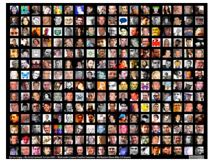
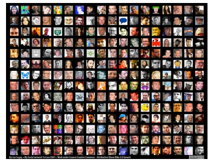


Social and Technological Network Analysis

Lecture 1: Networks, Random Graphs and Metrics

Dr. Cecilia Mascolo

UNIVERSITY OF CAMBRIDGE



About Me

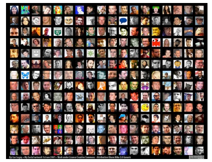
- Reader in Mobile Systems
  - NetOS Research Group
- Research on Mobile, Social and Sensor Systems
- More specifically,
  - Human Mobility and Social Network modelling
  - Geo-social recommendation systems
  - Mobile Sensor Systems and Networks

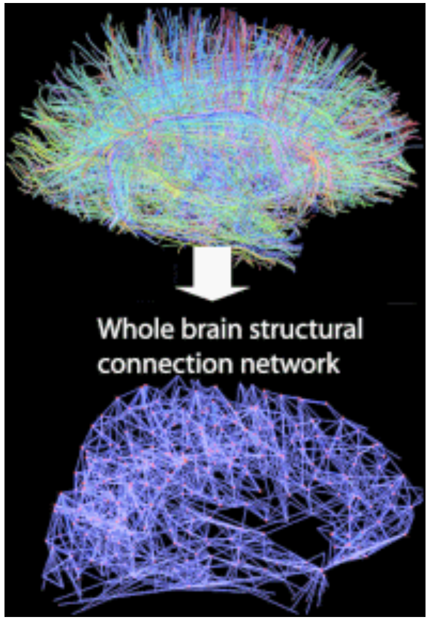
twitter

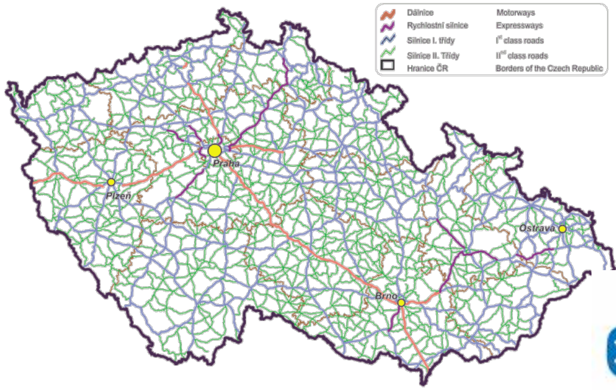
foursquare

UNIVERSITY OF CAMBRIDGE

Networks are Everywhere

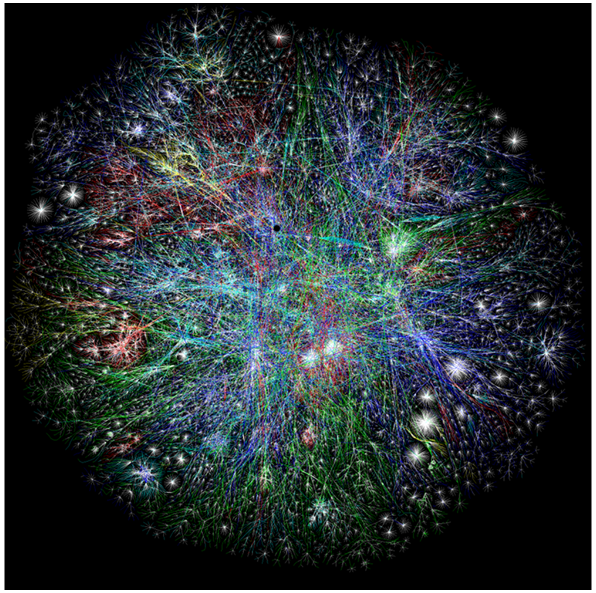






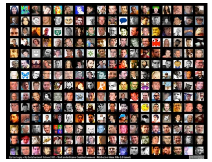
facebook

Bluetooth



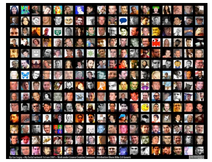
UNIVERSITY OF CAMBRIDGE

Facebook Friendship Network



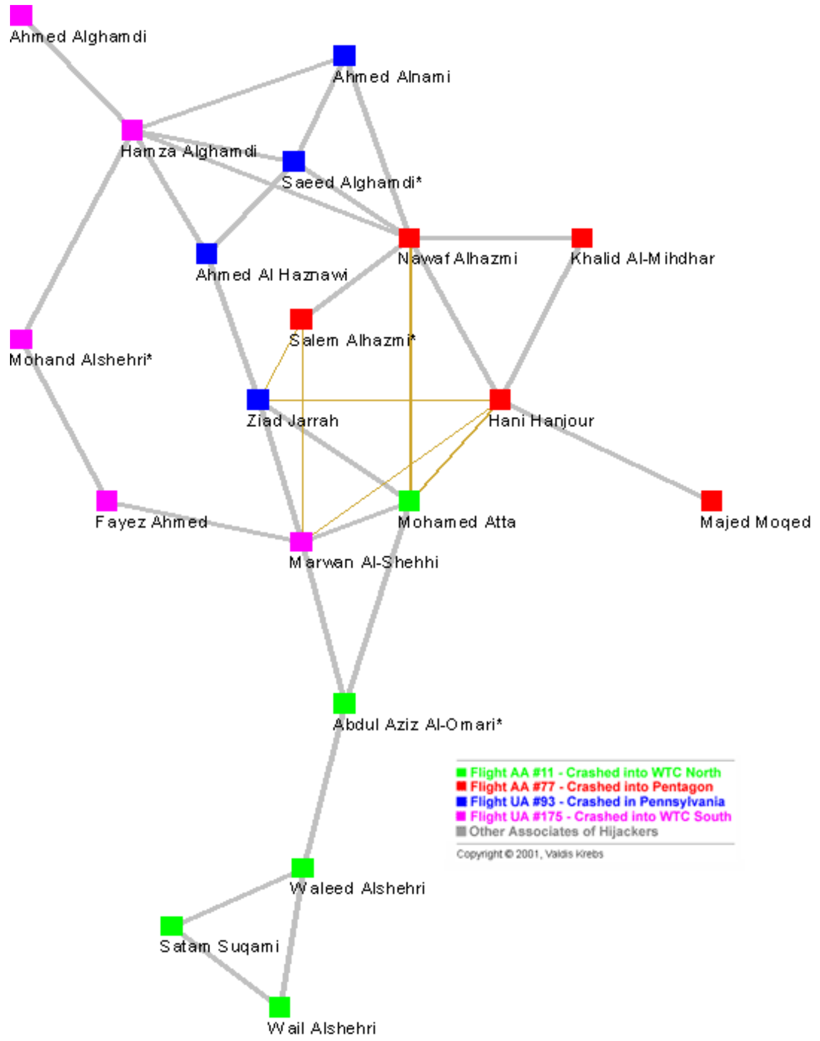
UNIVERSITY OF
CAMBRIDGE

Terrorist Network



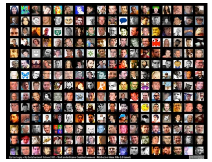
“Six degrees of Mohammed Atta”

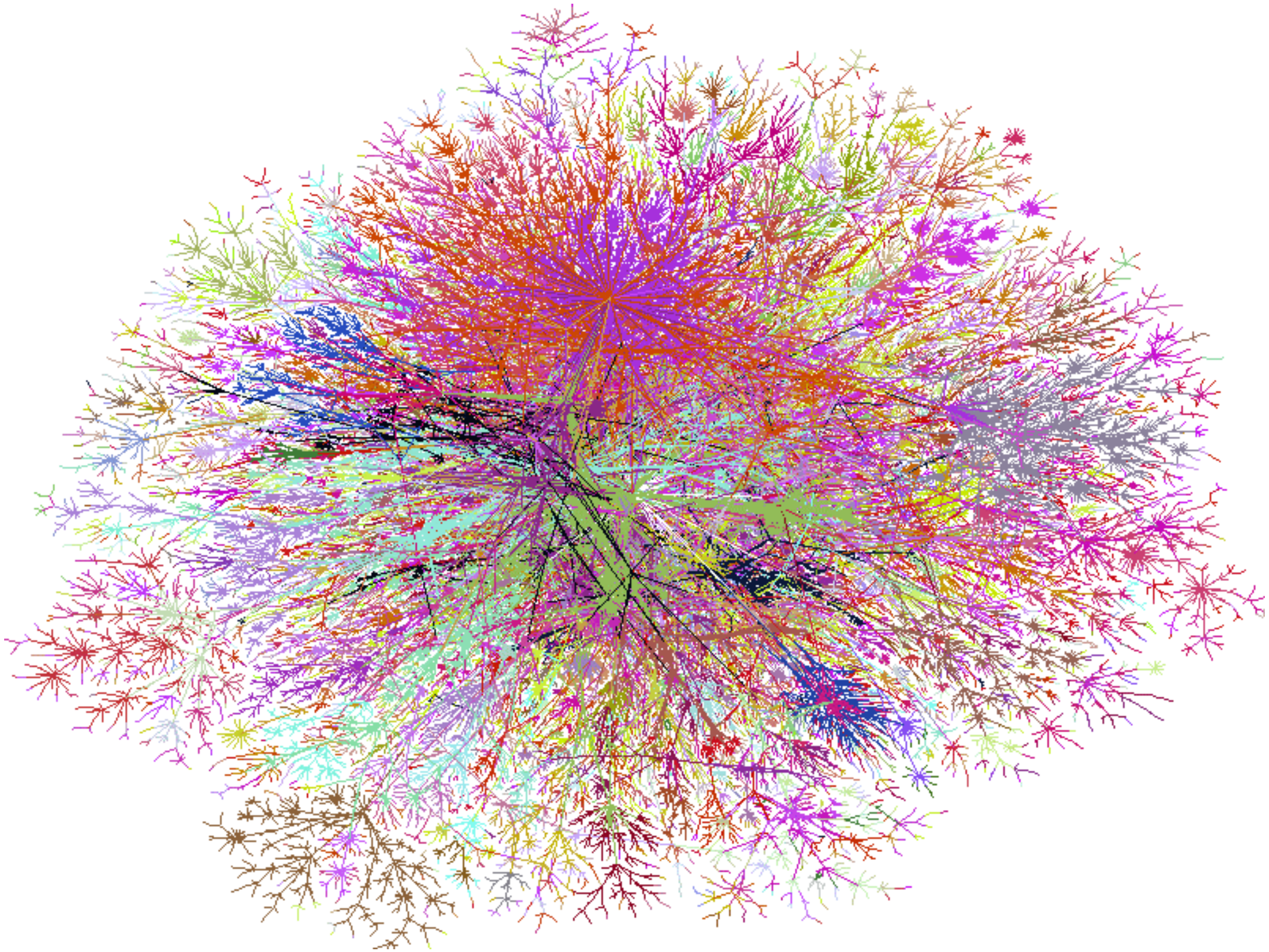
Uncloaking Terrorist Networks, by Valdis Krebs



UNIVERSITY OF CAMBRIDGE

The Internet

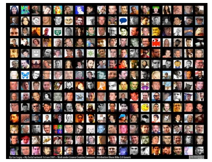


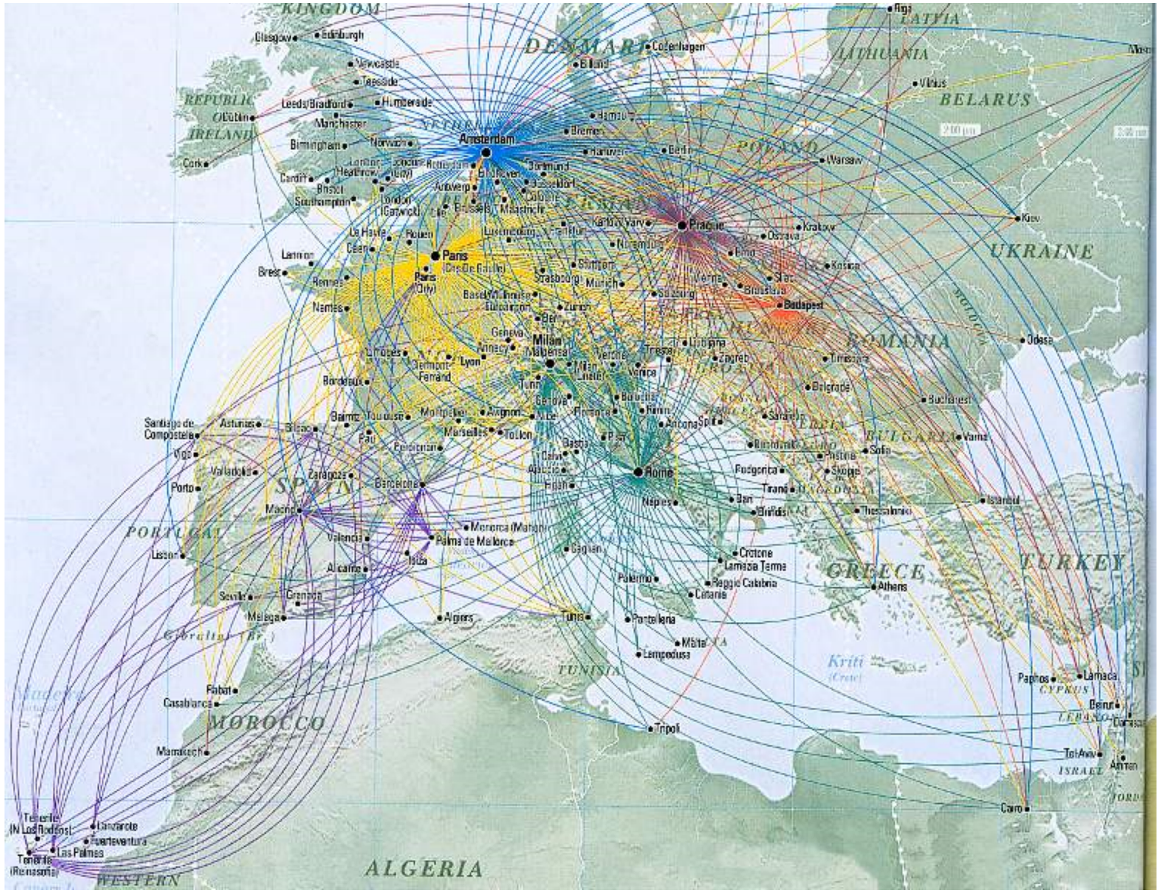


UNIVERSITY OF CAMBRIDGE

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

Airline Network

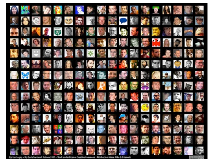


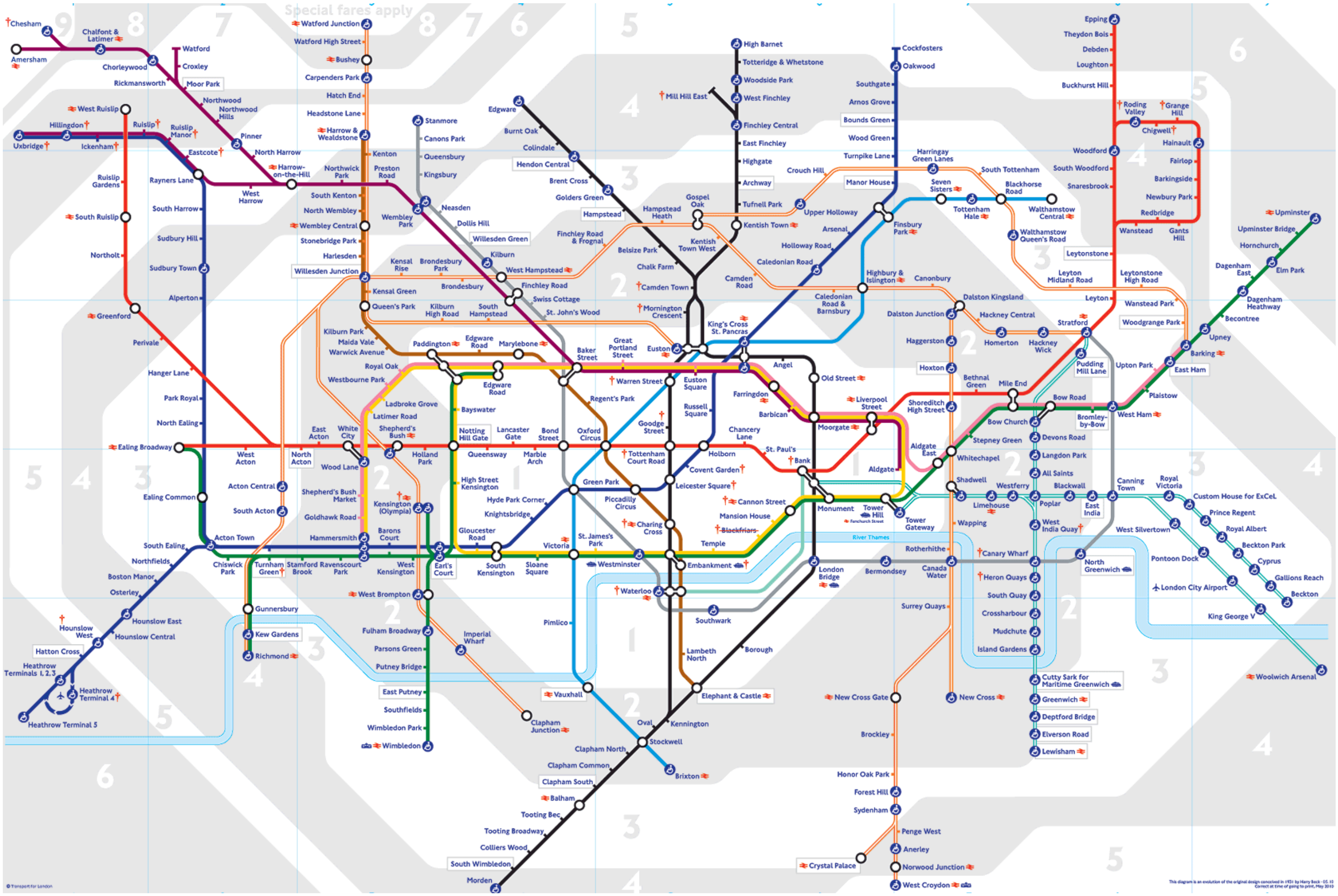


UNIVERSITY OF CAMBRIDGE

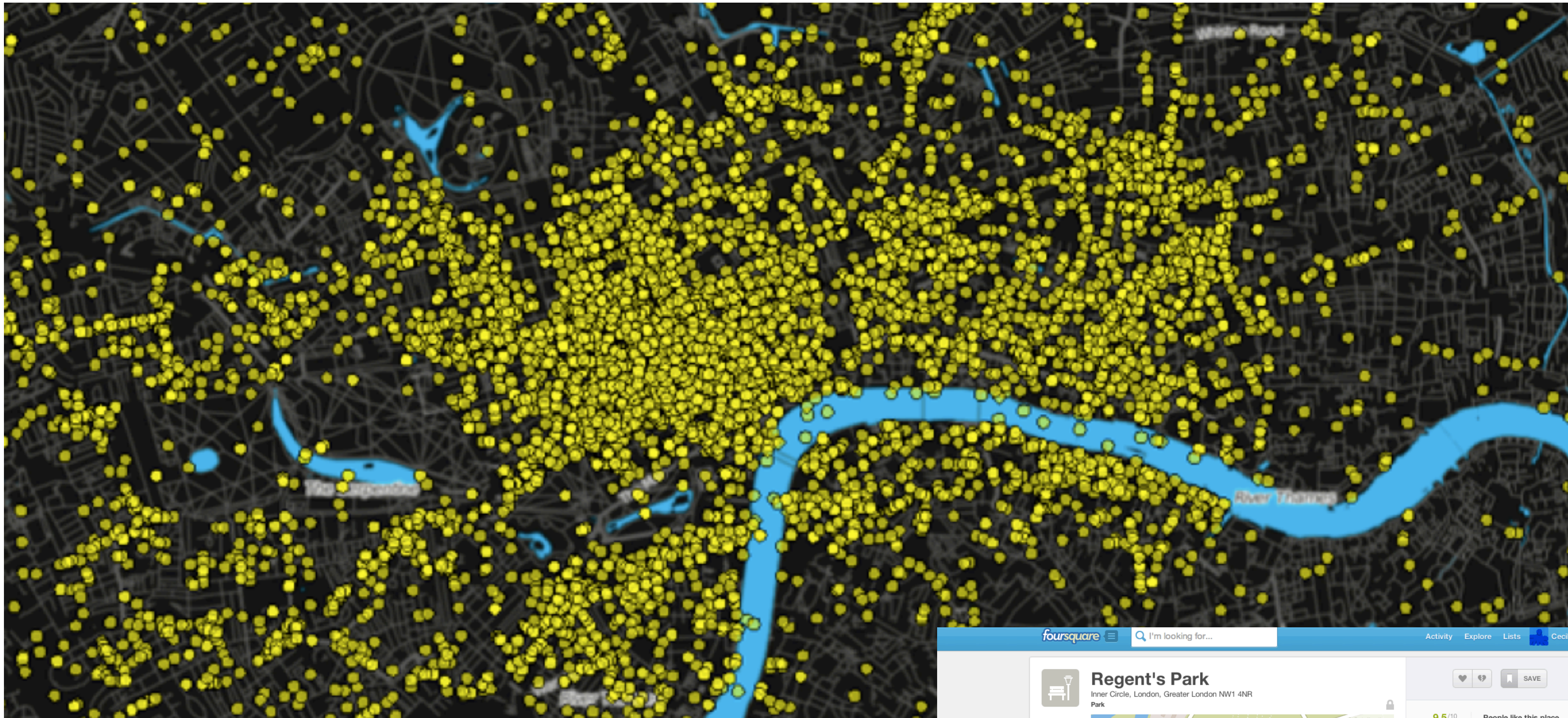
Source: Northwest Airlines WorldTraveler Magazine

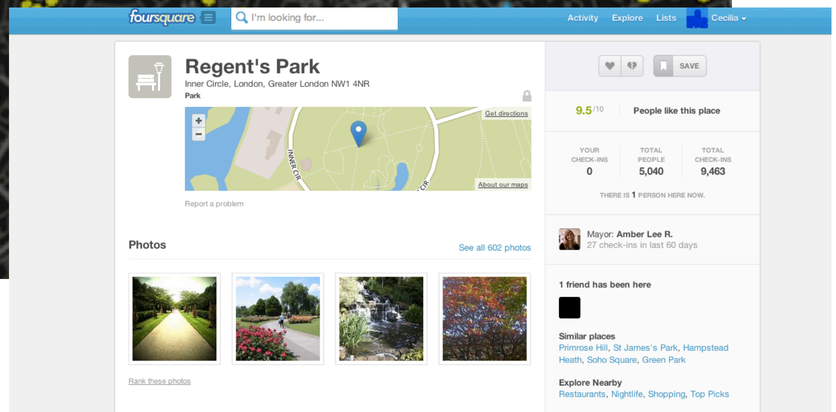
Railway/Metro Network





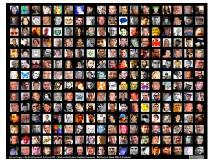
Geo-social Networks





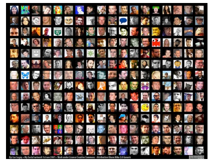
UNIVERSITY OF CAMBRIDGE

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

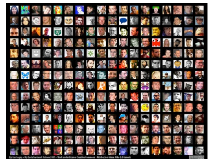


In This Course

- 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 Applications
- Lecture 4: Community Detection, Modularity, Overlapping Communities
- Lecture 5: Structure of the Web, Search and Power laws
- Lecture 6: Network Robustness and Applications
- Lecture 7: Information Cascades on Networks
- Lecture 8 and 9: 2h Practical Tutorial
- Lecture 10: Epidemic Dissemination on Networks
- Lecture 11: Cascades and Epidemics Applications
- Lecture 12: Time Varying Social Networks
- Lecture 13: Geo-Social Networks
- Lecture 14: Industrial Presentation
- Lecture 15-16: Student Presentations

UNIVERSITY OF CAMBRIDGE

Assessment

- All information on the course page: http://www.cl.cam.ac.uk/teaching/1213/L109/materials.html
- **One report** (of approximately 1,500 words) on one assigned research paper. The report is due on **12th February (noon)** 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**: 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 **7th March** and will be worth 50% of the final mark.
- Presentation of the findings of 2nd assignment on **8th March.** The presentation is worth 20% of the final mark.

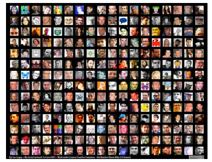
UNIVERSITY OF CAMBRIDGE

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

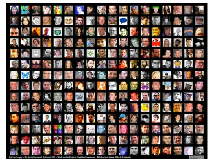


- First assignment: send me email immediately with a choice of paper **(and a backup choice)** from the website [assignment is first come first served: you will receive an email confirming which paper you are assigned to].
  - **Do this in the next TWO days.**

- Second assignment: same. At that point you will receive an email with a link to the dataset.
  - **Do this by 8th February.**

UNIVERSITY OF CAMBRIDGE

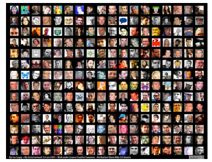
In This Lecture



- 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.

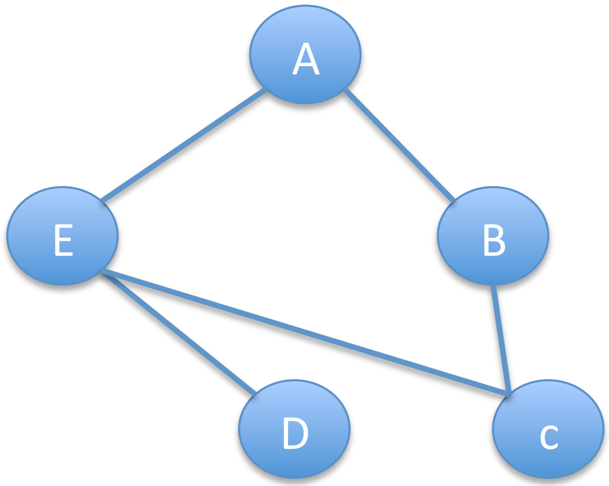
A **neighbour set** N(v) is the set of vertices adjacent to v:

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

UNIVERSITY OF CAMBRIDGE

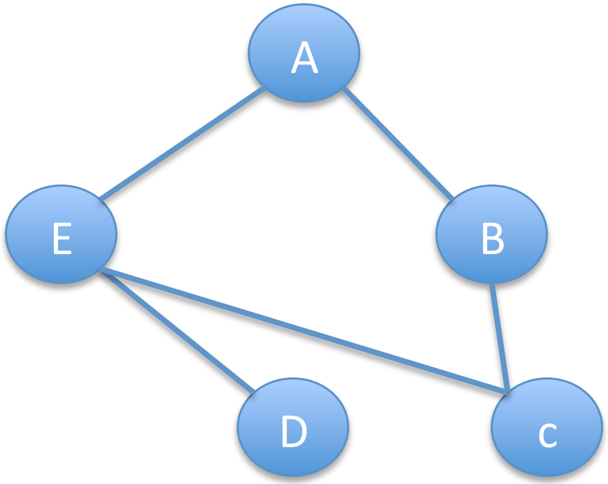
Node Degree

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

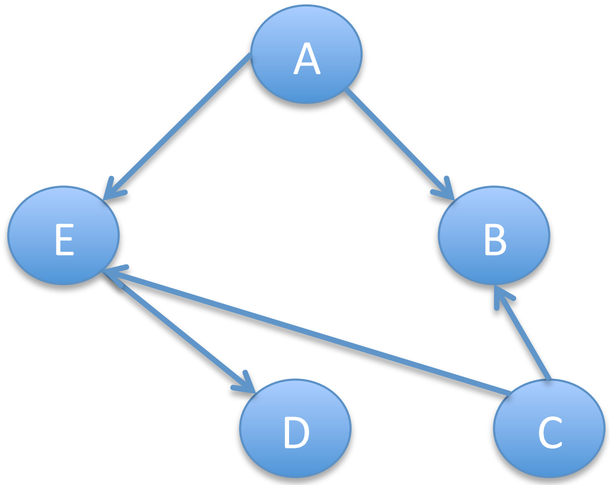


The study of the degree distribution of networks allows the classification of networks in different categories

Directed & Undirected Graphs



Undirected Graph



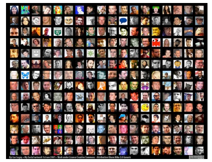
Directed Graph

Example of Undirected Graphs: Facebook, Co-presence

Examples of Directed: Twitter, Email, Phone Calls

UNIVERSITY OF CAMBRIDGE

Paths and Cycles



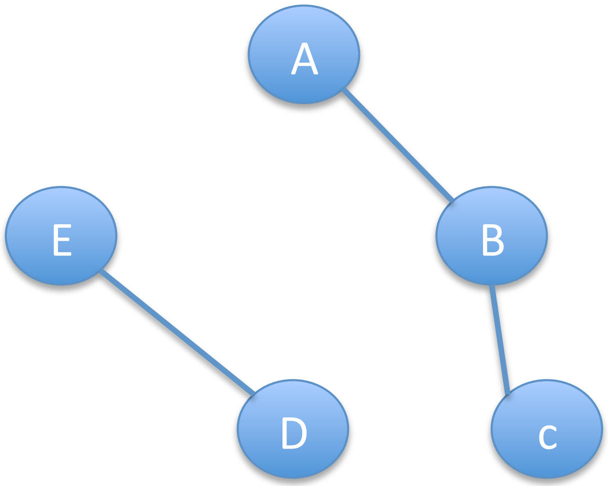
- A **path** is a sequence of nodes in which each pair of consecutive nodes is connected by an edge.
  - If graph is directed the edge needs to be in the right direction.
  - E.g. A-E-D is a path in both previous graphs
- A **cycle** is a path where the start node is also the end node
  - E.g. E-A-B-C is a cycle in the undirected graph

UNIVERSITY OF CAMBRIDGE

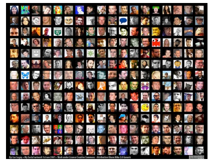
Connectivity

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

- Example of **disconnected** graph:



UNIVERSITY OF CAMBRIDGE

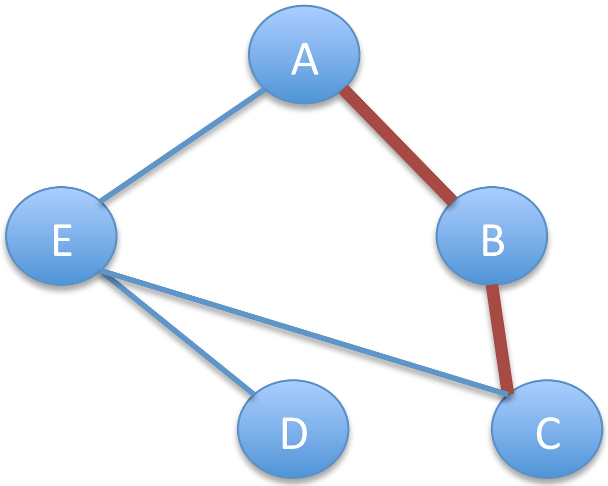


Components

- 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.

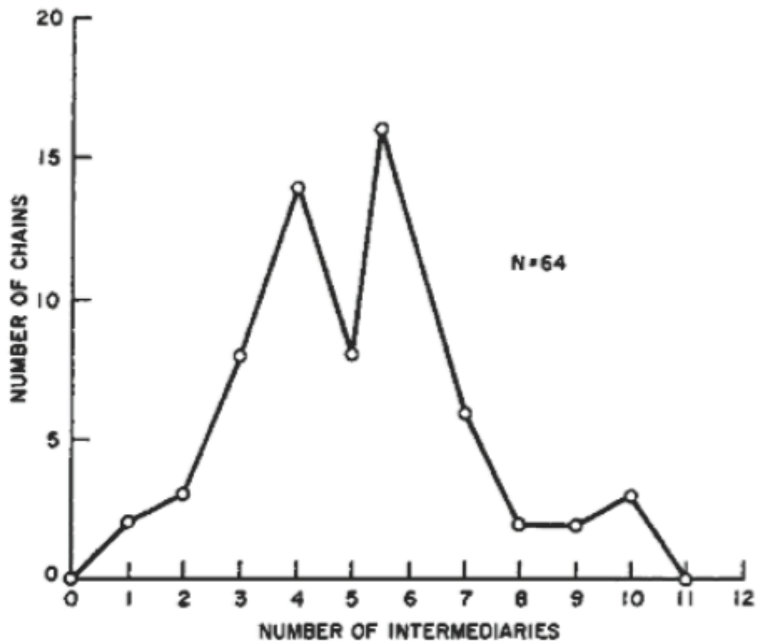


d(A,C)=2

UNIVERSITY OF CAMBRIDGE

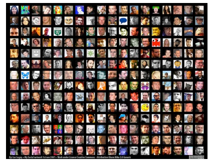
Small-world Phenomenon
Milgram's Experiment

- 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

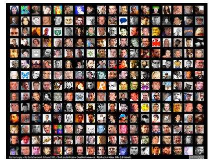
Recent Remake



- 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).

UNIVERSITY OF CAMBRIDGE

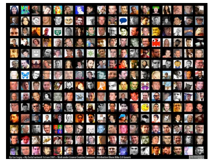
Findings



- 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 (eg professor)

UNIVERSITY OF CAMBRIDGE

It’s not resolved…




YAHOO! RESEARCH
SMALL WORLD EXPERIMENT

About the Experiment

The Small World Experiment is designed to test the hypothesis that anyone in the world can get a message to anyone else in just "six degrees of separation" by passing it from friend to friend. Sociologists have tried to prove (or disprove) this claim for decades, but it is still unresolved.

Now, using Facebook we finally have the technology to put the hypothesis to a proper scientific test. By participating in this experiment, you'll not only get to see how you're connected to people you might never otherwise encounter, you will also be helping to advance the science of social networks.

Become a Sender

We have already recruited a number of Target Persons from around the world.

Now we want you to try to reach them by becoming a Sender

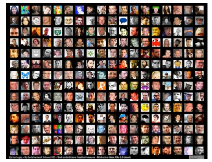
Click on the Participate Button below, and you'll be shown your assigned target. Then you'll get to choose a friend to pass the message to. That person will then get the same instructions, and so on....

If everyone passes the messages along, your message will reach the target. How many steps will it take? There's only one way to find out.

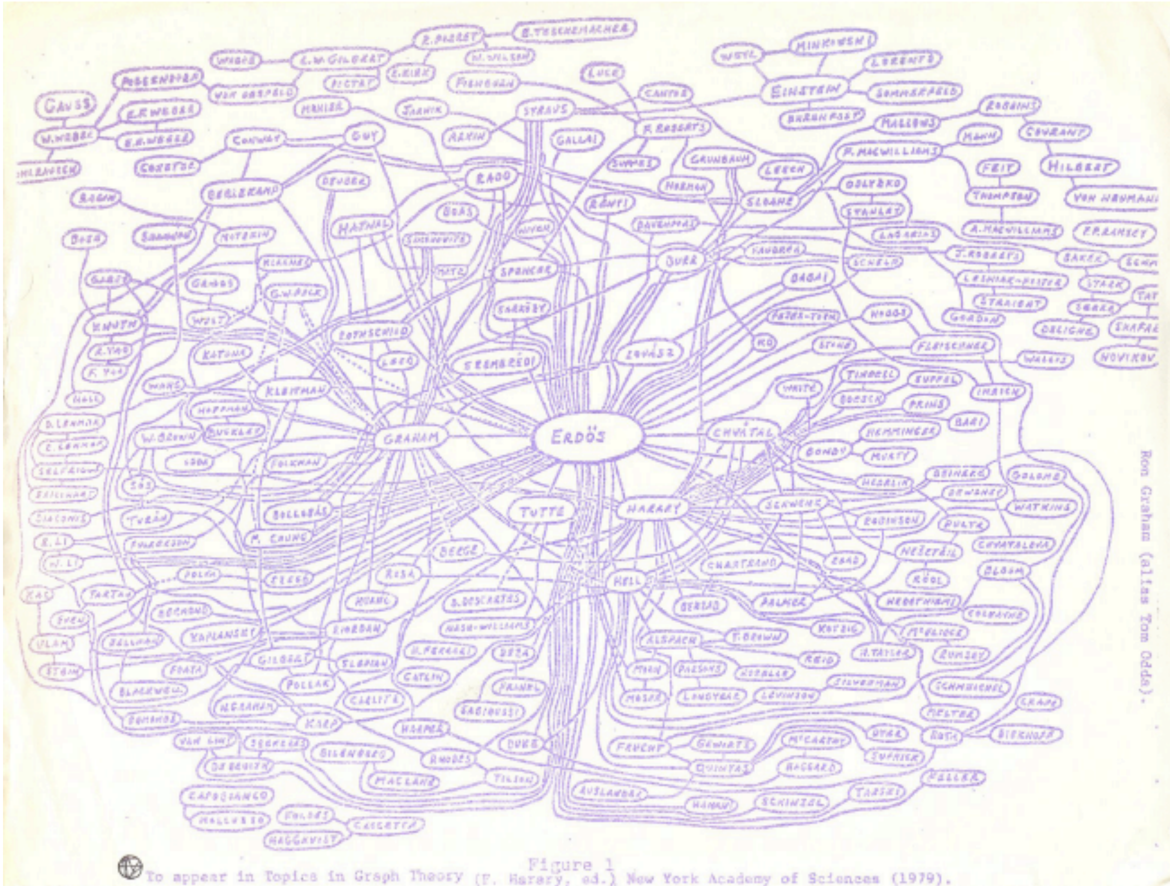
Continue

Privacy Policy | Discontinue from Experiment | Become a Target | My Chains | Terms of Service | Contact Us

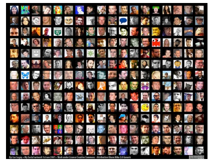
Erdos Number



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



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.

UNIVERSITY OF CAMBRIDGE

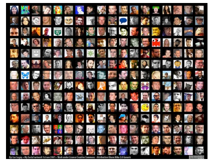
Random Graphs

- 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:

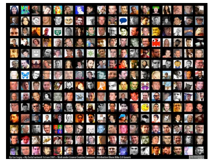
$$P(k_i = k) = C_{N-1}^{k} p^{k} (1-p)^{N-1-k}$$

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

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

UNIVERSITY OF CAMBRIDGE

Random Graphs Properties

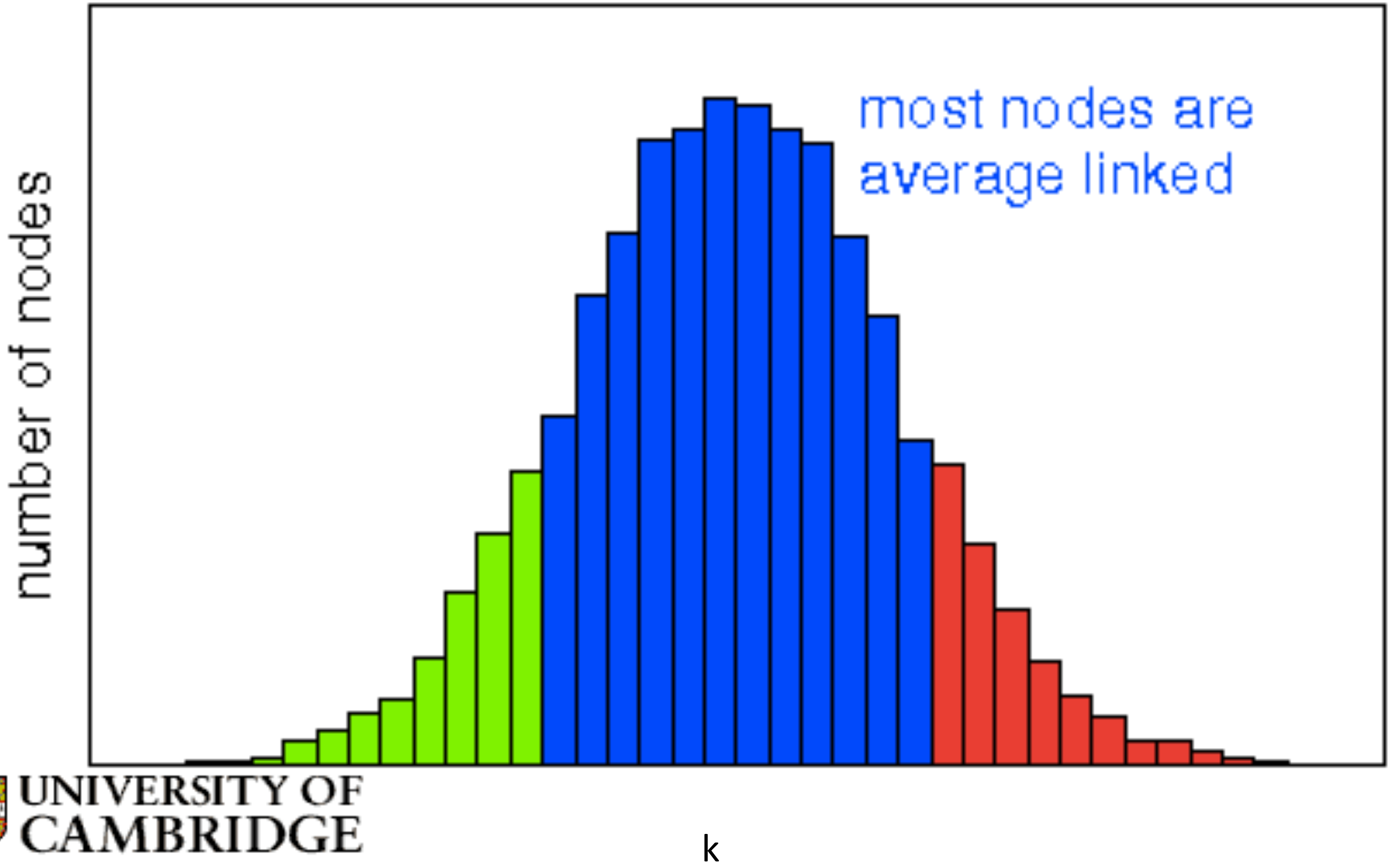


- For large N this is approximated by the Poisson distribution with

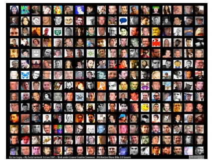
$$P(k) \approx e^{-Np}\frac{(Np)^k}{k!} = e^{-\langle k \rangle}\frac{\langle k \rangle^k}{k!}$$

UNIVERSITY OF CAMBRIDGE

Degree Distribution of Random Graphs



UNIVERSITY OF CAMBRIDGE



Random Graph Diameter

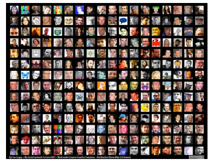
- The **diameter** of random graph and the **average path length** of the graph have been demonstrated to be:

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

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

UNIVERSITY OF CAMBRIDGE

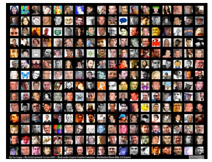
Relationship of <k> and connectivity

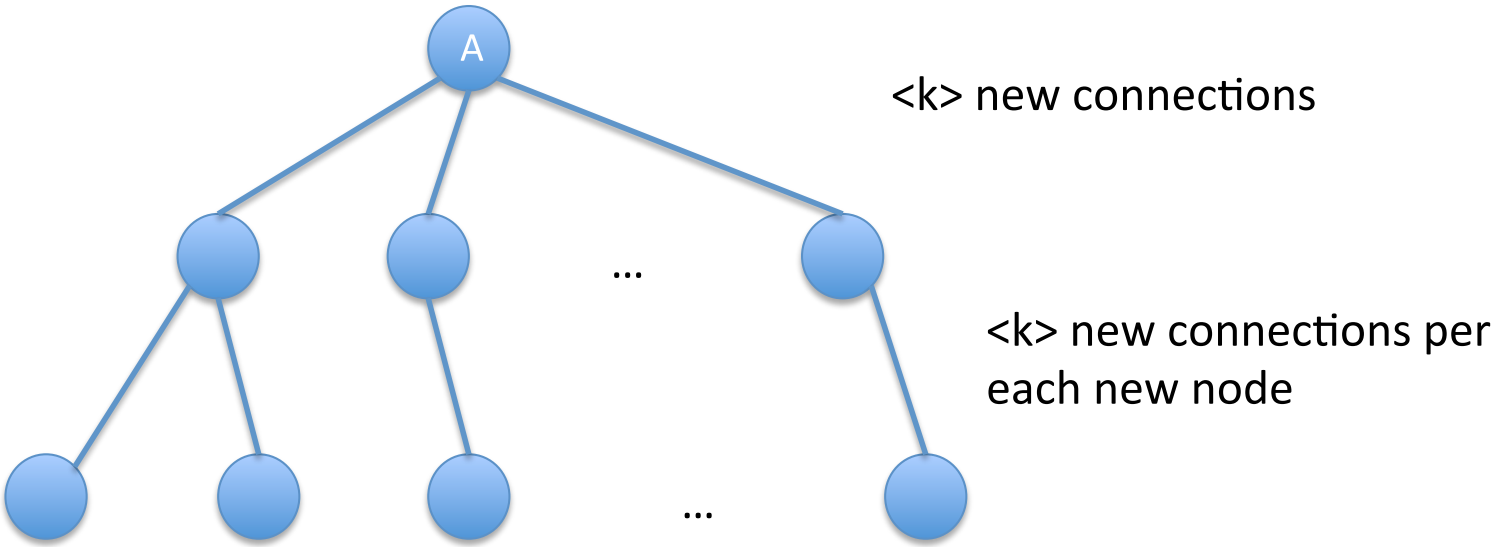


- <k> = average degree (np)
- If <k> < 1 disconnected network
- If <k> > 1 a giant component appears
- If <k> >= ln(N) graph is totally connected

UNIVERSITY OF CAMBRIDGE

Random Graph Diameter: An Intuition





- The nodes at distance l from A will be $\sim \langle k \rangle^{l}$

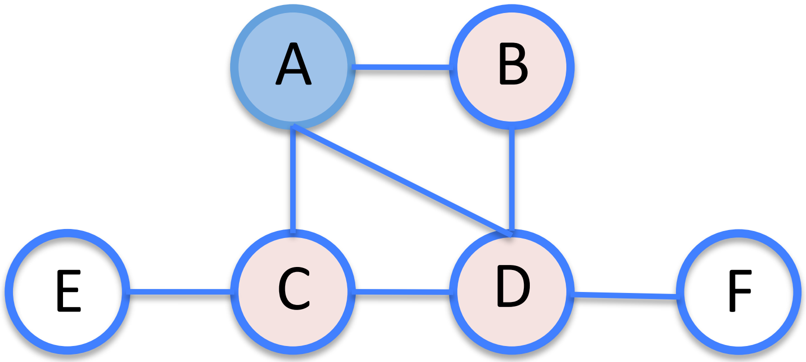
$N = k^l \quad \log N = l \log k$

$l = \log N / \log k$

UNIVERSITY OF CAMBRIDGE

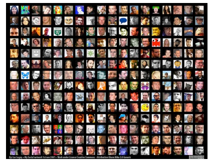
Clustering Coefficient

- 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

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

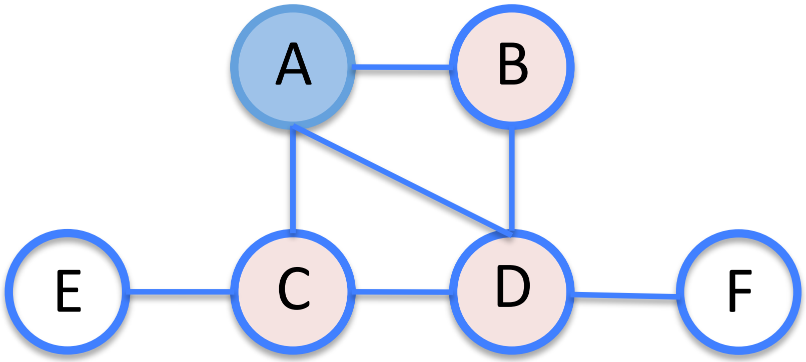
Network Clustering Coefficient

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

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

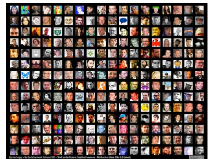
$$C_{rand} = p = \frac{\langle k \rangle}{N}$$

$$p * \binom{n_v}{2} / \binom{n_v}{2} = p$$

- 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?

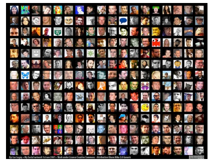
UNIVERSITY OF CAMBRIDGE

Question



- Are Random Graphs representatives of Real Networks?

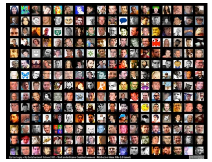
UNIVERSITY OF CAMBRIDGE



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.**
- 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 .

UNIVERSITY OF CAMBRIDGE