# Social Distance Policies in ECON-EPI Networks*

Marina Azzimonti
Stony Brook University
and NBER

Alessandra Fogli
FRB of Minneapolis
and CEPR

Fabrizio Perri
Mark Ponder
University of Minnesota

July 2020

## Abstract

We analyze the economic consequences of social distancing policies designed to mitigate and contain the spread of COVID19 and evaluate a series of re-opening strategies in a network model of social interaction, the 'ECON-EPI network model.' We construct a city where households interact in their homes, with friends, or through work, school, public transportation, and shopping. We calibrate the model to New York City, and impose lock-down policies implemented there between February and April 2020, to match both disease transmission and individual mobility. We illustrate how the trade-off between disease mitigation and output loss is affected by the timing and structure of alternative policies, emphasizing the efficacy of targeting high-contact sectors of the economy. We then analyze alternative re-opening strategies, and find significant output gains in re-opening the manufacturing sector with almost no increase in infections. Opening large retail establishments (such as malls and stadiums), on the other hand, causes a second wave of infections in the Fall.

*Keywords*: COVID19, Epidemiology, SIR, Complex Networks, Pandemic.

*We thank Maria Cristina De Nardi, as well participants at several seminars and conferences for great comments and suggestions. Also many thanks to Dhananjay Ghei and Thomas Gill for outstanding research assistance.

# 1 Introduction

The United States has been one of the most affected countries during the COVID-19 pandemic, with a vast majority of the cases (to date) concentrated in large urban areas. State government officials enacted stricter policies as cases continued to raise involving, eventually, a complete lock-down of non-essential workers in the New York metro area by March 20th, 2020. These measures severely depressed economic activity, which declined up to 80% in the NY manufacturing and retail sectors the following month. The rapid spread of the disease presented an important challenge to policymakers, as the trade-off between the epidemiological cost (lives) and the economic cost (livelihoods) became apparent. The key question that motivates this paper is how to design *smart* mitigation policies and re-opening strategies that are effective in reducing the spread of the disease, while at the same time minimizing the output loss.

Our point of departure is that the spread of infections and economic activity happen through the same network of human interactions. We develop an 'ECON-EPI network model,' characterized by three components. The first one is the *network* of human interactions that specifies contacts among individuals through different network layers. The second one is the *econ* component, which describes how economic activity is created on the network. Finally, we have the *epi* component, specifying how the disease spreads through the network. We embed the standard SEIR model—which extends the one developed by Kermack and McKendrick (1927)—into a network model of disease progression and economic activity in the spirit of Fogli and Veldkamp (2020).

The spread of COVID19 within our multi-layered network is simulated in an economy where individuals are heterogeneous in terms of age, family structure, occupation, shopping behavior, public transforation, and social interactions at home, school, and with friends. These characteristics, which are modeled through different links between agents in the network, affect the frequency and intensity of interactions, and determine individual-specific probabilities of infection. These probabilities are affected endogenously by the progression of the disease (as sick agents quarantine) and exogenously by lock-down policies. Our detailed micro-foundation of agent-based infection risk is important when designing targeted policies, as imposing stay-at-home orders has very different epidemiological and economic consequences depending on which groups are affected by them. The topology of the network (i.e. the arrangement of its elements and the strength of their connections) determines the rate of transmission of COVID19, which in turn implies a particular time-path for the number of susceptible, exposed, infected, and recovered individuals in the aggregate.

The econ component is comprised of two sectors, the Low-contact and the High-contact sectors. The former captures production activities organized in stable teams that meet frequently and do not need to interact with customers. Examples are the manufacturing sector, the agricultural sector, and mining. Within this sector, some workers are able to work from home even when stay-at-home orders are imposed (such as software developers and educators). Because they have lower contact with other individuals, workers in this sector exhibit, in equilibrium, lower infection risk. The high contact sector, on the other hand, is characterized by higher proximity to customers. Links are more unstable and drawn from a wider set of individuals in the network. Examples are retail, hospitality, entertainment, and the health care sectors. A distinctive characteristic is that production in this sector requires both, a worker and a shopper. This is important because when either is under quarantine or subject to a stay-at-home order, production declines. We consider two types of H-sector establishments: 'mom and pop shops,' who draw geographically close and frequent customers, and 'large establishments,' who draw a large number of customers from the whole city (such as shopping malls, stadiums, and concerts). The economic consequences of lock-down policies and their effectiveness depends importantly on which occupations are ultimately targeted by them.

We calibrate our ECON-EPI network model to the New York metro area in the first few weeks of the pandemic. We reduce the number of network links in order to match the sharp decline in home, retail, and work mobility as reported by the Google mobility indexes in these three categories. This strategy allows us to capture, parsimoniously, lock-down policies and the endogenous response of individuals to the disease. It also allows us to pin down some unobservable parameters of the disease, such as the probability of contagion conditional on contacting an infectious individual, and the proportion of asymptomatic agents. Using sectoral data on wages and capital income, we estimate sector-specific production functions which allow us to quantify the economic impact of alternative mitigation policies and re-opening strategies.

The main takeaway of our paper is that smart mitigation and re-opening strategies must take into consideration the infection risk associated to occupations in different sectors of the economy and the network topology of a city. Employees at large retail establishments act as super-spreaders in our model, because of their high contact rate with many customers from distant locations in the network. Through shopping activities, individuals become infected and slowly spread the disease in their closer social cliques (home and neighbors). Because the production structure of the manufacturing and construction sectors involves

stable teams working together (and little to no interaction with customers) re-opening these sectors poses the lowest risk of a second wave, while at the same time significantly increasing output. The reason being that these workers tend to be more productive, on average, than those in many high-contact sectors (such as the retail). We find that re-opening schools and camps in August poses significant risks of a spike in infections, unless contact among kids is significantly reduced. Finally, we show that interaction among social layers generates amplification effects on the spread of the disease. For example, re-opening schools and retail simultaneously, would bring us back to peak levels of infection in a very short period of time.

Section 2 summarizes the related literature, the multi-layered network model is presented in Section 3. Section 4 describes our calibration strategy, whereas the model fit is discussed in Section 5. Policy counter-factual are discussed in Section 7 and re-opening strategies presented in Section 8. Concluding remarks and extensions are presented in Section 9.

# 2 Literature Review

There is a new and growing literature studying the consequences of mitigation policies and re-opening strategies for COVID19 on different aspects of the economy. As in our paper, the basic trade-off under consideration is that controlling the spread of the disease through social distancing measures (e.g. 'flattening the curve' gains) comes at the expense of reduced economic activity, and hence private consumption and welfare. Along these lines, Alvarez et al. (2020) discuss the optimal duration of lockdown strategies chosen by a planer who would like to reduce fatalities while minimizing output costs. Full macroeconomic models incorporating important dimensions of heterogeneity and feedback effects between infections and demand and/or supply are analyzed in Eichenbaum et al. (2020) and Kaplan et al. (2020). Alon et al. (2020) and Glover et al. (2020) focus on the redistributive effects of lockdown policies. Berger et al. (2020) and Chari et al. (2020) extend the set of mitigation strategies to include testing and case-dependent quarantine policies. A commonality of these papers is the underlying assumption of homogeneous random mixing, standard SIR (or SIER) models. In other words, that every susceptible individual in the population is equally likely to become infected by everybody else in society.

Modern epidemiology research has moved away from this classic homogeneous and highly stylized framework to incorporate heterogeneities that depend on between- and within- city population characteristics, the structure of social interaction, and the mobility of agents. Data driven models of geographical location allow consideration of spatially heterogeneous

interventions (see Colizza et al. (2007) or Benzell et al. (2020)), whereas 'individual-based' models, by simulating micro level details of transmission and following agents through space and time, permit more targeted non-pharmaceutical interventions (NPI). The effectiveness of general lockdown policies can then be assessed against more targeted interventions, such as quarantine within a specific geographical area instead. Ferguson et al. (2020) and Klepac et al. (2020) apply these modeling strategies to predict the progression of COVID19. [1] Moro and Bisin (2020) extend a spatial SIR model to incorporate behavioral responses. Rampini (2020), Favero et al. (2020) and Acemoglu et al. (2020) focus instead on individual, rather than geographical, heterogeneity. By introducing an age structure, they can analyze the effectiveness of targeted age-dependent mitigation policies.

All these models feature varying degrees of contagion risk, but they all retain the assumption of random mixing (and just allow for different mixing rates). As Keeling and Eames (2005) point out, however, the number of contacts each individual has in a given period of time is significantly smaller than the population size. In such case, mixing is not completely random. The stability of human interactions, mobility, and contacts patterns are best represented in the form of networks (see Jackson (2010) and Vespignani (2009)). In such models, individuals are affected at a rate proportional to the number of neighbours (e.g. family, friends, co-workers). A main contribution of our paper is to consider a network model of interaction among economic agents to estimate the costs and benefits of targeted non-pharmaceutical interventions to COVID19.

Epidemic spreading in classical network models have been studied over lattices (see Ganesh et al. (2005)), Barabasi-Albert (BA) scale-free or small-world networks (see Kuperman and Abramson (2001)). Recent applications to COVID19 of such models include Zhao (2020) and Karaivanov (2020). In our paper, we consider instead a mode complex community structure, known in a the literature as a 'multilayered network' (see Pastor-Satorras et al. (2015) and Nowzari et al. (2016) for surveys of the literature). A key difference lies in the fact that an agent can belong to several networks, each with different attachment properties, frequency of interactions, and meeting stability. A worker, for example, belongs to a family (small stable network), goes to work most days of the week (larger stable network), and sometimes shops at a mall (infrequent, unstable interaction in a large network). Relative to recent papers in economics, this allows us to assess effectiveness of non-pharmaceutical

[1] Fang et al. (2020), Dave et al. (2020), Hsiang et al. (2020) take a different approach by empirically evaluating the effect of anti-contagion policies on the spread of the disease. These studies aim at estimating the effectiveness of policies already in place. Our environment is constructed to forecast and design these policies through synthetic models of human interaction instead.

interventions in the lines of Glass et al. (2006)—who focused on the spread of influenza.—

A key determinant of infection dynamics in network models is the frequency and intensity of interactions across agents, typically characterized by contact matrices that may vary by age and occupation. Cross-walking occupation to industry, and the ability to work from home, allows us to study sector-specific mitigation policies and re-opening strategies. Recent related studies include Baqaee et al. (2020) and Akbarpour et al. (2020). We contribute to this literature by incorporating a production structure with a feedback loop between infections and the economy through the behavior of shoppers and workers. This allows us to better quantify the economic consequences of government policy in the calibrated model.

# 3 A ECON-EPI network

We now proceed to describe the details of the ECON-EPI network, a model designed to capture human and economic interaction in a typical US metro area. We first present the network structure, i.e. the links that connect individuals in their different activities. We then proceed to specify the EPI component, i.e. how infections travel through the network, and finally we describe the ECON component, i.e. how interactions in the network produce output. In this part we will first specify a pre-pandemic steady state equilibrium, which is meant to describe the normal state of economic affairs before the pandemic. We then discuss how the appearance of a disease and the adoption of containment measures affects economic activity during the pandemic period.

## 3.1 A multilayered network

We construct a multilayered network where individuals of different characteristics (age, employment status, public transportation usage, etc.) interact with each other. The set-up is necessarily stylized, nevertheless it has enough richness to capture key aspects of the social distancing policies that have been implemented during the 2020 COVID-19 pandemic. Time is discrete and the network is generically represented by a $M \times M$ adjacency matrix $\mathbf{G}$. Each node represents an individual. Individuals are heterogeneous in several dimensions. First, in terms of age, there are Adults and Kids. Kids are a fraction $\nu_K$ of total population and go to schools. Second, in terms of employment characteristics, as Adults may work in different sectors or be out of the workforce. Individuals interact in households, neighborhoods, and via their use of public transportation. We now proceed to describe the various layers connecting individuals. These layers affect the probability of contracting and spreading the

disease of each individual in the network.

**Households and Neighbours:** Households can be single member (composed by an adult) or two-member, composed by an adult and a kid. Members of the same household are fully connected through intra-household links. These links form the first layer of our network, contained in the adjacency matrix $\mathbf{G}^H$. The left panel of Figure 1 shows an example of household links in a city with 12 households, where the circles represent adult members and the stars represent kids. Households are placed next to each other on a ring (as in Watts and Strogatz (1998)), and each household member is connected to $N$ neighboring households on the left and on the right. We further assume that each household member is connected to each member of the neighbor households. The neighborhood links form the second layer of our network and are recorded in the adjacency matrix $\mathbf{G}^N$. Household and neighbourhood links are 'short-stable links,' meaning that they are active at every point in time, and link individuals who are close to each other.

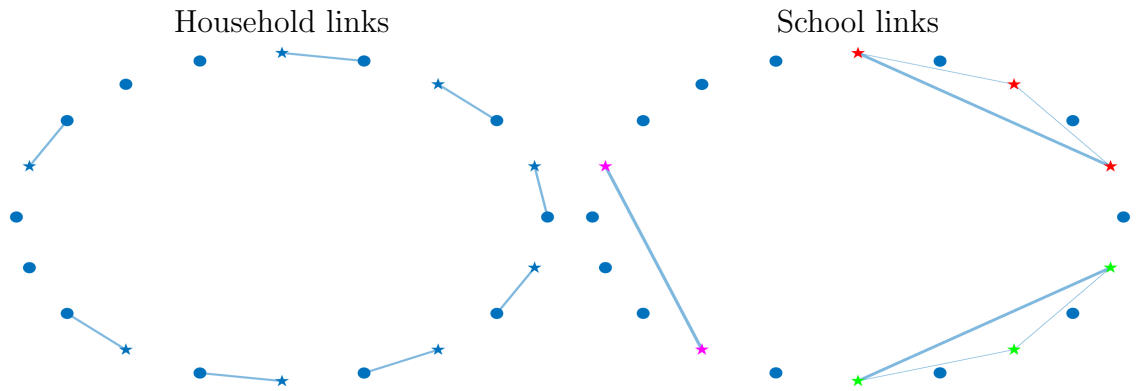


**Figure 1:** Households and Schools

**Schools:** Our third layer involves a school system where each day every kid interacts with a subset of other kids in her classroom. Each classroom draws kids which live close to each other, and the classroom size $Q$ determines the pool of potential interactions of each kid. We refer to these links as "potential" school links and their associated adjacency matrix is denoted by $\mathbf{G}^S$. The right panel of Figure 1 we show the potential links for three classrooms (two of size 3 and one of size 2) in the same 12 households network used in the left panel of this figure. Note that we refer to school links as "potential" because, differently from the links in the first two layers, they are not always active: every day each kid interacts only

with a subset of her classmates, which is randomly drawn every period. In our example, even though the size of one class is 3, each kid in the class may interact each day with just one other kid. We define school links as 'short-unstable' meaning that they link individuals who are geographically close to each other (short), and that they change over time (unstable). The main objective of introducing this layer is to later evaluate the effect of school closures on disease progression.

**Public Transportation** The next layer of the network specifies interactions through public transportation. A fraction $\phi$ of individuals uses public transportation. Each public transportation vehicle has a capacity of seating $P$ individuals, and we assume that agents living close to each other use the same public transportation vehicle. This implies that each individual using public transportation is potentially connected to locally close individuals who also use public transportation. Potential public transportation links are summarized in the adjacency matrix $\mathbf{G}^P$. During each public transportation trip, each individual interacts with a random subset of the vehicle occupants. Notice that individuals which use public transportation will be more exposed to the disease that those who use private means of transportation (see, Harris (2020)), and since use of public transportation is connected to work, the explicit modelling of public transportation is important to assess the epidemic impact of workers shutdown measures. Like school links, public transportation links are short and unstable. The difference between the two is that school links involve only kids, while public transportation connects both adults and kids.

**Workplace** A difference between adults in the network is that a fraction of them work and the rest do not. The workplace layer describes how working adults interact with each other and with the rest of society. The city features two distinct workplaces, which we label L (for Low contact) and H (for High contact). In the L-workplace (which is meant to capture sectors like manufacturing) there are stable teams of L-workers which are all connected to each other. In the H workplace (which is meant to capture sectors like retail or hospitality) there are similar teams of workers, but in addition the workers are also connected to a time varying subset of customers. We now describe in more detail the two work places.

**L-Workplace** L-Workers are a share $\nu_L$ of adults. There is an important heterogeneity across them, in that some of them (e.g. software developers) have the opportunity of working from home, which they will use in different intensity before and after the pandemic. The remaining members (e.g. assembly line workers) cannot work from home, and they are all

connected to each other when working.[2] The light green connected nodes in the left panel of Figure 2 depicts a team of L-workers.

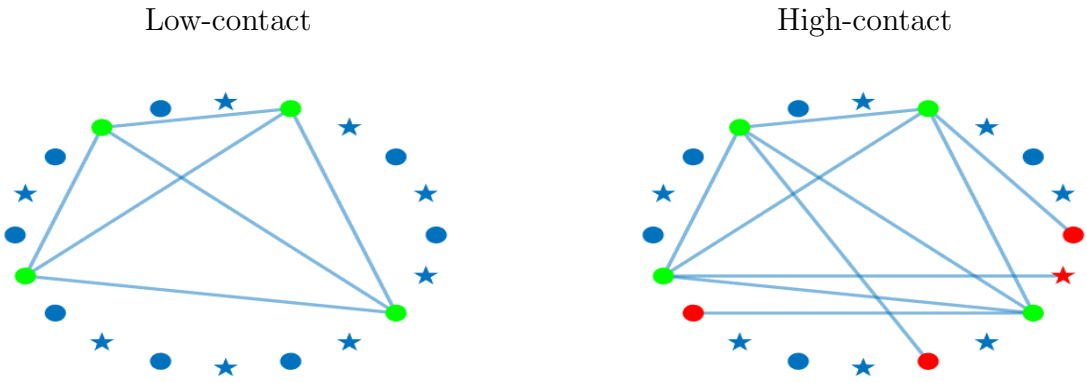


**Figure 2:** Worker links

**H-Workplace** H-Workers are a share $\nu_H$ of adults and they represent occupations that, for the purpose of production, involve stable contacts with co-workers (just like the L-workers) as well as unstable contacts with external customers (such as retail). As we assume that all persons shop, the number of links between customers and H-workers is given by the entire city population $M$ times $s$, which is the average number of shopping trips per person. The locking down of retail establishments has been at the centerpiece of the policy discussion during the 2020 COVID19 epidemics. Although it has been widely acknowledged that larger retail establishments lead to fast spreading of the disease, there has been much less emphasis on the fact that large retail establishments are, on average, more productive (see, for example, Foster et al. (2006)), and thus shutting down workers in those establishment might be more costly. In order to capture this trade-off we introduce heterogeneity in the workplace of the 'H-Workers'. We consider two types of establishments: small and large. Small establishments (mom and pop corner store) are less productive, employ less labor and have a local customer and employee base drawn from individuals in a geographically close area. Large establishments (e.g. large shopping malls, concert venues, and stadiums) are more productive, employ more labor and have a larger customer and employee base drawn from the entire network. To sum up, H-workers have some stable links that are always active (just like the L-workers), but they also have additional unstable links with customers. The right

[2]We will discuss the details of production in subsection 3.3.

panel of Figure 2 illustrates an establishment of H-workers. The light green nodes represent the workers, and the red nodes that are connected to each worker represent the customers on a given day. Over time the members of the team stay the same, but customers are drawn randomly each period. Note that customers are not connected to each other. This captures the fact that individuals from certain professions (doctors, bartenders, shop clerks) may see several clients during a day, sequentially, so their visits do not overlap. Note finally that due to the nature of their work, H-workers do not have the opportunity of working from home.

Worker links are, thus, intrinsically different from school links or public transportation links for two reasons. First, they include long links, connecting individuals in the network who are not necessarily in the same geographical area. Second, the number of connections of each individual worker can be different, with L-workers having only stable links, and H-workers having stable and unstable links. For convenience, we record worker links in three adjacency matrices: $\mathbf{G}^W$, which records all co-worker links (stable) in the H and L setor, $\mathbf{G}^{W_1}$ which records worker to shopper links in small establishments in the H-sector (unstable) and $\mathbf{G}^{W_2}$ which records worker to shopper links in large establishments in the H-sector (unstable).

**Network Clock** An important network feature, for the purpose of disease spreading, is the presence of unstable links. Connections in the household, in the neighborhood, and among co-workers have a stable nature and individuals are linked with the same set of people every period. On the other hand, school, public transportation, and workplace links between shoppers and workers are characterized by meeting with sets of people that are randomly drawn every period from a large set. To model this, we incorporate a *clock* in the spirit of Acemoglu et al. (2010) and Acemoglu et al. (2013). More specifically, for all $t \geq 1$, we associate a *clock* to every link of the form $(i,j)$ in the original adjacency matrix $\mathbf{G}^i$ (where $i = S, P, W_1, W_2$) to determine whether the link is activated or not at time $t$. The ticking of all clocks at any time is dictated by i.i.d. samples from a Bernoulli Distribution with fixed and common parameter $\varrho_i \in (0, 1]$, meaning that if the $(i,j)$-clock ticks at time $t$ (realization 1 in the Bernoulli draw), the connections between agents $i$ and $j$ is active at time $t$. This is meant to capture two kids in the same classroom having lunch together on a given day, two persons sitting next to each other in the subway, or a customer and a cashier interacting over a transaction. The Bernoulli draws are represented by the $M \times M$ matrix of zeros and ones $c_t^i$. Thus, the adjacency matrices for school, public transportation and worker-customers networks evolve stochastically across time according to

$$\begin{aligned} \mathbf{G}_t^i &= \mathbf{G}^i \circ c_t^i \\ \text{where } i &= S, P, W_1, W_2 \end{aligned}$$

**City Network:** We finally over-impose the layers described so far to construct a meta network which corresponds to our synthetic city, as illustrated in Figure 3. The adjacency matrix capturing all links within a city, $\mathbf{G_t}$, is constructed as a weighted sum of the different layers. The weights correspond to the relative importance of each layer, capturing that individuals spend different amounts of time interacting with others in these different social spheres. In particular, we have that

$$\mathbf{G_t} = \omega^H \mathbf{G}^H + \omega^N \mathbf{G}^N + \omega^W \mathbf{G}^W + \omega^S \mathbf{G}_t^S + \omega^P \mathbf{G}_t^P + \omega^{W_1} \mathbf{G}_t^{W_1} + \omega^{W_2} \mathbf{G}_t^{W_2}. \tag{1}$$

Each element in $\mathbf{G_t}$, denoted by $g_{i,j,t}$, summarizes the link between two individuals $i$ and $j$ at time $t$, weighted by the strength of their relationship. Figure 3 gives a graphical representation of all the layers of a city network.

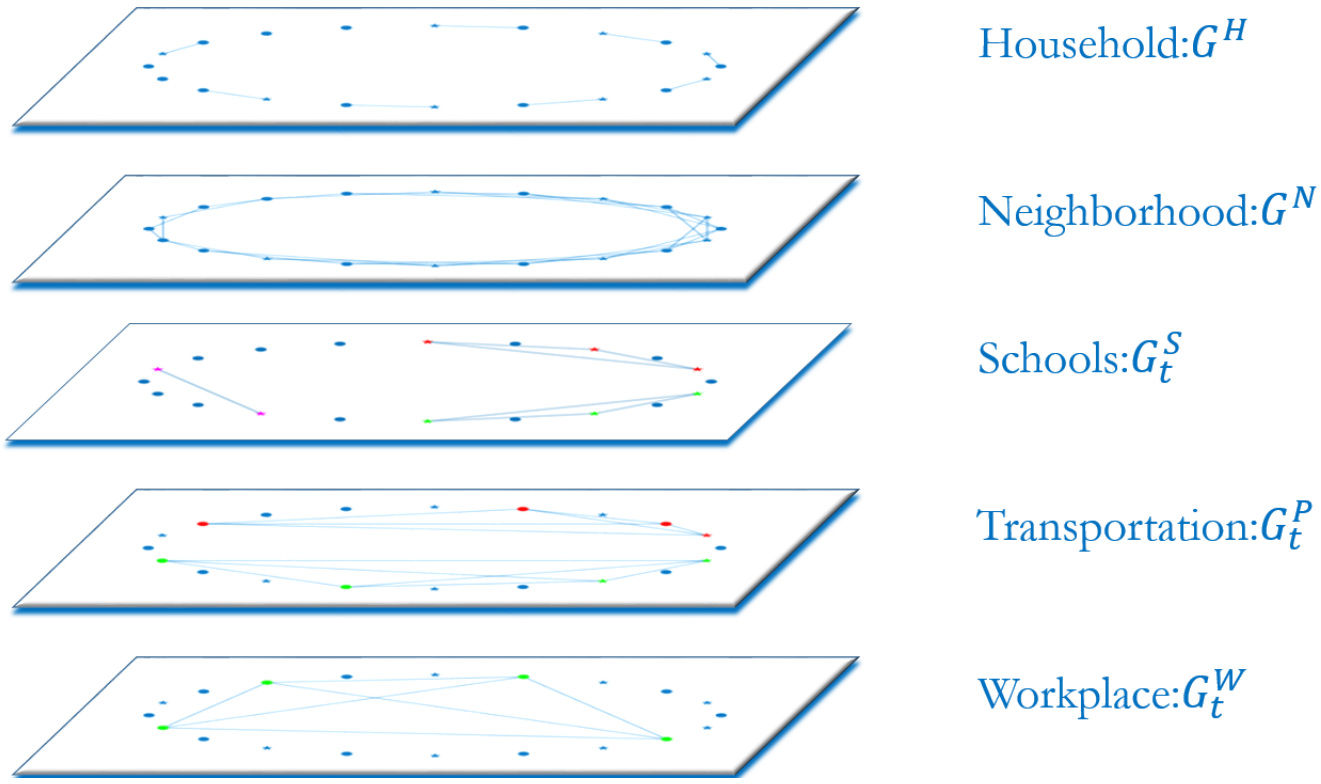


**Figure 3:** Network Layers

## 3.2 The EPI component

The spread of disease within our multilayered network is the result of two types of events: the person to person transmission of the disease (which depends on the network) and the progression of the disease for a given infected person, which is independent from the network structure. Our modelling of the disease progression follows closely the standard SEIR structure in the epidemiological literature, with the exception that we added the possibility of an "asymptomatic" branch. This assumption is motivated by the fact that during the COVID-19 epidemic many infection cases went undetected, either because symptoms were mild, or because testing was not available. These cases were never officially recorded as infected, and transited directly to the recovered stage. However, according to several studies (see, for example, Li et al. 2020) they significantly contributed to the spread of the disease.

Each individual node can be, at each point in time, in one of six health states: **S**usceptible, **E**xposed, **I**nfected-**A**symptomatic, **I**nfected-**P**re-symptomatic, **I**nfected-**S**ymptomatic, and **R**ecovered.

(1) Susceptible (S): a node which has not been exposed to the disease, but may contract it in the future.

(2) Exposed (E): a node which has been in contact with an infected node and has contracted the disease. Exposed nodes are not infectious and continue to perform normal activities. However they will transit with certainty to one of the infectious states the day following the exposure.

(3) Infected Pre-symptomatic (IP): a node which is infected and will show symptoms in the future. Nodes at this stage do not know they are infected, so they continue to perform normal activities. They transmit the disease with probability $\pi$.

(4) Infected Symptomatic (IS): a node which is infected and shows symptoms. IS nodes are removed from all layers of the network of the network, with the exception of the household layer. They transmit the disease with probability $\pi$.

(5) Infected Asymptomatic (IA): a node which is infected, but does not and will not show severe symptoms. These nodes do not know they are infected, so they continue to perform normal activities. IA nodes, when in contact with an S node, transmit the disease with probability $\eta\pi <= \pi$.

(6) Recovered (R): a node which is no longer infected. Recovered nodes are immune to the disease and can resume normal activities.

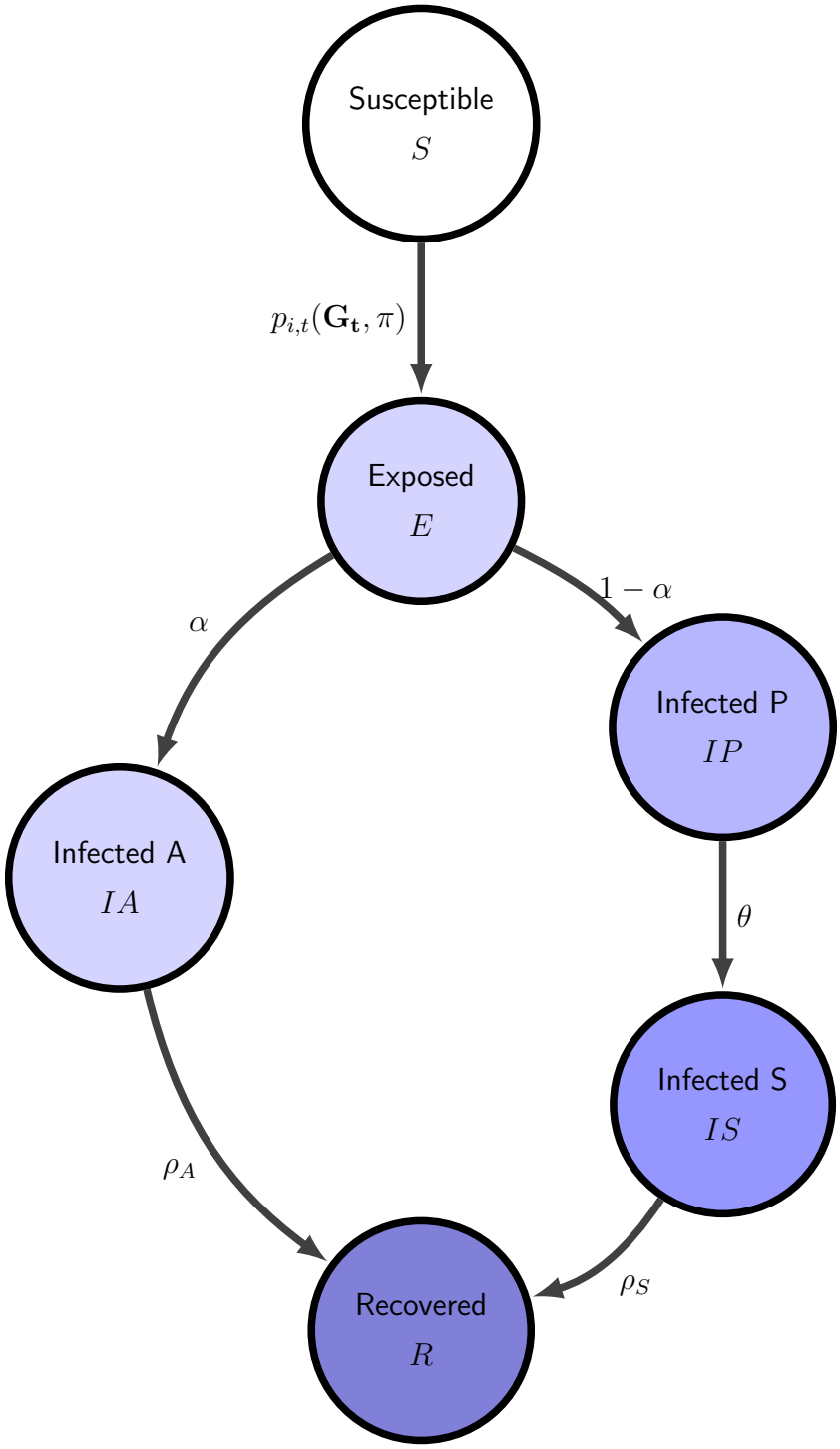


**Figure 4:** Transition between health states

Note that all nodes in an infected state can transmit the disease to susceptible nodes, although the infected asymptomatic are less likely to transmit. The transition between states is illustrated in Figure 4. A susceptible node $i$ contracts the disease at time $t$ with probability $p_{i,t}$ and if it does so it moves to the exposed state. An exposed node transits to the asymptomatic stage with probability $\alpha$ and to the pre-symptomatic stage with probability $1-\alpha$. A pre-symptomatic node moves to the symptomatic stage in each period with probability $\theta$ and symptomatic node moves to the recovered stage with probability $\rho_S$. An asymptomatic node, on the other hand, has probability $\rho_A$ in each period of moving directly to the recovered stage. Finally, recovery is an absorbing state. The key object of our analysis is $p_{i,t}$, the probability that a susceptible node $i$ contracts the disease in period $t$. The probability $p_{i,t}$ is a function of the odds of contracting the disease conditional on meeting an infected node (governed by the parameters $\pi$ and $\eta$), of the active contacts of node $i$ at time $t$ (encoded in $\mathbf{G_t}$) and of their health status. In particular we can write

$$p_{i,t}(\mathbf{G_t}, \pi) = 1 - \prod_{j=1}^{M} (1 - \pi\eta^{A(j,t)})^{g_{i,j,t} I(j,t)}. \tag{2}$$

where $g_{i,j,t}$ is the $ith, jth$ element of $\mathbf{G_t}$, the indicator function $I(j,t)$ equals 1 when node $j$ is infected (either pre-symptomatic, a-symptomatic or symptomatic) at time $t$, and zero otherwise, and $A(j,t)$ is a similar indicator function for the infected-asymptomatic status. This equation makes it clear that the spreading of the disease in the economy depends not only on the disease prevalence (captured by $I(j,t)$ and $A(j,t)$) and on the biological transmissibility (captured by $\pi$ and $\eta$), but also on the network structure summarized by $g_{i,j,t}$.

## 3.3 The ECON Component

Individual nodes, together with the network structure, produce, at each point in time new infections and economic output. This section describes how output gets produced over the network and how it is affected by social distance policies and by behavioral changes that result from the progression of the infection. The two workplaces described in Section 3.1 map into two sectors where output is produced. Both sectors produce the same homogenous good (which is also the numeraire) and production is organized in establishments. Low contact workers work in the L-sector where there are $Q_L$ ex-ante identical establishments, each endowed with the same amount of fixed capital $K_L$. High contact workers work in the H-Sector, which also has a fixed number $Q_H$ of establishments. However, in this sector,

we allow capital to be potentially different across establishments. As we think of these two sectors as having a substantially different occupational mix, we assume that workers cannot move across sectors.[3] In the L-sector, production requires L-workers and capital, while in the H-sector, production requires H-workers, capital, and customers. We first describe a pre-pandemic steady state equilibrium, where there are no infected nodes and the level of economic activity is stable over time, and then move on to describe how economic activity evolves as the disease hit the city and containment measures are adopted.

### 3.3.1 Pre-pandemic equilibrium

**L-Sector** Recall that establishment in this sector are homogenous. Each establishment produces $y_L$ units of output according to

$$y_L = K_L^{\alpha} n_L^{1-\alpha}$$

where $n_L$ denotes units of labor. Notice that $n_L$ is labor input which is not necessarily the same as employment, as not all L-workers supply the same amount of labor input. In particular, consistently with recent empirical work by Dingel and Neiman (2020) and Leibovici et al. (2020), we assume that a fraction $\omega$ of L-workers can work from home, and the labor input (or productivity) of these workers is $\delta_\omega\%$ higher than the labor input of those that cannot work from home. Given the wage rate per unit of L-work $w_L$, the establishment manager chooses labor input to maximize profits, which are given by $y_L - n_L w_L$. This implies that per establishment labor demand is given by

$$n_L = K_L \left( \frac{1-\alpha}{w_L} \right)^{\frac{1}{\alpha}}. \tag{3}$$

Labor supply of the L-workers is inelastic and is simply given by the total numbers of L-workers times their average labor input. A pre-pandemic equilibrium is then a wage rate $w_L$ and quantity of L-labor per plant $n_L$ such that, i) given wages, $n_L$ is chosen optimally by the plant manager and ii) labor market for L-workers clear. Equation 4 summarizes these two conditions

$$\underbrace{Q_L n_L = Q_L K_L \left( \frac{1-\alpha}{w_L} \right)^{\frac{1}{\alpha}}}_{\text{Labor Demand}} = \underbrace{\Big[\omega(1+\delta_\omega) + (1-\omega)\Big] \nu_L (1-\nu_K) M}_{\text{Labor Supply}}, \tag{4}$$

[3]The details of the mapping of these two model sectors to sectors in the data is discussed in Section 4.

where $\nu_L$ denotes the share of adults which work in the L and $(1-\nu_K)$ the share of adults in the population, implying that $\nu_L(1-\nu_K)$ is just the total number of individuals who work in the sector.

**H-Sector** Establishments in this sector are of two different types, thus we index them by $i$. Each establishment of type $i$ in the H-sector produces $y_{Hi}$ units of output according to

$$y_{Hi} = K_{Hi}^{\alpha}\left(\min\left\{\frac{K_{Hi}d}{\mu}, n_{Hi}\right\}\right)^{1-\alpha},$$

where $K_{Hi}$ denotes establishment capital, $n_{Hi}$ denotes the number of workers (which in this sector are homogenous) employed, $\mu$ is the number of customers that a H-worker can attend to and $d$ represents the number of customers (per unit of capital) which shows up at establishment $i$. This assumption captures that in the H-sector, if a customer does not go to the establishment, a sale does not materialize and output is reduced. In addition, if there are too few workers, they may not be able to serve all the customers that come to the establishment. The establishment manager takes as given the wage rate $w_H$ and demand $d$ and hires workers to maximize profits, which are given by $y_{Hi} - n_{Hi}w_H$. This implies that labor demand in establishment $i$ is given by

$$n_{Hi} = K_{Hi}\min\left\{\left(\frac{1-\alpha}{w_H}\right)^{\frac{1}{\alpha}}, \frac{d}{\mu}\right\}. \tag{5}$$

Similarly to the L-sector, the labor supply of the H-workers is inelastic and is given by the total numbers of H-workers, which is equal to $M(1-\nu_K)\nu_H$. The last element that is needed to define a pre-pandemic equilibrium is the determination of $d$. Recall that in our model city there are $M$ persons, and each person makes $s$ shopping trip every period. It follows that the total number of customers of the H-sector is $Ms$. The customer capacity of the H-sector is instead given by the sum of all workers employed in that sector, times the number of customers a worker can attend, $\mu$. Since in equilibrium the sum of all workers employed in the H-sector is the labor supply in the sector, equilibrium customer capacity is given by $\mu M(1-\nu_K)\nu_H$. We then assume that in a pre-pandemic equilibrium, the number of shopping trips per person is such that total shopping trips equals customer capacity of the H-sector, that is $s = \mu(1-\nu_K)\nu_H$.

To sum-up, a pre-pandemic equilibrium in sector H is a wage rate $w_H$, a quantity of H-labor per establishment $n_{Hi}$ and an amount of customers per capital $d$, such that, i) given wages and customers, $n_{Hi}$ is chosen optimally by the establishment manager, ii) labor market for H-workers clear and iii) the total number of shopping trips equals the customer capacity

of the H sector.[4] Note that our concept of equilibrium guarantees that in every pre-pandemic period every shopper in each of her/his shopping trip is assigned to an H-worker that can attend to her/him. We conclude the description of the sector by noting that we consider two types of H-establishments: a number $Q_{H2}$ with high capital $K_{H2}$ and a number $Q_{H1}$ with low capital $K_{H1}$. It follows directly from the maximization at the establishment level, that the high capital establishment will employ more labor and make more sales. Moreover, as we discussed in Section 4.4, the potential customers of the large H-establishments are randomly drawn across the entire network, while small H-establishments draw their customers locally. Large H-establishments are more productive, but more likely to spread the disease.

### 3.3.2 Production during the pandemic

The model specified so far implies that, in absence of a pandemic, aggregate economic activity in both sectors is constant. During the pandemic, however, output can change over time. As discussed in Section 3.2 nodes that are infected and show symptoms are prevented from working and shopping. Moreover, as the disease spreads, policies are introduced that may prevent also a fraction of healthy workers from working. We denote by $n_{Lit}$ the number of L-workers that show up at work in establishment $i$ in period $t$, $n_{Hit}$ the same number for H-workers, and finally by $d_{it}$ the number of customers (per unit of capital) that will show up to shop at H-establishment $i$ in period $t$. By assumption, in the short run establishments can not replace workers, therefore when the number of workers falls, establishment output will also fall. Moreover, when a customer assigned to an H-establishment is sick and does not show up to shop, the output of that establishment also is reduced. We can now define $Y_t$, i.e. the total production of the city in period $t$ as

$$Y_t = \underbrace{\sum_{i=1}^{Q_{H1}} \left[ K_{H1}^{\alpha} \left( \min\{ \frac{d_{it} K_{H1}}{\mu}, n_{Hit} \} \right)^{1-\alpha} \right]}_{\text{Output of small H establishments}} + \underbrace{\sum_{i=Q_{H1}+1}^{Q_{H1}+Q_{H2}} \left[ K_{H2}^{\alpha} \left( \min\{ \frac{d_{it} K_{H2}}{\mu}, n_{Hit} \} \right)^{1-\alpha} \right]}_{\text{Output of large H establishments}}$$

$$+ \underbrace{\sum_{i=1}^{Q_L} K_L^{\alpha} n_{Lit}^{1-\alpha}}_{\text{Output of L establishments}}$$

[4] For simplicity, we do not develop an explicit theory of the individual choice of shopping trips. A possible way of doing so, that would be consistent with our equilibrium restriction, would be to have the individual benefit of shopping trips to be decreasing in the tightness of the shopping market, i.e. in the ratio between shoppers and customer capacity

# 4 Calibration

In this section, we describe how we set the parameters that characterize the city's model and the transmission of the disease.

## 4.1 Demographics, and Public Transportation

We calibrate our model to a 5% synthetic version the New York-Newark-Jersey City (NY-NJ-PA) metro area, which in 2019 has a population of approximately 20 million. The percentage of kids in the population $\nu_K$ is set to 28% so that the synthetic city has 40% of households with kids, which matches the percentage of households with kids in NY-NJ-PA metro from the 2014-2018 American Community Survey. The percentage of non working adults $\nu_N$ is set to 37%, to match the employment to population ratio for persons over 18 in the NY-NJ-PA metro in 2019.[5] The share of agents using public transportation, $\phi$, is set to 0.32 in order to match the percentage of individuals who report commuting to work using public transportation in the NY-NJ-PA metro from the 2014-2018 American Community Survey.

**Table 1: Demographics and Public Transportation**

| Parameter Name | Symbol | Value | Source |
|---|---|---|---|
| *Demographics* | | | |
| Total Population | $M$ | 1,000,000 | Census Data: ACS 2018 |
| Share of Kids | $\nu_K$ | 28% | American Community Survey |
| Share of Non-working Adults | $\nu_N$ | 37% | American Community Survey |
| *Public Transportation* | | | |
| Share using Public Transportation | $\phi$ | 32% | American Community Survey |

[5]Employment figures are from the BLS and population figures are from the Census.

## 4.2 Workplace

An important aspect of the calibration is to map workers in the data to workers in the two sectors in the model: the High and Low contact. In order to do so, we first work with occupations. Recall that there are two key features that characterize the high contact sector: one is the physical proximity with other people (so that infection can be transmitted) and the second is the instability of the contact with customers (which also speeds up the spread of the disease). To capture these two features in an occupation, we use two questions in the ONET database. The first one asks about physical proximity to other people on the job, while the second one asks about the importance of interactions with external customers. [6]

The answers to these questions can be used two construct two indexes, both ranging from 0 to 100, that give, for each 6-digit occupation, measures of physical proximity and external interactions. Next, using a standard cross walk we compute similar indexes for all the private sectors at the 2-Digits NAICS level, where the index for sector $i$ is the average of the index of each occupation $j$ in that sector, weighted by the national employment share of occupation $j$ in sector $i$. This procedure yields indexes of physical proximity and external interactions for all the 2-digits NAICS sector. Figure 5 shows these (standardized) indexes for all the NAICS 2 digits private sectors.

The shaded northwest quadrant highlights the 5 sectors which have both indexes above the mean; we thus construct the H-contact sector by aggregating them, and the low contact by aggregating all the others.[7] In Table 2, we report key characteristics of workers in the two sectors using employment figures from the Census Statistics of US Business (SUSB) for the NY-NJ-PA metro for 2016. Note that the L-sector employs more workers, and that workers in that sector have higher average wages. In the last two columns we compute the fraction of workers in each sector that work from home and a measure of their annual wages. Note that in the L sector there are many more workers that work from home and that the annual wage of the home L-workers is roughly 16% higher of the non home workers in that sector.[8]

[6]Specifically the first question (ONET question 21) is "How physically close to other people are you when you perform your current job?" and the second question (ONET question 8) is "In your current job, how important are interactions that require you to deal with external customers (as in retail sales) or the public in general (as in police work)?."

[7]We do not include public administration because the Census Statistics of US Business, which we use below, does not report wage and employment data for that sector. See the appendix for more details on the procedure

[8]To measure the share of workers that work from home we first use ACS data to compute the share of home workers in each 2-digits NAICS sector and then take a weighted average of these percentages, where the weights are the employment shares of each NAICS sector in our 2 macro sectors. Similarly to compute

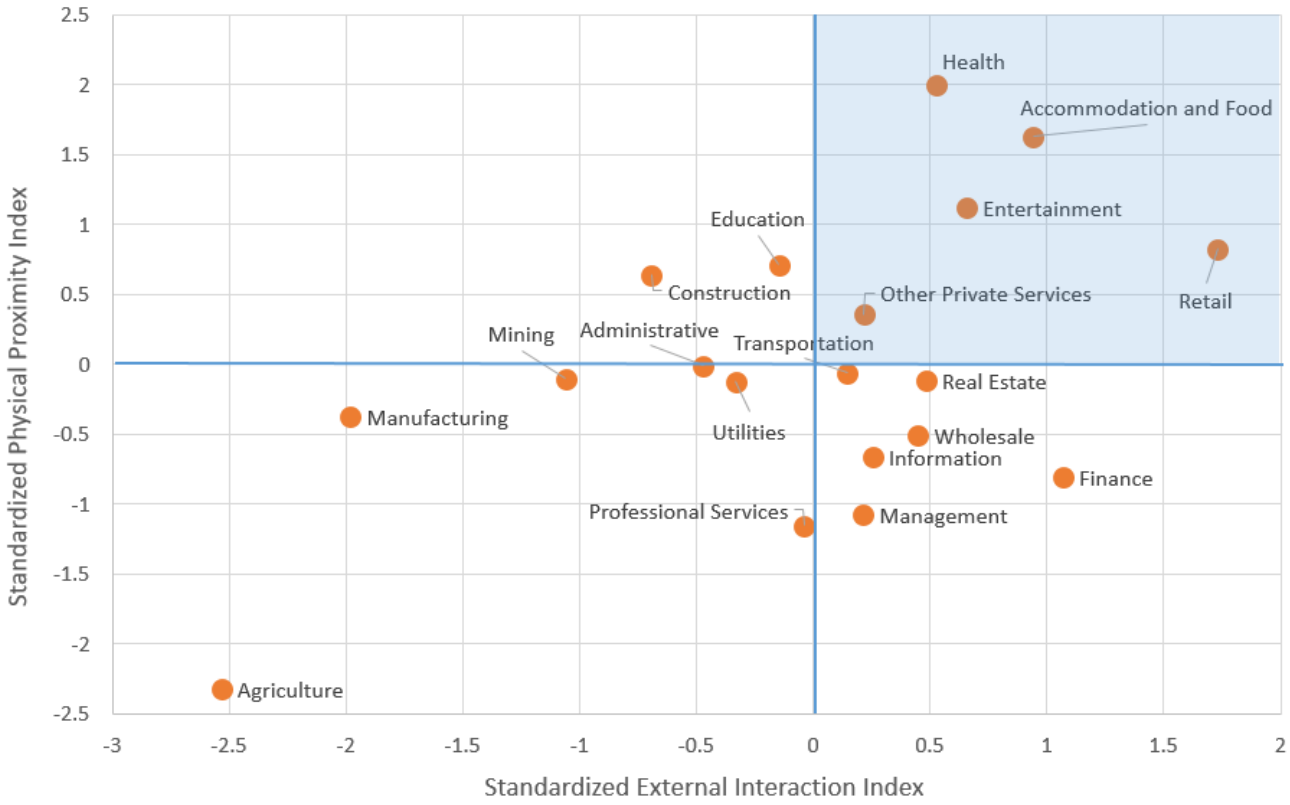


**Figure 5:** Identifying high contact sectors

**Table 2: Characteristics of workers in H and L Sectors**

| | Share | Avg. yearly payroll ($) | % Home workers | Home wage premium |
|---|---|---|---|---|
| L-sector | 54.1% | $94k$ | 6.9 | 16% |
| H-sector | 45.9% | $40k$ | 3.4 | 6% |

In the next section, we use these numbers to pin down the labor supply and the technological differences across these two sectors.

## 4.3 Labor and Technology

The general logic of this section is that restrictions from the pre-pandemic equilibrium (see Section 3.3.1), plus data from firms and workers as described above, pin down the parameters that characterize the labor supply and the technology in the L and H sectors. All the parameters are reported in Table 3 below.

wages of home workers we take a weighted average of the wages in each sector, where the weight are the shares of home workers coming from each NAICS sector.

We first use statistics on home workers and on their wages reported in Table 2 to pin down the parameters $\omega$ and $\delta_\omega$, which determine: (i) the fraction of workers that can work from home and (ii) the ratio of their wage relative to the wage of those who cannot work from home in the L-sector. In order to determine the fraction of workers who can work from home, we use two types of information. In Table 2, we report that in the L-sector 6.9% of workers are working from home: this is a lower bound for the fraction of workers that can actually start tele-commuting once the pandemic hits. Dingel and Neiman (2020) estimate the fraction of workers that can work from home in each occupation based on occupational characteristics. We compute their measure for each sector, and aggregating by sector, we find that the fraction of L-workers that can tele-commute is 49.7%. We view this number as an upper bound, as in the short run it is unlikely that such a large percentage of workers can switch to tele-commuting. For this reason, we set the fraction of workers who can actually work from home, $\omega$, to 28.3% (which is the mid-point between the lower and upper bound). We assume that home workers supply 1 unit of labor input and home workers supply $1+\delta_\omega = 1.16$ units, in order to match the wage differences between the two groups in the L-sector. We then use demographic statistics from Table 1, plus worker statistics from Table 2, to pin down the parameters $\nu_L$ and $\nu_H$, which denote the share of adults working in the L and H sectors, respectively. All these parameters determine the total labor supply (in units of labor input) in both sectors.[9]

Both sectors share constant returns to scale production functions, where capital share is common and given by $\alpha$. We estimate $\alpha$ using the standard methodology outlined in Cooley and Prescott (Cooley and Prescott), using 2018 data for the New York Metro Area.[10]

Given $\alpha$ we can normalize the wage of a unit of labor (which is equivalent to the wage of a non-home worker) in the L-sector to 1 and use the establishment labor demand (Equation 3) to pin down the labor demand per unit of capital. We then use the labor market clearing (Equation 4) to pin down the total capital in the sector $Q_L K_L$. Note that in the pre-pandemic equilibrium the number of establishments $Q_L$ is not determined separately from the capital per establishment $K_L$, so we simply pick $Q_L$ so that the number of workers per establishment

[9]Recall that, since few workers in the L-sector work from home, we assumed none of the workers in that sector can work from home

[10]The estimate is the ratio between capital income (consumption of fixed capital plus rent, interest and dividend income) and the sum of capital income plus labor income (compensation of employees). Data for rent, interest and dividend income, and for compensation of employees is available from BEA regional tables. Consumption of fixed capital is computed by first taking the ratio of consumption of fixed capital to GDP on national data for 2018 (the ratio is 16%) and then multiplying it by the metro area GDP.

**Table 3: Labor and Technology Parameters**

| Parameter Name | Symbol | Value |
|---|---|---|
| Share of Capital Income | $\alpha$ | 33% |
| *L-Sector* | | |
| Share of adults working in L | $\nu_L$ | 34% |
| Share of workers that can work from home | $\omega$ | 28% |
| Home premium | $\delta_\omega$ | 16% |
| Number of establishments (14 wkrs) | $Q_L$ | 18,400 |
| Capital per unit of labor | $\frac{K_L}{n_L}$ | 3.4 |
| *H-sector* | | |
| Share of adults working in H | $\nu_H$ | 29% |
| Number of small establishments (4 wkrs) | $Q_{H1}$ | 12,000 |
| Number of large establishments (50 wkrs) | $Q_{H2}$ | 3,500 |
| Number of customers per H-worker | $\mu$ | 10 |
| Capital per unit of labor | $\frac{K_H}{n_H}$ | 0.26 |

in the model is 14, which matches the number of workers per establishment in the NAICS sectors that compose our L-sector.[11]

Now moving to the H-sector, we use, as we did in the L-sector, the establishment labor demand (Equation 5) to pin down the labor demand per unit of capital. We then use the labor market clearing to pin down the total capital in the sector $Q_{H1}K_{H1} + Q_{H2}K_{H2}$. We pick $Q_{H1}$ and $Q_{H2}$ to match features of the establishment size distribution in the NY-NJ-PA metro in the NAICS sectors that comprise our H-sector. In particular, we choose the size of the small establishments to match the average establishment size of the firms in the H-

[11]In the model, the number of workers per establishment is smaller than the quantity of effective labor, as the average worker, due to higher productivity of home workers supplies more than one unit of effective labor.

sector that have less than 20 employees. This gives a number of 4 employees for the small establishment and 50 employees for the large establishments. This choice, together with equilibrium restrictions, implies that in the model 22% of H-workers are in 4-employee shops (so that we match the employment share of small establishments). We denote this share as $\nu_{H1}$. The values of $Q_{H1}$ and $Q_{H2}$ are reported in Table 3.

The remaining parameter to be determined in the H sector is $\mu$, that is the number of customers that a worker can attend to in a day. Recall that, in a pre-pandemic equilibrium, in the H-sector the number of customers is equal to the total customer capacity. In the next section, we calibrate the equilibrium shopping trips ($s$) to be 2 per person, so that the total number of customers in a day is $2M$. This implies, given the share of H-workers in the population, that the parameter $\mu$ is approximately 10; that is, an H-sector worker attends to an average of 10 customers per day. One final important statistic reported in Table 3 is the capital per unit of labor, which is higher in the L-sector (3.4) than in the H-sector (0.26). The magnitude of this gap is identified from the data on the wage differential (see Table 2) between workers in the two sectors. The reason why a unit of labor used in the production of the final good in the L-sector receives a higher compensation than a unit of labor used in the production of the final good in the H-sector, is that labor in the L-sector works with more capital.[12]

## 4.4 Network Contacts and Weights

The number of contacts each person has on each layer, and the weights of different layers play an important role in the spreading of the disease through the network. Our main reference for setting these in the model is the work by Mossong et al. (2008), which, using a common paper-diary methodology, has collected data on various characteristics of daily face to face interactions for a sample of over 7000 persons in 8 European countries.

The number of contacts of various individuals in different layers in the model and the targets from Mossong et al. (2008) are reported in the first two columns of Table 4.

Mossong et al. (2008) reports that on average each individual has 5.2 contacts in the household and during leisure activities. We map these contacts with model's contacts that take place within the household and neighborhood layers. Since the average household size

[12]In our set-up, we have abstracted from differences in human capital among the workers in the two sectors, and attributed all the differences in wages to differences in physical capital. Since in the short run physical capital is fixed, the results concerning output losses from shutting down workers in the two sectors are independent on whether we attribute wage differences to physical or human capital.

in the model is 1.6, we impose that each household has 3 neighbors, so that each individual has on average $1.6 + 3 \cdot 1.6 - 1 = 5.2$ household/neighbour contacts. Mossong et al. (2008) also reports that, on average, each individual experiences 2 contacts during shopping and 0.4 while travelling. We set the number of shopping trips per person and the number of meetings while using public transportation in the model to match these two figures. Moving now to the differences between kids and adults, Mossong et al. (2008) reports that kids between the ages of 0 and 19 have on average 15.3 contacts, and adults have on average 12.4 contacts. In the model, we set the number of school contacts (which are specific to kids) to match total kids contacts. For adults, the number of contacts is more heterogenous. A fraction of adults (the non-workers) has no contacts resulting from work. Another fraction (the L-workers) has contacts resulting from meeting their team (of size $T_L$) of co-workers, where the team of workers is meant to capture the set of co-workers with which a worker interacts more closely. Finally, the H-workers have contacts resulting from the team of co-workers (of sizes $T_{H1}$ and $T_{H2}$) and from meeting with customers ($\mu$). Since we do not have much hard evidence on the size of workers teams, we simply set the size of the team of H-workers in the small establishments $T_{H1}$ to 2 (mom and pop stores) and set $T_{H2} = T_L = 5$ so to match the total number of adult contacts. [13]

Mossong et al. (2008) also reports information on the average duration of contacts, by contact type (daily, weekly and first time). We identify contacts in the household, work and school layers as daily, with an associated average duration of 3 hours. We identify shopping and neighborhood with weekly contacts, with an associated average duration of 1.4 hours and finally we think of contacts during travel as first time contacts, with an average duration of 0.5 hours. These figures results in weights of each layer (normalized to sum to 1) which are reported in the third column of Table 4. These weights are then used to identify the parameters used in equation 1 that capture the weight of each layer $\omega^H = \omega^W = \omega^S = 22\%$, $\omega^P = 6\%$ and $\omega^{W_1} = \omega^{W_2} = \omega^N = 14\%$.

The final column of Table 4 reports the potential pool of contacts for those layers where the actual contacts are drawn randomly every day. This information is not available in Mossong et al. (2008), however it is an important determinant of the spread of infection, and therefore we pin it down using the network structure, as well as additional information. For the shopping links, every person does (on average) $2v_{H1}$ and $2(1 - v_{H1})$ shopping trips to the small and large establishments, respectively. When shopping at the local mom and pop

[13] A team size of 5 will result in 4 co-worker contacts for the H-worker, and only 3.7 contacts for the L-worker because a small fraction of the workers work from home

**Table 4: Network Contacts and Weights**

| Person type | Layer | Actual Contacts | | Weight | Contact Pool |
|---|---|---|---|---|---|
| | | Model | Mossong | | |
| *All:* | Home and Neighbor | 5.2 | 5.2 | [ 22%, 10%] | - |
| | Shopping overall | 2 | 2 | 10% | |
| | small | $2v_{H1}$ | | | 4 |
| | large | $2(1-v_{H1})$ | | | $\tilde{v}v_H M$ |
| | Public Transport | 0.3 | 0.4 | 4% | 54 |
| *Kids:* | Total | 15.6 | 15.3 | | |
| | School | 6 | | 22% | 26 |
| *Adults:* | Total | 12.3 | 12.4 | | |
| | Work | | | | |
| | H-small (co-workers) | 1 | | 22% | - |
| | H-small (customers) | 10 | | 22% | 56 |
| | H-large (co-workers) | 4 | | 22% | - |
| | H large (customers) | 10 | | 22% | $M$ |
| | L (co-workers) | 3.7 | | 22% | - |

store, the pool of potential sales people that a shopper meets is given by 4 (the employment size of the store). When shopping at large establishments, the pool of potential contacts is given by $\tilde{v}M$, with $\tilde{v} = (1-v_K)(1-v_{H1})$, which is the total number of workers that work in large establishments in the H-sector. Note that pool of contacts (i.e. potential sales people) when shopping at large establishments is much larger than the contacts when shopping at the small stores. The reason is the assumption of a different customer base: when shopping at a small establishment, a person always visits the same local store whereas when shopping at a large establishment, the individual is randomly assigned to an establishment in the city. For adults working in local small establishments in the H-sector, the pool of potential customers is given by the local customer base which is equal to the size of the population divided by the number of workers in the small establishments, $\frac{M}{M(1-v_K)v_H v_{H1}} \simeq 56$. Workers in large establishments in the H sector draw their potential customers from the whole city, so their pool of contacts is the city population $M$.

For public transportation, we choose the number of potential contacts equal to 54 to match the seating capacity of the R160 New York City subway car. Finally, for schools, we proxy the pool of potential contacts with the class, so we set the size of the pool to 26 to

match average class size (across grades) in New York City public schools for 2018-19. The ratio between the actual contacts and the contact pool for the unstable layers (shopping, public transportation, school and H-work place) is then used to set the Bernoulli parameter $\rho_i$ in the network clocks described in equation 1.

## 4.5 Disease Transmission

The final parameters to be determined are those regulating the diffusion of the disease, described in Sub-Section 3.2. We set some parameters based on epidemiological studies on COVID 19, and set the remaining, for which there is less evidence, to match the early stages of infection diffusion in the New York metro. Parameters are reported in Table 5.

Starting on the symptomatic branch, we set $\theta$ to 0.25 and $\rho_S$ to 0.071, in order to match a duration of the pre-symptomatic and symptomatic stages of the disease to 4 and 14 days, respectively (see, among others, Guan et al. 2020). Going now to the asymptomatic branch, we follow Li et al. (2020) and set $\eta$ to 0.5, capturing the finding that asymptomatic are half as infectious as the patients showing symptoms. Also, following Li et al. (2020), we set $\rho_A$ to match a duration of the asymptomatic stage to 4 days. The three remaining parameters are $\pi$, the infectiousness of the symptomatic cases, $\alpha$, the fraction of exposed that transit to the asymptomatic stage, and $r_{as}$, the initial ratio of asymptomatic to symptomatic. These are calibrated, as described in the next section.

**Table 5: Disease Transmission Parameters**

| Parameter Name | Symbol | Value | Target |
|---|---|---|---|
| Infection Probability | $\pi$ | 0.52 | Calibrated (see text) |
| Relative Infectiousness of $IA$ | $\mu$ | 0.50 | Li et al. (2020) |
| Prob. of transition from $E$ to $IA$ | $\alpha$ | 0.6 | Calibrated (see text) |
| Initial Ratio of Asy. to Sym. | $r_{as}$ | 1.7 | Calibrated (see text) |
| Prob. of transition from $IP$ to $IS$ | $\theta$ | 0.25 | Incubation 4 days |
| Prob. of transition from $IS$ to $R$ | $\rho_S$ | 0.071 | Duration of disease 14 days |
| Prob. of transition from $IA$ to $R$ | $\rho_A$ | 0.25 | Duration of asymptomatic stage of 4 days |

# 5 Model's validation

In this section we first describe how we calibrate the epidemiological parameters $\pi$, $\alpha$, and $r_{as}$. We then assess the ability of the network model to replicate the evolution of infection in the New York Metro area, and contrasts it with another popular model of infection spreading, i.e. the standard random mixing SIR model. We focus on the period from March 8th, 2020, where the first 160 cases where reported in the New York metro area, until May 25, 2020. We start our model city with the same number of infected symptomatic per million in the New York MSA on March 8th. The progress of the infection in the model does not only depend on initial conditions and epidemiological parameters, but also on the network structure which, as the pandemic spreads, evolves. In order to capture this evolution we use both information on actual regulatory changes and data on mobility, as reported by Google.[14] In particular Google reports three mobility series that track the visits and length of stay of individuals at workplaces, retail and residences. These series have a natural mapping into our model: workplace mobility maps into presence of L-workers at their establishments, retail mobility maps into presence of workers and shoppers at H-establishments and finally residential mobility captures the time individuals spend at home. These three measures for New York City are reported in the top panel of Figure 6. The panel shows that initially workplace/retail mobility sharply falls, then it stays constant at a depressed level and partially recovers towards the end of the period. Residential mobility displays the opposite pattern. This evolution is most likely the result of both changes in policy and in behavior. Our strategy is to match this evolution by furloughing a time varying fraction of both L and H workers. In particular in each period, we match the observed decline in workplace mobility in two ways. In the first days of the pandemic we match the decline in workplace mobility by having all L-workers that can work from home starting to do so. As time progresses and workplace mobility continues to decline we match the further decline by furloughing a fraction of L-workers each day. Third we furlough a fraction of H-workers each day so to match the decline in retail mobility. As far as the relative decline for large retail and small retail we impose a larger percentage decline of the employment in large establishments to be consistent with the fact that authorities in New York shut down events with more than 500 attendees by March 12th. When a worker is furloughed, their time is reallocated to their household and neighborhood networks. A fraction of the work hours are assigned to the household and neighborhood layer so that the increase in home hours matches the increase in residential mobility. Note that when we furlough H-workers we also

[14] See appendix B for a timeline of the pandemic related policies in New York

cut a number of shopping links, as shoppers assigned to furloughed workers are not able to shop. We also close schools in the model on March 14th, which is the date in which K-12 schools are shut-down statewide.[15] The mobility patterns suggest a division of our period in three subperiods. The first (labeled "closing", from March 8th to April 3rd) is the period in which mobility sharply declines, the second (labeled "lock-down", from April 3rd to April 26th) is the period in which mobility stays low, and the last (labeled "re-opening", from April 26th to May 25th) when mobility picks up. Our calibration strategy is to pick the epidemiological parameters $\pi$, $\alpha$, and $r_{as}$ so that the infection curve in the model exactly matches the data in the first subperiod, and then assess the ability of the model of replicating the infection curve in the two subsequent periods.

The bottom panel of Figure 6 shows the cumulative infection curve generated by the network model against the data. The periods after April 3rd constitute a validation of our model. The network model is close to the empirical epidemiological curve all throughout the lock-down phase, and shows more growth in infection (relative to the data) as the city starts to re-open. For comparison purposes, we also report the infection curve predicted by a standard SIR model, where each individual has the same number of contacts as in the network, but the contacts are randomly drawn across the entire population. We calibrate the epidemiological parameters in the SIR to match the data infection curve in the first sub-period (exactly as we did for the network model), and we change the number of random contacts in the SIR so to match the average change in Google mobility. Possibly the most important message of Figure 6 is that even when the two models (Network and SIR) are put on equal footing, as they generate the same initial surge of infection and have similar containment measures, they have sharply different predictions for the evolution of the pandemic. In the network model, the infection naturally slows down, as reduction of contacts is enough to keep infection local and prevent it from reaching the entire population. The SIR model, however, predicts that despite the reduction in contacts, the infection takes off in an exponential fashion. This is due to the random nature of contacts: in the SIR model, an individual is equally likely to meet any other individual in the city, whereas in the network model contacts are more clustered and less random. Before we move on to policy experiments we use our calibrated model to quantify the contribution of several layers to the infection.

[15]Schools were announced to be closed on March 16th, a Monday, so we shut down schools effectively on Friday 14th.

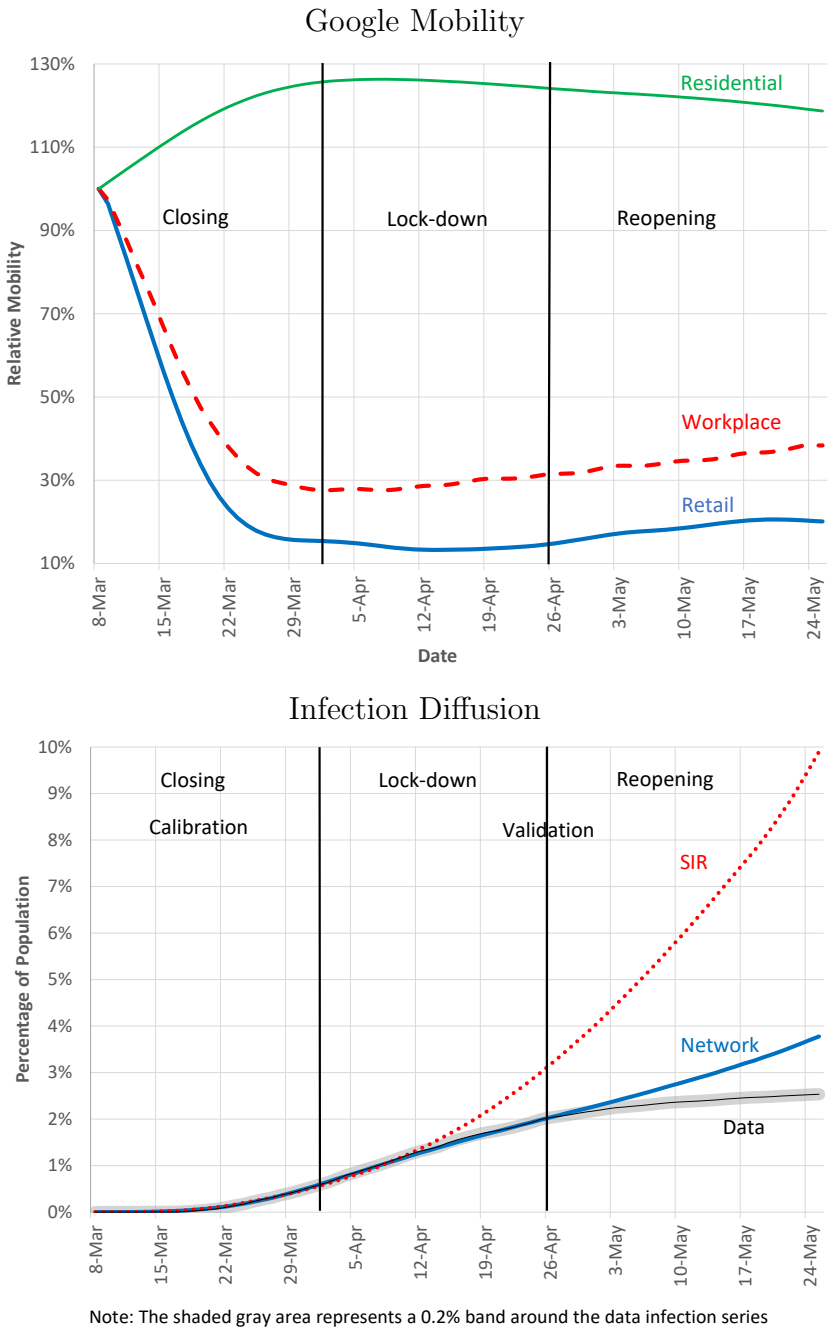


**Figure 6:** Network v/s SIR

# 6 Infection Decomposition and complementarities

In this section we study the effect of shutting different layers of the network, and how this shutdown interacts with the transmissibility of the disease. In order to do so we sequentially set to 0 the weights of separate layers of the network, and assess the impact of shutting

down one single layer on the evolution of the infection. An important issue in assessing the impact of a given layer is the presence of mitigation policies (for example school closures) or endogenous reduction of contacts (as captured by Google mobility). If contacts in a layer are already substantially reduced, we might find that shutting that layer down completely does not have much impact on infection; this, obviously, does not reflect the importance of the layer, but rather the fact that the layer was already almost closed. For this reason, we conduct this experiment in the fully open (pre-pandemic) network.

Figure 7 shows the evolution of the disease under different scenarios. In both panels, we show epidemic curves for the network with all layers open (benchmark), with the H-contact large establishments shut down, with the L-contact establishments shut down and finally with schools shut down.[16] The panel on the left uses the infection probability parameter $\pi$ from our benchmark calibration, while the right panel plots the same curves with a lower infection probability parameter, which we use later in our re-opening experiments.

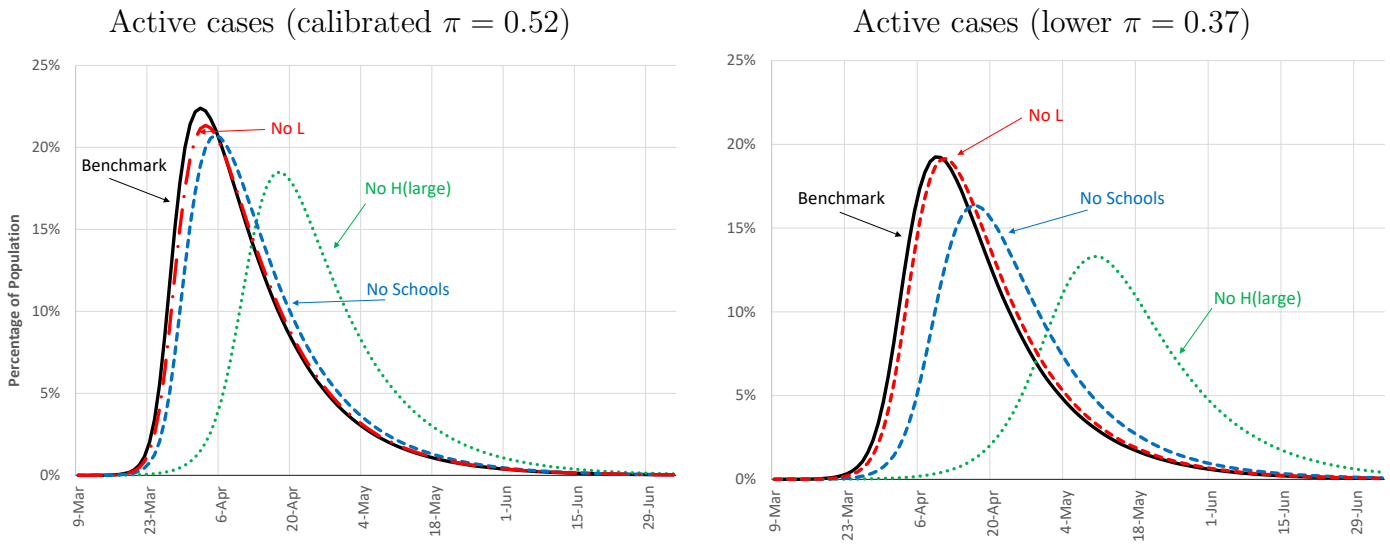


**Figure 7:** Infection Decomposition

Both panels show that the component of the network which has the biggest impact on infections is the large H-establishments. Shutting those establishments down achieves a substantial delay of the peak of infection and a substantial reduction of total infections.

[16] We do not plot curves for public transportation, and H-contact small establishments as they have a very small impact.

There are two reasons for this. The first is that, as shown in Table 4, workers in the large establishments in the H-sector have the highest number of contacts, so they are obviously more likely to get sick and spread the infection. The second is that the customers of these workers are randomly drawn from the entire city; this makes their layer very close to a random mixing set-up, and thus very conducive to the rapid spread of infections. Another layer that is quantitatively relevant is the one of schools. Table 4 shows that kids in schools have a high number of contacts, also randomly drawn, albeit from a smaller set. We find it interesting to compare the right and the left panel of Figure 7. The curves in the right panel are drawn with a smaller infection probability parameter and thus, not surprisingly, they are lower, as there is less infection spreading. Note however that with a smaller $\pi$ the impact of shutting down high contact layers gets magnified. To see why consider the case of a susceptible node with many contacts that are infected. If $\pi$ is close to 1 the probability of infection increases rapidly to 1 with the first contacts but additional contacts do not have much impact. In this case shutting down a layer (and thus reducing the number of contacts) does not affect much infection dynamics, which is always very fast. On the other hand, when $\pi$ is lower (but sufficiently far from 0), Equation 2 implies that shutting down a layer with many contacts can significantly reduce the infection probability. Therefore in this case infection dynamics are more sensitive to the network structure, and mitigation policies that reduce the number of contacts are effective. This highlights an important point, namely the *complementarity* between mitigation policies that reduce the transmission of the disease (e.g. face masks) and mitigation policies that reduce the number of contacts (i.e. shutting down malls). If the transmissibility of the disease is high ($\pi$ close to 1), then a moderate reduction in contacts is not very effective in reducing infection. Similarly, if individuals have a large number of contacts, a moderate reduction in transmissibility is not effective. However, if the transmissibility is lower, then the same reduction of contacts can have a large impact on the spread of infection, and similarly if the number of contacts is lower, the same reduction in the transmissibility of the disease can have a large impact on infection. We will return to these considerations later when we analyze re-opening policies.

Having established that the network model constitutes a good benchmark to study the evolution of the pandemic, and having elucidated the importance of various layers we now use the model to conduct two types of policy experiments. The first set (in Section 7) studies how counterfactual policies would have affected ECON-EPI outcomes at the outbreak of the COVID 19 crisis in New York City. These experiments are also helpful to evaluate different options, should a second wave of infection hit. The second set of experiments, in Section 8,

studies different strategies for reopening the city, as the infection subsides.

# 7 Policy Counterfactuals

As figure 6 shows, after the draconian lockdown of March and April, infections in the New York metro area stopped increasing by mid May. The question that is commonly asked is whether the lock-down was 'too draconian.' To analyze this claim, we perform a series of experiments that relax lock-down restrictions in the first four weeks of the pandemic (e.g., between March 8th and April 5th). With the lessons drawn from those experiments, we design a counter-factual *smart* mitigation policy that targets sectors with higher risk of spreading. We show that this policy could have reduced infections and increased output relative to the benchmark case.

We start from our benchmark case and compare it with three counterfactuals in which gradually bring back the same number (around 1% of the pre-pandemic total employment) of shutdown workers in each sector (L, H-small, and H-large). We then compare the epidemiological and economic outcomes to our benchmark case. Starting first with the epidemiological outcomes (the top panel) we see that increasing workers in the H-large sector has a very large impact (over 1.5% of the population) on infection, extra workers in the H-small sector have a moderate impact (0.5% of population), while extra workers in the L-sector have almost no impact on infection. The large increase in infection brought about by the additional H-large workers is not surprising; as discussed earlier, these workers have a lot of random contacts thus they function as spreaders. The increase in infection coming from the H-small workers is more surprising as, in section 6 we discussed that shutting down the H-small workers when the whole economy is open had no impact on infection dynamics. The reason for this difference is the starting point of the experiment. Adding H-small workers when the economy is substantially shutdown contributes to the spreading, while the marginal contribution of the H-small workers when the economy is fully open is small. Finally the L-workers constitute highly clustered groups in their respective productive units, who meet frequently and do not interact with customers. For these reasons a relatively small increase in the number of these workers does not affect infections on the margin.

Moving now to the economic outcomes (the bottom panel), we first observe that the largest output gain (around 2% of GDP) is obtained by adding the H-large workers, followed by an output gain of 1.5% of GDP, obtained by adding L-workers; the smallest output gain (around 1%) is obtained by bringing back the H-small workers. To understand this ranking

consider that the marginal productivity of a worker is increasing in the capital labor ratio of the worker's sector. In the pre-pandemic equilibrium the capital labor ratio for the L-workers is higher than the one of H-workers (small and large). During the pandemic however, workers in the H-large sector are mostly shutdown, hence their capital output ratio is the highest: that explains why bringing them back gives the highest output gain. L-workers and H-small workers are instead shutdown in roughly the same proportion, and therefore, because L-workers have a higher capital to start with, bringing them back results in a larger gain (relative to the H-small).

The results so far suggest that a reallocation of workers from the H to the L sectors during the shutdown might actually achieve a reduction in infection *and* an increase in output, relative to the benchmark. In Figure 9, we consider the effect of such a policy, which we label smart mitigation. More specifically, we impose stricter lock-down measures in the H-sector (mostly in the H-small sector) while relaxing those in the L-sector. We impose that the total number of individuals going to work is the same as the benchmark and that the amount of workers reallocated across sectors is around 1% of pre-pandemic employment (the same amount considered in the experiments in Figure 8. A concrete example of such a policy would be to allow more workers in manufacturing plants to go to work, while furloughing an equal number of retail workers that are allowed to go to work in the benchmark. The figure shows that the smart mitigation achieves a substantial double-gain. The top panel shows that it reduces the number of infections by 1.5% of the population of the metro area (300 thousands fewer cases) and the bottom panel shows that at the same it increases output by an average of 1.0% each day.

The reason why we can realize this double gain is that most of the reallocation is from the H-small sector into the L-sector. We can think of our policy as a two steps procedure. The first step is to add workers to the L-sector. As Figure 8 shows this step involves an increase in output and virtually no change in infection. The second step involves a reduction of (mostly) H-small workers. Figure 8 suggests that this causes a substantial reduction in infection and a reduction in output that is smaller than the gain obtained in the first step: hence the double gain.

# 8 Re-opening strategies

Results in Section 5 suggest that the network model captures well infection dynamics in the lock-down period. However, as the city starts to reopen in the month of May, the

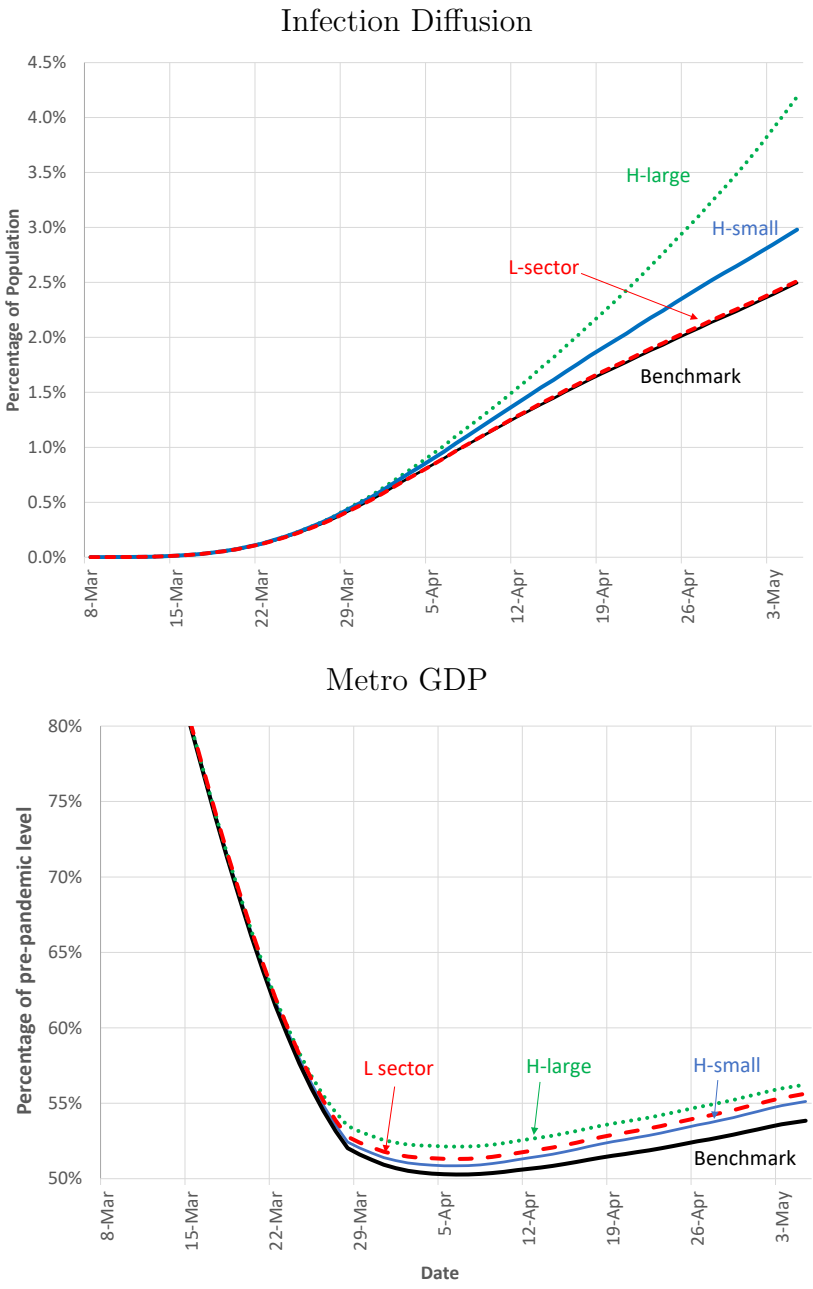


**Figure 8:** Policy counterfactual

model predicts a level of infection that is higher than the data. One possible reason for this discrepancy is that we keep the epidemiological parameters constant throughout our period, while the much broader availability of PPE and of testing, together with social distancing (e.g. requiring individuals to meet 6 feet apart form each other) has reduced the transmissibility of the disease. As this issue is critical to analyze reopening scenarios, we incorporate changes in transmissibility by assuming that in the post lock down period (after

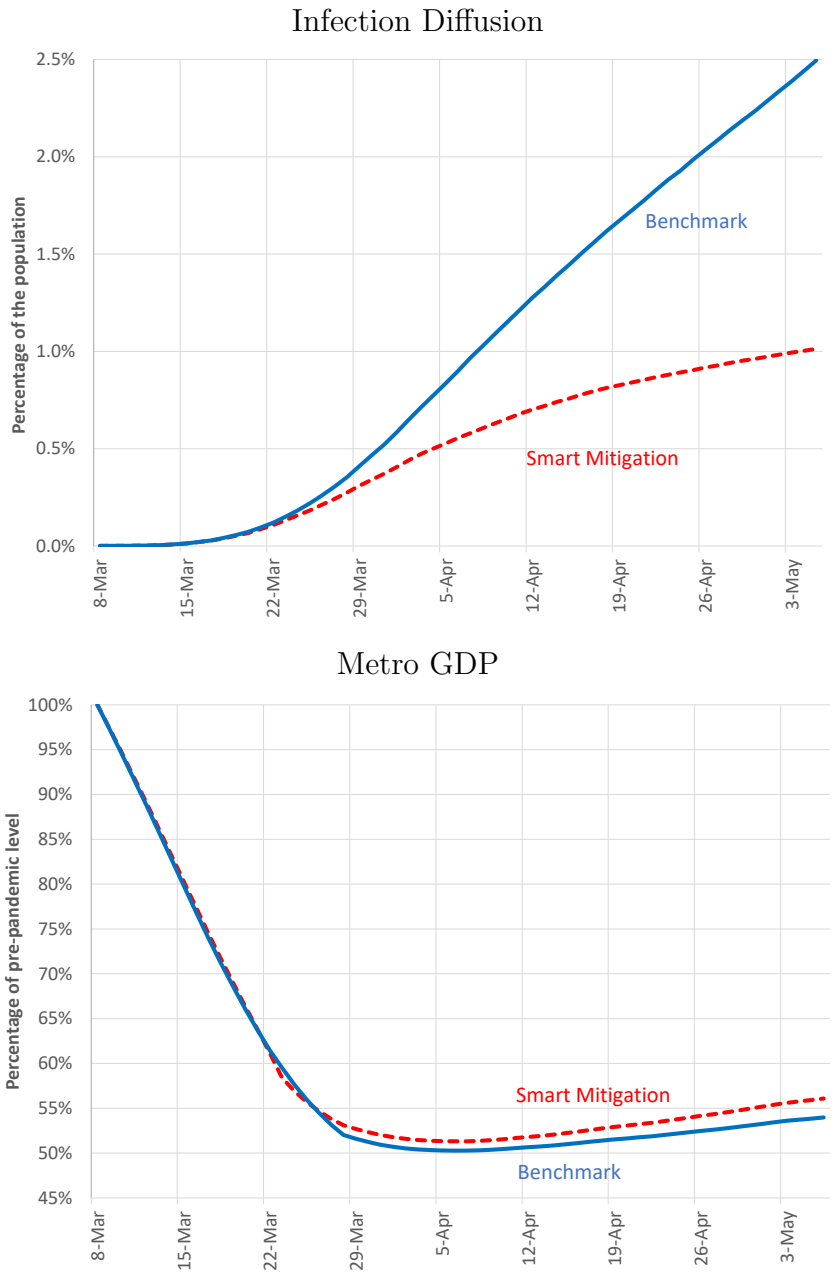


**Figure 9:** Smart Mitigation

4/26) there is a one-time decline in the parameter $\pi$ and we calibrate this decline (from 0.52 to 0.37,) so that the infection curve in the reopening period (4/26 through 5/25) matches the data. The result of this procedure are illustrated in figure 10. The figure suggests that the network model with the recalibrated $\pi$ can be a good starting point to study reopening strategies, that is to predict the evolution of infection and output under different assumptions for the evolution of mobility.

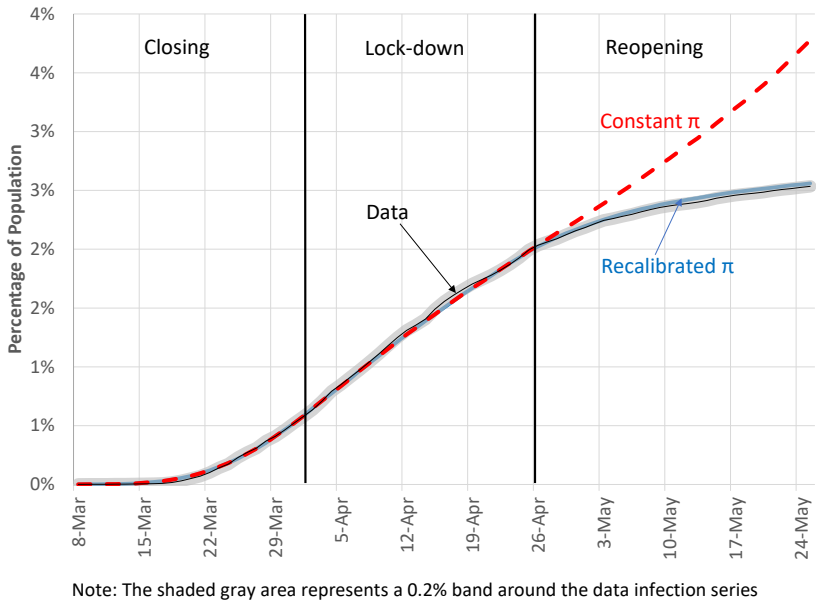


**Figure 10:** The impact of lower $\pi$

## 8.1 Re-opening workplaces

In Figure 11 we consider three scenarios for the New York metro. The first (labeled no re-opening) is the one in which mobility stops increasing on May 25th. The second (labelled L-reopening) is the one in which only the L-establishments are allowed to substantially re-open, and the third (labelled H and L re-opening) is the one in which both L and large H-establishments are allowed to substantially reopen.[17] The top panel depicts the infection curves, while the bottom panel shows GDP. Under the no-reopening scenario GDP remains severely depressed; on the positive side, the cumulative infection curve becomes flat, suggesting that a prolonged shutdown can stop the growth and thus eradicate the disease.[18] The dashed lines show that a substantial reopening of the L-sector comes at virtually no infection cost, and with large GDP gains, as the metro area GDP would recover almost up to 25% of the pre-pandemic level. Finally the dotted lines, displaying the consequences of a re-opening of both the L and large H-sector, suggest that this scenario is potentially troublesome. GDP would recover more substantially, but infection would show a dramatic second wave, with total infected reaching over 7% of the population by early September, and thus it is doubtful

[17] Across the three experiments the re-opening pattern of the small H-establishments is kept constant, and schools are kept closed.

[18] In appendix D we show how to compute another popular statistic that tracks the growth of infection, namely the $R_t$ or basic reproduction rate, in our model. We show that in the case of no reopening the $R_t$ drops and stays below 1.

whether in such a scenario the GDP recovery can be sustained.

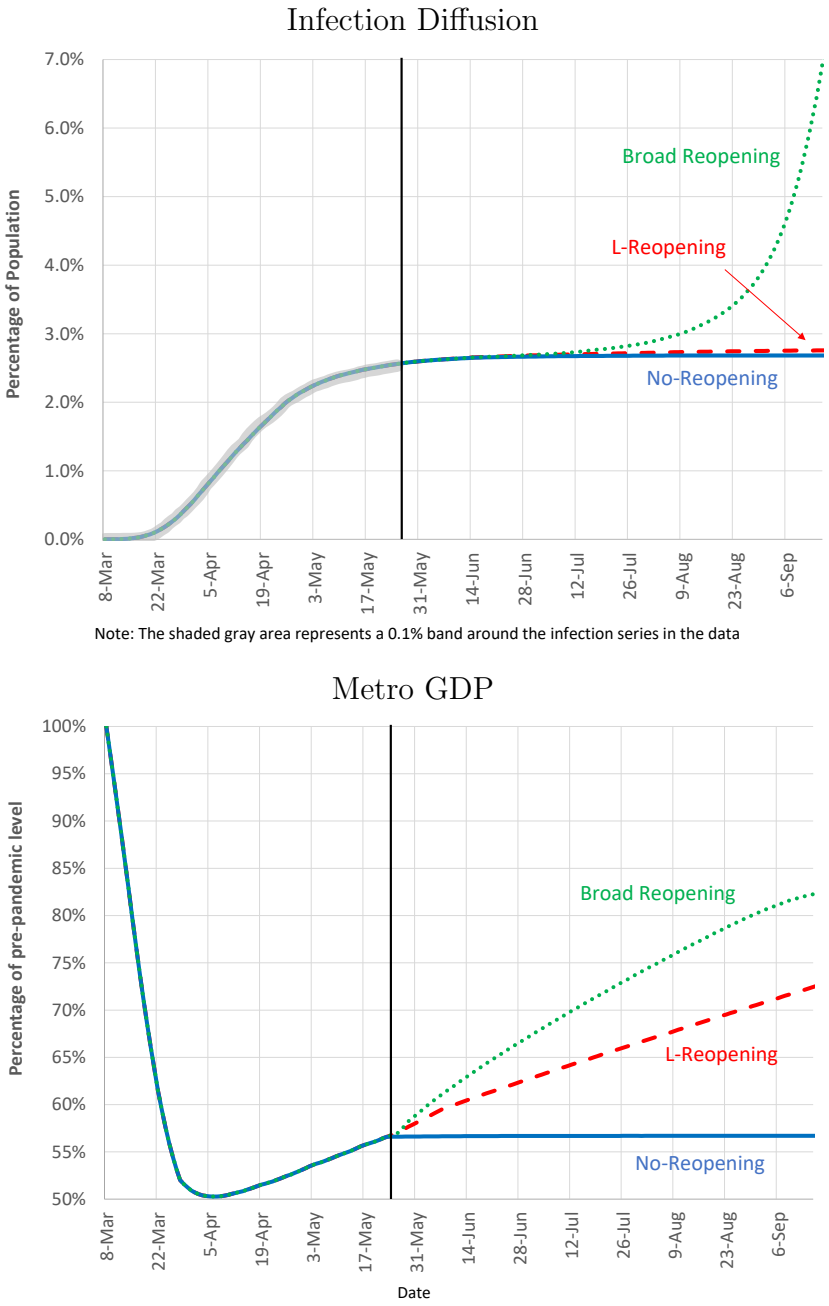


**Figure 11:** Reopening scenarios

## 8.2 Re-opening schools

In the experiments above, we assumed that schools remained closed until the end of the year. One important issue during the COVID pandemics is the impact of re-opening activities for

the kids, such as schools and summer camps. The impact of reopening these activities on infections depends significantly on the current level of infections, and hence on the date in which the re-opening happens. In the top panel of Figure 12, we depict the effects of school reopenings in different dates assuming that mobility in the L and H sectors stays constant at the level of May 25th. On the right scale of the panel, we plot the curve depicting the level of active cases under the scenario of no-reopening of schools. This curve shows that in July there is still a positive mass of active cases; thus if schools were to re-open in July the large increase in contacts brought about by the open schools would imply a fast growth of infection which would reach 10% by mid November. If, on the other hand, schools were to re-open in August, when the mass of active cases is minimal, the addition of contacts from schools would not be causing a rapid takeoff of infection, so the infection would be manageable.

In our final experiment, we consider re-opening scenarios that combine increases in mobility of the L-sector (such as those in Figure 11) with school re-openings. We consider two scenarios: the benchmark case in which schools reopen with the normal number of contacts, and a socially distanced scenario where contacts are reduced to 2. We assume that schools open on August 1st, and that school-related activities end by Thanksgiving day. The progression of infections is shown in the bottom panel of Figure 12. The key result is that, even in the case of strongly socially distanced schools, the infection takes off rapidly and exceeds 5% of the population by November. The reason why this happens is that the impact of reopening the L-sector, despite being fairly modest in itself, is sufficient to keep the number of active cases high enough, so that when schools are re-opened in August (and the L-sector also stays open) infections take off. This experiment highlights the importance of interactions among social layers in our network. Reopening just the L-sector or just schools (in August) results in a relatively small increase in infections, but opening both simultaneously results in a distinctive second wave during the Fall.

# 9 Conclusion

We developed an ECON-EPI network model of human interaction and disease progression to study the consequences of alternative mitigation policies and re-opening strategies associated with the COVID-19 pandemic. The model, calibrated to the NY metro area, suggests that re-opening sectors in which workers interact with each other in stable teams (such

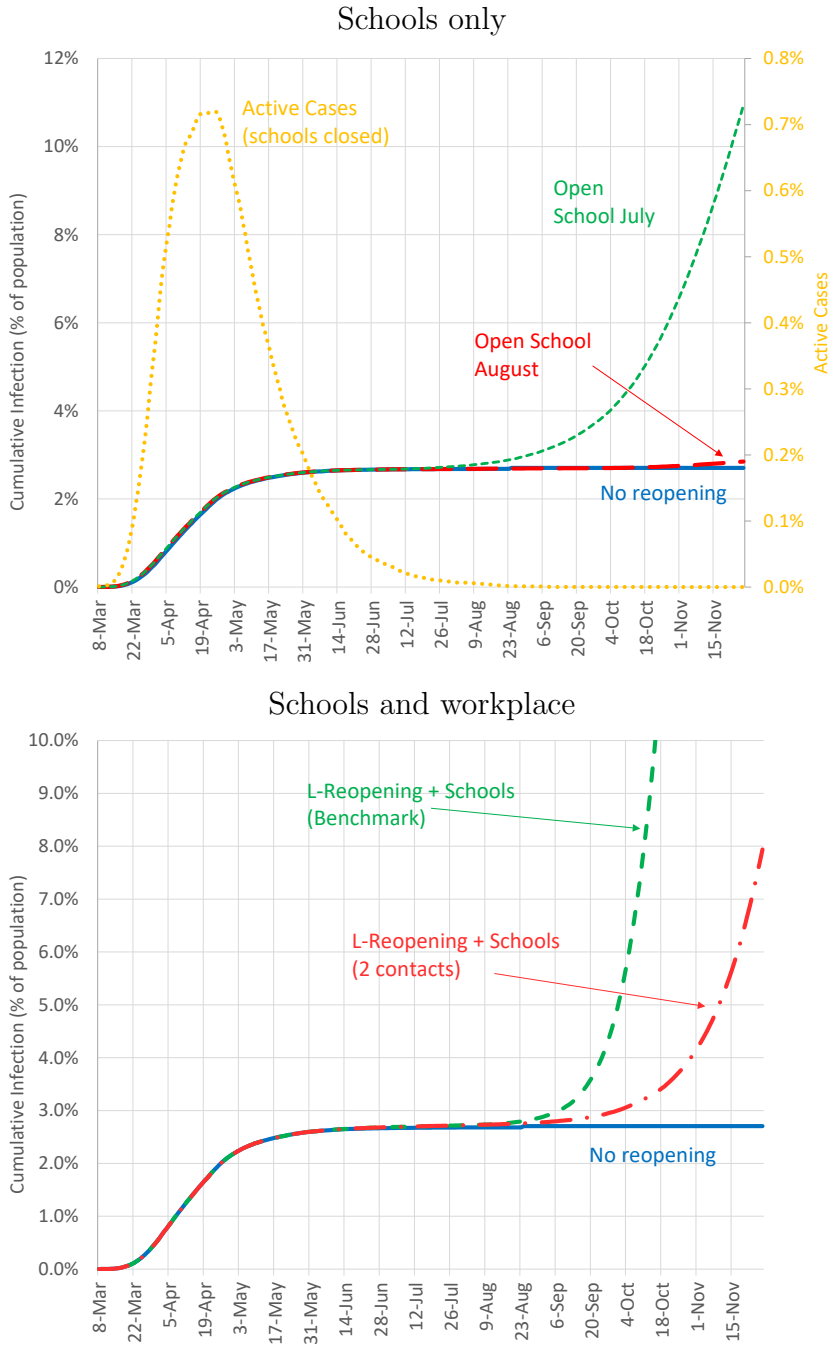


**Figure 12:** Reopening Schools

as manufacturing) is the best strategy to output recovery, while at the same time keeping the spread of the disease under control. A broad re-opening strategy including sectors whose workers interact with a large customer base (such as malls and large retail stores), on the other hand, would significantly increase the risk of a second wave. According to our

model, under a broad re-opening strategy infections would be increase marginally until mid-August, but significantly accelerate afterwards, with cumulative infections surpassing 7% of the population by early September. We also show that re-opening the low contact sector together with schools would be a feasible strategy, but only if social distancing measures are strictly observed and density in the classroom can be significantly reduced. This highlights the importance of interactions between the work, school, and home layers in our network. Our analysis can be further enriched by incorporating endogenous decisions by individuals to reduce their mobility when increases in infections are observed. Another important extension would involve tying the ability to return to work with school re-openings. In other words, to incorporate the fact that many workers would not be able to join the labor force unless they are able to send their kids back to school. Finally, the analysis can be extended to other properly calibrated metro areas. This could allow us to investigate the role of population density, shopping patterns, and sectoral productivity to the costs and benefits of mitigation and re-opening strategies.

# References


Acemoglu, D., V. Chernozhukov, I. Werning, and M. D. Whinston (2020, May). A multi-risk sir model with optimally targeted lockdown. Working Paper 27102, National Bureau of Economic Research.

Acemoglu, D., G. Como, F. Fagnani, and A. E. Ozdaglar (2013). Opinion fluctuations and disagreement in social networks. *Mathematic of Operations Research*.

Acemoglu, D., A. Ozdaglar, and A. ParandehGheibi (2010). Spread of (mis)information in social networks. *Games and Economic Behavior 70*(2), 194 – 227.

Akbarpour, M., C. Cook, A. Marzuoli, S. Mongey, A. Nagaraj, M. Saccarola, P. Tebaldi, S. Vasserman, and H. Yang (2020). Socioeconomic network heterogeneity and pandemic response. Working paper.

Alon, T. M., M. Doepke, J. Olmstead-Rumsey, and M. Tertilt (2020, April). The impact of covid-19 on gender equality. Working Paper 26947, National Bureau of Economic Research.

Alvarez, F., D. Argente, and F. Lippi (2020). A simple planning problem for covid-19 lockdown. Technical report, Manuscript.

Baqaee, D., E. Farhi, , D. Allen, M. Mina, J. Stock, V. De Falco, M. Droste, A. Fernandes, K. Holston, S. Kestelman, E. Kong, D. Maroz, C. Okafor, and L. Xu (May 2020). Evaluating the economics and epidemiology of strategies for reopening the economy. Technical report, Manuscript.

Benzell, S., A. Collis, and C. Nicolaides (2020). Rationing social contact during the covid-19 pandemic: Transmission risk and social benefits of us locations. Working paper, SSRN 3579678.

Berger, D., K. Herkenhoff, and S. Mongey (2020). An seir infectious disease model with testing and conditional. Technical report, Manuscript.

Chari, V. V., R. Kirpalani, and C. Phelan (2020). An economic examination of social distancing. Working paper.

Colizza, V., A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani (2007). Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Medicine 4*(13).

Cooley, T. F. and E. C. Prescott. Economic growth and business cycles. Princeton, New Jersey.

Dave, D. M., A. I. Friedson, K. Matsuzawa, and J. J. Sabia (2020, May). When do shelter-in-place orders fight covid-19 best? policy heterogeneity across states and adoption time. Working Paper 27091, National Bureau of Economic Research.

Dingel, J. I. and B. Neiman (2020). How many jobs can be done at home? Technical report, National Bureau of Economic Research.

Eichenbaum, M. S., S. Rebelo, and M. Trabandt (2020, March). The macroeconomics of epidemics. Working Paper 26882, National Bureau of Economic Research.

Fang, H., L. Wang, and Y. Yang (2020, March). Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china. Technical report, Manuscript.

Favero, C., A. Ichino, and A. Rustichini (2020). Restarting the economy while saving lives under covid-19. Working paper.

Ferguson, N., D. Laydon, G. Nedjati-Gilani, and N. Imai (2020, March 16). Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. Technical report, Nature.

Fogli, A. and L. Veldkamp (2020). Germs, social networks and growth. *Review of Economic Studies Forthcoming*.

Foster, L., J. Haltiwanger, and C. Krizan (2006). Market selection, reallocation, and restructuring in the u.s. retail trade sector in the 1990s. *The Review of Economics and Statistics 88*(4), 748–758.

Ganesh, A., L. Massoulie, and D. Towsley (2005). The effect of network topology on the spread of epidemics. *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*.

Glass, R. J., L. M. Glass, W. E. Beyeler, and H. J. Min (2006). Targeted social distancing design for pandemic influenza. *Emerging infectious diseases 12*(11).

Glover, A., J. Heathcote, D. Krueger, and V. Rios-Rull (2020). Health versus wealth: On the distributional effects of controlling a pandemic. Technical report, Manuscript.

Guan, W.-j., Z.-y. Ni, Y. Hu, W.-h. Liang, C.-q. Ou, J.-x. He, L. Liu, H. Shan, C.-l. Lei, D. S. Hui, B. Du, L.-j. Li, G. Zeng, K.-Y. Yuen, R.-c. Chen, C.-l. Tang, T. Wang, P.-y. Chen, J. Xiang, S.-y. Li, J.-l. Wang, Z.-j. Liang, Y.-x. Peng, L. Wei, Y. Liu, Y.-h. Hu, P. Peng, J.-m. Wang, J.-y. Liu, Z. Chen, G. Li, Z.-j. Zheng, S.-q. Qiu, J. Luo, C.-j. Ye, S.-y. Zhu, and N.-s. Zhong (2020). Clinical characteristics of coronavirus disease 2019 in china. *New England Journal of Medicine 382*(18), 1708–1720.

Harris, J. E. (2020, April). The subways seeded the massive coronavirus epidemic in new york city. Working Paper 27021, National Bureau of Economic Research.

Hsiang, S., S. Annan-Phan, D. Allen, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, A. Hultgren, L. Y. Huang, E. Krasovich, P. Lau, J. Lee, E. Rolf, J. Tseng, and T. Wu (2020). The effect of large-scale anti-contagion policies on the coronavirus (covid-19) pandemic. Technical report, Manuscript.

Jackson, M., M. (2010). *Social and Economic Networks*. Princeton University Press.

Kaplan, G., B. Moll, and G. Violante (2020). Pandemics according to hank. Technical report, Manuscript.

Karaivanov, A. (2020, April). A social network model of covid-19. Technical report, Manuscript.

Keeling, M. J. and K. T. Eames (2005). Networks and epidemic models. *Journal of the Royal Society 2*.

Kermack, W. O. and A. G. McKendrick (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London 115*.

Klepac, P., A. Kucharski, A. Conlan, S. Kissler, M. Tang, H. Fry, and J. R. Gog (2020, March 5). Contacts in context: large-scale setting-specific social mixing matrices from the bbc pandemic project.

Kuperman, M. and G. Abramson (2001, Mar). Small world effect in an epidemiological model. *Physical Review Letters 86*, 2909–2912.

Leibovici, F., A. M. Santacreu, and M. Famiglietti (2020). Social distancing and contact-intensive occupations.

Li, R., S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science 368*, 489–493.

Moro, A. and A. Bisin (2020). Learning epidemiology by doing: The empirical implications of a spatial sir model with behavioral responses. Working paper.

Mossong, J., N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, J. Scalia Tomba, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine 5*(3).

Nowzari, C., V. M. Preciado, and G. J. Pappas (2016). Analysis and control of epidemics: A survey of spreading processes on complex networks. *Nowzari 36*.

Pastor-Satorras, R., C. Castellano, P. Van Mieghem, and A. Vespignani (2015). Epidemic processes in complex networks. *Reviews of Modern Physics 87*.

Rampini, A. A. (2020, April). Sequential lifting of covid-19 interventions with population heterogeneity. Working Paper 27063, National Bureau of Economic Research.

Vespignani, A. (2009). Predicting the behavior of techno-social systems. *Science 325*(5939,).

Watts, D. and S. Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature 393*.

Zhao, P. J. (2020). A social network model of the covid-19 pandemic. Technical report, Manuscript.

# A Mapping ONET Data to BLS and NAICS Codes

We calculate the work intensity indices based on O*NET data. In particular, we look at the following questions from the Work Context Module of the ONET data:

1. **Question 1** How often does your current job require face-to-face discussions with individuals and within teams?

2. **Question 8** In your current job, how important are interactions that require you to deal with external customers (as in retail sales) or the public in general (as in police work)?

3. **Question 21** How physically close to other people are you when you perform your current job?

4. **Question 29** How often does your current job require that you be exposed to diseases or infection? This can happen with workers in patient care, some laboratory work, sanitation control, etc.

The table shows available responses for Questions 1, 8, 21, and 29 from the Work Context Module of ONET survey.

Table 6 shows the available responses for these questions in the ONET survey. Each of these questions measure a different characteristic of the job as – (1) Face-to-Face interactions with individuals or teams, (2) Dealing with external public, (3) Physical proximity to colleagues, and (4) Exposure to diseases at the job.

We construct four intensity indices, at the 2010 O*NET Standard Occupation Classification (SOC, hereafter) following the framework as in Leibovici et al. (2020). Subsequently, we go from the O*NET 2010 SOC to the 2018 BLS SOC and the NAICS industry level using the method described in Dingel and Neiman (2020) and calculate indices based on BLS occupation level and the NAICS industry level.

Section A.1 describes the data and Section A.2 describes the methodology to construct the indices and the crosswalk from ONET 2018 SOC to (1) BLS SOC and (2) NAICS Industry level.

| Response | Question 1 | Question 8 | Question 21 | Question 29 |
|---|---|---|---|---|
| 1 | Never | Not important at all | I don't work near other people (beyond 100 ft.) | Never |
| 2 | Once a year or more but not every month | Fairly important | I work with others but not closely (e.g. private office) | Once a year or more but not every month |
| 3 | Once a month or more but not every week | Important | Slightly close (e.g. shared office) | Once a month or more but not every week |
| 4 | Once a week or more but not every day | Very important | Moderately close (at arm's length) | Once a week or more but not every day |
| 5 | Every day | Extremely important | Very close (near touching) | Every day |

**Table 6:** Available responses for Questions 1, 8, 21, and 29 from the ONET survey.

## A.1 Data

The questions mentioned above are extracted from the US Occupational Information Network (O*NET) survey covering the "work context" module. The O*NET 24.2 database (the most current release) is used for this analysis [19]

1. 2010 ONET SOC to 2010 OES[20] SOC : It can be downloaded here.
2. 2010 OES SOC to 2018 & 2019 OES SOC : It can be downloaded here.
3. 2018 OES SOC to NAICS 2-digit level : It can be downloaded here.

The O*NET 24.2 release is based on the O*NET 2010 SOC. We do a crosswalk with the occupations in the 2018 OES SOC that is used by BLS in two steps – first, we match the O*NET 2010 SOC in the `onet_*.csv` files with the 2010 OES SOC using the first file and second, we merge the 2010 OES SOC with the 2018 OES SOC using the second file. This is the procedure followed in Dingel and Neiman (2020).

Finally, we go from the 2018 OES SOC to the NAICS 2-digit industry level using the third file. This gives us a mapping from the occupation level to the industry level. This

[19]The link to O*NET data can be found here: overview

[20]Occupation Employment Statistics

procedure is also the same as in Dingel and Neiman (2020). There is one caveat here: We are using the May 2019 OES data for New York which is based on the 2019 OES SOC[21]

## A.2 Methodology

For each question described above, we give each response a score and then aggregate the scores, weighted by the percentage of responses by occupation, to calculate the intensity index. The responses are scored as follows:

1. 1 – Score of 0
2. 2 – Score of 25
3. 3 – Score of 50
4. 4 – Score of 75
5. 5 – Score of 100

This is the same methodology used by Leibovici et al. (2020). However, we calculate these indices at the 6-digit level instead of the 4-digit level as in their blog post. After the BLS crosswalk, we re-do the same calculations as above for each occupation at the 2018 OES SOC level.

After the NAICS crosswalk, we calculate the weighted average of the index (calculated above) at the NAICS level where the weights are the total employment numbers at the national level divided by the total employment in that industry at the national level. This gives us the indices at three different levels – (1) O*NET 2010 SOC (Occupation level), (2) BLS 2018 OES SOC (Occupation level), and (3) NAICS 2-digit (Industry level).

## A.3 Intensity and Proximity

Using this information, we construct two indexes: the Proximity Index and the Intensity Index. The high and low sectors are summarized in the table below.

We plot the correlation between the two indexes in Figure 13

[21] See overview

| | | External Index | |
|---|---|---|---|
| | | High | Low |
| Proximity Index | High | Retail Trade | Educational Services |
| | | Accommodation and Food Services | Transportation and Warehousing |
| | | Health Care and Social Assistance | Construction |
| | | Arts, Entertainment, and Recreation | Administrative and Support |
| | | Federal, State, and Local Government | |
| | | Other Services | |
| | Low | Real Estate and Rental and Leasing | Management of Companies and Enterprises |
| | | Finance and Insurance | Professional, Scientific, and Technical Services |
| | | Wholesale Trade | Mining, Quarrying, and Oil and Gas Extraction |
| | | Information | Manufacturing |
| | | | Agriculture, Forestry, Fishing and Hunting |
| | | | Utilities |

| NAICSCode | labels |
|---|---|
| Retail Trade | Retail |
| Accommodation and Food Services | AFS |
| Health Care and Social Assistance | Health Care |
| Arts, Entertainment, and Recreation | AER |
| Federal, State, and Local Govt. | Federal, State, and Local Govt. |
| Educational Services | Education |
| Other Services (except Public Administration) | Others |
| Real Estate and Rental and Leasing | Real Estate |
| Finance and Insurance | F&I |
| Transportation and Warehousing | Transport |
| Wholesale Trade | Wholesale |
| Construction | Construction |
| Information | Information |
| Administrative and Support and Waste Management and Remediation Services | Admin Support |
| Utilities | Utilities |
| Management of Companies and Enterprises | Management |
| Mining, Quarrying, and Oil and Gas Extraction | Mining |
| Professional, Scientific, and Technical Services | Scienctific Services |
| Manufacturing | Manufacturing |
| Agriculture, Forestry, Fishing and Hunting | AFFH |

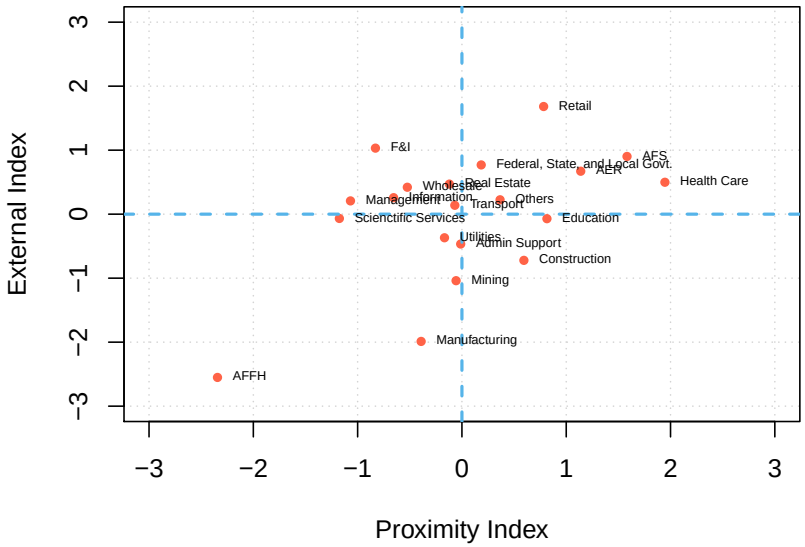


**Figure 13:** Intensity and Proximity Indexes.

## A.4 Employment shares

The share of employment in the L-Sector is computed using information from the '2016 SUSB Annual Data Tables by Establishment Industry' from the US Census[22].In particular, we obtain the percentage of a subset of low-contact occupations (as defined in the previous section) in different sectors in the New York MPS, and multiply this by the share of workers to adult population. To do this, we obtain the level of employment from the the table 'Number of Firms, Number of Establishments, Employment, and Annual Payroll by Enterprise Employment Size for Metroplitan Areas, NAICS Sectors: 2016,' and consider the following sectors: 'Agriculture, Forestry, Fishing and Hunting;' 'Mining, Quarrying, and Oil and Gas Extraction;' 'Utilities;' 'Construction;' 'Manufacturing;' 'Wholesale Trade;' 'Information;' 'Finance and Insurance;' 'Real Estate and Rental and Leasing;' 'Professional, Scientific, and Technical Services;' 'Management of Companies and Enterprises;' 'Administrative and Support and Waste Management and Remediation Services;' 'Educational Services;' and 'Health Care and Social Assistance (enterprise employment $>$ 500 employees).'

The share of employment in the H-Sector are: 'Retail Trade;' 'Transportation and Warehousing;' 'Arts, Entertainment, and Recreation;' 'Accommodation and Food Services;' and 'Health Care and Social Assistance (enterprise employment $<$ 500 employees).'

[22]See https://www.census.gov/data/tables/2016/econ/susb/2016-susb-annual.html

# B Changes in regulation in the New York Metro

The sequence of measures imposed by the NY government, aimed at slowing down the spread of the disease, is summarized in Figure 14. Increasingly stricter mitigation policies reducing gatherings, retail and production activities were implemented in a short span of time.

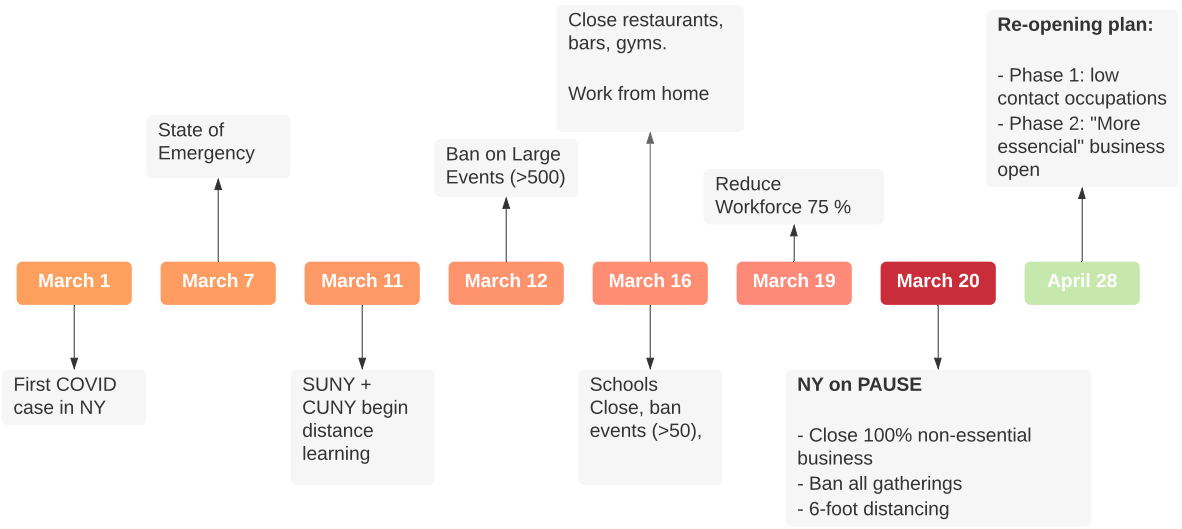


**Figure 14:** Timeline of lock-down Policies

# C Network Characteristics Computation

We compute: Average Degree, Global Clustering Coefficient, Average Path Length and Max Triangles. Their definitions follow

**Degree** The degree $d_i$ of node $i$ satisfies

$$d_i = \sum_{j=1}^{M} g_{i,j}$$

The average degree is just

$$D = \frac{1}{M} \sum_{i=1}^{M} d_i$$

**Global Cluster** A clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. The global clustering coefficient is based on triplets of nodes. A triplet is three nodes that are connected by either two (open triplet) or three (closed triplet) undirected ties. A triangle graph therefore includes three closed triplets, one centered on each of the nodes (n.b. this means the three triplets in a triangle come from overlapping selections of nodes). The global clustering coefficient is the number of closed triplets (or 3 x triangles) over the total number of triplets (both open and closed). The global clustering coefficient is defined as:

$$GC = \frac{\text{number of closed triplets}}{\text{number of all triplets (open and closed)}}$$

The number of closed triplets has also been referred to as 3 $\times$ triangles in the literature, so:

$$GC = \frac{3 \times \text{number of triangles}}{\text{number of all triplets}}$$

**Avg Path Length** We take a sample of $Q = 1,000$ 'source nodes' out of our total sample of M=1,000,000 nodes. We compute the shortest-path between each source node and all nodes in the network using the Dijkstra's algorithm (see Wikipedia for a nice summary of the algorithm). Let $a_{ij}$ denote the shortest path between nodes $i$ and $j$, then the average path length is

$$APT = \frac{1}{Q} \sum_{i}^{Q} \left\{ \frac{1}{M-1} \sum_{j=1}^{M} a_{ij} \right\}$$

Note that we divide by $M - 1$ because $a_{ii} = 0$ by construction.

The effects of these reductions in mobility on the structure of the network can be analyzed by computing network characteristics at different periods in time. In Table 7, we report: the average degree (e.g., the average number of contacts per person), the global clustering coefficient, and the average path length. In the first column, these statistics are computed in the pre-pandemic equilibrium in a network with full mobility. In the second column, we maintain full working and shopping mobility, but close schools down. In the last two columns, we compute them at the end of the calibration period, emphasizing the effect of schools (e.g. in the third column we assume schools are still open, whereas in the fourth one we shut them down).

**Table 7: Network Characteristics**

| Statistic | Pre-pandemic | | Pandemic (April 3rd) | |
|---|---|---|---|---|
| | Open Schools | Closed Schools | Open Schools | Closed Schools |
| Degree | 12 | 11 | 7 | 6 |
| Global Clustering Coeff. | 0.16 | 0.14 | 0.33 | 0.43 |
| Avg. Path Length | 6 | 6.2 | 16 | 17 |

The first thing to note is that lock-down policies significantly reduce the average number of contacts per person, regardless of whether schools remain opened or closed. This can be seen by comparing the first two columns, where the average number of contacts on a given day declines almost by half. A lower number of daily contacts slows the spread of the disease by reducing each individual's probability of becoming infected. These policies also make the network more clustered. Global clustering gives us a sense of the density of the graph by capturing the probability that a friend of your friend is also your friend. In a clustered network, individuals are more stably connected to each other, which allows for a rapid local depletion of those who are susceptible. The reduction in mobility also affected the topology of the network by increasing the average path length from 6, in the pre-pandemic equilibrium, to 16, when the lock-down period starts (or 17 when we, in addition, shut down schools). Because ours is a 'small world network' it displays, under full mobility, a number of long range contacts linking distant nodes. This property, also known as 'six degrees of separation,' allows the disease to spread rapidly through the population.[23] Policies that reduce the number of long links increase the average path length, which ultimately results in the infection to slowing down.

# D $R_t$ under different reopening scenarios

The $R_0$ is defined as the average number of secondary cases produced by an infectious individual in a totally susceptible population (e.g. when $S = M$). As time goes by, we can

[23]The benchmark network features the realistic characteristic that any two individuals in the synthetic city are connected through at most six contacts.

capture the same concept by calculating the ratio of new infections (e.g. number of newly exposed individuals) to those exiting a state where the disease can be transmitted. Clearly, an $R_t = 1$ indicates that the disease is at a steady state, whereas a value larger than one indicates that the disease is spreading. If $R_t < 1$, the disease eventually dies out. In our network model, the basic reproductive number can be calculated as

$$R_t = \frac{\sum_{i \in x_{it}=S} Pr(x_{it+1} = E / x_{it} = S)}{H_t}, \tag{6}$$

where the numerator denotes the total number of new infections (e.g. exposed individuals who were susceptible last period, as defined by eq. 2) and the denominator, $H_t$, denotes the number of individuals exiting the state at which they are contagious. Figure **??** displays the $R_t$ under the three re-opening scenarios discussed in section 8