Digital Media and Social Networks

Laurissa Tokarchuk

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

WEEK 1: NETWORKS, RANDOM GRAPHS AND METRICS

ABOUT ME

Email: laurissa.tokarchuk@qmul.ac.uk

Office: CS412 (book appointment: https://laurissat.youcanbook.me)

About Me | Research | Stuff.jar | Teaching | Activities | Publications

Laurissa Tokarchuk

Profile

Laurissa Tokarchuk
Lecturer
Member of: Cognitive Science and The Centre for Intelligent Sensing (CIS)
School of Electronic Engineering and Computer Science
Mile End Road, E1 4NS.

Email: laurissa dot tokarchuk at qmul.ac.uk

Research Interests

Mobile Sensing, Social Computing, Social Sensing, Mobile Gaming, Recommendation.
Examples and things I have been up to: Stuff.jar

Interested Students:

Please take a look at the topics of my current research students and examples of things I have been up to: Stuff.jar. I welcome any proposals related to my research from prospective students and am happy to discuss potential funding opportunities.

News:

- Nov 2017: Our paper, Piecing together the puzzle: Improving event content coverage for real-time sub-event detection using adaptive microblog crawling. PLOS ONE
- Oct 2017: Our paper titled Effects of Valence and Arousal on Working Memory Performance in Virtual Reality Gaming has been presented at ACII 2017 –Int'l Conf. on Affective Computing and Intelligent Interaction.

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

QMplus

ECS637U/ECS757P - DIGITAL MEDIA AND SOCIAL NETWORKS - 2017/18

> ECS637U/ECS757P - Digital Media and Social Networks - 2017/18

The fast rise in adoption of Online Social Networks (OSN) and digital media has evolved the way users interact on the Internet, in a manner that most personal communication is now taking place via such tools. The adoption of services such as Facebook, Twitter and YouTube affect the traffic patterns on the Internet as well. Recently, there have been a large number of studies on measurement and analysis of user connectivity, data sharing and traffic patterns on the Internet focusing on OSNs. This course covers different aspects of the concepts of OSN, recommender systems, user behavior, advertising and privacy. In this module, I aim to introduce the concepts around measurement, analysis, usability and privacy aspects of OSNs.

Planned schedule (subject to amendments):

	Topic	Lecturer
Lecture 1	Networks and Random Graphs	Laurissa Tokarchuk
Lecture 2	Small World and Weak Ties	Laurissa Tokarchuk
Lecture 3	Network Centrality and Applications	Minos Katevas
Lecture 4	Practical Network Analysis Tutorial	Minos Katevas
Lecture 5	Community Detection & Overlapping Communities	Laurissa Tokarchuk
Lecture 6	Structure of the Web, Search and Power Laws	Laurissa Tokarchuk
	REVISION WEEK	
Lecture 7	Cascades and Behaviour Influence	Laurissa Tokarchuk
Lecture 8	Epidemic Spreading and Mobility	Laurissa Tokarchuk
Lecture 9	Online Advertising	Minos Katevas
Lecture 10	Privacy and Ethics	Minos Katevas
Lecture 11	Revision	Minos Katevas

STRUCTURE OF THE SESSIONS

One lecture per week (two hour)

- Learn **fun**damental concepts of digital media and social networks
- Monday 15:00 - 17:00
- Location: Bancroft:Mason-LT

One seminar per week (one hour)

- Learn research aspects of digital media and social networks
- Labs/tutorials on courework
- Guest lectures by researchers in the field
- Tuesday 17:00 - 18:00
- Location: Bancroft:Mason-LT

ASSESSMENT

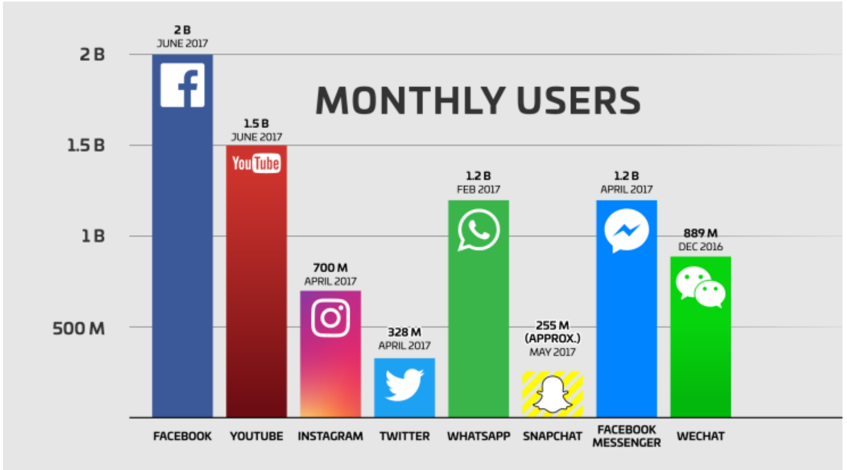
All information on the course page. All questions should be posted to the forums there and **not** emailed directly to the lecturers.

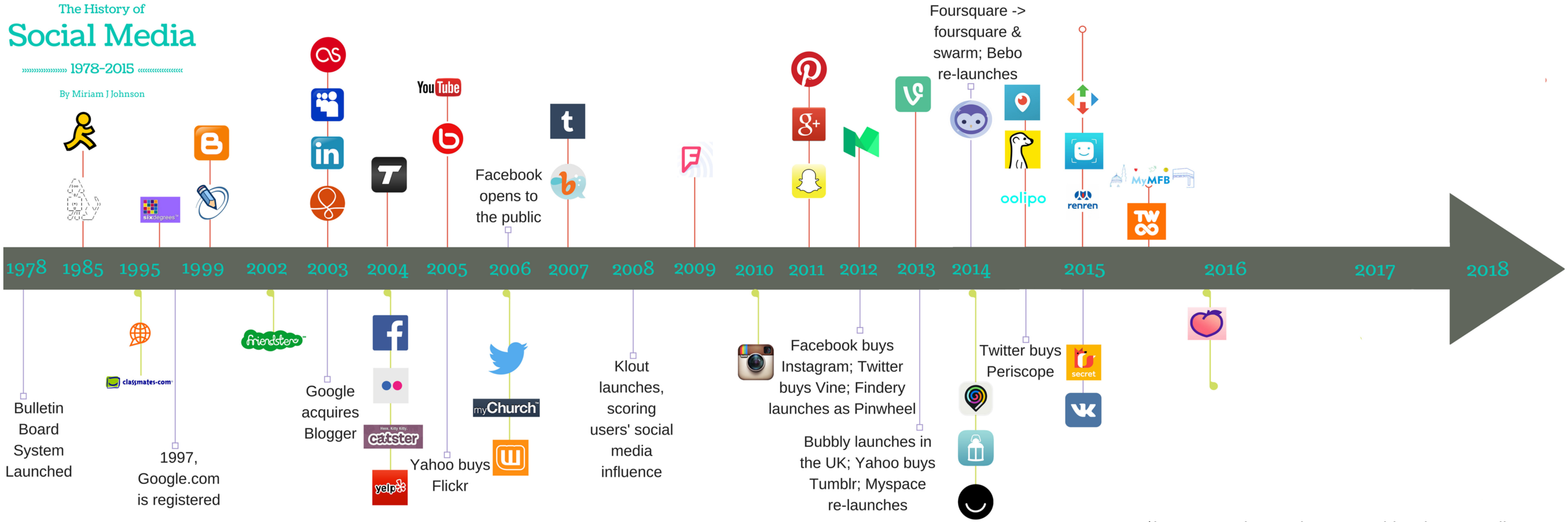
Two assignments due at the end of the term (TBA) **One report**(of approximately 1,000 words) on collection of profiles info and tweets from your user account using the twitter API. Worth 15% of the final mark.

The second assignment will consist of **analysis of the collected dataset according to some indicated network measures**: the analysis should be reported in a document of about 1,000 words where the results are commented and justified. Worth 15% of the final mark.

And… Final Exam in May (?) worth 70%.

HOW DID WE GET THERE?





*images techcrunch.com and booksaresocila.com

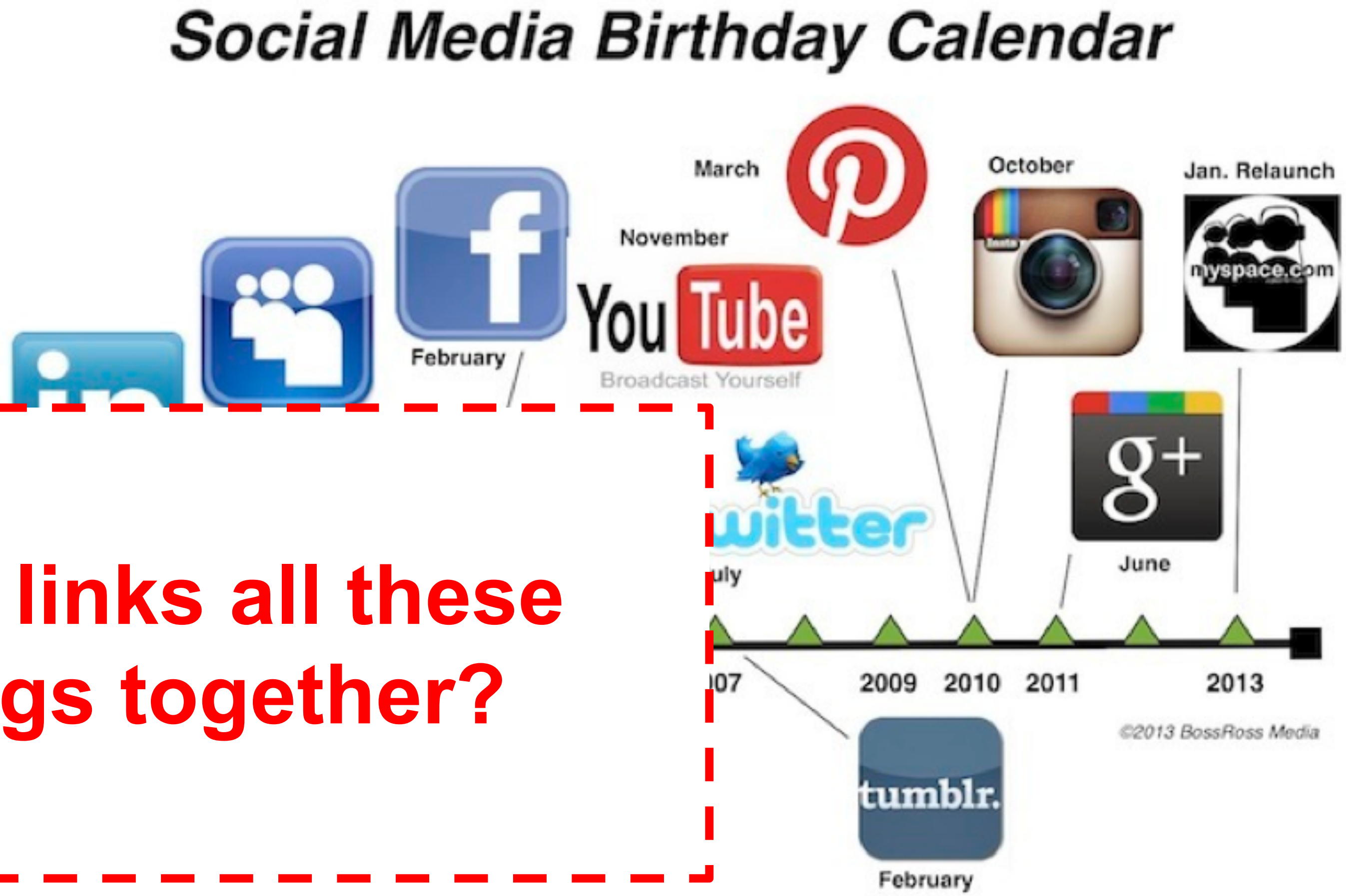
HOW DID WE GET THERE?

https://www.youtube.com/watch?v=95-yZ-31j9A

https://www.youtube.com/watch?v=X84muuaySVQ

BBS (BULLETIN BOARD SYSTEMS) AND IRC (INTERNET RELAY CHAT)

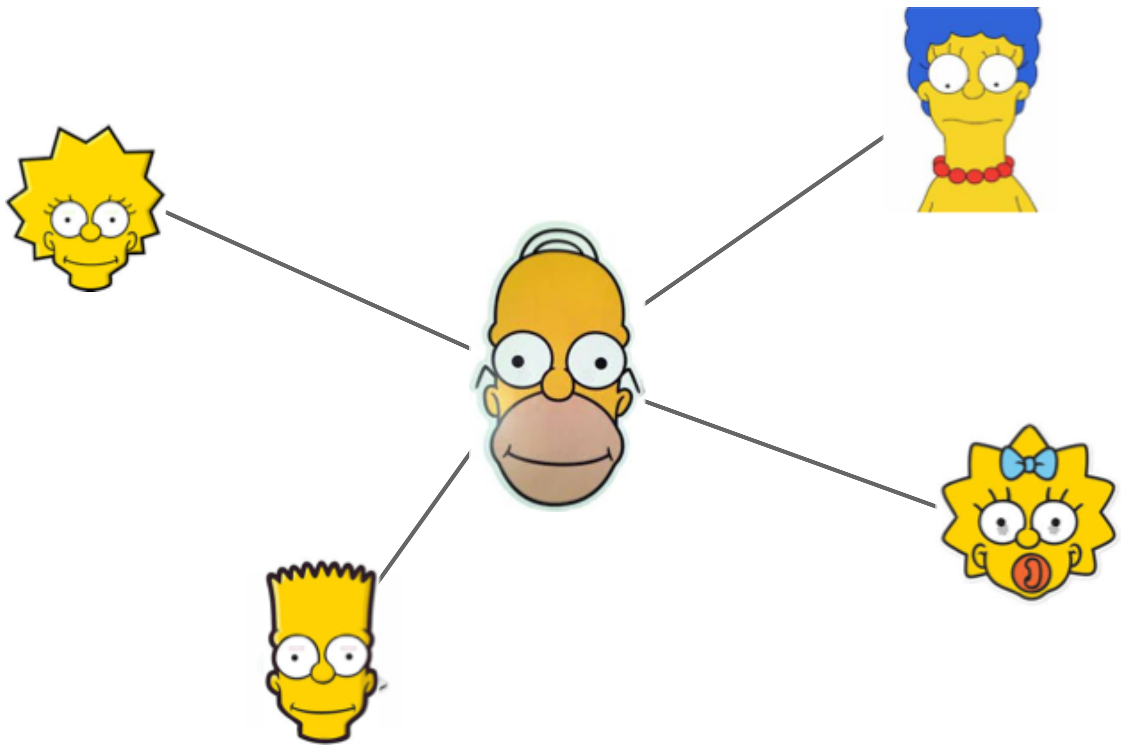
HOW DID WE GET THERE?



What links all these things together?

NETWORKS ARE EVERYWHERE

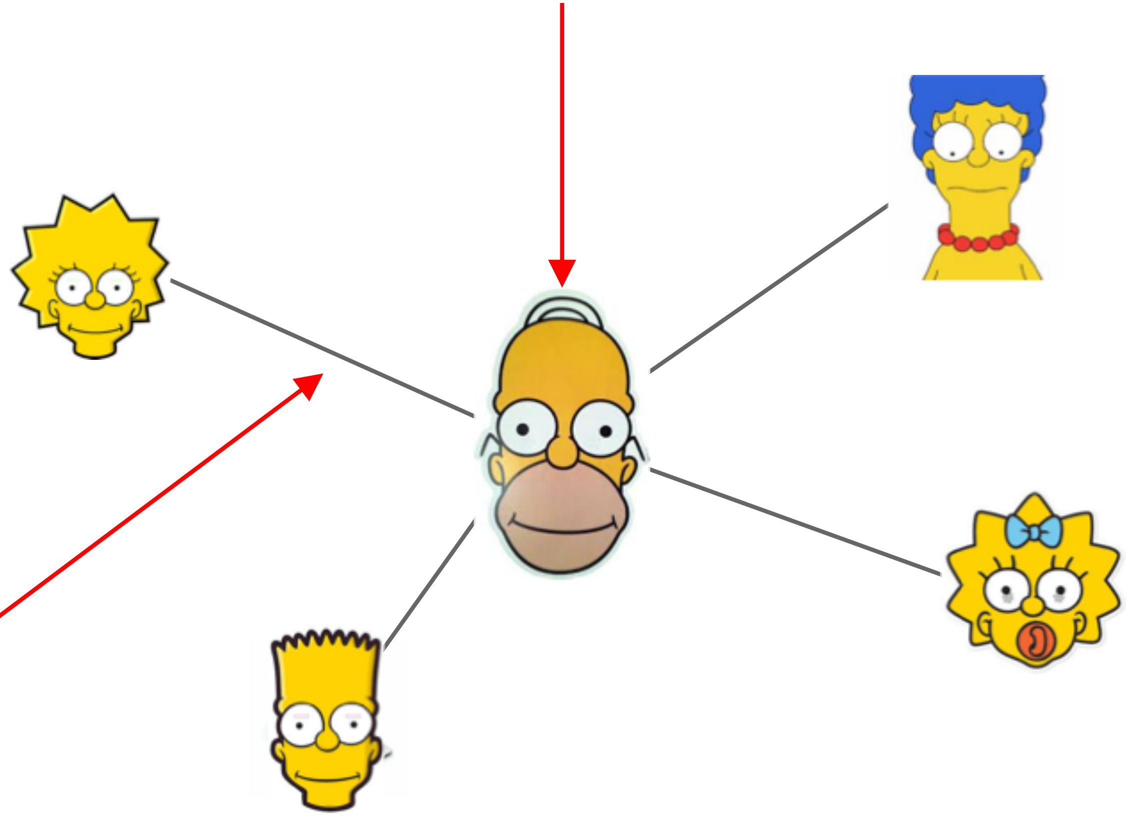
facebook



NETWORKS ARE EVERYWHERE

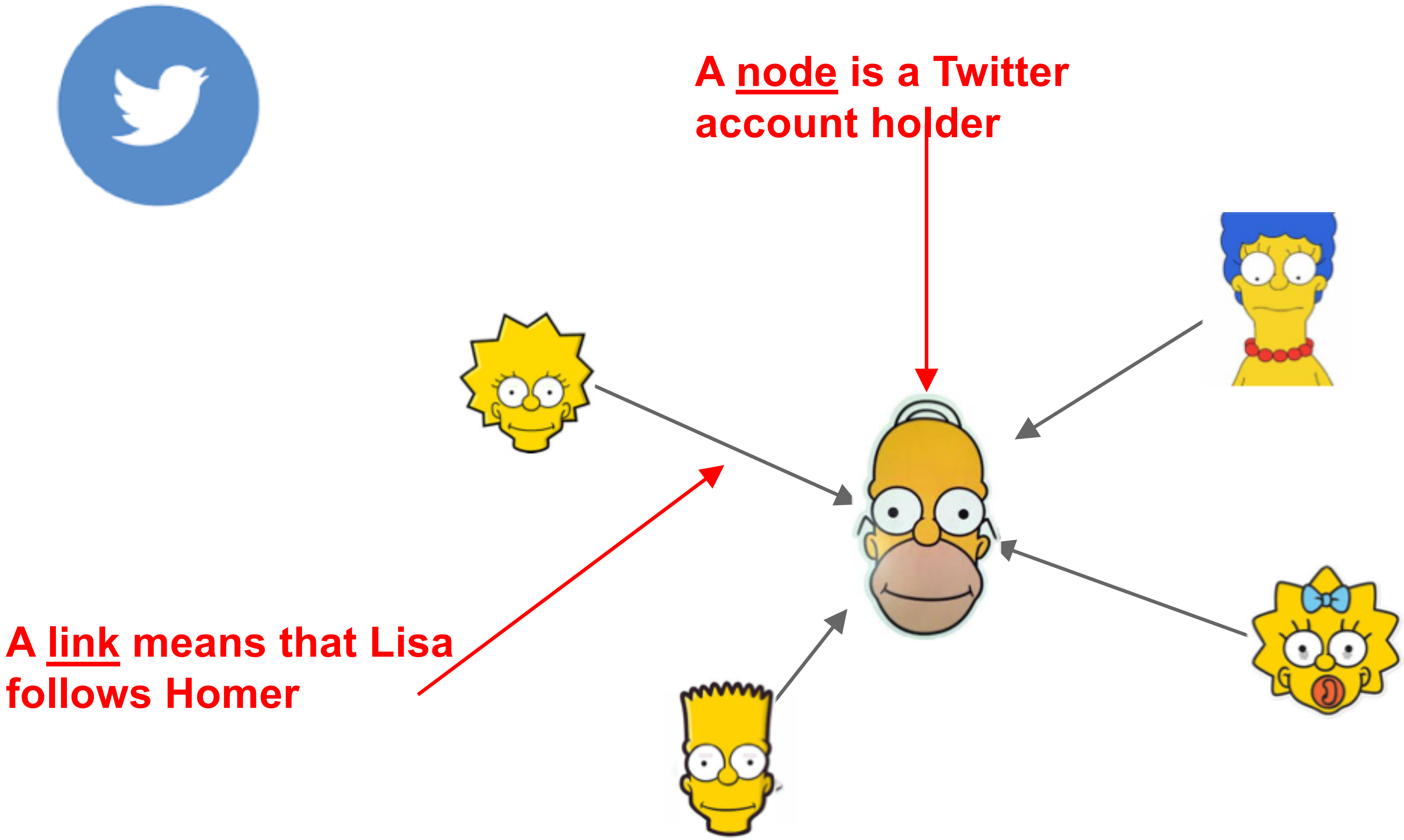
facebook

A node is a person



A link means that the two people are friends

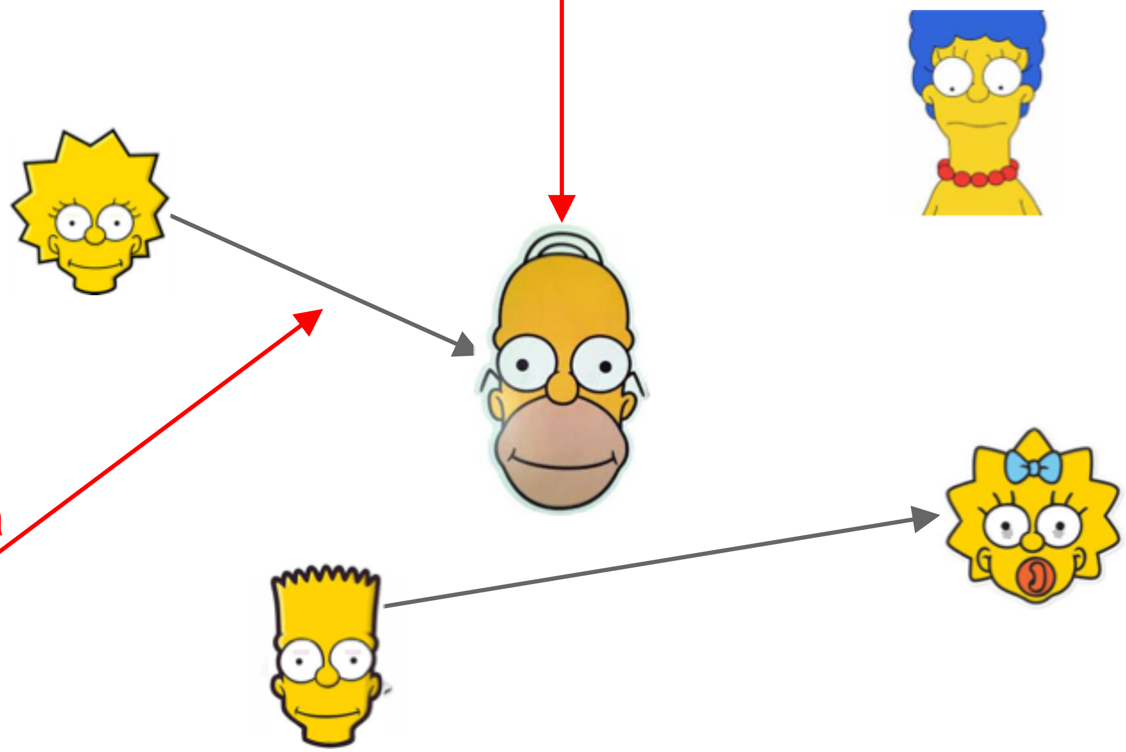
NETWORKS ARE EVERYWHERE



NETWORKS ARE EVERYWHERE

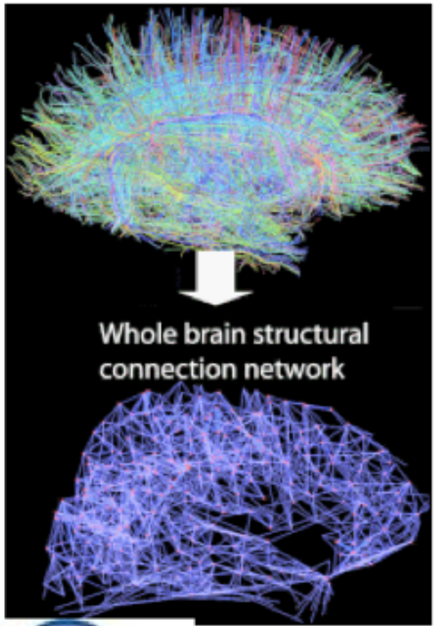
You Tube

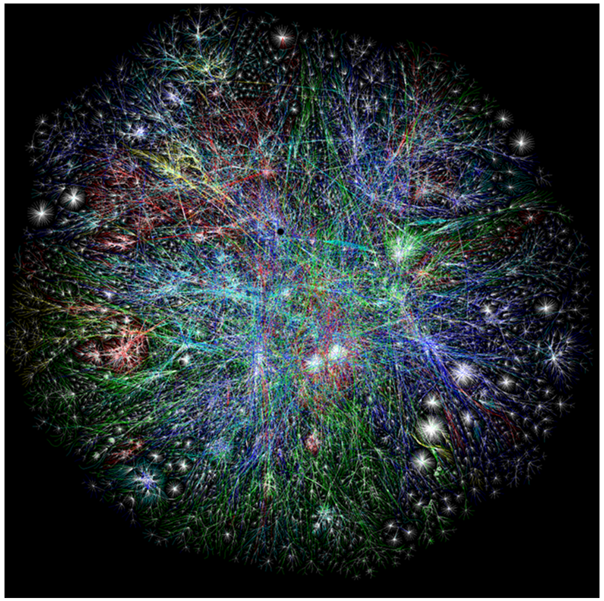
A **node** is a YouTube publisher



A **link** means that Lisa has watched a video of Homer

NETWORKS ARE EVERYWHERE



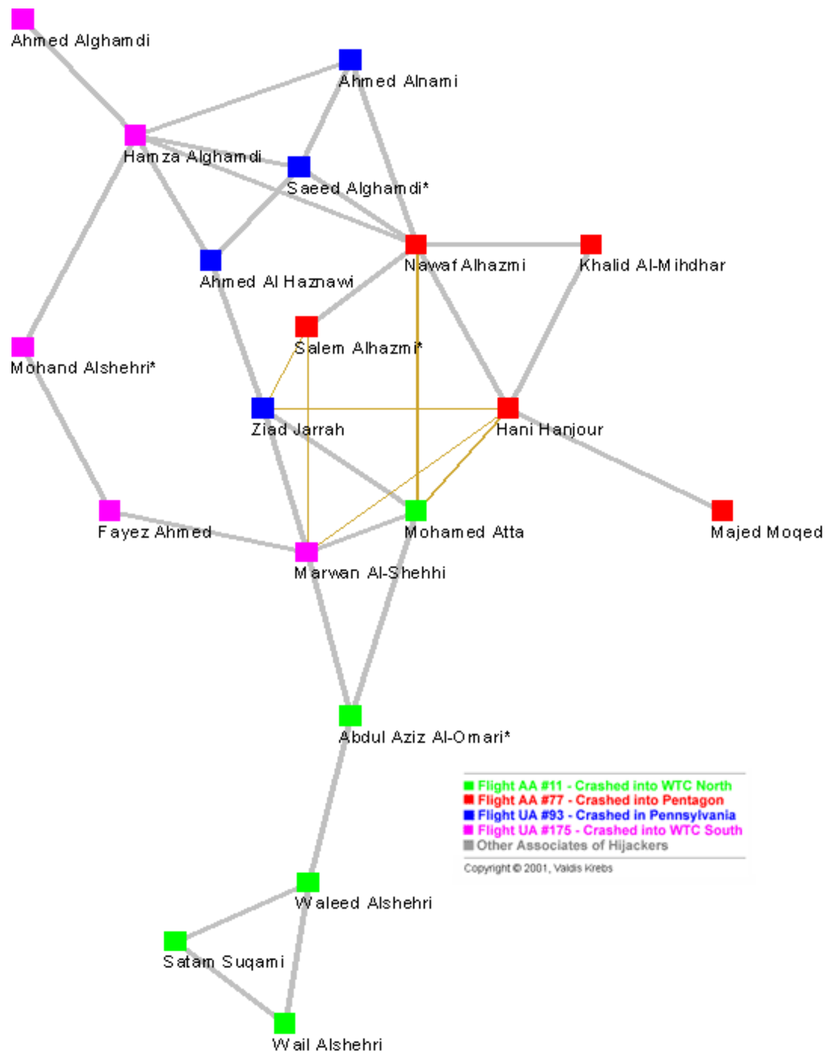


FACEBOOK FRIENDSHIP NETWORK

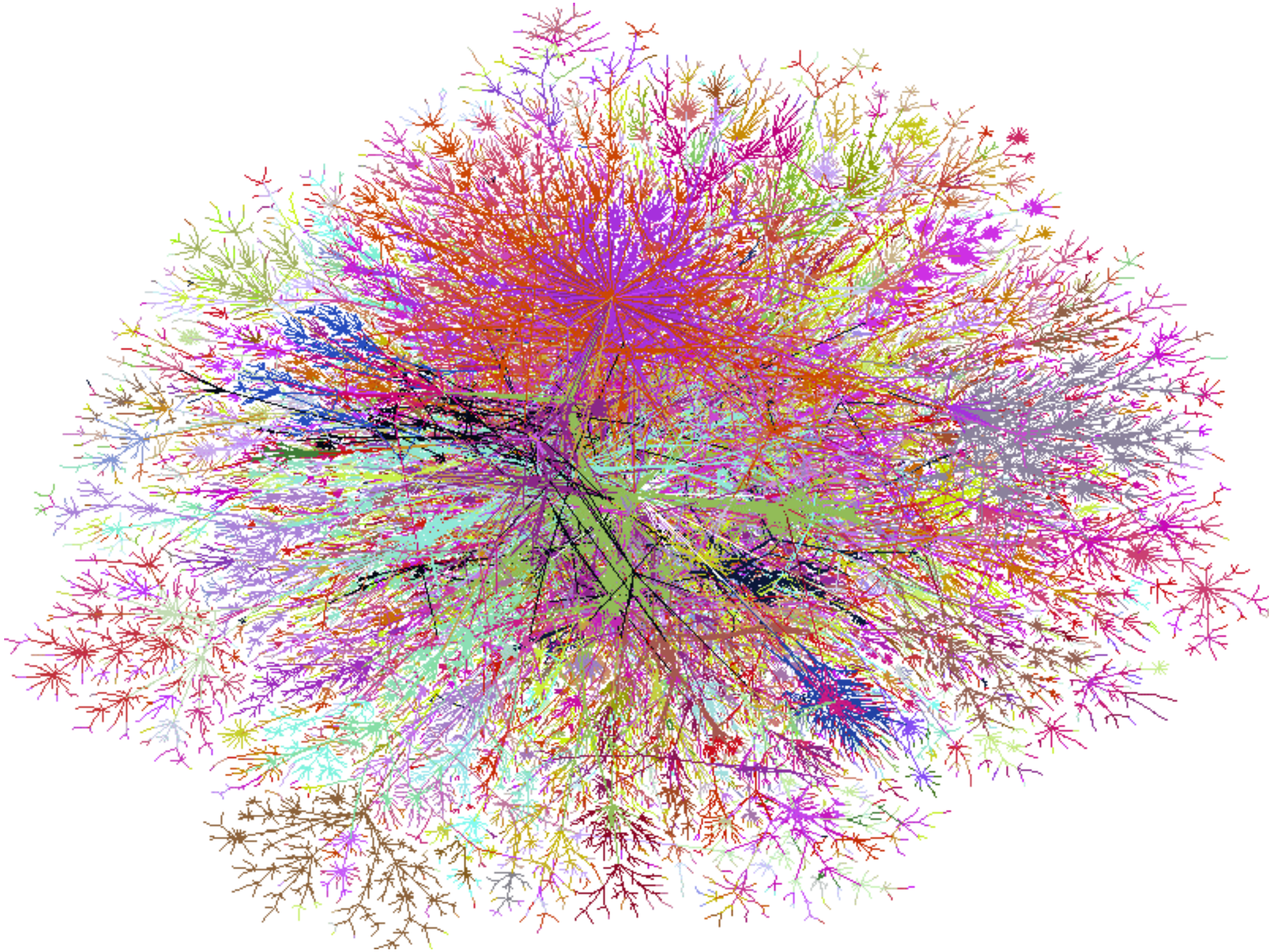
TERRORIST NETWORK

“SIX DEGREES OF MOHAMMED ATTA”

UNCLOAKING TERRORIST NETWORKS, BY VALDIS KREBS

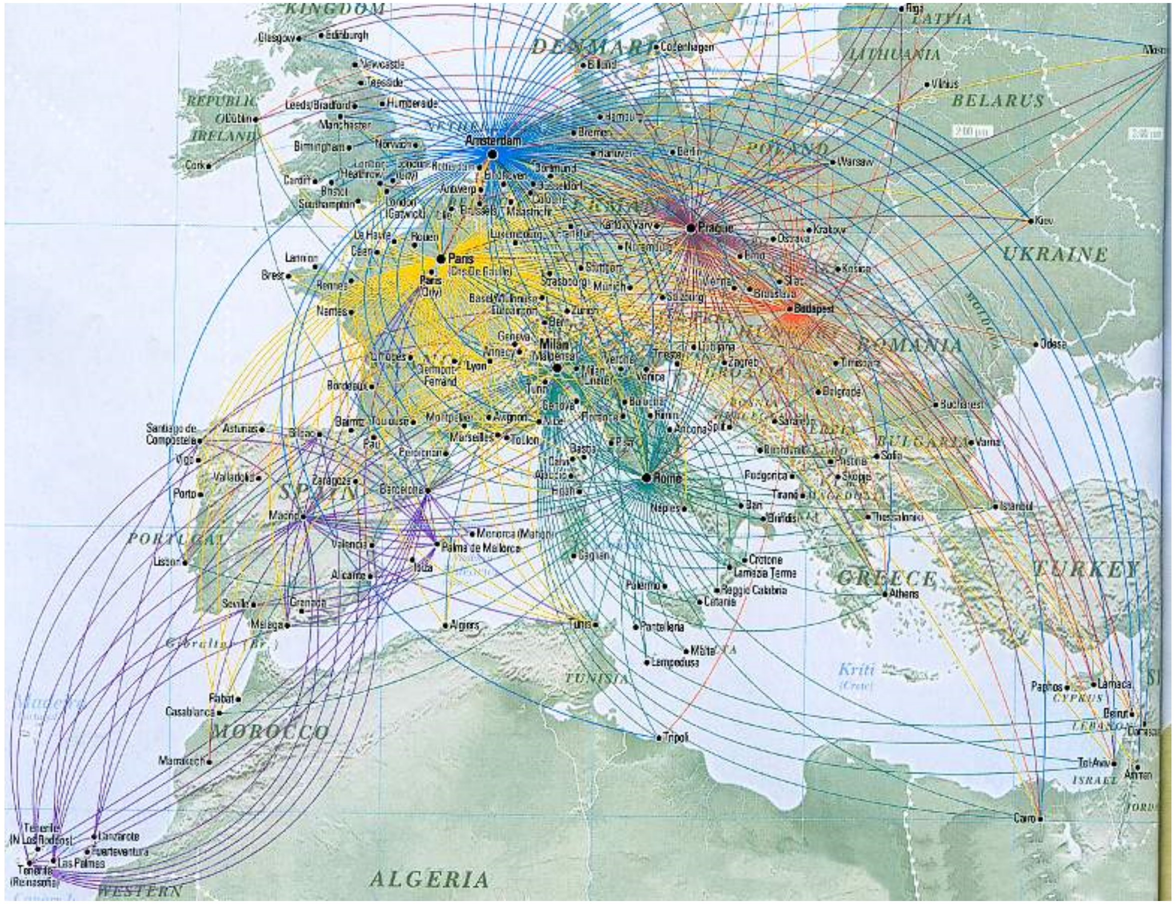


THE INTERNET



Source: Bill Cheswick http://www.cheswick.com/ches/map/gallery/index.html

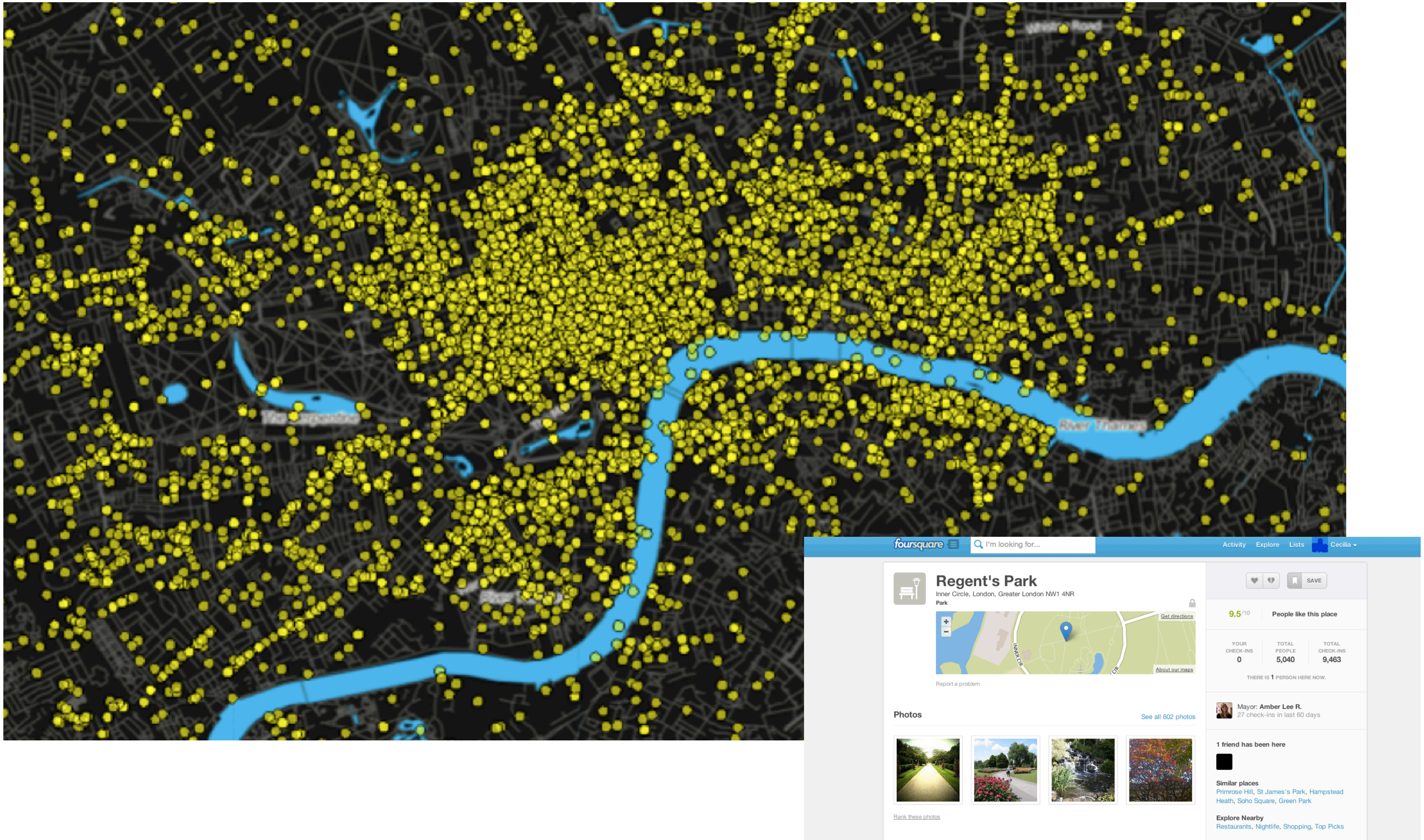
AIRLINE NETWORK



Source: Northwest Airlines WorldTraveler Magazine

RAILWAY/METRO NETWORK

GEO-SOCIAL NETWORKS



WHAT KIND OF NETWORKS?

- Who talks to whom?
- Who is friend with whom?
- What leads to what?
- Who is a relative of whom?
- Who eats whom?
- Who sends messages to whom?

IN THIS COURSE

- We will study the models and metrics which allow us to understand these phenomena.

- We will show analysis over large datasets of real social and technological networks.

- Mondays Lectures, Tuesday seminars both days mandatory!!
  (check qmplus page for seminar schedule – updated regularly)

- USE THE FORUM! NOT EMAIL!

IN THIS LECTURE

In this lecture, we will introduce:

- Networks/graphs
- Basic network measures
- Random Graphs
- Examples

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.

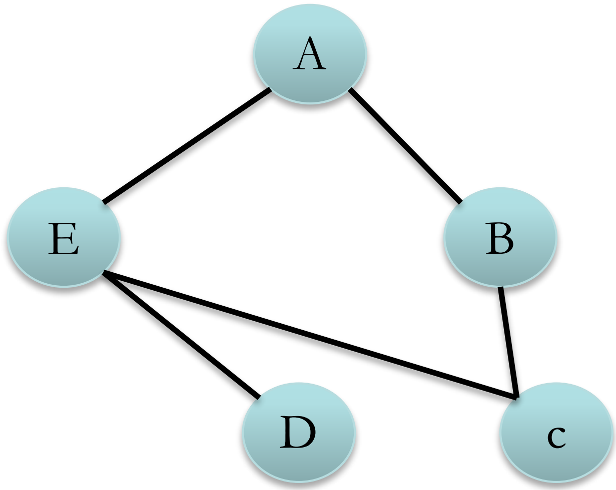
A **neighbour set** N(v) is the set of vertices adjacent to v:

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

NODE DEGREE

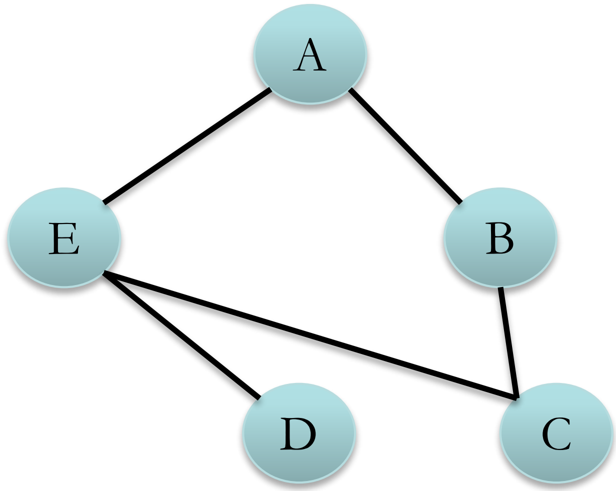
The **node degree** is the number of neighbours of a node

E.g., Degree of A is 2

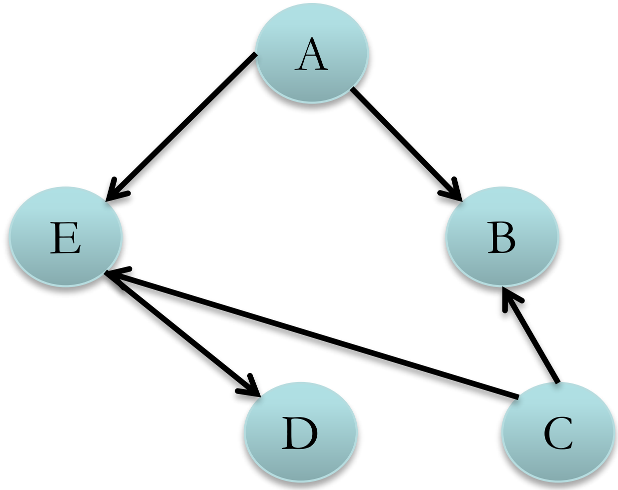


The study of the degree distribution of networks allows the classification of networks in different categories

DIRECTED & UNDIRECTED GRAPHS



Undirected Graph

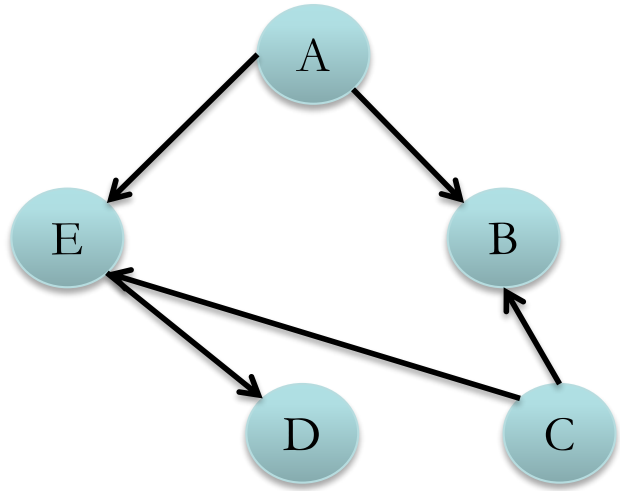


Directed Graph

Example of Undirected Graphs: Facebook, Co-presence

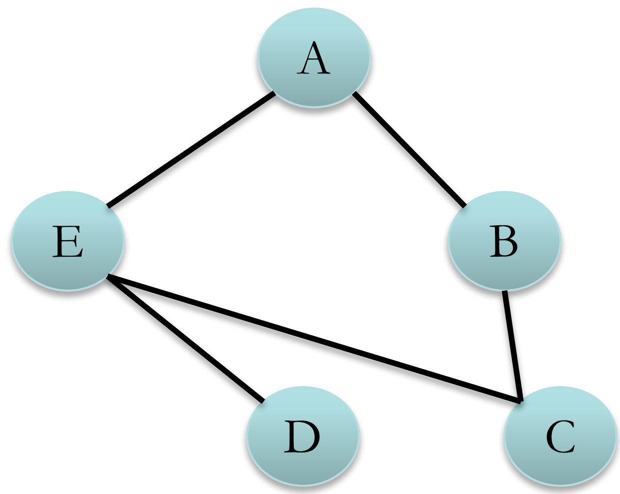
Examples of Directed: Twitter, Email, Phone Calls

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



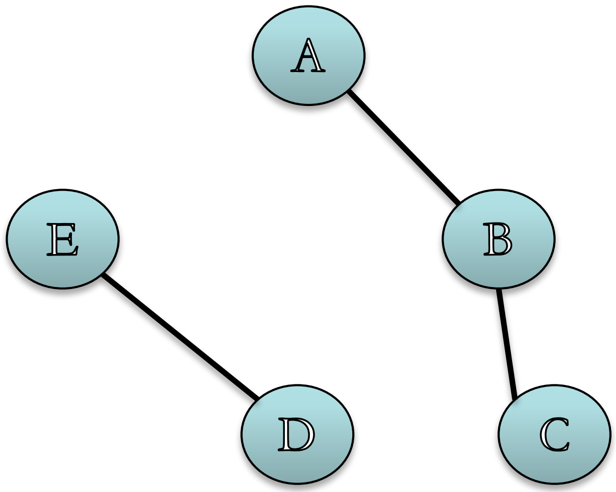
A **cycle** is a path where the start node is also the end node

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

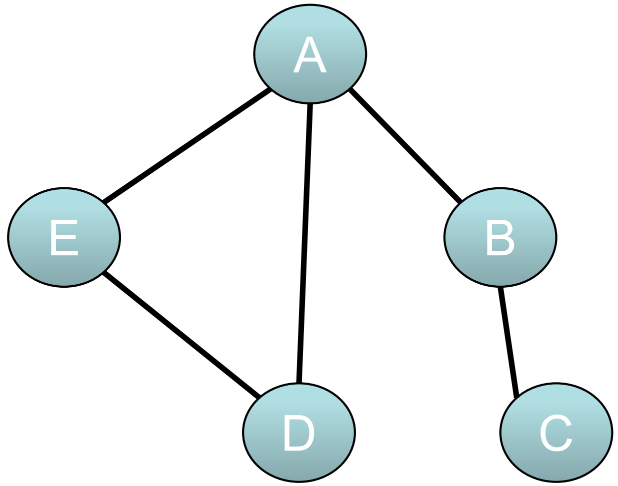
CONNECTIVITY

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

Example of **disconnected** graph:



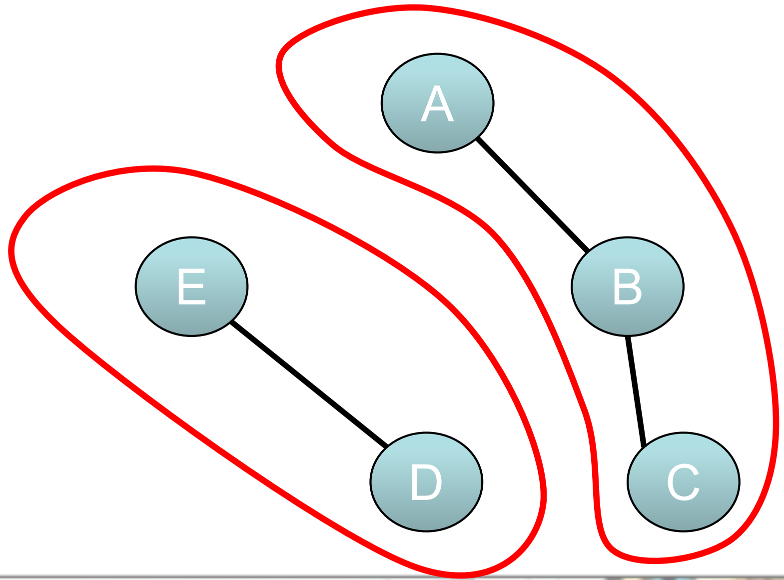
Is Facebook connected?



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).

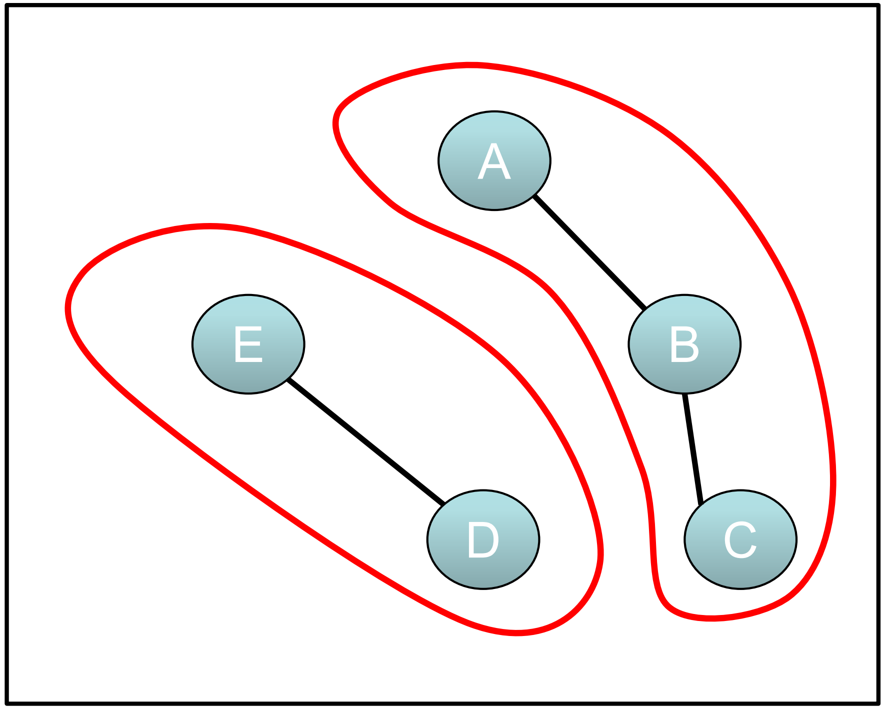
Connected components: A-B-C and E-D



COMPONENTS

A **giant component** is a connected component containing a significant fraction of nodes in the network.

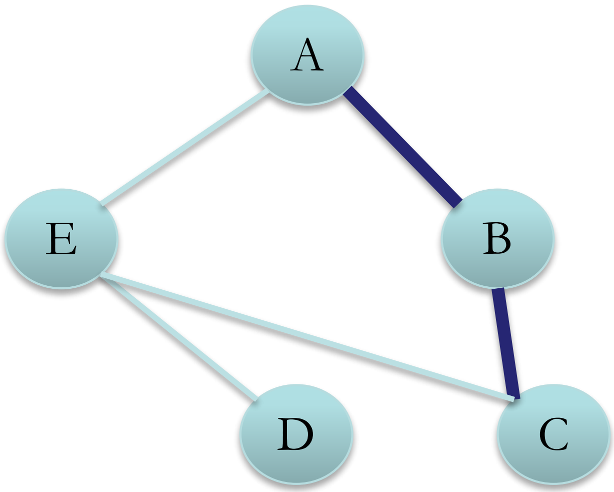
- Real networks often have one unique giant component: **largest connected component**



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 (d) between any pair of its nodes.



d(A,C)=2

Diameter???

Max of: :

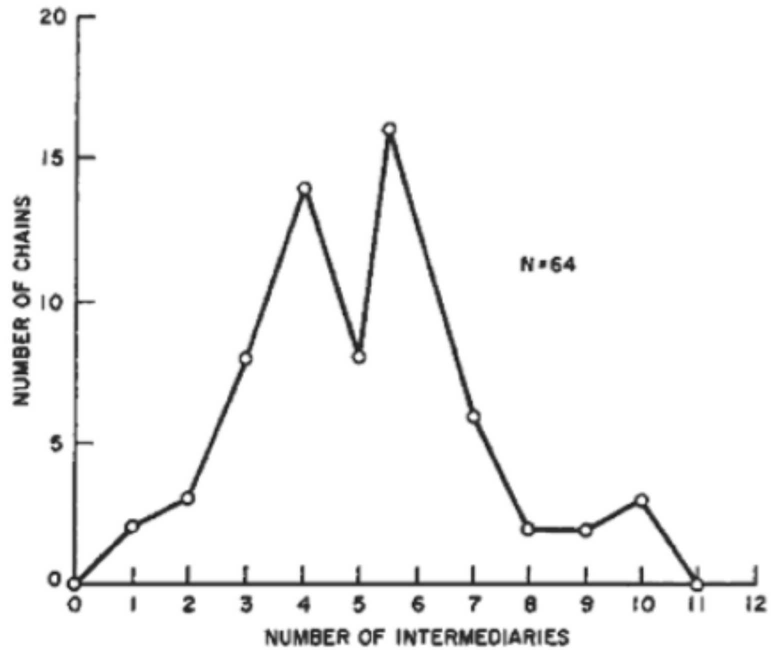
d(A.B) d(B,D)
d(A.C) d(B,E)
d(A.D) d(C.D)
d(A.E) d(C.E)
d(B.C) d(D.E)

SMALL-WORLD PHENOMENON MILGRAM'S EXPERIMENT

Two random people are connected through only a few (6) intermediate acquaintances.

Milgram's experiment (1967) shows the known "six degrees of separation":

- Choose 300 people at random
- Ask them to send a letter through friends to a stockbroker near Boston.
- 64 successful chains.



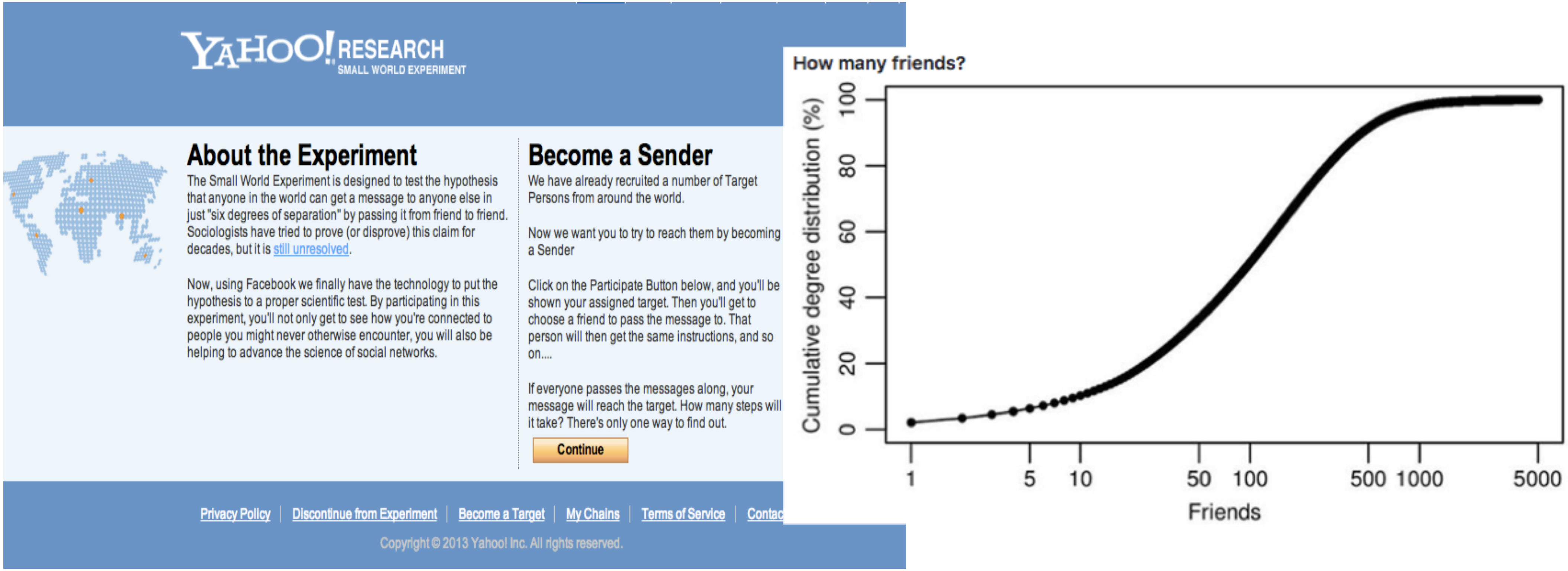
RECENT REMAKE

- In 2003 the experiment was redone.
- Choice of 18 targets over the world. Choice of senders (60K) from a commercially obtained email list.
- Website to control the "email" contact from one participant to the next (two weeks to choose next hop)
- Verification of relationship by receiver (to avoid cheating by web search).

FINDINGS

- Use of “weak ties” and professional relationships
- Median of 5-7 steps
- “Network structure alone is not everything”
- Some different incentives had a high impact on completion rate of chains
  - If the target was in a prominent place (e.g. professor)

IT'S NOT RESOLVED…



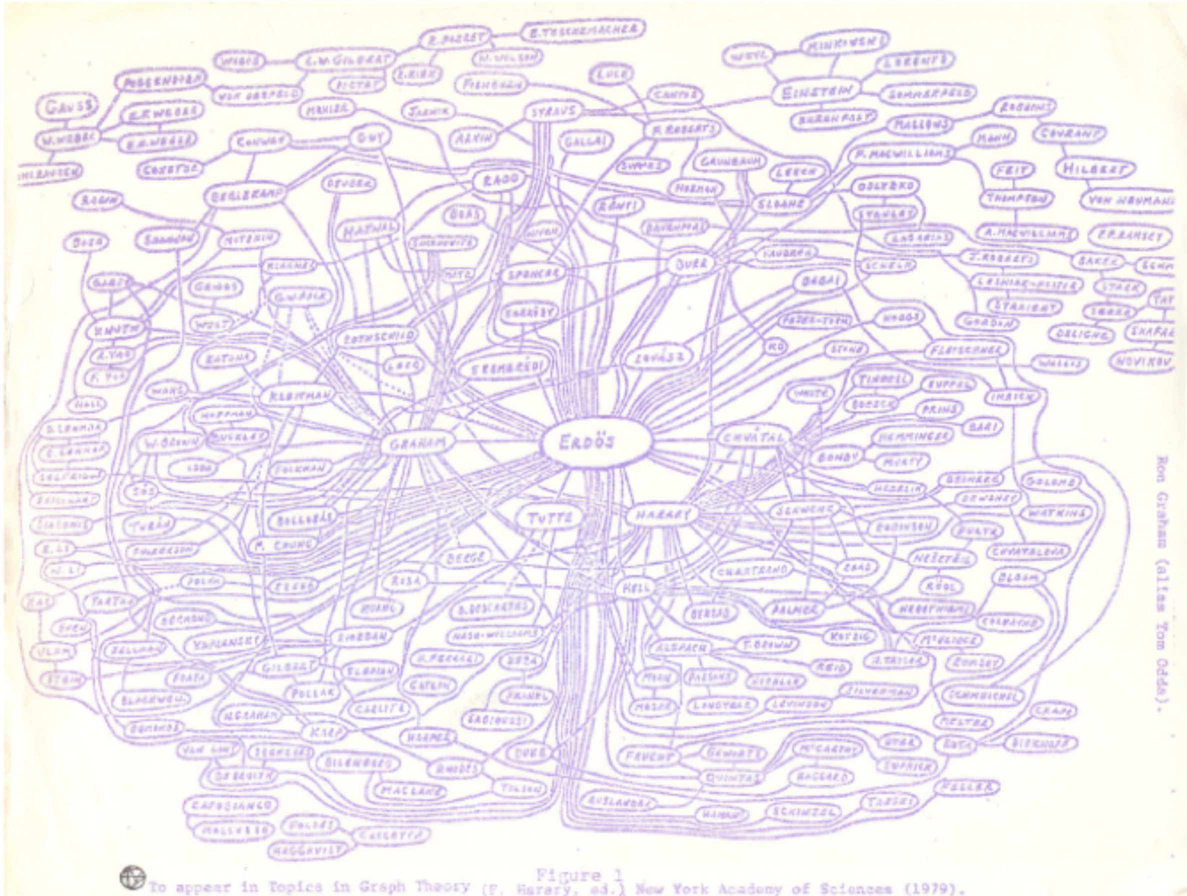
And the Facebook study that shows the The average distance in 2008 was 5.28 hops, while in 2011 it was 4.74.

https://www.facebook.com/notes/facebook-data-team/anatomy-of-facebook/10150388519243859

ERDOS NUMBER

Erdos Number: distance from the mathematician Paul Erdos (LEGEND!)

(most people are 4-5 hops away) based on collaboration.



Laurissa N. Tokarchuk
co-authored 6 papers with
Hamed Haddadi
co-authored 1 paper with
Andrew Thomason
co-authored 14 papers with
Béla Bollobás
co-authored 7 papers with
Paul Erdős
distance = 4

https://www.csauthors.net/distance

BACON NUMBER

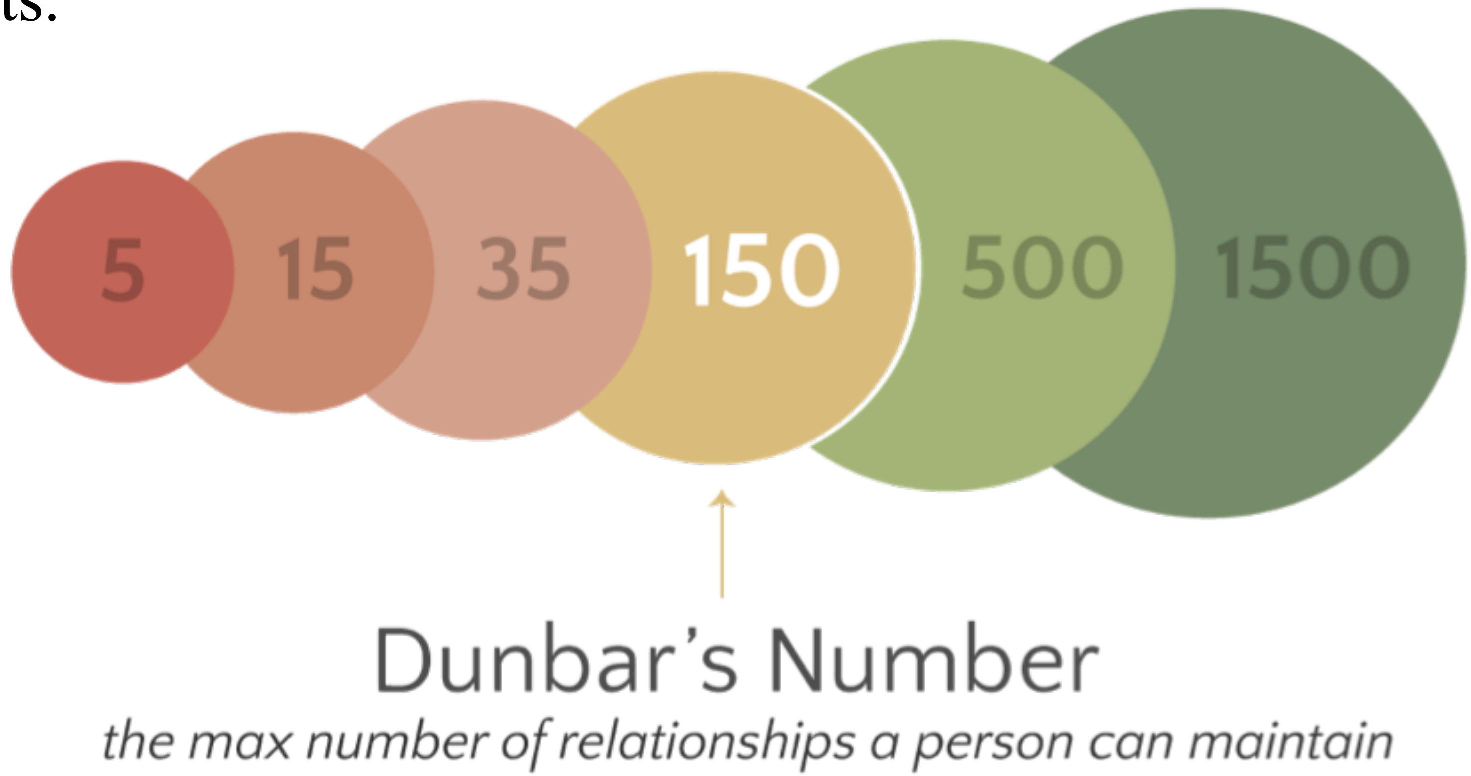
A network of actors who costarred in a movie.

Most actors are no more than ~3 hops from Kevin Bacon.

One very obscure movie was at distance 8.

DUNBAR NUMBER

British anthropologist Robin Dunbar theorized that there's a limit to the amount of relationships a person can maintain—roughly 150. Although the exact number is widely debated, the theory stands. Some people have embraced this concept on social media platforms by self-imposing friend limits.



MODELLING GRAPHS

Can we auto-generate a graph that "looks" like a real graph?

Lets scientists/nerds have fun playing with graphs that match certain characteristics

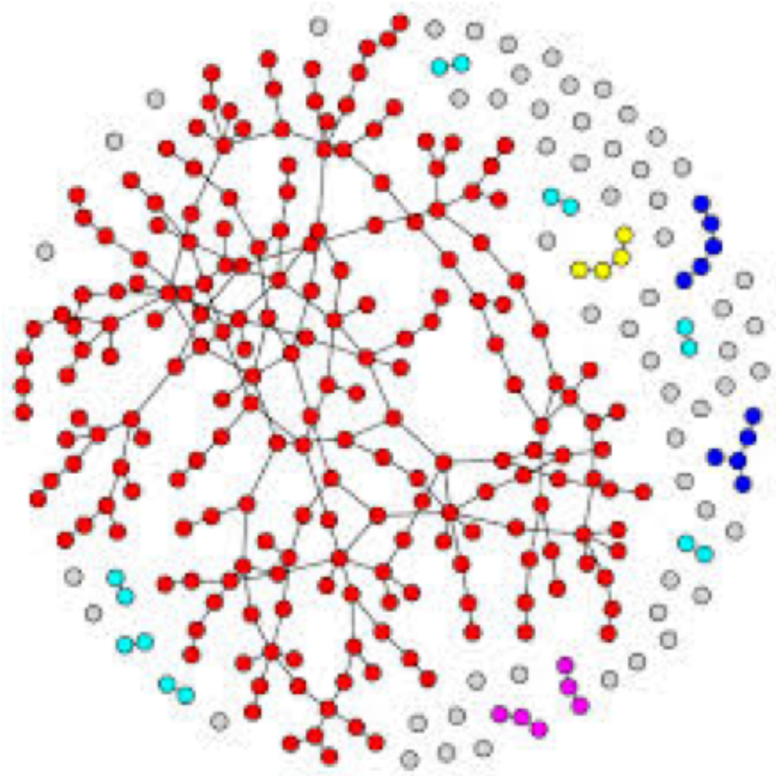
E.g. we can understand how graphs evolve over time

💡 Different random graph models produce different probability distributions on graphs.

RANDOM GRAPHS

First way to model these networks:

Erdos-Renyi Random Graph [Erdos-Renyi '59]:



G(n,p): graph with n vertices where an edge exists with independent random probability 0<p<1 for each edge.

RANDOM GRAPH MODEL

- For each node n1, an edge to node n2 exists with probability p.
- Degree distribution is **binomial.**
- The probability of a node to have degree k:

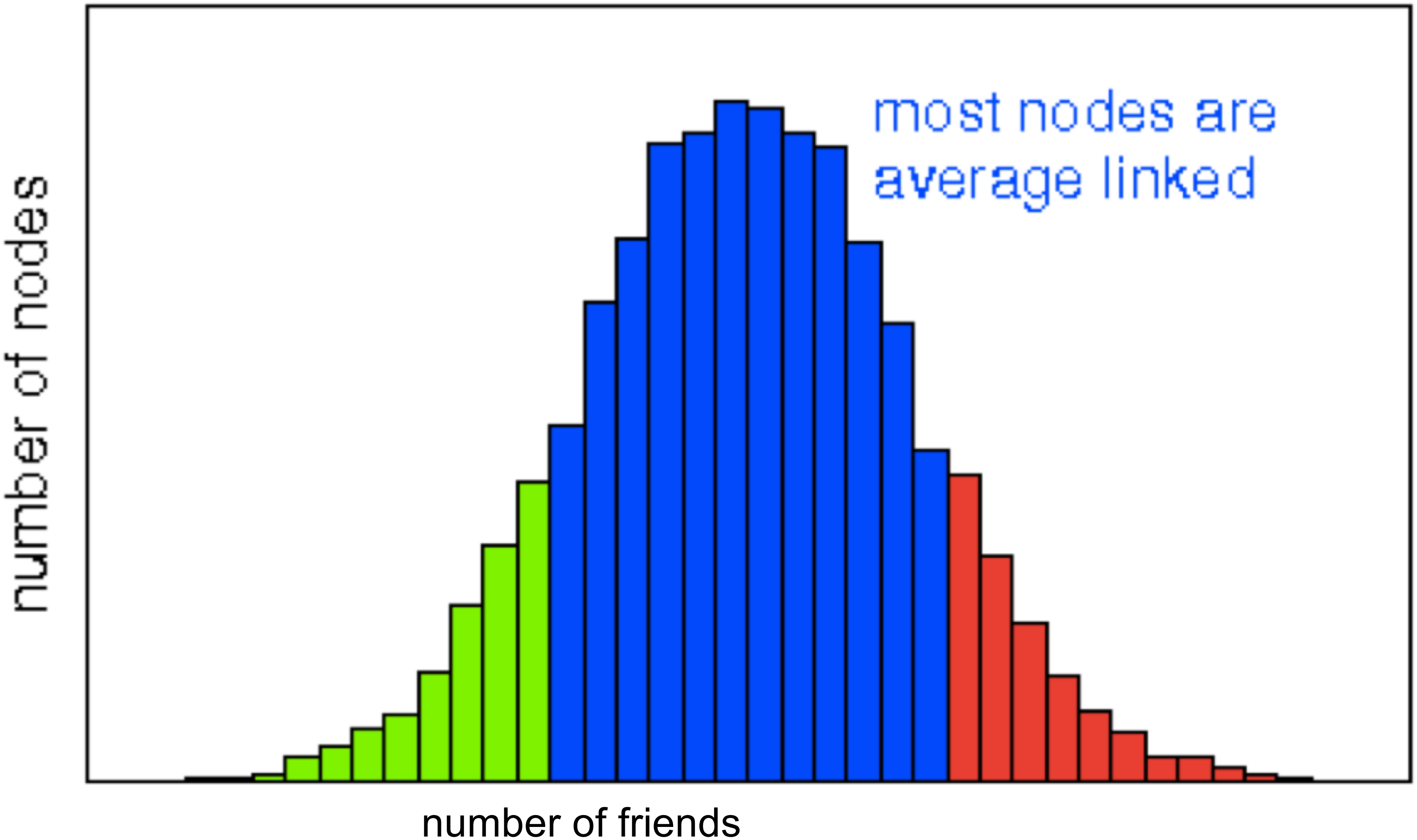
$$P(k_i = k) = C_{N-1}^{k} p^{k} (1-p)^{N-1-k}$$

- Where $C_{N-1}^{k} = \binom{N-1}{k}$

- Expected Degree of a node: $(N-1)p \approx Np$

*

DEGREE DISTRIBUTION OF RANDOM GRAPHS



RANDOM GRAPH DIAMETER

The **diameter** of random graph and the **average path length** of the graph have been demonstrated to be:

$$d = \frac{\ln(N)}{\ln(pN)} = \frac{\ln(N)}{\ln(\langle k \rangle)} \approx l_{rand}$$

The average distance between two nodes is quite small wrt to the size of the graph.

RELATIONSHIP OF <K> AND CONNECTIVITY

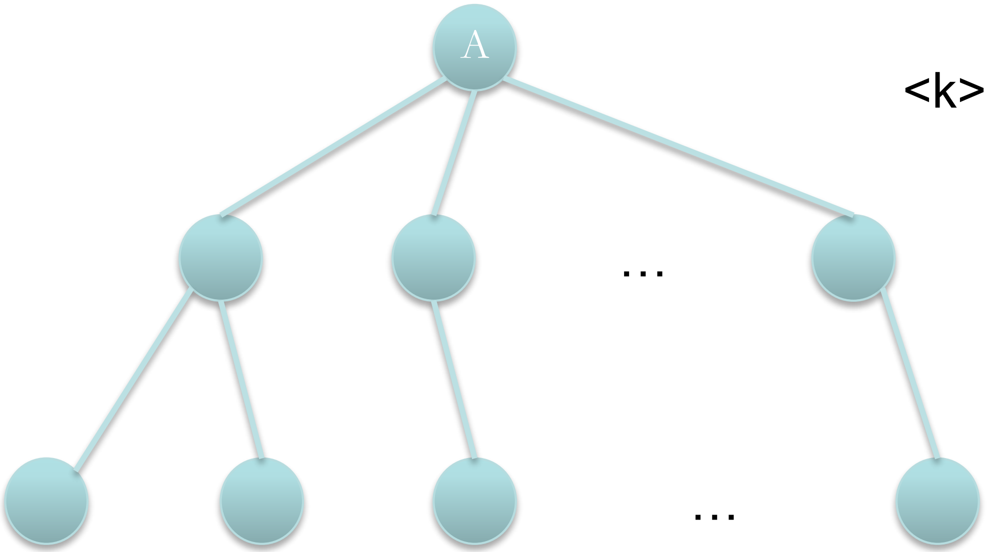
<k> = average degree (np)

If <k> < 1 disconnected network

If <k> > 1 a giant component appears

If <k> >= ln(N) graph is totally connected

RANDOM GRAPH DIAMETER: AN INTUITION



<k> new connections

<k> new connections per each new node

The nodes at distance l from A will be ~ $\langle k \rangle^l$

N=k^l log N= l log k
l = log N/log k

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

FORMALLY: CLUSTERING COEFFICIENT

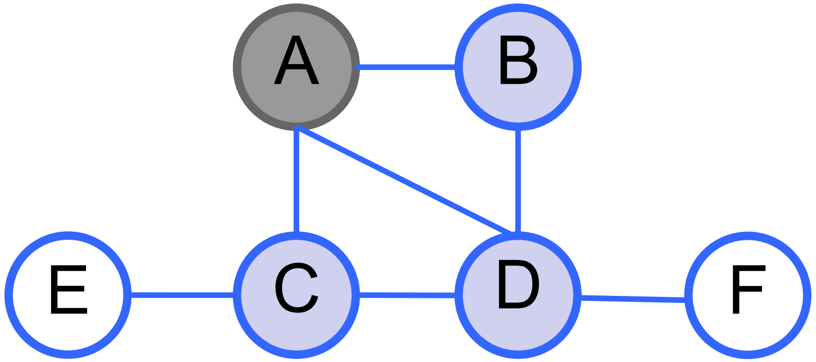
Local Clustering Coefficient

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

Network Clustering Coefficient

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

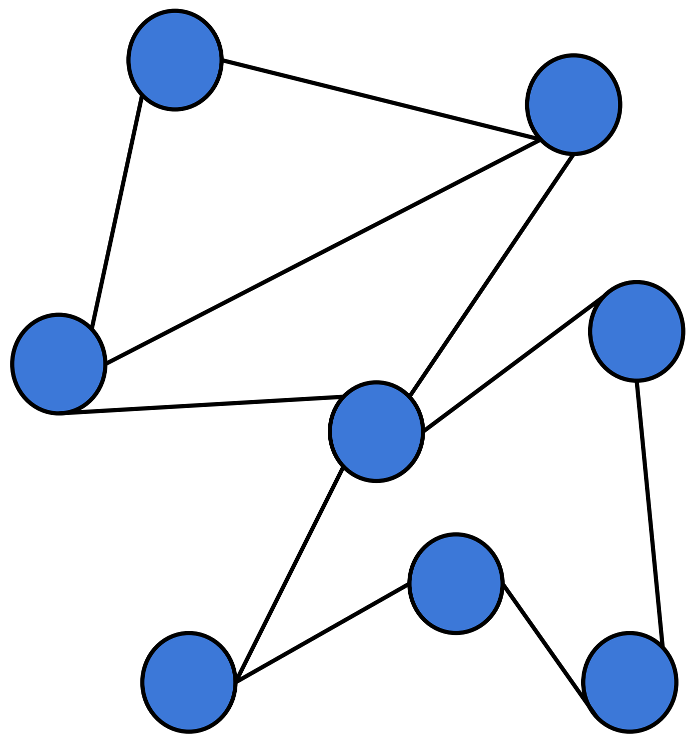
CLUSTERING COEFFICIENT



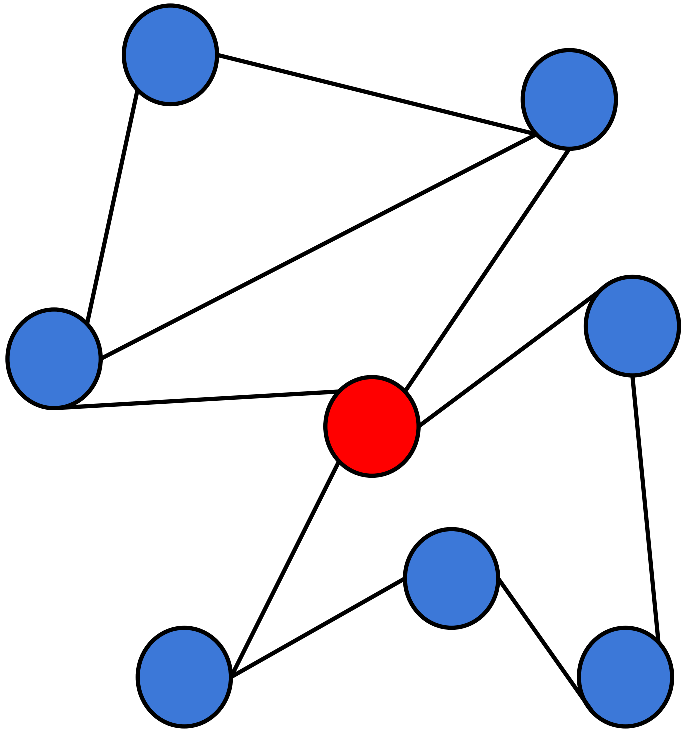
The **clustering coefficient** defines the proportion of A's neighbours (N(A)) which are connected by an edge (are friends).

The number of triangles in which A is involved wrt to the ones it could be involved in.

CLUSTERING COEFFICIENT: EXAMPLE

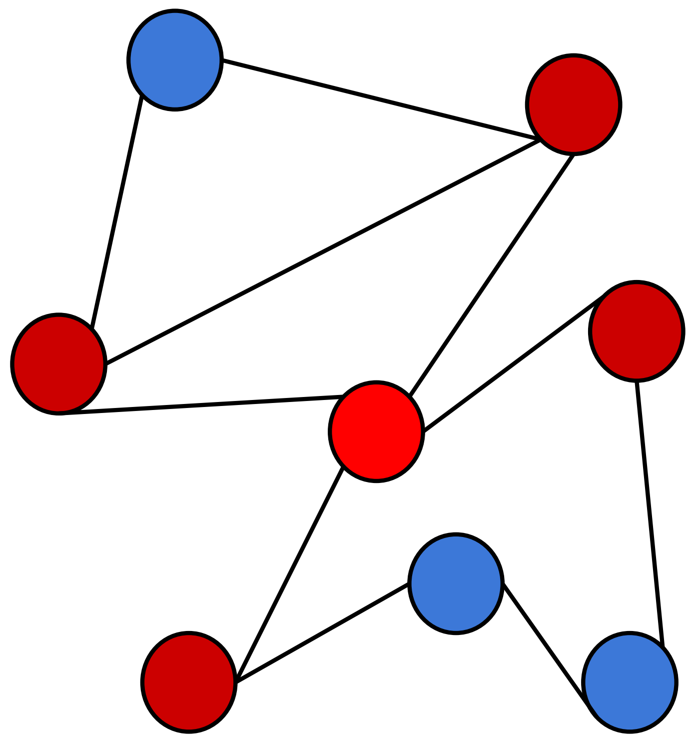


CLUSTERING COEFFICIENT: EXAMPLE



CLUSTERING COEFFICIENT: EXAMPLE

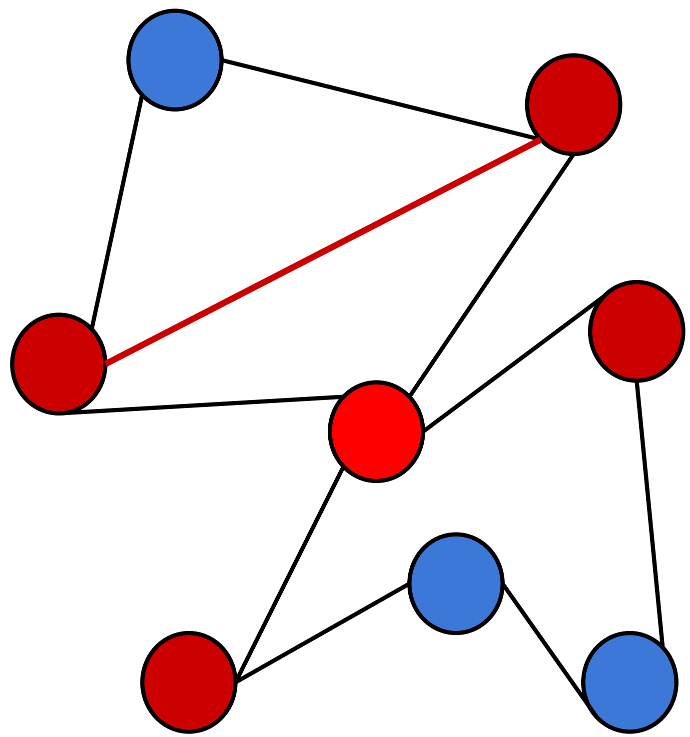
Degree = 4



CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

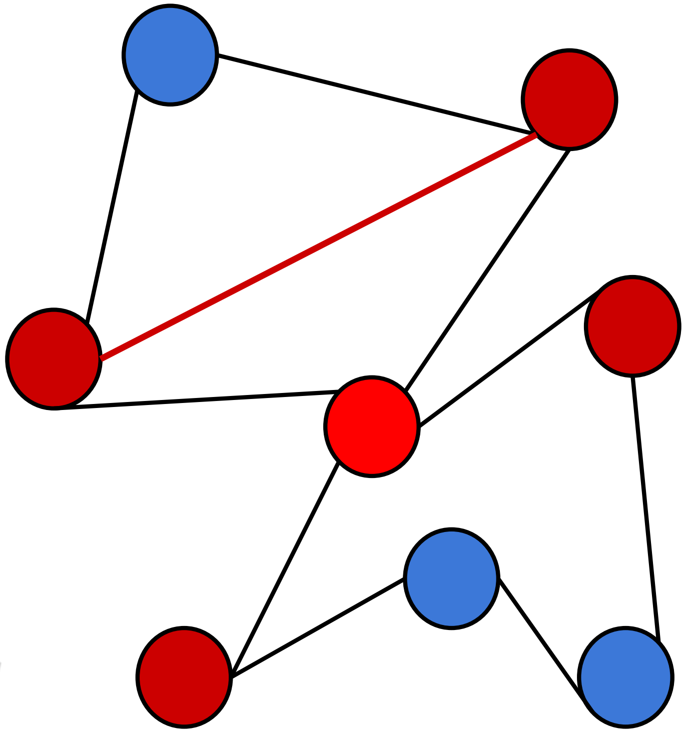


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

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

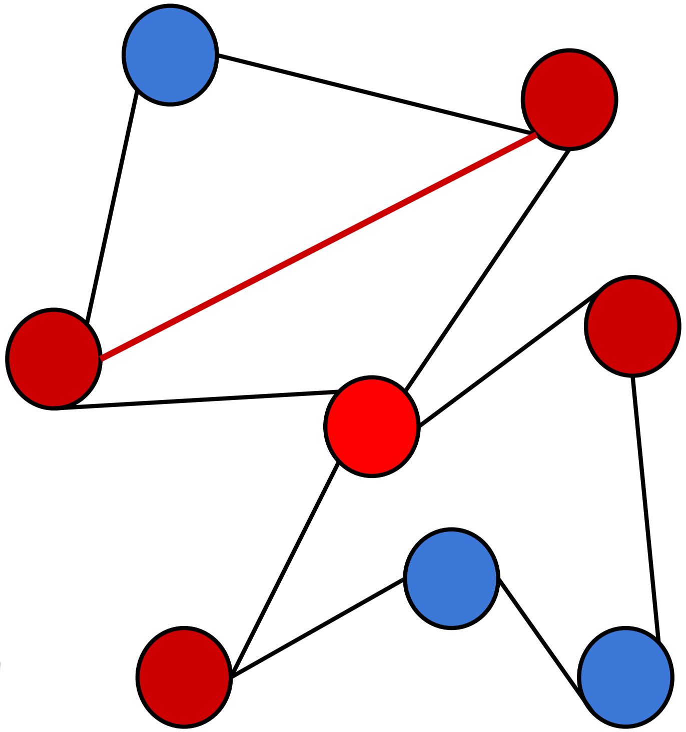


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

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

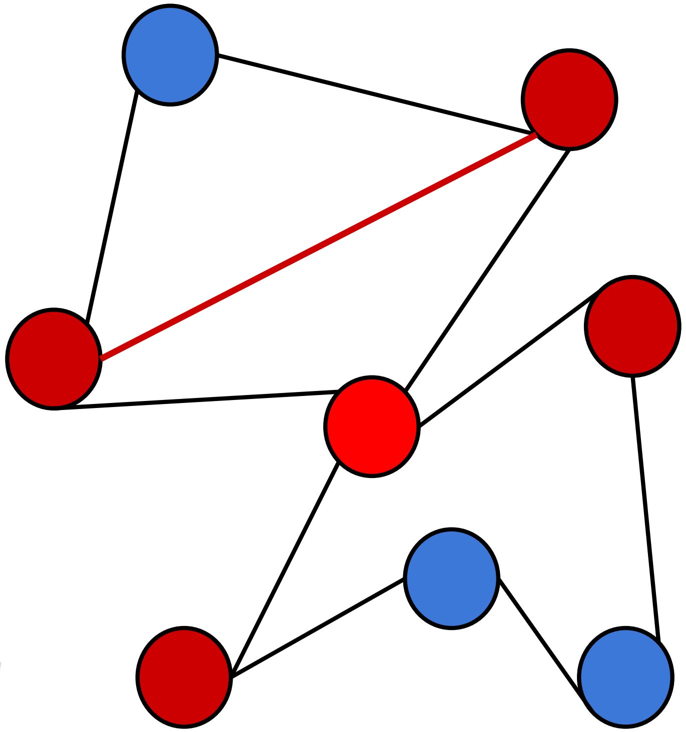


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

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



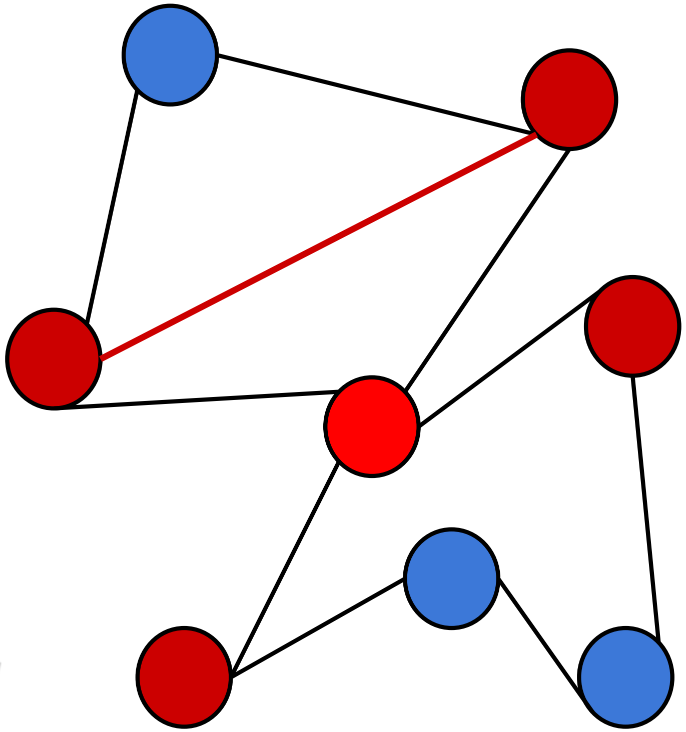
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

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

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

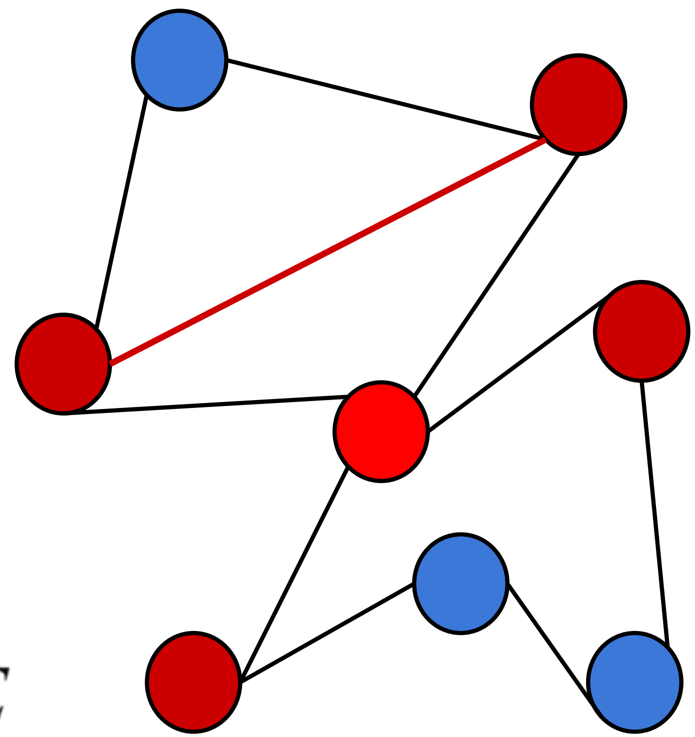


CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

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



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

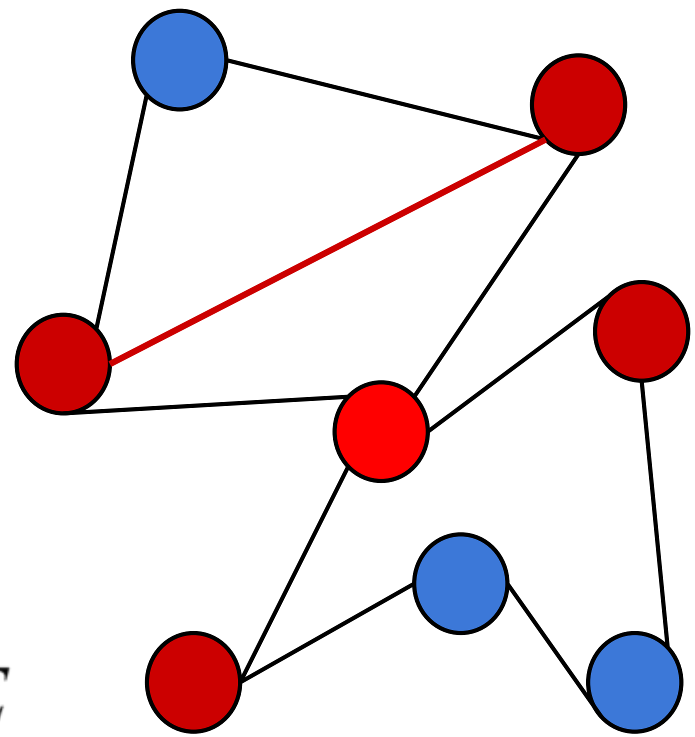
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

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

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



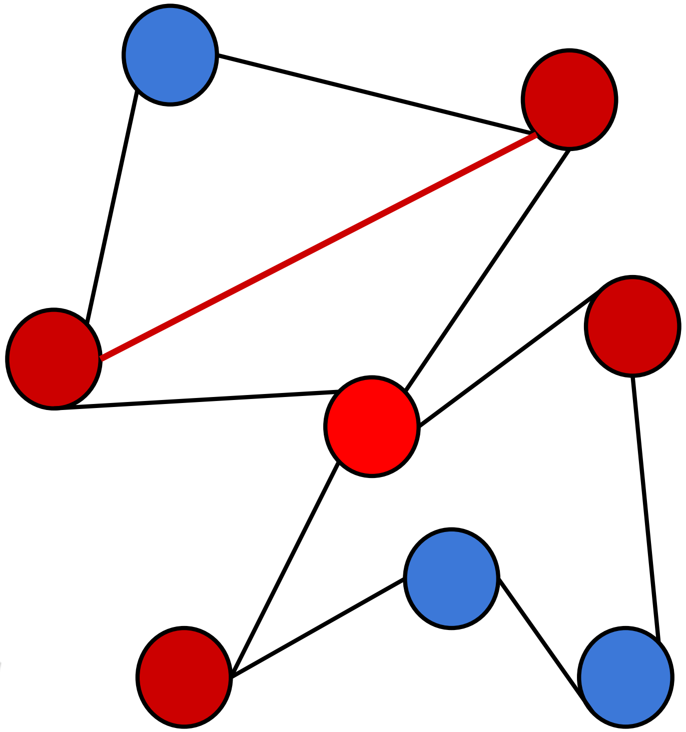
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

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

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



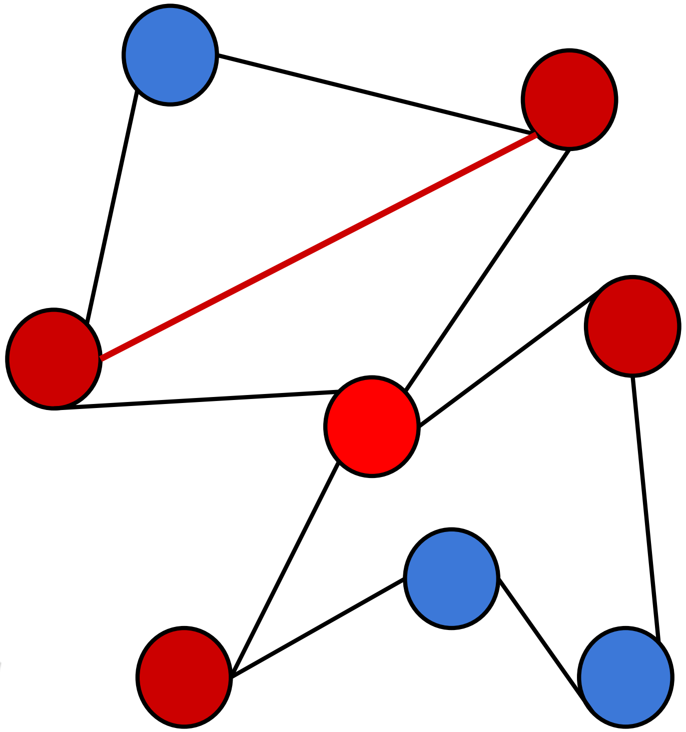
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

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

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



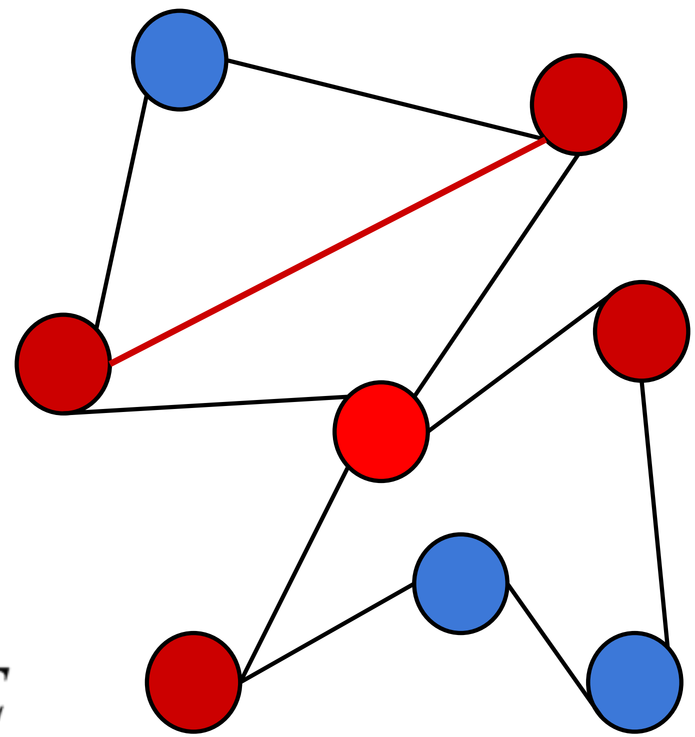
CLUSTERING COEFFICIENT: EXAMPLE

Degree = 4

Links between neighbours = 1

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

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

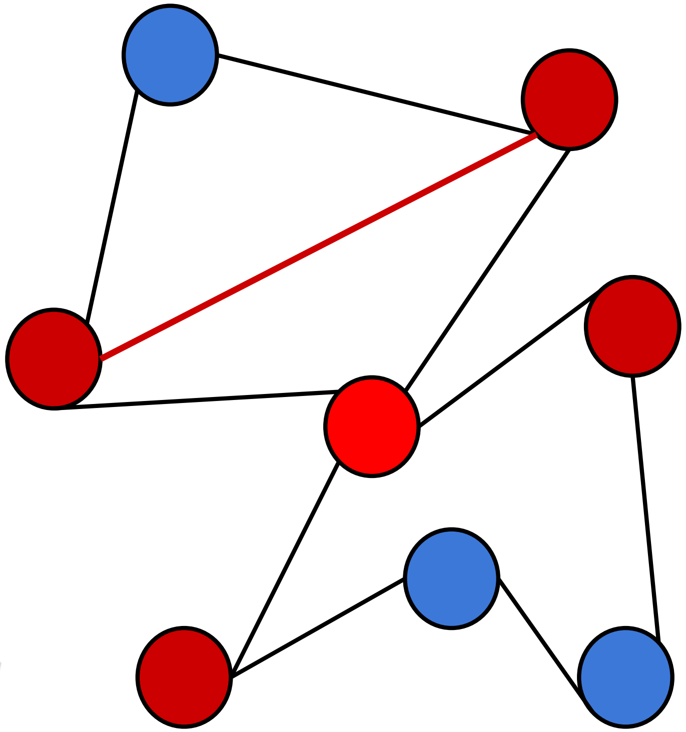


CLUSTERING COEFFICIENT: EXAMPLE

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 OF A RANDOM GRAPH

The clustering coefficient of a random graph is

$$C_{rand} = p = \frac{\langle k \rangle}{N}$$

The probability that 2 neighbours of a node are connected is equal to the probability that 2 random nodes are connected

Is this mirroring the clustering coefficient of real networks?

QUESTION & SUMMARY

Are Random Graphs representatives of Real Networks??

We have introduced graphs definitions and measures.

Random graphs are a first examples of models for networks. But don't very representative of real networks!

References

- Material from Chapter 1, 2 of
  - **D. Easley, J. Kleinberg. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press, 2010.**
- P. Sheridan Dodds, R. Muhamad, and D. J. Watts. An Experimental Study of Search in Global Social Networks. Science 8. August 2003: 301 (5634), 827-829.
- R. Albert, A. Barabasi. Statistical Mechanics of Complex Networks. Reviews of Modern Physics (74). Jan. 2002.
- S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex Networks: Structure and Dynamics Physics Reports 424 (2006) 175 .