Using Contact Tracing Data to inform inference of transmission networks

Finlay Campbell

July 12, 2016

1 Introduction

- The value and importance of outbreak reconstruction
- The problems and issues with outbreak reconstruction
  - Integrating multiple data sources in non-arbitrary ways
  - Existing data sources having an insufficient information content
- Describe various previous approaches to outbreak reconstruction, with their respective benefits and drawbacks
- Introduce Outbreaker as an integrated Bayesian approach
  - Describe general approach of Outbreaker
  - Describe situations in which Outbreaker struggles (mutation rate too low, generation time too broad, etc.)
  - CTD as an additional data source is able to fill many of these gaps
- Describe the availability and types of CTD available, and the benefit of CTD over other types of genetic and epidemiological data
- Describe previous uses of CTD in informing outbreak reconstruction
- Describe the aim of this project: integrating CTD as an additional data source to inform outbreak reconstruction in a Bayesian framework

2 Methods

2.1 Simulating CTD

CTD is simulated from known transmission networks and used to test the potential of CTD in informing outbreak reconstruction. Our simulation function incorporates the probability of observing contact between a transmission pair (ϵ) in order to simulate CTD of variable coverage and sensitivity, as well as a scaling factor (ξ) to describe the probability observing contact between individuals not in a transmission pair, relative to ϵ. This allows the simulation of misinformative CTD. The probability of observing contact (c_{ij}=TRUE) between any two individuals is calculated as follows:

$$\text{For transission pairs} — p(c_{ij} = \text{TRUE}) = \epsilon \tag{1}$$

$$\text{For non-transission pairs} — p(c_{ij} = \text{TRUE}) = \epsilon * \xi \tag{2}$$

2.2 Contact Tracing Likelihood

The contact tracing likelihood Ω_3 describes the probability of obtaining the observed contact relationship between two individuals (c_{ij}=TRUE or FALSE), given the proposed transmission network. It uses a contact reporting probability ϵ as its only parameter, describing the probability of observing contact (c_{ij}=TRUE) between a transmission pair. The probability of not observing contact (c_{ij}=FALSE) is therefore the compliment of ϵ. Formally, the contact tracing likelihood is described as follows:

$$\text{For } c_{ij} = \text{TRUE} \text{ — } \Omega_3 = \epsilon \tag{3}$$

$$\text{For } c_{ij} = \text{FALSE} \text{ — } \Omega_3 = 1 - \epsilon \tag{4}$$

3 Results

Figure 1: Performance changes upon incorporating contact tracing data into Outbreaker2 (ϵ=1, ξ=0.01)

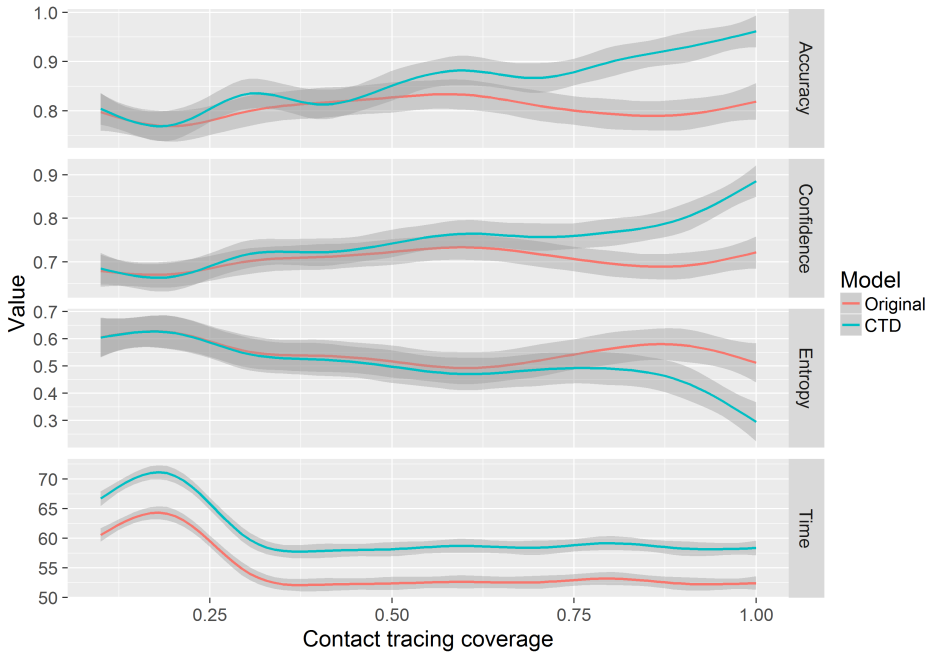

Figure 2: Relative performance changes of CTD.outbreaker with contact tracing coverage of simulated CTD

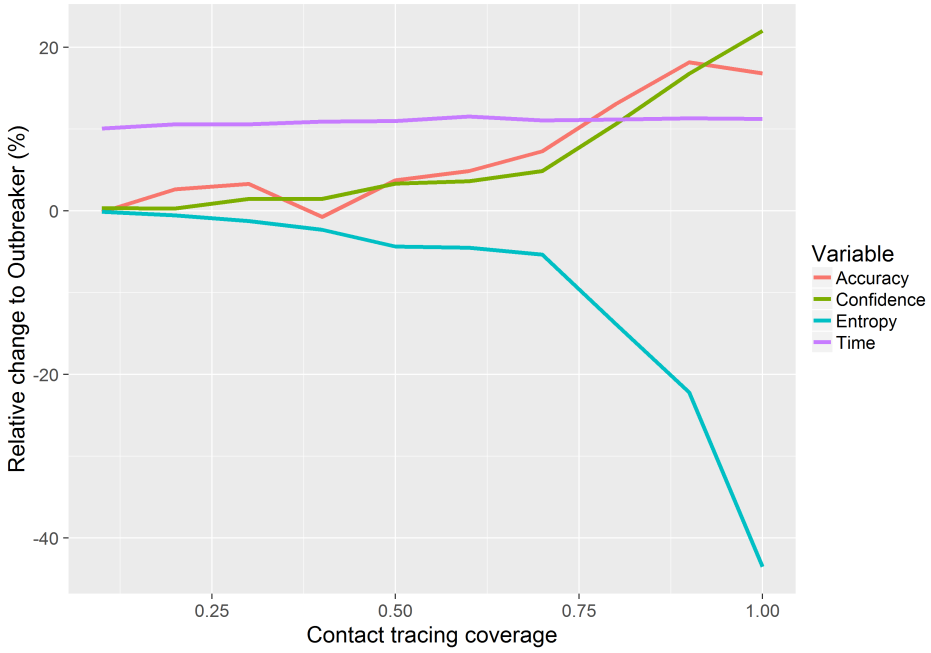

Figure 3: Performance changes of CTD.outbreaker with contact tracing coverage of simulated CTD

4 To do:

- Do a grid of xi/eps
- Make table of things to look at and then decide

- Look at large/small outbreak
- Look at high/low mutation
- SimOutbreak allows superspreading (0.1 of superspreader R0 or 5 times higher)
- Value for R0/generation time
  - Use Ebola values
  - Or SARS (but strong genetic signature)
  - But also look at high/low R0 (might make it more difficult)
  - Central setup being similar to Ebola
  - Then do theoretical examples for other situations
- Add prior for eps in description
- (DONE) Label order for relative.performance
- (DONE) Don't enforce scaling on violinplot
- (DONE) Add uncertainty to absolute.plot