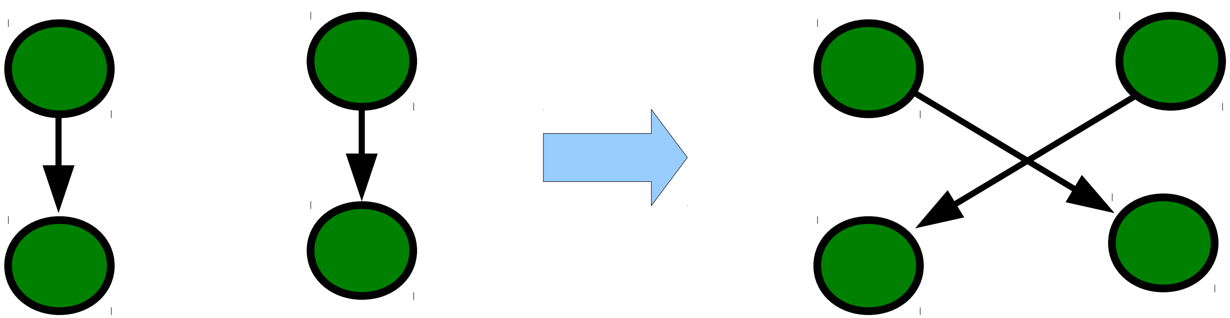
# Network Analysis and Models of Random Graphs

# Network Analysis – Physical Sciences

- Typical approach:
  - Measure properties of a network
  - Compare to some suitably chosen reference model and notice if deviations are significant
  - E.g.:
    - Compare given network to an Erdos-Renyi random graph
    - Compare given network to configuration model
    - Randomization techniques
  - Problems:
    - Findings can depend on the reference model if this is not well chosen (or it is unclear which reference model should be used)

# Network Randomization

- Aim: Generate ensemble of graphs that retain some property P but are otherwise random
- Apply local rewiring steps that preserve P (either exactly – “microcanonical” or on average – “canonical”)
- Example – rewiring that preserves degree sequence



# Network Randomization (2)

- There are problems with ergodicity (sequences of rewirings may not sample the ensemble appropriately)
- Q: Could you use this to generate random regular graphs?

  (i.e. graphs which are random but every node has the same degree)

# Network Analysis – (Quantitative) Social Sciences

- Typical approach:
  - Approach is often to “fit” to models of exponential random graphs
  - Philosophy like the derivation of ensembles in Stat. Phys.
    - Suppose we know one property of a graph (e.g. link density)
    - Construct and analyse ensemble of graphs that have this property on average (in the spirit of the canonical ensemble)
    - Add properties to graph Hamiltonian until network is described satisfactorily and use coefficients to classify it

# Exponential Random Graphs (1)

- We are interested in a probability distribution P over all graphs $\Gamma$
- Which matches some constraints $x_i$ on average, i.e.:

$$\sum_{\Gamma} P(\Gamma) = 1$$

$$\sum_{\Gamma} P(\Gamma) x_i(\Gamma) = \langle x_i \rangle$$

- This distribution should not have any bias other than meeting the given constraints
- A suitable choice maximises the Gibbs entropy

$$S = -\sum_{\Gamma} P(\Gamma) \ln P(\Gamma)$$

# Exponential Random Graphs (2)

- We are interested in a probability distribution P over all graphs $\Gamma$
- Which matches some constraints $x_i$ on average, i.e.:

$$\sum_{\Gamma} P(\Gamma) = 1$$

$$\sum_{\Gamma} P(\Gamma) x_i(\Gamma) = \langle x_i \rangle$$

- This distribution should not have any bias other than meeting the given constraints
- A suitable choice maximises the Gibbs entropy

$$S = -\sum_{\Gamma} P(\Gamma) \ln P(\Gamma)$$

# Exponential Random Graphs (3)

- Using the method of Lagrange multipliers

$$-\sum_\Gamma P(\Gamma)\ln P(\Gamma)-\alpha\left(1-\sum_\Gamma P(\Gamma)\right)-\sum_i \beta_i\left(\langle x_i\rangle-\sum_\Gamma P(\Gamma)x_i(\Gamma)\right)=max$$

- Differentiation with regard to P:

$$\ln P(\Gamma)-1+\alpha+\sum_i \beta_i x_i(\Gamma)=0 \longrightarrow P(\Gamma)=\exp\left(-1+\alpha+\sum_i \beta_i x_i(\Gamma)\right)$$

- Or:

$$P(\Gamma)=\frac{1}{Z}\exp(-\beta H(\Gamma))$$

$$Z=\sum_\Gamma \exp(H(\Gamma))$$ "partition function", fixed by normalisation

$$H(\Gamma)=\sum_i \beta_i x_i(\Gamma)$$ "Graph Hamiltonian"

- Can then calculate other quantities y of interest

$$\langle y\rangle=\frac{1}{Z}\sum_\Gamma \exp(H(\Gamma))\,y(\Gamma)$$ ("best estimate of y given x")

# An Example

- Let's say we only know the average number of edges m in a graph
- In terms of the adjacency matrix: $m = \sum_{i<j} a_{ij}$
- Our ERG model is then:

$$Z = \sum_{\{a_{ij}\}} \exp\left(\beta \sum_{i<j} a_{ij}\right)$$

$$= \sum_{\{a_{ij}\}} \prod_{i<j} \exp(\beta a_{ij}) = \prod_{i<j} \sum_{a_{ij}=0,1} \exp(\beta a_{ij}) = (1+\exp(\beta))^{\binom{n}{2}}$$

- To calculate network averages ...

$$\langle m \rangle = \frac{1}{Z} \sum_{\Gamma} m \exp(\beta m) = \frac{1}{Z} \frac{\partial Z}{\partial \beta} = \frac{\partial lnZ}{\partial \beta}$$

# An Example (2)

- "free energy" $\ln Z = \binom{n}{2} \ln\left(1+e^{\beta}\right)$

$\longrightarrow$ $\langle m\rangle = \frac{\partial \ln Z}{\partial \beta} = \binom{n}{2} \frac{1}{1+\exp(-\beta)}$

- This allows to determine the Lagrange multiplier, i.e.: $\beta = -\ln \frac{1}{\langle m\rangle / \binom{n}{2} - 1}$

- Could now calculate an avg. connection prob.

$$p_{ij} = \frac{1}{Z} \sum_{\{a_{kl}\}} a_{ij} \exp\left(\beta \sum_{k<l} a_{kl}\right) = \frac{\sum_{a_{kl}=0,1} a_{kl} \exp\left(\beta a_{kl}\right)}{\sum_{a_{kl}=0,1} \exp\left(\beta a_{kl}\right)} = \frac{\exp(\beta)}{1+\exp(\beta)} = \langle m\rangle / \binom{n}{2}$$

(this recovers the Erdos-Renyi model of random graphs!)

# Exponential Random Graphs

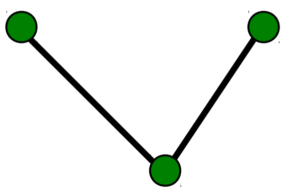
- Not many analytical results yet (and probably difficult to obtain in many cases)
- Numerical implementation through Monte-Carlo techniques, e.g.
  - Start with a random graph
  - Suggest "moves" (rewiring/addition, deletion of links)
  - Accept moves with probability $\propto \exp(\beta(H(\Gamma')-H(\Gamma)))$
  - Repeat until equilibrium … average.

# Exponential Random Graphs

- “Fitting” procedure:
  - Start with H=beta m and then add local/global graph properties, e.g.

$$H = \beta \sum_{i<j} a_{ij} + \gamma \sum_{i,j,k} a_{ij} a_{ik}$$

“2 stars”



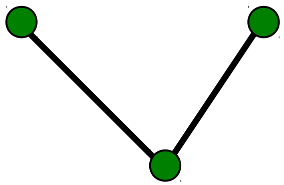
- Drawbacks:
  - There are many aspects of these models that are not well understood yet
  - Many of these models exhibit phase transitions (i.e. “fitting” becomes ambiguous
  - Difficult to handle for large systems

# Exponential Random Graphs

- “Fitting” procedure:
  - Start with H=beta m and then add local/global graph properties, e.g.

$$H = \beta \sum_{i<j} a_{ij} + \gamma \sum_{i,j,k} a_{ij} a_{ik}$$

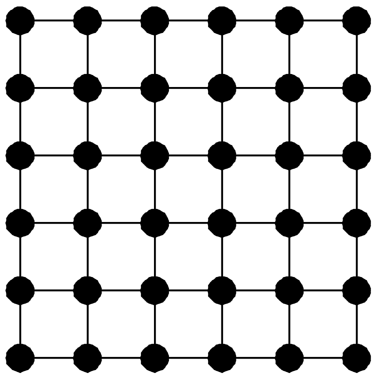
“2 stars”



- Drawbacks:
  - There are many aspects of these models that are not well understood yet
  - Many of these models exhibit phase transitions (i.e. “fitting” becomes ambiguous
  - Difficult to handle for large systems

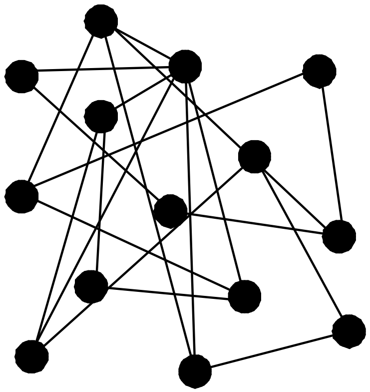
# Some prominent network types

- Spatial grids



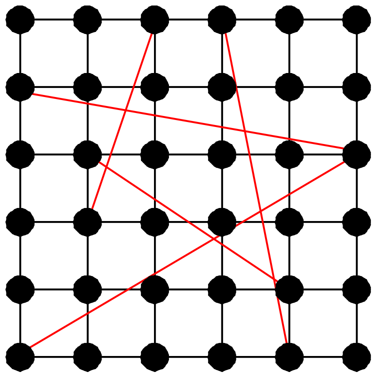
  - regular
  - large
  - cliquish

- Random graphs



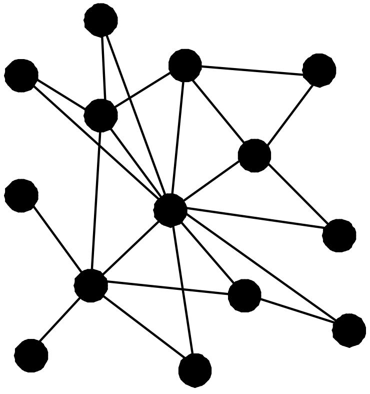
  - ~regular
  - small
  - not cliquish

- Small worlds



  - ~ regular
  - small
  - cliquish

- Scale-free networks



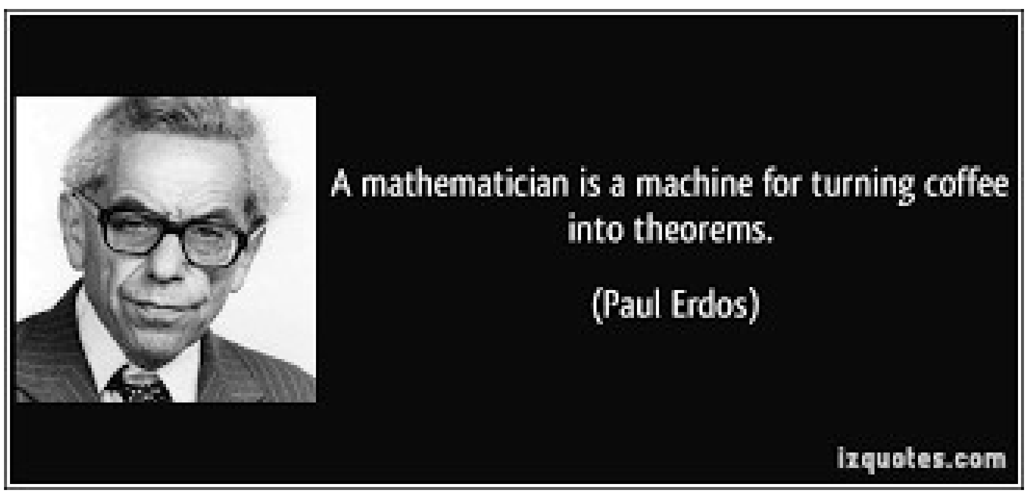
  - v. hetero-geneous
  - ultrasmall

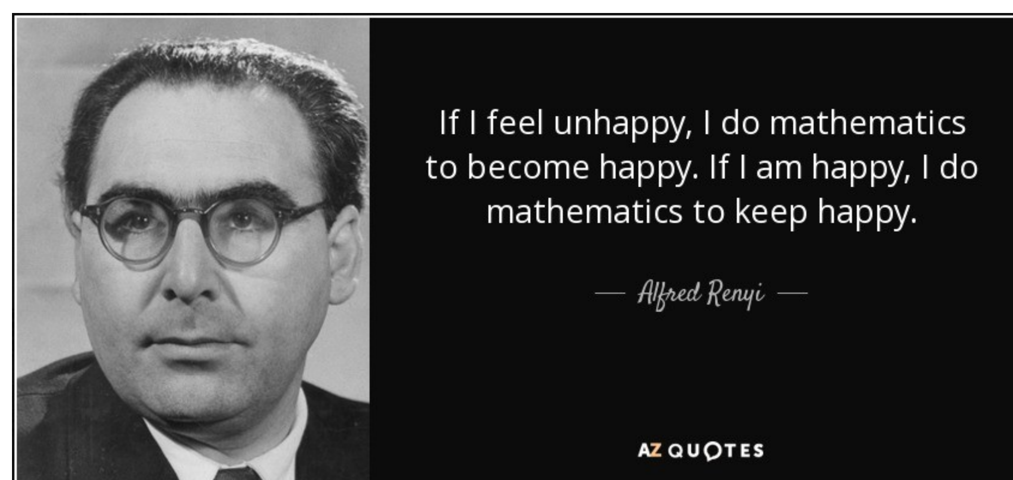
# Network Models

- (Erdos Renyi) Random Graphs
  - The configuration model
  - Random geometric graphs
- The small world model
- Scale-free networks
  - Preferential attachment
  - Vertex copying
  - Optimization
  - Appolonian networks, planarity preservation

# Erdos-Renyi Random Graphs

- First studied by Solomonoff and Rapaport, but named after Paul Erdos and Alfred Renyi to honour their contributions in the 1950s and 60s
- One of the best studied graph models



# ER Random Graphs (2)

- Want to have networks for which some parameters have certain values, but that are otherwise random
- Simplest idea is to specify number of nodes N and edges L, this gives the **ensemble** G(N,L)
- (Strictly speaking we define a pdf over the set of all graphs and set P(G)=1/K for all graphs with N nodes and L edges)
- For network measures we then consider averages over the ensemble, e.g. for the diameter

$$\langle D \rangle = \sum_G P(G) D(G) = 1/K \sum_G D(G)$$

# ER Random Graphs (3)

- More tractable model is G(N,p) – consider N nodes and connect each pair with prob. p

$p = 0.1$ $\qquad$ $p = 0.25$ $\qquad$ $p = 0.5$

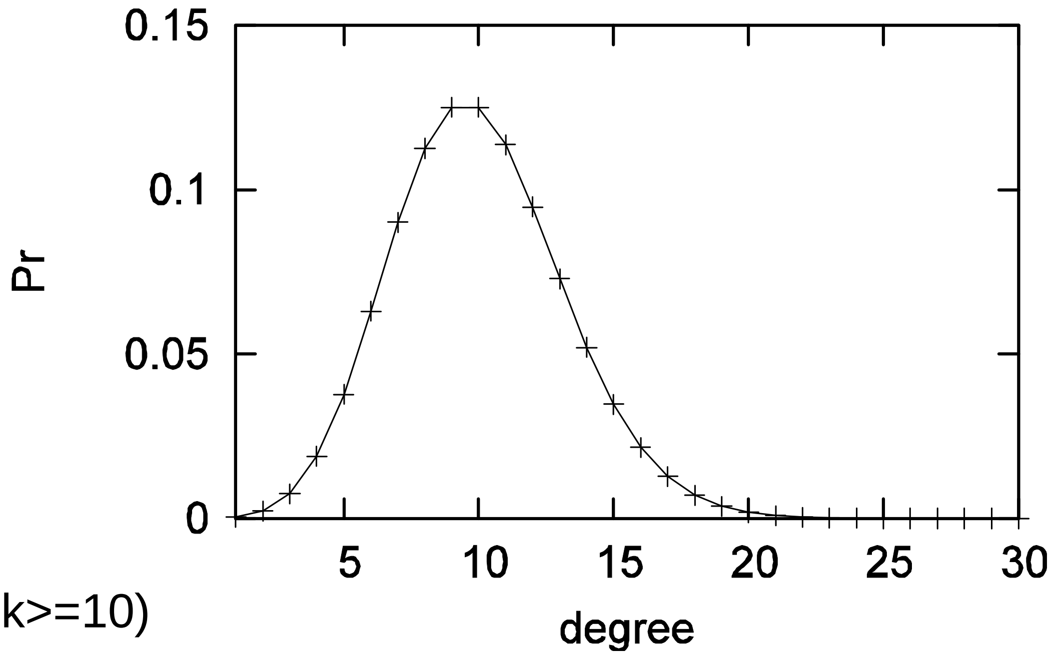
- Each simple graph G with L edges appears with

$$P(G) = p^L (1-p)^{\binom{N}{2} - L} \longrightarrow P(L) = \binom{\binom{N}{2}}{L} p^L (1-p)^{\binom{N}{2} - L}$$

$$\longrightarrow \langle L \rangle = \sum_{L=0}^{\binom{N}{2}} L P(L) = N(N-1)/2\, p \qquad \langle k \rangle = p(N-1)$$

# ER Random Graphs (4)

- Degree Distribution




(<k>=10)

$$Pr\{d(v)=n\}=\binom{N}{n}p^n(1-p)^{N-n}$$

- strongly concentrated around mean
- no vertices with degree much higher than mean

- Sometimes the limit N >>1 and <k>=const. is of interest

$$Pr\{d(v)=n\} \rightarrow e^{-\langle k\rangle}\langle k\rangle^n/n!$$

- Clustering coefficient: $C=p=\frac{\langle k\rangle}{N-1}\rightarrow 0 \; for \; N\rightarrow\infty$

# ER Random Graphs (5)

- Threshold property: property Q, such that for some $p_Q$
  for $p>p_Q$ a.e. r.g. has Q while for
  $p<p_Q$ a. no r.g. has Q
- Giant component: $p_{gc}=1/(N-1)$

| $p>p_{gc}$ | $p<p_{gc}$ |
|---|---|
| - Giant component with $N^{2/3}$ vertices | - every component O(log N) |
| - Loops of every order in giant comp. | - almost no loops, |
| - G\{GC} still tree-like | essentially tree-like |

- Connectedness: $p_{conn}= \ln N/(N-1)$
- K-connectedness: $p_{conn}^{(k)}= (\ln N + k \ln \ln N)/(N-1)$

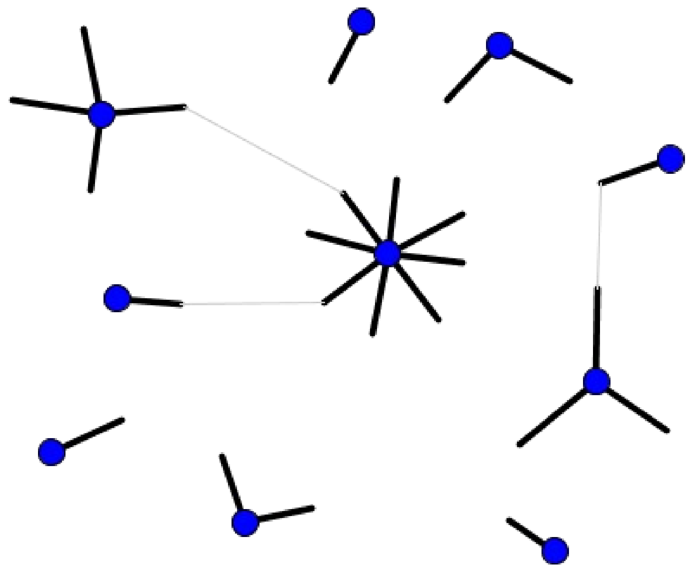
(graph k-connected iff for every pair of vertices i,j there are at least k independent paths from i to j)

# ER Random Graphs (6)

- Shortcomings (as models of real world Nws)
  - No clustering/transitivity
  - Degree distribution is very “narrow” – no heterogeneity
  - No correlations between adjacent degrees

# The Configuration Model

- Random graph with a given degree sequence
- “Stub” matching

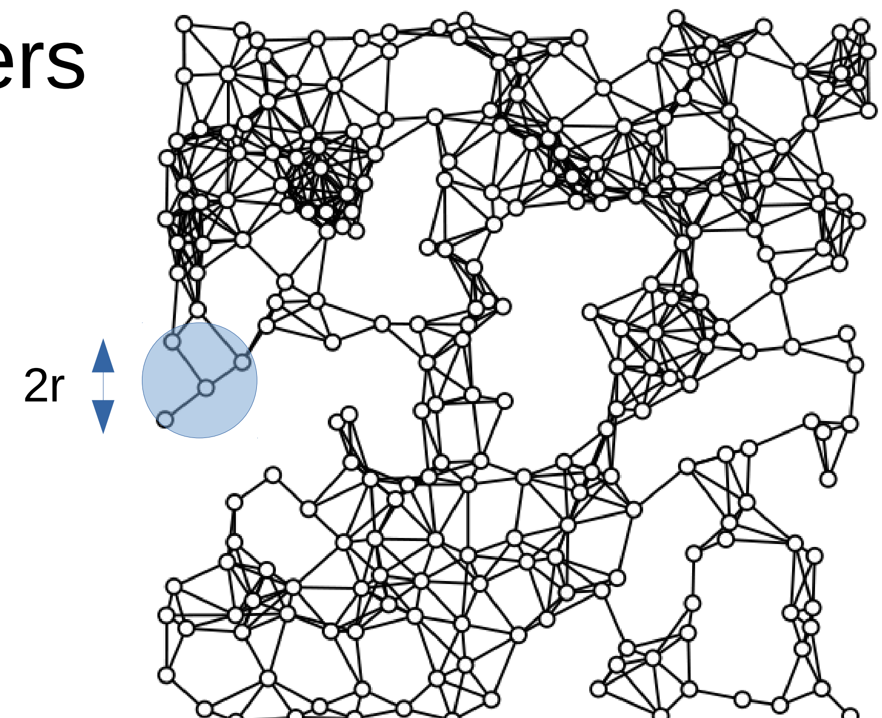


- Potential problems:
  - Odd number of edges?
  - Self-edges?

- Can answer questions whether certain network properties are artifacts of a given degree sequence

# Random Geometric Graphs

- Simplest model of a spatial random graph
- Place nodes randomly in some d dimensional space, connect nodes to all other nodes within some range r
- Avg. distances and diameters scales as on grids
- Narrow degree distribution
- High clustering coefficients




256 nodes on [0,1]*[0,1] with r=0.1 (wikipedia)