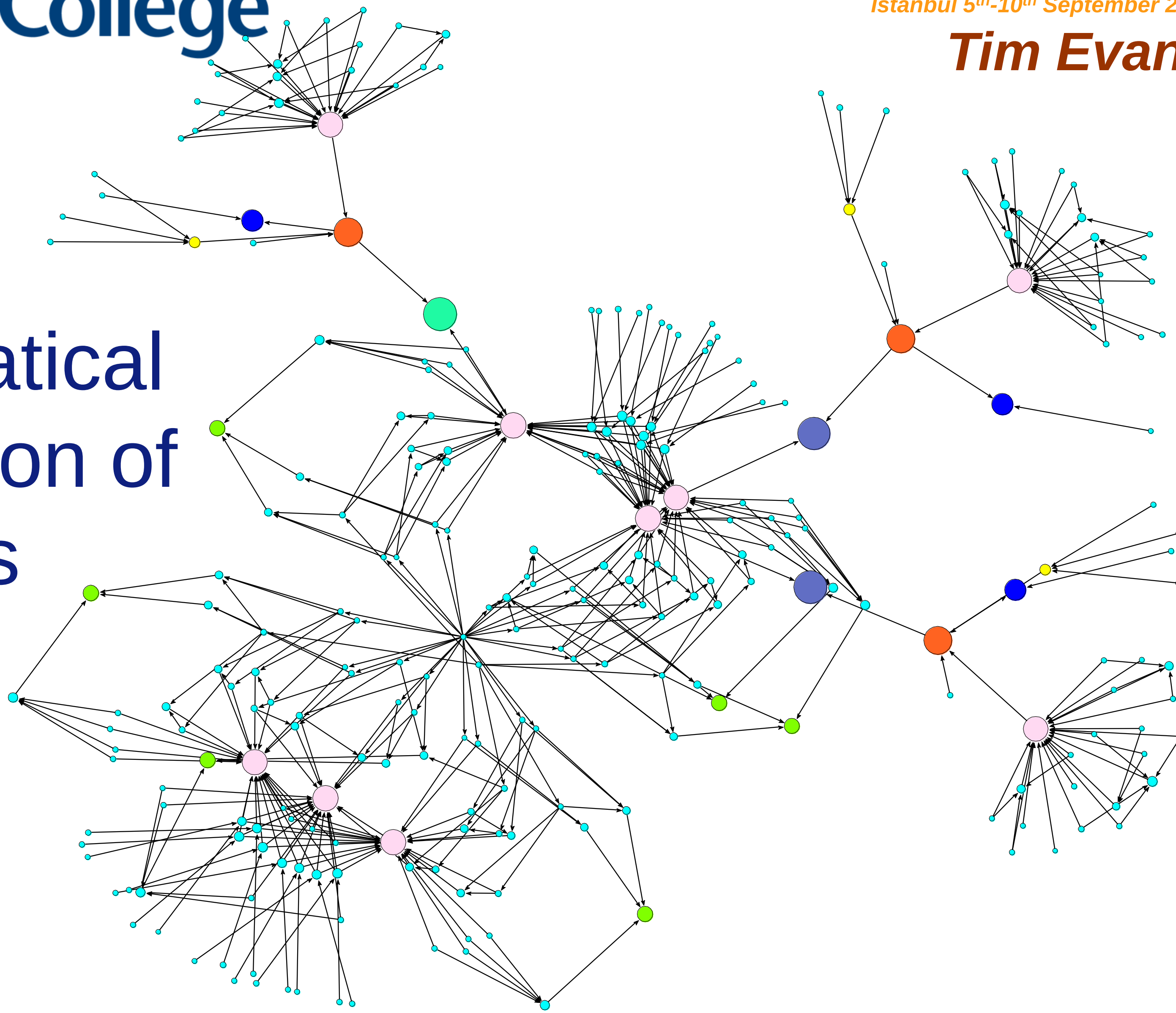
Imperial College London

Feza Gürsey Institute - Imperial College International Summer School and Research Workshop on Complexity, Istanbul 5th-10th September 2011

Tim Evans

The Mathematical Description of Networks

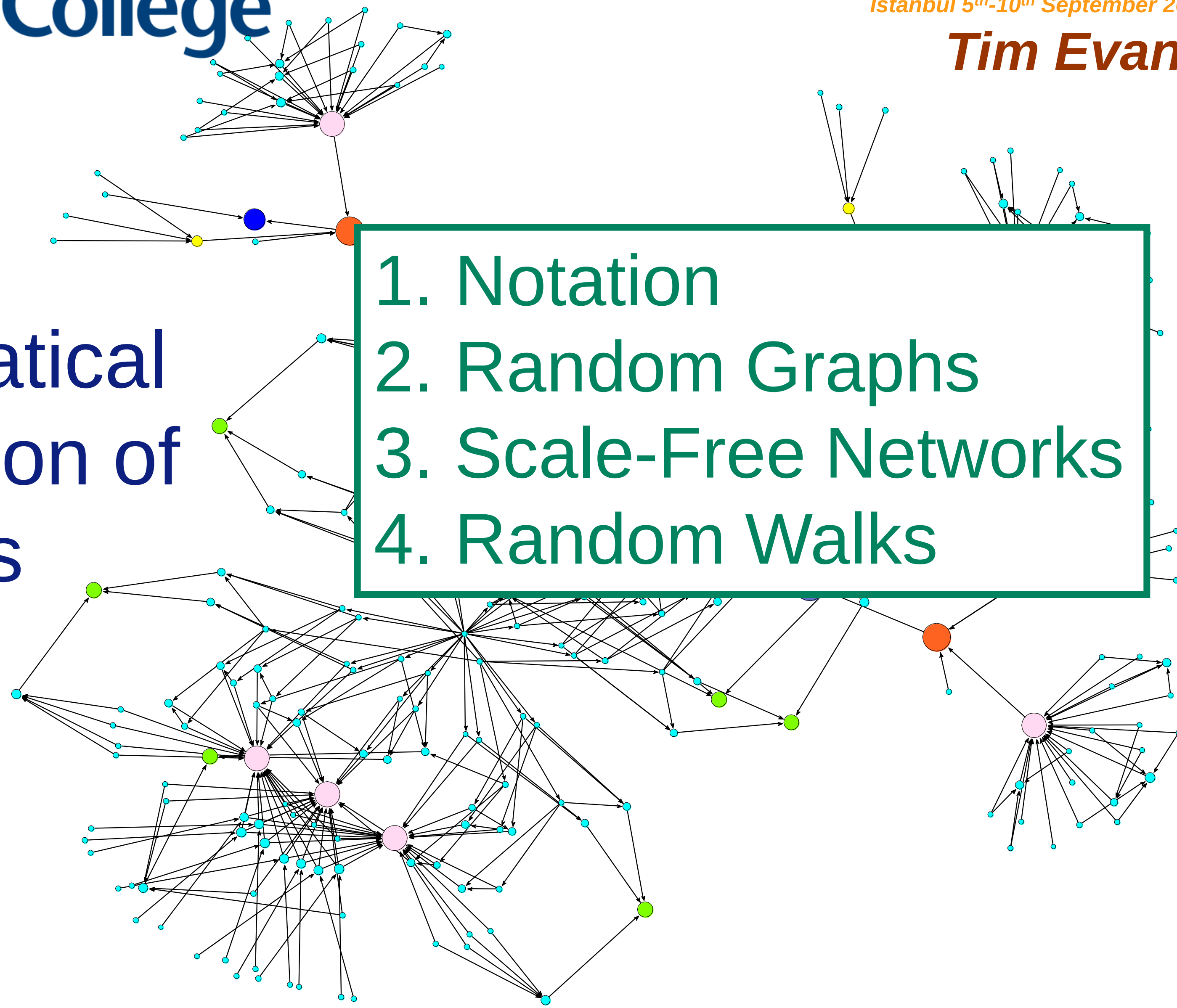


Imperial College London

Feza Gürsey Institute - Imperial College International Summer School and Research Workshop on Complexity, Istanbul 5th-10th September 2011

Tim Evans

The Mathematical Description of Networks



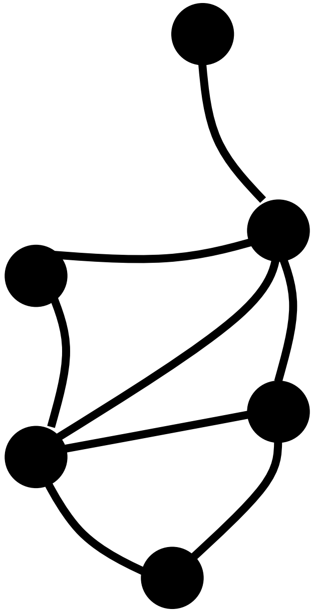
1. Notation
2. Random Graphs
3. Scale-Free Networks
4. Random Walks

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*



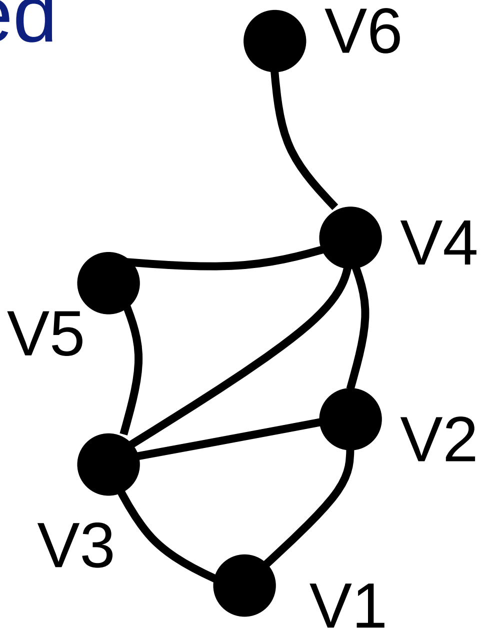
N=6
E=8

Notation - Adjacency Matrix

The ***Adjacency Matrix*** $\boldsymbol{A_{ij}}$ is

- **1** if vertices ***i*** and ***j*** are attached
- **0** if vertices ***i*** and ***j*** are not attached

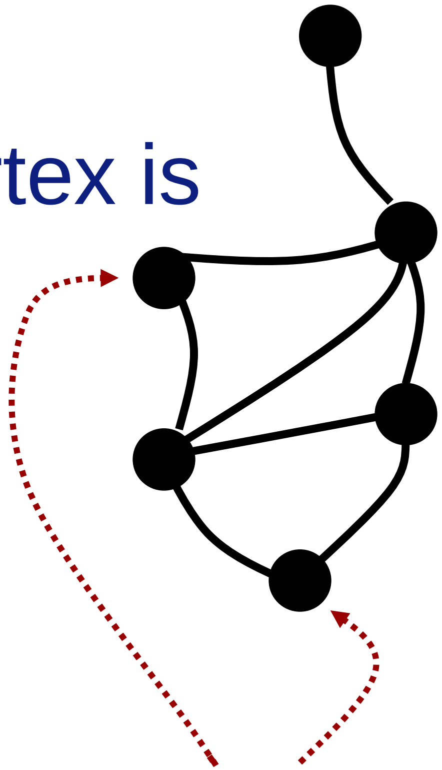
vertices	V1	V2	V3	V4	V5	V6
V1	0	1	1	0	0	0
V2	1	0	1	1	0	0
V3	1	1	0	1	1	0
V4	0	1	1	0	1	1
V5	0	0	1	1	0	0
V6	0	0	0	1	0	0



Notation – degree of a vertex

Number of edges connected to a vertex is called the ***degree*** of a vertex

- ***k*** = *degree of a vertex*
- ***<k>*** = *average degree* = ***(2E / N)***



- Degree Distribution

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

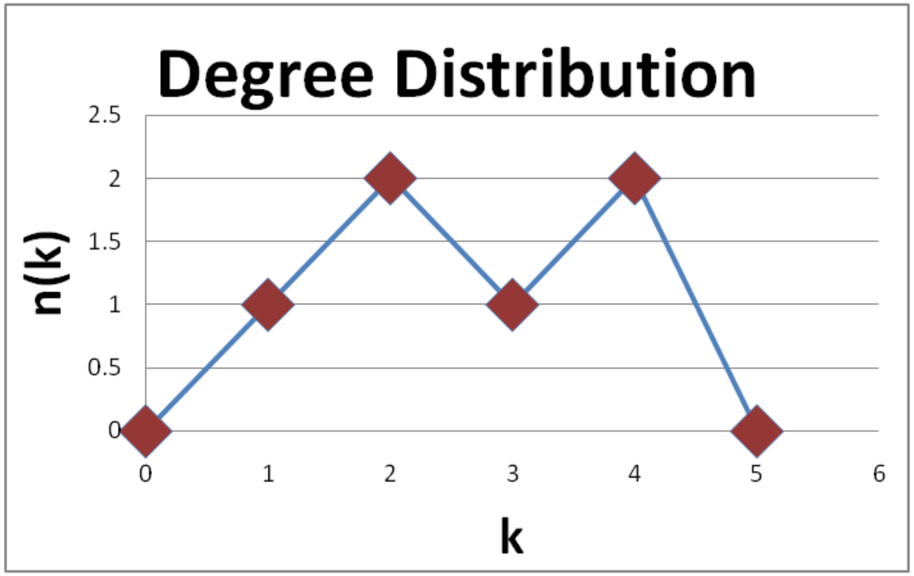
  p(k) = ***n(k)/N*** = *normalised distribution*

  = *probability a vertex chosen at random (uniformly) has degree* ***k***

Notation – degree distribution

Degree Distribution is

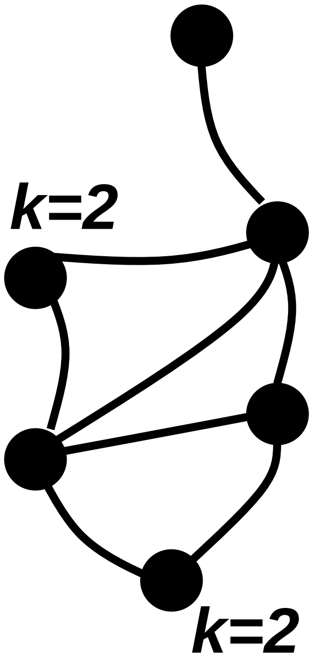
n(k) = *number of vertices with degree* ***k***

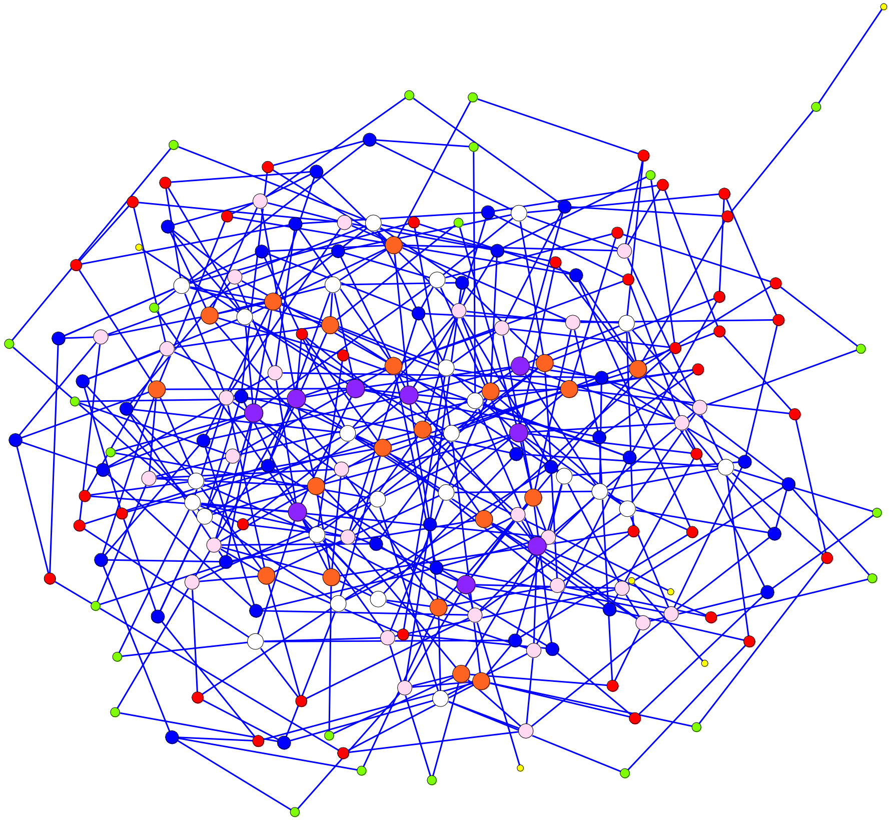


The normalised degree distribution is

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

= *probability a vertex chosen at random (uniformly) has degree* ***k***





How to excite a Mathematician –
give them the simplest model

RANDOM GRAPHS

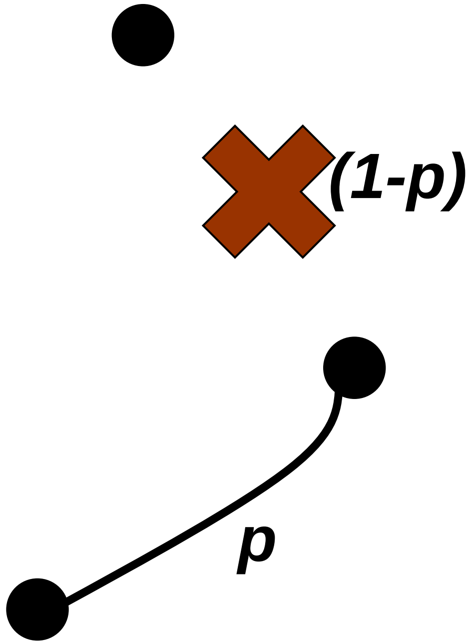
Classical Random Graphs

[Solomonoff-Rapoport ‘51, Erdős-Réyni ’59]

For every pair of distinct vertices add a single edge with probability

$p = \langle k \rangle/(N-1)$,

otherwise with probability $(1-p)$ no edge is added



Classical Random Graph

- Gives Binomial Degree Distribution

$$p(k) = \binom{\Omega}{k} p^k (1-p)^{\Omega - k}$$

with Ω **= *N(N-1)/2*** number of possible edges
and ***<k>=(N-1)p***

Classical Random Graph

- which is an approximate Normal Distribution

$$p(k) \approx \frac{\exp\left(\langle k \rangle\right)\langle k \rangle^{k}}{k!}$$

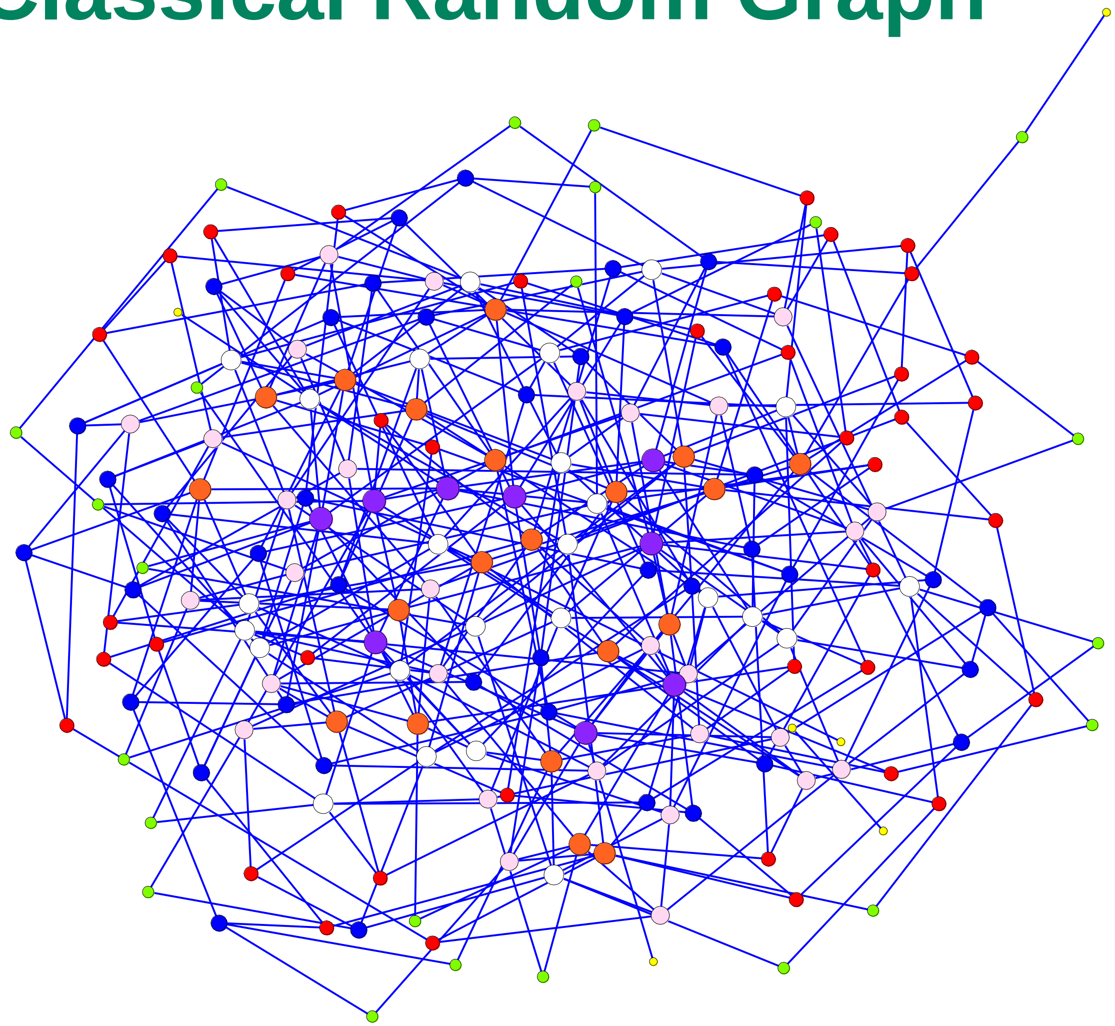
with ***<k>=(N-1)p***

- Exponential cutoff so no 'hubs'

  e.g. ***N=10^6***, ***<k>=4.0***, typically has ***k <17***

Example of Classical Random Graph

- ***N=200 <k>~4.0***
- ***k < 11***
- In figure vertex size $\propto$ ***k***
- ***Diffuse, no tight cores***

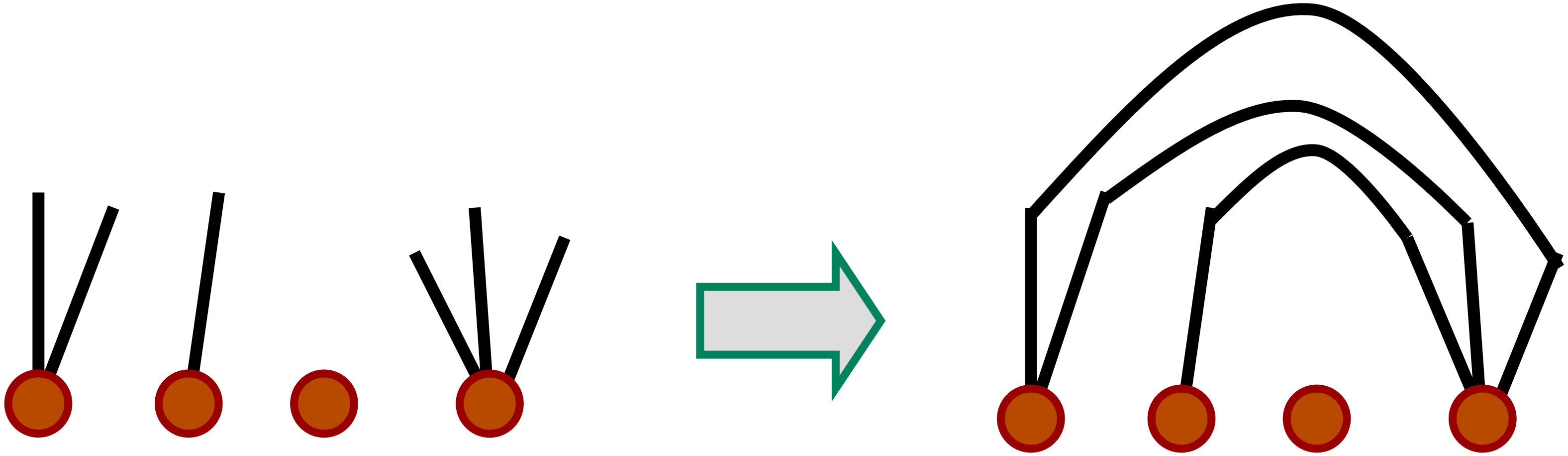


Generalised Random Graphs – The **Molloy-Reed** Construction [1995,1998]

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



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

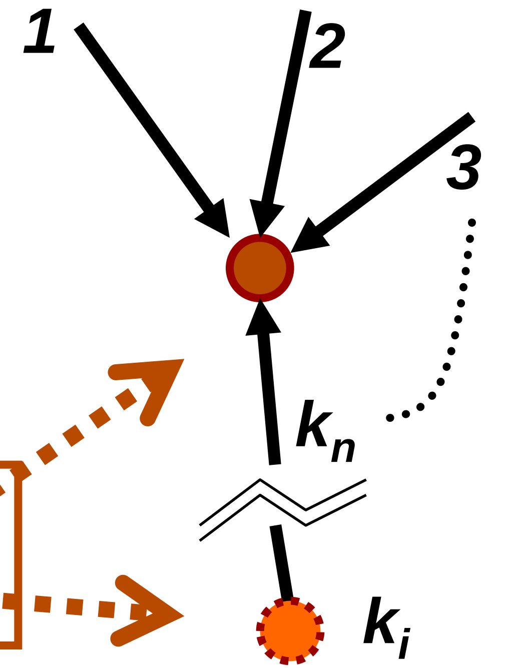
Random Walks on Random Graphs

The degree distribution of a neighbour is not simply ***p(k)***

You are more likely to arrive at a high degree vertex than a low degree one

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

Degree of neighbour k_n ***independent of degree of starting point*** k_i



A random friend is more popular than you

$$\langle k_n \rangle = \sum_{k_n} p(k_n \mid k_i) = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

(Number of friends neighbour has)

-

(Number of your friends)

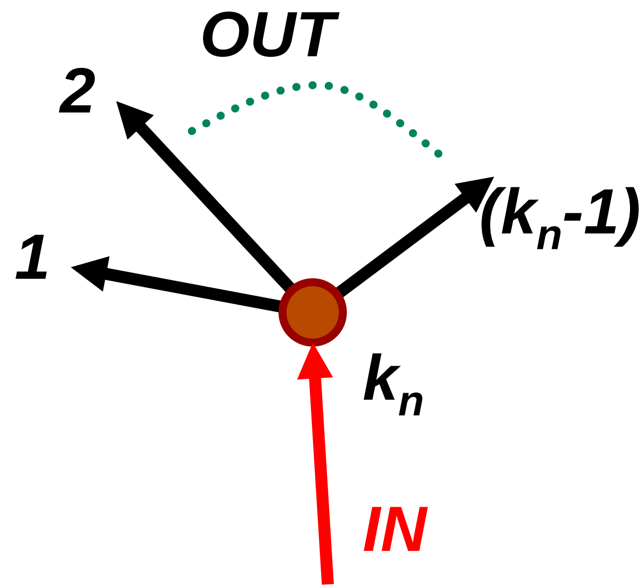
$$\equiv \langle k_n \rangle - \langle k \rangle = \frac{\sigma_k^2}{\langle k \rangle} \geq 0$$

Give a random friend that life saving vaccine

(if social networks are random and uncorrelated)

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 $(N \longrightarrow \infty)$



$\Rightarrow$ Walks always end if $\langle k_n \rangle < 2 \Leftrightarrow$ No GCC

$\Rightarrow$ Walks never end if $\langle k_n \rangle > 2 \Leftrightarrow$ GCC

(GCC= Giant Connected Component)

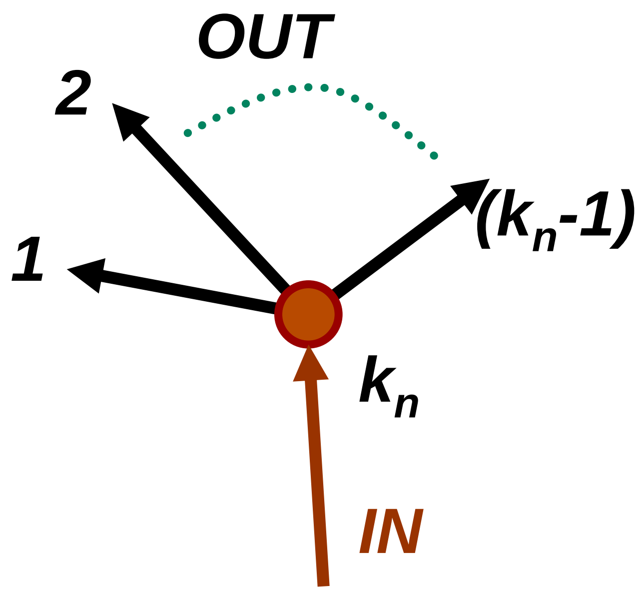
Length of Random Walks on Random Graphs

PROVIDED there are no loops.

True for sparse random graphs in limit of infinite size ***($N \longrightarrow \infty$)***

$\Rightarrow$ Walks always end if

$\langle k_n \rangle < 2 \Leftrightarrow$ No GCC



$\Rightarrow$ Walks never end if

$\langle k_n \rangle > 2 \Leftrightarrow$ GCC

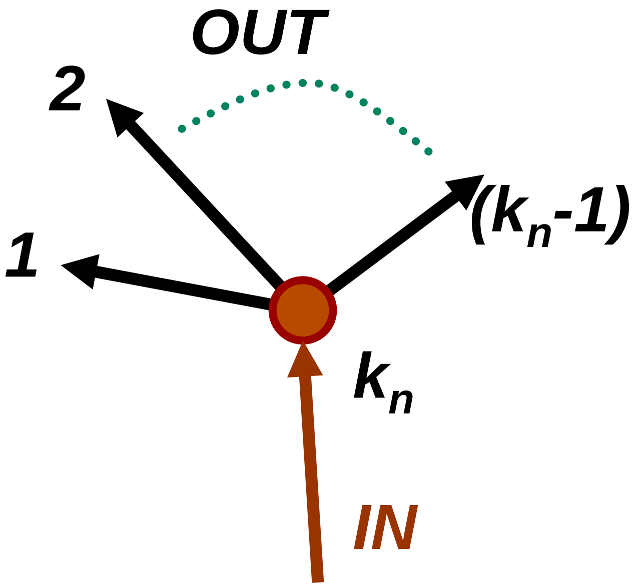
(GCC= Giant Connected Component)

GCC (Giant Connected Component) transition

GCC= Giant Connected Component, where a finite fraction of vertices in infinite graph are connected

GCC exists if ***z>1*** where

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



= Fractional measure of how much more popular your friends are

Other properties of General Random Graphs

All global properties depend on same

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

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

Average Path Length in MR Random Graph

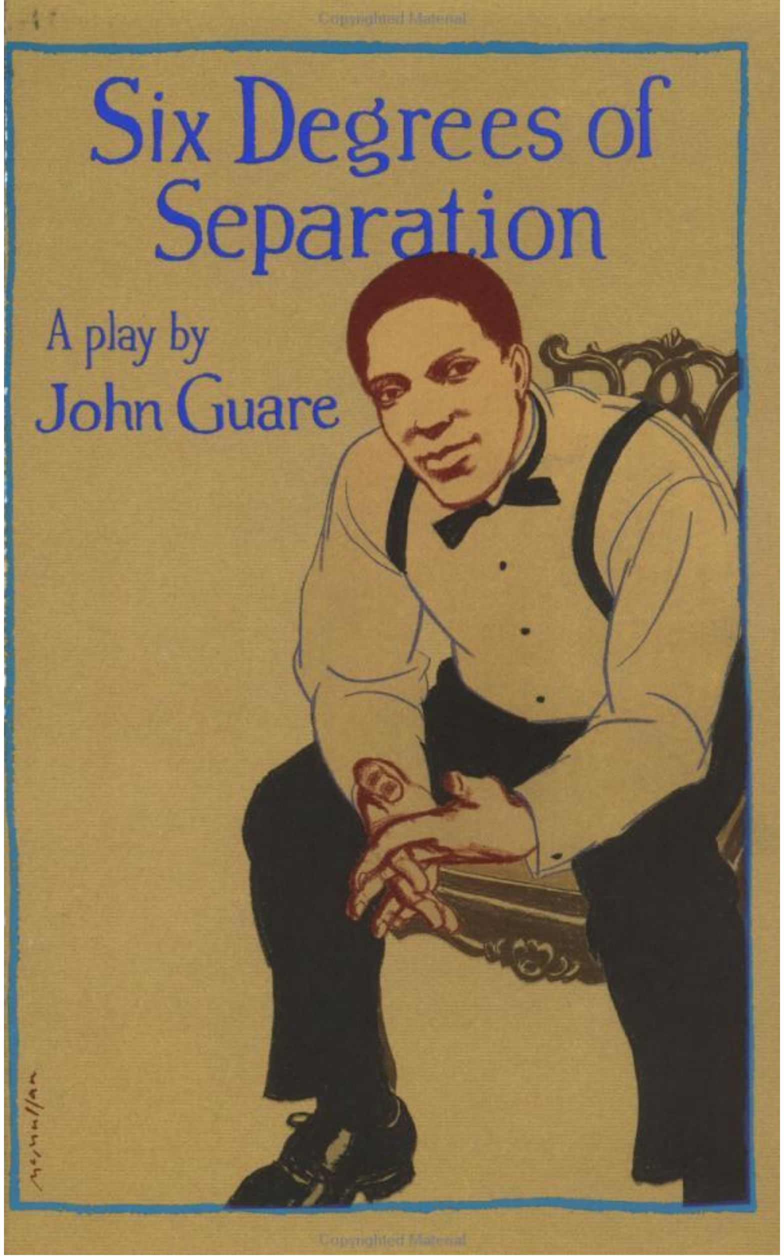
- For ***any*** random graph has an average shortest length which scales as

$$\langle d \rangle \approx \frac{\ln(N)}{\ln(z)} + c$$

Six Degrees of Separation

[John Guare 1990]

"I read somewhere that everybody on this planet is separated by only six other people. Six degrees of separation."

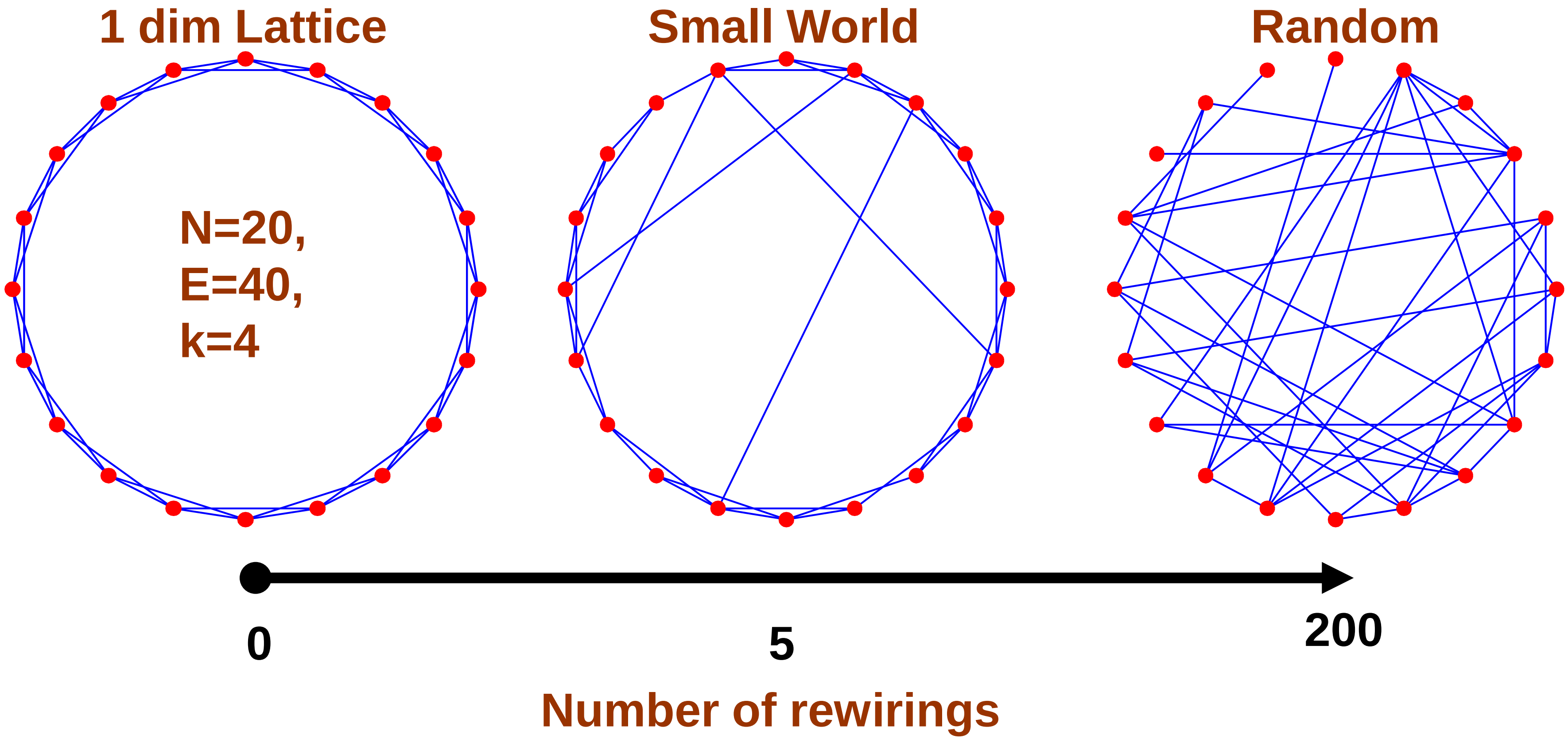


Small World

- A ***Small World*** network is one where the average shortest distance is $\langle d \rangle \sim O(\ln(N))$
- All random graphs are small world
- In fact most complex networks are small world
- c.f. a regular lattice in d-dimensions where the distance scales as $\langle d \rangle \sim O(N^{1/d})$

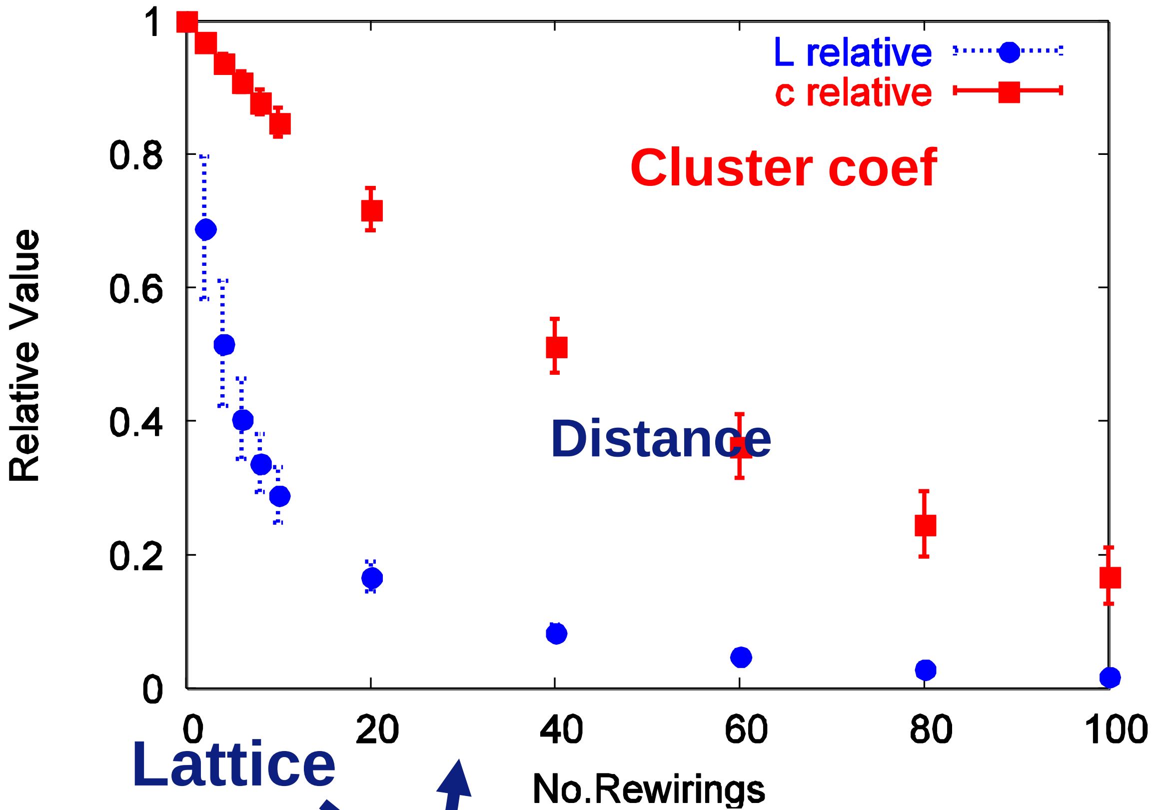
Watts and Strogatz's Small World Model (1998)

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



Clustering and Length Scale in WS network

- Average distance drops very quickly,
- Loss of local lattice structure much slower



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

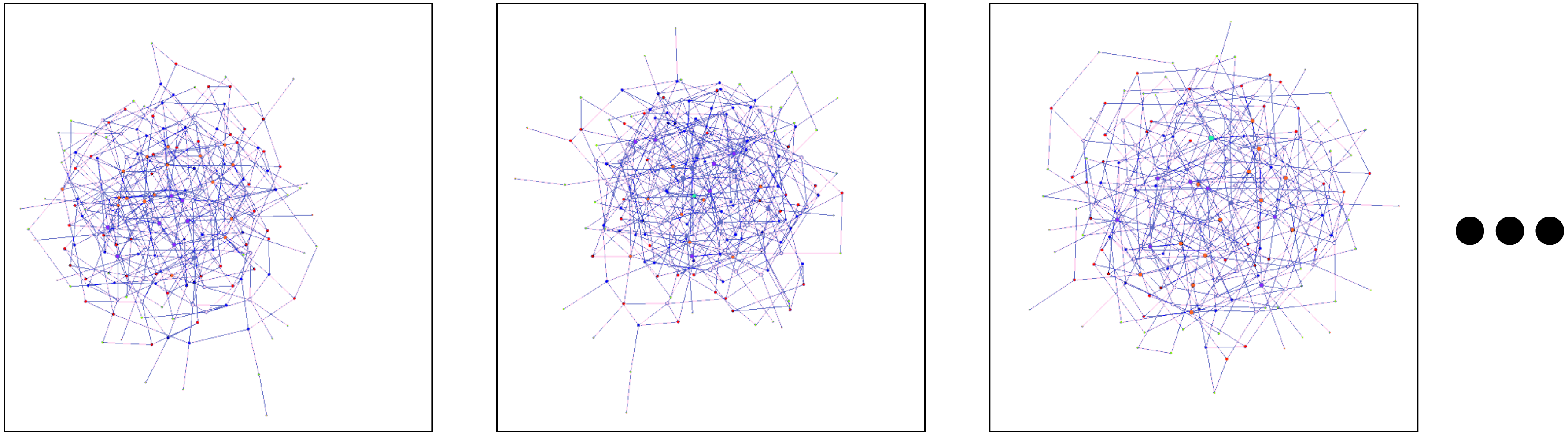
Lattice

Classical Random Graph

Small World

Ensembles of Graphs

Mathematically we ***do not*** consider a single instance of a random graph but an ***ensemble*** of random graphs

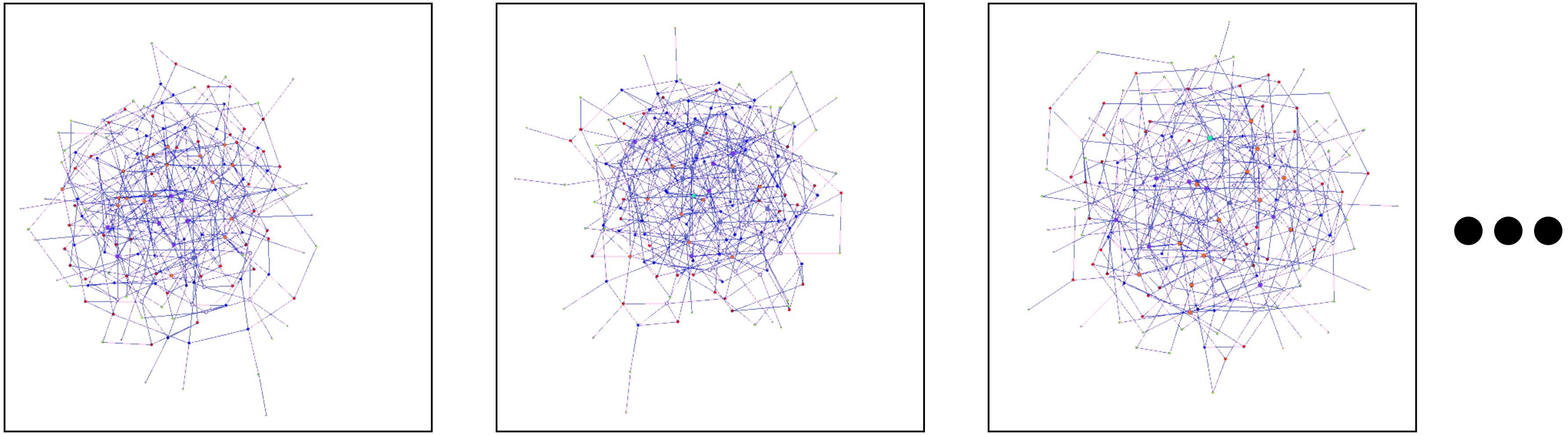


Ensembles of Graphs

e.g. The probability of creating a particular simple graph with ***E*** edges and ***Ē*** empty edges is

$$P(G) = p^{E}(1-p)^{\bar{E}}$$

Classical Random Graphs



Ensemble Averages

Averages of quantities are strictly over both

a) different graphs and

b) over some element of a graph e.g. vertices

$$\langle k \rangle = \sum_{G} P(G) \left(\frac{1}{N} \sum_{i \in V(G)} k_i \right)$$

Exponential Random Graphs (p* models)

General ensemble of graphs, those with highest probability obey any given constraints

$$\langle f \rangle = \sum_{G} P(G) f(G)$$

$$P(G) = \frac{1}{Z} e^{H(G)}$$

H(G) chosen so that graphs with preferred properties are most likely

Example Graph Hamiltonians ***H(G)***

- $H(G) = \beta E$

  Classical random graph with ***p=2E/(N(N-1))***

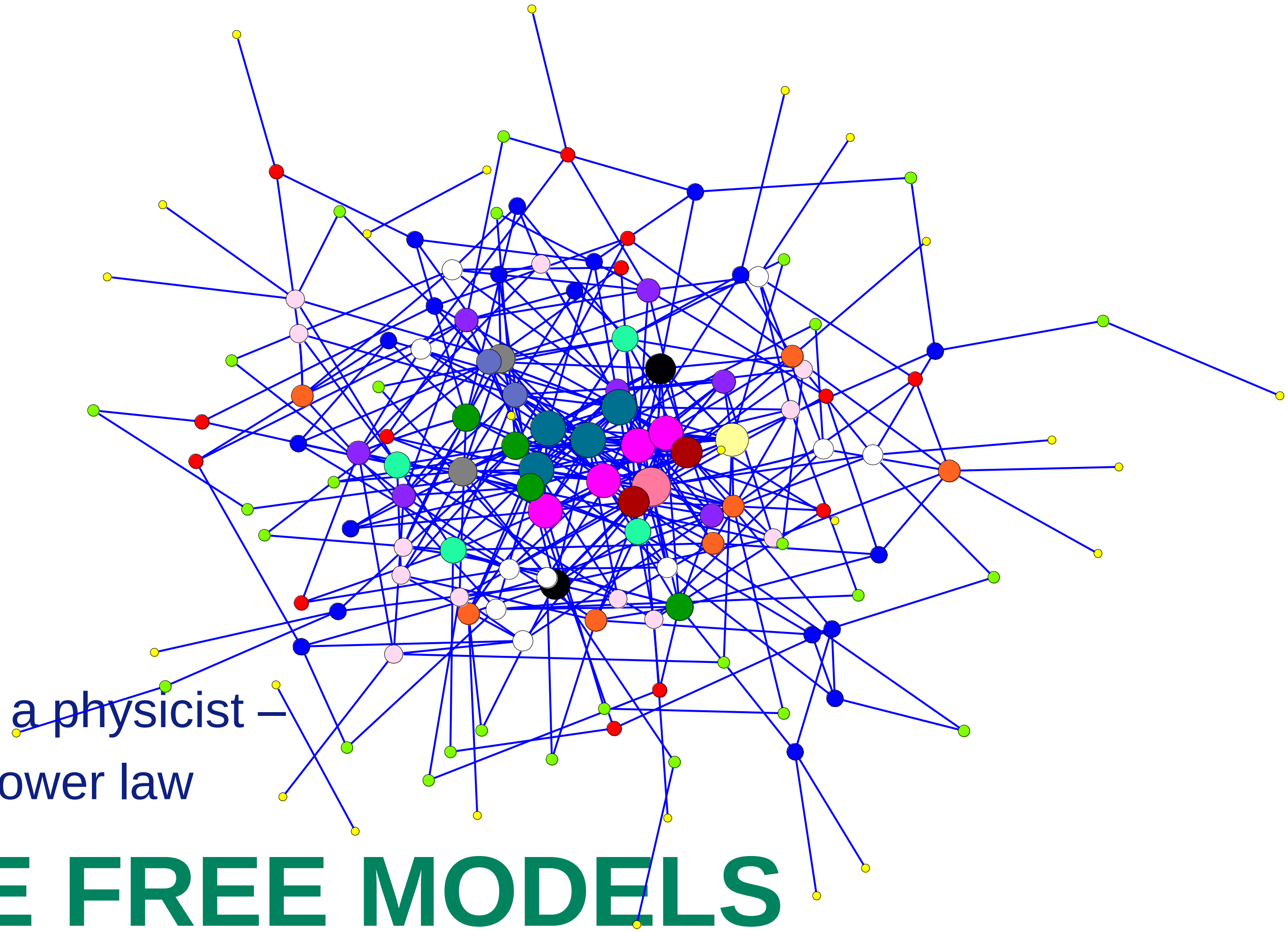
- $H(G) = \sum_{v \in V} \beta_v k_v$

  Random Graph with given degree distribution.

In both cases Lagrange multipliers β, β_v fixed by specifying desired values of $< E >$ and $< k_v >$

Summary of Random Graphs

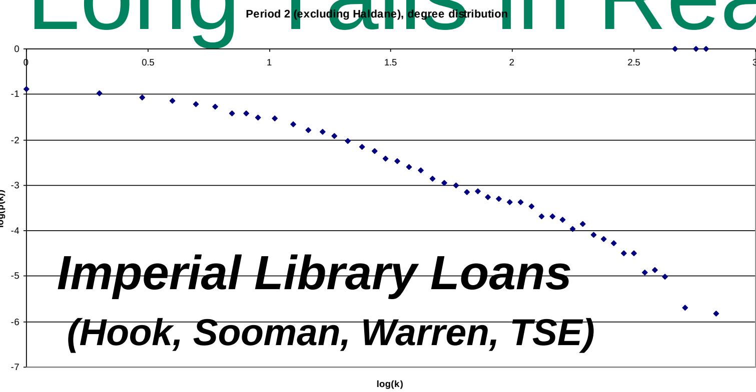
- Calculations work because
  - lack of correlations between vertices
  - few loops for large sparse graphs, graphs are basically trees
- Accessible analytically so can suggest typical behaviour even if very weak e.g. diameter vs N
- These can be reasonable approximations for many theoretical models
- Probably not for real world so then use these as a ***null model***.

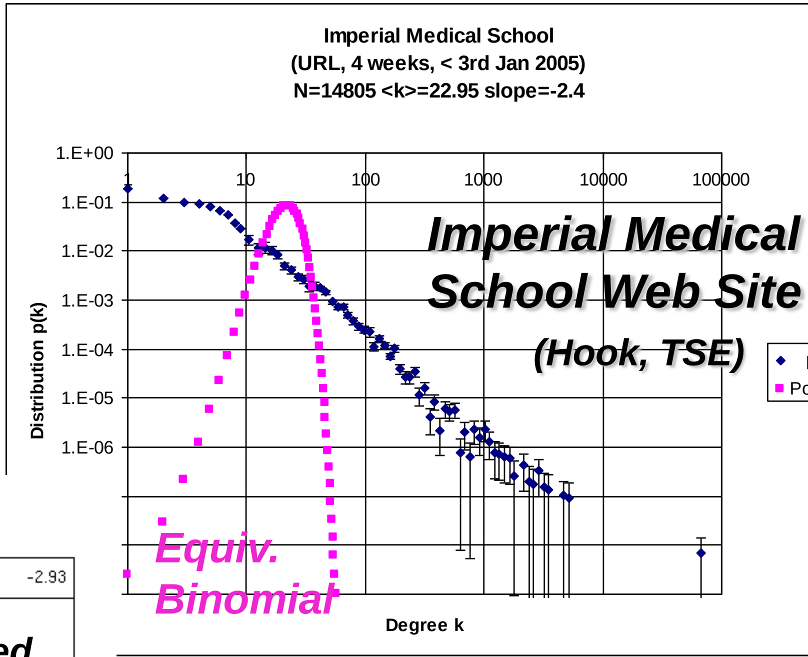


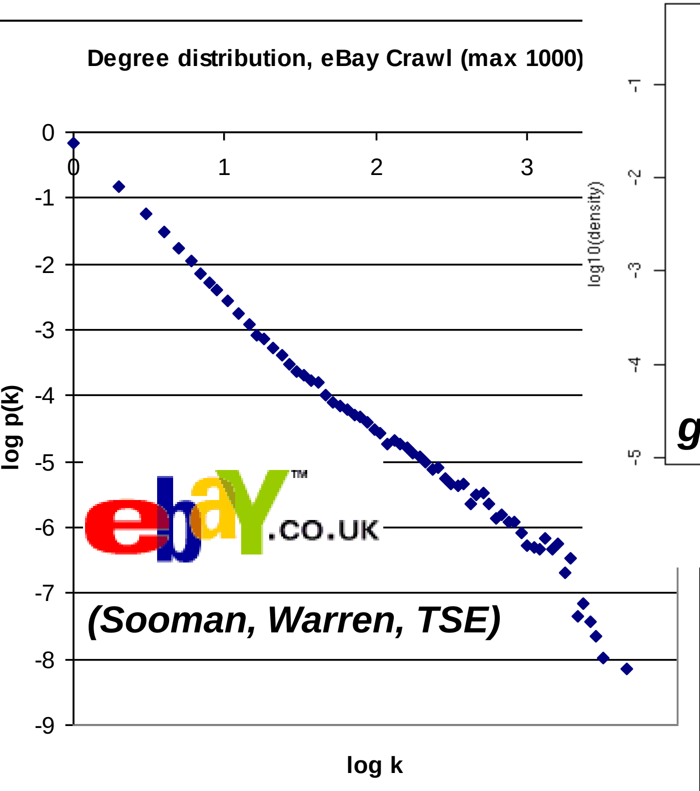
How to excite a physicist –
give them a power law

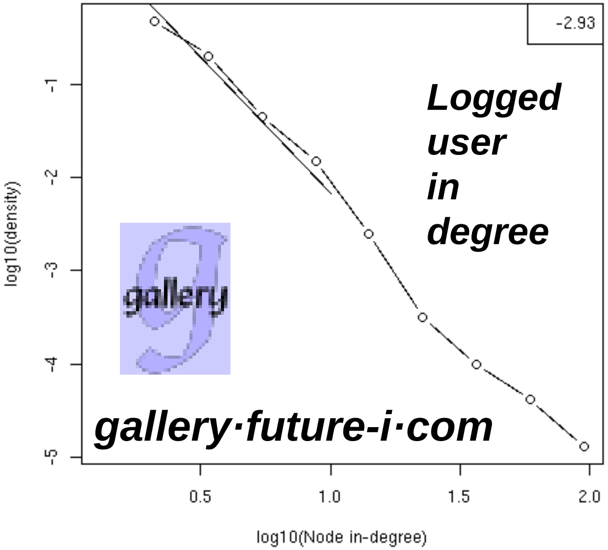
SCALE FREE MODELS

Long Tails in Real Data

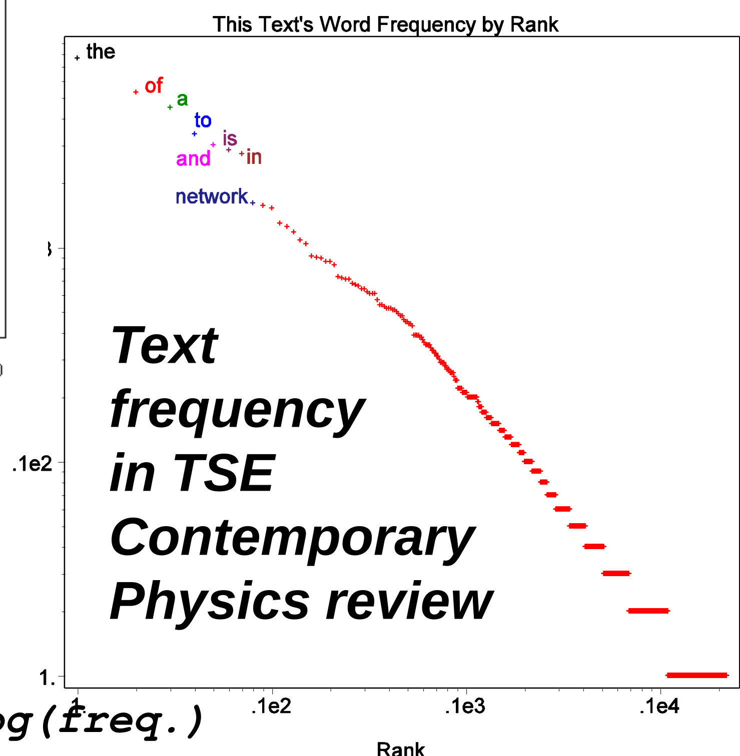








[Barabasi, Albert, 1999]



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

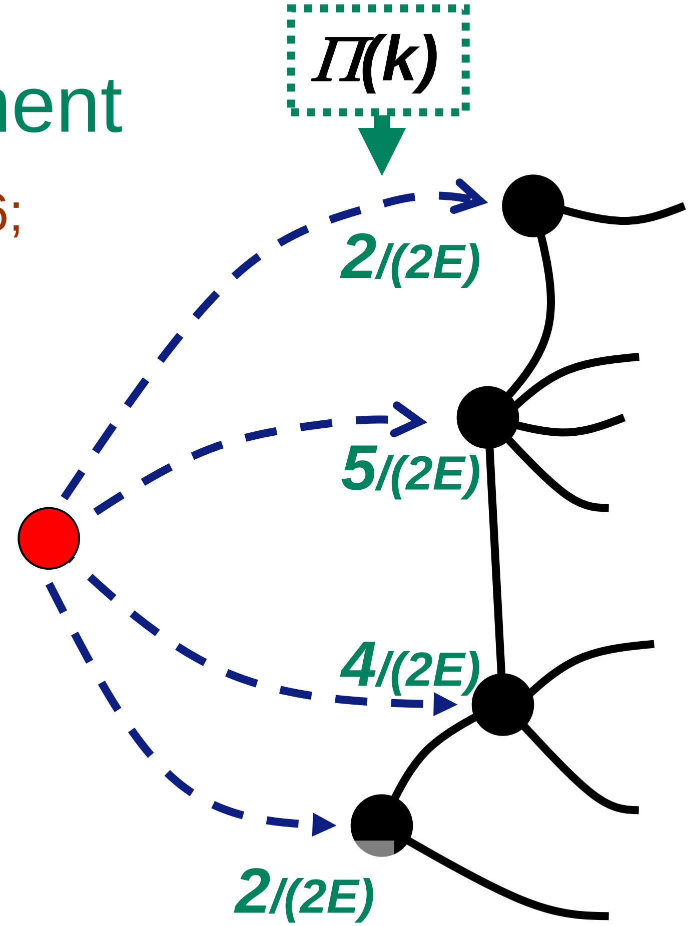
Growth with Preferential Attachment

[Yule 1925, 1944; Simon 1955; Price 1965,1976; Barabasi,Albert 1999]

1. Add new vertex attached to one end of ***m=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
$n(k) \sim k^{-\gamma}$
$\gamma = 3$

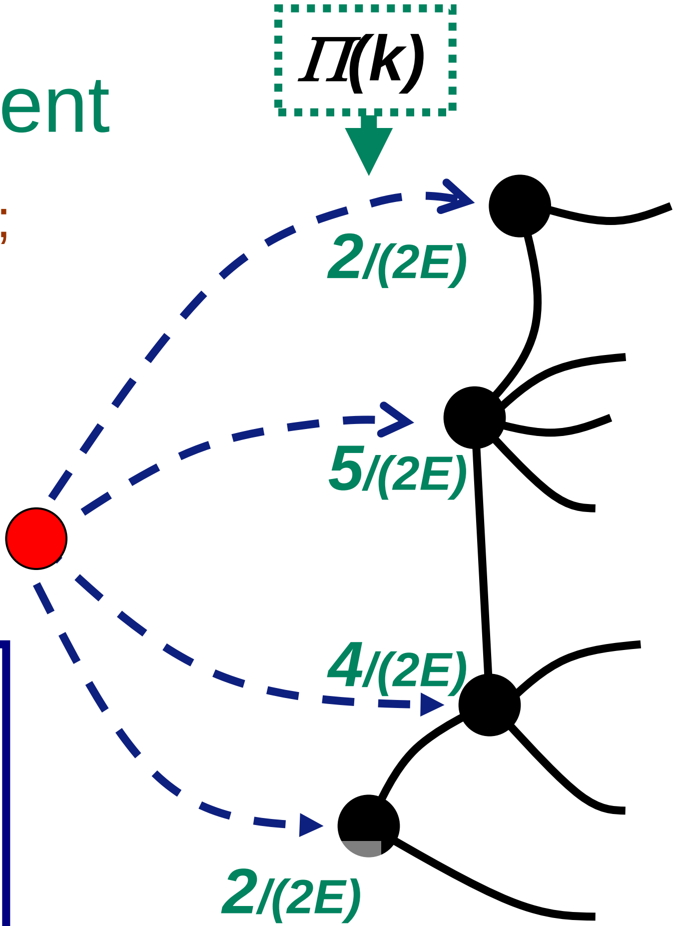
Growth with Preferential Attachment

[Yule 1925, 1944; Simon 1955; Price 1965,1976; Barabasi,Albert 1999]

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

Preferential Attachment

“Rich get Richer”



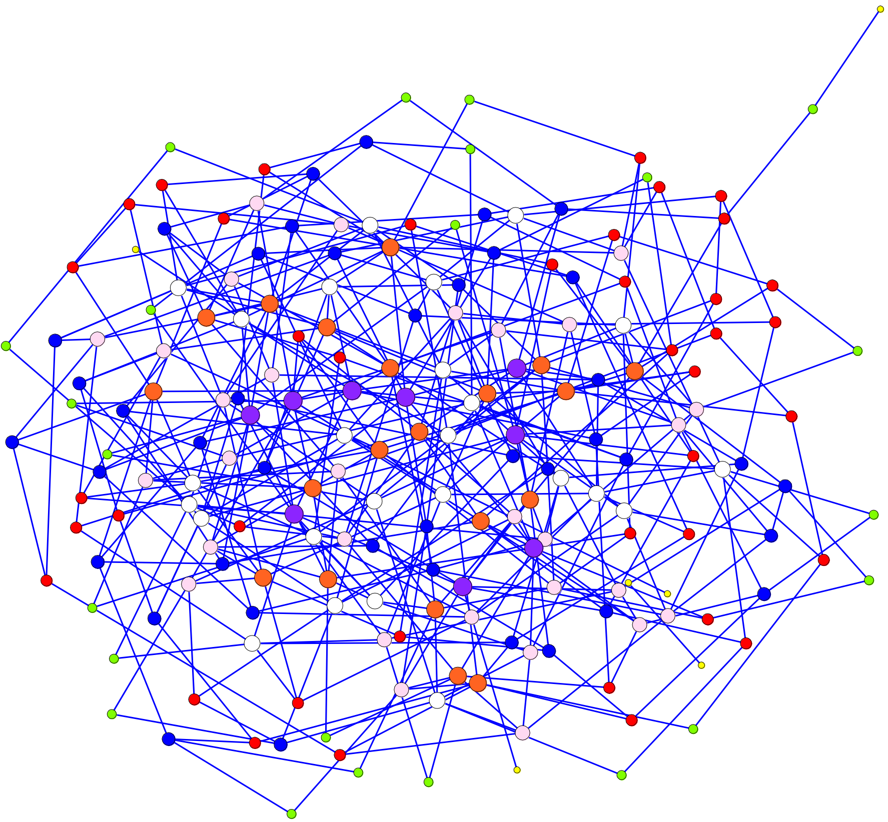
Result: Scale-Free Network

$$n(k) \sim k^{-\gamma}$$

$$\gamma = 3$$

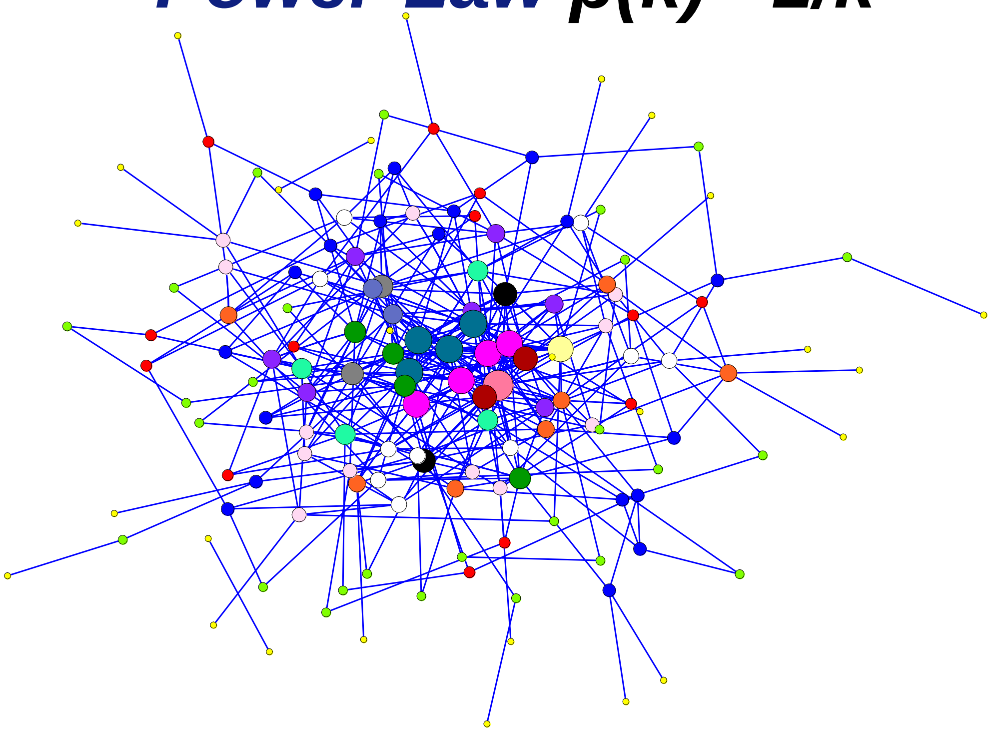
N=200, $\langle k \rangle$~4.0, vertex size ∝ k

Classical Random



Diffuse, small degree vertices $k_{max}=O(\ln(N))$

Scale-Free = Power-Law $p(k) \sim 1/k^3$



Tight core of large hubs $k_{max}=O(N^{1/2})$

Master Equation Approach

Let ***n(k,t)*** represent the *average* number of vertices at time ***t***. (I should really use **<*n(k,t)*>**)

Again average means we look at an ensemble of such networks.

The master equation the equation for evolution of the degree distribution averaged over different instances of network in the ensemble ***n(k,t)*** to ***n(k,t+1)***

Master Equation Processes

n(k,t) changes in one of three ways:-

- Increases as we add an edge to existing vertex of degree ***(k-1).***

  $k \rightarrow (k-1)$

- Decreases as we add an edge to existing vertex of degree ***k.***

  $(k-1) \rightarrow k$

- Number of vertices of degree ***k=m= ½<k>*** always increase by ***1*** as add new vertex.

  new vertex

Mean Field Degree Distribution Master Equation

$$n(k,t+1) - n(k,t) = +n(k-1,t)\, m\, \Pi(k-1)$$
$$- n(k,t)\, m\, \Pi(k)$$
$$+ \delta_{k,m}$$

k→(k-1)

(k-1)→k

new vertex

Π(k)* =** Probability of attaching to a vertex of degree ***k

$\propto$ ***k*** in simplest preferential attachment models

The Mean Field Approach is an Approximation

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

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

Ensembles over many instances *i* at *one time t*

Normalisation of probabilities not usually same for different *i*

If ***Π(k)*** is a function of degree ***k*** then normalisation of this probability is different in each instance of a network in the ensemble at a single time ***t***.

Ensemble Invariants

$$n(k,t+1)-n(k,t) = +n(k-1,t)\Pi(k-1) - n(k,t)\Pi(k) + \delta_{k,m}$$

Adding one vertex and ***m= ½<k>*** edges at each time means that the

- number of edges ***E(t) = mt + E(0)***
- number of vertices ***N(t) = t + N(0)***

are the same for all instances of network in the ensemble at any one time ***t***.

The Mean Field Approach Can Be Exact

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

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

Ensembles over many instances *i* at one time *t*

Normalisation of probabilities the same for different *i* if β=0 or 1

YES only if

$$\sum_k n_i(k)k^{\beta} = \left\langle \sum_k n_i(k)k^{\beta} \right\rangle$$

***β*=0 or *β*=1**

Exact Solution of Master Equation

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

Preferential Attachment ***Random Attachment***

- Note probability so $0 \le \Pi(k) \le 1$ **&** $p_p + p_r = 1$
- Lowest degree is $1 \le k_{\min} \le m = \langle k \rangle / 2$

- Thus $$0 \le p_p \le \frac{\langle k \rangle}{\langle k \rangle - k_{\min}} \le \mathbf{1}$$ **!**

Exact Solution of Master Equation

- Look for asymptotic solutions

$$n(k,t) = N(t)\,p(k)$$

- Find for ***k* > *m* = ½<*k*>**

$$\frac{p(k)}{p(k-1)} = \frac{N.\Pi(k-1)}{1+N.\Pi(k)} = \frac{(1/2)\,p_p(k-1)+p_r m}{1+(1/2)\,p_p k + p_r m}$$

Exact Solution of Master Equation

Hence $$p(k) = A\frac{\Gamma(k+a)}{\Gamma(k+1+a+b)}$$

where $$a = \frac{p_r\langle k\rangle}{p_p}, b = \frac{2}{p_p},$$

Large k limit:-

$$\lim_{k\to\infty} p(k) = \frac{A}{k^\gamma}$$ **&** $$\gamma = 1 + \frac{2}{p_p} \geq 2$$

Scale-Free Growing Model comments

- Illustrates use of master equations and their approximations ⇒ statistical physics experience
- Exact solutions for ensemble average asymptotic value of degree distribution ***p(k)*** if

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

- Interpretation of parameters – $\boldsymbol{p_p > 1}$ allowed
- Finite Size effects? – real networks are mesoscopic

[TSE, Saramäki 2004]

- Fluctuations in ensemble?
- Network not essential – ***k***=frequency of previous choices
- Growth not essential – network rewiring ⇒ [illegible] ~ ***1.0***

[Moran model, see TSE,Plato, 2008]

Scale-Free in the Real World

Attachment probability used was

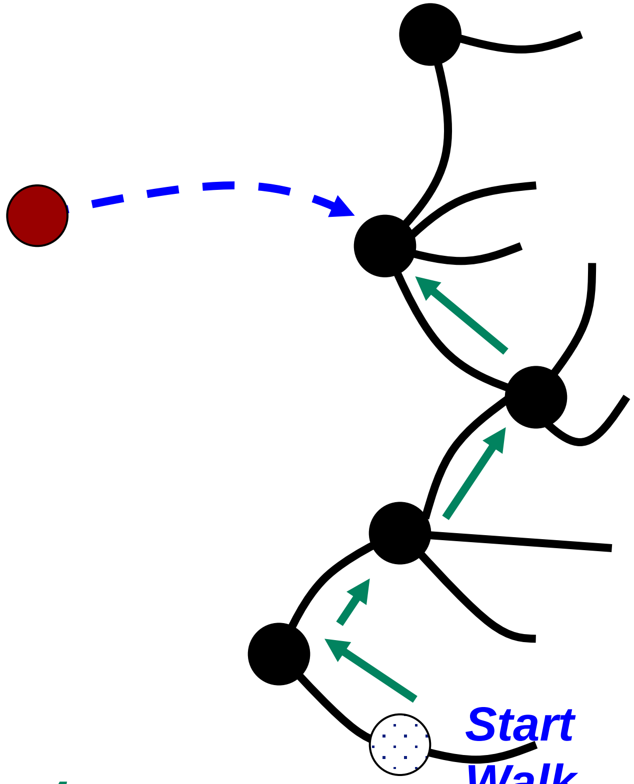
$$\Pi(k) = p_p \frac{k}{2E} + p_r \frac{1}{N},$$

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!*

Preferential Attachment for Real Networks

[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*** $\Rightarrow \gamma = 3$

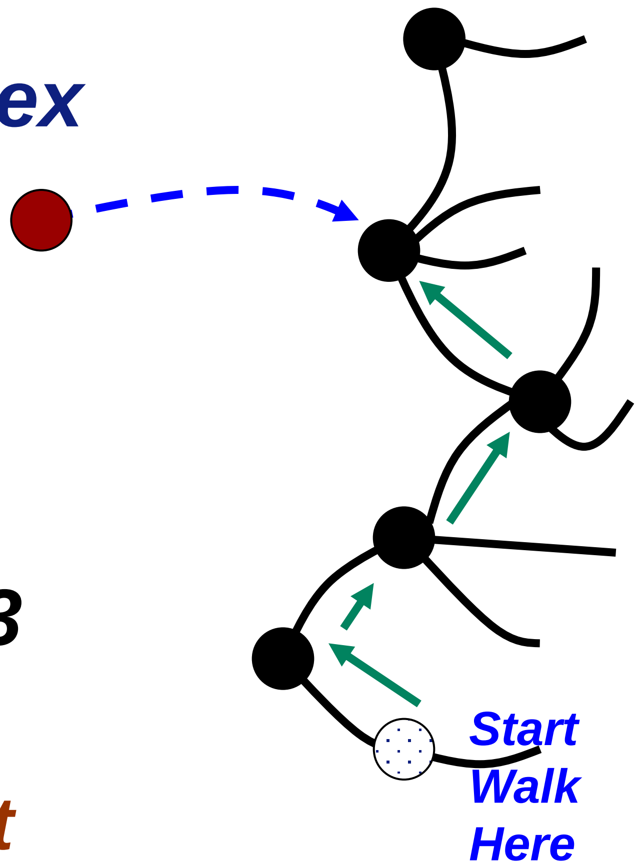
(Can also mix in random attachment with probability p_r)

Preferential Attachment for Real Networks

➔ ***Probability of arriving at a vertex***
$\propto$ number of ways of arriving at vertex
= k, the degree

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

Can also mix in random attachment with probability p_r



Naturalness of the Random Walk algorithm

- Gives preferential attachment from any network and hence 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

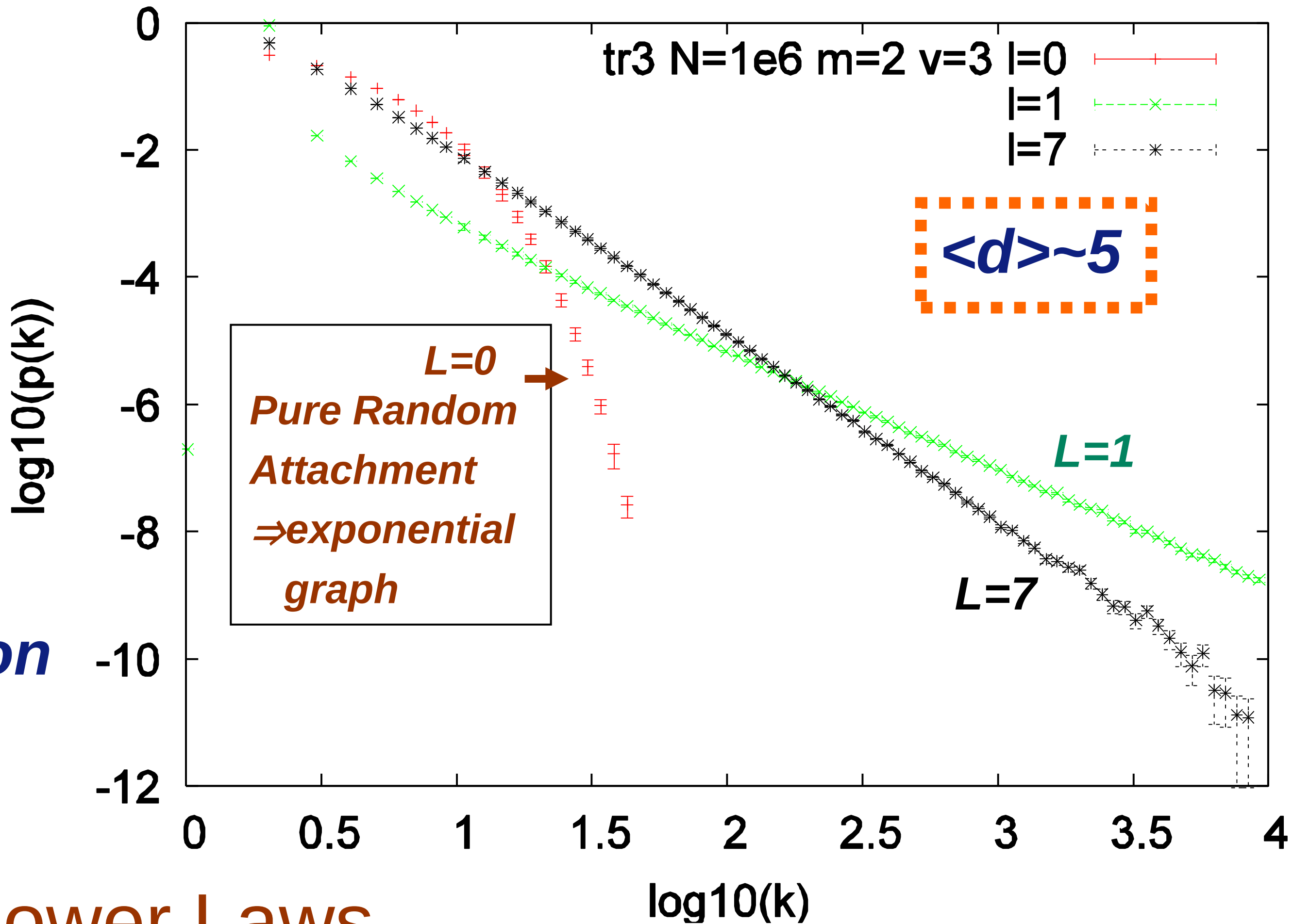
    e.g. informal requests for work on the film actor's social network

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

Is the Walk Algorithm Robust?

I varied:

- ***Length of walks***
- <k>
- ***Starting point of walks***
- ***Length distribution of walks***
- ***…..***



YES - Good Power Laws

but NOT Universal values - 10% or 20% variation

Finite Size Effects

Networks are ***mesoscopic*** systems

In practice a network of ***N ~ 1*** million is still ***not large*** since many quatitities scale with the logarithm of system size
e.g. Diameter scales as ***log(N) ~ 6***.

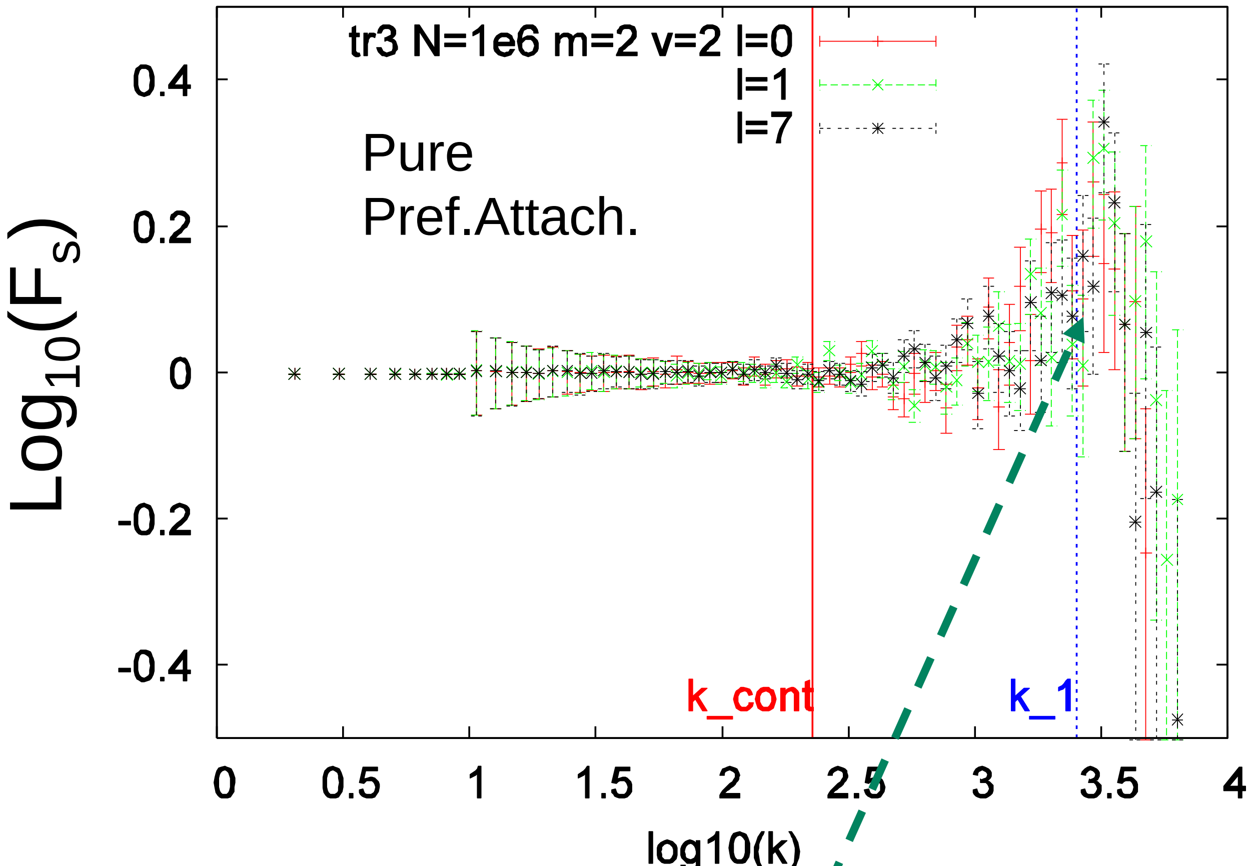
Finite Size Effects for pure preferential attachment

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

Scaling Function

$$F_S(x) \approx 1$$

$$if \quad x < 1$$

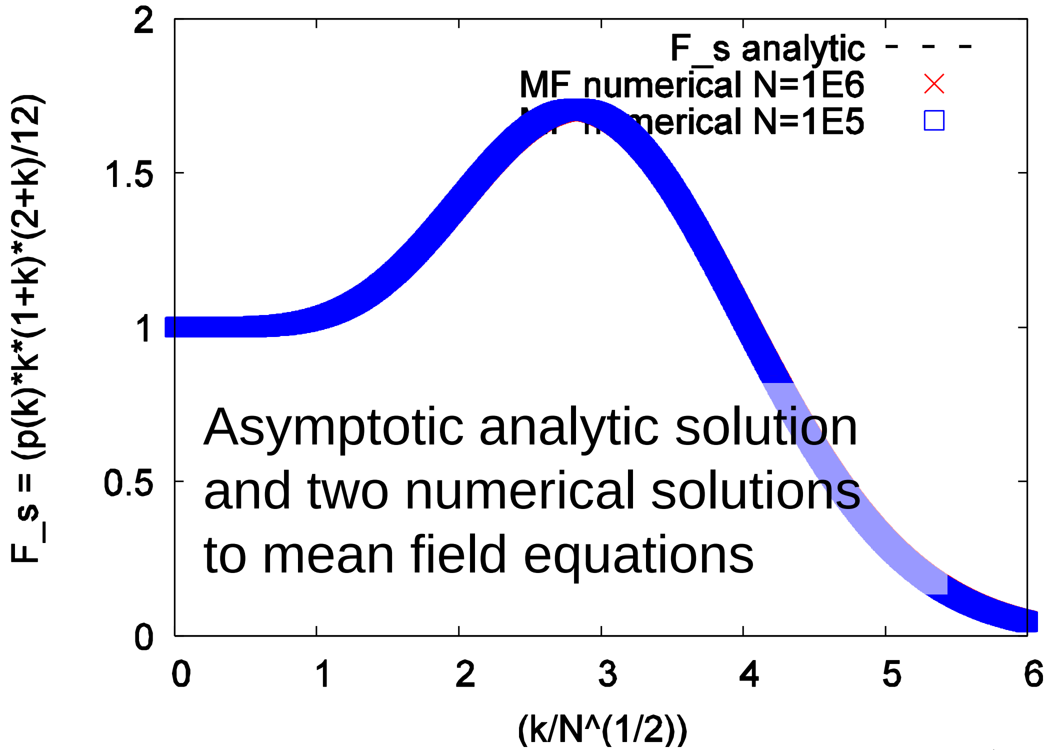


100 runs to get enough data near k_1

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



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

2m=<k>

Hermite Polynomials

(TSE+Saramäki, 2005;

generalisation of Krapivsky and Redner, 2002)

What can physicsts and mathematicians do well?

RANDOM WALKS

Properties of irreducible non-negative matrices (1)

Will phrase this in terms of

Adjacency Matrix A_{ij} for a network

- $A_{ij} = A_{ji} = 1$ for edges in ***simple graphs***

- A_{ij} is the weight of edge from j to i for ***weighted network***

- $A_{ij} \neq A_{ji}$ (symmetric matrix) if ***directed network***

Definition of irreducible non-negative matrices (1)

In terms of an Adjacency Matrix $\boldsymbol{A_{ij}}$ for a network

- A ***non-negative*** matrix is $\boldsymbol{A_{ij} \geq 0}$

- ***Irreducible*** if there is a path from each vertex to every other vertex

$$\forall\, i, j \quad \exists\, n > 0 \quad s.t.\ (A^n)_{ij} > 0$$

Properties of irreducible non-negative matrices (2)

- Largest eigenvalue (λ_1) is real and positive
- Largest eigenvalue is bounded by largest and smallest sums of each row and each column
- Eigenvector of largest eigenvalue has only positive entries
- Entries in all other eigenvectors differ in sign

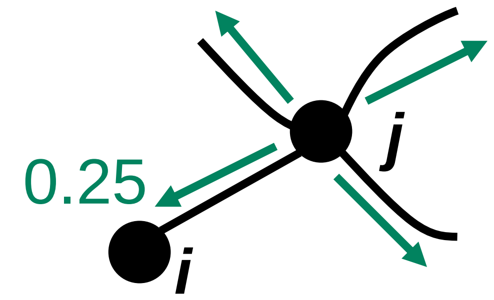
Random Walk Transition Matrix

The transition matrix for a simple unbiased random walk on a network is ***T*** where the probability of moving *from* vertex ***j*** *to* vertex ***i*** is

i ← j

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

with *strength* $k_j = \sum_i A_{ij}$



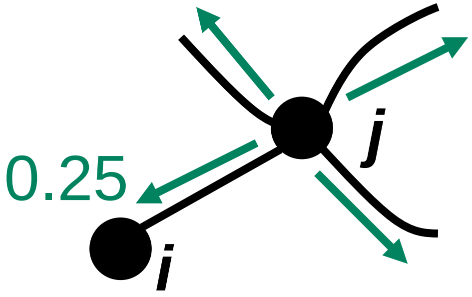
Probability of following an edge ***from j to*** any vertex ***i*** is 0.25

Random Walk Transition Matrix (2)

Another useful form is

$$T = A \bullet D^{-1}$$

$$D_{ij} = \delta_{ij} k_j$$



Probability of following an edge ***from j to*** any vertex ***i*** is 0.25

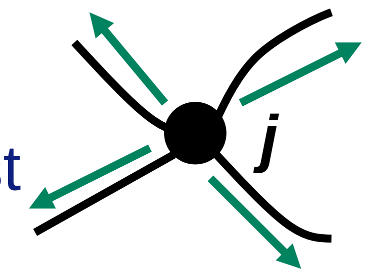
Transition matrix properties (1)

- Adjacency matrices of networks are non-negative (almost always)
- Irreducible if network fully connected

  (or add some weak links to make it so)
- Transition matrix is also non-negative and irreducible

$$\sum_i T_{ij} = \sum_i A_{ij} \Big/ k_j = \frac{k_j}{k_j} = 1$$

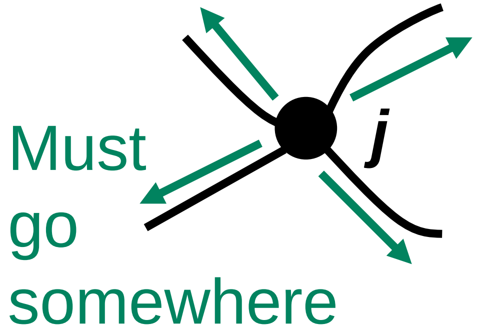
Must go somewhere



Transition matrix properties (2)

- Transition matrix columns always sum to 1

$$\sum_i T_{ij} = \sum_i A_{ij} \Big/ k_j = \frac{k_j}{k_j} = 1$$



$\Rightarrow$ Transition matrix has unique largest eigenvalue equal to $\mathbf{1=\lambda_1}$

Transition matrix properties (3)

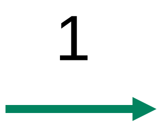
- Eigenvector of largest eigenvalue of transition matrix, $\boldsymbol{v_1}$, of ***undirected*** network is just $\boldsymbol{k_i}$.

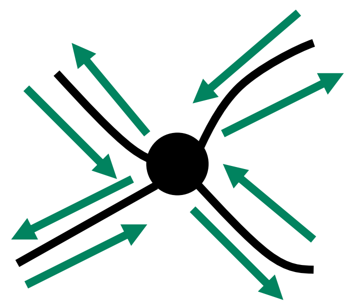
$$\left(T\vec{v}_1\right)_i = \sum_j \frac{A_{ij}}{k_j} k_j = \sum_j A_{ij} = k_j = \left(\vec{v}_1\right)_i$$

i.e. ***Flow in = Flow Out*** is equilibrium reached if flow along each edge is equal to the weight of the edge

Transition matrix properties (4)

- ***Flow in = Flow Out*** is equilibrium reached if flow along each edge is equal to the weight of the edge





For simple graph,
ONE walker passes
along each edge
in each direction at
each time step

$\Rightarrow$***k*** walkers arrive
and leave each vertex

*** NOT solution if number of in- and out-edges different

Random walk as linear algebra

Let ***$w_i(t)$*** be the number of random walkers at vertex ***i*** at time ***t***

(or the probability of finding one walker at ***i***)

then the evolution is simply

$$\vec{w}(t+1) = T.\vec{w}(t)$$

$$w_i(t+1) = \sum_j T_{ij} w_j(t)$$

Random walk as linear algebra

Decompose $\boldsymbol{w_i(t)}$ in terms of eigenvectors $\boldsymbol{v_n}$ as

$$\vec{w}(t=0) = \sum_n c_n \vec{v}_n$$

then the evolution is simply

$$\vec{w}(t) = \sum_n c_n (\lambda_n)^t \vec{v}_n$$

Equilibrium

Equilibrium reached is eigenvector with largest eigenvalue as $1 = \lambda_1 > |\lambda_n| \quad \forall\, n > 1$

$$\vec{w}(t \to \infty) \propto \vec{v}_1$$

So for simple networks we have

$$w(t \to \infty)_i \propto k_i$$

PageRank

- Google ranks web pages using $\boldsymbol{v}_1$
- Follows links between web pages like a random walker
- Google makes money because the web is a *directed* graph so largest eigenvector, $\boldsymbol{v}_1$, is not trivial

PageRank for Mathematicians

Using MacTutor bibliography of over 200 mathematicians finds [Clarke, TSE, Hopkins, 2010]

Rank	Degree	Closeness	Betweenness	Page Rank
1st	Newton	Newton	Euclid	Euclid
2nd	Hilbert	Hilbert	Newton	Newton
3rd	Euclid	Riemann	Euler	Laplace
4th	Riemann	Euler	Riemann	Hilbert
5th	Euler	Euclid	Van der Waerden	Lagrange

Centrality Measures

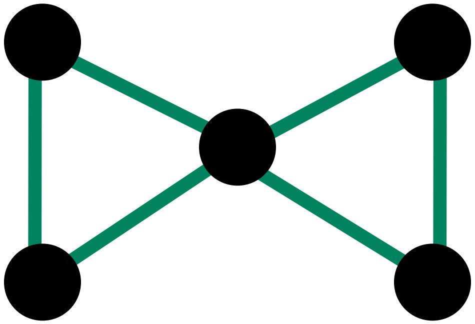
The closer a vertex is to the *"centre"* of a network, the higher its ***Centrality Measures***:-

- Degree
- PageRank
- **Betweenness**
  Simple Betweenness = number of shortest paths passing through each vertex
- etc.

Simple Betweenness

1. Calculate the shortest paths between all pairs of vertices.
2. **Betweenness** = number of shortest paths passing through each vertex

Example



BUT ONLY defined for simple graphs

Electric Current Betweenness

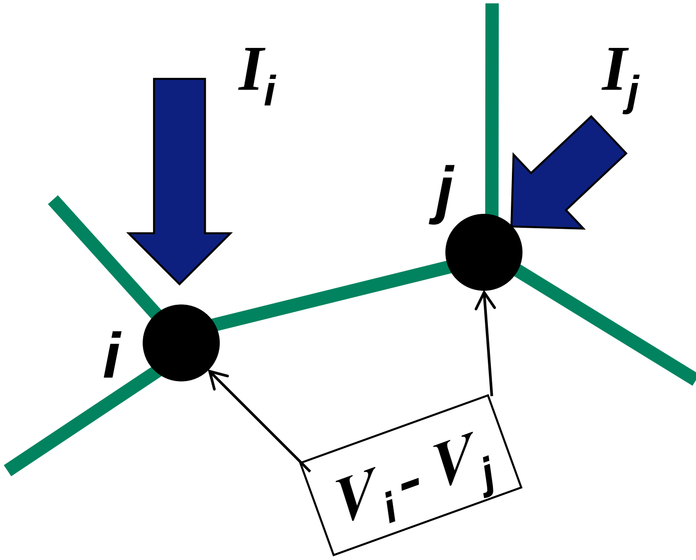
[Newman 2005]

$$I = \frac{V}{R}$$ Ohm's Law

Treat *undirected* network as resistors, with

- Conductivity of resistor edge weights = $A_{ij} = A_{ij}$
- Voltage at vertex i = V_i
- External current flowing into vertex i = I_i

$$\sum_j A_{ij}(V_i - V_j) = I_i$$



Betweenness and Currents (2)

$$\sum_j A_{ij}(V_i - V_j) = I_i$$

$$(D - A)\vec{V} = I$$

where $\boldsymbol{D_{ij} = k_i \delta_{ij}}$ a diagonal matrix using degree

$$\Rightarrow \vec{V} = (D - A)^{-1} \vec{I}$$

Wait till later to see why inverse is OK

Betweenness and Currents (3)

$$\vec{V} = (D - A)^{-1}\vec{I}$$

$$= D^{-1}(1 - AD^{-1})^{-1}\vec{I}$$

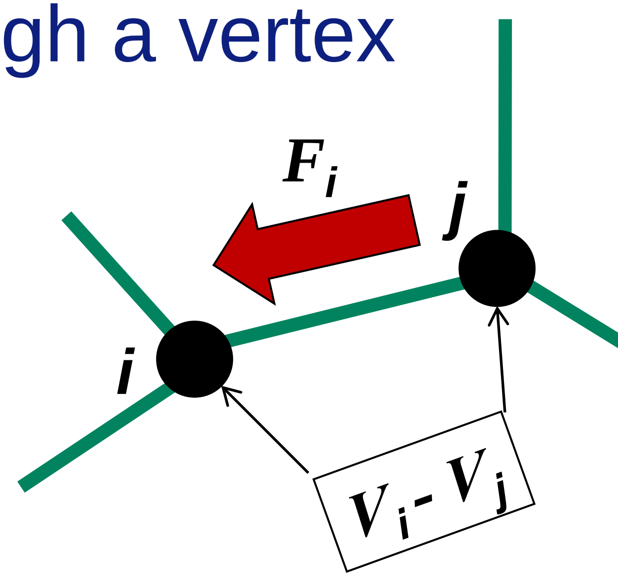
$$= D^{-1}(1 - T)^{-1}\vec{I}$$

where $\boldsymbol{T = AD^{-1}}$ is the random walker transition matrix

Betweenness and Currents (4)

Define the ***net flow*** of current through a vertex to be $\boldsymbol{F_i}$ so

$$F_i = \frac{1}{2}\sum_j \left| A_{ij}(V_i - V_j) \right|$$



then using $\vec{V} = D^{-1}(1-T)^{-1}\vec{I}$ we find that

$$F_i = \frac{1}{2}\sum_j \left| T_{ij}((1-T)^{-1}I)_j - T_{ji}((1-T)^{-1}I)_i \right|$$

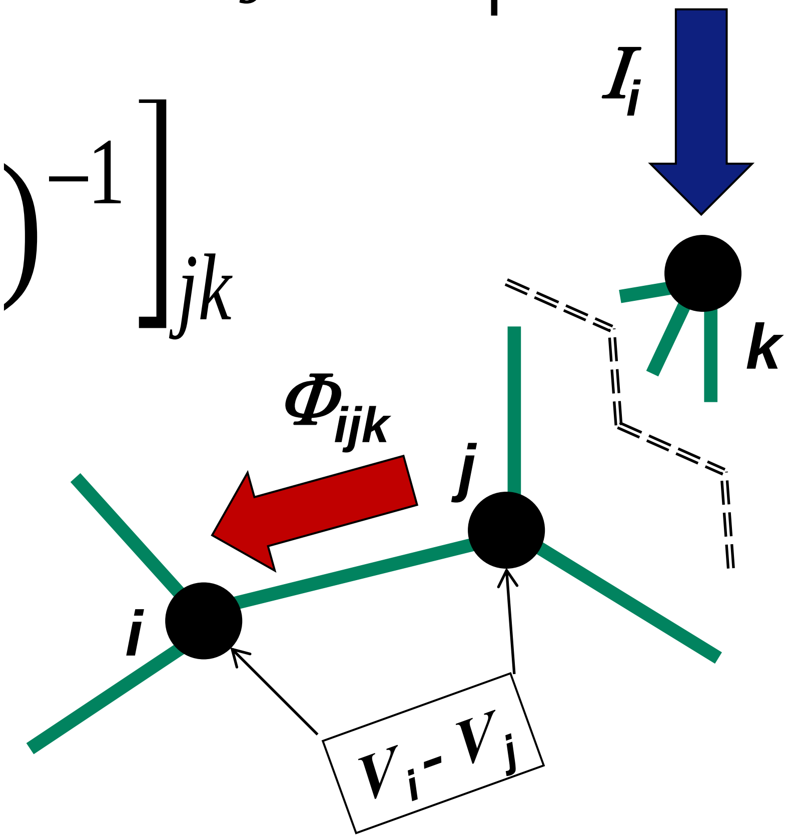
Betweenness and Currents (5)

So net flow of current F_i through vertex i is

$$F_i = \frac{1}{2}\sum_{j,k}\left|\Phi_{ijk} I_k - \Phi_{jik} I_k\right|$$

where $$\Phi_{ijk} = \sum_k T_{ij}\left[(1-T)^{-1}\right]_{jk}$$

is the current flowing from j to i due to external current put in at k



Betweenness and Currents (6)

However in terms of random walkers

$$(1 - pT)^{-1} = \sum_n (pT)^n \qquad \text{if } |p\lambda_1| < 1$$

So $\boldsymbol{\Phi_{ijk}}$ counts the number of random walkers *starting* at ***k***, *arriving* at ***j*** after ***n*** steps, followed by a move ***to i***

$$\Phi_{ijk} = \sum_n T_{ij} \left[T^n \right]_{jk}$$

Betweenness and Currents (7)

The total current put into the circuit must match the current taken out

$$\sum_i I_i = 0$$

In terms of the transition matrix, this means this vector $\boldsymbol{I_i}$ does ***not*** contain the equilibrium eigenvector with eigenvalue ***1***

Thus $$\Phi_{ijk} = \sum_n T_{ij} \left[T^n \right]_{jk}$$ is well defined

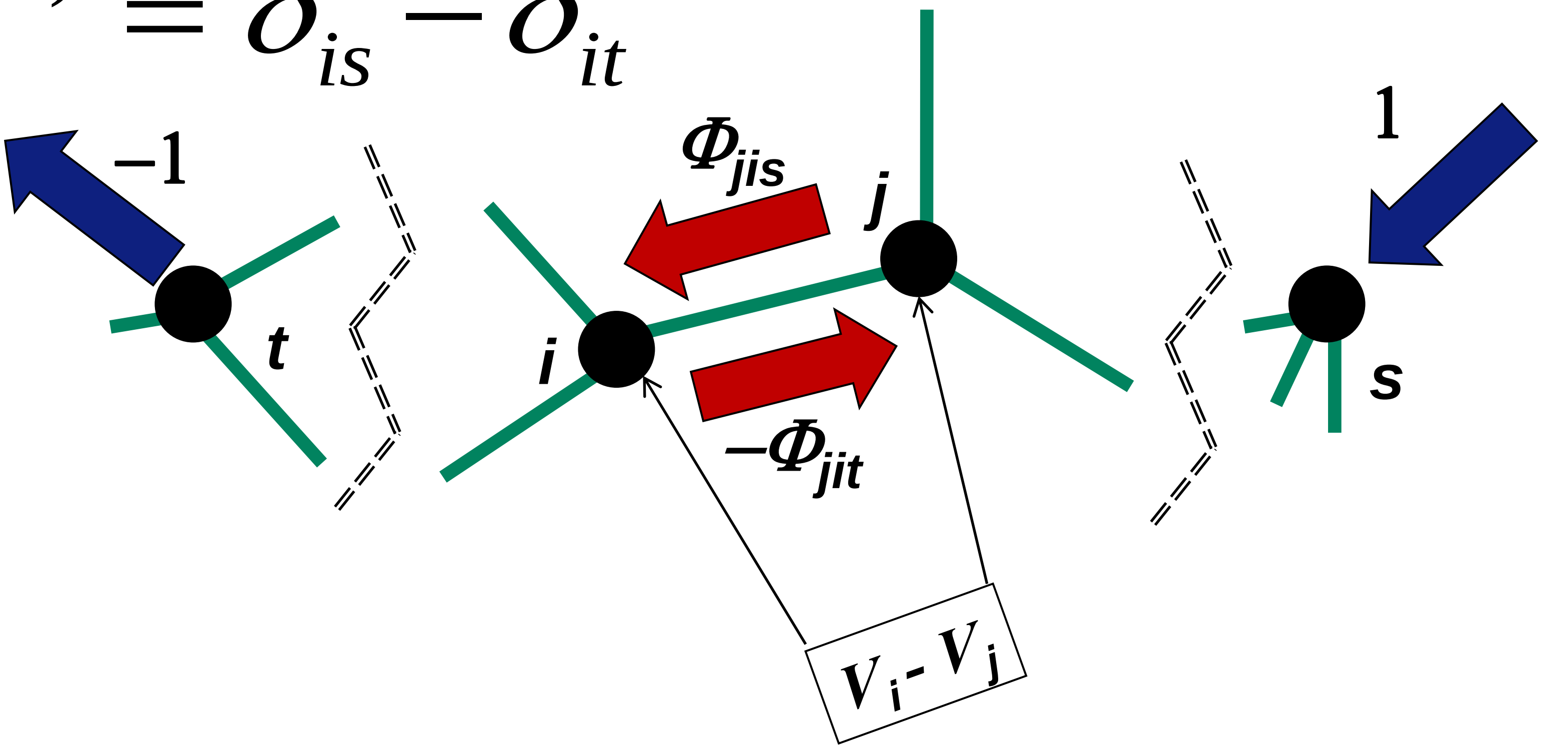
if ***p<1*** or if acts on $\boldsymbol{I}$.

Betweenness and Currents (8)

Suppose

- we put one unit of current in at ***source*** vertex ***s***
- we take one unit of current out at ***target*** vertex ***t***

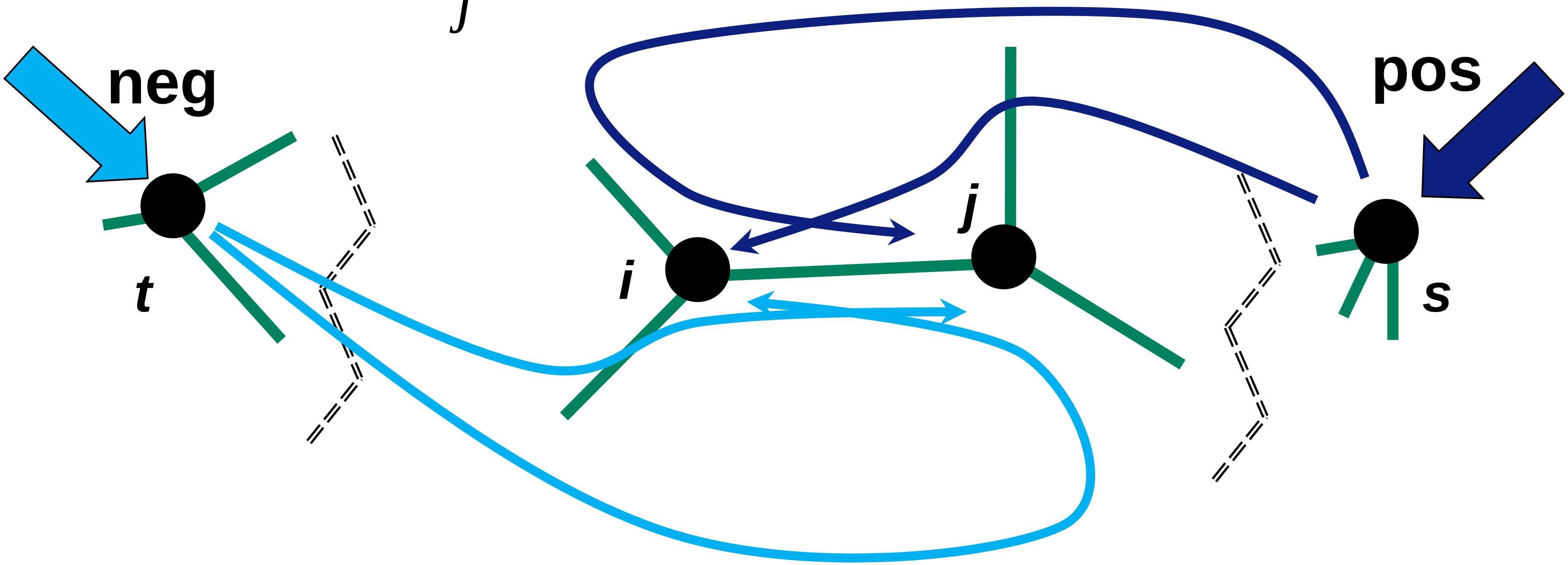
$$I_i^{(st)} = \delta_{is} - \delta_{it}$$



Betweenness and Currents (9)

The net flow in terms of positive (from ***s***) and negative (from ***t***) random walkers is

$$F_i^{(st)} = \frac{1}{2}\sum_j \left| \Phi_{ijs} - \Phi_{ijt} - \Phi_{jis} + \Phi_{jit} \right|$$



Betweenness and Currents (10)

Newman suggests a centrality measure of summing over all possible source and sink currents

$$F_i = \sum_{s,t} F_i^{(st)}$$

with $F_i^{(st)} = \frac{1}{2}\sum_j \left| \Phi_{ijs} - \Phi_{ijt} - \Phi_{jis} + \Phi_{jit} \right|$

and $\boldsymbol{\Phi_{ij}}$ is the number of random walkers starting from ***k*** passing from ***j*** to ***i*** after ***n*** steps

Betweenness and Currents Summary

- Uses negative random walkers
- Random walker picture works for directed graphs
- Does not use equilibrium eigenvector
  - unlike PageRank, Modularity for community detection, ...
- Can introduce distance scale ***d= p/(1-p)***
  - Walkers move on with probability ***p***, stop with probability ***(1-p)***. Replace $T \rightarrow pT$ and $(1-T)^{-1} \rightarrow (1-p)/(1-pT)$
- Can introduce biased random walks
  - e.g. $T_{ij} = \frac{\alpha_i A_{ij}}{z_j}, \quad z_j = \sum_i \alpha_i A_{ij}$ [Lambiotte et al. 2011]

Betweenness and Currents Summary

Newman suggests a centrality measure of summing over all possible source and sink currents

$$F_i = \sum_{s,t} F_i^{(st)}$$

with $F_i^{(st)} = \frac{1}{2}\sum_j \left| \Phi_{ijs} - \Phi_{ijt} - \Phi_{jis} + \Phi_{jit} \right|$

and $\boldsymbol{\Phi_{ij}}$ is the number of random walkers starting from ***k*** passing from ***j*** to ***i*** after ***n*** steps

THANKS

Bibliography: Free General Network Reviews

- Evans, T.S. “Complex Networks”, *Contemporary Physics,* **2004***, 45*, 455-474 [arXiv.org/cond-mat/0405123]
- Easley, D. & Kleinberg, J. Networks, “Crowds, and Markets: Reasoning About a Highly Connected World Cambridge University Press, 2010” [http://www.cs.cornell.edu/home/kleinber/networks-book/]
- van Steen, M. “Graph Theory and Complex Networks” Maarten van Steen, 2010 [http://www.distributed-systems.net/gtcn/]
- Hanneman, R. A. & Riddle, M. “Introduction to social network methods” 2005 Riverside, CA: University of California, Riverside [http://faculty.ucr.edu/~hanneman/]
- Brandes, U. & Erlebach, T. (ed.) “Network Analysis: Methodological Foundations” 2005 [http://www.springerlink.com/content/nv20c2jfpf28/]

Bibliography in more detail

- Evans, T.S. “Complex Networks”, Contemporary Physics, 2004, 45, 455-474 [arXiv.org/cond-mat/0405123]
- M.Newman, “Networks: An Introduction”, OUP, 2010
- 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.
- Watts, D. J. & Strogatz, S. H., “Collective dynamics of 'small-world' networks”, Nature, 1998, 393, 440-442.
- Zachary, W. “Information-Flow Model For Conflict And Fission In Small-Groups” Journal Of Anthropological Research, 1977, 33, 452-473
- Evans, T.S. & Lambiotte, R., “Line Graphs, Link Partitions and Overlapping Communities”, Phys.Rev.E, 2009, 80, 016105 [arXiv:0903.2181]
- Evans, T.S. “Clique Graphs and Overlapping Communities”, J.Stat.Mech, 2010, P12037 [arxiv.org:1009.063]
- Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R. & Lefebvre, E., “Fast unfolding of community hierarchies in large networks”, J STAT MECH-THEORY E, 2008, P10008 [arXiv.org:0803.0476]

- Also try `arXiv.org` and search for words *networks* with *school* or *review* in whole entry

Bibliography in more detail (2)

- Watts, D. J. & Strogatz, S. H. "*Collective dynamics of 'small-world' networks"* Nature, 1998, 393, 440-442
- T.S.Evans, J.P.Saramäki "Scale Free Networks from Self-Organisation " Phys.Rev.E 72 (2005) 1 [cond-mat/0411390]
- Krapivsky, P. L. & Redner, S. "Finiteness and Fluctuations in Growing Networks" J. Phys. A, 2002, 35, 9517
- M.Newman, "A Measure of Betweenness centrality based on random walks" Social networks 27 (2005) 39-54
- Lambiotte, R.; Sinatra, R.; Delvenne, J.-C.; Evans, T.; Barahona, M. & Latora, V., "Flow graphs: interweaving dynamics and structure", Phys. Rev. E 84, 017102 (2011)
- T.S.Evans & J.Saramäki, "Scale-free networks from self-organization" Phys.Rev.E. 72 (2005) 026138 [arXiv: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 [arXiv:cond-mat/0612214]

Bibliography: General Network Reviews

- Newman, M. “Networks: An Introduction”, OUP, 2010

Conclusions

Considerable input possible from from a mathematical approach to networks

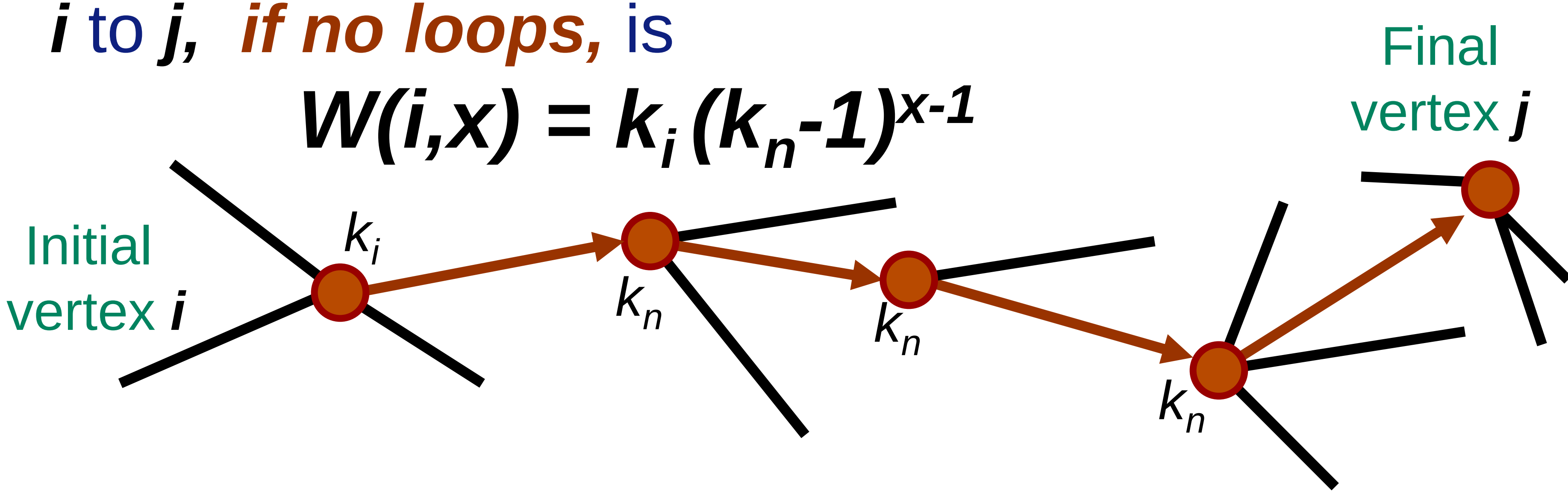
Google “Tim Evans Networks”
to find my web pages on networks

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- (***k_j*** **/2*E***)

⇒ 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 ***$p_{ij}(x-1)$* - *$p_{ij}(x)$***

⇒Average path length from ***i*** to ***j*** is

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

⇒ Average path length **<*d*>** is (after some work)

[Fronczak et al,2005]

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

Average Path Length in MR Random Graph (4)

- For ***any*** random graph has an average shortest length which scales as

$$\langle d \rangle \approx \frac{\ln(N)}{\ln(z)} + c$$