

BAYESIAN INFERENCE OF EPIDEMIOLOGICAL CHARACTERISTICS IN A PARTIALLY OBSERVED EPIDEMIC*

J. Ray[†] and Y. M. Marzouk
Sandia National Laboratories,
Livermore, CA 94550-0969

ABSTRACT

The spread of a disease in a population depends on its inherent transmission rate (a function of the pathogen) and the social mixing between the members of the population, commonly expressed in terms of a social network. Often the social network is unrecorded, especially for historical epidemics, and is generally imputed for analytical purposes. In this paper, we pose and solve a Bayesian inverse problem that infers both the transmission rate and the social network, using data from a 1967 smallpox epidemic in Abakaliki, Nigeria. The chain of transmission is identified as links on the inferred social network with a high probability of transmission.

INTRODUCTION

Communicable diseases, especially emerging ones, pose a particular threat to military populations. Their repeated deployments expose them to exotic pathogens while stressful, crowded and sometime communal living conditions (e.g. on ships) can amplify the spread of diseases. The spread rate of a disease depends on the intrinsic transmissibility of the pathogen and the social/mixing patterns of the host population. A military society with its well known structure and efficient medical surveillance can provide a fertile ground for advanced, data-driven epidemiological modeling, which can help uncover mechanisms of disease spread. In this paper, we show how the two components of disease spread may be inferred from observations of an ensuing epidemic. It also allows the automatic inference of the chain of transmission, which to date are uncovered via contact tracing. The inference technique is Bayesian and estimates the inferred epidemiological variables as probability density functions, thus quantifying the uncertainty in the inference. This allows the method to be used with incomplete observations e.g. during the early days of an epidemic, allowing inference of its characteristics with little data and allowing progressive improvement as more data becomes available.

THE MODEL

The model presented here is adapted from the approach developed in [1]. It adds a structured propulsion, differential transmission rates and uses a SEIR epidemic model instead of SIR (as in [1]). We consider a structured population consisting of a small number of groups. Individuals mix strongly within a group,

*Presented at the Chem-Bio Defense Physical Science and Technology Conference, November 2007, New Orleans, LA. SAND 2008-5049C.

[†]Corresponding author; Email: jairay@somnet.sandia.gov

and weakly across groups. The social network within a group is modeled using a binomial graph with link probability p_{in} ; social connections across groups are also modeled as a binomial graph, but with link probability p_{cross} , $p_{cross}/p_{in} = \rho \leq 1$. Disease transmission along a social link is modeled as a Poisson process, with rates β_{in} and β_{cross} inside and across groups, $\beta_{in} = (1 + r)\beta_{cross}$, $r \geq 0$.

Let the size of the population be N , of whom M show symptoms on dates S_i , $S_i \in \mathbf{S}$, $i = 1 \dots M$ during an epidemic. Let their infection and recovery dates be I_i , $I_i \in \mathbf{I}$ and R_i , $R_i \in \mathbf{R}$ respectively. Let \mathcal{G} be a social network (an undirected graph) connecting the N individuals, and let it be a collection of binomial graphs, one for each group and one spanning all groups. Let \mathcal{P} , a directed graph, denote an infection pathway, connecting the source of an infection to the targets. The incubation and recovery periods ($\mathbf{S} - \mathbf{I}$ and $\mathbf{R} - \mathbf{S}$ respectively) of a disease are modeled as Γ -distributed random variables, with means and standard deviations calculated from historical outbreaks. Assuming that the dates of exhibition of symptoms \mathbf{S} are known (i.e. they are observed), one can adapt [1] to derive the probability $\pi(\mathbf{S}|\mathbf{I}, \mathbf{R}, r, \beta_{cross}, p_{in}, \rho, \mathcal{G}, \mathcal{P})$ of observing \mathbf{S} , the data, given all other epidemiological ($\mathbf{I}, \mathbf{R}, r, \beta_{cross}$) and social ($p_{in}, \rho, \mathcal{G}, \mathcal{P}$) parameters. Using Bayes' theorem, this can be inverted for the conditional probability $\pi(\mathbf{I}, \mathbf{R}, r, \beta_{cross}, p_{in}, \rho, \mathcal{G}, \mathcal{P}|\mathbf{S})$. The spread rates β_{in} , β_{cross} and the infection pathway \mathcal{P} , form the principal quantities of interest and can be obtained by sampling from the conditional probability. Uniform distributions are used as priors for $\mathbf{I}, \mathbf{R}, r, \beta_{cross}, p_{in}$ and ρ . We use a single-component random walk Metropolis-Hastings algorithm [2] to sample for all variables except \mathcal{G} and \mathcal{P} , for which we use an independence sampler. Some implementation details can be found in [1].

RESULTS

In April 1967, a smallpox outbreak occurred in the town of Abakaliki, Nigeria [3, 4]. The outbreak was largely confined to the members of a fundamentalist Christian sect, the Faith Tabernacle church (FTC). 120 FTC members lived distributed among 9 compounds, along with 177 non-FTC members. The FTC members generally did not mix with their non-FTC brethren; however, they mixed strongly with their compound members and somewhat less strongly with FTC members who did not live in the same compound, via social visits and church gatherings (four times a week). The disease was introduced into the FTC community by a small girl. Thereafter, the disease was seen to spread strongly in a compound, followed by sporadic ‘‘hops’’ to other compounds. The FTC members refused medical treatment (i.e. vaccinations) and did not even separate the sick (and therefore contagious members). All in all, 7 compounds (consisting of 74 FTC members) and 32 people were affected, of which 30 were FTC members. The World Health Organization recorded the outbreak [3], in the form of dates of exhibition of symptoms of the 32 victims, their compound affiliations etc, but did not record their fate (i.e. dates of recovery or death). The outbreak lasted 3 months. Our aim is to estimate β_{in} , β_{cross} and the expected infection pathway, $\langle \mathcal{P} \rangle$ from the partially observed epidemic described above. We will only consider the spread of the disease among 74 FTC members, distributed in 7 compounds, who accounted for 30 (out of 32) smallpox victims.

We investigate the problem under two separate simplifications. We first assume that the FTC members were fully connected i.e. $p_{in} = \rho = 1.0$ and the difference in disease dynamics could be expressed using different β_{in} and β_{cross} . We infer these β from data. Using data from the entire epidemic we find β_{in} to be $6.7 \times 10^{-3} \text{day}^{-1}$ (median; 90% CI: $\{4.1 \times 10^{-3}, 10.3 \times 10^{-3}\}$) while $\beta_{cross} = 1.6 \times 10^{-3} \text{day}^{-1}$ (median; 90% CI: $\{0.9 \times 10^{-3}, 2.7 \times 10^{-3}\}$). We re-estimated the same quantities using data from the first 40 days, by which time smallpox cases had been observed in 4 different compounds. The corresponding values are $4.8 \times 10^{-3} \text{day}^{-1}$ (median; 90% CI: $\{2.6 \times 10^{-3}, 7.7 \times 10^{-3}\}$) and $0.8 \times 10^{-3} \text{day}^{-1}$ (median; 90% CI: $\{0.3 \times 10^{-3}, 1.6 \times 10^{-3}\}$). We see from the median values that $\beta_{in} \approx 4\beta_{cross}$.

Our second simplified approach assumes that smallpox spreads at the same rate inside and across compounds (i.e. $\beta_{in} = \beta_{cross} = \beta$) but the slower cross-compound transmission is due to a paucity of ‘‘strong’’ cross-compound social links. In this case, we assume complete mixing inside a compound (i.e.

$p_{in} = 1$), but only partial connectivity exists across compounds (i.e. $\rho < 1$). In this problem, we infer β , ρ , the social graph \mathcal{G} and the infection pathway \mathcal{P} . We used data from the first 40 days of the epidemic only. Fig. 1 shows the expected infection pathway $\langle \mathcal{P} \rangle$. Individuals are colored by their compound associations. The estimate of β was $3.3 \times 10^{-3} \text{day}^{-1}$ (median; 90% CI: $\{2.0 \times 10^{-3}, 5.3 \times 10^{-3}\}$).

CONCLUSIONS

We have presented a Bayesian method by which transmission rates, social networks and infection pathways can be inferred from a partially observed epidemic. The uncertainty in the values are quantified as a part of the inference procedure. This method can be used to estimate the characteristics of emerging infectious diseases from outbreaks, which are generally poorly observed and recorded; any significant change in these characteristics can signal an important step in its evolution and human virulence. The primary challenge in the method lies in inferring the social network and any informative prior belief regarding their structure (which may be possible for military societies) has the potential to significantly simplify the problem.

ACKNOWLEDGMENTS

This work was supported by Sandia National Laboratories' LDRD (Laboratory Directed Research and Development) program, with funds allocated under the "Enabling Predictive Simulation" Investment Area. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

References

- [1] Tom Britton and P. O'Neill. Bayesian inference for stochastic epidemics in populations with random social structure. *Scandinavian Journal of Statistics*, 29:375–390, 2002.
- [2] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1996.
- [3] D. Thompson and William Foege. Faith Tabernacle smallpox epidemic. Technical Report WHO/SE/68.3, World Health Organization, 1968.
- [4] Martin Eichner and Klaus Dietz. Transmission potential of smallpox: Estimates based on detailed data from an outbreak. *American Journal of Epidemiology*, 158(2):110–117, 2003.

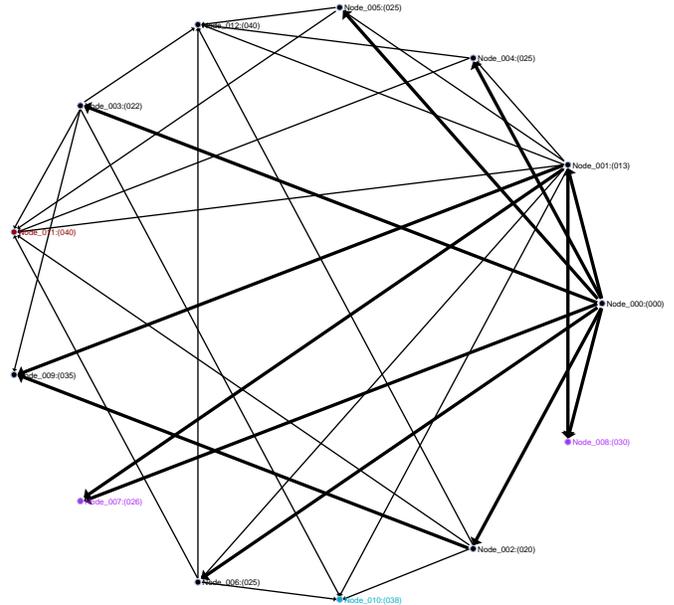


Figure 1: The expected infection pathway (or the chains of transmission) $\langle \mathcal{P} \rangle$. Only links with probability greater than 0.1 are plotted; those with probability greater than 0.3 are plotted with thick lines.