3. Unveiling Invisible Barriers

Minwei Zhao <ucfamz0@ucl.ac.uk>
With: Stephen Law and Ivica Obadic

ABSTRACT

SITUATION: Jane Jacobs urban theory is one of the most influential in urban design. It proposes, in addition to the core elements of vitality, such as diversity of land use and concentration of people, that there are certain urban elements that are ambivalent with respect to vitality. Such elements are called border vacuums and they can have both positive and negative influence on vitality, and also on residents' health. Border vacuums are areas within a city or between urban and suburban areas that lack vitality, activity, or development. Despite the spatial autocorrelation observed in nearby areas, invisible barriers such as hills, highways, or other physical features may lead to significant disparities in health and socio-economic outcomes between adjacent areas.

COMPLICATION: The suggested interplay of border vacuums with various sociodemographic and health outcomes has not been studied empirically before.

PROPOSAL: This research project investigates the concept of "border vacuums" (see Jane Jacobs urban theory) in urban areas and their influence on health and socio-economic outcomes. This study proposes to develop a model that predicts health and socio-economic outcomes considering the distance between areas, while also analyzing residual values to identify and visualize these invisible barriers using satellite data.

We aim to do so by predictive modelling integrating satellite image and other types of features in linear models or spatially explicit models such as Graph Neural Networks (GNNs). By constructing a graph for the Lower Layer Super Output Area (LSOA) network in MedSat, incorporating connections between LSOAs, the project explores the potential of GNNs to improve prediction accuracy compared to the models reported in the MedSat paper, such as SLM and LightGBM. The spatial network can either use adjacency of LSOA, potentially incorporate the street network weights or apply the percolation theory [3].

An area's health and socioeconomic outcomes are then predicted from the outcomes of its neighboring areas. For example, calculate the average outcome for Area A as the sum of outcomes for its neighboring areas divided by the total number of neighboring areas (|n_A|), where n_A represents Area A's neighboring areas. By studying the residual value for each area, representing the difference between the predicted outcomes based on its neighbors and observed health and socio-economic outcomes, we aim to explain what drives these unexpected differences.

Upon extracting relevant features from satellite images, such as land cover types, built environment characteristics, and natural barriers (e.g., rivers, mountains), we will correlate the residual values with features extracted from satellite data to identify potential invisible barriers contributing to the observed disparities. We can also explore the value of the percolation theory on the street network and on its intersections method.

The novelty of the project lies not only in the application of a novel model to satellite imagery but also in its exploration of underutilized multispectral imagery in public health, and, notably, in grounding our results in urban theory (i.e., that of Jane Jacobs).

Check IC2S2 abstract with barriers using mobile phone data.
If in an area you have more border vacuums, it is negative. It might be related to poorer sociodemographic conditions.
If LSOA is a good unit?

STEPS

1. Familiarize yourself with MedSat papers and datasets.

    a. MedSat paper [1] data [2]
2. Defining spatial network for prediction:
    a. Adjacent neighbors or K-nearest neighbors
    b. Incorporating street network connection as weights.
    c. Spatial Network based on Percolation theory [3].
3. Predict medical prescriptions using:
    a. [Spatial] Regression: Area A's health and socioeconomic outcomes based on the outcomes of its neighboring areas. For simplicity, calculate the average outcome for Area A as the sum of outcomes for its neighboring areas divided by the total number of neighboring areas (|n_A|), where n_A represents Area A's neighboring areas.
    b. GNN.
    c. Other?
4. Calculate the residual value for each area, representing the difference between the predicted and observed health and socio-economic outcomes. More formally, $r_i = y_i - \hat{y}_i$ where r is the residual and y is the outcome variable.
5. Extract relevant features from satellite images, such as land cover types, built environment characteristics, and natural barriers (e.g., rivers, mountains).
6. Correlate the residual values with features extracted from satellite data to identify potential invisible barriers contributing to the observed disparities.
7. Write the paper.

The idea: for A, neighboring B, C, and D areas. The outcome for A is the average for B, C, and D. If the difference between the prediction and true value is large (accuracy is low).

With Elsa's method, we know for 2 points if they are disconnected (IC2S2 abstract did it based on mobile phone data). The detection of disconnected points is the KEY issue. Do we have mobility data?

We can ask Elsa for her data.

In the end we could crowdsource or do overlay OSM and similar data.

Resources

[1] Scepanovic, S., Obadic, I., Joglekar, S., Giustarini, L., Nattero, C., Quercia, D., & Zhu, X. (2024). MedSat: A Public Health Dataset for England Featuring Medical Prescriptions and Satellite Imagery. Advances in Neural Information Processing Systems, 36.

[2] https://mediatum.ub.tum.de/1714817

[3] Arcaute, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, A. P., & Batty, M. (2016). Cities and regions in Britain through hierarchical percolation. Royal Society open science, 3(4), 150691.