The Battle of Neighborhoods – Social Media Mapping

Romain Bocher

December 9, 2019

1. Introduction

1.1 Background

The digital revolution has dramatically changed the way people decide and plan their leisure activities, such as going to a restaurant or a bar. Today, numerous applications enable individuals to find the right venue (e.g. TripAdvisor, Booking.com, Airbnb, Facebook, Google, etc.). The game changer has been the massive adoption of such applications by so-called Millennials for checking and booking. But every person who has visited an unknown city knows that online reviews might not be fully reliable. Moreover, Millennials tends to check multiple sites, including blogs, in order to find the right place to visit. Indeed, existing reviews platforms fail to fully address people's expectations and there are at least two reasons for that. First, reviews are heterogeneous in terms of demographic cohorts since they can also be published by older or younger people. Second, if you place the same person in two different contexts (e.g. a romantic evening vs a bachelor party), then he/she is likely to have distinct expectations.

1.2 Problem

Assuming Fourthsquare and Instagram to be popular platforms among Millennials, this project aims to compare neighborhoods of an unknown city based on the social media popularity of their restaurants. The goal is to target relevant areas to visit without having to spend time on reviews sites or blogs. The focus will be set on Paris, France.

1.3 Interest

This work might gain interest from social travel and booking applications since their current search and recommendation algorithms do not target any specific demographic cohort. It could also prove useful for social media platforms. In the offline world, it could make sense for businesses such as

restaurants, bars and nightclubs, or event planners looking for the right location. Last, although it is not the initial purpose of this paper, network analysis could lead to interesting findings about urban organization, online interactions and Millennials preferences.

2. Data Acquisition and Cleaning

2.1 Data Sources

Paris is divided into 20 administrative areas known as *arrondissements*, and each arrondissement is divided into 4 *quartiers* (i.e. neighborhoods). Therefore, there are 80 neighborhoods to analyze. Data such as latitudes, longitudes or perimeters can be imported for free from Paris official web platform. Restaurants search is done calling Fourthsquare API, while social media popularity metrics are retrieved calling Fourthsquare API once again and scrapping Instagram pages.

2.2 Data Selection

We are using a free personal account on Fourthsquare API. Thus, there are two main limits to consider:

1. We can only get up to 50 venues per location.
2. We can only make up to 500 premium calls per day.

Premium calls are a critical part of the process since they enable us to retrieve important details on restaurants like Fourthsquare rating or Instagram username if one. To overcome this issue, data on restaurants are selected as follows:

Step 1: Get up to 50 recommended restaurants for each neighborhood calling Fourthsquare API. Please note that for each neighborhood we must set a radius that is consistent with the neighborhood's surface. Moreover, we need to limit the request to the neighborhood administrative limits using 'near' option.

Step 2: Select 500 restaurants displaying the highest eigenvector centralities. This method is closed to the famous Google PageRank algorithm: restaurants with high eigenvector centrality are geographically closed to restaurants which are also closed to other restaurants. This technique enables us to detect relevant areas for our study and to focus on a limited number of venues.

Step 3: Import venues details such as Fourthsquare rating, Fourthsquare likes count and Instagram followers count.

2.3 Data Cleaning and Features Selection

Neighborhoods data are downloaded from one single source with few changes to make to get the following table:

	Neighborhood	PostalCode	Latitude	Longitude	Radius
0	Javel	75015	48.839060	2.278076	1669.482927
1	Saint-Thomas-d'Aquin	75007	48.855263	2.325588	791.864176
2	Faubourg-Montmartre	75009	48.873935	2.343253	576.539498
3	Palais-Royal	75001	48.864660	2.336309	448.322129
4	Clignancourt	75018	48.891668	2.345979	1242.550732
5	Saint-Germain-des-Prés	75006	48.855289	2.333657	530.888347
6	Porte-Saint-Martin	75010	48.871245	2.361504	671.579562
7	Maison-Blanche	75013	48.823128	2.352433	1409.903551
8	Halles	75001	48.862289	2.344899	539.271421
9	Chaillot	75016	48.868434	2.291679	1077.345335

For each neighborhood, coordinates and radius are used as variables for an 'explore' request (step 1), leading to a table of 4000 restaurants. Then, we calculate a distance matrix using each restaurants latitude and longitude, and generate a graph using Networkx package in order to calculate eigenvector centralities and filter 500 restaurants (step 2). For each selected venue, we get the Fourthsquare rating, the likes count and the Instagram username (step 3). There might be several restaurants with no Instagram username, so we need to handle that case setting the followers count to zero. Last, we import Instagram followers count for each restaurant with a username using a simple web scrapping algorithm. Note that we could also use the Instagram API for that purpose, but the registration process is not straightforward.