Multiscale mobility networks and the spatial spreading of infectious diseases

Duygu Balcan[a,b], Vittoria Colizza[c], Bruno Gonçalves[a,b], Hao Hu[d], José J. Ramasco[b], and Alessandro Vespignani[a,b,c,1]

[a]Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, IN 47408; [b]Pervasive Technology Institute, Indiana University, Bloomington, IN 47404; [c]Computational Epidemiology Laboratory, Institute for Scientific Interchange Foundation, 10133 Torino, Italy; and [d]Department of Physics, Indiana University, Bloomington, IN 47406

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved October 13, 2009 (received for review June 19, 2009)

Among the realistic ingredients to be considered in the computational modeling of infectious diseases, human mobility represents a crucial challenge both on the theoretical side and in view of the limited availability of empirical data. To study the interplay between short-scale commuting flows and long-range airline traffic in shaping the spatiotemporal pattern of a global epidemic we (*i*) analyze mobility data from 29 countries around the world and find a gravity model able to provide a global description of commuting patterns up to 300 kms and (*ii*) integrate in a worldwide-structured metapopulation epidemic model a timescale-separation technique for evaluating the force of infection due to multiscale mobility processes in the disease dynamics. Commuting flows are found, on average, to be one order of magnitude larger than airline flows. However, their introduction into the worldwide model shows that the large-scale pattern of the simulated epidemic exhibits only small variations with respect to the baseline case where only airline traffic is considered. The presence of short-range mobility increases, however, the synchronization of subpopulations in close proximity and affects the epidemic behavior at the periphery of the airline transportation infrastructure. The present approach outlines the possibility for the definition of layered computational approaches where different modeling assumptions and granularities can be used consistently in a unifying multiscale framework.

complex networks | computational epidemiology | human mobility | multiscale phenomena

Computational approaches to the realistic modeling of spatial epidemic spread make use of a wide array of simulation schemes (1) ranging from very detailed agent-based approaches (2–6) to structured metapopulation models based on data-driven mobility schemes at the interpopulation level (7–10). All these approaches integrate a wealth of real-world data. However, it is not yet clear how to discriminate the effects of the inclusion/lack of real-world features in specific models. This limitation is mainly related to our incomplete knowledge of human interactions and mobility processes, which are fundamental aspects to describe a disease spread. Although recent efforts started to make available massive data on human mobility from different sources and at different levels of description (11–20), the multiscale nature of human mobility is yet to be comprehensively explored. Human mobility can be generally described by defining a network of interacting communities where the connections and the corresponding intensity represent the flow of people among them (13, 14). Global mobility flows therefore form very complex multiscale networks (21) spanning several orders of magnitude in intensity and spatiotemporal scales ranging from the long-range intercontinental air traffic (13, 15) to the short range commuting flows (17–19). A multitude of heuristic models for population structure and mobility patterns have been proposed, but they all depend on the specific mobility process under consideration (22, 23). The limited understanding of the interrelations among the multiple scales entailed in human mobility and their impact on the definition of epidemic patterns constitute a major road block in the development of predictive large-scale data driven epidemic models. In this context, two questions stand out: (*i*) Is there a most relevant mobility scale in the definition of the global epidemic pattern? and (*ii*) At which level of resolution of the epidemic behavior does a given mobility scale become relevant, and to what extent?

To begin addressing these questions, we use high-resolution worldwide population data that allow for the definition of subpopulations according to a Voronoi decomposition of the world surface centered on the locations of International Air Transport Association (IATA)-indexed airports (www.iata.org). We have then gathered data on the commuting patterns of 29 countries in five continents, constructing short-range commuting networks for the defined subpopulations. Extensive analysis of these networks allows us to draw a general gravity law for commuting flows that reproduces commuting patterns worldwide. This law, valid at the scale defined by the tessellation process, is statistically stable across the world because of the globally homogeneous procedure applied to build the subpopulations around transportation hubs. The multiscale networks we obtain are integrated into the global epidemic and mobility (GLEaM) model, a computational platform that uses a metapopulation stochastic model on a global scale to simulate the large-scale spreading of influenza-like illnesses (ILI). To fully consider the effect of multiscale mobility processes in the disease dynamics, we develop a timescale-separation technique for evaluating the force of infection due to different mobility couplings and simulate global pandemics with tunable reproductive ratios. The results obtained from the full multiscale mobility network are compared with the simulations in which only the large-scale coupling of the airline transportation network is included. Our analysis shows that although commuting flows are, on average, one order of magnitude larger than the long-range airline traffic, the global spatiotemporal patterns of disease-spreading are mainly determined by the airline network. Short-range commuting interactions have, on the other hand, a role in defining a larger degree of synchronization of nearby subpopulations and specific regions, which can be considered weakly connected by the airline transportation system. In particular, it is possible to show that short-range mobility has an impact in the definition of the subpopulation infection hierarchy. The techniques developed here allow for an initial understanding of the level of data integration required to obtain reliable results in large-scale modeling of infectious diseases.

Results and Discussion

Simulations of worldwide epidemic spread are generally based on structured metapopulation models that consider data-driven schemes for long-range mobility at the interpopulation level coupled with coarse-grained techniques within each subpopulation

Author contributions: D.B., V.C., B.G., H.H., J.J.R., 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.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: alexv@indiana.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0906910106/DCSupplemental.

Downloaded from https://www.pnas.org by 37.19.197.239 on October 16, 2023 from IP address 37.19.197.239.

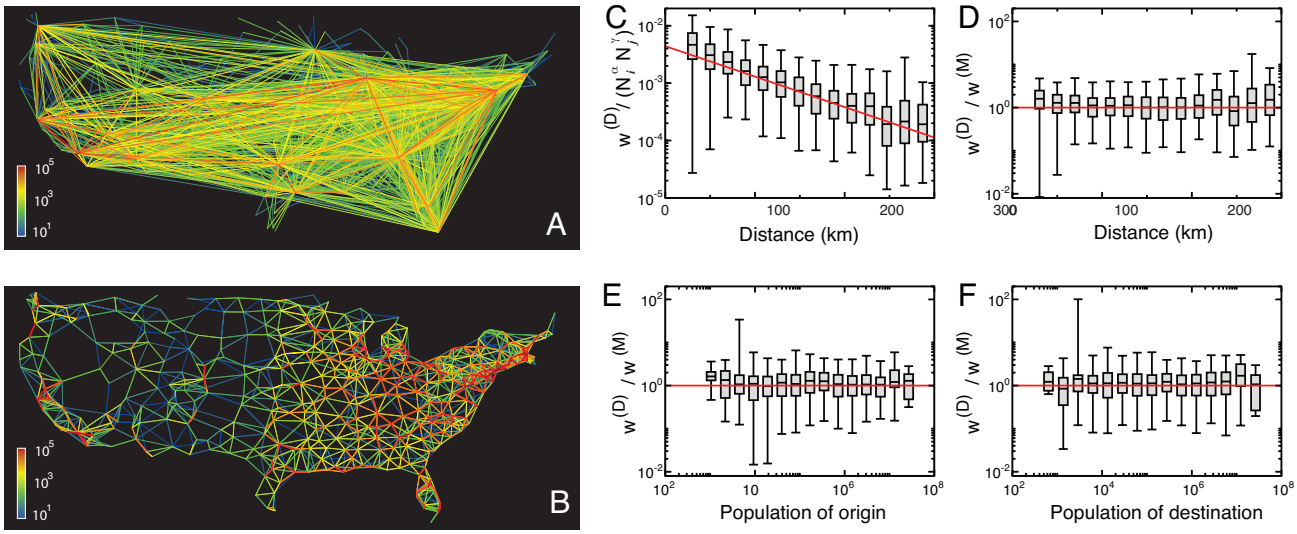


Fig. 1. Multiscale mobility networks and gravity law fit. (*A*) Continental U.S. airline transportation network. (*B*) Continental U.S. commuting network. The width and color (from blue to red) of the edges represent on a logarithmic scale the intensity of the mobility flow. (*C*) Commuting flux obtained from data ($w^{(D)}$) rescaled by the gravity law's dependence on origin and destination populations ($N_i^{\alpha}N_j^{\gamma}$), as a function of the distance between subpopulations. The number of people commuting between different urban areas decreases exponentially with distance up to 300 kms. (*D–F*) Ratio of commuting flux obtained from data ($w^{(D)}$) to corresponding commuting flux predicted by the gravity model with fitted parameters ($w^{(M)}$), as a function of distance, population of origin and population of destination, respectively. The three plots provide values spread $\approx$1, showing that the synthetic networks generated by the functional form (see Table 1) reproduce well the commuting fluxes obtained from data. Solid lines in all frames are guides to the eye.

(7–10, 24–26). In this paper, we use the GLEaM computational scheme based on a georeferenced metapopulation approach. The model consists of three data layers. The population and mobility layers allows the partition of the world into geographical census regions coupled by population movements. This partition defines the subpopulation network where the connections between subpopulations represent the fluxes of individuals due to the transportation infrastructures and mobility patterns. Superimposed on this subpopulation network is the epidemic layer that defines inside each subpopulation the disease dynamics that depends on the specific etiology of the disease considered (see *Material and Methods*).

Multiscale Mobility Networks. The basic structure of GLEaM is based on high-resolution population data[†] that estimates the population with a resolution given by cells of 15 $\times$ 15 minutes of arc, covering the whole planet. This population data allows the construction of Voronoi tassels around transportation hubs in the world, defining the subpopulations structure of the metapopulation model (see *SI Appendix*). In particular, we identify 3,362 subpopulations centered around IATA airports in 220 different countries. The air-traffic network among the defined subpopulations is obtained from the IATA databases that contain the list of worldwide airport pairs connected by direct flights and the number of available seats on any given connection. The high level of geographical resolution of the subpopulation database enables us to integrate also the mobility flows due to commuting patterns between subpopulations (see *Material and Methods*) and construct the corresponding commuting network. The main difficulty in defining a commuting network worldwide is the lack of a global database as opposed to the case of the air-traffic flow. Data are scattered in different national and international databases that use different administrative and geographical granularities, and several definitions of commuting flows. We have collected commuting data from 29 countries (a full list of countries and the database properties are reported in the *SI Appendix*) in five different continents. Each dataset was mapped into the GLEaM Voronoi tessellation constructing the commuting networks at the subpopulation level.

In Fig. 1, we show the commuting network of the continental U.S. as obtained by mapping the county commuting data onto the subpopulations used by GLEaM. Commuting data do not consider airline flows that are accounted for by the IATA dataset. On the same scale, we also report the airline traffic network, readily highlighting the difference in scale and spatial structure of the two networks. The commuting network appears as an almost grid-like lattice connecting neighboring subpopulations, whereas the airline traffic network is dominated by long range connections. The wide range of scales is evident also in the intensities of the mobility flows, spanning several orders of magnitude, with the average commuting flow being one order of magnitude larger than the average airline traffic flow. Finally, it should be noted that, in general, commuting flows refer to round trip processes with a characteristic time of the order of 1/3 day (average duration of a work day) compared with much longer characteristic times for airline travel (average value around two weeks at the end.[‡]

To gain general insight on the commuting flow, we use the general gravity model from transportation theory (22, 23) as a starting point. This model assumes that the commuting flow w_{ij} between subpopulation i (with population N_i) and subpopulation j (with population N_j) takes on the form:

$$w_{ij} = C\,\frac{N_i^{\alpha}N_j^{\gamma}}{f(d_{ij})}, \qquad [1]$$

where C is a proportionality constant, α and γ tune the dependence with respect to each subpopulation size, and $f(d_{ij})$ is a distance-dependent functional form. Gravity laws usually consider power or exponential laws for the behavior of f(d_{ij}). The results reported in the literature are variable and generally depend on the way the subpopulations are defined. In our case,

[†]The Gridded Population of the World and The Global Rural-Urban Mapping Projects, Socioeconomic Data and Applications Center of Columbia University, http://sedac.ciesin.columbia.edu/gpw.

[‡]Travel Trends 2007, Office for National Statistics, www.statistics.gov.uk.

Downloaded from https://www.pnas.org by 37.19.197.239 on October 16, 2023 from IP address 37.19.197.239.

Table 1. Exponents of the gravity law as obtained by applying a multivariate analysis to global commuting data

d (km)	α	γ	r (km)
≤300	0.46 ± 0.01	0.64 ± 0.01	82 ± 2
>300	0.35 ± 0.06	0.37 ± 0.06	N/A

we can take advantage of the statistical similarity of the subpopulations centered on major transportation hubs. We tested the gravity law by the statistical analysis of more than 10^4 flows worldwide, and we found that the best fit is obtained by using an exponential function $f(d_{ij}) = \exp(d_{ij}/r)$, where r is the characteristic length that governs the decay of commuting flows. In Table 1, we report the estimated values for the exponents α, γ, and r.

Noticeably, we can validate the gravity law at the level of geographical area or country as shown in Fig. 1, where the commuting flows obtained from the data are compared with the synthetic ones predicted by the model, as functions of the various variables of the gravity law. In the *SI Appendix*, we report similar tests at the level of geographical areas and countries, that confirm the generality and goodness of the obtained law. It is worth remarking that although the present gravity law works well at the granularity defined by our Voronoi decomposition, these results cannot be extrapolated to different granularities. For instance, we notice that the exponents obtained by our approach are quite close to those obtained by Viboud et al. (17), although the spatial decay has a completely different form. We do not see any obvious reason for the observed scaling form, but it must be noted that we are working with subpopulations, which are by construction statistically more homogeneous and of larger size than the county level used in ref. 17. Administrative regions might indeed impose boundaries that define subpopulations not clearly associated to centers of gravity for mobility processes, as e.g., large urban areas cut in multiple counties. In general, both approaches are usually tested in transportation theory (as in ref. 19), and in some cases the exponential function was found to better fit migration phenomena (27).

The above gravity law allows us to work with two different worldwide commuting networks. An entirely synthetic one, generated by using the gravity law fitted to the empirical data, and one integrating the empirical data. The synthetic network considers only neighboring subpopulations. We have empirically observed that second-neighbor interactions are on average one order of magnitude smaller than the nearest-neighbor interactions and their contribution can be neglected in the first-order estimation of the effective force of infection. In the following, we will report the results obtained only with the synthetic network and we leave to the *SI Appendix* a demonstration that no significant differences are observed when we compare the results obtained by using the synthetic and the real data networks.

Epidemic Simulations. To study the effect that commuting networks have on the overall spread of an emerging disease, we consider the simulation of an ILI and compare the results with a simulation in which we include only airline traffic as in previous works (10). For the sake of clarity, in the following we compare the results for the simulations of a hypothetical pandemic influenza with $R_0 = 1.9$ starting in Hanoi on April 1 (the *SI Appendix* also reports results for October 1). The model includes a seasonal dependence of the transmission. In the *SI Appendix* file, we report simulations of the 2001–2002 seasonal influenza timeline and compare those with the real surveillance data. The simulations of the realistic model and comparison with real data confirm the analysis presented here for the synthetic pandemic model.

GLEaM is fully stochastic and takes into account the discrete nature of individuals both in the travel coupling and in the compartmental transitions. The transmission model within each urban area follows a compartmentalization specific to the disease under study. Here we use the classic ILI compartmentalization in which each individual is classified by one of the discrete states such as susceptible (S), latent (L), infectious (I), and permanently recovered (R). Infectious persons are further subdivided into asymptomatic, symptomatic traveling, and symptomatic nontraveling as detailed in *Material and Methods*. The discrete nature of individuals is implemented by introducing binomial and multinomial processes inside each urban area for the stochastic evolution of the infection. A detailed exposition of the stochastic approach to disease evolution is provided in the *SI Appendix*.

In GLEaM, the airline mobility is integrated explicitly as individuals are allowed to travel from one subpopulation to another by means of the airline transportation network (28) similarly to the models in refs. 7 and 8 and the stochastic generalizations of ref. 9. In each subpopulation j, the number of individuals is $N_j(t)$ and $X_j^{[m]}(t)$ is the number of individuals in compartment $[m]$ of the disease evolution at time t; therefore $N_j(t) = \Sigma_m X_j^{[m]}(t)$. The dynamics of individuals traveling between cities is described by the stochastic transport operator Ω_j ($\{X^{[m]}\}$) representing the net balance of individuals in a given state $[m]$ that entered or left each city j. This operator is a function of the traffic flows per unit time $\omega_{j\ell}$ with the neighboring cities and of the city population N_j. In particular, the number of passengers of each category traveling from city j to city ℓ is an integer random variable, in that each of the potential travelers has a probability $p_{j\ell} = \omega_{j\ell}\,\delta t/N_j$ to go from j to ℓ in the time interval δt. In city j, the number of passengers traveling on each connection (j,ℓ) at time t defines a set of stochastic variables that follow a multinomial distribution (10). The calculation can be extended to include transit traffic up to one connecting flight (29).

The introduction of commuting flows represents a numerical challenge as it acts at a very different time scale with individuals that have very short visit duration in the neighboring subpopulation. Although the airline traffic finds a natural time scale of one day, an equivalent mechanistic simulation of the commuting mobility would require working on a much smaller time scale hardly compatible with the airline traffic. This problem can be solved by relying on the results of refs. 30 and 31. Commuting flows govern subpopulation interactions by defining the visiting rate of an individual in subpopulation i to subpopulation j as $\sigma_{ij} = w_{ij}/N_i$. Visiting individuals have a very short visit time associated to high return rates, τ, to their original subpopulations ($\tau \geq 3$ day^{-1}), corresponding to $\sigma_{ij} \ll \tau$ for all of the populations. It can be shown that the system is then described by stationary quantities $X_{ij}^{[m]}$ in each compartment $[m]$ for time scales larger than τ^{-1} (the time scale governing the relaxation time to equilibrium across subpopulations) in which $X_{ij}^{[m]}$ is defined as the number of visitors currently in j coming from subpopulation i in compartment $[m]$. When the disease duration is significantly larger than τ^{-1}, as is the case for the ILI with characteristic time $\mu^{-1} \simeq 3$ days, we can use a time scale separation approximation in which at each time step, the force of infection experienced by susceptible individuals in each subpopulation is a function of the stationary values I_{ii} and I_{ji}. The full force of infection used in the model is reported in the *Material and Methods* section, and a full derivation of the time separation approximation is reported in the *SI Appendix*.

Fig. 2 shows that the global and regional timing and size of the epidemic is weakly affected by also considering the commuting network in GLEaM. Both the probability of having a global outbreak and the overall profiles are very similar in the two cases, with the global epidemic size at the end of the first year almost unaffected by the inclusion of commuting. We perform a sensitivity analysis by testing the outset variation of the intensity of the commuting flows and varying the return time rate τ of more than one order of magnitude. In the *SI Appendix*, we show that the profiles do not show significant variations, the results being very robust against strong fluctuations in the commuting mobility process. The effect of commuting flows is, however, noticeable during

Downloaded from https://www.pnas.org by 37.19.197.239 on October 16, 2023 from IP address 37.19.197.239.

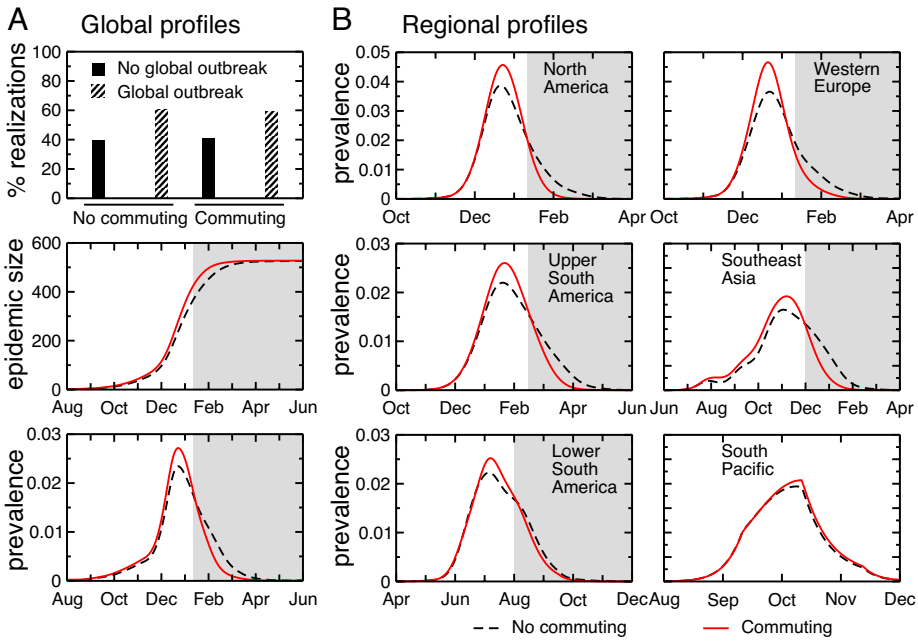


Fig. 2. Comparison of GLEaM predictions at the global and regional level obtained with and without commuting flows. Results refer to a pandemic influenza with $R_0 = 1.9$ starting in Hanoi on April 1. (*A*, *Top*) Probability of outbreak. About 40% of the realizations leads to an extinction at the source (Hanoi), whereas the remaining 60% causes a pandemic reaching more than 100 countries (i.e., a global outbreak). (*Middle* and *Bottom*) Global profiles for the epidemic size (number of cases per 1,000) and the prevalence, averaged over global outbreaks. (*B*) Regional profiles for the prevalence averaged over all runs that led to an outbreak in the given region. All results show a very limited impact of the commuting on the simulated patterns, more evident in the faster decay in the prevalence profiles as highlighted by the shaded areas. Reported results are averaged over 10^3 outbreak realizations.

the tail of the epidemic event. As presented in Fig. 2, many regions of the world show a broader tail in the absence of commuting, showing that the commuting coupling enhances the synchronization of the local epidemic profiles. The observed broadening of an epidemic profile that includes multiple subpopulations is due to the different timing of the outbreak that reaches the various subpopulations. The effect is more pronounced in the lack of short range coupling, as highlighted in the example reported in Fig. 3*D* and *E* of an air transportation hub loosely connected by air travel flow to the surrounding subpopulations. As expected, no significant change is observed in the hub profile, whereas the time delay in neighboring locations with limited airline connections is dramatically reduced by the coupling due to local commuting flows. After infecting the hub, the epidemic radiates out to the neighboring geographical census areas in a pattern reminiscent of the physical process of diffusion. This effect naturally leads to a much stronger correlation and synchrony in the evolution of the pandemic at the local level. In the *SI Appendix*, we present the model informed with the realistic

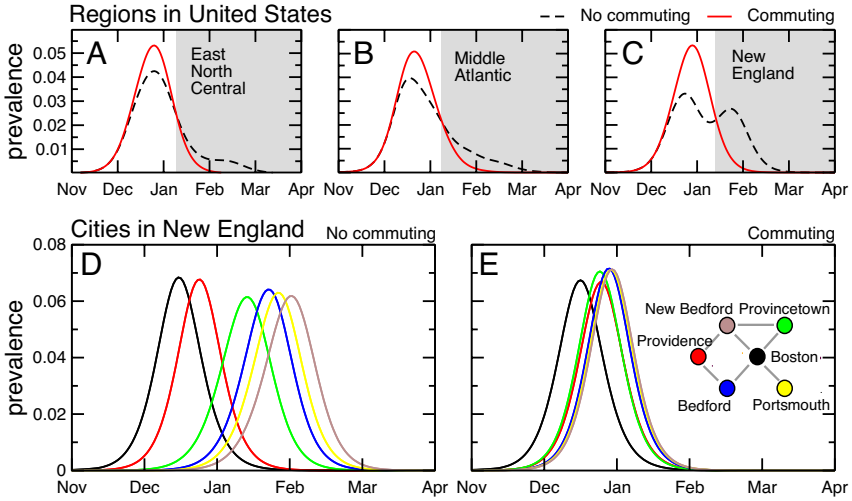


Fig. 3. Comparison of GLEaM predictions at the local level obtained with and without commuting. (*A*–*C*) Prevalence profiles of three continental U.S. regions. The effect of commuting is visible in the faster decay (as highlighted by the shaded areas) and absence of multiple peaks. (*D* and *E*) Prevalence profiles for Boston area and the surrounding cities with no commuting (*D*) and with commuting (*E*). A schematic network representation of the short-range connections is shown for guidance. The synchronization among the prevalence profiles is considerably increased when commuting is considered, with a reduction of over one month in the time interval between peaks in neighboring cities. Reported profiles are averaged over 10^3 outbreak realizations.

Downloaded from https://www.pnas.org by 37.19.197.239 on October 16, 2023 from IP address 37.19.197.239.

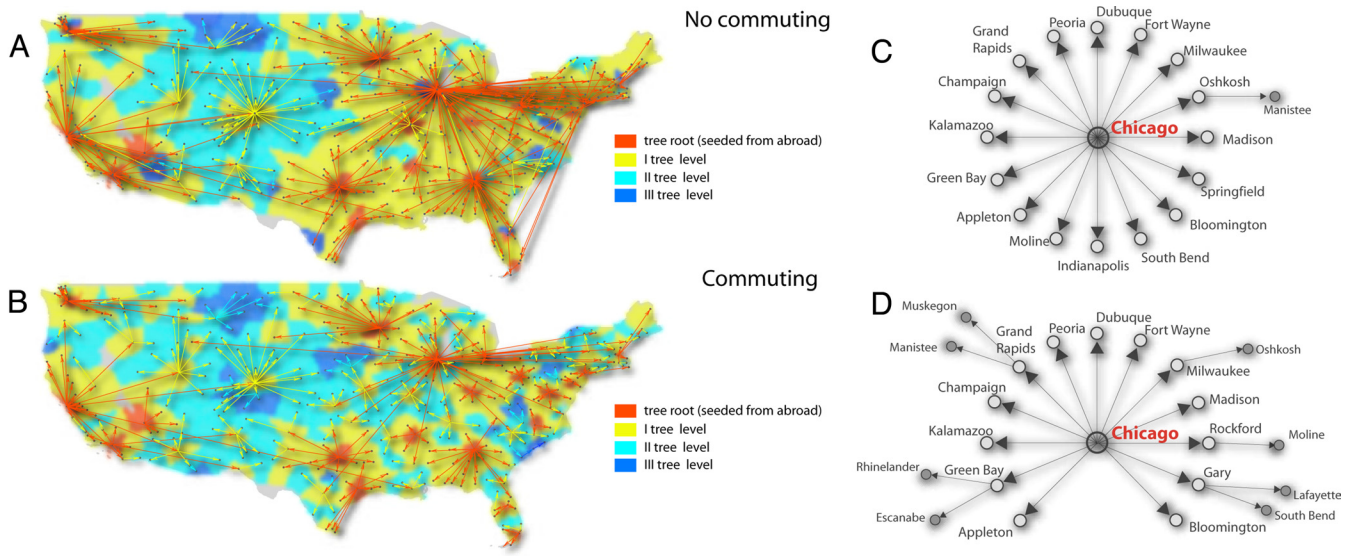


Fig. 4. Epidemic invasive tree. (*A* and *B*) Geographical representation of the continental U.S. epidemic invasion tree with only airline traffic (*A*) and when both airline traffic and commuting are considered (*B*). Red represents the roots (i.e., the first cities that were seeded from abroad), and, as we move down the tree, the colors change from yellow to dark blue. The arrows representing the edges of the tree are colored as the parent node. (*C* and *D*) We also provide a schematic representation of the invasion tree rooted at Chicago when only flights are considered (*C*) and with both air traffic and commuting (*D*). As demonstrated in both examples, the spreading pathway is completely dominated by the airline hubs as the only sources of imported seeds. However, the hierarchy is broken by the introduction of commuting flows as the number of shells around the airline hubs and the branches at the secondary nodes increase.

parameters and initial conditions of 2001–2002 seasonal influenza season and compare the obtained epidemic pattern with the real data. The results confirm the synchronization effect and the better agreement with the actual data at the regional level when commuting flows are introduced in the model.

Commuting flows therefore alter the hierarchy of epidemic transmission from region to region. This hierarchical organization can be inferred by constructing the epidemic invasion tree that represents the transmission of the infection from one subpopulation to the other during the history of the epidemic. The stochastic nature of the epidemic process implies that each realization will produce a different tree. An overall epidemic invasion network can be constructed by defining weighted, directed links, T_{ij}, that denote the probability that the epidemic in subpopulation j is seeded by individuals belonging to the subpopulation i. This probability is defined by the ratio between the number of realizations in which we have a seeding $i \rightarrow j$ and the total number of realizations. When constructing the epidemic invasion tree we use averages over 10^3 realizations. Finally, to highlight only the most likely infection tree, we construct the minimum spanning tree from the world-seeding subpopulation where we minimize the distance defined on each link as $\sqrt{1 - T_{ij}}$. In Fig. 4, we show the infection arrival time hierarchy in the two considered scenarios for the continental U.S. In the absence of commuting (Fig. 4*A*), airline hubs have a predominant role and are completely responsible for spreading the disease to every other location through direct air connections. This feature leads to the counterintuitive effect that locations near a large airport, but with no frequent direct flight to that airport, can be infected only much later through a convoluted sequence of flights. On the other hand, when we superimpose the commuting network we obtain the expected effect of reducing the importance of large airports and increasing the locality of the epidemic spread (Fig. 4*B*). The inclusion of commuting patterns is therefore relevant in the evaluation of the epidemic invasion path and timing.

Conclusions

Data collected from 29 countries in five continents were used to fit a gravity law that was then used to model commuting behavior between the Voronoi geographical census areas built around every airport indexed by IATA. The effect of adding this short-range commuting network to a worldwide epidemic model including all airline traffic flowing among 3,362 airport locations allows us to discriminate the main contribution of the long- and short-range mobility flows. The impact of the epidemic does not change as the competition between the long- and short-range coupling acts only at the beginning of the epidemic in each subpopulation. Both coupling terms become a second-order effect once the epidemic ramps up and the major force of infection is endogenous to the subpopulation. Therefore, both coupling mechanisms affect just the hierarchy of epidemic progression and its timing. On the one hand, the global epidemic behavior is governed by the long-range airline traffic that determines the arrival of infectious individuals on a worldwide scale. At the local level, however, the short-range epidemic coupling induced by commuting flows creates a synchrony between neighboring regions and a local diffusive pattern with the epidemic flowing from subpopulations with major hubs into the neighboring subpopulations. These results clearly show that the level of detail on the mobility networks can be chosen according to the scale of interest. Neglecting local coupling for instance does not produce a dramatic effect if one is mainly interested in the global overall pattern at the granularity level of a large geographical area or country. On the other hand, more refined strategies that require access to finer granularity can be implemented by the progressive addition of details without radically altering the perspective achieved at the larger scales. This is extremely important in the balance between computational time and flexibility of models and becomes very relevant when computational approaches are used in real time to aid the decision process for a public health emergency. The present analysis opens the path to quantitative approximation schemes that calibrate the level of data resolution and the needed computational resources with respect to the accuracy in the description of the epidemics.

Materials and Methods

Voronoi Tessellation Around Main Transportation Hubs. We define the geographical census areas centered around IATA airports by assigning the population of each cell of 15 × 15 minutes of arc to the closest airport within the same country. Such a procedure defines a Voronoi-like tessellation (32) for the popu-

Downloaded from https://www.pnas.org by 37.19.197.239 on October 16, 2023 from IP address 37.19.197.239.

lated cells of the world with a cut-off scale for the tassels size of 200 kms (see also the *SI Appendix* for further details).

Disease Structure, Seasonality and R_0. In each urban area, the evolution of the disease is governed by the compartmental scheme of the baseline scenario of ref. 10. A susceptible individual S in contact with a symptomatic (I^t,I^{nt}, traveling or nontraveling, respectively) and asymptomatic (I^a) infectious individual contracts the infection at rate β or $r_\beta\,\beta$, respectively, and enters the latent (L) compartment, where he/she is infected but not yet infectious. At the end of the latency period, individuals in the latent class enter one of the symptomatic infectious compartments (I^t, I^{nt}) with probability $1-p_a$ or become asymptomatic (I^a) with probability p_a. Symptomatic individuals are further divided between those who are allowed to travel (I^t) with probability p_t and those who are prevented from doing so (I^{nt}) with probability $1-p_t$, depending on the severity of symptoms. All infectious individuals enter the permanently recovered/removed compartment (R) at a rate of μ per day. The latent period has an average duration of $\varepsilon^{-1} = 1.9$ days and is assumed to be followed by an infectious period with a mean duration of $\mu^{-1} = 3$ days (3, 10, 33). Given that infection has occurred, we assume that individuals become asymptomatic with probability $p_a = 0.33$ (3, 10, 33). The relative infectiousness of asymptomatic individuals is $r_\beta = 0.5$ (10) and symptomatic individuals are allowed to travel with probability $p_t = 0.5$. The contagion process (i.e., the generation of new infections through the transmission of the disease from infectious to susceptible individuals) and the spontaneous transitions (e.g., from latent to infectious or from infectious to recovered) are modeled with binomial and multinomial distributions (see the *SI Appendix* for a detailed description of the processes). The threshold parameter of the disease that determines the spreading rate of infection is called basic reproduction number (R_0) and is defined as the average number of infected cases generated by a typical infectious individual when introduced into a fully susceptible population (34). For our compartmental model we have $R_0 = \beta\mu^{-1}\,[1-p_a + r_\beta\, p_a]$. The R_0 values indicated in the figures and discussed in the paper do not consider the effect of seasonality and the commuting in the force of infection. We take into account the seasonal behavior of influenza by adopting the scheme from ref. 10. The transmission rate β_j in each geographical census area is adjusted by a scaling factor that varies monthly according to the city's climatic zone. For example, cities in the tropical zone have a scaling factor that is always 1, independent of the season. See ref. 10 and its supporting information for details.

Effective Force of Infection Generated by Commuting Flows. The effect of commuting in the spread of infection can be considered implicitly by evaluating the force of infection between subpopulations coupled by commuting flows. In the case of $\tau \ll \sigma_j$, the relaxation time to equilibrium values in the populations is dominated by the return rate τ to the origin subpopulation, as shown in the *SI Appendix* file. We can therefore use the equilibrium values of population sizes in our calculations of force of infection, because the τ^{-1} is much smaller than the time scales of disease evolution (i.e., ε^{-1} and μ^{-1}). New infections in a subpopulation are due to the transmission between susceptibles and infectious individuals occurring in the subpopulation or during a visit to a neighboring subpopulation. Taking this into account, it is possible to derive the force of infection λ_j in j as

$$\lambda_j = \frac{\beta_j}{(1+\sigma_j/\tau)N_j^*}\left[I_j^{nt} + \frac{I_j^t + r_\beta I_j^a}{1+\sigma_j/\tau}\right] + \frac{1}{(1+\sigma_j/\tau)\tau}\sum_{i\in v(j)}\left[\frac{\beta_j\sigma_{ij}}{N_j^*}\frac{I_i^t + r_\beta I_i^a}{1+\sigma_i/\tau} + \frac{\beta_i\sigma_{ji}}{N_i^*}\left(I_i^{nt} + \frac{I_i^t + r_\beta I_i^a}{1+\sigma_i/\tau} + \sum_{\ell\in v(i)}\frac{\sigma_{\ell i}}{\tau}\frac{I_\ell^t + r_\beta I_\ell^a}{1+\sigma_\ell/\tau}\right)\right]. \quad [2]$$

A detailed derivation is provided in the *SI Appendix*. In the above expression, $N_j^* = N_{jj} + \Sigma_{i\in v(j)} N_{ij}$ is the actual number of individuals in j due to commuting. The first terms of the right-hand side of Eq. **2** takes into account the transmission of the infection from the local infectious individuals in j. The second term considers the transmission due to the infectious individuals during their visits to j with local susceptible persons. The third and fourth terms consider the interactions of susceptible individuals during their visits to neighboring subpopulations i with the local infectious persons and the infectious visitors of i, respectively. Here we have also considered that the transmission rate β may be different in each population. The last expression includes second-order commuting terms (e.g., $\sigma_{ji}\,\sigma_{\ell i}/\tau^2$), which are neglected in the actual computation. The probability of new infections to be generated in city j is finally given by $\lambda_j\,\delta t$ in the time interval δt, acting on a pool of susceptible individuals S_j.

ACKNOWLEDGMENTS. We are grateful to the International Air Transport Association for making the airline commercial flight database available to us. This work has been partially funded by the National Institutes of Health Grant R21-DA024259, the Lilly Endowment Grant 2008 1639–000. A.V. was supported by Defense Threat Reduction Agency Grant 1–0910039, A.V. and V.C. was supported by Future Emerging Technologies contract no. 231807 (EPIWORK), and V.C. was supported by European Research Council Ideas contract no. ERC-2007-Stg204863 (EPIFOR).

1. Riley S (2007) Large-scale spatial-transmission models of infectious disease. *Science* 316:1298–1301.
2. Eubank S, et al. (2004) Modelling disease outbreaks in realistic urban social networks. *Nature* 429:180–184.
3. Longini IM, et al. (2005) Containing pandemic influenza at the source. *Science* 309:1083–1087.
4. Ferguson NM, et al. (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437:209–214.
5. 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.
6. 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.
7. Rvachev LA, Longini IM (1985) A mathematical model for the global spread of influenza. *Math Biosci* 75:3–22.
8. Grais RF, Hugh Ellis J, Glass GE (2003) Assessing the impact of airline travel on the geographic spread of pandemic influenza. *Eur J Epidemiol* 18:1065–1072.
9. Hufnagel L, Brockmann D, Geisel T (2004) Forecast and control of epidemics in a globalized world. *Proc Natl Acad Sci USA* 101:15124–15129.
10. Colizza V, Barrat A, Barthélemy M, Valleron A-J, Vespignani (2007) Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PloS Medicine* 4:e13.
11. Pentland A (2009) Reality mining of mobile communications: Toward a New Deal on data. In *The Global Information Technology Report 2008–2009*, eds Dutta S, Mia I (World Economic Forum, Geneva) p. 75.
12. Chowell G, Hyman JM, Eubank S, Castillo–Chavez C (2003) Scaling laws for the movement of people between locations in a large city. *Phys Rev E* 68:066102.
13. Barrat A. Barthelemy M. Pastor–Satorras R. Vespignani A (2004) The architecture of complex weighted networks. *Proc Natl Acad Sci USA* 101:3747–3752.
14. Wu Z. Braunstein LA. Havlin S. Stanley HE (2006) Transport in weighted networks: Partition into superhighways and roads. *Phys Rev Lett* 96:148702.
15. Guimera R, Mossa S, Turtschi A, Amaral LAN (2005) The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc Natl Acad Sci USA* 102:7794–7799.
16. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439:462–465.
17. Viboud C, et al. (2006) Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* 312:447–451.
18. De Montis A, Barthélemy M, Chessa A, Vespignani A (2007) The structure of interurban traffic: A weighted network analysis. *Environ Planning B* 34:905–924.
19. Patuelli R, Reggiani R, Gorman SP, Nijkamp P, Bade F-J (2007) Network analysis of commuting flows: A comparative static approach to German data. *Networks Spatial Econ* 7:315–331.
20. González MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. *Nature* 453:779–782.
21. Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512.
22. Erlander S, Stewart NF (1990) *The Gravity Model in Transportation Analysis* (VSP, Utrecht).
23. Ortúzar J de D, Willumsen LG (2001) *Modelling Transport* (Wiley, Chichester, United Kingdom).
24. Flahault A, Valleron A-J (1991) A method for assessing the global spread of HIV-1 infection based on air-travel. *Math Popul Stud* 3:1–11.
25. Cooper BS, Pitman RJ, Edmunds WJ, Gay NJ (2006) Delaying the international spread of pandemic influenza. *PloS Medicine* 3:e12.
26. Epstein JM, et al. (2007) Controlling pandemic flu: The value of international air travel restrictions. *PLoS ONE* 2:e401.
27. Kulldorf G (1955). *Migration Probabilities*, Lund Studies in Geography, Series B. (Lund University, Lund, Sweden), No. 14.
28. Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The role of airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci USA* 103:2015–2020.
29. Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The modeling of global epidemics: Stochastic dynamics and predictability. *Bull Math Bio* 68:1893–1921.
30. Keeling MJ, Rohani P (2002) Estimating spatial coupling in epidemiological systems: A mechanistic approach. *Ecol Lett* 5:20–29.
31. Sattenspiel L, Dietz K (1995) A structured epidemic model incorporating geographic mobility among regions. *Math Biosci* 128:71–91.
32. Okabe A, Boots B, Sugihara K, Chiu SN (2000) *Spatial Tessellations—Concepts and Applications of Voronoi Diagrams* (Wiley, New York).
33. Longini IM, Halloran ME, Nizam A, Yang Y (2004) Containing pandemic influenza with antiviral agents. *Am J Epidemiol* 159:623–633.
34. Anderson RM, May RM (1992), *Infectious Diseases in Humans* (Oxford Univ Press, Oxford).

Downloaded from https://www.pnas.org by 37.19.197.239 on October 16, 2023 from IP address 37.19.197.239.