Phylogenetic inference of contact network parameters with kernel approximate Bayesian computation

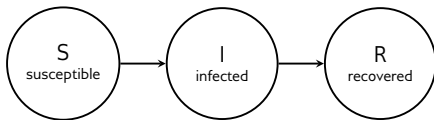Rosemary M McCloskey[1]   Richard H Liang[1]   Art FY Poon[1,2]

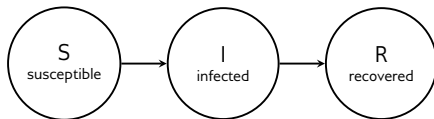HIV Dynamics & Evolution, Woods Hole, USA, April 25, 2016

[1]BC Centre for Excellence in HIV/AIDS, Vancouver, Canada
[2]Department of Medicine, University of British Columbia, Vancouver, Canada
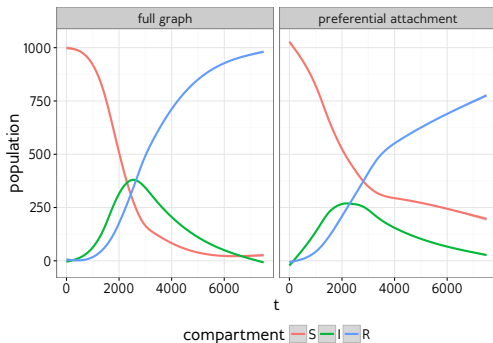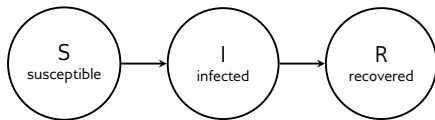
Most epidemiological models assume homogeneous mixing
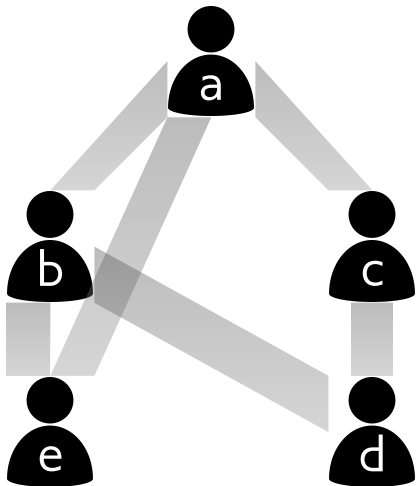

- Often provide a reasonable approximation in practice.

- Often provide a reasonable approximation in practice.

- Can be inaccurate when substantial contact heterogeneity exists.
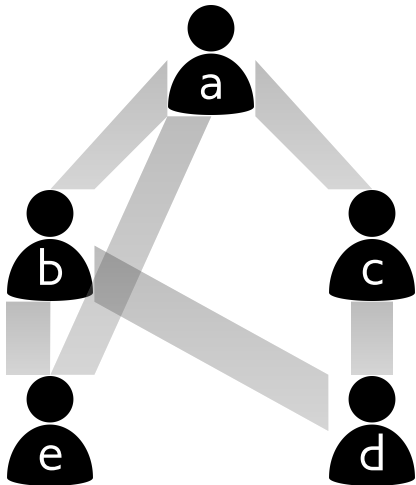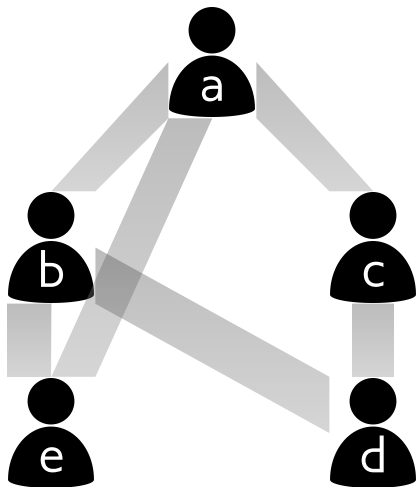
- May offer more accurate predictions for highly structured populations.

- May offer more accurate predictions for highly structured populations.

- Network parameters may be of interest for their own sake, *e.g.* are there superspreaders?

- May offer more accurate predictions for highly structured populations.

- Network parameters may be of interest for their own sake, *e.g.* are there superspreaders?
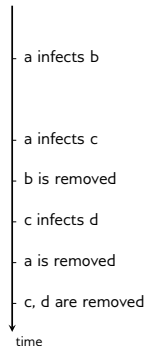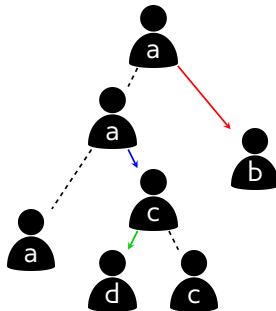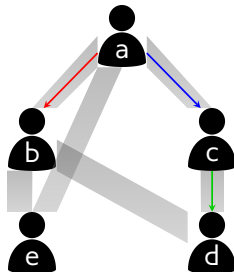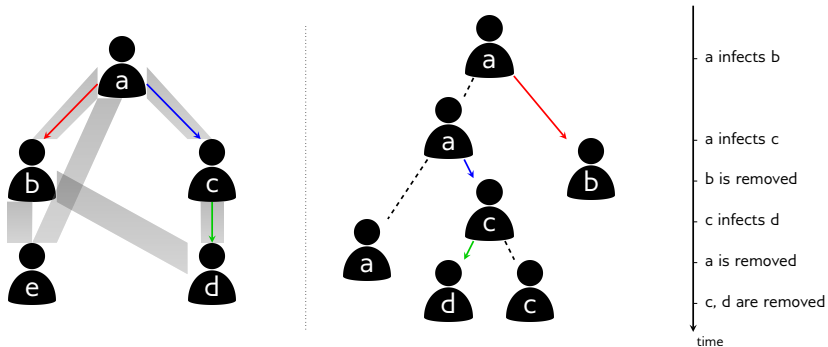
- Extremely difficult to estimate in practice.

a infects b

a infects c

b is removed
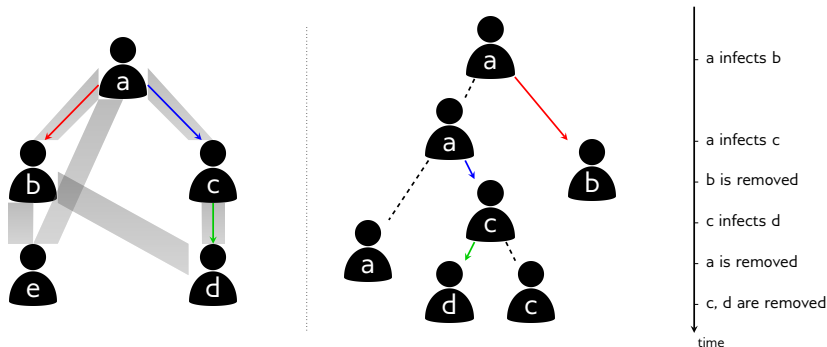
c infects d

a is removed

c, d are removed

time

4

Contact networks shape transmission trees

a infects b

a infects c

b is removed

c infects d
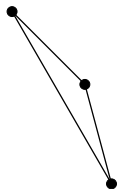
a is removed

c, d are removed

time

- Transmission trees in turn shape viral phylogenies.
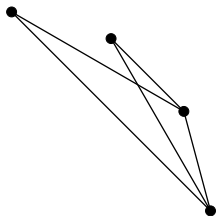
Contact networks shape transmission trees

- Transmission trees in turn shape viral phylogenies.
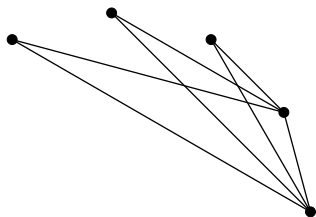- **Aim:** estimate contact network parameters from viral phylogenies.

- Start with a small number of connected nodes.

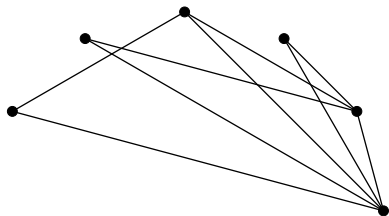# Barabási-Albert model incorporates preferential attachment

- Start with a small number of connected nodes.
- Attach new nodes with *m* edges.

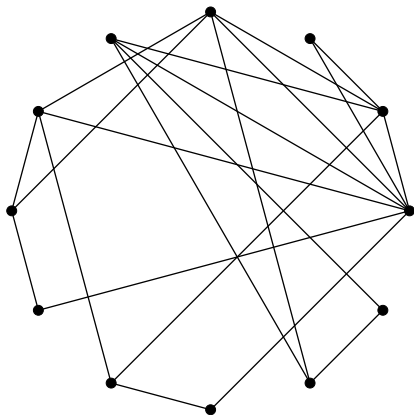# Barabási-Albert model incorporates preferential attachment

- Start with a small number of connected nodes.
- Attach new nodes with *m* edges.
- Other endpoints of degree *d* are chosen with probability $\propto d^{\alpha} + 1$.

Barabási-Albert model incorporates preferential attachment

- Start with a small number of connected nodes.
- Attach new nodes with *m* edges.
- Other endpoints of degree *d* are chosen with probability $\propto d^{\alpha} + 1$.
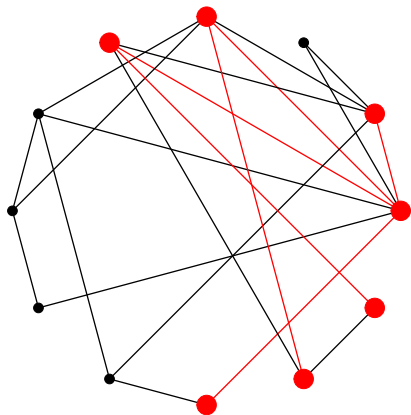
Barabási-Albert model incorporates preferential attachment

- Start with a small number of connected nodes.
- Attach new nodes with m edges.
- Other endpoints of degree d are chosen with probability $\propto d^{\alpha} + 1$.
- Continue until network has N nodes.

- Start with a small number of connected nodes.
- Attach new nodes with m edges.
- Other endpoints of degree d are chosen with probability $\propto d^{\alpha} + 1$.
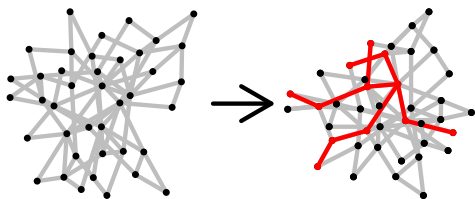- Continue until network has N nodes.
- Also consider the prevalence l.
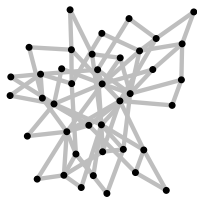
- Generate networks under different parameter values (number of nodes *N*, number of edges per vertex *m*, pereferential attachment power α).

---

[1]Art FY Poon et al. (2013). "Mapping the shapes of phylogenetic trees from human and zoonotic RNA viruses". In: *PLoS ONE* 8.11.

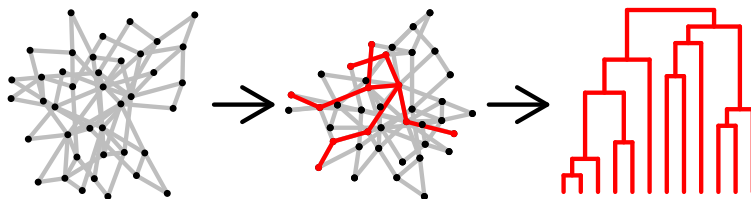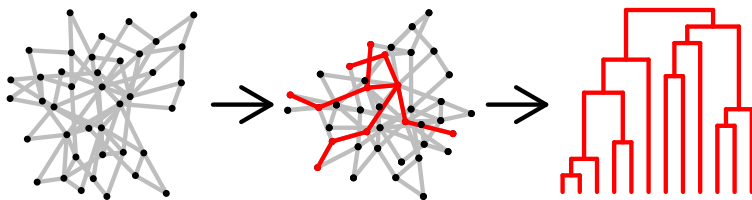# Do network parameters measurably affect tree shape?

- Generate networks under different parameter values (number of nodes N, number of edges per vertex m, pereferential attachment power α).
- Simulate epidemic over each network until I nodes are infected.

---

[1]Art FY Poon et al. (2013). "Mapping the shapes of phylogenetic trees from human and zoonotic RNA viruses". In: *PLoS ONE* 8.11.

- Generate networks under different parameter values (number of nodes N, number of edges per vertex m, pereferential attachment power α).
- Simulate epidemic over each network until I nodes are infected.
- Randomly subsample to form transmission trees.

---

[1]Art FY Poon et al. (2013). "Mapping the shapes of phylogenetic trees from human and zoonotic RNA viruses". In: *PLoS ONE* 8.11.

# Do network parameters measurably affect tree shape?

- Generate networks under different parameter values (number of nodes N, number of edges per vertex m, pereferential attachment power α).
- Simulate epidemic over each network until I nodes are infected.
- Randomly subsample to form transmission trees.
- Compare trees pairwise using tree kernel[1].

---

[1]Art FY Poon et al. (2013). "Mapping the shapes of phylogenetic trees from human and zoonotic RNA viruses". In: *PLoS ONE* 8.11.

Preferential attachment power α affects tree shape
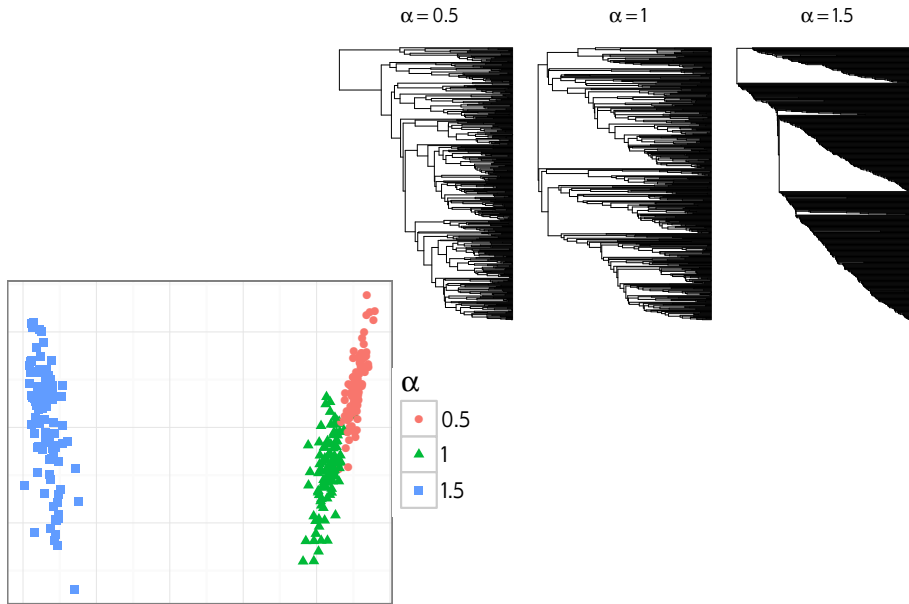
Prevalence *I* affects tree shape
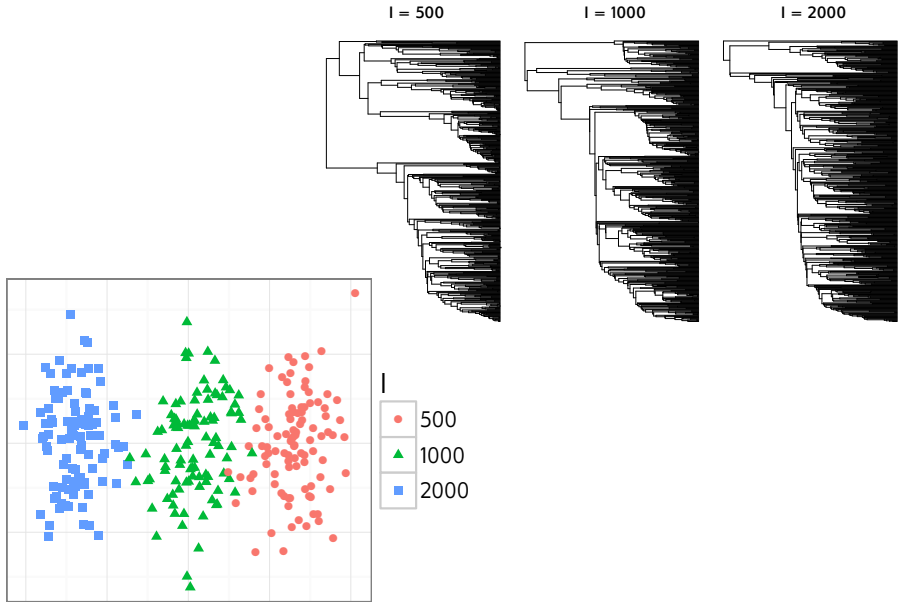
Number of edges per vertex *m* does not affect tree shape
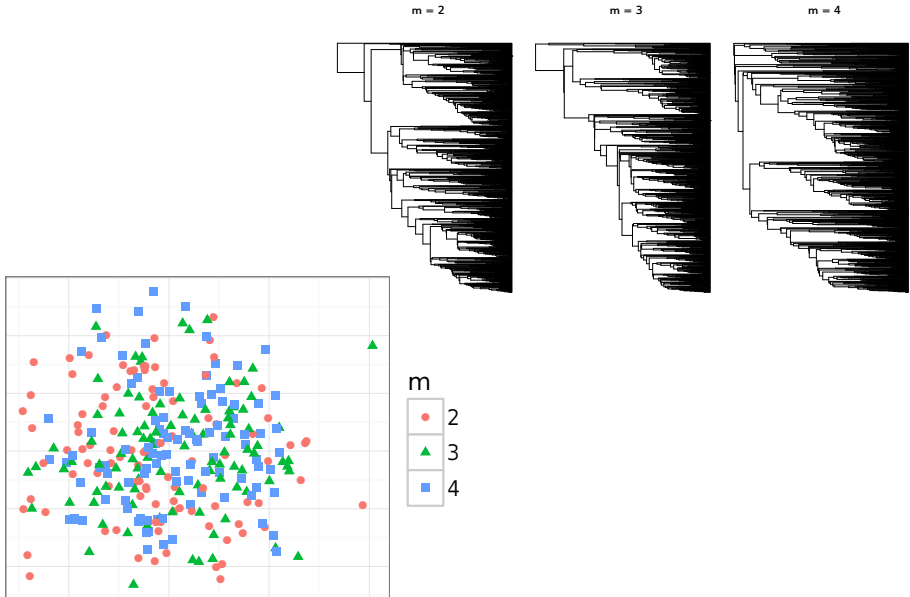
N = 3000   N = 5000   N = 8000

N
- 3000
- 5000
- 8000

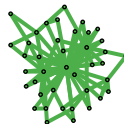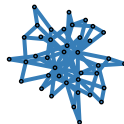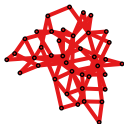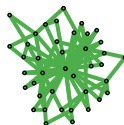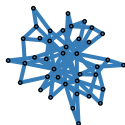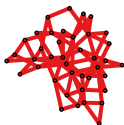sample parameters θ_1 θ_2 θ_3

sample
parameters

θ_1

θ_2

θ_3
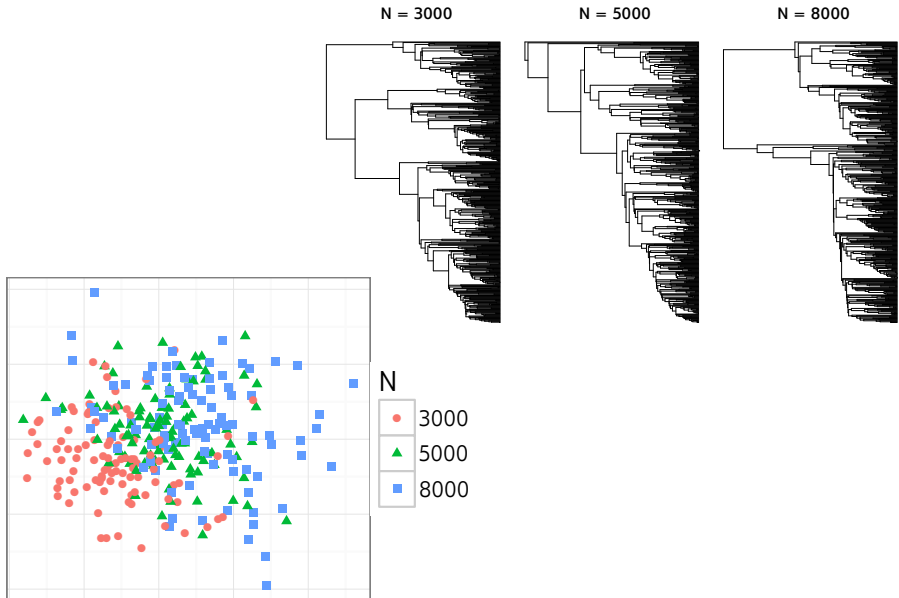
simulate
networks

Simulated trees can be used to fit a network model

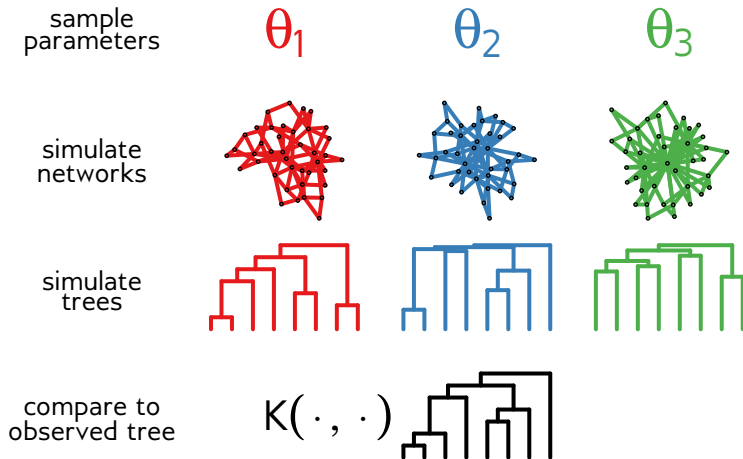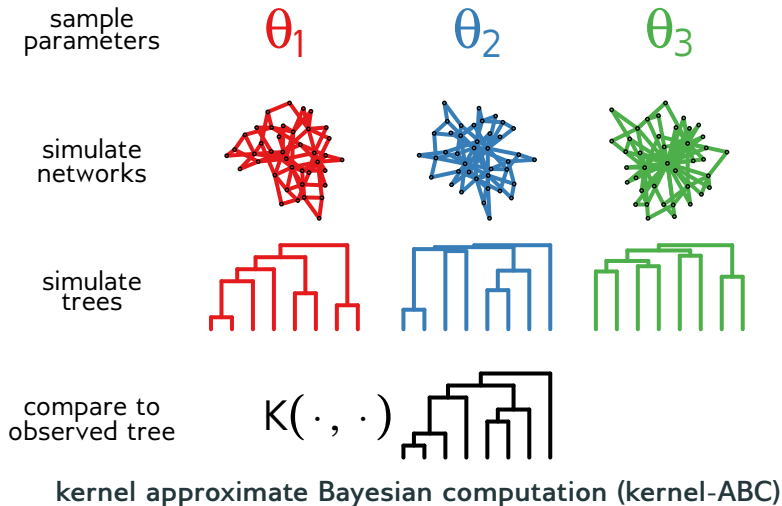sample parameters θ_1 θ_2 θ_3

simulate networks

simulate trees

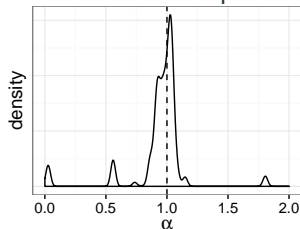compare to observed tree $K(\cdot, \cdot)$

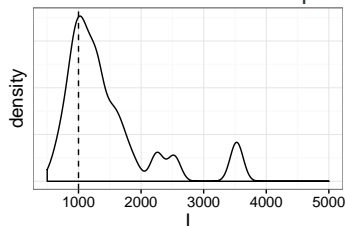# Simulated trees can be used to fit a network model

Example simulation shows α and l can be reconstructed
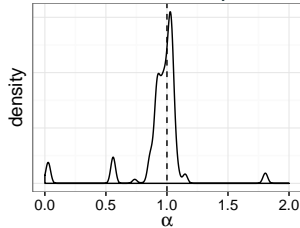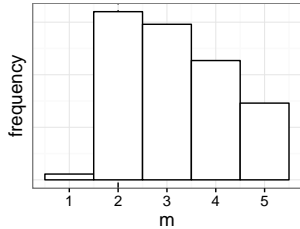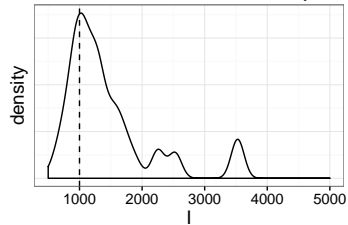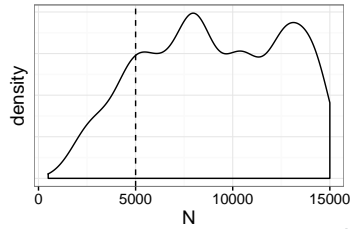
preferential attachment power

prevalence

preferential attachment power

prevalence

edges per vertex

total nodes
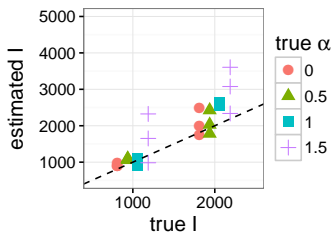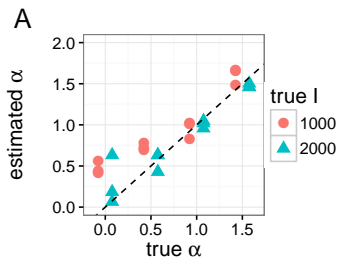
preferential attachment power

prevalence

preferential attachment power

edges per vertex

A

B
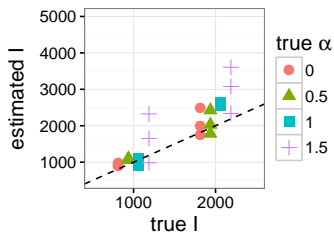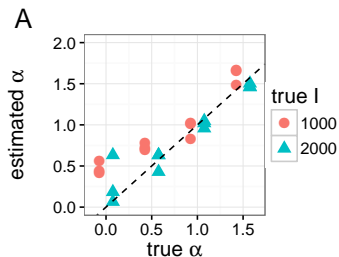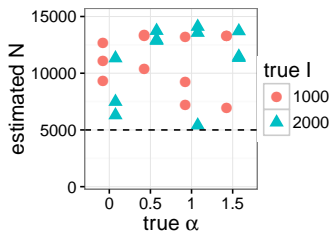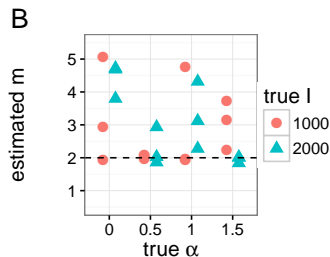
prevalence

total nodes

Real world HIV datasets exhibit network heterogeneity
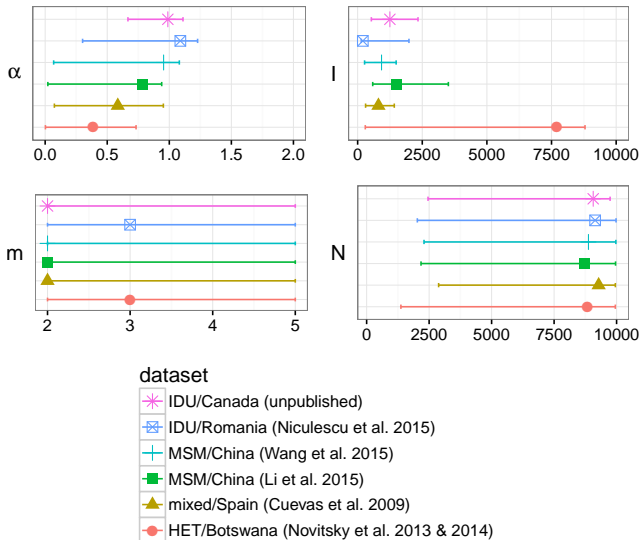
Conclusions

- Kernel-ABC is the first phylodynamic method to fit contact network models to phylogenetic data.

Conclusions

- Kernel-ABC is the first phylodynamic method to fit contact network models to phylogenetic data.
- The preferential attachment power of the Barabási-Albert network model, which is challenging to estimate by traditional epidemiological methods, can be estimated with kernel-ABC.

Conclusions

- Kernel-ABC is the first phylodynamic method to fit contact network models to phylogenetic data.
- The preferential attachment power of the Barabási-Albert network model, which is challenging to estimate by traditional epidemiological methods, can be estimated with kernel-ABC.
- The networks underlying real epidemics are heterogeneous, underscoring the importance of considering network structure in phylodynamic analyses.

Conclusions

- Kernel-ABC is the first phylodynamic method to fit contact network models to phylogenetic data.
- The preferential attachment power of the Barabási-Albert network model, which is challenging to estimate by traditional epidemiological methods, can be estimated with kernel-ABC.
- The networks underlying real epidemics are heterogeneous, underscoring the importance of considering network structure in phylodynamic analyses.

 github.com/rmcclosk/netabc

Acknowledgements

BC Centre for Excellence in HIV/AIDS

Art Poon

Jeff Joy

Richard Liang

Thuy Nguyen

P. Richard Harrigan

University of British Columbia

Sarah Otto

Alexandre Bouchard-Côté
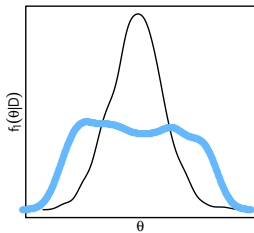
Classifiers for network parameters

- kSVR using tree kernel
- SVR using normalized lineages-through-time[2]
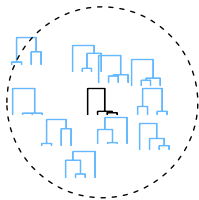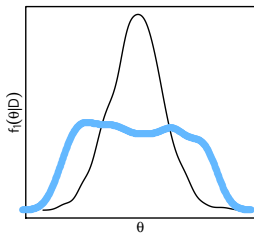- linear regression using Sackin's index[3]
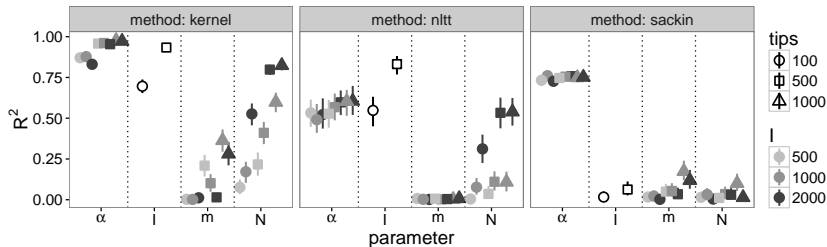
[2] Thijs Janzen, Sebastian Höhna, and Rampal S Etienne (2015). "Approximate Bayesian computation of diversification rates from molecular phylogenies: introducing a new efficient summary statistic, the nLTT". . In: *Methods in Ecology and Evolution* 6.5, pp. 566–575.

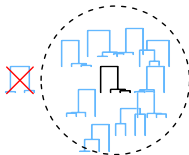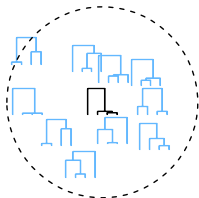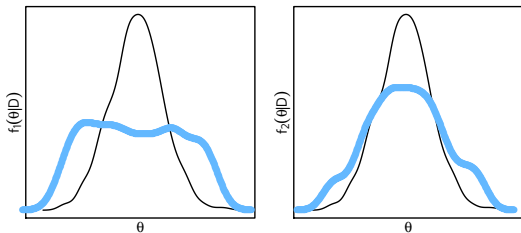[3] Kwang-Tsao Shao (1990). "Tree balance". In: *Systematic Biology* 39.3, pp. 266–276.

Sequential Monte Carlo for ABC