Contagious diseases stimulate genetic diversity

An analysis of the evolutionary effects of contagious diseases using the mathematical SIS model

Author: Henk-Jan van der Molen – 2023-03-05

Summary

Since the 1960's science has been baffled by the biodiversity paradox: the biodiversity we see everywhere in the food chain seems incompatible with Darwin's principle of *the survival of the fittest*. Because the fittest species have the best chance to pass their genes to next generations, they should become dominant over their less adaptive competitors.
The purpose of this research is to strengthen the evolutionary link between contagious diseases and the diversity of species. Using the mathematical Susceptible-Infected-Susceptible (SIS) model, contagious diseases prove to be a significant factor in the evolutionary creation of diversity. This article supports the "Kill the Winner" (KtW) hypothesis with contagious diseases as actors.

Key words: contagious diseases, SIS model, diversity, evolution, KtW, herd immunity, MacArthur-Wilson.

1. Introduction

Darwin's theory of evolution is often summarized as: "*It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is most adaptable to change*". This "*survival of the fittest*" principle suggests that the most adaptive species will become dominant. Based on this principle, Hutchinson presented in 1961 the so-called Paradox of the Plankton: why do so many species coexist that feed on the same nutrients, instead of one species out-competing all the others[1]? This latter expectation has been formulated as the so-called competitive exclusion principle[2]. The Plankton Paradox has been generalized as the biodiversity paradox.
Recent research is making progress to reconcile this paradox with Darwin's principle, stating that populations with a large additive genetic variance have a better change to adapt to environmental changes[3].

This article investigates the evolutionary effects of contagious diseases on the genetic diversity of a population. It extends prior research that used simulation to generate populations with 1, 2 or 100 genotypes[4]. The researchers used the Susceptible-Infected-Recovered (SIR) model to calculate the total number of infected individuals in each sub-population as a measure for the impact of epidemics. The conclusion of the researchers is that populations with low genetic diversity should endure epidemics less frequently. But once they do, the impact of the epidemic is more severe.

The analysis in this article uses the Susceptible-Infected-Susceptible (SIS) model, because its steady state can be calculated when the population contains more than one genotype, even when the system of differential equations cannot be solved[5]. The value of the steady state is used as a measure for impact of epidemics. As this analysis will show, the diversity of the population is a significant factor on the impact of epidemics, which determines the herd immunity threshold[6] for that disease.

The structure of the article is as follows. Section 2 compares the standard Predator-Prey model to the SIS model for the spread of contagious diseases, to spot similarities and differences. This involves relating predators to infected individuals and prey to susceptible individuals.

The next section describes evolutionary events linked to high mortality, which follow a Poisson density function over time. For contagious diseases this translates as: in due time an epidemic with a pathogen will occur, which will affect the number of susceptible species. When the circumstances of species increase the impact of (repeating) epidemics, this can have an evolutionary effect on the diversity of ecosystems.
Using the homogeneous and heterogeneous SIS model, sections 4 and 5 show that less diverse populations are more impacted by contagious diseases than populations with more genetic diversity. Because the SIS model describes the exponential behavior of contagious diseases, the results of this analysis are significant in the same order of magnitude.

Section 6 considers the design choices for the simulation. Individual nodes in two equally sized sub-populations are simulated, which have a different susceptibility for a particular infectious disease.

Section 7 summarizes the simulation results and shows that they are consistent with the predictions of the heterogeneous SIS model.

The final section summarizes the conclusions and points to areas for further research.

Annex A contains the mathematical calculations that show that diversity has a significant, decreasing effect on the impact of contagious diseases. The Python simulation script is listed in Annex B[7].

2. From *Predator–Prey* to the *Susceptible–Infected–Susceptible* model

The standard mathematical Lotka–Volterra Predator–Prey model consists of two differential equations to describe the change over time of the size of both the predator and prey populations connected in a food chain[8]. Each species will experience evolutionary pressure, because they must adapt to successful mutations of its predators, competitors and preys in the food chain network. In a study aimed at finding a cause for the high diversity in existing ecosystems, the mechanism of co-evolution of predator-prey species proved to keep the diversity level stable[9]. Specifically, a significant inflow of genetically different members is necessary to avoid extinction of the population. The study introduces mutation to create genetic diversity, using the same rate of mutation for both predator and prey.

There are many similarities between disease ecology and macro ecology[10]. Focusing on disease transmission, diseases can be seen as a parasitic predator in an infected host preying on susceptible species. Despite the similarities, a pathogen as predator is substantially different from its multi-cell prey. For example, the rate of mutation for each genome is roughly the same (less than 1 mutation per billion copies[11]), but a bacterium or virus reproduces much faster than e.g. mammals. In the same time period pathogens will have accumulated much more mutations then the higher life forms they prey on.

Pathogens can evolve when in their environment conditions change that are related to their survival. Phages as bacterial predators exchange genetic material and thus, stimulate diversity of bacterial species and contribute to an accelerated evolution[12]. Other studies have demonstrated that vaccines causes pathogens to evolve[13] [14] [15] [16] [17]. Novel adaptations can change the genome of pathogens significantly and lead to rapid evolution[18]. This relation goes both ways: in areas where malaria is endemic, the frequency of mutations in red blood cells remains high[19]. Nature is very flexible to explore new possibilities, even multi-year periodicity in epidemics occurs[20].

Research revealed that tight interactions between microbial species could lead to ecological disasters [21] [22]. Based on this result, one could suppose that the balance between various pathogens is an important stabilizing factor, but this probably is an oversimplification. Researchers tried to verify if reduced bacterial diversity (dysbiosis) is a cause of chronic diseases, since this could lead to new treatments for infected patients.

A study in 2012 revealed that Chronic Sinusitis patients have dysbiosis and that *Lactobacillus sakei* could be a cure for some forms of sinusitis[23]. In 2013 a study suggested that inflammatory diseases reduce the diversity of the ecological system of the human gut micro-biome[24]. A study in 2014 on Inflammatory Bowel Disease concludes that it is not yet clear if dysbiosis is a cause or an effect of this disease[25]. An overview of metagenomic studies concludes in 2016 that science is still at the starting phase of this research and that current methods are far from perfect. A better understanding of the effect of changes to microbiota depends on the development of better methods and techniques[26]. Another study on Chronic Inflammatory Lung Diseases in 2018 states that information on host-microbiota interactions is still very limited, so that causes and effects cannot yet be determined[27]. A study of Crohn's disease in 2018 suggests that dysbiosis is a consequence of this disease[28]. Research in 2020 and 2021 has determined that medical treatment sometimes causes dysbiosis, but that due to differences in human lifestyle and genetics, dysbiosis is not always associated with autoimmune, inflammatory and pernicious diseases [29] [30].

3. Evolutionary Survival: avoid Extinction

On the survival of species, Robin Dunbar stated[31]: *Species change through the gradual failure of some lineages to reproduce, resulting in a subtle but steady drift in the species' genetic make-up towards that of lineages that are more successful. Although in most cases these processes are quite slow, an entire species can go extinct catastrophically if none of its various lineages can reproduce fast enough to offset unusually high levels of mortality, there is always a steady trickle of such extinctions over time – there have been literally dozens within our own lineage during the course of our six-million-year evolutionary history. Sometimes, however, environmental conditions conspire to produce a rapid burst of extinctions.*

This echoes the insights provided by the MacArthur-Wilson theory of Island Bio-geography[32]. In this theory the area of the island is one of the key factors that determine how many different species that island can support. One of the arguments that smaller islands have less biodiversity is that smaller populations are

more prone to go extinct. So, places with high diversity have a small influx of new species and vice versa. The same result was witnessed in bacterial communities: the more genetically diverse a microbial community becomes, the more likely antagonistic interactions are[33]. Another study concluded that there is an inverse relationship between soil microbial diversity and survival of invading species[34].

The theory of Island Bio-geography has inspired a lot of criticism, but research has indeed provided evidence that links the island size and the stable number of species on that island. Furthermore, the assumption relating population size to extinction likelihood is very well supported by data[35]. This could indicate that in the short term, mono-cultures have a better chance to survive, because populations of species less successful in the competition are (much) smaller and therefore, more likely to go extinct.

Generally speaking, an organism can use available resources to increase growth, for reproduction or to counteract stress (e.g. predators, parasites, decay, unfavorable environmental conditions, or competitive ability)[36]. Resistance to contagious diseases has evolutionary benefits for a species, but on the long term it seems that higher lifeforms cannot avoid the risk of epidemics, no matter how fit and balanced they are compared to their competitors and predators in the food chain[37].

Because pathogens tend to specialize on specific hosts, they are more likely to reduce the populations of common species and therefore maintain species diversity[38]. Because virulent diseases are by far the most effective eradicator of species[39], contagious diseases have an effect on the evolution of species, independent of the regular Predator – Prey interaction, see Figure 1 [40].

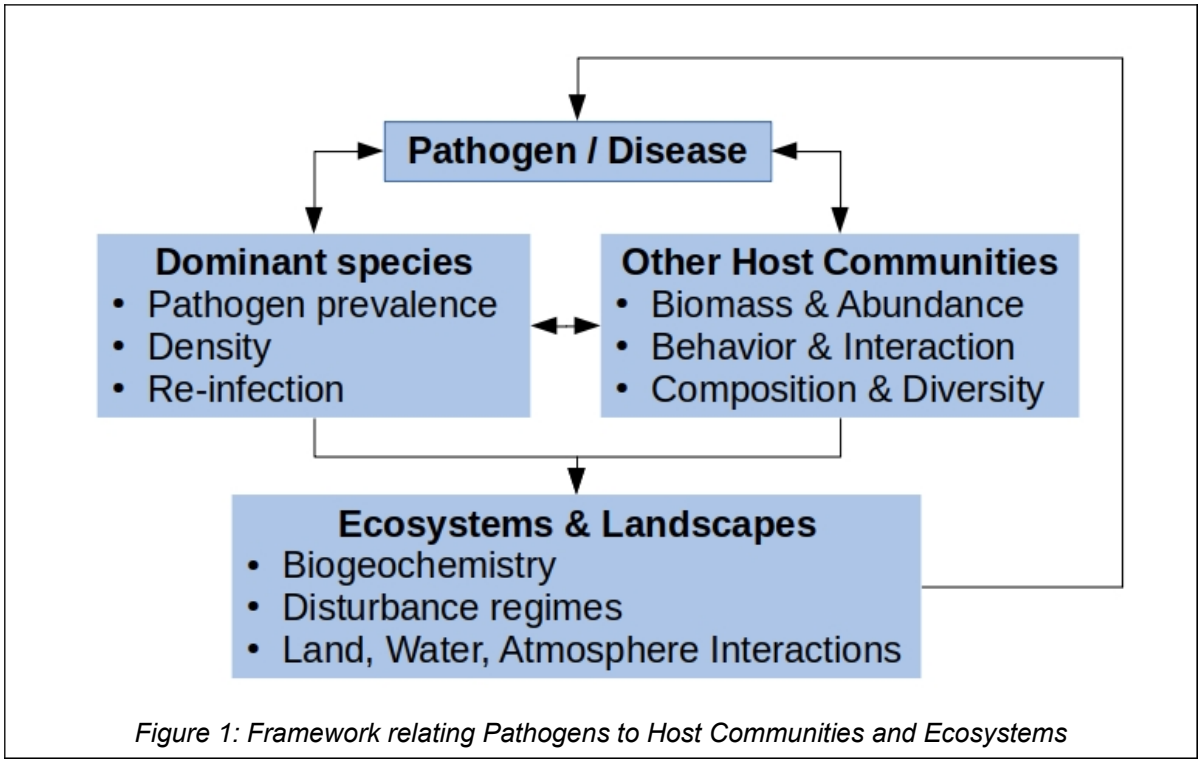

Figure 1: Framework relating Pathogens to Host Communities and Ecosystems

A study has divided pathogens in three classes, see Table 1 for examples[41]. The characteristics of class A are typical of flu-like diseases with high diversity and a low Basic Reproduction Number (R_0), which translates to a low Herd Immunity threshold (HIT) for epidemics. Class B is typical for STD's with a lower R_0 (and a corresponding lower HIT for epidemics). Because of the high diversity of Hepatitis B, several variants develop over time[37], so no distinctive HIT can be determined. Pathogens in class C are characterized by a high R_0 and low diversity, so that vaccines provide a lifelong protection.

By changing their genetic code, contagious diseases in class A are able to exploit new genetic susceptibilities to infect species. Murphy's law[42] predicts that a lethal epidemic will occur during a long enough period of time. Some contagious diseases have a combined effect, because the immune system can be impacted by previous, other infections. Research has found support for the hypothesis that an infection with measles decreases the resistance to other infectious diseases up to 5 years[43].

Another source of resistance is diversity in and between species. The following sections contain the analysis of the effect of diversity on the impact of epidemics using the mathematical SIS model.

Infection	Class	Diversity	R_0	HIT(%)
Influenza H1N1[44]	A	High	1.3 - 2	23 - 50
SARS[45]	A	High	2 - 4	50 - 75
STD: HPV[46]	B	High	0.5	0
STD: Hepatitis B[47]	B	High	?	?
Mumps[48]	C	Low	4 - 7	75 - 86
Polio[12]	C	Low	5 - 7	80 - 86
Rubella[12]	C	Low	6 - 7	83 - 86
Pertussis[12]	C	Low	12 - 17	92 - 94
Measles[12]	C	Low	12 - 18	83 - 94

Table 1: Classes of contagious diseases with Basic Reproduction Number (R_0) and corresponding Herd Immunity Threshold (HIT)

4. Diversity in the homogeneous SIS Model

The SIS model describes how contagious diseases spread in a population assuming that each node in the network is in one of two states: healthy (and susceptible to infection), or infected. Once a node is infected, it can immediately spread the disease to other susceptible nodes. Infected nodes can 'heal', but then immediately become susceptible again for the same disease. So the SIS model is only fit for diseases like tuberculosis or the Class A diseases in Table 1 that do not create permanent immunity after infection.

The SIS model contains the parameters β^*k and δ, with specific values for every combination of a contagious disease and susceptible species. The β^*k parameter expresses the ease with which an infected host can transfer the disease to a susceptible one. The δ parameter expresses how soon an infected host can recover from the disease. In its simplest form, both the β^*k and δ parameters in the SIS model are constant and uniform for the whole population. Assuming $\beta^*k > \delta$, solving the differential equation produces the characteristic S-curve of the logistic function, showing how the disease spreads over time.

Because the time scale is unimportant, the Basic Reproduction Number R_0 (= β^*k / δ) is used to simplify calculations. The parameter R_0 tells on average how many other individuals an infected individual will infect. This simplification is justified since the analysis focuses on the steady state of the model, which only depends on the value of R_0. The SIS model predicts an epidemic when the $R_0 > 1$ of a contagious disease.

The homogeneous SIS model only permits situations were one sub-population is susceptible ($R_1 > 0$ for fraction n_1 of the population) and the other sub-population is not ($R_2 = 0$ for fraction $n_2 = 1 - n_1$). Diversity in this context starts off with a mutation that introduces immunity for the disease in the population. If that immunity increases the changes for survival of the species, the immune sub-population will increase over time. When {n_1 or $n_2 = 0$} this constitutes a mono-culture by definition, so the maximum values of n_1 and n_2 are set at 0.5. To avoid bias when increasing the immune fraction in the population, the value of R_0 for the population as a whole must remain constant. If the genetic drift would increase or decrease the value R_0 for the population, the steady state of the model would follow, making the analysis unfair. Keeping the R_0 constant for the population as a whole, leads to a diversity factor which only depends on the value of n_2. The restriction on the maximum value of n_2 prevents that the value of R_1 reaches infinity.

The analysis shows that when diversity increases, the infected fraction of the population decreases monotonously. This result is verified in the next section with the heterogeneous SIS model, which allows each sub-population to have different parameters (β^*k, δ) for a disease all sub-populations are susceptible.

5. Diversity in the heterogeneous SIS Model

In this analysis the least complex heterogeneous SIS model is used, with two sub-populations. This enables a pair wise comparison between any pair of subspecies that share a susceptibility for a contagious disease. Similar to the homogeneous model, an epidemic will occur when $R_0 > 1$. The R_0 for the whole population can be determined by adding the weighed R_0 factors of both sub-populations, but only if complete symmetry is

assumed[5]. Each node must see the same fraction of nodes from every sub-population, so a fraction n_1 of neighbors from sub-population #1 and a fraction n_2 of neighbors from sub-population #2, and so on.

Based on the Gini-Simpson index[49] the population is set up as neutral as possible, with two equally sized sub-populations to analyze the effects of diversity ($n_1 = n_2 = 0.5$). The analysis assumes a genetic drift that increases the gap between the R_0 of both sub-populations, while the R_0 for the population as a whole remains constant. This can be seen as a redistribution of *susceptibleness* elements in the population.

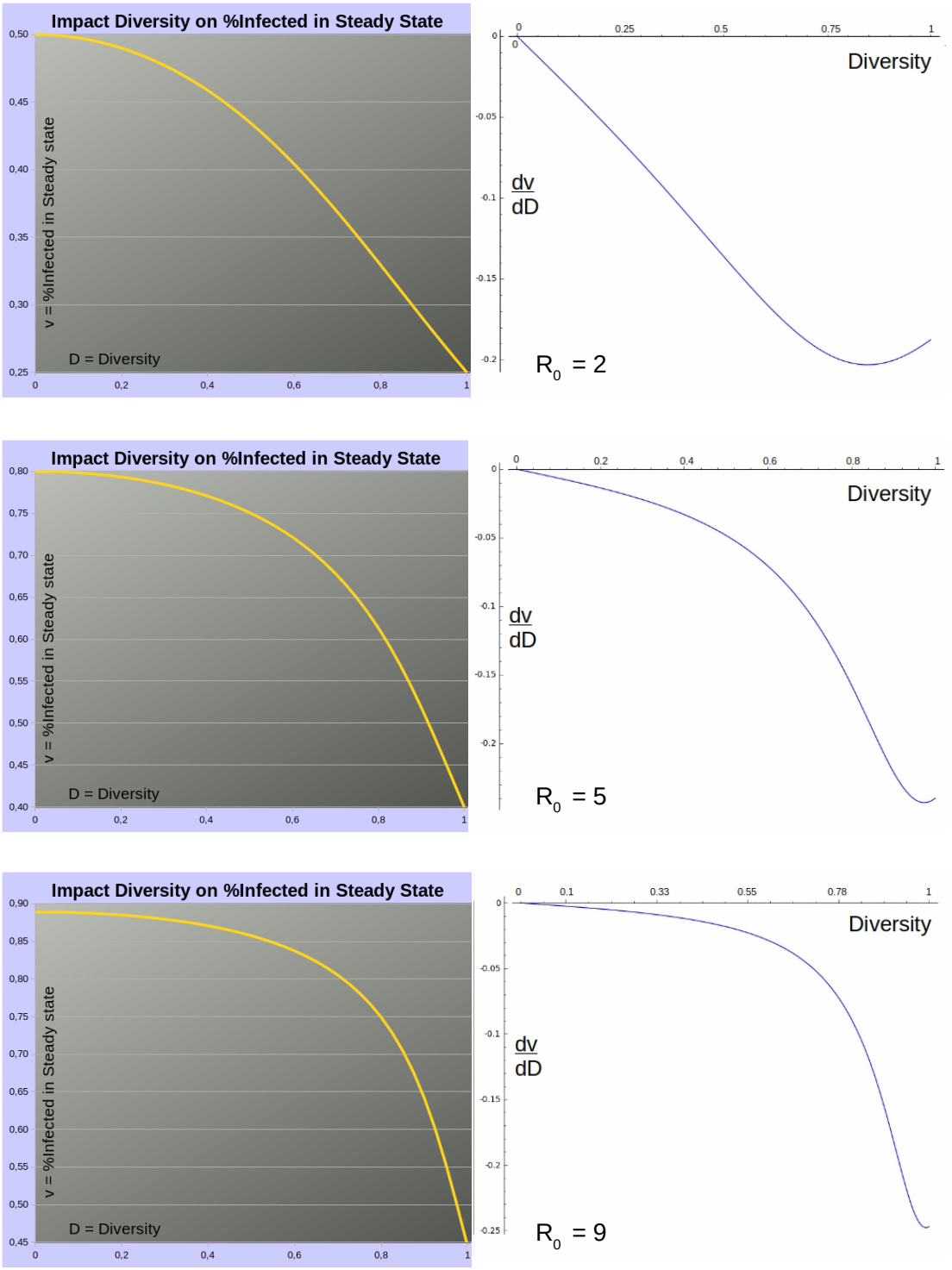

Figure 2. Steady State v (left) en dv/dD (right) of the heterogeneous SIS Model with $n_1 = n_2 = 0.5$, $R_0 = 2$ (top), 5 (middle), 9 (bottom)

As an example, assume a value of $R_0 = 2$ for the whole population. For a mono-culture the diversity is zero, so for both sub-populations the $R_0 = 2$. Now the effect of diversity can be analyzed by increasing the distance between the R_0 values (indicated by R_1, R_2) of both sub-populations, keeping the $R_0 = 2$ for the population as a whole. So when $R_1 = 3$ for the first sub-population, then $R_2 = 1$ for the second sub-population. Maximum diversity is reached when $R_1 = 4$ for the first sub-population and $R_2 = 0$ for the other sub-population.

In steady state, the influence of diversity on the infected fraction (v) is shown in Figure 2 by the graphs on the left. The derivative for diversity of (v) shows that an increasing diversity has an consistent negative impact on the fraction of infections in steady state[50]. So, for a diversifying population the impact of the epidemic decreases. Therefore, epidemics have more impact on populations with dominant species. For non-dominant species the impact of epidemics and therefore the HIT is lower, improving their chance of survival.

For almost the entire diversity interval $\{0 \leq D \leq 1\}$ it holds that the *more diverse* a population becomes, the *faster* the impact of pathogens decreases. Therefore, diversity presents a slippery slope for the impact of epidemics. On an evolutionary time scale, this results in populations that become more diverse. The subtle force of contagious diseases opposes the mechanism to produce a “winner takes all” genome that scores the best in the food chain network. The next two sections describe the setup and results of simulations to test the predictions of the mathematical SIS model in Annex A.

6. Designing the simulation of the heterogeneous SIS model

The simulation is set up with 500 individual nodes in a regular graph, each with a fixed number of connections (k) with other nodes in an undirected network. The population is divided in two sub-populations of 250 nodes each ($n_1 = n_2 = 0.5$), maintaining complete symmetry: each node connects to the same number of nodes from every sub-population.

A simulation regularly includes 750 *linear* time steps, because the analysis only focuses on the steady state. To reach the steady state as fast as possible, β and δ are maximized under the constraint that the chances β.k, δ must be ≤ 1. Each simulation is repeated 50 times, because this decreases the standard deviation and increases the chance that repeated calculations will show similar results. See Figure 3 for a simulation run using the script in annex B, where $R_0 = 4$ and $D = 1$.

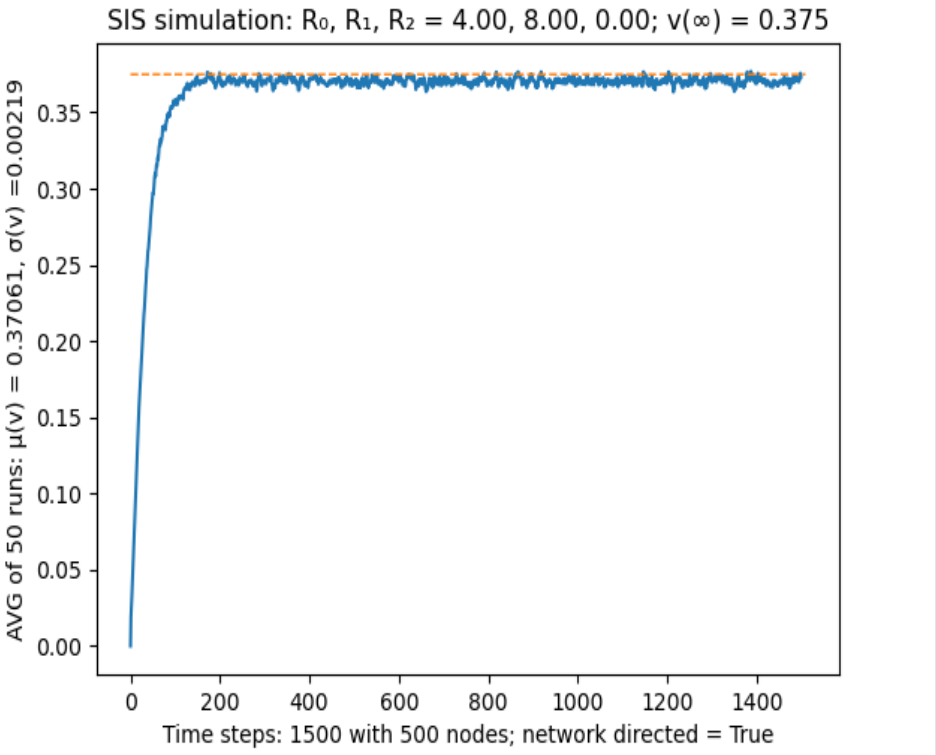

Figure 3. Result of a simulation run 50 times

To initialize the simulation, 2% randomly chosen nodes are infected (node value => 1), the remaining 98% of the nodes are made susceptible (value = 0). It is easy to verify that the initially infected fraction (2% = 10 nodes) is enough to trigger an epidemic – the impact of initially seeding more infections is small. Simulations that start with 100% of the population infected, lead to a comparable mean steady state value and standard deviation[51]. Note that the initial "seeding" of infected nodes is random, so the particular numbers of nodes do not matter. What matters is that the network is symmetrical and whether its edges are (un)directed.

In a single time step the condition of all nodes is evaluated within their context. For instance, in every time step an infected node has a chance δ to "cure" and become susceptible again. This rule is straightforward to implement: if a random sample from the uniform {0..1} distribution is less than δ, the node is cured. For every time step, a susceptible node has a chance β to become infected when precisely one of its contacts is infected. With multiple infected contacts the infection chance can be calculated with the following statistical formula: 1 – (chance of remaining susceptible). This chance can be written as $1 – (1 – β)$ with 1 infected contact and as $1 – (1 – β)^3$ with 3 infected contacts.

To get a feel for the predicted values, the parameters are given values: β = 0.125; k = 8; the fraction of infected nodes (v) in steady state = 0.4. This means that on average in steady state every susceptible node links to k*v = 3.2 infected nodes. Because 3.2 is not an integer, this leads in steady state to 80% of the susceptible nodes having 3 infected contacts and 20% with 4 infected contacts (check: 0.8*3 + 0.2*4 = 3.2). Thus, the chance to become infected is defined as: $1 – 0.8*(1 – β)^3 – 0.2*(1 – β)^4 = 0.347$, which is approximately equal to $1 – (1 – β)^{3.2}$. However, simulations using this formula produce steady state values that are significantly lower than the heterogeneous SIS model predicts.

To test the statistical formula, it is compared with the differential equation of the SIS model, which states that a susceptible node has a chance of β*(k*v) to become infected. The component (k*v) is defined as: how many of the k contacts of that particular node are infected. The chance to become infected can now be calculated as 3.2*β = 0.4, which is significantly higher than the value from the statistical formula[52]. Simulations with this formula produce steady state values that closely match the SIS model predictions. So the formula used to calculate the infection chance of a susceptible node with z infected contacts is: β*z.

In the simulations the value of k = 8, so each node links to 4 other nodes in its own sub-population and 4 nodes in the other sub-population. In an undirected network, for each new edge its mirror edge is also added (e.g. add edge from node[0] to node[1] + the mirror edge from node[1] to node[0]). For instance, node[0] is linked to nodes [248, 249, 1, 2] in both sub-populations, node[1] to nodes [249, 0, 2, 3], and so on. This implies that all links are bi-directional, but the infection rate in each direction may be different.

However, a directed network can also be completely symmetrical. An example of such a network: node[0] links to nodes [1, 2, 3, 4], node[1] to nodes [2, 3, 4, 5] … node[246] to nodes [247, 248, 249, 0]. The next section shows that a directed network better matches the predictions of the heterogeneous SIS model.

7. Results of simulations on the heterogeneous SIS model in a directed network

With the formulas in section 6, populations are simulated with the R_0 list [2, 2.25, 2.5, 3, 4, .. , 9] and Diversity index values D list [0.0, 0.25, 0.5, 0.75, 1.0]. For each population in steady state, the size of the infected fraction is calculated. As in section 5, when D = 0 the R_0 values of both sub-populations are equal to the R_0 of the population; when D = 1, the R_0 values of sub-populations #1 and #2 are $2R_0$ and 0, respectively.

Figure 4 shows steady state values of all simulations executed using a directed network, for the R_0 list [2 .. 9] and for the Diversity index values D [0.0, 0.25, 0.50, 0.75, 1.0]. Predictions of the deterministic model are indicated with *v(∞)* and its corresponding simulation as *Sim μ*. This diagram (steady state vs. R_0 for several values of D) is orthogonal to Figure 2 (steady state vs. D for several values of R_0).

As can be seen, the simulation results accurately match the predictions of the deterministic model for high R_0 values. For lower R_0 values and higher D values, the accuracy decreases and the simulation results steadily become lower than the lines of the deterministic model.

Table 2 summarizes the results of simulations for the R_0 list [2, 5, 9]. The columns R_0, R_1 and R_2 respectively contain the Basic Reproduction Numbers of the population as a whole, the first and second sub-population. The column *v(∞)* shows the values calculated with the deterministic model in Annex A for the discrete values of D (diversity). The columns *Sim μ* and *Sim σ* list the calculated average and standard deviation of the steady state part, based on 50 simulation runs.

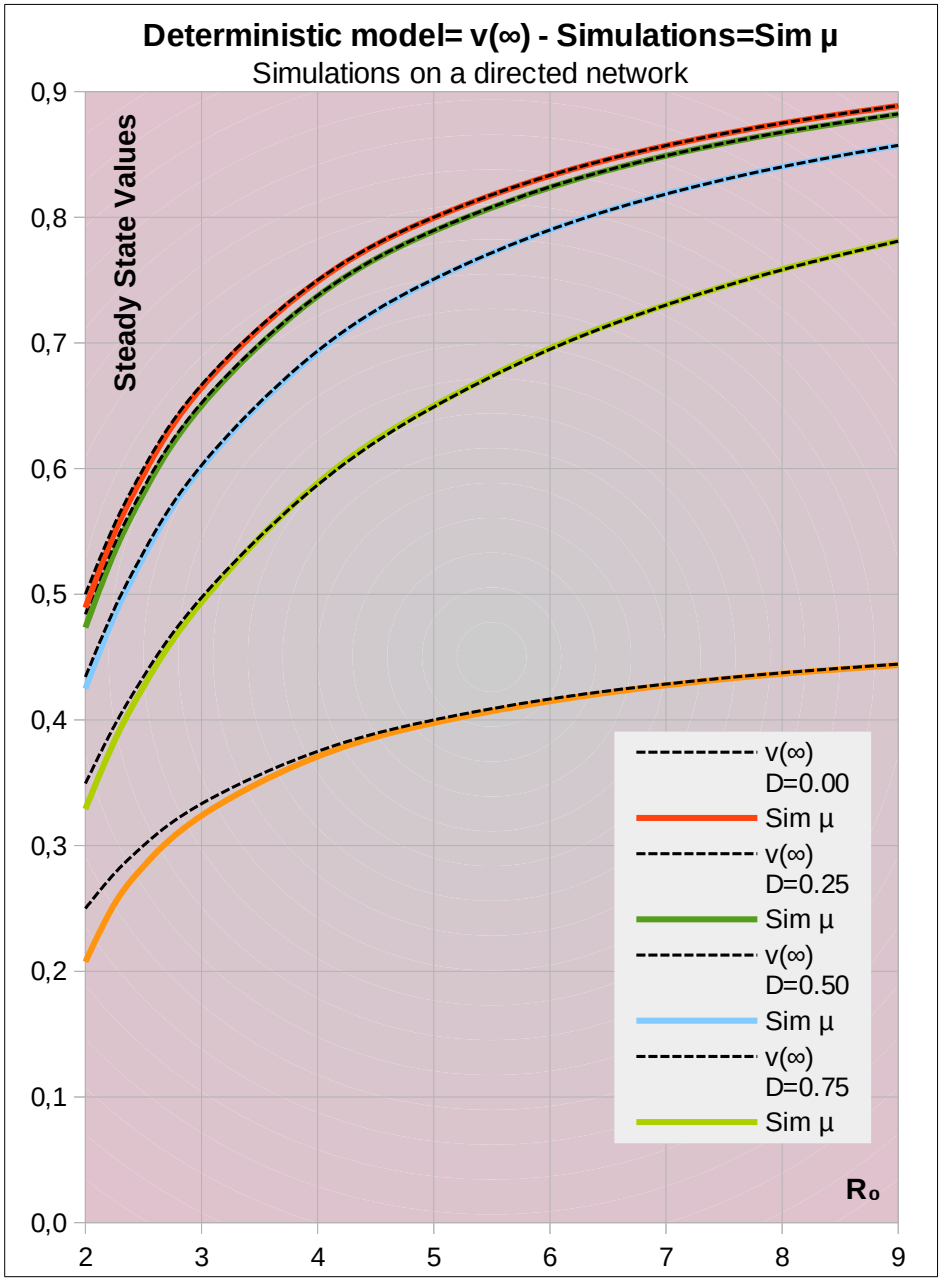

Figure 4. Steady State values of the deterministic model and simulations

The values derived from the simulations with high R_0 values correlate better with the deterministic values. For example, for the R_0 list [5, 6, 7, 8, 9] the correlation between the steady state simulations and the theoretical values of is 0.99999. With -0.72 the standard deviation also correlates with R_0. When the R_0 list is completed with the values [2, 2.25, 2.5, 3, 4], the correlation drops slightly to 0.99958. The correlation between the standard deviation and R_0 weakens to a value of -0.59. The conclusion is that the simulations confirm the predictions of the deterministic model on the impact of diversity.

The steady state values of simulations in Table 2 are mostly lower then predicted by the deterministic model (exceptions in **red**). Because infection and recovery occur by chance and some stochastic models show greater variations, this likely leads to more epidemics to die out[53]. Stochasticity could explain why the simulation results are mostly smaller then theory predicts at lower R_0 values and higher D values.

For instance, when $R_0 = 2$ and the Diversity index (D) is 0.75, the deterministic model predicts an infected fraction of 0.35 in steady state. It takes this simulation 1500 time steps (twice as long) to reach its maximum of 0.32912 (-5.8%). When $R_0 = 2$ and D = 1 the maximum drops to 0.20749 (-17% of 0.25). Therefore, in simulations of directed networks with lower R_0 values and high diversity the impact of diversity is amplified.

R_0	D	R_1	R_2	v(∞)	Sim μ	Sim σ
2.00	**0.00**	2.00	2.00	0.50000	0.48929	0.00147
2.00	**0.25**	2.50	1.50	0.48414	0.47350	0.00191
2.00	**0.50**	3.00	1.00	0.43426	0.42482	0.00235
2.00	**0.75**	3.50	0.50	0.34946	0.32912	0.00291
2.00	**1.00**	4.00	0.00	0.25000	0.20749	0.00644
5.00	**0.00**	5.00	5.00	0.80000	0.79968	0.00136
5.00	**0.25**	6.25	3.75	0.78950	0.78909	0.00104
5.00	**0.50**	7.50	2.50	0.75081	0.75044	0.00119
5.00	**0.75**	8.75	1.25	0.64912	**0.64996**	0.00126
5.00	**1.00**	10.00	0.00	0.40000	0.39769	0.00163
9.00	**0.00**	9.00	9.00	0.88889	0.88874	0.00078
9.00	**0.25**	11.25	6.75	0.88236	0.88203	0.00078
9.00	**0.50**	13.50	4.50	0.85731	0.85731	0.00094
9.00	**0.75**	15.75	2.25	0.78108	**0.78137**	0.00100
9.00	**1.00**	18.00	0.00	0.44444	0.44368	0.00113

Table 2: Simulating steady state values from 2 equally sized sub-populations with different Basic Reproduction Numbers (R_1, R_2), while the R_0 of the population remains constant.

At high R_0 values the results of simulations with undirected networks are only slightly less accurate when compared to simulations with a directed network[54]. But at lower R_0 values and higher Diversity index values undirected networks amplify the impact of diversity even more than directed networks[55]. This is surprising, since both networks are regular graphs. Because all nodes have the same degree (k) of contacts, the viral conductivity of both networks is equal to *k/2* [56].

8. Conclusion and open questions

If the principle of Darwinian evolution is *the survival of the fittest,* usually the arms race between predator and prey is viewed as the most important factor. As stated in section 2, contagious diseases can be seen as fast mutating predators. Since contagious diseases usually mutate much faster as compared to susceptible multi-cell organisms, these organisms cannot adapt quickly enough to prevent all epidemics.

The MacArthur – Wilson theory of Island Bio Geography states that species in big populations are less likely to become extinct. But the more dominant a species becomes, the bigger its genetically uniform population will be and the more this species is impacted by epidemics with contagious diseases. With two opposing forces that are both based on the size of the population, an equilibrium is the likely result.

For mono-cultures the impact of an epidemic with contagious diseases is at maximum, because there is no diversity within the population to quell the propagation of the infections. Dominant species suffer epidemics as frequent as more diverse populations, but the epidemics are more severe. This conclusion partly contradicts previous research that suggested that mono-cultures should suffer less epidemics in number[4].

The SIS model can be used to determine the effect of diversity, when comparing two groups of species that differ in their shared susceptibility for a contagious disease. Because the model does not specify or restrict which two species should be compared, this result holds for *any pair* of species that by their mutual contacts can transfer a contagious disease.
The model and the simulations show that the more genetically different those two groups become, the lower the impact of epidemics gets and the better the chances are for survival of the population as a whole. Therefore the conclusion of this analysis is that contagious diseases stimulate genetic diversity.

This analysis has revealed several areas for further research. First, it should be noted that this analysis relies on the SIS model, which simplifies the real world greatly. It should be applied with care for its limitations. The SIS model requires a static population that is completely symmetrical, i.e. the Infected and Susceptible nodes are distributed evenly in the network. Because the SIS model predicts the emergent (exponential) behavior of contagious diseases, the results of this analysis are significant in the same order of magnitude.

To relax some of the restrictions of the SIS model, studies could use more complex simulations or mathematical models that better resemble reality to determine the impact of diversity on epidemics. For instance, by simulated populations with mutating pathogens and multiple Predator – Prey relations that are not perfectly mixed. The pathogens should exploit genetic vulnerabilities unrelated to the fitness of species in the food chain.

Next, research could be aimed at finding the reason why in simulations with low values of R_0 the effect of diversity is amplified – by undirected networks even more than directed networks. Diversity decreases the steady state more than can be contributed to the number of epidemics that die out, and thus amplify the effects of diversity. This amplification may be caused by stochasticity, or (yet unknown) design flaws in the simulation model design. A possible explanation for the difference between directed and undirected networks could be that undirected networks have nodes with 1-on-1 recursive relations, which could slow down the propagation of infections – whereas the healing of infected nodes does not depend on the conductivity of contacts. This issue may be solved by using more sophisticated simulation techniques.

Finally, a study could investigate if the impact of contagious diseases combined with the MacArthur-Wilson theory of Island Bio-geography, leads to an optimal size of populations of species on an evolutionary time scale. If so, it could be interesting to determine the optimal population size of domestic species, like humans.

Henk-Jan van der Molen CISSP CISM CISA has a background in Electrical Engineering, Audits, Information Systems Design, Software Development, Project Management, Cybersecurity and Privacy. He is a freelance teacher at Security Academy (henk-jan.van.der.molen@securityacademy.org)

The author would like to thank anyone who has supported this research and those who are willing to support it in the future. After all, research will always be "work in progress".

Annex A – mathematical calculations

1. Diversity in the homogeneous SIS model

The homogeneous SIS model is described by the following differential equation:

$$\frac{dv}{dt}=\beta k\, v(1-v)-\delta v \;\rightarrow\; \frac{dv}{dt}=\delta\left(R_0 v(1-v)-v\right) \tag{1}$$

The curing rate is denoted as δ, and the infection rate of all incoming links as β. The fraction of infected and susceptible nodes is represented by v and $(1 – v)$, respectively. The Basic Reproduction Number R_0 is used to replace $\beta.k / \delta$. When $R_0 > 1$, the steady state $h(v)$ can be calculated by setting the differential equation to zero and ignoring the solution $v = 0$:

$$\frac{dv}{dt}=0 \;\Rightarrow\; R_0 v(1-v)-v=0 \;\Rightarrow\; v\left(R_0-1-R_0 v\right)=0 \;\Leftrightarrow\; h(v)=1-\frac{1}{R_0} \tag{2}$$

If a fraction n_2 of the population becomes immune to a disease ($R_2 = 0$) and the rest ($n_1 = 1 – n_2$) remains susceptible ($R_1 > 0$), the susceptible part of the population in (2) changes from $(1 – v)$ to $(1 – n_2 – v)$. Thus:

$$R_0 v(1-n_2-v)-v=0 \;\Leftrightarrow\; h(v)=1-n_2-\frac{1}{R_1}; \quad \text{when} \quad n_2 \geqslant \left(1-\frac{1}{R_1}\right) \quad \text{the HIT is reached} \tag{3}$$

To do a fair analysis on the effect of diversity on the steady state of v using the SIS model, the value of R_0 for the population as a whole must be kept constant, above the epidemic threshold. If R_0 is not kept constant, $h(v)$ will fluctuate with the value of R_0 and this would invalidate the analysis.

When n_1 decreases, the value of R_0 for the population as a whole remains constant if: $R_1 = R_0 / n_1$ (4)

When $\{n_1 = 0\}$ or $\{n_1 = 1\}$, this by definition constitutes a mono-culture, so Diversity (D) is defined as: $D = 2n_1n_2R_1 / R_0$, with $\{0 < n_1, n_2 \leq ½\}$, so $\{0 \leq D \leq 1\}$. (5)

Inserting (4) for $h(v)$ in eq. (3) and D in eq. (5): $h(v)=1-n_2-\frac{(1-n_2)}{R_0}; \quad D=\frac{2n_1n_2R_1}{R_0}=2n_2$ (6)

To investigate how v behaves as a function of diversity in steady state, its derivative v' is computed:

$$v' = \frac{dv}{dD} = \frac{1}{2}\frac{dv}{dn_2} = -\frac{1}{2}\left(1-\frac{1}{R_0}\right) \tag{7}$$

Therefore when diversity increases in the interval $[0 \rightarrow 1]$, v decreases monotonously between the extremes of $v(D)$:

$$v(0)=1-\frac{1}{R_0}; \quad v(1)=\frac{1}{2}-\frac{1}{2R_0} \tag{8}$$

2. Diversity in the heterogeneous SIS model with 2 sub-populations

To be able to read this article independently, some formulas for the heterogeneous SIS model are restated of a previous publication[5]. The general heterogeneous SIS model is described by the following equations:

$$\frac{dv_i}{dt}=\delta_i\left(R_i v(1-v_i)-v_i\right); \quad \sum_{i=1}^{m}(n_i)=1; \quad v=\sum_{i=1}^{m}(n_i v_i); \quad R_0=\sum_{i=1}^{m}(n_i R_i) \tag{9}$$

In a population with m sub-populations, the curing rate of sub-population i is denoted as δ_i, and the Basic Reproduction Number as R_i, for $i = \{1, 2, \ldots, m\}$. The fraction of infected nodes of sub-population i and the population as a whole are represented by v_i and v, respectively. Like in the homogeneous SIS model, a disease becomes epidemic when its $R_0 > 1$. The steady state $h(v)$ is calculated, ignoring $v = 0$ as a solution.

$$\frac{dv_i}{dt}=0 \;\rightarrow\; v_i=\left(\frac{R_i v}{R_i v+1}\right);\;\; v=\sum_{i=1}^{m} n_i v_i \;=\; v\sum_{i=1}^{m}\left(\frac{n_i R_i}{R_i v+1}\right) \;\rightarrow\; (\forall\, v\neq 0)\;\; \sum_{i=1}^{m}\left(\frac{n_i R_i}{R_i v+1}\right)=1 \qquad (10)$$

For 2 sub-populations (m = 2), the steady state of v only depends on the parameters n_1, R_1 and R_2:

$$h(v)=av^2+bv+c \;\Rightarrow\; v=\frac{-b+\sqrt{d}}{2a} \qquad (11)$$
$$with \;\; a=R_1R_2;\;\; b=R_1+R_2-R_1R_2;\;\; c=1-n_1R_1-n_2R_2;\;\; d=b^2-4ac$$

When {$R_0 > 1$} then {$c < 0$}, so the heterogeneous model always yields a solution for $h(v)$ with $v > 0$.

Diversity (D) is defined similar to (5): $D = 2n_1n_2|R_1 - R_2| / R_0$, with {$0 \le n_1, n_2 \le ½$}, so {$0 \le D \le 1$}.

Keeping the value of R_0 for the population as a whole constant, the parameter n_1 is set to a constant value which guarantees maximum diversity and select R_1 and R_2 as variables, i.e.:

$$n_1=n_2=\frac{1}{2} \;\rightarrow\; R_1+R_2=2R_0 \;\rightarrow\; R_2=2R_0-R_1 \qquad (12)$$

If the condition {$R_0 \le R_1 \le 2R_0$} holds, then $D = (R_1 - R_0) / R_0 \equiv R_1 / R_0$.

In the following R is used instead of R_1 and $(2R_0 - R)$ for R_2, thus:

$$a=-R^2+2R_0R;\;\; b=R^2-2R_0R+2R_0;\;\; c=1-R_0;$$
$$d=b^2-4ac=R^4-4R_0R^3+4R_0^2R^2+4R^2-8R_0R+4R_0^2 \qquad (13)$$

To investigate how the steady state of v behaves as a function of diversity, its derivative v' is computed:

$$a'=\frac{da}{dD}=R_0\frac{da}{dR}=2R_0(R_0-R);\quad b'=-a';\;\; d'=R_0\frac{dd}{dR}=4R_0(R^2-2RR_0+2)(R-R_0)$$

$$v'=R_0\frac{dv}{dR}=R_0\left(\frac{-2ab'+2a'b}{4a^2}+\frac{ad'}{4a^2\sqrt{d}}-\frac{2a'\sqrt{d}}{4a^2}\right)=R_0\frac{(4R_0a'\sqrt{d}+ad'-2a'd)}{4a^2\sqrt{d}} \;\rightarrow$$

$$v'=2R_0^2(R-R_0)\frac{(R^2-2R_0R+2R_0^2-R_0\sqrt{d})}{R^2(R-2R_0)^2\sqrt{d}} \qquad (14)$$

Because {$R_0 > 1$} and {$R_0 \le R < 2R_0$} and therefore {$R^2 - 2R_0R + 2R_0^2 > 0$}, the sign of v' only depends on:

$$(R^2-2R_0R+2R_0^2)^2-R_0^2d \;=\; -(R_0^2-1)R^2(R-2R_0)^2 \qquad (15)$$

When diversity increases in the interval [0 → 1], v decreases monotonously between the same extremes of $v(D)$ as stated in (8).

Annex B – Python source code of the simulation

```
''' Run a simulation of the heterogeneous SIS model with 2 subpopulations.
Graph implementation based on: https://stackabuse.com/courses/
graphs-in-python-theory-and-implementation/lessons/representing-graphs-in-code/

*** Formulas - rnd = random sample from U[0..1] ***
chance node is infected from k infected contacts  = beta * k  = (rnd < beta * k)
chance infected node recovers from infection      =     delta = (rnd < delta)
chance infected node does NOT recover             = 1 - delta = (rnd > delta)

If  node = 0    # Susceptible, chance of (re)infection =
    node = 1 * (rnd < beta * count_infected_contacts)
Else            # Infected, chance to remain infected =
    node = 1 * (rnd > delta)

Author: Henk-Jan van der Molen, 2023-01-15
Last source code versions on: https://github.com/HJvdMolen/SIS-model.git
'''

def calc_stats(floatlist):  # calculate μ, σ  from a list of values
    from math import sqrt
    total = total2 = 0
    for value in floatlist:
        total  += value
        total2 += value ** 2

    n = len(floatlist)
    mu = total / n
    sd = sqrt((total2 - total ** 2 / n) / n)
    return mu, sd

def plot_function(function, title, xlabel, ylabel, v_8):
    import matplotlib.pyplot as plt
    plt.plot(function)
    plt.plot([v_8 for _ in range(10 + len(function))], linestyle='dashed', linewidth=1)
    plt.title(title)
    plt.xlabel(xlabel)
    plt.ylabel(ylabel)
    plt.show()
    return

def calc_beta_delta(r0, k): # calculates β, δ as big as possible
    assert k % 4 == 0,      f"Variable k must be a multiple of 4, got {k}"
    assert r0 >= 0,         f"r0 must be >= 0, got {r0}"

    if r0 > 1:
        delta = 1 / r0
    else:
        delta = 1
    assert 0 < delta <= 1,  f"0 < δ <= 1 expected, got: {delta}"

    beta = r0 * delta / k
    assert beta * k <= 1,   f"β.k <= 1 expected, got: {beta * k}"

    print(f" r0 = {r0}; k = {k}; beta = {beta:.3f}; delta = {delta:.3f}")
    return beta, delta

class Graph:
    # Constructor
    def __init__(self, network_size, n1, no_of_steps, r0, D):
        self.network_size = network_size
        self.n1 = n1
        self.n2 = 1 - n1
        self.no_of_steps = no_of_steps
        self.r0 = r0
        self.D  = D
```

```
        # Calculated later...
        self.v_8 = 0.0
        self.mu = self.sd = 0.0

        # Create simulation + average simulation list
        self.sim_tot = [0 for j in range(self.no_of_steps)]
        self.sim_avg = [0 for j in range(self.no_of_steps)]
        return

    def calc_steady_state(self, p1, p2):
        from math import sqrt

        a           = p1.r0 * p2.r0       # calculate steady state of v
        if a != 0:
            b       = p1.r0 + p2.r0 - a
            c       = 1 - self.r0
            d       = b * b - 4 * a * c
            self.v_8 = (sqrt(d) - b) / (2 * a)
        elif p1.r0 == 0:
            self.v_8 = self.n2 - 1 / p2.r0
        elif p2.r0 == 0:
            self.v_8 = self.n1 - 1 / p1.r0
        return

    # Execute complete simulation
    def run_simulation(self, p1, p2):
        p1.init_simulation()
        p2.init_simulation()
        self.sim_tot = [0 for j in range(self.no_of_steps)]

        for j in range(1, self.no_of_steps):
            p1.simulation_step(p2, j)
            p2.simulation_step(p1, j)

            p1.node = p1.temp[:]
            p2.node = p2.temp[:]

            avg_v = self.n1 * p1.sim[j] + self.n2 * p2.sim[j]
            self.sim_tot[j]  = avg_v
            self.sim_avg[j] += avg_v

        # Calculate (μ, σ) from steady state part = right half of sim[]
        self.mu, self.sd = calc_stats(self.sim_tot[self.no_of_steps//2:])
        return

class SubPopulation:
    def __init__(self, num_of_nodes, beta, k, delta, no_of_steps, directed):
        self.num_of_nodes = num_of_nodes
        self.no_of_steps = no_of_steps
        self.directed = directed

        self.beta = beta
        self.k = k
        self.delta = delta
        self.r0 = beta * k / delta

        # Create the nodes list
        self.node = [0 for j in range(self.num_of_nodes)]

        # Create Temp copy of self.node[]
        self.temp = [0 for j in range(self.num_of_nodes)]

        # Create simulation list for SubPopulation
        self.sim = [0 for j in range(self.no_of_steps)]

        # Initialize & fill the adjacency list
        self.adj_list = {node: set() for node in range(num_of_nodes)}
```

```
        # adj_list is also used for contacts between SubPopulations = /2
        if self.directed:   # directed graph
            contacts_out = self.k // 2
        else:               # undirected = include "mirror" edges
            contacts_out = self.k // 4

        for key in self.adj_list.keys():
            for j in range(contacts_out):
                # Link node(n) to node(n+1, n+2, .. , n+contacts_out)
                self.add_edge(key, (key + 1 + j) % self.num_of_nodes, directed)
        return

    # Add the edge from node1 to node2 and v.v. in undirected graphs
    def add_edge(self, node1, node2, directed):
        assert node1 != node2, f"node {node1} cannot connect to itself"

        self.adj_list[node1].add(node2)
        if not (self.directed or node1 == node2):   # also add "mirror" edge
            self.adj_list[node2].add(node1)
        return

    def init_simulation(self):
        import random

        # Set all nodes to be Susceptible (== 0)
        for j in range(self.num_of_nodes):
            self.node[j] = 0
            self.temp[j] = 0

        if False:   # switch to start with 100% infected
            self.node = [1 for j in range(self.num_of_nodes)]
            self.nodes_infected = self.num_of_nodes
        else:       # Randomly infect 2% of susceptible nodes (0 => 1)
            self.nodes_infected = self.num_of_nodes // 50
            for j in range(self.nodes_infected):
                while True:
                    p = random.randrange(0, self.num_of_nodes)
                    if self.node[p] == 0 : break
                self.node[p] = 1    # node is infected

        # Fill in first value in simulation list
        self.sim = [0 for j in range(self.no_of_steps)]
        self.sim[0] = self.nodes_infected / self.num_of_nodes
        return

    # #infected contacts (node[]==1); p1 + p2 have identical adjacency list
    def count_infected_contacts(self, other, key):
        cic = 0
        for nd in self.adj_list[key]:
            cic +=  self.node[nd]   # If node[nd] = 1 -> Infected
            cic += other.node[nd]
        return cic

    # Execute single time step in the simulation for all nodes
    def simulation_step(self, other, step_no):
        import random
        self.nodes_infected = 0
        for j in range(self.num_of_nodes):
            rnd = random.uniform(0,1)
            if self.node[j] == 0:    # Susceptible node, can be infected
                contacts = self.count_infected_contacts(other, j)
                self.temp[j] = 1 * (rnd < self.beta * contacts)
            else:                    # Infected node, can recover
                self.temp[j] = 1 * (rnd > self.delta)
            self.nodes_infected += self.temp[j]

        self.sim[step_no] = self.nodes_infected / self.num_of_nodes
        return
```

```
# Main program
if __name__ == '__main__':
    network_size  = 500               # no. of nodes in the network
    k             = 8                 # no. of contacts for each node
    n1            = 0.5               # fraction of SubPopulation 1
    n2            = 1 - n1            # fraction of SubPopulation 2
    directed      = True              # False = bidirectional network
    assert n1 == n2 == 0.5, f"n₁, n₂ must be 0.5, got n1 = {n1}, n2 = {n2}"

    r0            = 3                 # basic reproduction number R₀
    assert r0 > 1, f"R₀ > 1 expected, got: {r0}"

    D             = 0.00              # Diversity index, [0 = min, 1 = max]
    assert 0 <= D <= 1, f"0 <= Diversity <= 1, got {D}"

    no_of_steps   = 750               # no. of time steps in 1 simulation
    g             = Graph(network_size, n1, no_of_steps, r0, D)

    # SubPopulation #1 & #2
    num_of_nodes1 = int(network_size * n1)
    num_of_nodes2 = network_size - num_of_nodes1
    r0_1          = r0 * (1 + D / (4 * n1 * n2))
    r0_2          = (r0 - n1 * r0_1) / n2
    # β: infection chance from 1 infected contact | δ: recovery from infection
    beta1, delta1 = calc_beta_delta(r0_1, k)
    beta2, delta2 = calc_beta_delta(r0_2, k)
    p1            = SubPopulation(num_of_nodes1, beta1, k, delta1, no_of_steps, directed)
    p2            = SubPopulation(num_of_nodes2, beta2, k, delta2, no_of_steps, directed)

    g.calc_steady_state(p1, p2)   # calculates D and v(∞)
    g.run_simulation(p1, p2)

    title  = f"SIS simulation: R₀, R₁, R₂ = {g.r0}, {p1.r0}, {p2.r0}; v(∞) = {g.v_8}"
    xlabel = f"#time steps: {no_of_steps} with {network_size} nodes, directed network:
{directed}"
    ylabel = f"μ(v) = {g.mu:.5f}, σ(v) ={g.sd:.5f}"
    plot_function(g.sim_avg,  title, xlabel, ylabel, g.v_8)
```

Funding: this research was not funded by any organization. It was supported by individuals investing time in reviewing draft versions of the article. The reviewers had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1 G.E. Hutchinson, the American Naturalist 95, page 137 (1961)
2 G. Hardin, “The Competitive Exclusion Principle”, Science vol. 131, page 1292 (1960)
3 T. Bonnet, “Genetic variance in fitness indicates rapid contemporary adaptive evolution in wild animals” Science vol 376, page 1012 (2022)
4 The Contribution of Genetic Diversity to the Spread of Infectious Diseases in Livestock Populations, A. J. Springbett e.a., july 28, 2003.
5 The article “On the Malware Front”, International Journal of Computer Networks (2012) uses the heterogeneous SIS model to determine why diversity decreases the propagation of contagious diseases.
6 Herd Immunity occurs when a significant part of the population becomes immune for an infectious disease. When the Herd Immunity Threshold (HIT) is reached, the spread of this disease becomes very limited. This provides indirect protection for the complete population.
7 Last version of the source code is placed on https://github.com/HJvdMolen/SIS-model.git
8 The periodic Predator-Prey Lotka-Volterra Model, Julián López-Gómez e.a., Advances in Differential Equations Volume 1, Number 3, May 1996, pp. 403 – 423
9 Co-evolution Maintains Diversity in the Stochastic “Kill the Winner” Model, Chi Xue and Nigel Goldenfeld, June 9, 2017
10 Is Infectious Disease Just Another Type of Predator-Prey Interaction? S.R. Hall e.a., Infectious Disease Ecology ISBN 978-0-69112485-8, page 235
11 From Bacteria to Bach and back – Daniel C. Dennett, page 7 – ISBN 9780141978048
12 Phages of Staphylococcus aureus and their impact on host evolution, Guoqing Xia and Christiane Wolz, Infection, Genetics and Evolution 21 (2014) 593–601
13 Imperfect Vaccination Can Enhance the Transmission of Highly Virulent Pathogens, Andrew F. Read et al, July 27, 2015
14 Adaptation of Bordetella pertussis to Vaccination: A Cause for Its Reemergence? Frits R. Mooi et al, June 2001
15 Rapid Increase in Pertactin-deficient Bordetella pertussis Isolates, Australia, Connie Lam et al, April 2014
16 Hepatitis B surface antigen escape mutations: Indications for initiation of antiviral therapy revisited, Jennifer Leong et al, March 16, 2016
17 Tracking a Hospital Outbreak of Carbapenem-Resistant Klebsiella pneumoniae with Whole-Genome Sequencing, Evan S. Snitkin et al, 2012
18 Core Genes Evolve Rapidly in the Long-Term Evolution Experiment with Escherichia coli; Rohan Maddamsetti e.a., April 3, 2017
19 Genetic variation and susceptibility to infection: the red cell and malaria; D. J. Weatherall, British Journal of Haematology, 2008
20 Evolution toward multi-year periodicity in epidemics, Masashi Kamo and Akira Sasaki, 2005
21 Stability Criteria for Complex Microbial Communities, Stacey Butler and James O’Dwyer, April 2, 2018
22 The ecology of the microbiome: Networks, competition, and stability, Katharine Z. Coyte et al, November 6, 2015
23 Sinus Microbiome Diversity Depletion and Corynebacterium tuberculostearicum Enrichment Mediates Rhinosinusitis Nicole A. Abreu et al, Science Translational Medicine, 2012
24 Metagenomic Analysis of the Human Gut Microbiome, M.B.Q. dos Santos, 2013
25 Microbiota biodiversity in inflammatory bowel disease, Comito e.a., Italian Journal of Pediatrics 2014.
26 An overview of major metagenomic studies on human microbiomes in health and disease, Hongfei Cui e.a., Quantitative Biology 2016
27 Contribution of Host Defence Proteins and Peptides to Host-Microbiota Interactions in Chronic Inflammatory Lung Diseases, Anne M. van der Does e.a., July 28, 2018
28 The microbiome of Crohn’s disease aphthous ulcers, Claire L. O’Brien e.a., Gut Pathogens 2018
29 Inflammatory Bowel Diseases: Host-Microbial-Environmental Interactions in Dysbiosis, C. Colquhoun e.a., May 10 2020.
30 Interconnections between the Oral and Gut Microbiomes: Reversal of Microbial Dysbiosis and the Balance between Systemic Health and Disease; Micro organisms 2021, 9, 496.
31 How Many Friends Does One Person Need? – Dunbar's Number and Other Evolutionary Quirks, page 143, Robin Dunbar – ISBN 9780571258291
32 MacArthur, R. H., and E. O. Wilson. The Theory of Island Biogeography. Princeton, NJ: Princeton University Press, 1967.
33 Antagonism influences assembly of a Bacillus guild in a local community and is depicted as a food-chain network, 2013, Rocío-Anaís Pérez-Gutiérrez e.a.
34 Microbial diversity determines the invasion of soil by a bacterial pathogen J.D. van Elsas e.a., 2011
35 The Theory of Island Biogeography Revisited, 2010, Princeton University Press, page 65
36 Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the “Killing the Winner” Hypothesis Revisited; C. Winter e.a., Microbiology and Molecular Biology Reviews ,

References

Mar. 2010, p. 42–57.

37 Pathogenic Responses among Young Adults during the 1918 Influenza Pandemic, February 2012, G. Dennis Shanks and John F. Brundage

38 Red Queen Communities, K. Clay e.a., Infectious Disease Ecology ISBN 978-0-69112485-8, page 156

39 Ancient oncogenesis, infection and human evolution, 2017, R. F. Rifkin e.a.

40 Effects of Disease on Ecosystems, V.T. Eviner; Effects of Disease on Keystone species, Dominant Species and Host Communities S.K Collinge e.a. - Infectious Disease Ecology ISBN 978-0-69112485-8, page 126 + 140 (combined)

41 Antigenic Diversity, Transmission Mechanisms and the Evolution of Pathogens, Alexander Lange and Neil Ferguson, PLoS Computational Biology, www.ploscompbiol.org, October 2009, Volume 5, Issue 10

42 Murphy's law is based on the cumulative Poisson distribution where the chance of infection with a changing pathogen within a certain time period is non-zero; during a long enough time period, the chance that no infections occurs approaches zero.

43 Impact and longevity of measles-associated immune suppression: a matched cohort study using data from the THIN general practice database in the UK, Kartini Gadroen e.a., November 8, 2018

44 Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings, Christophe Fraser e.a., 2009

45 Consensus document on the epidemiology of severe acute respiratory syndrome (SARS), World Health Organization, 2003

46 A Mathematical Model of Human Papilloma Virus (HPV) in the United States and its Impact on Cervical Cancer; Shernita L. Lee e.a., 2012

47 Sero epidemiologic Survey for Hepatitis B Virus Infection in Taiwan: The Effect of Hepatitis B Mass Immunization Hsu-Mei Hsu, Chih-Feng Lu, Shin-Chwen Lee, Sheue-Rong Lin & Ding-Shinn Chen

48 Herd Immunity: History, Theory, Practice; Paul E. M. Fine, 1993

49 The Gini-Simpson index for diversity specifies the chance that two individuals are of different sub-populations. In this context, diversity is maximized when $n_1 = n_2 = 0.5$

50 Sage-math was used to calculate dv/dR, which is equivalent to the derivative of v for diversity.

51 For example, when $R_0 = 3$ and D = 0, the simulation returns steady state values for (2% initially infected) $\mu = 0.66427$, $\sigma = 0.00201$ vs. (100% infected) $\mu = 0.66388$, $\sigma = 0.00185$

52 It is easy to verify that predictions using the statistical formula are significantly lower for any $k^*v > 1$. For example, when $\beta = 0.125$, and $k^*v = 3.2$, the value from the statistical formula (= 0,347), minus the differential equation in steady state ($\beta kv = 0.4$) is: $\beta^2(\beta^2(3 - kv) + \beta(3kv - 8) + 6 - 3kv) = -0,053$

53 Integrating stochasticity and network structure into an epidemic model, C. E. Dangerfield e.a., 30 October 2008

54 For the R_0 list [5, 6, 7, 8, 9], the correlation between the predictions of the deterministic model and the results of simulating undirected networks is 0.99997, so slightly less than simulations of a directed network.

55 For instance, when $R_0 = 2$ and D = 0.75, the undirected network with 500 nodes needed1500 time steps to reach its steady state of 0.2817 (-19% of the predicted value 0.35). The situation becomes even more dramatic for $R_0 = 2$ and D = 1: in one simulation run the steady state collapses to 0.0044 (-98% of the predicted value 0.25), with 43 of 50 epidemics (-86%) in that simulation run die out. Note that the decrease of the steady state cannot be caused solely by the number of epidemics dying out. Although less dramatic, the same applies to simulations of a directed network with $R_0 = 2$ and D = 1: no epidemics die out, but the steady state is decreased to 0.203 (-19% of the predicted value of 0.25). Repeated simulations show comparable results.

56 The viral conductance of a network; Piet Van Mieghem, Computer Communications 35, 2012