nature communications

Article | Article in Press

https://doi.org/10.1038/s41467-026-78115-6

Block-fitness modeling of the global air mobility network

Received: 11 February 2026

Accepted: 8 September 2026

Published online: 12 October 2026

Check for updates

Giulia Fischetti[1,2,9], Anna Mancini[3,4,9], Giulio Cimini[3,4,5] ✉, Jessica T. Davis[6], Abby O. K. Leung[6], Alessandro Vespignani[6] & Guido Caldarelli[1,5,7,8] ✉

Accurate representations of origin-destination (OD) passenger flows on the World Air Transportation Network (WAN) are fundamental inputs to models of global mobility, epidemic risk, and infrastructure planning. However, OD data are largely commercial and proprietary, therefore often inaccessible to the research community. Here we introduce a generative model that treats air mobility as a stochastic process within a sparse maximum-entropy framework, informed by aggregate traffic volumes of individual locations and geographic regions. The model accurately reconstructs both network structure and mobility flows, yielding epidemic spreading trajectories that closely match those obtained using real mobility data. Furthermore, the method can be effectively implemented using open data on flight routes and airport traffic, reducing dependence on proprietary datasets. Our approach provides a scalable, interpretable, and computationally efficient framework for scenario analysis and policy design in global mobility systems.

Statistical physics provides a powerful framework to describe collective behavior in complex systems. In recent years, this approach has extended beyond traditional domains to capture socio-economic phenomena, where large-scale patterns emerge from the interactions of many heterogeneous agents[1–8]. A central example is epidemic spreading, in which underlying mobility networks shape the pathways of disease transmission across populations [9–17]. In particular, the World Air Transportation Network (WAN) drives global contagion dynamics by rapidly connecting distant regions [18–22].

Despite their key importance in computational epidemiology[23–25], relevant datasets quantifying air passenger mobility are difficult to access. Openly available data consist of scheduled flights and seat capacities[26,27], which describe the physical transportation infrastructure, but provide information only on direct flight connections and misestimate realized passenger flows[28,29]. Indeed, traffic measured on individual legs tends to overestimate mobility importance at transit hubs and underestimate it at actual destinations. Origin-destination (OD) datasets derived from air ticket sales instead capture how many individuals travel from a source population to a final destination, regardless of intermediate flight segments. Therefore, they provide a more accurate representation of the population mixing underlying epidemic spread, improving large-scale forecasting performance[30,31]. As such, their use has become standard in recent literature[32–35].

However, OD datasets are typically proprietary, expensive for research purposes, and difficult to share openly. This limitation motivates the development of methods capable of reconstructing global air mobility from more accessible information. Existing approaches infer OD flows using gravity-like models[36–40], which combine demographic, socio-economic and geographic variables to estimate movement between locations. Machine-learning approaches (see[41] and references therein) extend this strategy by exploiting richer information sets, improving predictive performance at the expense of interpretability. While successful in many settings, these approaches rely on strong modeling assumptions and on the selection of exogenous informative covariates. Moreover, unless additional sparsity assumptions are introduced (such as conditioning on the topology extracted from

[1]DMSN, Ca' Foscari University of Venice, Mestre, Italy. [2]IAMP, ZHAW School of Engineering, Winterthur, Switzerland. [3]Physics Dept. and INFN, University of Rome Tor Vergata, Rome, Italy. [4]Enrico Fermi Research Center, Rome, Italy. [5]CNR-Institute for Complex Systems (ISC), Rome, Italy. [6]MOBS Lab, Northeastern University, Boston, MA, USA. [7]ECLT Ca' Bottacin, Venice, Italy. [8]LIMS, Royal Institution, London, UK. [9]These authors contributed equally: Giulia Fischetti, Anna Mancini. ✉e-mail: giulio.cimini@roma2.infn.it; guido.caldarelli@unive.it

flight schedules[42,43]), they typically produce dense or effectively fully connected systems–at odds with the observed sparsity of air connectivity arising from operational, economic, and geographical constraints.

Here we aim to overcome these limitations by proposing a generative framework to reproduce both structure and flows of the OD global air mobility network using a minimal set of endogenous information. To this end, we draw on network science approaches for reconstructing weighted networks.

Focusing on topology extracted from flight routes, previous studies show that the global air mobility network exhibits heavy-tailed degree distributions, small-world properties, and a modular organization characterized by dense regional communities connected by long-range intercontinental routes[44,45]. Several generative approaches have been proposed to reproduce these features. Preferential attachment modulated by geographic distance captures the emergence of hubs and local clustering[46]. Entropy-based spatial formulations assign connection probabilities as functions of node degree and distance, reproducing similar large-scale connectivity patterns[47–49]. However, these approaches largely neglect link weights, which are not trivially determined by connectivity[26]. Weighted maximum-entropy approaches that constrain passenger volumes at each location have not been explored, as they generate dense connectivity patterns analogous to gravity models[7].

We instead build on a two-step reconstruction framework developed for economic networks[50]. First, we infer the existence of connections through a maximum-entropy fitness model[51,52]. Second, we assign passenger volumes conditional on those links through a density-corrected gravity mechanism[53]. We extend this approach by introducing constraints that reflect the modular organization of global air mobility into geographic communities[54,55]. Specifically, we use traffic-by-region at individual locations as fitness governing their connectivity (see Methods). The resulting block-fitness model provides a probabilistic framework to generate synthetic realizations of the network while preserving analytical tractability.

We validate the proposed framework against empirical OD data and compare its performance with alternative formulations. The model accurately recovers individual connections, traffic statistics, and large-scale mobility patterns of the empirical network. When employed in metapopulation epidemic simulations, synthetic networks reproduce spreading trajectories that closely match those obtained on real data, including their spatial heterogeneity and temporal dynamics. Importantly, the model is informed solely by aggregate mobility patterns, which we extract from a proprietary OD dataset, but can also be efficiently calibrated using open information on flight routes and airport passenger counts. Beyond reconstruction performance, the framework remains interpretable and scalable while providing insight into the mechanisms underlying the formation of global air mobility networks, offering a practical tool for researchers and practitioners in network epidemiology.

Results

WAN representation and properties

In this work, we define the WAN as the worldwide origin-destination (OD) air passenger network. We thus consider air-travel data from the Official Aviation Guide (OAG)[56], which reports monthly counts of passengers traveling from an origin airport to their final destination, independently of intermediate layovers. We group airports into "basins" by dividing the world into over 3200 geographic subpopulations generated using a Voronoi tessellation of the Earth's surface, where subpopulations are centered around major transportation hubs (details in Supplementary Methods). Passenger flows are then summed across each basin pair. We take the monthly basin networks from September to November 2019 and compute edge weights by averaging over the number of daily travelers. The resulting network is undirected, with link weight w_{ij} given by the flux of passengers in both directions between basins i and j. For each basin i, the passenger volume is measured by the strength $s_i = \sum_j w_{ij}$ and the number of connections by the degree $k_i = \sum_j a_{ij}$ ($a_{ij} = 1$ if $w_{ij} > 0$ meaning that i and j are connected, and zero otherwise). Degree and strength distributions are well approximated by power-laws with exponential cut-offs[57], reflecting capacity constraints that limit the number of connections and passengers a single basin can accommodate (Fig. 1A). The relation between k and s is sub-linear, with some dispersion (Fig. 1B). The distance d_{ij} between basins i and j is computed using the Haversine formula applied to the centroids of the two basins. Link occurrence does depend on geographical distance, as shown by the frequency of connections at different distance intervals (Fig. 1C). Basins are partitioned into 11 geographical regions derived from the United Nations geoscheme and the World Bank regional classification (Fig. 1D). Such partition induces a block structure in the network with very high modularity[58] (Fig. 1E).

Network reconstruction

The central element of our modeling framework is the fitness ansatz, for which node strengths serve as parameters that control expected degrees. This assumption is justified when degrees are strongly correlated with strengths, as is the case for the WAN (Fig. 1B). Consistent with this premise, the model accurately reconstructs node degrees as ensemble expectations (Fig. 2A), while node strengths are reproduced by construction (see Supplementary Methods). Concerning the inference of individual links, we measure the empirical link probability as a function of the model link probability p_{ij}, by binning p_{ij} values and counting the fraction of linked (i, j) pairs in each bin. We obtain a very good agreement between the two quantities (Fig. 2B). Besides, the model recovers the weights of existing WAN connections as aggregate ensemble averages $\langle w_{ij} \rangle$ (Fig. 2C).

We further employ the model link probability p_{ij} for a link prediction exercise that determines the outcome of a binary classification test on the existence of a link, given by a_{ij}. Values of precision ($\sum_{ij} a_{ij} p_{ij} / \sum_{ij} p_{ij}$) and recall ($\sum_{ij} a_{ij} p_{ij} / \sum_{ij} a_{ij}$) of the classifier are statistically equivalent in our case, since by construction model networks have on average the same number of links as the empirical one (see Methods). We obtain a very high recall of 75%, meaning that the model is able to correctly identify about three out of four existing links, despite the dataset being highly unbalanced with only a few "positive" existing links: the network density is 1.8%, corresponding to the expected recall of a random classifier.

The model can also reconstruct the signature properties of the WAN. It captures the trend and the full dispersion of the k vs s relation (Fig. 2D). It reproduces, to a large extent, the empirical dependence of link occurrence on geographic distance (Fig. 2E), despite not using any explicit spatial information. By construction, it also recovers the network's block structure (Fig. 2F).

Spreading simulations

Next we evaluate whether the block-fitness model reproduces epidemic dynamics on the empirical WAN. We use a stochastic metapopulation model for global epidemic spreading, where the compartmental equations in each basin are coupled through the passenger flows set by the WAN travel matrix[9,59] (see Methods for further details). Fig. 3 illustrates the spatiotemporal dynamics of contagion on both the real and model WAN, for a representative case in which the epidemic outbreak originates in London. Prevalence maps (Fig. 3A and Fig. 3B for the real and model networks, respectively) show the global distribution of basin-level prevalence $i_m(t)$ at a fixed time, where $i_m(t)$ denotes the fraction of infected individuals in basin m at time t relative to its total population. The temporal evolution of the total prevalence $i(t)$–defined as the fraction of infected individuals worldwide (Fig. 3C)–exhibits the characteristic epidemic peak. To quantify the spatial

https://doi.org/10.1038/s41467-026-78115-6

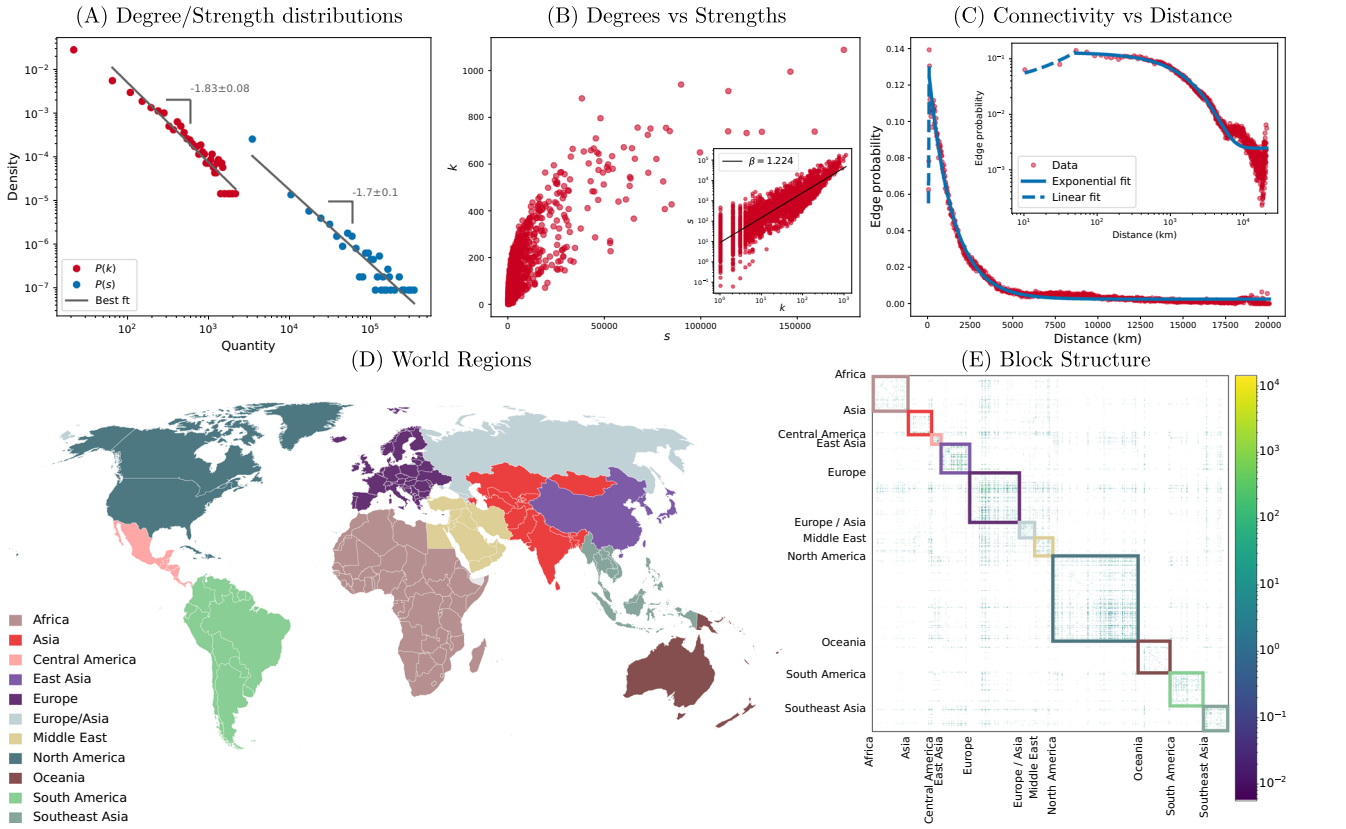


Fig. 1 | Topological features of the WAN. A Degree distribution with power-law fit $P(k) \sim k^{-\gamma_k}$ and exponent $\gamma_k = 1.8(3)$, and strength distribution with power-law fit $P(s) \sim s^{-\gamma_s}$ and exponent $\gamma_s = 1.(7)$ (50 data bins are used to represent each distribution). **B** Degree-strength relation of individual basins, with OLS fit $s \sim k^{\beta}$ and exponent $\beta = 1.2(2)$. **C** Link probability as a function of geographic distance (1000 data bins of approximately 20 km width are used). The curve is fitted with a linear function $l(d) = m \cdot d + q$ up to a characteristic distance $d^* = 50km$ and then with an exponential $l(d) = \exp[-a(d - d_0)] + b$. Best-fit parameters are $m = 0.003$, $q = -0.0013$ for the linear term and $d_0 = -2.(9) \cdot 10^3$, $a = 7.(1) \cdot 10^{-4}$, $b = -2.(4) \cdot 10^{-3}$ for the exponential decay ($R^2 = 0.99$). **D** World partition into regions: Africa, Asia, Central America, East Asia, Europe, Europe/Asia, Middle East, North America, Oceania, South America, Southeast Asia. **E** Community structure in the WAN adjacency matrix induced by the world partition, with modularity $M = 0.67$.

heterogeneity in disease prevalence, we compute the normalized spatial distribution $\rho_m(t) = i_m(t)/\sum_n i_n(t)$ and its associated entropy $H(t) \propto -\sum_m \rho_m(t) \log \rho_m(t)$. This quantity vanishes when the epidemic is fully localized within a single basin and is maximized when the epidemic is homogeneously spread across the globe[21] (a condition that typically occurs in correspondence of the epidemic peak, Fig. 3D).

Although visual inspection already suggests strong agreement between contagion dynamics on real and model networks (the prevalence maps and the epidemic peak are very well reproduced), we consider several complementary metrics to quantitatively compare simulations on real and model WANs. We measure the overall mismatch between the prevalence curves using the area between curves and their L2 distance, while we capture differences in outbreak timing, intensity, and duration by measuring the difference in peak time, peak magnitude, and epidemic length. Further, we compute the Root Mean Square Error (RMSE) of the basin-level prevalence values $i_m(t)$, averaged over time, and the L2 distance between the entropy curves. Table 1 reports values of these quantities, averaged over several simulations on empirical or model networks. Note that we always compare simulations with the same initial seed, but we average across various seeds to evaluate model performance across multiple starting points of the dynamics. We obtain very low values for all metrics that, notably, are very close to the values computed only among simulations on the real network, thus quantifying the intrinsic variability of the dynamics. Overall, the epidemic spreading patterns on the model network closely reproduce those observed on the real WAN.

Open data calibration

Finally, we show how the model performs when informed with openly available data on air traffic, rather than aggregate OD information from OAG. We first construct an open dataset by combining two independent data sources: flight records from OpenSky[60] used to extract network topological information, and passenger traffic statistics from Wikipedia used to infer basin strengths (see Methods). As open data coverage is incomplete, we can collect information only for a restricted set of 739 basins (about 25% of the total), which is yet representative (45%) of the total mobility derived from OAG data. We then compare model networks calibrated on the collected open data with the empirical OAG network for the restricted basin set.

Table 2 contains an overview of the results. By definition, the model now reproduces strength values from Wikipedia, which are not perfectly aligned with OAG values. Also, because of the smaller sample size leading to a higher statistical error, structural reconstruction performance slightly deteriorates overall with respect to the OAG-informed model (especially at the level of individual link weights), but recall stays as high as 0.58, and large-scale topological properties of the network remain well reproduced. In terms of epidemic simulations, the gap between the intrinsic variability for empirical data and the difference between model and empirical scenarios increases. However, model values remain generally compatible with empirical quantities (note that RMSE[i_m] and L2[H] are sensitive to the cardinality of the node set). Hence, the model remains accurate even for imperfect input information on aggregate OD flows.

https://doi.org/10.1038/s41467-026-78115-6

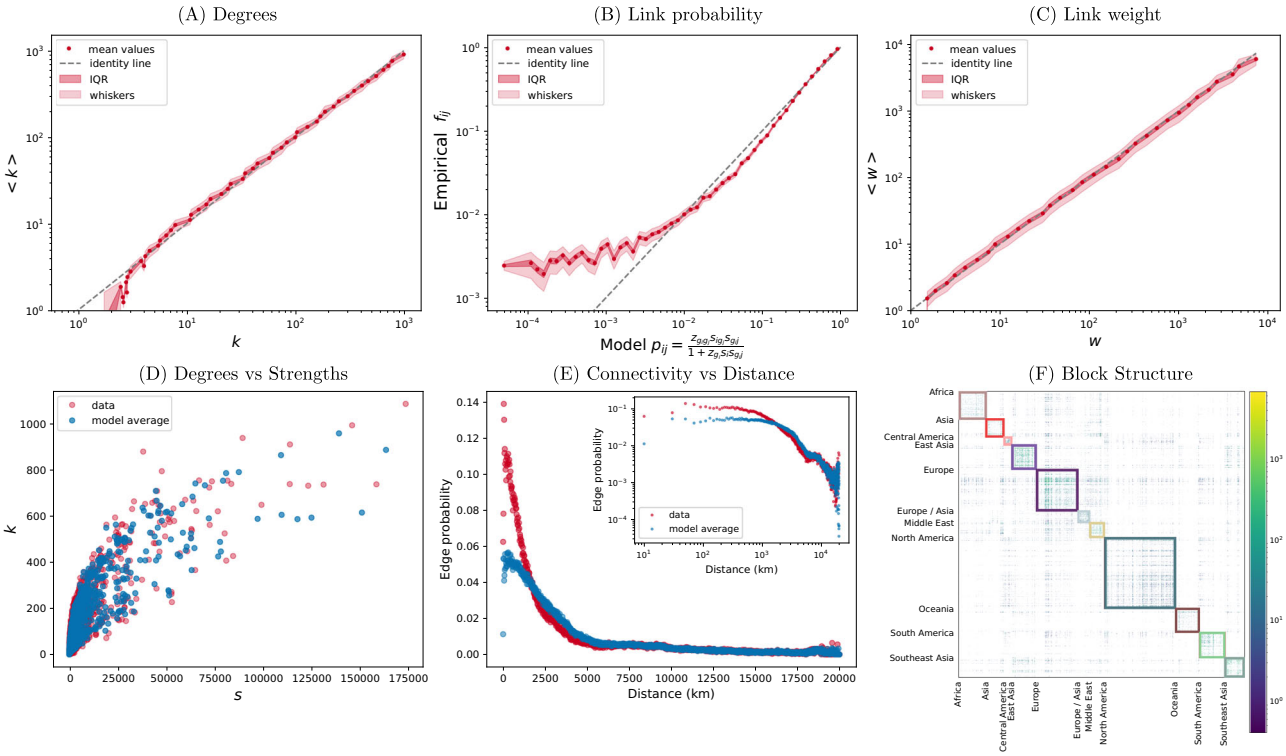


Fig. 2 | Reconstruction of topological properties of the WAN. A–C Scatter plots of model vs empirical values (represented by binning abscissa values into 50 uniformly-spaced intervals in log scale). The dashed identity line represents perfect agreement while the shaded regions mark the inter-quartile range [Q1, Q3] and the whiskers of the model values set (with the lower limit truncated at zero). Root Mean Squared Relative Error (RMSRE, definition in Supplementary Table 2) values are computed for binned values to focus on the statistical reconstruction of aggregate trends rather than individual points. **A** Model degrees $\langle k_i \rangle = \sum_{j \neq i} p_{ij}$ versus empirical degrees k_i, with RMSRE = 0.18. Quartiles are computed empirically over the ensemble values falling in each bin. **B** Empirical link probability (given by the fraction f_{ij} of ij pairs that are actually connected in each bin of p_{ij}) as a function of the model link probability p_{ij}, with RMSRE = 0.016. As each bin contains a single value of f_{ij}, quartiles are obtained from the normal approximation of the binomial distribution mean, with variance $f_{ij}(1-f_{ij})$ divided by the number of ij pairs in the corresponding bin. **C** Average model weights $\langle w \rangle_{ij}$ versus empirical weights w_{ij}, with RMSRE = 0.063. Quartiles are computed empirically over the ensemble values falling in each bin. For this exercise, we limit the test set to links $w_{ij} > 0$ that are actually realized. **D–F** Empirical vs model network features. **D** Degrees vs strengths relation $s \sim k^{\beta}$ for individual basins, with OLS fit exponent $\beta = 1.2(2)$ on empirical data and $\beta = 1.17(0)$ on model data. **E** Link probability as a function of distance (binning as in Fig 2C), with Mean Squared Error (MSE) between empirical and model curve of 1.61. **F** Block structure of a sample reconstructed network. Modularity is 0.63 on average over the model ensemble for the imposed partition ($M = 0.64$ for real data).

Table 1 | Similarity of epidemic simulations on real and model networks, in terms of: area between and L2 distance of prevalence curves; differences of peak time, peak magnitude, epidemic length; RMSE of prevalence vectors; L2 distance of entropy curves (definitions in Supplementary Table 3)

	Δ area	L2[i]	Peak time	Peak magn.	Epid. length	RMSE[i_m]	L2[H]
Model	0.56 ± 0.31	0.073 ± 0.040	2.3 ± 1.8	0.0048 ± 0.0035	4.5 ± 3.2	0.050 ± 0.006	14.5 ± 6.1
Data	0.43 ± 0.32	0.058 ± 0.042	2.0 ± 1.7	0.0043 ± 0.0034	2.2 ± 2.2	0.024 ± 0.005	13.1 ± 6.1

Values on the "model" row are averages over pairwise comparisons of one simulation on the empirical network and one simulation on a model network, while "Data" values are computed with simulations on the empirical network only. Uncertainties are given as standard deviations.

Discussion

In this paper, we introduced a maximum-entropy framework to model origin-destination (OD) flows of the worldwide air mobility network (WAN). The model extends classical fitness approaches to spatially embedded systems with community structure through a *block-fitness* formulation[61], in which passenger volumes at individual locations determine connectivity while aggregate passenger fluxes across geographic communities are preserved. This construction enables the probabilistic generation of synthetic air mobility networks using only coarse-grained information on the system.

The model faithfully reproduces the large-scale topological properties of the empirical WAN, the presence and traffic statistics of individual connections (achieving a link prediction recall of 75%), as well as the dynamical patterns of epidemic spreading. The block-fitness approach outperforms alternative formulations that account for the super-linear relation between strengths and degrees[26] and for the dependence of connection probabilities on geographical distance[48] (see Supplementary Methods for model definitions and Supplementary Note 1 for model comparison). Our conclusions remain robust when considering global air mobility at a finer geographical resolution, where nodes correspond to individual airports rather than basins of metropolitan areas (details in Supplementary Note 2).

The community partition is specified a priori using the UN geoscheme and World Bank regional classification, rather than inferred from network data (see Supplementary Note 3 for another model specification with communities corresponding to ICAO regions). Hence we do not optimize the model with respect to the chosen partition, and a data-driven refinement of the community structure could

https://doi.org/10.1038/s41467-026-78115-6

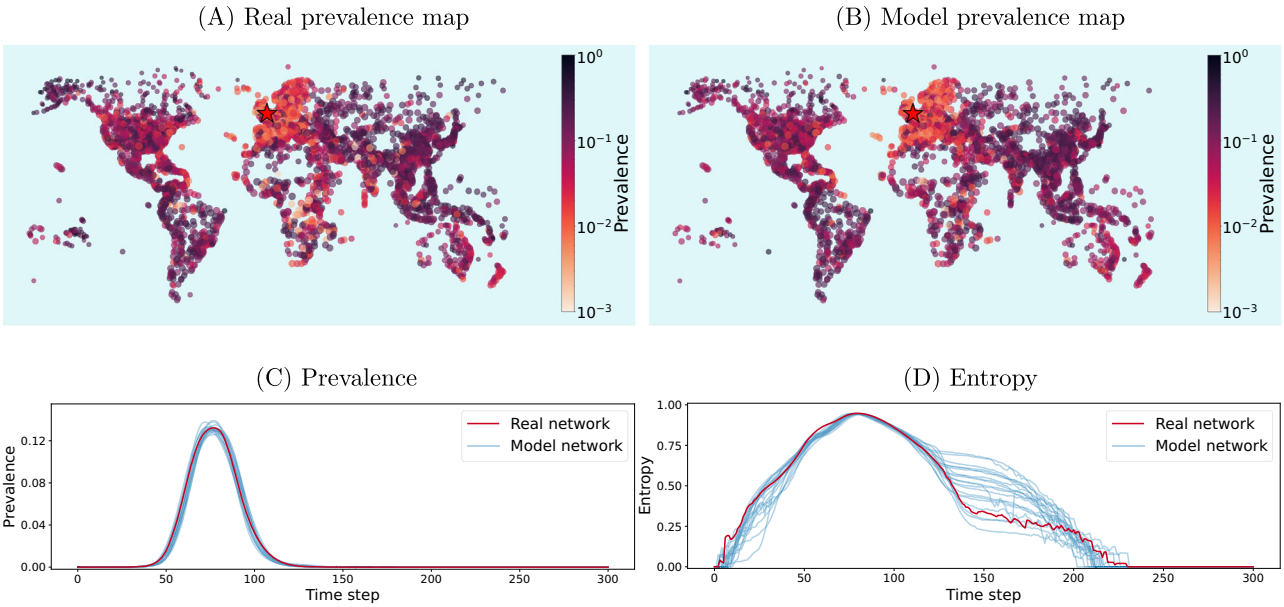


Fig. 3 | Epidemic simulations starting from London. A,**B** Prevalence map at simulation step $t = 80$ for the empirical (**A**) and model (**B**) WAN. Each point represents a basin placed at its geographical coordinates, with dimension proportional to degree and color representing the prevalence (only basins with a prevalence of at least 10^{-3} are displayed). **C**,**D** Time evolution of the prevalence (total fraction of infected individuals) (**C**) and of the entropy of the prevalence distribution (**D**), for one real simulation and 20 model simulations.

further improve performance. Future work may also develop extensions in which binary and weighted constraints jointly inform the reconstruction[62,63], or evaluate models for their ability to reproduce the features of more realistic contagion models.

Finally, we emphasize that our inference framework is analytic (*white-box*) and therefore complements, rather than competes with, *black-box* machine-learning approaches. Unlike methods that require access to complete network realizations for training, our framework reconstructs realistic WAN surrogates directly from aggregate information, without ever observing the full network: reconstruction relies only on low-dimensional OD information at the level of individual locations and geographic communities. Importantly, the model can also be informed with open data on individual flight legs and transit passenger counts, with limited loss of reconstruction accuracy (further information in Supplementary Note 4). Overall, this work shows that combining degree-strength constraints with coarse-grained spatial organization can be used to generate realistic WAN surrogates for forecasting and policy analysis.

Methods

Model formulation

Our modeling approach belongs to the class of Exponential Random Graphs (ERGs)[47,64–66], which build on the principle of Maximum Entropy[67] to generate ensembles of networks that are maximally random but preserve, on average, some features of the empirical network[7,68,69]. This recipe allows obtaining the most unbiased probability distribution describing a network with the imposed features, while model parameters can be efficiently obtained with scalable algorithms[70]. For this approach to work in the context of sparse weighted networks, the features to be imposed are both node strengths and degrees[62,71]. Indeed, without constraining degrees, the generated network would be fully connected, because strengths alone do not contain information on how many links should be actually realized (see the *irreducibility conjecture* in[72]): unconstrained entropy maximization distributes the available weight over all possible (in fact all) connections. When degree information is not available or for the sake of computational cost, a heuristic *fitness ansatz*[51] can be used to replace the corresponding model parameters with other node properties that are positively correlated with the degree. Typically the strengths themselves are used[52], with the underlying assumption is that nodes with large strength are likely to have many connections. The model can then be formulated as a two-step inference procedure[53,72], where links are first sampled independently according to a connection probability that depends on node strengths, and then assigned a weight according to a *corrected* gravity model prescription. This recipe thus generates sparse topologies while preserving node strengths.

Here we build on this formulation by further imposing a community structure on the model network while keeping the global strength constraints. We thus split nodes into a set $\{g\}$ of communities and, for each node i, we constrain the strength values computed with respect to each community g: $s_{ig} = \sum_{j(\neq i)\in g} w_{ij}$. The probability of connection for nodes i and j (with $i \neq j$), belonging respectively to community g_i and g_j, thus reads:

$$p_{ij} = \frac{z_{g_i g_j} s_{ig_j} s_{jg_i}}{1 + z_{g_i g_j} s_{ig_j} s_{jg_i}} \tag{1}$$

which depends on s_{ig_j} and s_{jg_i} (the required inputs of the model). The free parameter $z_{g_i g_j}$ is determined by setting the expected number of connections among communities g_i, g_j equal to the empirical value: $\sum_{i\in g_i}\sum_{j\in g_j} p_{ij} = \sum_{i\in g_i}\sum_{j\in g_j} a_{ij}$ (details in Supplementary Methods). Then, each sampled link ij (realized with probability p_{ij}) is assigned the weight:

$$w_{ij} = \frac{s_{ig_j} s_{g_i j}}{p_{ij} W_{g_i g_j}} \tag{2}$$

while $w_{ij} = 0$ otherwise. $W_{g_i g_j} = \sum_{i\in g_i}\sum_{j\in g_j} w_{ij}$ is the total weight of connections among nodes in g_i and g_j. The average link weight thus reads $\langle w_{ij}\rangle = s_{ig_j} s_{g_i j} / W_{g_i g_j}$, hence the model preserves, as ensemble averages, the strength by block of each node: $\langle s_{ig_j}\rangle = \sum_{j\in g_j}\langle w_{ij}\rangle \equiv s_{ig_j}$ (and thus the total strength). This model formulation can be seen as a composition of fitness models for each block[72] or as a fitness-based version of discrete latent variable models[54,55] (see Supplementary Methods for the model formulation tailored for directed networks).

Table 2 | Model calibrated on open data. Top: Reconstruction of topological properties (same quantities of Fig. 2)

	$RMSRE_s$	$RMSRE_k$	$RMSRE_p$	$RMSRE_w$	Recall	β	$MSE_{l(d)}$	M
Model	8.36	0.28	0.057	4.53	0.58	1.3(8)	0.56	0.18
	Δ area	**L2[i]**	**Peak time**	**Peak magn.**	**Epid. length**		**RMSE[i_m]**	**L2[H]**
Model	1.2 ± 0.5	0.17 ± 0.07	4.3 ± 2.4	0.0071 ± 0.0049	7.2 ± 2.4		0.0086 ± 0.0010	13.2 ± 2.8
Data	0.55 ± 0.34	0.076 ± 0.047	1.7 ± 1.4	0.0042 ± 0.0039	3.3 ± 2.7		0.0055 ± 0.0004	7.4 ± 3.6

For empirical filtered data, $\beta = 1.2(4)$ and $M = 0.26$. Bottom: Similarity of epidemic simulations on real and model networks (same quantities of Table 1).

Epidemic simulations

We implement metapopulation simulations of epidemic spreading by coupling a Susceptible-Infected-Recovered (SIR) dynamics taking place in each basin with mobility among different basins according to WAN passenger flows[9,73]. We use the *Epidemiology 101* package[74], available at https://github.com/DataForScience/Epidemiology101/tree/master. For each basin m, the number of susceptible, infected and recovered individuals, denoted S_m, I_m, and R_m (with $S_m + I_m + R_m = N_m$ the total population m), evolves according to a stochastic SIR dynamics, described by the mean-field equations

$$\begin{cases} \dfrac{dS_m}{dt} = -\dfrac{\beta I_m S_m}{N_m} \\ \dfrac{dI_m}{dt} = \dfrac{\beta I_m S_m}{N_m} - \gamma I_m \\ \dfrac{dR_m}{dt} = \gamma I_m \end{cases} \tag{3}$$

where β is the transmission rate and γ the recovery rate. The dynamics of the various basins are coupled through the OD travel matrix of the WAN[9], whose link weights determine the passenger flows among basins. Here we set $\pi_{mn} = w_{mn}/(N_m + N_n)$ as the probability that an individual travels from m to n in a given time step. Simulations start with an initial number of infected $I_0 = 10$ in a designated seed basin and evolve with a transmission rate $\beta = 0.3$ and a recovery rate $\gamma = 0.1$, for 300 time steps.

We select initial seeds as the four most interconnected basins from each continent (Africa, Asia, Europe, North America, South America, and Oceania), resulting in a set of 24 highly connected basins all over the world: Johannesburg, Cairo, Cape Town, Algiers, Hong Kong, Shanghai, Beijing, Bangkok, London, Paris, Moscow, Milan, New York, Los Angeles, San Francisco, Miami, Sydney, Melbourne, Brisbane, Auckland, Sao Paulo, Rio de Janeiro, Bogota, Santiago. For each seed, we performed 5 simulations on the real network and 20 simulations on different network realizations of the model, for a total of 504 simulations.

Note that by using OD data to model passenger flows among basins, we are not taking layovers into account within the spreading dynamics. Indeed, transmission during transit is expected to contribute less to epidemic dynamics than movements between origin and destination populations, for several reasons. In-flight transmission of respiratory pathogens is relatively uncommon due to high cabin air exchange rates and HEPA filtration[75]. Moreover, infected travelers are often still in the latent stage during travel, making progression to infectiousness after arrival the dominant mechanism shaping geographical spread. Transit dwell times are also substantially shorter than time spent at the destination.

Open data

Topological information. We use the OpenSky COVID-19 Flight Dataset[60], part of the OpenSky Network's crowdsourced ADS-B sensor data covering global air traffic from January 2019 onward. To match the temporal scope of the OD basin dataset, we restrict our analysis to the period Sept.-Nov. 2019. Each record in the OpenSky data reports the first and last detected position of a given flight; we use these coordinates to approximate the corresponding origin and destination airports, which we then map to their nearest basin. This procedure yields an open-data edge list, from which we derive topological information for each basin. Note that direct flight connections from OpenSky cannot fully capture the connectivity patterns of OD data, as routes with layovers are decomposed into individual legs. Indeed, degrees inferred from OpenSky are systematically underestimated; therefore, we rescale them to match the observed network density of OAG data.

Traffic statistics. We employ Wikidata to obtain passenger traffic volumes of all available airports for the period Sept.-Nov. 2019. We map (and sum) these values to the corresponding basins to obtain a proxy of basin strengths. In order to improve consistency with OD values from OAG, we filter out a handful of small outliers. Note that since Wikidata counts include transit passengers, obtained strengths are systematically higher than original OD values, but this effect is absorbed by automatic recalibration of model parameters. Lastly, to fully inform the model, we need passenger flows among regions. We obtain them by splitting Wikidata strengths according to rescaled connectivity-by-region obtained from OpenSky.

Supplementary Note 4 contain full details on the procedures described above, as well as results of the different model formulations when calibrated with open data.

Data availability

Raw OD data are available under commercial license from OAG. OpenSky data are available at http://www.opensky-network.organd Wikidata at https://www.wikidata.org/wiki/Wikidata:Main_Page.

Code availability

Open data and codes to implement the network generation models are available at the github repository https://github.com/mnlknt/WAN-fitness-modeling.

References

1. Durlauf, S. N. How can statistical mechanics contribute to social science? *Proc. Natl. Acad. Sci. USA* **96**, 10582–10584 https://doi.org/10.1073/pnas.96.19.10582 (1999).
2. De Martino, A. & Marsili, M. Statistical mechanics of socio-economic systems with heterogeneous agents. *J. Phys. A Math. Gen.* **39**, R465 https://doi.org/10.1088/0305-4470/39/43/R01 (2006).
3. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646 https://doi.org/10.1103/RevModPhys.81.591 (2009).
4. Perc, M. et al. Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51 https://doi.org/10.1016/j.physrep.2017.05.004 (2017).
5. Caldarelli, G., Wolf, S. & Moreno, Y. Physics of humans, physics for society. *Nat. Phys.* **14**, 870–870 https://doi.org/10.1038/s41567-018-0266-x (2018).
6. Barthelemy, M. The statistical physics of cities. *Nat. Rev. Phys.* **1**, 406–415 https://doi.org/10.1038/s42254-019-0054-2 (2019).

7. Cimini, G. et al. The statistical physics of real-world networks. *Nat. Rev. Phys.* **1**, 58–71 https://doi.org/10.1038/s42254-018-0002-6 (2019).
8. Bouchaud, J.-P. From statistical physics to social sciences: the pitfalls of multi-disciplinarity. *Ind. Corp. Change* **34**, 235–248 https://doi.org/10.1093/icc/dtaf002 (2025).
9. Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* **106**, 21484–21489 https://doi.org/10.1073/pnas.0906910106 (2009).
10. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 https://doi.org/10.1126/science.1245200 (2013).
11. Tizzoni, M. et al. On the use of human mobility proxies for modeling epidemics. *PLOS Computational Biol.* **10**, 1–15 https://doi.org/10.1371/journal.pcbi.1003716 (2014).
12. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 https://doi.org/10.1103/RevModPhys.87.925 (2015).
13. Oliver, N. et al. Mobile phone data for informing public health actions across the covid-19 pandemic life cycle. *Sci. Adv.* **6**, eabc0764 https://doi.org/10.1126/sciadv.abc0764 (2020).
14. Sills, J. et al. Aggregated mobility data could help fight covid-19. *Science* **368**, 145–146 https://doi.org/10.1126/science.abb8021 (2020).
15. Wesolowski, A. et al. Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270 https://doi.org/10.1126/science.1223467 (2012).
16. Kostandova, N. et al. A systematic review of using population-level human mobility data to understand sars-cov-2 transmission. *Nat. Commun.* **15**, 10504 https://doi.org/10.1038/s41467-024-54895-7 (2024).
17. Lu, X. et al. Human mobility in epidemic modeling. *Phys. Rep.* **1157**, 1–45 https://doi.org/10.1016/j.physrep.2025.10.010 (2026).
18. Flahault, A. & Valleron, A. A method for assessing the global spread of hiv–1 infection based on air travel. *Math. Popul. Stud.* **3**, 161–171 https://doi.org/10.1080/08898489209525336 (1992).
19. Grais, R. F., Hugh Ellis, J. & Glass, G. E. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *Eur. J. Epidemiol.* **18**, 1065–1072 https://doi.org/10.1023/A:1026140019146 (2003).
20. Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci.* **101**, 15124–15129 https://doi.org/10.1073/pnas.0308344101 (2004).
21. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. USA* **103**, 2015–2020 https://doi.org/10.1073/pnas.0510525103 (2006).
22. Hollingsworth, T. D., Ferguson, N. M. & Anderson, R. M. Will travel restrictions control the international spread of pandemic influenza? *Nat. Med.* **12**, 497–499 https://doi.org/10.1038/nm0506-497 (2006).
23. Lessler, J., Kaufman, J. H., Ford, D. A. & Douglas, J. V. The cost of simplifying air travel when modeling disease spread. *PLoS ONE* **4**, 1–5 https://doi.org/10.1371/journal.pone.0004403 (2009).
24. Wardle, J., Bhatia, S., Kraemer, M. U., Nouvellet, P. & Cori, A. Gaps in mobility data and implications for modelling epidemic spread: a scoping review and simulation study. *Epidemics* **42**, 100666 https://doi.org/10.1016/j.epidem.2023.100666 (2023).
25. National Academies of Sciences Engineering and Medicine. *Improving the CDC Quarantine Station Network's Response to Emerging Threats* (The National Academies Press, 2022).
26. Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA* **101**, 3747–3752 https://doi.org/10.1073/pnas.0400087101 (2004).
27. Zanin, M. & Lillo, F. Modelling the air transport with complex networks: a short review. *Eur. Phys. J. Spec. Top.* **215**, 5–21 https://doi.org/10.1140/epjst/e2013-01711-9 (2013).
28. Johansson, M. A. et al. On the treatment of airline travelers in mathematical models. *PLOS ONE* **6**, 1–7 https://doi.org/10.1371/journal.pone.0022151 (2011).
29. Meslé, M. M. I., Hall, I. M., Christley, R. M., Leach, S. & Read, J. M. The use and reporting of airline passenger data for infectious disease modelling: a systematic review. *Eurosurveillance* **24**, 1800216 https://doi.org/10.2807/1560-7917.ES.2019.24.31.1800216 (2019).
30. Nicolaides, C., Cueto-Felgueroso, L., González, M. C. & Juanes, R. A metric of influential spreading during contagion dynamics through the air transportation network. *PLoS ONE* **7**, e40961 https://doi.org/10.1371/journal.pone.0040961 (2012).
31. Gómez, S., Fernández, A., Meloni, S. & Arenas, A. Impact of origin-destination information in epidemic spreading. *Sci. Rep.* **9**, 2315 https://doi.org/10.1038/s41598-019-38722-4 (2019).
32. Wang, L. & Wu, J. T. Characterizing the dynamics underlying global spread of epidemics. *Nat. Commun.* **9**, 218 https://doi.org/10.1038/s41467-017-02344-z (2018).
33. Davis, J. T. et al. Cryptic transmission of sars-cov-2 and the first COVID-19 wave. *Nature* **600**, 127–132 https://doi.org/10.1038/s41586-021-04130-w (2021).
34. Klamser, P. P. et al. Inferring country-specific import risk of diseases from the world air transportation network. *PLoS Comput. Biol.* **20**, 1–26 https://doi.org/10.1371/journal.pcbi.1011775 (2024).
35. Wardle, J., Bhatia, S., Cori, A. & Nouvellet, P. Temporal variations in international air travel: implications for modelling the spread of infectious diseases. *J. Travel Med.* **31**, taae062 https://doi.org/10.1093/jtm/taae062 (2024).
36. Zipf, G. K. The p_1 p_2/d hypothesis: on the intercity movement of persons. *Am. Sociol. Rev.* **11**, 677–686 https://doi.org/10.2307/2087063 (1946).
37. Grosche, T., Rothlauf, F. & Heinzl, A. Gravity models for airline passenger volume estimation. *J. Air Transp. Manag.* **13**, 175–183 https://doi.org/10.1016/j.jairtraman.2007.02.001 (2007).
38. Barbosa, H. et al. Human mobility: models and applications. *Phys. Rep.* **734**, 1–74 https://doi.org/10.1016/j.physrep.2018.01.001 (2018).
39. Simini, F., Barlacchi, G., Luca, M. & Pappalardo, L. A deep gravity model for mobility flows generation. *Nat. Commun.* **12**, 6576 https://doi.org/10.1038/s41467-021-26752-4 (2021).
40. Cabanas-Tirapu, O., Danús, L., Moro, E., Sales-Pardo, M. & Guimerà, R. Human mobility is well described by closed-form gravity-like models learned automatically from data. *Nat. Commun.* **16**, 1336 https://doi.org/10.1038/s41467-025-56495-5 (2025).
41. Ehsani, S., Sergeeva, E., Murdy, W. & Fox, B. Predicting the skies: a novel model for flight-level passenger traffic forecasting, https://arxiv.org/abs/2401.03397 (2024).
42. Huang, Z., Wu, X., Garcia, A. J., Fik, T. J. & Tatem, A. J. An open-access modeled passenger flow matrix for the global air network in 2010. *PLoS ONE* **8**, 1–9 https://doi.org/10.1371/journal.pone.0064317 (2013).
43. Mao, L., Wu, X., Huang, Z. & Tatem, A. J. Modeling monthly flows of global air travel passengers: an open-access data resource. *J. Transp. Geogr.* **48**, 52–60 https://doi.org/10.1016/j.jtrangeo.2015.08.017 (2015).
44. Guimerà, R., Mossa, S., Turtschi, A. & Amaral, L. A. N. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. USA* **102**, 7794–7799 https://doi.org/10.1073/pnas.0407994102 (2005).
45. Diop, I. M., Cherifi, C., Diallo, C. & Cherifi, H. Revealing the component structure of the world air transportation network. *Appl. Netw. Sci.* **6**, 92 https://doi.org/10.1007/s41109-021-00430-2 (2021).

46. Guimerà, R. & Amaral, L. A. N. Modeling the world-wide airport network. *Eur. Phys. J. B - Condens. Matter* **38**, 381–385 https://doi.org/10.1140/epjb/e2004-00131-0 (2004).
47. Bianconi, G. The entropy of randomized network ensembles. *Europhys. Lett.* **81**, 28005 https://doi.org/10.1209/0295-5075/81/28005 (2008).
48. Bianconi, G.*Information theory of spatial network ensembles*, chap. 5 (Edward Elgar Publishing, 2021).
49. Picciolo, F., Squartini, T., Ruzzenenti, F., Basosi, R. & Garlaschelli, D. The role of distances in the world trade web. *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems* 784–792 https://doi.org/10.1109/SITIS.2012.118 (2012).
50. Squartini, T., Caldarelli, G., Cimini, G., Gabrielli, A. & Garlaschelli, D. Reconstruction methods for networks: The case of economic and financial systems. *Phys. Rep.* **757**, 1–47 https://doi.org/10.1016/j.physrep.2018.06.008 (2018).
51. Caldarelli, G., Capocci, A., Rios, P. D. L. & Mun^oz, M. A. Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* **89**, 258702 https://doi.org/10.1103/PhysRevLett.89.258702 (2002).
52. Garlaschelli, D. & Loffredo, M. I. Fitness-dependent topological properties of the world trade web. *Phys. Rev. Lett.* **93**, 188701– https://doi.org/10.1103/PhysRevLett.93.188701 (2004).
53. Cimini, G., Squartini, T., Garlaschelli, D. & Gabrielli, A. Systemic risk analysis on reconstructed economic and financial networks. *Sci. Rep.* **5**, 15758 https://doi.org/10.1038/srep15758 (2015).
54. Karrer, B. & Newman, M. E. J. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 https://doi.org/10.1103/PhysRevE.83.016107 (2011).
55. Fronczak, P., Fronczak, A. & Bujok, M. Exponential random graph models for networks with community structure. *Phys. Rev. E* **88**, 032810 https://doi.org/10.1103/PhysRevE.88.032810 (2013).
56. Official aviation guide (OAG) worldwide flight and passenger data (2019).
57. Serafino, M. et al. True scale-free networks hidden by finite size effects. *Proc. Natl. Acad. Sci. USA* **118**, e2013825118, https://doi.org/10.1073/pnas.2013825118 (2021).
58. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103**, 8577–8582 https://doi.org/10.1073/pnas.0601602103 (2006).
59. Desiderio, A., Salina, G. & Cimini, G. Multiplex mobility network and metapopulation epidemic simulations of Italy based on open data. *J. Phys. Complex.* **3**, 04LT01 https://doi.org/10.1088/2632-072X/ac9a29 (2022).
60. Strohmeier, M., Olive, X., Lübbe, J., Schäfer, M. & Lenders, V. Crowdsourced air traffic data from the opensky network 2019–2020. *Earth Syst. Sci. Data* **13**, 357–366 https://doi.org/10.5194/essd-13-357-2021 (2021).
61. Bernaschi, M., Celestini, A., Guarino, S., Mastrostefano, E. & Saracco, F. The fitness-corrected block model, or how to create maximum-entropy data-driven spatial social networks. *Sci. Rep.* **12**, 18206 https://doi.org/10.1038/s41598-022-22798-6 (2022).
62. Gabrielli, A., Mastrandrea, R., Caldarelli, G. & Cimini, G. Grand canonical ensemble of weighted networks. *Phys. Rev. E* **99**, 030301 https://doi.org/10.1103/PhysRevE.99.030301 (2019).
63. Parisi, F., Squartini, T. & Garlaschelli, D. A faster horse on a safer trail: generalized inference for the efficient reconstruction of weighted networks. *N. J. Phys.* **22**, 053053 https://doi.org/10.1088/1367-2630/ab74a7 (2020).
64. Park, J. & Newman, M. E. J. Statistical mechanics of networks. *Phys. Rev. E* **70**, 066117 https://doi.org/10.1103/PhysRevE.70.066117 (2004).
65. Cover, T. M. & Thomas, J. A.*Elements of Information Theory* (Wiley, 2005).
66. Fronczak, A.*Exponential Random Graph Models*, 810–826 (Springer New York, 2018).
67. Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 https://doi.org/10.1103/PhysRev.106.620 (1957).
68. Squartini, T. & Garlaschelli, D. Analytical maximum-likelihood method to detect patterns in real networks. *N. J. Phys.* **13**, 083001 https://doi.org/10.1088/1367-2630/13/8/083001 (2011).
69. Squartini, T. & Garlaschelli, D.*Maximum-Entropy Networks* (Springer International Publishing, 2017).
70. Vallarano, N. et al. Fast and scalable likelihood maximization for exponential random graph models with local constraints. *Sci. Rep.* **11**, 15227 https://doi.org/10.1038/s41598-021-93830-4 (2021).
71. Mastrandrea, R., Squartini, T., Fagiolo, G. & Garlaschelli, D. Enhanced reconstruction of weighted networks from strengths and degrees. *N. J. Phys.* **16**, 043022 https://doi.org/10.1088/1367-2630/16/4/043022 (2014).
72. Cimini, G., Squartini, T., Gabrielli, A. & Garlaschelli, D. Estimating topological properties of weighted networks from limited information. *Phys. Rev. E* **92**, 040802 https://doi.org/10.1103/PhysRevE.92.040802 (2015).
73. Balcan, D. et al. Modeling the spatial spread of infectious diseases: the global epidemic and mobility computational model. *J. Comput. Sci.* **1**, 132–145 https://doi.org/10.1016/j.jocs.2010.07.002 (2010).
74. Gonçalves, B. Epidemiology 101. https://bgoncalves.medium.com/list/epidemiology-101-185b95a0f345.
75. Hertzberg, V. S. et al. Behaviors, movements, and transmission of droplet-mediated respiratory diseases during transcontinental airline flights. *Proc. Natl. Acad. Sci. USA* **115**, 3623–3627 https://doi.org/10.1073/pnas.1711611115 (2018).

Author contributions

Designed the study: G.C.i., A.V. and G.C.a. Curated the OAG data: J.D. and A.L. Collected the Open data: A.M. and G.C.i. Contributed code and repository: G.F. and A.M. Carried out the analysis: G.F., A.M. and G.C.i. Analyzed results: G.F., A.M., G.C.i., J.D., A.L., A.V. and G.C.a. Prepared the manuscript and figures: G.F., A.M. and G.C.i. Revised the manuscript: G.F., A.M., G.C.i., J.D., A.L., A.V. and G.C.a.

Funding

A.M. and G.Ci. acknowledge funding from the Italian Ministry of University and Reseach, under the program PRIN 2022 (grant number 2022MTBB22) and MUR PRIN 2022 PNRR (grant number P2022E93B8). G. Ca. acknowledge funding from the Italian Ministry of University and Reseach, under the program PRIN 2022 (grant number 2022KWTEB7), and thanks the COGNIA visiting scheme at LIMS.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-78115-6.

Correspondence and requests for materials should be addressed to Giulio Cimini or Guido Caldarelli.

Peer review information : *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.