Degree truncation and its impact on spreading process outcomes

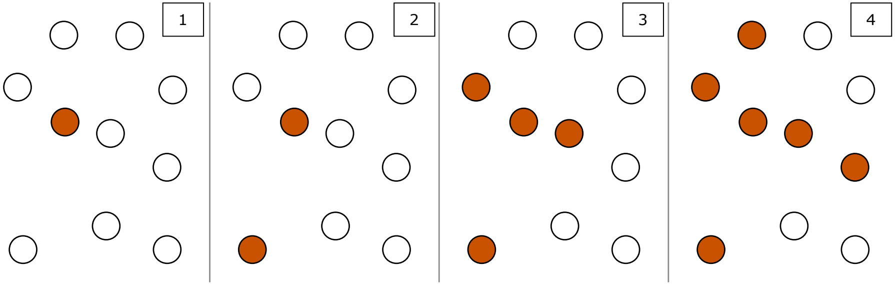
Guy Harling, Jukka-Pekka Onnela

Departments of Global Health and Population & Biostatistics
Harvard T.H. Chan School of Public Health
Harvard University

25 June 2015
Brighton, England

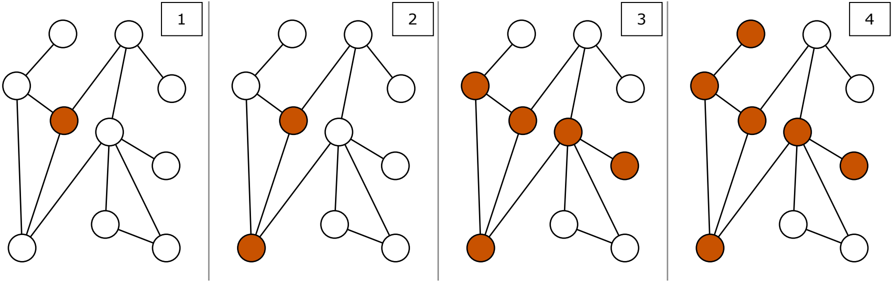
Goal

To predict in advance how a process will spread through a population



Requirement

Knowledge of population contact structure



Elucidating contact structure

- Common approaches
  1. Ask everyone for the names of all their contacts
  2. Ask some people about their contacts & contacts' interrelations
- Problems
  1. The former exhausts resources
  2. The latter exhausts respondents (& resources)

So what do we do?

Fixed choice designs (FCD)

Limit the number of contacts asked about

- Truncation of nodal degree

$$k_i^o = \begin{cases} k_i & \text{if } k_i < k_{fc} \\ k_{fc} & \text{otherwise} \end{cases} \tag{1}$$

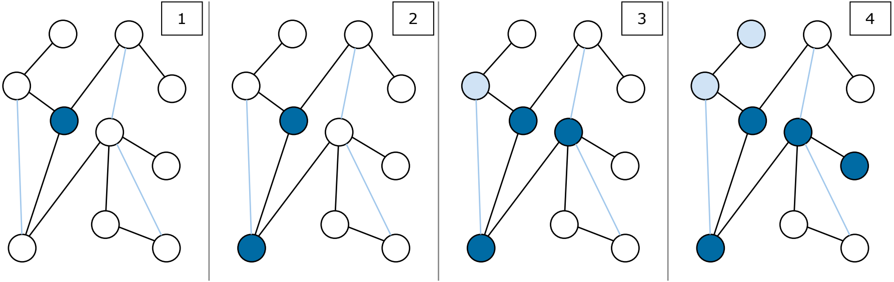
- Different truncation methods
  - Strength: "Who are the five people you spend most time with"
  - Random: "Tell me everyone you spoke to in the past week"
- If undirected, both nodes must non-report to drop an edge

Burt 1984. Network Items and the General Social Survey. Social Networks

Motivation

Impact of fixed choice designs

Change in information will change our predictions



Research Questions

- How does truncation by fixed-choice design affect predictions of:
  1. Speed of spread
  2. Final size?
- How do these changes vary by network characteristics:
  1. Variability of nodal degree distribution
  2. Degree assortativity of nodes
  3. Clustering ?

Working hypotheses

FCD truncation impacts on spreading process outcomes:

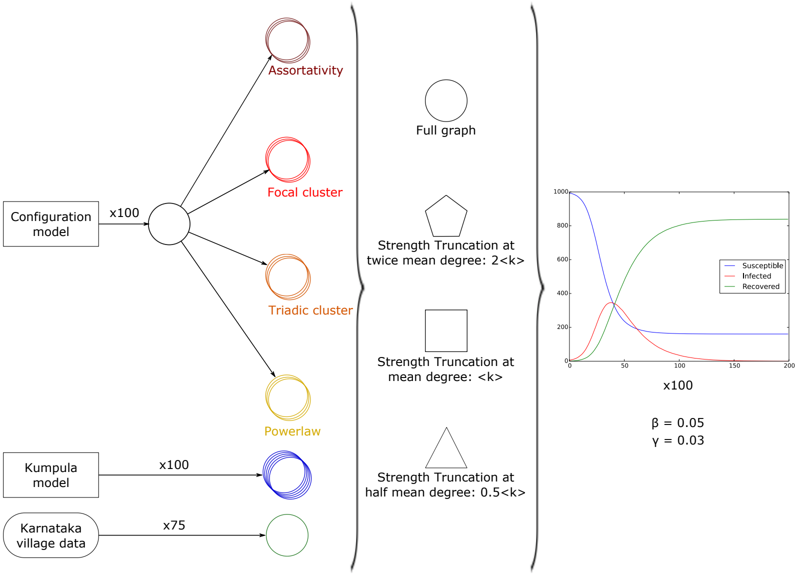
- Reduced number of connections will reduce speed of spread
  - Unless degree variance reduced significantly too
  - Since greater variance leads to faster take-off
- Higher assortativity networks leads to smaller, faster spread
  - Assortativity rises under strength truncation
  - Assortativity often leads to dense core, sparse periphery
- Increased clustering reduces speed of spread and final size
  - Clustering rises under strength truncation
  - Clustering means more redundant ties/smaller LCC

Methods: Datasets

1. Synthetic datasets
    - Degree-assortative
    - Clustered
    - Powerlaw degree distribution
2. Empirical datasets: 75 villages in Karnataka state, India
    - Respondents reported 12 types of interactions
    - Tie present if any report of interaction by either party

Watts & Strogatz 1998. Collective dynamics of 'small-world' networks. Nature.
Saramäki et al. 2007. Generalizations of the clustering coefficient to weighted complex networks. Phys Rev E.
Kumpula et al. 2007. Emergence of Communities in Weighted Networks. Phys Rev Lett.
Banerjee et al 2013. The Diffusion of Microfinance. Science.

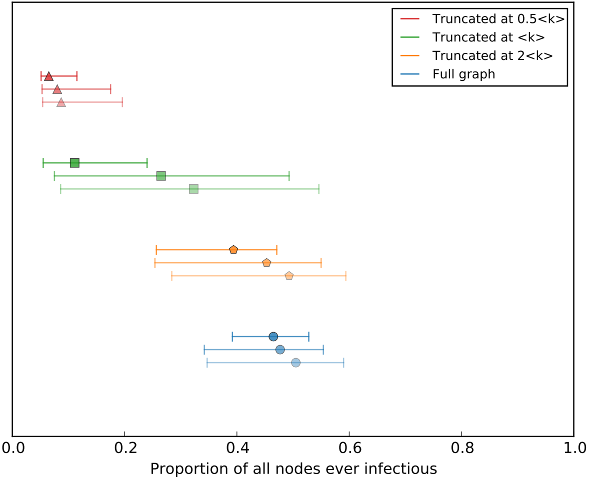
Methods: Simulation



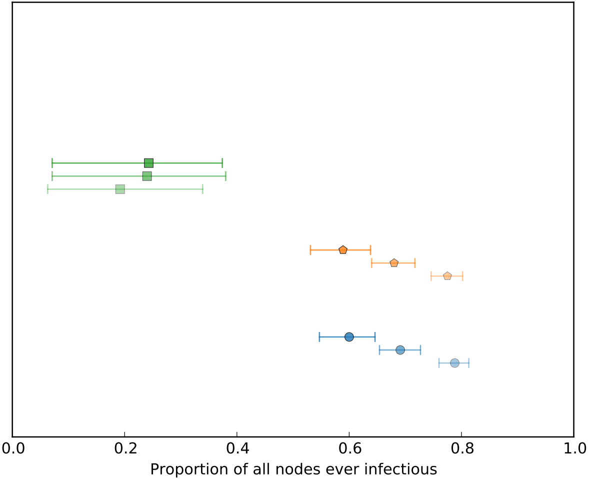
Results: Impact of truncation on network

1. Consistent reduction in:
   - Mean degree
   - Triadic clustering
   - Degree assortativity (except for powerlaw networks)
2. Modularity
   - Consistent increase for Powerlaw networks
   - Increase only for $0.5\langle k \rangle$ truncation for clustered networks
   - Falls for Assortative networks (from very high)
   - Small, steady increase for Karnataka networks

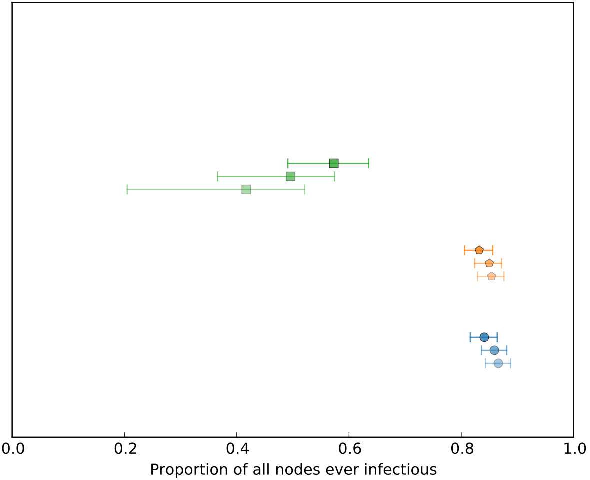
Results: Final Size: Assortative



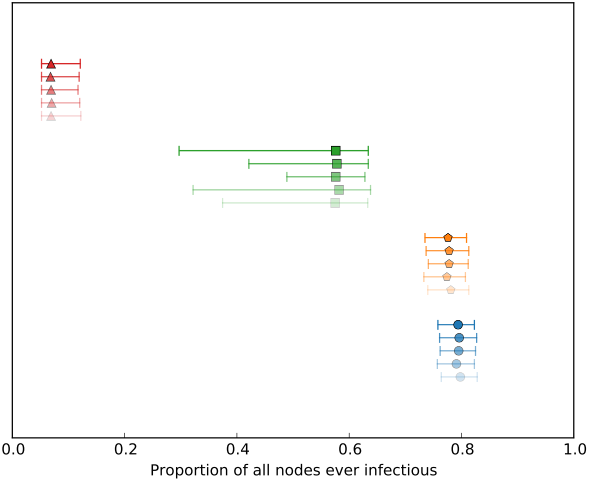
Results: Final Size: Focal clustering



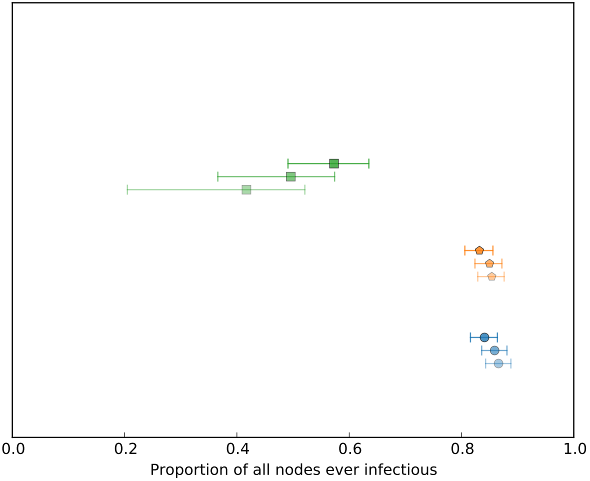
Results: Final Size: Triadic clustering



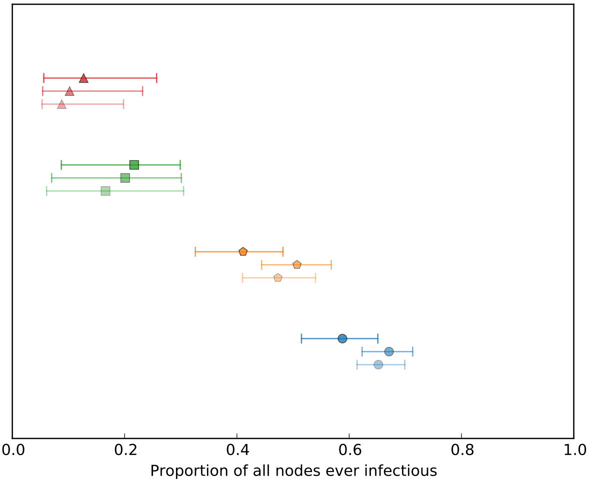
Results: Final Size: Community structure



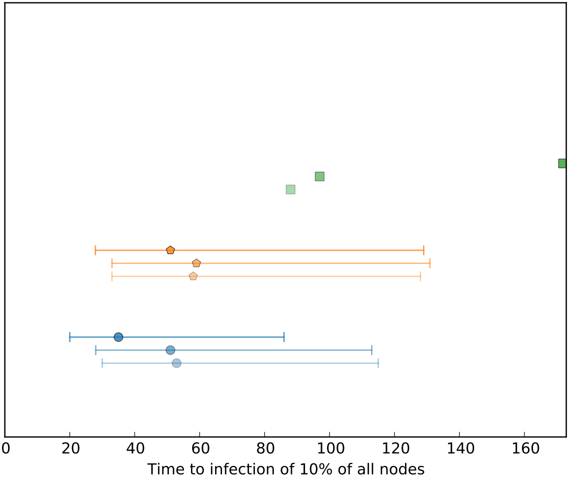
Results: Final Size: Karnataka



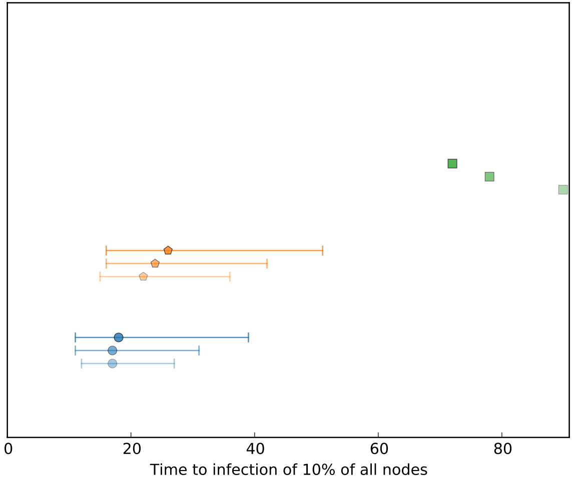
Results: Final Size: Powerlaw



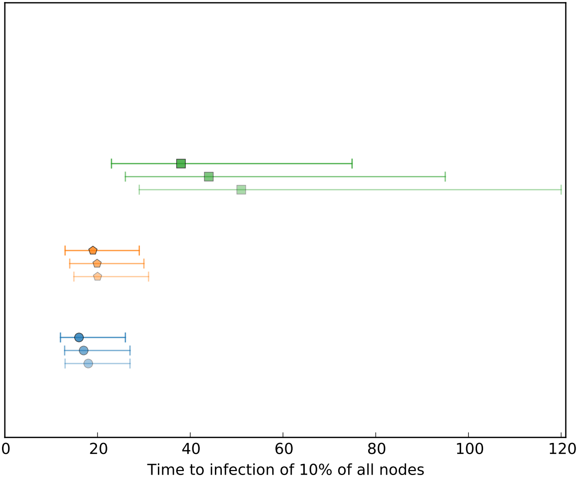
Results: Speed of spread: Assortative



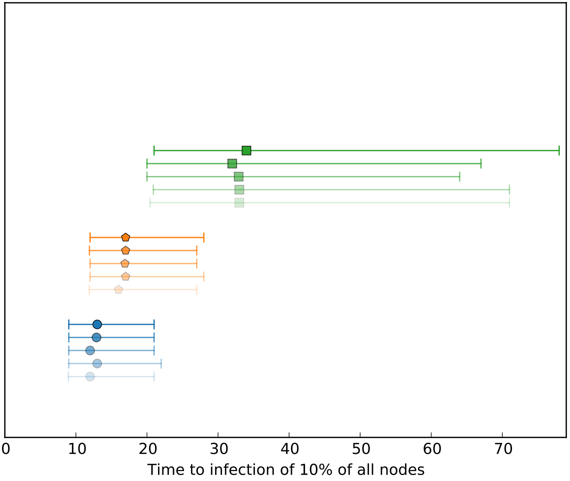
Results: Speed of spread: Focal clustering



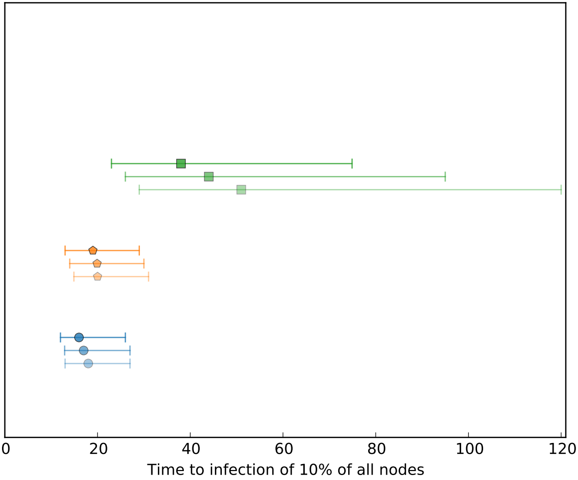
Results: Speed of spread: Triadic clustering



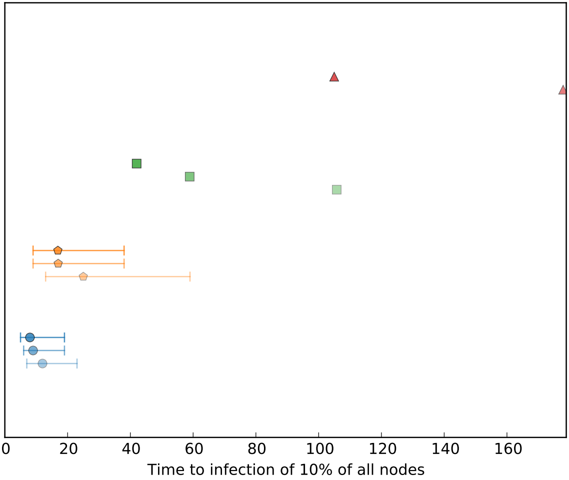
Results: Speed of spread: Community structure



Results: Speed of spread: Karnataka



Results: Speed of spread: Powerlaw



Discussion

- Truncating at mean degree significantly affects our predictions
- Reduced connections overwhelm other network features
  - Except in powerlaw networks: extreme disassortativity?

Possible next steps

- Reproduce these results using empirical spreading processes?
- Considering truncation as a specific kind of sampling:
  - Can we validly infer characteristics of missing ties?
  - Can we validly predict process outcomes from truncated data?
  - Do we need the former to achieve the latter?

Acknowledgements

- This work is supported by P30-AG-034420 from NIA (PI Nicholas Christakis)
- This work is conducted in the Onnela lab: `http://www.hsph.harvard.edu/onnela-lab/`

You can reach me at `gharling@hsph.harvard.edu` & `@harlingg`