Digital Media and Social Networks

Gareth Tyson

http://www.eecs.qmul.ac.uk/~tysong/

WEEK 1: NETWORKS, RANDOM GRAPHS AND METRICS

SOME SLIDES COPYRIGHT CECILIA MASCOLO (CAMBRIDGE) AND HAMED HADDADI (QMUL)

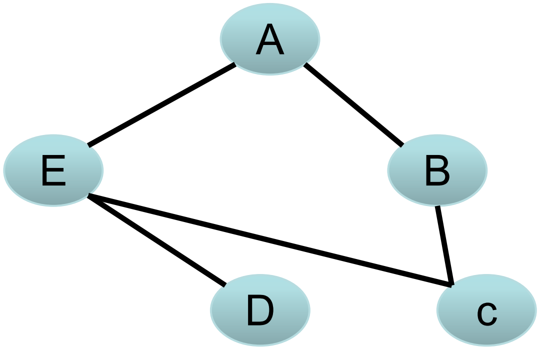
WHAT IS A GRAPH?

WHAT IS A GRAPH?

A **set of objects** where some pairs of objects are **connected by links**

A way of modelling the **connectivity** of **things**

Usually represented in either visual manner of using set theory



WHAT IS A SET?

WHAT IS A SET?

A **set** is a collection of **distinct objects**

Notation uses { }

V = {Gareth, Laurissa, Bob}

WHAT IS A SET?

Element of is an element that exists in a given set

Notation uses $a \in \mathrm{A}$

V={Gareth,Laurissa,Bob}

Gareth $\in$ V

...means that Gareth is part of set V

WHAT IS A SET?

Cardinality is the number of elements in a set

Notation uses $|V|$

V={Gareth,Laurissa,Bob}

$|V|=3$

...there are 3 elements in the set

WHAT IS A SET?

Such that means that a value is subject to certain constraints

Imagine you want to extract all people who are over the age of 65

Notation is |

$Y = \{x \mid x \in B, x < 0\}$

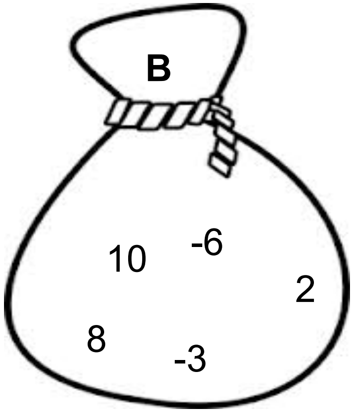
...Y contains all negative numbers in B

WHAT IS A SET?

Such that means that a value is subject to certain constraints

Imagine you want to extract all people who are over the age of 65

Notation is |



$Y = \{x \mid x \in B, x < 0\}$

...Y contains all negative numbers in B

WHAT IS A SET?

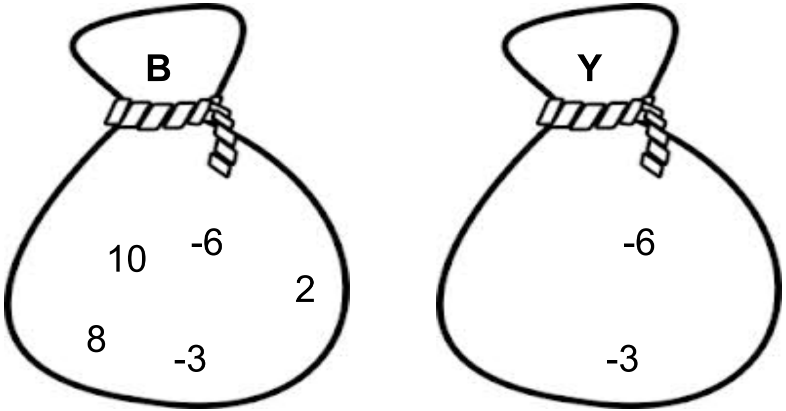
Such that means that a value is subject to certain constraints

Imagine you want to extract all people who are over the age of 65

Notation is |

$Y = \{x \mid x \in B, x<0\}$

...Y contains all negative numbers in B



WHAT IS A SET?

Such that means that a value is subject to certain constraints

Imagine you want to extract all people who are over the age of 65

Notation is |

We usually use uppercase letters indicate a set

$Y = \{x \mid x \in B, x < 0\}$

...Y contains all negative numbers in B

WHAT IS A SET?

Such that means that a value is subject to certain constraints

Imagine you want to extract all people who are over the age of 65

Notation is |

$\mathrm{Y} = \{x \mid x \in \mathrm{B},\ x<0\}$

We usually use lowercase letters indicate an individual element from a set

...Y contains all negative numbers in B

HOW TO DESCRIBE A GRAPH

A NETWORK IS A GRAPH

A **graph** G is a tuple (V,E) of a set of vertices V and edges E. An edge in E connects two vertices in V.

A NETWORK IS A GRAPH

V = {Gareth, Laurissa, Bob} for the vertices

E = {(Gareth,Laurissa), (Laurissa,Bob)} for the edges

G =?

A NETWORK IS A GRAPH

V = {Gareth, Laurissa, Bob} for the vertices

E = {(Gareth,Laurissa), (Laurissa,Bob)} for the edges

G = (V, E)

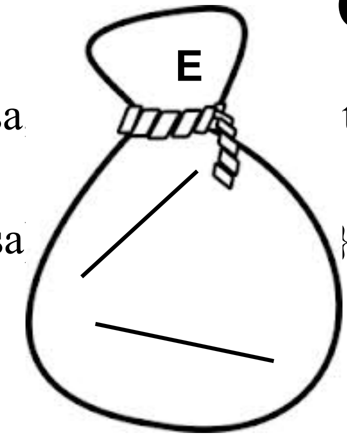
A NETWORK IS A GRAPH

V = {Gareth, Laurissa tices

E = {(Gareth,Laurissa } for th

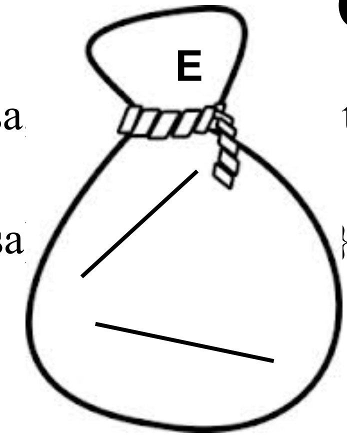
G = (V, E)

G



A NETWORK IS A GRAPH

G



V = {Gareth, Laurissa … tices

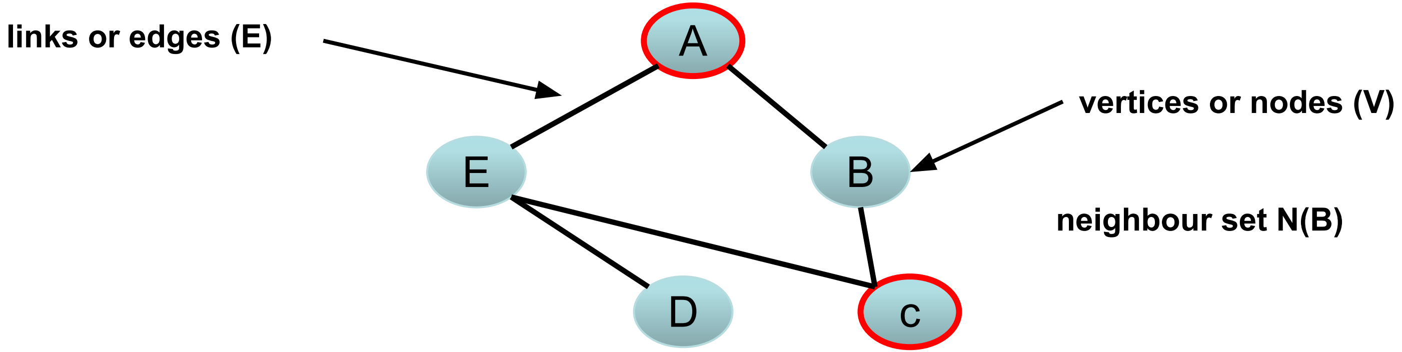
E = {(Gareth,Laurissa … } for th

G = (V, E)

|**V**| = ?

HOW TO MEASURE A GRAPH

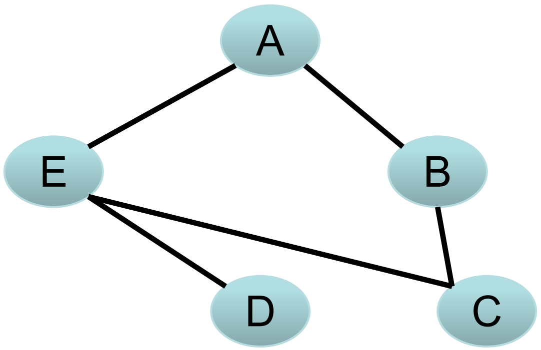
NEIGHBOUR SETS



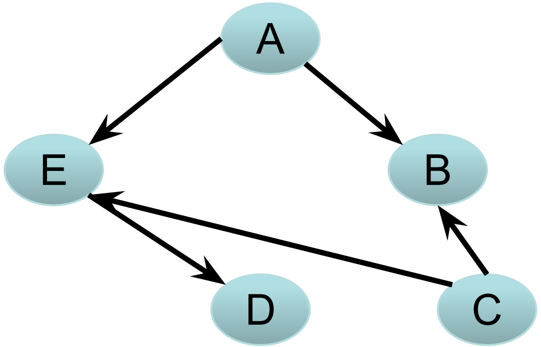
A **neighbour set** N(v) is the set of vertices adjacent to v:

$$N(v) = \{u \in V \mid u \neq v, (v,u) \in E\}$$

DIRECTED & UNDIRECTED GRAPHS



Undirected
Graph

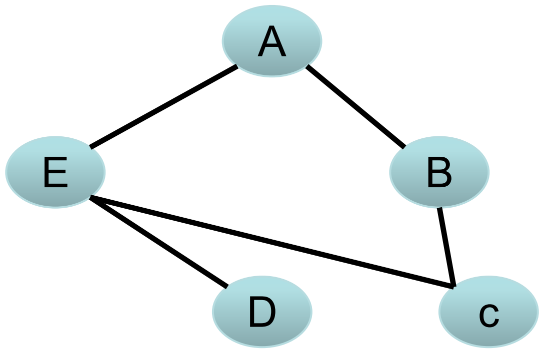


Directed
Graph

A directed graph has
arrowheads!

NODE DEGREE

The **node degree (k)** is the number of neighbours of a node

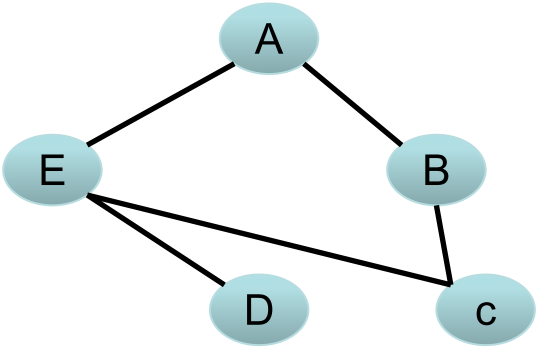


The study of the degree distribution of networks allows the classification of networks in different categories

NODE DEGREE

The **node degree (k)** is the number of neighbours of a node

If we wanted to have fun, we could say **|N(v)|**

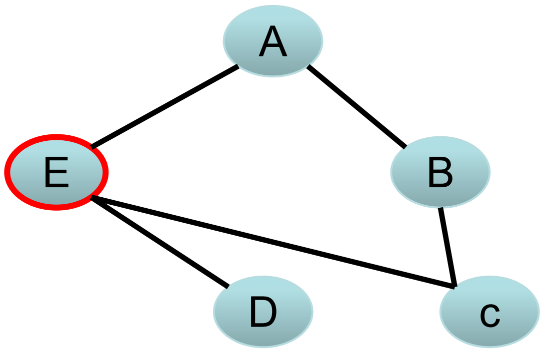


The study of the degree distribution of networks allows the classification of networks in different categories

NODE DEGREE

The **node degree (k)** is the number of neighbours of a node

Degree of E?

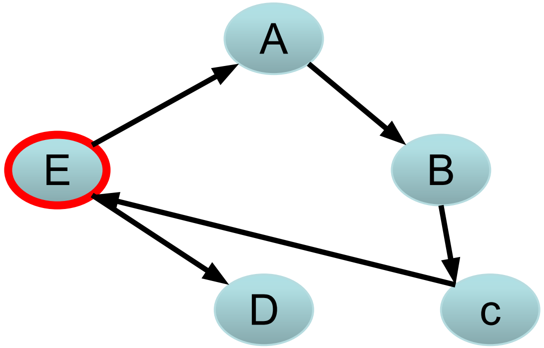


The study of the degree distribution of networks allows the classification of networks in different categories

NODE DEGREE

The **node degree (k)** is the number of neighbours of a node

In-degree of E?

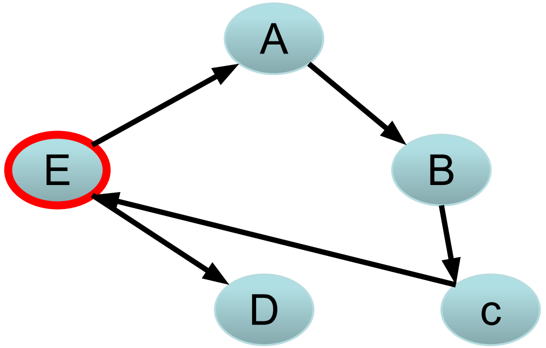


The study of the degree distribution of networks allows the classification of networks in different categories

NODE DEGREE

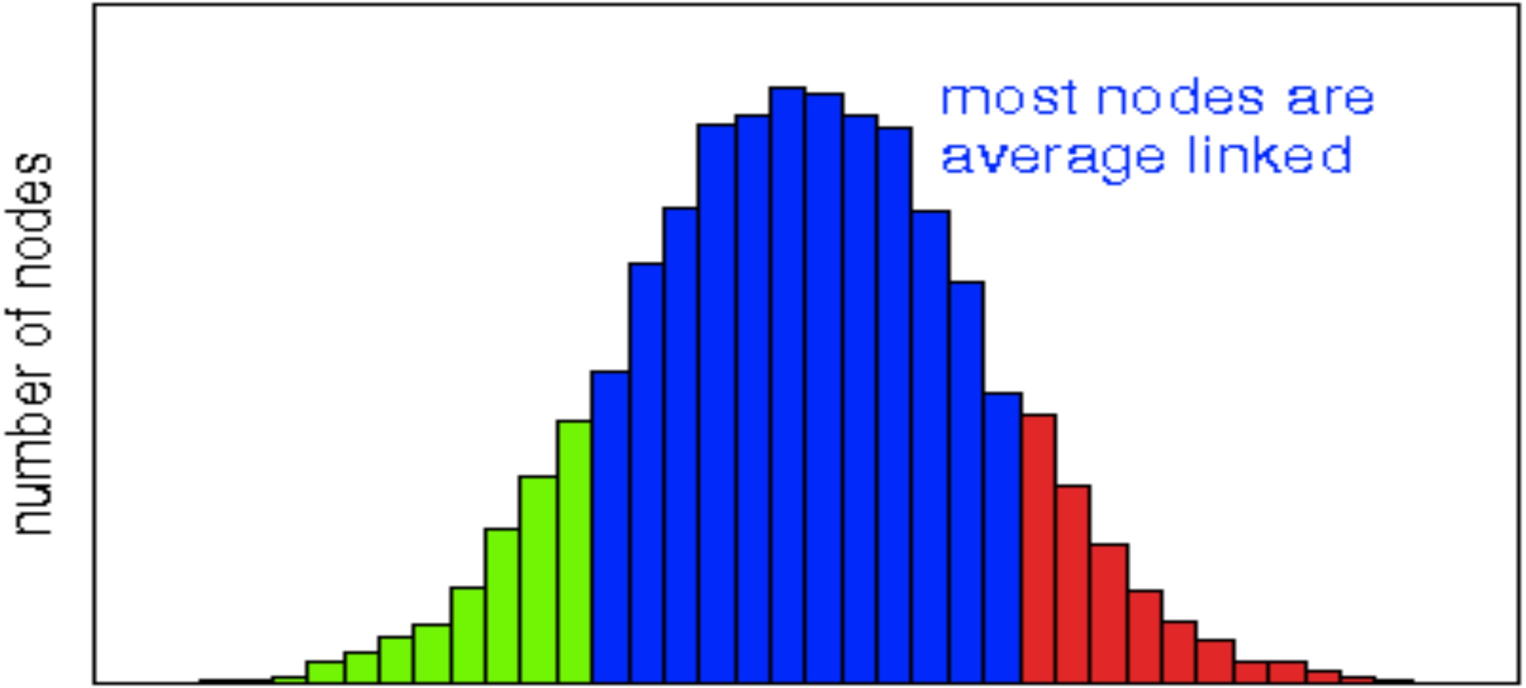
The **node degree (k)** is the number of neighbours of a node

Out-degree of E?

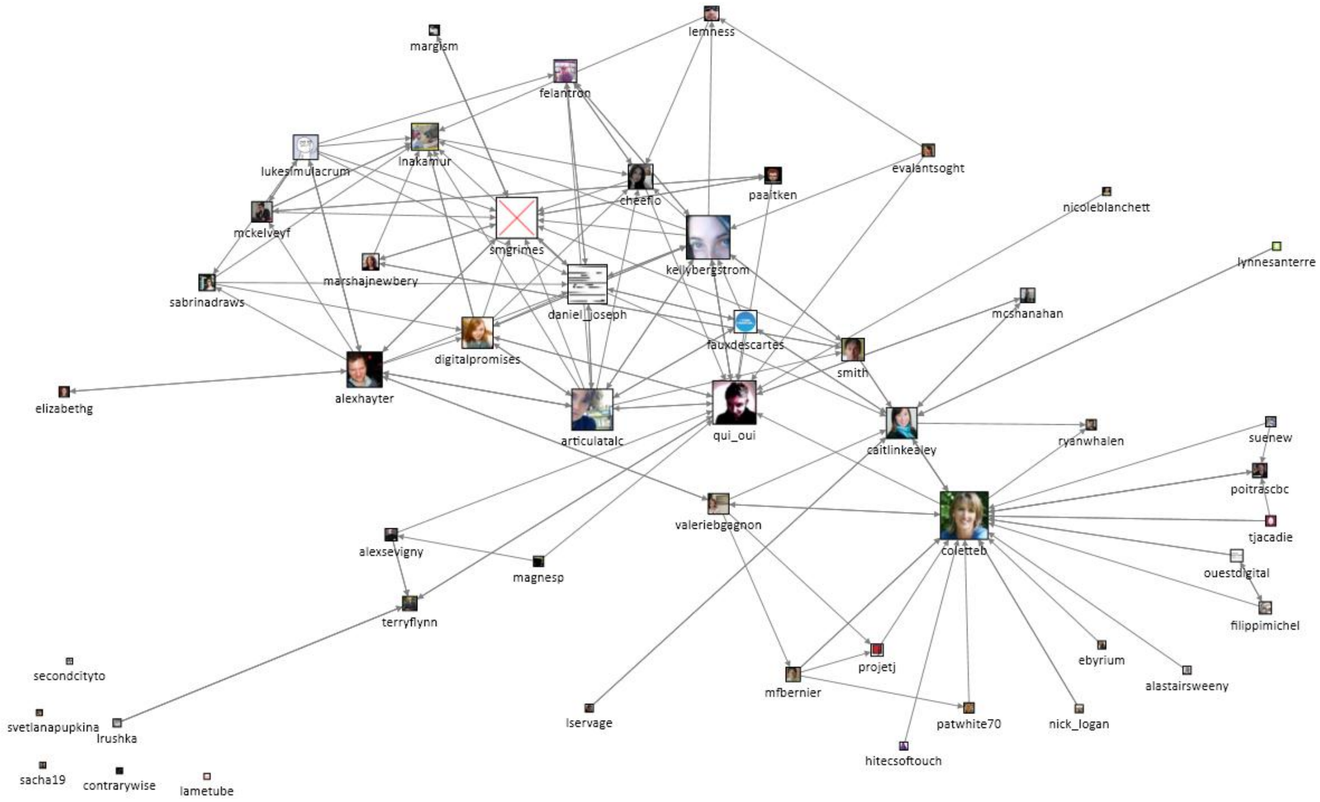


The study of the degree distribution of networks allows the classification of networks in different categories

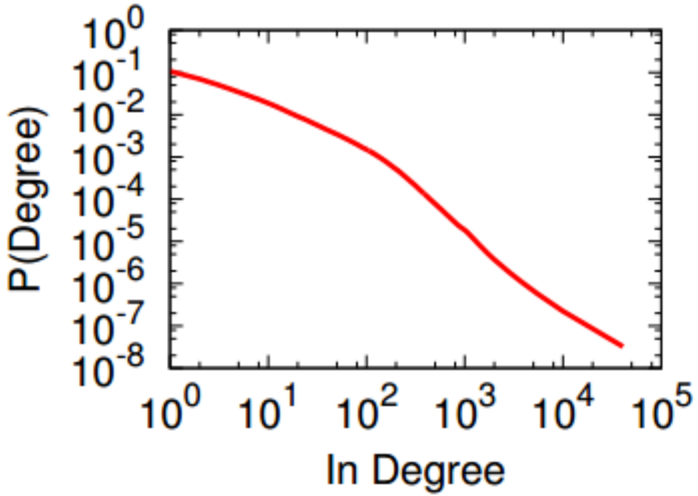
NODE DEGREE



NODE DEGREE

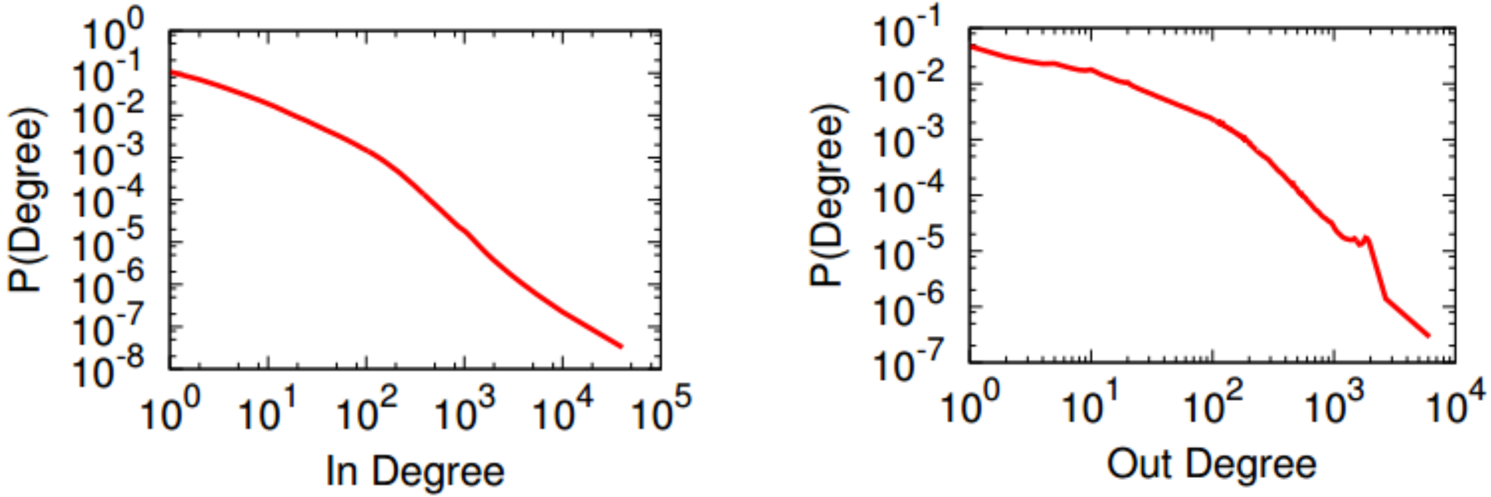

margism
lemness
felantron
lnakamur
lukesimulacrum
cheeflo
paaltken
evalantsoght
nicoleblanchett
mckelveyf
smgrimes
kellybergstrom
lynnesanterre
marshajnewbery
sabrinadraws
daniel_joseph
mcshanahan
fauxdescartes
digitalpromises
smith
elizabethg
alexhayter
articulatalc
qui_oui
caitlinkealey
ryanwhalen
suenew
poitrascbc
valeriebgagnon
coletteb
tjacadie
alexsevigny
magnesp
ouestdigital
terryflynn
filippimichel
ebyrium
secondcityto
projetj
alastairsweeny
mfbernier
lservage
patwhite70
nick_logan
svetlanapupkina
lrushka
hitecsoftouch
sacha19
contrarywise
lametube

NODE DEGREE



P(Degree)
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^0 10^1 10^2 10^3 10^4 10^5
In Degree

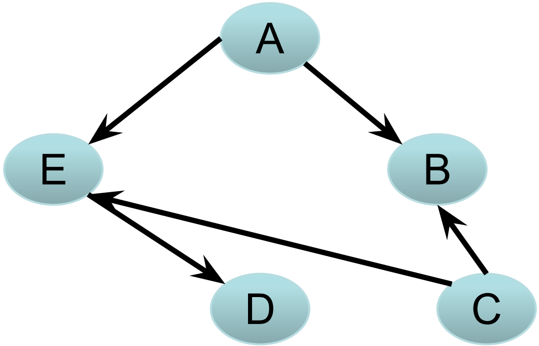
NODE DEGREE



The Structure of the Twitter Follower Graph

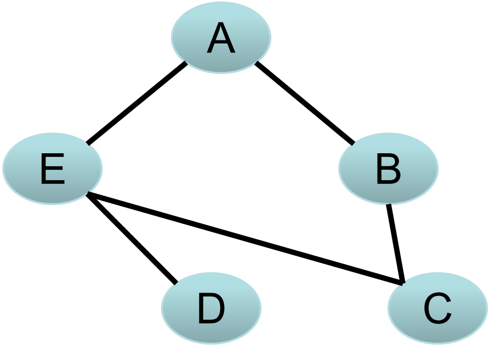
http://wwwconference.org/proceedings/www2014/companion/p493.pdf

PATHS AND CYCLES

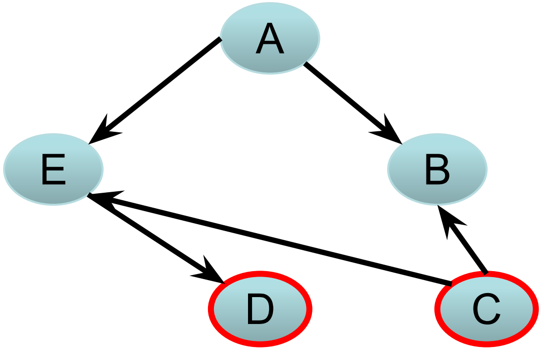


A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.

- If graph is directed the edge needs to be in the right direction.
- E.g. A-E-D is a path in both previous graphs

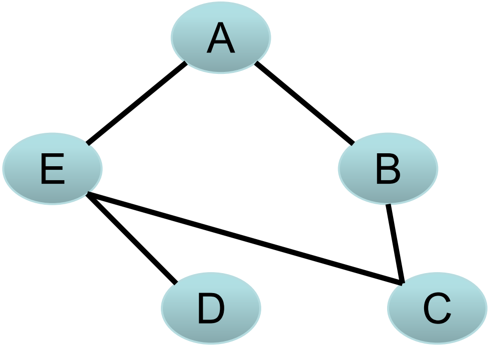


PATHS AND CYCLES

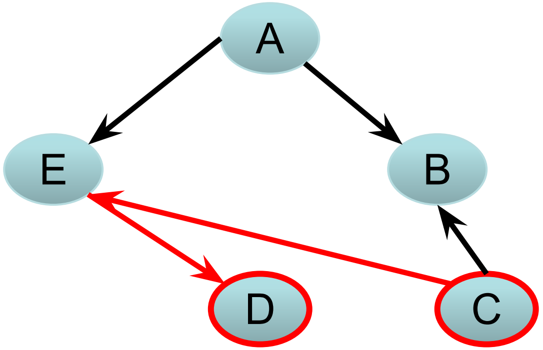


A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.

- If graph is directed the edge needs to be in the right direction.
- E.g. A-E-D is a path in both previous graphs

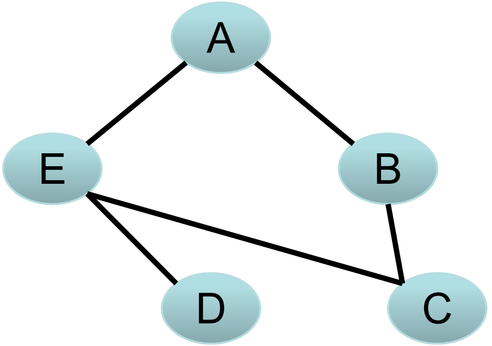


PATHS AND CYCLES

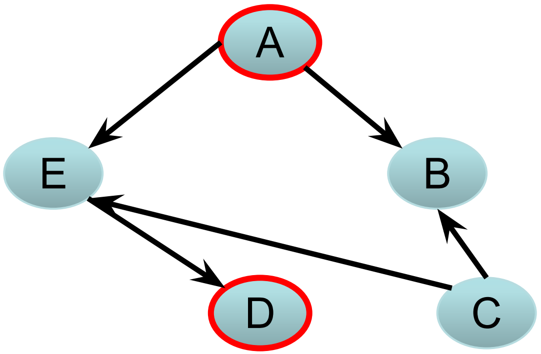


A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.

- If graph is directed the edge needs to be in the right direction.
- E.g. A-E-D is a path in both previous graphs

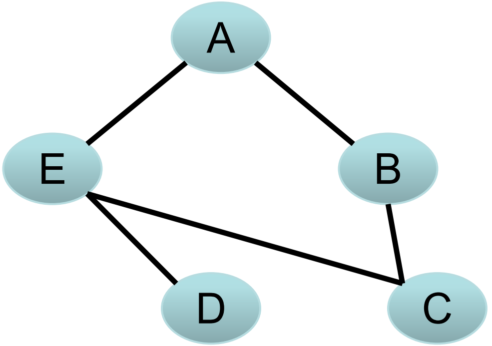


PATHS AND CYCLES

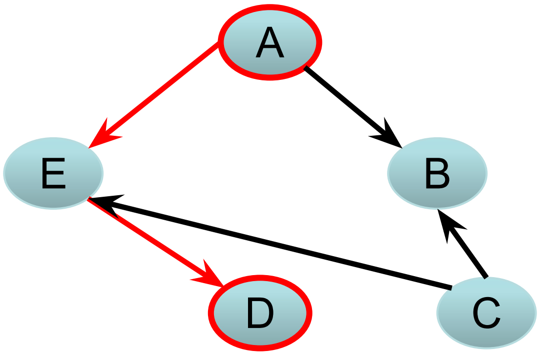


A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.

- If graph is directed the edge needs to be in the right direction.
- E.g. A-E-D is a path in both previous graphs

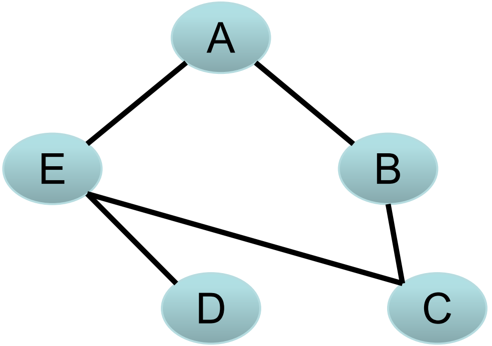


PATHS AND CYCLES

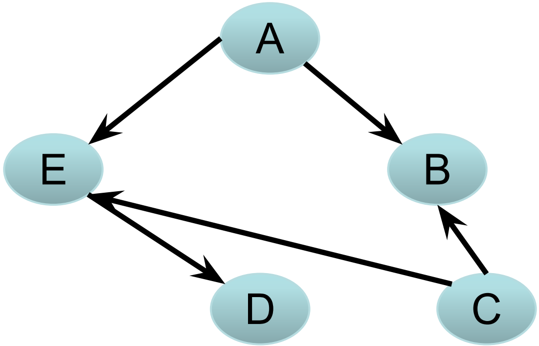


A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.

> If graph is directed the edge needs to be in the right direction.
>
> E.g. A-E-D is a path in both previous graphs



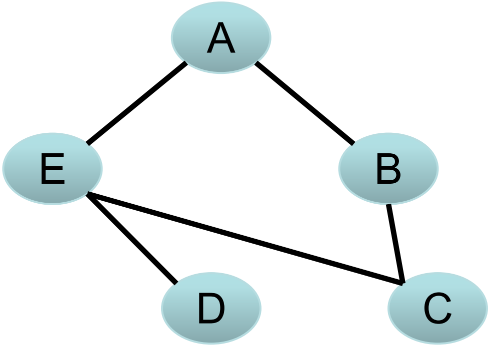
PATHS AND CYCLES



A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.

If graph is directed the edge needs to be in the right direction.

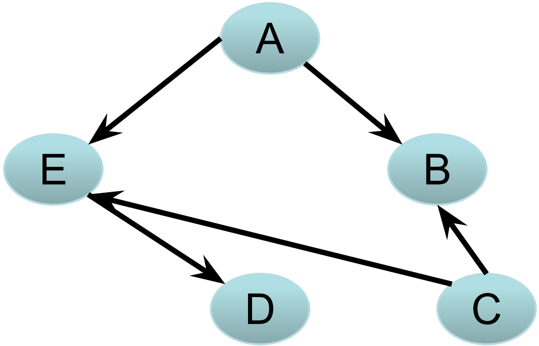
E.g. A-E-D is a path in both previous graphs



A **cycle** is a path where the start node is also the end node

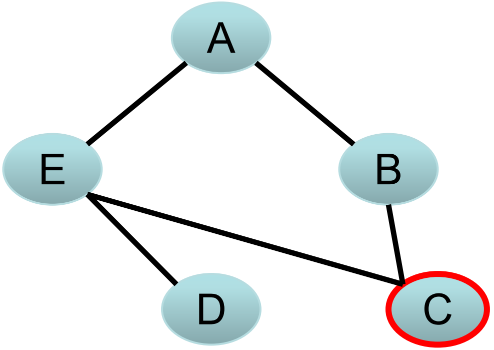
E.g. E-A-B-C is a cycle in the undirected graph

PATHS AND CYCLES



A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.

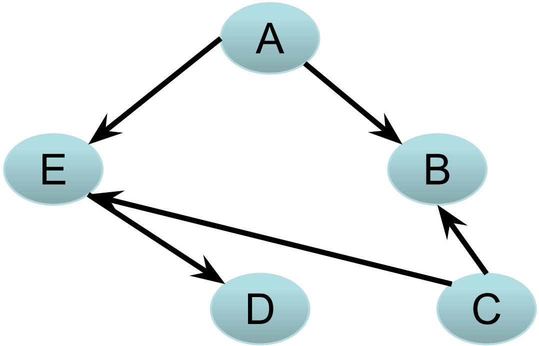
- If graph is directed the edge needs to be in the right direction.
- E.g. A-E-D is a path in both previous graphs



A **cycle** is a path where the start node is also the end node

- E.g. E-A-B-C is a cycle in the undirected graph

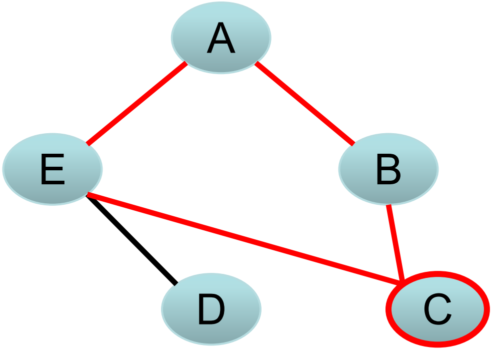
PATHS AND CYCLES



A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.

If graph is directed the edge needs to be in the right direction.

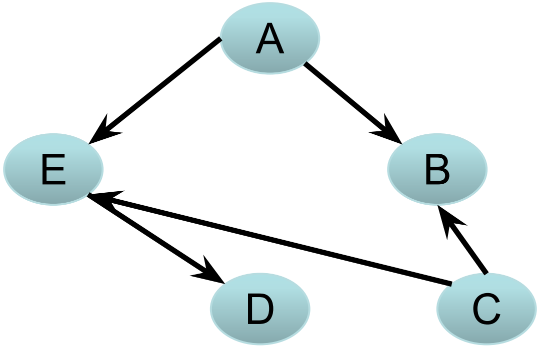
E.g. A-E-D is a path in both previous graphs



A **cycle** is a path where the start node is also the end node

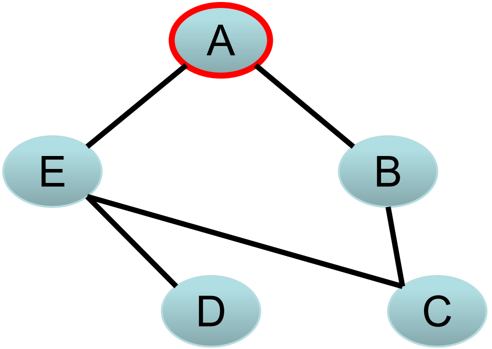
E.g. E-A-B-C is a cycle in the undirected graph

PATHS AND CYCLES



A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.

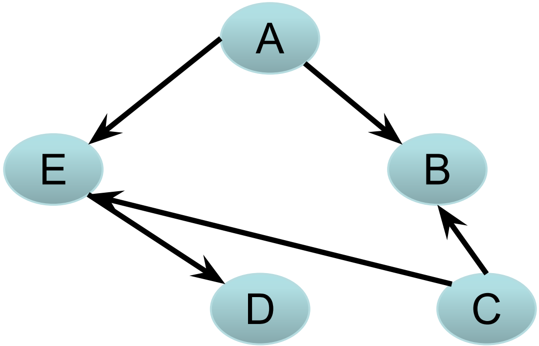
- If graph is directed the edge needs to be in the right direction.
- E.g. A-E-D is a path in both previous graphs



A **cycle** is a path where the start node is also the end node

- E.g. E-A-B-C is a cycle in the undirected graph

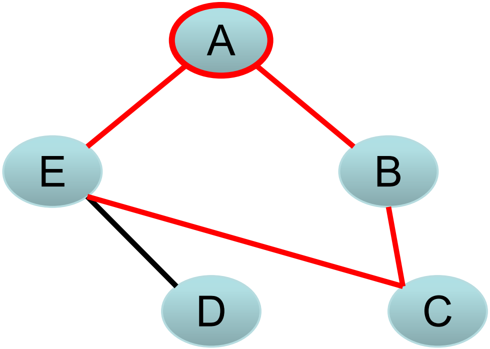
PATHS AND CYCLES



A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.

If graph is directed the edge needs to be in the right direction.

E.g. A-E-D is a path in both previous graphs



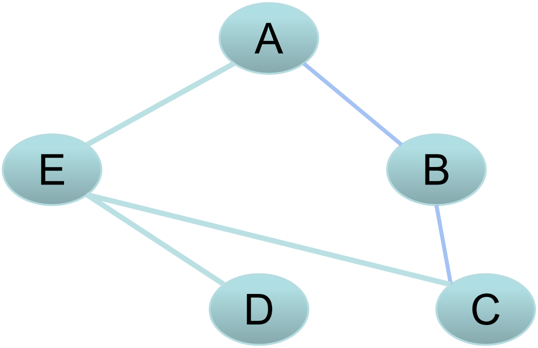
A **cycle** is a path where the start node is also the end node

E.g. E-A-B-C is a cycle in the undirected graph

PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

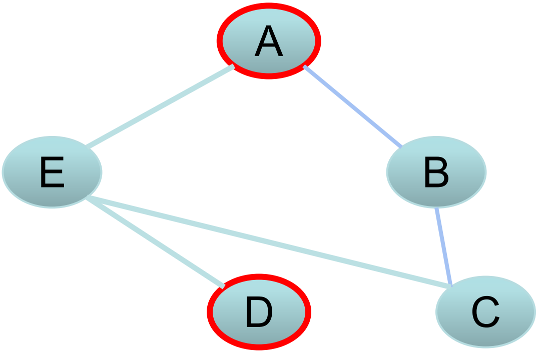
The **diameter** of the graph is the maximum distance between any pair of its nodes.



PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

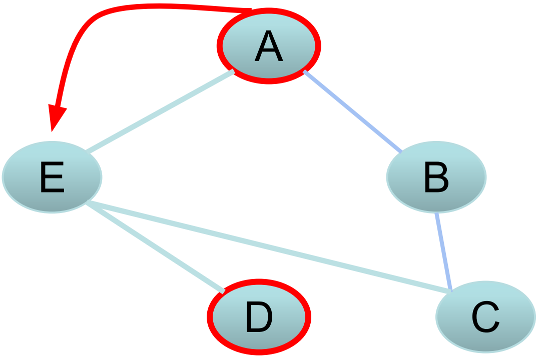
The **diameter** of the graph is the maximum distance between any pair of its nodes.



PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

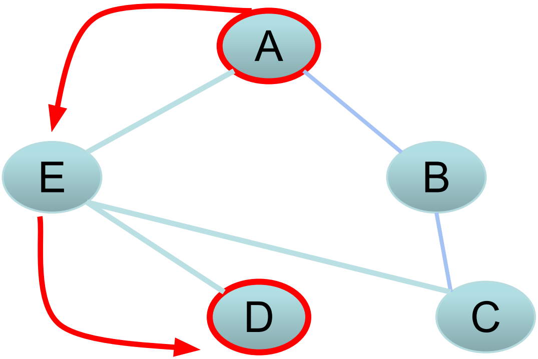
The **diameter** of the graph is the maximum distance between any pair of its nodes.



PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

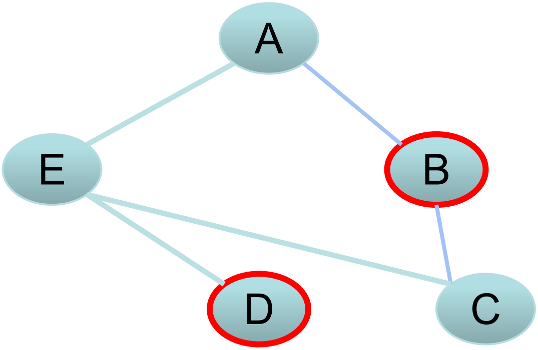
The **diameter** of the graph is the maximum distance between any pair of its nodes.



PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

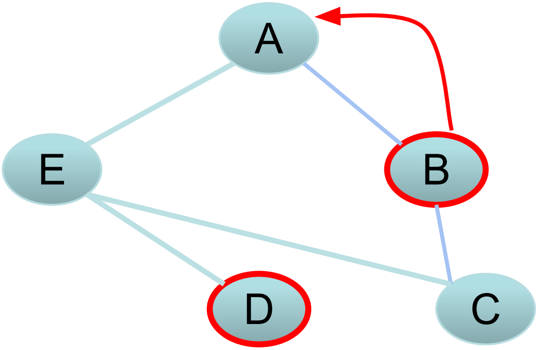
The **diameter** of the graph is the maximum distance between any pair of its nodes.



PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

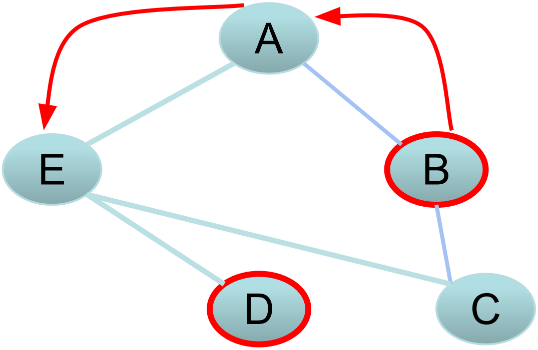
The **diameter** of the graph is the maximum distance between any pair of its nodes.



PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

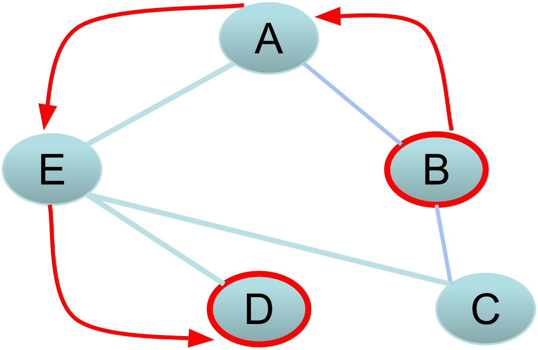
The **diameter** of the graph is the maximum distance between any pair of its nodes.



PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

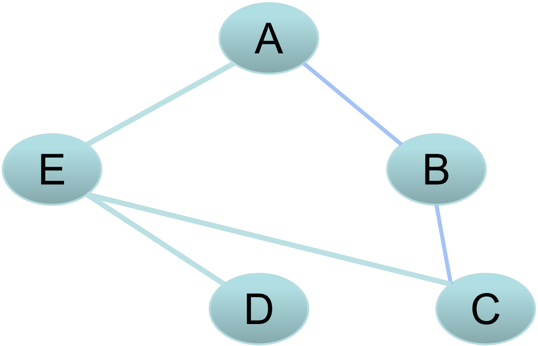
The **diameter** of the graph is the maximum distance between any pair of its nodes.



PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

The **diameter** of the graph is the maximum distance between any pair of its nodes.

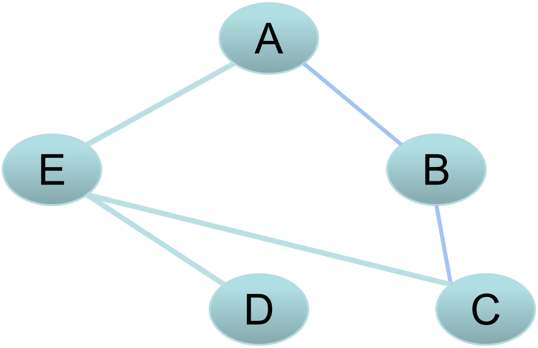


What’s the diameter?

PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

The **diameter** of the graph is the maximum distance between any pair of its nodes.



d(A,B) = 1

d(A,C) = 2

d(B,D) = 3

.

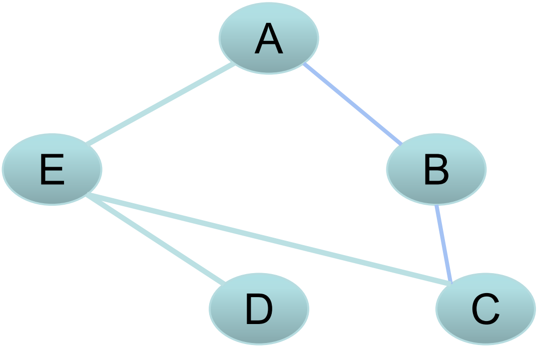
.

.

PATH LENGTH/DISTANCE

The **distance** (d) between two nodes in a graph is the length of the shortest path linking the two graphs.

The **diameter** of the graph is the maximum distance between any pair of its nodes.



d(A,B) = 1

d(A,C) = 2

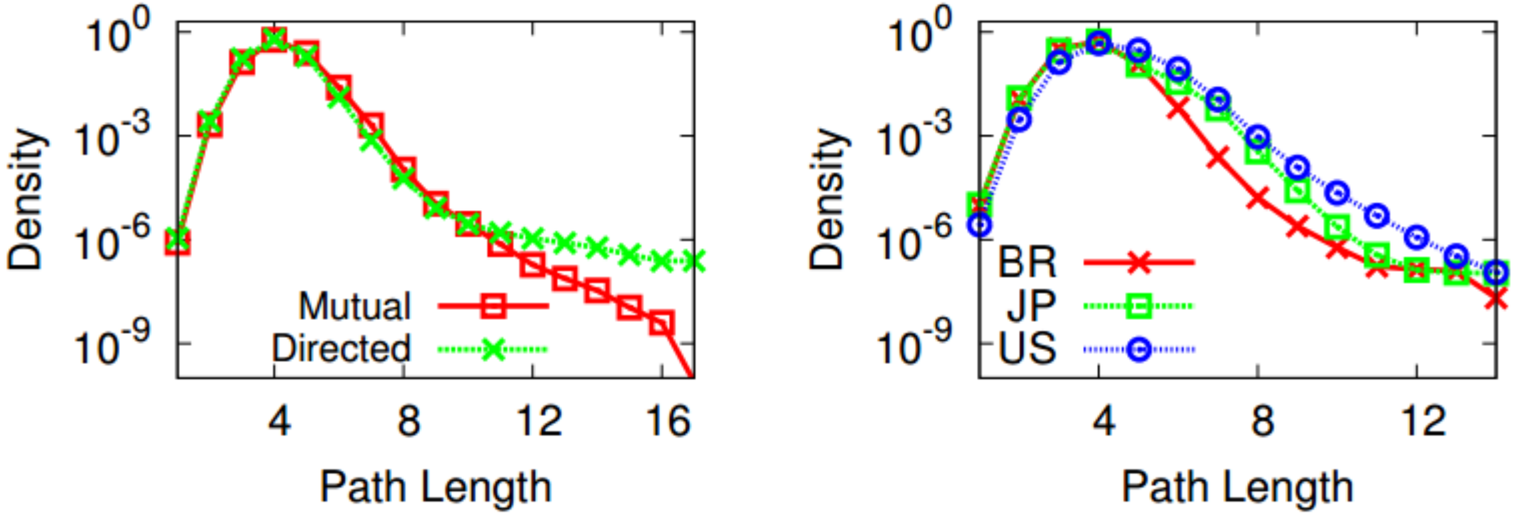
d(B,D) = **3**

.

.

.

PATH LENGTH/DISTANCE



CONNECTIVITY

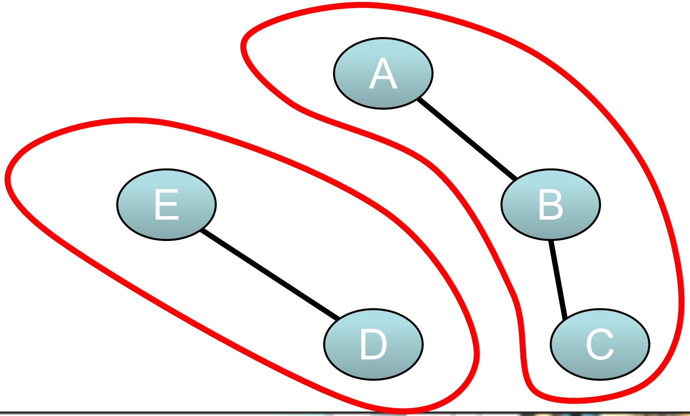
A graph is **connected** if there is a path between *each pair* of nodes.

Example of **disconnected** graph:

COMPONENTS

A **connected component** of a graph is the subset of nodes for which each of them has a path to all others (and the subset is not part of a larger subset with this property).

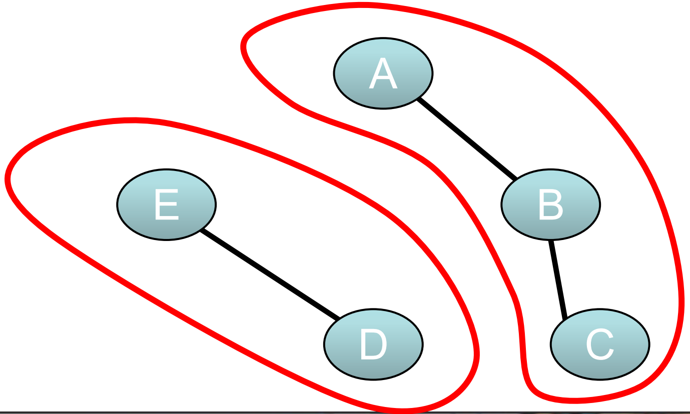
Connected components: A-B-C and E-D



COMPONENTS

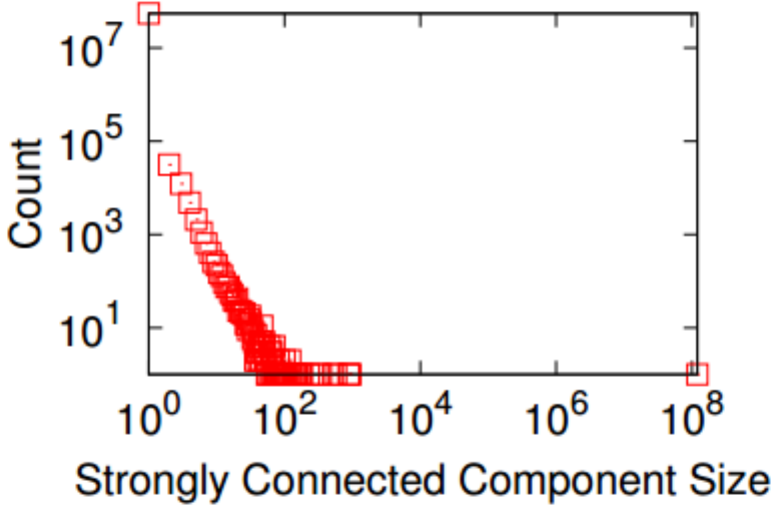
A **connected component** of a graph is the subset of nodes for which each of them has a path to all others (and the subset is not part of a larger subset with this property).

Connected components: A-B-C and E-D



What's the largest connected component?

COMPONENTS



Count
10^7
10^5
10^3
10^1
10^0
10^2
10^4
10^6
10^8
Strongly Connected Component Size

FORMALLY: CLUSTERING COEFFICIENT

Local Clustering Coefficient

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

Network Clustering Coefficient

$$CG = \frac{1}{N}\sum_i C_i$$

FORMALLY: CLUSTERING COEFFICIENT

Local Clustering Coefficient

$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

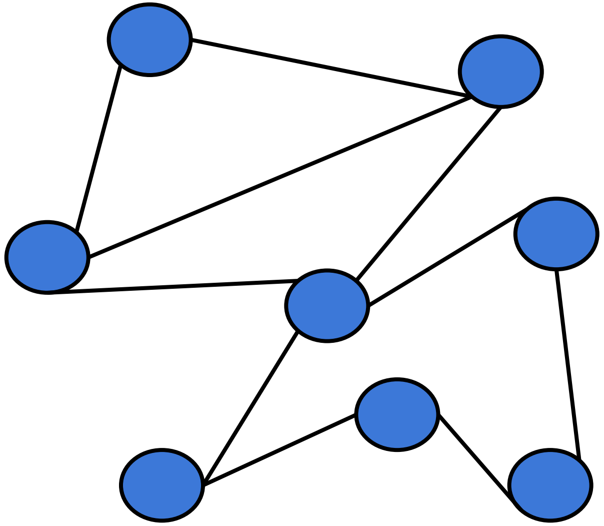
Proportion of my friends who are also friends with my other friends...

Network Clustering Coefficient

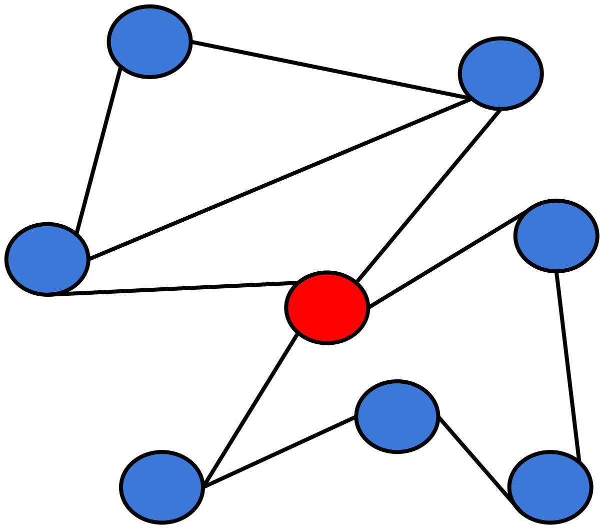
$$CG = \frac{1}{N}\sum_i C_i$$

The average all all the node’s local clustering coefficients

CLUSTERING COEFFICIENT: EXAMPLE

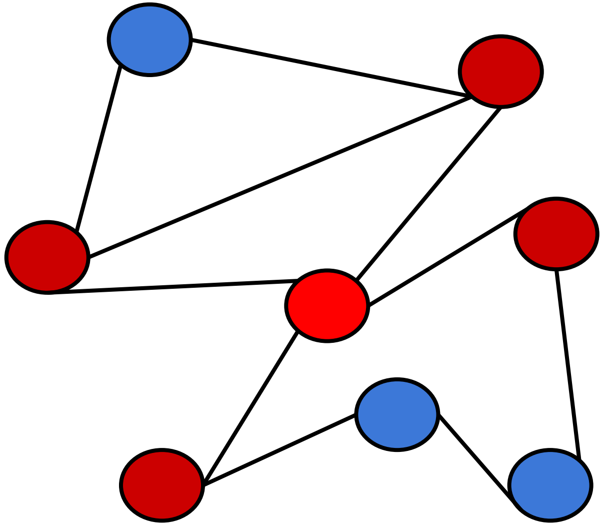


CLUSTERING COEFFICIENT: EXAMPLE



CLUSTERING COEFFICIENT: EXAMPLE

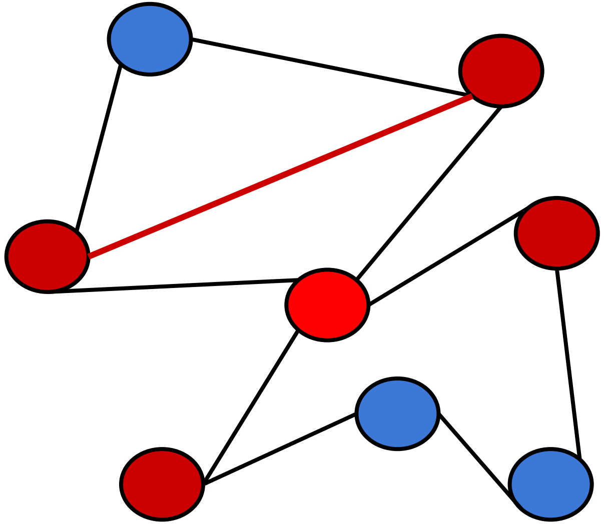
Degree = 4



CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

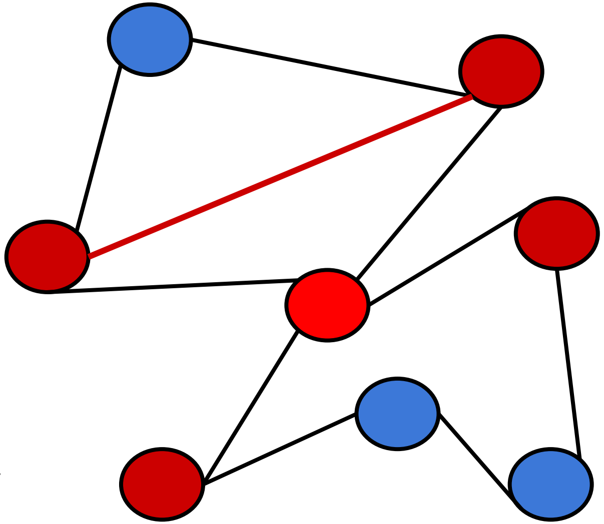


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

$$C_i = \frac{2\,|\,\{e_{jk}\}\,|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

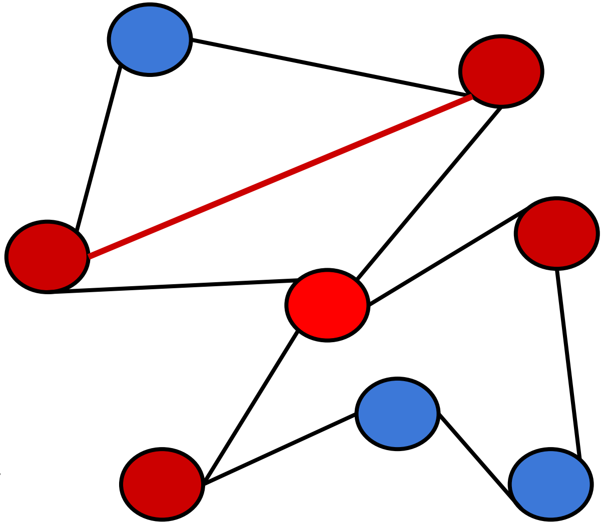


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

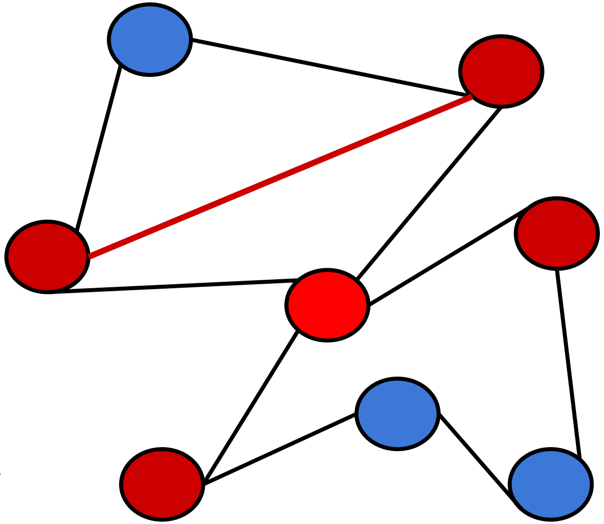


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{j,k} \in E$$



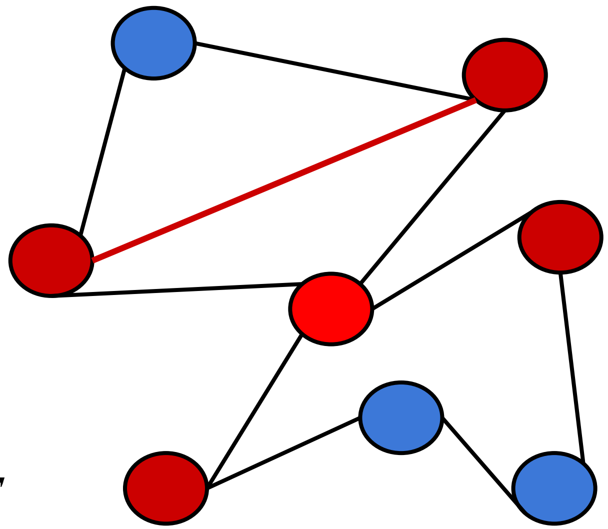
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

$$C_i = \frac{\boxed{2}\,|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

$$C_i = \frac{2}{\quad}$$



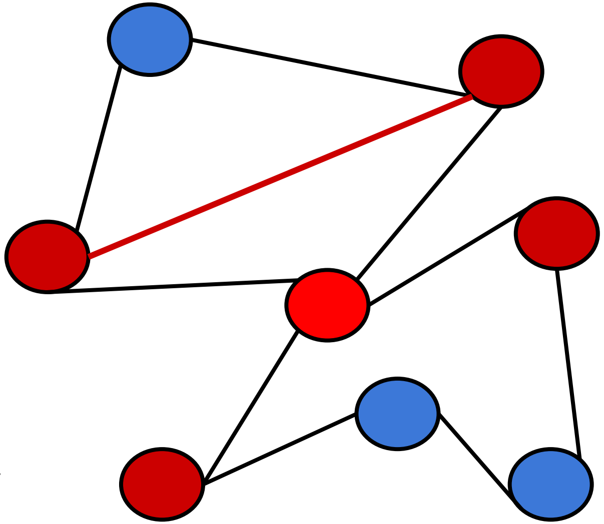
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

$$C_i = \frac{2\boxed{|\{e_{jk}\}|}}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

$$C_i = \frac{2 * 1}{}$$



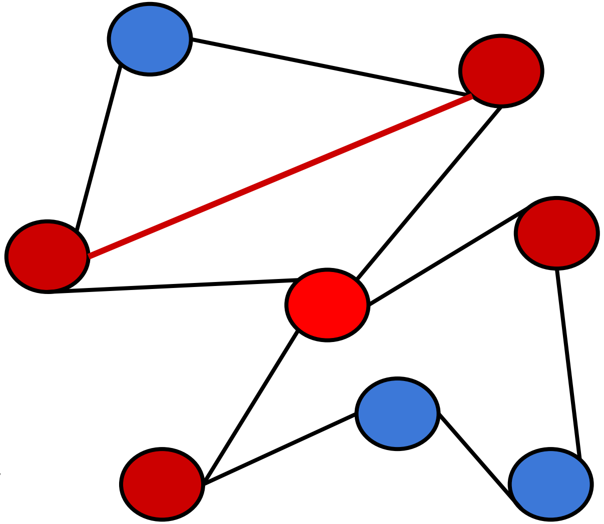
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

$$C_i = \frac{2\,|\{e_{jk}\}|}{\boxed{k_i}(k_i - 1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

$$C_i = \frac{2 * 1}{4}$$



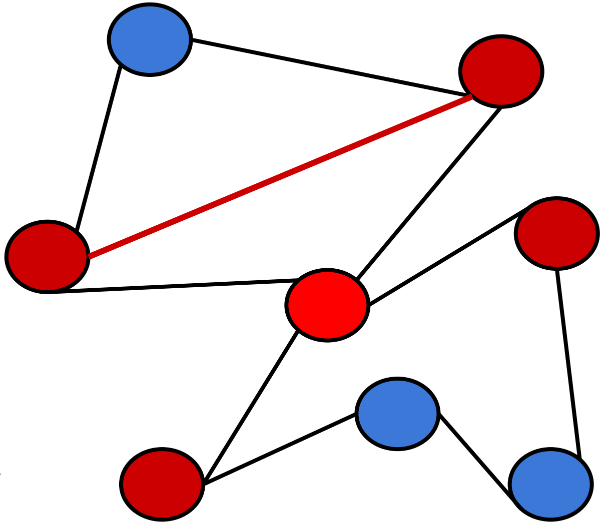
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

$$C_i = \frac{2|\{e_{jk}\}|}{k_i\boxed{(k_i - 1)}} : v_j, v_k \in N_i, e_{j,k} \in E$$

$$C_i = \frac{2 * 1}{4 * 3}$$

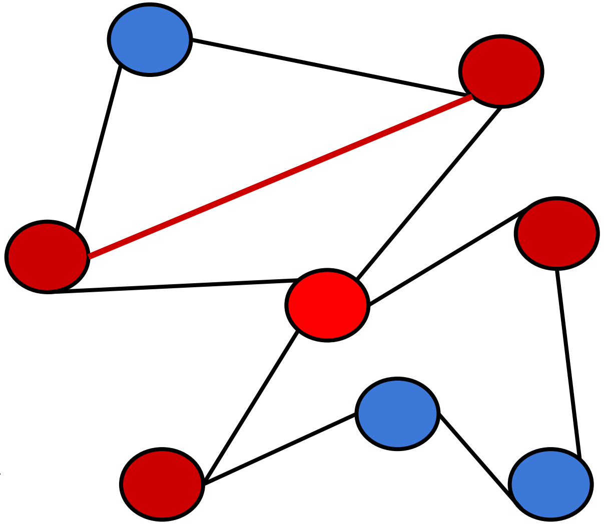


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{j,k} \in E$$



$$C_i = \frac{2}{12}$$

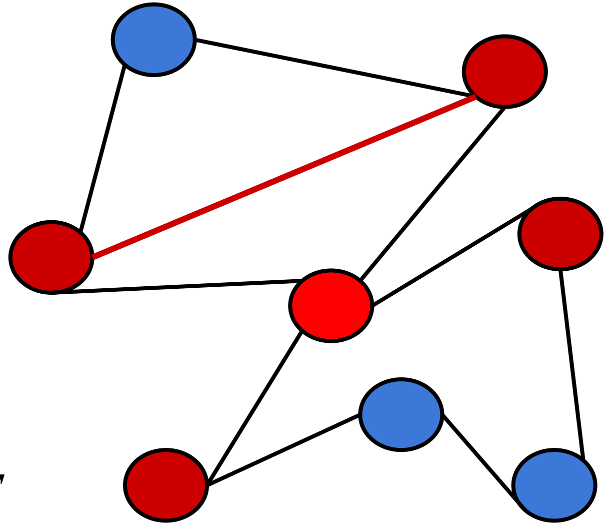
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

$$C_i = \frac{1}{6}$$

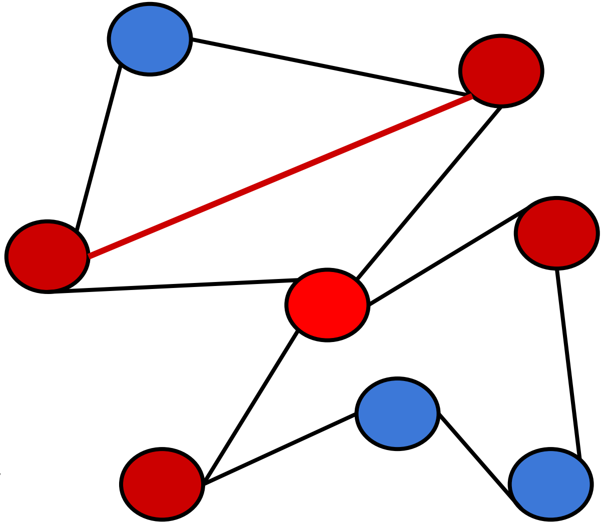


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{j,k} \in E$$



$$C_i = \frac{1}{6}$$

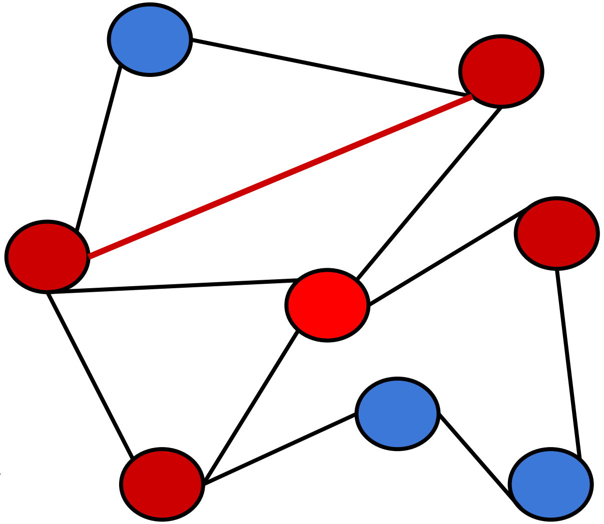
Fraction of possible interconnections between my neighbour!

CLUSTERING COEFFICIENT: EXAMPLE

Degree =

Links between neighbours =

$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

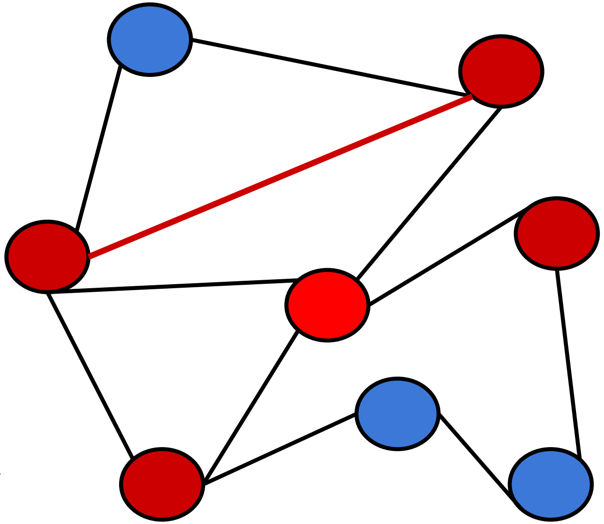


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours =

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

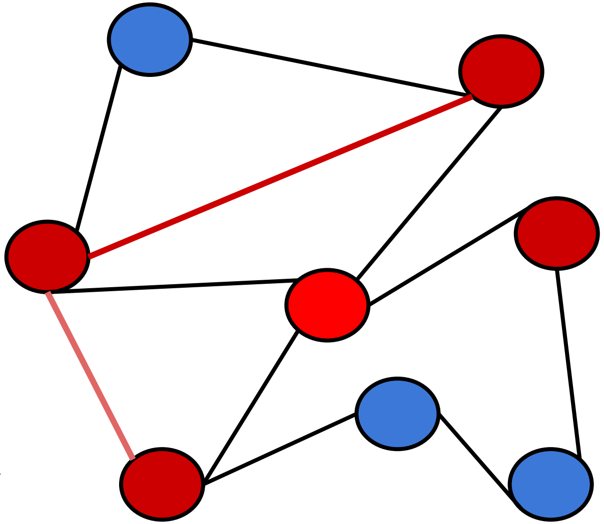


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 2

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{j,k} \in E$$



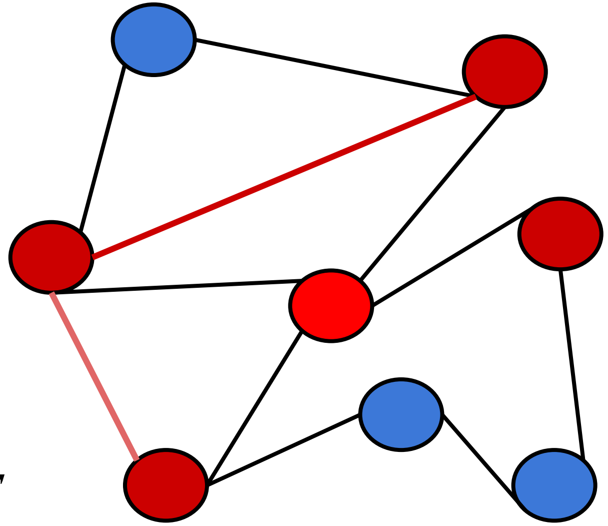
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 2

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{j,k} \in E$$

$$C_i = \frac{2 * 2}{4 * 3}$$

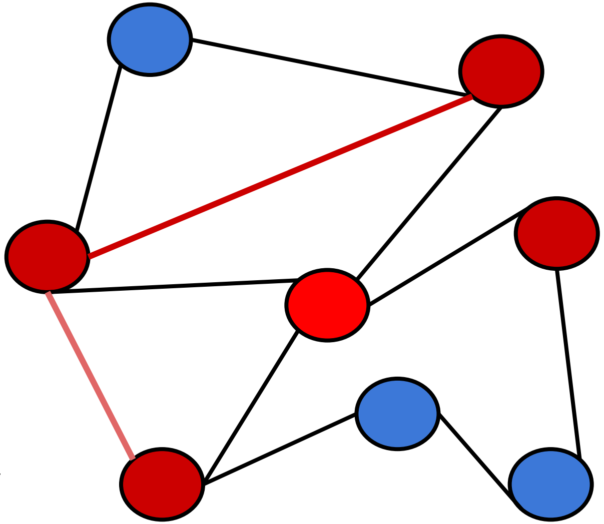


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 2

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{j,k} \in E$$



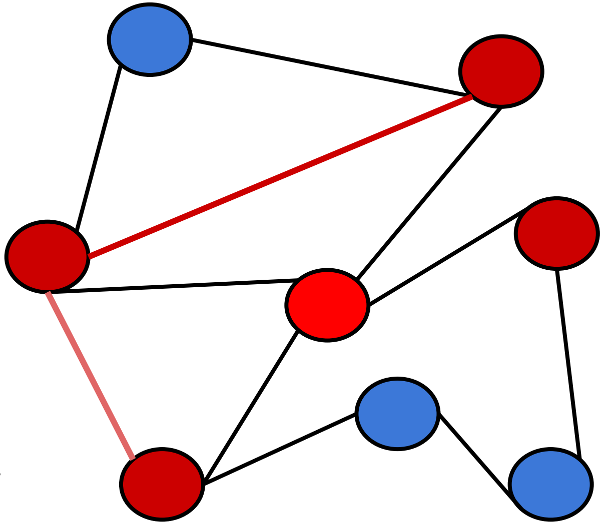
$$C_i = \frac{4}{12}$$

CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 2

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{j,k} \in E$$



$$C_i = \frac{1}{3}$$

CLUSTERING COEFFICIENT: EXAMPLE

Average clustering coefficient for degree = 5 is about **0.4** in the Facebook graph

compared to Twitter mutual graph **0.23**