Managing Trajectories and Interactions During a Pandemic: A Trajectory Similarity-based Approach (Demo Paper)

Edward Buckland
edwardnbuckland@gmail.com
Victoria, Australia

Egemen Tanin
etanin@unimelb.edu.au
The University of Melbourne
Parkville, Victoria, Australia

Nicholas Geard
nicholas.geard@unimelb.edu.au
The University of Melbourne
Parkville, Victoria, Australia

Cameron Zachreson
cameron.zachreson@unimelb.edu.au
The University of Melbourne
Parkville, Victoria, Australia

Hairuo Xie
xieh@unimelb.edu.au
The University of Melbourne
Parkville, Victoria, Australia

Hanan Samet
hjs@cs.umd.edu
The University of Maryland
College Park, Maryland, United States

ABSTRACT

COVID-19 has brought about substantial social, economic and health related burdens, motivating different control measures from policy makers worldwide. Contact tracing plays a pivotal role in the COVID-19 era. However, contact tracing is by nature entirely retrospective: it can only identify contacts of known or suspected cases. Our proposed system is prospective, aiming to 'create' networks that will ultimately make contact tracing and pandemic management easier. As contact tracing seeks to reconstruct the underlying interaction network, we can improve the process by reducing the complexity of contact network structure; we introduce a method for reducing contact network complexity through strategic scheduling. The method functions through pairwise comparison of individual trajectories in a coordinate space of activities, locations, and time intervals. We demonstrate the method through a simulated scenario where individuals (students) register for activities using a mobile application in a campus. The application then applies our algorithm to provide individuals with schedules that reduce the complexity of the overall network, without compromising individual privacy.

CCS CONCEPTS

• **Applied computing → Health informatics**; • **Mathematics of computing →** *Spectra of graphs*; *Approximation algorithms.*

KEYWORDS

COVID-19, pandemic, transmission, contact tracing, interaction networks, spatial databases, graphs

ACM Reference Format:
Edward Buckland, Egemen Tanin, Nicholas Geard, Cameron Zachreson, Hairuo Xie, and Hanan Samet. 2021. Managing Trajectories and Interactions During a Pandemic: A Trajectory Similarity-based Approach (Demo Paper). In *29th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '21), November 2–5, 2021, Beijing, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3474717.3484206

1 INTRODUCTION

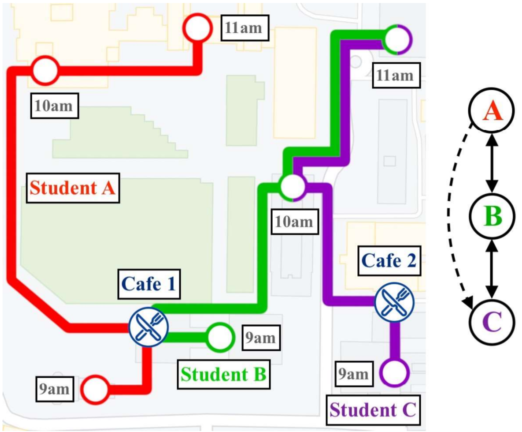
The novel coronavirus disease 2019 (COVID-19) is a respiratory illness caused by severe acute respiratory syndrome coronavirus 2. Declared a pandemic on March 11th 2020, COVID-19 continues to spread globally (see visualization in [11]), with over 206 million confirmed cases and 4.3 million deaths as of August 17th 2021[13]. Existing treatments are predominantly limited to symptom management, oxygen therapy and mechanical ventilation[2]. Though vaccines have been launched, herd immunity is still unlikely in the immediate future, cementing the need for non-pharmaceutical interventions[5]. Controlling disease spread often necessitates aggressive control measures such as mass lock-downs, travel restrictions and border closures; these interventions often carry great socioeconomic burdens and in many cases are politically not viable[10]. Thus, we need alternate solutions that are more effective and deployable to stop further outbreaks. In other words, it is important to establish a new 'COVID normal' that is conducive to both societal productivity and early-stage outbreak management, minimizing the chances of new cases developing into more widespread outbreaks.

COVID-19 has caused a surge in epidemiological research; recent advances in spatial systems pertain to better modeling, analyzing and contact tracing[1, 3, 12]. Contact tracing in particular has proven to be one of the most effective targeted management strategies during the early stages of outbreaks[4]; it facilitates the identification and isolation of high risk individuals with minimal impositions to the rest of society. However, these approaches are naturally retrospective. We adopt an approach orthogonal to such research, proactively reshaping the spatio-temporal pattern of contacts to yield more efficient contact tracing thereafter and limit the spread of disease. Our work is inspired by the fact that the efficiency of contact tracing is positively correlated with the strength of assortative community structure in the underlying interaction network[7]. From the perspective of spatial algorithms, we aim to solve a spatial query on moving object trajectories, where the trajectories of people who will participate in an activity are split into multiple clusters such that dissimilar trajectories are separated. Thus, contact tracing is easier and disease spread across distances and communities is limited. Figure 1 demonstrates. Although cafe 1 is nearest to student B, assigning in kind yields one big component where A may infect C through B (Figure 1(a)). B and C already come in contact from 10 to 11am; thus, assigning B to cafe 2 yields the

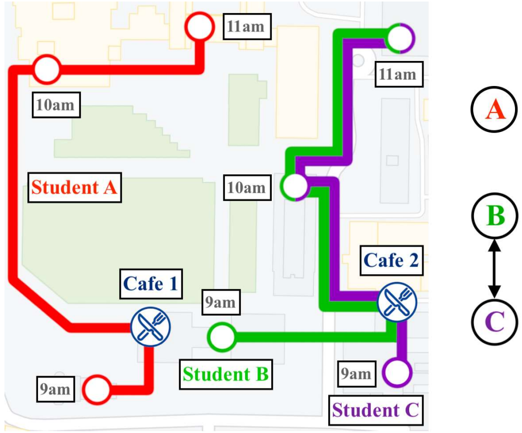
---

SIGSPATIAL '21, November 2–5, 2021, Beijing, China

ACM ISBN 978-1-4503-8664-7/21/11.
https://doi.org/10.1145/3474717.3484206



(a) Distance minimizing scheduling



(b) Trajectory similarity clustering scheduling

Figure 1: Three students, each attending three classes (open circles), register to drink coffee. The activity of 'drinking coffee' is partitioned by assigning students to cafes. Different scheduling yields different trajectories (left) and associated transmission networks (right). Using trajectory similarity clustering scheduling, the transmission network is disjoint.

best clustering, where { A } and { B, C } remain isolated (Figure 1(b)). Intuitively, our work is not only a spatial query but is biased to encourage assortative community structure. This concept of 'passive' control, managing the pandemic without blanket bans but rather better scheduling, allows for effective, targeted outbreak response.

We propose a realistically implemented scheduling framework that augments existing contact tracing solutions. It should be noted that scheduling already takes place in varying capacities under a variety of contexts, such as universities and workplaces, for instance scheduling tutorials, labs, and classes. Additionally, many existing schedule constraint systems allow degrees of freedom in choosing locations of activities; as an example, larger university courses often run multiple tutorials concurrently in different classrooms, with arbitrary allocation of students therein. Integrating our approach with such systems therefore poses very little imposition, utilizing the degrees of freedom to optimize spatio-temporal interaction patterns for contact tracing. Our solution may be particularly useful where case numbers of transmissible diseases are low and sporadic. Our solution could allow institutions such as universities and workplaces to resume more physical activities without a need for mass restrictions; at the very least, if new outbreaks do occur and site access restrictions become necessary, contact tracing will be able to contain such outbreaks faster and more efficiently.

Global optimization generally requires global information (such as detailed trajectories of all traced individuals), meaning that such solutions would almost certainly involve severe privacy infringements; in Section 2.3 we show how our optimization can be achieved without compromising individual privacy, exposing no identifying data to a central system nor other users.

We have developed a full working prototype consisting of a user-facing mobile application and an administrator desktop application. The mobile application allows users to register for activities, which are then added to their timetables in a 'pending approval' state. Administrators may then periodically harvest, evaluate and schedule activities and locations using our privacy-aware, spectral scheduling framework. The demonstration application will be discussed in more detail in Section 3.

2 ACTIVITY SCHEDULING FRAMEWORK

2.1 Definitions

1. An **activity** is something abstract to be completed over a defined time span and location, for example 'studying'.
2. **Registrants** are individuals registering in activities.
3. **Equivalent activities** are the realizations of an activity (for example, a scheduled study session in a fixed location) similar enough that exchanging them would not be of unjust imposition to registrants. The distinction may solely be spatial, as with identical study sessions in different locations.
4. A **trajectory** is a sequence of a registrant's (possibly not yet approved) activities, i.e. a sequence of time-location tuples.
5. **Registrant trajectory similarities** are similarity metrics calculated between the trajectories of pairs of registrants.

2.2 Optimization Through Clustering

Let R be the set of all registrants, A be the set of all activities, and $A_r \subseteq A$ notate the set of activities for a registrant $r \in R$. Given two registrants r_0 and r_1, their trajectory similarity is therefore the Jaccard index $s(r_0, r_1) = \left|A_{r_0} \cap A_{r_1}\right| / \left|A_{r_0} \cup A_{r_1}\right|$[6]. At the core of our framework is the interpretation of scheduling as a partitioning of registrants from one big activity to smaller equivalent activities, each of which corresponds to a subset of the registrants. Consider an activity $a \in A$ with registrants $R_a \subseteq R$, which we plan to split into k equivalent activities. An example might be university students registering for a subject's tutorial, where there are k tutors/classrooms. To maximize assortative community structure, we want to partition R_a into k groups such that students with similar trajectories are kept together and students with dissimilar trajectories are kept apart. Construct a similarity graph S where nodes are the registrants in R_a and edges are weighted by the trajectory similarity of registrants at each end; the objective of clustering is then

	$n = 500$	$n = 1000$	$n = 1500$
$k = 2$	703 ms	3374 ms	23311 ms
$k = 5$	474 ms	3493 ms	23443 ms
$k = 10$	510 ms	3729 ms	24166 ms

Table 1: Running time of clustering n random trajectories of length 8 into k partitions on an AMD Ryzen 5 3600X CPU.

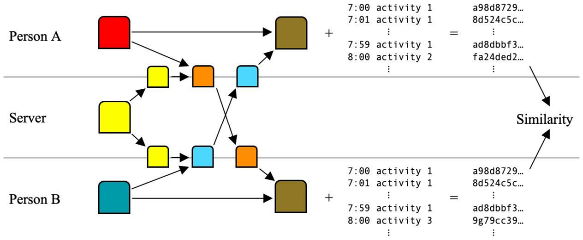


Figure 2: Decentralized, privacy-aware calculation of registrant trajectory similarities. Two registrants establish a common key hidden to the server, analogized through colors. This is then used to encode the registrants’ trajectories into digests which are sent to the server to calculate similarity.

to partition S into k components such that the total weight of edges between partitions is minimized. In this form, the optimization problem is NP-hard[9]. However, a variety of heuristic approaches are very well studied and are known to produce very high quality splits in polynomial time, the most common of which is arguably spectral partitioning[8], which runs in $O\left(|R_a|^{2.373}\right)$. Benchmark running times for our implementation are given in Table 1.

2.3 Decentralized Similarity Calculation

We have developed a privacy-aware protocol that allows the decentralized calculation of registrant trajectory similarities using Diffie-Hellman key exchange. Figure 2 illustrates this through the analogy of paint mixing. Assume that we want to calculate the similarity between the trajectories of two registrants. First, each registrant generates his or her own hash or ‘paint’. Then the server generates another random hash and sends the hash to the two registrants, both of whom return a combination of their own hash and the server’s hash. The server then sends these combinations to the opposite registrants, who finally combine the received hashes with that of their own. Assuming that we cannot unmix paint (i.e. the hash is not invertible in reasonable time), the registrants now have a common key (a product of the three hashes) not known nor calculable by the server. Each registrant’s mobile application will then discretize their schedule by concatenating time values with activities, then form digests by hashing with the common key. These digest lists are finally submitted to the central system, where a simple sort-merge join facilitates a quick calculation of the similarity. Note that no personal information, including non-encrypted trajectory information, is ever shared in this protocol.

2.4 Recursive Splitting

The operational flow of the framework is shown in Figure 3. The key concept employed is that of continual revision, as activities are recursively partitioned and evaluated to ensure that a satisfactory level of trajectory similarity between the registrants in each partition is always maintained. Activities have four statuses: ‘pending approval’, ‘rejected’, ‘scheduled’, and ‘approved’. After registrations are received, activities are given a ‘pending approval’ status. Administrators may then select such an activity for partitioning into any number of groups. The resultant partitions may then be examined and evaluated, to be approved, rejected or further partitioned.

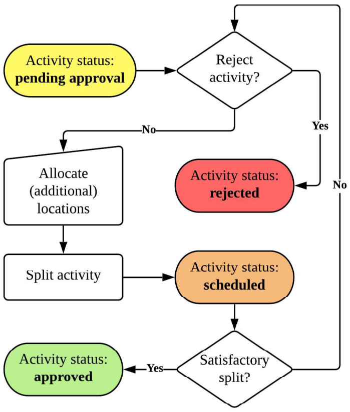


Figure 3: Transitions between activity statuses as administrators split activities to perform scheduling.

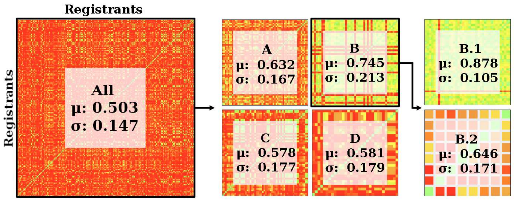


Figure 4: Similarity matrices as an activity is recursively split through spectral partitioning. The average value of registrant trajectory similarities μ increases with splits, while the standard deviation of similarities σ is minimized in B.1.

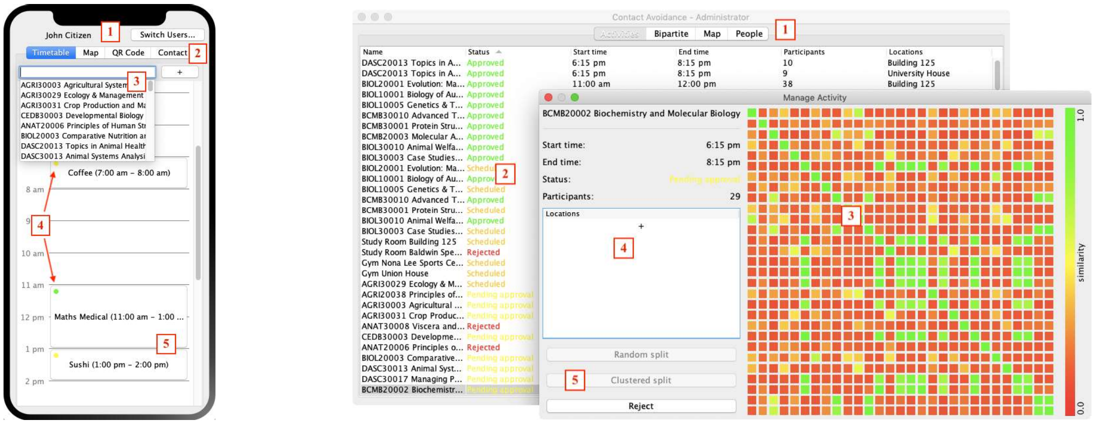
Figure 4 demonstrates this recursive activity splitting in action by showing one example of the evolution of similarity matrices. Cell colors represent pairwise similarities of individuals as defined in section 2.3; red (dark) indicates highly dissimilar individuals, while yellow (light) indicates highly similar individuals. This particular example shows a machine generated set of 300 trajectories split into four, with one partition split further into two; in a real application these parameters (such as the number of splits and acceptable similarity threshold) would be chosen at the administrator’s discretion. The initial average similarity is quite low, but as more splits take place, the average similarity progressively increases; partition B.1 finally captures a very tight cluster of individuals with highly similar trajectories, while partitions A, C and D would likely benefit from further downsizing (not shown).

3 DEMONSTRATION APPLICATION

We implemented the framework described above in a user-facing mobile application and an administrator-facing desktop application[1]. The mobile application is shown in Figure 5(a): (1) Account management; (2) Additional tabs (map, QR check-in, administrator

---

[1] Source code for demonstration application available on request



Name	Status	Start time	End time	Participants	Locations
DASC20013 Topics in A...	Approved	6:15 pm	8:15 pm	10	Building 125
DASC20013 Topics in A...	Approved	6:15 pm	8:15 pm	9	University House
BIOL20001 Evolution: Ma...	Approved	11:00 am	12:00 pm	38	Building 125
BIOL10001 Biology of Au...	Approved				
BIOL10005 Genetics & T...	Approved				
BCMB30010 Advanced T...	Approved				
BCMB30001 Protein Stru...	Approved				
BCMB20003 Molecular A...	Approved				
BIOL30010 Animal Welfa...	Approved				
BIOL30003 Case Studies...	Approved				
BIOL20001 Evolution: Ma...	Schedu				
BIOL10001 Biology of Au...	Approv				
BIOL10005 Genetics & T...	Scheduled				
BCMB30010 Advanced T...	Approved				
BCMB30001 Protein Stru...	Scheduled				
BIOL30010 Animal Welfa...	Approved				
BIOL30003 Case Studies...	Scheduled				
Study Room Building 125	Scheduled				
Study Room Baldwin Spe...	Rejected				
Gym Nona Lee Sports Ce...	Scheduled				
Gym Union House	Scheduled				
AGRI30029 Ecology & M...	Scheduled				
AGRI20038 Principles of...	Pending approval				
AGRI30003 Agricultural ...	Pending approval				
AGRI30031 Crop Produc...	Pending approval				
ANAT30008 Viscera and...	Rejected				
CEDB30003 Developme...	Pending approval				
ANAT20006 Principles o...	Rejected				
BIOL20003 Comparative...	Pending approval				
DASC30013 Animal Syst...	Pending approval				
DASC30017 Managing P...	Pending approval				
BCMB20002 Biochemistr...	Pending approval				

BCMB20002 Biochemistry and Molecular Biology

Start time: 6:15 pm
End time: 8:15 pm
Status: Pending approval
Participants: 29
Locations

(a) User interface (b) Administrator view

Figure 5: Demonstration application; see text for further detail on numbered application components.

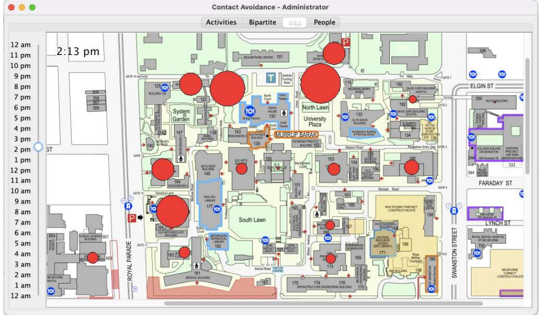


Figure 6: The map tab allows administrators to visualize how the spatial distribution of registrants evolves over time.

contact); (3) List of activities to register; (4) Activity status: yellow (pending approval), orange (scheduled), green (fully approved), red (rejected); (5) Viewing and deletion of activities. Through the app, users register for activities in advance. At specific intervals, registrations are collected and processed. The administrator application is shown in Figure 5(b): (1) Administrator tools such as a map-based visualization tool (Figure 6); (2) Activity statuses; (3) An $n \times n$ similarity matrix constructed pairwise for the activity's registrants in a decentralized, privacy-aware manner; (4) Administrators allocate candidate locations to the activity; the system may also suggest locations, e.g. previous locations of recurrent activities; (5) Activities are split across locations, accepted if satisfactory, or rejected.

4 CONCLUSION AND FUTURE RESEARCH

Using clustering, our framework facilitates decentralized scheduling that proactively improves the efficiency of subsequent contact tracing and also reduces the spread of disease. This solution is privacy-aware, effective even in the absence of known positive cases, and computationally efficient and thus scalable. This combination of qualities is largely absent from existing solutions but is likely to be increasingly crucial in establishing the best 'COVID normal'. Future research is possible in extending the proposed clustering to incorporate other biases such as predicted epidemiological risk. As an example, variations on the similarity proxy may be tailored to account for factors such as environmentally mediated transmission and registrants' age, health and preexisting conditions. In addition, this research leads to numerous spatial data management problems such as in semantic trajectory management, where spatio-temporal constraints in clustering are investigated. It could also be of merit to study how the framework's performance degrades with reductions in data completeness and user compliance.

ACKNOWLEDGMENTS

This work was sponsored in part by the NSF under grants IIS-18-16889, IIS-20-41415, and IIS-21-14451.

REFERENCES

[1] R. Agarwal and A. Banerjee. 2020. Infection risk score: identifying the risk of infection propagation based on human contact. In *SIGSPATIAL '20: Proceedings of the 28th International Conference on Advances in Geographic Information Systems.* https://doi.org/10.1145/3423459.3430754

[2] X. Cao. 2020. COVID-19: immunopathology and its implications for therapy. *Nat Revs Immunology* 20 (4 2020). https://doi.org/10.1038/s41577-020-0308-3

[3] Y. Chang, J. Qi, E. Tanin, X. Ma, and H Samet. 2021. Sub-trajectory similarity join with obfuscation. In *33rd International Conference on Scientific and Statistical Database Management* (Tampa, FL, USA) *(SSDBM 2021)*. Association for Computing Machinery, New York, NY, USA, 181–192. https://doi.org/10.1145/3468791.3468822

[4] K. T. D. Eames and M. J. Keeling. 2003. Contact tracing and disease control. *The Royal Society Publishing* (12 2003). https://doi.org/10.1098/rspb.2003.2554

[5] N. M. Ferguson et al. 2020. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College London* (3 2020). https://doi.org/10.25561/77482

[6] P. Jaccard. 1907. La Distribution de la Flore dans la Zone Alpine. *Revue Générale des Sciences* (12 1907).

[7] T. Leng et al. 2021. The effectiveness of social bubbles as part of a COVID-19 lockdown exit strategy. *Wellcome Open Res* (2021). https://doi.org/10.12688/wellcomeopenres.16164.2

[8] U. von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and Computing* 17 (8 2007). https://doi.org/10.1007/s11222-007-9033-z

[9] R. M. MacGregor. 1978. *On partitioning a graph: a theoretical and empirical study.* Ph.D. Dissertation. EECS Department, University of California, Berkeley. http://www2.eecs.berkeley.edu/Pubs/TechRpts/1978/8368.html

[10] M. Nicolaa et al. 2020. The socio-economic implications of the coronavirus pandemic (COVID-19). *International Journal of Surgery* 78 (6 2020). https://doi.org/10.1016/j.ijsu.2020.04.018

[11] H. Samet, Y. Han, J. Kastner, and H. Wei. 2020. Using Animation to Visualize Spatio-Temporal Varying COVID-19 Data. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19* (Seattle, WA, USA). 53–62. https://doi.org/10.1145/3423459.3430761

[12] G. Thakur et al. 2020. COVID-19 joint pandemic modeling and analysis platform. In *SIGSPATIAL '20: Proceedings of the 28th International Conference on Advances in Geographic Information Systems.* https://doi.org/10.1145/3423459.3430760

[13] World Health Organisation. 2021. Weekly epidemiological update on COVID-19 - 17 August 2021. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---17-august-2021