UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics (LM-17)

Final Dissertation

Modelling COVID-19 spreading in a network

Candidate

Riccardo Milocco

Thesis supervisor

Prof./Dr. Marco Baiesi

Academic Year 2020/2021

Contents

3

Abstract

The usual simplified description of epidemic dynamics predicts an exponential growth. This is due to the mean field character of the dynamical equations. However, a recent paper (Thurner S, Klimek P and Hanel R 2020 Proc. Nat. Acad. Sci. 117, 22684) [13] showed that in a network with fixed connectivity, the nodes become infected at a rate that increases linearly rather than exponentially. Experimental data for COVID-19 seem to validate this approach. In this thesis we plan to study this model by tuning its parameters. In particular, we monitor the effect induced by a significant presence of hubs in the network.

Motivations

Last Update: 06-07-2021

2.1 Motivations

"The COVID-19 pandemics has led to a dramatic loss of human life worldwide and presents an unprecedented challenge to public health, food systems and the world of work" [3]. In this scenario, monitoring and forecasting the diffusion of the virus is a essential tool for policymakers to handle the health-care resources and restrictions among individuals. To this purpose, many academics, e.g. UniPD against COVID[1], have been trying to tackle the COVID-19 spreading as well as its side-effects, which have to be considered on the "path to normality". This drastic shift with respect to the "path to herd immunity" comes from a recent study [5] which shed light on the fact that "herd immunity", due to delay on vaccinations, could probably not be reached for the end of 2021. In other words, the recovery of the pre-COVID situation could be slow even in the presence of effective vaccines.

So, an interesting path to cover is to describe the evolution of the COVID-19 pandemic, e.g. with the SIR model [11], on different network topologies in order to simulate the different containment policies, i.e. "lockdowns".

---

[1]https://web.unipd.it/covid19/en/

Introduction

3.1 Gentle Introduction

Infectious diseases are responsible for a significant health and economic burden on society. New diseases appear frequently and old diseases persist. Therefore, many mathematical models have been used recently to guide policy makers to mitigate the impact of new infectious diseases (e.g. Polio, H1N1 influenza and Ebola) or established ones (such as HIV, cholera and seasonal influenza). On the other hand, with the phrase *All models are wrong; some models are useful* [2], George Box wanted to highlight that no model, no matter how complicated[1], is perfect. Thus, a way to evaluate a "successefull" model has to be recovered.

Inside the "ecosystems" of models, there are the simple models which allow to analytically explain how the primary mechanisms influence important characteristics (e.g. epidemic threshold, epidemic size[2], how long it will last). To be more accurate, it is also possible to develop models that incorporate much more details about both the disease and individual-level interactions [7].
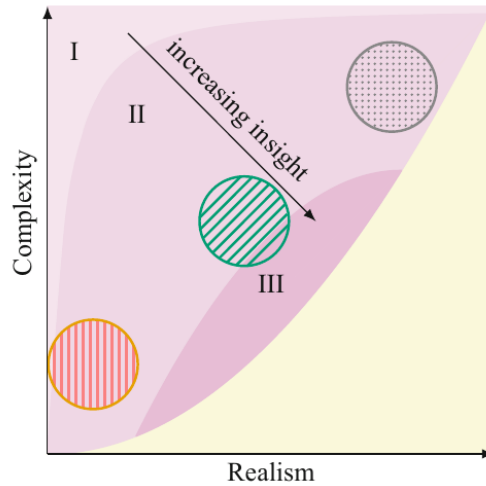

Figure 3.1: A caricature of the relation between model realism, complexity and the insight the model provides [7].

As suggested by the Figure 3.1[3], models provide maximum insight where the right balance of realism (circle filled with dots, e.g. Graph-Neural-Network to predict R_t [4]) and simplicity (circle with vertical lines e.g. a mean-field approximation of the SIR model [8]) is met[4] (central circle with oblique lines).

In this middle region, there could be find several works that use network analysis as the underlying

---

[1]Instead, the complexity of a model could challenge its possibility to be predictive in many scenarios. A related topic is the "overfitted model" in statistical learning or "Ocham's razor" in Philosophy

[2]how large it will be

[3]*"In medio stat virtus"*

[4]The equilibrium point may depend on the specific posed question

structure for a phenomenon([13] [10] [11]); while other, more involved, that incorporate human mobility data in several hybrid models [12, 14] in order to predict its diffusion. For the sake of simplicity, the present work of thesis is going to deepen the paper of [13]. In particular, Thurner et al. [13], by focusing on the COVID-19 spreading in Austria and in the USA, are able to find a scenario, i.e. a combination of the spreading parameters[5], such that the outbreak is (quasi-)linear for a value of the average number of contact D. This manifestation of the SIR model is, at least, qualitatevely showed by the curves reported by the JHU University. So, inspired by [13], the aim of this project is to analyze how the SIR model behave by changing its "innern" epidemic parameters but also the topology of the underlying social network, i.e. the susceptible people[6].

3.2 Technical Introduction

At 8 May 2020, none of the affected COVID-19 states have reached "herd immunity"[7], but still they reached the "epidemic peak" due to the containment restrictions on social contacts, e.g. Austria New Daily Infected. The most striking observation is that the number of total cases, after early-stages when Non-Pharmaceutical Interventions (NPIs), i.e. lockdown, were not at work, exhibits a linear growth for an extended time interval in contrast with the "S-shaped" logistic curve predicted by the standard compartmental models. The extension of the linear regime depends on the onset of the measures; while for early stages, as it was the case for many countries [13] (8 May 2020), an exponential growth dominates the spreading of the disease. By taking care of the modification of the underlying social network structure, in this present work of thesis, we want to (qualitatively) recover the spreading trends for the states that have or have not applied the NPIs measures.
Specifically, an infection may occur for two reasons:

1. interaction between an infected and a susceptible person;

2. contact is "intense" (e.g. long, close,...) enough to lead to a disease transmission.

So, the rationale behind the social distancing is that it takes to a reduction of both of these factors. On the other hand, the analitically solvable SIR model assumes that, defined N, D as the total number of individuals in the population and the average number of contacts respectively, there is the same probability that an individual encounters an infected person ("well-mixed population") and all the nodes have the same number of neighbours ($N - 1$ or D). Therefore, as claimed in [10], there is the need of studying how the underlying network affects the spreading of a disease, but still no focus is put on the spreading below the epidemic threshold [13] as it is the case if nodes were separated thurough NPIs. In this thesis, the main goal is to grasp the relevant features of a complex social network in presence of NPIs and compare them with respect to the mean-field approach which annihilate the network structure in favor of the analytical treatment of the SIR model.
In detail, we let the SIR Model evolve in different network topologies by changing the epidemic parameters. On the other hand, it has been considered both the standard models of the graph-theory ("Erdoes-Rényi","Scale-Free", "Caveman Model") but also a more involved topology, i.e. a Poissonian Small-World Network. As a benchmark of a "well-mixed" population, it has been used an "annealed" mean-field (see) which is a natural realisation of a mean-field network with a fixed average number of contacts per node.

---

[5]the infection rate β, recover rate μ, the long-range attachment probability p

[6]In graph theory, a graph G is a pair of node/vertices (V) and edges(E) which connnect the nodes, i.e. $G = (V, E)$. Thus, in the present context the nodes are the (susceptible) people. Therefore, "node" and "people" words may be used interchangebly according to the context.

[7]The total infected cases at the first peak were 0.16% of the total population, remarkably low with respect to the SARS-COV-2 "herd-immunity" level that are 0.8% of the entire population.

The Models

To this end, we proceeded with the following steps. At first, we pruned a fully connected graph of $N = 1000$ nodes, by keeping fixed N while having D, i.e. the average of nearest neighbors. In parallel, since $R_0 = \frac{\beta \cdot D}{\mu}$, we doubled the β [1] parameter to keep the power of the epidemic R_0 fixed (**??**).

Since the aim is to understand the difference with respect to the "well-mixed" population model, we put in the same plot the number of total infected cases with the epidemics that spreads either on a specific network (in green) or in a "mean-field" fashion (in orange)[2]. More precisely, the "mean-field" approximation is a "D-neighbors-mean-field" where every node has D nearest neighbors picked at random on the whole set of nodes. In this way, the hypothesis of the "well-mixed" population is recovered, i.e. everyone has the same probability to get in touch with an infected, and the spreading could be modeled with the exact same parameters (β and μ) as the network epidemics.

Then, we introduced the "rewiring probability" p which enables the distant nodes to connect faster, thus, obtaining a Erdoes-Rényi (E-R) graph with non-zero p. In this case, we could recover the "(pseudo) linearity" of the SIR infection, as it is shown in **??**, using two similar methods of choosing the spreading parameters β and μ:

- pick a series of $\beta \in [0,1]$ and $\mu \in [0,1]$ coefficients at random;

- choosing $\mu \in [0,1]$ and fixing $p = 0.1$. Then, as in the previous point, while halving D, the β is doubled.

For conciseness, in **??**, it is reported only the epidemics obtained by the "pruning procedure".
To further test the recovered (pseudo) linearity with respect to the mean-field benchmark, we used a Poissonian "small-world" network which enabled us to taking care of degree heterogeneity (i.e. different social contacts), family-clusters and their overlaps, small-world feature, given the fact that "leisure" activities may connect nodes that are diametrically opposite. Moreover, by changing the number of the average degree D we may pass from light ("exponential spreading" for $D \simeq 5$) to severe ("linear spreading" for $D \simeq 2.5$) NPIs measures, i.e. a severe lock-down. The Poissonian "small-world" network could be realized by imposing, at first, a Poissonian degree sequence on all nodes ("Configuration Model"); then, re-link the nodes to their closest neighboring nodes, i.e. a nearest neighbors rewiring (NNR) procedure is applied. Thus, obtaining a network with poissonian degree distribution and the small-world property. The (pseudo) linearity obtain with this graph is reported in the **??**.

Furthermore, the work focused on analyzing the network spreading on a fully-connected cavemen graph, where the families cores could be taken into consideration; and also enabling the possibility of a "long-range" connection via the rewiring probability $p = 0.1$ (cfr. **??**). Finally for what concerns the different network models, it has been studied the case of "(pseudo) linear" epidemics on a Barabasi-Albert (B-A) Model, i.e. with the presence of hubs (cfr. **??**).

---

[1] $\beta \in [N, 2]$
[2] Moreover, the daily infected of the network (in blue) and of the mean-field (in magenta)

Moreover, the critical average degree D_c [3] , is going to be recovered for the quoted network topologies. A final up-to-date point, would be to use a Graph Neural Network architecture, based on the previously described scenarios and networks, in order the make epidemics forecasting. For this purpose, a relevant reference, is [4]

4.1 Goals Outline

1. Assuming that a typical social network has an average number of contact (degree) $D \simeq 16.5$([9][4]); while lockdown measures reduce it to the household size ($D \simeq 3.5$ [9]), our goal is to reproduce the dynamics of an epidemic spreading on a sparser network due to the Non-Pharmaceutical Interventions (NPIs) generally known as lockdown[5]. A direct comparison of the obtained curves with the states of Austria and the USA [13] is satisfactory even at a plain-eye inspection. However, the project aims at the study of how the topology of a network could characterize a paradigmatic model as the SIR one. So, no statistical fit has been performed, since a relevant work on the available data has to be performed, e.g. removing out-of-range values, impossible data combinations, and missing values on *GitHub John Hopkins University*[6]. In addition, for the fitting procedure, it would be possible to extend the SIR model in accordance to the other reported compartments, e.g. death (SIRD), exposed (SEIR),... Anyway, these states express the (quasi-)linearity in the cumulative number of cases similarly to the simulated ones. Moreover, it has to be pointed out that the previous quoted states are representative of two opposite political policies, undertaken by many others states, in facing the COVID-19 pandemic: a immediate "severe lockdown" (Austria) and "light restriction" on contacts (USA).

2. As claimed in [13], we want to find that for any fixed transmission rate, recovery rate and "long-range" probability (β, μ, p) there exists a critical number of social contacts (D_c) which describes a first-order phase transition of the disease growth obtained with the SIR model:

   - for $D < D_c$, linear growth, i.e. low infection prevalence, occurs;

   - for $D > D_c$, "classical" exponential growth as predicted by the SIR model.

   The rationale behind is that, since $R_0 := \beta\,D/\mu$, by augmenting the average degree of the udnerlying network also the R_0 is enanched, resulting in a more powerful spreading.

3. On the way to obtain D_c, the simulated plots show a strongly dependence on p, i.e. the probability of a node to have a distant neighbor ("long-range" attachment), which is as crucial as β, μ and D in the characterisation of a epidemic spreading. Therefore, a new definition of $R_0 := R_0$ is proposed, which integrate the possibility that same effective parameter but with different long-range attachment (e.g. one local and the other random attached) could drive to different scenarios. Finally, by setting our parameters to the empirical estimates of β and μ, we are going to estimate D_c.

---

[3]D_c as reported in [13] describes a "first-order-phase transition" between exponential and linear growth of an epidemic

[4]A "contact" is defined as per touching or conversationing at a certain distance. For a spreading of a SARS disease, this open the discussed topic on the validity of the "2 meters" distance policy [6]

[5]For brevity, the word "lockdown" is an alias for all the "lockdown-like" measures undertaken to prevent the diffusion of COVID-19 without requiring Pharmaceutical drug treatment, e.g. confinment (the direct translation of "lockdown") but also contact tracing, higher hygiene, face masks,

[6]https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

4.2   Relevant Figures

FIGURES ARE AFTER "\textbf{\end{document}}"

Model

The main assumptions of the model are:

- fixed number of total individuals N connected by a social link;

- "undirected" adiaciency matrix A, where $A_{i,j} = 1$ if there's a link between the ith- and jth-node or $A_{i,j} = 0$ otherwise;

- transmission per day probability r and d days for an infected individual to be recovered, i.e. immune or naturally dead;

- average degree D and shortcut (or rewiring) probability ϵ

As one may expect, the r and d parameters are in a one-to-one map with to the SIR ones, respectively, for the recovery and transmission rates:

$$\gamma = 1/d, \quad \beta = rD/N. \tag{5.1}$$

At each time step (day), the infection curve of positive cases, $P(t)$, is the cumulative sum of the number of new cases $C(t)$. So, if $D > D_c$ the "well-mixed" (or homogeneous/mean-field) approximation may be applied, with the result that $C(t)$ is associated to $R(t)$ up to a time-shift of d, since all the infected are predicted to become recovered after d days.

Results

6.1 Infection Dynamics

To gain some intuition about the importance of modelling the underlying network, we've made the disease spread for two different values of D but fixing $N_{total} = 1000$, $D = 8$, $\epsilon = 0.1$, $d = 6$ days, $r = 0.1$ and 10 initially infected nodes as seeds ($N_{seeds} = 10$).
The results are the following:

- In the large D and ϵ limit, the mean-field condition is fulfilled and the model resembles the compartmental SIR model. By mapping the parameters as done in 5.1, we may compare $P(t)$ from the network model with the compartmental $R(t)$ by shifting it of $-d$ days. As a relevant reference for the comparison to the "small D" evolution, we've plotted the histogram of the daily cases which shows the typical peak in the early-exponential stages and, then, a decrease towards herd immunity, which for [13] has been reached at the 98% of the overall population.

- By reducing the average degree to $D = 3$, we've find a changing in the spreading behaviour, since $P(t)$ increases almost linearly for a remarkable timespan. At the end of the infection, nearly 1/5 of the entire population has been infected which is noticeably smaller than the SIR predicted herd immunity nearly at 80%.

With this rough perception of the main results by considering the underlying network structure, we've further analyzed the parameter dependence.

6.2 Parameter Dependence and Phase Transition

The aim of this section is to obtain the value of the "critical average degree" D_c, such that, for $D < D_c$, the infection curve is (roughly) linear. In order to take care of the two growth trends, we've defined an order parameter

$$\mathbb{O} := SD(C(t)) \tag{6.1}$$

as the standard deviation of the daily new cases, having removed the days without new infected. In this way, $\mathbb{O}! = 0$ signals the presence of a nonlinear increase of the daily new cases and, therefore, of its cumulative sum $P(t)$. Other way around, $\mathbb{O} = 0$ corresponds to a linear growth of $P(t)$.

Appendix

7.1 Notes on Network book of A.Barabasi

7.1.1 Compartimental Models 10.2

SI model

At (10.1) β is the probability of 1 spread only. So, $\frac{\beta \langle k \rangle S(t)}{N}$ expresses the probability of 1 spread over $\langle k \rangle S(t)/N$ neighbors.

$C := \ln \frac{i_0}{1 - i_0}$

SIS model

In (10.7), define $v_{SIS} := 1 - \frac{\mu}{\beta * \langle k \rangle} = 1 - R_0^{-1}$ the "characteristic velocity". If $v > 0$ spread; otherwise it dies out. So, $\tau^{SIS} = \frac{1}{\beta \langle k \rangle v_{SIS}} = \frac{\tau^{SI}}{v_{SIS}}$ In, SI model $v_{SIS} = 1$ [1]

The endemic fraction of infected is given by $\frac{\beta - \mu}{\beta}$ slides 16 of the LDE course.

7.1.2 Epidemics on Nets 10.3

SI Model

10.16 :

$$\tau_n^{SI} = \frac{\tau_c^{SI}}{\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle^2}}$$

10.17 : is obtained by imposing the initial condition $i(t = 0) = i_0$. Anyway, $i_k = i_0 + i_0 f(t)$

SIS

$$\tau_N^{SIS} = \frac{\tilde{k}^2}{\beta \langle k \rangle - \mu \tilde{k}^2}$$

where $\tilde{k}^2 := \frac{\langle k \rangle^2}{\langle k^2 \rangle}$ and ???? if the network is homogeneous $\langle k \rangle^2 = \langle k^2 \rangle$. Check better what's the behaviour of $\langle k^2 \rangle$ wrt the considered network.

7.1.3 10.B

If a nets lacks degree of correlation, i.e. $e_{ij} = q_i q_j$, the probability $p_{kk'}$ is independent on k. Thus,

$$p_{kk'} = \frac{k' p'_k}{\langle k \rangle}.$$

Typo under (10.50) multiply by $k p_k$ not $k - 1) p_k$

Tight Binding model

8.1 Second Quantization formulatio of the SIR Model

Bibliography

[1] Albert-László Barabási and Márton Pósfai. *Network science.* Cambridge University Press, 2016.

[2] George E. P Box. *Statistics for experimenters : design, innovation and discovery.* Wiley-Interscience, Hoboken, N.J., 2nd ed. edition, 2005.

[3] Kimberly Chriscaden. Impactt of covid-19 on people's livelihoods, their health and our food systems. https://www.who.int/news/item/13-10-2020-impact-of-COVID-19-on-people%27s-livelihoods-their-health-and-our-food-systems, 2021.07.01.

[4] Mohammad Reza Davahli, Krzysztof Fiok, Waldemar Karwowski, Awad M. Aljuaid, and Redha Taiar. Predicting the dynamics of the covid-19 pandemic in the united states using graph theory-based neural networks. *International Journal of Environmental Research and Public Health*, 18(7), 2021.

[5] Youyang Gu. Path to normality website, 2021.01.07. https://COVID19-projections.com/path-to-herd-immunity/.

[6] Nicholas R Jones, Zeshan U Qureshi, Robert J Temple, Jessica P J Larwood, Trisha Greenhalgh, and Lydia Bourouiba. Two metres or one: what is the evidence for physical distancing in covid-19? *BMJ*, 370, 2020.

[7] Istvan Z Kiss, Joel C Miller, and Peter Simon. *Mathematics of epidemics on networks: from exact to approximate models.* Springer, 2017.

[8] Istvan Z Kiss, Joel C Miller, and Peter Simon. *Mathematics of epidemics on networks: from exact to approximate models.* Springer, 2017.

[9] Carol Y. Liu, Juliette Berlin, Moses C. Kiti, Emanuele Del Fava, André Grow, Emilio Zagheni, Alessia Melegaro, Samuel M. Jenness, Saad Omer, Benjamin Lopman, and Kristin Nelson. Rapid review of social contact patterns during the covid-19 pandemic. *medRxiv*, 2021.

[10] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, 86(14):3200–3203, April 2001.

[11] Clara Pizzuti, Annalisa Socievole, Bastian Prasse, and Piet Van Mieghem. Network-based prediction of covid-19 epidemic spreading in italy. *arXiv*, 2020.

[12] Johannes Stübinger and Lucas Schneider. Epidemiology of coronavirus covid-19: Forecasting the future incidence in different countries. *Healthcare*, 8(2), 2020.

[13] Stefan Thurner, Peter Klimek, and Rudolf Hanel. A network-based explanation of why most covid-19 infection curves are linear. *Proceedings of the National Academy of Sciences*, 117(37):22684–22689, 2020.

[14] Abdelhafid Zeroual, Fouzi Harrou, Abdelkader Dairi, and Ying Sun. Deep learning methods for forecasting covid-19 time-series data: A comparative study. *Chaos, Solitons & Fractals*, 140:110121, 2020.