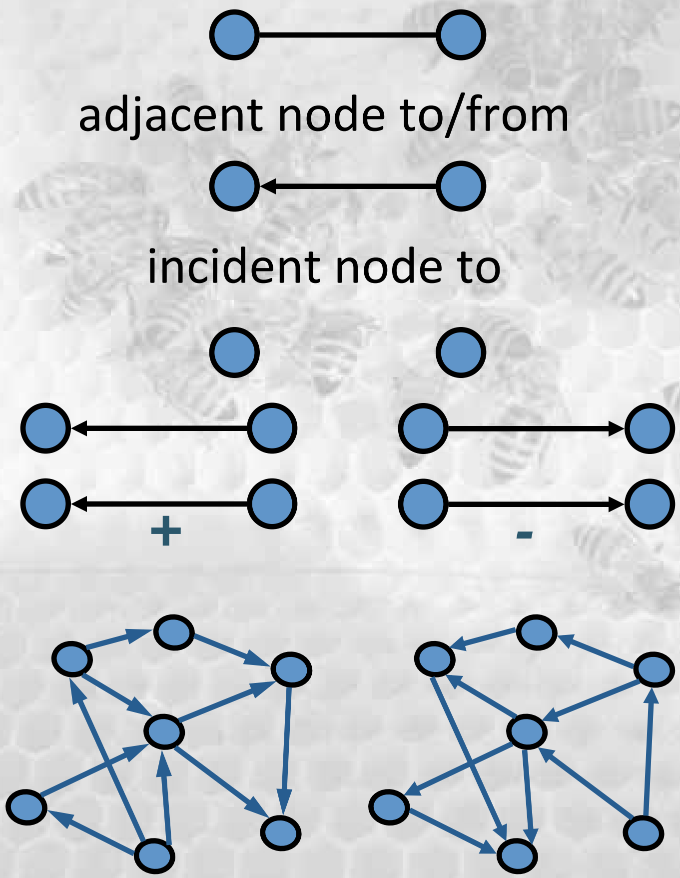
# Strength of a Tie




- *Social network*
  - finite set of actors and relation(s) defined on them
  - depicted in *graph/ sociogram*
    - *labeled graph*
- *Strength of a Tie*
  - *dichotomous* vs. *valued*
    - depicted in *valued graph* or *signed graph* (+/-)

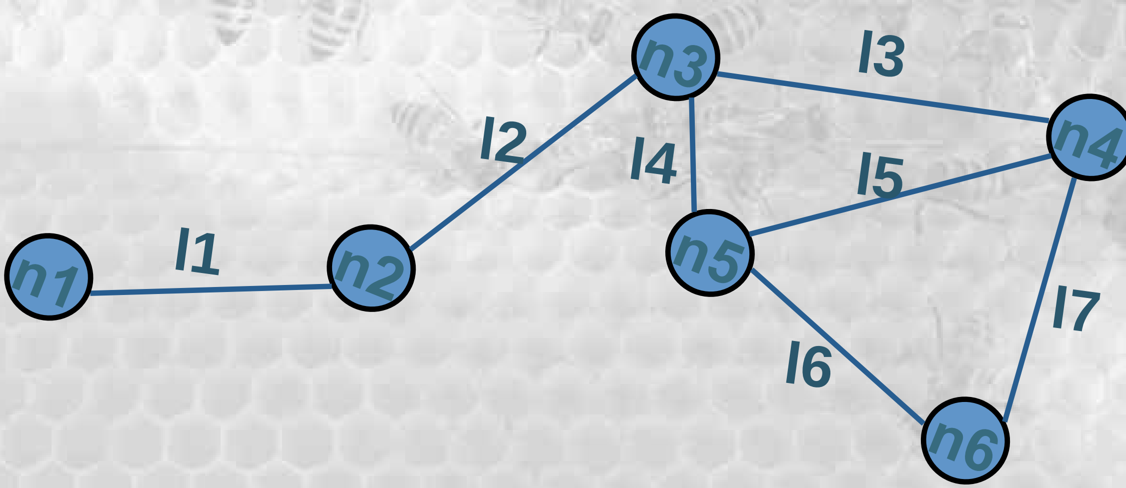
# Strength of a Tie




- *Strength of a Tie*
  - *nondirectional* vs.
    *directional*
    - depicted in *directed graphs (digraphs)*
    - nodes connected by *arcs*
    - 3 *isomorphism classes*
      - *null dyad*
      - *mutual / reciprocal / symmetrical dyad*
      - *asymmetric / antisymmetric dyad*
    - *converse of a digraph*
      - *reverse direction of all arcs*

# Reachability, Distances and Diameter

- *Reachability*
  - If there is a path between nodes $n_i$ and $n_j$
- *Geodesic*
  - Shortest path between two nodes
- *(Geodesic) Distance* d(i,j)
  - Length of Geodesic (also called „degrees of separation“)



# Mathematical Notation and Fundamentals

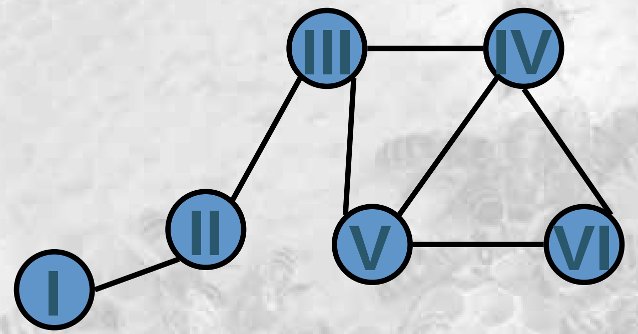
# Three different notational schemes

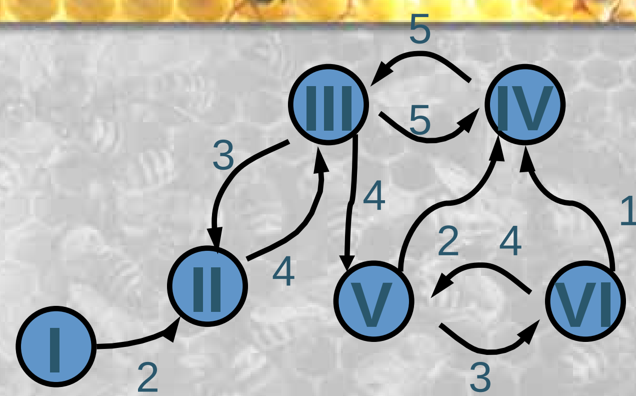
1. Graph theoretic
2. Sociometric
3. Algebraic

# 1. Graph Theoretic Notation

- N Actors $\{n_1, n_2, \ldots, n_g\}$
- $n_1 \rightarrow n_j$ there is a tie between the ordered pair $<n_i, n_j>$
- $n_1 \rightarrow n_j$ there is no tie
- $(n_i, n_j)$ nondirectional relation
- $<n_i, n_j>$ directional relation
- g(g-1) number of ordered pairs in $<n_i, n_j>$ directional network
- g(g-1)/2 number of ordered pairs in nondirectional network
- L collection of ordered pairs with ties $\{l_1, l_2, \ldots, l_g\}$
- G graph descriped by sets (N, L)
- *Simple graph* has no *reflexive ties, loops*

# 2. Sociometric Notation - From Graphs to (Adjacency/ Socio)-Matrices




Binary, undirected

| | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|
| I | - | | | | | |
| II | 1 | - | | | | |
| III | | 1 | - | | | |
| IV | | | 1 | - | | |
| V | | | 1 | 1 | - | |
| VI | | | | 1 | 1 | - |

symmetrical

Valued, directed

| | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|
| I | | 2 | 0 | 0 | 0 | 0 |
| II | 0 | 0 | 4 | 0 | 0 | 0 |
| III | 0 | 3 | 0 | 5 | 4 | 0 |
| IV | 0 | 0 | 5 | 0 | 0 | 0 |
| V | 0 | 0 | 0 | 2 | 0 | 3 |
| VI | 0 | 0 | 0 | 1 | 4 | 0 |

# 2. Sociometric Notation

- X g × g *sociomatrix* on a single relation
  g × g × R *super-sociomatrix* on R relations

  - $X_R$ sociomatrix on relation R

- $X_{ij(r)}$ value of tie from $n_i$ to $n_i$ (on relation $\chi_r$) where $i \neq j$

# 2. Sociometric Notation – From Matrices to Adjacency Lists and Arc Lists

| | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|
| I | - | 1 | | | | |
| II | 1 | - | 1 | | | |
| III | | 1 | - | 1 | 1 | |
| IV | | | 1 | - | 1 | 1 |
| V | | | 1 | 1 | - | 1 |
| VI | | | | 1 | 1 | - |

## Adjacency List

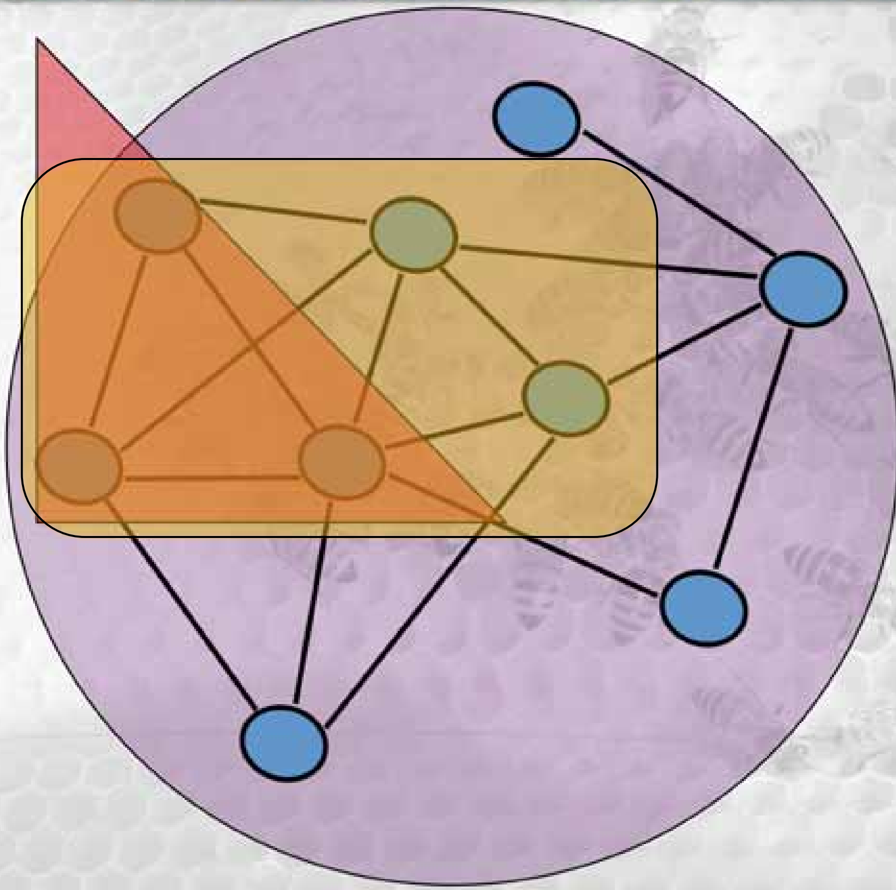
| | |
|---|---|
| I | II |
| II | I III |
| III | II IV V |
| IV | III V VI |
| V | III V VI |
| VI | IV V |

## Arc List

I II
II I
II III
III II
III IV
III V
IV III
IV V
IV VI
V III
V IV
V VI
VI IV
VI IV

# Network Statistics

# Different Levels of Analysis



- Actor-Level
- Dyad-Level
- Triad-Level
- Subset-level (cliques / subgraphs)
- Group (i.e. global) level

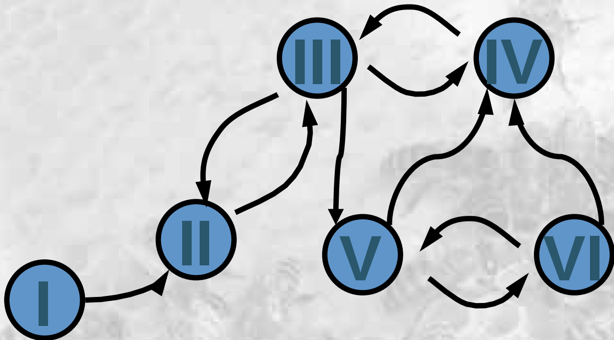
# Measures at the Actor-Level:

# Measures of Prominence: Centrality and Prestige

# Degree Centrality

- *Who knows the most actors? (Degree Centrality)*
- Who has the shortest distance to the other actors?
- Who controls knowledge flows?
- ...

# Degree Centrality I




| | I | II | III | IV | V | VI | |
|---|---|---|---|---|---|---|---|
| I | | 1 | 0 | 0 | 0 | 0 | 1 |
| II | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| III | 0 | 1 | 0 | 1 | 1 | 0 | 3 |
| IV | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| V | 0 | 0 | 0 | 1 | 0 | 1 | 2 |
| VI | 0 | 0 | 0 | 1 | 0 | 1 | 2 |
| | 0 | 2 | 2 | 3 | 1 | 2 | |

**Marginals of adjacency matrix**

- *Indegree $d_I(n_i)$*
  - *Popularity, status, deference, degree prestige*

$$C_{DI}\left(n_i\right) = d_I(n_i) = \sum_j^g x_{ji} = x_{+i}$$

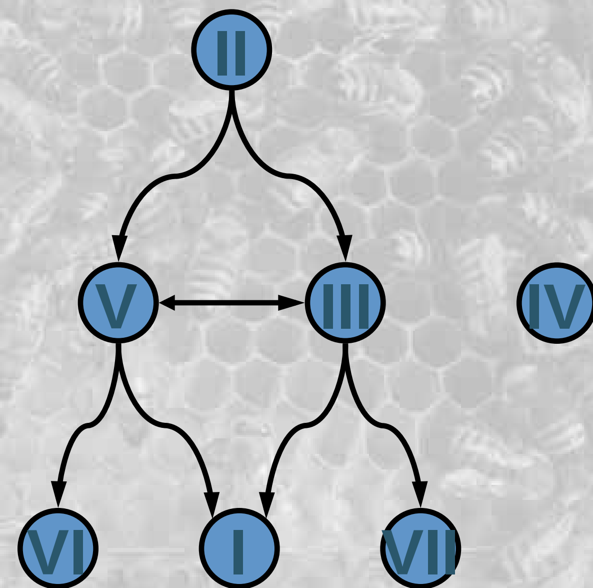
- *Outdegree $d_O(n_i)$*
  - Expansiveness

$$C_{DO}\left(n_i\right) = d_o(n_i) = \sum_j^g x_{ij} = x_{i+}$$

- Total degree ≡ 2 x number of edges

# Degree Centrality II

- Interpretation: opportunity to (be) influence(d)
- Classification of Nodes
  - *Isolates*
    - $d_I(n_i) = d_O(n_i) = 0$
  - *Transmitters*
    - $d_I(n_i) = 0$ and $d_O(n_i) > 0$
  - *Receivers*
    - $d_I(n_i) > 0$ and $d_O(n_i) = 0$
  - *Carriers / Ordinaries*
    - $d_I(n_i) > 0$ and $d_O(n_i) > 0$

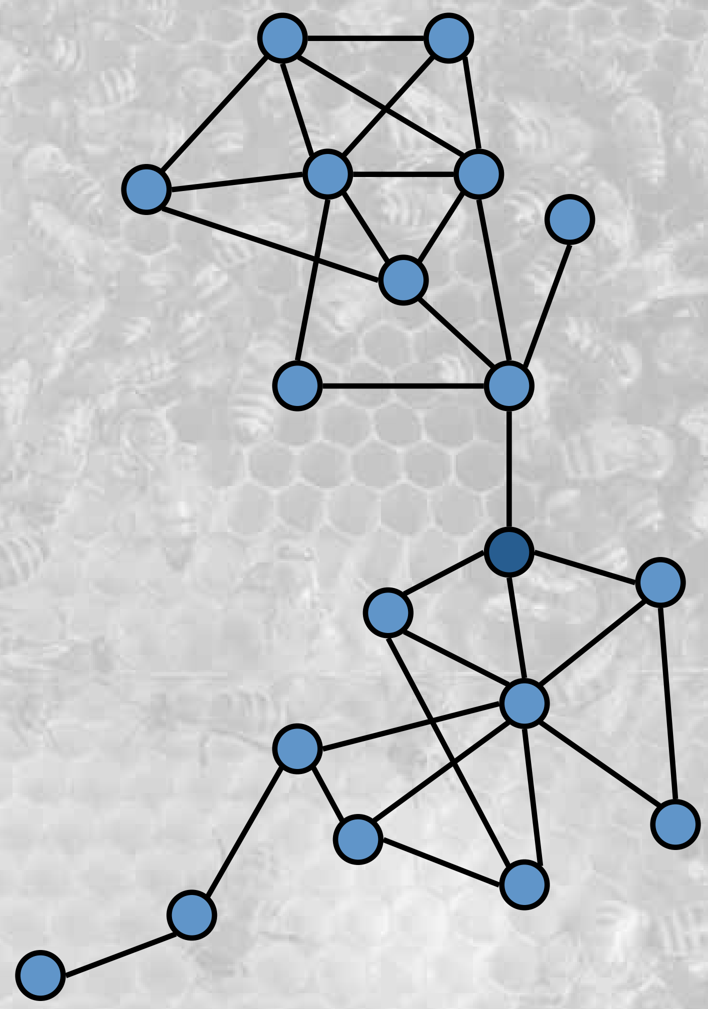



- Standardization of $C_D$ to allow comparison across networks of different sizes: divide by ist maximum value

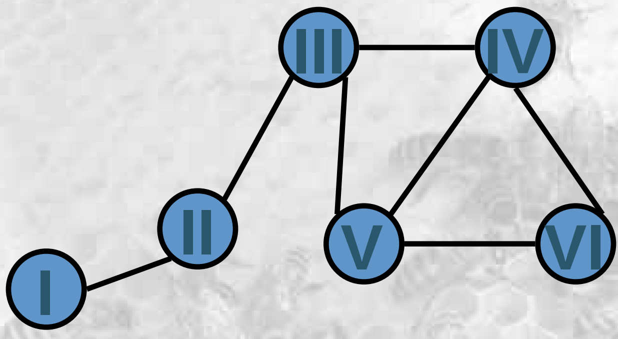
$$C'_D(n_i) = \frac{d(n_i)}{g-1}$$

# Closeness Centrality

- Who knows the most actors?
- *Who has the shortest distance to the other actors? (Closesness Centrality)*
- Who controls knowledge flows?
- …



# Closeness Centrality




| | I | II | III | IV | V | VI | |
|---|---|---|---|---|---|---|---|
| I | - | 1 | 2 | 3 | 3 | 4 | 13 |
| II | 1 | - | 1 | 2 | 2 | 3 | 9 |
| III | 2 | 1 | - | 1 | 1 | 2 | 7 |
| IV | 3 | 2 | 1 | - | 1 | 1 | 8 |
| V | 3 | 2 | 1 | 1 | - | 1 | 8 |
| VI | 4 | 3 | 2 | 1 | 1 | - | 11 |

- Index of expected arrival time

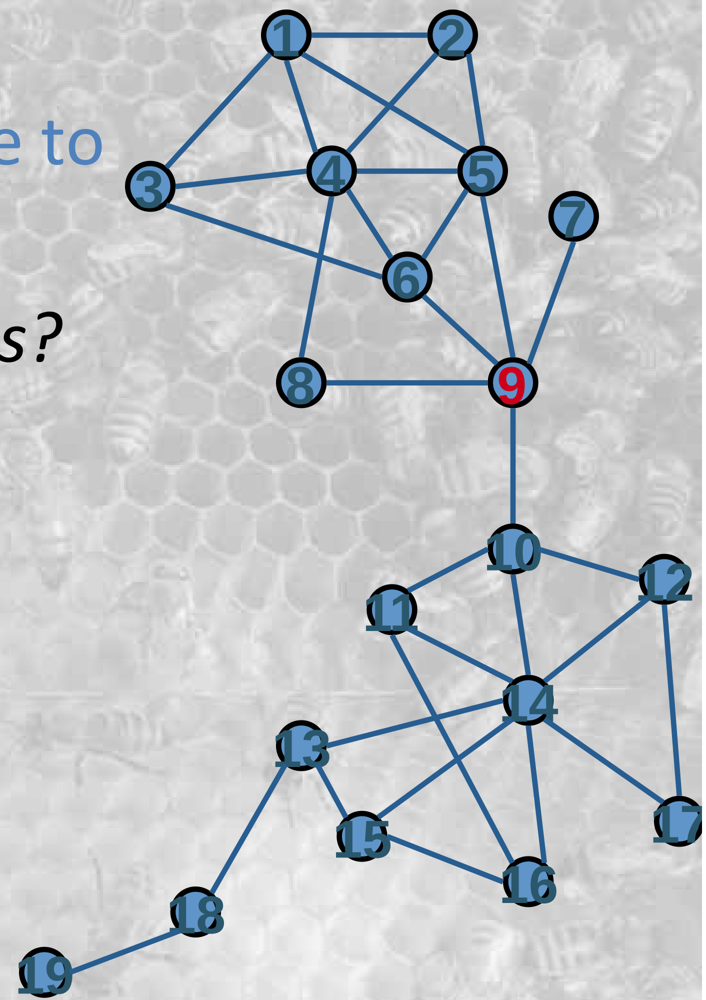
$$C_C(n_i) = \frac{1}{\sum_{j=1}^{g} d\left(n_i, n_j\right)}$$

**Reciprocal of marginals of geodesic distance matrix**

- Standardize by multiplying (g-1)
- Problem: Only defined for connected graphs

# Betweenness Centrality

- Who knows the most actors?
- Who has the shortest distance to the other actors?
- *Who controls knowledge flows? (Betweenness Centrality)*
- ...

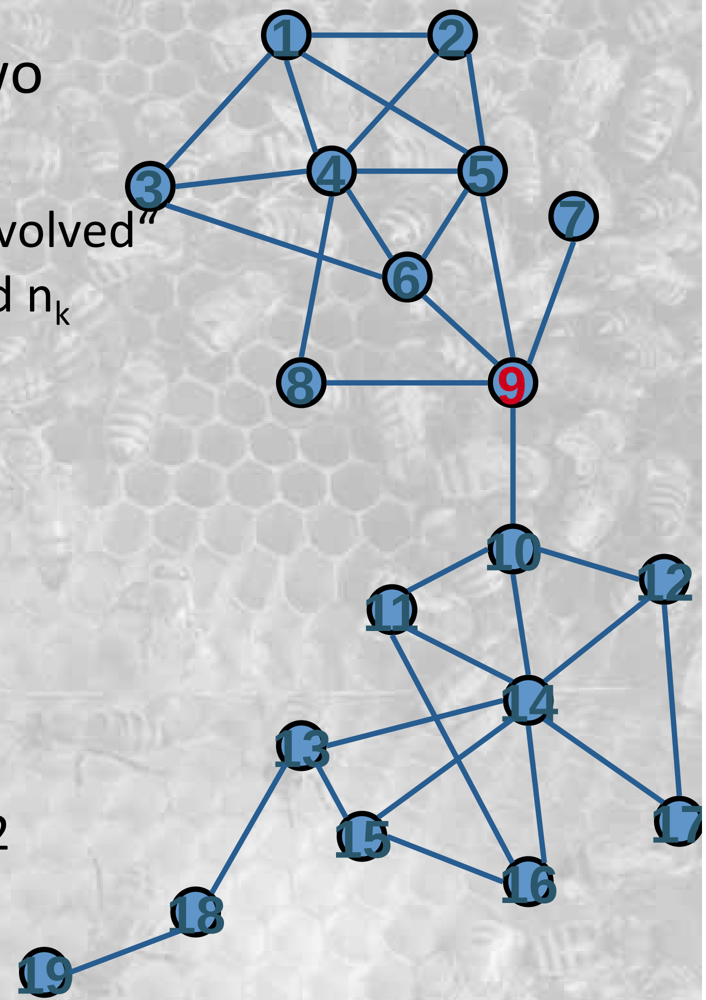


# Betweenness Centrality

- How many geodesic linkings between two actors j and k contain actor i?
  - $g_{jk}(n_i)/g_{jk}$ probability that distinct actor $n_i$ „involved“ in communication between two actors $n_j$ and $n_k$

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

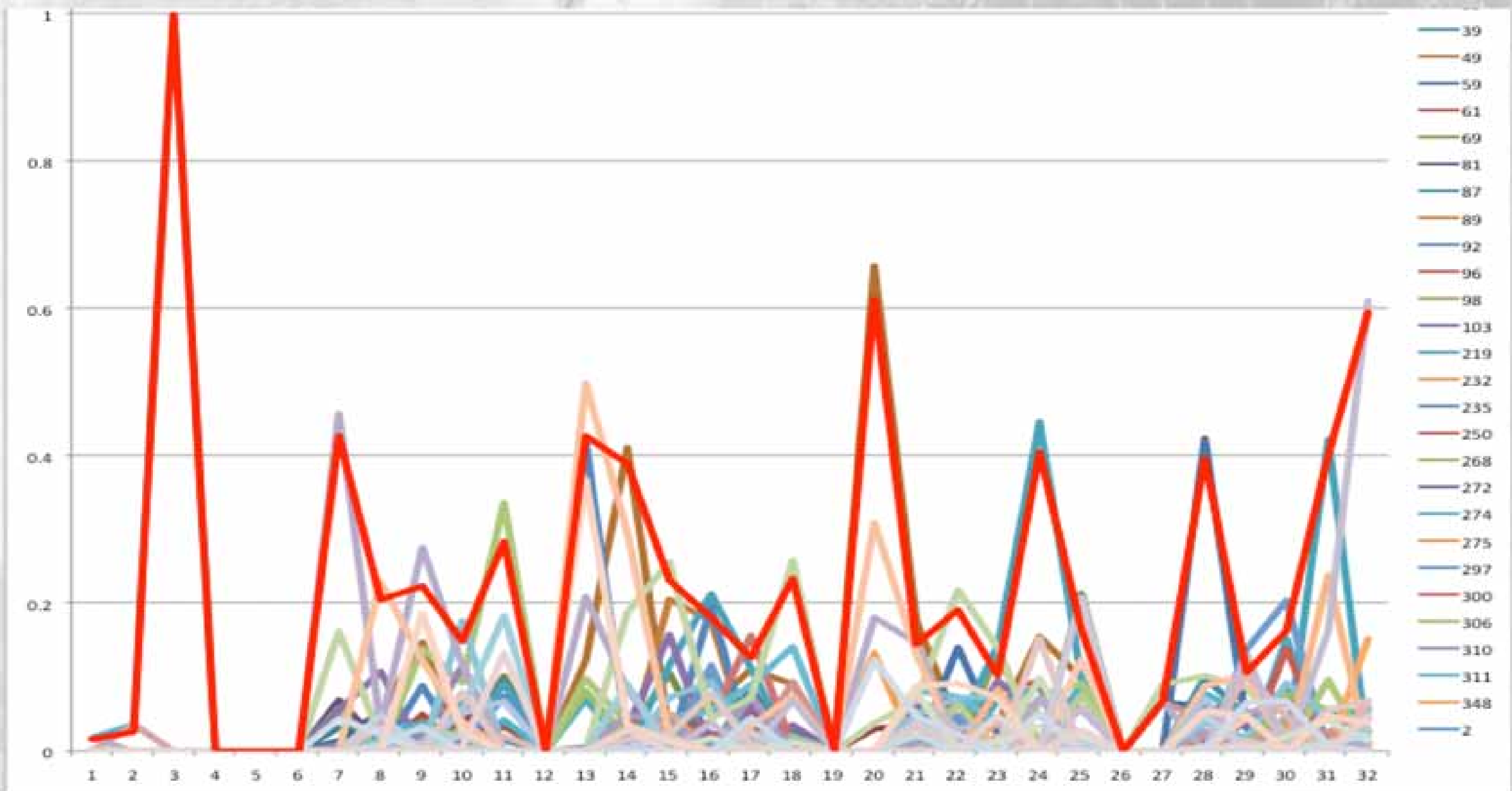
  - standardized by dividing through (g-1)(g-2)/2



# Several other Centrality Measures

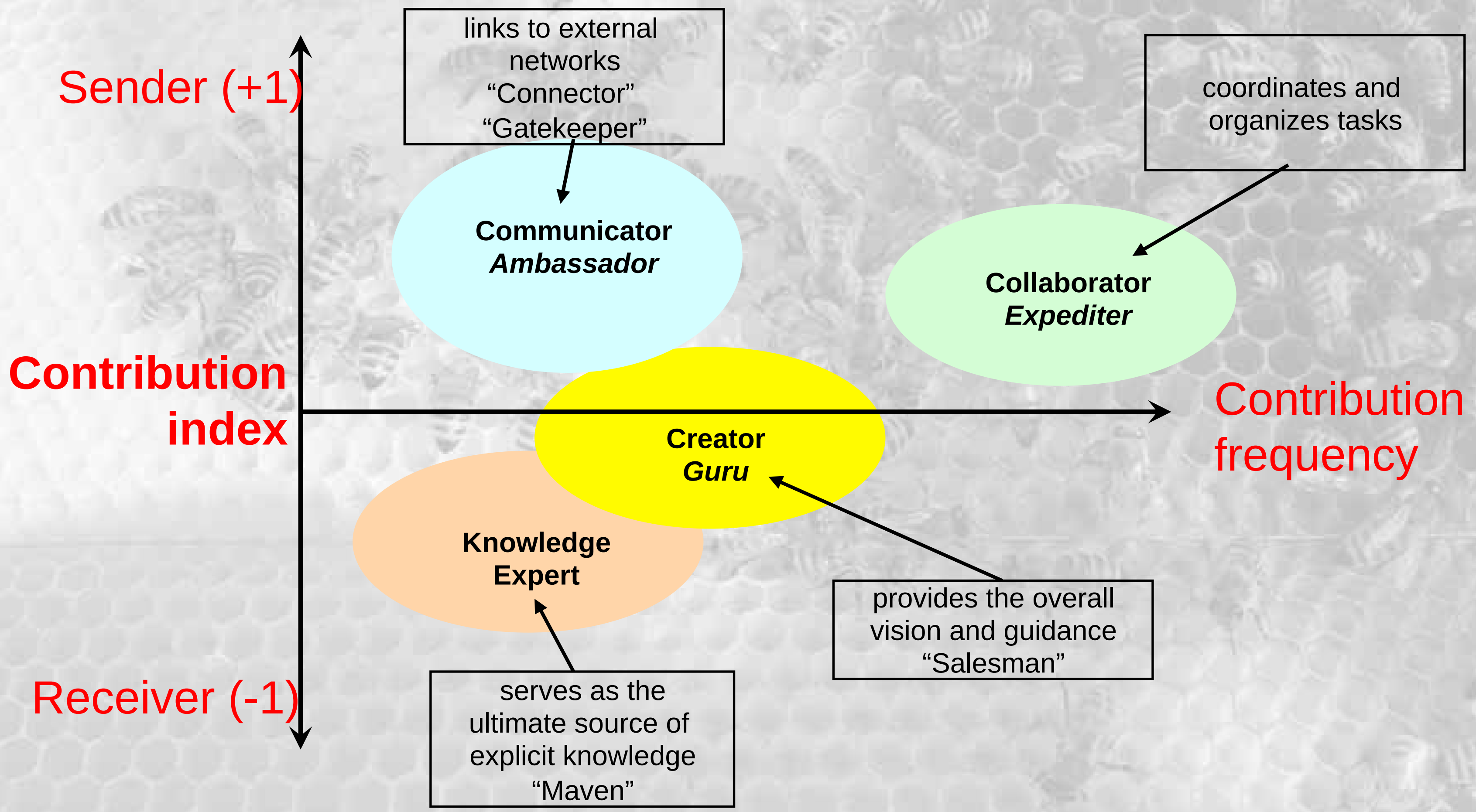
- …beyond the scope of this lecture
  - *Status or Rank Prestige, Eigenvector Centrality*
    - also reflects status or prestige of people whom actor is linked to
    - Appropriate to identify *hubs* (actors adjacent to many peripheral nodes) and *bridges* (actors adjacent to few central actors)
      - attention: more common, different meaning of bridge!!!
  - *Information Centrality*
  - see Wasserman & Faust (1994), p. 192 ff.
  - *Random Walk Centrality*
    - see Newman (2005)

# Condor – Betweenness Centrality



# (Actor) Contribution Index

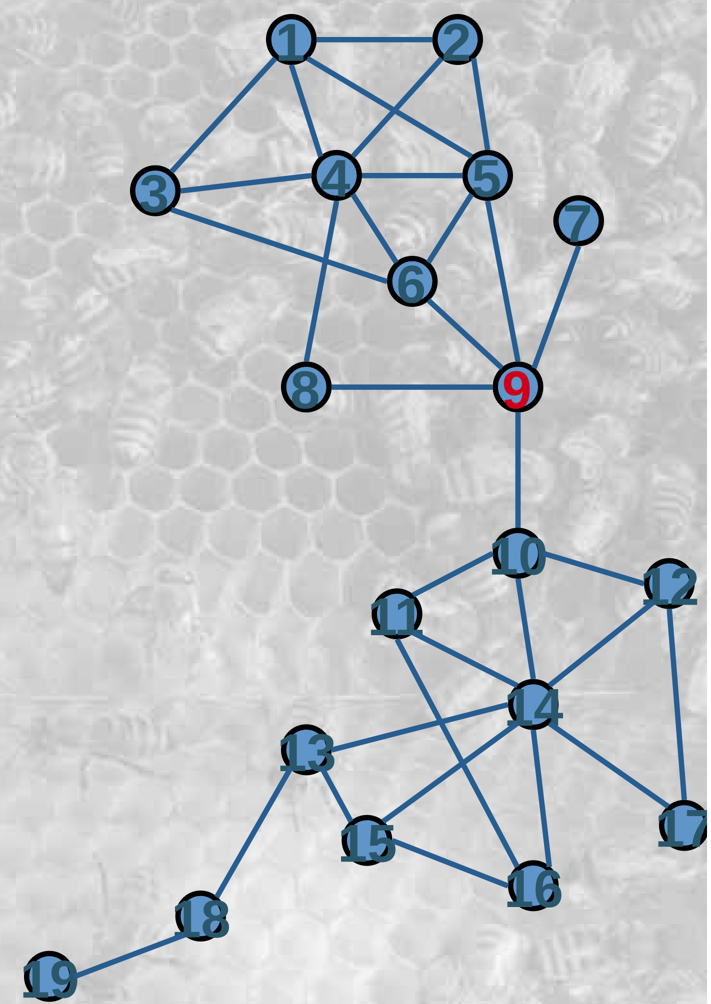
$$\frac{messages\_sent - messages\_received}{messages\_sent + messages\_received}$$



# Measures at the Group-(Global-)Level and Subgroup-Level

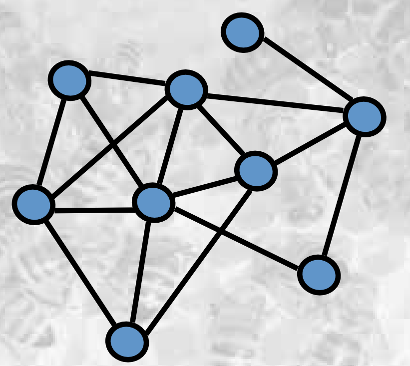
# Diameter of a Graph and Average Geodesic Distance

- Diameter
  - Largest geodesic distance between any pair of nodes
- Average Geodesic Distance
  - How fast can information get transmitted?

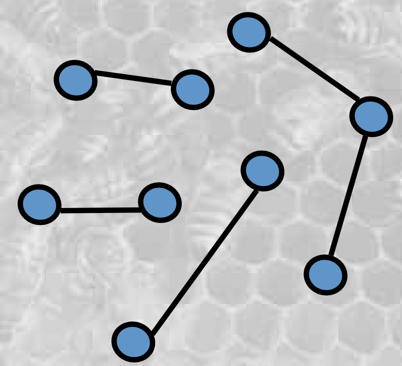


# Density

- Proportion of ties in a graph

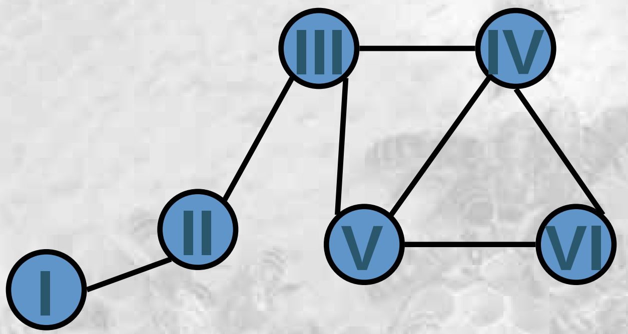


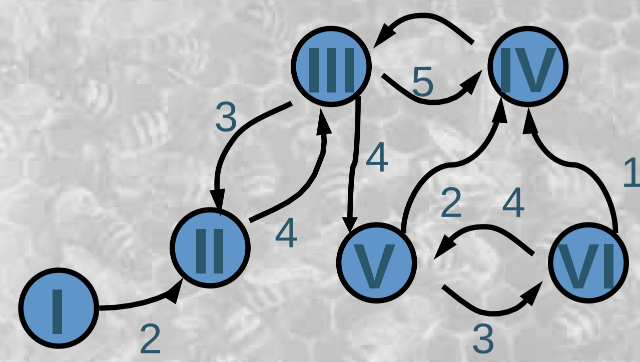
High density (44%)



Low density (14%)

# Density




$$\Delta = \frac{L}{g(g-1)/2} = \frac{L}{\binom{g}{2}}$$

**In undirected graph:**

**Proportion of ties**

$$\Delta = \frac{\sum_{i=1}^{g}\sum_{j=1}^{g} x_{ij}}{g(g-1)}$$

**In valued directed graph:**

**Average strength of the arcs**

# Group Centralization I

- How equal are the individual actors' centrality values?
  - $C_A(n_i^*)$ actor centrality index
  - $C_A(n^*)$ $\max_i C_A(n_i^*)$
  - $\sum_{i=1}^{g}\left[C_A\left(n^*\right)-C_A\left(n_i\right)\right]$ sum of difference between largest value and observed values
- General centralization index:

$$C_A = \frac{\sum_{i=1}^{g}\left[C_A\left(n^*\right)-C_A\left(n_i\right)\right]}{\max \sum_{i=1}^{g}\left[C_A\left(n^*\right)-C_A\left(n_i\right)\right]}$$

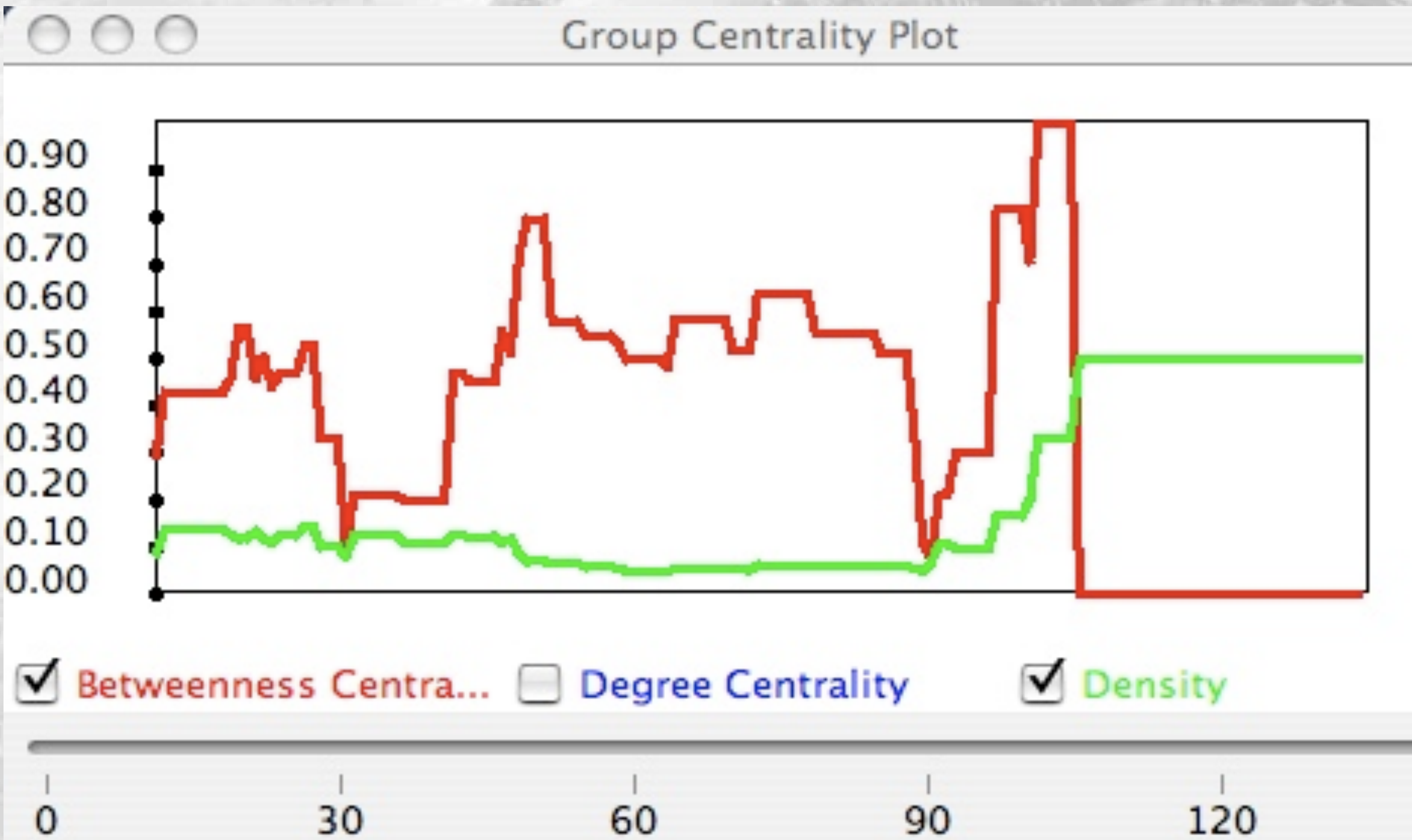
# Group Centralization II

$$C_D = \frac{\sum_{i=1}^{g}\left[C_D\left(n^*\right) - C_D\left(n_i\right)\right]}{(g-1)(g-2)}$$

$$C_C = \frac{\sum_{i=1}^{g}\left[C_C^{'}\left(n^*\right) - C_C^{'}\left(n_i\right)\right]}{\left[(g-1)(g-2)\right](2g-3)}$$

$$CB = \frac{\sum_{i=1}^{g}\left[C_B\left(n^*\right) - C_B\left(n_i\right)\right]}{(g-1)^2(g-2)} = \frac{\sum_{i=1}^{g}\left[C_B^{'}\left(n^*\right) - C_B^{'}\left(n_i\right)\right]}{(g-1)}$$

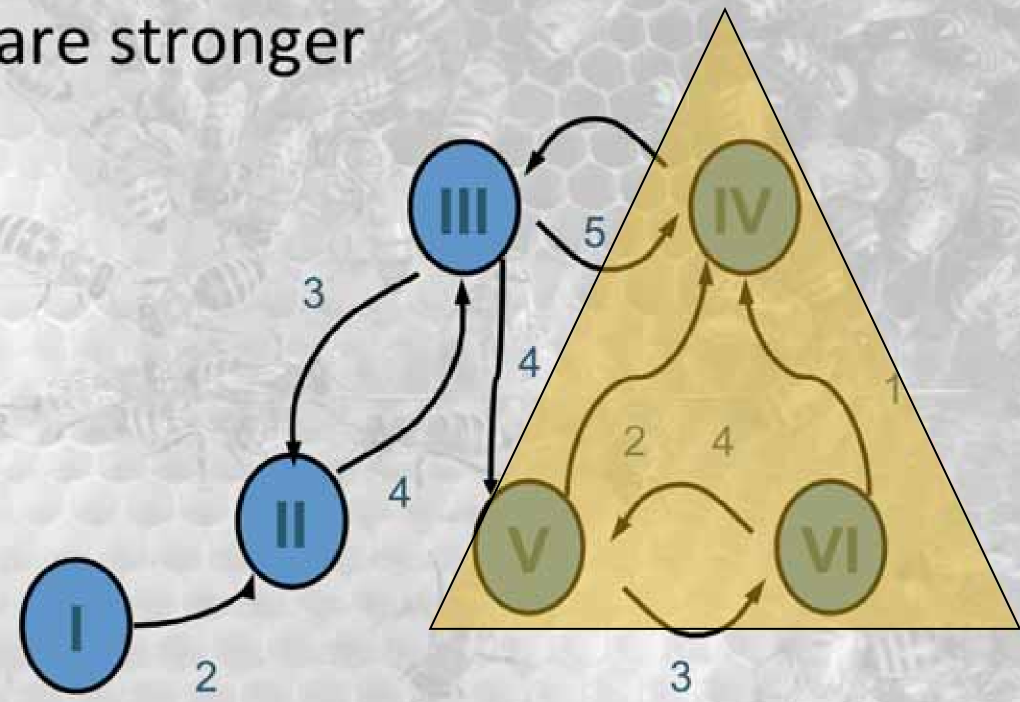
# Condor – Group Centralization



# Subgroup Cohesion

- average strength of ties within the subgroup divided by average strength of ties that are from subgroup members to outsiders
- >1 → ties in subgroup are stronger

$$\frac{\dfrac{\sum_{i \in N_s} \sum_{j \in N_s} x_{ij}}{g_s(g_s - 1)}}{\dfrac{\sum_{i \in N_s} \sum_{j \notin N_s} x_{ij}}{g_s(g - g_s)}}$$

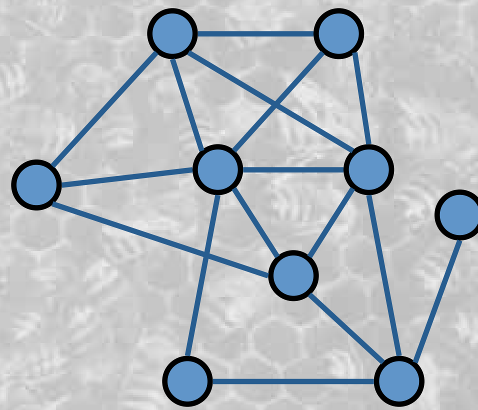


# Connectivity of Graphs and Cohesive Subgroups

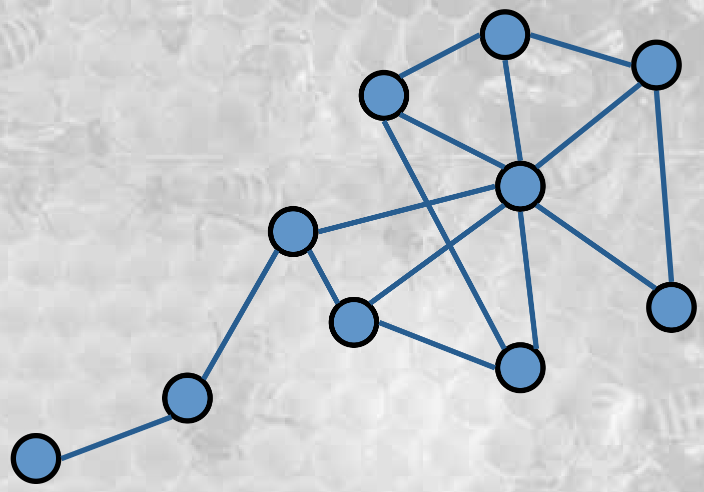
# Connectivity of Graphs

# Connected Graphs, Components, Cutpoints and Bridges

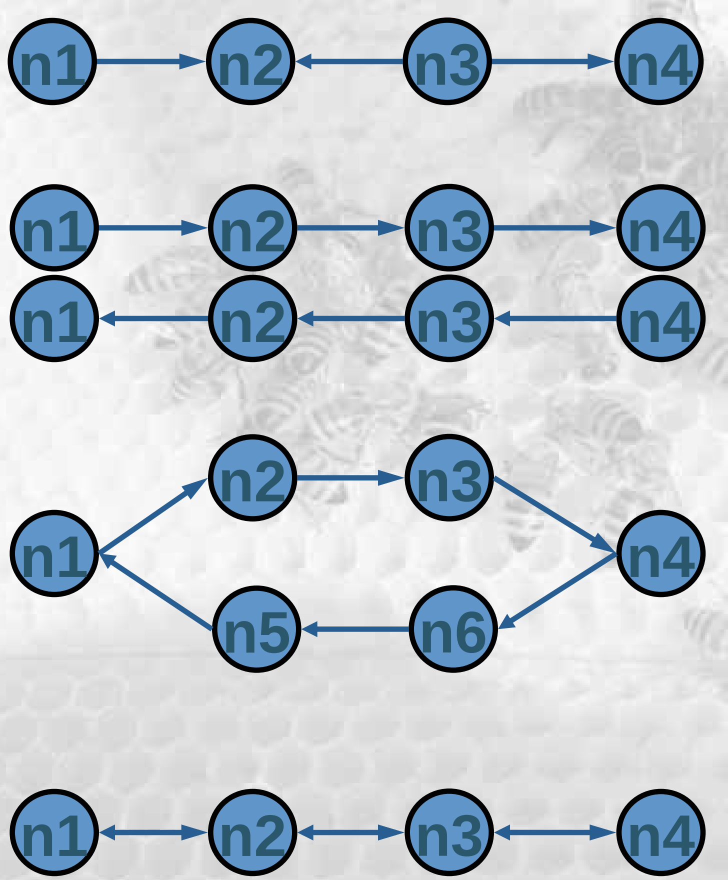
- *Connectedness*
  - A graph is connected if there is a path between every pair of nodes



- *Components*
  - Connected subgraphs in a graph
  - Connected graph has 1 component
  - Two disconnected graphs are one social network!!!



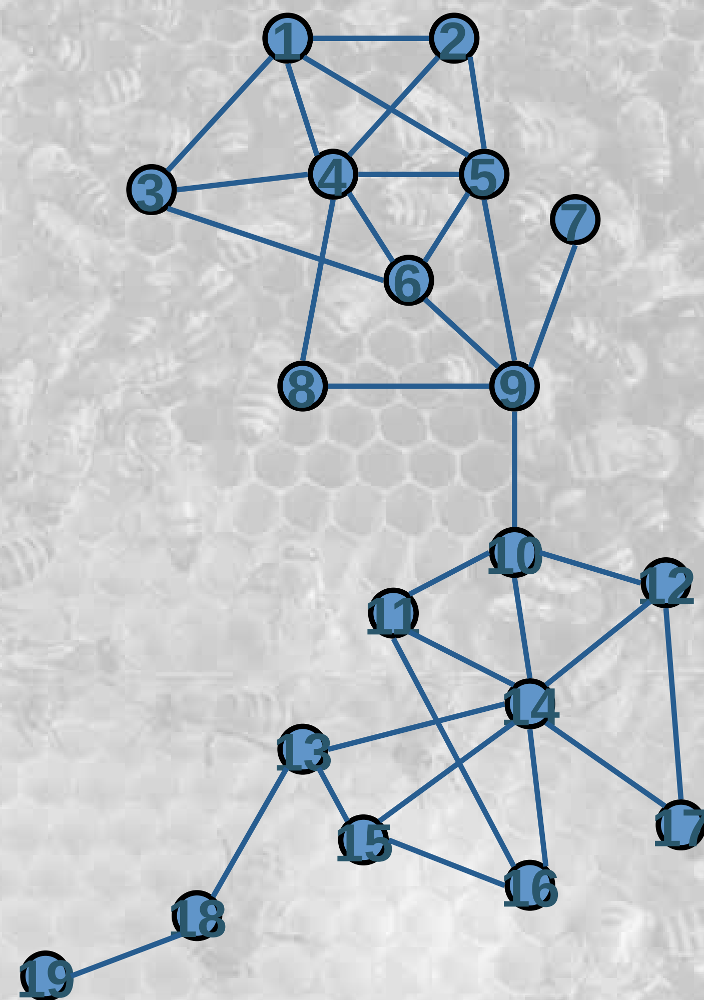
# Connected Graphs, Components, Cutpoints and Bridges




- Connectivity of pairs of nodes and graphs
  - *Weakly connected*
    - Joined by semipath
  - *Unilaterally connected*
    - Path from $n_j$ to $n_j$ or from $n_j$ to $n_j$
  - *Strongly connected*
    - Path from $n_j$ to $n_j$ *and* from $n_j$ to $n_j$
    - Path may contain different nodes
  - *Recursively Connected*
    - Nodes are strongly connected and both paths use the same nodes and arcs in reverse order

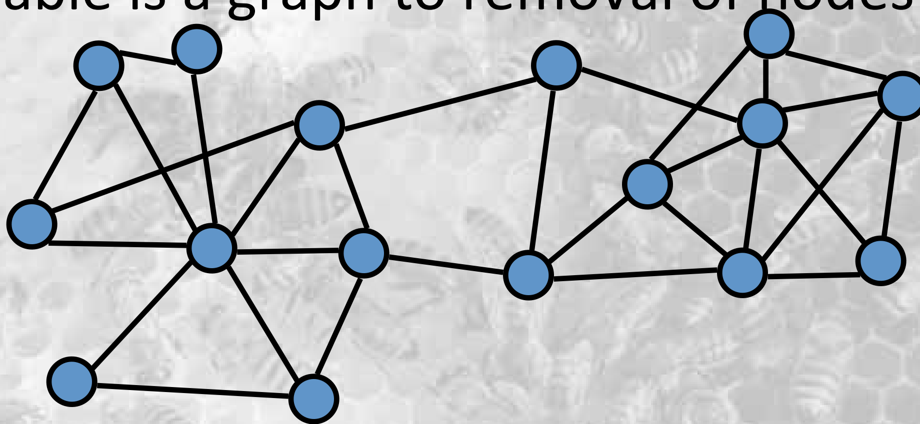
# Connected Graphs, Components, Cutpoints and Bridges

- *Cutpoints*
  - number of components in the graph that contain node $n_j$ is fewer than number of components in subgraphs that results from deleting $n_j$ from the graph
- *Cutsets (of size k)*
  - *k*-node cut
- *Bridges / line cuts*
  - Number of components… that contain line $l_k$



# Node- and Line Connectivity

- How vulnerable is a graph to removal of nodes or lines?



*Point connectivity /*
*Node connectivity*

- Minimum number of *k* for which the graph has a *k*-node cut
- For any value <k the graph is *k-node-connected*

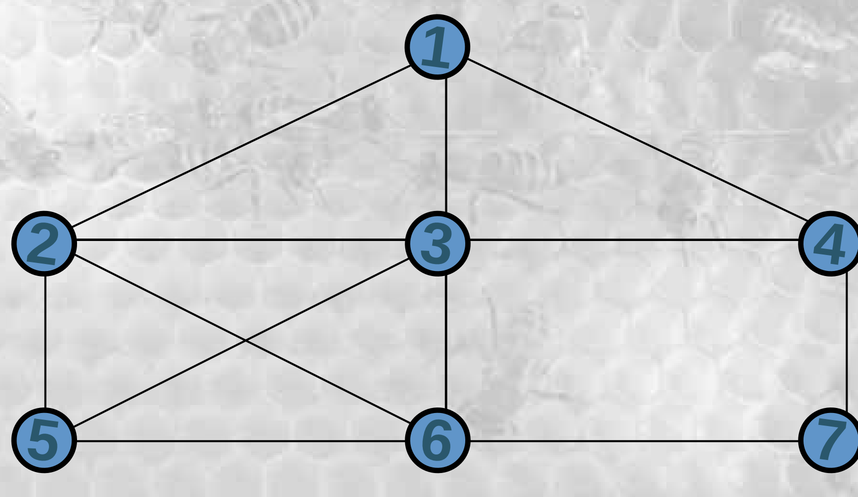
*Line connectivity / Edge connectivity*

- Minimum number λ for which for which graph has a λ-line cut

# Cohesive Subgroups

# Cohesive Subgroups, (n-)Cliques, n-Clans, n-Clubs, k-Plexes, k-Cores

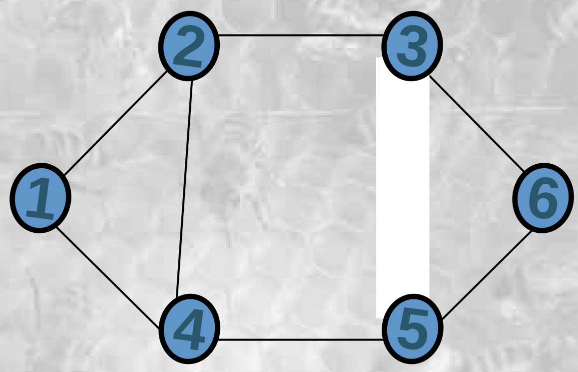
- *Cohesive Subgroup*
  - Subset of actors among there are relatively strong, direct, intense, frequent or positive ties
- *Complete Graph*
  - All nodes are adjacent
- *Clique*
  - Maximal complete subgraph of three or more nodes
  - Cliques can overlap
  - {1, 2, 3}
  - {1, 3, 4}
  - {2, 3, 5, 6}



# Cohesive Subgroups, (n-)Cliques, n-Clans, n-Clubs, k-Plexes, k-Cores

- *n-clique*
  - maximal subgraph in which d(i,j) ≤ n for all $n_i$, $n_j$
  - 2: cliques: {2, 3, 4, 5, 6} and {1, 2, 3, 4, 5}
  - intermediaries in geodesics do not have to be n-clique members themselves!

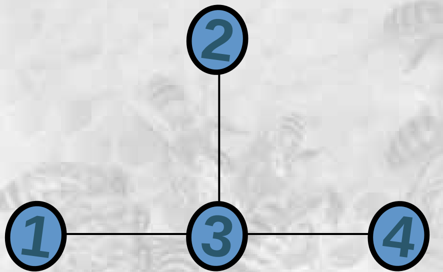
- *n-clan*
  - *n-clique* in which the d(i,j) ≤ n for the subgraph of all nodes in the n-clique
  - 2-clan: {2, 3, 4, 5, 6}

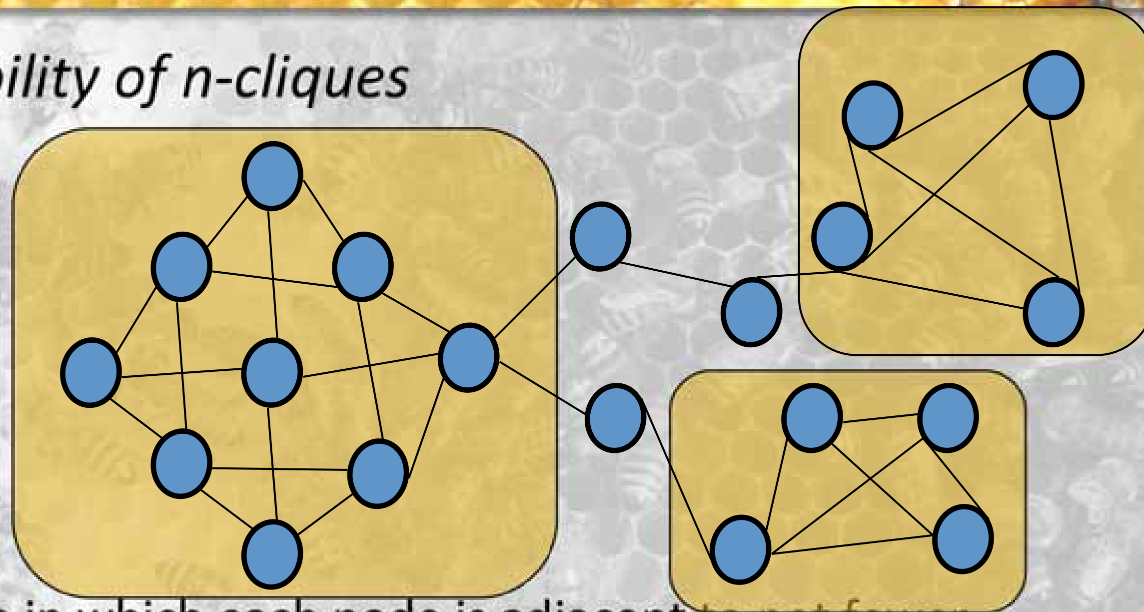



- *n-club*
  - maximal subgraph of diameter n
  - 2-clubs: {1, 2, 3, 4} ; {1, 2, 3, 5} and {2, 3, 4, 5, 6}

# Cohesive Subgroups, (n-)Cliques, n-Clans, n-Clubs, k-Plexes, k-Cores

- *Problem: vulnerability of n-cliques*



- *k-plexes*
  - maximal subgraph in which each node is adjacent to not fewer than $g_s$-k nodes („maximal": no other nodes in subgraph that also have $d_s(i) \geq (g_s\text{-}k)$ ]
- *k-cores*
  - subgraph in which each node is adjacent to at least k other nodes in the subgraph

# Analyzing Affiliation Networks

## Affiliation Matrix, Bipartite Graph and Hypergraph, Rate of Participation, Size of Events

*Two-mode network / affiliation network / membership network / hypernetwork*

- nodes can be partitioned in two subsets
  - N (for example g persons)
  - M (for example h clubs)
- depicted in *Bipartite Graph*
- lines between nodes belonging to different subsets