

# Epidemic thresholds for infections on networks

Laura Zager

EECS, MIT

25 September 2007



# Outline

Compartmental models and  $R_0$

The deterministic framework and threshold results

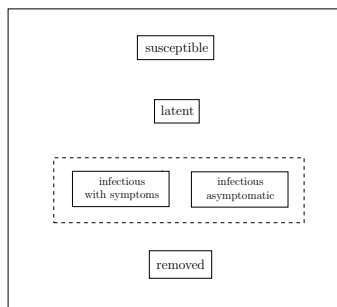
Threshold estimation with unknown network structure

Ongoing work



## Compartmental models

The idea: individuals transition between different disease compartments corresponding to their infectious state.

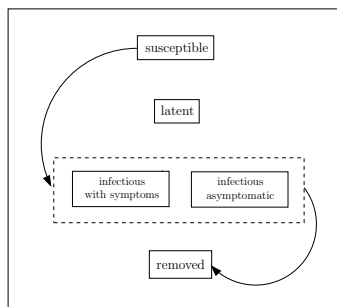


### Assumptions:

- ▶ all individuals in a compartment are identical
- ▶ compartments can represent heterogeneities in interaction pattern, disease susceptibility and transmissibility, immunization state, etc.

## Compartmental models

The idea: individuals transition between different disease compartments corresponding to their infectious state.



**SIR model** - appropriate for diseases which confer immunity or death, e.g. chicken pox

Assumptions:

- ▶ all individuals in a compartment are identical
- ▶ compartments can represent heterogeneities in interaction pattern, disease susceptibility and transmissibility, immunization state, etc.

## Threshold tests on $R_0$

- ▶ the basic reproductive ratio, defined as 'the expected number of secondary cases produced by a typical infected individual during its entire period of infectiousness in a completely susceptible population' [Diekmann, 1990]
- ▶ the largest eigenvalue of the *next-generation matrix*,  $K$ , where

$K_{ij}$  = the average number of direct infections of individuals of type  $j$  from an initial infective of type  $i$  (averaged over the entire population)

- ▶ represents the rate of growth of the epidemic in a population with infinitely replaceable susceptibles
- ▶ **common interpretation: an epidemic will occur if and only if  $R_0 > 1$**

AIDS	2 – 5
smallpox	3 – 5
measles	16 – 18
malaria	> 100

Estimated ranges of  $R_0$  for several diseases from [Keeling, 2001]

## Threshold tests on $R_0$

- ▶ the basic reproductive ratio, defined as 'the expected number of secondary cases produced by a typical infected individual during its entire period of infectiousness in a completely susceptible population' [Diekmann, 1990]
- ▶ the largest eigenvalue of the *next-generation matrix*,  $K$ , where

$K_{ij}$  = the average number of direct infections of individuals of type  $j$  from an initial infective of type  $i$  (averaged over the entire population)

- ▶ represents the rate of growth of the epidemic in a population with infinitely replaceable susceptibles
- ▶ **common interpretation: an epidemic will occur if and only if  $R_0 > 1$**

AIDS	2 – 5
smallpox	3 – 5
measles	16 – 18
malaria	> 100

Estimated ranges of  $R_0$  for several diseases from [Keeling, 2001]

$R_0$  heavily dependent upon population structure

## The deterministic framework

- ▶ formulated for continuous-time in [van den Driessche, 2002] and adapted here for discrete-time (rates are measured per time step).

### Model elements

- ▶  $x = (x_1, \dots, x_n)$  counts the individuals in each of  $n$  disease compartments, the first  $m$  of which are infected.
- ▶  $\mathcal{F}_i(x)$  is the rate of appearance of new infections in compartment  $i$ .
- ▶  $\mathcal{V}_i^+(x)$  is the rate of movement of individuals into compartment  $i$  by means other than infection.
- ▶  $\mathcal{V}_i^-(x)$  is the rate of removal of individuals from compartment  $i$  by any means.

$$x_i \leftarrow h_i(x) = x_i + \mathcal{F}_i(x) + \mathcal{V}_i^+(x) - \mathcal{V}_i^-(x) = x_i + \mathcal{F}_i(x) - \mathcal{V}_i(x)$$

# The deterministic framework

- ▶ formulated for continuous-time in [van den Driessche, 2002] and adapted here for discrete-time (rates are measured per time step).

## Model elements

- ▶  $x = (x_1, \dots, x_n)$  counts the individuals in each of  $n$  disease compartments, the first  $m$  of which are infected.
- ▶  $\mathcal{F}_i(x)$  is the rate of appearance of new infections in compartment  $i$ .
- ▶  $\mathcal{V}_i^+(x)$  is the rate of movement of individuals into compartment  $i$  by means other than infection.
- ▶  $\mathcal{V}_i^-(x)$  is the rate of removal of individuals from compartment  $i$  by any means.

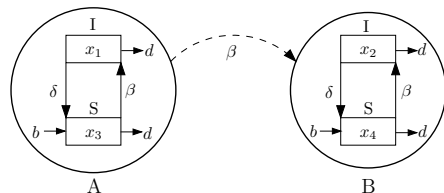
$$x_i \leftarrow h_i(x) = x_i + \mathcal{F}_i(x) + \mathcal{V}_i^+(x) - \mathcal{V}_i^-(x) = x_i + \mathcal{F}_i(x) - \mathcal{V}_i(x)$$

## Constraints

- ▶ for nonnegative  $x$ , the rate functions are nonnegative.
- ▶ no more individuals can leave a compartment than currently occupy it.
- ▶ no new infections can arise in non-infected compartments.
- ▶ if there are no infected individuals, then no new infections can arise.
- ▶ we'll call  $\bar{x}$  a *disease-free equilibrium* (DFE) if there are no infected individuals and the population dynamics at that equilibrium are stable.



## An example



- ▶ an SIS model
- ▶  $A$  and  $B$  are two different subpopulations
- ▶ the dashed arrow from  $A$  to  $B$  indicates disease transmission
- ▶ births are susceptible, and deaths remove individuals from both compartments

$$x_1 \leftarrow x_1 + \underbrace{\beta x_1 x_3}_{\mathcal{F}_1} - \underbrace{(\delta x_1 + dx_1)}_{v_1^-}$$

$$x_2 \leftarrow x_2 + \underbrace{\beta x_2 x_4 + \beta x_1 x_4}_{\mathcal{F}_2} - \underbrace{(\delta x_2 + dx_2)}_{v_2^-}$$

$$x_3 \leftarrow x_3 + \underbrace{b + \delta x_1}_{v_3^+} - \underbrace{(\beta x_1 x_3 + dx_3)}_{v_3^-}$$

$$x_4 \leftarrow x_4 + \underbrace{b + \delta x_2}_{v_4^+} - \underbrace{(\beta x_2 x_4 + \beta x_1 x_4 - dx_4)}_{v_4^-}$$

- ▶ a DFE is given by  $\bar{x} = (x_1, x_2, x_3, x_4) = (0, 0, b/d, b/d)$

## $R_0 < 1$ equivalent to local DFE stability

Let  $\bar{x}$  be a DFE and let  $F$  and  $V$  denote the Jacobian matrices of the rate functions  $\mathcal{F}$  and  $\mathcal{V}$  evaluated at  $\bar{x}$ .

1. The next-generation matrix  $K$  is given by  $K = FV^{-1}$ , so  $R_0 = \rho(FV^{-1})$ .
2. The DFE  $\bar{x}$  is locally asymptotically stable if and only if  $R_0$  is less than 1.

(DT version of CT result in [van den Driessche, 2002])

# Unknown network structure

## Some partial information scenarios

1. A subgraph is known.
  - ▶ e.g. in egocentric social network studies, only local information is gathered (i.e., the neighborhood of individual network nodes)
2. A generation/evolution mechanism for the network can be hypothesized.
  - ▶ e.g. Barabasi and Albert's preferential attachment model [Barabasi, 1999] and Chung's duplication model [Chung, 2006]
3. Ranges for certain network parameters can be estimated.
  - ▶ e.g. in a family tree, we can assume that no person has more than twenty children

# Unknown network structure

## Some partial information scenarios

1. A subgraph is known.
  - ▶ e.g. in egocentric social network studies, only local information is gathered (i.e., the neighborhood of individual network nodes)
2. A generation/evolution mechanism for the network can be hypothesized.
  - ▶ e.g. Barabasi and Albert's preferential attachment model [Barabasi, 1999] and Chung's duplication model [Chung, 2006]
3. Ranges for certain network parameters can be estimated.
  - ▶ e.g. in a family tree, we can assume that no person has more than twenty children

## The idea

### **Modeling the network as a probabilistic ensemble of all graphs that satisfy the known conditions**

- ▶ the approach taken in social network analysis, which uses field data to fit the coefficients of network models drawn from a family of distributions (most commonly the  $p^*$  or exponential random graphs)



## Bounding and approximating a random $R_0$

Let  $\mathbf{A}$  denote a random adjacency matrix, which describes the interaction patterns in the network. In general,  $\mathbf{K} = \mathbf{F}(\mathbf{A})\mathbf{V}(\mathbf{A})^{-1}$ .

However, if individuals only differ in their mixing patterns, then we can show that

$$\mathbf{R}_0 = \rho(\mathbf{F}\mathbf{V}^{-1}) = R_{0,h}\rho(\mathbf{A})$$

where  $R_{0,h}$  is the basic reproductive ratio for a *homogeneous* population.

- ▶ interpretation: recovery and demographic changes are not dependent on network location

## Some observations

$$\mathbf{R}_0 = R_{0,h\rho}(\mathbf{A})$$

## Some observations

$$\mathbf{R}_0 = R_{0,h\rho}(\mathbf{A})$$

- ▶ In most deterministic models, the random matrix  $\mathbf{A}$  is replaced with  $E[\mathbf{A}]$ .

## Some observations

$$\mathbf{R}_0 = R_{0,h}\rho(\mathbf{A})$$

- ▶ In most deterministic models, the random matrix  $\mathbf{A}$  is replaced with  $E[\mathbf{A}]$ .
- ▶ However, if connections are undirected, then

$$\rho(E[\mathbf{A}]) \leq E[\rho(\mathbf{A})].$$

leads us to underestimate the the mean of the distribution of  $R_0$

## Some observations

$$\mathbf{R}_0 = R_{0,h}\rho(\mathbf{A})$$

- ▶ In most deterministic models, the random matrix  $\mathbf{A}$  is replaced with  $E[\mathbf{A}]$ .
- ▶ However, if connections are undirected, then

$$\rho(E[\mathbf{A}]) \leq E[\rho(\mathbf{A})].$$

leads us to underestimate the the mean of the distribution of  $R_0$

- ▶ in some cases,  $\rho(E[\mathbf{A}])$  and  $E[\rho(\mathbf{A})]$  diverge from each other as the number of nodes increases, e.g. [Chung, 2003]
- ▶ the median, mode or upper bound on the support of the distribution of  $\mathbf{R}_0$  may be more epidemiologically relevant
- ▶ if the network structure is dynamic, our conclusions should hold for any time sample
- ▶ these estimates should come from structural characteristics of the network

## Some observations

$$\mathbf{R}_0 = R_{0,h}\rho(\mathbf{A})$$

- ▶ In most deterministic models, the random matrix  $\mathbf{A}$  is replaced with  $E[\mathbf{A}]$ .
- ▶ However, if connections are undirected, then

$$\rho(E[\mathbf{A}]) \leq E[\rho(\mathbf{A})].$$

leads us to underestimate the the mean of the distribution of  $R_0$

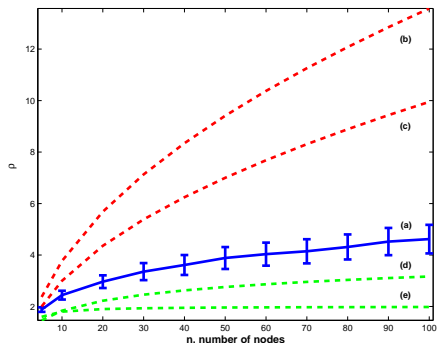
- ▶ in some cases,  $\rho(E[\mathbf{A}])$  and  $E[\rho(\mathbf{A})]$  diverge from each other as the number of nodes increases, e.g. [Chung, 2003]
- ▶ the median, mode or upper bound on the support of the distribution of  $\mathbf{R}_0$  may be more epidemiologically relevant
- ▶ if the network structure is dynamic, our conclusions should hold for any time sample
- ▶ these estimates should come from structural characteristics of the network

apply results from spectral graph theory



## An example

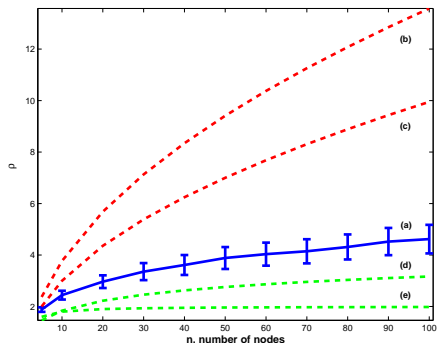
Goal: estimate the spectral radius of the adjacency matrix of a network generated by preferential attachment [Barabasi, 1999].



- (a) the mean ( $\pm$  std. dev.) of the spectral radius of the adjacency matrix of a simple preferential attachment model on  $n$  nodes ( $n - 1$  edges), taken over 100 trials

## An example

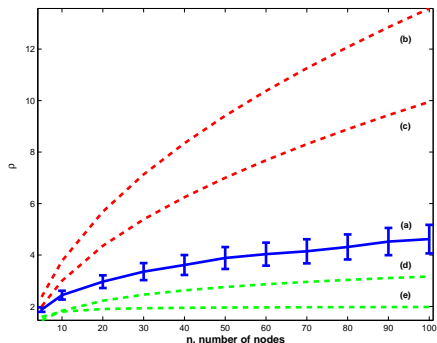
Goal: estimate the spectral radius of the adjacency matrix of a network generated by preferential attachment [Barabasi, 1999].



- (a) the mean ( $\pm$  std. dev.) of the spectral radius of the adjacency matrix of a simple preferential attachment model on  $n$  nodes ( $n - 1$  edges), taken over 100 trials
- (b) an upper bound on  $\rho(\mathbf{A})$  obtained using the number of edges via [Stanley, 1987]
- (c) an upper bound obtained using the number of nodes, edges, minimum degree and maximum degree via [Das, 2004]

## An example

Goal: estimate the spectral radius of the adjacency matrix of a network generated by preferential attachment [Barabasi, 1999].



- (a) the mean ( $\pm$  std. dev.) of the spectral radius of the adjacency matrix of a simple preferential attachment model on  $n$  nodes ( $n - 1$  edges), taken over 100 trials
- (b) an upper bound on  $\rho(\mathbf{A})$  obtained using the number of edges via [Stanley, 1987]
- (c) an upper bound obtained using the number of nodes, edges, minimum degree and maximum degree via [Das, 2004]

- (d) approximation assuming a degree distribution  $\sim k^{-3}$ , corresponding to preferential attachment, without degree correlations.
- (e) approximation assuming a homogeneous network on  $n$  nodes with  $n - 1$  edges distributed identically

## Ongoing work

- ▶ assembling results from spectral graph theory to aid in bounding  $R_0$
- ▶ analyzing the spectral properties of the family of exponential network models
- ▶ constructing an analogous stochastic framework and threshold
- ▶ exploring spatial patterning in outbreaks (e.g. correlation and clustering)
- ▶ validating these approaches on test data

# Acknowledgments



- ▶ George Verghese
- ▶ Martin Family Fellowship for Environmental Sustainability
- ▶ Women's Technology Program at MIT
- ▶ National Science Foundation
- ▶ DoD AFOSR URI for 'Architectures for Secure and Robust Distributed Infrastructures'