Towards a General Framework for Co-Location

Keiran Suchak
School of Geography, University of Leeds, UK

Ed Manley
School of Geography, University of Leeds, UK

Abstract

Previous studies into co-location exist in a variety of fields such as epidemiology and human mobility. In each field, researchers are interested identifying points of co-location amongst members of a population. In each of these fields, however, the definition of what it means for members of the population to be co-located may differ; furthermore, the ways in which data are collected vary. This piece of work aims to provide an initial outline of a general framework for identifying points of co-location. It demonstrates that the identification of co-location points between individuals is sensitive to the way in which co-location is defined in each context, as well as the types of data used. Furthermore, it highlights the impact that uncertainty in observations can have on our ability to reliably identify co-location.

2012 ACM Subject Classification Computing methodologies → Modeling methodologies

Keywords and phrases human mobility, co-location, contact tracing

Digital Object Identifier 10.4230/LIPIcs.COSIT.2024.24

Category Short Paper

Funding This work was undertaken as part of the i-sense project (EP/R00529X/1) and the EPSRC Digital Hub for AMR (EP/X031276/1).

1 Introduction

The study of any system involving movement is strongly influenced by how we observe and capture instances of mobility. We can consider these systems from a Lagrangian or Eulerian perspective, i.e. we can consider a system by either focusing on individuals or places [5]. This distinction arises in studies of fluid phenomena, in which an Eulerian perspective considers a fixed frame of reference and a Lagrangian perspective considers a frame of reference which moves with the fluid [13, 16]. A Eulerian perspective in the study of mobility focuses on the observation and modelling of fixed places whilst a Lagrangian perspective focuses on the observation and modelling of individuals as they move between locations. In each case, one of the key considerations is the way in which individuals come together – how they co-locate. A Lagrangian perspective may provide us with the ability to study the movements of individuals, identifying not only where they co-located but also where they were before and after. Whereas the Eulerian perspective, on the other hand, may provide us with the ability to study the places where individuals co-locate; this may offer richer contextual information focused on the characteristics of places that draw individuals to them.

We can define co-location of two individuals as them spending some degree of time in the *same space as each other*. We can additionally specify that the individuals share the same space *at the same time*. We make this distinction to highlight that there may be contexts in which we may wish to identify whether individuals are in the same place at the same time (e.g. identifying potential instances of virus spread), whilst in others we may only wish to establish that the individuals visit the same place (e.g. establishing whether the individuals are part of the same community).

16th International Conference on Spatial Information Theory (COSIT 2024).
Editors: Benjamin Adams, Amy Griffin, Simon Scheider, and Grant McKenzie; Article No. 24; pp. 24:1–24:10
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

As such, when undertaking studies into co-location, and devising a framework for its identification, we should consider the following three, interconnected factors:

1. **What constitutes co-location in our context?** – Here we consider how we define co-location in a given context. This may place constraints on whether or not individuals must be co-located spatially, or also spatiotemporally. Additionally, it may place constraints on the scale or proximity between individuals to constitute co-location.
2. **How are our location data generated?** – By which we consider the way in which the observations were generated. Variations in the data generation process can have significant impact on what we are able to do with the data and the contexts in which they can be useful. For example, mobility data may be generated from mobile phone through both is GPS traces [2] or geotagged social media check-ins. Each of these types of data lend themselves to different perspectives. GPS trace data allow us to frame co-location through the lens of the individuals' movements, whereas geotagged social media check-ins allow us to frame co-location through the lens of the places that people visit; [5] describe this distinction as the difference between Lagrangian and Eulerian conceptualisations.
3. **What mathematical approach are we using to identify co-location?** – Where we consider the specific method by which co-location between individuals is identified. This can be largely influenced by both the way in which we define co-location (Point 1) and the types of data which we have regarding mobility (Point 2).

Arising from these definitions are several constraints that require specification during the course of estimating co-location. These are referred to henceforth as ΔX and ΔT, spatial and temporal tolerances respectively, which define our "accepted" measure of co-location (elaborated in Section 2). The classification of co-location can subsequently be defined based on binary or probabilistic functions (see Section 2). The definitions made here are determined by our need for granularity in identifying co-location, as well as restrictions of our data collection protocol or method (in turn impacted by technical and ethical issues). Related to each observation is uncertainty, σ^2, associated with each spatial and temporal observation, which impacts the accuracy of the detection of co-location. Uncertainty is inherent to the estimation of co-location, and strongly influences our detection of co-location, and thus specification of ΔX and ΔT (which we explore in Section 3).

This paper aims to present the problem of identifying points of co-location in a general form such that elements of the approach can be interchanged to fit different contexts. By specifying the parameters required to estimate co-location, we aim to provide clarity on what constitutes – and what data granularity is necessary to identify – co-location in different contexts. We use simulations to demonstrate the implications of different choices of tolerance, and the implications of locational uncertainty.

Previous investigations focusing on identifying points of co-location have been undertaken in fields including epidemiology, ecology, and human mobility. Thus, prior to elaboration of the general framework, we briefly explore two areas of use where capturing of co-location is important but different.

1.1 Epidemiology

In an epidemiology setting, researchers are interested in co-location from the perspective of disease transmission and contact tracing. During the COVID-19 pandemic, multiple parties attempted to devise digital contact tracing schemes that would make use of Bluetooth Low Energy (BLE) and Global Positioning System (GPS) facilities on mobile phones [18]; these approaches varied from country to country [9], and faced challenges in relation to public acceptance. Other recent work has made use of routinely collected data from within hospitals (which include electronic medical records and door access logs) to explore mobility of healthcare workers and their contact with patients [21].

In this context, definitions of co-location varied considerably – the US-based Center for Disease Control (CDC) defined a COVID-19 contact as an encounter in which an individual spends more than 15 minutes within 6 feet of another person [3]; whereas this distance was taken as equivalent to 2 metres in the United Kingdom [7]. A variety of different technical solutions were proposed to identify instances of co-location, each with varied reliability of estimation [12]. Other evidence suggests risk of infection can vary between settings (e.g. indoor vs. outdoor) and based on levels of ventilation [17]. Furthermore, the risk of infection spread may remain following the departure of an infectious individual – in some cases, such airborne diseases can remain in aerosol form for hours [19]. An approach to identifying contacts should be flexible enough to handle these variations.

1.2 Human mobility

In the field of human mobility, we may be interested in identifying co-location between individuals as a means of exploring a number of different issues. The co-location of individuals with specific Points of Interest (POIs) has been used to identify likely home and work locations and inform work on the predictability of human mobility patterns [8]. Other investigations have made used of human mobility data to identify patterns of segregation between different groups – in essence, cases where there are a lack of co-locations between individuals from different groups [15]. Mobility data have also been used to identify co-locations between individuals as means of detecting in-person social networks within populations [11], and exploring the influence of place on how socialising occurs between individuals [4].

Studies in to human mobility make use of a wide range of data-sources which include location-based check-ins from social media, GPS and Call Detail Record (CDR) data from mobile phones and transaction data from public transport networks [20]. In this context, there can also be large variations in how co-location is defined. In scenarios where co-location is used to identify cases where individuals have socialised directly with each other, spatial and temporal constraints can be stricter, requiring that individuals to be within a couple of meters of each other. Alternatively, studies which focus on the segregation may have looser temporal constraints [14], and may not require precise spatiotemporal co-location but only that individuals visit the same location.

2 Defining co-location

Given the variations in contextual constraints and types of data outlined in Sections 1.1 and 1.2, we seek to propose a generalised approach to identifying points of co-location between two individuals. This requires that we collect time-series data consisting of locations for individuals in the population; given the ith individual in a population, we define the time-series data pertaining to the individual, d_i, as a collection of location-time records:

$$d_i = [(l_0, t_0), (l_1, t_1), \dots (l_N, t_N)], \tag{1}$$

where l_0 and t_0 are the location and time of the initial record. We may collect information regarding individuals' spatial locations in 2- or 3-dimensions as $l = (x, y)$ pairs or $l = (x, y, z)$ triples; the relationships between the locations at which these data are collected may be described based on discrete grid-based space or continuous space.

When identifying points of co-location, we check that observations of individuals are in proximity to each other in both time and space. The definition of what constitutes proximal varies based on context. As part of this process, we define spatial and temporal tolerances, ΔX and ΔT. Given such tolerances, we may make a decision regarding whether the result of the identification process is a binary output or a probabilistic output.

2.1 Spatial proximity

The binary classification of whether two individuals are proximal to each other in space may be defined based on a function such as the Heaviside step function, $\Theta(\Delta X, l_a, l_b)$:

$$\Theta(\Delta X, l_a, l_b) = \begin{cases} 1, & \text{if } dist(l_a, l_b) \leq \Delta X, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Such an approach would indicate that individuals a and b are in proximity to each other if the distance between them is within the tolerance ΔX. Alternatively, we may wish to output a real value on the interval $[0, 1]$ as part of constructing some probabilistic score of how likely individuals are to be co-located. This can be achieved using a kernel function, $K(\Delta X, l_a, l_b)$; a simple example might be to use a triangular kernel:

$$K(\Delta X, l_a, l_b) = \begin{cases} 1 - \frac{dist(l_a, l_b)}{\Delta X} & \text{if } dist(l_a, l_b) \leq \Delta X, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

This approach would provide a score of proximity on the interval $[0, 1]$ within the spatial tolerance which varies linearly with distance. Other kernels such as Gaussian can also be used to weight scores more favourably in cases where individuals are closer together.

2.2 Temporal proximity

These scoring approaches are applied not only to the spatial proximity of observations of individuals, but also to temporal proximity. We may, therefore, calculate a binary indicator of whether observations of two individuals are close to each other in time using the same approach, calculating the Heaviside step function $\Theta(\Delta T, t_a, t_b)$.

In both the spatial and temporal cases, indications of proximity are dependent on measures of distance. The way in which distances are calculated will in turn be dependent on the way in which space and time are represented. Given a continuous description of space, we may choose to calculate Euclidean distances between locations; we may, however, choose to use other representations of space such as discrete grids (which would require that we use something such as Manhattan distance) or a network representation (which would require that we use some sort of network distance).

2.3 Spatiotemporal proximity

Based on the general approach outlined above, we can score observations of two individuals on whether they are close to each other in space and time, or how close they are. Given a spatial score and a temporal score, we would then wish to create a composite score to indicate spatio-temporal proximity. This is often achieved by taking the product of the two scores. For example, given an observation of individual a and individual b – (l_a, t_a) and (l_b, t_b) respectively – we could calculate a score of whether or not the two individuals are co-located in space and time as:

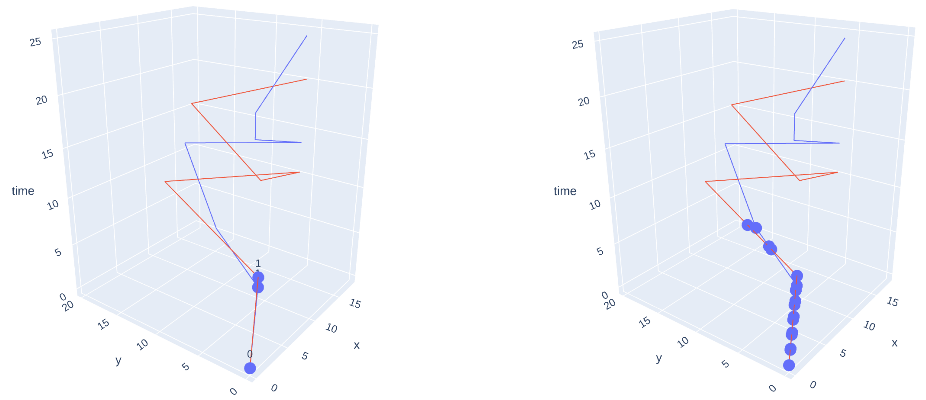
$$\Theta(\Delta X, l_a, l_b) \times \Theta(\Delta T, t_a, t_b),$$

which would provide a binary indicator. The same approach could be applied with kernel-based scores for time and space to return a continuous indicator.

Given the approach outlined thus far, we can see that the number of instances of co-location identified would be strongly dependent on the spatial and temporal tolerances, ΔX and ΔT, which are in turn dependent on the context of study and the stipulations on what constitutes co-location for that context. In cases where tolerances are high in comparison to the frequency at which the data are sampled, we expect to identify a larger number of instances of co-location; conversely, when tolerances are small in comparison to the frequency at which data are sampled, we expect to identify a smaller number of co-locations. Again, the term frequency here is applied in both the spatial and temporal sense. This issue is highlighted in Figure 1 which shows the trajectories of two individuals in blue and red, as well as identified points of co-location. In both subfigures, the spatial and temporal tolerances are set to:

$$\Delta X = 1, \tag{4a}$$

$$\Delta T = 1. \tag{4b}$$



(a) Identified points of co-location from asynchronously collected data at discrete locations.

(b) Identified points of co-location from trace-like data collected with constant temporal frequency.

Figure 1 Identifying points of co-location.

In Figure 1a, observations are only sampled at discrete locations and are generated asynchronously; this is akin to the location check-in data mentioned in Section 1. Such observations lend themselves to analysis from an Eulerian perspective. In Figure 1b, observations are sampled with a constant temporal frequency, generating (x, y, t) co-ordinates for each observation. These may be considered more similar to GPS traces, and as such lend themselves to analysis from a Lagrangian perspective. Despite the two figures showing the same patterns of mobility, the former indicates far fewer instances of co-location that the latter. Specifically, in Figure 1a, we find that the two individuals are co-located on two occasions: at the outset of the data collection at location 0, and later at location 1. In Figure 1b, however, we find that they are also co-located along their journey between the two locations, as well as being co-located for a brief period after leaving location 1. The difference in the number of identified instances of co-location is not indicative of either approach being better than the other, but instead highlights the difference between the Eulerian and Lagrangian perspectives. In considering the system from an Eulerian perspective, we focus our analysis on the places that are visited, allowing us to explore the characteristics of the places and how these lead different types of individuals to congregate in these places. In considering the system from a Lagrangian perspective, we instead focus our analysis on the individuals in the population which allows us to explore the way in which they interact while in motion.

3 Co-location under uncertainty

In many scenarios, observations of mobility will be impacted by some degree of uncertainty; all observations have an associated degree of uncertainty, and mobility data may have additional degrees of noise added in order to preserve privacy [6]. The presence of noise in data can impact the extent to which it can reliably be used to identify instances of co-location.

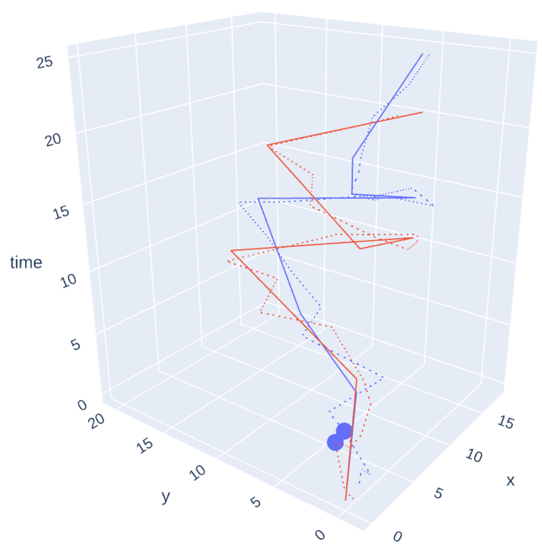


Figure 2 Identified points of co-location from trace-like data collected with constant temporal frequency with the addition of normally distributed noise.

In Figure 2, we see the same initial trace-like data as in Figure 1b, but with the addition of normally distributed noise at each sample point (resulting in the dashed lines). The same co-location identification process has been applied to these trajectories. This results in fewer instances of co-location being identified in comparison to the original data as a consequence of the additional noise. This can distort not only how often the two individuals are co-located, but also when and where these co-locations take place.

In a context such as epidemiology, such errors and distortions can result in either potential points of infection spread being missed (false negatives) or incorrect indications of potential infection spread resulting in additional testing (false positives). It is, therefore, critical that the approach be robust to noise. Given that many contexts place specific constraints on what constitutes co-location, this may inform the degree of uncertainty that is acceptable in data that are used to identify points of co-location.

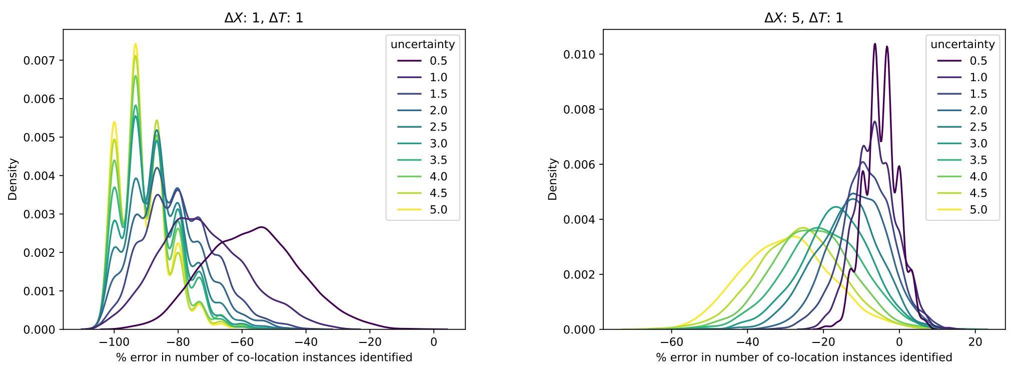
3.1 Evaluating the impact of varying degrees of observation uncertainty

In this section, we explore the relationship between the process of identifying instances of co-location, the uncertainty attached to observations, and the spatial constraints imposed on co-location. To do this, we consider a scenario in which the solid line trajectories in Figure 2 constitute the ground truth regarding the mobility of two individuals, and as such instances of co-location that occur between the two trajectories are considered to have happened.

Observations of these trajectories are then generated, just as previously, by adding normally distributed noise each of the states making up the ground truth trajectories. This noise is generated based on varying values of variance, $0.5 \leq \sigma^2 \leq 5.0$ (increasing in 0.5 steps). For each of these values of σ^2, we generate noise and attach it to the states making up the trajectories to generate observations. Observations are generated with a temporal frequency of 1. We then consider two spatial constraints for co-location: in the first scenario, we set $\Delta X = 1$; in the second scenario, we set $\Delta X = 5$; in each case, we have a fixed temporal constraint of $\Delta T = 1$. In each case, we try to identify instances of co-location between the two trajectories in the noisy observations, counting the number of instances found and comparing against the "true" number found in the ground truth data. This allows us to assess how many of the instances of co-location the process is able to identify when handling noising observations. This process is repeated $1,000$ times with each combination of σ^2 and ΔX to allow us to explore the level of variability in the results.

3.2 Results

Having run the co-location process $1,000$ times for each combination of σ^2 and ΔX, we calculate the percentage error in the number of instances of co-location identified in each case, plotting the distribution of these errors in Figure 3. Figure 3a shows the results for $\Delta X = 1$ and Figure 3b shows the results for $\Delta X = 5$. For each of the values of ΔX, we see that as the degree of uncertainty (i.e. the value of σ^2) increases, the degree of error increases.



(a) $\Delta X = 1, \Delta T = 1$. **(b)** $\Delta X = 5, \Delta T = 1$.

Figure 3 Percentage error in number of co-location events identified under varying degrees of noise attached to observations of traces.

In Figure 3a, we see that with an uncertainty of $\sigma^2 = 0.5$ the process is often returns an error of at least -50%, indicating that it often only captures half of the instances of co-location that have occurred. Similarly, we ca see that with an uncertainty of $\sigma^2 = 5.0$, the process often returns an error of at least -90%, indicating that in the scenario in which the uncertainty is much larger than the spatial threshold for co-location, the process is not able to reliably identify co-location. In Figure 3b, we are able to see the results of cases in which the uncertainty attached to observations is much smaller than the spatial threshold for co-location, e.g. $\sigma^2 = 0.5$. In this case, we find that the process reliably identifies the majority of the co-location instances. If we wish to identify at least 80% of the instances of co-location under this spatial constraint, we would consider using observations with at most $\sigma^2 = 2.5 - 3.0$, i.e. $\sigma^2 \approx \frac{1}{2}\Delta X$.

If we consider this in terms of the relationship between the spatial constraints imposed by the contexts outlined in Section 1 and the uncertainties found in different types of sensor observations, we can try to identify the the types of sensors that are appropriate for reliably detecting co-location for the context. In the context of indoor epidemiological contact tracing, we would define our spatial constraint as $\Delta X = 2m$, and would then seek to identify the type of indoor mobility sensors that are capable of producing observations with an appropriate level of uncertainty. In this context, we may consider using Bluetooth Low Energy (BLE) sensors ($1m - 2m$) or RFID sensors ($< 1m$) [10]. The increased accuracy of RFID-based position systems may provide appropriately accurate observations, but the limited sensing range on some sensors may mean that observations can only be generated close to the sensor; this may be appropriate for producing observations for analysis under an Eulerian perspective. The reduced accuracy of BLE sensors may result in a reduction in the reliability with which we can identify instances of co-location, but the increased range of the sensors will mean that we are able to gain better spatial coverage of a room.

4 Conclusions

In this paper, we have proposed an initial outline for a generalised approach to identifying points of co-location. This approach allows users to place context-specific constraints on what constitutes co-location, as well as allowing users to define different functions through which to indicate co-location. The aim for this approach is that it can be modular, i.e. users can swap in and out different building blocks such as different distance measures or indicator functions, without a requirement that the approach change substantially. We have further demonstrated how uncertainty – which runs hand-in-hand with observations of mobility – impacts our ability to reliably estimate co-location.

Future work on this general approach will focus on two avenues. The first of these will follow on from Figure 2 and explore the way in which measures of co-location vary when noise is present in observations. As seen in when comparing Figures 1b and 2, the addition of noise can impact the identification of instances of co-location, with the potential of false positive identifications and missing identifications. It is expected that this impact will depend on the relationship between the degree of uncertainty in the observations and the spatial and temporal tolerances that are used to define co-location for a context. This may, in turn, place constraints on the types of data that can be used to reliably identify points of co-location.

The second related avenue will focus on the relationship between these quantities – uncertainty and tolerances – and the characteristic spatial scales of different contexts [1]. This may help to standardise the relationship between uncertainty in observations and tolerances.

References

1 Laura Alessandretti, Ulf Aslak, and Sune Lehmann. The scales of human mobility. *Nature*, 587(7834):402–407, 2020.

2 Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.

3 Martin Z Bazant and John WM Bush. A guideline to limit indoor airborne transmission of covid-19. *Proceedings of the National Academy of Sciences*, 118(17):e2018995118, 2021.

4 Chloë Brown, Neal Lathia, Cecilia Mascolo, Anastasios Noulas, and Vincent Blondel. Group colocation behavior in technological social networks. *PloS one*, 9(8):e105816, 2014. `doi:10.48550/arXiv.1408.1519`.

5 Urška Demšar, Jed A Long, Fernando Benitez-Paez, Vanessa Brum Bastos, Solène Marion, Gina Martin, Sebastijan Sekulić, Kamil Smolak, Beate Zein, and Katarzyna Siła-Nowicka. Establishing the integrated science of movement: bringing together concepts and methods from animal and human movement analysis. *International Journal of Geographical Information Science*, 35(7):1273–1308, 2021. `doi:10.1080/13658816.2021.1880589`.

6 Zhigang Gao, Yucai Huang, Leilei Zheng, Huijuan Lu, Bo Wu, and Jianhui Zhang. Protecting location privacy of users based on trajectory obfuscation in mobile crowdsensing. *IEEE Transactions on Industrial Informatics*, 18(9):6290–6299, 2022. `doi:10.1109/TII.2022.3146281`.

7 Aritra Ghosh, Srijita Nundy, Sumedha Ghosh, and Tapas K Mallick. Study of covid-19 pandemic in london (uk) from urban context. *Cities*, 106:102928, 2020.

8 Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779–782, 2008.

9 Majid Hatamian, Samuel Wairimu, Nurul Momen, and Lothar Fritsch. A privacy and security analysis of early-deployed covid-19 contact tracing android apps. *Empirical software engineering*, 26:1–51, 2021.

10 SJ Hayward, Katherine van Lopik, Christopher Hinde, and Andrew A West. A survey of indoor location technologies, techniques and applications in industry. *Internet of Things*, 20:100608, 2022. `doi:10.1016/j.iot.2022.100608`.

11 Hsun-Ping Hsieh, Rui Yan, and Cheng-Te Li. Where you go reveals who you know: Analyzing social ties from millions of footprints. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1839–1842, 2015. `doi:10.1145/2806416.2806653`.

12 Douglas J Leith and Stephen Farrell. Measurement-based evaluation of google/apple exposure notification api for proximity detection in a light-rail tram. *Plos one*, 15(9):e0239943, 2020.

13 Michael S Longuet-Higgins. Eulerian and lagrangian aspects of surface waves. *Journal of Fluid Mechanics*, 173:683–707, 1986.

14 Esteban Moro, Dan Calacci, Xiaowen Dong, and Alex Pentland. Mobility patterns are associated with experienced income segregation in large us cities. *Nature communications*, 12(1):4633, 2021.

15 John RB Palmer, Thomas J Espenshade, Frederic Bartumeus, Chang Y Chung, Necati Ercan Ozgencil, and Kathleen Li. New approaches to human mobility: Using mobile phones for demographic research. *Demography*, 50(3):1105–1128, 2013.

16 Vilas J Shinde and Datta V Gaitonde. Lagrangian approach for modal analysis of fluid flows. *Journal of Fluid Mechanics*, 928:A35, 2021.

17 Chanjuan Sun and Zhiqiang Zhai. The efficacy of social distance and ventilation effectiveness in preventing covid-19 transmission. *Sustainable cities and society*, 62:102390, 2020.

18 C Troncoso, M Payer, JP Hubaux, M Salathé, JR Larus, W Lueks, T Stadler, A Pyrgelis, D Antonioli, L Barman, et al. Decentralized privacy-preserving proximity tracing. *IEEE Data Engineering Bulletin*, 43(2):36–66, 2020. URL: `http://sites.computer.org/debull/A20june/p36.pdf`.

19 Neeltje Van Doremalen, Trenton Bushmaker, Dylan H Morris, Myndi G Holbrook, Amandine Gamble, Brandi N Williamson, Azaibi Tamin, Jennifer L Harcourt, Natalie J Thornburg, Susan I Gerber, et al. Aerosol and surface stability of sars-cov-2 as compared with sars-cov-1. *New England journal of medicine*, 382(16):1564–1567, 2020.

20 Jinzhong Wang, Xiangjie Kong, Feng Xia, and Lijun Sun. Urban human mobility: Data-driven modeling and prediction. *ACM SIGKDD explorations newsletter*, 21(1):1–19, 2019. doi:10.1145/3331651.3331653.

21 Jared K Wilson-Aggarwal, Nick Gotts, Wai Keong Wong, Chris Liddington, Simon Knight, Moira J Spyer, Catherine F Houlihan, Eleni Nastouli, and Ed Manley. Investigating healthcare worker mobility and patient contacts within a uk hospital during the covid-19 pandemic. *Communications Medicine*, 2(1):165, 2022.