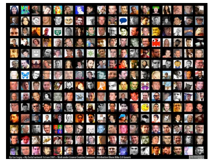


Social and Technological Network Analysis

Lecture 1: Networks, Random Graphs and Metrics

Dr. Cecilia Mascolo

UNIVERSITY OF CAMBRIDGE

About Me

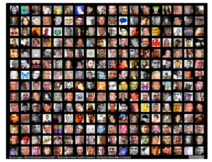
- Reader in Mobile Systems
  - NetOS Research Group
- Research on Mobile, Social and Sensor Systems
- More specifically,
  - Human Mobility and Social Network modelling
  - Opportunistic Mobile Networks
  - Mobile Sensor Systems and Networks

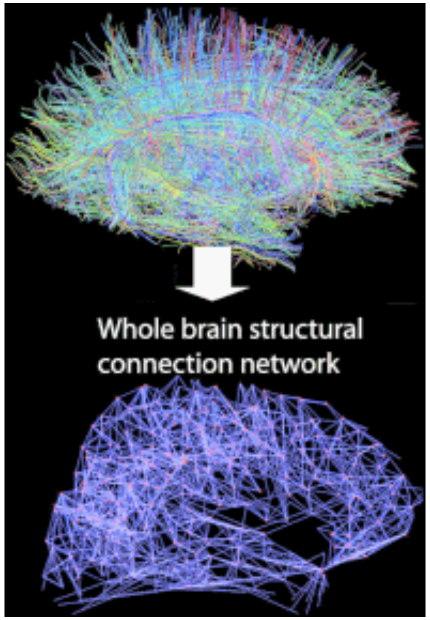
twitter

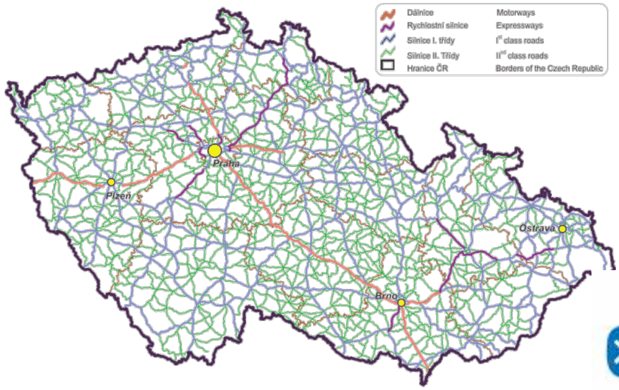
foursquare

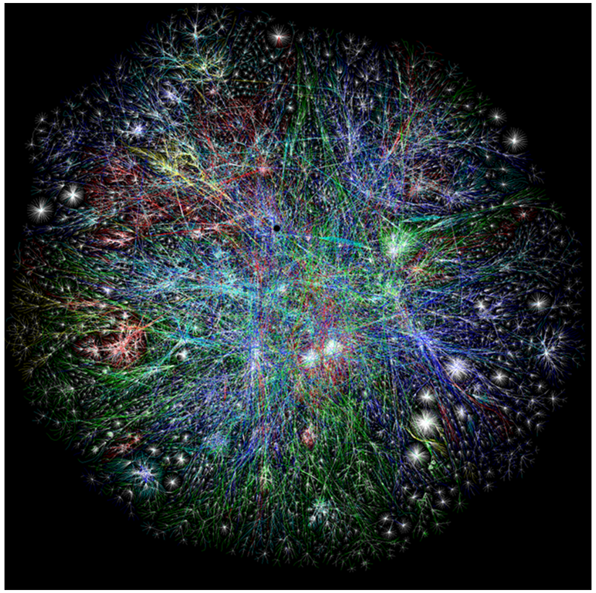
UNIVERSITY OF CAMBRIDGE

Networks are Everywhere

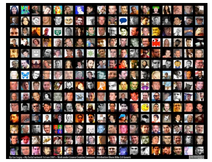






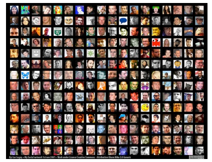


Facebook Friendship Network



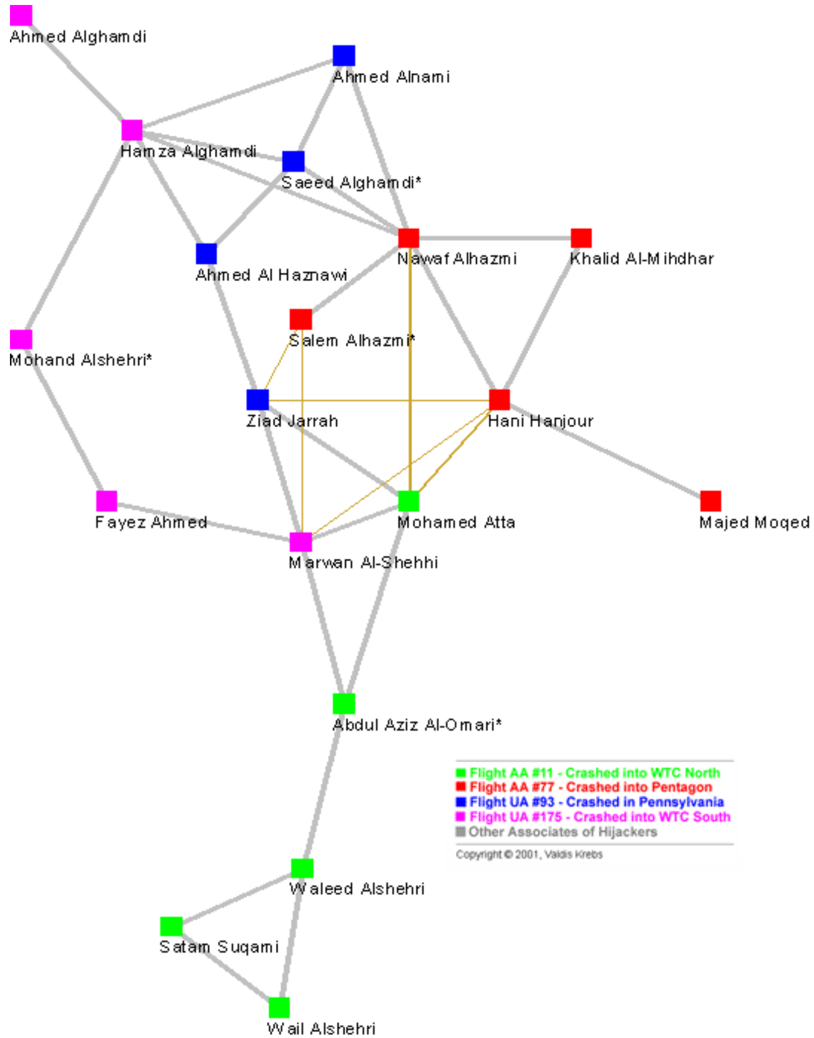
UNIVERSITY OF CAMBRIDGE

Terrorist Network



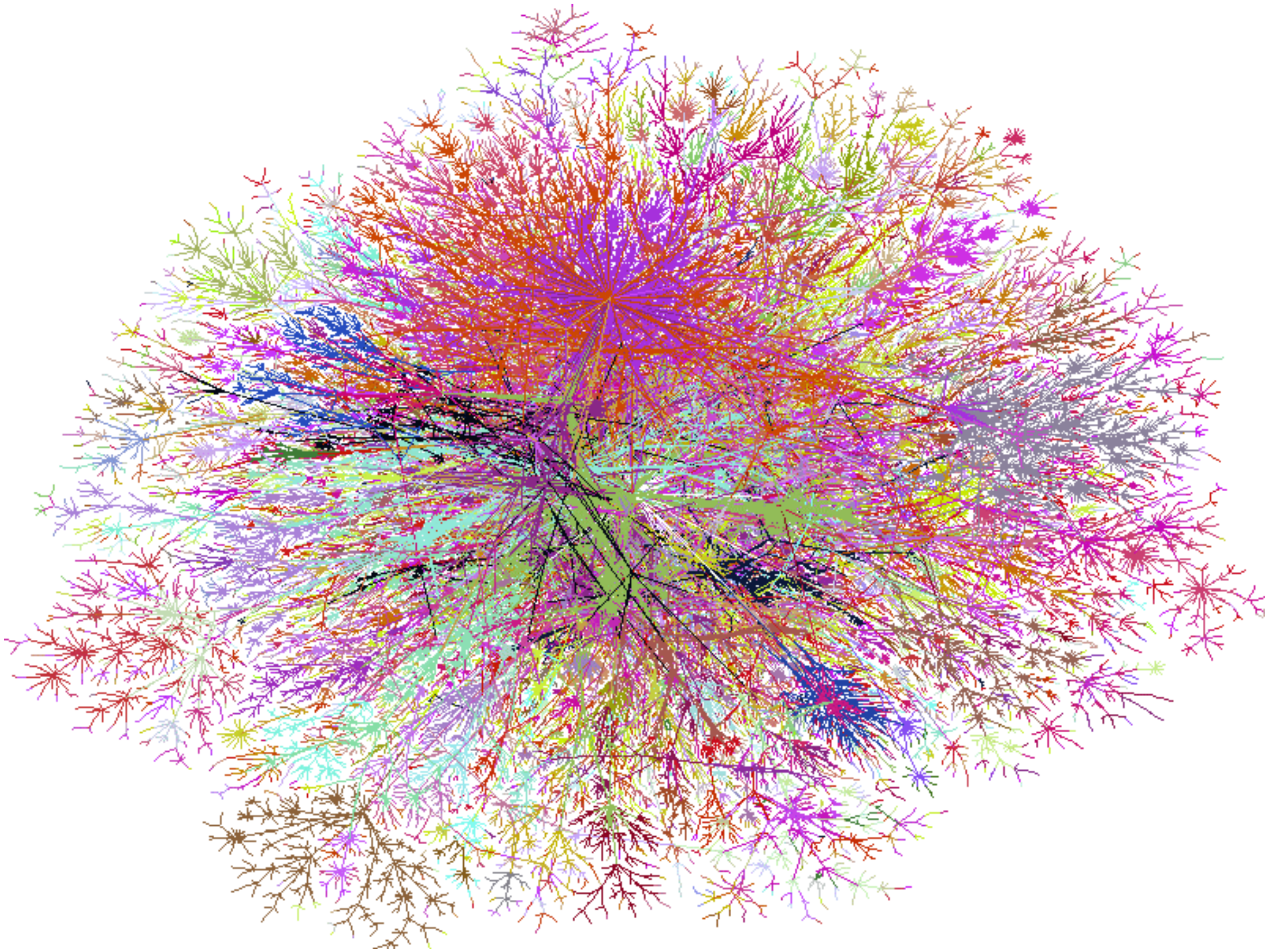
“Six degrees of Mohammed Atta”

Uncloaking Terrorist Networks, by Valdis Krebs



UNIVERSITY OF CAMBRIDGE

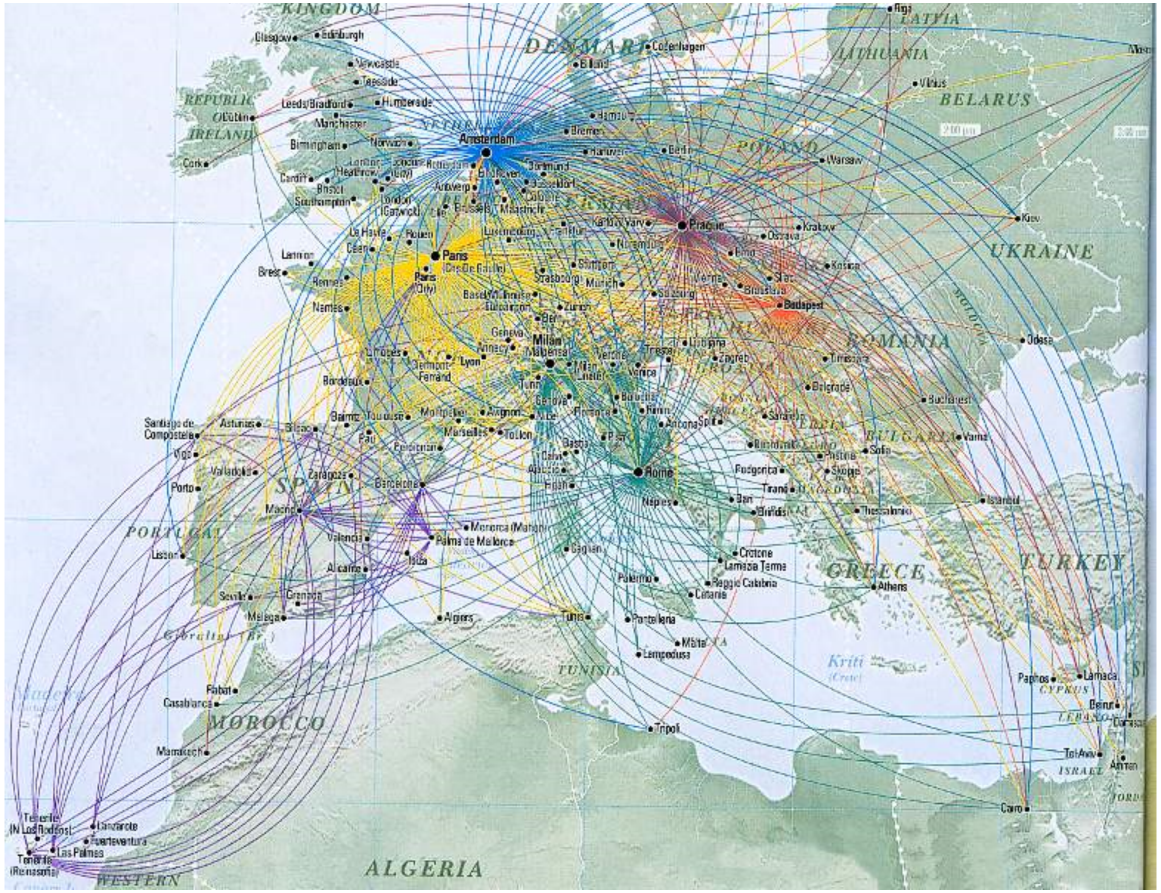
The Internet



UNIVERSITY OF CAMBRIDGE

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

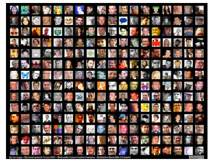
Airline Network

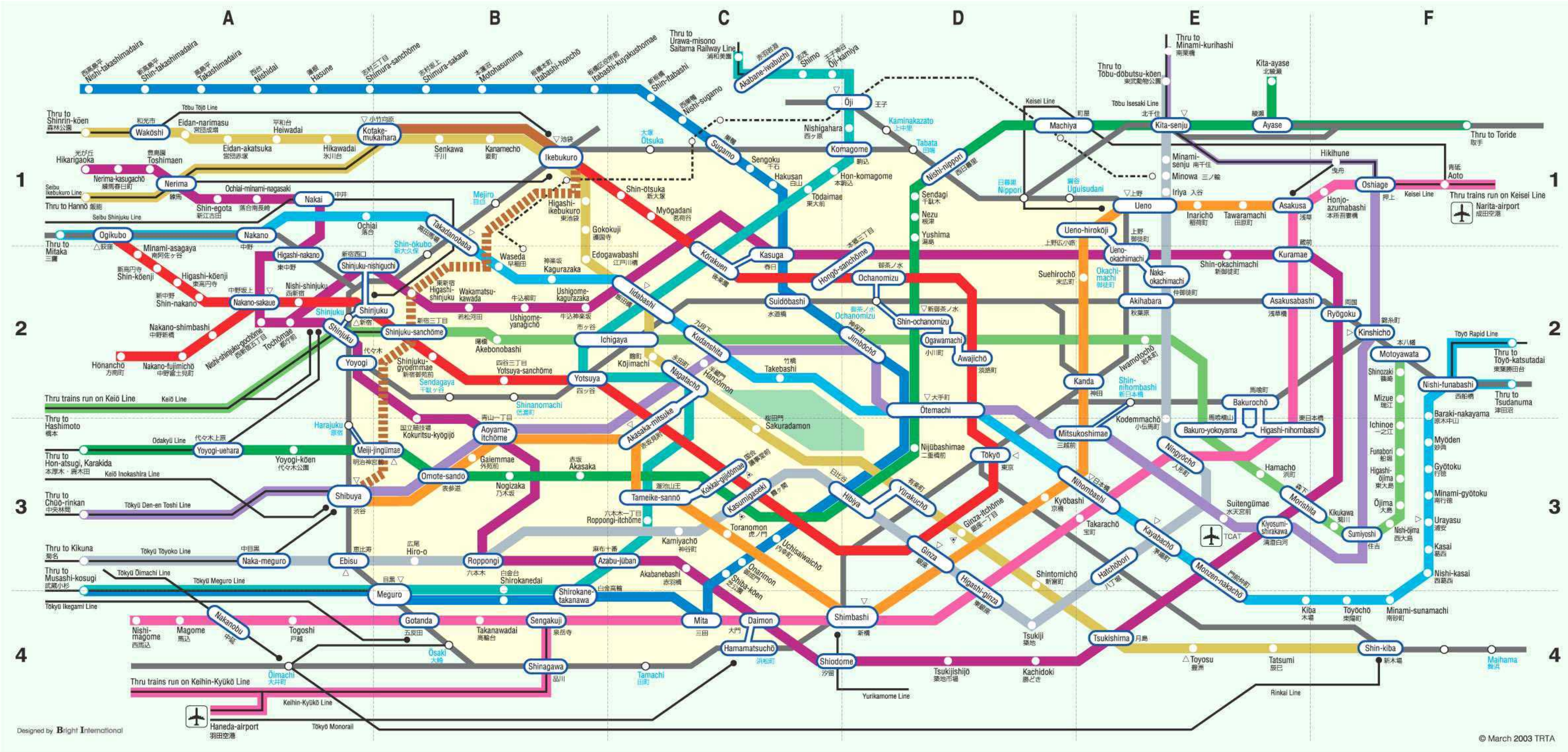


UNIVERSITY OF CAMBRIDGE

Source: Northwest Airlines WorldTraveler Magazine

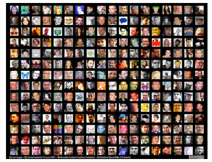
Railway Network





Source: TRTA, March 2003 - Tokyo rail map

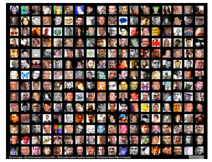
What Kind of Networks?



- 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?

UNIVERSITY OF CAMBRIDGE

In This Course

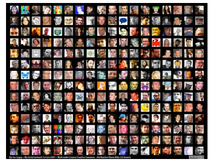


- We will study the models and metrics which allow us to understand these phenomena.
- We will show analysis over large datasets of real social and technological networks.

UNIVERSITY OF CAMBRIDGE

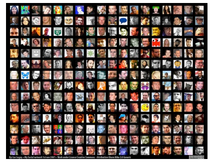
List of Lectures

- Lecture 1: Networks and Random Graphs
- Lecture 2: Small World and Weak Ties
- Lecture 3: Centrality and community detection
- Lecture 4: Modularity, overlapping communities
- Lecture 5: Structure of the Web and Power laws
- Lecture 6: The Internet and Robustness
- Lecture 7: Information Cascades on Networks
- Lecture 8: Epidemic Dissemination on Networks
- Lecture 9: Cascades and Epidemics Applications
- Lecture 10: Time Varying Social Networks
- Lecture 11-12: Practical Tutorial
- Lecture 13: Geo-Social Networks
- Lecture 14-15: Student Presentations
- Lecture 16: Industrial Presentation



Assessment

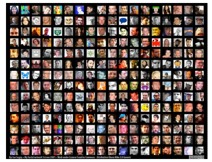
- All information on the course page: http://www.cl.cam.ac.uk/~cm542/teaching/2011/stna2011.html
- **One report** (of approximately 1,500 words) on one assigned research paper. The report is due on **10th February** and it is worth 30% of the final mark.
- The second assignment will consist of **analysis of an assigned dataset according to some indicated network measures** using NetworkX: the analysis should be reported in a document of about 1,500 words where the results are commented and justified. This should be handed in by **6th March** and will be worth 50% of the final mark.
- Presentation of the findings of 2nd assignment on **8-12thst March.** The presentation is worth 20% of the final mark.



Structure of First Report

- **The first report should be approximately 1,500 words.** The report will contain two parts of about **750 words each**:
  – Critical analysis of the papers including, possibly, comparisons and references to other material presented in the course or found by the student and comments on how solid the result obtained are (e.g., comments on the evaluation methods or on the analysis applied can be included);
  – Discussion of possible future research ideas in the area.

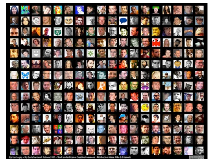
UNIVERSITY OF CAMBRIDGE



Choices

- First assignment: send email immediately with a choice of paper **(and a backup)** from the website [assignment is first come first served].
  - **Do this in the next TWO days.**

- Second assignment: same. At that point you will receive an email with a link to the dataset.
  - **Do this by 13th February.**

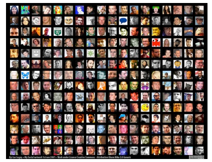


In This Lecture

- We will introduce:
  - Networks/graphs
  - Basic network measures
  - Random Graphs
  - Examples

UNIVERSITY OF CAMBRIDGE

A Network is a Graph



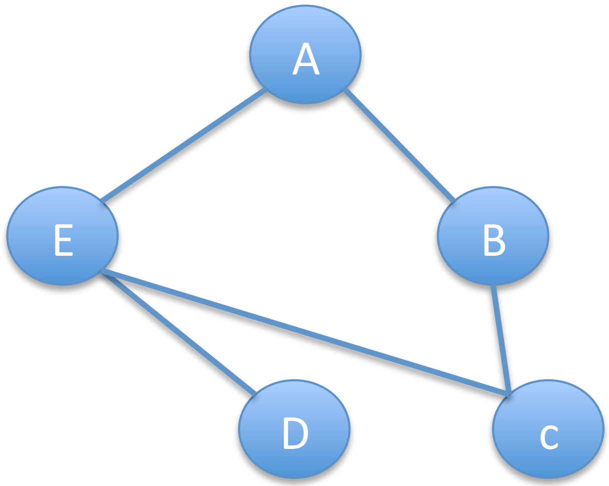
A **graph** G is a tuple (V,E) of a set of vertices V and edges E. An edge in E connects two vertices in V.

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

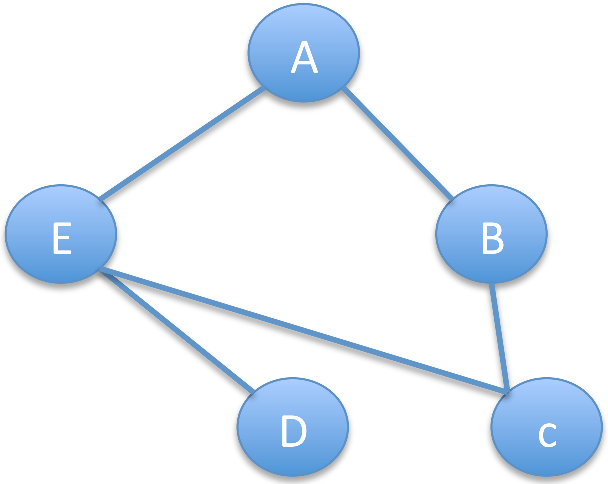
- The **node degree** is the number of neighbours of a node
- E.g., Degree of A is 2



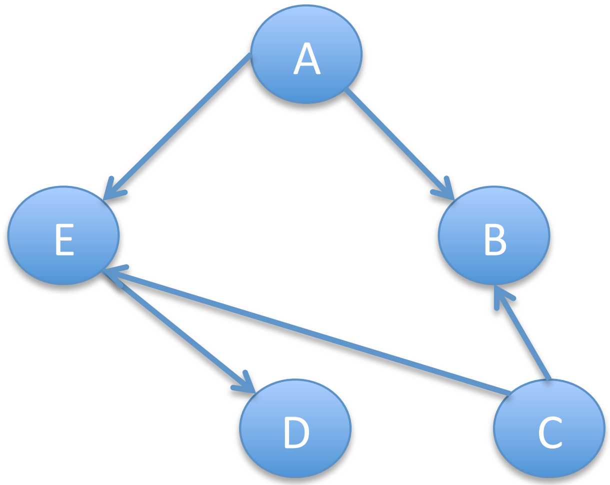
The study of the degree distribution of networks allows the classification of networks in different categories

UNIVERSITY OF CAMBRIDGE

Directed & Undirected Graphs



Undirected Graph



Directed Graph

Example of Undirected Graphs: Facebook, Co-presence

Examples of Directed: Twitter, Email, Phone Calls

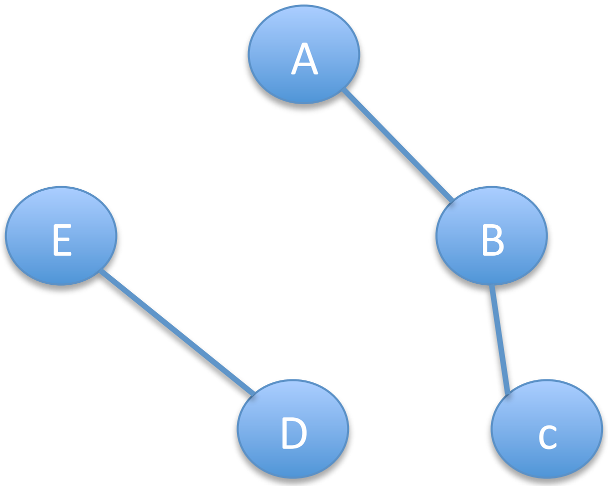
Paths and Cycles

- 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

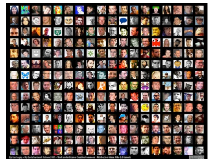
Connectivity

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

- Example of **disconnected** graph:



UNIVERSITY OF CAMBRIDGE

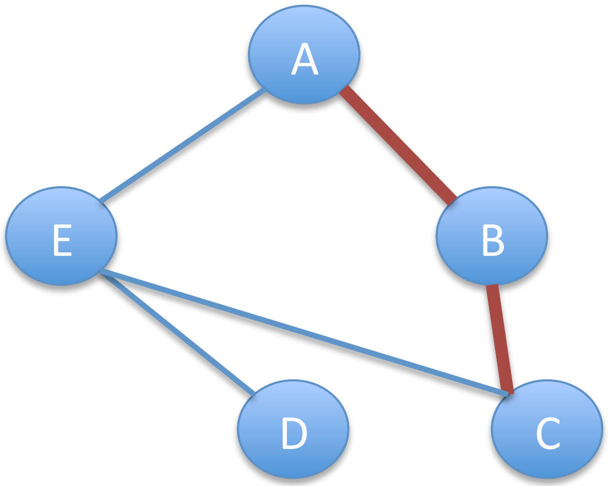


Components

- 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
- A **giant component** is a connected component containing a significant fraction of nodes in the network.
  - Real networks often have one unique giant 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 between any pair of its nodes.

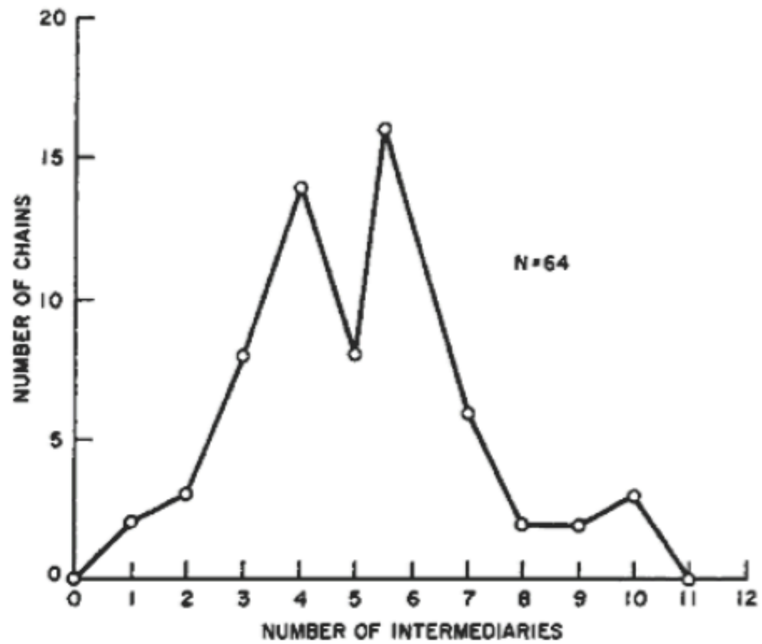


d(A,C)=2

UNIVERSITY OF CAMBRIDGE

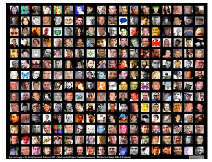
Small-world Phenomenon
Milgram's Experiment

- 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.

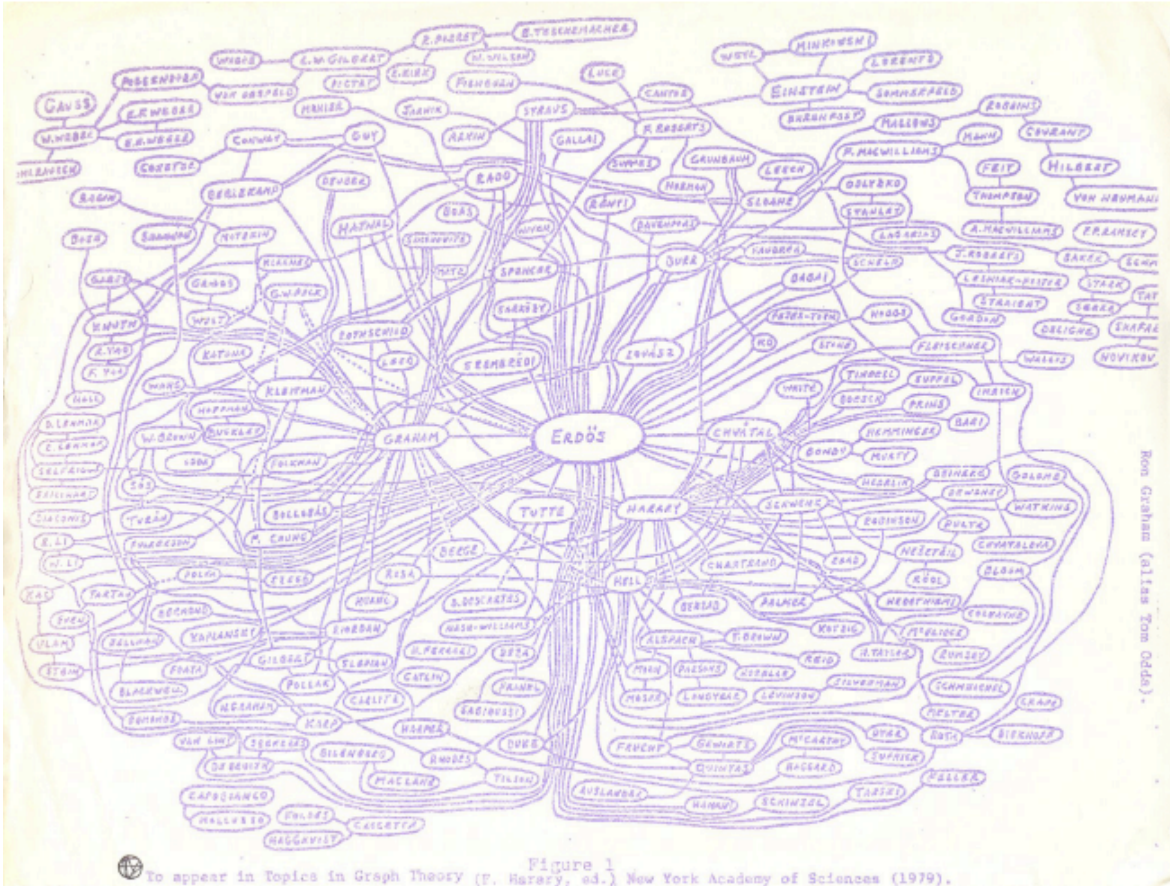


UNIVERSITY OF CAMBRIDGE

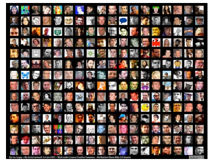
Erdos Number



- Erdos Number: distance from the mathematician (most people are 4-5 hops away) based on collaboration.



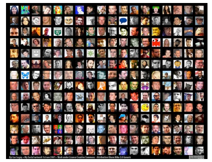
UNIVERSITY OF CAMBRIDGE



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.

Random Graphs



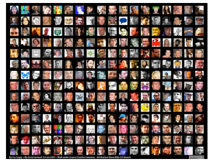
- First way to model these networks:

- **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.

UNIVERSITY OF CAMBRIDGE

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:

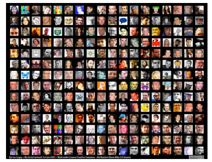
$$P(k_i = k) = C_{N-1}^{k} p^{k} (1-p)^{N-1-k}$$

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

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

UNIVERSITY OF CAMBRIDGE

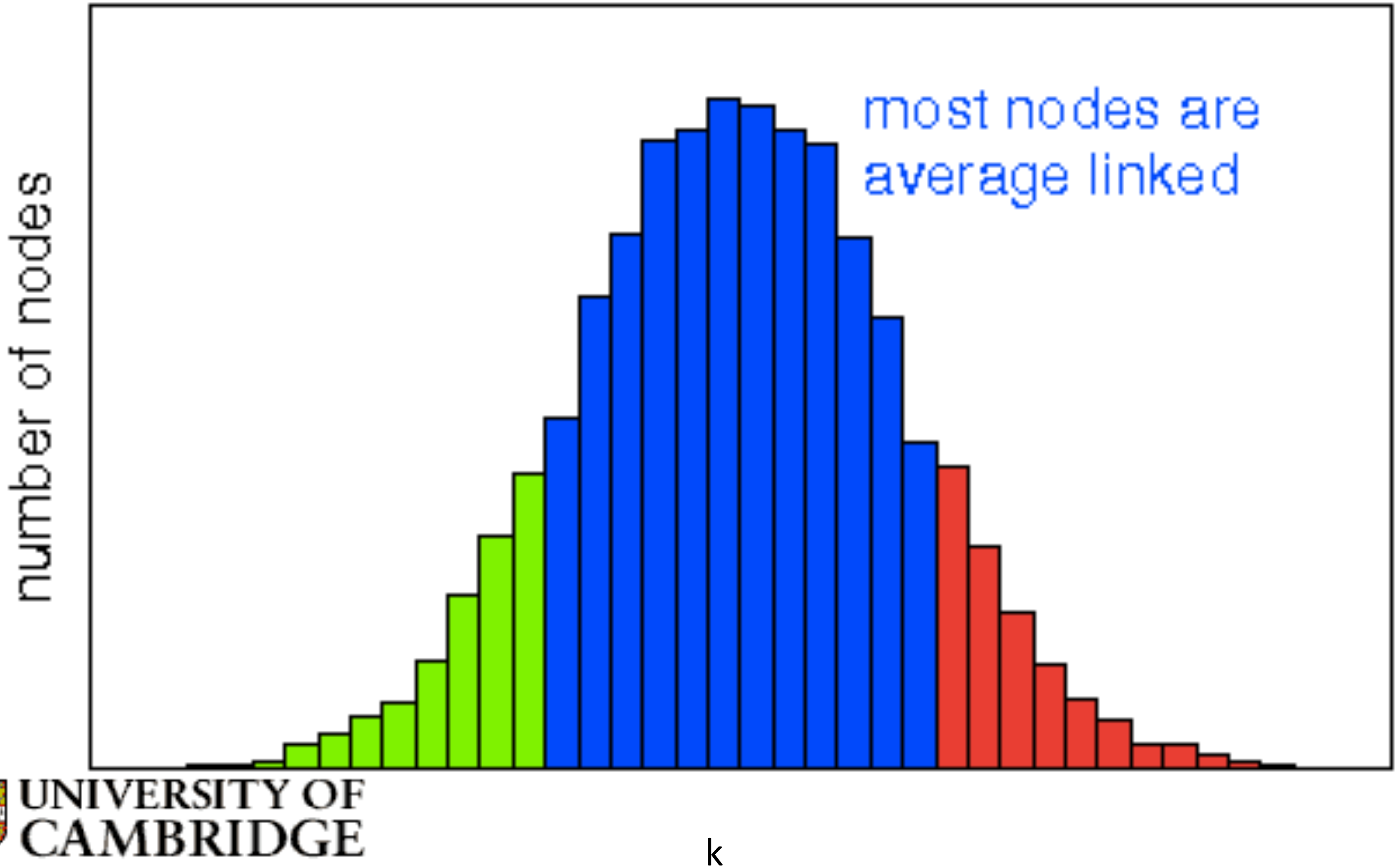
Random Graphs Properties



- For large N this is approximated by the Poisson distribution with

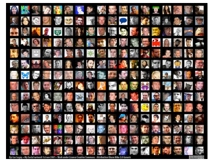
$$P(k) \approx e^{-Np}\frac{(Np)^k}{k!} = e^{-\langle k \rangle}\frac{\langle k \rangle^k}{k!}$$

Degree Distribution of Random Graphs



UNIVERSITY OF CAMBRIDGE

Random Graph Diameter



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

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

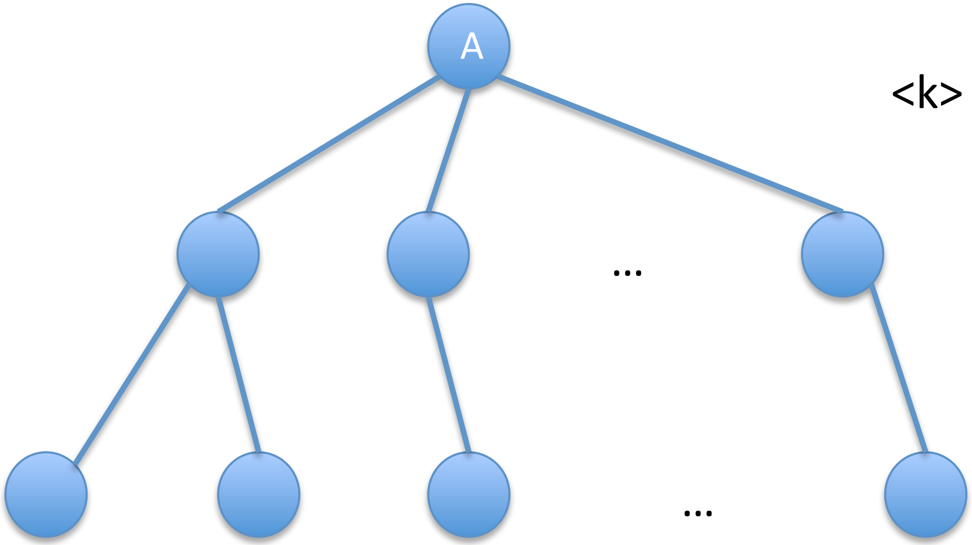
The average distance between two nodes is quite small wrt to the size of the graph.

UNIVERSITY OF CAMBRIDGE

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 ~ $\langle k \rangle^l$

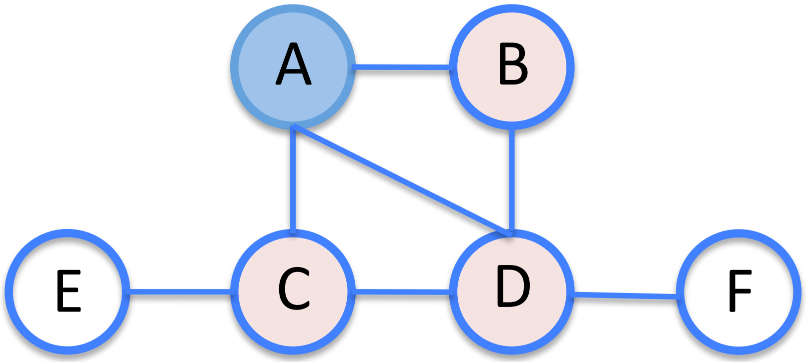
$N=k^l \quad \log N = l \log k$

$l = \log N / \log k$

UNIVERSITY OF CAMBRIDGE

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.



UNIVERSITY OF CAMBRIDGE

Formally: Clustering Coefficient

Local Clustering Coefficient

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

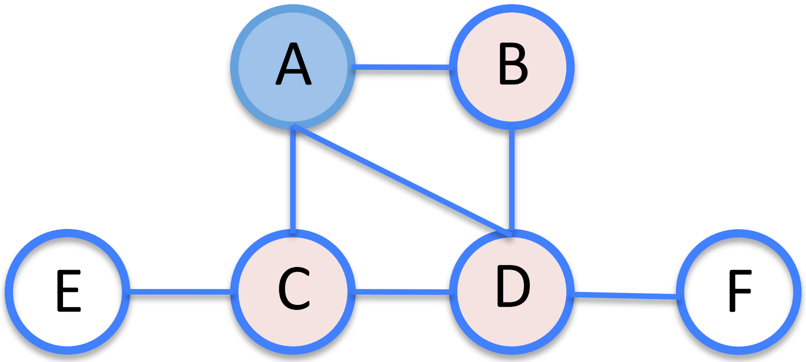
Network Clustering Coefficient

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

UNIVERSITY OF CAMBRIDGE

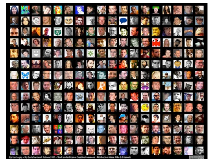
Clustering Coefficient: Example

- Ki=3
- Nominator= 2*2=4
- Denominator=6
- Ci=4/6=2/3



UNIVERSITY OF
CAMBRIDGE

Clustering Coefficient of a Random Graph



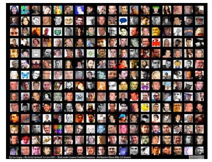
- The clustering coefficient of a random graph is

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

$$p * \left(\frac{n_v}{2}\right) / \left(\frac{n_v}{2}\right) = p$$

- 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?

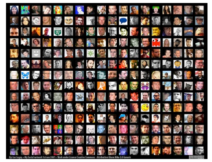
UNIVERSITY OF CAMBRIDGE



Question

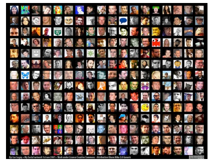
- Are Random Graphs representatives of Real Networks?

Summary



- We have introduced graphs definitions and measures.
- Random graphs are a first examples of models for networks.

UNIVERSITY OF CAMBRIDGE



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.**
- 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 .