How to vaccinate nodes of a network?

Winfried Just   Department of Mathematics, Ohio University

The goal of this poster

This poster is all about **formulating mathematically precise questions.** The answers, if such can be found, would then translate into predictions about **optimal vaccination strategies** in network-based models. These questions take into account aspects of disease modeling that are often ignored (see third panel).

- There is already a literature at least on some versions of the first of these questions.
- In other cases, the questions may be **combinatorial optimization problems** that have not previously been studied.
- At least some of the latter are **NP-hard.**
- The author conjectures that when there is some randomness in the set of nodes targeted by vaccination or uncertainty about the network, these *a priori* NP-hard optimization problems become computationally tractable. **This may be an instance of a more general known phenomenon in computational complexity theory.** The author hopes for some pointers to the literature.

Vaccination strategies

A major objective of mathematical epidemiology is to compare expected effectiveness of various feasible control strategies and derive policy recommendations for their implementation. For vaccinations, such recommendations can be phrased as answers to the following questions:

1. What proportion of hosts need to be immunized to prevent major outbreaks?
2. Which subpopulations of hosts should be vaccinated if vaccine is scarce?

Compartment-level models that are based on the assumptions of uniform mixing and homogeneity of hosts give straightforward answers to the first question by predicting that when vaccination coverage exceeds the **herd immunity threshold** $1 - \frac{1}{R_0}$, no major outbreaks will occur. However, such models cannot address the second question.

Network-based models assume that transmission can occur only along the edges of a given contact network. These models are in principle capable to provide more nuanced predictions and can address the second question.

Things aren't that simple

For example, when the contact network has an approximately scale-free degree distribution, more protection for the entire population can be achieved by preferentially vaccinating **K-hubs,** that is nodes with degrees $\geq K$ in the contact network. That much is intuitively clear and entirely unsurprising. However, for real contact networks things aren't quite as simple:

- The cost of achieving $x\%$ vaccination coverage for a given group of hosts will not be linear in x. For a given limited budget, what is the optimal tradeoff between near perfect vaccination coverage for the highest priority group and a more moderate level of coverage for a larger group?
- Apart from degree distribution, there may be other structural properties of the network that may influence the choice of the optimal target group for vaccinations.
- For real populations we don't have perfect knowledge of the actual contact network. What strategy is optimal when we have only partial knowledge of the network?

Some terminology

For simplicity, let us consider disease transmission models of type SIR for a fixed population of N **hosts** without demographics. Hosts who are in states S (susceptible), I (infectious), or R (removed) are said to reside in the respective **compartments. Compartment-level models** study how the proportions of hosts in these compartments change over time t and do not distinguish between hosts. In **network-based models** each host is represented by a node of a graph that represents the **contact network**, and it is assumed that transmission can occur only by direct contact between two hosts that are connected by an edge.

We focus on initial states with one infectious host, called the **index case,** in an otherwise susceptible population. The resulting dynamics represents an **outbreak,** which in an SIR model always terminates in a state with no infectious hosts. All hosts that were in state I at some time will have **experienced infection.** Their proportion is called the **final size** F of the outbreak.

Control measures aim at reducing the expected final size as much as possible. **Vaccination** can be conceptualized as changing the state of some hosts to R prior to any secondary infections.

Two general questions

Suppose we are given a (stochastic) model with distinguishable disjoint subpopulations $P_1, \dots, P_k$.

Question 1: What levels of vaccination coverage $x_1, \dots, x_k$ of these subpopulations will guarantee that the probability of outbreaks with final size $F > \varepsilon$ is at most δ?

In a network-based model we may be able to choose the subpopulations $P_1, \dots, P_k$ themselves based on optimality. For a given nonlinear cost function of vaccination coverage we may then ask:

Question 2: Given tolerance levels ε and δ, what choice of k, subpopulations P_i, and levels of vaccination coverage x_i will guarantee, at the lowest possible cost, that the probability of outbreaks with final size $F > \varepsilon$ is at most δ?

Two specific questions

Suppose we are given a model with a scale-free contact network and we decide to vaccinate a proportion x of randomly chosen K-hubs. Then we can ask the following version of Question 2:

Question 3: Given tolerance levels ε and δ, what choice of K and x will guarantee, at the lowest possible cost, that the probability of outbreaks with final size $F > \varepsilon$ is at most δ?

Now suppose we are given a model with an arbitrary contact network and the cost function is linear so that we may aim at 100% vaccination coverage for any chosen set of nodes. Then Question 2 becomes:

Question 4: What is the set of nodes of the smallest size so that vaccination of these nodes will guarantee that the probability of outbreaks with final size $F > \varepsilon$ is at most δ?

This may be intractable!

If we want to prevent with certainty any secondary infections whatsoever, then ε is the reciprocal of the network size and $\delta = 0$. In this case, Question 4 is the **vertex cover problem,** a known NP-hard combinatorial optimization problem.

Question 5: What kind of combinatorial optimization problems do we obtain from Question 4 for other choices of ε and δ? Is there a translation into previously studied problems?

Question 6: For what choices of ε and δ does Question 4 remain NP-hard? Does the problem become computationally tractable if we allow some randomness in the sense that $\delta > 0$?

Question 7: Can analogues of Question 4 be NP-hard when we ask for a specified level x of less than perfect coverage of a subset of nodes?

Imperfect knowledge

So far, we have only considered the scenario when perfect knowledge of the network is assumed. This is not realistic in practice.

One can model the inherent uncertainty about the contact network by treating it as a distribution of random graphs, where each given potential edge is either included or excluded with probability $p \approx 1$. We can then ask analogues of the previous questions for ensembles of such models.

Question 8: Does any of these questions remain NP-hard for a fixed $p < 1$?

Contact Information:

Web http://www.ohio.edu/people/just/
Email mathjust@gmail.com
Phone +1 (740) 593 1260