Journal of Theoretical Biology 407 (2016) 387–400

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Compact pairwise models for epidemics with multiple infectious stages on degree heterogeneous and clustered networks

N. Sherborne, K.B. Blyuss *, I.Z. Kiss

Department of Mathematics, University of Sussex, Brighton BN1 9QH, UK

HIGHLIGHTS

- A compact pairwise model for epidemics with a non-Markovian infectious period is constructed.
- The size of the model does not depend on the degree distribution.
- The model is extended to analyse the impact of clustering in the contact network.

ARTICLE INFO

Article history:
Received 10 February 2016
Received in revised form
28 May 2016
Accepted 10 July 2016
Available online 14 July 2016

Keywords:
Pairwise approximation
Moment closures
Clustering
Non-Markovian recovery

ABSTRACT

This paper presents a compact pairwise model describing the spread of multi-stage epidemics on networks. The multi-stage model corresponds to a gamma-distributed infectious period which interpolates between the classical Markovian models with exponentially distributed infectious period and epidemics with a constant infectious period. We show how the compact approach leads to a system of equations whose size is independent of the range of node degrees, thus significantly reducing the complexity of the model. Network clustering is incorporated into the model to provide a more accurate representation of realistic contact networks, and the accuracy of proposed closures is analysed for different levels of clustering and number of infection stages. Our results support recent findings that standard closure techniques are likely to perform better when the infectious period is constant.

1. Introduction

Mathematical models of infectious diseases have proven to be an invaluable tool in understanding how diseases invade and spread within a population, and how best to control them (Anderson and May, 1991; Diekmann et al., 2012; Pastor-Satorras et al., 2015). Given a good understanding of the biology of the disease and of the behaviour and interaction of hosts, it is possible to develop accurate models with good predictive power, which provide the means to develop, test and deploy control measures to mitigate the negative impacts of infectious diseases, a good example being influenza (Ferguson et al., 2006). However, as has been highlighted by the recent Ebola outbreak in West Africa (Chowell and Nishiura, 2014), models can be very situation-specific and can become highly sophisticated or complex depending on intricacies of the structure of the population and the characteristics of the disease.

In the last few decades the use of networks to describe interactions between individuals has been an important step change in modelling and studying disease transmission (Keeling, 1999; Danon et al., 2011; Keeling and Eames, 2005; Pastor-Satorras et al., 2015). There is now overwhelming empirical evidence that in many practical instances individuals interact in a structured and selective way, e.g. in the case of sexually transmitted diseases (Liljeros et al., 2001). Thus, the well-mixed assumption of early compartmental models (Kermack and McKendrick, 1927) has to be relaxed or models need to be refined by including multiple classes and mixing between classes. However, in some cases a network representation could be more realistic than a description based on compartmental models. Conventionally, nodes in network-based models represent individuals, and the edges describe connections between people who have sufficient contact to be able to transmit the disease (Keeling and Eames, 2005; Danon et al., 2011; Pastor-Satorras et al., 2015). This study focuses on static undirected networks, in which the edges of the network do not change over time, and all connections are sufficient to transmit the disease in either direction. The total number of edges a node has is known as its degree, and the frequency of nodes with different degrees is determined by a specific *degree distribution* $P(k)$ which can either be empirically measured or given theoretically. In either case $P(k)$ is the probability of a randomly chosen node having degree k. Early

* Corresponding author.
E-mail address: k.blyuss@sussex.ac.uk (K.B. Blyuss).

http://dx.doi.org/10.1016/j.jtbi.2016.07.015
0022-5193/

network models often assumed regular networks where all nodes have the same degree, or well-studied networks from graph theory, such as the Erdős–Rényi random graphs (Erdős and Rényi, 1959). However, empirical research showed that real biological, social or technological networks do not conform to such idealised models. In fact, many studies on human interactions ranging from sexual contact networks (Liljeros et al., 2001) to using the travel of banknotes as an indicator of human activity (Brockmann et al., 2006), or even internet connectivity (Caldarelli et al., 2000) have observed *wide-tail distributions*, with the majority of nodes having a low number of contacts, and a few nodes in the network having a much higher degree. This structure is most closely approximated by scale-free networks described by a power-law degree distribution $P(k) \sim k^{-\alpha}$ with some positive exponent α, which for most accurately described human contact patterns lies in the range $\alpha \in [2, 3]$ (see, for example, Pastor-Satorras and Vespignani, 2001). The impact of contact heterogeneity on the spread of epidemics is significant, and studies have highlighted the disproportionate role which may be played by a few highly connected nodes (James et al., 2007).

Another striking feature of real social contact patterns is the presence of small and highly interconnected groups which occur much more frequently than if edges were to be distributed at random. This is known as *clustering*, and its presence in empirical data (Newman et al., 2001; Foster et al., 2011) has driven the need to consider network models that include this feature. Perhaps, one of the most well-known and parsimonious theoretical models with tuneable clustering is the small-world network (Watts and Strogatz, 1998), where nodes are placed on a ring, and the network is dominated by local links to nearest neighbours with a few links rewired at random, which means that the average path length is not too large and comparable to that found in equivalent random networks. For a summary of numerous alternative algorithms that can be used to generate clustered networks see, for example, Green and Kiss (2010) or Ritchie et al. (2014). It is well known that modelling epidemic spread on such networks is more challenging, although some models have successfully incorporated clusterings (Miller, 2009; Karrer and Newman, 2010b; Volz et al., 2011; Ritchie et al., 2015, and references therein). However, it is often the case that such models only work for networks where clustering is introduced in a very specific way, e.g. by considering non-overlapping triangles or other subgraphs of more than three nodes.

Besides the details of the network structure, another major assumption that significantly reduces the mathematical complexity of models and makes them amenable to analysis with mean-field models of ordinary differential equations and tools from Markov chain theory is the assumption that the spreading/transmission of infection and recovery processes are Markovian. However, it has long been recognised that this is often not the case, and, for example, the infectious periods are typically far from exponential, and, perhaps, are better described by a normal-like or peaked distribution (Gough, 1977; Lloyd, 2001; Wearing et al., 2005). Modelling non-Markovian processes can be challenging and often leads to delay differential or integro-differential equations that are much more difficult to analyse. Recently, Kiss et al. (2015) have put forward a generalisation of a pairwise model for Markovian transmission with a constant infectious period for a susceptible–infected–recovered (SIR) dynamics, with a further recent extension by the same authors to an arbitrary distribution of the recovery time (Röst et al., 2016). The first generalisation resulted in a model given by a system of delay differential equations with discrete and distributed delays which makes it possible to gain insight into how the non-Markovian nature of the recovery process affects the epidemic threshold and the final epidemic size. Other important recent research in this direction includes the message passing formalism (Karrer and Newman, 2010a; Wilkinson and Sharkey, 2014) and an approach based on renewal theory (Cator et al., 2013).

In light of the importance of the above-mentioned network properties (i.e. degree heterogeneity and clustering) and the non-Markovian nature of the spreading and/or recovery processes, in this paper we generalise our recent research on a multi-stage *SIR* epidemics (Sherborne et al., 2015) and focus on modelling a Markovian spreading process with gamma-distributed infectious period on networks that account for heterogeneous degree distribution and clustering. This is achieved within the framework of pairwise models (Keeling, 1999), and we show that the additional model complexity induced by degree heterogeneity and non-Markovian recovery can be effectively controlled via a reduction procedure proposed by Simon and Kiss (2015). This allows one to derive an approximate deterministic model that helps numerically determine the time evolution of the epidemic and the final epidemic size. Moreover, the model allows us to gain insights into the interactions of the three main model ingredients, namely, degree heterogeneity, clustering and non-exponential recovery and the agreement between the model and the stochastic network simulation. The paper is organised as follows. In the next section we derive a compact pairwise model for unclustered networks whose size is independent of the range of degrees and derive and discuss some analytical results for this model. All results are validated by comparing the numerical solution of the pairwise model to results from direct stochastic network simulation. In Section 3, we investigate the case when the same epidemic unfolds on clustered networks. The corresponding pairwise model is derived, and we discuss the extra complexities necessary to more accurately approximate the spread of the disease. More importantly, we investigate how clustering and the non-Markovian recovery affect the agreement between the pairwise model and simulations. Finally, in Section 4 we conclude with a discussion of our results and future work.

2. Disease dynamics in the absence of clustering

As a first step in the analysis of the spread of epidemics on unclustered networks, we introduce the necessary concepts from multi-stage infections and pairwise models (Sherborne et al., 2015). In the SI^KR model, once a susceptible individual S becomes infected, they progress through K equally infectious stages denoted as $I^{(i)}$, $1 \le i \le K$. The transition rates between successive stages are given by $K\gamma$. Thus, in simulation the times spent in each of the K stages are independent exponentially distributed random numbers. The total time of infection is, therefore, the sum of K exponential distributions, which is a gamma distribution with the mean time of γ^{-1} (Durrett, 2010). In order to describe the dynamics of an epidemic we consider the state of the nodes in the network and the edges connecting them. Since a susceptible individual can only become infected upon a transmission across an $S - I^{(i)}$ link we need to consider the expected number of edges connecting susceptible and infected individuals in any stage i from 1 to K at time t over the whole network, to be denoted as $[SI^{(i)}](t)$. Here we have taken $[SI^{(i)}]$ independently of the degrees of the nodes in state S and $I^{(i)}$, i.e. $[SI^{(i)}] = \sum_{a,b} [S_a I_b^{(i)}]$ where a and b denote the degrees in the range between the minimum and maximum degrees in the network, denoted as k_{min} and k_{max}, respectively. This definition applies to all pairs, i.e. $[AB]$ stands for the population level count of all $A - B$ edges taken across all possible connections between nodes of different degrees;

$$[AB] = \sum_{a,b} [A_a B_b], \quad \text{and} \quad A, B \in \{S, I^{(1)}, I^{(2)}, \ldots, I^{(K)}, R\} := \mathbb{S}.$$

Here and henceforth $\mathbb{S}$ will denote the set of all possible states for a node. The expected number of $S - S$ edges depends on the

Please cite this article as: Sherborne, N., et al., Compact pairwise models for epidemics with multiple infectious stages on degree heterogeneous and clustered networks. J. Theor. Biol. (2016), http://dx.doi.org/10.1016/j.jtbi.2016.07.015