SCORE: Scalable Contact Tracing over Uncertain Trajectories

Avinaba Mistry(✉), Xichen Zhang, Suprio Ray, and Sanjeev Seahra

University of New Brunswick, Fredericton, Canada
{avinaba.mistry,xichen.zhang,sray,sseahra}@unb.ca

Abstract. In the context of a global pandemic, mitigating contagion risk requires an integrated analysis of global positioning data from location-based services and complex disease dynamics varying across geography and demography. However, the mobility datasets have inherent issues of imprecision and of being high volume. This is compounded by the challenges of changing pharmacological and non-pharmacological context of contagion behaviour, geography, demography, public health strategies across the globe.

In this paper, we propose a comprehensive framework, SCORE, to provide new analytical tools for public health strategy and planning. We also propose a novel data structure, DisCoUnt, which serves as a distributed uncertain trajectory index for moving objects as well as infection event data. We conduct extensive experiments to demonstrate the scalability of our query workflow for an infection risk measure over uncertain trajectories.

Keywords: location-based-service applications · mobile data · multi-dimensional index · Contact Tracing

1 Introduction

As of September 27, 2023, the cumulative positively identified infections due to COVID-19 (SARS-CoV- 2) had risen to 770,875,433 with a death toll of 6,959,316 as reported by WHO Coronavirus (COVID-19) Dashboard [55]. The original contagion strain showed rapid mutation and exhibited changes in infection propagation, latent period, symptoms and fatality proportions [45]. COVID-19 is not an isolated phenomenon. In fact, throughout human history a number of major pandemics affected millions of human beings, and thus changing the course of human civilization. For instance, the Plague known as the Black Death, during 1346 to 1353 resulted in the death of one third of the entire population of Europe [44]. Besides Covid-19, in recent years there have been a number of major outbreaks, including Middle East Respiratory Syndrome coronavirus (MERS-CoV) [39], the Ebola virus (EVD) [41], and the Zika virus (ZIKV) [42]. Epidemiologists predict that future outbreaks are highly likely [43].

During the course of a pandemic, such as Covid-19, public health agencies need to conduct real-time infection risk-assessment and develop mitigation

A. Zaslavsky et al. (Eds.): MobiQuitous 2023, LNICST 593, pp. 68–93, 2024.
https://doi.org/10.1007/978-3-031-63989-0_4

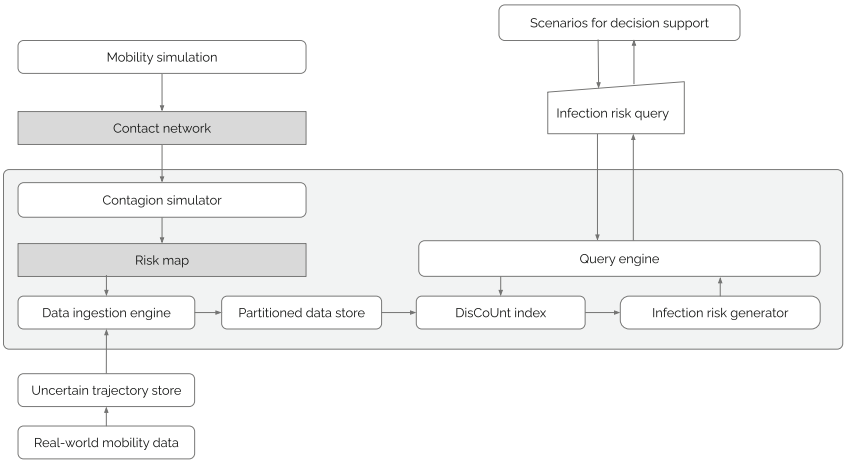

Fig. 1. Scalable Contagion Risk Estimation (SCORE) architecture

strategies utilising interventions to contain the risk of infection by obstructing the propagation of contagion. These strategies include effective pharmaceutical interventions like vaccination strategies as reported by Pritchard et al. [23], Makhoul et al. [24], Layan et al. [25] and Bian et al. [26]. Non pharmaceutical interventions (NPIs) have also been effective in mitigating public health risk. NPIs include *community masking*, *physical distancing*, *social bubbles*, *closures* or *lockdowns* as stated by Abueg et al. [31], Hens et al. [32], Lai et al. [33], Cheng et al. [34] and de Souza Santos et al. [35].

However, depending on the selection of intervention measures, the time of implementation and the coverage area of its effects, the efficacy of the intended outcome may exhibit significant variance. In addition to the technical complexities, executing these interventions requires mobilisation of large amount of resources and has psycho-social consequences. Therefore it is of utmost importance for public health policy makers to be informed with optimal timely intervention strategies for a given scenario. And more often than not this involves determining short-term estimates for multiple different mitigation strategies before context drifts farther away from initial model assumptions. Contagion modelling has been an integral part of any decision support system for public health to inform their strategies. Contagion modelling utilising ordinary differential equation (ODE) based disease propagation models abstracts the infection propagation of contagion over groups of individuals, which is computationally fast but does not capture spatio-temporal dynamics and demographic heterogeneity. On the other hand, there are individual level agent based infection propagation models, where each human's movement is simulated over a real world geo-spatial area but it is computationally expensive to model both movement and contagion propagation. A mid-way approach between these two is to extract a contact network of individuals at POIs based on detailed simulation

and simulate contagion propagation over said network, which preserves the heterogeneity of a real-world community. In other words, contagion modelling may be considered as construction of a risk of infection map across a geography with reasonable spatial and temporal resolution. We have witnessed efforts in identification of real-world contact networks with large scale contact tracing efforts from public health agencies.

It has been well established that contact among people due to human mobility plays a significant role in epidemic spreading. The issue of contagion risk identification and assessment at scale may be formulated as a problem of identifying sufficient spatio-temporal overlaps between moving objects. In other words, for a given geographic area and overarching temporal bound, we need to filter moving objects that may have the opportunity to transfer infection inducing viral load. For e.g. Public Health of Canada defined a temporal overlap of 15 min over 24 h within a proximity of 2 m for contact identification and management purposes [2]. Utilisation of mobility data to assess contagion risk has been explored in the following works [3–7]. However, these approaches have limited applicability due to imprecise nature and very high volume of spatio-temporal data. Cicala et al. [8] note that proximity based contact tracing (PCT) protocols, which work by logging distances between smartphones with PCT apps, may estimate the distance with a degree of uncertainty. For one, Bluetooth Low Energy (BLE)-based PCT techniques use the signal strength for estimating the distance and two, the individuals may be carrying their smartphones further from their position, for instance in their luggage instead of in their pockets.

To address the limitations of existing approaches, we propose a decision support system for public health agencies called SCORE (Scalable COntagion Risk Estimator). The overall architecture of SCORE is shown in Fig. 1. The aims of SCORE include providing decision support to public health efforts for estimating risk posed by a contagion, for a specific socioeconomic and geographic context with respect to changing intervention scenarios through individual level models in combination with real-world trajectory data. Our framework includes a contagion simulator (as shown in Fig. 1) that captures real-world spatio-temporal complexities of modelling a rapidly changing contagion with respect to intervention strategies and geo-spatial variance.

The key contributions of this paper are as follows:

- We propose a novel, easy to parallelise distributed spatio-temporal index DisCoUnt, to enable efficient uncertain trajectory indexing as well as infection data.
- We illustrate how infection data may be dis-aggregated into specific points of interests (POIs) of a geography utilising contagion simulators.
- Our experimental evaluation demonstrate the efficacy of DisCoUnt in a distributed environment.

The rest of the paper is organized as follows. In Sect. 2, we outline the research questions. We describe our overall architecture of SCORE in Sect. 3. We discuss the essential elements of a contagion simulator in Sect. 4. In Sect. 5 we describe in-depth our infection risk query processing and distributed index. In Sect. 6

we describe our experimental evaluation. In Sect. 7 we discuss the limitations of this paper. In Sect. 8, we present related work. Finally, we draw conclusions in Sect. 9.

2 Research Questions

In this paper we focus on the following three questions:

RQ1 How could we combine analyses of infection data and trajectory with an appropriate spatial resolution?
RQ2 Given an appropriate combination of trajectory and infection data, how could we define a risk measure for a given moving object?
RQ3 Given the definition of a risk measure of a moving could, how might we efficiently compute it?

3 SCORE Architecture

As shown in Fig. 1, SCORE relies on two sources of data, namely the *Contact network* and *Uncertain trajectory data* pre-processed from *Real world mobility data*. It utilises a network based contagion simulator with an appropriate disease model and the input *Contact network* to produce a *Risk map* for a community. The *Data ingestion engine* is responsible for distributing *Risk map* and *Uncertain trajectory store* into the available nodes in the distributed network. The proposed Distributed Contact detection over Uncertain trajectory *(DisCoUnt)* index is responsible for distributed execution of a *Infection risk query* and gathering of filtered data points. The data points are passed to *Infection risk generator* for a final output, which is passed back to *Query engine* and finally back to the user.

In the next sections we describe each component of our system.

3.1 Mobility Simulation

Human mobility simulation entails analysis of available mobility trace datasets derived from cellular network, location based services, public transportation and survey to produce a model that can output reasonable movement patterns of a community. For example, Toronto Area Scheduling Model for Household Agents (TASHA) is a prototype activity scheduling micro-simulation model to generate a typical 24-hour workday or *schedule* for individuals of a household proposed by Miller et al. [46]. The model has been subsequently validated by Roorda et. al [47] for the Greater Toronto Area, Canada for the 1996 travel survey data from which the model is derived from and the comparison with forecast for 2001 travel survey data.

3.2 Contact Network

A contact network with respect to network-based individual-level contagion model is an abstraction that captures the co-locations of individuals for a given time-step. Here each vertex is an individual and each edge represents an event of shared time and space. A representative example is show in Fig. 2.

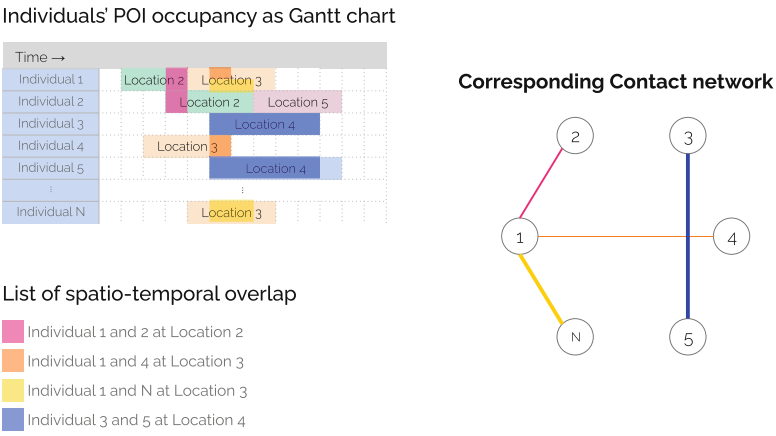

Fig. 2. Generation of contact network for 6 individuals from their POI occupancy data

3.3 Contagion Simulator

A contagion simulator that operates on contact networks, infers the propagation of a contagion based on a disease model, intervention strategies and underlying health factors associated with the demography and POIs contained within the geography. Their output includes time series data of infection events across POIs. We describe parts of a typical network-based contagion simulator relevant to this framework in Sect. 4.

3.4 Risk Map

Based on the output of a calibrated contagion simulator, the infectious individuals' POI occupancy data is dis-aggregated across the spatial bounds of a POIs as discrete tiles.

3.5 Uncertain Trajectory Store

Given any two-dimensional trajectory data as a sequence of observations $o_1, o_2, ...o_n$, where each o_i is position data and the associated timestamp: $o_i = (x_i, y_i, t_i)$, it may be encoded as uncertain trajectory segments whenever appropriate. In our framework we choose a shape-based encoding of uncertainty of moving objects called *Beads* and *Necklaces*.

A **Bead** is a shape-based model for an uncertain trajectory as proposed by Trajcevski et al. [10]. The construction of a Bead relies on a pair of observation (o_i, o_{i+1}) of a moving object m_k and a maximum possible velocity m_k may attain for the particular trajectory segment v^i_{max} dictated by real-world constraints.

Mathematically, a Bead B_i may be constructed from an observation pair (o_i, o_{i+1}) is an ellipse with the following equation:

$$\frac{(2x - x_i - x_{i+1})^2}{{v^i_{max}}^2(t_{i+1} - ti)^2} + \frac{(2y - y_i - y_{i+1})^2}{{v^i_{max}}^2(t_{i+1} - ti)^2 - (x_{i+1} - x_i)^2 - (y_{i+1} - y_i)^2} = 1 \quad (1)$$

A sequence of Beads B_1, B_2, ... B_n created from a sequence of observations o_1, o_2, ... o_{n+1} is referred to as a **Necklace**. In [10] the authors recommend the following strategies to approximate Beads and Necklaces with reasonable geometric constraints to balance the conflict between imprecision and encoding and analysis efficiency:

- **Large Bound** (LB): The entirety of a necklace is bounded by a single minimal vertical cylinder.
- **Individual Bead Bound** (IBb): Each Bead B_i is approximated by a circle with their radius r_i as the semi-major axis of the underlying Ellipse. The r_i is dependent on the choice of v^i_{max}, the maximum possible velocity between observations o_i and o_{i+1}.
- **Uniform Bead Bound** (UBb): Similar to IBb, each Bead B_i is approximated by a circle with their radius r_i as the semi-major axis of the underlying Ellipse. But the choice of v^i_{max}, is constant as the maximum possible velocity between any two points o_i and o_{i+1}.

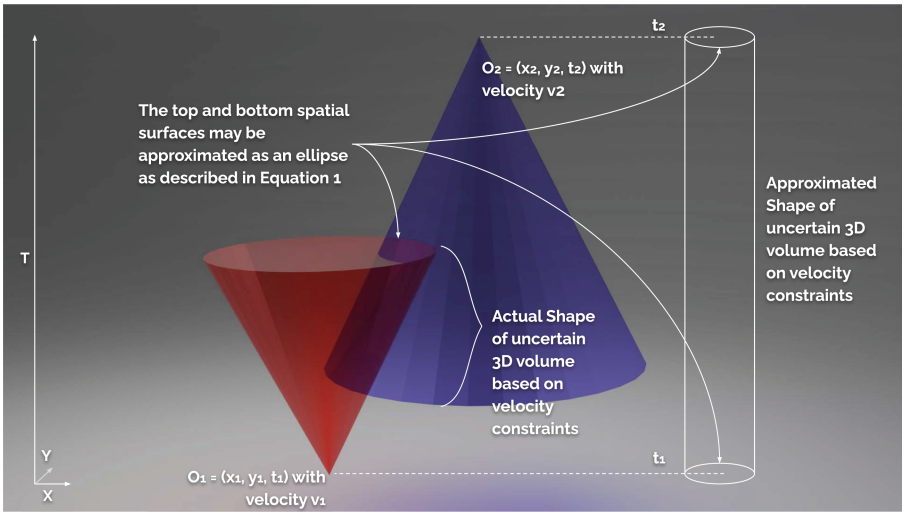

Fig. 3. Approximating trajectory segments as ellipses extruded over T axis as proposed by Trajcevski et al. [10]

Please note, even with the least lossy bead encoding such as, **Individual Bead Bound,** the spatial bound of a moving object can be approximated by an ellipse across the time bound of two subsequent observations, as shown in Fig. 3. For a detailed consideration of pruning strategies we refer the reader to Section III-B of the paper by Trajcevski et al. [10]. For example of IBb Bead

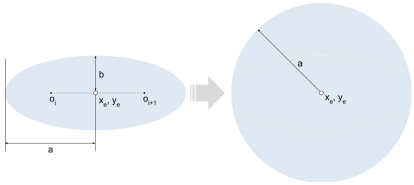

Fig. 4. Approximating spatial bound of Beads from ellipses to circles as proposed by Zhang et al. in [1]

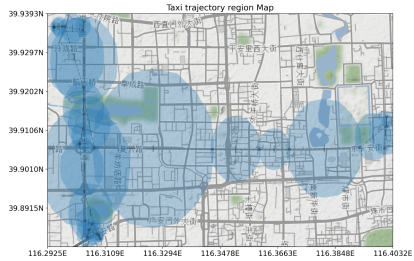

Fig. 5. A necklace with approximated circular beads distorted on Equirectangular projection with GIS mobility trace data provided by Yuan et al. [11]

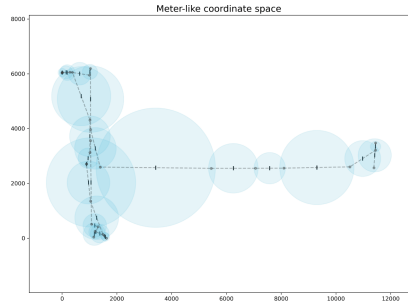

Fig. 6. Same necklace as shown in Fig. 5 projected on Cartesian plane

Necklace based on real-world trajectory data as provided by Yuan et al. [11], please refer to Figs. 5 and 6. Note that, although the spatial bounds of beads are approximated as circles from ellipses as shown in Fig. 4, they appear as ellipses due to distortions introduced by cartographic projection.

3.6 Combining Risk Map and Uncertain Trajectory Store

Taking a similar approach to model spatial-congestion in different POIs as illustrated by Alix et al. [21], we chose the discrete tiles for infectious individual POI occupancy to be hexagonal with appropriate resolution for humans. Next we approximate the spatial bounds of hexagons with its corresponding circumscribed circle or circumcircle to get a cylinder in $X \times Y \times T$ space. This is a data structure equivalent to our encoding of an uncertain trajectory. Thus, a Bead may be utilised to store both data points of real-world mobility as well as space-time occupancy of infectious individuals in a POI. Please note, although the trajectory data of humans are technically in 3 spatial and 1 temporal dimensions, most of human activities may be considered to be taking place on 2D planes with connectors like stairs or elevators and therefore without loss of generality we may consider the human trajectories as points in $X \times Y \times T$ space.

3.7 Data Ingestion Engine and Partitioned Data Store

The *Data ingestion* component of the framework is responsible for creating a *Partitioned data store* for data points of both *Riskmap* and *Uncertain trajectory store*. Two partitioning heuristics are discussed in Sect. 5.

3.8 Infection Risk Query

This query involves the identification of infection risk associated with a moving object present in the *Uncertain trajectory store*.

For example, given a scenario "How risky is to visit this mall?" one may translate it into a SQL-like query as follows:

```
INFECTION_RISK(*) FROM UNCERTAIN_TRAJECTORY_STORE, SIMULATION_STORE,
WHERE RISK_MAP = SIMULATION_STORE.RISK_MAP,
UNCERTAIN_TRAJECTORY_BEADS = UNCERTAIN_TRAJECTORY_STORE.BEADS AND
POLYGON == "MALL_01" AND
TIME BETWEEN YESTERDAY AND TODAY
```

The details of infection risk query processing is described in Sect. 5.

3.9 Query Engine

A query engine is the component responsible for interpreting a user's query into a query plan along with execution details taking into account the indices and possible speedups. Here, the proposed query engine augment interprets a spatio-temporal query and produces a query execution plan using *DisCoUnt index*.

3.10 DisCoUnt Index

This is our proposed novel distributed index, which facilitates efficient search and retrieval from *Partitioned data store*. In particular, the index is designed to support scalable evaluation of infection risk queries. This is described in Sect. 5.2.

3.11 Infection Risk Generator

This component computes the probability of infection risk with respect to the retrieved data points of a *Infection risk query*. Details of infection risk measure and its computation is described in Sect. 5.1.

4 Contagion Simulator

A detail discussion of contagion simulation in general or even network-based stochastic contagion simulation in specific is out of scope for this paper. Therefore, here we briefly discuss how an individual level network based geo-spatial contagion simulator combines demography data, geo-spatial data and disease models.

4.1 Combining Demography and Geo-Spatial Data

Due to privacy and security issues being among the primary concerns, the resolution of trajectory or trajectory traces of humans is limited and may not be suitable to determine the spatio-temporal overlaps required for infection propagation. Therefore one technique to generate human mobility data that captures the variations of the age distribution and the network of POIs is to simulate it via human mobility models. A simplified description of this process is outlined below.

Step 1 First, a statistically similar individual level population data is synthesised based on the available aggregated census and labour data for a community.

Step 2 In parallel, road-network and POI data are extracted for the same community from sources like OpenStreetMaps.

Step 3 A mobility simulation is performed with a relevant time-step, say 24 h, with the data obtained in **Step 1** and **Step 2** to obtain movement patterns of the individuals.

This movement pattern data is transformed into individual occupancy data consisting of POIs and finally transformed into a contact network.

4.2 Infectious Disease Modelling as Compartments

Over the years, several infectious disease models have been proposed. According to Keeling et al. [48], in these models, individuals of a population are put into compartments in relation to the contagious disease, which are denoted by labels such as, *Susceptible*: individuals who may get infected, *Infectious*: individuals who may infect susceptible individuals, *Latent*: an individual who is infected but not yet infectious, *Removed*: Individuals who are not participating in the propagation of disease due to immunity or death and so on. At any given time, a person may be in exactly one of the compartments. One of the simplest of contagion models comprises of three compartments, namely Susceptible, Infectious and Recovered and is called the SIR model.

For our simulation purview, we invert this perspective. That is, we do not consider mixing of groups of individuals (as compartments) and investigating the parameters from the real world to fit into the measures of central tendencies. Instead we simulate the movement of the population to the best of our abilities and capture the POI occupation events as an emergent property of the simulation to generate the contact network. Then we simulate disease propagation on that network, without loosing individual variations of health statuses and keeping a portion of the geo-spatial heterogeneity.

4.3 Interpreting Infection Propagation

Irrespective of the design of an infectious disease model, it must always contain a mechanism of transition from a non-infectious state to an infectious state. As

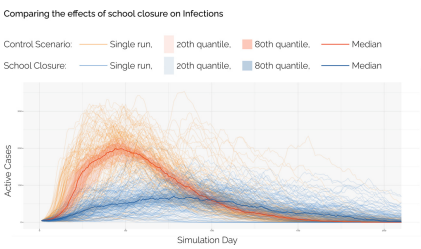

Fig. 7. Simulation of the effect of school closure on the infection propagation in a community from [38]

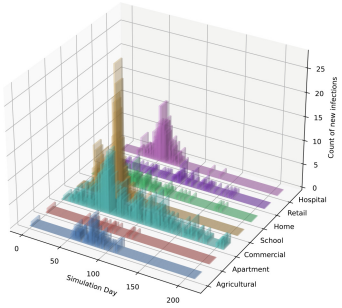

Fig. 8. Simulation of relative infection counts with respect to the venue types in a community from [38]

per the mechanistic model of viral load transmission by Challenger et al. [49], each infectious individual may be thought of as a emitters of air-born virus by virtue of exhalation. If and only if there is sufficient transfer of said virus (a.k.a viral load) from an infectious to susceptible individual, there is a possibility that, infection may spread to the susceptible individual. This interaction is modelled as a infection probability function f, which in the simplest case is proportional to the shared time t between a pair of susceptible and infectious individual in a POI.

4.4 Example Outputs of Contagion Simulation Without Calibration from [38]

As part of separate research activities on contagion simulation over real-world communities in [38] we have utilised CitiSketch am [56] to generate representative contact networks. The figures are example outputs of infection spread in a community of 4602 individuals from the Campbellton region of New Brunswick, Canada. The demography and labour data were procured from Statistics Canda [57] and the Road network and POIs from OpenStreetMap [58] respectively.

Figure 7 shows relative infection risk with school closure as a mitigation strategy. Figure 8 illustrates the relative infection risks for venue types.

Upon calibration, an individual level contagion simulator can provide POI level occupancy data of expected number of infectious individuals for an outbreak.

5 Infection Risk Query Processing

Before we begin to describe the query workflow in-depth, we answer **RQ2** with the following definitions.

5.1 Infection Risk Measure

Taking inspiration from Zhang et al. [1], we approach to define the probability of transmission of contagion from an infectious individual described by an occupancy encoding bead to a susceptible individual described by an uncertain trajectory encoding bead as follows:

We begin with the premise that two beads B_1 and B_2 describing object O_1 and O_2 respectively have spatio-temporal overlap with area of spatial intersection ρ_{12} and maximum possible time overlap of τ. Let us also assume that the probability of finding an object within the spatial bound defined by a bead is uniform and remains constant across time.

Definition 1 ***Probability of spatial contact.*** *This probability may be decomposed into three events as follows:*

- $P(E_1)$*: When sampled at random the probability of* O_1 *to lie within spatial intersection.*
- $P(E_2)$*: When sampled at random the probability of* O_2 *to lie within spatial intersection.*
- $P(E_3)$*: Given the a priori probabilities of* E_1 *and* E_2*, the probability that the sampled points are at most* δ *distance apart.*

Since the events are independent of each other the resulting probability is as follows:

$$P_{space}(B_1 \cap B_2) = P(E_1) \cdot P(E_2) \cdot P(E_3) \tag{2}$$

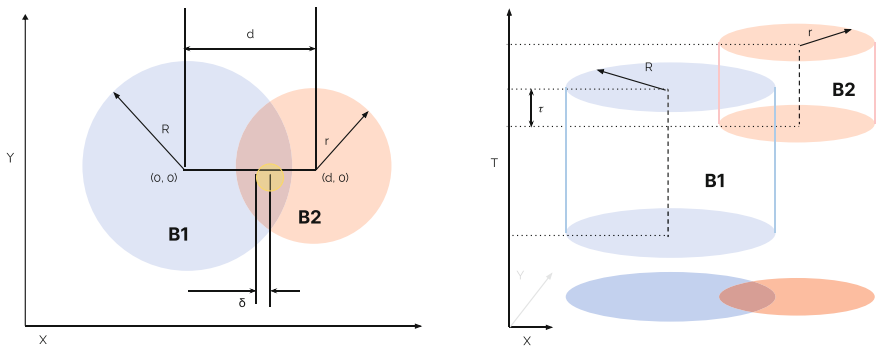

Fig. 9. Schematic describing probability of spatial contact and temporal overlap between beads

Without loss of generality, consider the two Beads in Fig. 9. Here, B_1 *is translated to* $(0, 0)$ *and has a radius of* R *with* B_2 *at* d *distance with radius* r*. Then, by construction, we have the following results:*

$$P(E_1) = \frac{\rho_{12}}{\pi R_1^2} \textit{ first object is on the overlapping area} \tag{3}$$

$$P(E_2) = \frac{\rho_{12}}{\pi r_1^2} \textit{ second object is on the overlapping area} \tag{4}$$

$$max(P(E_3)) = \frac{\pi\delta^2}{\rho_{12}} \textit{ both of them are "close enough"} \tag{5}$$

And the spatial area of intersection ρ may be computed as follows:

$$\begin{aligned}\rho_{12} &= r^2 \cos^{-1}\left(\frac{d^2 + r^2 - R^2}{2dr}\right) \\ &+ R^2 \cos^{-1}\left(\frac{d^2 + R^2 - r^2}{2dr}\right) \\ &- \frac{1}{2}\sqrt{(-d + r + R)(d + r - R)(d - r + R)(d + r + R)}\end{aligned} \tag{6}$$

Putting it all together, we have the following probability of spatial contact between the objects O_1 and O_2 described by two beads B_1 and B_2 with a threshold of minimum distance of effective contact as δ.

$$\begin{aligned}P_{space}(B_1 \cap B_2) &= \frac{\rho_{12}}{\pi R_1^2} \cdot \frac{\rho_{12}}{\pi r_1^2} \cdot \frac{\pi\delta^2}{\rho_{12}} \\ &= \frac{\rho_{12} \cdot \pi\delta^2}{\pi R_1^2 \cdot \pi r_1^2}\end{aligned} \tag{7}$$

Definition 2 ***Probability of infection propagation***
Let us also assume that this pair (O_1, O_2) is a pair of susceptible and infectious objects with a maximum possible temporal overlap be τ_{12}.

A mechanistic model of infection transmission may be generalised as a function of time overlap t between two individuals in close proximity with a minimum cutoff time t_{min}: $f(t, t_{min})$. Therefore, the final probability may be decomposed into two events as follows:

- *$P_{space}(B_1 \cap B_2)$: A priori probability of spatial contact.*
- *$P(transmit_{12})$: $f(\tau_{12}, t_{min})$*

Therefore the probability of infection from O_1 to O_2, for any O_I who is infectious and O_S who is susceptible is as follows:

$$\begin{aligned}P(O_I \xrightarrow{infects} O_S) &= P_{space}(B_1 \cap B_2) \cdot P(transmit_{12}) \\ &= \frac{\rho_{12} \cdot \pi\delta^2}{\pi R_1^2 \cdot \pi r_1^2} \cdot f(\tau_{12}, t_{min})\end{aligned} \tag{8}$$

Thus we have a risk measure for a single space-time overlap. When extended for an complete uncertain trajectory of a moving object with multiple beads we have the following definition.

Definition 3 ***Infection risk measure.*** *for a susceptible moving object O_i with a trajectory T_i may be defined as follows:*

- *Let T_i be decomposed into a Necklace of Beads as* $N_i = [B_{i1}, B_{i2}..., B_{ik}]$
- *Let the set of all infectious beads be* $S_{inf} = B_{inf1}, B_{inf2}, ...B_{infm}$
- *Then, we find the set of all the overlap of pairs of objects from* S_{inf} *and* N_i *as* $S_{overlap}$.
- *Any object* $s \in S_{overlap}$ *is a pair of beads* $(B_i, B_j) \mid B_i \in N_i, B_j \in N_{inf}$
- *Then to compute the overall infection risk of* O_i *with respect to* T_i*, we must compute every* $P(O_j \xrightarrow{infects} O_i)$ *for each* $(B_i, B_j) \in S_{overlap}$

5.2 DisCoUnt Index

To accommodate efficient search over a partitioned Bead store we took inspiration from UTM-tree proposed by Zhang et al. [1]. UTM-tree is constructed by extending a metric-index M-tree [12] by including the radius of a bead as part of the data point and including it in the metric distance function for space based pruning and including a range-tree, proposed by Bentley et al. [50] for time based pruning in a single traversal of the resultant tree. However, UTM-tree, like that of a M-tree relies on a bottom-up insertion approach. This step ensures a height-balance and overlap reduction of the nodes of the tree and therefore subsequent spatial indexing efficiency. Therefore, it is prudent to partition the trajectories according to children or sub-trees of a node to ensure efficiency in parallel or distributed context. However, the tree creation process is bottom up, resulting in shuffling of trajectories among partitions making it unsuitable for partition based scaling up as is. Also, UTM-tree, like M-tree, utilises ball-type partitioning that introduces overlaps between sibling nodes (nodes that are at the same depth) with overlaps that may not be effectively mitigated as it is dependent on stochastic selection of routing objects during tree creation. This in turn poses a non-deterministic challenge of selection of partitions for further data insertion, which may cause data replication across partitions raising memory cost of indexing to an unsatisfactory degree. Although, it is possible to partition the trajectory segments before constructing individual UTM-tree indices out of each partition, the resultant data structure would be a hybrid index, for which ensuring overall efficiency would become a multi-part challenge. Instead, we considered alternative candidates for indexing based on the systematic evaluation of metric-space indices by Chen et al. [13] of range query on main-memory. The highest ranking technique reported was Multiple Vantage Point tree (MVPT) [14], followed by Extreme Pivot Table (EPT*) [15], Bisector Tree (BST) [16] and Geometric Near-Neighbour Access tree (GNAT) [17] and its variants. Since MVPT and partitioning of BST and EPT also utilises ball partitioning, it posed similar challenges to that of an M-tree in a setting of a distributed cluster of nodes. To support efficient uncertain trajectory indexing and infection data, we propose a novel distributed spatio-temporal index DisCoUnt, Next we describe the components of this index.

The membership of Beads and DisCoUnt node, and their attributes from an example index instance are shown in Fig. 10.

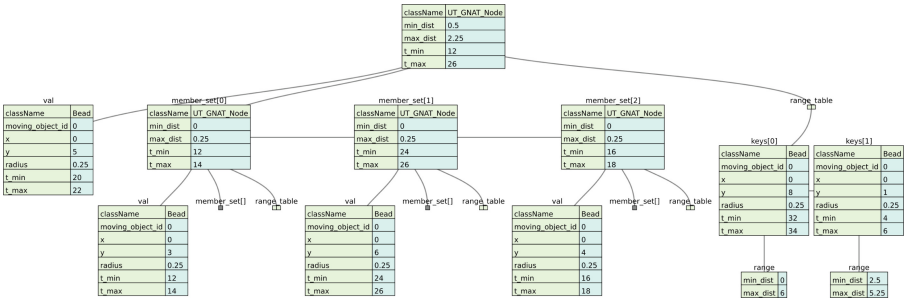

Fig. 10. An example DisCoUnt index instance visualised with JSON tree visualiser [36]

DisCoUnt Node Structure. The following parameters are sufficient for tree construction and pruning both spatial and temporal ranges.

- **Bead:** Contains full information of an Uncertain Trajectory segment:
  - m_{id}: moving object id
  - x, y, $radius$: spatial range of $radius$, centred on (x, y)
  - t_{start}, t_{end}: temporal range
- min_dist: minimum distance to its member set
- max_dist: maximum distance to its member set
- $*t\_max$: maximum of all t_ends of its member set
- $*t\_min$: minimum of all t_starts of its member set
- $member\_set$: set of member DisCoUnt nodes
- $range\_table$: set of member DisCoUnt nodes

The fields marked with an asterisk are introduced to keep a temporal range tree like secondary index structure for efficient temporal pruning during spatio-temporal range query computation.

5.3 Distance Function for Symmetric Range in Metric Space

Since we are dealing with distances between symmetric ranges of space, the distances between tWo Beads, B_1 and B_2 may be defined as Eqs. 9, 10 and 11 below and illustrated in Fig. 11:

$$c2c\_dist(B_1, B_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{9}$$

$$min\_dist(B_1, B_2) = max(0, c2c\_dist(B_1, B2) - r_1 - r_2) \tag{10}$$

$$max\_dist(B_1, B_2) = c2c\_dist(B_1, B2) + r_2 \tag{11}$$

Here, min_dist is a stand-in for overlap detection and therefore has the "symmetry" property, i.e.:
$min\_dist(B_1, B_2) = min\_dist(B_2, B_1)$
But it, does not follow "triangle inequality" for all B_1, B_2.

Therefore, we need *max_dist*, which does follow "triangle-inequality": $max\_dist(B_1, B_2) \geq max\_dist(B_1, B_3) + max\_dist(B_2, B_3)$.
This is true by construction as, $r_2, r_3 \geq 0$ and $c2c\_dist(B_1, B_2)$ is Euclidean distance between (x_1, y_1) and (x_2, y_2).

Please note, *min_dist* does not strictly have the "identity" property, since any two overlapping pair of *Beads* yields a 0 and neither does *max_dist* as, $max\_dist(B_1, B_2) > 0$ for all $r_2 > 0$.

For split point selection and membership assignment, we utilise the *min_dist* between DisCoUnt nodes and construct the DisCoUnt index. And, utilise *max_dist* for relative range based pruning and as a stand-in for maintaining minimum bounding circle for *member_set* of a *Bead*. Thereby making sure during each operation, only one kind of distance calculation is performed.

5.4 Algorithmic Complexity of a Single Index Construction

In this section we discuss the algorithmic complexity of our proposed index, particularly, the space and time complexity.

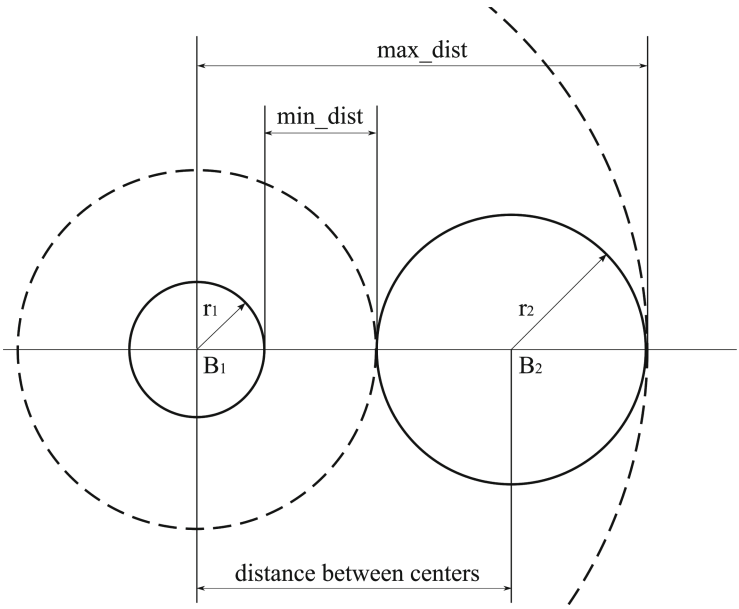

Fig. 11. Max and min spatial distances between two Beads

Space Complexity. Space complexity of DisCoUnt is the same as the data structure GNAT. It is as reported by the author of GNAT in [17] the space complexity is $O(nk_2 + Ns)$ where

1. n: number of nodes.
2. k_2: the average of the squares of the degrees.
3. N: number of data points (in a partition).
4. s: maximum space requirement for a data point.

Time Complexity. Time complexity of DisCoUnt construction is the same as GNAT. The time complexity for a completely balanced GNAT [17], is: $O(Nklog_kN)$, where k is the average degree (N/n).

Please note, our approach does not change the number of distance calculations. Hence, the O of space and time complexity of index building is the same in spite of accounting for maintaining an orthogonal range tree like structure for temporal comparison.

5.5 Bead Overlap Detection Query

The spatio-temporal range search starts from the root node of a DisCoUnt index; it utilises the pre-computed distance and temporal range information present the current node and its siblings *range_table* with respect to the query bound to prune the sub-trees. Please note, bead overlap detection through a spatio-temporal range search is accomplished by setting the querying each Bead of a susceptible moving object and looking for overlaps with an infectious moving object or vice versa.

This process is detailed in Algorithm 1.

The double-range search is an extension of the simplified algorithm described in Sect. 4.1 of the author of GNAT in [17] and re-framed by authors of EGNAT [52] in Algorithm 1. Since we added one additional computation per candidate solution and pivot point, it does not adversely effect the time complexity of search. Furthermore, Kimmo Fredriksson in their pre-print submission [51] states that for range-searching in a GNAT, for any surviving non-pruned pivot point without further child nodes, the search algorithm takes a $O(k)$ distance comparison in average, where k is the average degree. In our case it would be another $O(k)$ for the time overlap comparison, but given sufficiently larger k, the complexity of searching within a DisCoUnt node will still be $O(k)$.

Algorithm 1: Range Search to detect bead overlap

Input : P, query
Output: R : set of overlapping Beads

Let P be the list of root level DisCoUnt nodes
query contains both spatial and temporal bound to search
Randomly select a split point $p \in P$
Check if p spatio-temporally overlaps with *query*, then add to R
$d \leftarrow dist(query, p)$ $r \leftarrow query.spatial\_range$ $relative\_range \leftarrow [d - r, d + r]$
$t \leftarrow query.temporal\_range$ **while** $q \in P - p$ **do**
    if $relative\_range \cap range(p, D_q) = \phi$ **then**
        Remove q from P
                                                    // Space Pruning
    if $t \cap [q.t\_min, q.t\_max] = \phi$ **then**
        Remove q from P
                                                    // Time Pruning
Double Range Search for each D_p for p left in P
return R

5.6 Partitioning Strategies

To gauge the efficiency gain of the index, the following partitioning strategies were considered:

1. **Clustered:** Beads closest to a split point were assigned to the same index.
2. **Round-robin:** Beads that would have been part of the same partition in the clustered strategy are allocated across all available partitions in round-robin fashion.
3. **Random:** Random allocation of Beads for comparison.

5.7 Infection Risk Query Processing

For SCORE, we build two DisCoUnt instances, one for indexing uncertain trajectories and the other for indexing risk maps. Since every infection risk query has a implicit spatio-temporal range, of say a city or a town, depending upon the scope of the risk map, we decompose the query as follows:

Step 1 Gather all the trajectory Beads in the overarching query range of the relevant moving object id i as S_i.
Step 2 Gather all the risk map Beads in the same range S_inf.
Step 3 Pass the sets to infection risk generator to detect pair-wise overlap
Step 4 Compute probability of infection propagation for each overlap and return the result set to the user.

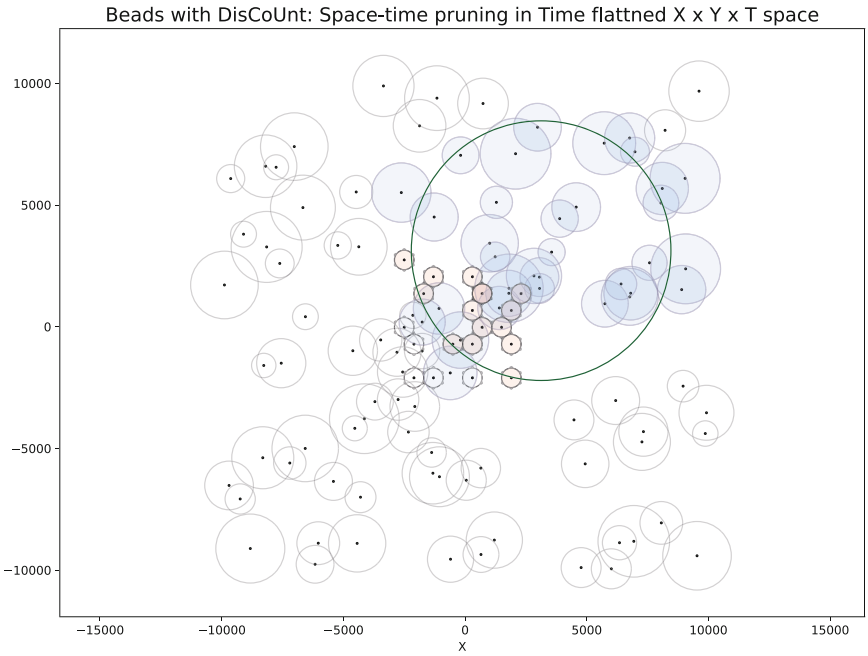

Fig. 12. Schematic diagram describing the gathering of trajectory Beads in blue and risk Beads in red within the scope of range query outlined in green (Color figure online)

A simplified spatial description of **Step 1** and **Step 2** is presented in Fig. 12, where the overarching spatial query range is outlined in green. The hexagonal tiled occupancy data has been approximated to its corresponding circle. The trajectory beads are constructed as Individual Bound Beads to encode movement with variable velocity in between trajectory observations. The result set S_i and S_inf are passed to the infection risk generator to detect pair-wise overlap. This is the proposed solution to **RQ3**.

6 Experimental Evaluation

In this section DisCoUnt index is evaluated, firstly in a single machine and secondly on a cluster of machines. We chose to perform these experiments on synthetic data for both trajectory and infection risk map. The synthetic data was generated to evaluate effects of varying *data distribution*, *data volume*, *data partitioning*, and *query selectivity* on the proposed index that we expect to find in real-world mobility data and risk maps.

6.1 Experiment Setup

The components of SCORE system and DisCoUnt index were implemented in Python 3. First, the experiments involving standalone DisCoUnt index were conducted on a single machine with Intel(R) Core(TM) i7-10875H CPU @ 2.3 GHz running Ubuntu 20.04.2 LTS x86_64-bit OS with 32 GB RAM. An uncertain trajectory data store was generated for testing by increasing data volumes from 1K to 1M Beads (Fig. 13).

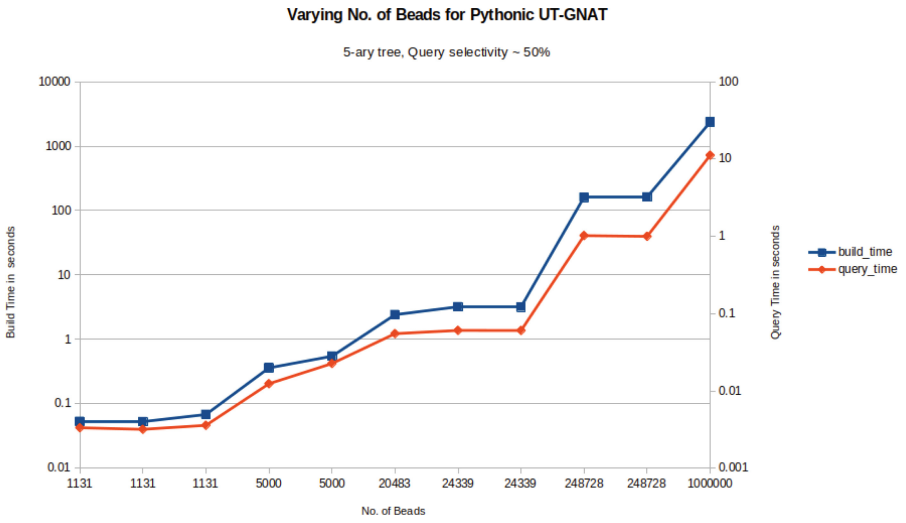

Fig. 13. Time complexity of single DisCoUnt index. Please note that there are separate logarithmic time axes for building and querying on the left and right side respectively

The results of searching in parallel DisCoUnts for partitioned data store in a local machine shows clear efficiency gain over a single index as shown in Fig. 14.

Second, a two instance DisCoUnt index was set-up over a partitioned bead store with an associated risk map of 20% of total number of trajectory Beads in a cluster of 4 machines with Intel(R) Core(TM) i5-2400 CPU @ 3.10 GHz running Ubuntu 14.04.5 LTS x86_64-bit OS with 8 GB RAM. The machines were networked via wired Ethernet TP-LINK(R) TL-SG1005D Gigabit Switch. Initial implementation of DisCoUnt was extended for processing of query in a distributed environment utilising Dask [37]. Specifically, the *Actor* API of Dask, which enables stateful computation at worker nodes. Please note that for the scope of this paper, we conclude our evaluation at gathering the relevant Beads from the partitioned store and do not proceed with computing the overlaps and the subsequent probabilities. Since the Infection risk measure calculation as

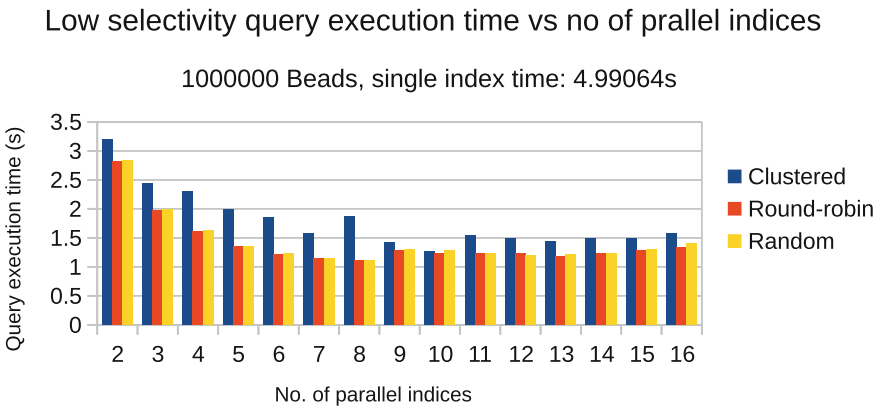

Fig. 14. Low selectivity, parallel query on 1000000 Beads showing expected speedup on increasing partitions

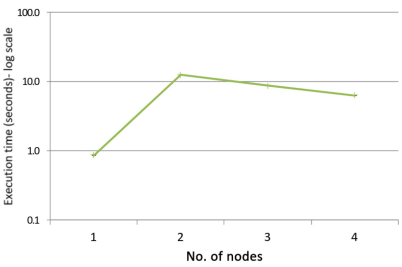

Fig. 15. Query time vs. number of nodes in cluster for 100K Beads

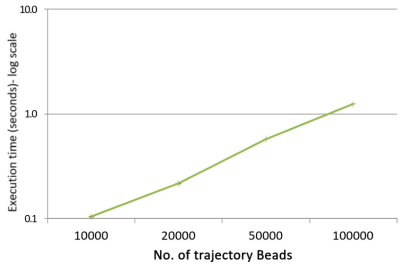

Fig. 16. Query time vs. No of Beads

described by Definition 3 is a sum of multiplications with a sampling of PRNG, it is expected to not exceed the delay of data transfer over network and affect the evaluation metrics.

6.2 Effect of Data Distribution

In this experiment, we vary the number of nodes as 1, 2, 3 and 4, and we report the corresponding query execution times in Fig. 15. The figure shows an increase in efficiency of query execution with an increase in the number of nodes. However, it is not clear how many nodes may be needed to outweigh the delays introduced due to data transmission over a network.

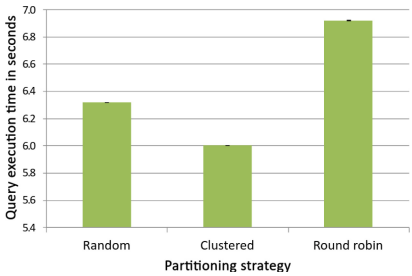

Fig. 17. Query time vs. Partitioning strategy for 100K Beads across 4 nodes

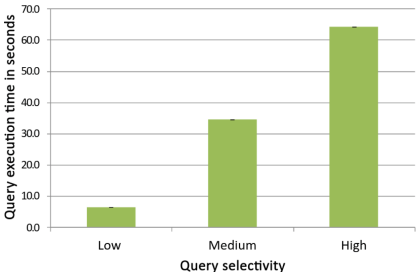

Fig. 18. Query time vs. Query selectivity of High (0.6), Medium (0.3) and Low (0.04) for 100K Beads across 4 nodes

6.3 Effect of Data Volume (Number of Beads)

We vary the number of beads from 10k to 100k and measure the query execution time. Figure 16 shows expected trend in increase in query execution time with respect to data volume, similar to that of a non-networked implementation of DisCoUnt.

6.4 Effect of Partitioning Heuristics

For this experiment, the trajectory Beads and the risk maps beads were partitioned separately with respect to two sets of split points. We evaluate the impact of the partitioning strategies (i.e. clustered, random and round robin, mentioned in Sect. 5.6) on the query execution time. The result in Fig. 17 suggests that the clustering partitioning scheme works slightly better than random allocation of Beads and the round-robin scheme performs worse.

6.5 Effect of Query Selectivity

In this experiment, we vary the query selectivity and evaluate its impact on the query execution time keeping the data volume constant at 100K Beads. We used three selectivity values: low, medium and high with selectivity of 0.04, 0.3 and 0.6 respectively. Therefore the query results of low medium and high selectivity would contain 400 Beads, 30K Beads and 60K Beads respectively. Figure 18 demonstrates that query execution time is linearly proportional to volume of result. This is to be expected for a networked environment of our experimental setup where over 90% of the total query execution time is spent transferring query result from worker nodes.

7 Discussions and Future Work

In this section we discuss the limitations of the experimental evaluation, the scope of SCORE, and future lines of enquiry.

- Our experiments on the distributed DisCoUnt index demonstrates that it has the potential to scale similar to its parallel variant as shown in Fig. 14. However, as shown in Fig. 15 we will need additional experiments to determine optimum multi-machine configuration.
- The experiments are performed on synthetic data that varies query load and data volume. Although it is representative of worst case scenarios, performing similar suite of experiments on real-world trajectory store remains as a future work.
- At a framework level, comparing the accuracy and efficiency of SCORE would require multiple data sources for a well-documented outbreak of a particular contagion in a specific community. It would also require a calibrated contagion simulator. Calibration of a contagion simulator is out of scope of this paper and by extension a full framework evaluation of SCORE remains as future work.

Please note, in this paper we focus on a "uncertain proximity detection" framework and the key definitions, algorithms, and data structures to compute it in a distributed context. This enables the proposed SCORE architecture to use outputs of any mechanistic or proximity based disease modelling system.

8 Related Work

The literature surveyed during the course of this research were found to focus on a part of the problems we aim to tackle through the SCORE architecture. In this section, we mention some of the relevant research.

Authors of [3–5] and [6] have utilised mobility data to produce outputs analogous to the proposed *Risk map*, however the spatial resolution of their analysis spans areas as large as entire cities and districts. Hence, these models do not incorporate individual level mobility. Alix et al. [21], also represents single POIs with multiple hexagonal tiles, as in SCORE. They also assign spatial dispersion within a POI with respect to different categories of POI. Similarly, Anastasiou et al. [7], also spatially divide a POI into square tiles rather than hexagonal tiles. They run micro-simulation of expected movement of individuals within a POI with respect to ingress, exists and other architectural features. They aggregate them to produce regions of high-congestion and assign predictive risk scores future trajectory plans. However, both risk measures only take into account the occupancy data of individuals and not the disease dynamics of the contagion question, nor the expected changes due to intervention measures.

An et al. [53] explore agent based contagion simulation but they consider their spatial boundaries to be at the scale of countries and the co-location data was inferred from synthetic contact networks instead of real-world mobility data. Lee et al. [54] propose an efficient graph transformation to simulate reactive and preventative vaccination strategies. Their spatial resolution was of an administrative division of a US state and their co-location graph of choice was of routers and P2P file sharing network. We believe that approaches like these would benefit from distributed index, DisCoUnt proposed by us. At the time of writing this

manuscript, we did not find any previous work that integrates calibrated individual level agent based contagion modelling using real-world mobility data, with uncertain trajectory analysis to assess infection risk as an overarching framework.

9 Conclusion

In this paper, we propose a holistic framework called SCORE that combines simulated disease propagation model with real-world trajectory data to facilitate infection risk analysis for a community. We formally define a probability measure of spatial contact given a pair of uncertain trajectory segments. We combine this probability of spatial contact with a mechanistic model of infection transmission to define the probability of infection risk. We propose a novel index DisCoUnt, to efficiently search and retrieve data points to calculate infection risk measure for a moving object given their trajectory. Finally, we demonstrate with our experimental results that ours is a viable approach for computing infection risk measure using a distributed data store. However, our approach is incumbent on the calibration of an agent based contagion simulator, which is a non-trivial task. Our future directions include the exploration of different partitioning techniques and conducting a comparative evaluation of alternate candidates for distributed spatio-temporal indices.

Acknowledgements. We thank Bilal Husain, Gaia Noseworthy, Jarod Kelly, and Robert Santacruz, along with Dane Sheppard of The Black Arcs for their inputs and inspiration. We also thank Suvam Kumar Das a member of our Big data systems and analytics lab for his insightful discussions on metric indices that helped in planning early research activities. Finally we thank the reviewers for their comments that improved this paper.

References

1. Zhang, X., Ray, S., Shoeleh, F., Lu, R.: Efficient contact similarity query over uncertain trajectories. In: EDBT, pp. 403–408 (2021)
2. Canada, P. Public health management of cases and contacts associated with COVID-19 - Canada.ca. (2021). https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals/interim-guidance-cases-contacts.html. Accessed 19 May 2023
3. Rambhatla, S., Zeighami, S., Shahabi, K., Shahabi, C., Liu, Y.: Toward accurate spatiotemporal COVID-19 risk scores using high-resolution real-world mobility data. ACM Trans. Spat. Algorithms Syst. (TSAS). **8**, 1–30 (2022)
4. Zeighami, S., Shahabi, C., Krumm, J.: Estimating spread of contact-based contagions in a population through sub-sampling. ArXiv Preprint ArXiv:2012.06987 (2020)
5. Kuddus, M., Rahman, A.: Analysis of COVID-19 using a modified SLIR model with nonlinear incidence. Results Phys. **27**, 104478 (2021)

6. Chiang, W., Liu, X., Mohler, G.: Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. Int. J. Forecast. **38**, 505–520 (2022)
7. Anastasiou, C., Costa, C., Chrysanthis, P., Shahabi, C., Zeinalipour-Yazti, D.: ASTRO: reducing COVID-19 exposure through contact prediction and avoidance. ACM Trans. Spat. Algorithms Syst. (TSAS) **8**, 1–31 (2021)
8. Cicala, F., et al.: PURE: a framework for analyzing proximity-based contact tracing protocols. ACM Comput. Surv. (CSUR) **55**, 1–36 (2021)
9. Pfoser, D., Jensen, C., Theodoridis, Y., et al.: Novel approaches to the indexing of moving object trajectories. In: VLDB, pp. 395–406 (2000)
10. Trajcevski, G.: Uncertainty in spatial trajectories. In: Zheng, Y., Zhou, X. (eds.) Computing With Spatial Trajectories, pp. 63–107. Springer, Cham (2011). https://doi.org/10.1007/978-1-4614-1629-6_3
11. Yuan, J., Zheng, Y., Xie, X., Sun, G.: Driving with knowledge from the physical world. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 316–324 (2011)
12. Ciaccia, P., Patella, M., Zezula, P.: M-tree: an efficient access method for similarity search in metric spaces. VLDB **97**, 426–435 (1997)
13. Chen, L., et al.: Indexing metric spaces for exact similarity search. ACM Comput. Surv. (CSUR) **55**, 1–39 (2020)
14. Bozkaya, T., Ozsoyoglu, M.: Distance-based indexing for high-dimensional metric spaces. In: Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data, pp. 357–368 (1997)
15. Chen, L., Gao, Y., Zheng, B., Jensen, C., Yang, H., Yang, K.: Pivot-based metric indexing. VLDB Endow. (2017)
16. Kalantari, I., McDonald, G.: A data structure and an algorithm for the nearest point problem. IEEE Trans. Softw. Eng. 631–634 (1983)
17. Brin, S.: Near neighbor search in large metric spaces. VLDB **95**, 574–584 (1995)
18. Mahmood, I., et al.: FACS: a geospatial agent-based simulator for analysing COVID-19 spread and public health measures on local regions. J. Simul. **16**, 355–373 (2022)
19. Horni, A., Nagel, K., Axhausen, K.: The Multi-Agent Transport Simulation MATSim. Ubiquity Press (2016)
20. Müller, S., et al.: Predicting the effects of COVID-19 related interventions in urban settings by combining activity-based modelling, agent-based simulation, and mobile phone data. PLoS ONE **16**, e0259037 (2021)
21. Alix, G., Yanin, N., Pechlivanoglou, T., Li, J., Heidari, F., Papagelis, M.: A mobility-based recommendation system for mitigating the risk of infection during epidemics. In: 2022 23rd IEEE International Conference on Mobile Data Management (MDM), pp. 292–295 (2022)
22. Dong, E., Du, H., Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. **20**, 533–534 (2020). https://www.sciencedirect.com/science/article/pii/S1473309920301201
23. Pritchard, E., et al.: Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. Nat. Med. **27**, 1370–1378 (2021)
24. Makhoul, M., et al.: Epidemiological impact of SARS-CoV-2 vaccination: mathematical modeling analyses. Vaccines **8**, 668 (2020)
25. Layan, M., et al.: Impact of BNT162b2 vaccination and isolation on SARS-CoV-2 transmission in Israeli households: an observational study. Am. J. Epidemiol. **191**, 1224–1234 (2022)

26. Bian, L., et al.: Effects of SARS-CoV-2 variants on vaccine efficacy and response strategies. Expert Rev. Vaccines **20**, 365–373 (2021)
27. Glasser, J., Feng, Z., Omer, S., Smith, P., Rodewald, L.: The effect of heterogeneity in uptake of the measles, mumps, and rubella vaccine on the potential for outbreaks of measles: a modelling study. Lancet. Infect. Dis. **16**, 599–605 (2016)
28. Killingley, B., et al.: Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. Nat. Med. **28**, 1031–1041 (2022)
29. Klomjit, N., et al.: COVID-19 vaccination and glomerulonephritis. Kidney Int. Rep. **6**, 2969–2978 (2021)
30. Puhach, O., et al.: Infectious viral load in unvaccinated and vaccinated individuals infected with ancestral, Delta or Omicron SARS-CoV-2. Nat. Med. 1–10 (2022)
31. Abueg, M., et al.: Modeling the effect of exposure notification and non-pharmaceutical interventions on COVID-19 transmission in Washington state. NPJ Digit. Med. **4**, 1–10 (2021)
32. Hens, N., Vranck, P., Molenberghs, G.: The COVID-19 epidemic, its mortality, and the role of non-pharmaceutical interventions. Eur. Heart J. Acute Cardiovasc. Care **9**, 204–208 (2020)
33. Lai, S., et al.: Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature **585**, 410–413 (2020)
34. Cheng, V., et al.: The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. J. Infect. **81**, 107–114 (2020)
35. Souza Santos, A., et al.: Dataset on SARS-CoV-2 non-pharmaceutical interventions in Brazilian municipalities. Sci. Data **8**, 1–6 (2021)
36. Ivanov JSON to Tree Diagram Converter (2014). https://github.com/ivan111/vtree. Accessed 10 Nov 2022
37. Dask Development Team Dask: Library for dynamic task scheduling (2016). https://dask.org
38. Mistry, A.: Dynamic contagion spread modelling over emergent spatio-temporal contact networks. Masters thesis (2022)
39. Groot, R., et al.: Middle east respiratory syndrome coronavirus (MERS-CoV): announcement of the coronavirus study group. J. Virol. **87**, 7790–7792 (2013)
40. Great Britain. Magna Carta, 1215 June 15 [manuscript] (1215)
41. Kourtis, A., Appelgren, K., Chevalier, M., McElroy, A.: Ebola virus disease: focus on children. Pediatr. Infect. Dis J. **34**, 893–897 (2015)
42. Noorbakhsh, F., et al.: Zika virus infection, basic and clinical aspects: a review article. Iran. J. Public Health **48**, 20–31 (2019)
43. Marani, M., Katul, G., Pan, W., Anthony J.: Parolari intensity and frequency of extreme novel epidemics. Proc. Natl. Acad. Sci. **118**, e2105482118 (2021). https://www.pnas.org/doi/abs/10.1073/pnas.2105482118
44. Glatter, K., Finkelman, P.: History of the Plague: an ancient pandemic for the age of COVID-19. Am. J. Med. **134**, 176–181 (2020)
45. Rambaut, A., et al.: A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. **5**, 1403–1407 (2020)
46. Miller, E., Roorda, M.: Prototype model of household activity-travel scheduling. Transp. Res. Rec. **1831**, 114–121 (2003)
47. Roorda, M., Miller, E., Habib, K.: Validation of TASHA: a 24-h activity scheduling microsimulation model. Transp. Res. Part A: Policy Pract. **42**, 360–375 (2008)
48. Keeling, M., Eames, K.: Networks and epidemic models. J. R. Soc. Interface **2**, 295–307 (2005)

49. Challenger, J., et al.: Modelling upper respiratory viral load dynamics of SARS-CoV-2. BMC Med. **20**, 1–20 (2022)
50. Bentley, J., Friedman, J.: Data structures for range searching. ACM Comput. Surv. (CSUR) **11**, 397–409 (1979)
51. Fredriksson, K.: Geometric near-neighbor access tree (GNAT) revisited. ArXiv Preprint ArXiv:1605.05944 (2016)
52. Navarro, G., Uribe-Paredes, R.: Fully dynamic metric access methods based on hyperplane partitioning. Inf. Syst. **36**, 734–747 (2011)
53. An, T., Kim, H., Joo, C.: Prediction of COVID-19 infection spread through agent-based simulation. In: Proceedings of the Twenty-Third International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing, pp. 247–252 (2022). https://doi.org/10.1145/3492866.3557735
54. Lee, C., Tenneti, S., Eun, D.: Transient dynamics of epidemic spreading and its mitigation on large networks. In: Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 191–200 (2019). https://doi.org/10.1145/3323679.3326517
55. WHO COVID-19 Dashboard. Geneva: World Health Organization (2020). https://covid19.who.int/. Accessed 30 Sept 2023
56. The Black Arcs Inc. Citisketch (2018). https://blackarcs.org/. Accessed 18 Oct 2022
57. Government of Canada. Statistics Canada: Canada's national statistical agency (2022). https://www.statcan.gc.ca/en/start. Accessed 18 Oct 2022
58. OpenStreetMap contributors. Planet dump (2017). https://planet.osm.org