Introduction to mathematical epidemiology

Julien Arino

arinoj@cc.umanitoba.ca

U of M – April 2007

Outline

What is mathematical epidemiology?

Use of mathematical techniques to understand the mechanisms leading to the spread of diseases (mostly infectious) in populations.

- deterministic vs stochastic models
- many different modelling paradigms: ODEs, PDEs, DDEs, SDEs, integral equations, branching processes, chain stochastic models..

Outline

Outline

A susceptible-infected-susceptible model

Take a population with $N(t)$ individuals at time t, and consider an infectious disease. Suppose that individuals are

- susceptible to a disease, with number $S(t)$ at time t,
- infectious with the disease, with number $I(t)$ at time t.

Need to describe several processes:

- demography,
- infection,
- other disease related events.

Demography

For the whole population $N(t) = S(t) + I(t)$,

$$N' = bN - dN$$

with b the birth rate and d death rate. Need to split this between S and I. Assume no vertical transmission (from mother to child), then

$$S' = bN - dS$$
$$I' = -dI$$

Infection

At time t, there are $S(t)$ susceptible and $I(t)$ infectious individuals.

- Use a *mass action* formulation: there are $S(t)I(t)$ contacts.
- Or use *proportional incidence*: the population is large, and no individual can meet every other individual, there are $S(t)I(t)/N$ contacts.

Take proportional incidence. For each contact, there is a probability β that the contact leads to infection,

- $\Rightarrow$ rate of apparition of new infectious individuals is $\beta S(t)I(t)/N$.

So

$$S' = -\beta \frac{SI}{N}$$
$$I' = \beta \frac{SI}{N}$$

So the model with demography and infection is

$$S' = bN - \beta \frac{SI}{N} - dS$$

$$I' = \beta \frac{SI}{N} - dI$$

Additional disease processes

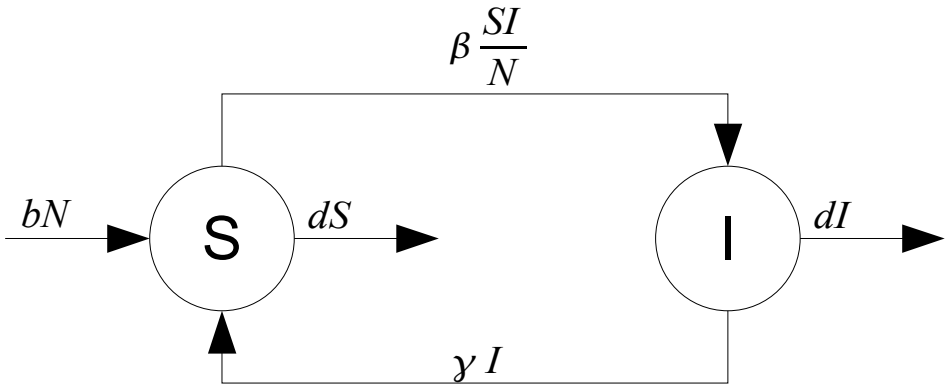
We can choose to add additional properties:

- disease induced death, at the rate δ,
- recovery from the disease, at the rate γ.

An SIS model:

$$S' = bN + \gamma I - \beta \frac{SI}{N} - dS$$

$$I' = \beta \frac{SI}{N} - (d + \delta + \gamma) I$$



For simplicity, assume equal birth and death ($b = d$) and no disease induced mortality ($\delta = 0$), so N is constant:

$$S' = dN + \gamma I - \beta \frac{SI}{N} - dS \tag{1a}$$

$$I' = \beta \frac{SI}{N} - (d + \delta) I \tag{1b}$$

Diseases

- SI case ($\gamma = 0$):
  - ghonorrhea
  - herpes
  - all chronic infections
- SIS case ($\gamma > 0$):
  - common cold

Outline

Equilibria

An equilibrium is a point such that $S' = I' = 0$. We solve for (S, I):

$$dN + \gamma I - \beta \frac{SI}{N} - dS = 0$$

$$\beta \frac{SI}{N} - (d + \gamma) I = 0$$

Find the *disease free equilibrium* (DFE)

$$(S, I) = (N, 0)$$

and the *endemic equilibrium* (EEP)

$$(S, I) = \left(\frac{d + \gamma}{\beta} N, \left(1 - \frac{d + \gamma}{\beta} \right) N \right)$$

which exists only when $(d + \gamma)/\beta < 1$.

Stability of the equilibria

Theorem 1

Define the basic reproduction number,

$$\mathcal{R}_0 = \frac{\beta}{d+\gamma}.$$

If $\mathcal{R}_0 < 1$, then the EEP does not exist, and for all positive initial conditions,

$$\lim_{t\to\infty}(S(t), I(t)) = (N, 0).$$

If $\mathcal{R}_0 > 1$, then for all positive initial conditions,

$$\lim_{t\to\infty}(S(t), I(t)) = \left(\frac{d+\gamma}{\beta}N, \left(1 - \frac{d+\gamma}{\beta}\right)N\right)$$

The basic reproduction number

$$\mathcal{R}_0 = \frac{\beta}{d+\gamma}$$

is product of

- β, the probability of contracting the disease when a potentially infecting contact occurs,
- $\frac{1}{d+\gamma}$, the mean time spent in the infectious class when subject to the competing risks of natural death and recovery.

$\Rightarrow \mathcal{R}_0$ is the average number of infectives produced when one infective individual is introduced into a wholly susceptible population.

Reinterpretation of the result

Given the average number of new infectives generated by one infective individual introduced in a wholly susceptible population,

$$\mathcal{R}_0 = \frac{\beta}{d+\gamma},$$

we have that,

- if $\mathcal{R}_0 < 1$, then the disease goes extinct,
- if $\mathcal{R}_0 > 1$, then the disease becomes established in the population, for all positive initial conditions,

$$\lim_{t\to\infty}(S(t), I(t)) = \left(\frac{1}{\mathcal{R}_0}N, \left(1-\frac{1}{\mathcal{R}_0}\right)N\right)$$

Outline

Modification of our model to incorporate vaccination

We vaccinate a proportion p of the newborns.

These vaccinated newborns do not enter the susceptible population, so only a proportion $1-p$ of newborns are susceptible. The model becomes

$$S' = (1-p)d(S+I) + \gamma I - \beta\frac{SI}{N} - dS \tag{2a}$$

$$I' = \beta\frac{SI}{N} - (d+\gamma)I \tag{2b}$$

The DFE becomes

$$(S, I) = ((1-p)N, 0)$$

Stability of the equilibria

Theorem 2

Define the basic reproduction number with vaccination,

$$\mathcal{R}_v = (1-p)\mathcal{R}_0.$$

If $\mathcal{R}_v < 1$, then the EEP does not exist, and for all positive initial conditions,

$$\lim_{t\to\infty}(S(t), I(t)) = ((1-p)N, 0)\,.$$

If $\mathcal{R}_v > 1$, then for all positive initial conditions,

$$\lim_{t\to\infty}(S(t), I(t)) = \left((1-p)\frac{1}{\mathcal{R}_0}N, (1-p)\left(1-\frac{1}{\mathcal{R}_0}\right)N\right)$$

Herd immunity

$\mathcal{R}_0$ is the "natural" reproduction number, $\mathcal{R}_v$ is the one on which we have an influence. The objective of vaccination is thus to bring $\mathcal{R}_v$ to a value less than 1.

We have

$$\mathcal{R}_v < 1 \Leftrightarrow (1-p)\mathcal{R}_0 < 1 \Leftrightarrow p > 1 - \frac{1}{\mathcal{R}_0}$$

This is an extremely important concept, called *herd immunity*. To drive the disease to extinction, it is only necessary to immunize a fraction of the population larger than $1 - 1/\mathcal{R}_0$.

$\mathcal{R}_0$ for measles

Geographical location	Time period	$\mathcal{R}_0$	$1 - \frac{1}{\mathcal{R}_0}$
Cirencester, England	1947-50	13-14	0.923-0.928
England and Wales	1950-68	16-18	0.937-0.944
Kansas, USA	1918-21	5-6	0.8-0.833
Ontario, Canada	1912-3	11-12	0.909-0.916
Willesden, England	1912-3	11-12	0.909-0.916
Ghana	1960-8	14-15	0.928-0.933
Eastern Nigeria	1960-8	16-17	0.937-0.941

Outline

Impact of an epidemic

Two important questions:

1. How can we evaluate the *impact* (human, economic) of an infectious disease?
2. Does an epidemic involve all the individuals in the population?

$\Rightarrow$ size of an epidemic

Kermack-McKendrick, 1927

In 1927, Kermack and McKendrick started publishing a series of papers on epidemic models. In the first of their papers, they have this model as a particular case:

$$\begin{aligned} S' &= -\beta SI \\ I' &= \beta SI - \gamma I \\ R' &= \gamma I \end{aligned} \tag{3}$$

This is an *epidemic* model (whereas the previous model was an *endemic* model). It describes one wave of epidemic through a population.

Here,

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

Analyzing the system

First, note that the total population in the system is constant. This is deduced from the fact that

$$N' = (S + I + R)' = -\beta SI + \beta SI - \gamma I + \gamma I = 0.$$

Since this is true for all values of t, we have N constant.

Let us ignore the R equation, since R can be deduced from I. We compute

$$\frac{dI}{dS} = \frac{dI}{dt}\frac{dt}{dS} = \frac{I'}{S'} = \frac{\gamma}{\beta S} - 1$$

This gives, with initial condition (S_0, I_0),

$$I(S) = S - \frac{\gamma}{\beta}\ln S + I_0 - (S_0 - \frac{\gamma}{\beta}\ln S_0).$$

This gives a curve in the (S, I) plane.

$$I(S) = S - \frac{\gamma}{\beta} \ln S + I_0 - (S_0 - \frac{\gamma}{\beta} \ln S_0).$$

Typically, assume $S \approx N$ and $I > 0$ small. Let us denote

$$S_\infty = \lim_{t \to \infty} S(t)$$

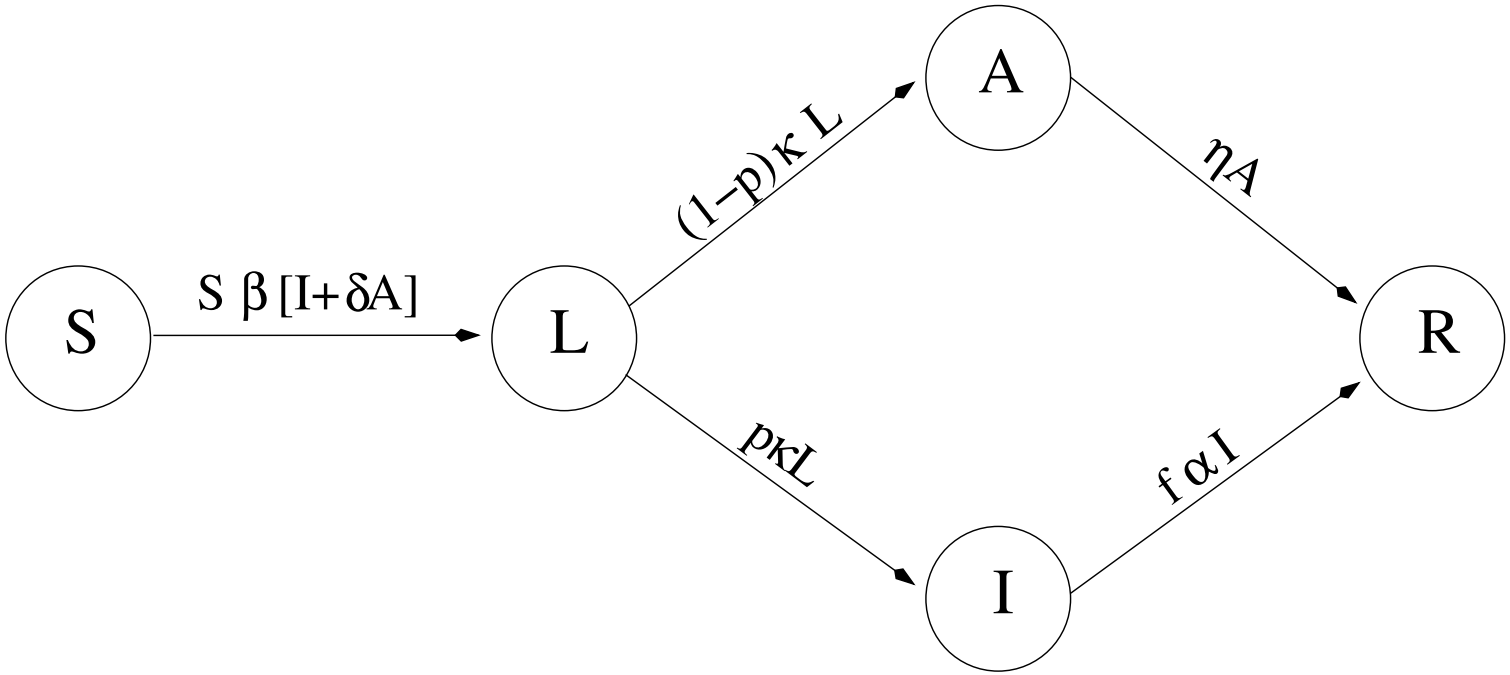
S_∞ is the *final size* of the epidemic, that is, the number of individuals who are still susceptible after the epidemic has moved through the population.

We want to find the value of S when $I \to 0$. Then

$$I_0 - \frac{\gamma}{\beta} \ln S_0 = S_\infty - \frac{\gamma}{\beta} \ln S_\infty$$

A more complicated model

Arino et al, J. Royal Society Interface, 2006: model for pandemic influenza.



$$\mathcal{R}_0 = \frac{S_0 \beta \rho}{\alpha}$$

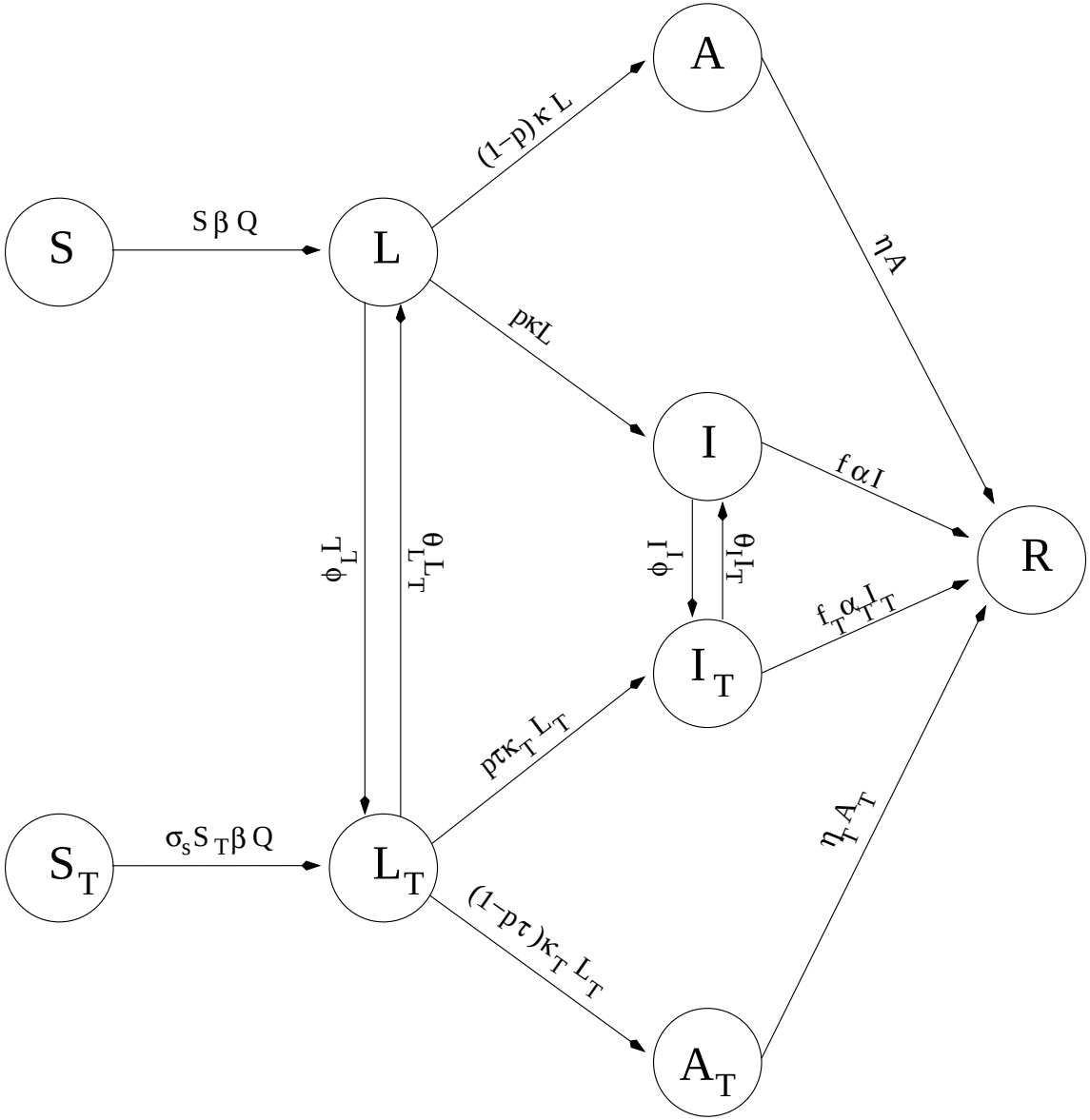
where

$$\rho = \alpha \left[\frac{p}{\alpha} + \frac{\delta(1-p)}{\eta} \right]$$

The *final size relation*, essential for determining the total number of infections $S_0 - S_\infty$, is given by

$$S_0[\ln S_0 - \ln S_\infty] = \mathcal{R}_0(S_0 - S_\infty) + \frac{\mathcal{R}_0 I_0}{\rho}.$$

The same .. with treatment



It gets messy..

$$
\begin{aligned}
S' &= -S\beta Q \\
S'_T &= -\sigma_S S_T \beta Q \\
L' &= S\beta Q - \kappa L - \varphi_L L + \theta_L L_T \\
L'_T &= \sigma_S S_T \beta Q - \kappa_T L_T + \varphi_L L - \theta_L L_T \\
I' &= p\kappa L - \alpha I - \varphi_I I + \theta_I I_T \\
I'_T &= p\tau\kappa_T L_T - \alpha_T I_T + \varphi_I I - \theta_I I_T \\
A' &= (1-p)\kappa L - \eta A \\
A'_T &= (1-p\tau)\kappa_T L_T - \eta_T A_T \\
N' &= -(1-f)\alpha I - (1-f_T)\alpha_T I_T,
\end{aligned}
\tag{4}
$$

with

$$Q = I + \delta A + \sigma_I I_T + \delta\sigma_A A_T.$$

The controlled reproduction number

For the pandemic model with treatment, the controlled reproduction number is

$$\mathcal{R}_c = (1-\gamma)\mathcal{R}_u + \gamma\mathcal{R}_v,$$

where $\mathcal{R}_u$ and $\mathcal{R}_v$ are the reproduction numbers in the case of no individuals and all individuals vaccinated, respectively.

$$\mathcal{R}_u = \frac{S_0\beta\left[(\alpha_T+\theta_I+\sigma_I\varphi_I)p\kappa(\kappa_T+\theta_L)+(\theta_I+\sigma_I(\alpha+\varphi_I))p\tau\kappa_T\varphi_L\right]}{\Delta_I\Delta_L}$$
$$+\frac{\delta S_0\beta}{\Delta_L}\left(\frac{(1-p)\kappa(\kappa_T+\theta_L)}{\eta}+\frac{\sigma_A(1-p\tau)\kappa_T\varphi_L}{\eta_T}\right)$$

and

$$\mathcal{R}_v = \frac{\sigma_S S_0\beta\left[(\alpha_T+\theta_I+\sigma_I\varphi_I)p\kappa\theta_L+(\theta_I+\sigma_I(\alpha+\varphi_I))p\tau\kappa_T(\kappa+\varphi_L)\right]}{\Delta_I\Delta_L}$$
$$+\frac{\delta\sigma_S S_0\beta}{\Delta_L}\left(\frac{(1-p)\kappa\theta_L}{\eta}+\frac{\sigma_A(1-p\tau)\kappa_T(\kappa+\varphi_L)}{\eta_T}\right)$$

where

$$\Delta_L = (\kappa+\varphi_L)(\kappa_T+\theta_L)-\varphi_L\theta_L$$
$$\Delta_I = (\alpha+\varphi_I)(\alpha_T+\theta_I)-\varphi_I\theta_I.$$

The final size relation for the treatment model is

$$S_0[\ln S_0 - \ln S_\infty] = \mathcal{R}_c(S_0 - S_\infty) + \frac{\mathcal{R}_0 I_0}{\rho_T},$$

where $\rho_T \geq \rho$ is determined by the model parameters, with $\rho_T = \rho$ if there is no treatment of infectives.

Outline

Outline

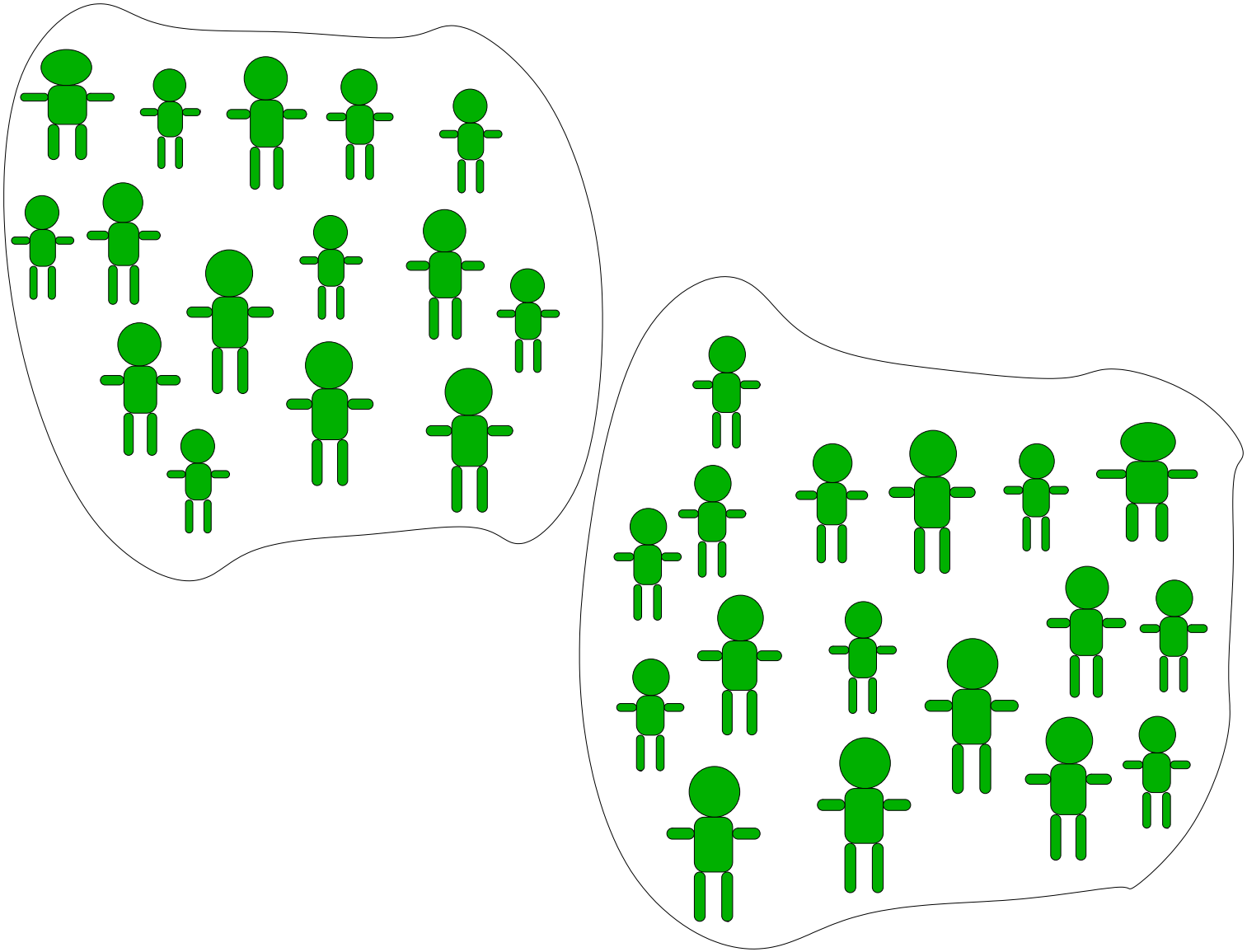
What is a patch?

A *patch* is a unit (typically geographical) within which the population is considered homogeneous.

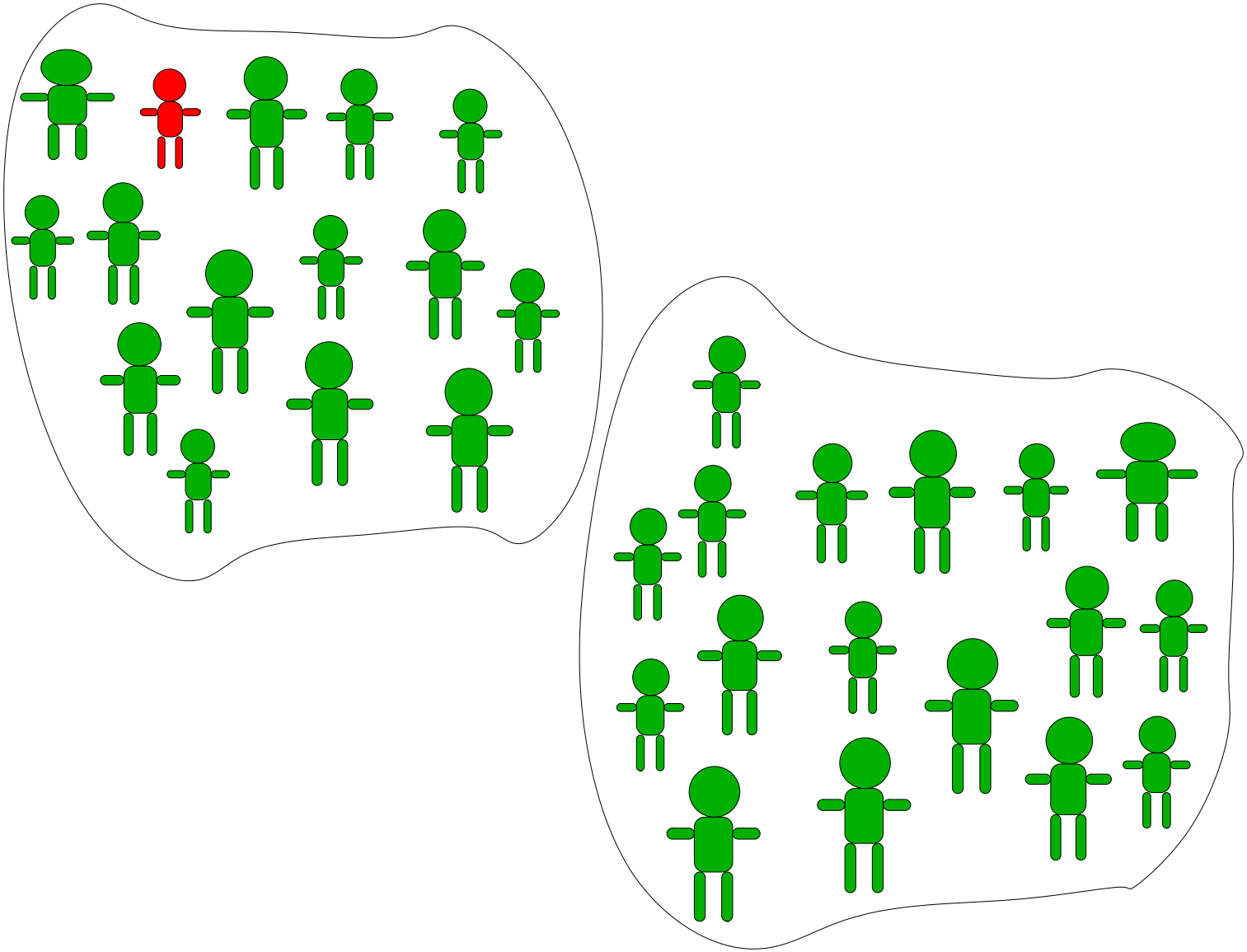
- City,
- region,
- country,
- but also, location where a given species lives (for example, forest, swamp, etc.).

Patches may or may not overlap.

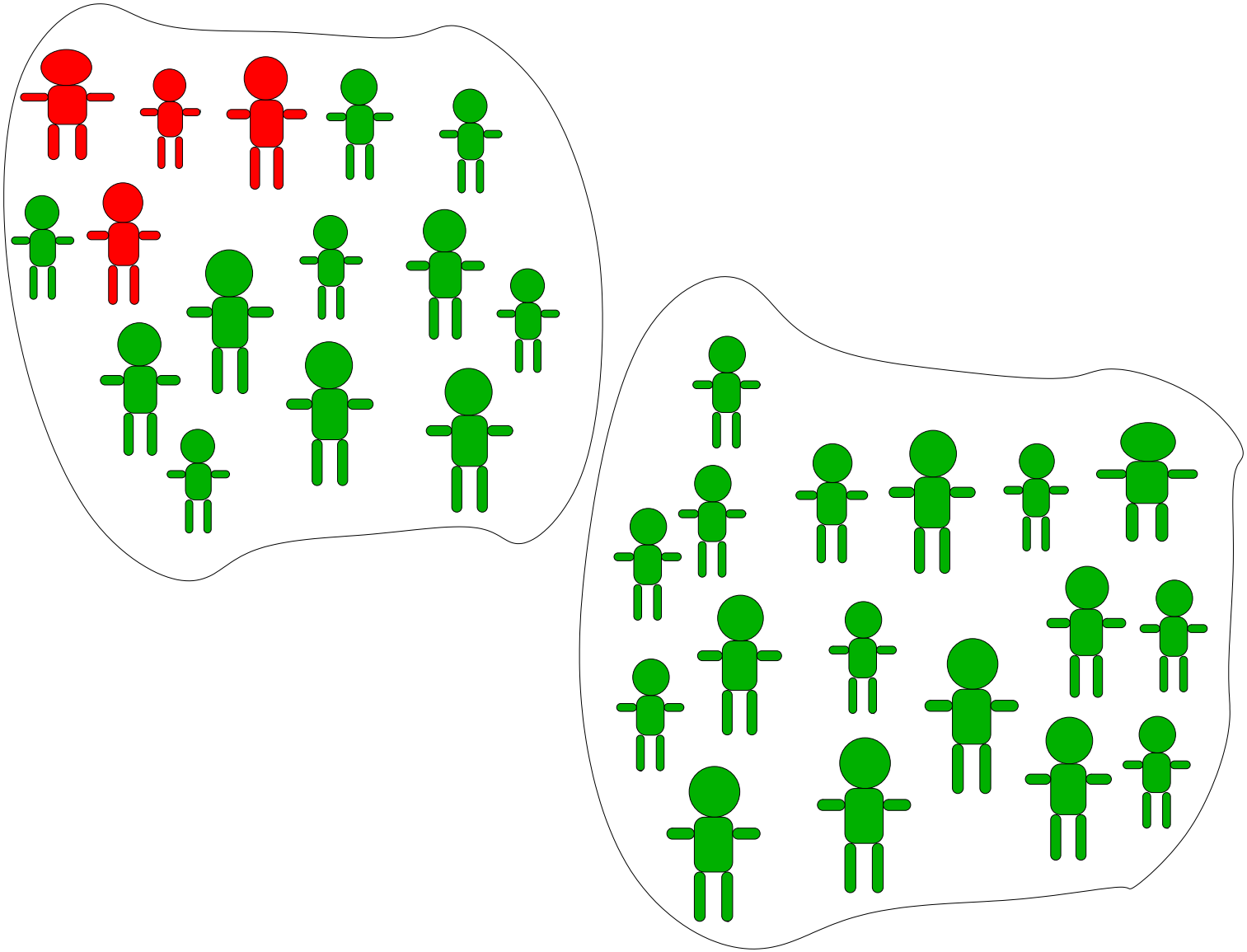
How a disease jumps between patches



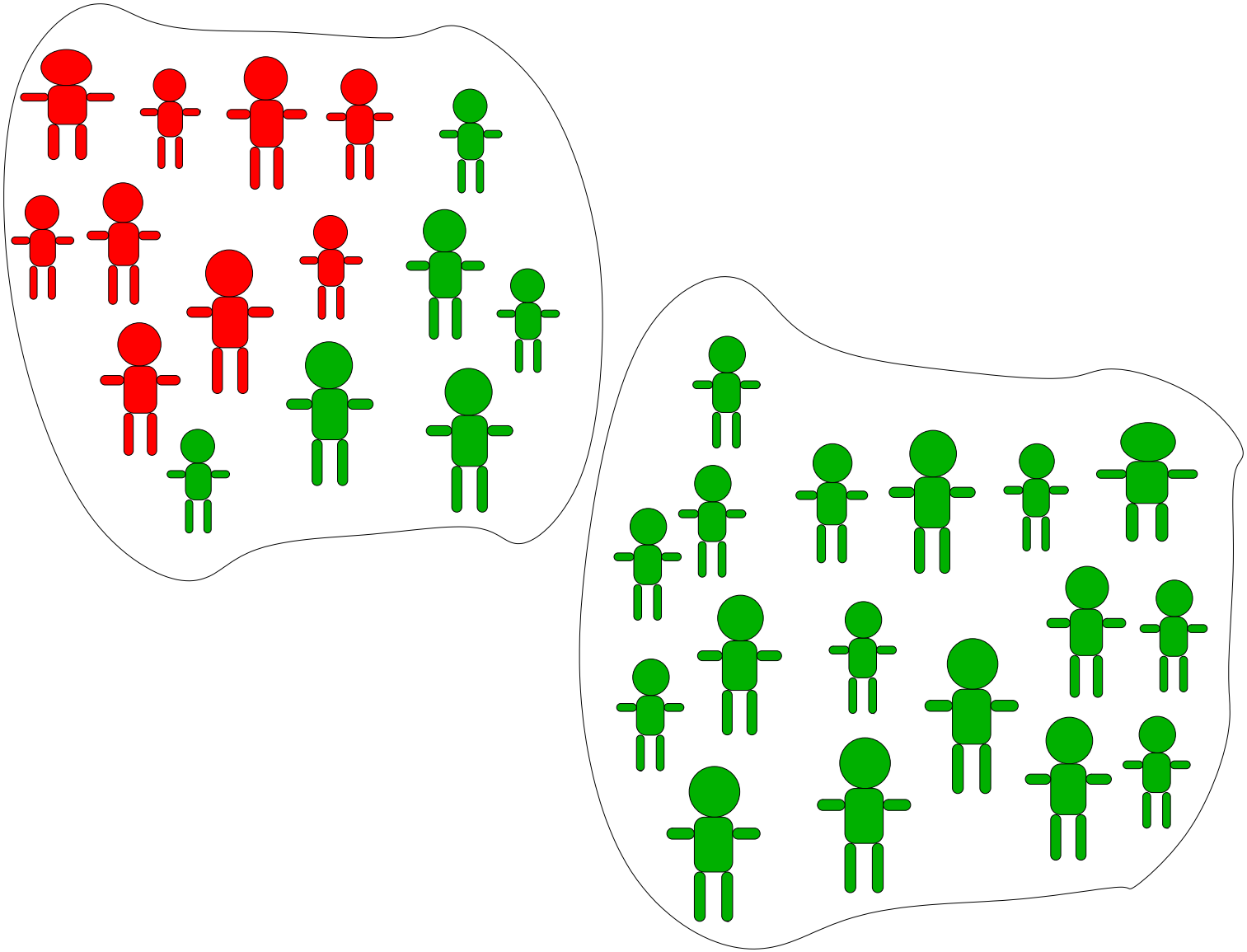
How a disease jumps between patches



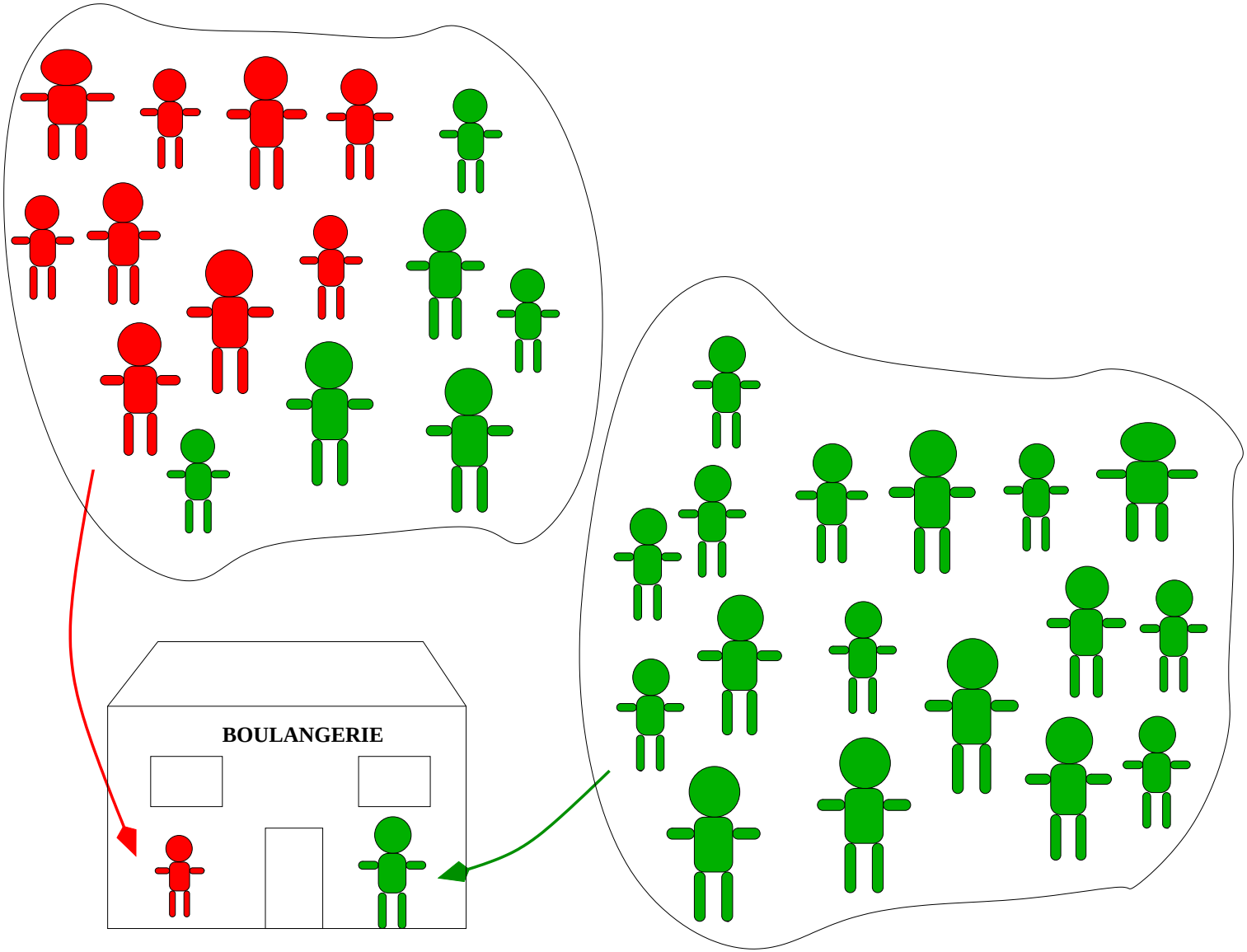
How a disease jumps between patches



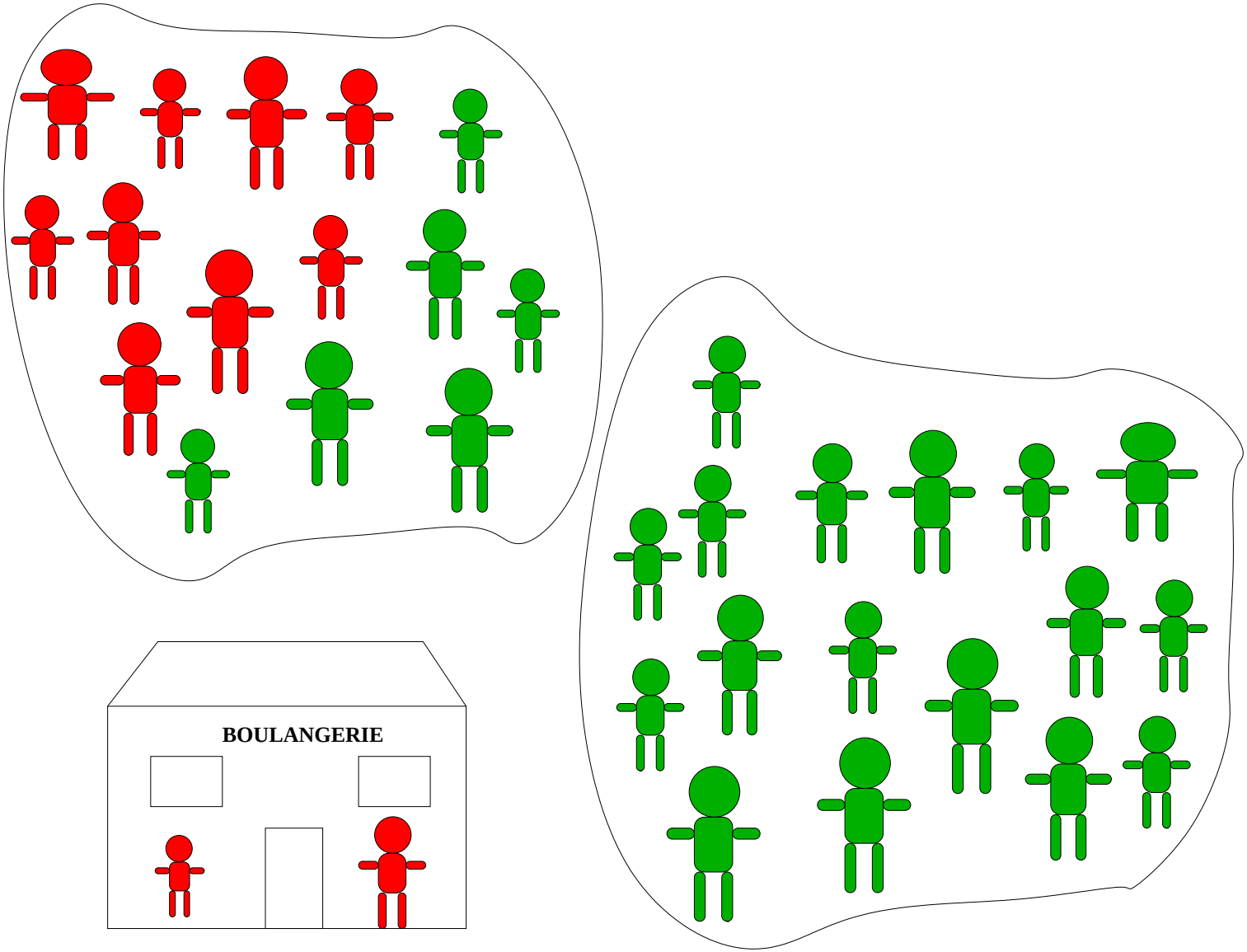
How a disease jumps between patches



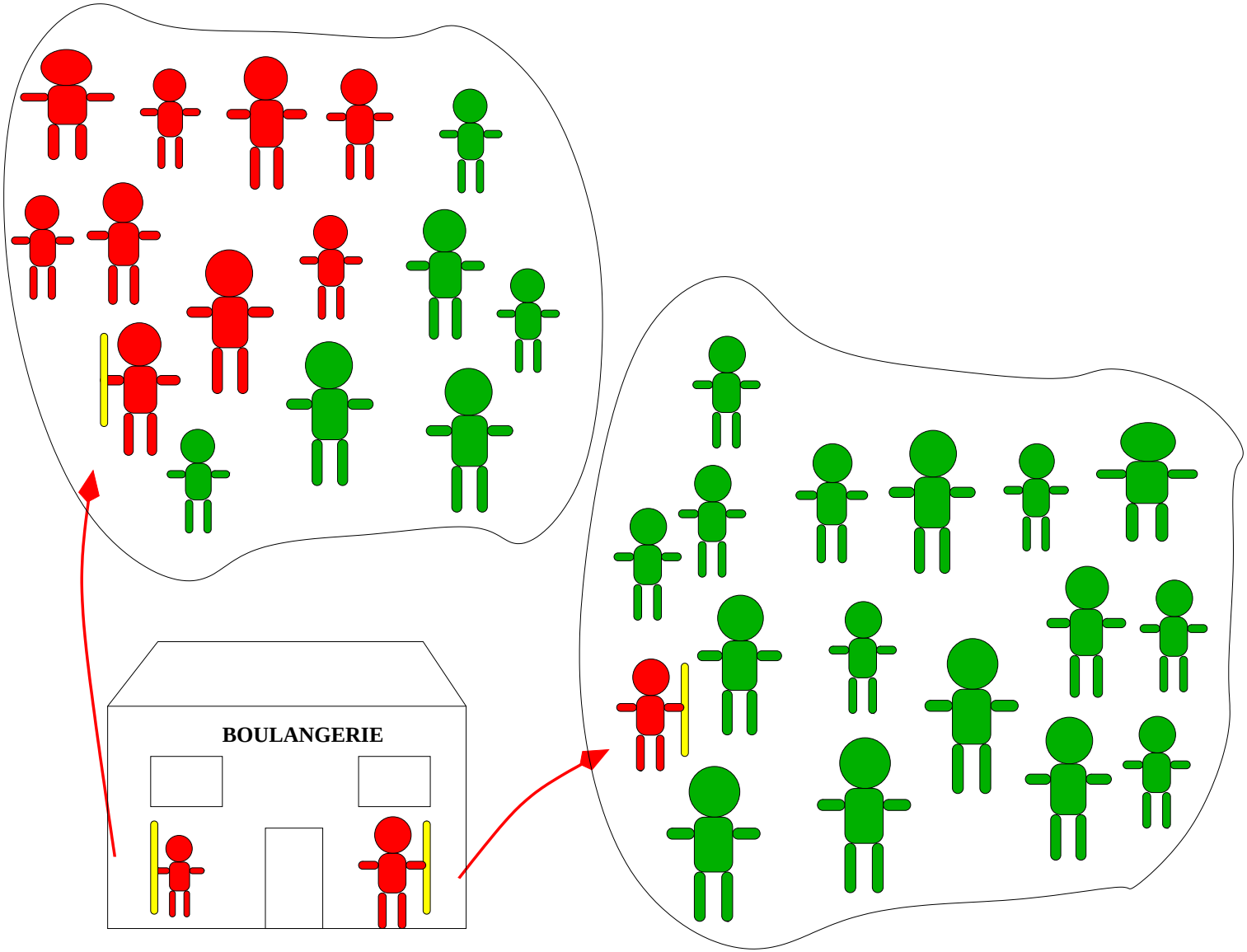
How a disease jumps between patches



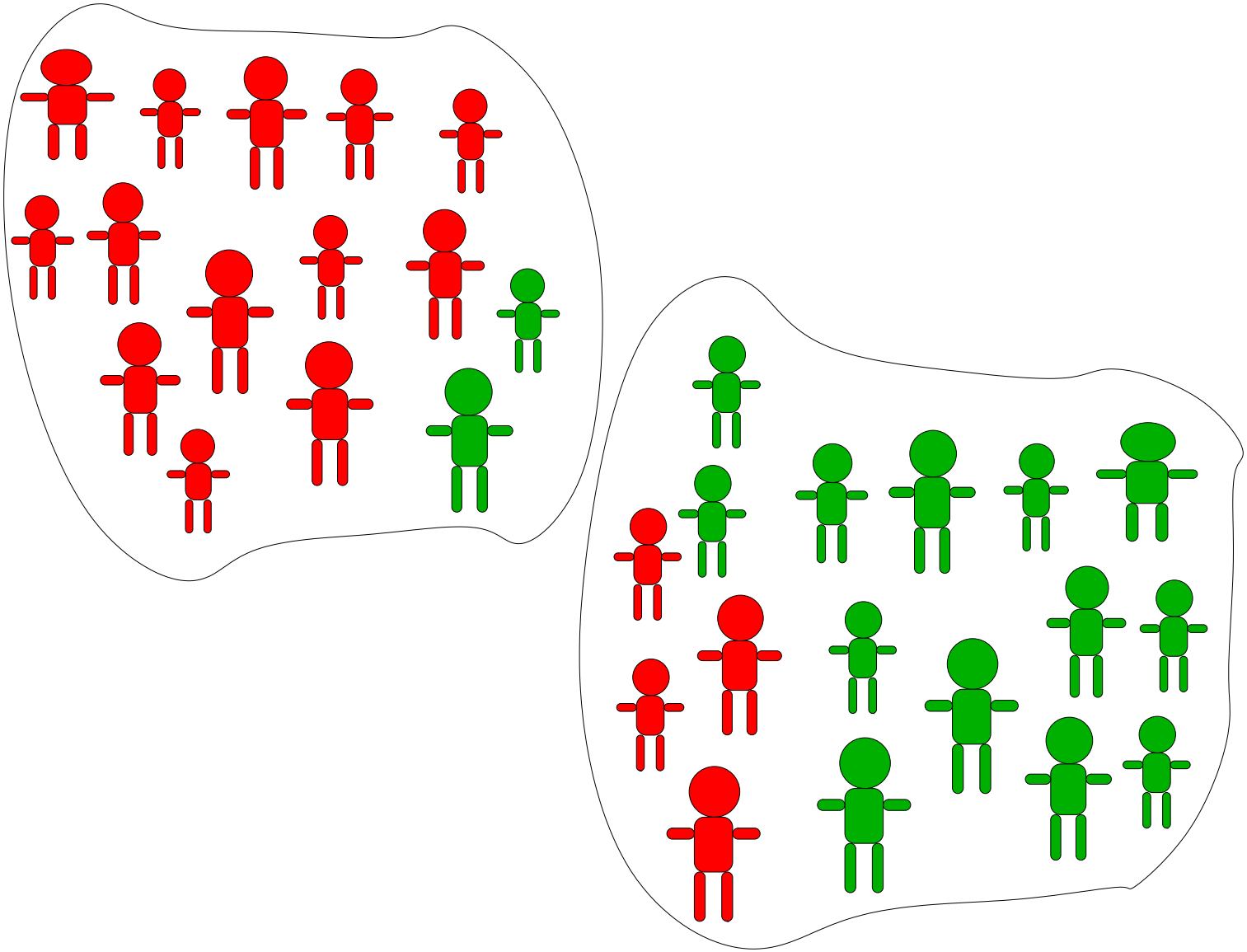
How a disease jumps between patches



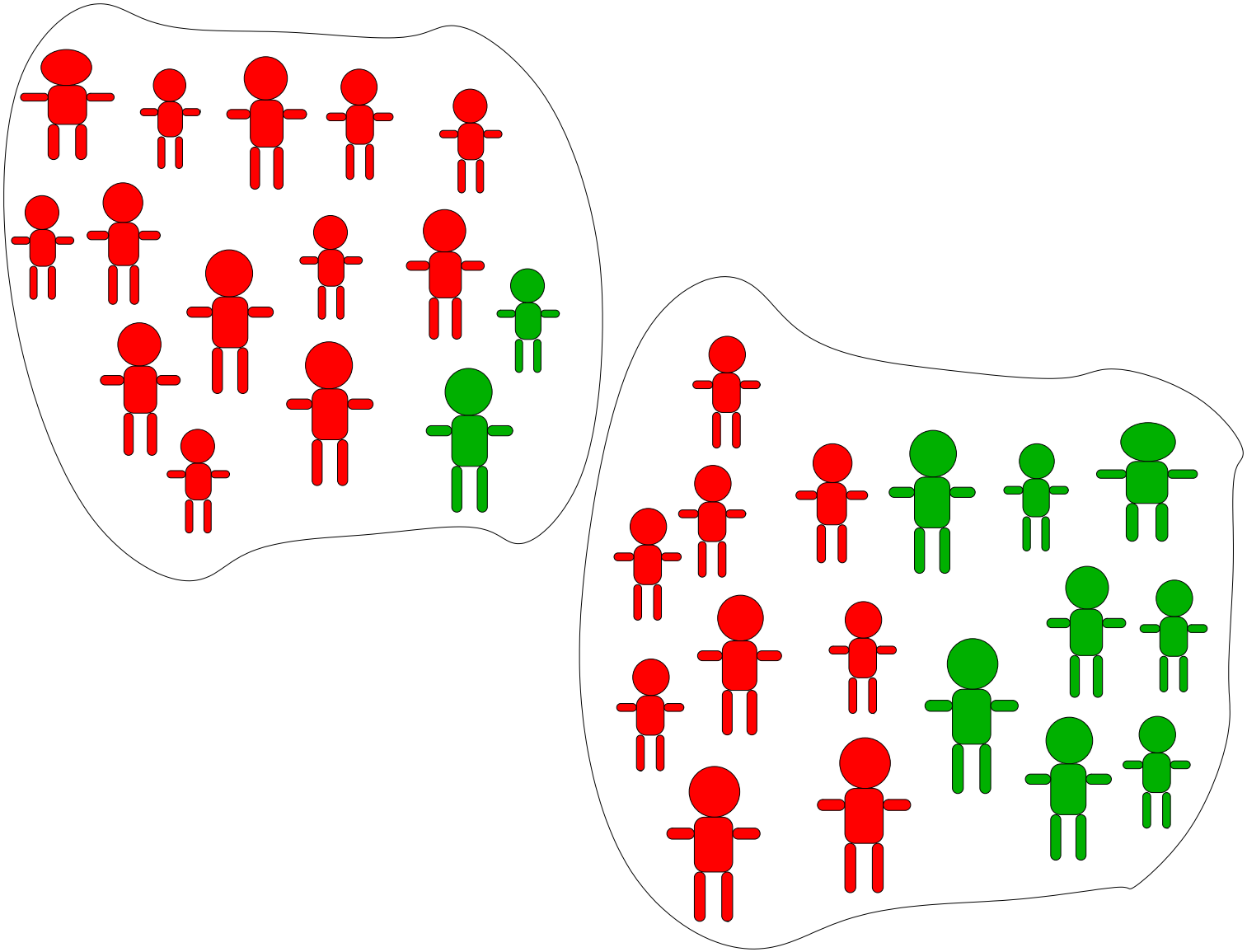
How a disease jumps between patches



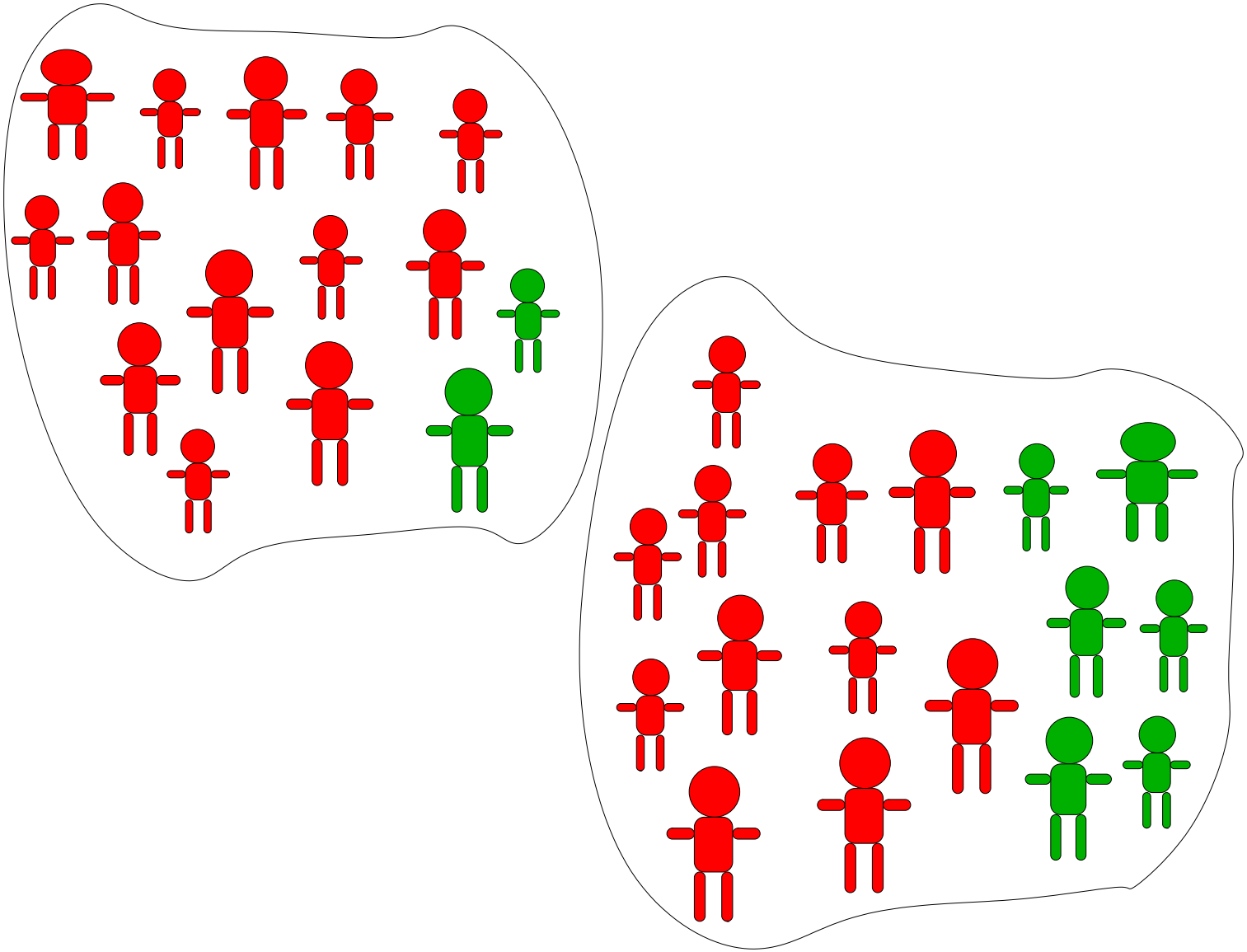
How a disease jumps between patches



How a disease jumps between patches



How a disease jumps between patches



Outline

Boulangeries as markers of population density

- France has 44,900 bakeries, or 1 bakery per 15 km^2.
- Canada has 2,710 Tim Hortons, or 1 Tim Hortons per 3684.4 km^2.

Hence, if shops are uniformly distributed in space,

- every bakery is the center of a circle of radius 2.19 km in France,
- every Tim Hortons is the center of a disk of radius 34.25 km in Canada.

Number of inhabitants "depending" on a bakery:

- 1,427 in France,
- 12,131 in Canda.

Different population densities lead to different modeling paradigms:

- in France, the patches overlap or are close $\Rightarrow$ diffusion
- in Canada they are well separated $\Rightarrow$ patches

In real life, disease propagation occurs as a mix of

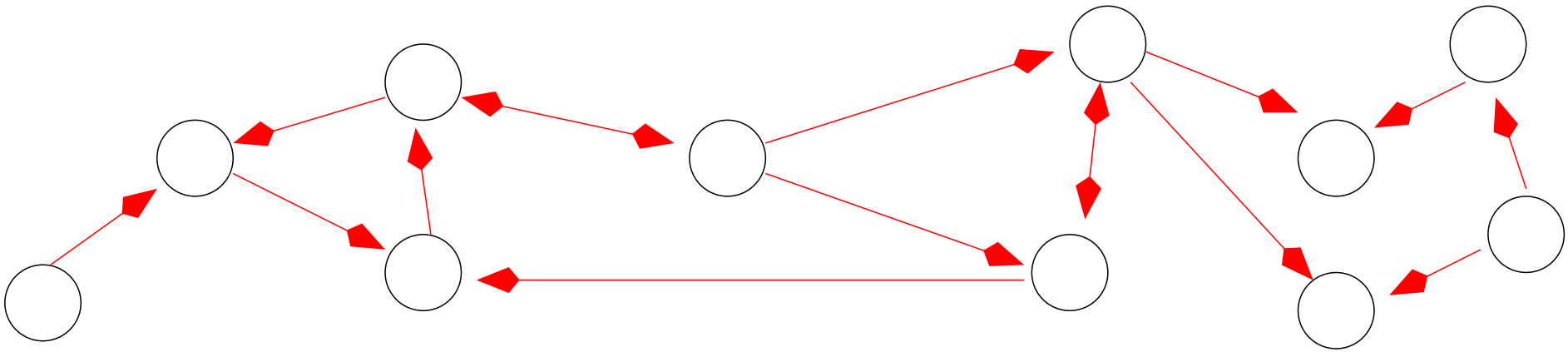
- local propagation
- long range events.

But one component of propagation at least is well captured by patches: air (by extension, high speed) transportation between large cities.

Collaboration with Kamran Khan (Division of Infectious Diseases, St. Michaels Hospital, University of Toronto) on modelling and simulation of these problems.

Models for the spread between discrete spatial locations

- Define a graph $\mathcal{G}$.
- In each vertex, use a system of ODEs to describe the dynamics of the disease.
- Connect the patches (vertices) with arcs representing the rates of movement between the patches.



$$\frac{d}{dt}S_p = \mathcal{B}_p(N_p) - \Phi_p - d_p S_p + \nu_p R_p + \sum_{q=1}^{\bar{p}} m^S_{pq} S_q - \sum_{q=1}^{\bar{p}} m^S_{qp} S_p \quad (5a)$$

$$\frac{d}{dt}E_p = \Phi_p - (\varepsilon_p + d_p) E_p + \sum_{q=1}^{\bar{p}} m^E_{pq} E_q - \sum_{q=1}^{\bar{p}} m^E_{qp} E_p \quad (5b)$$

$$\frac{d}{dt}I_p = \varepsilon_p E_p - (\gamma_p + d_p + \delta_p) I_p + \sum_{q=1}^{\bar{p}} m^I_{pq} I_q - \sum_{q=1}^{\bar{p}} m^I_{qp} I_p \quad (5c)$$

$$\frac{d}{dt}R_p = \gamma_p I_p - (\nu_p + d_p) R_p + \sum_{q=1}^{\bar{p}} m^R_{pq} R_q - \sum_{q=1}^{\bar{p}} m^R_{qp} R_p \quad (5d)$$

What can we do with this?

- Nature of equilibria.
- Value of $\mathcal{R}_0$.
- Is it possible to have mixed equilibria, with some patches without disease and other with disease?
- Are there periodic solutions?
- What is the effect of isolating one patch from the others?
- What is the effect of reducing travel from one patch to the others?
- What is the effect of treating/vaccinating one patch and not the others? (or vice versa: effect of one patch interrupting vaccination)

Outline

Another context quite similar to sparsely populated areas

Consider the airline transportation network.

- Each node is set at a given spatial location.
- Each node (roughly) has a local population attached to it.
- Communication between nodes is fast, however large the distance.

Airport data

We have

- 3,545 IATA codes for international airports
- location of these airports (GIS)
- population information on nearest urban centres

Flight data

Between these 3,545 airports, for 2003 (and 2006), we have

- total number of *seats*, in each direction, day by day.

One caveat:

- We do not have the **number** of travellers. This is confidential airline data.

WHO data on SARS is bad

K. Khan has been trying to establish the nature of cases,

- *imported* case or *locally generated* case,
- true case or misclassified case (probable SARS later typed as non SARS).

In doing so, realized that the WHO data is current as of the end of the epidemic, and in most cases,

- does not mention misdiagnosed patients,
- has little or no information on point of entry (POE).

Khan's data

Because we did not have access to the data we needed, Khan set out to get it:

- Investigations with each of the 27 (?) countries that had reported SARS cases in the WHO data.
- Obtain classification of cases as locally generated or imported.
- In case of importation, POE, date of entry (DAE), origin.

Currently, has complete information all but one country.

143 imported cases to 41 airports outside of China. (By comparison, 9 cases in total crossed borders by land, 8 in Mongolia and 1 in Russia).

Assumptions

- Airport "population" large, equal to N_i
- Incidence is mass action,

$$\beta_i \frac{S_i I_i}{N_i}$$

We want to model short term effect of disease, and are interested by initial apparition of cases, so

- $S_i \simeq N_i$.

As a consequence, incidence is

$$\beta_i I_i$$

- Since $S_i \simeq N_i$ and N_i constant, we can consider S_i as fixed.
- Short term behavior wanted, so also neglect R_i.

Truncated SEIRS model

$$\frac{d}{dt}E_i = \beta_i I_i - \varepsilon_i E_i + \mathbf{T}_i^E(t, \mathbf{E}) \tag{6a}$$

$$\frac{d}{dt}I_i = \varepsilon_i E_i - \gamma_i I_i + \mathbf{T}_i^I(t, \mathbf{I}) \tag{6b}$$

In airport i,

- $\mathbf{T}_i$ transport operators
- β_i disease transmission coefficient
- $1/\varepsilon_i$ average duration of latent phase
- $1/\gamma_i$ average duration of infectious stage

Transport operators

Can be time-dependent. Considered two types. For $X \in \{E, I\}$,

$$\mathbf{T}_i^X(t, \mathbf{X}) = \sum_{j=1}^{n} m_{ij}^X X_j - m_{ji}^X X_i \tag{7}$$

and

$$\mathbf{T}_i^X(t, \mathbf{X}) = \Delta_T(t) \times \text{dispersion kernel}, \tag{8}$$

with $\Delta_T(t) = \sum_{k=0}^{\infty} \delta(t - kT)$ Dirac comb for Dirac delta function δ.

Analysis

Basic steps

1. Well-posedness of the system.
2. Existence of disease free equilibria (DFE).
3. Computation of a reproduction number $\mathcal{R}_0$, study local asymptotic stability of DFE.
4. If DFE unique, prove global asymptotic stability when $\mathcal{R}_0 < 1$.

Additional steps

5. Existence of *mixed* equilibria, with some patches at DFE and others with disease.
6. Computation of some bounds on $\mathcal{R}_0$.

The model is a particular case of models in (Salmani & VdD 2005, Arino & VdD 2006, Arino – submitted), so all these points are easily established.

FMIPW workshop work

(Fields Mitacs Industrial Problem solving Workshop, Fields Institute, Toronto, August 2006).

- Worked with anonymised data set of 800 largest airports, because of data sharing agreements.
- Airport #7 = Hong Kong (the only one we knew), point of introduction of the disease into the airport network.

Airport data

Airport,totalVolume,intlVolume,RgVolume,numRg,
numCountry,numCity,numApt,numAirline

1,90649886,81835465,43679506,8,85,165,167,80,3
2,67876564,61530217,27171173,8,96,182,190,106,7
3,69497938,57382648,28191231,8,93,273,280,112,8
4,47443739,47192487,17473041,8,71,189,198,77,0
5,45649047,45649047,15748174,5,35,89,90,51,9
6,44799389,43005819,21144203,6,32,79,80,55,0
7,48885296,40522344,10071174,7,32,62,62,56,0
8,47993316,36565731,11525099,6,50,91,93,69,9
9,29926760,28766790,4152938,5,22,44,44,28,21

and 791 more airports.

Connection data

Connection data takes the form

```
DepartureAirport,ArrivalAirport,minStops,distance
1,2,0,346.05
1,3,0,653.79
1,4,0,369.86
1,5,0,10883.46
1,6,0,9593.88
1,7,0,9633.00
1,8,0,9550.68
1,9,1,9784.58
```

and 639,192 more lines.. Distance is distance on a sphere between the locations (even in the case of no direct connection).

Flight intensity data

We used the annual number of seats, for simplicity.

```
1 2 0 346.05 1576500
1 3 0 653.79 1183216
1 4 0 369.86 1253553
1 5 0 10883.46 1000660
1 6 0 9593.88 772567
1 7 0 9633.00 764256
1 8 0 9550.68 569643
1 9 1 9784.58 0
```

and 639,192 more lines.. The matrix is pattern-symmetric, but not symmetric. E.g.,

```
2 1 0 346.05 1576507
3 1 0 653.79 1183502
4 1 0 369.86 1253427
```

The matrices

From this information, we deduce a connection matrix

$$C = [c_{ij}]$$

with $c_{ij} = 1$ if there is a direct flight from airport j to airport i, 0 otherwise (also, $c_{ii} = 0$). Note that in this case, the matrix C is irreducible (equivalently, the induced graph is strongly connected).

Also, we have the matrix of seats,

$$S = [s_{ij}]$$

with s_{ij} the number of seats available for flights from airport j to airport i, 0 if there is no direct flight.

Finally, we have the matrix of distances,

$$D = [d_{ij}]$$

with $d_{ij} = d_{ji}$ the distance between airports i and j.

The movement matrix

In the following, chose

$$m_{ij} = \frac{s_{ij}}{\sum_{i=1}^{800} s_{ij}}$$

that, proportion of seats out of airport j that go to airport j.

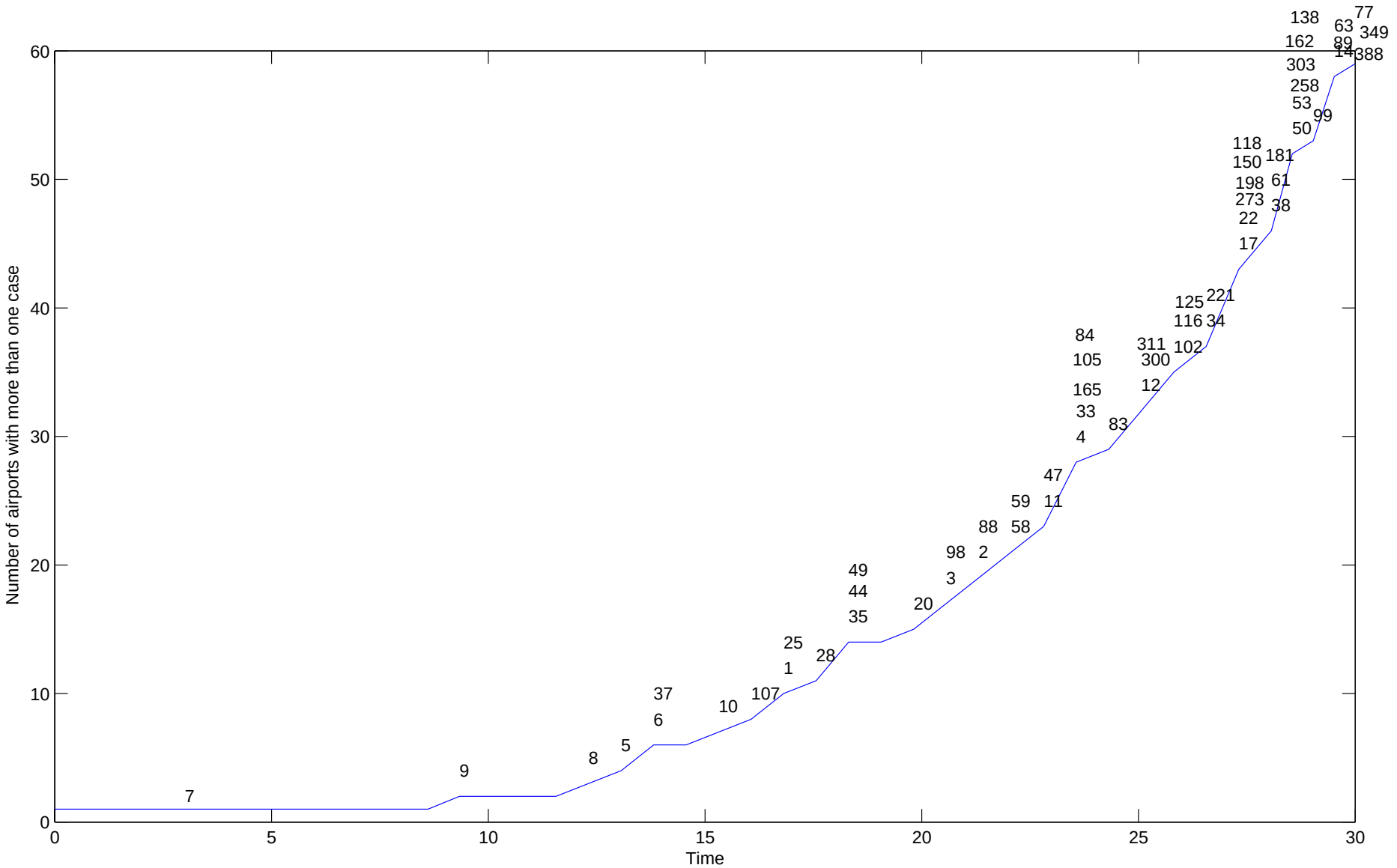
Also tried

$$m_{ij} = \frac{s_{ij}}{\sum_{i=1}^{800} s_{ij}} \frac{1}{d_{ij}^2},$$

which gave about the same results, and

$$m_{ij} = \frac{1}{d_{ij}^2}$$

which gave much worse results.


Number of airports with more than one case
Time
0
10
20
30
40
50
60
0
5
10
15
20
25
30
7
9
8
5
37
6
10
107
25
1
28
49
44
35
20
98
3
2
88
58
59
11
47
84
105
165
33
4
83
311
300
12
102
125
116
34
221
17
22
273
198
150
118
181
61
38
138
162
303
258
53
50
99
63
89
14
77
349
388

In the previous graph, an airport i is *active* if $I_i \geq 1$.

Comparing to number of cases per airport, which was provided, we got a relatively good agreement: over 70% of the airports that activate in the first 45 days of simulation had cases in "real life".

Used the same type of approach with the stochastic approach, but chose one airport at random (with uniform distribution), to which the cases were sent.

Parameter values used:

- Transmission coefficient $\beta = 0.5$,
- mean incubation period $1/\alpha = 7$ days,
- mean sojourn time in the infectious stage $1/\gamma = 21$ days.

Conference in June

Second International Conference of the French-speaking Society for Theoretical Biology.
Winnipeg, June 4-6.
`http://euromedbiomath.free.fr/w2007`