Urban data and spatial segregation: analysis of food services clusters in St.Petersburg, Russia

Aleksandra Nenko, Artem Konyukhov, Sergey Mityagin
Institute for Design and Urban Studies ITMO University
Saint-Petersburg, Russia.

Abstract

This paper presents an approach to study spatial segregation through clusterization of food services in St.Petersburg, Russia, based on analysis of geospatial and user-generated data from open sources. We consider a food service as an urban place with social and symbolic features and we track how popularity (number of reviews) and rating of food venues in Google maps correlate with formation of food venues clusters. We also analyze environmental parameters which correlate with clusterization of food services, such as functional load of the surrounding built environment and presence of public spaces. We observe that main predictors for food services clusters formation are shops, services and offices, while public spaces (parks and river embankments) do not draw food venues. Popular and highly rated food venues form clusters in historic city centre which collocate with existing creative spaces, unpopular and low rated food venues do not form clusters and are more widely spread in peripheral city areas.

Keywords: Urban environment, Urban data, Spatial segregation, Food services, Food venue popularity, Food venue rating

1. Introduction

In the context of postindustrial economy mass consumption of services and goods is the utmost component of the everyday life of the city dwellers and the quality of services has become an unquestionable value. Contemporary urban lifestyle demands accessible, available, and variable services. Economic development follows the tendency of services complication and diversification.

Digitalization of urban life and growth of possibilities to communicate about the urban services online between the producer and consumer allows producers to promote and position their services more efficiently, clients to browse and translate feedback about the services, and researchers to trace trends in service development based on urban data coming from various data sources. Online platforms which are expressing location, type and score of urban services by expert or peer-to-peer reviews, such as Google maps, Foresquare, Instagram, become significant guides of consumption processes. Our paper lies in line with the corpus of research which digs into the possibilities of applying user-generated geolocated data to analyze urban processes.

Service diversity and equal distribution across the urban space is considered to be a value, as far as it facilitates satisfaction of different users and improves overall subjective well-being (Jacobs, 1961, Sennett, 1977). At the same time urban space is segregated, while it is subject to the laws of capitalistic economy (Lefebvre, 1991; Harvey, 2001, 2012; Zukin, 1989). For this reasons considering segregation of services in the city acquires topicality.

We pursue the thesis that an urban service is not just an economic facility or an organization which exist in vacuum, but a "place" in urban space which attracts people and proposes a certain lifestyle. Services as places contain social, economic, symbolic and environmental characteristics. In this paper to analyze classes of services we consider symbolic parameters such as popularity (number of reviews) and rating and environmental parameters of venue location. The object of this study are food services as one of the most representative FMCG (fast moving consumer goods) services spread around urban space.

2. Literature Review

Inequality of food services distribution has won attention of scholars from 1990s. Cummins and Macintyre (1999) food stores in urban areas are distributed unequally. Wrigley (2009) proves that cities and regions might even contain "food deserts" - areas with restricted access to food stores and food venues, which emerge due to unequal distribution of food stores. The latter tend to locate more in central areas of the cities, than in non-central quarters, forming "ghettos" devoid from food. Global processes of urbanization, such as gentrification, can lead to emergence of paradoxical situations, such as expensive shops and food venues next to the dwellings of the poor or disadvantaged (McDonald and Nelson, 1991). Kelly and Swindel (2002) argue that sufficient diversity in services, in particular, food venues, leads to the improvement of quality of life and citizen satisfaction. Public and urban development policies should be focused on provision of the equal access to basic goods and services based on indexes of price and availability (Donkin et. al., 2000). Porta (Porta et. al. 2009; 2011) shows that street centralities are correlated with the location of economic activities and that the correlations are higher with secondary than primary activities.

3. Research Design and Dataset

We analyze spatial distribution of food services based on clusterization techniques and on correlation analysis with environmental characteristics described above. Formation of different classes of food venues is considered an indicator of service variety as well as of spatial segregation in the city in terms of quality of services.

The key research questions are: is there any evidence for correlation between the environmental characteristics and spatial clasterization of food services? Is there any evidence for spatial clusterization of food services according to symbolic properties of the places, in particular, parameter of venue popularity or rating?

Hypothesis 1. Clusterization of food services is correlated with characteristics of the built environment: food venues tend to collocate with (a) commercial function, (b) historical areas and (c) public spaces.

Hypothesis 2. Popular high rated venues cluster in favorable environmental conditions (collocate with objects described above) while unpopular low rated venues locate in unfavorable environmental conditions.

Hypothesis 3. Popular high rated venues tend to cluster while they “keep” the quality of service together.

The dataset on St.Petersburg food venues was parsed from Google maps open source via API and contains 4496 items. Google maps were chosen as a resource while it is well-spread in Russia: 85% of Russian smartphones run on Android system with preinstalled Google maps (Sharma, 2017). Usually places records provide venue ID, its name, geographical coordinates, rating, and number of reviews. However for our dataset only 2327 records contain information on venue rating and number of reviews. We consider formation of venues clusters by analyzing their “popularity” defined through 2 parameters: (a) number of reviews given by venue clients, (b) rating of a food establishment from 1 to 5 points given by its clients. Every time client leaves th place Google asks her to leave a review and rate the venue. Google company doesn't explain how the actual calculation of the rating is processed, it might be calculated as a Bayesian average (Blumenthal, 2014). Number of reviews in the dataset distribute as follows: mean - 54.24, standard deviation - 203, median - 9. During data processing venues with number of reviews beyond 3 sigma (mean average deviation) distance from the mean were removed. Ratings distribute as follows: mean - 4.26, std - 0.74, median - 4.4. It was detected that venues with rating less than 4,8 points consistently have few reviews (for 5 points ratings no more than 15 reviews) and they were removed from the dataset.

To allocate spatial clusters of food places we have applied DBSCAN (density-based spatial clustering of applications with noise) algorithm. DBSCAN allows to deliberately chose the size of the clusters and exclude single objects and is often used for spatial clusterization tasks (Ester et. al., 1996; Kisilevich et. al., 2010). The main parameters of the algorithm are (a) minimal number of objects in a cluster and (b) neighborhood radius. To define optimal parameters for clusterization we have allocated already existing food and drinking clusters in different parts of St.Petersburg city (Lenpoligraphmash, Loft project Etaji, Golitsyn Loft, etc.) which were to appear during clusterization procedure. Minimal number of objects in existing clusters is 5. Optimization procedure has shown the optimal radius of 100 meters (when existing officially defined clusters appear on the map).

The dataset for functional objects was derived from Foursquare with its prescribed categories of venues: “Arts & Entertainment”, “College & University”, “Food”, “Nightlife”, “Office”, “Residence”, “Shop & Service”, “Travel & Transport”. The overall dataset for St.Petersburg is 166 thousand functional objects. Additionally polygons of industrial territories, rivers and parks were parsed from OpenStreetMap open source.

For each food venue functional objects in 100 m radius were defined and assigned categories (1 - objects of this category are present in the list of neighbors, 0 - objects of this category are absent).

To define probabilities of food venue location next to the functional objects the mean was calculated for each of functional categories. For each functional category Spearman correlation coefficient was calculated between venue rating and this category object presence in 100 m radius from the venue.

To define collocation of food venues with certain popularity and rating (i.e. classes of venues) we have calculated average rating for their neighboring food venues.

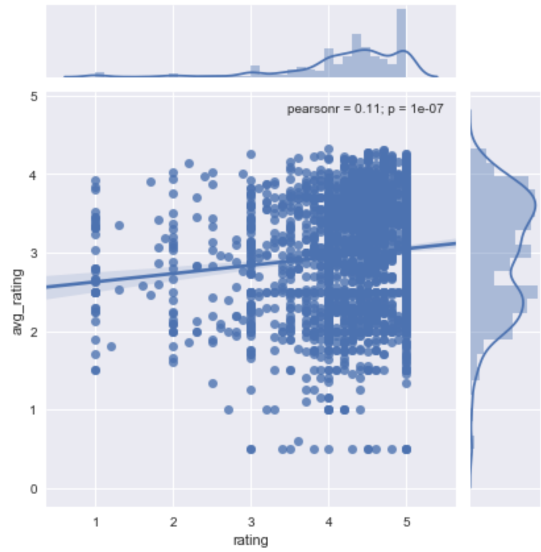

Figure 1. Distribution of food venues rating and average rating of their neighbours

Figure 1 shows that at least two different classes of venues are present in the dataset. We have applied Ward's Hierarchical Clustering Method and have set "rating" and "average rating" parameters for clusterization (Ward, 1963; Murtagh & Legendre, 2014) and have received three classes of food venues: (1) low rated venues which collocate with low rated neighbors, (2) high rated venues which collocate with high rated neighbors, (3) high rated venues which collocate with low rated neighbors (Figure 3).

4. Results

Calculation of the proportion of food venues which collocate with functional objects give the following results: shops & services - 0.776023, offices - 0.608319, (other) food venues - 0.557829, residential areas - 0.273577, arts & entertainment venues - 0.232651, travel & transport - 0.200845, outdoor activities - 0.199066, parks - 0.169706, colleges & universities - 0.144795, nightlife - 0.134342, industrial territories - 0.081851, river embankments - 0.051379. Hypothesis 1 has partly proved: functional objects such as shops, services and offices tend to collocate with food venues, while public spaces do not. As for the historical objects - see Figure 4 below.

Hypothesis 2 has not proved: no correlation was detected between presence of discussed functional objects and venue rating and popularity (Figure 2).

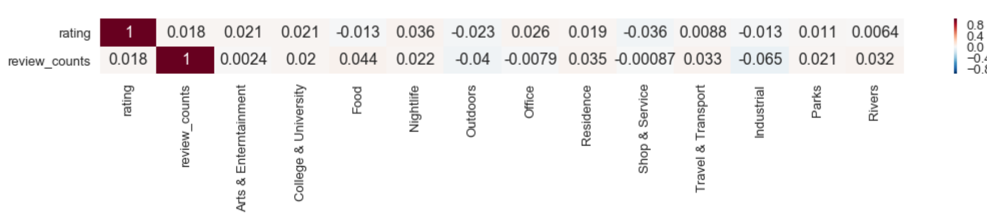

Fig. 2. Matrix of correlation between environmental parameters, food venue rating and popularity

Figure 3 shows the three classes of food venues described above and how they collocate with functional objects.

Venue class	№ 1	№ 2	№ 3
rating (mean)	2,435841	4,542263	4,316224
average rating (mean)	2,858794	3,566354	2,350592
Arts & Enterntainment	0,261062	0,297082	0,237192
College & University	0,132743	0,196286	0,145161
Food	0,615044	0,609195	0,610057
Nightlife	0,119469	0,184792	0,128083
Outdoors	0,243363	0,170645	0,225806
Office	0,588496	0,596817	0,634725
Residence	0,216814	0,263484	0,29981
Shop & Service	0,867257	0,729443	0,833966
Travel & Transport	0,234513	0,263484	0,191651
Industrial	0,053097	0,016799	0,055977
Parks	0,154867	0,167993	0,175522
Rivers	0,057522	0,056587	0,048387

Figure 3. Venue classes and proportions of their collocation with functional objects

To check hypothesis 3 we have applied DBSCAN algorithm and have received clusters of highly rated venues (class 2) (Figure 4) and highly rated venues with low rated neighbors (class 3). Low rated venues (class 1) do not tend to form clusters. High rated food clusters collocate with existing creative spaces (they reside on their territory) as well as shops and office zones and appear to be located mostly in historical city centre. Their clustering might be explained by the fact that venues control quality of their services together and are controlled by their renters.

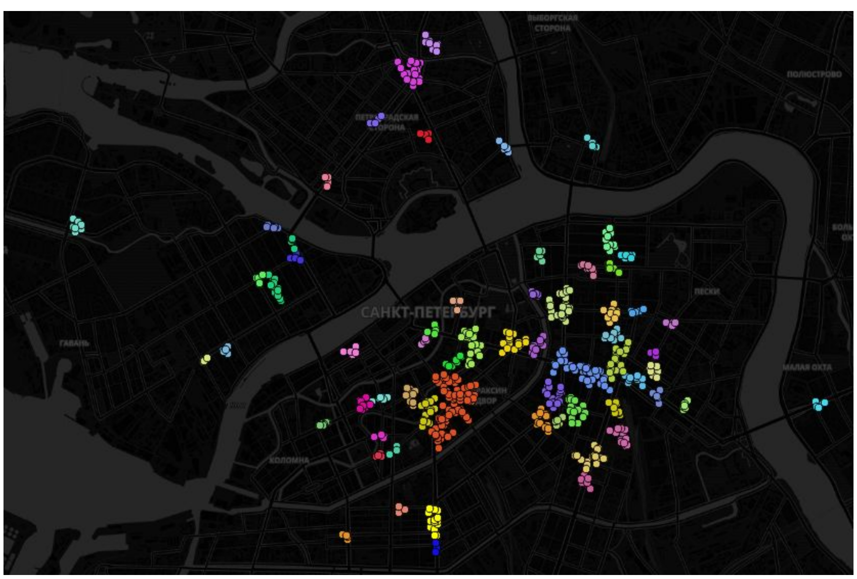

Figure 4. Clusters of highly rated food venues (radius 100 m, minimal number of neighbors 5)

5. Conclusions

The paper shows that user-generated data on urban services together with geospatial data on functional objects can be successfully used together for analysis of spatial segregation, in particular, spatial clusterization of food services. Results received show that food services tend to collocate with certain functional objects in urban environment. The most important environmental feature for location of food venues is shops and services, second - offices and business centers. Public spaces as attractors are not significant. This hints a problem that open public space in St. Petersburg is underdeveloped, in particular, left without sterling food service. Functional objects do not impact location of venues with high (or low) ratings. Food venues tend to form clusters of popular highly rated places in historical city centre, peripheral city areas are more occupied by less popular and less rated venues. These conclusions can be interpreted as spatial segregation of urban space: monocentricity, lack of diversity.

The advantages of using different data sources to analyze food services have to be analyzed further. In this paper we have argued that Google maps has became an important datasource for Russian cities due to the abundance of Android mobile phones, however we plan to conduct a comparative survey with data from Google places, Foursquare and other location based platforms to check their applicability and resourcefulness.

The environmental characteristics should be researched further, in particular, a more detailed account should be given on *mobility* as a predictor for food services appearance and clustering. Space Syntax analysis of street network centrality and network analysis of pedestrian flows could be conducted.

A more accurate calculation is needed to define if *public spaces* (parks, streets, embankments) play any role in attracting services: while we have not detected any importance in our analysis, we are going to compose an indicator of landscape attractivity to check if high-rated popular places are clustered in locations with a good view on a river or a beautiful street.

Class formation of food venues should be also explored more for *social* parameters, such as demographic, economic, and cultural features of their users. This can be conducted based on user-specific check-in and reviews data.

Based on analysis of service-driven spatial segregation recommendation might be formed on normalization of service distribution, planning of inclusive and diverse chains of services, optimization of urban space use and improvement of the perceived quality of life.

6. Literature

1. Jacobs J.: The Death and Life of Great American Cities. Random House, New York. (1961)
2. Sennett R.: The Fall of Public Man. Knopf. (1977)
3. Lefebvre H.: The Production of Space. Basil Blackwell, Oxford. (1991)
4. Harvey D.: Spaces of Capital: Towards a Critical Geography. (2001)
5. Harvey D.: Rebel Cities: From the Right to the City to the Urban Revolution. (2012)
6. Zukin S.: Loft Living: Culture and Capital in Urban Change. Rutgers University Press. (1989)
7. Cummins S., Macintyre S.: The location of food stores in urban areas: a case study in Glasgow. British Food Journal, Vol. 10, Issue 7, pp. 545-553. (1999)
8. Wrigley N.: Food deserts in British cities: Policy context and research priorities. Urban Studies, Vol. 39, Issue 11, pp. 2029-2040. (2002)

9. McDonald J. M. and Nelson P. E.: Do the poor still pay more? Food price variations in large metropolitan areas. Journal of Urban Economics, Vol. 30, pp. 344-359. (1991)
10. Kelly J. M., Swindel D.: Service Quality Variation across Urban Space: First Steps Toward a Model of Citizen Satisfaction. Journal of Urban Affairs, Vol. 24, Issue 3. (2002)
11. Donkin A.J., Dowler E.A., Stevenson, S.J., Turner, S.A.: Mapping access to food in a deprived area: the development of price and availability indices. Public Health Nutrition, Vol. 3, pp. 31-38. (2000)
12. Porta S., Strano E., Iacoviello V., Messora R., Latora V., Cardillo A., Wang F., Scellato S.: Street Centrality and Densities of Retail and Services in Bologna, Italy. Environment and Planning B: Urban Analytics and City Science, Vol 36, Issue 3, pp. 450-465. (2009)
13. Porta S., Latora V., Wang F., Rueda S,. Strano E., Scellato S., Cardillo A., Belli E., Càrdenas F., Cormenzana B., Latora L.: Street Centrality and the Location of Economic Activities in Barcelona. Urban Studies, Vol. 49, Issue 7, pp. 1471-1488. (2011)
14. Apple & Xiaomi strengthen position in Russia in Q3 2017, https://www.counterpointresearch.com/apple-posted-record-shipments-and-xiaomi-became-the-5thlarest-smartphone-brand-in-russia, last accessed 2018/01/10 10.01
15. Why does Google show a 4.8 rating when I have all 5-star reviews?, http://localu.org/blog/google-show-4-8-rating-5-star-reviews, last accessed 2018/01/10
16. Battersby S. E., Finn M. P. Usery E. L., Yamamoto K. H.: Implications of Web Mercator and Its Use in Online Mapping. Cartographica, Vol. 49, Issue 2. (2014)
17. Ester M., Kriegel H., Sander J., Xu X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Association for the Advancement of Artificial Intelligence, https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf , last accessed: 2018/01/10
18. Kisilevich S., Mansmann F., Keim D.: P-DBSCAN: A Density Based Clustering Algorithm for Exploration and Analysis of Attractive Areas Using Collections of Geo-tagged Photos. Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application. New York: Association for Computer Machinery. (2010)
19. Ward, J.H.: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, Vol. 58, pp. 236–244. (1963)
20. Murtagh, F., Legendre, P.: Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion? Journal of Classification, Vol. 31, Issue 3, pp. 27-295. (2014)