Exploring epidemic spreading using network models

Simon Dobson

School of Computer Science
University of St Andrews
Scotland UK

simon.dobson@st-andrews.ac.uk
https://simondobson.org

University of St Andrews | FOUNDED 1413

Introduction

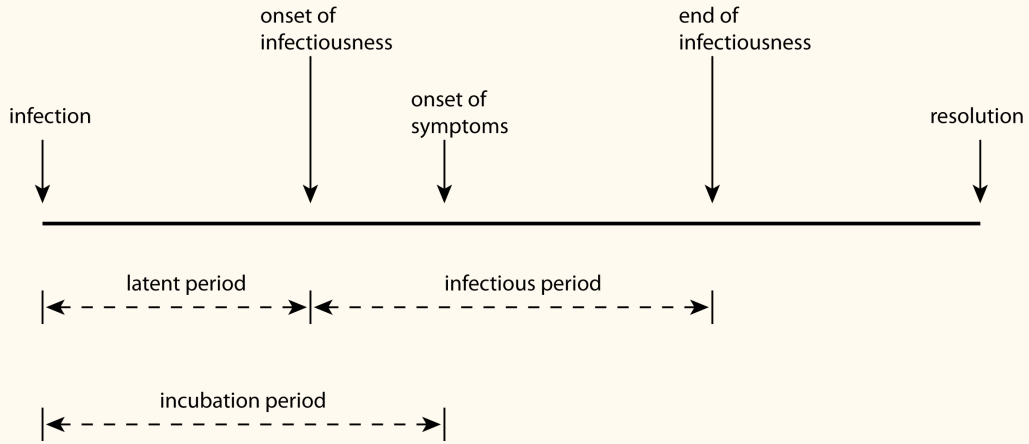
Epidemic modelling has become topical

A huge field drawing upon mathematical, statistical, and computational techniques

- Explore one part of the space: epidemic processes working over complex contact networks
- What possibilities can this show us?
- Can we make the tools and techniques more accessible?
- Can we generate insight for later empirical investigation?

Structure of this talk

Real diseases – general structure

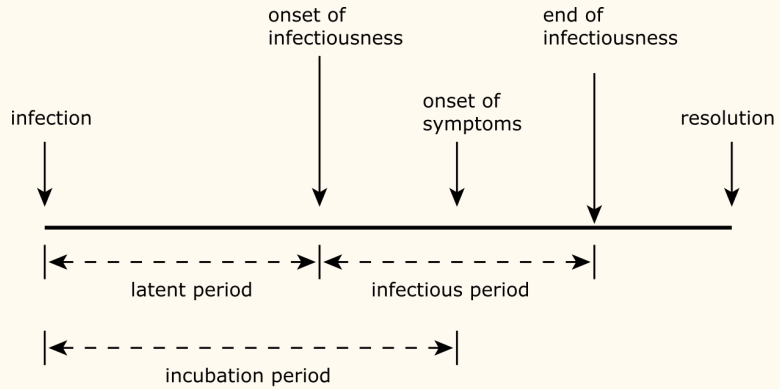


Different periods

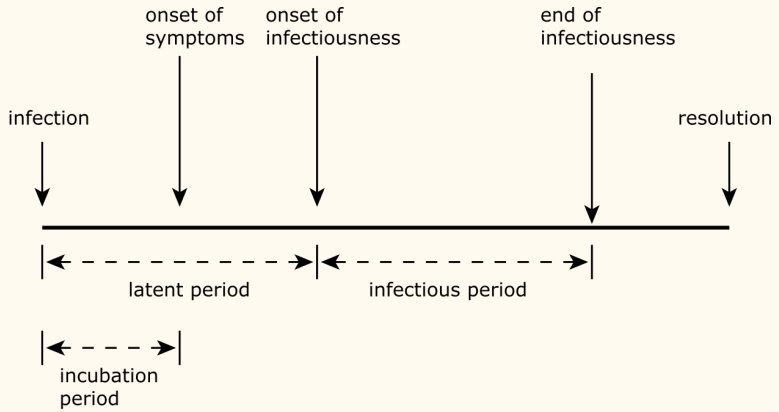
- *Incubation*: from infection to onset of symptoms
- *Latent*: from exposure to infectiousness
- *Infectious*: overlapping with symptoms (usually)

Periods defined by biology, of both disease and host

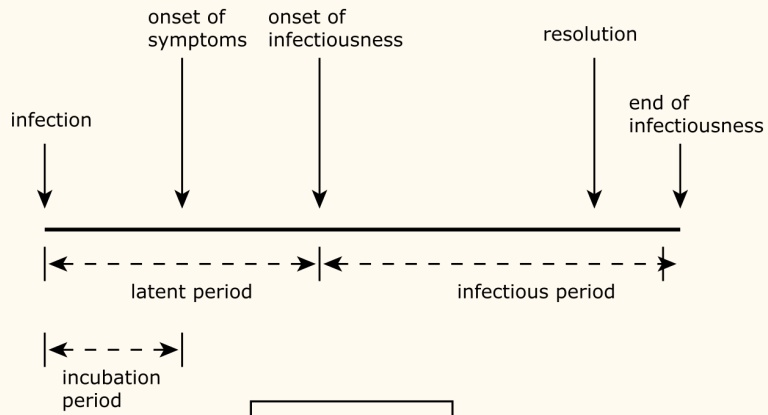
Real diseases – examples



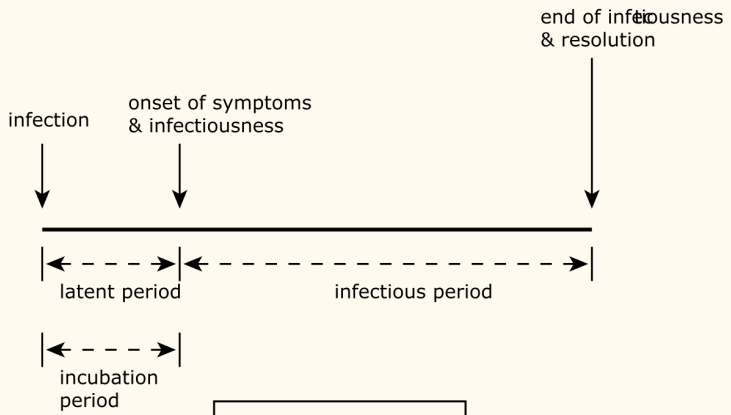
Measles (Type A)



SARS (Type B)



Ebola (Type B)



Zombie apocalypse

Real diseases – evolution

A person infected at the *end* of an epidemic doesn't get the same disease as a person infected at the *start*

- Pathogen is constantly mutating
- Lateral gene transfer from co-infecting pathogens
- Another reason to work to reduce transmission

Selection pressures often (but don't necessarily) introduce a particular dynamics

- More transmissible
- Less severe

$\mathcal{R}$ and all that [1]

$\mathcal{R}$, the case reproduction number

- Number of secondary cases per primary
- Especially $\mathcal{R}_0$, reproduction absent countermeasures

r, the case reproduction rate

- Doubling time for an epidemic
- Also sometimes see T_g, the inter-generation time

Typically averages over (unknown) distributions

- Details may be very significant

[1] Royal Society SET-C group. Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation, August 2020. URL https://royalsociety.org/-/media/policy/projects/set-c/set-covid-19-R-estimates.pdf

The "wickedness" of covid-19 – 1

$\mathcal{R}_0 \approx 3$ is not particularly infectious

- Straightforward to get $\mathcal{R} \approx 1.5$; harder to get $\mathcal{R} < 1$
- A more transmissible new variant may be emerging
- Significant overdispersion ("superspreaders")
- Infection may convey only temporary immunity

Substantial asymptomatic transmission

- Asymmetric costs (spreading *vs* dying, "long covid")
- Effective countermeasures are collective (and expensive)

The "wickedness" of covid-19 – 2

Infection fatality rate is about 1%

- Too large to comfortably ignore, too small to generate a universal consensus about its seriousness
- The numbers can be misrepresented

The goals of modelling

What are we trying to find out?

- Concrete: how will this *particular* outbreak behave, in this *particular* population?
- Abstract: how can diseases behave *in general*? Are there common mathematical structures?

Compartmented models

Traditional epidemic modelling uses the framework of a *compartmented model* of a disease

- A number of compartments holding some fraction of the population
- Can also think of a compartment as the state of each individual within the population (we'll come back to this)
- Rules on how these fractions change over time

Continuous SIR

The model

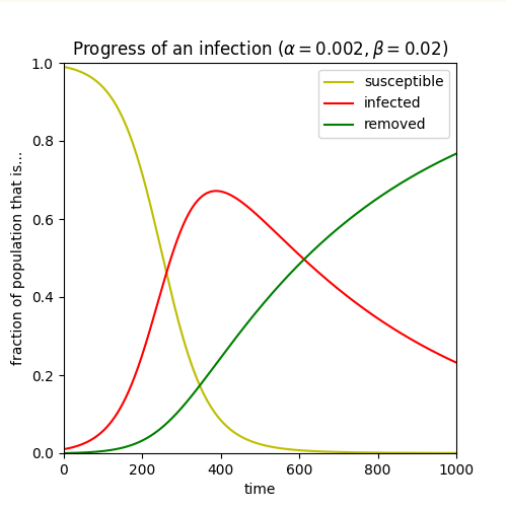
- **S**usceptible individuals can catch the infection from **I**nfected individuals
- …who then are **R**emoved from the dynamics by recovery (or death)

Epidemic dynamics

- Susceptibles infected per contact with probability β
- Infecteds removed with probability α
- Gives rise to $\mathcal{R} = \frac{\beta}{\alpha}$

$$\frac{dS}{dt} = -\beta SI \qquad \frac{dI}{dt} = \beta SI - \alpha I \qquad \frac{dR}{dt} = \alpha I$$

Solution



Different disease structures [2]

- SIR – complete immunity post-infection
- SIS – infection confers no immunity
- SEIR – exposed individuals are infectious before symptoms
- MSEIR – initial immunity passed from mother to child
- SEIRS – immunity wears off with time
- …

[2] H. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, December 2000. URL doi://10.1137/S0036144500371907

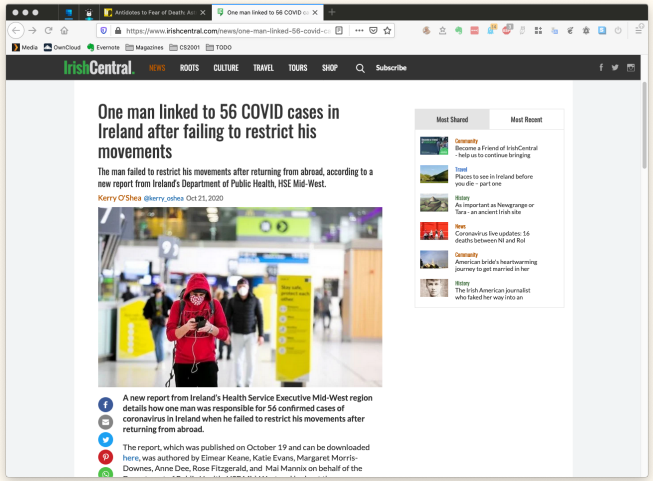
Benefits and limitations

Flexible and scalable

- Can model large populations
- Complete mixing

Limited heterogeneity

- Get heterogeneity using sub-populations and flows between them [3]
- Makes system stochastic



[3] K. Prem, A. Cook, and M. Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*, 13(9), 2017. URL https://doi.org/10.1371/journal.pcbi.1005697Ed

Structure of this talk

The case for using network science

Use a network as the substrate for the epidemic [4]

- Only adjacent nodes can interact
- Compartment = label on node
- Number of **SI** edges is the "locus" for infection

Pros and cons

- Doesn't scale as well as ODEs (explicit individuals)
- Can build contact structures and systems of equations we can't solve (but can simulate)

[4] M. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66, July 2002. URL doi://10.1103/PhysRevE.66.016128

Basic treatment – networks

Network degree distribution

- Probability p_k of randomly-chosen node having degree k

Often start with a *mean field* approach

- The mean degree $\langle k \rangle$ is "representative"
- Solve equations *as if* all nodes have degree $\langle k \rangle$

Add fine structure

- Loops, assortativity, modules, layers, nesting, …
- Adaptive behaviour to change features over time and/or in response to the disease

Basic treatment – processes

Assign a state vector to each node

- For epidemics, this might be the node's compartment

Process defines changes to state vectors

- A function of current states of the node and its immediate neighbours
- Generally stochastic, applied with some probability

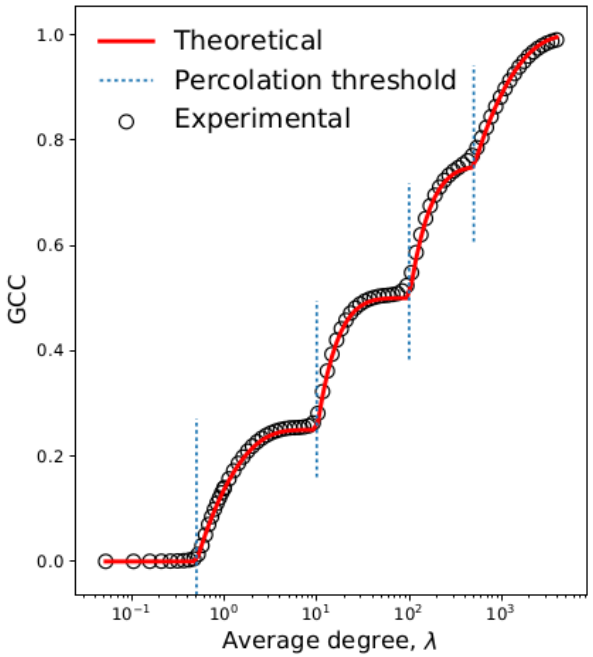
Seed the network with initial state vectors

- For SIR, mainly susceptible with a few infected

How to do analysis

The "gold standard" is an analytic model with numerical validation

- Find an analytic description for what happens under different infection parameters
- Run process on random networks with different topologies
- Lots of repetitions to squeeze out variance
- (Hopefully) sample points land on solutions to the equations [5]



[5] P. Mann, V. A. Smith, J. Mitchell, and S. Dobson. Random graphs with arbitrary clustering and their applications. *Physical Review E*, 2020. URL http://arxiv.org/abs/2006.08427. To appear

Gillespie simulation – 1

Originally developed for *ab initio* chemistry [6]

- Define basic events e and their probabilities $P(e)$
- When will the next event occur? What will it be?

Consider SIR as a model

- **I** infects a **S** neighbour, $P(infect) = \beta SI$
- **I** is removed, $P(remove) = \alpha I$
- Each event changes the sizes of the loci

[6] D. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25): 2340—2361, 1977

Gillespie simulation – 2

Define $P(\tau, e)\, d\tau$ as the probability that an event e occurs in the next interval $(t + \tau, t + \tau + d\tau)$

Define $P(\tau) = \sum_e P(\tau, e)$ as the probability that *some* event happens in the next interval τ.

Then re-arrange the joint probability distribution

$$P(\tau, e) = P(\tau)P(e|\tau)$$

We want to draw a (τ, e) pair from this distribution

Gillespie simulation – 3

Turn the probability density function into a *cumulative* density function

$$P(x \leq x_0) = \int_{-\infty}^{x_0} P(a)\, da$$

If we draw a value r uniformly from $[0, 1]$ then we can compute $x = P^{-1}(x \leq r)$ to get a value distributed according to $P(x)$

Letting $a = \sum_e P(e)$, much maths then yields

$$\tau = \frac{1}{a} \ln\left(\frac{1}{r_1}\right) \qquad e = argmax_e\left(\sum_{e'=e_0}^{e} P(e') \leq r_2 a\right)$$

Tooling

There wasn't any standard tooling, so we built some

A flexible way to express networks and processes

- `epydemic`, a simulation framework using `networkx`
- Reference epidemic (and other) processes
- Network generators

A way to perform repeated, repeatable, experiments

- `epyc`, a computational experiment manager
- Experiment submission, parallelism, remote evaluation
- Immutable datasets with metadata, stored in HDF5

Example code

```
import numpy
import pandas
from epyc import ClusterLab, HDF5LabNotebook, RepeatedExperiment
from epydemic import ERNetwork, SIR, StochasticDynamics

# notebook for results and lab with connection to compute cluster
nb = HDF5LabNotebook('test.h5', description='My experiments in networking')
lab = ClusterLab(profile='hogun', notebook=nb)

# set up the experimental parameters
lab[ERNetwork.N] = 10000
lab[ERNetwork.KMEAN] = 40
lab[SIR.P_INFECTED] = 0.001
lab[SIR.P_REMOVE] = 0.002
lab[SIR.P_INFECT] = numpy.linspace(0.00001, 0.0002, num=50)

# construct the experiment: a process and a class of networks
m = SIR()
g = ERNetwork()
e = StochasticDynamics(m, g)

# repeat runs across the parameter space
lab.runExperiment(RepeatedExperiment(e, 100))

# retrieve for analysis
df = nb.current().dataframe(only_successful=True)
```

Structure of this talk

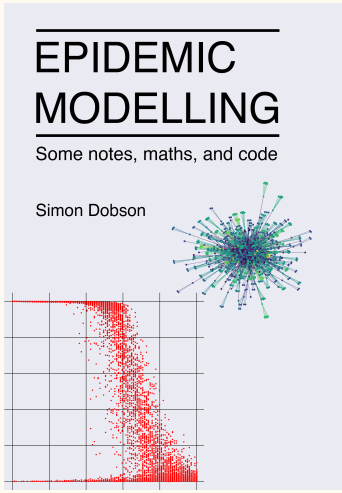
Explorations

We've been experimenting with different network structures

- Especially interested in "clustered" networks: friends-of-friends and larger cycles
- Fine structure affects how processes evolve

Make the science more accessible

- With available and re-usable code
- With explanations [7]

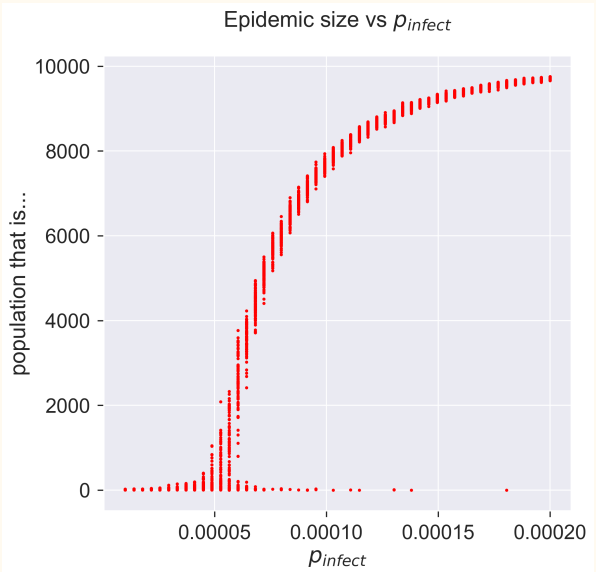


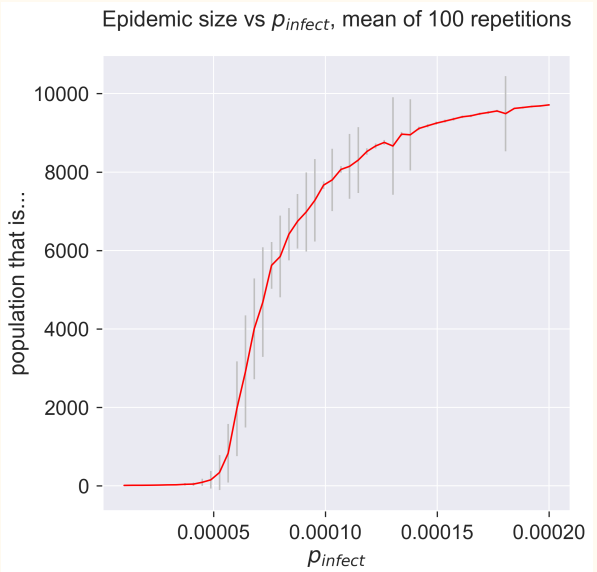
[7] S. Dobson. *Epidemic modelling – Some notes, maths, and code*. Independent Publishing Network, 2020. ISBN 978-183853-565-0. URL https://simoninireland.github.io/introduction-to-epidemics/

The epidemic threshold

Erdős-Rényi (ER) networks

- For N nodes build the complete network K_N
- For each edge, retain ("occupy") it with probability p_{infect}
- Leads to p_k normally distributed around $\langle k \rangle = p_{infect} N$

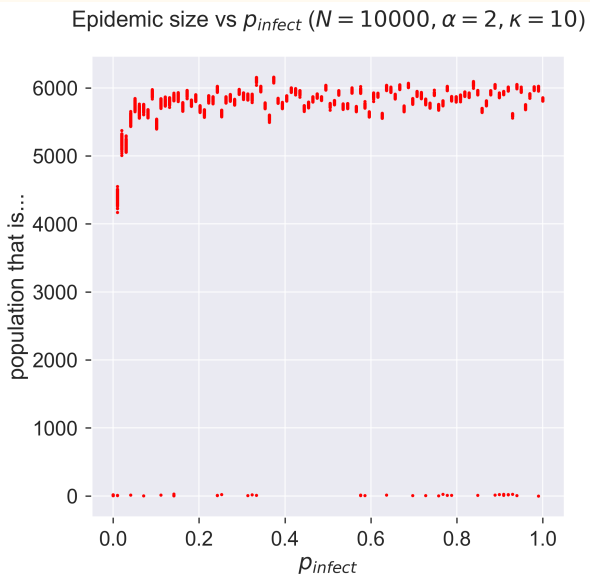


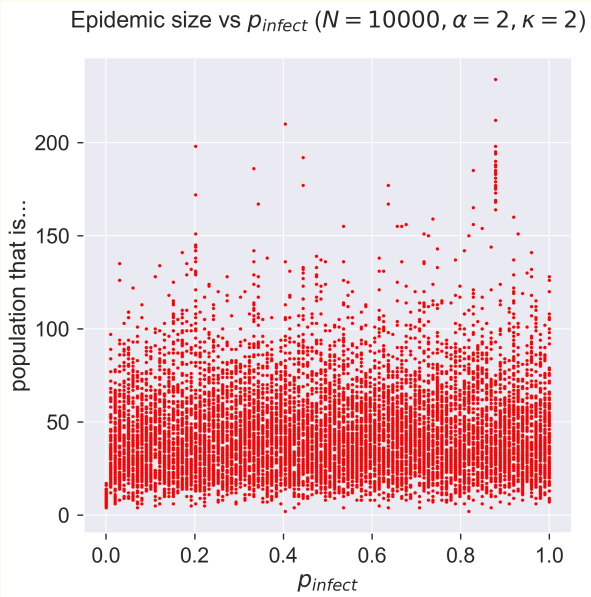


Not all networks behave like this

Too "even" to be a good model of human contacts

- Powerlaw with cutoff, $p_k \propto k^{-\alpha}\, e^{K/\kappa}$



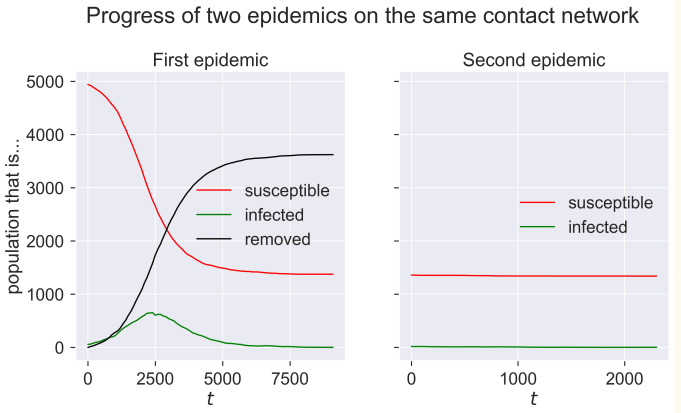


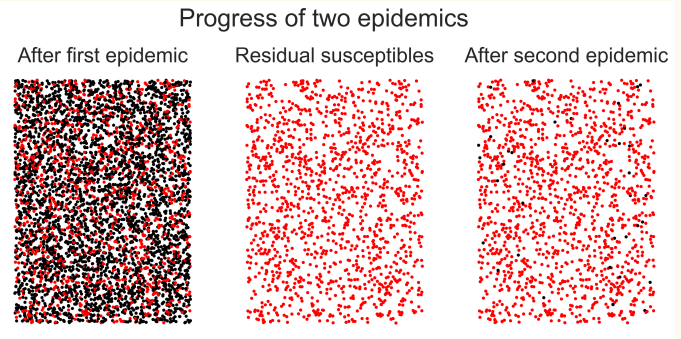
- Relatively insensitive to p_{infect}, but sensitive to α and κ

Herd immunity

Sufficient immune/recovered individuals to stop an epidemic propagating

- Infecteds never adjacent to enough susceptibles
- First epidemic changes the effective topology
- "Effective" $\langle k \rangle$ drops from 20 to 5.5





Why pursuing herd immunity is a bad idea

Herd immunity has been seriously advocated as a strategy for covid-19[8]

Ignores some rather inconvenient facts

- A 1% death rate = 700K UK deaths, about one year's excess
- At a rate that would collapse health services
- Immunity may not be permanent – which makes herd immunity behave differently (or not appear at all)
- Long covid not accounted for in the costs

[8] See the "Great Barrington Declaration", https://gbdeclaration.org

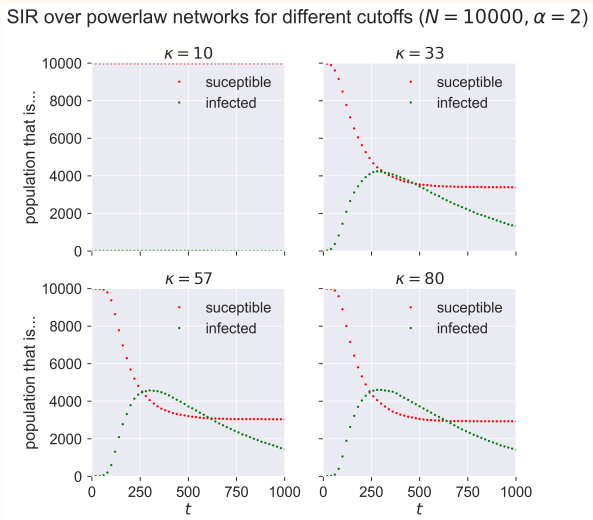
Vaccination

"Herd immunity without the bad bits"

- Aim for the herd immunity threshold, generally about 60% of the population
- …without anyone actually being infected

Epidemic proceeds at different rates depending on topology

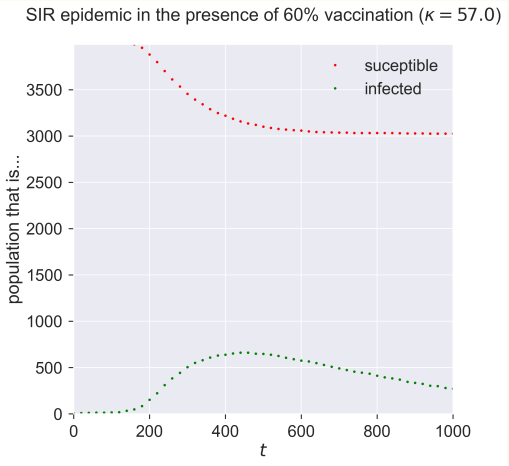
- "Enough" contacts stabilise the size of outbreak



Vaccination strategies

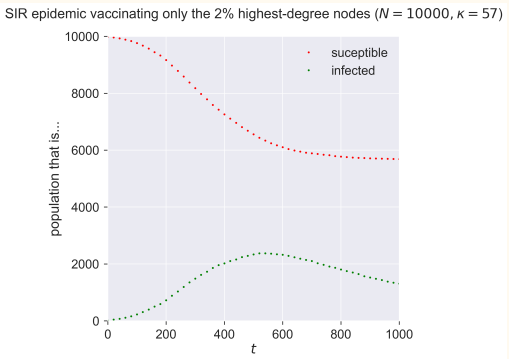
Randomly vaccinate 60% of the population

- Massive reduction in epidemic size
- Only catching high-degree nodes at random
- Sensitive to missing people



If we target vaccination we can reduce the threshold needed to get the same effect

- Target highest-degree 2% of nodes
- Take out the super-spreaders



Physical distancing

What does a physically-distanced contact network look like?

- Good question: needs *lots* of assumptions, especially when considering compliance

One possible model

- Normally-distributed, fully connected family "bubbles" of mean size 4
- A couple of members with outside contacts
- Exponentially-distributed connections between the contacts in different bubbles

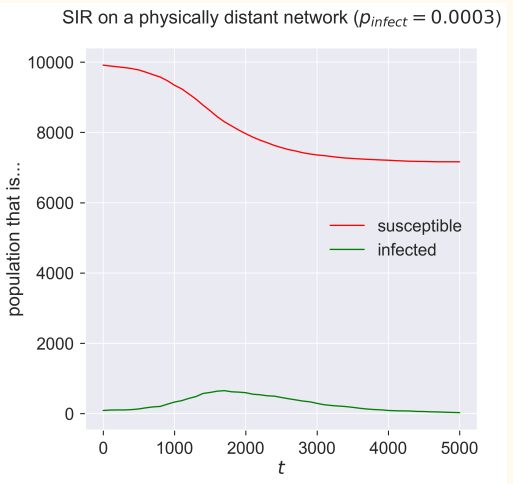
Lockdown changes propagation

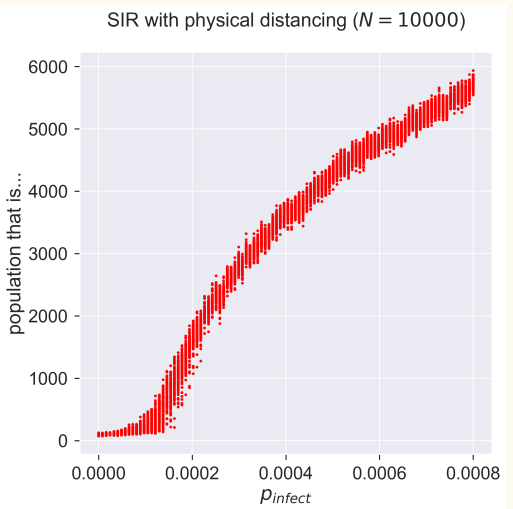
Changes the epidemic threshold compared to an ER network

- Needs a higher infectivity to take off

Slower take-off

- Not like a powerlaw network
- Get bursts of infection if the infection gets into a bubble





Asymptomatic transmission

Because covid-19 is essentially SEIR (or maybe SEIRS) it invites other countermeasures

- Self-isolating on showing symptoms is ineffective
- Try to find the asymptomatic carriers

This is the basis for track-and-trace

- Identify contacts of that person
- Quarantine them if they're infected – means we catch infecting individuals before they knew to self-isolate
- Quarantine the symptomatic individual too

Track and trace in practice

A large-scale procedure, unlike the local procedure of self-isolation when symptomatic

- Requires organisation by some authority
- What can possibly go wrong?…

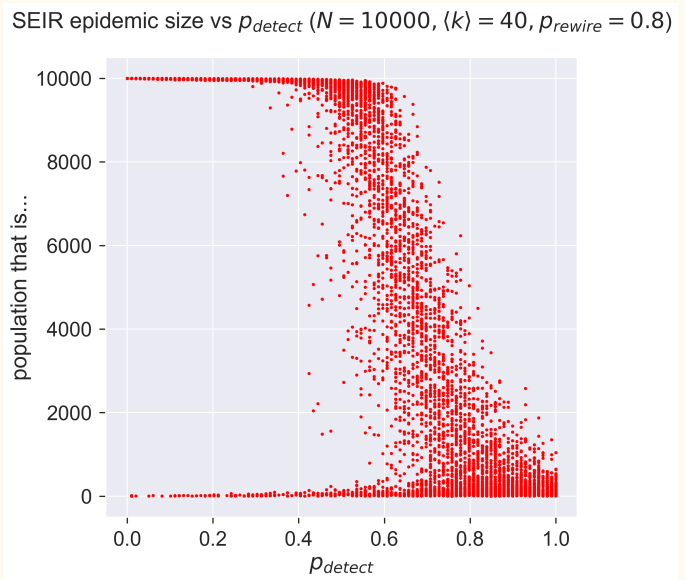
Unlikely to be fully accurate even if done competently

- Some proportion of people don't quarantine? (p_{rewire})
- Only test some proportion of contacts? (p_{detect})

The impact of detection rates

Hold p_{rewire} constant and vary p_{detect}

- High detection is very effective
- Need to check at least 40% to have any effect at all
- Lower rates are unstable
- All sizes of epidemic are possible
- Possibly a "smeared" phase transition [9]
- Possibly an artefact



[9] L. Hébert-Dufresne and A. Allard. Smeared phase transitions in percolation on real complex networks. *Physical Review Research*, 1, August 2019. URL https://doi.org/10.1103/PhysRevResearch.1.013009

Structure of this talk

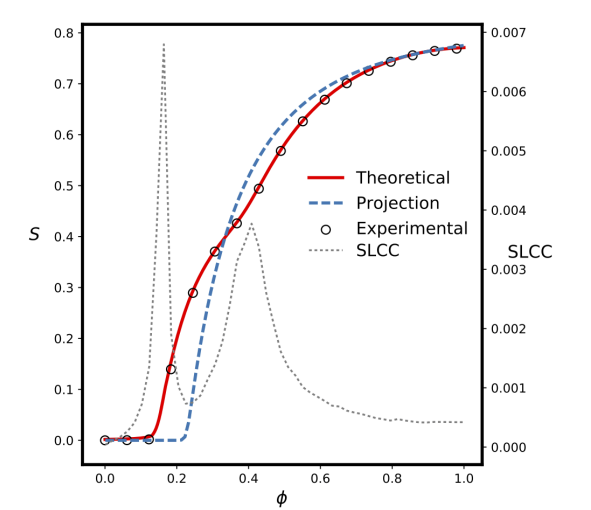
Research directions

Multiple diseases

- What happens when disease evolve?
- Co-infection dynamics, when one affects susceptibility to another

We're now very interested in network fine structure

- Disrupt processes by disrupting small local features?
- Local phenomena as leading indicators of global changes [10]



[10] P. Mann, V. A. Smith, J. Mitchell, and S. Dobson. Random graphs with arbitrary clustering and their applications. *Physical Review E*, 2020. URL http://arxiv.org/abs/2006.08427. To appear

Three things to take away

1. Epidemic spreading still isn't fully understood – there's lots of exciting work still to do, mathematically and computationally
2. Interactions between network and process can be very subtle, and may have significant effects
3. We can explore the space of public policy decisions as "citizen scientists", and also counter misinformation

References

S. Dobson. *Epidemic modelling – Some notes, maths, and code*. Independent Publishing Network, 2020. ISBN 978-183853-565-0. URL https://simoninireland.github.io/introduction-to-epidemics/.

D. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25): 2340—2361, 1977.

L. Hébert-Dufresne and A. Allard. Smeared phase transitions in percolation on real complex networks. *Physical Review Research*, 1, August 2019. URL https://doi.org/10.1103/PhysRevResearch.1.013009.

H. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, December 2000. URL doi://10.1137/S0036144500371907.

P. Mann, V. A. Smith, J. Mitchell, and S. Dobson. Random graphs with arbitrary clustering and their applications. *Physical Review E*, 2020. URL http://arxiv.org/abs/2006.08427. To appear.

M. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66, July 2002. URL doi://10.1103/PhysRevE.66.016128.

K. Prem, A. Cook, and M. Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*, 13(9), 2017. URL https://doi.org/10.1371/journal.pcbi.1005697Ed.

Royal Society SET-C group. Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation, August 2020. URL https://royalsociety.org/-/media/policy/projects/set-c/set-covid-19-R-estimates.pdf.