Digital Media and Social Networks

Laurissa Tokarchuk

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

WEEK 2:
SMALL WORLD AND WEAK TIES

IN THIS LECTURE

- We will compare random networks with real networks
- We will introduce the concept of small world networks
- We will introduce the concept of weak ties and illustrate their importance

LAST WEEK in review

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.

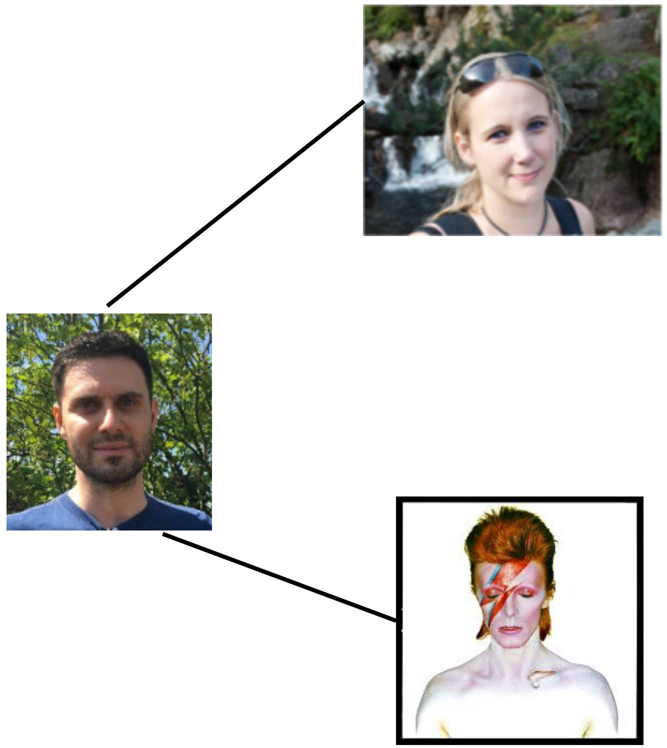
Used to model **connectivity** between **things**

in review

V = { Laurissa, Minos, } for the vertices

E = {(Laurissa, Minos), (Minos,David)} for the edges

G = (V, E)



in review

We have **Directed** and **Undirected** graphs

Your **Neighbour Set** is the set of nodes that neighbour you in the graph

Node Degree is the number of nodes in your Neighbour Set

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

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

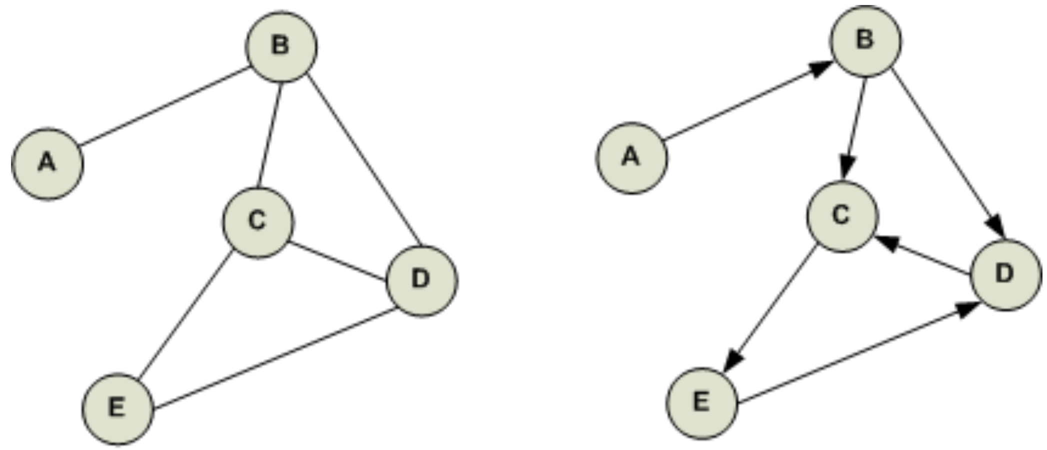


Fig 1. Undirected Graph

Fig 2. Directed Graph

Image: https://algorithmsinsight.wordpress.com/graph-theory-2/

in review

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

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

A **connected component** of a graph is the subset of nodes for which each of them has a path to all others

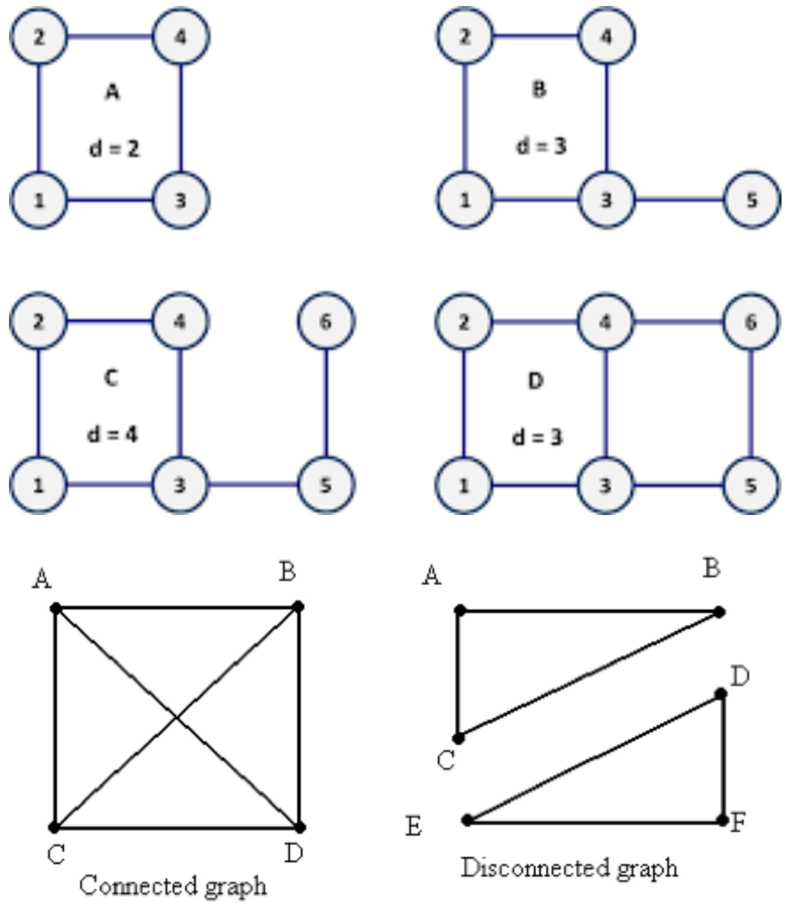
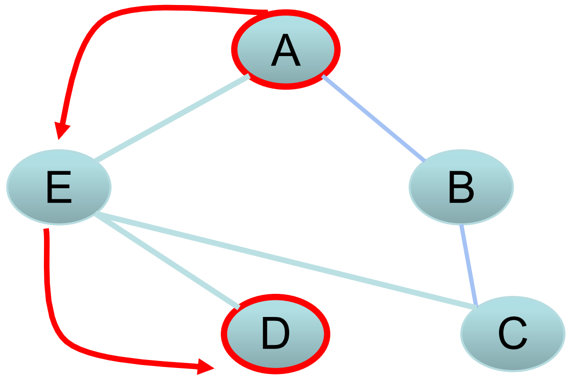


Image: https://people.hofstra.edu/geotrans/eng/methods/diameter2.html and http://aix1.uottawa.ca/~jkhoury/graph.htm

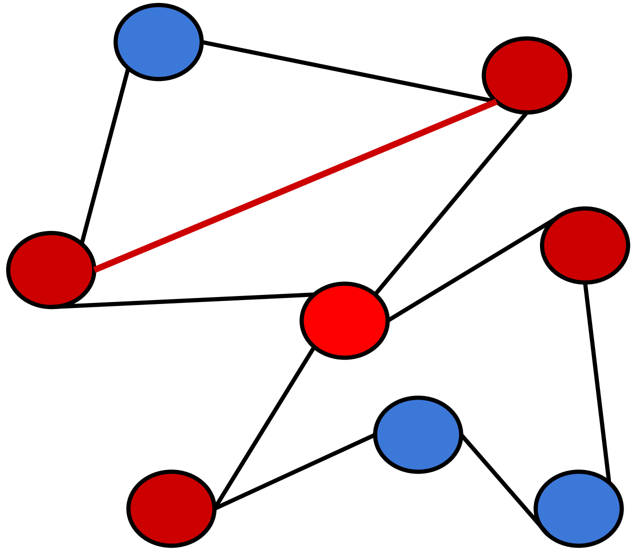
in review



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

in review

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



The **Clustering Coefficient** is the proportion of triangles formed between my friends (from the maximum number of triangles)

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

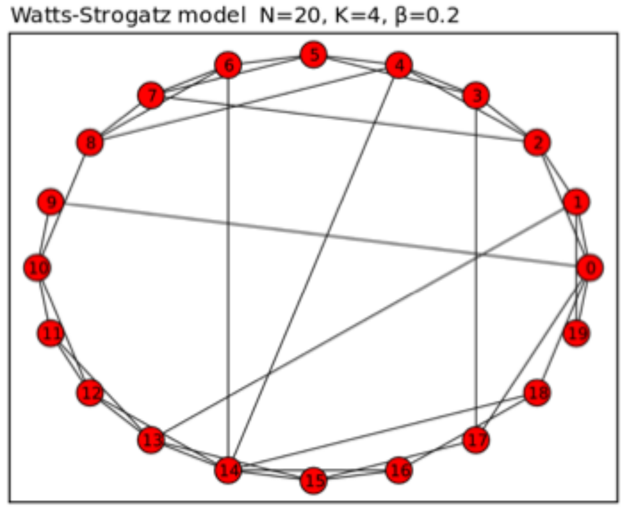
SMALL WORLD AND WEAK TIES

WHAT DO REAL NETWORKS LOOK LIKE?

- Real world example.. Social networks, computer networsk, etc. What metrics does this map into?

  - Short path length
    - We have a wide reach in our networks without requiring that many hops
  - High clustering coefficient
    - Yet we form tight groups of friends often

CLUSTERING COEFFICIENT OF REAL NETWORKS



- [Watts and Strogatz, Nature1998]
- short average path lengths and high clustering.
- Characteristic path length and clustering coefficient for some real networks and for random networks with same number of nodes and average number of edges per node.

- Aim is to check if random graphs can model real networks.

REAL NETWORKS VS RANDOM NETWORKS

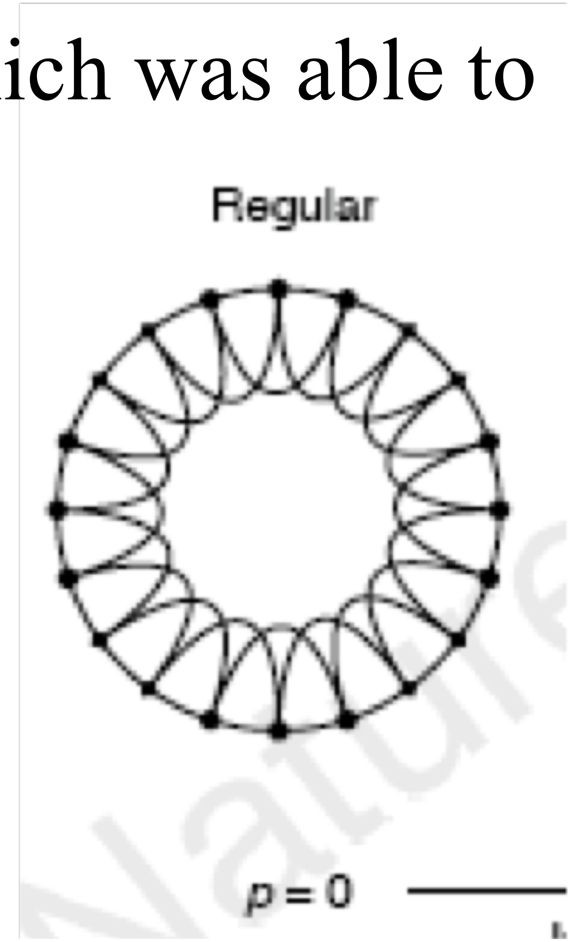
- Film Actors: actors in movies together
- Power grid: the network of the electricity generators
- C. elegans: network of neurons of a worm
    - **L is comparable while C is very different**

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

SMALL WORLD MODEL

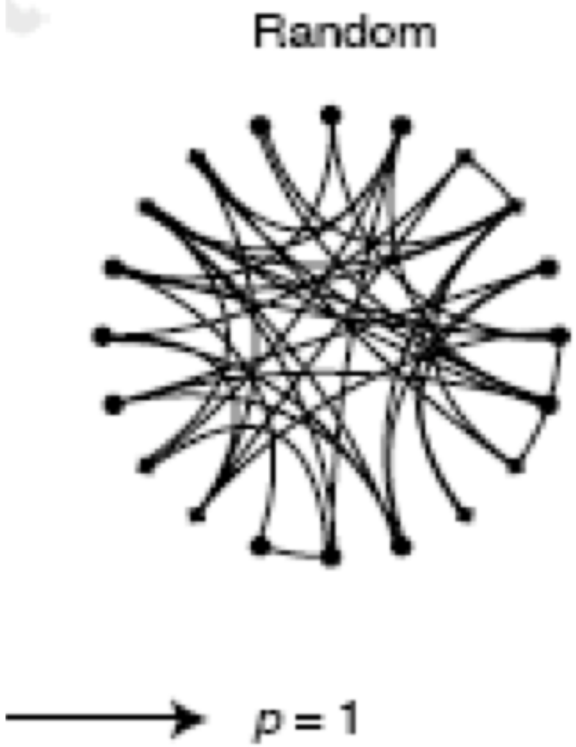
- Watts & Strogatz built a model which was able to capture these characteristics.
- How?
  - Start with regular lattice:



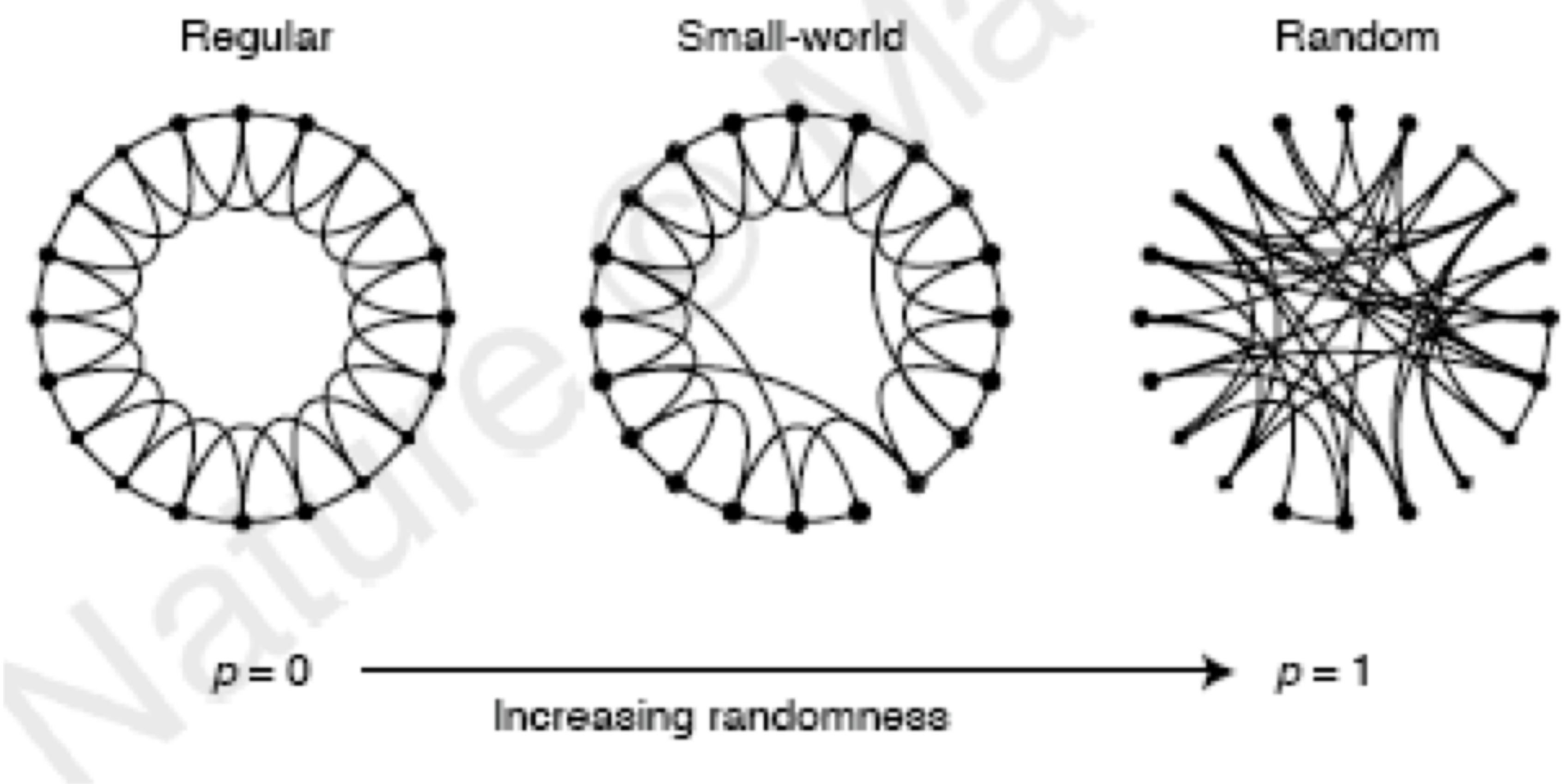
SMALL WORLD MODEL

How?

- *Increase a probability p of "rewiring" a node to another node.*
    1. *Choose node and edge to nearest neighbour.*
    2. *With prob p, reconnect it to random node, else leave edge alone.*
    3. *Repeat for all original edges*
- *When p very high the lattice would become a random graph.*



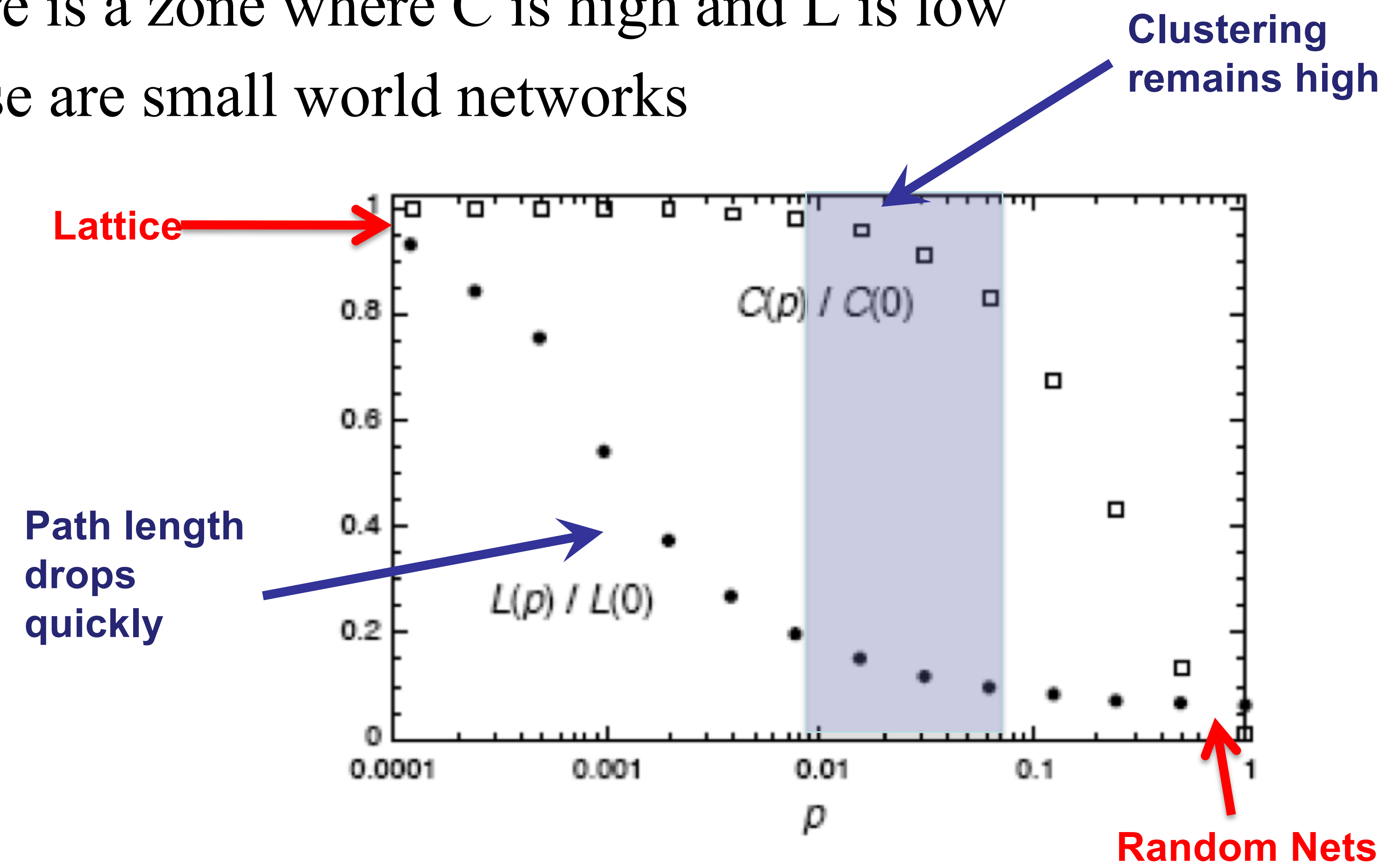
SMALL WORLD MODEL (2)



HOW ARE L AND C IN THIS MODEL?

There is a zone where C is high and L is low

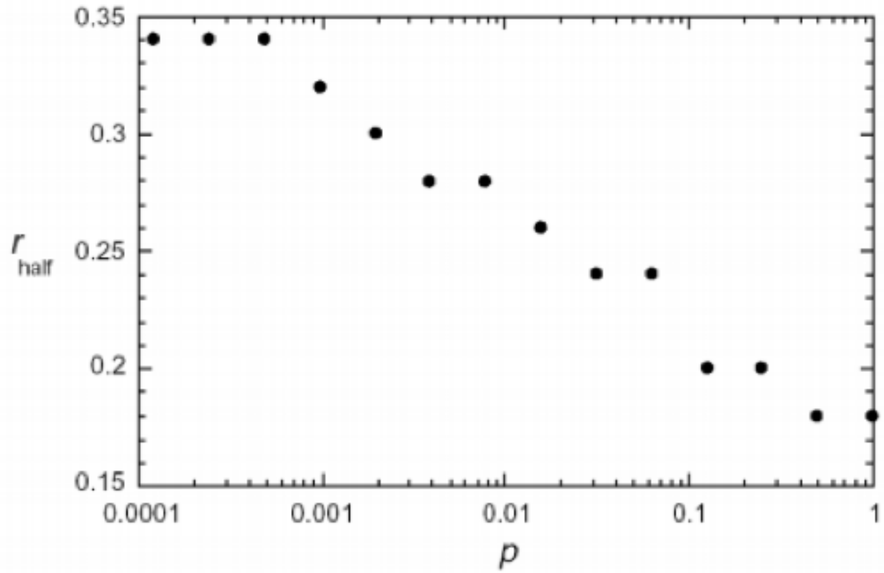
These are small world networks



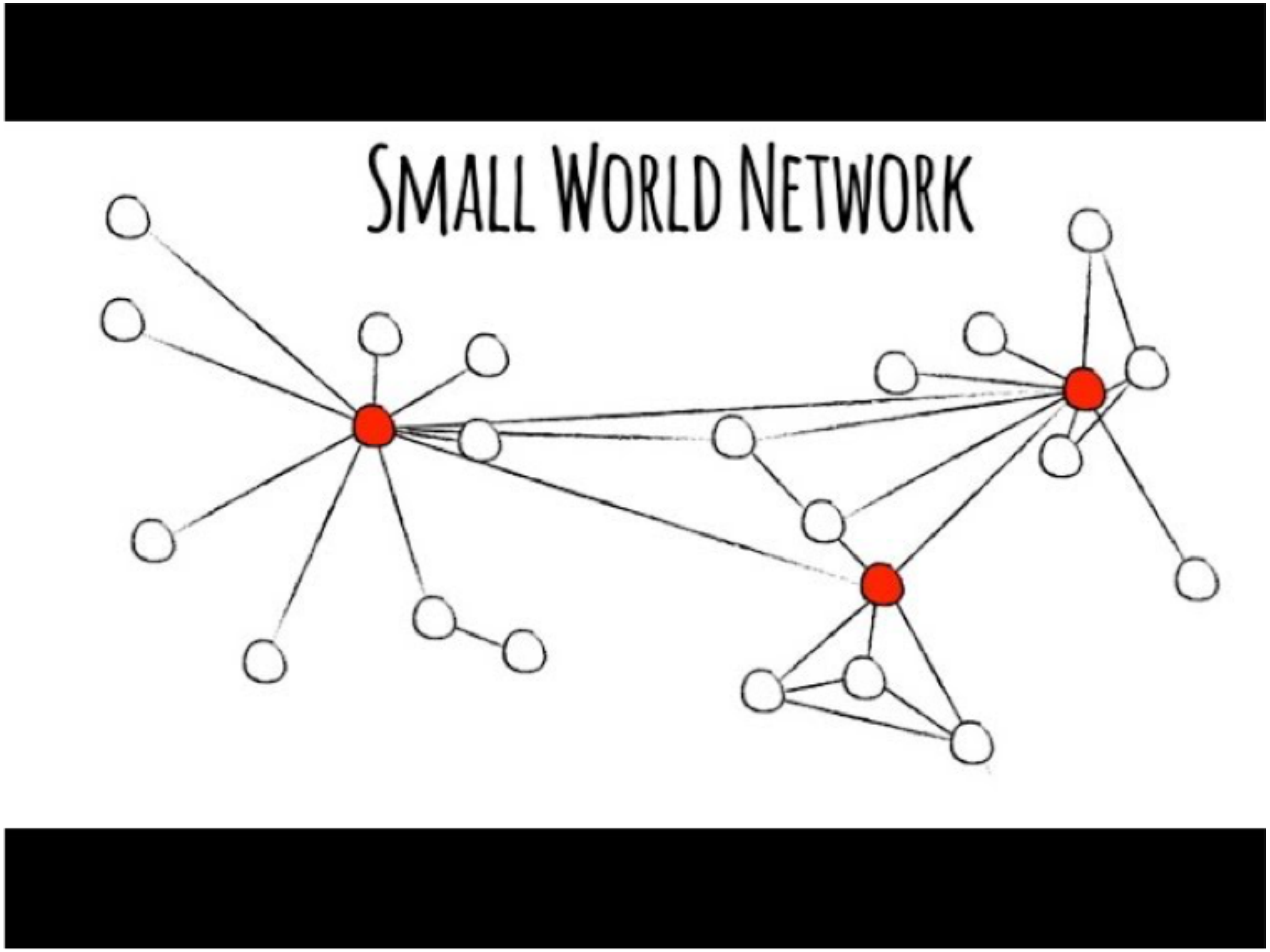
WHAT WOULD THIS GRAPH BE GOOD AT?

- Discovering files in a Gnutella topology
- Spreading information in Twitter
- Headhunting a great candidate for a job
- Passing on diseases

Disease spreads faster as *p* increases



WHAT WOULD THIS GRAPH BE GOOD AT?



https://youtu.be/Lq5hIsJAOfc

OTHER REAL NETWORKS EXAMPLES

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

SMALL-WORLD PHENOMENON:

L(real) ≅ L(random)

The real networks still have very have much higher C!

ANALYSIS OF MESSENGER NETWORK

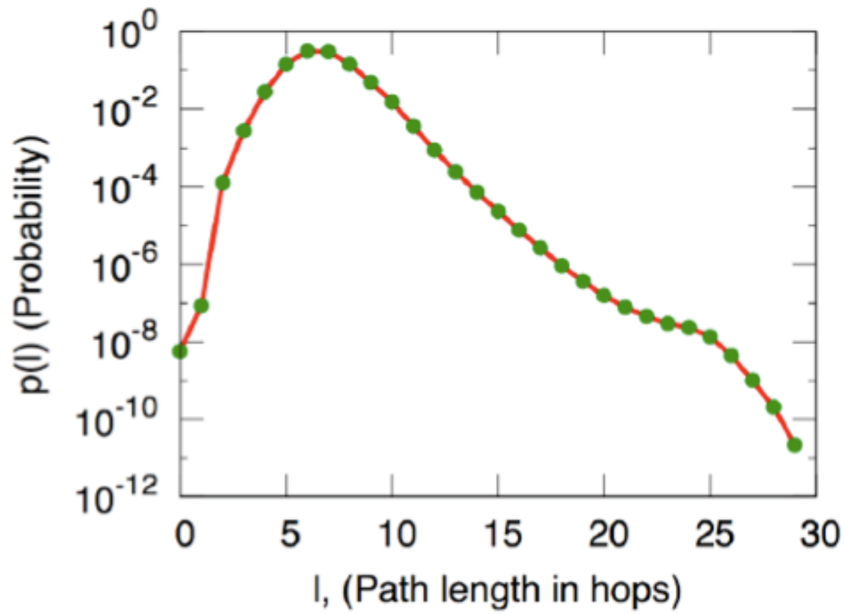
- [Leskovec and Horvitz 2008] analyzed a large dataset of the Microsoft Messenger.
- Communication Network contained 180 million users and 1.3 billion conversations in 1 month.
- Buddy Network contained 240 million users.
- *99.9% users belonged to a connected component.*

ANALYSIS OF A MESSENGER NETWORK

Average shortest path is 6.6 (confirming Milgram's study).

Although some longer paths up to 29.

Average clustering coefficient is quite high: 0.137.

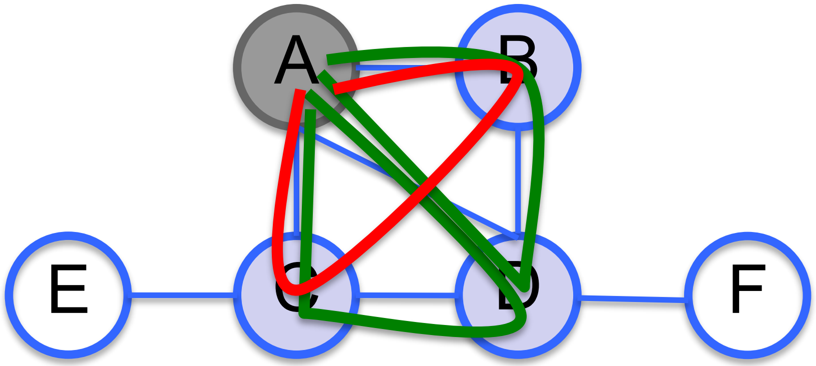


AGAIN ON CLUSTERING COEFFICIENT

We have introduced the clustering coefficient. This indicates:

- The number of triangles including node A.
- How connected the friends of A are.

Triadic closure: if C and B are connected to A there is an increased likelihood that they will be connected among themselves in future.



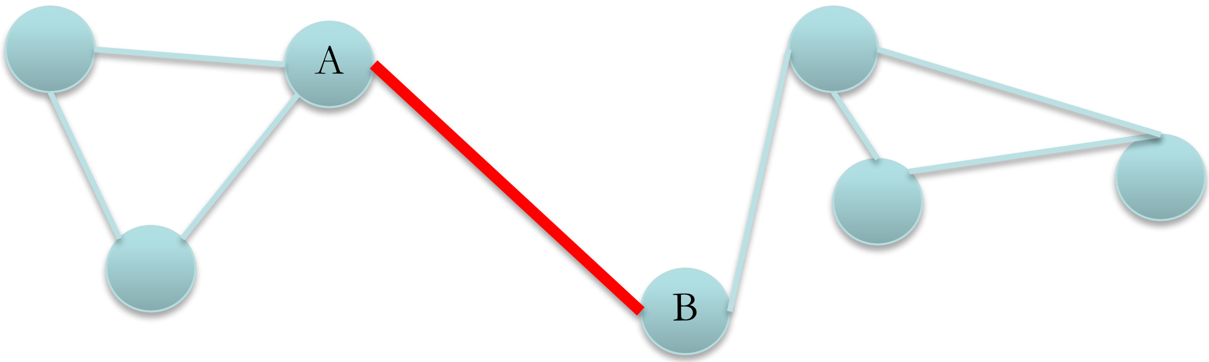
[GRANOVETTER'74]

- Granovetter interviewed people about how they discovered their jobs
  - Most people did so through personal contacts
  - Often the personal contacts described as acquaintances and not close friends

- Basic intuition on this is: close friends are part of triad closures and would know what you know and would know others who would know what you know
- We will explain this more formally…

BRIDGES

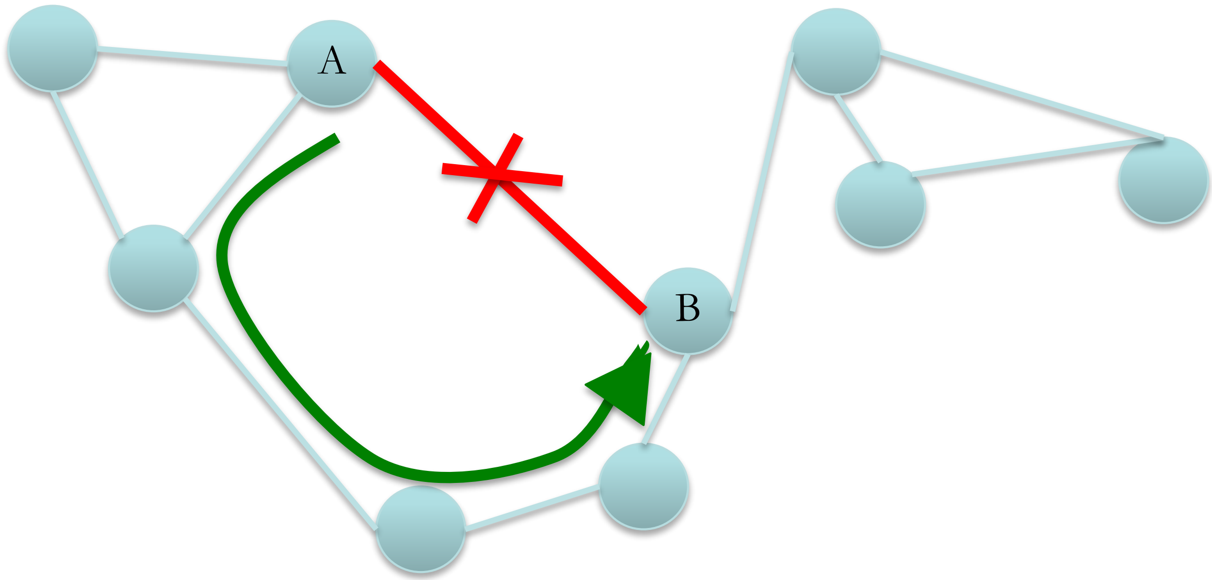
Edge between A and B is a **bridge** if, when deleted, it would make A and B lie in 2 different components



LOCAL BRIDGES

An edge is a local bridge if its endpoints have no friends in common

- If deleting the edge would increase the distance of the endpoints to a value more than 2.



STRONG TRIADIC CLOSURE PROPERTY (STPC)

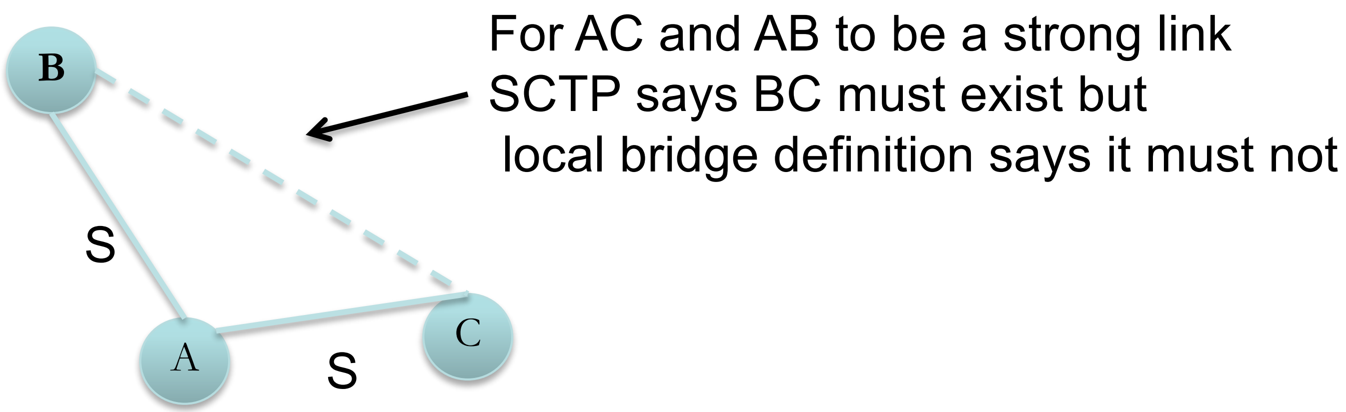
Links between nodes have different “value”: **strong and weak ties**

E.g: Friendship vs acquaintances

Strong Triadic Closure Property (Granovetter): If a node A has two strong links (to B and C) then a link (strong or weak) must exist between B and C.

LOCAL BRIDGES AND WEAK TIES

If node A satisfies the SCTP and is involved in at least two strong ties then any local bridge it is involved in must be a weak tie. (Proof by contradiction)



(assuming SCTP) If there are enough strong ties in the network then local bridges must be weak ties

REAL DATA VALIDATION

Granovetter's theory about the importance of weak ties remained not validated for years for large social networks due to the lack of data.

[Onnela et al '07] tested it over a large cell-phone network (4 millions users):

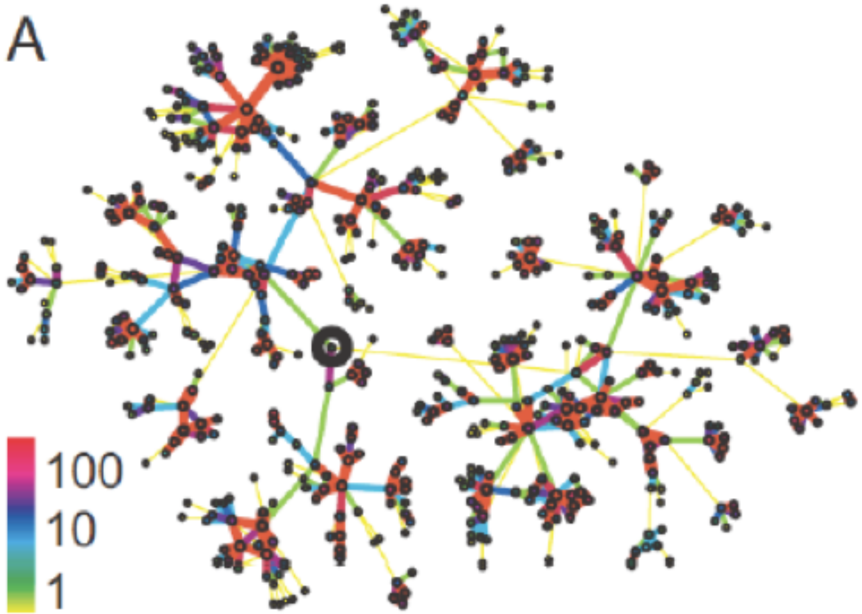
Edge between two users if they called each other within the 18 months period.

Edge weight: time spent in conversation.

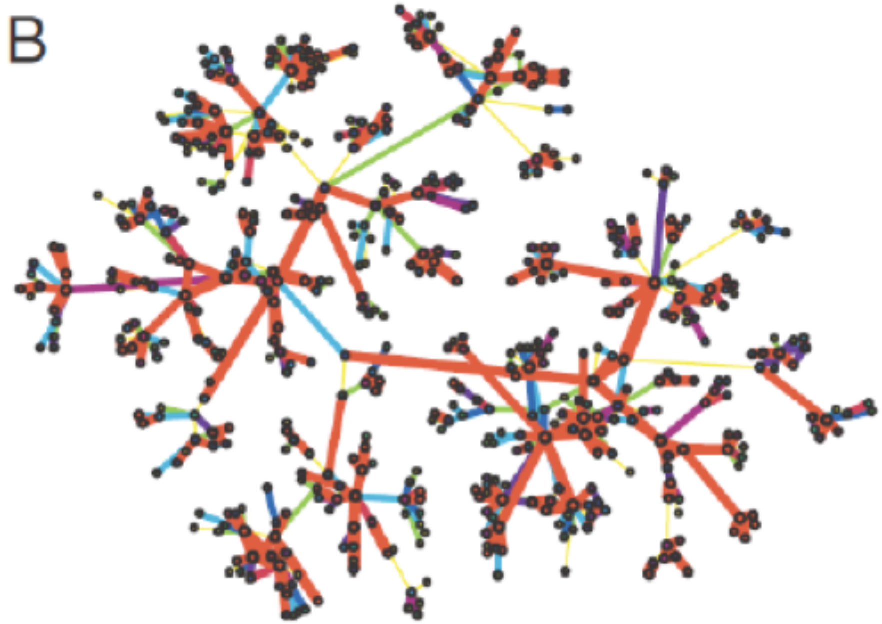
Data exhibits a giant component (84%).

http://www.pnas.org/content/104/18/7332.long

REAL TIE WEIGHTS IN A PORTION OF THE GRAPH (AROUND A RANDOM NODE)



A= Real
B= Randomly shuffled



WHAT DOES IT MEAN?

(INTERCOMMUNITY TIES(<u>BRIDGES</u>) VS INTRACOMMUNITY TIES(LOCAL ROADS)

<u>Extending the defini</u>tion of local bridge

Given:

Neighbourhood overlap:

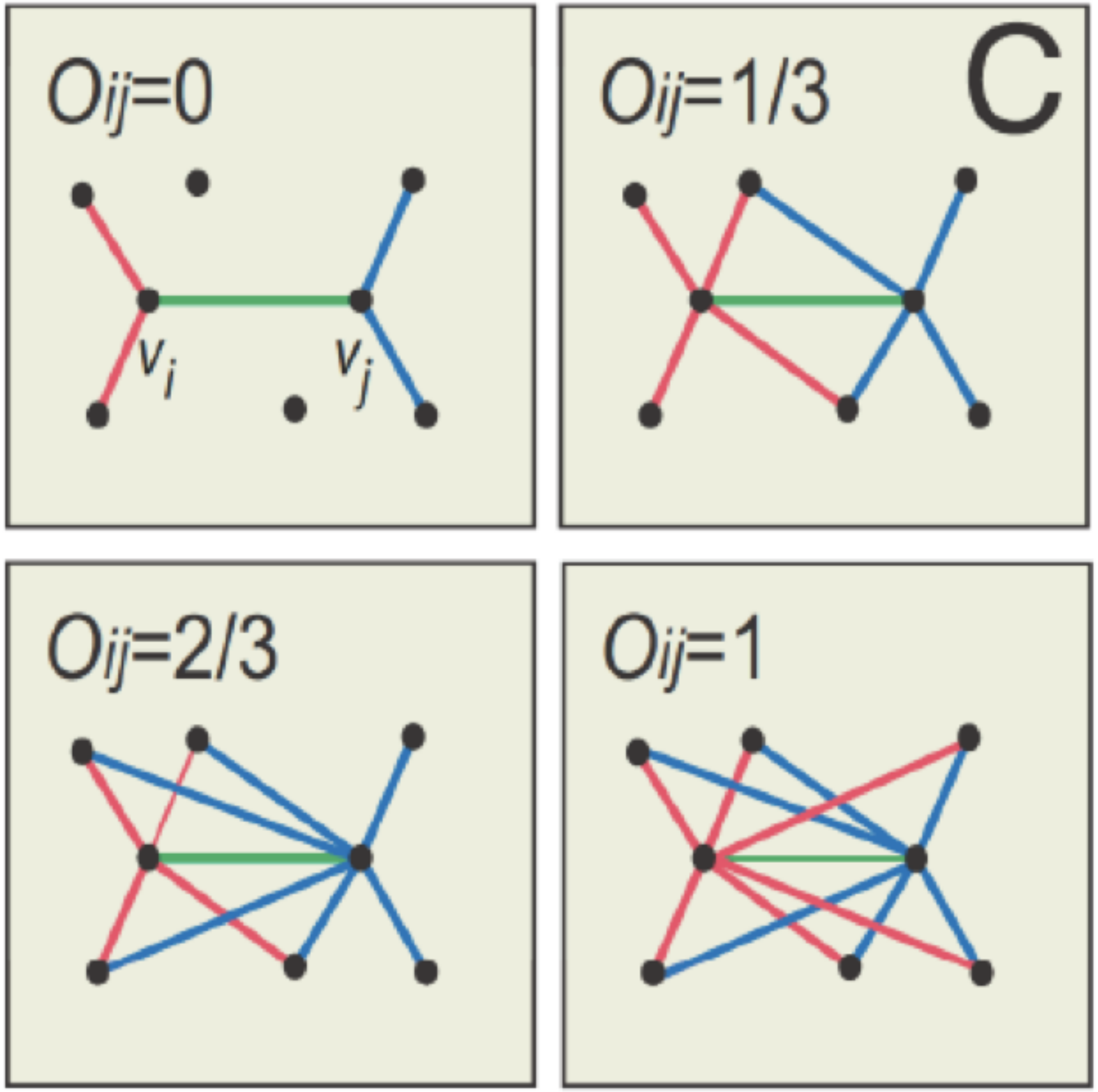
$$\frac{\text{Number of nodes who are neighbours of both A \& B}}{\text{Number of nodes who are neighbours of at least A or B}}$$

When the numerator is 0 the quantity is 0.

Numerator is 0 when AB is a local bridge

The definition finds “almost local bridges” (~0)

NEIGHBOURHOOD OVERLAP

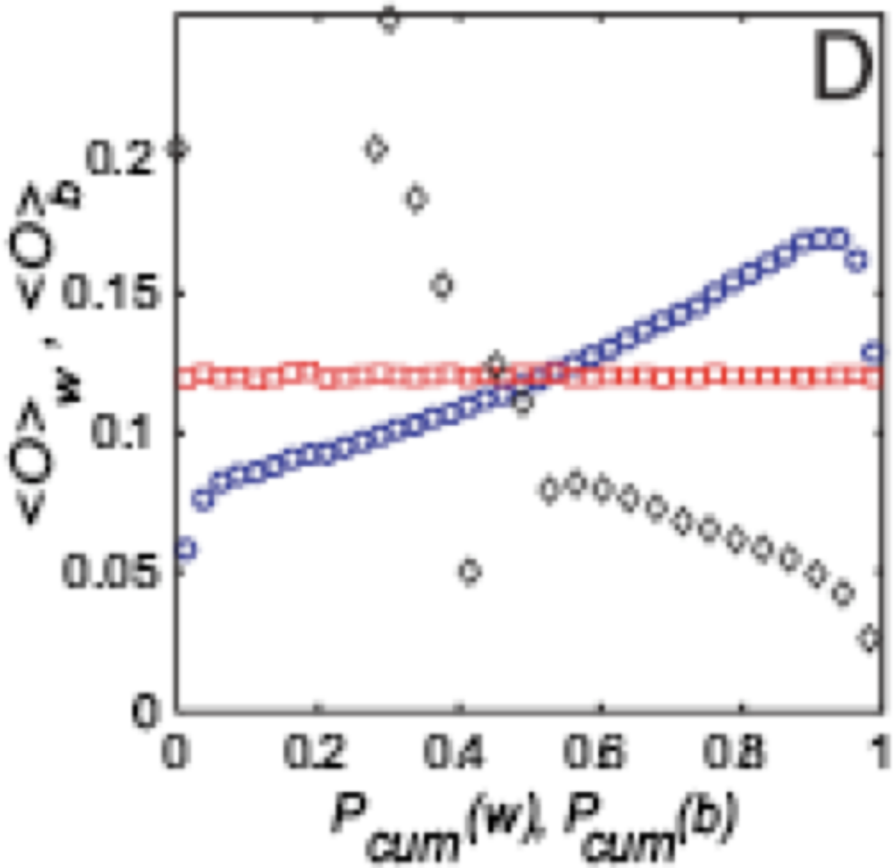


RELATIONSHIP OF OVERLAP WITH TIE STRENGTH

Red: random shuffled weights over links.

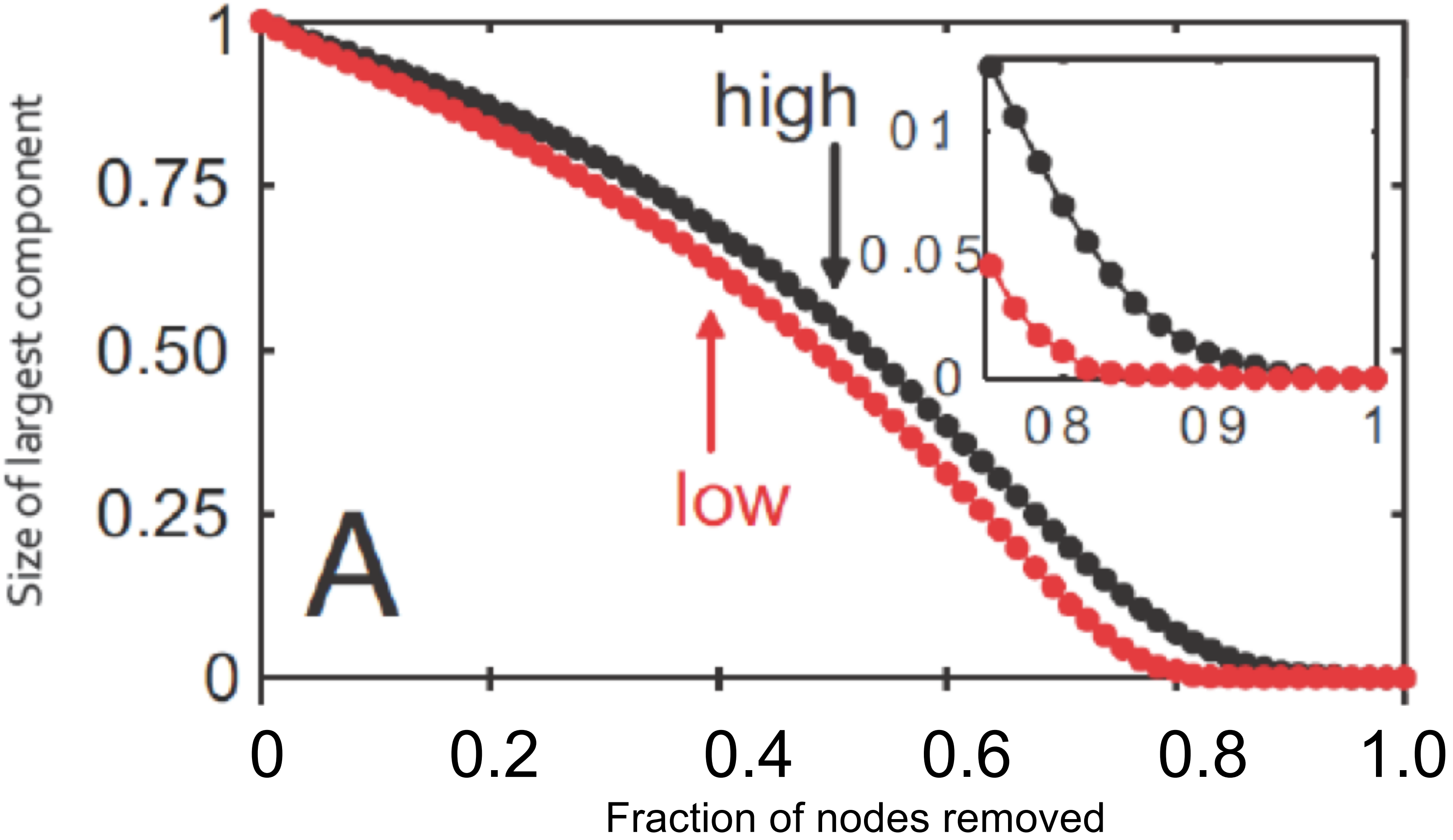
Blue: real ones. Correlation with tie strength.

Overlap

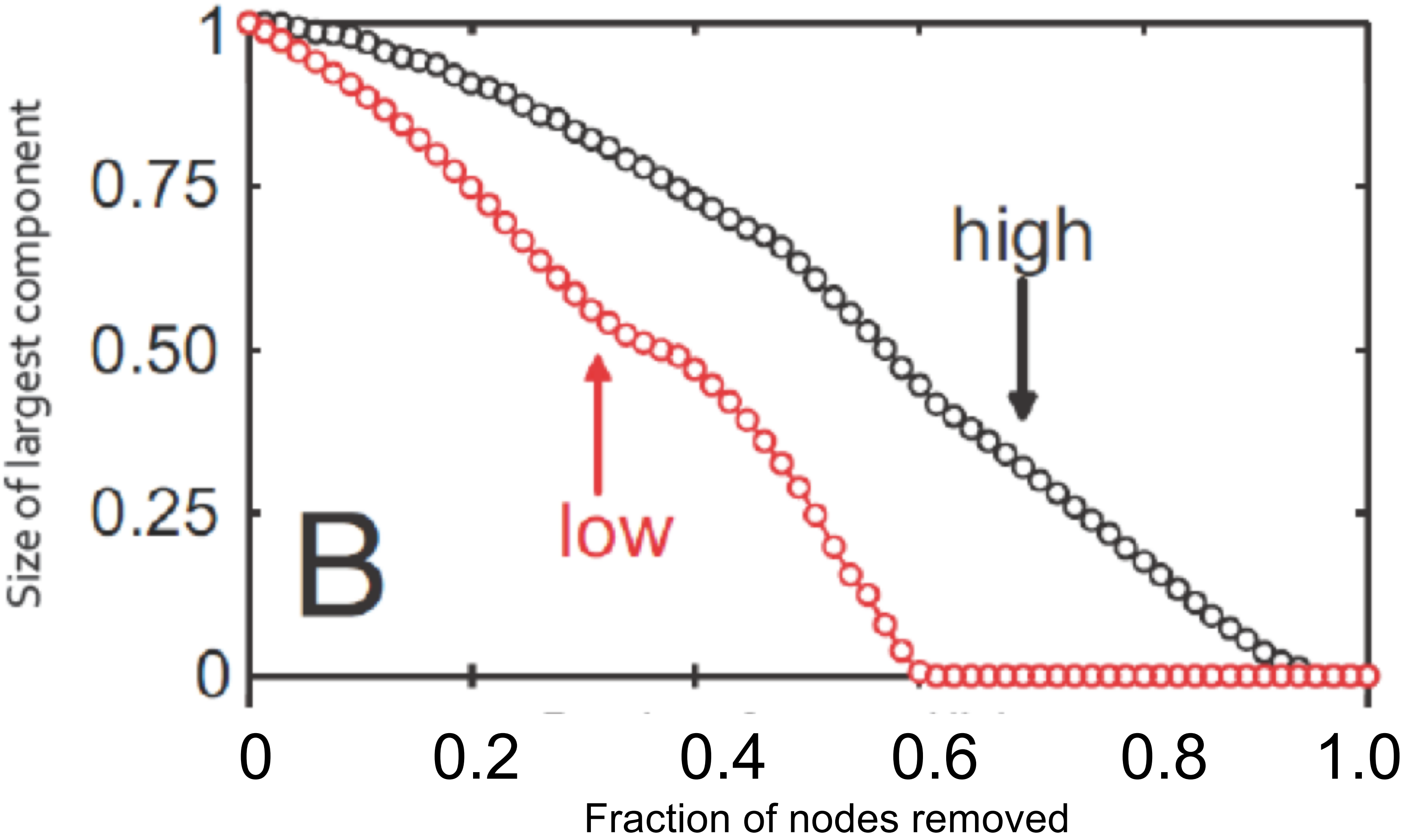


Tie strength: cumulative tie strength smaller than w

EFFECT OF EDGE REMOVAL
(BASED ON EDGE WEIGHT)



EFFECT OF EDGE REMOVAL
(BASED ON OVERLAP)



WEAK TIES MATTER!

- We have just seen that weak ties matter and if they are removed, they lead to a breakdown in the network.

- If strong ties are removed they lead to a smooth degrading of the network

DIFFERENCE OF IMPORTANCE OF WEAK TIES IN SOCIAL AND OTHER NETWORKS

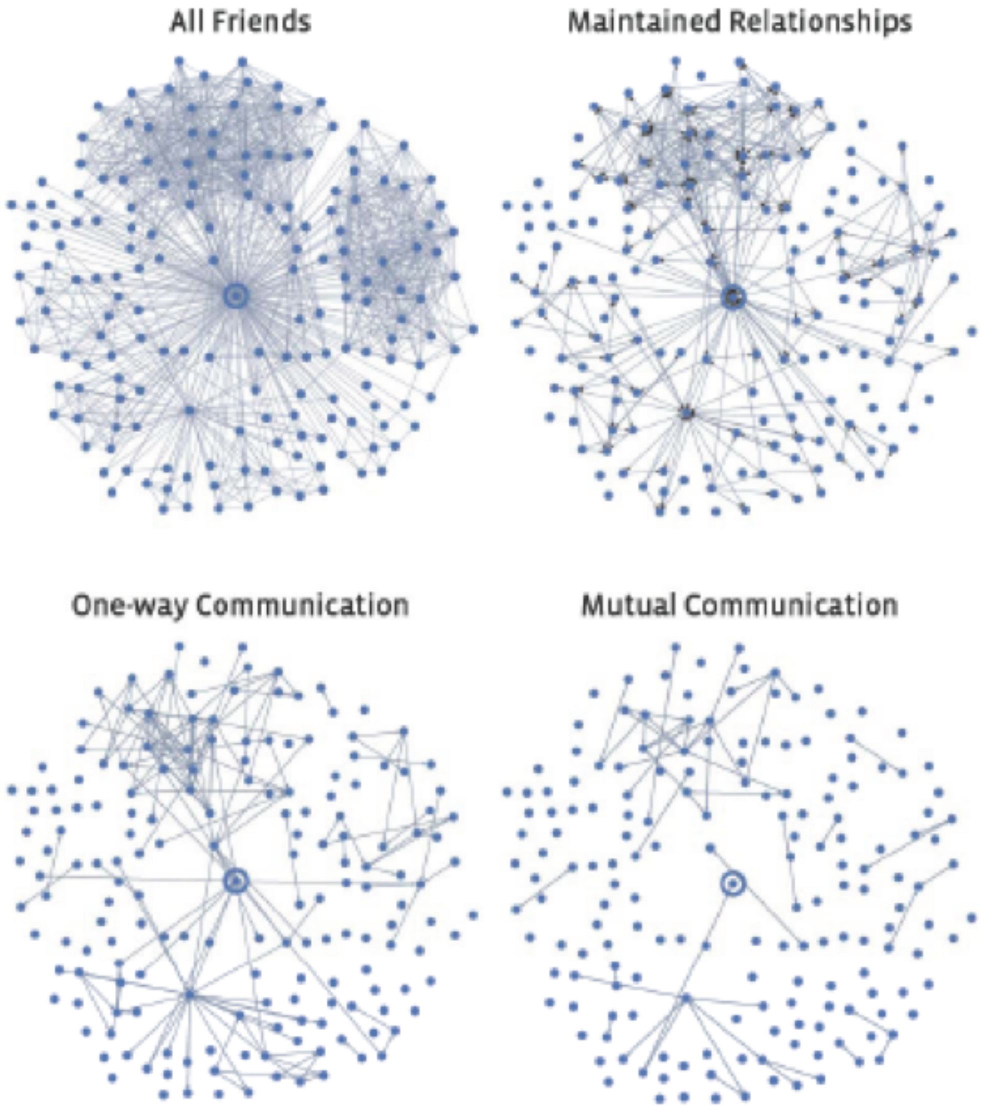
- The importance of weak ties is specific to social networks
- In biological and spatial networks:
  - Deleting an important road [strong tie] damages the network more
  - A central vein in a leaf is more important than smaller veins

TIE STRENGTH MATTERS: FACEBOOK EXAMPLE

- Facebook data analysis of one month of data
- Four networks:
  - Declared friendship
  - Reciprocal communication (messages)
  - One way communication
  - Maintained relationship: clicking on content on news feed from other friend or visiting profile more than once.

WHAT DOES IT LOOK LIKE? (ONE RANDOM USER)

- Couple friendship groups
- Plenty of stragglers

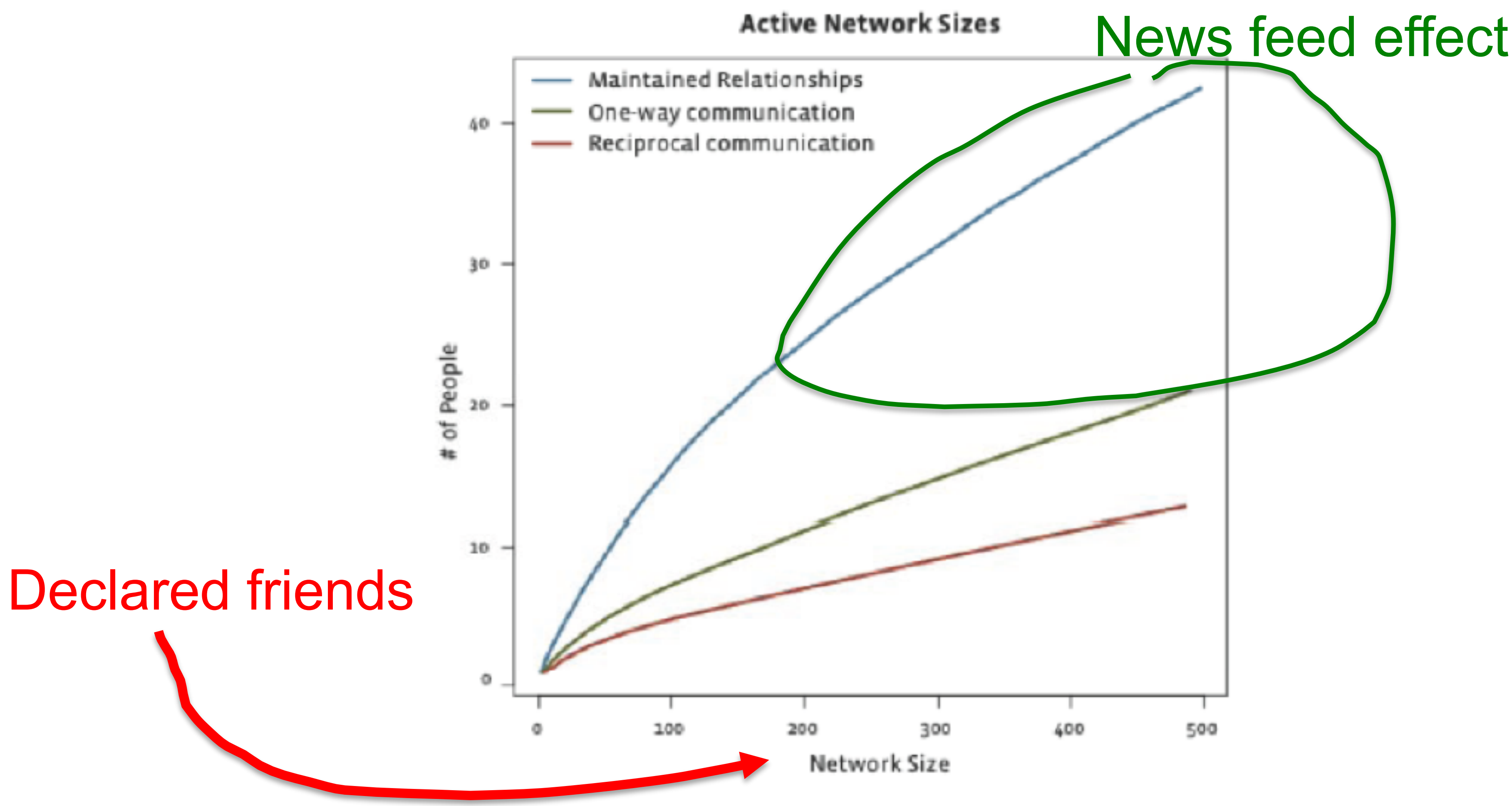


- Not bad
- Stragglers become less important

- Only talks to (at) a few

- Even smaller mutua communications

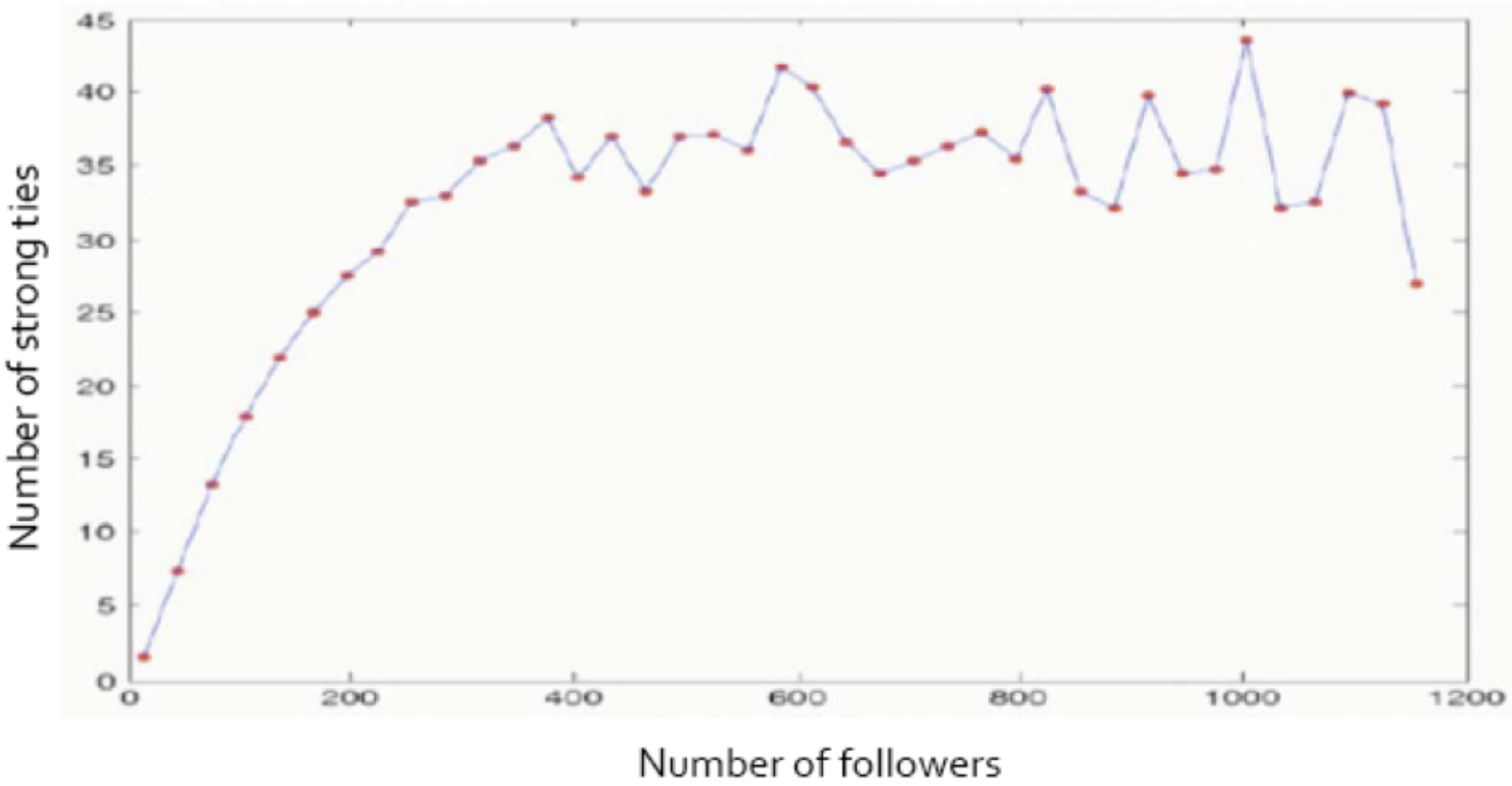
ACTIVE NETWORK SIZE: NUMBER OF LINKS



TWITTER ANALYSIS

- Huberman at al. have analyzed strong and weak ties in Twitter.
- The “followers” graph in Twitter is directed
  - Someone can follow someone else who does not follow him
- Messages of 140 chars can be posted
- Messages can be addressed to specific users (although they stay readable to all)
- **Weak ties**: users followed
- **Strong ties**: users to whom the user sent at least 2 messages in the observation period

TWITTER



Number of strong ties stays below ~50

SUMMARY

Small world network models are able to capture a good quantity of real networks

- They have characteristic path length comparable to random networks.
- But much higher clustering coefficient.

We have introduced weak and strong ties and shown example of application on real networks

References

- Kleinberg's book: Chapter 3 and 20.

- **Collective dynamics of 'small-world' networks.** Watts, D.J.; Strogatz, S.H. (1998). *Nature* 393 (6684): 409–10.
- **Structure and tie strengths in mobile communication networks.** J. P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, A. L. Barabasi. *Proceedings of the National Academy of Sciences*, Vol. 104, No. 18. (13 Oct 2006), pp. 7332-7336.
- **Maintained relationships on facebook.** Cameron Marlow, Lee Byron, Tom Lento, and Itamar Rosenn. 2009. On-line at https://www.facebook.com/notes/facebook-data-science/maintained-relationships-on-facebook/55257228858/.
- **Social networks that matter: Twitter under the microscope.** Bernardo A. Huberman, Daniel M. Romero, and Fang Wu. First Monday, 14(1), January 2009.