Network-based targeting of interventions

in stochastic SIR epidemic models

Eben Kenah

Departments of Biostatistics and Global Health,
University of Washington, Seattle

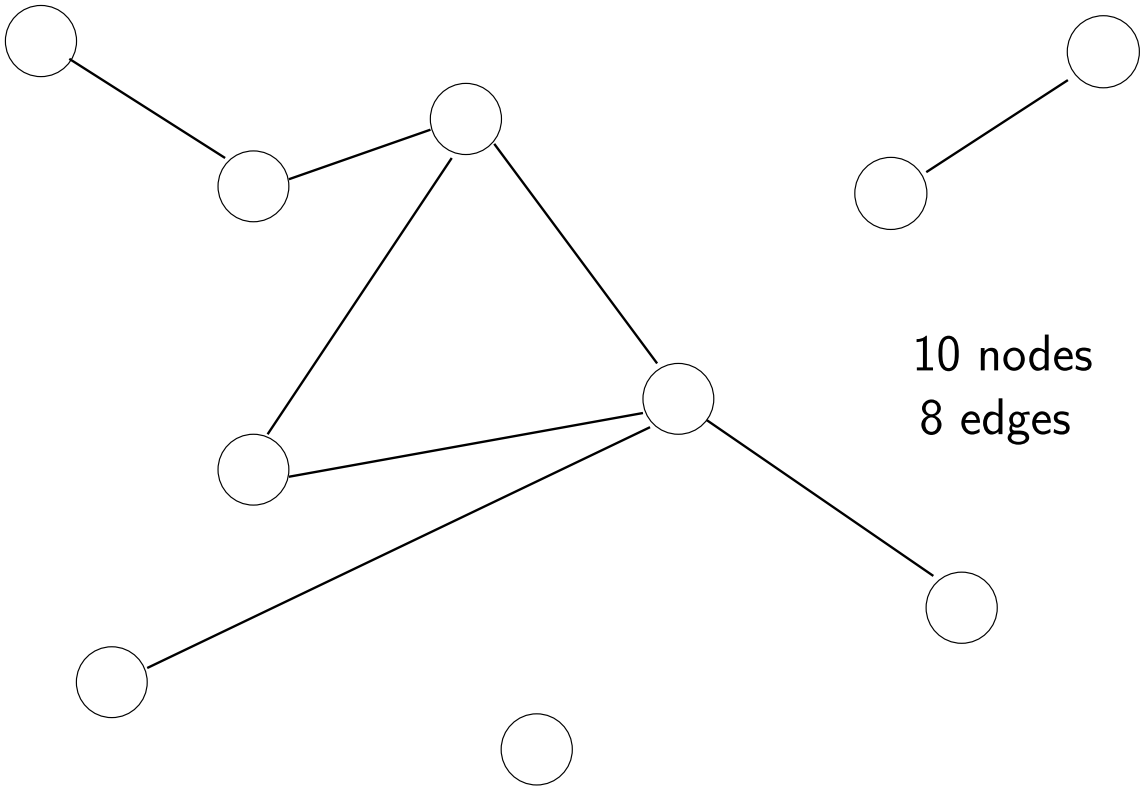
October 22, 2008

Overview

- Outcomes of a stochastic SIR epidemic model can be mapped onto a random directed network that we call the *epidemic percolation network* (EPN).
- The effects of vaccination and other interventions can be modeled by deleting edges from the EPN.
- Disconnection of the giant strongly-connected component of the EPN is a necessary and sufficient condition for the elimination of a disease.

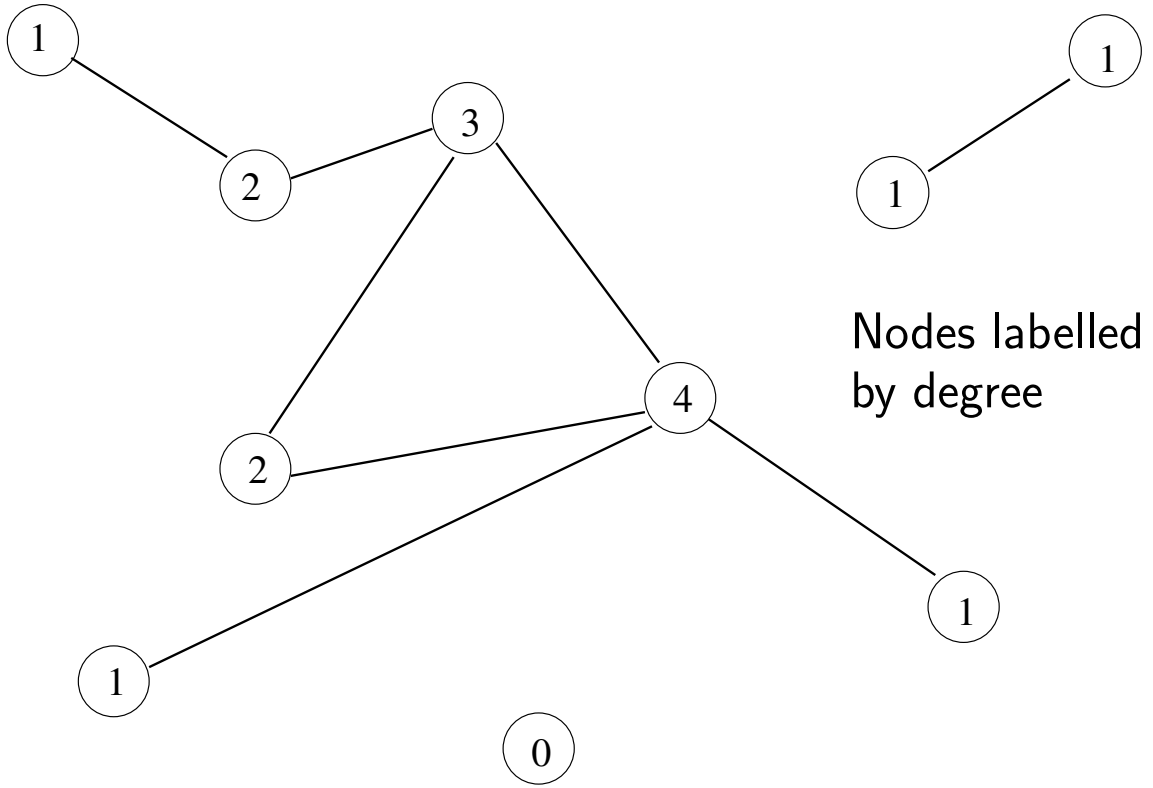
Contact networks

In network-based epidemic models, infection is transmitted across the edges of a *contact network*.



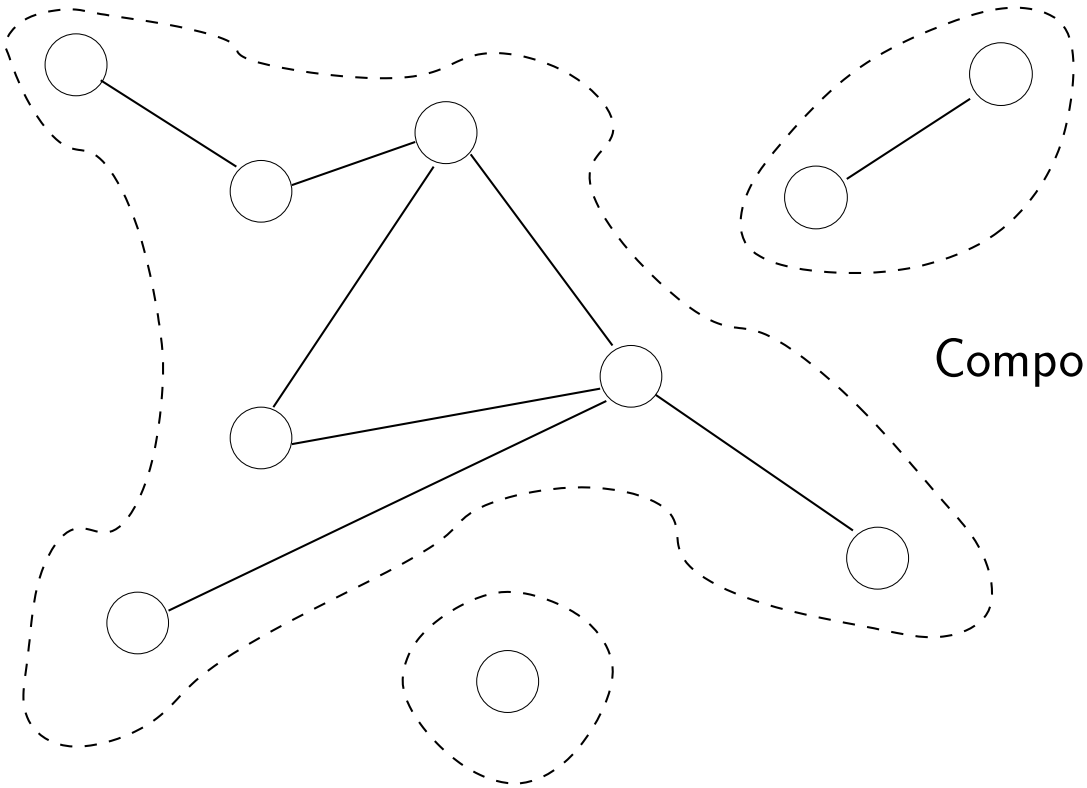
Contact networks

Degree : the number of edges (equivalently, nodes) connected to a node.



Contact networks

Component : a maximal group of nodes in which each node is connected to every other node by a series of edges.



Components circled

Giant components

As we add edges to a large contact network, a unique *giant component* emerges.

- In the limit of a large population, it is the only component that contains a positive proportion of the population.
- In a contact network with no giant component, large epidemics are not possible.

Vaccinating a contact network

The node for a person vaccinated with a perfect vaccine loses all of its edges.

- Transmission to and from that individual is no longer possible.
- By vaccinating enough individuals, we can break apart the giant component of the contact network.
- In general, the most efficient way to do this is to target the nodes with highest degree.[1]

[1] R. Cohen *et al.* [*Phys Rev Lett* **85**, 4626 (2000)] and R. Albert, H. Jeong, and A.-L. Barabási [*Nature* **406**, 378 (2000)]

So what's the problem?

- Should we really have the same vaccination strategy for all diseases spreading on the same network?
- If not, then how do we take disease-specific characteristics into account? Will a tailored vaccination strategy really be more effective?
- What about epidemic models that are not network-based?

To search for an improvement, we begin with a very general stochastic SIR epidemic model...

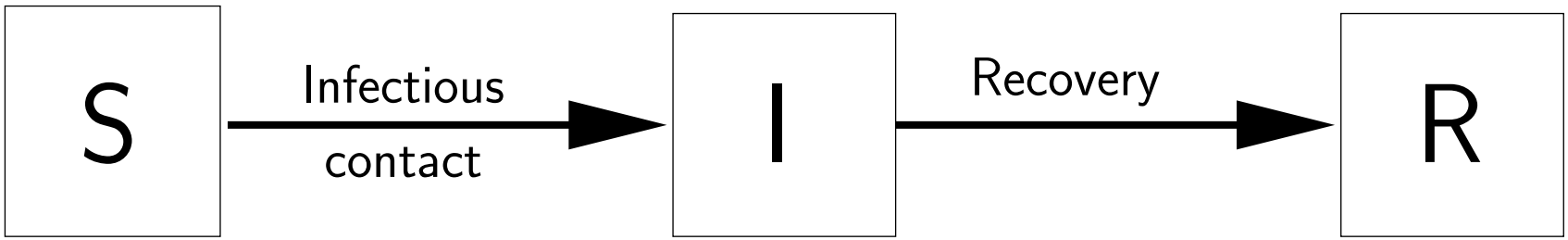
Susceptible-Infected-Removed (SIR)

At any given time, each node exists in one of three states:

Susceptible (S) : can be infected through *infectious contact* with a node in the *I* state.

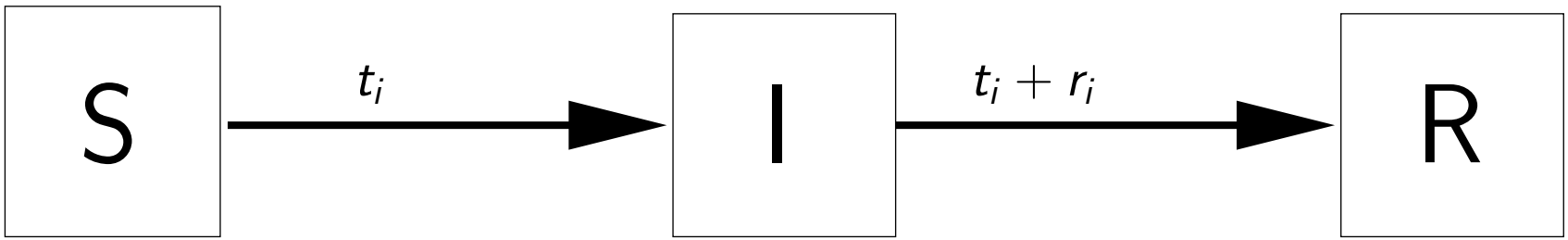
Infectious (I) : can make infectious contact with other nodes.

Removed (R) : can no longer make infectious contact or be infected.



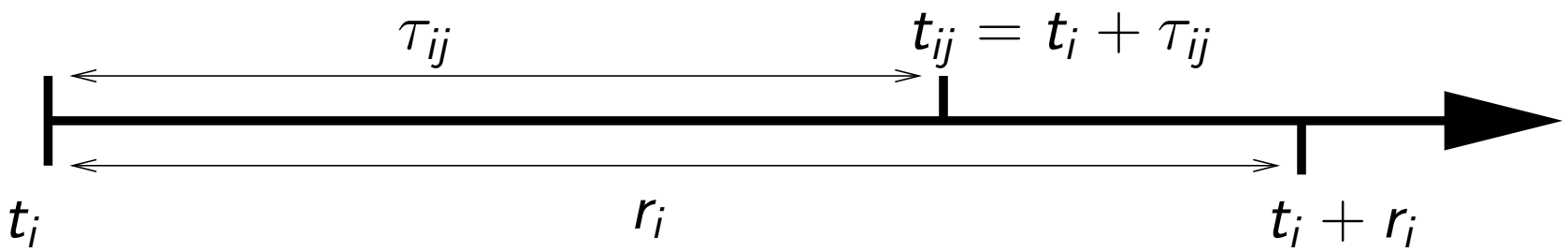
Infection and recovery

1. Node i enters the I state at its *infection time* t_i.
   - $t_i = \infty$ if infection never occurs.
2. Node i enters the R state at its *recovery time* $t_i + r_i$.
   - r_i is a positive random variable called the *recovery period*.
   - $r_i < \infty$ with probability one, so all nodes end up in S or R.



Infectious contact

1. After t_i, node i makes infectious contact with $j \neq i$ after an *infectious contact interval* τ_{ij}.
   - τ_{ij} is a positive random variable, with $\tau_{ij} = \infty$ if infectious contact never occurs.
   - $\tau_{ij} \in (0, r_i)$ or $\tau_{ij} = \infty$.
2. Node j receives infectious contact from i at the *infectious contact time* $t_{ij} = t_i + \tau_{ij}$.



Generality of this SIR model

Specifying (joint) distributions for r_i and τ_{ij} gives us any possible time-homogeneous SIR model:

Stochastic Kermack-McKendrick model[2]

- $r_i \sim$ exponential(μ^{-1})
- $\tau_{ij} \sim$ exponential$(\frac{\beta}{n-1})$ truncated at r_i with remaining probability mass at ∞

Network-based version of Kermack-McKendrick

- $r_i \sim$ exponential(μ^{-1})
- $\tau_{ij} \sim$ exponential(β) truncated at r_i with remaining probability mass at ∞

[2] W. O. Kermack and A. G. McKendrick [*Proc Roy Soc Lond A* **115**, 700-721 (1927)]; reprinted in *Bull Math Biol* **53**, 33-55 (1991).

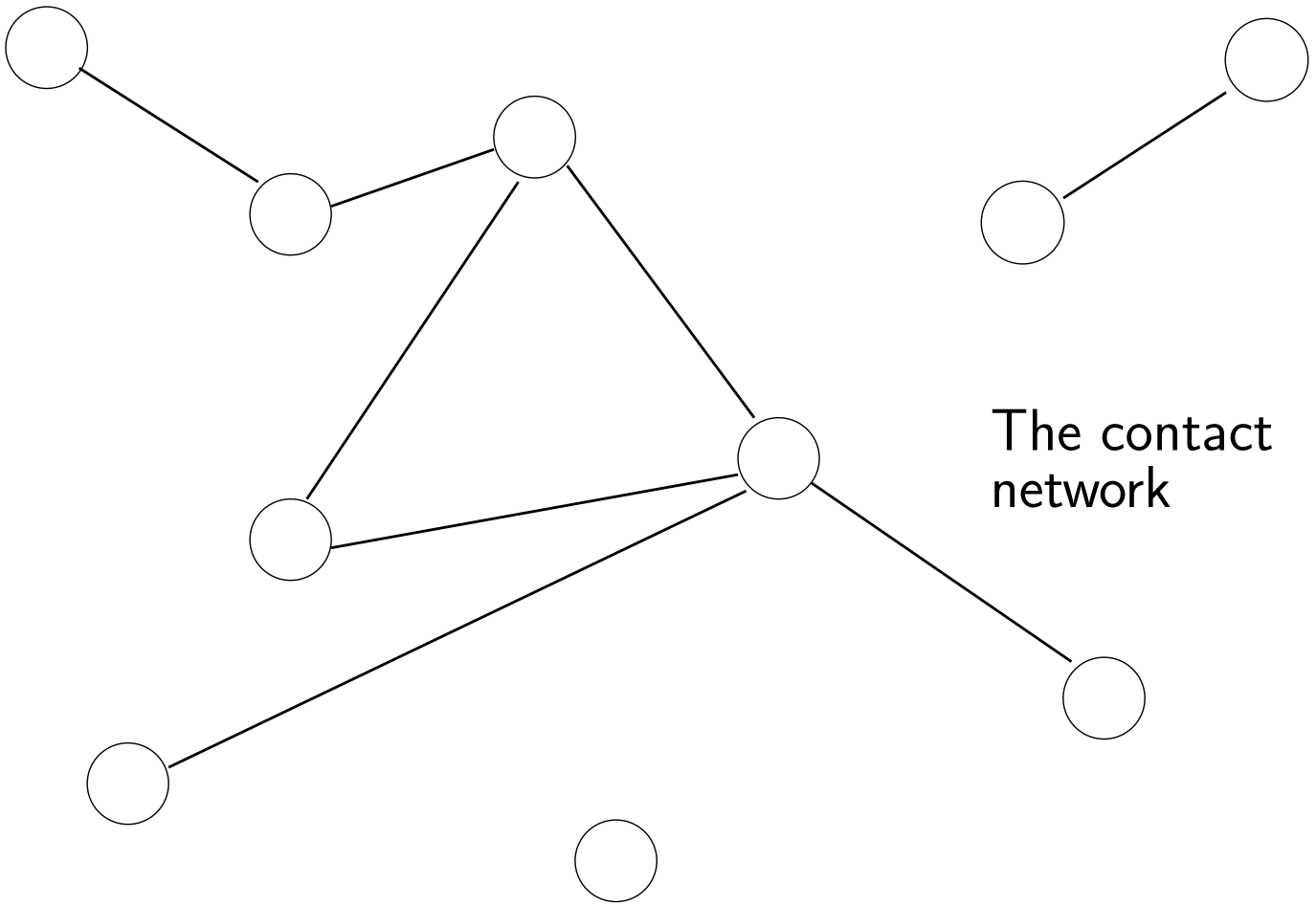
Mapping final outcomes onto networks

In a time-homogeneous model, it does not matter if r_i and τ_{ij} are sampled "on the fly" or *a priori*.

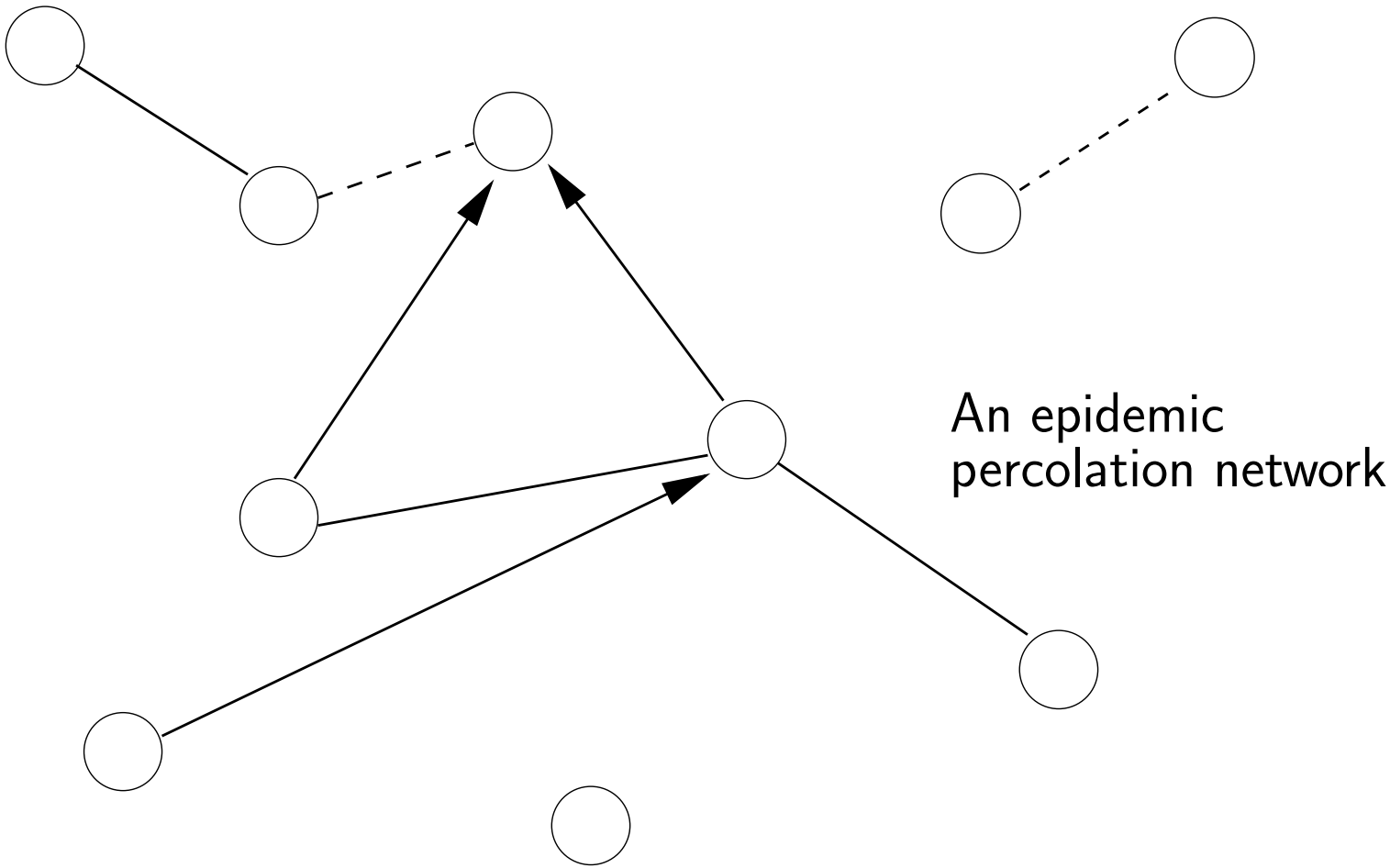
1. Sample $\mathbf{r} = (r_1, \ldots, r_n)$ and then sample $\boldsymbol{\tau} = [\tau_{ij}]_{i,j=1,\ldots,n}$ from its conditional distribution given $\mathbf{r}$.
2. For each ordered pair ij, draw one of the following four edges between nodes i and j:
   - $i \longleftrightarrow j$ if $\tau_{ij} < \infty$ and $\tau_{ji} < \infty$ (infectious contact both ways).
   - $i \longrightarrow j$ if $\tau_{ij} < \infty$ and $\tau_{ji} = \infty$ (infectious contact from i to j).
   - $i \longleftarrow j$ if $\tau_{ij} = \infty$ and $\tau_{ji} < \infty$ (infectious contact from j to i).
   - $i \quad\quad j$ if $\tau_{ij} = \tau_{ji} = \infty$ (no infectious contact).

The directed network with the edge set $\{ij : \tau_{ij} < \infty\}$ is a single realization of the EPN.

EPN examples

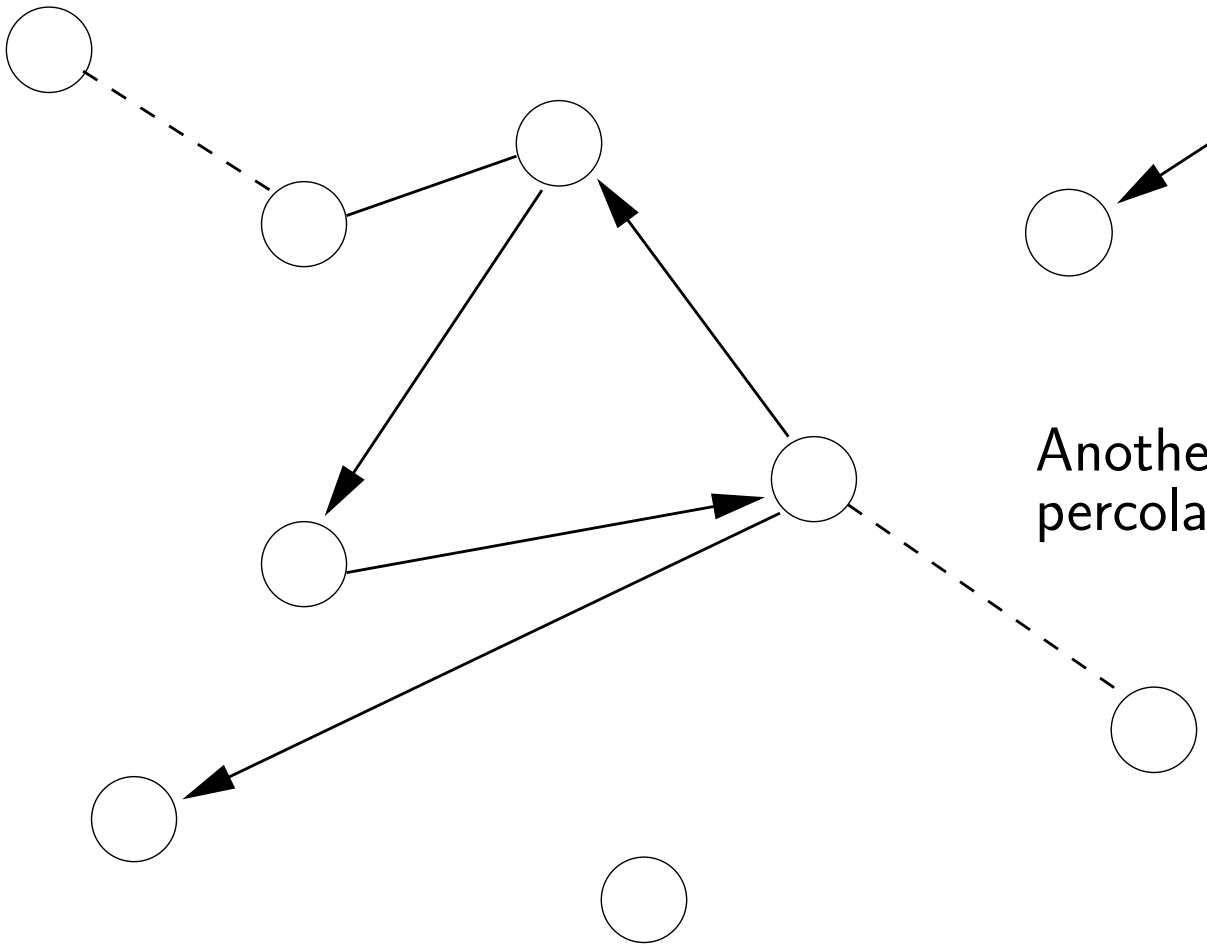


EPN examples



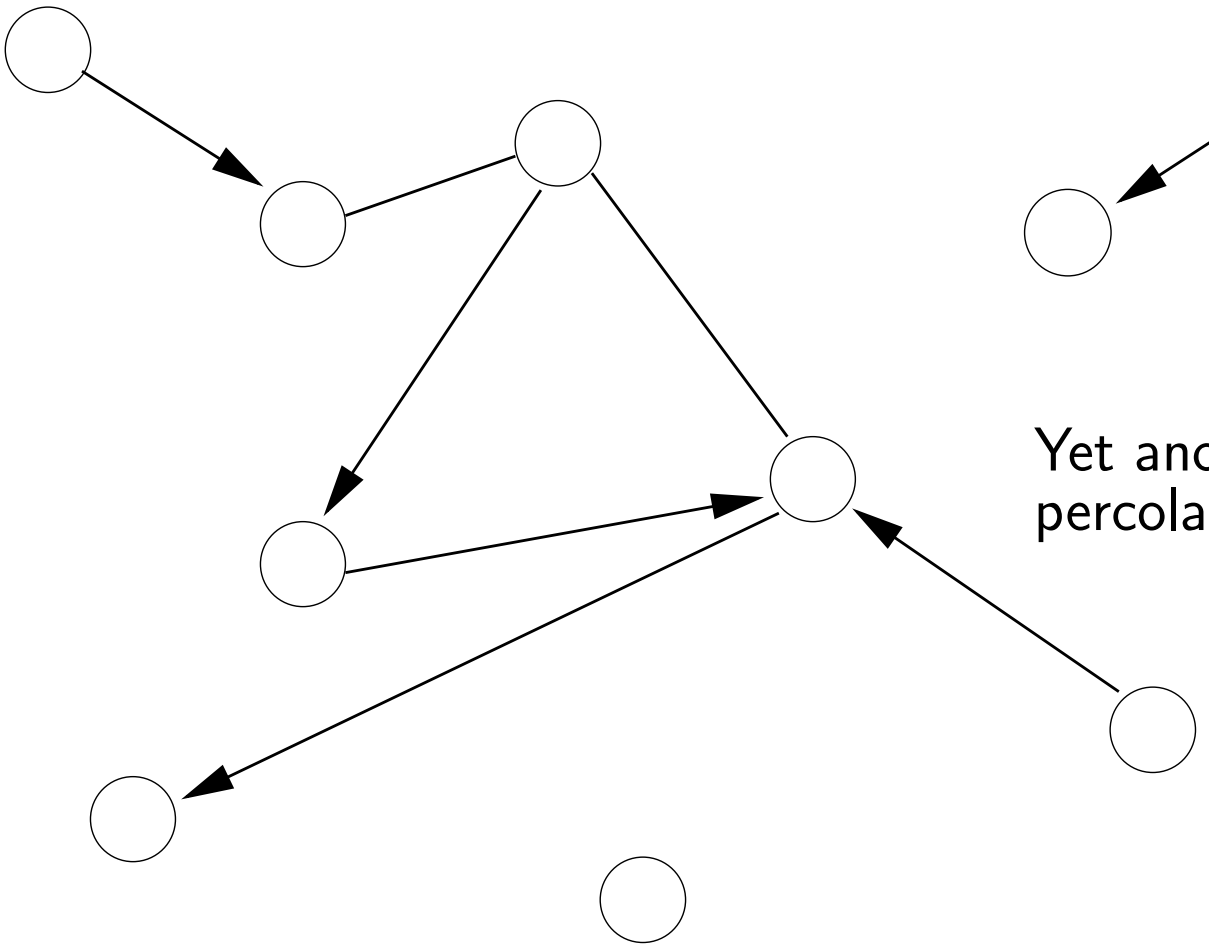
An epidemic percolation network

EPN examples



Another epidemic percolation network

EPN examples

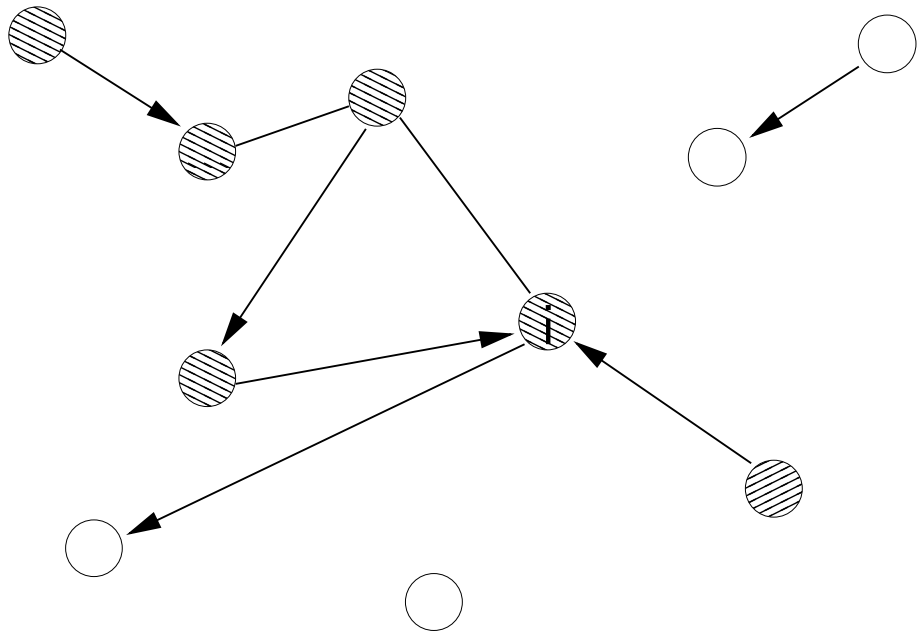


Yet another epidemic percolation network

Components in a directed network

There are three types of components in a directed network:

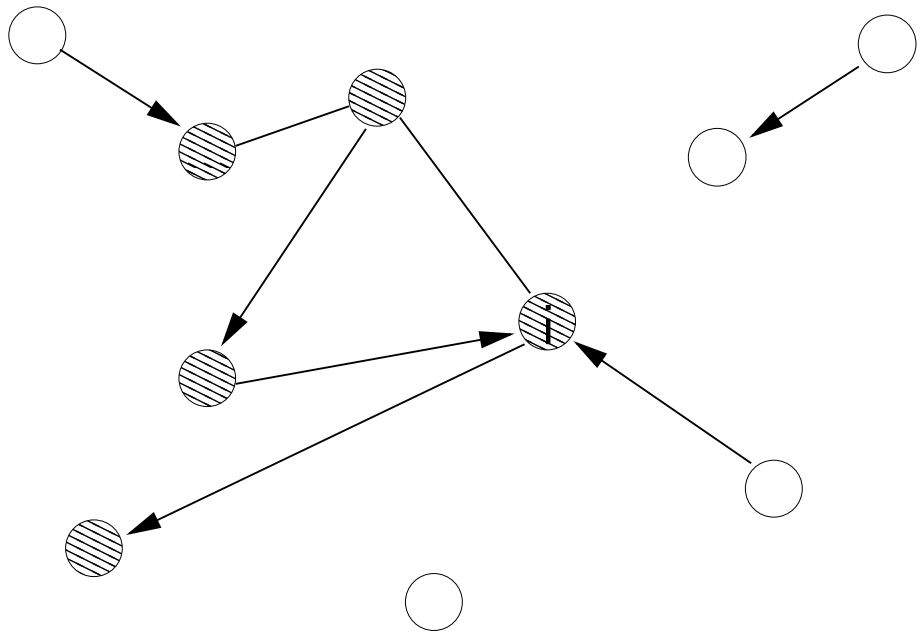
In-component (of node i): the set of nodes from which i can be reached by following edges.



Components in a directed network

There are three types of components in a directed network:

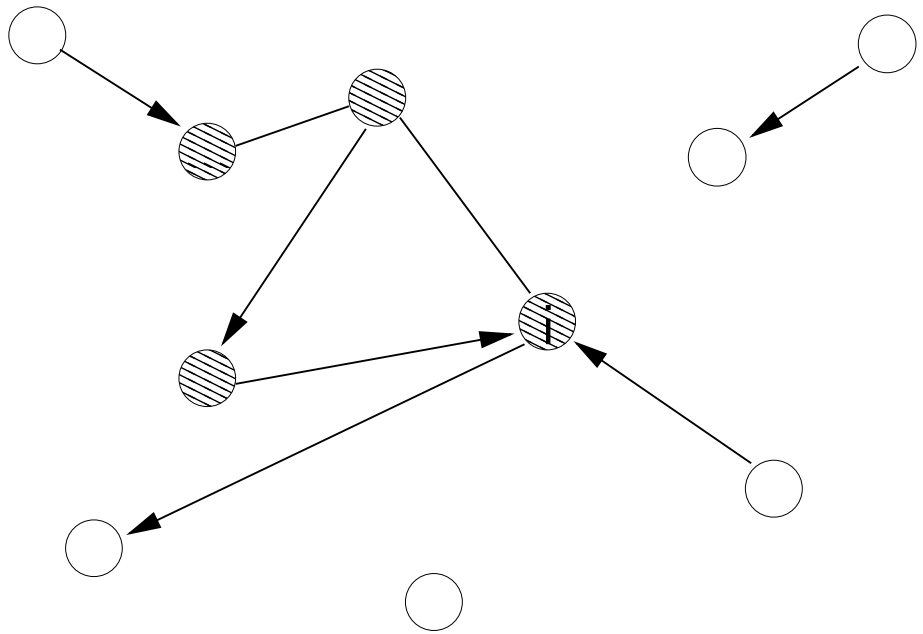
Out-component (of node i): the set of nodes that can be reached from i by following edges.



Components in a directed network

There are three types of components in a directed network:

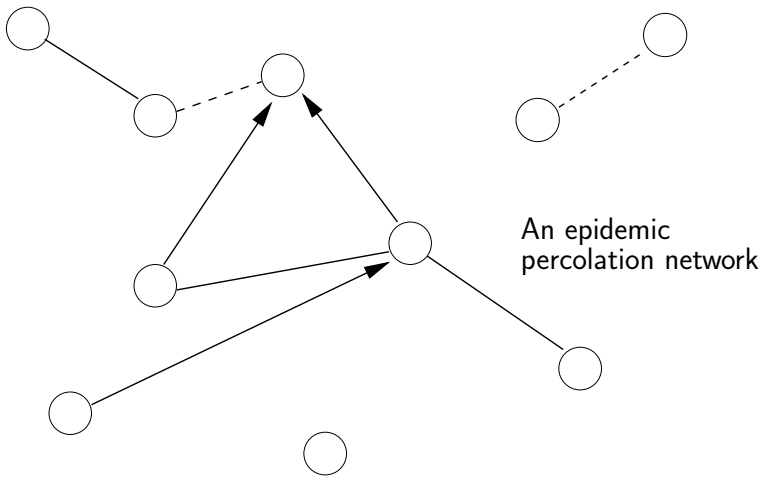
Strongly-connected component (including node i): the intersection of the in- and out-components of i.



Outbreaks and out-components in the EPN

Given **r** and τ, a node is infected eventually if and only if it is in the out-component of an imported infection in the EPN.

$\Rightarrow$ **The distribution of outbreak sizes** starting from person i in a stochastic SIR model **is equal to the distribution of out-component sizes** of node i in the EPN.[3]

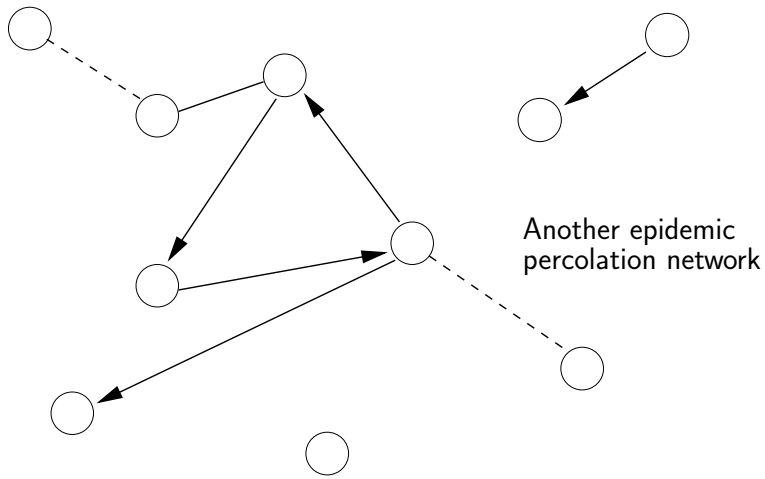


[3] E. Kenah and J. M. Robins [*Phys Rev E* **76**, 036113 (2007)].

Outbreaks and out-components in the EPN

Given **r** and τ, a node is infected eventually if and only if it is in the out-component of an imported infection in the EPN.

$\Rightarrow$ **The distribution of outbreak sizes** starting from person i in a stochastic SIR model **is equal to the distribution of out-component sizes** of node i in the EPN.[3]

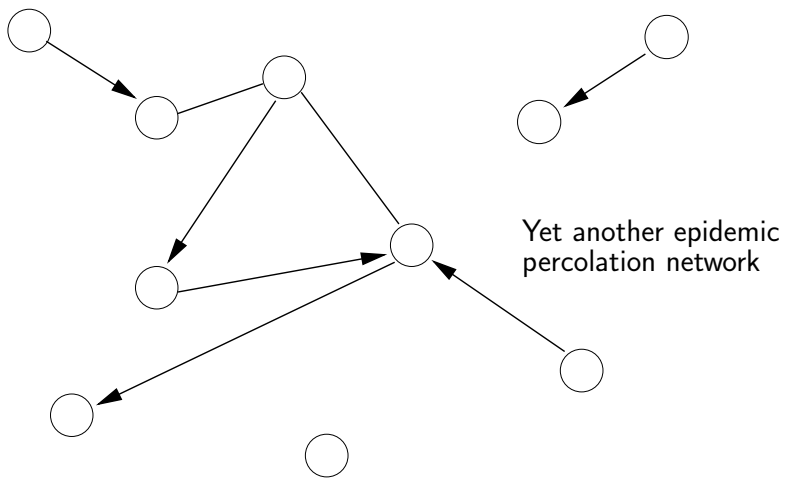


[3] E. Kenah and J. M. Robins [*Phys Rev E* **76**, 036113 (2007)].

Outbreaks and out-components in the EPN

Given **r** and τ, a node is infected eventually if and only if it is in the out-component of an imported infection in the EPN.

$\Rightarrow$ **The distribution of outbreak sizes** starting from person i in a stochastic SIR model **is equal to the distribution of out-component sizes** of node i in the EPN.[3]



[3] E. Kenah and J. M. Robins [*Phys Rev E* **76**, 036113 (2007)].

Phase transition in directed networks

As we add edges to a large directed network, three giant components emerge simultaneously:

Giant strongly-connected component (GSCC): unique largest strongly-connected component

Giant in-component (GIN): in-component of the GSCC.

Giant out-component (GOUT): out-component of the GSCC.

(Note that the nodes in any strongly-connected component all share the same in-component and the same out-component.)

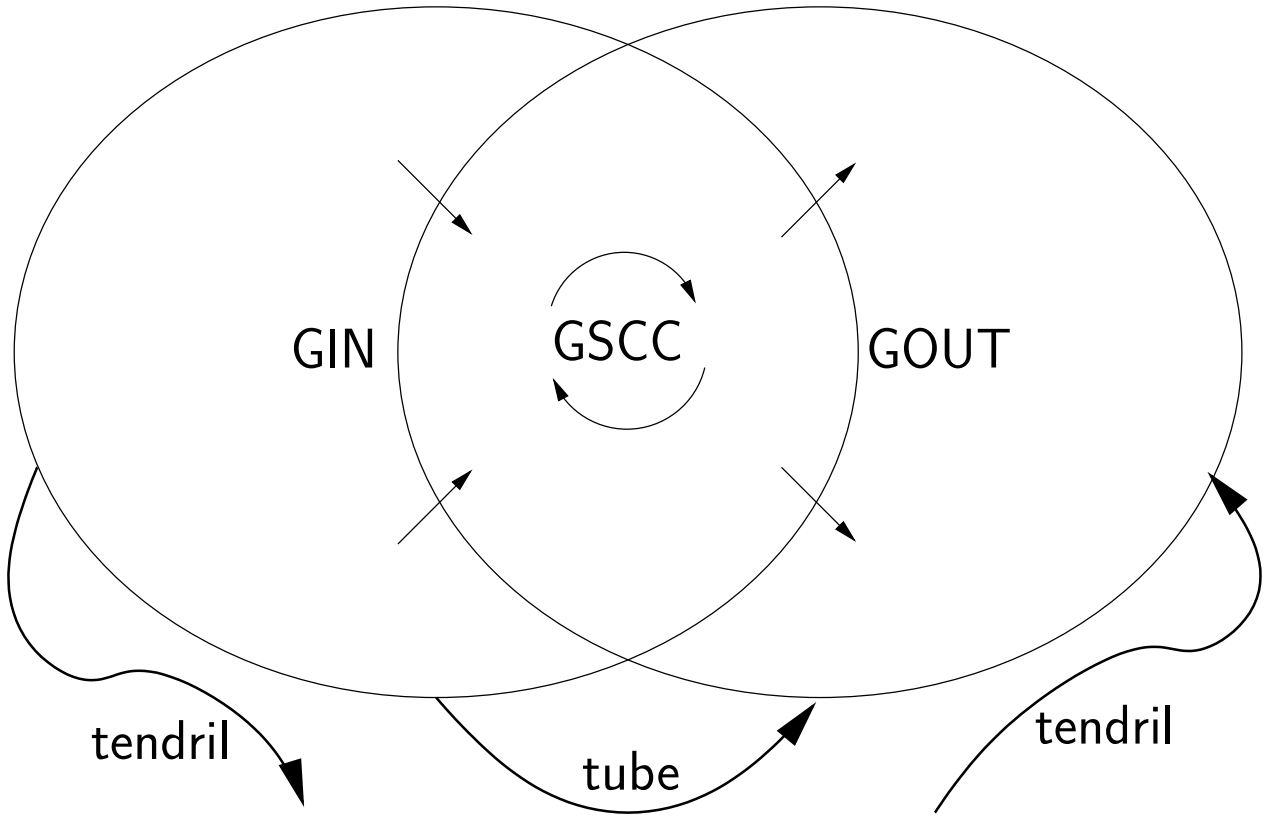
Meaning of the GIN and GOUT

In the limit of a large population, the GIN and the GOUT tell us about the probability and final size of an epidemic:

	SIR model		**EPN**
Pr(infection of i starts an epidemic)		=	Pr(node i is in the GIN)
Pr(i is infected in an epidemic)		=	Pr(node i is in the GOUT)

But what is the meaning of the GSCC?

"Bow-tie" schematic[4]



[4] Adapted from A. Broder *et al.* [*Comput Netw* **33**, 309 (2000)] and S. N. Dorogovtsev *et al.* [*Phys Rev E* **64**, 025101(R) (2001)]

Vaccination and the GSCC

If we break apart the GSCC by vaccinating nodes, then no large epidemics can occur.

- Disconnecting the GSCC is *necessary* and *sufficient* for driving the population below the epidemic threshold.
- Applies to network-based, fully-mixed, and all other time-homogeneous stochastic SIR models.

Targeting the GSCC

The most efficient way to disconnect an undirected network is by "vaccinating" nodes with the highest degree. By analogy, we consider the following method of targeting vaccination:

1. Generate an EPN and erase all edges except those between nodes within the GSCC.
2. Turn all remaining edges (i.e., edges between nodes in the GSCC) into undirected edges.
3. Target the nodes with the highest degree in the resulting network.[5]

[5] More generally, target the nodes with the highest *expected* degree as a result of this process in the probability space of EPNs.

Network-based models[6]

Strategies compared

In a series of network-based models, we made two ranked lists of vaccination targets: One *by contact network degree* and another *by degree within the GSCC* in a single realization of the EPN. We consider the effects of:

- Different degree distributions in the contact network
- Increasing heterogeneity in infectiousness and susceptibility
- Positive, negative, and zero correlation between infectiousness and susceptibility

We look at the probability and final size of an epidemic versus the vaccination fraction under each strategy.

[6] **This work was done with Joel C. Miller at Los Alamos National Laboratory.**

Network-based models [7]

We studied models on two different contact networks:

Erdős-Rényi network with mean degree 5 ($p_k = \frac{5^k}{k!}e^{-5}$).

Scale-free network with $\alpha = 2$ and an exponential cutoff around 50 ($p_k = k^{-2}e^{-\frac{k}{50}}$).

For neighbors i and j in the contact network, the probability of transmission from i to j was

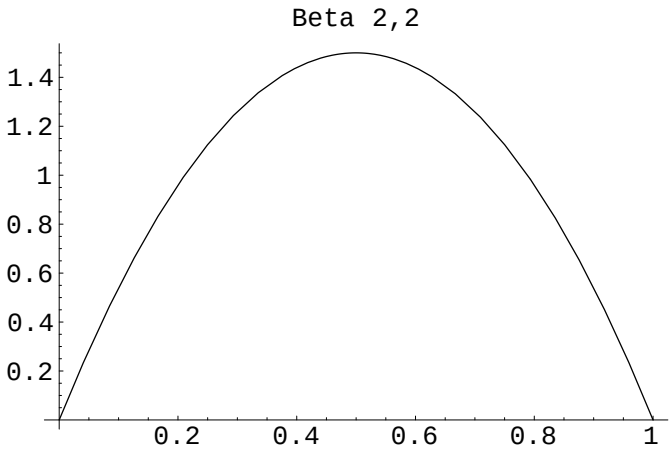
$$1 - e^{-100 * inf_i * sus_j},$$

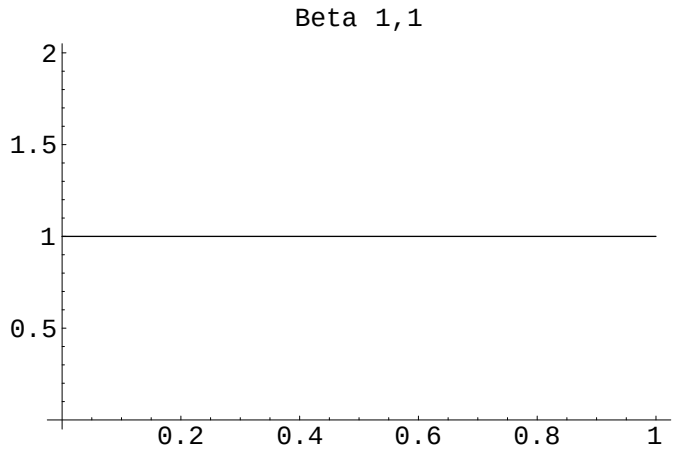
where inf_i, sus_i were drawn from a beta distribution.

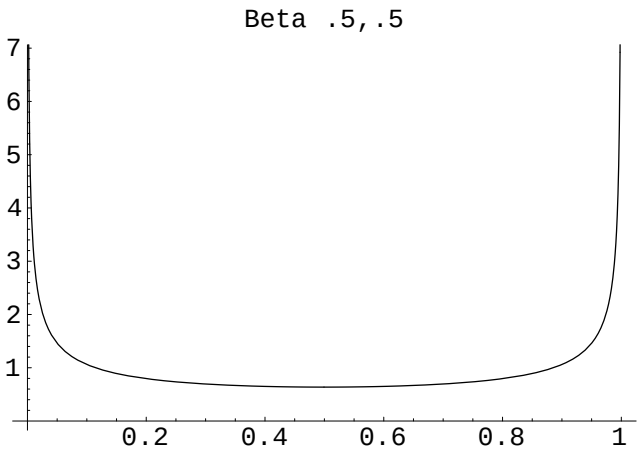
[7] Simulations implemented in Python 2.5.1 (www.python.org) using the NetworkX package (networkx.lanl.gov).

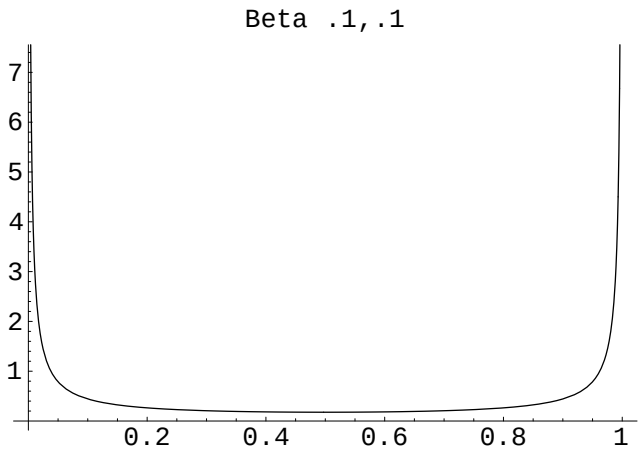
Heterogeneity

Beta distributions for infectiousness and susceptibility









Correlations

We used the following relationships between inf_i and sus_i to obtain independent or correlated infectiousness and susceptibility:

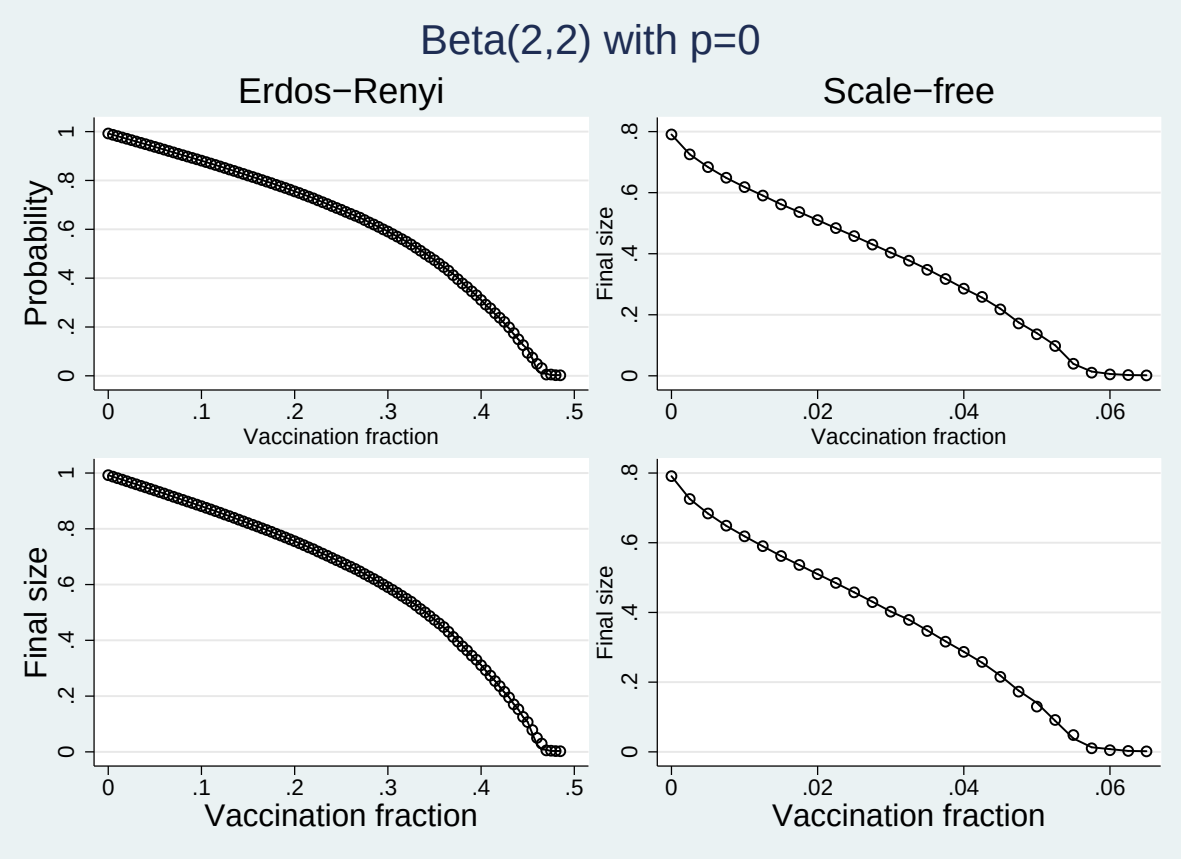
Independent: inf_i and sus_i are independent draws from the same beta distribution

Positive correlation: $inf_i = sus_i$

Negative correlation: $inf_i = 1 - sus_i$

Simulation results [8]

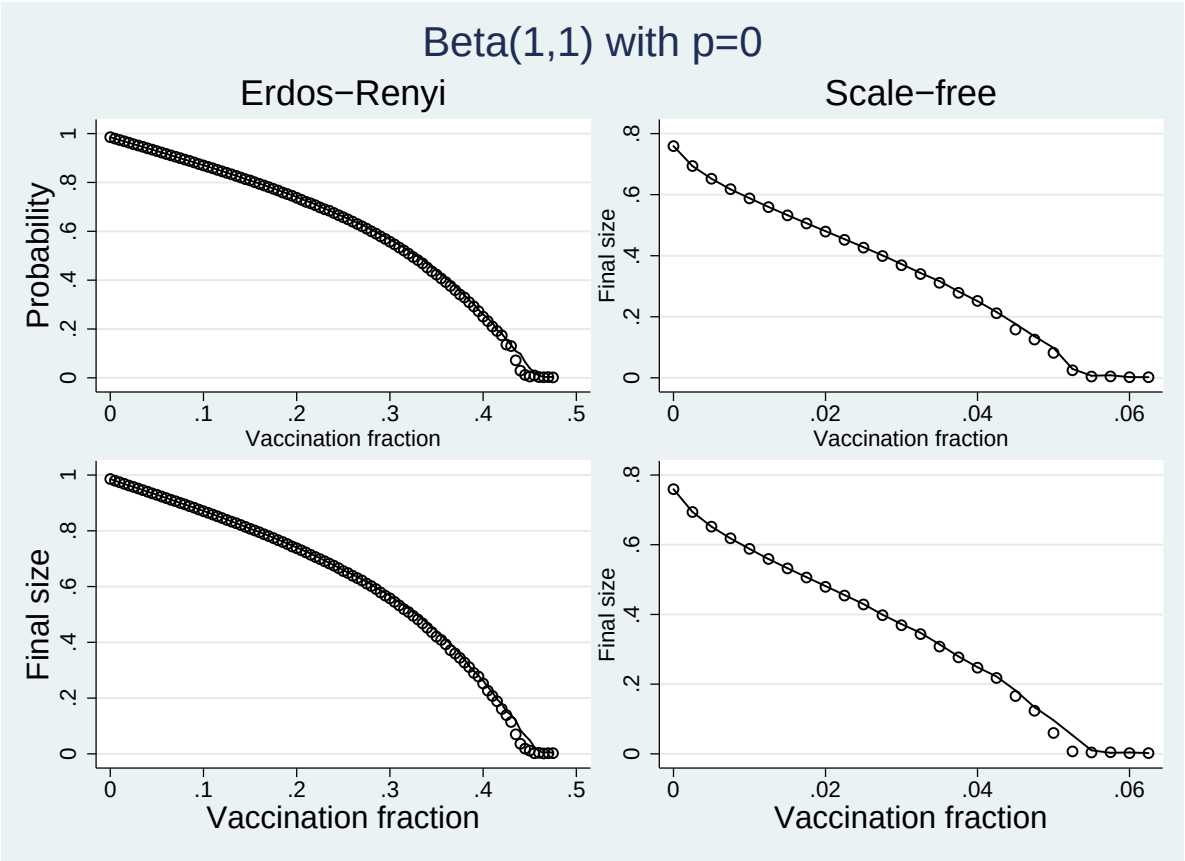
Independent infectiousness and susceptibility



[8] Lines represent targeting by contact network degree; circles represent targeting by degree within the GSCC. Graphs produced in Stata 9.2 (©StataCorp LP).

Simulation results [8]

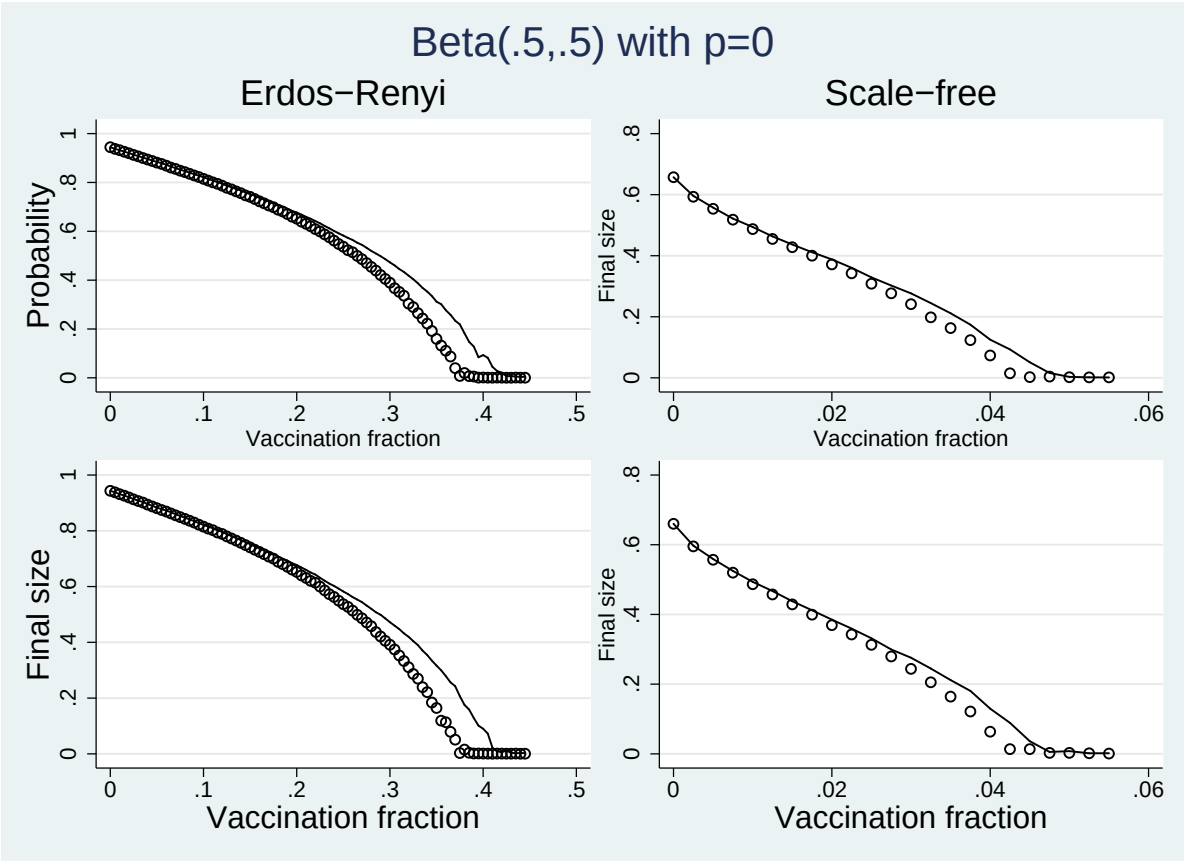
Independent infectiousness and susceptibility



[8] Lines represent targeting by contact network degree; circles represent targeting by degree within the GSCC. Graphs produced in Stata 9.2 (©StataCorp LP).

Simulation results [8]

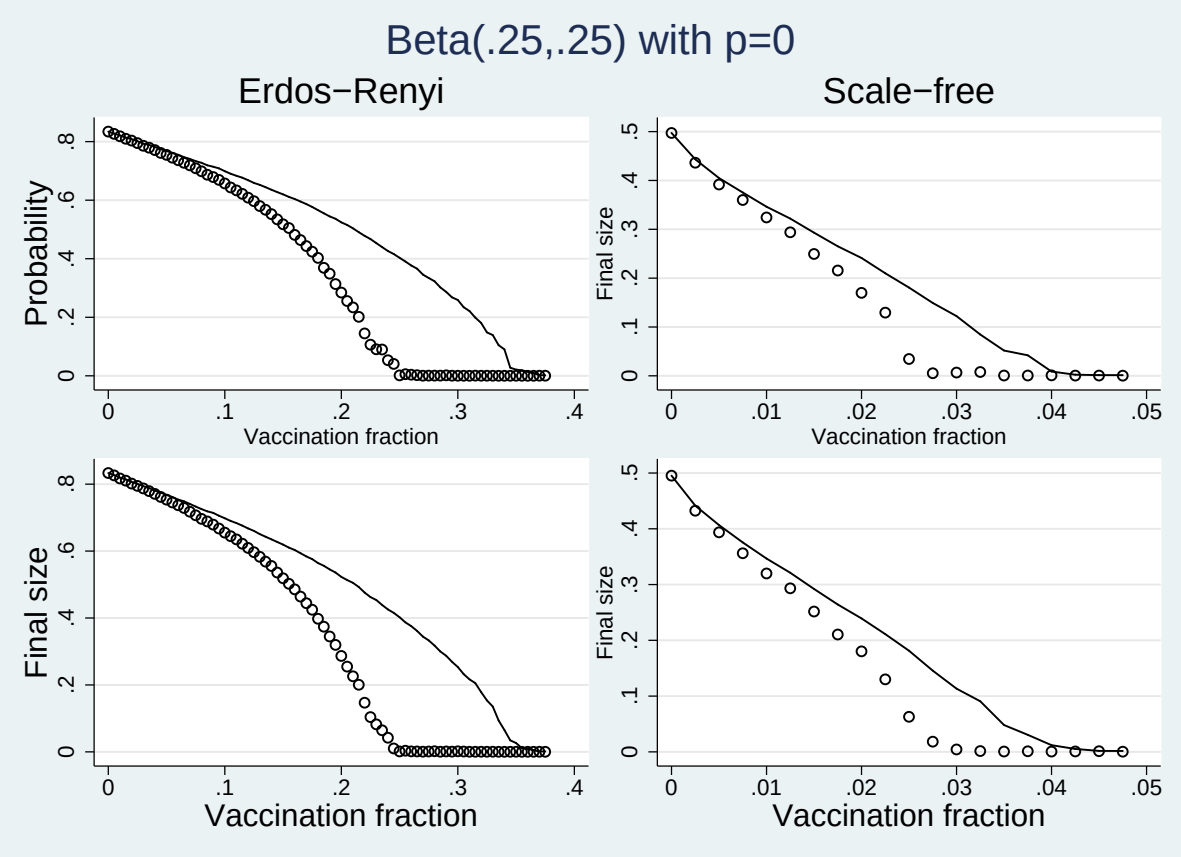
Independent infectiousness and susceptibility



[8] Lines represent targeting by contact network degree; circles represent targeting by degree within the GSCC. Graphs produced in Stata 9.2 (©StataCorp LP).

Simulation results [8]

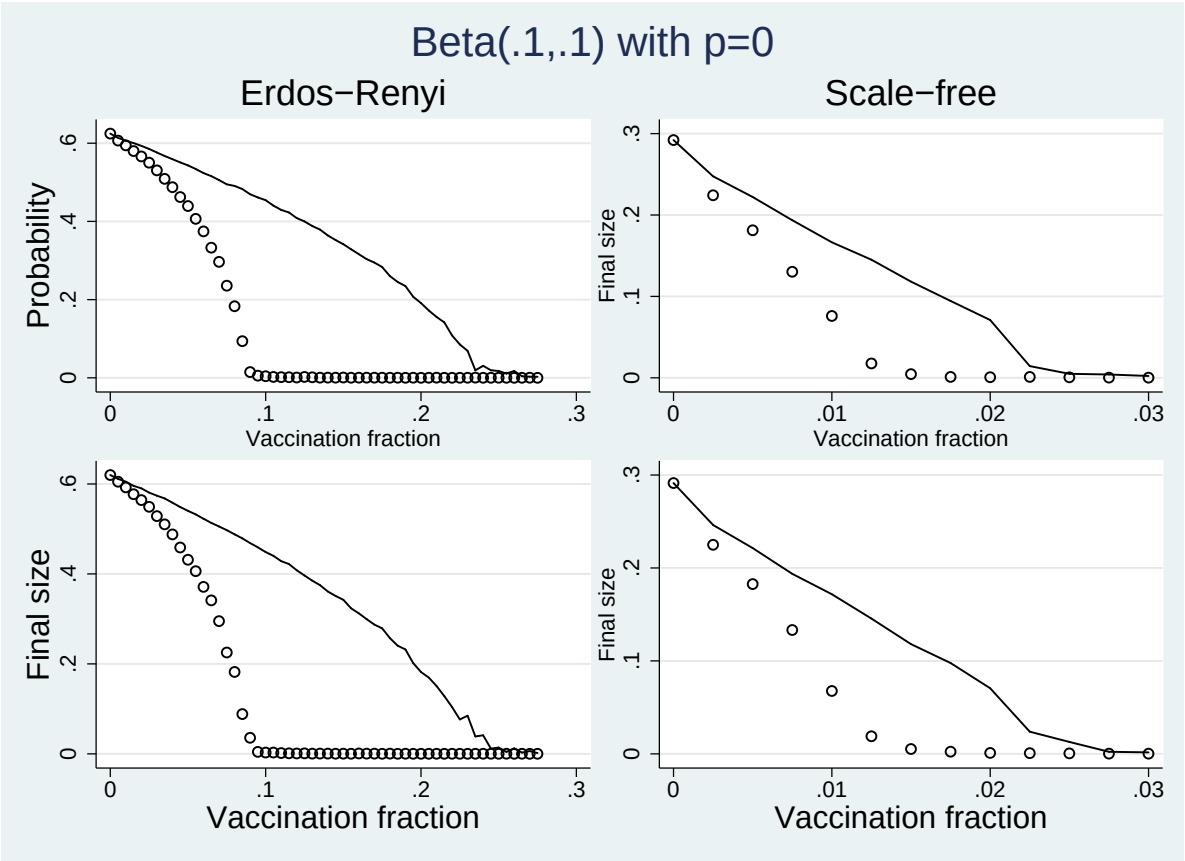
Independent infectiousness and susceptibility



[8] Lines represent targeting by contact network degree; circles represent targeting by degree within the GSCC. Graphs produced in Stata 9.2 (©StataCorp LP).

Simulation results [8]

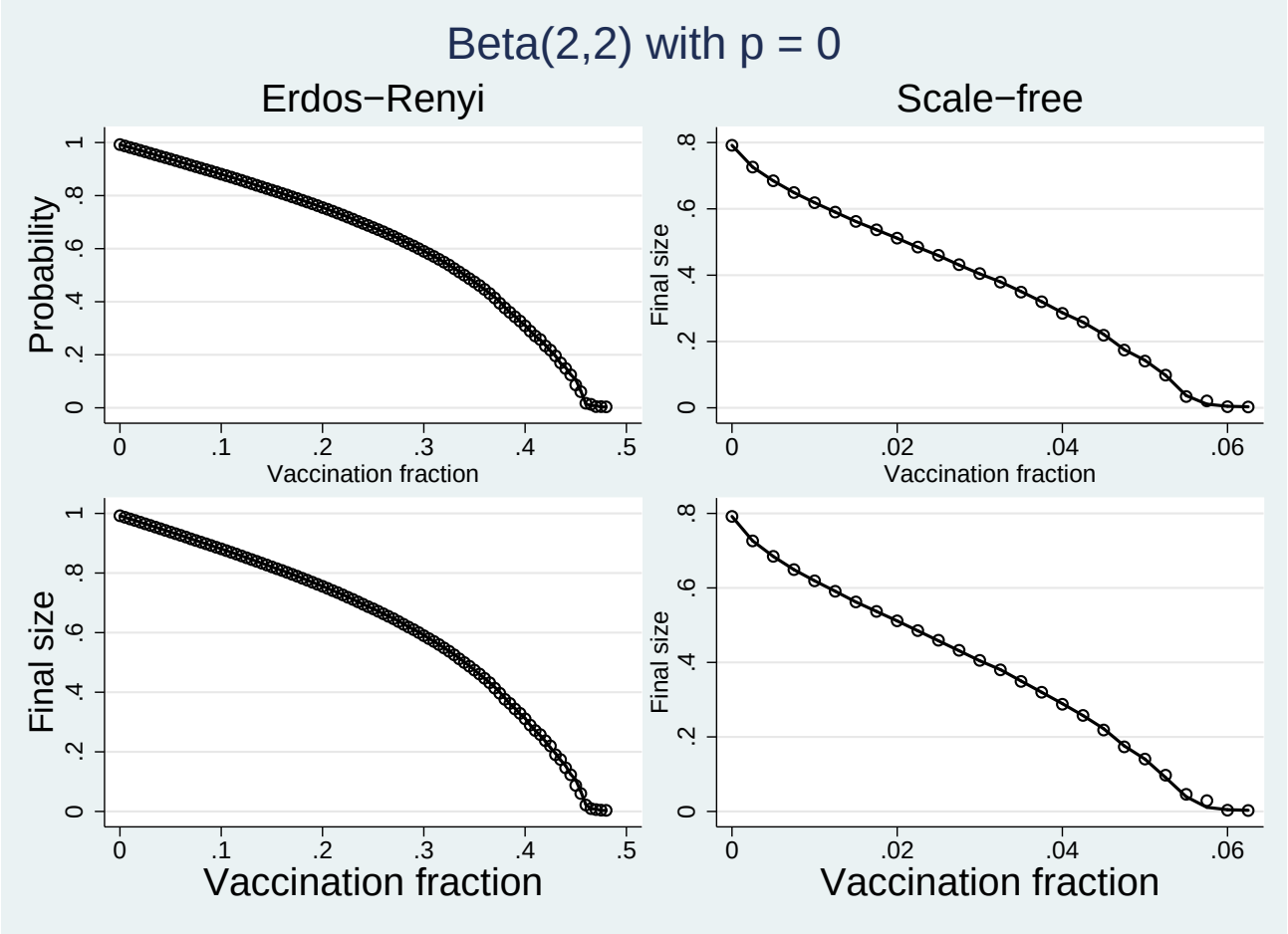
Independent infectiousness and susceptibility



[8] Lines represent targeting by contact network degree; circles represent targeting by degree within the GSCC. Graphs produced in Stata 9.2 (©StataCorp LP).

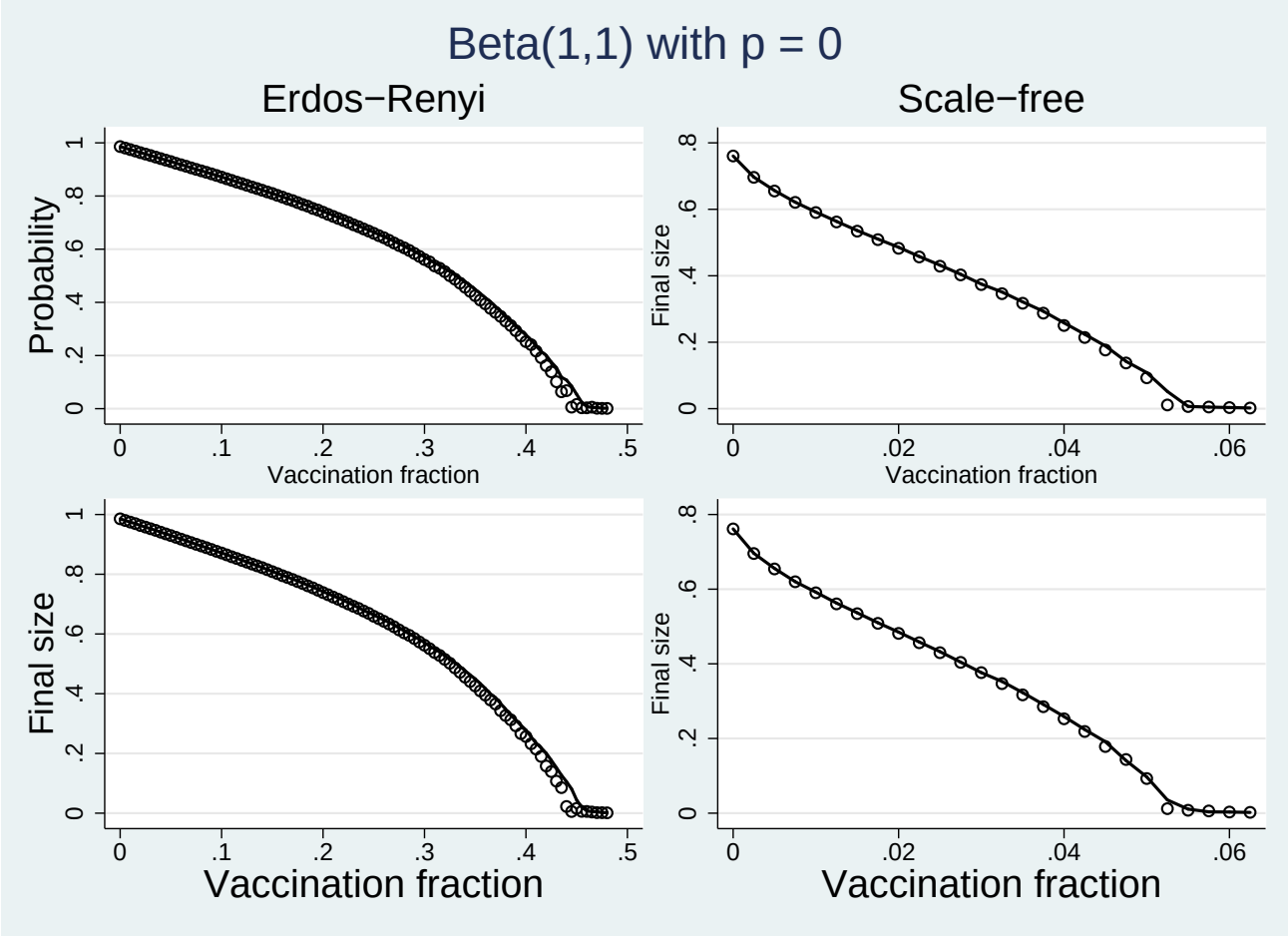
Simulation results

Positively correlated infectiousness and susceptibility



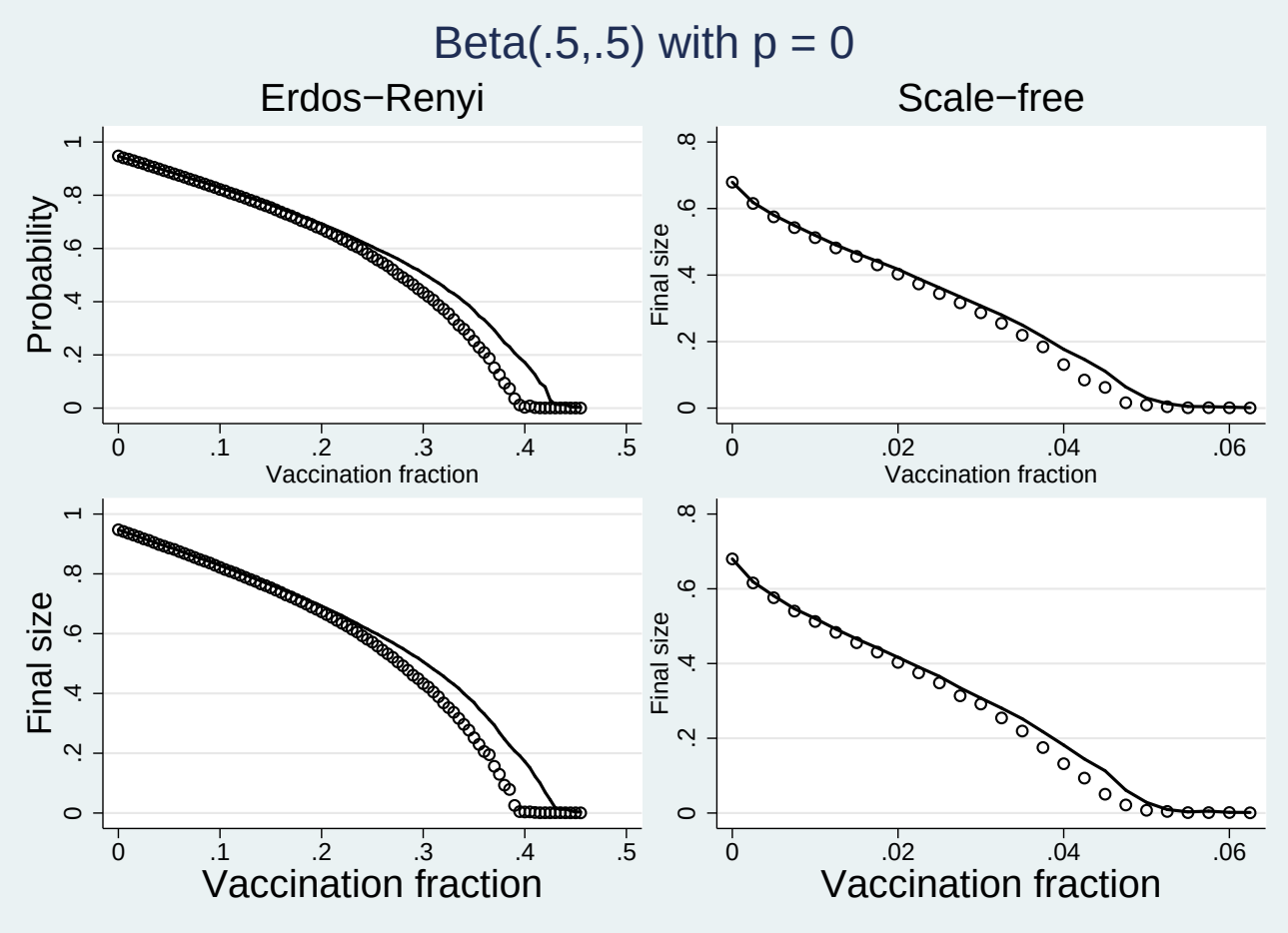
Simulation results

Positively correlated infectiousness and susceptibility



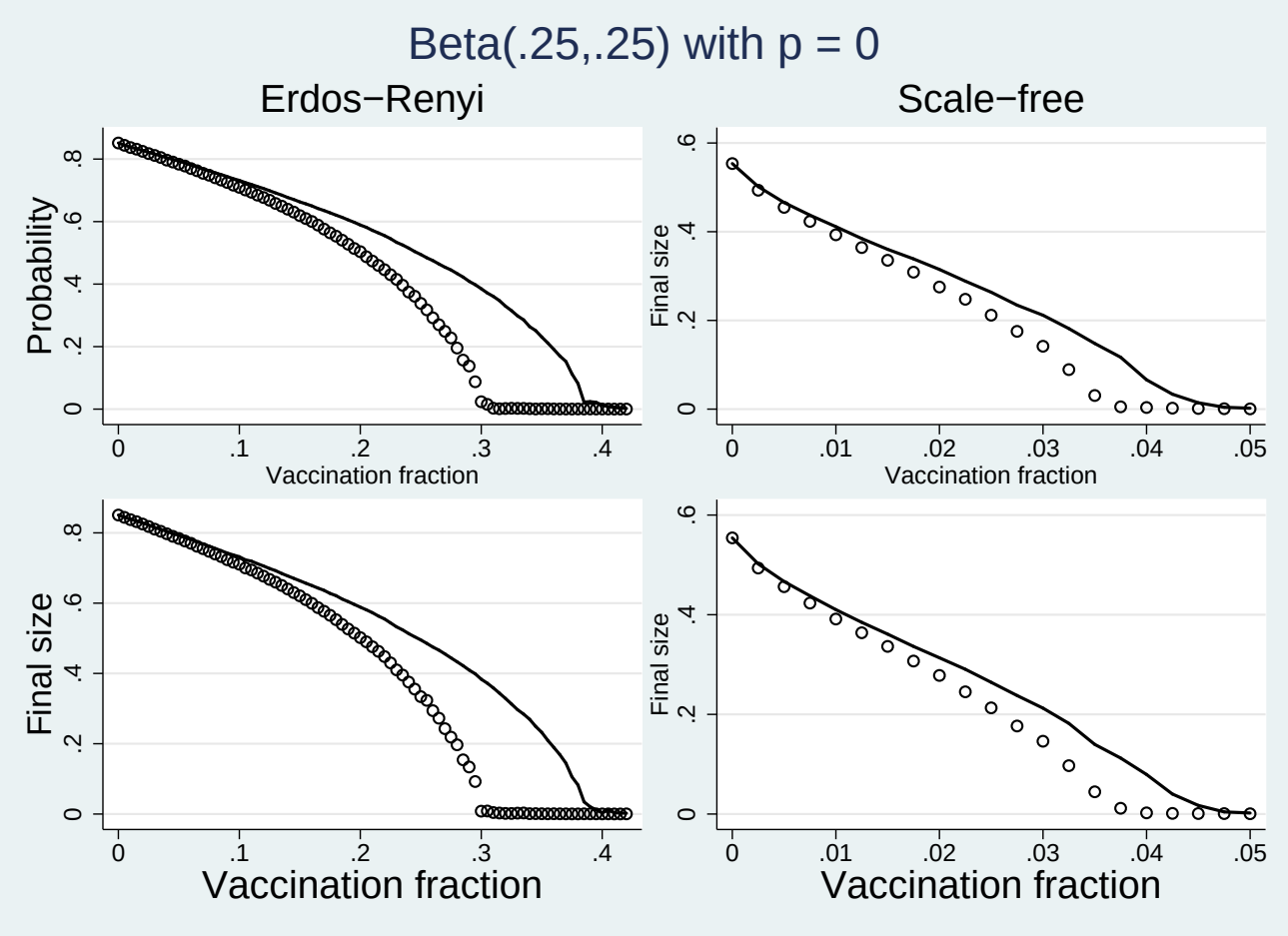
Simulation results

Positively correlated infectiousness and susceptibility



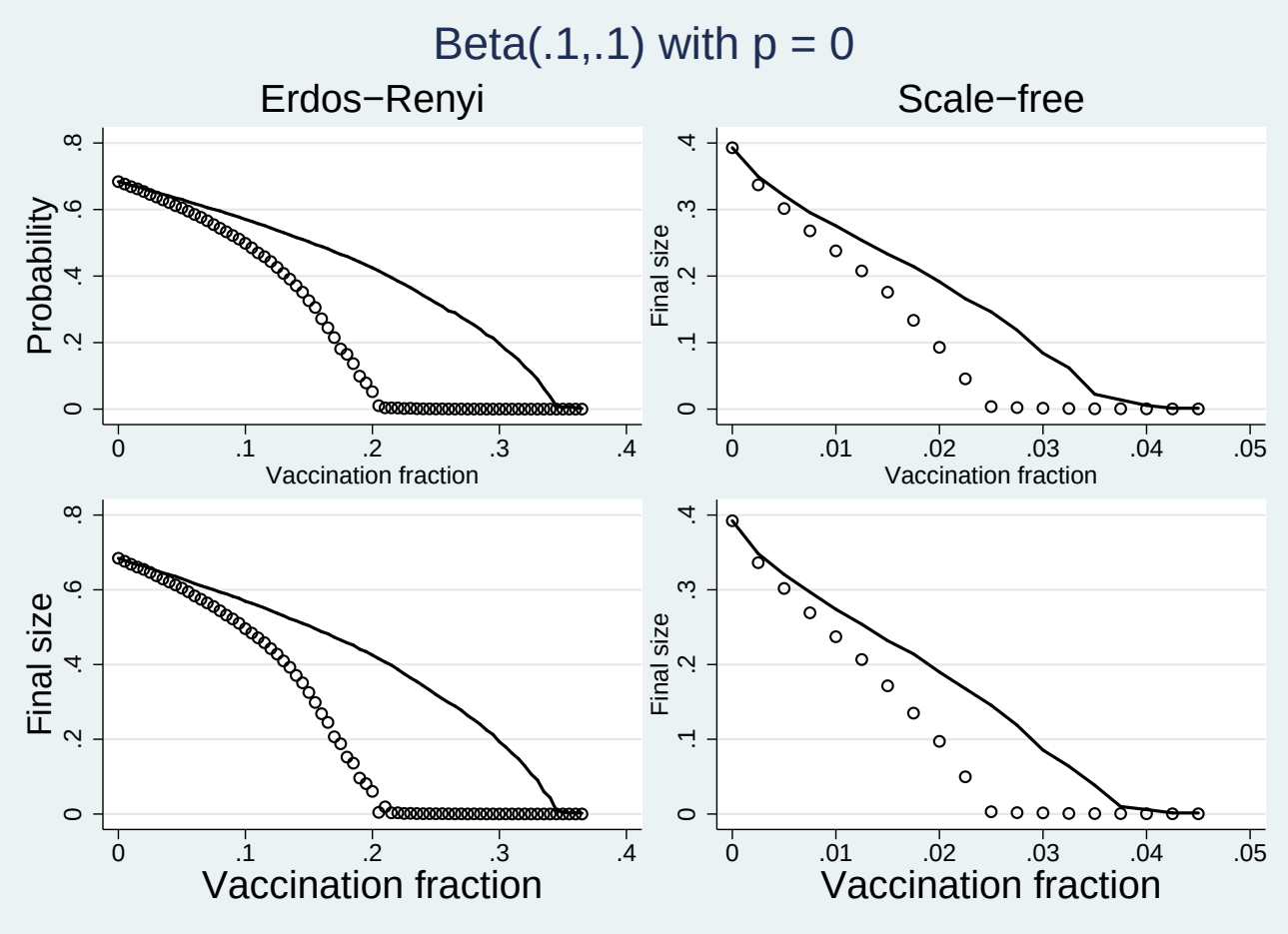
Simulation results

Positively correlated infectiousness and susceptibility



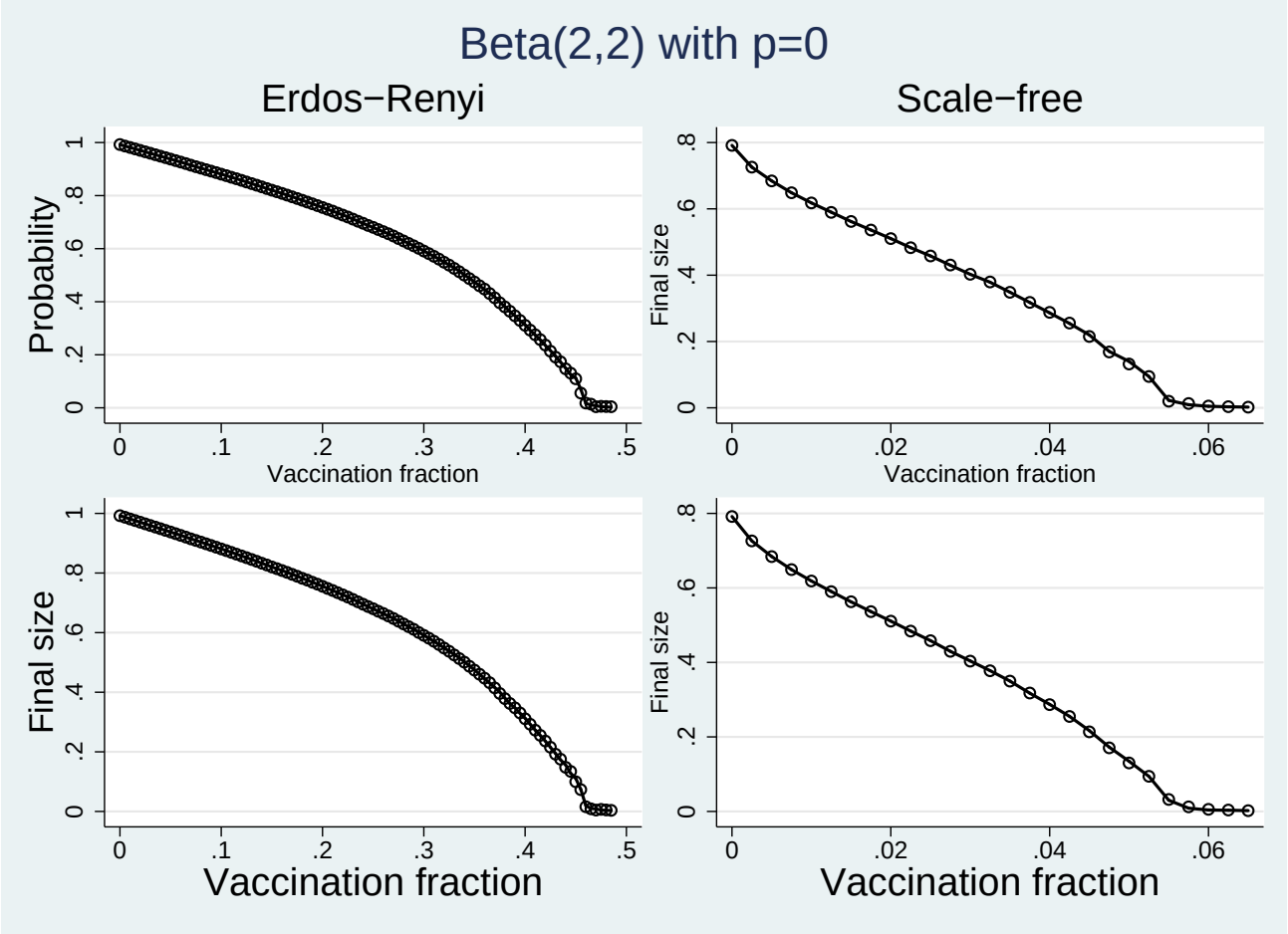
Simulation results

Positively correlated infectiousness and susceptibility



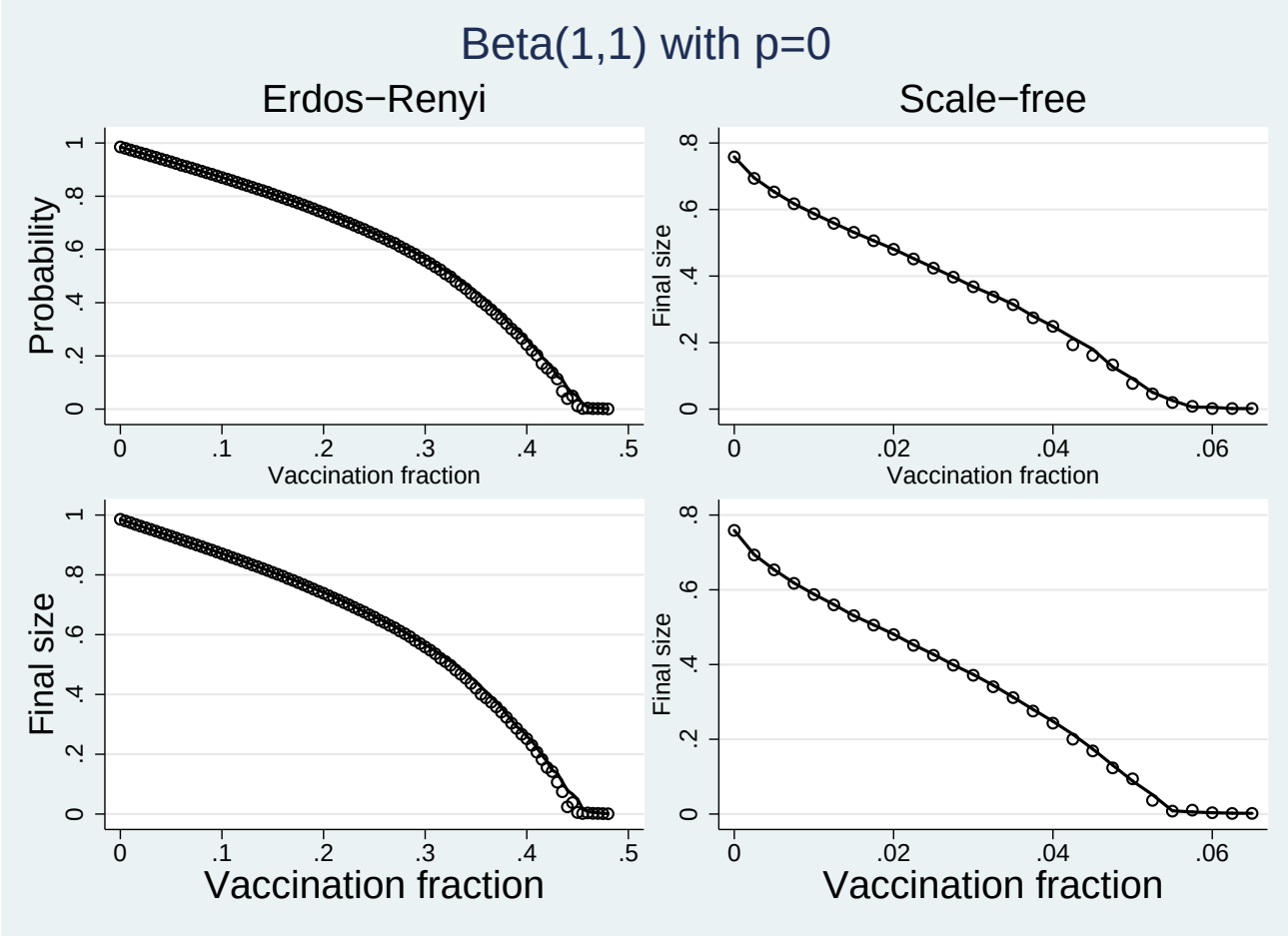
Simulation results

Negatively correlated infectiousness and susceptibility



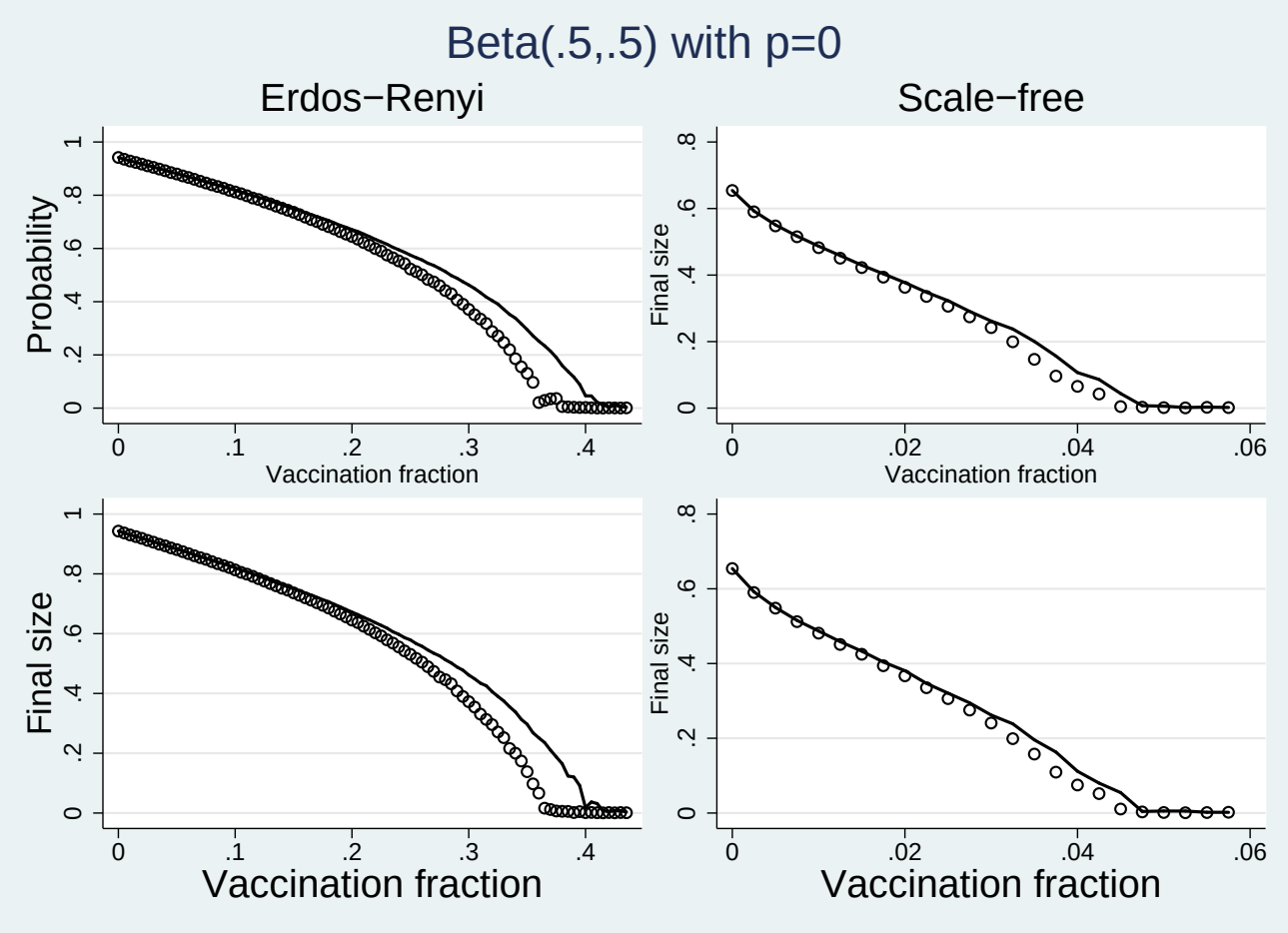
Simulation results

Negatively correlated infectiousness and susceptibility



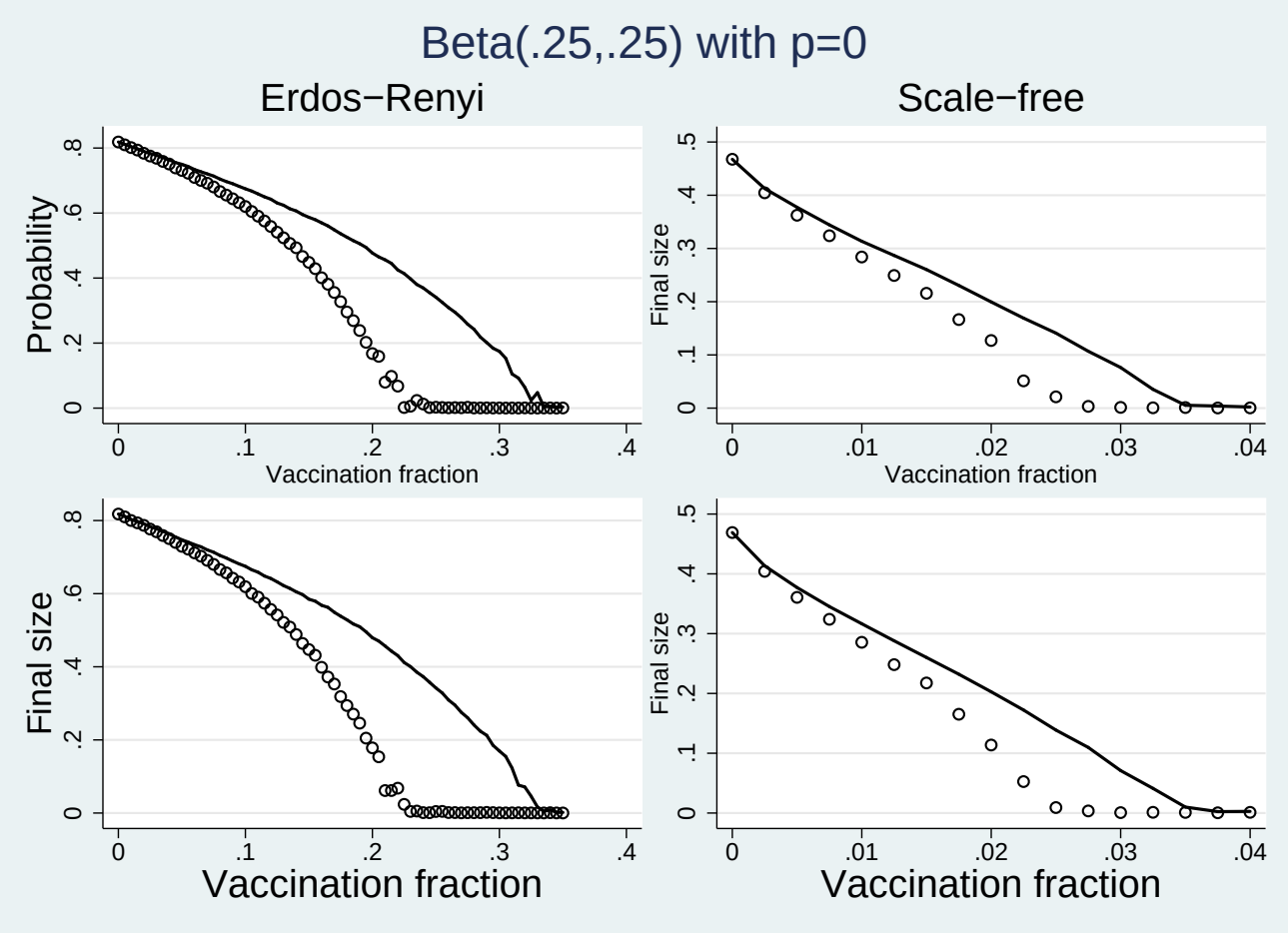
Simulation results

Negatively correlated infectiousness and susceptibility



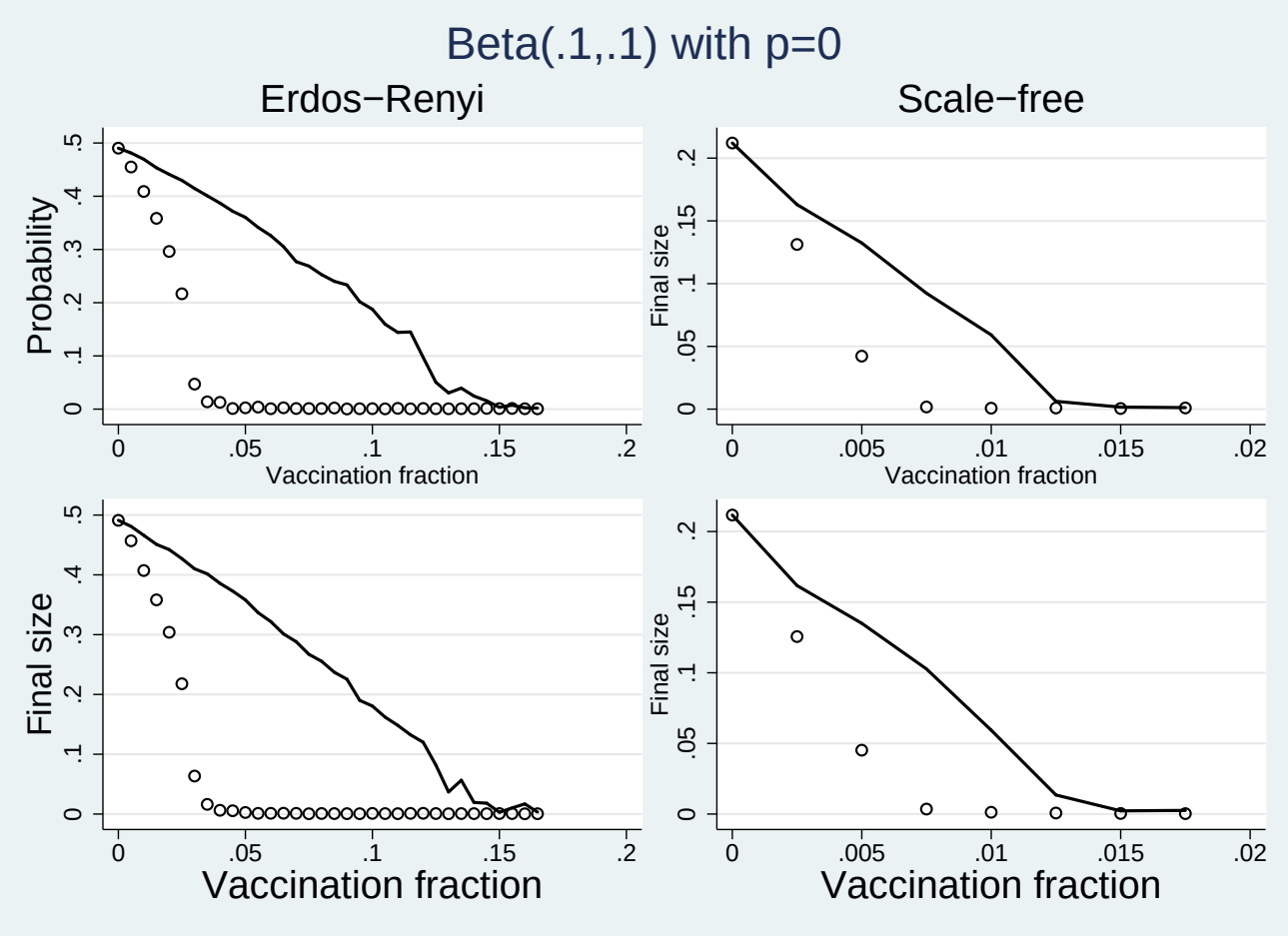
Simulation results

Negatively correlated infectiousness and susceptibility



Simulation results

Negatively correlated infectiousness and susceptibility



Fully-mixed models

Strategies compared

In a fully-mixed model with three equal subpopulations:

Subpopulation A has high infectiousness but low susceptibility, so it has the highest probability of being in the GIN.

Subpopulation B has average infectiousness and susceptibility but the highest probability of being in the GSCC.

Supopulation C has low infectiousness but high susceptibility, so it has the highest probability of being in the GOUT.

We look at the probability and final size of an epidemic versus the vaccination fraction in each subpopulation.

Fully-mixed model [9]

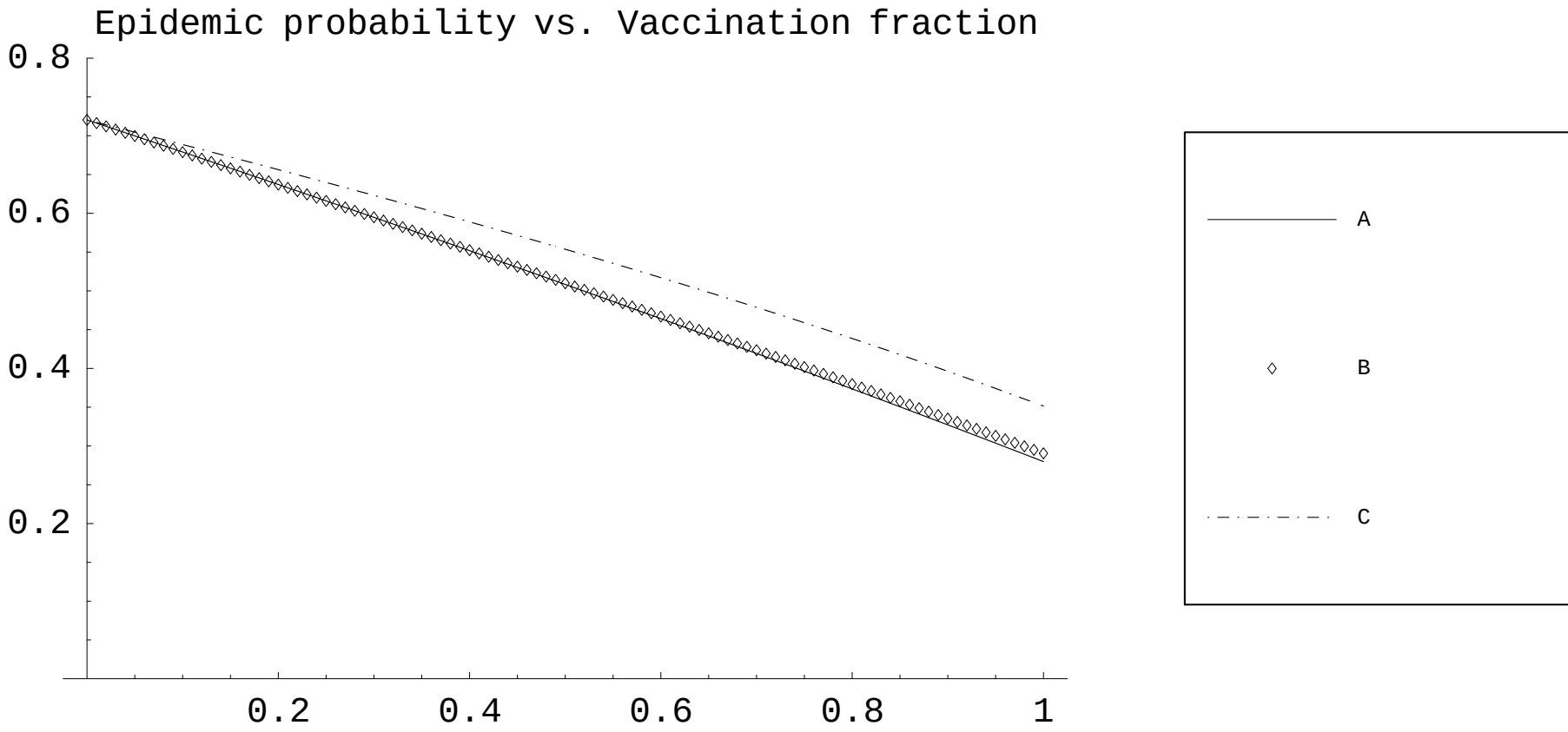
Subpopulations A, B, and C each constitute one-third of the overall population.

Subpopulation	**A**	**B**	**C**
Mean outdegree (infectiousness)	5	2.5	1.25
Mean indegree (susceptibility)	1.25	2.5	5
Pr(causes epidemic)	.951	.779	.430
Pr(infected in epidemic)	.430	.779	.951
Pr(in GSCC)	.409	.607	.409
Mean degree within GSCC	.835	.942	.835

[9] Calculations and graphs done in Mathematica 5.0.0.0 (©Wolfram Research, Inc) based on E. Kenah and J. M. Robins [*J Theor Biol* **249**, 706-722 (2007)].

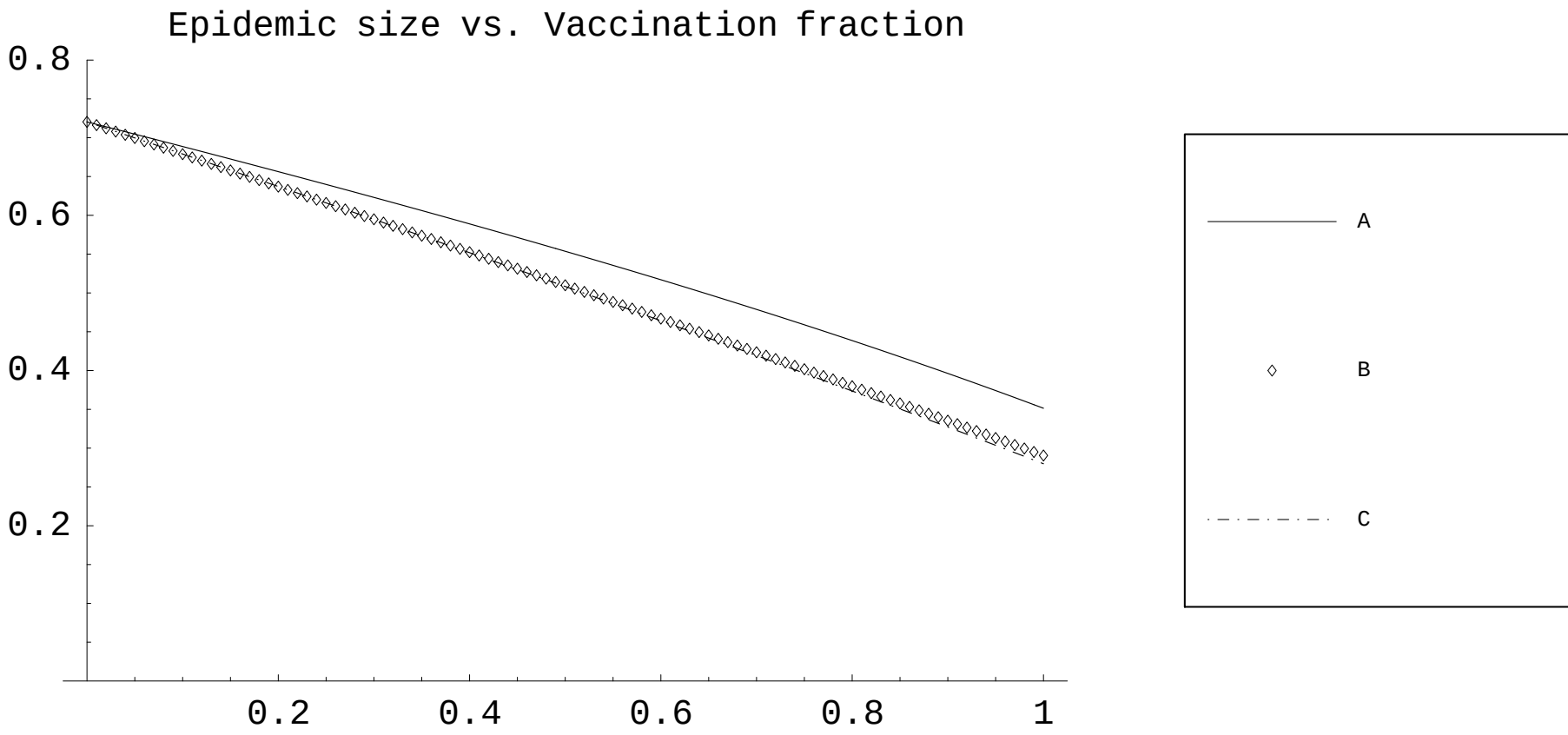
Analytical results

Effects of vaccination



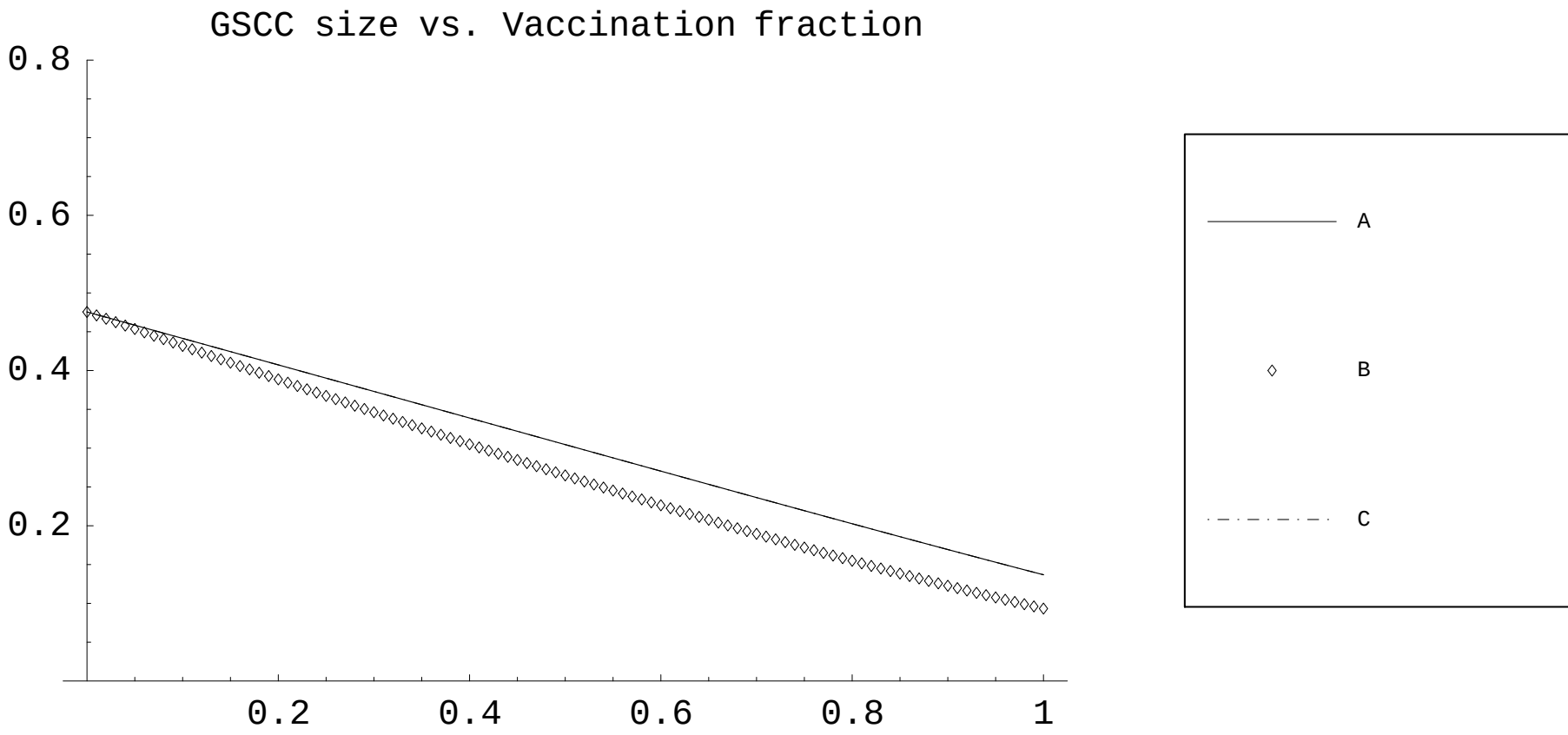
Analytical results

Effects of vaccination



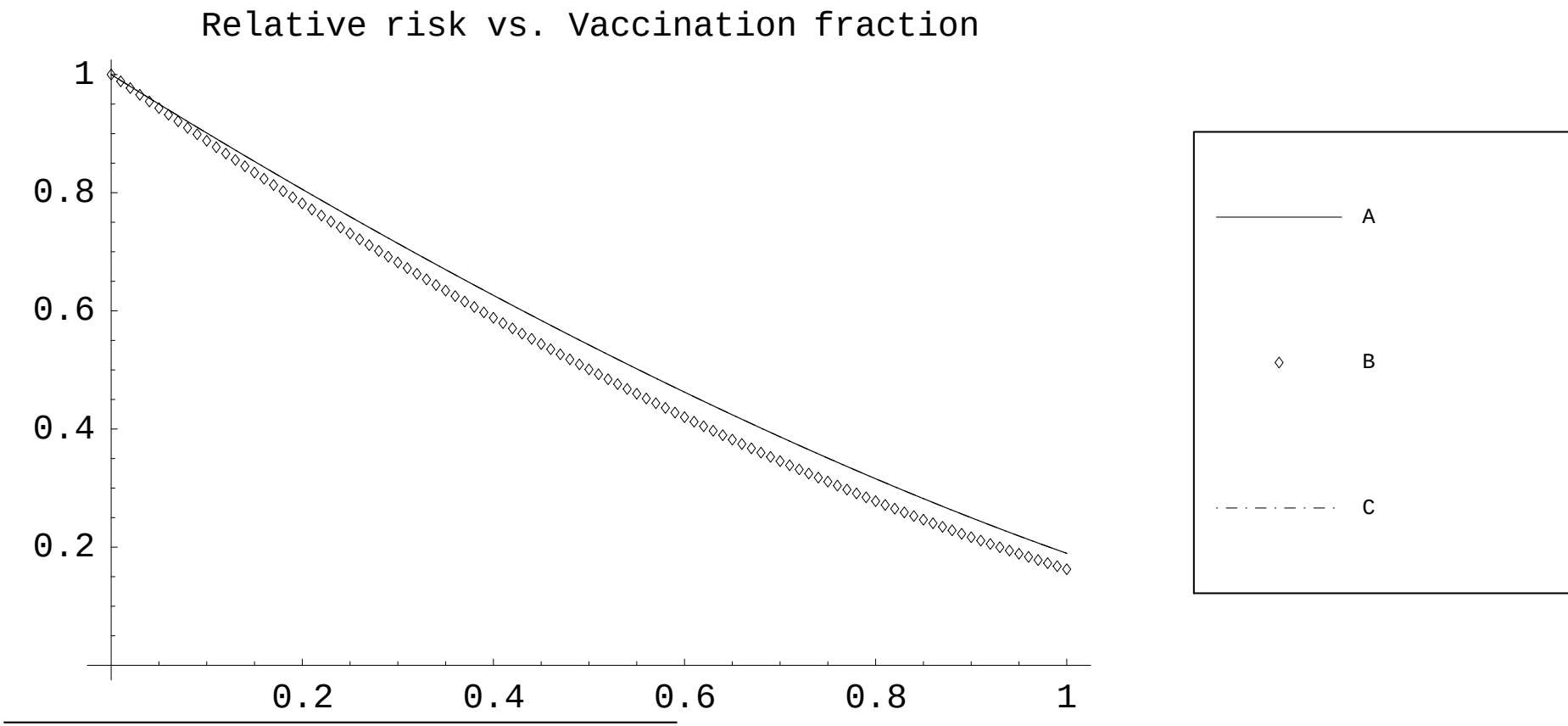
Analytical results

Effects of vaccination



Analytical results

Effects of vaccination[10]



[10] Relative risk of being infected eventually given a single randomly chosen initial infection in the population

Summary

EPNs provide a useful and intuitive framework for thinking about interventions in SIR epidemic models.

- Targeting the GSCC was an effective vaccination strategy in both network-based and fully-mixed epidemic models.
- In the network-based models, it was never inferior to the strategy of targeting highly-connected nodes in the contact network. In models with great heterogeneity of infectiousness and susceptibility, it was a superior strategy.
- In the fully-mixed model, the best vaccination strategies for reducing the probability and final size of an epidemic were different, but targeting the GSCC was very close to the most effective strategy for both.

Clarifications and extensions

- **The properties by which nodes should be targeted in the GSCC need to be better defined.**
  - Personally, I suspect the key lies in the stable distribution of a Markov process defined by transmission within the GSCC.
- The GSCC can be targeted by other types of interventions (building closure, vector control, etc.).
  - In models where each transmission is associated with a location or a vector breeding site, we could target locations or sites that account for the greatest number of edges within the GSCC.
- An understanding of the EPNs of complex epidemic models, such as EpiSimS,[11] would be extremely useful.
  - Violations of time-homogeneity may (or may not) have important consequences.

[11] S. Eubank *et al.* [*Nature* **429**, 181 (2004)]

Collaborators

I thank the following people for their comments, questions, encouragement, and support at various stages of this research:

- James M. Robins and Marc Lipsitch *(HSPH)*
- Joel C. Miller *(British Columbia Center for Disease Control)*
- Carl Bergstrom *(University of Washington)*
- Jacco Wallinga *(National Institute for Public Health and the Environment, the Netherlands)*
- Aric Hagberg *(Los Alamos National Laboratory)*
- Leon Arriola *(University of Wisconsin, Whitewater)*
- Eduardo López *(Oxford University)*
- My wife, Asma Aktar, and my son, Rafi

Financial support

This research received financial support from:

- National Institute of General Medical Sciences (NIH) grants:
  U01GM076497 "Models of Infectious Disease Agent Study"
  (PI: Marc Lipsitch, HSPH)
  F32GM085945 "Linking transmission models and data analysis in infectious disease epidemiology"
  (Host: Ira Longini, University of Washington)
- Los Alamos Mathematical Modeling and Analysis Summer Program *(T-7 Division and Center for Nonlinear Studies, Los Alamos National Laboratory)*
- Institute for Quantitative Social Science *(Harvard University)*