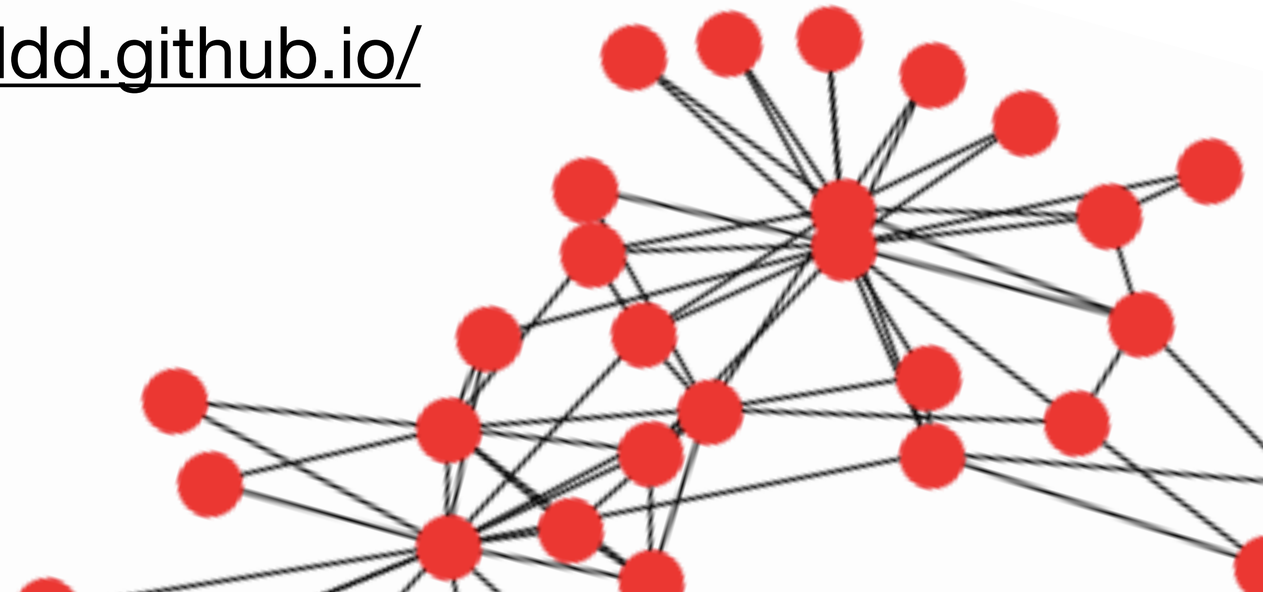
Week 8: Cascades/CW Tutorial

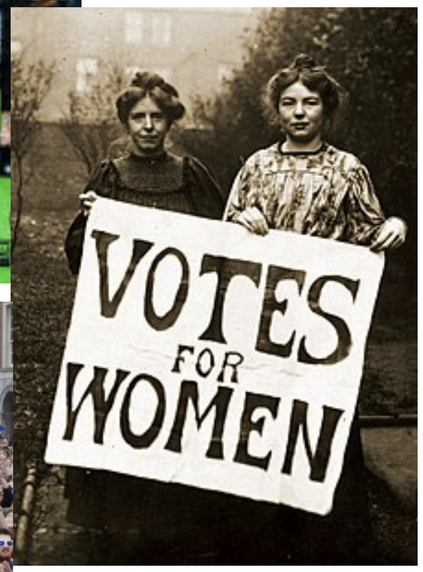
Naomi Arnold
https://narnolddd.github.io/



Tutorial aims

- **Recap** the cascades lecture
- **Go over** a cascade threshold example
- **Get some pointers for** where to start with the network analysis coursework, with Gephi demo
- If **time_left > 0**, answer some coursework questions

What can be modelled as a cascade?



Movements to change a consensus

Uptake of new technologies/products

Important features

- Each agent (node) is in **one of two states** — adopter or non-adopter
- **State** of a node depends only on that of its **neighbours**. I.e. grass-roots and no outside influence
- Each node acts **rationally**

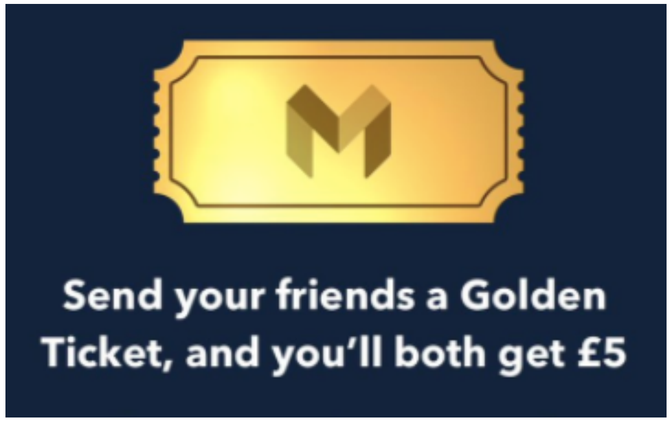
Game-theoretic formulation

- Pair of nodes ***v*** and ***w*** in a social graph
- Behaviours **A** and **B**
- If both adopt **A**, they each get a payoff **a** > 0
- If both adopt **B**, they each get a payoff **b** > 0
- If they adopt **opposite** behaviour, both get a payoff of 0

v \ ***w***	**A**	**B**
A	a,a	0,0
B	0,0	b,b

What can payoff mean?

Social cohesion: easier (maybe!) to be friends with someone you agree with on something



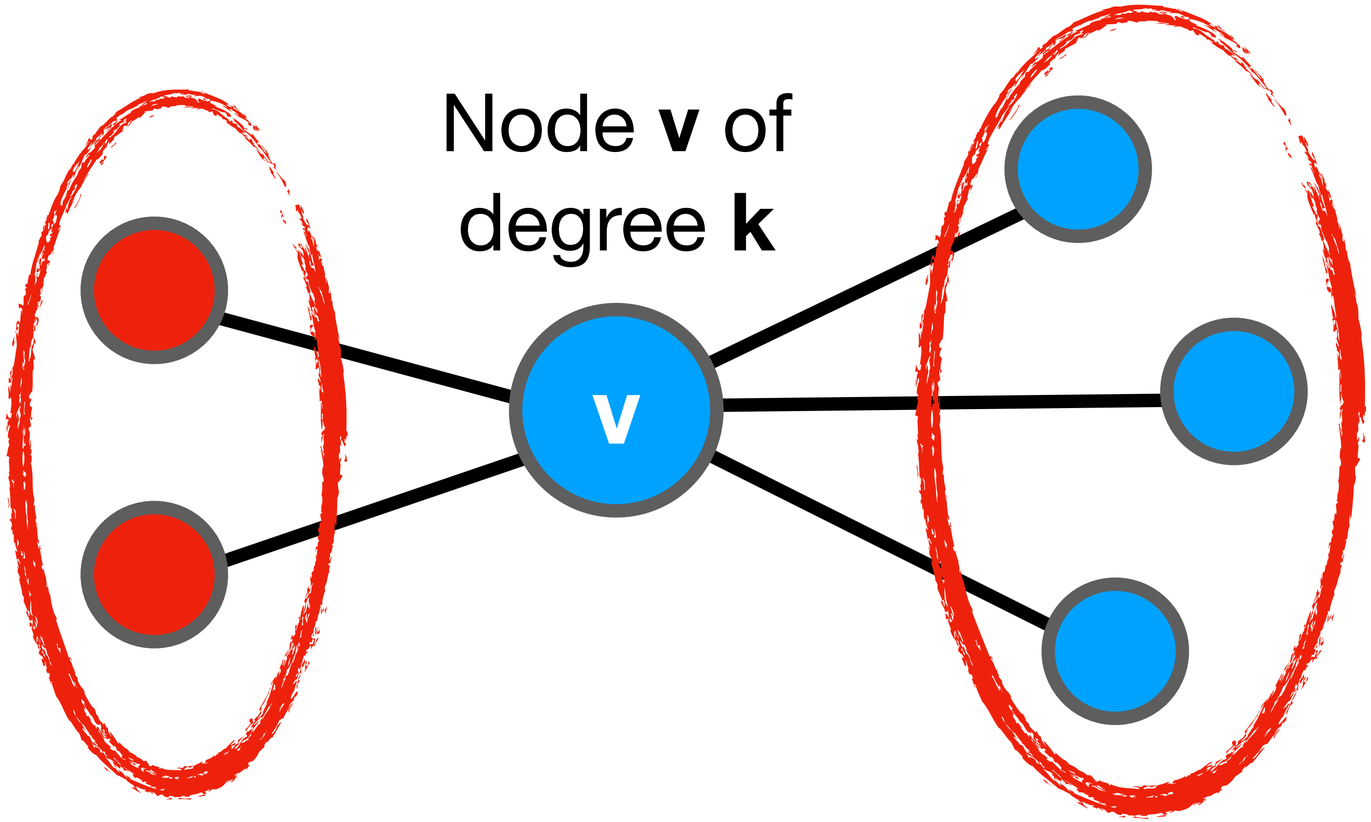
(Monzo bank scheme)

Literal cash: Receive incentives for adopting or encouraging others to adopt

Other benefits: using same technologies/ platforms

Cascade threshold

Proportion **p** of **v**'s neighbours adopt **new** behaviour **A**



The rest (proportion **1-p**) stick with **old** behaviour **B**

v's payoff for adopting **A:**

$$pka$$

v's payoff for adopting **B:**

$$(1-p)kb$$

Cascade threshold

Adopting new behaviour
A is a better* option if:

$$pka \geq (1-p)kb$$

$$p \geq \frac{b}{a+b}$$

*assuming nodes are rational and want option with best payoff — not always the case in reality!!

Cascade threshold

Adopting new behaviour **A** is a better* option if:

$$pka \geq (1-p)kb$$

$$p \geq \frac{b}{a+b}$$

Cascade threshold — proportion of neighbours you need to **convince** you to adopt behaviour **A!**

*assuming nodes are rational and want option with best payoff — not always the case in reality!!

Cascade threshold interpretation

Small threshold

$< 1/2$

New behaviour appealing, little convincing needed

Threshold equal to

$1/2$

People will simply follow majority of their friends

High threshold

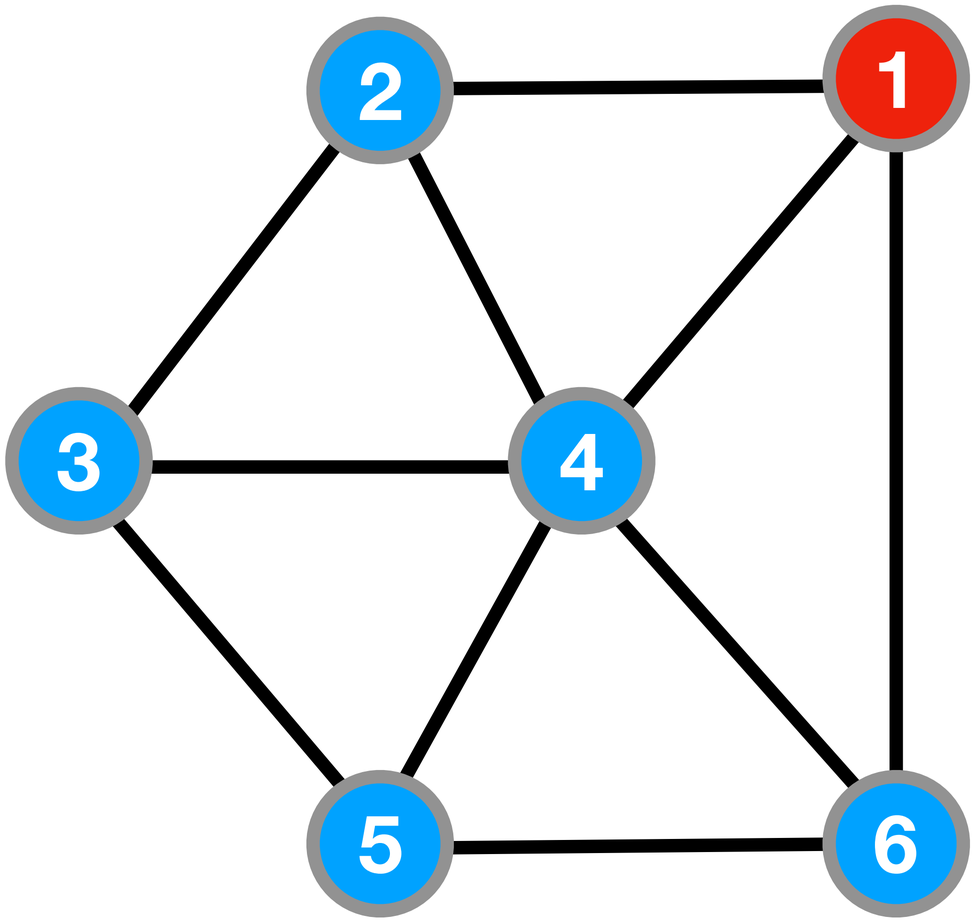
$> 1/2$

New behaviour less appealing, need critical mass of initial adopters

Example

a = 3, b = 1

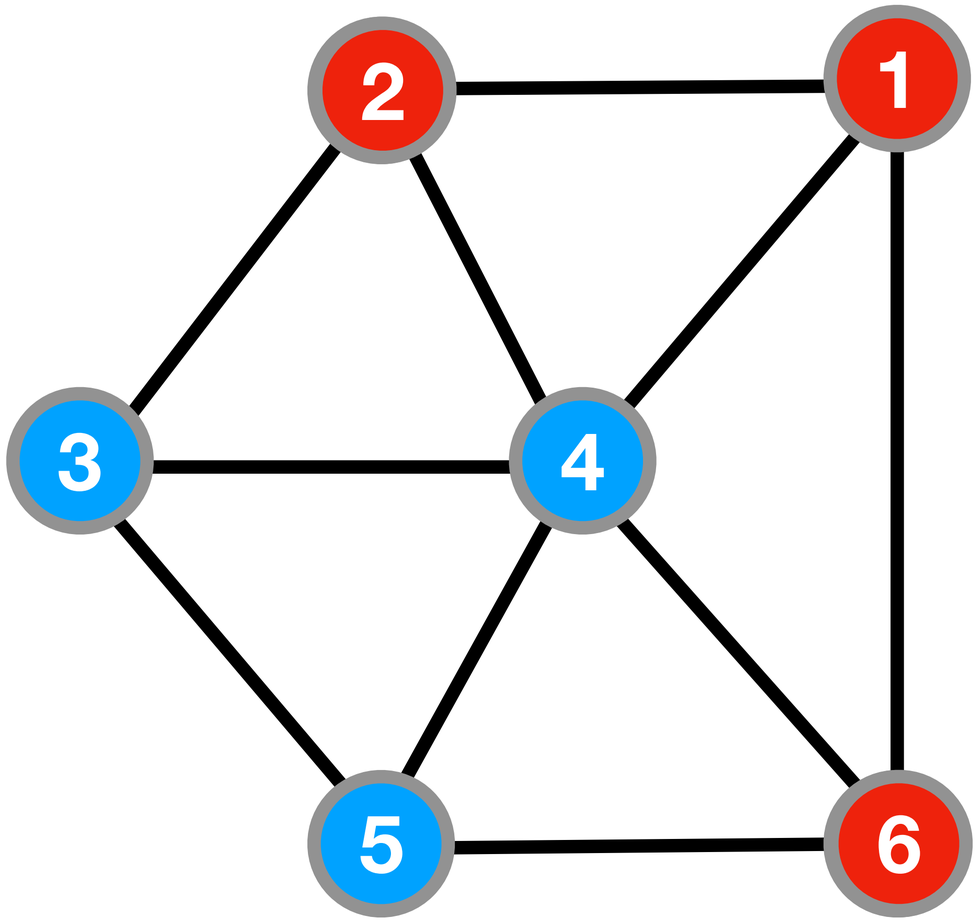
Threshold
= b/(a+b)
= 1/(1+4)
= 1/4



Example

a = 3, b = 1

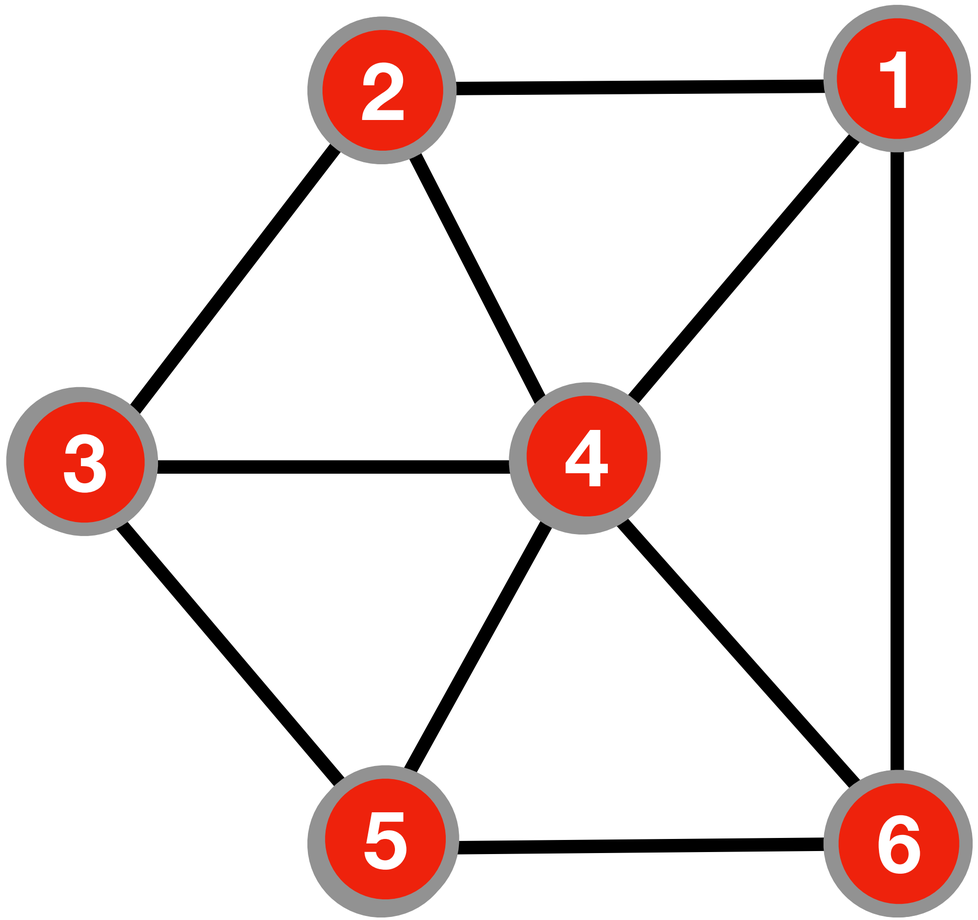
Threshold
= b/(a+b)
= 1/(1+4)
= 1/4



Example

a = 3, b = 1

Threshold
= b/(a+b)
= 1/(1+4)
= 1/4

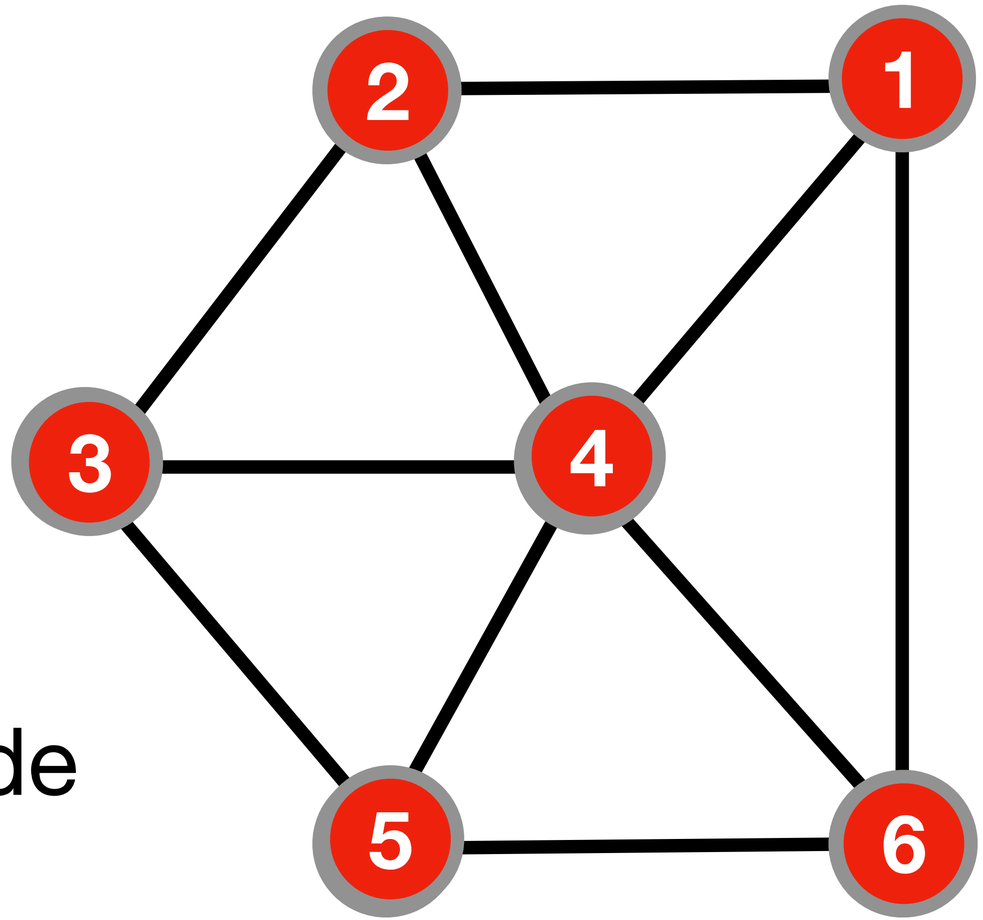


Example

a = 3, b = 1

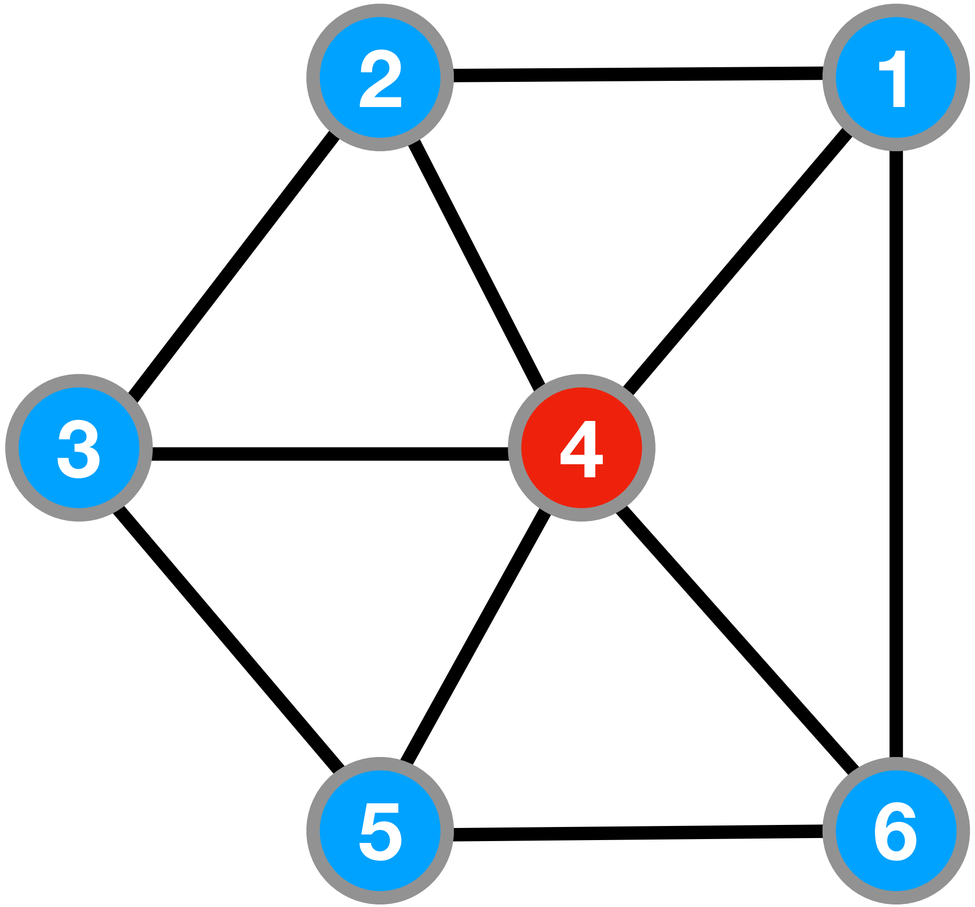
Threshold
= b/(a+b)
= 1/(1+4)
= 1/4

Leads to **complete** cascade



Example

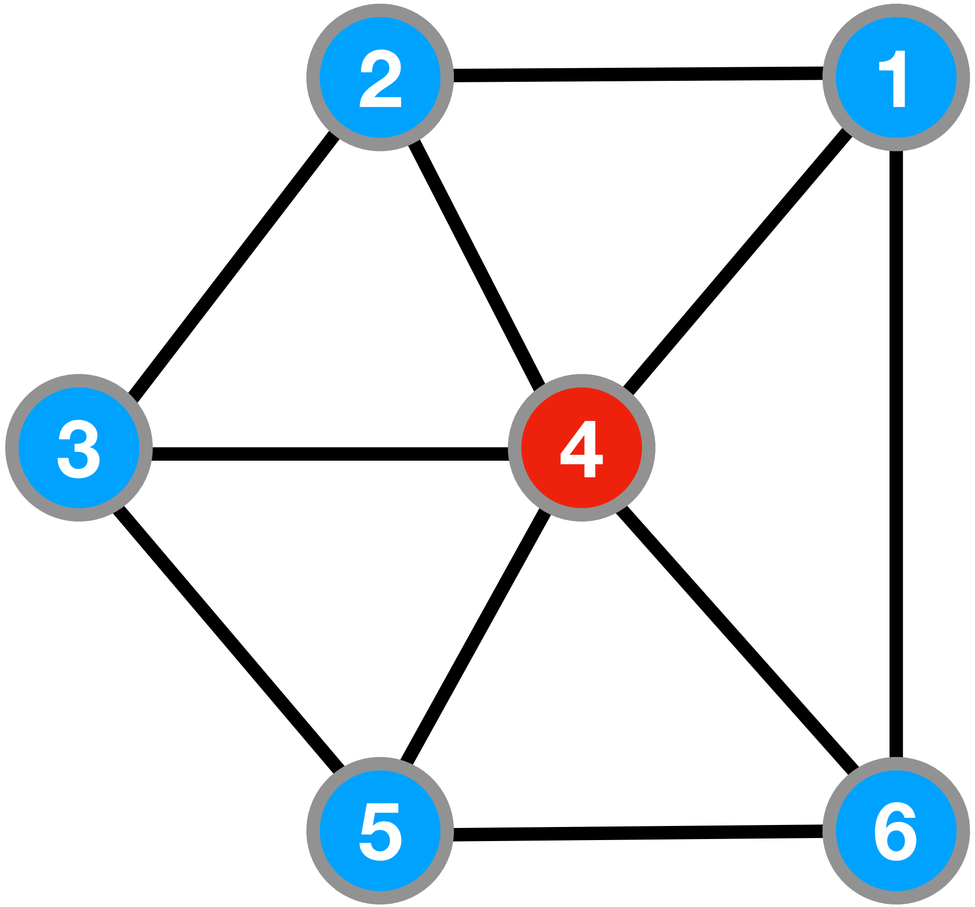
What threshold would you need to generate a cascade from this node?



Example

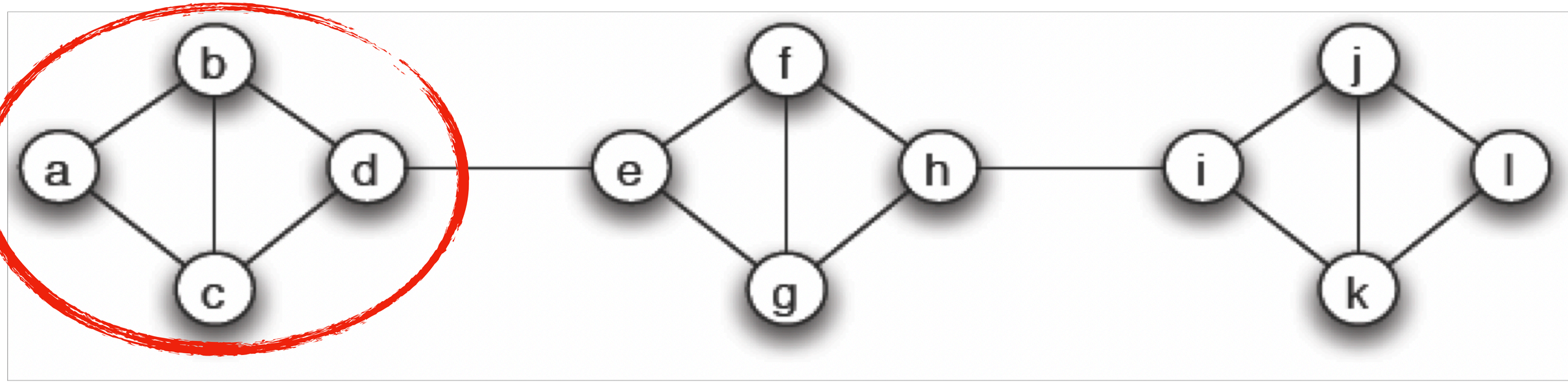
What threshold would you need to generate a cascade from this node?

Each neighbour has degree **3**, so anything lower than **1/3** will generate cascade



Effects of network structure

Weak ties between **tightly clustered** communities **stop** **cascades**



Cascade threshold would need to be **lower than 1/3**
for **e** to be convinced to **pick up new behaviour**

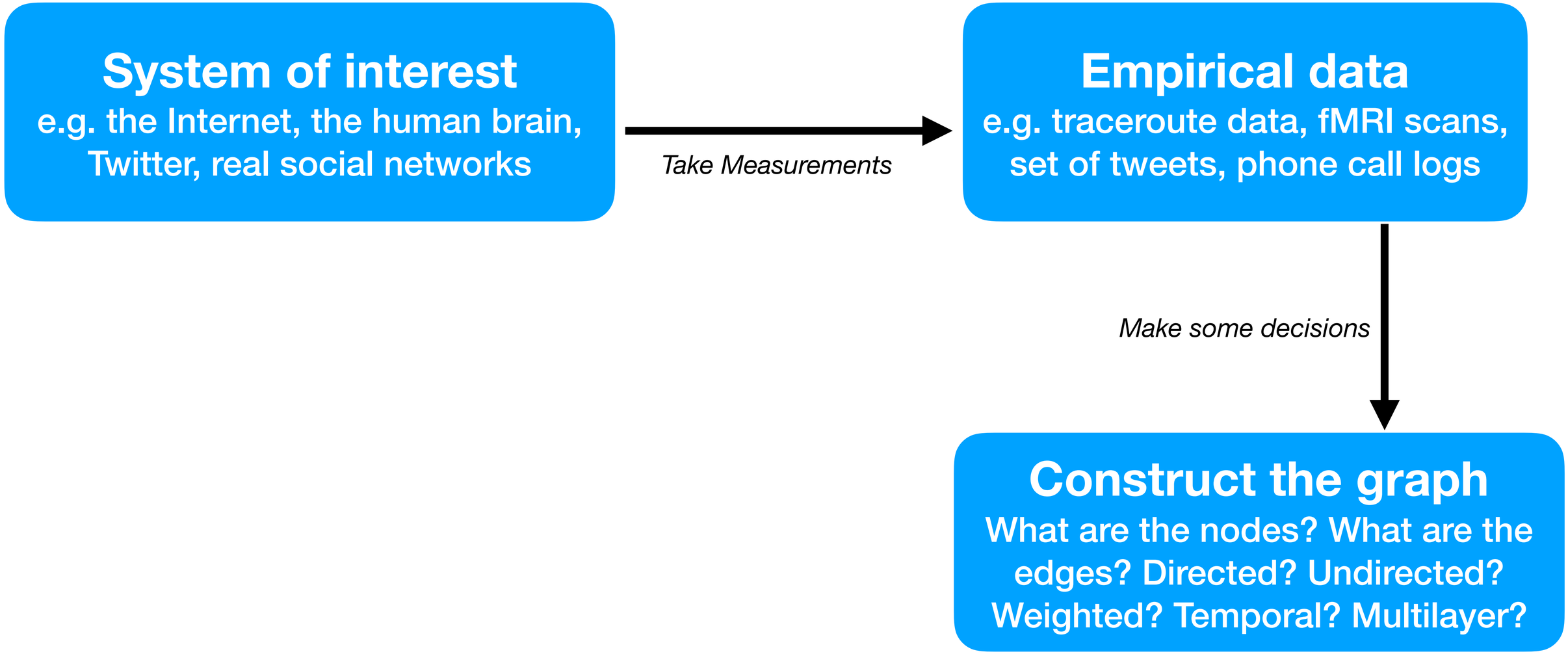
Studying systems as networks

Studying systems as networks

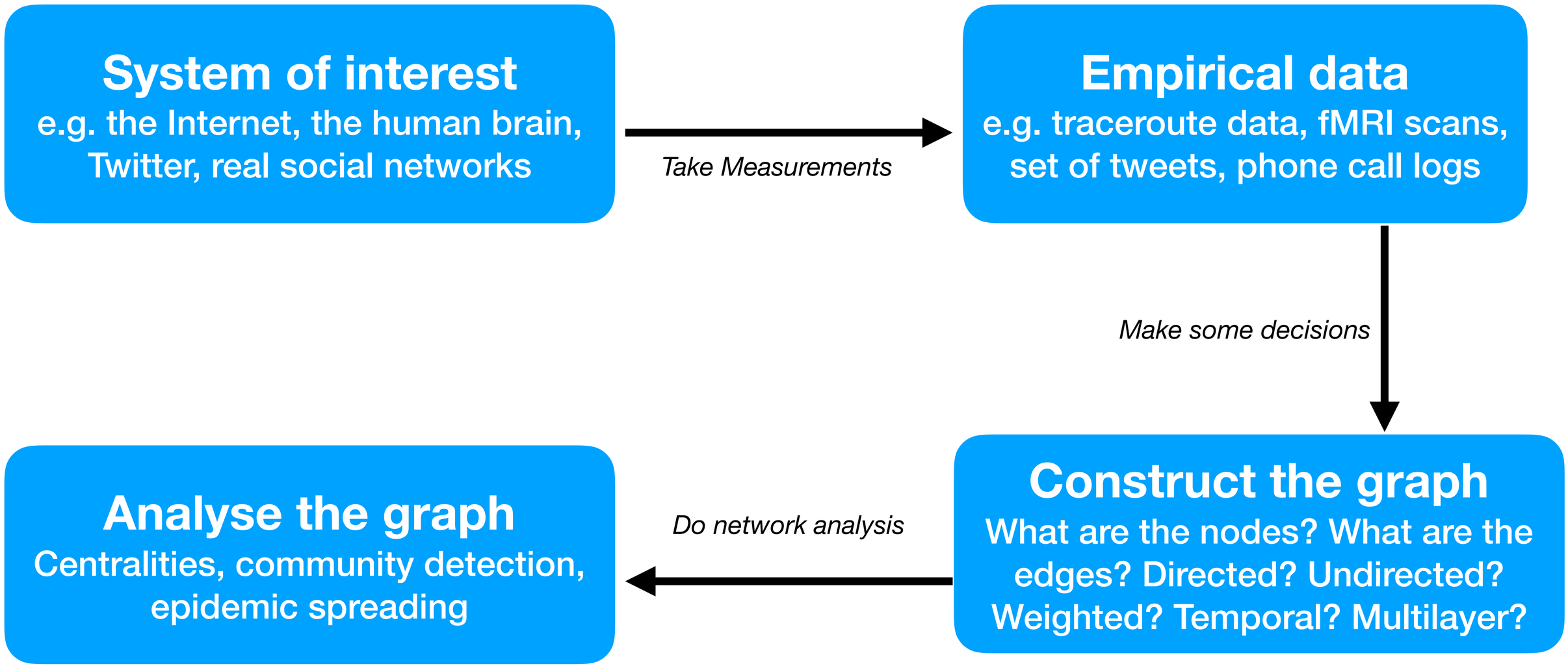
System of interest
e.g. the Internet, the human brain, Twitter, real social networks

Studying systems as networks

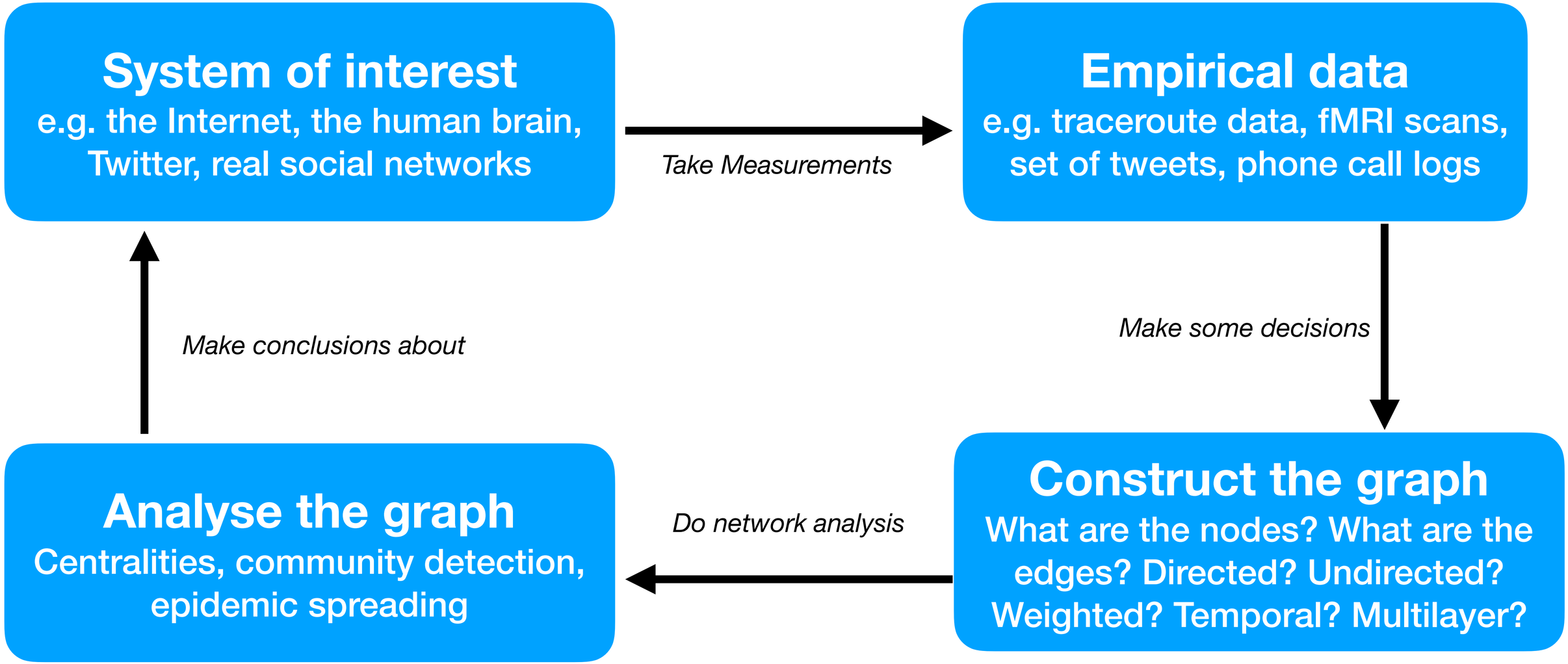
Studying systems as networks



Studying systems as networks

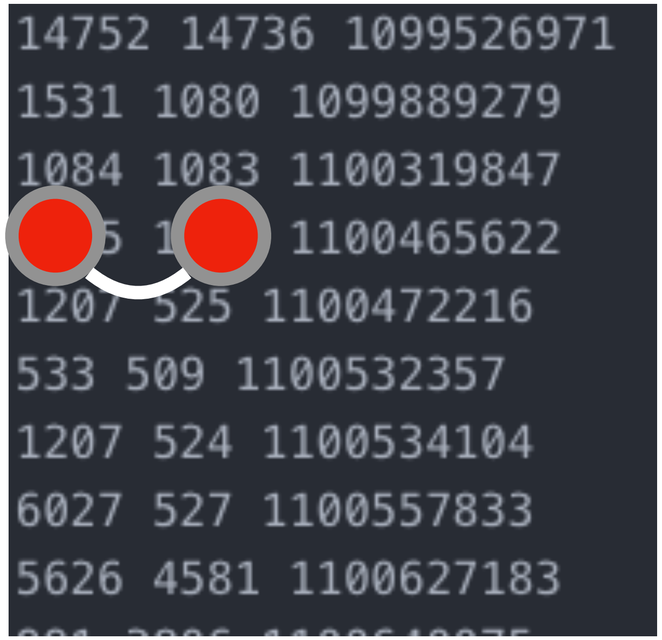


Studying systems as networks



Data Types: Edge List

- Usually a list with each row having **2 or more** comma/tab separated values
- First two values are **source** and **destination** nodes of edge
- Any extra values are **metadata** e.g. timestamp, edge weight

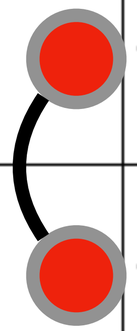


e.g. Facebook wall post dataset has user ids of wall poster and postee resp., and UNIX timestamp of when post was created

[Bimal Viswanath et al On the evolution of user interaction in Facebook. In *Proc. Workshop on Online Social Networks*, pages 37--42, 2009.]

Data types: Databases

eventName	eventSec	id	matchId	matchPeriod	playerId	positions	subEventId	subEventName	tags
Pass	1.656214	258612104	2057954	1H	122671	[{'y': 50, 'x': 50}, {'y': 53, 'x': 35}]	85	Simple pass	[{'id': 1801}]
Pass	4.487814	258612106	2057954	1H	139393	[{'y': 53, 'x': 35}, {'y': 19, 'x': 75}]	83	High pass	[{'id': 1801}]
Duel	5.937411	258612077	2057954	1H	103668	[{'y': 81, 'x': 25}, {'y': 83, 'x': 37}]	10	Air duel	[{'id': 703}, {'id': 1801}]



More complicated example: FIFA dataset where edges could be passes, specific types of passes, tackles etc.

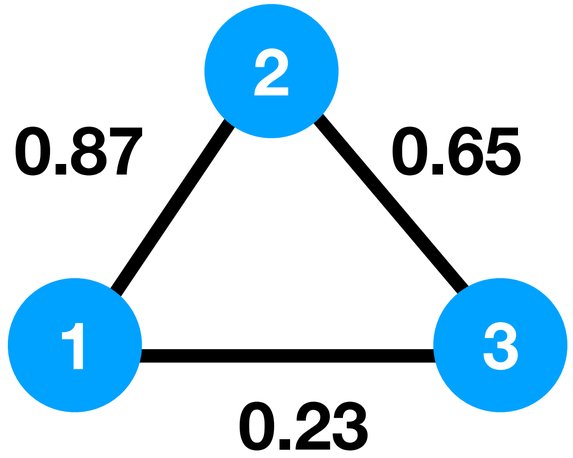
[Luca Pappalardo et al, A public data set of spatio-temporal match events in soccer competitions, 2018, Nature]

Data Types: Adjacency/Weight Matrix

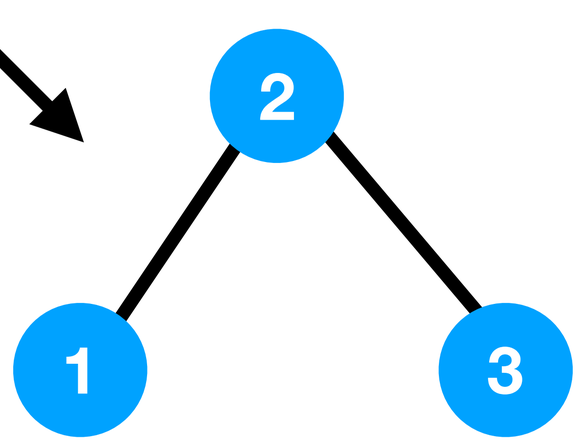
Dest Node

Source Node

	1	2	3
1	0	0.87	0.23
2	0.87	0	0.65
3	0.23	0.65	0

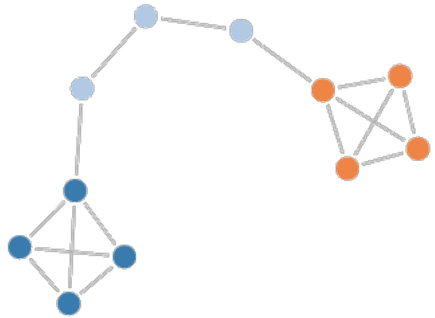


Threshold 0.6

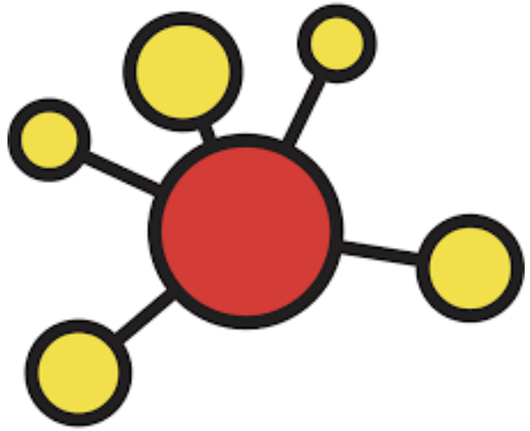


Value at **row i** and **column j** is the weight between node i and j

Software/libraries for network visualisation



NetworkX



(iGraph)

Gephi

Aaagh, my network is too big — thinning

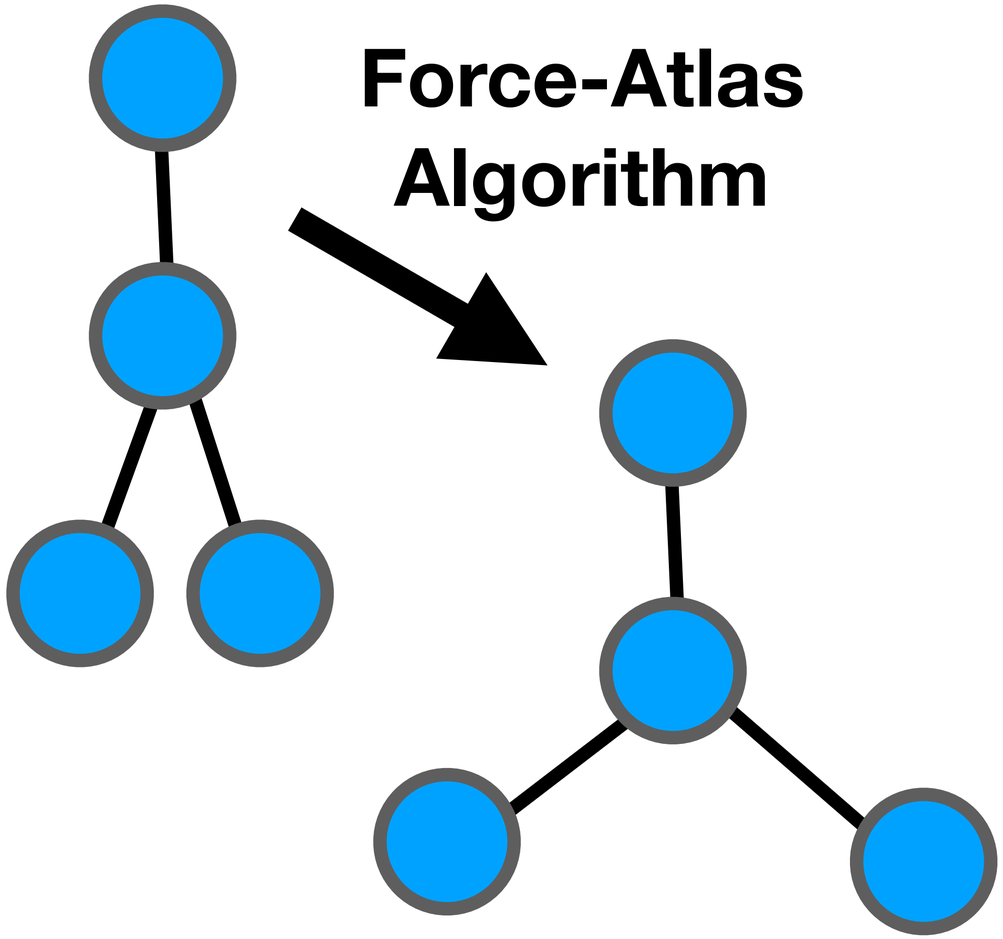
- **Windowing** — if the edges in your dataset have timestamps, look only at edges within a certain time window.
- **Random edge sample** (requires preprocessing in e.g. Python) — take a random sample of the edges in the dataset.
- **Random node sample** — take the network you get from a random sample of the nodes
- **Degree filter** — only include nodes of above a certain degree (e.g. get rid of nodes of degree 0,1)

Network visualisation ingredients

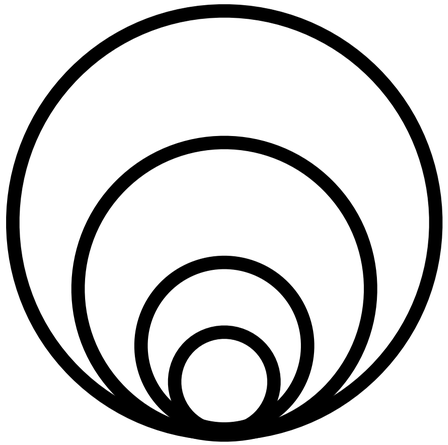
- **Node positions** — where the nodes are placed in the space?
- **Node features** — size/colour/shape?
- **Edge features** — thickness, colour?

Node Positions

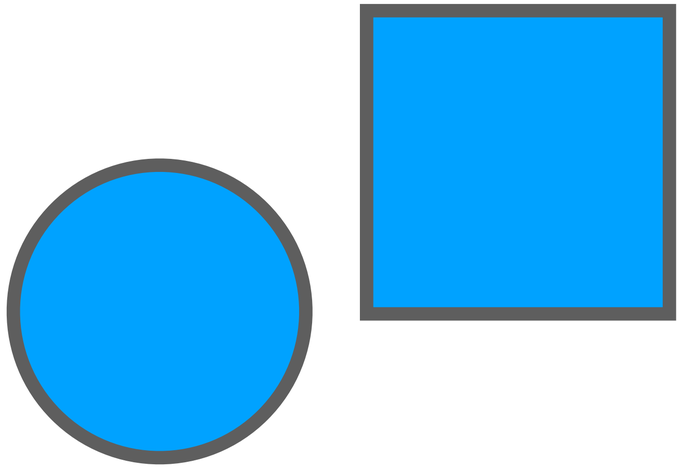
- Want to minimise **edge crossover**
- Put nodes that are **close to each other** by hops close to each other in the space
- **Layout algorithms** help with this (but are not perfect)



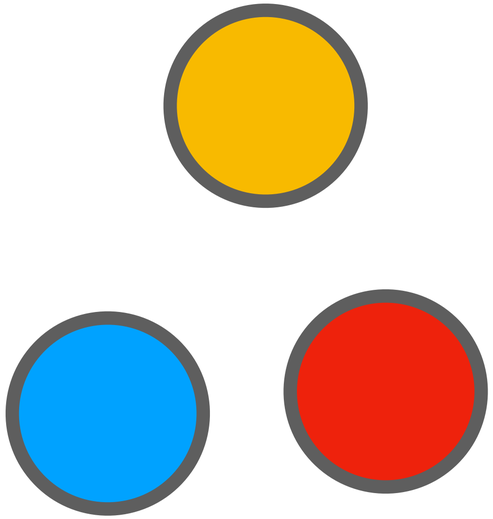
Node Properties



Size (usually some centrality measure)



Shape (different type of nodes in the graph?)



Colour (usually community-related)

Rest of tutorial: Gephi Demo

Dataset of Twitter interactions during a conference centred around the #NetSci2018 hashtag, collected by Ulf Aslak
https://gist.github.com/ulfaslak/2686ebe674b761e7947aacd2780b8384