Halting SARS-CoV-2 by Targeting High-Contact Individuals

Gianluca Manzo*

GEMASS

CNRS & Sorbonne University

Arnout van de Rijt

European University Institute

Utrecht University

ABSTRACT

Two decades ago network scientists proposed that infectious diseases may be effectively halted by targeting interventions at a minority of highly connected individuals. Can this strategy be effective in combating a disease transmitted through physical proximity such as the contemporary coronavirus? The strategy's effectiveness critically depends on high between-person variability in close-range contact. We analyze population survey data showing that indeed the frequency of close-range contact across individuals is characterized by a small fraction of people reporting very high frequencies of close-range contact. We simulate a population with empirical contact distributions coming out of lockdown. Simulations show that targeting a small fraction of high-degree individuals dramatically improves containment. Our results further suggest two concrete procedures for identifying high-contact individuals: acquaintance sampling and employment-based targeting.

Corresponding author: gianluca.manzo@cnrs.fr

1. Introduction

The world is currently plagued by the new coronavirus SARS-CoV-2 and the social and economic consequences of its containment. Many countries are in the process of gradually lifting a lockdown initially imposed to contain spread, reintroducing measures to contain a second wave. Most containment measures that are being lifted combine general interventions aimed at broad groups of citizens (children, elderly, contact professions) or categories of meeting places (schools, restaurants, airports) (Zhang et al. 2020). The economic costs of such measures are tremendous, expressed in percentages of GDP. At the same time, easing restrictions on interaction achieved through such measures could spur new waves of infections.

Two decades ago, network scientists proposed that viruses may be effectively halted by targeting interventions at a minority of highly connected individuals who are jointly responsible for the majority of viral traffic potential (Desző & Barabási 2001; Pastor-Satorras & Vespignani 2002). Compartmental models in use in the SARS-CoV-2 crisis context do not incorporate this idea (Manzo 2020). By minimizing the risk of spread through "hubs" – through repeated medical testing, quarantining-if-positive, extra protections for cashiers and flight attendants, and perhaps eventually vaccination – the current epidemic could theoretically be brought to a halt. Such targeted intervention would be economically efficient if achieved with relatively limited financial and medical resources.

Could this approach be made to work against a virus like SARS-CoV-2 that primarily spreads through droplet transmission and direct contact (Mittal 2020)? The theoretical analysis relies on the spreading network being scale-free, while empirical networks often deviate from this assumption (Jones & Handcock 2003; Clauset et al. 2009; Stumpf & Porter 2012). Nevertheless, degree-targeting may still be an effective strategy in the fight against SARS-CoV-2 if close-range contact exhibits high skew, with the majority of close-range contacts in society involving a small minority of individuals, as has been found for online contact (Barabási & Albert 1999; Adamic & Huberman 2002; Vázquez et al. 2002) and sexual contact (Liljeros et al. 2001; Trewick et al. 2013; Little et al. 2014). Available data suggests this is indeed the case for close-range contact as well. A study of short-range Bluetooth data showed high interpersonal variability among 700 university students, with a long tail (Mones et al. 2018; Sapiezynski et al. 2019). Evidence from contact-tracing in early stages of past outbreaks of SARS, Typhoid, Tuberculosis, Ebola, HIV, and also SARS-CoV-2 suggests that "super spreaders", working jobs that requires close-range contact with many others, often play an important role in initial dissemination (Glasser et al. 2011; Stein 2011; Wong et al. 2015; Sun et al. 2020). The literature notes a general tendency across viral diseases for 20 percent to generate roughly 80 percent of total infections (James et al. 2007: fig. 1). However, we do not know if this is primarily due to variability in contact or other forms of heterogeneity, e.g. in infectiousness (Woolhouse et al. 1997; Galvani & May 2005).

Here we draw on a large-scale population survey conducted some years ago in France (Béraud 2015). 2,033 people were interviewed and asked to report on all close-range contacts (< 2 meters) over the course of 2 days. From the survey data we derive the degree distribution for close-range contact in France (section 2). We then impose this empirical degree distribution on a synthetic social network (section 3.1). In this network, we introduce a virus mimicking the main empirical features of SARS-CoV-2, and let it spread through the network according to the common SEIR parametrization (section 3.2). We design different ways of reaching the best-connected nodes (section 3.3), and calculate how the trajectory of the epidemic varies under these interventions (section 4). We find that random acquaintance sampling is an

effective method to reach high-contact individuals and can considerably help "flattening" the curve (section 5). We conclude that, possibly coupled with job-specific targeting, this is a feasible practical method by which governments could identify high-degree nodes in contact networks, thus efficiently coping with the epidemic on the long-run.

2. Data analysis

We draw on data from COMES-F, a survey conducted in 2012 with data on the number of close-range contacts for a representative sample of about two thousand French residents (Béraud et al. 2015). The leading epidemiological compartmental models of SARS-CoV-2 spread in France currently rely on these data (see, for instance, Di Domenico et al. 2020; Roux et al. 2020; Salje et al. 2020a; in a comparative perspective, see Patrick et al. 2020). In these models, COMES-F data are routinely employed to build social contacts matrices, i.e. average contacts between age groups by places (like school, public transportation, home, etc.) (for an example of this methodology more generally, see, for instance, Prem et al 2017). In contrast to prior use of the data, we rely on the entire cross-individual heterogeneity of the observed distribution of close-range contacts. We implement this distribution in a social network model of disease propagation (section 3), so that we can evaluate the effectiveness of interventions targeted at high-connectivity individuals (section 4).

COMES-F is a large-scale population survey conducted in France during the first half of 2012. An initial sample of 24,250 was drawn from the French population excluding overseas territories through random-digit dialing of landline and mobile numbers. Using quotas for age, gender, days of the week and school holidays, 3,977 people who accepted to participate were sent a diary to complete. 2,033 (51%) contact diaries were returned (participants' age and household size were used as sampling weights to maintain representativeness). In these diaries, participants were asked to keep track of all short-range contacts over the course of 2 full days, and report on sex and age of these contacts, meeting context, and contact duration. A short-range contact was defined as talking to someone at less than 2 meters (possibly including physical contact). COMES-F made a special effort to collect high-quality data also on social contacts of children, defined as respondents aged less than 15 (753 respondents). For them, an adult member of the household completed the diary.

To relieve the reporting burden, respondents were asked to record up to 40 close-range contacts. In addition, respondents currently in employment were asked whether they regarded their occupation as especially exposed to short-range contacts. This turned out to concern 257 respondents. These respondents had to indicate the average number of persons they estimated to meet every day because of their job. Should this number be higher than 20, those specific respondents were asked to enumerate, over the course of the

2 surveyed days, only non-professional contacts. Detailed multivariate statistical analyses of how the number of reported close-range contacts varies as a function of respondents' socio-demographic features, meeting places and time can be found in Béraud et al. (2015). Here we descriptively focus on one specific, under-investigated aspect of the observed distribution of close-range contacts: their within- and between-group heterogeneity.

The distribution of contacts is shown in figure 1. The 2,033 individuals reported a total of 19,728 per-day close-range contacts (left panel). The median number of contacts is 8 whereas the average is approximately 9.5, with strong right-skewness. Respondents reporting a number of close-range contacts greater than twice (n=175) or even three times (n=36) the mean are not rare. This is even more visible when considering respondents who declared having an occupation especially exposed to social contacts (n=257) (figure 1's right panel). Overall they reported 14,971 additional job-related contacts. For those contacts the median is 30 whereas the average is approximately 58. One still finds 20 and 10 cases respectively at twice and three times the mean. The central tendencies of both distributions are consistent with those found in other contact surveys (Hoang et al. 2019: 727-728). In both cases, averages are clearly driven by a small fraction of individuals reporting (extremely) high numbers of short-range contact.

Other studies tend to discard outliers as they may heavily impact statistical estimation of central tendencies (see Béraud et al. 2015: 5, 9). We explicitly leave in outliers, as they are characteristic of the distribution of contact and are a crucial characteristic of empirical networks. It is not unusual for a small fraction of people to have jobs with very high frequencies of contact. The feature of high distributional skew is also visible in recent smaller-scale contact surveys conducted in China (Zhang et al. 2019; Zhang et al. 2020) and in a recent six-country survey permitting high self-reported values that range into the 4 and 5 digits. There are three suspicious cases where 999 was recorded. We judged it best to leave these values unaltered. The survey documentation does not identify these cases as missing values nor as maximum values due to a limit of 3 on the number of digits. While these values greatly impact the average level of contact, they do not make or break the skewed nature of the distribution. Accordingly, analyses (available upon request) that exclude 999 cases and upwardly adjust per-contact transmission probabilities accordingly to achieve the same R_0 show intervention effects that are qualitatively unchanged. Vice versa, analyses that substitute the 999 cases with higher numbers imputed following the shape of the distributional tail and downwardly adjust per-contact transmission probabilities produce similar intervention effects as well. To obtain a single measure of daily close contact for each respondent we simply combined the number of per-day contacts tracked in the two-day contact diary with the number of supplementary professional contacts estimated by respondents. This is the measure we use in section 3.1 to construct a synthetic network.
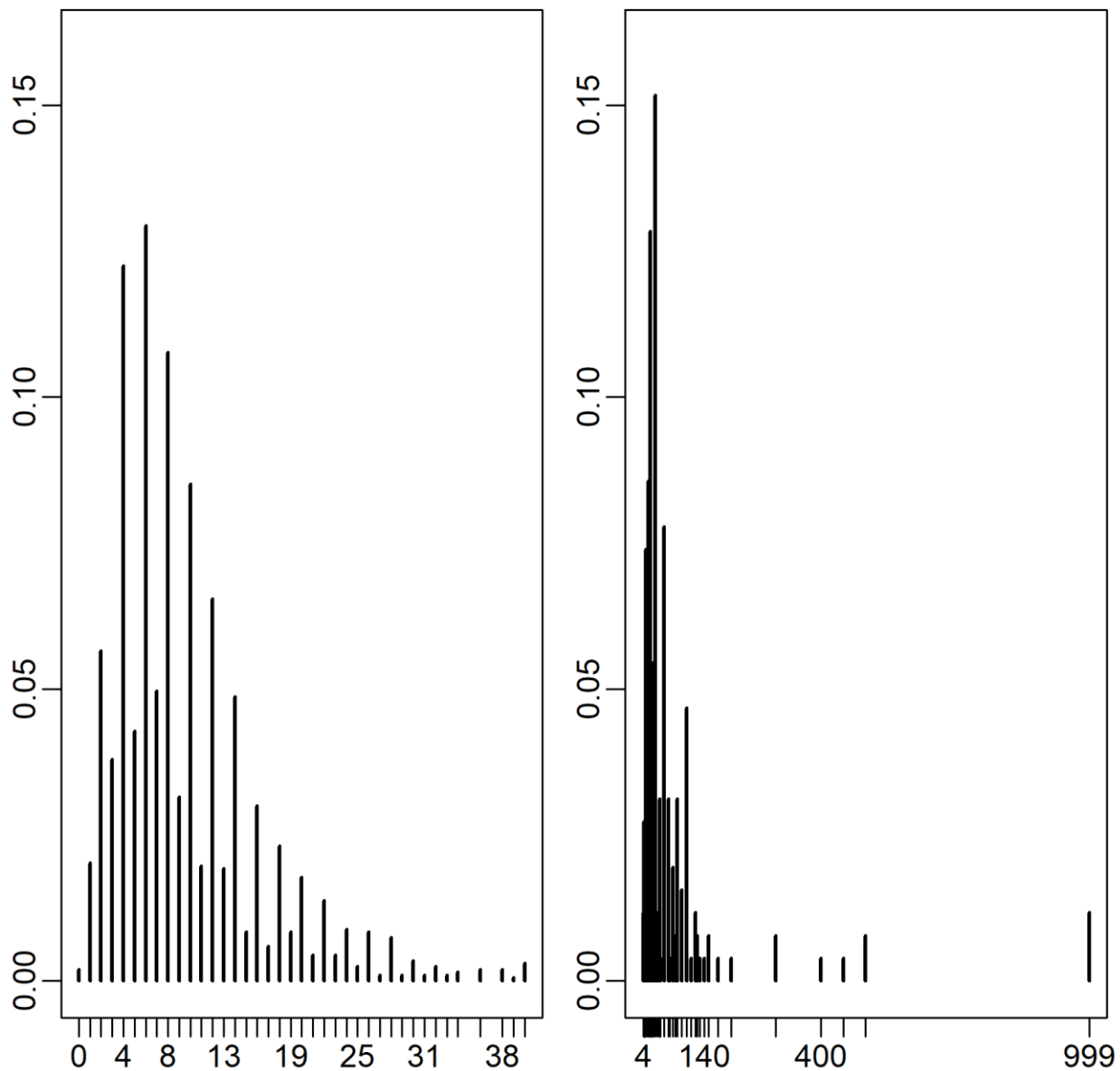
Figure 1. *Left*: Fraction of cases (y-axis) reporting a given number of close-range contacts (averaged over the two days) (x-axis) (n=2,033); *Right*: Fraction of cases (y-axis) reporting a given number of daily job-related contacts (x-axis) among respondents regarding their occupation as especially exposed to social contacts (n=257)

This high variability in the number of social contacts persists within major demographic categories that current interventions instead tend to treat as a whole. Figure 2's upper panel shows the distribution of per day self-reported close-range contacts by respondent's gender. Past multivariate analyses of these data found that women (mainly adult women) tend to have a higher average number of contacts than men (see

Béraud et al. 2015: 6 and table 1). Figure 2 shows that, behind this main mean effect, there is a large degree of variation within genders. For both men and women numbers of contacts (far) higher than the median occur frequently.
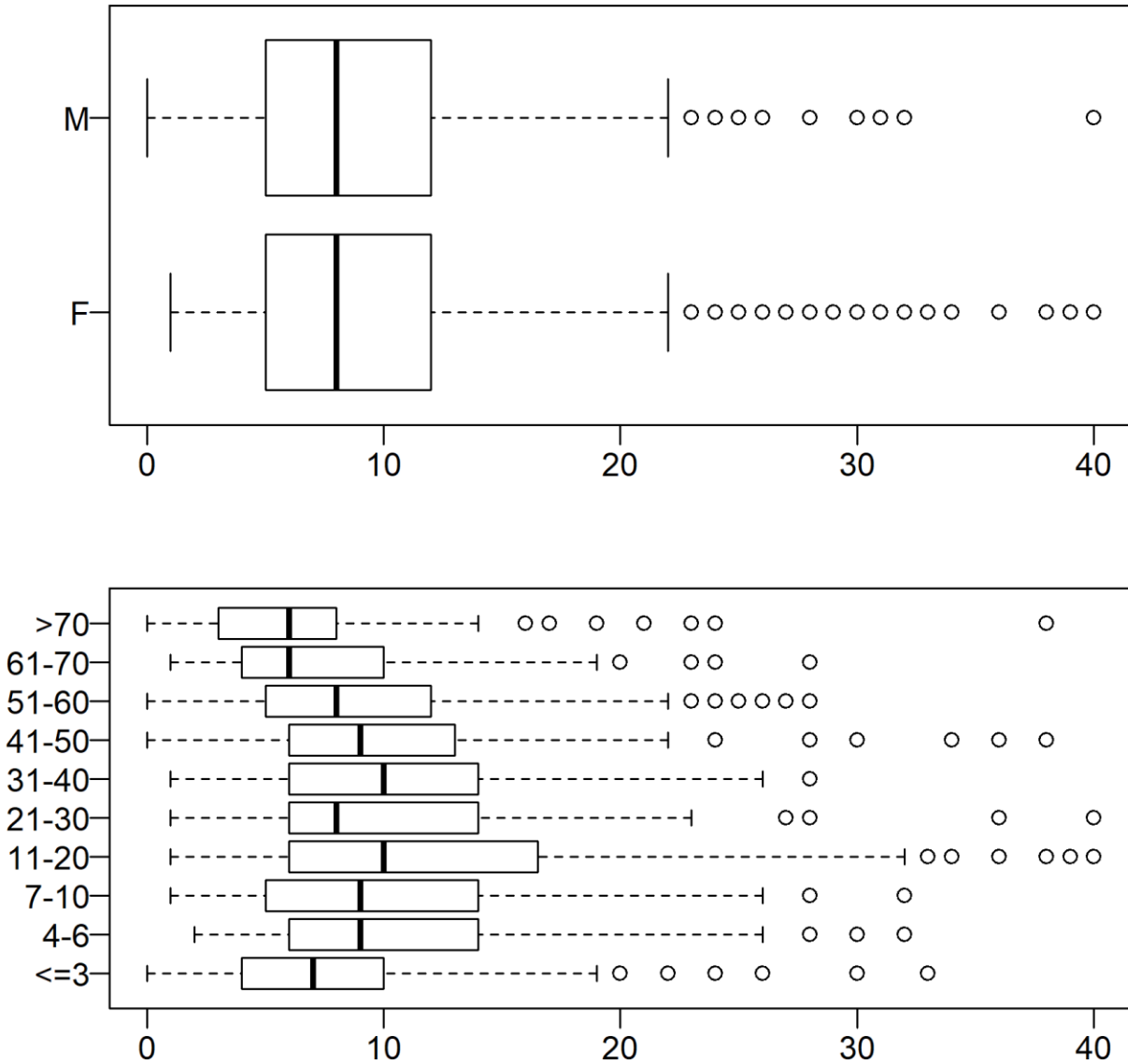

Figure 2. Distribution of self-reported per day close-range contacts (x-axis) by sex (top panel; F [1136], M [897]) and age groups (bottom panel;<=3 [240], 4-6 [169], 7-10 [196], 11-20 [276], 21-30 [155], 31-40 [109], 41-50 [135], 51-60 [195], 61-70 [357], >70 [201]) (y-axis).

Figure 2's bottom panel shows the distribution of close-range contacts per day by respondent's age. Age is the most recurrent variable used in epidemiological models to represent socially structured social

interactions. Age assortativity (and dissortativity at home) is found to be one of the most robust empirical regularities in epidemiological social contact surveys, as also found in multivariate analyses of the COMES-F data (see Béraud et al. 2015: 7-8). This motivates the use of average contacts per (more or less disaggregate) age-groups in age-structured compartment models (see, for some recent examples, Di Domenico et al. 2020; Roux et al. 2020; Salje et al. 2020: 3-4; in a comparative perspective, see Patrick et al. 2020). Now, apart from positive (more or less linear) effects of age on the likelihood of having more social contacts (which is indeed found in these data, see Béraud et al. 2015: table 1), figure 2's bottom panel again shows high variability within age-groups. Rather than averaging this variability, we propose to exploit it using the existence of high-contact individuals as leverage for effective intervention in viral diffusion dynamics.

Focusing the analysis on adult respondents in employment, we find a similar pattern for broad occupational groups (see figure 3). Among both "normal" respondents (top panel) and those regarding their occupation as especially exposed to social contacts (bottom panel), there is a great deal of within-group variation across individuals in the number of self-reported close-range contacts. High-contact individuals seem especially concentrated among high (e.g. elementary school teacher, teaching assistant) and low routine non-manual workers (e.g. bank teller, teller in public administration) and service class (e.g. university professor, politician, journalist or doctor). Data (available upon request) on an eleven-category variable describing respondents' occupational sector also suggest that high-contact individuals tend to concentrate in the service sector and non-manual routine labor, specifically shops, services to persons, education, health, and administration. However, while these data suggest that social contacts within occupations are much more dispersed than one could expect under a distribution symmetrically centered around the mean, COMES-F does not provide a detailed list of jobs. The appendix shows that in data from another survey with less precise measures of close social contacts, more fine-grained distinctions in professional categories can indeed account for a substantial portion of person-to-person variability in close-contact frequency.
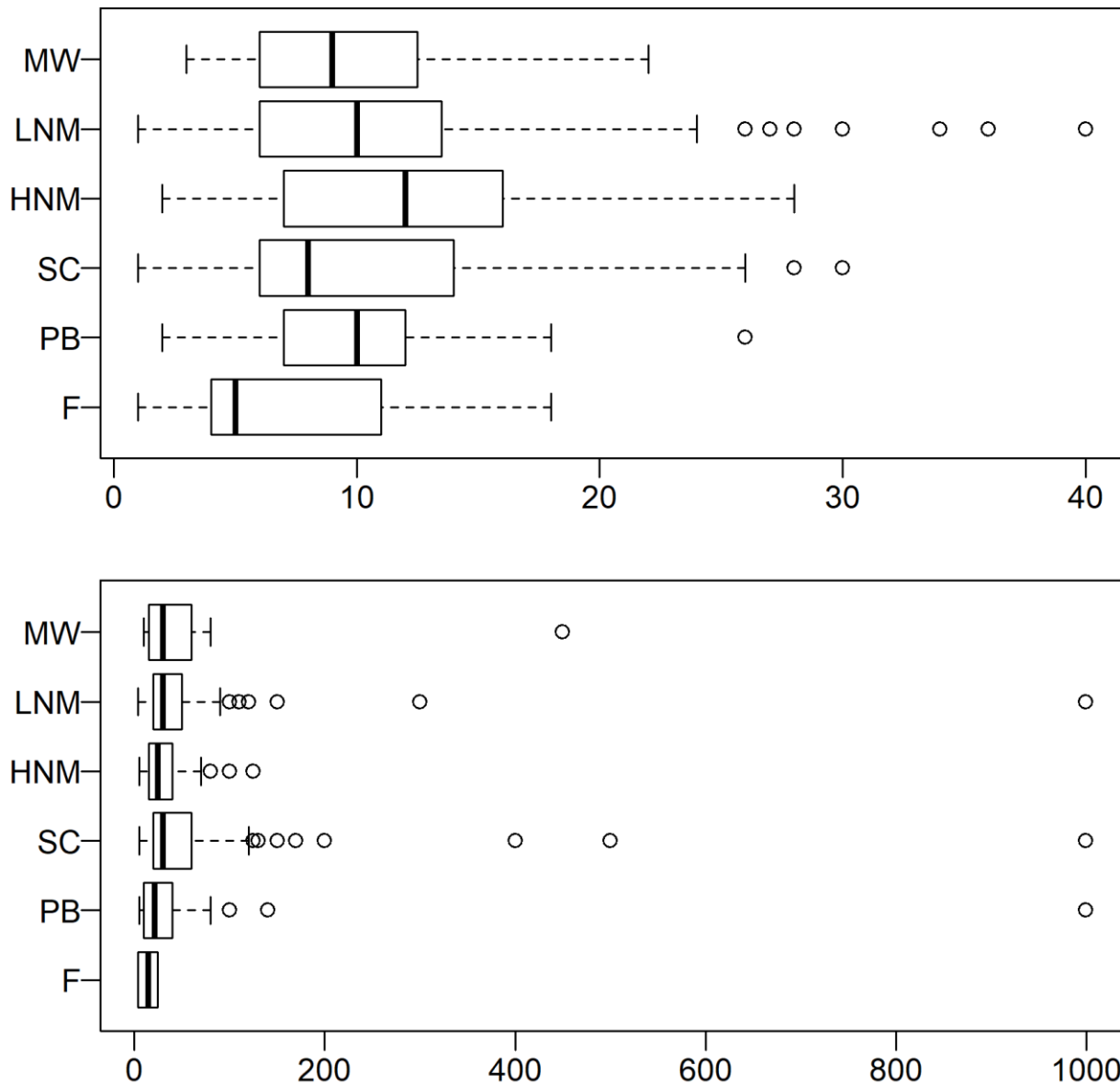
Figure 3. *Top*: Distributions of daily close-range contacts among employed respondents (n=436) by occupational category (y-axis); *Bottom*: Distribution of daily job-related contacts among respondents regarding their occupation as especially exposed to close-range contact (*n*=257) by occupation category (y-axis). Occupational categories (*n* in parenthesis, top panel first): F=farmers (n_{TOP}=14, n_{BOTTOM}=2), PB=petty bourgeoisie (craftsmen and shopkeepers) and entrepreneurs (n_{TOP}=26, n_{BOTTOM}=18), SC= Service class (managers, high-skilled administrators, intellectual, scientific and liberal professions) (n_{TOP}=123, n_{BOTTOM}=91), HNM=High routine non-manual worker (n_{TOP}=47, n_{BOTTOM}=41), LNM=Low routine non-manual workers (n_{TOP}=183, n_{BOTTOM}=93), MW=manual workers (43,n_{BOTTOM}=12).

3. Model

Using COMES-F survey data we build an agent-based computational model in which the synthetic network through which the virus diffuses is calibrated on the French contact data (for other work using empirical network data in agent-based diffusion models, see Smith and Burrow 2018; Manzo et al. 2018). Our aim is to study the macroscopic consequences of cross-individual variability in close-range contact frequencies empirically observed in France and assess whether this variability can be exploited for effective intervention in the ongoing epidemic. As such, we simulate a population the size of the COMES-F sample from which we eliminate four respondents who reported no close-range contacts. (We find that populations much larger than our sample produce qualitatively similar results but naturally take much longer to simulate.) We now explain how we connect these 2,029 agents (section 3.1), present the transmission model (3.2), and describe the interventions we simulate (3.3).

3.1 Network construction and features

We connected the agents according to two social network models. The first, which we will refer to as the "empirical" model, is the focus of our simulation; the second, which we will refer to as the "random" model, constitutes a benchmark against which to compare dynamics and effects of interventions.

In the *empirical* network model, agents are first given a degree (number of network ties) precisely equal to the number of per day close contacts reported by each respondent in the survey (see figure 1). Then, to connect agents to one another, we adapted the configuration model, an algorithm that was proposed to generate random networks with arbitrary degree distributions (for a discussion, see Jackson 2008: 83-85). To avoid duplicate links and self-links while ensuring an exact match between each virtual agent and an empirical respondent, we considered source agents in descending order of the to-be-generated degree rather than in random order, and then randomly picked available destination agents. Every time a connection was made, the degree of the two newly connected agents naturally increased by one. As soon as an agent reached the to-be-generated degree, it was excluded from the search algorithm. We found this procedure to always converge, achieving the intended empirical degree distribution.

Under the random mixing assumption common to epidemiological models, social contacts are assumed to happen at random within certain categories (usually age-groups): only the (within-category) average number of contacts is of interest. From a network perspective, this amounts to postulating a random network where contact probabilities across individuals have little variability and the degree of each node can be assumed being well approximated by the average degree (Newman 2002; Barthélemy et al. 2005). We therefore also study an Erdős–Rényi random graph with the average degree observed in the survey, as a benchmark distribution. This *random* network is characterized by low variability in contact across agents.

Table 1 shows network statistics computed over 100 realizations of the two networks. By construction, the empirical network has the precise mean, median and standard deviation of the degree distribution observed from the survey. The degree distribution of the empirical network has strong right-skewness, with the mean well above the median. By contrast, the degree distribution of the random network has essentially equal mean and median. The dispersion of the degree also strongly differs between the two networks, with the empirical network exhibiting variation in the nodal number of links that is five times higher than in the random network.

Table 1 also shows several other network statistics. We cannot calibrate or verify these network features, as the survey data lack data on connections between respondents' contacts. Nonetheless, the empirical network generated by our modified configuration model shares various features with social networks studied elsewhere (for a detailed discussion, see Appendix A2).

Network	Average degree	Median degree	Stdev degree	Degree corr	Clustering coef	Deg-clust corr	Av path length	Diameter
Empirical	23.99 (0.000)	16 (0.000)	48.01 (0.047)	-0.15 (0.001)	0.14 (0.003)	-0.26 (0.005)	2.22 (0.002)	4.00 (0.000)
Random	23.99 (0.155)	23.96 (0.196)	4.86 (0.082)	-0.00 (0.005)	0.01 (0.000)	-0.00 (0.023)	2.74 (0.003)	4.00 (0.000)

Table 1. Topological features of the simulated contact networks. Mean values across 100 network realizations (standard deviation in parentheses). Degree corr=Pearson correlation coefficient computed over the degrees of all pairs of linked nodes; Clustering coef=clustering coefficient; Deg-clust corr=Pearson correlation coefficient between nodes' degree and their clustering coefficient; Av path length=Average of the shortest path lengths; Diameter=Maximum of the shortest path lengths.

3.2 SEIR model

We model disease propagation through the network by following the logic of compartmental models, which have been previously applied to the COVID-19 outbreak (Brethouwer et al. 2020; Kucharski et al. 2020; Qun et al. 2020; Prem et al. 2020). In particular, we adopt a SEIR model (Martcheva 2015) and follow recent French studies (and parametrization therein, see Salje et al. 2020a, b) to determine how agents unidirectionally move from being (S)usceptible, to (E)xposed, (I)nfected, and eventually (R)ecovered. In

particular, agents are supposed to stay within a given state as many iterations as days current empirical studies show real-world individuals remain within those states. One iteration thus represents one day.

Upon infection, agents first enter E where they stay 4 days; during this period, they are not infectious (see Salje et al. 2020a: 10). They then move to I where they become infectious, and can contaminate other agents over the course of 4 days (see Salje et al. 2020a: 10). Infected agents move to R with probability following a normal distribution with average 0.993 (and possible range at the agent-level between 0.990 and 0.996) (see again Salje et al. 2020b) provided they spent a number of days in I at least equal to a given recovery time. The recovery time follows a Poisson distribution centered on 2 weeks (with possible range at the agent-level between 1 and 6 weeks) (for these values, see empirical estimates in World Health Organization 2020: 14).

We combine this basic compartment structure with network topology such that the agents an infectious-infected agent can infect are determined by the network of close-range contact (see Barrat et al. 2008: ch. 9). During each day, an infectious-infected agent can only transmit the disease to its direct contacts. The agent-to-agent per-day-per-contact transmission probability is assumed to be normally distributed around 0.0300 or 0.0186 (and possible range at the agent-level between 0.0005 and 0.12). This parameterization at the dyadic level generates at the aggregate level an average basic reproduction rate R_0 equal to 2.9 and 1.785 respectively. The value 2.9 corresponds to the empirically estimated R_0 in France before lockdown (see Salje et al. 2020b): we adopt it to assess whether the model may have generated an epidemic in the absence of any intervention that is in line with existing counterfactual studies (figure 4). The 1.785 value corresponds to the average between pre- and post-lockdown estimates (the latter being 0.7, see Salje et al. 2020b). We switch to 1.785 to study the impact of targeted interventions in a scenario after partial lifting of the French lockdown on May 11. It represents a reduction in R_0 down from the original pre-lockdown level of 2.9 reflecting post-lockdown continuation of general policy measures such as social distancing, group size restrictions and sanitary improvements. At $R_0 = 1.785$, further person-specific interventions are needed to bring down the effective R_0 and mitigate a second wave. We study how the targeting of these interventions impacts their effectiveness in doing so.

All the simulations we performed start with five (randomly chosen) initially infectious infected agents. Given the size of our model population, this is the lowest number of seeds that prevents excessive variability across simulation trials. Proportionally to the size of the French population, this number at 160000 roughly approximates, for our basic testing scenario (i.e. $R_0=2.9$), estimates of infected cases around March 7 (see Salje et al. 2020b: figure 3E); on the other hand, for our intervention scenario (i.e. $R_0=1.785$), this number of initial seeds would roughly correspond to targeted interventions commencing a week after partial lockdown release, with $R_0=1.785$ having allowed infections to climb back up by a factor of ten from current levels at about 4000 new cases per day (see Salje et al. 2020b: figure 3G). In simulations on larger

networks, which take much longer to run, in which we implemented the same degree distribution and used the same number of seeds (smaller fraction), we find that peaks naturally occur later, while the interventions we present next show qualitatively the same relative effects.

Parameter	Average Value	Ref
R_0	2.9 (1.785)	Salje et al. 2020b
Dyadic transmission probability	0.0300 (0.0186)	$$\frac{R_0}{Av\ Degree\ \times IE}$$
Exposed time	4 days	Salje et al. 2020a
Infectiousness length (IE)	4 days	Salje et al. 2020a
Recovery time	2 weeks	WHO 2020
Recovery chance	0.993	Salje et al. 2020b

Table 2. Covid-19 SEIR model parameters

3.3 Interventions

We consider a scenario in which France progressively exits the strict March-17/May-11-2020 lockdown and has a fixed daily (medical / technological / financial / ethical) capacity to immunize a portion of susceptible/infected individuals. In the absence of a vaccine, "immunization" describes an effort equivalent to the medical act of 100% effective vaccination. This could be any combination of preventive interventions, such as medical testing, quarantining-if-positive, protective measures in high-risk professions and targeted informational campaigns. Accordingly, we assume that the French government has a daily budget b with which it can immunize b agents, preventing future infection of these and / or spread from these targeted agents to other agents. On day 1, spread from b agents is permanently prevented, on day 2 spread from an additional b agents is permanently prevented, and so on. This is implemented as follows: Each iteration, representing one day, b S, E or I agents are selected and moved to R.

We study four budgets for each intervention: $b = 1, 3, 5,$ and 10. We consider three methods for selecting agents for intervention. The first method, "NO-TARGET", simply randomly samples b agents for immunization each day, and is intended as a benchmark against which to contrast the other two methods. The second method, "CONTACT-TARGET", follows the strategy described in Cohen & Havlin (2010),

whereby each day *b* random agents are sampled who each select one random contact (without replacement) for immunization. Because of the friendship paradox (Feld 1991), these targets have above-average expected degree (Galvani et al. 2005; Christakis & Fowler 2010; Kitsak et al. 2010). This is so because high-degree nodes are by definition overrepresented among other nodes' contacts (Feld 1991). The CONTACT-TARGET strategy is implementable in practice as a government could in fact randomly sample from the known population and have sampled individuals suggest their contacts. The third method, "HUB-TARGET", assumes that agents' numbers of contacts are perfectly observed. During each iteration, nodes are targeted in strictly decreasing order of their network degree, starting with the *b* largest hubs.

4. Results

Figure 4 shows the number of concurrently infected individuals over time for the two networks when no interventions are taken, and R_0 is 2.9, the pre-lockdown estimated value early March in France. The epidemic peaks 13 days earlier in the empirical network than in the random network, day 30 vs. day 43, demonstrating the impact of hubs: Highly connected individuals are more likely connected to the seeds and their neighbors. Once infected, they expose others early on, thus catalyzing viral diffusion. In the ER network, by contrast, there are no hubs to accelerate spread.

If we translate the simulated peaks in real-world ICU admissions proportionally to the size of French population (see appendix A4 for the details), the model generates, under both network regimes, a massive number of cases requiring hospitalization. The order of magnitude of this sub-population, between 100,000 and 200,000 individuals, is consistent with other studies that have counterfactually estimated COVID-19 spread in France without any intervention between early March and Mid-April (see, in particular, Roux et al. 2020). Having established that the model produces an epidemic of appropriate magnitude, timing and shape, we move on to our post-lockdown scenario to assess the effect of interventions under a lower average reproduction rate (1.785) which allows to simulate a second wave of about half of the fist wave's size by mid-June (figure's 4 H=824 *versus* table 3's H=470).
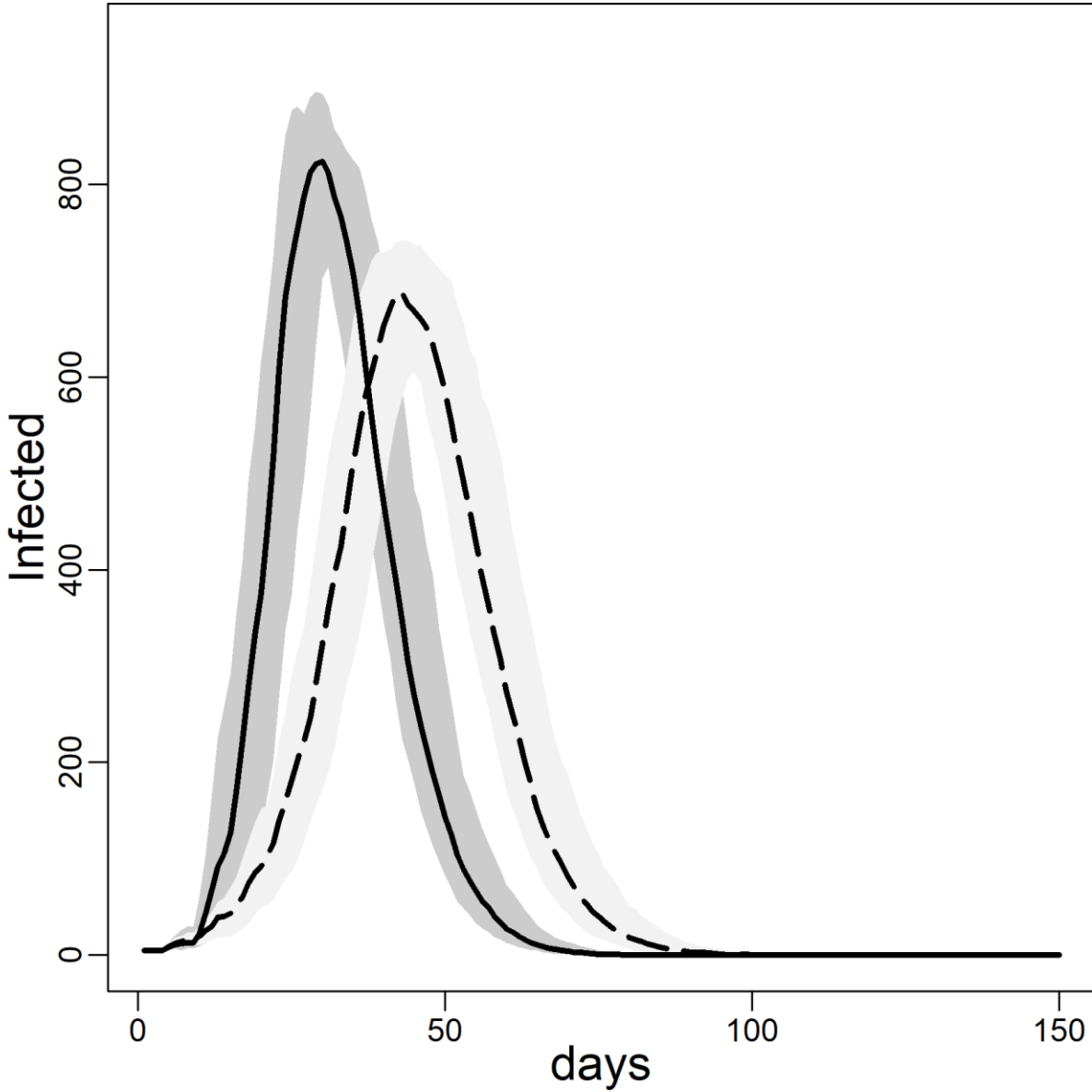
Figure 4. Number of infected agents (y-axis) by days (x-axis) (median of 100 replications). Lower and upper bounds of the shaded areas correspond to the 5th percentiles and 95th percentiles of the 100 replications. n = 2,029 agents. R_0 = 2.9. Solid line: Network with French long-tailed degree distribution (H = 824 [703; 894]; ICU_{FRANCE} = 180274 [153801; 195588]; T = 30); dashed line: Erdős–Rényi network with the same average degree as the empirical network (H = 688 [578; 742]; ICU_{FRANCE} = 150520 [126454; 162334; T = 43).

Figure 5 shows the impact of the three intervention methods on viral diffusion in the empirical network. Peak reductions and delays are reported in table 3. Spread under intervention regimes is displayed as dashed curves in figure 5. Solid curves represent the no-intervention scenario, for contrast. Panel A

shows results for the NO-TARGET procedure, whereby each day b randomly selected susceptible, exposed or infected agents are immunized. The NO-TARGET procedure's maximally achievable impact, using the most generous budget considered, $b = 10$, corresponding to 10% of the population being immunized during the first 20 days, leaves the peak at 69% (Table 3: 326 / 470) of what it would have been without any intervention. Also, the peak is not delayed.

Panels B show results for the CONTACT-TARGET procedure, which assumes that no global information on connectivity is available. Lacking this information, it attempts to find high-degree nodes by drawing a random sample of agents with unknown degree and selecting a random neighbor of each sampled agent for immunization. This procedure is much more effective than the NO-TARGET intervention, as the figure shows. A budget of $b = 3$ CONTACT-TARGET immunizations produces an impact comparable to a NO-TARGET immunization regime with $b = 10$ daily immunizations. At $b = 10$, the CONTACT-TARGET intervention achieves a reduction down to only 5% of the peak in the no-intervention scenario (Table 4: 25 / 470). The peak occurs two days earlier, after 30 days (CONTACT-TARGET) instead of 32 days (NO-TARGET).

The CONTACT-TARGET method would be more effective if randomly chosen agents would be able to select a random neighbor for immunization among relatively high degrees at a higher chance than network structure per se allows. Survey data suggest that targeting of certain professions may also effectively identify high-degree agents (see appendix A1). To evaluate the maximally achievable impact of any degree-based intervention, panels C of figure 5 show the impact of the HUB-TARGET policy, whereby each day the b non-immune agents with highest degree are immunized. Even a single agent per day ($b = 1$) reduces the peak down to only 25% (Table 2: 118 / 470) and delays it by 17 days. This peak reduction exceeds what NO-TARGET immunization achieves with ten agents per day ($b = 10$). Three agents per day accelerates the peak again, but this happens as peak height is reduced to only 4%. With five agents (5% of the population after 20 days) virus spread is effectively halted. In other words, if one fully protects a small vulnerable fraction of the population, one can prevent nearly all infections that would have otherwise occurred.
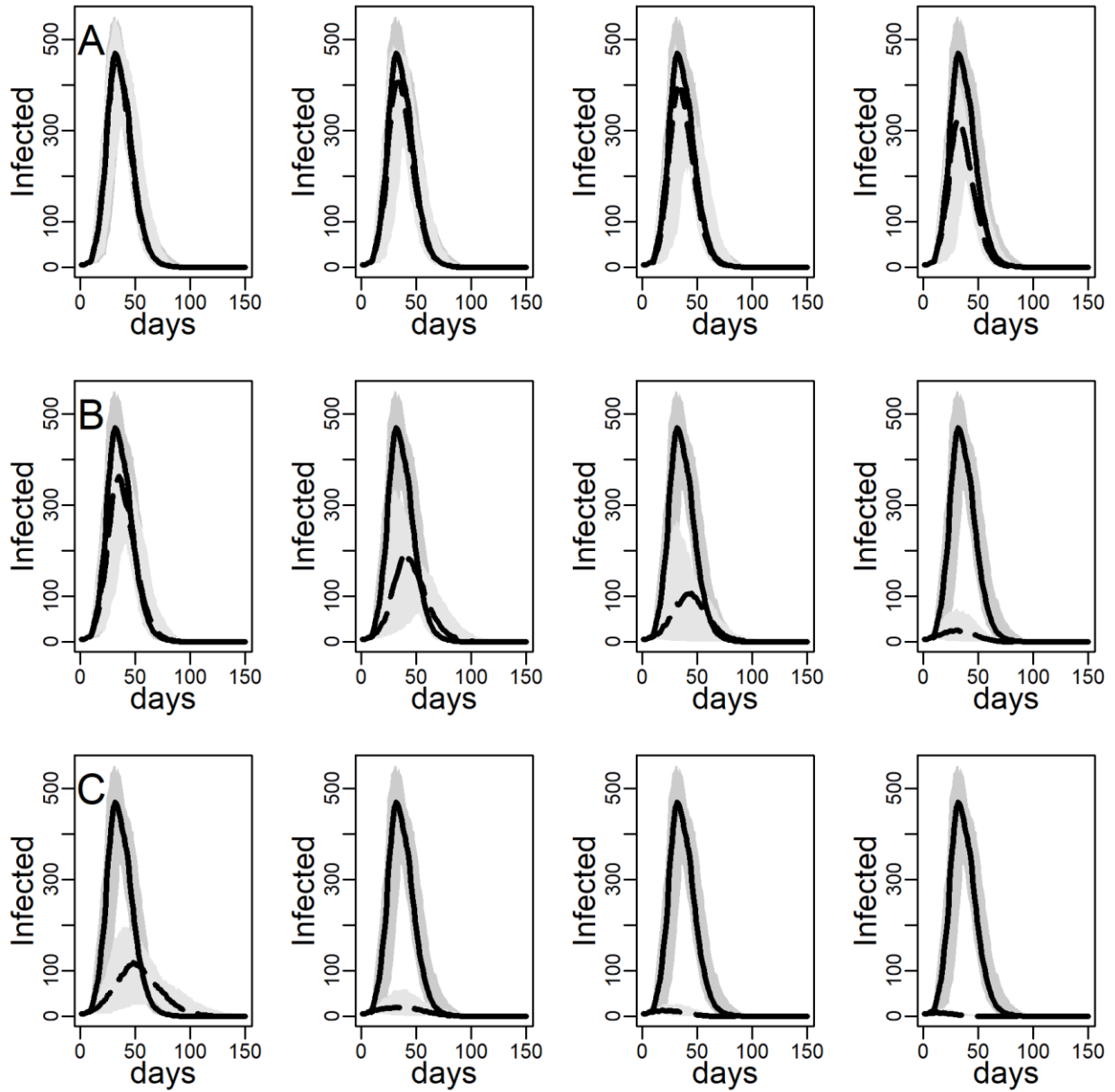
Figure 5. Number of infected agents (y-axis) by days (x-axis) (median of 100 replications) under three different interventions (rows) targeting 1, 3, 5, or 10 agents per day (columns). A – NO-TARGET immunization; B – CONTACT-TARGET immunization. C – HUB-TARGET immunization. Lower and upper bounds of the shaded areas correspond to the 5th percentiles and 95th percentiles of the 100 replications. Solid line: *empirical* network; dashed line: interventions. $n = 2,029$ agents. $R_0 = 1.785$.

Empirical network							no intervention: H = 470 [203; 549]; T = 32	
		b = 1		*b* = 3		*b* = 5		*b* = 10
	peak height	time	peak height	time	peak height	time	peak height	time
NO-TARGET	447 [203; 539]	31	407 [173; 470]	34	391 [72; 475]	32	326 [88; 411]	32
CONTACT-TARGET	362 [153; 442]	35	195 [38; 278]	41	106 [45; 164]	45	25 [1; 72]	30
HUB-TARGET	118 [25; 182]	49	20 [1; 56]	32	13 [3; 25]	18	9 [4; 16]	13

Table 3. Peak height (maximum # concurrently infected agents) and time (in days) under three interventions (rows) and four budgets (column) on the *empirical* network. Shown are median, 5% and 95% percentiles across 100 iterations. R_0=1.785.

5. Implications for contemporary policy

According to estimates available on 11 May 2020, when lockdown progressively started to be eased in France, only between 2.8% and 7.2% of the French population had acquired immunity against COVID-19 (Salje et al. 2020b). This would mean that the vast majority of the population is still susceptible. As a consequence, in absence of a vaccine, many observers consider the only alternative to be massive testing, possibly coupled with digital contact tracing (Sustained Suppression, *Nature Biomedical Engineering*, 2020).

Contemporary French post-lockdown policy is indeed based on this strategy, and policy in several other European countries is similar. French government and health authorities ask persons with symptoms to call their city doctor, who should prescribe a free-of-charge test. Should the person be positive, the patient is asked to describe all her/his close contacts during the last seven days; each is also asked to be reached by phone, interviewed, and, if established that contact with the focal individual took place, this contact-case too is tested and quarantined if positive (Info Coronavirus COVID-19, 2020). The government early May 2020 claimed to be able to perform up to 10,000 tests per day (which amounts to 0.015% of the French population). This advertised capacity is widely questioned (France Info 2020). Even if this capacity were available, it still remains below the agreed effective fraction of the population to be daily tested, i.e. at least 2% (Humanity tested, *Nature Biomedical Engineering*, 2020). Moreover, as acknowledged by experts

advising the French government (Conseil Scientifique Covid-19, 2020: 29), this strategy requires unrealistic levels of human resources, in particular in case of citizen resistance to and suspicion of contact tracing, as is the case in France (Casilli et al. 2020). Last but not least, the proposed strategy focuses on self-reported symptoms, which exclude asymptomatic cases. Asymptomatic cases are thought to be a non-negligible fraction of COVID-19 infected persons, especially among children (Heneghan et al. 2020).

Our results suggest a third way in addition to contemporary resource-intensive testing and contact tracing strategies and future mass vaccination. This alternative strategy is based on four pillars: First, interventions should concern both susceptible and infected persons; Second, in both cases, given current practical limitations, search over randomly selected portions of the population can fruitfully replace exhaustive search; Third, sampling should be performed in such a way that high-contact persons are more likely selected for intervention; Fourth, high-contact persons should be "treated" with priority. The rationale behind the third and fourth pillar is that high-contact individuals, if susceptible, are simultaneously at higher risk of being contaminated because of their high social exposure, and, when they are infected, also transmit the virus to a larger fraction of the population.

Our model results suggest this strategy can be effective. When a virus spreads through a network that includes individuals with dramatically more close-range contacts than the average (as observed in France), the CONTACT-TARGET method (which attempts to find high-degree nodes by drawing a random sample of agents with unknown degree and selecting one random neighbor of each sampled agent for immunization) can reduce the concurrent number of infected agents by a factor of 1.3, 2.4, 4.4 or 18.8 depending on whether one samples 1, 3, 5 or 10 agents per iteration (our budget "b"). If simulated peak heights are transformed into ICU admissions proportionally to French population (see appendix, table A4, and computational details therein), this would amount to reducing ICU admissions from 102826 without intervention to 79197, 42661, 23191 or 5469 respectively (5000 ICU beds being normal French national capacity). If b-values are also expressed proportionally to the French population, our CONTACT-TARGET method leads to the largest ICU-admission reduction by targeting approximately 320000 individuals per day. In contrast, for the same budget, the NO-TARGET method (which searches agents for immunization at random within the population) can generate at best a peak reduction of a factor of 1.4. Thus, keeping budget constant, systematic targeting of high-contact individuals is much more effective than random search of average-contact persons.

The sampling burden could be further reduced if one efficiently exploits the empirical observation that high-contact individuals tend to be concentrated within certain occupations (see appendix A1). Following this logic, the sampling person may be asked preferentially to report her/his social contacts

within a given list of highly socially exposed professions. Under this condition, our CONTACT-TARGET method would get closer to the HUB-TARGET one. With the latter method a sample of approximately 95000 or 160000 individuals per day would suffice to reduce the peak, and corresponding ICU admissions, under the level of normal French hospital capacity (i.e. 5000 ICU admissions).

Our simulations are silent about the specific content of the actions to be performed on each finally selected individual. We only provide a method to maximize the efficiency of that selection. However, we do believe that, in the absence of a vaccine, and given the relative scarcity of tests, the skeleton of a sustainable recipe could be as follows: a) if the selected high-contact person is symptomatic, directly quarantine; b) if the selected high-contact person is not symptomatic, test with probability equal to the daily number of available tests divided by the number of daily sampled people. In both cases, the selected high-contact person should receive a targeted informational message explaining why, given her/his social exposure, it is more crucial for her/him than for the "average" person to adopt basic protective gestures during all his social and professional interactions. The feasibility of this strategy thus comes from the fact that it relies more on actions, i.e. targeted messages, that can be performed at distance and does not require mobilizing huge human resources across a country. This is in line with recent calls for approaching COVID-19 in terms of massive, but targeted, preventive interventions based on pedagogic actions rather than on actions constraining individuals' freedom, which have proven not to be effective in the long-run for other aggressive infectious diseases (Marcus 2020).

6. Discussion

As countries exit the Covid-19 lockdown many have limited capacity to prevent flare-ups of the coronavirus. With medical, technological, and financial resources to prevent infection of only a fraction of its population, which individuals should countries target for testing and tracking? Together, our results suggest that targeting individuals characterized by high frequencies of short-range contacts dramatically improves the effectiveness of interventions. An additional known advantage of targeting hubs with medical testing specifically is that they serve as an early-warning device that can detect impending or unfolding outbreaks (Christakis & Fowler 2010; Kitsak et al. 2010).

This conclusion is reached by moving away from the standard compartmental models that rely on random mixing assumptions toward a network-based modeling framework that can accommodate person-to-person differences in infection risks stemming from differential connectedness. The framework allows us to model rather than average out the high variability of close-contact frequencies across individuals observed in contact survey data. Simulation results show that consideration of realistic close-contact

distributions with high skew strongly impacts the expected impact of targeted versus general interventions, in favor of the former.

A key limitation of our approach is that we have not considered the duration of close-range contact. A negative relationship may exist between the number of close-range contacts and the per-contact transmission probability if the per-contact duration decreases in the number of contacts. In this case our results could overstate the importance of high-contact individuals and the effectiveness of the targeted intervention strategies we have proposed. Too little is still known about Covid-19 transmission through droplet "clouds" (see Bourouiba 2020; Mittal 2020) to say anything definite in this direction. Nonetheless, limited evidence on super-spreading events involving large crowds interacting in close range (for an overview, see Kay 2020) suggest that the risk posed by large numbers of interactions with many people is not necessarily mitigated by the brevity of such contacts.

How could public policy effectively target high-contact individuals? We propose two concrete methods. First, one could test and track workers in professions characterized by high frequencies of contact. Survey data show that some professions involve ten times as much close-range contact than others, with elementary school teachers and cashiers topping the list. The occupational categories used in the survey follow a common international standard used by the US Bureau of Labor Statistics. Legislation could straight-forwardly be set on its basis. A second method is entirely agnostic of who the high-degree individuals are and targets random acquaintances of random individuals, who statistically have high expected degree. In our simulations this second method shows to be effective yet at the same time is conservative in assuming no knowledge of degree or use thereof. If individuals nominated contacts they knew to have many other contacts, the difference targeted intervention could make would be even greater.

References

Adamic, L. A., & Huberman, B. A. (2002). Zipf's law and the Internet. *Glottometrics*, *3*(1), 143-150.

Albert R., A.-L. Barabási (2002). Statistical mechanics of complex networks, *Review of Modern Physics*, 74, 47-97.

Barabási, A-L. (2014). *Linked*, New York, Basics Books.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509-512.

Barrat Alain, Marc Barthélemy, Alessandro Vespignani, *Dynamical Processes on Complex Networks*, Cambridge, Cambridge University Press, 2008.

Barthelemy Marc, Alain Barrat, Romualdo Pastor-Satorras, Alessandro Vespignani (2005. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks, *Journal of Theoretical Biology*, 235, 275–288.

Belot, Michèle, Syngjoo Choi, Julian C. Jamison, Nicholas W. Papageorge, Egon Tripodi, Eline van den Broek-Altenburg. 2020. *Six-Country Survey on Covid-19*. Unpublished manuscript.

Béraud, G., Kazmercziak, S., Beutels, P., Levy-Bruhl, D., Lenne, X., Mielcarek, N., ... & Dervaux, B. (2015). The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. *PloS one*, *10*(7).

Bourouiba L. (2020). Turbulent Gas Clouds and Respiratory Pathogen Emissions: Potential Implications for Reducing Transmission of COVID-19, *JAMA*, 323(18):1837–1838.

Brethouwer, J. T., van de Rijt, A., Lindelauf, R., & Fokkink, R. (2020). "Stay Nearby or Get Checked": A Covid-19 Lockdown Exit Strategy. *arXiv preprint arXiv:2004.06891*.

Casilli A., Dehaye P-O, Soufron J-B (2020), StopCovid est un projet désastreux piloté par des apprentis sorciers, *Le Monde*, 25 avril 2020.

Heneghan C., Brassey J., Jefferson R. (2020), COVID-19: What proportion are asymptomatic https://www.cebm.net/covid-19/covid-19-what-proportion-are-asymptomatic/

Christakis, N. A., & Fowler, J. H. (2010). Social network sensors for early detection of contagious outbreaks. *PloS one*, *5*(9).

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, *51*(4), 661-703.

Cohen, R., & Havlin, S. (2010). *Complex networks: structure, robustness and function*. Cambridge University Press.

Conseil scientifique COVID-19 (2020), Sortie progressive de confinement: prerequis et mesures phares, avis n° 6, 20 avril 2020

(https://solidarites-sante.gouv.fr/actualites/presse/dossiers-de-presse/article/covid-19-conseil-scientifique-covid-19).

Dezső, Z., & Barabási, A. L. (2002). Halting viruses in scale-free networks. *Physical Review E*, *65*(5), 055103.

Di Domenico Laura, Giulia Pullano, Chiara E. Sabbatini, Pierre-Yves Boëlle, Vittoria Colizza (2020) Expected impact of reopening schools after lockdown on COVID-19 epidemic in Île-de-France (Report#10) www.epicx-lab.com/covid-19.html

Feld, S. L. (1991). Why your friends have more friends than you do. *American Journal of Sociology*, *96*(6), 1464-1477.

Feser, E. J. (2003). What regions do rather than make: A proposed set of knowledge-based occupation clusters. *Urban Studies*, *40*(10), 1937-1958.

France Info (11 May 2020). Dépistage du Covid-19 : pourquoi la France est encore loin de l'objectif de 700 000 tests virologiques par semaine (https://www.francetvinfo.fr/sante/maladie/coronavirus/)

Galvani, A. P., & May, R. M. (2005). Dimensions of superspreading. *Nature*, *438*(7066), 293-295.

Hoang, T., Coletti, P., Melegaro, A., Wallinga, J., Grijalva, C. G., Edmunds, J. W., Hens, N. (2019). A systematic review of social contact surveys to inform transmission models of close-contact infections. *Epidemiology*, *30*(5), 723-736.

Info Coronavirus COVID-19 - Stratégie de déconfinement. Gouvernement.fr https://www.gouvernement.fr/infocoronavirus/ strategie-de-deconfinement.

Jackson M. O. (2008) *Social and Economic Networks*, Princeton University Press, Princeton.

James, A., Pitchford, J.W., & Plank, M.J. (2007). An Event-Based Model of Superspreading in Epidemics, Proc. Biol. Sci. Proceedings B, 274(1610), 741–747. https://doi.org/10.1098/rspb.2006.0219.

Jones, J. H., & Handcock, M. S. (2003). Sexual contacts and epidemic thresholds. *Nature*, *423*(6940), 605-606.

Johnson, S, Torres, JJ, Marro, J, Munoz, MA Entropic Origin of Disassortativity in Complex Networks *Physical Review Letters*, 104, 10: 108702.

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature physics*, *6*(11), 888-893.

Kucharski, Adam J., Timothy W. Russell, Harlie Diamond, and Yang Liu et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 2020.

Li, Qun, Xuhua Guan, and Wu Peng et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *The New England Journal of Medicine*, 382(13):1199–1207, 2020.

Liljeros, F., C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg (2001). The web of human sexual contacts. *Nature* 411 (6840), 907–908.

Little, S. J., Pond, S. L. K., Anderson, C. M., Young, J. A., Wertheim, J. O., Mehta, S. R., ... & Smith, D. M. (2014). Using HIV networks to inform real time prevention interventions. *PloS one*, *9*(6).

Kay, Jonathan (2020). COVID-19 Superspreader Events in 28 Countries: Critical Patterns and Lessons, *Quillette* (April 23, 2020) (https://quillette.com/2020/04/23/)

Manzo, G. (2020). The Role of Complex Social Networks is Lost in Dominant Epidemic COVID-19 Models*, Sociologica*, 14, 1.

Manzo G., Gabbriellini S., Roux & V., M'Mbogori F. N. (2018). Complex Contagions and the Diffusion of Innovations: Evidence from a Small-N Study, *Journal of Archaeological Method and Theory*, 25, 4, 1109–1154

Martcheva, M. (2015). *An introduction to mathematical epidemiology* (Vol. 61). New York: Springer.

Marcus, J. (2020) Quarantine Fatigue is Real, *The Atlantic*, May 11 2020.

Mittal, R., Ni, R., & Seo, J. (2020). The flow physics of COVID-19. *Journal of Fluid Mechanics, 894*, F2. doi:10.1017/jfm.2020.330

Mones, E., Stopczynski, A., Pentland, A. S., Hupert, N., & Lehmann, S. (2018). Optimizing targeted vaccination across cyber–physical networks: an empirically based mathematical simulation study. *Journal of The Royal Society Interface*, *15*(138), 20170783.

Newman M. E. J. (2002). The spread of epidemic disease on networks. *Physical Review E*, 66, 1, 1095-3787.

Newman M. E. J. (2003a). Mixing patterns in networks, *Physical Review E* 67, 026126.

Newman M. E. J. (2003b). The Structure and Function of Complex Networks *SIAM Review*, 45, 2, 167–256.

Pastor-Satorras, R., & Vespignani, A. (2002). Immunization of complex networks. *Physical review E*, *65*(3), 036104.

Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., ... & Abbott, S. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*.

Prem K, Cook AR, Jit M (2017) Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Comput Biol 13(9): e1005697.

Roux, J., Massonnaud, C. & Crépey, P. COVID-19: One-month impact of the French lockdown on the epidemic burden. http://medrxiv.org/lookup/doi/10.1101/2020.04.22.20075705 (2020).

Sapiezynski, P., Stopczynski, A., Lassen, D. D., & Lehmann, S. (2019). Interaction data from the Copenhagen Networks Study. *Scientific Data*, *6*(1), 1-10.

Salje, Henrik, Cécile Tran Kiem, Noémie Lefrancq, Noémie Courtejoie, Paolo Bosetti, et al. (2020a) Estimating the burden of SARS-CoV-2 in France (pasteur-02548181).

Salje H. et al. (2020b), Estimating the burden of SARS-CoV-2 in France, *Science* 10.1126/science.abc3517

Smith, Jeffrey A., Jessica Burrow. 2018. Using Ego Network Data to Inform Agent-Based Models of Diffusion, *Sociological Methods & Research. doi: 10.1177/0049124118769100.*

Stein, R. A. (2011). Super-spreaders in infectious diseases. *International Journal of Infectious Diseases*, *15*(8), e510-e513.

Stumpf, M. P., & Porter, M. A. (2012). Critical truths about power laws. *Science*, *335*(6069), 665-666.

Sun, L., Axhausen, K. W., Lee, D. H., & Cebrian, M. (2014). Efficient detection of contagious outbreaks in massive metropolitan encounter networks. *Scientific reports*, *4*, 5099.

Trewick, C., Ranchod, P., & Konidaris, G. (2013, August). Preferential Targeting of HIV Infected Hubs in a Scale-free Sexual Network. In *the Annual Conference of the Computational Social Science Society of the Americas*.

Vázquez, A., Pastor-Satorras, R., & Vespignani, A. (2002). Large-scale topological and dynamical properties of the Internet. *Physical Review E*, *65*(6), 066130.

Walker, Patrick GT, Charles Whittaker, Oliver Watson et al. The Global Impact of COVID-19 and Strategies for Mitigation and Suppression. Imperial College London (2020), doi: https://doi.org/10.25561/77735

World Health Organization (2020), Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

Woolhouse, M. E., Dye, C., Etard, J. F., Smith, T., Charlwood, J. D., Garnett, G. P., Watts, C. H. (1997). Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proceedings of the National Academy of Sciences*, *94*(1), 338-342.

Wong, G., Liu, W., Liu, Y., Zhou, B., Bi, Y., & Gao, G. F. (2015). MERS, SARS, and Ebola: the role of super-spreaders in infectious disease. *Cell host & microbe*, *18*(4), 398-401.

Zhang, J., Klepac, P., Read, J. M., Rosello, A., Wang, X., Lai, S., Yang, J. (2019). Patterns of human social contact and contact with animals in Shanghai, china. *Scientific reports*, *9*(1), 1-11.

Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S., Ajelli, M. (2020). Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* April 29, 2020.

Appendix

A1. Segregation of network degree by profession

One potential method for finding individuals with high contact frequencies is to target professions characterized by high contact. The professional categories in the COMES-F data (Figure 3 in main text) are too coarse to evaluate the effectiveness of such a method. We therefore exploit a recent survey that has a somewhat less thorough measurement of close-range contact (Belot et al. 2020) but fine-grained professional categories. The Belot et al. (2020) survey was conducted in the third week of April, 2020 in the midst of the Covid-19 epidemic in six countries: China, South Korea, Japan, Italy, the UK and four states in the US: California, Florida, New York, and Texas. The sample consists of roughly 1,000 individuals from each country for a total of 6,082 respondents. Data was collected using market research companies Lucid and dataSpring, using gender and income quota. With regard to close-range contact, instead of being asked to keep a two-day-long diary, respondents were asked: "On a typical working day (before the outbreak of Covid-19), with how many people would you have close social contact (at less than 1 metre distance) and how long would you interact with them? (indicate approximate numbers - leave blank if the answer is zero)". Respondents' professions were classified in terms of the O-Net classification used by the US Bureau of Labor Statistics.
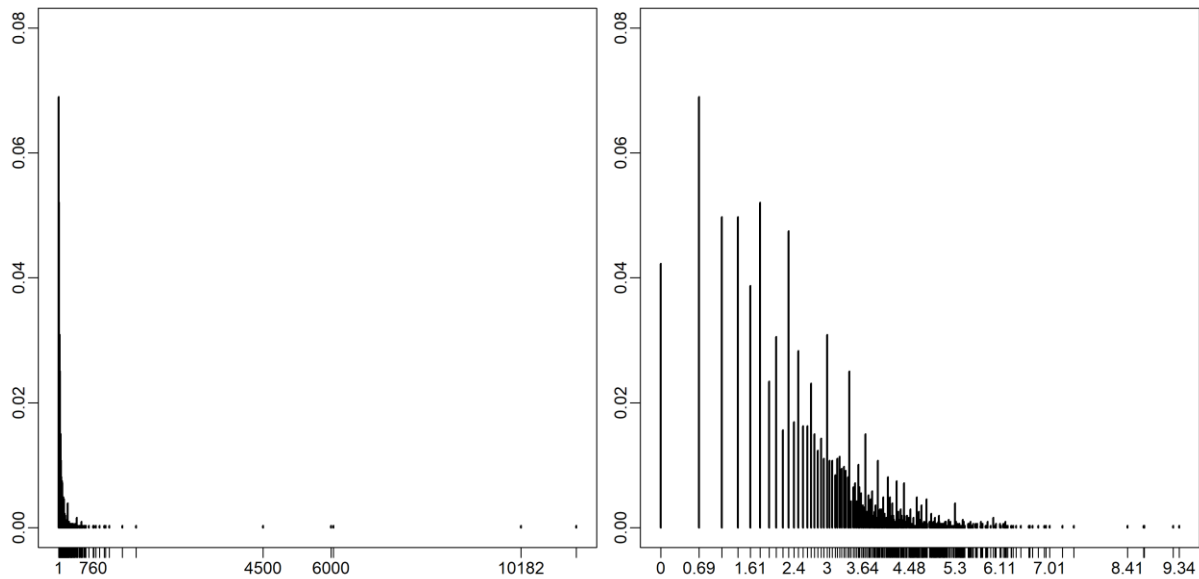

Figure A1. Fraction of cases (y-axis) reporting a given number of close-range contacts (x-axis) in the Belot et al. (2020) data (n=4,103). Left: Linear scale. Right: Logarithmic scale.

Profession	Mean # contacts	Median # contacts	N
Elementary School Teacher	120	50	17
Cashier	76	40	20
Order Clerks	70	34	34
Teacher Assistants	67	40	18
Retail Salespersons	62	17	29
Administrative Services Managers	59	16	47
Childcare Workers	49	30	19
Bill and Account Collectors	45	19	21
Sales Managers	45	17	20
Computer and Information Systems Managers	34	14	33
Financial Analysts	34	19	25
Customer Service Reps	31	18	49
Audio and Video Equipment Technicians	31	19	27
Construction Managers	30	10	33
Construction Laborers	28	12	24
Architectural Drafters	25	11	16
File Clerks	24	12	38
Civil Engineers	23	16	53
Data Entry Keyers	22	16	17
Credit Checkers	22	14	16
Ophthalmic Laboratory Technicians	21	14	18
Computer Network Support Specialists	19	12	28
Financial Managers, Branch or Department	17	7	25
Financial Examiners	17	6	20
Computer Programmers	15	9	16

Table A1. Close contact by profession in Belot et al. data (2020)

The distribution of close contact frequency in the Belot et al. (2020) data is displayed in Figure A1. The distribution is severely right-skewed, as we also observed for the French survey data. This provides confidence that the existence of hubs is not a measurement artifact but a robust feature of contact networks: Using different methods for measuring contact the same distributional characteristic is obtained.

Table A1 shows the mean and median number of close-range contacts by profession, in descending order of mean contact frequency, combining short- and long-duration contacts, excluding zero answers and professions with 15 or fewer cases. Table A1 has face validity, topped by professions that clearly involve close contact with many individuals -- elementary school teachers, cashiers -- and at the very bottom individuals who mostly work from home -- computer programmers. Some professions have an order of magnitude greater mean close-range contact than others. The spread is substantial especially when considering the ambiguity in the possible interpretation of the phrase "social contact" used in the questionnaire, translated into different languages, and the difficult task of estimating such numbers without use of a contact diary, which will produce noise that suppresses measured occupational differences. These results suggest that targeting select professions may be an effective strategy for finding hubs in contact networks.

A2. Topology of the empirical network

Table 1 in the main text shows a number of topological features of the empirical network that we generated to match the empirical distribution of close contacts. Here we discuss the topological features of the empirical network in greater detail and compare them with those observed for social networks measured elsewhere. The first is *degree assortativity*. Social networks tend to be characterized by positive degree assortativity, i.e. nodes with similar degree are more likely to be connected (see Newman 2003a). However, in presence of broad degree distributions, the network can turn to be disassortative. This results from the fact that, while high-degree nodes still tend to be connected to one another, thus pushing assortativity up, low-degree nodes are preferentially attached to high-degree nodes, thus pushing assortativity down. Depending on the specific distributions of high- and low-degree nodes these two effects will be more or less balanced (see Johnson et al. 2010). In our case, the simulated empirical network is slightly disassortative, which suggests that the highly connected individuals are not only connected amongst themselves but they also widely connect to individuals with less close-range contacts. The corresponding random network has assortativity close to zero. Degree disassortativity renders targeting of high-degree nodes more effective: "(...) those vertices are strewn far apart across the network, so that attacking them attacks all parts of the network at once" (Newman 2003a: 10).

The *clustering coefficient* of the empirical network is 0.12, which is in line with values observed for many social networks (see Newman 2003b: table 3.1; Jackson 2008: 3.2.2). It is about 7 times higher than the corresponding value for our benchmark random network. Analysis of the entire distribution of nodal clustering coefficients (available upon request) suggests systematic variation across nodes with different degrees. In particular, low-degree nodes tend to have higher clustering than high-degree nodes. On the overall simulated empirical network, this produces a negative *correlation* of -0.26 between nodal

degree and *clustering* coefficient (meaning that the higher the degree of a node the lower the fraction of connections among its own neighbors). This means that, in addition to a broad degree distribution, our simulated empirical network contains hubs that span across the network rather than clustering together. This is a feature shared with some previously studied networks with high-degree nodes (Barabási 2014: 232-237). It is also a topological feature that makes hub-centered interventions especially effective: attacking the hubs means interrupting (or slowing down) communication among the modules (*ibid*: 236). By contrast, the correlation between nodal degree and clustering in the random network we study tends toward zero.

Finally, the network we generated based on the empirically observed broad degree distribution of close-range contacts in France exhibits the typical high reachability of small-world topologies: its *average path length* and *diameter* are comparable to that of an Erdős–Rényi network with the same size and degree, and is consistent with what is usually observed in pure scale-free models (see Albert and Barabasi 2002:74). Highly connected individuals are effective in bringing many parts of the network close.

Thus, overall, the network that we consider as the possible relational infrastructure of the virus spread in France is anchored to the French long-tailed distribution of close-range contacts and, at the same time, exhibits reasonable topological features in terms of degree assortative, community structure, and reachability.

A3. Results for the random network

Figure 5 and table 3 show large differences in the effectiveness of interventions that do and do not target high-contact individuals for immunization. Here we explore how instrumental the skewness in the empirical distribution of close-range contact is for the effectiveness of hub targeting. We do so by recalculating the figure 5 and table 3 results for the random network, respectively figure A2 and table A2.
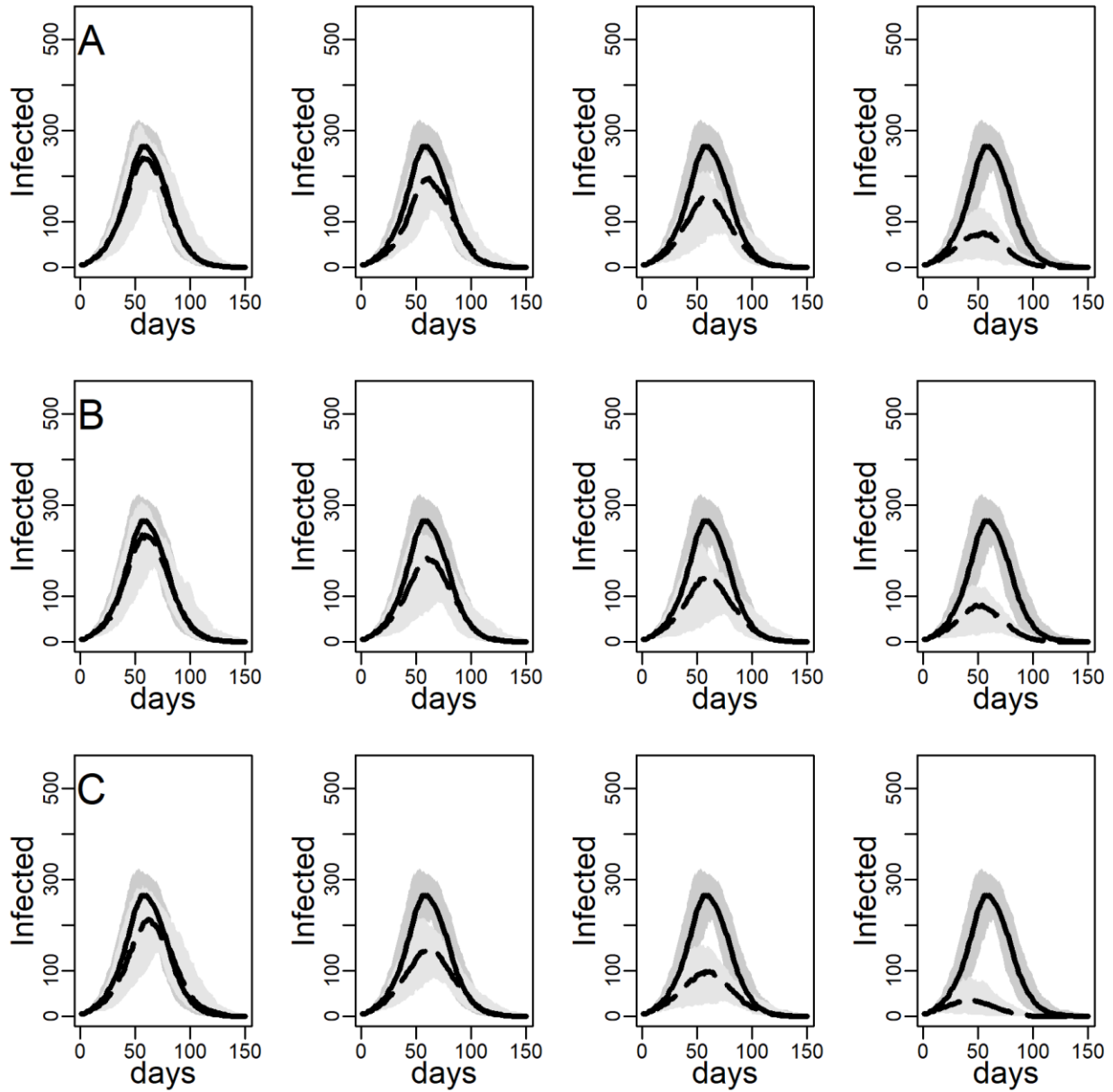
Figure A2. Number of infected agents (y-axis) by days (x-axis) (median values across 100 replications) under three different interventions (rows) targeting 1, 3, 5, or 10 agents per day (columns). A – NO-TARGET immunization; B – CONTACT-TARGET immunization. C – HUB-TARGET immunization. Lower and upper bounds of the shaded areas correspond to the 5th percentiles and 95th percentiles of the 100 replications. Solid line: *random* network; dashed line: interventions. n = 2029 agents. R_0=1.785.

A comparison of panels A between figures A2 and 5 shows that NO-TARGET interventions are less effective in empirical networks with high degree skew than in random networks with low degree variance. This suggests that models that do not account for empirical network structure may overestimate the expected impact of interventions. Comparing panels B and C across figures we find that HUB-TARGET

and CONTACT-TARGET interventions are much more effective in the empirical network than in random networks, where the to-be-immunized agents have lower network degree.

Random network							no intervention: H = 265 [207; 310]; T = 60		
		$b = 1$		$b = 3$		$b = 5$		$b = 10$	
	peak height	time	peak height	time	peak height	time	peak height	time
NO-TARGET	239 [135; 307]	57	196 [111; 243]	62	153 [53; 206]	57	78 [17; 128]	55
CONTACT-TARGET	234 [121; 307]	56	183 [66; 235]	60	146 [40; 188]	60	81 [17; 119]	54
HUB-TARGET	213 [111 ; 275]	63	148 [80; 197]	63	98 [27; 149]	58	34 [7; 79]	38

Table A2. Peak height (maximum # concurrently infected agents) and time (in days) under three interventions (rows) and four budgets (column) on the *random* network. Shown are median, 5% and 95% percentiles across 100 iterations. $R_0=1.785$.

A4. ICU admissions at the national scale

Simulated peak within the artificial population can be transformed into real-world ICU admissions according to the following transformation rule: $((H_{ABM} \times P_{HOSP} \times P_{ICU})/N_{ABM}) \times N_{FRANCE}$, where H_{ABM} is the peak of infected agents within a given simulated scenario, P_{hosp} is the empirical probability for an infected person to be hospitalized, P_{ICU} is the empirical probability for an infected-hospitalized person to enter ICU, and N_{ABM} and N_{FR} respectively are the sizes of our artificial population and of French metropolitan population (at the 1st of january 2020 according to the national bureau of statistics's official counting). P_{hosp} and P_{ICU} are based on estimates produced by Salje et al. (2020b) who found, for France, that 3.6% (95% CrI: 2.1-5.6) of infected individuals are likely to require hospitalization, and, once hospitalized, 19% (95% CrI: 18.7%-19.4%) of those individuals are likely to enter ICU. Our translation is based on central values.

Empirical network						no intervention: H = 102826 [44412; 120110]; T = 32		
	$b = 1$			$b = 3$		$b = 5$		$b = 10$
	peak height	time	peak height	time	peak height	time	peak height	time
NO-TARGET	97794 [44412; 117922]	31	89043 [37849; 102826]	34	85542 [15752; 103920]	32	71322 [19253; 89918]	32
CONTACT-TARGET	79197 [33473; 96700]	35	42661 [8314; 60820]	41	23191 [9845; 35880]	45	5469 [219; 15752]	30
HUB-TARGET	25815 [5470; 39818]	49	4376 [291; 56]	32	2844 [656; 5469]	18	1969 [857; 3500]	13

Table A4. ICU admissions at the national scale corresponding to Peak height (maximum # concurrently infected agents) generated by the model (see table 3) under three interventions (rows) and four budgets (column) on the *empirical* network. Shown are median, 5% and 95% percentiles across 100 iterations. R0=1.785.