

Network Science Cheatsheet



Made by
Remy Cazabet

Spreading Processes

Spreading - Diffusion

In many real world situations, networks can be seen as **support of spreading or diffusion** processes. Typical examples are diffusion of information, innovation, rumors, biological or computer virus, adoption of ideas or products, etc. The support can be social networks of interactions, social media platforms of computer networks, to name a few.

A typical way to model such a process is to assign **properties** to nodes, categorical or numerical, to represent the current status of the node relatively to the process (e.g., *Susceptible*, *Infected*). We can also call such a process **Dynamic On Networks**, as opposed to **Dynamics Of Networks**.

SI - SIR - SIS

Three of the most popular models of diffusion in epidemiology are the **SI**, **SIR** and **SIS** models. Letters correspond to the states in which individuals can be according to the model:

- **Susceptible:** Individual is not Infected
- **Infected:** Individual is Infected
- **Recovered/Removed.** Individual cannot be infected again (Considered cured or dead)

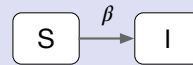
All individuals are in one of the states allowed by the model, and we define:

$s(t)$	Fraction of individuals in Susceptible state at time t
$i(t)$	Fraction of individuals in Infected state at time t
$r(t)$	Fraction of individuals in Recovered state at time t
i_0	Initial ($t = 0$) fraction of infected individuals

SI

In the SI model, individuals can be only in two states, Susceptible and Infected. Susceptible ones can become Infected, and Infected individuals rest in this state indefinitely. Parameters are:

- τ **Infectivity:** probability that the contact between an *Infected* individual and a *Susceptible* one results in the infection of the Susceptible.
- \hat{c} **Contact rate:** average number of contact per person per time
- β **Effective contact rate.** $\beta = \tau\hat{c}$, number of newly infected individuals by each infected individual in a population in which everyone else is susceptible.



SI - characteristics

Each of the i infected individuals infects in average β contacts, but only $s = (1 - i)$ of its contacts are indeed susceptible. More formally using differential equations:

$$\frac{di}{dt} \quad \text{Rate of new infection: } \frac{di}{dt} = \beta i s = \beta(1 - i)i$$

$$i(t) \quad \text{Infected fraction}^a: i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}$$

$$s(t) \quad \text{Susceptible fraction: } 1 - i(t)$$

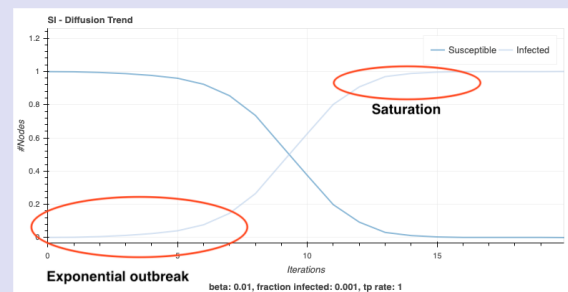
The process can be separated in three steps:

- At first, the fraction of infected individuals **Grows exponentially** until a large fraction of the population is infected. (i is small, $\frac{di}{dt} \approx \beta i \Rightarrow$ exponential)
- Due to **saturation**, the infection of the last individuals is slow
- The growth is faster and faster until half the population is infected ($\text{argmax}_{x,y}(x(1-x)) : x = y = 0.5$).

If $\beta > 0$, everyone is infected at the end of the process.

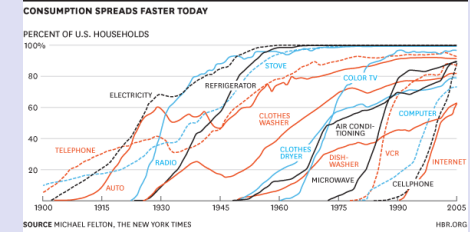
^aBarrat, Barthelemy, and Vespignani 2008.

SI - Sketch



SI - Application

An example in which this model can be appropriate is for diffusion of innovation: being infected means buying or using a new service, product or technology whose usage becomes widespread in society, e.g., television, cell-phone, internet, Netflix, etc.

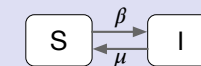


SIS

When modeling real viruses, it is often useful to consider that infected individuals can **recover**, i.e., go back to the susceptible state, without becoming immune, such as for common cold or influenza. This can be modeled using the **SIS** model.

Additionally to β , the SIS model requires another parameter:

- μ **recovery rate:** probability that an *Infected* individual go back to the susceptible state per unit of time.



SIS - characteristics

Intuitively, the fraction of infected individuals is now reduced by those switching to the susceptible state, more formally:

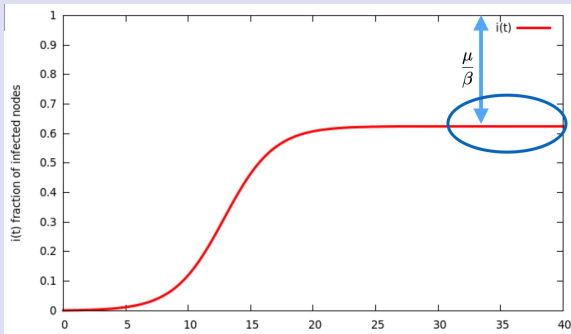
$$\frac{di}{dt} \quad \text{Rate of new infection: } \beta i(1 - i) - \mu i = i(\beta - \mu - \beta i)$$

$$i(t) \quad \text{Infected fraction}^a: \left(1 - \frac{\mu}{\beta}\right) \frac{C e^{(\beta - \mu)t}}{1 + C e^{(\beta - \mu)t}}$$

For large times, $i(t) \rightarrow 1 - \frac{\mu}{\beta}$, i.e., the fraction of infected individuals stabilize around a value which depends only of parameters μ and β .

^aBarrat, Barthelemy, and Vespignani 2008.

SIS - Sketch



λ ratio or (R_0)

In the SIS model, an important notion is the λ ratio, also called R_0 .

$$R_0 = \frac{\beta}{\mu}$$

R_0 can be understood as the average number of individuals that will be infected by an infected individual, **in a population in which all other nodes are Susceptible**. R_0 is a property of the model and **do not change with time**.

Looking at the R_0 is important in the early stage of the epidemic:

- if $R_0 > 1$, **there will be an outbreak**
- if $R_0 < 1$, **the epidemic will disappear naturally**.

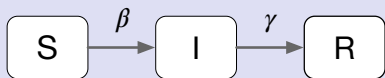
If R_0 is just above 1, the outbreak also can stop naturally by chance in the early stage.

SIR

In many spreading situations, infected individuals can themselves switch to a new state, usually called **Recovered**, which can represent either an acquired immunization or its removal (death, computer failure, etc.). In any case, Recovered individuals cannot infect nor be infected.

Additionally to β , the SIR model requires another parameter:

γ | **recovery rate**: probability that an *Infected* individual switch to the *Recovered* state per unit of time.



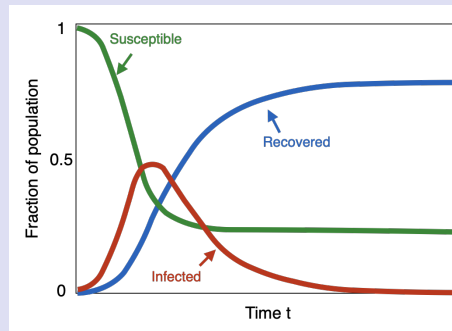
SIR - characteristics

Intuitively, the fraction of infected individuals is now reduced by those switching to the recovered state, more formally:

$$\frac{ds}{dt} = -\beta is, \quad \frac{di}{dt} = \beta is - \gamma i, \quad \frac{dr}{dt} = \gamma i$$

- The initial steps of the outbreak still follow an exponential growth
- The fraction of infected nodes reach a peak and then decreases
- The fraction of recovered can saturate below 1
- The λ ratio is defined as $\lambda = \frac{\beta}{\gamma}$

SIR - sketch



Other non-network models

Epidemic modeling is a large and rich scientific topic, thus those models are nowadays considered *toy models*, too simple to model real epidemics. Most used model thus include other factors, such as **natural population dynamic** (birth, natural death, etc. for long term dynamics), **population segmentation** (τ might differ among subsets of populations, e.g., elderly, married couples, etc.), **population mixing** (\hat{c} might vary between members of a subpopulation and another), etc.

Network models

A natural way to add more details to a spreading process is, instead of considering an homogeneous population, or a population composed of homogeneous subpopulations, to study the diffusion on a network representing the structure of the population. \hat{c} has no meaning in networks (its role is played by the network structure), so by convention we use $\beta = \tau$ the probability for a node to infect each of its neighbor at each step.

Homogeneous Networks

If we consider an **homogeneous random network** in which all nodes have degree exactly k , then we can consider the spreading on this network as similar to the non-network models, with $\hat{c} = k$. For instance, the SI model becomes:

$$\frac{di}{dt} = \beta \langle k \rangle (1 - i) i$$

Note that in practice, there are a few differences, such as a switch from a continuous to a discrete setting, but they have no consequences on large graphs.

ER Networks

In ER random graphs, for large graphs, we have seen that $k \approx \langle k \rangle$, thus the same model description still holds, with the approximation being better for larger networks.

Heterogeneous Degrees - Degree blocks

We have seen that most real networks have an heterogeneous degree distribution. To study analytically the effect of this property, a method is to use a **degree block approximation**: we consider all nodes with a given degree as an homogeneous groups (degree block), and analyze each of these groups separately.

i_k, s_k, r_k : fractions of nodes of degree k that are infected, susceptible, recovered, respectively.

The global average is the average for all degree block, weighted by the fraction of nodes in each block. $i = \sum_k P(k) i_k$ with $P(k)$ the fraction of nodes having degree k .

The same holds for global s and r .

Heterogeneous Degrees - SI

For the SI model, we know that all nodes are infected in the end, but what may vary is **speed** of the process.

The speed of diffusion by degree block can be expressed as:

$$\frac{di_k}{dt} = \beta k (1 - i_k) \Theta_k$$

with Θ_k being the fraction of infected neighbors of a node with degree k .

Heterogeneous Degrees - Θ_k

Θ_k represents the fraction of infected neighbors of a node with degree k . If the network is homogeneous, $\Theta_k(t) = i(t)$. If the graph is heterogeneous, nodes with different degrees have different probabilities of being infected: higher degree nodes are, by definition, more exposed, since they have more chances of being infected at each step.

This is important for two reasons:

- Due to the friendship paradox, nodes are more likely to be connected to large nodes than to small ones
- In real networks, we have seen that there is often a degree assortativity, thus nodes of a given degree have different degrees in their neighborhood.

We can thus define Θ_k as the probability to be connected to nodes of a given degree and or their respective probability of being infected $\Theta_k = \sum_{k'} P(k'|k) i_{k'}$, with $P(k'|k)$ the probability that a node with degree k connects to a node with degree k' .

For simplicity, **we assume no degree-degree correlation.**

$P(k'|k)$ can then be expressed as the fraction of all edge stubs attached to nodes of degree k' , independent of the k under study:

$$P(k'|k) = \frac{k' P(k')}{\sum_{k''} k'' P(k'')} = \frac{k' P(k')}{\langle k \rangle}$$

And:

$$\Theta_k = \Theta = \frac{\sum_{k'} k' P(k') i_{k'}}{\langle k \rangle}$$

Heterogeneous Degrees - SI - time scale

From previous equations, it can be shown^a that the **time scale** τ of the process, i.e., a measure inversely proportional to its speed, is $\tau = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}$.

Thus, for a given average degree $\langle k \rangle$ and a given β , **the more heterogeneous the degrees, the faster the diffusion.**

If the degree distribution follows a power law of exponent $\alpha < 3$, we have seen that $\langle k^2 \rangle$ diverge towards infinity, thus τ tends toward 0, thus the diffusion is nearly instantaneous.

This can be understood as follows: if a node is connected to nearly every other node, then it has an extremely high probability to become infected immediately, and can then infect the rest of the network extremely quickly.

^aBarrat, Barthelemy, and Vespignani 2008.

Heterogeneous Degrees - λ

For SIS and SIR models, it can also be shown^a that the epidemic threshold λ (or R_0) is not reached when $\lambda = \frac{\beta \langle k \rangle}{\mu} > 1$ as in homogeneous networks, but when $\lambda > \frac{\langle k \rangle}{\langle k^2 \rangle}$.

This means that in a very heterogeneous network, **an outbreak can start even if λ is very small, and below 1.** Intuitively, even if people recover faster than they spread the virus in average, some nodes (hubs) will nevertheless become infected, and since they can infect many others, the contagion will spread. Fortunately, it can also be shown that such an outbreak will stop way before reaching all nodes

^aBarrat, Barthelemy, and Vespignani 2008.

SIR - Experimental

When analytical solutions cannot be simply derived, empirical simulations can be used to observe the effect of network properties on diffusion processes.

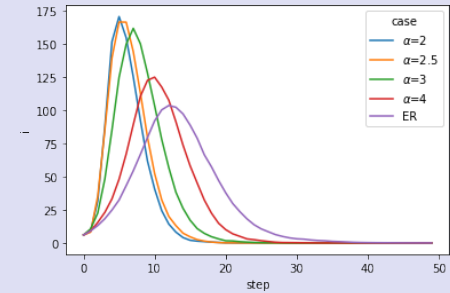
In particular, these properties can be used to assess the effect of typical heterogeneity: degree-heterogeneity, belonging to blocks, spatial heterogeneity, etc.

SIR - Scale Free

In this experiment, we compare an ER network to Configuration Models with power law degree distributions.

Network parameters: $n = 1000$, $\langle k \rangle = 5$. We vary the exponent of the distribution, while keeping $\langle k \rangle = 5$ constant.

SIR parameters: $\theta = 0.2$, $\gamma = 0.5$. The initial number of infected nodes is 5, all of them in the same community structure.



The highest the exponent of the degree distribution, the faster is the diffusion.

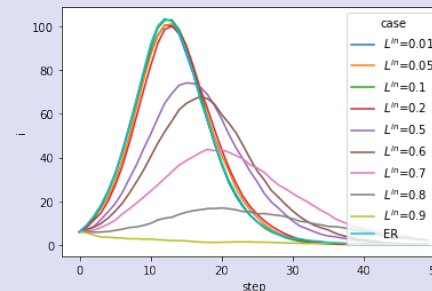
SIR - Community Structure

In this experiment, we compare an ER network to Stochastic Block Models.

Network parameters: $n = 1000$, $\langle k \rangle = 5$.

SBM parameters Number of blocks $|C| = 100$. We vary L^{in} , the fraction of all edges that are inside blocks. When $L^{in} = 0.01$, $p^{in} \approx p^{out} = 0.005$. When $L^{in} = 0.9$, $p^{in} = 0.5$, $p^{out} \approx 0.0005$

SIR parameters: $\theta = 0.2$, $\gamma = 0.5$. The initial number of infected nodes is 5, all of them in the same community structure.



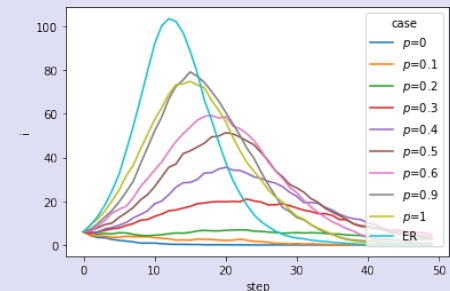
We observe that the more marked the communities, the less efficient the spreading process.

SIR - Spatial effect - WS

In this experiment, we compare an ER network to Watts Strogatz random graphs, varying the probability of rewiring edges. It can be understood as a model of spatial proximity: with $p = 0$, each node is connected only to its direct neighbors in the 1 dimensional space. If $p = 1$, each node is connected to exactly k random nodes.

Network parameters: $n = 1000$, $\langle k \rangle = 5$

SIR parameters: $\theta = 0.2$, $\gamma = 0.5$. The initial number of infected nodes is 5, being 5 direct neighbors.



The more nodes tend to be connected to direct neighbors in space, the slower the diffusion.

Application of diffusion models

Diffusion models can be used for several applications:

- Model fitting: better understand an actual epidemic by fitting parameters on real observations
- Predicting trends of evolution
- Control of epidemics: Given an epidemic model and a supporting network, find an optimal solution to control (accelerate or slow-down) the epidemic.

Optimal node/edge removal

One way to slow-down an epidemic consist in **removing nodes** (e.g., vaccination). The problem can be formulated as a budget constrained removal, i.e., if we can remove only x nodes/edges, which one should we choose? Based on theoretical and experimental results, heuristic solutions consist in removing: **highest degree nodes**, **highest betweenness** nodes/edges (isolating communities), long-distance edges (shortcuts) in spatial networks.

Friendship paradox and node removal

It has been proposed^a that the friendship paradox could be used to apply budget-constrained high degree nodes preferential vaccination in real networks where finding such nodes is not possible because the whole network is unknown: instead of targeting random individuals, one could vaccinate random contacts of random individuals, thus greatly increasing the average degree of vaccinated persons.

^aCohen, Havlin, and Ben-Avraham 2003.

Going further

Book Dynamic processes on Networks: Barrat, Barthelemy, and Vespignani 2008
Surveys:
Analysis and Control of Epidemics: Nowzari, Preciado, and Pappas 2016
Diffusion in networks: Lamberson 2016
Impact of community structure: Stegehuis, Van Der Hofstad, and Van Leeuwen 2016

References

- [1] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge university press, 2008.
- [2] Reuven Cohen, Shlomo Havlin, and Daniel Ben-Avraham. "Efficient immunization strategies for computer networks and populations". In: *Physical review letters* 91.24 (2003), p. 247901.
- [3] PJ Lamberson. "Diffusion in networks". In: *The Oxford handbook of the economics of networks*. Oxford University Press, Oxford (2016), pp. 479–503.
- [4] Cameron Nowzari, Victor M Preciado, and George J Pappas. "Analysis and control of epidemics: A survey of spreading processes on complex networks". In: *IEEE Control Systems Magazine* 36.1 (2016), pp. 26–46.
- [5] Clara Stegehuis, Remco Van Der Hofstad, and Johan SH Van Leeuwen. "Epidemic spreading on complex networks with community structures". In: *Scientific reports* 6.1 (2016), pp. 1–7.