

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}, and Alessandro Vespignani^{c,g,1}

^aWeb Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, People's Republic of China; ^bBig Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, People's Republic of China; ^cLaboratory for the Modeling of Biological and Socio-Technical Systems, Northeastern University, Boston, MA 02115; ^dBruno Kessler Foundation, 38123 Trento, Italy; ^eInstitute for Biocomputation and Physics of Complex Systems, University of Zaragoza, 50018 Zaragoza, Spain; ^fDepartment of Theoretical Physics, University of Zaragoza, 50009 Zaragoza, Spain; and ^gISI Foundation, 10126 Turin, Italy

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved October 16, 2018 (received for review June 27, 2018)

The basic reproduction number is one of the conceptual cornerstones of mathematical epidemiology. Its classical definition as the number of secondary cases generated by a typical infected individual in a fully susceptible population finds a clear analytical expression in homogeneous and stratified mixing models. Along with the generation time (the interval between primary and secondary cases), the reproduction number allows for the characterization of the dynamics of an epidemic. A clear-cut theoretical picture, however, is hardly found in real data. Here, we infer from highly detailed sociodemographic data two multiplex contact networks representative of a subset of the Italian and Dutch populations. We then simulate an infection transmission process on these networks accounting for the natural history of influenza and calibrated on empirical epidemiological data. We explicitly measure the reproduction number and generation time, recording all individual-level transmission events. We find that the classical concept of the basic reproduction number is untenable in realistic populations, and it does not provide any conceptual understanding of the epidemic evolution. This departure from the classical theoretical picture is not due to behavioral changes and other exogenous epidemiological determinants. Rather, it can be simply explained by the (clustered) contact structure of the population. Finally, we provide evidence that methodologies aimed at estimating the instantaneous reproduction number can operationally be used to characterize the correct epidemic dynamics from incidence data.

computational modeling | infectious diseases | multiplex networks | reproduction number | generation time

Mathematical and computational models of infectious diseases are increasingly recognized as relevant quantitative support to epidemic preparedness and response (1–3). Independent of the type of modeling approach, our understanding of epidemic models is generally tied to two fundamental concepts. One is the basic reproduction number R_0 , which is the average number of secondary cases generated by a typical infectious individual over the entire course of the infectious period in a fully susceptible population (4). The other is the generation time T_g , the time interval between the infection time of the infector and her/his infectees (5, 6). These quantities are the cornerstones of our understanding of basic epidemic models, as they encompass the condition for the occurrence of an epidemic outbreak ($R_0 > 1$) and the epidemic doubling time. Both the reproduction number and generation time are determined by the biological characteristics of the pathogen (e.g., probability of transmission given a contact), the pathogen–host system (e.g., timeline of pathogen replication inside the host), and the contact patterns of the population in which the infection spreads (4, 7). The concept of the reproduction number has been extended to stratified models (8) and to heterogeneous contact networks (9) to account for more complex interaction patterns. Furthermore, the definition of R_0 has been generalized by the effective reproduction number $R(t)$ (i.e., the average number of secondary cases generated by

an infectious individual at time t), thus relaxing the hypothesis of a fully susceptible population (10).

R_0 and T_g are mathematically well defined in the early stages of the epidemic in homogeneous models and are widely used in predictive approaches (10, 11). However, several studies have cautioned on the importance of the local contact structures (e.g., households, extended families, communities) in estimating the value of R_0 (12–17). Theoretical work has explored the definition of the reproduction number in models entailing community and household structure (18–24), although an operational way to compute R_0 is still lacking. For more complex models closely resembling the actual structure of the human populations (e.g., accounting for households, schools, workplaces) (25–34), the estimation of R_0 mainly relied on the direct count of secondary cases generated by the index case of the outbreak (R_0^{index}) and/or the analysis of the growth rate of the simulated epidemics. However, both methods have limitations and show marked differences in the estimated values (31, 32). Indeed, R_0^{index} is computed from the analysis of the index case of an outbreak, which is not necessarily representative of a “typical” infectious individual. In addition to this, growing evidence shows that the classic exponential early growth of the epidemic is an oversimplification often not backed by real-world data (35, 36). For this reason, statistical methods were developed for the analysis of $R(t)$, assuming that the variations of this quantity are mostly due to the impact of the performed intervention strategies and behavioral changes in the population (10, 16, 17, 37). Unfortunately, for these methods, disentangling the role of

Significance

The analysis of real epidemiological data has raised issues of the adequacy of the classic homogeneous modeling framework and quantities, such as the basic reproduction number in real-world situations. Based on high-quality sociodemographic data, here we generate a multiplex network describing the contact pattern of the Italian and Dutch populations. By using a microsimulation approach, we show that, for epidemics spreading on realistic contact networks, it is not possible to define a steady exponential growth phase and a basic reproduction number. We show the operational use of the instantaneous reproduction rate as a good descriptor of the transmission dynamics.

Author contributions: Q.-H.L., M.A., A.A., S.M., Y.M., and A.V. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence should be addressed. Email: a.vespignani@northeastern.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1811115115/-DCSupplemental.

Published online November 21, 2018.

Finally, without loss of generality, we fix the removal time to 3 d (15, 47, 48).

In *SI Appendix*, we report all of the details of the transmission model (*SI Appendix*, section 1.1). In the text, we also report the corresponding analysis of $R(t)$ and $Tg(t)$ for a stochastic implementation of a homogeneous mixing SIR model at the individual level, where all individuals are identical and are in contact with the same fixed probability. Along with the homogeneous SIR model, we also studied a set of alternative null models with an increasing level of complexity from the homogeneous model to the data-driven model (details are in *SI Appendix*, sections 1.2–1.5). Null models include annealed (edges are constantly rewired) and quenched (edges are fixed over time) configuration models. Results reported in the text refer to the synthetic population for Italy.

Effective Reproduction Number and Generation Time. The first quantities generally investigated in epidemic models are the incidence (of new infections) as a function of time and the associated growth rate of the epidemic. In homogeneous models, the number of new cases increases exponentially at a nearly constant rate r (5, 49, 50) during the early phase of the epidemic. This is not the case in the data-driven model, where we find a non-monotonous behavior: an increasing trend over the initial phase of the epidemic occurs followed by a marked decrease about 20 d before the epidemic peak (Fig. 2A). Such a result is in sharp contrast with the classic theory, where the epidemic growth rate is expected to slowly and monotonically decrease over time in the early epidemic phase (Fig. 2A). This suggests that, in contrast to simple SIR models where the basic reproduction number can readily be defined through the relation $R_0 = 1 + rTg$ (5), it is difficult to find a proper definition of the basic reproduction number in populations characterized by realistic connectivity patterns.

The daily effective reproduction number and generation time can be computed from the microsimulations by keeping track of the exact number of secondary infections generated by each individual infected at time t in the simulations (Fig. 1C). We find that $R(t)$ increases over time in the early phase of the epidemic, starting from $R_0^{\text{index}} = 1.3$ to a peak of about 2.1 (Fig. 2B). In contrast, in the homogeneous model, which lacks the typical structures of human populations, $R(t)$ is nearly constant in the early epidemic phase and then rapidly declines before the epidemic peak (Fig. 2A) as predicted by classical mathematical epidemiology theory (4). The pattern found in the data-driven model can also be partly explained by the variation of the average degree induced by the infection of individuals with a higher number of adequate contacts [an effect already observed in heavy-tailed networks (51)], thus leading to an average growth of the reproduction number. The temporal dynamics of $R(t)$ does not show a constant phase, implying that R_0 loses its meaning as a fundamental indicator in favor of $R(t)$. Although we report the results averaged over 50,000 realizations of the model, this result is supported by our analysis of the outcome of each single simulation, highlighting that the early time increase of $R(t)$ is a common pattern in the performed simulations (*SI Appendix*, section 2.1). In Fig. 2C, we show an analogous analysis of the estimated generation time in the data-driven model. We find that the generation time is considerably shorter than the duration of the infectious period (i.e., 3 d in our simulations). The estimated average Tg over the whole epidemic is 2.67 d ($\approx 11\%$ shorter than the theoretical value), with a more marked shortening right before the epidemic peak (Fig. 2C). This differs from what is predicted by the classic theory and the analysis of the homogeneous model (Fig. 2C), where the length of the infectious period corresponds to the generation time (6).

A closer look at the transmission process in the different layers of the multiplex network helps in understanding the origin

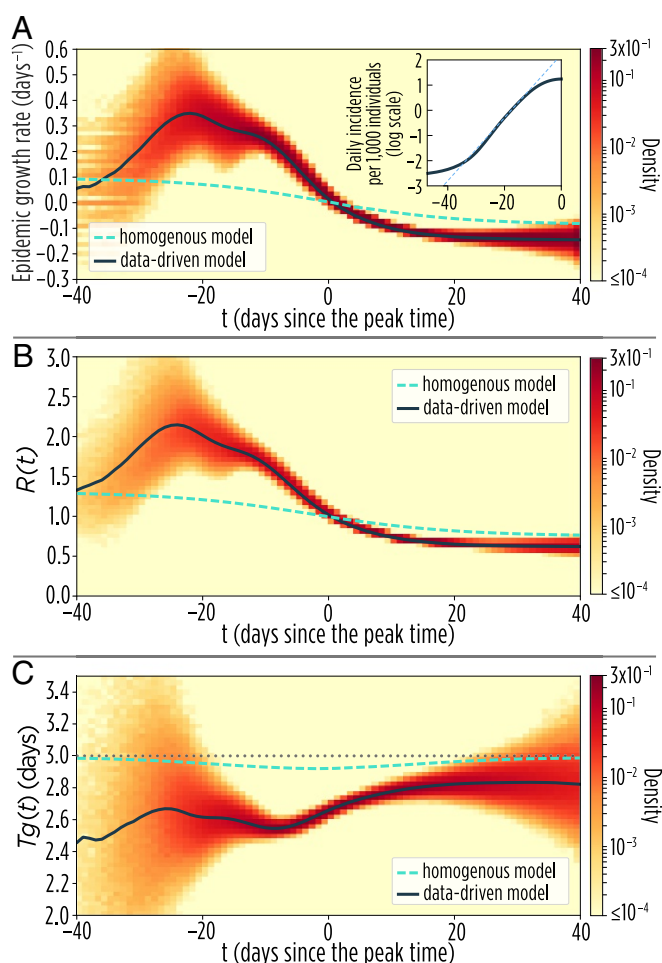


Fig. 2. Fundamental epidemiological indicators. (A) Mean daily exponential epidemic growth rate, r , over time of the data-driven and homogeneous models. The colored area shows the density distribution of $r(t)$ values obtained in the single realizations of the data-driven model. Results are based on 50,000 realizations of each model. Results are aligned at the epidemic peak, which corresponds to time $t = 0$. Inset shows the logarithm of the mean daily incidence of new influenza infections over time, which does not follow a linear trend. (B) Mean $R(t)$ of data-driven and homogeneous models. The colored area shows the density distribution of $R(t)$ values obtained in the single realizations of the data-driven model. (C) The three lines represent the mean $Tg(t)$ of data-driven and homogeneous models. The colored area shows the density distribution of $Tg(t)$ values obtained in the single realizations of the data-driven model. The horizontal dotted gray line represent the constant value of the duration of the infectious period.

of the deviations of $R(t)$ and $Tg(t)$ from the classical theory (Fig. 3). Specifically, we found that the average degree of infectious nodes as well as $R(t)$ tend to peak in the workplace layer (*SI Appendix*, section 2.2), and at least to some extent, the same happens in the school layer. However, $R(t)$ generally decreases in the household and community layers (Fig. 3A). Indeed, $R(t)$ in the household layer tends to be more uniform across the different nodes and thus, follows a general decreasing trend simply led by the depletion of susceptible contacts. We also found that, in the household layer, Tg is remarkably shorter than in all other layers (Fig. 3B)—with an average fluctuating around 2.6 d, close to the value reported by analyzing real data for household transmission (26). To provide a simple illustration of the saturation effect in households, let us consider a household of three, with one index case and two susceptible members. If, at time t , the index case infects exactly one of the two susceptibles,

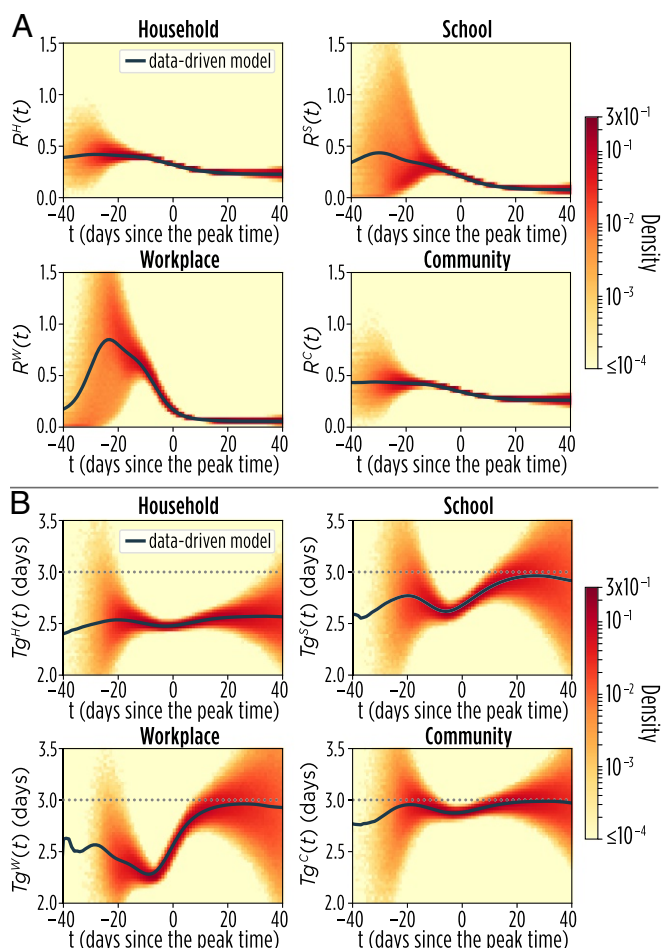


Fig. 3. Layer-specific patterns. (A) Mean $R(t)$ for the data-driven model in the four layers. The colored area shows the density distribution of $R(t)$ values obtained in the single realizations. (B) Mean $Tg(t)$ for the data-driven model in the four layers. The colored area shows the density distribution of $Tg(t)$ values obtained in the single realizations.

then at time $t + 1$, the index case has to compete with the other infectious individual to transmit the infection; the resulting generation time is shorter, simply because she/he cannot infect any other household member, although she/he may still be infectious. A similar argument was used in ref. 52 to explain why observed generation times are shorter than the infectious period. This evidence calls for considering within household competition effects when providing empirical estimates of the generation time (or serial interval) from household studies. Saturation effects are also responsible for shortening Tg in the other transmission settings, with the exception of the general community (Fig. 3B). All of the observed patterns of $R(t)$ and $Tg(t)$ are robust with respect to changes in the sociodemographic structure of the population (i.e., we simulated the infection spreading in both the Italian and Dutch synthetic populations), influenza transmission intensity (measured in terms of R^{index}), and the distribution of the removal time (we tested exponential and gamma distributions). These analyses are reported in *SI Appendix, sections 2.4–2.6*.

The 2009 H1N1 Influenza Pandemic in Italy. To test the robustness of the results in a more realistic epidemic transmission model, we used the data-driven modeling framework to model the 2009 influenza pandemic in Italy. One of the characteristic signatures of the 2009 H1N1 pandemic was the presence of a differential susceptibility by age (34, 53, 54); this is included in the model

by using values estimated for Italy as reported in the literature (55). We also consider prepandemic immunity by age in the population according to serological data (55). Vaccination is not considered, as vaccination started only during the tail of the pandemic and had a very limited uptake in Italy (vaccination coverage $< 1\%$) (56). As in the previous section, the model has four unknown parameters: the four layer-specific transmission rates. They are calibrated through a Bayesian Markov chain Monte Carlo (MCMC) approach on seroprevalence data by age collected in Italy before and after the 2009 H1N1 influenza pandemic (55). Model details are provided in *SI Appendix, section 1.1*.

The calibrated model is able to well capture the seropositive rates by age at the end of the pandemic (Fig. 4A). The estimated growth rate from the influenza-like illness (ILI) cases reported in Italy over the course of the 2009 H1N1 influenza pandemic clearly shows an increasing trend during the early phase of the epidemic followed by a sharp drop about 3 weeks before the epidemic peak (Fig. 4B). The trend observed in the data is consistent with that obtained in model simulations (Fig. 4C). Fitting a

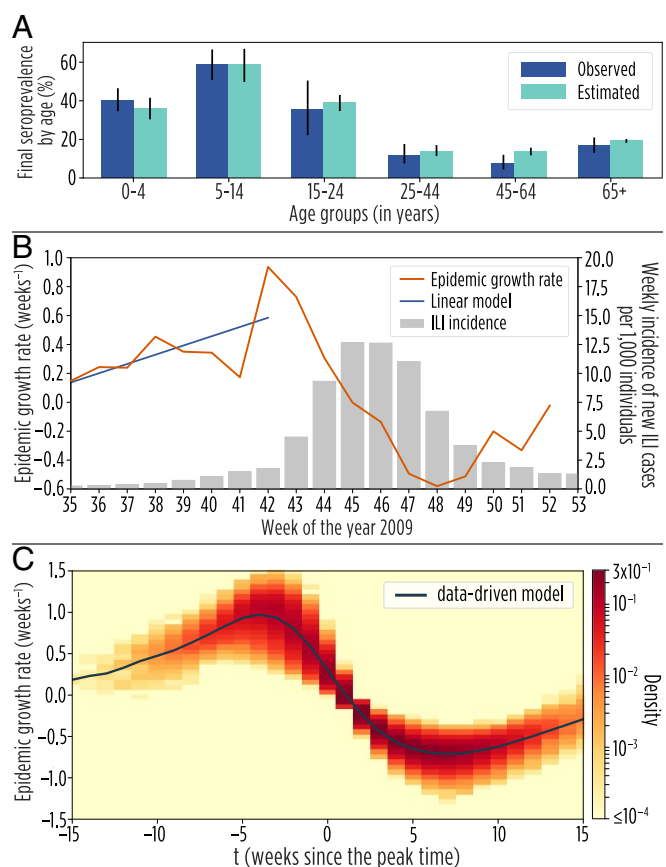


Fig. 4. The 2009 H1N1 influenza pandemic in Italy. (A) Seroprevalence rates by age as observed in a serosurvey conducted at the end of the 2009 H1N1 influenza pandemic in Italy (55) and as estimated by the calibrated model. (B) Epidemic growth rate over time $r(t)$ as estimated from the weekly incidence of new ILI cases in Italy over the course of the 2009 H1N1 influenza pandemic and the best-fitting linear model from week 35 to week 41 in 2009 (scale on the left axis). Weekly incidence of new ILI cases in Italy over the course of the 2009 H1N1 pandemic (scale on the right axis). Data are available at the ISS Influnet website (old.iss.it/flue/). Note that, over the period from week 35 to week 51 in 2009, schools were regularly open in Italy. (C) Temporal pattern of the mean weekly exponential epidemic growth rate (r) resulting from the analysis of the data-driven model calibrated on the 2009 H1N1 influenza seroprevalence data. The colored area shows the density distribution of $r(t)$ values obtained in the single realizations of the data-driven model.

$$L = \prod_{t=1}^T P\left(C(t), R(t) \sum_{s=1}^t \phi(s) C(t-s)\right),$$

where here, $P(k, \lambda)$ is the probability mass function of a Poisson distribution (i.e., the probability of observing k events if these events occur with a known rate λ). The posterior distribution of $R(t)$ is then explored using MCMC sampling.

1. Viboud C, et al. (2018) The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics* 22:13–21.
2. Biggerstaff M, et al. (2016) Results from the Centers for Disease Control and Prevention's Predict the 2013–2014 Influenza Season Challenge. *BMC Infect Dis* 16:357.
3. Chretien JP, Riley S, George DB (2015) Mathematical modeling of the West Africa Ebola epidemic. *eLife* 4:e09186.
4. Anderson RM, May RM, Anderson B (1991) *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ Press, Oxford).
5. Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Biol Sci* 274:599–604.
6. Roberts M, Heesterbeek J (2007) Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *J Math Biol* 55: 803–816.
7. Lessler J, et al. (2009) Incubation periods of acute respiratory viral infections: A systematic review. *Lancet Infect Dis* 9:291–300.
8. Diekmann O, Heesterbeek JAP, Metz JA (1990) On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J Math Biol* 28:365–382.
9. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A (2015) Epidemic processes in complex networks. *Rev Mod Phys* 87:925–979.
10. Wallinga J, Teunis P (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol* 160:509–516.
11. WHO Ebola Response Team (2014) Ebola virus disease in West Africa: The first 9 months of the epidemic and forward projections. *N Engl J Med* 371:1481–1495.
12. Cintrón-Arias A, Castillo-Chávez C, Bettencourt LM, Lloyd AL, Banks H (2009) The estimation of the effective reproductive number from disease outbreak data. *Math Biosci Eng* 6:261–282.
13. Burr T, Chowell G (2009) The reproduction number $R(t)$ in structured and nonstructured populations. *Math Biosci Eng* 6:239–259.
14. Britton T, Tomba GS (2018) Estimation in emerging epidemics: Biases and remedies. arXiv:1803.01688.
15. 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* 108:2825–2830.
16. Ajelli M, et al. (2015) The 2014 Ebola virus disease outbreak in Pujehun, Sierra Leone: Epidemiology and impact of interventions. *BMC Med* 13:281.
17. Ajelli M, et al. (2016) Spatiotemporal dynamics of the Ebola epidemic in Guinea and implications for vaccination and disease elimination: A computational modeling analysis. *BMC Med* 14:130.
18. Pellis L, et al. (2015) Eight challenges for network epidemic models. *Epidemics* 10:58–62.
19. Becker NG, Dietz K (1995) The effect of household distribution on transmission and control of highly infectious diseases. *Math Biosci* 127:207–219.
20. Fraser C (2007) Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One* 2:e758.
21. House T, Keeling MJ (2008) Deterministic epidemic models with explicit household structure. *Math Biosci* 213:29–39.
22. Goldstein E, et al. (2009) Reproductive numbers, epidemic spread and control in a community of households. *Math Biosci* 221:11–25.
23. Ross JV, House T, Keeling MJ (2010) Calculation of disease dynamics in a population of households. *PLoS One* 5:e9666.
24. Pellis L, Ferguson NM, Fraser C (2011) Epidemic growth rate and household reproduction number in communities of households, schools and workplaces. *J Math Biol* 63:691–734.
25. Longini IM, et al. (2005) Containing pandemic influenza at the source. *Science* 309:1083–1087.
26. Ferguson NM, et al. (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437:209–214.
27. Ferguson NM, et al. (2006) Strategies for mitigating an influenza pandemic. *Nature* 442:448–452.
28. Ciofi Degli Atti ML, et al. (2008) Mitigation measures for pandemic influenza in Italy: An individual based model considering different scenarios. *PLoS One* 3:e1790.
29. Halloran ME, et al. (2008) Modeling targeted layered containment of an influenza pandemic in the United States. *Proc Natl Acad Sci USA* 105:4639–4644.
30. Fumanelli L, Ajelli M, Merler S, Ferguson NM, Cauchemez S (2016) Model-based comprehensive analysis of school closure policies for mitigating influenza epidemics and pandemics. *PLoS Comput Biol* 12:e1004681.
31. Ajelli M, Merler S (2008) The impact of the unstructured contacts component in influenza pandemic modeling. *PLoS One* 3:e1519.
32. Germann TC, Kadau K, Longini IM, Macken CA (2006) Mitigation strategies for pandemic influenza in the United States. *Proc Natl Acad Sci USA* 103:5935–5940.
33. Merler S, Ajelli M (2010) The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc Biol Sci* 277:557–565.
34. 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:e1002205.
35. Chowell G, Viboud C, Simonsen L, Moghadas SM (2016) Characterizing the reproduction number of epidemics with early subexponential growth dynamics. *J R Soc Interface* 13:20160659.
36. Viboud C, Simonsen L, Chowell G (2016) A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics* 15:27–37.
37. Lipsitch M, et al. (2003) Transmission dynamics and control of severe acute respiratory syndrome. *Science* 300:1966–1970.
38. Kivela M, et al. (2014) Multilayer networks. *J Complex Netw* 2:203–271.
39. De Domenico M, et al. (2013) Mathematical formulation of multilayer networks. *Phys Rev X* 3:041022.
40. Cozzo E, Ferraz de Arruda G, Rodrigues F, Moreno Y (2018) *Multiplex Networks: Basic Formalism and Structural Properties*, Springer Briefs in Complexity (Springer, Cham, Switzerland).
41. Ajelli M, Poletti P, Melegaro A, Merler S (2014) The role of different social contexts in shaping influenza transmission during the 2009 pandemic. *Sci Rep* 4:7218.
42. Cauchemez S, Carrat F, Viboud C, Valleron A, Boelle P (2004) A Bayesian MCMC approach to study transmission of influenza: Application to household longitudinal data. *Stat Med* 23:3469–3487.
43. Mossong J, et al. (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 5:e74.
44. 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:e1002673.
45. Ajelli M, Litvinova M (2017) Estimating contact patterns relevant to the spread of infectious diseases in Russia. *J Theor Biol* 419:1–7.
46. 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* 14:480.
47. Cowling BJ, Fang VJ, Riley S, Peiris JM, Leung GM (2009) Estimation of the serial interval of influenza. *Epidemiology* 20:344.
48. Cowling BJ, et al. (2010) Comparative epidemiology of pandemic and seasonal influenza A in households. *N Engl J Med* 362:2175–2184.
49. Chowell G, Hengartner NW, Castillo-Chavez C, Fenimore PW, Hyman JM (2004) The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda. *J Theor Biol* 229:119–126.
50. Nishiura H, Chowell G, Safan M, Castillo-Chavez C (2010) Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009. *Theor Biol Med Model* 7:1.
51. Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A (2004) Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys Rev Lett* 92:178701.
52. Scalia Tomba G, Svensson Å, Asikainen T, Giesecke J (2010) Some model based considerations on observing generation times for communicable diseases. *Math Biosci* 223:24–31.
53. Fraser C, et al. (2009) Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* 324:1557–1561.
54. Cauchemez S, et al. (2009) Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *N Engl J Med* 2009:2619–2627.
55. Merler S, et al. (2013) Pandemic influenza A/H1N1pdm in Italy: Age, risk and population susceptibility. *PLoS One* 8:e74785.
56. Rizzo C, et al. (2010) Response to the 2009 influenza a (H1N1) pandemic in Italy. *Euro Surveill* 15:19744.
57. Bettencourt LM, Ribeiro RM (2008) Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One* 3:e2185.
58. White LF, et al. (2009) Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA. *Influenza Other Respir Viruses* 3:267–276.
59. Nguyen VK, Rojas CP, Hernandez-Vargas E (2018) The 2017 plague outbreak in Madagascar: Data descriptions and epidemic modelling. bioRxiv:10.1101/247569.
60. Poletti P, Ajelli M, Merler S (2011) The effect of risk perception on the 2009 H1N1 pandemic influenza dynamics. *PLoS One* 6:e16460.