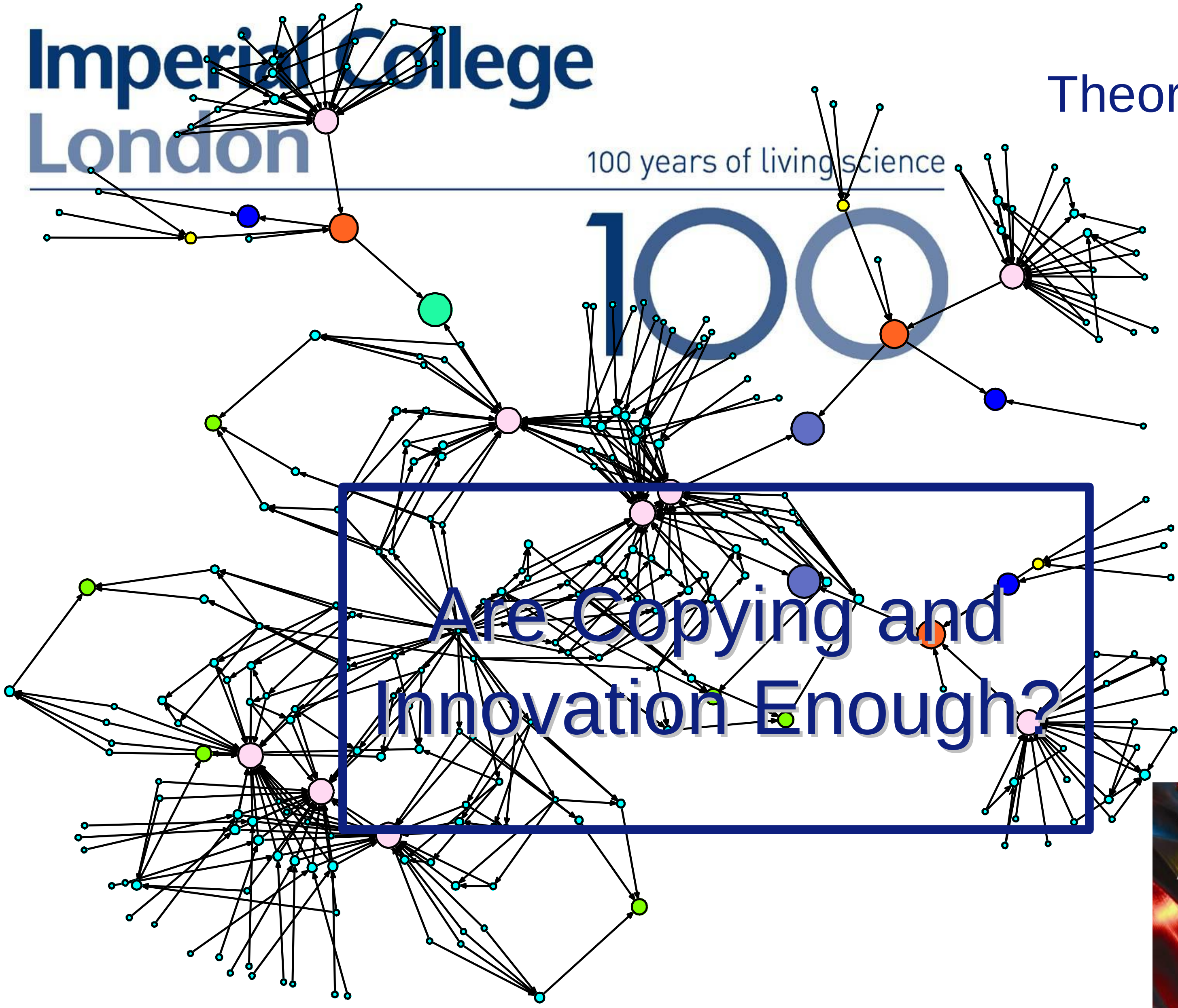
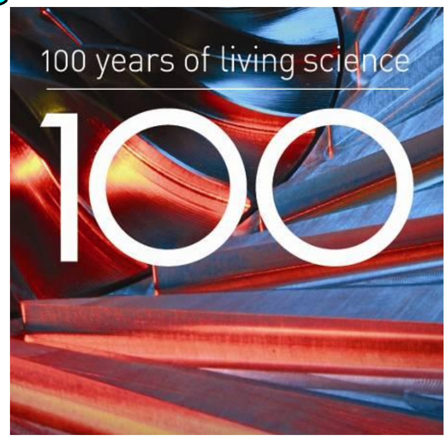
Brunel University 25th October 2007
Tim Evans
Theoretical Physics

Imperial College London

100 years of living science



Are Copying and Innovation Enough?



Are Copying and Innovation Enough?

Main authors on different aspects of work discussed here:-

- Tim Evans, Physics Dept.
- Doug Plato, Inst. Mathematical Sciences
- Tevong You, Physics Dept.

All above at Imperial College London

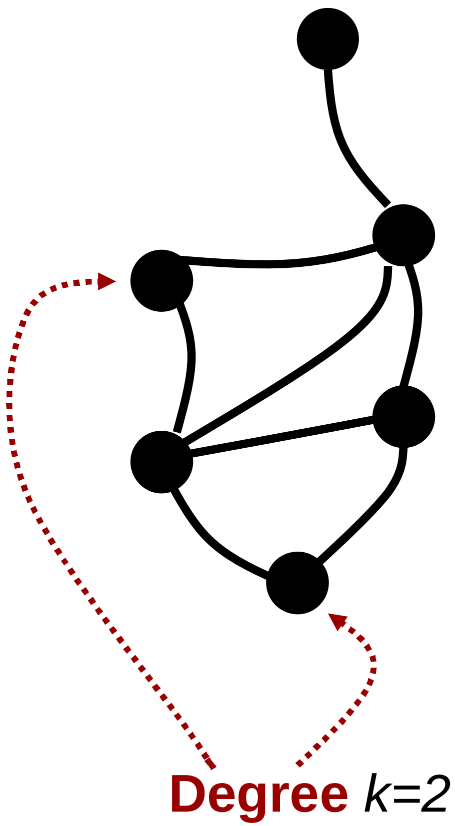
- Jari Saramäki, Helsinki University of Technology

Notation

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

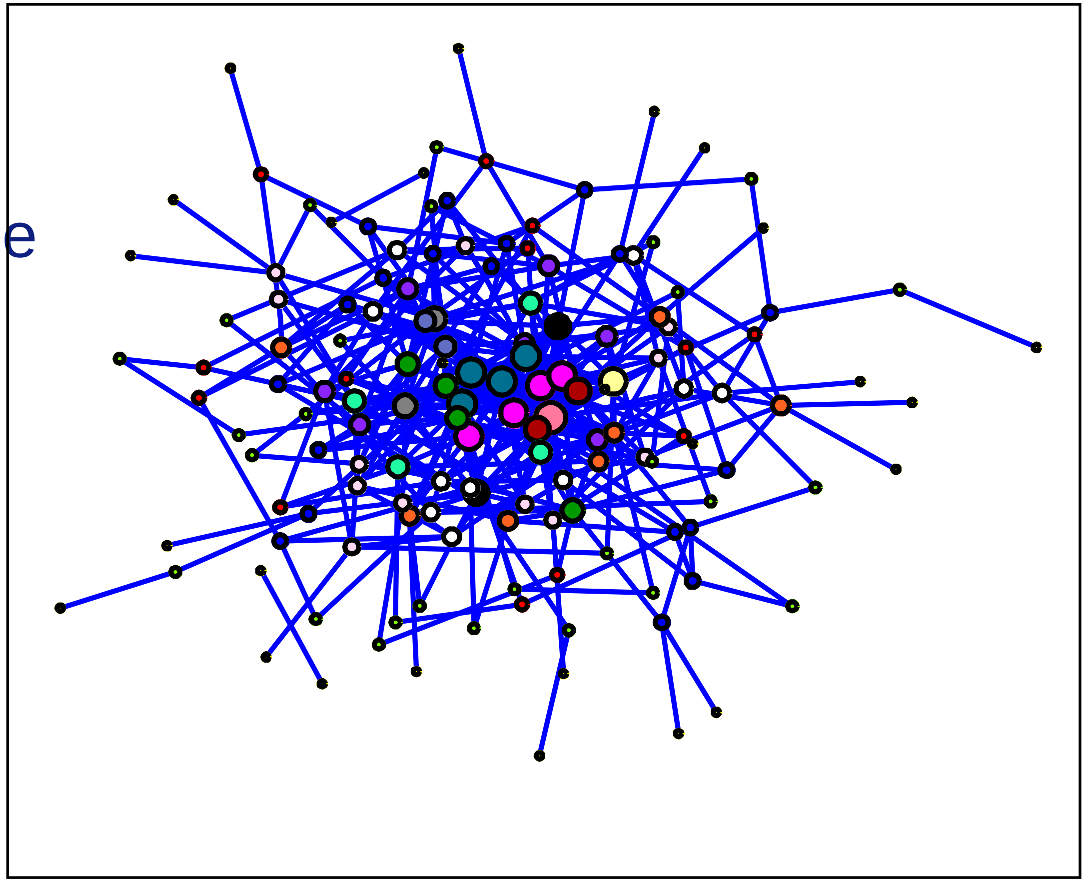
(no values or directions on edges, no values for vertices)

- ***N*** = *number of vertices in graph*
- ***E*** = *number of edges in graph*
- ***k*** = *degree of a vertex*
- ***<k>*** = *average degree* = ***2E/N***
- Degree Distribution
  n(k) = *number of vertices with degree* ***k***
  p(k) = ***n(k)/N*** = *normalised distribution*

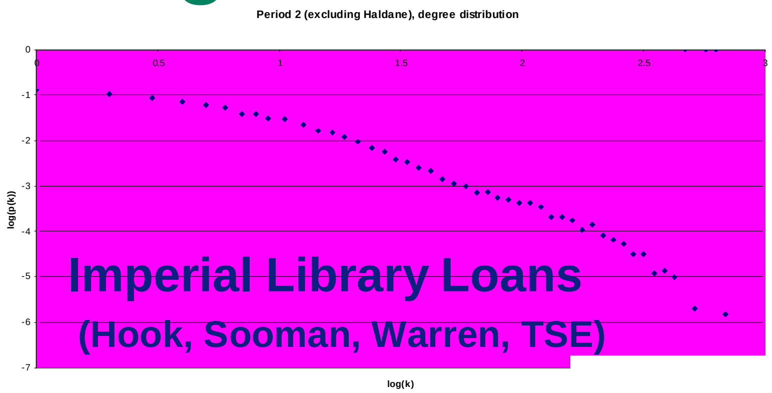


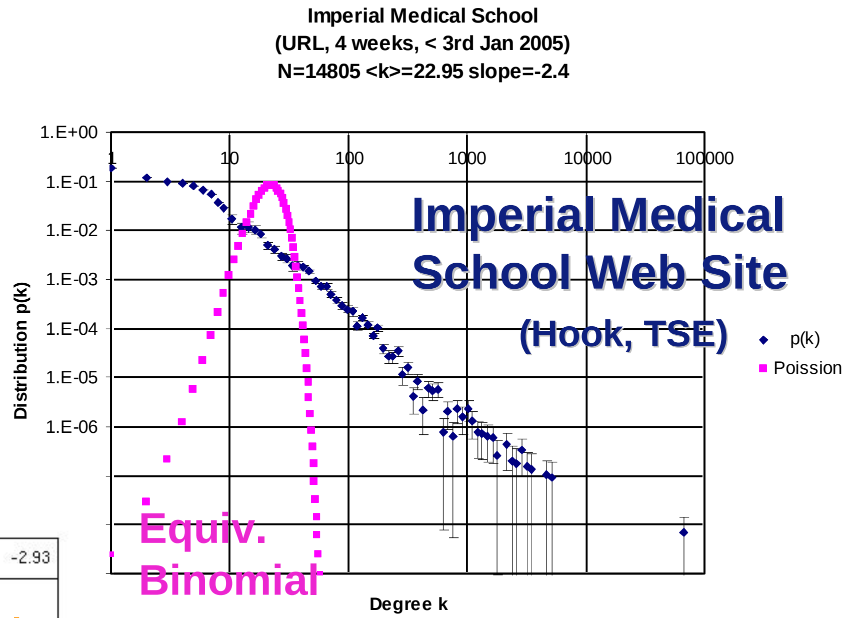
Growing Scale-Free Networks

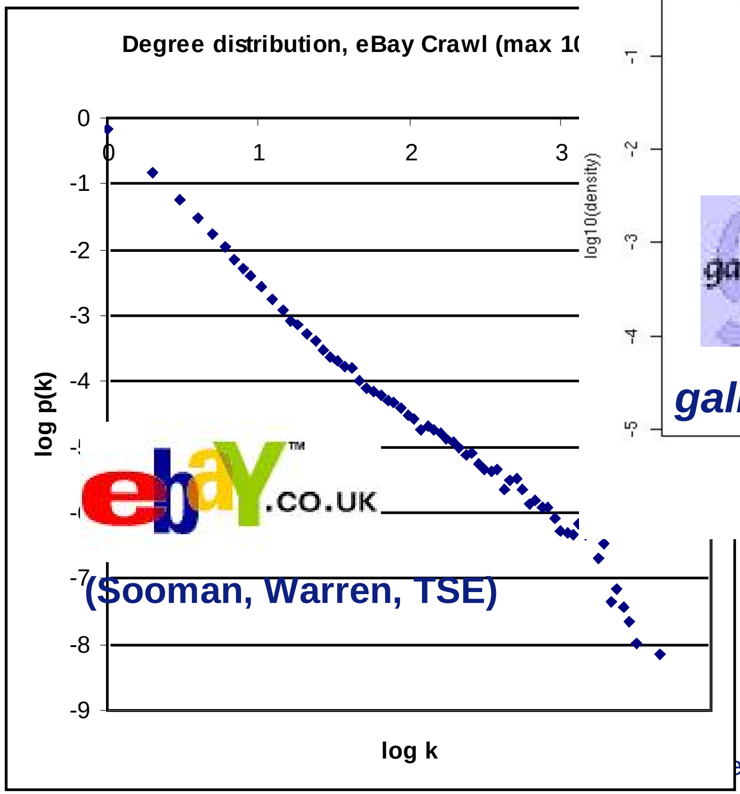
- Random Walks and Copying
  - The Origin of Scale-Free Networks?
- Copying Model
  - General Features
  - Equilibrium Solutions
  - Time Evolution
- Summary

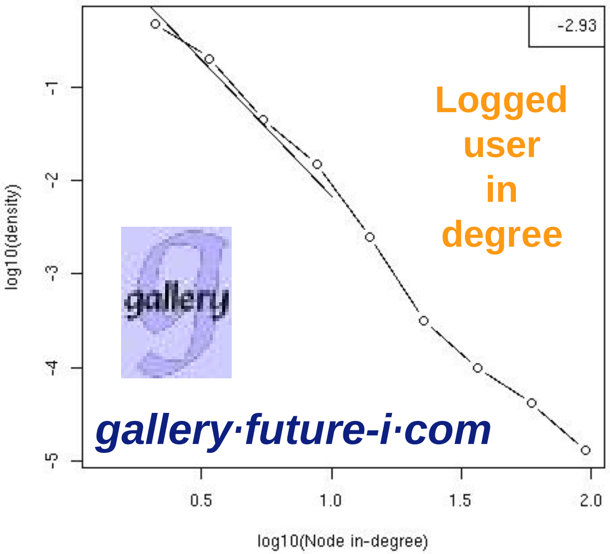


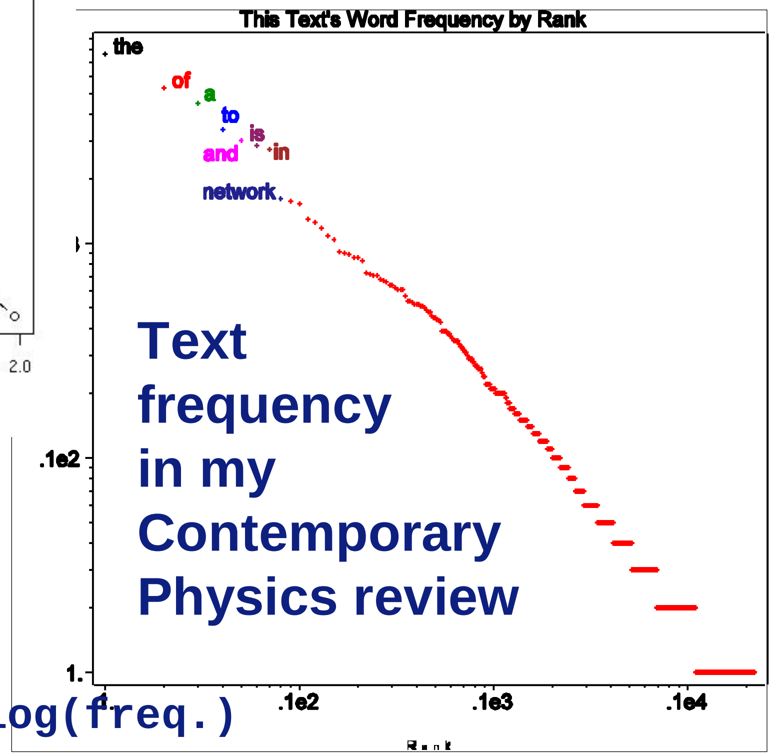
Long Tails in Real Data











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,

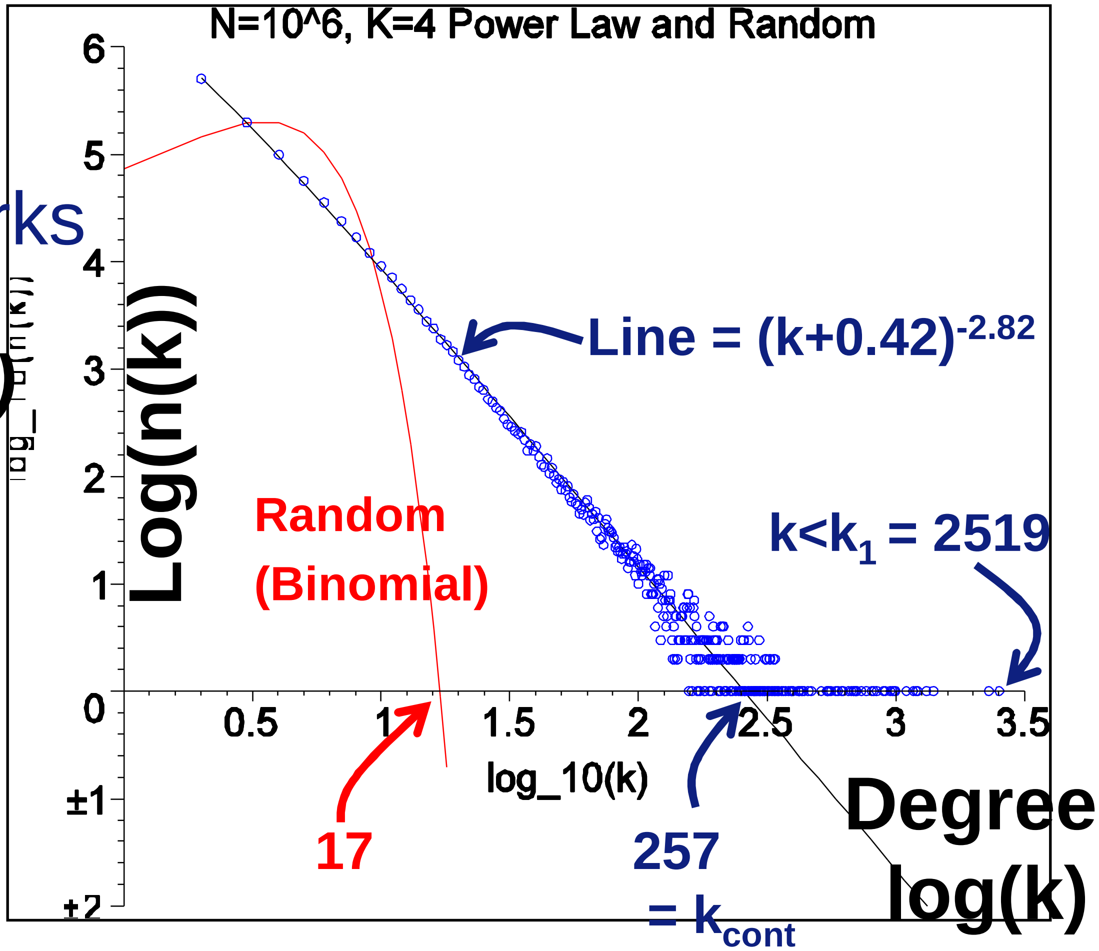
  $k \leq k_1 \sim O(ln(N))$

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

e.g. POWER LAW

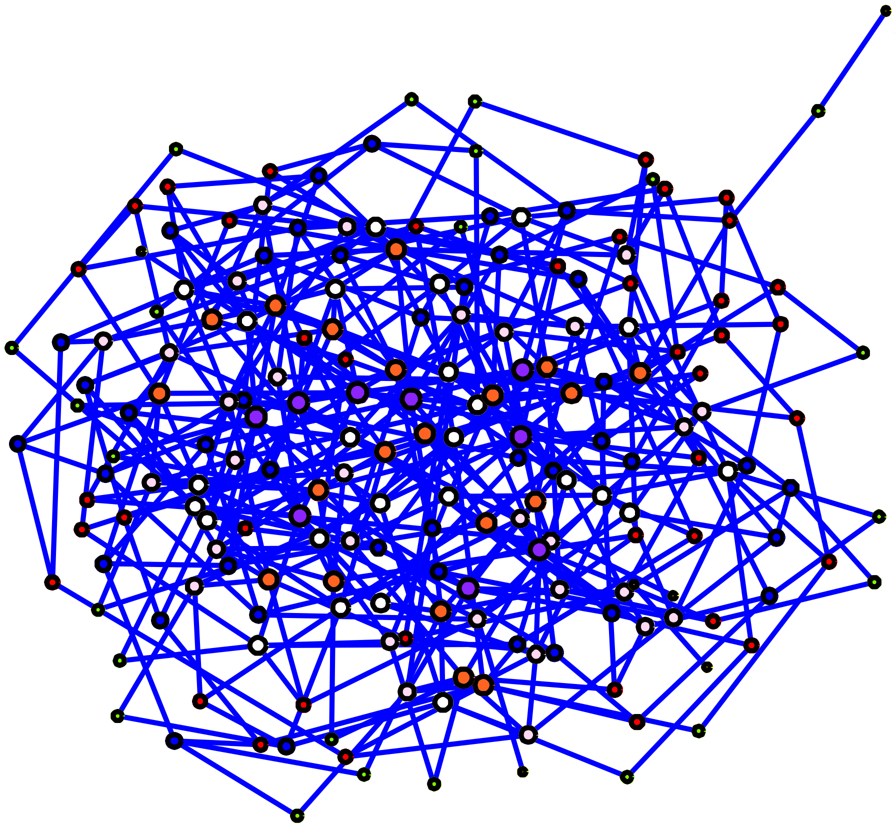
$n(k) \sim 1/k^3$

$k \leq k_1 \sim O(N^{1/2})$



N=200, <*k*>~*4.0*, vertex size ∝ ***k***

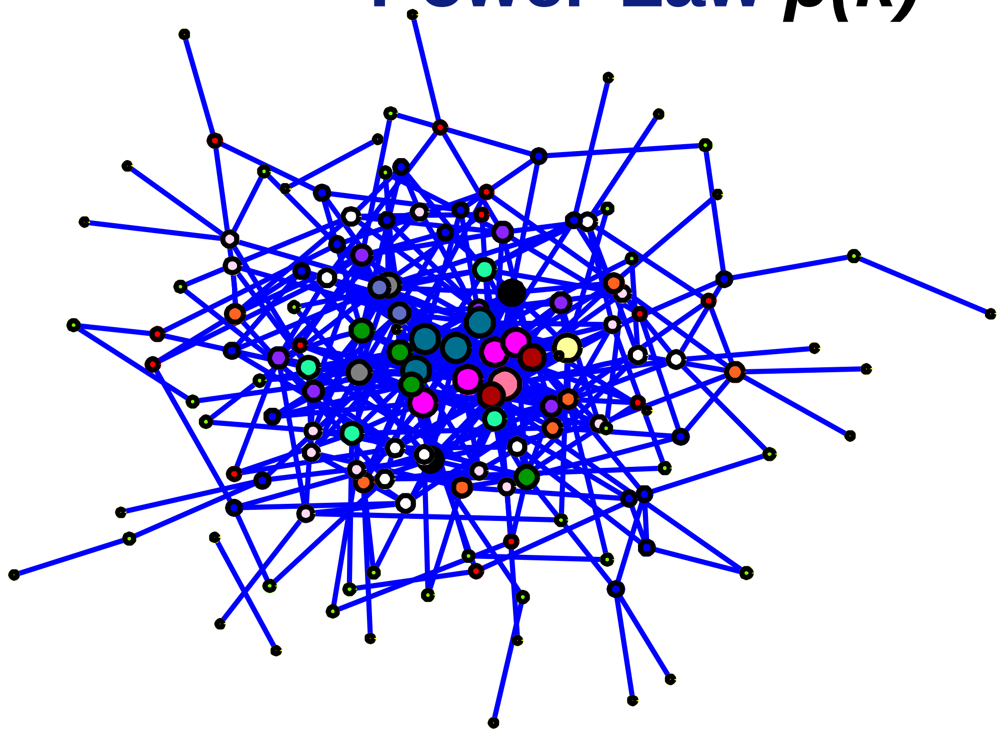
Classical Random



Diffuse centre of small degree vertices

Scale-Free

= Power-Law *p(k)*



Tight core of large hubs

Growth with Preferential Attachment

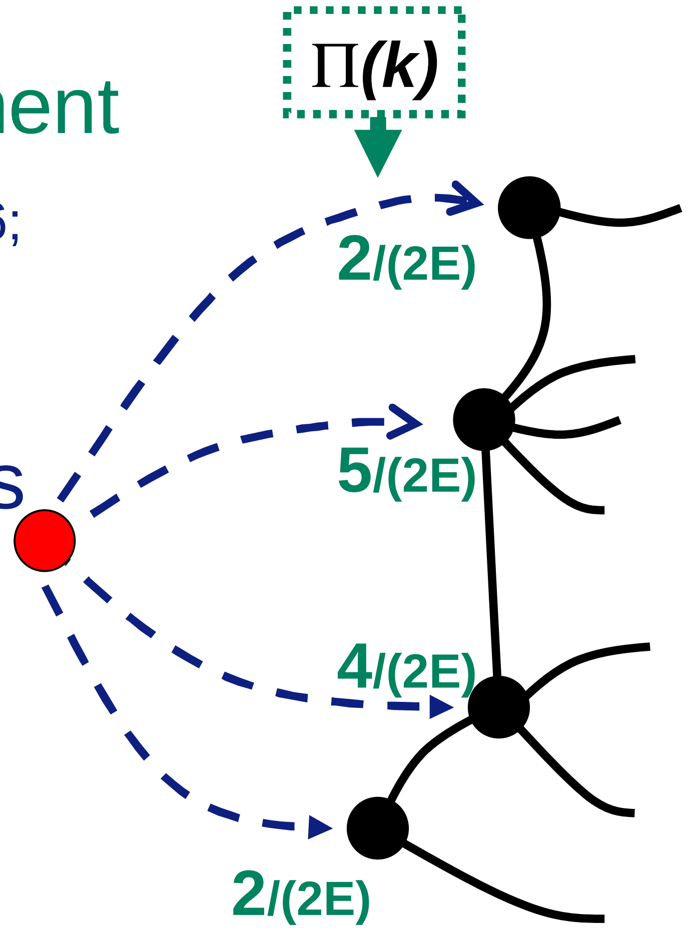
(Yule 1925, 1944; Simon 1955; Price 1965,1976; Barabasi,Albert 1999)

1. Add new vertex attached to one end of ½$\langle k \rangle$ 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

- Network not essential – ***k***=frequency of previous choices
- Generalised attachment probability

$$\Pi(k) = (1 - p_r)\frac{k}{2E} + p_r \frac{1}{N},$$

Preferential Attachment (first term) — Random Attachment (second term)

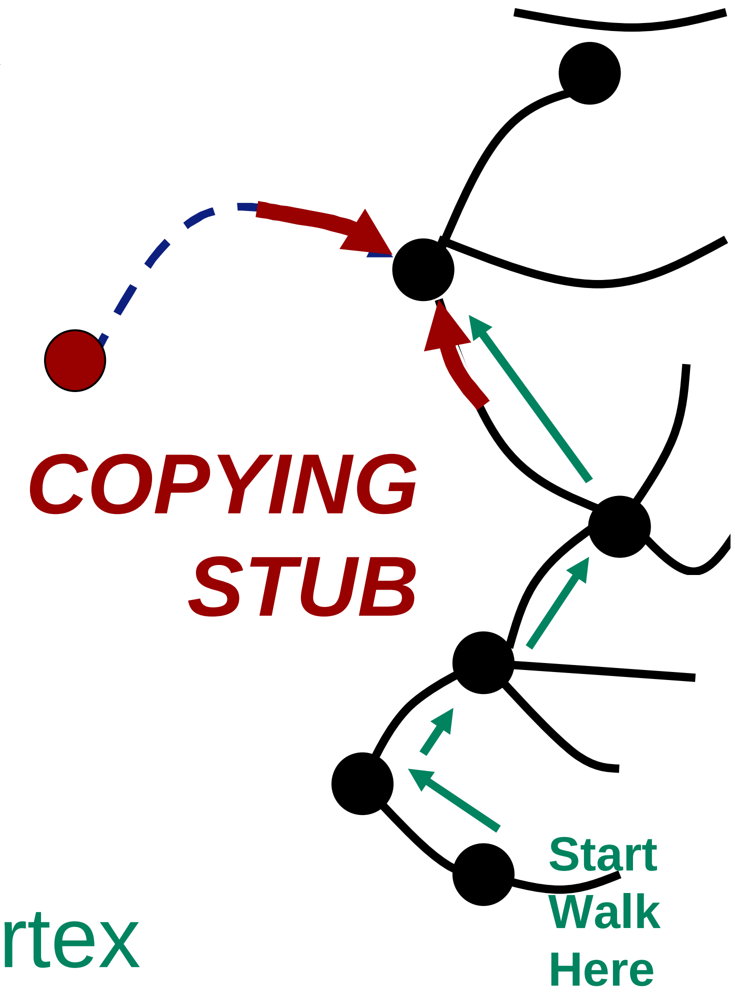
with new vertex added fraction ε of the time gives power laws with powers from ***2*** to ∞

- BUT if $\lim_{k} \Pi(k) \propto k^{\alpha}$ for any $\alpha \neq 1$ 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



è Probability of arriving at a vertex

$\propto$ number of ways of arriving at vertex

= *k*, the degree

$\Rightarrow$ Preferential Attachment γ **= *3***

(Can also mix in random attachment with probability p_r)

Naturalness of the Random Walk algorithm

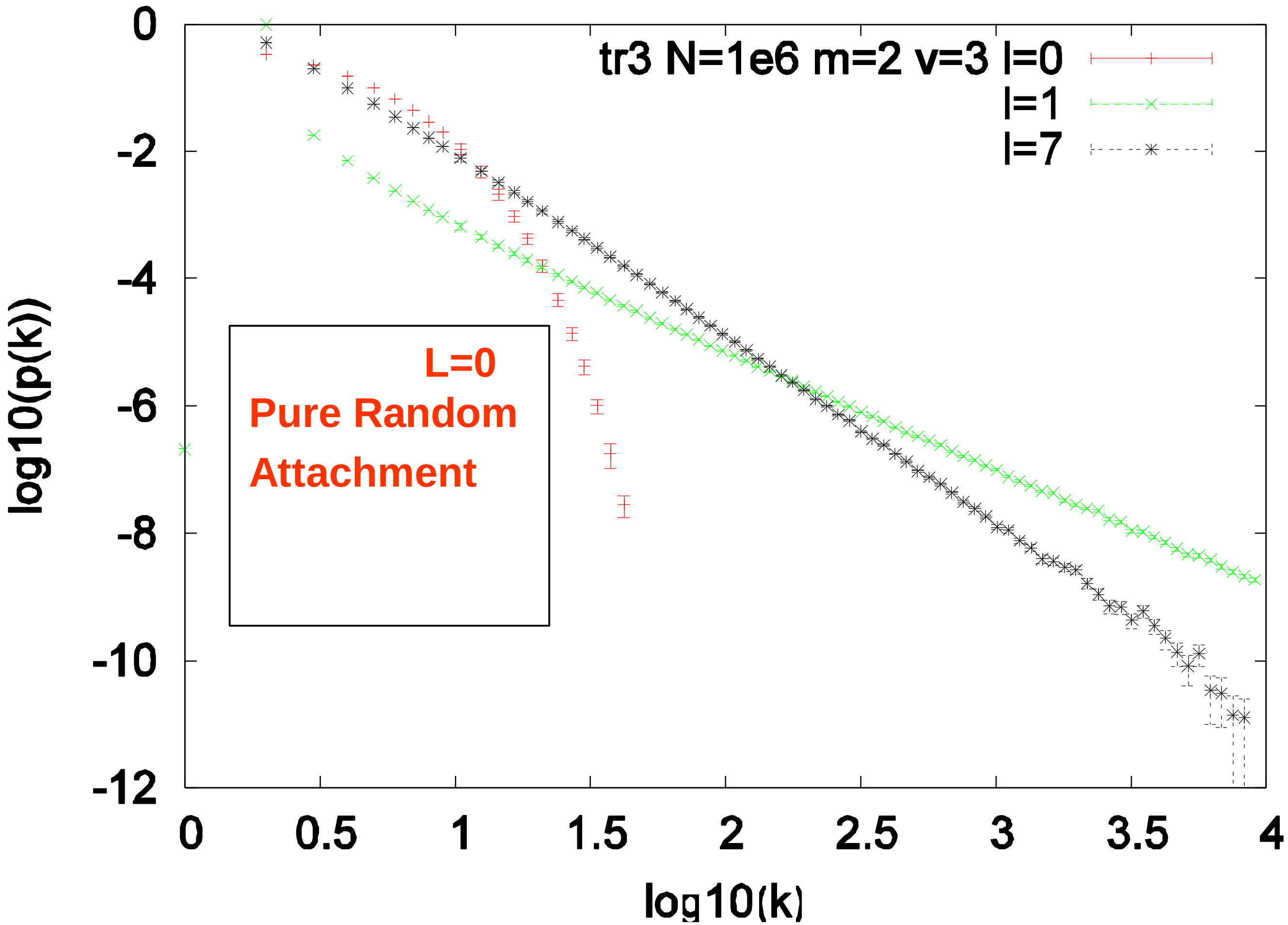
Automatically gives preferential attachment for any shape network and hence tends to a scale-free network

- 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

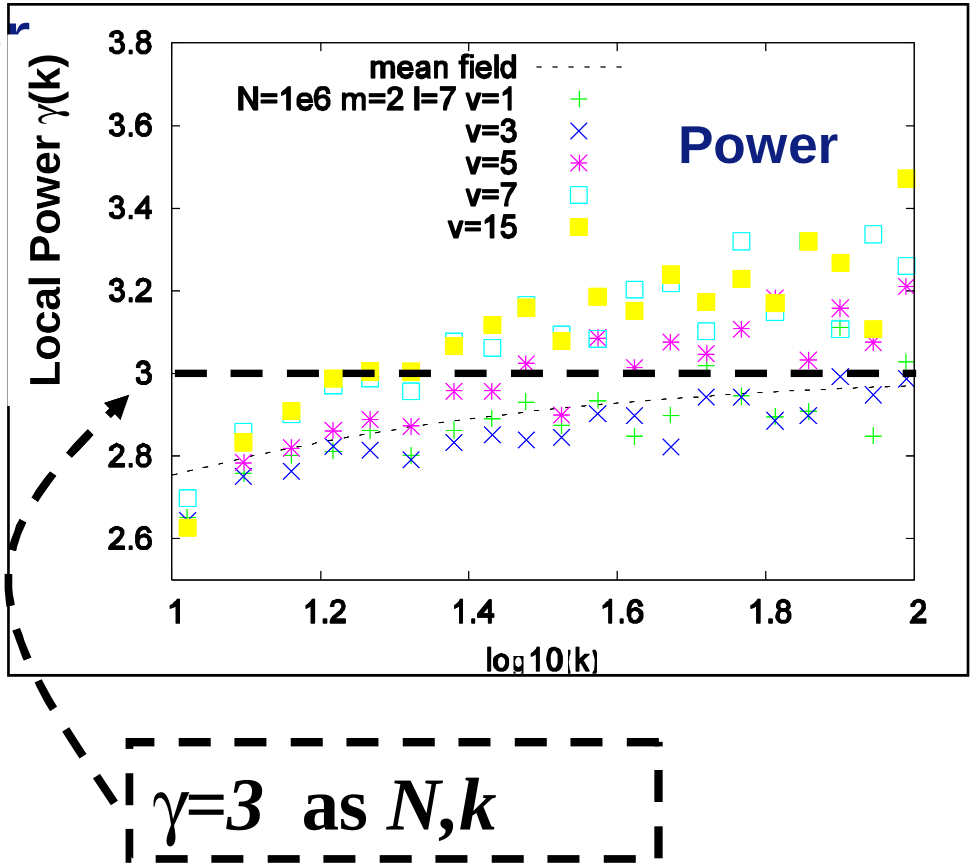
  $\Rightarrow$ Degree Correlation Length ξ ***< 1 < d*** (any distance scale)

Is the Walk Algorithm Robust?
YES

- Different starting points
- Vary length of walks per edge *keep* ***L*=<*L*>** *fixed*
- Vary edges added per vertex *keep* <***k***> *fixed*
- Allow multiple edges

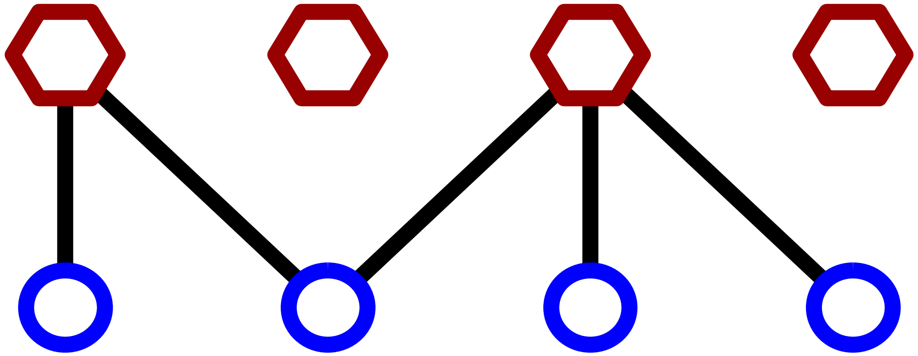
Good Power Laws
but power varies by
10% or 20%

Long Walk Variants, L=7



Copying Model – General Features

- Random Walks and Copying - The Origin of Scale-Free Networks?
- Copying Model
  - – General Features
  - – Equilibrium Solutions
  - – Time Evolution
- Summary



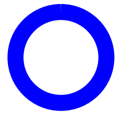
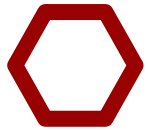
Copying

Copying is an intrinsically local process, no global information used yet produces distinctive macroscopic features.

- Preferential Attachment ⇔ Copying

  e.g via random walk (TSE+Saramäki 2005)
- Rewiring of Networks of fixed size (***N***, ***E***) vs. Growing Networks
- 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

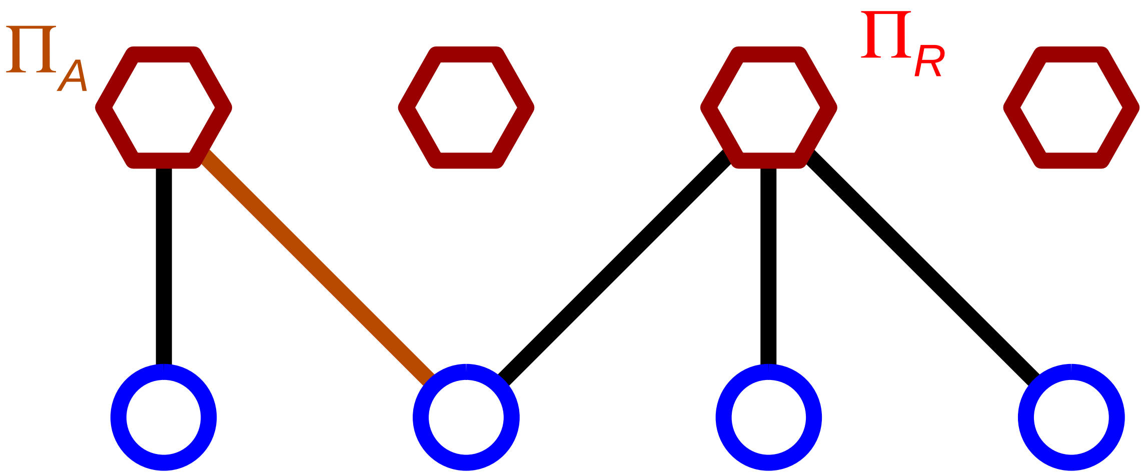
The Model as Network Rewiring

- **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

$$\Pi_A(k) = (1-p_r)\Pi_{\text{copy}} + p_r \Pi_{\text{innovate}}$$

copying probability   innovation probability

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



N artifacts

n(k) artifact degree dist.

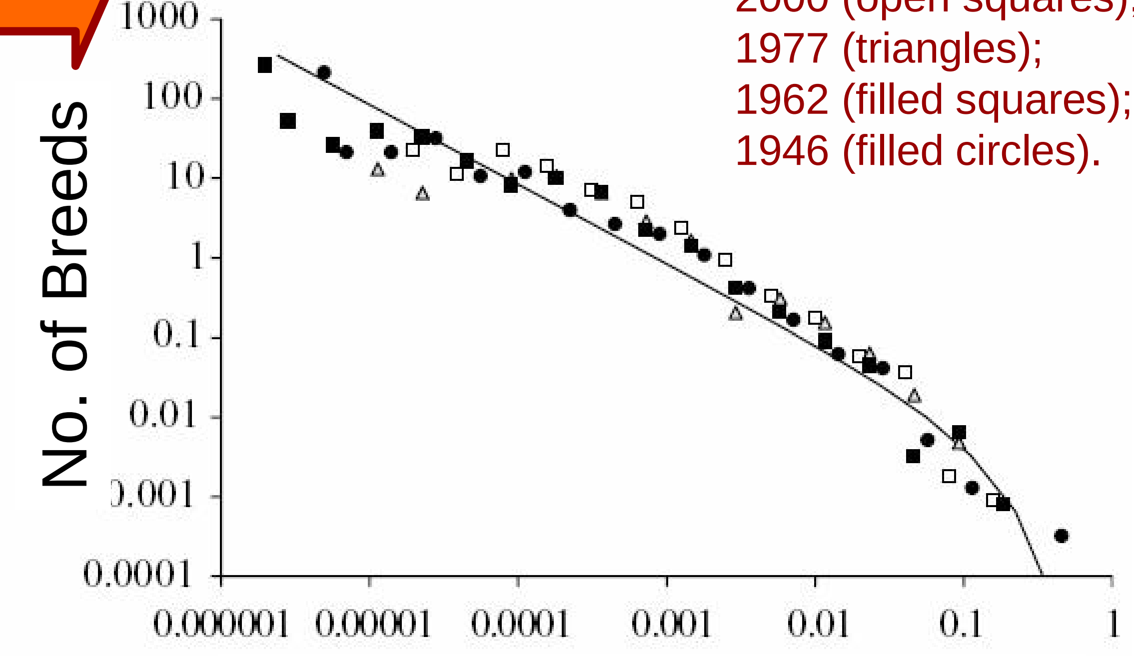
E individuals

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]

2000 (open squares);
1977 (triangles);
1962 (filled squares);
1946 (filled circles).



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 (genes not memes)
- Speciation ['Tangled Nature' Christensen et al 2002]
- Family Names [Zanette & Manrubia, 2001]
  - Inheritence and New Immigrants
- Language Extinction [Stauffer et al. 2006]
- Minority Game variant [Anghel et al, 2004]

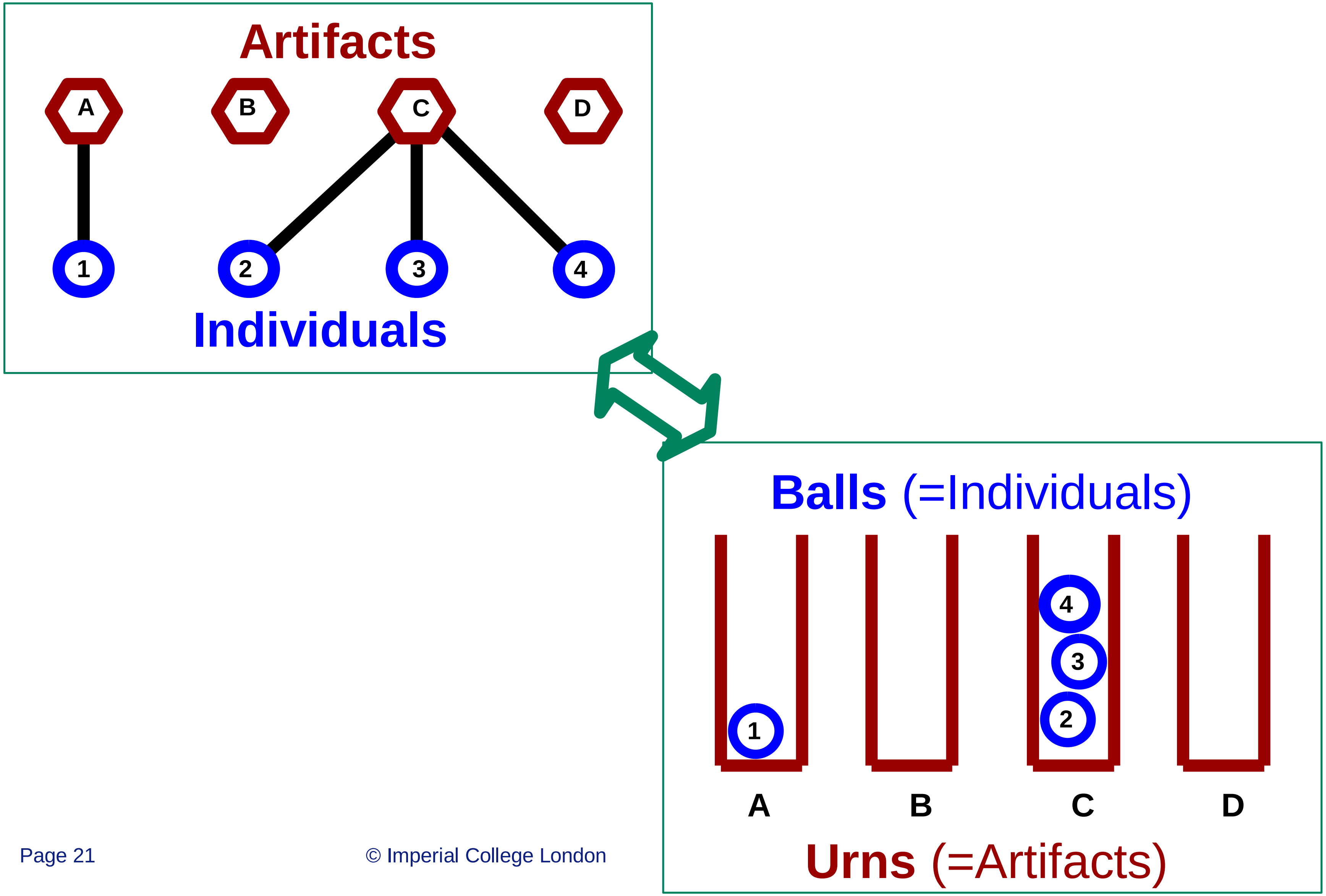
Relationship to Statistical Physics Models

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

Urn models as networks



Mean Degree Distribution Master Equation

Usually one also uses a mean field *approximation* very accurate for many models (low vertex correlations)

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

(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 Degree Distribution equation be exact?

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

Normalisation of probabilities not usually same for different instances ***i***

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

Random or preferential attachment only

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

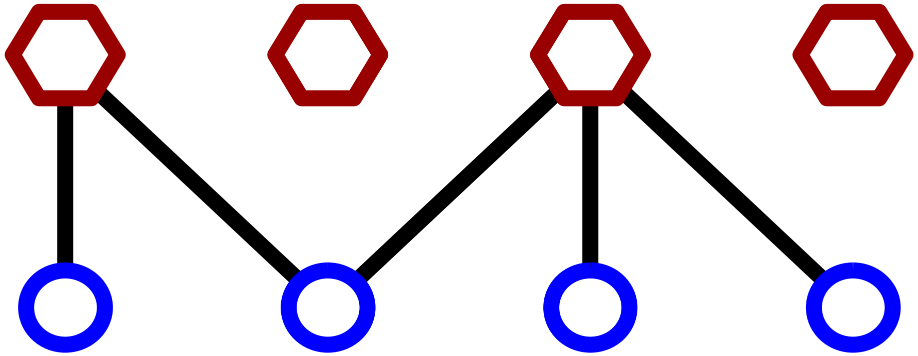
$$a = \frac{p_r}{p_p}\langle k \rangle, \quad b = -E, \quad c = 1 + a + b - \frac{p_r}{p_p} E$$

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

Copying Model – General Features

- Random Walks and Copying - The Origin of Scale-Free Networks?
- Copying Model
  - General Features
  - Equilibrium Solutions
  - Time Evolution
- Summary



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 if* *<k> << 10*

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

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

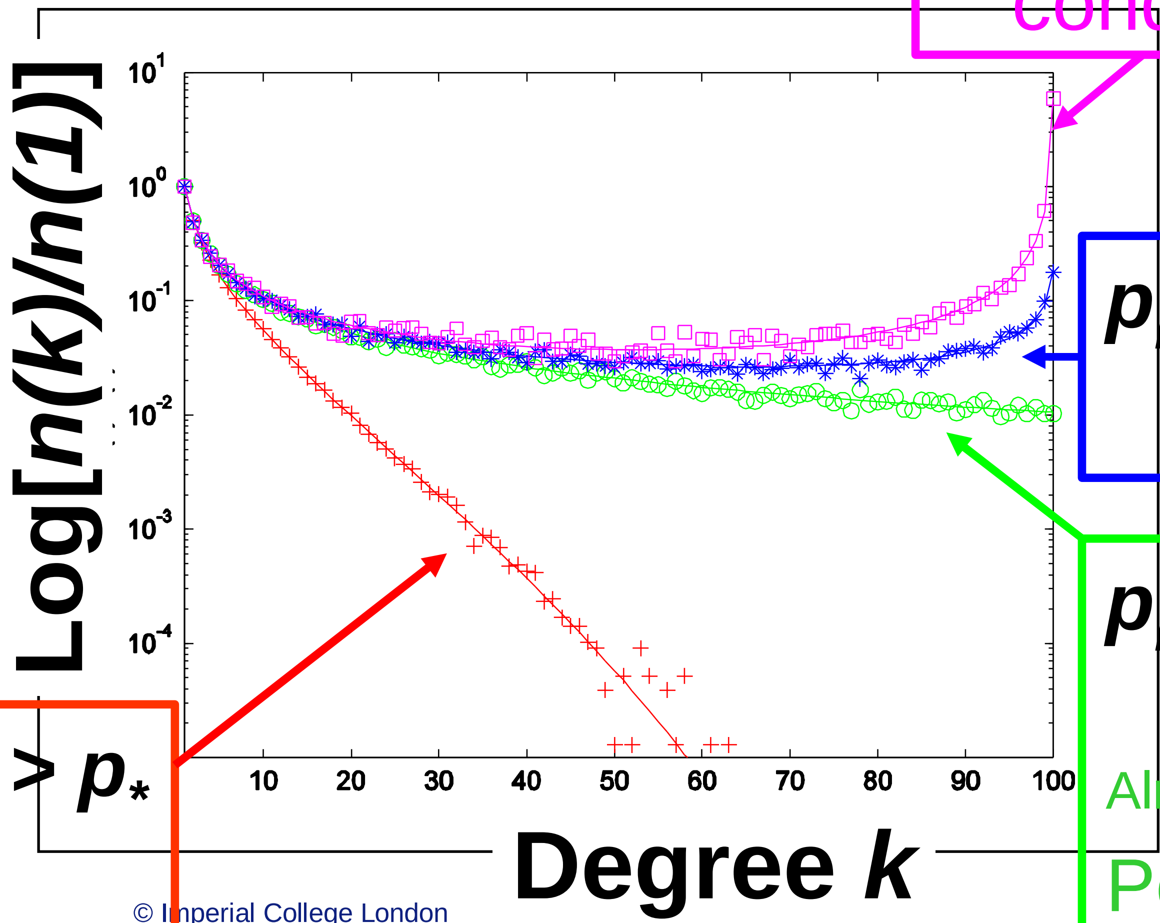
$$n(k) = A\left(\frac{\Gamma(k+\bar{K})}{\Gamma(k+1)}\right)\left[\frac{\Gamma(E-\bar{E}-\bar{K}-k)}{\Gamma(E+1-k)}\right]$$

Equilibrium Behaviour Results

N=E=100

Points: *10^5* data runs

Lines: exact mean field solution



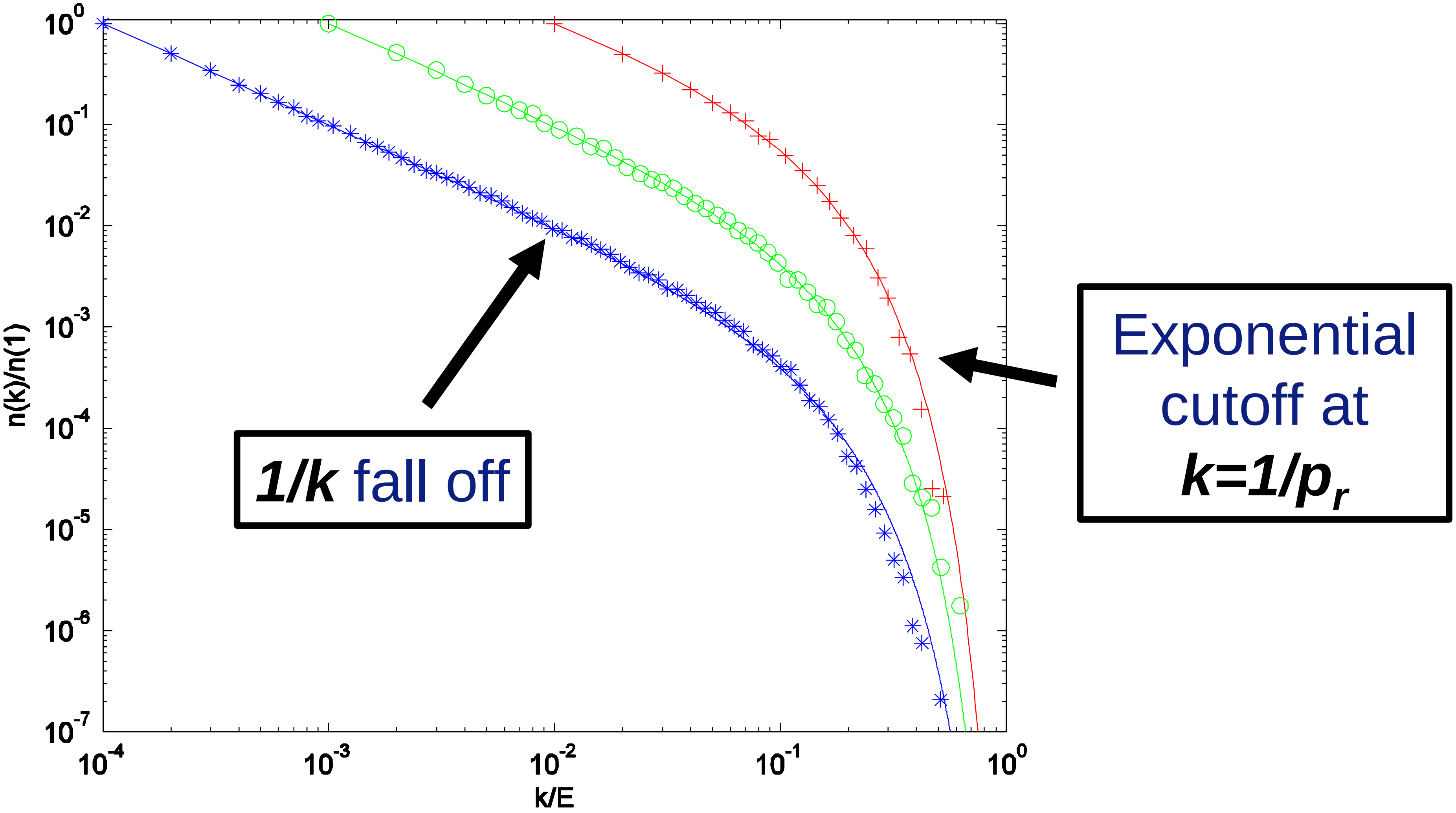
$p_r = 0.001 < p_*$
condensate

$p_r = 0.005 < p_*$

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

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

Log-Log plot of typical ***p(k)***



Tangled Nature

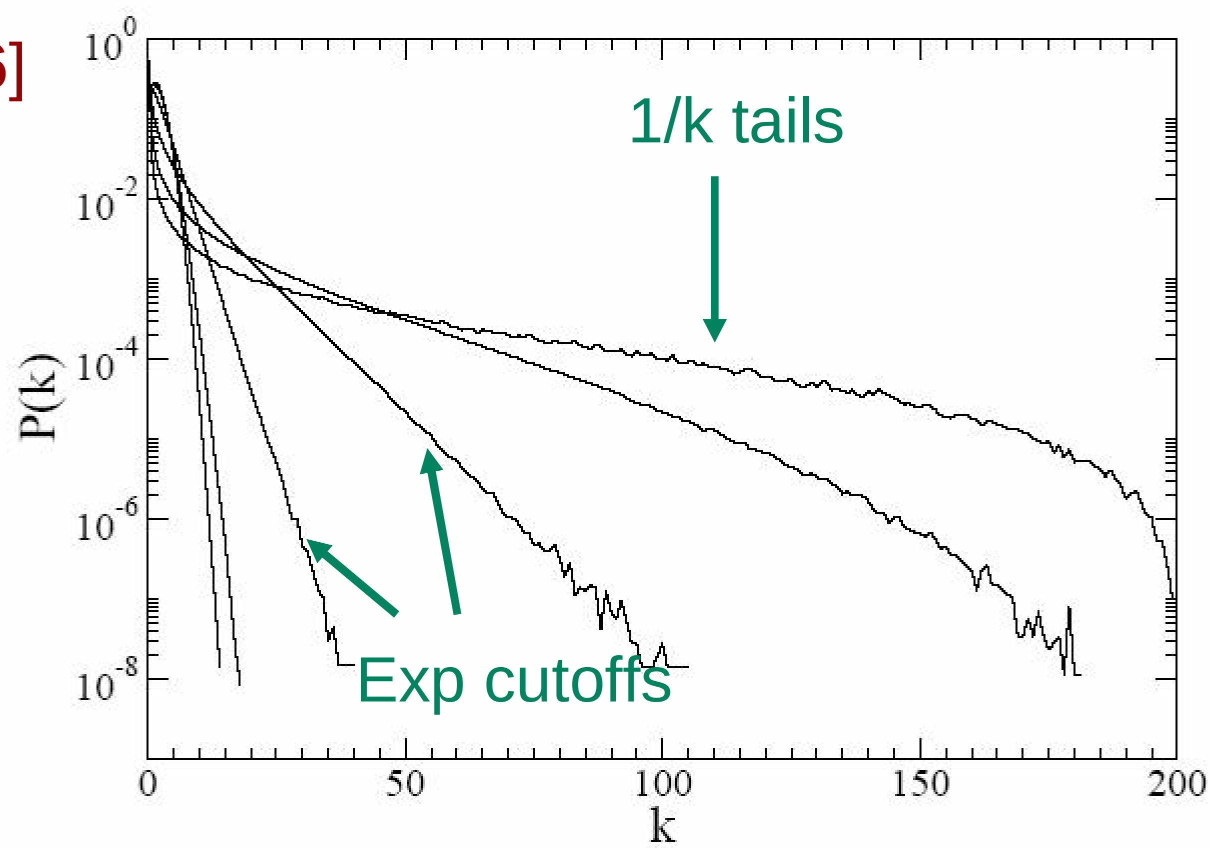
- Tangled Nature model of speciation with a particular focus on extinctions and intermittency in extinction patterns [Christensen et al, 2002; Laird & Jensen 2007]

- Each species (= artifact) identified by genes. Species die and reproduce (with inheritance/copying) and mutation (random attachment). Reproduction probability depends on interaction with other species.

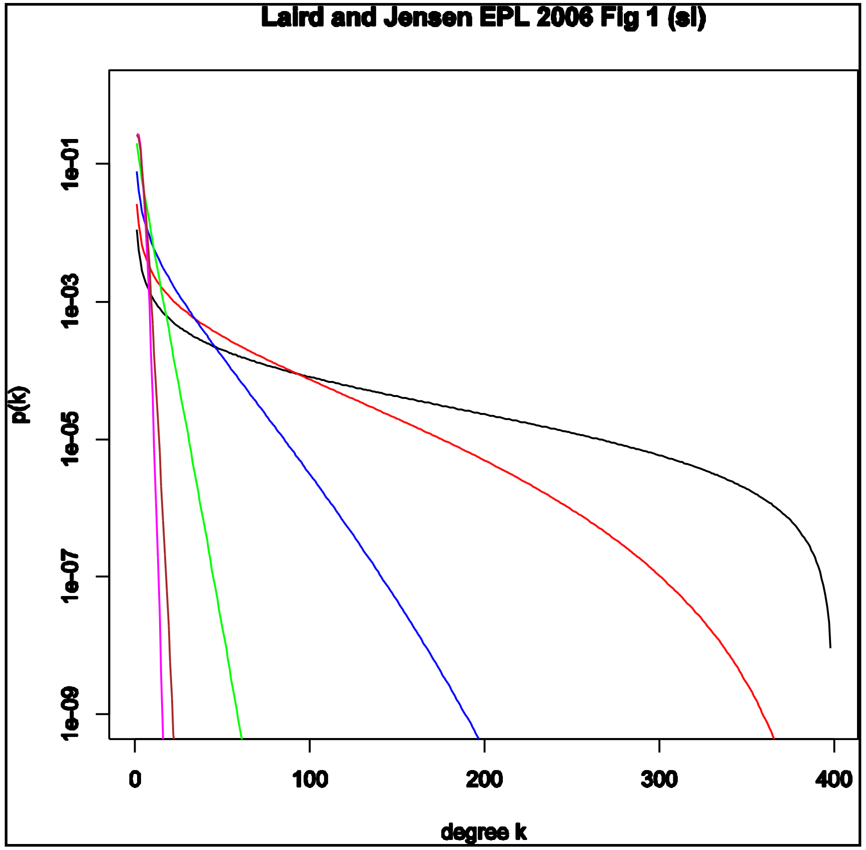
Network Version of Tangled Nature

- Simplified network model of Tangled Nature deletes whole nodes (species) and creates new ones with some inheritance = copying of old links
  [Laird, Jensen 2006]

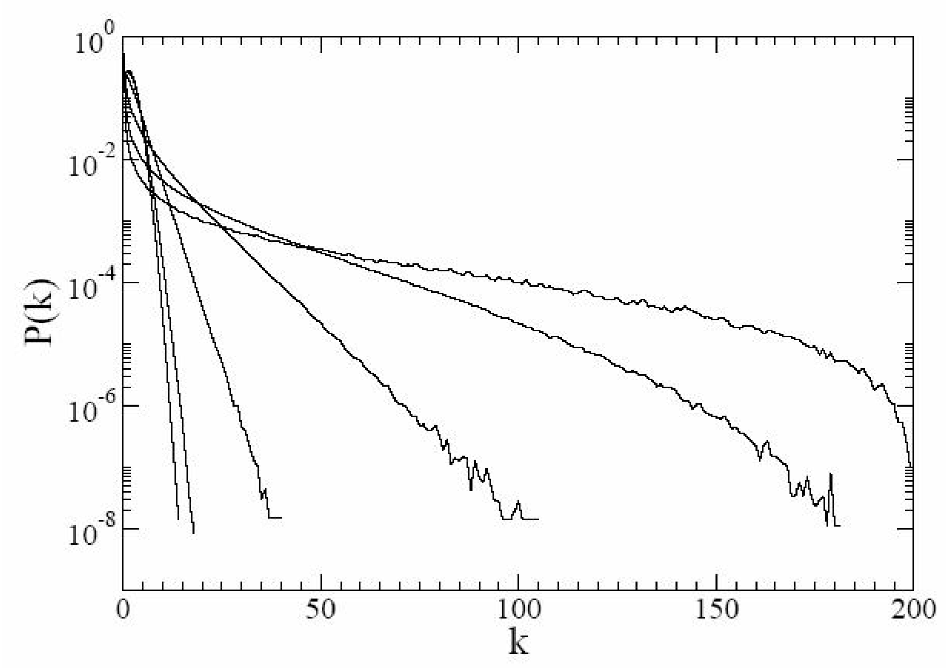
  Shows same ***p(k)*** as model here



Network Version of Tangled Nature (2)



This Model



Laird & Jensen 2006

Minority Game - 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 – Leaders and Followers

Question:

How many individuals - *followers* – are using the same strategy, one which belongs to a particular individual – a *leader*?

Minority Game - Leaders and Followers

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

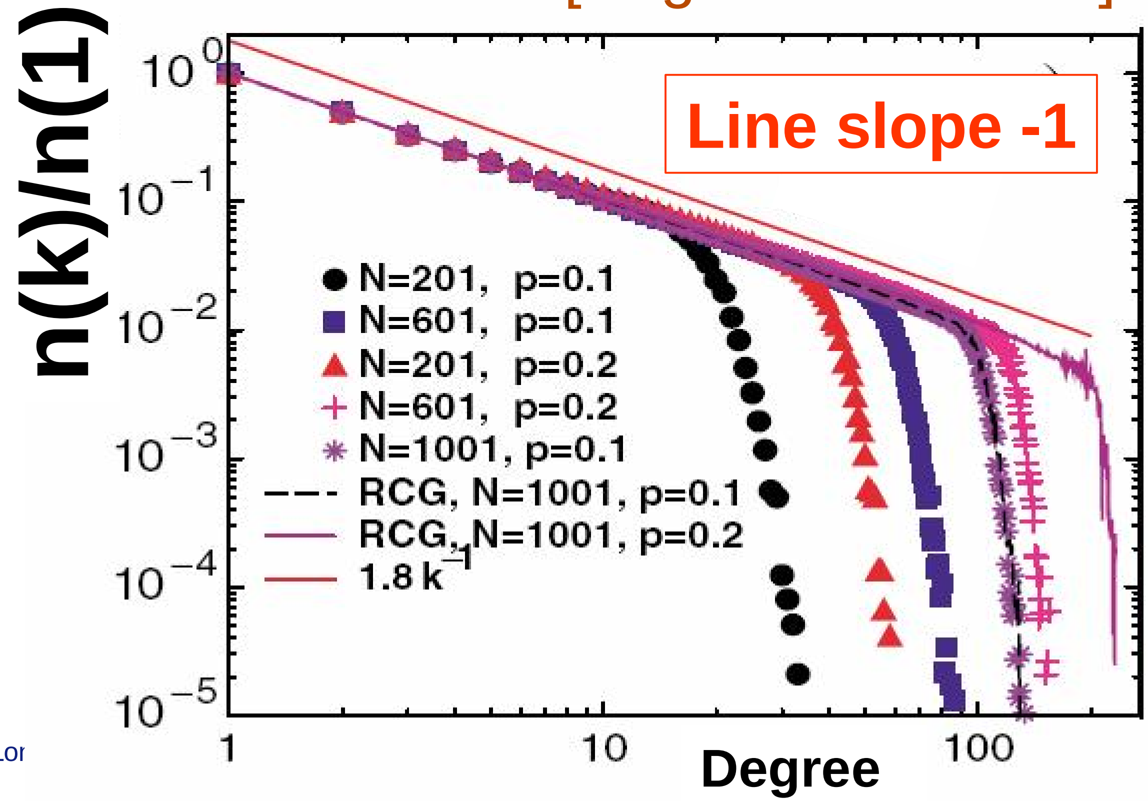
Various system sizes and various ER random graphs.

[Anghel et al. 2004]

Result exactly as in our model

Strategies = Artifacts

⇒ Simple copying of strategies

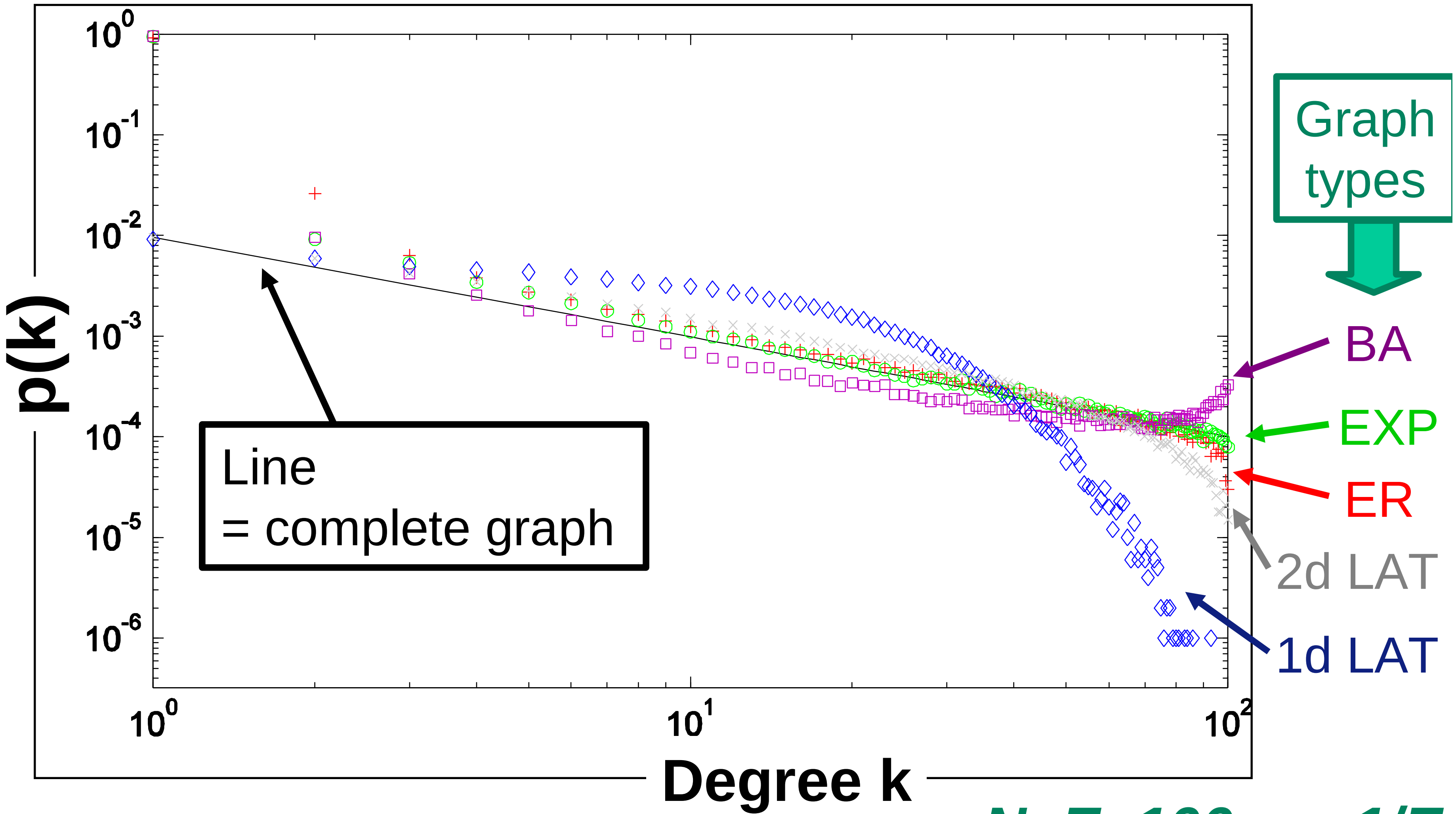


Minority Game Example - Leaders and Followers

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

Equilibrium with a Network of Individuals

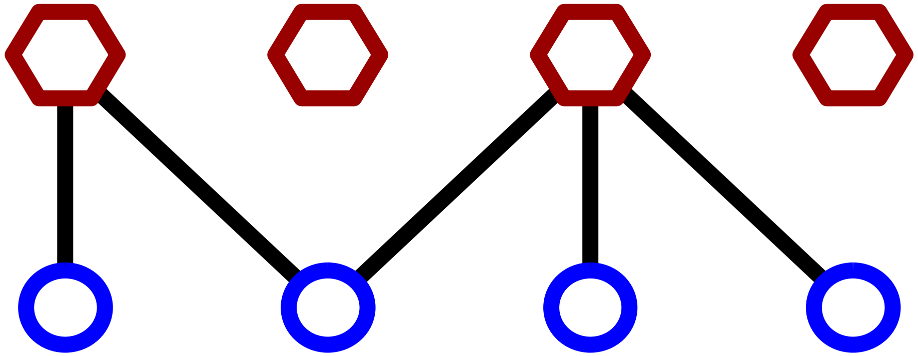
Qualitative behaviour largely unchanged except for 1d Lattice



N=E=100, p_r=1/E

Copying Model – General Features

- Random Walks and Copying - The Origin of Scale-Free Networks?
- Copying Model
  - General Features
  - Equilibrium Solutions
  - Time Evolution
- Summary



Exact Time Evolution Known

Exact solution for generating function known at all times and any finite parameter in terms of standard functions

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

constants c_m fixed by initial conditions

Eigenvalues

Eigenfunctions are ratios of Gamma functions

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

Homogeneity Measures F_n

***n*-th derivatives of generating function gives measures of homogeneity**

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

- **These are simple known ratios of Γ functions**
- **Related to *m*-th moments μ_n via Stirling numbers**

$$F_n(t) = N \frac{\Gamma(E+1+n)}{\Gamma(E+1)} \sum_{m=0}^{n} S_m^n \mu_m$$

F_n Homogeneity Measures

$$F_n(t) = \sum_{k=0}^{E} \frac{k}{E}\frac{(k-1)}{(E-1)} \cdots \frac{(k-n+1)}{(E-n+1)} \mathrm{n}(k)$$

- **These equal the probability of choosing *n* different individuals connected to the same artifact**

$\Rightarrow$ **$F_n = 0$ if no artifact chosen more than once**
$F_n = 1$ if all individuals attached to same artifact

Time Dependence and F_n Homogeneity Measures

F_n = probability that ***n*** 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

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

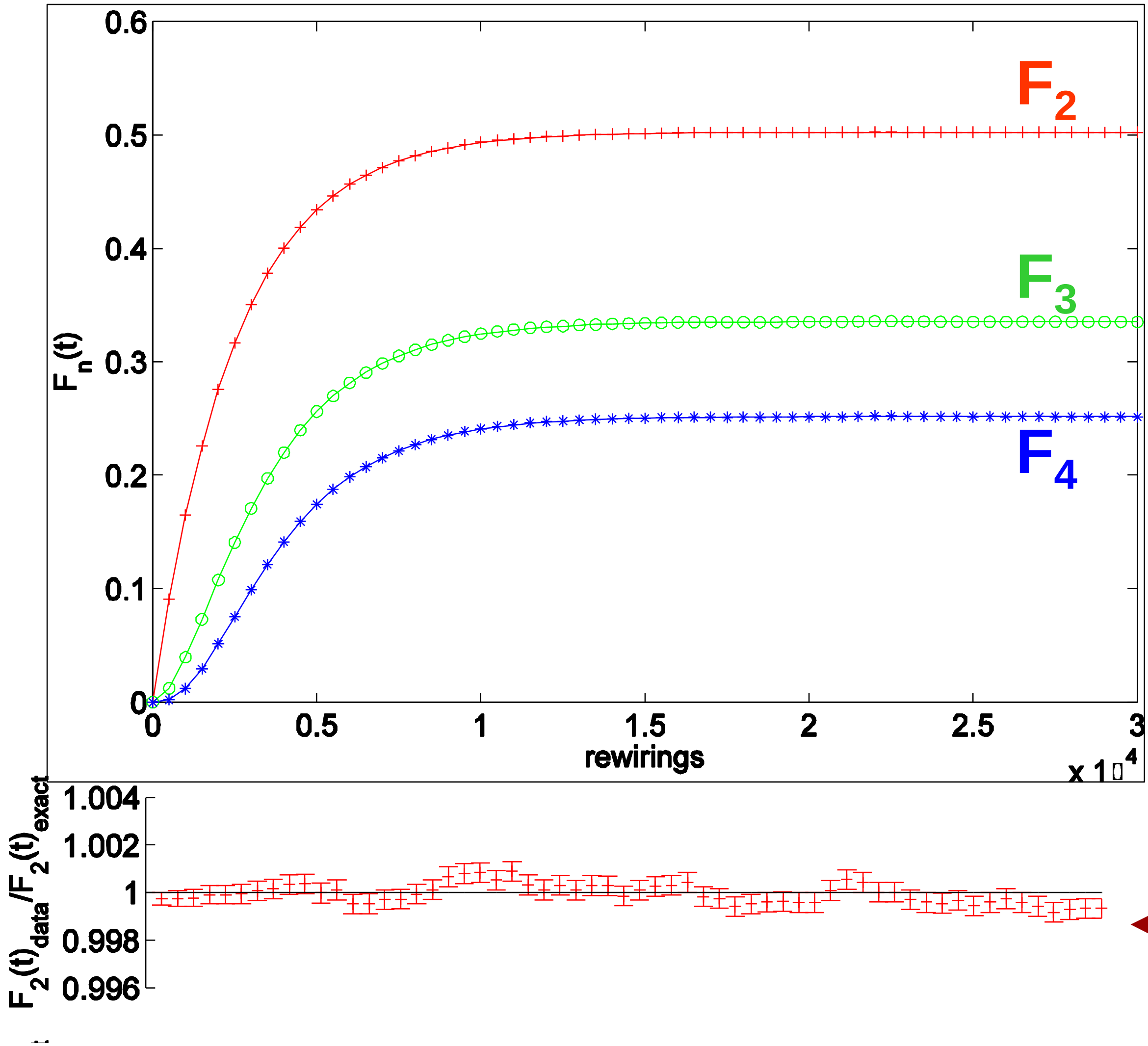
$\approx [2(p_r/E) + 2(1-p_r)/E^2]^{-1}$

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

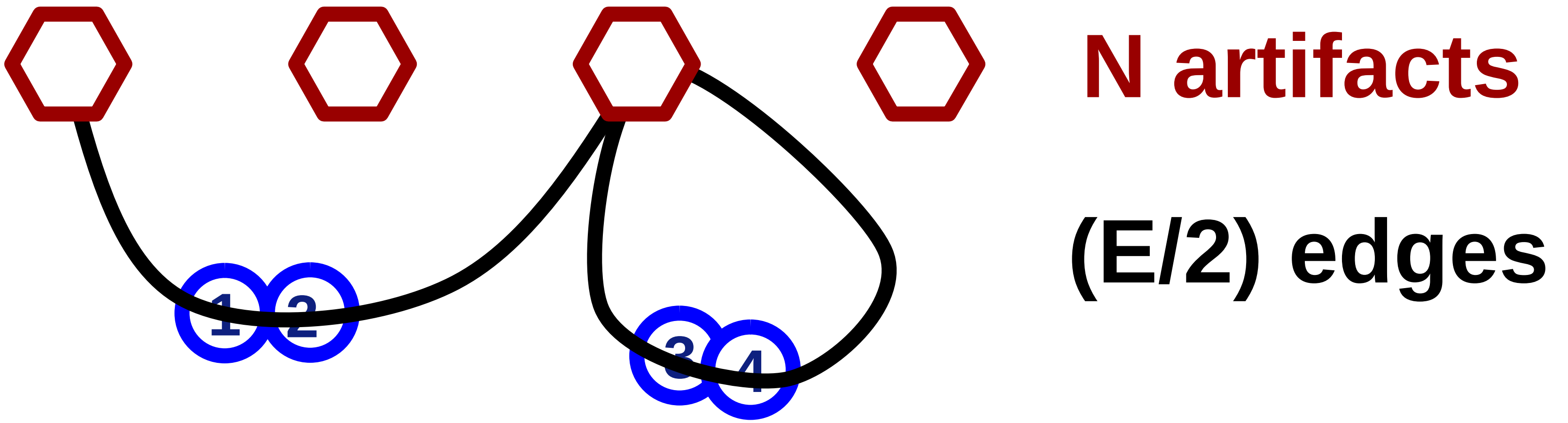


F_n increase as homogeneity increases with time

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



⇒ This is a *Molloy-Reed* projection

Graph Transition in Real Time

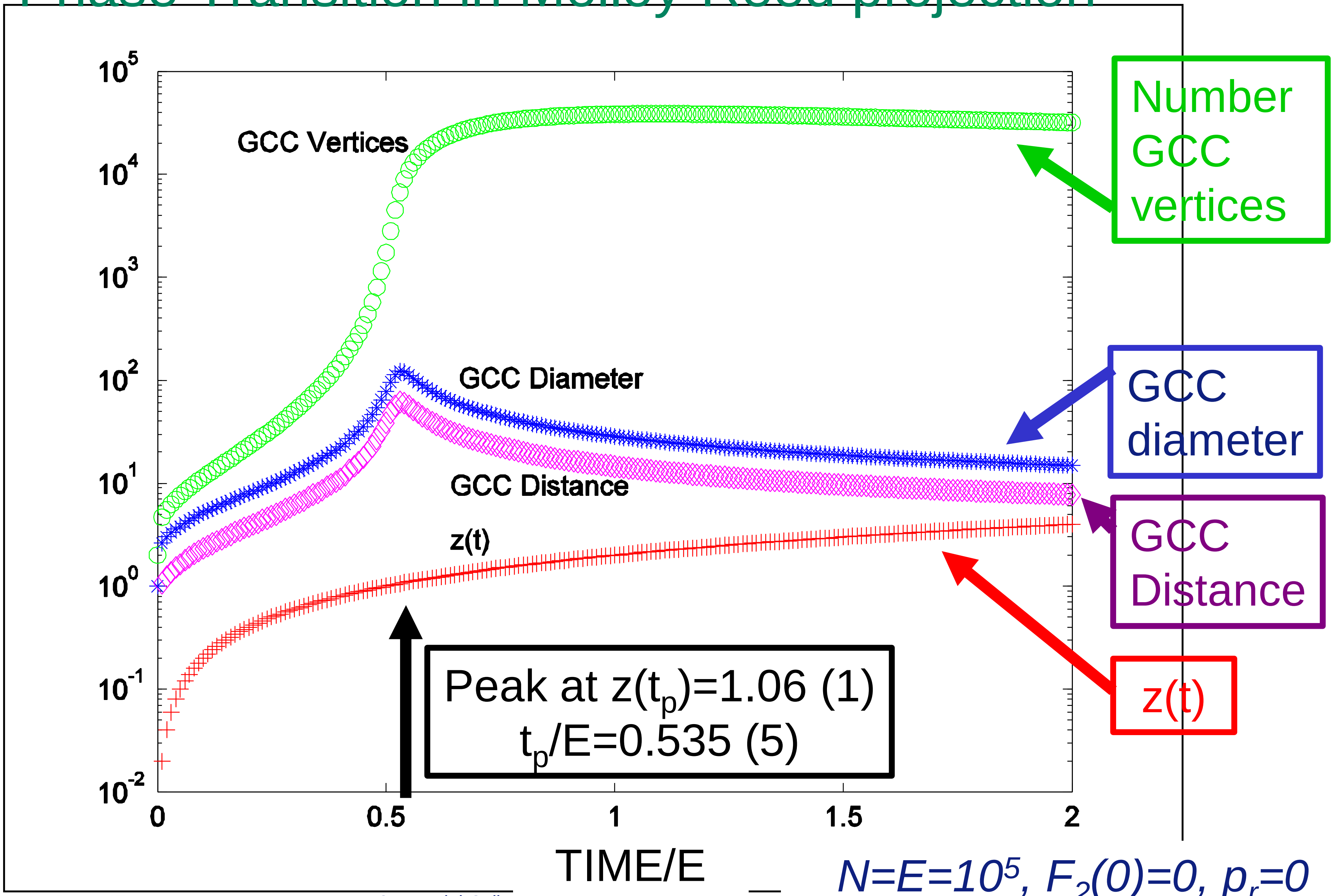
Infinite Random Graphs have a phase transition (e.g. appearance of **GCC** - Giant Connected Component) at ***z(t)=1*** where

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

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

F_2 is the probability that two randomly chosen stubs are attached to the same (artifact) vertex

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)

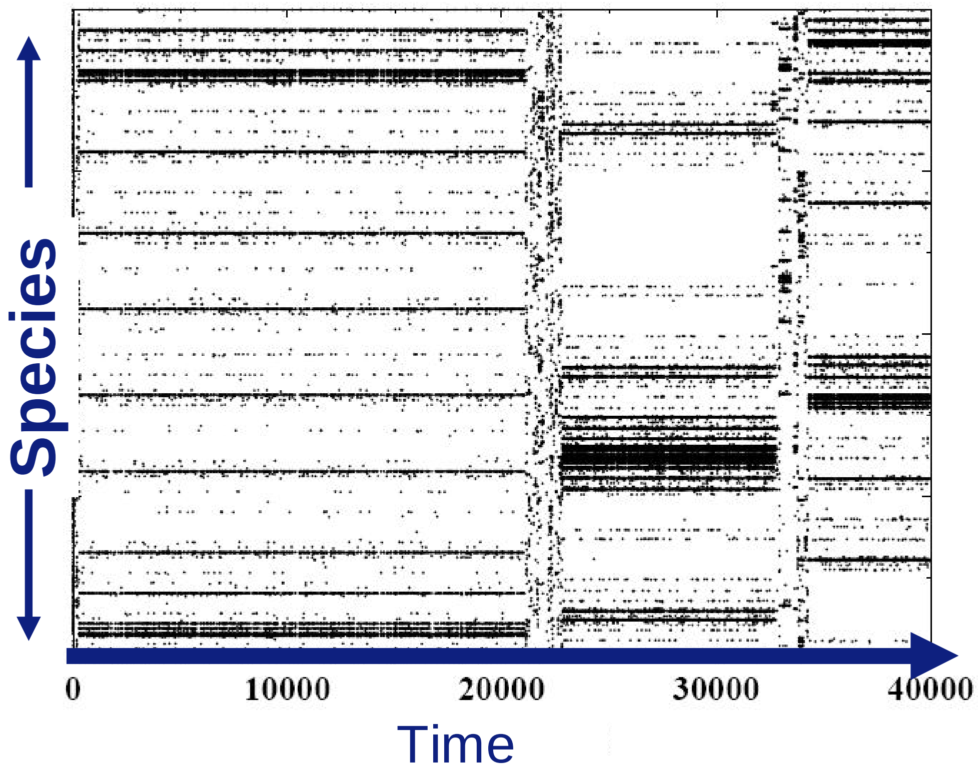
$\Rightarrow$ **Finite size effects clearly present**

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

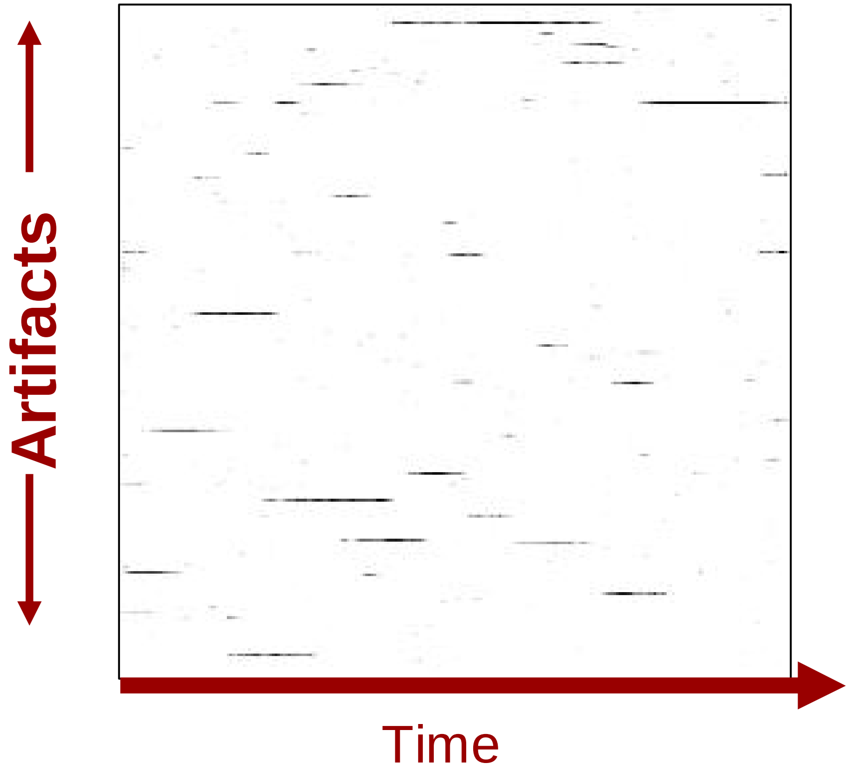
Tangled Nature

Tangled nature models can show periodic extinctions’

Simple copying model seems to show show less stability



Hall et al. `cond-mat/0202047`



Generalisations of Model

- Different ways to update the model
  – Exact analytic results still possible
- Add a graph to the individual vertices
  – choose who to copy using individual's network
- Add a graph to the artifact vertices
  – mutations/innovations limited by metric in an artifact space
- Different types of individual
  – update their choice and copy/innovate at different rates

Summary

- Random Walks and Copying - The Origin of Scale-Free Networks?
- Copying Model
  - General Features
  - Equilibrium Solutions
  - Time Evolution
- 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