# Optimal Risk-aware POI Recommendations during Epidemics

Nina Yanin and Manos Papagelis
{nina27@my.,papaggel@eecs.}yorku.ca
York University – Toronto, Ontario, Canada

## ABSTRACT

The movement of people can influence the spread of diseases, especially in populated areas. While measures like quarantine can curb disease spread by restricting the movement of those infected, they come with socioeconomic consequences. Furthermore, not everyone might adhere to these restrictions, undermining their effectiveness. A more effective strategy is to educate people on the risks tied to their movement habits and recommend safer options. In this research, we introduce the problem of *optimal risk-aware point-of-interest (POI) recommendations during epidemics*, where people get recommendations on what POI to visit that reduces the risk of getting infected. The risk of infection at a POI is modeled based on its capacity and visit patterns over time. Then, we present a method that provides personalized recommendations which, when universally adopted, the overall risk is minimized. Unlike existing strategies, our method considers simultaneous user requests made in the same time period, which might influence the relative risk at POIs. An extensive evaluation was conducted, using *real-world data coming from three major cities in Canada*, which showed that our method outperforms the current state of practice method and other sensible baselines, on varying settings. Specifically, our method presented a decrease in the relative added risk of infection by 99.87%, 71.56% and 61.54% at each city, respectively. We also examined how the optimal solution is impacted if only a specific portion of the population follows the recommendation. Our optimal risk-aware recommendation method has the potential to reduce infection risk by promoting responsible behaviors within communities.

## CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computing methodologies** → **Modeling and simulation**; • **Applied computing** → **Health informatics**.

## KEYWORDS

human mobility, POI recommendation, optimization, epidemics

**ACM Reference Format:**
Nina Yanin and Manos Papagelis. 2023. Optimal Risk-aware POI Recommendations during Epidemics. In *4th ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi '23), November 13, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3615898.3628256


*SpatialEpi '23, November 13, 2023, Hamburg, Germany*

ACM ISBN 979-8-4007-0360-7/23/11…$15.00
https://doi.org/10.1145/3615898.3628256


**Figure 1: An illustrative example of how different models form POI recommendations. The *black pins* show all the POI options available to two users within their search radius (RANDOM); the *pins bounded by a blue square* represent the POIs closest to each user (CLOSEST); the *pin bounded by an orange circle* represents the least risky POI at that time (LOCAL); the *pins bounded by a green diamond* represent our method (GLOBAL). If both people follow LOCAL, then they will visit the same place at the same time, leading to suboptimal results (i.e., higher risk), as opposed to following our globally optimal risk-aware recommendation model GLOBAL.**

## 1 INTRODUCTION

**Motivation & the problem of interest.** Human mobility can significantly affect the spread of epidemics. When people move from one place to another, they can carry the pathogen with them and introduce it to new populations. Densely populated areas and crowded places are associated with higher risk of transmission that can lead to outbreaks and the rapid spread of the disease. Mobility can also play a role in containing epidemics. For instance, quarantine and isolation measures can be implemented by controlling the movement of infected individuals and preventing further spread of the disease [13, 30]. However, such policies have significant drawbacks — they can lead to decreased economic activity, they can have a psychological impact, and they can infringe the privacy of individuals and our civil liberties. They can also be ineffective, as individuals might not conform with them. A better approach is to inform people about the risks associated with their mobility patterns, present them with alternatives, allow them to weight their choices and help them to make more informed decisions. These observations, along with the advancement of location tracking technologies [17, 28, 38], have highlighted the need for location-based services (or mobile apps) that can suggest points of interest (POI) with reduced risk of infection. The task is to provide personalized recommendations that minimize the overall risk of infection.

**Current approaches & their limitations.** Current approaches have been based on the assumption that individuals can reduce their risk of contracting an infection by opting for less crowded locations. Therefore, given as input their current location and a search radius, they provide a ranked list of the safest POIs. An illustrative example is shown in Figure 1, where multiple users, located at different locations, submit queries about POIs in their vicinity (defined by a radius parameter) that satisfy specific risk-aware preferences. The safety of a POI is determined by mining historical visit patterns, and providing predictions about a POI's occupancy at the time of visit [3, 31], or by mining historical real-world check-in datasets [19]. A critical limitation of this approach is that it does not consider concurrent user requests, requests made by different people within the same geographic area and time period. Concurrent queries may lead to suboptimal results as multiple individuals might end up at the same place, at the same time. The problem can be more stressed in busy urban areas, where suboptimal recommendations might end up turning supposedly safe POIs to risky ones. The mere consequence of this is that current methods fail to minimize the risk for certain people, and the community as a whole.

**Our approach & contributions.** In contrast to this local optimization strategy, our proposed solution follows a global optimization strategy by thoroughly exploring the entire solution space and recommending POIs to individuals that minimize the overall risk of the community. A summary of our contributions is provided below:

- We present a model that incorporates mobility data to assess the risk of infection at POIs. The risk model factors in (i) the expected occupancy of a POI at a certain time using historical data, and (ii) the expected number of visits in the near future due to real-time recommendations.
- We introduce the novel optimization problem of risk-aware POI recommendations during epidemics. We formally characterize the computational complexity of the problem.
- We model the optimization problem as a *generalized assignment problem* (GAP) where individuals are assigned to POIs. We solve the problem using a *mixed-integer quadratic programming* (MIQP) solver. Compared to a locally optimal recommendation method (LOCAL), our globally optimal recommendation method (GLOBAL) minimizes the overall risk of infection for all individuals.
- We present a comprehensive experimental evaluation that demonstrates the superiority of GLOBAL against state of practice and other sensible baselines. For instance, we witness that GLOBAL yields a 99.87%, 71.56% and 61.54% improvement compared to LOCAL, in the relative added risk of infection for Toronto, Montreal, and Calgary, respectively.
- We make our source code publicly available to encourage the reproducibility of our work[1].

**Paper organization.** The remainder of the paper is organized as follows: section 2 provides preliminaries and a formal definition of the problem. Section 3 presents our proposed recommendation model and scientific approach. Our experimental evaluation and a discussion of the results are presented in section 4. We review the related work in section 5, and conclude in section 6.

[1]https://github.com/Nina1234y/RiskAwarePOIRecommender

# 2 PRELIMINARIES AND PROBLEM DEFINITION

In this section, we briefly introduce definitions and notations (refer to Table 1). Then, we formally define the problem of interest, present its *computational complexity* and characterize its *hardness*.

## 2.1 Preliminaries

**Definition 1 (Map)**. Let $\mathcal{M}$ be a map over a predefined, finite, and continuous geographical area.

**Definition 2 (Block)**. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{|\mathcal{B}|}\}$ be a set of disjoint hexagonal shaped blocks that fully tessellate map $\mathcal{M}$. All $\mathbf{b}_k \in \mathcal{B}$ are polygons with the same arbitrary area $X$.

**Definition 3 (POI)**. Let $\mathcal{P} = \{p_1, p_2, \ldots, p_{|\mathcal{P}|}\}$ be a set of points of interests on a map $\mathcal{M}$. Each $p_j \in \mathcal{P}$ has a a surface area $a_j$, average dwell time of $adt_j$, and an expected, time-varying occupancy $o_j^t$ at time $t$. Due to the tessellation of the map $\mathcal{M}$ to blocks, each $p_j$ consists of $\lceil \frac{a_j}{X} \rceil$ finite disjoint hexagonal blocks of area $X$. Furthermore, we assume availability of surface areas and average dwell-times, which are generally available by third-party data providers, as discussed in section 4. In addition, we assume a taxonomy of POIs, such that individuals can query for a certain POI type, including restaurants, grocery stores, pharmacies, and so on.

**Definition 4 (Contact)**. Let $n$ be the number of individuals in the same POI and in the same block $b_k \in \mathcal{B}$. These individuals form contacts with each other, which can be represented as the number of edges in a complete graph $n(n-1)/2$. As individuals move around over time, this value changes accordingly. Let $\Gamma_j^t$ be the number of contacts at time $t$ for POI $p_j$. Let $o_j^t$ be the number of people in $p_j$ at time $t$. As previously defined, $p_j$ consists of $\lceil \frac{a_j}{X} \rceil$ hexagons. We make the simplifying assumption that people are uniformly distributed in these blocks. As such, the number of people per block at time $t$ can be represented as $n = \lfloor \frac{o_j^t X}{a_j} \rfloor$. Finally, we formally present the number of contacts in a block at $p_j$:

$$\Gamma_j^t = \frac{\frac{o_j^t X}{a_j}\left(\frac{o_j^t X}{a_j} - 1\right)}{2}, \quad \forall j \in 1 \ldots |\mathcal{P}| \tag{1}$$

**Definition 5 (Invocation frequency)**. Let invocation frequency, $\eta$, be the rate at which a model is invoked. As queries arrive, they wait to be processed as a batch in the next invocation of the model. A lower $\eta$ provides more information to the model, i.e., more concurrent queries, which justifies the benefit of optimization. However, as the size of the problem increases, the model's run-time cost increases as well. We further investigate this trade-off in section 4.

## 2.2 Problem Definition

**Problem 1 (Optimal risk-aware POI recommendation)**. Let a map $\mathcal{M}$, a tessellation of it by blocks in $\mathcal{B}$, and a set of POIs $\mathcal{P}$. Also, let a set of user queries $\mathcal{N} = \{< s_1, r_1 >, \ldots, < s_n, r_n >\}$ over a period $t \in [t_s, t_e]$, where $s_i$ is the source location of user $i$, $r_i$ is the search radius applied by user $i$, and $t_s$, $t_e$ are the start and end time of the sample period, respectively. Given $\mathcal{M}$, $\mathcal{B}$, $\mathcal{P}$, and $\mathcal{N}$, find the top-$K$ POIs to recommend to each user, such that the overall risk at all POIs is minimized. POI risk is defined in section 3.1.

**Table 1: Summary of Notations**

| Symbol | Description |
|---|---|
| $\mathcal{M}$ | enclosed geographical map |
| $t_s, t_e$ | start and end time of the sample period (in seconds) |
| $\mathcal{N}^{\Delta t}$ | a set of queries during duration $\Delta t$ |
| $\mathcal{P}$ | a set of POIs $\{p_1, p_2, \ldots, p_{\lvert\mathcal{P}\rvert}\}$ at $\mathcal{M}$ |
| $i, j, k$ | $i \in [0, \lvert\mathcal{N}\rvert]$, $j \in [0, \lvert\mathcal{P}\rvert]$, $k \in [0, K]$ |
| $o_j^{t_e}$ | occupancy at $p_j$ at time $t_e$ |
| $\overline{o}_j^{ts}$ | average occupancy of $p_j$ over $t \in [ts, ts + adt_j]$ |
| $\mathcal{B}$ | a set of hexagonal blocks |
| $X$ | recommended social distancing area in $m^2$ |
| $a_j$ | area of $p_j$ in $m^2$ |
| $d_{ij}$ | eligibility of user $i$ to go to $p_j$ |
| $y_j$ | number of users assigned to $p_j$ |
| $S_{ij}$ | Selection likelihood matrix |
| $I_i$ | number of recommendation user $i$ gets |
| $Q_{ijh}$ | assign user $i$ to $p_j$ at rank $h$ |
| $risk_j^t$ | risk of infection of $p_j$ at time $t$ |
| $s_i$ | source coordinate for user $i$ |
| $r_i$ | search radius in Km for user $i$ |
| $K$ | number of POI recommendation |
| $adt_j^t$ | average dwell time of $p_j$ at time $t$ |
| $arisk_j^t$ | added risk at $p_j$ in time $t$ |
| $\eta$ | invocation frequency |
| $\Gamma_j^t$ | number of contacts at POI $p_j$ at time $t$ |

## 2.3 Complexity Analysis

In this section, we evaluate the complexity of concurrent POI recommendations and show that the problem is NP-hard. Let $\mathcal{N}$ be a set of queries, $\mathcal{P}$ be a set of POIs, and $K$ be the number of recommendations each user receives. We want to minimize a specific utility function that is affected by the size of these three components: $\mathcal{N}$, $\mathcal{P}$ and $K$. The number of possibilities are a $K$-permutation of $|\mathcal{P}|$ POIs per user (i.e., $O(P_{|\mathcal{P}|,K})$), to the power of $\mathcal{N}$ users, bringing the overall time and space complexity to $O((P_{|\mathcal{P}|,K})^{|N|})$.

*2.3.1 Problem Hardness.* We establish the hardness of our problem by showing it is at least as hard as the *Generalized Assignment Problem* problem, which is a known NP-complete problem [29].

Theorem 2.1. *Problem 1 (optimal risk-aware POI recommendation) is NP-hard.*

Proof. We prove the claim by reducing the GAP problem to our problem. The GAP is defined as follows. Let $i \in 1 \ldots n$ be a set of workers, $j \in 1 \ldots m$ be a set of tasks, $b_i \in B^n$ be the total work capacity of worker $i$, $r_{ij} \in R^{n \times m}$ be the weight of task $j$ if assigned to worker $i$, $h_j \in H^m$ be the total number of workers needed for task $j$. It can be set to $h_j \leq n$ to indicate that anywhere from 0 to $n$ workers can be assigned to task $j$. Finally, let $c_{ij} \in C^{n \times m}$ be the cost associated to assign worker $i$ to task $j$. The GAP is to determine an optimal assignment of workers to tasks, where each worker has a fixed capacity. Given an instance of GAP, we build an instance of our problem as follows. Let $i \in \mathcal{N}$ be a set of users that correspond to the $n$ set of workers, and let $j \in \mathcal{P}$ be a set of POIs that correspond to the $m$ set of tasks. Then, let $d_{ij} \in D^{|\mathcal{N}| \times |\mathcal{P}|}$ correspond to $r_{ij} \in R^{n \times m}$. We set $d_{ij}$ to be a binary matrix that shows the feasibility of user $i$ to go to $p_j$. Let $l_i \in min(\sum_{j \in \mathcal{P}} d_{ij}, K)$ be a set of the POI recommendations user $i$ gets and correspond to $b_i \in B^n$. To clarify, this capacity constraint ensures that every user can go to a maximum of up to $K$ POIs. We know this constraint prior to running the algorithm, and its value remains constant, as defined by Eq. 7 in the manuscript. We let $g_j \leq |\mathcal{N}|$ correspond to $h_j \leq n$ to indicate that POIs have no capacity restrictions/ can accommodate all users. Let $y_j$ be the number of people assigned to $p_j$ and assume the number of recommendations $K$ a user gets is 1 in order to view the problem in a two dimensional space. Finally, let $c_{ij} \in C^{n \times m}$ be the cost associated to assign user $i$ to $p_j$. This reduction is polynomial-time and shows that solving our problem is at least as hard as solving the GAP, which is NP-hard, implying that our problem is also NP-hard. This concludes our proof. □

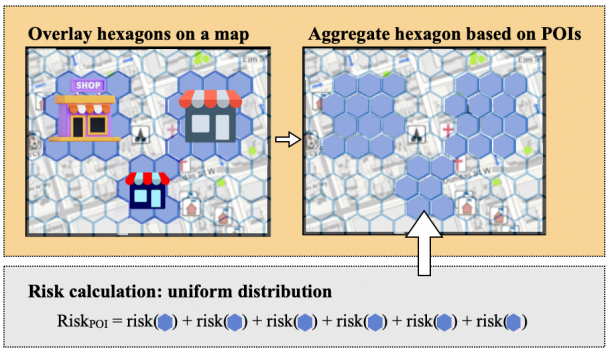



**Figure 2: Architecture diagram for POI risk calculation.**

# 3 METHODOLOGY

In this section, we first present information related to POI modeling. Then, we present two risk-aware recommendation models that make use of this information.

## 3.1 Modeling POI Information

We first present an extrapolation method to model POI occupancy as a function of a POI's extrapolated temporal occupancy and average dwell-time for a given timestamp $t$. Second, we present a method for computing the risk associated with each block due to the number of individuals (and contacts) in it. Moreover, we use this block risk to define a POI's **added risk** due to **added users**. Third, we model a selection likelihood function that associates each rank of a recommended POI with a specific weight. This weight represents the likelihood a user will commit to a POI at that level of risk. Finally, we present a user-POI eligibility matrix that shows which POIs are accessible to which users. It is worth mentioning that we treat all POI categories uniformly, without considering their size, location, or function. Additionally, any hierarchical structure of POIs, where one POI might encompass another (e.g., a multi-store shopping mall), is out of the scope of this study. This limitation stems from the data collection process, which is beyond our influence.

**POI occupancy estimation**. In this work, we assume that past human mobility traces and average dwell-times are available (see details on data in section 4). When a recommendation model considers potential POIs, it needs to take into account the expected

duration a user will stay there based on $adt_j$ of POI $p_j$. More specifically, it needs to consider the change in risk over the dwell-time, and not only the risk at the time of query. Therefore, we model POI occupancy as a function of POIs' extrapolated temporal occupancies and average dwell-times. Let $\overline{o}_j^{ts}$ be the average occupancy of $p_j$ over the duration $t \in [ts, ts + adt_j]$. Formally, it is:

$$\overline{o}_j^{ts} = \frac{\sum_{t=ts}^{ts+adt_j} o_j^t}{adt_j}, \quad \forall j \in 1 \dots |\mathcal{P}| \tag{2}$$

**Modeling POI infection risk**. Let $risk_j^t$ be the risk of infection at POI $p_j$ at time $t$. The risk can be represented as the number of pairwise contacts of all individuals in a hexagon $b_j$. Fig. 2 illustrates the risk calculation process for each POI. This risk definition is the same as the one presented in Eq. 1: $risk_j^t = \Gamma_j^t$.

To normalize the POI risks (i.e., range between zero and one), we define the *relative risk*, $rrisk_j^t$, of POI $p_j$ at time $t$. Formally, it is:

$$rrisk_j^t = \frac{risk_j^t}{\sum_{w=1}^{|\mathcal{P}|} risk_w^t}, \quad \forall j \in 1 \dots |\mathcal{P}| \tag{3}$$

**Modeling POI added risk**. To properly model the risk of a POI, in addition to historical visit patterns, we need to consider any visits to the POI in the near future that can be attributed to real-time recommendations. Building on Eq. 1, an addition of $y_j$ users to POI $p_j$ at time $t_s$ contributes an additional risk to $p_j$ that is equivalent to the POI's risk before and after the addition. Formally, it is:

$$arisk_j^t = \frac{(\overline{o}_j^t + y_j)X}{2a_j}\left(\frac{(\overline{o}_j^t + y_j)X}{2a_j} - \frac{1}{2}\right) - risk_j^t, \quad \forall j \in 1 \dots |\mathcal{P}| \tag{4}$$

To normalize the added risks across all POIs, we define the *relative added risk* $rarisk_j^t$ of POI $p_j$ at time $t$. Formally, it is:

$$rarisk_j^t = \frac{arisk_j^t}{\sum_{w=1}^{|\mathcal{P}|} arisk_w^t}, \quad \forall j \in 1 \dots |\mathcal{P}| \tag{5}$$

**Modeling POI selection likelihood**. Given a recommendation list, we do not know which one the end user will choose. Despite that, we make the assumption that order matters; thus, users have the highest likelihood to select the least risky recommended POI, represented by rank 1, and the lowest likelihood to select the most risky recommended POI, represented by rank $K$. Additionally, we want the sum of all users' $K$ weights to be equal to 1 in order to illustrate a complete selection. Therefore, we model the user selection likelihood as an exponential decay function, where $K$ represents the maximum number of recommendations a user gets, and $k$ represents the rank of the recommended POI. Formally:

$$f(k) = \begin{cases} \frac{(1-\frac{1}{K})^{Kk}}{\sum_{w=1}^{K}(1-\frac{1}{K})^{Kw}} & \forall k \in [1, K], \quad K > 1 \\ K & K \in \{0, 1\} \end{cases} \tag{6}$$

To reduce computation space and accommodate users whose number of POI options is less than $K$, we compute a matrix $\mathcal{S}^{(K+1)\times K}$, where each row corresponds to the number of POI recommendations (i.e., row $h \in [0, K]$ produces $h$ recommendations). Moreover, $\mathcal{S}$ holds unique weight distributions per each row following Eq. 6, where column $c$ corresponds to the rank and holds the weighted choice $f(c)$. This process is shown in algorithm 1.

**Algorithm 1:** Computing POI selection likelihood by rank

**Input:** $K$: Maximum number of POI recommendations
**Output:** $S$: a matrix of length $K + 1$ over $K$

```
S ← []
for k ∈ [0, K] do
    if k = 0 then
        S.append([0] * K)
    if k = 1 then
        t ← [0] * K
        t[0] ← 1
        S.append(t)
    else
        t ← [0] * K
        sum-val ← Σ_{q=1}^{k} (1 − 1/k)^{k*q}
        for q ∈ [1, k] do
            w ← (1 − 1/k)^{k*q}
            t[q − 1] ← w / sum-val
        S.append(t)
Return S
```

Let $\mathcal{I}^{\mathcal{N}}$ be a vector of size $\mathcal{N}$, where each index $i \in \mathcal{N}$ corresponds to a unique user id. The value of the cell contains either the number of candidate POIs of user $i$ or the number $K$, whichever is smaller. This is depicted in Eq. 7. As such, $I[i]$ holds an integer value representing the number of POI recommendations user $i$ gets, and $\mathcal{S}[\mathcal{I}[i]]$ is a vector containing the ordered POI weight distributions of user $i$ based on the number of recommendation they receive.

$$I_i = min(d_i, K) \tag{7}$$

**Eligibility matrix**. Let $d^{|\mathcal{N}|\times|\mathcal{P}|}$ be a binary eligibility matrix, where each row corresponds to user $i$ and each column corresponds to POI $p_j$. When $d_{ij}$ equals one, it means that $p_j$ is a candidate to user $i$. In this case, the candidate POI satisfies the constraints of the users (i.e., search radius and POI type). We pre-compute and use this matrix in the GLOBAL method described in section 3.2.

## 3.2 Risk-aware POI Recommendation Methods

We present two risk-aware POI recommendation methods. The first one, LOCAL, performs a local optimization by recommending the safest POI to every individual. This method is agnostic to concurrent queries in the vicinity, and therefore can lead to a suboptimal solution. The second one, GLOBAL, performs a global optimization by recommending POIs that minimize the overall risk of infection to the overall community, taking into account concurrent queries.

**LOCAL**. The current state of practice way to solve this problem is by assigning to each user the safest POIs ordered by risk, without taking into account concurrent queries that cover the same POIs. Formally, given $\mathcal{N}$ queries over duration $t \in [t_s, t_e]$ and $\mathcal{P}$ POIs, each person $i \in \mathcal{N}$ gets *I[i]* safest POIs within their search radius $r_i$. Algorithm 2 shows the recommendation process for user $i$.

**GLOBAL**. We want to expand on the LOCAL model by considering concurrent user queries in order to minimize the *rrisk*. We treat this

**Algorithm 2:** LOCAL risk-aware recommendation method

**Input:** $\overline{o}$: POI average occupancy vector of size $|\mathcal{P}|$,
$a$: a vector of size $|\mathcal{P}|$ that represents POI area,
$X$: a constant that represents block size,
$I$: a vector of length $|\mathcal{N}|$,
$S$: a matrix of length $K+1$ over $K$,
$i$: user's unique index

**Output:** up-to K recommendations for user $i$

$risk \leftarrow \frac{\frac{\overline{o}X}{a}(\frac{\overline{o}X}{a}-1)}{2}$ // compute risk for each POI

$id \leftarrow 1..|P|$ // numerate POI id

$p \leftarrow [id, risk].transpose()$ // create lookup by id

$p \leftarrow p.sortby(col = 1, ascending = True)$ // sortby risk

Return $p.transpose()[:, 1][I[i]]$

problem as a variation to the GAP, a *quadratic mixed integer ordered many-to-many assignment problem.* Additionally, we care about the order of recommendations and want to apply the selection likelihood in Eq. 6 to each rank accordingly. Formally, given $|\mathcal{N}|$ queries over duration $t \in [t_s, t_e]$ and $|\mathcal{P}|$ POIs, each person $i$ gets up to $I[i]$ ranked POI recommendations that minimize the objective function (*rarisk*) in Eq. 8, such that the following constraints are satisfied: (**i**) the number of recommendations is restricted by either $K$ or the total number of feasible options for user $i$, whichever is smaller (Eq. 9); (**ii**) the number of users assigned to each POI along with their weights is constrained (Eq. 10); (**iii**) the POI recommendations to each user $i$ are unique (Eq. 11); and (**iv**) for each user $i$, different POIs cannot be assigned the same rank (Eq. 12). The details of each variable is summarized in Table 1.

$$min \quad \sum_{j=1}^{|\mathcal{P}|} \frac{X^2}{2a_j^2}[(\overline{o}_j^t + y_j)(\overline{o}_j^t + y_j - \frac{a_j}{X^2}) - \overline{o}_j^t(\overline{o}_j^t - \frac{a_j}{X^2})] \tag{8}$$

$$s.t. \quad \sum_{j=1}^{|\mathcal{P}|}\sum_{h=1}^{K} Q_{ijh}d_{ij} = min(\sum_{j} d_{ij}, K), \forall i \in [1, |\mathcal{N}|] \tag{9}$$

$$y_j = \sum_{i=1}^{|\mathcal{N}|}\sum_{h=1}^{K} Q_{ijh}d_{ij}S_{I_i,h}, \forall j \in [1, |\mathcal{P}|] \tag{10}$$

$$\sum_{h=1}^{K} Q_{ijh} \leq 1, \forall i \in [1, |\mathcal{N}|], \forall j \in [1, |\mathcal{P}|] \tag{11}$$

$$\sum_{j=1}^{|\mathcal{P}|}\sum_{h=1}^{K} Q_{ijh}d_{ij}S_{I_i,h} = 1, \forall i \in [1, |\mathcal{N}|] \tag{12}$$

$$Q_{ijh} \in \{0,1\}, \forall i \in [1, |\mathcal{N}|], \forall j \in [1, |\mathcal{P}|], \forall h \in [1, K|] \tag{13}$$

$$y_j \in \mathbb{N}_{\geq 0}, \forall j \in [1, |\mathcal{P}|] \tag{14}$$

# 4 EXPERIMENTAL EVALUATION

In this section, we present a comprehensive experimental evaluation of our model. We begin by listing the research questions we aim to explore. Then, we present details of the data employed and the methods to be evaluated. Finally, we present the results and discuss broader insights. Table 1 lists the notation used in the experiments.

**Table 2: Summary of datasets**

| city | # POIs | # queries | # blocks | # visits |
|---|---|---|---|---|
| Toronto | 905 | 119,799 | 187,996 | 583,758 |
| Montreal | 536 | 78,771 | 79,675 | 388,525 |
| Calgary | 743 | 100,047 | 139,050 | 636,355 |

## 4.1 Experimental Scenarios

Our experiments aim to answer the following questions:

**(Q1) Effect of recommendation model on risk**. How different recommendation models, with different compliance levels, affect the relative added risk of infection at POIs?

**(Q2) Effect of different invocation frequencies on the relative added risk of infection at POIs**. How different invocation frequencies affect the relative added risk of infection at POIs?

**(Q3) Effect of $K$ on relative added risk**. How does $K$ affect the relative risk of infection at POIs?

**(Q4) Effect of different invocation frequencies on model run-time**. How different invocation frequencies affect the run-time of our proposed model?

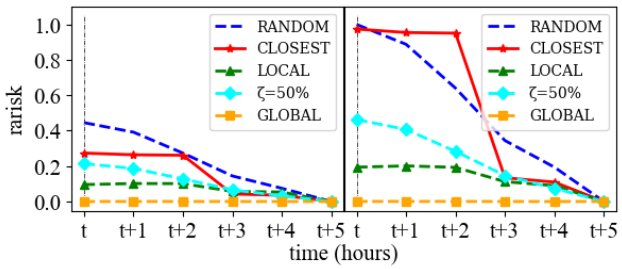
**(Q5) Effect of $K$ on the run-time of our model**. How does $K$ affect the run-time of our proposed model?

**(Q6) Effect of $\mathcal{N}$, $\mathcal{P}$ and $K$ on computation time**. How does $\mathcal{N}$, $\mathcal{P}$ and $K$ individually affect the run-time of GLOBAL model?

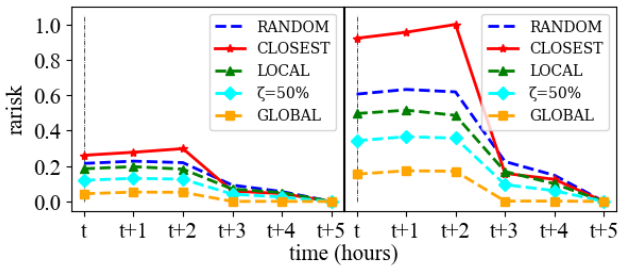
## 4.2 Datasets & Computation Environment

**Visit pattern extraction**. Our methods assume availability of human mobility data, POI geometry, and average dwell-time per POI. Collecting such data is out of the scope of the current paper, but several datasets are available for research purposes. In our evaluation, we use SafeGraph's *geometry* and *patterns* datasets to get POI real-world temporal visit patterns, POI average dwell-times and POI areas for three major cities in Canada: *Toronto*, *Montreal* and *Calgary*. We want to recommend POIs in real-time, however having access to real-time POI visits is unfeasible. Therefore, we use four weeks of real-world mobility patterns, covering the time period between March 29 to April 25, 2021, inclusive. Then, we apply a moving average to estimate the fifth week, beginning at April 26, 2021. Finally, we look for the busiest time period in the fifth week for POIs of type *Restaurants* in order to mimic the environment, and set it as the start-time for our experiments. As such, there are 905, 536, and 743 open POIs that we consider over the Toronto, Montreal, and Calgary region, respectively. We set the start-time of all our experiments to be at 1pm on April 28 in Toronto, 1pm on May 1 in Montreal, and 12pm on May 1 in Calgary. This is represented by the variable $t$ on the $x$-axis of relevant plots. Additionally, we mark the start of the experiment with a vertical line on the relevant figures below. A summary of the datasets can be seen in Table 2. We provide ethical considerations in section D in the appendix.

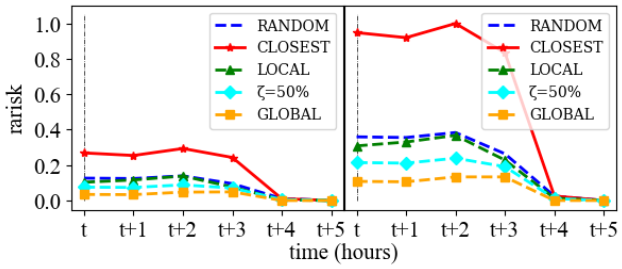
**Query generation**. We generated 2000 queries per second, uniformly distributed to each of Toronto, Montreal and Calgary, over the span of one minute. We removed queries that yield no POI candidates. At the end, the total number of queries is 119,799 in Toronto, 78,771 in Montreal, and 100,047 in Calgary.




(a) Toronto




(b) Montreal




(c) Calgary

**Figure 3: Rarisk tracking at 30 and 60 seconds on the left and right respectively;** $K = 1, \eta = 1/2$

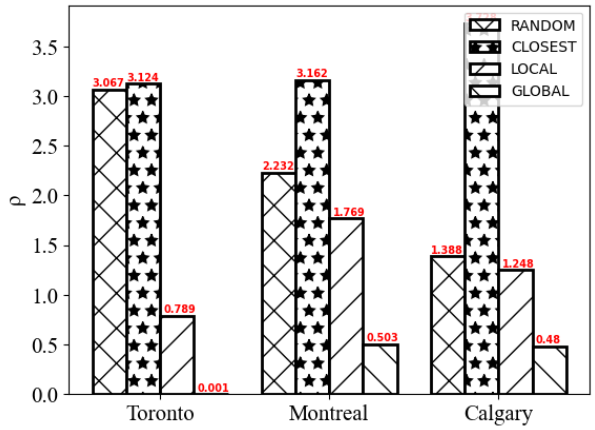



**Figure 4: Aggregate rarisk of each model at three major cities.**

**Parameter tuning**. For our experiments, we use parameter values related to the COVID-19 infectious disease. Health care professionals in Ontario define close contact when individuals are standing within $2m$ from each other for 15 minutes or more, or multiple shorter lengths of time without proper personal protective equipment[2]. We set the block area $X$ to be $4m^2$ in order to represent these social distancing guidelines. We set the search radius to 5 Km, and set POI type to *Restaurants*.

**Computing environment**. All experiments were conducted on a 2.2GHz 6-core Intel Core i7, 16 GB 2400 MHz DDR4 machine. We use IBM CPLEX solver to solve our optimization problem.

## 4.3 Methods

We consider four different recommendation methods. The first two are intuitive baselines that people tend to follow in their everyday mobility. The third and fourth are representing a local and a global optimization approach, respectively, to address the risk-aware POI recommendation problem, as described in section 3.2.

**RANDOM**. This model recommends $K$ POIs selected uniformly at random from the set of the eligible ones.

**CLOSEST**. This model recommends $K$ POIs selected based on proximity, where a POI located closer (to the location of the individual) is preferred to a POI located further away.

**LOCAL**. This model recommends $K$ POIs following the locally optimal risk-aware recommendation method described. It represents the current state of practice approach.

**GLOBAL**. This model recommends $K$ POIs following our proposed globally optimal risk-aware recommendation method.

## 4.4 Results and Discussion

**(Q1) Effect of recommendation model on risk.** In this experiment, we want to learn how effective our proposed model is compared to other models. We set $\eta = 1/2$, $K = 1$, and evaluate the performance at 30 and 60 seconds. Overall, 119,799, 78,771, and 100,047 queries were served in Toronto, Montreal and Calgary, respectively. To consider users' preferences, we let 50% of people follow the GLOBAL model, and 50% of people follow the RANDOM model, denoted by $\zeta = 50\%$. The results can be seen in Fig. 3. The figures on the left and right represent the 30 and 60 second marks, respectively. In both cases, GLOBAL yields the best results, successfully minimizing the overall risk at all POIs. Moreover, it is clear that the benefit of GLOBAL against the remaining models increases with the number of queries served each time. In the left figures, the benefit is relatively small; however the benefit between GLOBAL and the other models significantly increases at the 60 seconds mark, as can be seen on the right. This trend can be also observed for $K = 2$ in section B in the appendix. Moreover, it is clear that a partial compliance at $\zeta = 50\%$ outperforms both RANDOM and CLOSEST. A more comprehensive sensitivity analysis of how users' preferences affect risk is covered in section C in the appendix. Next, we let $\rho$ be the sum of all the aggregated POIs *rarisk* per recommendation model for the duration of 60 seconds, incurred from the time $t$ to $t+5$. The results can be seen in Fig. 4. The figure shows that GLOBAL has the smallest incurred added risk among all models, reducing the risk level by 99.87% in Toronto, 71.56% in Montreal, and 61.54% in Calgary with respect to the next best performing model. Section A in the appendix covers a more comprehensive evaluation of GLOBAL at different settings for $K$, $\eta$, and length of the experiment.

[2]https://www.ontario.ca/page/public-health-measures-and-advice

**(Q2) Effect of different invocation frequencies on the relative added risk of infection at POIs**. In this experiment, we study how different invocation frequencies affect GLOBAL *rarisk*. We set $K$ to 1 and run the experiment over 60 seconds, covering 119,799 queries in Toronto, 78,771 in Montreal, and 100,047 in Calgary. We test for invocation frequencies $\eta = \{1, 1/2, 1/4, 1/8, 1/10\}$ Hz. Then, we compare these frequencies with each other on a more granular scale by subtracting *rarisk* resulted from an invocation frequency of 1Hz with each one of the remaining invocation frequencies, separately. The results can be seen in Fig. 5. In the figures, a data point above zero means that the *rarisk* of that invocation frequency is lower than the one for invocation frequency of 1Hz. From the figures, it is clear that $\eta_{1/2}, \eta_{1/4}, \eta_{1/8}$ and $\eta_{1/10}$ are positive, hence are less risky than $\eta_1$. In fact, we can see that a lower invocation frequency tends to yield in general a better result, where $\eta_{1/10}, \eta_{1/8}, \eta_{1/4}, \eta_{1/2}$ and $\eta_1$ can be arranged from most to least effective parameters in reducing risk. However, for $\eta = 1/10$ Hz in Fig. 5a, this trend does not hold and the *rarisk* sharply increases. This might be due to Toronto having more POIs and more queries compared with Calgary and Montreal. This affects the computation space required to process the model, and results in a decreased performance.

**(Q3) Effect of $K$ on relative added risk**. Next, we evaluate how different $K$ values affect the *rarisk*. We set $\eta = 1/2$ Hz and test for $K = \{1, 2, 3, 4\}$. Then, we run the experiment over 10 seconds, addressing 19,963 queries in Toronto, 13,166 in Montreal, and 16,692 in Calgary. Fig. 6a, 6b, and 6c show the results of the experiment for Toronto, Montreal, and Calgary, respectively. It is interesting to note that there is a direct relation between risk and $K$ and it is clear that the model yields the best result for $K = 1$. This is intuitive due to our incorporated POI selection-likelihood. For $K = 1$, the most optimal result at rank 1 is chosen 100% of the time. However, at $K = 2$, that value drops to 80%, while the remaining 20% assigned to the second most optimal POI at rank 2. As $K$ increases, the likelihood to select the safest option at rank 1 decreases. Therefore, the *rarisk* of infection increases. Moreover, since $K$ proportionally affects the *rarisk* of our model, we want to verify that GLOBAL remains the most optimal model for a large $K$ value. We test for $K = 4$ and show the results in Fig. 6d, 6e, and 6f, which confirm that GLOBAL continues to provide the most optimal recommendations.

**(Q4) Effect of different invocation frequencies on model run-time**. In this experiment, we examine the relative run-time of GLOBAL for different invocation frequencies. We set $K = 1$ and run the experiment over the span of 30 seconds, covering 59,892 queries in Toronto, 39,327 in Montreal, and 50,062 in Calgary. We text for $\eta = \{1, 1/2, 1/4, 1/8, 1/10\}$ Hz. As noted, there are roughly 2,000 queries per second. This means that the rough number of queries a model gets from each invocation frequency setting of $1/\alpha$ Hz is $\sim$2,000$\times\alpha$. As such, the input of the model in this experimental setting increases linearly. Fig. 7 shows the relative computation time per invocation frequency per city. We can see that the output is a linear relationship between the invocation frequency and the computation time. Therefore, we can conclude that the processing time of the model depends on the number of queries it serves. For a uniform flow of queries, the computation time increases linearly. Note that according to our complexity analysis, constants $K$ and $\mathcal{P}$, and a varying $\mathcal{N}$ should increase the computation time exponentially rather than linearly with respect to $\mathcal{N}$. This difference can be attributed to implementation optimizations of the solver we use.

**(Q5) Effect of $K$ on model run-time**. In this experiment, we evaluate how $K$ values affect GLOBAL run-time. We set $\eta = 1/2$ Hz and test for $K = \{1, 2, 3, 4\}$. We run the experiment over 10 seconds, addressing 19,963 queries in Toronto, 13,166 in Montreal, and 16,692 in Calgary. Fig. 8 shows the results of the experiment. We observe that the computation of the model increases exponentially with $K$.

**(Q6) Effect of $\mathcal{N}$, $\mathcal{P}$ and $K$ on computation time**. Finally, we want to study how the number of users, POIs, and recommendations affect GLOBAL's computation time. This test can provide insights regarding possible limitations and mitigation approaches for our model. We simulate a sparse eligibility matrix, following a normal distribution, where the mean is 25 and the standard deviation is 15. This is done to represent a percentage of randomly selecting eligible POIs, where, on average, a user has 25% candidate POIs. Then, we evaluate how a change in any of $\mathcal{N}$, $\mathcal{P}$ or $K$ affect the computation time. In Fig. 9a, we fix the number of POIs to 900 and $K = 2$. We observe a polynomial increase in computation time as the number of users increases. In Fig. 9b, we fix the number of users to 1000 and $K = 2$. We observe that there is a linear relationship between the number of POIs and the computation time of GLOBAL. Finally, in Fig. 9c, we fix the number of users to 500 and the number of POIs to 400. We observe an exponential increase in time complexity.

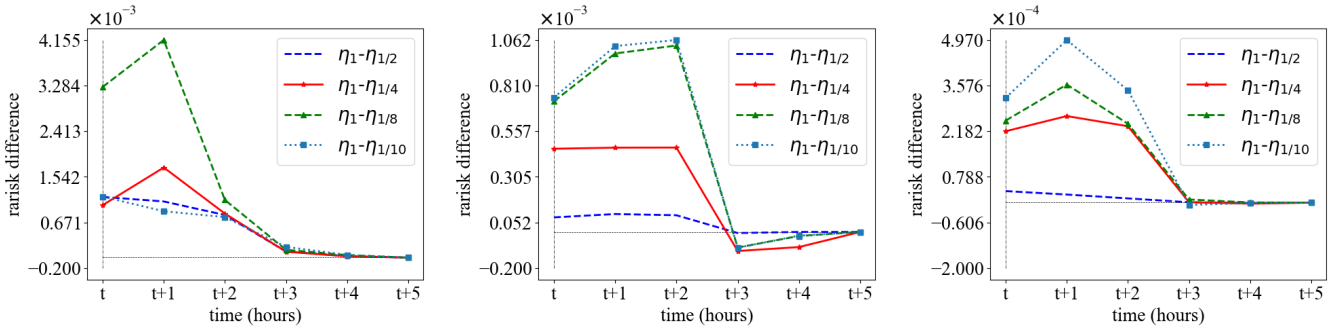
## 5 RELATED WORK

Our research is related to (i) *trajectory data mining* and (ii) *mobility and epidemics*. We cover below some of the most significant efforts relevant to our work. Note that some related works have already been cited throughout the manuscript to keep the discussion focused, so they are mostly omitted here.

### 5.1 Trajectory Data Mining

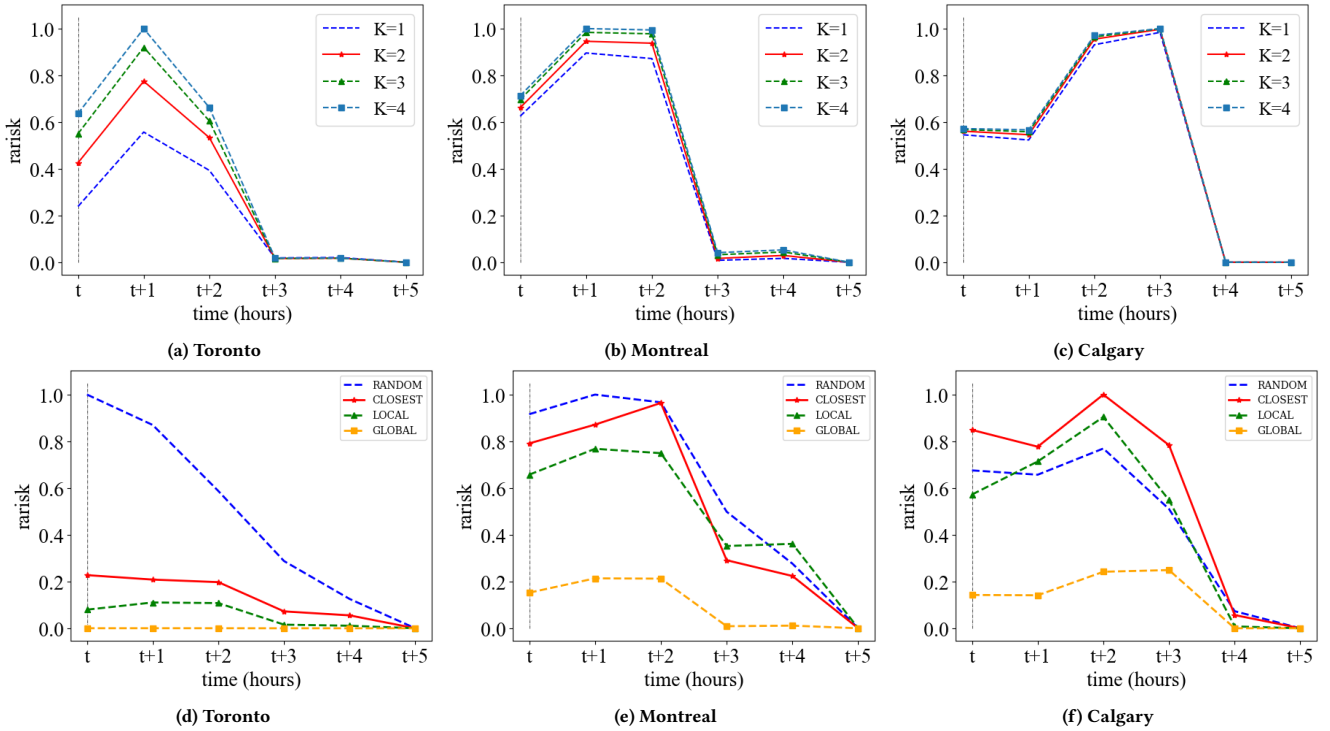
Trajectory data mining is the process of analyzing the mobility patterns of individuals or objects over time to gain insights and make decisions. It has been an active research direction for a long time and many comprehensive surveys on the topic exist [6, 16, 40]. This high interest can largely be attributed to the rapid development and prominence of geospatial technologies, the abundance of location-based services [23], and deep-learning based methods to represent trajectory data (see surveys [20, 36, 37]). The focus is on popular technical problems, including trajectory similarity [12], simplification [2], classification [4], and clustering [21]. Other popular topics include pedestrian group mining [35] and semantic analysis of city neighborhoods [26]. Our research relies on trajectory data mining methods to process and analyze geolocation data related to POIs and to provide trip recommendations.

**Location-based recommendations**. Location-based recommendations are personalized suggestions for places to visit, based on an individual's current location or intended location [7]. These recommendations use information about the person's location and their preferences to suggest points of interest (POIs) such as restaurants, shops, tourist attractions, and more. This information can be obtained through GPS data, or by manually entering a location, and can be used to suggest POIs that are nearby or within a certain




(a) Toronto (b) Montreal (c) Calgary

**Figure 5: GLOBAL performance related to invocation frequency of 1 over 60 seconds; $\eta = 1/2$**




(a) Toronto (b) Montreal (c) Calgary

(d) Toronto (e) Montreal (f) Calgary

**Figure 6: Effects of $K$ on POI rarisk. (a)-(c) GLOBAL model at different K values. (d)-(e) Relative POI aggregated rarisk for K=4.**

distance. The goal of location-based recommendations is to provide a convenient and relevant experience for the user by suggesting POIs that are most relevant to their interests and current location. Deep learning methods have also been proposed for providing location-based recommendations [24, 39]. These works cannot be easily adapted to the domain of epidemics. They do not model the POI risk of infection, and they focus on analysis of historical transactions to provide recommendations that optimize different type of preferences (i.e., restaurants, touristic attractions, and more).

## 5.2 Mobility and Epidemics

Mobility is related to epidemics because the movement of individuals and their social interactions can impact the topology of the underlying contact network on which an infectious disease spreads [33]. For example, increased mobility can facilitate the rapid spread of a disease from one location to another. On the other hand, reducing mobility, such as through travel restrictions or quarantine measures, can slow the spread of a disease and contain outbreaks.

**Digital contact tracing**. Digital contact tracing has been proposed as a technology-based approach for tracking the spread of infectious diseases, especially during outbreaks or epidemics. It involves the use of digital devices, such as smartphones, to identify and track close contacts between individuals who have been infected with a disease [9, 11, 15]. Digital contact tracing can complement traditional public health measures, such as manual contact tracing, and can provide a more efficient and scalable way of identifying and alerting individuals who have been in close contact with an infected person. For instance, Aleta et al. [1] synthesized contact networks

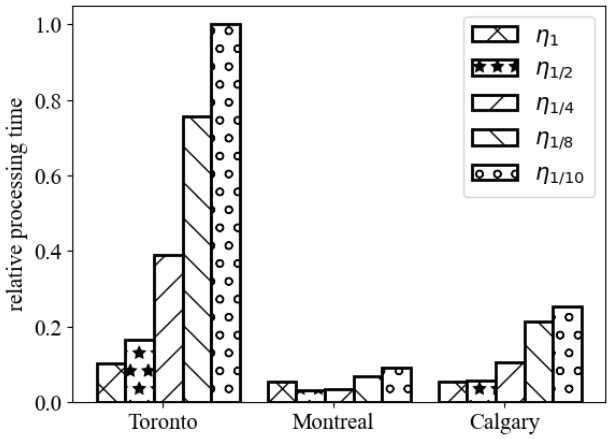



**Figure 7: Different invocation frequencies $\eta$; $K = 1$ over 60 seconds.**

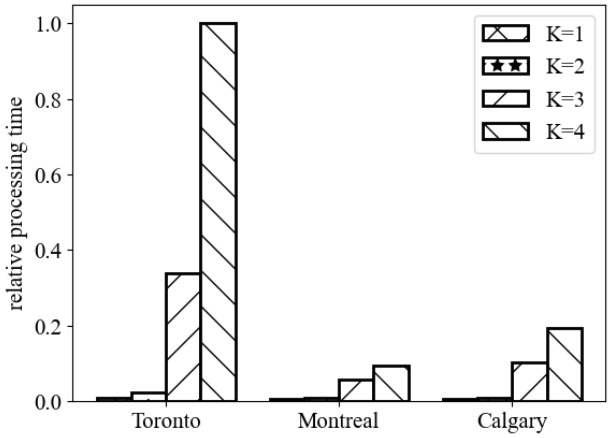



**Figure 8: Different $K$ values; $\eta = 1/2$ over 10 seconds.**

and modeled SARS-CoV-2 transmission in the Boston metropolitan area using census and contact tracing data. They showed how vital contact tracing is in preventing a second wave of spreading when complete isolation is relaxed.

**Mobility-based interventions for epidemics**. Pechlivanoglou et. al [31] emphasized the need for more accurate microscopic spatiotemporal epidemic modeling that allows to design targeted non-pharamaceutical intervention strategies that aim to control the epidemic spreading. For example, Block et al. [8] and Pechlivanoglou et al. [32] proposed more moderate distancing strategies including limiting contacts to *similar*, *community-based* or *repetitive* contacts. Similarly, Miralles-Pechuán et al. [27] used reinforcement learning to suggest high-level intervention strategies rewarding fewer infections and less severe lockdowns. Fan et al. [14] showed that mobility inspired interventions, such as limited long distance trips, can notably reduce epidemic spread. Hébert-Dufresne et al. [22] argued the importance of modeling mobility-based population heterogeneity for predicting an infectious disease's outbreak size through digital contact tracing technologies. Changruenngam et al. [10] studied the effects of individual human mobility on disease transmission dynamics. Lloyd-Smith et al. [25] showed how the basic $R_0$ in traditional epidemic analyses is only a population-level estimate; thus, more targeted control interventions would be more effective. These works are complementary (orthogonal) to our study and can be used together to enhance the intervention strategies.

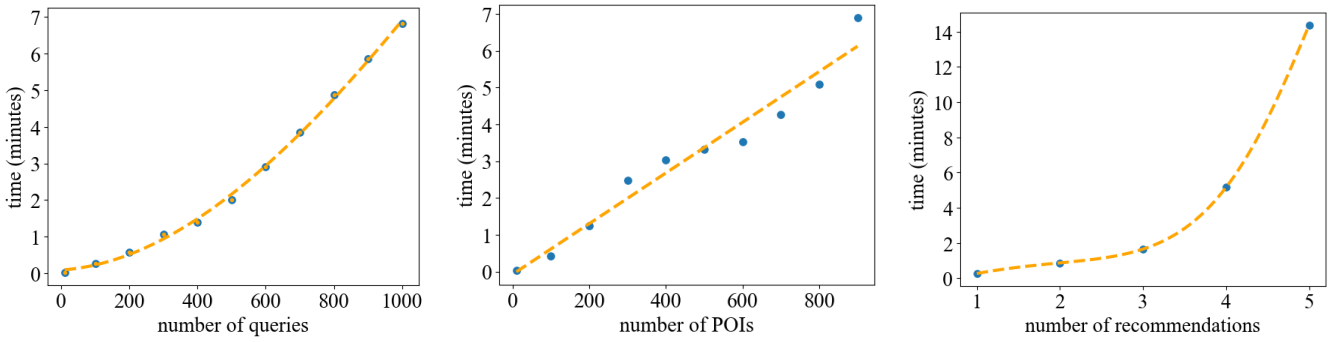
**Trip recommendations and epidemics**. Our study is mostly related to research that focuses on providing POI recommendations that satisfy domain-specific constraints. These methods work by factoring in infection risk in the recommendation models. For example, Fotsing et al. [18] designed an epidemic-aware socio-spatial POI recommender model. Graph-based path search algorithm have also been proposed for reducing an individual's COVID-19 exposure risk taking into account accessibility constraints, outdoor exposure thresholds, and congestion tolerance [5, 34]. Similarly, Alix et al. [3] presented a system for recommending safer trips to POIs on a geographic map based on alternative trip risk evaluation methods.

## 6 CONCLUSIONS

Understanding the relationship between human mobility and disease spreading is crucial for developing effective public health strategies to prevent and control the spread of diseases [32]. There are many mobility-related approaches to contain and mitigate epidemics. One such approach is by informing individuals about the risks associated with their trips via recommendation systems. This approach can alleviate economic downturns and reduce or eliminate intrusive approaches, such as quarantines. In this work, we presented a novel optimal risk-aware POI recommendation model to use during epidemics. To the best of our knowledge, no previous research has explored the application of the GAP to address a POI recommendation problem. Throughout this paper, we studied the relative risk of infection as a result of following specific POI recommendation strategies. Through a series of extensive experiments, we showed that (GLOBAL) is superior to current state of practice approach (LOCAL) and sensible baseline models, i.e. CLOSEST and RANDOM. This is because our model makes POI recommendations that minimize the overall risk at all POIs. We also performed parameter sensitivity analysis that showed that GLOBAL consistently outperformed the other models in various settings, such as different number of recommendations provided and different invocation frequencies. Moreover, we showed that the GLOBAL's impact increases over time as the difference in added risk increases with time. We further evaluated the computational trade-offs associated with different dimensions of the problem, including the number of users, POIs and recommendations, in order to better understand its limitations. We further showed that GLOBAL reduced the relative added risk of infection by 99.87%, 71.56% and 61.54% compared with the second best model in three major cities in Canada. Our results provide clear evidence that our recommendation model can be a vital tool that informs individuals of the risks associated with their mobility patterns, and can help them make informed decisions.

## REFERENCES

[1] Alberto Aleta, David Martin-Corral, et al. 2020. Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. *medRxiv* (2020).
[2] Gian Alix and Manos Papagelis. 2023. PathletRL: Trajectory Pathlet Dictionary Construction using Reinforcement Learning. In *Proc. of the 31st ACM SIGSPATIAL*. Accepted. In Press.
[3] Gian Alix, Nina Yanin, Tilemachos Pechlivanoglou, Jing Li, Farzaneh Heidari, and Manos Papagelis. 2022. A Mobility-based Recommendation System for Mitigating the Risk of Infection during Epidemics. In *23rd IEEE MDM*. IEEE, 292–295.




**(a) $\mathcal{N}$ changes, while $K$ and $\mathcal{P}$ are constants.** **(b) $\mathcal{P}$ changes, while $K$ and $\mathcal{N}$ are constants.** **(c) $K$ changes, while $\mathcal{N}$ and $\mathcal{P}$ are constants.**

**Figure 9: Relationship between $\mathcal{N}$, $\mathcal{P}$ and $\mathcal{K}$ and computation time.**


[4] Mahmoud Alsaeed, Ameeta Agrawal, and Manos Papagelis. 2023. Trajectory-User Linking using Higher-order Mobility Flow Representations. In *24th IEEE MDM*. 158–167.

[5] Chrysovalantis Anastasiou, Constantinos Costa, Panos K Chrysanthis, Cyrus Shahabi, and Demetrios Zeinalipour-Yazti. 2021. ASTRO: reducing COVID-19 exposure through contact prediction and avoidance. *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 8, 2 (2021), 1–31.

[6] Gowtham Atluri, Anuj Karpatne, and Vipin Kumar. 2018. Spatio-Temporal Data Mining: A Survey of Problems and Methods. *ACM Comp. Surveys* 51, 4 (2018).

[7] Jie Bao, Yu Zheng, David Wilkie, and Mohamed Mokbel. 2015. Recommendations in location-based social networks: a survey. *GeoInformatica* 19 (2015), 525–565.

[8] Per Block, Marion Hoffman, Isabel J Raabe, Jennifer Beam Dowd, Charles Rahal, Ridhi Kashyap, and Melinda C Mills. 2020. Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behaviour* (2020), 1–9.

[9] Giulia Cencetti, Gabriele Santin, Antonio Longa, Emanuele Pigani, Alain Barrat, Ciro Cattuto, Sune Lehmann, Marcel Salathe, and Bruno Lepri. 2020. Digital Proximity Tracing in the COVID-19 Pandemic on Empirical Contact Networks. *medRxiv* (2020).

[10] Suttikiat Changruenngam, Dominique J Bicout, and Charin Modchang. 2020. How the individual human mobility spatio-temporally shapes the disease transmission dynamics. *Scientific Reports* 10, 1 (2020), 1–13.

[11] Forrest W. Crawford, Sydney A. Jones, Matthew Cartter, Samantha G. Dean, et al. 2022. Impact of close interpersonal contact on COVID-19 incidence: Evidence from 1 year of mobile device data. *Science Advances* 8, 1 (2022).

[12] Liwei Deng, Yan Zhao, Zidan Fu, Hao Sun, Shuncheng Liu, and Kai Zheng. 2022. Efficient Trajectory Similarity Computation with Contrastive Learning. In *Proc. of the 31st ACM CIKM*. 365–74.

[13] Ken TD Eames and Matt J Keeling. 2003. Contact tracing and disease control. *Proc. of the Royal Soc. of London. Series B: Biolog. Sciences* 270, 1533 (2003), 2565–2571.

[14] Chao Fan, Ronald Lee, Yang Yang, and Ali Mostafavi. 2021. Fine-grained data reveal segregated mobility networks and opportunities for local containment of COVID-19. *Scientific Reports* 11, 1 (2021), 1–6.

[15] Katayoun Farrahi, Remi Emonet, and Manuel Cebrian. 2014. Epidemic contact tracing via communication traces. *PloS one* 9, 5 (2014), e95133.

[16] Zhenni Feng and Yanmin Zhu. 2016. A Survey on Trajectory Data Mining: Techniques and Applications. *IEEE Access* 4 (2016), 2056–2067.

[17] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, et al. 2020. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 368, 6491 (2020).

[18] Cedric Parfait Kankeu Fotsing, Ya-Wen Teng, Guang-Siang Lee, Chih-Ya Shen, Yi-Shin Chen, and De-Nian Yang. 2022. On Epidemic-aware Socio Spatial POI Recommendation. In *23rd IEEE Intern. Conf. on Mobile Data Mgt (MDM)*. 169–178.

[19] Abhirup Ghosh and Tong Xia. 2021. Mobility-based individual poi recommendation to control the covid-19 spread. In *IEEE BigData*. 4356–4364.

[20] Ali Hamdi, Khaled Shaban, Abdelkarim Erradi, Amr Mohamed, Shakila Khan Rumi, and Flora D. Salim. 2021. Spatiotemporal data mining: a survey on challenges and open problems. *Artificial Intelligence Review* 55, 2 (2021), 1441–1488.

[21] Nan Han, Shaojie Qiao, Kun Yue, Jianbin Huang, Qiang He, Tingting Tang, Faliang Huang, et al. 2022. Algorithms for Trajectory Points Clustering in Location-Based Social Networks. *ACM Trans. Intell. Syst. Technol.* 13, 3 (2022).

[22] Laurent Hébert-Dufresne, Benjamin M Althouse, Samuel V Scarpino, and Antoine Allard. 2020. Beyond R0: Heterogeneity in secondary infections and probabilistic epidemic forecasting. *medRxiv* (2020).

[23] Haosheng Huang, Georg Gartner, Jukka M. Krisp, Martin Raubal, and Nico Van de Weghe. 2018. Location based services: ongoing evolution and research agenda. *Journal of Location Based Services* 12, 2 (2018), 63–93.

[24] Zhenhua Huang, Xiaolong Lin, Hai Liu, Bo Zhang, Yunwen Chen, and Yong Tang. 2020. Deep representation learning for location-based recommendation. *IEEE Transactions on Computational Social Systems* 7, 3 (2020), 648–658.

[25] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. 2005. Superspreading and the effect of individual variation on disease emergence. *Nature* 438, 7066 (2005), 355–359.

[26] Saim Mehmood and Manos Papagelis. 2020. Learning Semantic Relationships of Geographical Areas based on Trajectories. In *21st IEEE MDM*. 109–118.

[27] Luis Miralles-Pechuán, Fernando Jiménez, Hiram Ponce, and Lourdes Martínez-Villaseñor. 2020. A methodology based on deep q-learning/genetic algorithms for optimizing covid-19 pandemic government actions. In *Proc. of the 29th ACM CIKM*. 1135–1144.

[28] Mohamed Mokbel, Sofiane Abbar, and Rade Stanojevic. 2020. Contact tracing: Beyond the apps. *SIGSPATIAL Special* 12, 2 (2020), 15–24.

[29] Temel Öncan. 2007. A survey of the generalized assignment problem and its applications. *INFOR: Inform. Syst. and Operat. Research* 45, 3 (2007), 123–141.

[30] World Health Organization, Centers for Disease Control, Prevention, et al. 2015. *Implementation and management of contact tracing for Ebola virus disease: emergency guideline*. Technical Report. World Health Organization.

[31] Tilemachos Pechlivanoglou, Gian Alix, Nina Yanin, Jing Li, Farzaneh Heidari, and Manos Papagelis. 2022. Microscopic modeling of spatiotemporal epidemic dynamics. In *Proc. of the 3rd ACM SIGSPATIAL/SpatialEpi*. 11–21.

[32] Tilemachos Pechlivanoglou, Jing Li, Jialin Sun, Farzaneh Heidari, and Manos Papagelis. 2022. Epidemic Spreading in Trajectory Networks. *Big Data Research* 27 (2022).

[33] Tilemachos Pechlivanoglou and Manos Papagelis. 2018. Fast and Accurate Mining of Node Importance in Trajectory Networks. In *IEEE BigData*. 781–790.

[34] Vasilis Ethan Sarris, Constantinos Costa, and Panos K Chrysanthis. 2022. ASTRO-K: Finding Top-k Sufficiently Distinct Indoor-Outdoor Paths. In *2022 23rd IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 372–377.

[35] Abdullah Sawas, Abdullah Abuolaim, Mahmoud Afifi, and Manos Papagelis. 2019. A versatile computational framework for group pattern mining of pedestrian trajectories. *GeoInformatica* 23, 4 (2019), 501–531.

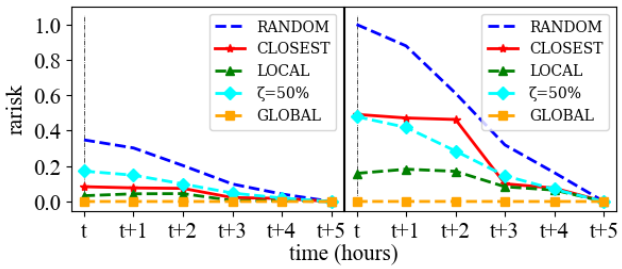
[36] Sheng Wang, Zhifeng Bao, J. Shane Culpepper, and Gao Cong. 2021. A Survey on Trajectory Data Mgt., Analytics, and Learning. *ACM Comp. Survey* 54, 2 (2021).

[37] Senzhang Wang, Jiannong Cao, and Philip Yu. 2022. Deep Learning for Spatio-Temporal Data Mining: A Survey. *IEEE Trans. on Knowl. & Data Eng.* 34, 08 (2022), 3681–700.

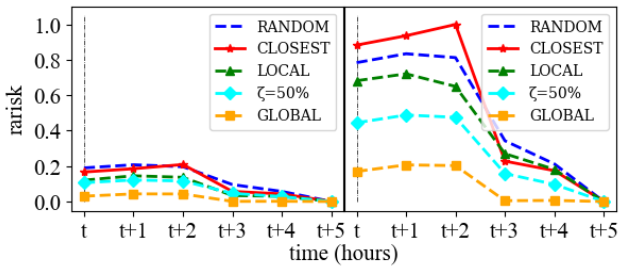
[38] Samuel Woodhams. 2020. COVID-19 digital rights tracker (2020).

[39] Min Xie, Hongzhi Yin, Hao Wang, Fanjiang Xu, Weitong Chen, and Sen Wang. 2016. Learning graph-based poi embedding for location-based recommendation. In *Proc. of the 25th ACM CIKM*. 15–24.

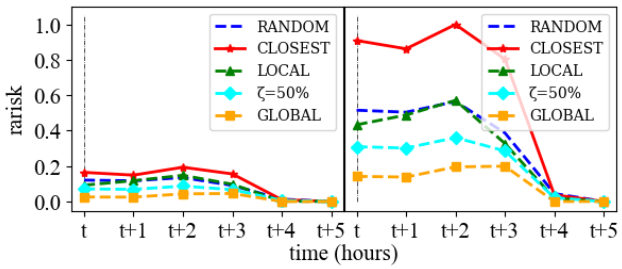
[40] Yu Zheng. 2015. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 3 (2015), 1–41.




(a) Toronto




(b) Montreal




(c) Calgary

**Figure 10: Relative added risk tracking at 10 seconds on the left and 30 seconds on the right;** $K = 2, \eta = 1/2$

# APPENDIX

## A EXPERIMENTAL RESULTS

In this section, we want to understand how GLOBAL performs for different invocation frequencies $\eta$, number of recommendations $K$, and the total number of queries served at each instance. For each of these settings, we check the relative aggregated added risk incurred at all POIs per city. Finally, we want to quantify the amount of benefit GLOBAL brings. Table 3 shows the results of this experiment. The bolded results represent the optimal recommendation model that added the minimum amount of risk. It is clear that the GLOBAL model consistently yields the best outcome. The underlined values represent the second best results. We quantified the percentage of improvement by comparing our proposed method GLOBAL with the second best method. Overall, our model yielded an average of 99.88%, 77.33% and 65.71% improvement in Toronto, Montreal, and Calgary, respectively, compared with the second best performer. This shows that the GLOBAL model is not affected by the number of recommendations, invocation frequencies, number of queries, nor duration of the experiment.

## B MODEL PERFORMANCE ANALYSIS OVER TIME

In this section, we want to learn how the relative added risk is affected over time per each model for a different set of parameters than the ones that were covered in **Q1** in section 4.4. We set $K$ to 2 and invocation frequency $\eta$ to 1/2. We run the experiment for 30 seconds, but also check the results at the 10 seconds mark. Overall, this experiment serves 59,892 queries in Toronto, 39,327 queries in Montreal, and 50,062 queries in Calgary. Fig. 10a, Fig. 10b, and Fig. 10c show the results of this experiment. The figures on the left hand side show the 10 second mark, while the ones on the right hand side show the 30 second mark. The figures clearly show that as time increases, the relative added risk increases as well, regardless of the model. Moreover, the figures show that GLOBAL is the best performer, consistently yielding the least relative added risk. Additionally, it shows that as time progresses, the rate of increased added risk from the 10 second mark to the 30 second mark is much more subtle for GLOBAL compared with the other models. This implies that our model provides a compounding reduction in risk over time.

## C EFFECT OF POPULATION COMPLIANCE LEVEL ON RISK

In this section, we want to consider users' preferences and evaluate how they affect risk. We let a certain percentage of people follow the GLOBAL recommendation and the remaining people follow the RANDOM recommendation. Furthermore, we want to see how it is compared with the other models: CLOSEST and LOCAL. Let $\zeta$ and $100-\zeta$ represent the percentage of people following the GLOBAL and RANDOM models, respectively. We set the number of recommendation $K$ to 2, invocation frequency $\eta$ to 1/2, and run the experiment for 30 seconds. Altogether, this experiment serves 29,939 queries in Toronto, 19,704 queries in Montreal, and 25,056 queries in Calgary. The results can be seen in Fig. 11. It is clear that GLOBAL performs the best by adding the least amount of risk. Additionally, we can observe that as $\zeta$ increases, more people follow the GLOBAL model, and the relative added risk decreases. Therefore, we can conclude that the ideal scenario is one where all people follow the recommendation provided by the GLOBAL model. However, even a partial compliance from the community can provide significant benefit in mitigating the risk of infection.

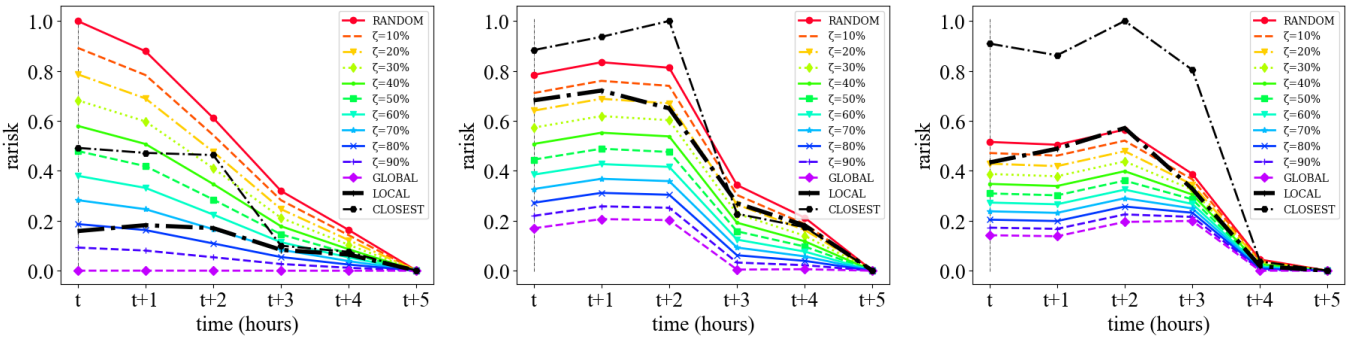
## D PRIVACY AND ETHICS STATEMENT

In order to uphold privacy and ethical conduct, all datasets were preprocessed to remove identifying POI information. The original datasets, which include POIs' `Places`, `Geometry`, and `Patterns`, became available through SafeGraph[3]. We used an archived version from SafeGraph, however it is important to note that recently Advan[4] began providing the `Patterns` datasets. The `Patterns` datasets are anonymized and collected with proper informed consent from a list of opted-in mobile devices. Throughout our work, we adhered to all the terms and conditions stated by the dataset provider, including properly citing their work.

[3]https://www.safegraph.com/
[4]https://advanresearch.com/

**Table 3: *rarisk* per city at different settings**

| | | $K=1, \eta=1/2$ | | | | | | $K=2, \eta=1/2$ | | | $K=4, \eta=1/2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 10 sec | 20 sec | 30 sec | 40 sec | 50 sec | 60 sec | 10 sec | 20 sec | 30 sec | 10 sec |
| Toronto | # queries | 19,963 | 39,924 | 59,892 | 79,865 | 99,836 | 119,799 | 19,963 | 39,924 | 59,892 | 19,963 |
| | RANDOM | 0.14826 | 0.27747 | 0.42552 | 0.62514 | 0.79820 | 0.98176 | 0.16287 | 0.28986 | 0.43429 | 0.15429 |
| | CLOSEST | 0.04428 | 0.13449 | 0.28047 | 0.47114 | 0.70841 | 1.00000 | 0.04138 | 0.12025 | 0.24225 | 0.04089 |
| | LOCAL | 0.02468 | 0.06120 | 0.13079 | 0.16297 | 0.20826 | 0.25273 | 0.01961 | 0.07490 | 0.09969 | 0.01743 |
| | GLOBAL | **0.00002** | **0.00005** | **0.00010** | **0.00018** | **0.00028** | **0.00040** | **0.00003** | **0.00006** | **0.00012** | **0.00003** |
| | Improvement | 99.92% | 99.92% | 99.92% | 99.89% | 99.87% | 99.84% | 99.85% | 99.92% | 99.88% | 99.83% |
| Montreal | # queries | 13,166 | 26,232 | 39,327 | 52,532 | 65,682 | 78,771 | 13,166 | 26,232 | 39,327 | 13,166 |
| | RANDOM | 0.07007 | 0.15116 | 0.25635 | 0.38151 | 0.53314 | 0.70591 | 0.06414 | 0.14172 | 0.25677 | 0.06676 |
| | CLOSEST | 0.05925 | 0.15867 | 0.29772 | 0.48807 | 0.71829 | 1.00000 | 0.05678 | 0.14924 | 0.27697 | 0.05733 |
| | LOCAL | 0.07279 | 0.13525 | 0.21666 | 0.31097 | 0.42321 | 0.55941 | 0.03998 | 0.13455 | 0.21530 | 0.05270 |
| | GLOBAL | **0.00943** | **0.02493** | **0.04748** | **0.07828** | **0.11609** | **0.15918** | **0.01010** | **0.02668** | **0.05069** | **0.01092** |
| | Improvement | 84.08% | 81.57% | 78.09% | 74.83% | 72.57% | 71.55% | 74.74% | 80.17% | 76.46% | 79.28% |
| Calgary | # queries | 16,692 | 33,420 | 50,062 | 66,696 | 83,348 | 100,047 | 16,692 | 33,420 | 50,062 | 16,692 |
| | RANDOM | 0.03150 | 0.07516 | 0.13293 | 0.20197 | 0.28023 | 0.37223 | 0.03163 | 0.07574 | 0.13387 | 0.03238 |
| | CLOSEST | 0.05127 | 0.14652 | 0.28697 | 0.47070 | 0.72423 | 1.00000 | 0.04464 | 0.12434 | 0.23983 | 0.04179 |
| | LOCAL | 0.02847 | 0.06976 | 0.11728 | 0.18722 | 0.24236 | 0.33476 | 0.03032 | 0.07643 | 0.12204 | 0.03313 |
| | GLOBAL | **0.00896** | **0.02340** | **0.04381** | **0.06713** | **0.09388** | **0.12876** | **0.00919** | **0.02392** | **0.04480** | **0.00934** |
| | Improvement | 68.53% | 66.46% | 62.64% | 64.14% | 61.26% | 61.54% | 69.69% | 68.42% | 63.29% | 71.16% |




**(a) Toronto** **(b) Montreal** **(c) Calgary**

**Figure 11: relative added risk for different models and compliance levels over 30 seconds;** $K=2, \eta=1/2$