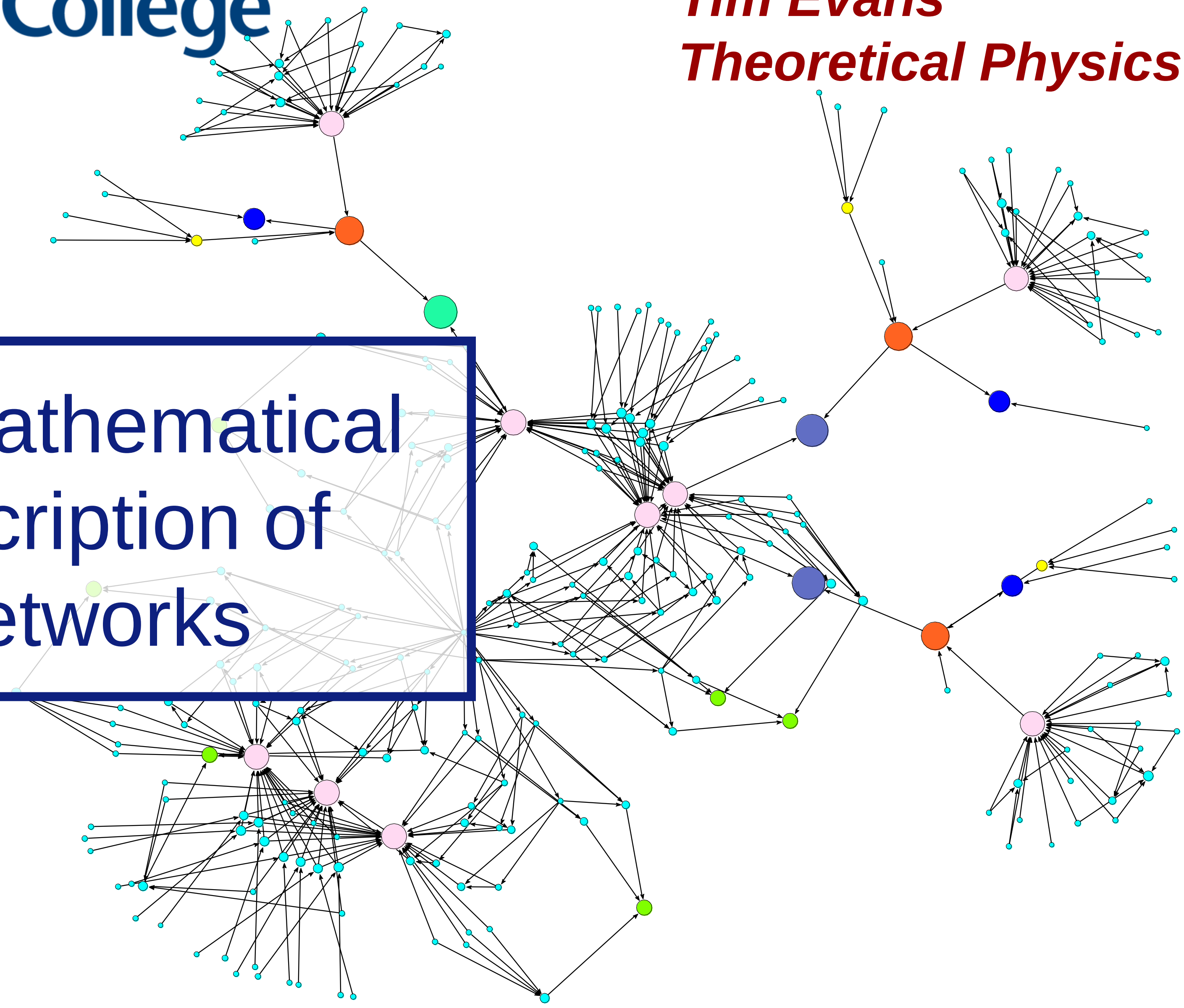
Imperial College London

Modelling Complex Systems
University of Manchester, 21st – 23rd June 2010

Tim Evans
Theoretical Physics

The Mathematical Description of Networks

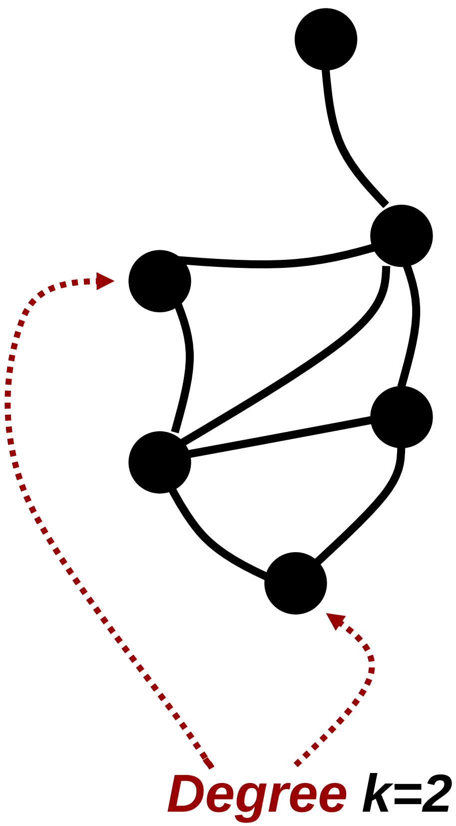


Notation

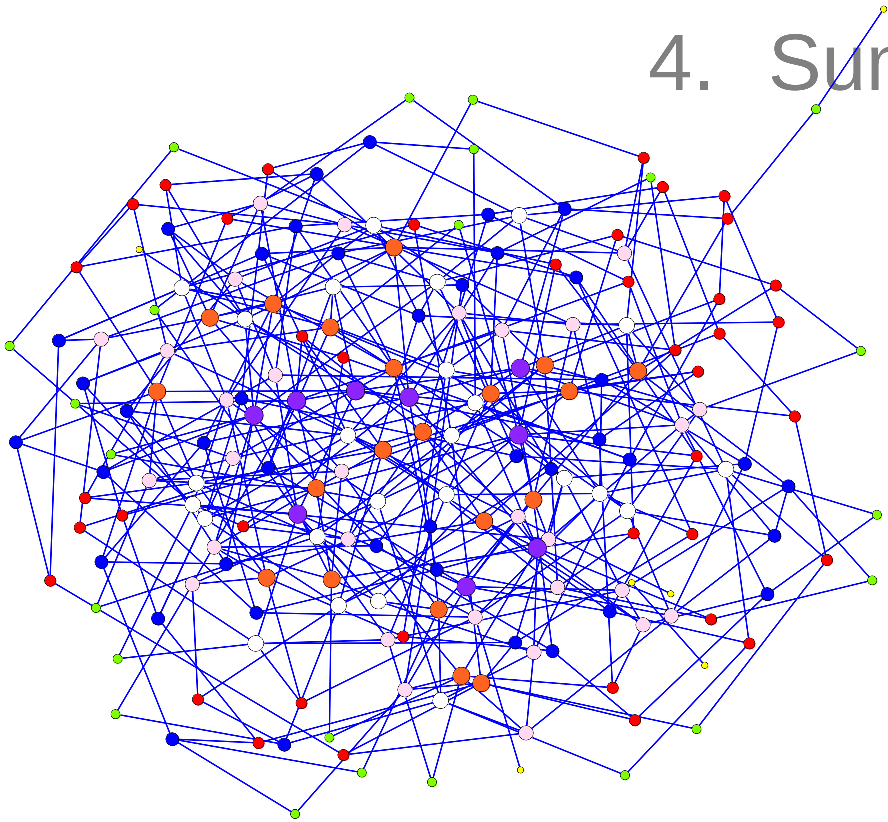
I will focus on **Simple Graphs** with multiple edges allowed

(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*



1. Random Graphs
2. Scale-Free Models
3. Different Network Views
4. Summary



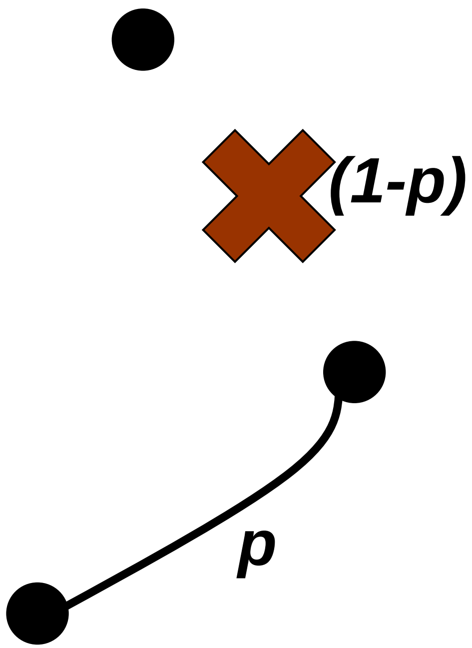
How to excite a Mathematician – give them the simplest network model

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

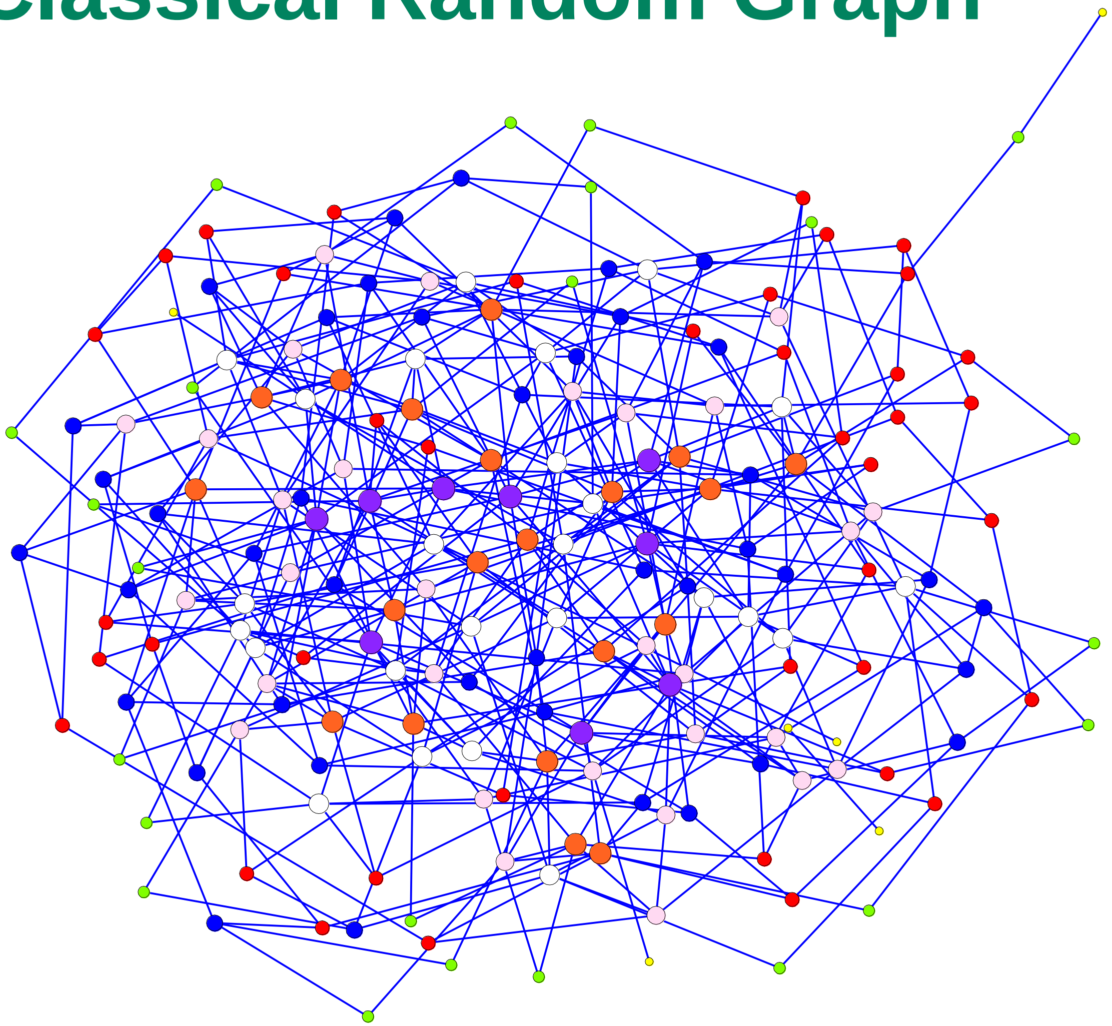
$$p(k) \approx \frac{\exp(\langle k \rangle)\langle k \rangle^{k}}{k!}$$

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***

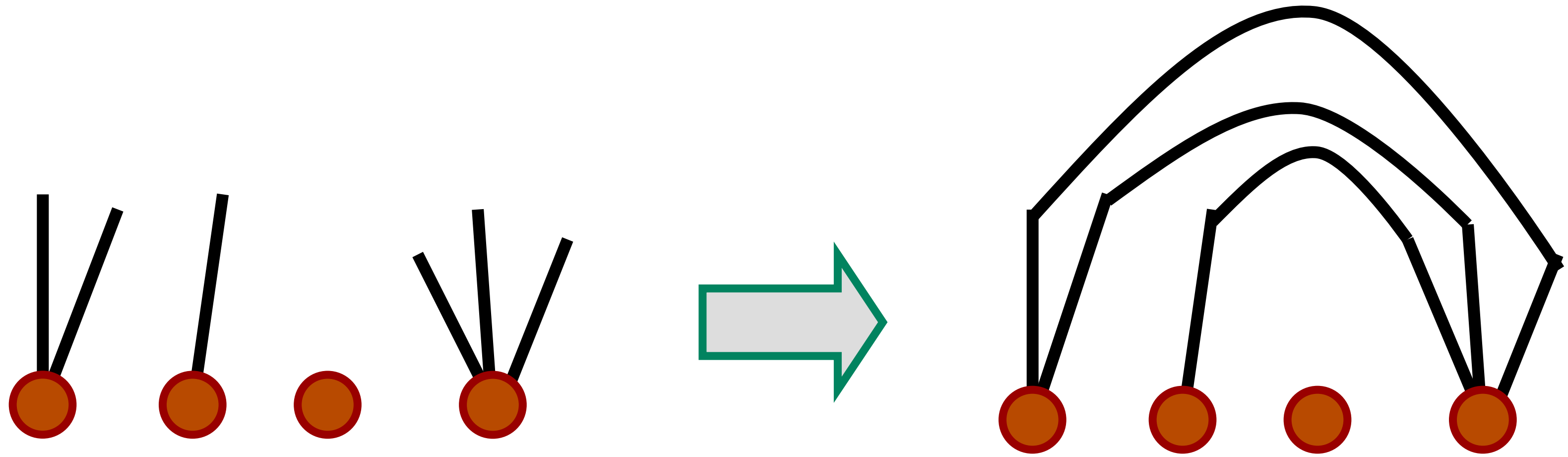


Generalised Random Graphs –
The **Molloy-Reed** Construction [1995,1998]

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

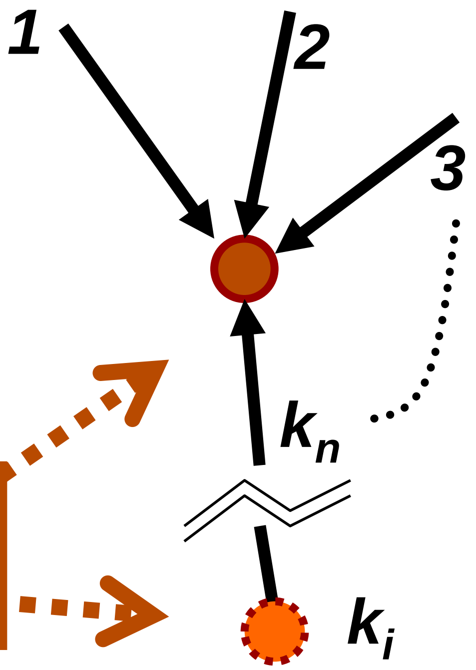
For any given source vertex, the properties of neighbouring vertices independent of properties of the source vertex

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

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)

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

Give a random friend that life saving vaccine

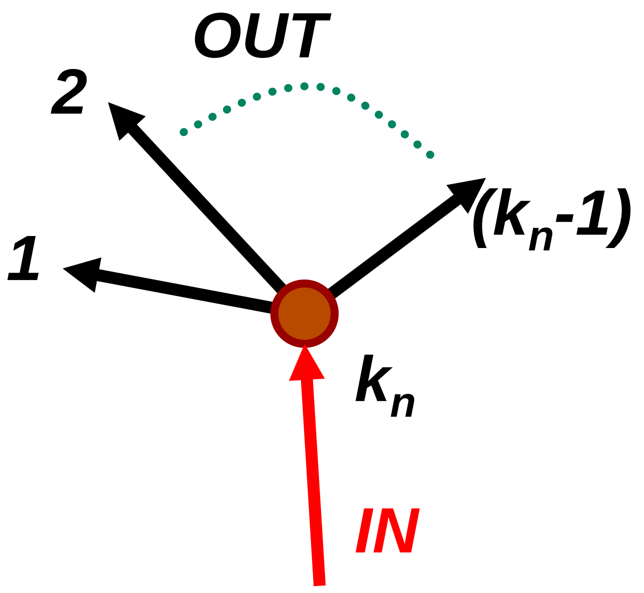
(if social networks are random and uncorrelated)

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 $\boldsymbol{(N \longrightarrow \infty)}$

$\Rightarrow$ 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.

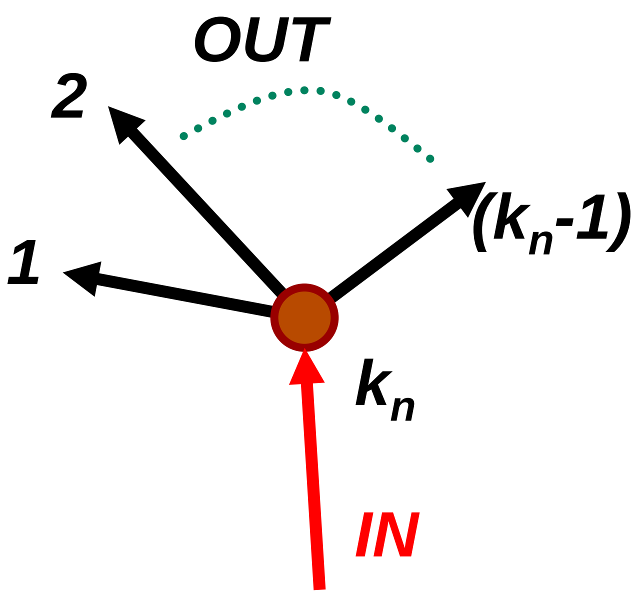
True for sparse random graphs in limit of infinite size ***(N*** $\longrightarrow \infty$***)***

$\Rightarrow$ Walks always end if

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

$\Rightarrow$ Walks never end if

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



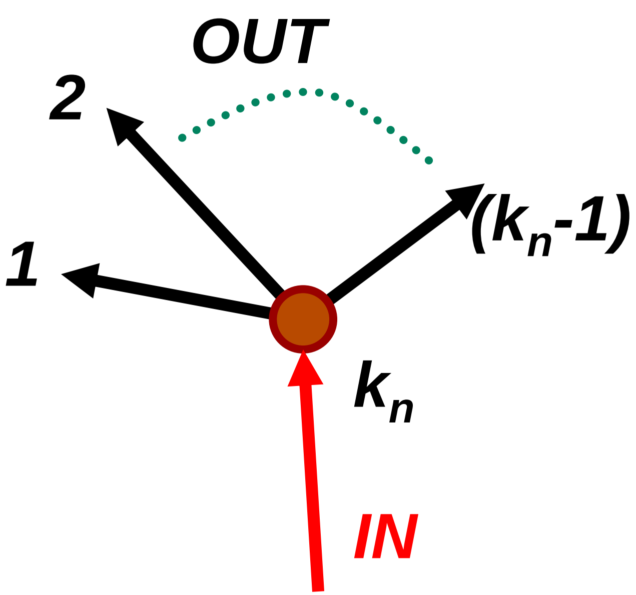
(GCC= Giant Connected Component)

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

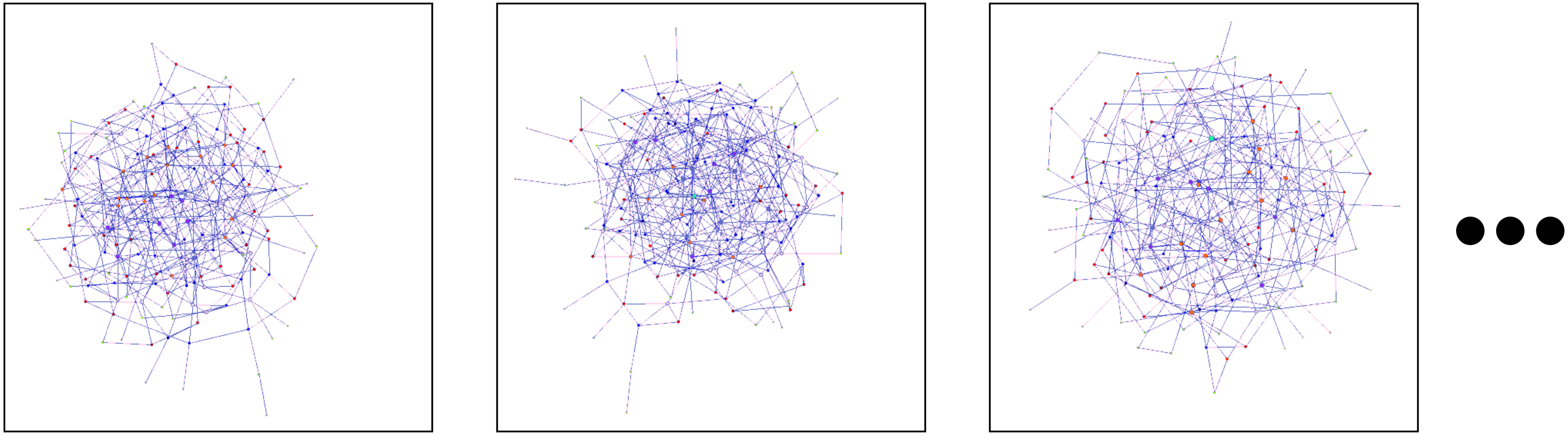
All global properties depend on same

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

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

Ensembles of Graphs

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

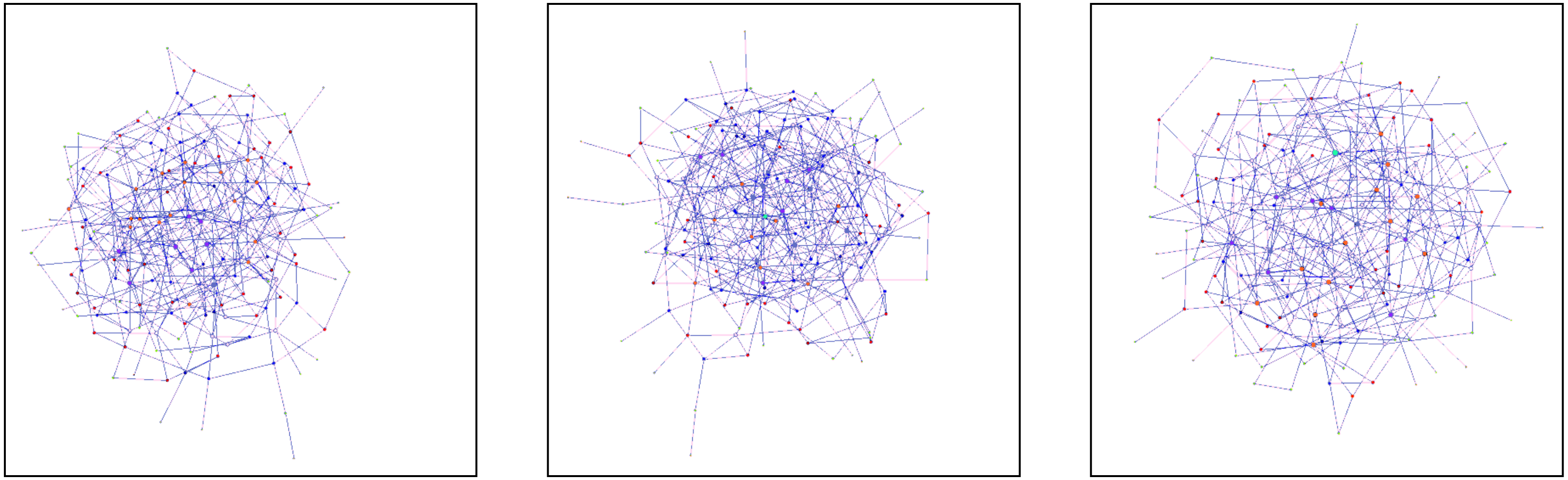


Ensembles of Graphs

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

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

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 >**

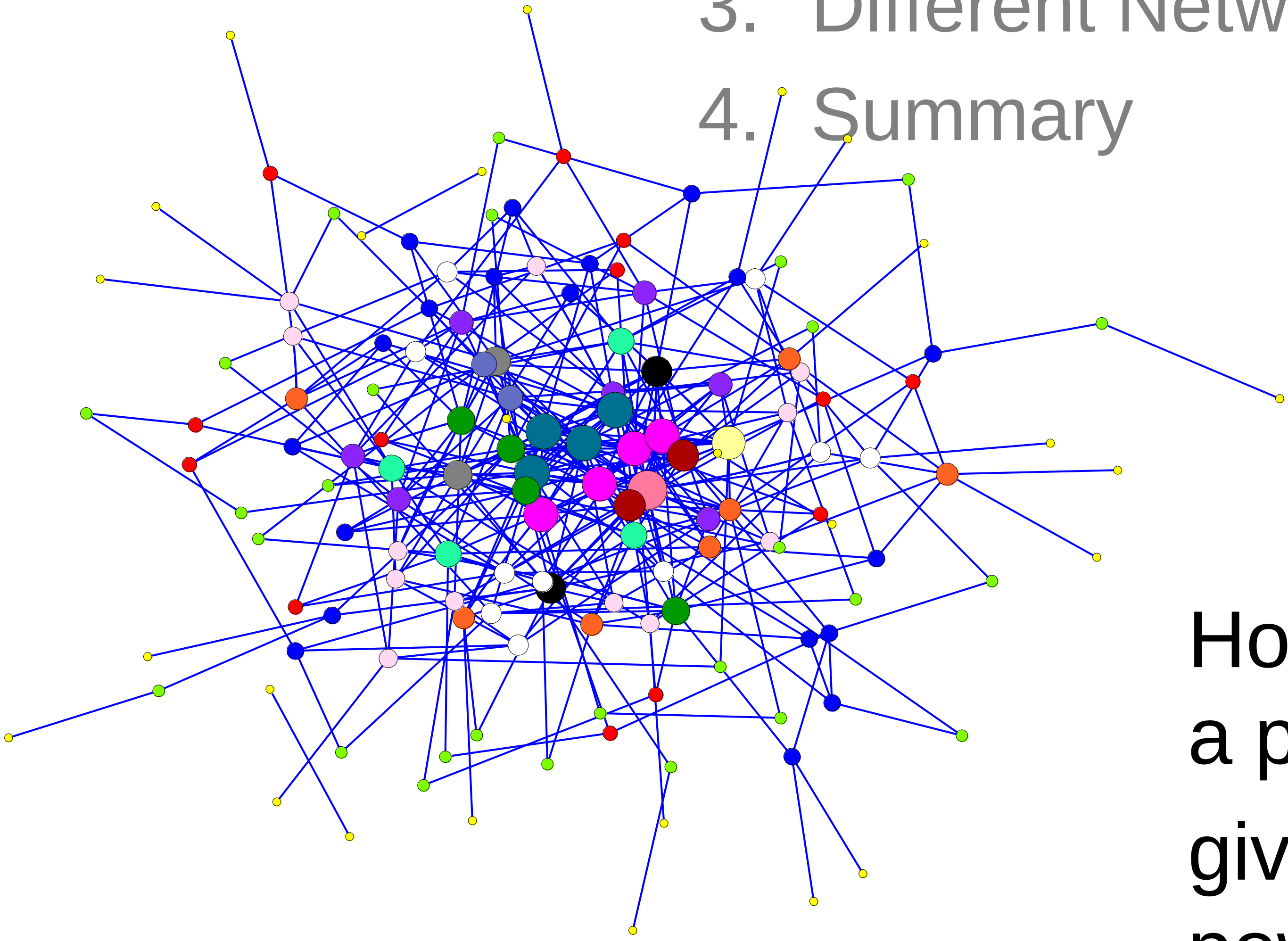
MR Random Graph Calculations

- Calculations like those above work because
  - lack of correlations between vertices
  - few loops for large sparse graphs, graphs are basically trees
- These can be reasonable approximations for many models and perhaps for a few real graphs too. Otherwise use as a ***null model***.

Summary of Random Graphs

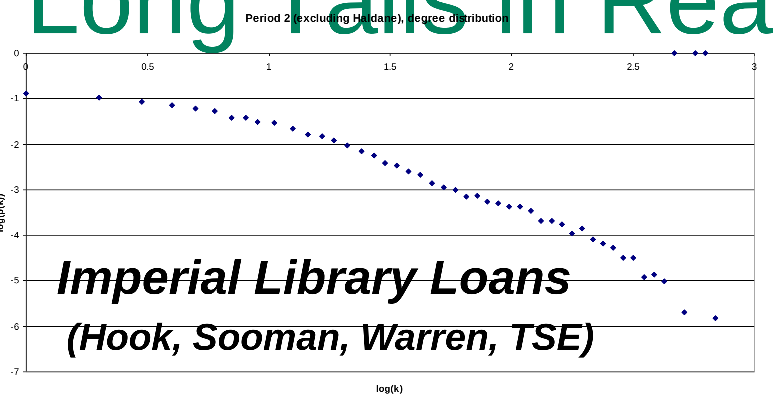
- Calculations work because
  - lack of correlations between vertices
  - few loops for large sparse graphs, graphs are basically trees
- Accessible analytically so can suggest typical behaviour even if very weak e.g. diameter vs N
- These can be reasonable approximations for many theoretical models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

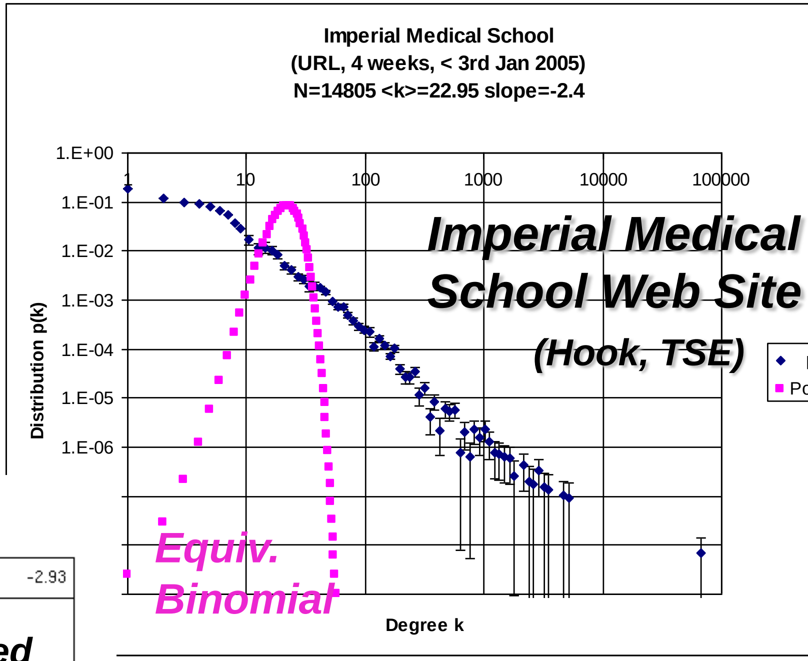
1. Random Graphs
2. Scale-Free Models
3. Different Network Views
4. Summary

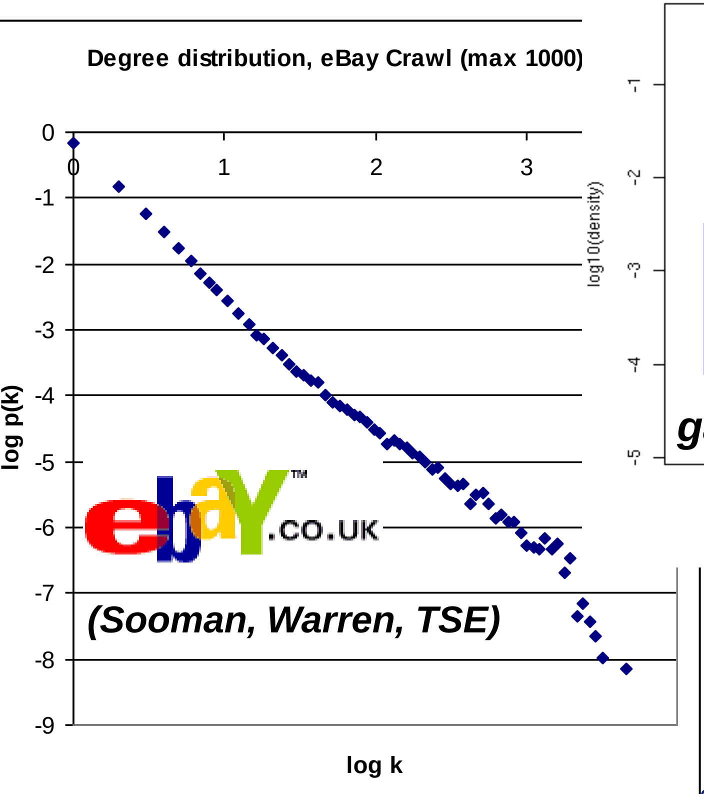


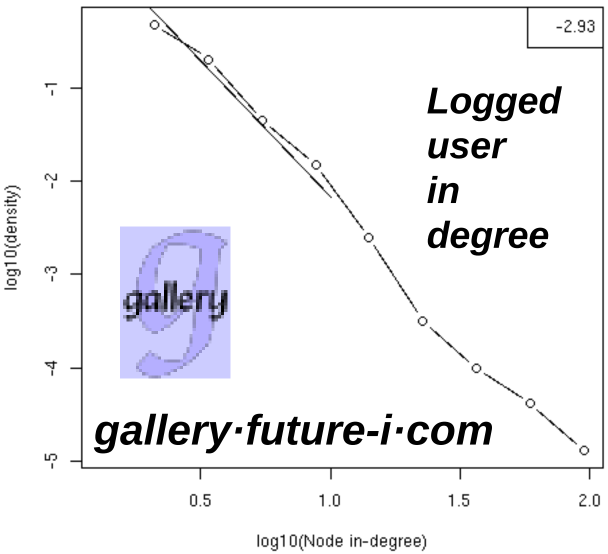
How to excite a physicist – give them a power law

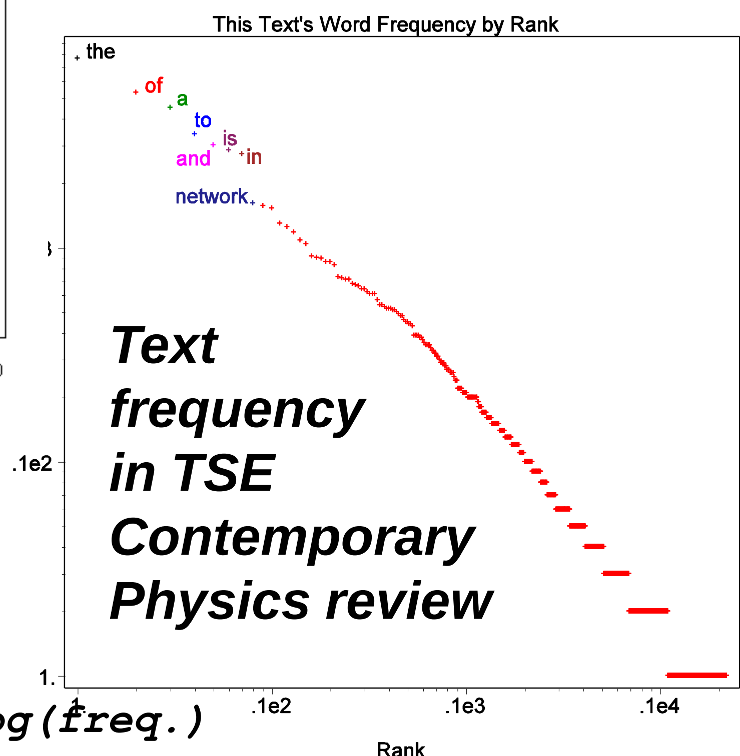
Long Tails in Real Data











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

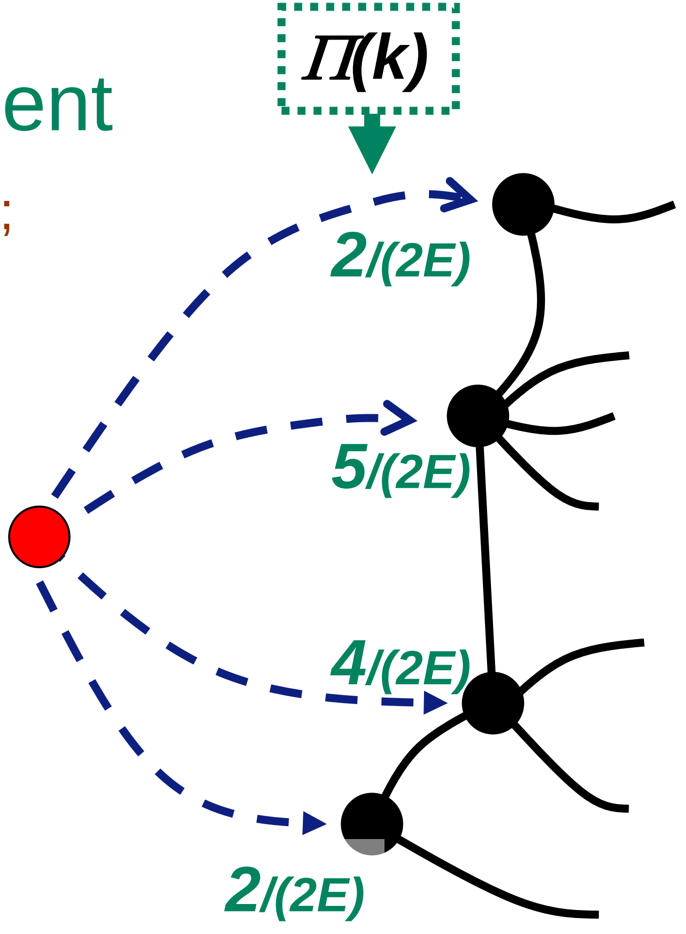
Growth with Preferential Attachment

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

1. Add new vertex attached to one end of **$m=\frac{1}{2}\langle k \rangle$** 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"



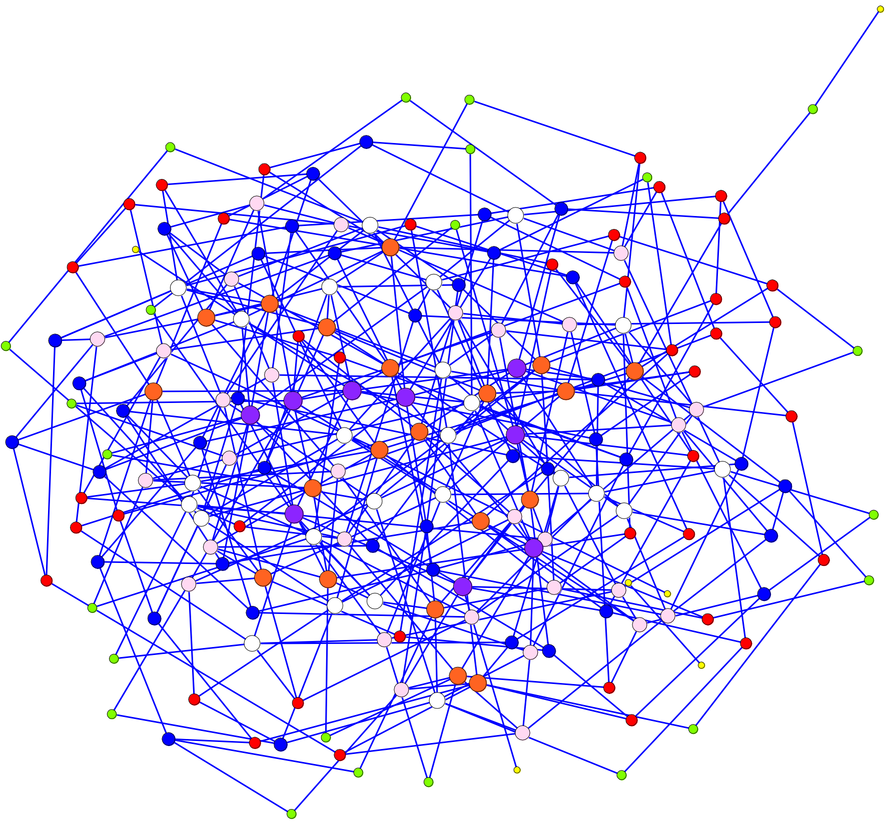
Result: Scale-Free

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

$$\gamma = 3$$

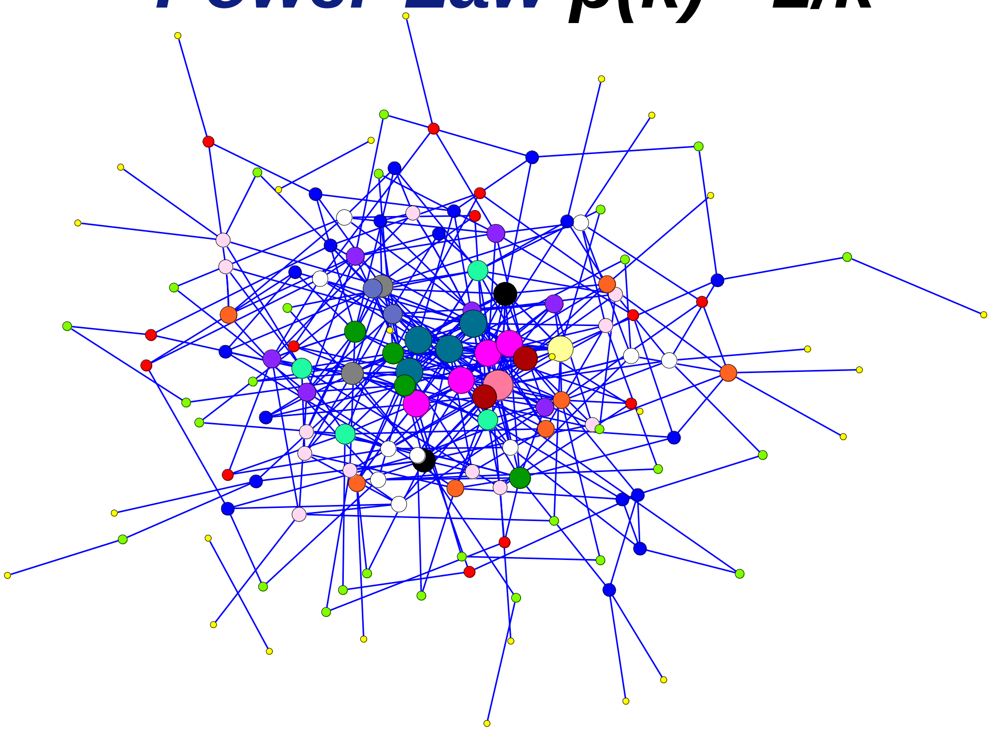
N=200, <*k*>~4.0, vertex size ∝ *k*

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).***
- Decreases as we add an edge to existing vertex of degree ***k.***
- Number of vertices of degree $k=m=\frac{1}{2}\langle k\rangle$ always increase by ***1*** as add new vertex.

Mean Field Degree Distribution Master Equation

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

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 $\Pi(k)$ 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

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

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

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

$\beta=0$ or $\beta=1$

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 $\Rightarrow$ 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 – $p_p > 1$ 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 $\Rightarrow$ $\gamma \sim 1.0$

[Moran model, see TSE,Plato, 2008]

Scale-Free in the Real World

Attachment probability used was

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

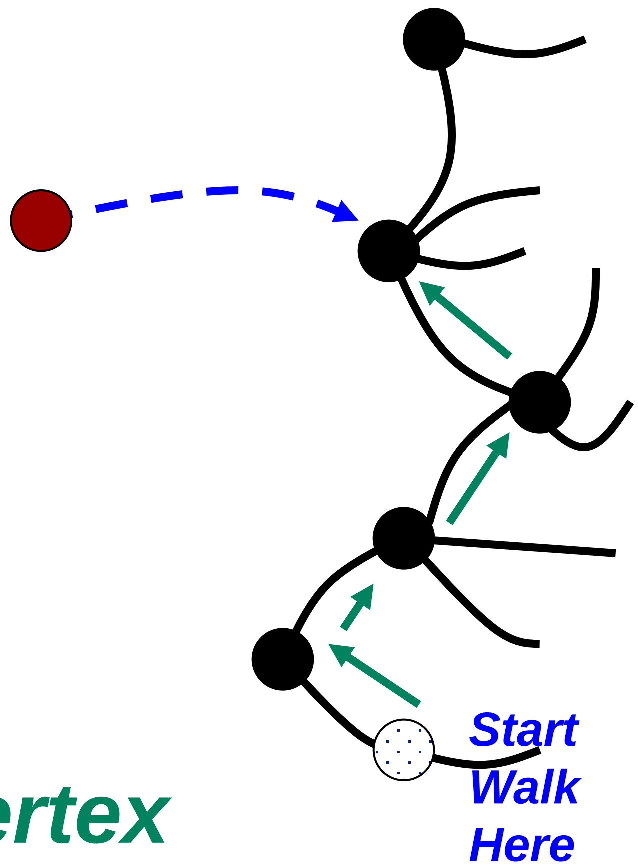
BUT if $\lim_{k\to\infty} \Pi(k) \propto k^{\alpha}$ for any $\alpha \neq 1$ then a *power law degree distribution is not produced!*

A

Preferential Attachment for Real Networks

[Saramäki, Kaski 2004; TSE, Saramäki 2004]

1. Add a new vertex with $\frac{1}{2}\langle k \rangle$ 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)

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

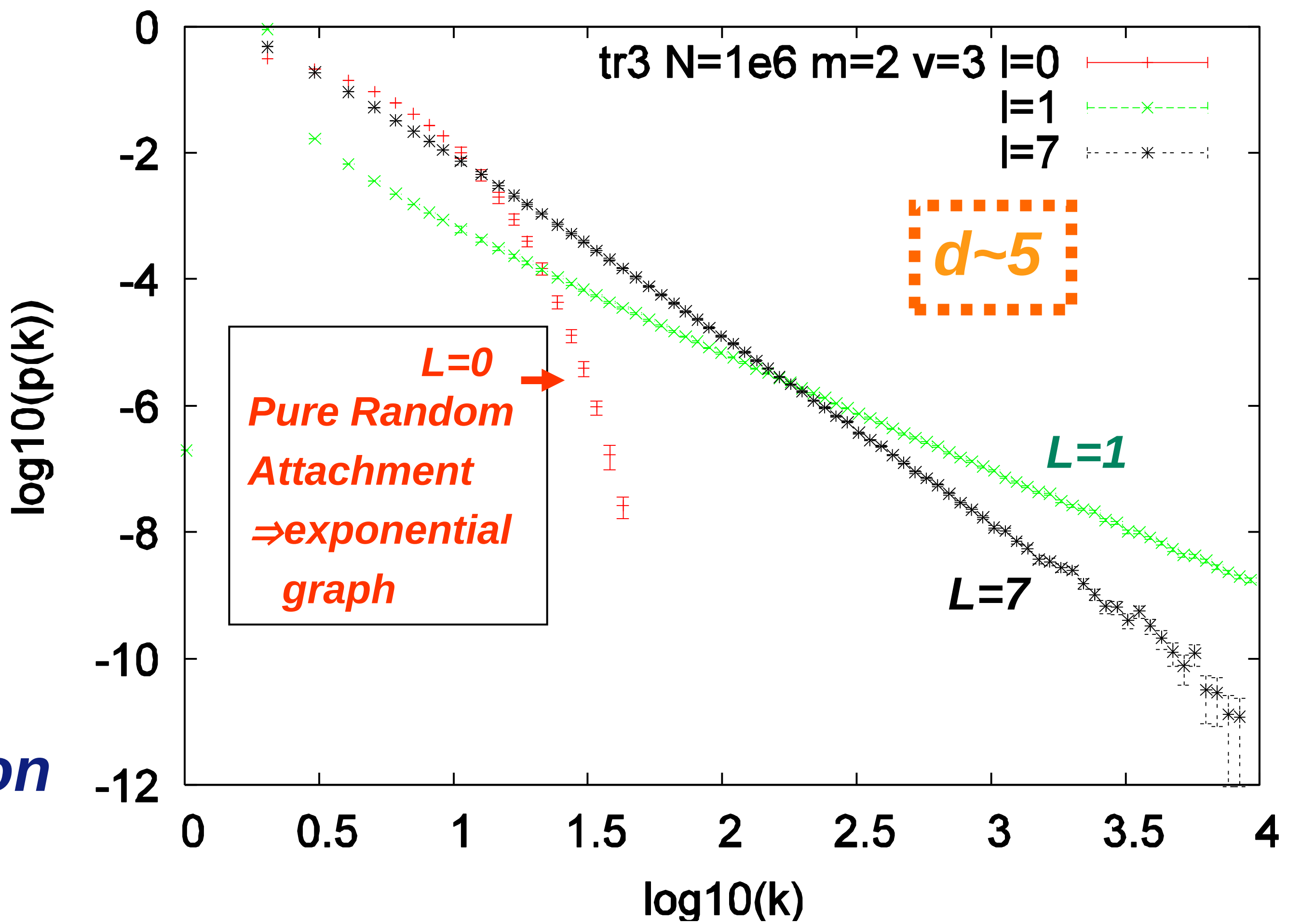
    e.g. informal requests for work on the film actor’s social network

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

Is the Walk Algorithm Robust?

I have varied:

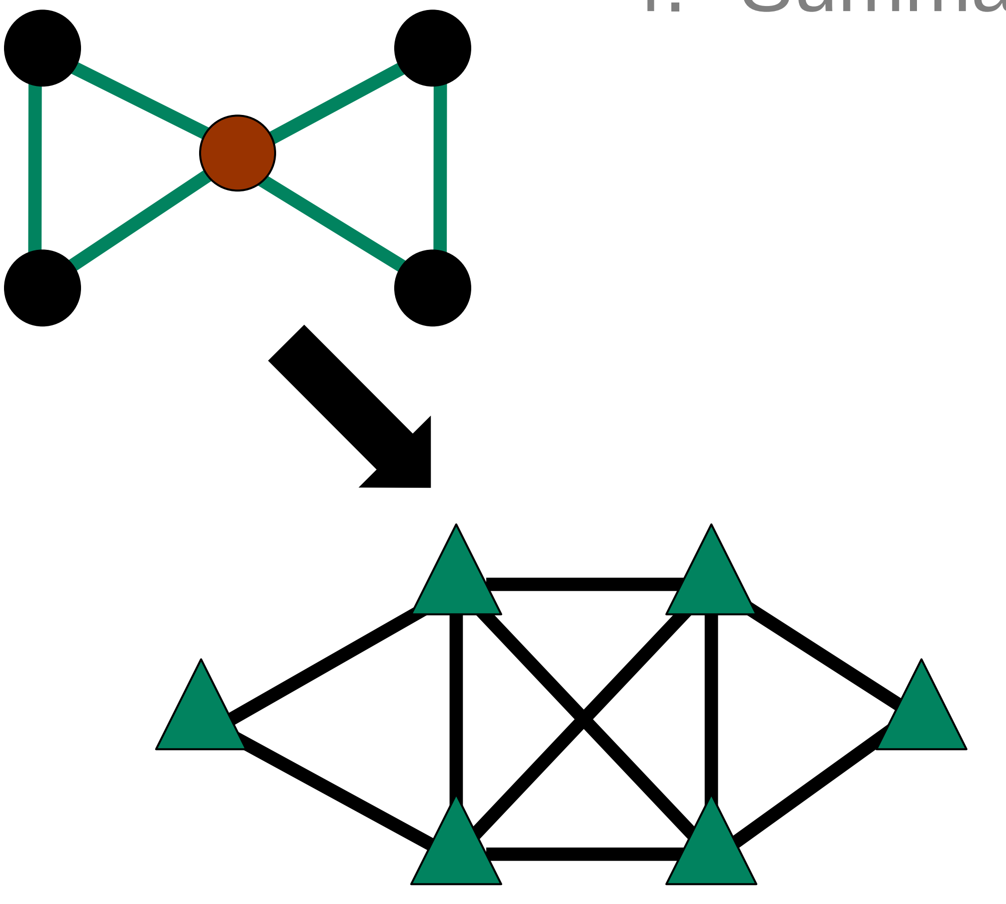
- ***Length of walks***
- <k>
- ***Starting point of walks***
- ***Length distribution of walks***
- *.....*



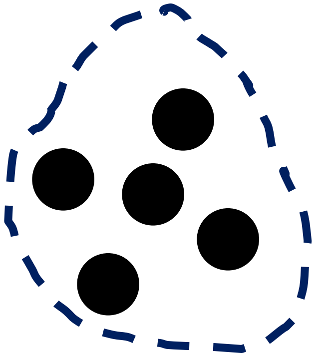
YES - Good Power Laws

but NOT Universal values - 10% or 20% variation

1. Random Graphs
2. Scale-Free Models
3. Different Network Views
4. Summary



What is a Network?

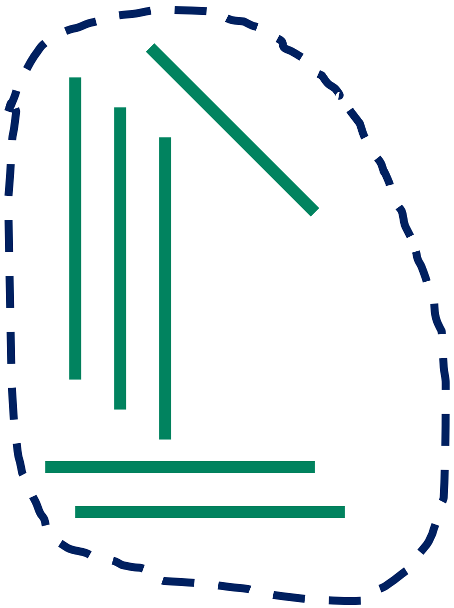


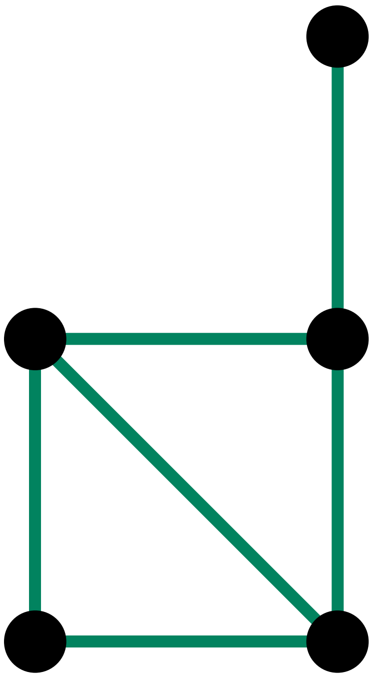
A Network is **BOTH**

a set of vertices

AND

a set of edges





Vertex Centric Viewpoint

YET we tend to have a very VERTEX centred viewpoint:-

- Degree of a ***vertex***
- Communities usually a partition of ***vertices***
- Distances are ***vertex*** to ***vertex*** distances
- Word frequencies in Network Review

[TSE, Contemporary Physics 2004]

Network Words	Vertex Words	Edge Words
6	3	2

Word Frequencies in Network Review

Word	Rank	Count	Word	Rank	Count
network	1	254	distribut	21	34
vertic	2	107	scale	21	34
edg	3	86	problem	24	33
random	3	86	simpl	24	33
graph	5	81	idea	26	30
degre	6	78	physic	26	30
power	7	68	size	26	30
lattic	8	67	find	29	29
law	9	65	real	29	29
vertex	10	61	type	31	27
number	11	58	case	32	26
distanc	12	48	hub	33	25
model	13	47	show	33	25
connect	14	46	area	35	24
data	15	40	neighbour	35	24
link	16	38	studi	35	24
world	16	38	point	38	23
larg	18	37	term	38	23
small	19	36	figur	40	22
averag	20	35	form	40	22
comput	21	34	site	40	22

Stop words removed and stemmed [TSE, Contemporary Physics 2004]

Edge Centric Viewpoint?

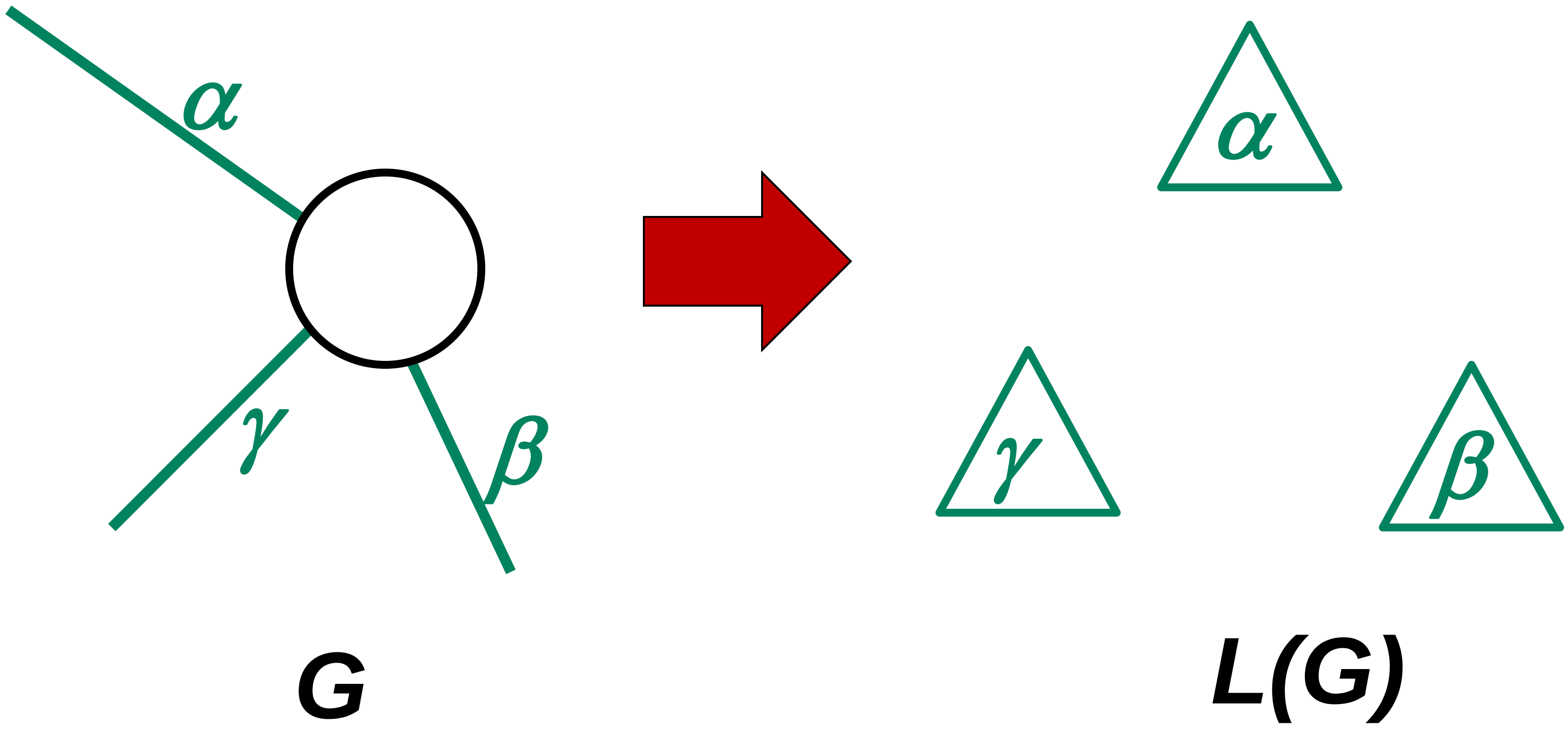
Create a new graph ***L(G)*** whose vertices represent the edges of the old graph, ***G***, and whose edges record the overlap of the old edges i.e. the common vertices in ***G***.

This is called the LINE GRAPH

Name given by Harary who was using the word *"line"* for *edges* but this construction predates this work.

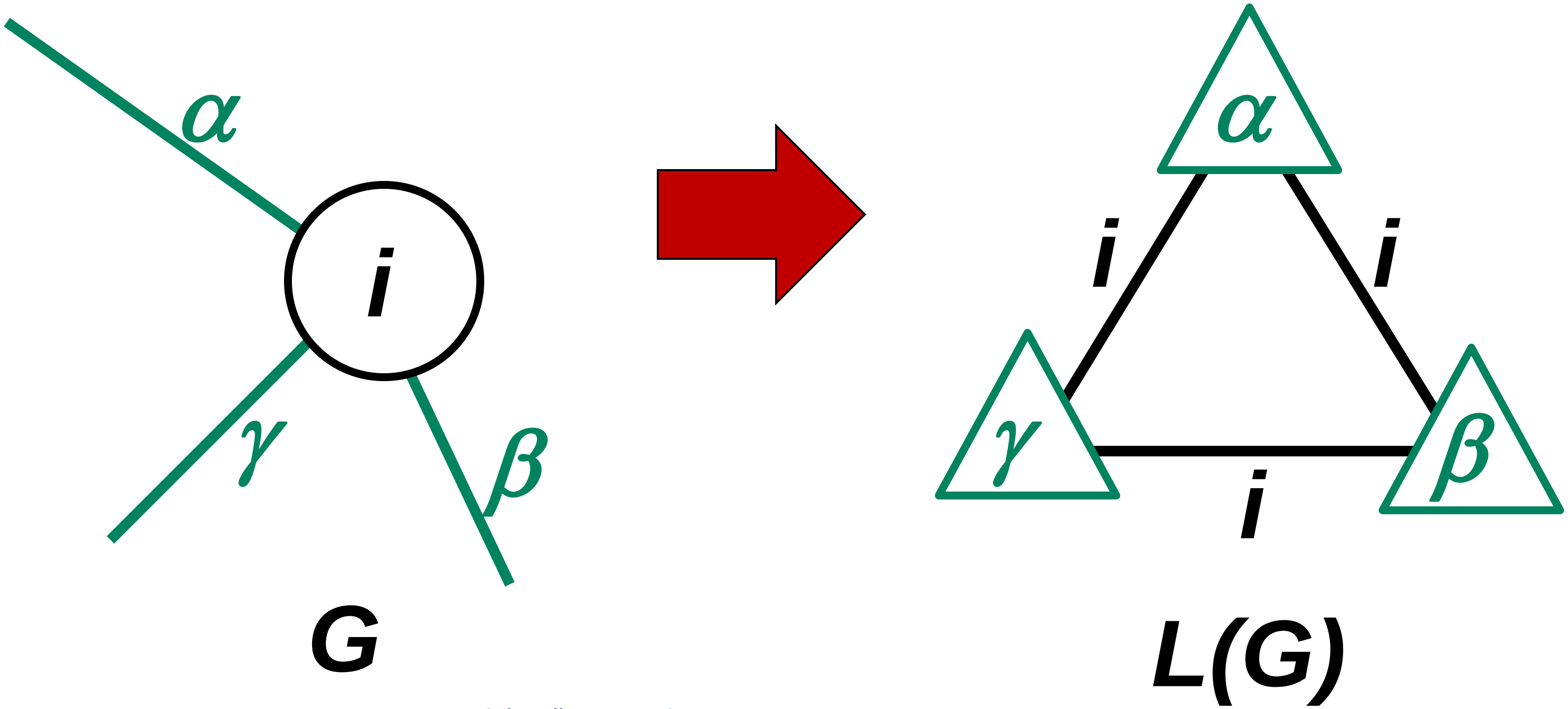
Vertices of a Line Graph

1. For every edge α in original graph ***G*** create a vertex α in the line graph ***L(G)***



Edges of a Line Graph

2. Connect the vertices α and β in the line graph **L(G)** if the corresponding edges in original graph ***G*** were coincident



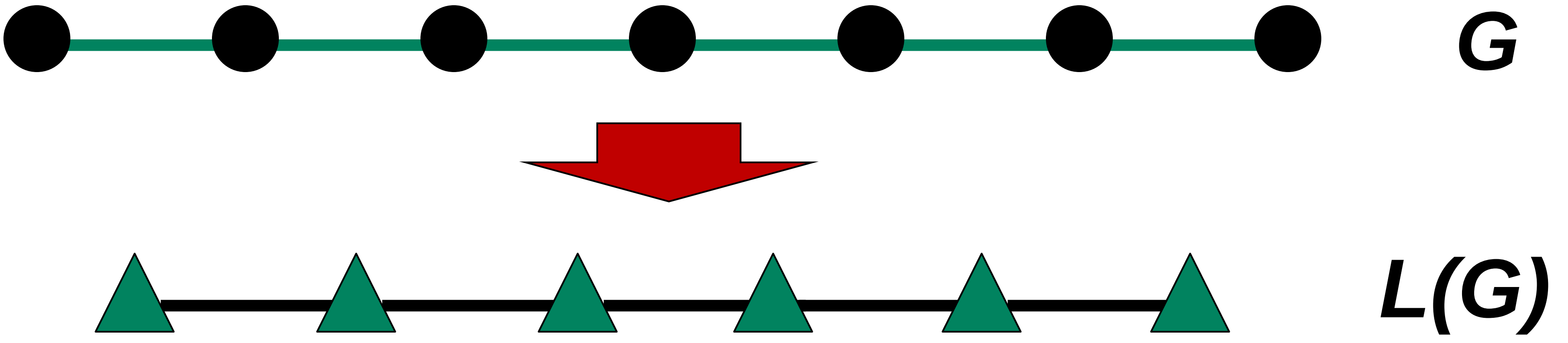
Properties of a Line Graph

- Not usually a duality transformation

$$L(L(G)) \neq G$$

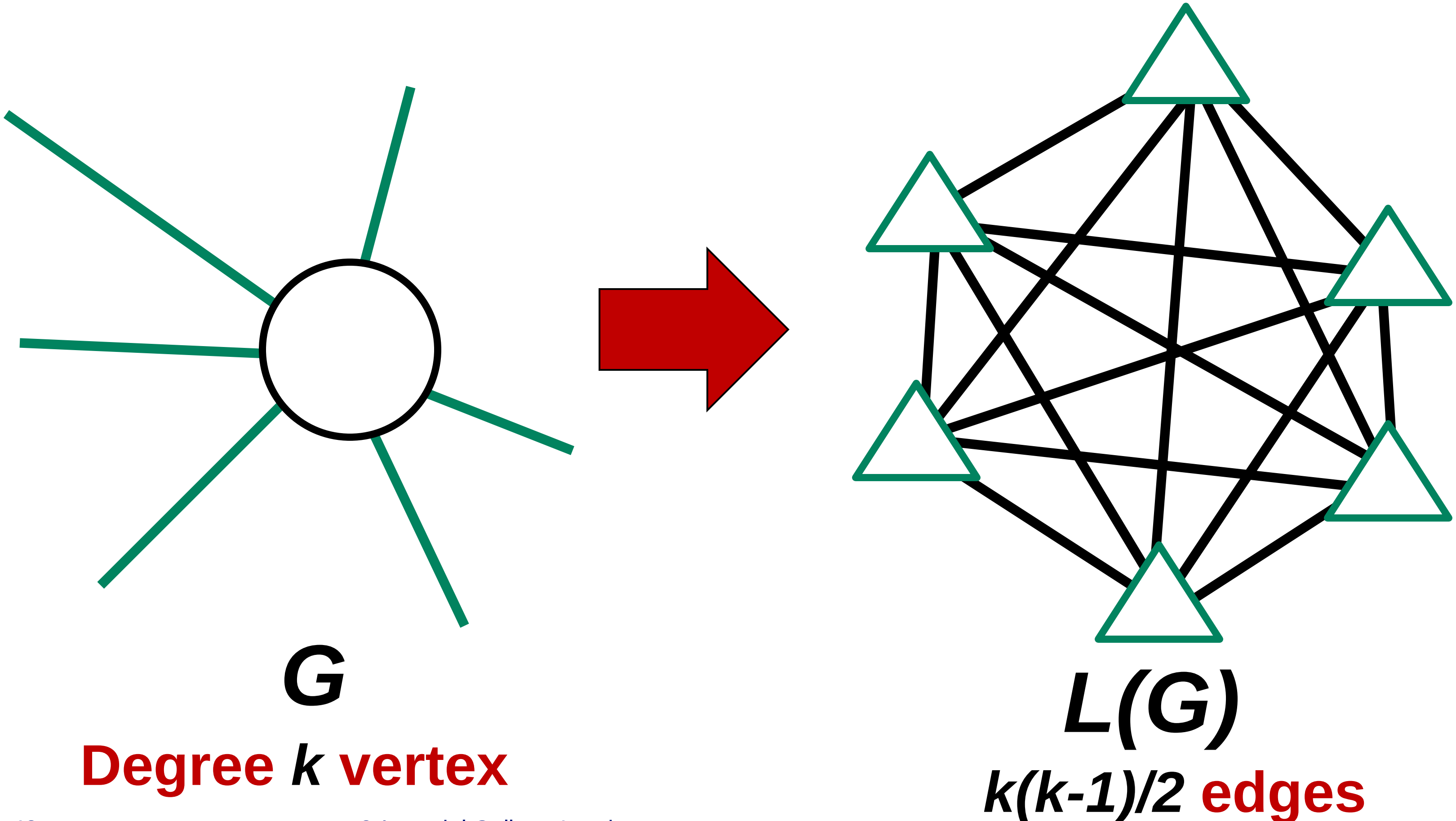
- (Almost) always reversible [Whitney 1932]

$$L(G) \rightarrow G$$



The Problem with a Standard Line Graph

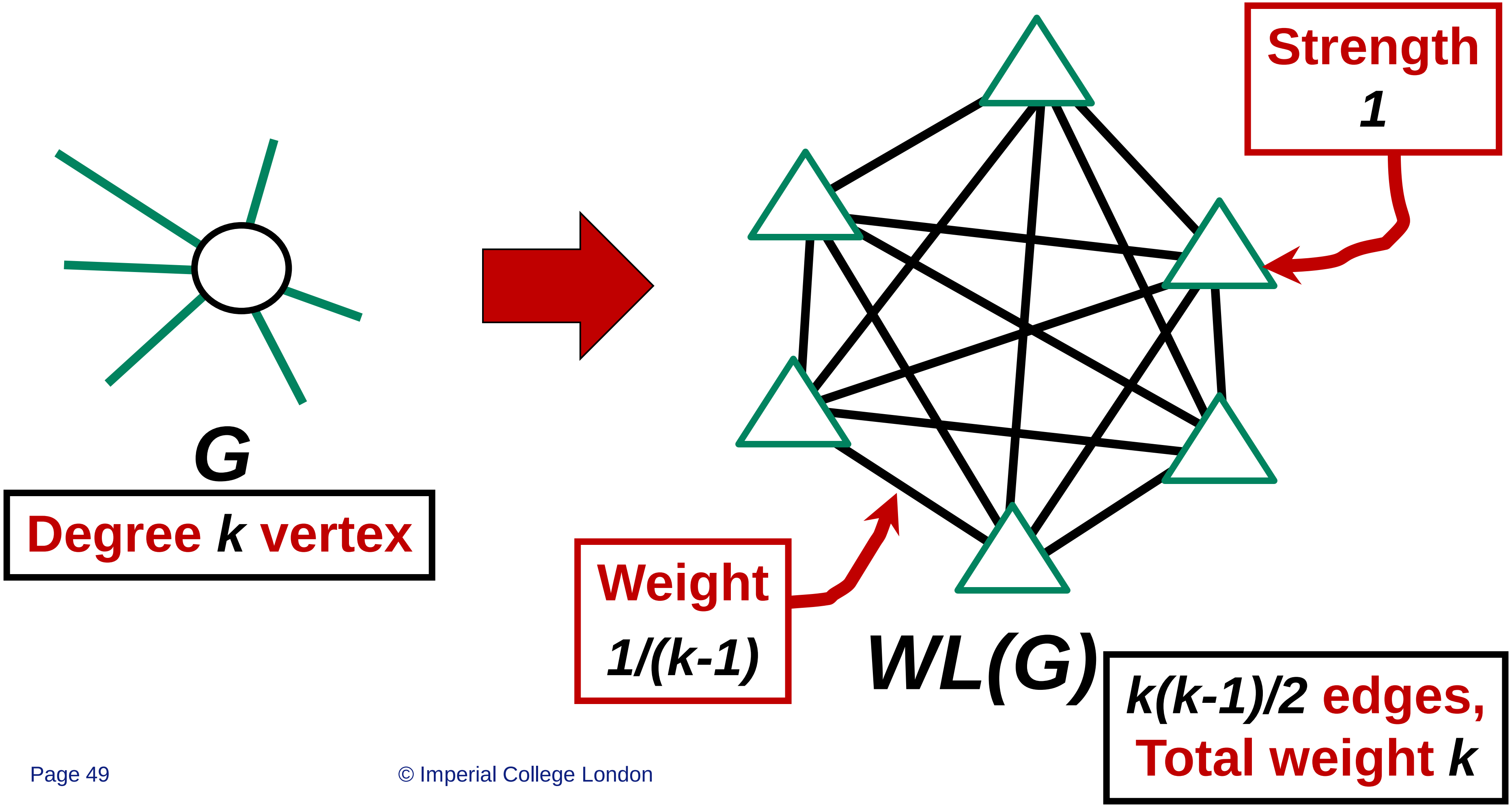
High degree vertices in original graph ***G*** over represented by edges in Line Graph ***L(G)***.



[Evans & Lambiotte 2009]

Solution – Weighted Line Graphs

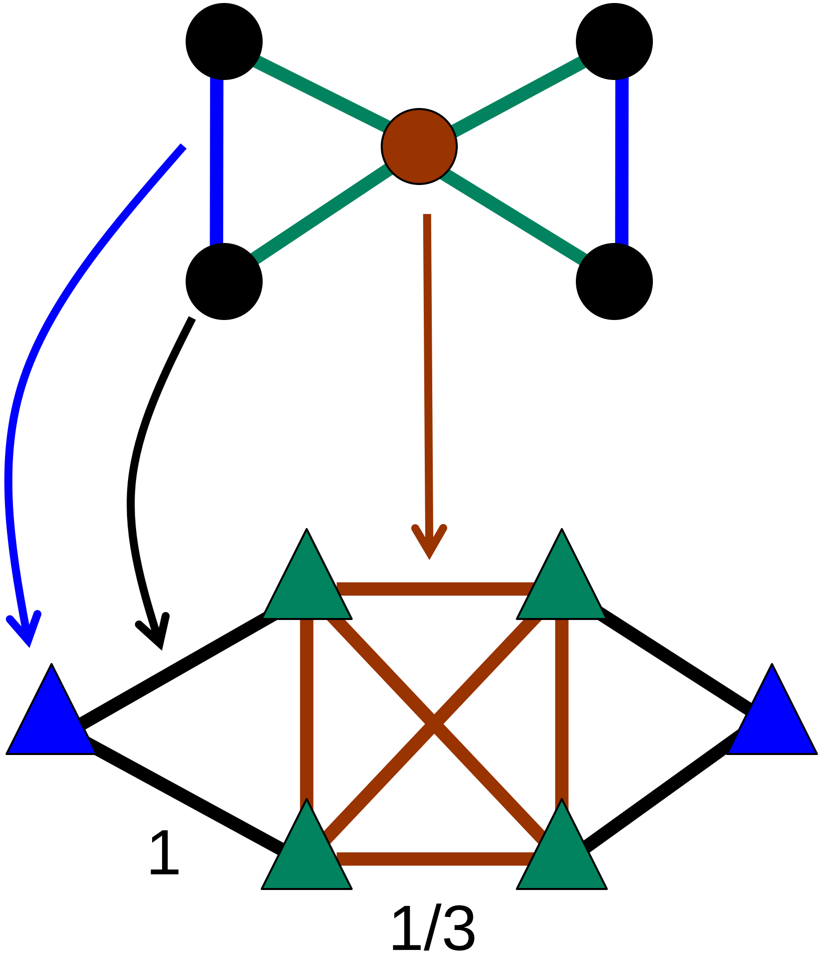
- Original graph vertex of degree ***k*** produces line graph edges of weight ***1/(k-1)***



Example – Bow Tie Graph

Graph *G*

Line Graph *L(G)*



Community Detection - Vertex Centric version

- Partition vertices into communities
- Perform random walk on vertices
- Compare number of random walkers which stay within community after one step against number which remain within communities after infinite number of steps

= Optimisation of Modularity

[Girvan & Newman 2002;
Lambiotte, Delvenne & Barahona 2008]

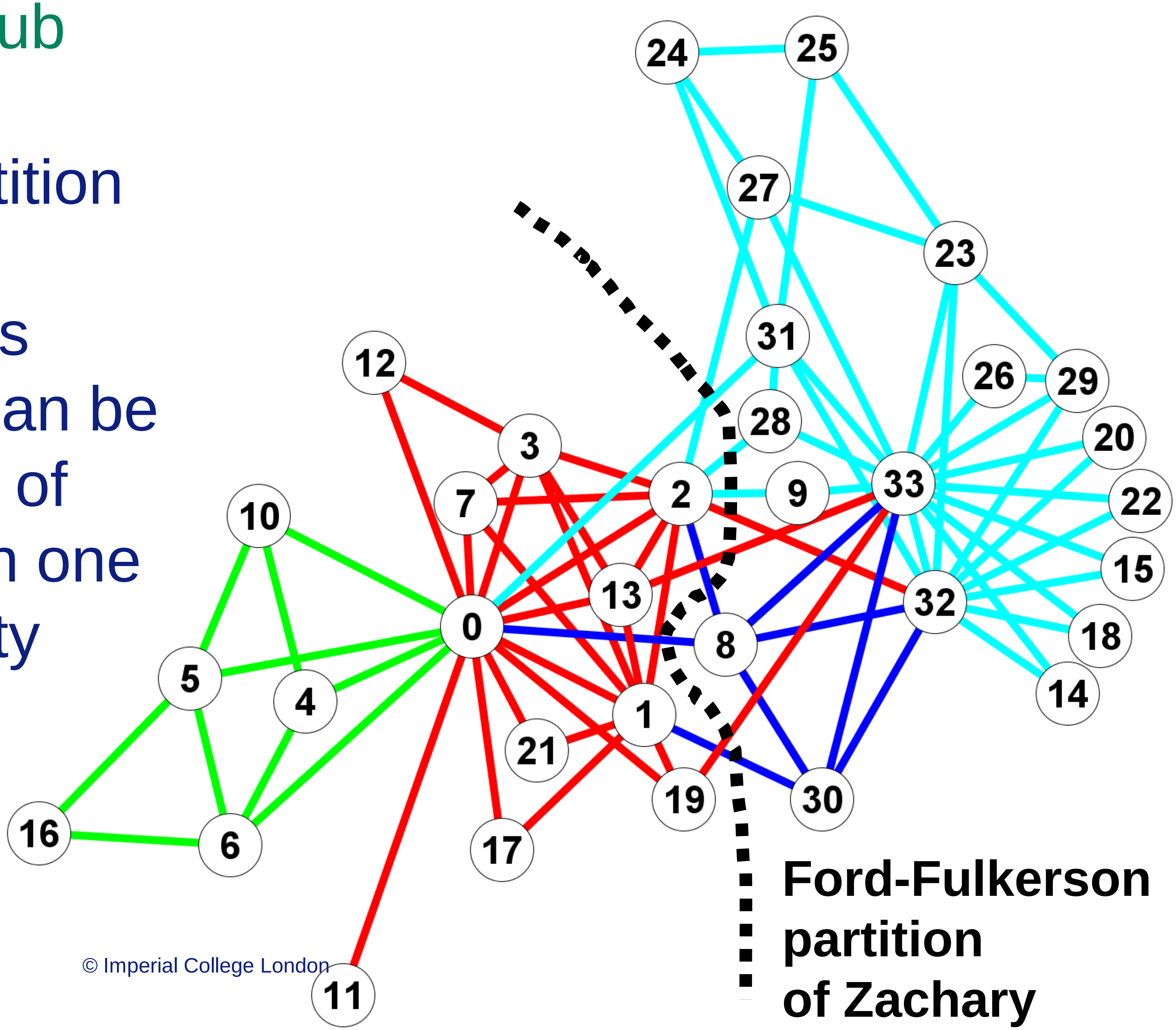
Community Detection - Edge Centric version

- Partition edges into communities
- Perform random walk on edges
  = Random walk on line graph vertices
- Compare number of random walkers which stay within community after one step against number which remain within communities after infinite number of steps

= Optimisation of Modularity of Weighted Line Graph [Evans & Lambiotte 2009]

Karate Club

Edge partition means individuals vertices can be members of more than one community

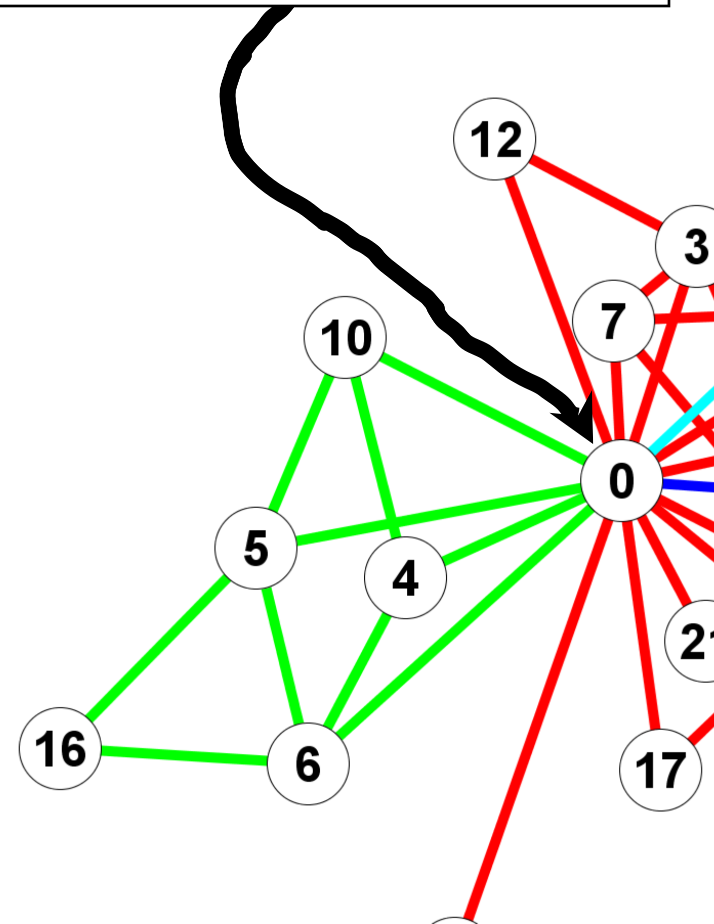


Karate Club Analysis

Vertices in One Edge Community

#	k	Fraction k In Green C
5	4	100%
6	4	100%
10	3	100%
4	3	100%
16	2	100%
0 (Mr_Hi)	16	25%

Mr Hi (the Instructor) bridges several groups

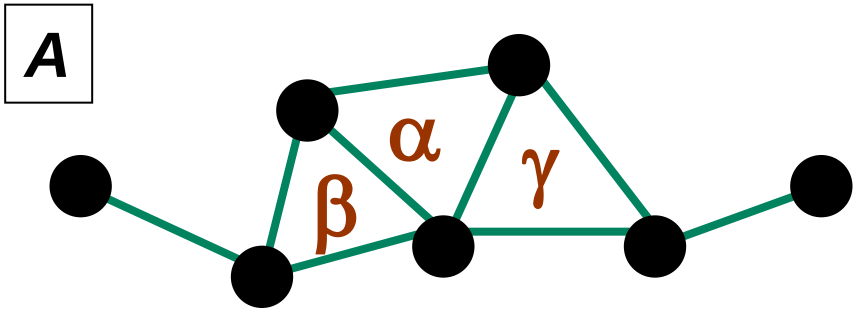


Give Cliques a Chance – a clique centric viewpoint

Can we shift out view point from vertices to cliques?

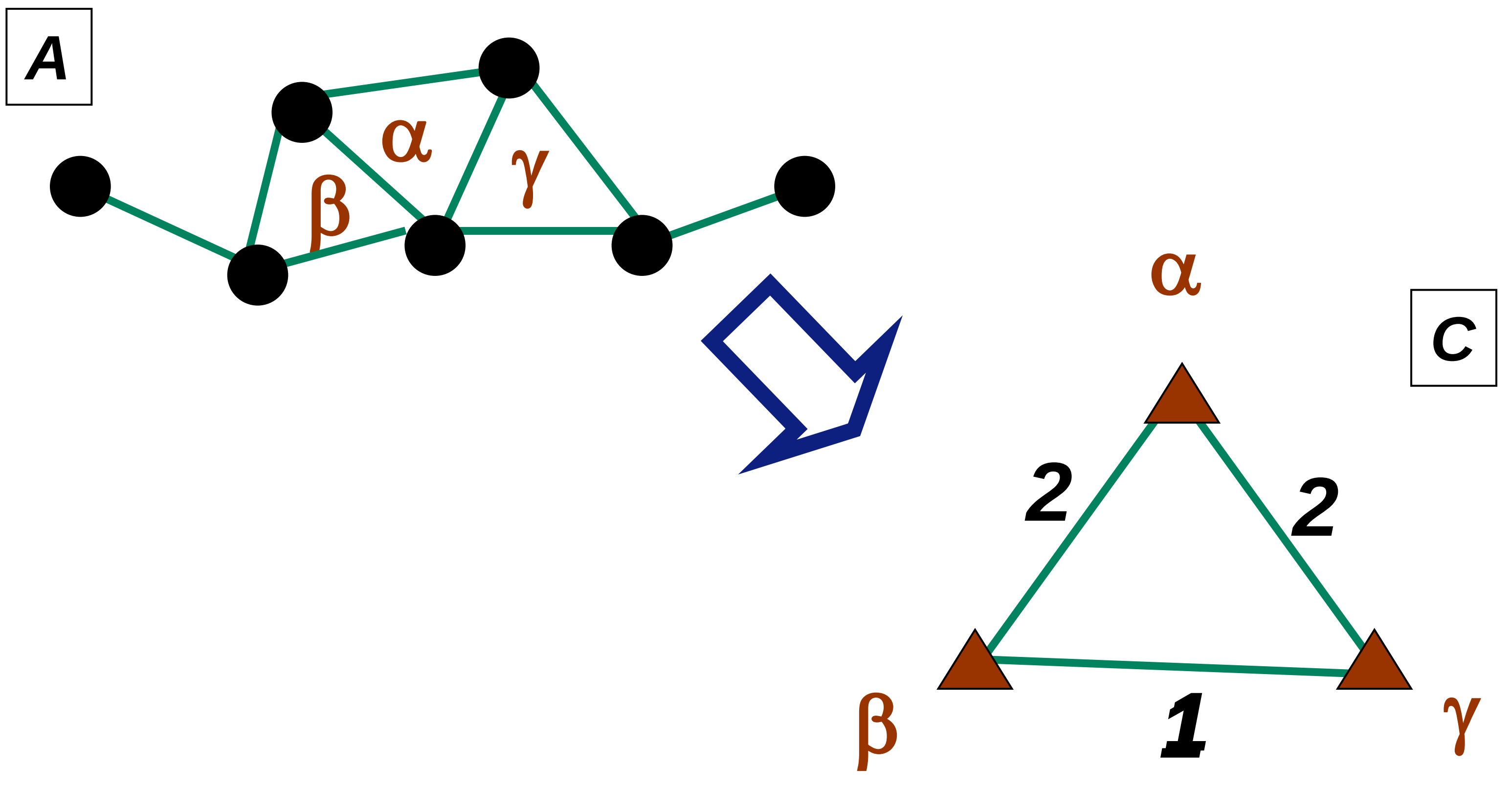
Use a **Clique Graph** to represent the way that ***n***-cliques overlap.

Example
Graph ***A*** has three 3-cliques = triangles



Edge Weights in Basic Clique Graph

Weight of an edge in the basic Clique Graph ***C*** records the number of vertices common to two ***n***-cliques in in the original graph ***A***.



Summary

- Line graphs, clique graphs, and graphs of the overlap of other structures, move focus from vertices to edges/cliques etc with minimal effort
- Weighted line graphs avoid problem of over representation of high degree vertices
- *Example:* Community detection on line graph produces overlapping vertex communities for original graph

1. Random Graphs
2. Scale-Free Models
3. Different Network Views
4. Summary

Conclusions

Considerable input possible from from a mathematical approach to networks

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