Developing a method for inferring transmission chains of MRSA in a hospital setting
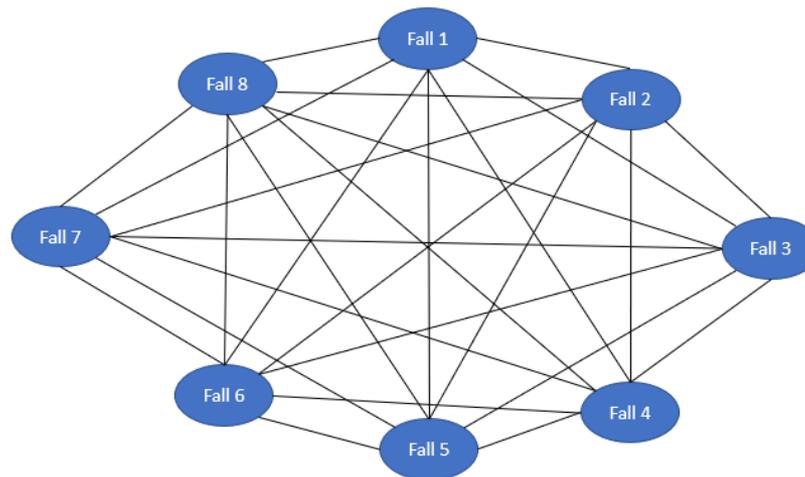
Philip Gerlee & Jon Edman Wallér
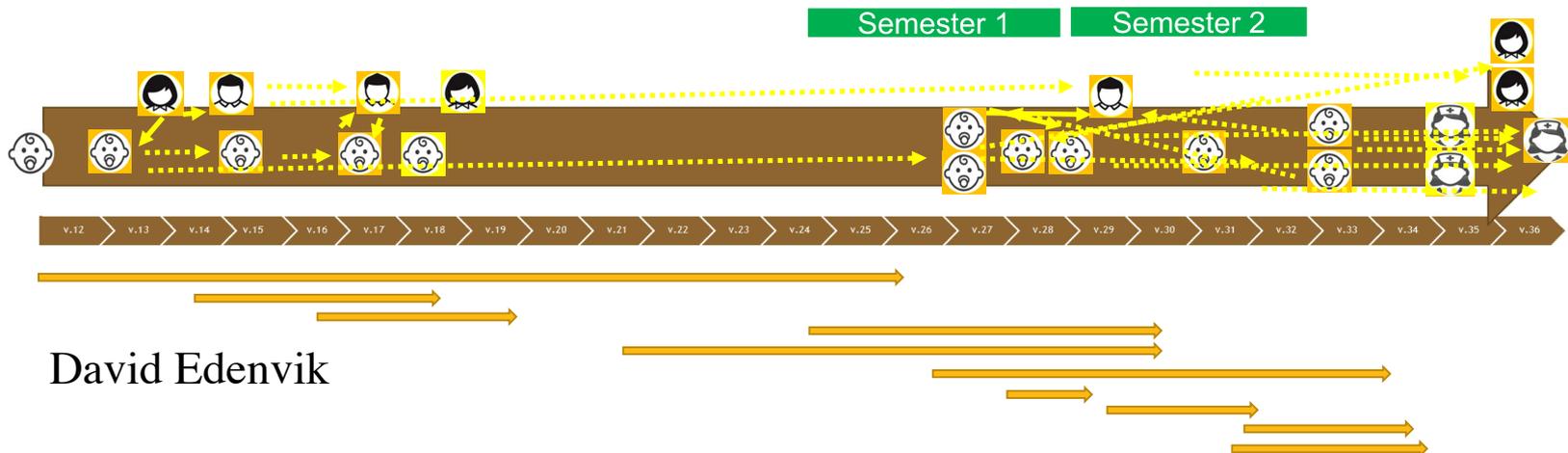
Motivation

- MRSA outbreak in a neonatal ward (Ryhov, Jönköping)

- Improved hygiene routines requires knowledge about routes of transmission

- Hypothesis: mathematical modelling can aid in finding the most likely chain of transmission

David Edenvik

Introduction

- Mathematical modelling can aid infection control in many ways:
  - From outbreak data we can estimate disease-specific parameters, e.g. infectivity, effect of isolation
  - With a validated model we can simulate interventions, e.g. improved hygiene, screening programmes
  - We can infer the most likely transmission chain given outbreak data, e.g. contacts, timing of symptoms

Goal of study

- We aim to build a mathematical model for inference of transmission chains

- Input: time of symptoms, genetic sequences, contact data, swab results, healthcare worker schedules, generation & incubation time

- Output: Transmission trees

RESEARCH ARTICLE

Bayesian inference of transmission chains using timing of symptoms, pathogen genomes and contact data

Finlay Campbell[1]*, **Anne Cori**[1], **Neil Ferguson**[1], **Thibaut Jombart**[1,2,3]*

Bayesian framework

- In a Bayesian model the parameters θ (including transmission trees) are considered random variables

- A prior probability distribution of parameters $P(\theta)$ is updated using data D according to:

Likelihood

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}$$

Likelihood: an example

- Assume that have an unfair coin that shows Heads with probability *p* and Tails with *1-p*

- Assume we have data: $D = HTHTT$

- What value of *p* should we choose?

- The likelihood of the data
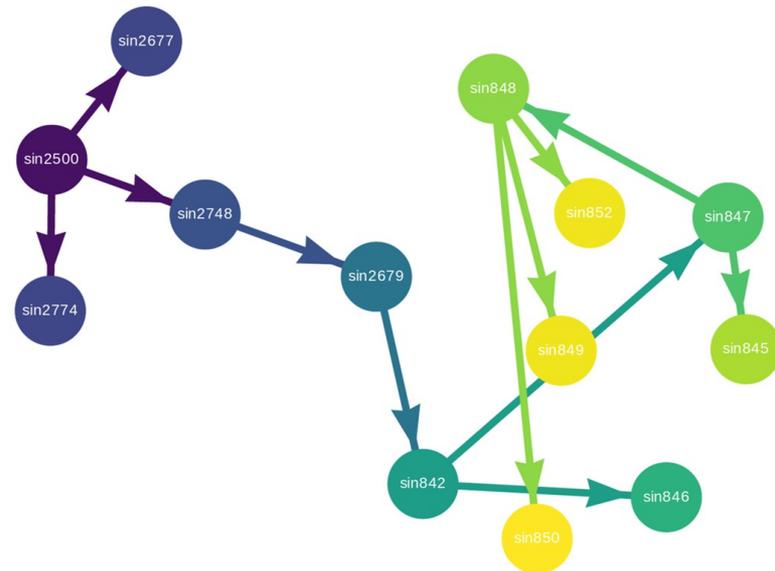$$P(D \mid p) = p^2(1-p)^3$$

- Likelihood maximised at
$$\hat{p} = \frac{2}{5}$$
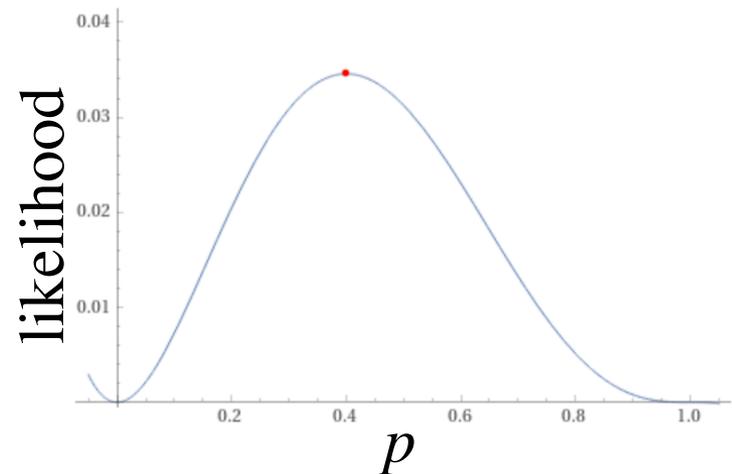
The outbreaker model (Campbell et al.)

- A transmission tree is described by α_i the most recent sampled ancestor of case i and κ_i the number of generations between ancestor and case

- Remember: α_i is a random variable e.g. with 5 cases we might get

$$\alpha_1 = (0, 0.1, 0.1, 0.1, 0.7)$$

D) Posterior ancestries (TCG, λ = 0)

Likelihood in outbreaker

- The likelihood of a transmission tree is composed of four terms:
  - Genetic likelihood
  - Temporal likelihood
  - Reporting likelihood
  - Contact likelihood

Genetic likelihood

- This takes into account the sequence data
- The likelihood of case j being the ancestor of i given
  - genetic similarity
  - assumed mutation rate
  - generations separating the cases

Temporal likelihood

- This takes into account the time of symptom onset and depends on an inferred time of infection

- The likelihood of case j having infected i given
  - symptom time
  - generation time distribution (serial interval)
  - incubation period distribution

Reporting likelihood

- Describes the probability of unobserved intermediate cases

- Depends on
  - number of generations separating the cases
  - probability of observation

Contact likelihood

- Depends on a contact matrix where $c_{ij} = 1$ if there is a reported between case i and j and $c_{ij} = 0$ otherwise
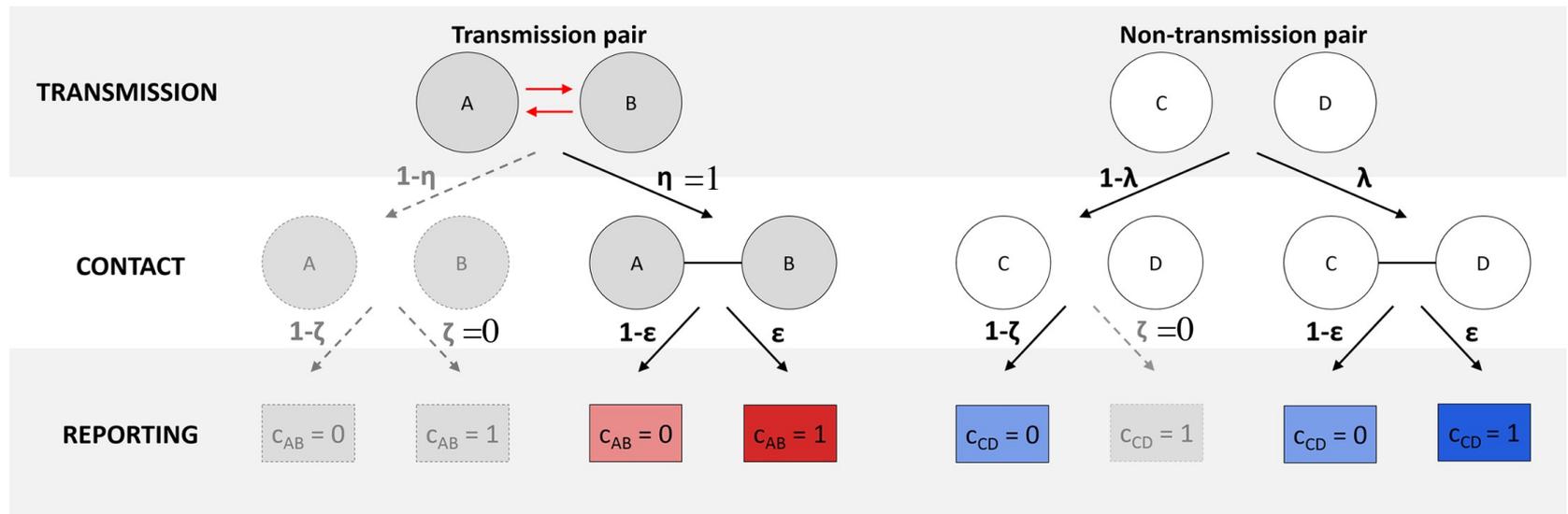
Extensions

- Disease transmission via healthcare workers can be accounted for by:
  - Including them into the contact data
  - Defining a new type of indirect contact
- Routine swabs of patients, parents, HCWs and environment should affect the reporting likelihood

Questions & suggestions?