

5-2015

Hiding Private Locations by Anonymizing Data

Kelly J. Clifton
Portland State University, kclifton@pdx.edu

Steven R. Gehrke
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/trec_briefs



Part of the [Civil and Environmental Engineering Commons](#), and the [Transportation Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Clifton, Kelly J. and Gehrke, Steven R., "Hiding Private Locations by Anonymizing Data". OTREC-RR-489. Portland, OR: Transportation Research and Education Center (TREC) 2015.

This Report is brought to you for free and open access. It has been accepted for inclusion in TREC Project Briefs by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.



NITC

NATIONAL INSTITUTE
for TRANSPORTATION
and COMMUNITIES

Web: <http://nitc.us>

HIDING PRIVATE LOCATIONS BY ANONYMIZING DATA

Researchers explore ways of masking private locations in the interest of making useful data publicly available.

The Issue

Public agencies spend vast amounts of money collecting information in household travel surveys. In exchange for their participation, survey respondents are guaranteed anonymity. Detailed geospatial referencing of the home, work and other travel destinations is common practice. Such data can be of enormous use to planning professionals in creating travel demand models, but this must be balanced with the need to keep individuals' information confidential. To protect confidentiality, data are often aggregated to a geographic level before they are publicly shared.

This limits the usefulness of the information. Details are lost when data are aggregated. For example, walking trips can be affected to a large degree by the built environment of a particular neighborhood. If all walking trips are aggregated up to a larger zone, then questions about how people's mode choices were affected by their immediate surroundings cannot be answered. To allow more precise data to be widely distributed without sacrificing participants' anonymity, researcher Kelly Clifton of Portland State University examined other methods of masking sensitive locations.

The Research

Clifton and her research team looked into techniques known as geomasking, or geographical perturbation. One promising method, known as the donut masking technique, ensures that a masked record is moved a minimum distance from its original location so that the new position is not randomly assigned too close to its original location.

THE ISSUE

Household travel survey data can be useful, but its use must be balanced against the need for confidentiality.

THE RESEARCH

Researchers tested "donut masking," a technique that:

- Preserves the spatial sensitivity of survey data;
- Minimizes the risk of data disclosure;
- Requires careful application on the part of the data custodian.

IMPLICATIONS

When used with parameters appropriate to the urban context, the donut masking technique stands to provide planners with useful data while preserving the privacy and anonymity of survey participants.

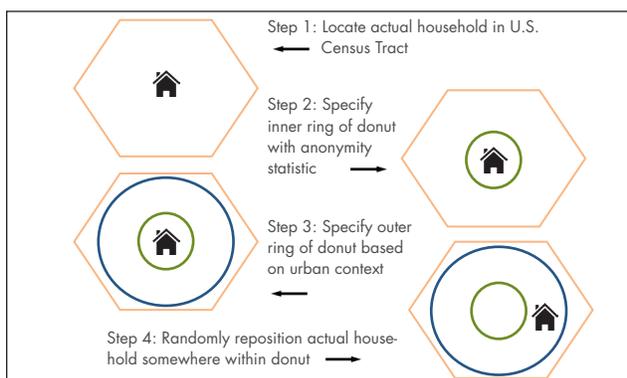
Photo: Home with address numbers anonymized

In the technique, a double circle or “donut” is defined around each protected point. The inner ring of the donut, with the protected location in the center, is the anonymity zone: the public dataset will not show the point anywhere within that circle. The radius of the inner ring is defined by an “anonymity statistic” determined by the data custodian, while the boundary of the donut’s outer ring is selected based on the context and population density of the area. In dense urban areas, the outer ring may be defined as being an accessible walking distance from the central location. In rural areas, it may need to extend further out to ensure minimal disclosure risk. Once the donut is defined, the data points are then randomly redistributed on a map so that they fall somewhere within the donut, between the inner and outer rings.

To test the donut masking technique empirically, Clifton’s team used it to explore the connections between 4,824 households and five measures of the built environment in the Portland, Oregon metropolitan area. They found that data custodians using this method must be aware of a sensitive “tipping point” between disclosure risk and data utility, and must define the donut accordingly. When used with appropriate parameters, the donut masking technique successfully masked geographic locations while at the same time retaining the data’s spatial relevance and fine resolution.

Implications

This project offers an introductory examination into the relationship between disclosure risk and data utility in the context of household travel survey data. The use of a geographic perturbation method such as the donut masking technique offers an exciting prospect for researchers who are interested in investigating the complex links between non-motorized travel and a household’s surrounding built environment. Yet, before disaggregate household travel survey data may be widely disseminated for public use, individuals with access to these fine-scaled



Steps in applying the donut masking technique

This graphic provides an illustration of how the donut masking technique may be applied to a dataset. Considerable attention must be given to the selection of the radius of the two rings.

data sources have

a responsibility to introduce a fitting level of “noise” into the dataset. It should be high enough to preserve respondent anonymity while not so high that researchers using the data would be led toward inaccurate conclusions. Clifton’s report takes a deep look into these trade-offs.

The research uncovered several trends regarding the interplay between data retention, disclosure risk and data utility. It also resulted in the development of some rules of thumb for future application of this technique, including the recommendation for its use mainly in urban contexts, as it is in dense, multimodal areas that the donut masking technique yields the lowest disclosure risk and the highest level of data utility.

PROJECT INFORMATION

TITLE: Wider Dissemination of Household Travel Survey Data Using Geographical Perturbation Methods

LEAD INVESTIGATOR: Kelly Clifton, Ph.D., Portland State University

PROJECT NUMBER: 2011-489

CONTACT: Justin Carinci, 503-725-2843 carinci@pdx.edu

MORE INFORMATION

<http://nitc.trec.pdx.edu/research/project/489>