Characterizing US spatial connectivity: implications for geographical disease dynamics and metapopulation modeling

Giulia Pullano[1], Lucila G. Alvarez-Zuzek[2], Vittoria Colizza[3], and Shweta Bansal[1*]

[1]Department of Biology, Regents Hall, Georgetown University, 37th and O Streets NW, Washington, DC, 20057-1229, USA
[2]Fondazione Bruno Kessler, Trento, Italy
[3]INSERM, Paris France
[*]Corresponding author: Shweta Bansal, email: sb753@georgetown.edu

Summary

Background Human mobility is expected to be a critical factor in the geographic diffusion of infectious diseases, and this assumption led to the implementation of social distancing policies during the early fight against the COVID-19 emergency in the United States. Yet, because of substantial data gaps in the past, what still eludes our understanding are the following questions: 1) How does mobility contribute to the spread of infection within the United States at local, regional, and national scales? 2) How do seasonality and shifts in behavior affect mobility over time? 3) At what geographic level is mobility homogeneous across the United States? Addressing these questions is critical to developing accurate transmission models, predicting the spatial propagation of disease across scales, and understanding the optimal geographical and temporal scale for the implementation of control policies.

Methods We address this problem using high-resolution human mobility data measured via mobile app usage. We compute the daily connectivity network between US counties to understand the spatial clustering and temporal stability of mobility patterns. We then integrate our mobility data into a spatially explicit transmission model to reproduce the national invasion of the first wave of SARS-CoV-2 in the US, and characterize the impact of the spatio-temporal scale of mobility data on disease predictions.

Findings Temporally, we observe that intercounty connectivity is annually stable, and was unperturbed by mobility restrictions during the early phase of the COVID-19 pandemic, despite significant changes in overall activity. Spatially, we identify 104 geographic clusters of US counties that are highly connected by mobility within the cluster and more sparsely connected to counties outside the cluster. Together, these results suggest that intercounty connectivity in the US is relatively static across time and is highly connected at the sub-state level. We find that the stability in temporal patterns allows static mobility data to effectively capture infection dynamics. On the other hand, spatial uniformity at the sub-state (cluster)-scale does not capture spatial dynamics; instead, mobility data at the county-scale is necessary to better predict spatial disease diffusion.

Interpretation Our work demonstrates that intercounty mobility was negligibly affected outside the lockdown period of Spring 2020, explaining the broad spatial distribution of COVID-19 outbreaks in the US during the early phase of the pandemic. Such geographically dispersed outbreaks place a significant strain on national public health resources and necessitate complex metapopulation modeling approaches for predicting disease dynamics and control design. We thus inform the design of such metapopulation models to balance high disease predictability

with low data requirements.

Introduction

Human mobility plays a crucial role in the spread of respiratory diseases [17, 41]. The combination of regional travel and local commuting represents the spatial connectivity between locations, serving as the main driver in the geographic diffusion of infectious diseases. Characterizing the spatial dynamics of pathogen transmission is, therefore, intricately tied to unraveling human mobility patterns. Such a task has proven to be challenging due to the inherent complexity and privacy-related limitations on collecting mobility data [7]. Over the past few decades, researchers have extensively relied on mobility data obtained from census records, surveys, transportation statistics, commuting data, and international air traffic data. Such datasets have widely contributed to a better understanding of human mobility patterns and their impact on the epidemic spread [1, 6, 3, 11, 39, 2, 33], but can be limited in their resolution or scale. More recently, this gap has been filled by the use of mobile phone data [10, 14], primarily based on phone records, but no such data has been available in the United States.

The global health crisis triggered by COVID-19 has underscored the critical need for swift access to mobility to help mitigate the spread of the virus. The urgency of the situation prompted an unprecedented sharing of data by private companies worldwide, through legally and ethically compliant agreements. This data was based on mobile location-based app usage and thus provided incomparable access to high-resolution, large-scale, and near-real-time mobility data and has expanded human mobility science [37], and computational epidemiology [19, 18]. The availability of this data has especially represented a shift in US public health and it has been used to inform epidemic models and reveal the impact of mitigation strategies on behavior[27, 38, 35, 32, 20, 26, 15, 25]. While the association between mobility patterns and COVID-19 transmission in the USA has been extensively studied, no studies have been devoted to assessing when the underlying mobility network needs to be embedded into models to characterize epidemic spread.

Moreover, the effects of control measures on human mobility at mesoscale (i.e., intermediate or regional level of geographical granularity) and long-range (i.e., entire countries, continents), as well as the most suitable geographical and temporal granularity for implementing these measures, still lack clarity. This gap in understanding the characteristic spatio-temporal scale of mobility not only limits target control policies but also our ability to model transmission dynamics effectively. To date, mobility data have been integrated into epidemic models without due consideration for the optimal geographical (e.g. municipalities, regions, states) and temporal resolution (e.g., day, week, month) required to accurately capture epidemic spread. The level of granularity used in these models has consistently been dictated by a priori assessments from data providers [28, 34].

To address these gaps, we aim to characterize the spatiotemporal characteristic scale of human mobility in the United States, for the periods before and after the pandemic emergency by using mobile phone data. Furthermore, to assess if mobility was relevant in the spread of COVID-19 and which mobility scale drove the invasion, we integrate mobility data into spatially-explicit transmission models to reproduce the national invasion of the first wave of SARS-CoV-2 in the US. More specifically, we observe the daily intercounty connectivity in the US in time and space. To assess the role of mobility, we evaluate how the predictability of the model depends on the underlying mobility data scale, and how model predictive power is impacted by not accounting for fine-grain mobility.

Methods

Our study has two objectives: i) to characterize the temporal and spatial scale of variation in the connectivity between U.S. counties influenced by human mobility and ii) to inform metapopulation models with mobility data at different scales, thereby informing the scale of data required for epidemic predictability. To achieve this, we analyze human mobility data measured via mobile app usage across spatial and temporal scales. We then integrate these mobility data into an epidemic model parameterized with national COVID-19 public health data to evaluate the performance of the model at different scales of mobility to predict epidemic spatial invasion.

Characterizing intercounty connectivity with mobility data

We access data from SafeGraph [45] (now called Advan Patterns), which collects and shares mobility data based on location-based mobile app usage. In particular, we use the daily Social Distancing dataset provided by Safegraph (see Supplementary Information for dataset selection), and use information on the number of mobile devices with a home in an origin census block group that visit a given destination census block group or staying in the origin one for at least a minute. Data span the period from January 2019 to April 2021 on a daily timescale. We aggregate the data to the US county level to capture a common geographic scale for disease surveillance and public health decision-making. To address the spatial and temporal heterogeneity in the observed devices obs_i within each county i (Figure S1), we developed a correction factor:

$$\gamma_i(t) = \left(\frac{obs_i(t)}{pop_i} \right) \tag{1}$$

where pop_i is the population in the county i. To reduce sampling biases, we exclude the bottom 25% of counties by population size, running all analyses on 2327 geographical counties within the continental US of a population size greater than 11,000.

After rescaling for the correction factor, to quantify intercounty connectivity monthly, we normalize the number of visits between an origin and destination county by the daily total number of visits originating in the origin county. We then compute the average daily visits each month from Jan 2019 to March 2021 for all pairs of counties. In Figure 1A is shown for illustrative proposes the spatial connectivity network computed in March 2020. We therefore obtain a time-evolving connectivity network between US counties, in which the links represent the daily coupling probabilities p_{ij} between any pair i and j of US counties. p_{ij} is normalized as follows: $\sum_{i \neq j} p_{ij} + p_{ii} = 1$. A comparison between the monthly and daily dataset provided by SafeGraph is reported in Figure S2.

Early phase of COVID-19 in the US

The initial confirmed case of COVID-19 in the United States was reported in Washington state on January 21, and within a few weeks, local transmission was established. Guidelines advocating for social distancing and the avoidance of gatherings were released on March 16. During the early stages of the COVID-19 pandemic, while most European countries implemented national restrictions, lockdowns and stay-at-home orders in the United States -like in Chile - were implemented locally and at diverse times. The peak of lockdowns was in April 2020, when more than 40 states had issued some form of stay-at-home or shelter-in-place order [16]. However, the spatial spread of COVID-19 was not contained, and at the end of June 2020, most US counties reported COVID-19 cases. Facing a rebound of cases in the fall of 2020, social distancing was recommended (but not mandated) to maintain epidemic activity at low levels.

COVID disease incidence data was derived from CDC data [**empty citation**]. In this work, we use COVID-19-related daily new reported cases and the time of arrival in any US county,

i.e. the day when at least 10 cases have been reported in that area. daily new reported cases and time of arrivals have been corrected by county accounting for undereporting. We estimate the level of ascertainment using reported data on COVID-19 cases and fatalities globally, as described in [**russell“˙reconstructing“˙2020**]. (Figure 1 B-C).

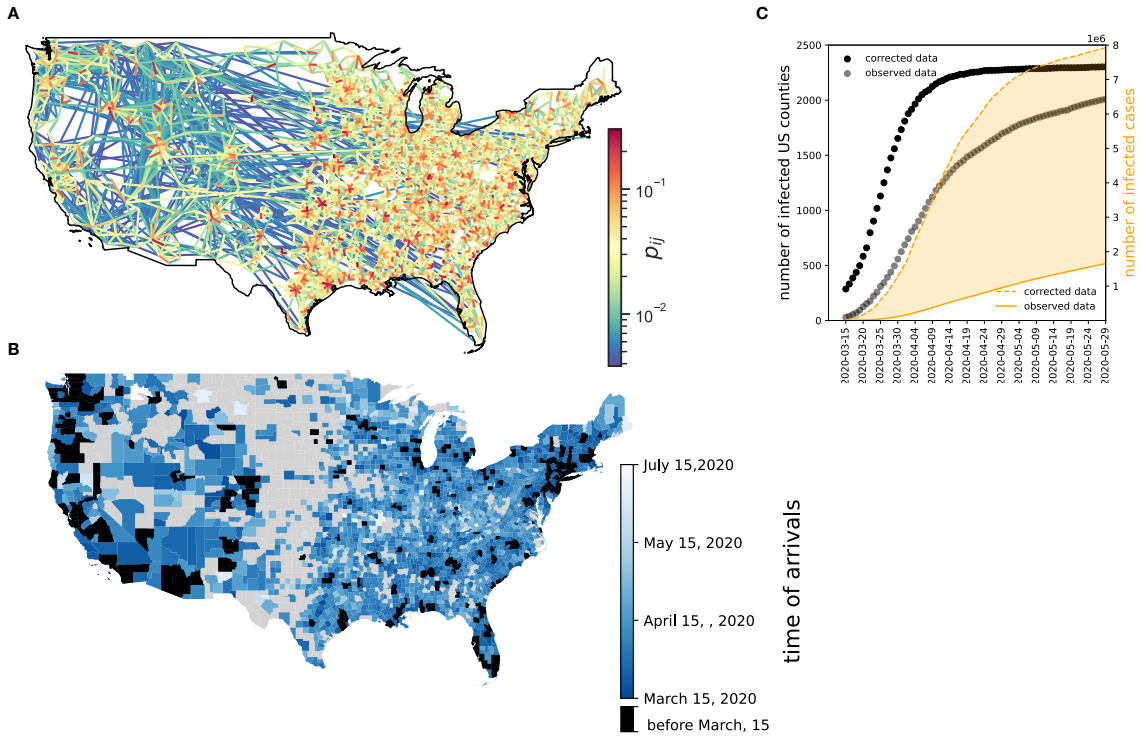

Figure 1: Data sources and epidemic context. (A) Mobility data. Figure shows the spatial connectivity network between US counties on March, 2020. Only links with top 1% of coupling probability are shown. (B-C) Public health data. (B) Color coded map shows the corrected time of arrival by county i.e. the time when at least 10 infected cases are reported. Black colored counties are counties that have been infected before March, 15. (C) Black dots show the corrected number of infected counties over time. A county is considered infected if it reports at least 10 confirmed COVID-19 cases. Grey dots show the observed number of infected counties over time. A county is considered infected if it reports at least 10 confirmed COVID-19 cases.Orange solid line shows the daily number of new confirmed cases nationally, while the orange dotted line shows the real number of cases accounting for underreporting.

Describing temporal and spatial variability in the mobility network

We examine the monthly network structure to evaluate the temporal dynamics of mobility patterns. We quantify the degree by county i.e. the number of connections (edges) the counties in the network have to other counties. We also define link persistence as the probability that links between counties that exist with non-zero mobility during the month of 2019 remain present in the same month during 2020 and 2021.

We also fitted a gravity model to the intercounty connectivity network for each month. For technical details and model performance, see Supplementary Information (SI).

To detect the different geographical communities generated by human mobility patterns, we performed a community detection analysis using the stochastic InfoMap algorithm [13]. We aim to identify regions where within movements occur more frequently compared to movements to other regions. To account for stochasticity, we use a bootstrap resampling method (see SI for details).

The classification of urban and rural areas was determined based on the NCHS Urban-Rural Classification Scheme for counties [31].

All network analyses were done using Python's `networkX` library, and the gravity model fitting was done using the Python `scikit-mobility` library.

Incorporating human mobility into infectious disease models

We used a stochastic non-Markovian transmission model with a metapopulation structure at the US county level [44]. In each county, the model accounted for disease transmission proportional to (i) infected residents not moving (ii) infected visitors coming from other counties and (iii) returning residents previously infected in other counties. The resulting force of infection in the county i is defined as follows:

$$\lambda_i = \lambda_{ii} + \sum_{i \neq j} \lambda^v_{ji} + \sum_{i \neq j} \lambda^r_{ij} \qquad with \begin{cases} \lambda_{ii} = \beta p^2_{ii} \frac{I_i}{\hat{N}_i} \\ \lambda^v_{ji} = \beta p_{ii} p_{ji} \frac{I_j}{\hat{N}_i} \\ \lambda^r_{ij} = \beta p_{ij} \frac{\hat{I}_j}{\hat{N}_j} \end{cases} \tag{2}$$

where p_{ij} is the coupling probability between patches i and j extracted from the intercounty connectivity network. The effective population, and effective number of infections, are respectively defined as follows:

$$\hat{N}_i = p_{ii} N_i + \sum_{i \neq j} p_{ji} N_j \tag{3}$$

$$\hat{I}_i = p_{ii} I_i + \sum_{i \neq j} p_{ji} I_j \tag{4}$$

We considered SEIR (Susceptible - Exposed - Infectious - Recovered) epidemic dynamics specific to COVID-19. Epidemics parameters are informed by [29].

The detailed mathematical framework, model calibration and implementation details can be found in the Supplement.

Inference framework and Goodness of fit

To effectively calibrate the epidemic pathway, we utilize national-level data on the count of counties with a minimum of 10 reported infected cases over time (see Figure 1C). The calibration process was conducted within the time frame spanning March 14, 2020, to Sept 15, 2020, a period during which all counties had reported instances of infection. During this process, we derive the parameter estimates β_{pre-LD} for the time interval from March 15 to March 31, and $\beta_{post-LD}$ for the period from March 31 to May 15, 2020. Inference is based on maximum likelihood, assuming a Poisson distribution to model the reported number of infected cases over time.

To assess model performance, we computed the goodness of fit to compare the estimated invasion probability $p_{i,inv}(t)$ with the observed early phase COVID-19 spatial invasion. $p_{i,inv}(t)$ denotes the likelihood for a county i in a day t to report at least 10 infected cases in the simulation [9]. The goodness of fit is defined as follows:

$$\begin{aligned} G(t) = \sum_i I_i log p_{i,inv}(t)) + \\ (1 - I_i) \, log \, (1 - p_{i,\ inv}(t)) \end{aligned} \tag{5}$$

$I_i = 1$ if the county i have been at least 10 infected cases at the day t, and $I_i = 0$ otherwise.

Comparing models across geographical scales

In designing the metapopulation structure across various spatial scales, we maintain the spatial scale of the model at the county level. This approach enables us to make meaningful comparisons of the results across different scales. To specify a metapopulation model at a spatial scale R (e.g., cluster, state), we randomize the mobility links that occur among all counties within a patch at scale R, preserving the number of links. We then set the coupling probabilities for the connected counties within each patch R equal to the average coupling probability of the links within the patch, as follows:

$$p_{ij}^{R} = \left(\frac{\sum_{l,k \in R}^{N} p_{lk}}{N} \right) \tag{6}$$

with $i, j \in R$, while the coupling probabilities across patches are not changed.

Results

Human mobility is expected to be a pivotal driver in infectious disease transmission. Understanding the impact of mobility on infection spread at local, regional, and national scales is therefore imperative for precision in transmission models, predicting disease spread, and optimizing targeted control strategies. By analyzing US county-level spatial connectivity using mobile phone data, we assess the temporal and geographical variability of human mobility and identify the geographical scale that drove the early phase of COVID-19 spatial invasion. As we address public health questions, we reveal the characteristic scale to design metapopulation modeling.

Temporal stability of the intercounty connectivity network

Limited changes in mobility are observed from January 2019 to March 2021, except for a significant impact localized to April 2020 (Figure 2A). This notable transition coincided with the implementation of lockdown measures, causing a nationwide decline in mobility from roughly 45 million daily visits to about 25 million visits post-lockdown enforcement. The mobility shock extended throughout the month, encompassing a transitional period (Figure 2A-B). Analyzing the temporal evolution of the intercounty connectivity network, we discovered a consistent seasonal pattern in the degree distribution and the persistence of mobility connections. Local variations were observed, only in April 2020, with a 23% reduction in degree and a 26% reduction in link persistence, respectively. Surprisingly, no variation was observed in November 2020, despite the strong recommendations for social distancing ahead of the winter surge of SARS-CoV-2. The reduction in Rural-Urban connections is particularly pronounced, with a 25% decrease compared to the pre-reduction value in February 2020. This decrease stabilized in May and beyond. Notably, Urban-Urban connections exhibited greater resilience over time when compared to connections involving rural areas. Moreover, while coupling probabilities stay consistent over the study period (Figure 1C), the likelihood of staying in the home location exhibited larger variability (See SI).

In spite of sporadic extreme events leading to local variability, the intercounty connectivity network demonstrated temporal stability and exhibited a high level of predictability through a gravity fit model as shown in Figure S7-9. Indeed, the Spearman coefficient between the original and modeled intercounty connectivity network remains constant over time, averaging 0.55.

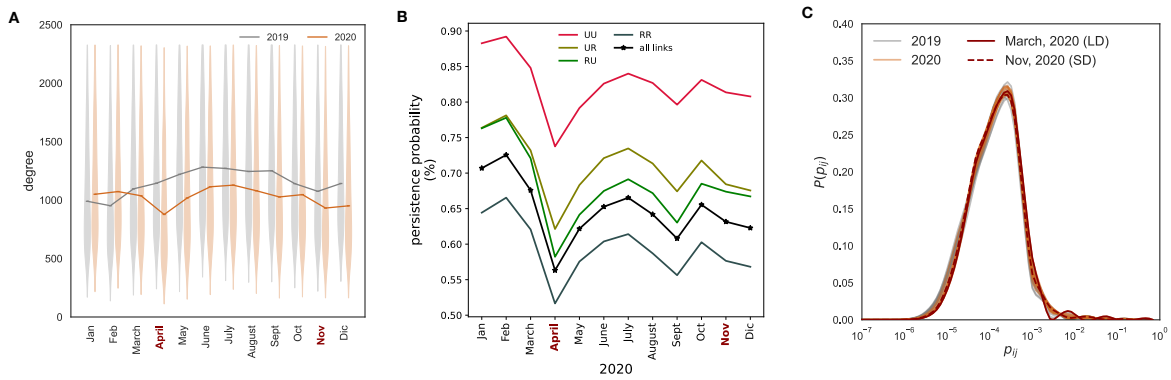

Figure 2: Temporal stability in intercounty mobility. (A) The monthly degree distribution of the intercounty network show significant variability but are consistent across months. (B) The persistence probability of links is illustrated, denoting the likelihood that a connection existing in 2019 remains present in 2020 and 2021. The plot provides a breakdown for different link types: urban-urban (UU), urban-rural (UR), rural-urban (RU), and rural-rural (RR) links. (C) Distribution of coupling probabilities in the connectivity network by month. We highlighted in darkred i) April, 2020 (LD) that represents the peak time of number of US states in lockdown, and November, 2020 (SD) that represents the period when social distancing recommendations were in place.

Spatial stability of the intercounty connectivity network

To identify the geographic scale at which mobility is highly connected, we detect clusters of counties that are more connected via mobility within the cluster than outside the clusters, we use a network community detection algorithm. Our hypothesis was that this partitioning of the US would be at a geographic scale larger than 3143 US counties but smaller than 50 US states or 10 HHS regions. Indeed, we find that based on human mobility, the US can be partitioned into around 100 regions that split most US states into multiple clusters (Figure 3). We also find that these clusters are highly spatially contiguous and respect state boundaries (with a similarity measured by normalized mutual information (NMI) as 0.82). Furthermore, these regions demonstrate stability over time (NMI = 0.95) despite the perturbations of the early phases of the COVID pandemic (Figure 3B and in Figure S9-10). Thus, we identify a persistent geographic partitioning of the US in which clusters are more connected within than between, and hypothesize that the relevance of mobility to the spatial diffusion of infectious diseases occurs at a mesoscale.

Implications for metapopulation disease models

After analyzing the stability of mobility patterns in both space and time, we evaluated how the spatiotemporal scale of human mobility affects our ability to effectively model metapopulation dynamics of disease. To address this, we integrate the connectivity network into a spatially explicit metapopulation model, with the goal of simulating the national spread of the initial wave of SARS-CoV-2 in the US. To understand the role of the geographic scale of mobility on disease dynamics, we integrate networks into the disease model that always at the US county level, but are homogenized at different spatial scales to represent missing information at different scales. To investigate the influence of temporal scale on model effectiveness, we inform the model with either a time-evolving connectivity network or a static connectivity network representing mobility from March, 2020 (without loss of generalizability given the temporal stability of the network we discuss above). In all cases, we measured goodness of fit by comparing the model predicted time of arrival of disease in a county to the observed time of arrival of the disease.

In Figure 4A, we demonstrate that a metapopulation disease model informed by a county-level intercounty connectivity network is highly predictive of observed early COVID-19 spatial

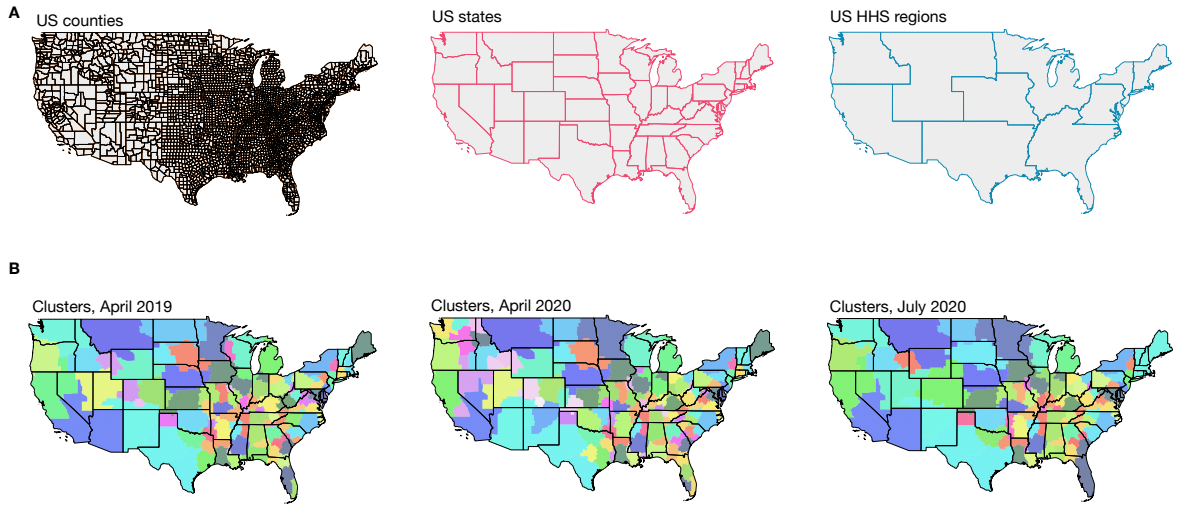

Figure 3: Analysis of Spatial Stability. (A) Geographical subdivisions at the county and state levels within the US, as well as the division into HHS (Health and Human Services) regions, used for health administration purposes. (B) We partition the intercounty mobility network so that each cluster of counties is more strongly connected to each other via mobility than to counties in other clusters. We find highly consistent partioning based on mobility networks from April 2019, April 2020 (during the mandated lockdown period of the early COVID pandemic) and July 2020 (after the lockdown period of the early COVID pandemic). Clusters are colored to delineate cluster boundaries and do not represent any other information. Counties colored in gray have populations of fewer than 11,000 inhabitants and are excluded from the analysis.

diffusion. Relative to an Erdős–Rényi random intercounty connectivity network, the empirical mobility network has a stronger goodness of fit throughout the early phase of the pandemic. This emphasizes the crucial role of human mobility in the spatial spread of the initial SARS-CoV-2 wave and underscores the necessity of accessing mobility data for constructing more reliable models. Additionally, we find that the spatial invasion predicted by a static mobility network is highly consistent with the prediction from a time-varying mobility network, suggesting that static mobility data is sufficient to accurately reproduce epidemic spatial heterogeneity.

In Figure 4B, we demonstrate the impact of the spatial scale of metapopulation structures on predictions of spatial diffusion. We compared the ability of county-level mobility data to predict spatial invasion, relative to mobility data at three spatial scales: US HHS regions, US states, and mobility data-based clusters (as defined in Figure 3). When comparing spatial scales, we find that all scenarios perform very similiarly to the random network. These results suggest that the state, cluster and HHS region-level mobility data lack the necessary granularity and fail to provide adequate insights into the diversity of mobility patterns.

Discussion

Since the onset of the COVID-19 pandemic, mobile phone data has played a crucial role in addressing the public health crisis [27, 38, 35, 32, 20, 26, 34, 22, 43, 40, 29, 21, 19, 18]. During this period, numerous network operators and private enterprises have made considerable efforts to swiftly share their data within the confines of legal regulations. Consequently, researchers worldwide have embarked on working with this data, monitoring human behavior caused by containment measures and adaptive responses to the epidemic, and utilizing it to enhance epidemic models in order to increase their reliability. [22, 43, 40, 29, 26, 21, 20, 34].

While static mobility data have predominantly been analyzed and integrated into models over the past decades [4, 5, 6], the current accessibility to real-time human behavior data prompts an

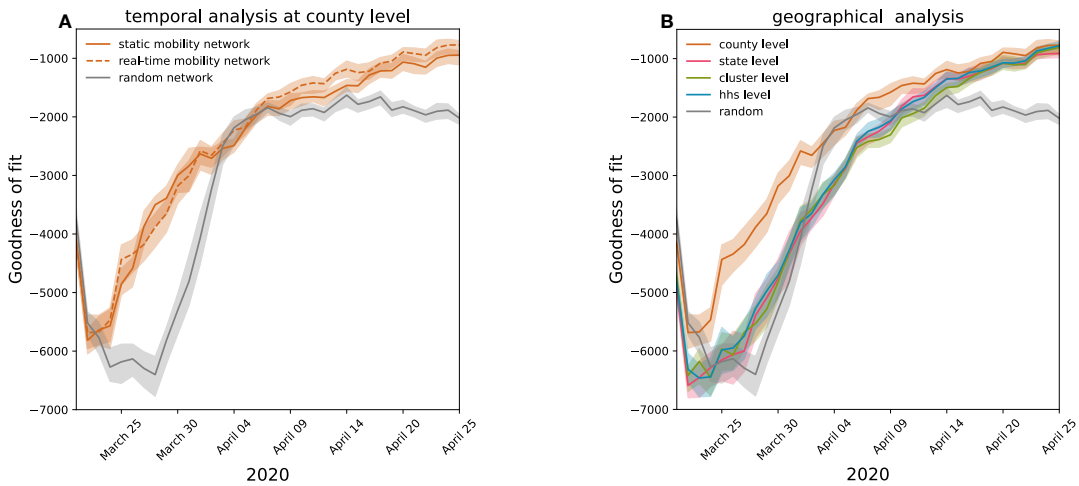

Figure 4: Implications of temporal and spatial scale of mobility data for prediction with metapopulation disease models. (A) The goodness of fit (median and 95% confidence interval) for time of arrivals for metapopulation models at a county level, informed by a time-evolving intercounty connectivity network, a static intercounty connectivity network, and a random intercounty network. (B) Solid lines show the goodness of the fit (median and 95% confidence interval) for metapopulation models informed with state-, cluster-, and HHS level mobility data.

essential investigation into the optimal scenarios for utilizing this dynamic information versus relying solely on static representations of reality [30]. Equally important is the exploration of the characteristic mobility scale to comprehensively capture the intricate coupling between different locations, a consideration with potential implications for target control policies to reduce epidemic activity, and for improving epidemic model forecasting. Furthermore, numerous researchers have emphasized the pressing necessity to implement standardized strategies that facilitate rapid data access while upholding stringent data privacy measures [12].

To answer this gap in the literature, in this study, we investigate the spatial connectivity of US counties during the early phase of the COVID-19 pandemic using high-resolution real-time human mobility data obtained from mobile phone usage. Our findings reveal significant insights into the dynamics of human mobility and their implications for infectious disease modeling. In contrast to findings from other countries (e.g., France [22], India [23], Germany [24]), we observe that despite the implementation of local social distancing measures and lockdowns, intercounty connectivity remained largely unperturbed, leading to rapid geographic diffusion of SARS-CoV-2. Mobility patterns experience only marginal changes before and after the early-stage COVID-19 pandemic in March-April 2020. The most notable disruption occurred during the first lockdown period in April 2020, when mobility sharply declined. However, this reduction was short-lived, and mobility patterns quickly rebounded. Notably, even during periods of social distancing recommendations, the mobility network remains relatively stable. Assuming the lockdown represents the most extreme form of mobility disruption, the temporal stability findings suggest that global human mobility demonstrates resilience against short-term changes.

We also assess the spatial stability of the intercounty connectivity network by detecting spatial communities based on mobility patterns. Our results indicate that mobility-driven clusters align closely with state boundaries, reflecting the influence of administrative and geographical factors on human movement, accordingly with [8, 36]. These clusters exhibited remarkable stability over time, reinforcing the idea that spatial mobility patterns are deeply ingrained and relatively resistant to abrupt changes. The fact that mobility patterns are highly correlated with state boundaries suggests that state-level structures could be effective for designing target

public health interventions based on travel reductions. Our findings underscore the importance of considering mobility patterns when designing interventions, resource allocation, and disease control strategies.

As shown in the context of COVID-19 pandemic in France [29], we also demonstrated that incorporating high-resolution human mobility data is crucial for accurately capturing the spatial spread of infectious diseases. Our findings indicate that county-level, daily mobility data offer the most accurate representation of the spatial spread of disease in the US. Notably, static county-level mobility data achieves similar model performance to real-time data, suggesting that an undisturbed representation of reality is adequate for reproducing spatial spread. More interestingly, our exploration of various spatial scales for metapopulation models underscores the significance of aligning the model's structure with the inherent spatial scale of human mobility. While county-level mobility data yields the most accurate depiction, mobility data-based clusters, US states and HHS regions don't capture the heterogeneity of the COVID-19 geographical diffusion.

While our study provides valuable insights, it is not without limitations. Our work focused on the early phase of the pandemic, during which response measures (e.g., social distancing, closures, (lack of) masking) were largely homogeneous in the US, and pharmaceutical measures (e.g., vaccination, antivrials) were not available; thus these findings are not generalizable to later stages. Furthermore, we assume homogeneity within US counties. Additionally, Safegraph mobility data, like all mobile phone datasets, exhibits sampling biases. Ongoing efforts to comprehend these biases are crucial for developing better correction methods [**coston“˙leveraging“˙2021**]. An independent analysis by Safegraph revealed the underrepresentation of older and non-white individuals in POI-specific analyses, though the panel is representative of race, education, and income [**coston“˙leveraging“˙2021**, 42].

While characterizing the key role of mobility in the spatial invasion of the COVID-19 pandemic in the US, our study sheds light on the global stability of human mobility patterns, and the relevant information needed to design a reliable predictive model. This result may be specific to countries such as the US in which mobility restrictions were not stringent, specified for intercounty mobility, nor enforced. Metapopulation models that incorporate accurate mobility data can provide valuable insights into disease dynamics and enhance our ability to predict and control the spread of future infectious disease outbreaks. Moreover, standardized data extraction and sharing we introduced might help facilitate the timelines associated with legal agreements for data sharing, which do not always align with the rapid spread of epidemics, thus diminishing the feasibility of timely responses to such outbreaks.

Acknowledgments

The authors thank the teams at Safegraph/Advan Patterns for sharing mobility data.

References

[1] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. en. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2015–2020, Feb. 2006. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0510525103. URL: http://www.pnas.org/content/103/7/2015 (visited on 01/11/2018).

[2] C. Viboud, O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell. Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza. *Science*, 312(5772):447–451, Apr. 2006. DOI: 10.1126/science.1125237. URL: https://www.science.org/doi/

full/10.1126/science.1125237 (visited on 11/02/2023). Publisher: American Association for the Advancement of Science.

[3] V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. eng. *PLoS medicine*, 4(1):e13, Jan. 2007. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.0040013.

[4] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani. Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study. *BMC Medicine*, 5(1):34, Nov. 2007. ISSN: 1741-7015. DOI: 10.1186/1741-7015-5-34. URL: https://doi.org/10.1186/1741-7015-5-34 (visited on 04/13/2021).

[5] V. Colizza, R. Pastor-Satorras, and A. Vespignani. Reaction–diffusion processes and metapopulation models in heterogeneous networks. en. *Nature Physics*, 3(4):276–282, Apr. 2007. ISSN: 1745-2481. DOI: 10.1038/nphys560. URL: https://www.nature.com/articles/nphys560 (visited on 04/13/2021). Number: 4 Publisher: Nature Publishing Group.

[6] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. en. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, Dec. 2009. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0906910106. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0906910106 (visited on 01/16/2020).

[7] Y.-A. Montjoye, C. Hidalgo, M. Verleysen, and V. Blondel. Unique in the Crowd: The Privacy Bounds of Human Mobility. *Scientific reports*, 3:1376, Mar. 2013. DOI: 10.1038/srep01376.

[8] S. Sobolevsky, M. Szell, R. Campari, T. Couronné, Z. Smoreda, and C. Ratti. Delineating Geographical Regions with Networks of Human Interactions in an Extensive Set of Countries. en. *PLOS ONE*, 8(12):e81707, 2013. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0081707. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0081707 (visited on 12/15/2023). Publisher: Public Library of Science.

[9] M. Tizzoni, P. Bajardi, A. Decuyper, G. K. K. King, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. González, and V. Colizza. On the Use of Human Mobility Proxies for Modeling Epidemics. en. *PLOS Computational Biology*, 10(7):e1003716, July 2014. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003716. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003716 (visited on 01/15/2020).

[10] V. D. Blondel, A. Decuyper, and G. Krings. A survey of results on mobile phone datasets analysis. En. *EPJ Data Science*, 4(1):10, Dec. 2015. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-015-0046-0. URL: https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-015-0046-0 (visited on 01/11/2018).

[11] H. Barbosa-Filho, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini. Human Mobility: Models and Applications. en. *Physics Reports*, 734:1–74, Mar. 2018. ISSN: 03701573. DOI: 10.1016/j.physrep.2018.01.001. URL: http://arxiv.org/abs/1710.00004 (visited on 01/16/2020). arXiv: 1710.00004.

[12] Y.-A. de Montjoye, S. Gambs, V. Blondel, G. Canright, N. de Cordes, S. Deletaille, K. Engø-Monsen, M. Garcia-Herranz, J. Kendall, C. Kerry, G. Krings, E. Letouzé, M. Luengo-Oroz, N. Oliver, L. Rocher, A. Rutherford, Z. Smoreda, J. Steele, E. Wetter, A. ". Pentland, and L. Bengtsson. On the privacy-conscientious use of mobile phone data. en. *Scientific Data*, 5(1):180286, Dec. 2018. ISSN: 2052-4463. DOI: 10.1038/sdata.2018.286. URL: https://www.nature.com/articles/sdata2018286 (visited on 04/16/2021). Number: 1 Publisher: Nature Publishing Group.

[13] J. Zeng and H. Yu. A Distributed Infomap Algorithm for Scalable and High-Quality Community Detection. In *Proceedings of the 47th International Conference on Parallel Processing*, ICPP '18, pages 1–11, New York, NY, USA. Association for Computing

Machinery, Aug. 2018. ISBN: 978-1-4503-6510-9. DOI: 10.1145/3225058.3225137. URL: https://dl.acm.org/doi/10.1145/3225058.3225137 (visited on 08/17/2023).

[14] A. Maxmen. Can tracking people through phone-call data improve lives? eng. *Nature*, 569(7758):614–617, 2019. ISSN: 1476-4687. DOI: 10.1038/d41586-019-01679-5.

[15] H. S. Badr, H. Du, M. Marshall, E. Dong, M. M. Squire, and L. M. Gardner. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. en. *The Lancet Infectious Diseases*, 20(11):1247–1254, Nov. 2020. ISSN: 1473-3099. DOI: 10.1016/S1473-3099(20)30553-3. URL: https://www.sciencedirect.com/science/article/pii/S1473309920305533 (visited on 01/18/2023).

[16] CDC. COVID Data Tracker. en, Mar. 2020. URL: https://covid.cdc.gov/covid-data-tracker (visited on 01/25/2023).

[17] S. Changruenngam, D. J. Bicout, and C. Modchang. How the individual human mobility spatio-temporally shapes the disease transmission dynamics. en. *Scientific Reports*, 10(1):11325, July 2020. ISSN: 2045-2322. DOI: 10.1038/s41598-020-68230-9. URL: https://www.nature.com/articles/s41598-020-68230-9 (visited on 05/18/2021). Number: 1 Publisher: Nature Publishing Group.

[18] K. H. Grantz, H. R. Meredith, D. A. T. Cummings, C. J. E. Metcalf, B. T. Grenfell, J. R. Giles, S. Mehta, S. Solomon, A. Labrique, N. Kishore, C. O. Buckee, and A. Wesolowski. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. en. *Nature Communications*, 11(1):4961, Sept. 2020. ISSN: 2041-1723. DOI: 10.1038/s41467-020-18190-5. URL: https://www.nature.com/articles/s41467-020-18190-5 (visited on 04/16/2021). Number: 1 Publisher: Nature Publishing Group.

[19] N. Oliver, B. Lepri, H. Sterly, R. Lambiotte, S. Deletaille, M. D. Nadai, E. Letouzé, A. A. Salah, R. Benjamins, C. Cattuto, V. Colizza, N. d. Cordes, S. P. Fraiberger, T. Koebe, S. Lehmann, J. Murillo, A. Pentland, P. N. Pham, F. Pivetta, J. Saramäki, S. V. Scarpino, M. Tizzoni, S. Verhulst, and P. Vinck. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. en. *Science Advances*, 6(23):eabc0764, June 2020. ISSN: 2375-2548. DOI: 10.1126/sciadv.abc0764. URL: https://advances.sciencemag.org/content/6/23/eabc0764 (visited on 04/02/2021). Publisher: American Association for the Advancement of Science Section: Editorial.

[20] S. Pei, S. Kandula, and J. Shaman. Differential effects of intervention timing on COVID-19 spread in the United States. eng. *Science Advances*, 6(49):eabd6370, Dec. 2020. ISSN: 2375-2548. DOI: 10.1126/sciadv.abd6370.

[21] E. Pepe, P. Bajardi, L. Gauvin, F. Privitera, B. Lake, C. Cattuto, and M. Tizzoni. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. en. *Scientific Data*, 7(1):230, July 2020. ISSN: 2052-4463. DOI: 10.1038/s41597-020-00575-2. URL: https://www.nature.com/articles/s41597-020-00575-2 (visited on 04/02/2021). Number: 1 Publisher: Nature Publishing Group.

[22] G. Pullano, E. Valdano, N. Scarpa, S. Rubrichi, and V. Colizza. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. English. *The Lancet Digital Health*, 2(12):e638–e649, Dec. 2020. ISSN: 2589-7500. DOI: 10.1016/S2589-7500(20)30243-0. URL: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30243-0/abstract (visited on 05/06/2021). Publisher: Elsevier.

[23] J. Saha, B. Barman, and P. Chouhan. Lockdown for COVID-19 and its impact on community mobility in India: An analysis of the COVID-19 Community Mobility Reports, 2020. *Children and Youth Services Review*, 116:105160, Sept. 2020. ISSN: 0190-7409. DOI: 10.1016/j.childyouth.2020.105160. URL: https://www.sciencedirect.com/science/article/pii/S0190740920310422 (visited on 12/15/2023).

[24] F. Schlosser, B. F. Maier, O. Jack, D. Hinrichs, A. Zachariae, and D. Brockmann. COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceed-*

ings of the National Academy of Sciences, 117(52):32883–32890, Dec. 2020. DOI: 10.1073/pnas.2012326117. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2012326117 (visited on 12/15/2023). Publisher: Proceedings of the National Academy of Sciences.

[25] H. S. Badr and L. M. Gardner. Limitations of using mobile phone data to model COVID-19 transmission in the USA. English. *The Lancet Infectious Diseases*, 21(5):e113, May 2021. ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(20)30861-6. URL: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30861-6/fulltext (visited on 09/25/2023). Publisher: Elsevier.

[26] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec. Mobility network models of COVID-19 explain inequities and inform reopening. en. *Nature*, 589(7840):82–87, Jan. 2021. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2923-3. URL: https://www.nature.com/articles/s41586-020-2923-3 (visited on 01/17/2023). Number: 7840 Publisher: Nature Publishing Group.

[27] R. Garnier, J. R. Benetka, J. Kraemer, and S. Bansal. Socioeconomic Disparities in Social Distancing During the COVID-19 Pandemic in the United States: Observational Study. EN. *Journal of Medical Internet Research*, 23(1):e24591, Jan. 2021. DOI: 10.2196/24591. URL: https://www.jmir.org/2021/1/e24591 (visited on 09/01/2023). Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

[28] G. Pullano. *Human mobility and epidemics*. en. PhD thesis, Sorbonne Université, Sept. 2021. URL: https://theses.hal.science/tel-04099742 (visited on 09/18/2023).

[29] G. Pullano, L. Di Domenico, C. E. Sabbatini, E. Valdano, C. Turbelin, M. Debin, C. Guerrisi, C. Kengne-Kuetche, C. Souty, T. Hanslik, T. Blanchon, P.-Y. Boëlle, J. Figoni, S. Vaux, C. Campèse, S. Bernard-Stoecklin, and V. Colizza. Underdetection of cases of COVID-19 in France threatens epidemic control. en. *Nature*, 590(7844):134–139, Feb. 2021. ISSN: 1476-4687. DOI: 10.1038/s41586-020-03095-6. URL: https://www.nature.com/articles/s41586-020-03095-6 (visited on 06/30/2023). Number: 7844 Publisher: Nature Publishing Group.

[30] R. E. Baker, A. S. Mahmud, I. F. Miller, M. Rajeev, F. Rasambainarivo, B. L. Rice, S. Takahashi, A. J. Tatem, C. E. Wagner, L.-F. Wang, A. Wesolowski, and C. J. E. Metcalf. Infectious disease in an era of global change. *Nature Reviews. Microbiology*, 20(4):193–205, 2022. ISSN: 1740-1526. DOI: 10.1038/s41579-021-00639-z. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8513385/ (visited on 09/01/2023).

[31] Data Access - Urban Rural Classification Scheme for Counties. en-us, Oct. 2022. URL: https://www.cdc.gov/nchs/data_access/urban_rural.htm (visited on 01/25/2023).

[32] K. Sparks, J. Moehl, E. Weber, C. Brelsford, and A. Rose. Shifting temporal dynamics of human mobility in the United States. *Journal of Transport Geography*, 99:103295, Feb. 2022. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2022.103295. URL: https://www.sciencedirect.com/science/article/pii/S0966692322000187 (visited on 09/01/2023).

[33] E. Valdano, J. T. Okano, V. Colizza, H. K. Mitonga, and S. Blower. Use of mobile phone data in HIV epidemic control. English. *The Lancet HIV*, 9(12):e820–e821, Dec. 2022. ISSN: 2352-3018. DOI: 10.1016/S2352-3018(22)00332-0. URL: https://www.thelancet.com/journals/lanhiv/article/PIIS2352-3018(22)00332-0/fulltext (visited on 12/15/2023). Publisher: Elsevier.

[34] M. Zhang, S. Wang, T. Hu, X. Fu, X. Wang, Y. Hu, B. Halloran, Z. Li, Y. Cui, H. Liu, Z. Liu, and S. Bao. Human mobility and COVID-19 transmission: a systematic review and future directions. *Annals of GIS*, 28(4):501–514, Oct. 2022. ISSN: 1947-5683. DOI: 10.1080/19475683.2022.2041725. URL: https://doi.org/10.1080/19475683.2022.2041725 (visited on 09/01/2023). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/19475683.2022.2041

[35] R. Arambepola, K. L. Schaber, C. Schluth, A. T. Huang, A. B. Labrique, S. H. Mehta, S. S. Solomon, D. A. T. Cummings, and A. Wesolowski. Fine scale human mobility changes within 26 US cities in 2020 in response to the COVID-19 pandemic were associated with distance and income. en. *PLOS Global Public Health*, 3(7):e0002151, 2023. ISSN: 2767-3375. DOI: 10.1371/journal.pgph.0002151. URL: https://journals.plos.org/globalpublichealth/article?id=10.1371/journal.pgph.0002151 (visited on 09/01/2023). Publisher: Public Library of Science.

[36] B. Klein, H. Hartle, M. Shrestha, A. C. Zenteno, D. B. S. Cordera, J. R. Nicolas-Carlock, A. I. Bento, B. M. Althouse, B. Gutierrez, M. Escalera-Zamudio, A. Reyes-Sandoval, O. G. Pybus, A. Vespignani, J. A. Diaz-Quiñonez, S. V. Scarpino, and M. U. G. Kraemer. Spatial scales of COVID-19 transmission in Mexico, Jan. 2023. DOI: 10.48550/arXiv.2301.13256. URL: http://arxiv.org/abs/2301.13256 (visited on 12/15/2023). arXiv:2301.13256 [physics, stat].

[37] L. Pappalardo, E. Manley, V. Sekara, and L. Alessandretti. Future directions in human mobility science. en. *Nature Computational Science*, 3(7):588–600, July 2023. ISSN: 2662-8457. DOI: 10.1038/s43588-023-00469-4. URL: https://www.nature.com/articles/s43588-023-00469-4 (visited on 09/01/2023). Number: 7 Publisher: Nature Publishing Group.

[38] Z. Susswein, E. C. Rest, and S. Bansal. Disentangling the rhythms of human activity in the built environment for airborne transmission risk: An analysis of large-scale mobility data. *eLife*, 12:e80466, Apr. 2023. N. Hens, D. M. Harper, and G. Béraud, editors. ISSN: 2050-084X. DOI: 10.7554/eLife.80466. URL: https://doi.org/10.7554/eLife.80466 (visited on 09/01/2023). Publisher: eLife Sciences Publications, Ltd.

[39] Eurostat. URL: https://ec.europa.eu/eurostat/ (visited on 01/16/2020).

[40] Facebook Data for Good. en-US. URL: https://dataforgood.fb.com/ (visited on 05/14/2021).

[41] A. Findlater. Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel. en:12.

[42] R. Fox. What about bias in your dataset?": Quantifying sampling bias in SafeGraph Patterns. URL: https://colab.research.google.com/drive/16eBT1dqnUK76GnBMgPMY5rJO4vBD2gdQ?usp=sharing.

[43] Google.com. COVID-19 Community Mobility Report. COVID-19 Community Mobility Report https://www.google.com/covid19/mobility?hl=fr. URL: https://www.google.com/covid19/mobility/ (visited on 02/13/2021).

[44] G. Pullano. *Epidemics and Human Mobility, PhD Thesis.*

[45] Safegraph. Link: https://www.safegraph.com. URL: https://www.safegraph.com/lp/get-poi-data?utm_campaign=gg_nam_dg_search_brand_exact_acquisition&utm_source=google&utm_medium=paid_search&utm_content=safegraph_exact&utm_term=safegraph&ad_headline=&ad_name=&gad=1&gclid=CjwKCAjwivemBhBhEiwAJxNWN7291rtb_DgSon1DjhlDNkwxgKDrAgMWFNOsMRXataKQ75GxAf1reRoC8dEQAvD_BwE (visited on 08/17/2023).