Public Health Interventions in the Face of Pandemics: Network Structure, Social Distancing, and Heterogeneity

Mohammad Ghaderi

July 2020

Barcelona GSE Working Paper Series

Working Paper nº 1193

Public health interventions in the face of pandemics: network structure, social distancing, and heterogeneity

Mohammad Ghaderi * † ‡
mohammad.ghaderi@upf.edu

June 20, 2020

Abstract

Complexity, resulting from interactions among many components, is a characterizing property of healthcare systems and related decisions. Such complexity scales up quickly in the face of pandemics, where multiple sources of uncertainty are involved and various contextual factors interacting with policy parameters yield outcome distribution. This paper presents a unified framework to assist and inform policy decisions in confronting pandemics. The general framework consists of a model of contagion that makes the policy-relevant variables explicit and exogenous, establishes links between them and the main features of the environment in which the policy is going to be implemented, and treats various sources of uncertainty at different layers of the system. At the macro level, special attention is devoted to the network structure, for which we provide a simple characterization based on two constructive factors. Our results show that by conditioning on these two factors, a large proportion of the stochasticity resulted from the inherent randomness in the network can be captured. Components of the model are synthesized in a broader agent-based model that enables accounting for heterogeneous individual-level attributes that collectively yield the macro-level outcomes. Using several stylized examples and a comprehensive controlled experiment, insights on the overall tendency of the complex system in terms of multidimensional outputs are derived across a range of scenarios and under various types of policy conditions.

Key words: Public health interventions, social contagion, random networks, social distancing, simulation

---

JEL Codes: C6, C54, C32, I1

* Department of Economics and Business, Pompeu Fabra University, Barcelona, Spain

† Barcelona Graduate School of Economics

‡ UPF Barcelona School of Management

1. Introduction

The novel coronavirus disease (Covid-19) outbreak has been declared as a public health emergency of international concern by the World Health Organization (WHO). The pandemic has challenged the governments, public, and private sectors by its multifaceted impacts. On one hand, the rapid spread of the disease has oppressed the public health system, which in response to, different intervention strategies across the countries have been adopted, for the most part, based on containment and non-pharmaceutical measures such as household quarantine, workplace closure, and mobility restrictions. On the other hand, the pandemic and adopted confrontation strategies have caused serious economical damages and raised concerns regarding their long-lasting impacts, for instance, declines in employment, disrupting global supply chains, negative supply shocks, and demand shortage (Fornaro and Wolf 2020, McKibbin and Fernando 2020, Fernandes 2020), with asymmetric influences across sectors (Guerrieri et al. 2020). Although the Covid-19 pandemic has received substantial global attention due to ample media coverage, it is only one of the many disease outbreaks in recent decades, for instance, Ebola in 2014 or influenza in 2009, among others. While wishful, but unfortunately it is unrealistic to expect similar incidents to be rare in the coming years, especially in view of accelerated growth of worldwide transportation which facilitates the rapid circulation of goods and people, but also infectious diseases. Hence, developing effective intervention strategies to mitigate the impacts of a pandemic is a top global priority now.

Deciding upon elements of an effective response in confronting pandemics, however, is a highly complex task for at least three main reasons. First, a high level of uncertainty involved, and a lack of sufficient data available, impede reliable predictions on the effectiveness of alternative strategies. Moreover, the multidimensional nature of the consequences of intervention methods, from the healthcare, economic, and social perspectives, escalates the complexity of the task since it requires to make a compromise between various types of conflicting criteria.

Second, assessing the efficacy of alternative policy choices in relation to one another is a cumbersome task. For example, concerning the popular intervention methods based on self-isolation,

despite that they contribute in detaining the virus spread, nevertheless, they also cause substantial and inevitable economic damages. In this case, for policymaking purposes, it is crucial to have an idea on their extent of effectiveness, especially compared to alternative intervention methods such as healthcare capacity expansions, health service quality improvements, testing capacity enlargements, extensive and regular public area decontamination, or any combination of those options subject to the budget constraints. Specifically, it is essential to develop insights on how different intervention methods compare and how they interact if were to combine in a unified response plan.

Third, the efficacy of an intervention method might be largely determined by the environment in which it is embedded. The mechanism under which the policy parameters affect the outcome distribution can, to a considerable extent, be shaped by the dynamic characteristics of the embedding environment. For instance, the patterns of contacts among individuals combined with heterogeneous social distancing behavior would control the frequency of contacts between pairs of individuals, hence determining the infection transmission rate. Similarly, population overall health status and healthcare capacity combined with the total number of hospitalized individuals, which in turn is determined by the infection transmission rate and policy parameters such as testing capacity, control the load on healthcare system which affects the health service quality and subsequently influences the mortality rate. From this perspective, contextual factors such as population overall health status, patterns of contacts among individuals, current healthcare system characteristics, and even income distribution in the population, interact with the policy parameters and determine the elements of an effective response plan.[1] In the same vein, the level of heterogeneity among the population is among other influential factors interacting with the policy parameters. For instance, some views speculate that remaining at home as a policy is only accessible to those with income above a certain level, hence individuals' ability to self-isolate differs across the population (Chiou and Tucker 2020), making low-income citizens more prone to the crisis and invoking heterogeneous

[1]In Europe, for instance, different countries have adopted different strategies in response to the health crisis caused by Covid-19, with some relying largely on extensive testing strategy, e.g. Germany, some others focusing on strict household isolation measures, e.g Spain, and some choosing other different paths, e.g. Sweden.

response to the adopted measures. An enhanced understanding of the contingent relationships between features of the intervention strategies and characteristics of the population, as well as contextual factors that shape the environment in which the strategy would be implemented, is essential for policymaking. Likewise, the sensitivity of different strategies to the changes in such factors, thus robustness of the adopted intervention method, has clear implications for policymaking. For instance, it is important to understand how the effectiveness of a strategy based on confinement would be undermined if a small fraction of the population breach.

This paper contributes to the above challenges by presenting a general model that integrates policy-relevant variables in a unified framework with various components describing idiosyncratic features of the environment and population. The model is primarily intended to inform policy decisions for developing effective solutions that are customized to the characteristics of the population and features of the context. To this objective, a probabilistic contagion model is developed where dissemination of infection throughout the population is determined by various types of policy-relevant variables which are projected on a network representing patterns of physical contacts between individuals (also known as network structure) in the population. The role of the network is captured through two primary exogenous factors characterizing its structure, namely centrality in its topology and sparsity in its connections. Components of the model are synthesized in a broader agent-based model that enables accounting for heterogeneous individual-level attributes that collectively determine the macro-level outcomes.

It is important, however, to note that the objective of this framework is not to make accurate numerical predictions regarding the specific outcomes by using historical data but to produce insights regarding the overall tendency of the complex system under various conditions. To this aim, stylized examples are presented and an extensive controlled experiment is conducted. From this perspective, the model is suggestive and directional, rather than conclusive. With this clarification, caution is advised when projecting directly the derived quantities.

2. Related Literature

Perhaps the most well-known compartmental model of infectious disease is the class of SIR models (Kermack and McKendrick 1927) that examines the flow of a population among the three phases– susceptible, infected, and recovered.[2] This class of models are deterministic nonlinear based on differential equations and can be used to describe the spread of a virus, to compute the total number of people infected, and to estimate various epidemiological parameters such basic reproduction number (Hethcote 2000). Recent extensions of the model include attempts to enable accounting for intervention policies and make inferences on their consequences in such context. For instance, Atkeson (2020) examines the effectiveness of quarantine measures, Berger et al. (2020) considers the role of testing and diagnosis, Obiols-Homs (2020) accounts for social distancing by placing an upper limit on the number of interactions between individuals. In order to make predictions regarding the total number of infected or recovered individuals, these models build on assumptions that impose specific mechanisms of transmission between the phases.

The model presented in this paper is developed based on a stochastic agent-based model so as to account for major sources of uncertainty and capture heterogeneity in individuals attributes. In the context of social epidemiology and public health, agent-based modeling is used to promote population-level inference from micro-level rules in simulated populations over time and space (Bonabeau 2002, Epstein 2006, Rahmandad and Sterman 2008, El-Sayed et al. 2012, Tracy et al. 2018). Contrary to the contagion models based on differential equations that assume homogeneity and perfect mixing within compartments (Riley 2007, Mahajan 2010), agent-based models allow including heterogeneity in individuals attributes (Rahmandad and Sterman 2008), such as social distancing behavior, health status, and network structure of their interactions. Such flexibility, moreover, enables us to treat the transmission probabilities between phases and epidemic duration as endogenous variables that are determined by the policy in place combined with additional

[2]Several variants of the basic SIR model have been developed, where one of the most well-known among which is the SEIR model that considers an *exposed* state between the susceptible and infected states.

contextual factors.[3] The questions posed in this paper, therefore, are closer to those of Piguillem et al. (2020) and Obiols-Homs (2020), for instance, whether and to what extent sever social distancing or testing should be considered in the design of an optimal response to infectious disease. The present model considers a broad range of policy-relevant parameters in a unified framework, hence enables exploring a wide range of policies such as decisions related to healthcare capacity and health service improvements, investment in testing and diagnosis, or even public areas decontamination. Furthermore, by accounting for heterogeneity in individuals attributes, we aim to produce more realistic insights. For instance, individuals exhibit a heterogeneous response to self-isolation measures since the method is not equally accessible to everyone depending on several factors including income (Chiou and Tucker 2020).[4] Moreover, accounting for heterogeneity is crucial in policy analysis for developing robust strategies, for instance, to understand how the effectiveness of costly measures based on self-isolation diminishes when a small fraction of population breach as opposed to when the entire population fully collaborate. Similar to other agent-based models, however, the outputs of our model should be interpreted purely at the qualitative level (Bonabeau 2002), hence applicable for deriving insights and deciding on general directions of the policies.

Our model is related to the literature on network theory and random graphs (Gilbert 1959, Easley et al. 2010, Abbe 2017), which are leveraged to account for structured interaction patterns among individuals. The prevailing importance of network structure and its impacts have been demonstrated in various contexts including social contagion, epidemiology, and diffusion process (Davis 1991, Abrahamson and Rosenkopf 1997, Rahmandad and Sterman 2008, El-Sayed et al.

[3]Agent-based models, however, are computationally and cognitively demanding, both at the level of model design since it should be built at the right level of description, with just the right amount of detail (Bonabeau 2002), and the level of results interpretation and linking the behavior of the model to its structure (Rahmandad and Sterman 2008). Moreover, the computational cost of agent-based models limits the size of the population in analysis–e.g. few hundreds compared to some millions in the case of differential equation based models (Rahmandad and Sterman 2008).

[4]In somewhat related work, Van den Bulte and Stremersch (2004) discuss the role of income heterogeneity and cultural dimensions in the diffusion of new products and social contagion.

2012, Muller and Peres 2019, Manshadi et al. 2020). In the spread of disease, individuals contact network plays an important role since it directly influences patterns of interactions among individuals and hence transmission rate of infections. For instance, in contact networks consisting of several linked communities, the pace and scale of the spread might be different depending on the community-size distribution and population density within each community. Moreover, the topology of the network can have a large impact too (Davis 1991, Muller and Peres 2019, Manshadi et al. 2020). For instance, the presence of a large central component in the network, e.g. social hubs (Goldenberg et al. 2009, Muller and Peres 2019), versus a decentralized network consisting of several fairly equal-in-size communities might exhibit different patterns of spread. This has clear implications for policy analysis, especially for developing context-sensitive solutions.

In the present paper, a random network model is used to represent human contact networks. While many types of real networks have been constructed empirically, for instance in the diffusion of innovation or opinion through social contagion (Iyengar et al. 2011), with diseases, on the other hand, the process is sufficiently complex and unobservable at the person-to-person level that it is most useful to model it as random (Easley et al. 2010, pages 568-569). The random network model considered in this paper first generates a random community-size distribution based on the literature in integer partitioning problem from number theory (Fristedt 1993) and by introducing a new algorithm. It then constructs the network within each community according to the Erdős-Rényi random graph model (Erdős and Rényi 1960, Gilbert 1959) and links the communities based on a stochastic block model (Abbe 2017).

The intra-community connections resulted from the Erdős-Rńyi model leads to a Binomial degree distribution within the community. Recent literature in network science has approached this property with skepticism since it has been observed that several types of empirical networks follow a different pattern of the degree distribution, namely power-law distributions in scale-free networks (Barabási and Albert 1999, Barabási et al. 2000, Barabási and Bonabeau 2003). In such networks, it is asserted that the fraction of nodes with degree k follows a power law $k^{-\alpha}$, a property that has broad implications for the structure of the network. While such patterns ostensibly

appear in many types of networks such as world wide web, online social networks, citation networks, networks of organizations, or biological networks (Albert et al. 1999, Barabási et al. 2000, Newman 2003b), nevertheless, many others refute their assumptions and challenge their credibility on statistical and theoretical grounds (Broido and Clauset 2019, Willinger et al. 2009). On the other hand, assumptions of Erdős-Rńeyi combined with stochastic block models appear plausible for representing networks of human contacts, especially that in our model heterogeneity in the degree distribution is directly invoked by the dispersion in community-size distribution– see, for instance, (Newman 2003a, Karrer and Newman 2011). For a review of the models of network structure, see Burt (1980); for a detailed survey of statistical network models, see Goldenberg et al. (2010); for a general discussion on models of random networks, see Newman (2018) and sections 4.1-4.2 in Jackson (2010).

3. Model Construction

Let $\mathcal{I}$ be a finite set of individuals and $\mathcal{T} = \{1, 2, \cdots, |\mathcal{T}|\}$ be a finite set of time periods with an unknown size. At any time t, an individual i's state, denoted by the random variable S_{it}, can take one of the five possible values–susceptible, infected, detected, recovered, and perished. By default, the neutral state of an individual is susceptible, i.e healthy or uninfected, analogous to the naïve or uninformed individual in the context of new product diffusion in marketing, and is indicated by $S_{it} = 0$. A susceptible person can become infected if he interacts with an infectious agent. The state of an infected (informed) individual is indicated by $S_{it} = 1$. Every time a susceptible individual i meets an infectious agent j, the infection would transmit with a probability δ. Hence, the infection transmission in direct contact occurs according to a Bernoulli experiment with a parameter δ. If the two individuals meet multiple times within a specific period, each experiment takes place independently of the others. Formally, suppose that individual i, $S_{it} = 0$, meets individual j, $S_{jt} = 1$, for Y_{ijt} times in time period t. Assuming independence, and conditional on Y_{ijt}, the binary random variable Z_{ijt}, with $Z_{ijt} = 1$ indicating transmission of infection from individual j to individual i at time period t, follows a Bernoulli distribution with parameter $1 - (1 - \delta)^{Y_{ijt}}$, i.e.

$$Z_{ijt}|Y_{ijt} \sim Bern\big(1 - (1 - \delta)^{Y_{ijt}}\big). \tag{1}$$

In turn, number of direct contacts between two individuals i and j at a time period t, i.e. Y_{ijt}, is a random variable itself and its realized value at each trial depends on the distance between the two individuals. The distance, on the other hand, depends on the network structure and social distancing behavior of the two individuals. Therefore, notion of distance here serves as a latent factor that links social network structure and social distancing behavior to the frequency of direct contacts between two individuals in a unit of time. Construction process of these two components will be discussed in the next subsection. At this stage, suppose that the vector $\mathbf{Y}_{it}$, with its jth component representing Y_{ijt}, follows an arbitrary discrete joint probability mass functions G_{it}. The infection risk for i, therefore, is obtained as following:

$$Pr\big(S_{it+1}=1|S_{it}=0\big)=1-\sum_{\mathbf{y}_{it}}\prod_{j\in\mathcal{I}\setminus i}\Big(\mathbb{I}[S_{jt}=0]+(1-\delta)^{y_{ijt}}\mathbb{I}[S_{jt}=1]\Big)G(\mathbf{y}_{it}) \tag{2}$$

where $\mathbb{I}[.]$ is an indicator function that takes value 1 if its input argument hold true, and 0 otherwise.[5]

Proof: By conditioning on Y,

$$Pr\big(S_{it+1}=0|S_{it}=0,Y_{ijt}=y_{ijt}\big)=\textstyle\prod_{j\in\mathcal{I}\setminus i}\Big(1-[1-(1-\delta)^{y_{ijt}}]\mathbb{I}[S_{jt}=1]\Big)$$
$$\Rightarrow Pr\big(S_{it+1}=1|S_{it}=0\big)=1-\textstyle\sum_{\mathbf{y}_{it}}\prod_{j\in\mathcal{I}\setminus i}\Big(1-[1-(1-\delta)^{y_{ijt}}]\mathbb{I}[S_{jt}=1]\Big)G(\mathbf{y}_{it}).$$

By expanding the argument under the product, and considering that $1-\mathbb{I}[S_{jt}=1]=\mathbb{I}[S_{jt}=0]$, equation (2) is derived. QED

The main assumptions in (2) are that, in a given t, the Y_{ijt} Bernoulli trials are independent from each other, and Z_{ijt} random variables are independent across j. In other words, the outcome of Z_{ijt} is not informative to Z_{ikt} for $k\neq j$. The interaction patterns of i in the network, however, are represented by the joint probability mass function G. From this perspective, G is the main source of differentiation among members of the population when updating the individual's status. Such differentiation is based on the network structure and social distancing behavior, where the two

[5]Note that the product in (2) is taken over all the population, because if $y_{ijt}=0$ for some j, then the resulting value would be equal to $\mathbb{I}[S_{jt}=0]+\mathbb{I}[S_{jt}=1]=1$ which is a neutral element of multiplication.

factors collectively determine the likelihood of contact between an individual i and other members of the population.

An infected individual i remains asymptomatic for a random period of T_i, before been detected.[6] A detected individual is separated from the population, i.e. hospitalized or quarantined, for a random period of H_i before turning to either recovered or perished. Recovered individuals remain immune to the infection during the time window of the analysis.

3.1. Social Distancing

The frequency of direct contact between two individuals depends on whether they are directly linked in the social network or not, and on their social distancing behavior. Let us denote by $d \in \mathbb{R}^+$ the distance between two individuals i and j, and suppose that the number of direct contacts between i and j in a single period of time is inversely related to the square of d_{ij}. If the two individuals are not linked in the network, then $d_{ij} \to \infty$. Conversely, $d_{ij} \to 0$ when the two individuals meet an infinite number of times in a period, e.g. living together. Assuming independence of the number of direct contacts in any non-overlapping subperiods within a time interval, by the law of rare events the number of direct contacts between i and j would follow a Poisson distribution, i.e.

$$Y_{ijt}|i \text{ and } j \text{ directly linked } \sim Poisson(\lambda_{ij}) \tag{3}$$

where $\lambda_{ij} \propto 1/d_{ij}^2$.

The distance between two individuals is not only determined by whether they are directly linked in the network or not, but by their respective social distancing behavior. For two individuals linked in the network, the distance is mainly determined by the one with a greater degree of social distancing. Let us denote by $\alpha_i \in (0,1)$ the degree of social distancing for an individual i, where a larger α_i implies a greater degree of social distancing by $i \in \mathcal{I}$. For $i, j \in \mathcal{I}$ that are directly linked in the network, if $\alpha_i \to 0$ and $\alpha_j \to 0$, then $d_{ij} \to 0$. On the other hand, if $max(\alpha_i, \alpha_j) \to 1$, it means that at least one of the two is practicing a strict self-isolation, hence $d_{ij} \to \infty$. Moreover, for

[6] Note that the stochastic incubation period varies across the individuals.

identification purpose, the middle point on the scale of α is presumed to correspond, in expectation, to one direct contact in a unit of time, i.e. $\mathbb{E}(Y_{ijt}) = 1$ when $max(\alpha_i, \alpha_j) = 1/2$. Following this standardization, the scaling factor between the Poisson parameter and distance vanishes, hence $\lambda_{ij} = 1/d_{ij}^2$. The following specification for d satisfies the above properties:

$$d_{ij} = \frac{max(\alpha_i, \alpha_j)}{1 - max(\alpha_i, \alpha_j)}. \tag{4}$$

The above specification relies solely on the largest of the two social distancing parameters when determining the distance between two individuals, thus does not differentiate, for instance, between $(\alpha_i, \alpha_j) = (0.1, 0.9)$ and $(0.8, 0.9)$. Nevertheless, one might expect the distance in the latter case to be slightly smaller than the other. For this reason, in our simulation computations, the max function is replaced by the smooth approximation $max(\alpha_i, \alpha_j) = (\alpha_i e^{n\alpha_i} + \alpha_j e^{n\alpha_j})/(e^{n\alpha_i} + e^{n\alpha_j})$, where n is a smoothing factor and the approximation error becomes smaller as n grows.[7]

Individuals in the network are different in terms of practicing social distancing behavior. Such heterogeneity can have a large impact on the overall outcome of the system. Moreover, the effectiveness of a response to the virus spread might heavily rely on effective collaboration among individuals. In such settings, it is important to examine how the outcomes vary by the level of heterogeneity in individuals' social distancing behavior.[8] To capture this phenomenon, parameter α is modeled as a finite mixture of Beta distributions with support $(0, 1)$. The parameters of the components of this mixture distribution are defined in such a way that each corresponding density function is either strictly increasing or decreasing. This way, by varying the model parameters, the individuals' degree of social distancing can be pushed towards one of the extremes of the support

[7] $n = 10$ in our computational analysis, which provides a very accurate approximation of the max function within the range of parameters values.

[8] A good proxy for such heterogeneity may be the income differences across the population (Chiou and Tucker 2020), which can be accounted for using the macro-level indicators such as Gini index (Van den Bulte and Stremersch 2004). Further empirical investigations are needed in this direction for a more comprehensive characterization of this component.

range without a mass concentration in the vicinity of the middle point, hence directly controlling the heterogeneity in social distancing behavior. Specifically,

$$\alpha_i \sim \pi \text{Beta}(a, 1) + (1 - \pi)\text{Beta}(1, a') \tag{5}$$

in which $0 \leq \pi \leq 1$ is the mixing factor and represents a fraction of the population that exercises a large degree of social distancing, including a within-group variation controlled by $a \in \mathbb{Z}$, $a > 1$. The second component, with a population fraction of $1 - \pi$, represents the sub-population that do not conform to the social distancing behavior, with $a' \in \mathbb{Z}$, $a' > 1$ being the shape parameter of the distribution. Density function of the first (second) component is strictly increasing (decreasing) in a (a'). By varying a and a', compliance with the social distancing behavior within each group can be modeled, whereas π controls the social conformity level to the social distancing behavior. The larger value of π, the greater is collaboration level among the population in practicing social distancing.[9]

3.2. Stochastic Incubation

An important characteristic in the spread of many types of infectious disease, including Covid-19, is the presence of a notable incubation period.[10] In the case of Covid-19, for instance, many people become symptomatic only after several days of being transmitters, and few others might even remain asymptomatic til recovered. While the incubation period differs largely among people, depending on various factors such as their demographics and health status, testing capacity plays an important role in shortening this period. Extensive testing as a containment strategy helps to detect and isolate asymptomatic infectious agents, hence curb the spread of the virus.

[9]The density function for α_i, given π, is $f_{\alpha_i|\pi}(x) = \pi a x^{a-1} + (1-\pi)a'(1-x)^{a'-1}$. By deriving first and second moments of α_i, it can be shown that $\mathbb{E}(\alpha_i) = \pi\mu + (1-\pi)\mu'$, and $Var(\alpha_i) = \pi\sigma^2 + (1-\pi)\sigma'^2 + \pi(1-\pi)(\mu-\mu')^2$, where $\mu = \frac{a}{a+1}$ and $\sigma^2 = \frac{\mu(1-\mu)}{a+2}$ are, respectively, the mean and variance of the first component of the mixture distribution. Similarly $\mu' = \frac{1}{a'+1}$ and $\sigma' = \frac{\mu'(1-\mu')}{a'+2}$ are mean and variance of the second component, respectively.

[10]This is related to the concept of time-delay in diffusion and social contagion, which can introduce a memory element to the contagion process and produce non-Markovian characteristic.

Let us denote by T_i the incubation period of the infectious agent i. An infectious agent is disclosed according to a Bernoulli trial at the end of each time period. Therefore, T_i follows a geometric distribution, i.e.

$$T_i|i \text{ infected } \sim Geometric(\tau) \tag{6}$$

where parameter τ reflects the testing capacity. The larger value of τ, the shorter is likely to be the incubation period T_i. One way to interpret the parameter τ is to imagine that there is a test center that in each period randomly selects some individuals, each with a probability τ, for test.[11] Now by doubling the testing capacity, there will be two identical centers which, for simplicity, operate independently and select individuals with replacement, hence changing the parameter of the geometric distribution to $1-(1-\tau)^2$. In general, raising the testing capacity by a factor of n will change the distribution parameter in (6) from τ to $1-(1-\tau)^n$.

3.3. Hospitalization and Recovery Chance

An infected individual i, if detected, would be hospitalized for a period of H_i. Similar to incubation period, H_i differs across the population, i.e.

$$H_i|i \text{ detected } \sim Geometric(\eta)\cdot \tag{7}$$

At the end of hospitalization period, individual i is dismissed either as recovered or perished, according to a Bernoulli trial with individual-specific parameter.[12] The recovery chance depends on the general health status of the individual, denoted by θ_i, and contribution of the healthcare system and health service quality, denoted by ϕ. The value of θ_i can be interpreted as the recovery chance of the individual without receiving any health service, i.e. self-recovery chance. Suppose that

[11] We assume that there is no error in the test results, i.e. an infected individual will certainly be detected if tested, and a negative case is certainly uninfected.

[12] It is assumed that the recovered cases are no longer susceptible to the virus during the time window of the analysis. This can produce the so-called herd immunity that occurs when the immune proportion exceeded $(1-1/\mathcal{R}_0)$, where $\mathcal{R}_0$ is the reproduction number and represents the average number of people to which a single infectious agent will transmit the virus. An epidemic with a reproduction number below 1 will gradually disappear.

$\theta_i \sim \text{Beta}(\theta_0, \theta_1)$, hence parameters θ_0 and θ_1 reflecting the overall health status of the population. The health service contributes to the recovery chances by adding $\phi \times (1-\theta_i)$ to these probabilities. The parameter ϕ represents the reduction percentage in mortality risk of an infected individual when receiving health service.[13]

The better the health service quality, ϕ, the greater would be the chance of recovery for a hospitalized individual. However, health service quality would deteriorate as the number of hospitalized individuals increases. To account for this endogenous effect, health service quality is discounted with respect to the extent of demand overload:

$$\begin{aligned} &Pr\big(S_{it+T_i} = \text{recovered } | i \text{ detected at } t\big) = Bern\big(\theta'_{it+T_i}\big) \\ &\theta'_{it+T_i} = \theta_i + \phi^{eff}_{t+T_i}(1-\theta_i), \\ &\phi^{eff}_{t+T_i} = \phi \cdot (max\{\tfrac{\sum_{i\in\mathcal{I}} \mathbb{I}[S_{it+T_i} = \text{ hospitalized }]}{C\cdot|\mathcal{I}|}, 1\})^{-1} \end{aligned} \tag{8}$$

where ϕ^{eff}_t is the effective health service quality at time t, and C is the health service capacity.[14] The health service quality, hence, has an adaptive property and is corrected via a discounting factor that captures the demand overload on the healthcare system at any time. Therefore, the healthcare system contributes to the recovery rate through two main factors: service quality and service capacity. The above formulation allows us to disentangle the effects of these two factors. As a strategic decision, therefore, the policymakers might decide to improve the health service quality (ϕ), e.g. investing on new medical equipment and medical training programs, or to increase the healthcare capacity (C), e.g. constructing additional hospitals so to be able to cope with demand shock without compromising the health service quality.[15]

[13]In other words, $\phi = \frac{(1-\theta')-(1-\theta)}{1-\theta}$, where θ and θ' are the recovery chance without and with receiving health service, respectively. For instance, if an infected individual's mortality risk is 30%, i.e. $\theta = 0.70$, by receiving health service of quality $\phi = 0.8$, his mortality risk will be reduced by 80%, i.e. from 30% to 6%, hence a $\theta' = 0.94$.

[14]Precisely, $C \cdot |\mathcal{I}$ is defined as demand that can be responded without overloading the healthcare system.

[15]Note that healthcare capacity C does not necessarily need to be equal to the total physical capacities of the hospitals since not all the detected cases are hospitalized. Many detected cases, if not in a severe condition, might be advised to self-isolate themselves at home till full recovery. Hence, this parameter may be construed as an adjusted

4. Network Topology

Individuals' physical contacts network is modeled as a random network consisting of communities of different sizes, where intra-community densities are larger than inter-community densities. The network is constructed based on the following three main steps. First, community-size distribution and number of communities are randomly generated. For this purpose, an algorithm for random integer partitioning is developed. Second, each community, given its size from the previous step, is constructed according to the Erdős-Rényi random graph model. Finally, the overall network is constructed by linking the communities based on a stochastic block model, specifically the assortative planted partition model.

In this section, an algorithm for partitioning $|\mathcal{I}|$ into some positive integers is introduced, where the number of parts is itself a random variable. Specifically, the objective is to generate positive integer numbers $i_1 \geq i_2 \geq \cdots \geq i_K$, where K is a random number, representing the total number of communities, i_k is the size of community k, and $|\mathcal{I}| = \sum_{k=1}^{K} i_k$. This is one of the classical NP-hard problems of combinatorial optimization.

In our model, the network of contacts is characterized based on two factors, namely variation in the community size distribution, i.e. dispersion of i_k values, and the density of links within each community. Moreover, we would like to avoid generating communities with absurdly small sizes, hence a lower bound on i_k values is at work. To generate a random partition of $|\mathcal{I}|$ that satisfies the above properties, the following probabilistic algorithm is used.

Suppose that there are $|\mathcal{I}|$ baskets, labeled 1 to $|\mathcal{I}|$, each containing a single ball. Given the parameter values for the variation in the community-size distribution and lower bound on i_k, we aim to redistribute the balls among the baskets in such a way that i) the number of non-empty baskets represents the number of communities K, and ii) the number of balls in each basket represents the size of the corresponding community, i_k. This is performed based on the following stages. At

value concerning the fraction of the detected cases that are required to be hospitalized. If in a city, for instance, only 5% of the detected cases are hospitalized, then the value of the parameter C reflects the total physical capacity multiplied by 20.

the first stage, each ball placed in the baskets 2 to $|\mathcal{I}|$ is transferred to basket 1 with a probability λ, and then basket 1 will be closed. At the second stage, each ball placed in the baskets 3 to $|\mathcal{I}|$, if any, is transferred to basket 2 with a probability λ, and then basket 2 will be closed too. The process continues until all the remaining open baskets are empty. The empty baskets are removed and the closed ones are sorted, in descending order, based on the number of balls in them. The sorted baskets are labeled as [1], [2], $\cdots$, $[n]$, $n \leq |\mathcal{I}|$, where basket [1] contains the greatest number of balls and basket $[n]$ contains the smallest number of balls.

Now let i_{min} be the a lower bound on the admissible parts, i.e. constraint $i_k \geq i_{min} \forall c$. In other words, i_{min} represents the smallest admissible size for a randomly generated community. To satisfy the constraint, the number of balls in basket $[n]$ is compared with i_{min}. If the constraint is satisfied, i.e. $i_{[n]} \geq i_{min}$, then the process terminates. Otherwise, all the balls in the basket $[n]$ are transferred to basket $[n-1]$, and basket $[n]$ is dismissed. The process is repeated for basket $[n-1]$ with an updated number of balls in it, and if this number is smaller than i_{min}, all the balls in this basket will be transferred to basket $[n-2]$, and basket $[n-1]$ will be dismissed. The process continues until all baskets contain at least i_{min} balls. At this stage the algorithm terminates, the number of remaining baskets in the process represents the number of communities, i.e. K, and the number of balls in each basket specifies the size of each community, i.e. $i_1 \geq i_2 \geq \cdots \geq i_K$. This procedure is described in Algorithm 1, which takes as input the population size, network centrality parameter γ, and minimum community size i_{min}, and returns the number of communities and community-size distribution, satisfying the constraint on the community minimum size, and that parts sizes sum up to $|\mathcal{I}|$. The parameter γ controls variation in the community-size distribution. For instance, for $|\mathcal{I}| = 1000$ and $i_{min} = 10$, two generated instances of $(i_1, \cdots, i_K)$ by Algorithm 1 when $\gamma = 0.7$ are $(725, 191, 56, 17, 11)$ and $(720, 193, 54, 33)$, while an instance when $\gamma = 0.3$ is $(318, 176, 140, 98, 72, 60, 43, 33, 20, 19, 11, 10)$[16].

[16]The numbers are reproducible in Matlab by setting the random number generator to the default seed, i.e. 0.

Algorithm 1 Random integer partitioning with restricted dispersion

procedure CommunityGeneration

Input: $|\mathcal{I}|$, γ, i_{min}

Output: K, $i_1 \geq i_2 \geq \cdots \geq i_K \geq i_{min}$

$X \leftarrow \text{ones}(|\mathcal{I}|)$

for all $j < i \leq |\mathcal{I}|$ **do**

$r \leftarrow Rand(0,1)$

$X[i] \leftarrow X[i] + \mathbb{I}[r < \gamma] * (X[j] > 0)$

$X[j] \leftarrow X[j] - \mathbb{I}[r < \gamma] * (X[j] > 0)$

end for

$X \leftarrow sort(X,\ descend)$

$X[X == 0] \leftarrow [\ ]$

while $X[end] < i_{min}$ **do**

$X[end-1] \leftarrow X[end-1] + X[end]$

$X[end] \leftarrow [\ ]$

$X \leftarrow sort(X,\ descend)$

end while

$X \leftarrow sort(X,\ descend)$

$K \leftarrow |X|$, $(i_1, \cdots, i_C) \leftarrow X$

Return $K, i_1, \cdots, i_K$

end procedure

Finally, to construct the entire network, the random adjacency matrix A of size $|\mathcal{I}| \times |\mathcal{I}|$ is generated where $A[ij] = 1$, $i < j$, indicates that individuals i and j are directly linked in the network. It is important to note that a link between two individuals does not necessarily imply friendship or any association between the two individuals, except for the possibility of physical contact. Two individuals can be linked in the network because they might visit the same supermarket in a

neighborhood, use the same bus, or cash note within a reasonably short period of time, hence capable of transmitting the virus directly to one another. The matrix A is divided into K parts where the first i_1 rows and columns correspond to the individuals in the first community, the next i_2 rows and columns correspond to the individuals in the second community, and so on. The submatrix A_k, corresponding to community k with size i_k, has an Erdős-Rény distribution, with $A_k[ij] \overset{iid}{\sim} Bern(p)$, $i < j$. In other words, any two individuals within the community are linked according to an independent Bernoulli trial with a parameter p. Two individuals i and j from two different communities, on the other hand, are directly linked according to a Bernoulli trial with parameter q, $q < p$. An illustration of the generated networks with centrality parameters $\gamma = 0.7$ and 0.3 is presented in Figure 1 by setting $p = 0.20$, and $q = p/100$ or $q = p/20$ for the same generated community-size distribution.

4.1. a Prototype Analysis

In this section, a series of sequentially related examples are presented to illustrate the model's usefulness in a set of selected settings. The objectives of these examples are threefold. First, to demonstrate the model flexibility in capturing and disentangling the effects of various components of the complex system. Second, to illustrate the types of outcomes that can be expected from the model. Third, to instantiate types of what-if analyses that can be performed using the model to inform policy analysis and decision making.

Consider a population of $|\mathcal{I}| = 1000$ individuals, distributed in a decentralized network, $\gamma = 0.3$, and $i_{min} = |\mathcal{I}|/100$, with densely connected within-community individuals, $p = 0.50$, and sparsely connected communities, $q = p/100$. In terms of behavioral aspects, suppose that the population exercises a great degree of social distancing and with a relatively low level of within-group variation, hence large value for the parameter a in (5). First, we assume a high degree of social conformity, i.e. $\pi = 1$. Later, for the same generated network, we compare the results in presence of a small heterogeneity, i.e. $\pi < 1$, depicting a scenario where a small fraction of the population breach ($\pi = 0.95$, $a' = 5$). At this stage, let us assume $\alpha_i \sim \text{Beta}(5, 1)$. This would result in an average of $\lambda \approx 0.03$ times

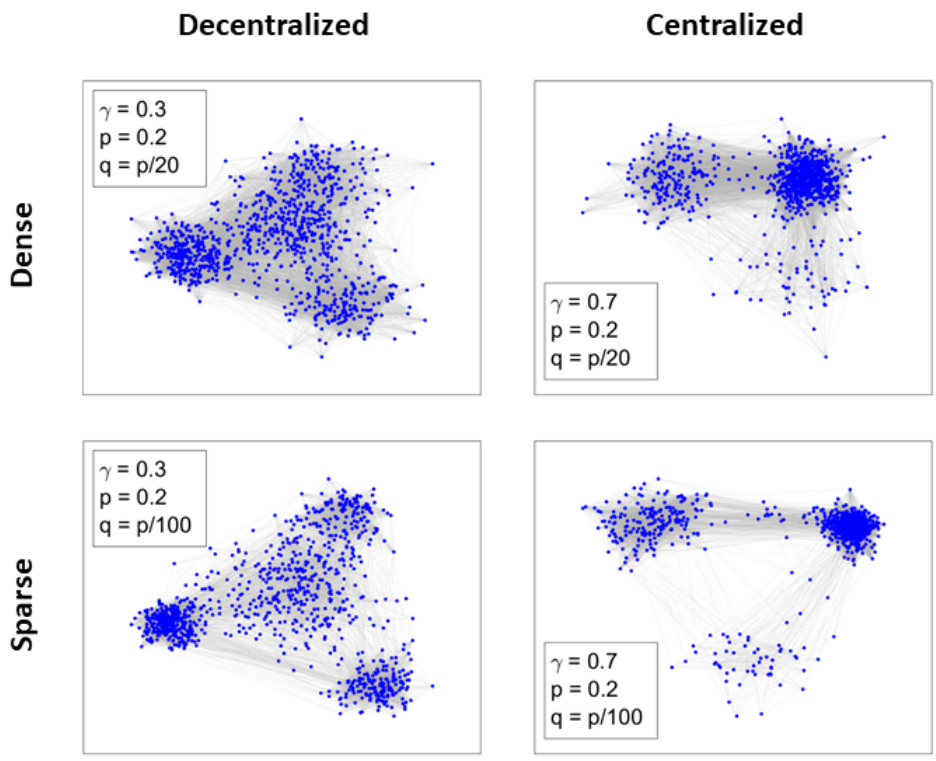

Figure 1 **Constructed random networks using Algorithm 1 for different network structures (left: decentralized – $\gamma = 0.3$, right: centralized – $\gamma = 0.7$) and different levels of inter-community connections (left: densely connected communities – $q = p/20$, right: sparsely connected communities – $q = p/100$), and fixed intra-community intensity ($p = 0.20$).**

of physical contacts in a single time period between two directly linked individuals in the network, and with an average of ≈ 0.28 number of interactions for an individual in the 5% rightmost extreme of the distribution.[17] In terms of population characteristics, suppose that $\theta \sim \text{Beta}(8, 2)$, i.e. 80% chance of self-recovery, in expectation, with a moderate heterogeneity, and $H_i \sim Geommetric(0.5)$,

[17]Precisely, define $F_\lambda(x) = \Pr(\lambda \leq x)$ and $Q_\lambda(p) = \inf\{x : F_\lambda(x) \geq p\}$. Then, in the above example, $\mathbb{E}\ (\lambda \mid \lambda \geq\ Q_\lambda(0.95)) \approx 0.28$. Note that the number of physical contacts between two individuals i and j in a time period t, i.e. Y_{ijt}, is determined by the latent distance d_{ij} that, in turn, depends on $max(\alpha_i, \alpha_j)$. Hence, assuming that the two individuals are directly linked in the network, the expected value of Y_{ijt} can be approximated numerically using the distribution of $max(\alpha_i, \alpha_j)$, derived, e.g., in Sculli and Wong (1985). Specifically, suppose that $\alpha = max(\alpha_i, \alpha_j)$, then $1/d_{ij} = (1 - \alpha)/\alpha$, hence $\mathbb{E}(\lambda_{ij}) = \mathbb{E}(1/d_{ij}^2) = 1 + \mathbb{E}(1/\alpha^2) - 2\mathbb{E}(1/\alpha)$. These expectations, and other quantities with respect to λ_{ij}, can be approximated numerically using Monte Carlo simulation and drawing from the distribution of $max(\alpha_i, \alpha_j)$, which is known under the iid assumption.

i.e. two weeks of hospitalization in expectation. In terms of contagion characteristic, assume that $\delta = 0.2$, i.e. 20% risk of transmission in a direct contact. Finally, in terms of the healthcare system, assume a high level of healthcare quality, $\phi = 0.8$, a large healthcare capacity, $H = 50$, and a relatively low testing capacity, $\tau = 0.30$, i.e. more than 3 time periods in expectation to detect an infectious agent. For the initial condition, suppose that at time $t = 0$ there are 5 infectious agents in the population. The computations continue until no infected or hospitalized case remains in the system, hence the epidemic duration, $|\mathcal{T}|$, is determined endogenously.

Holding the generated network constant and repeating the simulation several times, we can compute various types of outputs together with their uncertainty levels, hence accounting for the inherent stochasticity in the system. Thus, computations for the above setting are repeated 50 times, where each time several types of outputs are measured, including the fraction of population infected or perished, as well as epidemic duration. In each round of the simulation, the total number of infected, detected, recovered, and perished cases are calculated for each period. Moreover, the new number of infected cases at each time, as well as the total number of active cases, i.e. infected but not detected, are stored. Besides, the basic reproduction number ($\mathcal{R}_0$) that is the average number of people to which a single infected person will transmit the virus is calculated and stored for each period. This is a widely used metric in epidemiology for tracking the pace of the virus spread. Finally, the epidemic recovery pace, defined as the proportion of time between the epidemic peak to the system full recovery with reference to the total epidemic duration, is measured for each setting. This measure is important because it indicates the relative position of maximum load on the healthcare system over the entire period. To enhance the practical relevance of this measure, a peak time in our analysis is identified using the detected cases, instead of infected cases, since the former is observable while the other is latent. Specifically, the peak time is defined as $t_{peak} := \sup\{\tau : d_\tau \geq d_t \ \forall t \in \mathcal{T}\}$, where d_t is the number of new detected cases in a time period t. Subsequently, the ratio $\nu = t_{peak}/|\mathcal{T}|$ indicates the recovery speed of the system. For instance, $\nu = 0.5$ indicates that the recovery time is of the same length as the time duration of the epidemic from its start to its

peak in terms of detected cases, while $\nu = 0.2$ indicates that the recovery duration is 4 times longer than start to the peak duration of the epidemic. For all these measures, 99% confidence intervals are computed by bootstrapping, using the bias-corrected and accelerated percentile method and $10,000$ bootstrap samples. Results for the above setting are presented in Figure 2– top row (note different right-side vertical axis units for perished cases).

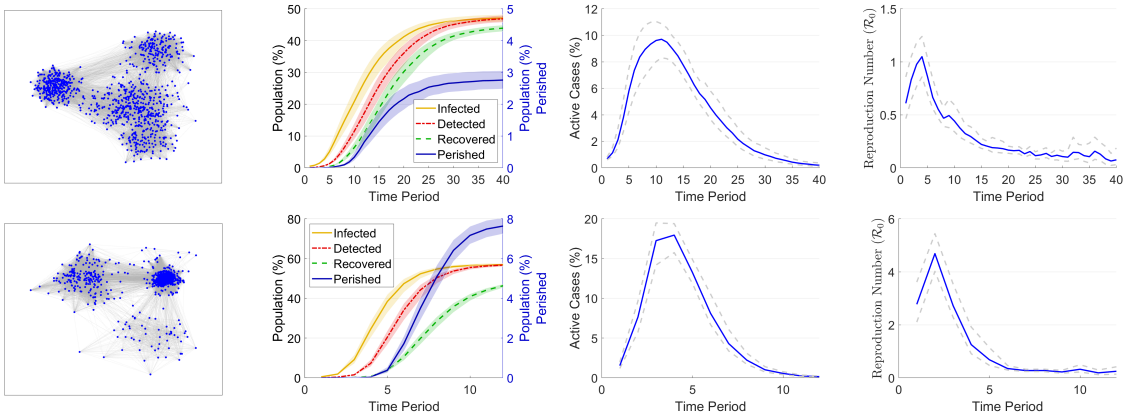

Figure 2 **(Color online) Results of the simulation corresponding to setting 1 (top) and setting 4 (bottom) in the prototype analysis. The leftmost figure represents the generated random network, held constant in the 50 replications. The other figures visualize the state of the network over time and represent, respectively from left to right: population fraction that is infected, detected, recovered, or perished at each period (note different right-side vertical axis units for the perished cases); population fraction that is active, i.e undetected infectious agents, at each time; basic reproduction number at each time. Shaded areas and grey dashed line intervals represent 99% bootstrap confidence intervals, computed using the bias-corrected and accelerated percentile method, and $10,000$ bootstrap samples.**

In the setting described above, the epidemic lasts, in expectation, for $|\mathcal{T}| = 42.8$ periods (with a 99% confidence interval of 38.6 to 46.8) and in total affects 45.4 percent of the population (with a 99% confidence interval of 39.1 to 47.7), resulting to passing of 2.7 percent of the population (with a 99% confidence interval of 2.3 to 3.0). Moreover, the recovery pace is as small as 0.27 (with a 99% confidence interval of 0.24 to 0.30). Even though the basic reproduction number barely exceeds its critical value of 1, nonetheless the combination of a considerable incubation period and low testing capacity leads to an outbreak.

In the next stage, we depict a setting related to the previous one by holding constant all the characterizing parameters of the setting, as well as the generated random network, but assuming a low level of heterogeneity in practicing social distancing behavior, specifically assuming that 5% of the population breach ($\pi = 0.05$). By conducting the same analysis as in the previous setting, results indicate a moderate increase in the total number of infected cases ($\sim 15\%$), but a dramatic increase in the number of perished cases ($\sim 50\%$).[18] Moreover, the epidemic duration is decreased ($\sim 0.18\%$), while the recovery pace remains unchanged. For the same setting, increasing testing capacity fourfold, i.e. $\tau = 1 - (1 - 0.3)^4 = 0.76$, compensates the small heterogeneity introduced in the system in terms of the number of perished cases, brings this number down to 2.5 (99% confidence interval 2.1 to 2.8) that is close to the full social conformity results in the previous setting 1. Moreover, this change further reduces the number of infected cases (30 versus 45.4 in setting 1) and shrinks the epidemic period (19.2 versus 42.8 in setting 1), also slightly increases the recovery pace. Detailed descriptions of the settings and summaries of the results are presented in Tables 1 and 2, respectively. Changes at each stage are highlighted.

	Network structure			Infection	Health status			Healthcare system			Social distancing		
Setting	γ	p	q	δ	θ_0	θ_1	η	τ	ϕ	C	a	a'	π
1	0.3	0.50	$p/100$	0.2	8	2	0.5	0.30	0.8	50	5	5	1
2	0.3	0.50	$p/100$	0.2	8	2	0.5	0.30	0.8	50	5	5	**0.95**
3	0.3	0.50	$p/100$	0.2	8	2	0.5	**0.76**	0.8	50	5	5	0.95
4	**0.7**	0.50	$p/100$	0.2	8	2	0.5	0.76	0.8	50	5	5	0.95
5	0.7	0.50	$p/100$	0.2	8	2	0.5	0.76	0.8	**100**	5	5	0.95
6	0.7	**0.25**	$p/100$	0.2	8	2	0.5	0.76	0.8	100	5	5	0.95
7a	0.7	0.25	$p/100$	0.2	8	2	0.5	0.76	0.8	100	5	5	**1**
7b	0.7	0.25	$p/100$	0.2	8	2	0.5	0.76	0.8	100	5	5	**0.80**

Table 1 Parameter values of the sequentially related settings, identical except for the highlighted values, in the prototype analysis ($|\mathcal{I}| = 1000$).

[18] In order to be able to attribute the changes in the outcomes to the change in the parameter value, before simulating the system we have restored the random number generator seed, hence the two scenarios and all the generated random numbers are identical in all aspects except for the modified parameter value.

Setting	Condition change	Infected (%)			Perished (%)			Epidemic duration $\lvert\mathcal{T}\rvert$			Recovery pace ν		
		Mean	99% Conf. Int.		Mean	99% Conf. Int.		Mean	99% Conf. Int.		Mean	99% Conf. Int.	
1	—	45.4	39.1	47.7	2.7	2.3	3.0	42.8	38.6	46.8	0.27	0.24	0.30
2	$\pi: 1.00 \rightarrow 0.95$	51.8	50.8	52.7	4.0	3.7	4.3	36.3	34.6	38.5	0.27	0.25	0.30
3	$\tau: 0.30 \rightarrow 0.76$	30.0	26.1	31.4	2.5	2.1	2.8	19.2	18.0	20.5	0.34	0.30	0.39
4	$\gamma: 0.3 \rightarrow 0.7$	57.1	56.3	57.8	7.7	7.4	8.2	18.0	17.1	18.9	0.27	0.25	0.29
5	$C: 50 \rightarrow 100$	58.2	57.5	59.2	5.9	5.6	6.2	17.4	16.8	18.3	0.26	0.25	0.28
6	$p: 0.50 \rightarrow 0.25$	43.2	42.0	44.2	3.0	2.8	3.3	18.0	17.2	18.8	0.29	0.27	0.31
7a	$\pi: 0.95 \rightarrow 1.00$	37.0	32.7	38.2	2.1	1.8	2.3	19.0	17.7	20.1	0.34	0.31	0.36
7b	$\pi: 0.95 \rightarrow 0.80$	55.6	54.9	56.4	5.6	5.3	5.9	16.4	15.5	17.5	0.24	0.22	0.26

Table 2 Results summaries of the sequentially related settings, identical except for the parameter in condition change column, in prototype analysis ($|\mathcal{I}| = 1000$). The 99% confidence intervals are computed by bootstrapping and using the bias-corrected and accelerated percentile method, and $10,000$ bootstrap samples.

In the next stage, let us project the previous setting to a centralized network, hence maintaining all the parameter values constant except for changing γ from 0.3 to 0.7. Such a configuration in a centralized network leads to a very different picture than its decentralized counterpart. Specifically, by changing the parameter value, we observe a dramatic increase in the number of infected and perished cases, doubled and tripled respectively. Nevertheless, the epidemic duration remains unchanged. For a better comparison, additional details are provided in Figure 2– bottom row (note different scale ranges compared to the top row). It is observed that the number of active cases, i.e. undetected infectious agents, grows faster and reaches a higher peak in the centralized network, which is also reflected by the values of the basic reproduction number over time. These results demonstrate the prominent role of network structure. While variations in outputs due to stochasticity in agents' behavior are marginal (narrow confidence intervals), variations due to changes in the network structure are of the order of magnitude.

In the subsequent stages, different public health intervention alternatives are explored, one at each stage, so to mitigate the adverse impacts introduced by the increased network centrality. At the first stage, the merits of increasing healthcare capacity are explored via doubling C, i.e. changing from 50 to 100. This has reduced the number of perished cases by about 23% while had no impact on the number of infected cases, epidemic duration, or recovery pace. Since a major cause of the scaled-up numbers in a centralized network might be attributed to the shorter paths length between

pairs of individuals, due to the presence of a "hub", weakening contact connections in the network might be an effective response. Therefore, in the next step, constant decontamination of surfaces and public areas is considered. This would decrease the rate of mediated virus transmission since a method of communication of the virus from one person to another is through physical contact to common objects and surfaces. Hence, such intervention should be translated into the parameter related to the intra-community density p. Changing p from 0.50 to 0.25, not only resulted in a decline in the number of infected cases, as was expected, also substantially reduced the number of perished cases ($\sim 50\%$).

In the next stage, in addition to the previous measures, surveillance is tightened by enforcing maximal conformity to social distancing, hence updating π from 0.95 to 1 (setting $7a$). In a parallel opposite scenario, we look into the results of a further de-escalation of social distancing restrictions, changing π from 0.95 to 0.80 (setting $7b$). In contrast to the decentralized environment, tightening surveillance had no impact on epidemic duration, but increased the recovery pace. It implies that a small heterogeneity in social distancing can shift the peak time to the left side in both centralized and decentralized networks, but the shift in the centralized network is attributed to diminishing recovery pace, while in the decentralized network it is due to shortening the epidemic duration– see Figure 3.

The above results exhibit the interrelated complex nature of the interactions between the contextual factors and policy parameters, as well as the nonlinear behavior of the system. Nevertheless, these results are obtained by holding constant the generated random network for each setting. Even though relatively narrow confidence intervals demonstrate low variation in the measured outputs due to agents' stochastic behavior and attributes, to enhance the external validity of the results, randomness in the generated network must be accounted for, conditioning on the network structure parameters. This is discussed in the next session where each setting is repeatedly analyzed, in the same way as before, but also across a range of randomized manifestations of the network.

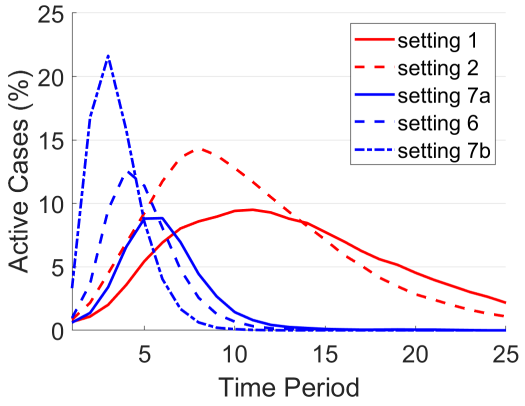

Figure 3 **(Color online) Mean value of active cases over the time for settings** 1 **and** 2**, red solid and dashed lines, respectively, corresponding to** $\pi = 1$ **and** $\pi = 0.95$ **in decentralized network; as well as settings** $7a$**,** 6**, and** $7b$**, solid, dashed, and dash-dotted lines, respectively, corresponding to** $\pi = 1$**,** 0.95 **and** 0.80 **in centralized network. Note that red and blue lines should not be compared in absolute values since they differ in other setting parameters too.**

5. Experimental Analysis

This section presents an extensive experimental analysis. The main objective of this analysis is to examine interrelationships among the characterizing parameters of the system and its outputs, hence to derive insights for informing policy decisions across a range of settings. The design parameters of the experiment and their corresponding levels are provided in Table 3. For each network structure, characterized by the triple $\langle \gamma, p, q \rangle$, the baseline condition is defined as $(\theta_0, \theta_1) = (5, 5)$, $\tau = 0.30$, $\phi = 0.40$, $C = 50$, and $\pi = 0.80$. Since we are interested in the *improvements*, in the respective measures, made by various health intervention methods, rather than the absolute values per se, the outputs of each setting are compared with its respective baseline condition in the same network, thus differences in the values are stored for further analysis. More importantly, such a measurement method enables us to obtain consistent estimates of the intervention effects by eliminating the idiosyncratic impacts of the specific generated random network in each instance of a scenario.

Network structure			Infection	Health status		Healthcare system			Social distancing		
γ	p	q	δ	(θ_0, θ_1)	η	τ	ϕ	C	a	a'	π
{ 0.30, 0.70 }	{ 0.25, 0.50 }	$p/100$	0.20	{ (5, 5), (8, 2) }	0.50	{ 0.30, 0.76 }	{ 0.40, 0.80 }	{ 50, 100 }	5	5	{ 0.80, 0.95, 1 }

Table 3 Parameters and their levels in the experimental analysis ($|\mathcal{I}| = 1000$).

Throughout the analysis, population size is fixed at $|\mathcal{I}| = 1000$. Each setting is specified by a combination of the above levels. The experiment is performed based on the following main steps.

Step 1: Specify γ and $i_{min} = |\mathcal{I}|/100$, and generate number of communities, K, and community-size distribution, $i_1, \cdots, i_K$, according to Algorithm 1. Moreover, specify p and q.

Step 2: Generate a random network based on the stochastic block model. Hold the generated network constant in steps 3 and 4.

Step 3: Set π and generate individual-specific social distancing parameter values α_i by drawing from the distribution in 5. Set θ_0 and θ_1 and generate individual-specific self-recovery chance parameter values θ_i. Set τ and η and generate incubation and hospitalization period values for each individual, i.e. T_i and H_i, respectively. Finally, specify value of ϕ, that is the health service quality in the absence of demand overload, and C, that is the healthcare system capacity. Simulate the system and compute total fraction of infected and perished cases, as well as epidemic duration and recovery pace.[19] Hold the generated network constant and repeat the analysis 50 times and store the mean values of the respective outputs over the 50 replications.

[19]For each time period t in the time window $t = 1, \cdots, |\mathcal{T}|$ do the followings. Construct $\mathcal{I}' \subseteq \mathcal{I}$ by excluding the recovered and perished cases from $\mathcal{I}$. Count the number of currently hospitalized individuals and update ϕ_t^{eff} and θ'_{it} according to (8). For each pair of individuals $i, j \in \mathcal{I}'$ with $S_{it} = 0$ (i is healthy at t) and $S_{jt} = 1$ (j is infected but not detected at t) generate number of physical contacts between them at period t, only if they are directly linked in the generated network, according to $Y_{ijt} \sim Poisson(\lambda_{ij})$, $\lambda_{ij} = 1/d_{ij}^2$, and $d_{ij} = \frac{\alpha_i e^{n\alpha_i} + \alpha_j e^{n\alpha_j}}{e^{n\alpha_i} + e^{n\alpha_j} - \alpha_i e^{n\alpha_i} + \alpha_j e^{n\alpha_j}}$, and update the status of i according to $Bern(1 - (1-\delta)^{Y_{ijt}})$. For an already infected individual j, $S_{jt} = 1$, update the status to Detected if incubation period T_i has been reached and update the number of hospitalized individuals. For a detected individual j, update the status if time period H_i of been hospitalized has been reached, to recovered with probability θ'_{it}, or perished with a probability of $1 - \theta'_{it}$. Update $\mathcal{I}'$.

Step 4: Repeat step 3 for all the configurations defined by the levels of the parameters that are required to be specified at that step, including the baseline condition. Compute and store differences in the measured outputs with reference to the baseline condition.

Step 5: Repeat steps 2 to 4 for 20 times, each time with a different generated network.

Step 6: Go back to step 1, update levels of the parameters that are required to be specified in that step, i.e. network structure parameters, and repeat steps 2 to 5.

Technical details: Overall, $192,000$ instances of scenarios were simulated. All simulations were implemented in MATLAB $R2019b$. The computations were performed on a computer cluster with 27 computing nodes (all nodes connect to IBM Spectrum Scale with Infiniband), with a total of 720 cores, and approximately 7.4 TB of RAM. Total execution time was 355 hours.

5.1. Results

This section presents the main results of the experimental analysis. The impacts of the policy parameters, for example, enforcing social distancing or expanding testing capacities, on each of the four measures that are used describe the epidemic outcome, for example, population fraction of perished cases or epidemic duration, are estimated using ordinary least squares regression. As mentioned before, the measurements of the above outcomes are performed with reference to the baseline condition, specified above, hence the parameters estimates represent changes in improvements in the respective outputs obtained by adapting a health intervention method.

Results for reductions in the number of perished cases and the number of infected cases, in terms of population percentages, are summarized in Figures 4 and 5, respectively. The results concerning the perished cases are separated across the four types of network structure, based on levels of network centrality and community sparsity, and also for the population's overall health status. For the infected cases, similar patterns were observed for the healthy and frail populations, hence parameters are estimated by combining observations from both these groups. The same holds for the results of the epidemic duration, presented in Table 4, implying that the impact of intervention methods on epidemic duration is not contingent on the population's overall health status. For

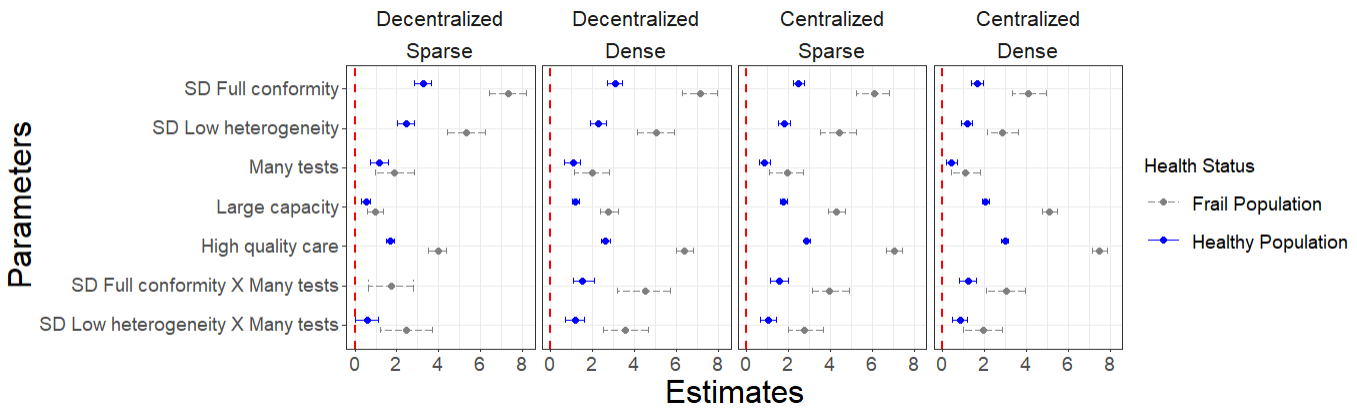

Figure 4 **(Color online) Coefficient plots of OLS parameter estimates of the impact of intervention methods on reduction in the total number of perished cases (cutback in the number of perished cases with reference to the baseline condition). Sample size $n = 460$ for each of the 4 regression models corresponding to the frail population, and $n = 480$ for each of the 4 regressions corresponding to the healthy population. The 99% confidence intervals are computed by bootstrapping and using the bias-corrected and accelerated percentile method and 1000 bootstrap samples. Models' R^2 corresponding to decentralized networks: for dense, 92% and 90%, and for sparse, 87% and 82%, for healthy and frail population respectively. These values are $\sim 92\%$ for all the 4 models corresponding to the centralized networks.**

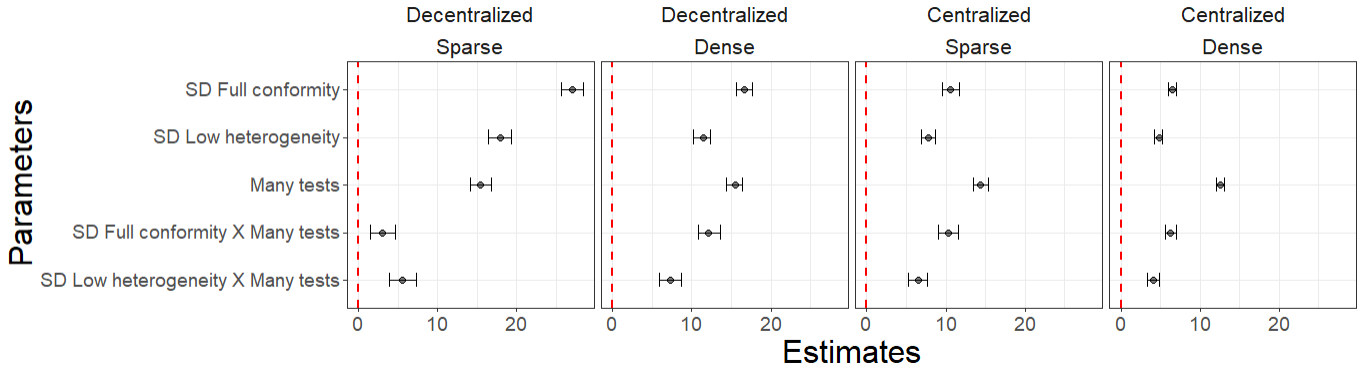

Figure 5 **(Color online) Coefficient plots of OLS parameter estimates of the impact of intervention methods on reduction in the total number of infected cases (cutback in the number of infected cases with reference to the baseline condition). Sample size $n = 940$ for all the 4 models. The 99% confidence intervals are computed by bootstrapping and using the bias-corrected and accelerated percentile method and 1000 bootstrap samples. Model $R^2 \geq 98\%$ for all the 4 models.**

	Decentralized						Centralized					
	Sparse			Dense			Sparse			Dense		
	Est.	99% Conf. Int.		Est.	99% Conf. Int.		Est.	99% Conf. Int.		Est	99% Conf. Int.	
SD Full conformity	6.5	5.1	7.9	7.6	6.7	8.5	6.9	6.2	7.7	3.8	3.2	4.4
SD Low heterogeneity	4.9	3.4	6.1	4.1	3.5	4.8	3.6	2.8	4.3	2.3	1.8	2.9
Many Tests	-15.6	-16.7	-14.5	-14.6	-15.2	-14.0	-14.2	-14.9	-13.5	-13.5	-14.0	-12.9
SD Full conformity × Many tests	-11.3	-13.0	-9.8	-5.0	-6.0	-4.0	-4.6	-5.5	-3.8	-0.8	-1.4	-0.1
SD Low heterogeneity × Many tests	-6.0	-7.4	-4.5	-2.1	-3.0	-1.3	-1.9	-2.7	-1.1	-0.8	-1.4	-0.2

Table 4 OLS parameter estimates of the impact of intervention methods on changes in epidemic duration (change in $|\mathcal{T}|$ with reference to the baseline condition). Model $R^2 \geq 98\%$ and sample size $n = 940$ for all the 4 models.

	Decentralized			Centralized		
	Est.	99% Conf. Int.		Est.	99% Conf. Int.	
SD Full conformity	11	10.1	12.0	4	3.7	4.6
SD Low heterogeneity	8	7.1	8.6	3	3.0	3.6
Many Tests	9	8.4	9.8	8	7.2	7.9
SD Full conformity × Many tests	-9	-10.6	-8.1	2	1.6	3.0
SD Low heterogeneity × Many tests	-5	-6.1	-4.1	1	0.1	1.1

Table 5 OLS parameter estimates ($\times 10^2$) of the impact of intervention methods on changes in recovery pace (change in ν with reference to the baseline condition). Model $R^2 = 93\%$ for centralized, and 72% for decentralized setting, and sample size $n = 1880$ for both models.

the recovery pace, on the other hand, the only factor introducing contingencies to the impact of intervention methods was the network centrality– see Table 5.

Concerning the population fraction of perished cases, it can be observed that, in general, a system with frail populations, e.g. with a constrictive population pyramid, is more responsive to the health intervention methods. Results show that as network centrality and community density increase, importances of healthcare quality and capacity increase too, while that of social distancing decreases unless it is combined with a large testing capacity. This is particularly important given the trouble and considerable economical and social consequences of enforcing social distancing in centralized and dense networks. In other words, such a costly intervention method, when network centrality and community density increases, can be justified only in the presence of an extensive testing capacity. Moreover, it is observed that the joint effect of social distancing combined with many tests in the centralized networks decreases by 25% when only 5% of the population stop complying with the social distancing requirements. Moreover, this observation is valid for both healthy and frail populations. Hence, these results suggest shifting the focus toward healthcare

capacity and quality when operating in a more centralized network and with denser communities. Furthermore, in such contexts, when measures based on social distancing are going to be adopted, first, they must be combined with expanding testing capacities, second, they must be implemented with strict surveillance.

The results demonstrate several important direct and indirect benefits of investing in expanding testing capacity. It can be observed that, despite the other intervention methods, the direct impact of increasing the number of tests is independent of the population health status in all the respective measures used to describe epidemic behavior. Notably, while the direct impact of increasing the number of tests on reduction in the number of perished cases is relatively small, yet it is a highly effective intervention method since it moderates the impact of social distancing; plays an important role in reducing the number of infected cases and increasing the recovery pace, e.g. shifting the peak to the right side of the time scale. Moreover, it remains the prevailing factor in reducing epidemic duration across all types of networks, which is essential to mitigate the adverse economical consequences of a strict self-isolation policy.

6. Conclusion

An effective response to the outbreak of infectious disease requires a meticulous consideration of the features of the embedding environment. In such domains, contextual factors such as characteristics of the population, their overall health status, compliance behavior, patterns of contacts, as well as the level of heterogeneity in individual-level attributes largely interact with the policy parameters to yield the outcome. Policy measures, on the other hand, can alter some features of the environment, at least partly, such as by enforcing compliance behavior via punitive methods. From this perspective, there is no one-size-fits-all solution, and elements of an effective strategy should be designed in the light of interrelationships among these groups of factors.

This paper presents a unified framework that integrates key components of the contagion process with the key characteristics of the embedding environment and by accounting for heterogeneity in individual-level attributes, as well as sources of uncertainty in different layers of the system.

The model is primarily intended to serve as a tool for informing policy decisions concerning public health interventions; an objective that establishes the main rationale behind the model construction process. To this aim, the main features of the contagion model are integrated with the characteristics of the environment and the relevant policy parameters. Components of the model are subsequently synthesized in a broader agent-based model that enables accounting for heterogeneity in individual-level attributes that collectively yield the macro-level outcomes.

Among the contextual factors, special attention is devoted to the critical role of the patterns of contacts between individuals, i.e. network structure. First, such patterns directly control the virus transmission disease spread in the population. Secondly, supported by our results, network structure largely interacts with policy parameters, making their effectiveness contingent on the characteristics of the network. In our model, the network structure is characterized by two simple factors, namely the degree of centrality in the network topology, and density of connections within components of the network called communities. Conditioning on these two factors, random networks are generated using an algorithm that is introduced in this paper. Our empirical results, with reasonably narrow confidence intervals when conditioning on centrality and density, show that these two factors can capture a large proportion of the variation in the outcome stemming from the inherent randomness in the network. The results demonstrate the critical role of the network structure as a substantial element in the design of an effective response to the outbreak of infectious diseases.

Aside from the methodological contributions, we presented a series of stylized examples to illustrate the applicability of the model and conducted an extensive controlled experiment to derive insights concerning the overall tendencies of various types of outcomes and effectiveness of different types of policy parameters in relation to the contextual factors. The results highlight the substantial importance of social distancing and self-isolation but also showing that when network centrality and density increase, its impact progressively becomes contingent on the presence of a large testing capacity. Moreover, the results suggest that expanding testing capacity is essential in order to mitigate the inevitable impact of social distancing on epidemic duration, which subsequently might translate into economic losses. Finally, in terms of the number of perished cases,

we observed that when network centrality and density increase, healthcare system capacity and care quality dominate other factors, while social distancing prevails in decentralized and sparse networks.

Beyond what has been discussed in this paper, the proposed framework can be extended in various directions. A specifically interesting venue is to incorporate macro-level covariates to explain heterogeneity in the individual-level attributes. Among those, income distribution, population pyramid type, access to, and usage patterns of high-speed internet, lifestyle parameters and dimensions of national culture in the same manner measured by the Hofstede model, might be among the propitious candidates. Other important directions for future research include improving the computational demand of the model, which arguably is a major limitation of the practical use of agent-based models as a whole. Moreover, most recent criticisms towards the agent-based modeling paradigm are directed to the calibration and validation challenges, which can restrict the scope of its finding to qualitative and suggestive levels (Leombruni and Richiardi 2005, Marks 2007). This is a promising direction to explore. In this regard, there is a growing body of research focusing on the estimation techniques based on simulated minimum distance (Fabretti 2013). These methods attempt to minimize a type of distance between the observed and simulated data, e.g. as in (Marks 2013), and they usually involve a considerable computational cost. In the context of the model developed in this paper, it would be interesting to exploit the method of simulated moments (McFadden 1989, Banerjee et al. 2013) and explore potentials of nonparametric simulated maximum likelihood combined with Bayesian inference due to the computational benefits and the possibility to fully exploit informational content of the data (Grazzini et al. 2017, Kukacka and Barunik 2017).

References

Abbe E (2017) Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research* 18(1):6446–6531.

Abrahamson E, Rosenkopf L (1997) Social network effects on the extent of innovation diffusion: A computer simulation. *Organization science* 8(3):289–309.

Albert R, Jeong H, Barabási AL (1999) Diameter of the world-wide web. *nature* 401(6749):130–131.

Atkeson A (2020) What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios. Technical report, National Bureau of Economic Research.

Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO (2013) The diffusion of microfinance. *Science* 341(6144).

Barabási AL, Albert R (1999) Emergence of scaling in random networks. *science* 286(5439):509–512.

Barabási AL, Albert R, Jeong H (2000) Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: statistical mechanics and its applications* 281(1-4):69–77.

Barabási AL, Bonabeau E (2003) Scale-free networks. *Scientific american* 288(5):60–69.

Berger DW, Herkenhoff KF, Mongey S (2020) An seir infectious disease model with testing and conditional quarantine. Technical report, National Bureau of Economic Research.

Bonabeau E (2002) Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences* 99(suppl 3):7280–7287.

Broido AD, Clauset A (2019) Scale-free networks are rare. *Nature communications* 10(1):1–10.

Burt RS (1980) Models of network structure. *Annual review of sociology* 6(1):79–141.

Chiou L, Tucker C (2020) Social distancing, internet access and inequality. Technical report, National Bureau of Economic Research.

Davis GF (1991) Agents without principles? the spread of the poison pill through the intercorporate network. *Administrative science quarterly* 583–613.

Easley D, Kleinberg J, et al. (2010) *Networks, crowds, and markets*, volume 8 (Cambridge university press Cambridge).

El-Sayed AM, Scarborough P, Seemann L, Galea S (2012) Social network analysis and agent-based modeling in social epidemiology. *Epidemiologic Perspectives & Innovations* 9(1):1.

Epstein JM (2006) *Generative social science: Studies in agent-based computational modeling* (Princeton University Press).

Erdős P, Rényi A (1960) On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* 5(1):17–60.

Fabretti A (2013) On the problem of calibrating an agent based model for financial markets. *Journal of Economic Interaction and Coordination* 8(2):277–293.

Fernandes N (2020) Economic effects of coronavirus outbreak (covid-19) on the world economy. *Available at SSRN 3557504* .

Fornaro L, Wolf M (2020) Covid-19 coronavirus and macroeconomic policy .

Fristedt B (1993) The structure of random partitions of large integers. *Transactions of the American Mathematical Society* 337(2):703–735.

Gilbert EN (1959) Random graphs. *The Annals of Mathematical Statistics* 30(4):1141–1144.

Goldenberg A, Zheng AX, Fienberg SE, Airoldi EM, et al. (2010) A survey of statistical network models. *Foundations and Trends® in Machine Learning* 2(2):129–233.

Goldenberg J, Han S, Lehmann DR, Hong JW (2009) The role of hubs in the adoption process. *Journal of marketing* 73(2):1–13.

Grazzini J, Richiardi MG, Tsionas M (2017) Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control* 77:26–47.

Guerrieri V, Lorenzoni G, Straub L, Werning I (2020) Macroeconomic implications of covid-19: Can negative supply shocks cause demand shortages? Technical report, National Bureau of Economic Research.

Hethcote HW (2000) The mathematics of infectious diseases. *SIAM review* 42(4):599–653.

Iyengar R, Van den Bulte C, Valente TW (2011) Opinion leadership and social contagion in new product diffusion. *Marketing Science* 30(2):195–212.

Jackson MO (2010) *Social and economic networks* (Princeton university press).

Karrer B, Newman ME (2011) Stochastic blockmodels and community structure in networks. *Physical review E* 83(1):016107.

Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115(772):700–721.

Kukacka J, Barunik J (2017) Estimation of financial agent-based models with simulated maximum likelihood. *Journal of Economic Dynamics and Control* 85:21–45.

Leombruni R, Richiardi M (2005) Why are economists sceptical about agent-based simulations? *Physica A: Statistical Mechanics and its Applications* 355(1):103–109.

Mahajan V (2010) Innovation diffusion. *Wiley International Encyclopedia of Marketing* .

Manshadi V, Misra S, Rodilitz S (2020) Diffusion in random networks: Impact of degree distribution. *Operations Research* Doi: 10.1287/opre.2019.1945.

Marks RE (2007) Validating simulation models: a general framework and four applied examples. *Computational economics* 30(3):265–290.

Marks RE (2013) Validation and model selection: Three similarity measures compared. *Complexity Economics* 2(1):41–61.

McFadden D (1989) A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society* 995–1026.

McKibbin WJ, Fernando R (2020) The global macroeconomic impacts of covid-19: Seven scenarios .

Muller E, Peres R (2019) The effect of social networks structure on innovation performance: A review and directions for research. *International Journal of Research in Marketing* 36(1):3–19.

Newman M (2003a) Random graphs as models of networks. *Handbook of graphs and networks* 1:35–68.

Newman M (2003b) The structure and function of complex networks. *SIAM review* 45(2):167–256.

Newman M (2018) *Networks* (Oxford university press).

Obiols-Homs F (2020) Precaution, social distancing and tests in a model of epidemic disease. Technical report.

Piguillem F, Shi L, et al. (2020) The optimal covid-19 quarantine and testing policies. Technical report, Einaudi Institute for Economics and Finance (EIEF).

Rahmandad H, Sterman J (2008) Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science* 54(5):998–1014.

Riley S (2007) Large-scale spatial-transmission models of infectious disease. *Science* 316(5829):1298–1301.

Sculli D, Wong K (1985) The maximum and sum of two beta variables and the analysis of pert networks. *Omega* 13(3):233–240.

Tracy M, Cerdá M, Keyes KM (2018) Agent-based modeling in public health: current applications and future directions. *Annual review of public health* 39:77–94.

Van den Bulte C, Stremersch S (2004) Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science* 23(4):530–544.

Willinger W, Alderson D, Doyle JC (2009) Mathematics and the internet: A source of enormous confusion and great potential. *Notices of the American Mathematical Society* 56(5):586–599.