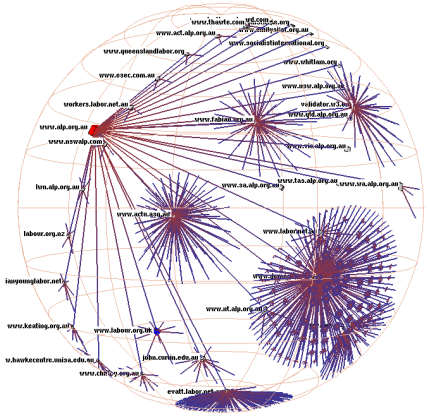
Social Network Analysis

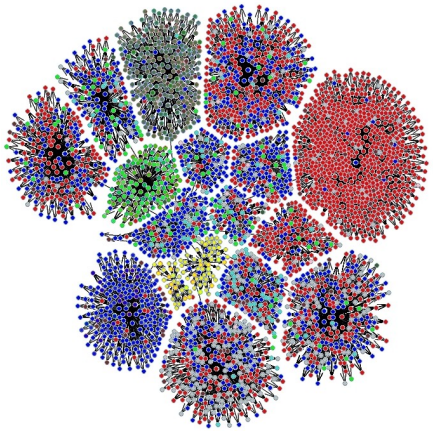


Robert Ackland

Australian Demographic and
Social Research Institute (ADSRI)
Australian National University

robert.ackland@anu.edu.au

http://voson.anu.edu.au



Notes prepared for “Big Data Analysis for Social Scientists” course
ACSPRI Winter Program, Brisbane, 29 June – 3 July 2015

Australian National University

VOSON
http://voson.anu.edu.au

Plan of lecture

- The Network Perspective
- Social network analysis – some terminology
- The Sociological Hierarchy
- Why do social ties form?
- Representing online interactions as interpersonal networks

The Network Perspective

- Wellman (1988) has identified 5 defining features of the “network perspective”
- 1. The *network perspective* puts emphasis on the *structure of relations*, rather than the *attributes of individual actors*, in determining behaviour and outcomes.
  - Example of getting a job:
    - Economic approach: human capital (experience, qualifications) is what matters
    - SNA approach: social network (in particular existence of “weak ties” [Granovetter] is what matters)

- 2. The *network perspective* focuses on relationships between actors (dyads) as the unit of analysis rather than the actors themselves
  - Ordinary Least Squares (OLS) – hallmark technique of “non-relational social science” - unit of observation is individual
  - Exponential Random Graph Models (ERGM) – unit of observation is the dyad

- 3. While OLS assumes that observations are **independent** (e.g. error term in labour market regression example assumed to be identical and independently drawn from the normal distribution), the *network perspective* explicitly assumes the **interdependence** of observations.
  - Three people: Sally, Jen, Andrew. Sally and Jen are friends and Jen and Andrew are friends. We want to model the probability of Sally and Andrew becoming friends.
  - If we assume observations (dyads) are independent, equivalent to saying that the probability of Sally and Andrew becoming friends is the same in this situation as it would be if Jen wasn't friends with either of them. Not plausible! Overlooks a basic aspect of social behaviour, *triadic closure* "a friend of my friend, is my friend too".

- 4. The n*etwork perspective* recognises that social networks can have both **direct** and **indirect** impacts on individual behaviour and outcomes.
  - The flow of information and resources between two people is not just dependent on their own relationship, but on their relationships with everyone else.
- 5. The *network perspective* recognises that people tend to belong to several overlapping social networks. That is, individuals tend to be members of several groups and the group boundaries are often fuzzy and hence hard to define.

Social network analysis – some terminology

What is a social network?

- A **network** is a set of *nodes* (or vertices or entities) and a set of *ties* (or edges or links) indicating connections or relations between the nodes.

- In a **social network**, the network nodes are people and the network ties are relations between people

How to represent a social network?

- Graphs: points represent nodes and lines represent ties
  - Can become visually complicated and difficult to see patterns.
  - Doing mathematical functions is very difficult.

- Data structures:
  - Matrix: rectangular arrangement of a set of elements, each element is either 0 = “no tie”, or >0 = “tie”
  - Edge list: each row shows one tie between nodes in column1 and column2

Networks as matrices of relations

TABLE 3.1* A Network Represented as a Matrix

	Ann	Bob	Carol
Ann	0	1	1
Bob	0	0	0
Carol	1	0	0

**This network is a directed network, as it is not symmetrical (i.e., Ann points to Bob in row 1, but Bob doesn't point to Ann in row 2). It is a simple binary network: either a tie exists (value = 1) or not (value = 0).*

TABLE 3.2* A Network Represented as an Edge List

Vertex1	Vertex2
Ann	Bob
Ann	Carol
Carol	Ann

**Individuals in the Vertex1 column "point to" those in the Vertex2 column in this directed network. The network is implied to be a binary network. Additional columns could be used to describe each edge. For example, an Edge Weight column could be added with values representing the strength of various ties.*

11

Networks as graphs

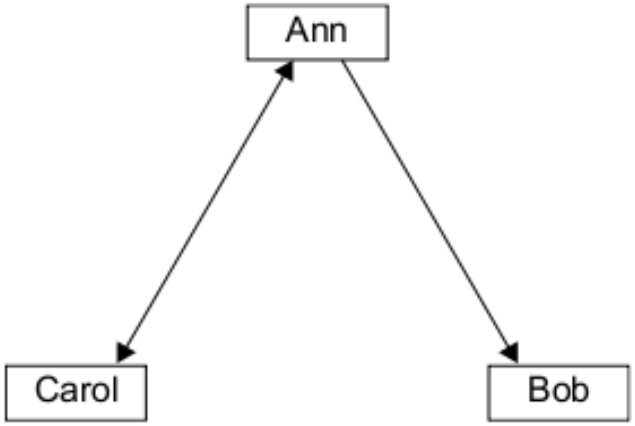
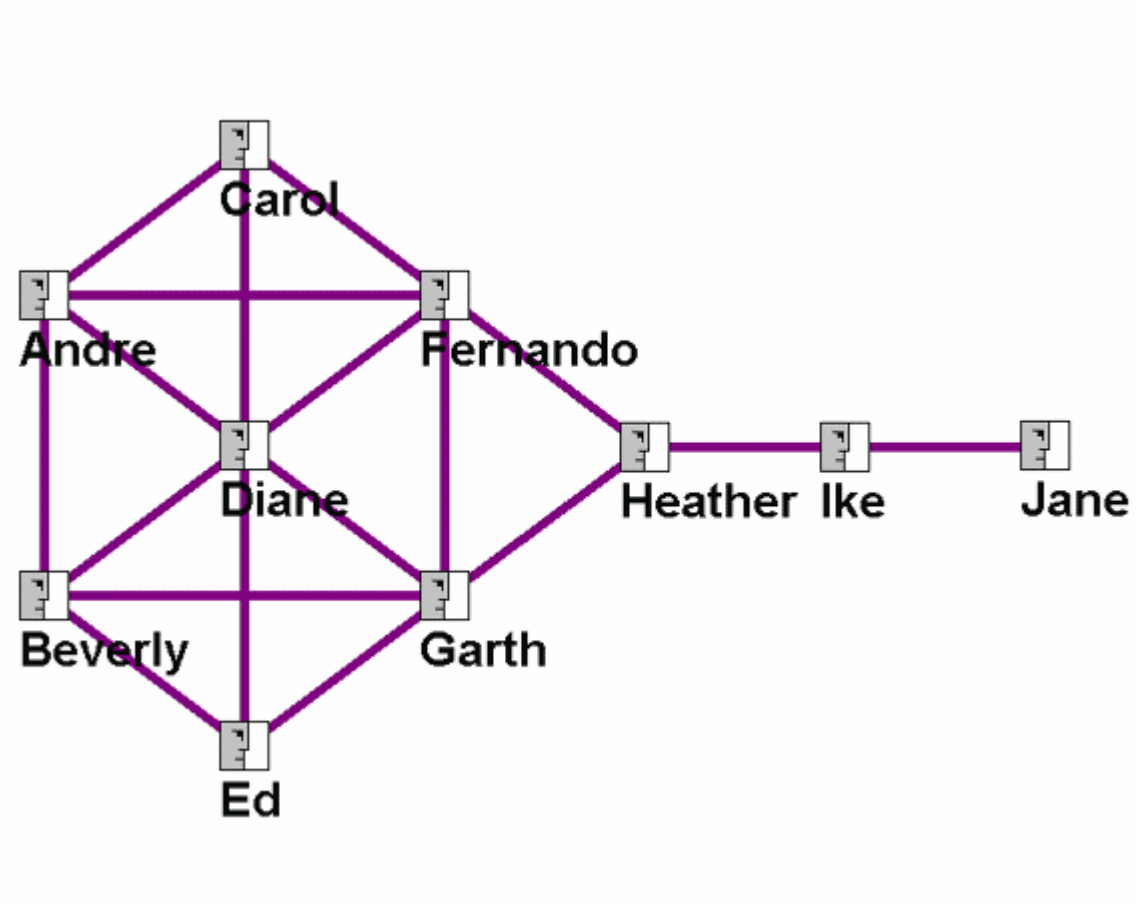


FIGURE 3.2 The directed, binary network described in Tables 3.1 and 3.2 represented as a network graph. Arrows indicate the direction of the connection (e.g., from Ann to Bob).

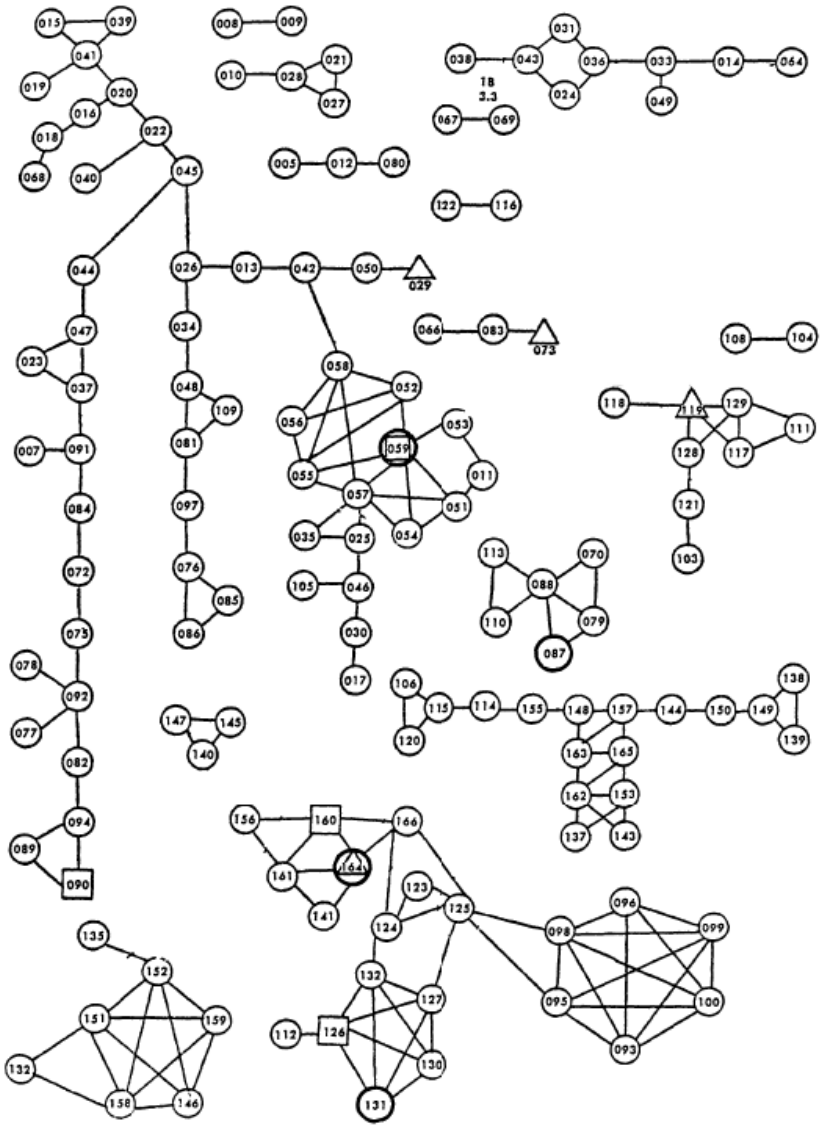
12

Andre	Beverly
Andre	Diane
Andre	Carol
Andre	Fernando
Beverly	Diane
Beverly	Ed
Beverly	Garth
Carol	Diane
Carol	Fernando
Diane	Ed
Diane	Fernando
Diane	Garth
Ed	Garth
Fernando	Garth
Fernando	Heather
Garth	Heather
Heather	Ike
Ike	Jane



"Kite Network" developed by David Krackhardt

- Reciprocated friendships among girls in Marketville (from *The Adolescent Society* by James C. Coleman, 1961)



14

- **Nodes** may have attributes:
  - *graph-theoretic attributes* e.g. number of connections (“degree”)
  - *non-graph-theoretic attributes* e.g. gender of person, race, age
- **Edges** in a network can represent different connections (collaboration, kinship, friendship, citations, transactions).

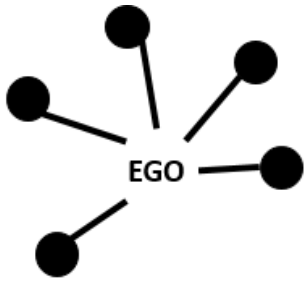
- Two major *types* of edges:
  - *directed* – clear origin and destination (e.g. Twitter user following another user, hyperlink from one page to another). May be reciprocated or not. Represented in sociogram as line with arrow head.
  - *undirected* – a mutual relationship with no origin/destination e.g. marriage, Facebook friend. Do not exist unless they are reciprocated.

- Edges can also have **weights**:
  - an **unweighted edge** is where the edge either exists or not e.g. Facebook friendship either exists or doesn't
  - a **weighted edge** has a value attached to it that indicates the strength of the relationship e.g. in a Twitter network a follower edge could be weighted by the number of re-tweets

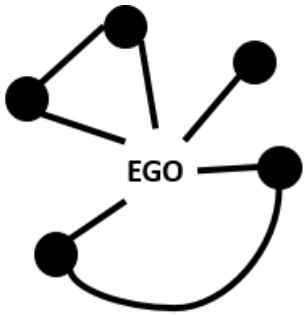
Types of networks (1)

Egocentric network (“egonet”)

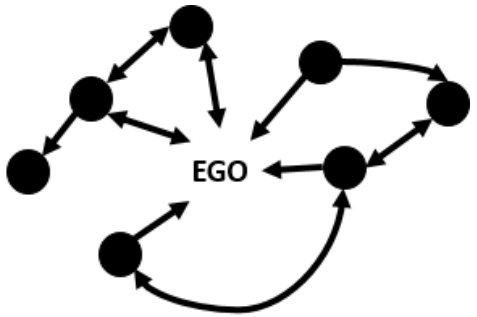
- a network consisting of the focal node/person (“ego”), the people he/she is connected to (“alters”)
  - 1 degree egonet: doesn't show connections between alters (this type is not network at all, in the matrix of relations all elements are zero except for one row)
  - 1.5 degree egonet: shows connections between alters
  - 2 degree egonet: shows friends of friends



1.0 Degree Egonet



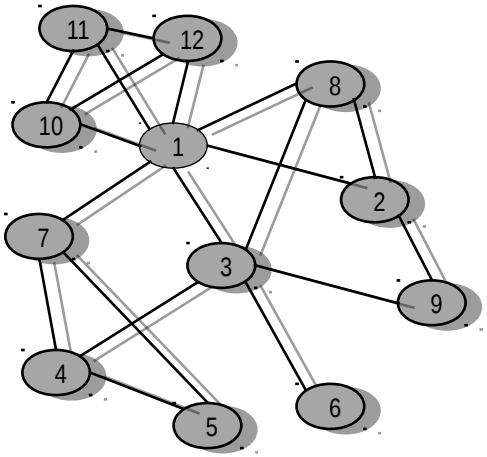
1.5 Degree Egonet



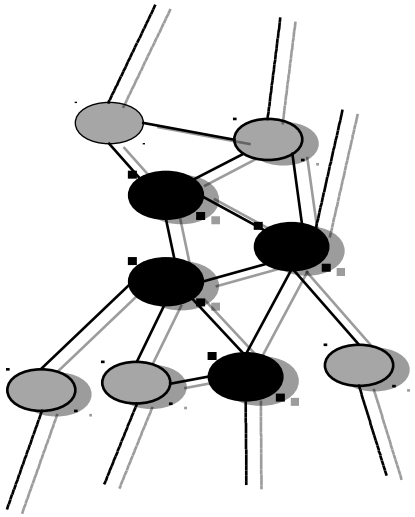
2.0 Degree Directed Egonet

Types of networks (2)

- **Complete network**
  - a network of all nodes and all ties; all egos are treated equally.
- **Partial network**
  - Partially observed network
  - Snowball sampling (ego-centric)
  - More realistic for social media networks



Complete network



Partial network

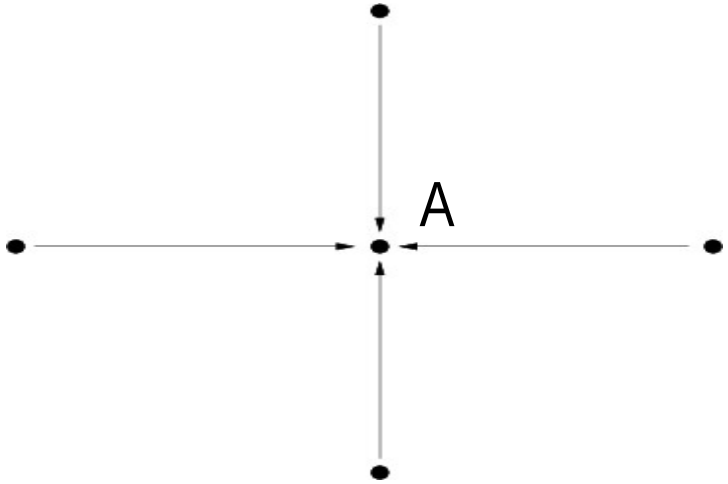
Types of networks (3)

- A *unimodal* network contains only one type of vertex. A *multimodal* network contains more than one type of node
- A *bimodal* network contains exactly two types of vertices e.g. an affiliation network consisting of people and the wiki articles they have edited
  - So in this network, people don't connect directly with people and wiki pages don't connect directly with wiki pages
  - However, a bimodal affiliation network can be transformed into 2 separate unimodal networks e.g. wiki editor-to-wiki editor and article-to-article

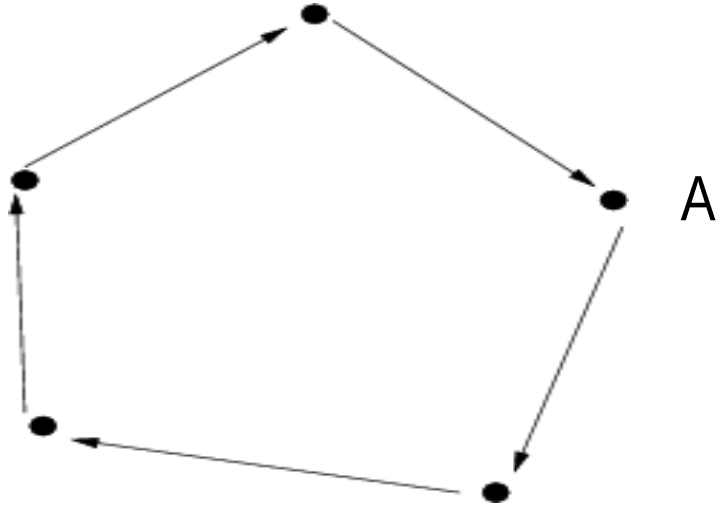
- *Multiplex networks* have multiple types of edges e.g. with Twitter we can have three types of directed edges: following relationships, “reply to” relationships, and “mention” relationships
  - Often multiplex ties are reduced to simplex tie e.g. tie exists if any of the multiplex ties exist

Social network analysis – studying power

- Power is based on relations, an individual has no power in abstract
- Structure of relations depicts how power is distributed between actors.
- Using SNA, power can be studied in the macro (network) or micro (individual) level.



Star network – perfectly **centralised** – what is the position of node “A”?



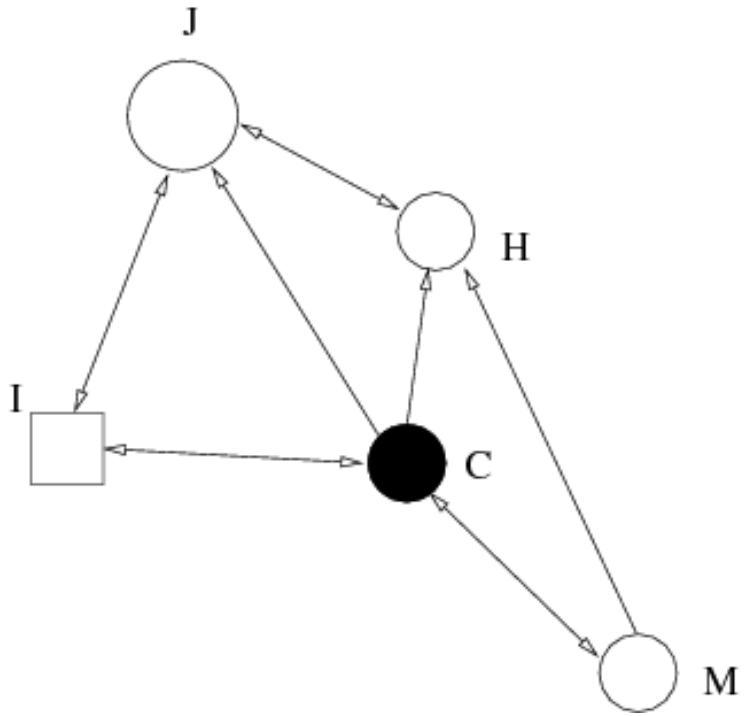
Circle network – perfectly **decentralised** – what is the position of node “A”?

Line network – what is the position of node “A”?

Network analysis metrics

- Two broad types of quantitative network metrics:
  - **network-level** (metrics describing the network as a whole)
  - **node-level** (metrics describing individual nodes within the network)

- We compute the metrics for the following small network



Network-level metrics

- **Network size** - number of nodes in the network. For our example network the network size is 5.

- **Network density** - number of network ties as proportion of the maximum possible number of network ties. The maximum number of ties in a directed network of size n is n(n - 1), while for an undirected network the maximum number of ties is n(n - 1)/2. So for our example network (which is directed), the maximum number of ties is 20, and the network density is 11/20 = 0.55.

- **Network inclusiveness** - number of non-isolates as a proportion of the network size. For our example network there are no isolates and so network inclusiveness is 1.

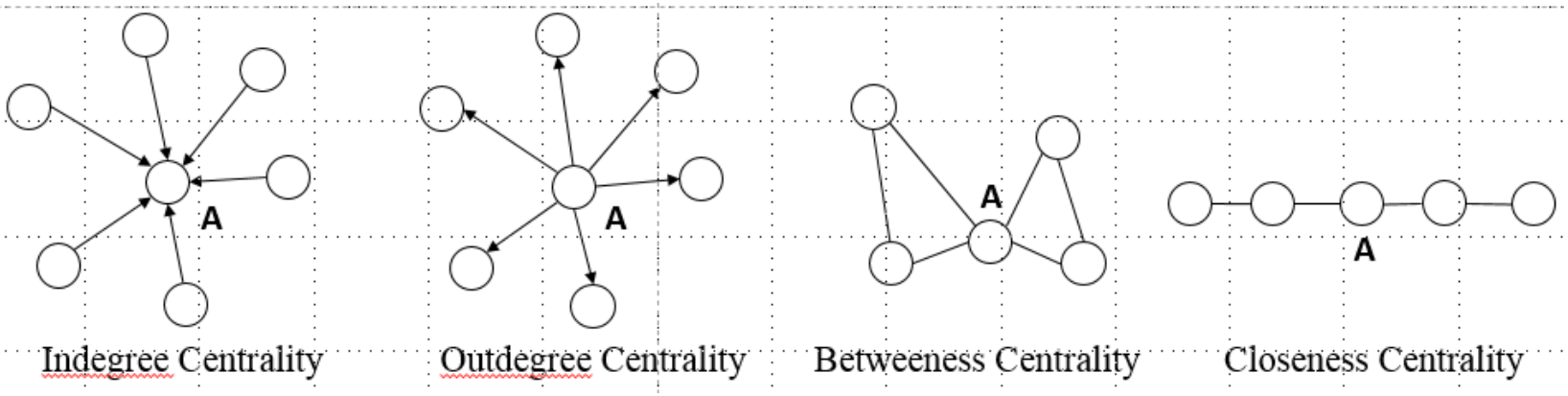
- **Centralisation** - a network-level property that is calculated for a given node-level property (betweenness, closeness, eigenvector) and it broadly measures the distribution of importance, power or prominence among actors in a given network i.e. the extent to which the network "revolves around" a single node or small number of nodes
  - Centralisation indicates: how central its most central node is in relation to how central all the other nodes are.

Node-level metrics

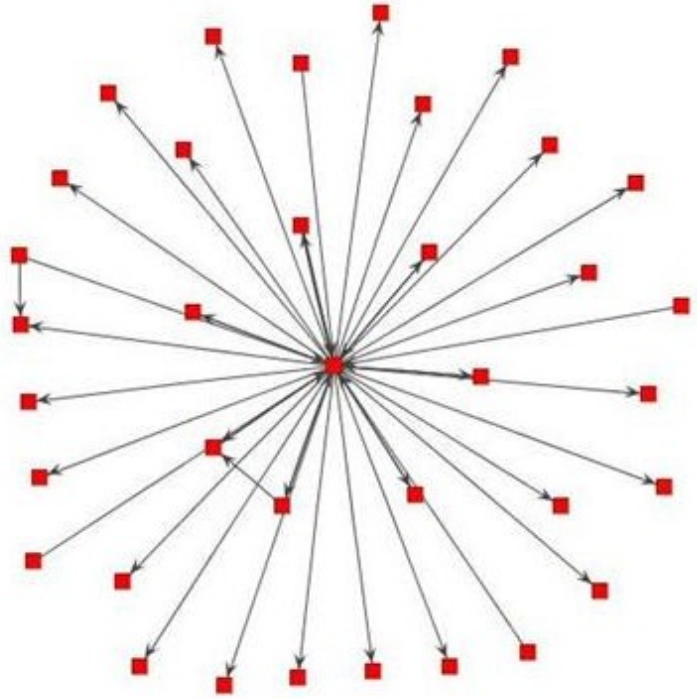
- **Degree (Prestige or Influence)** - number of ties (only defined for undirected network)
  - **Indegree -** number of inbound ties (only defined for directed network). For our example network, the indegree for node c is 1, while the indegree for node i is 2.
  - **Outdegree** - number of outbound ties (only defined for directed network). For our example network, the outdegree for node c is 1, while the outdegree for node i is 2.

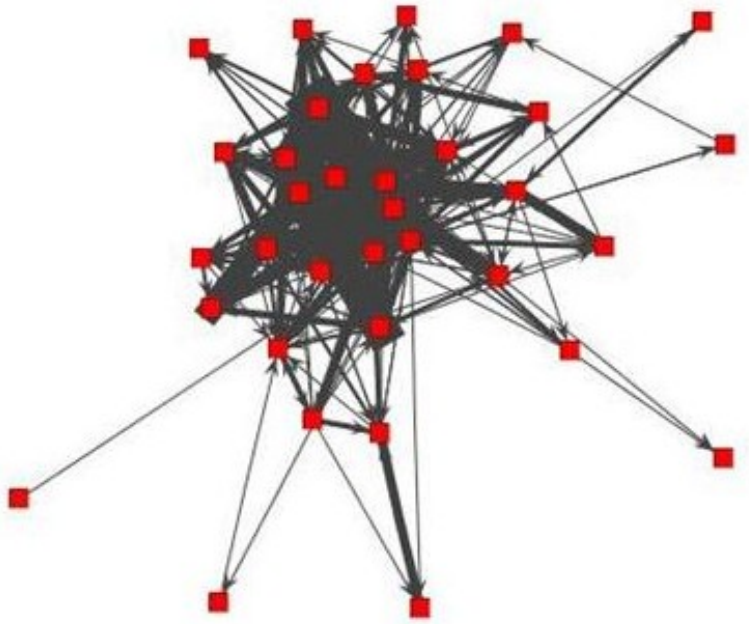
- **Betweenness centrality -** indicator of the extent to which an individual node plays a "brokering" or "bridging" role in a network and is calculated for a given node by summing up the proportion of all minimum paths within the network that "pass through" the node.

- **Closeness centrality** - indicator of the extent to which a given node has short paths to all other nodes in the graph and it is thus a reasonable measure of the extent to which the node is in the "middle" of a given network.
- **Eigenvector centrality** – in computing a node's centrality, this takes account of the centrality of node's alters i.e. you might have high eigenvector centrality even if you have few connections but the people you care connected are *themselves* well connected. Google's PageRank algorithm is a variant of eigenvector centrality.



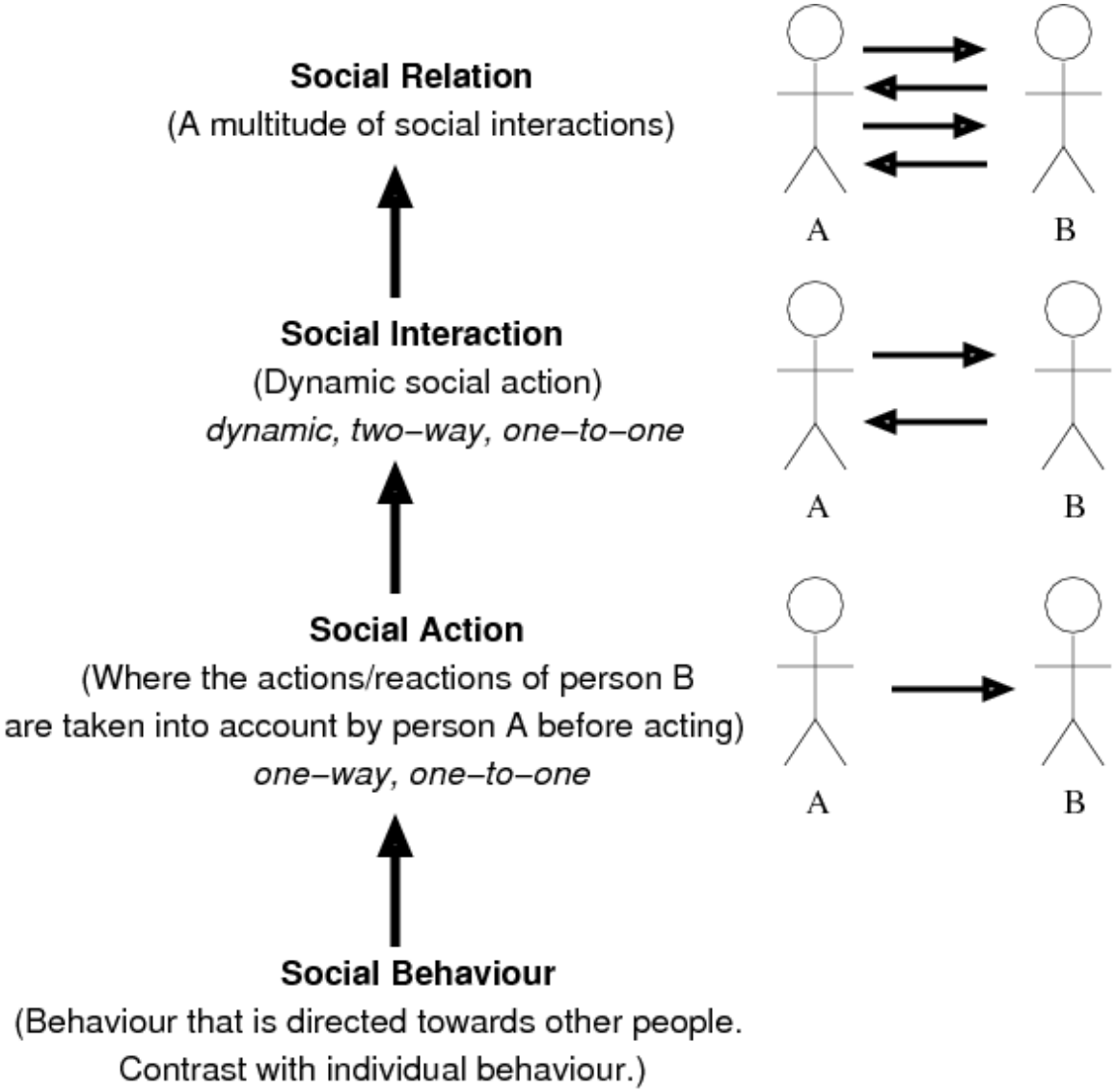
- **Clustering coefficient** – indicator of extent to which ego's friends are friends with one another i.e. it is a measure of the density of the 1.5 egonetwork. If your friends (alters) know each other then you have high clustering coefficient but if they don't know each other you have low clustering coefficient
- Egonet on LHS has low clustering coefficient while egonet on RHS has high clustering coefficient (from Welser et al. (2007))





The Sociological Hierarchy

The "Sociological Hierarchy"



See: http://en.wikipedia.org/wiki/Social_relation

- Consider each of the seven examples of computer-mediated interaction above: newsgroup, wiki, folksonomy, blog, SNS, Virtual World, micro-blog (Twitter)
- The Sociological Hierarchy may be used to assess whether each example comprises a social network
- ***When used in its typical mode***, does each application enable social behaviour?
  - If so, then does it enable social action?
  - If so, then does it enable social interaction?
  - If so, then does it enable social relations?
- Let's consider three of the examples: wiki, SNS, micro-blog

Social network service (e.g. Facebook)

- Enables social behaviour?
  - Yes
- Enables social action?
  - Yes, interactions on facebook can be both one-to-one and one-to-many
- Enables social interaction?
  - Yes, interactions on facebook can be two-way
- Enables social relations?
  - Yes, Facebook can enable a multitude of social interactions
- Conclusion: SNSs such as Facebook are examples of online social networks

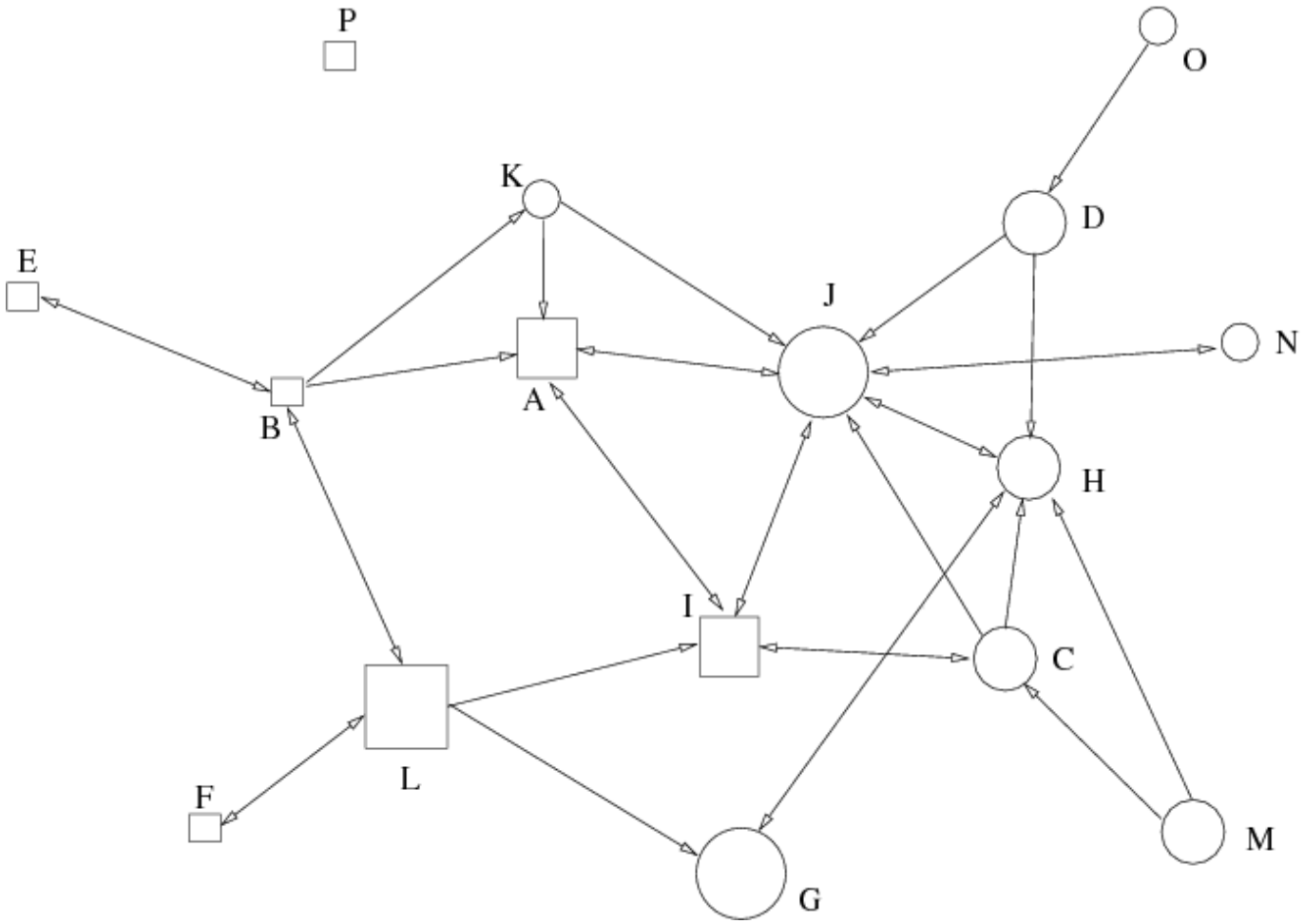
Micro-blogging (e.g. Twitter)

- *When used in its typical mode*, does Twitter enable social behaviour?
  - Yes
- What about social action?
  - No, if social action has to be one-to-one
- Conclusion: micro-blogging (e.g. Twitter) is not an example of an online social network
- Twitter themselves refer to the service as a “communication network” not a social network

Why do social ties form?

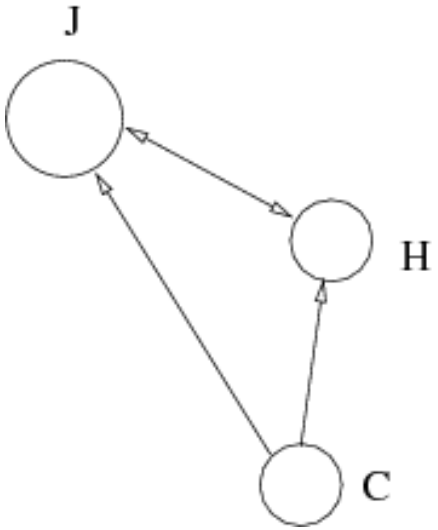
Controlling for endogenous network effects

- School friendship network (Fig 3.1 Web Social Science)
- square = boy circle = girl
- larger nodes are older



- Two types of features in social networks:
  - *Purely structural network effects*: network ties that have nothing to do with actor attributes. More to do with social norms, for example:
    - reciprocity (if A links to B, then more likely that B will link to A) – one generally reciprocates when someone extends the hand of friendship
    - transitivity (if A links to B, and B links to C, then for this to be a "transitive triad" then A must also link to C) - "a friend of my friend is also my friend"
  - *Actor-relation effects*: network ties that are created because of the characteristics or attributes of actors

Transitive triad extracted from the school friendship network



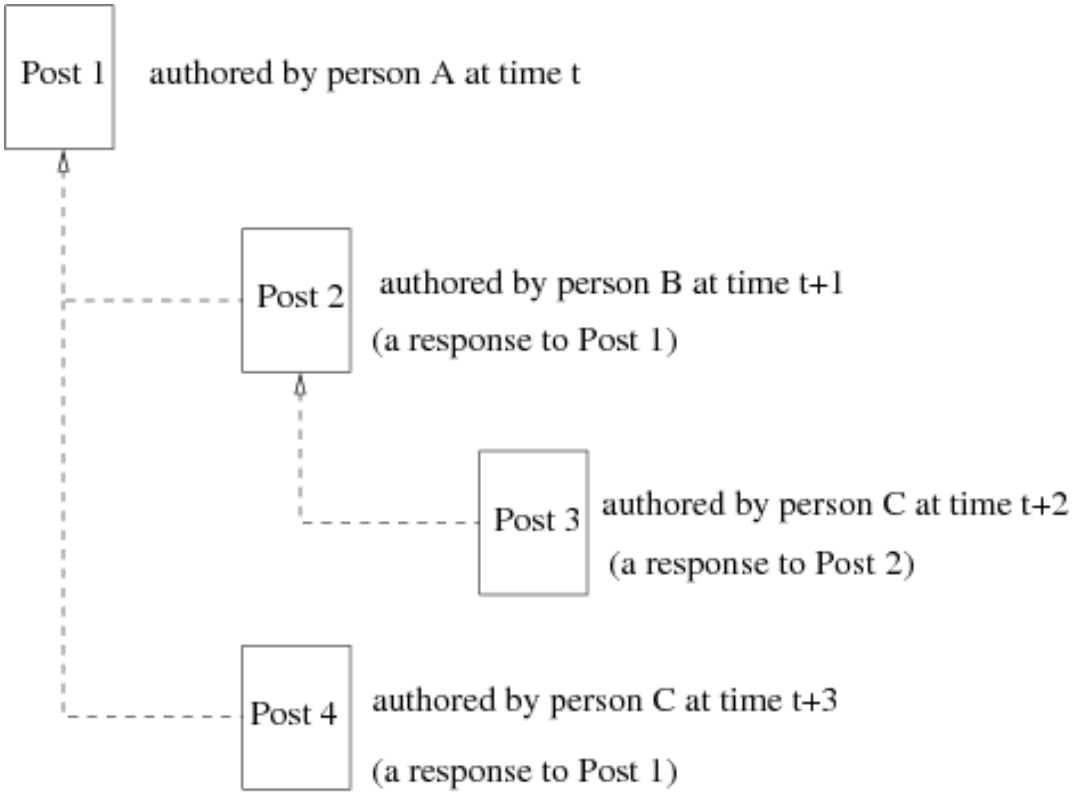
- Why has this transitive triad formed? In particular, why has C nominated J as a friend?
  - tie from C to J could be actor-relation effect (J is older “receiver effect”, J and C are both female “homophily”)
  - but tie from C to J could also be a purely structural effect: fact that *C* nominates H and H nominates J, means there is a higher likelihood that C will also nominate J (thus forming a transitive triad)
- Difficult to discern actor-relation effects from purely structural effects if only observe social network at single point in time
  - don't know if C was friends with J before H was friends with J
- Exponential Random Graph Models (ERGM) – method for statistically “unpacking” social networks

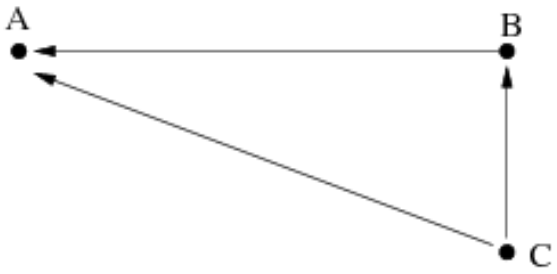
- There are many types of purely structural effects (we only focus here on reciprocity and transitivity)
- Actor-relation effects:
  - **Sender effects** – impact of presence or absence of particular actor attribute on propensity to send ties (significant and positive sender effect indicates actors with attribute send more ties *than expected by chance*)
  - **Receiver effects** – analogous to sender effects but refer to propensity of *receiving* ties
  - **Homophily effects** – positive and significant parameter indicates actors with attribute more likely than chance to send ties to actors sharing attribute

Representing online interactions as interpersonal networks

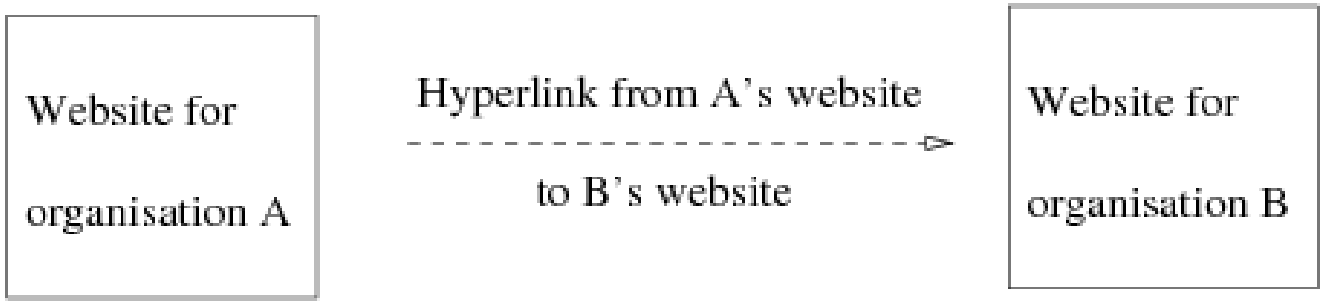
- Threaded conversations
- Web 1.0 websites
- Blogs
- Wikis
- Social network sites
- Microblogs
- Virtual worlds

Threaded conversations

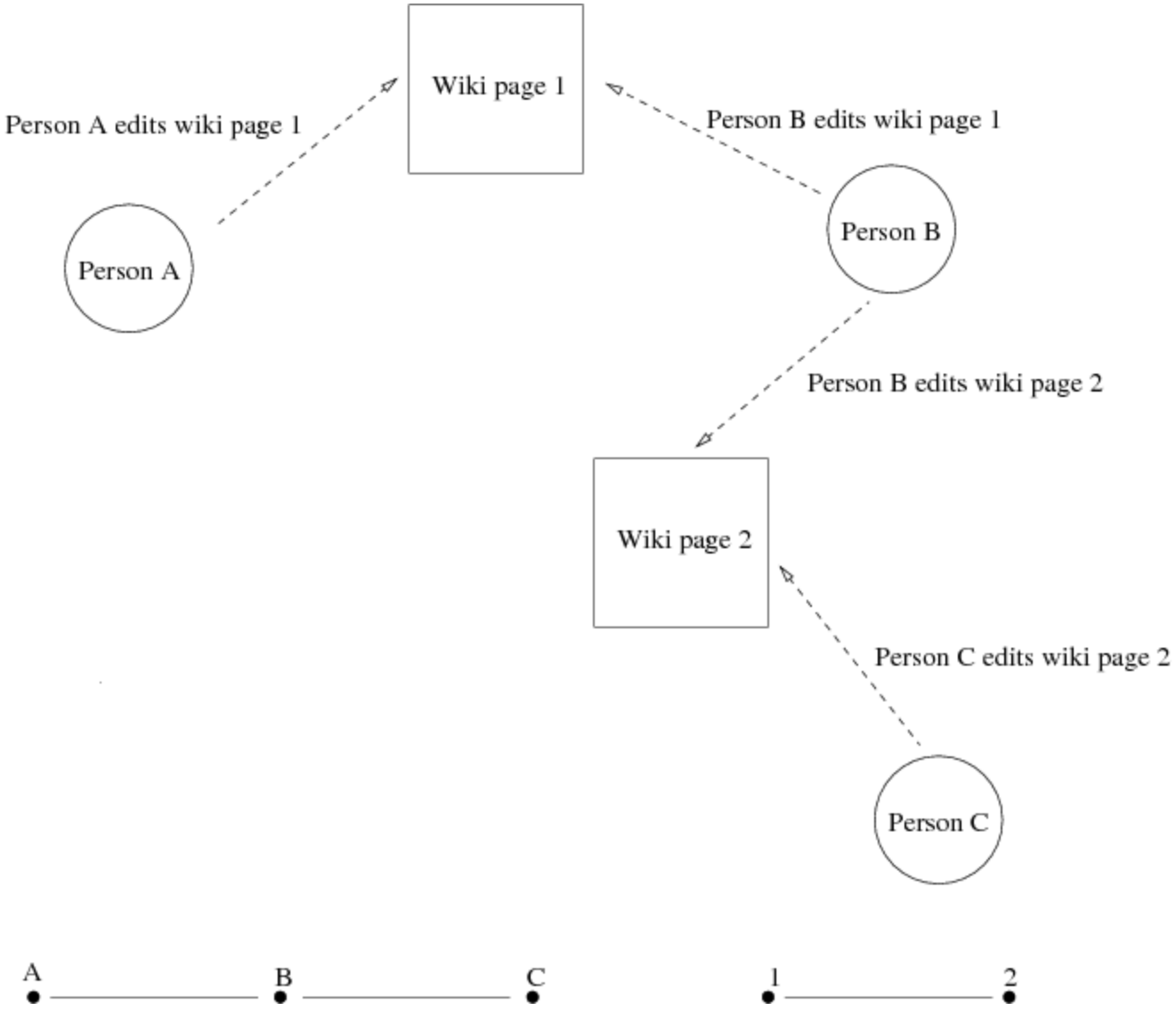




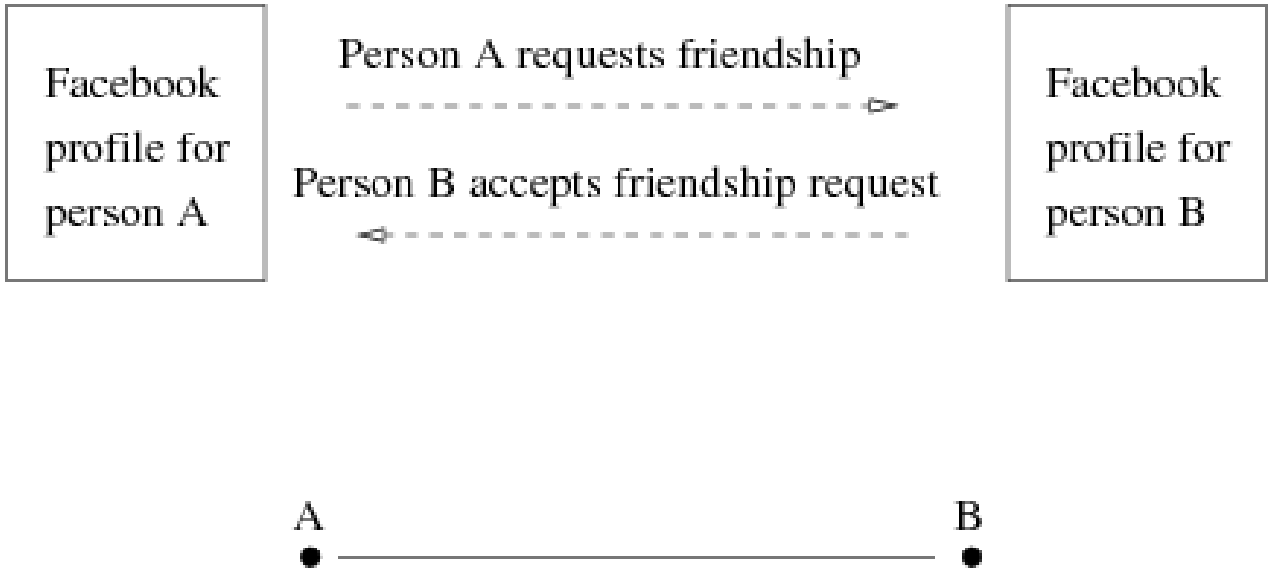
Web 1.0 websites and blogs



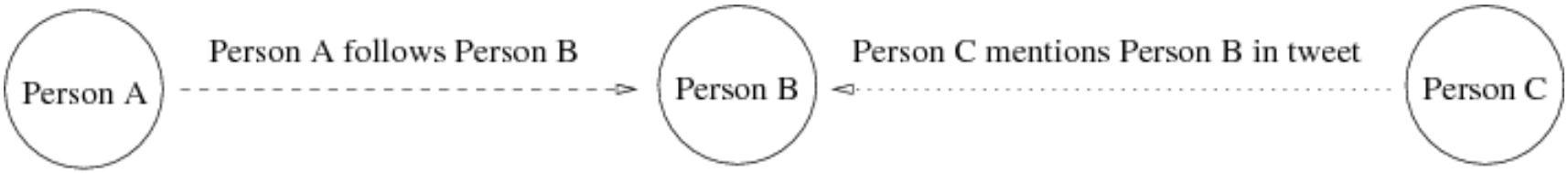
Wikis

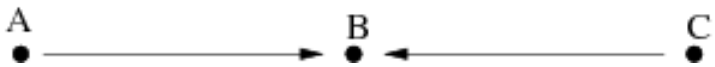

Wiki page 1
Person A edits wiki page 1
Person B edits wiki page 1
Person A
Person B
Person B edits wiki page 2
Wiki page 2
Person C edits wiki page 2
Person C
A
B
C
1
2

Social network site



Microblogs


Person A
Person A follows Person B
Person B
Person C mentions Person B in tweet
Person C


A
B
C

Virtual worlds

- Three people are playing World of Warcraft. Person A and person B both join raiding party 1. Person B also joins raiding party 2, which person C also belongs to.
- This can be represented as an affiliation network with exactly the same structure as the wiki network above (and the two unimodal networks for people and raiding parties).