Hide Your Distance: Privacy Risks and Protection in Spatial Accessibility Analysis

Liyue Fan
liyue.fan@charlotte.edu
University of North Carolina at Charlotte
Charlotte, NC, USA

Luca Bonomi
luca.bonomi@vumc.org
Vanderbilt University Medical Center
Nashville, TN, USA

ABSTRACT

Measuring spatial accessibility to healthcare resources and facilities has long been an important problem in public health. For example, during disease outbreaks, sharing spatial accessibility data such as individual travel distances to health facilities is vital to policy making and designing effective interventions. However, sharing these data may raise privacy concerns, as information about individual data contributors (e.g., health status and residential address) may be disclosed. In this work, we investigate those unintended information leakage in spatial accessibility analysis. Specifically, we are interested in understanding whether sharing data for spatial accessibility computations may disclose individual participation (i.e., membership inference) and personal identifiable information (i.e., address inference). Furthermore, we propose two provably private algorithms that mitigate those privacy risks. The evaluation is conducted with real population and healthcare facilities data from Mecklenburg county, NC and Nashville, TN. Compared to state-of-the-art privacy practices, our methods effectively reduce the risks of membership and address disclosure, while providing useful data for spatial accessibility analysis.

CCS CONCEPTS

• **Security and privacy → Data anonymization and sanitization**; • **Information systems → Geographic information systems**; • **Applied computing → Health informatics**.

KEYWORDS

Privacy, Spatial Accessibility, Health Informatics

ACM Reference Format:
Liyue Fan and Luca Bonomi. 2023. Hide Your Distance: Privacy Risks and Protection in Spatial Accessibility Analysis. In *The 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23), November 13–16, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3589132.3625656

1 INTRODUCTION

Spatial accessibility to resources and facilities has been of great importance in public health [4, 10, 15, 17, 19, 28, 31]. For example, understanding spatial access to healthy foods (e.g., grocery stores that sell fresh fruits and vegetables) could help reduce the risk of obesity and other chronic diseases and guide the development of progressive intervention strategies [4, 15]. Similarly, enhancing spatial accessibility of primary care may improve overall population health and benefit disadvantaged populations, e.g., in hypertension awareness and control [10, 17, 32]. Recently, the COVID-19 pandemic has seen healthcare resources (e.g., hospital beds, ventilators, testing resources) overwhelmed in a number of countries. Understanding the spatial accessibility for COVID-19 patients and population at risk has been crucial for allocating healthcare resources efficiently and effectively [13, 14, 22].

The two-step floating catchment area (2SFCA) method is widely used to measure spatial accessibility [10, 19]. It defines a "catchment area" for healthcare facilities such that individuals residing within the catchment area have utilization. The catchment is based on the travel distance[1] between the residential location and the facility location. Several variants of 2SFCA (such as E2SFCA [18] and G2SFCA[31]) have been proposed to incorporate distance decay, i.e., increasing travel distance would lead to less utilization.

However, sharing spatial accessibility data, e.g., travel distances to healthcare facilities, may raise individual privacy concerns. To our best knowledge, it has not been studied whether releasing those travel distances would result in unintended information leakage. One example of information leakage is membership leakage, where an individual's participation in a dataset can be inferred. In the context of spatial accessibility, membership would disclose sensitive information about the target individual, such as COVID-19 or diabetes diagnosis. Another example of information leakage is personal identifiable information (PII), as defined by the HIPAA de-identification standard. The disclosure of PII, such as name, ID, and street address, may incur severe damages to the target individual and the data publisher.

In this paper, we investigate unintended information leakage in sharing spatial accessibility data and propose privacy-enhancing methods to mitigate such leakage. The specific contributions of this work are:

- We formulate two privacy risk measures to quantify privacy risks associated with sharing spatial accessibility data. Intuitively, membership inference estimates a public individual's participation in the protected dataset; address inference estimates the residential street address of an individual in the protected dataset.
- We propose two private methods to release spatial accessibility data at the individual level, for travel distance to the

SIGSPATIAL '23, November 13–16, 2023, Hamburg, Germany

ACM ISBN 979-8-4007-0168-9/23/11…$15.00
https://doi.org/10.1145/3589132.3625656

[1] catchment can also be defined by travel time; travel distance is adopted in this work

nearest facility and travel distances to all facilities, respectively. We prove that both methods satisfy metric privacy, a generalized notion based on differential privacy; and releasing only distance to the nearest facility offers significant privacy savings.
- We conduct empirical evaluation with real population data and healthcare facilities data from Mecklenburg county, NC and Nashville, TN. We examine the feasibility of releasing spatial accessibility data with regard to both accuracy and privacy leakage; we further conduct a case study on spatial accessibility for COVID-19. Results show that our methods provide useful data for spatial accessibility analysis while providing strong privacy protection.

The rest of the paper is organized as follows: Section 2 briefly reviews recent literature most related to this work; Section 3 describes the problem setting, introduces travel distance-based spatial accessibility analysis, and presents exploratory analysis on travel distances in real datasets; Section 4 presents the definition of metric privacy, the proposed privacy methods, and theoretical guarantees; Section 5 presents two empirical privacy risk measures; Section 6 discusses empirical results; Section 7 concludes the paper with several working directions for future research.

2 RELATED WORK

Location Privacy. Location privacy has been extensively studied in literature, with a plethora of location data sharing methods to enable location-based applications, such as crowd-sourcing [25, 30], social networks [9, 16], and transportation [24, 26]. Recently, several surveys and empirical studies [7, 8, 11, 23] have categorized and analyzed existing methods. However, we do not consider this work as developing location privacy methods. As discussed later in detail, we adopt a practical assumption in which residential locations may be publicly available, e.g., via voter registration data; the privacy risks lie in the inference of individual participation in a protected dataset and the inference of residential addresses for those in the protected dataset. Interestingly, our work may be analogous to the technique of trilateration [29] (or multilateration), which estimates the target position by the distances between the target and a number of reference points (e.g., receivers). However the travel distances in our problem setting may be more challenging to model, as they are constrained by physical road networks.

Differential Privacy. Differential privacy (DP) [5] has become the state-of-the-art paradigm for privacy protection in statistical databases. It assumes a trusted data curator is responsible for data aggregation and guarantees that an adversary who observes the output results is not able to decide whether a particular record is included in the input database. While classic DP has been widely adopted for sharing dataset-level statistics, recent studies employ the local DP notion (LDP) in order to share individual-level data. LDP mechanisms can be built on randomized response techniques [6, 12], which provide strong privacy protection (i.e., in input indistinguishability) but may incur high utility loss. A generalized privacy notion, metric-based privacy [3], has been proposed to relax the privacy guarantees and to improve data utility, whereas the indistinguishability guarantee depends on the distance between input secrets (by a specific metric). It has been shown that differential privacy is a special instance of metric privacy. Geo-indistinguishability [1] is another application of metric privacy in 2D space. In this work, we adopt the metric privacy notion to protect the privacy of individual data contributors, while enable accurate computation of spatial accessibility.

3 PRELIMINARIES

3.1 Overview

Our solution aims to enable the analysis of spatial accessibility by public health researchers, with individually contributed travel distances. The proposed problem setting is presented in Figure 1. As an example, a public health study is interested in analyzing the spatial accessibility to healthcare resources for a specific cohort (e.g., COVID-19 or diabetic patients), in order to design effective interventions. Healthcare providers may share travel distances for those patients in the cohort to support the analysis. In the next section, we will discuss in detail how patient distance vectors are instrumental in spatial accessibility computation. Essentially, a patient's distance vector contains the travel distance from the patient's residential location to every hospital considered by the study.

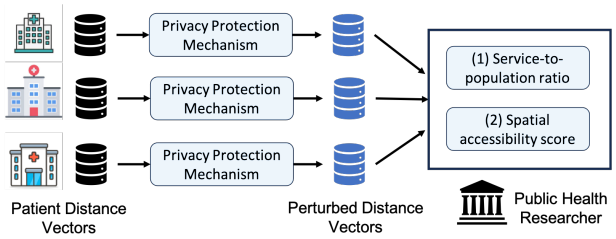


Figure 1: Problem Setting: private computation of spatial accessibility scores, e.g., for COVID-19 or diabetes patients.

A proposed privacy mechanism will sanitize each patient's distance vector and generate a perturbed vector; those perturbed vectors are shared with public health researchers for spatial accessibility computation (i.e., using a two-step approach as described in the next subsection). Note that we assume that a trusted data curator (e.g., healthcare providers) or a trusted personal device is available to run the proposed privacy protection mechanism. We will show that our proposed privacy methods only involve additive noises, which have low computational requirements.

We also consider a simplified setting for spatial accessibility, which requires only each individual's travel distance to the nearest hospital, as in a recent large-scale study [32]. In that case, the privacy mechanism generates a "noisy" shortest distance given a patient's distance vector, and public health researchers compute aggregate statistics of the noisy shortest distances, e.g., mean and standard deviation, as opposed to the two-step spatial accessibility.

3.2 Travel Distance-based Spatial Accessibility

In this section, we introduce a commonly used spatial accessibility measure, which is based on individuals' distance vectors, i.e., travel distances between an individual residence and every hospital considered in the study. Spatial accessibility captures the spatial interactions between the amount of supplies (e.g., the number of hospital beds, parks, healthy food stores) and demands along with the distance between the locations of health resources and those of

residential addresses. While spatial accessibility has been largely studied at area level, recent research has increasingly adapted spatial accessibility to individual-level analysis [2, 15, 17], to capture the individual variations in spatial access.

The two-step floating catchment area (2SFCA) method is one of the most widely used methods for measuring spatial accessibility to healthcare resources [20, 31]. Particularly, the generalized 2SFCA framework (G2SFCA) [31] accounts for *distance decay*, such that an increase in distance would lead to less service utilization. The G2SFCA method measures spatial accessibility in two steps and generates a score for spatial accessibility. A higher score indicates better accessibility.

In the first step, the G2SFCA method evaluates the catchment area of each facility, estimates the overall demands in the catchment area, and then generates the service-to-population ratio for each facility. In the second step, for each individual, the G2SFCA method identifies the facilities with the catchment area within which the individual lives and sums up the service-to-population ratios of all these facilities. A distance decay effect was assumed in both steps.

Specifically, for facility i, let S_i denote its capacity of supply (e.g., the number of hospital beds) and let $d_{k,i}$ denote the spatial distance between the residential address of individual k and the location of facility i. We further define the catchment area by imposing a threshold θ on the spatial distance. The service-to-population ratio of facility i is thus:

$$R_i = \frac{S_i}{\sum_{k \in \{d_{k,i} \le \theta\}} \mathbb{1} \cdot f(d_{k,i})} . \tag{1}$$

For individual k, the spatial accessibility score is computed as:

$$SA_k = \sum_{i \in \{d_{k,i} \le \theta\}} R_i \cdot f(d_{k,i}) . \tag{2}$$

Note that in both equations above, f is the distance decay function, which can be defined as:

$$f(d_{k,i}) = d_{k,i}^{-\beta} \tag{3}$$

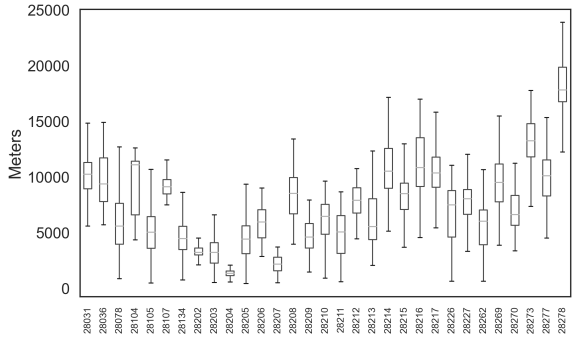
where $\beta > 0$.

As can be seen, an important prerequisite to computing the spatial accessibility scores is every individual's spatial distance to each facility, i.e., $d_{k,i}$'s. Let n denote the total number of facilities. We simplify the notation in the following whenever only one individual is concerned: each individual has a pre-computed distance vector $x = \{x_1, x_2, \ldots, x_n\}$, where each x_i denotes the individual's travel distance to facility i. In this study, all travel distances are computed as shortest routes on real road networks.

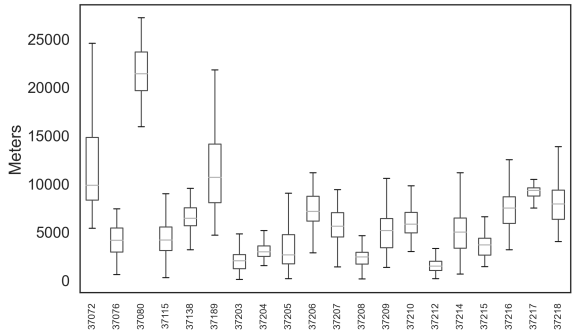
3.3 Exploratory Studies on Travel Distances

We present two exploratory studies to illustrate potential information leakage as a result of sharing travel distances. The following studies are conducted with population and healthcare facility data of Mecklenburg county, NC and Nashville, TN, the detail of which can be found in the Experiments section.

Distance to Nearest Hospital by Zip-code. Figure 2 plots the distribution of travel distances to the nearest hospital for individuals in each zip-code. We observe in both datasets that: (1) it is rare that two zip-codes have identical distributions; (2) some zip-codes have larger probability masses over certain ranges than other zip-codes (e.g., 28278 and 37080 in high distance ranges). Those observations indicate that travel distances to healthcare facilities may leak information about residential location, e.g., zip-code. For instance, if a Nashville individual must travel more than 20000 meters to reach the nearest hospital, it is very likely that the individual lives in zip-code 37080.

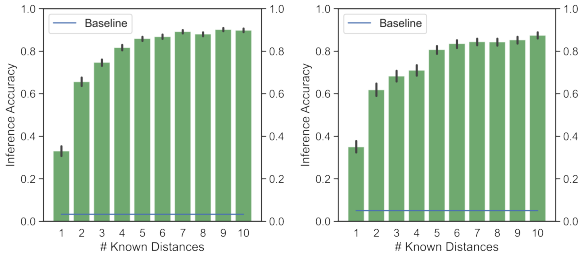


(a) Charlotte - Mecklenburg



(b) Nashville

Figure 2: Spatial travel distance (in meters) to the nearest hospital by zip-code.



(a) Charlotte - Mecklenburg **(b) Nashville**

Figure 3: Zipcode inference by spatial distances to a number of healthcare facilities.

Zip-code Inference with Known Travel Distances. In the following, we explore the information leakage associated with releasing distance vectors. Specifically, we would like to answer this question: *if knowing an individual's travel distances to k hospitals, how likely to infer the zip-code in which the individual lives?* To conduct this study, we form a public, population dataset B and a protected dataset D where $D \subset B$, to simulate a subset of individuals participating in a spatial accessibility study. Data sources and characteristics are described in detail in the Experiments section.

We generate a probability distribution of travel distances to each hospital in every zip-code using public data B. Given k, we randomly select k hospitals (i.e., a *query*), and estimate the zip-code for every individual in D with the most likely zip-code, which maximizes the joint probability of observing the travel distances to selected hospitals. We run 50 queries for each k (except for $k = 1$[2]) and report the accuracy results in Figure 3. The baseline accuracy with random guessing is $\frac{1}{\#zip-codes}$. It can be seen that knowing the travel distance to one specific hospital leads to above 30% accuracy for zip-code inference for both datasets. When increasing the knowledge to two hospitals, the accuracy grows to around 60%. It requires only distances to 4 hospitals for Mecklenburg and 5 hospitals for Nashville to reach above 80% accuracy.

Those results illustrate that sharing travel distances may leak information about individuals in the spatial accessibility study. That motivates us to develop provably private methods for sharing travel distances (in Section 4) and empirical privacy risk measures with powerful adversaries (in Section 5).

4 PRIVATE METHODS

In this section, we propose provably private algorithms to release travel distances for spatial accessibility studies. Specifically, we focus on sharing two computations: the first is the *distance vector*, i.e., travel distances from the input location to all facilities considered, which will be used to compute the spatial accessibility scores using variations of the 2SFCA method; the second is an input location's *distance to the nearest facility*, which will be aggregated as in [32]. While the latter is based on the former, we will show that it can be reported with significantly less privacy cost.

4.1 Metric Privacy

We first introduce the notion of metric privacy, which is adopted by our proposed privacy algorithms. In fact, differential privacy [5] has been widely adopted to protect individual records in statistical databases. Provable privacy protection is achieved with randomized mechanisms to provide a given level of indistinguishability between neighboring databases. However, a more general privacy notion is needed for sharing individual-level data, e.g., protecting input secrets which belong to an arbitrary domain. In such scenarios, it is meaningful to define a *distance* metric between secrets and guarantee a level of indistinguishability proportional of the distance. In [3], the authors extended the principle of differential privacy to arbitrary metrics. Let $\mathcal{X}$ denote an arbitrary set of secrets with a metric $d_{\mathcal{X}}$.

Definition 4.1. [3] A mechanism $K : \mathcal{X} \to \mathcal{P}(\mathcal{Z})$ satisfies $d_{\mathcal{X}}$-privacy, if and only if $\forall x, x' \in \mathcal{X}$

$$K(x)(Z) \leq e^{d_{\mathcal{X}}(x,x')} K(x')(Z) \quad \forall Z \in \mathcal{F}_{\mathcal{Z}} \tag{4}$$

where $\mathcal{Z}$ is a set of outcomes, $\mathcal{F}_{\mathcal{Z}}$ is a σ-algebra over $\mathcal{Z}$, and $\mathcal{P}(\mathcal{Z})$ is the set of probability measures over $\mathcal{Z}$.

Intuitively, $K(x)(Z)$ denotes the probability of mechanism K reporting Z given input x. With this generalized definition, a private mechanism K can be defined on any domain $\mathcal{X}$ and $\mathcal{Z}$. The authors of [3] argued that $d_{\mathcal{X}}$ can be derived by scaling a standard metric by a factor ϵ. For example, $d_{\mathcal{X}} = \epsilon \cdot d_p$ where d_p denotes the l_p metric. Furthermore, the authors showed that standard differential privacy is a special case of Def. 4.1, where $d_{\mathcal{X}} = \epsilon \cdot d_h$ and d_h denotes the Hamming distance between databases.

It is worth noting that geo-indistinguishability [1] is another instance of metric privacy. Specifically, the authors [1] proposed the Planar Laplace mechanism to randomize latitude and longitude coordinates, which satisfies the requirement of Equation 4 with $d_{\mathcal{X}}$ being the standard 2D Euclidean distance. In our empirical evaluation, we consider the Planar Laplace mechanism as an alternative solution, in which an individual's residential location is randomized first and the distance vector is computed based on the perturbed location.

4.2 Report Noisy Min Distance

Recall that for each individual, we can pre-compute a distance vector $x = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^n$, where x_i denotes the individual's travel distance to the i-th facility from home. Below, we present a randomized mechanism in Algorithm 1 to report the individual's distance to the nearest facility privately. Specifically, given parameter ϵ, each x_i is perturbed with a random noise drawn from Laplace distribution with 0 mean and $1/\epsilon$ scale; the facility index corresponds to the noisy shortest distance is reported. This algorithm has been inspired by the Report Noisy Max procedure for histograms [5], which satisfies standard differential privacy. It is important to note that the n noisy distances will not be released, except for the "winning" noisy distance which can be released at no extra privacy cost.

Algorithm 1 Report Noisy Min

Input: distance vector $x = \{x_1, x_2, \ldots, x_n\}$, privacy parameter ϵ
  for $i = 1 \cdots n$ **do**
    $\hat{x}_i \leftarrow x_i + Laplace(0, 1/\epsilon)$
  end for
  $i_{min} \leftarrow \underset{i=1\cdots n}{\operatorname{argmin}} \hat{x}_i$
Output: index of the noisy shortest distance i_{min}

The following theorem shows that our algorithm satisfies $(\epsilon \cdot d_1)$-privacy on $\mathbb{R}^n$, where d_1 denotes the L_1 metric.

Theorem 4.2. Report Noisy Min satisfies $(\epsilon \cdot d_1)$-privacy, where d_1 denotes the Manhattan Distance (or L_1 metric) for input vectors.

Proof. Let $x = \{x_1, x_2, \ldots, x_n\}$ and $x' = \{x'_1, x'_2, \ldots, x'_n\}$ denote any two input vectors and $x, x' \in \mathbb{R}^n$. Let r_i denote the random noise added to the i-th distance, $\forall i$. Fix any $i \in \{1, \ldots, n\}$, we will bound the ratio of the probabilities that i is selected by Algorithm 1 with x and with x'. Let r_{-i} denote a draw from $[Laplace(0, 1/\epsilon)]^{n-1}$ used for all the noisy distances, except for the i-th distance. We will argue for each r_{-i} independently. The notation $\Pr[i|e]$ denotes the probability that the output of Algorithm 1 is i, conditioned on e.

We define $r^* = \max_{r_i} : x_i + r_i < x_j + r_j, \forall j \neq i$. Having fixed r_{-i}, i will be the output of Algorithm 1 when input vector is x if and only if $r_i \leq r^*$, i.e., $\Pr[i|x, r_{-i}] = \Pr[r_i \leq r^*]$.

By definition of r^*, we have:

$$x_i + r^* < x_j + r_j \; \forall j \neq i \tag{5}$$

[2] for $k = 1$, we run a query for each hospital in the dataset.

$$\Rightarrow x'_i + r^* = (x'_i - x_i) + x_i + r^* < (x'_i - x_i) + x_j + r_j \quad (6)$$
$$= (x'_i - x_i) + (x_j - x'_j) + x'_j + r_j$$
$$\leq |x'_i - x_i| + |x'_j - x_j| + x'_j + r_j \; \forall j \neq i$$

Hence, if $r_i < r^* - |x'_i - x_i| - |x'_j - x_j| \; \forall j \neq i$, we have $x'_i + r_i < x'_j + r_j \; \forall j \neq i$. Furthermore, if $r_i < r^* - d_1(x, x')$, we have

$$r_i < r^* - \sum_{j=1}^{n} |x'_j - x_j| \leq r^* - |x'_i - x_i| - |x'_j - x_j| \; \forall j \neq i. \quad (7)$$

Thus, if $r_i < r^* - d_1(x, x')$, Algorithm 1 will report the i-th distance when input vector is x' and the noise vector is (r_i, r_{-i}), i.e., $\Pr[i|x', r_{-i}] \geq \Pr[r_i < r^* - d_1(x, x')]$.

As $r_i \sim Laplace(0, 1/\epsilon)$, the following probabilities yield:

$$\frac{\Pr[i|x, r_{-i}]}{\Pr[i|x', r_{-i}]} \leq \frac{\Pr[r_i \leq r^*]}{\Pr[r_i < r^* - d_1(x, x')]} \quad (8)$$
$$= \frac{exp(\epsilon \cdot r^*)}{exp(\epsilon \cdot (r^* - d_1(x, x')))}$$
$$= exp(\epsilon \cdot d_1(x, x')).$$

□

4.3 Distance Vector Laplace

To release the entire distance vector privately, we propose Algorithm 2 that adds Laplace noise to each distance. We show in Theorem 4.3 that the algorithm achieves $(n \cdot \epsilon \cdot d_1)$-privacy on $\mathbb{R}^n$.

Algorithm 2 Distance Vector Laplace

Input: distance vector $x = \{x_1, x_2, \ldots, x_n\}$, privacy parameter ϵ
  for $i = 1 \cdots n$ **do**
    $\hat{x}_i \leftarrow x_i + Laplace(0, 1/\epsilon)$
  end for
Output: noisy vector $\hat{x} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n\}$

Theorem 4.3. Distance Vector Laplace satisfies $(n \cdot \epsilon \cdot d_1)$-privacy, where d_1 denotes the Manhattan Distance (or L_1 metric) for input vectors.

Proof. Let $x = \{x_1, x_2, \ldots, x_n\}$ and $x' = \{x'_1, x'_2, \ldots, x'_n\}$ denote any two input vectors and $x, x' \in \mathbb{R}^n$. Let $\Pr[\hat{x}|e]$ denote the probability of Algorithm 2 reporting $\hat{x}$, conditioned on event e. We will show for any output $\hat{x} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n\}$, the ratio of probabilities of Algorithm 2 reporting $\hat{x}$ with x and with x' can be bounded.

Let $\{r_i\}$ and $\{r'_i\}$ denote the noise vectors drawn for x and x' respectively. As noise added to each distance is independently drawn from $Laplace(0, 1/\epsilon)$, we have:

$$\frac{\Pr[\hat{x}|x]}{\Pr[\hat{x}|x']} = \prod_{i=1}^{n} \frac{\Pr[r_i = x_i - \hat{x}_i]}{\Pr[r'_i = x'_i - \hat{x}_i]} = \prod_{i=1}^{n} \frac{exp(-\epsilon \cdot |x_i - \hat{x}_i|)}{exp(-\epsilon \cdot |x'_i - \hat{x}_i|)} \quad (9)$$
$$= \prod_{i=1}^{n} exp(\epsilon \cdot (|x'_i - \hat{x}_i| - |x_i - \hat{x}_i|))$$
$$= exp(n \cdot \epsilon) \cdot exp(\sum_{i=1}^{n} (|x'_i - \hat{x}_i| - |x_i - \hat{x}_i|))$$
$$\leq exp(n \cdot \epsilon) \cdot exp(\sum_{i=1}^{n} |x'_i - x_i|) = exp(n \cdot \epsilon \cdot d_1(x, x'))$$

□

5 EMPIRICAL PRIVACY MEASURES

Beyond theoretical guarantees, it is important to understand the level of practical privacy protection offered by our algorithms. In practice, spatial accessibility is evaluated for a set of participants who enrolled in a study; the distance vector or shortest distance can be either reported by the participant or computed by a trusted curator with the participant's residential address. In this section, we discuss two adversarial inferences that can be conducted by observing the shortest distance or distance vector. Similar to [27], we assume that study participants may be present in other publicly available data (i.e., attacker's background knowledge), such as voter registrations and open source datasets. In the following, we denote the set of study participants as D and the set of public individuals as B (such as voters), where $D \subset B$.

Membership Inference. A common privacy leakage is membership inference, in which an individual's participation in D is disclosed. In spatial accessibility studies, an individual's participation may leak sensitive health information, such as hypertension [17], COVID-19 [13], and diabetes [4]. As a result, membership inference may present a severe privacy risk to study participants.

By observing the released computations (either shortest distance or distance vector) for individuals in D, an informed adversary may launch the following membership inference attack against a target individual $t \in B$, in order to estimate whether t participates in D. Depending on the type of computation observed for D, the adversary computes the distance vector x^t (or distance to nearest facility x^t_{min}) for the target individual using the real residential address of t found in public data. The adversary then finds the best match among individuals in D to t, i.e., with a distance vector most similar to x^t (or shortest distance most similar to x^t_{min}) [3]. When there is a tie, one candidate is randomly chosen to be the best match for t, and the dissimilarity score is recorded for the match. Essentially, the match is recorded as $(t, s, \text{dissim}(t, s))$, where $t \in B$, $s \in D$, and $\text{dissim}(t, s)$ is computed for x^t and x^s (or x^t_{min} and x^s_{min}). The adversary repeats the same process for every target individual in B, generating $|B|$ matches along with dissimilarity scores, and selects $|D|$ matches with the lowest dissimilarities. To quantify the success of membership inference, we report the percentage of targets in the selected matches who actually participate in D.

Intuitively, the proposed membership inference may be more successful if all individuals in B (thus D) have unique distance vectors or shortest distances and are matched to themselves during inference. As a counter example, assume one participant $s \in D$ lives in an apartment building or shares a house with others. During inference, multiple individuals in B having the same travel distances as s could be matched to s with a dissimilarity score 0. The result of the membership inference is less accurate due to those false positives. While we cannot modify the geospatial distribution or population density in residential areas, we hypothesize that one effective solution to mitigate membership inference is to modify the released computations for D, such that targets in B are not matched to themselves during inference.

Address Inference. Residential addresses, such as street address and city, are considered identifiers of individuals and should not

[3] When measuring the dissimilarity between two distance vectors, the adversary may use L_1, L_2, or Dynamic Time Warping distances.

be shared according to HIPAA's Safe Harbor method. Therefore it is important to understand the risk of address inference for study participants in D, in order to protect their privacy.

In address inference, the goal of the adversary is to estimate the street address of a target individual t in D, upon observing the released computations for D. For every public individual $s \in B$, the adversary computes the distance vector x^s (or shortest distance x^s_{min}) using the real residential address of s. For a target individual $t \in D$, the adversary then finds the best match among all individuals in B, based on the measure dissim(t, s) also used for membership inference. The address of t is then estimated with the residential address of their best match. We adopt two quantitative measures to assess the risk of address inference. The first reports the percentage of participants in D whose street address is accurately inferred. The second reports the inference error (in meters) between the target's address and the estimated address, averaged among all participants in D.

Unlike membership inference, address inference may be successful even if a target individual t in D shares the same street address with multiple individuals in B. In that case, t would be matched to a fellow resident in B and the estimated street address would be accurate [4]. We hypothesise that modifying the released computations for participants in D will be a effective defense for address inference, as a target may be matched to a public individual with a different address during inference.

6 EXPERIMENTS

Data. The empirical evaluation adopts real population data of Mecklenburg county, NC and Nashville, TN. Voter registration data for Mecklenburg county is obtained from the North Carolina State Board of Elections, which contains the voter's name, registration status, street address, city, zip-code, along with other attributes. Open-source address data for Nashville is obtained from OpenAddresses [21], which contains distinct street addresses, zip-codes, and latitude/longitude coordinates. During pre-processing, we discard zip-codes with fewer than 1000 records in both datasets and further discard zip-codes dedicated to university campuses in Mecklenburg county (as individuals tend to use the same university address). The remaining zip-codes include 31 for Mecklenburg county and 20 for Nashville. Those zip-codes and their estimated boundaries are depicted in Figure 4. We randomly select 1000 records from each zip-code to form the final processed data. Hospital listings are obtained from the NC Division of Health Service Regulation and the TN Department of Health. We retrieve 21 hospitals for Mecklenburg county (in Mecklenburg and surrounding NC counties) and 16 hospitals for Nashville (in Davidson and Williamson counties). Hospital locations are highlighted in star in Figure 4.

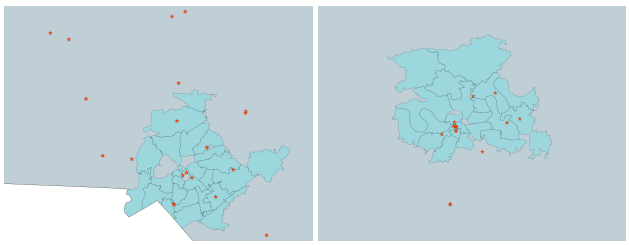
Geocoding and Routing. Hospital addresses and individual residential addresses are geo-coded with Nominatim. Travel distances in meters between individual address and each hospital are retrieved using OSRM APIs in driving mode. Both Nominatim and OSRM are based on OpenStreetMap data.

[4]Unit or apt. numbers are omitted in this study

Approaches. Let the complete processed data denote the public data B, which contains 1000 records from each zip-code in Mecklenburg county and Nashville. To simulate a small subset of individuals participating in research studies, we form D by randomly sampling 20 individuals from each zip-code in each dataset. The following approaches are applied to individuals in D to enhance the privacy of spatial accessibility studies. Among them, we consider `GeoInd` and `Clutering` as *input perturbation* approaches for perturbing the residential locations, and `Rounding` and our methods as *output perturbation* approaches for perturbing the distance vectors directly.

- GeoInd [1]: the Planar Laplace mechanism with parameter ϵ is applied to the residential location of each individual in D; the released distance vector and shortest distance of the individual are computed according to the perturbed location.
- Clustering: we devise this approach to generalize individual residential locations. We place residential locations in B into clusters of at least size k. For each individual in $D \subset B$, we perturb their residential location with the medoid of the corresponding cluster. The released distance vector and shortest distance of the individual are computed according to the perturbed location. With this approach, the adversary may not distinguish an individual from others in the same cluster. For empirical evaluation, we adopt hierarchical clustering with Euclidean distance, while other clustering methods and distances may also be adopted.
- Rounding: a common privacy-enhancing practice is to reduce the precision of released data. We apply rounding to the distance vectors and shortest distances for individuals in D such that only approximate information is preserved. For example, the spatial distance will be reported as multiples of a spacing parameter s, such as kilometers or miles. We assume that the adversary performs the same rounding on B in order to improve its inference success.
- Ours: we apply the proposed methods in Section 4 to individuals in D, denoted by `RNM` for releasing shortest distances and `DVLaplace` for releasing distance vectors. Both methods are associated with the privacy parameter ϵ.

Note that we repeat each experiment below for 20 runs to report average results and D is sampled independently for each run. We assume an informed adversary who knows the zip-code of each individual in D and thus can focus on the corresponding zip-code in B during inference.



(a) Charlotte - Mecklenburg **(b) Nashville**

Figure 4: Studied Zip-codes and Hospitals (shown in red stars) in Two Datasets.

6.1 Distance to Nearest Facility

We first examine the feasibility of privately releasing an individual's travel distance to the nearest healthcare facility. Results of Mecklenburg are discussed below and results of Nashville can be found in Appendix. In Figure 5, we report the MAE (mean absolute error) of the noisy shortest distances for individuals in D, with respect to their real shortest distances. We observe that stronger privacy levels (i.e., lower ϵ, higher s and k) incur higher accuracy loss for each method. `Rounding` incurs a predictable MAE around $s/2$ with s being the spacing parameter. Both `RMN` and `GeoInd` incur low MAEs (< 1m) with $\epsilon = 1$ or 2, as those randomized mechanisms introduce little noise to the input distances or coordinates. For high privacy settings, with $\epsilon = 0.001$ or 0.01, `GeoInd` incurs higher MAEs than `RMN`, which indicates a higher impact on accuracy for perturbing input locations. Lastly, increasing k for `Clustering` steadily increases the MAE from 50m ($k = 2$) to 860m ($k = 100$), and the amount of errors may depend on the density and spatial distribution of the population.

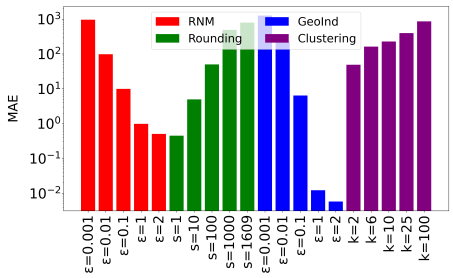
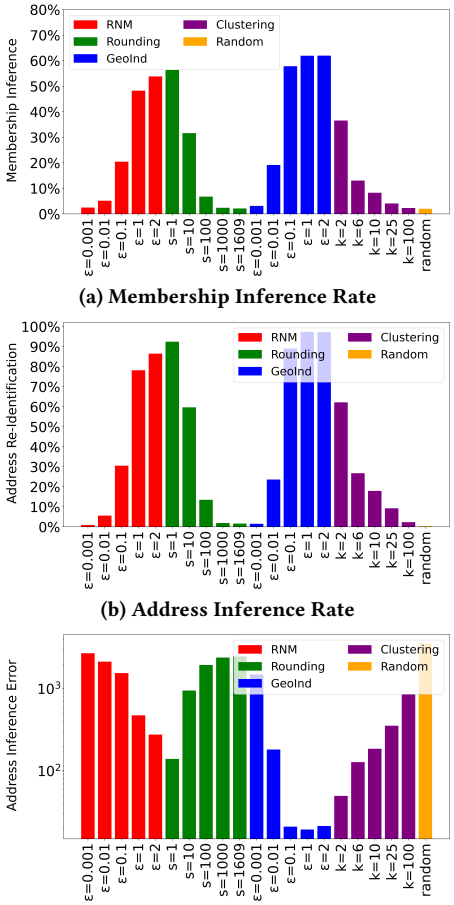


Figure 5: MAE (in meters) for Releasing Shortest Distances - Mecklenburg

Figure 6 reports the results of membership inference and address inference, using the shortest distances released by each privacy method for individuals in D. In addition, the baseline risk for each inference is included, denoted by `Random`, in which the adversary randomly pick individuals without considering the released shortest distances. Specifically, the baseline membership inference risk is 2% which is the ratio between $|D|$ and $|B|$; however, the baseline address re-identification risk and address inference error depend on the density and spatial distribution of the population, e.g., number of individuals in the same household and number of units per acre. Increasing the privacy levels (i.e., lower ϵ, higher s and k) will lead to lower membership inference risks, lower address re-identification risks, and higher address inference errors. `GeoInd` with $\epsilon = 1$ or 2 is seen to inflict highest membership inference and address re-identification risks (i.e., 62% and 97% respectively) and lowest address inference errors, as the perturbed location is likely to be truthful. `Rounding` the shortest distance to integers, i.e., $s = 1$, also incurs high privacy risks, i.e., 56% membership inference and 92% address re-identification.

To lower the empirical privacy risks $\leq 10\%$ in membership inference and address re-identification, we would need to adopt larger spacing $s \geq 1000$ for `Rounding`, $\epsilon \leq 0.01$ for `RMN`, $\epsilon = 0.001$ for `GeoInd`, and $k \geq 25$ for `Clustering`. An important observation is that input perturbation approaches lead to lower address inference errors than output perturbation. Specifically, the address inference errors for `GeoInd` with $\epsilon = 0.001$ and `Clustering` with $k = 100$ are 1494 meters and 354 meters, compared to 2707 meters (`RMN` with $\epsilon = 0.001$) and 2484 meters (`Rounding` with $s = 1609$). We consider the empirical privacy protection provided by `GeoInd` and `Clustering` weaker than that of `RMN` and `Rounding` in strong privacy settings. With $\epsilon \leq 0.01$, our method `RMN` outperforms `GeoInd` in both accuracy (i.e., MAE) and privacy.



(a) Membership Inference Rate

(b) Address Inference Rate

(c) Address Inference Error (in meters)

Figure 6: Privacy Evaluation for Releasing Shortest Distances - Mecklenburg

6.2 Distance Vector

In the following experiments, we examine the feasibility of publishing distance vectors for individuals in D using various privacy methods. Similarly, we report the accuracy and empirical privacy for Mecklenburg dataset in Figure 7 and Figure 8. Results of Nashville dataset can be found in Appendix.

The MAE for distance vectors is defined as $\frac{L_1(x,x')}{n}$, and averaged among all individuals in D. As it reflects the expected accuracy loss for each distance in the vector, results of `Rounding`, `GeoInd`, and `Clustering` in Figure 7 are similar to those of Figure 5. Our method `DVLaplace` incurs higher error than `RMN`, because `DVLaplace` perturbs each distance with parameter ϵ/n in order to guarantee $\epsilon \cdot d_1$-privacy for vectors. That illustrates the privacy saving (and accuracy gain) achieved by `RMN` to report only the shortest distance.

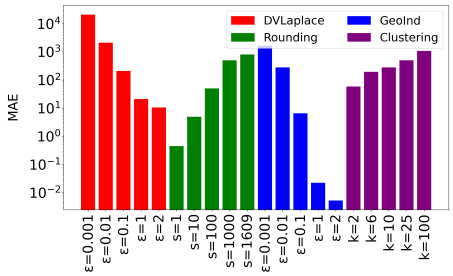
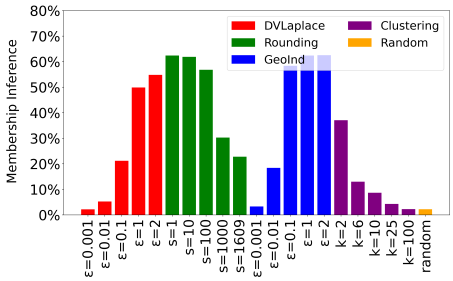
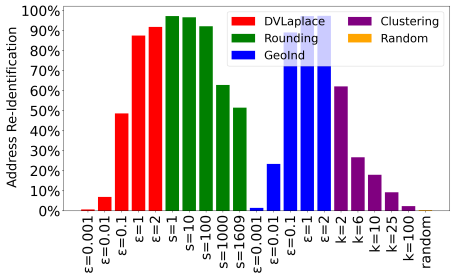


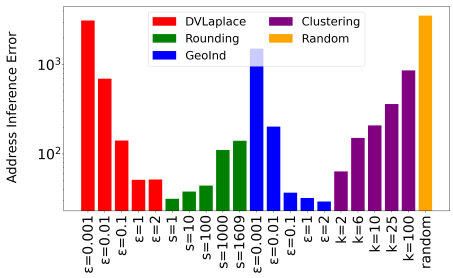
Figure 7: MAE (in meters) for Releasing Distance Vectors - Mecklenburg



(a) Membership Inference Rate



(b) Address Inference Rate



(c) Address Inference Error (in meters)

Figure 8: Privacy Evaluation for Releasing Distance Vectors - Mecklenburg

From Figure 8, it can be seen that releasing distance vectors, as opposed to releasing only shortest distances, does not significantly increase the empirical privacy risks for input perturbation approaches, i.e., `GeoInd` and `Clustering`. In those approaches, the released distance vector is calculated according to a perturbed location; therefore, the success of empirical privacy attacks is bounded by the privacy guarantees of location perturbation mechanisms. On the other hand, output perturbation via `Rounding` sees a significant privacy risk increase. Specifically, `Rounding` with $s = 1609$ incurs ~2% membership inference risk and address re-identification risk and 2848 meters address inference error when releasing only shortest distances (in Figure 6), and 23% membership inference risk and 51% address re-identification risk and 140 meters address inference error when releasing distance vectors (in Figure 8). Our approach `DVLaplace` does not incur higher privacy risks for releasing distance vectors, especially in strong privacy settings (i.e., $\epsilon = 0.001$), as it applies more stringent perturbation to each distance value. However, in weaker privacy settings, e.g., $\epsilon \geq 0.01$, the perturbed distance vectors may retain information about real data, e.g., via the dependence within a given vector. Specifically, we observer lower address inference errors with $\epsilon \geq 0.01$, e.g., 2136 meters for releasing shortest distances (in Figure 6) vs. 696 meters for releasing distance vectors (in Figure 8) with $\epsilon = 0.01$. Nonetheless, `DVLaplace` provides stronger privacy protection than `GeoInd` under the same ϵ. While `Clustering` with $k \geq 25$ provides good privacy protection, the method incurs a considerable overhead for computing clusters over the large population dataset B.

Table 1: MAE for SA Scores. *: parameter values that incur $> 100m$ in address inference error for both datasets.

DVLaplace			Rounding		
ϵ	Mecklenburg	Nashville	s	Mecklenburg	Nashville
0.01*	0.598	2.256	1000*	0.534	2.290
0.05*	0.321	1.147	500	0.192	1.413
0.1*	0.083	0.341	100	0.015	0.023
1	0.006	0.005	50	0.009	0.008

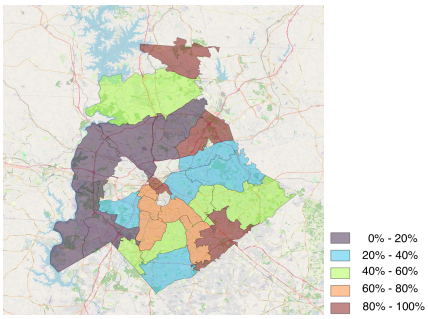
6.3 COVID-19 SA Scores

To evaluate the usefulness of the released distance vectors, we conduct a case study on spatial accessibility for COVID-19 patients. The case study simulates individual spatial accessibility scores for available hospital beds during the COVID-19 pandemic, using the G2SFCA method with $\theta = 10000$ meters and $\beta = 1$. Mean absolute error (MAE) is measured between real SA scores and privacy-enhanced SA scores, i.e., computed with perturbed distance vectors, among all individuals.

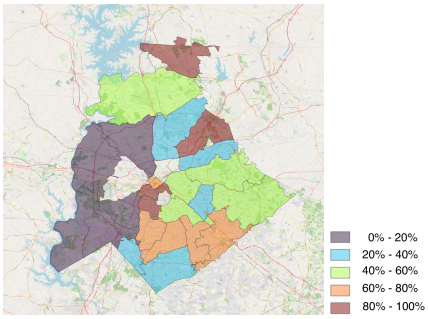
In the early months of the COVID-19 pandemic, 9730 cases were reported by the Mecklenburg County Health Department on June 28th 2020 and and 4427 cases were reported the Metro Nashville Public Health Department on May 26th 2020. With the published case counts by zip-code, we simulate COVID-19 patients by randomly sampling the target number of individuals in each zip-code and report the average results over 5 samples in Table 1. Note that `GeoInd` and `Clustering` are not included due to weaker empirical privacy protection and lower computational efficiency/usability, respectively. As privacy protection is relaxed, the MAE for SA scores is reduced, showing the improved usefulness for perturbed distance vectors. Linking the usefulness results to the privacy evaluation, we highlighted three ϵ settings of `DVLaplace` and one s setting of `Rounding`, in which an adversary's inference error on individual residential address is more than 100 meters on average. With similar usefulness results, `DVLaplace` is more private than `Rounding` in theoretical guarantees and empirical protection.

We visualize the real and privacy-enhanced SA scores by zip-code in Figure 9 and Figure 10. Zip-codes are classified into 5 quantiles based on SA scores, where higher SA scores indicate

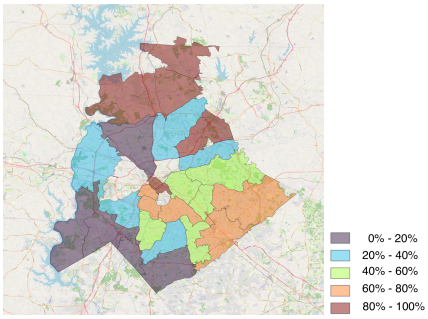
better access to health resources. Despite providing similar MAE errors, we observe that DVLaplace outperforms Rounding in preserving global SA patterns. For instance, in Figure 9, DVLaplace only mistakes zip-codes for adjacent color-codes/quantiles and misses one zip-code in the bottom 20% (color gray); Rounding places zip-codes in quantiles further from ground truth (e.g., from green to brown) and misses two zip-codes in the bottom 20% (color gray). In Figure 10, DVLaplace correctly classifies all zip-codes in the bottom 40% whereas Rounding makes several mistakes for the least resourced zip-codes, overestimating the spatial accessibility for zip-codes in north and west of Nashville. We hypothesize that DVLaplace preserves global patterns better because the aggregation tends to cancel out positive and negative perturbation noises, whereas Rounding introduces one-sided noise in distance values which may be propagated in aggregate analysis.



(a) Real

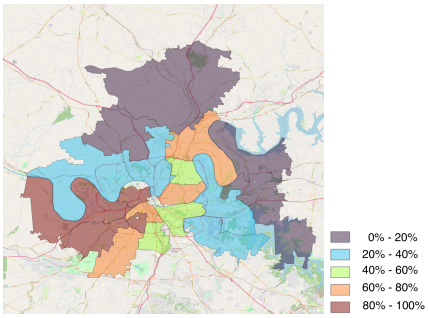


(b) DVLaplace with $\epsilon = 0.01$

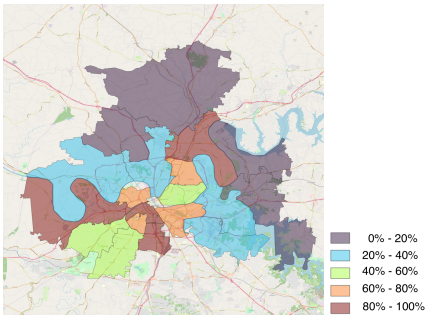


(c) Rounding with $s = 1000$

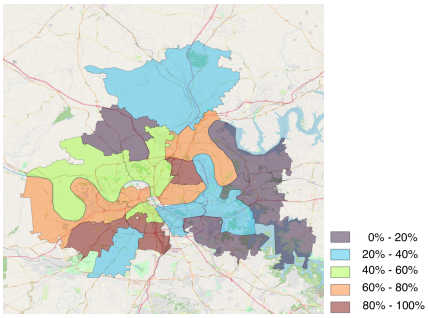
Figure 9: Spatial Accessibility Scores by Zip-Code - Mecklenburg: gray represents lowest SA scores and brown represents highest.



(a) Real



(b) DVLaplace with $\epsilon = 0.01$



(c) Rounding with $s = 1000$

Figure 10: Spatial Accessibility Scores by Zip-Code - Nashville: gray represents lowest SA scores and brown represents highest.

7 CONCLUSION

Our work takes a first step to enable private spatial accessibility studies by providing provable privacy protection to participants. Specifically, we have presented two empirical privacy risk measures to quantify the information leakage by sharing travel distances, which are essential in spatial accessibility analysis. To mitigate the privacy risks, we have proposed two privacy methods to report the shortest distance and distance vector at the individual level, and have analyzed their privacy guarantees in metric-based privacy. We have presented the empirical results obtained with real population and healthcare facilities data from two populous metro areas, which illustrate the usefulness and empirical privacy protection offered by our methods.

As for future work, we consider the following directions. Firstly, it is possible to develop secure aggregation protocols for computing spatial accessibility, i.e., to compute service-to-population ratios. It would be interesting to understand whether those intermediate

results incur any information leakage and to develop hybrid approaches that leverage both secure aggregation and metric privacy. Secondly, in this study, travel distances are estimated in the driving mode; future work may consider multiple transportation modes, e.g., walking and public transport, to estimate privacy risks and spatial accessibility scores. Lastly, it would be important to examine privacy risks in differential neighborhoods, e.g., at zip-code level, and the utility loss introduced by privacy methods in those neighborhoods. Our empirical results show that privacy methods may change the distribution of spatial accessibility scores. A large-scale study would be beneficial for an in-depth examination.

ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for their valuable feedback and Connor Burdick for participating in the early phase of this study. LF is supported in part by National Science Foundation grants CNS-1951430 and CNS-2144684. LB is supported in part by a National Human Genome Research Institute grant R00HG010493 and a National Library of Medicine grant R01LM013712. The opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2013. Geo-indistinguishability: Differential Privacy for Location-based Systems. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security* (Berlin, Germany) *(CCS '13)*. ACM, New York, NY, USA, 901–914. https://doi.org/10.1145/2508859.2516735

[2] James Bryant Jr and Paul L Delamater. 2019. Examination of spatial accessibility at micro-and macro-levels using the enhanced two-step floating catchment area (E2SFCA) method. *Annals of GIS* 25, 3 (2019), 219–229.

[3] Konstantinos Chatzikokolakis, Miguel E. Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi. 2013. Broadening the Scope of Differential Privacy Using Metrics. In *Privacy Enhancing Technologies*, Emiliano De Cristofaro and Matthew Wright (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 82–102.

[4] Andrew J Curtis and Wei-An Andy Lee. 2010. Spatial patterns of diabetes related health problems for vulnerable populations in Los Angeles. *International Journal of Health Geographics* 9, 1 (2010), 1–10.

[5] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.

[6] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. 1054–1067.

[7] Liyue Fan and Ishan Gote. 2021. A Closer Look: Evaluating Location Privacy Empirically. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems* (Beijing, China) *(SIGSPATIAL '21)*. Association for Computing Machinery, New York, NY, USA, 488–499. https://doi.org/10.1145/3474717.3484219

[8] Liyue Fan, Julius Marinak, and Ashley Bang. 2022. Co-Location and Air Pollution Exposure: Case Studies on the Usefulness of Location Privacy. In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on Location-Based Recommendations, Geosocial Networks and Geoadvertising* (Seattle, Washington) *(LocalRec '22)*. Association for Computing Machinery, New York, NY, USA, Article 6, 6 pages. https://doi.org/10.1145/3557992.3565991

[9] Dario Freni, Carmen Ruiz Vicente, Sergio Mascetti, Claudio Bettini, and Christian S Jensen. 2010. Preserving location and absence privacy in geo-social networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. 309–318.

[10] Mark F Guagliardo. 2004. Spatial accessibility of primary care: concepts, methods and challenges. *International journal of health geographics* 3, 1 (2004), 1–13.

[11] Hongbo Jiang, Jie Li, Ping Zhao, Fanzi Zeng, Zhu Xiao, and Arun Iyengar. 2021. Location privacy-preserving mechanisms in location-based services: A comprehensive survey. *ACM Computing Surveys (CSUR)* 54, 1 (2021), 1–36.

[12] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. 2014. Extremal mechanisms for local differential privacy. *Advances in neural information processing systems* 27 (2014).

[13] Jeon-Young Kang, Alexander Michels, Fangzheng Lyu, Shaohua Wang, Nelson Agbodo, Vincent L Freeman, and Shaowen Wang. 2020. Rapidly measuring spatial accessibility of COVID-19 healthcare resources: a case study of Illinois, USA. *International journal of health geographics* 19 (2020), 1–17.

[14] Kyusik Kim, Mahyar Ghorbanzadeh, Mark W Horner, and Eren Erman Ozguven. 2022. Assessment of disparities in spatial accessibility to vaccination sites in Florida. *Annals of GIS* 28, 3 (2022), 263–277.

[15] Jingjing Li and Changjoo Kim. 2020. Exploring relationships of grocery shopping patterns and healthy food accessibility in residential neighborhoods and activity space. *Applied Geography* 116 (2020), 102169.

[16] Rongxing Lu, Xiaodong Lin, Zhiguo Shi, and Jun Shao. 2014. PLAM: A privacy-preserving framework for local-area mobile social networks. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*. 763–771. https://doi.org/10.1109/INFOCOM.2014.6848003

[17] Jiajun Luo, Muhammad G Kibriya, Paul Zakin, Andrew Craver, Liz Connellan, Saira Tasmin, Tamar Polonsky, Karen Kim, Habibul Ahsan, and Briseis Aschebrook-Kilfoy. 2022. Urban Spatial Accessibility of Primary Care and Hypertension Control and Awareness on Chicago's South Side: A Study From the COMPASS Cohort. *Circulation: Cardiovascular Quality and Outcomes* 15, 9 (2022), e008845.

[18] Wei Luo and Yi Qi. 2009. An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & place* 15, 4 (2009), 1100–1107.

[19] Wei Luo and Fahui Wang. 2003. Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. *Environment and planning B: planning and design* 30, 6 (2003), 865–884.

[20] Matthew R McGrail. 2012. Spatial accessibility of primary health care utilising the two step floating catchment area method: an assessment of recent improvements. *International journal of health geographics* 11, 1 (2012), 1–12.

[21] OpenAddresses. [n. d.]. https://openaddresses.io/. Accessed: 2023-09-28.

[22] Jinwoo Park, Alexander Michels, Fangzheng Lyu, Su Yeon Han, and Shaowen Wang. 2023. Daily changes in spatial accessibility to ICU beds and their relationship with the case-fatality ratio of COVID-19 in the state of Texas, USA. *Applied Geography* 154 (2023), 102929.

[23] Vincent Primault, Antoine Boutet, Sonia Ben Mokhtar, and Lionel Brunie. 2018. The long road to computational location privacy: A survey. *IEEE Communications Surveys & Tutorials* 21, 3 (2018), 2772–2793.

[24] Apostolos Pyrgelis, Emiliano De Cristofaro, and Gordon J. Ross. 2016. Privacy-Friendly Mobility Analytics Using Aggregate Location Data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (Burlingame, California) *(SIGSPACIAL '16)*. Association for Computing Machinery, New York, NY, USA, Article 34, 10 pages. https://doi.org/10.1145/2996913.2996971

[25] Chenxi Qiu, Anna Squicciarini, Ce Pang, Ning Wang, and Ben Wu. 2022. Location Privacy Protection in Vehicle-Based Spatial Crowdsourcing via Geo-Indistinguishability. *IEEE Transactions on Mobile Computing* 21, 7 (2022), 2436–2450. https://doi.org/10.1109/TMC.2020.3037911

[26] Jinmeng Rao, Song Gao, and Xiaojin Zhu. 2021. VTSV: A privacy-preserving vehicle trajectory simulation and visualization platform using deep reinforcement learning. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. 43–46.

[27] Latanya Sweeney. 2002. k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems* 10, 05 (2002), 557–570.

[28] Zhuolin Tao, Yang Cheng, Qingjing Zheng, and Guicai Li. 2018. Measuring spatial accessibility to healthcare services with constraint of administrative boundary: A case study of Yanqing District, Beijing, China. *International Journal for Equity in Health* 17, 1 (2018), 1–12.

[29] Federico Thomas and Lluís Ros. 2005. Revisiting trilateration for robot localization. *IEEE Transactions on robotics* 21, 1 (2005), 93–101.

[30] Hien To, Gabriel Ghinita, Liyue Fan, and Cyrus Shahabi. 2016. Differentially private location protection for worker datasets in spatial crowdsourcing. *IEEE Transactions on Mobile Computing* 16, 4 (2016), 934–949.

[31] Fahui Wang. 2012. Measurement, optimization, and impact of health care accessibility: a methodological review. *Annals of the Association of American Geographers* 102, 5 (2012), 1104–1112.

[32] DJ Weiss, A Nelson, CA Vargas-Ruiz, K Gligorić, S Bavadekar, E Gabrilovich, A Bertozzi-Villa, J Rozier, HS Gibson, T Shekel, et al. 2020. Global maps of travel time to healthcare facilities. *Nature Medicine* 26, 12 (2020), 1835–1838.

A ADDITIONAL RESULTS

A.1 Distance to Nearest Facility

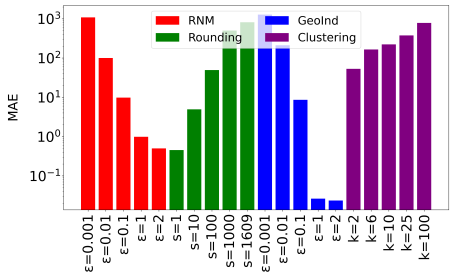
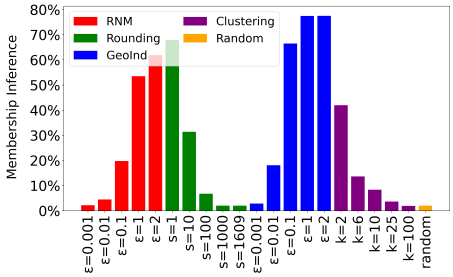
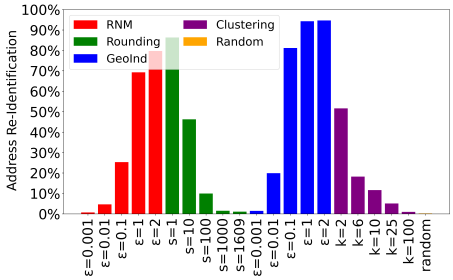


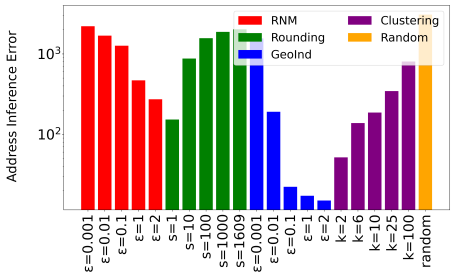
Figure 11: MAE (in meters) for Releasing Shortest Distances - Nashville



(a) Membership Inference Rate



(b) Address Inference Rate



(c) Address Inference Error (in meters)

Figure 12: Privacy Evaluation for Releasing Shortest Distances - Nashville

A.2 Distance Vectors

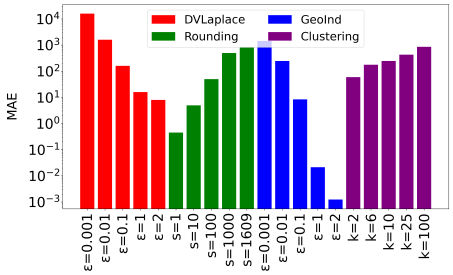
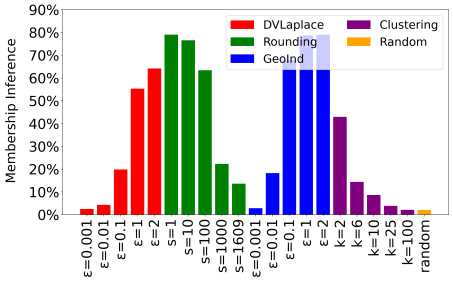
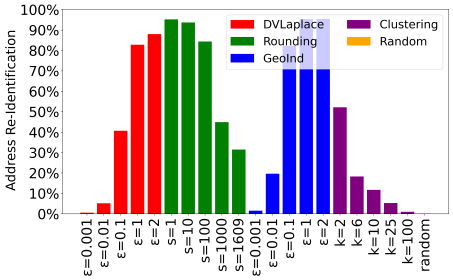


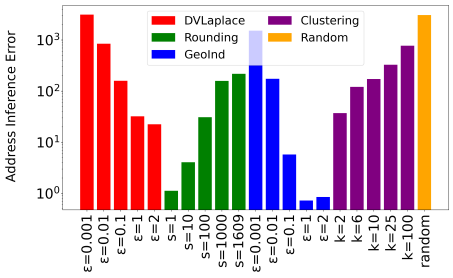
Figure 13: MAE (in meters) for Releasing Distance Vectors - Nashville



(a) Membership Inference Rate



(b) Address Inference Rate



(c) Address Inference Error (in meters)

Figure 14: Privacy Evaluation for Releasing Distance Vectors - Nashville