

Effect of manual and digital contact tracing on COVID-19 outbreaks: a study on empirical contact data

Supplementary Information

A. Barrat^{1,2}, C. Cattuto^{3,4}, M. Kivelä⁵, S. Lehmann⁶, and J. Saramäki⁵

¹Aix Marseille Univ, Université de Toulon, CNRS, CPT, Turing Center for Living Systems, Marseille, France

²Tokyo Tech World Research Hub Initiative (WRHI), Tokyo Institute of Technology, Tokyo, Japan

³University of Turin, Turin, Italy

⁴ISI Foundation, Turin, Italy

⁵Department of Computer Science, Aalto University, Finland

⁶Technical University of Denmark, Copenhagen, Denmark

S1 The parabola shape of the cost-benefit curve

The downward opening parabola shape of the Fig. 2C in the main text, giving the cost of the measures as a function of their benefit (reduction in epidemic size) can be understood via mathematical argumentation based on two stylistic assumptions about the behavior of the epidemic spread with regard to quarantine measures. Here we do not assume any specific epidemic model, but use the more easily understandable first-order behavior of the simulation (and the analytical) models as an input. Putting these inputs together leads to a second order polynomial shape for the number of quarantined versus reduction of the epidemic size, similar to the one observed for the full simulation model.

In order to simplify our computations we define auxiliary quantities that capture the overall behavior of the full simulation epidemics. These quantities describe the average epidemic observables and we disregard for simplicity all the variation and correlation between the various quantities. The variables are

- I_0 : fraction of infected population without quarantine measures;
- I_r : fraction of infected population when quarantine measures are implemented;
- q_r : fraction of quarantined individuals out of those who were in contact with an infected individual;
- k : the average number of individuals encountered by an infected person during the time they are infectious.

The two quantities we are interested were defined in the main text as:

- cost q : fraction of population that is quarantined at some point during the epidemics;
- benefit $r = \frac{I_0 - I_r}{I_0}$: relative reduction in the fraction of infected population.

We denote by f the function describing how the fraction of quarantined contacts q_r reduces the epidemic size (this function is strictly decreasing and is thus invertible):

$$I_r = f(q_r)I_0. \quad (\text{S1})$$

The probability of quarantine q_r , in addition to having the effect of decreasing the epidemic size, determines the number q of quarantined people, together with the other variables k and the epidemic size. Here we assume, in a mean-field hypothesis, that q is given by the product of the fraction of infected people, the average number of contacts k a person has while infective and the fraction q_r of quarantined individual out of those:

$$q = kq_r I_r. \quad (\text{S2})$$

The main assumption behind this equation is that the quarantine events take place independently of each other. This is the case if the epidemics is relatively small and the network neighborhoods of the nodes are not clustered: correlations between the states of individuals created by the dynamics of the spread are neglected. We note however that local clustering could be absorbed into k , independently from the epidemic size, which means that such correlations would not change the functional shape of the equations we derive next, but only the precise numerical values of the parameters.

From Eqs. (S1)-(S2) and the definition of $r = 1 - I_r/I_0$ we obtain:

$$q = kI_0(1 - r)f^{-1}(1 - r) \quad (\text{S3})$$

We further assume that the quarantine efficiency function is linear:

$$f(q_r) = 1 - \frac{I_0 - I_m}{I_0} \frac{q_r}{q_{\max}}, \quad (\text{S4})$$

where q_{\max} is the maximum value q_r can take and I_m is the fraction of people still infected when $q_r = q_{\max}$. Note that q_{\max} does not reach 1 in the simulations even at full quarantine rates as the quarantine procedures do not find all contacts as $\tau_{ct} > 0$, and also when the compliance is not perfect. Inverting f and inserting it into the equation (S3) we get:

$$q = \frac{kI_0^2 q_{\max}}{(I_0 - I_m)} r(1 - r). \quad (\text{S5})$$

From the above equation we see that the function $q(r)$ is a downwards opening second order parabola that always starts at $q(r = 0) = 0$ but is incomplete in a sense that the arch ends at $r = 1 - I_m/I_0$. The parabola reaches its peak (maximum cost) at $r = 1/2$ and the height of the peak is controlled by the combination of disease parameters and parameters related to interventions. The curve is illustrated in Fig. S1.

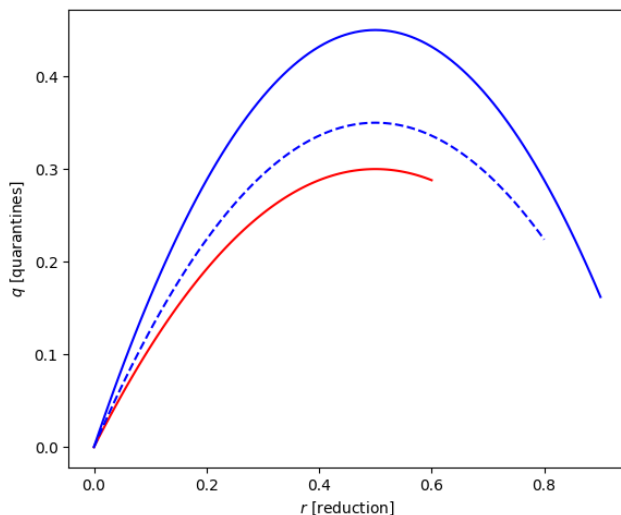


Figure S1: Some curves produced by the heuristic argument for the number of quarantined people (cost) versus the amount of epidemic size reduction (benefit). See equation (S5).

S2 Static network based approximation of the epidemic size

We develop a simplified epidemic model with manual contact tracing and app contact tracing in order to demonstrate the robustness of our results. Our aim is a minimal model that can exhibit all the phenomena we observe for the more complicated simulation model, but is simple enough so that we can write equations for the expected epidemic size. We thus consider an SIR model with two types of individuals: ones with the app and ones without the app. Further, the underlying contact network is assumed to be a static configuration model. We write the equations for the epidemic size for a generic degree distribution of the network, and we will illustrate our findings in the case of Erdős-Rényi (ER) random graphs.

Considering the SIR model at large times is approximately equivalent to studying the connectivity of the network when parts of the links are removed.¹⁻³ The idea of this approach is that the links that do not have time to pass the infection before the source node moves to the recovered state are removed, and after this the component structure of the network gives the distribution of possible final epidemic sizes. For a simple SIR epidemic model on ER networks, this approach is known to

yield exactly the same result as more conventional fully mixed population models.⁴ This solution is given by the equation $R_\infty = 1 - e^{-R_0 R_\infty}$, where R_0 is the expected number of individuals infected by a single infected individual.

S2.1 Model without manual tracing and no quarantine failure possibility

To illustrate our approach we will first derive a solution to the model with no manual contact tracing and where the digital contact tracing (app) prevents the infections with full certainty. This model with app can be solved by extending the standard techniques for solving the final epidemic size for the normal SIR model.³

More precisely, for a standard SIR model one would define u as the probability that following a link does not lead to the giant component, and then write a self-consistency equation for this probability. For our model, we have two different kinds of nodes (with or without app installed) and we need two define two analogous probabilities:

$$u_0 = \text{'Probability that following a link to a non-app user does not lead to the giant component'} \quad (\text{S6})$$

$$u_a = \text{'Probability that following a link to an app user does not lead to the giant component'} \quad (\text{S7})$$

Given that the excess degree of the node at the end of the followed link is k , we can write the conditional probabilities $u_0(k)$ and $u_a(k)$ using u_0 and u_a . If out of the k new neighbors the node has there are k' app users, then the probability $u_0(k)$ that the non-app user is not in the giant component is given by $u_a^{k'} u_0^{k-k'}$. For an app user, we do not follow the links leading to another app users, so the probability $u_a(k)$ is given by $u_0^{k-k'}$. The number of app user neighbors k' out of k neighbors follows the binomial distribution with probability p_{app} so we get

$$u_0(k) = \sum_{k'=0}^k \binom{k}{k'} p_{app}^{k'} (1 - p_{app})^{k-k'} u_a^{k'} u_0^{k-k'}, \quad (\text{S8})$$

$$u_a(k) = \sum_{k'=0}^k \binom{k}{k'} p_{app}^{k'} (1 - p_{app})^{k-k'} u_0^{k-k'}. \quad (\text{S9})$$

These probabilities can be rearranged and using the binomial theorem we can write

$$u_0(k) = [(1 - p_{app})u_0 + p_{app}u_a]^k, \quad (\text{S10})$$

$$u_a(k) = [(1 - p_{app})u_0 + p_{app}]^k. \quad (\text{S11})$$

Now taking into account the excess degree distribution $q(k)$ we can write the self-consistency

equations for u_0 and u_a as

$$u_0 = \sum_{k=0}^{\infty} q(k)u_0(k) = \sum_{k=0}^{\infty} q(k)[(1 - p_{app})u_0 + p_{app}u_a]^k, \quad (\text{S12})$$

$$u_a = \sum_{k=0}^{\infty} q(k)u_a(k) = \sum_{k=0}^{\infty} q(k)[(1 - p_{app})u_0 + p_{app}]^k. \quad (\text{S13})$$

By denoting the generating function of $q(k)$ as $g_1(z)$ and using the definition of generating functions [$g_1(z) = \sum_{k=0}^{\infty} q(k)z^k$] we can write

$$u_0 = g_1((1 - p_{app})u_0 + p_{app}u_a), \quad (\text{S14})$$

$$u_a = g_1((1 - p_{app})u_0 + p_{app}). \quad (\text{S15})$$

The two probabilities u_0 and u_a can be solved from the above equations. We can then compute the giant component size using these probabilities as

$$S = 1 - p_{app}g_0((1 - p_{app})u_0 + p_{app}) - (1 - p_{app})g_0((1 - p_{app})u_0 + p_{app}u_a), \quad (\text{S16})$$

where g_0 is the generating function for the degree distribution of the network. If the disease spreading is starting from a single node, the size of the epidemic spans a zero fraction of the network if it starts from a node in a non-giant component and a fraction S if it starts from the giant component. Thus, the expected final size of the epidemics is $R_{\infty} = 0 \times (1 - S) + S \times S = S^2$.

For Erdős-Rényi networks we can compute numerically the epidemic size by using and inserting $g_0(z) = g_1(z) = e^{R_0(z-1)}$.

S2.2 Model with manual tracing and possibility for quarantine failure

We next develop the above model further by considering that the quarantine has probability q_{succ} to succeed and by introducing a probability for manual contact tracing p_{ct} .

Let us first focus on how $u_0(k)$ (and therefore Eq. S14) needs to be updated. For a node that does not have the app, there are now two ways that a link leaving it does not lead to the giant component: either the link is manually traced and the quarantine works, or one of these fails and the link happens to lead to another node that doesn't belong to the giant component. The former happens with probability $p_{ct}q_{succ}$ for each link, and the latter with $1 - p_{ct}q_{succ}$ times the probability we wrote for the case without contact tracing (Eq. S8). That is, we write probabilities of there being k' failure in manual tracing or quarantine (out of k in total) and $k - k'$ success. Then out of those k' failures there are k'' links leading to an app user and $k' - k''$ individuals without the app. In total,

the formula becomes

$$u_0(k) = \sum_{k'=0}^k \binom{k}{k'} (p_{ct} q_{succ})^{k-k'} (1 - p_{ct} q_{succ})^{k'} \left[\sum_{k''=0}^{k'} \binom{k'}{k''} p_{app}^{k''} (1 - p_{app})^{k'-k''} u_a^{k''} u_0^{k'-k''} \right]. \quad (\text{S17})$$

Now we can apply the binomial theorem (exactly like we did going from Eq. S8 to Eq. S10) and write the part in the square brackets as $[(1 - p_{app})u_0 + p_{app}u_a]^{k'}$. We then apply the binomial theorem again to write:

$$u_0(k) = p_{ct} q_{succ} + (1 - p_{ct} q_{succ}) [(1 - p_{app})u_0 + p_{app}u_a]. \quad (\text{S18})$$

The equation for $u_a(k)$ is slightly trickier to write. We need to consider that the manual contact tracing and app contact tracing lead to the same tracing procedure that can fail. That is, if both manual contract tracing and app contact tracing would be applied to a link, it is not possible that one of them succeeds and the other one does not. They are both done with the same procedure, and either the quarantine is too late or it fails through some other way or not. What we do is to separate the links into ones where the quarantine would succeed if it was applied (with probability q_{succ}) and the ones where it does not succeed (with probability $1 - q_{succ}$).

To make the exposition clear we avoid writing multiple stacked binomial distributions (similar to Eq. S17 where we had two binomial distributions). Instead we describe the various cases separately and write down the final equation in the form of Eq. S18.

At the first level, we consider links where the quarantine would succeed if we applied it (either through manual or app contact tracing). This will fail with probability $(1 - q_{succ})$ and in this case the success of none of the quarantine measures make any difference and the probability of not going to the giant component is $(1 - p_{app})u_0 + p_{app}u_a$. Note that this is the same probability that appears inside Eq. S10, which also describes the case of no possibility of quarantine.

The possible quarantine would succeed with probability q_{succ} . In this case if the link is traced manually with probability p_{ct} it does not lead to the giant component with certainty. If the manual tracing fails there is still a possibility for the manual app tracing to work. This now happens with certainty if the neighbor has the app and otherwise the probability for the link to not lead to the giant component is u_0 . That is, in total the probability is $(1 - p_{app})u_0 + p_{app}$. Note that this is the same probability that appears inside Eq. S11, which describes the case of the app quarantine working if we are coming to a node with the app (and within a model where the quarantines always work).

Putting together the parts for the equation for $u_a(k)$ we get

$$u_a(k) = q_{succ} (p_{ct} + (1 - p_{ct})((1 - p_{app})u_0 + p_{app})) + (1 - q_{succ})((1 - p_{app})u_0 + p_{app}u_a). \quad (\text{S19})$$

Now that we have the updated equations for $u_0(k)$ and $u_a(k)$ the rest of the model derivation goes as before. We add in the generating function for the excess degree and get the final self-consistency equations:

$$u_0 = g_1(p_{ct}q_{succ} + (1 - p_{ct}q_{succ})[(1 - p_{app})u_0 + p_{app}u_a]), \quad (\text{S20})$$

$$u_a = g_1(q_{succ}(p_{ct} + (1 - p_{ct})((1 - p_{app})u_0 + p_{app}u_a))) + (1 - q_{succ})((1 - p_{app})u_0 + p_{app}u_a) \quad (\text{S21})$$

After solving these equations, we can calculate the giant component size S for the ER networks as before and from this we get the expected size of the epidemics R_∞ .

We now have a simple theoretical model that has the parameters R_0 , p_{app} , p_{ct} , and q_{succ} . Note that the R_0 probability (which is the average degree of our contact network) is the basic reproductive number without the manual tracing and application of quarantine measures. However, as we do not model any other isolation measures those must be absorbed in the R_0 value, which lowers it from the $R_0 = 3$ value discussed in the main text.

The main text explores the effects of the two parameters p_{app} and p_{ct} on the full model with realistic temporal contact sequences. In order to confirm the findings on the overall shapes of the epidemic size reduction curves as functions of p_{app} and p_{ct} , we need to match the two remaining parameters R_0 and q_{succ} to the simulation model. These two parameters describe the average behavior of the simulation model and are likely to have a complicated relationship with the temporal contact network structure and the simulation parameters. Instead of deriving these relationships we take an alternative approach of fitting the theoretical model prediction of R_∞ to the simulation results for R_∞ in two extreme cases.

The first extreme case is the one without any interventions through contact tracing, i.e., when $p_{ct} = p_{app} = 0$. This is the equivalent of fitting the conventional SIR model to the data, and effectively fixes the R_0 value to the expected number of infectious contacts when only the isolation intervention is applied. The second extreme case is when both manual contact tracing and app contact tracing are applied with 100% certainty, i.e., when $p_{ct} = p_{app} = 1$. Effectively this reduces the problem to solving the self-consistency equation $u_a = g_1(q_{succ} + (1 - q_{succ})u_a)$, which is equivalent to the usual SIR model with basic reproduction number scaled as $R_0(1 - q_{succ})$.

Fitting to the SD data (threshold $\theta_{ct} = \theta_{app} = 15$ min) we get $R_0 = 1.6$ and $q_{succ} = 0.30$. The epidemic size reduction curve we get by varying probabilities p_{app} and p_{ct} is shown in the main text Fig 3A.

S3 Supplementary results

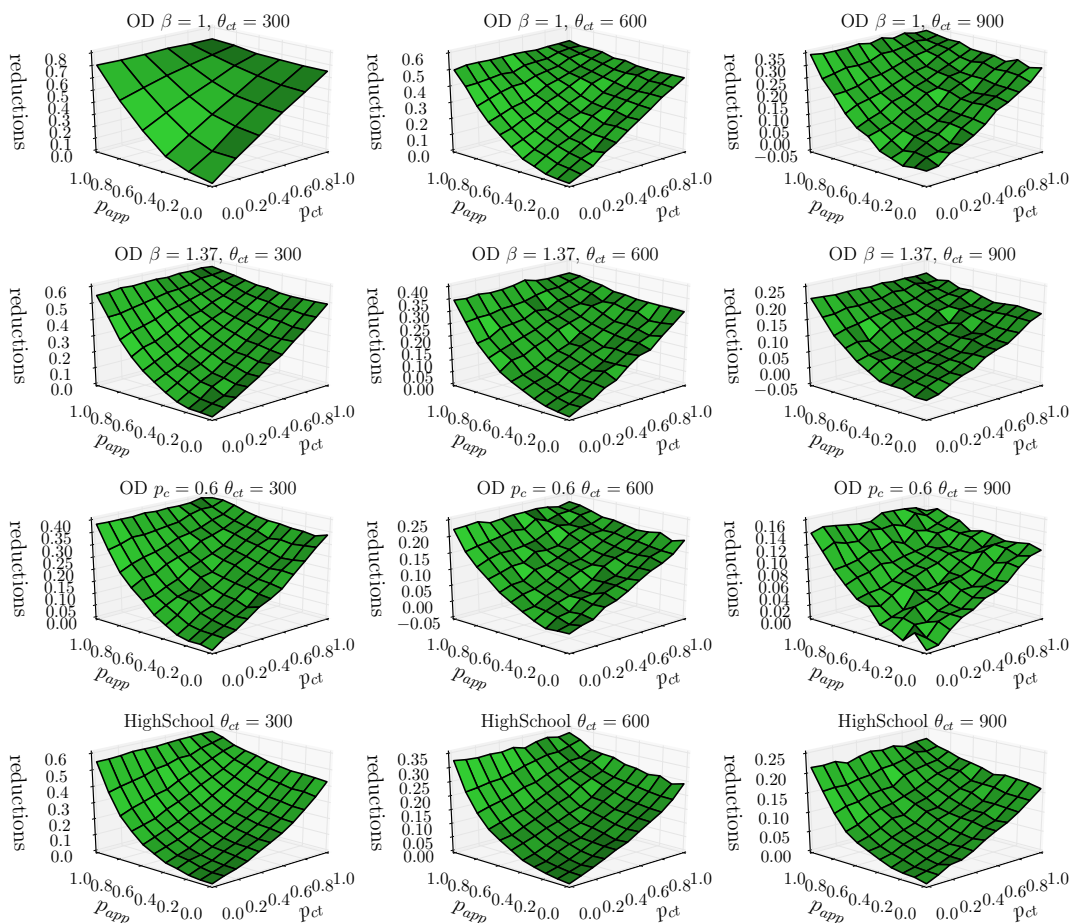


Figure S2: Average reduction in epidemic size as a function of p_{app} and p_{ct} for different parameters for the OD data and HighSchool data. Columns correspond to three different contact tracing thresholds ($\theta_{ct} = \theta_{app} = 5, 10, 15$ minutes). Top row: OD with $\beta = 1$ (corresponding to $R_0 \approx 2.36$). Second row: OD with $\beta = 1.37$ ($R_0 \approx 3$). Third row: OD with $\beta = 1.37$ and compliance rate $p_c = 0.6$. Bottom row: HighSchool data set with $\beta = 1.37$ and compliance rate $p_c = 1.0$.

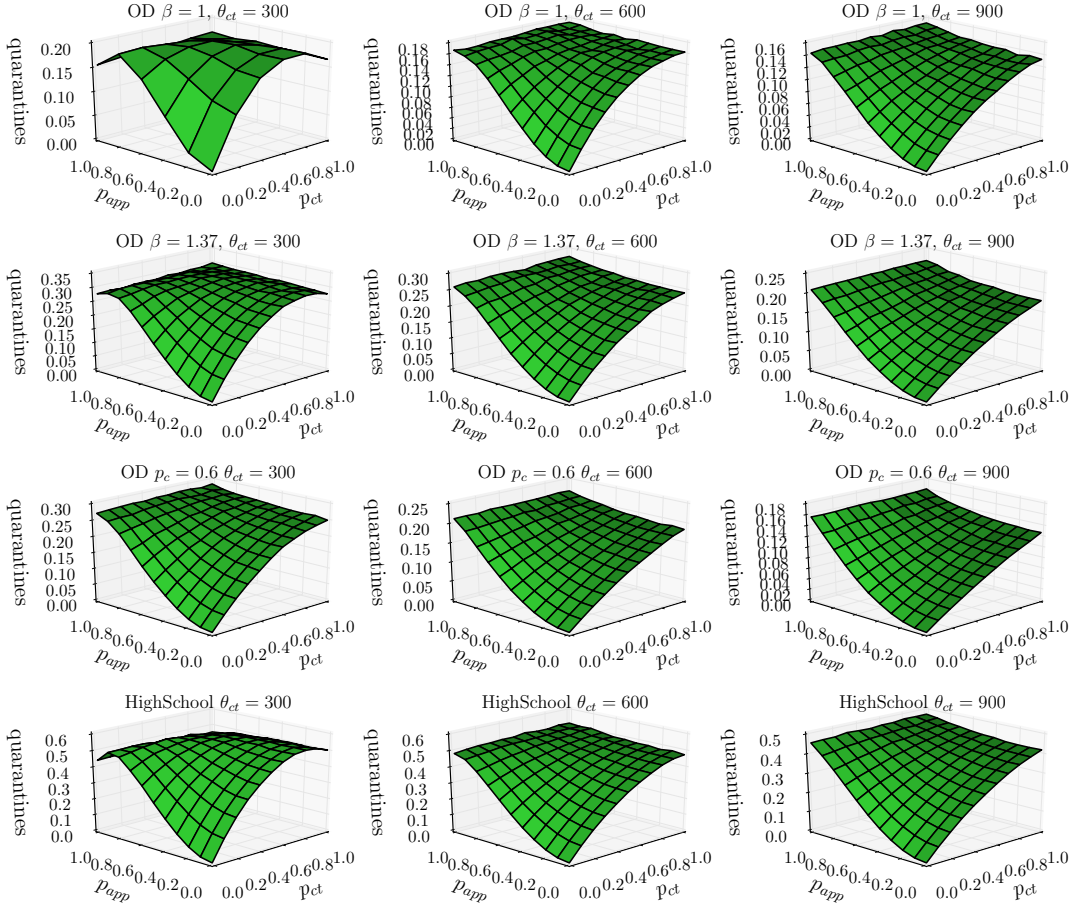


Figure S3: Average fraction of population in quarantine as a function of p_{app} and p_{ct} for different parameters for the OD data and HighSchool data. Columns correspond to three different contact tracing thresholds ($\theta_{ct} = \theta_{app} = 5, 10, 15$ minutes). Top row: OD with $\beta = 1$ (corresponding to $R_0 \approx 2.36$). Second row: OD with $\beta = 1.37$ ($R_0 \approx 3$). Third row: OD with $\beta = 1.37$ and compliance rate $p_c = 0.6$. Bottom row: HighSchool data set with $\beta = 1.37$ and compliance rate $p_c = 1.$

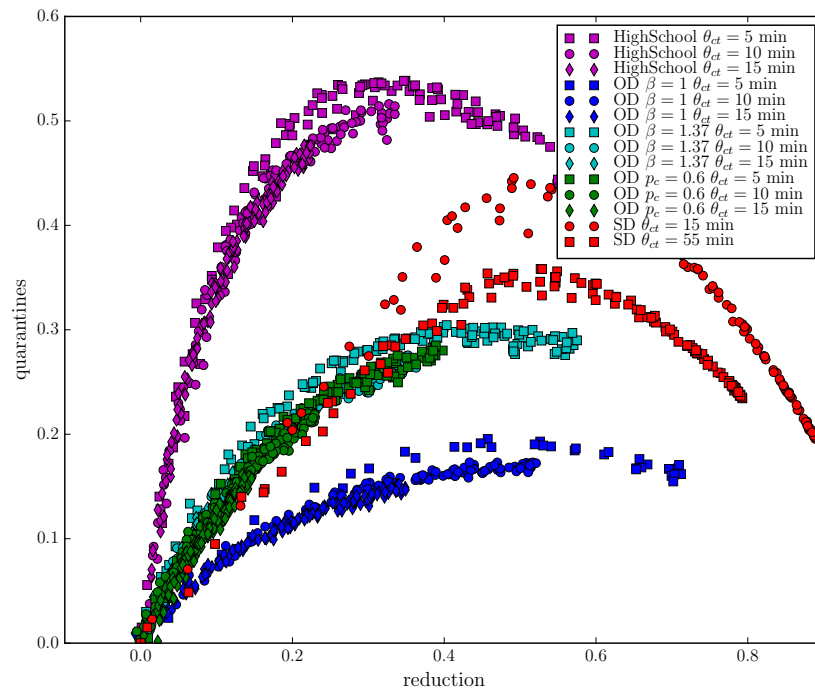


Figure S4: Cost-benefit curves: normalized number of quarantine events as a function of the relative reduction in average epidemic size, for different parameters for the OD data, HighSchool data, and SD data.

References

- [1] Denis Mollison. Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(3):283–313, 1977.
- [2] Peter Grassberger. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 63(2):157–172, 1983.
- [3] Mark EJ Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.
- [4] Andrew Barbour and Denis Mollison. Epidemics and random graphs. In *Stochastic processes in epidemic theory*, pages 86–89. Springer, 1990.