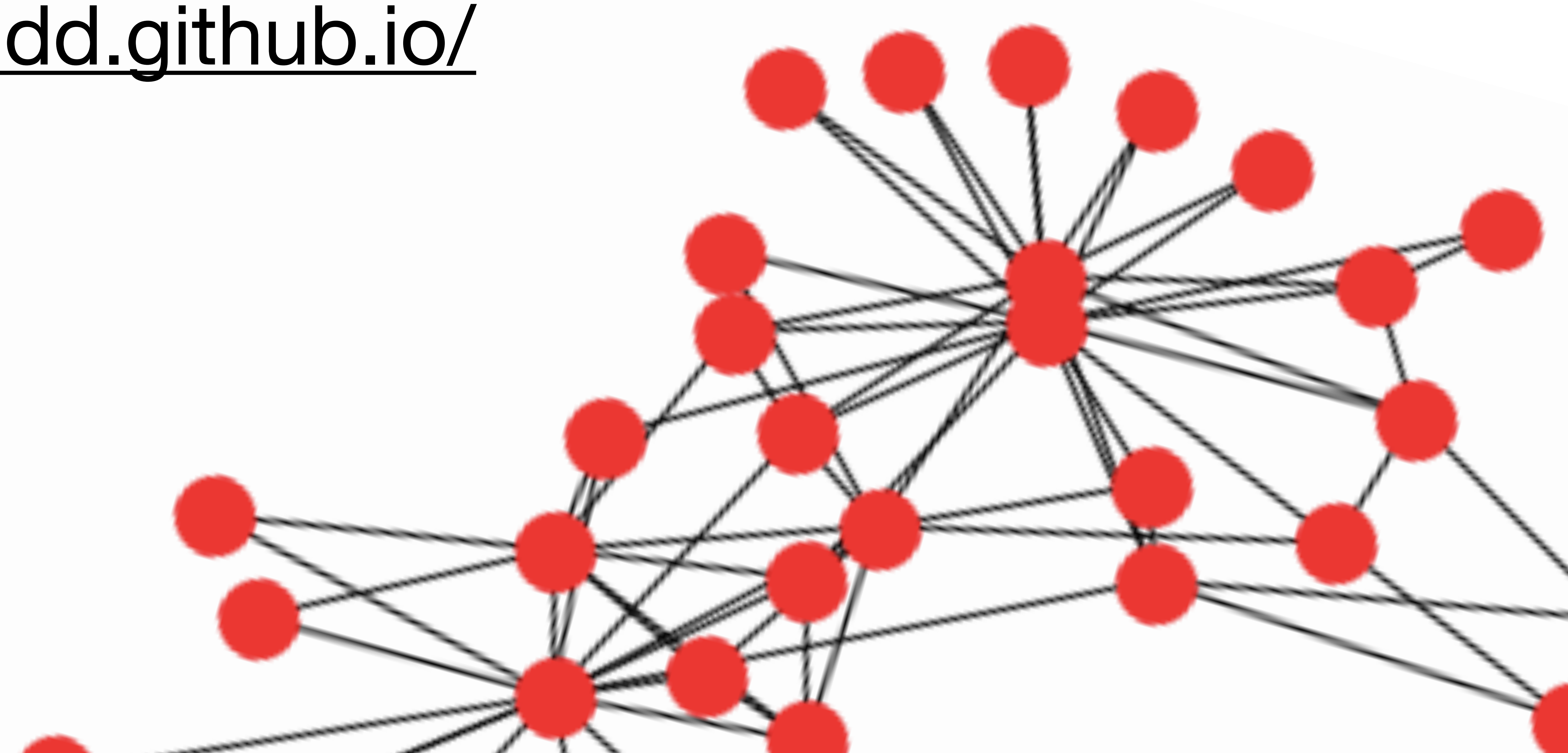
DMSN Tutorial 2: Small Worlds and Weak Ties

Naomi Arnold
https://narnolddd.github.io/

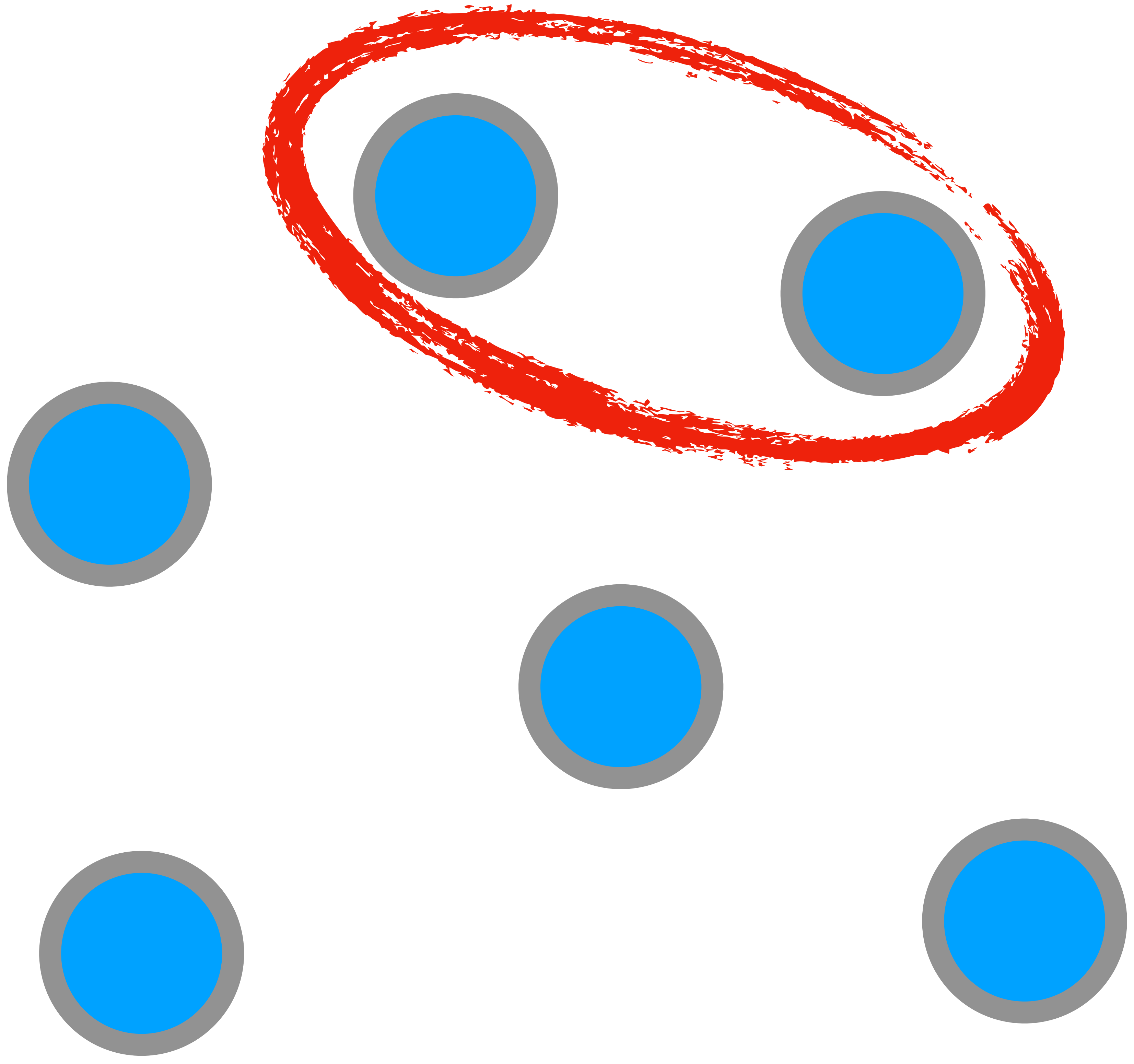


In this tutorial:

- **Recap** on real networks vs random graphs
- **Experiment with** Watts-Strogatz model
- **Understand** the role that weak ties play in networks

Real vs Random Networks

Erdos-Renyi G(n,p) Model



1. Start with an empty graph of **n** nodes

2. “Coin” with head probability **p**

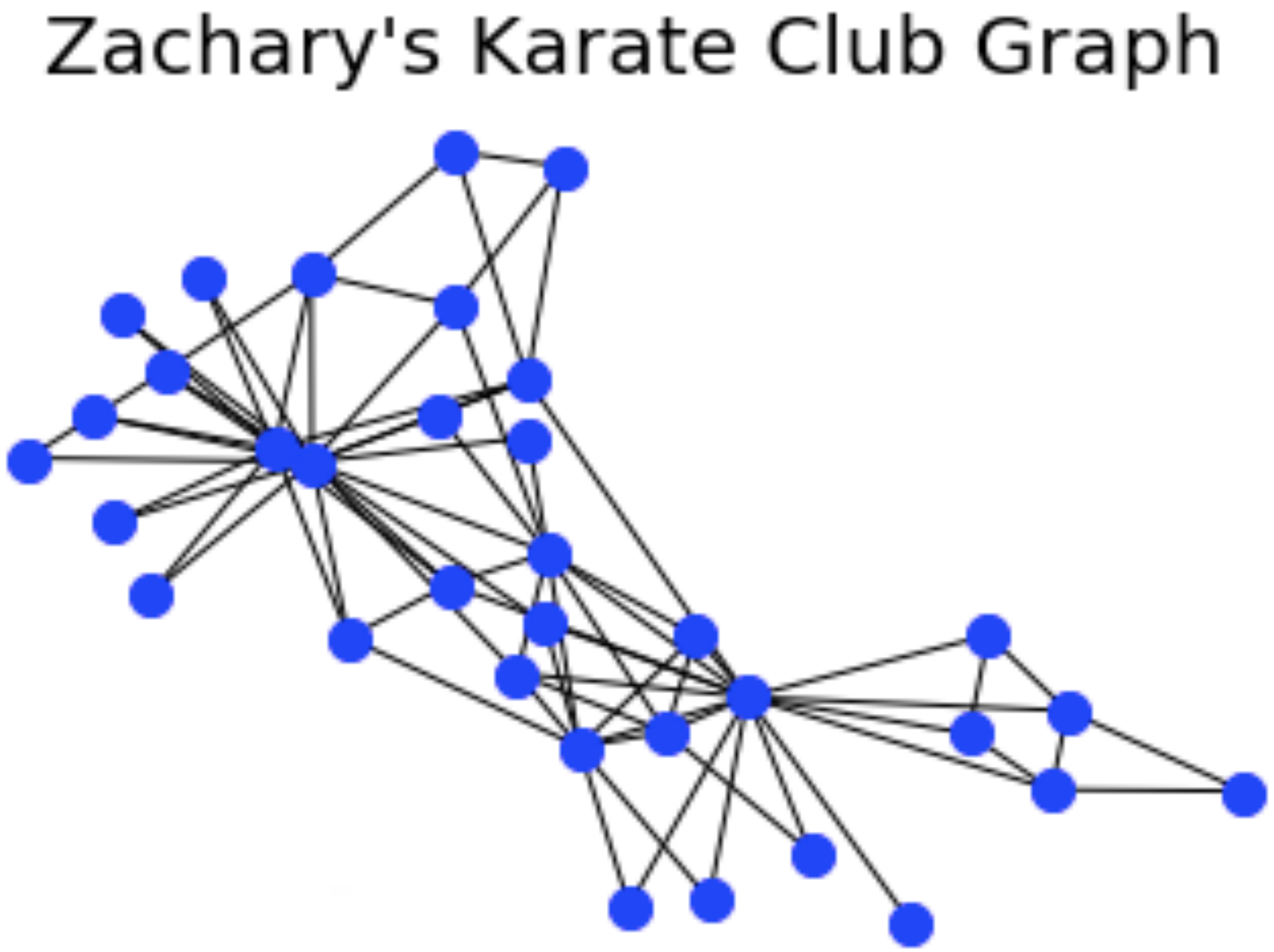
3. For each pair of nodes, do a coin toss. If heads, draw an edge between them. If not, move on.

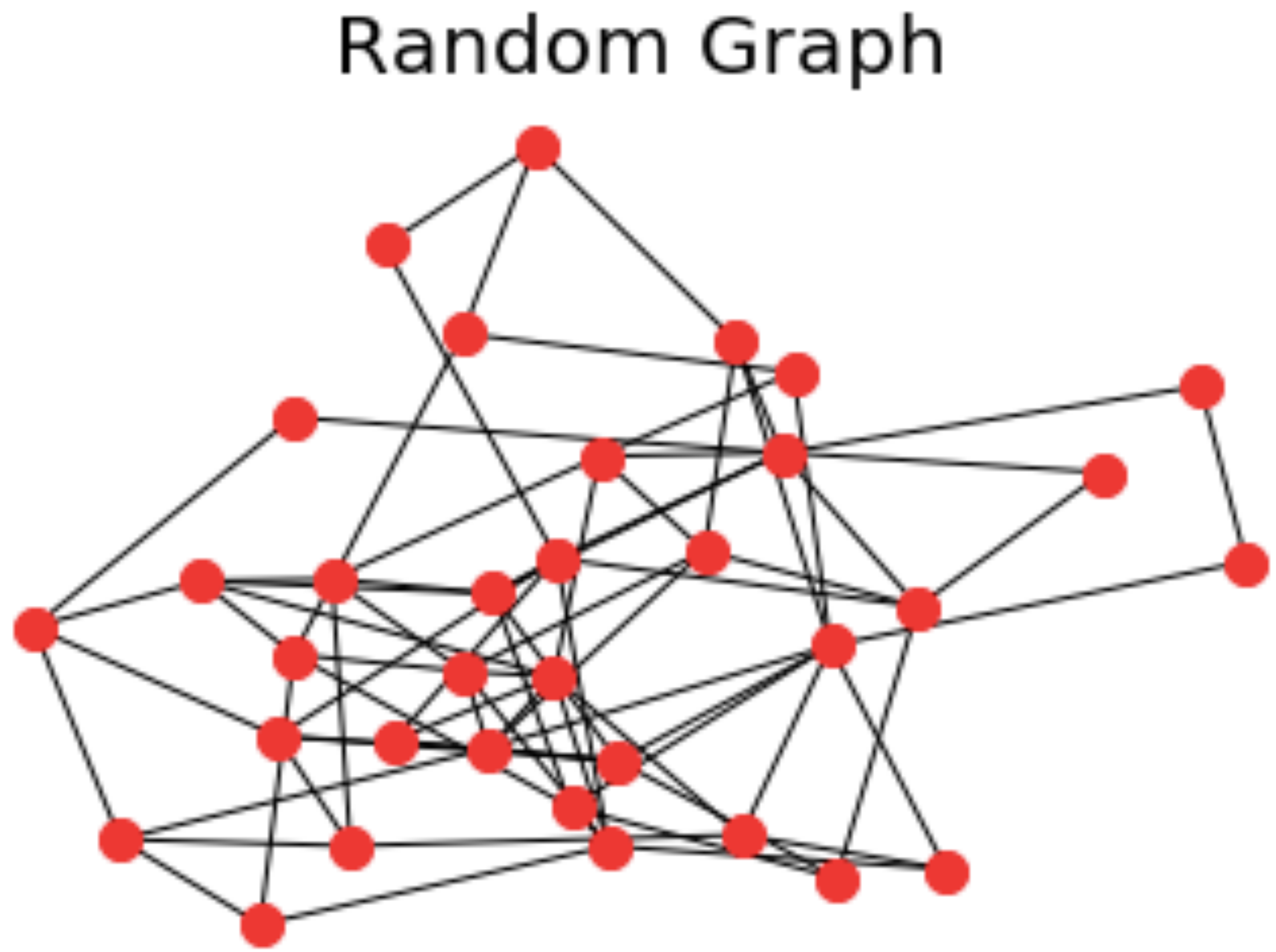
Erdos-Renyi G(n,p) model

$$p < \frac{1}{n} \qquad p = \frac{1}{n} + \epsilon \qquad p > \frac{\log(n)}{n}$$

Increasing **p**

Random Graphs vs Real Networks





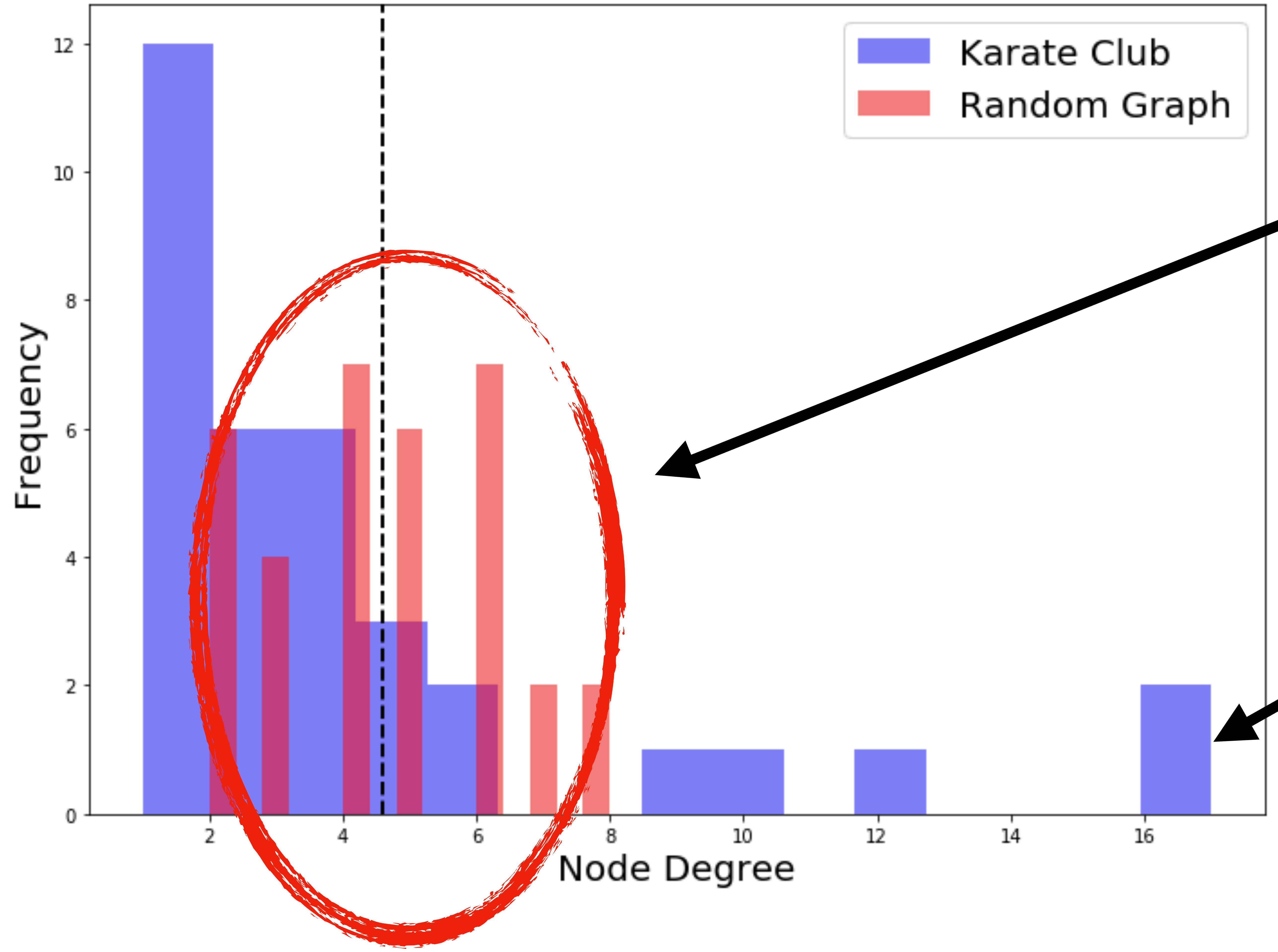
Apparent community structure

No community structure. “Blob”

Some high degree ‘hubs’

Nodes of similar degree

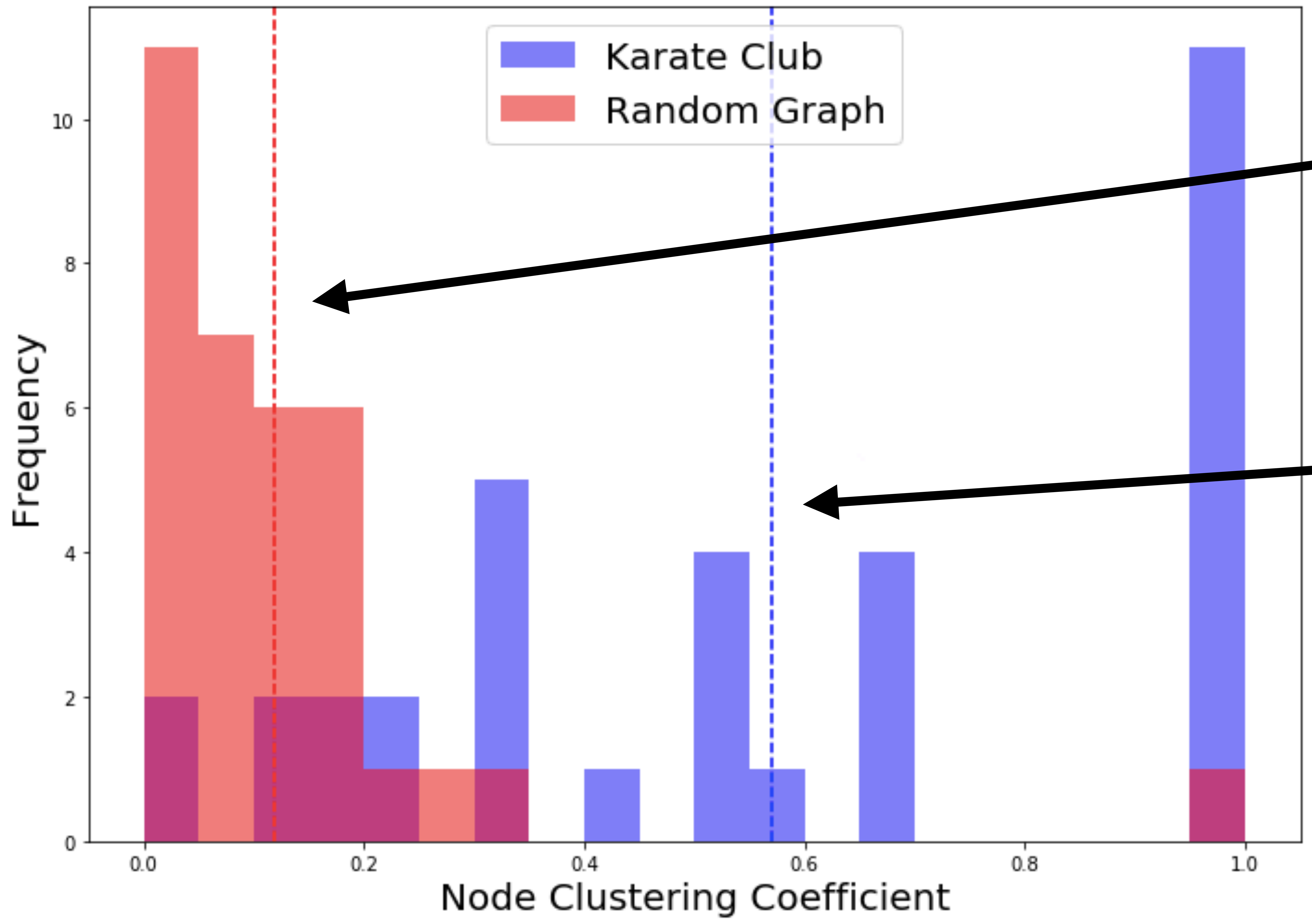
Random Graphs vs Real Networks: Degree



Random: node degrees all clustered round the average value

Real: "heavy tailed" — small number of high degree nodes, large number of low degree nodes

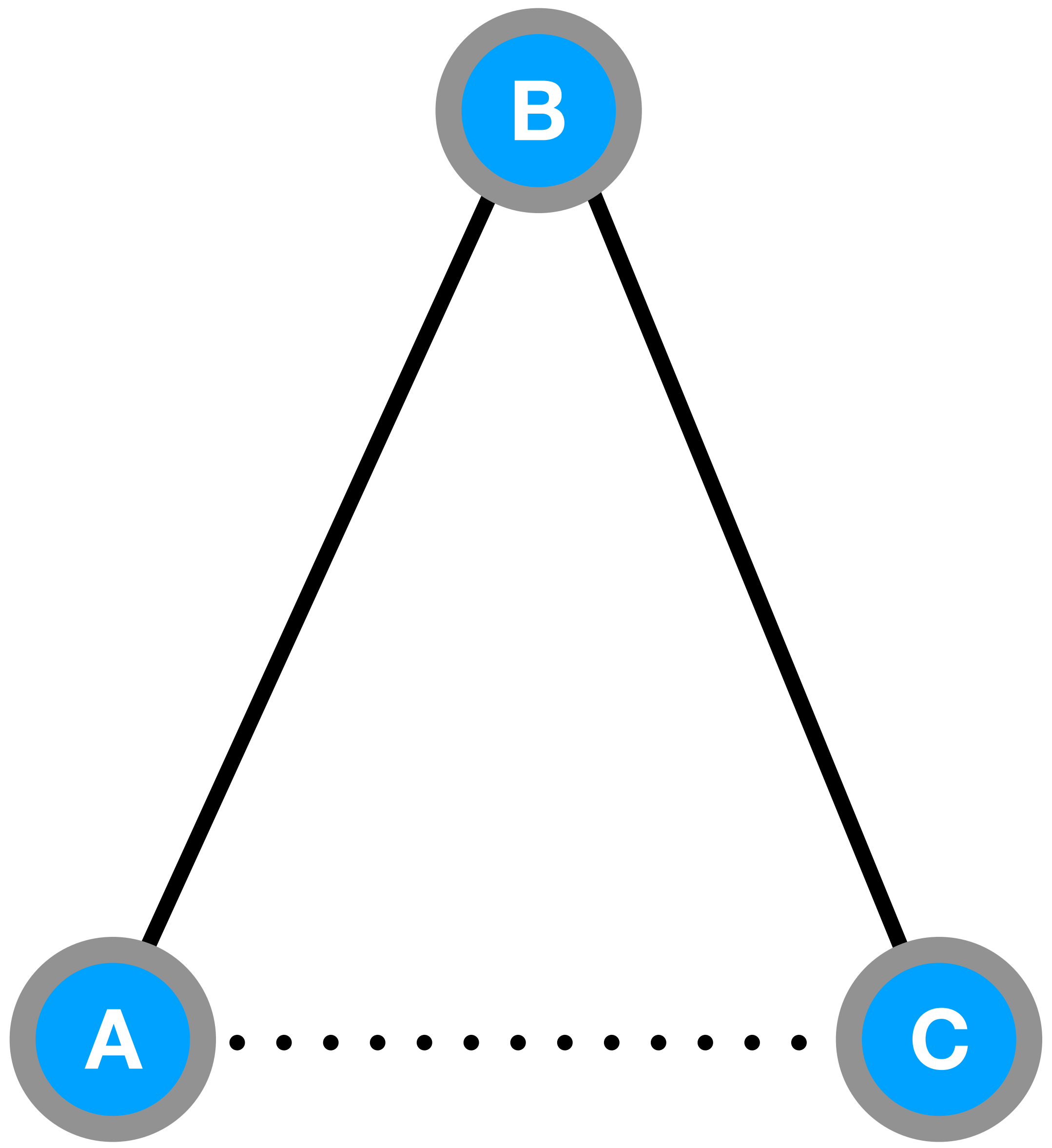
Random Graphs vs Real Networks: Clustering



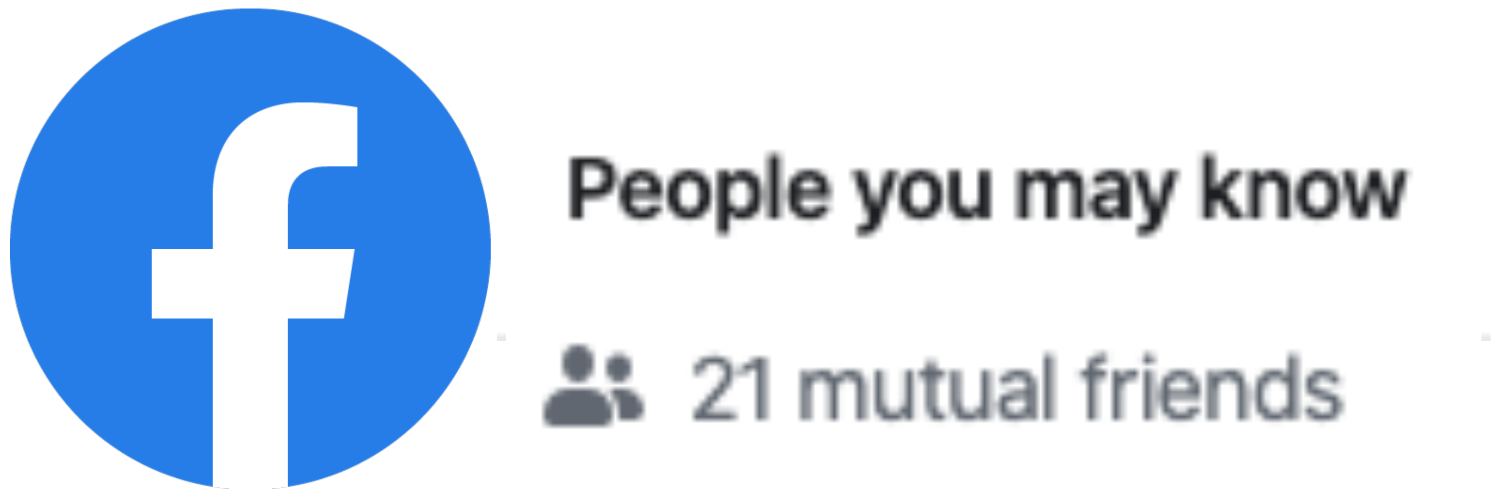
Random: very low average clustering coefficent

Real: much higher average clustering coefficient, with some nodes having very high values

Why is the clustering so high?

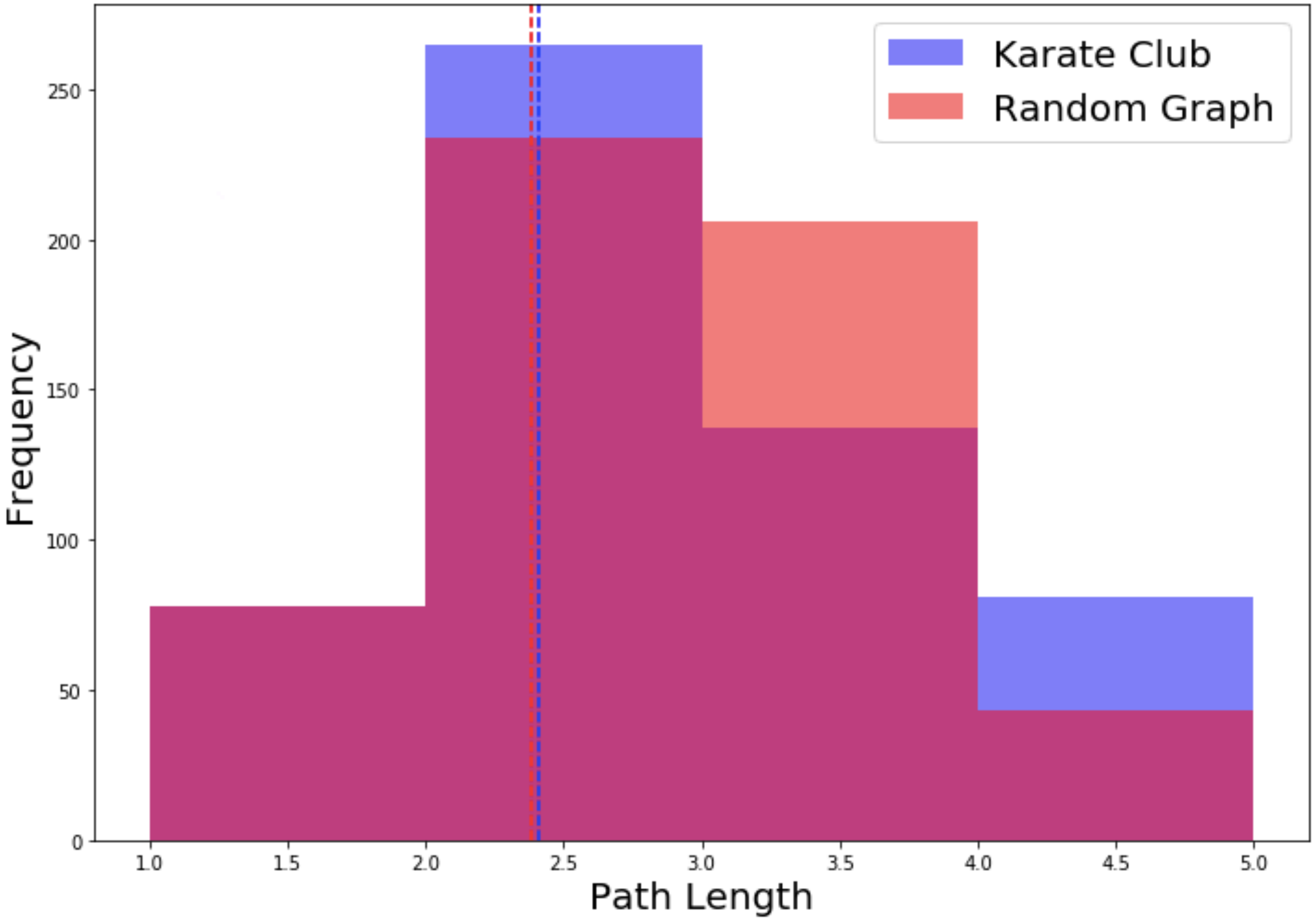


In **real life**, we meet many friends through mutuals



Online, this is "baked in" by friend/follow recommendation algorithms

Random Graphs vs Real Networks: Paths



Fairly spot on with **almost the same** average path length for each!

Are short path lengths unusual?

- If everyone in the world had 100 friends:
- My number of friends would be 100
- … friends of friends could be 100 x 100 = 10,000
- … friends of friends of friends could be 100 x 100 x 100 = 1,000,000
- With only 3 hops, already can reach 1 million people

Short path lengths can be a good thing

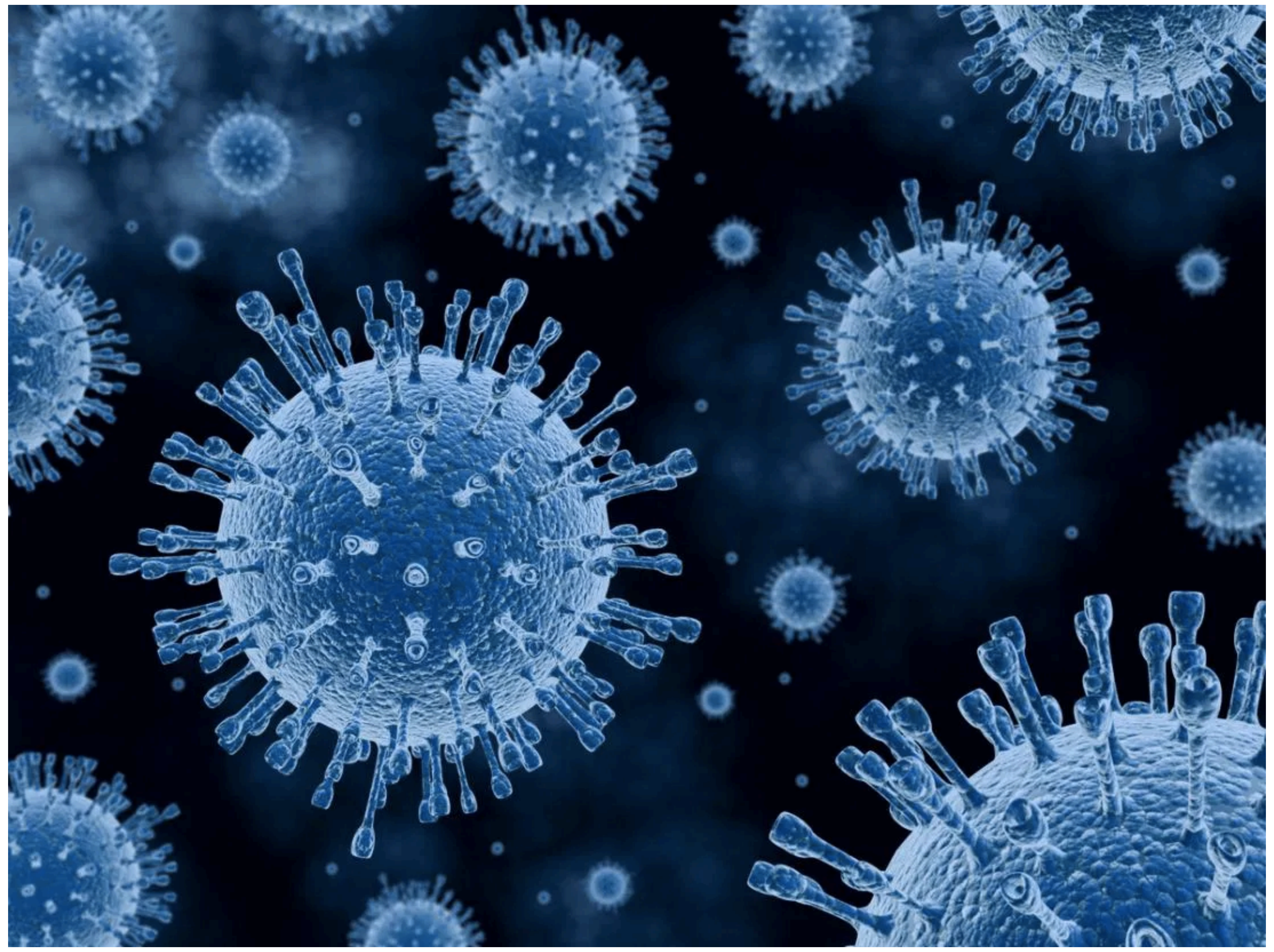
NETFLIX

Quick, efficient **distribution** of content

Discovering/spreading important **information**

Quick* **travel** across airport network

Short path lengths can be a bad thing



Epidemics can potentially spread very far very quickly

Fake news or misinformation can quickly be propagated

Other real-world networks

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

Random: very good at path lengths

But bad at clustering!

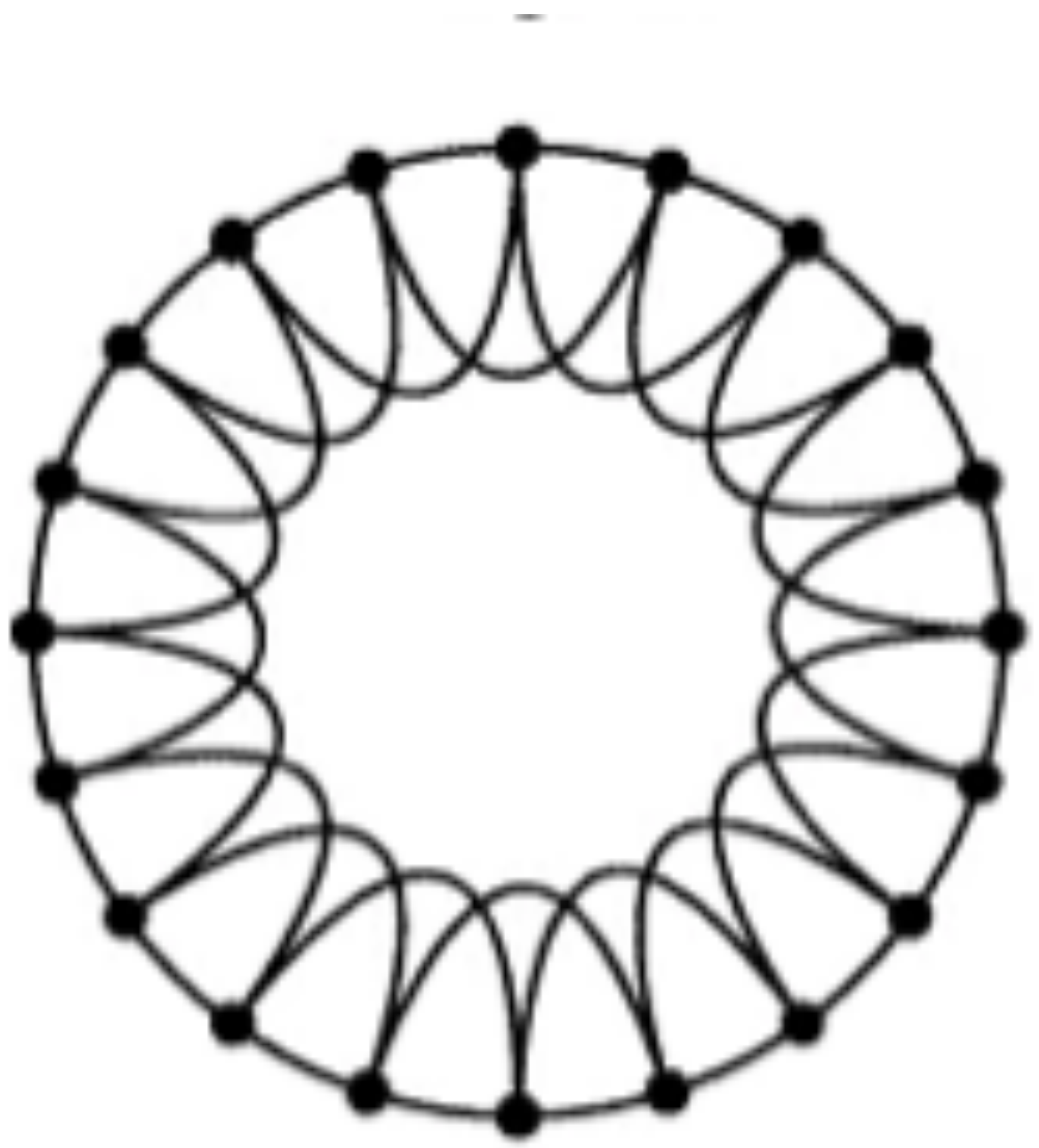
Summary: Random Graphs vs Real Networks

	Real Social Networks	Random Graphs	?
Degree Distribution	**Heavy Tailed** (most nodes have low degree, few with high degree)	**Light tailed** (all nodes have close to the average degree)	?
Clustering Coefficient	**High**	**Low**	?
Path Lengths	**Low**	**Low**	?
Communities?	**Yes**	**No**	?

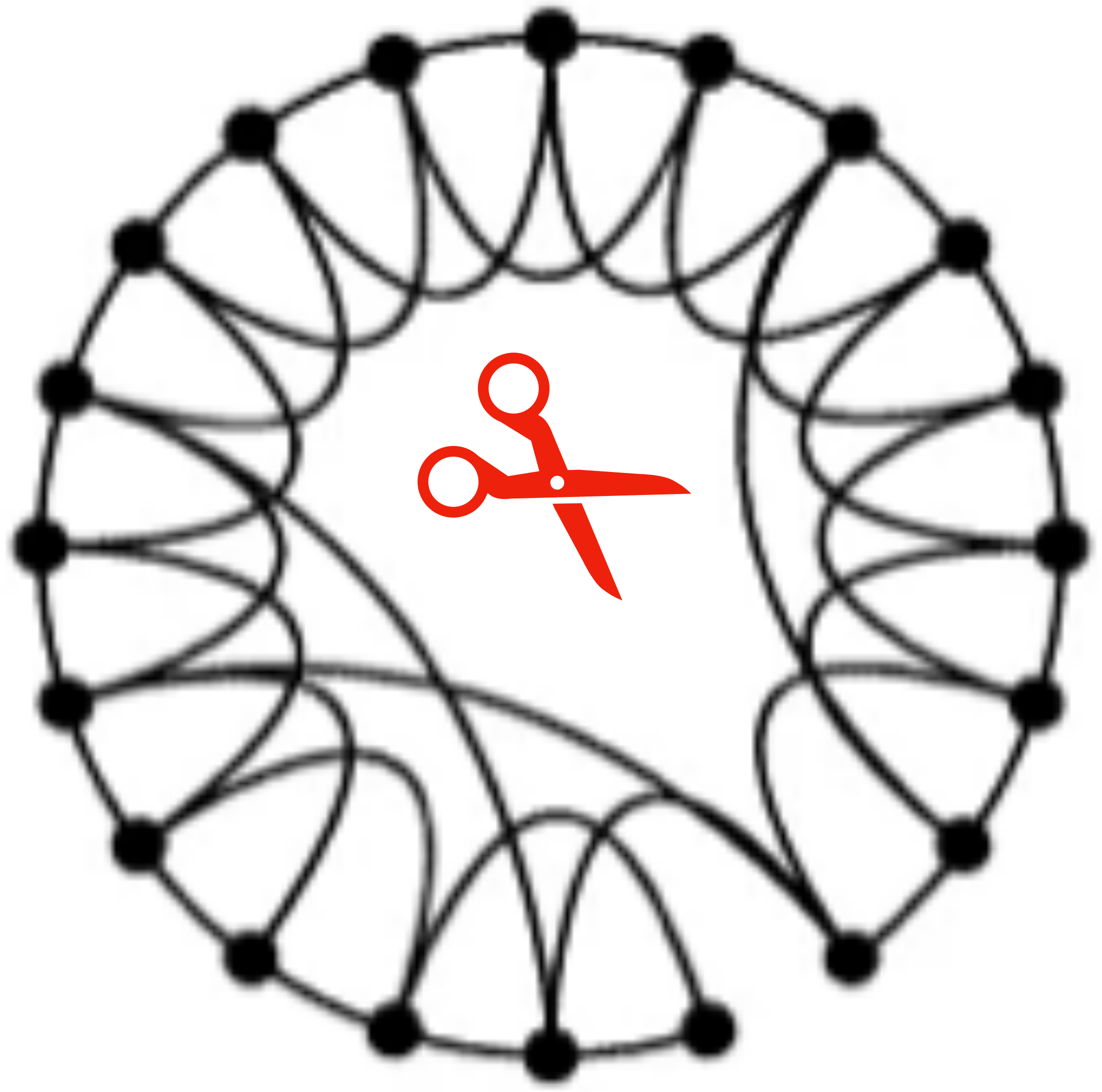
Questions so far?

Watts and Strogatz: “Can we keep the short path lengths but have higher clustering?”

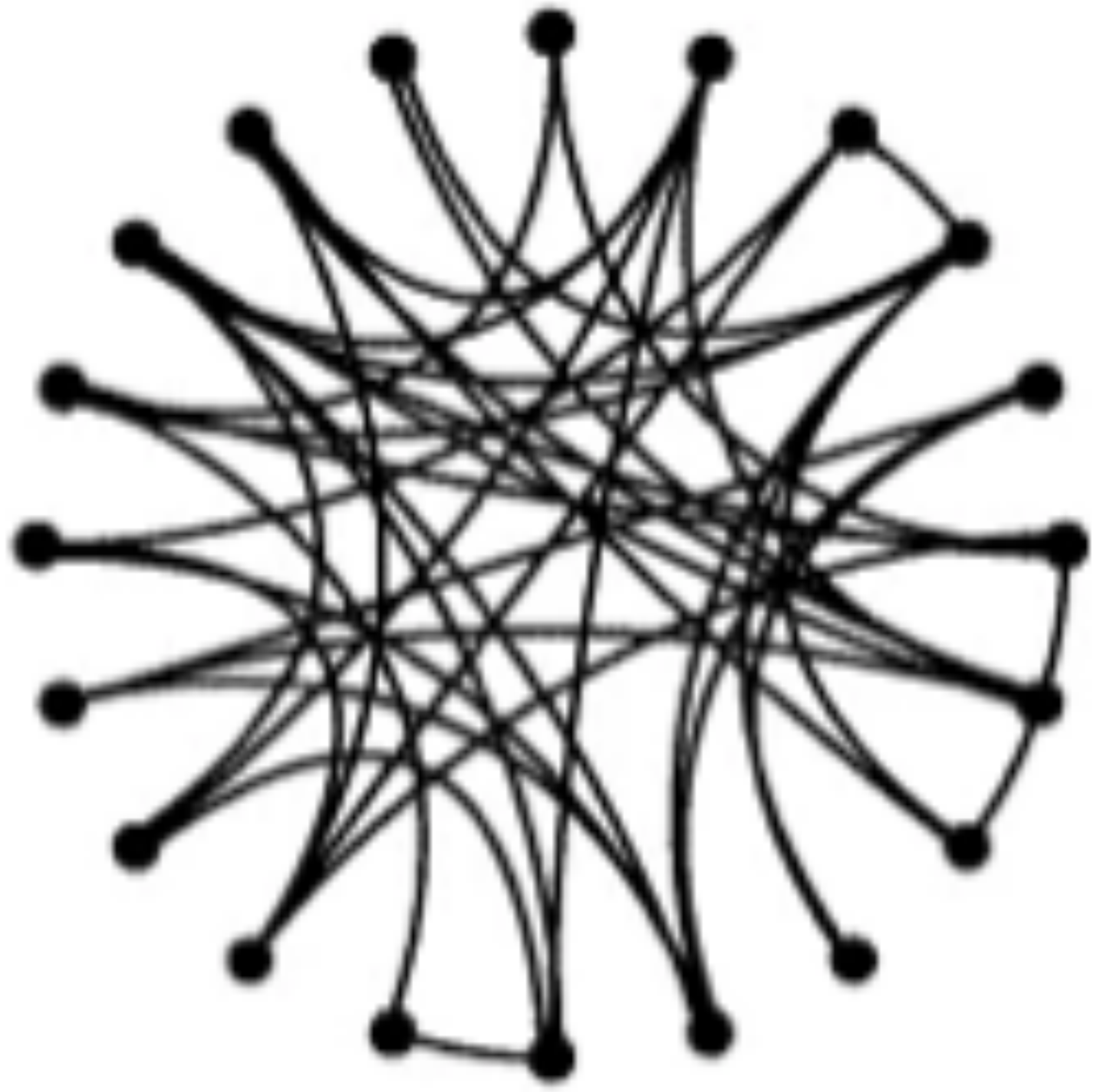
The model



Start with a **ring graph** where each node is connected to the **k** nodes closest to it. This has a **high clustering** coefficient.

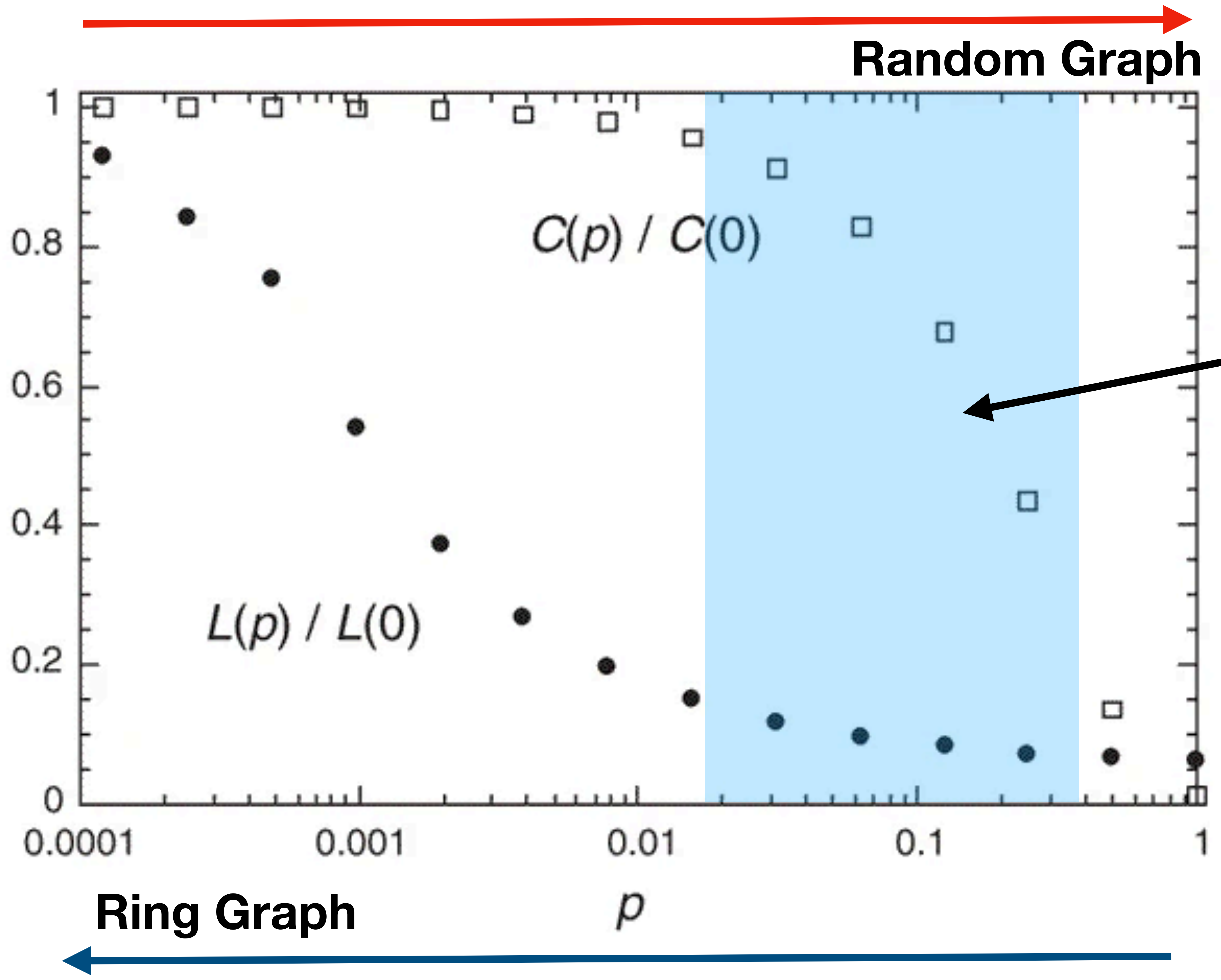


For each node and attached edge, with probability **p**, reconnect it to a randomly chosen node, otherwise leave alone.



When **p** is very high, this looks like a **random graph** again

Finding the happy medium



Zone where we have both **high clustering** and **low average path length**

“Goldilocks zone”

Graph models so far

	Real Social Networks	Random Graphs	Watts-Strogatz
Degree Distribution	**Heavy Tailed** (most nodes have low degree, few with high degree)	**Light tailed** (all nodes have close to the average degree)	**Light tailed** (all nodes have close to the average degree)
Clustering Coefficient	**High**	**Low**	**High**
Path Lengths	**Low**	**Low**	**Low**
Communities?	**Yes**	**No**	**No**

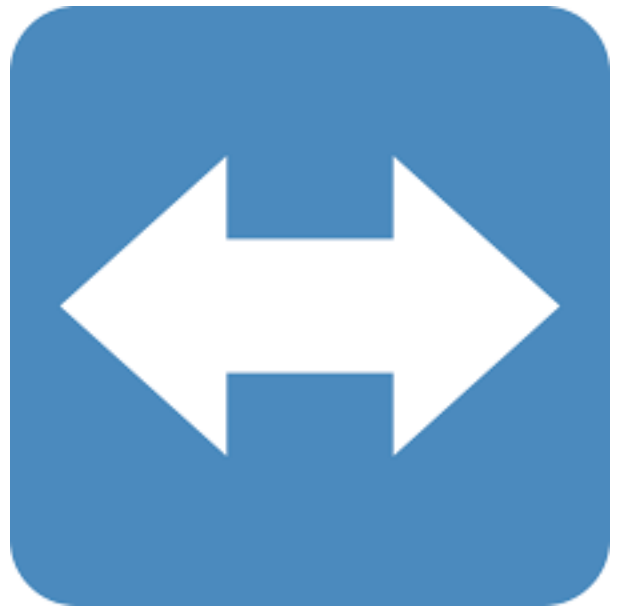
Tie Strength and Weak Ties

Tie strength

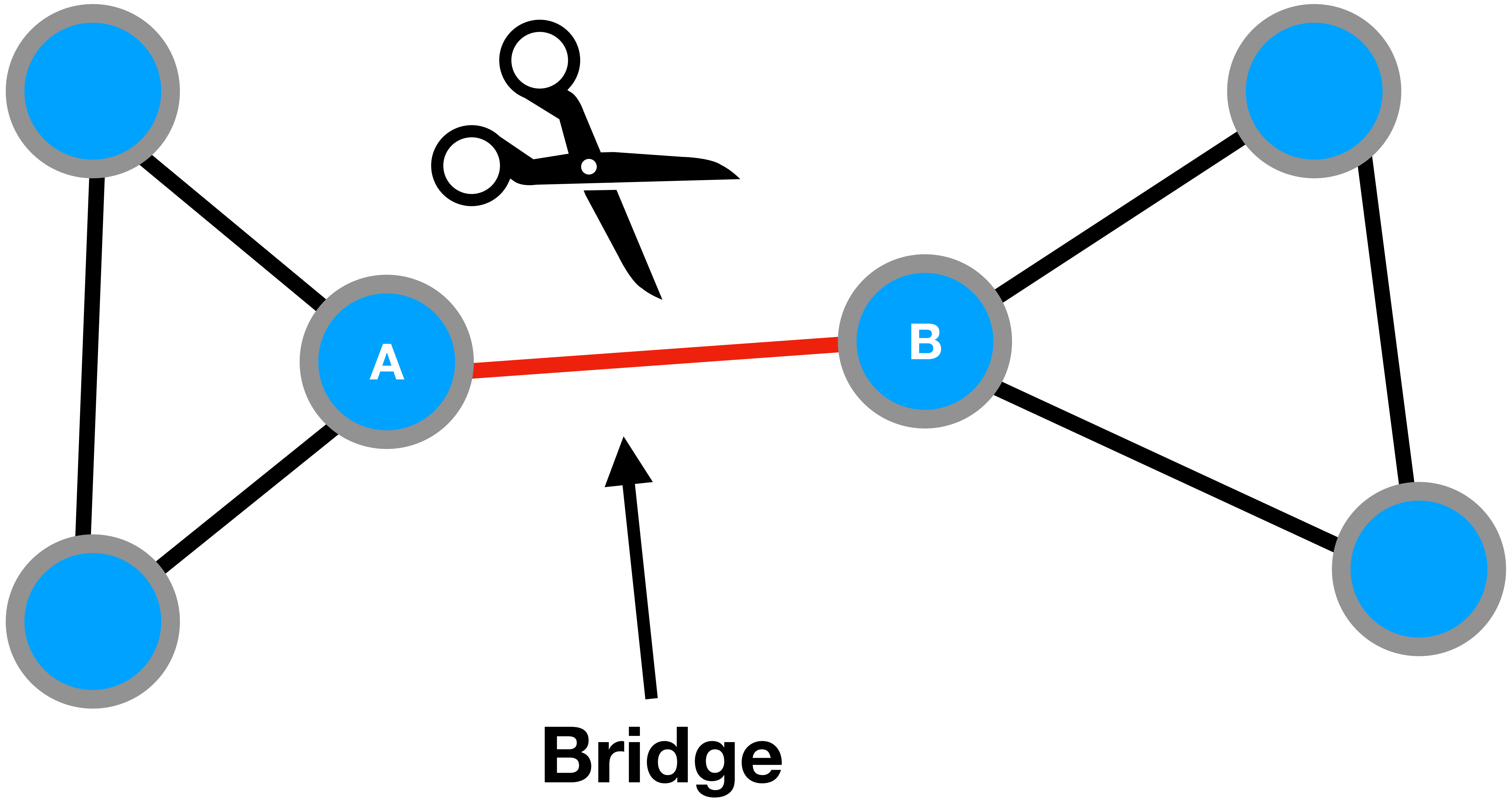
"combination of the ***amount of time****, the* ***emotional intensity****, the* ***intimacy*** *(mutual confiding) and* ***reciprocal services*** *which characterize the tie"*

Granovetter, 1973



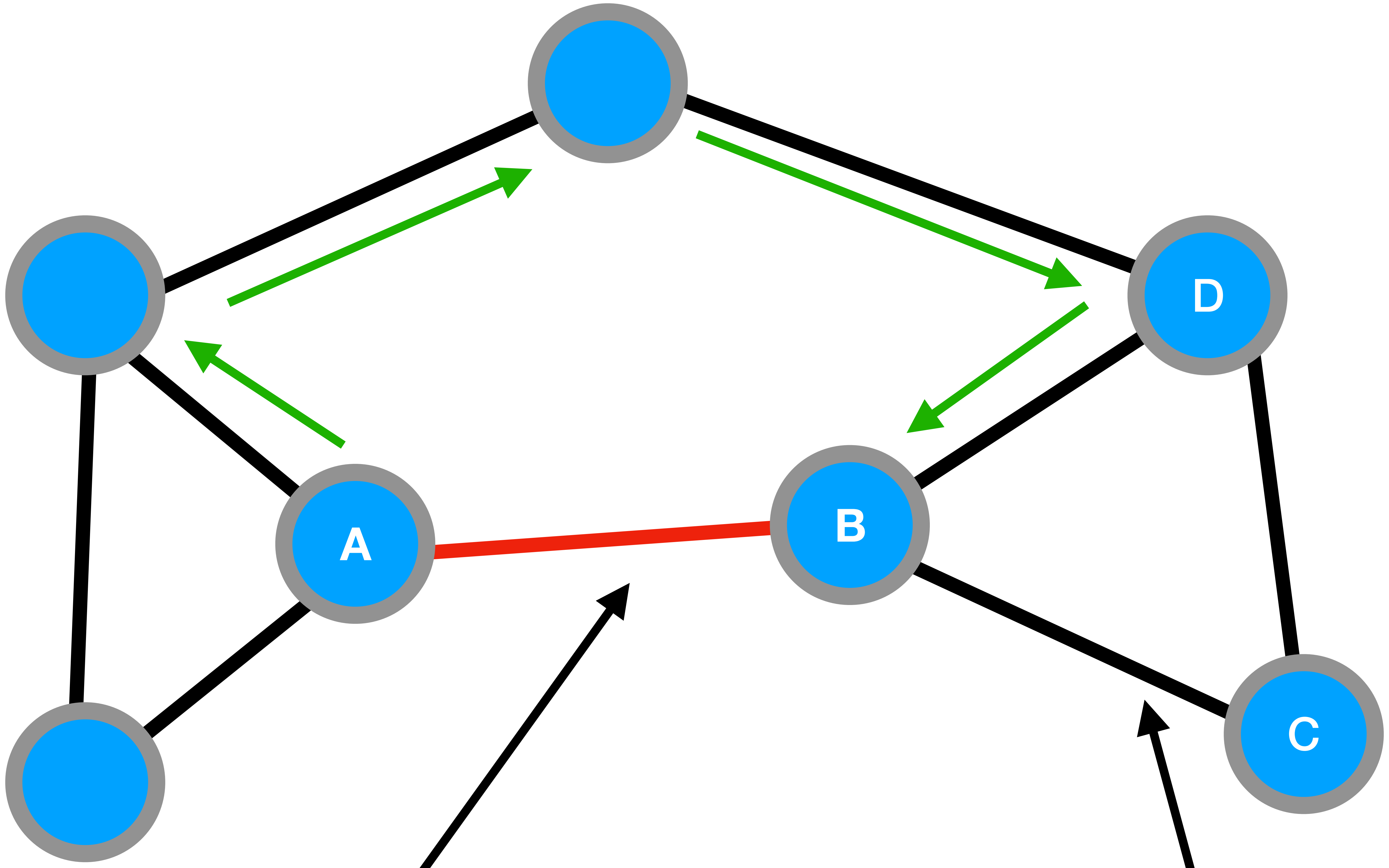
Weak ties: Bridges



One example of a weak tie is a **bridge**. A bridge is an edge which, if removed, would **disconnect** the network.

Fairly rare in big networks, as could be catastrophic

Weak ties: Local Bridge



An edge is a **local bridge** if removing it would make the distance between its endpoints **more than 2.**

Local bridge, as
d(A,B) = 4 without.

Not a local bridge, as
d(B,C) = 2 without.

A network measure of tie strength: Neighbourhood Overlap

Given an Edge, the **overlap** is:

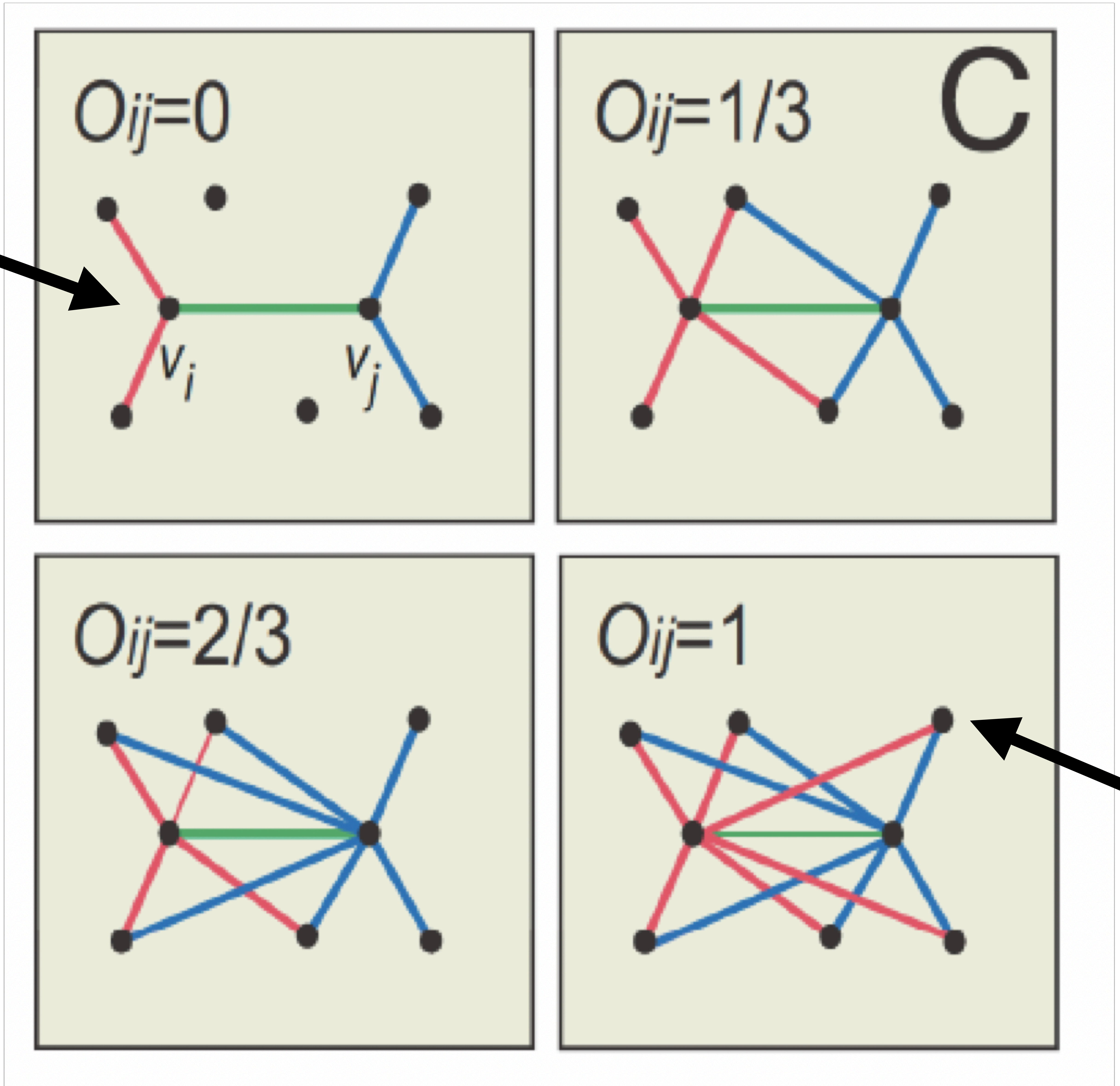
$$\frac{\text{Number of nodes who are neighbours of } \textbf{both A and B}}{\text{Number of nodes who are neighbours of } \textbf{at least A or B}}$$

$$= \frac{|N(A) \cap N(B)|}{|N(A) \cup N(B)|}$$

(If you enjoy set notation!)

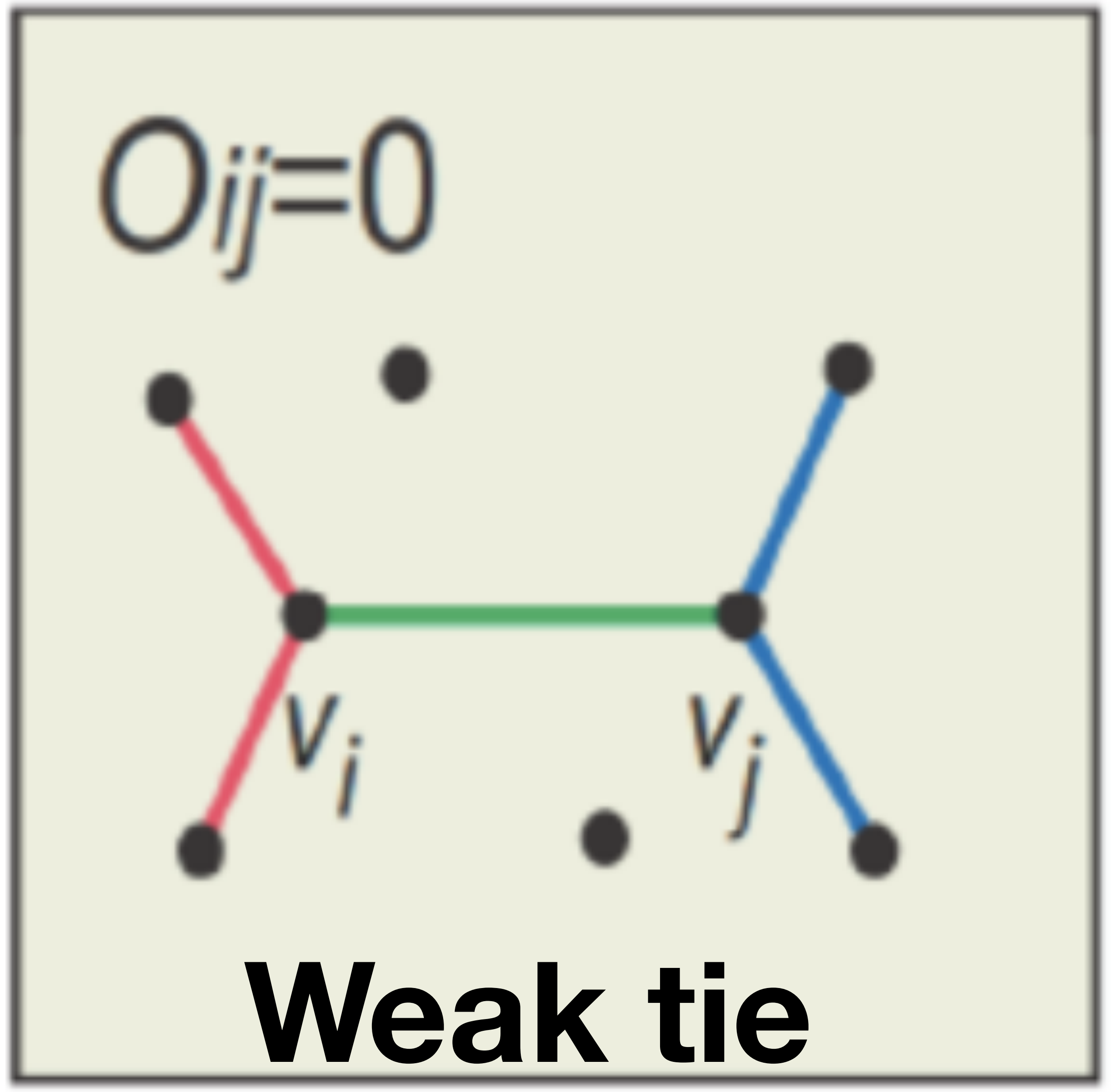
Examples

Local Bridge



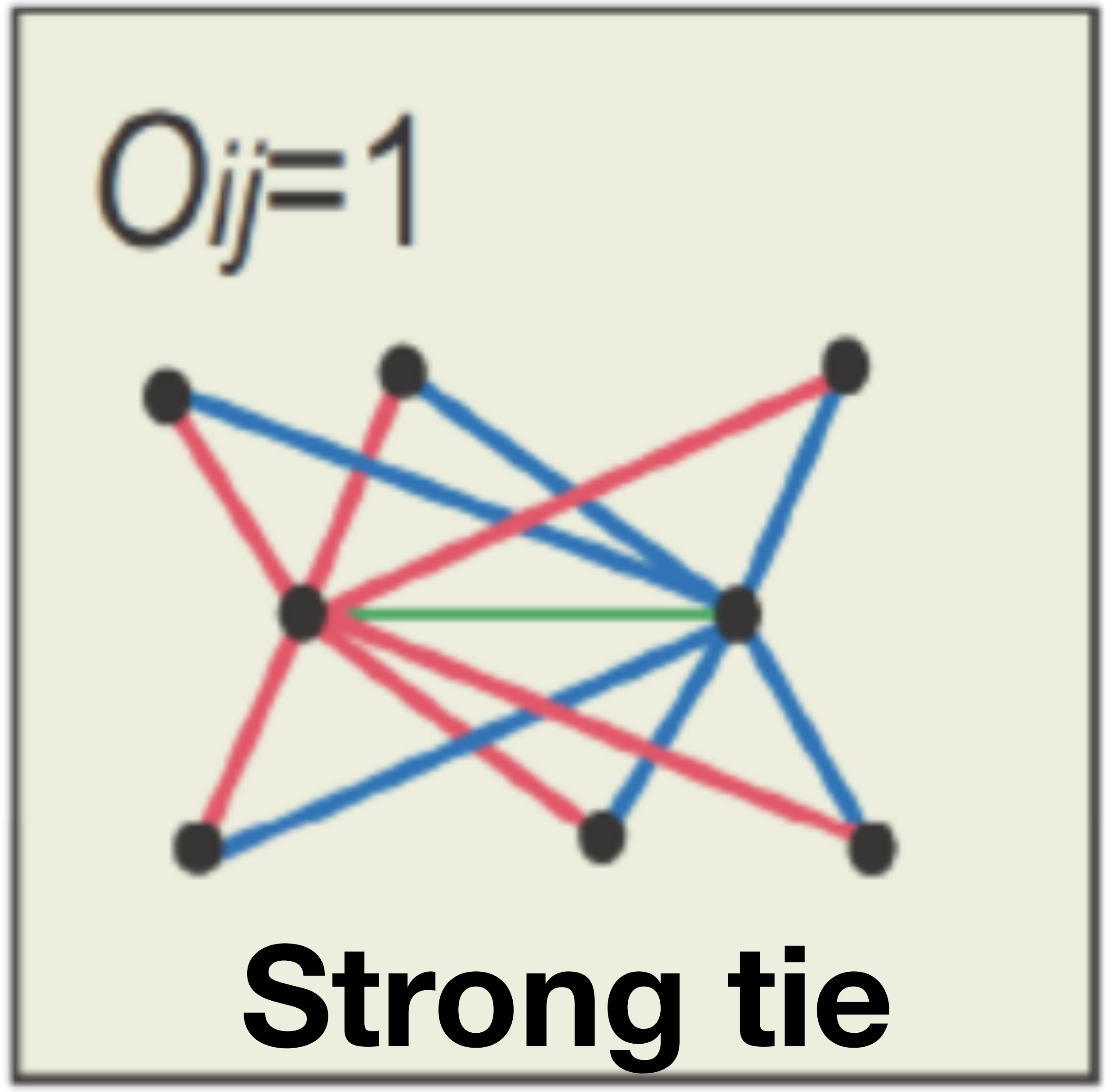
Strong tie

Significance of weak ties



May be the **only (short) path** between two communities

Important target for **epidemic prevention**

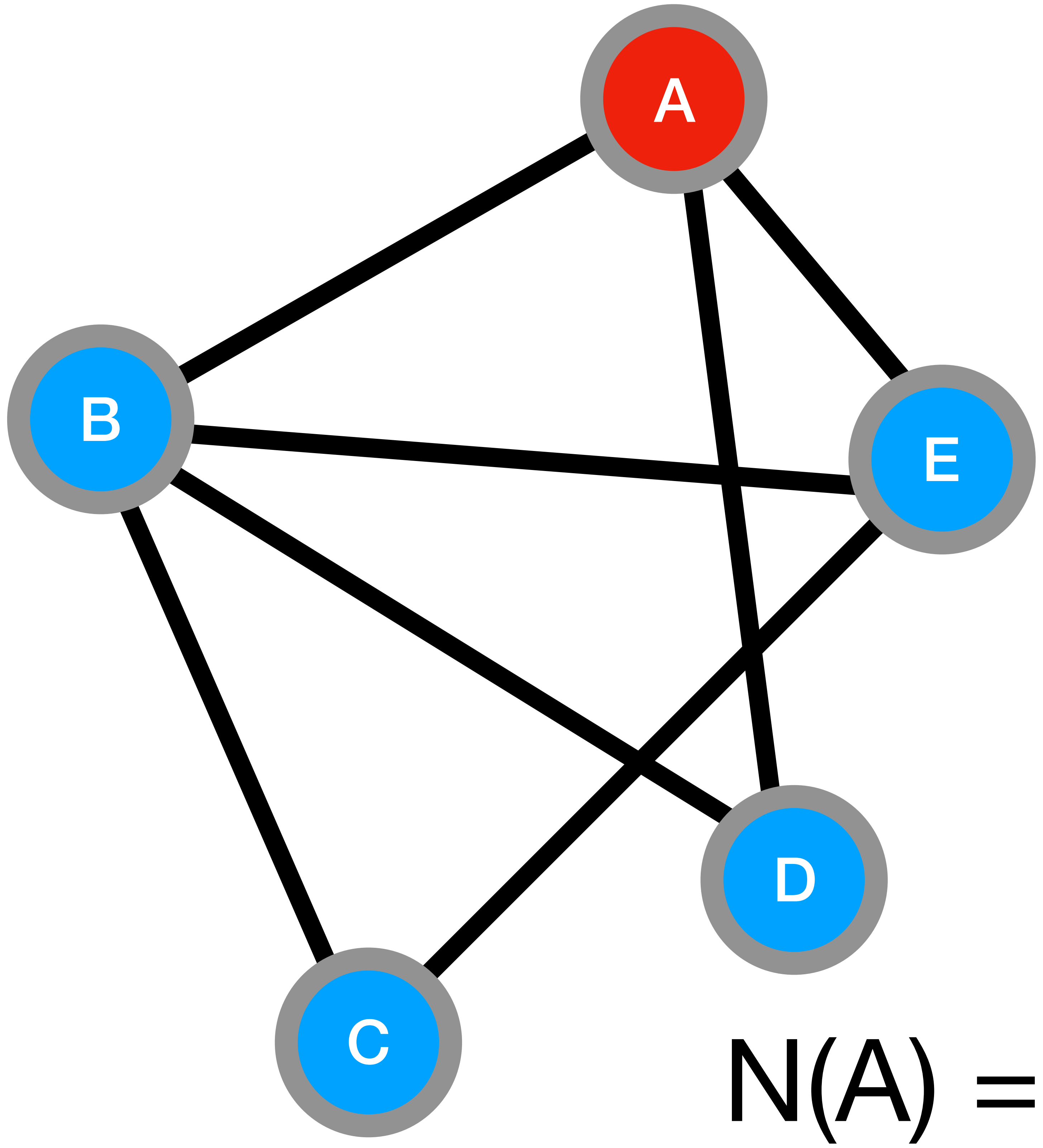


Strong ties **redundant** for information spread

Harder to **stop spread** of information/epidemic among densely connected graphs

Thanks for listening! What are your questions?

Recap: Node Clustering Coefficient



1. “Zoom in” on A’s neighbourhood and forget anything else.
2. Calculate the **bottom** of the fraction as $0.5* k(A)*(k(A) - 1)$
3. Count the **pairs of neighbours** of A that are connected

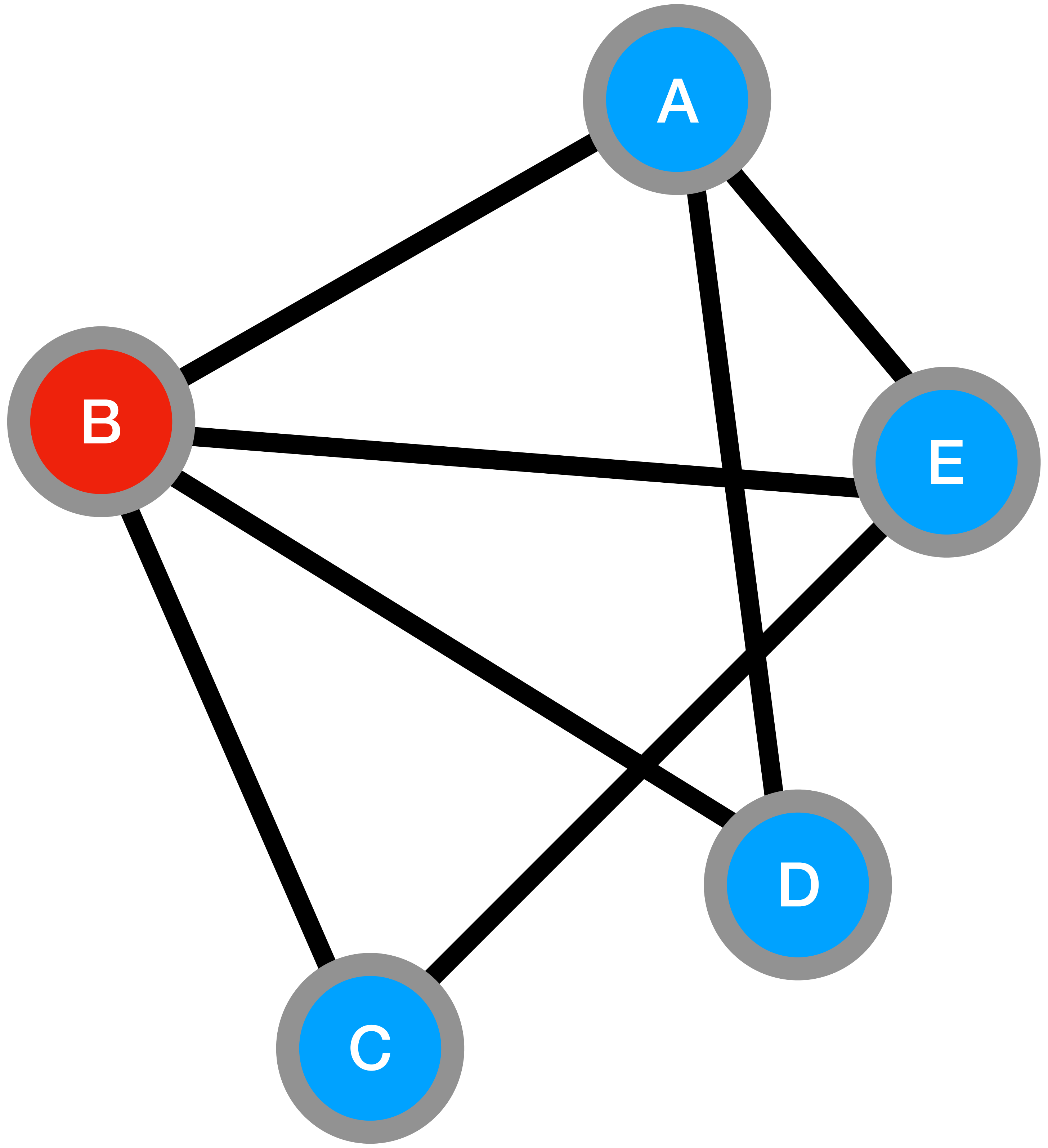
$N(A) = \{B, D, E\}$

$k(A) = 3$

$0.5*k(A)*(k(A) -1) = 0.5*3*2=3$

Pairs of connected neighbours: (B,E) , (B,D)

Recap: Node Clustering Coefficient



N(B) = {A,E,D,C}

k(B) = 4

0.5*k(B)*(k(B)-1) = 0.5*4*3 = 6

Pairs of connected neighbours of B:

(A,E), (A,D), (C,E)