Social and Graph Data Management: Spreading Phenomena

Pablo Mollá Chárlez

Contents

1 Introduction to Spreding Phenomena: Epidemics

On the night of **February 21, 2003** a physician from Guangdong Province in southern China checked into the Metropole Hotel in Hong Kong. He previously treated patients suffering from a disease that, lacking a clear diagnosis, was called atypical pneumonia. Next day, after leaving the hotel, he went to the local hospital, this time as a patient. He died there several days later of atypical pneumonia.

The physician did not leave the hotel without a trace: That night sixteen other guests of the Metropole Hotel and one visitor also contracted the disease that was eventually renamed Severe Acute Respiratory Syndrome, or SARS. These guests carried the SARS virus with them to Hanoi, Singapore, and Toronto, sparking outbreaks in each of those cities. Epidemiologists later traced close to half of the 8,100 documented cases of SARS back to the Metropole Hotel. With that the physician who brought the virus to Hong Kong become an example of a super-spreader, an individual who is responsible for a disproportionate number of infections during an epidemic.

A network theorist will recognize super-spreaders as hubs, nodes with an exceptional number of links in the contact network on which a disease spreads. As hubs appear in many networks, super-spreaders have been documented in many infectious diseases, from smallpox to AIDS. In this chapter, we introduce a network based approach to epidemic phenomena that allows us to understand and predict the true impact of these hubs. The resulting framework, that we call network epidemics, offers an analytical and numerical platform to quantify and forecast the spread of infectious diseases.

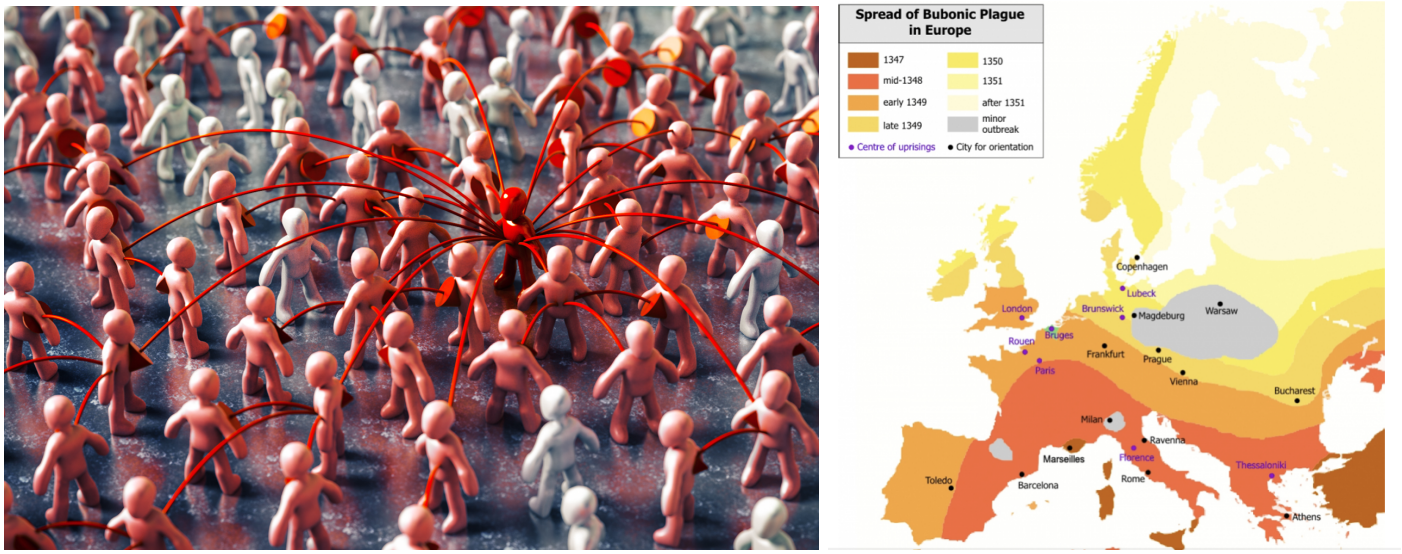

Figure 1: Super-spreaders in Epidemics

2 Modeling Spreading Phenomena

As previously mentioned, network epidemics provide a powerful framework for analyzing and forecasting the spread of infectious phenomena. These phenomena propagate through networks, facilitated by agents that transmit infections or information. The concept applies across a wide range of domains, from biology to digital systems and social behavior:

- **Biological epidemics**: The spread of pathogens like influenza, SARS, or COVID-19.
- **Digital infections**: The propagation of computer viruses and worms.
- **Social phenomena**: The spread of information, innovations, or memes within communication networks.

Different types of networks and agents are involved depending on the domain of the spreading phenomenon. Here's a mapping of the phenomena, their agents, and the networks that facilitate their spread:

Phenomenon	Agent	Network
Rumor spreading	Information, memes	Communication network
Innovation diffusion	Ideas	Communication network
Computer virus	Malware	Internet
Diseases	Pathogen	Human-human network
Bedbug infestations	Insects	Hotel-guest network

Table 1: Mapping of phenomena, agents, and networks.

2.1 Epidemic Modeling

Building on the concept of network epidemics, we delve deeper into the modeling approaches used to quantify and predict the spread of infectious phenomena. The spread of infections in networks often leads to cascading failures, as previously discussed, and this is especially prominent in epidemics, where the failure corresponds to the propagation of the disease or information. The modeling of epidemics in networks is based on two core assumptions:

- **Compartmentalization**

  Each individual (or node) in the network is classified into distinct compartments based on their state in the infection cycle. This classification simplifies the analysis and allows the development of models that capture the dynamics of disease spread. The primary compartments are:

  - **Susceptible (S)**: Healthy individuals who are not yet infected but are at risk of infection.
  - **Infectious (I)**: Contagious individuals who have contracted the infection and can spread it to others.
  - **Recovered (R)**: Individuals who were previously infected but have since recovered and are immune to reinfection (in most models).

  Compartmentalization enables the tracking of transitions between these states, which is fundamental for epidemic simulations like the **SIR model**.

- **Homogeneous Mixing**

  The second hypothesis assumes homogeneous mixing within the network, where any infectious individual has an equal probability of infecting any other susceptible individual. This assumption simplifies the calculations by treating the network as if all nodes interact uniformly, disregarding specific local structures or hubs in the network.

Bear in mind that, **homogeneous mixing** assumes equal probability of transmission for all infectious individuals, ignoring variations such as "super-spreaders" who carry and transmit a higher viral load. This simplification overlooks individual differences in infectivity, which can significantly influence real-world epidemic dynamics.

The core assumptions of epidemic models (compartmentalization and homogeneous mixing form the foundation of epidemic modeling by simplifying how infections spread in a population. While these assumptions allow for a clear framework, they abstract away the complexities of real-world interactions, such as the heterogeneity in individual behavior or the underlying contact network. To address these limitations, network-free models provide an alternative approach by focusing on population-level dynamics without explicit representation of network structures.

2.2 Network-Free Models

Network-free models use mathematical equations to describe infection dynamics in a well-mixed population, assuming that all individuals interact equally with one another. In the following, we will discuss three basic network-free models: the Susceptible-Infected (SI) Model, the Susceptible-Infected-Susceptible (SIS) Model and Susceptible-Infected-Recoverd (SIR) Model.

2.2.1 Susceptible-Infected (SI) Model

The SI model assumes a simple two-state system:

- Individuals are either Susceptible (S) (healthy and prone to infection) or Infected (I) (actively infected and contagious).
- Once infected, individuals remain infected indefinitely.

2.2.1.1 Infection Rate

Let us consider some assumptions and notations:

- **Notations**: N is the total number of individuals, each having $\langle k \rangle$ average contacts; $S(t)$ is the number of susceptible/healthy individuals at time t; $I(t)$ is the number of infected individuals at time t; β is the probability of infection during a contact (infection rate).
- **Assumptions**: Infection spreads over discrete time intervals and initially $S(0) = N$ and $I(0) = 1$.

The rate of change in the number of infected individuals, $I(t)$, is proportional to the product of the susceptible and infected populations:

$$\frac{\partial I(t)}{\partial t} = \beta \cdot \langle k \rangle \cdot \frac{S(t) \cdot I(t)}{N}$$

Since $S(t) + I(t) = N$, we rewrite $S(t) = N - I(t)$ and normalize $I(t)$ as $i(t) = \frac{I(t)}{N}$, meaning that:

$$S(t) = N - I(t) \iff \frac{S(t)}{N} = \frac{N - I(t)}{N} = 1 - i(t) \iff \frac{S(t)}{N} = 1 - i(t)$$

Substituting these, the equation becomes:

$$\frac{\partial I(t)}{\partial t} = \beta \cdot \langle k \rangle \cdot \frac{S(t) \cdot I(t)}{N} \iff \frac{\partial i(t)}{\partial t} = \beta \cdot \langle k \rangle \cdot \frac{S(t) \cdot I(t)}{N^2} = \beta \cdot \langle k \rangle \cdot [\frac{S(t)}{N}] \cdot [\frac{I(t)}{N}] = \beta \cdot \langle k \rangle \cdot [1 - i(t)] \cdot [i(t)]$$

Solving for $i(t)$ and using separation of variables and integrating, the solution is:

$$i(t) = \frac{i(0) \cdot e^{\beta \langle k \rangle t}}{1 - i(0) + i(0) \cdot e^{\beta \langle k \rangle t}}$$

Here, $\beta\langle k \rangle$ is called the transmission rate, which determines how quickly the infection spreads.

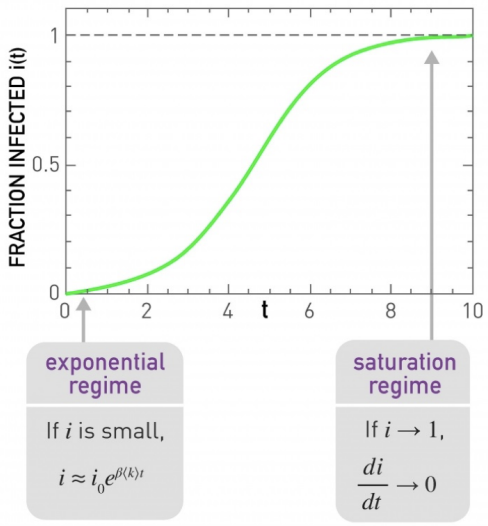

Figure 2: Progression of the fraction of infected individuals i(t) over time in the SI Model

As shown in the previous graph 2, we can distingush two main phases:

1. **Exponential Regime**: In the early stages of the infection, the fraction of infected individuals is small ($i \ll 1$). In this phase, the infection grows exponentially as:

$$i(t) \approx i(0) \cdot e^{\beta\langle k\rangle t}$$

   Where, $i(0)$ is the initial fraction of infected individuals, and $\beta\langle k\rangle$ is the transmission rate.

2. **Saturation Regime**: As $i(t)$ approaches 1 (i.e., most individuals are infected $\sim i(\infty) = 1$), the rate of new infections slows down significantly, causing the curve to plateau. In this phase, the change in the fraction of infected individuals over time becomes negligible ($\frac{\partial i}{\partial t} \to 0$).

This transition from exponential growth to saturation reflects the dynamics of an infection spreading in a population without recovery or immunity.

2.2.1.2 Characteristic Time

The characteristic time (τ) is the time it takes for the fraction of infected individuals to reach approximately $1/e$ ($\sim$36%) of all susceptible individuals. In other terms, the characteristic time measures how long it takes for the infection level $i(t)$ to get close to its steady state, which is given by:

$$\tau = \frac{1}{\beta\langle k\rangle}$$

The characteristic time τ is used as the main parameter for comparison in network-based epidemiological models because it succinctly quantifies the speed of the disease spread, irrespective of the absolute fraction of infected individuals. This is particularly useful for comparing networks with vastly different structures and infection dynamics.

2.2.2 Susceptible-Infected-Susceptible (SIS) Model

The Susceptible-Infected-Susceptible (SIS) model extends the SI model by allowing individuals to recover from the infection and return to the susceptible state at a recovery rate μ. This recovery creates a dynamic balance between infection and recovery.

2.2.2.1 Infection Rate

In this model, the rate of change is given by:

$$\frac{\partial i}{\partial t} = \beta\langle k\rangle \cdot i(t) \cdot [1 - i(t)] - \mu \cdot i(t)$$

Where $\beta\langle k\rangle \cdot i(t) \cdot [1 - i(t)]$ represents the infection spread and $\mu \cdot i(t)$ represents recovery. Solving the previous differential equation leads to:

$$i(t) = \left(1 - \frac{\mu}{\beta\langle k\rangle}\right) \cdot \frac{C \cdot e^{(\beta\langle k\rangle - \mu)t}}{1 + C \cdot e^{(\beta\langle k\rangle - \mu)t}}, \text{ where } C = \frac{i(0)}{1 - i(0) - \frac{\mu}{\beta\langle k\rangle}}$$

The following plot 3 illustrates the dynamics of the SIS model. It shows how the fraction of infected individuals $i(t)$ evolves over time, which can be split into two phases:

1. **Exponential Outbreak**: When the fraction of infected individuals (i) is small, the infection grows exponentially. This is modeled as:

$$i \approx i_0 e^{(\beta\langle k\rangle - \mu)t}$$

   Here, $\beta\langle k\rangle$ is the infection rate, and μ is the recovery rate.

2. **Endemic State**: Over time, the fraction of infected individuals stabilizes at a constant value:

$$i(\infty) = 1 - \frac{\mu}{\beta\langle k\rangle}$$

This plateau occurs when the infection and recovery processes balance out, creating a persistent infection level, where not everyone is infected but more than half.

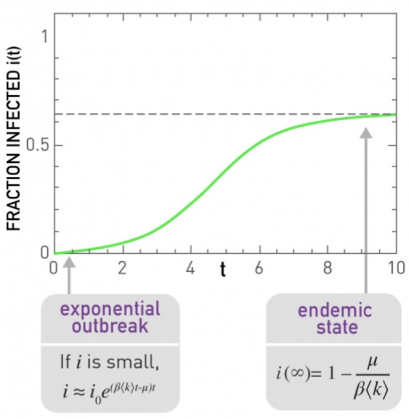

Figure 3: Progression of the fraction of infected individuals i(t) over time in the SIS Model

Note

- If $\beta\langle k\rangle > \mu$, the system reaches an Endemic State with a non-zero infection level.
- If $\beta\langle k\rangle \leq \mu$, the disease would instead die out over time (not shown in this plot), also known as Disease-Free State.

2.2.2.2 Characteristic Time and Basic Reproductive Number

The SIS model introduces the basic reproductive number $R_0 = \frac{\beta\langle k\rangle}{\mu}$, which determines the epidemic outcome, i.e., the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection. In other words, R_0 is the number of new infections each infected individual causes under ideal circumstances.

- If $R_0 > 1$: The epidemic reaches an endemic state.
- If $R_0 \leq 1$: The disease-free state is achieved.

The characteristic time τ for the system to reach equilibrium depends on R_0:

$$\tau = \frac{1}{\mu(R_0 - 1)}$$

Disease	**Transmission Mode**	R_0
Measles	Airborne	12-18
Smallpox	Social contact	5-7
HIV	Sexual contact	2-5
SARS	Airborne droplet	2-5
Influenza 1918	Airborne droplet	2-3

Table 2: Basic Reproductive Number and Disease Transmission

2.2.3 Susceptible-Infected-Recovered (SIR) Model

The Susceptible-Infected-Recovered (SIR) Model is an extension of the SI and SIS models, where individuals can move from being infected (I) to recovered (R) at a recovery rate μ. Once recovered, individuals are immune and cannot be reinfected. Reconsidering the compartmentalization previously mentioned the system is described by three coupled differential equations:

$$\frac{\partial s}{\partial t} = -\beta \cdot \langle k \rangle \cdot s(t) \cdot i(t); \quad \frac{\partial i}{\partial t} = \beta \cdot \langle k \rangle \cdot s(t) \cdot i(t) - \mu \cdot i(t); \quad \frac{\partial r}{\partial t} = \mu \cdot i(t)$$

Where:

- $s(t)$, $i(t)$, and $r(t)$ represent the fraction of susceptible, infected, and recovered individuals, respectively.
- $\beta\langle k \rangle$: Transmission rate (average contacts multiplied by infection probability).
- μ: Recovery rate.

2.2.3.1 Infection Rate and Properties

In the SIR Model, we observe that:

1. **No Closed Form Solution for** $i(t)$: The infection rate depends on $s(t)$, making it a non-linear system. This makes solving for $i(t)$ analytically difficult.
2. **Final State** $(i(\infty) = 0)$: Eventually, all individuals are either susceptible or recovered, and the infection dies out.
3. **Basic Reproductive Number** $(R_0 = \frac{\beta\langle k \rangle}{\mu})$: The disease spreads if $R_0 > 1$; otherwise, it dies out.

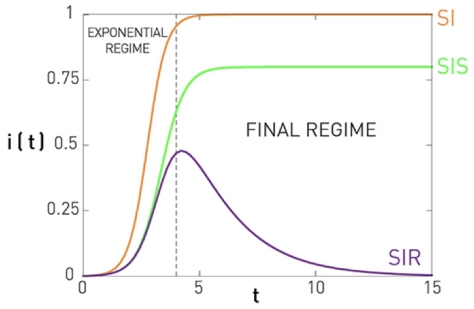

Figure 4: Rate of Change of Infected Individuals: SI vs. SIS vs. SIR

From the previous graph 4, and with respecto to the SIR Model, we observe an exponential growth followed by a decline. At the beginning, the infection grows exponentially if $R_0 > 1$. Later, it declines as $s(t)$ decreases. Unlike SI or SIS, the SIR Model does not define a simple time constant because recovery and infection rates interact in a more complex way.

3 Epidemics on Networks

In our previous discussions, we focused on **mean-field models** like SI, SIS, and SIR, where we only used the average degree ($\langle k \rangle$) of the network to describe the spread of a disease. These models assume that all nodes behave identically, which may oversimplify real-world scenarios, especially in networks where node connectivity (degree) is highly heterogeneous, such as scale-free networks. Real networks often have a complex structure:

- Some nodes (e.g., hubs) have far more connections than others.
- The assumption that all nodes are statistically equivalent is invalid for these networks.

To account for network structure, degree-block approximation is introduced, which incorporates the degree distribution into the dynamics of disease spread.

	SI	SIS	SIR
Exponential Regime: Number of infected individuals grows exponentially	$i = \frac{i_0 e^{\beta\langle k\rangle t}}{1 - i_0 + i_0 e^{\beta\langle k\rangle t}}$	$i = \left(1 - \frac{\mu}{\beta\langle k\rangle}\right) \frac{C e^{(\beta\langle k\rangle - \mu)t}}{1 + C e^{(\beta\langle k\rangle - \mu)t}}$	No closed solution
Final Regime: Saturation at $t \to \infty$	$i(\infty) = 1$	$i(\infty) = 1 - \frac{\mu}{\beta\langle k\rangle}$	$i(\infty) = 0$
Epidemic Threshold: Disease does not always spread	No threshold	$R_0 = 1$	$R_0 = 1$

Figure 5: Summary of Properties: SI vs. SIS vs. SIR

3.1 Degree-Block Approximation

The degree-block approximation categorizes nodes based on their degree k, assuming that all nodes with the same degree behave statistically similarly. This approach helps us model how a disease spreads differently across nodes with varying connectivity.

The idea behind is that nodes with higher degrees are more likely to get infected because they have more connections, making them more exposed to the disease. Let us define some of the notation that will be used in the upcoming sections:

- $i_k(t)$: Fraction of infected nodes with degree k at time t.
- N_k: Total number of nodes with degree k.
- p_k: Degree distribution of the network, i.e., the fraction of nodes with degree k. It satisfies $\sum_k p_k = 1$.
- $\langle k \rangle$: Average degree of the network.
- $\langle k^2 \rangle$: Second moment of the degree distribution.

3.2 Susceptible-Infected (SI) Model

In the SI model with degree-block approximation, the infection dynamics are written separately for nodes of degree k.

3.2.1 Infection Rate for Degree k

The infection rate is proportional to the transmission rate β and the fraction of degree-k nodes that are not yet infected, which is $(1 - i_k)$, multiplied by each node's actual degree k and the density function Θ_k which represents the fraction of infected neighbors of a susceptible node k, resulting in the following equation:

$$\frac{\partial i_k}{\partial t} = \beta(1 - i_k) \cdot k \cdot \Theta_k$$

Solving the above equation gives the fraction of infected nodes with degree k over time:

$$i_k(t) = i_k(0) \left(1 + \frac{k \cdot (\langle k \rangle - 1)}{\langle k^2 \rangle - \langle k \rangle} \cdot (e^{t/\tau^{SI}} - 1)\right)$$

where the characteristic time τ for the SI model is $\tau^{SI} = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}$. And, the total fraction of infected nodes is:

$$i(t) = \int_0^{k_{\max}} i_k \cdot p_k \, dk = i_k(t) = i_k(0) \left(1 + \frac{\langle k \rangle^2 - \langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \cdot (e^{t/\tau^{SI}} - 1)\right)$$

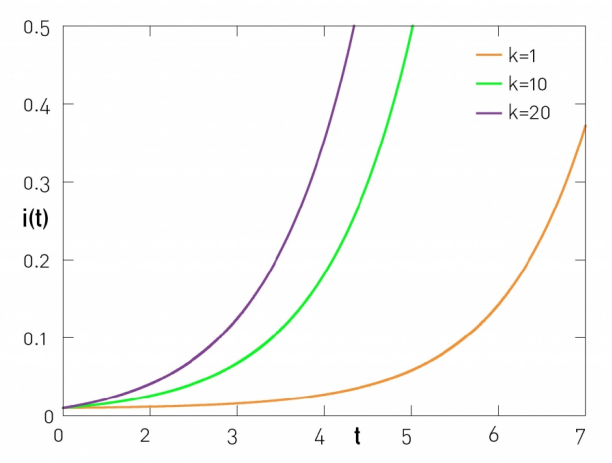

Figure 6: Susceptible-Infected Model with different Node Degrees

3.2.2 Spread Dynamics of SI Model Based on Network Type

Let us recall the different types of networks we have studied until now and apply the SI Model on them. Remember that, the key parameter for a valuable comparison between differents network models is the characteristic time τ.

3.2.2.1 SI Model on Random Networks

Random Networks: For random networks, $\langle k^2 \rangle = \langle k \rangle (\langle k \rangle + 1)$. The dynamics are similar to the homogeneous case (classic SI model) because the disease spreads with a finite characteristic time τ^{SI} (they share the same characteristic time but most importantly it's a finite value):

$$\tau^{SI} = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)} = \frac{\langle k \rangle}{\beta([\langle k \rangle(\langle k \rangle + 1)] - \langle k \rangle)} = \frac{1}{\beta \cdot \underbrace{\langle k \rangle}_{finite}} < \infty$$

3.2.2.2 SI Model on Scale-Free Networks

We need to distinguish both cases:

- Scale-Free Networks ($\gamma \geq 3$): In scale-free networks with $\gamma \geq 3$, both $\langle k \rangle$ and $\langle k^2 \rangle$ are finite. The disease spreads with a finite τ^{SI}, similar to random networks:

$$\tau^{SI} = \frac{\overbrace{\langle k \rangle}^{finite}}{\beta(\underbrace{\langle k^2 \rangle}_{finite} - \underbrace{\langle k \rangle}_{finite})} < \infty$$

- Scale-Free Networks ($\gamma < 3$): In scale-free networks with $\gamma < 3$, $\langle k^2 \rangle$ diverges due to the presence of high-degree hubs and $\langle k \rangle$ is finite. . As $\langle k^2 \rangle \to \infty$, the characteristic time $\tau^{SI} \to 0$. In other words, the spread of a pathogen on a scale-free network is instantaneous:

$$\tau^{SI} = \frac{\overbrace{\langle k \rangle}^{finite}}{\beta(\underbrace{\langle k^2 \rangle}_{infinite} - \underbrace{\langle k \rangle}_{finite})} \to 0$$

  The divergence of $\langle k^2 \rangle$ (the second moment of the degree distribution represents the variability or spread in the degrees of nodes), means that hubs (nodes with very high degrees) dominate the dynamics. Therefore, hubs are almost instantly infected, and they quickly spread the disease to the rest of the network. This leads to instantaneous spread in the limit of infinite hubs.

3.3 Susceptible-Infected-Susceptible (SIS) Model

The SIS Model differs from the SI Model in the fact that infected individuals can recover and return to a susceptible state, governed by the recovery rate μ. The dynamics for nodes with degree k are described by the equation:

$$\frac{\partial i_k}{\partial t} = \beta(1 - i_k(t)) \cdot k \cdot \Theta_k(t) - \mu \cdot i_k(t)$$

Where:

- β: Transmission rate.
- μ: Recovery rate.
- i_k: Fraction of degree-k nodes that are infected.
- $\Theta_k(t)$: Probability that a neighbor of a degree-k node is infected.

The recovery term $(-\mu \cdot i_k(t))$ introduces an additional dynamic absent in the SI Model, where infections can now decay over time, depending on the interplay of β, μ, and network structure.

3.3.1 Characteristic Time

The characteristic time τ for the SIS Model is given by:

$$\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \mu \langle k \rangle}$$

If μ (recovery rate) is sufficiently large, $\tau < 0$, meaning infections decay exponentially, and the pathogen dies out. The decay condition depends on both the network structure (captured by $\langle k \rangle$ and $\langle k^2 \rangle$) and the biological properties of the pathogen (through β and μ).

3.3.2 Spreading Rate λ and Epidemic Threshold λ_c

To predict whether a pathogen persists, we define the spreading rate:

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

Simply, we consider two cases:

- If $\lambda > \lambda_c$, the infection spreads and persists.
- If $\lambda < \lambda_c$, the pathogen dies out, leading to a disease-free state ($i = 0$).

3.3.3 Spread Dynamics of SIS Model on Network Type

3.3.3.1 SIS Model on Random Networks

For random networks, using $\langle k^2 \rangle = \langle k \rangle (\langle k \rangle + 1)$, the epidemic threshold is:

$$\lambda_c = \frac{1}{\langle k \rangle + 1}$$

Key points:

- λ_c is always non-zero for random networks because $\langle k \rangle$ is finite.
- If $\lambda > \lambda_c$, the infection persists in an endemic state, meaning a fraction of the population is always infected.
- If $\lambda < \lambda_c$, the pathogen dies out completely.

3.3.3.2 SIS Model on Scale-Free Networks

For scale-free networks with arbitrary degree distributions, the epidemic threshold is:

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

In scale-free networks with degree distribution ($P(k) \sim k^{-\gamma}$), for $\gamma < 3$, $\langle k^2 \rangle \to \infty$, leading to $\lambda_c \to 0$. This is called the vanishing epidemic threshold.

What is the vanishing epidemic threshold? It means that, no matter how small the spreading rate λ becomes, there is no critical value λ_c above which the infection persists. Any small $\lambda > 0$ can lead to the spread of the infection.

Why does the infection spread despite $\lambda_c \to 0$? In scale-free networks, highly connected nodes (hubs) significantly influence the spread. These hubs have many connections and are more likely to both become infected and spread the pathogen. Even when λ is very small, the hubs can sustain the infection and allow it to propagate throughout the network.

Note The vanishing epidemic threshold in scale-free networks is a hallmark of their heterogeneity. While random networks have a finite threshold that depends on their average connectivity, scale-free network's hubs ensure that even weakly contagious diseases can persist and spread.

3.4 Extra Key Points

We need to remind as well that degree correlations and community structures both influence how and where information or infections spread. Besides, the category of the network (Assortative, Dissortative and Neutral) strongly influences the spread. In particular, Assortativity intensifies local spreading, particularly in scale-free networks where hubs dominate, while weak ties in communities can slow or isolate the spread. Moreover, the type of contagion (simple vs. complex) determines whether the spread relies on single or multiple sources, emphasizing the interplay between network topology and transmission dynamics.

- **Effect of Degree Correlations**: Degree correlations (e.g., assortativity) affect the epidemic threshold λ_c. Assortativity (nodes preferentially connecting to similar-degree nodes) lowers λ_c for the SIS Model (as $\langle k^2 \rangle$ is larger), making it easier for a disease or information to spread. In scale-free networks, λ_c still vanishes, regardless of degree correlations, because hubs dominate the spread. Assortativity accelerates the spread since hubs are infected first and preferentially interact with other high-degree nodes.

- **Effect of Communities on Information Spread**: Strong ties within communities leads information or the infection to spread rapidly inside communities due to close and frequent interactions. However, weak ties between communities causes the infection/information to have difficulties escaping communities, leading to localized spread and slower global propagation.

- **Types of Information Spread** We can distinguish between two possible types of information spread:
  - **Simple contagion**: Information or infection spreads through a single contact (studied in models like SI, SIS and SIR).
  - **Complex contagion**: Information or infection requires reinforcement from multiple sources before spreading, making the dynamics more dependent on network structure and connections.

4 Network Immunization

Immunization strategies determine how vaccines, treatments, or drugs are distributed within a population. Ideally, if a treatment or vaccine is available, it would be administered to all infected individuals or those at risk of infection. However, factors such as costs, challenges in reaching all at-risk individuals, and real or perceived side effects often prevent full coverage. Given these limitations, immunization strategies focus on effectively distributing the available resources to minimize the risk of a pandemic.

Traditional epidemic models predict that reducing a pathogen's spreading rate λ below the critical threshold λ_c causes the infection to naturally die out . However, in scale-free networks, the epidemic threshold λ_c vanishes, raising doubts about the effectiveness of this approach. If λ_c is zero, it becomes impossible to reduce λ below the threshold.

This section explores immunization strategies to control the spread of pathogens in networks. First, we discuss random immunization, examining its effectiveness in both **homogeneous** (random networks) and **heterogeneous** (scale-free networks) topologies. We demonstrate how the fraction of immunized nodes needed to stop the spread depends on the network structure, highlighting the limitations of random immunization in scale-free networks due to their high heterogeneity. Next, we introduce selective immunization, a targeted approach designed to address the vanishing epidemic threshold in scale-free networks by prioritizing the immunization of hubs. This strategy leverages the network's topology to disrupt connectivity and effectively curb the spread of infection.

4.1 Random Immunization

The goal is to immunize a fraction g_c of nodes so that the spreading rate $\lambda = \frac{\beta}{\mu}$ is reduced below the epidemic threshold λ_c. Immunized nodes are considered "**invisible**," meaning they cannot contract or spread the infection. As it can be expected from its name, Random immunization involves selecting nodes at random to immunize. Let's study how it would work on Homogeneous ($\sim$ Random Networks) and Heterogeneous ($\sim$ Scale-Free Networks).

4.1.1 Homogeneous Networks: Random Networks

For random networks, the epidemic threshold is:

$$\lambda_c = \frac{1}{\langle k \rangle + 1}$$

To ensure $\lambda < \lambda_c$, we need:

$$\frac{(1 - g_c)\beta}{\mu} = \frac{1}{\langle k \rangle + 1}$$

The left term, $\frac{(1-g_c)\beta}{\mu}$, represents the **reduced spreading rate** λ after immunizing a fraction g_c of nodes. It must be equal to the epidemic threshold $\lambda_c = \frac{1}{\langle k \rangle + 1}$ because this is the condition for the disease to no longer spread in the network. Ensuring $\lambda < \lambda_c$ means the pathogen cannot sustain transmission, leading to its eventual extinction. Therefore, solving for g_c:

$$g_c = 1 - \frac{\mu}{\beta(\langle k \rangle + 1)}$$

The fraction g_c of immunized nodes increases when β (infection probability) increases and/or the recovery rate μ decreases. This means that the higher the infection rate or lower the recovery rate, the larger the fraction of nodes that must be immunized. However, g_c is always less than 1, meaning full immunization of all nodes is not necessary to stop the spread.

4.1.2 Heterogeneous Networks: Scale-Free Networks

In scale-free networks, the epidemic threshold is:

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

To ensure $\lambda < \lambda_c$, we need to solve the following equation:

$$\frac{(1 - g_c)\beta}{\mu} = \frac{\langle k \rangle}{\langle k^2 \rangle} \longleftrightarrow g_c = 1 - \frac{\mu \langle k \rangle}{\beta \langle k^2 \rangle}$$

Once again, g_c depends on the ratio $\langle k \rangle / \langle k^2 \rangle$. In heterogeneous networks, $\langle k^2 \rangle$ is much larger than $\langle k \rangle$, leading to a very small fraction. This implies that g_c approaches 1, meaning almost the entire network must be immunized

to stop the spread. For scale-free networks with $\gamma < 3$, $\langle k^2 \rangle$ diverges, making $g_c \to 1$. In such networks, random immunization is ineffective because the hubs are not targeted, and these hubs play a disproportionately large role in spreading the infection.

4.2 Selective Immunization

Selective immunization targets hubs to disrupt the network's connectivity. This approach is particularly effective in scale-free networks because the **epidemic threshold vanishes** ($\lambda_c \to 0$). Selective immunization uses a key strategy called the friendship paradox, which states that "your friends are more popular than you" (i.e., the neighbors of a randomly chosen node tend to have a higher degree).

4.2.1 Steps for Implementation

1. **Random Selection**: Choose a random fraction p of nodes from the network (Group 0).
2. **Neighbor Selection**: For each node in Group 0, randomly select one of its neighbors. Place the selected neighbors into Group 1.
3. **Immunization**: Immunize all nodes in Group 1.

Why It Works? Nodes in Group 1 (neighbors of Group 0) have a higher average degree than those in Group 0, due to the friendship paradox. This ensures that hubs are preferentially targeted for immunization, even without explicit knowledge of the network's structure. By immunizing hubs, the network's connectivity is disrupted, reducing the likelihood of disease spreading and effectively compensating for the vanishing epidemic threshold. **Key Insights** to remember:

- Random immunization is effective in random networks but fails in scale-free networks because it does not prioritize hubs.
- Selective immunization specifically targets hubs, addressing the challenges posed by the vanishing epidemic threshold in scale-free networks.

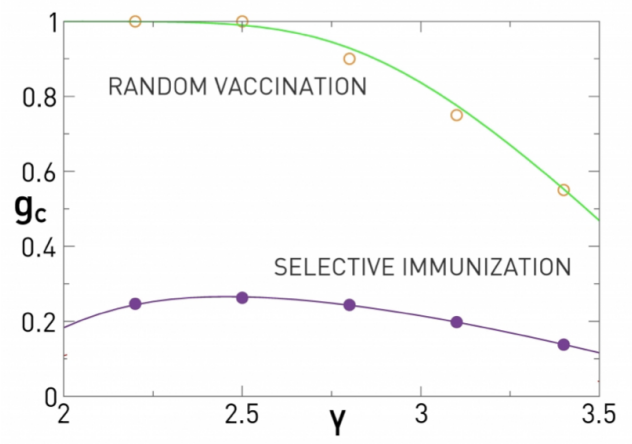

Figure 7: Random vs. Selective Immunization in Scale-Free Networks

The previous graph compares the critical fraction of nodes (g_c) that must be immunized to stop the spread of a disease for random vaccination and selective immunization strategies, as a function of the degree exponent γ in scale-free networks.

- Random Vaccination: Requires a much higher g_c (close to 1 for $\gamma \approx 2$) because hubs are not targeted, making the network still vulnerable to the disease's spread.
- Selective Immunization: Achieves a much lower g_c (0.2) by targeting hubs, effectively breaking the network's structure and preventing the spread even for low γ.