Mobility Mining – Graph Theory applied to assess COVID Pandemic effects on Users' Activity Space Relations using Social Media data

Marcos A. P. Motta

marcosmotta@edu.ulisboa.pt

–

Supervisor(s): Luis de Picado Santos1 [1], Claudia Soares2 [2], João Paulo Costeira [3]

1. CERIS/IST/ULISBOA
2. NOVA LINCS, Computer Science Department, NOVA School of Science and Technology
3. ISR/ULISBOA

MIT Portugal

2022 Annual Conference

1.Motivation: Traditional travel demand surveys have been feeding travel demand models (TDM) and can be considered the mainstream in transportation planning. Nevertheless, some scholars have been arguing some drawbacks derived from human memory failure in reporting travel patterns. Not to mention constantly rising costs.

Some alternative data sources have been tested to deal with both reported drawbacks, including Social Media as a primary or complementary data source. In addition, social media data presents some desirable characteristics for mobility studies, like longitudinal data and extensive samples, allowing the social network dimension to be considered.

This poster investigates COVID-19 Pandemic impact on users' activity space, using a Graph Theory Approach, highlighting changes in how users have spread within the city 's regions. The sample starts in 2019 (January) and ends in 2022 (June).

2.Research Problem: Activity Space is characterized *as the local areas that people travel within while conducting their daily activities* (Xujiao Wang & Yihong Yuan (2021)). In addition, Golledge and Stimsom (1997) pointed out that *activity spaces are determined by commonly visited locations and the links between them*.

There are many approaches to physically represent activity space, like: convex hull and ellipse-based methods. In addition, there are some indicators to assess and compare, like Radius of Gyration (ROG), Minimum Spanning Tree (MST), or Entropy (Xujiao Wang & Yihong Yuan (2021)).

We will physically represent the activity space for each user as the set of census tracks where users issue tweets. Using a Graphs Theory approach, we will group users that share all or part of their activity space, building geographic communities. The research question is: were the communities' structures affected by COVID 19 Pandemic?

The resulted geographic communities were assessed using *Wilcoxon rank sum test* with the follow Null Hypotheses:

h_0 - There is no difference in communities' structure (degree centrality), considering pandemic phases. Hence the samples belong to the same population.

$$ROG = \sqrt{\frac{1}{n}(r_i - r_m)}$$

Radius of Gyration

$$E = -\sum_1^N p_i log_2 p_i$$

Entropy

n - *Number of census track*

r_i - the geographical coordinate of each census track centroid

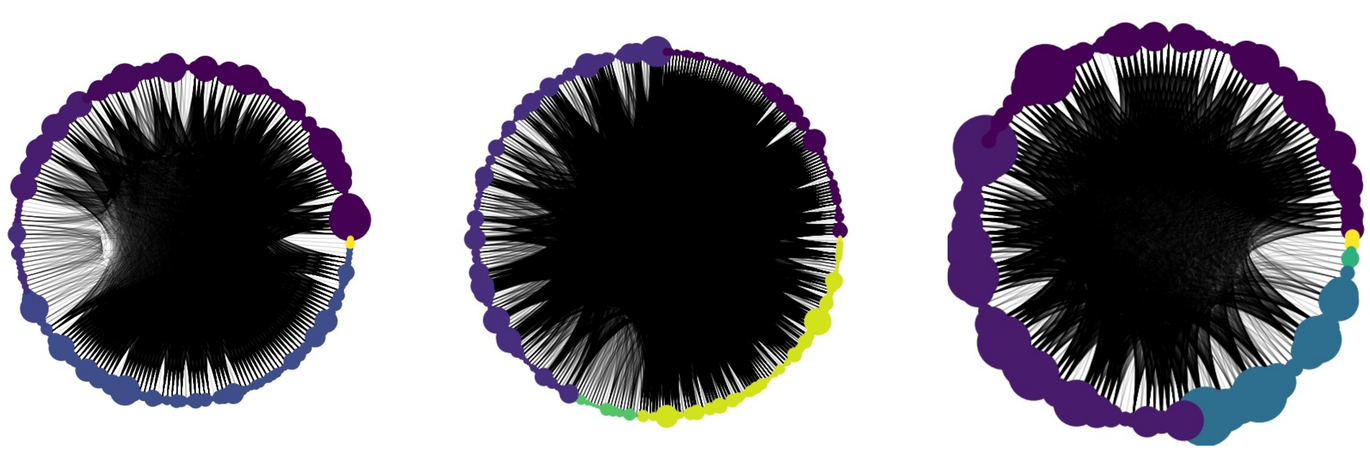
r_m - centroid of all check-in points of a user.

p_i - probability of a user checking in at the same census track i

N - the total number of census track where this user checked in.

3.Graph Theory: Bi-partite graphs represent the relationship between sets of different natures, like, clients x products, actors x movies or genes x diseases. One property of Bi-Partite Graph is projection, which is the relationship that arises in one set derived from the other. For example, a famous projection application is the Amazon problem, where a user's set connects a set of products. The projection of products' set over clients' set allows identifying clients that have purchased the same products. After placing clients that buy the same products inside clients' communities using an algorithm like Louvain Community detections, the communities' structure can be assessed based on centrality concepts. Here we have opted to use degree centrality to determine the community's structure.

4. Users Communities by Pandemic Phases:



Communities	*Pandemic Open Phase*	*Lockdown Pandemic Phase*	*Pre-Pandemic Phase*
Nodes	218	339	165
Edges	7,659	14,715	5,215
DG Median	0.5207	0.2307	0.5548
Periods (days)	08/02/20 – 01/14/21 (341) 09/30/21 – 06/22/22 (265)	03/19/20 – 08/01/20 (135) 01/15/2021 - 09/30/21 (258)	01/01/19 – 03/18/20 (442)

5.Results: The Wilcoxon rank sum test was conducted using all users inside each graph. Then, using only present users in both tested phases. The results indicate that with (α=5%) we can affirm that there is no difference between community structure in the pre-pandemic and open phase for both tests. However, the lockdown phase comparison presents different results for the selected samples. Therefore, we must reject H_0 for both comparisons using all users. Nevertheless, when using only the same users, we can't affirm that there is a difference inside the network structure between the lockdown and pre-pandemic phases.

Samples\P-values	PRE x LOCK	PRE x OPEN	LOCK x OPEN
All users	6.202141451460366e-07	0.8624368627880675	0.003284548102899408
Same Users	0.12494606062136765	0.05161506577569225	0.028025532021565346

6.References:

Xujiao Wang & Yihong Yuan (2021) - Modeling User Activity Space from Location-Based Social Media: A Case Study of Weibo, The Professional Geographer, 73:1, 96-114, DOI: 10.1080/00330124.2020.1803090
Golledge and Stimsom (1997) - Spatial behavior: A geographic perspective. New York: Guilford.

Funded by:

FCT Fundação para a Ciência e a Tecnologia

MIT Portugal

under the Doctoral Grant FRH/BD/151374/2021 | MIT Portugal Program Sustainable Cities