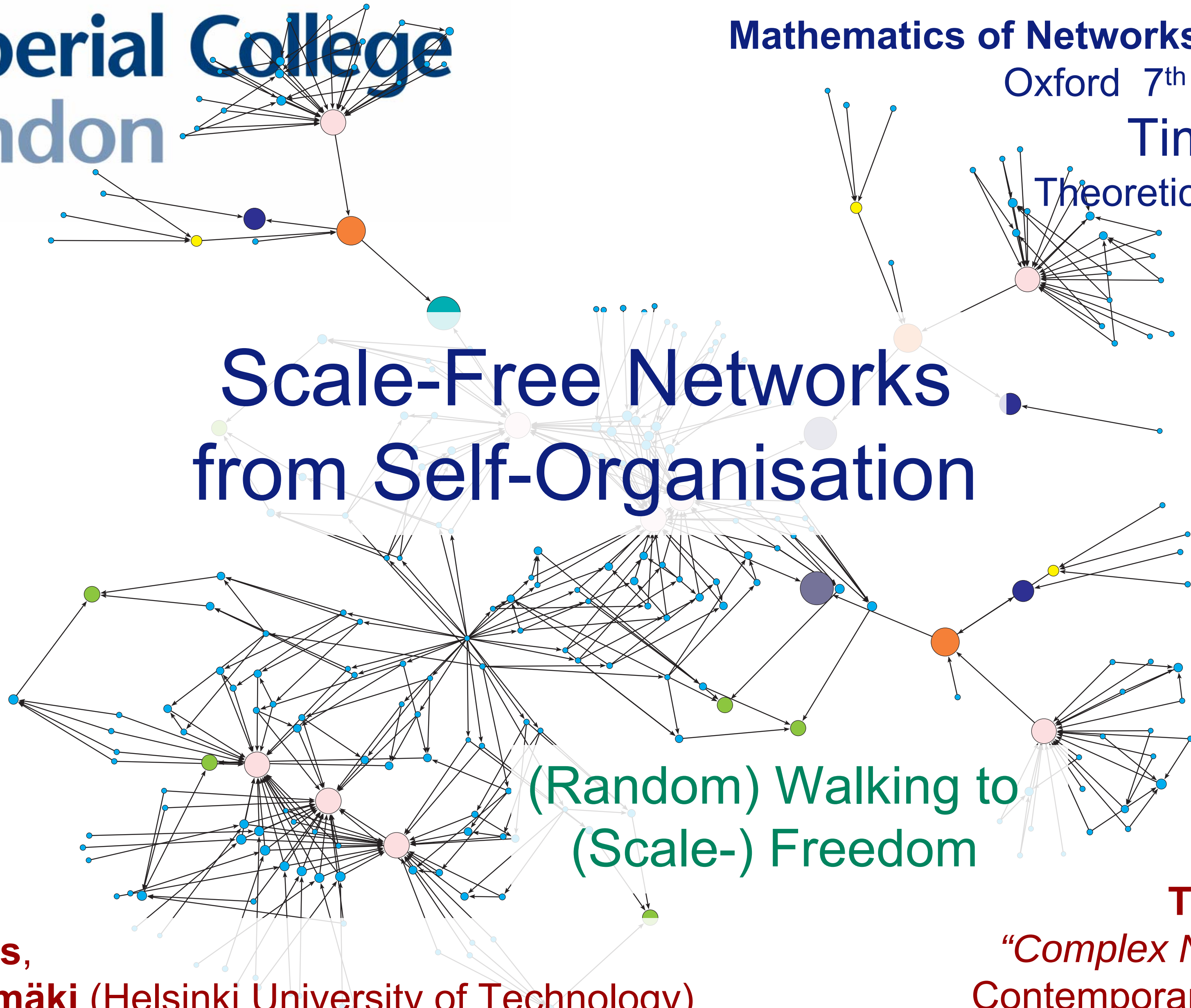
Imperial College London

Mathematics of Networks meeting,
Oxford 7th April 2006

Tim Evans
Theoretical Physics



Scale-Free Networks from Self-Organisation

(Random) Walking to (Scale-) Freedom

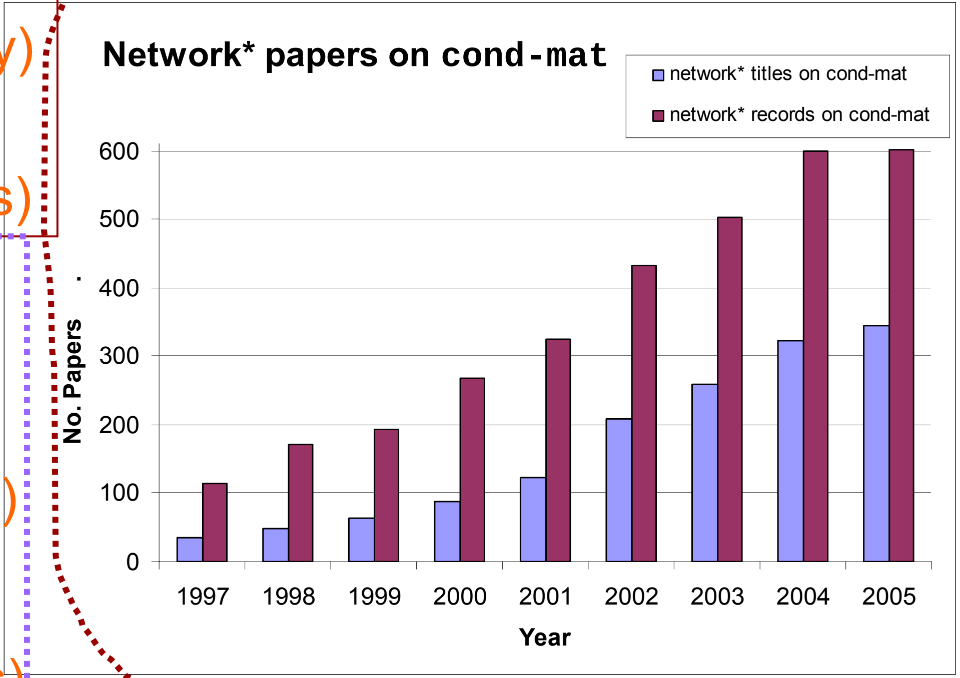
T.S.Evans,
J.P.Saramäki (Helsinki University of Technology)
"Scale Free Networks from Self-Organisation"
PRE 72 (2005) 026138 [cond-mat/0411390]

T.S.Evans,
"Complex Networks",
Contemporary Physics
45 (2004) 455 – 474
[cond-mat/0405123]

Multidisciplinary Nature

- Mathematics (Graph Theory, Dynamical Systems)
- Physics (Statistical Physics)
- Biology (Genes, Proteins, Disease Spread, Ecology)
- Computing (Search and ranking algorithms)
- Economics (Knowledge Exchange in Markets)
- Geography (City Sizes, Transport Networks)
- Architecture ("Space Syntax")
- Anthropology (Social Networks)
- Archaeology (Trade Routes)

For instance the condensed matter electronic preprint archives have gone from 35 papers in 1997 with a word starting with Network in their title to 344 last year, an increase of nearly 1000%

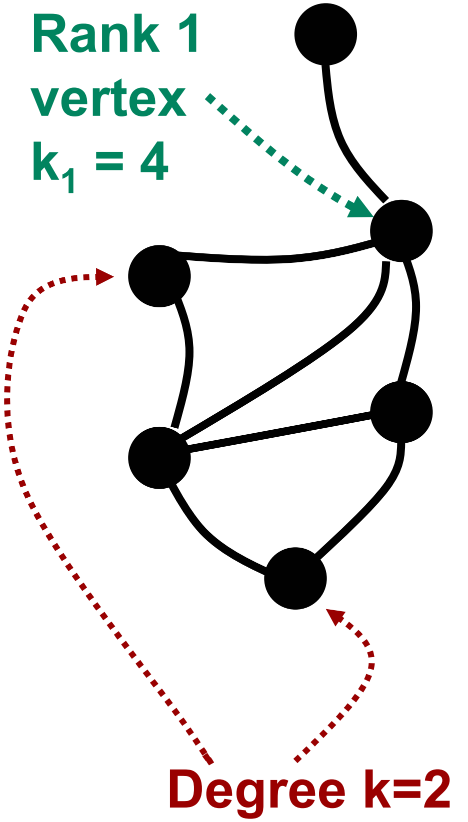


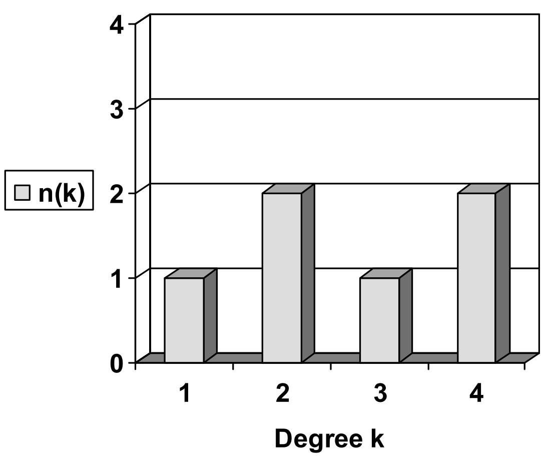
Notation

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_1 = *Maximum degree of graph*
  = *Degree of rank 1 vertex*
- K = <k> = *average degree* = 2E/N
- Degree Distribution
  n(k) = *number of vertices with degree k*
  p(k) = n(k)/N = *normalised distribution*



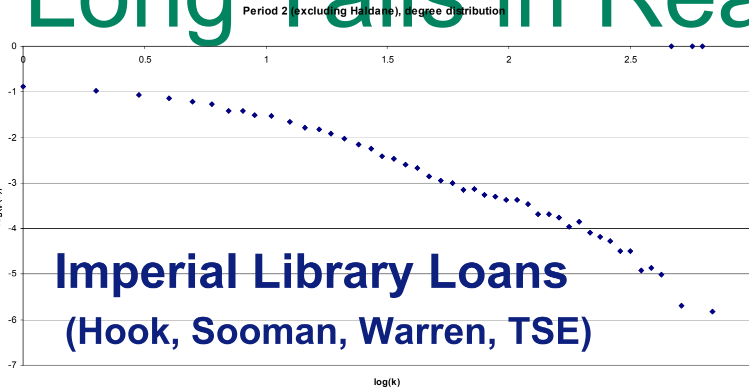


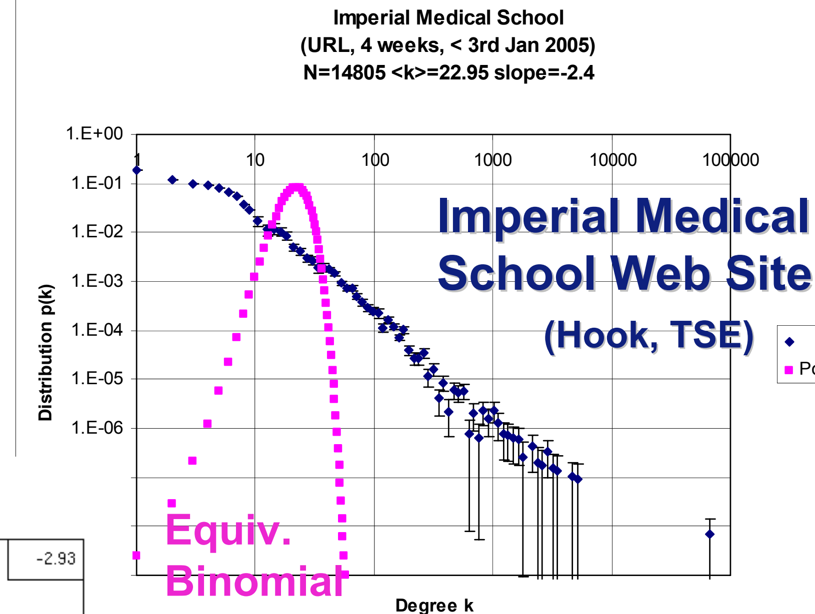
Real Networks

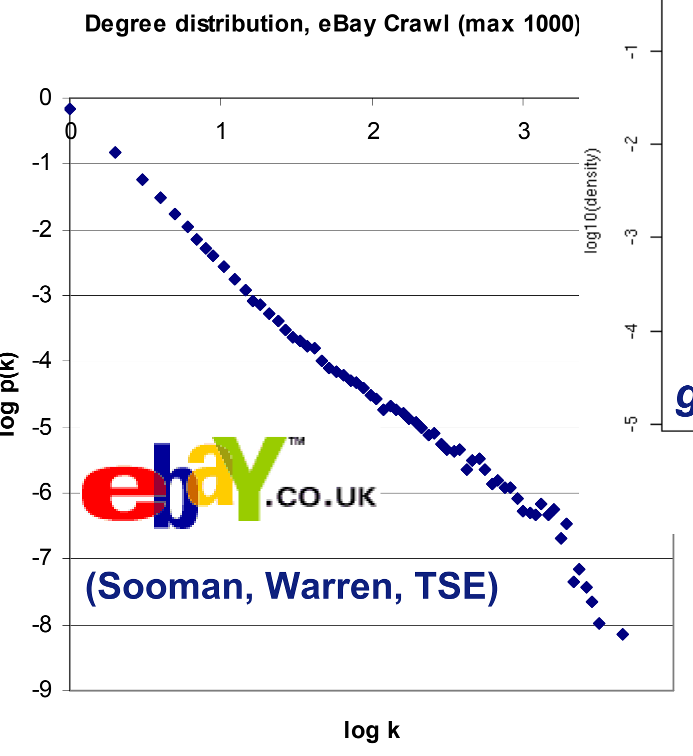
- **Short Distance Scales** $d \leq O(\ln(N))$
- **Long Degree Distributions** $k_1 > O(\ln(N))$

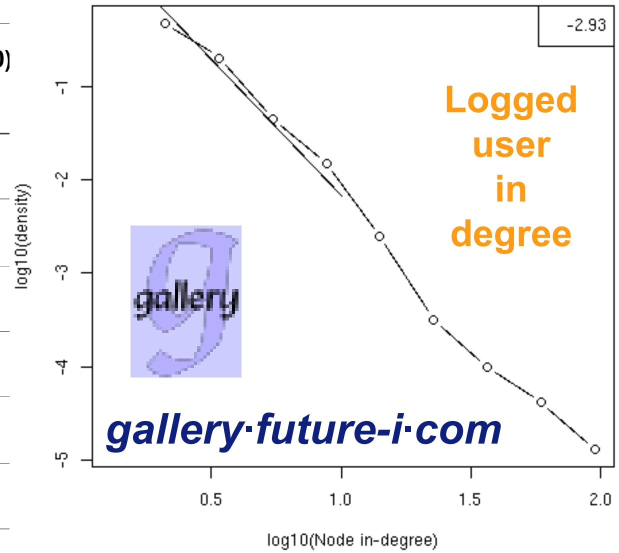
	Distance Scale d	Tail of Degree Distribribution	Maximum Degree k_1
Lattice	Large $d \sim N^{1/dim}$	No Tail $\delta(k-k_0)$	Fixed k_0
Watts-Strogatz Small World	Small $d \sim \log(N)$	No Tail $\sim \delta(k-k_0)$	V.Small $\sim k_0$
Erdős-Réyni Random	Small $d \sim \log(N)$	Short Tail $\langle k\rangle^k e^{-\langle k\rangle} / k!$ Poisson	Small $\sim\log(N)$
Scale-Free	Small $d \sim \log(N)$	Long Tail $\sim k^{-\gamma}$	Large = HUBS $\sim k^{1/(\gamma-1)}$

Long Tails in Real Data

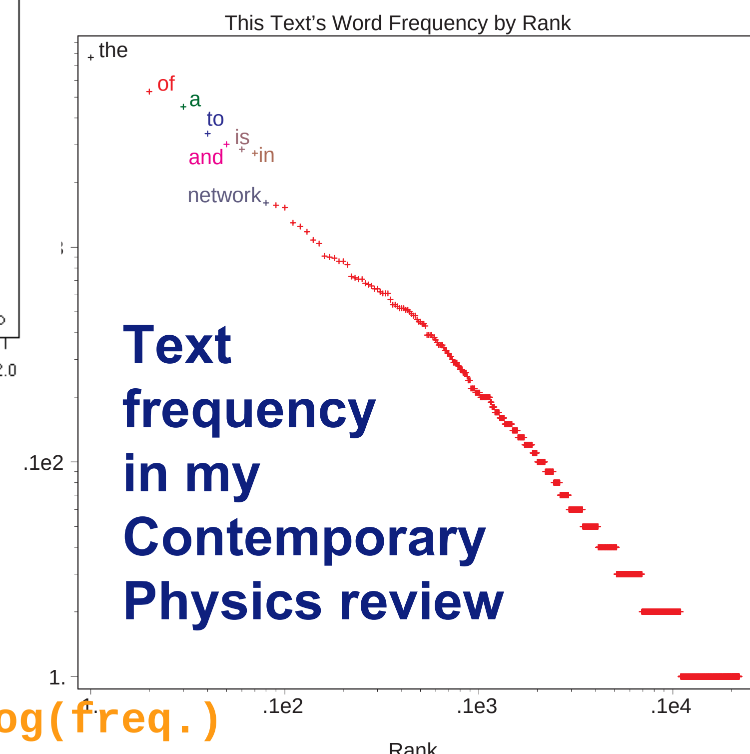








ege London



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

Long Tails = Hubs

Hubs are vertices of high degree

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

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

  rand.net. $N=10^6$, $<k>=4 \Rightarrow k_1 =$ **17**

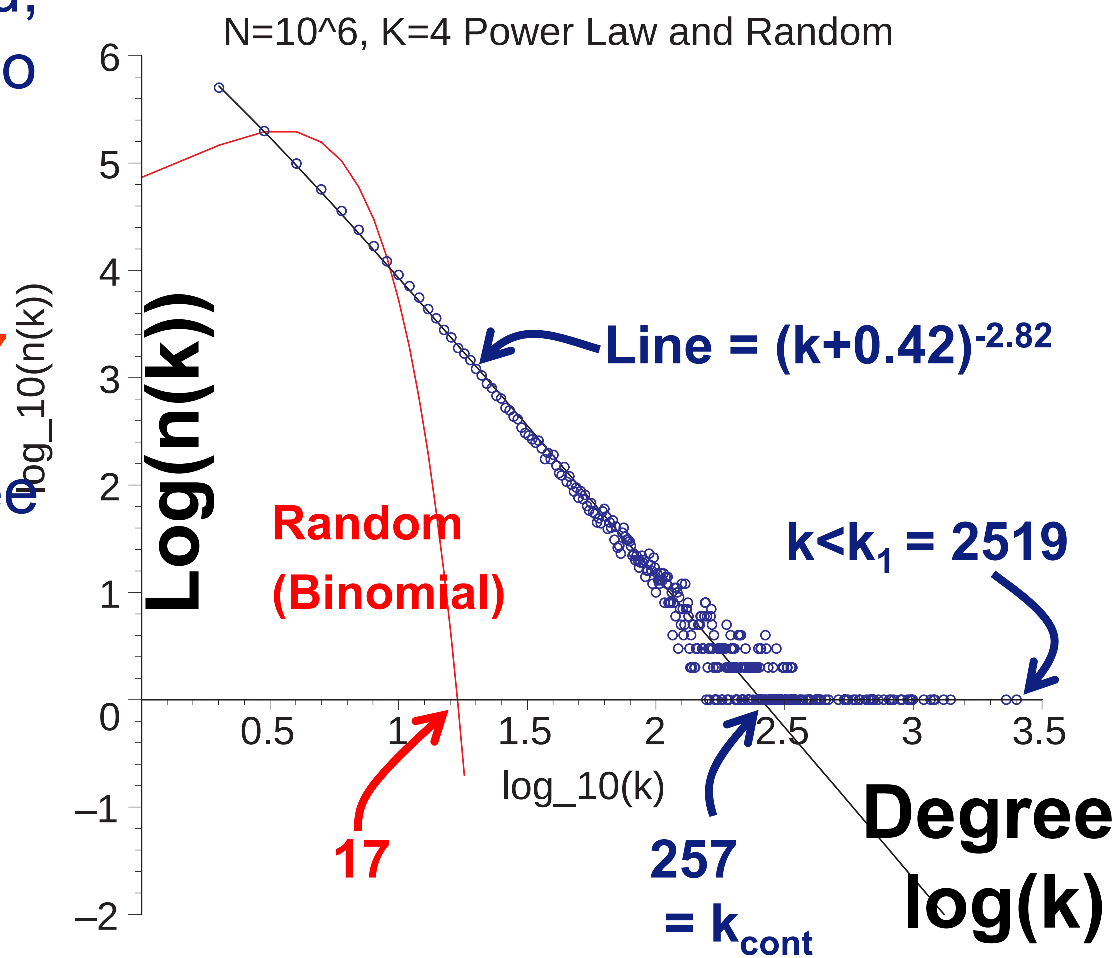
- Only a **long tailed** degree distribution has hubs

e.g. POWER LAW

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

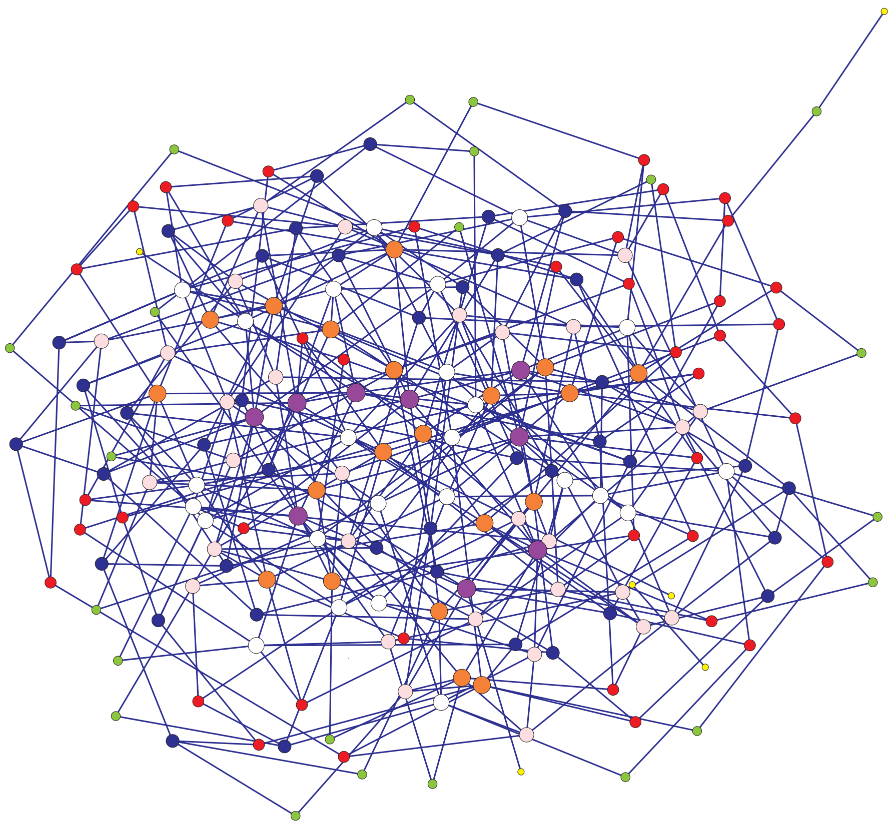
$k \leq k_1 = O(N^{1/2})$

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



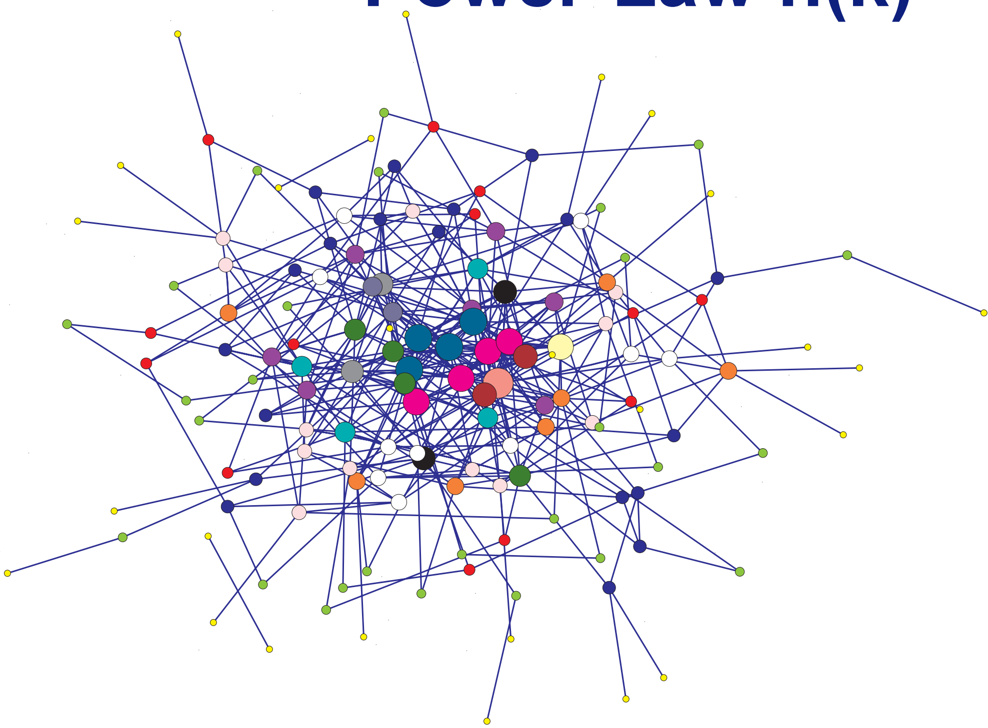
N=200, <k>~4.0, vertex size ∝ k

Random



Diffuse centre of small degree vertices

Scale-Free = Power-Law n(k)



Tight core of large hubs

Growth with Preferential Attachment

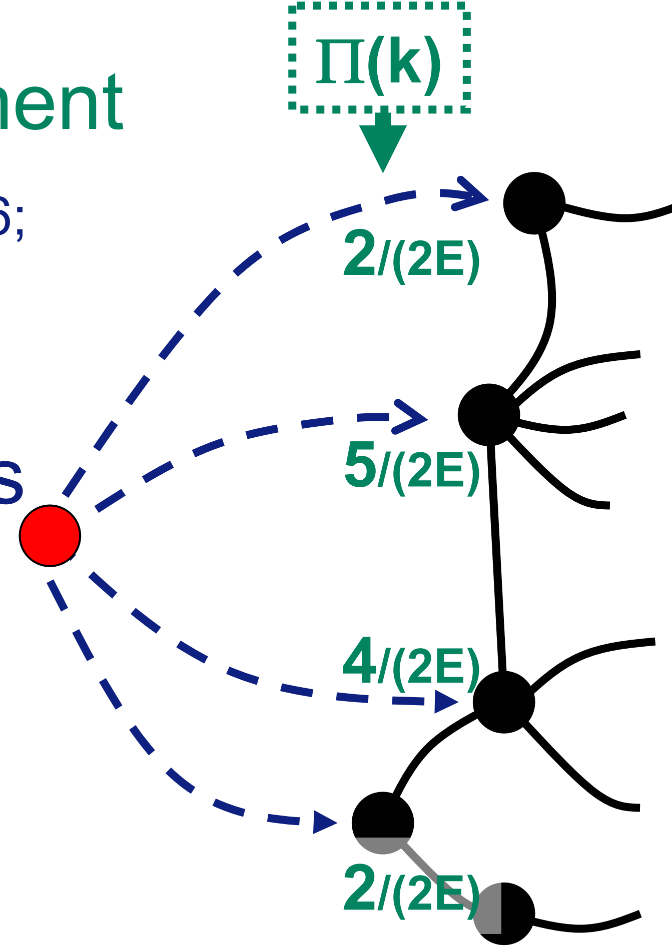
(Yule 1925, 1944; Simon 1955; Price 1965,1976; Barabasi,Albert 1999)

1. Add new vertex attached to one end of ½<k> new edges
2. Attach other ends to existing vertices chosen with probability Π proportional to their degree

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

Preferential Attachment

"Rich get Richer"



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

Equivalent to random edge selection

Scale-Free Growing Model comments

- Growth not essential
  – rewiring with reattachment probability $\Pi \Rightarrow \gamma \sim 1.0$
  – 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

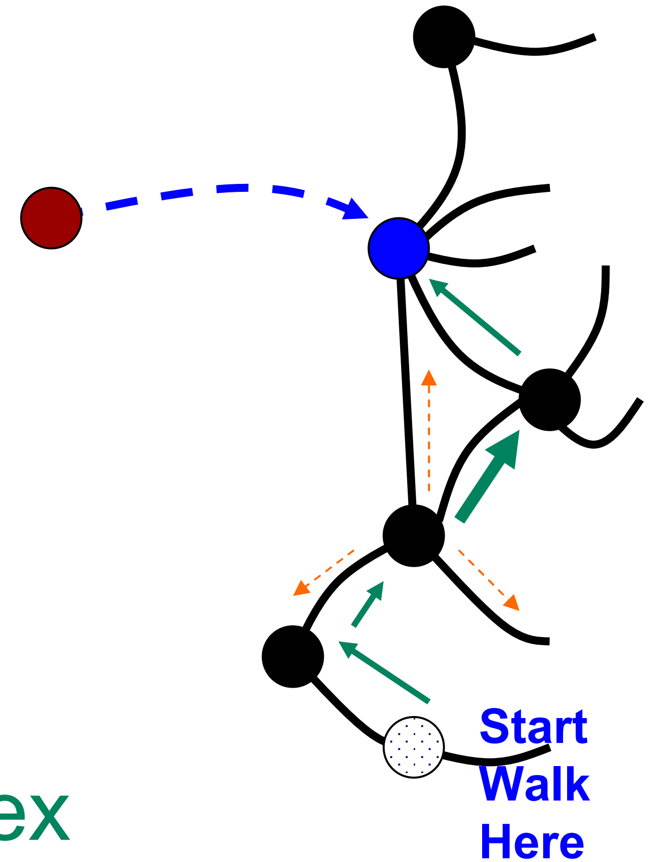
Random Attachment

ε = fraction of times add new vertex

- BUT if $\lim_{k\to\infty} \Pi(k) \propto k^{\alpha}$ for any $\alpha \neq 1$ then a *power law degree distribution is not produced!*

Walking to a Scale-Free Network

(TSE, Klauke 2002; 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

$\propto$ 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

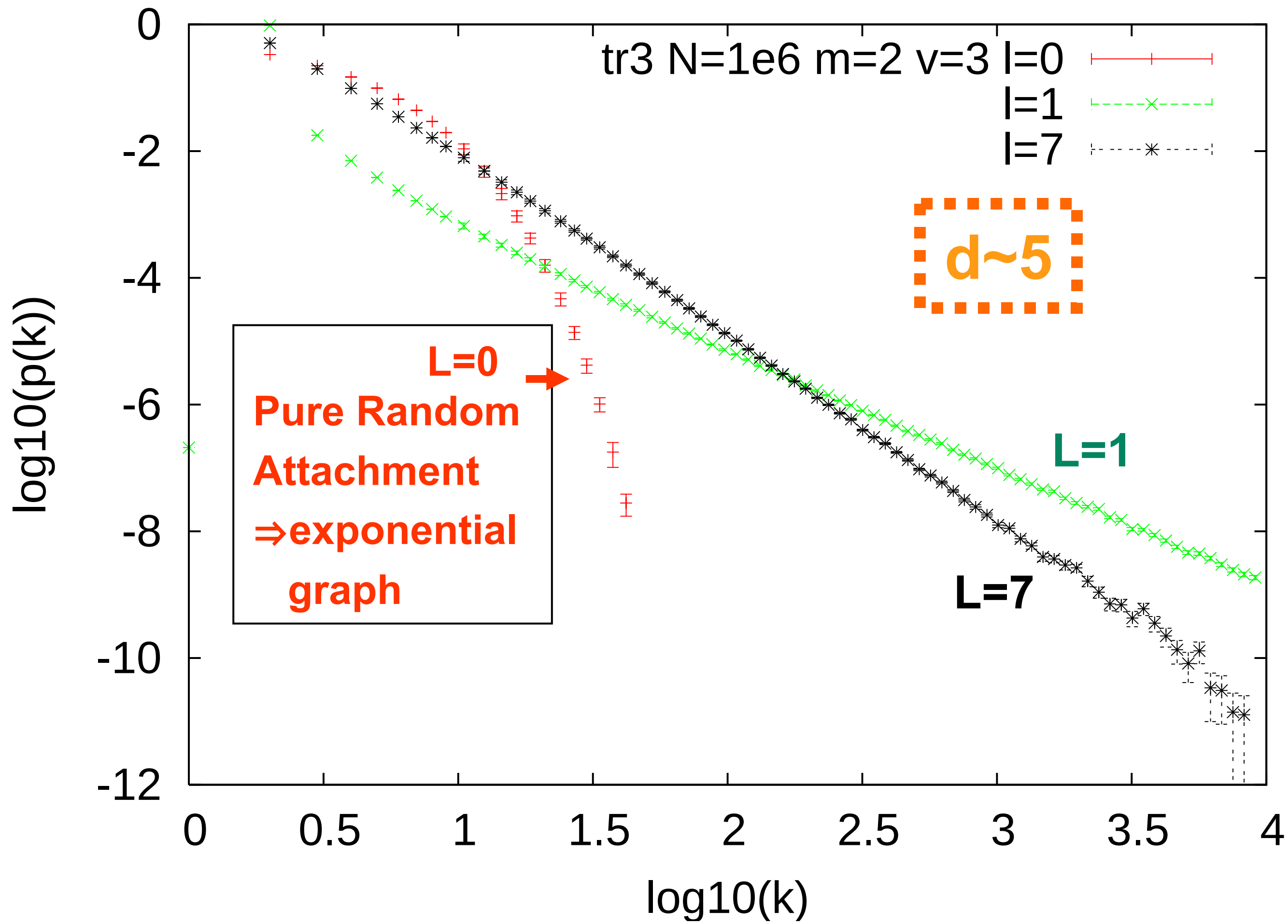
  Simon/Barabasi-Albert models use global information in their normalisation

- 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

Barabasi-Albert do NOT need a network, results and equations known from non-network work of Yule 1925; Simon 1955; Parker 1965; …

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

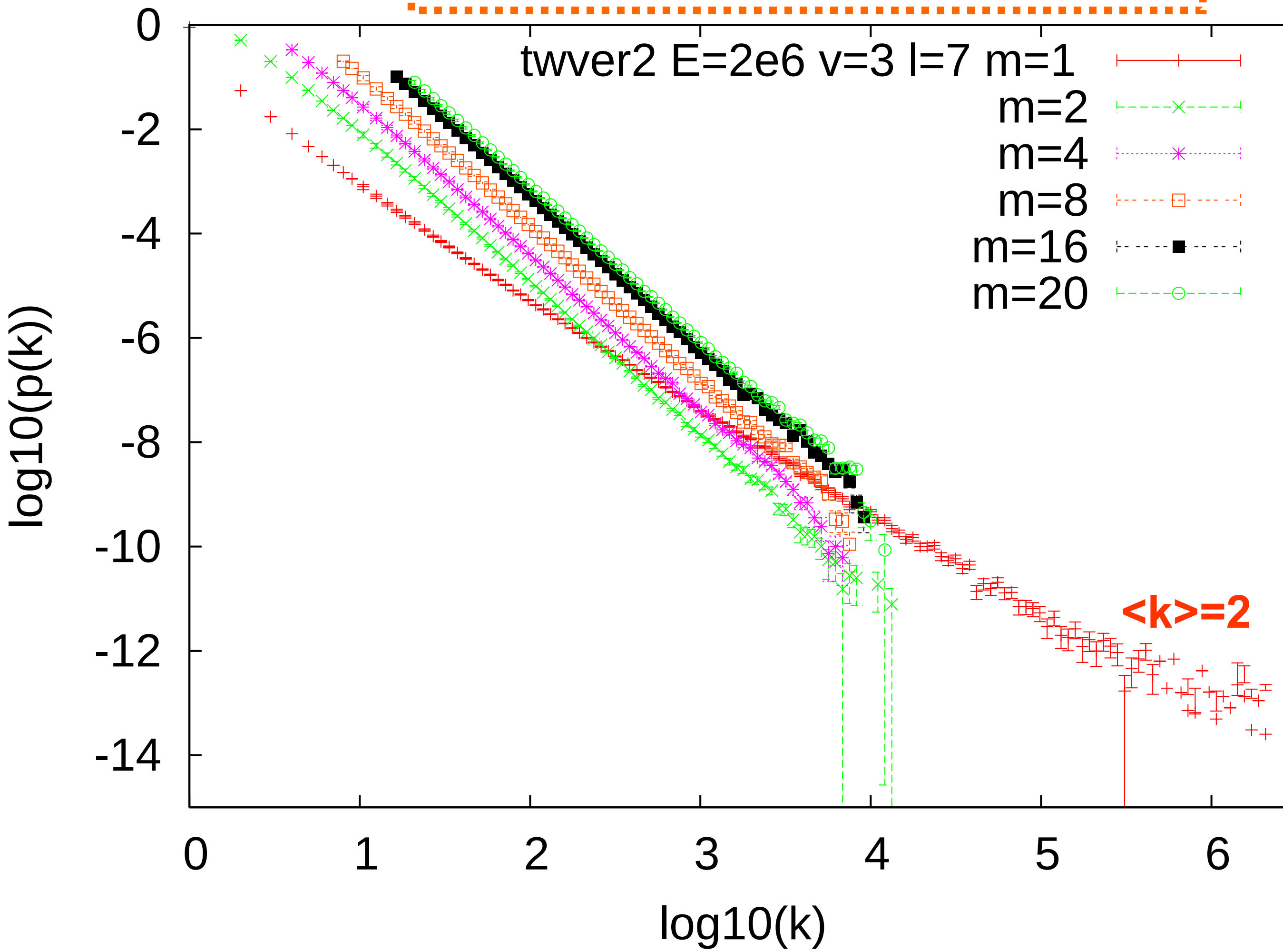


- Walks of length ONE are usually sufficient to generate reasonable scale-free networks

⇒ Degree Correlation Length < 1 < d (any distance scale)

Does the average degree <k> matter?

Simple walks, l=7, 2m=k



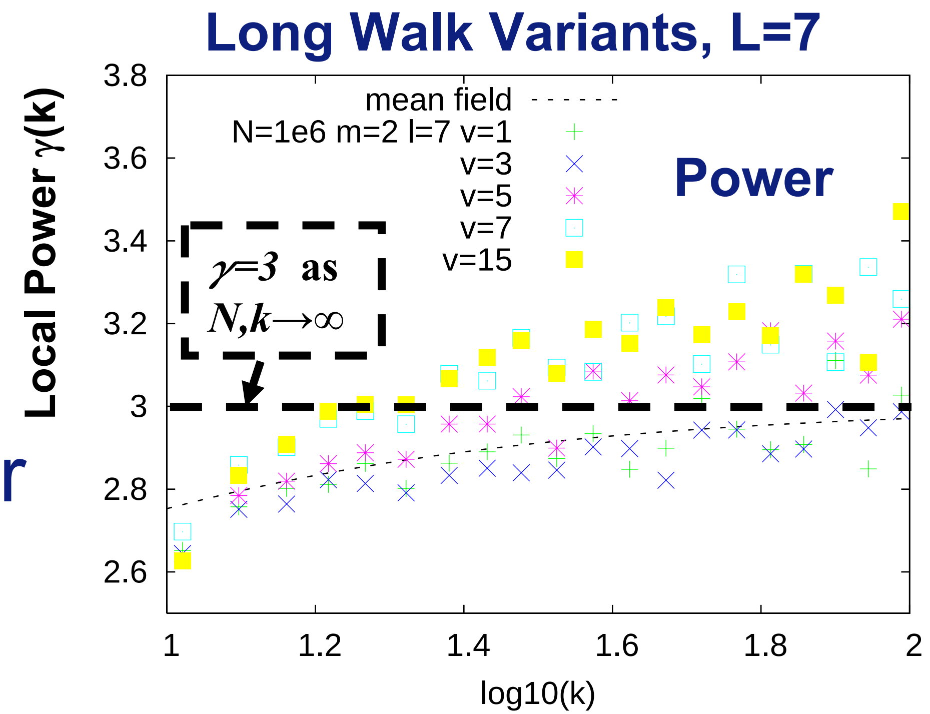
NO

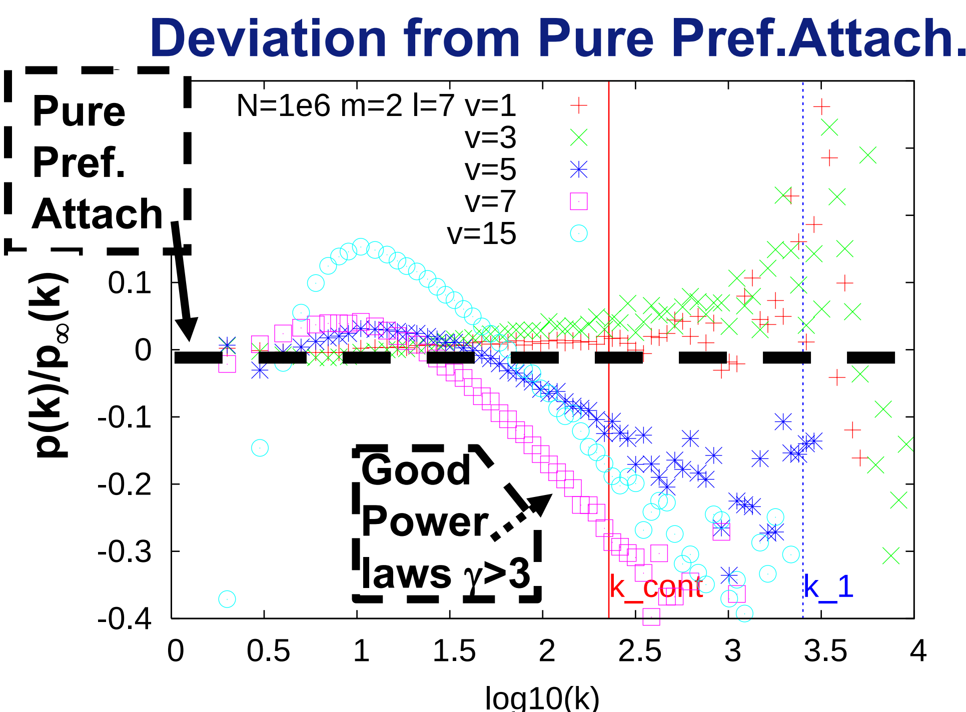
except for <k>=2 where a tree graph is generated

Is the Walk Algorithm Robust?
YES

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

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





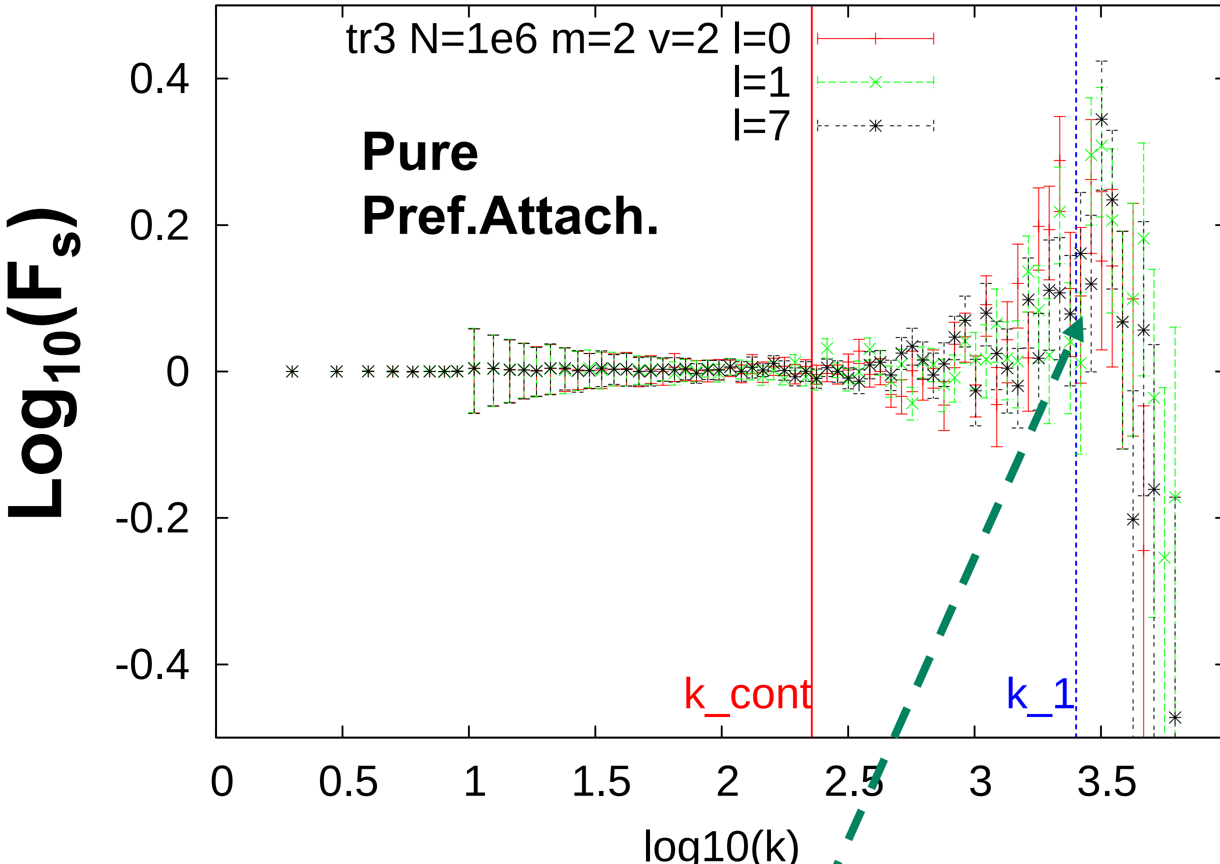
Finite Size Effects for pure preferential attachment

$$p(k) = p_\infty(k).F_S\left(\frac{k}{N^{1/2}}\right), \quad p_\infty(k) = \frac{\langle k\rangle(\langle k\rangle + 2)}{2k(k+1)(k+2)} \rightarrow \frac{1}{k^3}$$

Scaling Function

$$F_S(x) \approx 1$$

$$if \quad x < 1$$

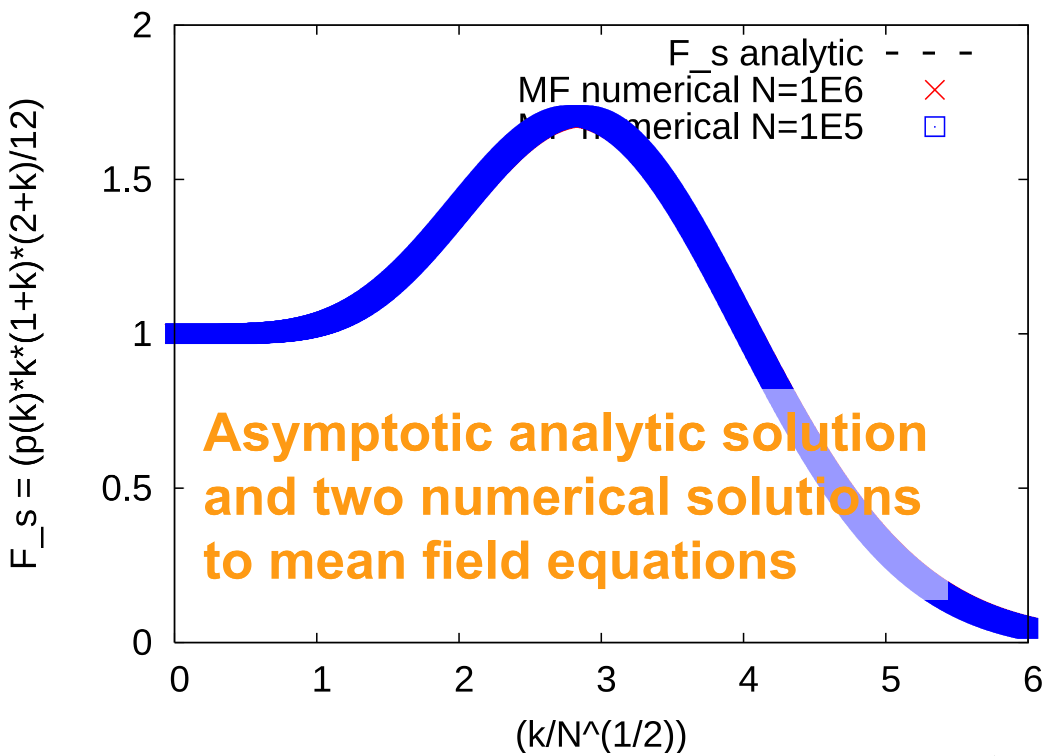


100 runs to get enough data near k_1

Mean Field Exact Finite Size Scaling Function F_s

(pure pref.attach.)

Can calculate the finite size effects in the mean field approximation to find



$$F_s(x) \approx \mathrm{erfc}(x) + \frac{\exp(x^2)}{\sqrt{\pi}}\left(2x + \sum_{n=3}^{m+2} \frac{8}{n!}\left[1 + \left((1+m)\delta_{m+1,n}\right)\right] x^n \, \mathrm{H}_{n-3}(x)\right)$$

(TSE+Saramäki, 2005;

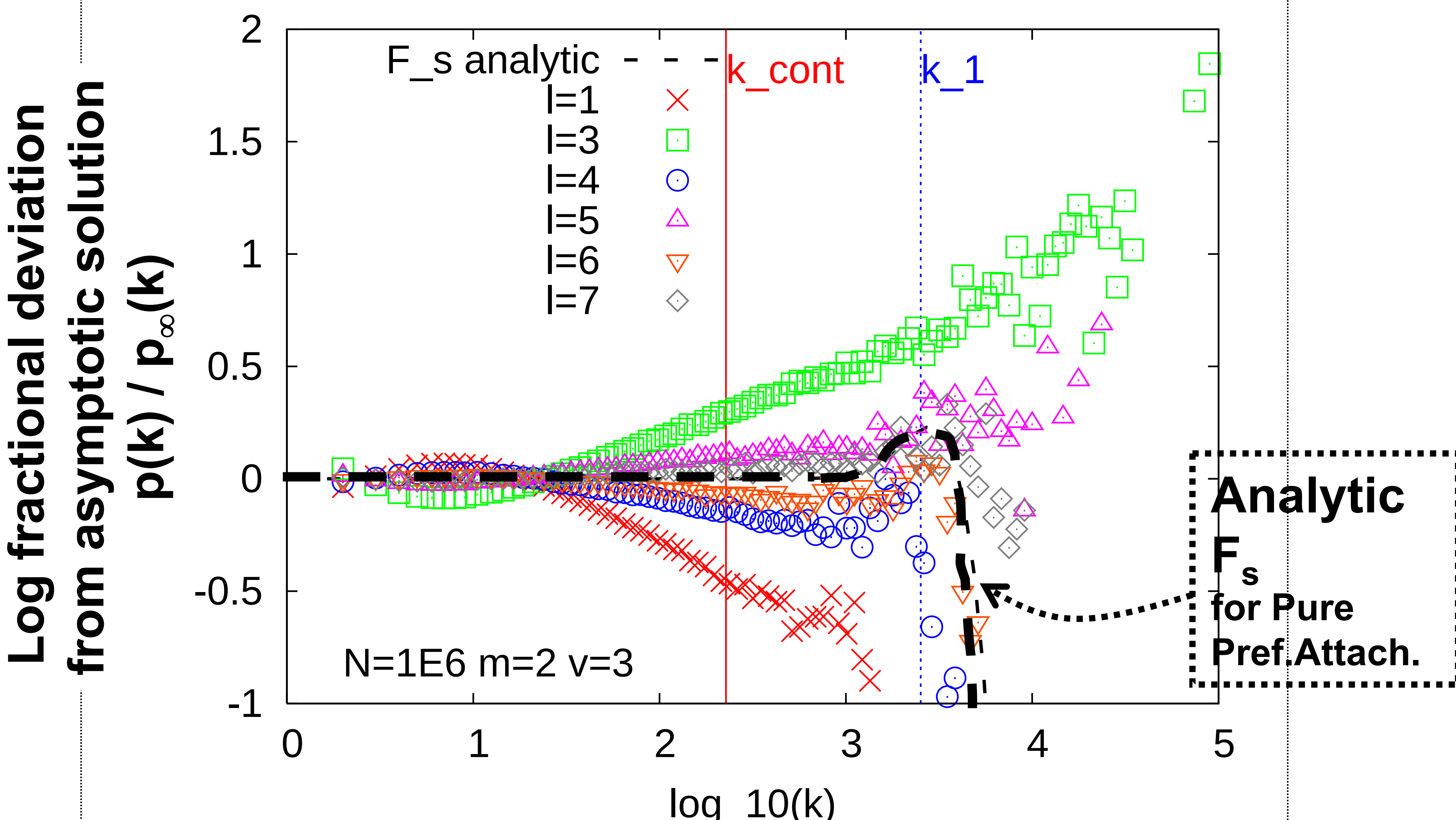
generalisation of Krapivsky and Redner, 2002)

2m=<k>

Hermite Polynomials

Walk Data & Finite Size Scaling Function F_s

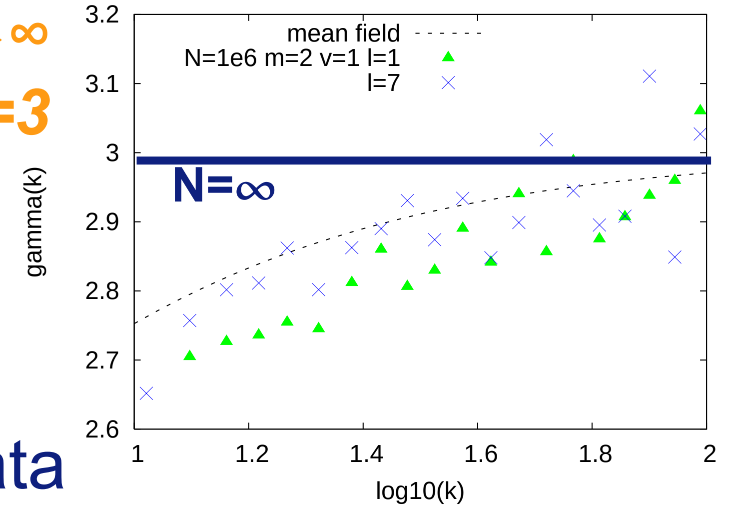
Similar shape but not exact fit



Powers and Finite Size Effects

as $N,k \to \infty$

$\gamma=3$



- Best data is for $k<k_{cont} = O(N^{1/3})$
- Finite N effects irrelevant to real data $F_s \neq 0$ only for largest degree vertex

  $k > k_1 \sim O(N^{1/2})$
- Power law only ever for large k $\Rightarrow$ *corrections for small k*
- Large networks are only *mesoscopic* systems, $\Rightarrow$ *k never large* e.g. $N=10^6$ $\gamma=3$ network $k_{cont} \sim 250$, $k_1 \sim 2500$

$\Rightarrow$ **Small k deviations vital for all known networks**

- Simple power law fits underestimate asymptotic power by $O(10\%)$

Random Walks as a SearchTool

Sample Networks via Random Walk

⇒ visit vertices with probability

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

so visit Hubs much more often,

⇒ find them very quickly

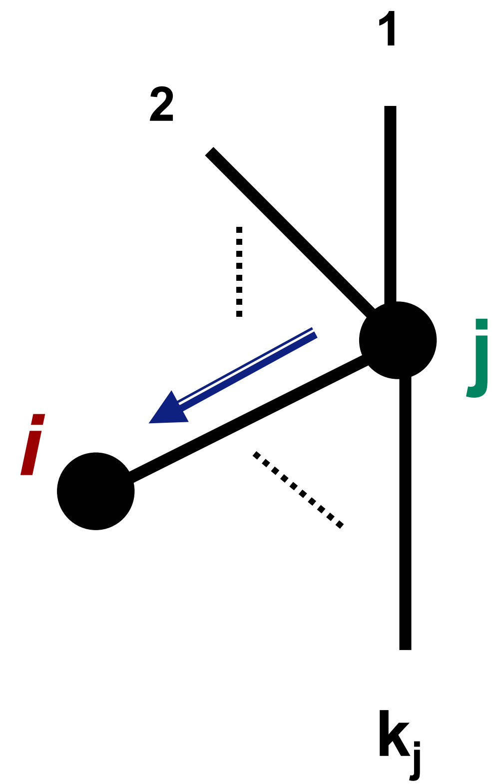
- Estimate tail of degree distribution very quickly
- Estimates of size of graph possible
- Other biased walks possible
  e.g. can sample vertices equally if slowly
  (Orponen, Schaeffer, 2004)

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

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



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

⟹ Solve Matrix Equation

$$\mathbf{v}(t) = [\mathbf{P}]^t\, \mathbf{v}(0)$$

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

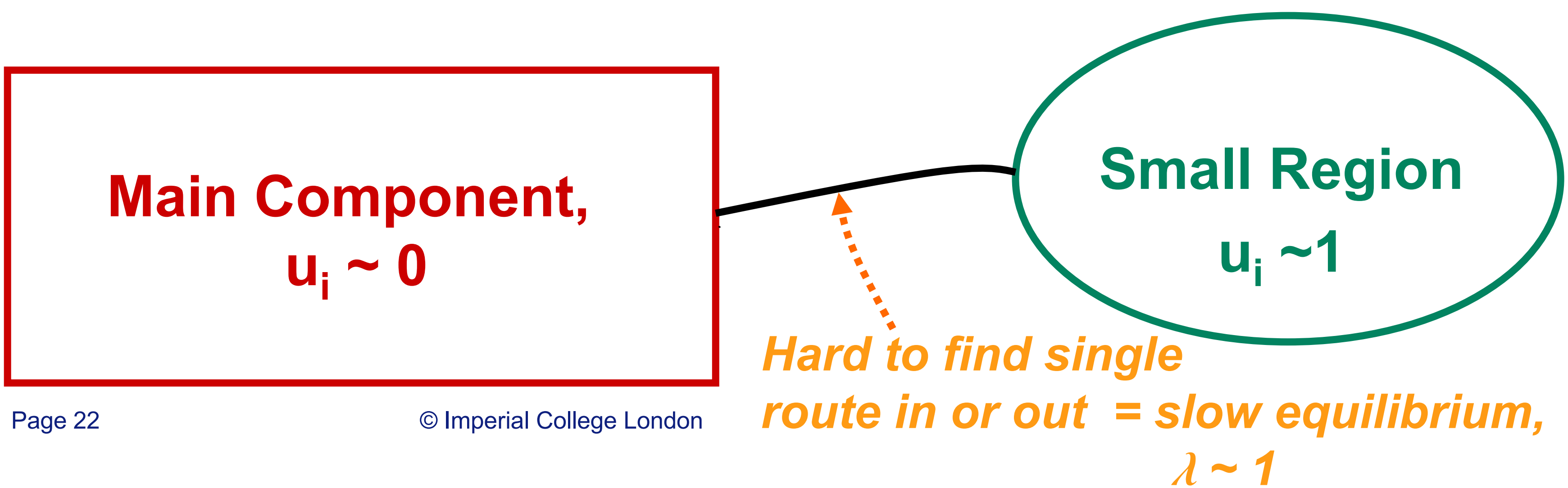
Simple Graph Diffusion Solution

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

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

Eigenvectors $u^{(n)}$ of largest eigenvalues tell us about small regions poorly connected to main component

(Eriksen et al 2003)



Diffusion as Ranking

- Long time solution gives a ranking of vertices

  Rank of vertex i = entry i of eigenvector of largest eigenvalue $u^{(1)}_{i}$

- Other types of walk

  = other types of diffusion

  = new weighted edges

  = new ranking scheme

  e.g. PageRank® (Google)

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

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

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

Conclusions

- Random Walk probes *global* structure of network but uses only *local* information
  $\Longrightarrow$ A Naturally Occurring Mechanism
  $\Longrightarrow$ Can lead to Self-Organisation
  $\Longrightarrow$ Useful Tool

- Used to grow network long power-law tails are a robust outcome with a wide variety of powers
  e.g. $N=10^6$ $\langle k \rangle=4$
  $\Longrightarrow$ $\gamma=3$ as $N,k \longrightarrow \infty$ in Simon/BA models
  Random Walk produces $2.4 < \gamma < 5$

I couldn't have done this without …

- Project Students
  Seb Klauke, JB Laloë, Christian Lunkes, Karl Sooman, Alex Warren
- ISCOM organisers
  David Lane, Sander van der Leeuw, Geoff West and all the ISCOM participants
- Collaborators
  Daniel Hook, Carl Knappet, Ray Rivers, Jari Saramäki

T.S.Evans, J.P.Saramäki
"Scale Free Networks from Self-Organisation"
Phys.Rev.E 72 (2005) 1 [**cond-mat/0411390**]

T.S.Evans,
"Complex Networks",
Contemporary Physics
45 (2004) 455 – 474
[**cond-mat/0405123**]

More Information

Following slides provide additional information.

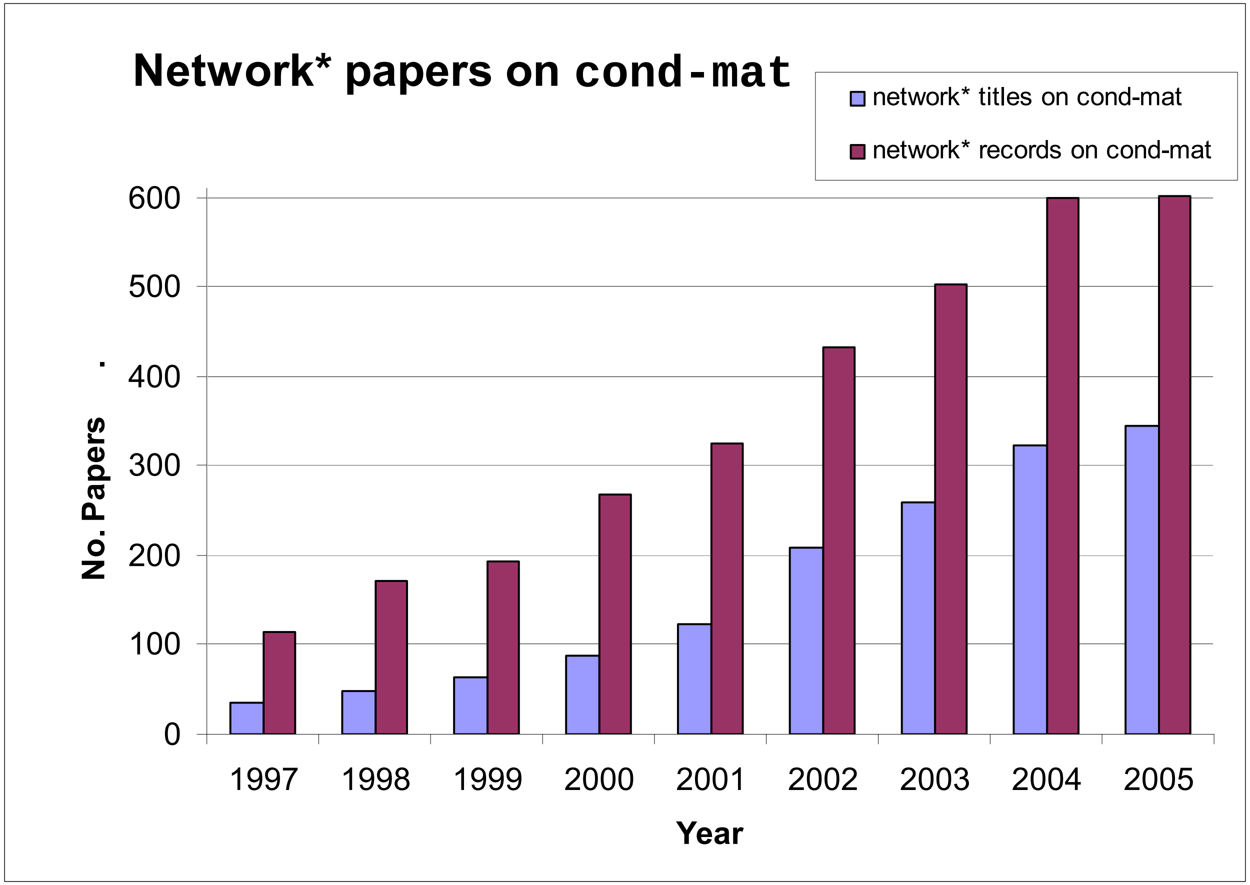
More Information

Following slides provide additional information.

Explosion of interest – WHY?

Since 1997 there has been an explosion of interest in networks by physicists.

For instance the condensed matter electronic preprint archives have gone from 35 papers in 1997 with a word starting with Network in their title to 344 last year, an increase of nearly 1000%



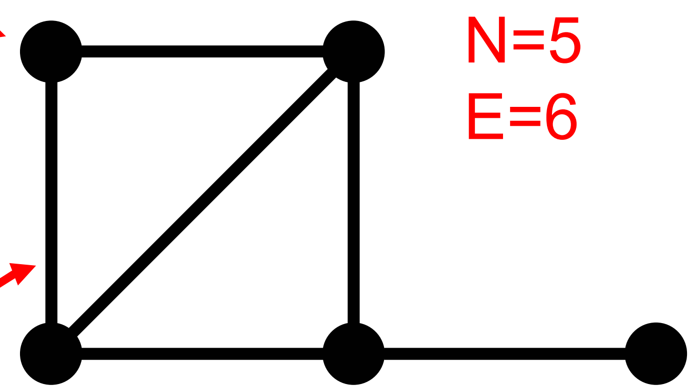
Updated from T.S.Evans, Contemporary Physics 45 (2004) 455 – 474 [cond-mat/0405123]

WHY?

Basic Definitions

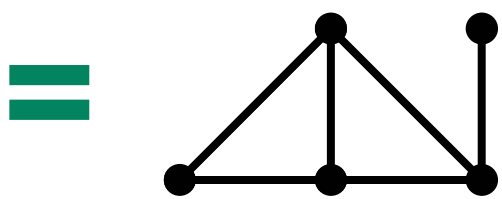
A **Network** or **Graph** is a collection of

N **Vertices** (nodes), pairs of which are connected by **E** **Edges**



This is a SIMPLE graph, it has no other information.

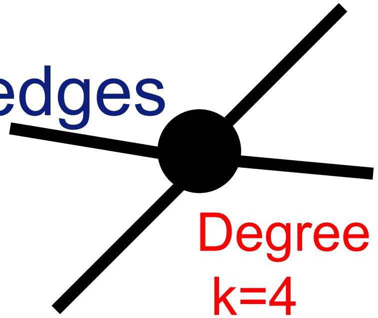
In particular the *same* network can be shown in several identical ways.



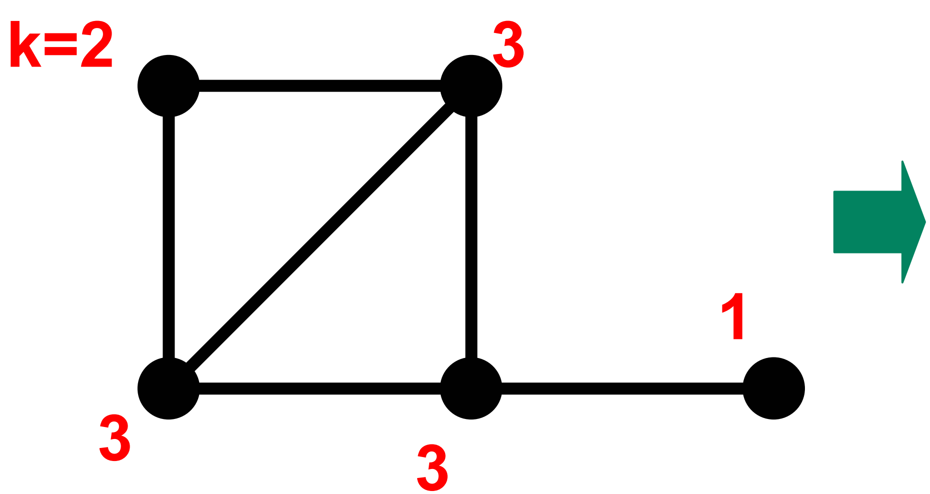
In general networks may have arrows on the edges (directed graphs), different values on edges (weighted graphs) or values to the vertices (coloured graphs).

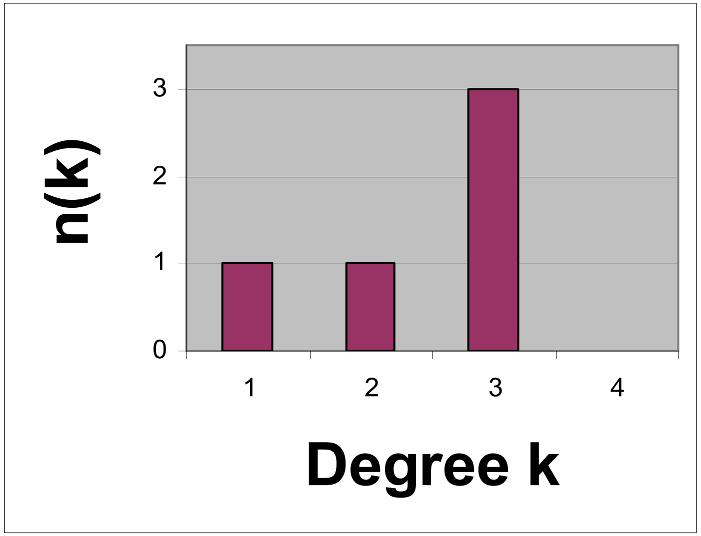
Degree (connectivity)

- The **Degree k** of a vertex is the number of edges attached to it.



- The **Degree Distribution n(k)** is the number of vertices with degree k

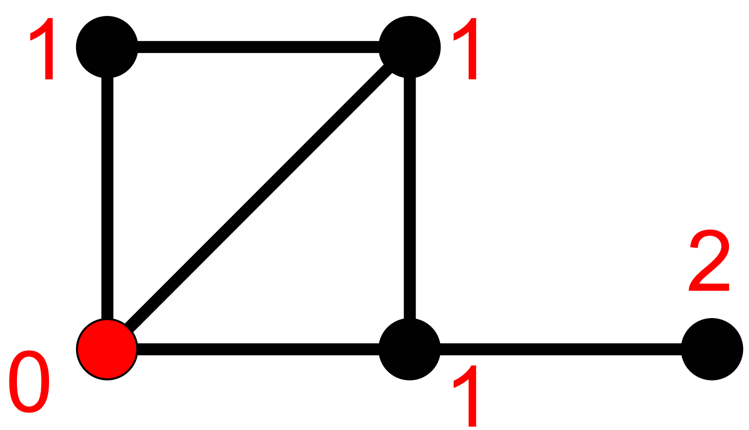




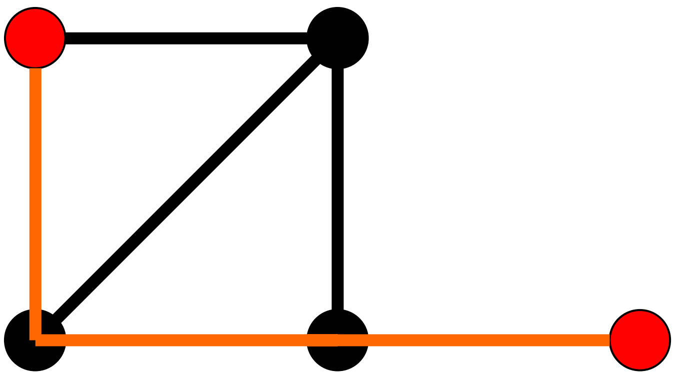
Network Distance

- Counting one for each edge traversed, we can find the shortest path between any two vertices, giving a distance between the two.

- The longest of these shortest paths is the diameter.



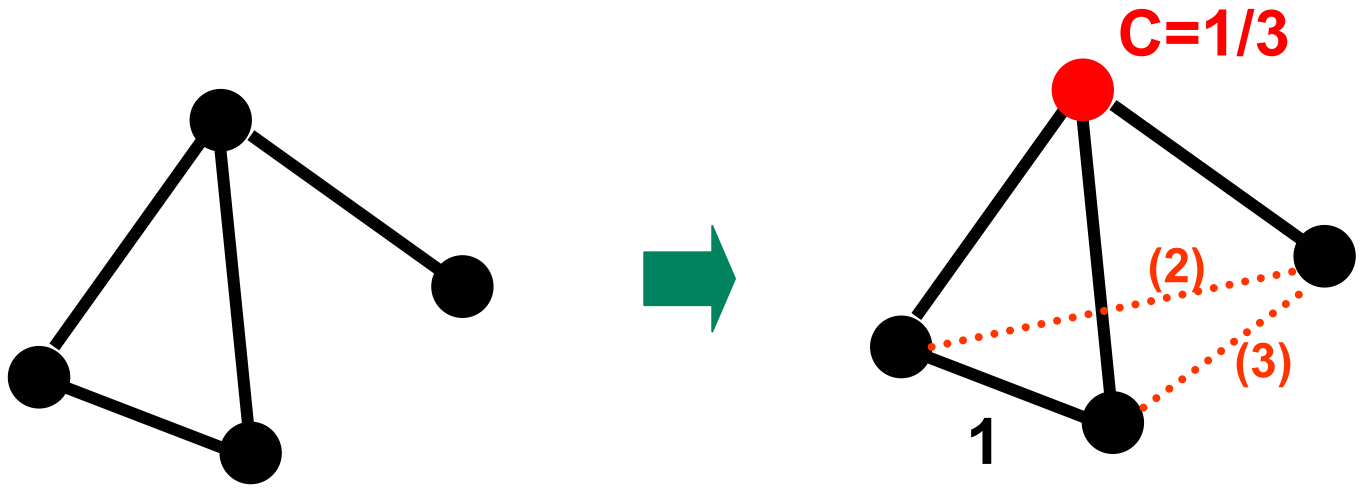
Distances from red vertex



Diameter is **3**, between red vertices

Cluster Coefficient

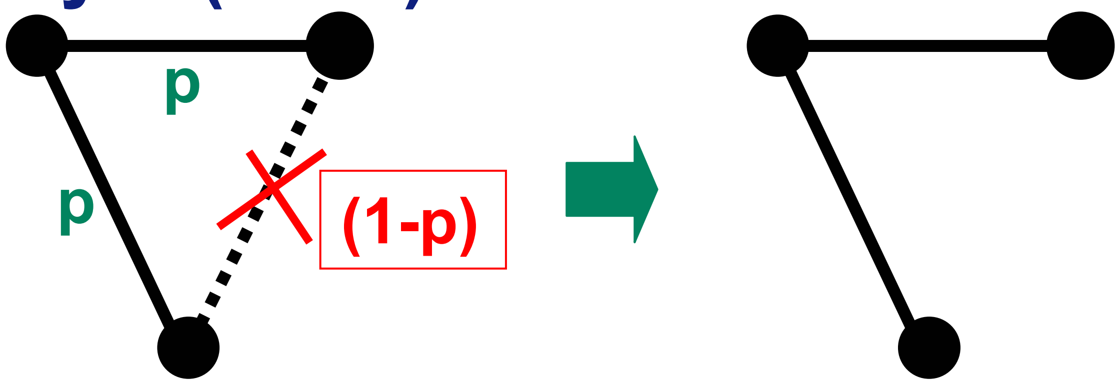
- **Clustering coefficient c**:
  Fraction of the neighbours which are themselves connected
  Simple measure of how much local structure there is in a network



Random Networks

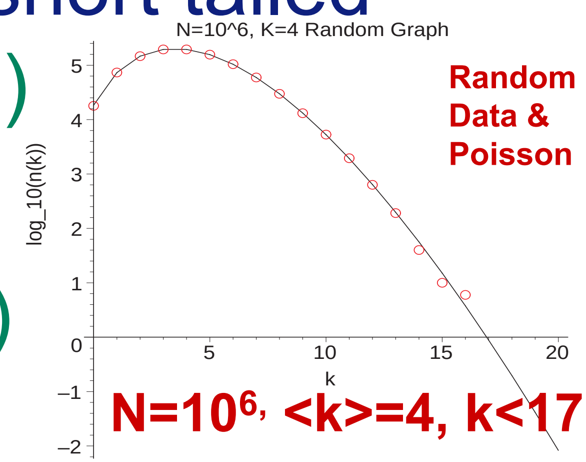
- Take **N** vertices then consider every pair of vertices and connect each with probability **p**
  Erdős-Reyní (1959).

N=3
p~2/3



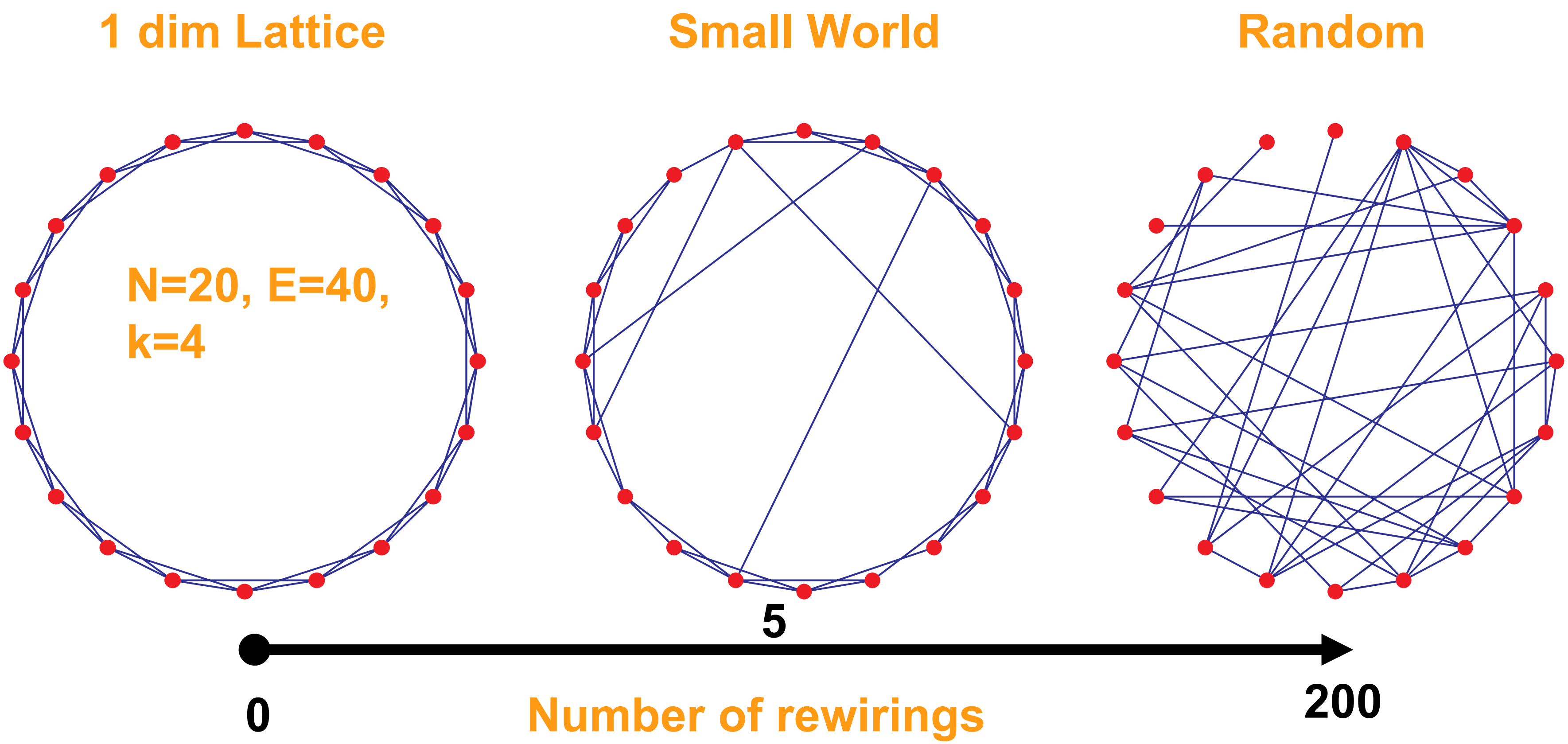
This is the opposite of the perfectly ordered lattice.

- Degree distribution is Poisson – short tailed
  ➡ Maximum Degree k_1 ~ ln(N)
- Little local structure c~1/N
- Short distances <d>~ln(N)



Watts and Strogatz's Small World Network (1998)

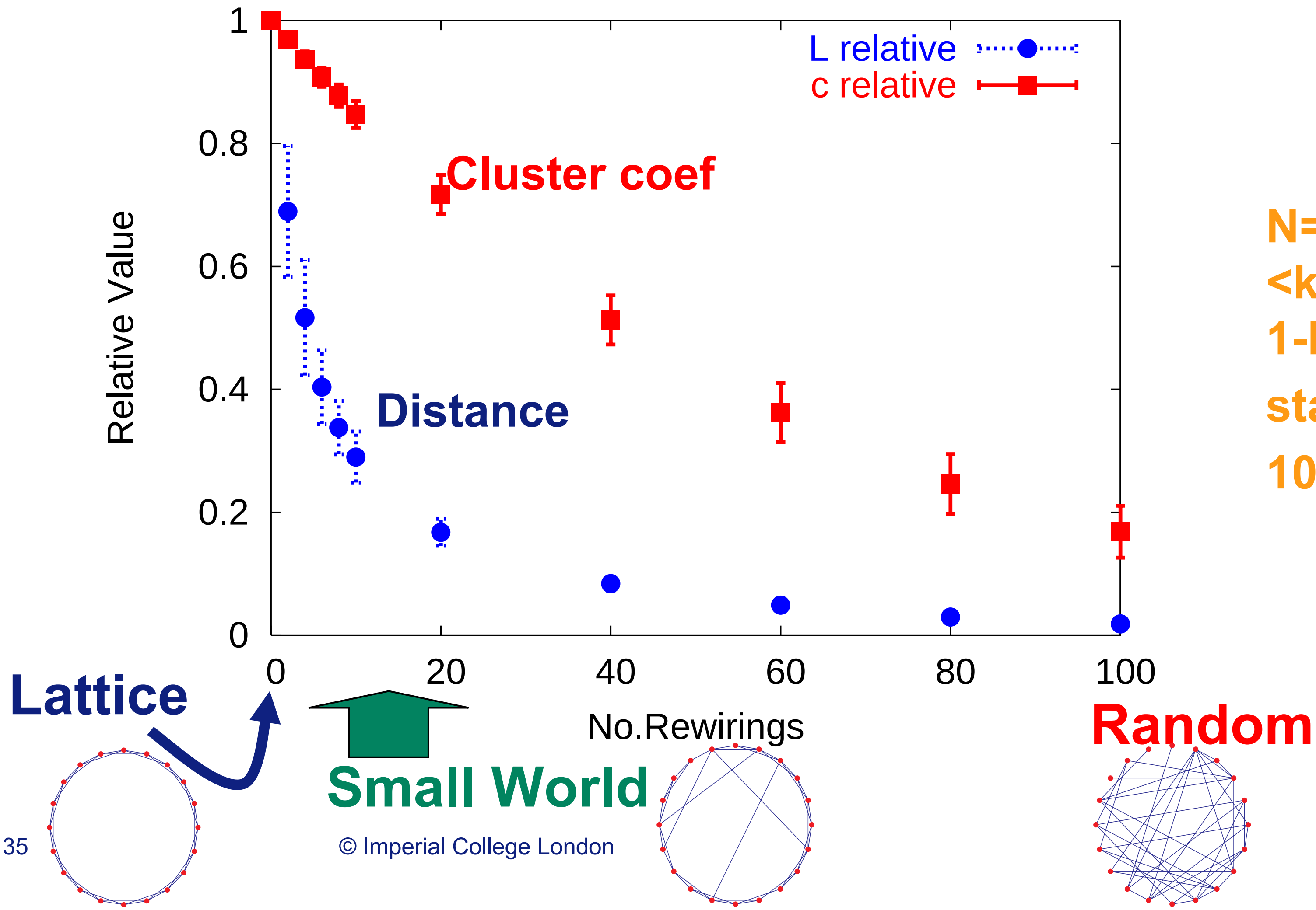
- Start with lattice, pick random edge and rewire it
  - – move ends to two new vertices chosen at random.



Clustering and Length Scale in WS network

- As you *rewire*, distance drops very quickly, clustering does not

➡ Find **SMALL WORLD NETWORKS** with short distances of random network, large clustering and local structure like a lattice



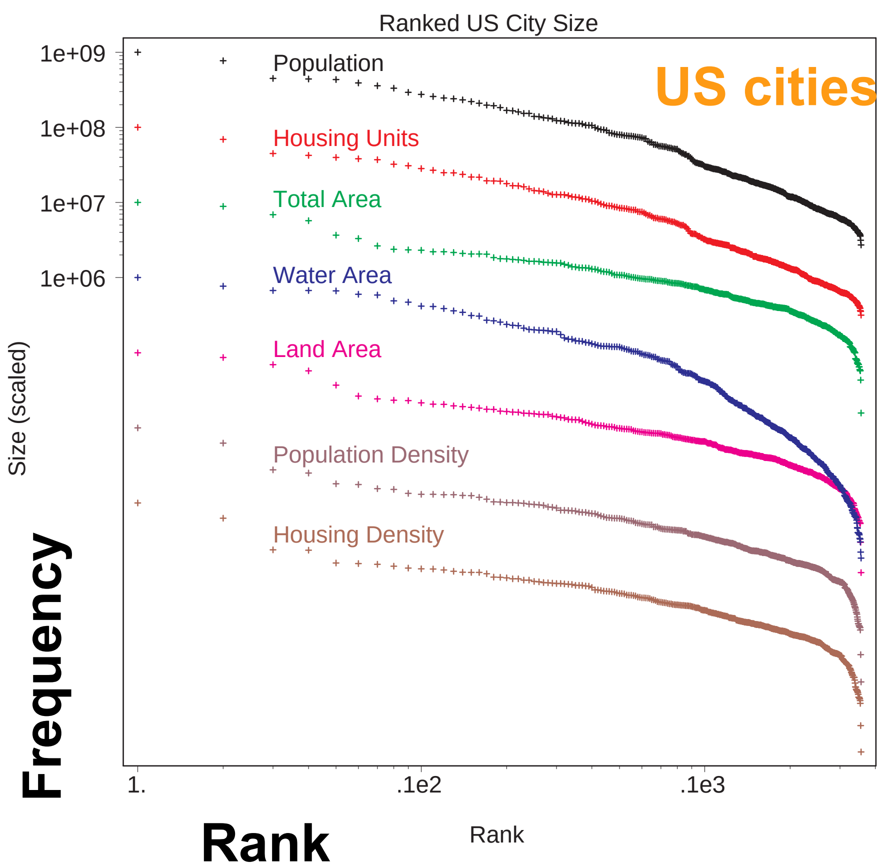
N=100, <k>=4, 1-Dim lattice start, 100 runs

Network Comparison

	Distance d	Degree Distrib. n(k)	Maximum Degree k_1	Cluster Coef. c
Lattice	Large $d \sim N^{1/dim}$	No Tail $\delta(k-k_0)$	Fixed k_0	~O(1)
Watts-Strogatz Small World	Small $d \sim \log(N)$	No Tail $\sim \delta(k-k_0)$	V.Small $\sim k_0$	~ O(1/N)
Erdős-Reyní Random	Small $d \sim \log(N)$	Short Tail Poisson $\langle k\rangle^k e^{-\langle k\rangle} / k!$	Small ~log(N)	~ O(1/N)
Scale-Free	Small $d \sim \log(N)$	Long Tail $\sim k^{-\gamma}$	Large = HUBS $\sim k^{1/(\gamma-1)}$	~ O(1/N)

Scaling in Social Sciences

- Zipf law (1949) – City Sizes, Text Frequencies,…
- Pareto's 80:20 rule (1890's)



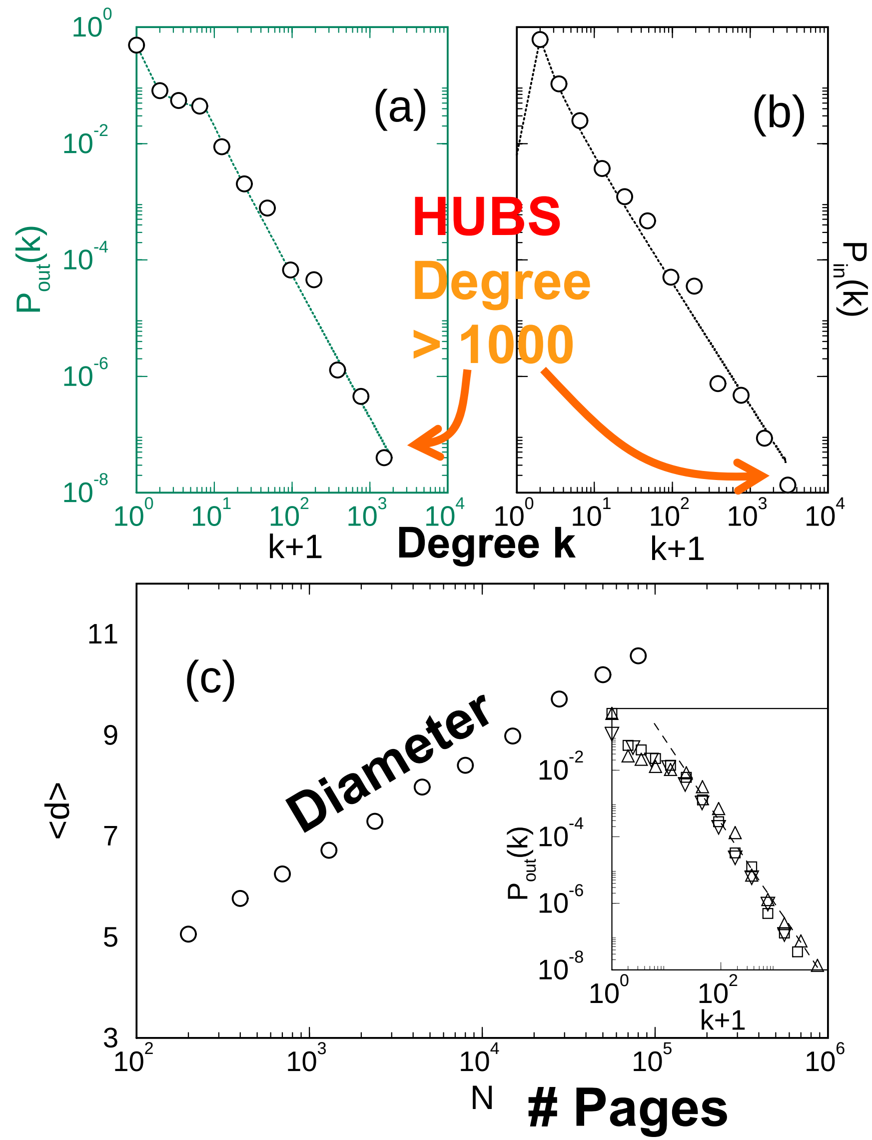
From T.S.Evans, Contemporary Physics 45 (2004) 455 – 474 [cond-mat/0405123]

The World Wide Web

- Every web page is a vertex, every link is an edge
- A few pages have a tremendous amount of links to them e.g. college home page, eBay, Google

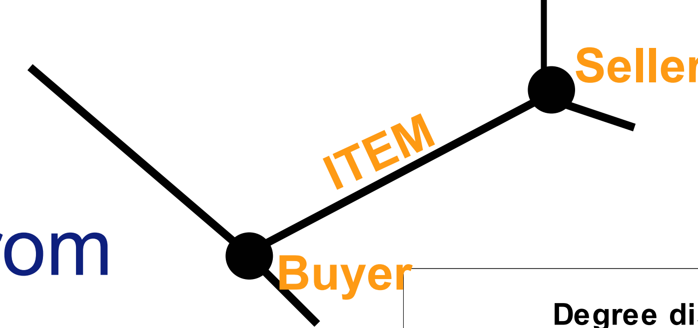
These are **Hubs** and they are a key aspect of how we navigate and use the web

Log-Log plot of degree distribution of nd.edu

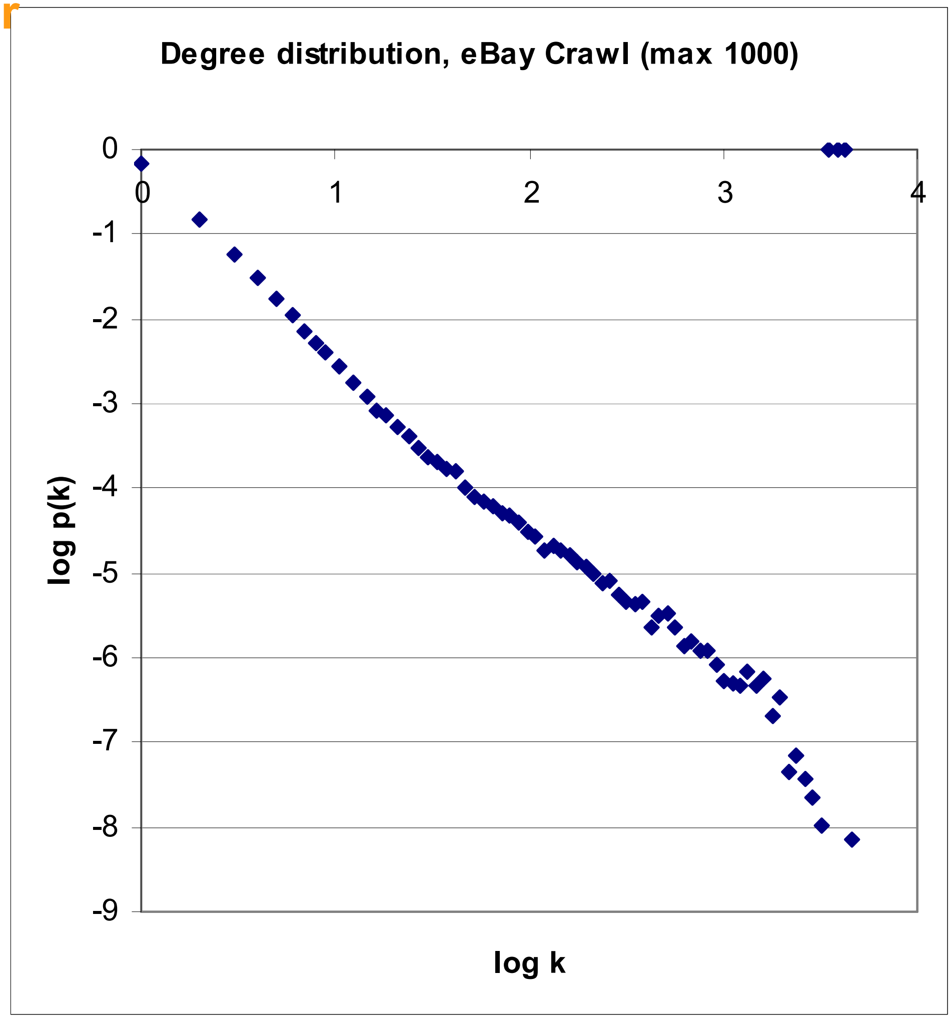


(Barabasi, Albert, Jeong 1999)

eBay

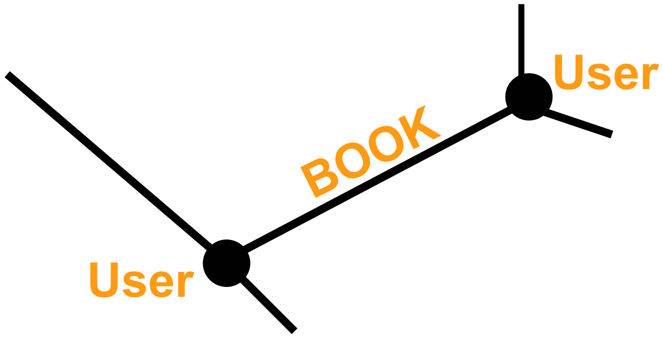


- Network from buyer/seller feedback links
- eBay is dominated by a few very large hubs.
  - The slight curvature due to crawling method.
  - Fetched 5,000 pages and built up a network of 318,000 nodes and 670,000 edges
- $\gamma \approx 2.3$

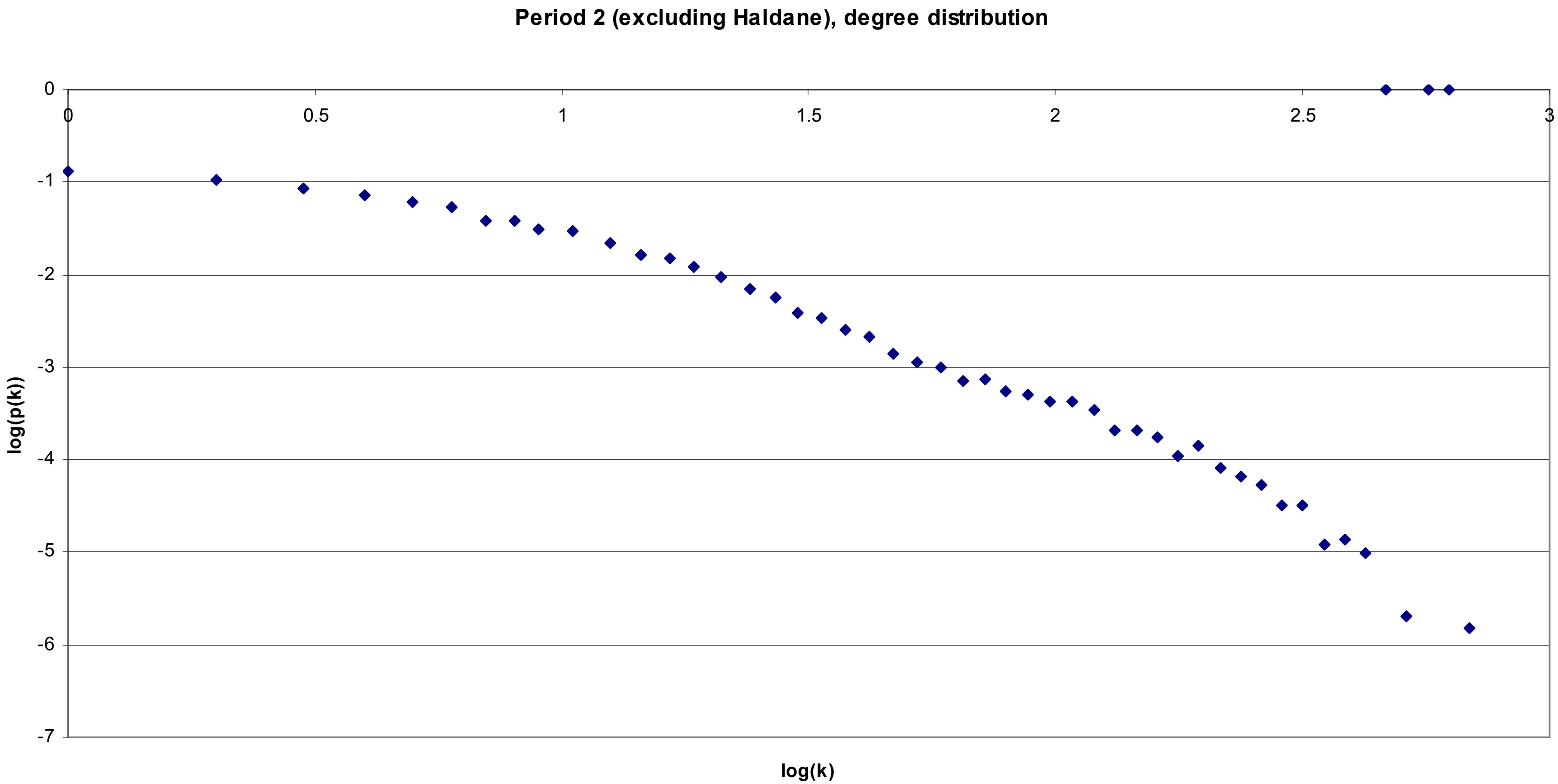


Sooman, Warren, TSE (2004)

Imperial Library



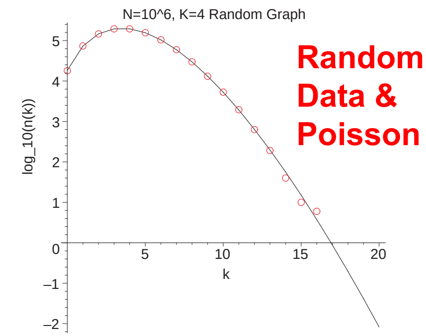
- Used to detect groups from lending patterns

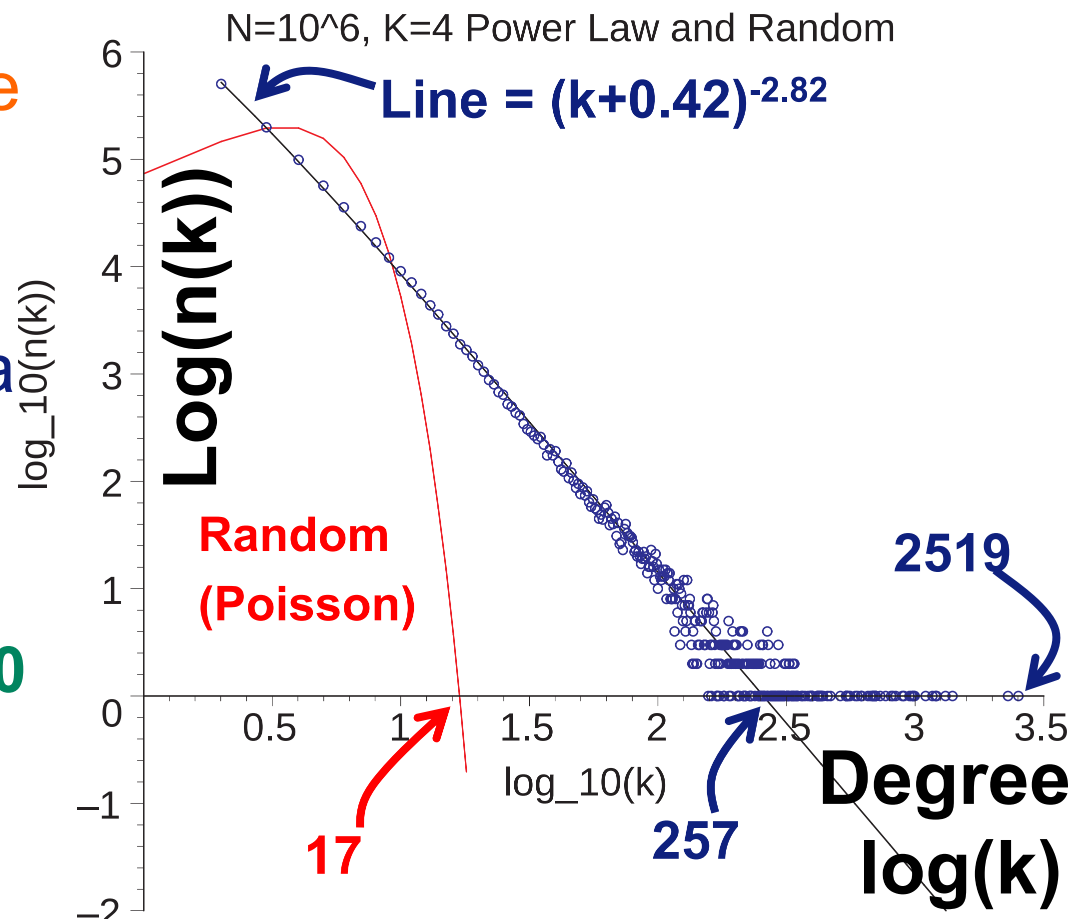


Sooman, Warren, Hook, TSE (2004)

What sort of network has hubs?

- Lattices, WS (Watts-Strogatz) Small World and random networks have no hubs, e.g. the largest degree is **17** for a random network with $N=10^6$, $<k>=4$
- Want a network with a **long tailed** degree distribution e.g. power law ~ k^{-3} has max. degree ~**2520** for $N=10^6$ $<k>=4$





From T.S.Evans, Contemporary Physics 45 (2004) 455 – 474 [cond-mat/0405123]

Long Tails Description

- Most data sets have "long" tails for degree distribution

  Characterised by a few vertices with many edges - **HUBS**

  e.g. maximum degree

$$k_{max} = O(N^{\nu}) >> O(\log(N)) \qquad \nu > 0$$

Power Law: $\nu = ½$ Poisson

- These can often be reasonably described by a power-law

$$n(k) \sim k^{-\gamma} \qquad (2 < \gamma < \infty \text{ if } N \rightarrow \infty, K < \infty)$$

BUT note that many other functions give reasonable fits too!

Models

- Short Exponential Tails

$$\lim_{k\to\infty} [n(k)] \sim \exp(k/k_{scale})$$

e.g. $N=10^6$, $<k>=4 \implies k_{max}=17 = O(\log(N))$

-Random Graph Erdős-Reyní (1959) (Poisson)

-Watts-Strogatz Small World (1998)

-Growing with Random Attachment

- Scale-Free = Long Power-Law Tails

$$\lim_{k\to\infty} [n(k)] \sim k^{-\gamma} \qquad 2<\gamma<\infty$$

e.g. $N=10^6$, $<k>=4 \implies k_{max}=2520 = O(N^{1/2})$

-Simon (1955) [graph can be added easily]

-Barabasi-Alberts (1999) [graph not required]

Scale Free Networks

- Any network with a **power law degree distribution** for large degrees

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

- Always have many large **Hubs** nodes with many edges attached – e.g. routers in the internet
- **Scale Free** means the number of vertices of degree 2k with those of degree k, always the same whatever k , that is there is no scale for degree

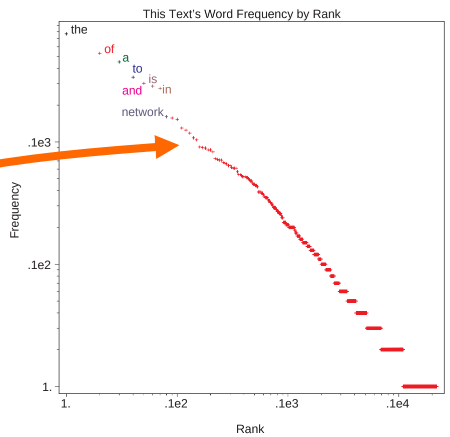
$$\frac{n(2k)}{n(k)} = \text{constant}$$

- In practice there are at least two scales for finite N:

$$O(1) \sim k_{min} \leq k \leq k_{max} \sim O(N^{1/(\gamma-1)})$$

Power Laws in the Real World

- 2^{nd} Order Phase Transitions
  (e.g. superconductors, superfluids,…)
  Long range order = no scale = physical insight
  Critical Phenomena – Renormalisation Group
- Scaling in Particle Physics
- Biology
  – Kleiber's Law (1930's) metabolic rate **r** $\propto$ **m**$^{3/4}$ body mass, explained (West, Brown, Enqvist 1997)
- Social Sciences
  – Zipf's Law (1949)
  City sizes,
  Word frequency, …
  ➔file compression



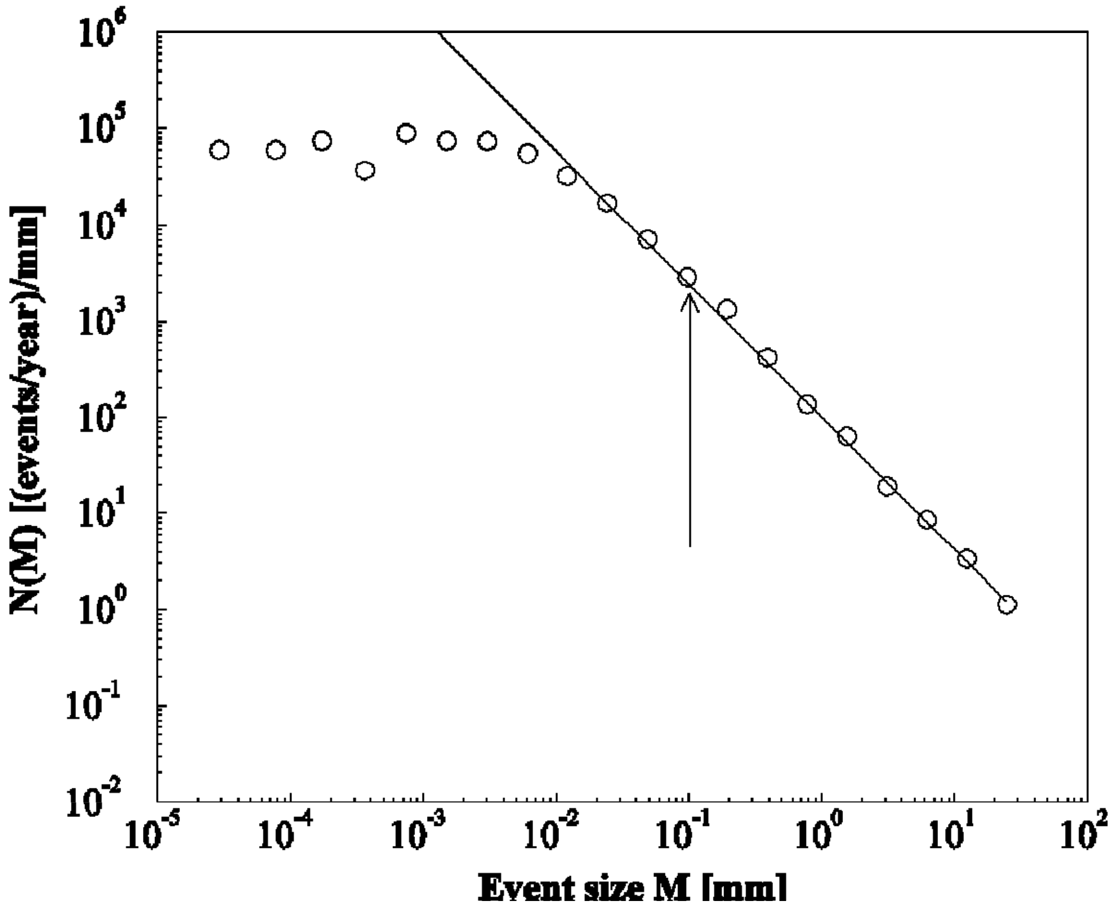
From **T.S.Evans, Contemporary Physics 45 (2004) 455 – 474 [cond-mat/0405123]**

Scaling in Complex Systems

- Earthquakes (Gutenberg-Richter Law), forest fires, rice piles, rainfall distributions, etc etc

  Self-Organised Criticality

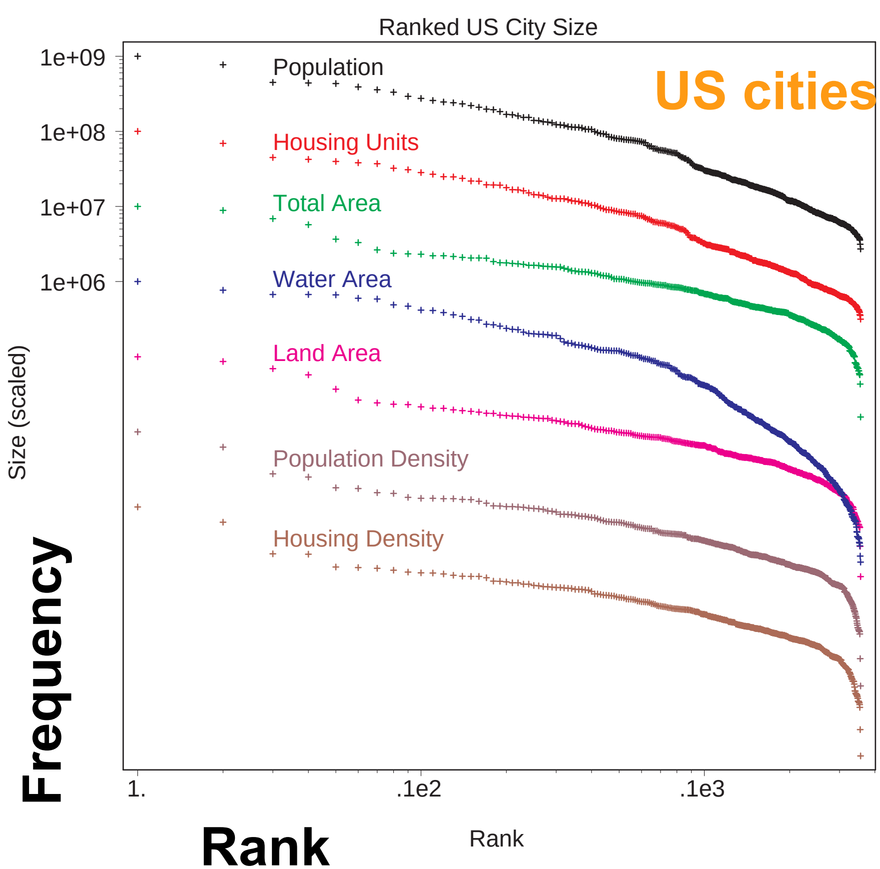
- **Still leading to further physical insights**

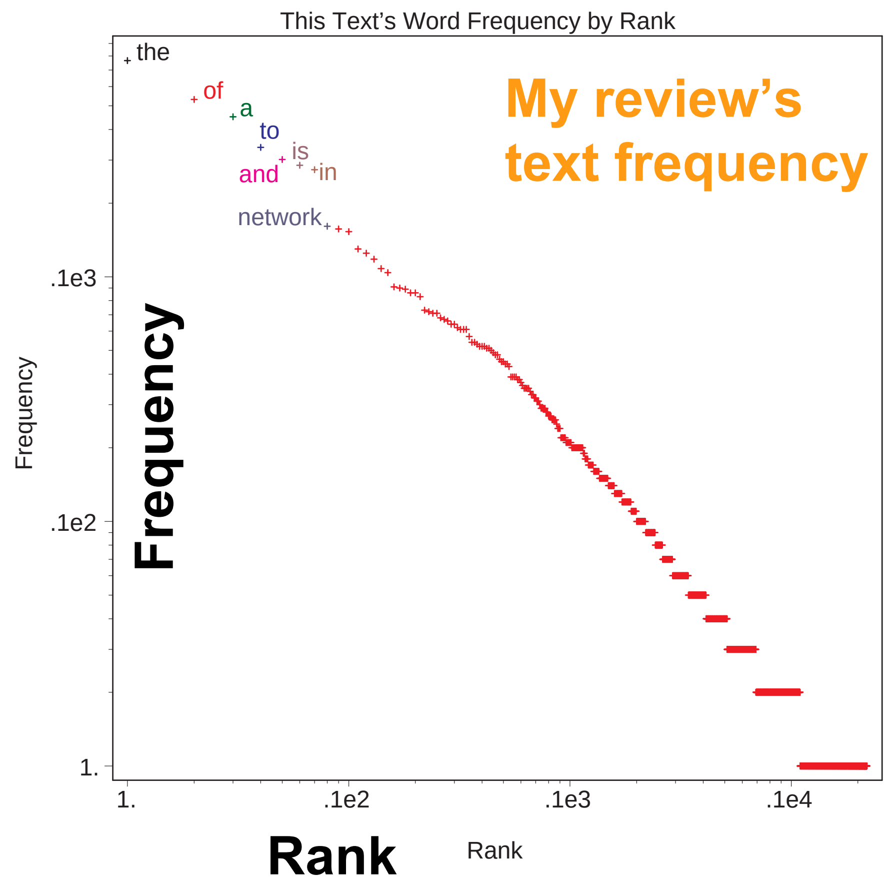


(Peters, Hertlein, Christensen 2002)

Scaling in Social Sciences

- Zipf law (1949) – City Sizes, Text Frequencies,…
- Pareto's 80:20 rule (1890's)





From T.S.Evans, Contemporary Physics 45 (2004) 455 – 474 [cond-mat/0405123]

Scaling with every network

- Friendship networks - Kevin Bacon game
- Scientific Collaboration Networks - Erdős number
- Scientific Citation Networks
- Word Wide Web
- Internet
- Food Webs
- Language Networks
- Protein Interaction Networks
- Power Distribution Networks
- Imperial Library Lending Data (Laloe, Lunkes, Sooman, Warren, Hook, TSE)
- eBay relationships (Sooman, Warren, TSE)
- Greek Gods
- Marvel Comic Heroes

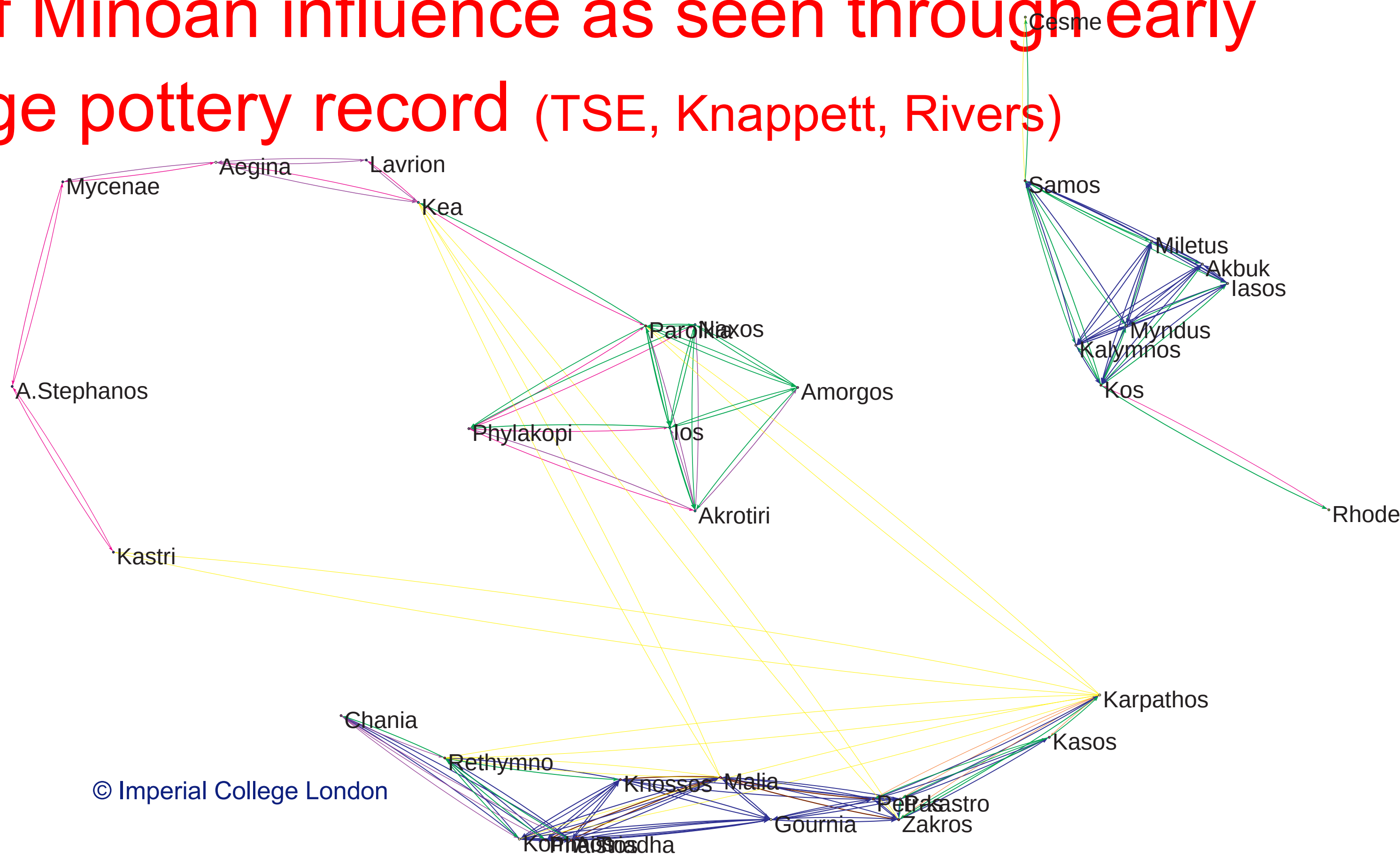
Scaling – a health warning

Almost every network is scale free if you believe the literature but

- Not many decades of data
  e.g. 10^6 vertex scale free network has largest vertex about 1000 so at most two decades of large degree scaling
- Data often a single data set no repeats
- Errors unknown in much social science data
- Other long tailed distributions have hubs too

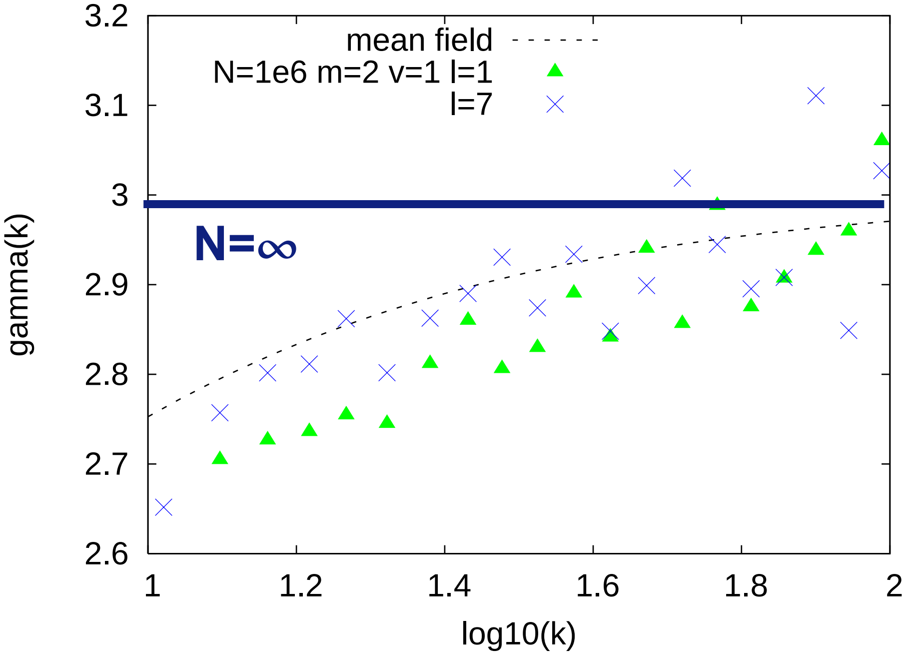
Applications: Archaeological Networks

- Ceremonial Pig exchange networks in Polynesia (Hage & Harary)
- Central role of Delos in ancient Greek culture (Davis)
- Spread of Minoan influence as seen through early bronze age pottery record (TSE, Knappett, Rivers)



What is the power?

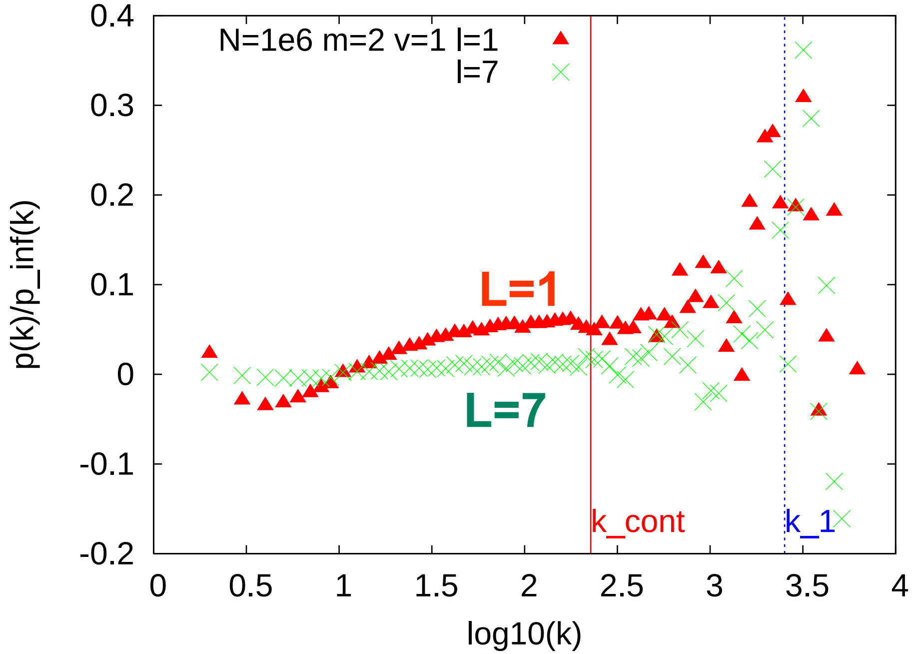
- Local power always below asymptotic value of 3



$$\gamma(k) = \frac{n(kr) - n(k)}{k(r-1)}$$

(r=1.1)

- However long walks fit mean field asymptotic solution very well



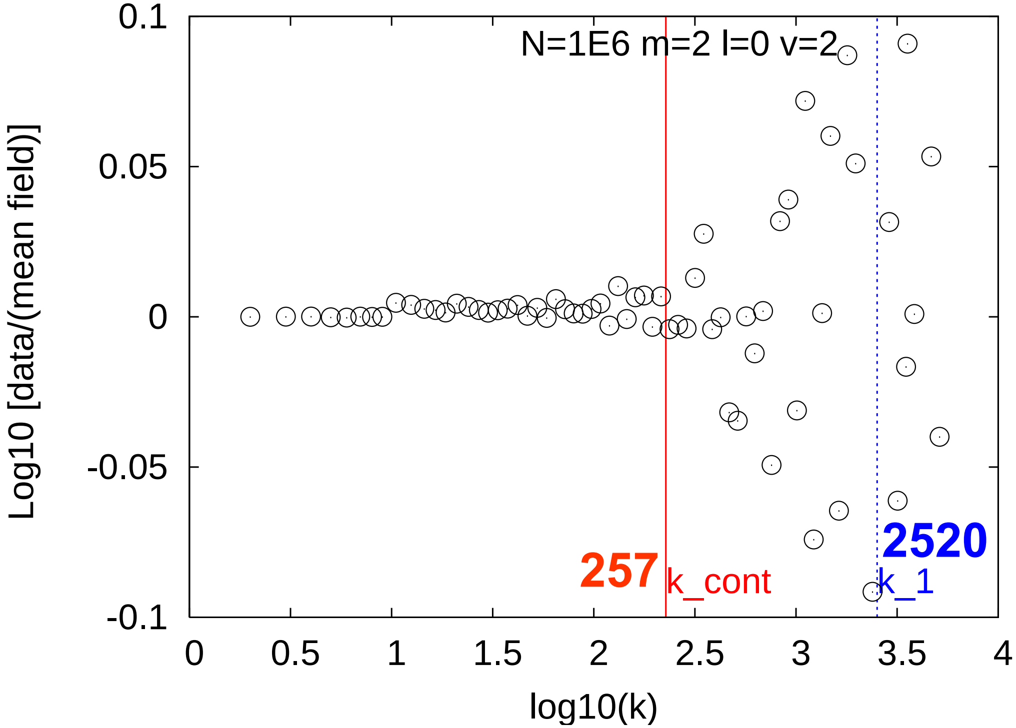
Data generated by walk method divided by predicted long time mean field value

From T.S.Evans, J.P.Saramäki **Phys.Rev.E 72 (2005) 1** [`cond-mat/0411390`]

Mean Field Solutions

- Assume behaviour of the average number of vertices of degree k given by the average properties of the network
- These are excellent for pure preferential attachment (Simon/BA)
$\Leftrightarrow$
correlations in degrees of neighbouring vertices insignificant

Fractional deviation of data from one run of pure pref. attachment model against mean field solution



$n(k_{cont}) := 1$

$k_{cont} = O(N^{1/3})$

Limit of good data

From T.S.Evans, J.P.Saramäki **Phys.Rev.E 72 (2005) 1** [cond-mat/0411390]

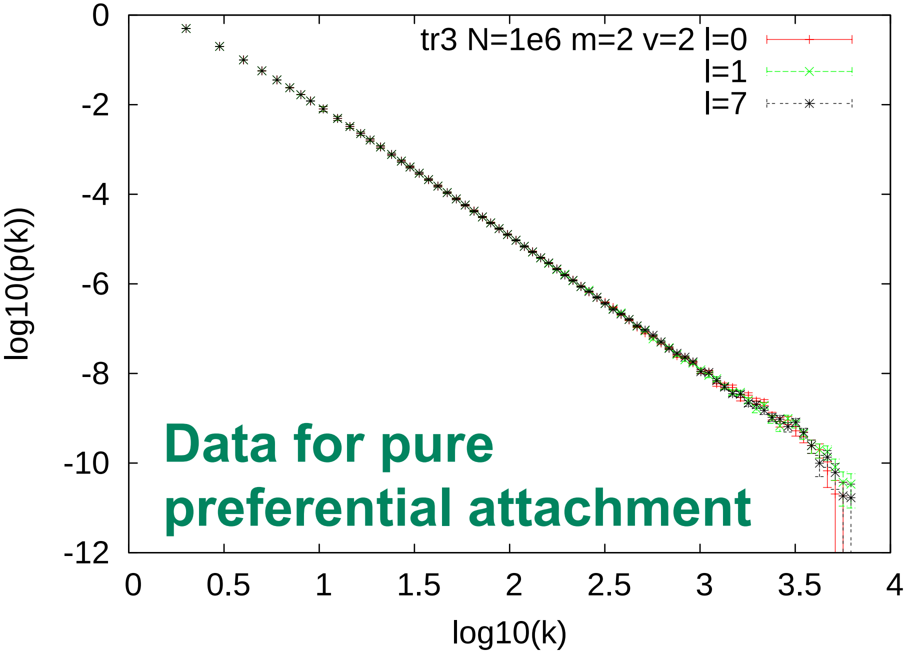
Finite Size Effects for pure preferential attachment

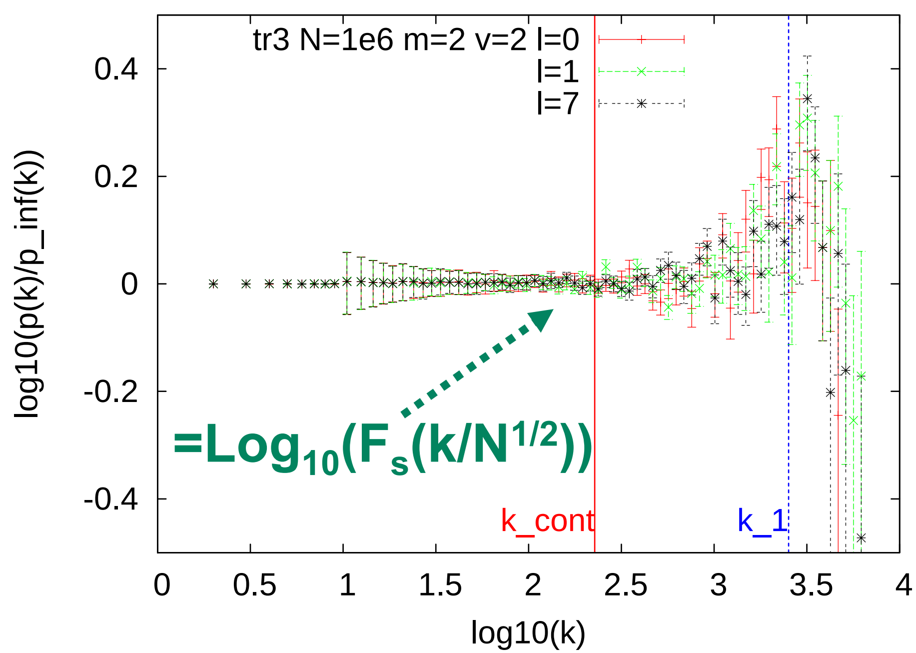
$$p(k) = p_\infty(k).F_S\left(\frac{k}{N^{1/2}}\right), \qquad p_\infty(k) = \frac{\langle k\rangle(\langle k\rangle+2)}{2k(k+1)(k+2)}$$

$$\rightarrow \frac{1}{k^3}$$

Scaling Function F_s

$$F_S(x) \approx 1 \quad if \quad x < 1$$





From T.S.Evans, J.P.Saramäki

Phys.Rev.E 72 (2005) 1 [cond-mat/0411390]

100 runs to get enough data near k_1