Measurability of the epidemic reproduction number in data-driven contact networks

Quan-Hui Liu[a,b,c], Marco Ajelli[c,d], Alberto Aleta[e,f], Stefano Merler[d], Yamir Moreno[e,f,g], Alessandro Vespignani[c,g]

[a] Web Sciences Center, University of Electronic Science and Technology of China, Chengdu, Sichuan, P.R. China
[b] Big Data Research Center, University of Electronic Science and Technology of China, Chengdu, Sichuan, P.R.China
[c] Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA, USA
[d] Bruno Kessler Foundation, Trento, Italy
[e] Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Zaragoza, Spain
[f] Department of Theoretical Physics, University of Zaragoza, Zaragoza, Spain
[g] ISI Foundation, Turin, Italy

SI Appendix

Contents

1 Methods Details

1.1 Data-driven model

1.1.1 Network structure

We developed a synthetic population of about 500,000 individuals resembling a subsample of the actual Italian population according to basic sociodemographic indicators (e.g., age structure, household size, employment rates, workplace size distribution, school attendance, school size distribution). We used the same algorithm we developed to infer contact patterns in the European populations [1]. As such, we refer the reader to Fumanelli et al. [1] for all technical details.

Briefly, the obtained synthetic population allows us to define a multilayer network of contacts that can be used to model the infection of the influenza virus in the population. Each layer represents one of the main four settings where influenza transmission occurs, namely households, schools, workplaces, and the general community [2, 3, 4]. The household layer is composed by n_h disconnected components, each one representing one household (i.e., the total number of households is n_h). Each household is composed by a certain amount of nodes, each one representing one single individual. The amount of nodes in each household is determined by sampling from the actual Italian household size distribution. An age is associated to each node by sampling from the actual Italian age distribution by household size using an algorithm tailored to match realistic age-gaps between household members (details in Fumanelli et al. [1]). Each node lives also in the community layer, which is composed by one single giant fully connected component. By sampling from the multinomial distribution of schooling and employment rates by age, for each node is determined whether it also lives in the school or workplace layer. Analogously to what has been described for the household layer, the workplace layer is composed by n_w disconnected components, each one representing one workplace. The number of nodes in each workplace is sampled from the actual Italian workplace size distribution. The school layer is composed by n_s disconnected components, each one representing one school. The number of nodes in each school is sampled from the actual Italian school size distribution by school level (e.g., pre-school, primary school, middle school, high school). Details in Fumanelli et al. [1].

Analogously, by using sociodemographic data specific for The Netherlands, we developed a multilayer network resembling a sample of about 500,000 individuals of the actual Dutch population.

1.1.2 Infection transmission process

We describe the influenza transmission process as a discrete-time Susceptible-Infectious-Removed (SIR) model with time-step t (corresponding to 1 day). Two types of transitions between states are possible: i) from susceptible (S) to infectious (I), and ii) from infectious (I) to removed (R).

Transition from S to I. Given that at time-step t node j is infectious and its neighbor node i is susceptible, the probability that j infects i (i.e, i changes its status from S to I) is given by:

$$\beta_i = pw_l\rho(a_i) \tag{1}$$

where

- p is the transmission probability per contact (day^{-1}).

- w_l is the contact weight in layer l, $l \in L = \{$h, s, w, c$\}$, where L is the set of layers; $h, s, w,$ and c represent household, school, workplace, and general community layers, respectively. Note that w_l embeds also the cluster size in that layers; for instance in the community w_l can be interpreted as $\hat{w}_l/(N-1)$ where N is the number of individuals in the community and $\hat{w}_l$ is the specific wight of one contact in the community.

- a_i is the age of individual i.

- $\rho(a_i)$ is the age-dependent susceptibility to infection.

Note that when node j infects node i at time-step t in layer l, we store all this information that will be useful for the computation of the reproduction number and the generation time.

Transition from I to R. Given that at time-step t node i is in the infectious state, it has a probability γ to recover at time-step $t+1$.

1.1.3 Model initialization and calibration

2009 H1N1 influenza pandemic At the beginning of each simulation, for each node i of age a_i, we determine whether it is initially susceptible or removed by sampling from a Bernoulli distribution with probability $P(a_i)$, where $P(a)$ is the fraction of seropositive individual of age a as observed before the 2009 H1N1 pandemic in Italy [5]. Simulations are initialized with one infectious individual and no further introductions of index cases is assumed.

One of the most striking features of the 2009 H1N1 influenza pandemic, observed in several studies focusing on different countries around the globe, was the presence of an age-specific susceptibility to infection, which was remarkably higher for young individuals (mainly corresponding to school-age individuals) [4, 5, 6, 7, 8, 9, 10]. In our model we use the estimates of age-specific susceptibility rates specific for the Italian population [5]:

$$
\rho(a) = \begin{cases} 1.00 & \text{if } a \in [0,4] \text{ years} \\ 0.85 & \text{if } a \in [5,19] \text{ years} \\ 0.25 & \text{if } a \in [20,64] \text{ years} \\ 0.10 & \text{if } a \geq 65 \text{ years} \end{cases}
$$

As direct estimates of the layer-specific transmission rates, we estimated them on the basis of age-specific seroprevalence data collected in Italy after the end of the 2009 H1N1 influenza pandemic [5]. Since the per-contact transmission probability p and the four layer-specific transmission rates w_l are multiplying factors in Equation (1) (and thus cannot be estimated simultaneously), we fixed $p = 1$ and estimated $w_h, w_s, w_w,$ and w_c.

The calibration of layer-specific transmission rates is performed by using Markov chain Monte Carlo (MCMC) sampling applied to the binomial likelihoods ($\mathcal{L}$) of the age-specific prevalence of H1N1 antibodies observed in the Italian population after the end of the 2009 H1N1 influenza
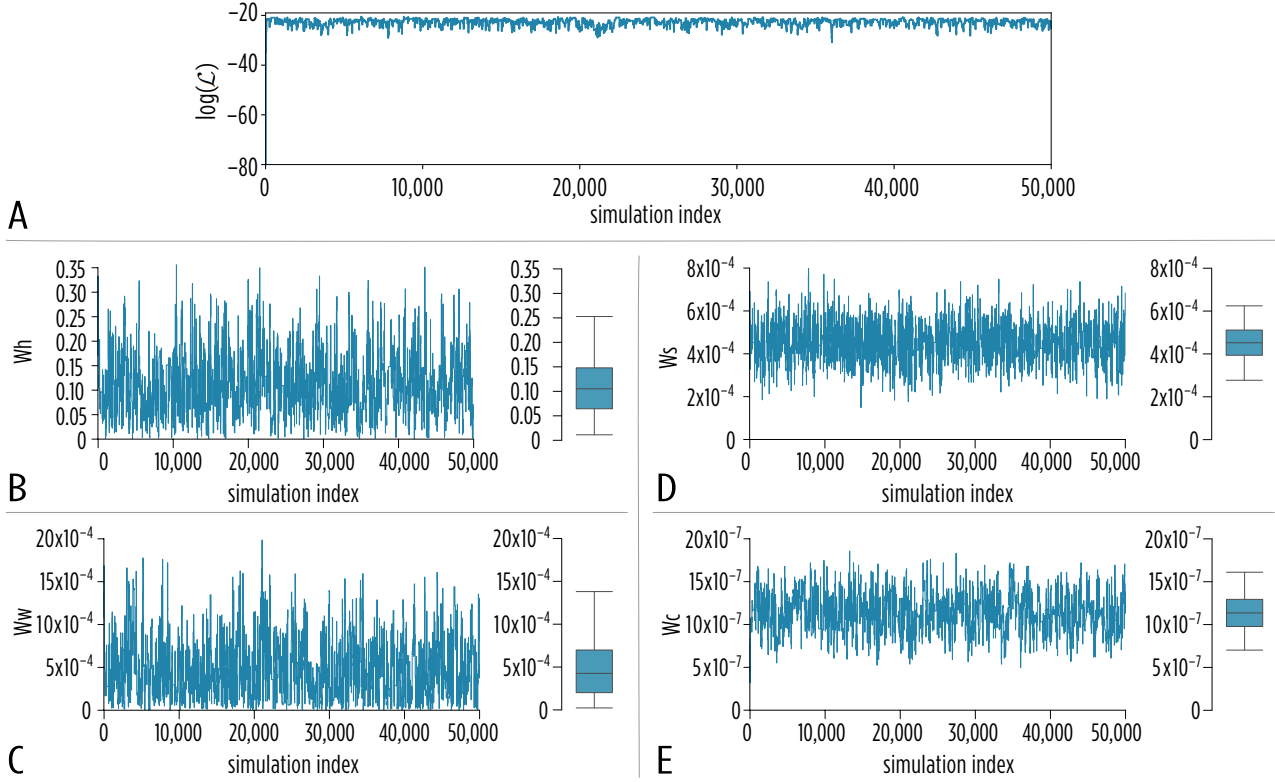
Figure S1: Logarithm of the value of the likelihood function (A) and of model parameters at each iteration of the MCMC algorithm (B-E). The four boxplots show the posterior distribution of model parameters.

pandemic [5]. We define $\Theta = (w_h, w_s, w_w, w_c)$ to be the vector of parameters to be estimated by MCMC and the likelihood function as

$$\mathcal{L}(n, r|\Theta) = \prod_{m \in M} \frac{n_m!}{r_m!(n_m - r_m)!}(\alpha_m(\Theta))^{r_m}(1 - \alpha_m(\Theta))^{n_m - r_m}$$

where

- M is the set of age groups considered in [5], namely 0-4, 5-19, 20-64, and 65+ years old;

- n_m is the number of individuals in the m-th age group in the dataset;

- r_m is the number of seropositive individuals (haemagglutination inhibition titre 1:40 or more) in the m-th age group in the dataset of Ref. [5];

- $\alpha_m(\Theta)$ is the fraction of removed individuals at the end of the pandemic in the m-th age group as resulting from model simulation performed by using parameter set Θ.

The posterior distribution of Θ is determined using random-walk Metropolis-Hastings sampling [11]. We performed 50,000 simulations and considered a burn-in period of 2,000 iterations. We assume no *a priori* knowledge on model parameters (i.e., flat prior distributions). Convergence was checked by considering different starting points and by visual inspection (see Fig. S1A-E). The estimated posterior distribution of Θ is reported in Fig. S1B-E.

"Uncharacterized" future influenza pandemic. At the beginning of each simulation, we assume that the population is fully susceptible. Simulations are initialized with one infectious individual and no further introductions of index cases is assumed.

In the classic SIR model, the removal rate corresponds to the inverse of the generation time [12]. According to the literature, influenza generation time is about 3 days [8, 13, 14]. As such, we fix the removal rate γ to $1/3$ days^{-1}, assuming that a future influenza pandemic will be characterized by a generation time similar to previous influenza pandemic and seasonal variants.

Predicting the future values of the age-specific susceptibility rates during a future influenza pandemic event is not possible. As such, we decided to set $\rho(a) = 1$ for each age a, and thus $\beta_i = \beta$. An equally impossible task is to predict the values of the layer-specific contact rates. Therefore, we decided to fix these weights as proposed by Ferguson and colleagues by analyzing past influenza seasons [2, 3]. Following Ferguson and colleagues approach we set w_l in such a way that, when simulating influenza pandemics with $R^{index} = 1.3$ the fraction of the final infection attack rate related to contacts in households, schools, workplaces, and the general community is equal to 0.3, 0.18, 0.19, and 0.33. respectively. The value of R^{index} (which is set to 1.3, in the ballpark for influenza [15]) is obtained by empirically choosing the per-contact transmission probability p such that the average reproduction number of the index case, averaged over 10,000 simulations is equal to 1.3.

Note that this scenario corresponds to the baseline scenario discussed in the main text.

1.2 Annealed configuration model

1.2.1 Network structure

We developed a configuration model with annealed order [16]. The degree distribution k characterizing this model is computed as follows. We consider the weighted number of contacts in each layer of the data-driven model describing the spread of an "uncharacterized" future influenza pandemic. Specifically, the weighted number of contacts of node i is

$$\sum_{l \in \{h,s,w,c\}} w_l c_l(i) \ ,$$

where $c_l(i)$ is the number of edges node i has to other nodes in layer l. We set w_l such that, when simulating influenza pandemics with $R^{index} = 1.3$ with the above described data-driven model, the two following conditions hold: i) the fraction of the final infection attack rate related to contacts in households, schools, workplaces, and the general community is equal to 0.3, 0.18,
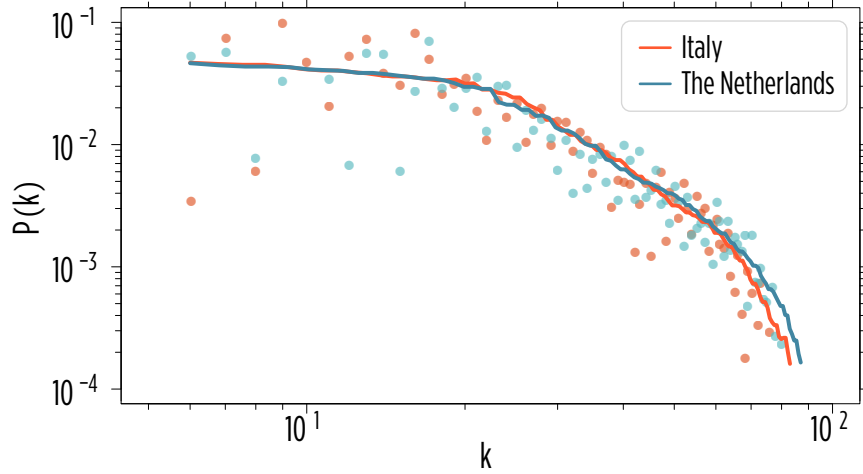
Figure S2: Degree distribution of the annealed configuration model as resulting from the weighted network of contacts for Italy and for The Netherlands in the case of an "uncharacterized" future influenza pandemic.

0.19, and 0.33, respectively; and ii) the average of the degree distribution $\langle k \rangle$ is

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} \sum_{l \in \{h,s,w,c\}} w_l c_l(i) = 19.8 \; ,$$

where N is the total number of nodes, and 19.8 is the number of face-to-face contacts recorded in the POLYMOD survey for Italy [17]. The degree distribution of the resulting network is shown in Fig. S2.

1.2.2 Infection transmission process

As for the data-driven model, we use the annealed configuration model to simulate influenza spread according to a discrete-time individual-based SIR model with time step $t = 1$ day.

Transition from S to I. Given that at time-step t node j is infectious and its neighbor node i is susceptible, the probability that j infects i (i.e., i changes its status from S to I) is given by the transmission probability per contact p (expressed in day^{-1}). As for the data-driven model, when node j infects node i at time-step t, we store this information useful for the computation of the reproduction number and the generation time.

Transition from I to R. Given that at time-step t node i is in the infectious state, it has a probability γ to recover at time-step $t + 1$.

1.2.3   Model calibration and initialization

At the beginning of each simulations we assume that the population is fully susceptible. Simulations are initialized with one infectious individual and no further introductions of index cases is assumed. We fix the removal rate γ to $1/3$ days^{-1}. We fix the per-contact transmission probability p such that $R^{index} = \frac{p\langle k \rangle}{\gamma} = 1.3$

1.3   Homogeneous mixing model

1.3.1   Network structure

We developed a homogeneous mixing model [18]. Basically, all nodes are identical and the network is fully connected (i.e., each node has $N-1$ neighbors, where N is the size of the population).

1.3.2   Infection transmission process, model calibration, and initialization

Infection transmission process, model calibration, and initialization are exactly the same for the annealed configuration model.

1.4   Quenched configuration model

1.4.1   Network structure

The structure is exactly the same of the one for the annealed configuration model; the only difference is that, instead of annealed, the network of contacts is quenched. In other words, the neighbors of all nodes remain unchanged for the entire duration of the simulation. This model misses the random component of the data-driven one, but preserve the degree and the topological correlation of the static realization of the configuration model.

1.4.2   Infection transmission process, model calibration, and initialization

Infection transmission process, model calibration, and initialization are exactly the same for the annealed configuration model.

1.5   Layer-shuffled data-driven model

1.5.1   Network structure

The structure is exactly the same of the one for the data-driven model; the only difference is that, the edges connecting different layers are shuffled, thus destroying the multiplex structure of the network.

1.5.2   Infection transmission process, model calibration, and initialization

Infection transmission process, model calibration, and initialization are exactly the same for the data-driven model.

1.6 Computation of the reproduction number

The basic reproductive number (R_0) is the average number of secondary cases generated by a typical infectious individual over the entire course of the infection in a fully susceptible population [18]. Its counterpart over time, the effective reproduction number $(R(t))$, is the number of secondary cases generated by an infectious individual at time t (and it does not require the hypothesis of a fully susceptible population) [19].

We define also the reproduction number in layer R^l as the average number of secondary infections generated by a typical infectious individuals over the entire course of the infection in layer l. From model(s) simulations we can compute R^l as

$$R^l(t) = \frac{\sum_{i \in \mathcal{I}(t)} C_l(i)}{|\mathcal{I}(t)|},$$

where

- $\mathcal{I}(t)$ represents the set of infectious nodes that acquired the infection at time t;

- $|\mathcal{I}(t)|$ represents the number of elements of set $\mathcal{I}(t)$;

- $C_l(i)$ represents the number of infections generated by infectious node i in layer l, $l \in L = \{h, s, w, c\}$.

Therefore, the overall reproduction number $R(t)$ at time t is given by

$$R(t) = \sum_{l \in L} R^l(t)$$

1.7 Computation of the generation time

The generation time Tg is defined as the average time interval between the infection time of infectors and their infectees [12]. Analogously to what we did for the reproduction number, we also define the generation time in each layer Tg^l as the average time interval between the infectors and their infectees occurred in layer l. From model(s) simulations we can compute Tg^l as

$$Tg^l(t) = \frac{\sum_{i \in \mathcal{I}(t)} \sum_{j \in \mathcal{I}'_l(i)} (\tau(j) - t)}{\sum_{i \in \mathcal{I}(t)} C_l(i)},$$

where

- $\mathcal{I}(t)$ represents the set of infectious nodes that acquired the infection at time t;

- $C_l(i)$ represents the number of infections generated by infectious node i in layer l, $l \in L$;

- $\mathcal{I}'_l(i)$ represents the set of nodes that individual i infects in layer l;

- $\tau(j)$ is the time when node j acquires the infection.

Therefore, the overall generation time $Tg(t)$ at time t is given by

$$Tg(t) = \frac{\sum_{l \in L} \sum_{i \in \mathcal{I}(t)} \sum_{j \in \mathcal{I}'_l(i)} (\tau(j) - t)}{\sum_{l \in L} \sum_{i \in \mathcal{I}(t)} C_l(i)}.$$

2 Additional Results

2.1 Variability in the initial growth of the reproduction number

As shown in the main text, we found that on average R increases over time in the early phase of the epidemic obtained by simulating the data-driven model. This is a common pattern to about 80% of all the performed simulations in the case of an "uncharacterized" future influenza pandemic in Italy (Fig. S3A) and to about 50% of all simulations in the scenario concerning the 2009 H1N1 pandemic influenza in Italy (Fig. S3B).
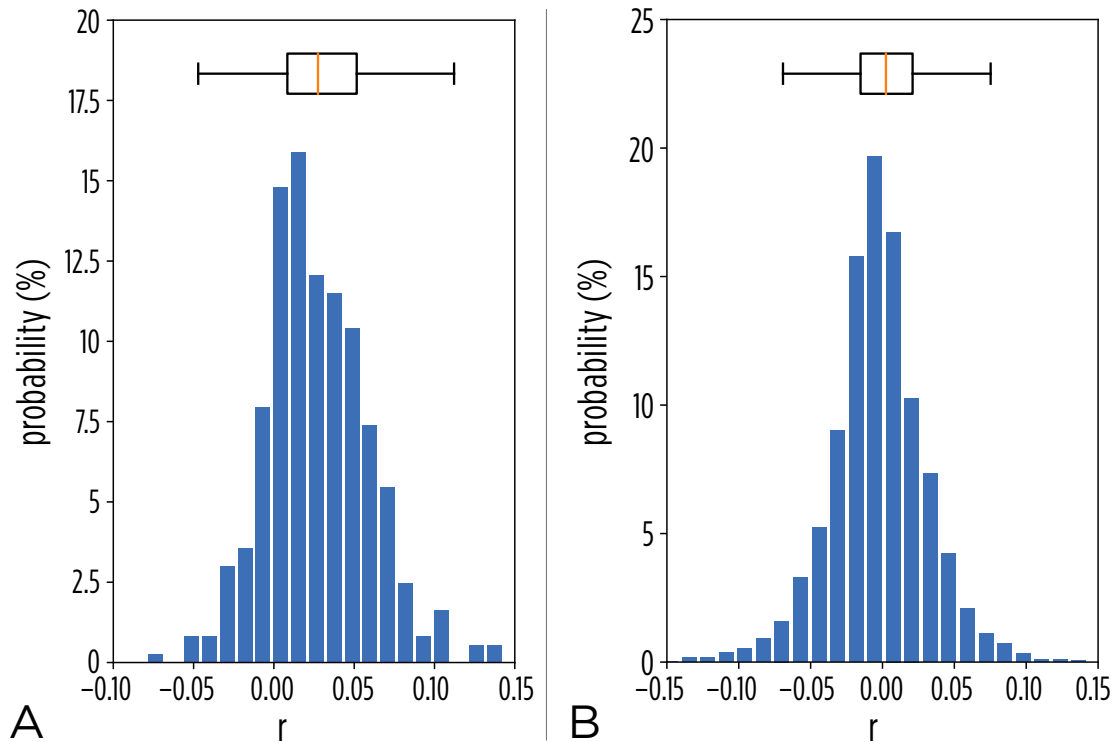

Figure S3: Probability distribution of the slope of $R(t)$ for each simulation of the data-driven model in the early epidemic phase (i.e., from 65 to 25 days before the epidemic peak). **A** Simulations refer to the spread of an "uncharacterized" future influenza pandemic in Italy. **B** Simulations refer to the spread of the 2009 H1N1 influenza pandemic in Italy.

2.2 Relation between the reproduction number and the average weighted degree of the infected nodes

We checked the relation between the reproduction number and the average weighted degree of the infected nodes by simulating the data-driven model for the scenario representing the spread of a "uncharacterized" future influenza pandemic in Italy. Results are shown in Fig.S4.
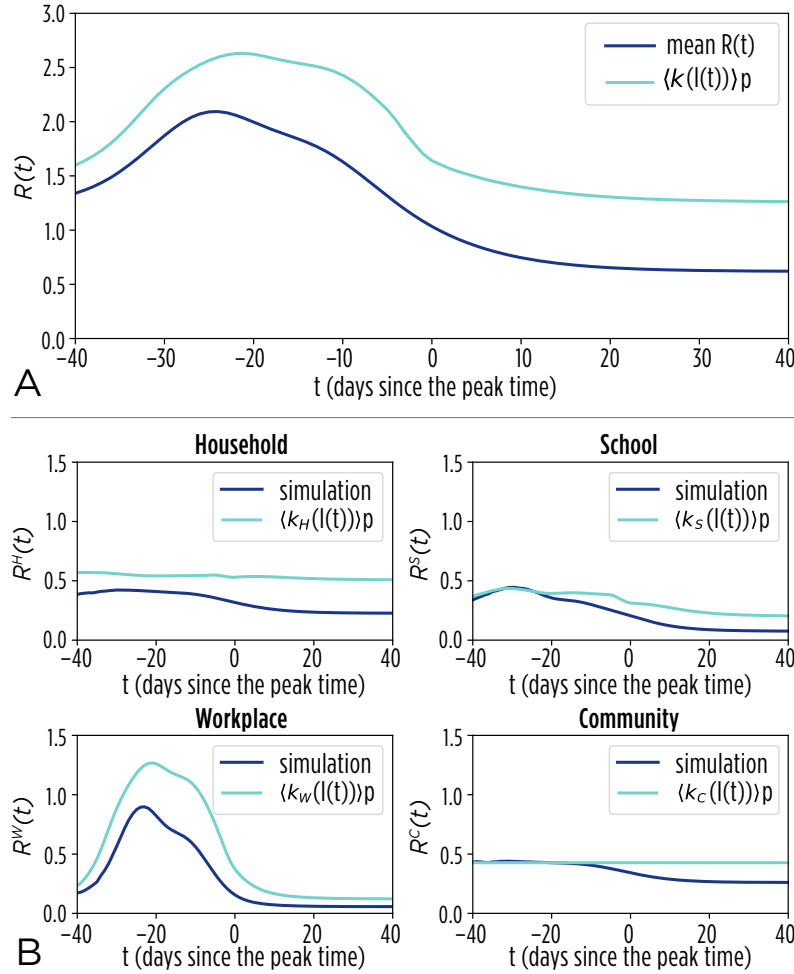

Figure S4: **A** Mean $R(t)$ of data-driven model and average weighted degree of the infected nodes multiplied by the per-contact transmission probability. **B** As A, but broken down in the four layers.

2.3   2009 H1N1 influenza pandemic in Italy

Figure S5 shows the dynamics of $R(t)$ and $Tg(t)$ as obtained by simulating the data-driven model calibrated on the 2009 H1N1 influenza pandemic in Italy.
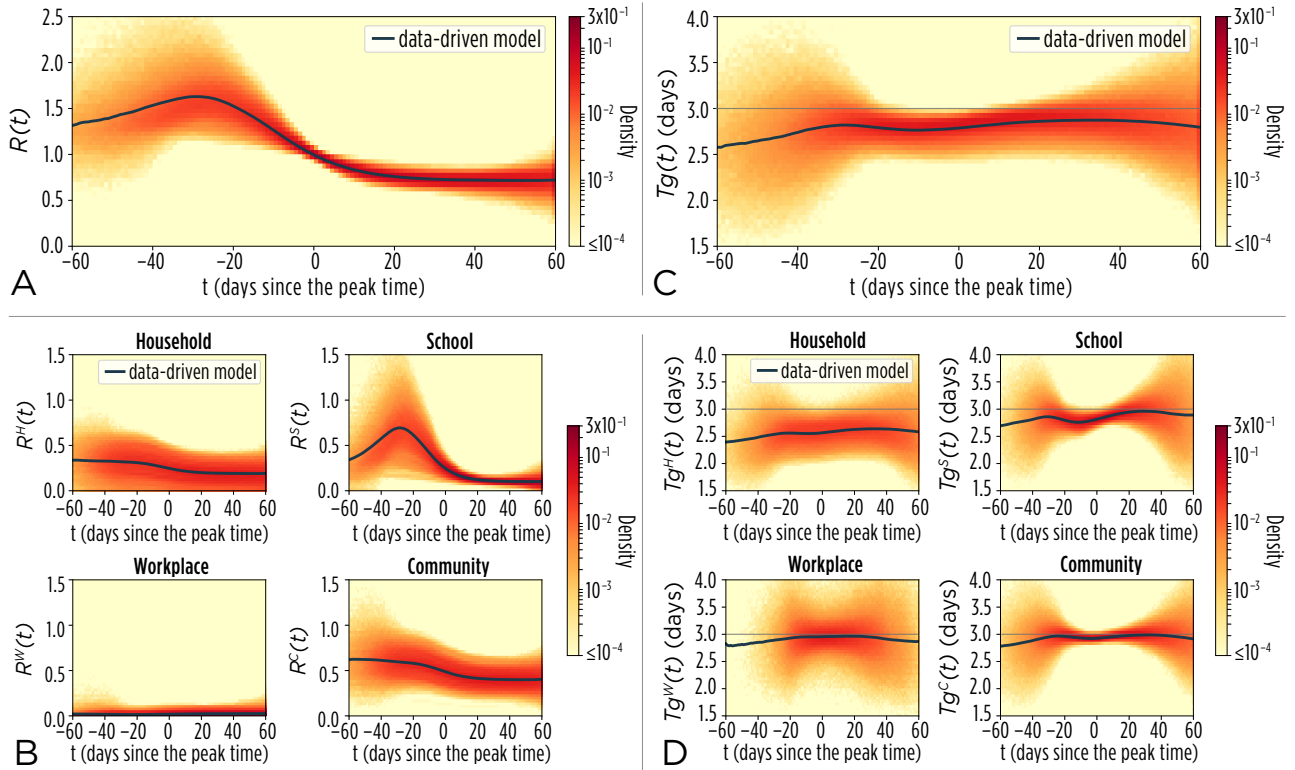

Figure S5: **A** Mean $R(t)$ of data-driven model. The colored area shows the density distribution of $R(t)$ values obtained in the single realizations of the data-driven model. Results are based on 50,000 realizations of the model. Note that only simulations with final attack rate $> 0.1\%$ are considered. Results are aligned at the epidemic peak, which corresponds to time $t = 0$. **B** As A, but $R(t)$ is broken down in the four layers. **C** As A, but for $Tg(t)$. **D** As B, but for $Tg(t)$.

2.4 Cross-country comparison of reproduction number and generation time

2.4.1 "Uncharacterized" future influenza pandemic in Italy and the Netherlands

Cross-country comparison of the dynamics of $R(t)$ and $Tg(t)$ as obtained by simulating an "uncharacterized" future influenza pandemic with the data-driven model assuming the synthetic population of Italy and of The Netherlands. For both countries we use the same parameters regulating the infection transmission process (e.g., the layer-specific transmission rates and the removal rate). Results are shown in Fig.S6.
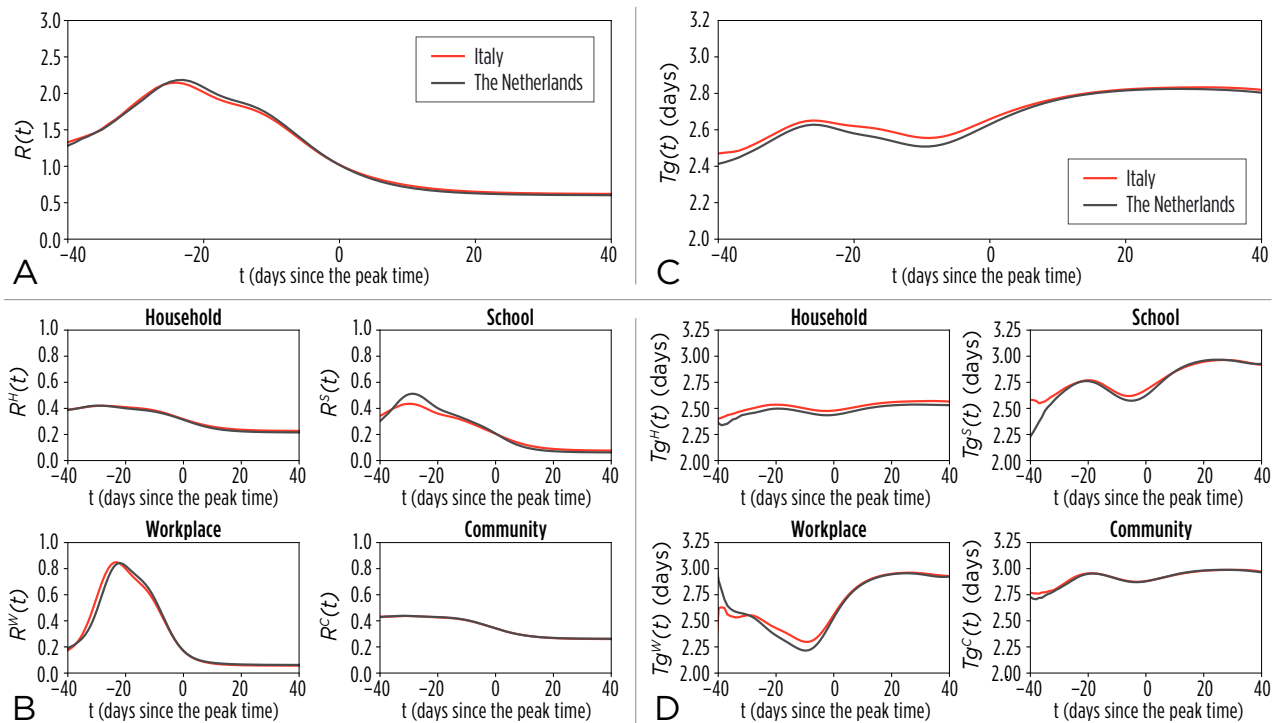

Figure S6: **A** Comparison between the mean $R(t)$ of data-driven model by using the Italian and the Dutch populations for the scenario representing the spread of a "uncharacterized" future influenza pandemic. **B** As A, but $R(t)$ is broken down in the four layers. **C** As A, but for $Tg(t)$. **D** As B, but for $Tg(t)$.

2.4.2 2009 H1N1 influenza pandemic scenario in Italy and the Netherlands

Cross-country comparison of the dynamics of $R(t)$ and $Tg(t)$ as obtained by using the data-driven model assuming the synthetic population of Italy and of The Netherlands. For both countries we use the same parameters regulating the infection transmission process (e.g., the layer-specific transmission rates and the removal rate) as obtained by calibrating the model for Italy to reproduce the spread of the 2009 H1N1 influenza pandemic in Italy. Results are shown in Fig.S7.
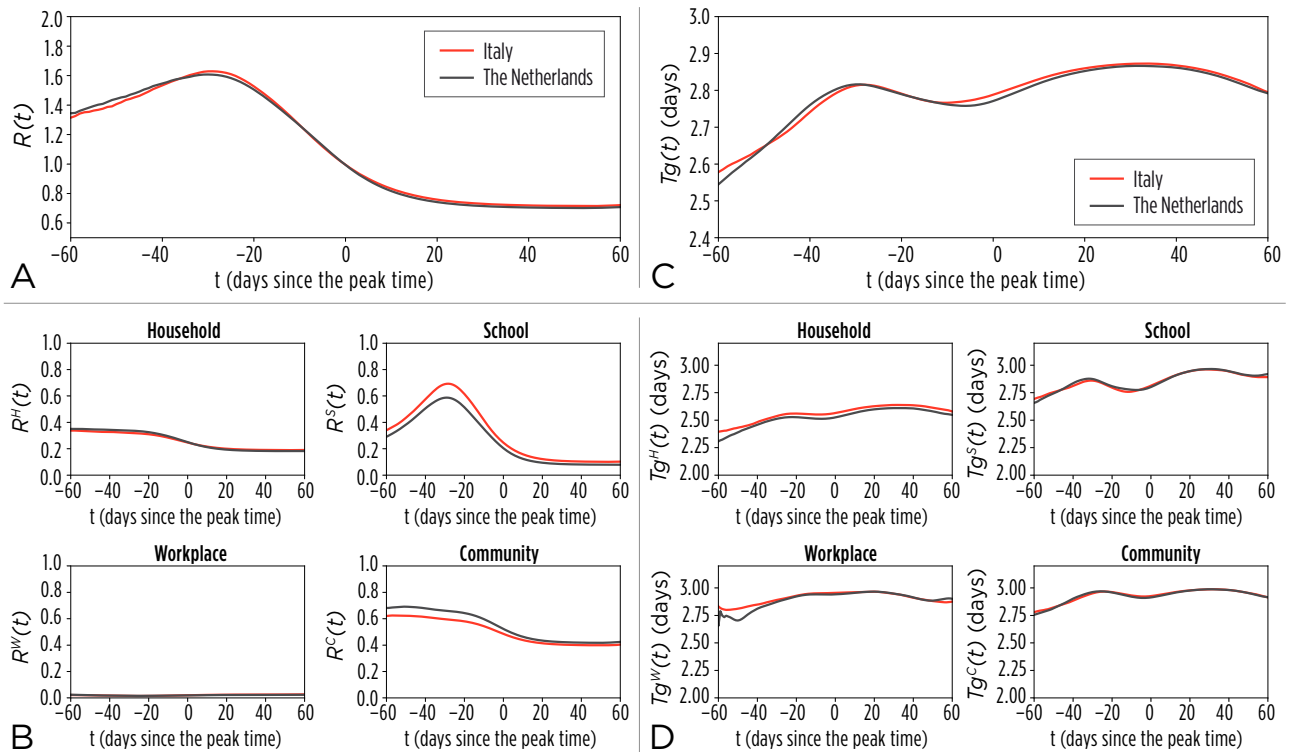

Figure S7: **A** Comparison between the mean $R(t)$ of data-driven model by using the Italian and the Dutch populations for the scenario representing the spread of a "uncharacterized" future influenza pandemic. **B** As A, but $R(t)$ is broken down in the four layers. **C** As A, but for $Tg(t)$. **D** As B, but for $Tg(t)$.

2.5 Sensitivity analysis on R^{index}

We checked the robustness of our findings with respect to changes in transmissibility of the infection, measure in terms of R^{index}. Figure S8 shows the obtained results for the scenario representing the spread of a "uncharacterized" future influenza pandemic in Italy as resulting from the simulation of the data-driven model.
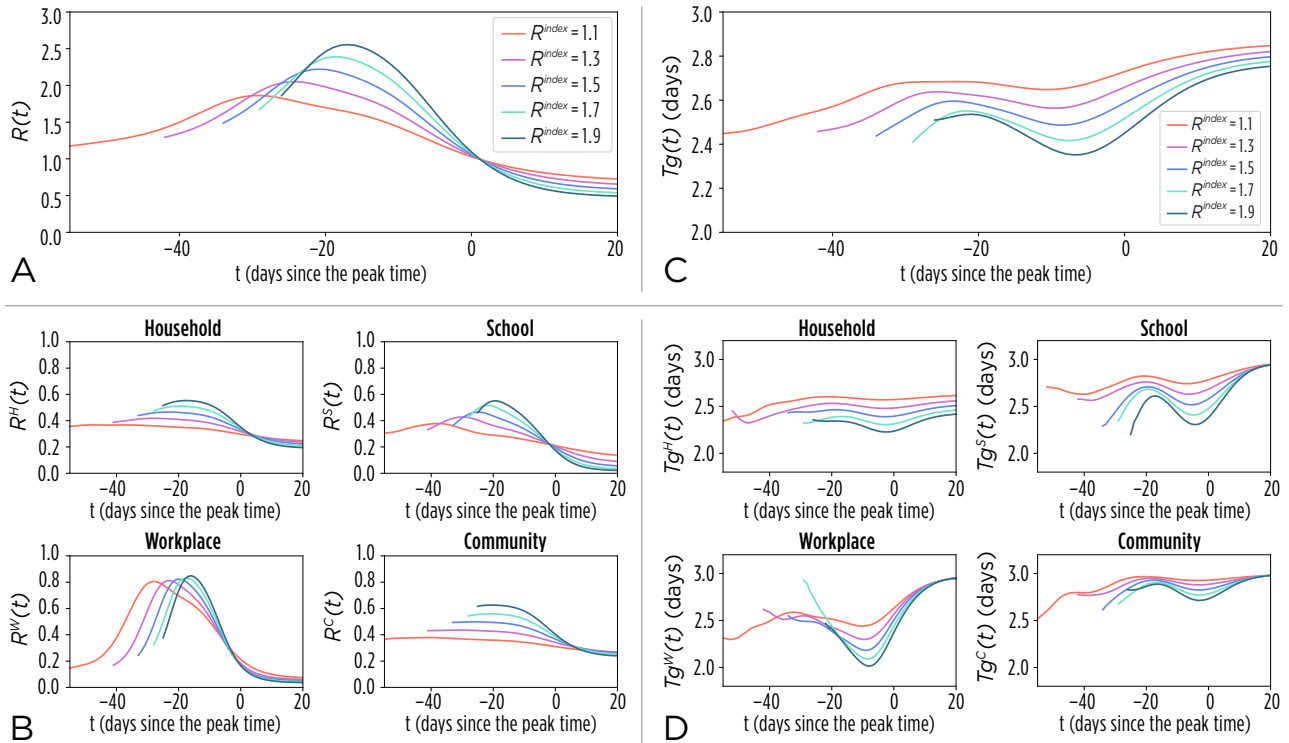

Figure S8: **A** Mean $R(t)$ of data-driven model by assuming different R^{index} values. **B** As A, but $R(t)$ is broken down in the four layers. **C** As A, but for $Tg(t)$. **D** As B, but for $Tg(t)$.

2.6 Sensitivity analysis on the distribution of the removal time

We checked the robustness of our findings with respect to changes in the distribution of the removal time. Figure S9 shows the obtained results for the scenario representing the spread of a "uncharacterized" future influenza pandemic in Italy as resulting from the simulation of the data-driven model.
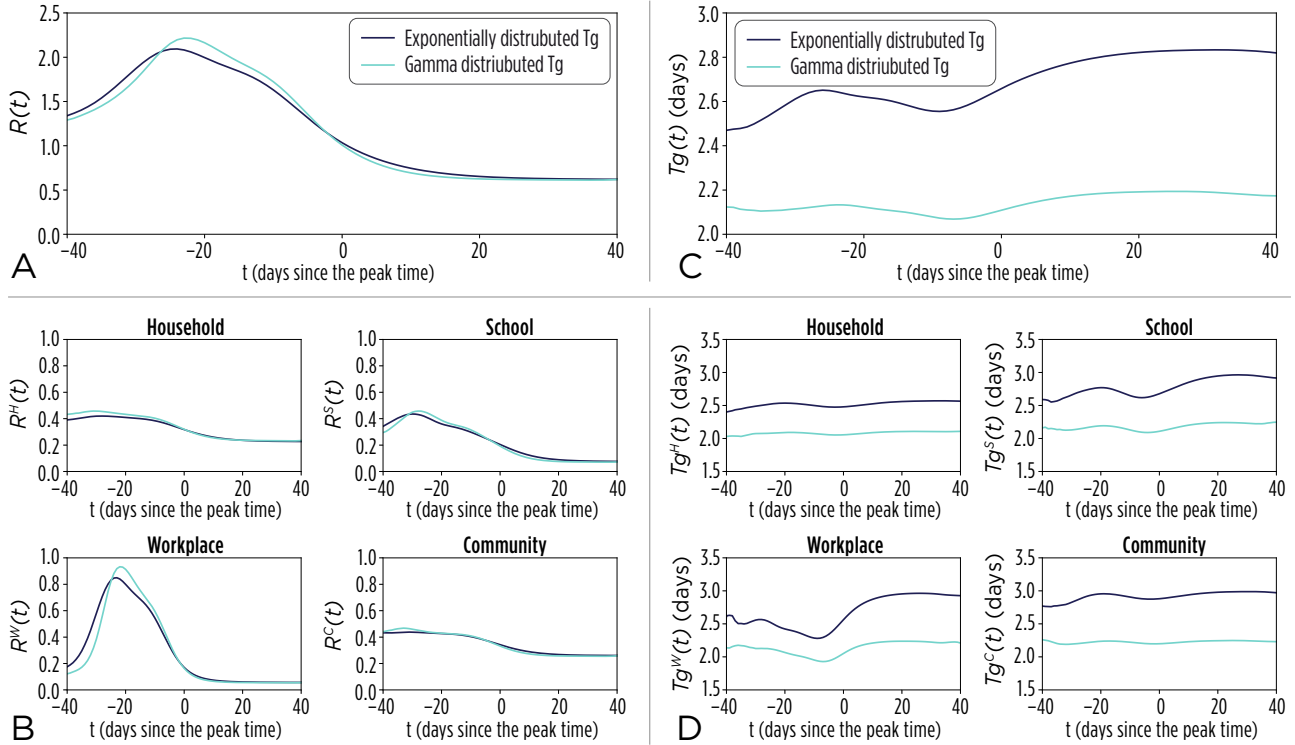

Figure S9: **A** Mean R(t) of data-driven model by assuming two different distributions of the removal time (namely exponential and gamma) whose average is equal to 3 days in both cases. Note that the gamma distribution we used has shape parameter 2 and scale parameter 1.5. **B** As A, but $R(t)$ is broken down in the four layers. **C** As A, but for $Tg(t)$. **D** As B, but for $Tg(t)$.

2.7 Comparison with null models of the contact network

To understand the main determinants of the observed patterns of $R(t)$ and Tg as resulting from the simulations of the data-driven model, we compare the data-driven model outcomes with those obtained by simulating a fully homogeneous mixing model and three alternative null models preserving the nodes' degree distribution. Specifically, the proposed null models are:

- Annealed configuration model: a model with annealed disorder and degree distribution derived from the data-driven model;

- Quenched configuration model: a model having the same network of contacts as the annealed configuration model, but with completely static connections;

- Layer shuffled data-driven model: a model having the same network of contacts on each layer but with reshuffled interlayer connections.

Details on the implementation of these models are reported in Sec. 1.2, 1.4, and 1.5 of this Supplementary Material file.

Our analysis of $R(t)$ reveals that the annealed configuration model and the layer-shuffled model exhibit behavior very close to the homogeneous mixing model (see Fig. S10A). Indeed, the same temporal trend of $R(t)$ can be observed in all layers, close to what happens for the data-driven model in the community layer (Fig. S10B). It is worth recalling that, although those models keep the degree distribution statistically equivalent to that of the data-driven model, the stochastic rewiring destroys temporal and topological correlations. The quenched configuration model, freezes the topological correlations of the model in time, and shows some deviations from the homogeneous model in terms of $R(t)$, although not as marked as in the data driven model. This suggests that the pattern observed in the data-driven model can be explained by the contact structure of the actual population.

Also, regarding Tg evolution over time, the annealed configuration model and the annealed data-driven model show the same behavior as the homogeneous mixing model (see Fig. S10C). The quenched configuration model shows a decrease of Tg only immediately before the epidemic peak, when the number of individuals that are infectious at the same time is higher and thus the competition to find susceptible nodes to infect is also higher (Fig. S10C). This is fairly different from the dynamics of the data-driven model that, mainly due to the consistently high competition in households, workplaces, and schools, shows a clear shortening of Tg since the beginning of the epidemic (see Fig. S10C and D). Similarly to what we saw for $R(t)$, none of the $Tg(t)$ patterns observed for the data-driven model are present when the configuration is annealed (Fig. S10D), although in the annealed data-driven model the contact structure is preserved.
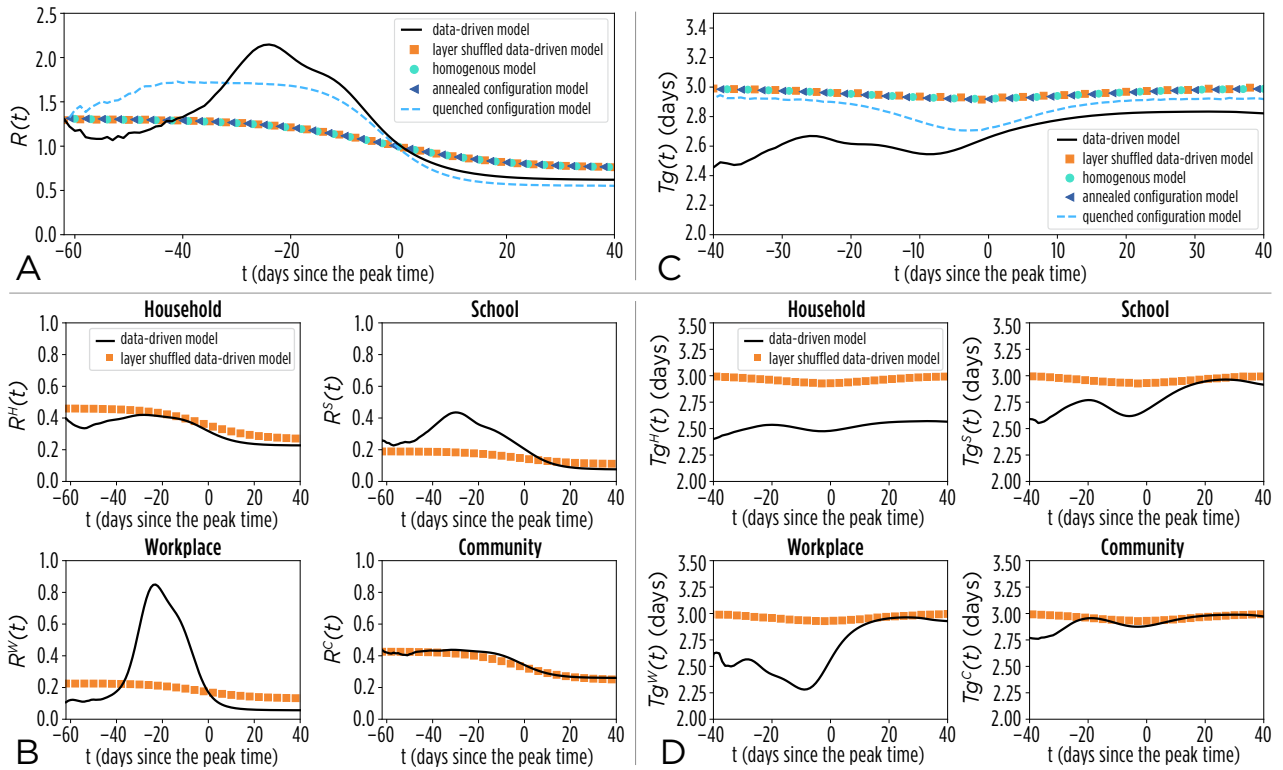
Figure S10: **A** Comparison between the mean $R(t)$ of data-driven model, of the homogeneous model, and of three null models for the scenario representing the spread of a "uncharacterized" future influenza pandemic in the Italian population. **B** As A, however $R(t)$ is broken down in the four layers. **C** As A, but for $Tg(t)$. **D** As B, but for $Tg(t)$.

2.8 Inferring $R(t)$ from incidence data

Figure S11 shows the comparison between $R(t)$ as inferred from the time-series of cases and as resulting from the microsimulation data of the transmission chain for two stochastic model realizations.
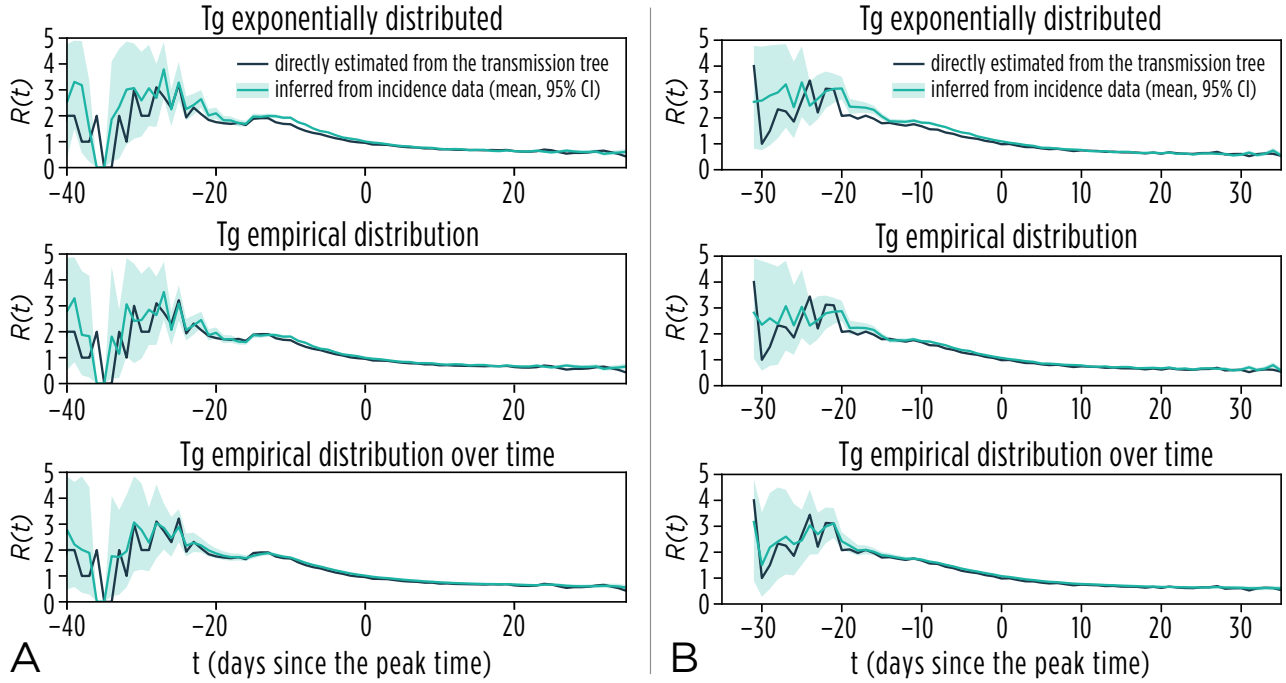

Figure S11: **A** Daily $R(t)$ as inferred from the daily incidence of new infections for one stochastic model realization. Tg is assumed to be: exponentially distributed with average 3 days (top); the distribution of Tg has been derived from the analysis of the transmission tree of the selected model simulation (middle); the distribution of Tg over time as derived from the analysis of the transmission tree of the selected model simulation (bottom). **B** as A, but for another stochastic realization. The two stochastic realizations used to produce this figure as well as Fig. 5 of the main text are randomly selected among the simulations run for scenario "Uncharacterized future influenza pandemic in Italy".

References

[1] Fumanelli L, Ajelli M, Manfredi P, Vespignani A, Merler S (2012) Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLOS Comput Biol* 8(9):e1002673.

[2] Ferguson NM et al. (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437(7056):209–214.

[3] Ferguson NM et al. (2006) Strategies for mitigating an influenza pandemic. *Nature* 442(7101):448–452.

[4] Merler S, Ajelli M, Pugliese A, Ferguson NM (2011) Determinants of the spatiotemporal dynamics of the 2009 H1N1 pandemic in Europe: implications for real-time modelling. *PLOS Comput Biol* 7(9):e1002205.

[5] Merler S et al. (2013) Pandemic influenza A/H1N1pdm in Italy: age, risk and population susceptibility. *PLOS One* 8(10):e74785.

[6] Fraser C et al. (2009) Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* 324(5934):1557–1561.

[7] Cauchemez S et al. (2009) Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *N Engl J Med* 2009(361):2619–2627.

[8] Cauchemez S et al. (2011) Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proc Natl Acad Sci USA* 108(7):2825–2830.

[9] Lau MS, Cowling BJ, Cook AR, Riley S (2015) Inferring influenza dynamics and control in households. *P Natl Acad Sci USA* 112(29):9094–9099.

[10] Marziano V, Pugliese A, Merler S, Ajelli M (2017) Detecting a Surprisingly Low Transmission Distance in the Early Phase of the 2009 Influenza Pandemic. *Sci Rep* 7(1):12324.

[11] Gilks W (2005) *Markov Chain Monte Carlo*. (Wiley Online Library).

[12] Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Royal Soc B* 274(1609):599–604.

[13] Cowling BJ, Fang VJ, Riley S, Peiris JM, Leung GM (2009) Estimation of the serial interval of influenza. *Epidemiol* 20(3):344.

[14] Cowling BJ et al. (2010) Comparative epidemiology of pandemic and seasonal influenza A in households. *N Eng J Med* 362(23):2175–2184.

[15] Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L (2014) Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infect Dis* 14(1):480.

[16] Newman ME (2002) Spread of epidemic disease on networks. *Phys Rev E* 66(1):016128.

[17] Mossong J et al. (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Med* 5(3):e74.

[18] Anderson RM, May RM, Anderson B (1991) *Infectious diseases of humans: dynamics and control*. (Oxford University Press; Oxford, UK).

[19] Wallinga J, Teunis P (2004) Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *Am J Epidemiol* 160(6):509–516.