Epidemics 16 (2016) 56–62

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

A fast algorithm for calculating an expected outbreak size on dynamic contagion networks

Jessica Enright[a,*], Rowland R. Kao[b]

[a] Computing Science and Mathematics, University of Stirling, Stirling FK9 4LA, United Kingdom
[b] Boyd Orr Centre, Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow G61 1QH, United Kingdom

ARTICLE INFO

Article history:
Received 16 October 2015
Received in revised form 12 May 2016
Accepted 18 May 2016
Available online 24 May 2016

Keywords:
Network modelling
Contagion on networks

ABSTRACT

Calculation of expected outbreak size of a simple contagion on a known contact network is a common and important epidemiological task, and is typically carried out by computationally intensive simulation. We describe an efficient exact method to calculate the expected outbreak size of a contagion on an outbreak-invariant network that is a directed and acyclic, allowing us to model all dynamically changing networks when contagion can only travel forward in time. We describe our algorithm and its use in pseudocode, as well as showing examples of its use on disease relevant, data-derived networks.

1. Introduction

Increasingly, models of contagion spread on highly structured populations are being used to inform disease control (Green et al., 2008; James et al., 2007; Eames et al., 2015; Danon et al., 2011). As these models become increasingly complex, simple and robust approaches to calculate the potential outbreak size become increasingly important. Here, we provide an algorithm which allows this to be done faster than current methods on a variety of real-world networks.

Previous research (Eames et al., 2015; Danon et al., 2011) uses simulation to estimate expected outbreak size on various different types of network. This is computationally intensive, and provides only stochastically derived estimates for the outbreak size. We present an exact method for calculating expected outbreak size. While our method does not apply to arbitrary contagion spread on general networks, which is a known NP-hard problem (Shapiro and Delgado-Eckert, 2012), it is relevant to the wide range of contagion examples that can be expressed on the well-studied class of directed acyclic graphs (DAGs): networks in which there are no directed cycles. This class is particularly useful for modelling temporally changing contact networks, and the notion that time (and therefore infection) only flows in one direction is central to our approach.

Our method has two advantages over simulation: it is computationally much faster, and it gives an exact answer rather than a statistical estimate. These two advantages are of particular importance in applications where a rapid estimate is important, without the requirement for a detailed behavioural or disease model, as in an outbreak situation with stringent externally-imposed timelines, or as an internal component in a larger software package that must complete a very large number of outbreak size calculations over a large number of different networks.

The method we describe here has much in common with several previously described methods: the novelty is largely in our algorithmic treatment and its use on a particular multi-layer directed acyclic graph (a structure also used in Kim and Anderson, 2012; Valdano et al., 2015) in order to incorporate a temporally changing network. We wish to highlight the relatedness of our approach to the methods of Rogers(Rogers, 2015), and Ludwig's method (Ludwig, 1975) as applied to a random network by House et al. (2012).

Rogers (2015) and Karrer and Newman (2010) describe the use of a cavity method on a network to calculate node risk and travel the development of an outbreak, as well as its final size. Rogers (2015) uses a tree approximation of a static network in its calculations of probability of given node's involvement in an outbreak; we apply a similar calculation to our directed acyclic graph.

Ludwig's method works on a system of pre-generated ranks in which nodes are assigned an order, and considered for infection in that order, and when applied to a network, requires the network be unchanged by an outbreak (Ludwig, 1975; House et al., 2012; Pellis et al., 2008). Given a starting node for the outbreak, nodes are

* Corresponding author.
E-mail address: jenright@gmail.com (J. Enright).

http://dx.doi.org/10.1016/j.epidem.2016.05.002

sorted by the length of their shortest paths to the starting node, with these shortest path lengths used as each node's "rank". Nodes are considered for infection by order of their ranks, with nodes closer to the starting node considered earlier. As described in House et al. (2012), an implementation of Sellke's construction (Sellke, 2012) on a network bears a close resemblance to Ludwig's method on a network.

As with Ludwig's method, we will consider nodes in a rough order of distance from an outbreak seeding set of nodes, though for our approach any topological ordering would suffice, and (again like in Ludwig's method) we will require our network to be invariant with respect to the outbreak.

We direct the reader to House et al. (2012) for a review of a wide variety of methods in use for calculating the probability mass function of a final outbreak size.

We present the method on a DAG derived from an infection network in Section 3. In Section 4 we show the method on several example networks, including two derived from real-world data. Section 5 compares our method to a series of simulations, demonstrating the advantages in speed and accuracy. We conclude with some adaptations which can be made to run the algorithm on more complex contagion networks, and some suggestions for further research. We provide open-source Python code which we hope will be of use in the future to other researchers.[1]

2. Overview of algorithm on a directed acyclic graph

We describe our approach in several steps: first, following Kim and Anderson (2012) we describe the production of a directed acyclic graph to describe a dynamically changing network. While we will focus on calculating on a dynamic network, it is possible to produce the directed acyclic graph required from a static network simply by repeating static contacts over many time steps.

Using this DAG as input, we then describe an efficient algorithm to calculate the expectation that any given node will be infected at a given time in an epidemic where individuals become immediately infectious and remain infections indefinitely (an SI model), or can recover and become immediately susceptible again (SIS). We allow an arbitrary choice, or distribution of choices, of starting nodes and times for the epidemic. Because expectations can be combined linearly (Hamming, 1991), this node-by-node expectation calculation enables us to calculate the expected size of an overall outbreak exactly at any fixed timepoint, again, either with a set starting node and time, or over a specified distribution of starting points.

2.1. Producing a directed acyclic graph from a dynamic network

In our preferred method for producing a DAG from a dynamic network, we essentially identify each agent at each time step with a node in the DAG, with an edge from one node (u, t) in the DAG to another $(v, t+1)$ if the state of the vertex u at time t can affect the state of the vertex v at time $t+1$. As in Kim and Anderson (2012) and Valdano et al. (2015) we use a multi-layered directed acyclic graph in which each layer is a time slice to encode a dynamically changing network of impulse edges. We assume throughout that disease cannot spread instantaneously, that is, an agent infected at t cannot infect another instantaneously, but is only able to infect others at $t+\epsilon$, where ϵ is an appropriately and arbitrarily small number. We also assume that the set of contacts that make up the network are known before the beginning of our calculation.

Let $G = (V, \vec{E})$ be a graph (or network) with vertices V and time-impulse directed edges $\vec{E}$. Let $\mathcal{T}$ be the relation between impulses and the times at which they occur. We assume that the range of $\mathcal{T}$ is a subset of the integers. Let $\mathcal{E}$ be the set of edges expressed as triples: (u, v, t) indicating an edge from u to v at time t, and let $Q : \mathcal{E} \to [0\ldots1]$ be the probability that, if the source of each impulse contact is infected, it will infect the destination of the edges.

Let V_T be the set: $\{(v, t)$ where $v \in V$, and $t \in [\min(range(\mathcal{T})) - 1 \ldots \max(range(\mathcal{T}))]\}$ Let $\vec{E_T}$ be the set:

- $\{((v, t) \to (u, s))$ where $u = v$ and $s = t+1\} \cup$
- $\{((v, t) \to (u, s))$ where $t = \mathcal{T}(u, v)$ and $s = t+1\}$

Let $\mathcal{P} : \vec{E_T} \to [0\ldots1]$ be a function from $\vec{E_T}$ to real-numbered probabilities between 0 and 1 such that:

- for edge $((u, t) \to (u, t+1))$, we set $\mathcal{P}(((v, t) \to (u, t+1)))$ to the probability that the disease persists at v from time t to $t+1$ and
- for edge $((v, t) \to (u, t+1))$ where $(v, u, t) \in \mathcal{T}$, we set $\mathcal{P}(((v, t) \to (u, s))) = Q(v, u, t)$

We have the building blocks of our directed acyclic graph in the form of a node set, an edge set, and probabilistic weights for the edges. Let graph $G_T = (V_T, \vec{E_T})$ be a directed graph: we know that G_T is acyclic because for every edge $((v, t) \to (u, s)) \in E_T$ we know that $s > t$; intuitively, the edges only go forward in time.

With the directed acyclic graph $G_T = (V_T, \vec{E_T})$ and the probability weighting function $\mathcal{P}$ we have the required input for our algorithm. Therefore, given a set of integer-time impulse contacts with probabilities of disease transmission associated with each contact, we can produce the graph we need, and use our algorithm to calculate expected outbreak size.

3. Expected outbreak size algorithm

While our algorithm below will work on any directed acyclic graph, we describe it in the context of a time-expanded graph as above, as this is the most relevant to our examples.

Let $G = (V, \vec{E})$ be a directed, acyclic graph as described above and $\mathcal{P} : \vec{E} \to [0\ldots1]$ be a function from the edges of G to probabilities such that $\mathcal{P}((u \to v))$ is the probability that u will infect v if it is, itself, infected. Note that, as described, there may be edges $((u, t) \to (v, t+1))$ where $u \neq v$ between different agents at successive times, as well as edges $((u, t) \to (u, t+1))$ between the same agent at successive times. The probability that an edge of the type $((u, t) \to (u, t+1))$ transmits is the probability that an infection of agent u at time t persists to time $t+1$. In general, the probabilities that edges transmit infection may differ: this is no impediment, so long as it is recorded in $\mathcal{P}$.

We start with a question: what is the expectation that (v, t) is infected in an epidemic with a known starting point (u, t_0)? If we consider all nodes at all times that could be infected in an epidemic stating at (u, t_0), we can identify the set of nodes that could directly infect (v, t): those that are the source of an edge leading into (v, t) that could, themselves, potentially be infected by an epidemic starting at (u, t_0). We call these the *parents* of (v, t), and due to the construction of the DAG we have used, we know that they are in time slice $t-1$. Let $A = \{(p_0, t-1), (p_1, t-1) \ldots (p_m, t-1)\}$ be the set of parents of (v, t) in a traversal of G from (u, t_0). Note that, if we are using a time-expanded graph as defined above, then exactly one p_i will be equal to u: exactly one parent of an agent at a time is that agent at the previous time. Then the probability that (v, t) is infected in an outbreak is the probability that at least one parent is infected and infects (v, t). Recall that $\mathcal{P}(((p_i, t-1) \to (v, t)))$

[1] https://github.com/magicicada/expected-outbreak-size.