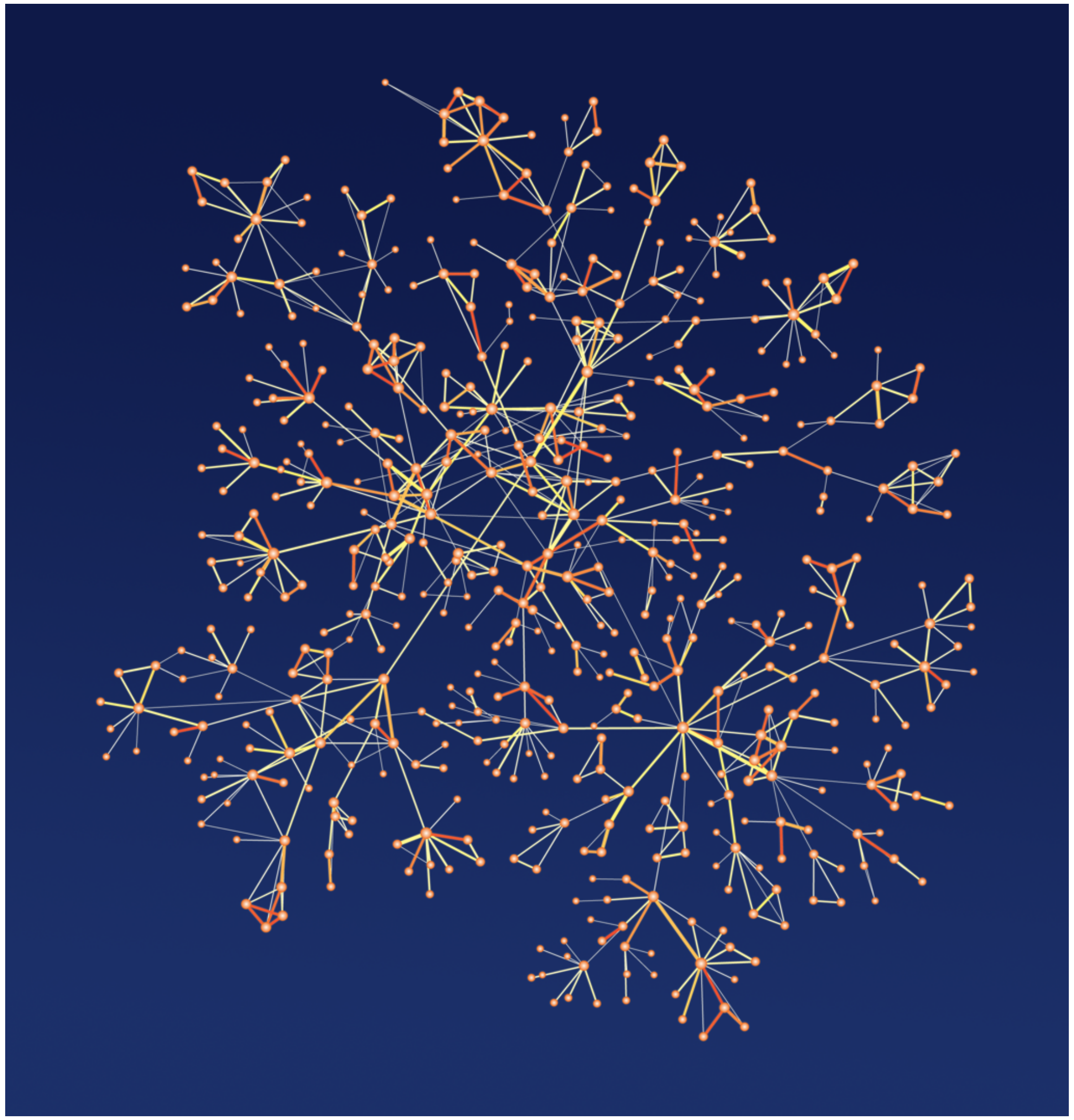
Introduction to Complex Networks



Dr. Jari Saramäki, Aalto University, Helsinki, Finland

http://www.lce.hut.fi/~jsaramak/

jari.saramaki@tkk.fi

Course Targets

After this week, you should

- know how to analyze and characterize networks
- understand the fundamental network models
- have insight into the evolution of networks
- know how network structure affects dynamic processes

Outline

Mon

Introduction, basic concepts, random networks

Tue

Small-world networks, Scale-free networks

Wed

Analytical techniques, Advanced network analysis

Thu

Weighted networks, Percolation on networks

Fri

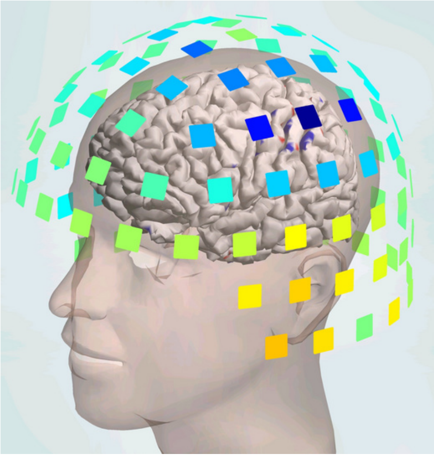
Communities and modularity, Dynamic networks

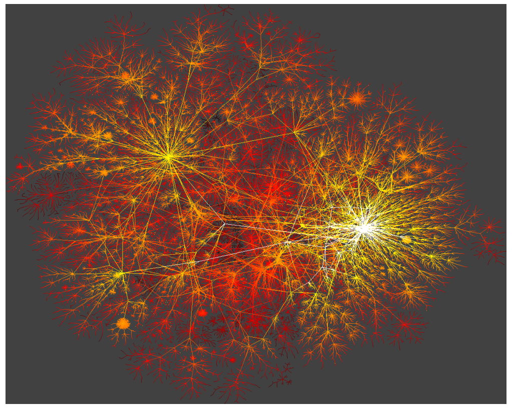
Lecture 1

The Very Basics of Complex Networks

Complex Systems

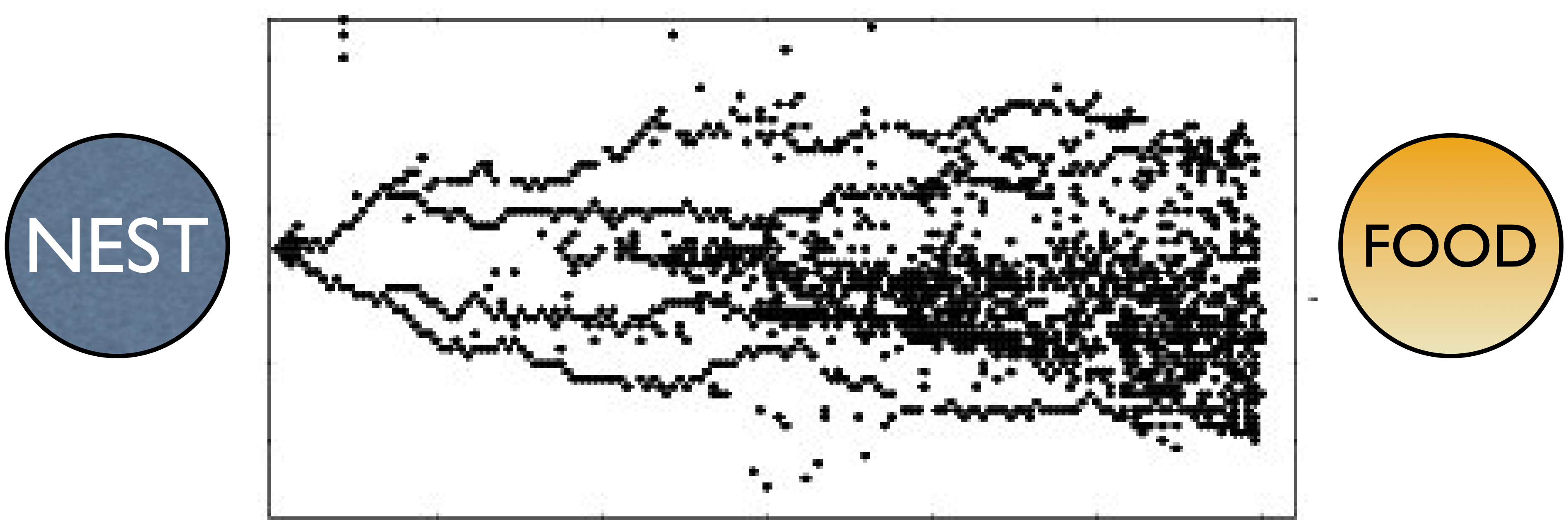
- Large number of interacting elements
- Interactions stochastic
- System adapts & evolves
- **Emergence:** elements may obey simple rules, yet the system behaves in a complex manner
- System behaviour arises from interaction structure: detailed understanding of elements in isolation won't help!





Emergence: an example

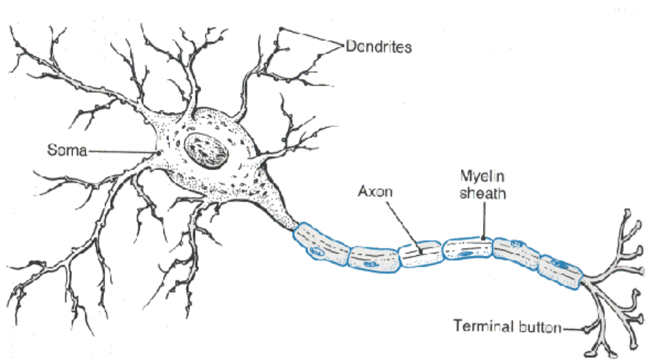
- Simple rules, complex behaviour
- **Army ant raid patterns:** rules for each ant
  1. Walk randomly, but follow scent of pheromone
  2. Deposit some pheromone while walking
  3. If food is found, carry it back to nest
  4. While carrying food, deposit lots of pheromone



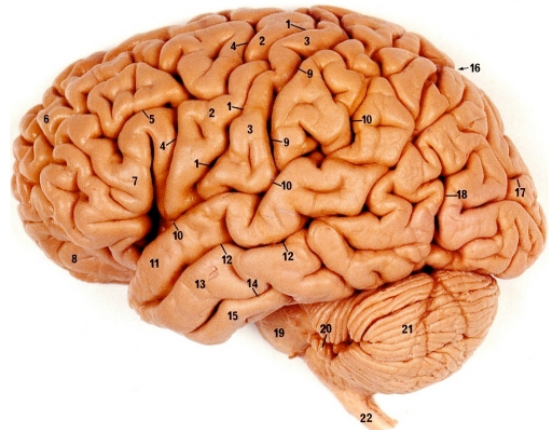
E. Bonabeau et al.

Complex Systems: How To Approach?

- Interactions between elements give rise to **emergent** behaviour
- This behaviour is apparent at the **system level**
- Studying isolated elements is not enough
- Variations in behaviour of elements often average out at the system level
- A "holistic", system-level viewpoint is needed!

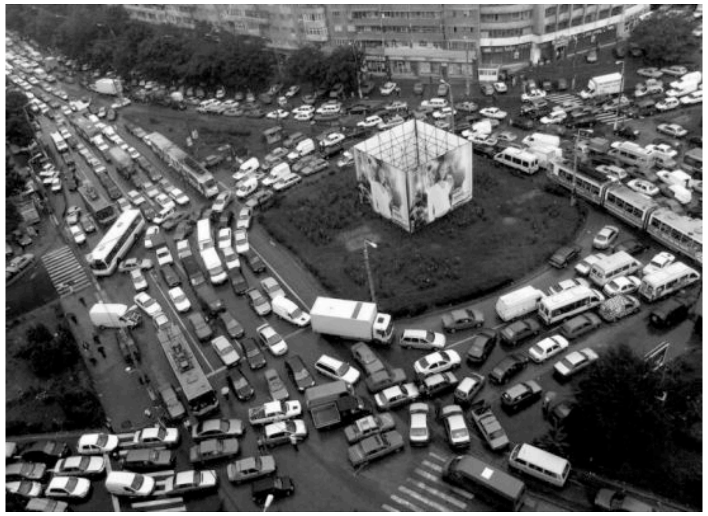


$\neq$



$\neq$

$\neq$



Complex Systems: How To Approach?

Analytical approach

- Write down (coupled) differential equations for interactions
- Attempt to solve
- Usually no closed-form solutions; numerical solutions, phase space analysis, etc

Empirical data analysis

- How to detect patterns and structure in information?
- How to characterize the system instead of its building blocks?
- Multivariate methods etc

Simulations

- Postulate rules (e.g. the ant raids)
- Simulate and observe system behaviour
- Try to match empirical observations

The network approach

- Contains elements from all these
- Disregards (unnecessary) details
- Focus on the structure of interactions
- Statistical characterization of system

The approach of (statistical) physics

The approach of (statistical) physics

1. Make observations on Nature

The approach of (statistical) physics

1. Make observations on Nature
2. Attempt to explain observations:
   2.1. Choose the right level of coarse-graining
      - Quarks ⇒ nuclear particles ⇒ atoms ⇒ molecules ⇒ ... ⇒ macroscopic bodies ⇒ planets ⇒ galaxies ⇒ Universe

   2.2. Strip the problem to its simplest form

   2.3. Formulate the problem in mathematical terms

The approach of (statistical) physics

1. Make observations on Nature
2. Attempt to explain observations:
   2.1. Choose the right level of coarse-graining
   - Quarks ⇒ nuclear particles ⇒ atoms ⇒ molecules ⇒ ... ⇒ macroscopic bodies ⇒ planets ⇒ galaxies ⇒ Universe

   2.2. Strip the problem to its simplest form

   2.3. Formulate the problem in mathematical terms
3. Check if calculations or simulations can
   - reproduce your observations
   - explain your observations

The approach of (statistical) physics

1. Make observations on Nature
2. Attempt to explain observations:
   2.1. Choose the right level of coarse-graining
      - Quarks ⇒ nuclear particles ⇒ atoms ⇒ molecules ⇒ ... ⇒ macroscopic bodies ⇒ planets ⇒ galaxies ⇒ Universe

   2.2. Strip the problem to its simplest form

   2.3. Formulate the problem in mathematical terms
3. Check if calculations or simulations can
   - reproduce your observations
   - explain your observations
4. Go back to 1. & 2. and rethink

The complex network approach

The complex network approach

1. Make observations on Nature

The complex network approach

1. Make observations on Nature
2. Attempt to explain observations:

   2.1. Choose the right level of coarse-graining

   - Vertices or nodes ⇔ interacting elements
   - Edges or links ⇔ interactions

   2.2. Strip the problem to its simplest form

   - Interaction structure ⇔ evolution and behaviour of system

   2.3. Formulate the problem in mathematical terms

   - Statistical analysis of network structure
   - Dynamics of processes taking place on networks

The complex network approach

1. Make observations on Nature
2. Attempt to explain observations:

    2.1. Choose the right level of coarse-graining

    - Vertices or nodes ⇔ interacting elements
    - Edges or links ⇔ interactions

    2.2. Strip the problem to its simplest form

    - Interaction structure ⇔ evolution and behaviour of system

    2.3. Formulate the problem in mathematical terms

    - Statistical analysis of network structure
    - Dynamics of processes taking place on networks

3. Check if calculations or simulations can
    - reproduce your observations
    - explain your observations

The complex network approach

1. Make observations on Nature
2. Attempt to explain observations:

    2.1. Choose the right level of coarse-graining

    - Vertices or nodes ⇔ interacting elements
    - Edges or links ⇔ interactions

    2.2. Strip the problem to its simplest form

    - Interaction structure ⇔ evolution and behaviour of system

    2.3. Formulate the problem in mathematical terms

    - Statistical analysis of network structure
    - Dynamics of processes taking place on networks

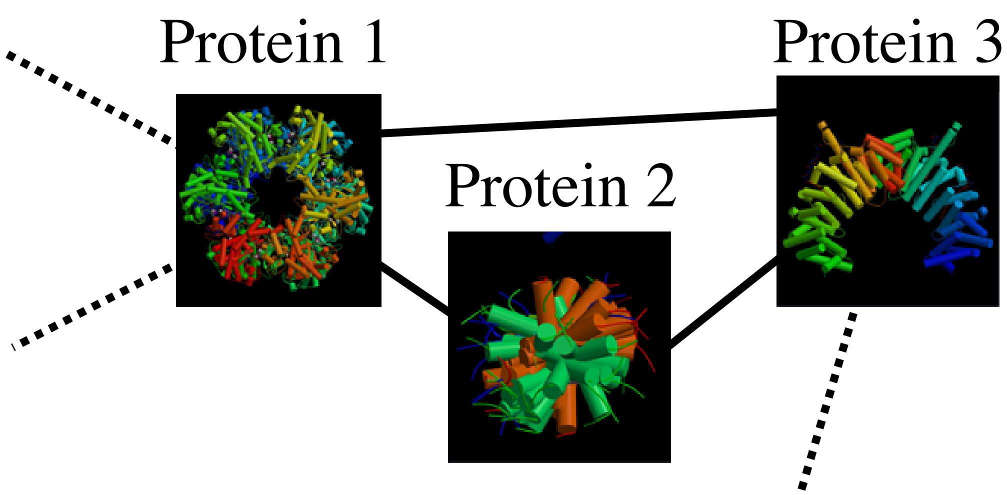
3. Check if calculations or simulations can
    - reproduce your observations
    - explain your observations
4. Go back to 1. & 2. and rethink

The Network View on Complex Systems

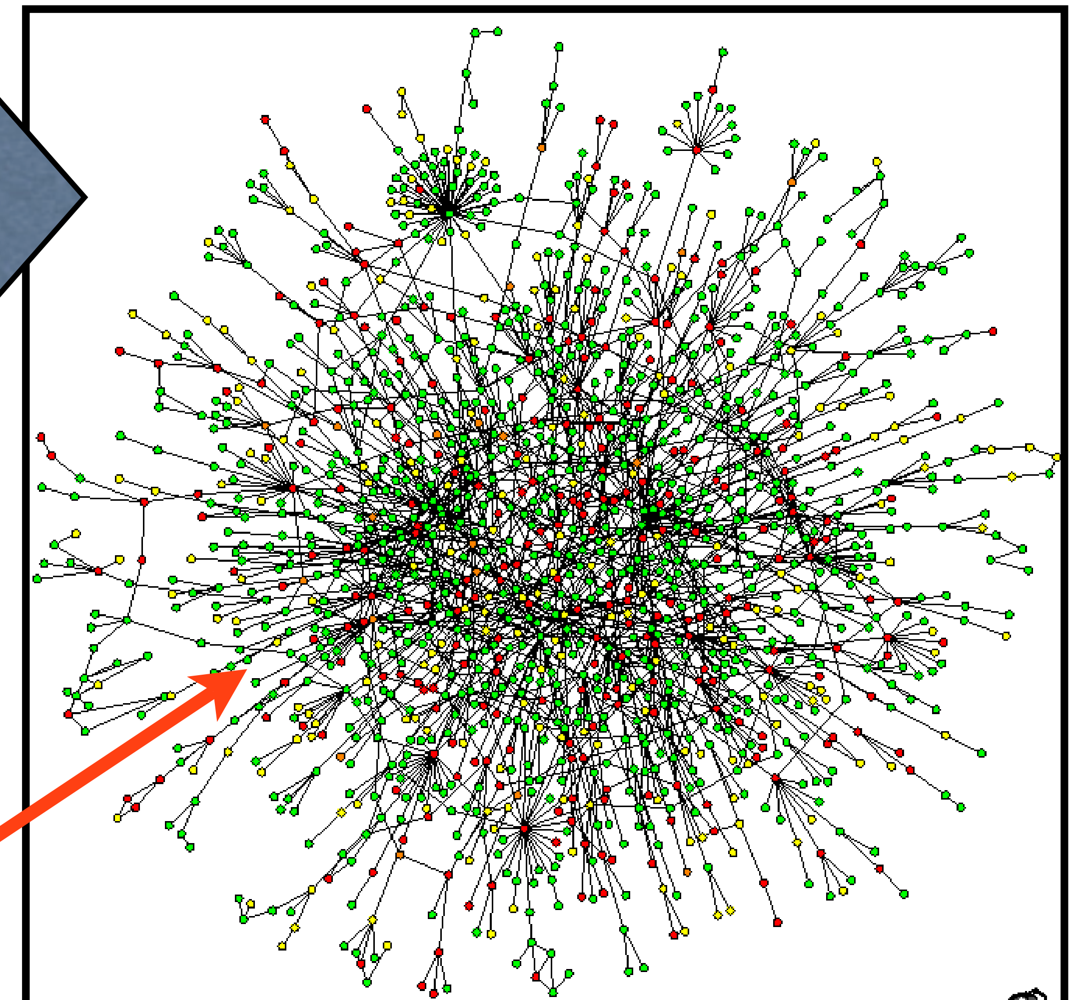
- Elements ⇔ vertices
- Interactions ⇔ edges
- An edge between v_i and v_j means v_i and v_j interact
- In reality, interactions can have different strengths, leading to *weighted networks* (to be discussed later)

Vertex	Edge
person	friendship
neuron	synapse
WWW	hyperlink
company	ownership
gene	regulation

The real question:



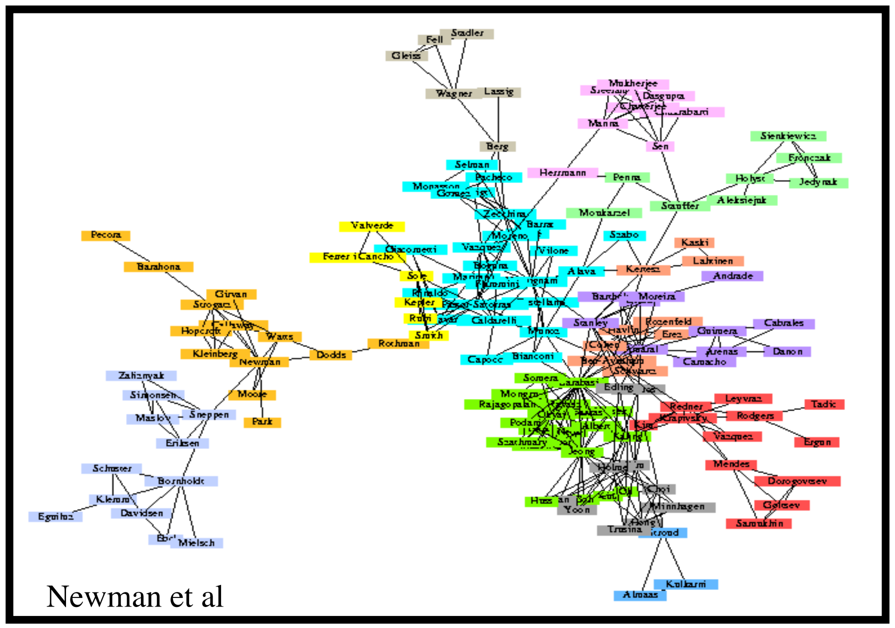
How do we deal with such things?



H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai, Nature 411, 41-42 (2001)

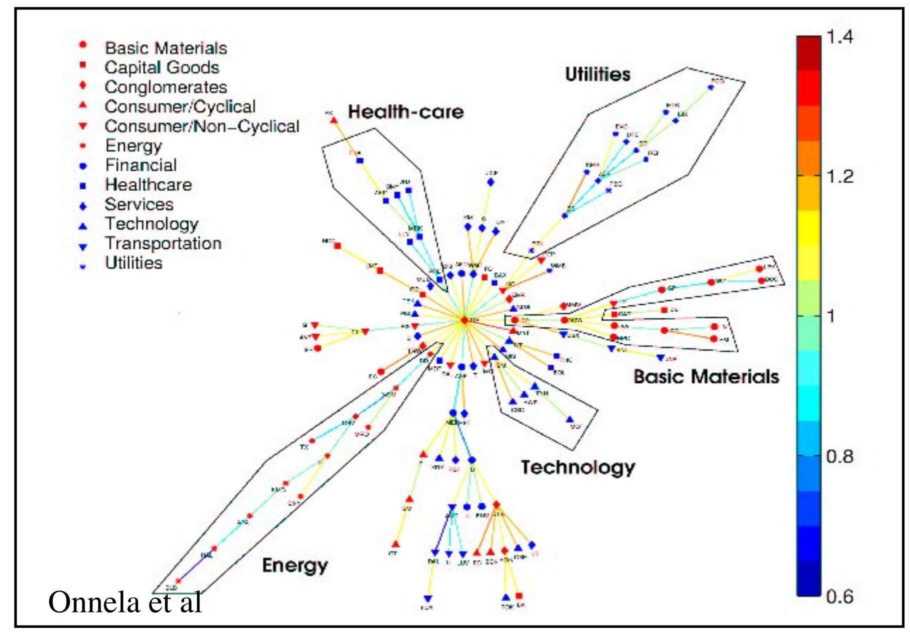
Examples of networks

scientific collaborations

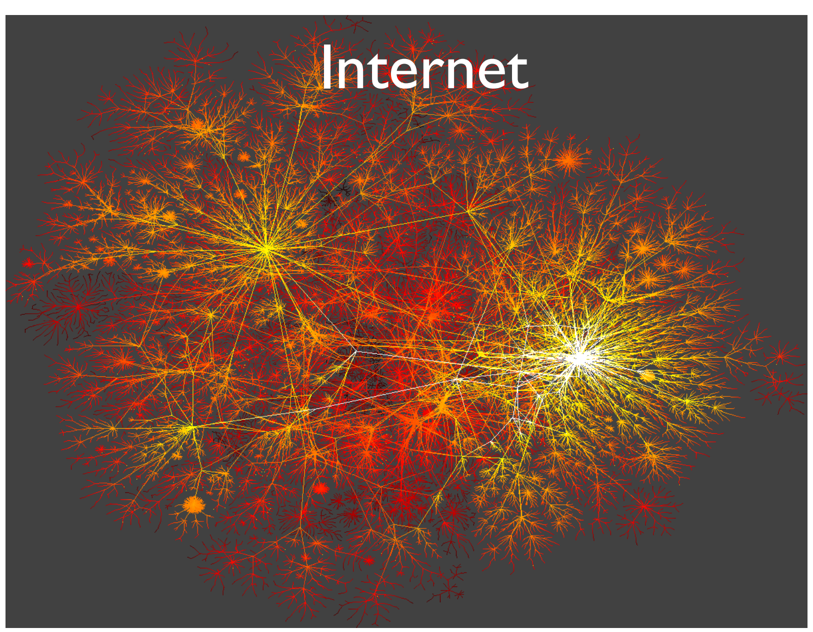


Newman et al

the stock market

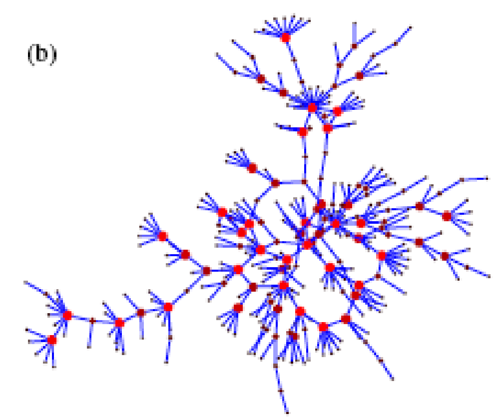


Onnela et al



Claffy et al

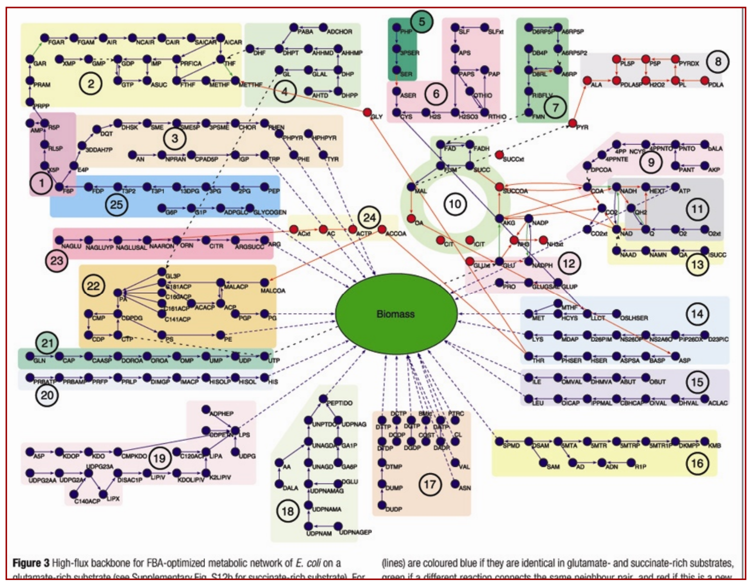
Sexual relationships



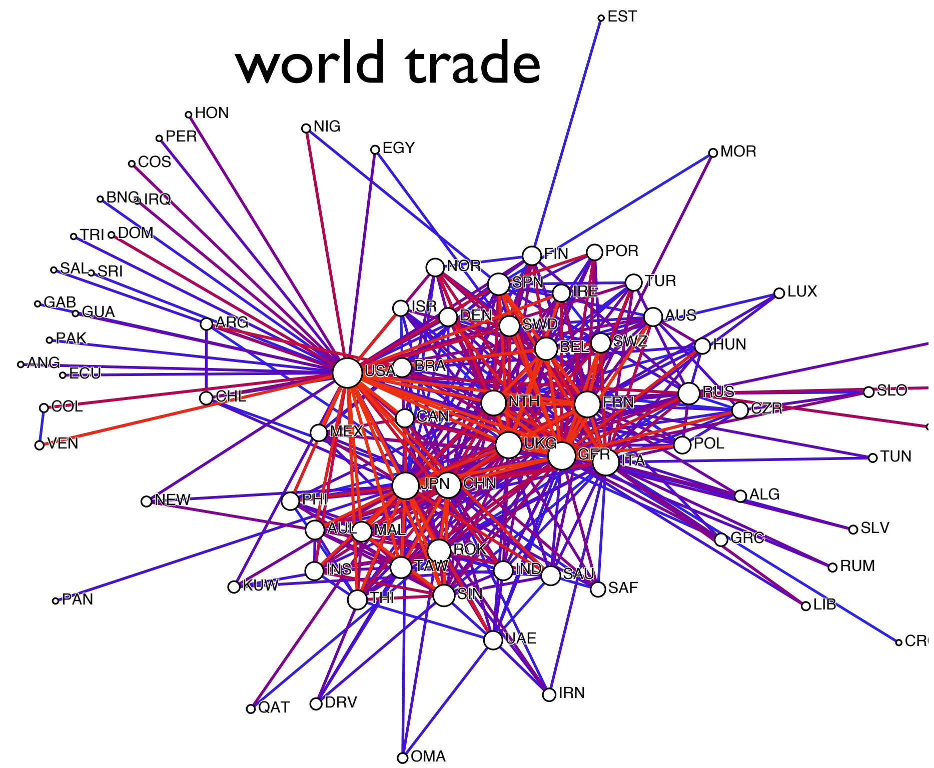
Potterat et al

More examples

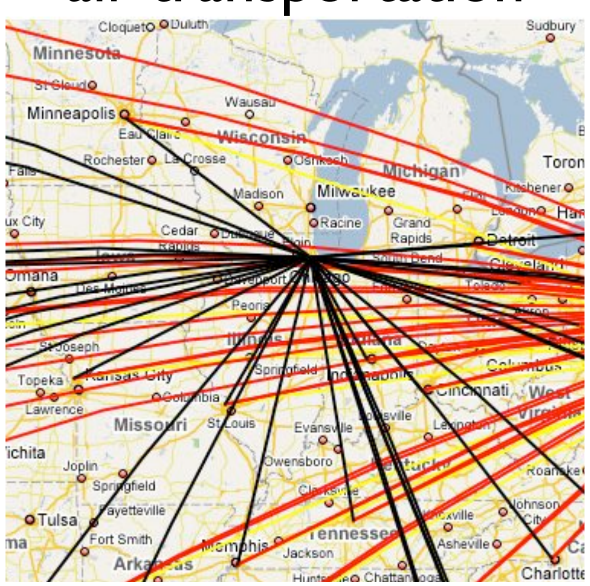
metabolic networks



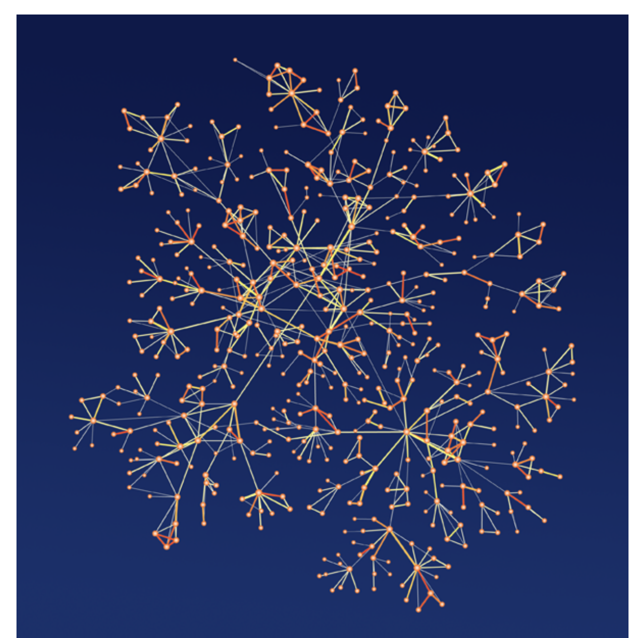
world trade



air transportation



electronic communication records



The Network View -
Why Does It Work So Well?

The Network View - Why Does It Work So Well?

- A common framework applicable to many systems

The Network View - Why Does It Work So Well?

- A common framework applicable to many systems
- Different systems can be studied with same methods

The Network View - Why Does It Work So Well?

- A common framework applicable to many systems
- Different systems can be studied with same methods
- A "birds-eye" view on the system

The Network View - Why Does It Work So Well?

- A common framework applicable to many systems
- Different systems can be studied with same methods
- A “birds-eye” view on the system

The Network View - Why Does It Work So Well?

- A common framework applicable to many systems
- Different systems can be studied with same methods
- A “birds-eye” view on the system

- **MANY NETWORKS SHARE SIMILAR CHARACTERISTICS**

The Network View - Why Does It Work So Well?

- A common framework applicable to many systems
- Different systems can be studied with same methods
- A “birds-eye” view on the system

- **MANY NETWORKS SHARE SIMILAR CHARACTERISTICS**

The Network View - Why Does It Work So Well?

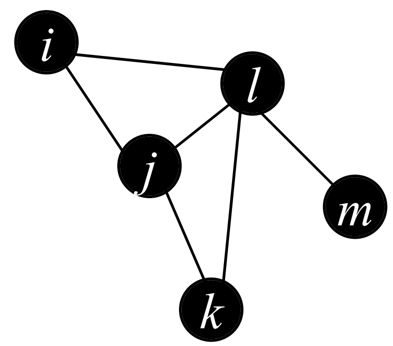
- A common framework applicable to many systems
- Different systems can be studied with same methods
- A "birds-eye" view on the system

- **MANY NETWORKS SHARE SIMILAR CHARACTERISTICS**

- These are because similar processes shape the networks

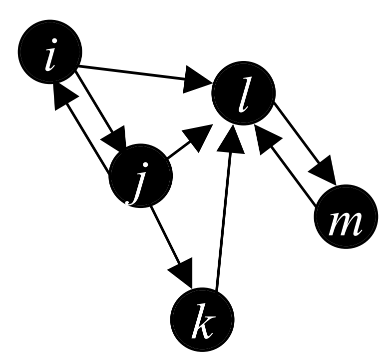
Fundamentals: network

- **Network** (*phys.*)
  = **graph** (*math, comp. sci.*)
- A network is a collection of vertices V and edges E
- E is a set of pairs of elements of V
- If the pairs are *ordered*, we have *directed* networks; otherwise they are *undirected*
- If no self-edges and no multiple edges are allowed, the network is *simple*
- We only deal with these!



undirected

$V=\{i,j,k,l,m\}$
$E=\{ij,il,jl,jk,lk,lm\}$



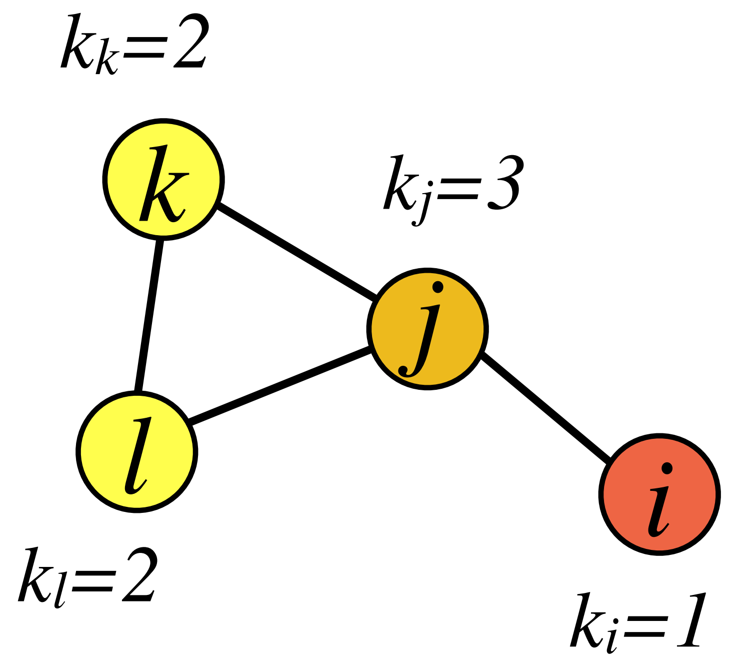
directed

$V=\{i,j,k,l,m\}$
$E=\{ij,ji,il,jl,jk,$
$kl,lm,ml\}$

Fundamentals: neighbours, degree

- Vertex v_i is a **neighbour** of vertex v_j if $\{i,j\} \in E$
- Vertex v_i is a 2nd (order) neighbour of vertex v_k if $\{i,j\},\{j,k\} \in E$, $\{i,k\} \notin E$
- The number of neighbours k of a vertex has is called its **degree**
- In directed networks, one can distinguish between in- and out-degrees k_{in}, k_{out}

- The probability distribution $p(k)$ of degrees is one of the central concepts in network analysis



Fundamentals: density, subgraphs

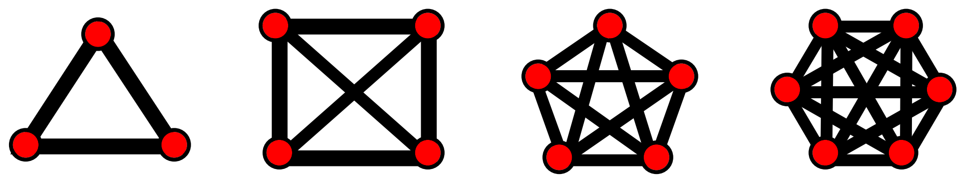
- The edge density of a network of N vertices is the ratio of numbers of existing edges $|E|$ and possible edges $\frac{1}{2}N(N-1)$:

$$\rho = \frac{2|E|}{N(N-1)}$$

- Real-world networks are typically **sparse**, *i.e.* their density is low

- If all edges exist, i.e. $\rho=1$, the network is **fully connected**

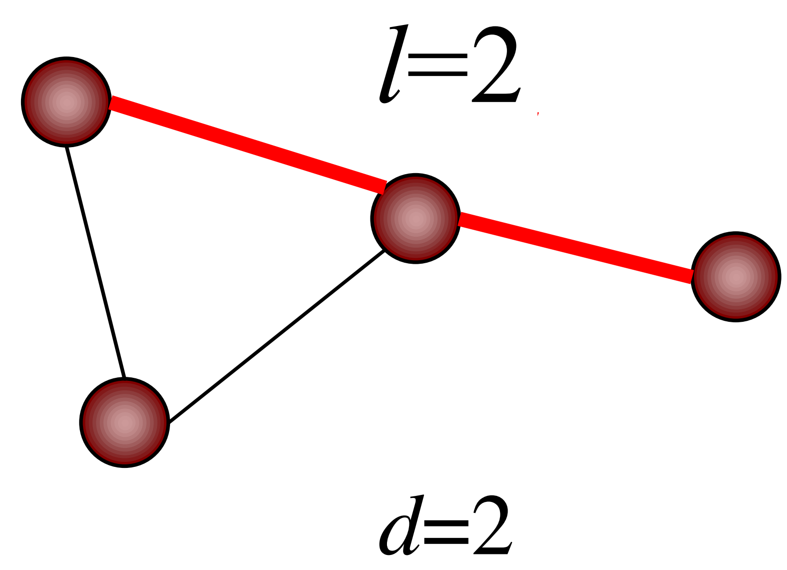
- Graph $\{V',E'\}$ is a subgraph of V, if $V' \in V, E' \in E$
- Fully connected subgraphs of k nodes are called k-cliques:



k=3 k=4 k=5 k=6

Fundamentals: paths, diameter

- The sequence of edges $\{i,j\},\{j,k\},\ldots,\{p,q\}$ defines a **path** between v_i and v_p
- The length l of the path is just its number of edges
- The distance d_{ij} between two vertices is the length of the shortest path connecting the vertices
- The longest shortest path of the network is its **diameter** d



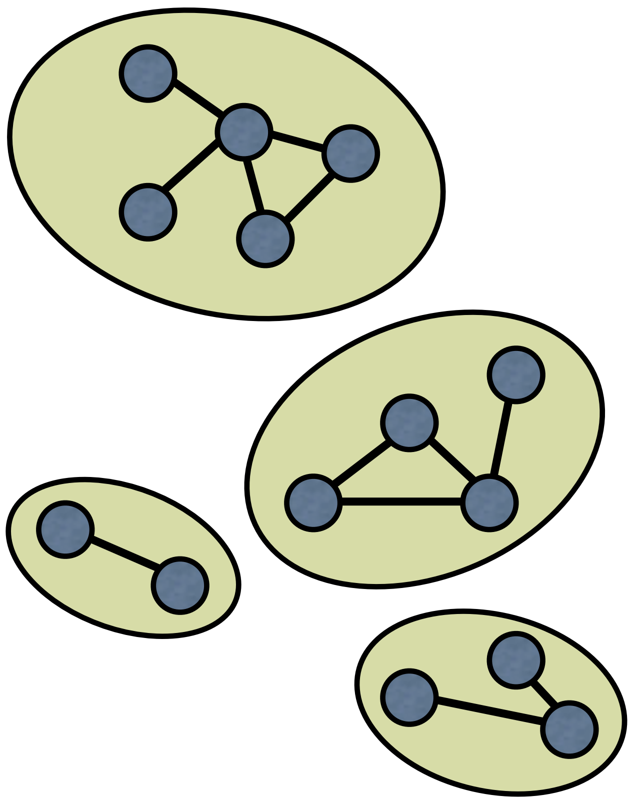
Fundamentals: components

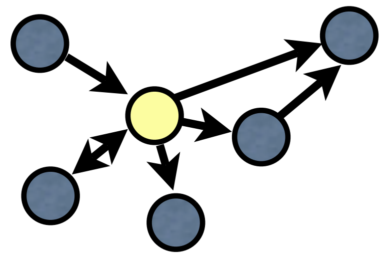
- A **connected component** is a subset of vertices with at least one path connecting each of them

- A network may consist of **a single connected component** (a connected network) or several of those

- Distances between nodes in disjoint components are not defined (infinite)

- For directed graphs, any vertex has an in-component (set of nodes with paths to it) and out-component (set of nodes with paths from it)

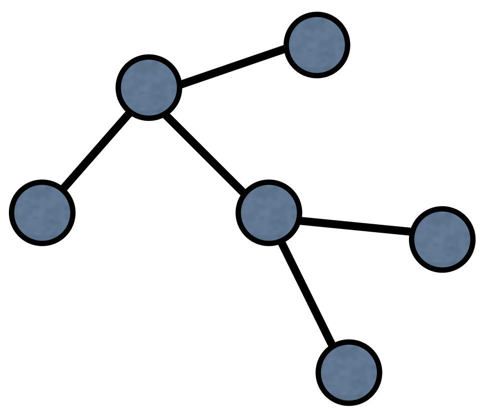




Special network types

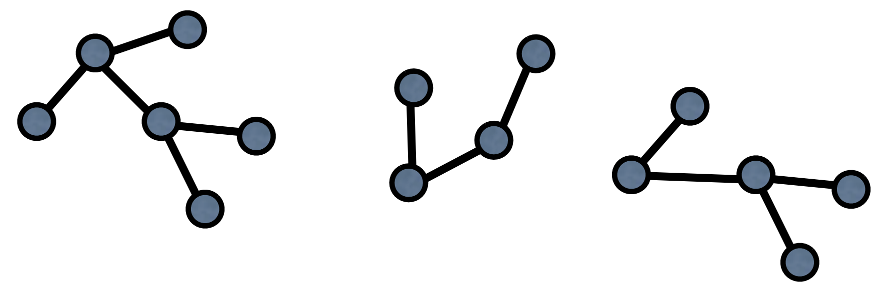
Tree

- there are no loops
- has N-1 edges



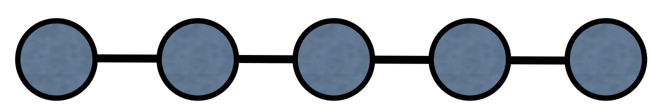
Forest

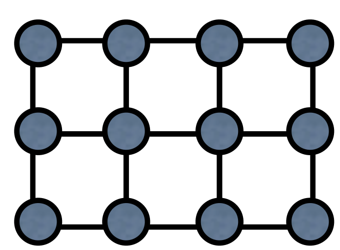
- collection of trees



Regular networks

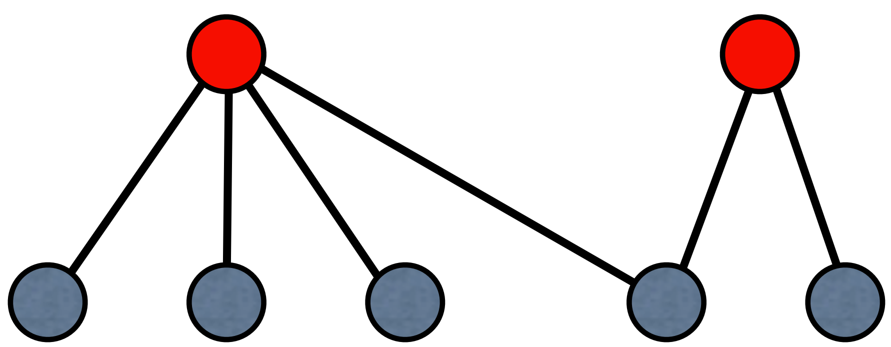
- "know one part, you know the rest"



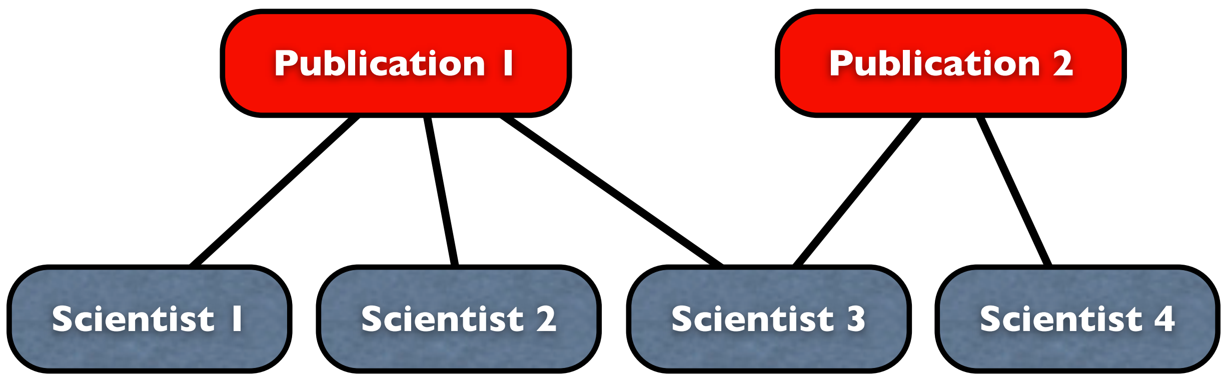


Bipartite networks

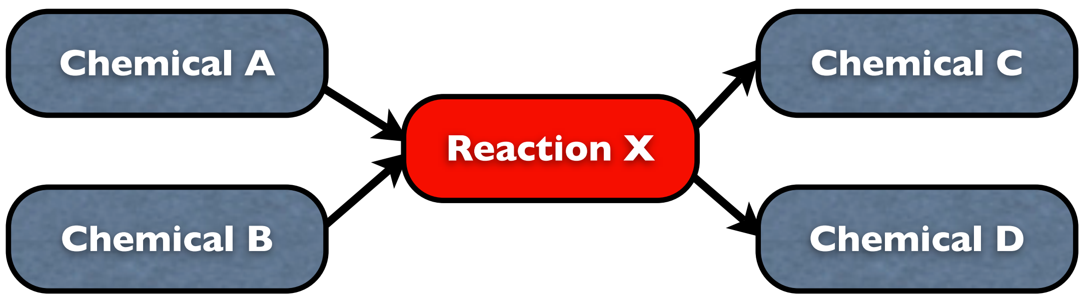
- Two types of nodes, links only between nodes of different type



Scientific collaborations

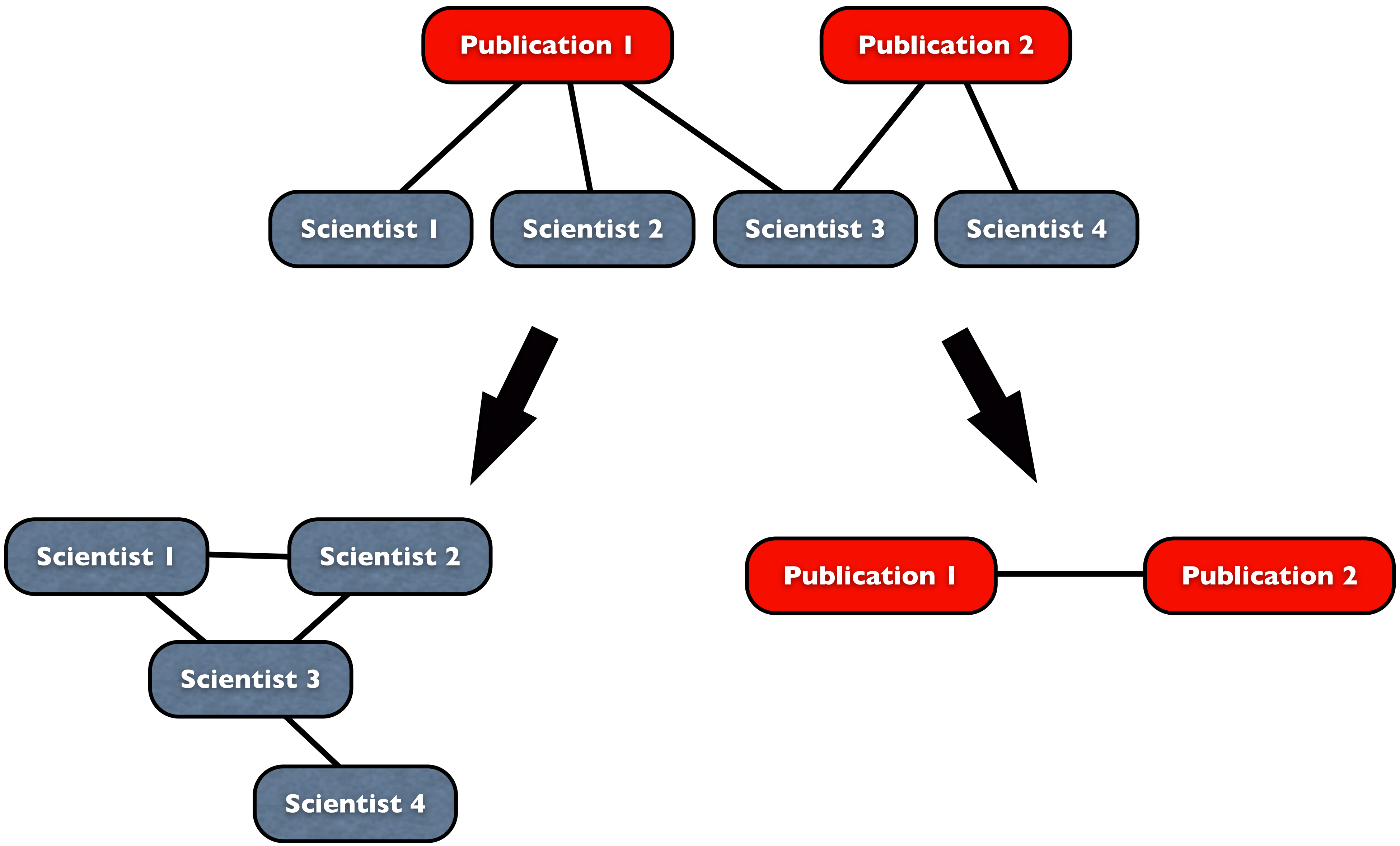


Metabolic reactions



Bipartite networks

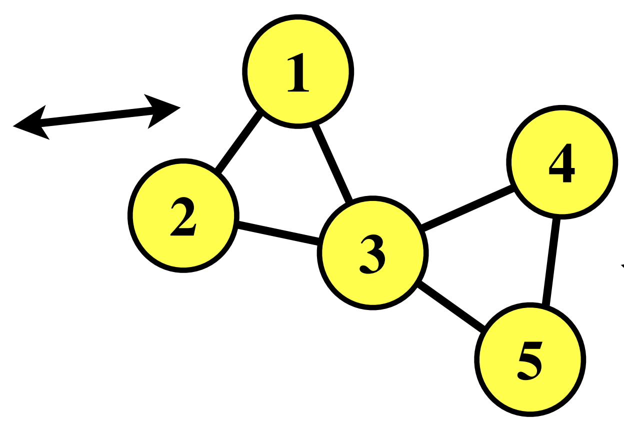
- Can be collapsed into nodes of one type



Storing & representing networks

- Let us label the vertices $1,2,...,N$
- The network can be represented as a list of edges:

i	j
1	2
1	3
2	3
3	4
3	5
4	5



- Sometimes neighbour lists are used:
`node(1).neighbours=[2 3]`

- Mathematically, one typically uses an **adjacency matrix** A:

$$\begin{aligned} A_{ij} &= 1, \text{if } \{i,j\} \in E \\ &= 0, \text{if } \{i,j\} \notin E \end{aligned}$$

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

- For computational purposes, directly representing the network as a matrix consumes too much memory!

On (Stochastic) Network Models

- Stochastic sets of rules for generating networks
- Target: to see what network features result from or can be explained by the rules
- Complex models can be viewed as agent-based models
  - "Agent" = node = e.g. individual
  - Rules mimick behaviour of agents

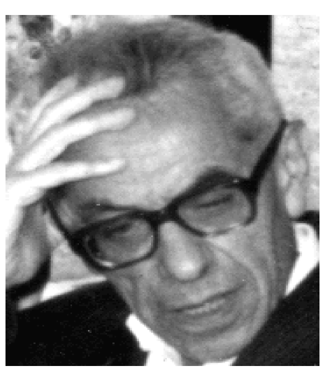
Erdös-Rényi Networks

Erdös-Rényi network:

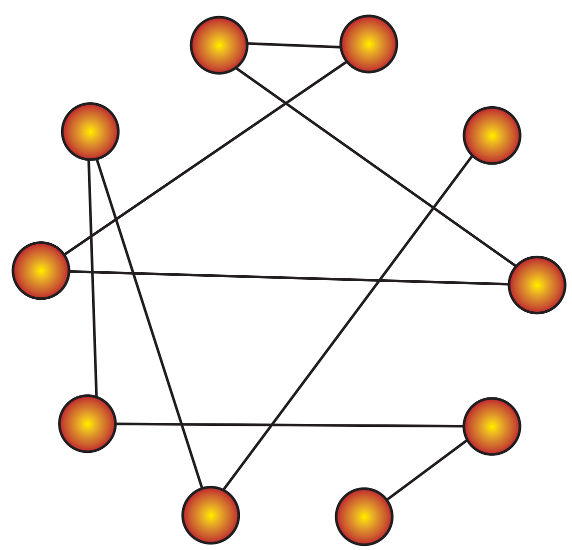
- An entirely random network of given size
- "Zeroth-order approximation" of real-world networks

Construction:

- Connect N vertices randomly
- Each pair is connected with probability p
- That's it!



Pál Erdös
(1913-1996)



$N = 10$

$p = 1/5$

$\langle k \rangle = 1.8$

Erdős, P.; Rényi, A. (1959). "On Random Graphs. I.". Publicationes Mathematicae 6: 290–297.

Erdös-Rényi Networks: Averages

Average number of edges:

- There are $N(N\text{-}1)/2$ pairs of nodes
- Each connected with probability p
- Hence the average number of edges

$$\langle E \rangle = pN\left(N-1\right)/2$$

Average degree:

$$\langle k \rangle = \frac{2\langle E \rangle}{N} = p\left(N-1\right) \approx pN$$

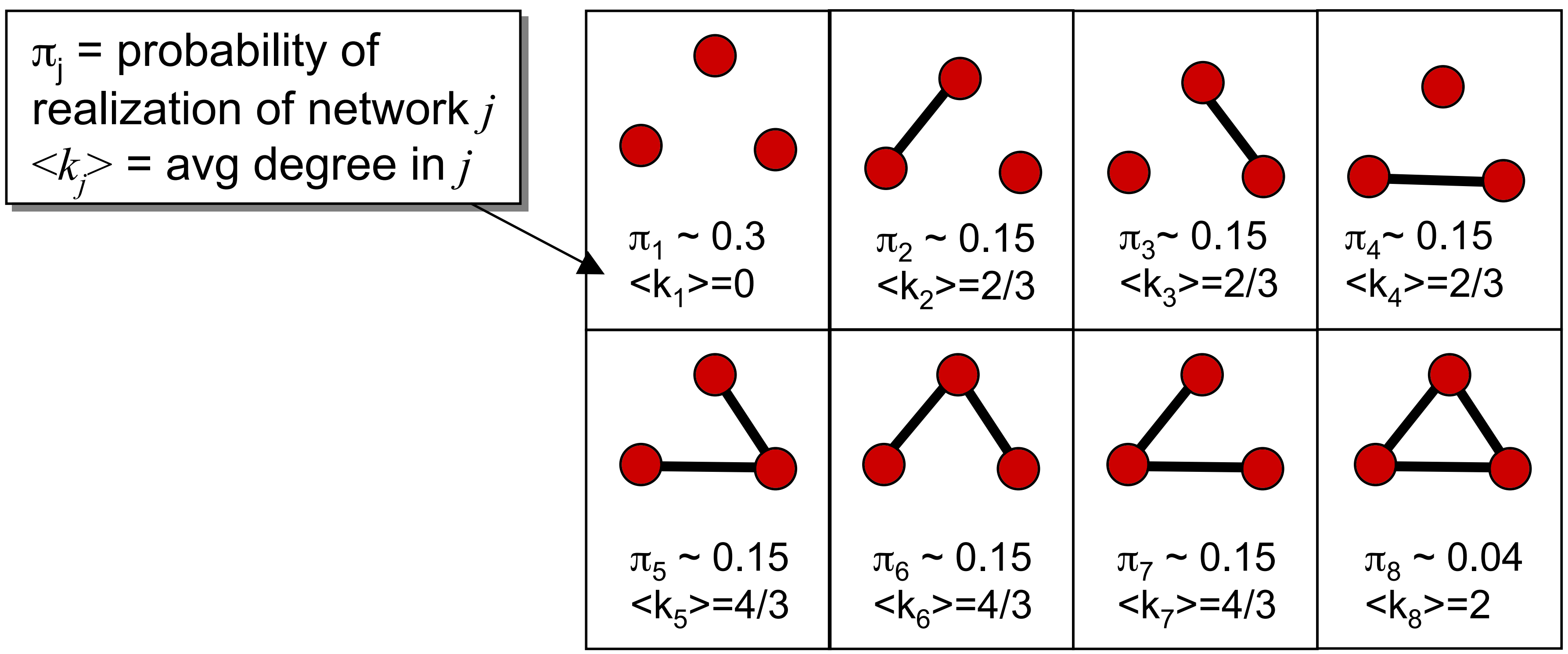
A Small Detour: Ensembles

- What do we exactly mean when we say that the average degree of an E-R network is $p(N\text{-}1)$?
- There is a multitude of possible realizations of an E-R network with N, p fixed
- These are drawn from the **ensemble** of all possible realizations

- Properties of model networks are to be considered as **averages within this ensemble**
- Each realization within the ensemble comes with its own probability
- For a *single* realization, the *expected* average degree is $p(N\text{-}1)$

Example ensemble

N=3, p=1/3



$$\langle k \rangle = \sum_{j=1}^{8} \pi_j \langle k \rangle_j = \frac{2}{3} = p(N-1)$$

The Degree Distribution

- "If a random vertex is picked, what is the probability that its degree equals k?"

- Denote by $p_i(k)$ the probability that vertex i has degree k

- For the whole network

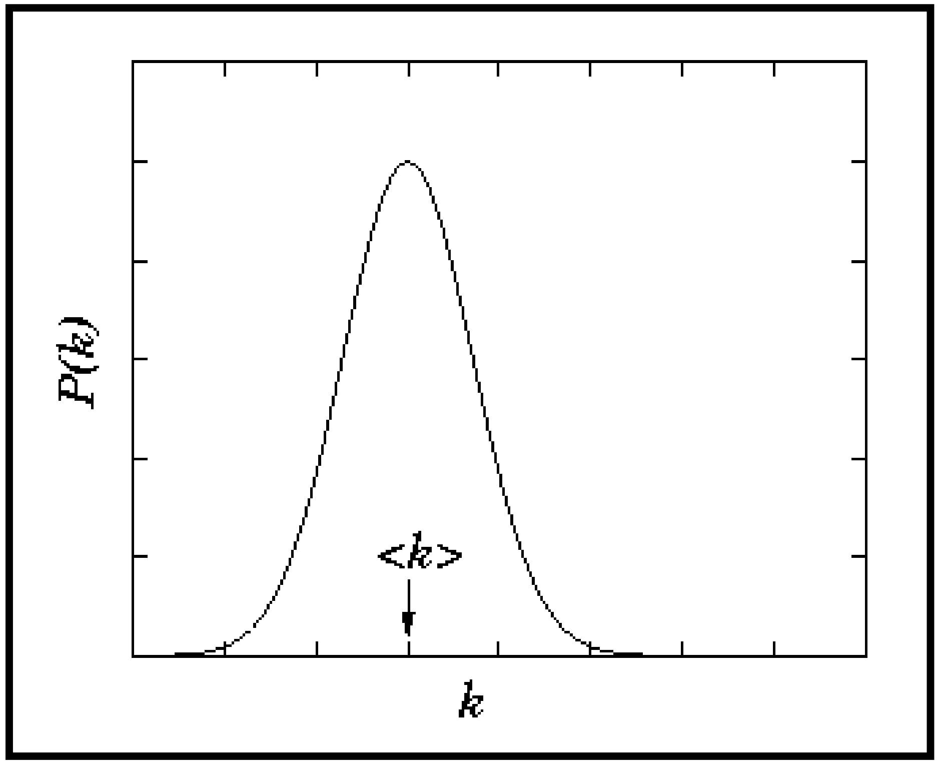
$$P(k) = \frac{1}{N}\sum_{i=1}^{N} p_i(k)$$

- For E-R networks, all vertices are alike, so $p_i(k) = P(k)$ for all i

- Erdös-Rényi networks:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

(for large N, such that $<k>=pN=const$)



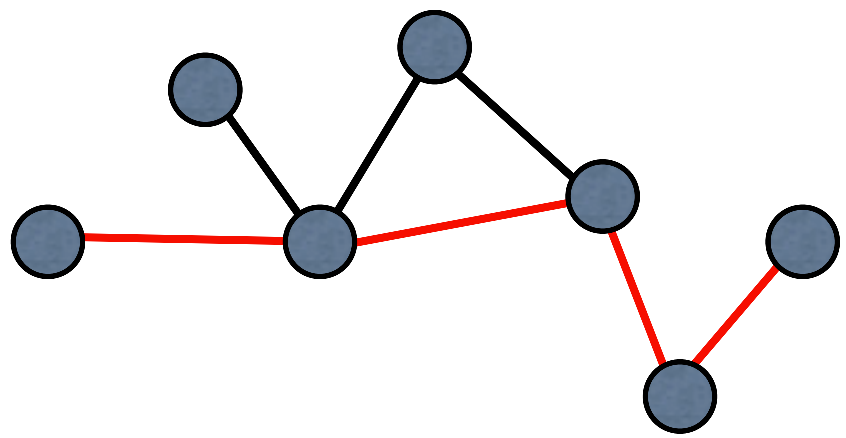
$$\langle k \rangle = \frac{2\langle E \rangle}{N} = p\,(N-1) \approx pN$$

Average shortest path length

- The average shortest path length $<l>$, characterizes the compactness of the network

$$\langle l \rangle = \frac{1}{N\,(N-1)} \sum_{i,j,i\neq j}^{N} l_{ij}$$

- Sometimes the diameter $d = \max(l_{ij})$ is used instead



Average shortest path length in E-R networks

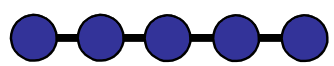
- Let us assume that there is a single connected component (a strong assumption, we'll get back to this)

- Then for E-R networks

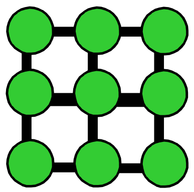
$$\langle l \rangle \propto \ln N$$

- The path length grows very slowly with network size; paths are short even for very large E-R networks

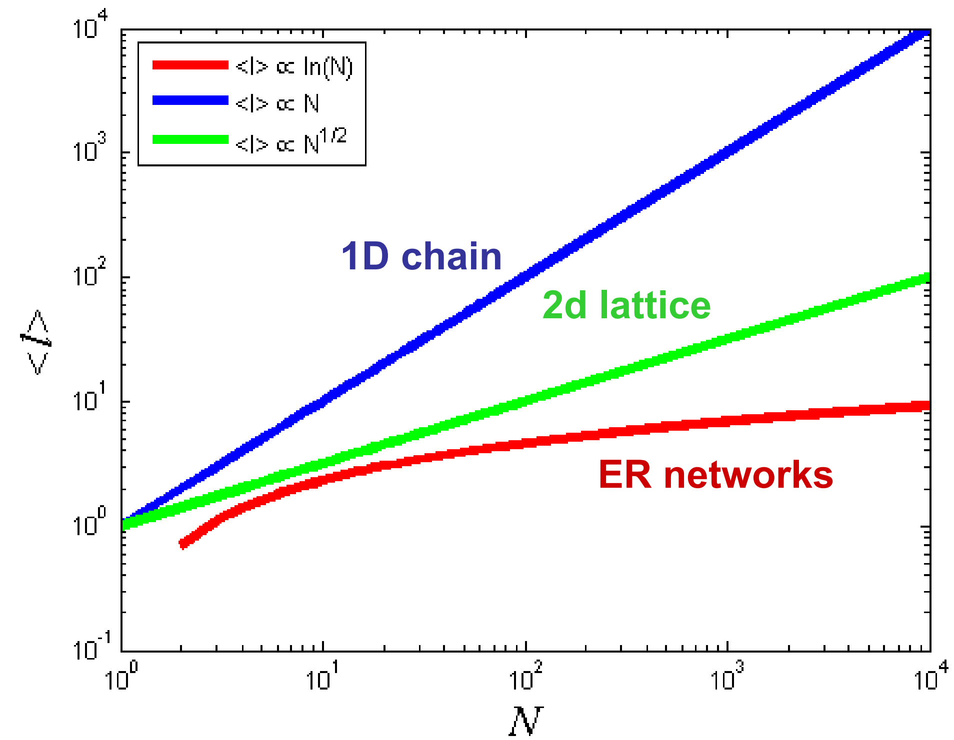
for comparison:



$\langle l \rangle \propto N$



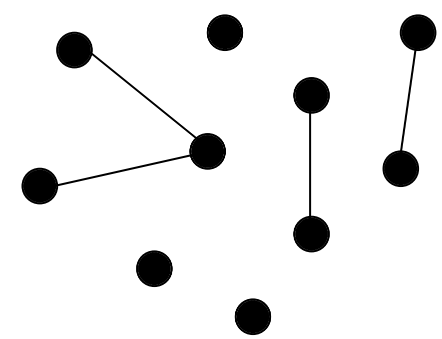
$\langle l \rangle \propto N^{1/2}$



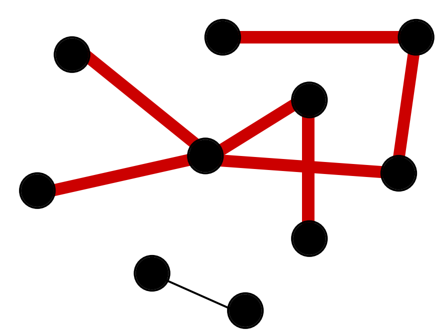
"E-R networks are infinite-dimensional"

Components in E-R networks

- When <*k*> is small, an E-R network consists of several disjoint components
- For a high enough value of <*k*>, a **giant connected component** appears
- The size of the giant component is of the order of network size, even when $N \rightarrow \infty$
- This transition from a fragmented to a connected phase is called the *percolation transition*



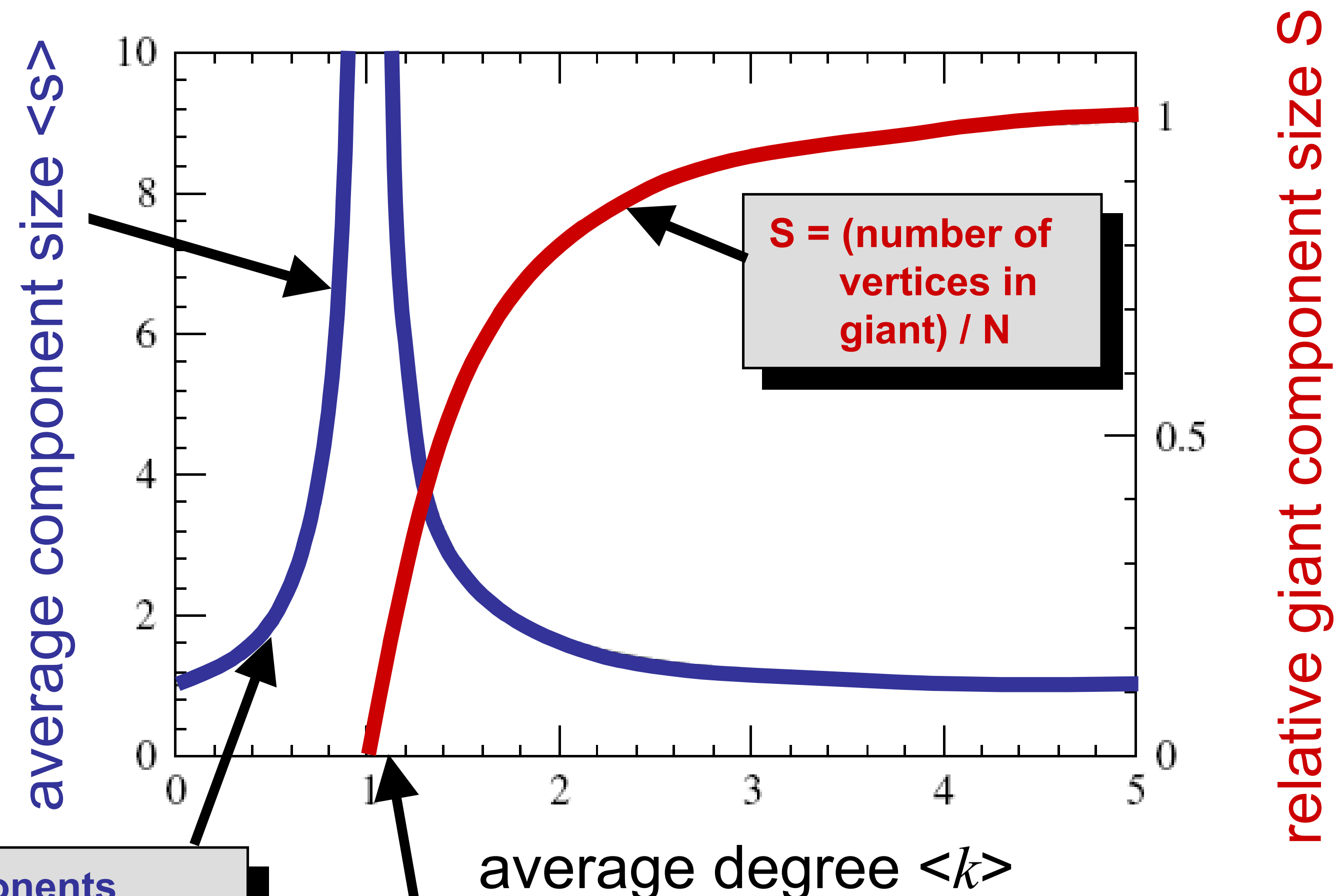
<*k*> small



<*k*> large

Connected component sizes

<s> = average number of vertices in components other than the giant



S = (number of vertices in giant) / N

small components grow in size until giant appears, then join the giant, leaving only very small disconnected parts

when <k>=1, giant component appears and starts growing in size ("the percolation transition")

this curve is, strictly speaking, valid only for ER networks where $N \to \infty$ so that pN=const

Erdös-Renyi Networks: Summary of features

- Degree distribution Poisson (degrees of all nodes close to average)
- Path lengths short, grow logarithmically with system size
- Connectivity depends on average degree - there is a percolation transition
- Overall: no correlations, all edges exist independently of each other, "as random as it gets"
- Very "homogeneous" networks