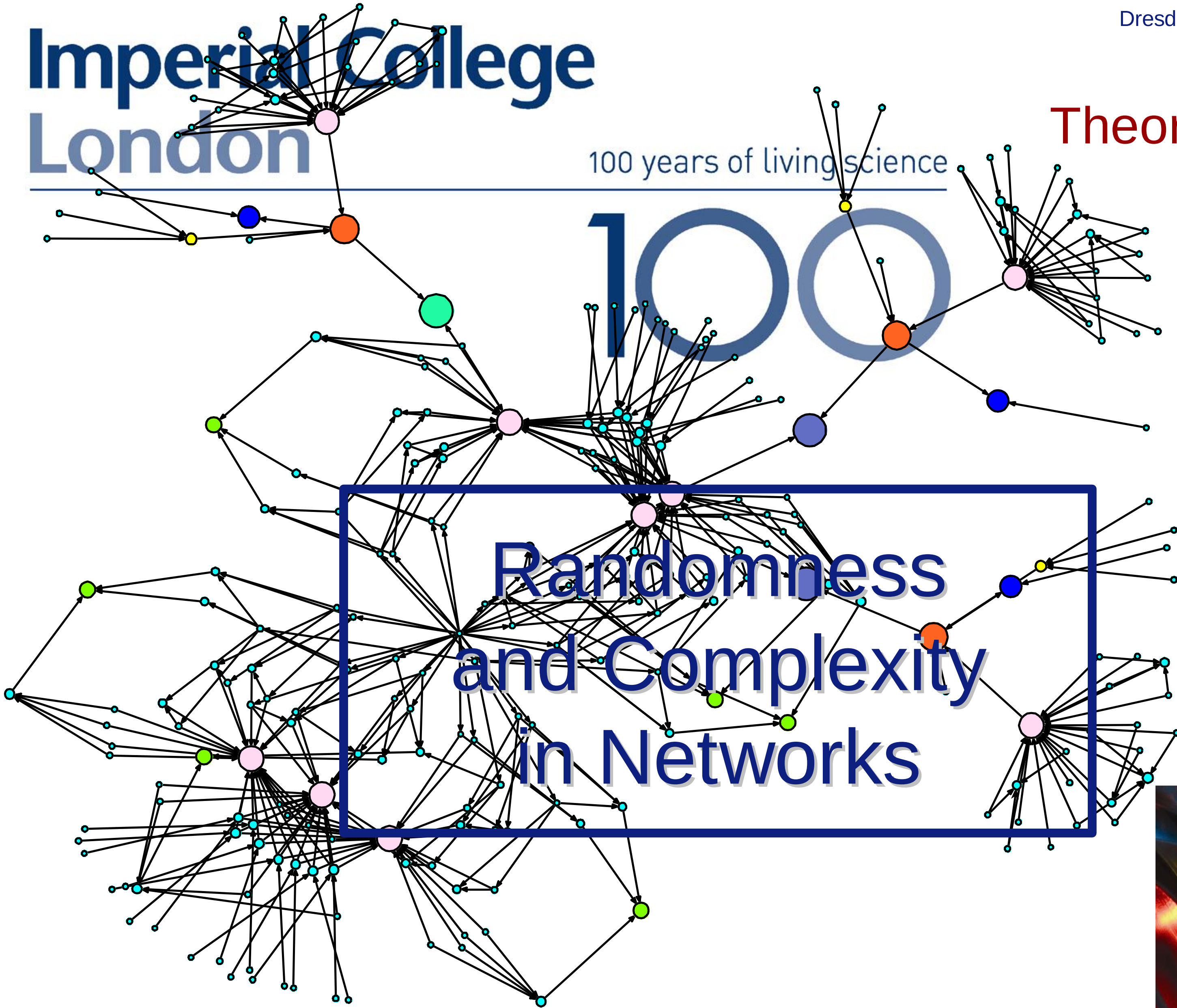
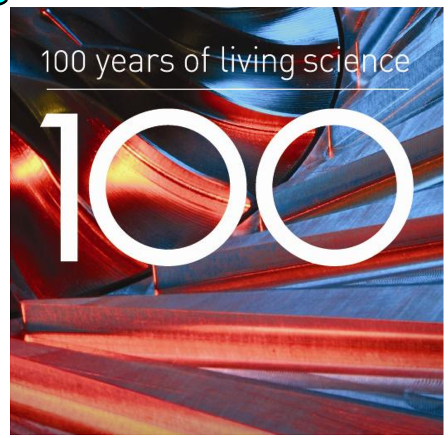
ECCS07
Dresden September 2007

Tim Evans
Theoretical Physics

Imperial College London
100 years of living science



Randomness and Complexity in Networks



Main authors on different aspects of work discussed here:-

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

All at Imperial College London

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+Saramaki 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

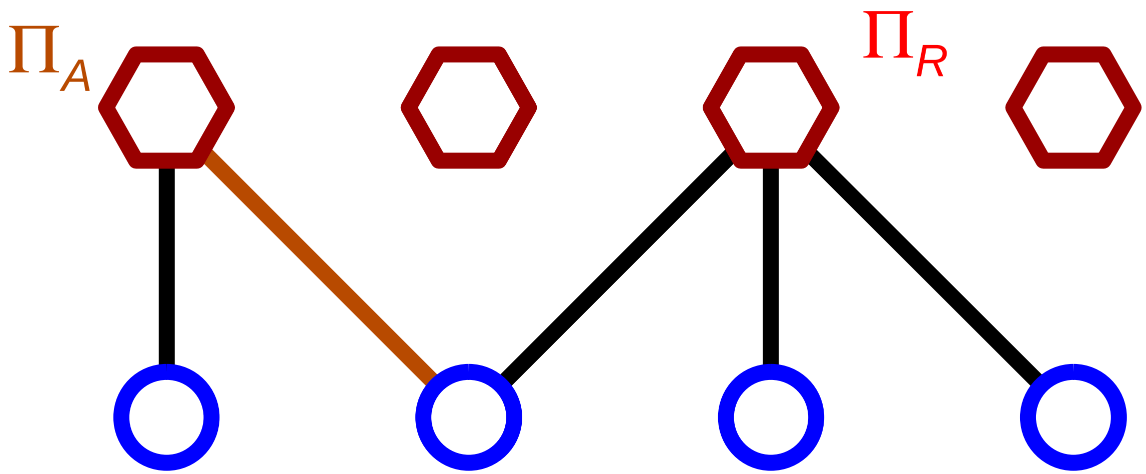
  (b) INNOVATING, picking an artifact at random
  it will be one no one else has chosen if *N* large

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

$$\Pi_A\,(k) = \underbrace{(1-p_r)}_{\text{copying probability}}\Pi_{\mathbf{copy}} + \underbrace{p_r}_{\text{innovation probability}}\Pi_{\mathbf{innovate}}$$

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



N artifacts

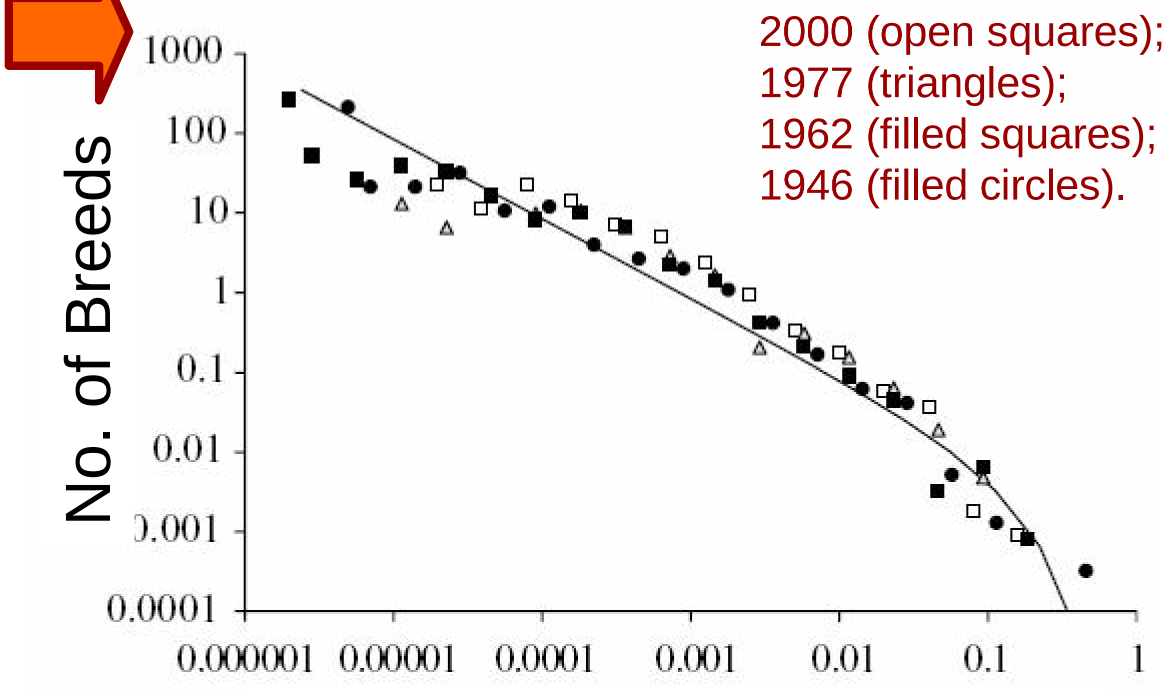
n(k) artifact degree dist.

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

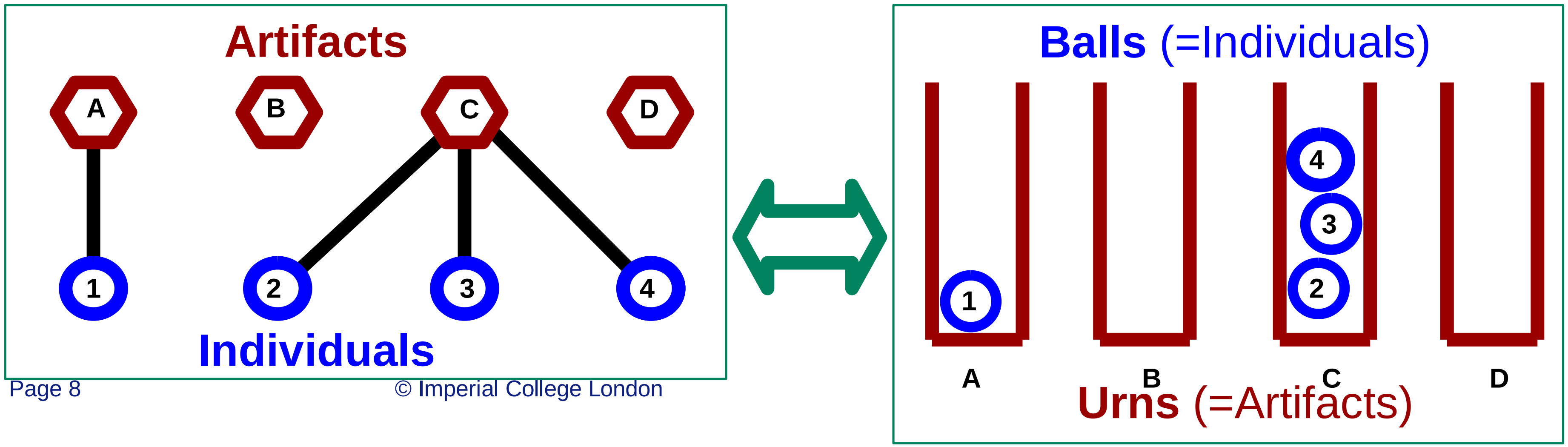
- Unipartite Graph Rewiring [Watts & Strogatz 1998]
- Gene Frequencies [Kimura & Crow, 1964]
  - Inheritence and Mutation (genes not memes)
- Speciation ['Tangled Nature' Christensen et al 2002]
- Family Names [Zanette & Manrubia, 2001]
  - Inheritence and New Immigrants
- Language Extinction [Stauffer 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 …



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:-

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

$\Rightarrow$ 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)

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

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

Power below one but in data *indistinguishable from one if*

<k> << 10

$$\boxed{\zeta = -\ln(1 - p_r)}$$

Exponential Cutoff

Large Degree Equilibrium Behaviour – Small p_r Case

For $\boldsymbol{p_r < p_* \sim 1/E}$

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

Degree distribution rises near $\boldsymbol{k=E}$

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

- a CONDENSATION or FIXATION

Blows up

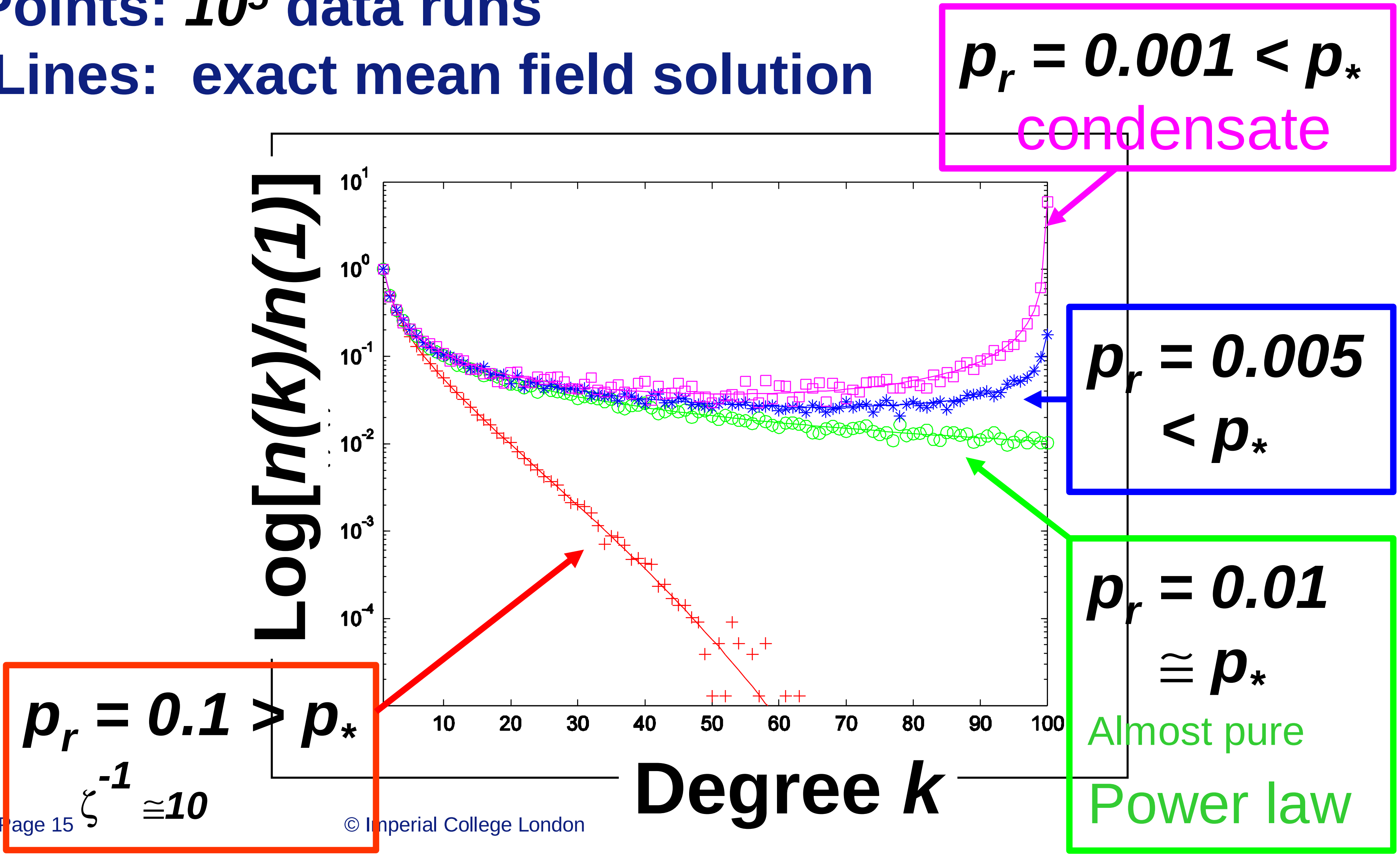
$$n(k) = A\left(\frac{\Gamma(k+\overline{K})}{\Gamma(k+1)}\right)\left[\frac{\Gamma(E-\overline{E}-\overline{K}-k)}{\Gamma(E+1-k)}\right]$$

Equilibrium Behaviour Results

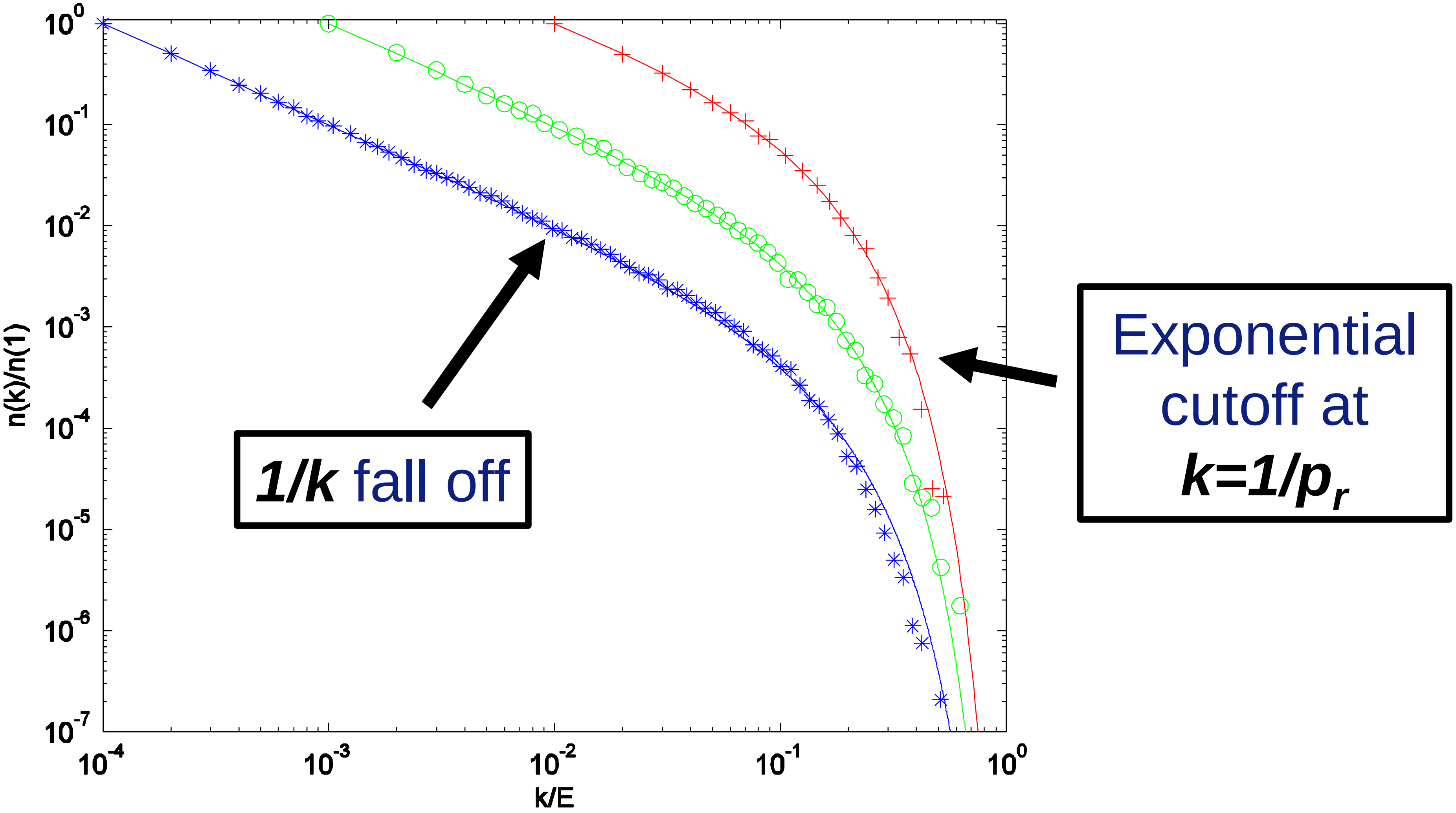
N=E=100

Points: *10^5* data runs

Lines: exact mean field solution



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



Tangled Nature

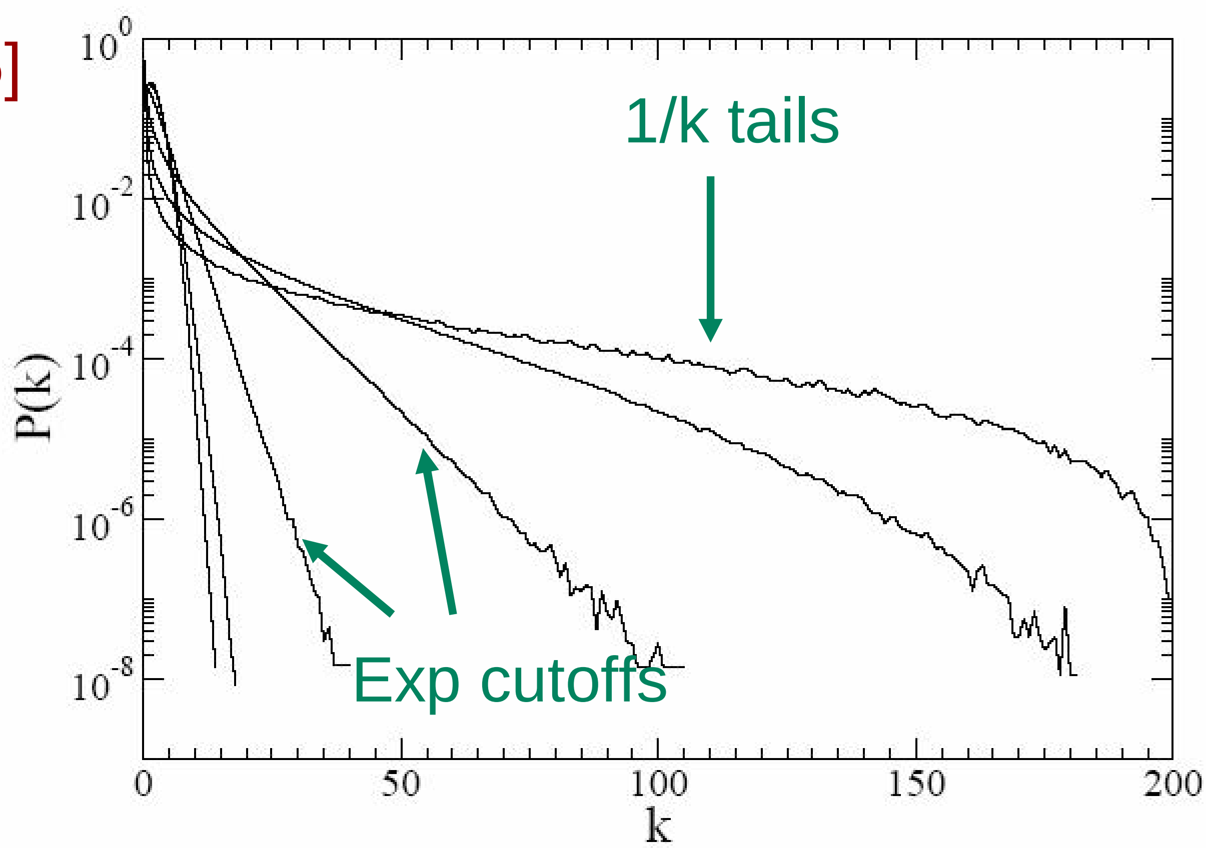
- Tangled Nature model of speciation with a particular focus on extinctions and intermittency in extinction patterns [Christensen et al, 2002; Laird & Jensen 2007]

- 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.

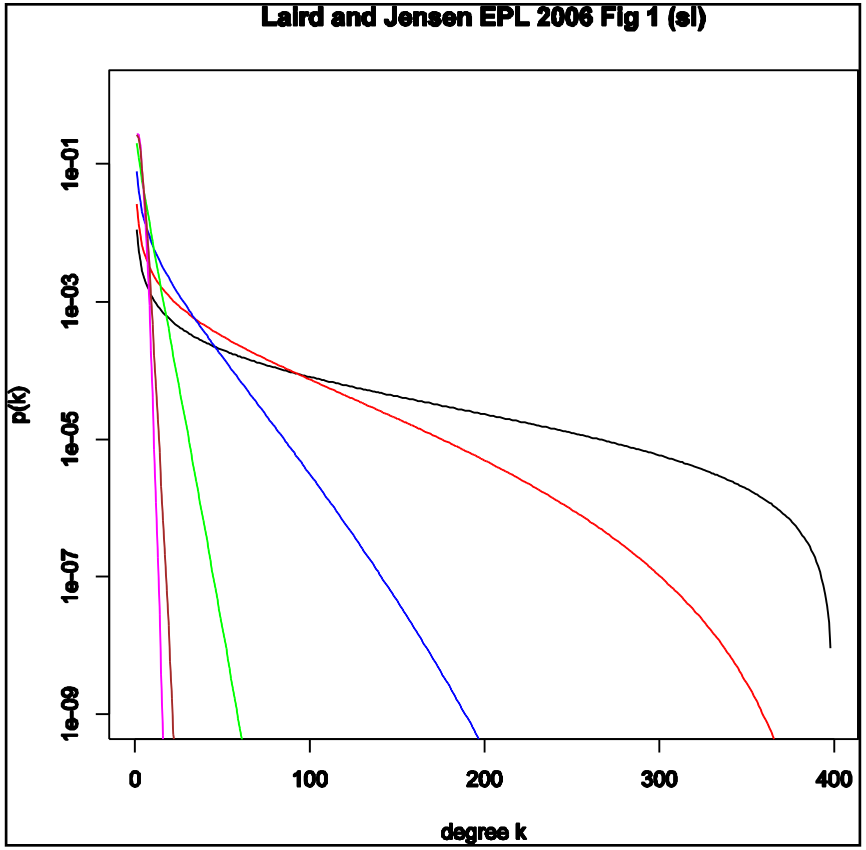
Network Version of Tangled Nature

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

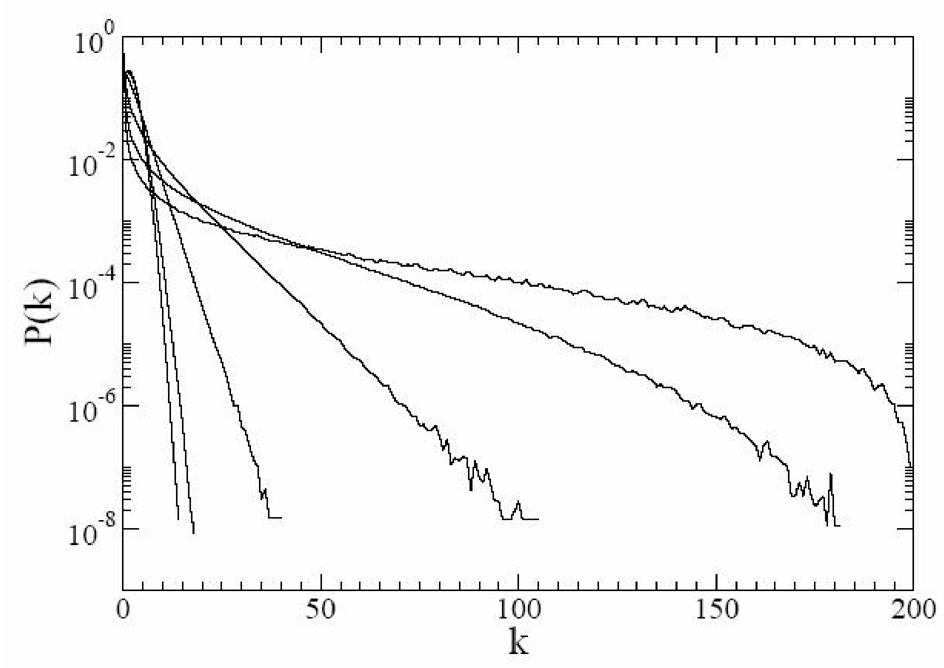
  Shows same ***p(k)*** as model here



Network Version of Tangled Nature (2)



This Model



Laird & Jensen 2006

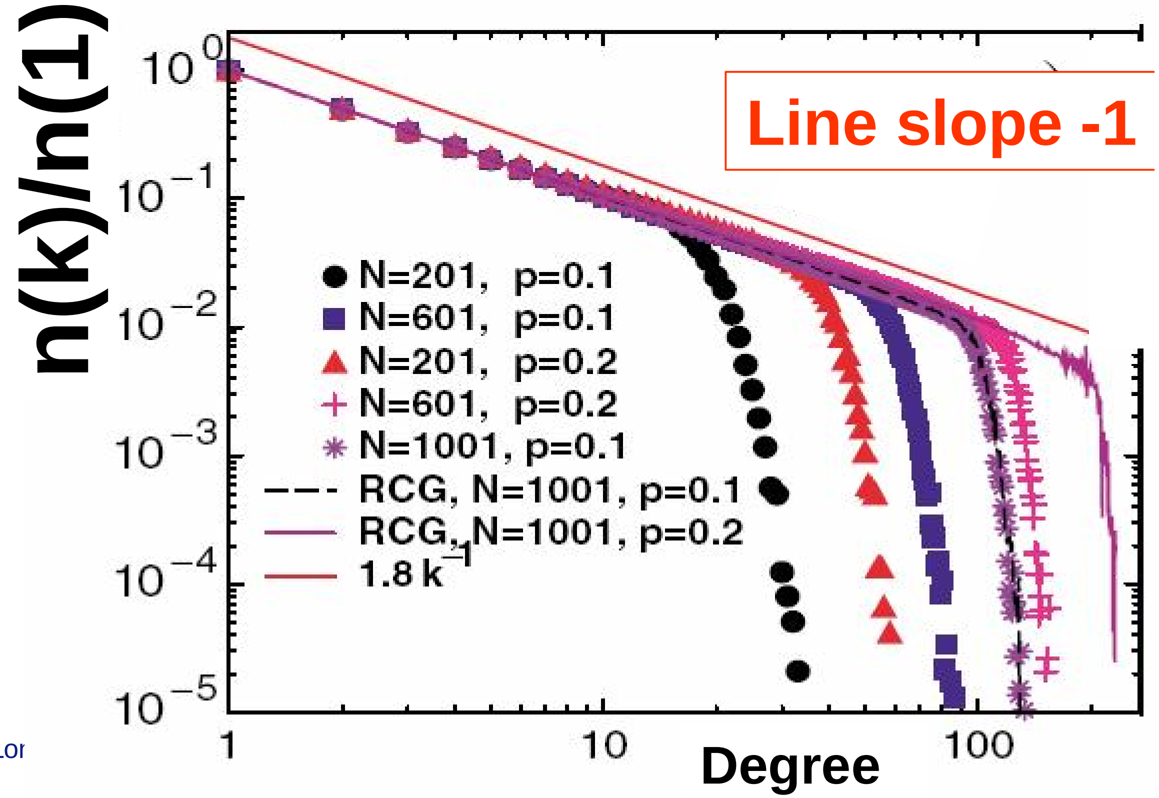
Minority Game Example – Leaders and Followers [Anghel et al. 2004]

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



Exact Time Evolution Known

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

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

constants c_m fixed by initial conditions

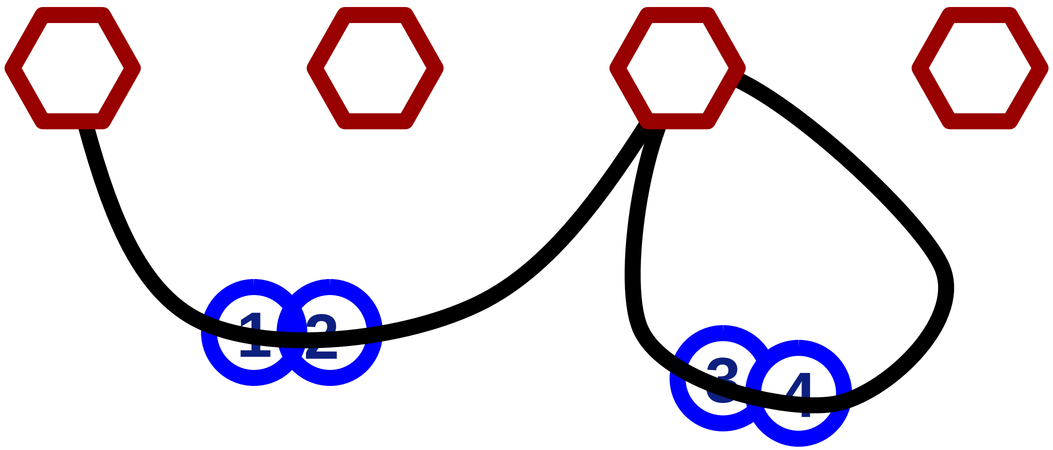
Eigenvalues

Eigenfunctions are ratios of Gamma functions

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

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

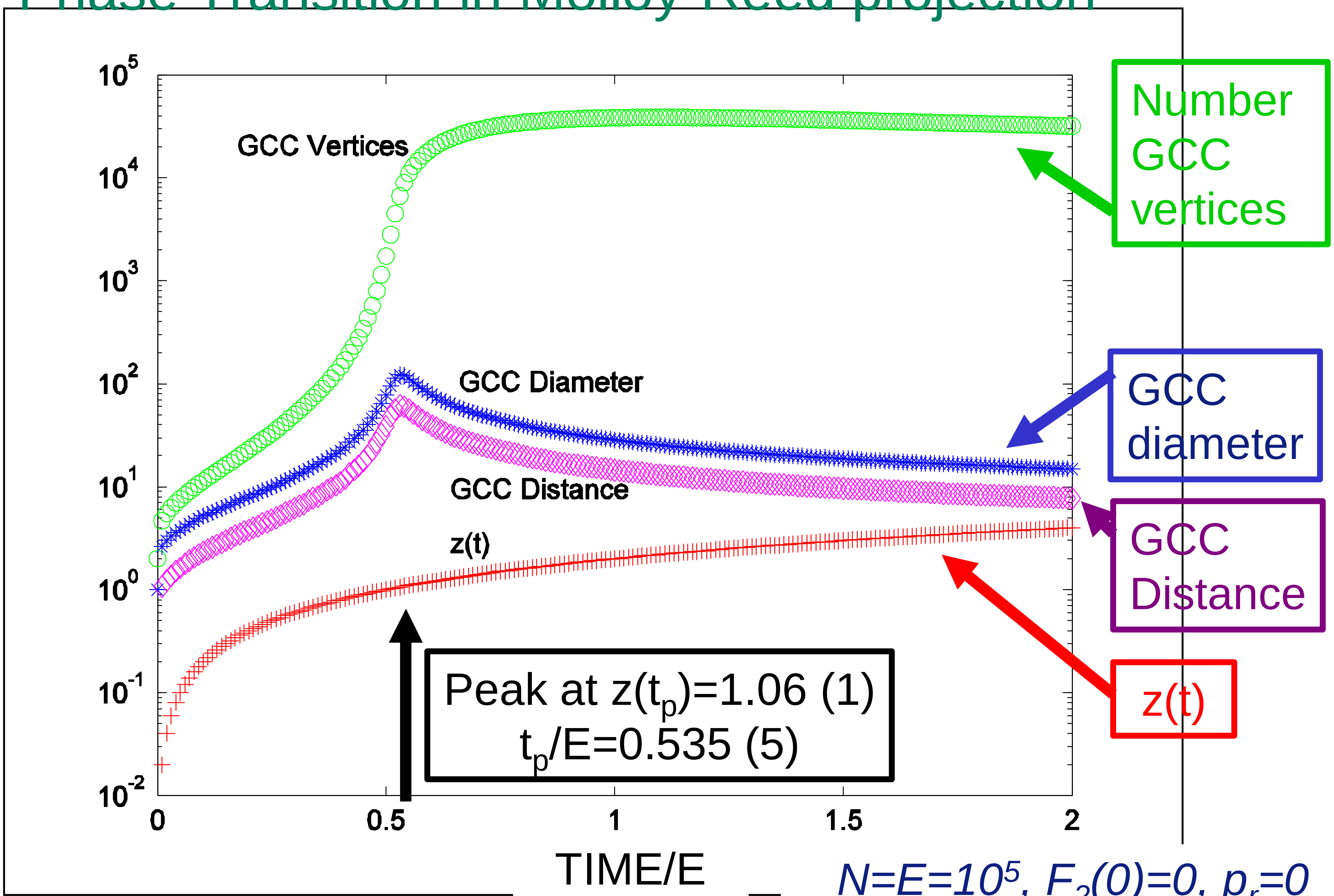
Infinite Random Graphs have a phase transition (e.g. appearance of **GCC** - Giant Connected Component) at ***z(t)=1*** where

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

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

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***

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

- 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

See my web page (*google Tim Evans Imperial*)

- 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
- T.S.Evans, A.D.K.Plato, & T.You (in prep)

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