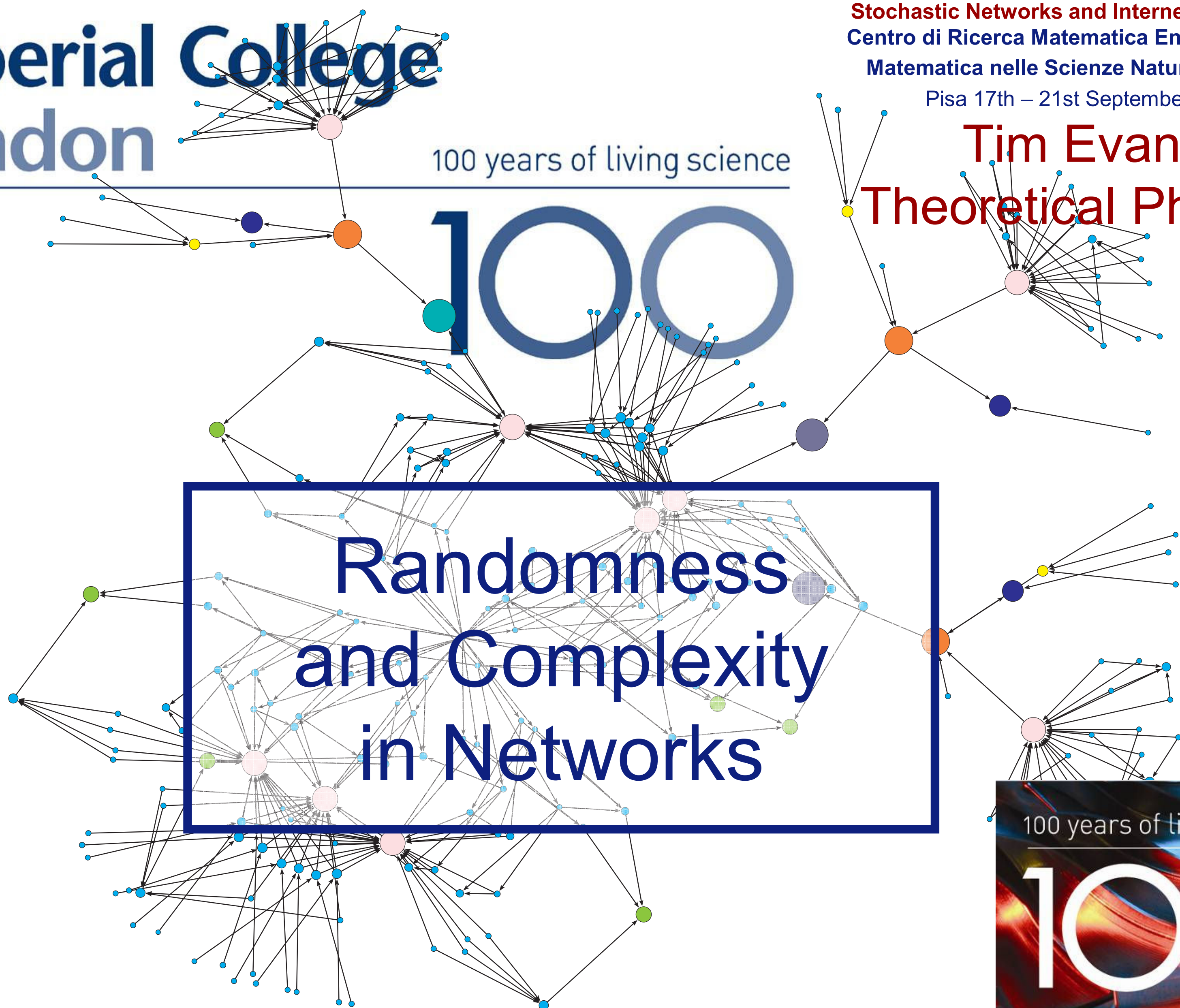


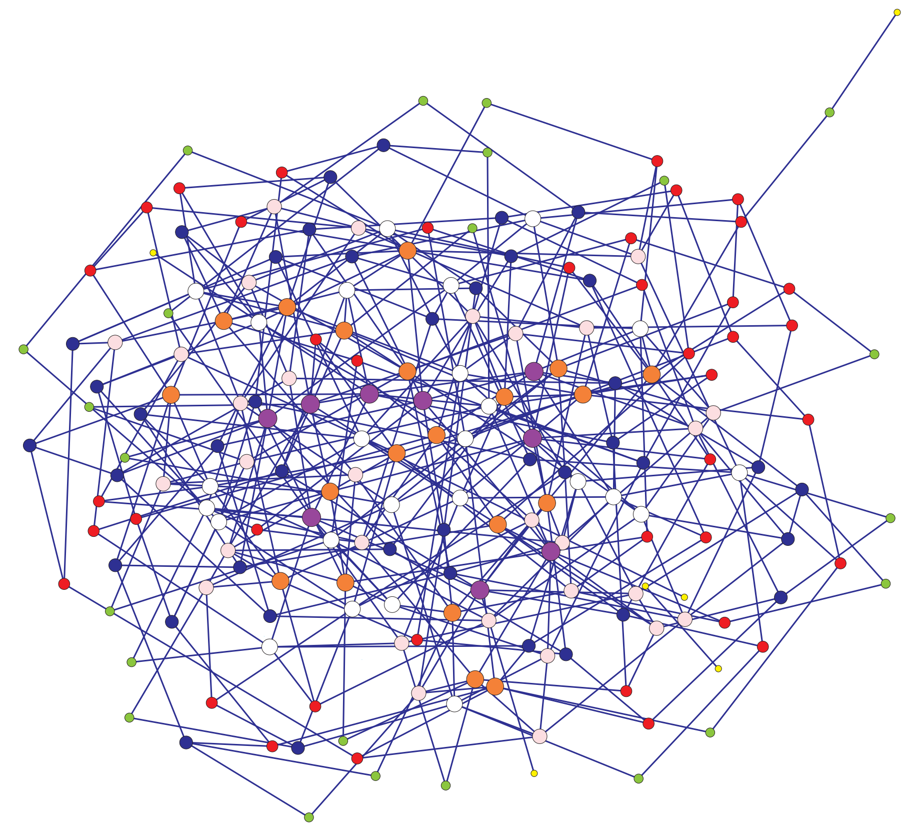
Workshop on
Stochastic Networks and Internet Technology
Centro di Ricerca Matematica Ennio De Giorgi.
Matematica nelle Scienze Naturali e Sociali
Pisa 17th – 21st September 2007

Tim Evans
Theoretical Physics

Randomness and Complexity in Networks

Random Graphs

- Random Graphs
- Random Walks
- Random Walks and Copying - The Origin of Scale-Free Networks?
- Copying and Culture
- Summary



- o **Generalised Random Graphs**
- o **Properties of Random Graphs**

Notation

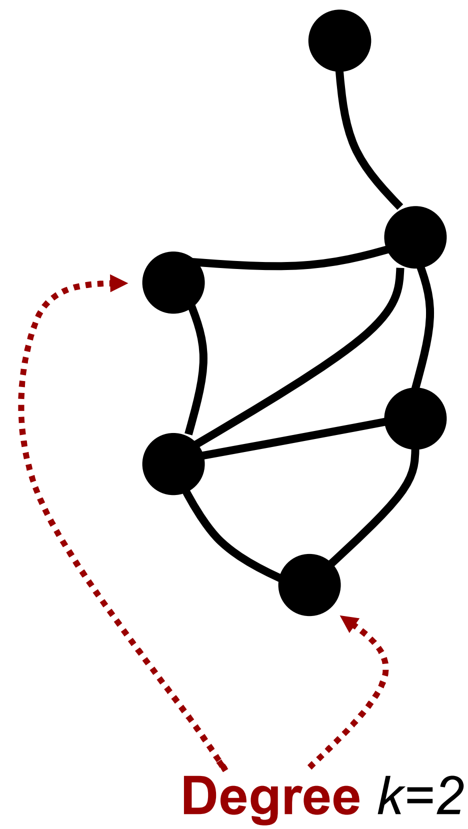
I will focus on **Simple Graphs** with multiple edges allowed

(no values or directions on edges, no values for vertices)

- ***N*** = *number of vertices in graph*
- ***E*** = *number of edges in graph*
- ***k*** = *degree of a vertex*
- **$\langle k \rangle$** = *average degree* = ***2E/N***
- Degree Distribution

  n(k) = *number of vertices with degree* ***k***

  p(k) = ***n(k)/N*** = *normalised distribution*



Classical or Erdős-Réyni Random Graphs

Either

a) For every pair of distinct vertices add a single edge with probability $p = E/N$, otherwise no edge is added

b) Add E vertices between randomly chosen vertex pairs

No difference for large N when sparse

$$2E/N = \langle k \rangle \sim O(1)$$

Similar to canonical vs microcanonical ensembles

Poisson degree distribution, exponential cutoff

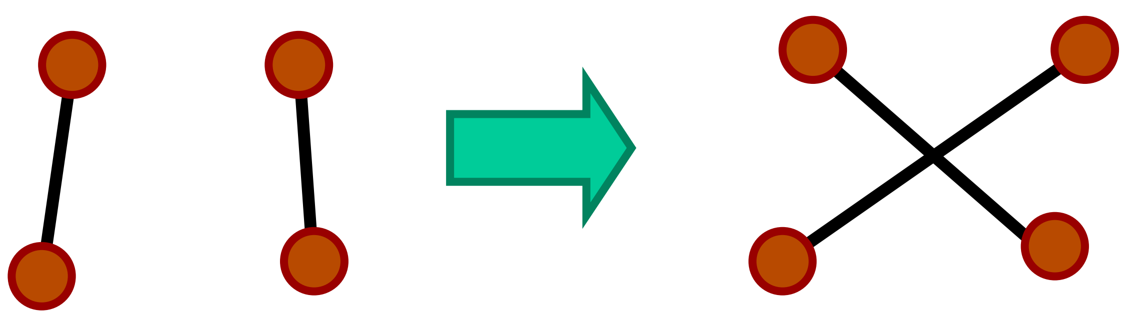
Generalised Random Graphs – The **Molloy-Reed** Construction [1995,1998]

i. Fix N vertices

ii. Attach k stubs to each vertex, where k is drawn from given distribution $p(k)$

iii. Connect pairs of stubs chosen at random

Or use **Maslov-Sneppen** [2002] rewiring to randomise graph of given $p(k)$



No Vertex-Vertex Correlations

Generalised Random Graphs have given ***p(k)*** but otherwise completely random in particular -

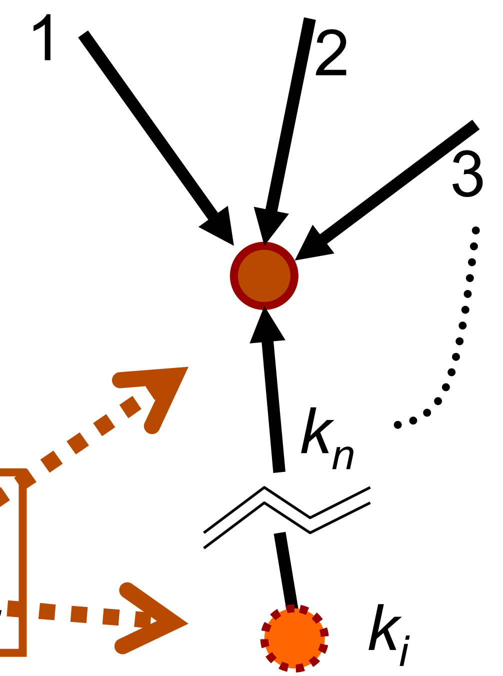
Properties of all vertices are the same

For any given source vertex, the properties of neighbouring vertices independent of properties of the source vertex

Random Walks on MR Graphs

The degree distribution of a neighbour is not simply p(k) because the higher the degree of a vertex the more likely you are to arrive at that vertex given there is no correlation between vertices

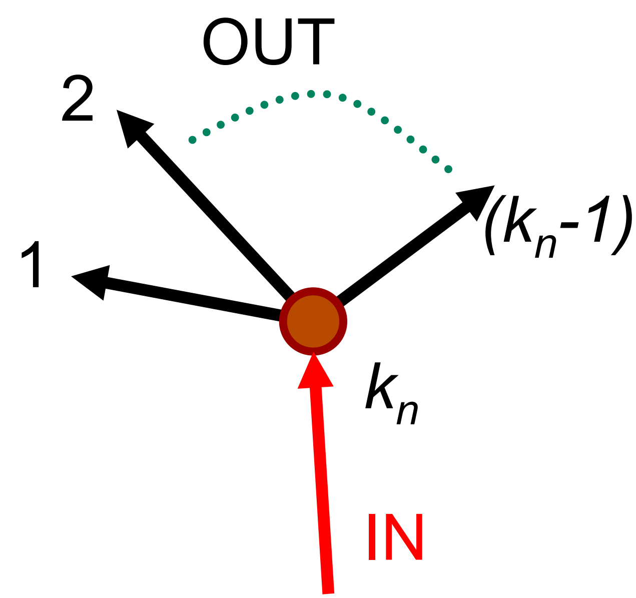
$$p(k_n \mid k_i) = \frac{k_n}{\langle k \rangle} p(k_n)$$



Degree of neighbour k_n independent of degree of starting point k_i

Length of Random Walks on Random Graphs

Suppose we follow a random walk where we never go back along the edge we just arrived on, then for infinite graphs ***(N*** $\to \infty$***)***



$\Rightarrow$ Walks always end if

$\langle \boldsymbol{k_n} \rangle < 2 \Leftrightarrow$ No GCC

$\Rightarrow$ Walks never end if

$\langle \boldsymbol{k_n} \rangle > 2 \Leftrightarrow$ GCC

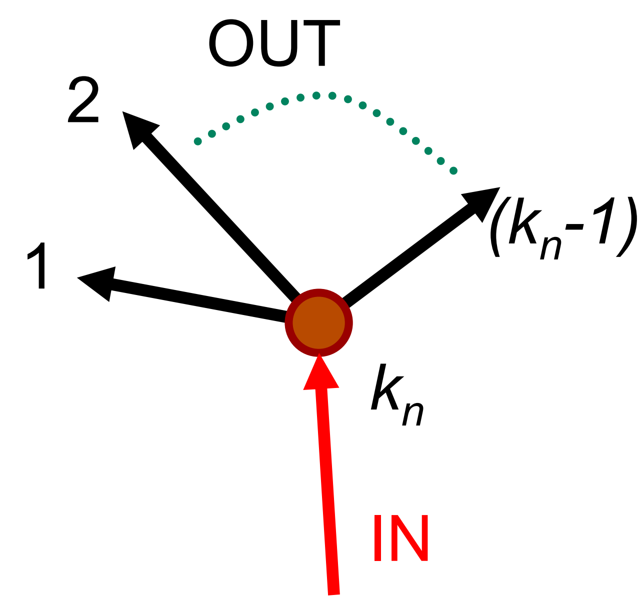
(GCC= Giant Connected Component)

GCC transition

Transition to GCC is at ***z=1*** where

$$z = \left\langle \frac{k_n}{\langle k \rangle} \right\rangle - 1 = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1$$

(GCC= Giant Connected Component)



Other properties of General Random Graphs

All global properties depend on same

$$z = (\langle k^2 \rangle / \langle k \rangle) - 1$$

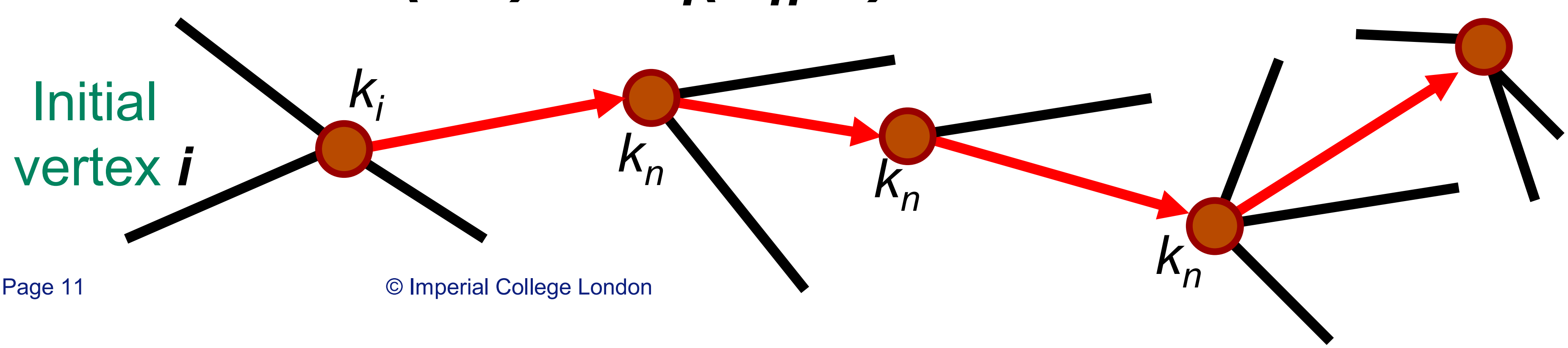
e.g. GCC size,
component distribution,
average path lengths

Average Path Length in MR Random Graph

[Fronczak et al, 2005]

- Let $p_{ij}(x)$ be the probability that a random walk (never returning along last step) starting at vertex i passes through vertex j at least once after x steps
- Number of different walks of length x from i to j *if no loops* is

$$W(i,x) = k_i(k_n-1)^{x-1}$$



Average Path Length in MR Random Graph (2)

- Probability of *not* arriving at ***j*** on any one step = 1- ($\boldsymbol{k_j / 2E}$)

$\Rightarrow$ Probability that a random walk does not arrive at ***j*** after ***x*** steps is

$$p_{ij}(x) = \left(1 - \frac{k_j}{2E}\right)^{W(i,x)} \approx \exp\left\{-\frac{k_i k_j}{2E} z^{x-1}\right\}$$

$$\left(z := \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 = \frac{\langle k_n \rangle}{\langle k \rangle} - 1 \right)$$

Average Path Length in MR Random Graph (3)

- Probability that walker first arrives after x steps is $p_{ij}(x-1) - p_{ij}(x)$

$\Rightarrow$ Average path length from ***i*** to ***j*** is

$$\ell_{ij} = \sum_{x=1} x\left[p_{ij}(x-1) - p_{ij}(x)\right] = \sum_{x=0} p_{ij}(x)$$

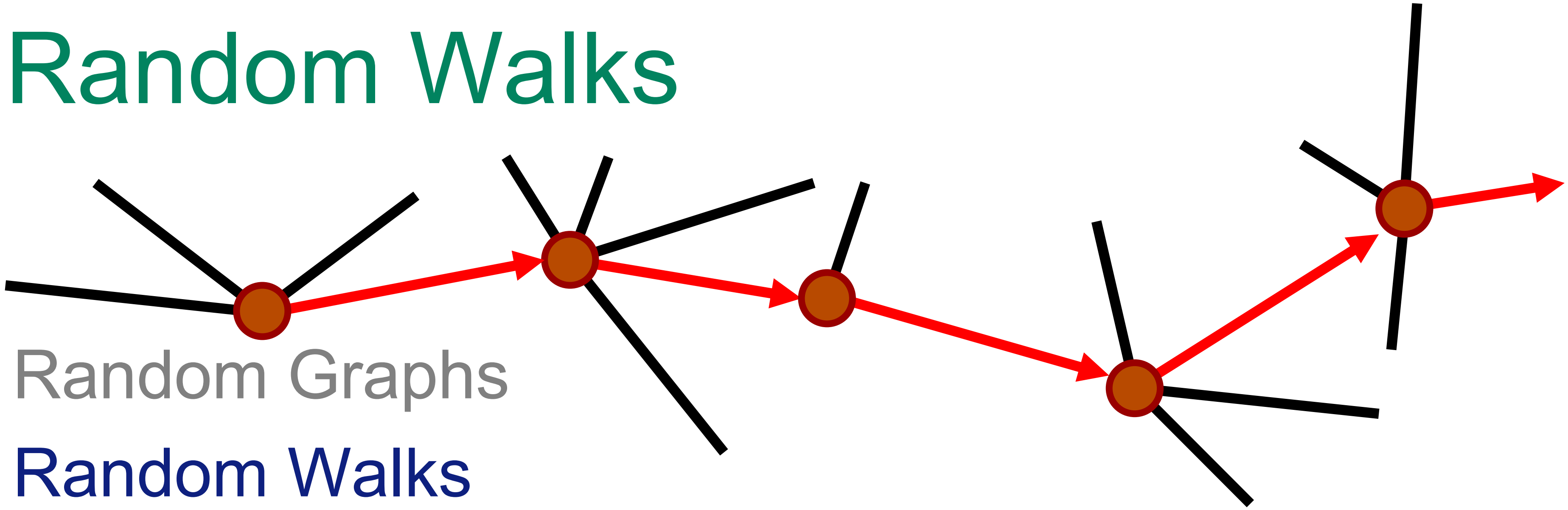
$\Rightarrow$ Average path length

$$\langle \ell \rangle \approx \frac{\ln(N) + \ln(z) - \ln(\langle k \rangle) - \gamma}{\ln(z)} + \frac{1}{2}$$

MR Random Graph Calculations

- Calculations like those above work because
  - lack of correlations between vertices
  - few loops for large sparse graphs, graphs are basically trees
- These can be reasonable approximations for many models and perhaps for a few real graphs too. Otherwise use as a ***null model***.
- Often these calculations are exact for closely related ***Urn models*** ***(see later discussions)***

Random Walks



- Random Graphs
- **Random Walks**
- Random Walks and Copying - The Origin of Scale-Free Networks?
- Copying and Culture
- Summary

  - o Uses
  - o Diffusion
  - o Origin of Scale-Free Graphs

Random Walks

- Random Walks are the extreme alternative to the use of Shortest Paths
  - calculate mean first passage time etc.
- Related to diffusion process and eigenvalues/vectors of Laplacian
- Used for
  - Calculations of Generalised Random Graph
  - Sampling graphs
  - Community detection
  - Natural Creation of Scale-Free Networks

Random Walks as a SearchTool

Sample Networks via Random Walk

$\Rightarrow$ visit vertices with probability roughly

$$p_{visit}(k) \approx k\, p(k) / (2E)$$

$\Rightarrow$ find Hubs (large k nodes) very quickly

$\Rightarrow$ Estimate tail of degree distribution very quickly

$\Rightarrow$ Estimates of size of graph possible

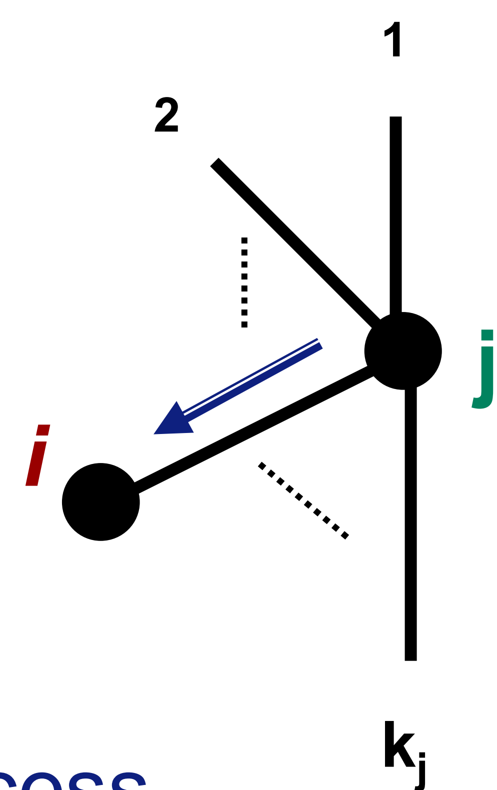
- Other biased walks possible but these do not share the same special properties
  e.g. can sample vertices equally if slowly
  [Orponen & Schaeffer, 2004; Sood & Grassberger, 2007]

Random Walks as Diffusion

Adjacency matrix A_{ij} **= 1 if edge from i to j, = 0 otherwise**

Probability of going from vertex j to vertex i

$$M_{ij} = \frac{A_{ij}}{k_j}$$



Number of random walkers at vertex i at time t $v_i(t)$

⇒ Solve Matrix Equation

$\mathbf{v}(t) = [\mathbf{M}]^t \, \mathbf{v}(0)$ Markov process

Simple Graph Diffusion Solution

$$v_i(t) = c_1\left(\frac{k_i}{2E}\right) + \sum_{n=2}^{N} c_n(\lambda_n)^t u_i^{(n)}$$

First Eigenvector

$$u_i^{(n=1)} = \left(\frac{k_i}{2E}\right)$$

Eigenvectors and eigenvalues of P

$$1 = \lambda_1 > |\lambda_2| \geq \ldots \geq |\lambda_j| \geq \ldots$$

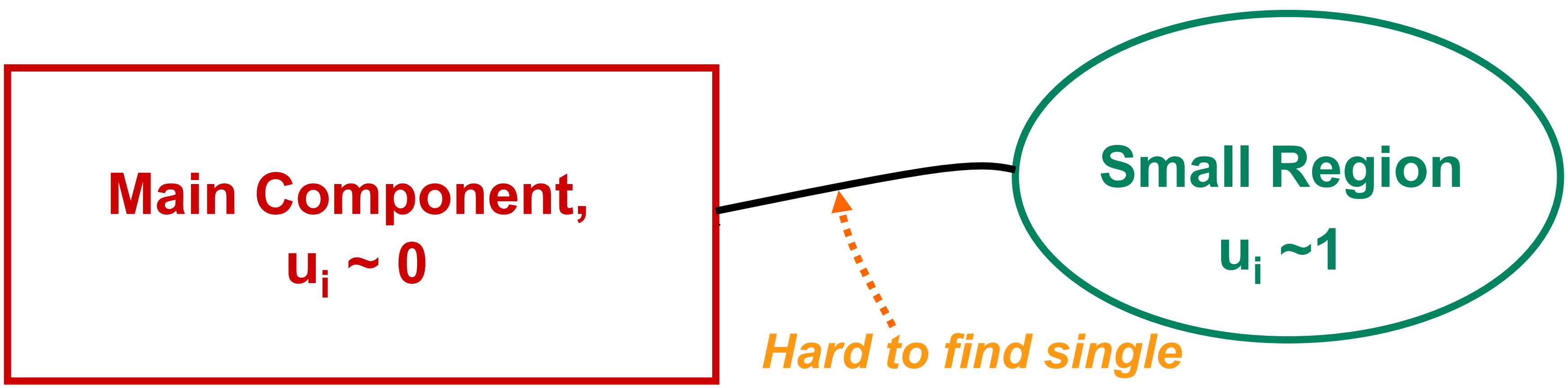
First eigenvalue = 1

Simple Graph Diffusion Solution

$$v_i(t) = c_1\left(\frac{k_i}{2E}\right) + c_2(\lambda_2)^t u_i^{(2)} + \ldots$$

$$1 = \lambda_1 > |\lambda_2| \geq \ldots$$

Eigenvectors $u^{(n)}$ of largest eigenvalues tell us about small regions poorly connected to main component - Community Detection (Eriksen et al 2003)



Spectral Analysis

- Eigenvalues/vectors of Markovian $\boldsymbol{M}$ matrix and Laplacian $L_{ij} = k_i\delta_{ij} - A_{ij}$ simply related
- Adjacency matrix eigenvectors and eigenvalues have no obvious physical interpretation or relationship to Markovian $\boldsymbol{M}$ matrix
- PCA (Principle Component Analysis)
- SVD (Singular Value Decomposition)

The idea is always to use a few eigenvectors to reduce the dimensionality of the graph from $O(N^2)$ to $O(N)$

Diffusion as Ranking

- Long time solution gives a ranking of vertices

  Rank of vertex ***i*** = entry ***i*** of eigenvector of largest eigenvalue $\boldsymbol{u^{(1)}_i}$

e.g. PageRank® (Google)

Jump to random vertex with probability p_v or use other constant vector

$$M_{ij} = (1 - p_v)\frac{A_{ij}}{k_j^{(out)}} + p_v \frac{1}{N}$$

$$\frac{k_i^{(in)}}{E} \; ?$$

Why is a random walk so useful?

Random Walk probes ***global*** structure of network but uses only ***local*** information

A process involving only local information is much more likely to occur naturally i.e. no external influence needed

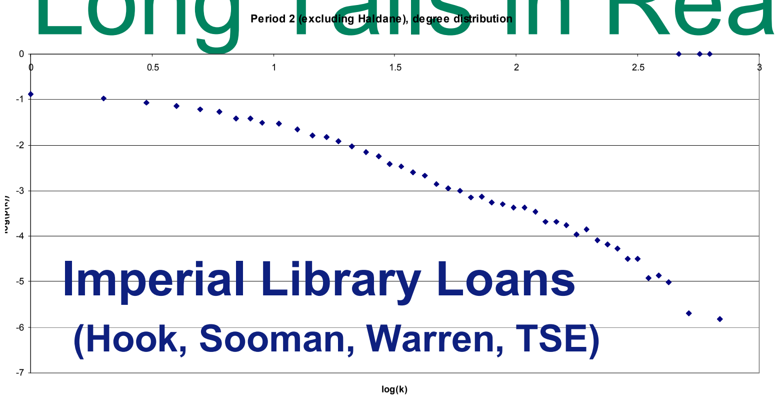
$\Rightarrow$New question…

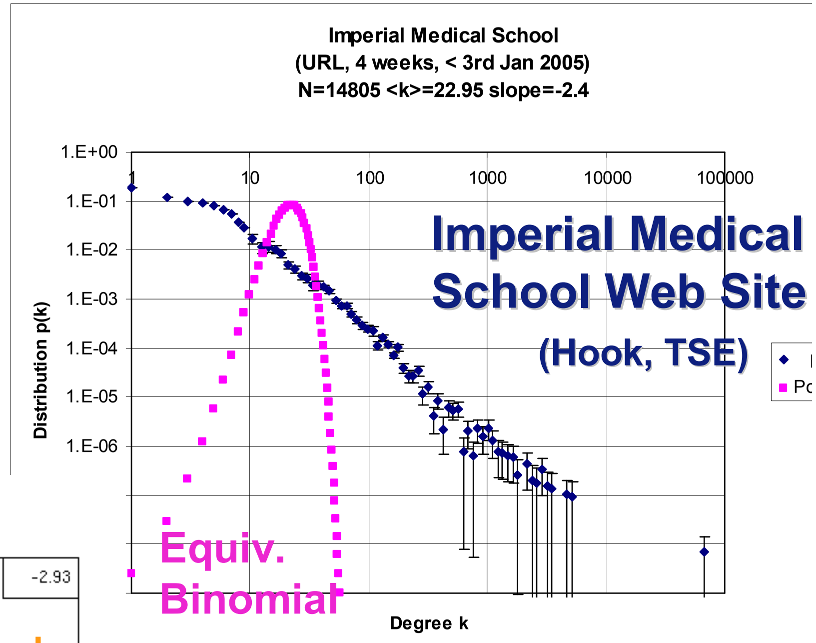
Can this explain the long tails in p(k) of many networks?

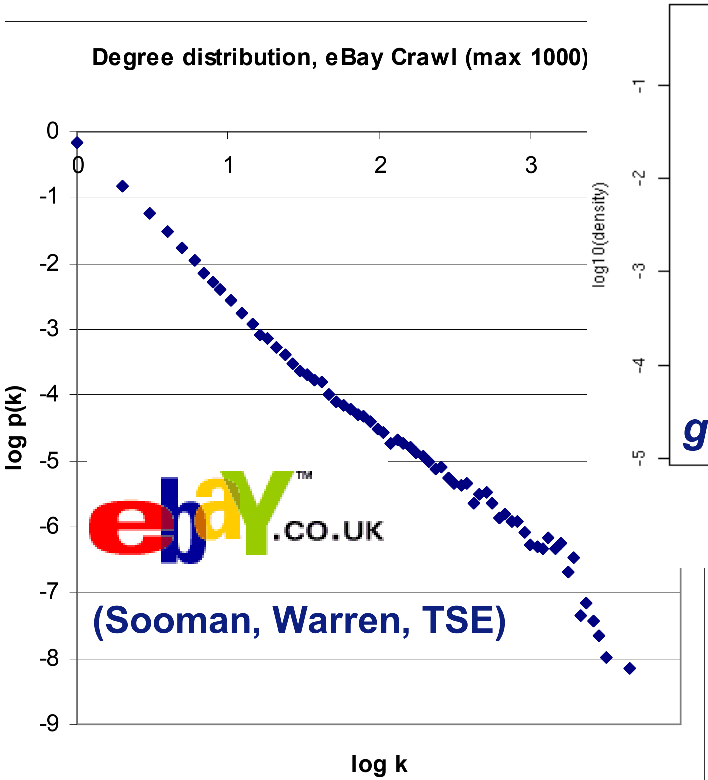
Do power laws emerge naturally?

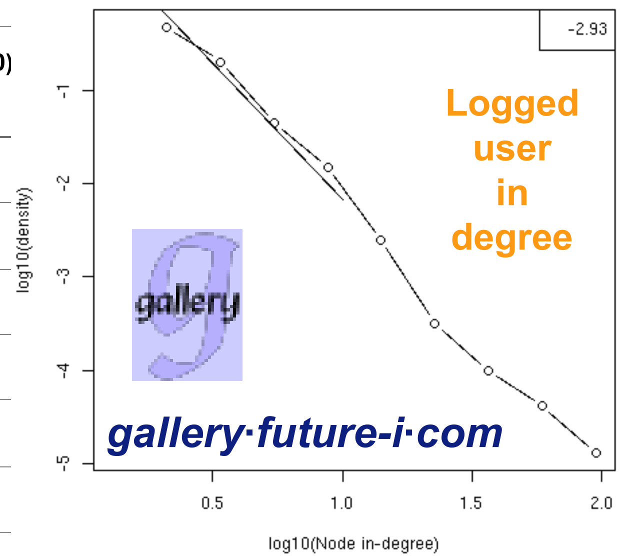
… Self Organised Criticality?

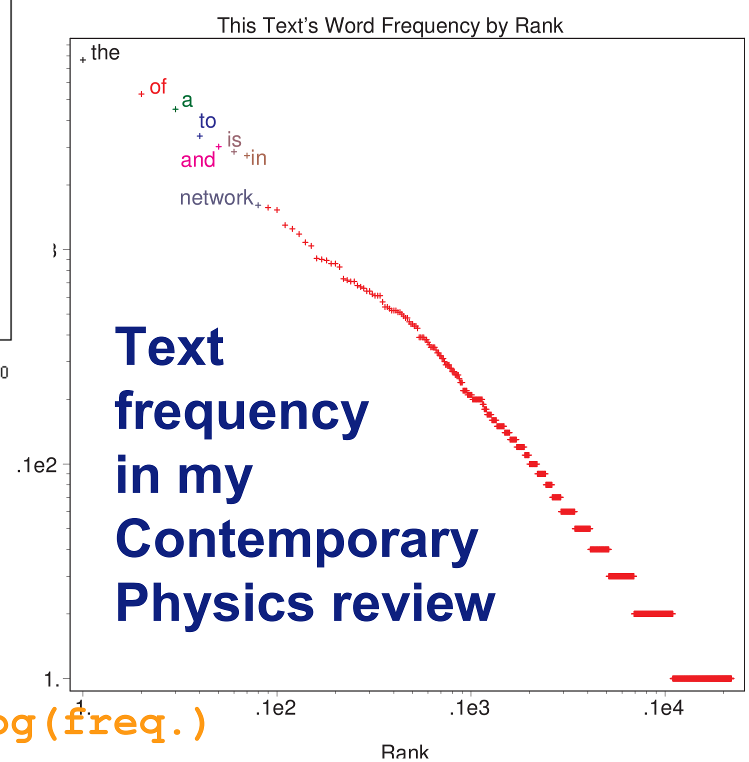
Long Tails in Real Data











ege London

All log(k) vs. log(p(k)) except text log(rank) vs. log(freq.)

Long Tails = Hubs

k_1 = Largest *Degree*

Hubs are vertices of high degree

- Lattices, WS Small World, random networks have no hubs,

  k O k_1 O $O(ln(N))$

  Classical Rnd $N=10^6$, $<k>=4$
  $\Rightarrow k_1 = 17$

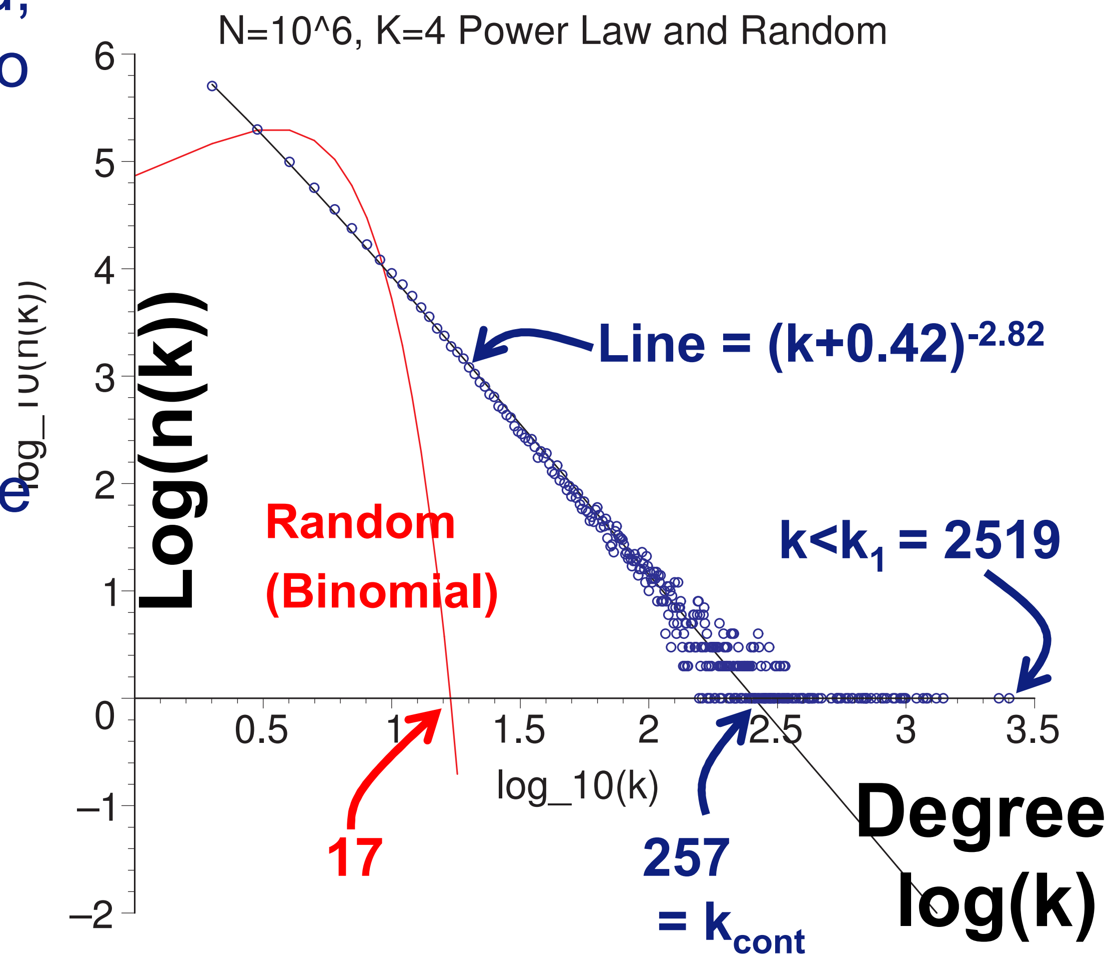
- Only a **long tailed** degree distribution has hubs

e.g. POWER LAW

$n(k) \sim 1/k^3$

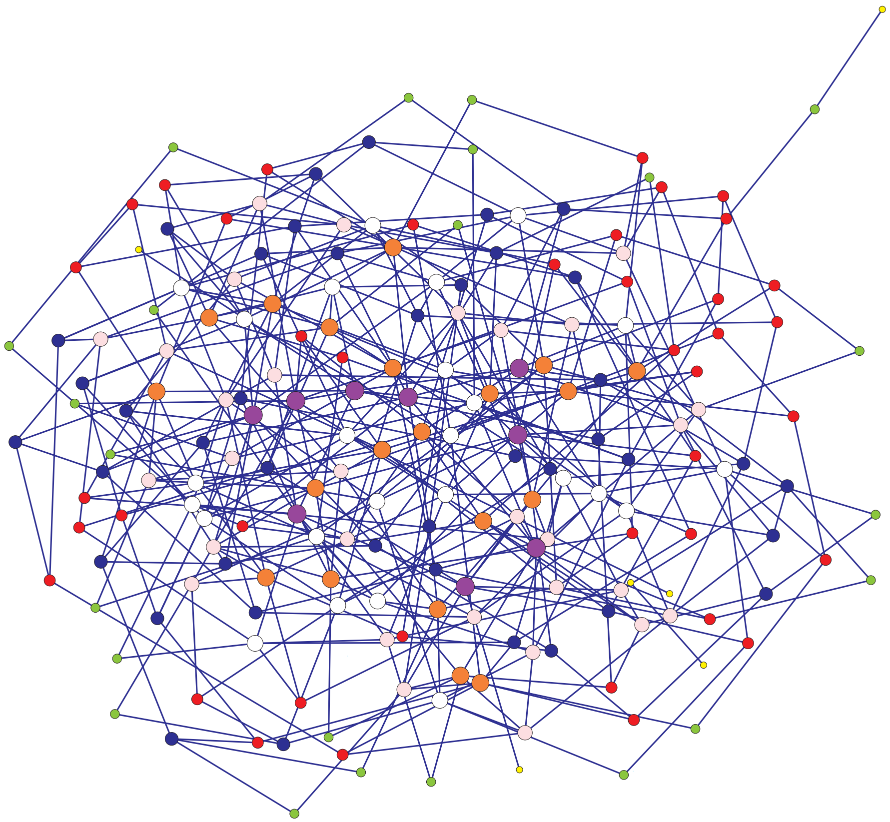
k O $k_1 = O(N^{1/2})$

has $N=10^6$, $<k>=4 \Rightarrow k_1 \sim 2520$



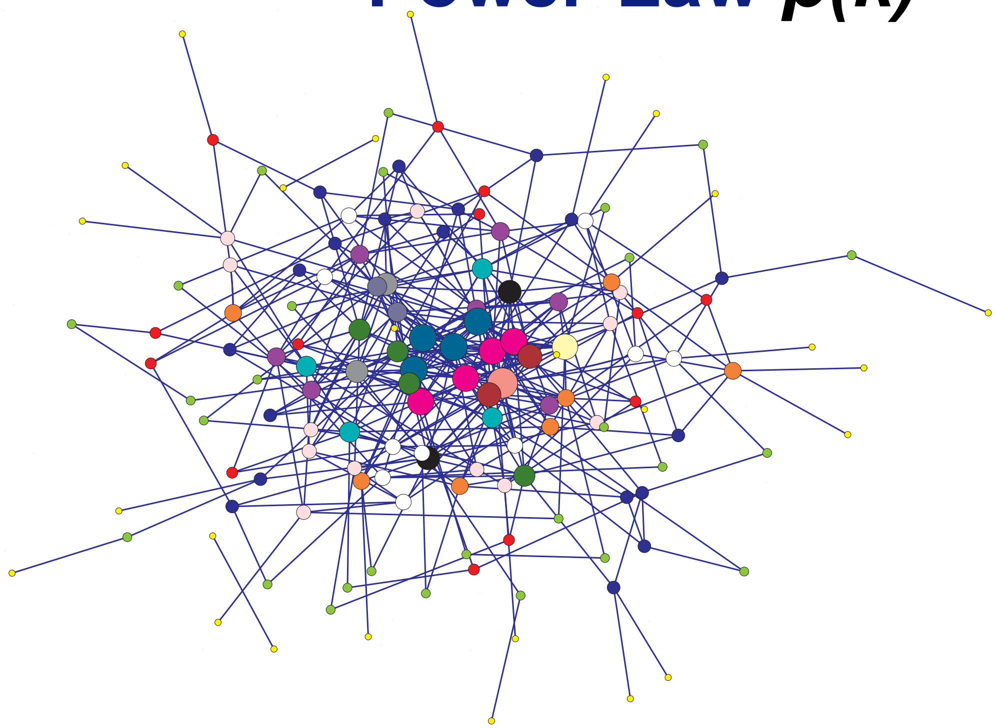
N=200, $<k>\sim 4.0$, vertex size ✪ k

Classical Random



Diffuse centre of small degree vertices

Scale-Free = Power-Law $p(k)$



Tight core of large hubs

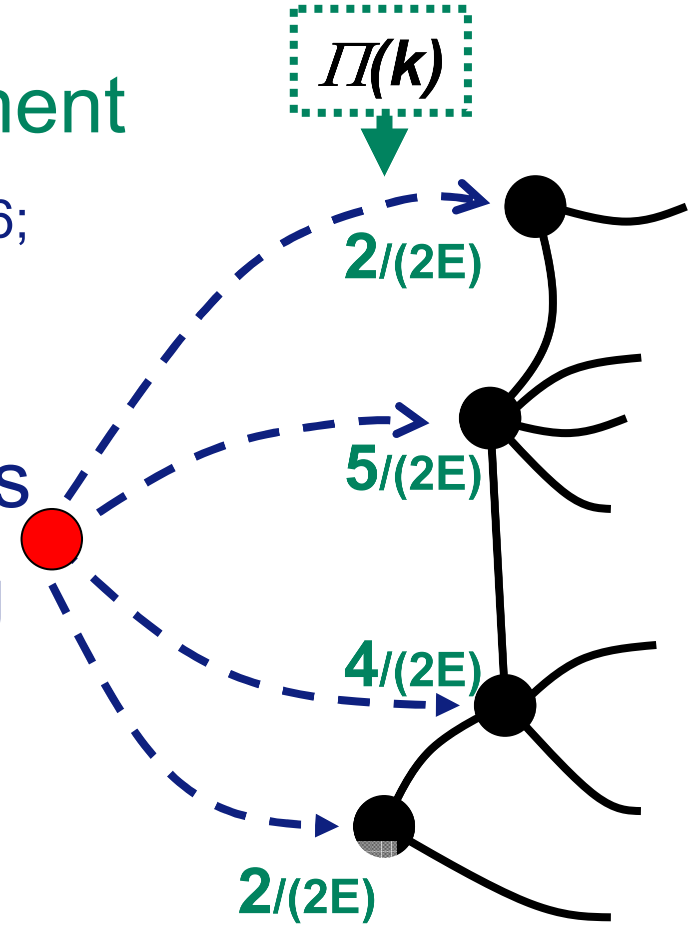
Growth with Preferential Attachment

(Yule 1925, 1944; Simon 1955; Price 1965,1976; Barabasi,Albert 1999)

1. Add new vertex attached to one end of **½<*k*>** new edges
2. Attach other ends to existing vertices chosen with by picking random end of an existing edge chosen randomly, so probability is

$$\Pi(k) = k \,/\, (2E)$$

Preferential Attachment
"Rich get Richer"



Result:
Scale-Free
$n(k) \sim k^{-\gamma}$
$\gamma = 3$

Scale-Free Growing Model comments

- Growth not essential
  – rewiring with reattachment probability ⚐ ⇒ ♑ ~ **1.0**
  – mixture of rewiring and new edges
  – Hamiltonian methods
- Network not essential – ***k***=frequency of previous choices
- Generalised attachment probability

$$\Pi(k) = (1-p_r)\frac{k}{2E} + p_r\frac{1}{N}, \qquad 2 < \gamma = 1 + \frac{2}{p_r(2-\varepsilon)} < \infty$$

Preferential Attachment — Random Attachment

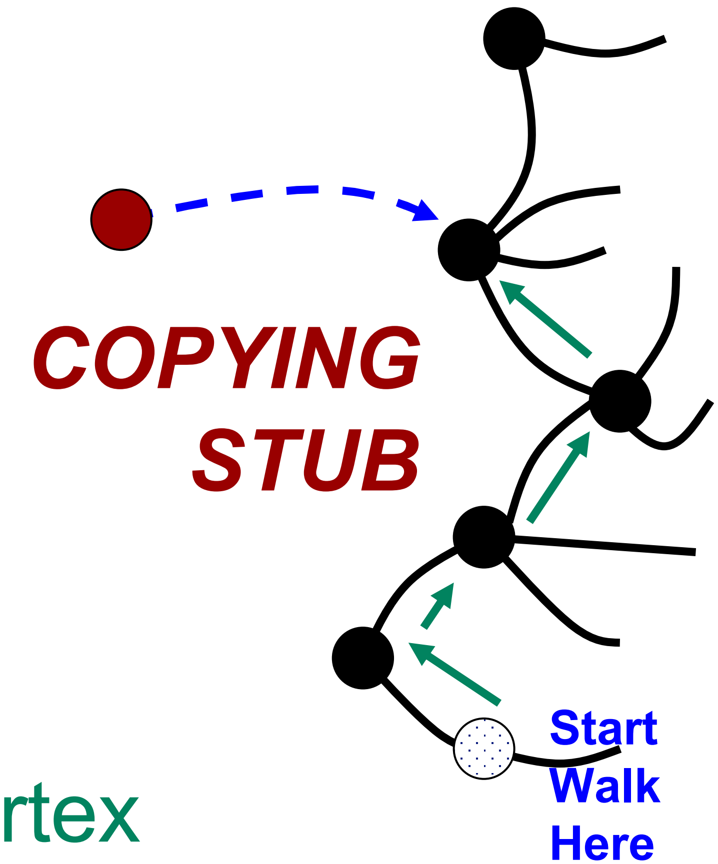
ε = fraction of times add new vertex

- BUT if ***lim***$_{k\to\infty}$ ⚐***(k)*** ✪ ***k***♋ ♋🕒🗁 for any then a *power law degree distribution is not produced!!*

Walking to a Scale-Free Network

[Saramäki, Kaski 2004; TSE, Saramäki 2004]

1. Add a new vertex with ***½<k>*** new edges
2. Attach to existing vertices, found by executing a random walk on the network of ***L*** steps



➔Probability of arriving at a vertex

✪ number of ways of arriving at vertex

= *k*, the degree

⇒ Preferential Attachment ***γ=3***

(Can also mix in random attachment with probability p_r)

Naturalness of the Random Walk algorithm

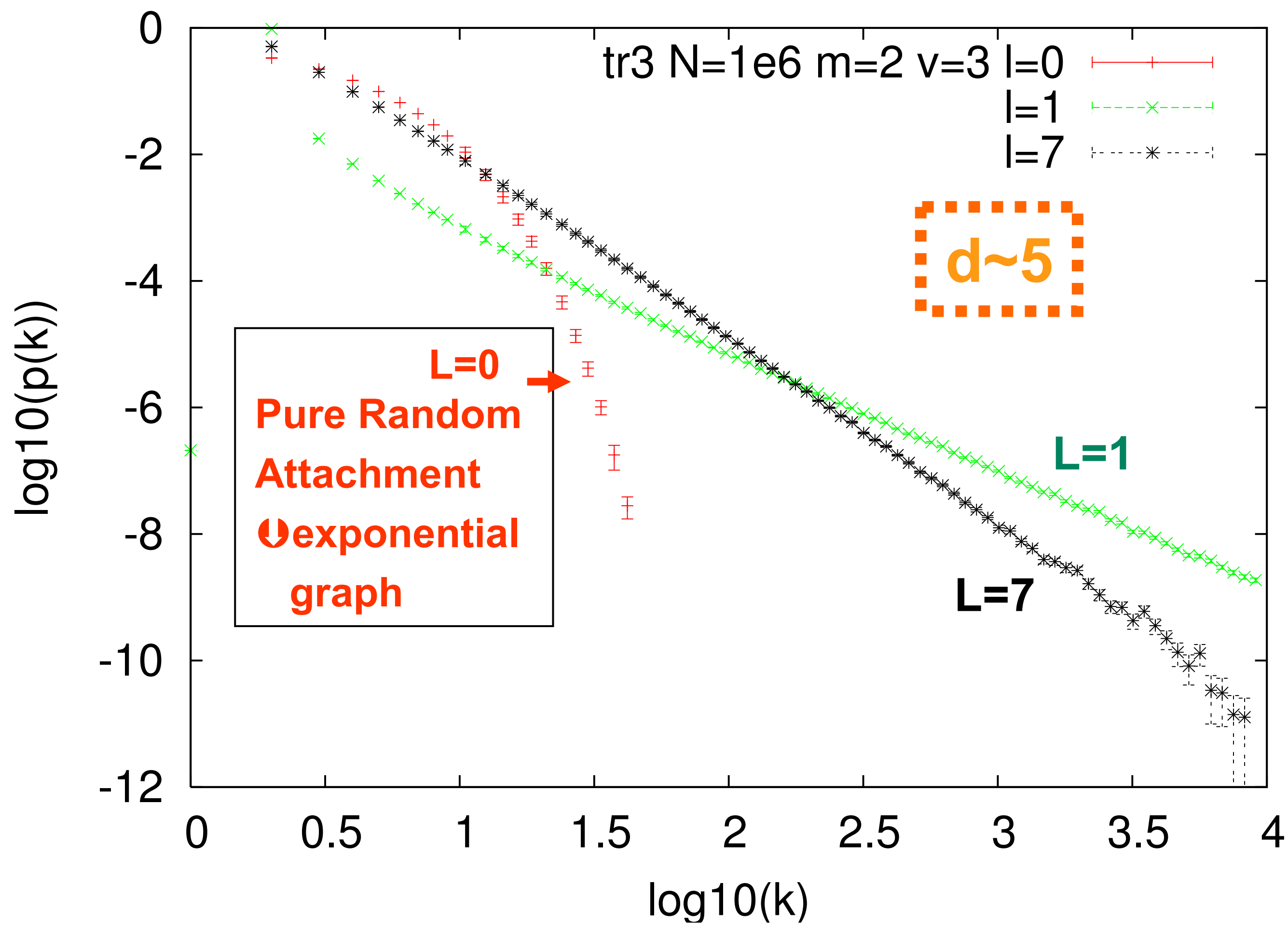
Automatically gives preferential attachment for any shape network and hence tends to a scale-free network

- Uses only **LOCAL** information at each vertex
  - Simon/Barabasi-Albert models use global information in their normalisation
- Uses structure of Network to produce the networks
  - a self-organising mechanism

    e.g. informal requests for work on the film actor’s social network

    e.g. finding links to other web pages when writing a new one

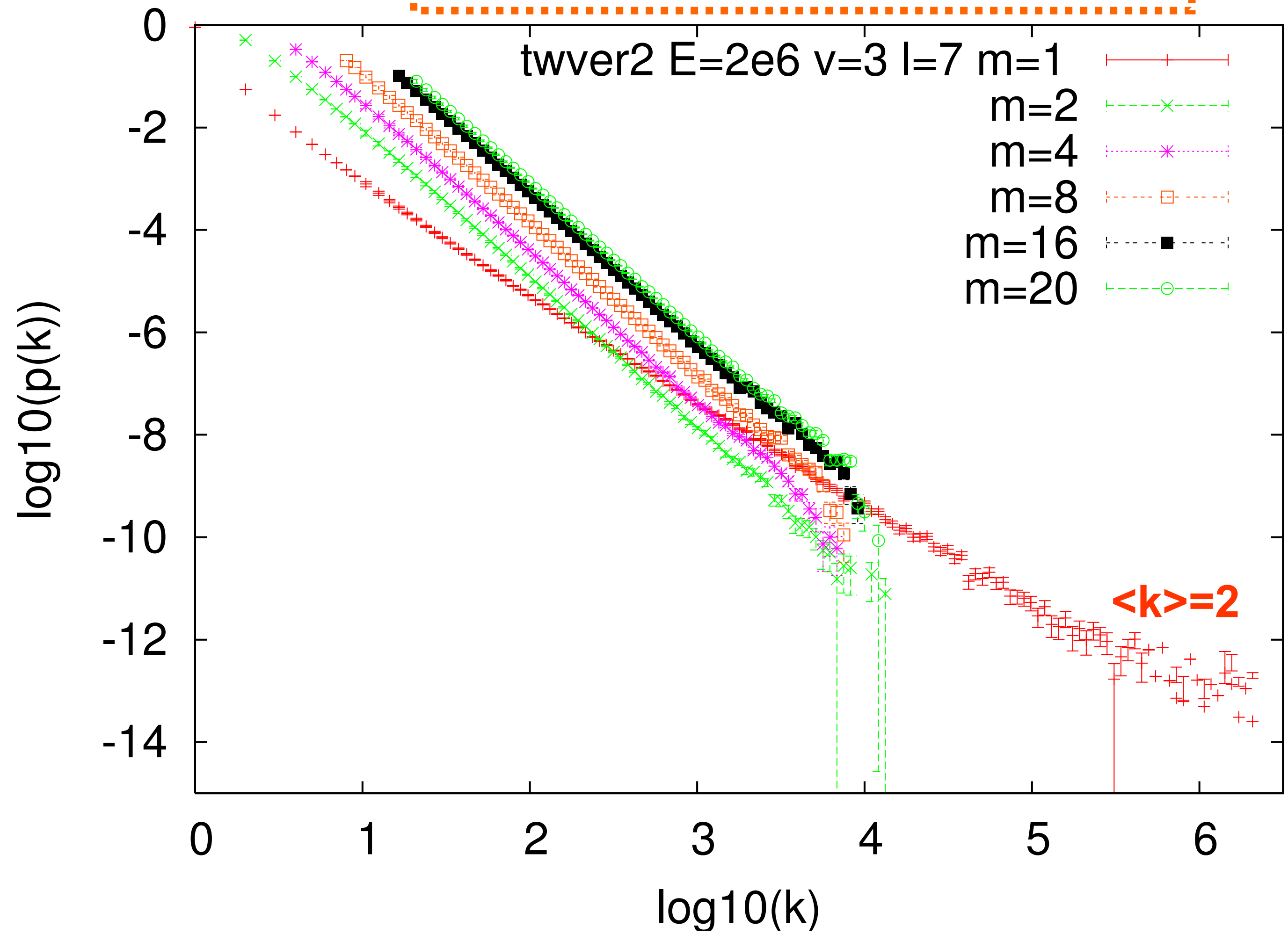
How long a walk is needed for a scale-free network?



- Walks of length ONE are usually sufficient to generate reasonable scale-free networks
  - ➔ Degree Correlation Length **_< 1 < d_** (any distance scale)

Does the average degree <k> matter?

Simple walks, l=7, 2m=k



NO

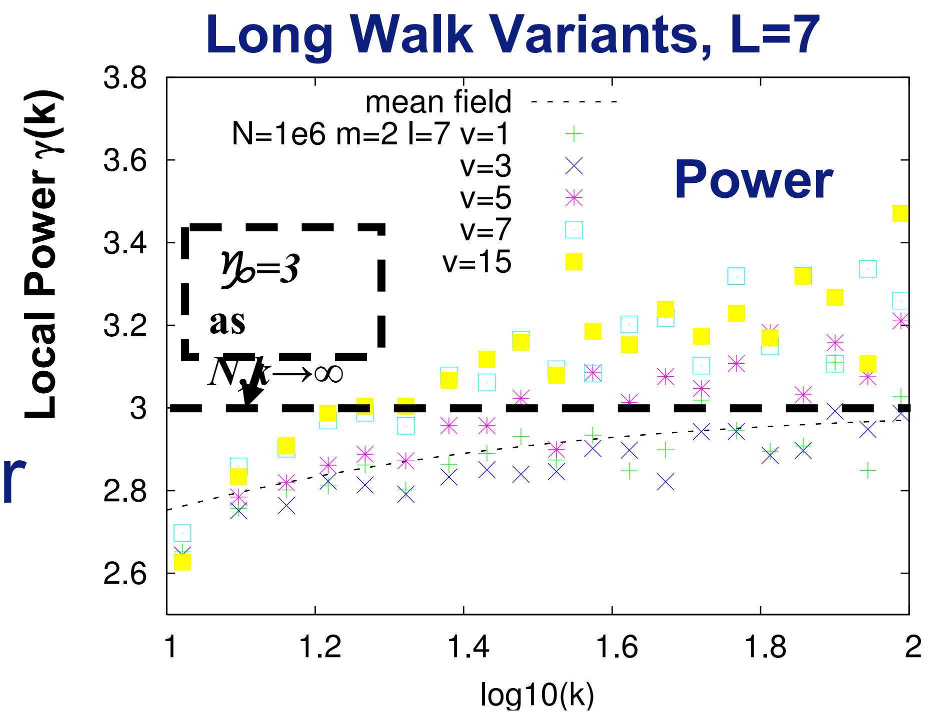
except for <k>=2 where a tree graph is generated

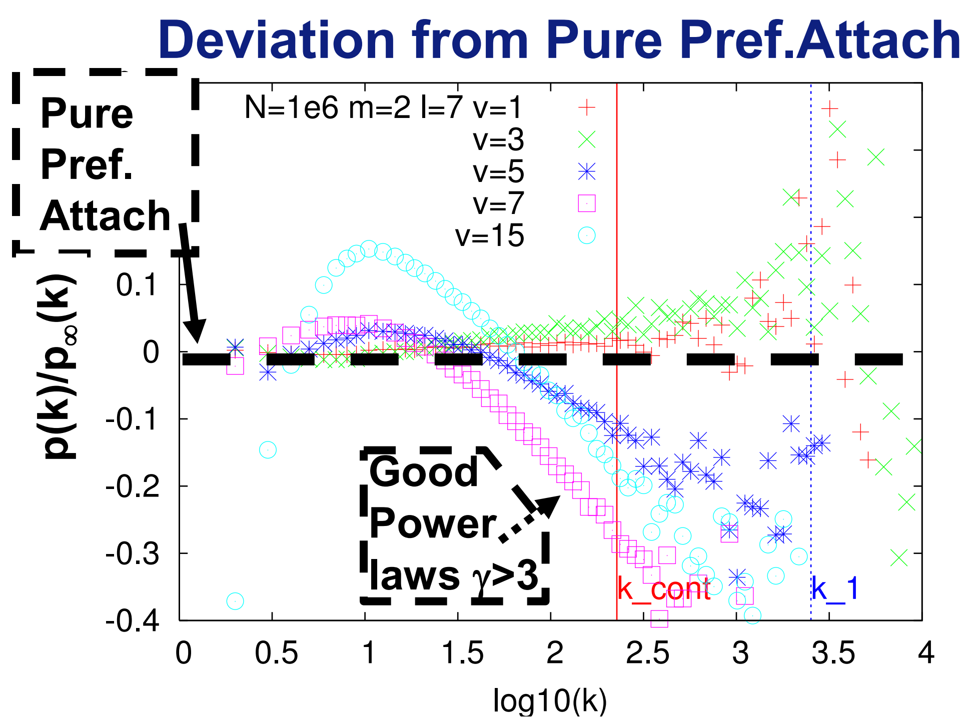
Is the Walk Algorithm Robust?

YES

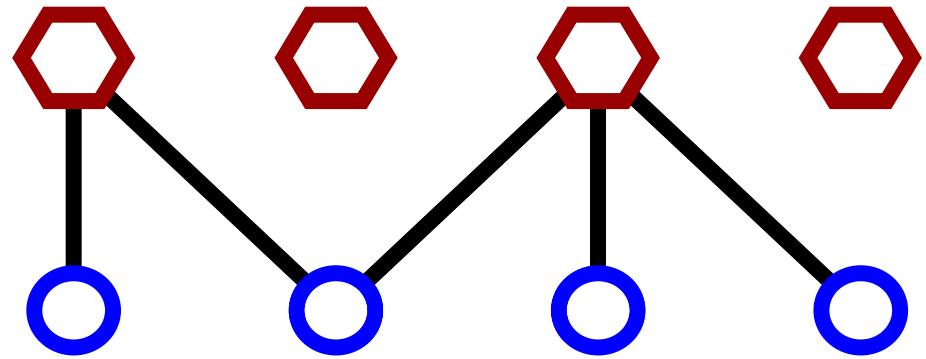
- Different starting points
- Vary length of walks per edge *keep* $L=\langle L\rangle$ *fixed*
- Vary edges added per vertex *keep* $\langle k\rangle$ *fixed*
- Allow multiple edges

Good Power Laws but power varies by 10% or 20%



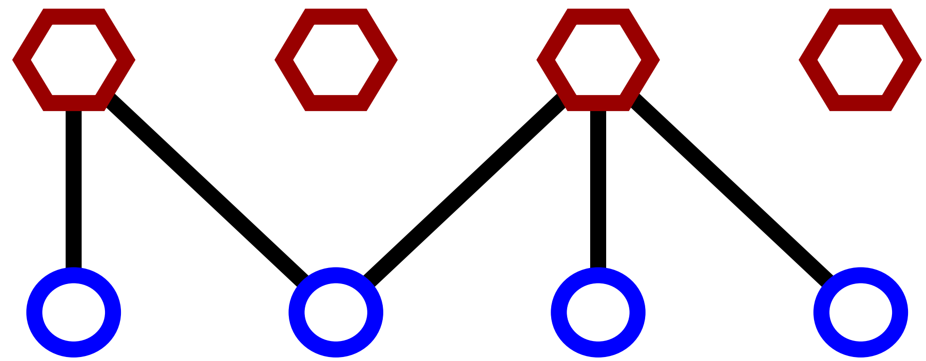


Copying and Culture



- Random Graphs
- Random Walks
- Random Walks and Copying - The Origin of Scale-Free Networks?
- Copying and Culture
- Summary

Copying and Culture



- Copying and Culture

o Preferential Attachment ⇔ Copying

e.g via random walk

o Rewiring of Networks of fixed size (***N***, ***E***)
vs. Growing Networks

o Example of how to get ***exact*** solutions for ***finite sized*** graphs at ***any time***

A Simple Model of Cultural Transmission

- Fixed population of ***E*** individuals
- Each person chooses one of ***N*** artifacts
  - Artifacts have no intrinsic benefit
  - e.g. pedigree dog, shoe style, name for baby
- At each time step, one random person updates their choice using one of two methods:-

  (a) COPYING someone else's choice

  (b) INNOVATING, picking an artifact at random
  it will be one no one else has chosen if *N* large

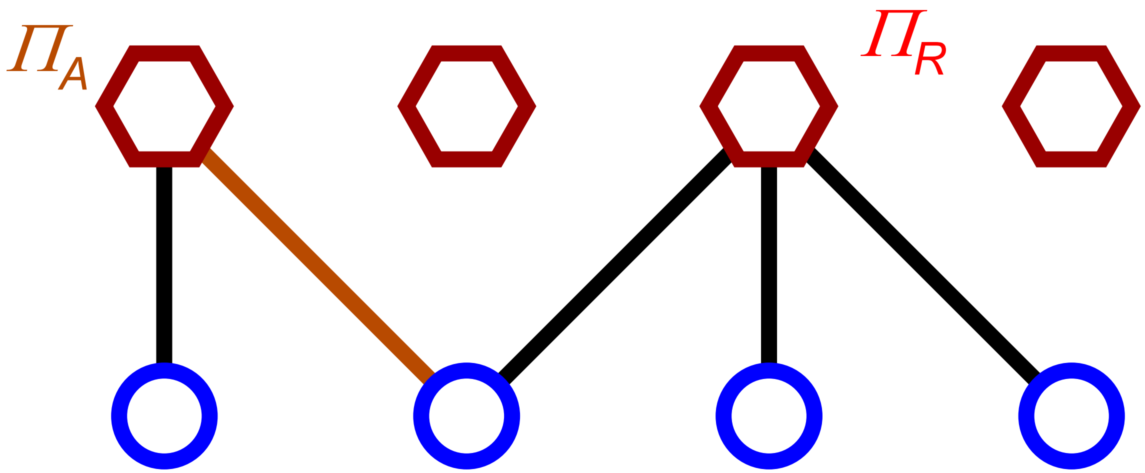
The Model as Network Rewiring

- **Removal**: Choose an individual at random
  = choosing departure artifact with probability $\Pi_R = (k/E)$
  = preferential removal from artifacts
- **Attachment**: Choose an arrival artifact with probability

$$\Pi_A(k) = [(1-p_r)k + p_r \langle k \rangle] / E \qquad \langle k \rangle = (E/N)$$

copying probability — innovation probability

- **Rewire**: Only *after* these choices are made.



N artifacts

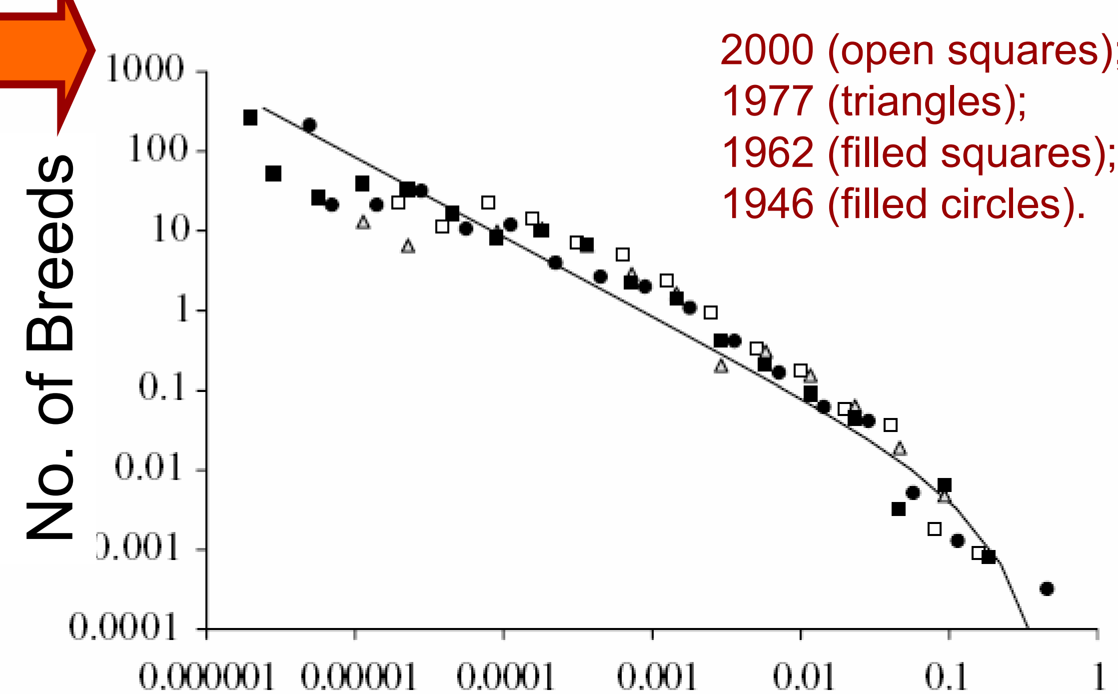
n(k) artifact degree dist.

E individuals

Evidence for this model

- Registrations of pedigree dogs
- Baby name registrations
- Changes in top 100 of popular music charts
- Applied to archaeological pot shards

[Herzog, **Bentley**, Hahn 2004]



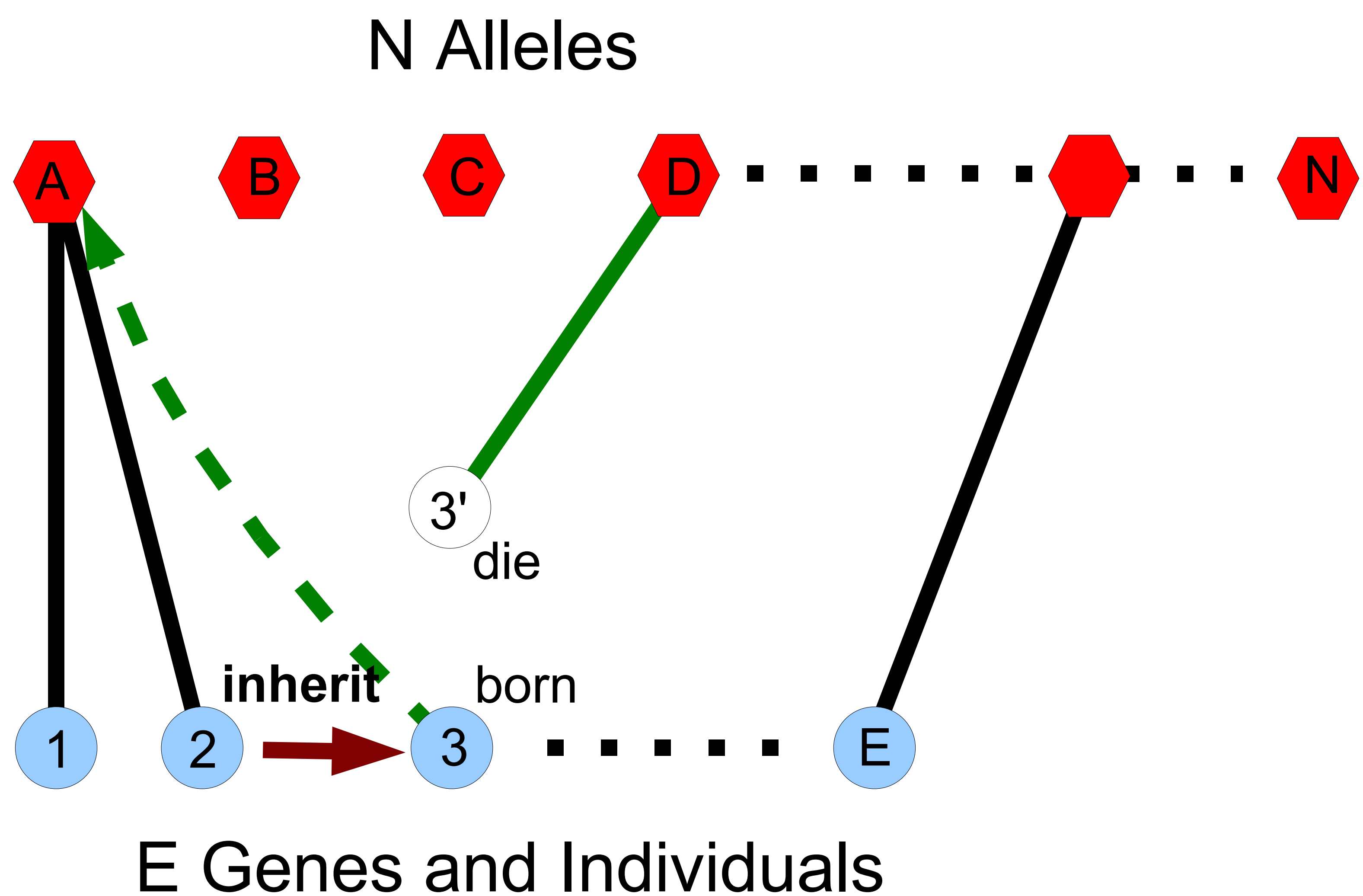
Frequency of registrations of each breed of pedigree dog

See Neiman (1995); Bentley, Maschner (2000,2001); Bentley, Hahn, Shennan (2004); Bentley, Shennan (2003,2005); Hahn, Bentley (2003); Herzog, Bentley, Hahn (2004); Bentley, Lipo, Herzog, Hahn (2007).

Relationship to Other Systems

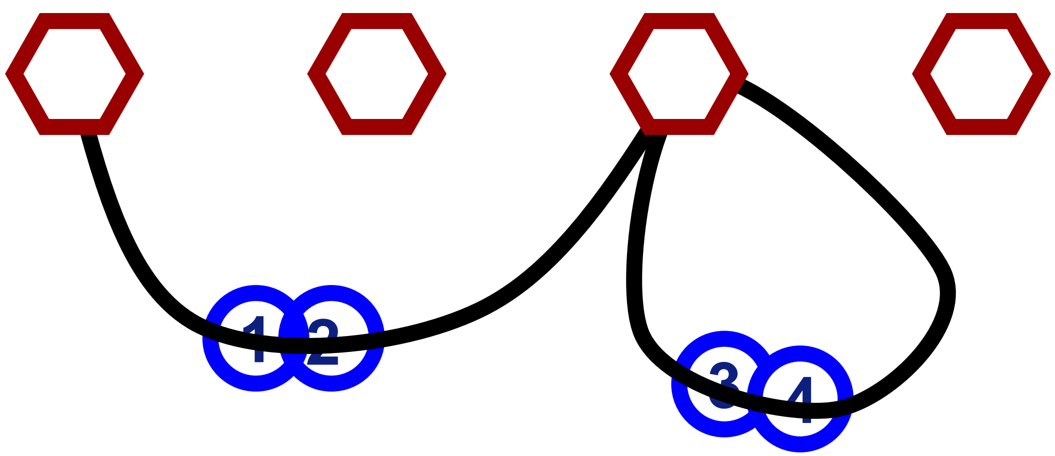
- Unipartite Graph Rewiring [Watts & Strogatz 1998]
- Gene Frequencies [Kimura & Crow, 1964]
  - Inheritence and Mutation
- Family Names [Zanette & Manrubia, 2001]
  - Inheritence and New Immigrants
- Language Extinction [Stauffer et al. 2006]
- Minority Game variant [Anghel et al, 2004]

Model as Population Genetics



Model as simple graph rewiring

- Project onto a unipartite graph of the artifact vertices of degree distribution ***p(k)***
  = degree distribution for a generalised random graph



N artifacts

(E/2) edges

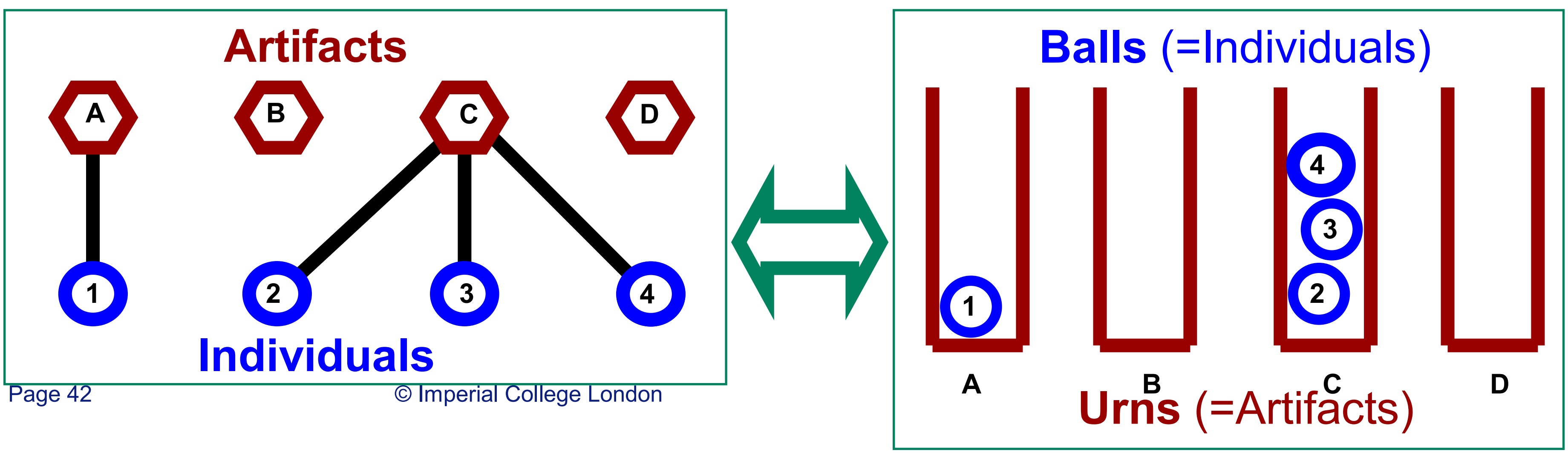
$\Rightarrow$ This is a Molloy-Reed projection

Relationship to Statistical Physics Models

Some parameter values of other models are equivalent to our model:

- Urn Models [Bernoulli 1713, …, Ohkubo et al. 2005]
- Zero Range Processes (Misanthrope version) [review M.R.Evans & Hanney 2005; Pulkkinen & Merikoski 2005]
- Voter Models [Liggett 1999, …, Sood & Redner 2005]
- Backgammon/Balls-in-Boxes

applied to glasses [Ritort 1995], wealth distributions, simplicical gravity …



Mean Field Degree Distribution Master Equation

Mean field approximation very accurate for many models (low vertex correlations)

$$\begin{aligned} n(k,t+1)-n(k,t) = &+ n(k+1,t)\Pi_R(k+1)[1-\Pi_A(k+1)] \\ &- n(k,t)\Pi_A(k)[1-\Pi_A(k)] \\ &- n(k,t)\Pi_R(k)[1-\Pi_R(k)] \\ &+ n(k-1,t)\Pi_A(k-1)[1-\Pi_R(k-1)] \end{aligned}$$

(1-Π) terms Invariably ignored

Number of edges attaching to a vertex of degree (k-1)

Probability of NOT reattaching to same vertex

Can the Mean Field equation be exact?

Distribution $n_i(k)$ different in each instance i

Ensembles over many instances i

$$\left\langle \frac{n_i(k)k^\beta}{\sum_k n_i(k)k^\beta} \right\rangle \neq \left\langle n_i(k)k^\beta \right\rangle \left\langle \frac{1}{\sum_k n_i(k)k^\beta} \right\rangle$$

Normalisation of probabilities not usually same for different i

YES only if $\sum_k n_i(k)k^\beta = \left\langle \sum_k n_i(k)k^\beta \right\rangle$ $\beta = 0$ or $\beta = 1$

Only Exactly Solvable Case

To be able to solve exactly we limit the attachment and removal probabilities, Π_R and Π_A, to be **linear** in degree exploiting only two constants of the motion, N and E

- $\Pi_R(k) = (k / E)$ Choose random edge to be rewired
- $\Pi_A(k) = [(1-p_r)k + p_r\langle k \rangle] / E$

Fraction $(1-p_r)$ of the time use preferential attachment

Fraction p_r of the time choose random attachment

Exact Solution

Use the generating function ***G(z,t)*** :-

$$\mathrm{G}(z,t) = \sum_{k=0}^{E} (z)^{k}\, \mathrm{n}(k,t)$$

Discrete version of a Mellin Transform

Properties of Generating Function ***G(z,t)***

- Derivatives related to moments

$$\langle k^n \rangle = \left. \frac{d^n \, \mathrm{G}(e^x, t)}{dx^n} \right|_{x=0} = \left. \left(z \frac{d}{dz} \right)^n \mathrm{G}(z,t) \right|_{z=1}$$

$\Rightarrow$ Equation for degree distribution becomes a second-order linear differential equation for the generating function

$$\frac{b(1+a-c)}{(1-z)} \left[G(z,t+1) - G(z,t) \right] =$$

$$z(1-z)G''(z,t) + [c-(a+b+1)z]G'(z,t) - abG(z,t)$$

where $a = \frac{p_r}{p_p} \langle k \rangle, \; b = -E, \quad c = 1 + a + b - \frac{p_r}{p_p} E$

Exact Solution

Exploit linearity and break into eigenfunctions:-

$$G(z,t) = \sum_{k=0}^{E} (z)^k \, n(k,t) = \sum_{m=0}^{E} c_m (\lambda_m)^t \, G^{(m)}(z)$$

$\Rightarrow$ Find Hypergeometric equations and solutions:-

Eigenfunctions $G^{(m)}(z) = (1-z)^m \, F(a+m, b+m; c; z)$

Hypergeometric function

$$a = \frac{p_r}{p_p}\langle k \rangle, \quad b = -E, \quad c = 1 + a + b - \frac{p_r}{p_p} E$$

Eigenvalues $\lambda_m = 1 - m(m-1)\frac{p_p}{E^2} - m\frac{p_r}{E}$

c_m are constants fixed by initial conditions

Exact Equilibrium Solution

$$n(k) = A \frac{\Gamma(k+\overline{K})}{\Gamma(k+1)} \frac{\Gamma(E-\overline{E}-\overline{K}-k)}{\Gamma(E+1-k)}$$

$$\overline{K} = \frac{p_r}{p_p}\langle k \rangle$$

$$\overline{E} = \frac{p_r}{p_p} E$$

A is ratio of four Γ functions

- Simple ratios of Γ functions
- Similar to those found for growing networks but second fraction is only found for network rewiring with correct master equation
- Only approximate solutions known previously

Large Degree Equilibrium Behaviour – Large p_r Case

For $\boldsymbol{p_r > p_* \sim 1/E}$
(on average at least one edge attached to a randomly chosen artifact per generation)

$$\lim_{k\to\infty}\left[n(k)\right] = k^{-\gamma}\exp(-\zeta k)$$

$$\gamma = 1 - \frac{p_r}{p_p}\langle k \rangle$$

Power below one but in data *indistinguishable from one*

$$\zeta = -\ln(1-p_r)$$

Exponential Cutoff

Large Degree Equilibrium Behaviour – Small p_r Case

For $p_r < p_* \sim 1/E$

(on average if all edges have been rewired once no edge is attached to a randomly chosen artifact per generation)

Degree distribution rises near $k=E$

⇒In extreme case $p_r=0$ all the edges are attached to ONE artifact

- a CONDENSATION or FIXATION

Blows up

$$n(k) = A\left(\frac{\Gamma(k+\overline{K})}{\Gamma(k+1)}\right)\left[\frac{\Gamma(E-\overline{E}-\overline{K}-k)}{\Gamma(E+1-k)}\right]$$

Equilibrium Behaviour Results

N=E=100

Points: *10^5* data runs

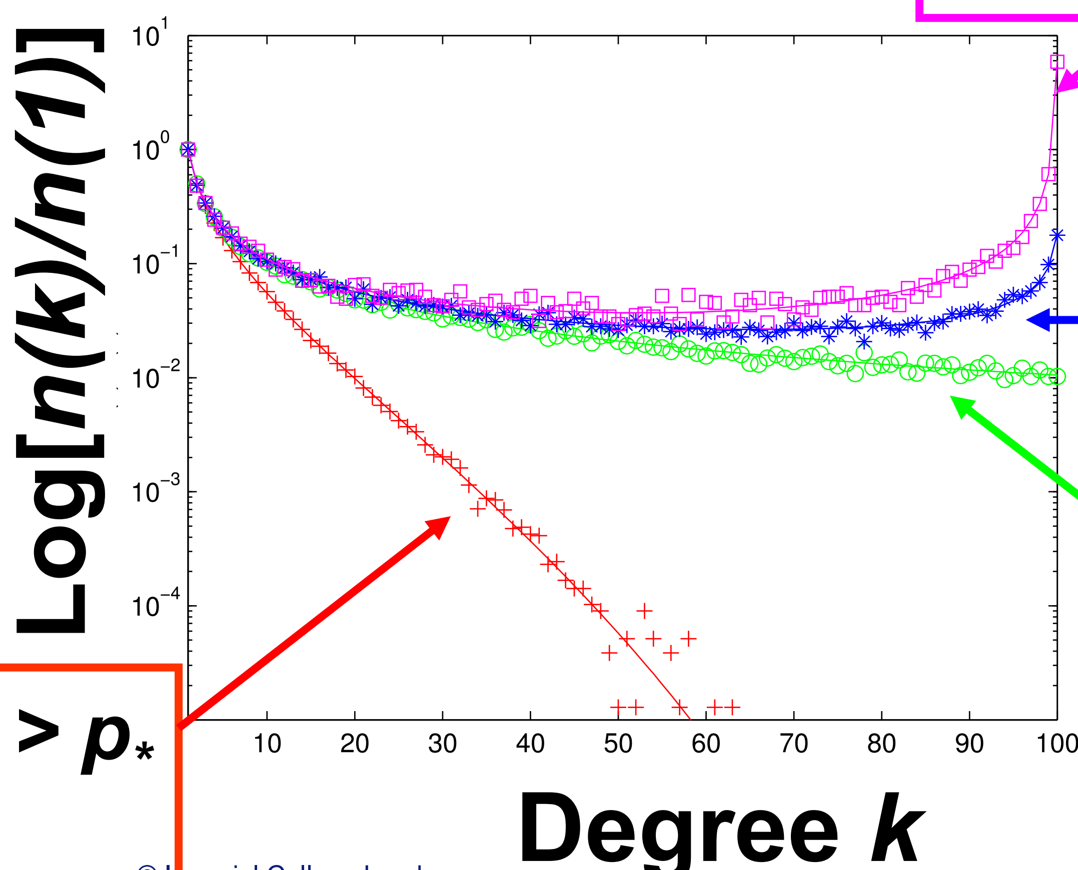
Lines: exact mean field solution

$p_r = 0.001 < p_*$
condensate

$p_r = 0.005 < p_*$

$p_r = 0.01 \cong p_*$
Almost pure
Power law

$p_r = 0.1 > p_*$
$\zeta^{-1} \cong 10$



Features of the Exact Solution

- Eigenfunction number zero is time independent ($\lambda_0=1$) $\Rightarrow$ equilibrium solution
- Eigenfunction number one *never contributes*
- Slowest time dependence comes from ***m=2*** eigenfunction, setting the equilibration time scale

$$\tau_2 = -1/\ln(\lambda_2) \approx [2(p_r/E) + 2(1-p_r)/E^2]^{-1}$$

- $\langle k^n \rangle$ *n-th* moment of degree distribution gets contributions only from eigenfunctions numbered *n* and lower

Exact Solution

$$\lambda_m = 1 - m(m-1)\frac{p_p}{E^2} - m\frac{p_r}{E}$$

(m=0,1,2,..E)

- Use generating function.
  - It splits into (E+1) eigenfunctions, given by Hypergeometric functions
  - Simple eigenvalues
- <k^n> *n-th* moment of degree distribution gets contributions only from eigenfunctions $m \leq n$ only
- $m=0$ eigenfunction number zero constant ($\lambda_0=1$)
  $\Rightarrow$ equilibrium solution
- $m=1$ eigenfunction *never contributes*
- Slowest time dependence comes from $m=2$ eigenfunction setting time scale $\tau_2 = -1/\ln(\lambda_2)$

Homogeneity Measures F_n

BEST WAY TO STUDY DEGREE DISTRIBUTION

- ***$F_n(t)$* = probability of choosing *n* different individuals connected to the same artifact**
  - ***F_n = 0* if no artifact chosen more than once**
  - ***F_n = 1* if all individuals attached to same artifact**
- Related to *m*-th moments ($m \leq n$) of degree distribution via Stirling numbers but ***F_n = 0 if n>E***
- *n*-th derivatives of generating function gives homogeneity measure ***F_n***

$$F_n(t) := \frac{\Gamma(E+1-n)}{\Gamma(E+1)} \frac{d^n G(z,t)}{dz^n}\bigg|_{z=1} = \sum_{k=0}^{E} \frac{k}{E} \frac{(k-1)}{(E-1)} \cdots \frac{(k-n+1)}{(E-n+1)} n(k)$$

Exact Solution for F_2 Homogeneity Measure

F_2 = probability that two different individuals have chosen the same artifact

$$F_2(t) = F_2(\infty) + (\lambda_2)^t \left(F_2(0) - F_2(\infty)\right)$$

$$F_2(\infty) = \frac{1 + p_r(\langle k \rangle - 1)}{1 + p_r(E - 1)}$$

3rd eigenfunction controls all time dependence

$\tau_2 = -1/\ln(\lambda_2)$

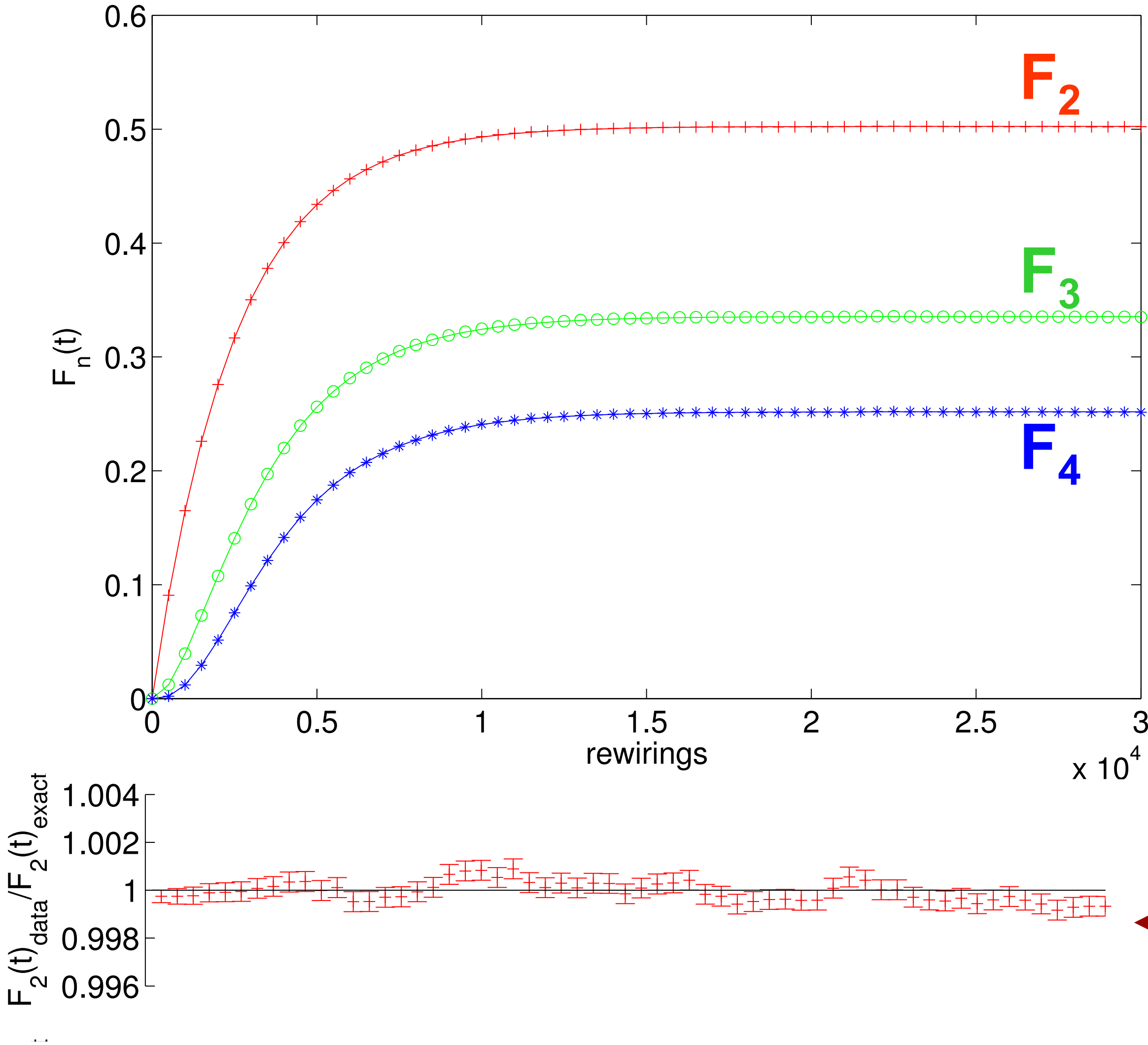
$\approx [2(p_r/E) + 2(1-p_r)/E^2]^{-1}$

Initial values fix $F_2(0)$

e.g. $F_2(0)=0$ if each individual starts attached to unique individual

F_n numerical results

E=N=100, $p_r=0.01 \cong p_*$,
Points: average of 10^5 simulations
Lines: exact mean field prediction
Start: $n(k)=\delta_{k,1}$

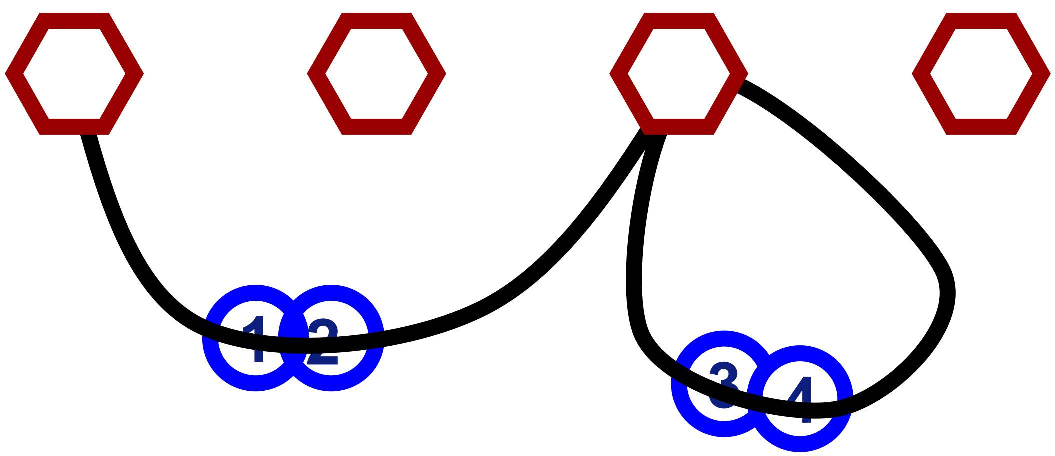


F increases as homogeneity increases with time

Time dependence of averages predicted very accurately, deviations less than 1%

Phase transitions in real time

- Bipartite graph can be projected onto a unipartite graph of the artifact vertices
- Artifact degree distribution ***p(k)*** is the degree distribution for a random graph



N artifacts

(E/2) edges

$\Rightarrow$ This is a *Molloy-Reed* projection

Graph Transition in Real Time

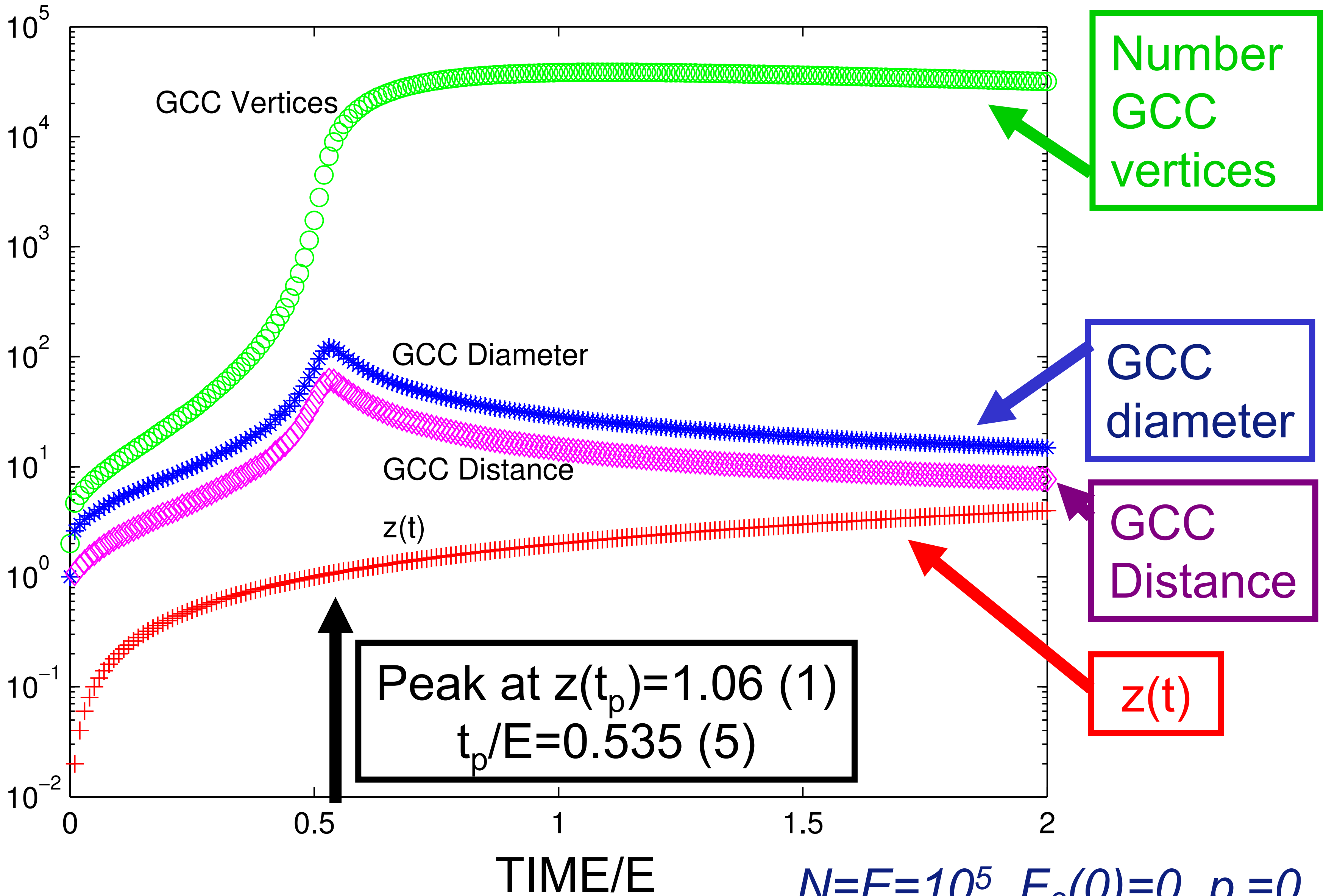
Infinite Random Graphs (given *p(k)* but otherwise completely randomised) have a phase transition (e.g. appearance of **GCC** - Giant Connected Component) at [Fronczak et al 2005, etc]

$$z(t)=1$$

where

$$z(t) = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 = (E-1)F_2(t)$$

Phase Transition in Molloy-Reed projection



$N=E=10^5$, $F_2(0)=0$, $p_r=0$

Phase Transition in Molloy-Reed projection

For $N=E=10^5$, $p_r=0$, initial $F_2(0)=0$

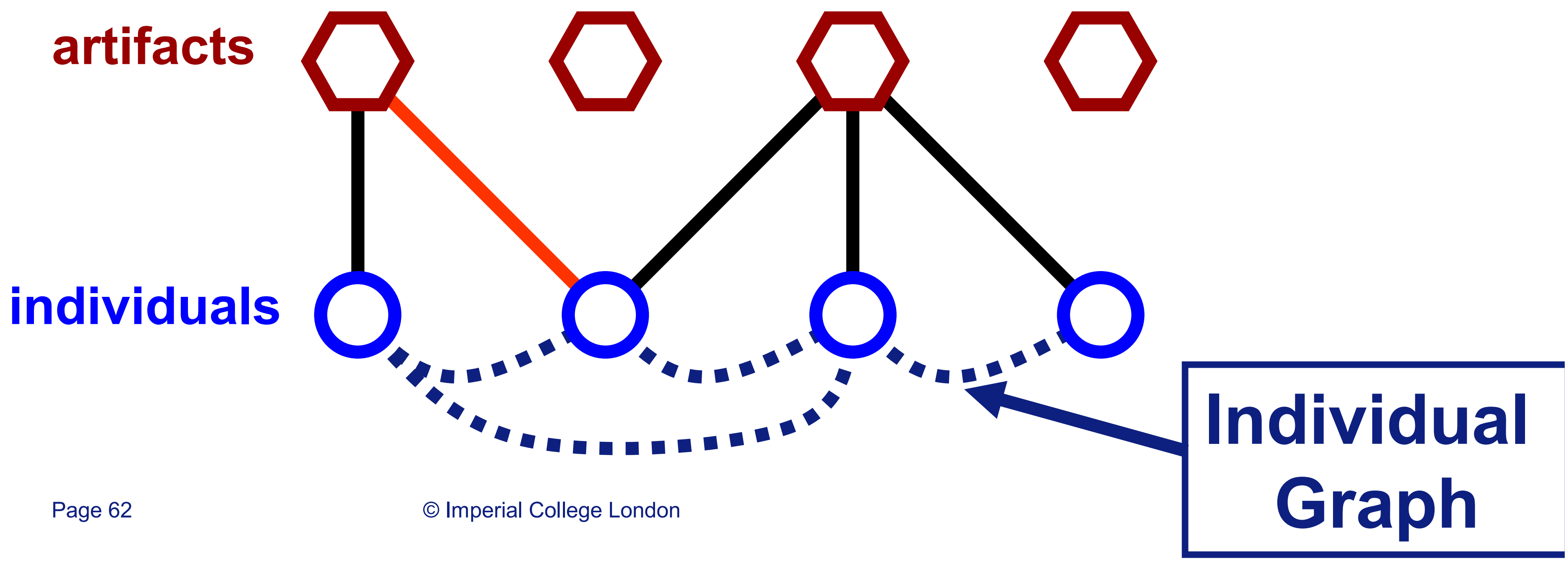
- $z(t)=1$ at $t=0.50000\ (2)$ as predicted
- Transition at $t/E = 0.535\ (5)$
- At transition $z(t)=1.06\ (1)$ not $z(t)=1$
- Average distance and diameter of GCC maximum at this point and second derivative of number of vertices in GCC zero at this point (within errors)

$\Rightarrow$ **Finite size effects clearly present**

$\Rightarrow$ **Can follow a system through a phase transition in time *exactly***

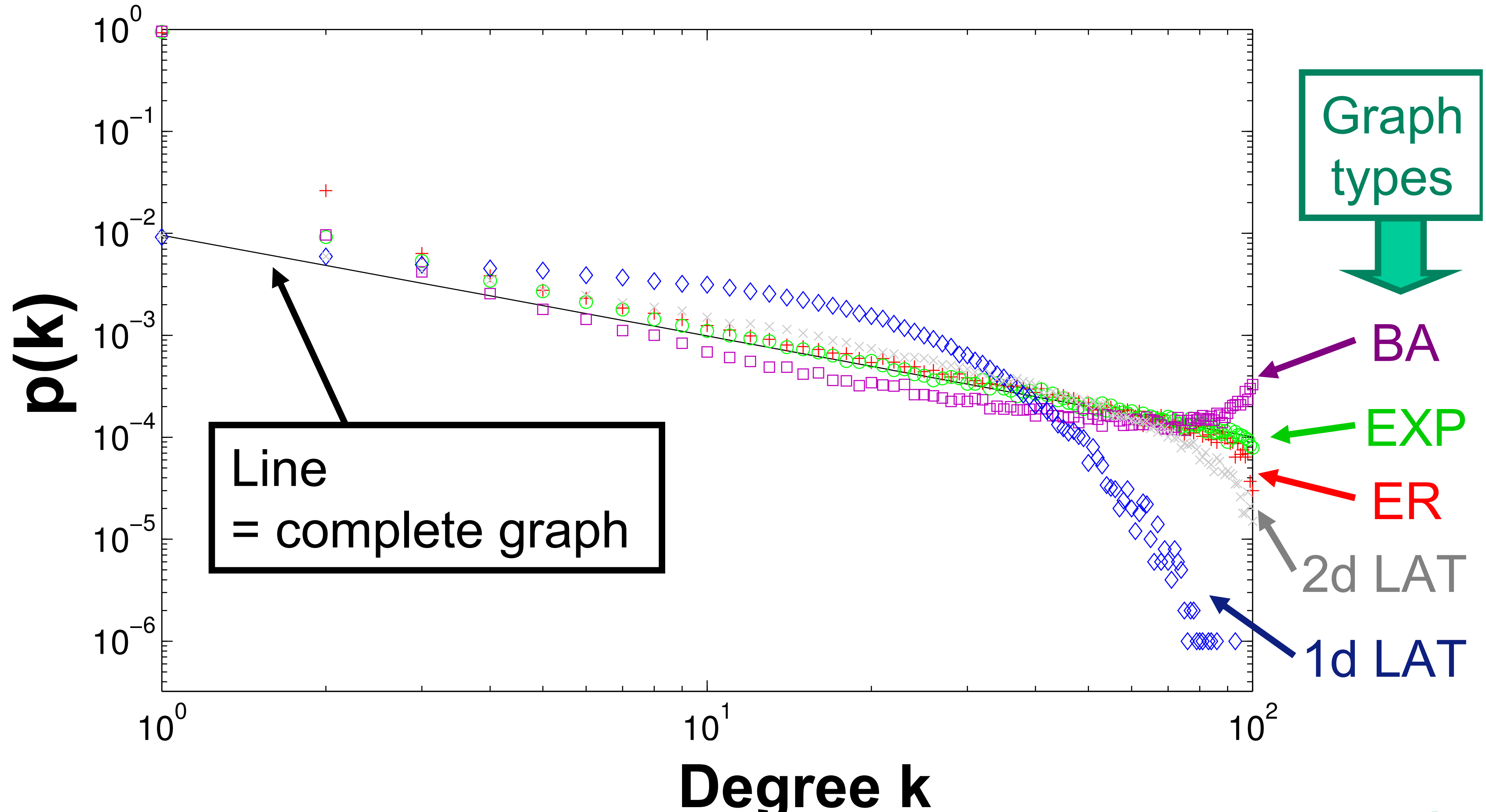
Adding a Network of Individuals

- **Removal**: Choose random individual as before
- **Attachment**:
  With probability $(1-p_r)$ the individual **copies** the existing choice of any **neighbouring** individual.
  With probability (p_r) the individual **innovates**



Equilibrium with a Network of Individuals

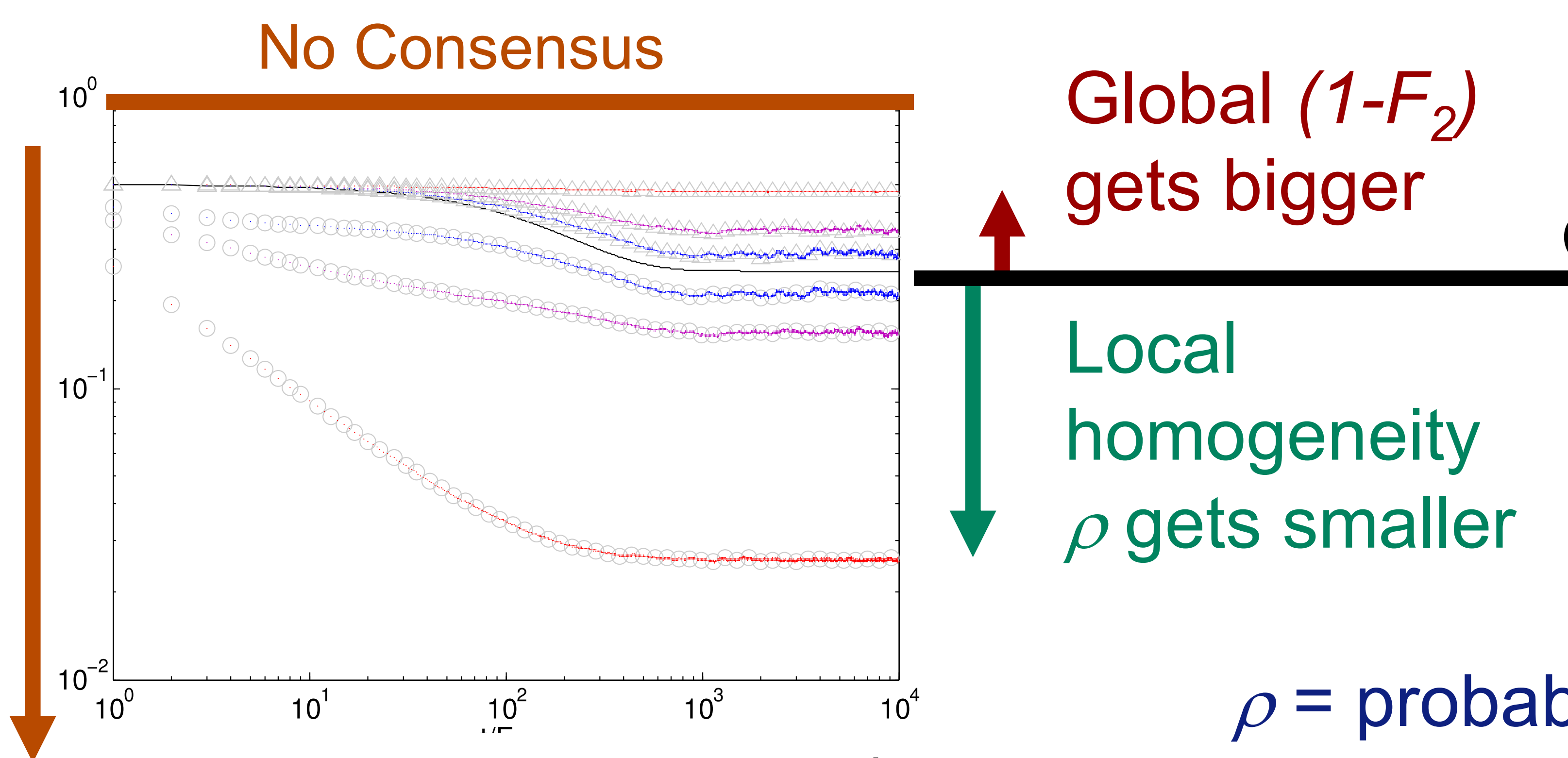
Qualitative behaviour largely unchanged except for 1d Lattice



$N=E=100,\ p_r=1/E$

Approach to Equilibrium for different Individual networks

- Results move away from complete graph as move from 3d -> 1d lattice



TIME $N=2$, $p_r=1/E$, $E=729$

ρ = probability that n.n. has made different choice

Voter Model [Liggett 1999; Sood & Redner 2005]

- At each time step an individual is chosen randomly who copies the choice of a neighbour in an individual network
- Equivalent to $N=2$, $p_r=0$ limit here
- Study time scales to come to complete **consensus** = condensation
- Used for models of language [Stauffer et al. 2006]

⇒We find approach to complete consensus is slow but a little randomness ($p_r > 0$) can speed this up while leaving a fairly complete condensation

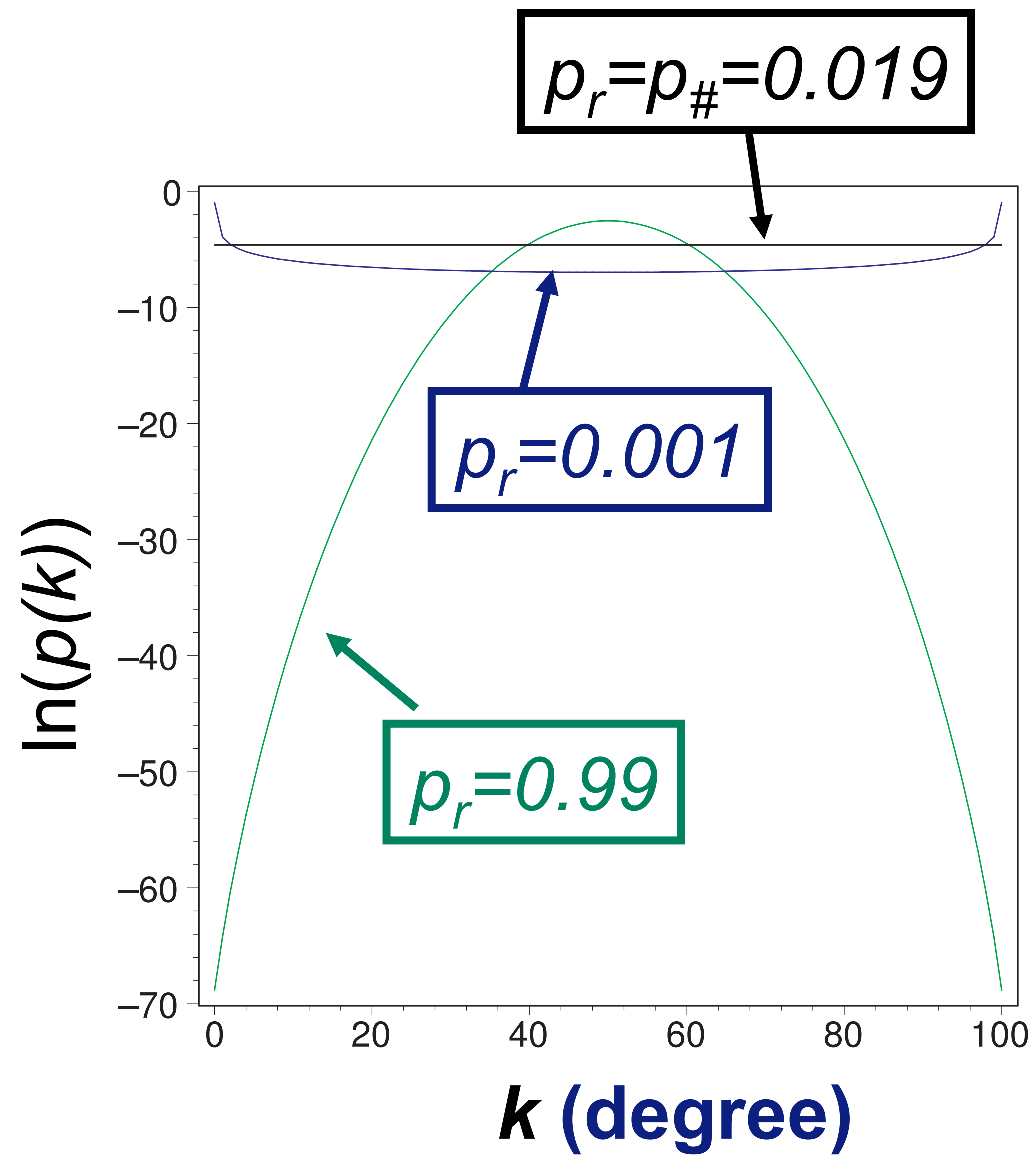
Phase Transition in the Generalised Voter Model

Here on a complete or random individual network

- N=2 so large <k> is a special case
- Transition occurs at

$$p_{\#} = (E+1+\langle k\rangle)^{-1} = (1+(E/2))^{-1}$$

- May be viewed as Z_2 symmetry breaking transition



Minority Game Example - Leaders and Followers

- At each step each individual chooses one or zero
  - the *minority* choice wins
- Choices are made based on one of a large but finite number of strategies using finite history
  - each strategy is a different artifact
- Individuals may follow their own prediction or they may follow the prediction from the most successful nearest neighbour in an ER random graph of individuals
  - i.e. they ***copy*** the strategy of a neighbour

[Anghel et al. PRL 92 (2004) 058701]

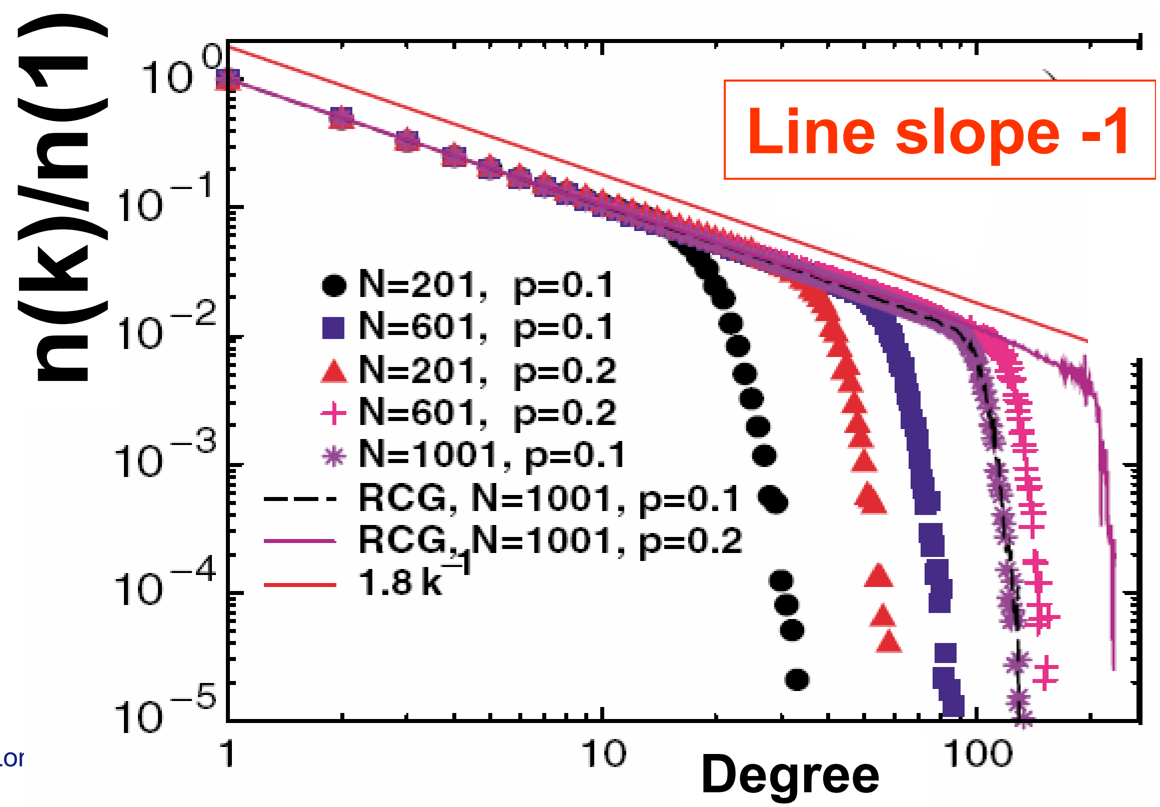
Minority Game Example – Leaders and Followers [Anghel et al. 2004]

Plot $n(k)$ the average of the number of strategies (of some leader) used by k individuals (followers).

Various system sizes and various ER random graphs.

Result exactly as in our model

⇒ Random Copying



Minority Game Example - Leaders and Followers

This Minority Game variant again shows how **copying** can arise naturally

c.f. preferential attachment in growing networks

[Saramäki & Kaski 2004, TSE & Saramäki 2005]

Different Update Methods

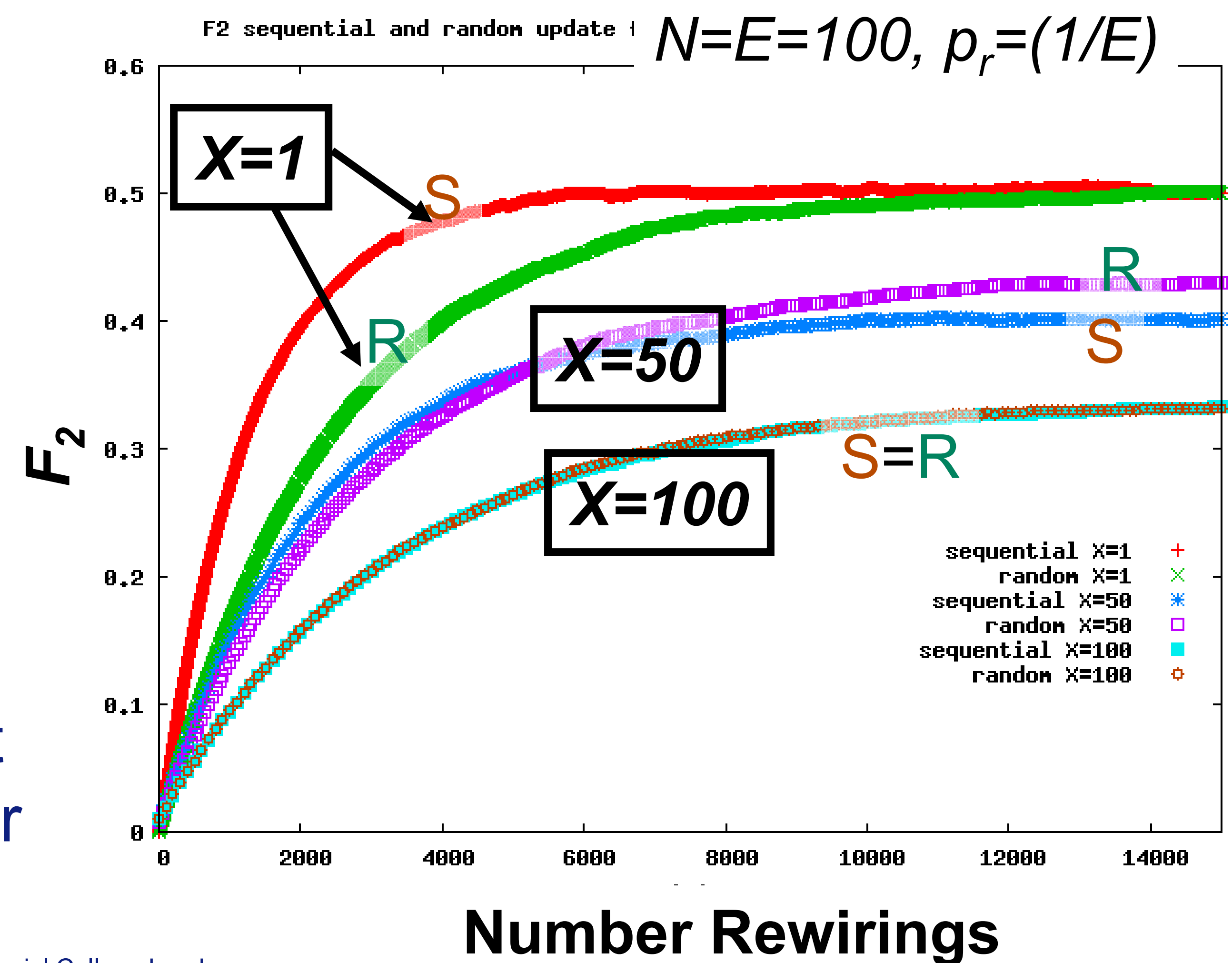
- First select X different individuals
  either (R) selected randomly
  or (S) in numerical sequence (1,2,3,…)
- They make their new artifact choices at the same time
- Only now we update the network and repeat

⇒ X=1 & *random selection* = model so far

⇒ X=E & *sequential selection* = models of Bentley et al.

Numerical Results for Update Variations

- *X=1:* sequential faster than random but equilibrium same
- *X=E/2*: time scales similar but equilibrium F_2 lower for sequential
- *X=100:* update all at once and get F_2 =1/3 not ½ as we get for *X=1*



[TSE + You (in progress)]

Analytic Results for Update Variations

$$F_2(t) = F_2(0) + (\lambda_2)^t \left(F_2(\infty) - F_2(0)\right)$$

(R) selected randomly
X=1
TSE & Plato

$$F_2(\infty) = \frac{1 + p_r\left(\langle k\rangle - 1\right)}{1 + p_r\left(E - 1\right)}$$

$$\lambda_2 = 1 - \frac{2p_r}{E} - \frac{2(1 - p_r)}{E^2}$$

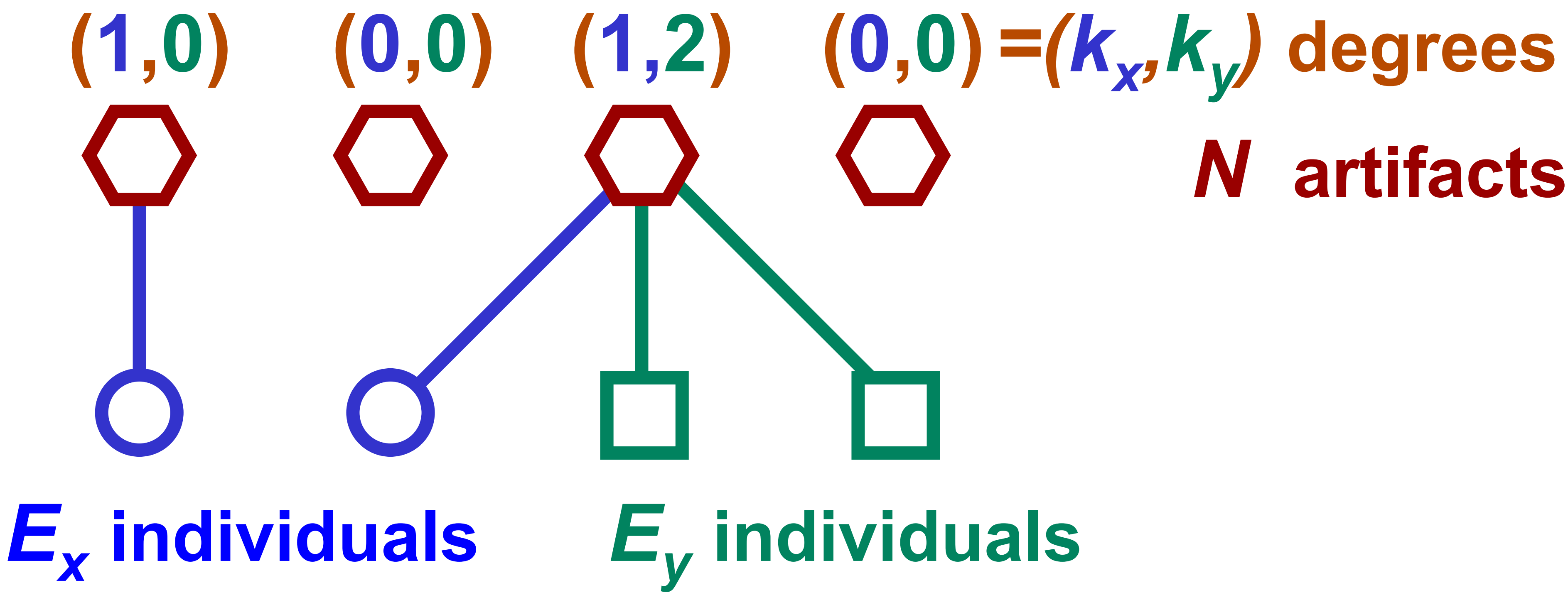
(S) Sequential Update
X=100
Bentley et al.

$$F_2(\infty) = \frac{(p_p)^2 + \left(1 - (p_p)^2\right)\langle k\rangle}{(p_p)^2 + \left(1 - (p_p)^2\right)E}$$

$$\lambda_2 = (1 - p_r)^2 (1 - 1/E)$$

Two Tribes

Change model so there are two types of individual, each type chooses new artifacts with their own probabilities for:- (A) copying from same type, (B) copying from different type, (C) innovation



Two Tribes

- Exact solutions for inhomogeneity measures $F_{2ab}(t)$ $[a,b \in \{X,Y\}]$ still possible
  - solutions of three-dimensional matrix
- 8 free parameters
  - difficult to draw general conclusions
- Might relate to *Freakonomics* type explanation for baby names in terms of different socioeconomic groups

- Random Graphs
- Random Walks
- Random Walks and Copying - The Origin of Scale-Free Networks?
- Copying and Culture
- Summary

Summary

- Preferential Attachment
  = Making Random Walk on Network
  = Copying Choice made by neighbour
- Applied to network rewiring can get ***exact*** solutions for ***any finite sized*** graph ***at any time***
  - Related to many other situations where reached size of system is constant (at least over short time averages) and where there are *two* processes
    1) copying/inheritance/preferential attachment
    2) innovation/mutation/random attachment

Bibliography

This work and most of their associated references are on my web page (google- *Tim Evans Imperial*) in:-

- T.S.Evans & J.Saramäki,
  Scale-free networks from self-organization.
  Phys.Rev.E. **72** (2005) 026138 [**cond-mat/0411390**]
- T.S.Evans & A.D.K.Plato
  "Exact Solution for the Time Evolution of Network Rewiring Models"
  Phys. Rev. E **75** (2007) 056101 [**cond-mat/0612214**]
- T.S.Evans & A.D.K.Plato
  "Network Rewiring Models" (for ECCS07)
  arXiv:0707.3783

Bibliography in more detail

- Eriksen, K. A.; Simonsen, I.; Maslov, S. & Sneppen, K., *Modularity and Extreme Edges of the Internet,* Phys. Rev. Lett., **90 (2003)** 148701
- T.S.Evans and A.D.K.Plato, *Exact Solution for the Time Evolution of Network Rewiring Models,* Phys. Rev. E **75** (2007) 056101 [`cond-mat/0612214`]
- T.S.Evans and A.D.K.Plato, *"Network Rewiring Models"* (for ECCS07) [`arXiv:0707.3783`]
- A.Fronczak, P. Fronczak and J.A. Holyst, *How to calculate the main characteristics of random uncorrelated networks* in "Science of Complex Networks: From Biology to the Internet and WWW; CNET 2004", (ed.s Mendes,J.F.F. et al.) **776** (2005) 52 [`cond-mat/0502663`]
- M. Molloy and B. Reed, *A critical point for random graphs with a given degree sequence,* Random Structures and Algorithms **6** (1995) 161-180.
- M. Molloy and B. Reed, *The size of the giant component of a random graph with a given degree sequence*, Combin. Probab. Comput. **7** (1998) 295-305.
- P.Orponen and S.E.Schaeffer, *Efficient Algorithms for Sampling and Clustering of Large Nonuniform Networks*, [`cond-mat/0406048`]
- Sood, V. & Grassberger, P., *Localization Transition of Biased Random Walks on Random Networks,* Phys.Rev.Lett **99** (2007) 098701

Bibliography in more detail (2)

- R.Bentley,M.Hahn, & S.Shennan, *Random Drift and Cultural Change,* Proc.R.Soc.Lon.B*, **271*** (2004) 1443
- M. Hahn & R.Bentley, *Drift as a Mechanism for Cultural Change: an example from baby names,* Proc.R.Soc.Lon.B*, **270*** (2003) S120
- R.A. Bentley, C.P. Lipo, H.A. Herzog, & M.W. Hahn, *Regular rates of popular culture change reflect random copying,* Evolution and Human Behavior ***28*** *(2007)* 151-158