Dynamic network sampling

Steve Thompson

Simon Fraser University
thompson@sfu.ca

Graybill Conference
Colorado State University
June 10, 2013

Dynamic network sampling

- The **population** of interest has spatial structure,
  often has network structure
  and moves or changes over time

Dynamic network sampling

- The **population** of interest has spatial structure,
  often has network structure
  and moves or changes over time
- Designs for selecting a **sample** units
  use spatial and network relationships and progress dynamically

Population and sample processes

Population: A stochastic process $\{Y_t\}$.

Sample: A stochastic process $\{S_t\}$.

Time t such as day, with $t = 0, 1, 2, \ldots$.

Values Y_t of units such as locations, states, and relationships between units at time t.

Sample S_t the set of units in the sample at time t.

Spatial-temporal population model

Purpose of dynamic population model is to evaluate the effectiveness of different sampling designs. We want it to be simple but to incorporate those characteristics that affect the effectiveness of sampling strategies.

Spatial-temporal population model

Purpose of dynamic population model is to evaluate the effectiveness of different sampling designs. We want it to be simple but to incorporate those characteristics that affect the effectiveness of sampling strategies.

- clustering, mixing, migration—point process

Spatial-temporal population model

Purpose of dynamic population model is to evaluate the effectiveness of different sampling designs. We want it to be simple but to incorporate those characteristics that affect the effectiveness of sampling strategies.

- clustering, mixing, migration—point process
- movements within and among groups—small random displacements, MCMC selections, autoregressive processes

Spatial-temporal population model

Purpose of dynamic population model is to evaluate the effectiveness of different sampling designs. We want it to be simple but to incorporate those characteristics that affect the effectiveness of sampling strategies.

- clustering, mixing, migration—point process
- movements within and among groups—small random displacements, MCMC selections, autoregressive processes
- insertions and deletions of objects—birth and death process, immigration and emigration.

Dynamic network model

- builds on spatial temporal point process
- link probabilities dependent on distance between nodes, node characteristics, and current degree or target degree distribution
- renewal process for link formation, persistence, and dissolution

Sampling process

- A sampling design in the static situation is a procedure for selecting units to include in the sample.
- For the dynamic situation we use a sampling process that includes

Sampling process

- A sampling design in the static situation is a procedure for selecting units to include in the sample.
- For the dynamic situation we use a sampling process that includes
  - an **acquisition process** by which units are added to the sample

Sampling process

- A sampling design in the static situation is a procedure for selecting units to include in the sample.
- For the dynamic situation we use a sampling process that includes
  - an **acquisition process** by which units are added to the sample
  - an **attrition process** by which units are removed from the sample

Equilibrium distributions

Many properties of the population and sample process are ergodic and have stationary and limiting distributions.

Equilibrium distributions

Many properties of the population and sample process are ergodic and have stationary and limiting distributions.

- Sample size tends to increase when acquisition rate exceeds attrition rate
- and decrease when attrition rate exceeds attrition rate

Uses of sampling designs

- **Inference** about population characteristics
- **Experiments** on sample units
- **Interventions** on sample units

Intervention strategy

- Select a sample of units from the population, make interventions on those units changing their values.

Intervention strategy

- Select a sample of units from the population, make interventions on those units changing their values.
- Objective is to change the population, not just sample units, in a desired way.

Intervention strategy

- Select a sample of units from the population, make interventions on those units changing their values.
- Objective is to change the population, not just sample units, in a desired way.
- One strategy interacts with another

Effect of an intervention

- An **intervention strategy** consists of a sampling design for finding units in the population on which to make interventions, and a plan for the types of interventions to be made, which may depend on sample unit characteristics.
- A simple way to view the **effect** of a strategy is the difference in the resulting equilibrium distribution, compared with the equilibrium distribution without the strategy, or with a different strategy.

Natural sampling strategies

- virus selects a sample of people with a link-tracing design
- insects select a sample of plants with a temporal spatial distance design.

Dynamic network sampling

Dynamic network sampling

- initially and at ongoing rate units are selected using conventional or spatial design

Dynamic network sampling

- initially and at ongoing rate units are selected using conventional or spatial design
- new units are added through link tracing tracing rate may depend on unit and link values

Dynamic network sampling

- initially and at ongoing rate units are selected using conventional or spatial design
- new units are added through link tracing tracing rate may depend on unit and link values
- units are removed from sample through removal probability or deletion from population.

Random walk in static network

The classic random walk in a graph starts with an arbitrary node and at each step selects at random one of the links out from the current node to reach the next node.

Random walk in static network

The classic random walk in a graph starts with an arbitrary node and at each step selects at random one of the links out from the current node to reach the next node.

- The current sample S_t consists of that one node.

Random walk in static network

The classic random walk in a graph starts with an arbitrary node and at each step selects at random one of the links out from the current node to reach the next node.

- The current sample S_t consists of that one node.
- When sampling is with replacement, the sequence S_0, S_1, S_2 is a Markov chain with constant transition matrix.

Random walk in static network

The classic random walk in a graph starts with an arbitrary node and at each step selects at random one of the links out from the current node to reach the next node.

- The current sample S_t consists of that one node.
- When sampling is with replacement, the sequence S_0, S_1, S_2 is a Markov chain with constant transition matrix.
- Connected components of networks form closed classes.

Random walk in dynamic network

- links between nodes change over time
- component structure changes
- a random walk temporarily stuck in one component can eventually reach nodes of other components

Random walk in dynamic network

- links between nodes change over time
- component structure changes
- a random walk temporarily stuck in one component can eventually reach nodes of other components
- deletions and insertions of nodes can interrupt the walk, requiring reseeding

Random set designs

- current sample S_t is a set of nodes
- acquisition of nodes includes tracing links out from sample
- attrition through removal of nodes from sample or deletion from population

Simple random set design

- with-replacement selection
- independent link tracing
- independent removals

Design options

- selection and removal probabilities depend on node and link values
- with or without replacent
- links followed from active sample units only

Design options

- selection and removal probabilities depend on node and link values
- with or without replacent
- links followed from active sample units only
- replacement, activeness values between 0 and 1

Desired features

- trace rapidly at first; reseeding.
- have a target sample size distribution.
- find units with high degree or "interesting" values.

Epidemic example

- HIV virus spreads with dynamic network sampling design
- seek and treat designs for interventions to reduce incidence.
- combination of interventions and counter-responses leads to new equilibrium distribution.

What influences equilibrium distribution

- sample volume = number of nodes
- sample surface = number of links out

What influences equilibrium distribution

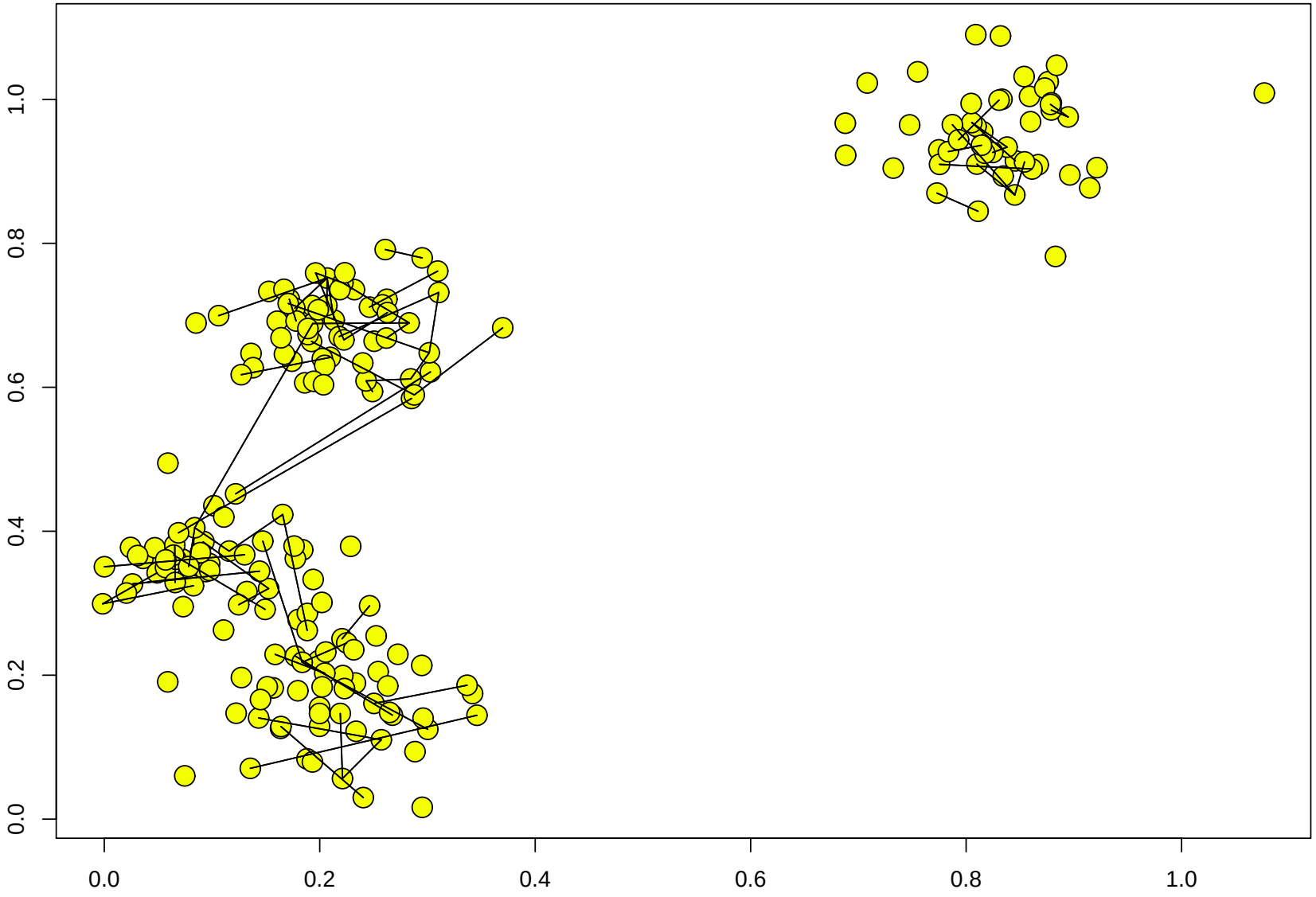
- sample volume = number of nodes
- sample surface = number of links out
- may be weighted by tracing probabilities

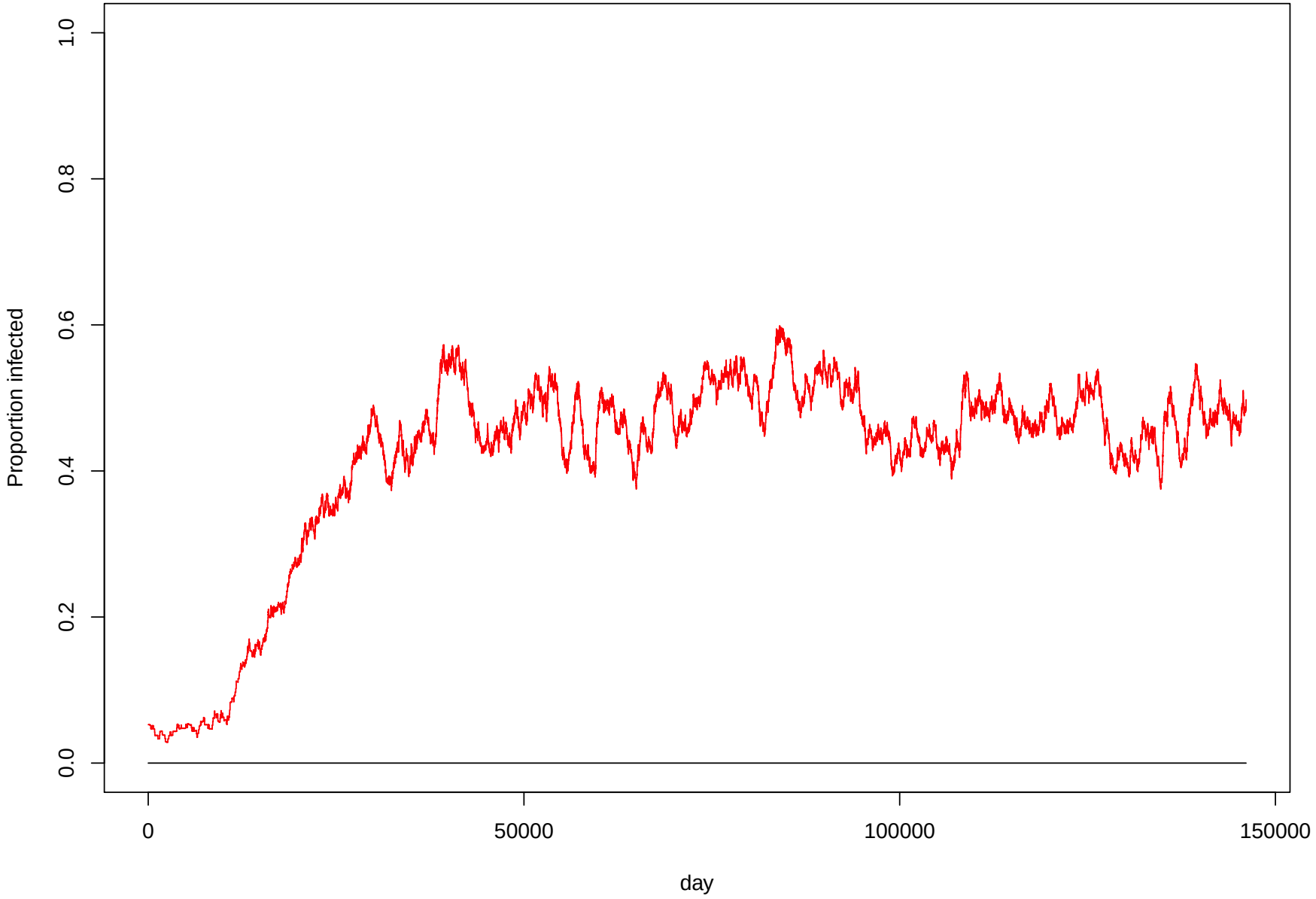
What influences equilibrium distribution

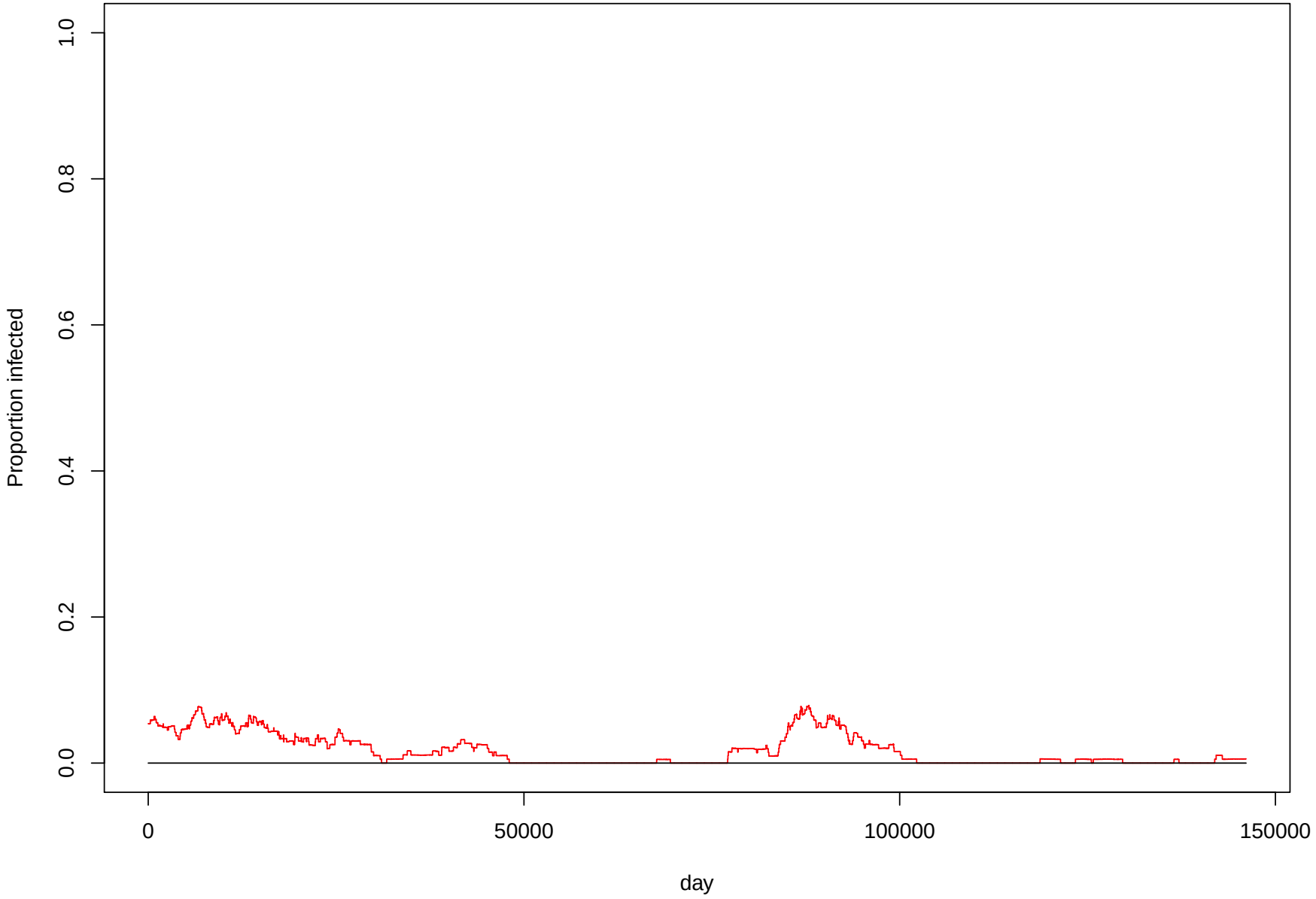
- sample volume = number of nodes
- sample surface = number of links out
- may be weighted by tracing probabilities
- surface to volume ratio tends to decrease as sample size increases

Also affecting equilibrium level

- new node entering sample (HIV) tends to have higher degree than average for population
- and higher proportion of links-out than average for sample
- especially in early stages of epidemic


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


Proportion infected
0.0
0.2
0.4
0.6
0.8
1.0
0
50000
100000
150000
day


Proportion infected
0.0
0.2
0.4
0.6
0.8
1.0
0
50000
100000
150000
day