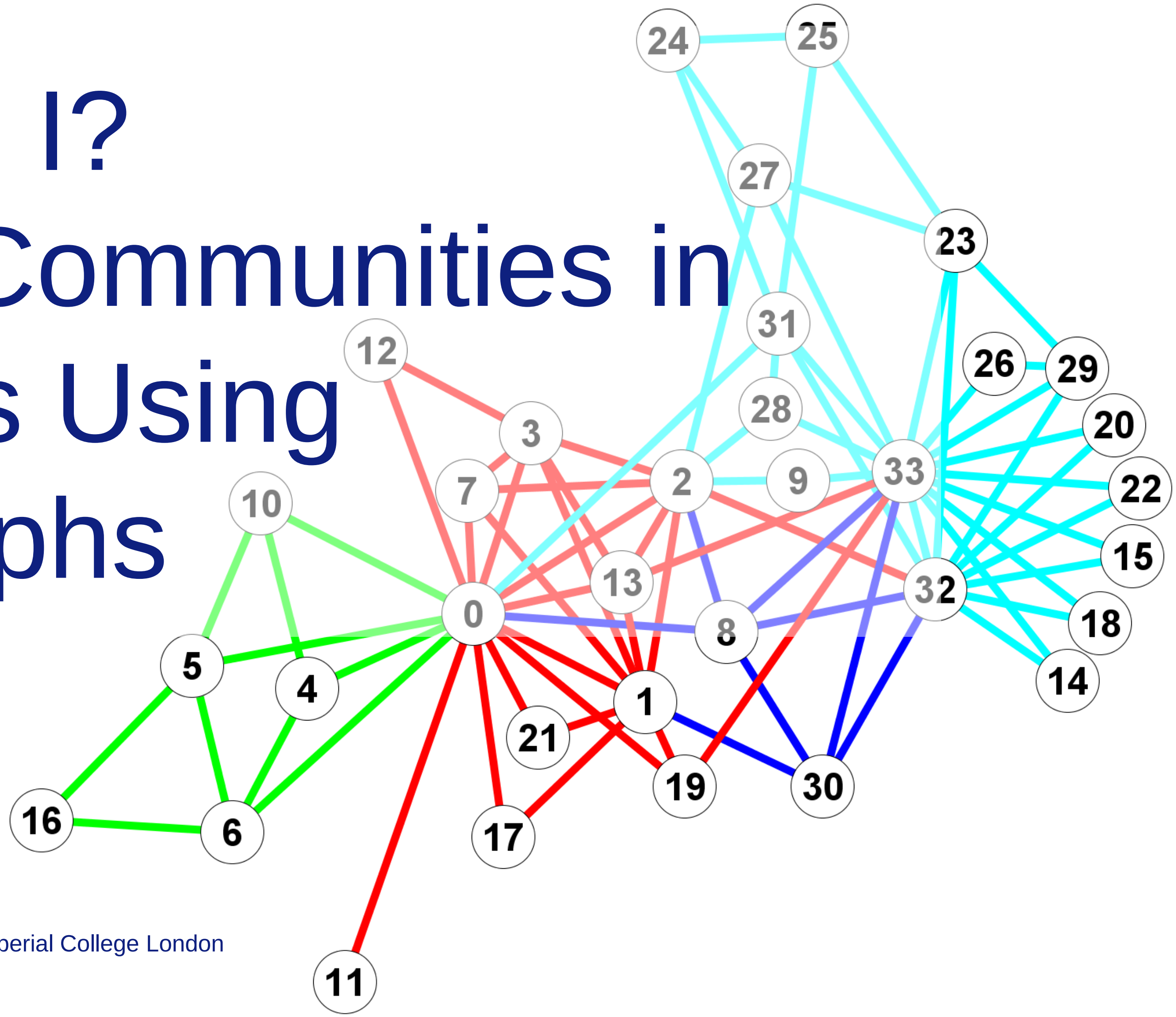
Imperial College London

Warwick Complexity Forum
28th October 2009

What am I? Finding Communities in Networks Using Line Graphs

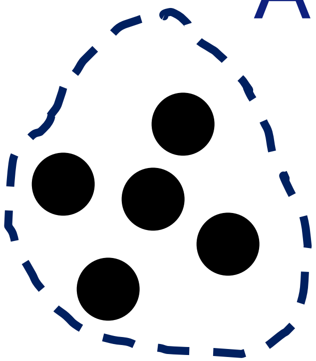
Tim Evans,
Theoretical Physics,
Imperial College
London



- Our Vertex Centric Viewpoint
- Communities in Networks
- Line Graphs
- Application to Community Detection
- Conclusions

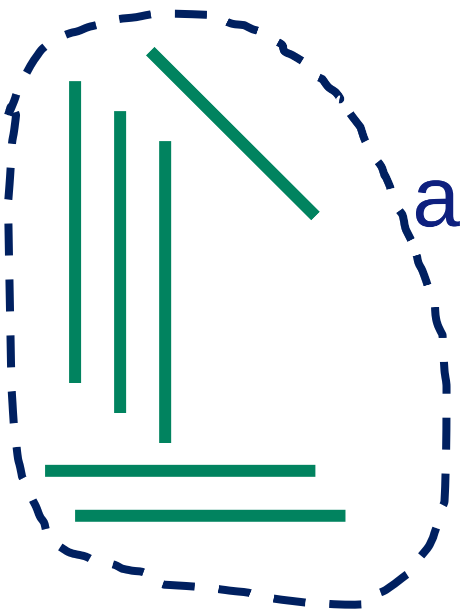
Vertex Centric Viewpoint

A Network is **BOTH**

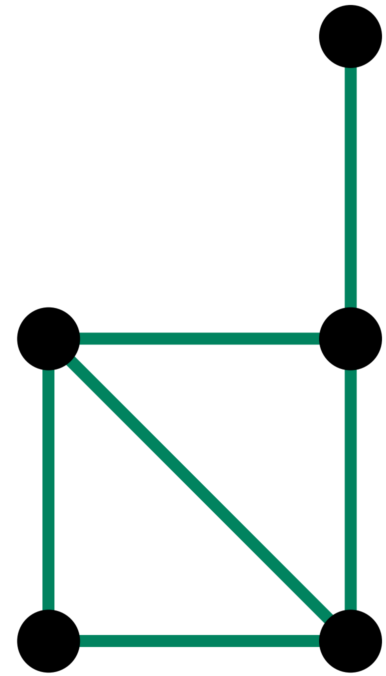


a set of vertices

AND



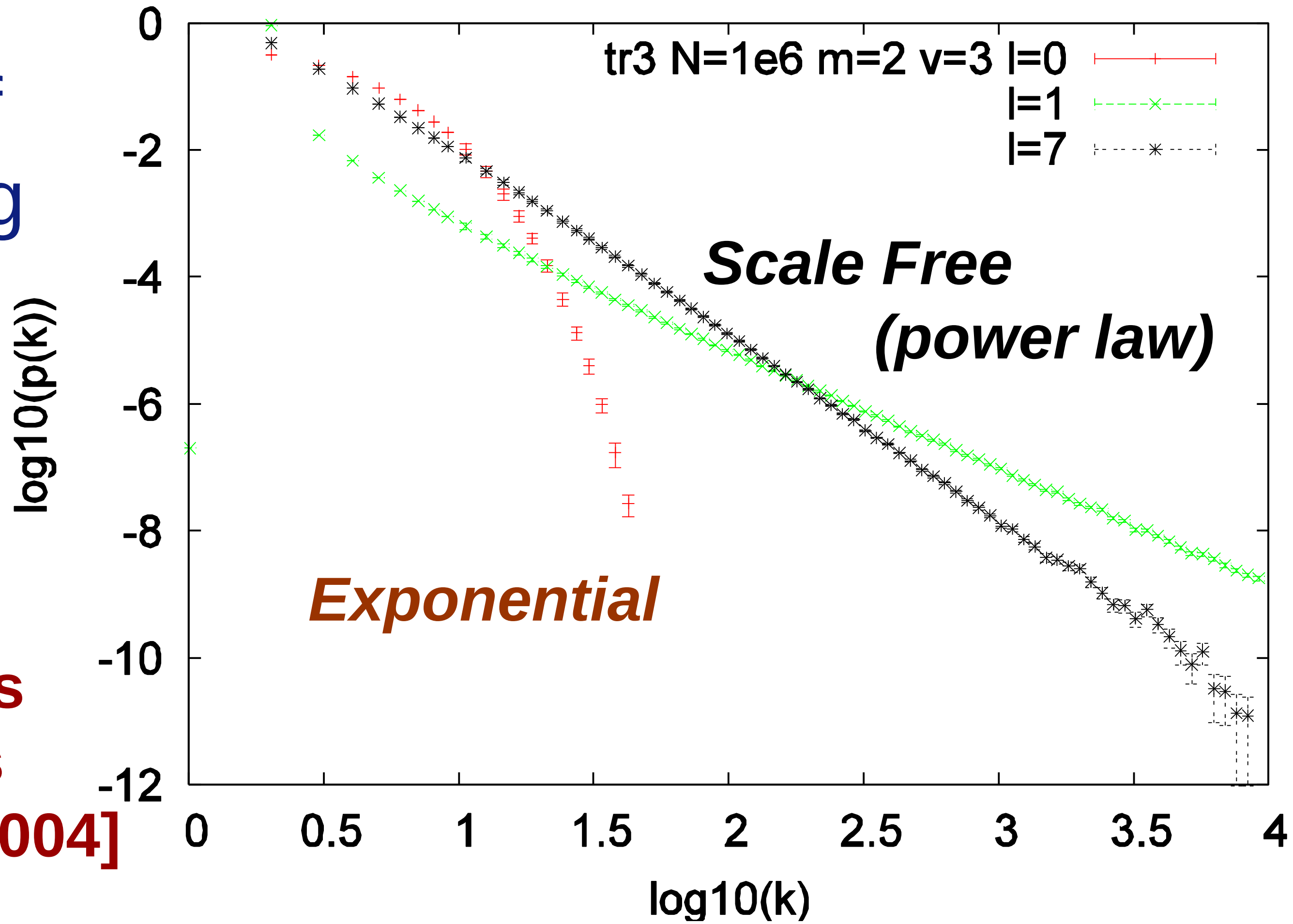
a set of edges



Yet focus is usually on vertices

Vertex Centric – Degree Distributions ***p(k)***

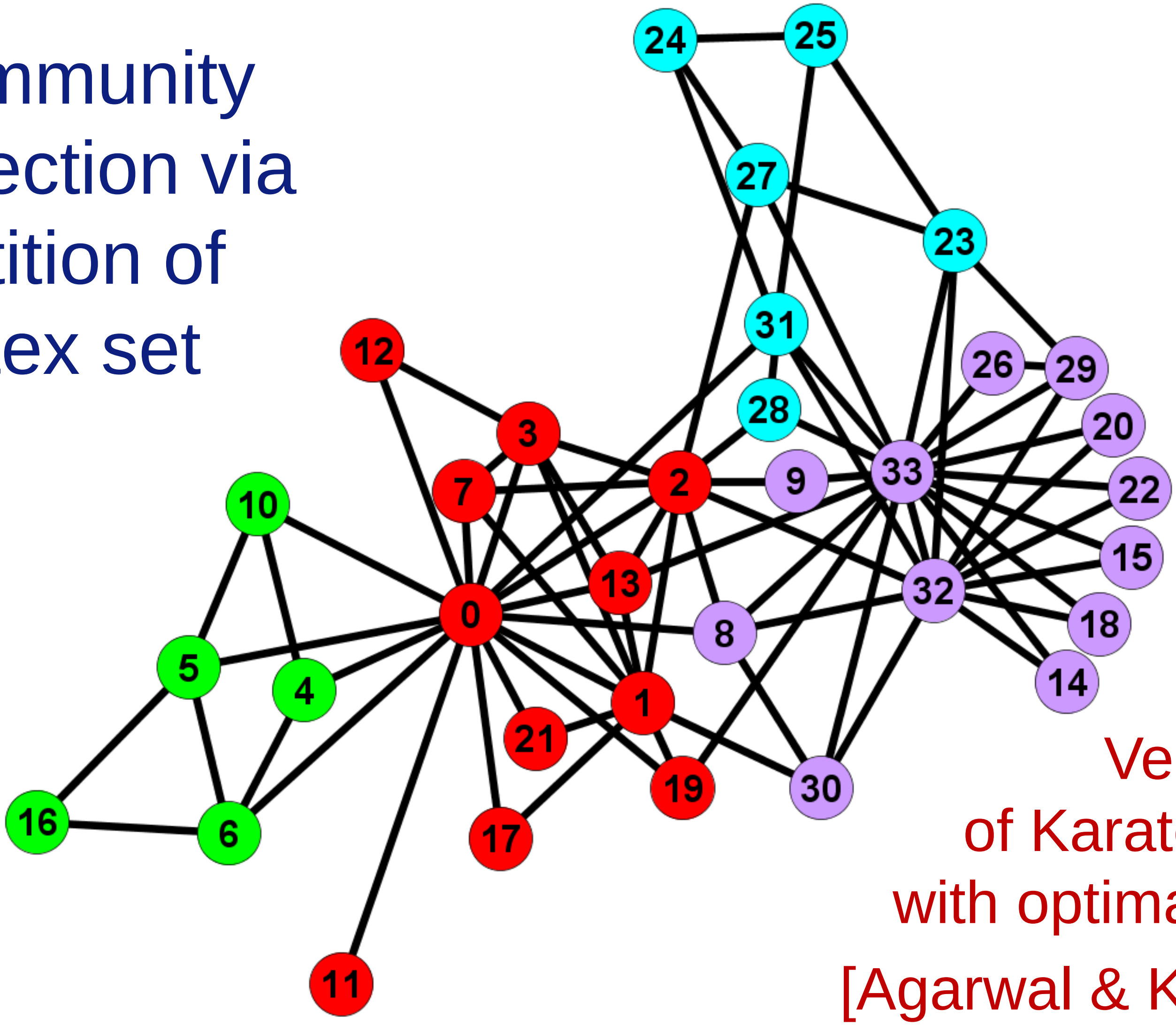
p(k) **=** probability of vertex having degree ***k***



Random walk on vertices produces scale free graphs [TSE, Saramäki 2004]

Vertex Centric – Vertex Partitions

Community detection via partition of vertex set



Vertex partition of Karate club graph with optimal modularity

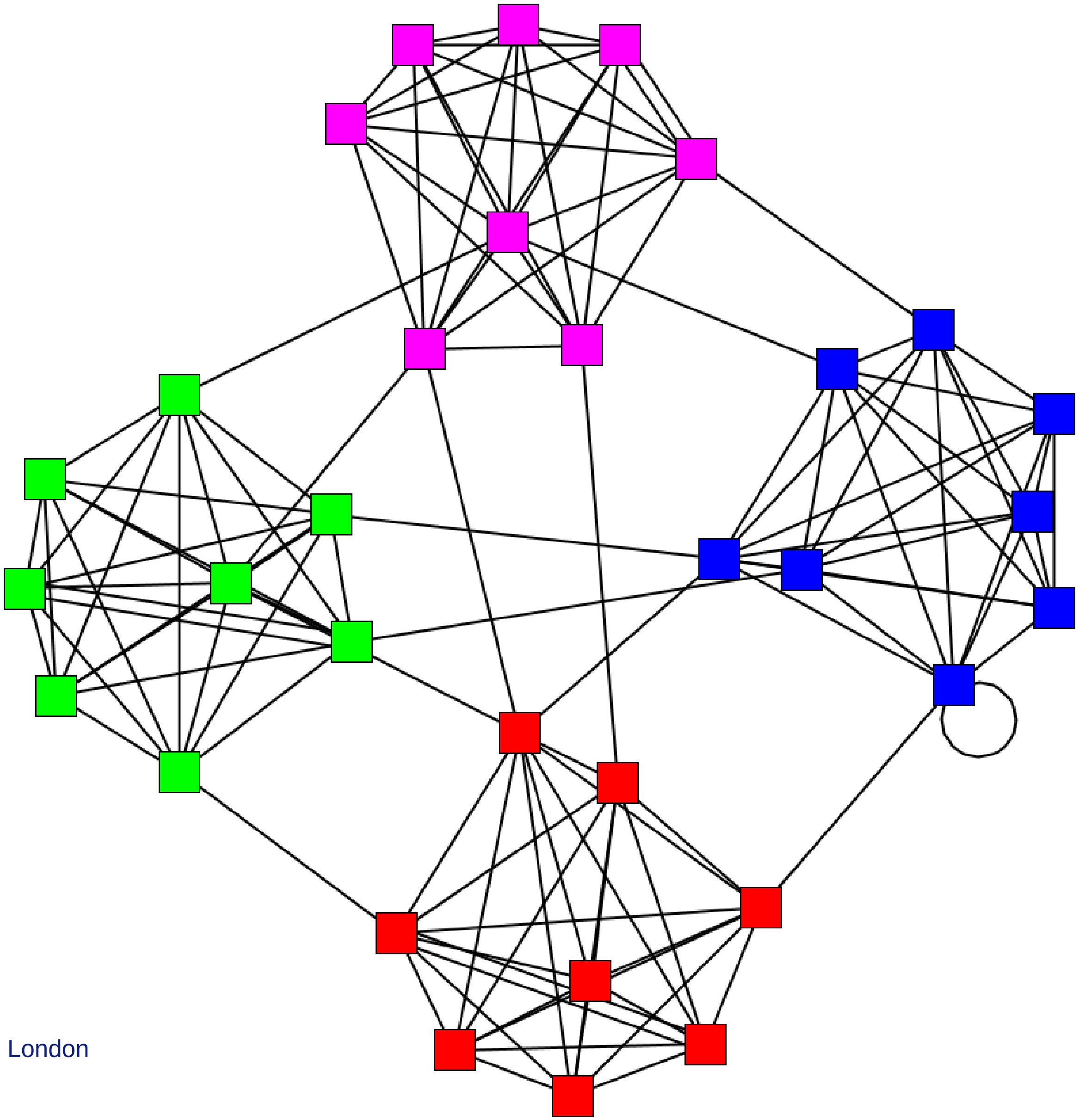
[Agarwal & Kempe 2007]

- Our Vertex Centric Viewpoint
- **Communities in Networks**
- Line Graphs
- Application to Community Detection
- Conclusions

Communities in Networks

Rough definition:-

A community is a subgraph which is more tightly connected than average

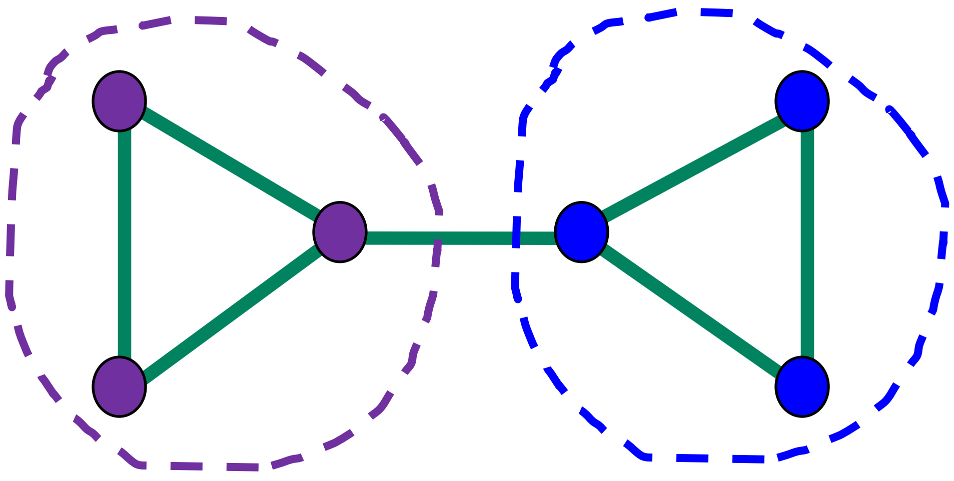


Vertex Partitions as Communities

Find a partition of the **vertices** which optimises a quality function ***Q*,**
e.g. Modularity [Girvan & Newman 2002]

***Q* =** (fraction of edges within vertex partitions)
- (fraction edges within vertex partitions in null model)

$$Q = (6/7) - (1/7) = 5/7$$



Advantages of Vertex Partition Communities

- Simplest way of assigning communities across whole graph
- Appropriate for some problems
  e.g. assigning pixels in image analysis
- Vast amount of development of theory and methods
  e.g. free code works on graph of 10^8 vertices
  in 20min
  (Louvain method, Blondel et al. 2008)

Limitations of Vertex Partition Communities

In many applications it is too simplistic to assign one vertex to one community

Examples:

- Friendship networks

- Academic papers

Who am I? Communities in Friendship Networks

Friendship networks have:-

- people as vertices
- edges if friends

We all have different types of friends:-

- Family
- Neighbours
- Work Colleagues

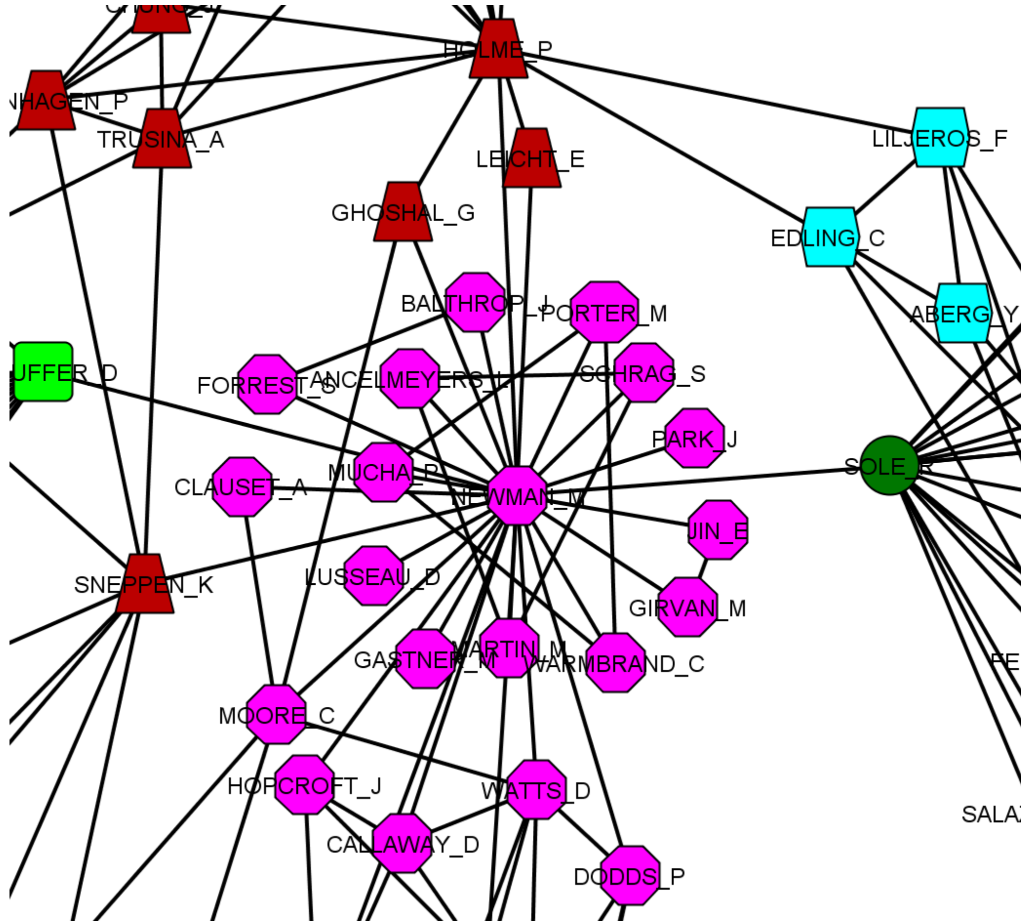
I am not just an Evans or an Imperial employee

What am I? Communities from Academic Papers

Coauthorship Networks:-

- authors as vertices
- edges between coauthors of a paper

Academics who work across boundaries always assigned to one community

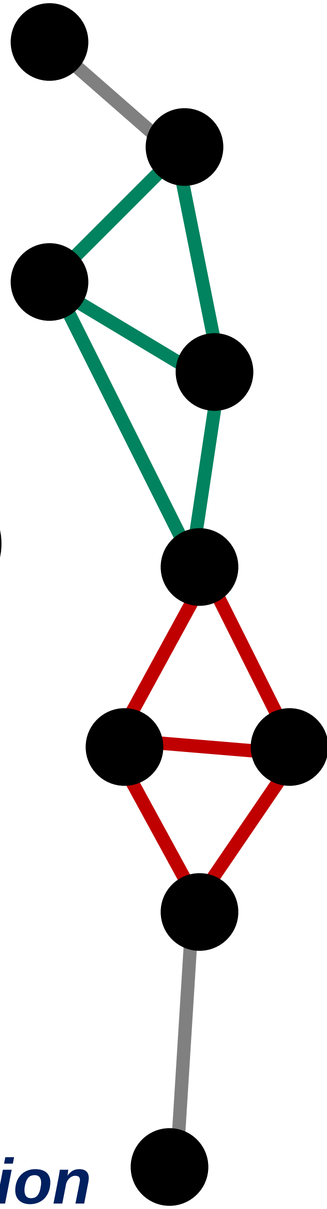


Overlapping Vertex Communities

Very few methods assign vertices to more than one community

e.g. k-clique percolation (Palla et al., 2005)

- Each community is a connected set of k-cliques (complete subgraphs of k vertices)
- Two cliques are connected if they have (k-1) vertices in common.
- Vertices can be in many communities, or none



3-clique percolation

Just Say NO to NODES

Can we shift our natural focus away from vertices and onto edges?

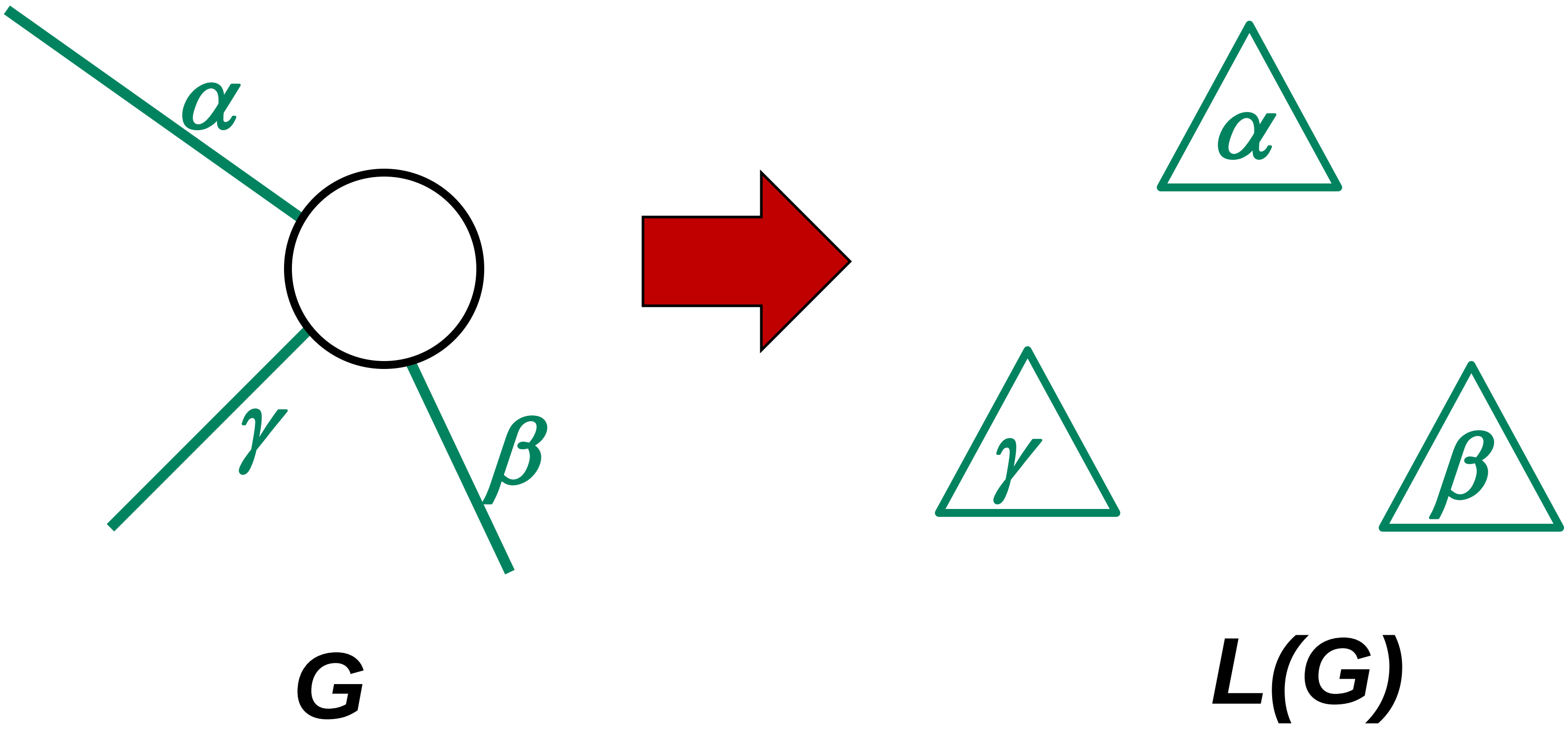
Can we focus on the relationships between agents?

For communities why do we not colour the edges not the vertices?

- Our Vertex Centric Viewpoint
- Communities in Networks
- **Line Graphs**
- Application to Community Detection
- Conclusions

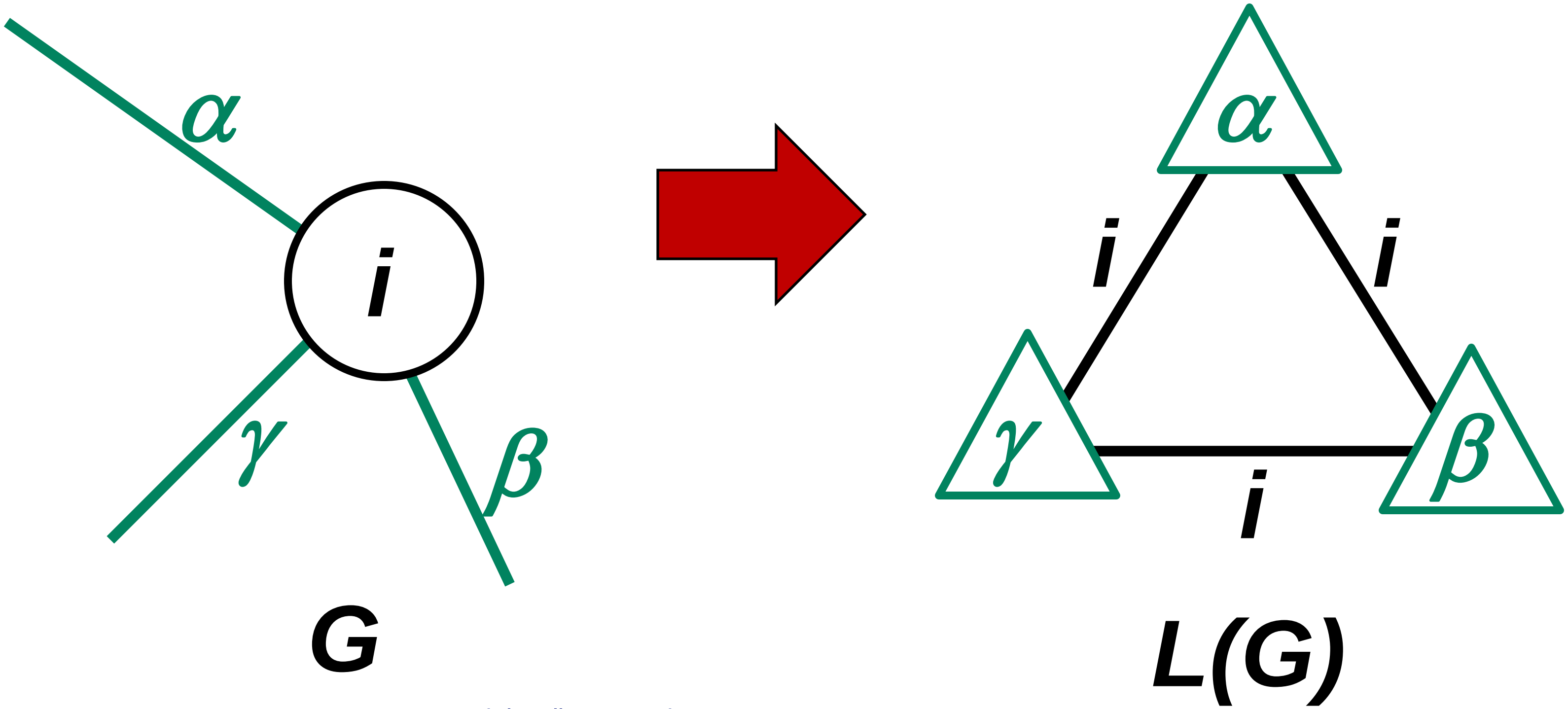
Vertices of a Line Graph

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



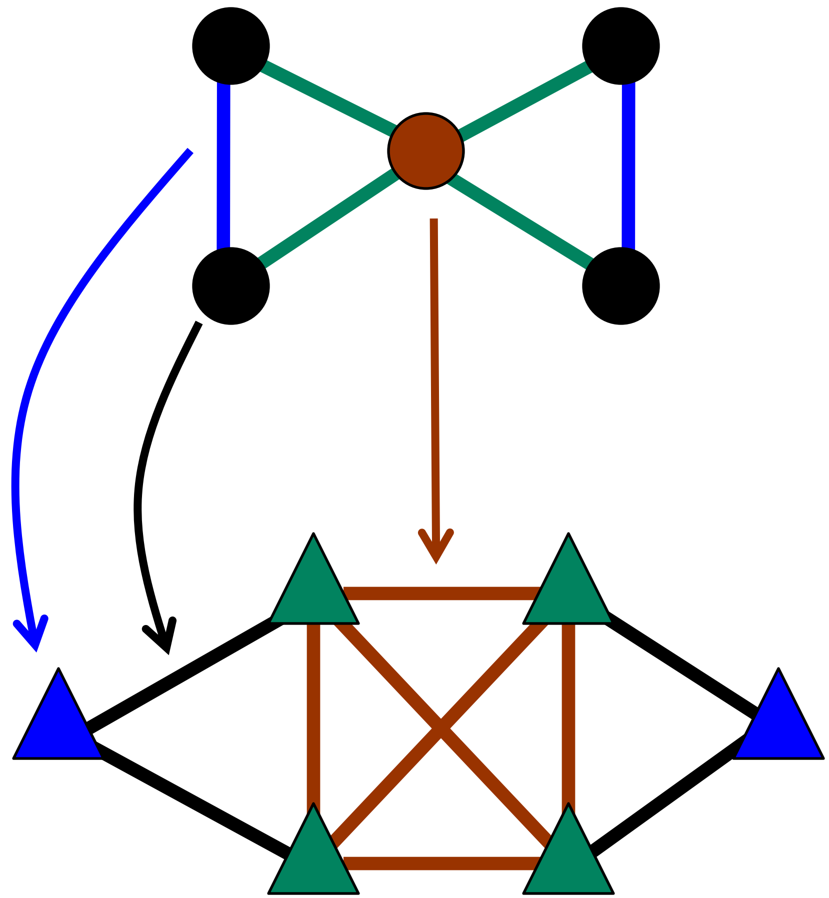
Edges of a Line Graph

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



Example – Bow Tie Graph

Graph *G*



Line Graph *L(G)*

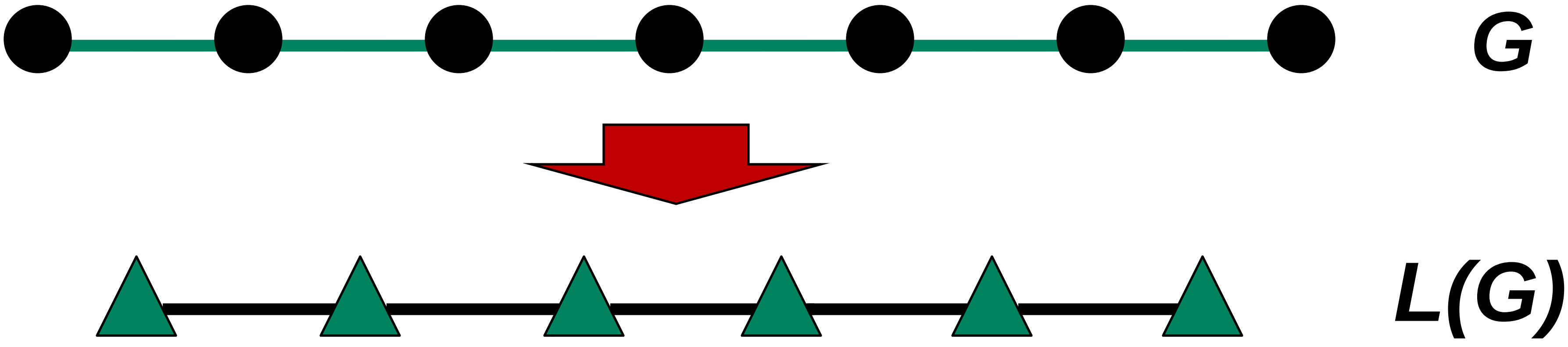
Properties of a Line Graph

- Not usually a duality transformation

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

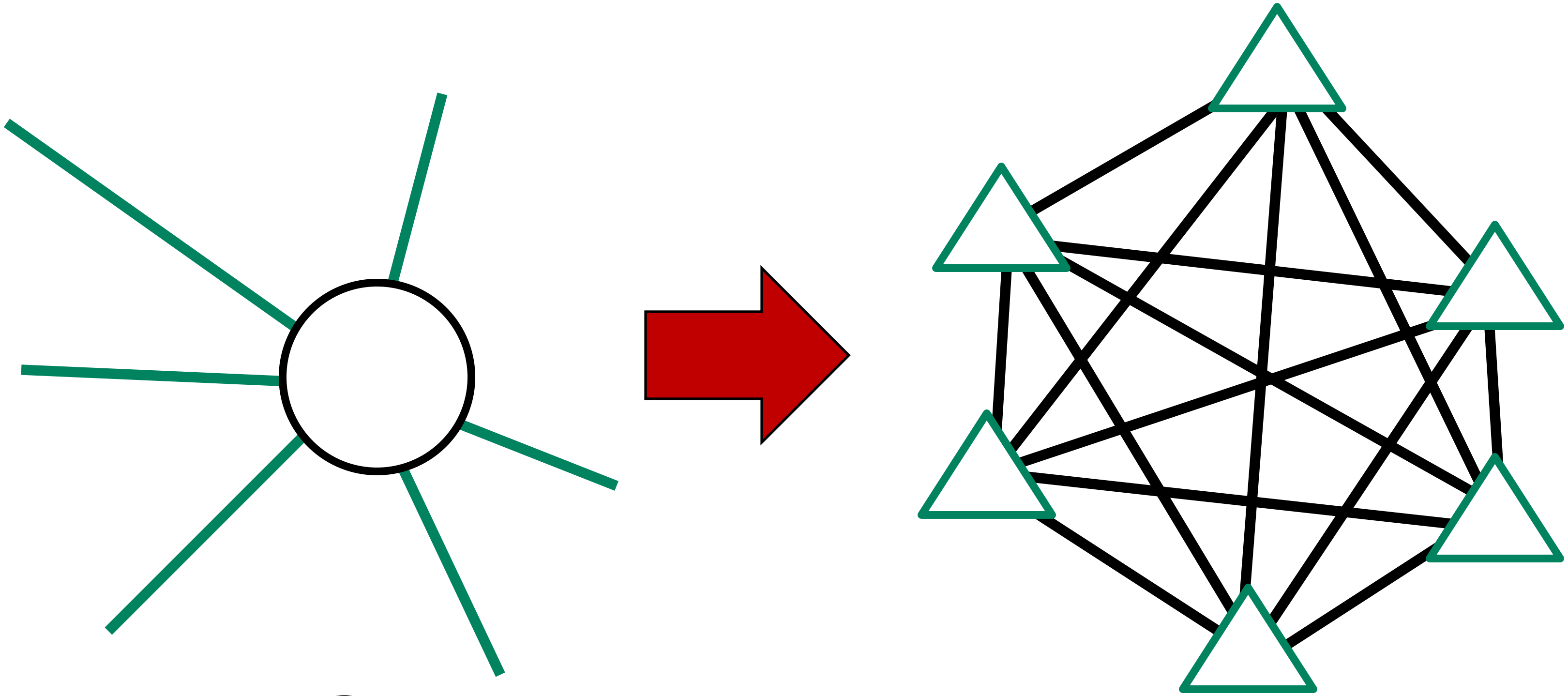
- (Almost) always reversible [Whitney 1932]

$$L(G) \rightarrow G$$



The Problem with a Standard Line Graph

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



G

Degree *k* vertex

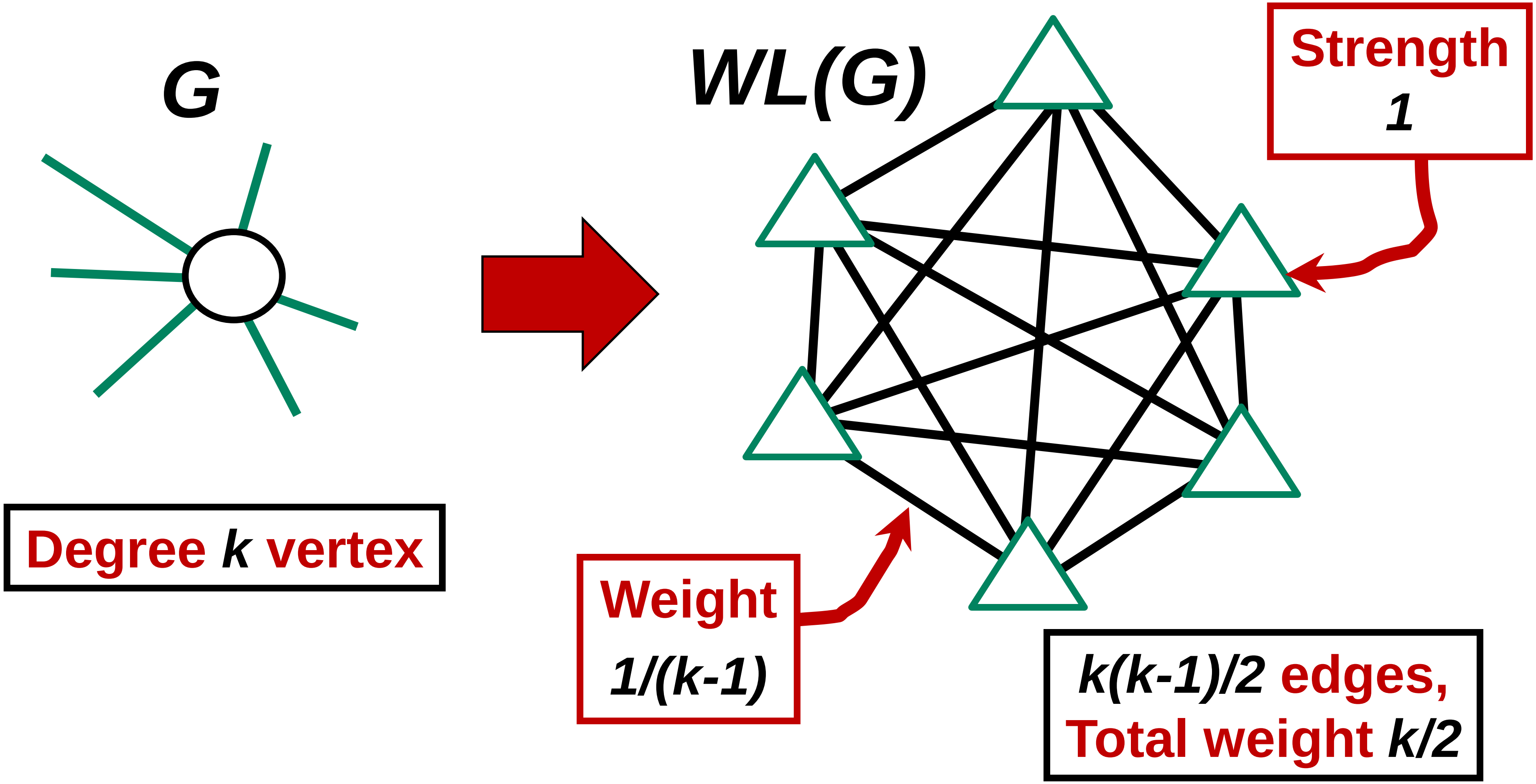
L(G)

k(k-1)/2 **edges**

[Evans & Lambiotte 2009]

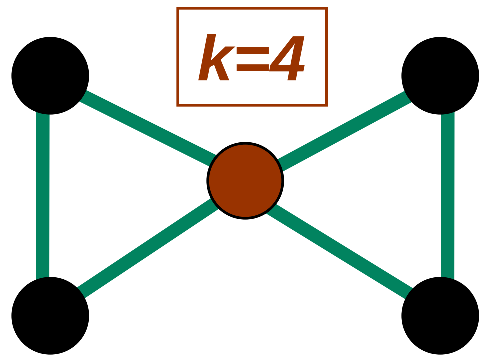
Solution – Weighted Line Graphs

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

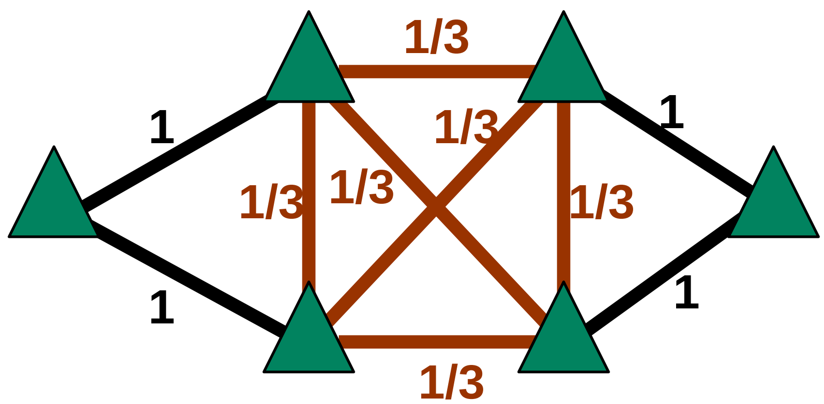


Example – Bow Tie Graph

Graph *G*



Weighted Line Graph
WL(G)

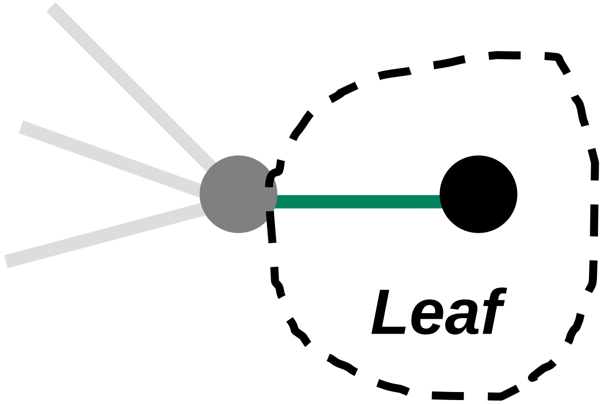


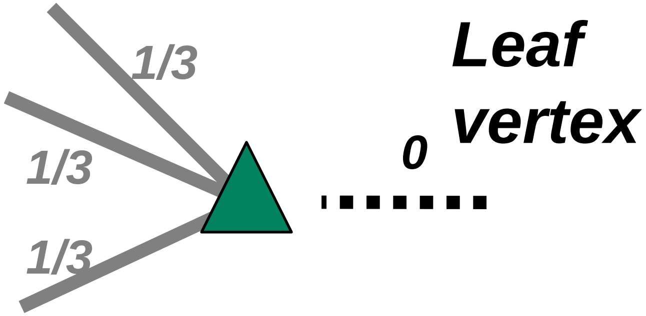
Edges 6
Strength 2

Nice Property of Weighted Line Graph

- Strength of each vertex in ***WL(G)*** is usually ***2*** since every edge in ***G*** has two ends

- Exception for edges of "leaves" in ***G*** which produce strength ***1*** vertices in ***WL(G)***.





From a Vertex to an Edge Centric Viewpoint

- Take your graph ***G*** with ***N*** vertices and $\langle k \rangle$ edges

- Make a Weighted Line Graph ***WL(G)*** with $N\langle k \rangle$ vertices and $O(N\langle k^2 \rangle)$ edges

- Run any vertex based algorithm on ***WL(G)*** and you are running it on the edges of ***G***.

- Our Vertex Centric Viewpoint
- Communities in Networks
- Line Graphs
- **Application to Community Detection**
- Conclusions

Community Detection - Vertex Centric version

To 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

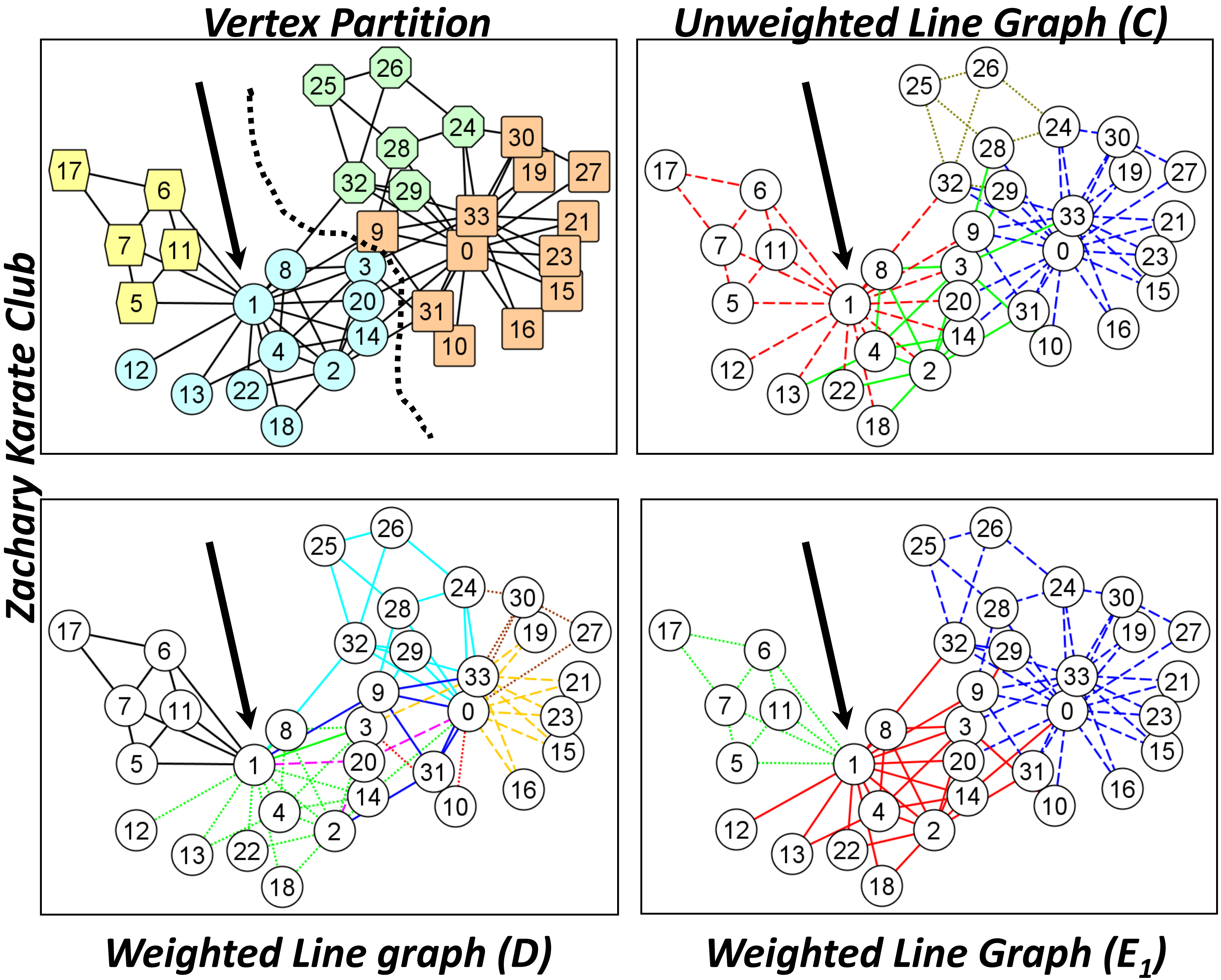
To partition edges into communities:-

- 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 [Zachary 1977]

- Community of 34 members of a Karate Club
- Split into two parts during study
  - Group centred on Club Officers
  - Group centred on instructor
- Zachary partitioned vertices into two subsets using Ford-Fulkerson algorithm (source-sink) which matched actual split except for one individual



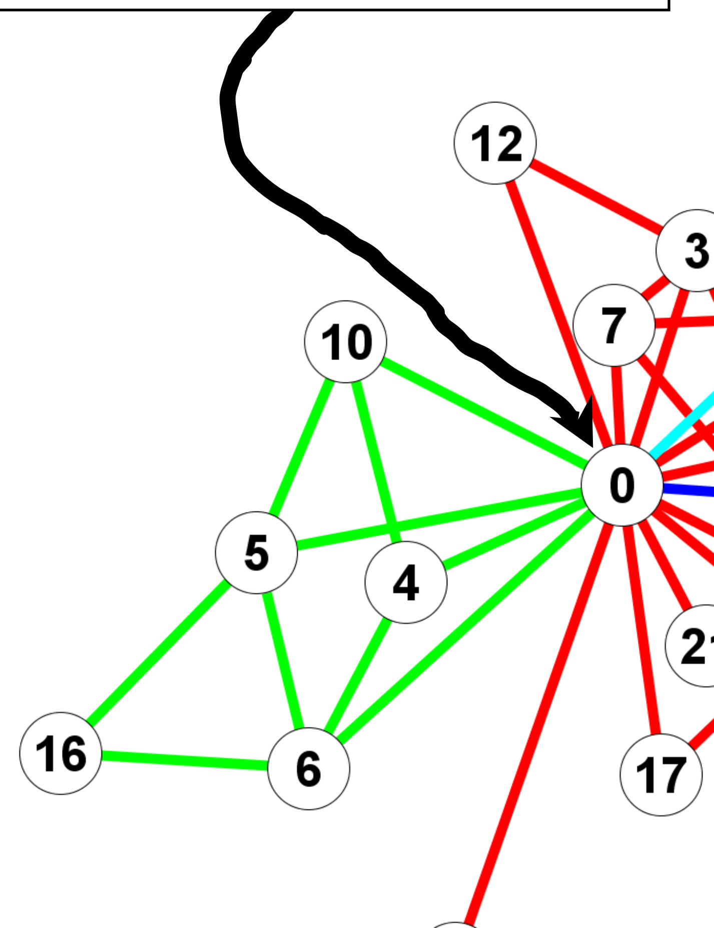
Zachary Karate Club
Vertex Partition
Unweighted Line Graph (C)
Weighted Line graph (D)
Weighted Line Graph (E1)
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

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



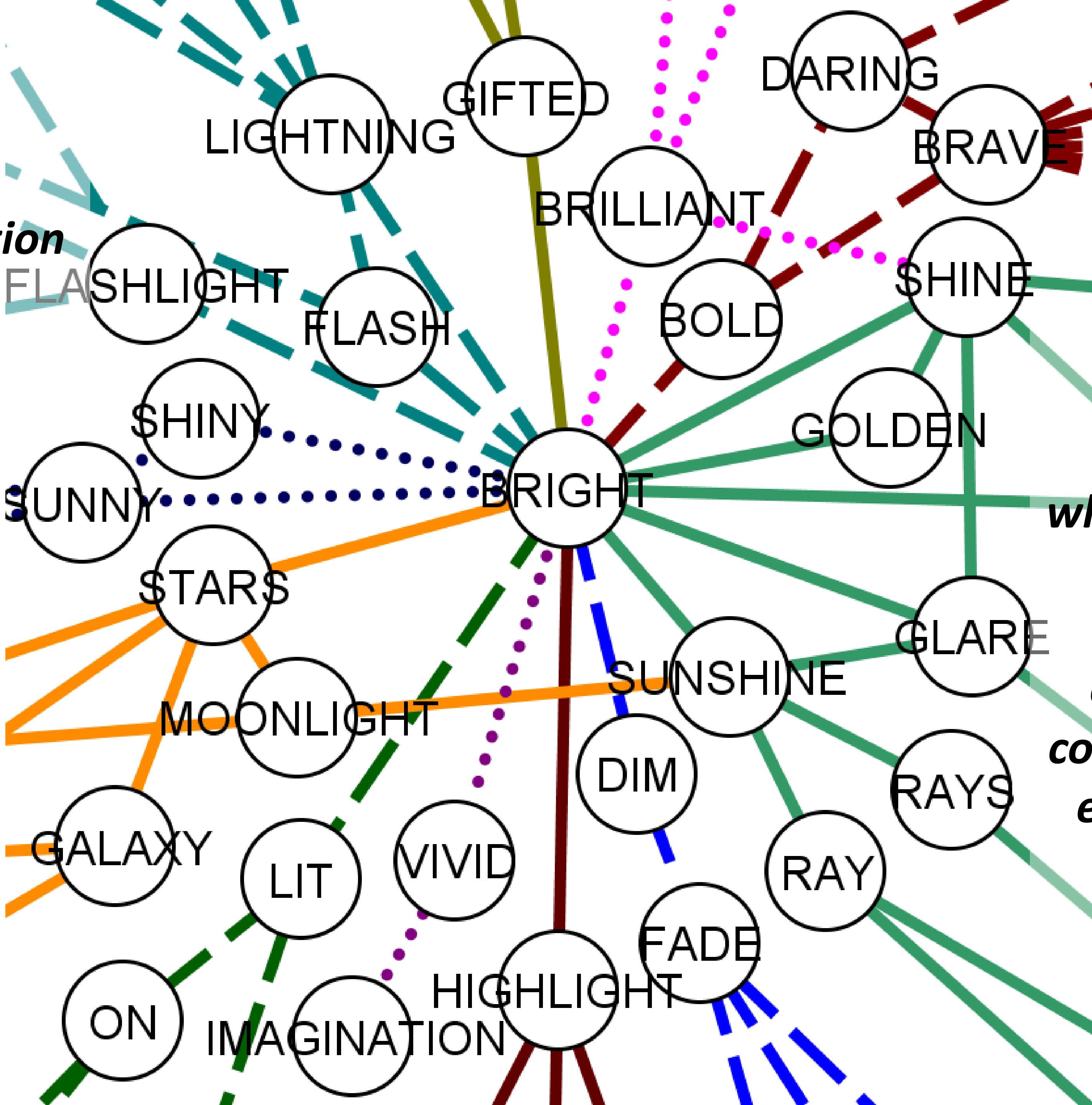
Karate Club Edge Partition

Vertices can be members of many communities

An overlapping community structure for vertices

Name	Community	Total k	k in C
0 Mr Hi	0	16	10
	1		4
	2		1
	3		1
33 John A	3	17	12
	0		3
	2		2

South Florida Word Association Data



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

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

- Our Vertex Centric Viewpoint
- Communities in Networks
- Line Graphs
- Application to Community Detection
- Conclusions

Conclusions

- Line graphs move focus from vertices to edges with minimal effort
- Weighted line graphs avoid problem of over representation of high degree vertices
- Community detection on line graph produces overlapping vertex communities for original graph

Evans & Lambiotte, Phys.Rev.E 80 (2009) 016105

http://theory.ic.ac.uk/~time/networks/

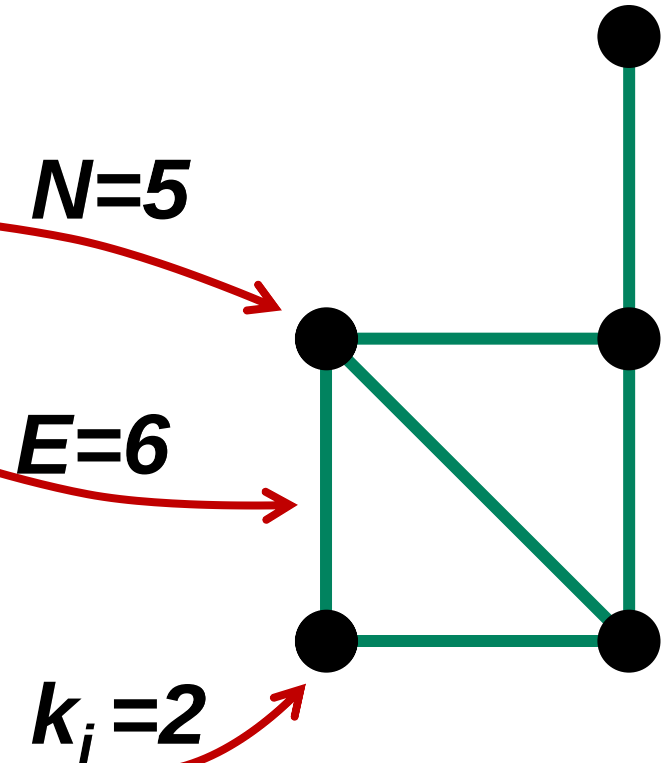
Additional Material

Basic Network Definitions

A **Network** or **Graph** is

a set of ***N*** **Vertices**

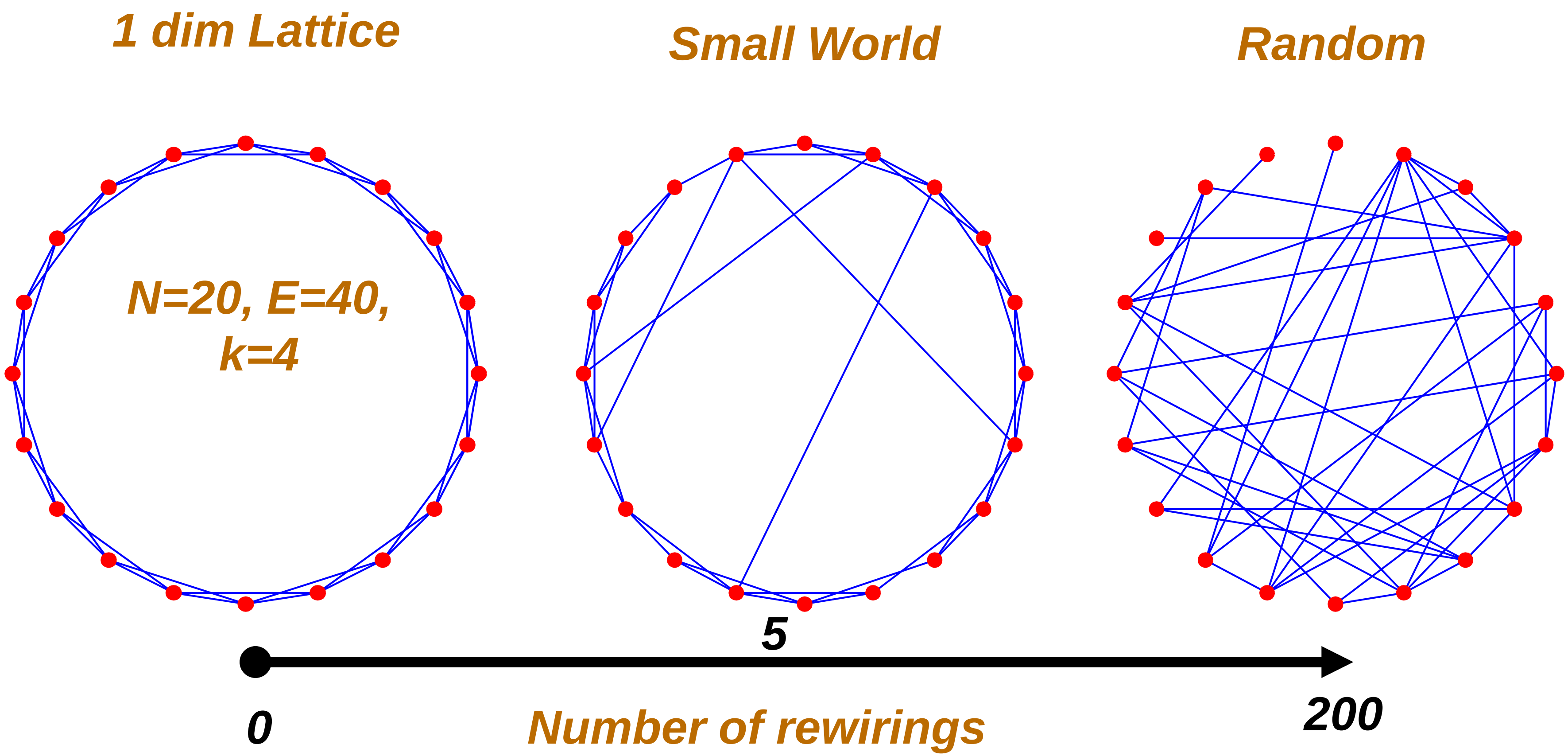
and a set of ***E*** **Edges**



Degree of vertex ***i*** is k_i

Vertex Centric - Small World networks

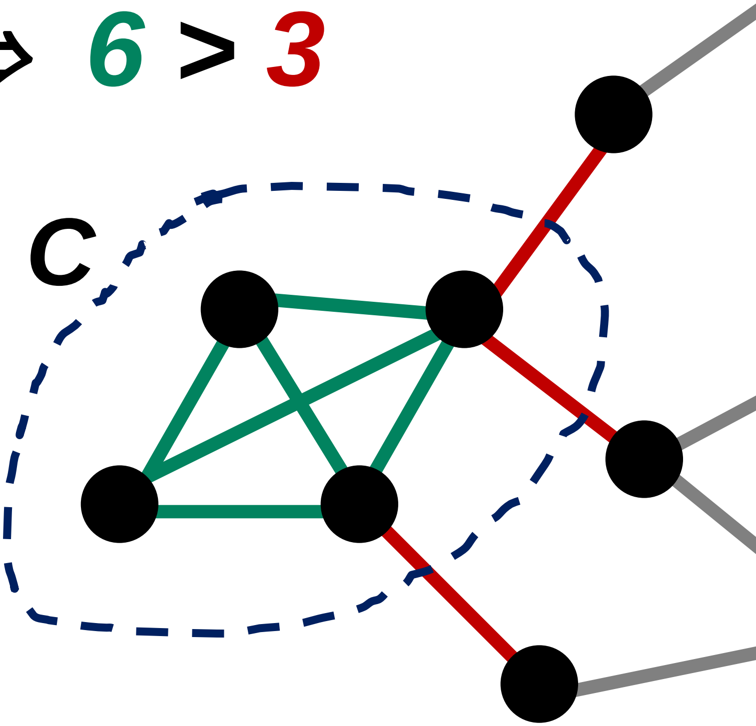
Short paths between NODES



[Watts, Strogatz 1998]

Example Community Criterion

$$\frac{1}{2}\sum_{i \in C} k_i^{in} > \sum_{i \in C} k_i^{out} \Rightarrow \mathbf{6 > 3}$$



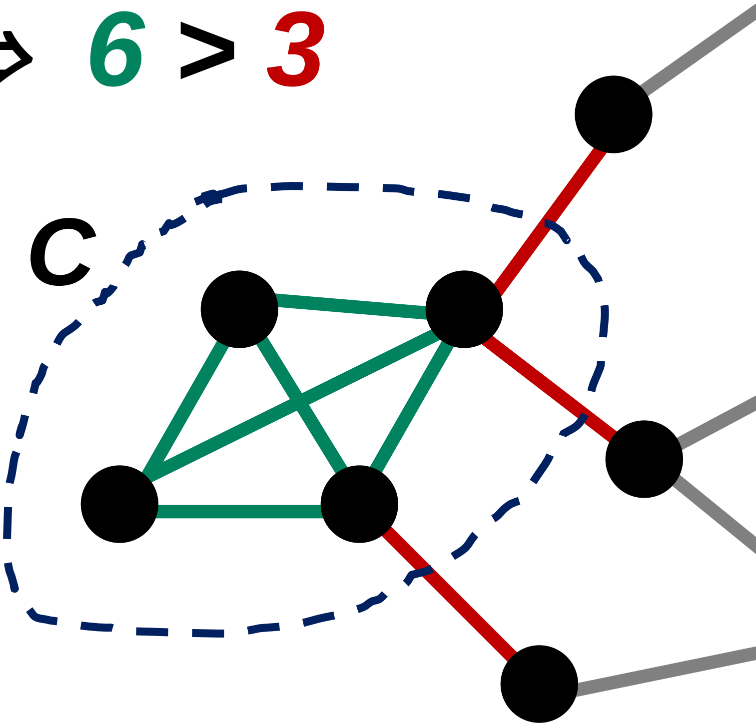
$k_i^{in(out)}$ = degree of vertex ***i*** within (leaving) community ***C***

- but there is no single rigorous definition

Example Community Criterion

The number of edges within a community is greater than number of edges leaving a community

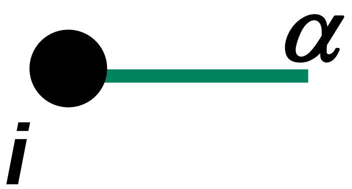
$\Rightarrow$ **6 > 3**



- but there is no single rigorous definition

Algebraic formulation – Incidence Matrix

$$B_{i\alpha} = 1$$



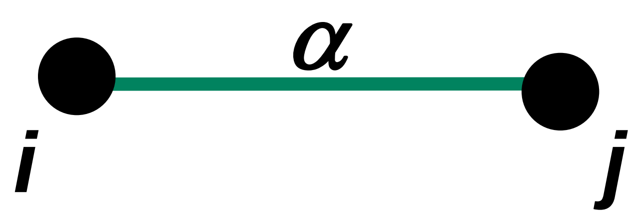
Incidence Matrix ***$B_{i\alpha}$*****:-**

1 if vertex i connected to edge α

0 if vertex i not connected to edge α

Algebraic formulation – Adjacency Matrix

$$A_{ij} = 1$$



Adjacency Matrix A_{ij} :-

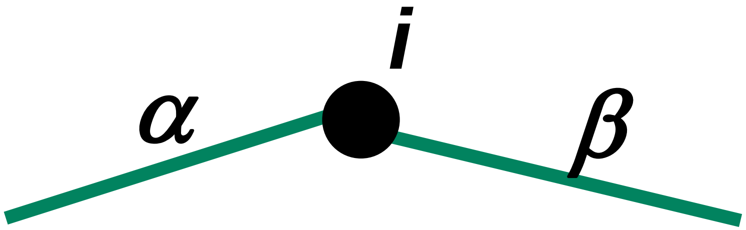
1 if vertices i and j connected

0 if vertices i and j not connected

$$A_{ij} = \sum_{\alpha} B_{i\alpha} B_{j\alpha} (1 - \delta_{ij})$$

Algebraic formulation – Line Graph

$$C_{\alpha\beta} = 1$$



Adjacency Matrix of Line Graph $C_{\alpha\beta}$:-

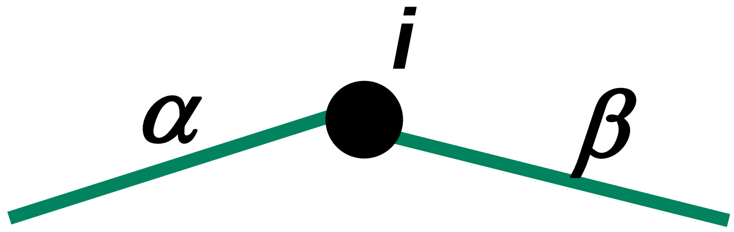
1 if edges α and β coincident at a vertex

0 if edges α and β not coincident at a vertex

$$C_{\alpha\beta} = \sum_i B_{i\alpha} B_{i\beta} (1 - \delta_{\alpha\beta})$$

Algebraic formulation – Weighted Line Graph

$$D_{\alpha\beta} = \frac{1}{k_i - 1}$$



Adjacency Matrix of Weighted Line Graph $D_{\alpha\beta}$

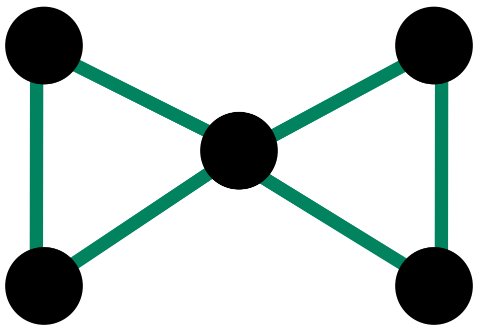
$(k_i - 1)^{-1}$ if edges α and β coincident at vertex i

0 if edges α and β not coincident at vertex i

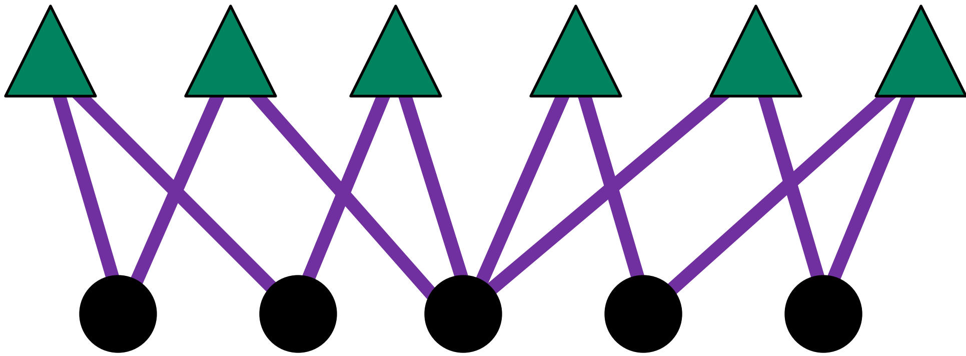
$$D_{\alpha\beta} = \sum_i \frac{B_{i\alpha} B_{i\beta}}{k_i - 1} (1 - \delta_{\alpha\beta})$$

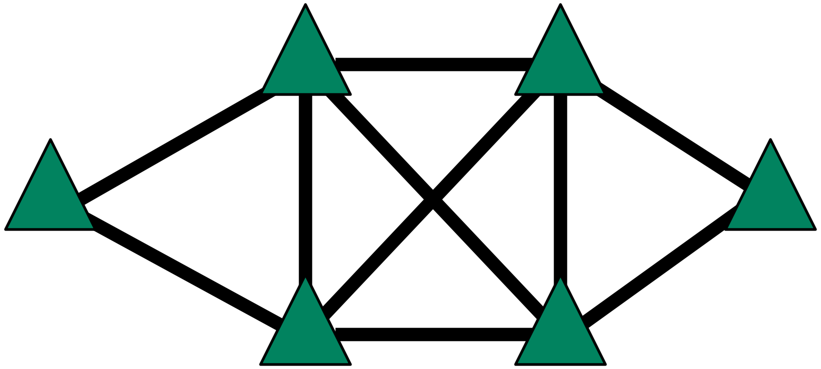
Example – Bow Tie Graph

Graph *G*, A_{ij}

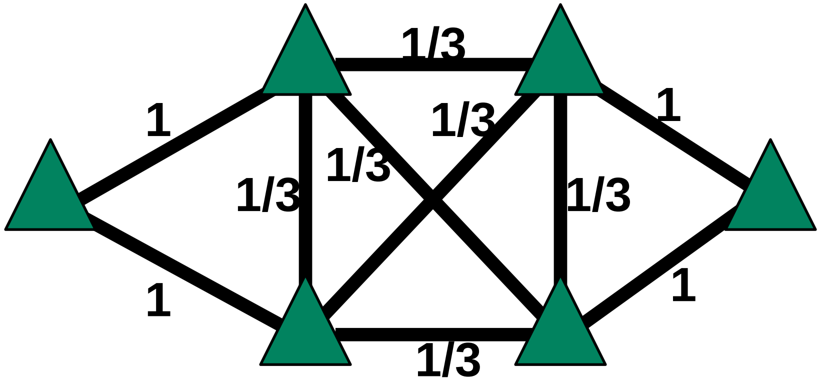


Incident Matrix $B_{i\alpha}$





Line Graph *L(G)*, $C_{\alpha\beta}$



Weighted Line Graph *WL(G)*, $D_{\alpha\beta}$