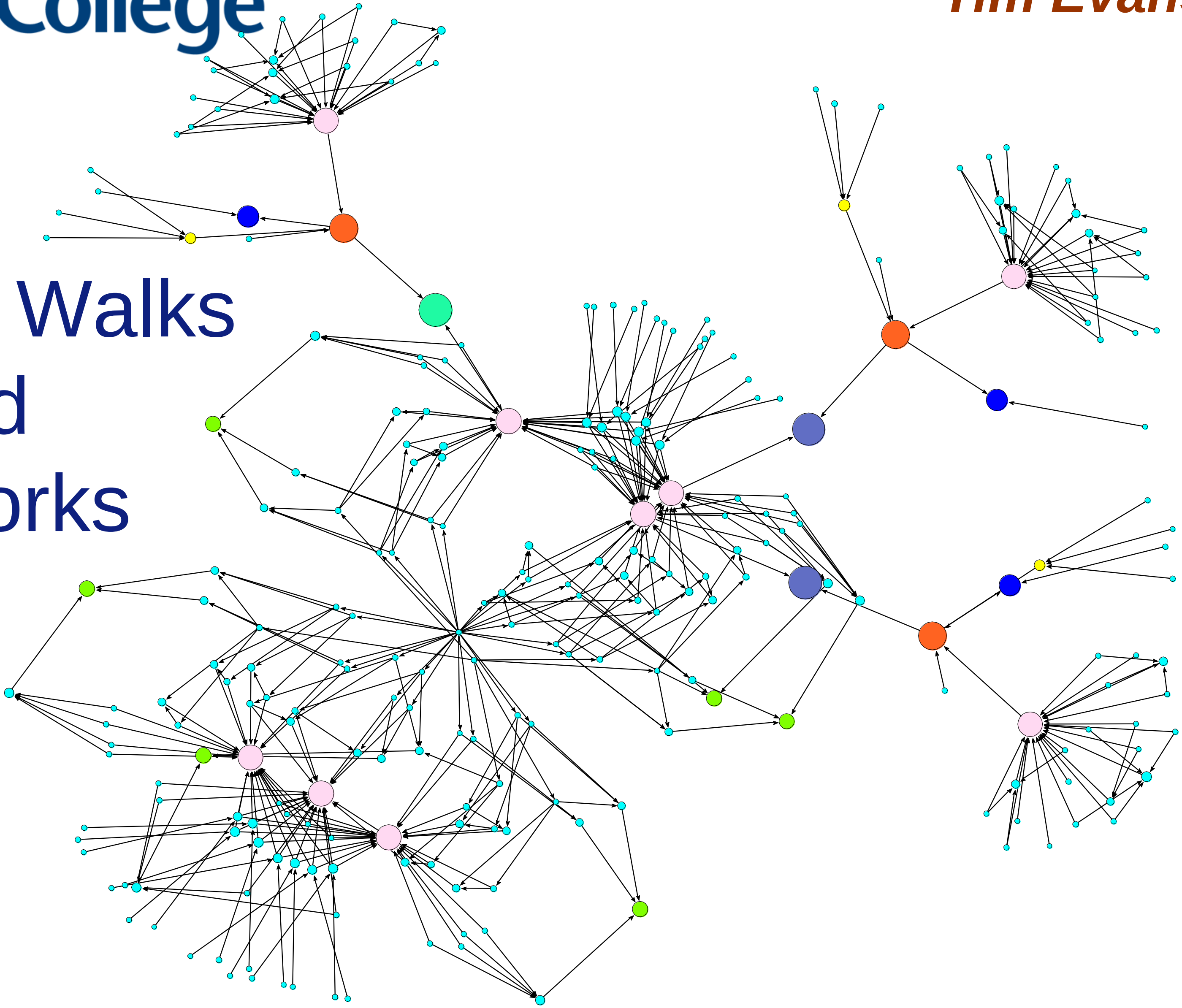
Tim Evans

Imperial College London

Random Walks and Networks



Outline

1. *Notation*
2. *Random Walks for Model Building*
3. *Random Walks and Community Detection*
   - *Modularity and Vertex Communities*
   - *Overlapping Communities*
4. *Random Walks for Everything*
   1. *Biased Random Walks*
   2. *Random Walks and the Map Equation*
5. *From Random Walks to Entropy*

$$\delta\Delta\prod_{a} x$$

$$z\sqrt[\Delta]{a/\zeta}$$

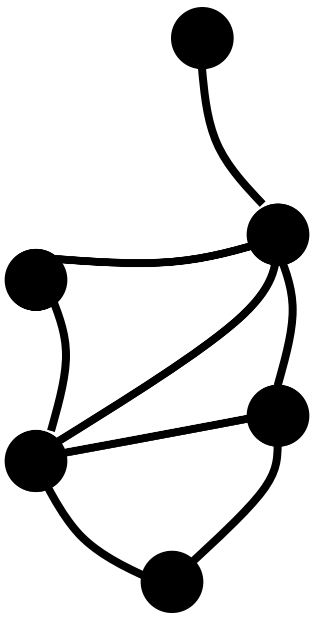
NOTATION

Notation

I will focus on ***Simple Graphs***

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

- ***N*** *= number of vertices in graph*
- ***i,j,..*** *= indices of vertices*
- ***E*** *= number of edges in graph*
- $\alpha, \beta, \ldots$ = for edge indices



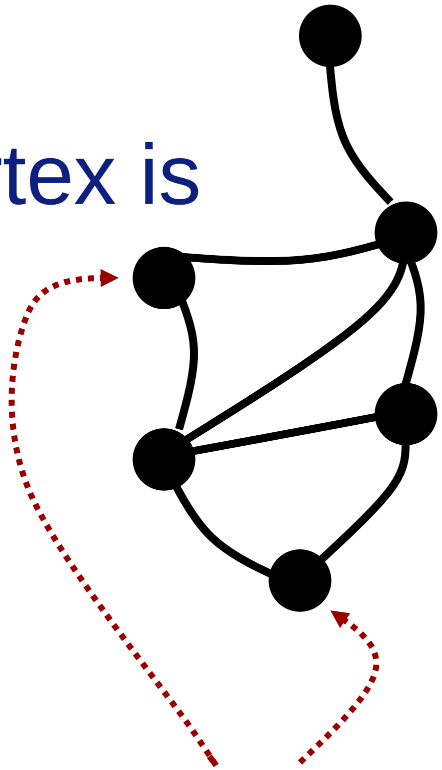
N=6

E=8

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)***



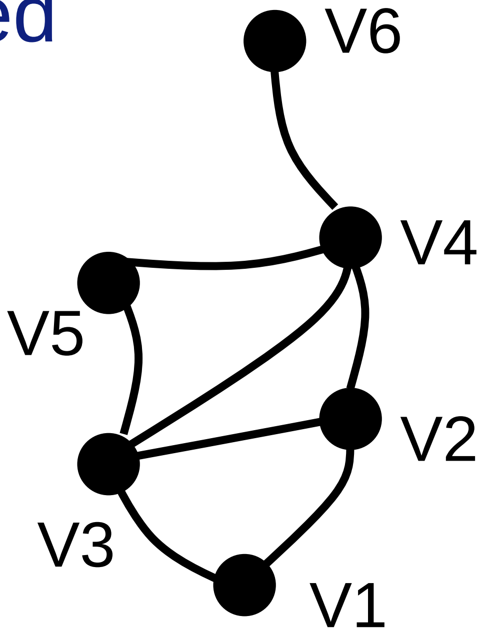
Degree Distribution

- ***n(k)*** = *number of vertices with degree* ***k***
- ***p(k)*** = ***n(k)/N*** = *probability a random vertex has degree* ***k***

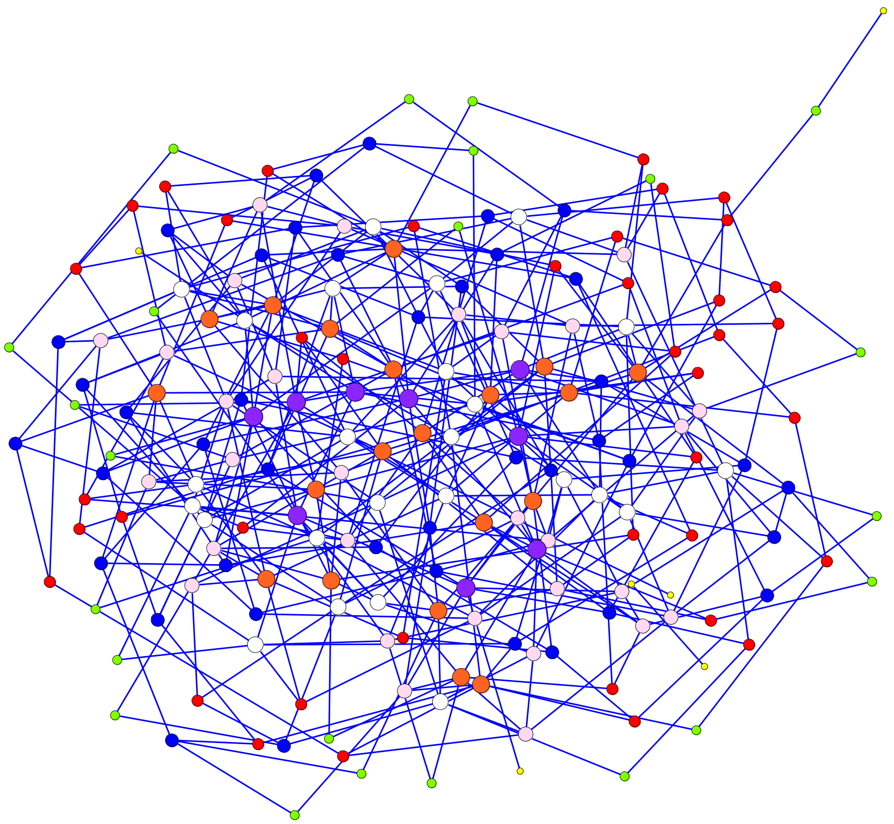
Notation - Adjacency Matrix

The ***Adjacency Matrix*** 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



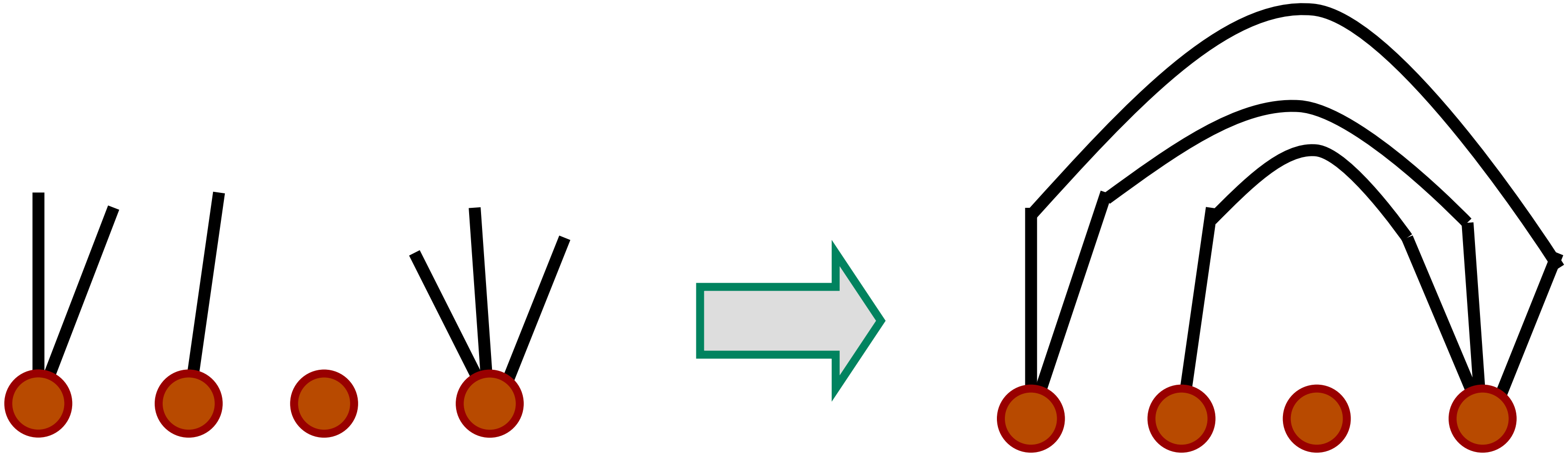
RANDOM WALKS FOR MODEL BUILDING

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

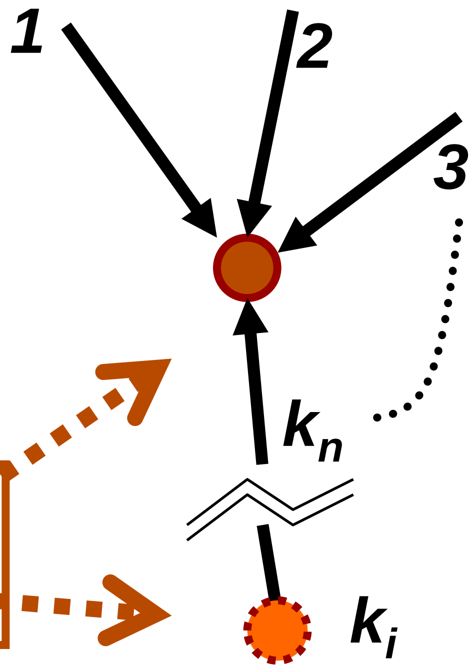
Random Walks on Random Graphs

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

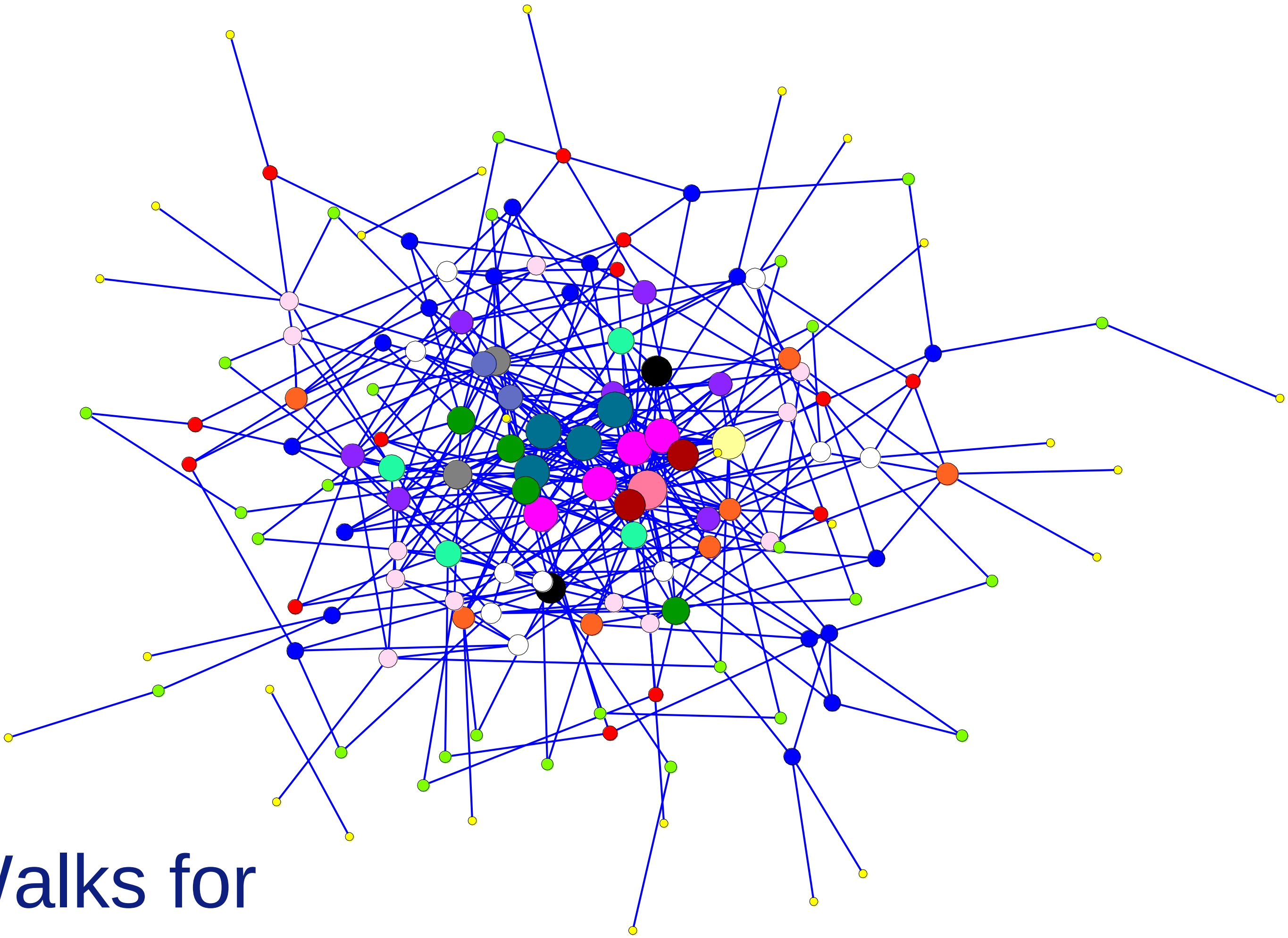
$$p(k_n \mid k_i) = \frac{k_n}{\langle k \rangle} p(k_n)$$

Degree of neighbour k_n independent of degree of starting point k_i



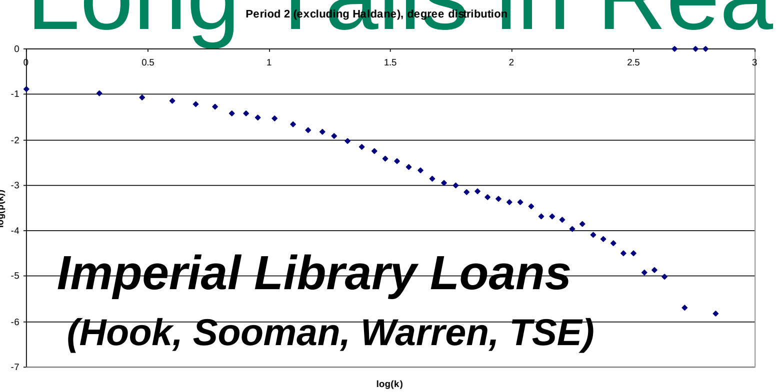
Summary of Generalised Random Graphs

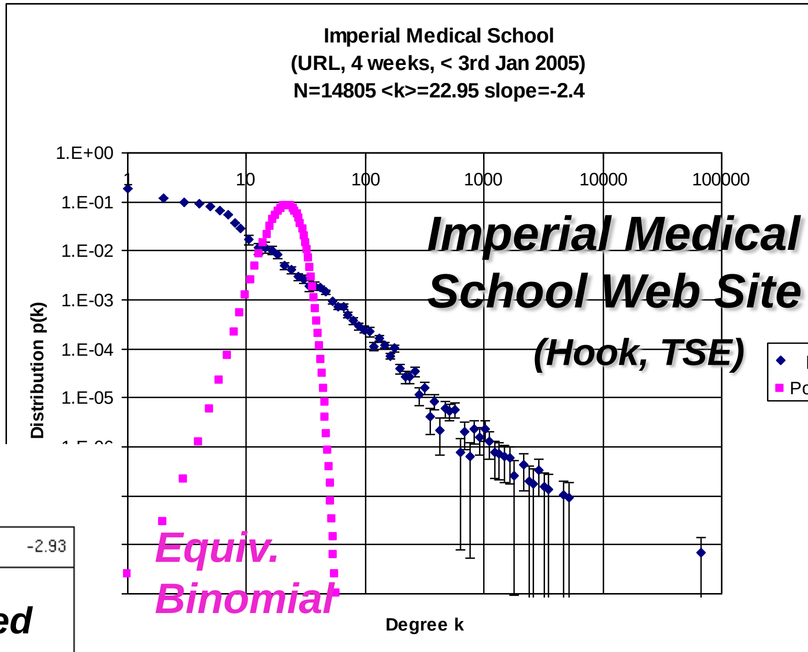
- These can be reasonable approximations for many theoretical models
- Probably not for real world so then use these as a ***null model***.
- Calculations with random graphs work because
  - lack of correlations between vertices
  - few loops for large sparse graphs, graphs are basically trees
- Accessible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

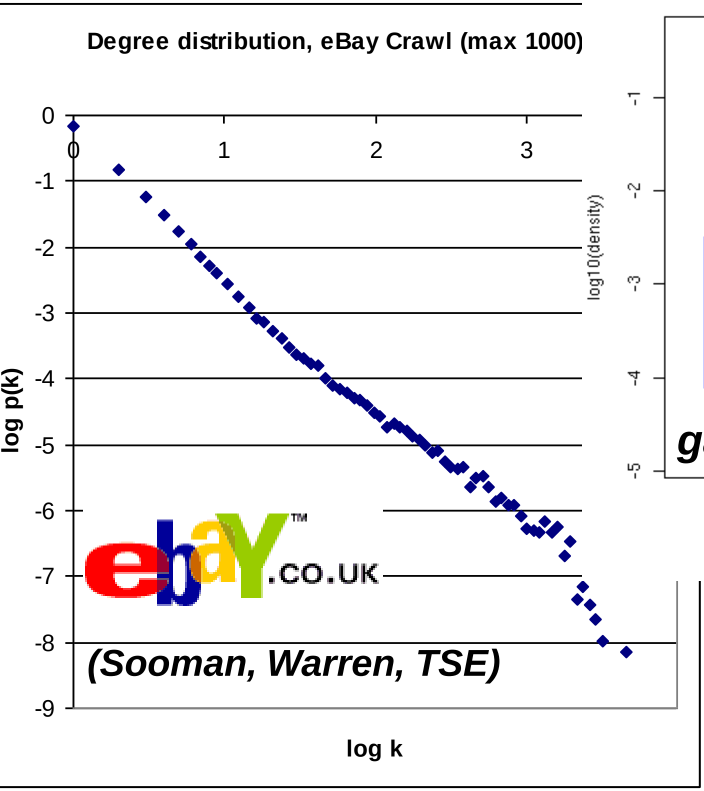


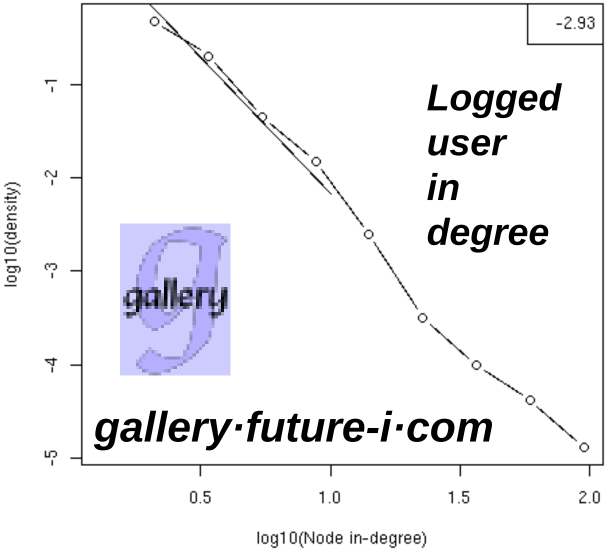
Random Walks for
Natural Scale Free Networks

Long Tails in Real Data

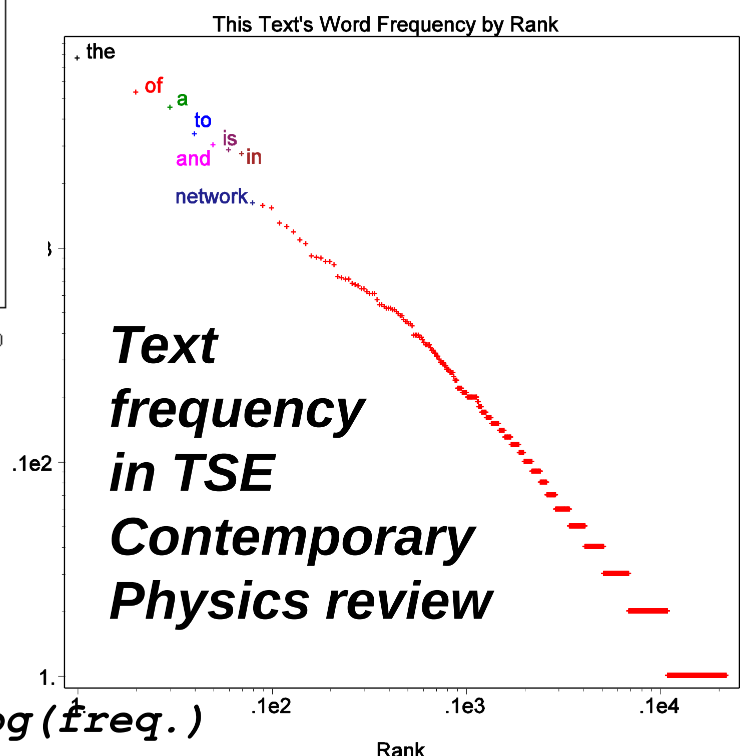








[Yule 1925,1944;
Simon 1955;
Price 1965,1976;
Barabasi, Albert 1999]



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

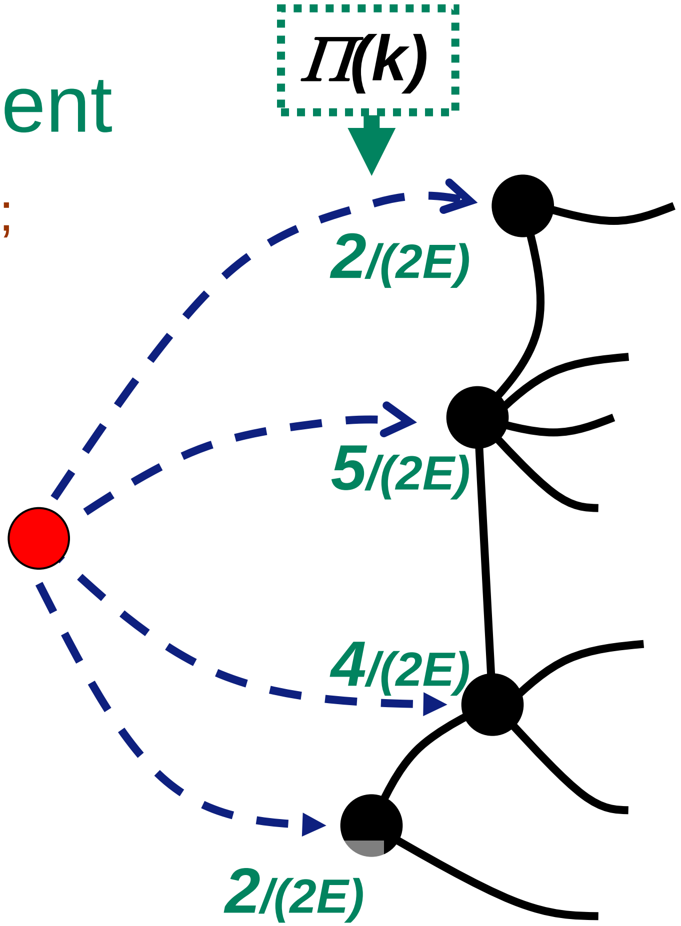
Growth with Preferential Attachment

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

1. Add new vertex attached to one end of **$m=½\langle k\rangle$** 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
$n(k) \sim k^{-\gamma}$
$\gamma=3$

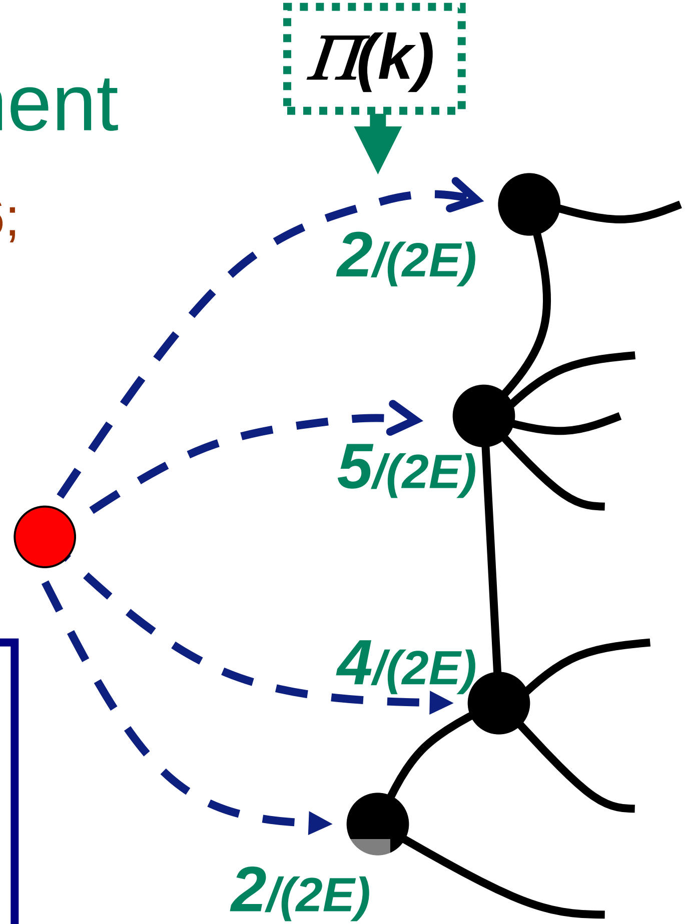
Growth with Preferential Attachment

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

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

Preferential Attachment

"Rich get Richer"



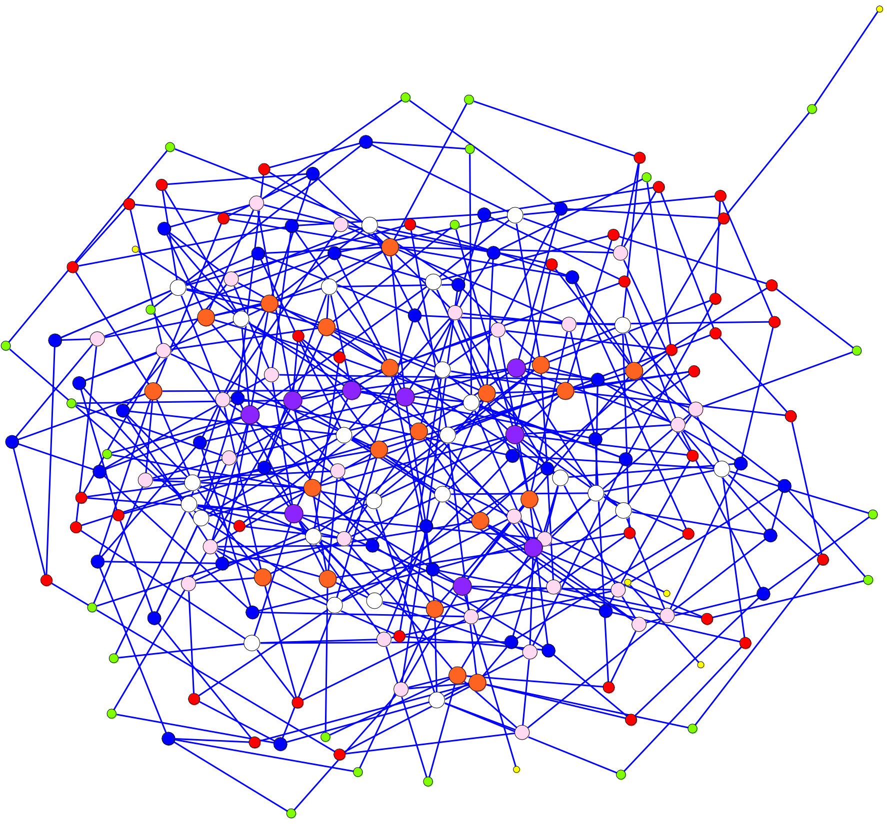
Result: Scale-Free Network

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

$$\gamma = 3$$

N=200, $<k>$~4.0, vertex size ∝ k

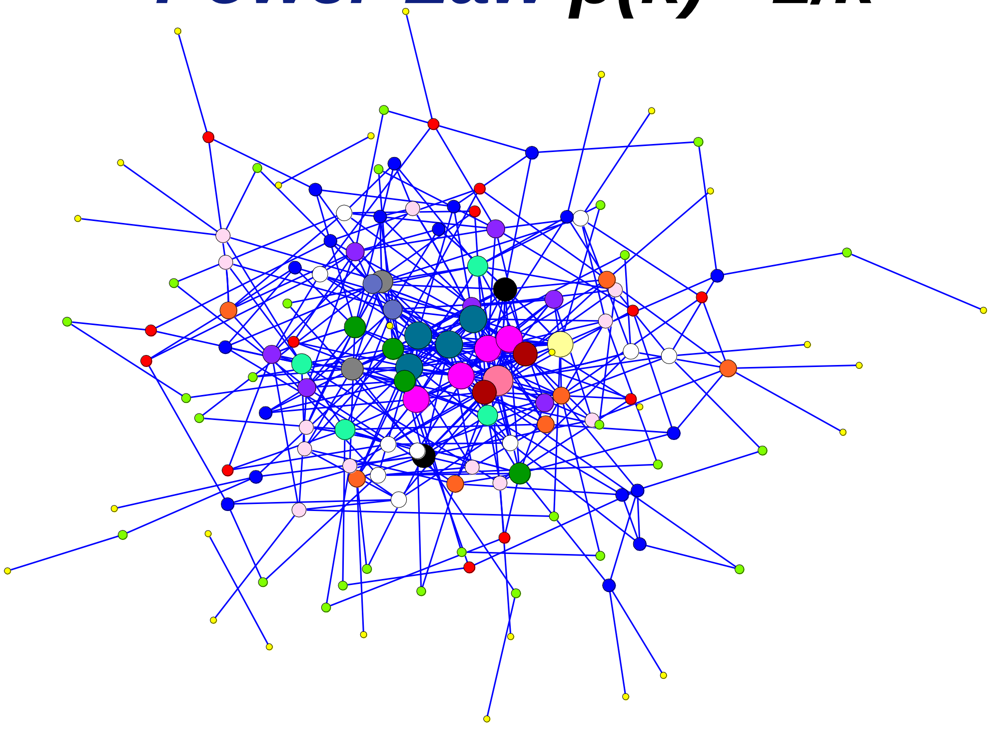
Classical Random



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

Scale-Free

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



Tight core of large hubs

$k_{max}=O(N^{1/2})$

Scale-Free in the Real World

Attachment probability used was

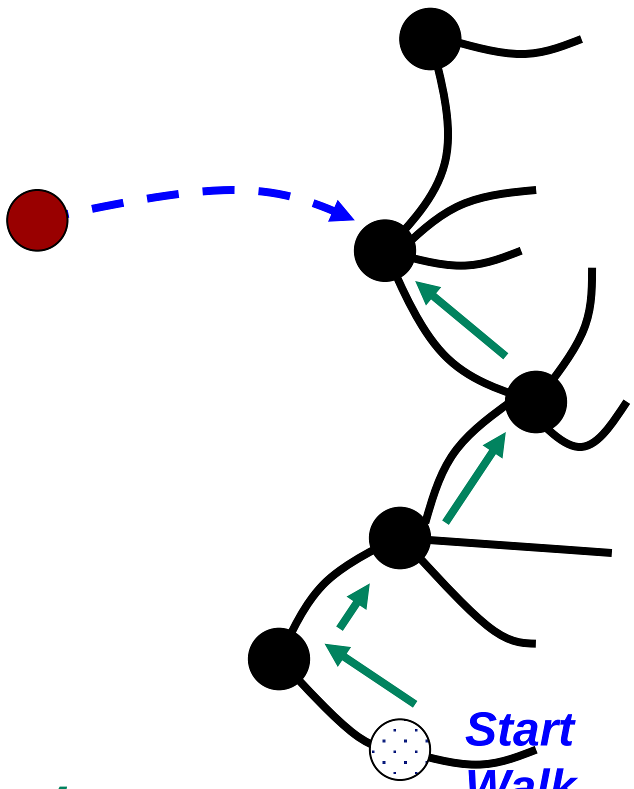
$$\Pi(k) = p_p \frac{k}{2E} + p_r \frac{1}{N},$$

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

⇒ Preferential Attachment ⇒ γ=3

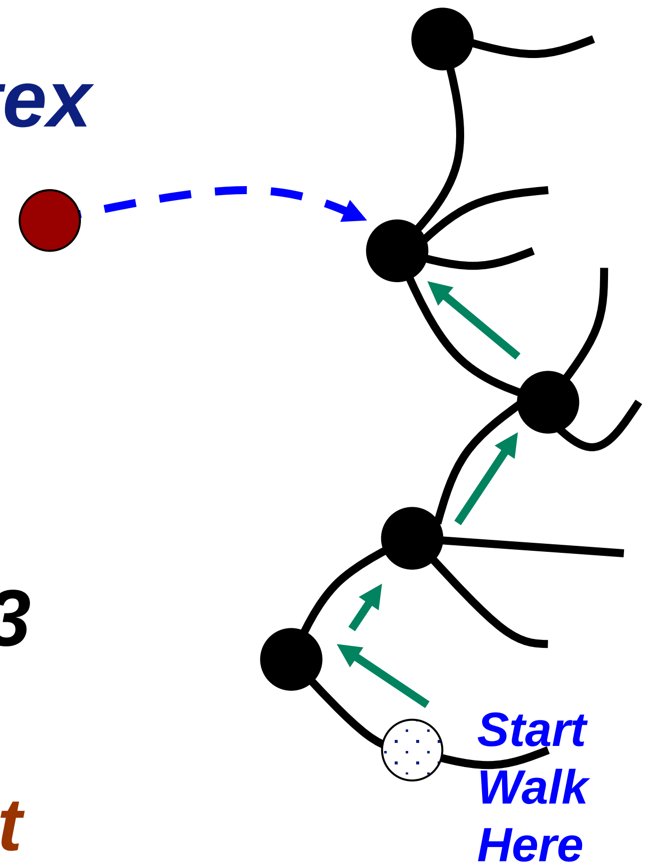
(Can also mix in random attachment with probability p_r)

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

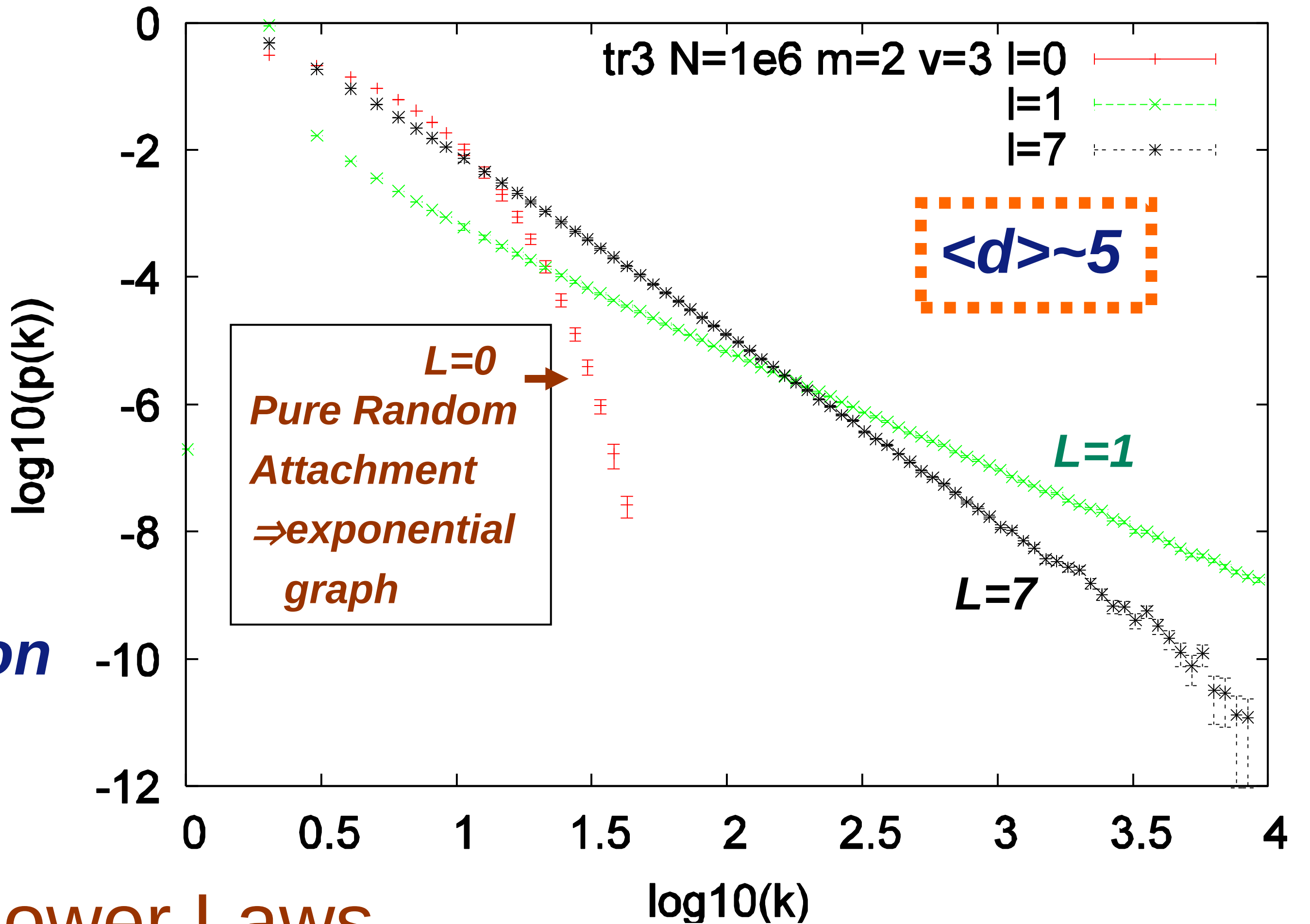
    e.g. informal requests for work on the film actor's social network

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

Is the Walk Algorithm Robust?

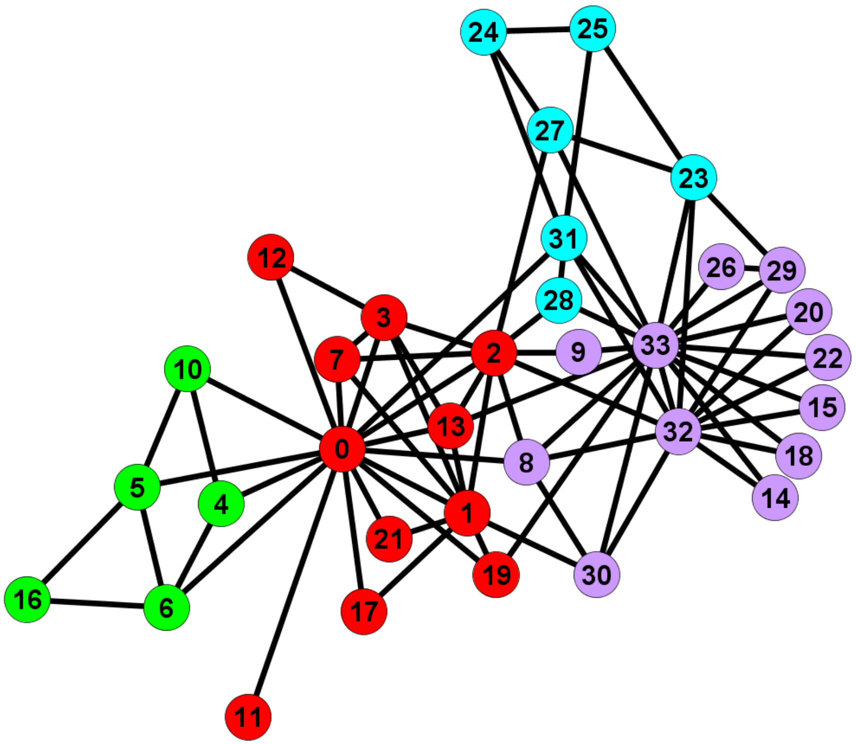
I varied:

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



YES - Good Power Laws

but NOT Universal values - 10% or 20% variation



RANDOM WALKS AND COMMUNITY DETECTION

Vertex Communities

Modularity Q [Girvan & Newman 2002]

1. Assign each vertex ***i*** to community $\boldsymbol{c_i}$ and for each community:-
2. Add the fraction of edges IN community $\boldsymbol{c_i}$
3. Subtract the number of edges in null model
   - generalised random graph with same degree distribution

$$Q(\{c_k\}) = \sum_{i,j} \left(\underbrace{\frac{A_{ij}}{W}}_{2} - \underbrace{\frac{k_i}{W}\frac{k_j}{W}}_{3} \right) \underbrace{\delta_{c_i, c_j}}_{1}$$

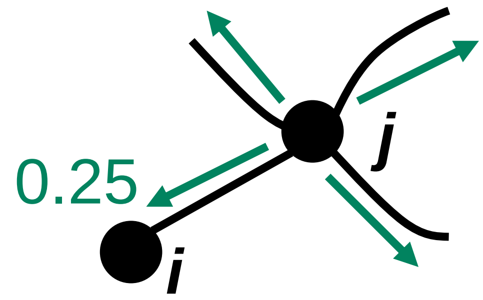
where $W = 2E = \sum_{i,j} A_{ij}$

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 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

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

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

Equilibrium

Equilibrium reached is eigenvector Π_1 with largest eigenvalue as $1 = \lambda_1 > |\lambda_n| \quad \forall\, n > 1$

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

For *simple networks* only we have trivial solution

$$w(t \to \infty)_i = \Pi_i = \frac{k_i}{W}$$

Modularity as Random Walk [Delvenne et al, 2008,2010]

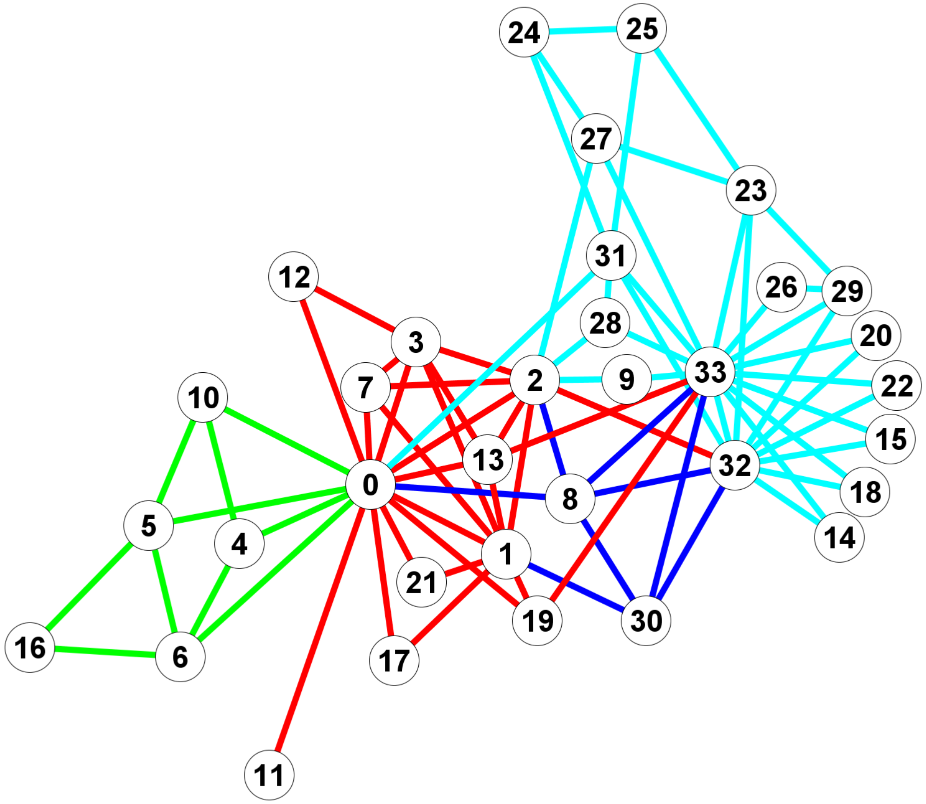
1. Assign each vertex ***i*** to community c_i and for each community:-
2. Add fraction of equilibrium walkers remaining in community after one step
3. Subtract fraction of equilibrium walkers in community after infinite number of steps

$$Q(\{c_k\}) = \sum_{i,j} \left(\underbrace{\frac{A_{ij}}{k_j} \frac{k_j}{W}}_{2} - \underbrace{\frac{k_i}{W} \frac{k_j}{W}}_{3} \right) \underbrace{\delta_{c_i, c_j}}_{1}$$

Modularity for any network

1. Assign each vertex ***i*** to community c_i
   and for each community:-
2. Add fraction of equilibrium walkers remaining in community after one step
3. Subtract fraction of equilibrium walkers in community after infinite number of steps

$$Q(\{c_k\}) = \sum_{i,j} \left(\underbrace{\frac{A_{ij}}{k_j^{out}}}_{2} \Pi_j - \underbrace{\Pi_i \Pi_j}_{3} \right) \underbrace{\delta_{c_i,c_j}}_{1}$$



RANDOM WALKS AND COMMUNITY DETECTION

Overlapping Communities

Vertex Centric Viewpoint

A network is

1. a set of vertices

AND

2. a set of edges

We tend to have a very VERTEX centred viewpoint

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

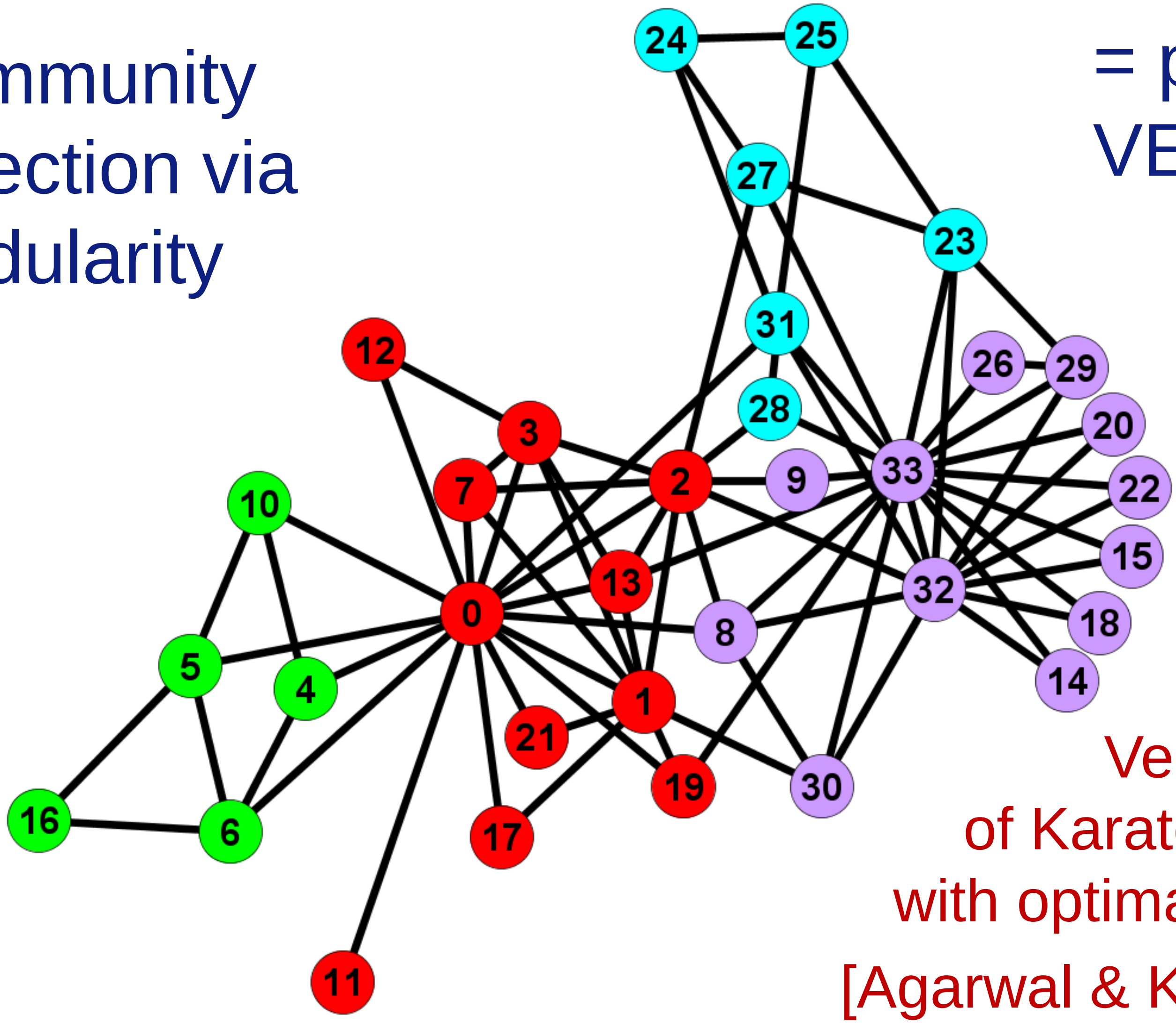
[TSE, Contemporary Physics 2004]

Stop words removed, stemmed, from T.S.Evans “Complex Networks” Contemporary Physics, 2004, **45**, 455-474, cond-mat/0405123

Vertex Centric Communities – Vertex Partitions

Community detection via modularity

= partition of VERTEX set

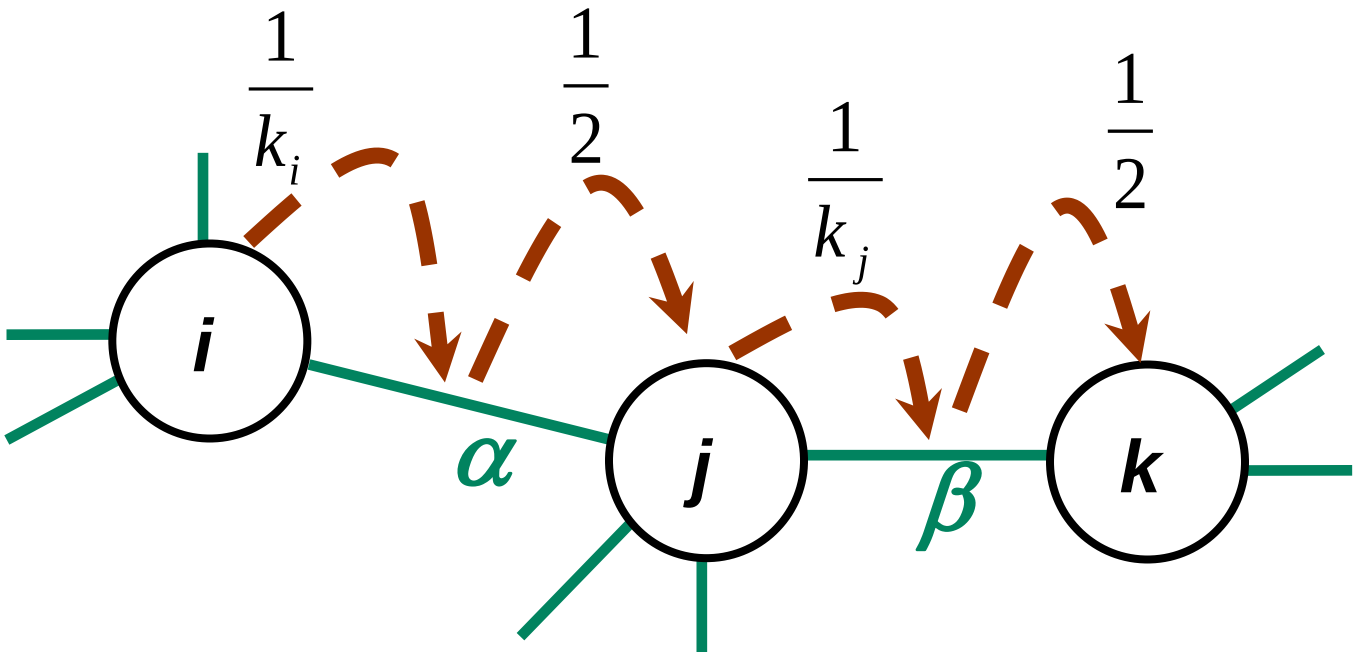


Vertex partition of Karate club graph with optimal modularity

[Agarwal & Kempe 2007]

Random walk on edges

Consider how random walkers pass through edges

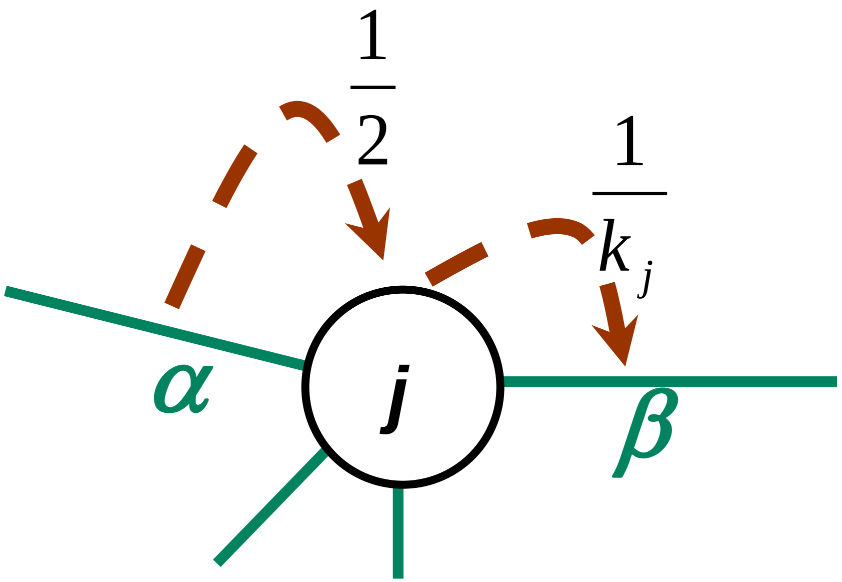


Using α, β, … for edge indices

Random walk on edges

Edge to Edge transition matrix $\boldsymbol{T}_{\alpha\beta}$ is just

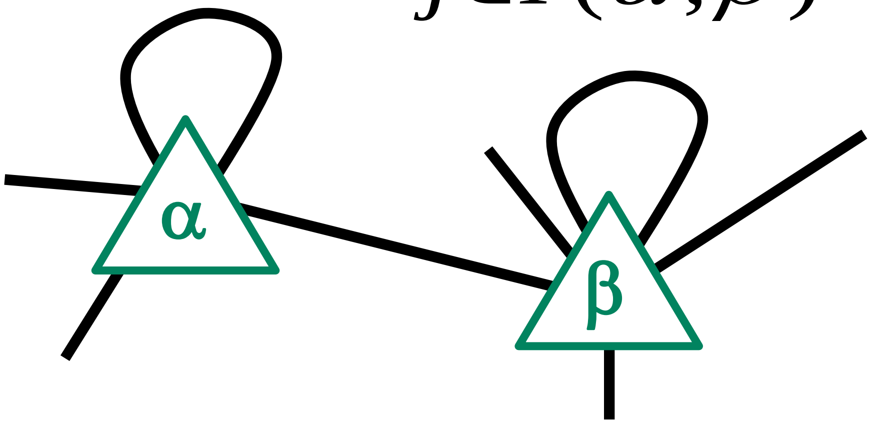
$$T_{\alpha\beta} = \overbrace{\frac{1}{2}}^{\alpha\to j}\overbrace{\frac{1}{k_j}}^{j\to\beta}$$



Random walk on edges

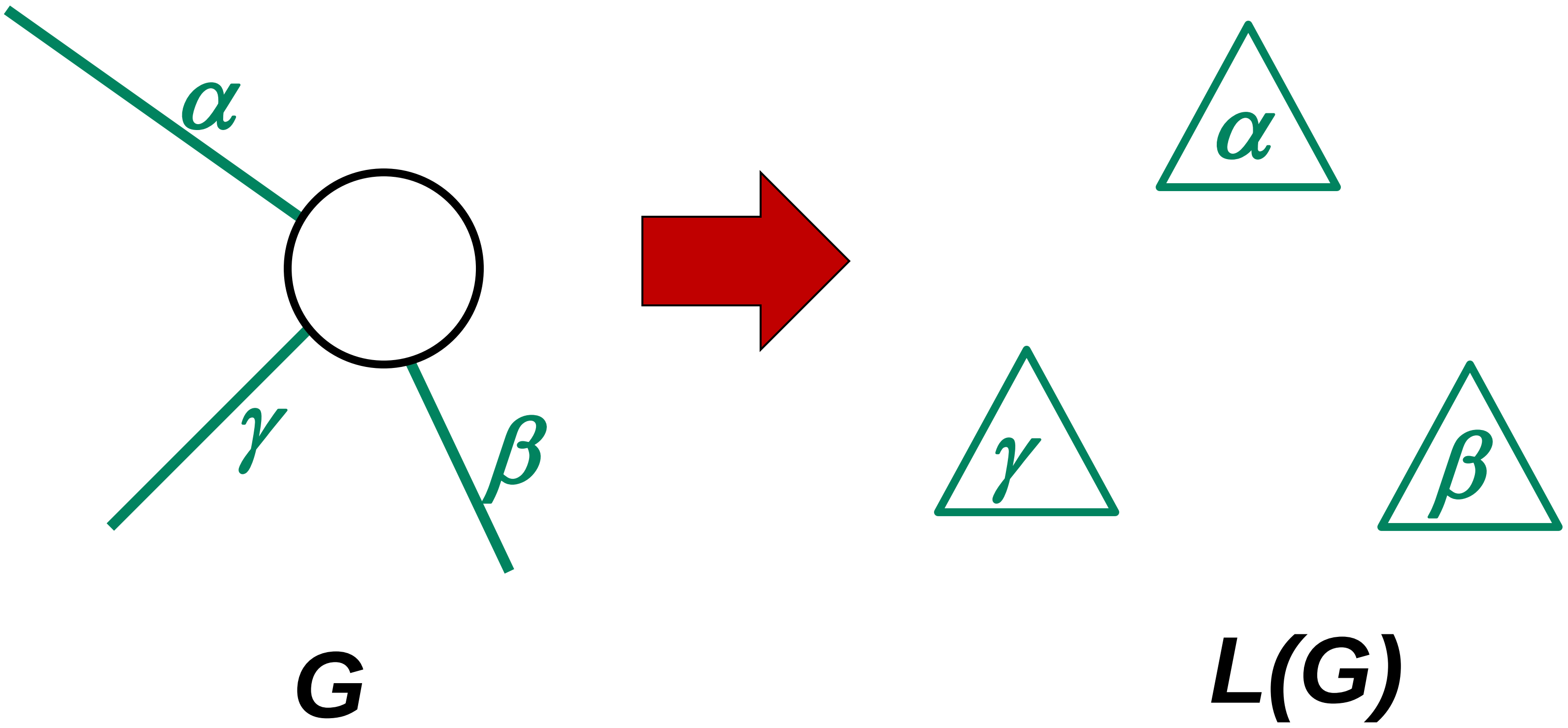
Edge-Edge transition matrix $T_{\alpha\beta}$ defines an adjacency matrix of a Weighted Line Graph ***WL(G)***

$$A_{\alpha\beta} = \sum_{j \in I(\alpha,\beta)} \frac{1}{2}\frac{1}{k_j}$$



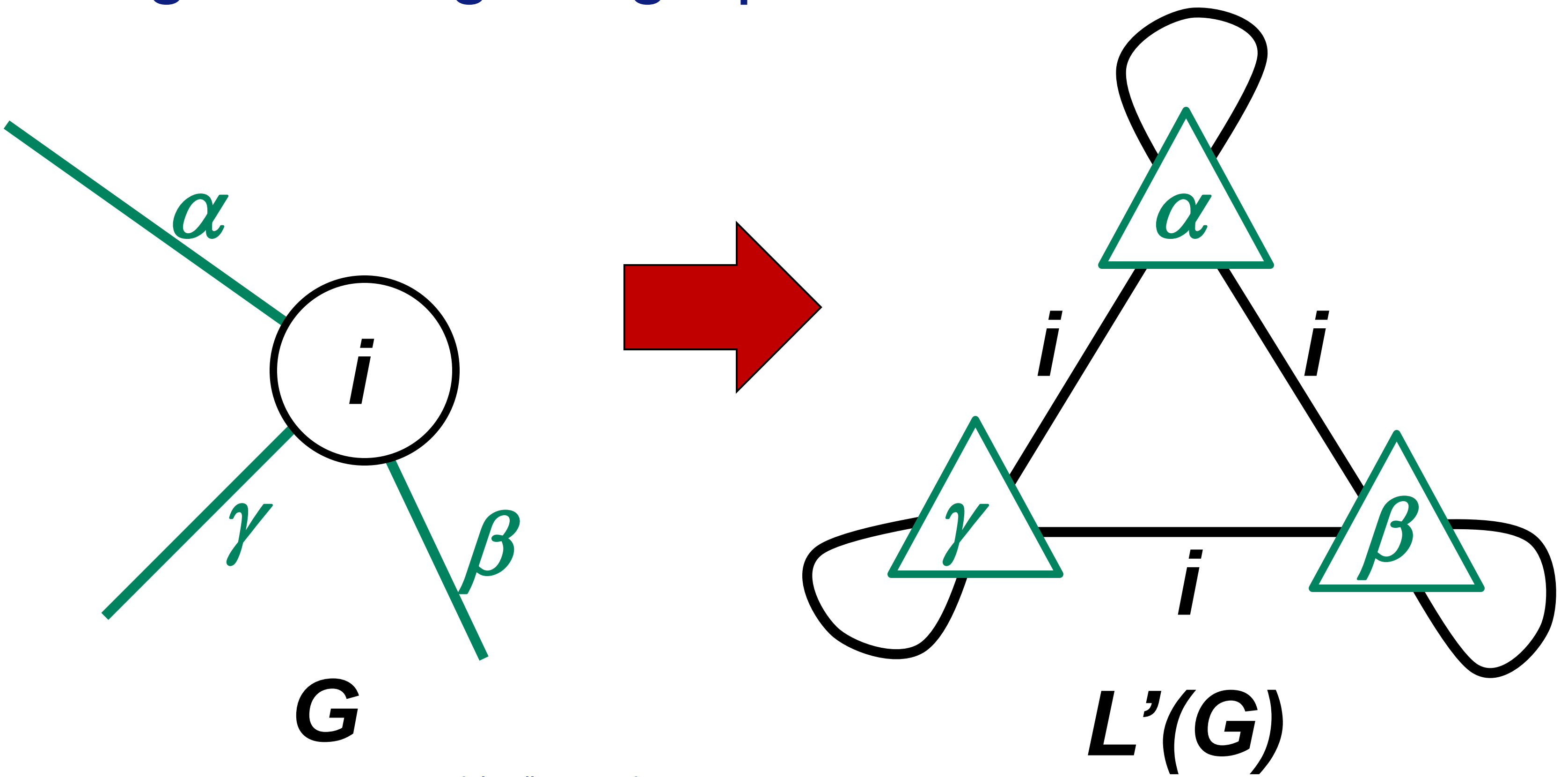
Vertices of a Line Graph

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



Edges of a Line Graph

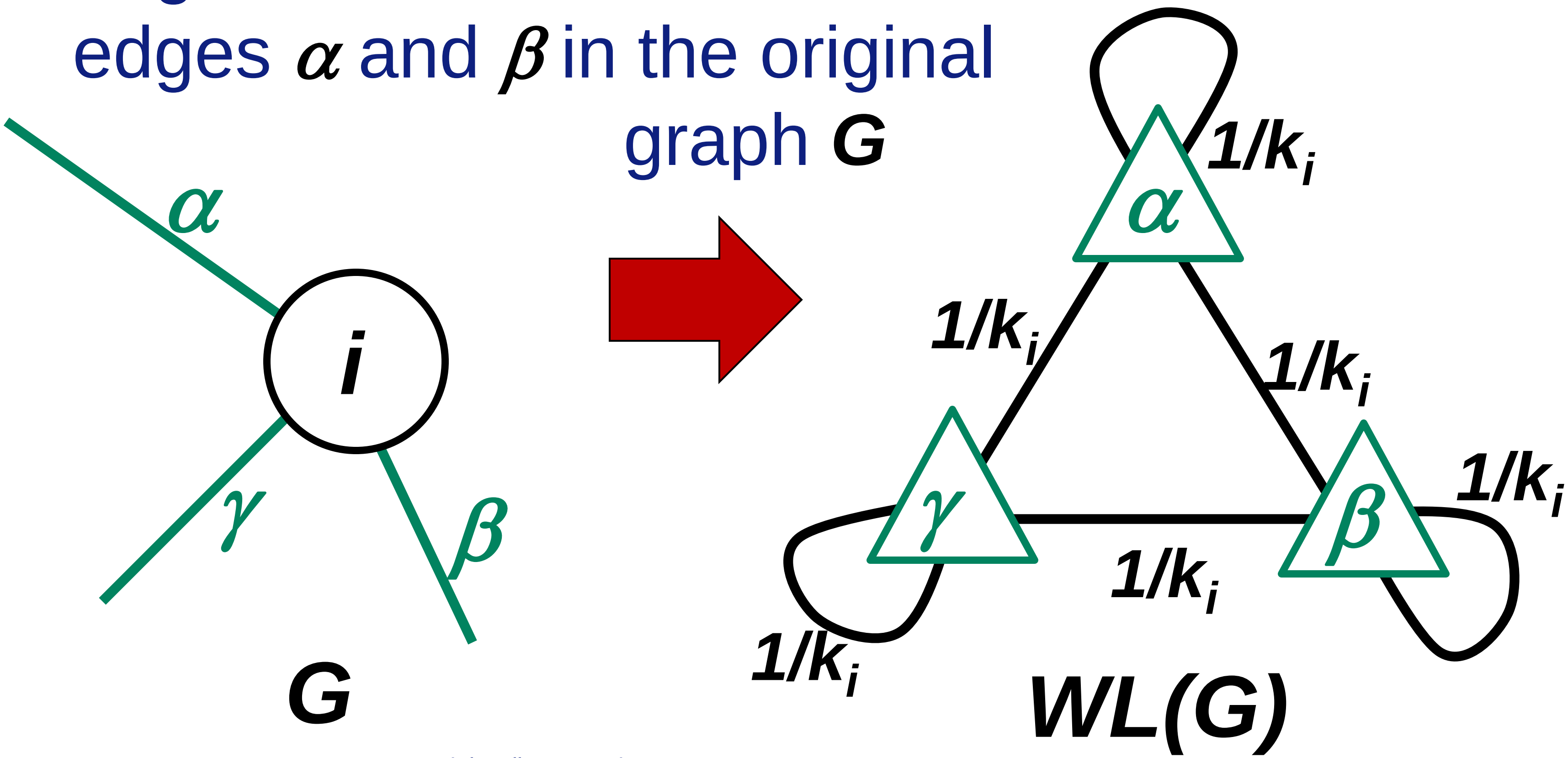
2. Connect the vertices α and β in the ***Line graph*** **L'(G)** if the corresponding edges in original graph ***G*** were coincident



Weights of a Weighted Line Graph

[Evans, Lambiotte PRE, 2009]

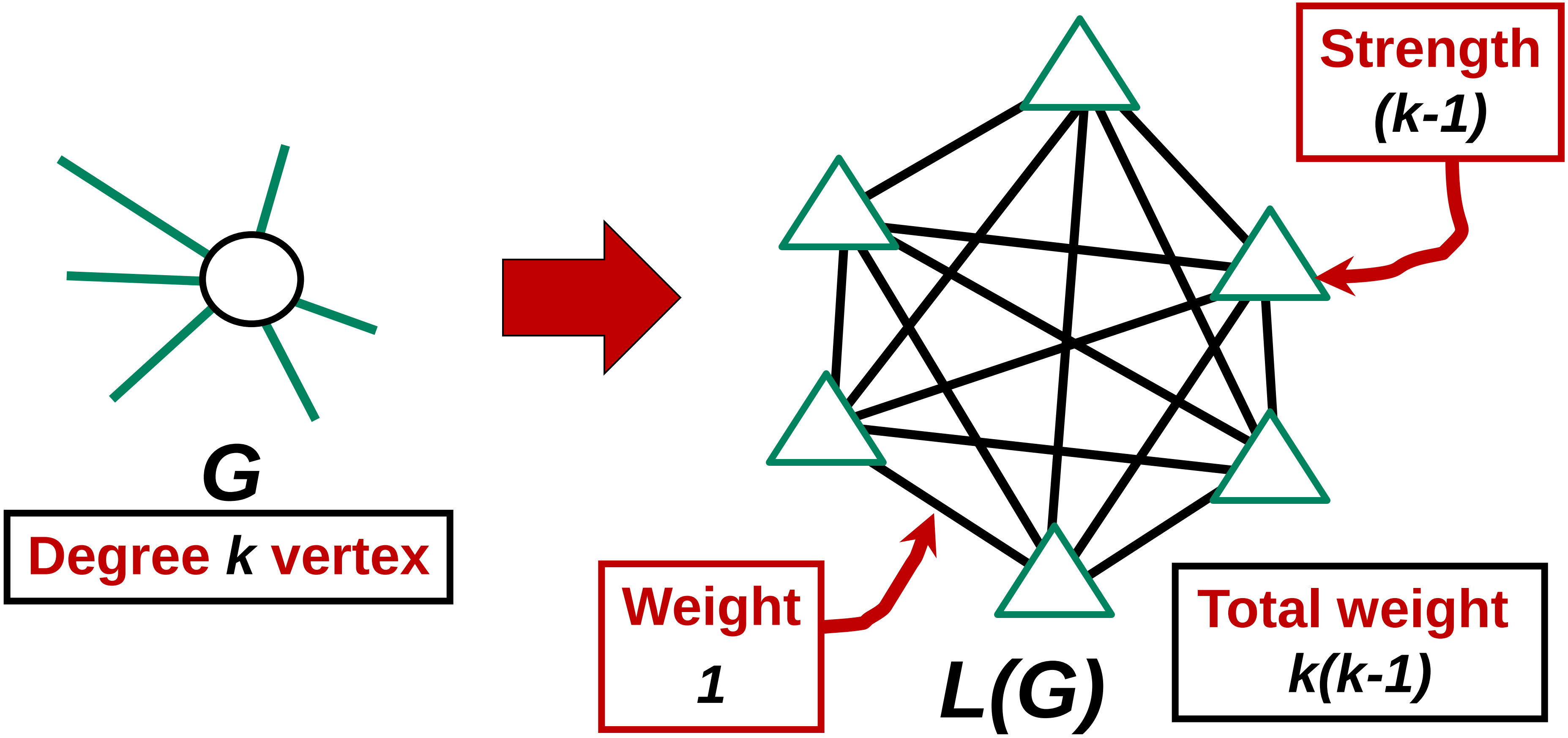
3. Weight the edge between the line graph vertices α and β by the inverse of the degree of the vertex coincident on both edges α and β in the original graph ***G***



[Evans & Lambiotte PRE 2009]

Traditional Line Graph Problem

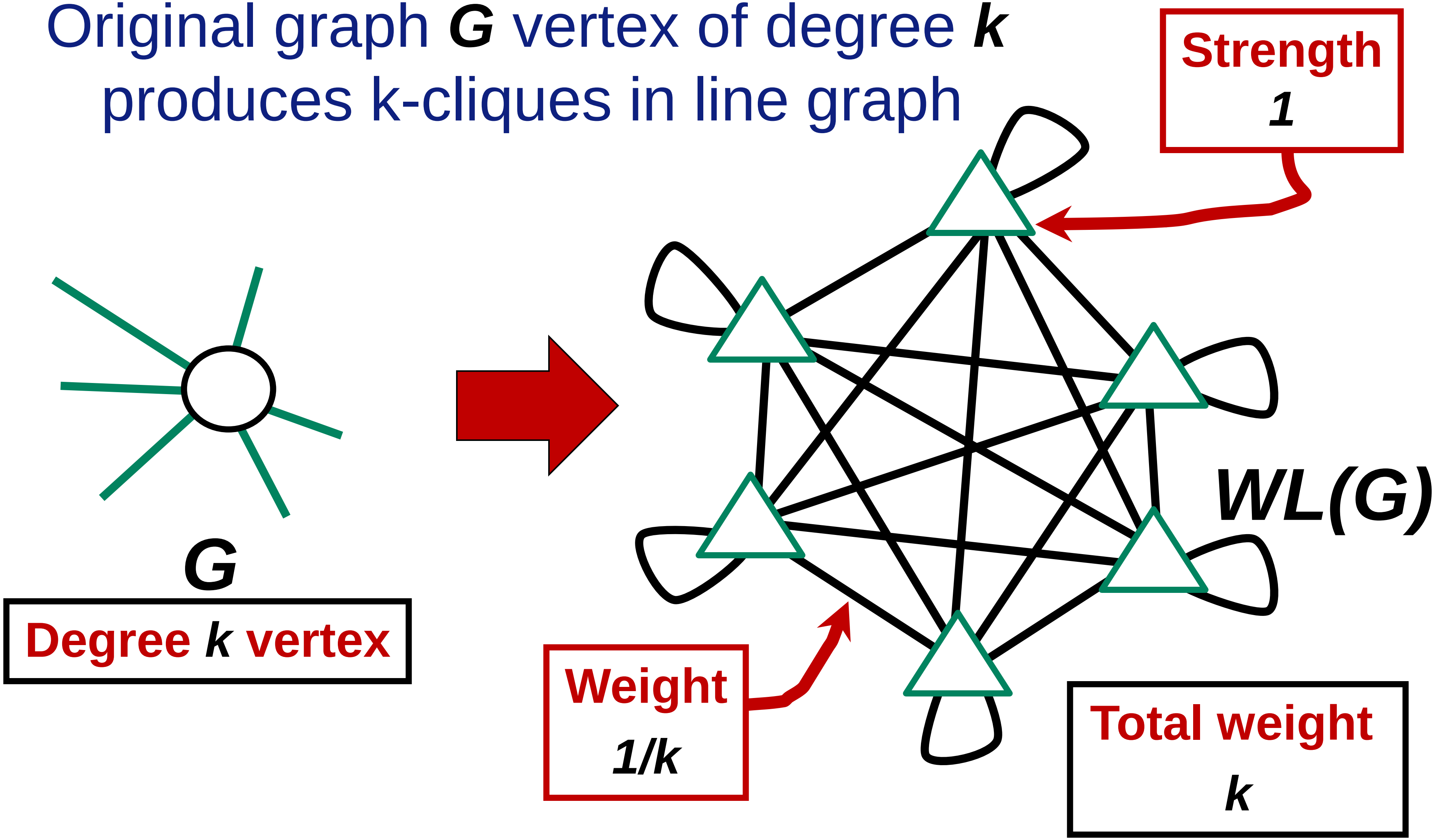
Original graph ***G*** vertex of degree ***k*** produces k-cliques in line graph



[Evans & Lambiotte, PRE 2009]

Traditional Line Graph Problem

Original graph *G* vertex of degree *k* produces k-cliques in line graph



Weighted Line Graph and Random Walks

[Evans & Lambiotte 2009]

Edges ***G*** ⟵ Random Walk ⟶ Vertices ***WL(G)***

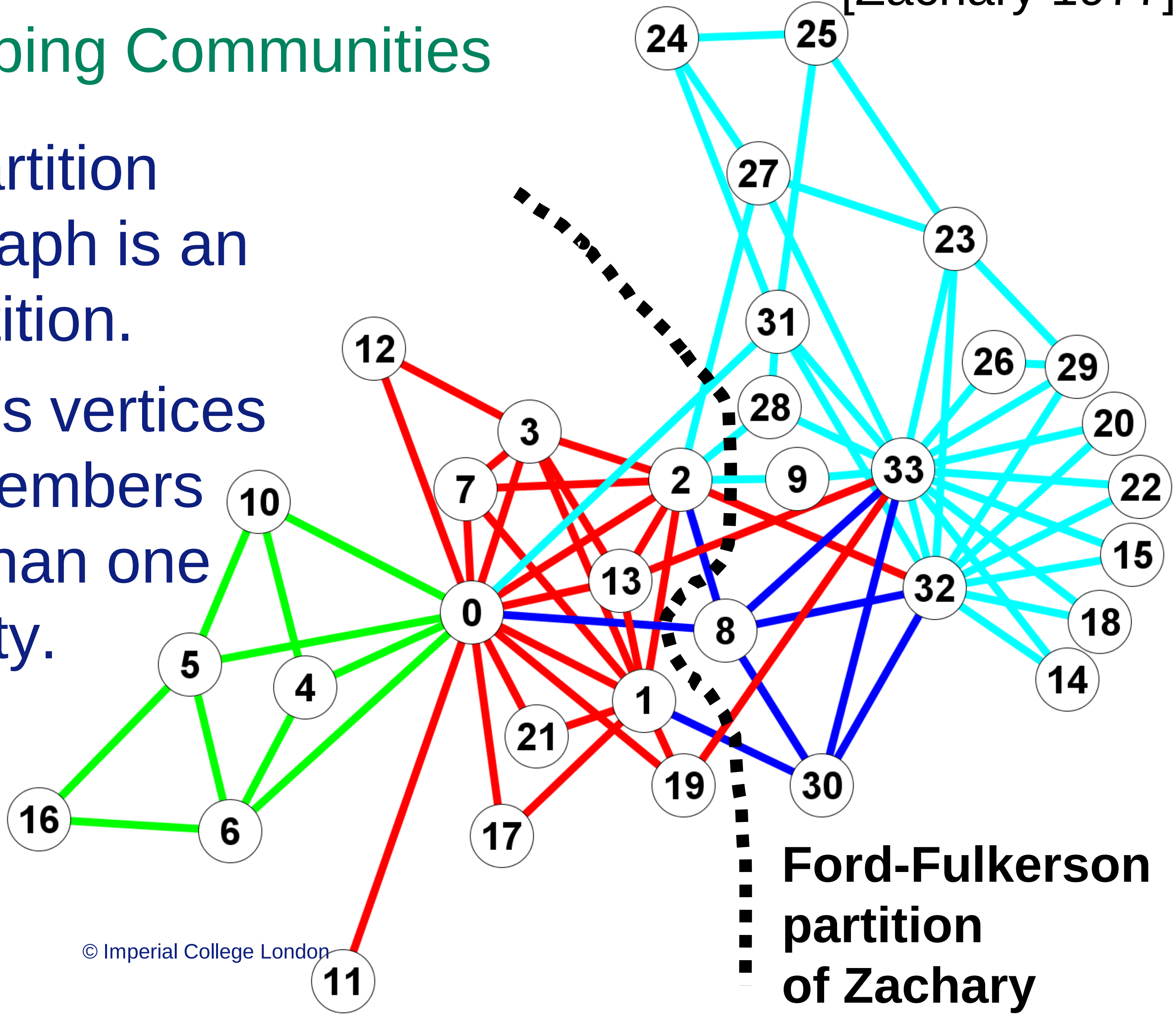
- Simple Random walk process on original graph ***G*** is reproduced ***exactly*** on Weighted Line graph ***WL(G)***
- Any vertex analysis tool using random walks can be used without bias on ***WL(G)*** but now this analyses the ***edges*** of original graph ***G***.
- Variations for slightly different random walks on original graph ***G***.

Zachary's Karate Club
[Zachary 1977]

Application:-
Overlapping Communities

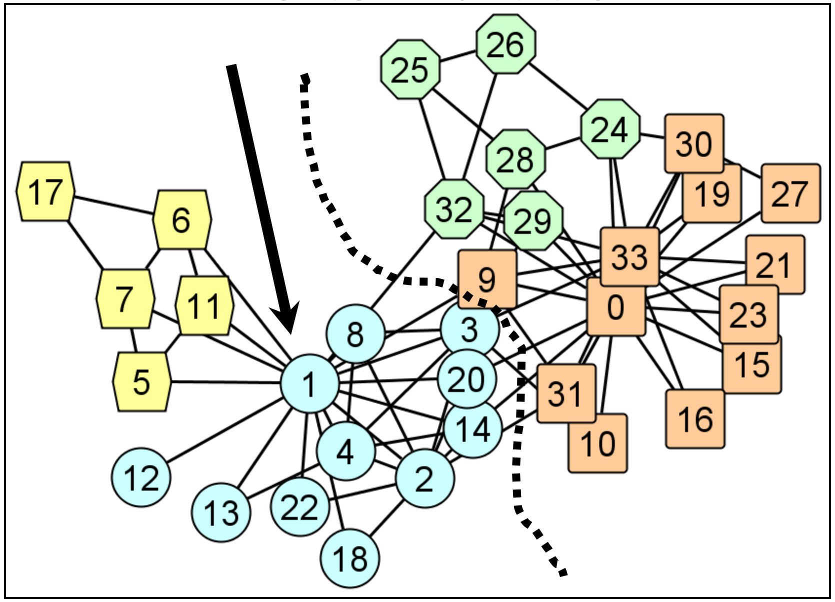
Vertex partition on line graph is an edge partition.

Individuals vertices can be members of more than one community.

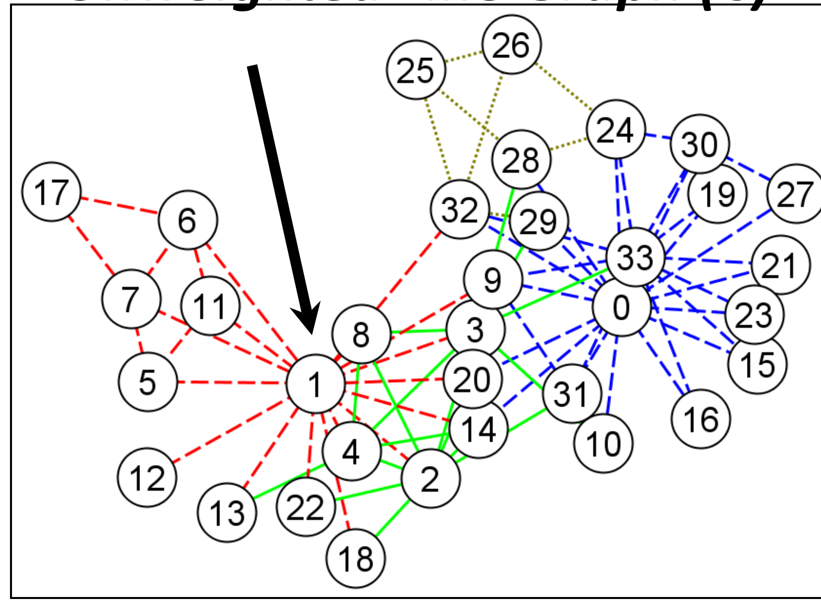


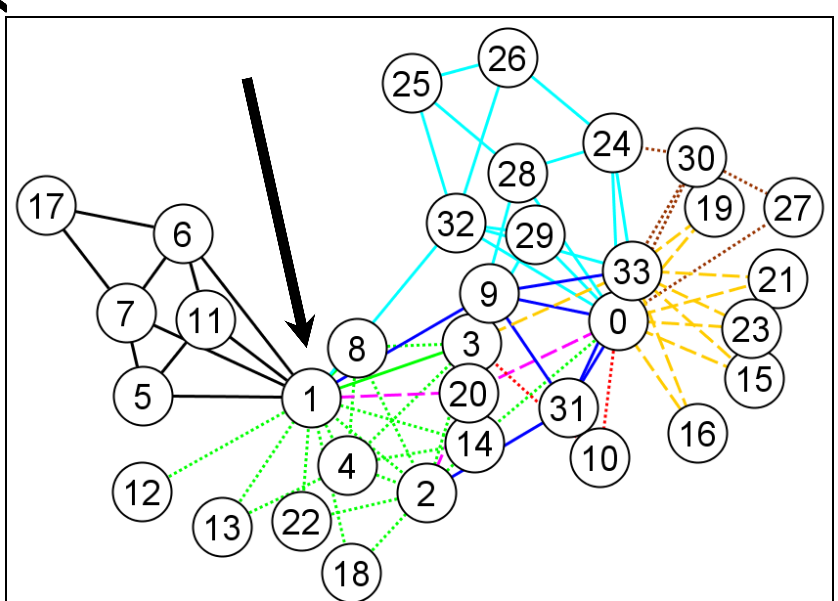
Zachary Karate Club

Vertex Partition

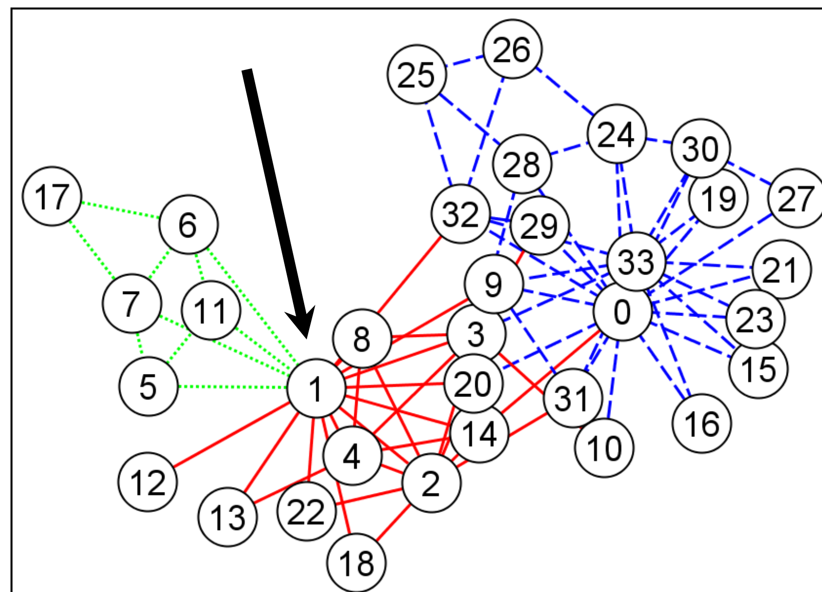


Unweighted Line Graph (C)





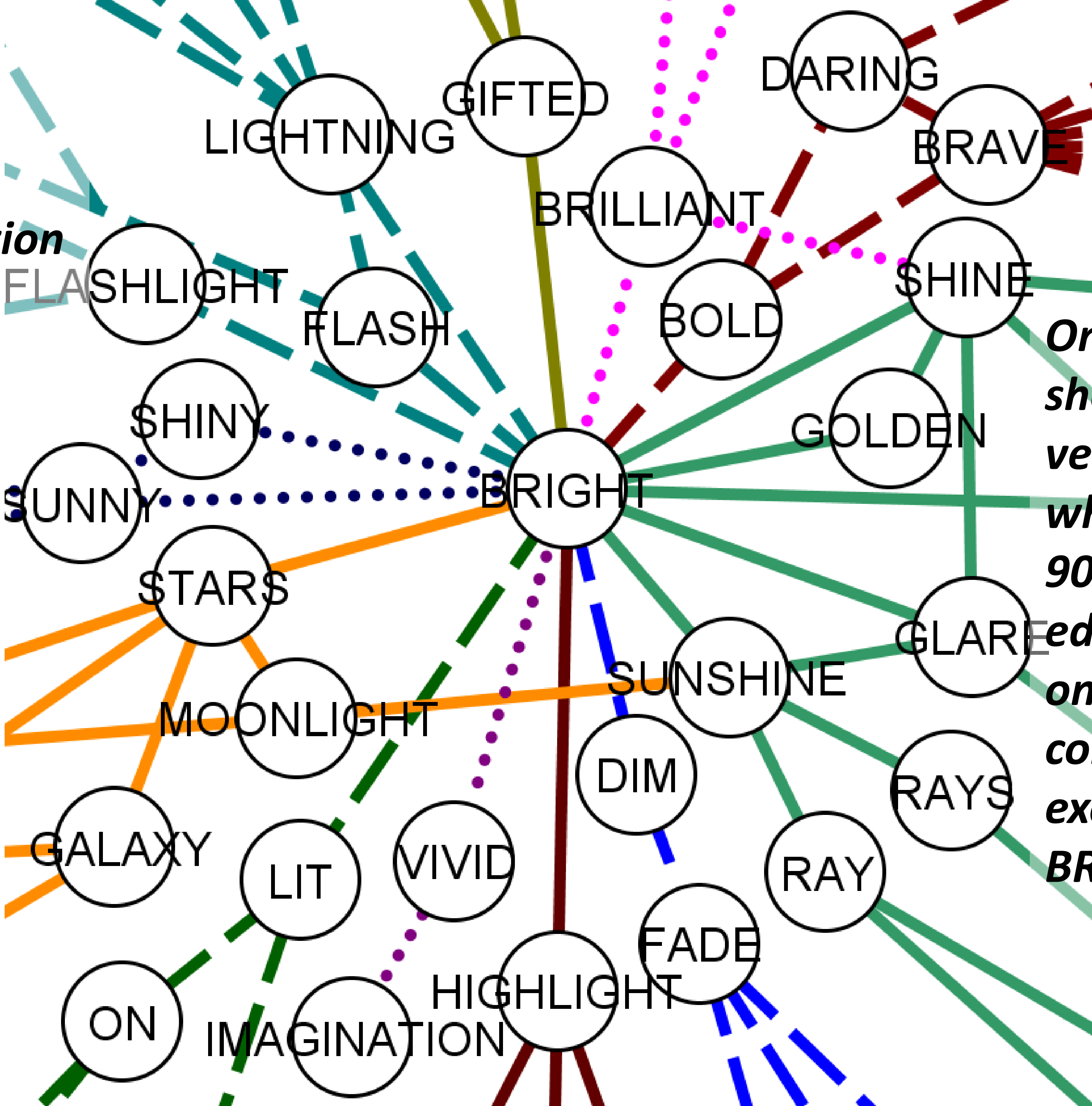
Weighted Line graph (D)



Weighted Line Graph (E_1)

[Evans & Lambiotte, 2009]

South Florida Word Association Data



Only showing vertices which have 90% of edges in one edge community except for BRIGHT

[TSE, Unpublished]

Edge Partition of Word in Paper Titles

- Some words have all edges in one partition
  – they define these communities e.g. cassini
- Other words have edges in several communities
  – stop words e.g. signature

Stem	Total k	k in C
interplanetari	78	78
cassini	62	62
heliospher	59	59
magnetopaus	53	53
spacecraft	52	52
signatur	91	32
solitari	30	10
radar	21	7
mhd	18	6

Weighted Line Graph and Random Walks

- Variations for slightly different random walks

  [Evans & Lambiotte, PRE 2009]
- Generalisation to any original graph ***G*** including weighted and directed

  [Evans & Lambiotte, EPJB 2010]
- Extensions to work in terms of overlap of units other than edges
  e.g. triads=triangles for social networks

  [Evans, J.Stat.Mech 2010]

RANDOM WALKS FOR EVERYTHING

Biased Random Walks

Not all random walks treat all vertices equally in a simple graph.

Consider a bias where probability of a random walker visiting vertex ***i*** in ***G*** is proportional to some bias factor b_i

$$T_{ij} = \frac{b_i A_{ij}}{Z_j} \qquad Z_j = \sum_i b_i A_{ij}$$

Biased Random Walks

Vertices are not identical on a simple graph.

Consider a bias where probability of a random walker visiting vertex $\boldsymbol{i}$ in $\boldsymbol{G}$ is proportional to some bias factor $\boldsymbol{b_i}$

Examples:-

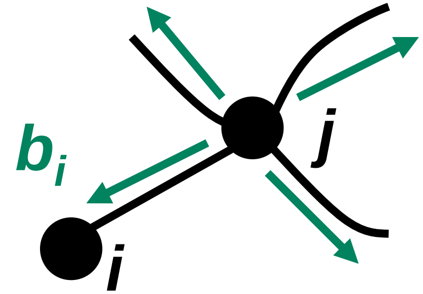
- $\boldsymbol{b_i}$ = (centrality measure of vertex $\boldsymbol{i}$)$^{\gamma}$
- $\boldsymbol{b_i}$ = (degree of vertex $\boldsymbol{i}$)$^{\gamma}$
- $\boldsymbol{b_i}$ = page rank of vertex $\boldsymbol{i}$

Biased Random Walk Transition Matrix

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

i ← j

$$T_{ij} = \frac{b_i A_{ij}}{Z_j}$$



with *normalisation*

$$Z_j = \sum_i b_i A_{ij}$$

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

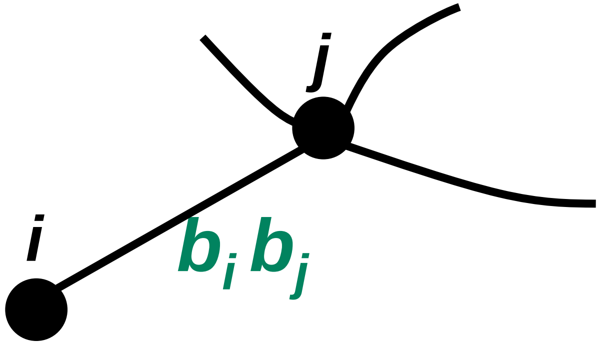
Biased Random Walk as Unbiased Walk

[Lambiotte et al, PRE 2011]

The biased random walk on graph ***G*** is an unbiased random walk on a ***flow graph*** ***F(G)*** whose adjacency matrix is

i ← j

$$F_{ij} = b_i A_{ij} b_j$$



- If ***G*** is symmetric then so is ***F(G)***

Random Walks and Other Network Tools

- Page Rank [Brin & Page 1998]
- Betweenness centrality based on electric current analogy [Newman 2005]
- Map equation (infomap) approach to community detection
  [Rosvall & Bergstrom 2008]

PageRank for Mathematicians

[Clarke, Hopkins, MSci Theses 2010]

Using bibliographies of over 2000 mathematicians we find a physicist is the best mathematician

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

N.B. Also Mark Newman is the “mathematician” with most recent citations

RANDOM WALKS AND ENTROPY

From Random Walks to Entropy

Find the adjacency matrix $\boldsymbol{F_{ij}}$ of null models by maximising Entropy constrained by given minimum information

$$S = \sum_{i,j} F_{ij}\left(\ln\left(F_{ij}\right) - 1\right)$$

$$+ \sum_{i} \alpha_i \left(u_i - \sum_{j} \left(F_{ij}\right) \right) + \sum_{j} \beta_j \left(v_i - \sum_{i} \left(F_{ij}\right) \right)$$

Flow $\boldsymbol{F_{ij}}$ distributed evenly

Input to vertex $\boldsymbol{i}$ fixed to be $\boldsymbol{u_i}$

Output to vertex $\boldsymbol{j}$ fixed to be $\boldsymbol{v_j}$

Maximum Entropy

Minimise entropy to find most likely configuration

$$\frac{\partial S}{\partial F_{ij}} = 0 \quad \Rightarrow F_{ij} = \frac{u_i v_j}{W}$$

Walkers are spread equally across all *edges* subject to constraints that input and output at each site are given by $\boldsymbol{u_i}$ and $\boldsymbol{v_i}$. Not always realisable as a simple random walk.

Random Walks and Entropy

1. Undirected graph
   $u_i = v_i = k_i$ degree
   $\Rightarrow F_{ij}$ randomised graph same degree distribution
   – usual modularity null model with random walk interpretation

$$F_{ij} = \frac{k_i k_j}{W}$$

Random Walks and Entropy

2. Directed graph

   $u_i = v_i = \Pi_i$ = location of random walkers after infinite number of steps

   $\Rightarrow$ F_{ij} randomised graph same degree distribution

   – good modularity null model, with random walk interpretation

$$F_{ij} = \frac{\Pi_i \Pi_j}{W}$$

Entropy and Spatial Constraints

3. Graph Constrained by Space
   F_{ij} = number of random walkers on edge (i,j) if average number of walkers travelling same distance as i to j is constrained to be equal to that found in the data

Entropy and Spatial Constraints

Provides best null models given minimum information

$$S = \sum_{i,j} F_{ij} \left(\ln\left(\frac{F_{ij}}{N_i N_j} \right) - 1 \right) + \sum_r \gamma_r \left(\sum_{i,j \in r} \left(A_{ij} - F_{ij} \right) \right)$$

Flow $\boldsymbol{F_{ij}}$ distributed evenly between people

Flow between all vertices r apart fixed to be equal to data

i,j* $\in$ *r means distance ***d*** lies in ***r***th interval of space

Entropy and Spatial Constraints

3. Graph Constrained by Space

$$F_{ij} = N_i N_j f\left(d_{ij}\right)$$

d_{ij} = distance between i and j

$$f\left(d\right) = \frac{\sum_{i,j \in r} A_{ij}}{\sum_{i,j \in r} N_i N_j}$$

$i,j \in r$ means distance d lies in r^{th} interval

$\Rightarrow$ Null model of Expert et al, PNAS 2011

THANKS

TSE Bibliography

- Evans, T.S. “Complex Networks”, *Contemporary Physics,* **2004***, 45*, 455-474 [arXiv.org/cond-mat/0405123]
- T.S.Evans, J.P.Saramäki “Scale Free Networks from Self-Organisation “ Phys.Rev.E 72 (2005) 1 [cond-mat/0411390]
- Evans, T.S. & Lambiotte, R., “Line Graphs, Link Partitions and Overlapping Communities”, Phys.Rev.E, 2009, 80, 016105 [arXiv:0903.2181]
- Evans, T.S. & Lambiotte, R. “Line Graphs of Weighted Networks for Overlapping Communities” *European Physical Journal B,* **2010***, 77*, 265–272 [arXiv:0912.4389]
- Evans, T.S. “Clique Graphs and Overlapping Communities”, J.Stat.Mech, 2010, P12037 [arxiv.org:1009.063]
- Lambiotte, R.; Sinatra, R.; Delvenne, J.-C.; Evans, T.S.; Barahona, M. & Latora, V. “Flow graphs: interweaving dynamics and structure” *Phys.Rev.E.,* **2011***, 84*, 017102 [arXiv:1012.1211]
- Expert, P.; Evans, T.S.; Blondel, V. D. & Lambiotte, R. “Uncovering space-independent communities in spatial networks” *PNAS,* **2011***, 108*, 7663-7668 [arxiv.org:1012.3409]

Bibliography (2)

- Brin, S. & Page, L.“The anatomy of a large-scale hypertextual Web search engine” *Computer networks and ISDN systems,* **1998***, 30*, 107-117
- Delvenne, J.; Barahona, M.; Yaliraki, S. & Lambiotte, R. “Dynamics and Modular Structure in Networks” **2008** [arXiv.org:0812.1770]
- Delvenne, J. C.; Yaliraki, S. N. & Barahona, M.”Stability of graph communities across time scales” *PNAS,* **2010***, 107*, 12755-12760
- Girvan, M. & Newman, M. E. J. “Community structure in social and biological networks” PNAS*,* **2002***, 99*, 7821-7826
- Molloy, M. and Reed, B. “A critical point for random graphs with a given degree sequence*”, Random Structures and Algorithms* **6** (1995) 161-180.
- Molloy, M. and Reed, B. “The size of the giant component of a random graph with a given degree sequence*”*, *Combin. Probab. Comput.* **7** (1998) 295-305.
- Newman, M. E. J. “A measure of betweenness centrality based on random walks” *Social Networks,* **2005***, 27*, 39-54 [arXiv:cond-mat/0309045]
- Zachary, W. “Information-Flow Model For Conflict And Fission In Small-Groups” Journal Of Anthropological Research, 1977, 33, 452-473

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