

# DMSN Tutorial 1: Networks and Random Graphs

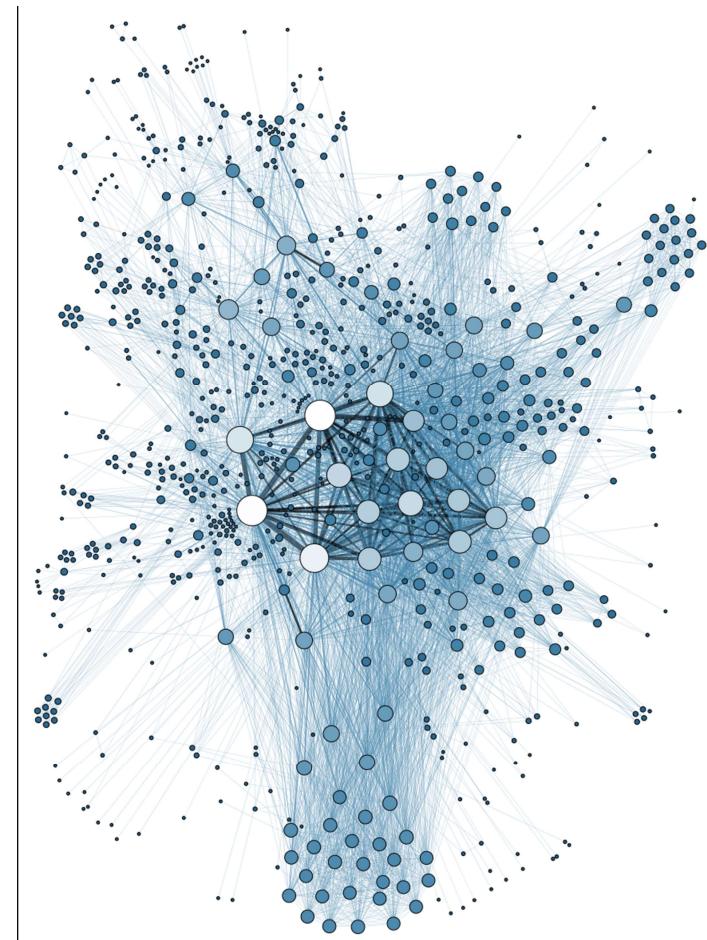
Naomi Arnold

<https://narnolddd.github.io/>



# Fully funded 3.5 year industrial PhD in complex networks

- Topic: Equality and social mobility in networks
- Are you interested in studying networks?
- Are you interested in equality?
- QMUL has a 3.5 year PhD fully funded (UK/EU applicants only).
- Funded by Moogsoft, includes fees and stipend.
- Closing date February 15th
- Find out more:
  - Contact supervisor: Richard Clegg
  - Email [r.clegg@qmul.ac.uk](mailto:r.clegg@qmul.ac.uk)



# In this tutorial:

- **Recap** on concepts and metrics covered in the lecture
- **Understand** the Erdos-Renyi random graph model
- **Understand** some of the key similarities and differences between random graphs and real networks

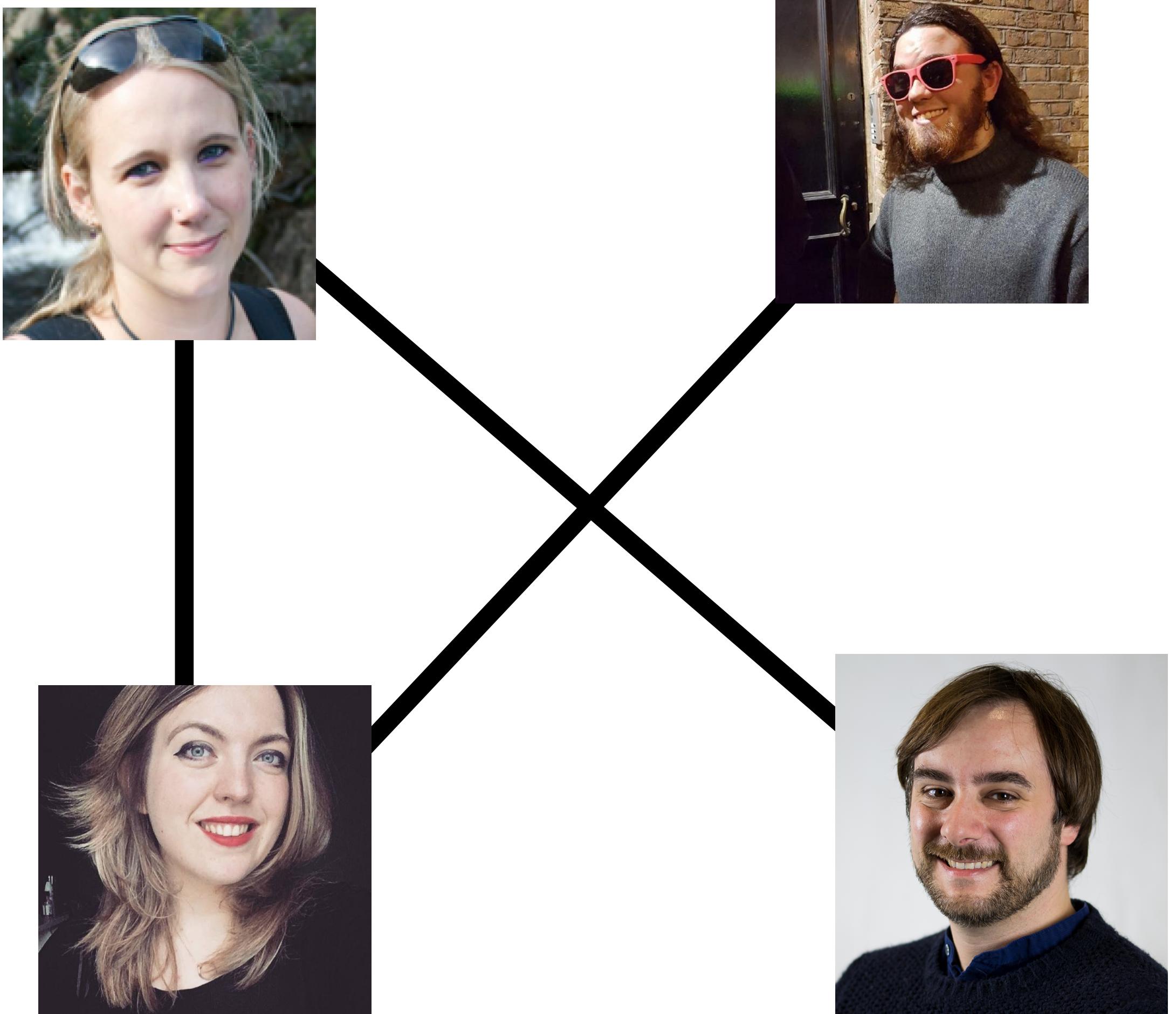
# Recap: (Undirected) Graph

A **graph** is a tuple **(V,E)** of a set **V** of vertices and **E** of edges

**Vertex** (node) set: {Laurissa, Ben, Naomi, Mathieu}

**Edge** (link) set: { (Laurissa, Naomi), (Laurissa, Mathieu), (Naomi, Ben)}

Here, order doesn't matter as graph is **undirected**

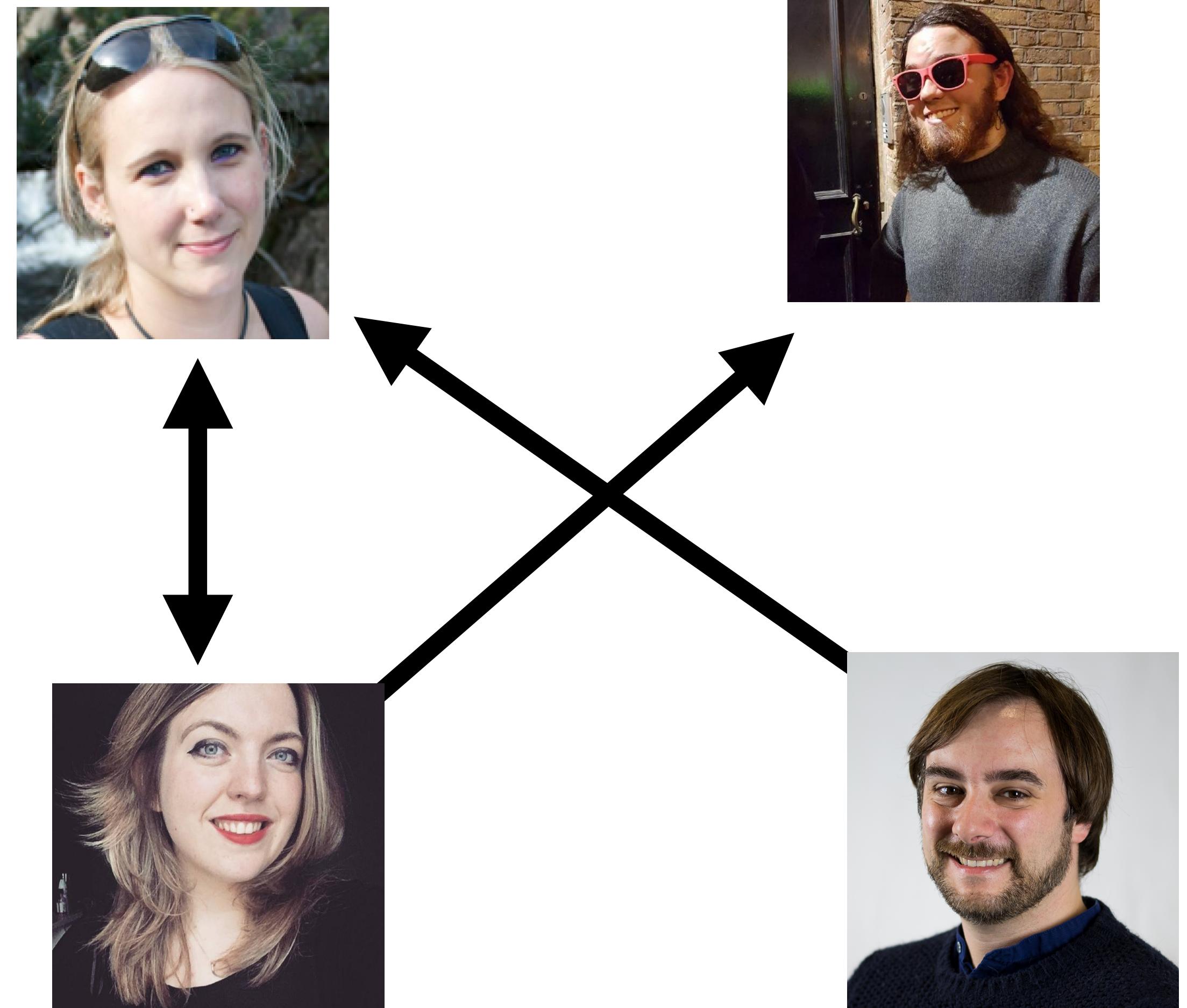


# Recap: Directed Graph

**Vertex (node) set:** {Laurissa, Ben, Naomi, Mathieu}

**Edge (link) set:** { (Laurissa, Naomi),  
(Naomi, Laurissa)  
(Mathieu, Laurissa),  
(Naomi, Ben)}

Here, order **does** matter as  
graph is **directed**



**How do we measure graphs?  
How do we compare them?**

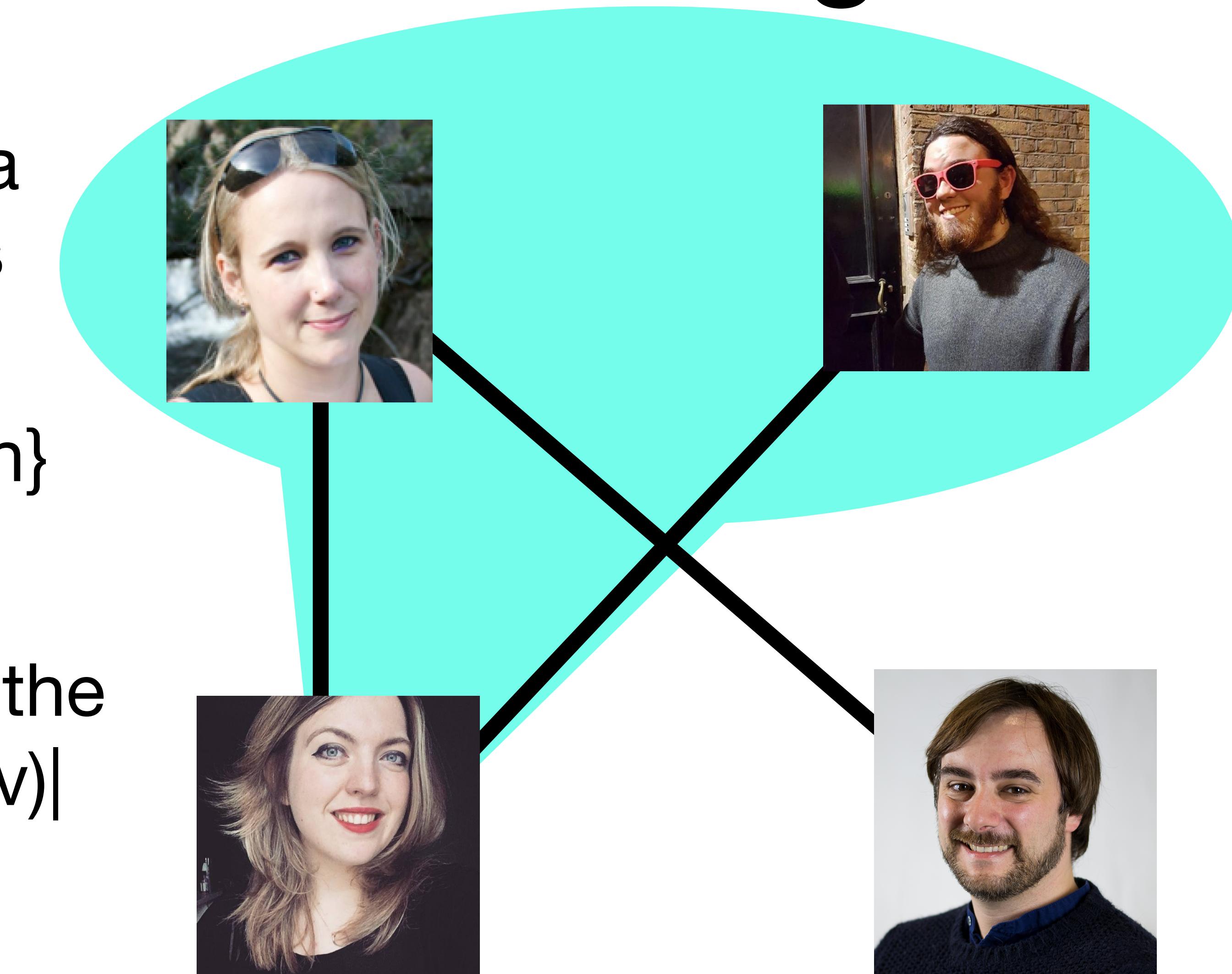
# Recap: Neighbourhood and Degree

The **neighbourhood**  $N(v)$  of a vertex  $v$  is the set of vertices adjacent to  $V$

e.g.  $N(Naomi) = \{Laurissa, Ben\}$

The **degree**  $k(v)$  of a vertex  $v$  is the size of the neighbourhood:  $|N(v)|$

e.g.  $k(Naomi) = 2$



# Recap: Degree Sequence/Average Degree

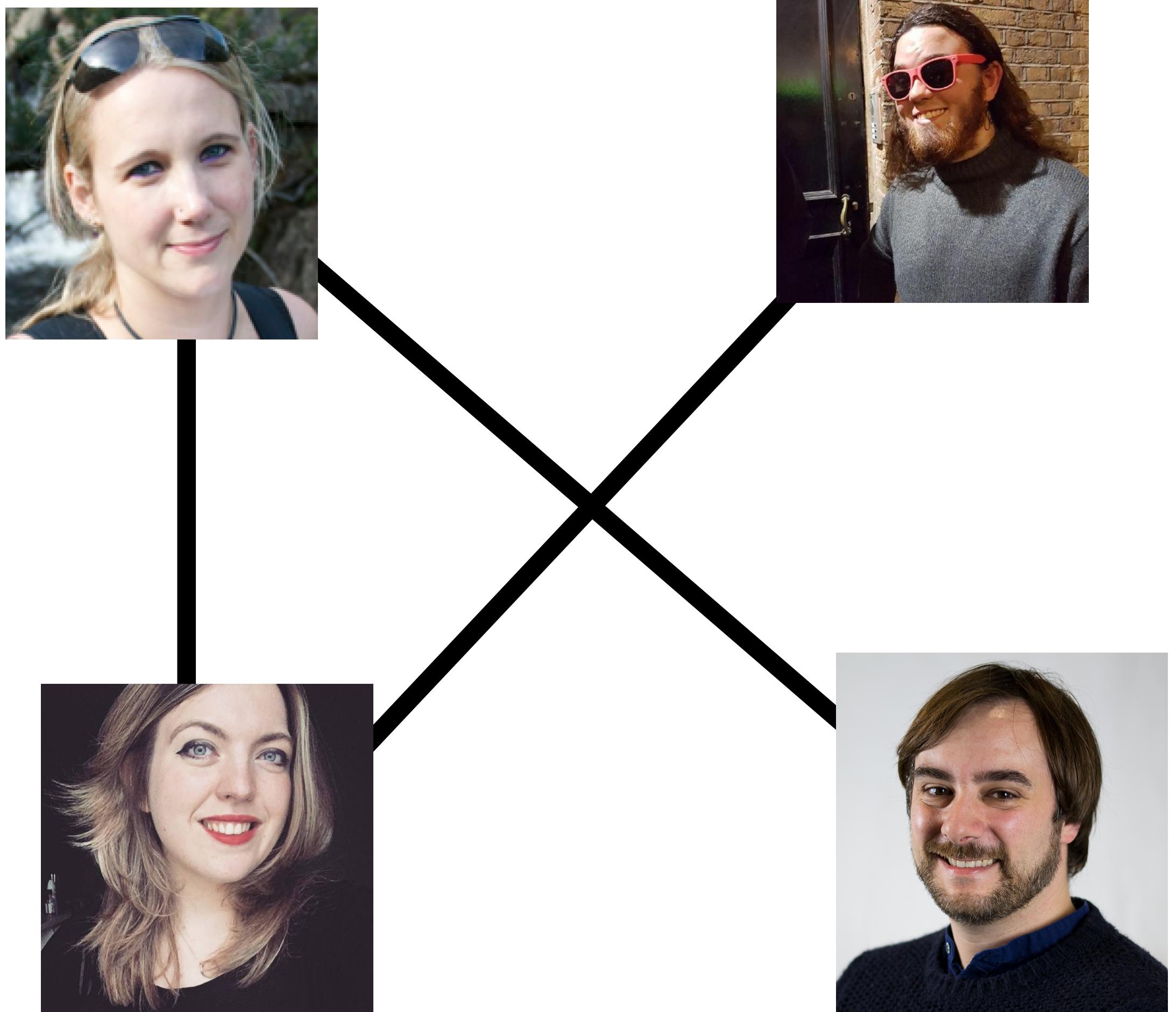
The **degree sequence** of a graph is the list of the vertex degrees for that graph (in decreasing order)

e.g. 2, 2, 1, 1

The **average degree** of a graph  $\langle k \rangle$  is the mean of the node degrees

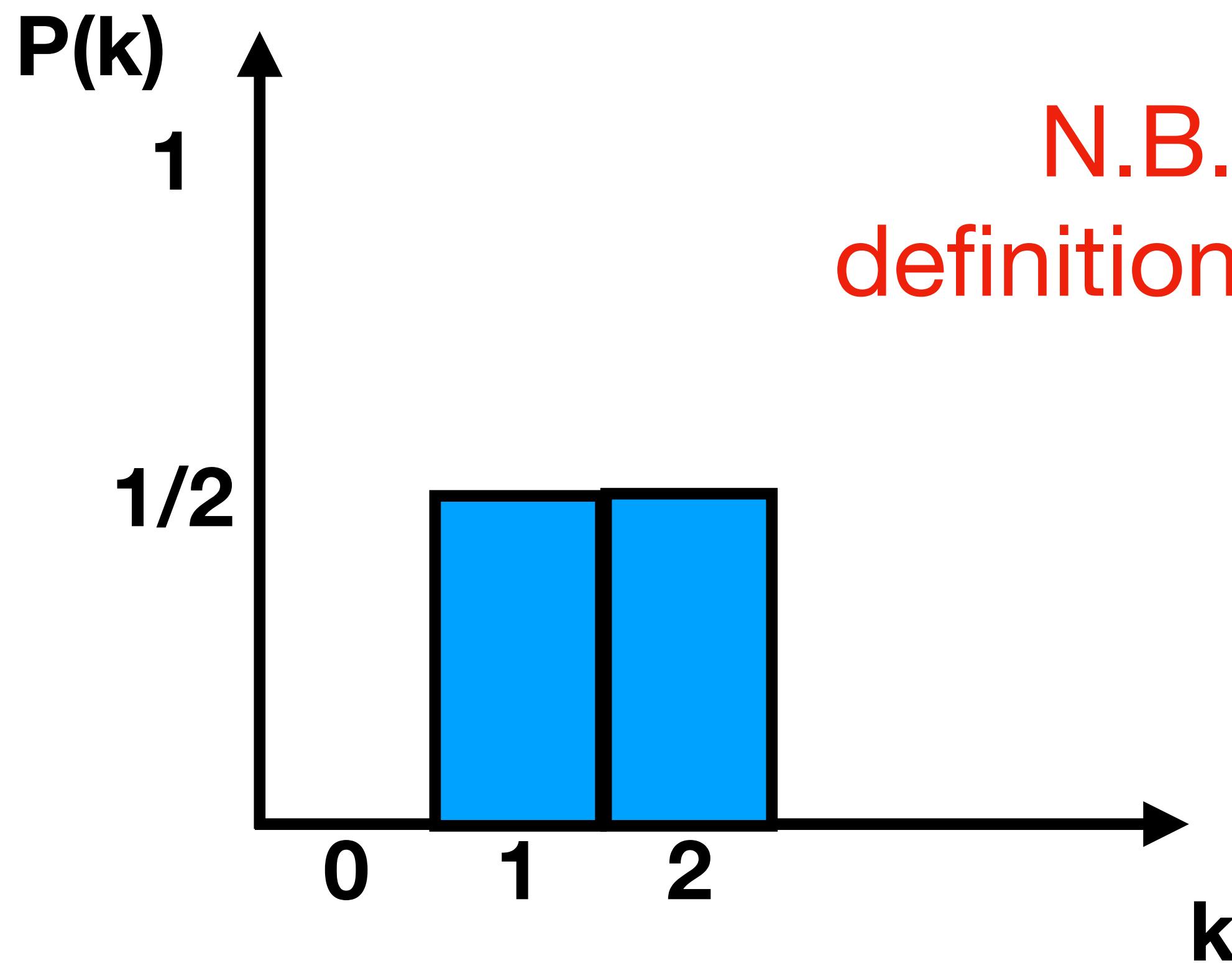
e.g.  $\langle k \rangle = (2 + 2 + 1 + 1)/4 = 1.5$

(also equal to  $2 * |\text{edges}| / |\text{nodes}|$   
... why?)

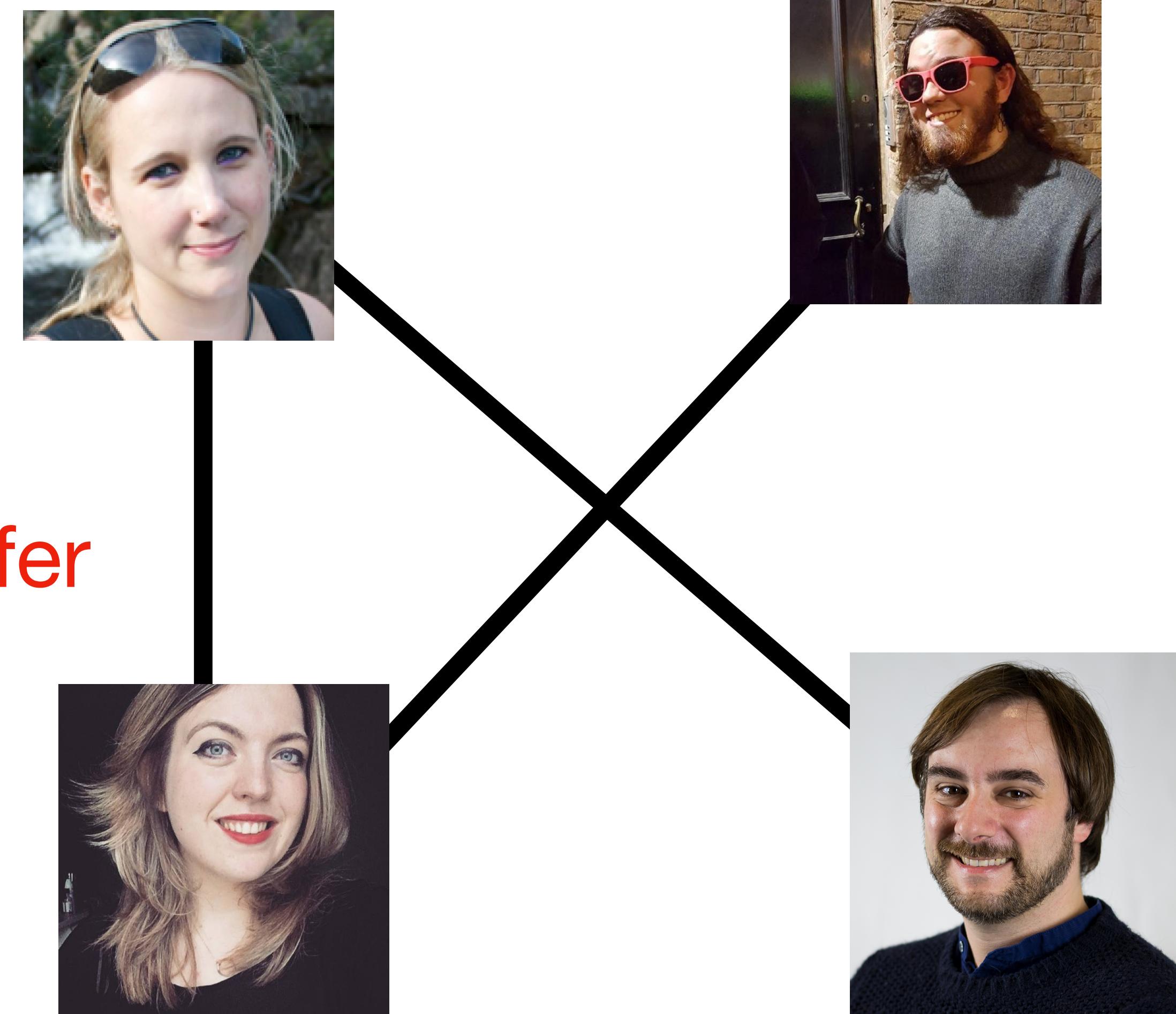


# Recap: Degree distribution

The degree distribution  $P(k)$  is the proportion of nodes with degree equal to  $k$

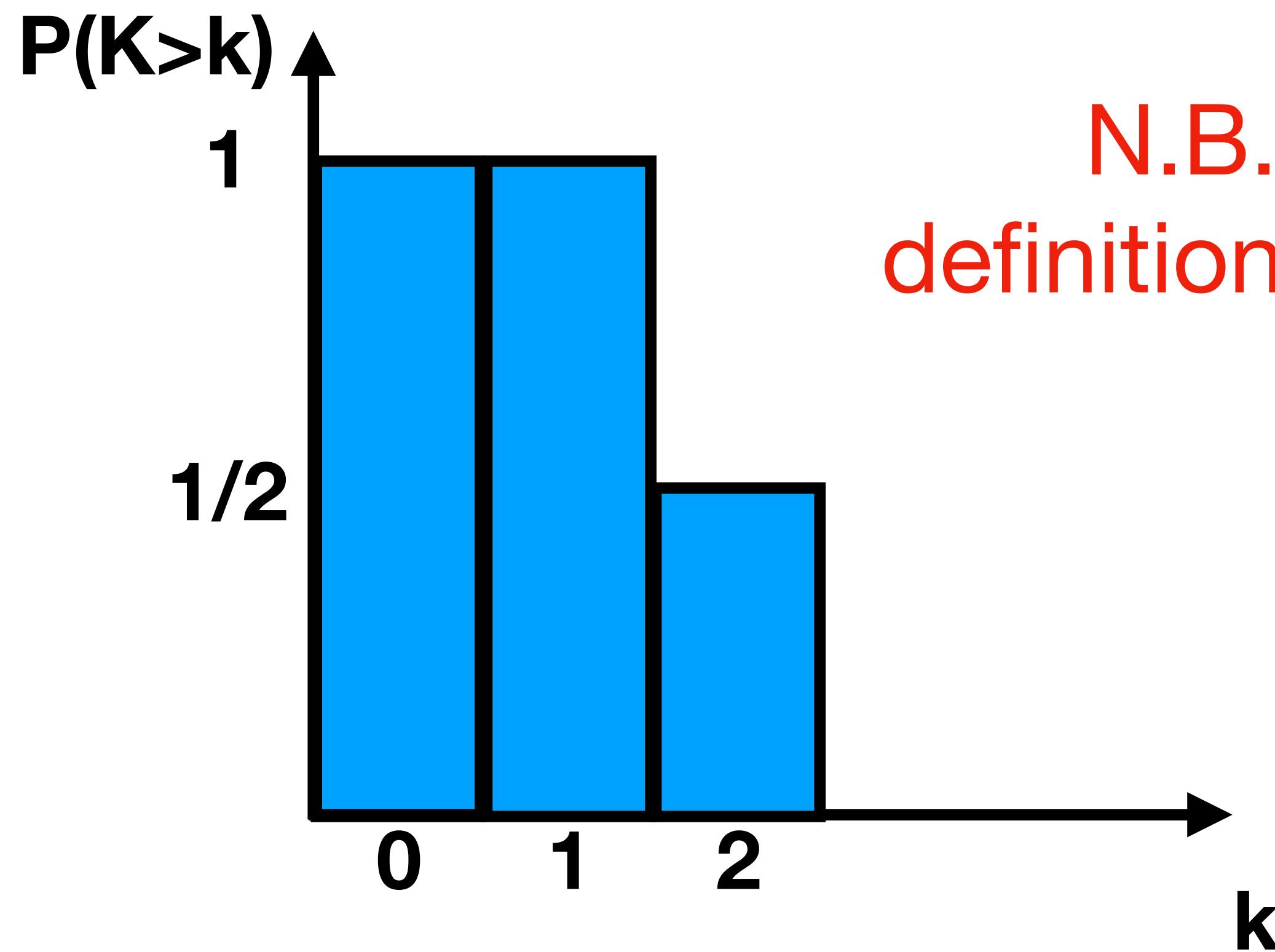


N.B. some  
definitions will differ

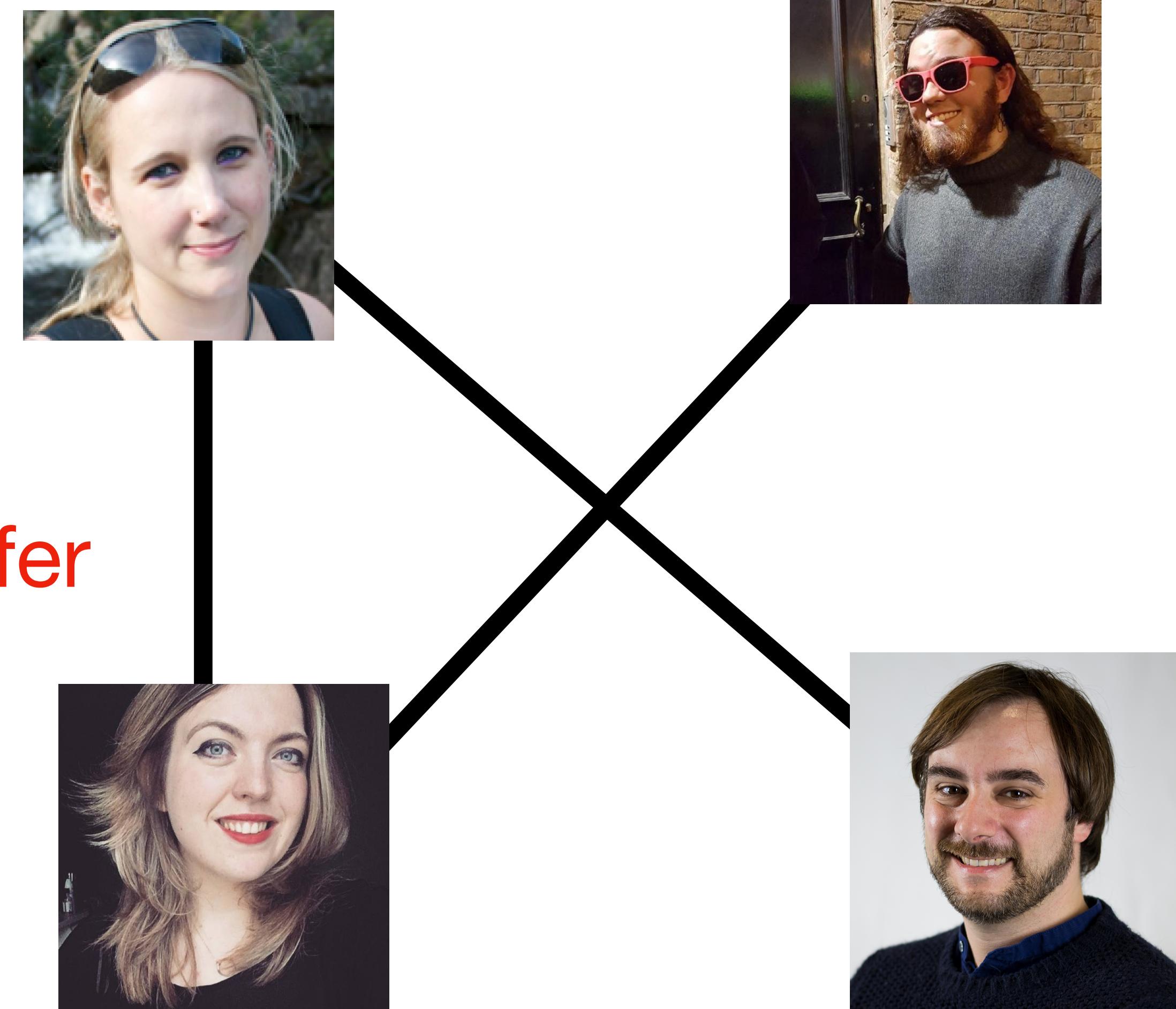


# Recap: Degree distribution

... but it's common to look instead at the proportion of nodes with degree **greater than or equal to k**



N.B. some definitions will differ



# Recap: Clustering Coefficient

The **clustering coefficient  $C(v)$**  is the proportion of possible connections between neighbours of  $v$  that are present

$$C(v) = \frac{|\{(u, w) \mid u, w \in N(v)\}|}{\frac{1}{2}k(v)(k(v) - 1)}$$

Pairs of neighbours of  $v$   
that are connected

Possible pairs of  $v$ 's  
neighbours, “ $k(v)$  choose 2”

**Special case:** if  
 $k(v) = 1$  or  $0$ ,  
 $C(v) = 0$

# Recap: Clustering Coefficient

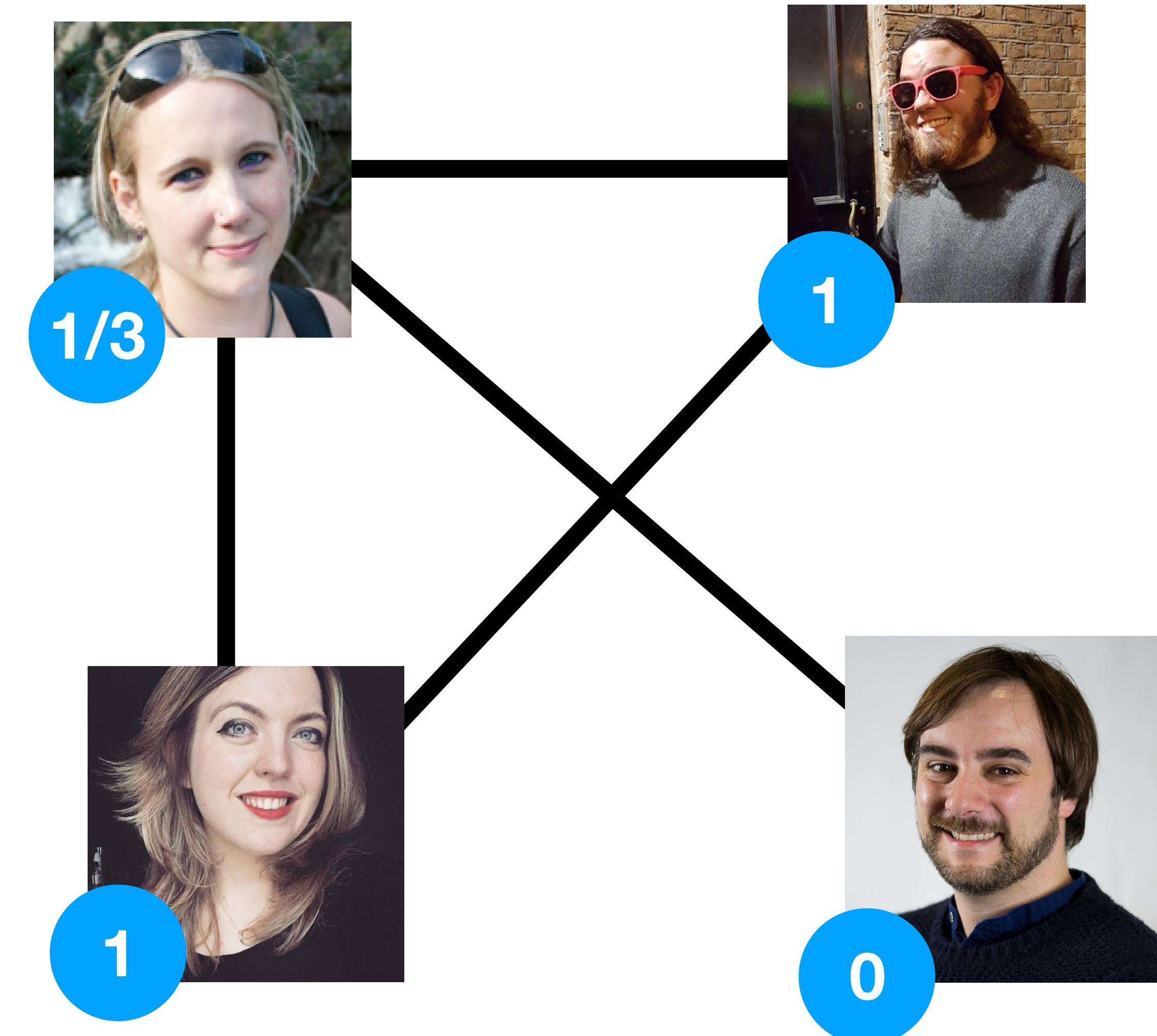
What is Laurissa's clustering coefficient?

**Denominator:** Laurissa's degree is 3, so  $0.5 \cdot 3 \cdot 2 = \underline{3}$

**Numerator:** Only one pair of Laurissa's neighbours are connected (Naomi, Ben)

So  $C(\text{Laurissa}) = \underline{1/3}$

Average clustering  $C(G) = 7/12$



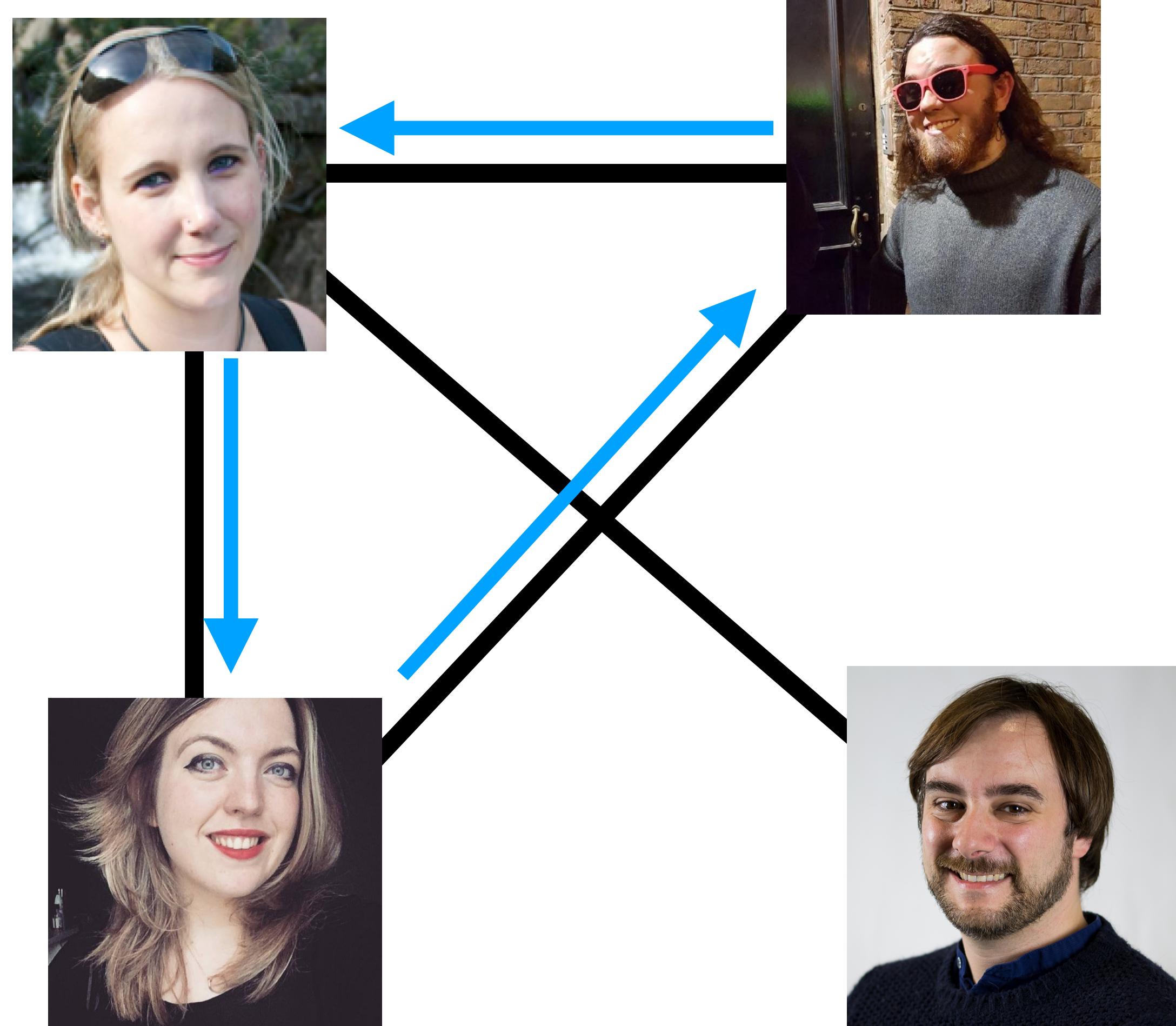
# Recap: Paths and Cycles

A **path** is a sequence of nodes where each consecutive pair of nodes is linked by an edge

Ben, Laurissa, Naomi

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

Ben, Laurissa, Naomi, Ben



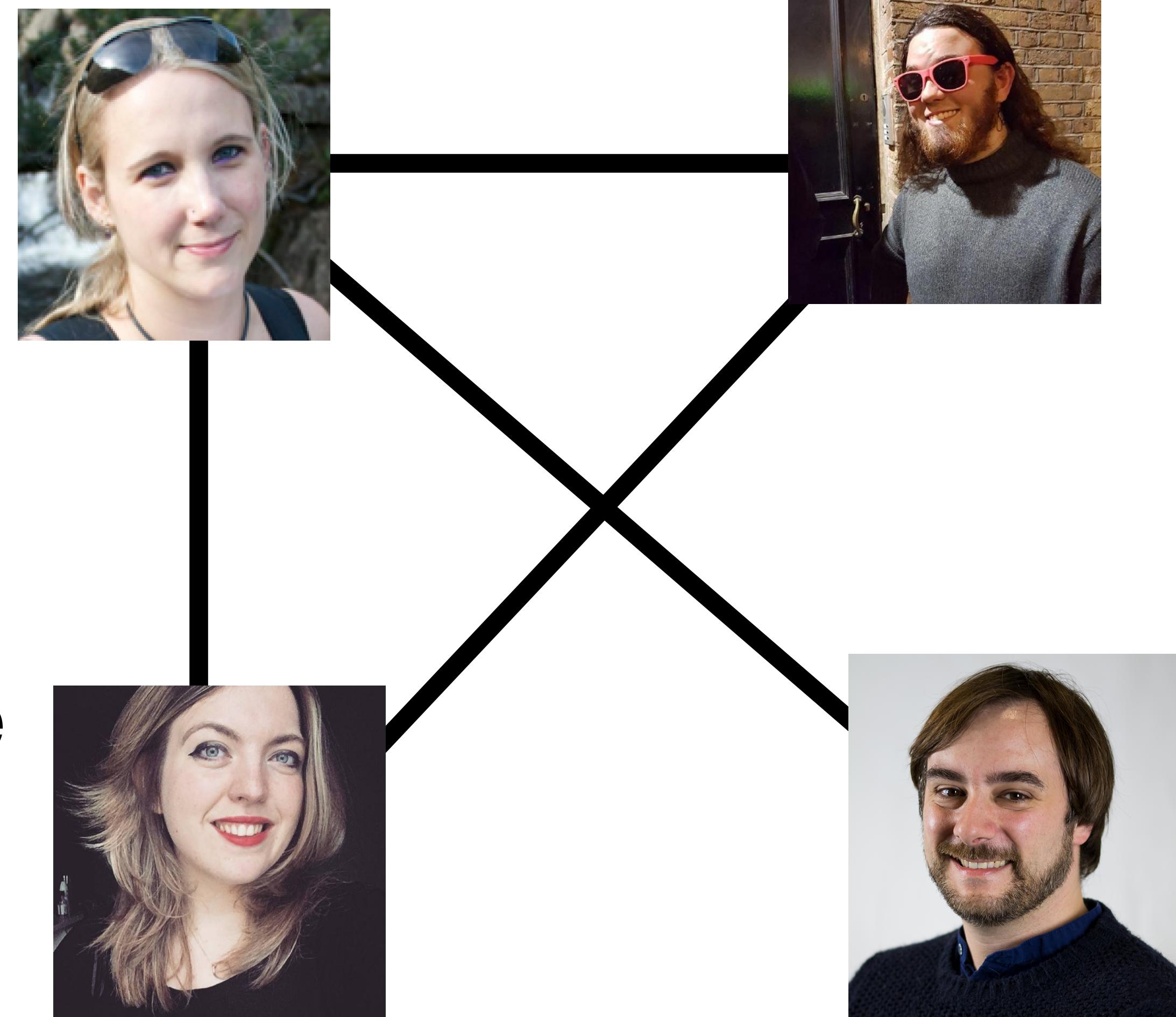
# Recap: Paths and Cycles

The **distance**  $d(u,v)$  between two nodes is the length of the shortest path connecting them

$$d(\text{Ben}, \text{Mathieu}) = 2$$

The **diameter** of a graph is the largest distance between a pair of nodes in the graph

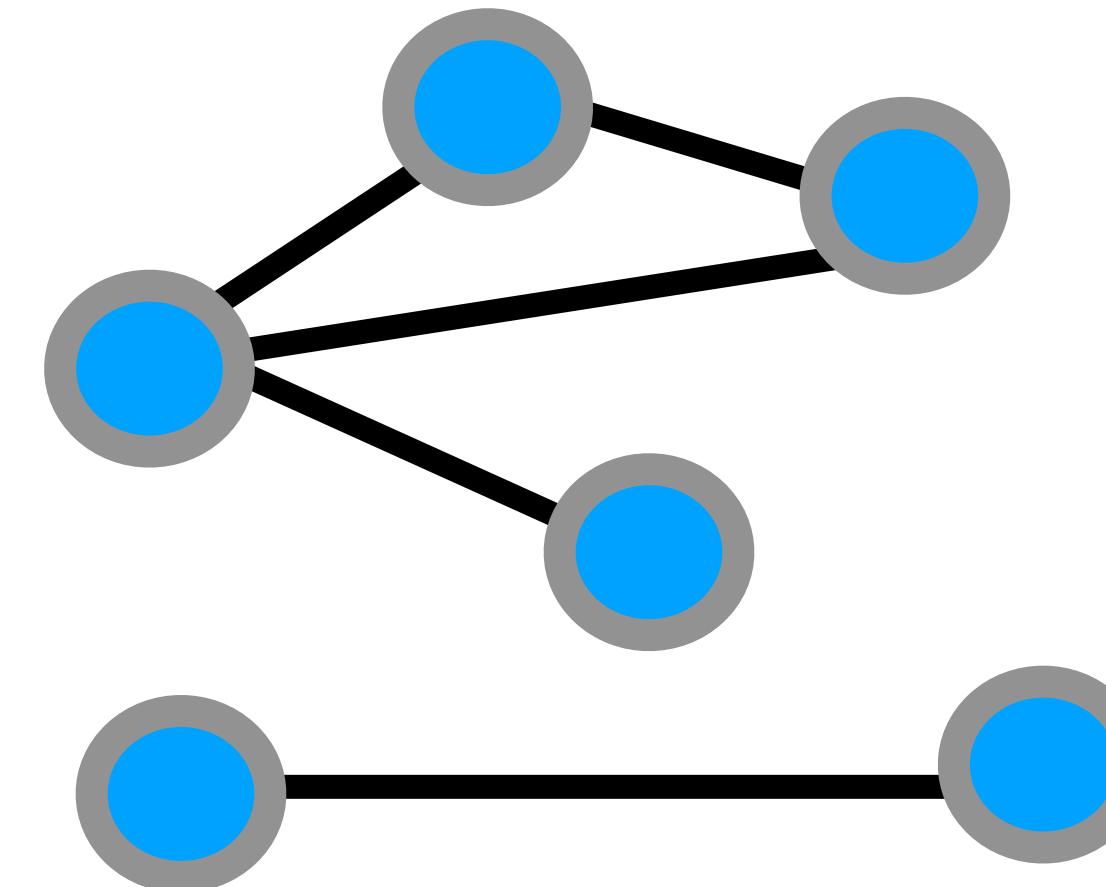
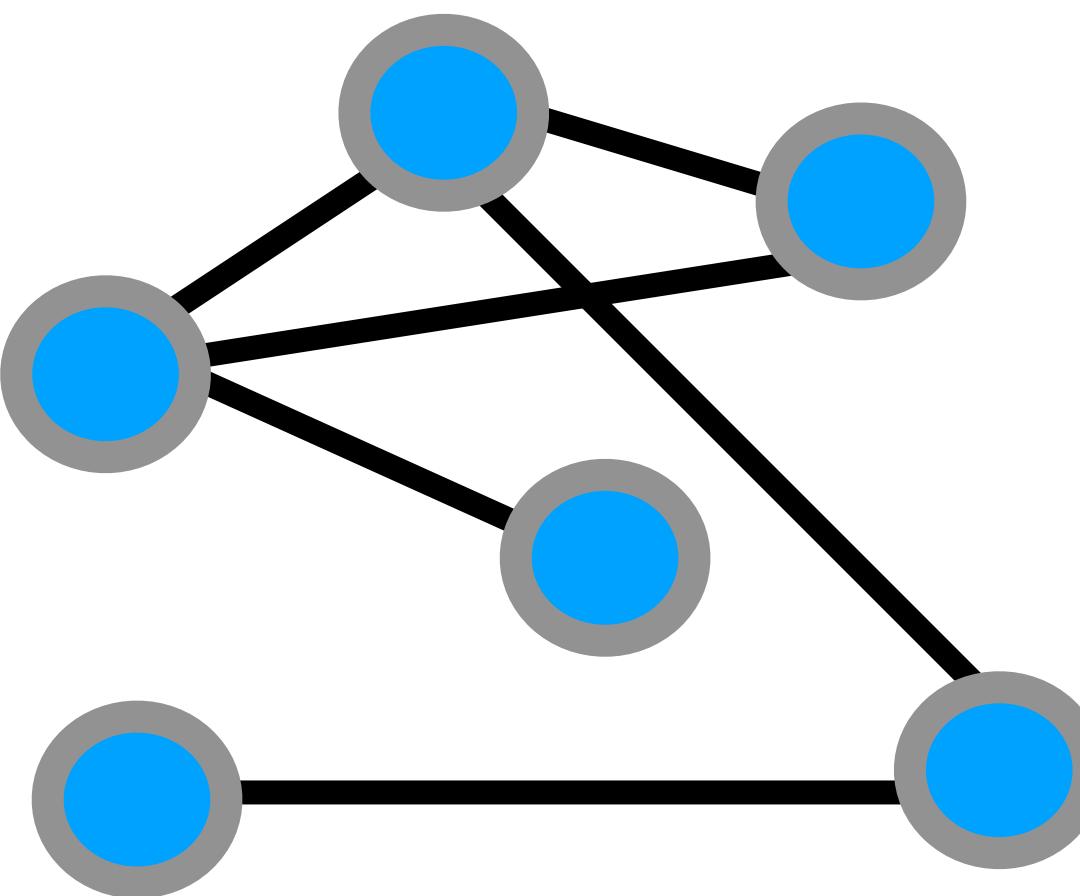
$$d(G) = 2$$



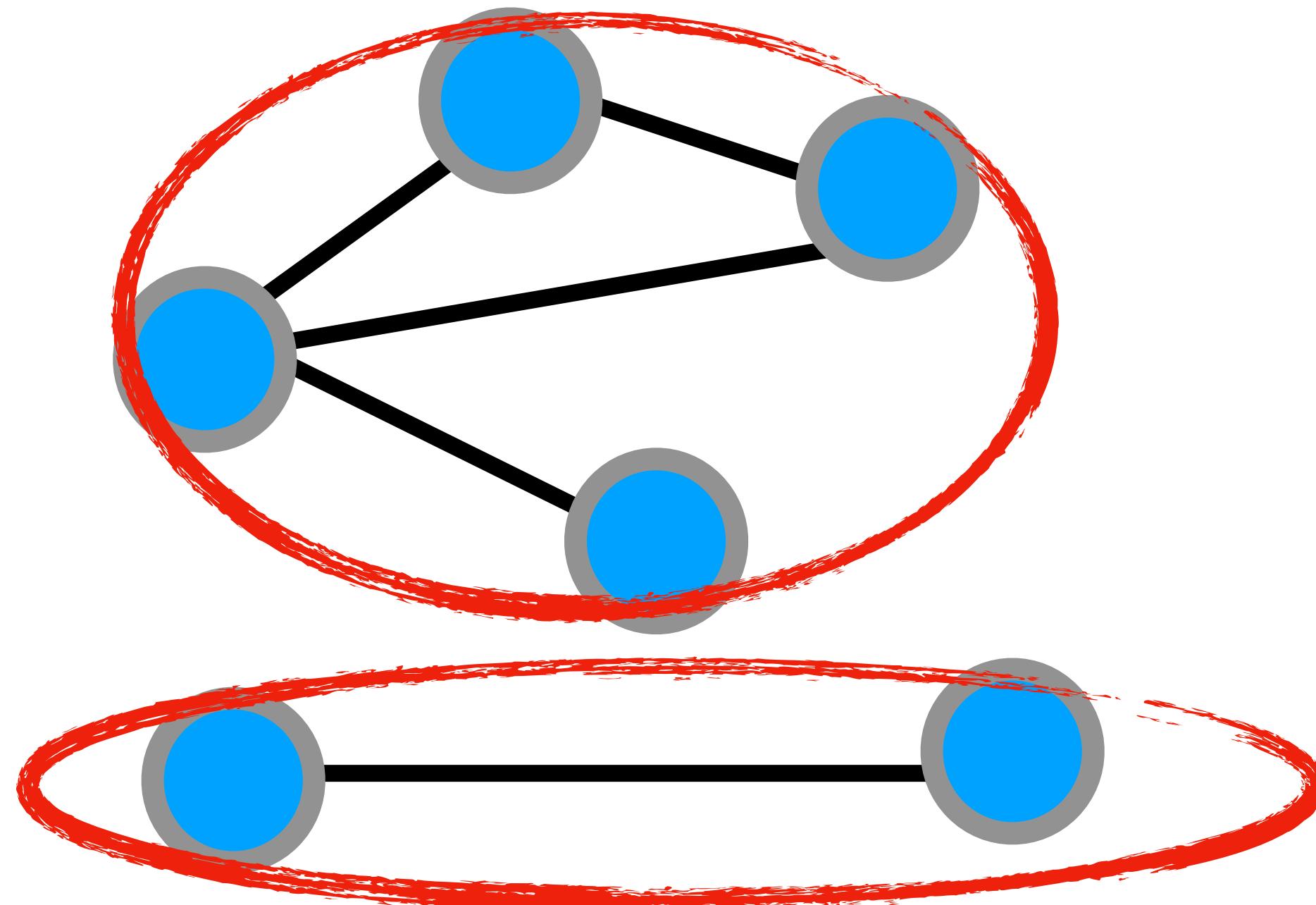
NB: often more meaningful to look at average path length

# Recap: Connected Graph

A graph is **connected** if there is a path between every pair of vertices



# Recap: Connected Components

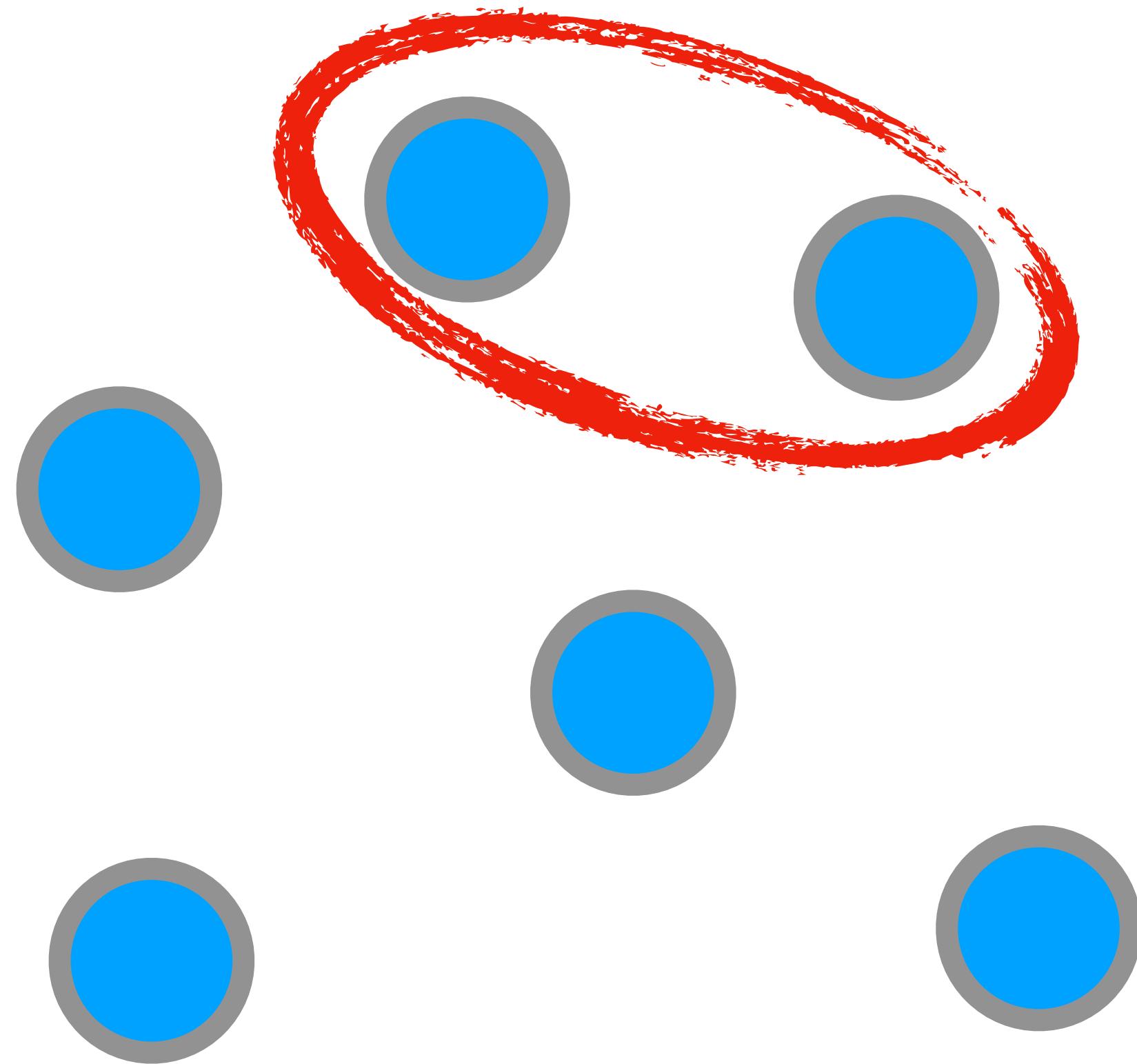


- A **connected component** of a graph  $G$  is a subgraph in which:
1. Any two vertices are **connected** by paths
  2. There are **no edges** to other vertices in  $G$ .

# Erdos-Renyi Random Graph Model

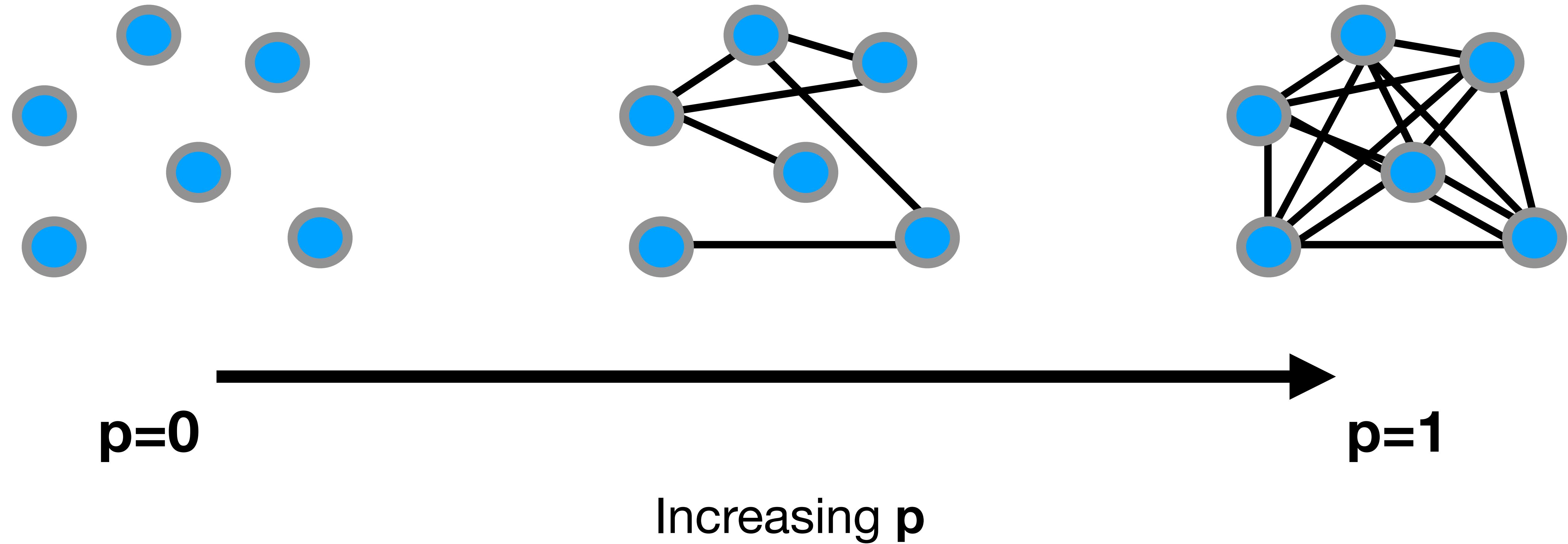
- Want to model real networks, have some **baseline** to compare
- “Is the value of this network metric unusual?” Want a **null model**
- What is the **very simplest** model formulation we can look at?

# Erdos-Renyi $G(n,p)$ Model

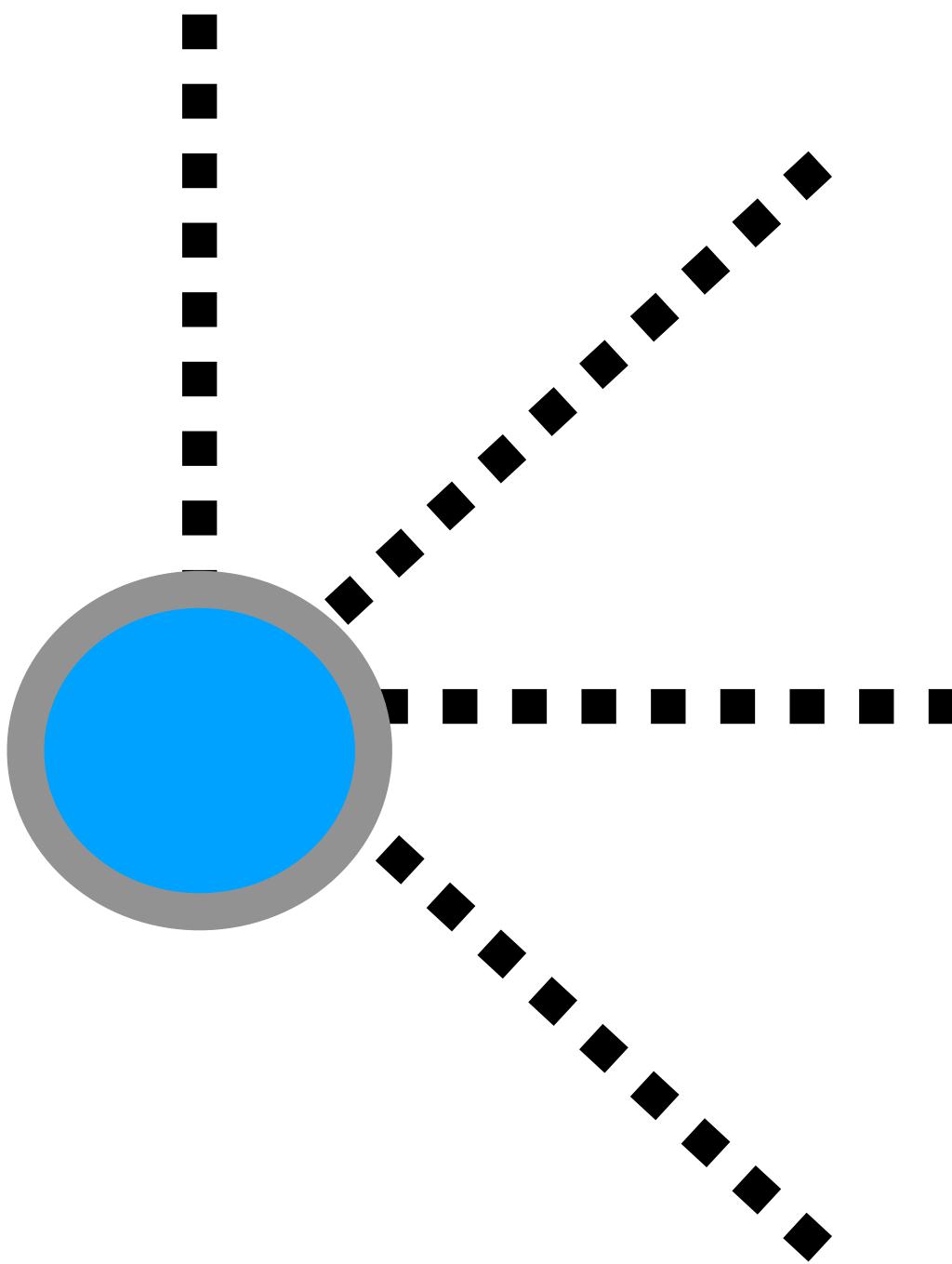


1. Start with an empty graph of  $n$  nodes
2. Acquire a biased coin with head probability  $p$
3. For each pair of nodes, do a coin toss. If heads, draw an edge between them. If not, move on.

# Erdos-Renyi $G(n,p)$ model



# Average degree of ER networks



For each node, there are  $n-1$  others in the graph it could connect to.

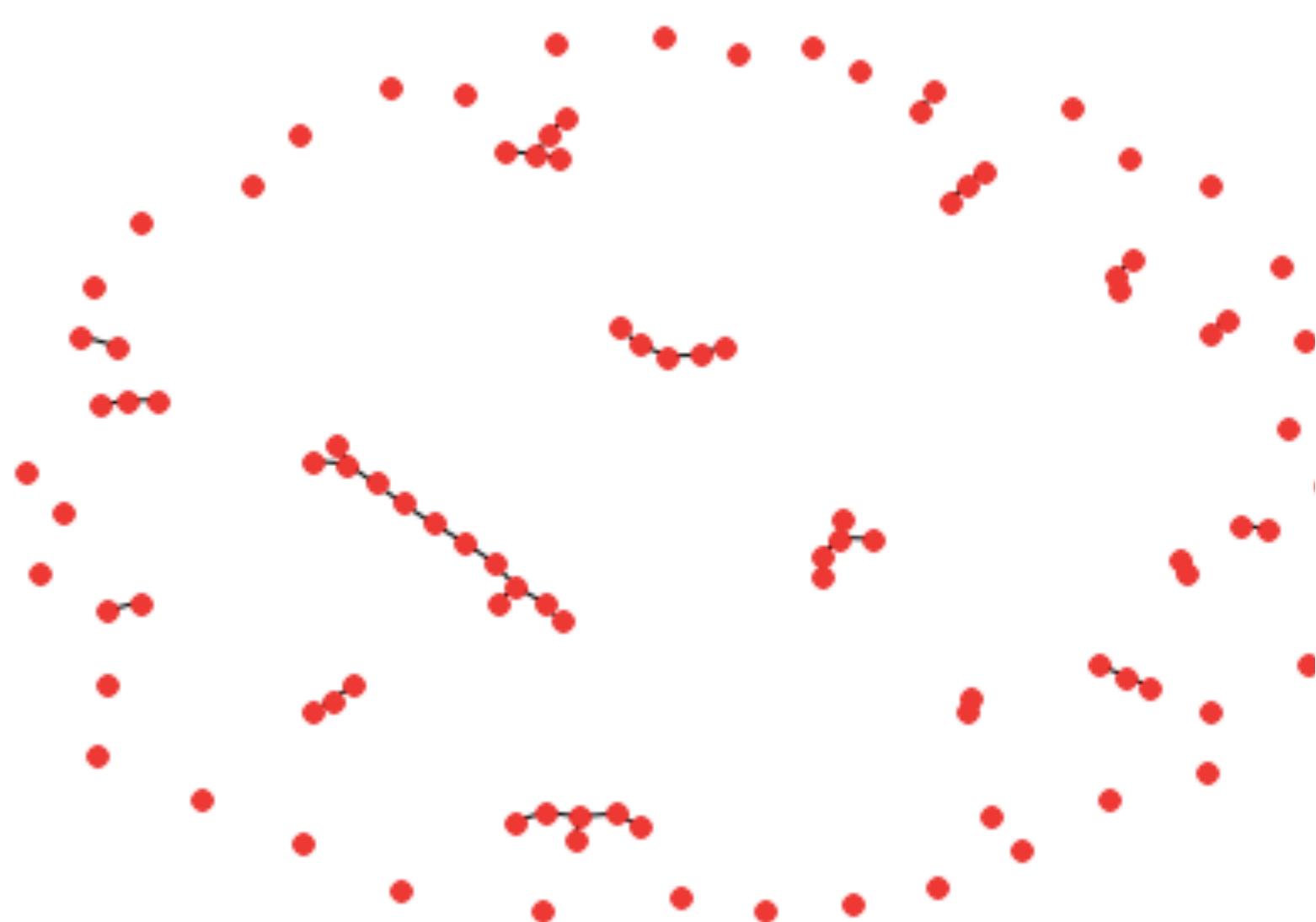
Each of those connections can happen with probability  $p$

(If you were a fan of Probability and Matrices, this is a binomial with  **$n-1$  trials** and **success probability  $p$** )

So average degree is  $(n-1)p$ , or approximately  $np$

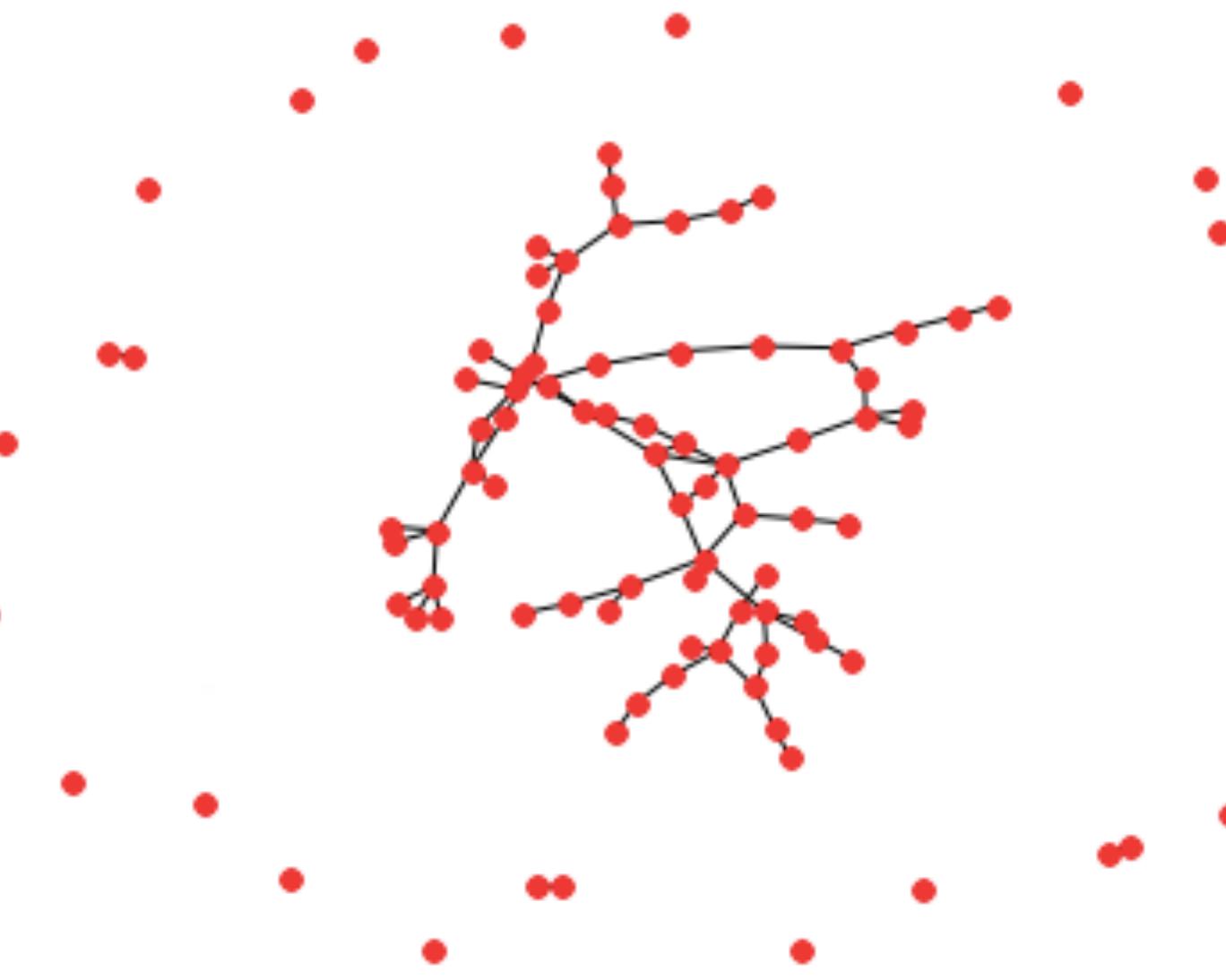
# What do ER graphs look like?

$$p < \frac{1}{n}$$



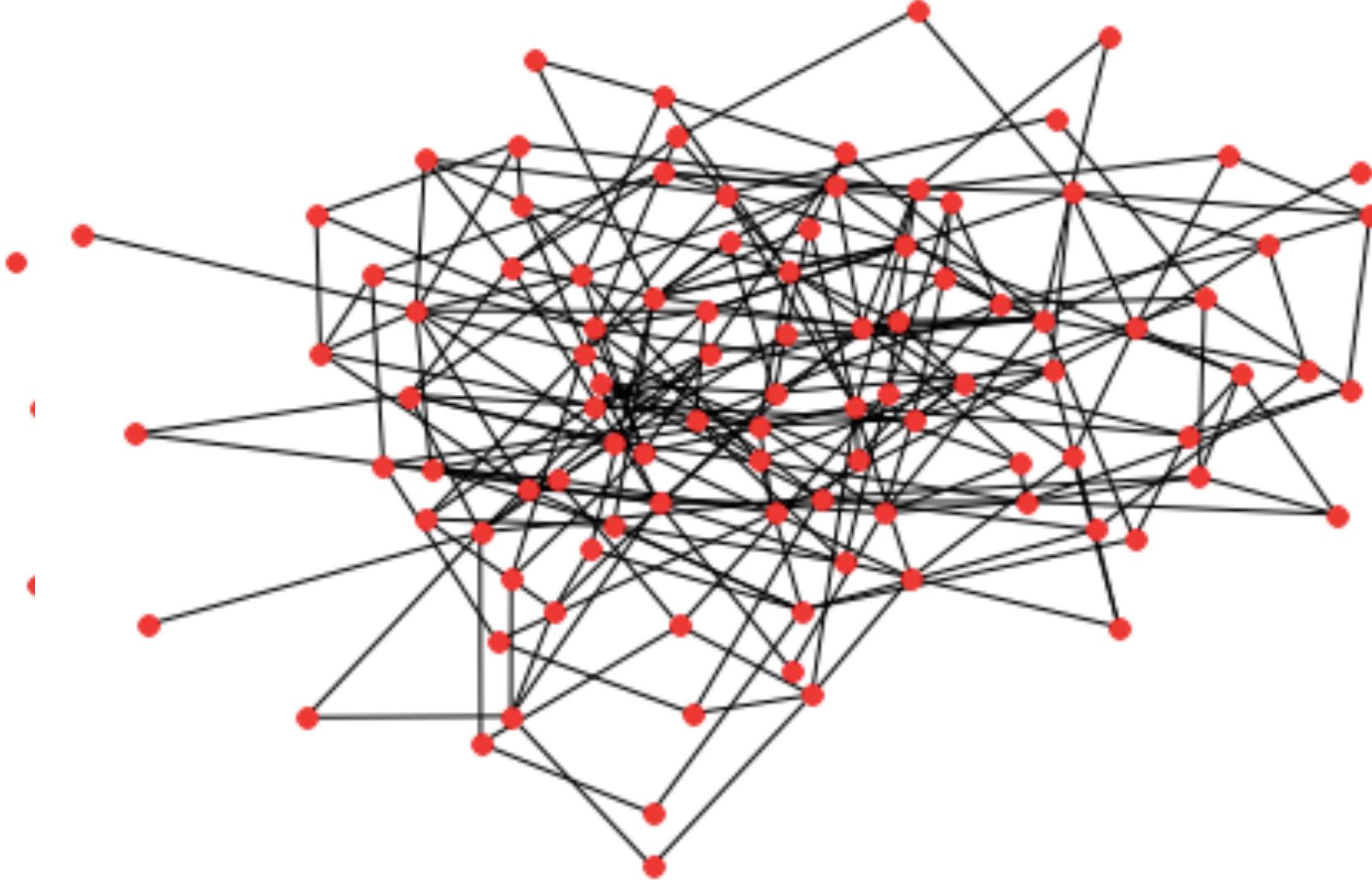
Very disconnected graph,  
only tiny connected  
components

$$p = \frac{1}{n} + \epsilon$$



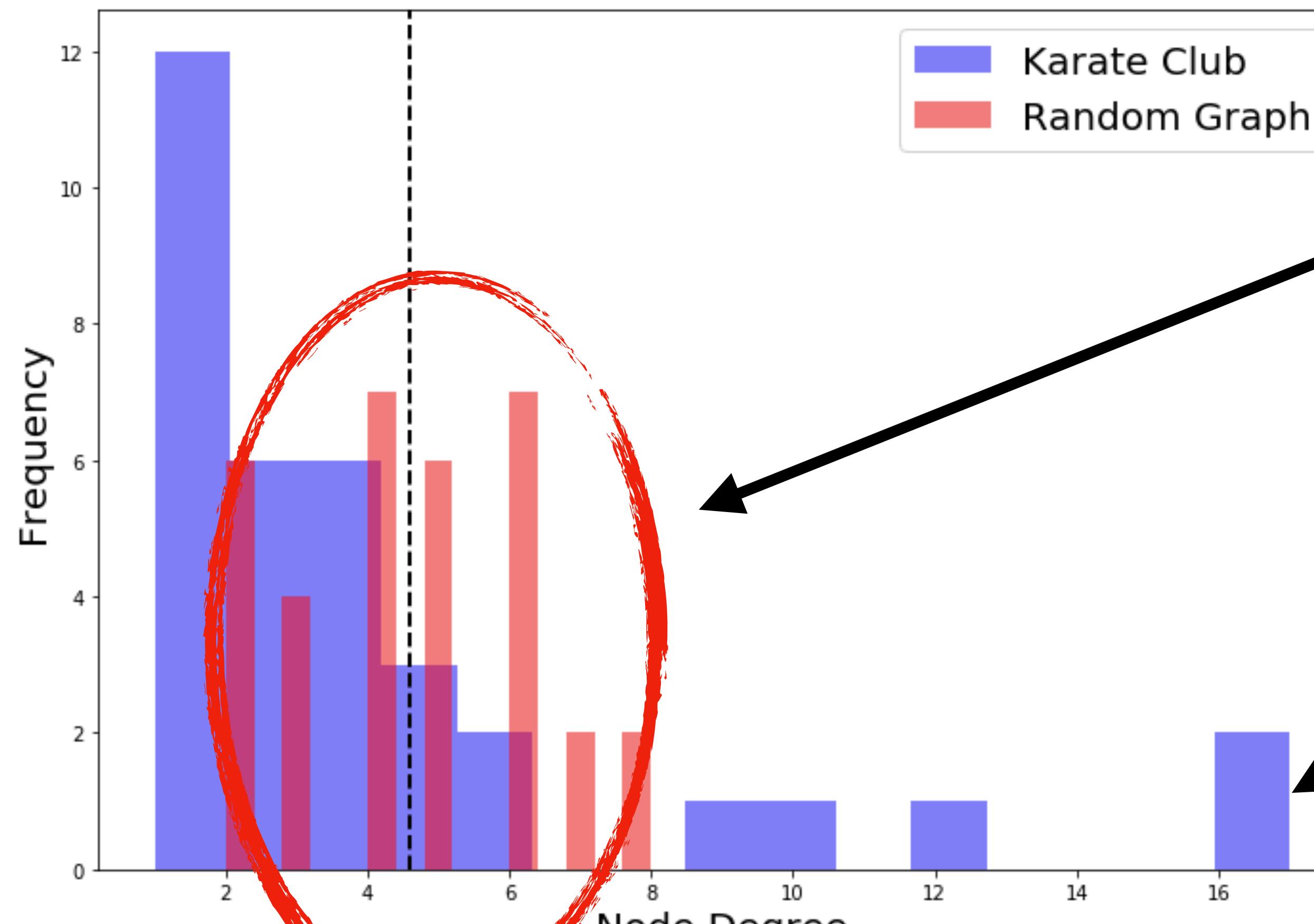
A giant component  
appears, no/very few  
cycles

$$p > \frac{\log(n)}{n}$$



Whole graph is connected,  
some cycles present

# Random Graphs vs Real Networks

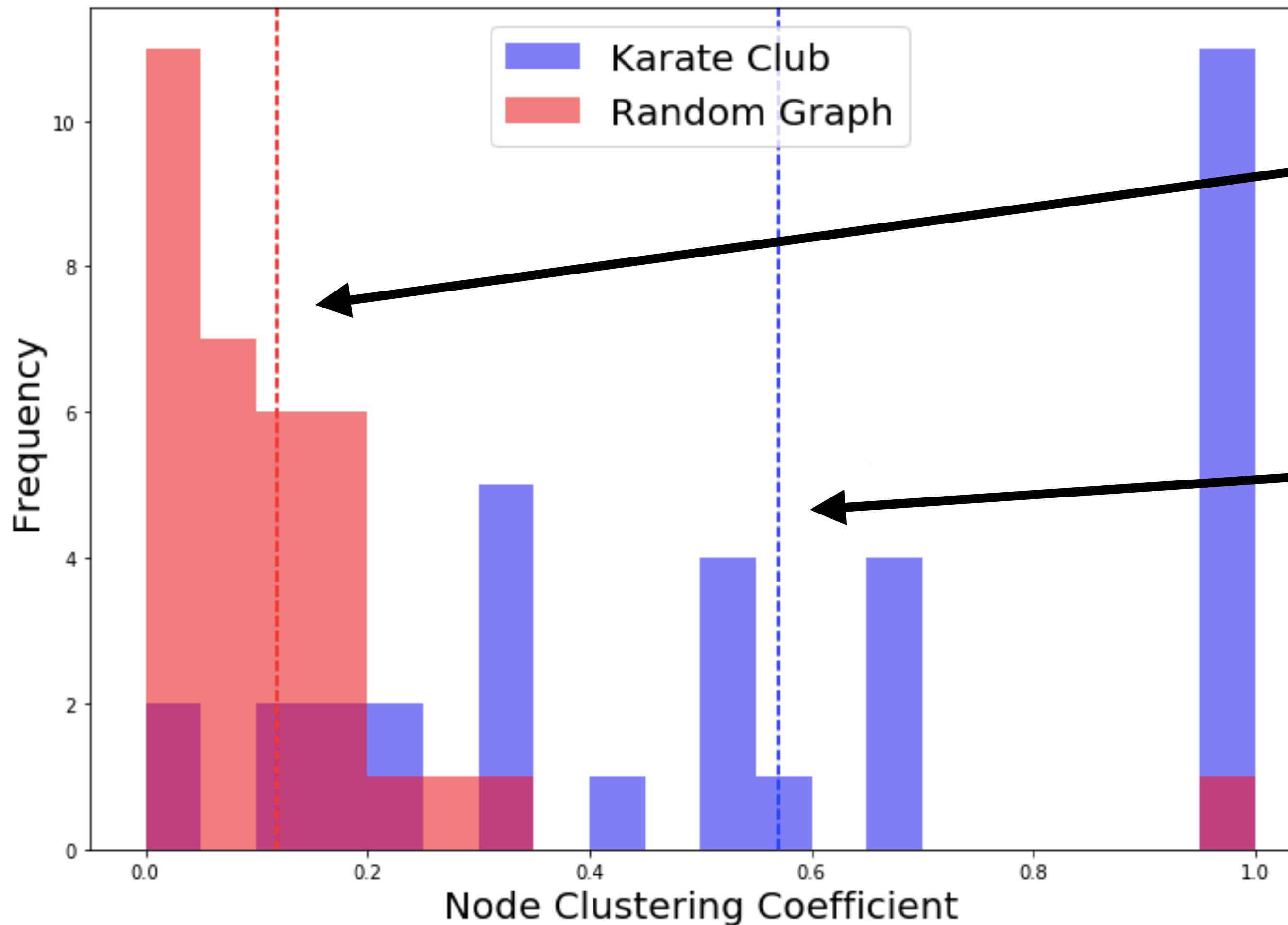


**Random:** node degrees all clustered round the average value

**Real:** small number of high degree nodes, large number of low degree nodes



# Random Graphs vs Real Networks

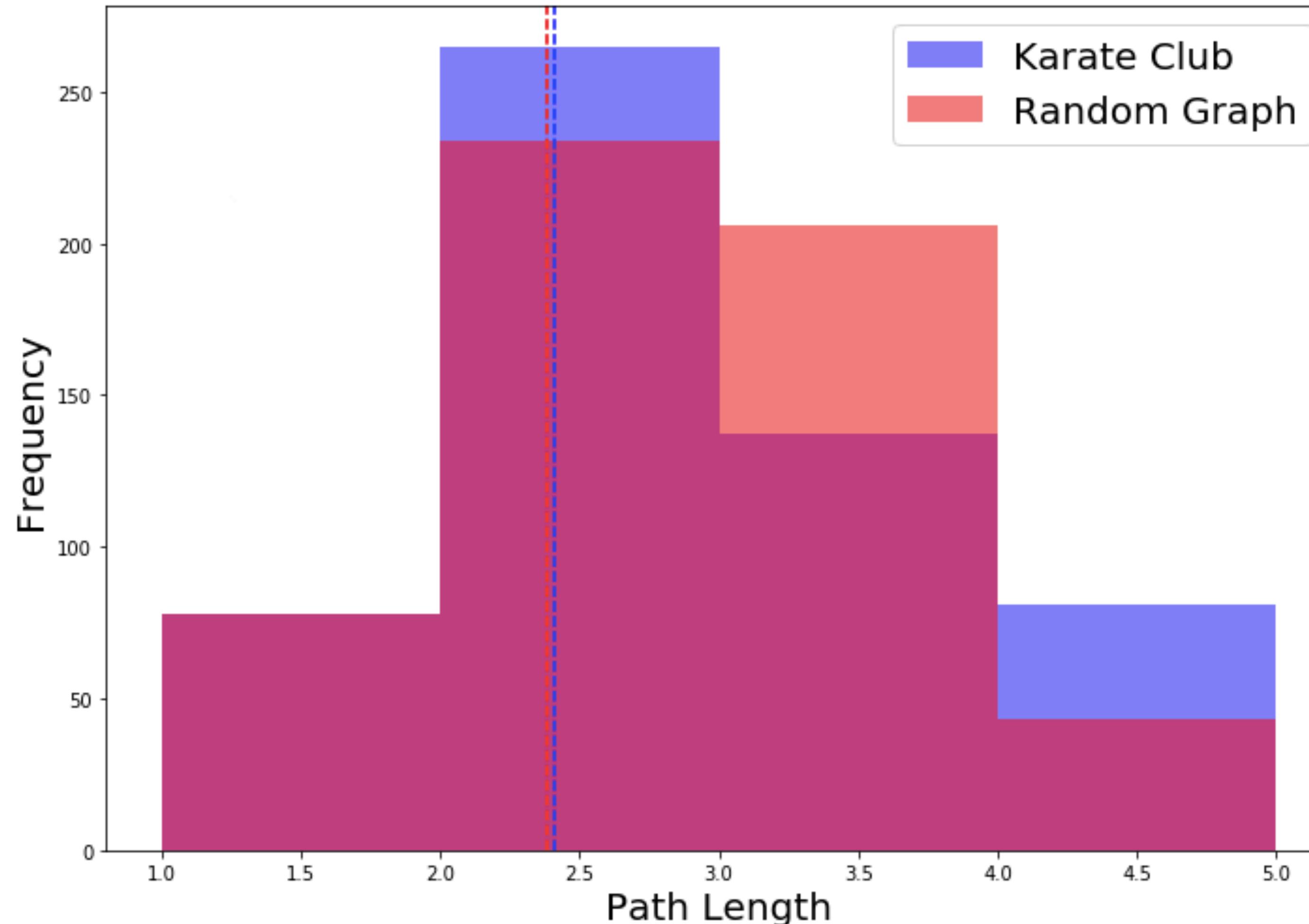


**Random:** very low average clustering coefficient

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



# Random Graphs vs Real Networks



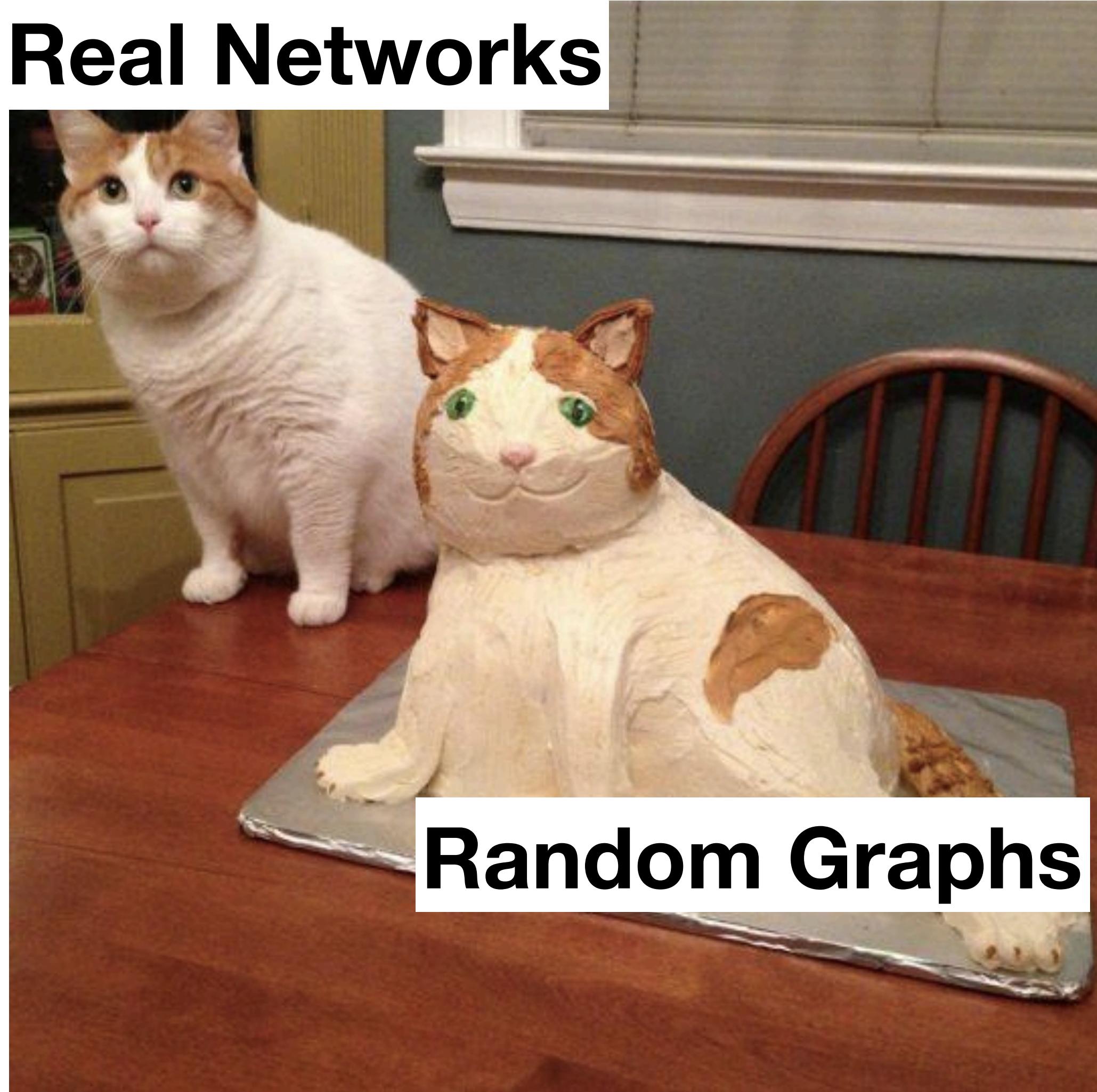
Fairly spot on with  
**almost the same**  
average path length for  
each!



# Summary: Random Graphs vs Real Networks

	Real Social Networks	Random Graphs	?
Degree Distribution	Heavy Tailed (most nodes have low degree, small few with high degree)	Light tailed (all nodes have close to the average degree)	?
Clustering Coefficient	High	Low	?
Path Lengths	Low	Low	?
?	?	?	?

## Real Networks



**Random Graphs**

**Thank you for  
listening! What are  
your questions?**