Social and Information Networks

CSCC46H, Fall 2022
Lecture 10

Prof. Ashton Anderson
ashton@cs.toronto.edu

Logistics

Blog posts A–J due Friday, Nov 11
Blog posts K–R due Friday, Nov 18
Blog posts S–Z due Friday, Nov 25

Today

A3 due today
A4 out tonight, due Weds, Nov 30

Today

Information Diffusion
Contagion & Epidemics

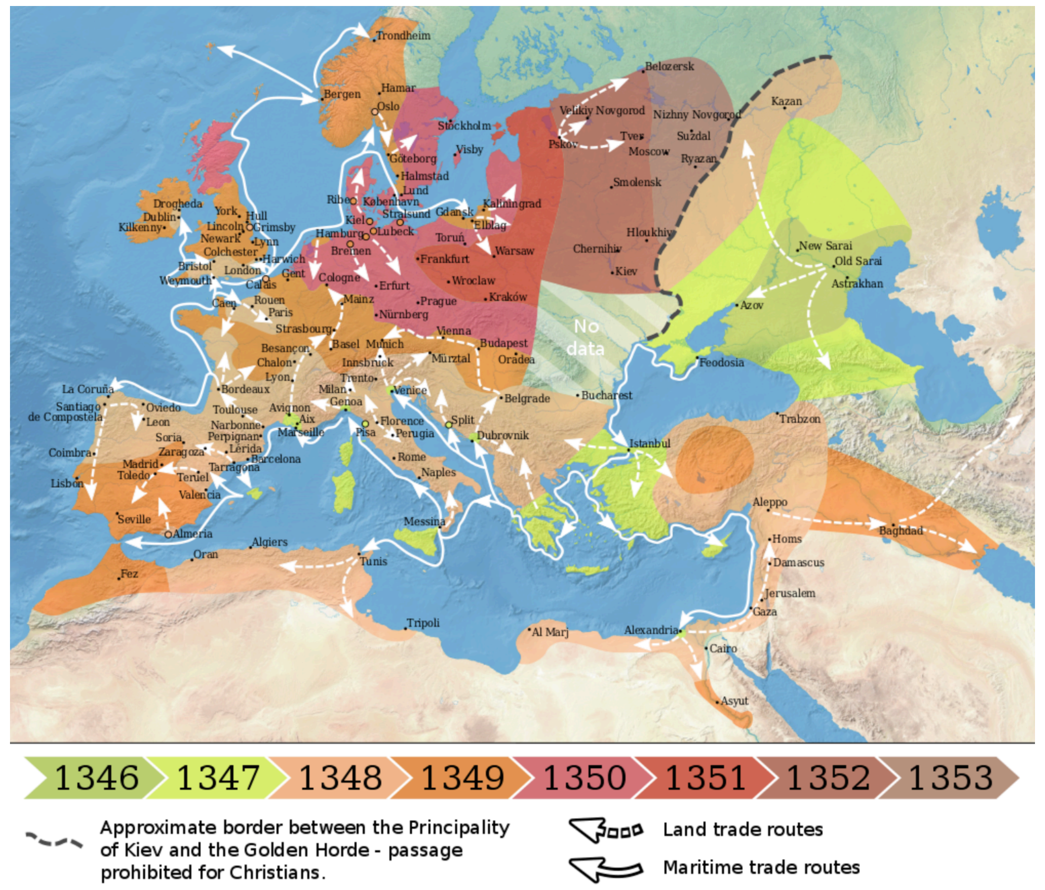
How Things Spread

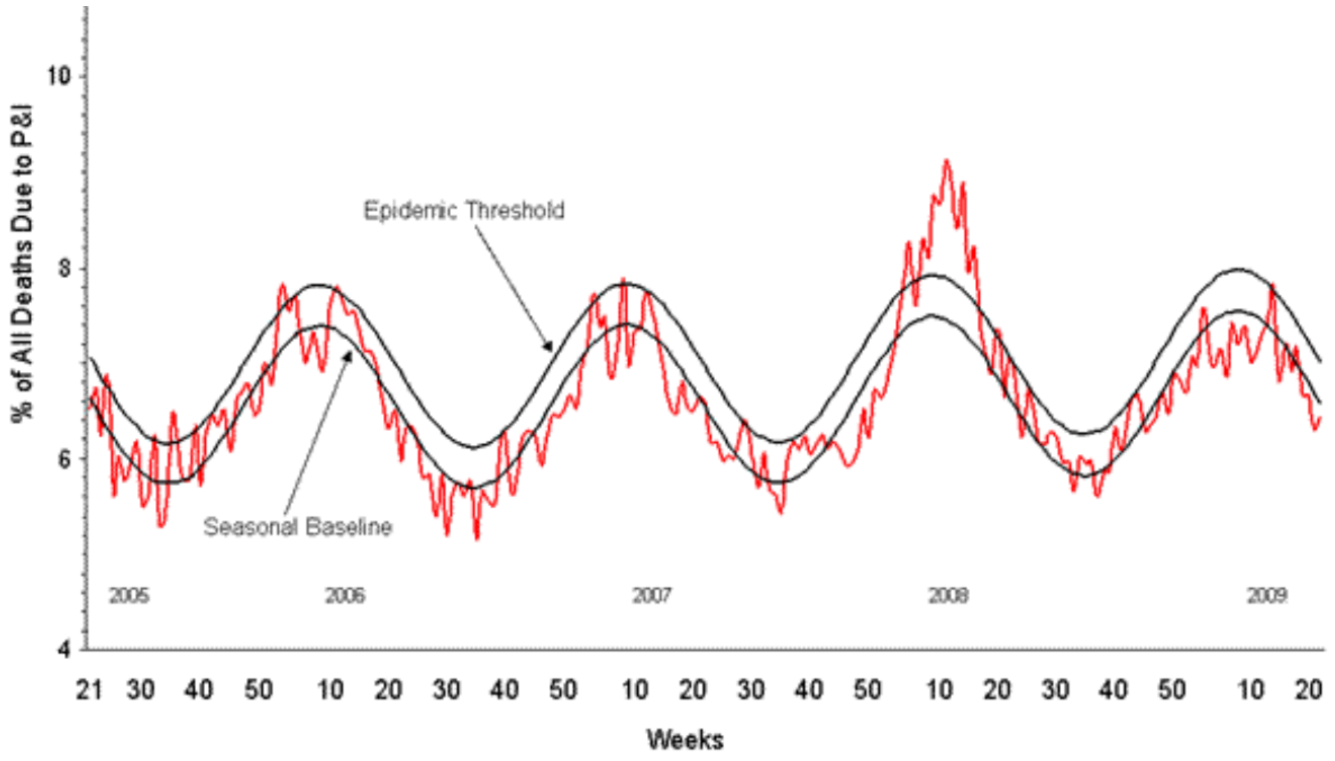
Networks define how behaviours, ideas, beliefs, diseases, etc. spread

Last class: behaviour (adoption of an innovation or technology)

Today:

- Information
- Diseases



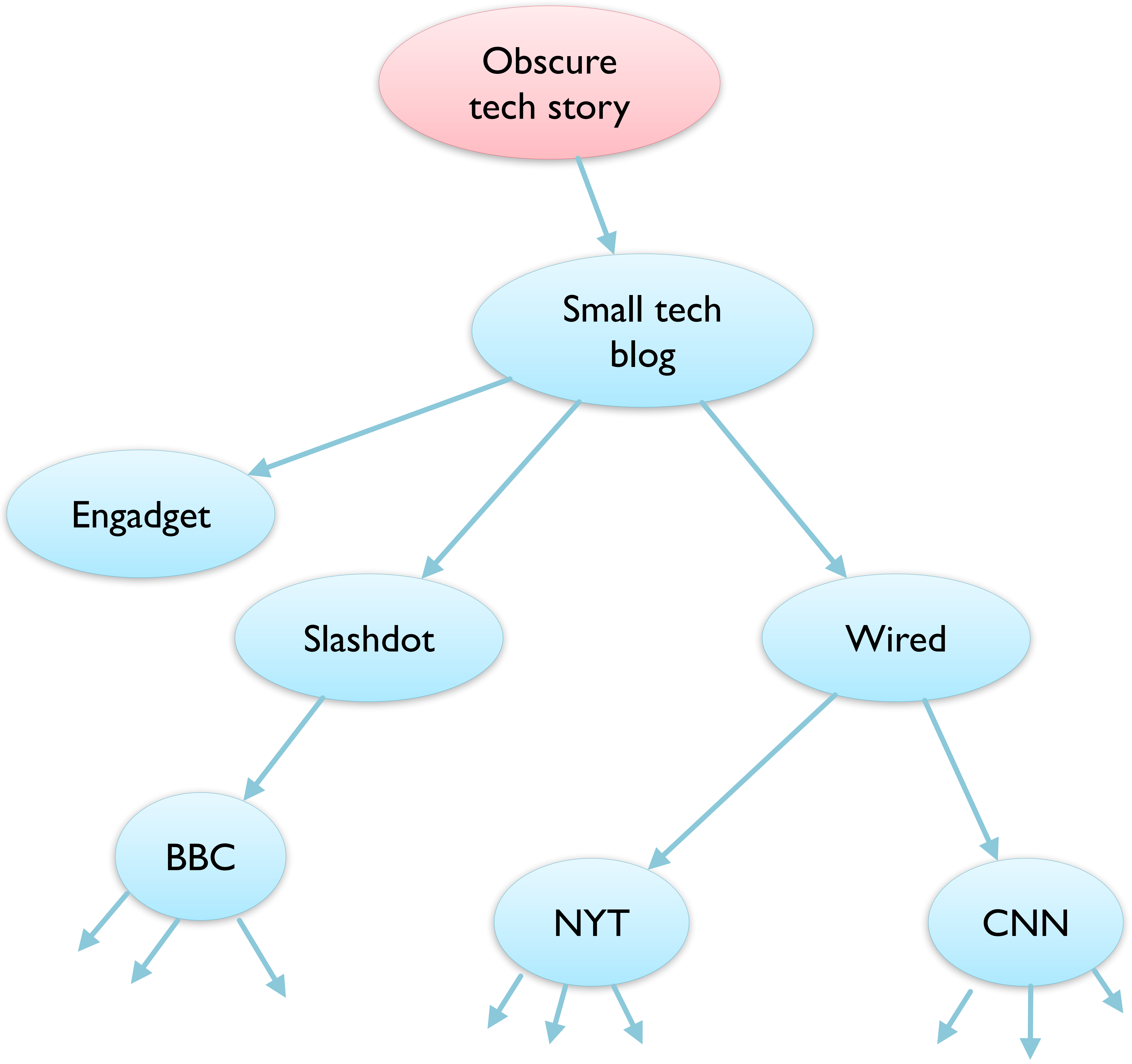


Information Diffusion

Influence Through Networks

- If people are connected through a network, it's possible for them to influence each other's behaviour and actions
- Today: why?
  - Direct benefit
  - Informational
  - Social conformity

Information Diffusion: Media



Barack Obama
@BarackObama

"No one is born hating another person because of the color of his skin or his background or his religion..."

8:06 PM · Aug 12, 2017 · Twitter for iPhone

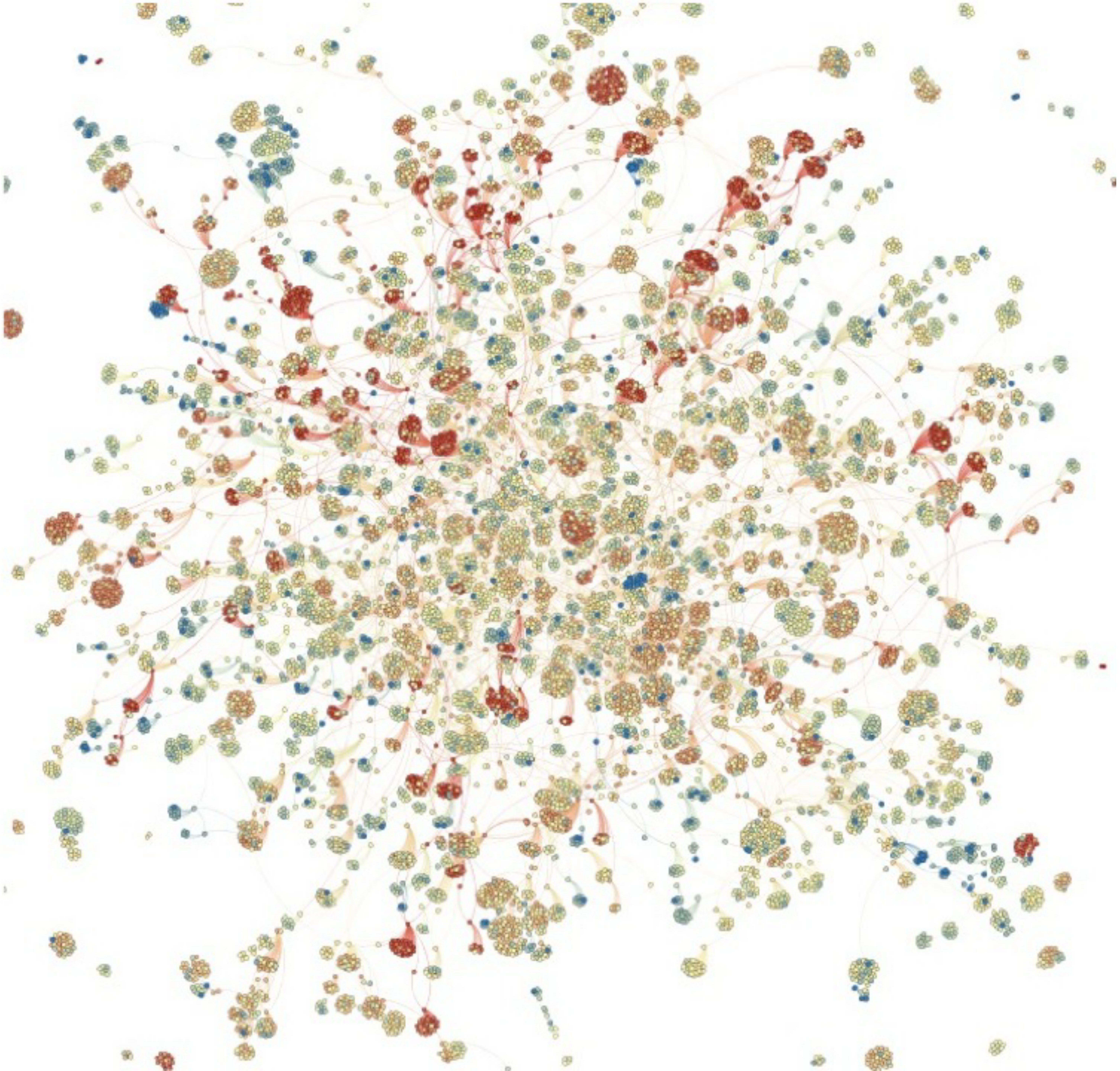
1.4M Retweets **73.1K** Quote Tweets **4M** Likes

Simple Herding Model: Lessons

erictucker @erictucker · Nov 9

Anti-Trump protestors in Austin today are not as organic as they seem. Here are the busses they came in. #fakeprotests #trump2016 #austin

16K 14K



Information-Based Model of Diffusion: Crowd Herding

People influencing each other

Almost infinite number of ways:

- Opinions
- Product purchases
- Political positions
- Technologies used
- etc…

Good reasons for this! Sometimes it's better to follow the crowd than trust your information

A simple example

Going to Yellowknife

Do some research, intend to eat at resto A

But you show up and no one's eating there, instead lots of people are in resto B!

A rational person may reason that those people know something he doesn't, and go with B as well

Sequential decision making

"Information cascade"

Imitation

In this example, people imitate others, but it's not mindless

Kinds of imitation/influence: informational, social pressure to conform, direct benefits

Sometimes hard to tell apart

Another example: social pressure or informational?

Experiment: bunch of people stand on a street corner and stare up into the sky
What fraction of passersby stop and look up?

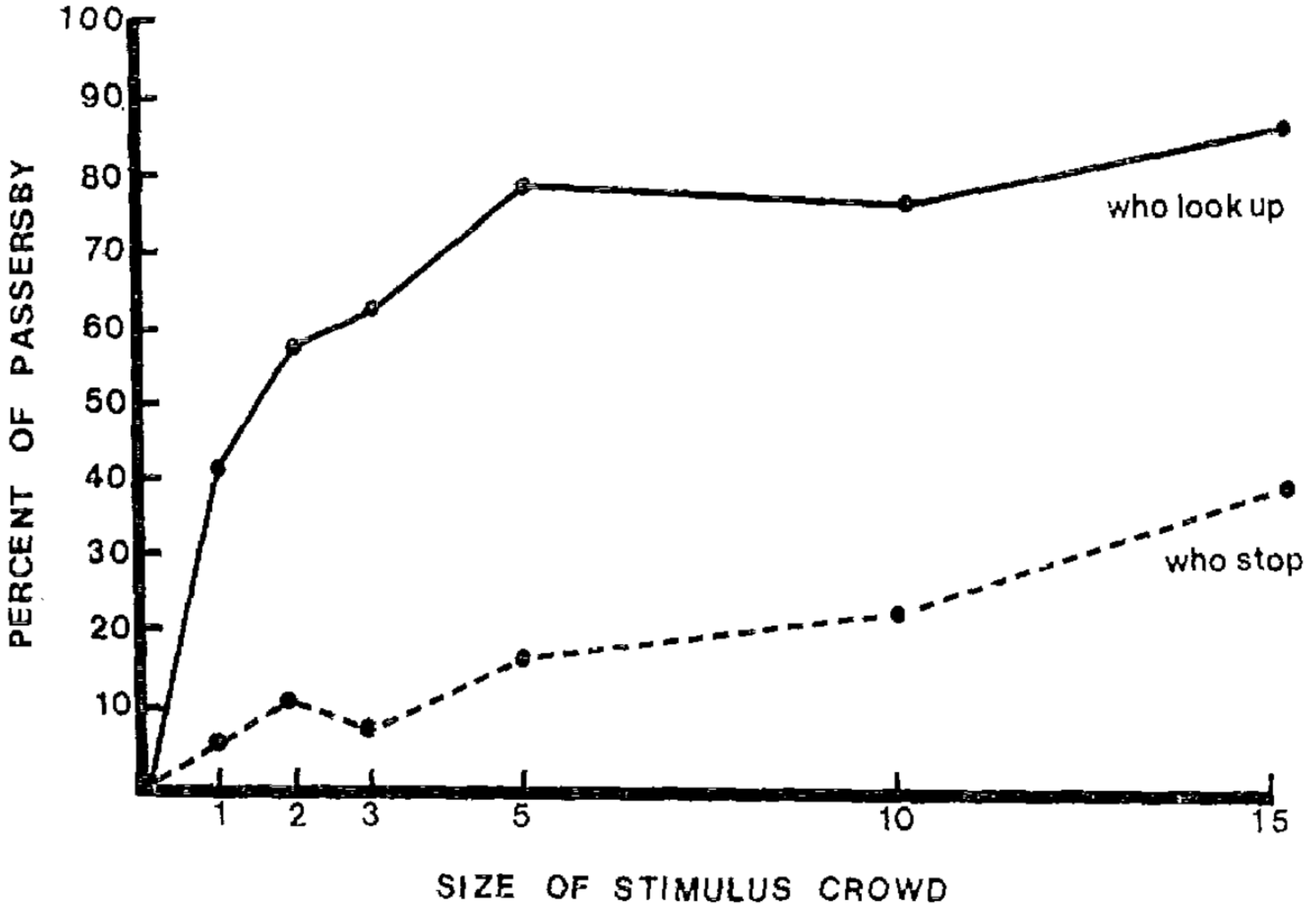


FIG. 1. Mean percentage of passersby who look up and who stop, as a function of the size of the stimulus crowd.

Another example: direct benefits

Joining Instagram
If no one else is on it, useless
But if lots of your friends are on it, helpful
Or phone brands, or WhatsApp, or gaming consoles, etc…

Simple Herding Model

Decision to be made (resto choice, adopt a new technology, support political position, etc)

People decide sequentially, and see all choices of those who acted earlier

Each person has some **private information** that can help guide their decision

People **can't** directly observe what others **know**, but **can** observe what they **do**

Simple Herding Model

Model: n students in a classroom, urn in front
Two urns with marbles:

- "Majority-blue" urn has 2/3 blue, 1/3 red
- "Majority-red" urn has 2/3 red, 1/3 blue
- 50%/50% chance that the urn is majority blue/red

One by one, each student privately gets to look at 1 marble, put it back without showing anyone else, and guess if the urn is Majority-blue or Majority-red

Simple Herding Model

Student 1: Just guess the colour she sees

Student 2:

If same as first person, guess that colour.

But if different from first, then since he knows first guess was what first person saw, then he's indifferent between the two. Guess what he saw

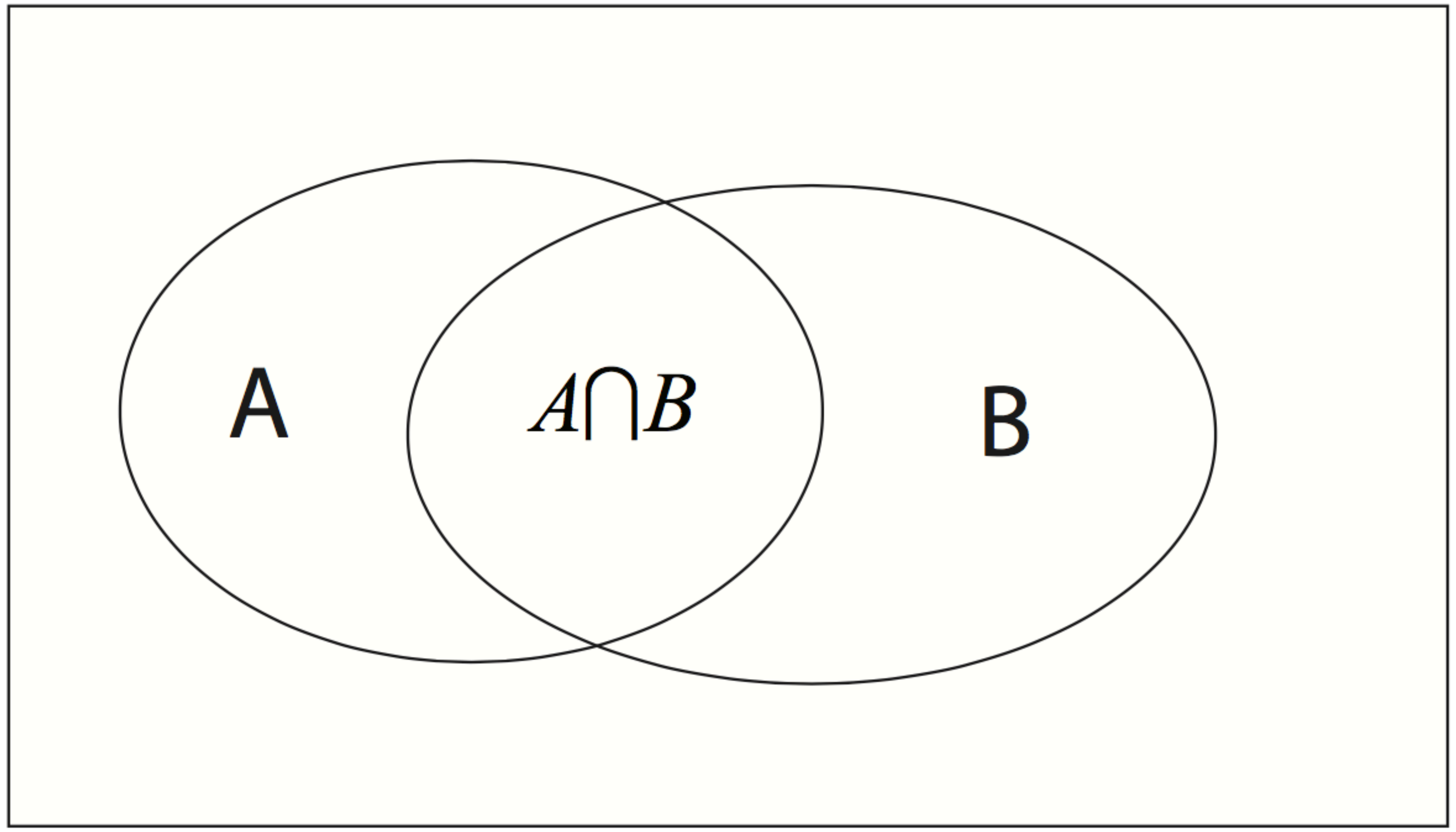
Student 3:

If first 2 are opposite colours, guess what she sees (tiebreaker)

If previous 2 are the same colour (blue) and S3 draws red, then it's like he has drawn three times and gotten two blue, so she should guess majority-blue, **despite her own private information!**

Bayes' Rule

P[A|B] = P[A and B] / P[B]
P[A|B] = P[B|A]P[A] / P[B]
Posterior = Update * Prior



A Student’s Decision

Say you’re one of the students. You go to the urn and pick a marble, say it’s blue.

What should you do?

A Student's Decision

Say you're one of the students. You go to the urn and pick a marble, say it's blue.

What should you do?

Don't just naively guess blue… you've heard a lot of information too! (what if everyone else said red?)

Guess blue if given you what you know **AND** the information you have from others leads you to believe the urn is majority-blue

Simple Herding Model

Student guesses blue if P[majority-blue | what she has seen/heard] > 1/2, red otherwise

Prior: P[majority-blue] = P[majority-red] = 1/2

And because of the marbles in the urns:

P[blue | majority-blue] = P[red | majority-red] = 2/3

Student 1: say she picks blue marble

P[maj-blue | blue] = P[maj-blue]*P[blue | maj-blue] / P[blue]

P[blue] = P[blue | maj-blue]*P[maj-blue] + P[blue | maj-red]*P[maj-red]

= (2/3)(1/2) + (1/3)(1/2) = 1/2

So P[maj-blue | blue] = (1/3)/(1/2) = 2/3

Simple Herding Model

Student 2 same as Student 1 (it's rational to guess what you see), so consider Student 3
Student 3 can reason that first two guesses are what the students actually saw (rationality)
Say she sees different from first two guesses: blue blue red

P[maj-blue | blue blue red]?

= P[maj-blue]P[blue blue red | maj-blue] / P[BBR]
= P[BBR | maj-blue] = (2/3)(2/3)(1/3) = 4/27

P[BBR] = P[BBR|maj-blue]P[maj-blue] + P[BBR|maj-red]P[maj-red]
= (2/3)(2/3)(1/3)(1/2) + (1/3)(1/3)(2/3)(1/2) = 1/9

Plug it all in: 2/3
Student 3 ignores what she sees and goes with what she heard before => **information cascade**

Same for all subsequent students!

Simple Herding Model: Lessons

Cascades can be wrong
Cascades can be based on very little information
Cascades are fragile

Be careful in drawing conclusions from the behaviour of a crowd: we just saw that the crowd can be wrong even if every individual is perfectly rational and takes the same action!

Simple Herding Model: Lessons

erictucker @erictucker · Nov 9

Anti-Trump protestors in Austin today are not as organic as they seem. Here are the busses they came in. #fakeprotests #trump2016 #austin

16K 14K

The Spread of Information

Friends tell their friends stuff

- Rumours/secrets
- Useful information (not homework answers though)
- Beliefs, hopes, desires, fears, …

Social media built to support this:

- Blogs (personal/professional)
- Social networks (Facebook)
- Microblogging (Twitter)

What is the structure of how information spreads?

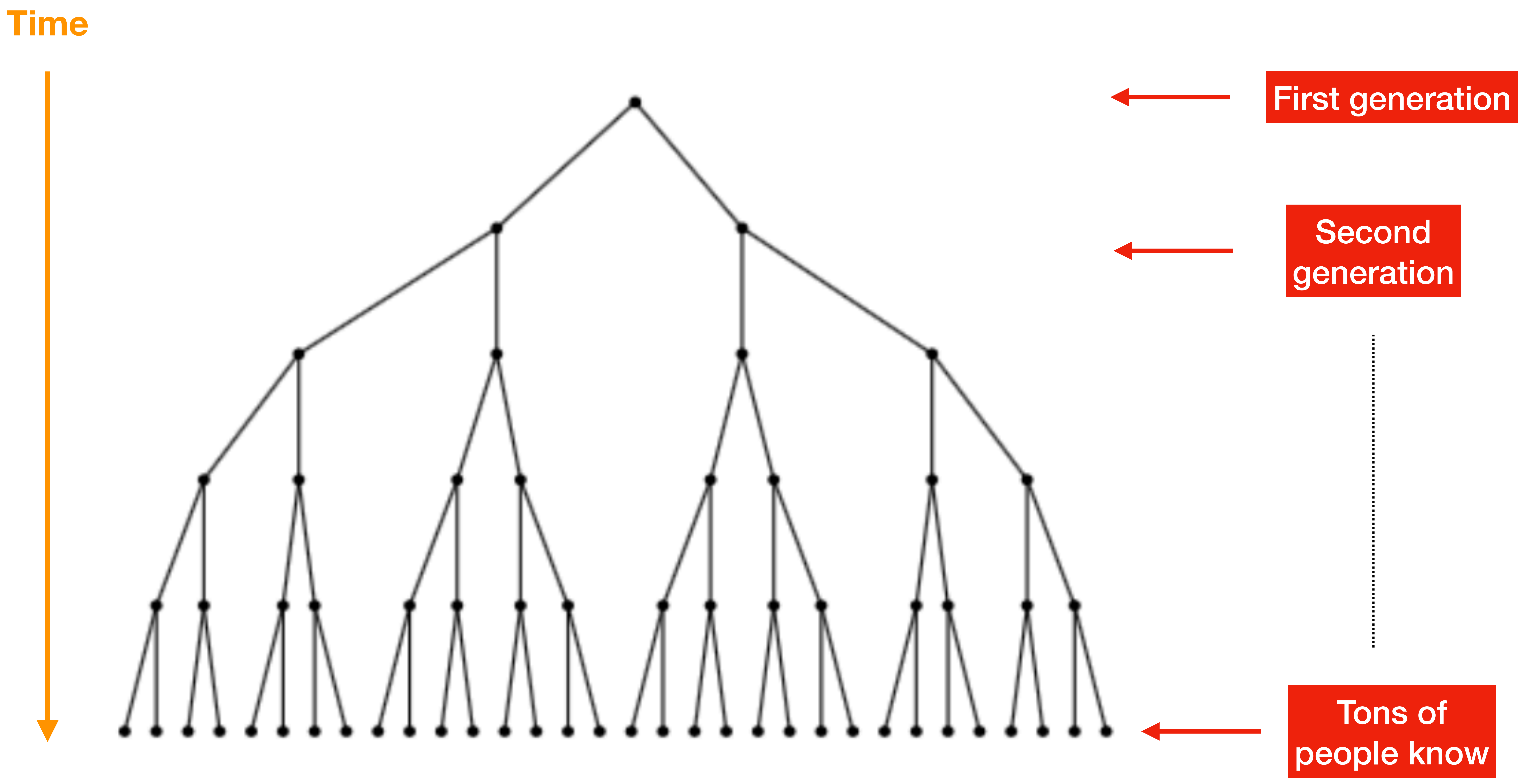
What does "go viral" mean?

People say stuff goes viral

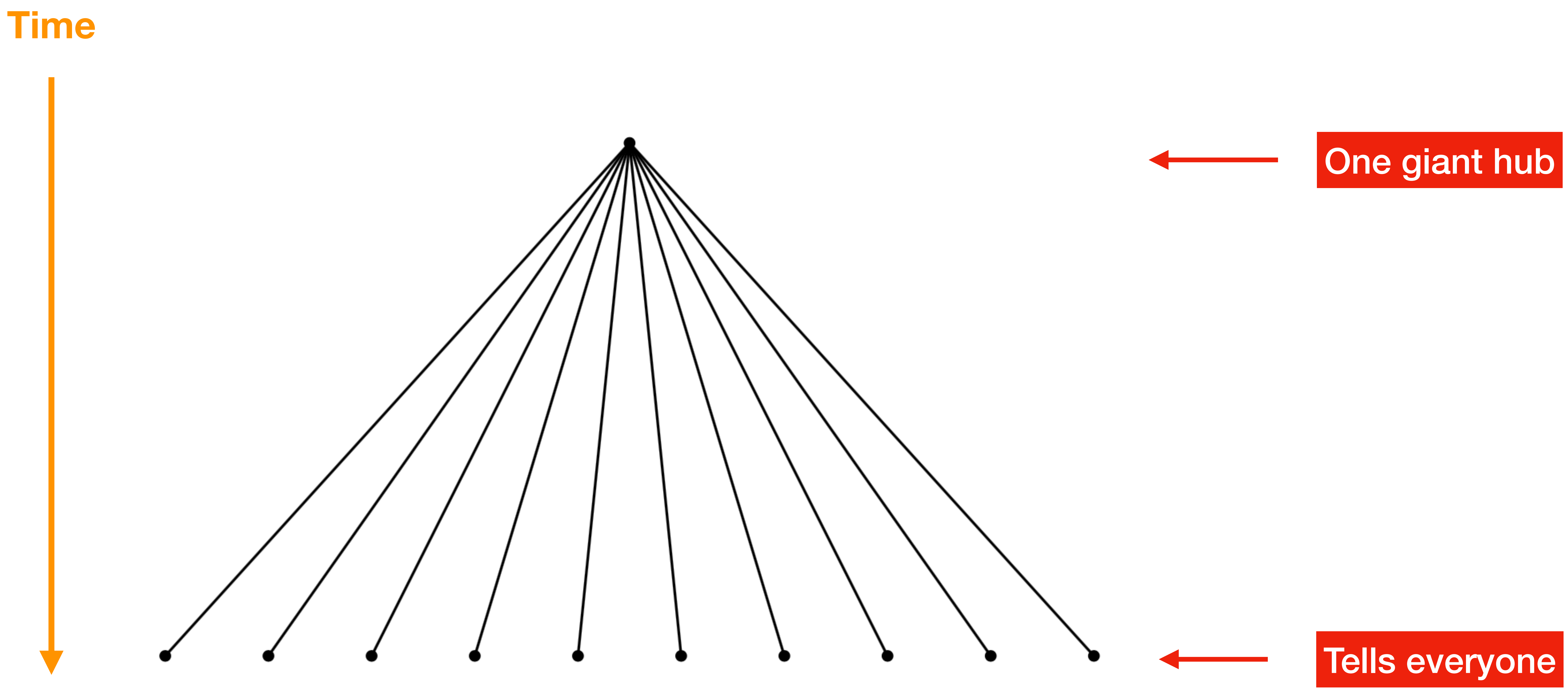
- Person-to-person transmission
- Deep branching structures
- Hypothesis: an idea, story, joke, etc. spreads like a **virus**, "infecting" minds like viruses infect the body

This implies a certain kind of structure!

What does “go viral” mean?

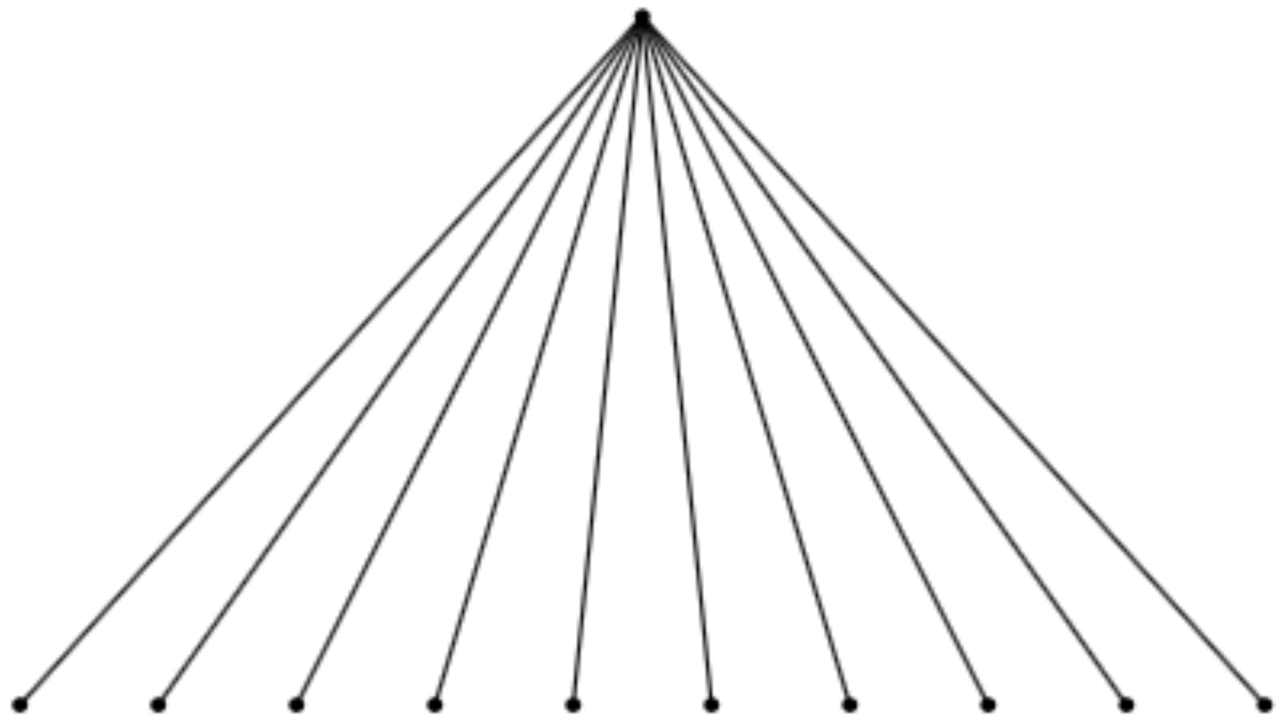


But another way

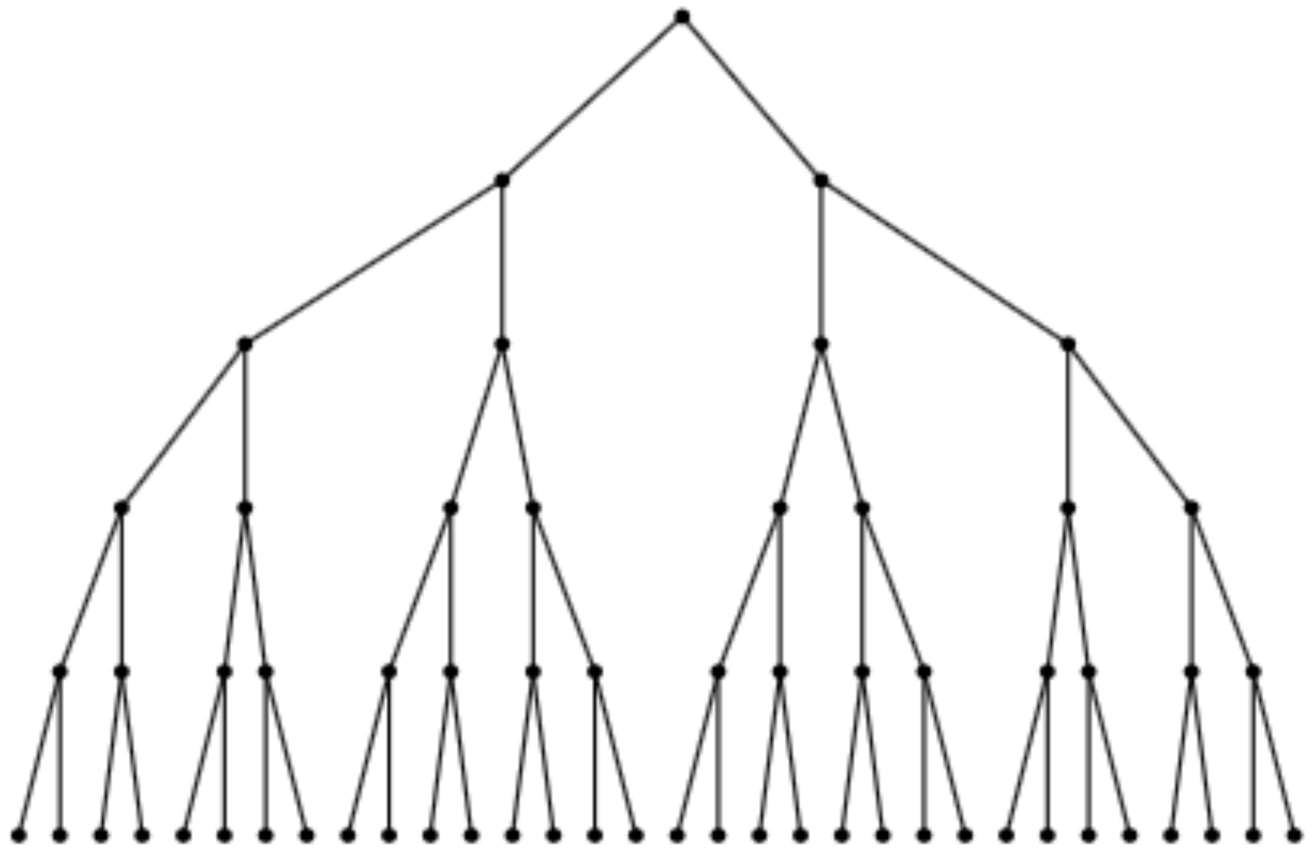


Which is it?

Big media (CNN, BBC, NYT, Fox)
Celebrities (Biebs, Taylor Swift)



or



"Broadcast"

"Viral"

- **Organically spreading content**
- **Chain letters**

How to study information spread?

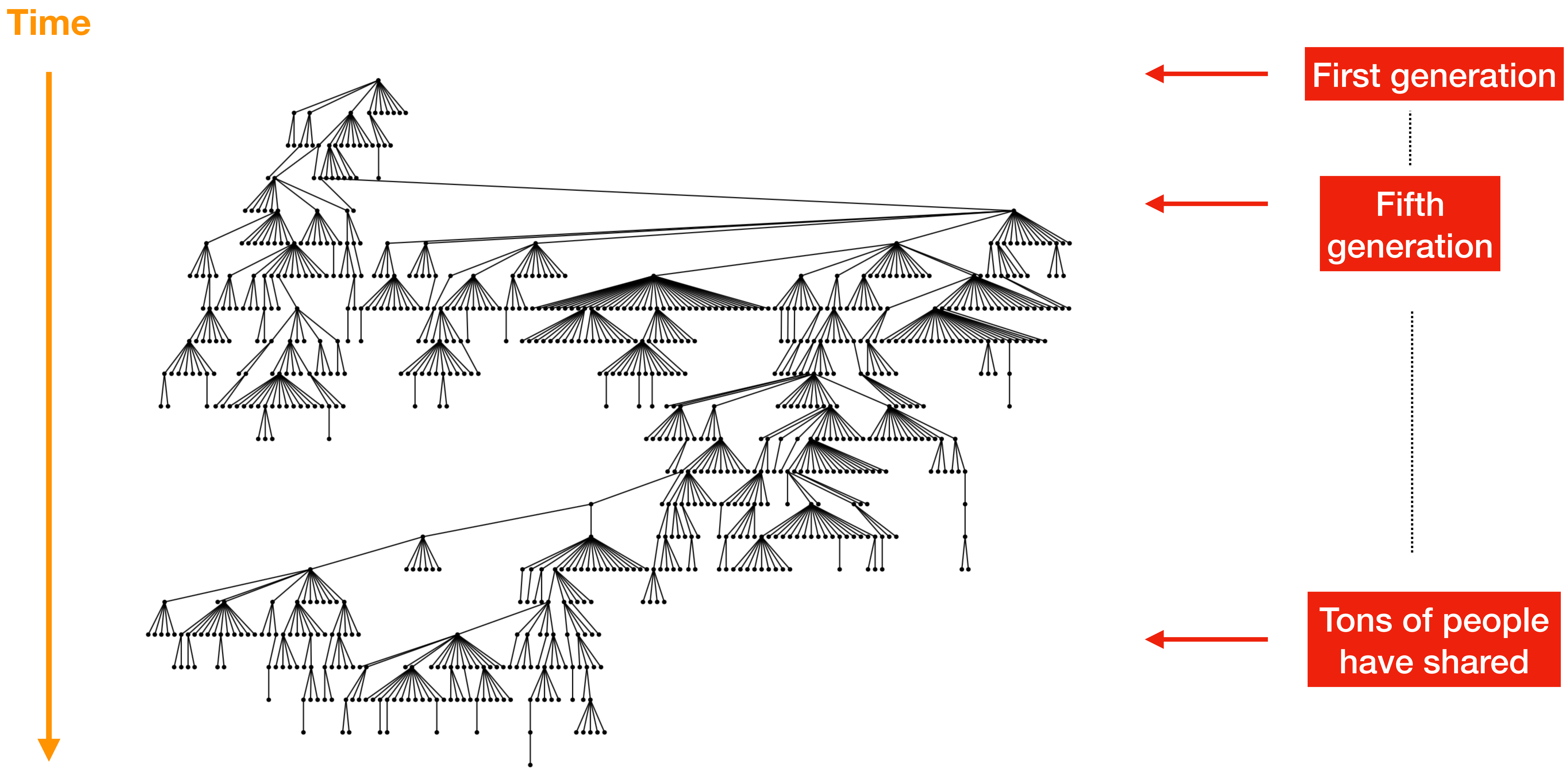
Hard to track "information" spreading from one mind to another

Online proxy: people sharing URLs

Twitter: person A tweets a URL, then a friend B tweets it (or directly retweets)
We say the URL passed from A to B

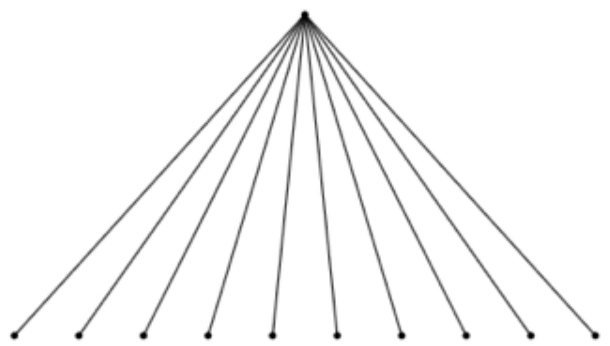
How to study information spread?

Connect these sharing edges into **trees**



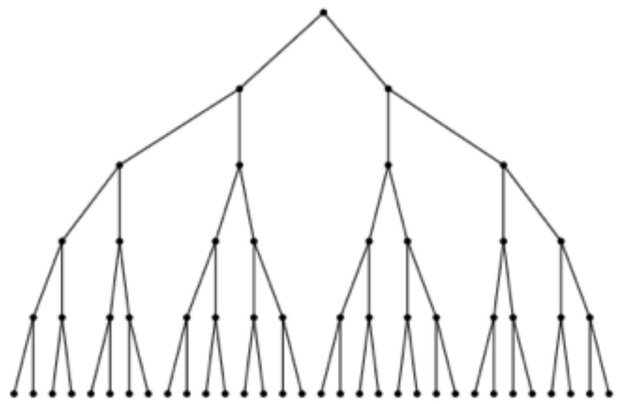
How to measure virality?

How **structurally viral** is a particular cascade?

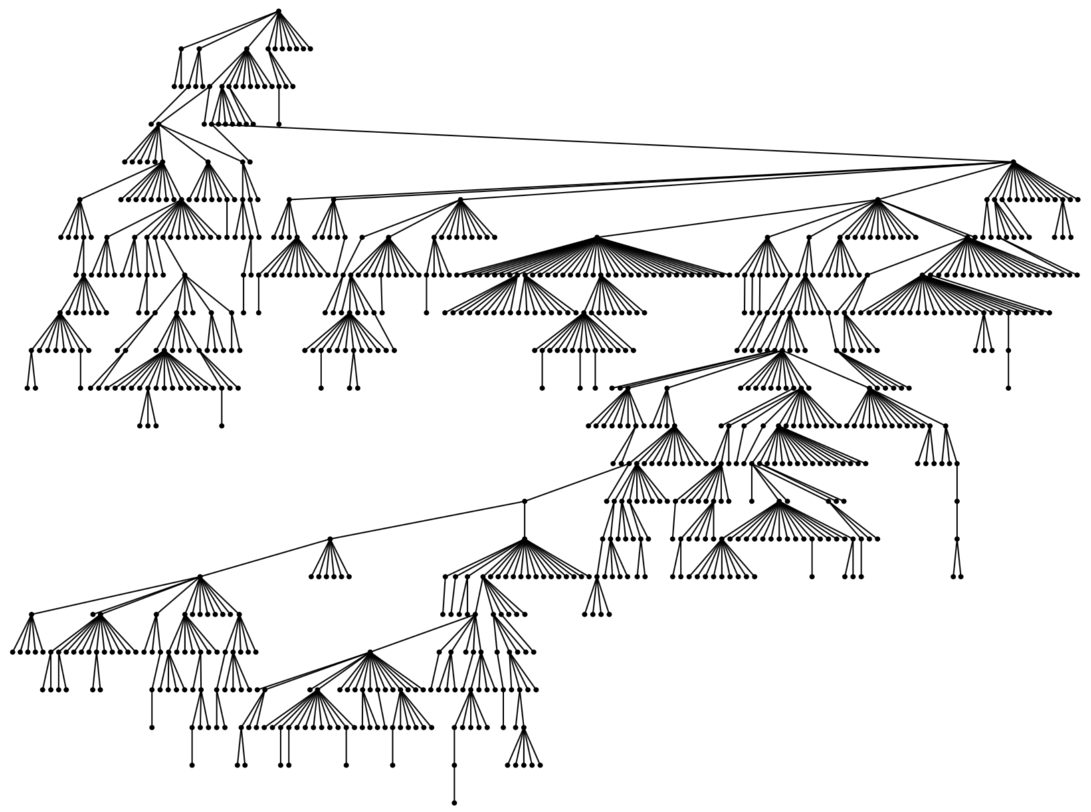


Not viral

?



Super viral



How to measure virality?

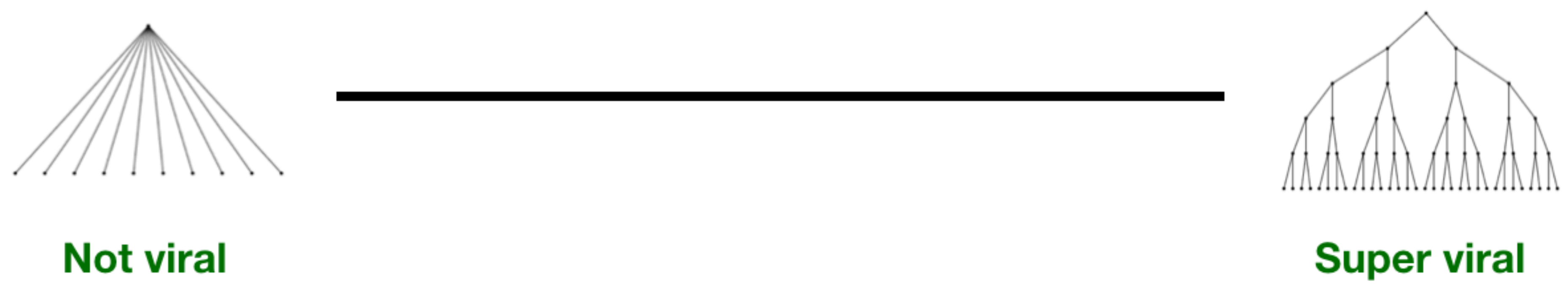
One idea: **depth of the cascade**

But this is **sensitive to a single long chain**

How to measure virality?

Another idea: **average depth of the cascade**

But even this **sometimes fails**: long chain then a big broadcast



How to measure virality?

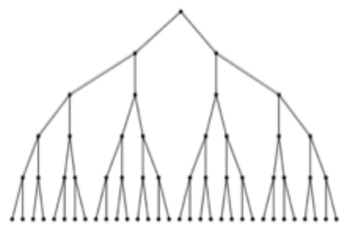
Solution: **average path length between nodes**

$$\nu(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}$$

Simple average!

Originally studied in mathematical chemistry [Wiener 1947] => “Wiener index”

Not viral



Super viral

Measure virality in data!

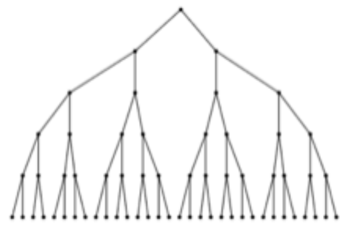
Now we have a way to construct information cascades on Twitter

And for each cascade we can compute a number that determines how "structurally viral" it is

So **how often does stuff go viral?**

Not viral

$$\nu(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}$$



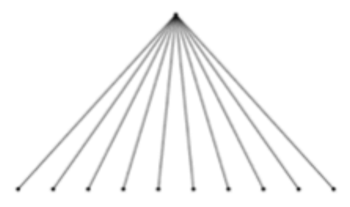
Super viral

Measure virality in data!

Looked at an **entire year of Twitter data**

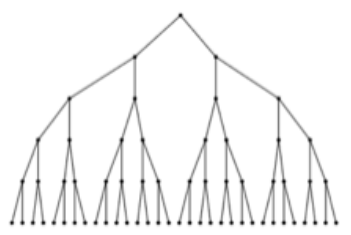
622 million unique URLs, 1.2 billion "adoptions" (tweets) of these URLs

Every URL is associated with a forest of trees

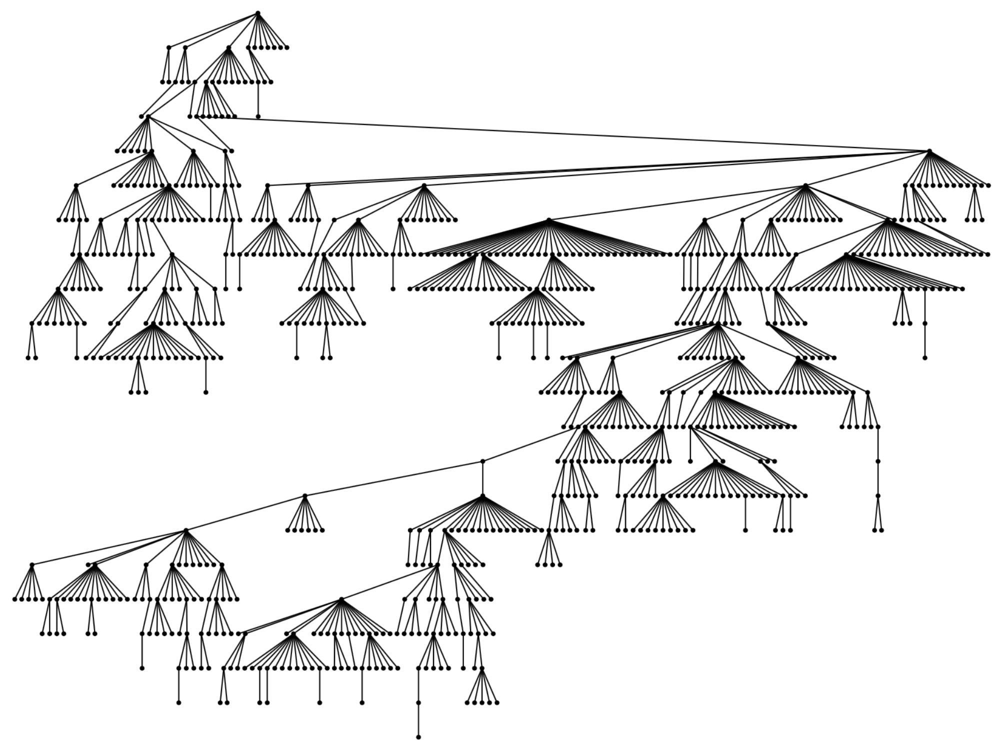


Not viral

$$\nu(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}$$



Super viral

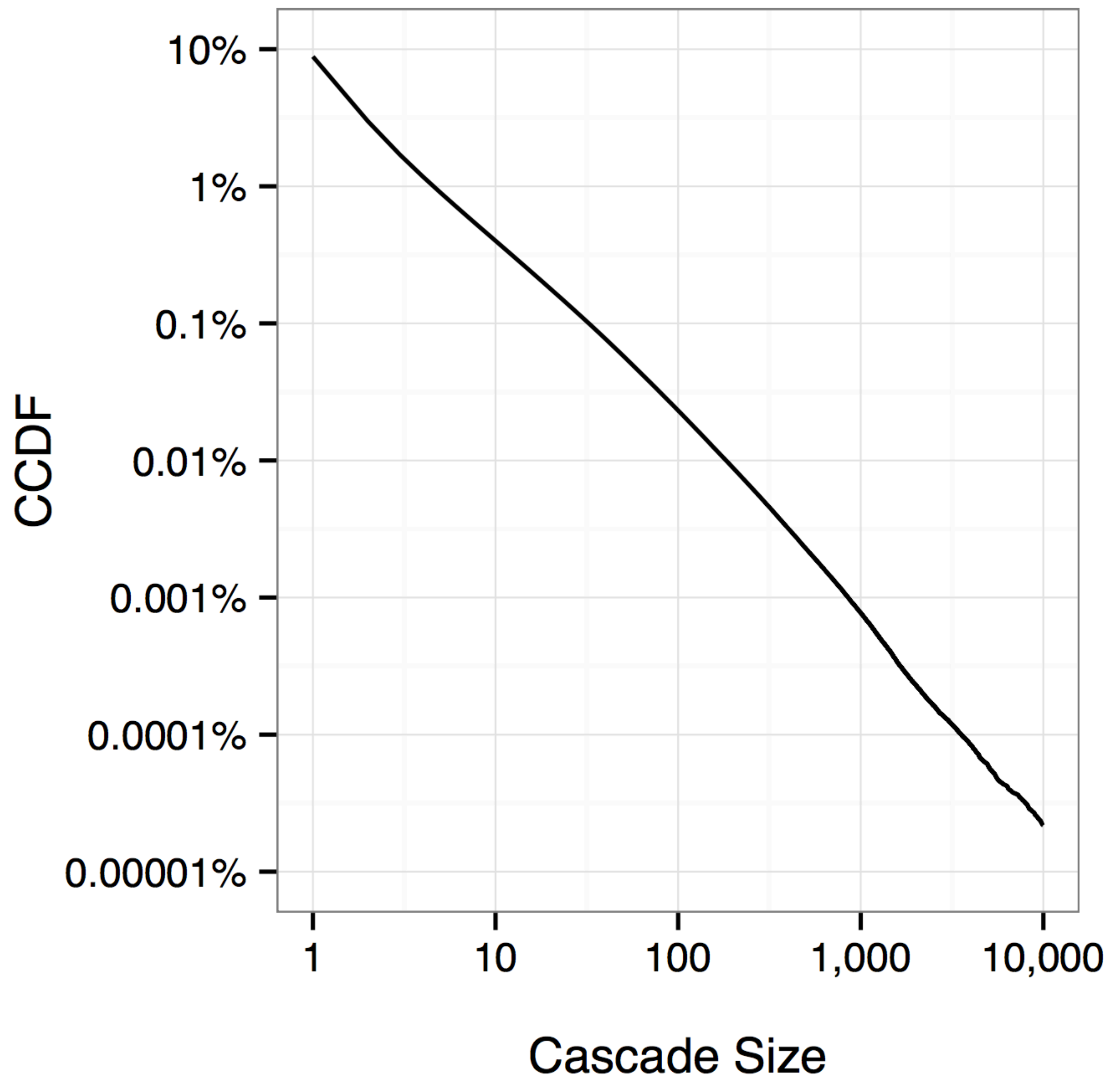


Measure virality in data!

First conclusion: most stuff goes nowhere

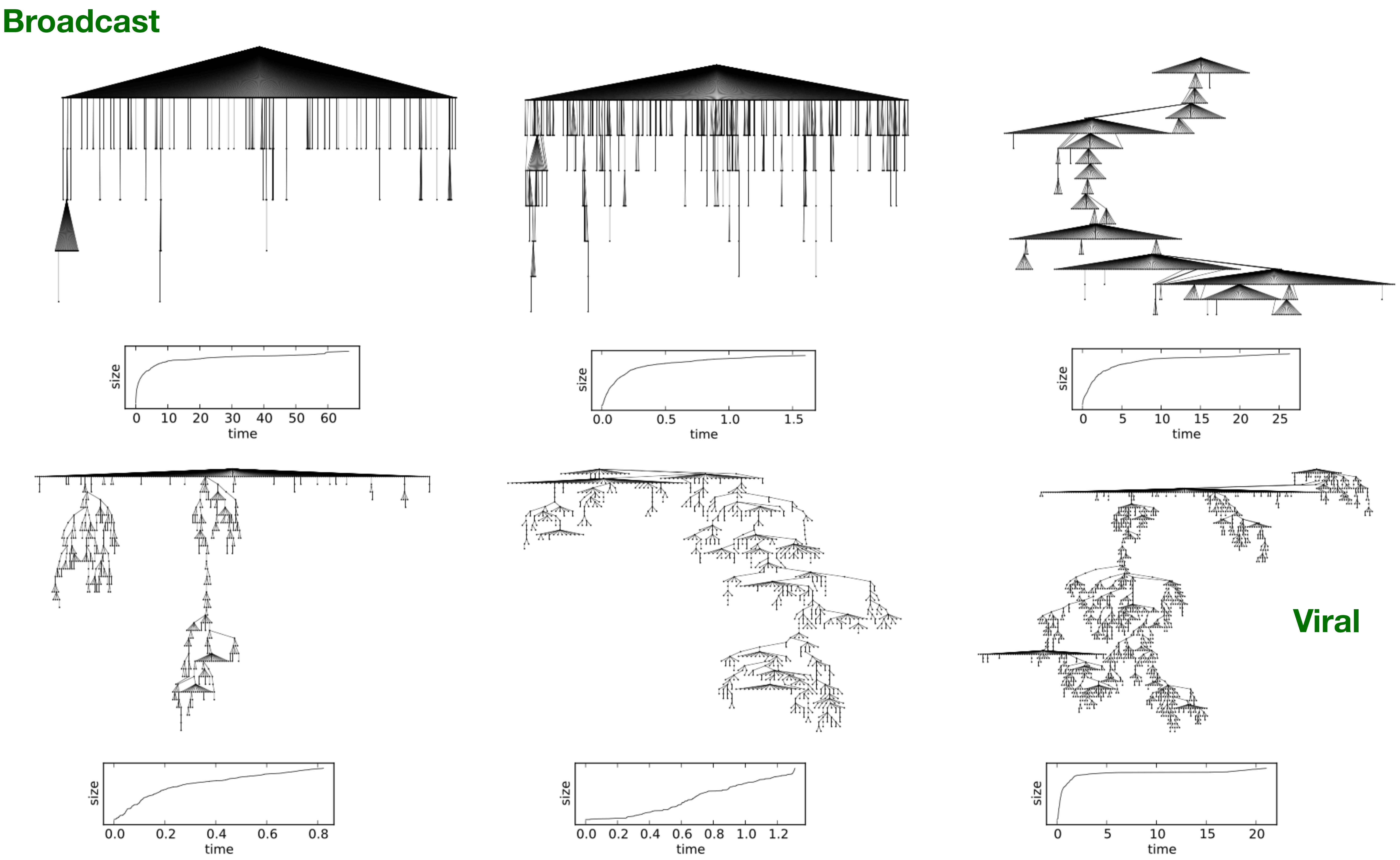
Average cascade size: 1.3

Not very interesting cascades: focus on trees of size at least 100 (empirically 1/4000)



Power law!

Surprising diversity



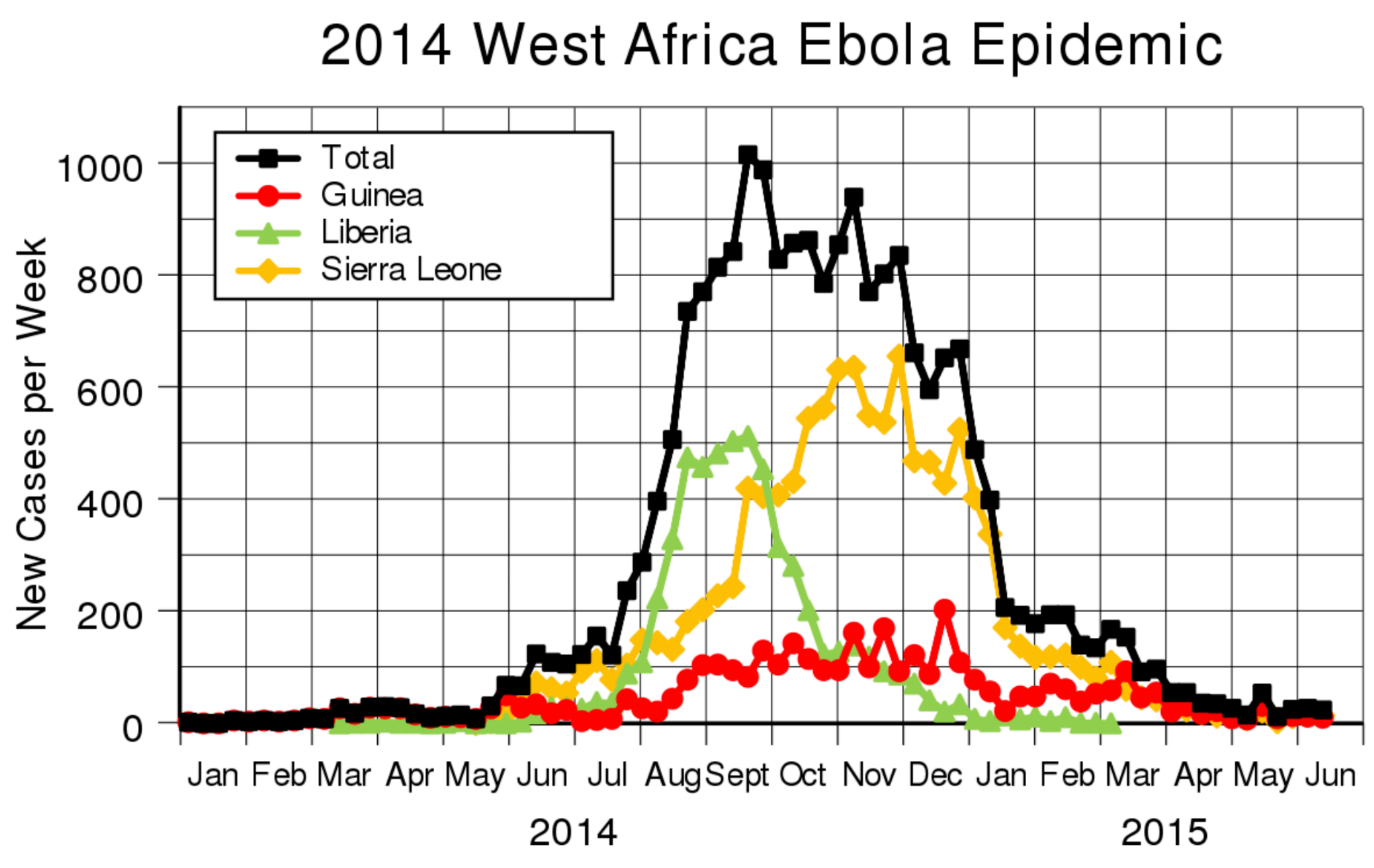
Epidemics

Epidemics

Why study epidemics in a computer science class?

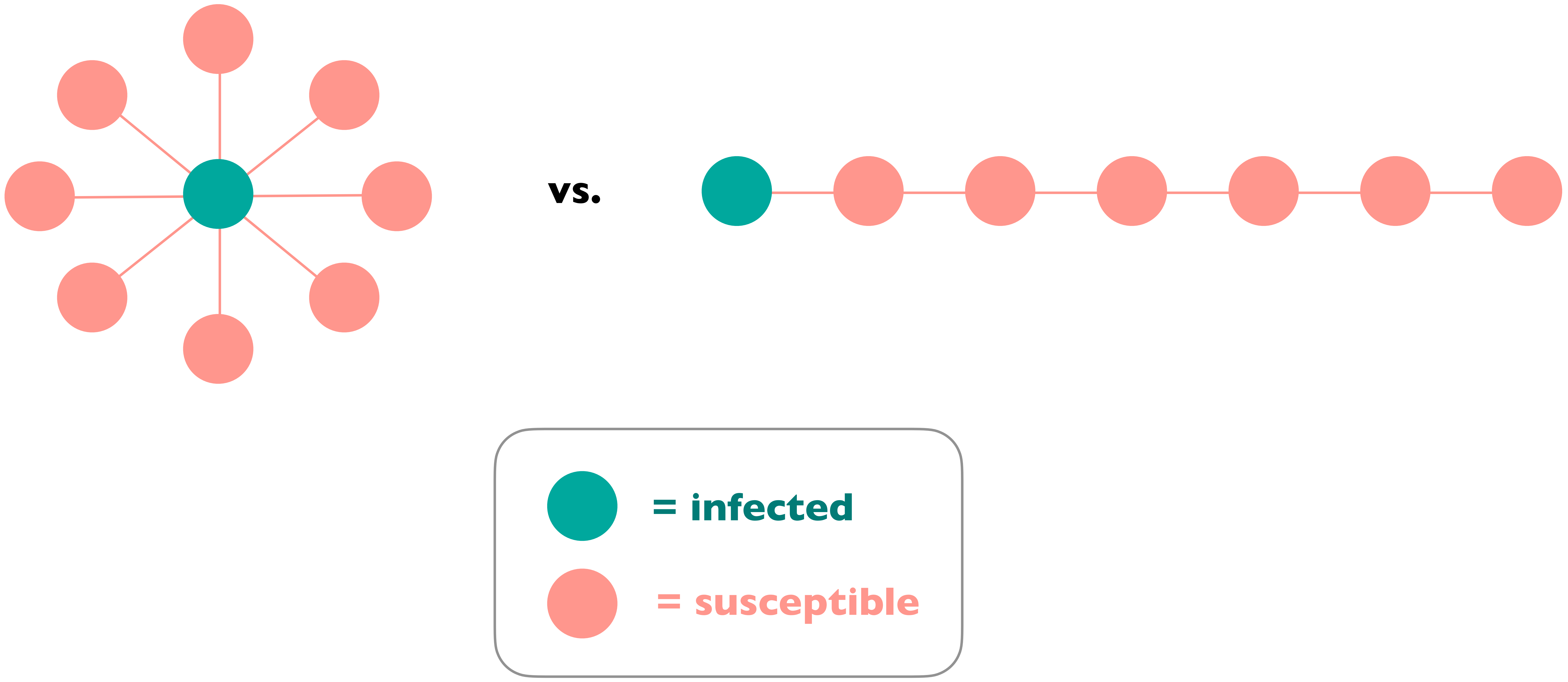
Epidemics are diseases that **travel socially**

The structure of social interaction networks determine the spread of disease



Epidemics

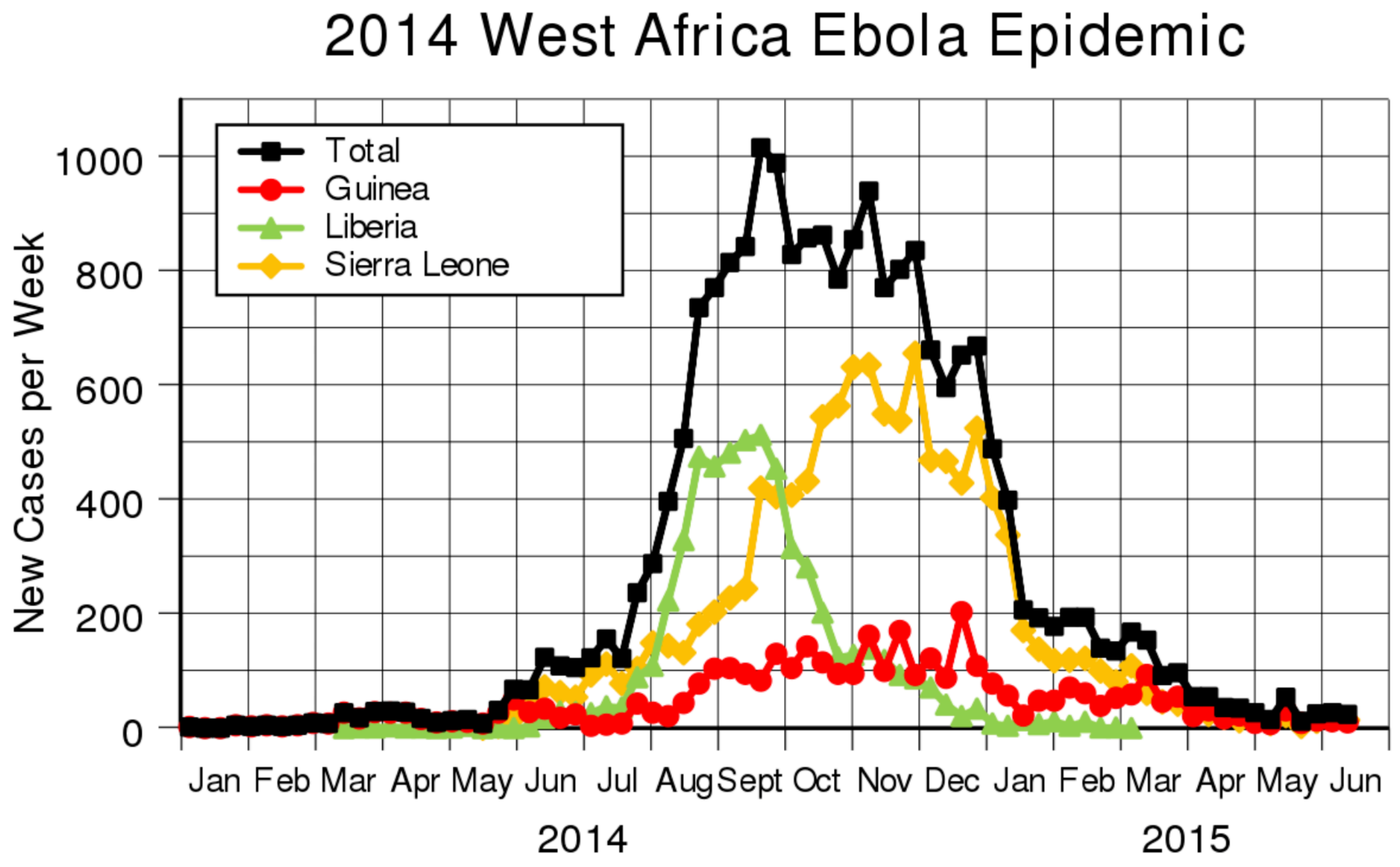
Which outbreak is more dangerous to the population?



Epidemics

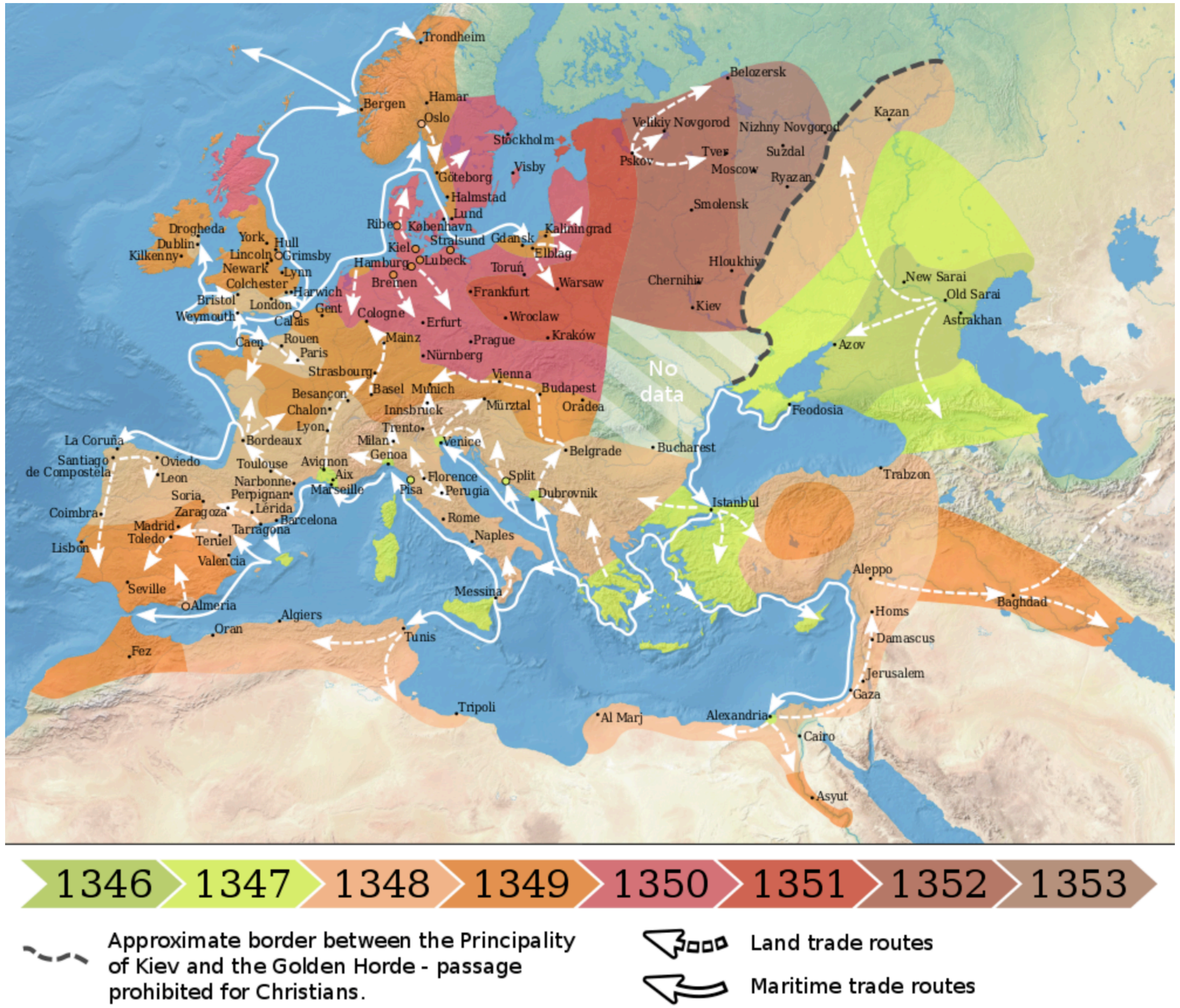
Types of epidemic diffusion:

- **Explosive spread** through a population
- **"Slow burn"** persistence over long periods of time
- Wave-like **cyclical patterns**



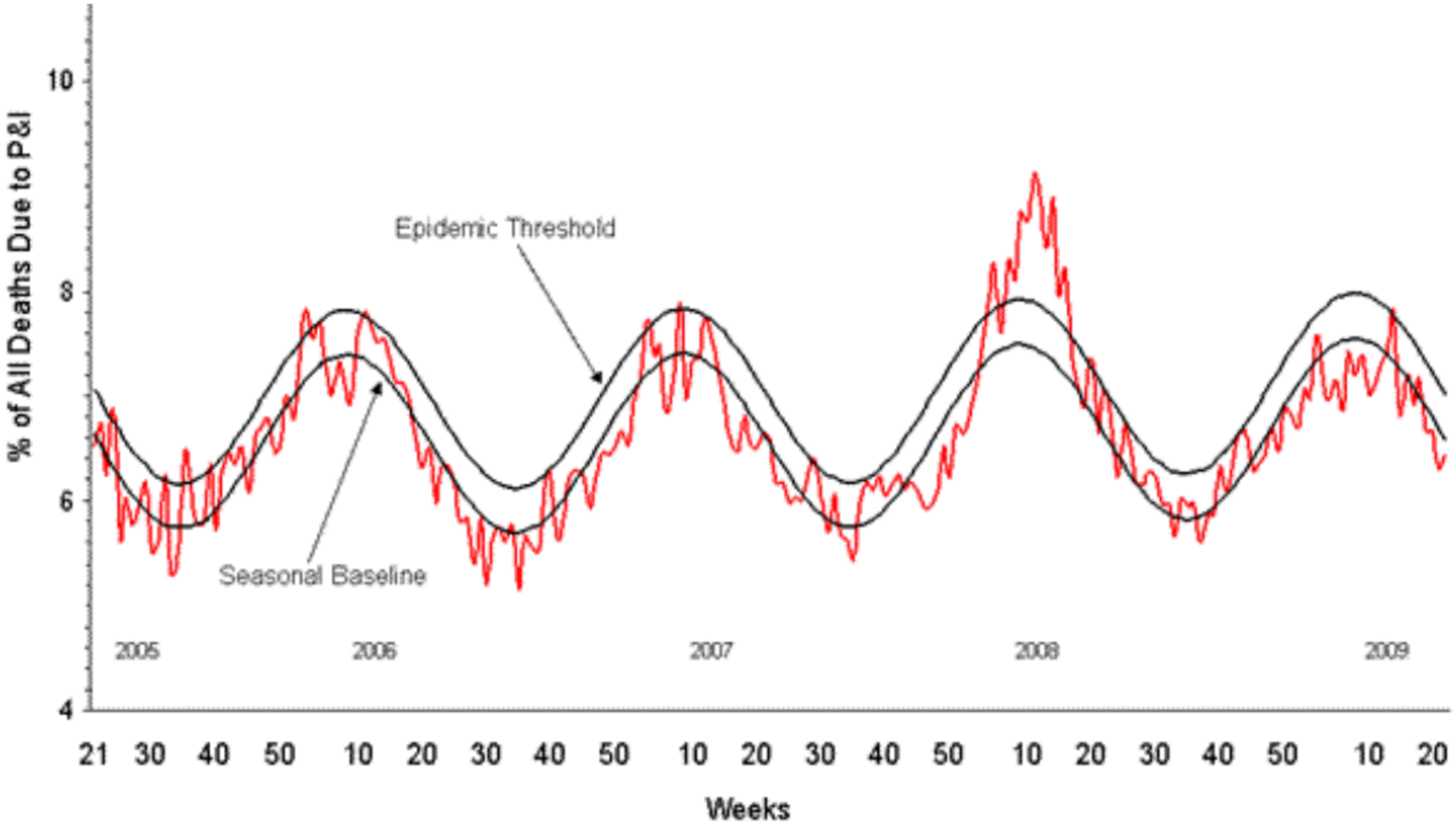
Epidemics

Explosive spread: Bubonic Plague (the "Black Death"): wiped out ~50% of the population in Europe (~150 million people) in 7 years



Epidemics

Other epidemics are cyclical

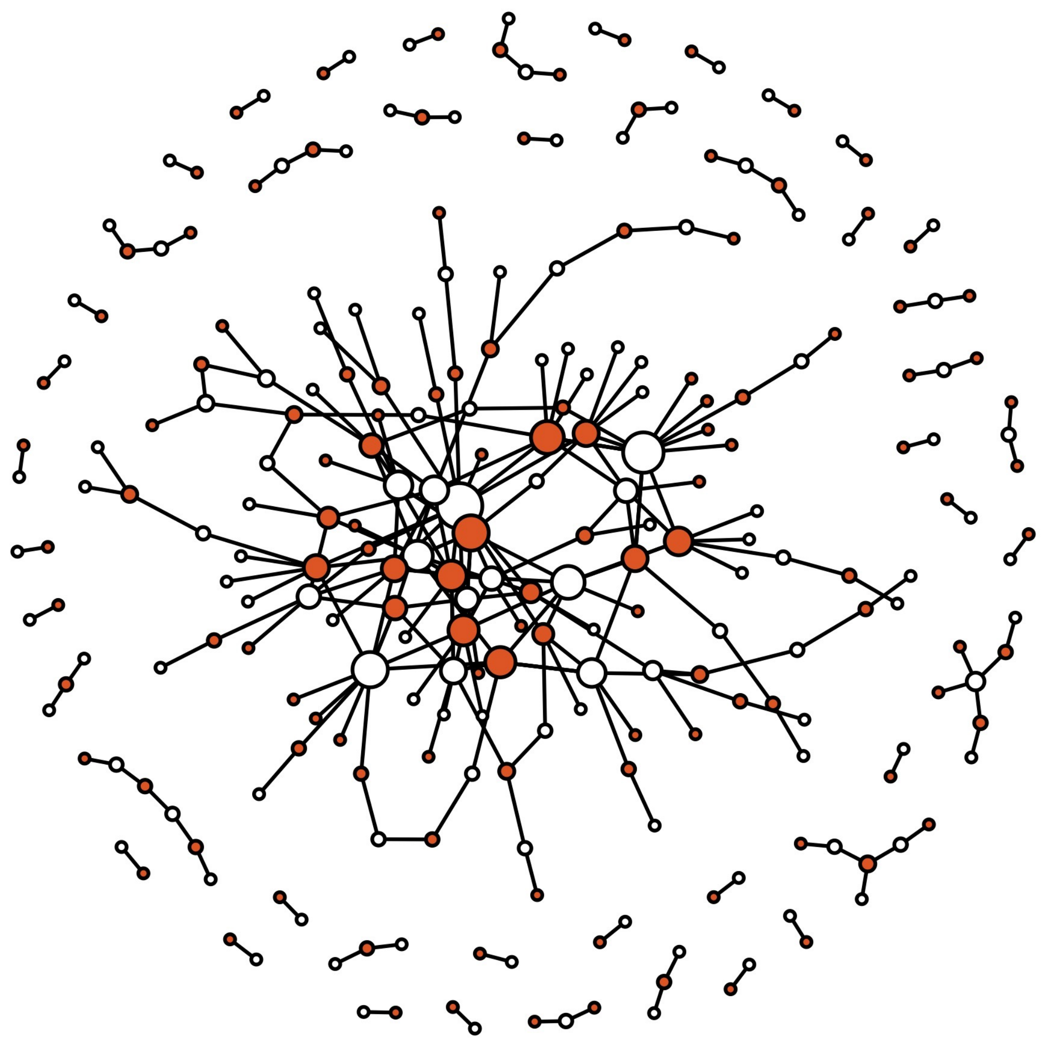


Epidemics

What determines how an epidemic might spread?

- Properties of the disease
- Structure of the network

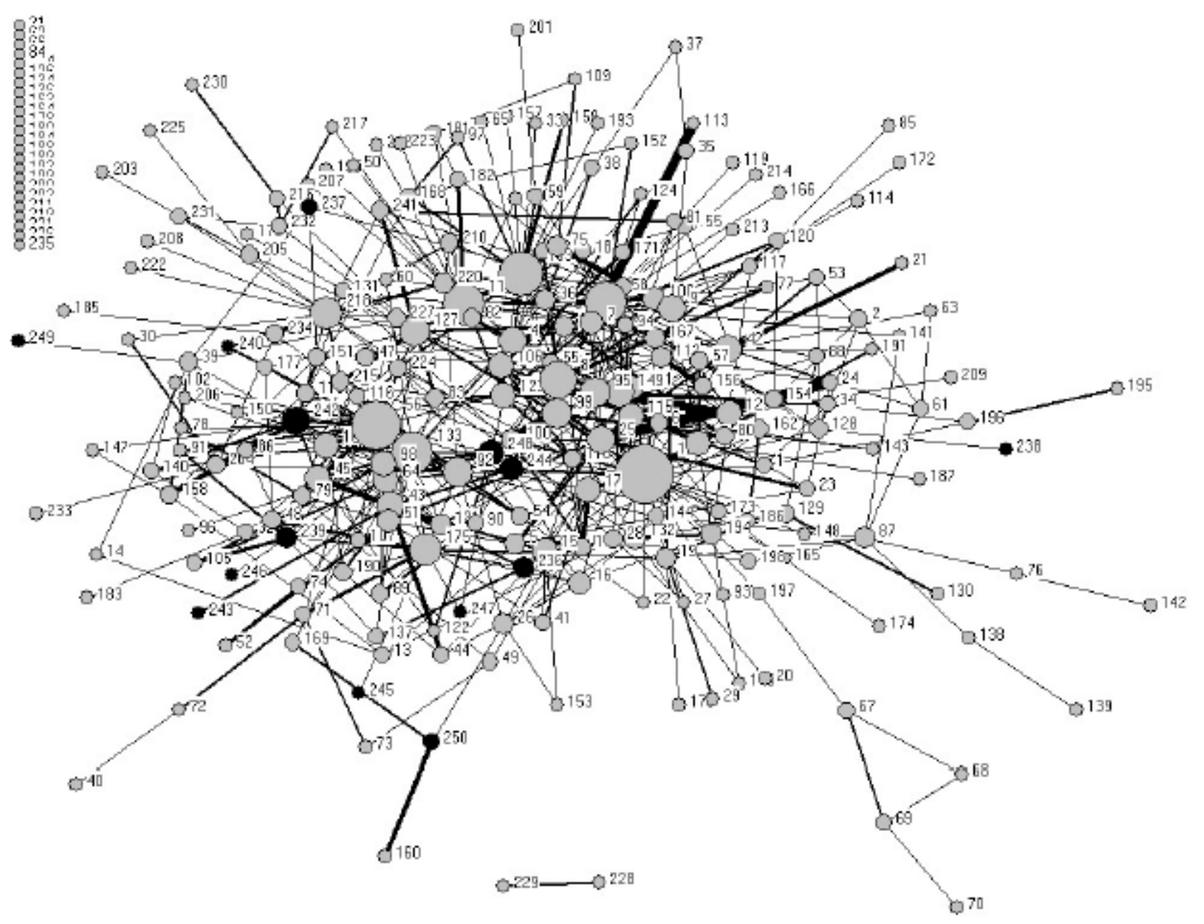
What network?



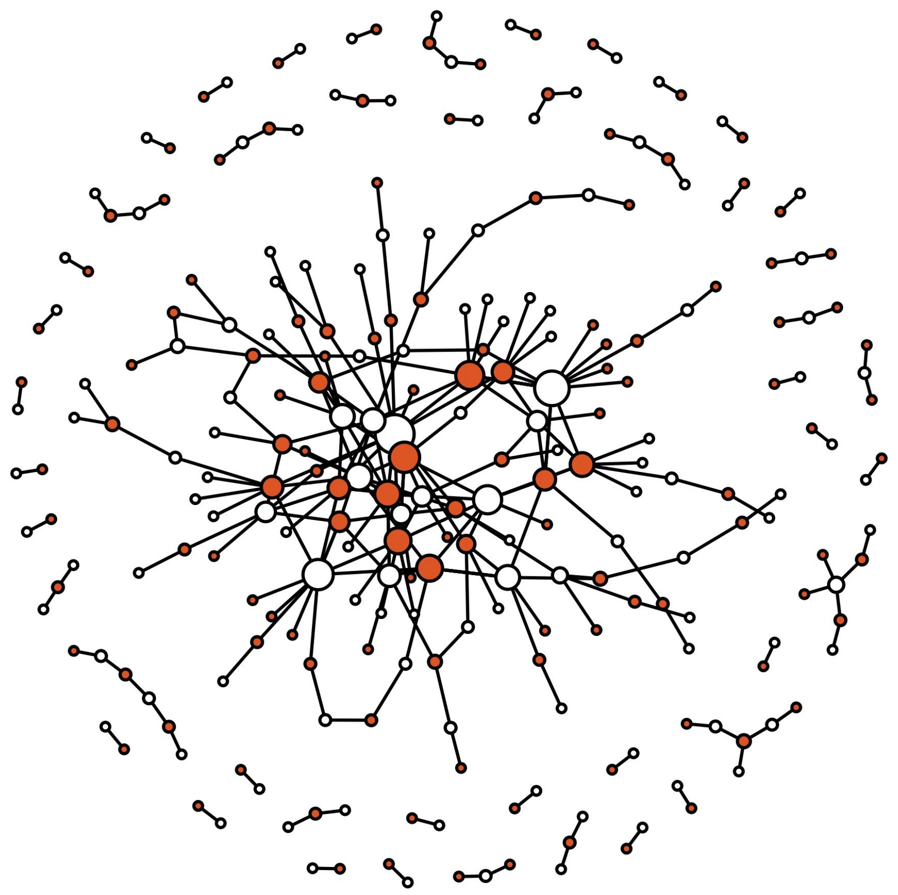
Contact Networks

Node for each person

Edge if two people come into contact with each other in a way that makes it possible for a disease to spread



UK fish farm exchanges



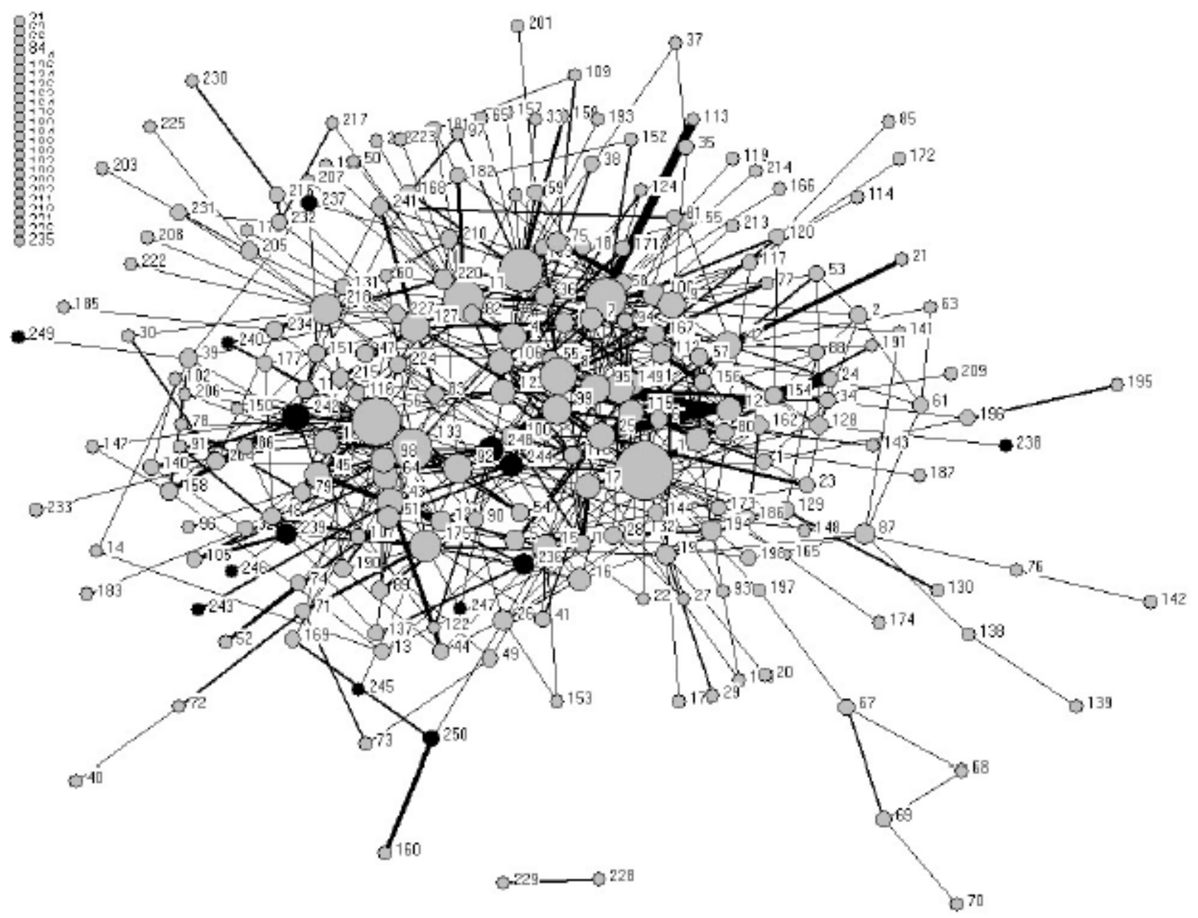
CH sexual contact network

Contact Networks

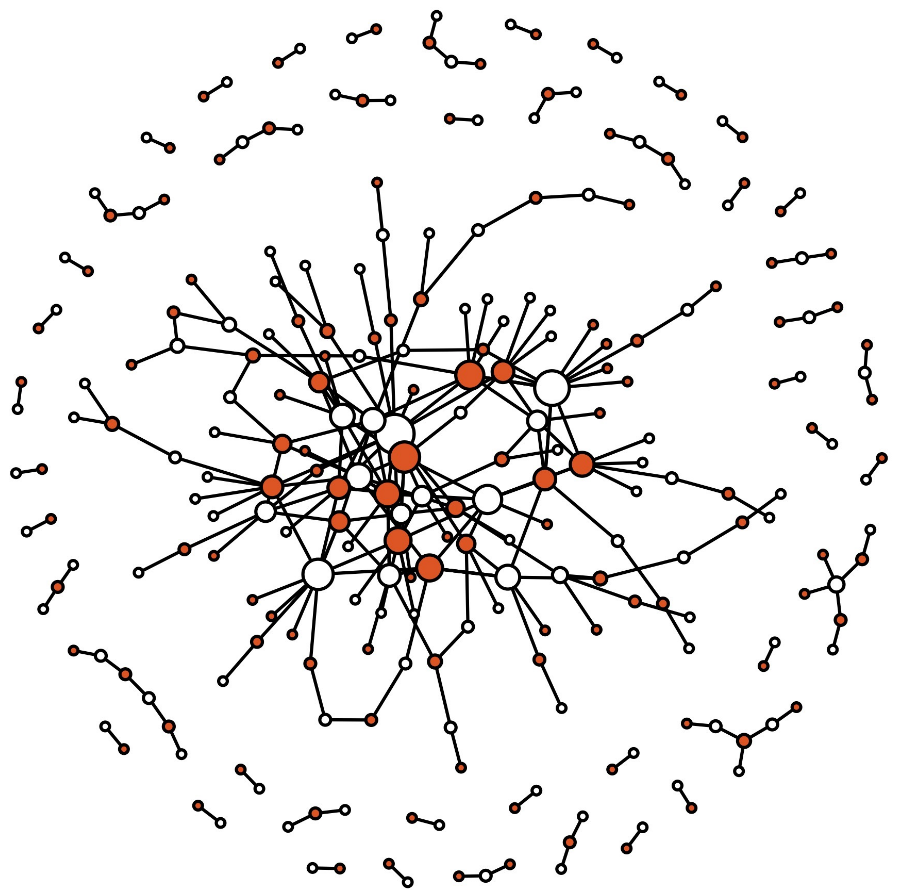
Once you've got through the laborious process of mapping out the contact network, can you us
disease?

No! Definition of "contact" depends on the disease

- Airborne transmission: edge between everyone who was in the same car, etc.) — many edges
- Close contact / sexual transmission: sparser graph



UK fish farm exchanges

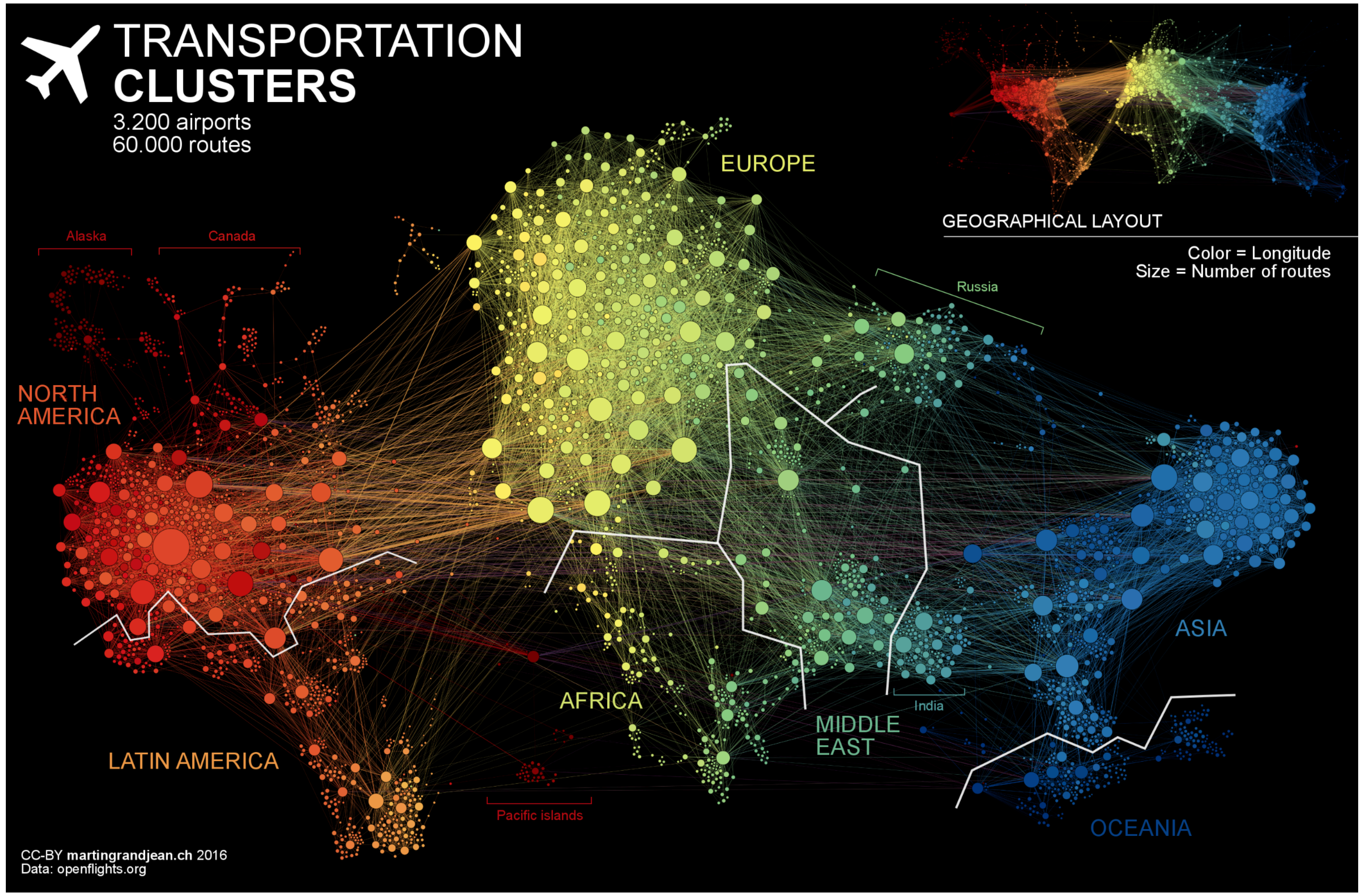


CH sexual contact network

Contact Networks

Big part of real-world epidemic research is constructing contact networks

Lots of work on travel patterns in cities, the worldwide airline network, etc. to understand how diseases can spread in today's world

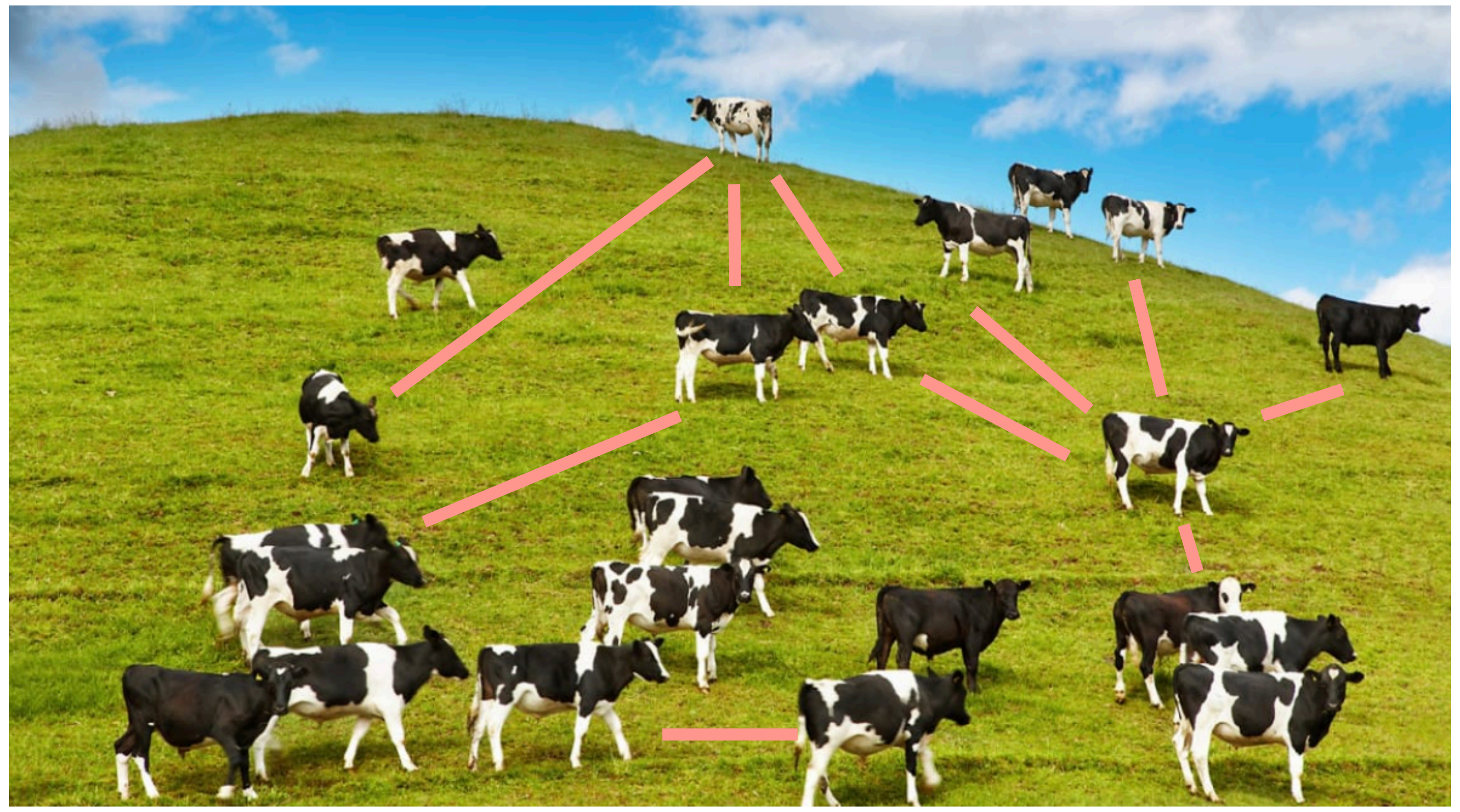


http://www.martingrandjean.ch/connected-world-air-traffic-network/

Contact Networks

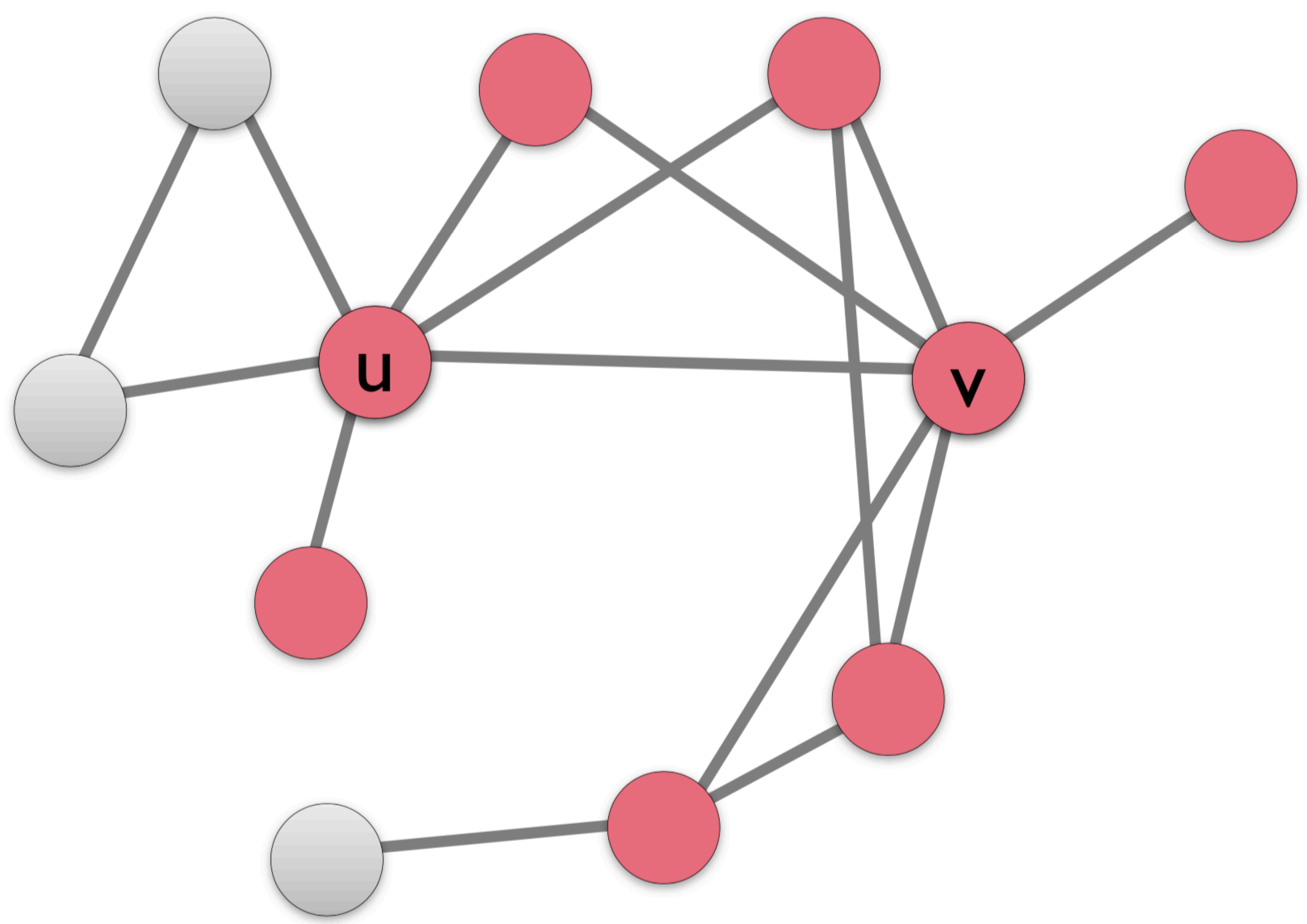
Not just human contact networks

Animal/livestock networks and plant networks

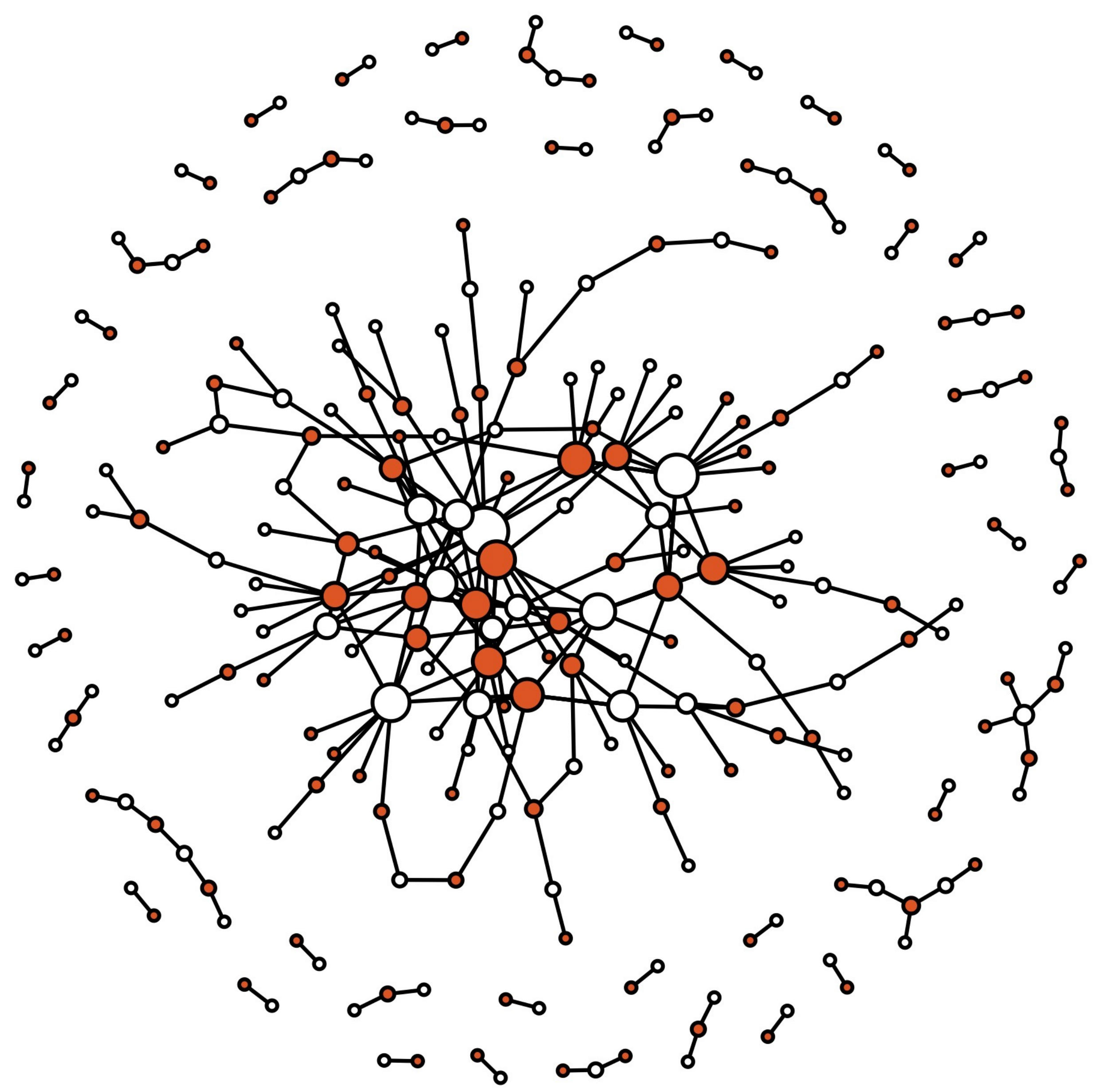


Behavioural vs. Biological Contagion

What's different between the spread of behaviour and viruses?



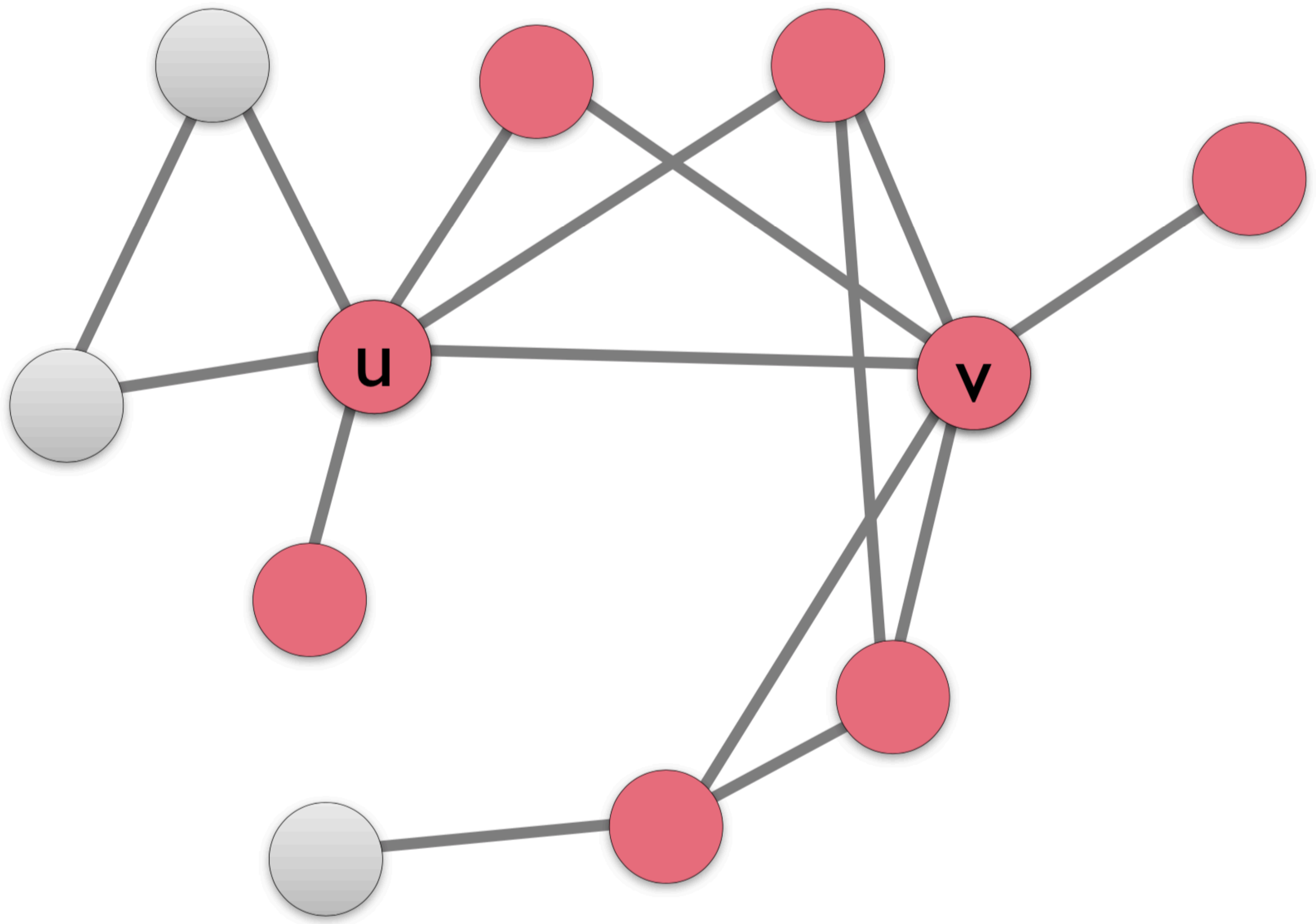
Decision cascade



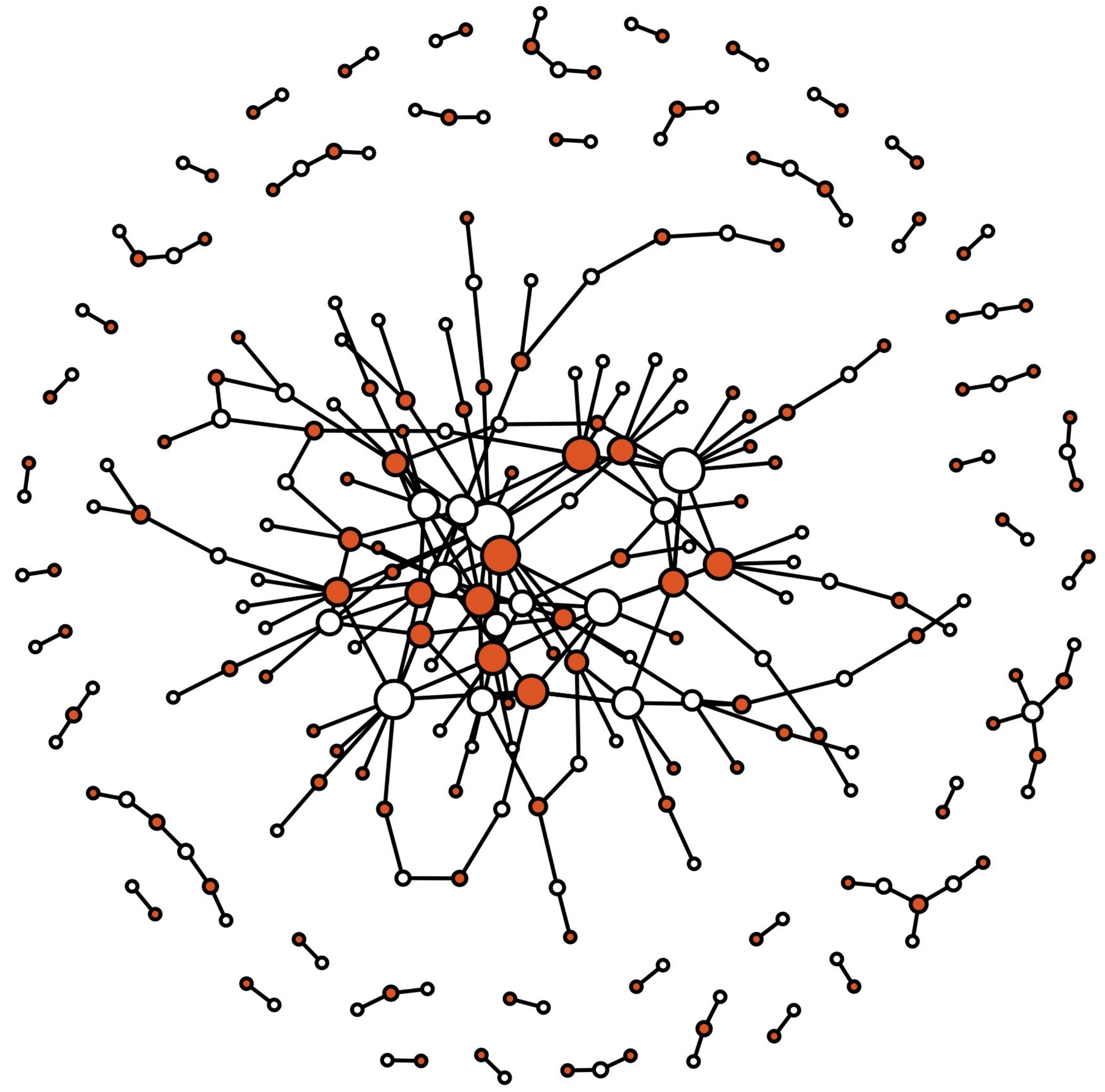
High school contact network

Behavioural vs. Biological Contagion

Biological/epidemic diffusion: **no decision-making!**



Decision cascade



High school contact network

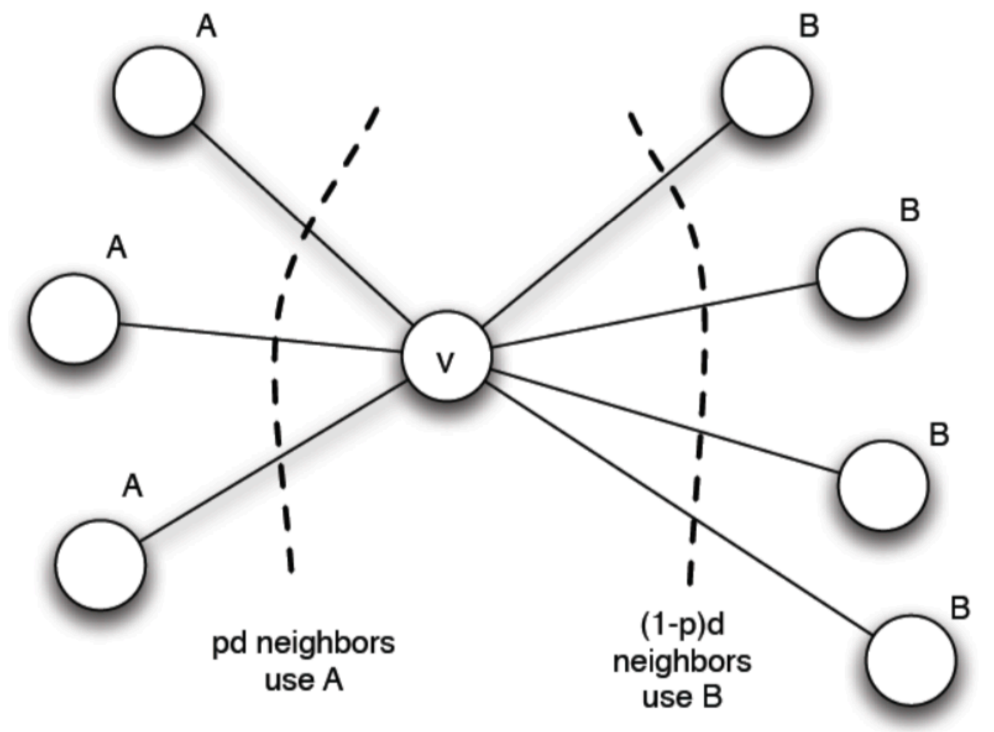
Modeling Epidemic Diffusion

Biggest difference: model transmission as **random**

No decision-making, but also the processes by which diseases spread from one person to another are so complex and unobservable at the individual level that it's most useful to think of them as random

Use randomness to abstract away difficult biological questions about the mechanics of spread

Behaviour (last class):



Epidemics (today):