

## Introduction

The network used in this work was obtained from a GitHub repository under the name Data for Good Lab, which, among many things, is dedicated to collecting datasets to share with the general public. [1] [2]

The downloaded file, which corresponds to research conducted by Fire, Puzis & Elovici (2013) [2], contains a series of 6 datasets in comma-separated value format. Among those 6 files, there are 2 of each company size, namely:

1. Small.
2. Medium.
3. Large.

Given the way in which the data were obtained (web crawling), the authors decided to anonymize the data, which is why it is impossible for me to provide information about the context of the network. However, considering the purpose of this paper, I believe that the information obtained in the dataset is sufficient.

The motivation of this project is to analyze how different characteristics of a pathogen can influence its ability to reproduce and/or to last longer. Likewise, we analyze how the use of vaccines, drugs and other methods of prevention/cure/immunity allow pathogens to be less dangerous to humans.

This and other projects studying the spread of diseases allow us to know the possible future behaviors at the societal level in order to take preventive action.

## Epidemics and Network Science

Antomaly (2014) provides the definition of epidemic in the first edition of the Oxford Companion to Medicine:

*“An outbreak of disease such that for a limited period of time a significantly greater number of persons in a community or region are suffering from it than is normally the case.... The extent and duration of an epidemic are determined by the interaction of such variables as the nature and infectivity of the causal agent, its mode of transmission, the degree of preexisting and newly acquired immunity, etc.” [3]*

Epidemics, as described in [4], spread through contact networks. In this case, we are dealing with a physical contact network, since its members had face-to-face interactions.

For its study, network science makes use of mathematical models capable of describing the behavior of infected, healthy, and recovered individuals. Barabasi explains 3 of the most commonly used models [5]:

- SI: Susceptible – Infected Model.
- SIS: Susceptible – Infected - Susceptible Model.
- SIR: Susceptible – Infected - Recovered Model.

Where:

- S: Healthy individuals who have not yet contacted the pathogen.
- I: Contagious individuals who have contacted the pathogen and hence can infect others.
- R: Individuals who have been infected before, but have recovered from the disease, hence are not infectious.

**Note:** Some authors may refer the R in SIR as *Removed*. \*

### Basic Characteristics of the Network

The network consists of 1429 nodes and 19357 links. The average degree of nodes is approximately 27, meaning that, on average, each person within the company has contact with 27 people.

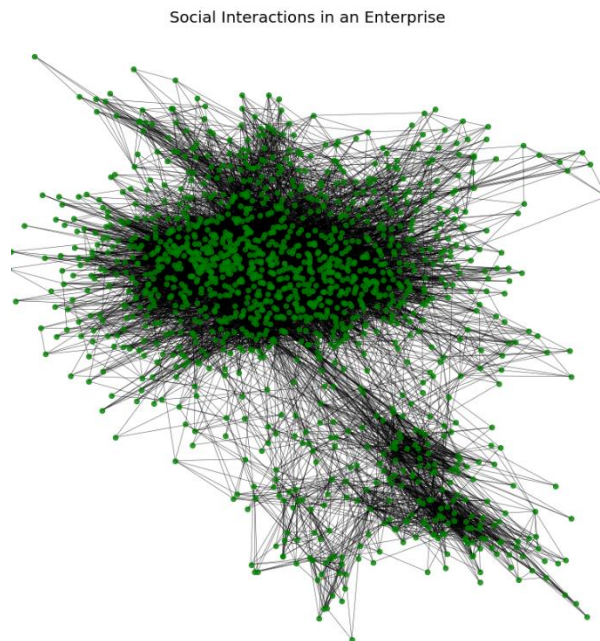


Figure 1. Network of Social Interaction within an Enterprise

Other basic metrics of interest are shown in the following table.

Metric	Value
Degree Centrality	438, 137, 489, 371, 487, 547, 640, 100, 1004, 877
Density	0.0189
Average Path Length	3.036
Clustering Coefficient	0.4211
Average Degree <K>	27
Nodes	1429
Links	19357

Note that the centrality measure is simply the number of direct links from the nodes. I decided to use this metric because it seemed to me to be a good indication of who are the individuals who, should they become ill, can infect the greatest number of people.

As you would expect from a real-world network, it is sparse, which can be confirmed with the degree distribution.

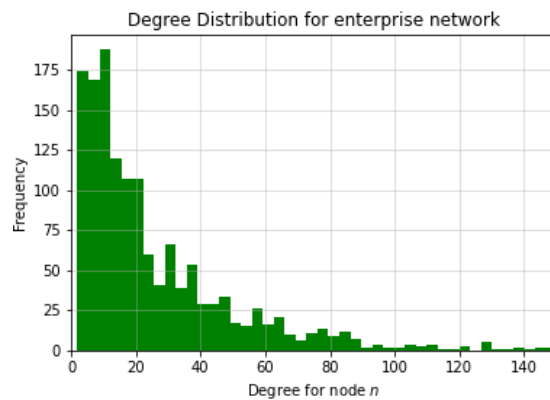


Figure 2. Degree Distribution for the network

## Model

This report makes use of the SIR model, so I am considering that patients may recover (or die) after suffering from the disease. Since the aim is to have a simulation as close to reality as possible, the parameters used throughout the models were taken from various research studies on Covid 19.

### Parameter estimation

The calculation and derivation of the most important parameters for modeling the epidemic in the SIR model was a combination of research and calculations done by myself, trying to get the most recent data possible.

For  $\mu$  calculation, we consulted the data available on the Statista website (intellectual property of John Elflein), which updates data constantly [7]. The figures show that as of August 4, 2021:

- There are 200,304,173 confirmed cases.
- 180,563,688 have recovered from the disease.

The recovery rate is the expressed ratio of the number of persons recovered to the number of total infected persons, which results in 0.901, which stands for 90.1% of people who get infected, recover from the disease. The calculation appears to agree with the data recorded in Mexico (and recollected by *worldometers* portal) as of August 5, 2021, which reports a recovery rate of 90.27%. [8]

Although, some pages in the medical area indicate that the recovery rate is even higher [9], I have decided to use the datum calculated and confirmed by other sources.

An article from the Global Barcelona Institute (published in May 2021) cites WHO estimates that the transmission rate ( $R_0$ ) of coronavirus varies between 1.4 and 2.5, although some estimates place this value in the 2.5 to 3 range [6]. The formula for its calculation is:

$$R_0 = \frac{\beta \langle k \rangle}{\mu}$$

Figure 3. Formula for  $R_0$

Where  $\beta$  is the likelihood that the disease will be transmitted from an infected to a susceptible individual in a unit time. [5] Fortunately, all the values are available to calculate the value of  $\beta$ , which turns out to be 0.083.

For the sake of simplicity, I have decided to take the average of the last two estimates, resulting in an  $R_0$  of 2.5.

The variable  $\tau$  is the characteristic time required to reach a  $\frac{1}{e}$  fraction (about 36%) of all susceptible individuals [5]. Its formula is shown below.

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

Figure 4. Formula for the characteristic time to reach 36% of the population

Thanks to having all the variables for its calculation, it was possible to obtain that  $\tau$  is 0.739.

The variable  $\rho$ , which represents the initial fraction of infected individuals, was retrieved from a study analyzing Covid-19 transmission dynamics from a data perspective. It should be mentioned that the rho value was not calculated by the authors of this study but was taken from the findings of other researchers. [10] The proposed value is 0.061.

The following table shows the parameter values required for the modeling.

Variable	Meaning	Value
$\mu$	Recovery rate.	0.901
$R_0$	Transmission rate.	2.5
$\beta$	Likelihood that the disease will be transmitted from an infected to a susceptible individual in a unit time.	0.083
$\tau$	Characteristic time required to reach a $\frac{1}{e}$ fraction (about 36%) of all susceptible individuals.	0.739
$\rho$	Initial fraction of infected individuals.	0.061

### First simulation: A Naive Approach

The data in the table were used as parameters for a simulation with the SIR model using a specialized Python module called EoN (Epidemics on Networks). The animation in [this link](#) shows how a pathogen with the properties of Covid-19 spreads in a work environment.

Before looking at the simulation, and with only the data collected so far, one can expect the pathogen to spread rapidly among the people in the company. This is justified by the fact that  $R_0$  has a relatively large value, in addition to being greater than 1. Although this is a disease that has a recovery rate close to 90% and  $\beta$  does not represent a high level of contagion (at least not in this strain), we must not forget how connected the network is.

Note that the contagion peak is reached in two-time units. It would be correct to think that the time units are represented in weeks. It is vital to emphasize that this simulation does not consider the use of masks or control of social distance, so it is to be expected that the virus will go viral at a high speed.

## Second Simulation: How Masks Help Prevent the Spread of a Disease

There are articles stating that the use of mouth covers by 50% of the population (assuming a 50% effective mouth cover) reduces the  $R_0$  value from 2.4 to 1.35, making the spread comparable to that of seasonal influenza [11].

Considering the formula for  $R_0$  given in *Figure 3*, We can do the math and get a new value for the  $\beta$  variable.

The average degree remains the same, as does the recovery rate of the disease, since they are intrinsic to the network and the disease, respectively. On the other hand,  $\beta$  can decrease or increase to the extent that preventive measures are taken, such as the use of masks.

The following lines shows how to compute the new  $\beta$ :

$$\begin{aligned} R_0 &= \frac{\beta \langle k \rangle}{\mu} \\ 1.35 &= \frac{(\beta)(27)}{0.90} \\ \beta &= \frac{(1.35)(0.90)}{27} \\ \beta &= 0.045 \end{aligned}$$

A simulation is then run again, using the newly calculated value for  $\beta$ .

It can be expected that the peak of contagion will be reached later, as the virus has difficulty to be transmitted. The animation in [this link](#) shows the curves for susceptible, infected, and recovered individuals.

## Third Simulation: Is Lockdown Really a Good Strategy to Prevent Disease Transmissions?

Studies in some countries show an  $R_0$  of 0.77 and a  $\beta$  of 0.055 in places where confinement was applied.[12] The creation of the model with these parameters is only intended to check if the work done so far follows a commonsense line.

Since the new value of  $\beta$  is similar to what it is when using the mouth covers, I did not expect major changes in the model. The maximum number of infections was reached in week 5, as in the previous model. No major changes were observed. The animation of the model can be accessed through [this link](#).

## Results

The graphs of susceptible, infected and recovered individuals respectively at a time interval of up to 17 weeks for different values of  $\beta$  are shown below.

- For  $\beta = 0.083$ .

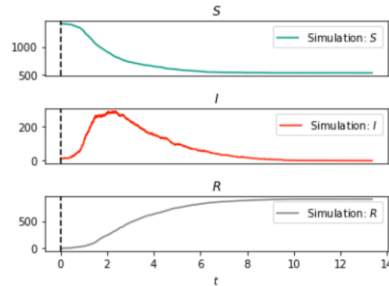


Figure 5. SIR graphs for  $\beta = 0.083$

- For  $\beta = 0.045$ .

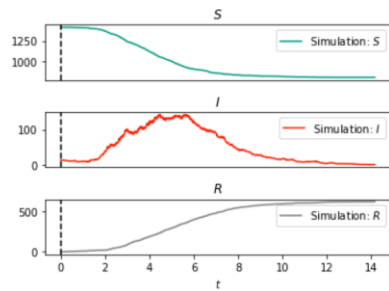


Figure 6. SIR graphs for  $\beta = 0.045$

- For  $\beta = 0.055$ .

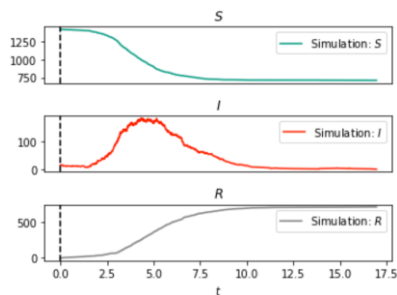


Figure 7. SIR graphs for  $\beta = 0.055$

Despite being a difference of only one and a half to two weeks, the last two curves show a less abrupt behavior. It can be seen that the peak in the second graph is much lower than that of the first graph, being almost half that of the first graph.

Complementarily, and as expected, the susceptibility curve tends to fall more smoothly as the value of  $\beta$  decreases.

The SIR plots with preventive methods have similar values, which is why there are not many differences in the models. Although a difference in  $\beta$  of 0.10 units seems small, a large difference can be observed in the number of infections at the peak of the epidemic, demonstrating that this is a parameter that should be minimized at any cost.

## Conclusions and Future Work

The models analyzed demonstrate the impact of preventive measures against disease transmission. Correct adherence to these measures, in addition to other measures not analyzed in this paper (such as social distancing and vaccinations) help to maintain stability in the number of susceptible, infected, and recovered persons, avoiding emergencies and overcrowding in the public health service.

Sadly, the present work does not distinguish between different strains of Covid-19; however, it would be interesting to analyze how more contagious strains spread in contact networks. Similarly, it is of interest to study the evolution of the pandemic once a large percentage of people have been vaccinated.



## References:

- [1] "Data Science for Social Good Lab - Datasets," *Github.io*. [Online]. Available: <https://data4goodlab.github.io/dataset.html>. [Accessed: 06-Aug-2021].
- [2] M. Fire and R. Puzis, "Organization mining using online social networks," *Netw. Spat. Econ.*, vol. 16, no. 2, pp. 545–578, 2016.
- [3] J. Anomaly, "What is an epidemic?," *The Journal of law, medicine & ethics: a journal of the American Society of Law, Medicine & Ethics*, vol. 42, p. 389+, 2014.
- [4] F. Menczer, S. Fortunato, and C. A. Davis, *A First Course in Network Science*. Cambridge, England: Cambridge University Press, 2020.
- [5] A.-L. Barabasi, *Network Science*. Cambridge, England: Cambridge University Press, 2016.
- [6] "ISGlobal - Coronavirus," *Isglobal.org*. [Online]. Available: <https://www.isglobal.org/coronavirus>. [Accessed: 06-Aug-2021].
- [7] "COVID-19 cases, recoveries, and deaths," *Statista.com*. [Online]. Available: <https://www.statista.com/statistics/1087466/covid19-cases-recoveries-deaths-worldwide/>. [Accessed: 06-Aug-2021].
- [8] "Mexico COVID: 2,922,663 Cases and 243,165 Deaths - Worldometer," *Worldometers.info*. [Online]. Available: <https://www.worldometers.info/coronavirus/country/mexico>. [Accessed: 06-Aug-2021].
- [9] "Coronavirus Recovery," *Webmd.com*. [Online]. Available: <https://www.webmd.com/lung/covid-recovery-overview>. [Accessed: 06-Aug-2021].
- [10] Y. Fang, Y. Nie, and M. Penny, "Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis," *J. Med. Virol.*, vol. 92, no. 6, pp. 645–659, 2020.
- [11] J. Howard *et al.*, "Face masks against COVID-19: An evidence review," *Preprints*, 2020.
- [12] S. A. Rella, Y. A. Kulikova, E. T. Dermitzakis, and F. A. Kondrashov, "Rates of SARS-CoV-2 transmission and vaccination impact the fate of vaccine-resistant strains," *Sci. Rep.*, vol. 11, no. 1, p. 15729, 2021.