Social Network Analysis

Master Sociology 2014-15

University of Amsterdam

Programme for today

Session 1: Monday, Jan 12, REC-H 2.01 (computer lab)

- Methods 1: Introduction to graph theory
- Methods 2: Cohesion

Session 2: Wednesday, Jan 14, REC-H 2.01 (computer lab)

- Methods 3: Cohesion / Centrality
- Methods 4: Centrality

Session 3: Thursday, Jan 15, REC-G S.09

- Theory 2: QAP correlations and regression

Session 4: Friday, Jan 16, 13-15, REC-G S.14

- Test 2

UCInet basics

- We learn the program by doing SNA;
  mechanics of the program are not the focus of this class
- Quick start guide to UCInet
  - https://sites.google.com/site/ucinetsoftware/document
- Working directory
  C:/Program Files (x86)/Analytic Technologies/Datafiles
- In this lab, the files we produce will be stored in C:/Temp

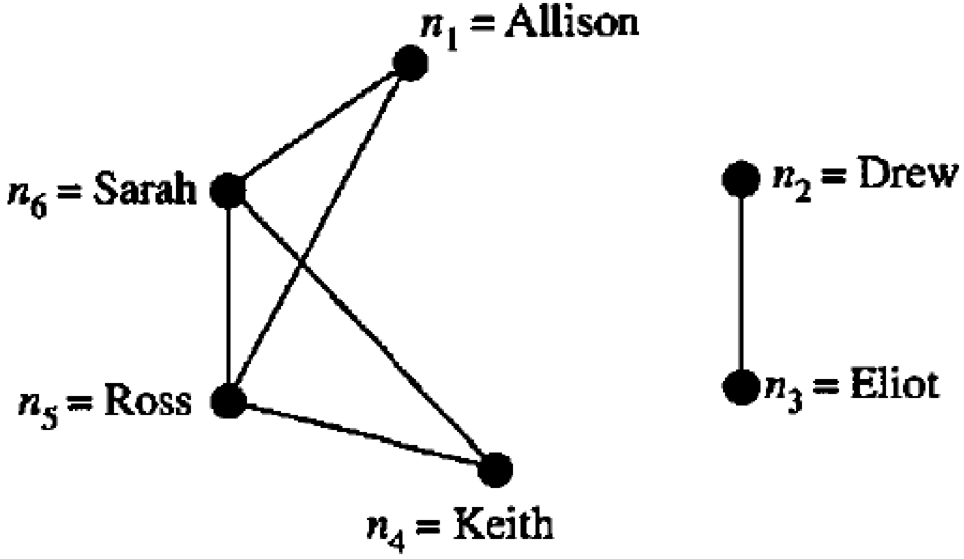
UCInet basics

- UCInet datasets
  - Each dataset consists of two files
    - Extension ##d are the actual data
    - Extension ##h contain information about the data
  - Many datasets are already stored in the folder Datafiles
- Import data
  - UCInet spreadsheet (manually or cut and paste)
  - From Excel
  - Data load format (DL)
  - See tutorials on the web

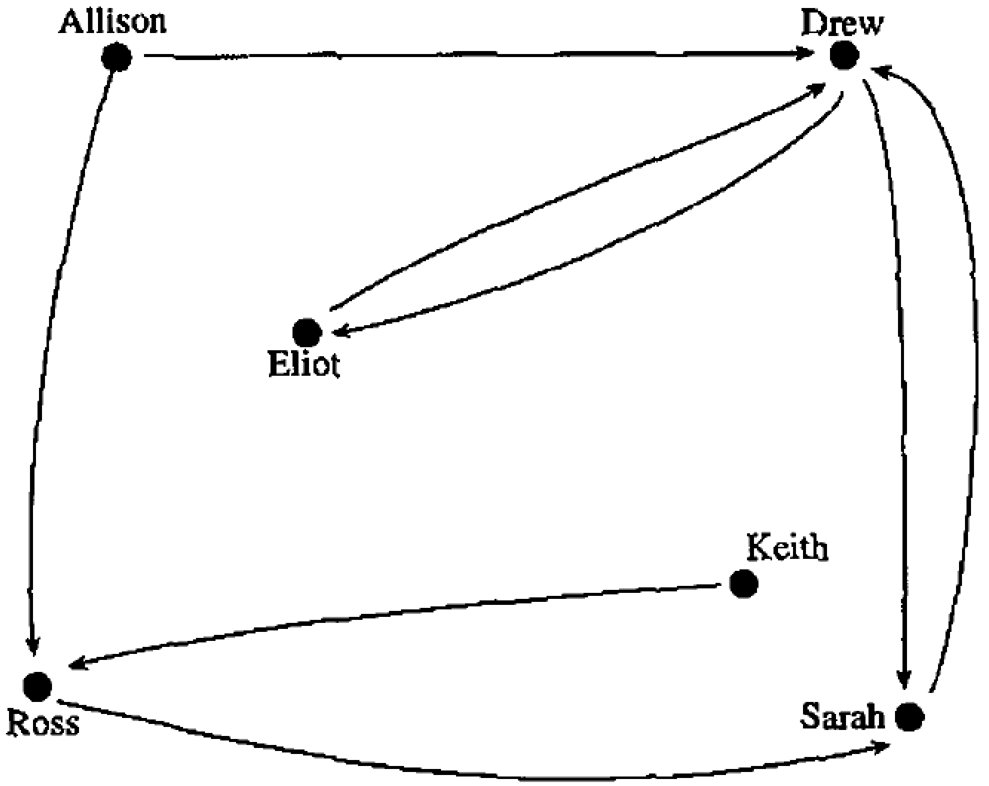
Basic Concepts of Graph Theory

Graph

- Graph G(V,E) consists of
  - Vertices V representing actors
  - Edges E representing ties
- Edge
  - Unordered pair of nodes (u,v)
  - Nodes u and v are adjacent if (u,v) ϵ E
  - E is a subset of all pairs of nodes (those which are tied)



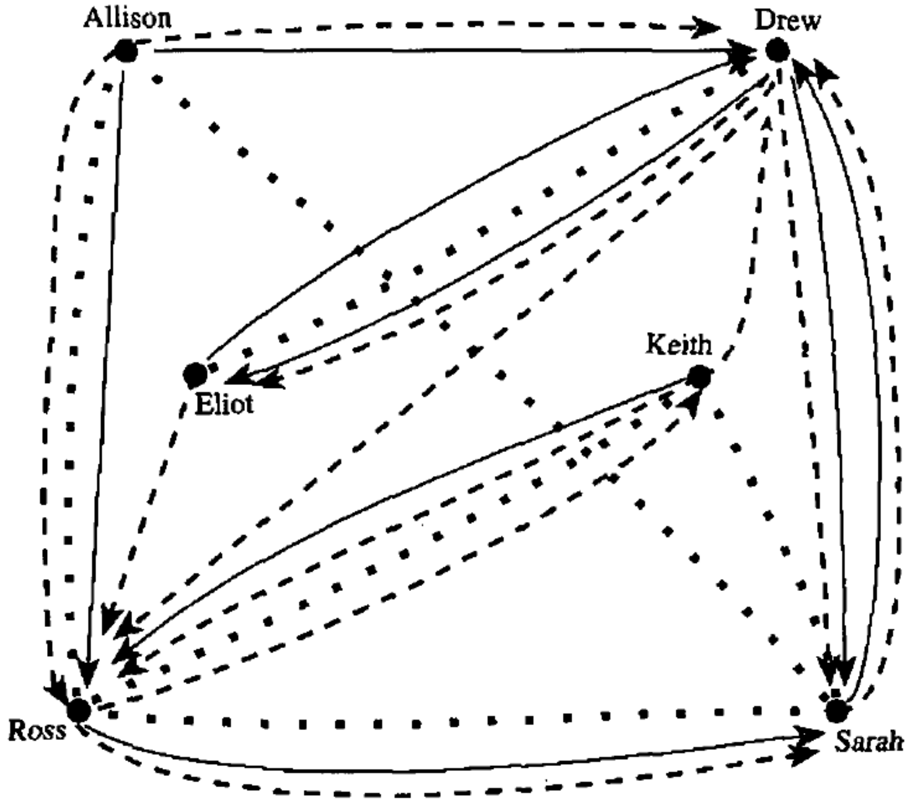
Directed graph



- Digraph D(V,E) consists of
  - Vertices V representing actors
  - Arcs E representing ties
- Arc
  - Ordered pair of nodes (u,v)
  - (u,v) ϵ E means that u sends arc to v
  - (u,v) ϵ E does not imply that (v,u) ϵ E

Multiple relations

Relation 1 Friendship at Beginning	Relation 2 Friendship at End	Relation 3 Lives Near
<Allison, Drew>	<Allison, Drew>	(Allison, Ross)
<Allison, Ross>	<Allison, Ross>	(Allison, Sarah)
<Drew, Sarah>	<Drew, Sarah>	(Drew, Eliot)
<Drew, Eliot>	<Drew, Eliot>	(Keith, Ross)
<Eliot, Drew>	<Drew, Ross>	(Keith, Sarah)
<Keith, Ross>	<Eliot, Ross>	(Ross, Sarah)
<Ross, Sarah>	<Keith, Drew>	
<Sarah, Drew>	<Keith, Ross>	
	<Ross, Keith>	
	<Ross, Sarah>	
	<Sarah, Drew>	



Sociometric notation

Friendship at Beginning of Year

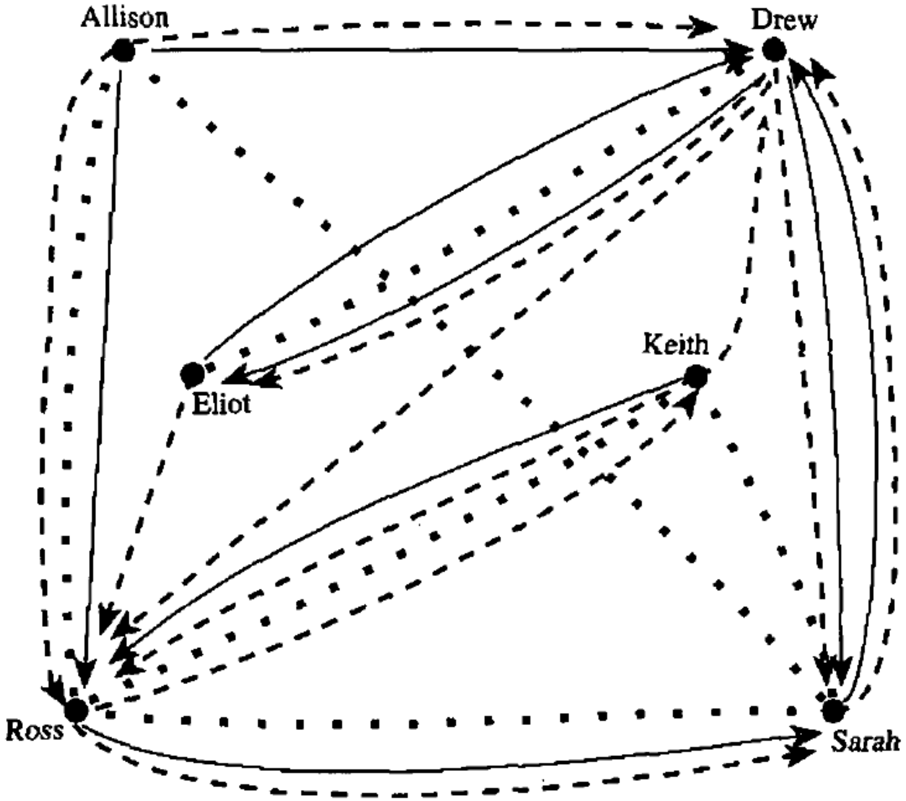
	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	0	1
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	0
Ross	0	0	0	0	-	1
Sarah	0	1	0	0	0	-

Friendship at End of Year

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	1	1
Eliot	0	0	-	0	1	0
Keith	0	1	0	-	1	0
Ross	0	0	0	1	-	1
Sarah	0	1	0	0	0	-

Lives Near

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	0	0	0	1	1
Drew	0	-	1	0	0	0
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	1
Ross	1	0	0	1	-	1
Sarah	1	0	0	1	1	-



Valued ties

- We can assign values to ties
  - Strength of affective relationship
  - Frequency of contact
  - Amount of money
  - Probability of passing on information
- These are called valued graphs

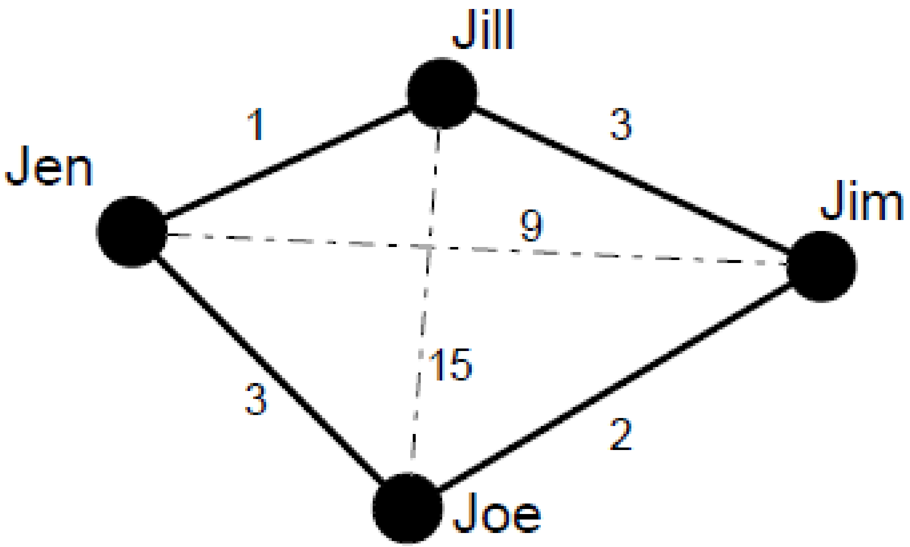
Adjacency matrices for binary and valued ties

Friendship

	Jim	Jill	Jen	Joe
Jim	-	1	0	1
Jill	1	-	1	0
Jen	0	1	-	1
Joe	1	0	1	-

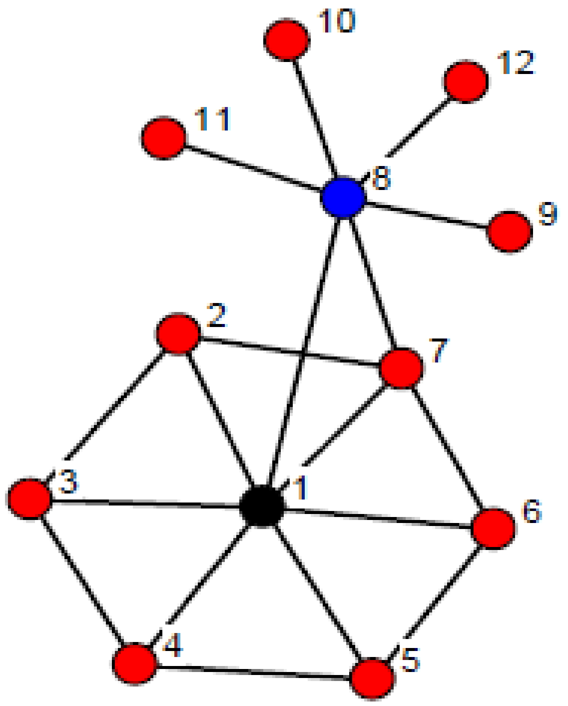
Proximity

	Jim	Jill	Jen	Joe
Jim	-	3	9	2
Jill	3	-	1	15
Jen	9	1	-	3
Joe	2	15	3	-



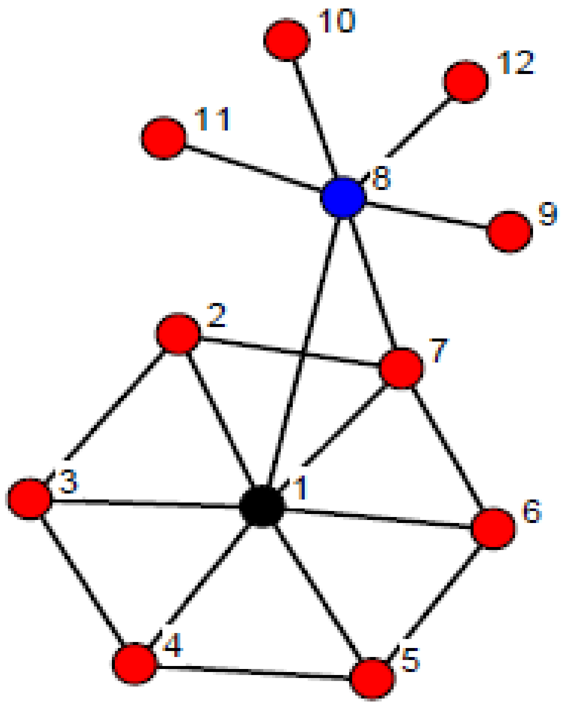
Degree

- The number of edges indicent upon a node



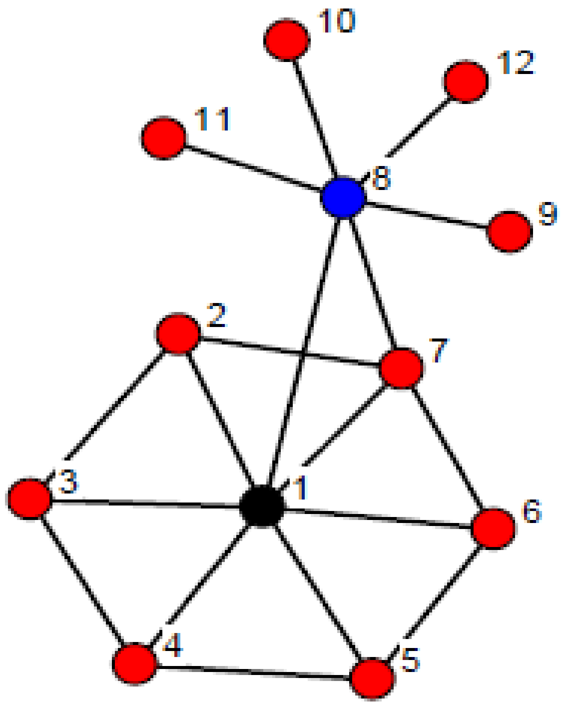
Walks, trails, and paths

- Walk: Unrestricted
  - The most general kind of sequence
  - 1-2-3-1-2-7-1-2
- Trail: Cannot repeat edge
  - Walk in which all edges are distinct
  - 1-2-3-1-7-8-10
- Path: Cannot repeat node
  - Walk in which all nodes and edges are distinct
  - 1-2-3-4-5-6-7-8



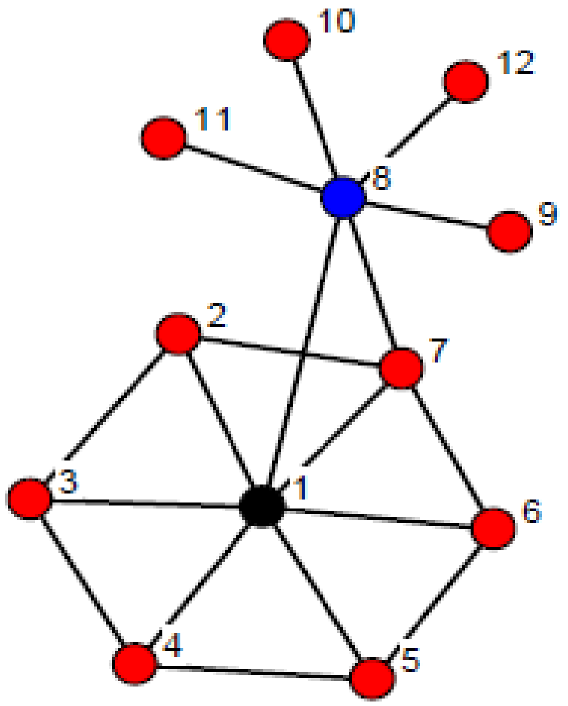
Walks, trails, and paths

- Walk: Unrestricted
  - The most general kind of sequence
  - 1-2-3-1-2-7-1-2
  - Example: ?
- Trail: Cannot repeat edge
  - Walk in which all edges are distinct
  - 1-2-3-1-7-8-10
  - Example: ?
- Path: Cannot repeat node
  - Walk in which all nodes and edges are distinct
  - 1-2-3-4-5-6-7-8
  - Example: ?

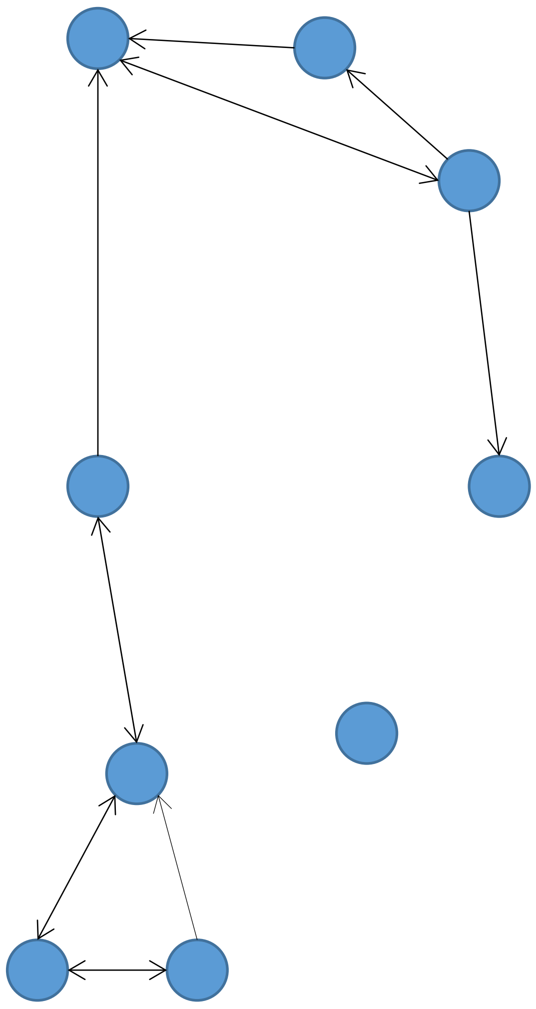


Walks, trails, and paths

- Walk: Unrestricted
  - The most general kind of sequence
  - 1-2-3-1-2-7-1-2
  - Example: Dollar bill
- Trail: Cannot repeat edge
  - Walk in which all edges are distinct
  - 1-2-3-1-7-8-10
  - Example: Gossip
- Path: Cannot repeat node
  - Walk in which all nodes and edges are distinct
  - 1-2-3-4-5-6-7-8
  - Example: Virus

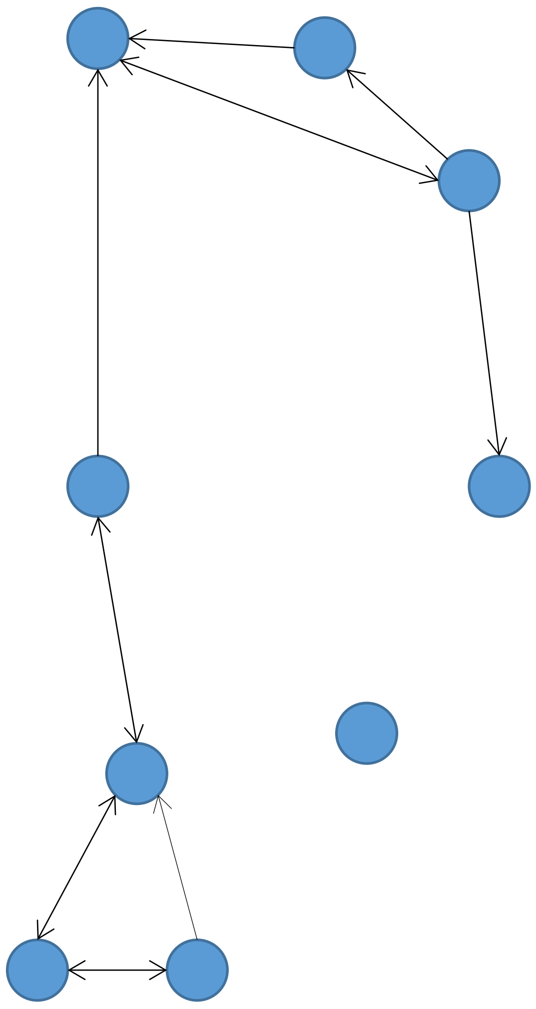


Indegree and Outdegree



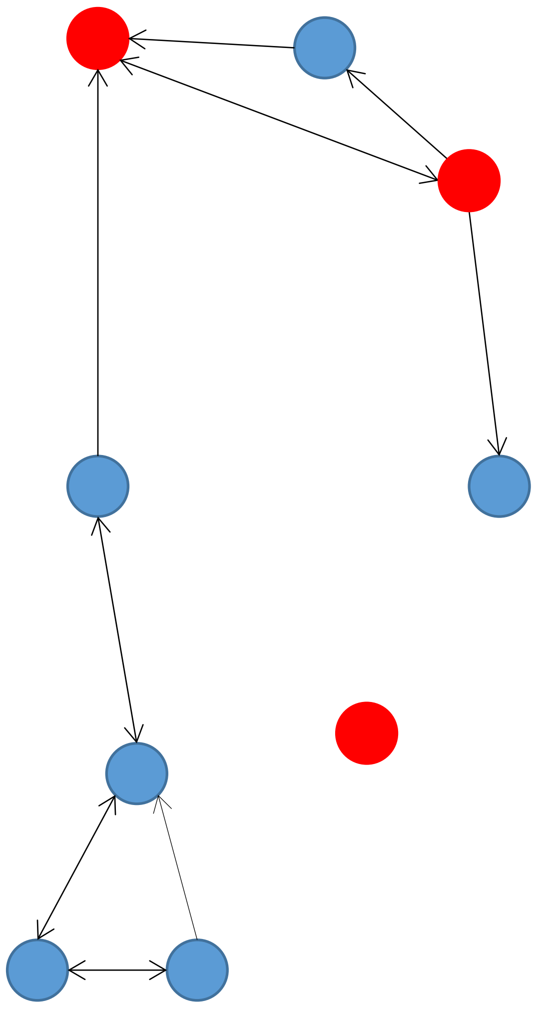
Indegree and Outdegree

- This applies to directed graphs only
- Indegree: Arcs that terminate at node
- Outdegree: Arcs that originate at node
- Average Indegree = Average Outdegree



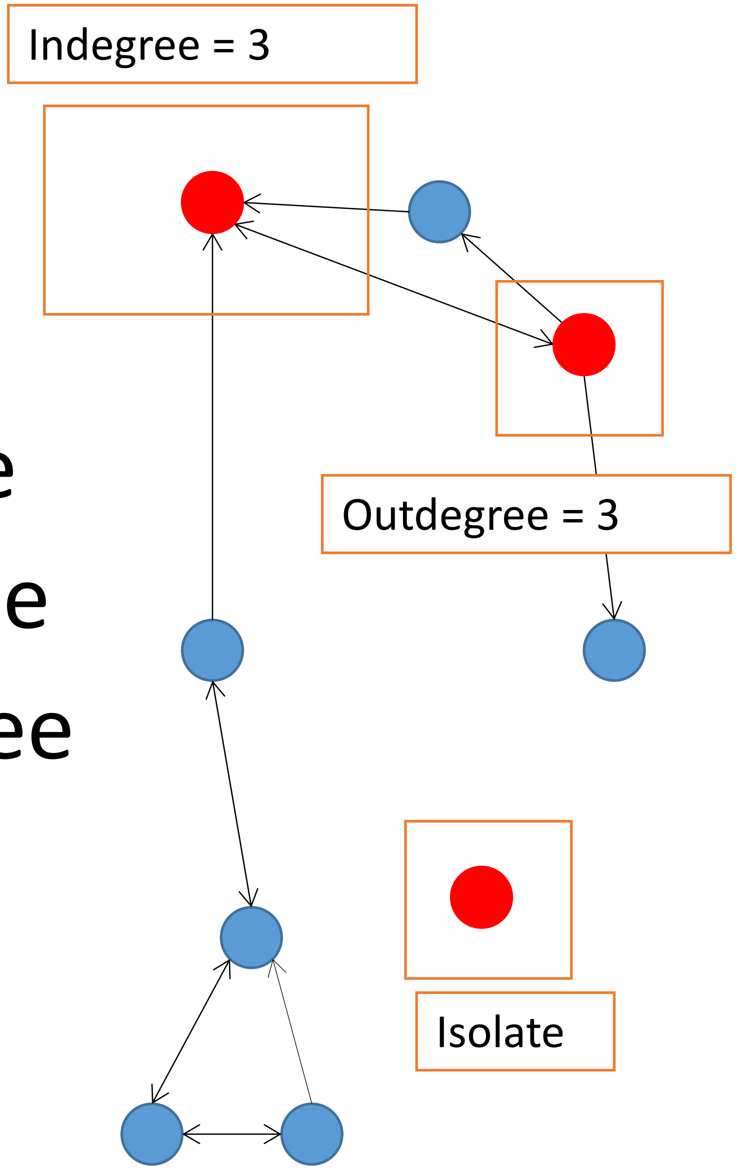
Indegree and Outdegree

- This applies to directed graphs only
- Indegree: Arcs that terminate at node
- Outdegree: Arcs that originate at node
- Average Indegree = Average Outdegree



Indegree and Outdegree

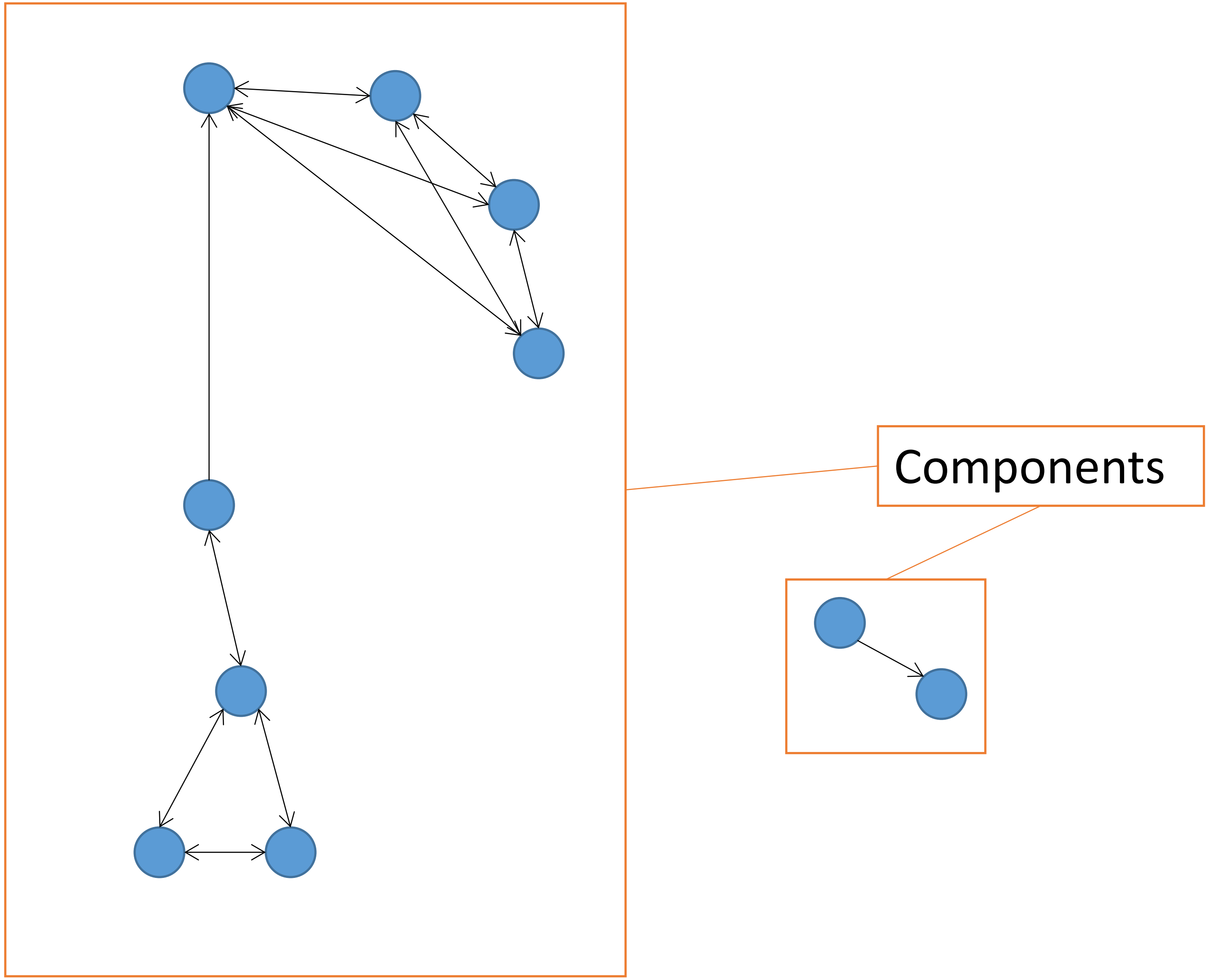
- This applies to directed graphs only
- Indegree: Arcs that terminate at node
- Outdegree: Arcs that originate at node
- Average Indegree = Average Outdegree



Component

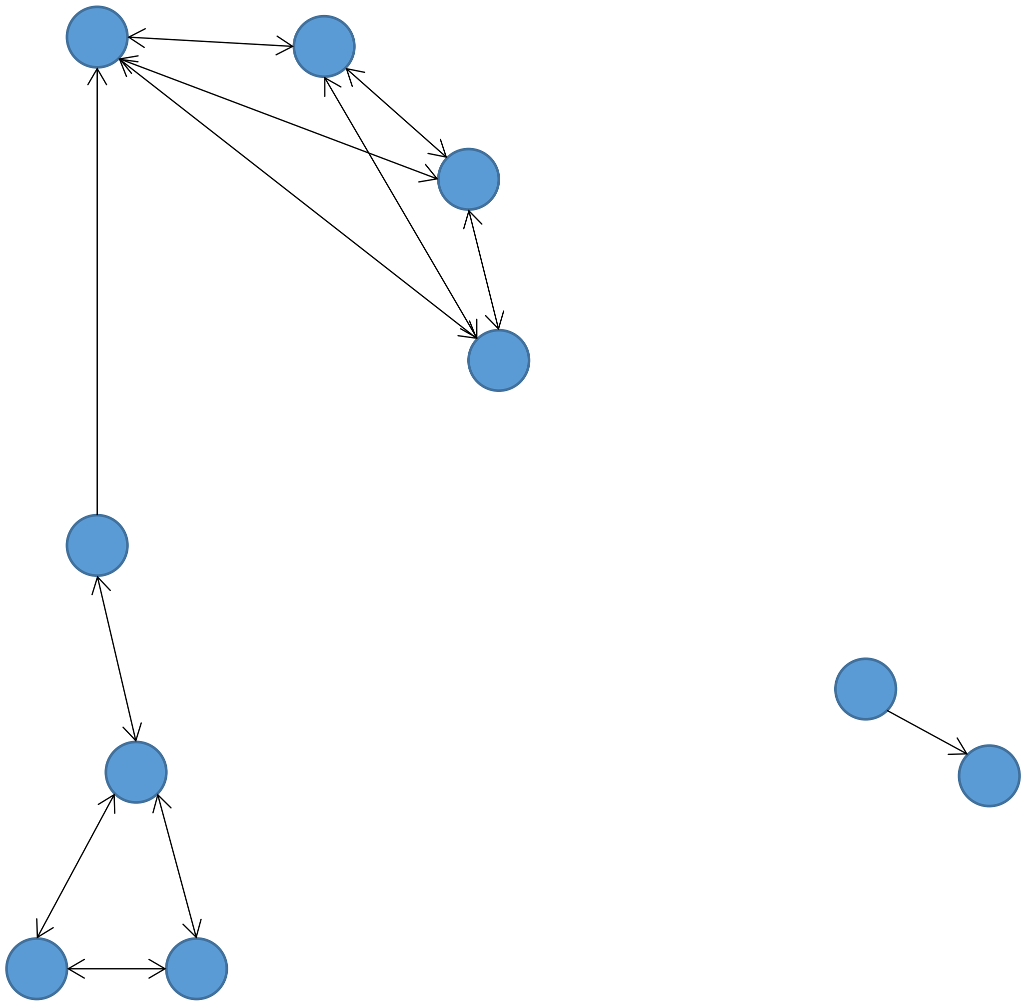
Component

- Maximal set of nodes in which every node can reach every other node (no matter how long the path)
- If something flows through a certain type of line, it cannot flow between components
- Connected graph: One component
- Weak component: Every node reaches the other, ignoring direction
- Beware: Components do not define networks



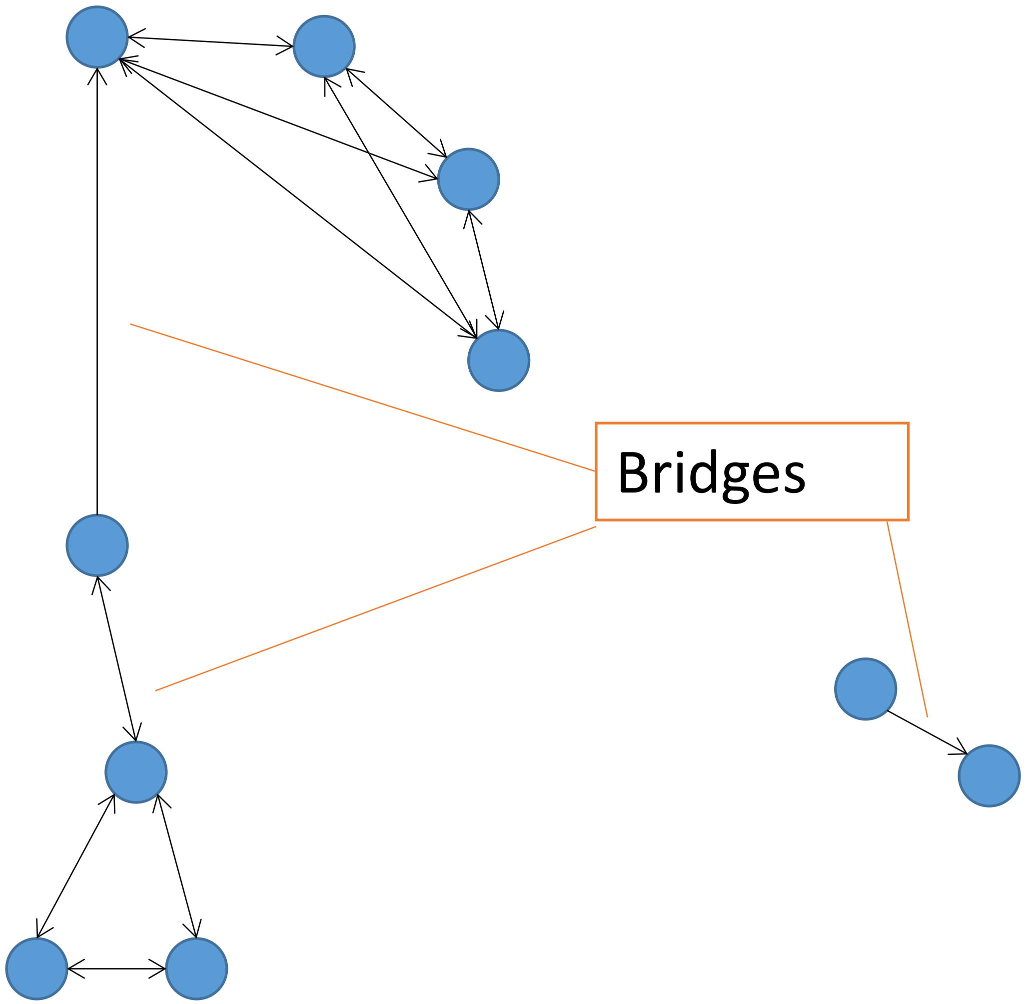
Components

Bridge



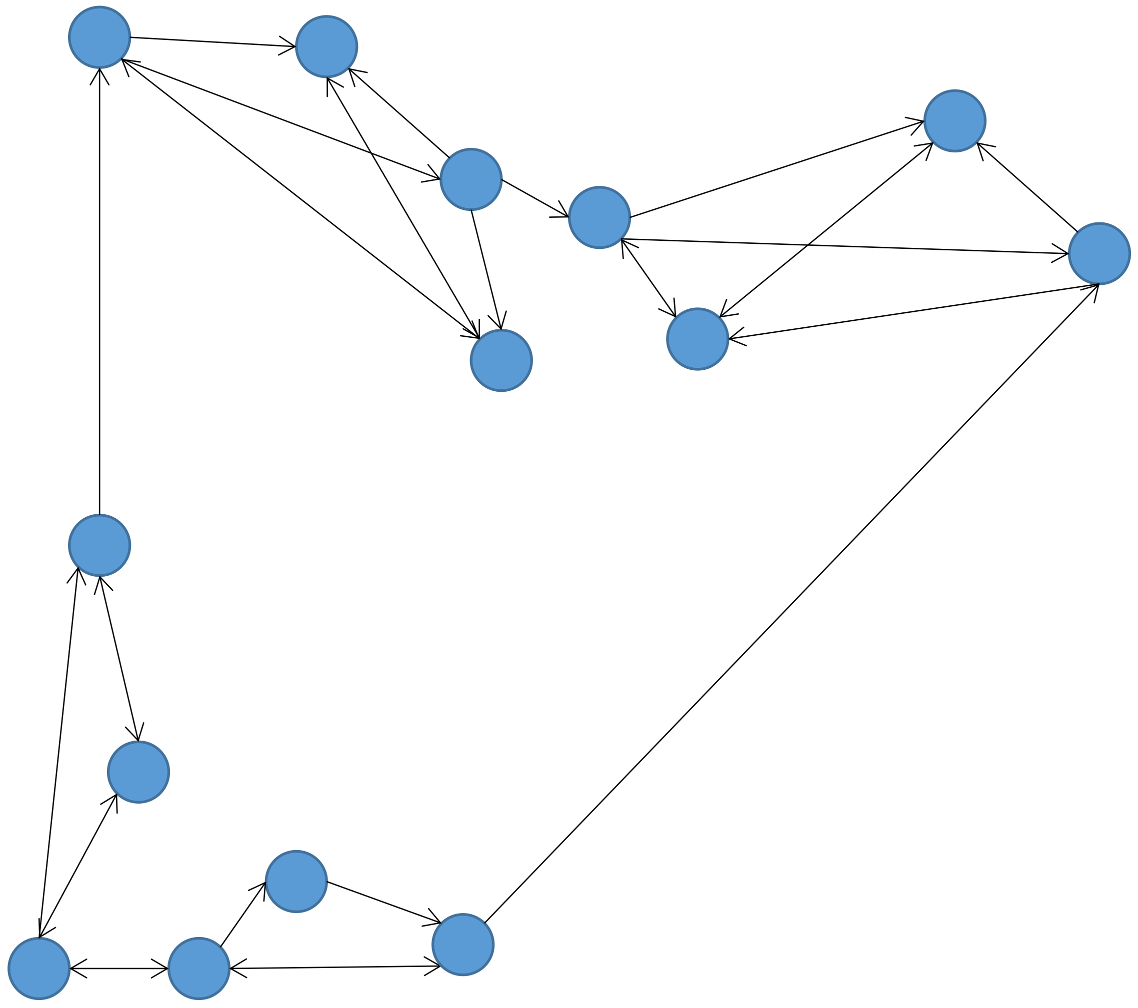
Bridge

- A tie that would increase the number of components if removed



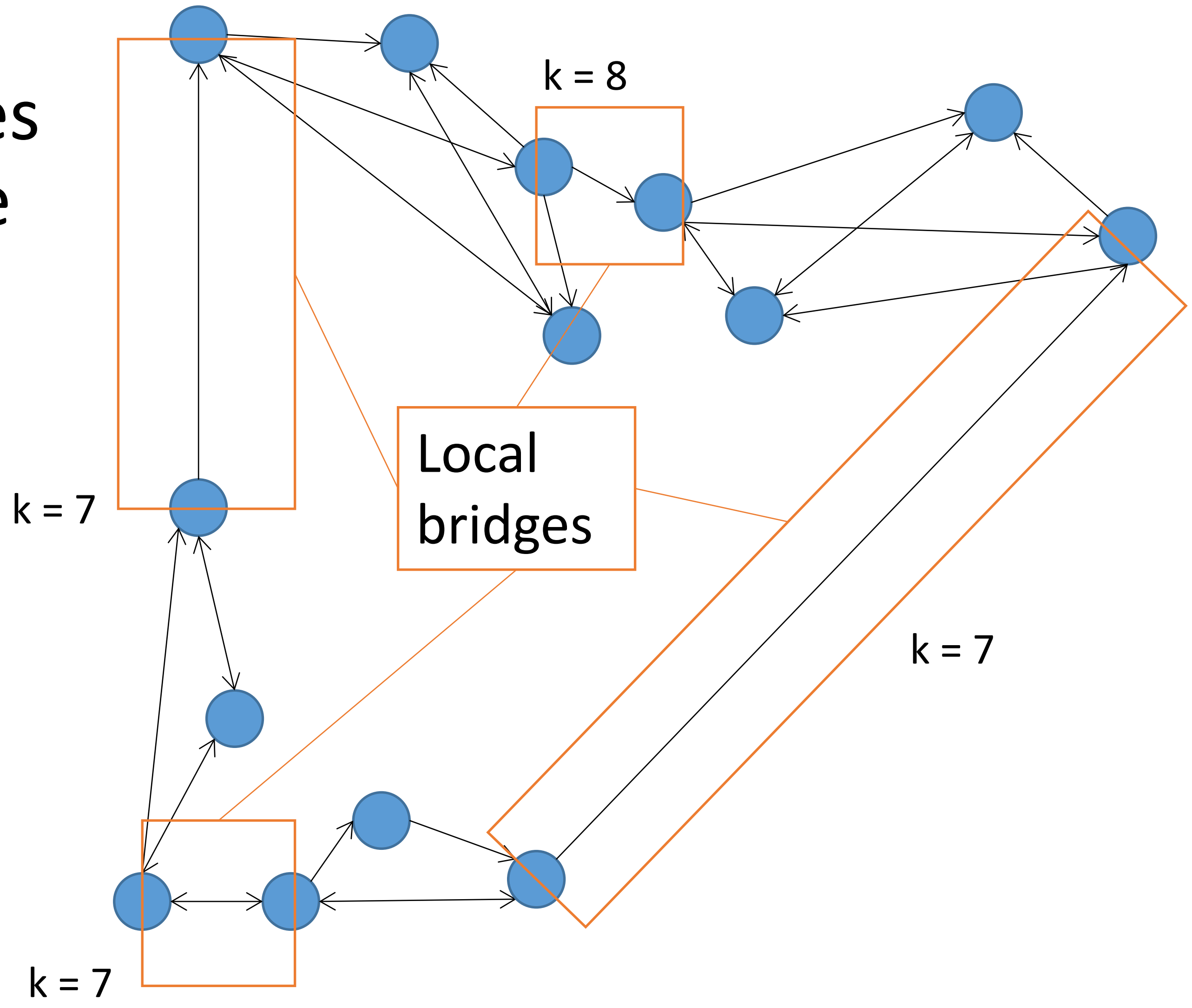
Local bridge of degree k

- A tie that connects nodes that would otherwise be at least k steps apart



Local bridge of degree k

- A tie that connects nodes that would otherwise be at least k steps apart



Cohesion

What is cohesion?

- Generally: How strongly connected a network is
- Why this that relevant? What happens in cohesive clusters?

What is cohesion?

- Generally: How strongly connected a network is
- Why this that relevant? What happens in cohesive clusters?
- Cohesive clusters
  - Enforce behavioral norms
  - Share and make sense of information
  - Develop an identity, ingroup-outgroup
  - Can coordinate for collective action
  - Share a common fate

Why care about cohesion?

- Why do some communities, organizations, families perform certain functions better than others?
- How does a community’s, family’s, organization’s internal structure affect its functions?
- Examples:
  - Charlestown resisting the urban renewal plan
  - Failure to establish a work council
  - Performance of R & D teams

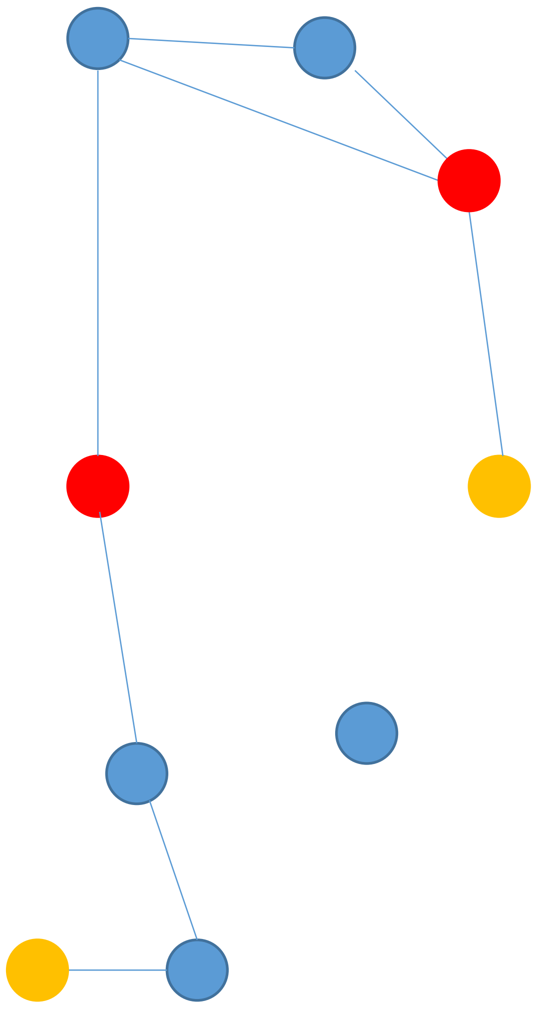
Investigating cohesion: A family of measures

- Distance
- Density
- Reciprocity
- Transitivity
- (Non-)Fragmentation
- …

Distance

- The shortest path between two nodes is the geodesic distance
- The longest shortest path in a network in the diameter
- Remember: Six degrees of separation

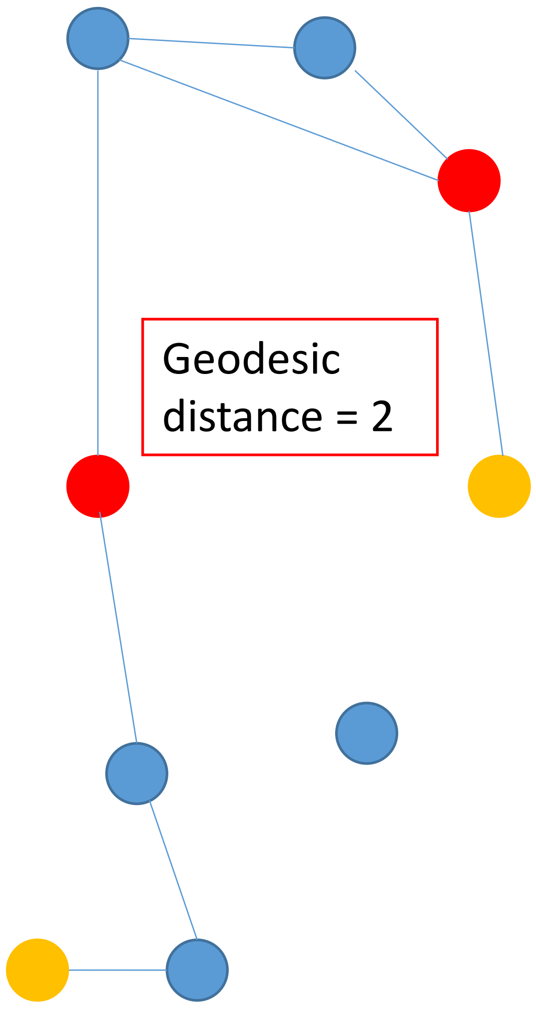
→ Geodesic distance is a measure of this



Distance

- The shortest path between two nodes is the geodesic distance
- The longest shortest path in a network in the diameter
- Remember: Six degrees of separation

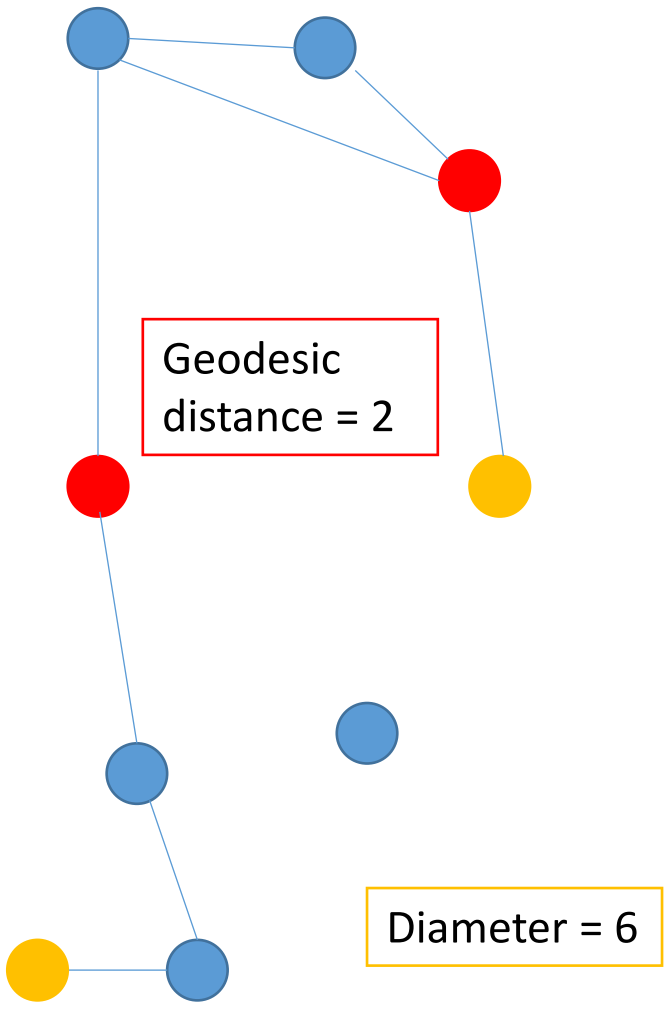
→ Geodesic distance is a measure of this



Distance

- The shortest path between two nodes is the geodesic distance
- The longest shortest path in a network in the diameter
- Remember: Six degrees of separation

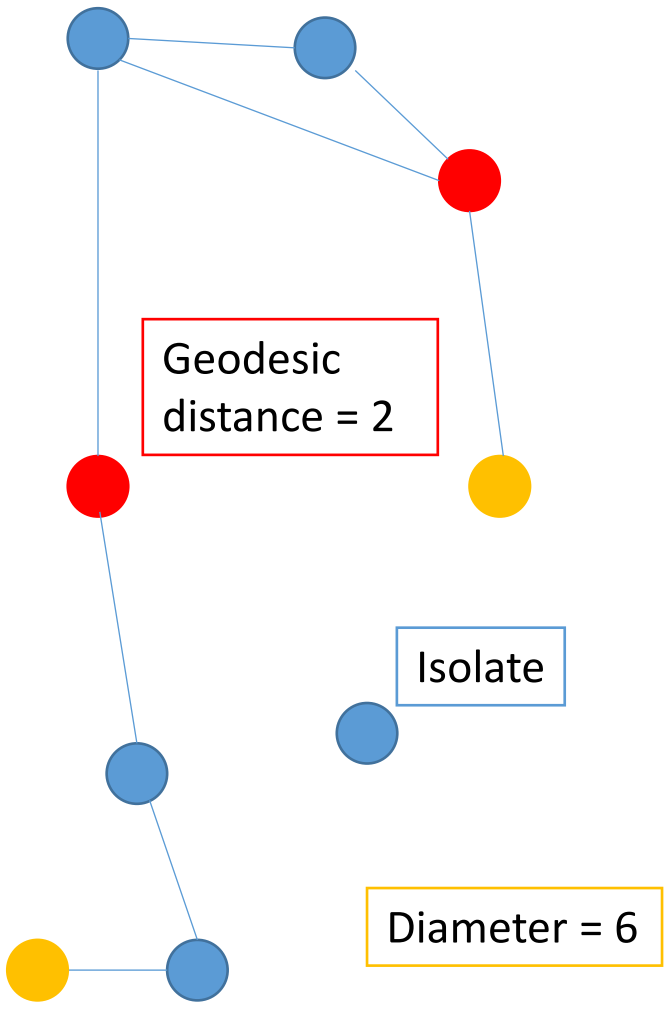
→ Geodesic distance is a measure of this



Distance

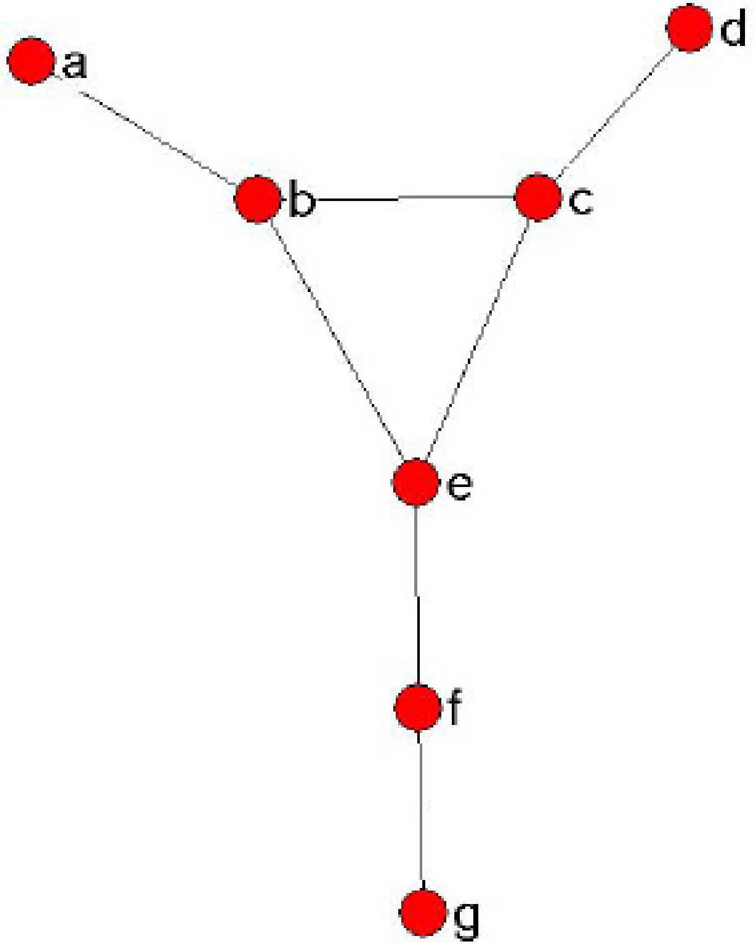
- The shortest path between two nodes is the geodesic distance
- The longest shortest path in a network in the diameter
- Remember: Six degrees of separation

→ Geodesic distance is a measure of this



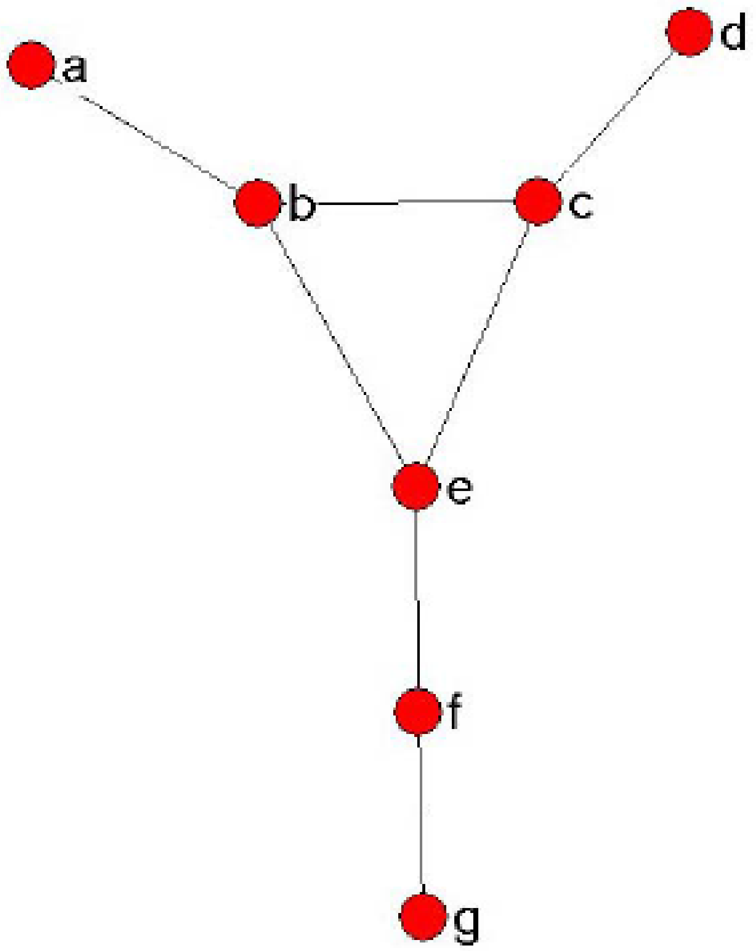
Geodesic distance matrix

	a	b	c	d	e	f	g
a							
b							
c							
d							
e							
f							
g							



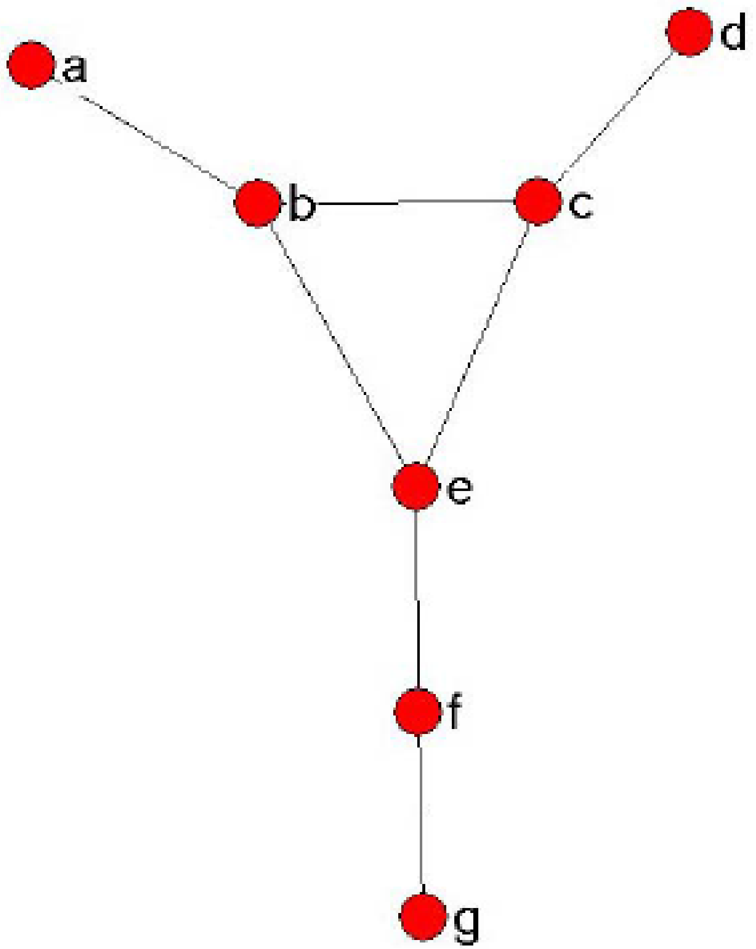
Geodesic distance matrix

	a	b	c	d	e	f	g
a	0	1	2	3	2	3	4
b							
c							
d							
e							
f							
g							



Geodesic distance matrix

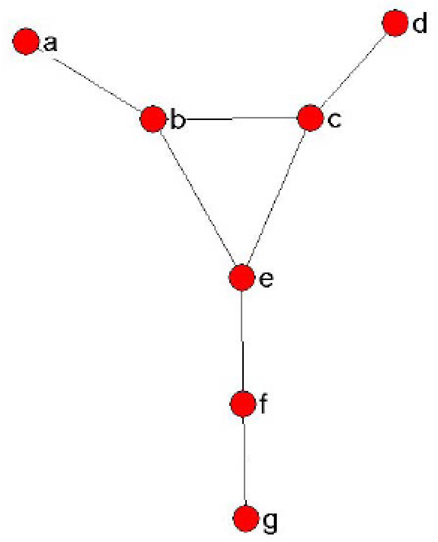
	a	b	c	d	e	f	g
a	0	1	2	3	2	3	4
b	1	0	1	2	1	2	3
c	2	1	0	1	1	2	3
d	3	2	1	0	2	3	4
e	2	1	1	2	0	1	2
f	3	2	2	3	1	0	1
g	4	3	3	4	2	1	0



Geodesic versus adjacency matrix

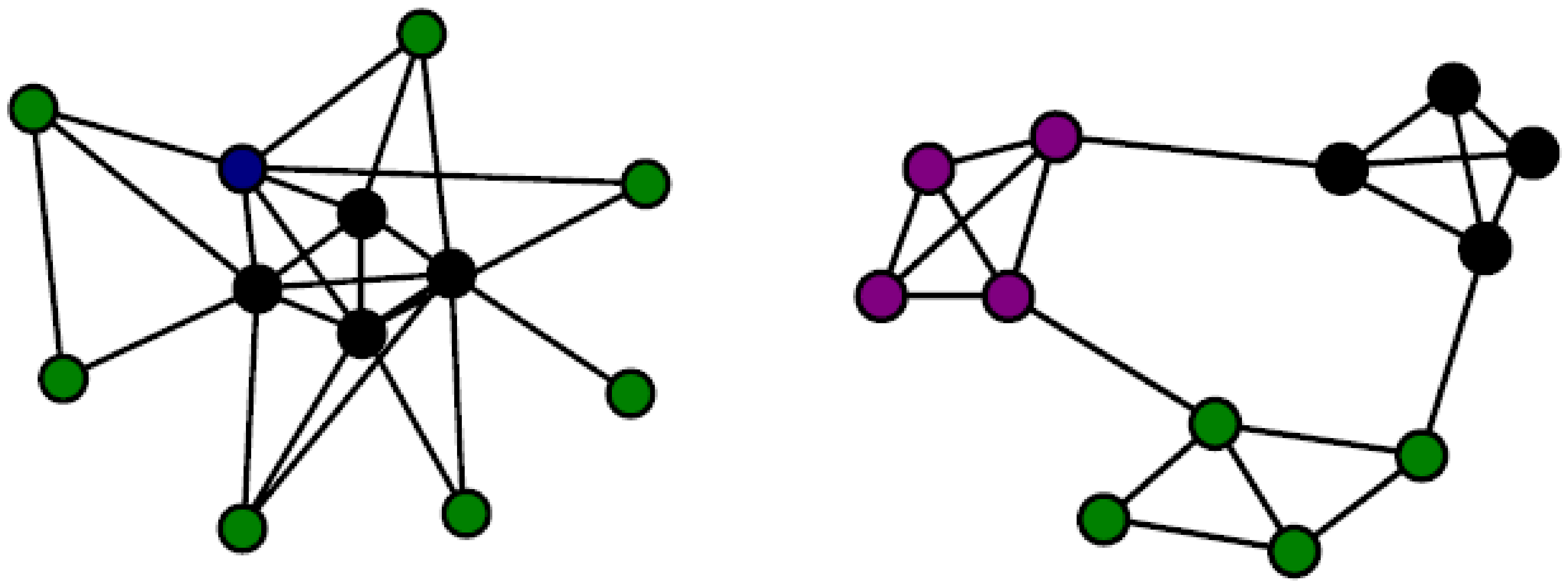
	a	b	c	d	e	f	g
a	0	1	2	3	2	3	4
b	1	0	1	2	1	2	3
c	2	1	0	1	1	2	3
d	3	2	1	0	2	3	4
e	2	1	1	2	0	1	2
f	3	2	2	3	1	0	1
g	4	3	3	4	2	1	0

	a	b	c	d	e	f	g
a	0	1	0	0	0	0	0
b	1	0	1	0	1	0	0
c	0	1	0	1	1	0	0
d	0	0	1	0	0	0	0
e	0	1	1	0	0	1	0
f	0	0	0	0	1	0	1
g	0	0	0	0	0	1	0



Average distance

- Average geodesic distance between all pairs of nodes



Core/Periphery
c/p fit = 0.97, avg. dist. = 1.9

Clique structure
c/p fit = 0.33, avg. dist. = 2.4

Distance in UCInet

Density

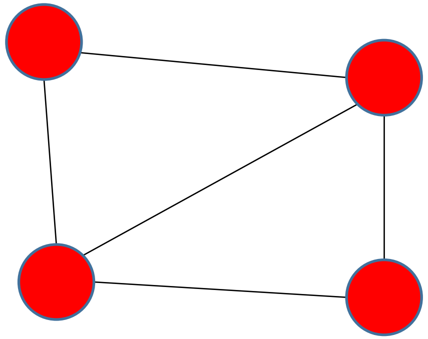
Density

- A measure for the connectedness of a network
- The sum of ties divided by the sum of possible ties
- Undirected graph: Number of possible ties = n*(n-1) / 2
    - The relationship A to B is the same as the relationship B to A
- Digraph: twice as many possible edges: n*(n-1)
    - The relationship A to B is not the same as the relationship B to A

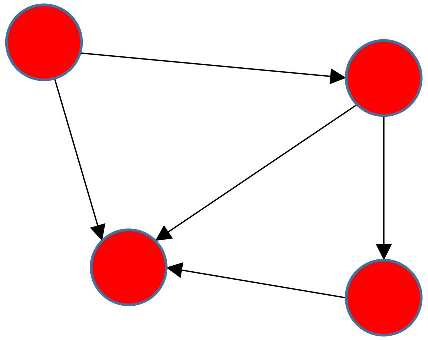
Density

- A measure for the connectedness of a network
- The sum of ties divided by the sum of possible ties
- Undirected graph: Number of possible ties = n*(n-1) / 2
  - The relationship A to B is the same as the relationship B to A
- Digraph: twice as many possible edges: n*(n-1)
  - The relationship A to B is not the same as the relationship B to A

Density = ?



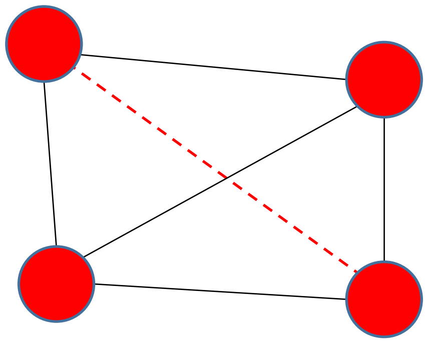
Density = ?



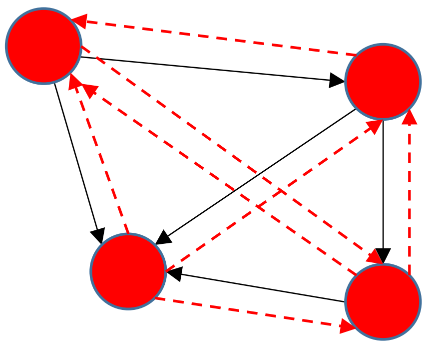
Density

- A measure for the connectedness of a network
- The sum of ties divided by the sum of possible ties
- Undirected graph: Number of possible ties = n*(n-1) / 2
  - The relationship A to B is the same as the relationship B to A
- Digraph: twice as many possible edges: n*(n-1)
  - The relationship A to B is not the same as the relationship B to A

Density = 5/6 = 0.83



Density = 5/12 = 0.42



Density in reflexive and non-reflexive ties

	Reflexive	Non-Reflexive
Undirected	$=\dfrac{T}{n^2/2}$	$=\dfrac{T}{n(n-1)/2}$
Directed	$=\dfrac{T}{n^2}$	$=\dfrac{T}{n(n-1)}$

Density

- What does it mean socially if a network is dense?

Density

- What does it mean socially if a network is dense?
  - High level of social capital
  - High degree of social control
  - High speed of information flow

Density in UCInet