DRAFT
EvoNetHIV User's Guide -- Software Package for Modeling HIV Epidemics and Viral Evolution in Sexual Networks

James Murphy, Steven Goodreau, Joshua Herbeck, and John Mittler

Version 1.0
November 15, 2017

University of Washington
Department of Microbiology
Department of Global Health
Center for Studies in Demography and Ecology

Preface

This manual describes a computer package (EvoNetHIV) that allows users to model the interactions between social networks, within-host dynamics, viral evolution, and HIV-1 spread. This software is intended for academic researchers in HIV epidemiology and evolutionary biology.

The software assumes familiarity with basic R commands. It is built using the API (application programming interface) of the R package *EpiModel*, so familiarity with that API and package is helpful. We provide an introduction below, as well as links to more information for advanced users. Users who wish to model new network structures will also need to be familiar with the methods and syntax from the *statnet* suite of R packages. We also provide a brief overview, and links for more information, within this Guide.

We acknowledge contributions from Neil Abernethy, Juandalyn Burke, Geoff Gottlieb, Kathryn Peebles, and And Sarah Stansfield.

This project was supported by a grant from the National Institutes of Allergy and Infectious Diseases Add grant number. (PIs Steven Goodreau, Joshua Herbeck and John Mittler) and the Center for Demography and Ecology. Findings from this research do not necessarily reflect the views of the funders.

For updates and further information about this package, please visit https://github.com/EvoNetHIV.

1. EVONET OVERVIEW

Introduction

EvoNetHIV is a stochastic, agent-based simulation model that incorporates within-host dynamics, sexual network structure, viral load variation, treatment, condom use, circumcision, and HIV evolution. *EvoNetHIV* is written as a series of interchangeable modules and user-specified parameters (written in the R programming language) that control different components of the system (e.g. HIV and CD4 progression, treatment initiation, coital acts). Network characteristics (e.g., age-related homophily, presence of risk groups) can be specified using standard statnet network commands. Advanced users have the option of writing additional modules (e.g., allowing circumcision probabilities vary with risk membership).

Evonet foundations Evonet uses a modular framework from the EpiModel API (www.epimodel). The network module is built using exponential random graph models in the statnet framework (www.statnet.org), which provides a language for specifying arbitrarily complex network models.

Experimental framework. Users may interact with Evonet in one of four principal ways to conduct their own experiments: (1) they may keep the model structure exactly as is, but vary some of the many dozen input parameters; (2) they may use some of the alternative functional forms included with the model through the use of flag arguments; (2) they may keep the model structure as is, but vary the network model and associated parameters through the use of *ergm* terms; (3) they may change the structure of the model through the writing of substitute modules. Each of these features requires understanding the current model, and being familiar with the default parameter values; these are described below. Option (3) requires familiarity with the mathematics and terminology of exponential random graph models, and option (4) requires familiarity with the EpiModel API; an introduction to each of these can be found here: http://statnet.github.io/nme/

Agent-based framework. EvoNet is a stochastic, agent based simulation model. Each agent (i.e., individual) is a discrete entity that has over 30 attributes. Important agent attributes include sex, age, infection status, time of infection, viral load, set-point viral load, and treatment status. Some attributes are regularly updated, such as age or viral load (for infected agents) while others (e.g. sex, set-point viral load), do not change once assigned to the agent. Agents are "born" (added to the model as uninfected agents), age, and depart from the model, either through death from AIDS, death from background mortality, or aging out. Agents can become infected and/or infect other agents.

Stochastic framework. EvoNet is a stochastic model, such that identical parameter values or starting conditions for different model runs will produce different results. Many input parameters are mean or variance values of probability distributions used to draw random numbers from the specified probability distribution. For example, the parameter for mean sex acts per ongoing relationship per time step is the mean value for a draw from a Poisson distribution.

Daily timesteps. Model dynamics occur on a daily timestep; thus fine-scale epidemiological and behavioral processes can be simulated.

Epidemiological and evolutionary dynamics. EvoNet models have four main inter-connected components that result in epidemiological and evolutionary dynamics: *network structure*, *sexual behavior*, *clinical phenomena*, and *SPVL and Viral progression.* All four components operate at the scale of the agent. The aggregate dynamics or values across agents give epidemiological and population level dynamics.

- ***Network structure*.** Agents can be either isolates or have one or more concurrent relationships with other agents. This network structure is dynamic and changes according to the specified network statistics.
- ***Behavior*.** Important behavioral dynamics include sex roles (for MSM), frequency of sex, condom use, HIV testing frequency, and adherence to treatment. Behavioral dynamics can be specified by the user.
    - ***MSM vs. Heterosexual models.*** MSM and heterosexual models are nearly identical. The most important differences are found in network structure. In MSM models, each agent is assigned either an "insertive", "receptive", or "versatile" role. Insertive agents can form relationships with versatile or receptive agents; receptive agents can form relationships with versatile or insertive agents; and versatile agents can form relationships with all three agent types. Heterosexual models assume a 50-50 sex ratio and only only relationships between opposite sexes are allowed.
- ***Clinical.*** *Treatment, CD4 stuff?*
- ***SPVL and Viral progression.*** Viral load progression is mostly a function of set-point viral load. SPVL for newly infected agents is determined by the donor's SPVL and by environmental (random) variation. The contribution of the donor's SPVL to the infectee's SPVL is determined by the assumed heritability of HIV SPVL.

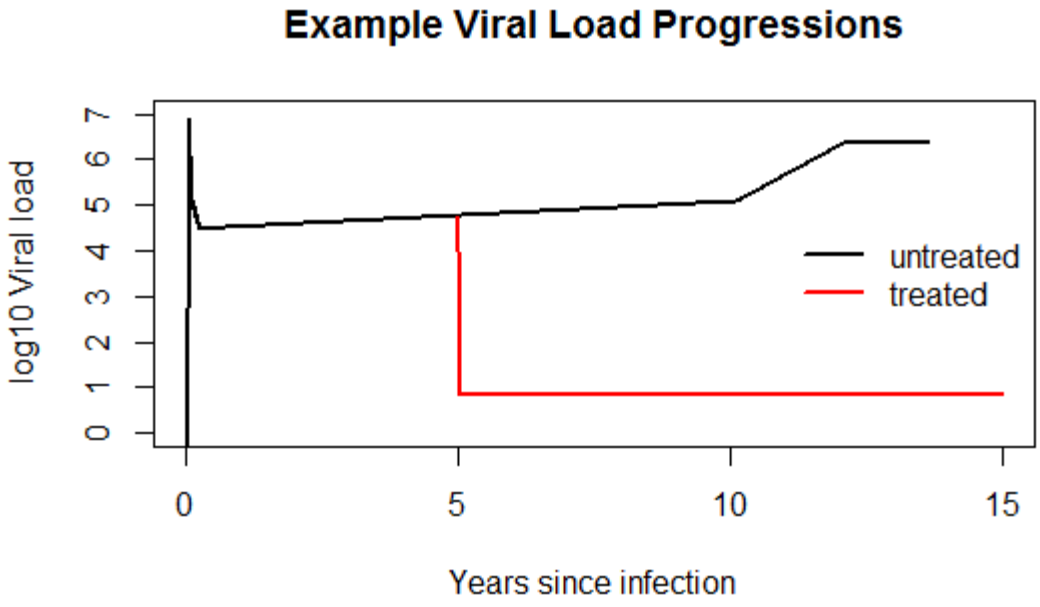


Figure 1. Example of viral load progression with and without ART. For this example, ART begins at 5 years since infection. Without ART, AIDS begins at about 10 years since infection and death at about 13 years. While acute phase duration is fixed at 90 days, start of AIDS and time of AIDS death are stochastic.

Virological, Epidemiological, and Demographic Processes

In this section we give an overview of the assumptions made by our default modules. Additional details and descriptions of alternative (nondefault) modules are given in the programming guide in section 2.

Viral Load and Setpoint Viral Load

- At infection, agents are assigned a set-point viral load (SPVL), using a function that incorporates infector SPVL, assumed heritability of SPVL, and a randomly determined host-specific factor (stochastic white noise), and heritability. The maximum possible viral load (all viral dynamics on log10 scale) is 7.0 and the minimum is 2.0.
- Initial VL starts out at -4.0 rises exponentially to a peak acute phase level, and then has a biphasic decay to SPVL. The acute phase last 50 days. Default value for peak acute phase VL is constant for all agents but can be set to be a function of an agent's SPVL. After the end of the acute phase, VL increases linearly, starting at SPVL, at a pre-specified chronic phase rate of increase. At the onset of AIDS, VL increases linearly at a pre-specified rate of increase for the AIDS phase until it reaches the assumed maximum VL value for AIDS at which it remains until agent's death. (The onset of AIDS is triggered by CD4 values described below.)
- With commencement of ART, VL will decrease linearly over 30 days to a "undetectable" level of 1.7 and remain there for the duration of treatment. If treatment stops, VL will then increase linearly over 30 days to the VL level at the start of treatment and then progress as if treatment did not occur.

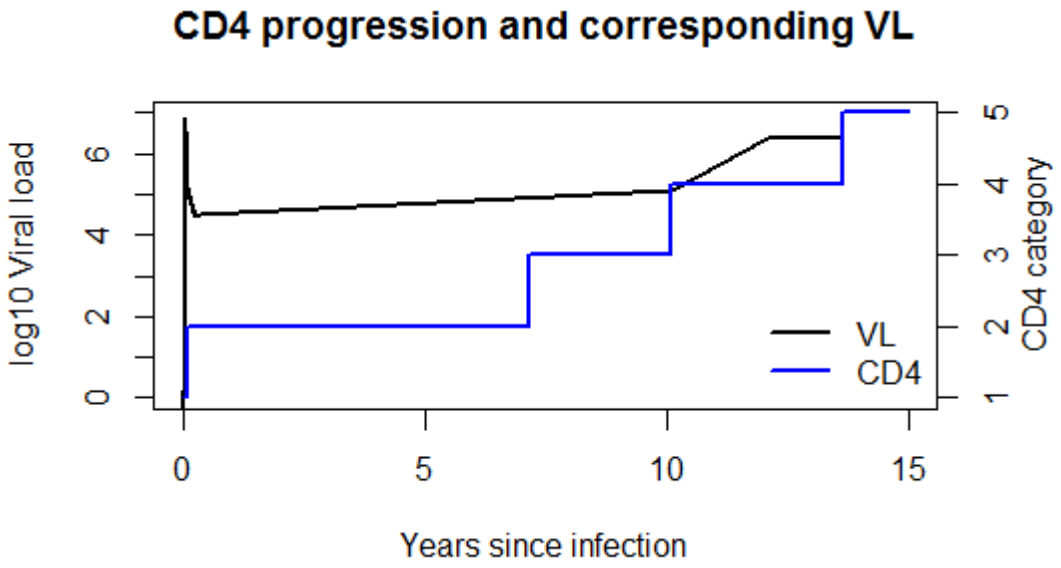


Figure 2. CD4 progression and corresponding V (no treatment).

CD4 Dynamics

- CD4 counts for an agent is a recorded as an ordinal variable with 4 categories; categories 1,2,3, and 4 represent CD4>500, 500>CD4>350, 350>CD4>200, 200>CD4>0, respectively. Category 4 represents AIDS stage.
- At infection, an agent is probabilistically assigned an initial CD4 category (1,2, or 3) where higher SPVL values increases the probability of assignment to categories 2 or 3. (Table 1). For each daily timestep, an infected agent in categories 1,2, or 3 can move to the following CD4 category (representing lower CD4 counts) or remain in the current category based on the outcome of a Bernoulli trial with the probability of success (moving to the next category) equal to the inverse of the mean passage time (in days) for a given SPVL and CD4 category (Table 2). Passage time in CD4 category 4 is fixed for each SPVL category; at the end of passage time in category 4, an agent is categorized as having died of AIDS.
- With commencement of ART, an agent's CD4 category is set to 1 and remains there until treatment stops or agent's death. If treatment stops before agent's death, the CD4 category of an agent is set equal to the value at the start of treatment.

Table 1. Probabilities of starting CD4 progression in categories 1,2, or 3.

Set point viral load (log_{10} copies/mL)	CD4 level (cells/mm^3) (category)		
	≥ 500 (1)	350 – 500 (2)	200 – 350 (3)
[2.0, 3.0]	0.88	0.12	0.00
(3.0, 3.5]	0.87	0.12	0.01
(3.5, 4.0]	0.85	0.12	0.03
(4.0, 4.5]	0.78	0.19	0.03
(4.5, 5.0]	0.73	0.21	0.05
(5.0, 5.5]	0.71	0.25	0.04
(5.5, 6.0]	0.64	0.27	0.09
(6.0, 6.5]	0.00	0.00	1.00
(6.5, 7.0]	0.00	0.00	1.00

Table 2. Mean passage times for categories 1,2, and 3. Fixed passage times for category 4.

Set point viral load (log_{10} copies/mL)	CD4 level (cells/mm^3) (category)			
	≥ 500 (1)	350 – 500 (2)	200 – 350 (3)	< 200 (4)
[2.0, 3.0]	6.08	5.01	3.60	4.67
(3.0, 3.5]	4.69	2.52	3.68	4.11
(3.5, 4.0]	3.94	4.07	2.38	3.54
(4.0, 4.5]	2.96	3.09	3.81	2.98
(4.5, 5.0]	2.25	2.32	3.21	2.42
(5.0, 5.5]	1.47	1.55	2.27	1.86
(5.5, 6.0]	0.95	1.19	1.00	1.29
(6.0, 6.5]	0.32	0.59	0.68	0.73
(6.5, 7.0]	0.30	0.46	0.37	0.17

SPVL, VL, and CD4 of initial population

- SPVL of ithe initial infected population is assumed to be normal distributed with a mean of 4.5 and variance of 0.8. SPVL values for initial population are drawn probabilistically from this distribution (user can change these values).
- Each model run begins with a specified number of infected agents. Each of these initially infected agents are randomly assigned an infection time 180-730 days before the start of the model run. Then VL and CD4 are projected forward based on infection time. Thus, each infected agent in the initial population will be in the chronic phase of HIV infection at the start of the model.

Table 3. VL and CD4 progression parameters

Description	**Value**	**Unit**
viral load at Day 1 of infection	0.0001	copies/ml
Viral Load at peak of Primary Infection	7.7 x 10^6	copies/ml
VL Maximum In AIDS	2400000	copies/ml
Daily Slope in VL Increase After AIDS Onset	1.0041122	unitless per day rate
Time To Peak VL during Acute Infection	21	Days
Time In Acute Inf	90	Days
Time until beginning of Acute Phase decline part 2	55	Days
Slope on ln scale in Acute Decline Phase 2	-0.03	NA
Slope of viral load increase during chronic phse	0.14	per year

Table 4. VL progression with treatment parameters.

Description	**Value**	**Unit**
Final VL After Complete Treatment	13	copies/ml
VL Undetectable Level	50	copies/ml
VL Exponential Decline Rate With Treatment	-0.6	per day

Table 5. SPVL and heritability parameters

Description	Value	Unit
Average Log SPVL for initial population	4.5	log10 copies/ml
Variance Log SPVL for initial population	0.8	NA
Mutation Variance	0.01	(log10 copies/ml)^2 / per year
donor-recipient heritability (h^2)	0.36	NA

Condom Use and Circumcision

- ***Condom use*:** Each sex act has a specified probability of condom use. Default is 0.50 but can be easily modified by user. An optional parameterization models condom use as a function of age and is described in the Treatment as Prevention Case Study in Section 2.xx.
- ***Circumcision:*** Each male agent has specified probability of having undergone circumcision. Default is 0.85 but can be easily modified by user.

Testing, Diagnosis, Treatment

- ***Testing and Diagnosis:*** Default testing interval is 1 year for both sexes, and sex specific intervals can be specified. All HIV+ agents that are tested receive a positive diagnosis, which is required for undergoing treatment.
- ***Treatment:*** Treatment campaigns begin at user specified start times (e.g., 5 years after start of model run). Once treatment campaign begins, all diagnosed agents are put on treatment, though this proportion can be modified. Agents in acute infections are assumed not to undergo treatment but this constraint can also be removed. Further constraints can be placed on treatment, such as requiring a minimum viral load or time since infection.

Coital Acts and Transmission

Sexual acts are determined among agents in a serodiscordant relationship at each daily time step. Among these partnerships, the number of sexual acts per partnership per daily time step is assigned according to a Poisson draw (mean = 0.43). Probability of sex can also be a declining function of age. Default condom use is determined for each sexual act with probability of 50% and default circumcision status is assigned to agents at model entry with 85% probability. When the infected partner is 47% past their expected time in AIDS, no sex occurs.

- For MSM models, agents are assigned one of three sexual roles: versatile, insertive, and receptive. Insertive agents only have sex with receptive agents and vice-versa. Versatile agents can have sex with any other agents. Intra-event versatility (when men switch roles and each engage in insertive and receptive anal intercourse (AI) during the same sexual contact) occurs with 40% probability when both partners are role versatile.

Transmission

HIV transmission probabilities are calculated as a function of relevant risk factors for each sex act according to characteristics of the sexual act and characteristics of the agents engaged in the sexual act. Following Hughes et al., the probability of transmission is calculated for each sexual act that occurs in a serodiscordant relationship, as:

$$P(transmission) = 1 - (1 - \lambda)^{e^{X\beta}}$$

where λ is the baseline per-act infectivity (0.000247).

Heterosexual transmission

For heterosexual partners,

$$X\beta = \ln(2.89) * (viral\ load - 4.0) + \ln(RR_{circumcised}) * circumcised + \ln(RR_{condom}) * condom + \ln(RR_{age}) * age_{base} + \ln(RR_{sti}) * sti\_status\_sus$$

where *circumcised* is an indicator variable for the circumcision status of the male, *condom* is an indicator variable for condom use for the male, *viral load* is the log10 viral load of infected agent, and *age_base* is the number of decades over 35 the susceptible agent is: (age-35)/10; and RR_x is the risk factor.

MSM Transmission

For MSM partners,

$$X\beta = \ln(RR_{VL}) * (viral\ load - 4.0) + \ln(RR_{insertive}) * insertive + \ln(RR_{receptive}) * receptive + \ln((RR_{circumcised})) * circumcised + \ln(RR_{condom}) * condom + \ln(RR_{sti}) * sti\_status\_sus$$

where *insertive* and *receptive* are indicator variables for role of susceptible agent and other variables are as for heterosexual partners.

Once the probability of transmission is calculated, a random uniform draw determines whether infection occurred.

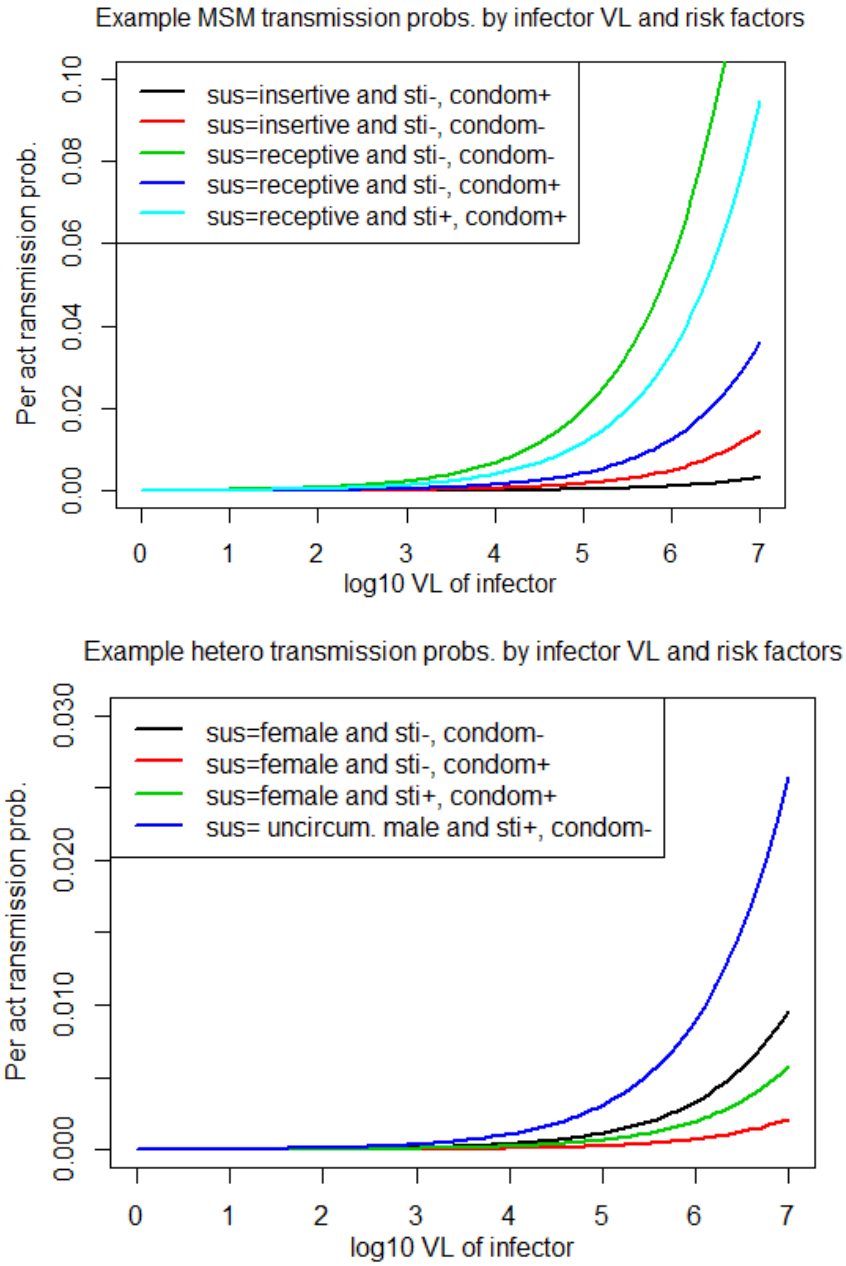


Figure 3. Example of transmission probabities as function of risk factors for MSM (top) and heterosexual partners (bottom).

Demographics

- **Aging:** age of each agent is updated each daily timestep. Default minimum age is 18 and default maximum age is 55.
- **Aging out:** agent is removed from population when maximum age is reached.
- **New agents ("births"):** New agents are added to the population assuming a 1% growth rate (in absence of an epidemic); for heterosexual model, new agents have a 50/50 sex ratio. Default age of new agents is 18.
- **Deaths:** Infected agents can die of
    - AIDS – when their passage time through CD4 category 4 is completed;
    - Natural mortality, independent of HIV – default sex and age specific natural mortality values are from recent U.S. census data
    - Natural mortality, HIV induced – infected agents have a heightened risk of non-AIDS death
- **Initial age distribution:** Default sex-specific initial age distribution is based on U.S. census data

Network Structure and Dynamics

- **Structure:** User specified network structure based on the statnet and EpiModel packages can be implemented (e.g., mean degree, age and other subgroup structure). Default structure is random mixing of agents with a network mean degree of 0.7.
- **Partnerships:** Each agent can exist as an isolate (no-partners) or have one or more sexual partners per timestep. Default mean partnership duration is 100 days.
- **Risk groups:** Various risk group substructures can be imposed.

2. USING EVONET WITH CASE STUDIES

Installing and running EvoNet (simple example)

Download/install/load EvoNet

```
if (!require("devtools")) install.packages("devtools")
install_github("EvoNetHIV/TestRepo",subdir="pkg")
library(evonet)
```

Load default parameters.

```
primary_parameters  <- input_parameters_primary()
cd4_data            <- input_parameters_cd4_data()
```

Combine individual parameters into single list. Parameter list can be viewed by entering 'evoparams' in RStudio console.

```
evoparams <- c(primary_parameters, cd4_data)
```

Change default parameters. In this example, we change initial population size to 200 (from 100); number of initially infected agents to 40 (from 20); and the duration of the model to 20 years.

```
evoparams$initial_pop      = 200
evoparams$initial_infected = 40
evoparams$n_steps = 365*20
```

Calculate derived parameters (parameters that are functions of other parameters)

```
evoparams  <- input_parameters_derived(evoparams)
```

Convert initial parameter list into EpiModel parameter list (so EpiModel recognizes it as a parameter list)

```
evoparams <- do.call(EpiModel::param.net,evoparams)
```

Check to make sure input parameters are valid (error returned if not)

```
input_parameters_qaqc(evoparams)
```

Create initial network (as a function of input parameters)

```
nw <- setup_initialize_network(evoparams)
```

Create list of arguments for EpiModel's network estimation function.

```
netest_arg_list <- list(
  nw           =  nw,
  formation    =  as.formula(evoparams$nw_form_terms),
  target.stats =  evoparams$target_stats,
  coef.form    =  evoparams$nw_coef_form,
  constraints  =  as.formula(evoparams$nw_constraints),
  verbose      =  FALSE,
  coef.diss    =  dissolution_coefs( dissolution =  as.formula(evoparams$dissolution),
                                     duration    =  evoparams$relation_dur,
                                     d.rate      =  3e-05) )
```

Estimate network (i.e., create desired network structure and dynamics)

```
estimated_nw <- do.call(EpiModel::netest, netest_arg_list)
```

Create vector of infection status (0/1) as an epimodel object for initial population

```
infected_list <- EpiModel::init.net(i.num=evoparams$initial_infected,
                                    status.rand = FALSE)
```

Create list with modules/functions to simulate epidemic with desired dynamics.

```
evo_module_list<- list(
  "initialize.FUN"     = initialize_module,
  "plot_nw.FUN"        = plot_network_fxn,
  "aging.FUN"          = vital_aging_module,
  "testing.FUN"        = social_testing_diagnosis_module,
  "treatment.FUN"      = social_treatment_module,
  "update_vl.FUN"      = viral_update_gamma,
  "update_cd4.FUN"     = viral_update_cd4_daily,
  "coital_acts.FUN"    = social_coital_acts_module,
  "trans.FUN"          = transmission_main_module,
  "trans_book.FUN"     = transmission_bookkeeping_module,
  "trans_cd4.FUN"      = transmission_cd4_module,
  "deaths.FUN"         = vital_deaths_module,
  "births.FUN"         = vital_births_module,
  "summary.FUN"        = summary_module,
  "resim_nets.FUN"     = EpiModel::resim_nets,
  "verbose.FUN"        = NULL)
```

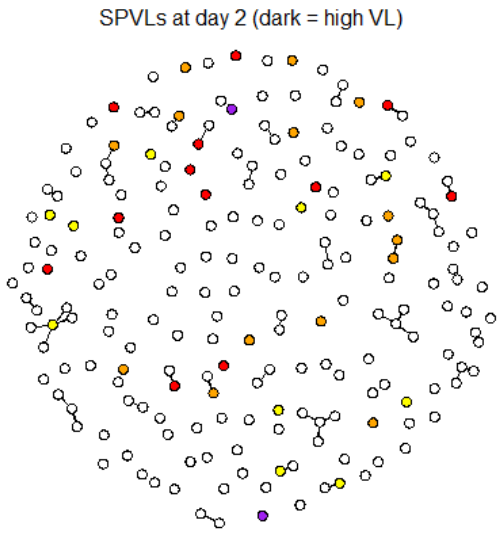
Create an EpiModel "control" object which contains both the input parameters and the modules

```
evocontrol <- setup_epimodel_control_object(evonet_params = evoparams,
                                            module_list   = evo_module_list)
```

Run the simulation using EpiMode's 'netsim' function

```
 evomodel  <- EpiModel::netsim(x = estimated_nw,
                               param = evoparams,
                               init = infected_list,
                               control = evocontrol)
```

Initial network plot will be plotted by default at simulation start (colored circles represent infected agents). (Day 2 is first day of simulation as model initialization is considered day 1.)



Save model output. Default location is current working directory.

```
save(evomodel,
    file = file.path(evoparams$output_path,"evomodel.RData"))
```

Create default output plots summarizing model run. Plots printed to screen and saved as pdf file (default name: "popsumm_figures.pdf") to working directory.

```
plots_popsumm(evomodel,outpath=evoparams$output_path,
          name=NULL,nw_stats=TRUE,max_points_rep=100,
          evoparams$popsumm_frequency)
```

Example of two default output plots