

Ecole Nationale des Sciences Géographiques- ENSG-Géomatique

Mastère Spécialisé

Décisions et Systèmes d'Informations Géolocalisés

Option : Data Science

Rapport du projet de l'analyse spatiale

Thème :

Paris

Restauration - réseaux routiers- et environnement

Dans le cadre de l'UE « Analyse spatiale »

Fait par :

Aymen Ghanem

Sabrine Roy

Houda Bouazzaoui

Lydia IMDNI

Chargé du cours/ TD

Mr. Juste Raimbaut

Année Universitaire 2021 – 2022

Introduction

Dans le cadre de ce projet autour de l'analyse spatiale sur l'accessibilité aux restaurants parisiens, plusieurs analyses réalisés dans le but de bien comprendre la problématique autour de ce sujet.

L'objectif est de comprendre les logiques, les causes et les conséquences de la localisation des peuplements et des activités des humains. Cependant, l'analyse spatiale met en évidence des structures et des formes d'organisation spatiale récurrentes, que résument par exemple les modèles centre-périphérie, les champs d'interaction de type gravitaire, les trames urbaines hiérarchisées, les divers types de réseaux ou de territoires, etc. Cette analyse part du postulat selon lequel l'espace est acteur organisé, tout en proposant une approche modélisée de l'espace géographique en mettant en évidence des formes récurrentes d'organisation spatiales et des théories, notamment à travers diverses notions-clés : distance, réseaux, structure, situation...

Il s'agit de prendre en compte un ensemble complexe de données physiques et humaines pour analyser les distributions spatiales de divers phénomènes, comme autour du sujet des restaurants parisiens en analysant leur évolution, les paramètres qui influencent leur croissance ou non. Tout en s'appuyant sur les données statistiques et sur des espaces isotropes (cela permet d'élaborer des théories qui sont ensuite appliquées à l'espace réel, par essence anisotrope, en tenant compte des particularités).

Nous mettons en valeur dans le cadre de ce travail les objets spatiaux, mesure des relations entre objets en fonction de leur distance, et l'interpolation spatiale. Notre étude se fait à travers les étapes suivantes :

- La reconnaissance des éléments constitutifs de l'espace : les localiser, les caractériser et les situer.
- Explication, détermination de chaque élément du système.

Tout en s'articulant autour des composantes suivantes :

- la visualisation
- l'explorer et de synthétiser les données

- la modélisation spatiale tente d'expliquer les configurations par la spécification d'un modèle statistique et l'estimation des différents paramètres.

Partie 1 : Analyse statistique

Dans cette partie, nous nous intéressons à l'aspect restauration de Paris. Selon les données de l'INSEE, Paris et ses alentours compterait plus de 17 000 restaurants¹. Les restaurants sont des espaces investis pour la consommation de nourriture mais aussi des lieux d'échanges économiques, sociaux et culturels participants à la représentation des altérités au sein de la ville de Paris.

Ce projet est l'occasion de se pencher sur la question de la spatialité de ces établissements et du lien avec d'autres éléments.

Données et packages :

Dans un premier temps, nous installons tous les packages nécessaires à l'accomplissement de l'analyse. Parmi les packages utilisées, on retrouve : units, sf, osmdata, cartography, spatialposition, ggplot2, tidyr, readr, rdgal, spatstat, maptools, raster.

Les données mobilisées proviennent de deux sources :

- Datagouv.fr : pour les quartiers, arrondissements et population de Paris
- OSM : avec le package osmdata et le site OSM DATA pour toutes les autres données : restaurant, la Seine, les parcs, les métros.

Nous avons donc les données suivantes :

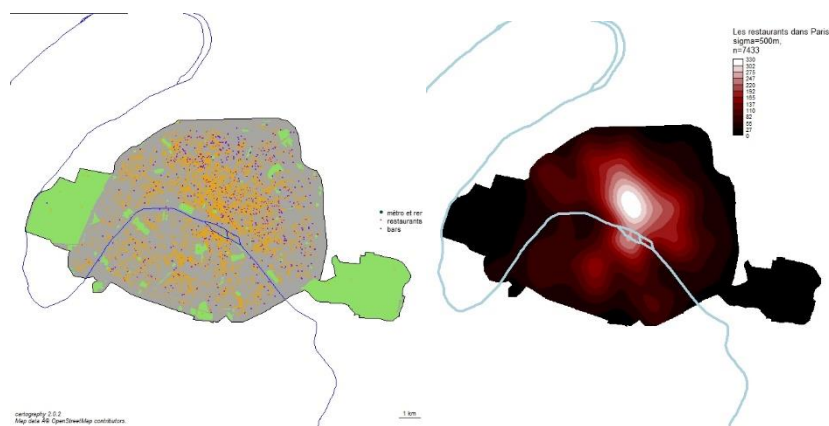


Fig. 1 : à droite l'ensemble des données chargées ; à gauche une carte de la densité des restaurants dans Paris.

^{1 1} Chiffre de l'INSEE en 2018, mais il faut garder en tête les effets de la crise du Covid et son influence sur le secteur de la restauration.

URL: <https://www.insee.fr/fr/statistiques/2011101?geo=DEP-75>

Pour la suite de l'étude, nous allons nous concentrer sur les restaurants (voir la figure à gauche). Les données de restaurants récupérées grâce au package « osm_data » nous permet d'avoir des informations qu'une grande partie des restaurants indiens sont situés à l'intersection du 9^e, 10^e et sud du 18^e arrondissement. Cela s'explique par les différentes vagues de migrations (au 18^e, début 20^e mais surtout autour des années 1990 dû aux réfugiés politique Tamouls du Sri Lanka et de l'Inde) des populations issues de l'Inde et de ses pays voisins (Pakistan, Sri-Lanka, Bangladesh) qui a participé à l'émergence d'une microgéographie indienne dans ces arrondissements. Pour la partie « statistique spatiale », nous allons nous intéresser aux restaurants de type « indien ». Le premier indice calculé est celui de l'autocorrélation spatiale qui a donné un résultat de 0.728. On est alors dans une tendance à l'agrégation. Un plot nous permet de voir cette répartition (voir figure 2). On peut voir sur la carte

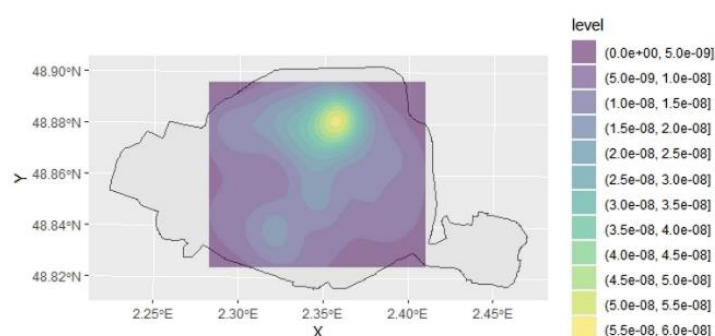


Fig2. Cartographie de la densité des restaurants indiens à Paris

Nous avons ensuite calculé des coefficients de corrélation des effectifs des restaurants indiens par rapport à la population parisienne. Nous avons calculé deux coefficients : celui de Spearman (0.49) et le coefficient de Pearson donne : 0.249 ce qui signifie que le lien entre les deux variables est non significatif. Un plot nous a permis d'avoir une représentation de cette relation (voir fig 3) et qui confirme que la tendance n'est ni monotone ni significative.

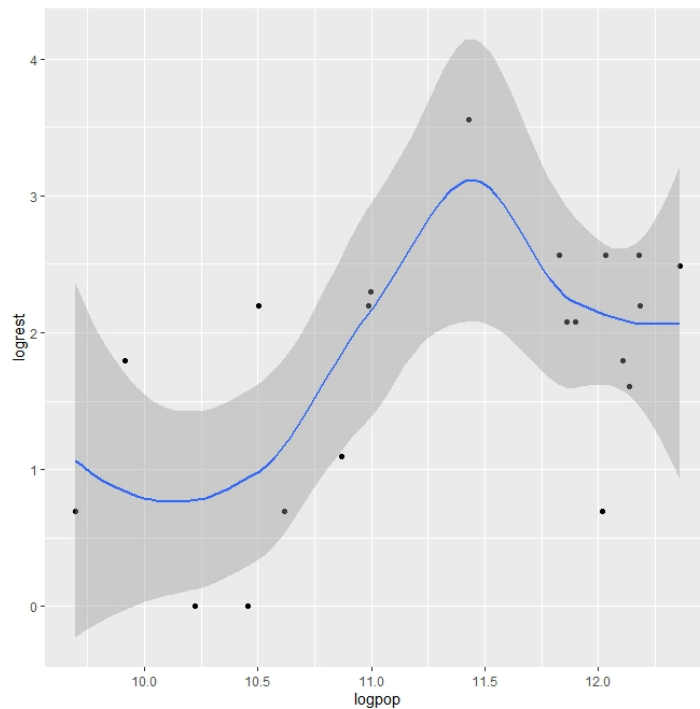


Fig.3. Plot de la relation entre les variables : nombre de restaurants indiens et population dans Paris.

Nous avons continué en calculant l'indice un indice d'autocorrélation spatiale pour mesurer la dépendance spatiale entre les deux variables sur Paris avec la fonction « spAutocorr » qui a été vue en cours mais le résultat qui en a découlé ne semble pas être juste « -21856889 » et est donc inutilisable.

La dernière opération tentée était un calcul des indices de concentration. Pour cela nous avons sélectionné les différents restaurants asiatiques comme les sushis, vietnamiens, chinois, thaïlandais, japonais, coréens mais la fonction censée faire le comptage de ces différents types de restaurants dans la couche arrondissement n'a pas fonctionné rendant impossible l'exécution de l'ensemble impossible (malgré les recherches intensives sur internet).

Partie 2 : Les chantiers perturbants de la ville de Paris

Cette deuxième partie vise à étudier l'ensemble des restaurants, notamment les restaurants indiens, qui ont un accès perturbé causé par des chantiers à proximité.

Cette étude consiste dans un premier temps d'affecter à chaque restaurant le chantier le plus proche et ensuite mesurer et cartographier son impact sur le restaurant.

Nb : nous avons filtré que les restaurants indiens afin que le traitement de calcul de chantier le plus proche ne soit pas trop long. En revanche, l'étude peut s'étaler sur l'ensemble des restaurants et de bars de Paris.

Données et packages :

Nous avons utilisé les mêmes packages de la partie précédente.

Concernant les données, deux couches de données essentielles ont été exploitées :

- ✓ Les restaurants de Paris traitées dans la première partie
- ✓ Les chantiers perturbants de la ville de Paris : des données surfaciques représentant « **Les chantiers principaux ayant un impact sur la circulation** »

Ces chantiers représentés répondent principalement aux critères suivants :

- Perturbent la circulation des voitures ou des vélos
- Durent plus d'une semaine
- Sont situés sur les voies principales, hors périphérique

Fiche descriptive de données :

Source	https://opendata.paris.fr
Système de coordonnées	RGF 93 Lambert 93
Thème	Mobilité et Espace Public
Producteur	Direction de la Voirie et des Déplacements - Ville de Paris
Mise à jour	Hebdomadaire
Licence	Open Database License (ODbL)

Description des codes présents dans la donnée :

NIVEAU_PERTURBATION		STATUT	
<u>Code</u>	<u>Valeur</u>	<u>Code</u>	<u>Valeur</u>
2	Perturbant	1	A venir
1	Très perturbant	2	En cours
		3	Suspendu
		4	Prolongé
		5	Terminé
NUMERO_STV		TYPOLOGIE	
<u>Code</u>	<u>Valeur</u>	<u>Code</u>	<u>Valeur</u>
9	Nord-Ouest	1	Ville
10	Nord-Est	2	Concessionnaire
11	Centre	3	Privé
12	Sud		
13	Sud-Ouest		
14	Sud-Est		

Analyse 1 :

L'objectif de cette étude est de cartographier les distances des chantiers les plus proches, ceci afin de mesurer l'effet de ces chantiers sur les restaurants, car plus le chantier est proche du restaurant plus il bouleverse l'accès à ce restaurant, est plus le chantier est loin, il est possible de trouver d'autres trajets alternatifs au restaurant et donc le risque d'y accéder diminue.

La cartographie des distances des chantiers perturbants les plus proches aux restaurants indiennes

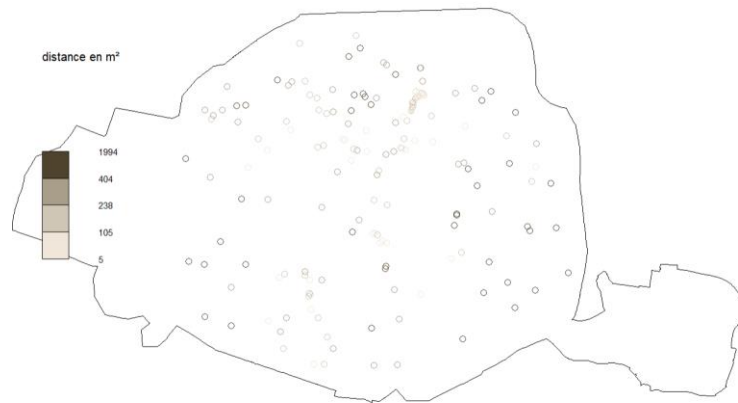


Fig.4.La cartographie des distances des chantiers aux restaurants.

Analyse 2 :

Cette deuxième analyse consiste à exploiter le champ « niveau de perturbation » de la donnée des chantiers afin de visualiser le degré d'impact d'un chantier sur un restaurant à côté.

Ce champ comprend deux codes de valeurs (Voir la partie « Description des codes présents dans la données »). Afin que ces codes soient plus parlants, ainsi que la légende de la carte soit plus compréhensible, nous avons les modifier par leurs significations.

Les restaurants indiennes ayant un accès perturbé

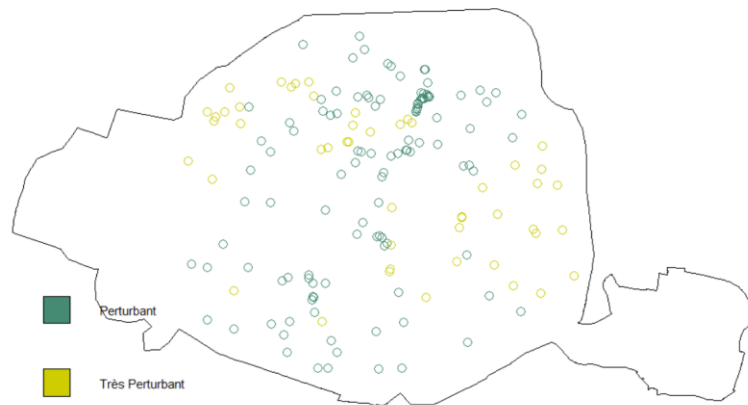


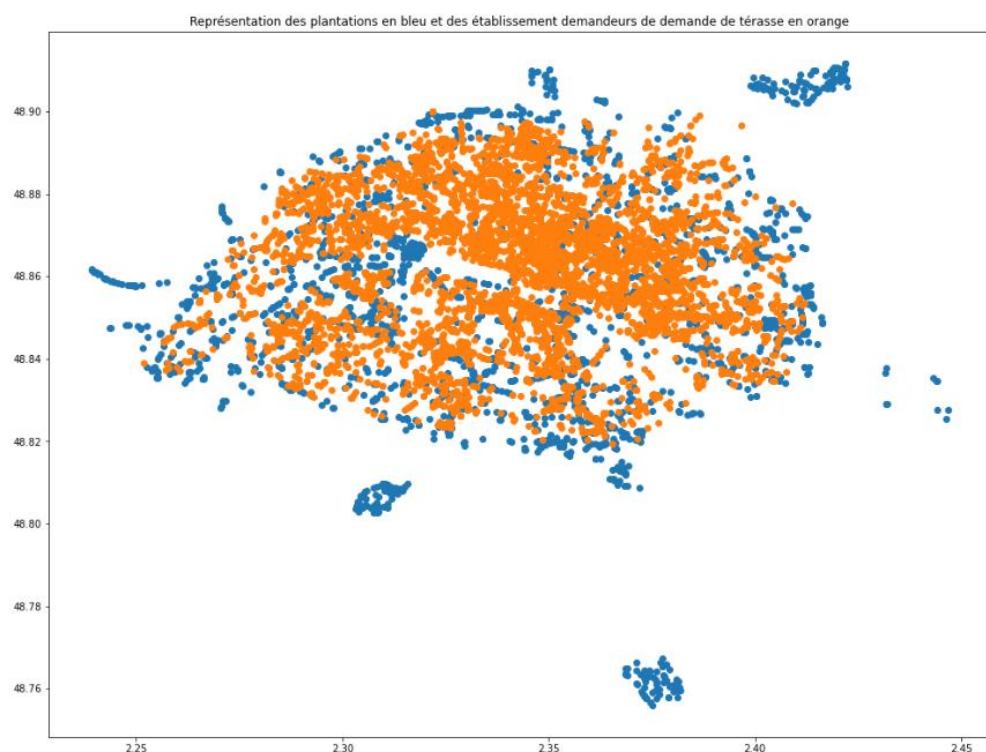
Fig.5. Cartographie des niveaux de perturbation des restaurants

Partie 3 : Analyse de l'interaction spatiale

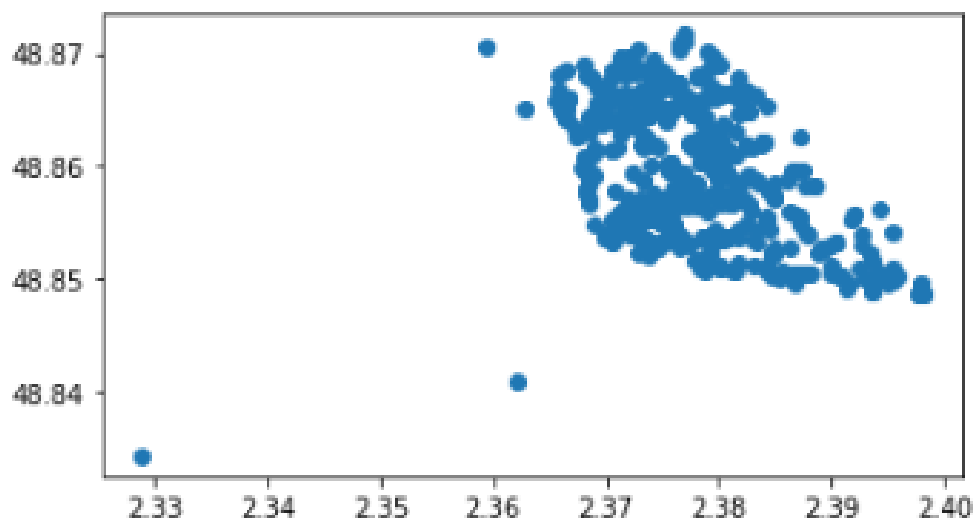
Cette partie consiste à analyser deux datasets : les arbres plantés dans la ville de Paris, ainsi que les restaurants présents dans cette ville.

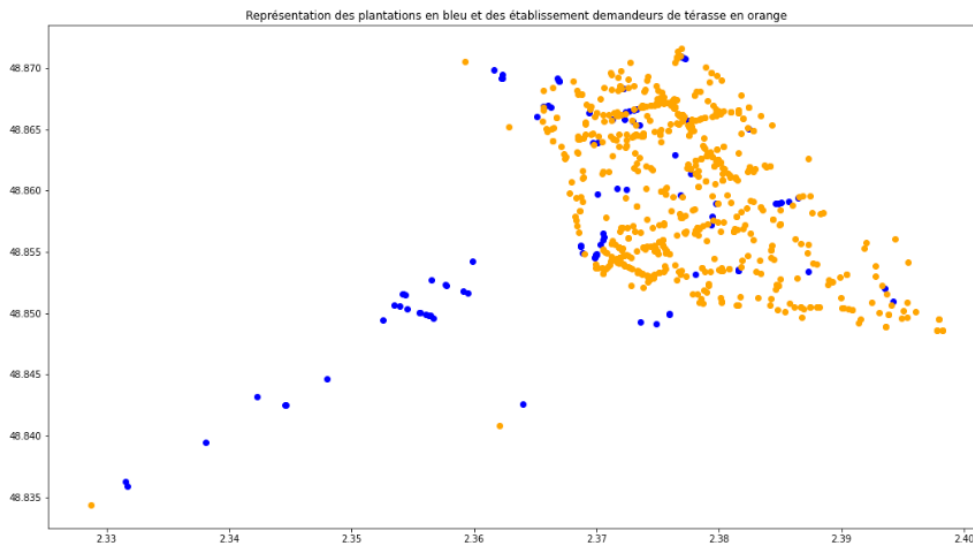
Les deux datasets mobilisés ont été récupérés depuis [opendata.paris](https://opendata.paris.fr/) et [data.gouv](https://data.gouv.fr/) (voir les données fournies pour cette partie)

Nous avons analysé la distribution des labels des arbres en fonction de la présence du plus grand nombre de restaurants dans les différents arrondissements de Paris.



Cette Carte nous montre la distribution des plantations avec les différents labels sur les arrondissements de Paris ayant la présence de restaurants. Ce qui nous intéresse dans cette analyse c'est bien la corrélation spatiale qui peut y avoir ou pas entre le plus grand nombre de restaurants présents dans les arrondissements et le type de labels des plantations présents dans ce meme arrondissement.





Après avoir calculer l'arrondissement où il y a une forte présence de restaurants (voir le code fourni), nous pouvons constater sur l'image ci-dessus la corrélation représentée des plantations en bleu et les établissements ayant déjà demander des terrasses en orange.

Nous avons constaté après cette étude que le plus grand nombre de restaurants se situent dans le 11ème arrondissement de Paris et le type label le plus présent des plantations c'est le type "Platanus".

Partie 4 : Réseau routier de la ville de Paris

Dans cette partie, on se prête à l'étude du réseau à l'aide de la théorie des graphes. On va utiliser un dataset disponible sur le site du [opendata.paris](https://opendata.paris.fr/). Le dataset se présente comme suit :

	Identifiant arc	Date debut dispo data	Date fin dispo data	Libelle	Identifiant noeud amont	Libelle noeud amont	Identifiant noeud aval	Libelle noeud aval
1	1	1 janvier 2005 02:00	1 juin 2019 04:00	Quai_du_Louvre	1	PT_Carrousel_Droite	0	Quai_du_Louvre_(Musee)
2	2	1 janvier 2005 02:00	1 juin 2019 04:00	Quai_du_Louvre	0	Quai_du_Louvre_(Musee)	2	Quai_du_Louvre-Pt_des_Arts
3	3	1 janvier 2005 02:00	1 juin 2019 04:00	Quai_du_Louvre	2	Quai_du_Louvre-Pt_des_Arts	3	Quai_du_Louvre-Am_de_Colig

Dans un premier temps, nous allons nous transformer le dataset suivant en un graph afin d'en extraire les propriétés qui le caractérisent.

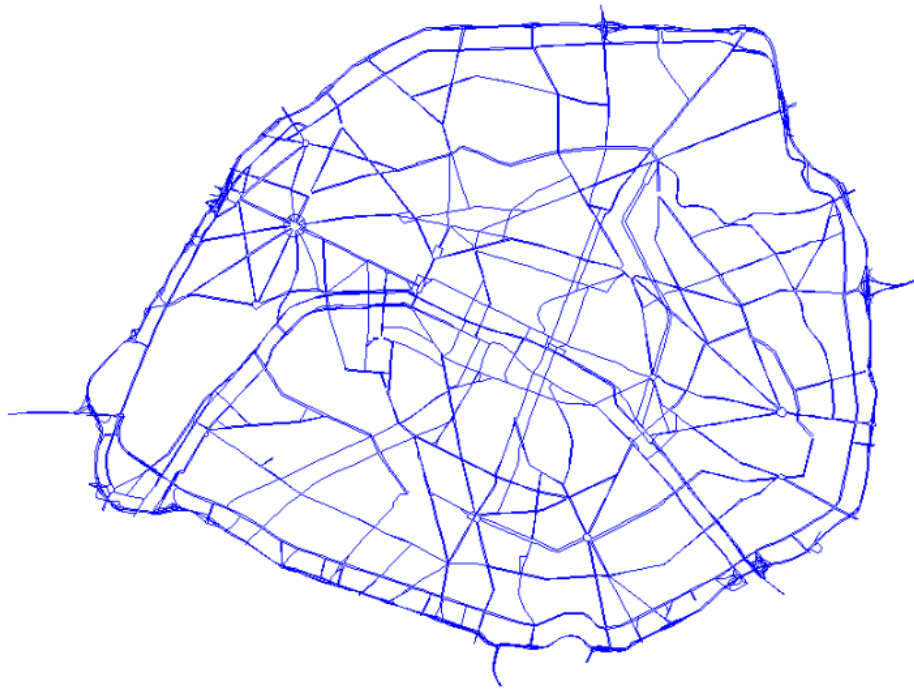


Fig.6 : Réseau routier parisien

Pour la représentation de ce dernier en graph, nous utilisons les nœud en amont et en aval pour caractériser les arrêtes du graph. Ainsi, lorsqu'on essaye de le représenter, on trouve :

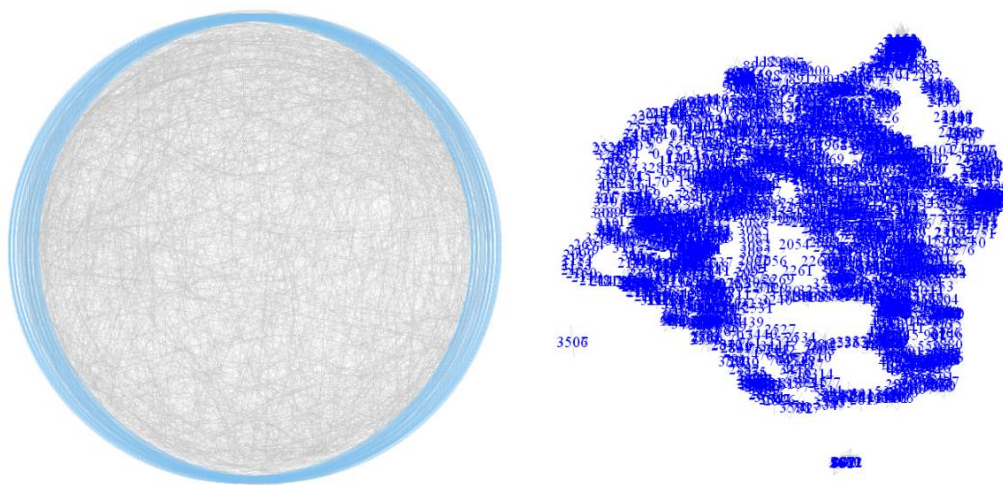


Fig.7 : Représentation du graph du réseau routier de Paris. On remarque deux clusters de route isolés

Analyse 1 : Diamètre du graph Parisien

Si on s'intéresse au diamètre du graph non pondéré, on trouve : 58. Ce chiffre n'a aucune signification physique. Cependant, en pondérant le graph par les distances en mètre de chaque route, on trouve que le diamètre du graph est : $d = 13,015$ km. Pour ce diamètre, si on calcule

la superficie de Paris (assumant que Paris soit circulaire), on trouve une surface de 132 km², sachant que la valeur réelle est de 105 km² (26% d'erreur relative).

Analyse 2 : Degré des sommets

En ce qui concerne l'analyse des degrés, on trouve la distribution suivante des degrés :

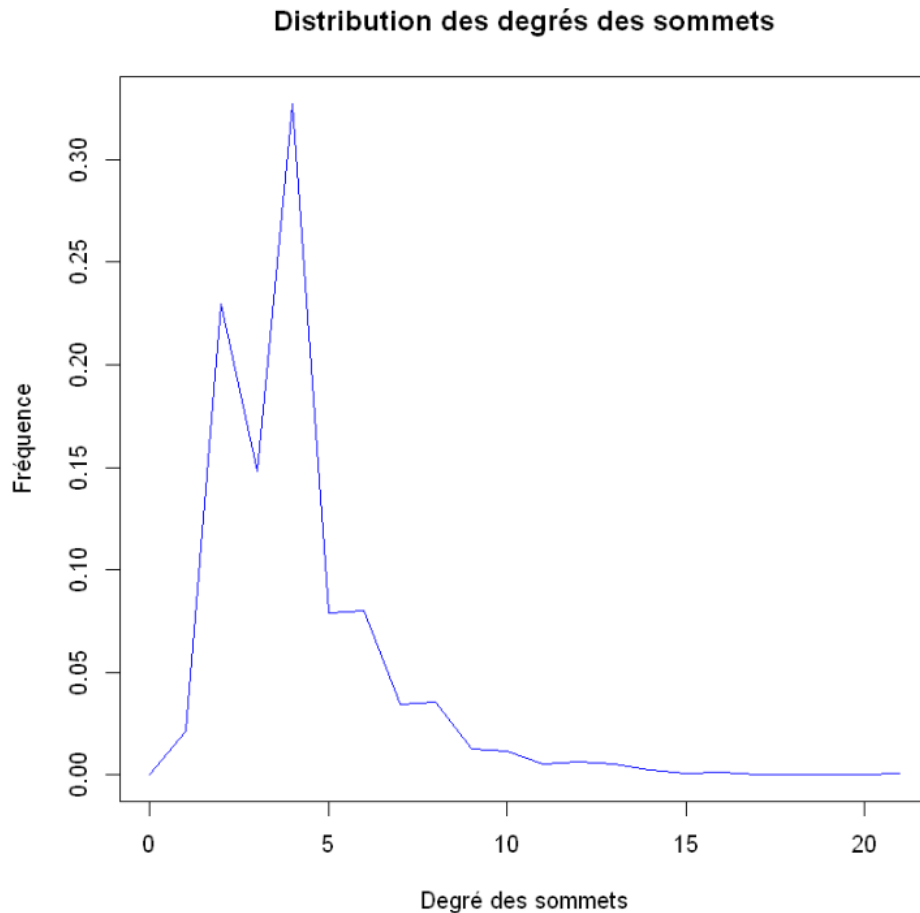


Fig 8. Distribution des degrés des sommets des graphes. Le degré le plus fréquent est 4.

4 est le degré le plus fréquent car ce dernier est dû au croisement de deux routes. Par pur curiosité, le sommet avec le degré le plus élevé (à savoir 21) est attribué à la Place Charles de Gaulle Etoile.

Analyse 3 : Plus court chemin de mon bar préféré à mon autre bar préféré

Pour cette analyse, nous allons extraire les coordonnées spatiales des lieux suivants à partir d'internet :

- The Highlander Bar - 48.856083, 2.339455

- The Freedom Pub - [48.872236, 2.303854](#)

Avec ces coordonnées, nous allons chercher les lieux sur le dataset qui leur sont le plus proche. Pour le Highlander, le Quai Conti correspond à la route la plus proche dans notre dataset. Tandis que pour le Freedom Pub, il s'agira de l'avenue des Champs Elysée.

Le chemin renvoyé est le suivant : 2623 2622 2600 2606 2603 2601 843 844 829 846 2120 2119 2116 2117 2114 2098 3199 2137 2138 2139 211 179 177 176 174 169 170 126 125 124 123 197 200 201 202 3042 3035 2925 3034 3023 2924 2923 2922 3014 3015 3016 3017 3149 3002 2996 644 643 588 647 636 637 634 635 610 2732.

Lorsqu'on représente le chemin suivi en carte, on se retrouve avec la carte suivante :



Fig 9. Carte de “l’itinéraire entre les deux bars”

Remarques : Le problème avec cette carte est qu'elle ne représente pas un itinéraire. En effet, ceci est dû à la non uniformité des identifiants des lieux dans le dataset choisis pour le travail. Ainsi, ce problème est plus un problème de qualité de donnée qu'un problème algorithmique.