Check for updates

What human mobility data tell us about COVID-19 spread

Laura Alessandretti ⬤

Peoples' movements — both local and long-distance — have driven the spread of COVID-19, within and between communities. At the same time, although most contagion events involve human travel, not all human travel leads to contagion events, and deriving information about virus spread from what is known about human mobility remains a challenge. In the past two years, new datasets and analyses have shed fresh light on the problem.

Human mobility is critical to the spread of airborne diseases: transmission occurs mainly through close-contact meetings and, to meet one another, people travel. Over the two years of the COVID-19 pandemic, unprecedented efforts have been made to understand the interplay between mobility and the spread of epidemics. These efforts have made it possible to assess the effectiveness and socio-economic impact of non-pharmaceutical interventions (such as national lockdowns) on different groups[1–3]; to develop models to predict disease spatial diffusion[4,5] and to assess the outcomes in different hypothetical future scenarios[6]. At the same time, these works have revealed that the causal mechanisms leading from human travel to close-contact meetings, and thus to disease transmission, are still not fully understood.

Research into disease transmission and mobility falls into two broadly distinct topics. The first deals with how long-distance travel drives the spatial diffusion of epidemics across areas (such as cities or regions) and the second with how local travel (within cities and neighbourhoods) drives local transmission. These two aspects are intertwined, such that they are often considered and modelled simultaneously[1,3]. With respect to long-distance travel, research developed throughout the COVID-19 pandemic has corroborated the understanding that travel across areas drives the importation risk in the early phases of an outbreak and the diffusion of emerging virus variants across communities (BOX 1).

Traditionally, close-contact meetings would be estimated using data sources such as censuses and surveys, but since the widespread diffusion of mobile phones, mobility data describing short-range movements has provided a new approach[7]. Data can be made available at different levels of aggregation. Most often, data consist of the number of trips (or unique travellers) within small spatial units, such as cities, zip codes or mobile-phone antenna ranges, aggregated by sufficiently small units of time, such as hours or days. Such data are typically provided by mobile phone operators (and now also by providers of location services like Google and Apple). The datasets can cover large population fractions; they preserve privacy due to their aggregated nature; and can be collected and analysed in near real-time. Data describing local movements are used as an input for epidemic models, often assuming that a simple functional form captures the relation between the volume of travel and the number of close-contact meetings at a location.

Evidence gathered during the COVID-19 pandemic has revealed that this assumption does not hold consistently. Data collected via contact surveys for ~3,300 individuals in four Chinese cities showed that intra-city travel volumes only correlate with the number of close contacts in some phases of an epidemic wave[8]. In particular, whereas imposing a lockdown led to a striking reduction in the number of contacts and corresponded to a decrease in mobility, the post-lockdown increase in movements was not associated

with an increase in contacts. In line with these findings, a study based on data from 52 countries showed that the relation between within-country mobility and the reproduction number R, which measures the number of secondary cases caused by an infected individual, has changed over time[9]. The introduction of non-pharmaceutical interventions generally led to a marked decrease in both travel and transmissibility (measured by R). However, after measures were relaxed, the relation between travel and transmissibility decoupled in most countries, likely due to people adopting other social distancing behaviours. Analyses that focus on the predictive power of mobility data for epidemic forecasting confirm these findings. Information on local travel volumes in the United States could consistently be used to anticipate downward trends in COVID-19 incidence, but not upward trends[5].

From the policy perspective, these findings suggest that epidemic control can be achieved even when restrictions on travel are partly lifted, provided other social distancing behaviours are maintained and alternative strategies such as complete contact-tracing are implemented. From the modelling perspective, the results suggest that epidemic modellers should be cautious in assuming specific functional relations between aggregated travel volumes and contacts, especially when the model aims at capturing widely different situations.

During the COVID-19 pandemic, access to more detailed data from GPS trajectories and location service providers has opened up a new research avenue for estimating close-contact meetings with an unprecedented

---

Key advances

- Research corroborated the understanding that long-distance travel regulates importation risk in the early phases of an outbreak and the diffusion of emerging variants.

- The relation between short-distance travel and transmission changes across epidemic phases. The reduction of local travel explains downward trends in epidemic incidence, but increases in travel volumes do not always lead to increased epidemic incidence.

- New approaches use GPS data to estimate realistic large contact networks and develop scenario projections to understand the drivers of disease spread.

Box 1 | **Long-distance travel: coupling across geographical areas**

Long-distance travel is well captured by data that describe the volume of travel between pairs of geographical areas over time, collected from sources such as mobile-phone or air travel data. The volume of travel between areas regulates importation risk in the early phases of an outbreak[3,10]. For example, it was found that, at the start of 2020, the number of COVID-19 cases in different Chinese provinces correlated with the volume of travel from Wuhan before the introduction of travel restrictions[3,10], and international travel restrictions effectively delayed global spread[3]. In addition, the resurgence of the virus in Europe in late summer 2020 was well predicted by data describing cross-country mobility[4]. When the outbreak is widespread, however, long-distance travel and transmission can decouple[3], due to the limited contribution of imported cases to the initiation of local chains. Given the critical role played by mobility across areas, especially in the early phases of an outbreak, epidemic models that capture the coupling of geographical units induced by mobility were shown to outperform models agnostic of travel[1] and capture well the spread of variants[4]. By revealing the critical role played by long-distance travel when the epidemic activity is limited, these works suggested that long-distance travel restrictions are crucial when the disease is confined in specific areas, whereas interventions targeting local travel are more effective when the disease is widespread.

level of detail[2,6]. One such study used data describing the flows of visitors between ~600,000 locations in ten US cities to estimate the probability for two individuals to meet in a given hour and a given location — such as a restaurant, shop or supermarket — based on their block-group of residence. The study used mobility data aggregated by block-group of residence and considered only block-groups with at least five individuals, thus ensuring k-anonymity with $k = 5$. Through a metapopulation epidemic model unfolding on top of the estimated contact network, the study showed that a small fraction of locations could account for the large majority of transmissions, and that reopening categories of places such as restaurants, gyms, hotels, cafes and places of worship could produce the largest increases in infections compared to other types of places.

Using a similar approach, a study analysed pseudonymized GPS trajectories of individuals based in Boston who opted-in to provide access to their location data through a GDPR-compliant framework. Focusing on ~2% of the Boston population, the study estimated the probability for any two individuals to be in close physical proximity on any given day, by measuring the fraction of time spent in the same location[6]. The authors developed a realistic epidemic model, in which the disease spreads on top of the empirical weighted contact network, with nodes representing individuals and links corresponding to close contact probabilities. Unlike the vast majority of epidemic models, which rely on simplified contact patterns, this approach accounts for characteristic aspects of human social behaviour, including wide heterogeneities in the number of contacts across people, and the existence of communities of highly connected individuals. The study used the model to explore different strategies for lifting social distancing interventions following the first epidemic wave, and revealed that the gradual reopening of activities in Boston would make it possible to maintain a low transmission incidence, provided extensive contact tracing and household quarantine were implemented.

Overall, these studies demonstrate the feasibility of using realistic contact networks estimated from GPS data to develop scenario projections for understanding drivers of disease spread. In contrast to other approaches, high-resolution mobility data makes it possible to explore how contacts occurring in different types of locations (such as restaurants, gyms or hotels) contribute to disease spreading. They can further enable us to assess the role played by super-spreading individuals, who infect a high number of other people, and study how epidemics spread within and across social communities.

However, the effects of the limitations of these new methods are not entirely understood. For example, they are based on population samples that are much smaller and that are biased compared to other approaches, and rely on several assumptions and choices underlying the estimation of contacts. Further, they can rely on personal data. More research is needed to validate the inferred contact networks against estimates made from other data sources[7,8], to determine the ability of realistic models to generalize across time and geographical areas, and to quantify the model uncertainties. Understanding these aspects will be crucial to assess the cost–benefit ratio of using high-resolution individual data for different purposes, as compared to aggregated data.

The work developed in the past two years has been crucial to monitor, forecast and design interventions to control the spread of disease. The challenges faced during the pandemic have also posed several questions that could drive future fundamental research. For example, what aspects mediate the relationship between local travel, close contacts, and transmissibility? What is the cost–benefit ratio of high-resolution mobility data for epidemic modelling?

Laura Alessandretti [ID]

Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark.

e-mail: lauale@dtu.dk

1. Gozzi, N. et al. Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nat. Commun.* **12**, 2429 (2021).
2. Chang, S. et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).
3. Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
4. Lemey, P. et al. Untangling introductions and persistence in COVID-19 resurgence in Europe. *Nature* **595**, 713–717 (2021).
5. Kogan, N. E. et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. *Sci. Adv.* **7**, eabd6989 (2021).
6. Aleta, A. et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **4**, 964–971 (2020).
7. Mistry, D. et al. Inferring high-resolution human mixing patterns for disease modeling. *Nat. Commun.* **12**, 323 (2021).
8. Zhang, J. et al. The impact of relaxing interventions on human contact patterns and SARS-CoV-2 transmission in China. *Sci. Adv.* **7**, eabe2584 (2021).
9. Nouvellet, P. et al. Reduction in mobility and COVID-19 transmission. *Nat. Commun.* **12**, 1090 (2021).
10. Kraemer, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).

Competing interests
The author declares no competing interests.