



# **Epigrass Documentation**

***Release 2.6.3***

**Flavio Codeco Coelho**

**February 21, 2021**



# CONTENTS

<b>1</b>	<b>Overview of Epigrass</b>	<b>3</b>
1.1	Modeling Approach . . . . .	3
1.2	Geographical Network Models . . . . .	3
1.3	The Simulation . . . . .	5
<b>2</b>	<b>Building and Installing</b>	<b>7</b>
2.1	Required Packages . . . . .	7
2.2	Installing Epigrass . . . . .	8
<b>3</b>	<b>Epidemic Models</b>	<b>9</b>
3.1	Introduction . . . . .	9
3.2	Disease models . . . . .	9
3.3	Network transportation models . . . . .	15
<b>4</b>	<b>Using Epigrass</b>	<b>17</b>
4.1	Data . . . . .	17
4.2	Specifying a model: the script . . . . .	19
<b>5</b>	<b>Analysis</b>	<b>27</b>
<b>6</b>	<b>Writing Custom Models</b>	<b>29</b>
6.1	Getting Started . . . . .	29
6.2	The Environment . . . . .	31
	<b>Index</b>	<b>35</b>



Epigrass is a framework for the construction and simulation of geo-referenced network epidemiology models. At least this is its main vocation. However, Epigrass can be used to simulate any population dynamical system on a set of nodes (connected into a network or not!). Epigrass is written in pure Python and is scriptable in Python, for your enjoyment!.

Contents:



## OVERVIEW OF EPIGRASS

Epigrass is a platform for geo network epidemiological simulation and analysis. It enables researchers to perform comprehensive spatio-temporal simulations incorporating epidemiological data and models for disease transmission and control in order to create sophisticated scenario analyses.

### 1.1 Modeling Approach

The geographical networks over which epidemiological processes take place can be very straightforwardly represented in a object-oriented framework. In such a framework, the nodes and edges of the geographical networks are objects with their attributes and methods.

Once the archetypal node and edge objects are defined with appropriate attributes and methods, then a code representation of the real system can be constructed, where cities (or other geographical localities) and transportation routes are instances of the node and edge objects, respectively. The whole network is also an object with a whole collection of attributes and methods.

This framework leads to a compact and hierarchical computational model consisting of a network object containing a variable number of node and edge objects. This framework also do not pose limitations to encapsulation, potentially allowing for networks within networks if desirable.

For the end user perspective, this framework is transparent since it mimics the natural structure of the real system. Even after the model is converted into a code object all of its component objects remain accessible to the user's models.

### 1.2 Geographical Network Models

Epigrass's geo-referenced models are built from two basic sources of data: a map (in shapefile or geopackage format) the cartographical base over which the models are represented and specific data about nodes and edges that are provided by the user for the network of interest.

#### 1.2.1 Defining the Cartographic Background

If the user has a map for the geo-referenced data, this can be passed to Epigrass. In this case, the cartographic background is defined by defining the name of the shapefile file (with path relative to the working directory) in the model'.epg file. Along with the path to the shapefile, the variable in the shapefile, which contains the geocode of localities and their name must also be specified:

```
shapefile = ['riozonas_LatLong.shp', 'nome_zonas', 'zona_trafe']
```

If the user does not have a map for his/her problem, he/she can still use Epigrass. In this case, the network topology data is read only from two .csv files (more on that ahead).

## 1.2.2 Defining Nodes

A graph has nodes and edges. Nodes can be cities or other localities depending on the model being constructed. The definition of nodes require the setting of many attributes listed below. The nodes will have many more attributes defined at run-time which will depend on other aspects of the model, these will be discussed later.

The data necessary at build time to create nodes come from a CSV (comma-separated-values) ASCII-text file, with the following columns, (in this order):

**Latitude, Longitude** This attribute will be used to geo reference the node. The coordinate system must match those used in the

cartographic base imported from GRASS. Coordinates can be coded in either decimal or sexagesimal format. *Name*

Used for identification purposes only. It can be a city name, for instance.

**Population** Since the simulation models will all be of a populational nature. Population size must be specified at build time.

**Geocode** A Unique Geocode (a number) is required. It will be used as a label/index to facilitate reference to specific nodes.

## 1.2.3 Defining Edges

In Epigrass' graphs, edges represent transportation routes. Similarly to nodes, edges are defined at build-time with a reduced set of attributes which will be extended at run-time. Epigrass also expects to get these attributes from a CSV file:

**Source** The name of the source node. The edges are bi-directional, but the nodes are labeled source and destination for

reference purposes. *Destination*

The name of the destination node.

**Forward migration** Migration rate from source to destination, in number of persons per unit of time.

**Backward migration** Migration rate from destination to source, in number of persons per unit of time.

**Length** Distance in kilometers (or another unit) from source to destination via the particular route (not straight line distance).

**Source's geocode** This is the unique numerical identifier used in the sites file.

**Destination's geocode** This is the unique numerical identifier used in the sites file.

## 1.2.4 Defining models

The word model in Epigrass can mean two distinct objects: The network model and the node's epidemic model.

Node objects, in an Epigrass model, contain well-mixed population dynamic models within them. These models determine the dynamics of epidemics within the particular environments of each node. Epigrass comes with a few standard epidemiological models `index{Models!epidemiological models}` to choose from when setting up your network. Currently, The same model type is applied to every node although their parameterization is node-specific. Besides the built-in model types, users can define their own, as shown in the chapter *Using Epigrass*.

Network models are specified in a ASCII-text script file (Called a `.epg` file). Epigrass comes with a few demo Network models for the user to play with until he/she is confident enough to build their own. Even then, it is advisable to use the demo scripts provided as templates to minimize syntax errors.



The script on the appendix specifies a network model with an stochastic SEIR (see chapter on epidemiological modeling) epidemic model in its nodes. The user should study this model and play with its parameters to understand the features of Epigrass. A step-by-step tutorial on how to edit the model script can be found in the chapter *Using Epigrass*.

## 1.3 The Simulation

A simulation run in Epigrass consists of a series of tasks performed at each time step<sup>1</sup>.

**Calculate migration** For all edges in the network, the number of persons traveling each way is determined for the current time-step.

**Run epidemic models** For each node in the network the epidemic demographics are updated based on the local number of infected and susceptible individuals which have been updated by the transportation system.

All aspects of the simulation such as number of passengers traveling on each edge, number of infected/susceptible on each node and etc., are recorded in a step-by-step basis. This complete record allows for the model to be analyzed after the simulation has been completed without having to recalculate it.

### 1.3.1 Output

The output of a simulation in Epigrass is three-fold: A web-based graphical dashboard to interact with outcome of the simulation, a written report, and a database table with numeric results.

#### Graphical display

Epigrass's Dashboard can be set to open immediately after the simulation has been completed:

```
$ epirunner -D mymodel.epg
```

or it can be invoked to explore simulations of a model already stored in the database, without simulating it first:

```
$ epirunner -V mymodel
```

#### Report Generation

The report contains a detailed analysis of the network model and the simulations ran with it. The report generates a LaTeX source file and compiles it to a PDF document for visualization.

Three types of report are currently available:

**Report = 1** Returns a set of descriptors of the network, described in chapter

**Report = 2** Returns a set of basic epidemiological measures and plots of the time series.

**Report = 3** Report 1 + Report 2

Report Generation is an optional, though recommended, step done at the end of the simulation. For the report, descriptive statistics are generated for the network. These have to do with network topology and properties. Additional sections can be added to the report with basic statistical analyses of the output of pre-selected nodes<sup>2</sup>.

<sup>1</sup> The number of time steps is defined in the model script

<sup>2</sup> Listed in the siteRep variable at the script

### Database output

Time series of **L**, **S**, **E**, and **I**, from simulations, are stored in a SQLite database. The results of each individual simulation is stored in a different table named after the model's script name, the date and time the simulation has been run. For instance, suppose you run a simulation of a model stored in a file named `script.epg`, then at the end of the simulation, a new table in the epigrass database will be created with the following name: `script_Wed_Jan_26_154411_2005`. Thus, the results of multiple runs from the same model get stored independently.

Epigrass also supports the SQLite database and CSV files as output for the time-series. The naming convention also applies to these other formats.

## BUILDING AND INSTALLING

This chapter will walk through all aspects of Epigrass installation. From obtaining, building and installing the prerequisites to the installation of Epigrass itself.

Most of the steps will be quite simple and similar since they will make use of standard tools for package installation on Debian GNU/Linux and derivatives. If you use a different distribution, you should check its documentation for package installation instructions.

If, on your distribution, a package is not available for the required version, you can try to obtain an updated version of the package at the web-sites provided. On the rare cases where pre-built packages are not available, instructions on how to build the software from source should also be available from its web-site.

### 2.1 Required Packages

#### 2.1.1 Python

**Web-site:** <http://www.python.org>

**Version required:**  $\geq 3.6$

Python is a simple yet powerful object-oriented language. Its simplicity makes it easy to learn, but its power means that large and complex applications can also be created easily. Its interpreted nature means that Python programmers are very productive because there is no edit/compile/link/run development cycle.

Python is probably installed automatically by your GNU/Linux distribution (it is on Debian). If not, it is best to use your distribution's standard tools for package installation. On Debian for example:

```
$ sudo apt-get install python build-essentials python3-pip
```

#### 2.1.2 Redis database

**Web-site:** <https://redis.io>

**Version required:**  $\geq 6.0$

Redis is an in-memory database that Epigrass uses as a message broker for parallel execution. Example installation:

```
$ sudo apt-get install redis-server
```

### 2.1.3 MySQL

**Web-site:** <http://www.mysql.com>

**Version required:** >5.0

MySQL is a fast, multi-threaded, multi-user SQL database server. If you have a MySQL server available in your LAN, you may skip this step after making sure you have permission to access and use it to store your data.

Example installation:

```
$ sudo apt-get install mysql-server mysql-client python-mysqldb
```

### 2.1.4 Redis

**web-site:** <http://redis.io>

**Version required:** >=2.8

Redis is a persistent Key-value database.

Example installation:

```
$ sudo apt-get install redis-server
```

### 2.1.5 Other Requirements

Some other requirements can be installed directly from the Python Package index using the *pip* tool:

```
$ sudo pip install cython networkx redis sqlalchemy requests sphinx osgeo ogr
$ sudo apt install libgdal-dev
```

#### Post-install configuration

After installing MySQL, you will need to create a new database in the server, called *epigrass* and a user with all privileges to access and modify it. This is the user Epigrass will use to interact with MySQL.

## 2.2 Installing Epigrass

If you got through all the steps above, it will be an easy task to install Epigrass. There is a *.deb* package for installing Epigrass on Debian and Ubuntu. However, since it is not maintained by the developers of Epigrass, It may very well be outdated. So we recommend that you install the latest version from PyPI, by typing:

```
$ sudo pip3 install -U epigrass
```

If the installation proceeds without errors, you will have three new executables available on your system:

***epirunner*** This is a command line version of Epigrass. With it you can run models without invoking the GUI. It's great for batch simulations and for remote use. for a quick help, try “*epirunner -h*”.

## EPIDEMIC MODELS

### 3.1 Introduction

Epidemiological models are implemented in the EpiGrass environment as a tool to predict, understand and develop strategies to control the spread of infectious diseases at different time/spatial scales. In this context, there are two direct potential applications. One is to model the spread of new diseases through an entirely susceptible population (ecological invasion). The velocity of spread of new diseases in a network of susceptible populations depends on their spatial distribution, size, susceptibility and patterns of contact. In a spatial scale, climate and environment may also impact the dynamics of geographical spread as it introduces temporal and spatial heterogeneity. Understanding and predicting the direction and velocity of an invasion wave is key for emergency preparedness.

Besides invasion, network epidemiological models can also be used to understand patterns of geographical spread of endemic diseases. Many infectious diseases can only be maintained in a endemic state in cities with population size above a threshold, or under appropriate environmental conditions (climate, availability of a reservoir, vectors, etc). The variables and the magnitudes associated with endemicity threshold depends on the natural history of the disease cite{MJKeelingandBTGrenfell1997}. These magnitudes may vary from place to place as it depends on the contact structure of the individuals. Predicting which cities are sources for the endemicity and understanding the path of recurrent traveling waves may help us to design optimal surveillance and control strategies. For measles, for example, models suggest that geographical spread tend to follow a hierarchical pattern, starting in big cities (core) and then spreading to smaller neighborhood cities (satellites) cite{BTGrenfellandONBjoernstadandJKaprey2001}.

EpiGrass implements a series of epidemiological models which, integrated with a set of analytical and visualization tools, may give us clues about the overall pattern of diseases spread in the geographical space. Besides, it can be used for scenario analysis to compare alternative intervention scenarios.

### 3.2 Disease models

Models of disease dynamics are quite diverse, ranging from caricatures to very detailed simulations. Traditional models of spread of diseases are based on the mean field assumption, i.e., that individuals interact randomly at a certain rate. Important references to the subject are cite{DiekmannOandHeesterbeekJAP2000,DaleyDJandGaniJandCanningsC2001,IshamVandMedleyG1996,AndersonRMandMayRMan}. These models are expressed mathematically as difference equations (discrete time) or differential equations (continuous time). In the simplest form, these models do not take into consideration either individual heterogeneity or the local nature of transmission events. Increased realism is achieved by structuring the population according to age, risk behavior, sex, susceptibility, or other category associated with different risk of getting or transmitting the disease. Within each sub-population, however, the assumption of well mixing must hold. When other species are involved in the transmission process (non-human hosts and vectors), these are also considered as compartments that may be sub-divided as well according to covariates associated with the risk of acquiring or transmitting the disease.

In this context, epidemiological models take the form of multi-compartmental models where each compartment is a well-mixed homogeneous population. The model describes the transition of the individuals in this population through

a sequence of disease-related stages. These stages could be *Susceptible*, *Infected*, *Recovered*, for example. And the transitions could be .. math:: Susceptible \rightarrow Infected .. math:: Infected \rightarrow Recovered

If only these transitions are allowed, then individuals in the recovered class never become susceptible again (lifelong immunity). If, on the other hand, immunity is only temporary (as in pertussis), then another transition should be included: .. math:: Susceptible \rightarrow Infected .. math:: Infected \rightarrow Recovered .. math:: Recovered \rightarrow Susceptible

One way to visualize these models is using state-flow diagrams, where boxes represent states (compartments) and arrows indicate the transitions. It is the state identity together with the transitions allowed that define the type of model in use.

### 3.2.1 Dispersal of infected individuals

A key feature of EpiGrass is to allow the simulation of geographical spread of infection from one locality to others. Dispersal of infection is modeled as in a ‘forest fire’ model cite{BTGrenfellandONBjoernstadandJKaprey2001}. An infected individual, traveling to another city, acts as a spark that may trigger an epidemic in the new locality. This approach is based on the assumption that individuals commute between localities and contribute temporarily to the number of infected in the new locality, but not to its demography. Further details about this approach can be found in Grenfell et al (2001).

In all models presented below, new infections in locality  $i$  arise from the contact between Susceptibles in  $i$  and Infectious individuals. Infectious individuals are of two types: those residents in  $i$  ( $I_t$ ) and those visiting  $i$  ( $\theta$ ). Mathematically, this is written as:

$$L_{t+1} = \beta S_t \frac{(I_t + \theta)^\alpha}{N_t + n_t}$$

where  $L_{t+1}$  is the number of new cases,  $\beta$  is the contact rate between Susceptibles and Infectious individuals,  $S_t$  is the number of susceptibles,  $I_t$  is the number of infectious individuals resident in the locality,  $\theta$  is the number of infectious individuals visiting the locality,  $N_t$  is the population residing in the locality and  $n_t$  is the total number of individuals visiting the locality.  $\alpha$  is a mixing parameter.  $\alpha = 1$  corresponds to homogeneous mixing (Finkenstadt and Grenfell, 2000).

### 3.2.2 Typology of infectious diseases and corresponding models

Here we present a brief description of the typology of infectious diseases models based on the main route of transmission, and type of immunity resulting from infection. These models correspond to the types of models built into EpiGrass.

#### SIR-like models

The natural history of many directly transmitted infectious diseases can be appropriately described by a SIR-like model. *SIR* stands for Susceptible ( $S$ ), Infected ( $I$ ) and Recovered ( $R$ ). Archetypal *SIRs* are measles or chickenpox, i.e., diseases that confer lifelong immunity (but see cite{KGlassandBTGrenfell2004}). An individual starts his life in the  $S$  state and may progress to the  $I$  state. The rate of progression of individuals from  $S$  to  $I$  is called the incidence rate or force of infection ( $\lambda$ ) which is a function of contact rate, probability of transmission per contact and density of infectious individuals. Individuals stay in the infectious period for a certain time and then move to the recovered state where they become immune to new infections. Generally, the removal rate from the infectious class is the inverse of the infectious period (i.e., it is assumed that the duration of infection is exponentially distributed).

Variations of this model allow cases where infected individuals do not acquire immunity after infection, thus returning to the susceptible pool (*SIS* model). Another variation is the inclusion of a latent stage to hold individuals that

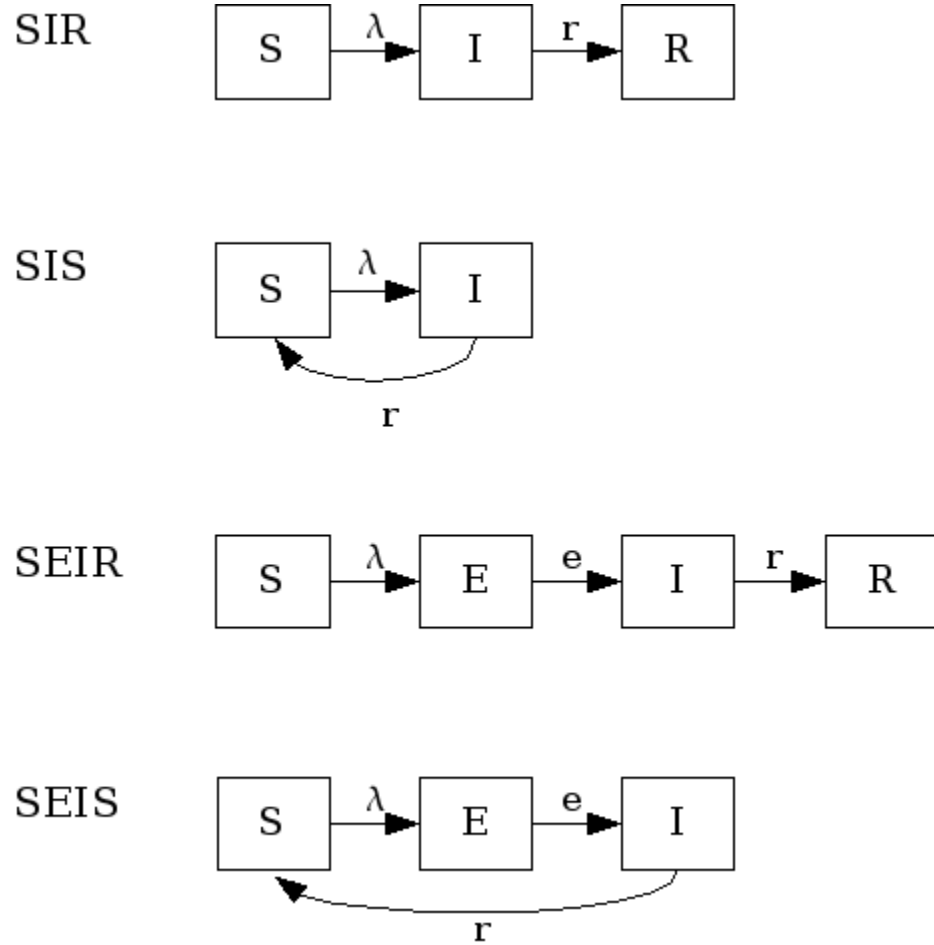


Fig. 1: Figure: SIR-like models

are infected but not infectious to others yet (incubation period). These are the *SEIR* (with immunity) and *SEIS* (no immunity) models.

Next, we describe in more detail each one of these models in their deterministic and stochastic versions, as used by EpiGrass.

Symbol	Meaning
$L_t$	number of newly infected individuals at time
$E$	number of exposed but not infectious individuals at time t
$I$	number of infectious individuals at time t
$R$	number of recovered individuals at time t
$\beta$	contact rate ( $t^{-1}$ )
$\theta$	number of infectious visitors
$\alpha$	mixing parameter ( $\alpha = 1$ means homogeneous mixing)
$n$	number of visitors
$N$   population ( $S + E + I + R$ )	
$B$   susceptible pool replenishment	
$r$	fraction of $I$ recovering from infection per unit of time ( $[0, 1]$ )
$e$	fraction of $E$ becoming infectious per unit of time ( $[0, 1]$ )
$\delta$	probability of acquiring immunity ( $[0, 1]$ )
$w$	probability of losing immunity ( $[0, 1]$ )
$p$	probability of recovered individual acquiring infection, given exposure ( $[0, 1]$ )

**SIR models** Examples of diseases represented by SIR models are measles, chickenpox. Some diseases that do not confer lifelong immunity may be represented by this model if only short term dynamics is of interest. In the scale of a year, influenza and pertussis, for example, could be described using SIR. The SIR model is implemented in EpiGrass as a system of four difference equations. Besides the three equations describing the dynamics of  $S$ ,  $I$  and  $R$ , a fourth equation explicitly defines the number of new cases per time step,  $L(t)$  (i.e., the incidence). In general, this quantity is embedded in the  $I$  equation (prevalence), but it is important to keep track of the incidence if one wishes to compare prediction with notification data.

$$\begin{aligned}
 L_{t+1} &= \beta S_t \frac{(I_t + \theta)^\alpha}{N_t + n_t} \\
 I_{t+1} &= L_{t+1} + (1 - r)I_t \\
 S_{t+1} &= S_t + B - L_{t+1} \\
 R_{t+1} &= N_t - (S_{t+1} + I_{t+1})
 \end{aligned}$$

This model can be easily extended to include diseases without recovery, for example AIDS, the so called SI models. Basically, the recovery rate is set to 0.

**SIS models** In the SIS model, individuals do not acquire immunity after the infection. They return directly to the susceptible class.

The only difference between  $\text{\$SIS\$}$  and  $\text{\$SIR\$}$  models is the absence of  $\text{\$R\$}$  and the flow of recovered individuals to the susceptible stage:

$$\begin{aligned}
 L_{t+1} &= \beta S_t \frac{(I_t + \theta)^\alpha}{N_t + n_t} \\
 I_{t+1} &= L_{t+1} + (1 - r)I_t \\
 S_{t+1} &= S_t + B - L_{t+1} + rI_{t+1}
 \end{aligned}$$



**SEIR models** These models have an extra compartment for those individuals who have acquired the infection but are still not infectious to others. This is the latent period and it is often parameterized as the inverse of the incubation period. Note, however, that for many diseases, initiation of infectiousness does not necessarily coincide with symptoms. In principle, any disease described by the SIR model can also be described by the SEIR model. The decision regarding the use of one or another depends on the magnitude of the latent period in relation to the time frame of other events in the simulation. The model has the form:

$$\begin{aligned} L_{t+1} &= \beta S_t \frac{(I_t + \theta)^\alpha}{N_t + n_t} \\ E_{t+1} &= (1 - e)E_t + L_{t+1} \\ I_{t+1} &= eE_t + (1 - r)I_t \\ S_{t+1} &= S_t + B - L_{t+1} \\ R_{t+1} &= N_t - (S_{t+1} + I_{t+1} + E_{t+1}) \end{aligned}$$

**SEIS models** These are SIS models with the inclusion of the latent stage.

$$\begin{aligned} L_{t+1} &= \beta S_t \frac{(I_t + \theta)^\alpha}{N_t + n_t} \\ E_{t+1} &= (1 - e)E_t + L_{t+1} \\ I_{t+1} &= eE_t + (1 - r)I_t \\ S_{t+1} &= S_t + B - L_{t+1} + rI_t \end{aligned}$$

### SIpR-like models

These are *SIR* models with immunity intermediary between full (*SIR*) and null (*SIS*). Some possibilities arise here: 1) Infection confers full immunity to a fraction of the individuals and the remaining ones return to the susceptible class again, after infection. (*SIpRpS*); 2) Infection provides only partial immunity and recovered individuals are partially susceptible to new infection (*SIpR*); 3) Immunity is full right after infection but wanes with time (*SIRS*). Each model is presented below. Figure fig:sipr illustrates them diagrammatically.

Related models, that included the latent state  $E$  are:  $SEIpRpS$ ,  $SEIpR$ ,  $SEIRS$ .

**SIpRpS model** This model assumes that a fraction  $\delta$  of infectious individuals acquire full immunity while the remaining  $(1-\delta)$  returns to the susceptible stage. The model is:

$$\begin{aligned} L_{t+1} &= \beta S_t \frac{(I_t + \theta)^\alpha}{N_t + n_t} \\ I_{t+1} &= L_{t+1} + (1 - r)I_t \\ S_{t+1} &= S_t + B - L_{t+1} + (1 - \delta)rI_t \\ R_{t+1} &= N_t - (S_{t+1} + I_{t+1}) \end{aligned}$$

**SIpR model** This model assumes that immunity is only partial and recovered individuals may acquire infection again (at a lower rate  $p\lambda$ , where  $0 \leq p \leq 1$ ). Two equations calculate the number of new infections.  $L_S$  calculates the number of susceptibles that become infected at  $t + 1$ .  $L_R$  calculates the number of recovered that become infected at  $t + 1$ . The latter are less susceptible to the disease when compared to susceptibles. The model is:

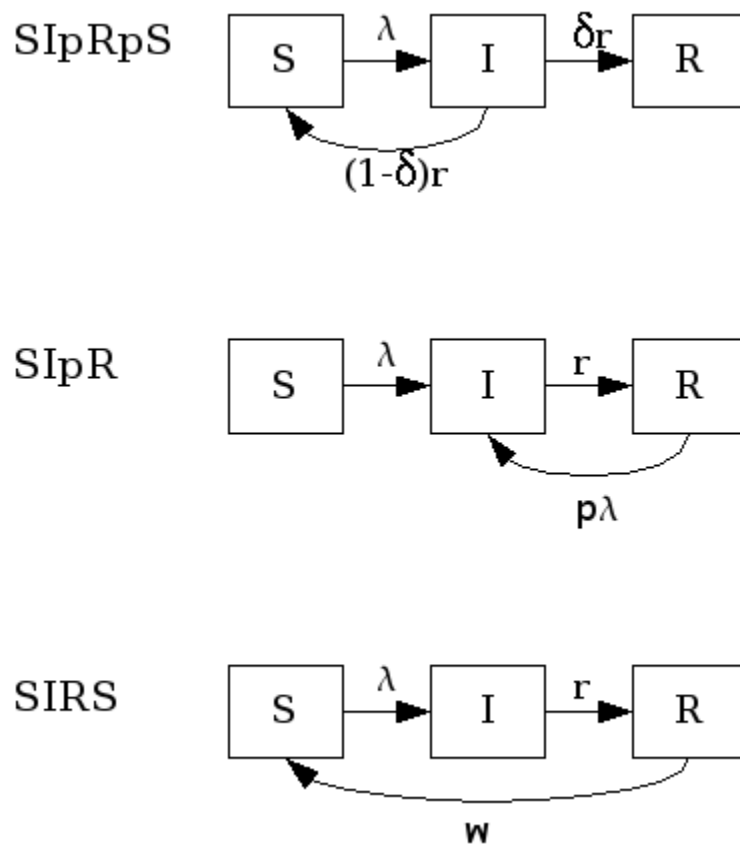


Fig. 2: SIpR-like models.

$$\begin{aligned}
L_{S,t+1} &= \beta S_t \frac{(I_t + \theta)^\alpha}{N_t + n_t} \\
L_{R,t+1} &= p\beta R_t \frac{(I_t + \theta)^\alpha}{N_t + n_t} \\
I_{t+1} &= L_{S,t+1} + L_{R,t+1} + (1 - r)I_t \\
S_{t+1} &= S_t + B - L_{S,t+1} \\
R_{t+1} &= N_t - (S_{t+1} + I_{t+1})
\end{aligned}$$

**SIRS model** Here, the immunity acquired by infection wanes with time. Pertussis is an example of this dynamic.

$$\begin{aligned}
L_{S,t+1} &= \beta S_t \frac{(I_t + \theta)^\alpha}{N_t + n_t} \\
I_{t+1} &= L_{S,t+1} + L_{R,t+1} + (1 - r)I_t \\
S_{t+1} &= S_t + B - L_{S,t+1} + wR_t \\
R_{t+1} &= N_t - (S_{t+1} + I_{t+1})
\end{aligned}$$

### SnInRn-like Models

These are models with more than one compartment for susceptibles, infected and recovered stages. They are used when infection involves more than one distinct populations. Vector borne diseases are classical examples where a SIR model is used to describe infection in humans and another SIR-like model is used to describe infection in the vector (and/or the reservoir(s)). Dengue fever and yellow fever are examples. Sexually transmitted diseases may also be modeled with SnInRn models if male and female populations are distinguished. These models can be define by the user as a custom model.

### 3.2.3 Stochastic models

All models described so far are deterministic. EpiGrass allows simulation of stochastic processes. This is done by assuming that  $L_{t+1}$  is a random variable with expected value given by the expressions found in the deterministic models. The user may choose the probability distribution for  $L_{t+1}$  between Poisson or Negative Binomial to draw realizations of  $L_{t+1}$ :

$$l_{t+1} \sim \text{Poisson}(L_{t+1})$$

or

$$l_{t+1} \sim \text{NegBin}(I_t, \frac{I_t}{I_t + L_{t+1}})$$

The Poisson distribution assumes independent events while the negative Binomial assume clustering of transmission events.

## 3.3 Network transportation models

The transmission models describe the dynamics of infection in a well-mixed population. EpiGrass allows the user to model the movement of infectious individuals between well-mixed populations, thus simulating the spread of disease through space. EpiGrass represents geographical space as a network where cities or localities are nodes and transportation routes are edges. The term network refers to the framework of routes within a system of locations, identified as nodes or sites. An edge is a single link between two sites (a road, a railroad, an air route or a river/sea corridor).

Transportation networks, like many networks, are generally embodied as set of locations and a set of links representing connections between those locations. The arrangement and connectivity of a network is known as its *topology*. Major types of topology are illustrated in figure [ref{fig:artgraphs}](#). Velocity and direction of disease spread depend on the topology and weight of the edges of the transport network and there are many properties of networks that may useful when analyzing the spread of diseases. EpiGrass calculates a set of these properties as described in chapter [Analysis](#).

In a transportation network, each edge (or link) is characterized by a variable *flow* which states the number of passengers that travel through that link per unit time. EpiGrass uses this information to calculate the number of passengers arriving at each city, per time step. For example, consider node *N1* in figure [A simple transportation network](#). At each time step, it receives 10 passengers from *N2*, 5 from *N5*, 1 from *N4*. Now suppose that, at this time step, 10% of the population within each site is infectious (*I* state), according to the epidemic model. Thus, a total of  $10\% \times 10 + 10\% \times 5 + 10\% \times 1 = 1.6$  infectious individuals are visiting site *N1*. In the epidemic model embedded in *N1*, EpiGrass sets  $n = 16$  and  $\theta = 1.6$ . This calculation of  $\theta$  is based on a deterministic argument. The other possibility, allowed by Epigrass, is to define  $\theta$  as a random variable, that follows a binomial distribution with parameters  $n, p$ , where  $n$  is as given in the deterministic version and  $p$  is the proportion of infectious individuals in the source population. From version 1.4 on, EpiGrass can also take into account the specific delays of each connection. When the average speed of the transportation system is set to a value greater than zero, epigrass calculates the time required to complete each trip and attributes this delay (in units of time) to the parameter  $\delta$ . So, the number of infectious passengers arriving at any given city, at time  $t$  corresponds to the number of infectious passengers that left the city of origin at time  $t - \delta$ .

Deterministic:

$$\Theta_t = n \times \frac{I_{t-\delta}}{N}$$

Stochastic:

$$\Theta_t \sim \text{Binomial}(N, \frac{I_{t-\delta}}{N})$$

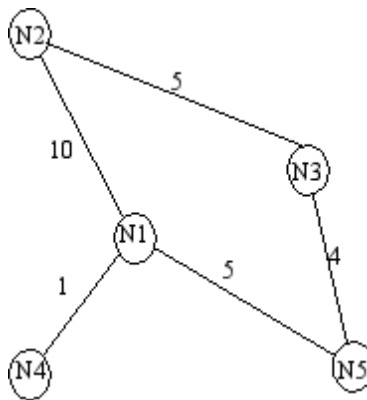


Fig. 3: A simple transportation network

## USING EPIGRASS

To simulate an epidemic process in Epigrass, the user needs to have in hand at least three files: Two files containing the site and edge data and a third file which is a script that defines what it is to be done. Here we go through each one of them in detail. The last part of this chapter is a step-by step guide the Graphical User Interface (GUI).

### 4.1 Data

#### 4.1.1 Site data file

See below an example of the content of a site file for a network of 10 cities. Each line corresponds to a site (except the first line which is the title). For each site, it is declared, in this order: its *spatial location* in the form of a pair of coordinates ([X,Y]); a site \$name\$ to be used in the output; the site's population; the site geocode (an arbitrary unique number which is used internally by Epigrass).

X	Y	City	Pop	Geocode
1	4	"N1"	1000000	1
2	4	"N2"	100000	2
3	4	"N3"	1000	3
4	4	"N4"	1000	4
5	4	"N5"	1000	5
1	3	"N6"	100000	6
2	3	"N7"	1000	7
3	3	"N8"	100000	8
4	3	"N9"	100000	9
5	3	"N10"	1000	10
1	2	"N11"	1000	11

In this example, the first site is located at  $[X,Y]=[1,4]$ , it is named N1, its population is 1000000 and its geocode is 1. This is minimum configuration of a site data file and it must contains this information in exactly this order.

In some situations, the user may want to add other attributes to the sites (different transmission parameters, or vaccine coverage or initial conditions for simulations). This information is provided by adding new columns to the minimum file. For example, if one wishes to add information on the vaccine coverage in cities N1 to N10 (\$vac\$) as well as information about average temperature (which hypothetically affects the transmission of the disease), the file becomes:

X	Y	City	Pop	Geocode	Vac	Temp
1	4	"N1"	1000000	1	0.9	32
2	4	"N2"	100000	2	0.88	29
3	4	"N3"	1000	3	0.7	25
4	4	"N4"	1000	4	0.2	34
5	4	"N5"	1000	5	0	26
1	3	"N6"	100000	6	0	27
2	3	"N7"	1000	7	0	31
3	3	"N8"	100000	8	0	30
4	3	"N9"	100000	9	0	24
5	3	"N10"	1000	10	0	31

During the simulation, each site object receives these informations and store them in appropriate variables that can be used later during model specification. Population is stored in the variable  $N$ ; while the extra columns (those beyond the geocode) are stored in a tuple named  $\{values\}$ . For example, for the city  $N1$ , we have  $N = 1000000$  and  $values=[0.9,32]$ . During model specification, we may use  $N$  to indicate the population size and/or we can use  $values[0]$  to indicate the level of vaccination of that city and  $values[1]$  to indicate the temperature.

It is up to the user, to know what means the elements of the tuple  $values$ . Note that the first element of the tuple has index 0, the second one has index 1 and so on.

When using real data, one may wish to use actual geocodes and coordinates. For example, for a network of Brazilian cities, one may build the following file:

latitude	longitude	local	pop	geocode
-16:19:41	-48:57:10	ANAPOLIS	280164	520110805
-10:54:32	-37:04:03	ARACAJU	461534	280030805
-21:12:27	-50:26:24	ARACATUBA	164449	350280405
-18:38:44	-48:11:36	ARAGUARI	92748	310350405
-21:13:17	-43:46:12	BARBACENA	103669	310560805
-22:32:53	-44:10:30	BARRA_MANSA	165134	330040705
-20:33:11	-48:34:11	BARRETOS	98860	350550005
-26:54:55	-49:04:15	BLUMENAU	241943	420240405
-22:57:09	-46:32:30	B.PAULISTA	111091	350760505

In this file, the coordinates are the actual geographical latitude and longitude coordinates. This information is important when using Epigrass with a map in shapefile format. The geocode is also the official geocode of these localities (the same one used in the shapefile). Despite the cumbersome size of the number, it may be worth using it because demographic official databases are often linked by this number.

### 4.1.2 Edge data file

The edge data file contains all the direct links between sites. Each line in the file (except for the first, which is the header) corresponds to an edge. For each edge (or link) one must specify (in this order): the *names of the sites* connected by that edge; the *number of individuals traveling from source to destination*; the *number of individuals travelling from destination to source* per time step; the *distance or length* of the edge. At last, the file must contain, in the fifth and sixth columns, the *geocodes* of the source and destination sites\*. This is very important as the graph is built internally connecting sites through edges and this is done based on geocode info.

**Warning:** It is required that the order of columns is kept the same.

See below the list of the 8 edges connecting the sites *N1* to *N10*. Let's look the first one, as an example. It links *N1* to *N2*. Through this link passes 11 individuals backwards and forwards per time step (a day, for example). This edge has length 1 (remember that *N1* is at  $[X,Y]=[1,1]$  and *N2* is at  $[X,Y]=[1,2]$ , so the distance between them is 1). The last two columns show the geocode of *N1* (geocode 1) and the geocode of *N2* (geocode 2).

Source	Dest	flowSD	flowDS	Distance	geoSource	geoDest
N1	N2	11	11	1	1	2
N2	N4	0.02	0.02	1	3	4
N3	N8	1.01	1.01	1	3	8
N4	N9	1.01	1.01	1	4	9
N5	N10	0.02	0.02	1	5	10
N6	N5	1.01	1.01	1	7	8
N7	N10	1.01	1.01	1	7	8
N9	N10	1.01	1.01	1	9	10

Note that it doesn't matter which site is considered a Source and which one is considered a Destination. I.e., if there is a link between *A* and *B*, one may either named *A* as source and *B* as destination, or the other way around.

If the edge represents a road or a river, one may use the actual metric distance as length. If the edge links arbitrary localities, one may opt to use euclidean distance, calculated from the *x* and *y* coordinates (using Pythagoras theorem).

## 4.2 Specifying a model: the script

Once the user has specified the two data files, the next step is to define the model to be executed. This is done in the .epg script file. The .epg script is a text file and can be edited with any editor (not word processor). This script must be prepared with care.

The best way to write down your own .epg is to edit an already existing .epg file. So, open Epigrass, choose an .epg file and click on the *Edit* button. Your favorite editor will open and you can start editing. Don't forget to save it as a new file in your working directory. Of course, there is an infinite number of possibilities regarding the elaboration of the script. It all depends on the goals of the user.

---

**Note:** Another way to edit an .epg file is to open it with the graphical editor provided with Epigrass. Just type `epgeditor yourmodel.epg`.

---

For the beginner, we suggest him/her to take a look at the .epg files in the demo directory. They are all commented and may help the user in getting used with Epigrass language and capabilities.

Some hints to be successful when editing your .epg:

- All comments in the script are preceded by the symbol `#`. These comments may be edited by the user as he/she wishes and new lines may be added at will. Don't forget, however, to place the symbol `#` in every line corresponding to a comment.
- The script is divided into a few parts. These parts have capital letter titles within brackets. Don't touch them!
- Don't remove any line that is *not* a comment. See below how to appropriately edit these command lines.

Let's take a look now at each part of a script (this is the script `.epg demo` file):

**Warning:** All variables defined in a .epg are **case-insensitive**. Consider this fact when naming your model's variables.

### 4.2.1 Part 1: THE WORLD

The first section of the script is titled: THE WORLD. An example of its content is shown:

```
shapefile = ['riozonas_LatLong.shp', 'nome_zonas', 'zona_trafe']
edges = edges.csv
sites = sites.csv
encoding =
```

where,

**shapefile** Is a list with 3 elements: the first is the path, relative to the working directory, of the shapefile file;

the second is the variable, in the shapefile, which contains the names of the localities (polygons of the map); the third and last is the variable, in the shapefile, which contains the geocode of the localities. If you don't have a map for you simulation, leave the list empty: *location = []*). Epigrass also accepts Maps in the *Geopackage* format. Epigrass assume that maps use the *EPSG:4326* projection. Using other projection may result in errors related to the visualizaton of results. **edges**

This is the name of the .CSV (comma-separated-values) file containing the list of edges and their attributes.

**sites** This is the name of the .CSV file containing the list of sites and their attributes.

**encoding** This is the encoding used in your sites and edges files. This is very important when you use location names which include non-ascii characters. The default encoding is *latin-1*. If you use any other encoding, please specify it here. Example: *utf-8*.

---

**Note:** All paths in the .epg file are relative to the working directory.

---

### 4.2.2 Part 2: EPIDEMIOLOGICAL MODEL

This is the main part of the script. It defines the epidemiological model to be run. The script reads:

```
modtype = SIR
```

Here, the type of epidemiological model is defined, in this case is a deterministic *SIR* model. Epigrass has some built-in models:

Name	Determ.	Stochastic
Susceptible-Infected-Recovered	<i>SIR</i>	<i>SIR_s</i>
Susceptible-Exposed-Infected-Recovered	<i>SEIR</i>	<i>SEIR_s</i>
Susceptible-Infected-Susceptible	<i>SIS</i>	<i>SIS_s</i>
Susceptible-Exposed-Infected-Susceptible	<i>SEIS</i>	<i>SEIS_s</i>
SIR with fraction with full immunity	<i>SIpRpS</i>	<i>SIpRpS_s</i>
SEIR with fraction with full immunity	<i>SEIpRpS</i>	<i>SEIpRpS_s</i>
SIR with partial immunity for all	<i>SIpR</i>	<i>SIpR_s</i>
SEIR with partial immunity for all	<i>SEIpR</i>	<i>SEIpR_s</i>
SIR with immunity wane	<i>SIRS</i>	<i>SIRS_s</i>

A description of these models can be found in the chapter *Epidemiological modeling*. The stochastic versions of models use *Poisson* distribution as default for the number of new cases ( $L(t+1)$ ), They are not full stochastic models. Besides these, the user may define his/her own model and access it by setting **modtype = Custom**.



### 4.2.3 Part 3: MODEL PARAMETERS

The epidemic model is defined by variables and parameter which require initialization:

```
#####
# They can be specified as constants or as functions of global or
# site-specific variables. Site-specific variables are provided
# in the sites .csv file. In this file, all columns after the 4th
# are collected into a values tuple, which can be referenced here
# as values[0], values[1], values[2], etc.
# Examples:
# beta = 0.001
# beta=values[0] #assigns the first element of values to beta
# beta=0.001*values[1]
# beta=0.001*N # N is a global name for total site population
# Currently, Epigrass requires that parameters beta, alpha, e, r, delta, B, w, p
# be present in the .epg even if they will not be used. Do not erase these lines.
# Just disregard them if they are not useful to you.

beta = 0.4 #transmission coefficient (contact rate * transmissibility)
alpha = 1 # clumping parameter
e = 1 # inverse of incubation period
r = 0.1 # inverse of infectious period
delta = 1 # probability of acquiring full immunity [0,1]
B = 0 # Birth rate
w = 0 # probability of immunity waning [0,1]
p = 0 #
```

These are the model parameters. Not all parameters are necessary for all models. For example,  $e$  is only required for SEIR-like models. Don't remove the line, however because that will cause an error. We recommend that, if the parameter is not necessary, just add a comment after it as a reminder that it is not being used by the model.

In some cases, one may wish to assign site-specific parameters. For example, transmission rate may be different between localities that are very distant and are exposed to different climate. In this case site specific variables can be added as new columns to the site file. All columns after the geocode are packed into a tuple named *values* and can be referenced in the order they appear. I.e., the first element of the tuple is *values[0]*, the second element is *values[1]*, the third element is *values[2]* and so on.

#### Part 4: INITIAL CONDITIONS

In this part of the script, the initial conditions are defined. Here, the number of individuals in each epidemiological state, at the start of the simulation, is specified. It reads:

```
#####
# Here, the number of individuals in each epidemiological
# state (SEI) is specified. They can be specified in absolute
# or relative numbers.
# N is the population size of each site.
# The rule defined here will be applied equally to all sites.
# For site-specific definitions, use EVENTS (below)
# Examples:
# S,E,I = 0.8*N, 10, 0.5*N
# S,E,I = 0.5*N, 0.01*N, 0.05*N
# S,E,I = N-1, 1, 0
S = N
E = 0
I = 0
```

Here,  $N$  is the total population in a site (as in the datafile for sites). In this example, we set all localities to the same

initial conditions (all individuals susceptible) and use an event (see below) to introduce an infectious individual in a locality. The number of recovered individuals is implicit, as  $R = N - (S + E + I)$

Another possibility is to define initial conditions that are different for each site. For this, the data must be available as extra columns in the site data file and these columns are referenced to using the *values* tuple explained above.

## 4.2.4 Part 5: EPIDEMIC EVENTS

The next step is to define events that will occur during the simulation. These events may be epidemiological (arrival of an infected, for example) or a public health action (vaccination campaign, for example):

```
#=====#
#   Specify isolated events.
#   Localities where the events are to take place should be Identified by the geocode,
#   ↪ which
#   comes after population size on the sites data file.
#   All coverages must be a number between 0 and 1.
#   Seed : [('locality1's geocode', epid state, n), ('locality2's geocode', epid state, ↪
#   ↪n),...]
#   N infected cases will be added to locality at time 0.
#   Vaccinate: [('locality1's geocode', [t1,t2,...], [cov1,cov2,...]), ('locality2's ↪
#   ↪geocode', [t1,t2,...], [cov1,cov2,...])]
#   Quarantine: [(locality1's geocode,time,coverage), (locality2's geocode,time,
#   ↪coverage)]
#seed = [(4550601,'ip20',10)] #santo cristo #
seed = [(4552110,'ip20',10)] #pechincha #
#seed = []
Vaccinate = [] #
Quarantine = []
```

The events currently implemented are:

**seed** One or more infected individual(s) are introduced into a site, at the beggining of the simulation. The notation for a single event is: *Seed* = [(‘locality1’s geocode’, epid state, n), (‘locality2’s geocode’, epid state, n), ...]. For example, *seed* = [(2,‘I’,10)] programs the arrival of 10 infected individuals at site geocode 2, at time 1.

**Vaccinate** Implements a campaign that vaccinates a fraction of the population in a site, at a pre-defined time-step. For multiple events, the notation is: [(‘locality1’s geocode’, [t1,t2,...], [cov1,cov2,...]), (‘locality2’s geocode’, [t1,t2,...], [cov1,cov2,...])], where the first element of every triplet is the geocode of the city, the second element is a list of the time(s) when the campaign is carried on, and the third element is the coverage(s). For example, the event [(2,[10],[0.7])] means that city 2, at time 10, has 70% of its population vaccinated. Mathematically, it means (in the model), the removal of individuals from the susceptible to the recovered stage.

**Quarantine** Prevents any individual from leaving a site, starting at *t*. Currently disabled.

## 4.2.5 Part 6: TRANSPORTATION MODEL

Here, there are two options regarding the movement of infected individuals from site to site (through the edges). If *stochastic* = 0, the process is simulated deterministically. The number of infected passengers commuting through an edge is a fraction *p* of the infected population that is traveling. *p* is calculated as *total passengers/total population*.

If *stochastic* = 1, the number of passengers is sampled from a Poisson distribution with parameter given by the expected number of travelling infectives (calculated as above):

```
#=====#
# If doTransp = 1 the transportation dynamics will be
```

(continues on next page)

(continued from previous page)

```
# included. Use 0 here only for debugging purposes.
doTransp = 1

# Mechanism can be stochastic (1) or deterministic(0).
stochastic = 1
#Average speed of transportation system in km per time step. Enter 0 for
↳instantaneous travel.
#Distance unit must be the same specified in edges files
speed =0 #1440 km/day -- equivalent to 60 km/h
```

That ends the definition of the model.

## 4.2.6 Part 7: SIMULATION AND OUTPUT

Now it is time to define some final operational variables for the simulation:

```
#=====#
# Number of steps
steps = 50

# Output dir. Must be a full path. If empty the output will be generated on the
# same path as the model script.
outdir =

# Output file
outfile = simul.dat

# Database Output
# SQLout can be 0 (no database output) or 1
SQLout = 1

# Report Generation
# The variable report can take the following values:
# 0 - No report is generated.
# 1 - A network analysis report is generated in PDF Format.
# 2 - An epidemiological report is generated in PDF Format.
# 3 - A full report is generated in PDF Format.
# siteRep is a list with site geocodes. For each site in this list, a detailed report
↳is apended to the main report.
report = 0
siteRep = []

# Replicate runs
# If replicas is set to an integer(n) larger than zero, the model will be run
# n times and the results will be consolidated before storing.
# RandSeed = 1: the seed will be randomized on each replicate
# RandSeed = 2: seeds are taken sequentially from the site's file
# Note: Replicate mode automatically turn off report and batch options.
Replicas = 10
RandSeed = 2
#Batch Run
# list other scripts to be run in after this one. don't forget the extension .epg
# model scripts must be in the same directory as this file or provide full path.
# Example: Batch = ['model2.epg', 'model3.epg', '/home/jose/model4.epg']
Batch = []#['sarsDF.epg', 'sarsPA.epg', 'sarsRS.epg']
```

where,

**step** Number of time steps for the simulation.

**outdir** Directory for data output (currently not in use)

**outfile** .csv filename that can be imported into R as a dataframe. This .csv file contains the simulated timeseries for all nodes.

**sqlout** Use `sqlout = 1` if simulated time series are to be stored. Time series of *L*, *S*, *E*, and *I*, from simulations, are stored in a SQLite database named *Epigrass.sqlite*. The results of each individual simulation is stored in a different table named after the model's script name, the date and time the simulation has been run. For instance, suppose you run a simulation of a model stored in a file named `script.epg`, then at the end of the simulation, a new table in the epigrass database will be created with the following name: `script_Wed_Jan_26_154411_2005`. Thus, the results of multiple runs from the same model get stored independently. **Batch=[]**

Script files included in this list are executed after the currently file is finished.

## 4.2.7 Operation

For running simulations, after all the information has been entered and checked on the first tab of the GUI, you can press the *Run* button to start the simulation or the `:guilabel`Exit`` button. When the *Run* button is pressed, the `~/ .epigrassrc` file is updated with all the information entered in the GUI. If the *Exit* button is pressed, all information entered since the last time the *Run* button was pressed is lost.

Epigrass also allows running simulation straight from the command line, with the **epirunner** executable. all you have to do is:

```
$ epirunner mymodel.epg
```

and the model will executed with the settings specified in the `~/ .epigrassrc` file. for help with *epirunner*, type:

```
$ epirunner -h
No Custom Model Available.
usage: epirunner [options] your_model.epg

Run epigrass models from the console

positional arguments:
  EPG                  Epigrass model definition file (.epg).

optional arguments:
  -h, --help            show this help message and exit
  -b <mysql|sqlite|csv>, --backend <mysql|sqlite|csv>
                        Define which datastorage backend to use
  -u DBUSER, --dbusername DBUSER
                        MySQL user name
  -p DBPASS, --password DBPASS
                        MySQL password for user
  -H DBHOST, --dbhost DBHOST
                        MySQL hostname or IP address
  --upload UPLOAD        Upload your models and latest simulation to Epigrass Web
  -P, --parallel        use multiprocessing to run the simulation
  -D, --dashboard        Open dashboard on browser after th run is done.
  -V, --view-only        Only Open dashboard.
```

## 4.2.8 The Web Dashboard

To interact and explore the simulations stored in the database, Epigrass offers an interactive dashboard. It allows you to navigate to any simulation previously done for a given model, and inspect it interactively. You can ask for it to open **right after a simulation is done**:

```
$ epirunner -D mymodel.epg
```

or you can simply open it to inspect previous simulations of given model by typing from the directory where the *epg* file is located:

```
$ epirunner -V mymodel.epg
```

The Dashboard shows various aspects of the Model output, such as Choroplectic map with the final state of the simulation,

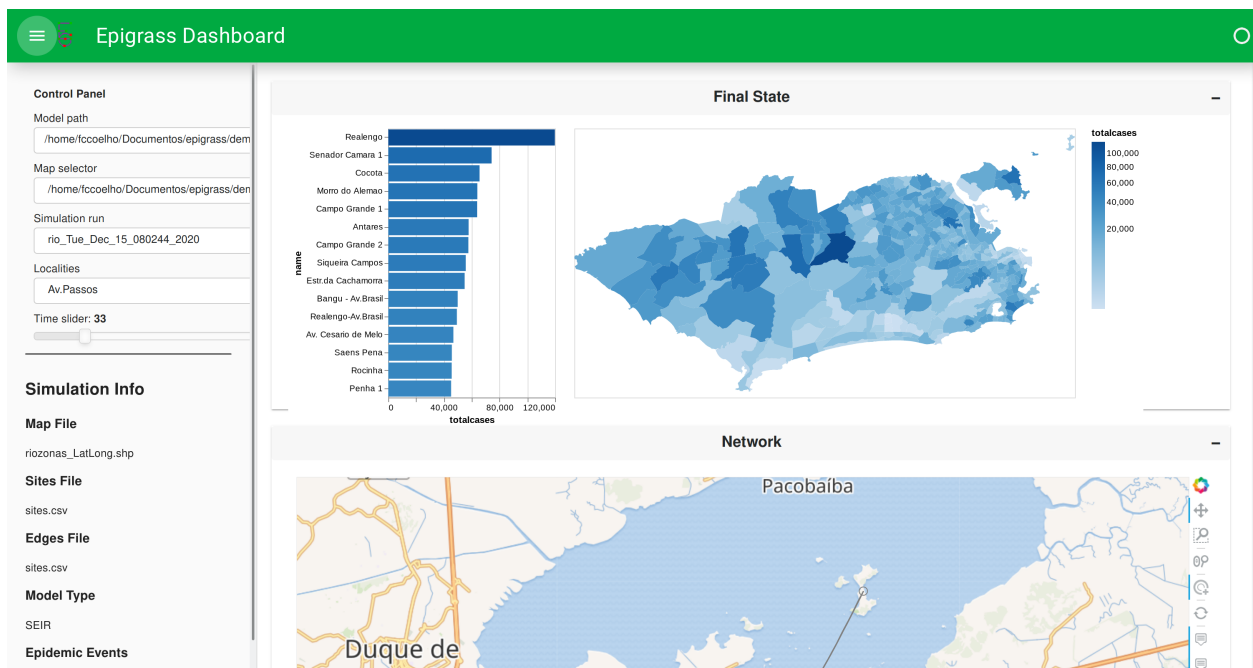


Fig. 1: Figure: Dashboard showing map with the final state of the model.

It also shows the network neighborhood of any of the localities of the models

It also shows a map which can be animated and timeseries plots of all the state variables

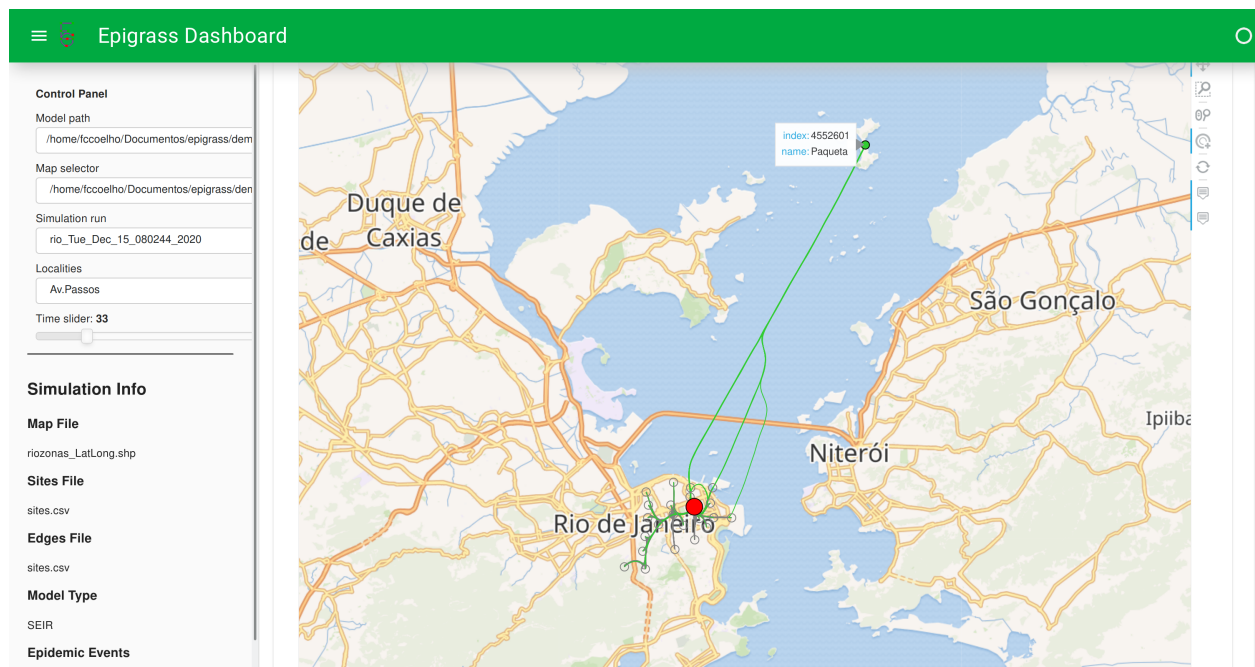


Fig. 2: Figure: Dashboard showing network neighborhood of a locality marked in red.

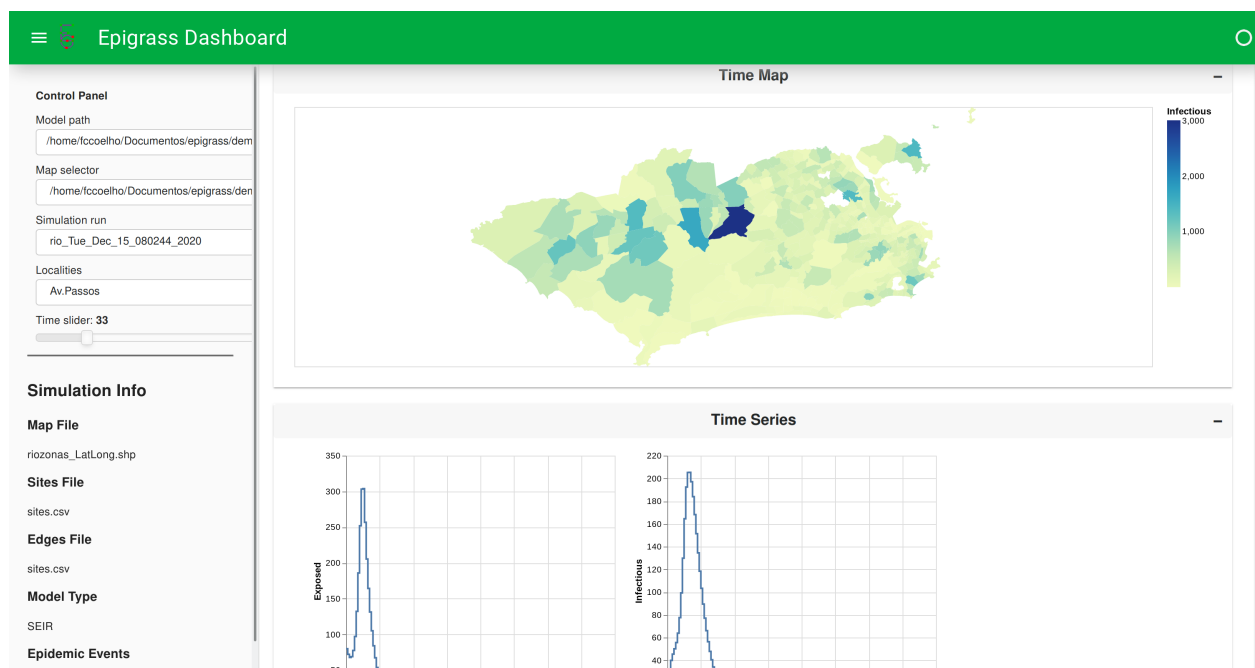


Fig. 3: Figure: Dashboard showing timeseries visualizations.

**ANALYSIS**





## WRITING CUSTOM MODELS

The most powerful feature of Epigrass is the ability to use custom models. It allows the user to specify intra-node dynamics and, in doing so, break away from the limitations imposed by the built-in models.

By learning to write his/her own models, the user begins to realize the full potential of Epigrass, which goes beyond being a platform to simulate networked epidemics. In reality Epigrass can be used to model any distributed dynamical system taking place on a set of nodes (connected or not).

### 6.1 Getting Started

The best way to get started in writing custom models is to look at the example distributed with Epigrass. It can be found in `demos/CustomModel_example.py`:

```
# This is a custom model to used in place of Epigrass' built-in models. Custom
# models must always be on a file named CustomModel.py and contain at least
# a function named Model. Both the File name and the function Names are case-
→sensitive,
# so be careful. Please refer to the manual for intructions on how to write your
# own custom models.

def Model(self,vars,par,theta=0, npass=0):
    """
    Calculates the model SIR, and return its values.
    * vars The state variables of the models
    * par The parameters (Beta, alpha, E,r,delta,B, w, p) see docs.
    * theta = infectious individuals from neighbor sites
    * npass = Total number of people arriving at this node
    """
    # Get state variables' current values

    if self.parentSite.parentGraph.simstep == 1: # if first step
        # Define variable names to appear in the output
        self.parentSite.vnames = ('Exposed','Infectious','Susceptible')
        # And get state variables's initial values (stored in dict self.bi)

        E,I,S = (self.bi['e'],self.bi['i'],self.bi['s'])
    else: # if nor first step
        E,I,S = vars

    # Get parameter values
    N = self.parentSite.totpop
    beta,alpha,e,r,delta,B,w,p = (self.bp['beta'],self.bp['alpha'],
    self.bp['e'],self.bp['r'],self.bp['delta'],self.bp['b'],
```

(continues on next page)

(continued from previous page)

```

self.bp['w'],self.bp['p'])

#Vaccination event (optional)
if self.parentSite.vaccineNow:
    S -= self.parentSite.vaccov*S

# Model
Lpos = beta*S*((I+theta)/(N+npass))**alpha #Number of new cases
Ipos = (1-r)*I + Lpos
Spos = S + B - Lpos
Rpos = N-(Spos+Ipos)

# Update stats
self.parentSite.totalcases += Lpos #update number of cases
self.parentSite.incidence.append(Lpos)

# Raise site infected flag and add parent site to the epidemic history list.
if not self.parentSite.infected:
    if Lpos > 0:
        self.parentSite.infected = self.parentSite.parentGraph.simstep
        self.parentSite.parentGraph.epipath.append(
            (self.parentSite.parentGraph.simstep,
             self.parentSite,self.parentSite.infector))

self.parentSite.migInf.append(Ipos)

return [0,Ipos,Spos]

```

Let's analyze the above code. The first thing to notice is that an Epigrass custom model is a Python program. So anything you can do with Python in your system, you can do in your custom model. Naturally, your knowledge of the Python programming language will define how far you can go in this customization. There are a few requirements on this Python program in order to make it a valid custom model from Epigrass's perspective.

1. **It must define a global function named Model. This function will be inserted as a method on every node object, at run time**

1. **This function must declare the following arguments:**

- *self*: reference to the model object.
  - *vars*: A list with the values of the model's state variables in time  $t-1$  in the same order as returned by this function.
  - *par*: The parameters of the model. Listed in the same order as defined in the .epg file.
  - *theta*: Number of infectious individuals arriving from neighboring sites. For disconnected models, it is 0.
  - *npass*: The total number of passengers arriving from neighboring sites. For disconnected models, it is 0.
2. In the beginning of the function you define a list of strings (`self.parentSite.vnames`) which will be the names used when storing the resulting time-series in the database. Choose strings that are not very long and are meaningful. You only need to do this once, at the beginning of the simulation so put it inside an *if* statement, which will be executed only at time-step 1 (see code above).
  3. After defining variable names, set their initial values in the same *if* clause. An *else* clause linked to this one will set variables values for the rest of the simulation.
  4. Define local names for the total population  $N$  and fixed parameters.

5. Proceed to implement your model anyway you see fit.
6. **Feed some site level variables (*incidence*,) with the result of the simulation.**
  - *incidence*: list of new cases per time step.
  - *infected*: Boolean stating if the site has been infected, i.e., it has had an autoctonous case.
  - *epipath*: This variable is at the graph level and contains the path of spread of the simulation.
  - *migInf*: Number of infectious individuals in this site per time-step.
7. Finally, this function must return a list/tuple with the values of the state variables in the same order as received in vars.

**Warning:** The strings in `self.parentSite.vnames` must be valid *SQL* variable names, or else you will have a insert error at the end of the simulation.

After defining this function with all its required features, you can continue to develop you custom model, writing other functions classes, etc. Note however, that only the *Model* function will be called by Epigrass, so any other code you add to your program must be called from within that function.

**Note:** Since `CustomModel` is imported from within Epigrass, any global code (unindented) in it is also executed. So you may add imports and other initialization code.

**Warning:** The name `CustomModel.py` is case-sensitive and cannot be changed. The same is true for the *Model* function.

## 6.2 The Environment

Nesting of the objects inside a `Simulate` object.

From quickly going through the example Custom model above it probably became clear, to the Python-initiated, that You can access variables at the node and graph levels. This is possible because *Model* becomes a method in a node object which in is turn is contained into a graph object (see figure).

Besides being nested within the *graph* object, *node* and *edge* contain references to their containers. This means that using the introspective abilities of Python the user can access any information at any level of the full *graph* model and use it in the custom model. In order to help you do this, Let's establish an API for developing custom models.

### 6.2.1 Model Development API

All attributes and methods (functions) from all around the simulation must be references from the model's perspective, denoted by *self*. The parent objects can be accessed through the following notation:

- *self.parentSite* Is the Site (node) containing the model.
- *self.parentSite.parentGraph* Is the Graph containing the parent site of the model.

The following attributes and methods can be accessed by appending them to one one the objects above. For example:

```
self.parentSite.parentGraph.simstep
```

## Site Attributes and Methods

Not all attributes and methods are listed, only the most useful. For a complete reference, look at the source code documentation.

### **class Site**

**self.parentSite.** Actually named *siteobj* in the source code.

#### **bi**

Dictionary with initial values for all of the model's state variables. Keys are the variable names.

#### **bp**

Dictionary with initial values for all of the model's parameters. Keys are the parameter names.

#### **totpop**

Initial total population

#### **ts**

List containing the model output time series (variables in the same order of the model)

#### **incidence**

Incidence time series

#### **infected**

Has the site been already infected? (logical variable)

#### **sitename**

Site's name (provided in the .csv)

#### **values**

Tuple containing extra-variables provided by .csv file

#### **parentGraph**

Graph to which Site belongs (see class Graph)

#### **edges**

List containing all edge objects connected to Site

#### **inedges**

List containing all inbound edges

#### **outedges**

List containing all outbound edges

#### **geocode**

Site's geocode

#### **modtype**

Type of dynamic model running in Site

#### **vaccination**

Time and coverage of vaccination event. Format as in .epg

#### **vaccineNow**

Flag indicating that it is vaccine day (0 or 1)

#### **vaccov**

Current vaccination coverage

#### **vaccinate** (*cov*)

At time *t*, the population is vaccinated with coverage *cov*

#### **getOutEdges** ()

Returns list of outbound edges

**getInEdges** ()  
Returns list of inbound edges

**getNeighbors** ()  
Returns a dictionary of neighbor sites as keys and distances as values

**getDistanceFromNeighbor** (*site*)  
Returns the distance in km from a given neighbor

**getDegree** (*site*)  
Returns degree of this site, that is, the number of sites connected to it

## Graph Attributes and Methods

Not all attributes and methods are listed, only the most useful. For a complete reference, look at the source code documentation.

**class Graph**  
self.parentSite.parentGraph

**simstep**  
Time-step of the simulation. Use it to keep track of the simulation progress.

**speed**  
The speed of the transportation system

**maxstep**  
Final time-step of the simulation

**episize**  
Current size of the epidemic, graph-wise.

**site\_list**  
Full list of nodes in the graph. Each element in this list is a real node object.

**edge\_list**  
Full list of nodes in the graph. Each element in this list is a real node object.

**getSite** (*name*)  
Returns an site object named *name*



## B

bi (*Site attribute*), 32  
bp (*Site attribute*), 32

## C

custom models, 29

## D

Database  
    epigrass database, 6  
    results table, 6

## E

edge, 4  
edge\_list (*Graph attribute*), 33  
edges (*Site attribute*), 32  
Epidemiological models, 9  
episize (*Graph attribute*), 33

## G

geocode (*Site attribute*), 32  
getDegree() (*Site method*), 33  
getDistanceFromNeighbor() (*Site method*), 33  
getInEdges() (*Site method*), 32  
getNeighbors() (*Site method*), 33  
getOutEdges() (*Site method*), 32  
getSite() (*Graph method*), 33  
Graph (*built-in class*), 33

## I

incidence (*Site attribute*), 32  
inedges (*Site attribute*), 32  
infected (*Site attribute*), 32

## M

maxstep (*Graph attribute*), 33  
Model, 9  
Models  
    SEIR, 12  
    SEIS, 13  
    SIpRpS, 13

SIR, 12  
SIRS, 15  
SIS, 12  
Stochastic, 15  
transportation, 15  
typology, 10  
models, 4  
    epidemiological models, 4  
    network models, 4  
modtype (*Site attribute*), 32

## N

node, 3

## O

outedges (*Site attribute*), 32

## P

parentGraph (*Site attribute*), 32

## S

shapefile, 3  
simstep (*Graph attribute*), 33  
site, 3  
Site (*built-in class*), 32  
site\_list (*Graph attribute*), 33  
sitename (*Site attribute*), 32  
speed (*Graph attribute*), 33

## T

totpop (*Site attribute*), 32  
ts (*Site attribute*), 32

## V

vaccinate() (*Site method*), 32  
vaccination (*Site attribute*), 32  
vaccineNow (*Site attribute*), 32  
vaccov (*Site attribute*), 32  
values (*Site attribute*), 32