



ENS DE LYON

Réseaux Complexes - Projet

ÉTUDE DES RÉSEAUX FERRÉS
FRANÇAIS ET ALLEMAND

Professeur : Rémy Cazabet

Encadrante : Juliana Du

Élèves :



Table des matières

1	Introduction	2
1.1	Sujet et questions abordées	2
1.2	Organisation du travail et répartition des tâches	3
2	Bases de données pour le réseau de TER français	3
2.1	Gare - fréquentation - population	3
2.1.1	Fusion des bases de données	3
2.1.2	Analyse quantitative de la fusion des bases de données	4
2.2	Trajet - tarif - distance - durée	4
2.2.1	Tarif	5
2.2.2	Distance	6
2.2.3	Durée	8
2.3	Conclusion	9
3	Création et étude du réseau de TER français	9
3.1	Création du réseau de TER français	9
3.2	Première étude du graphe comme représentation du réseau de TER français	10
3.3	Coefficient de clustering et propriété small-world	12
3.4	Étude des communautés	14
3.5	Étude des centralités	16
3.5.1	Degrés	16
3.5.2	Betweenness	16
3.6	Étude de la distance	18
3.7	Étude des corrélations entre attributs des nœuds	19
3.8	Conclusion	21
4	Création et étude du réseau ferré allemand	22
4.1	Base de données	22
4.2	Représentation "linéaire" du réseau ferré allemand	26
4.3	Représentation "small-world" du réseau ferré allemand	28
4.3.1	Étude préliminaire du réseau	28
4.3.2	Étude des communautés	31
4.3.3	Étude de centralités	33
5	Conclusion	34
A	Comparaison du réseau allemand avec un autre réseau français	37
A.1	Construction des graphes français sur le modèle des graphes allemands	37
A.2	Comparaison avec le réseau allemand	37

1 Introduction

1.1 Sujet et questions abordées

Originellement, notre objectif était d'essayer de "comprendre" les prix appliqués par la SNCF aux liaisons TER en France. Pour cela, nous voulions chercher des corrélations entre le prix et : la durée des trajets, la distance et l' "attractivité" des villes reliées (évaluée par la population des villes, par leur statut de chef-lieu ou non...). L'idée était donc de construire un réseau avec :

- **nœud** : une gare TER du réseau ferré ;
- **attributs de chaque nœud** : les coordonnées GPS (latitude, longitude) de la gare, la population de la ville dans laquelle se situe la gare, la fréquentation de la gare, le statut de chef-lieu ou non ;
- **lien** : si liaison TER directe entre les gares représentées par les nœuds ;
- **attributs ou poids (à choisir) des liens** : le prix du trajet, la distance kilométrique effectuée par le TER, la durée du trajet.

En plus de l'étude sur les prix, nous voulions comprendre les propriétés essentielles du réseau et entre autres les propriétés qui pourraient être spécifiques à un réseau de TER (vs. TGV) :

- **étude de la centralité** : quelles sont les gares centrales et pourquoi ?
Hypothèses : lien avec l'attractivité des villes à l'échelle régionale ? (population, chef-lieu, "capitale" économique de la région, ville transfrontalière...)?
- **étude des communautés** : à quoi ressemble le territoire national une fois découpé en communautés et pourquoi ?
Hypothèses : retrouve-t-on le découpage du territoire en régions comme attendu pour un réseau TER ? Le découpage correspond-t-il aux anciennes ou aux nouvelles régions ? Les communautés font-elles ressortir les territoires qui partagent une identité culturelle forte, au-delà des régions (Pays Basque, Alsace-Moselle, Bretagne et alentours etc.) ?

Enfin, nous souhaitions comparer notre réseau à un autre réseau ferré afin de juger de la robustesse des propriétés du réseau français. Deux possibilités étaient envisageables :

- **comparaison avec un "null model" de type gravité relaxé** : un modèle aléatoire où la probabilité de relier deux nœuds dépend de la population des deux villes et de la distance entre elles via une fonction de "deterrence" obtenue à partir de notre réseau.
- **comparaison avec le réseau ferré d'un autre pays** : l'Allemagne semblait être un bon choix de part son organisation fédérale. En effet, tous les pays ne font pas la distinction entre le réseau type TER et le réseau type TGV. En prenant l'Allemagne, même dans le cas où l'on ne pourrait pas séparer le réseau TER du réseau TGV, le fait que son territoire ne soit pas centralisé permettra quand même une comparaison avec la France (qui, du point de vue de la structure du réseau TER, ne devrait *a priori* pas apparaître comme centralisée).

Pour plusieurs raisons, l'étude n'a pas pu être réalisée telle qu'envisagée. Pour autant, il semblait intéressant de confronter dans ce rapport nos idées préliminaires à la "réalité de terrain". Dans les parties qui suivent, nous exposons les difficultés rencontrées et les ajustements qui ont été faits pour les surmonter ou les contourner.

1.2 Organisation du travail et répartition des tâches

Avant toute chose, nous avons vérifié la disponibilité des bases de données pour le réseau de TER français, ce qui n'a pas été difficile grâce à la [SCNF](#). Le travail s'est ensuite réparti comme suit :

- **Louhane Jacob** : a essentiellement travaillé sur le traitement des bases de données pour le réseau de TER en France. Elle a aussi aidé à l'interprétation des résultats pour le graphe français.
- **Gaspard Calvet** : a essentiellement construit et analysé le graphe du réseau de TER français à partir des bases de données préalablement préparées, et a construit un deuxième réseau français plus simple, en vue de comparer avec le réseau allemand.
- **Timothée Pouille** : a essentiellement travaillé sur le cas de l'Allemagne, en parallèle du travail sur le réseau français. Comme les bases de données étaient "prêtes à l'emploi", peu de temps a été nécessaire à leur préparation en vue de la construction du graphe. Timothée a donc pu travailler seul sur le cas allemand.

Ce découpage en tâches est bien sûr schématique puisque nous avons beaucoup collaboré pour résoudre certains problèmes, comparer les réseaux français et allemand, écrire le rapport, etc.

2 Bases de données pour le réseau de TER français

Pour étudier le réseau de TER français, nous avons exploité plusieurs bases de données (BDD) issues de sources différentes (SNCF et INSEE) :

1. une BDD donnant les trajets de TER sur tout le réseau ainsi que leurs tarifs pour différents abonnements : [BDD tarifs](#) ;
2. une BDD donnant les gares de voyageurs avec leurs coordonnées GPS et la commune : [BDD gares voyageurs](#) ;
3. une BDD donnant la fréquentation annuelle des gares (textiti.e. le nombre de visiteurs par an) : [BDD fréquentation](#) ;
4. une BDD avec la liste de toutes les gares (voyageurs et fret) avec les lignes ferrées sur lesquelles elles se trouvent ainsi que leur position sur la ligne, position donnée par le PK (point kilométrique) : [BDD PK](#) ;
5. une BDD donnant les horaires des lignes de TER : [BDD horaires](#) ;
6. une BDD donnant le relevé annuel jusqu'en 2020 de la population de chaque commune française : [BDD population INSEE](#).

2.1 Gare - fréquentation - population

2.1.1 Fusion des bases de données

Une première étape consiste à obtenir un document csv où chaque ligne correspond à une **gare de voyageurs** différente avec sa **fréquentation** (nombre de visiteurs/an) et la **population** de la commune dans laquelle elle se trouve. La population communale peut donner une idée de l'attractivité de la ville. Quant à la fréquentation, cela peut être utile pour interpréter

certains cas particuliers. Par exemple, certaines gares pourraient être situées dans des communes peu peuplées mais être très fréquentées. Ce serait le cas de gares TER desservant de grandes villes mais qui seraient situées dans une petite commune en périphérie.

Pour réaliser ce csv, il a fallu **fusionner trois BDD** : celle donnant les gares de voyageurs ([BDD gares voyageurs](#)), celle donnant les fréquentations ([BDD fréquentation](#)) et celle donnant les populations communales ([BDD population INSEE](#)). Les deux premières BDD sont faciles à manipuler car elles sont toutes les deux issues de la SNCF, qui attribue à chaque gare un numéro unique nommé "code UIC" (UIC pour 'Union internationale des chemins de fer'). Il suffit donc d'utiliser la fonction *merge* de *pandas*. En revanche, la troisième BDD est issue de l'INSEE, qui utilise un code géographique officiel (COG) qui lui est propre. Ce code est la contraction entre le code du département à 2 chiffres et le code de la commune à 3 chiffres. Pour autant, il est différent du code postal (plus d'informations [ici](#)). Par chance, la BDD des listes des gares mentionne les codes des départements et les codes des communes. Il a donc suffi de créer une nouvelle colonne avec la concaténation de ces deux codes pour obtenir le COG, permettant ainsi la fusion avec la BDD de l'INSEE.

2.1.2 Analyse quantitative de la fusion des bases de données

Analysons quantitativement la **qualité des fusions des BDD**. Pour commencer, la BDD avec la liste des gares de voyageurs compte 3 133 gares, celle avec les fréquentations en compte 3 021 et la BDD de l'INSEE donne la population de 34 957 communes. La fusion entre les deux premières BDD nous donne une nouvelle BDD avec 2 962 gares. Nous perdons donc environ 2% des données lors de la fusion. Puis la fusion entre la BDD obtenue par fusion et la BDD de l'INSEE nous ramène à 2925 gares. Nous perdons donc une quarantaine de gares supplémentaires. Pourquoi perdons-nous des données ?

Pour la première fusion, l'intersection entre les deux BDD ne donne pas 3021 gares. Cela s'explique par le fait que les BDD ne sont plus à jour (certaines gares ont fermé, ont ouvert...) mais aussi par le fait que des gares étrangères frontalières sont prises en compte dans la BDD des fréquentations mais pas dans la BDD des listes des gares. Ce faisant, nous ne pouvons pas améliorer la première fusion.

En revanche, la seconde fusion a pu être améliorée. En effet, nous avons remarqué que le COG des gares parisiennes pose problème car selon la BDD, la commune est soit définie comme l'arrondissement soit comme la ville de Paris. Nous avons donc uniformisé le COG et récupéré ainsi 27 gares. Il n'y a donc plus que 10 gares perdues en cours de la deuxième fusion. Nous avons regardé au cas par cas ces gares. Il s'agit souvent de cas atypiques, par exemple des communes qui changent de noms, qui disparaissent, d'autres qui fusionnent. Nous avons réussi à corriger deux COG.

Finalement, nous obtenons un document csv de 2 954 gares, avec pour chacune d'elle la fréquentation annuelle et la population de la ville dans laquelle elle se trouve. Un schéma récapitulatif est proposé en Fig. 1. À noter que le graphe construit à partir de cette BDD contiendra moins de 2 954 gares pour des raisons qui seront exposées au moment venu.

2.2 Trajet - tarif - distance - durée

En parallèle, nous créons un deuxième document csv où chaque ligne correspond à un **trajet effectué en TER, avec le tarif, la durée et la distance du trajet**.

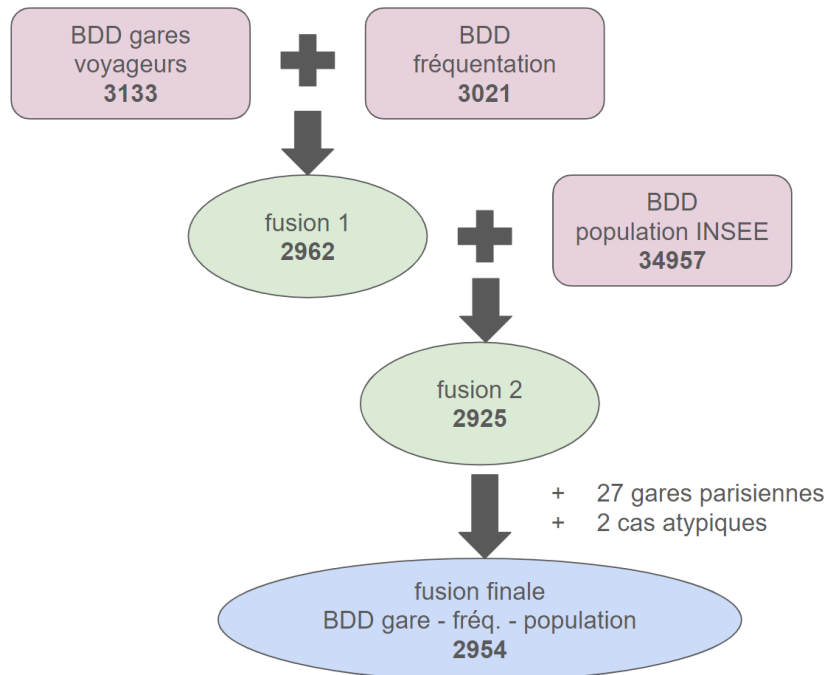


FIGURE 1 – Représentation schématique des fusions réalisées entre la BDD gares voyageurs, la BDD fréquentation et la BDD population INSEE. Les nombres en gras correspondent au nombres de lignes (de gares donc) qui apparaissent dans la BDD en question.

2.2.1 Tarif

Pour commencer, la liste des trajets et les **tarifs** sont donnés dans une BDD de la SNCF ([BDD tarifs](#)). La base de donnée comporte 32 093 entrées, mais chaque entrée ne correspond pas à un unique trajet. En effet, de nombreux trajets sont proposés à plusieurs tarifs différents : des tarifs normaux, des abonnements tout public, ainsi que des abonnements jeunes. En regardant les trajets uniques, on dénombre 20 115 trajets différents.

Afin de pouvoir exploiter ces données, il faudrait avoir pour chaque trajet un prix unique. Pour cela, nous avons décidé de garder pour chaque trajet le tarif le moins cher. Nous obtenons alors une nouvelle base de données avec 20 115 trajets TER différents et un tarif unique pour chacun d'eux.

Analysons cette base de données : fait-elle toujours figurer l'aller et le retour pour chaque liaison ? Nous avons un nombre impair de trajets, donc nous soupçonnions que certains trajets ne soient pas réciproques. Il s'avère que 6 549 trajets apparaissent comme étant à sens unique, ce qui représente tout de même un tiers des trajets listés. Les 13 566 trajets restants correspondent donc à des aller-retours (on a donc $13\,566 \div 2 = 6\,783$ liaisons).

La question qui vient ensuite est : observe-t-on des prix différents entre l'aller et le retour ? Parmi les 6 783 liaisons, 5 750 affichent des prix identiques. Pour les 1 033 liaisons restantes, les prix peuvent faire $\times 10$ entre l'aller et le retour ! Comment est-ce possible ? Déjà, pour 985 liaisons parmi celles-ci, les prix aller et retour ne correspondent pas aux mêmes abonnements. Cela est dû à l'absence de certains tarifs dans la [BDD tarifs](#) originale combinée

au choix qu'on a fait pour créer la nouvelle BDD (celle avec un tarif unique par trajet). On s'attend cependant à ce qu'un tarif normal corresponde à un prix plus élevé qu'un tarif dans le cadre d'un abonnement. Il n'en est rien, puisque que le tarif avec abonnement est toujours bien plus cher (jusqu'à $\times 10$) que le tarif sans abonnement. On a alors fait l'hypothèse que le tarif comprend le prix de l'abonnement à l'achat, de sorte que *prix dans la BDD tarifs = prix de l'abonnement + prix du trajet dans le cadre de l'abonnement*. Pour le vérifier, nous avons étudié des cas particuliers. Prenons par exemple le cas de ces deux liaisons : Lille - Santes et Lille - Don Sainghin. Les trois villes sont situées dans les Hauts-de-France. Pour la liaison Lille - Santes, nous savons que le prix de l'aller en tarif normal est de 2 euros, mais que le retour avec l'abonnement mensuel est de 22.70 euros. On intuite donc que l'abonnement coûte autour de 21 ou 22 euros. Sauf que pour la liaison Lille - Don Sainghin, l'aller en tarif normal coûte 3.20 euros et le retour avec l'abonnement mensuel coûte 34.40 euros, ce qui nous laisse à penser que l'abonnement coûte plus de 31 euros. Un schéma récapitulatif de l'analyse des prix est proposé en Fig.3. Au vue de la contradiction, le prix donné dans la base de données ne doit pas être construit ainsi : *prix dans la BDD tarifs = prix de l'abonnement + prix du trajet dans le cadre de l'abonnement*. N'ayant pas plus d'information, nous ne pouvons pas remédier à ce problème. Nous savons seulement qu'en moyenne, le prix est multiplié par environ 5.6 entre l'aller et le retour quand ils sont pris avec des abonnements différents. Il faudra prendre en compte cet artefact au moment de l'étude des prix. Un schéma récapitulatif du raisonnement sur cet exemple est proposé en Fig.2.

Par ailleurs, certains prix différent alors même que l'aller et le retour se font avec les mêmes abonnements. C'est le cas pour 48 liaisons. En moyenne, il y a une différence de 15% entre les prix aller et retour de ces liaisons.

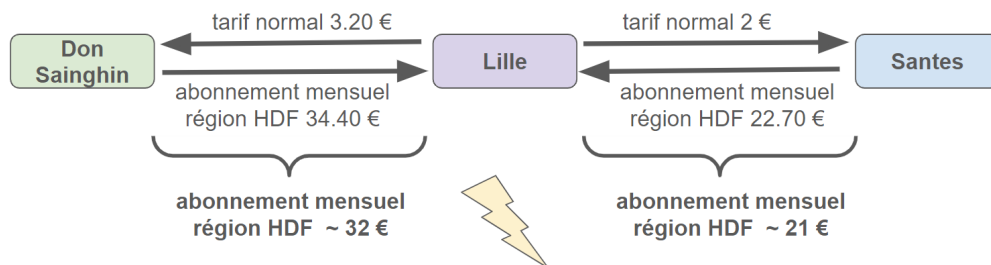


FIGURE 2 – Le prix des trajets en TER donné par la SNCF contient-il le prix de l'abonnement à l'achat ? Il semble que non, vu le cas de ces deux liaisons : Lille - Santes et Lille - Don Sainghin. Ce schéma récapitule le raisonnement. À noter que le prix de l'abonnement mensuel de la région Hauts-de-France est intuité. Pour les curieux.se, le prix de l'abonnement mensuel coûte actuellement [22 euros](#).

2.2.2 Distance

Passons désormais au **calcul de la distance effectuée en train**. La distance la plus intéressante *a priori* est celle qui sépare les gares d'origine et de destination en suivant les voies ferrées. En effet, on peut penser que l'usure des trains et des rails, le temps de trajet ou encore la consommation énergétique dépendent de cette distance et peuvent impacter les tarifs appliqués.

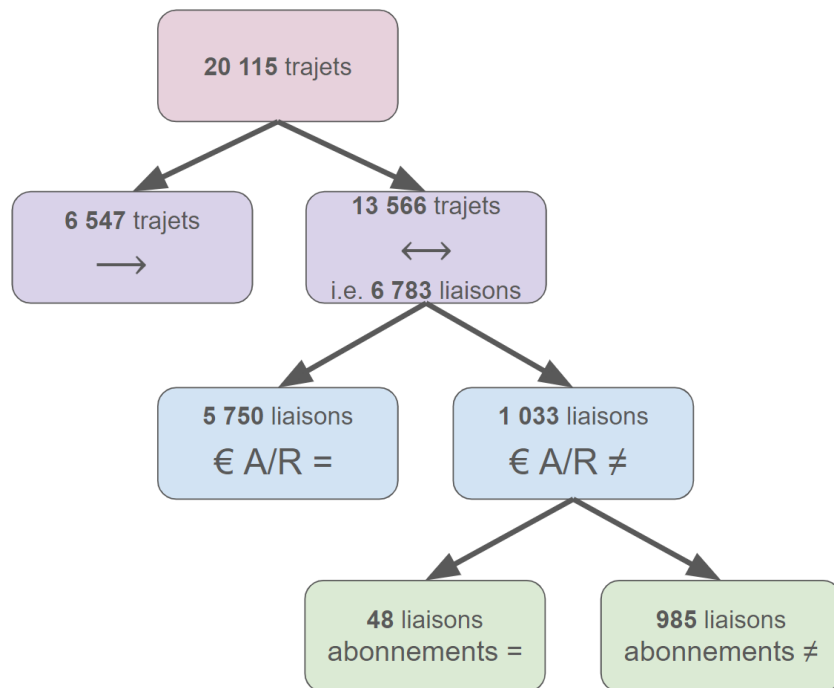


FIGURE 3 – Schéma récapitulatif de l’analyse de la [BDD tarifs](#). Les flèches indiquent si le trajet est à sens unique ou à double sens (aller-retour). Le symbole euro A/R indique si le prix de l’aller est identique au prix du retour ou non. Enfin, si les prix A/R sont différents, il est précisé si les abonnements pris à l’aller et au retour sont également différents ou non.

Il s’avère que le calcul de cette distance peut se faire grâce à la position des gares sur leur ligne, aussi appelée point kilométrique (PK). Le PK est défini par rapport à une ligne ferrée donnée, avec le PK 0 à un bout de la ligne et le maximum du PK à l’autre bout. La [BDD PK](#) nous donne cette information. Pour chaque gare, on a autant de lignes dans le tableau que de lignes ferrées qui passent par cette gare. Chaque ligne du tableau correspond donc à un couple (gare, ligne ferrée) et cette ligne donne le PK de cette gare sur cette ligne. La BDD en question compte 3 819 lignes, soit plus que le nombre de gares de voyageurs. Rien d’anormal puisque comme on vient de le dire, plusieurs lignes de la BDD peuvent correspondre à la même gare. De plus, certaines gares ne font que du fret.

Maintenant, à partir de la [BDD tarifs](#) qui liste les trajets et leur prix, on peut extraire les codes UIC des gares d’origine et de destination, trouver ces codes UIC dans la [BDD PK](#), et calculer la différence entre les PK des deux gares sur la ligne ferrée commune. Cela permet de connaître la distance de la liaison le long des voies.

Cette approche naïve se heurte à plusieurs problèmes. Pour commencer, sur les 40 230 gares à traiter (20 115 trajets TER avec une gare d’arrivée et une gare de départ), on est confrontés 428 fois à des gares de la BDD 1. qui ne se trouvent pas dans la BDD 4. D’autre part, sur les 20 115 trajets à traiter, il arrive 6 607 fois que les deux gares ne soient pas sur la même ligne ferrée.

Le premier problème ne peut pas être réglé car il vient du fait que les données sont partielles ou erronées. Le second problème quant à lui pourrait être réglé. Ce qu’on peut observer grâce à la visualisation proposée [ici](#), c’est que les trajets listés dans la [BDD tarifs](#) correspondent à des trajets sans correspondance mais avec changement de lignes ferrées. Les gares d’origine et

de destination ne sont donc pas forcément situées sur les mêmes lignes ferrées, et pire encore, le train peut emprunter des lignes ferrées intermédiaires, lignes qui ne passeraient ni par la gare d'origine ni par celle de destination. Un exemple d'un tel cas de figure est schématisé en Fig.4. Il apparaît que la seule façon de connaître la distance d'un trajet le long des voies est de réaliser une carte des liaisons comme [ici](#), puis de calculer le plus court chemin entre les deux gares. Cela permet d'en déduire les lignes ferrées empruntées et de faire les calculs sur les PK.

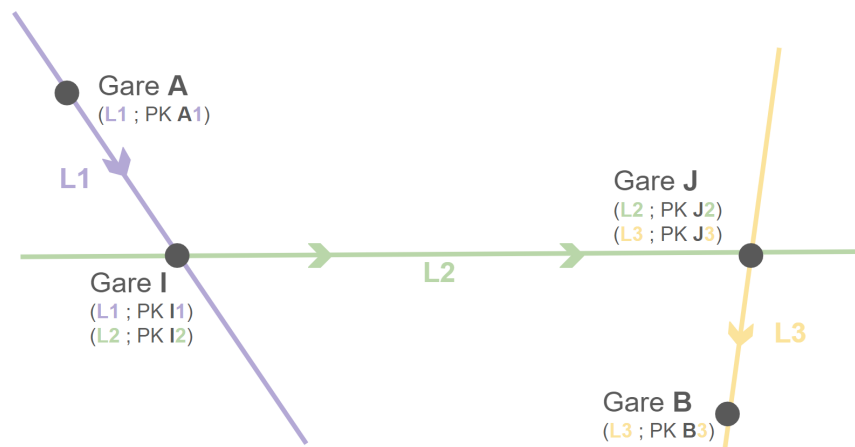


FIGURE 4 – Schéma explicatif d'un cas où la gare de départ (gare A) est sur une ligne ferrée (ligne 1), la gare d'arrivée (gare B) est sur une autre (ligne 3) et le trajet de TER emprunte une troisième ligne ferrée (ligne 2) qui ne peut pas être découverte seulement avec les données (Ligne, PK) des gares A et B.

Par manque de temps, nous avons décidé de procéder autrement. En effet, nous possédons les coordonnées GPS de chacune des gares grâce à la [BDD gares voyageurs](#). Nous avons donc décidé de simplifier l'étude des distances en calculant la distance à vol d'oiseau tout en prenant en compte la courbure de la terre. Celle-ci a pu être très facilement calculée une fois le réseau créé à partir de la fonction *haversine* du module python éponyme. Celle-ci calcule la distance du grand cercle entre deux points d'une sphère, à partir de leurs longitudes et latitudes. Il faudra donc à nouveau prendre des précautions au moment d'analyser le réseau.

2.2.3 Durée

Il nous reste à traiter la question de la **durée des trajets**. Étonnamment, aucune BDD en format csv n'est proposée par la SNCF. Il existe cependant les [BDD horaires](#), fournies par la SNCF sous différents formats : NeTEx, GFTS, GFTS-RT, SIRI SX LITE. Après une rapide recherche internet, il semble qu'il soit nécessaire d'utiliser des logiciels spécifiques à ces données pour les ouvrir et les traiter. Il a été décidé de ne pas s'aventurer sur ce terrain, d'autant plus que le traitement précédent des données avait déjà pris un temps considérable.

En conclusion, il apparaît à ce stade que les données auxquelles nous sommes parvenus ne permettront pas de réaliser une étude des tarifs des liaisons de TER. Nous avons donc décidé de recentrer le sujet sur une comparaison des réseaux ferrés (TER ou équivalent) français et allemand.

2.3 Conclusion

En conclusion, pour la création du réseau de TER français, nous sommes partis de 4 bases de données que nous avons fusionnées et nettoyées. Les 3 bases de données gares de voyageurs, fréquentation et population de l'INSEE ont été fusionnées pour obtenir une seule BDD. Dans celle-ci, chaque entrée correspond à une unique gare, avec son code gare, son nom, sa position, sa population et sa fréquentation. Le processus de fusion nous a fait passer de 3133 gares à 2954 gares pour différentes raisons susmentionnées. D'autre part, la BDD des trajets de TER avec leurs prix a été retraitée. Initialement, chaque entrée correspondait à un trajet d'une gare à une autre gare avec autant de tarifs que d'abonnements. Le premier nettoyage a consisté à récupérer un unique tarif pour chaque trajet. Il a été décidé de garder le tarif le moins cher. Ceci fait passer la BDD de 32 093 à 20 115 entrées. Le deuxième nettoyage consiste à regarder pour chaque couple de gares si le trajet aller et le trajet retour sont au même prix. Parmi les 20 115 trajets, 13 566 sont à double sens, soit 6 783 liaisons, mais seulement 5 750 liaisons ont le prix d'aller et de retour égaux (soit 1 033 non égaux). Cela vient du fait que pour de nombreux trajets, la BDD originale ne donnait pas le tarif normal (sans abonnement) pour l'aller ou bien le retour. Ce problème ne pouvant pas être réparé, et altérant trop significativement le réseau (environ 10% des liens), nous avons décidé de laisser de côté l'étude des prix. De même, nous n'avons pas pu trouver de données intéressantes concernant la durée et la distance effectuée par le train.

3 Création et étude du réseau de TER français

3.1 Création du réseau de TER français

La création de ce réseau se base donc sur les deux BDD obtenues dans la section précédente : une BDD avec les gares, leur fréquentation et la population des villes où elles se trouvent ; une autre BDD avec les trajets de TER, leur tarifs et leur distance à vol d'oiseau. Nous allons appeler la première "BDD des gares" et la seconde "BDD des trajets". Le réseau est construit en positionnant comme nœuds les entrées de la BDD des gares. Ensuite, nous créons les liens entre les gares à partir de la BDD des trajets. Nous relierons ainsi deux gares lorsqu'il existe un trajet TER reliant les 2 villes.

Le réseau se compose donc de cette manière :

- Les nœuds sont des gares, et les attributs de ces nœuds sont le nom de la gare, sa position géographique, sa fréquentation, ainsi que le nom, le COG, et la population de la commune dans laquelle se situe cette gare.
- Les liens sont les trajets de TER direct entre les villes, et leur attribut est la distance à vol d'oiseau entre les deux gares (l'attribut des prix n'ayant pas pu être exploité comme on l'a vu précédemment).

Une telle construction demande tout de même une étude préalable des gares présentes dans les 2 BDDs. Cette étude préalable montre qu'on trouve 2 277 gares différentes dans la BDD des trajets de TER, contre 2 954 gares dans la BDD des gares. Cela n'est pas un problème, puisqu'une gare pourrait ne servir qu'aux TGV et pas aux TER, par exemple. En revanche, en calculant l'ensemble d'intersection entre ces 2 jeux de données, on obtient seulement 2 074 gares

communes aux deux BDD. Entre autres, 9% des gares de la BDD des trajets ne se retrouvent pas dans la BDD des gares, ce qui est un problème. Autre problème, 5 gares parmi les 2 074 gares communes sont uniquement reliées à des gares faisant partie des 9% de gares "perdus". Puisqu'elles ne sont reliées à aucune autre gare, nous avons décidé de ne pas les inclure au graphe. Le graphe final comporte donc 2 069 nœuds (gares de TER communes aux deux BDD moins (-) les 5 gares problématiques) et 13 339 liens (liaisons TER directes entre ces gares).

Une première étude des composantes connexes permet d'observer 2 composantes connexes : la première qui concentre la quasi-totalité des gares ; la seconde qui est composée de deux gares de deux communes des Hauts-de-France, Crouy et Margival, qui ne sont reliées qu'entre elles. Par la suite, on étudiera les propriétés du réseau sur la composante connexe de 2 067 nœuds et 13 338 liens.

3.2 Première étude du graphe comme représentation du réseau de TER français

Une carte de France avec les 2 067 gares TER retenues dans notre étude est disponible ci-dessous en Fig. 5. Pour la réaliser, nous avons utilisé la projection Mercator appliquée aux coordonnées GPS (longitude et latitude) des gares. Cela permet d'éviter certaines déformations et ainsi d'obtenir une carte semblable aux cartes qu'on a l'habitude de voir. La formule utilisée pour la projection est la suivante (λ est la longitude, ϕ est la latitude, et R le rayon de la Terre) :

$$x = R\lambda \quad ; \quad y = R \ln \left[\tan \left(\frac{\pi}{4} + \frac{\phi}{2} \right) \right].$$

Pour commencer, la carte des gares (Fig. 5) fait apparaître les contours du territoire français. Cela signifie que la couverture TER s'étend bien jusqu'aux littoraux et aux frontières et donc que le réseau TER n'est pas cantonné à l'intérieur des terres. On peut aussi noter que le territoire dans sa globalité est bien couvert en gares TER, à l'exception de certaines zones :

- l'Île-de-France, ce qui s'explique par le réseau RER qui supplante le réseau TER ;
- des zones simplement peu peuplées (diagonale du vide, Lozère, etc.) ;
- des zones difficiles voire impossibles d'accès comme des îles, des isthmes ou encore des chaînes de montagne (Corse, Cotentin, Pyrénées, Alpes de Haute Provence, etc.).

Par ailleurs, le positionnement des gares sur le territoire français fait bien apparaître des "lignes" de TER. On pouvait s'y attendre pour un réseau de TER : le but du TER est d'obtenir un bon maillage du territoire à de petites échelles, notamment régionales. Ainsi, les lignes de TER proposent beaucoup d'arrêts, souvent rapprochés. C'est grâce à cela que les lignes ressortent aussi bien sur la représentation en Fig. 5. Au contraire, si on avait étudié le réseau de TGV, sans doute que les lignes seraient apparues moins nettement, car le but du TGV est de relier le plus rapidement possible des grandes villes à l'échelle nationale, ce qui implique moins d'arrêts intermédiaires. On aurait sans doute vu des concentrations de gares dans les grandes et moyennes villes et rien entre.

Enfin, les gares dessinent des motifs en étoile, motifs répartis assez uniformément sur le territoire. L'étoile la plus imposante se dessine autour de Paris, capitale de la France. Les autres étoiles, légèrement plus petites, s'organisent autour de métropoles de province. Ainsi, à une

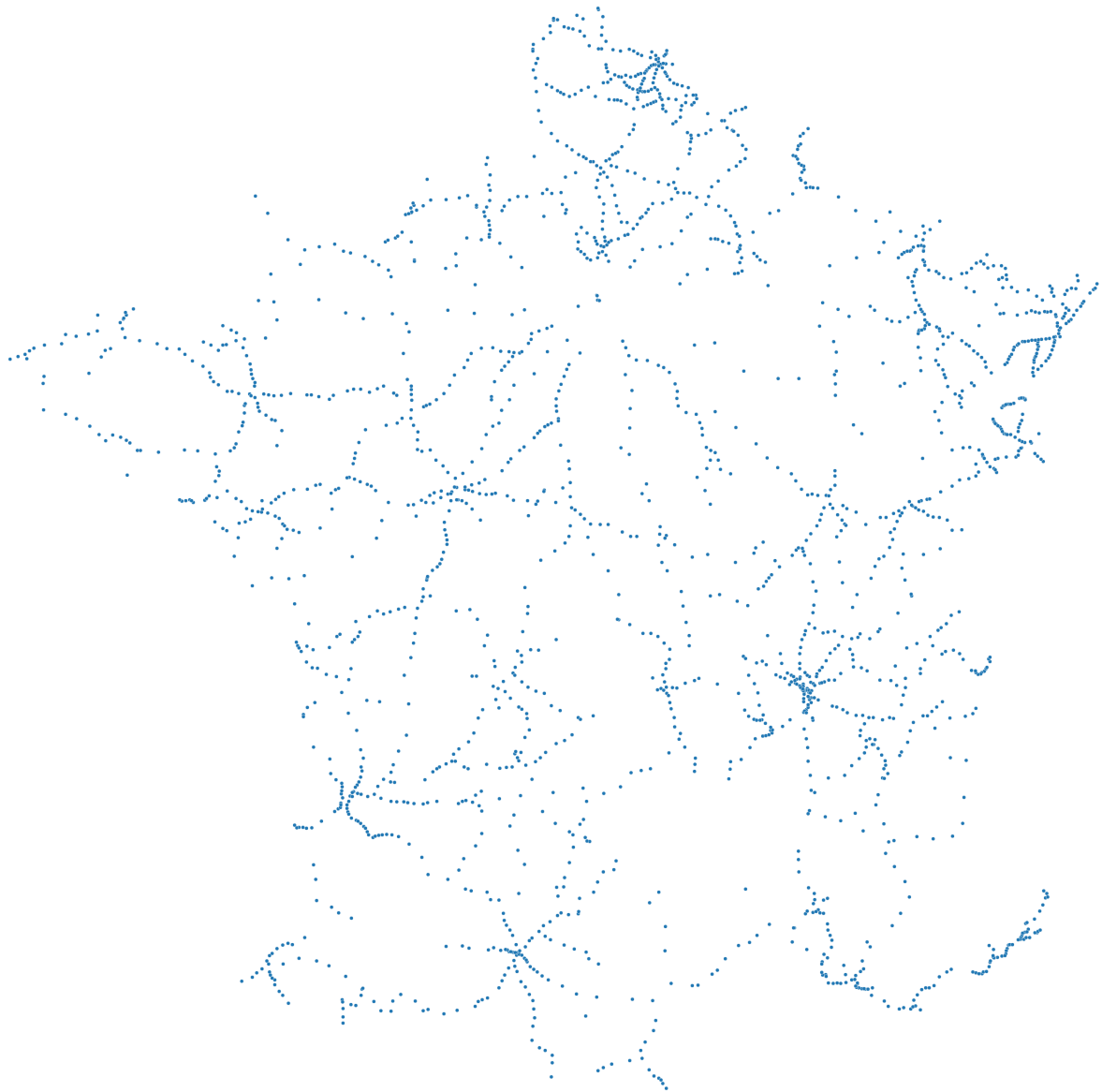


FIGURE 5 – Carte géographique des gares de TER retenues pour l'étude du réseau français. Les positions des gares ont été calculées en projection Mercator avec les coordonnées GPS.

échelle locale (celle des régions?), il semble y avoir des gares/villes centrales, par lesquelles passent une grande partie des lignes de TER environnantes. Cela motive une étude plus rigoureuse des centralités. Pour le réseau de TGV, on intuite que Paris ressortirait davantage en tant que ville centrale, très loin devant (en terme de centralité) les grandes villes de province comme Lyon et Bordeaux. Dans le cas du TER, on verra plus loin que cette différence de centralité entre Paris et les villes de province est certes visible mais pas si marquée.

Ajoutons désormais à notre représentation en Fig. 5 les liens entre les nœuds. Plus précisément, des liens sont ajoutés entre les gares pour lesquelles il existe un trajet de TER direct (men-

tionnés dans la [BDD tarifs](#)). Cela nous donne la représentation donnée en Fig. 6.

Sur le graphe en Fig. 6, on voit que le graphe représentant le réseau de TER est très dense dans certains endroits et peu dense sinon. Les zones du graphe les plus denses semblent être l'Alsace, les Hauts-de-France, la Côte d'Azur, les bords de l'embouchure de la Loire et les régions lyonnaise, bordelaise et toulousaine. Certains endroits comme la Normandie, la Bretagne ou la région PACA semblent même déconnectés du reste du réseau TER. De manière générale, il semble que les gares au sein des régions sont particulièrement bien connectées entre elles, contrairement aux gares de régions distinctes. Cela motive évidemment une étude des communautés. Enfin, Paris apparaît bien comme une ville particulièrement centrale, point de rencontre de liaisons venues de toute les directions et allant bien au delà des limites de la région Ile-de-France.

3.3 Coefficient de clustering et propriété small-world

Parmi les propriétés essentielles des graphes réels, il y a le fait d'être "small-world". Qu'en est-il du réseau de TER français ?

Pour commencer, le coefficient de clustering est de 0.3, ce qui est bien supérieur à la densité qui est de seulement 0.63%. Pour être plus précis, nous avons créé plusieurs graphes aléatoires de même nombre de nœuds et même nombre de liens. La méthode de création des graphes aléatoire s'est basé sur la fonction `gnm_random_graph` du module `networkx` qui choisit un graphe avec une probabilité uniforme parmi l'ensemble de tous les graphes à n nœuds et m arêtes. On obtient alors une moyenne du coefficient de clustering de 0,0063 pour ces graphes aléatoires. Le coefficient de clustering de notre graphe est donc environ 50 fois supérieur à celui d'un graphe aléatoire de même nombre de nœuds et de liens. Le premier critère pour être "small-world" est donc rempli.

De plus, la distance moyenne entre deux nœuds du graphe (*i.e.* la moyenne des plus courts chemins entre les paires de nœuds) est de 5.7, ce qui est relativement faible pour un graphe de cette taille (2 067 nœuds). À noter qu'on obtient un diamètre de 14. Ces valeurs peuvent sembler anormalement faibles, mais il faut se rappeler que les liens sont définis de telle sorte que c'est le graphe en Fig. 6 que l'on étudie, et non pas une carte des liaisons ferroviaires "classiques" (celle qu'on peut voir sur une carte de France ou sur Google Map).

On en conclut qu'on se situe bien dans le cas d'un réseau small-world. On peut supposer qu'en ajoutant les lignes de TGV, la distance moyenne et le diamètre seraient encore plus faibles puisqu'on rajoute des liens, qui plus est sur de longues distances.

Le fait d'être small-world nous dit que le passage d'un nœud à n'importe quel autre nœud du graphe se fait en empruntant très peu de liens, en moyenne 5.7, et au plus 14. Pour nous, chaque changement de lien correspond à une correspondance en gare. Par exemple, si je passe de Marseille à Paris en suivant les liens Marseille-Lyon puis Lyon-Paris, alors j'ai fait une correspondance à Lyon. Maintenant, on peut se demander quelles sont les gares de correspondance les plus empruntées. Autrement dit, on aimerait étudier les centralités des gares.

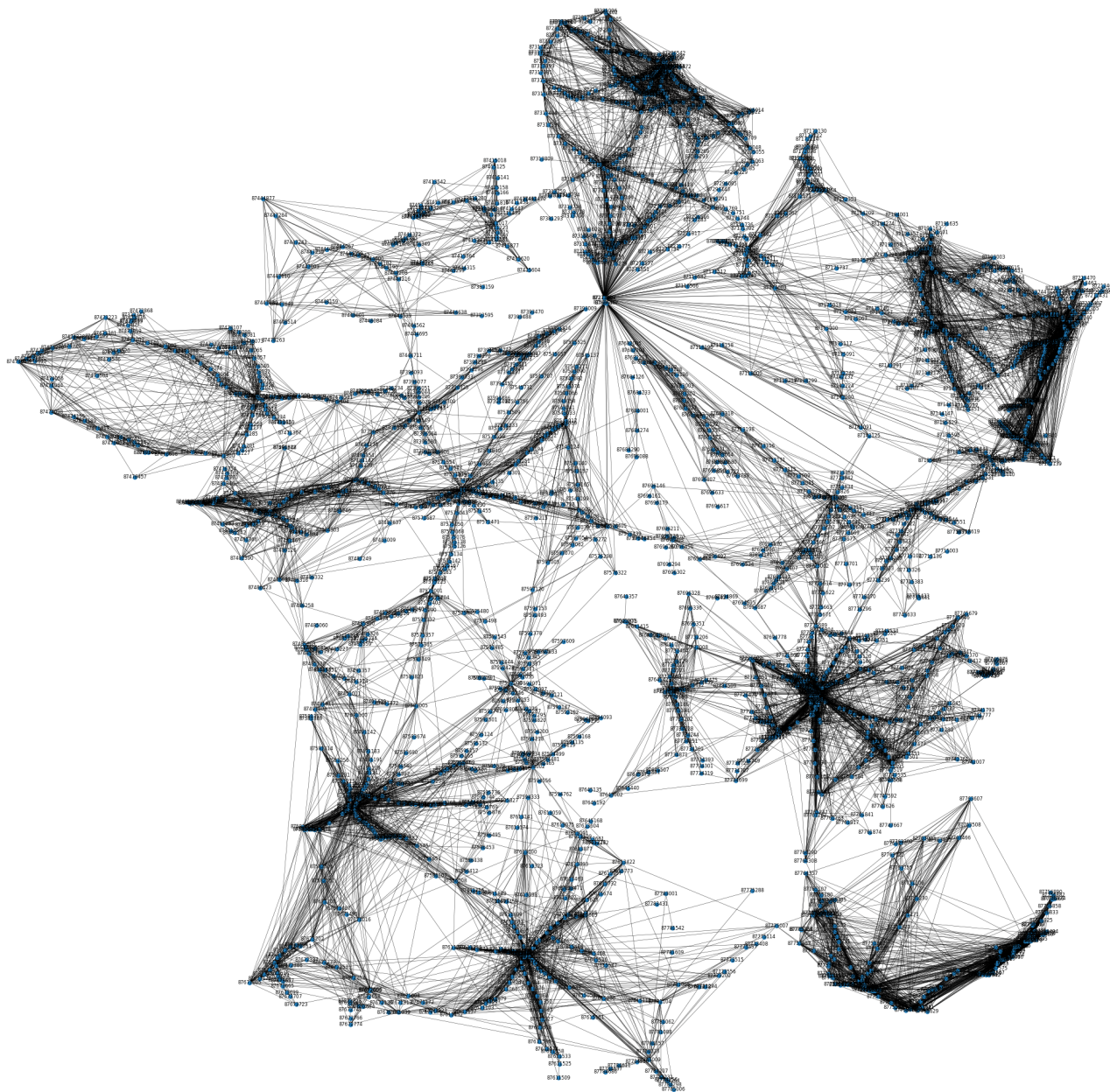


FIGURE 6 – Graphe du réseau de TER français. On rappelle que les nœuds sont les gares de TER et les liens sont les liaisons directes en TER entre ces gares. Les gares sont positionnées par leurs coordonnées GPS avec une projection Mercator.

3.4 Étude des communautés

L'étude des communautés du graphe a été faite avec l'algorithme de Louvain en tentant de minimiser le nombre de communautés. Il s'est trouvé que le nombre minimal de communautés est atteint pour un paramètre de résolution de 0.9, soit très proche de la valeur par défaut (1). Nous obtenons le découpage en communautés visible Fig. 7.

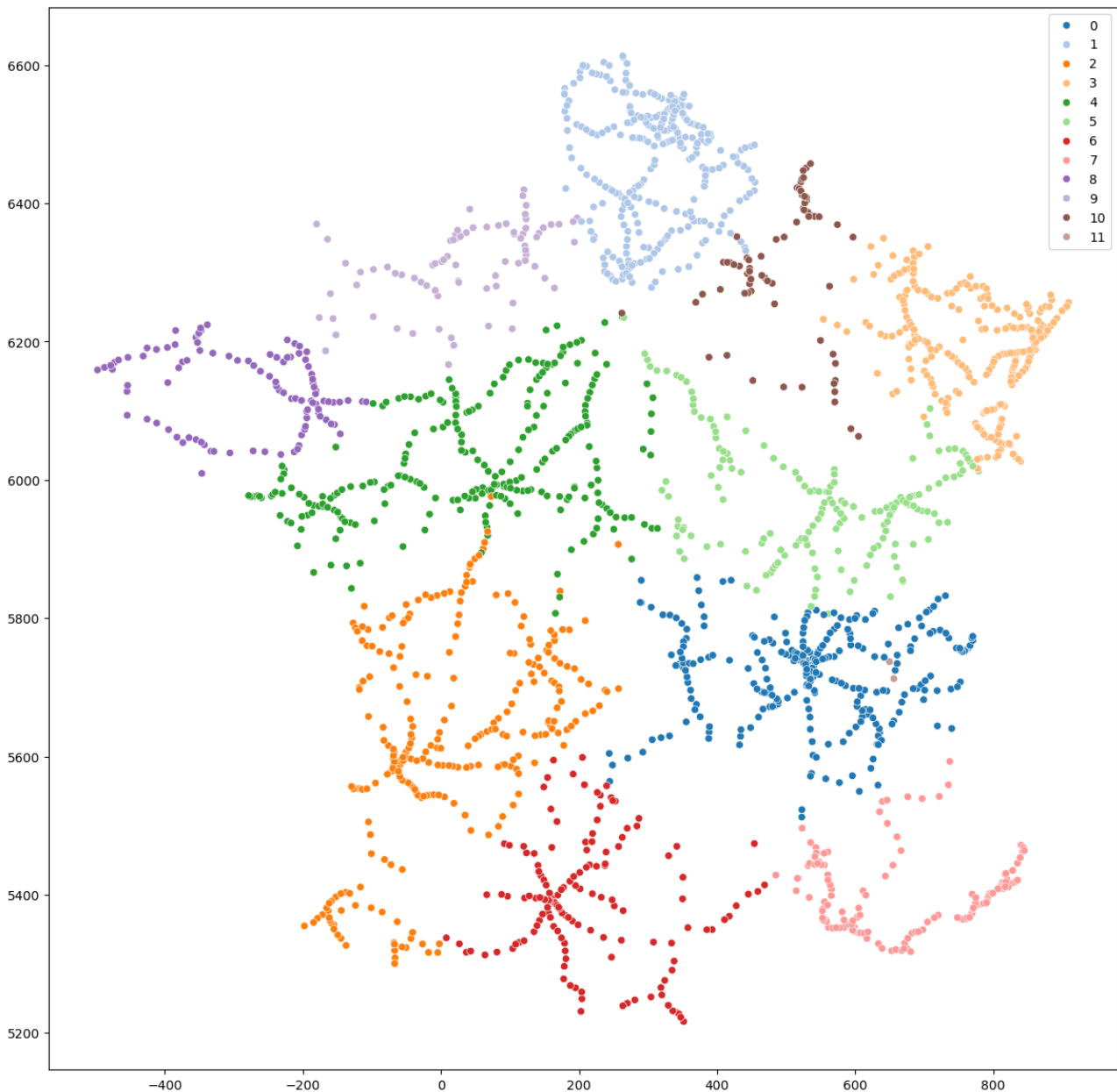


FIGURE 7 – Graphe du réseau de TER français faisant apparaître par des couleurs les communautés trouvées par la méthode de Louvain. Comme avant, les nœuds sont les gares de TER qui sont positionnées par leurs coordonnées GPS avec une projection Mercator. Les axes sont exprimés en kilomètres et ont été laissés pour l'échelle.

Comme attendu, les communautés correspondent à peu près aux régions. Cependant, il ne

s'agit pas des anciennes régions (Lorraine, Alsace, Champagne-Ardenne, etc.) mais des nouvelles régions (Grand-Est, etc.). Est-ce que le réseau de TER a évolué à la suite de la réforme des régions ? Ou a-t-il toujours été structuré ainsi, ce qui monterait quelque part la pertinence des choix de regroupement de régions qui ont été fait lors de la réforme ? Il est difficile de donner des réponses avec notre graphe comme seul support. Par ailleurs, on peut noter quelques différences particulièrement intéressantes entre le découpage en communautés et en nouvelles régions :

- L'Ile-de-France n'ayant pas un réseau TER très développé, elle est "partagée" entre les différentes régions adjacentes qui récupèrent les gares de Paris en fonction de leur destination (gare de Lyon, gare de l'Est, etc.).
- Les régions Pays-de-la-Loire et Centre-Val-de-Loire sont fusionnées. Cela ne paraît pas aberrant car lors de la formation des nouvelles régions, la région Centre projetait une fusion avec les Pays-de-la-Loire, région qui préférait une fusion avec la Bretagne. Cette dernière fusion a été refusée, menant au découpage actuel.
- La région Grand Est se retrouve amputée de l'ancienne région Champagne-Ardenne. Je (Louhane) suis de Lorraine et cela ne m'étonne pas. En effet, l'Alsace-Moselle, et par extension l'Alsace-Lorraine, possède une identité culturelle forte (notamment dû au passé commun avec l'Allemagne) qui n'est pas partagé avec la Champagne-Ardenne.
- deux gares en Rhône-Alpes sont systématiquement isolées dans une communauté de 2 représentants. Nous ne l'expliquons pas et pensons que cela pourrait être un artefact de la méthode.

Pour évaluer quantitativement la pertinence des communautés créées, nous avons calculé quelques caractéristiques des communautés, reportées dans la Table 1.

Nom	Coefficient de Clustering	Densité	Embeddeness
Moyenne sur les communautés	0.59	0.0768	0.961

TABLE 1 – Table de la moyenne de différentes caractéristiques des communautés créées avec la méthode Louvain

On peut notamment constater que le coefficient de clustering à l'intérieur des communautés est de 0.59, soit 2 fois supérieur à celui du réseau entier. De même, la densité moyenne de chaque communauté est de 0.0768, soit plus de 10 fois celle du réseau complet. Nous avons aussi calculé l'embeddeness des communautés. L'embeddeness d'un nœud est défini comme le ratio de son degré à l'intérieur de sa communauté divisé par son degré total. L'embeddeness E d'une communauté C correspond à l'embeddeness moyenne de tous ces nœuds (k_i étant les degré total du nœud i et k_i^C le degré du nœud i à l'intérieur de sa communauté) :

$$E = \frac{1}{|C|} \sum_{i \in C} \frac{k_i^C}{k_i}$$

Nous obtenons un moyenne de 0.961. Cela veut dire que 96% des liens du réseau sont expliqués par des liens à l'intérieur des communautés. Tous ces résultats permettent d'assurer la qualité des communautés créées.

Par ailleurs, on a précédemment remarqué que le découpage en communautés ressemble beaucoup aux nouvelles régions. Mais à quel point cela est-il vrai ? Nous avons souhaité quantifier cela en calculant l'information mutuelle normalisée qu'il y a entre les communautés créées

par la méthode de Louvain et les communautés basées sur les régions. L'information mutuelle normalisée NMI entre 2 partitions U et V d'un même réseau correspond simplement à l'information mutuelle MI (dans le sens d'entropie H) entre les deux partitions, ramenée à un résultat entre 0 et 1. La valeur 0 signifie qu'aucune information n'est partagée entre les 2 partitions, et la valeur 1 qu'elles sont identiques.

$$NMI(U, V) = \frac{MI(U, V)}{\max(H(U), H(V))}$$

Nous avons donc créé des communautés basées sur les nouvelles régions grâce au code géographique donnant le code département (les régions n'étaient pas directement disponibles dans les données). À partir d'un fichier Excel donnant la région à partir du code département, nous avons pu automatiser la procédure. Au final, nous obtenons une information mutuelle normalisée entre le découpage en régions et le découpage en communautés créées par l'algorithme de Louvain de 0.89. Cela signifie grossièrement que l'on retrouve 89% de l'information du découpage en région dans les communautés créées. Il s'agit d'un critère quantitatif de similarité entre les 2 découpages qui permet de conclure quant à la pertinence de considérer les communautés créées comme quasi-identiques aux régions.

3.5 Étude des centralités

3.5.1 Degrés

Une première façon de se donner une idée de la centralité est l'étude des degrés des nœuds. L'histogramme des degrés est visible en Fig. 8. On a tout d'abord tracé l'histogramme linéairement, puis en semi-log, et enfin en échelle log-log. L'échelle semi-log laisse penser que la distribution des degrés suit une loi géométrique : le nombre de nœuds chute d'un facteur 3 tous les 7 degrés. Cependant, on peut observer en échelle log-log que la distribution des degrés correspond à 2 droites. Une observation empirique des pentes permet de conclure que la distribution de degré commence ainsi avec une loi de puissance $r^{-\alpha}$, avec $\alpha = 0.75$ puis une autre loi de puissance $r^{-\beta}$ avec $\beta = 3$, après un cut-off à un degré environ égal à 20. La présence d'un échelle de degré caractéristique (ici, 20) montre que notre réseau n'est pas à proprement parler "scale-free" (qui par définition, ne comporte pas d'échelle caractéristique). Cependant, si l'on se cantonne à un certain intervalle de degrés, on constate que le réseau est "scale-free" à l'intérieur de cette intervalle de degrés.

Le tableau des degrés montre que les nœuds de plus forts degrés sont souvent des métropoles importantes (Strasbourg, Lyon, Bordeaux, Lille, Toulouse, etc.) ou un peu moins importantes mais situées dans les parties denses du réseau (Nancy, Douai, Amiens, etc.). Dans le deuxième cas, on peut voir les gares de ces villes comme des gares "relais" (par exemple Nancy relais de Strasbourg, et Metz relais de Nancy).

3.5.2 Betweenness

L'étude de la betweenness nous en apprend plus sur les gares jouant le rôle de "pont", c'est-à-dire permettant de relier différentes zones entre elles. Le tableau en Fig. 9 donne les 25 gares de betweenness les plus élevées. En plus des gares de grandes villes telles que Lyon, Bordeaux,

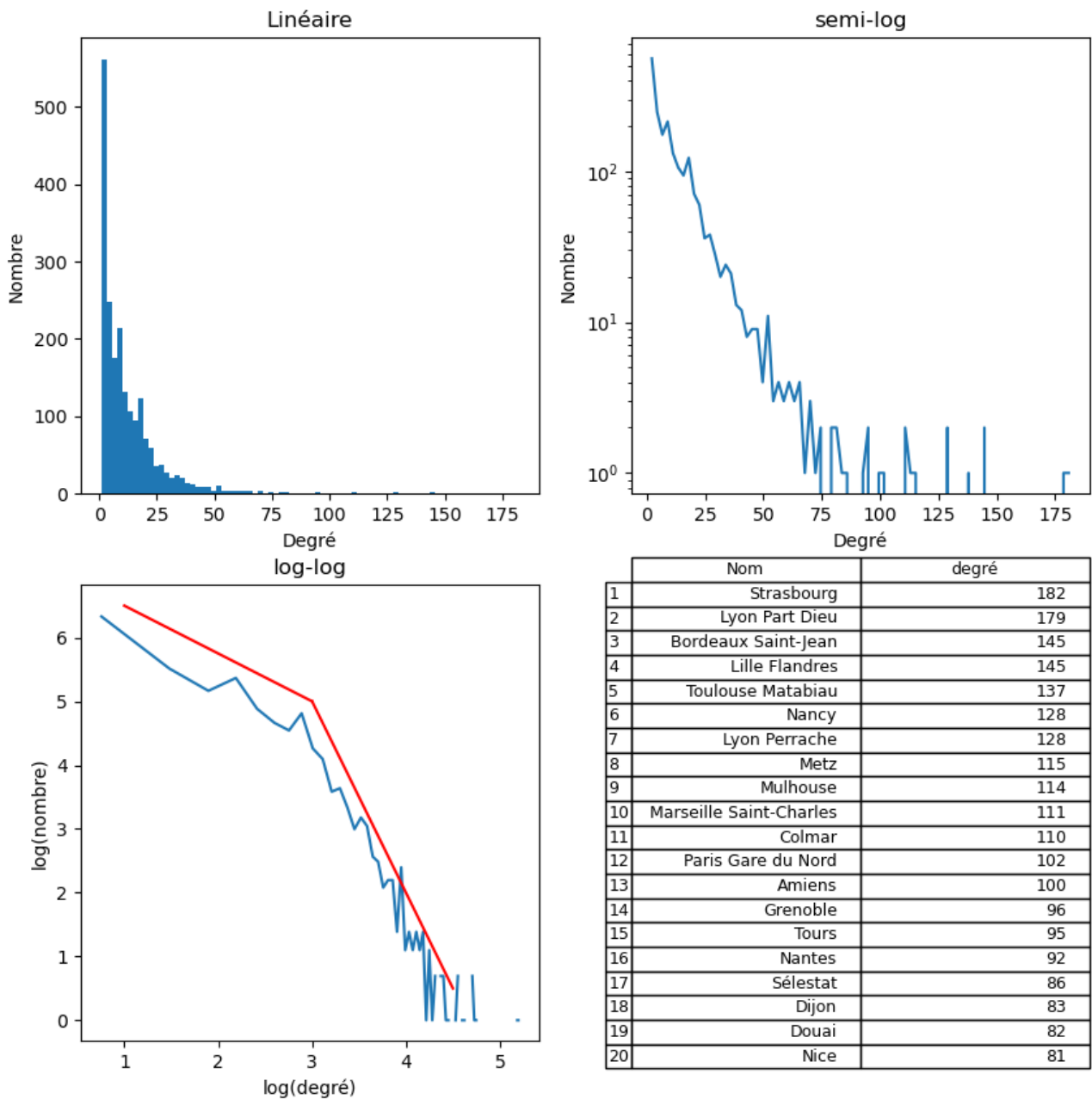


FIGURE 8 – Étude des degrés des nœuds du graphe français. En haut à gauche, l’histogramme des degrés des nœuds du graphe. En haut à droite, la distribution des degré en semi-log. En bas à gauche, la distribution des degrés en log-log. Les droites rouges positionnées au dessus de la distribution ont des pentes de -0.75 et -3 respectivement. En bas à droite, le tableau des 20 gares de plus haut degré, rangées par ordre décroissant.

Paris Est, Nantes, Toulouse etc., on retrouve également de nombreuses petites gares qui jouent un véritable rôle de pont entre régions :

- Besançon et Belfort sont des lieux de passage pour relier le Nord-Est au Sud-Est de la France ;
- Montluçon, situé au Nord de l’Auvergne, est la seule gare entre Paris et la Méditerranée reliant la France d’Est en Ouest ;

- Alençon est la seule gare à relier la Normandie sur toute sa frontière sud ;
- Avignon et Lorient-sur-Drôme relie la région Auvergne-Rhône-Alpes à la région PACA et est un passage obligé pour les itinéraires Nord-sud à l'Est de la France (à cause des Alpes et de la Lozère) ;
- La Rochelle relie la Nouvelle Aquitaine aux Pays-de-la Loire.

D'autres gares ont la chance de se situer au centre de la France et ont ainsi une forte betweenness. C'est notamment le cas de Bourges, Limoges, Amiens.

	Nom	Betweenness
1	Lyon Part Dieu	0.269
2	Besançon Viotte	0.266
3	Belfort	0.248
4	Bordeaux Saint-Jean	0.227
5	Montluçon	0.176
6	Paris Est	0.173
7	Bourges	0.143
8	Limoges Bénédictins	0.13
9	Tours	0.108
10	Paris Bercy Bourgogne - Pays d'Auvergne	0.103
11	Rouen Rive Droite	0.102
12	Alençon	0.091
13	Amiens	0.088
14	Nantes	0.078
15	Avignon Centre	0.076
16	Douai	0.075
17	Angers Saint-Laud	0.074
18	Toulouse Matabiau	0.072
19	Dijon	0.068
20	Lorient-sur-Drôme	0.066
21	La Rochelle	0.06
22	Nevers	0.058
23	Strasbourg	0.054
24	Culmont - Chalindrey	0.053
25	Le Mans	0.053

FIGURE 9 – Tableau des 25 gares de betweenness les plus élevées, classées par ordre décroissant.

3.6 Étude de la distance

On rappelle tout d'abord que l'attribut de distance entre les nœuds est la distance à vol d'oiseau. Les histogrammes de la distribution des distances entre les nœuds est visible Fig. 10. On observe en premier lieu une rapide croissance du nombre de paires de nœuds, puis une décroissance géométrique visible dans le graphe semi-log de la distribution (une augmentation de 100 km entraîne une diminution du nombre de nœuds d'un facteur 10).

La forme de cette distribution est conforme aux attentes. En effet, si la distance est de l'ordre du kilomètre, il n'y a pas lieu de prendre le TER pour se déplacer. Puis la probabilité

d'un trajet de TER direct entre deux gares augmente avec la distance entre ces gares jusqu'à atteindre un maximum à 15 km. À partir de ce maximum, la probabilité descend pour deux raisons couplées : il est peu probable que beaucoup de personnes souhaitent relier deux villes éloignées en prenant le TER ; et si un trajet existe, il ne sera pas direct car le TER est surtout destiné à desservir une région donnée. Ainsi, les trajets de TER directs sont principalement conçus pour relier des villes entre 10 et 200 km, ce qui correspond au sens commun.

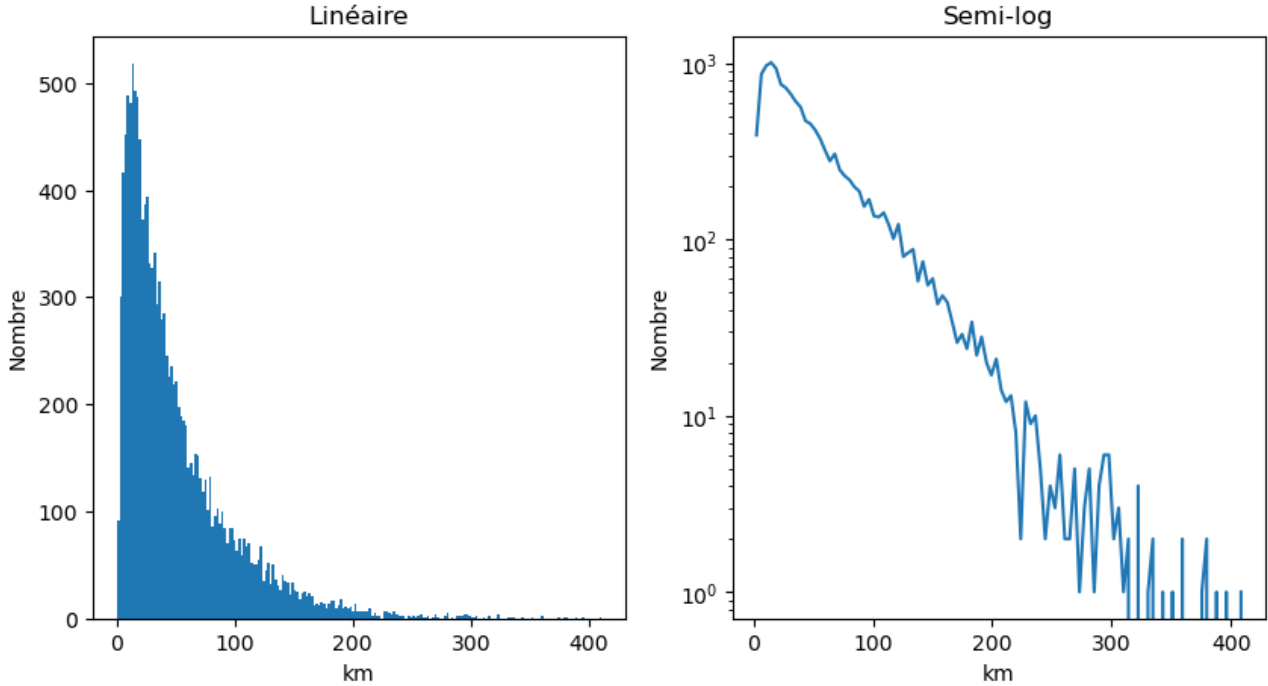


FIGURE 10 – Représentation de la distribution des distances des nœuds. À gauche, l'histogramme linéaire. À droite, le graphe en semi-log.

3.7 Étude des corrélations entre attributs des nœuds

Jusqu'à présent, nous avons étudié quelques attributs des nœuds comme les degrés ou la betweenness. Dans cette partie, nous souhaitons étudier le lien de corrélation entre ces 2 attributs et les attributs des gares comme la population des municipalités contenant les gares, ou encore la fréquentation des gares. Pour cela, nous avons calculé les corrélations entre toutes ces variables sur l'ensemble des gares. La Fig. 11 présente la matrice des corrélations sur l'ensemble des gares. Pour chaque couple X, Y d'attribut des gares, le coefficient de la matrice de corrélation est :

$$M_{X,Y} = \frac{\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle}{\sqrt{\langle (X - \langle X \rangle)^2 \rangle \langle (Y - \langle Y \rangle)^2 \rangle}}$$

Naturellement, la matrice est symétrique, avec des 1 en diagonale. Celle-ci comporte donc les attributs : degré, betweenness, fréquentation de l'année 2022, population de l'année 2020, et le Page rank que nous avons calculé pour tous les nœuds du réseau.

Certaines corrélations sont intéressantes à analyser :

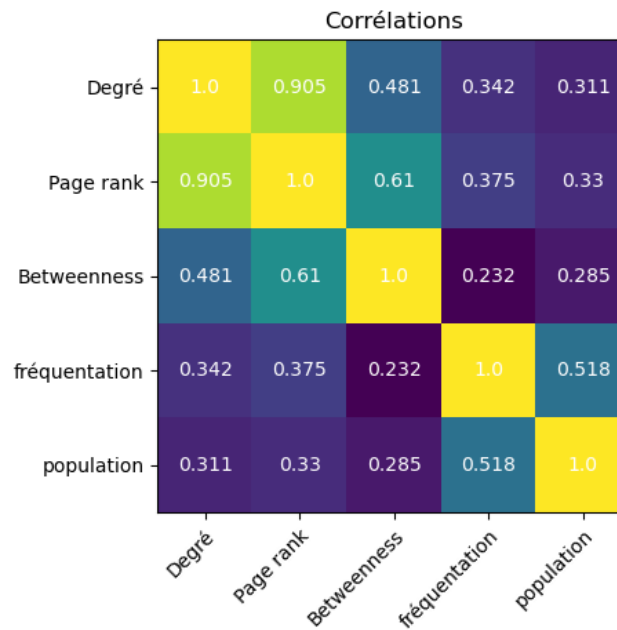


FIGURE 11 – Matrice de corrélations des attributs sur les nœuds.

- La corrélation fréquentation - population, qui est de 0.52 environ, est plus faible que ce que l'on prévoyait. Il est probable que la population soit la population municipale et non la population de toute l'aire urbaine. Ainsi, une gare située dans une petite commune de l'agglomération d'une grande métropole, très fréquentée, aura une population relativement faible, alors que sa fréquentation est bien plus grande que dans une ville de même taille mais isolée géographiquement.
- Nous pouvons observer que le degré et le Page rank sont les attributs les plus corrélés (0.9). Ceci s'explique facilement : comme on l'a vu, les gares de plus haut degré sont des métropoles centrales dans chacune de leur région. Elles sont reliées à un très grand nombre de gares de leur région. Elles sont donc très susceptibles d'être parcourues lors d'une marche aléatoire. De même, les gares de faible degré sont non seulement isolées par leur petit degré, mais aussi cloisonnées à l'intérieur de leur région, et donc encore moins susceptibles d'être visitées.
- La corrélation degré - betweenness, quant à elle, est plus faible, mais toujours bien présente (0.48). Certes, l'augmentation du degré augmente la probabilité que celle-ci serve à réaliser des ponts entre les parties du réseau, mais cela ne suffit pas car des gares de degré plutôt moyen, grâce à leur bonne position géographique, peuvent se retrouver propulsées à la tête du classement de la betweenness.
- La corrélation Page rank - betweenness (0.61) est en quelque sorte équivalente à la corrélation degré - betweenness puisque le degré et le Page rank sont très corrélés entre eux (0.9).
- D'une manière générale, les corrélations entre les attributs calculés tels que le degré, le Page rank, la betweenness sont très peu expliqués par la population ou la fréquentation (coefficients de corrélation un peu au-dessus de 0.3). Cependant, on peut voir que la corrélation est systématiquement plus faible avec la betweenness pour les mêmes raisons que celles susmentionnées.

3.8 Conclusion

Dans cette première partie sur le réseau de TER français, nous avons étudié un graphe de 2 069 nœuds et 13 339 liens.

Pour commencer, le réseau créé est bien de type small-world, avec un coefficient de clustering 50 fois supérieur à celui d'un graphe aléatoire de même nombre de nœuds et de liens, et une distance moyenne entre les nœuds de 5.7.

Nous avons également vu que la structure du réseau est cohérente avec le découpage en régions, ce qui correspond à l'intuition qu'on peut avoir pour un réseau de TER (R comme région). En effet, les communautés obtenues par l'algorithme de Louvain correspondent grossièrement (89% d'information mutuelle normalisée) aux nouvelles régions.

Au sein de ces communautés, on observe aussi des gares de haut degré et centrales, souvent situées dans de grandes métropoles, mais pas que. En effet, l'étude de la betweenness permet de déceler de nombreuses gares "ponts", de taille moyenne, mais bien placées géographiquement pour relier des parties importantes du réseau.

Par ailleurs, l'étude des degrés donne une distribution en 2 lois de puissances, l'une pour les faibles degrés ($r^{-\alpha}$ avec $\alpha = 0.75$), et une autre pour les grands degrés ($r^{-\beta}$ avec $\beta = 3$) après un "cut-off" à un degré égal à 20. Le réseau n'est donc, pour ainsi dire, scale-free qu'à certaines échelles, ce qui est une antithèse...

Enfin, l'étude de la distance et l'étude des corrélations entre les attributs sur les nœuds donnent des résultats conformes aux attentes, à ceci près que fréquentation et population sont moins corrélées qu'attendu. Cela s'explique peut-être par le fait que les gares TER de grandes métropoles se trouvent parfois en périphérie de la zone urbaine, c'est-à-dire dans des villes de population moyenne.

4 Création et étude du réseau ferré allemand

4.1 Base de données

Une seule base de données a été utilisée pour créer le réseau allemand. Elle provient directement du [site de la Deutsch Bahn](#), qui est l'équivalent de la SCNF en Allemagne.

Le document à disposition est un fichier csv qui contient l'ensemble des infrastructures au sens large du réseau. Dans la grande majorité des cas, ces infrastructures sont des gares mais cela peut être aussi des aiguillages ou des centres de maintenance. Chaque ligne du fichier csv correspond à une infrastructure et donne : la ligne sur laquelle elle se trouve, son point kilométrique (PK) sur la ligne, son nom, le type d'infrastructure et ses coordonnées GPS. La Fig. 12 montre à quoi ressemble le fichier csv.

Contrairement au cas de la France, il n'a pas été possible de trouver des bases de données avec les prix des trajets. De plus, l'absence de code géographique (COG) ne permet pas de remonter facilement à la commune de chaque gare pour la relier à une population. Cependant, il est possible de construire à partir de ces données, notamment grâce à l'indication des lignes, un graphe équivalent à celui étudié dans le cas du réseau TER de la France.

STRECKE_NR	RICHTUNG	KM_I	KM_L	BEZEICHNUNG	STELLE_ART	KUERZEL	GK_R_DGN	GK_H_DGN	GEOGR_BREITE	GEOGR_LAENGE
1000 0	117290016	172.9 + 16	Flensburg Weiche	Bf	AFW	3525691.331	6069974.991	54.75848402	9.39789295	
1000 0	117440000	174.4 + 0	Flensburg Friedensweg	Abzw	AFR	3525333.479	6071281.835	54.77024153	9.39244801	
1000 0	117440000	174.4 + 0	Flensburg Friedensweg	Abzw	AFR	3525333.479	6071281.835	54.77024153	9.39244801	
1000 0	117970084	179.7 + 84	Flensburg Grenze	Gp	AXFW	3523405.063	6076279.823	54.81523173	9.36288704	
1005 0	100000000	0.0 + 0	Flensburg	Bf	AF	3528132.808	6071715.352	54.7739873	9.43598426	
1005 0	100160086	1.6 + 86	Flensburg Süd	Awanst	AFS	3528716.414	6070532.501	54.76332888	9.44493623	
1005 0	100540095	5.4 + 95	Flensburg Friedensweg	Abzw	AFR	3525333.479	6071281.835	54.77024153	9.39244801	
1011 0	101410058	14.1 + 58	Ohrstedt	Awanst	AOR	3514739.787	6041428.881	54.50248278	9.22635174	
1011 0	102620091	26.2 + 91	Husum	Bf	AHM	3503635.17	6038048.565	54.47231461	9.05489896	
1012 0	104900086	49.0 + 86	Husum	Bf	AHM	3503632.025	6038038.466	54.47222391	9.05485033	
1020 0	100560000	5.6 + 0	Kronshagen	Hp	AKRH	3570547.122	6022844.919	54.33087847	10.08328685	
1020 0	100770050	7.7 + 50	Suchsdorf	Bf	ASU	3571002.097	6024873.725	54.3490387	10.09076172	
1020 0	101370000	13.7 + 0	Neuwittenbek	Bf	ANW	3566936.257	6027229.796	54.37075068	10.02877396	
1020 0	101910049	19.1 + 49	Gettorf	Bf	AGF	3563617.979	6031538.251	54.40987554	9.97864485	
1020 0	103100097	31.0 + 97	Eckernförde	Bf	AEC	3554199.841	6037876.008	54.46789424	9.83476629	
1020 0	103970036	39.7 + 36	Riesby	Hp	ARY	3553174.592	6046090.961	54.54179484	9.82043249	
1020 0	105130034	51.3 + 34	Süderbrarup	Bf	ASUE	3549873.269	6056619.297	54.63670267	9.77121319	
1020 0	106250003	62.5 + 3	Sörup	Bf	ASOE	3543207.206	6065471.207	54.71683049	9.66929042	
1020 0	107060039	70.6 + 39	Husby	Hp	AY	3537093.133	6070716.591	54.76443542	9.57508781	
1020 0	108060043	80.6 + 43	Flensburg	Bf	AF	3528150.877	6071750.32	54.77430041	9.43626842	
1022 0	100000094	0.0 + 94	Kiel Hbf	Bf	AK	3573701.043	6021115.14	54.31489643	10.13133244	
1022 0	100290000	2.9 + 0	Kiel-Hassee Citipark	Bf	AKH	3571666.628	6020388.812	54.30861143	10.09990379	
1022 0	100530037	5.3 + 37	Kiel-Russee	Hp	AKRIU	3569732.324	6020516.855	54.31007883	10.07022119	
1022 0	100790055	7.9 + 55	Melsdorf	Hp	AME	3567149.468	6020779.828	54.31278646	10.03060365	
1022 0	101250077	12.5 + 77	Achtenwehr	Hp	AACW	3562818.171	6020970.002	54.31504484	9.96410269	
1022 0	101380084	13.8 + 84	Felde-Brandsbek	Bf	AFEB	3561734.839	6020378.112	54.30985996	9.9473365	
1022 0	101910057	19.1 + 57	Bredenbek	Hp	ABRB	3565602.006	6020890.729	54.31506924	9.86704583	
1022 0	102730078	27.3 + 78	Schüllorf	Hp	ASC	3548727.734	6018986.199	54.29875959	9.74728795	
1022 0	103050088	30.5 + 88	Osterrönfeld Awanst	Awanst	AONA	3547260.528	6017048.155	54.28148753	9.72444814	
1022 0	103090055	30.9 + 55	Osterrönfeld	Bf Abzw	AON	3546957.412	6017255.029	54.2833738	9.71982705	
1023 0	100000047	0.0 + 47	Kiel Hbf	Bf	AK	3573738.548	6021114.633	54.31488647	10.1319085	
1023 0	100250037	2.5 + 37	Kiel Hbf (Ss)	Abzw	AKSS	3574569.204	6019010.213	54.29586389	10.14414528	
1023 0	100260021	2.6 + 21	Kiel-Wellsee	Awanst	AKEW	3574646.348	6018971.178	54.29560203	10.14532014	
1023 0	100510086	5.1 + 86	Kiel-Elmschenhagen	Bf	AKE	3576915.7	6018085.276	54.28720888	10.17993938	
1023 0	100800056	8.0 + 56	Raisdorf West	Awanst	ARSW	3579730.526	6017772.906	54.28397263	10.22306962	
1023 0	100950017	9.5 + 17	Raisdorf	Bf	ARS	3581067.19	6017473.161	54.28107051	10.24350727	
1023 0	101530076	15.3 + 76	Preetz	Bf	APZ	3583256.942	6012289.321	54.23415938	10.27566527	
1023 0	102660098	26.6 + 98	Ascheberg (Holst)	Bf	AAG	3587620.249	6002843.949	54.14859774	10.33983064	
1023 0	103300056	33.0 + 56	Plön	Bf	APLN	3592996.331	6004159.942	54.15947367	10.42248758	
1023 0	104240005	42.4 + 5	Bad Malente-Gremsmühlen	Hp Ust	AMAG	3601346.273	6005215.922	54.1673798	10.55061923	
1023 0	104770044	47.7 + 44	Eutin	Bf	AEU	3605284.432	6001765.437	54.13560109	10.60969035	
1023 0	106200020	62.0 + 20	Neustadt (Holst) Gbf	Bf	ANN	3617493.42	5997611.251	54.09565558	10.79477921	
1023 0	106320099	63.2 + 99	Neustadt (Holst)	Hp Awanst	ANT	3618327.677	5998544.224	54.10384262	10.80788866	
1031 0	110100000	101.0 + 0	Meimersdorf	Bf Abzw	AMR	3572244.992	6017794.999	54.28528183	10.10816197	
1032 0	110290076	102.9 + 76	Meimersdorf Ost	Bf Abzw	AMRO	3573937.899	6018625.84	54.29250297	10.13435623	
1033 0	110100000	101.0 + 0	Meimersdorf	Bf Abzw	AMR	3572248.778	6017786.338	54.2852035	10.108218	
1033 0	110290076	102.9 + 76	Meimersdorf Ost	Bf Abzw	AMRO	3573937.899	6018625.84	54.29250297	10.13435623	

FIGURE 12 – Structure de la base de données pour le réseau de train allemand

La base de données comporte 9 505 lignes. Cependant, ces 9 505 entrées ne correspondent pas à autant de gares pour plusieurs raisons.

Premièrement, une gare apparaît autant de fois qu'il existe de lignes ferrées passant par elle. Ainsi, les grosses gares régionales font apparaître plusieurs lignes dans le fichier. Une clé unique pour cette table est donc le tuple formé par le numéro de ligne et le nom de l'infrastructure.

Deuxièmement, cette base de données liste toutes les infrastructures existantes sur le réseau ce qui comprend des infrastructures pour le transport de marchandises (frets) mais aussi par

exemple les aiguillages, des voies d'évitements, etc. Ce travail s'intéressant au transport de voyageurs, il est donc nécessaire de nettoyer la base de données en enlevant les infrastructures qui ne sont pas des gares. Pour cela, il est possible d'utiliser la colonne "STELLE_ART" qui donne le type d'infrastructure correspondant à chaque ligne à partir d'abréviations. Il en existe 33 différentes, où cependant certaines sont des combinaisons d'autres. Par exemple, il existe les types "Hp", "Abzw", "Bk", "Awanst" mais aussi "Hp Abzw", "Hp Bk" et "Hp Awanst". Le tableau en Fig.13 récapitule les 33 abréviations différentes, leur nom allemand ainsi que la traduction littérale française. Dans ce tableau, le vert indique que ce type d'infrastructure a été gardé pour le graphe et le rouge indique le contraire. Une représentation géographique de ce choix est donnée par la Fig.14. Cependant, il est parfois difficile de trancher à partir d'une seule abréviation et les combinaisons ne rendent pas cette tâche plus facile. Par exemple, "Hp" signifie "point d'arrêt" et correspond à un quai sans bâtiment physique. Dans le langage courant, cela peut correspondre au mot gare bien qu'il n'existe pas de gare physique à proprement parlé. C'est typiquement le genre d'infrastructures qui existe en milieu rural quand le nombre d'usagers ne justifie pas la construction d'un bâtiment mais que l'arrêt est nécessaire. Cependant, un point d'arrêt peut aussi être utilisé pour le (dé)chargement de marchandises. "sst" signifie "Schiffslandestelle" et correspond à un point de chargement mais il existe aussi le "Hp sst". Tous les points d'arrêts "Hp" ne correspondent donc pas à du transport de voyageurs.

Ce filtrage (voir tableau en Fig.13) fait passer le nombre de nœuds de 7 757 à 6 546, ce qui n'est pas négligeable.

Abréviations	Nom allemand	Traduction
Bf	Bahnhof	Gare
Abzw	Abzweigstelle	Jonction (sans plateforme)
Gp	Grenzpunkt	Point frontière
Awanst	Ausweichanschlussstelle	Point de connexion alternatif
Hp	Haltepunkt	Point d'arrêt
Bft Abzw	Bahnhofsteil	Partie d'une station
šst	Schiffslandestelle	Point de chargement
Bft	Bahnhofsteil	Partie d'une gare
Anst	Anschlussstelle	Point de connexion
Bk	Blockstelle	Point de blocage
Strw	Streckenwechsel	Changement d'itinéraire
Hp Dkst	Haltepunkt deckungsstelle	Point de couverture du point d'arrêt
Abzw a.B.	Abzweigstelle	Point de branchement
Museum		Musée
Bf Abzw	COMPOSITION	
Hp šst	COMPOSITION	
Hp Awanst	COMPOSITION	
Bf a.B.	COMPOSITION	
Hp a.B.	COMPOSITION	
Hp Bk	COMPOSITION	
Hp Anst	COMPOSITION	
Awanst a.B.	COMPOSITION	
Hp Abzw	COMPOSITION	
Anst a.B.	COMPOSITION	
Awanst Bk	COMPOSITION	
Bft a.B.	COMPOSITION	
šst Awanst	COMPOSITION	
šst a.B.	COMPOSITION	
Hp Bk a.B.	COMPOSITION	
Hp Anst Bk	COMPOSITION	
Anst Bk	COMPOSITION	
Abzw Anst	COMPOSITION	
Gp a.B.	COMPOSITION	

FIGURE 13 – Liste des abréviations correspondant aux différents types d'infrastructures listés dans la base de données. Dans "Nom allemand", "COMPOSITION" signifie que l'abréviation est une composition d'abréviations détaillées plus haut dans le tableau.

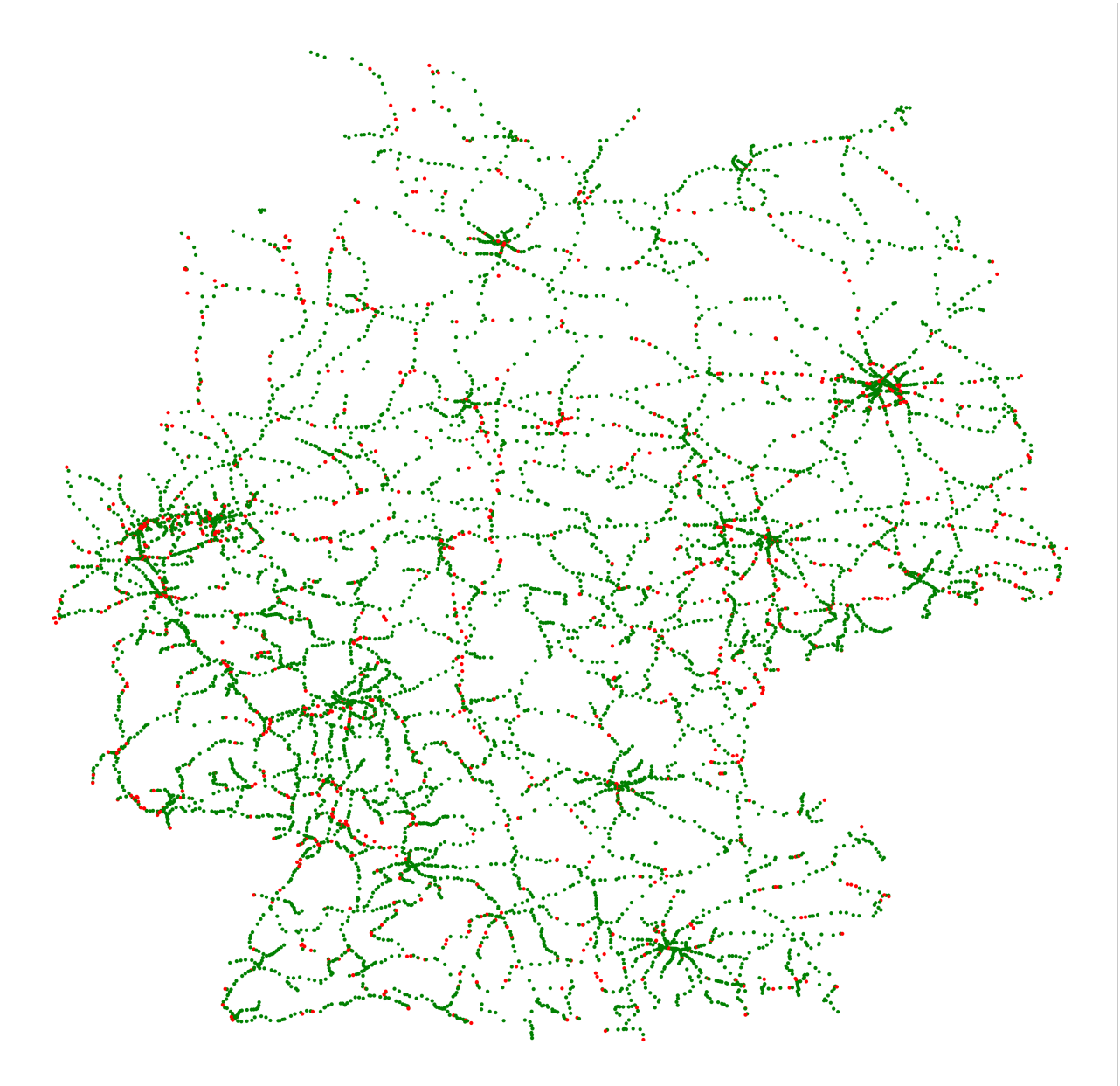


FIGURE 14 – Carte des nœuds du réseau, les points en vert correspondent aux nœuds gardés pour l'étude et les points rouges à ceux qui ont été supprimés.

4.2 Représentation "linéaire" du réseau ferré allemand

Pour ce réseau, on considère une arête entre deux gares (les nœuds) si et seulement si elles sont consécutives sur une même ligne. Avec cette structure, les arêtes représentent directement le réseau ferré, c'est-à-dire la voie physique qui relie les gares du réseau. De plus, la base de données contient les points kilométriques de chaque gare. Les distances entre les gares servent donc de poids sur ce graphe. C'est une manière de visualiser le réseau ferroviaire physique.

On peut remarquer que le graphe obtenu n'a pas qu'une seule composante connexe. Cela vient probablement de l'absence de certaines lignes dans la base de données. Il s'agit d'un point problématique pour calculer certaines grandeurs comme le diamètre. Afin de s'affranchir de ce problème, la plus grande composante connexe a été calculée. Le graphe linéaire possède 194 composantes et la plus grande comportait 87% de ses nœuds. Le graphe linéaire a un degré moyen légèrement supérieur à 2 comme attendu. En effet, la majorité des gares sont sur une unique ligne et sont donc reliées à la gare précédente et à la gare suivante. Seules quelques grandes gares régionales forment des hubs entre les lignes, d'où un degré moyen légèrement supérieur à 2.

De même, le nombre d'arêtes rapporté au nombre de nœuds est très faible (proche de 1) mais cela s'explique aussi par le fait que dans une ligne de n gares sans connexion avec d'autres lignes, il y a $n - 1$ arrêtes entre ces n nœuds. Comme la majorité des lignes ne présentent pas beaucoup de gares "hubs", cela explique le ratio de 1.

Enfin, le coefficient de clustering est aussi très faible de part la forme très "linéaire" du réseau. En effet, les lignes ferroviaires sont peu clusterisées, voir Fig.15.

L'ensemble des résultats est reporté dans le tableau 2.

	Graphe linéaire	Plus grande composante connexe
Nœuds	6546	5680
Arêtes	6636	5963
Nombre de composantes connexes	194	1
Degré moyen	2,03	2,10
Densité	0,00031	0,00037
Clustering moyen	0,0051	0,0058
Diamètre		1344,6 km
Plus court chemin moyen		467,9 km

TABLE 2 – Étude préliminaire du réseau ferré allemand représenté de façon linéaire et de sa plus grand composante connexe.

Les gares sont très bien réparties sur le territoire. Il est cependant possible d'observer une plus grosse densité autour des villes majeures comme Berlin, Munich, Stuttgart, Nuremberg ou encore Cologne ou Dortmund. Le nombre de gares très important dans ces grandes villes provient du fait qu'en Allemagne, ces dernières bénéficient d'un réseau de train de ville appelé S-Bahn bien développé et connecté autant à la banlieue qu'aux zones rurales proches.

Une région sort particulièrement du lot en terme de densité de gares : la Rhénanie du Nord - Westphalie. Cette dernière comporte un cluster de grosses villes avec de nombreuses gares dont Cologne, Dortmund, Düsseldorf, Essen ou Duisbourg. De plus, cette région est traditionnellement la région la plus industrialisée d'Allemagne de part ses mines de charbon et ses aciéries.

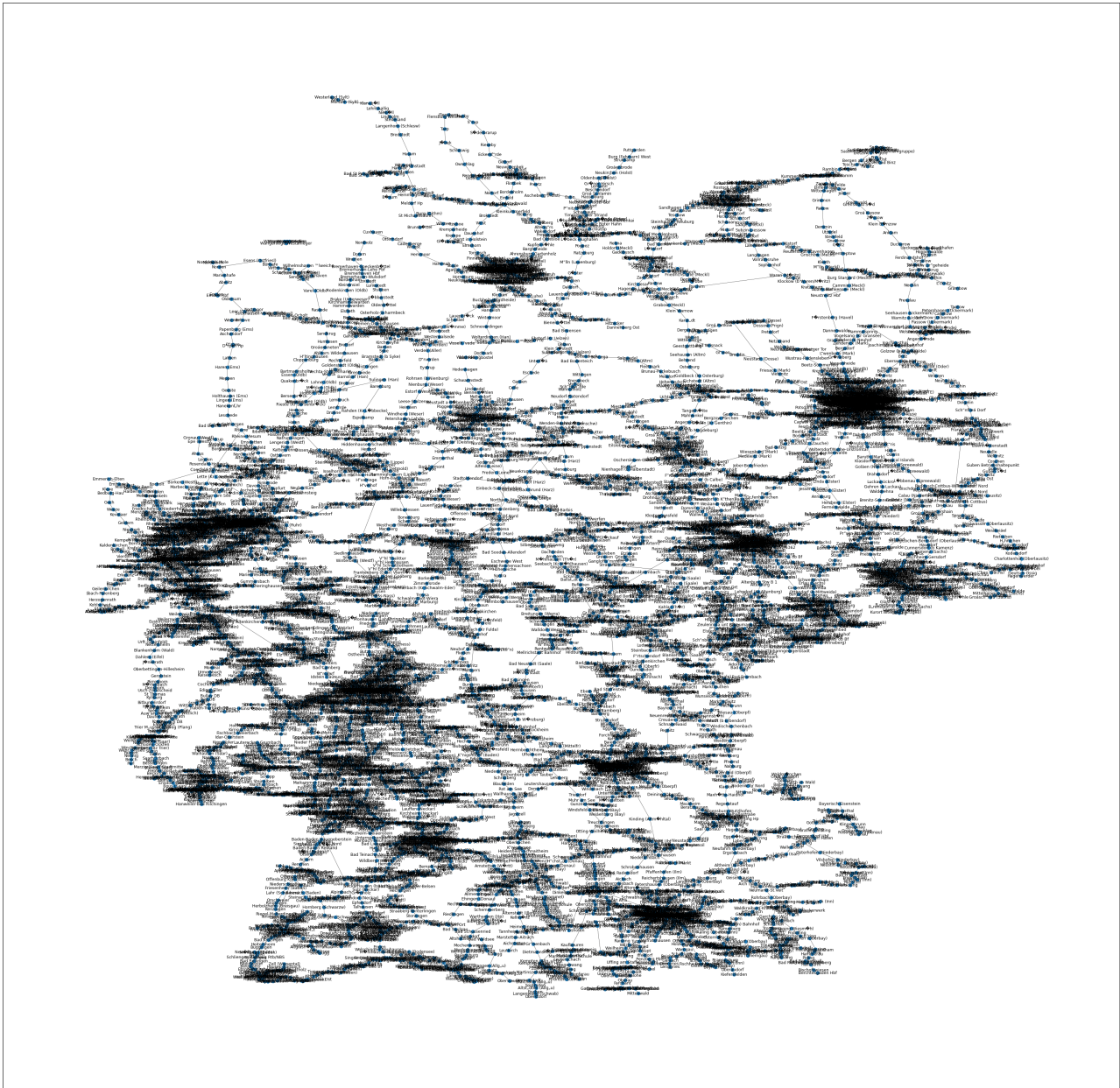


FIGURE 15 – Représentation dite "linéaire" du réseau ferroviaire de l'Allemagne.

Ces activités industrielles très demandeuses en réseau de transport ont donc favorisé l'apparition d'un réseau ferroviaire développé.

Les 10 gares avec les plus grands degrés et PageRank sont présentées tableau 3. Les noms des plus grandes villes allemandes sont présentes à l'exception de Francfort et de Berlin. Comment l'expliquer ? Sans doute que dans les grandes villes, le trafic ferroviaire se répartit entre plusieurs gares, qui sont donc chacune de taille moyenne. Or ce classement favorise les villes avec des grosses gares.

Graphe "linéaire"			
Top 10 Gares Degrés		Top 10 Gares PageRank	
Düsseldorf Hbf	10	Olching	0,000541
Nürnberg Hbf	8	Düsseldorf Hbf	0,000541
Halle (Saale) Hbf	7	Düsseldorf-Gerresheim	0,000537
München-Pasing	7	Forchheim (Oberfr)	0,000526
Mannheim Hbf	7	Hamburg-Harburg	0,000503
Oberhausen Hbf	7	Hockenheim	0,000487
Münster (Westf) Hbf	7	Offenburg	0,000482
Hamburg-Harburg	7	Dortmund Hbf	0,000479
Cottbus Hbf	6	Halle (Saale) Hbf	0,000478
München Ost Pbf	6	Landau (Pfalz) Hbf	0,000477

TABLE 3 – Liste des 10 gares de plus hauts degrés et respectivement de plus hauts PageRank.

4.3 Représentation "small-world" du réseau ferré allemand

4.3.1 Étude préliminaire du réseau

Pour ce réseau, on considère une arête entre deux gares (les nœuds) si et seulement si elles sont sur une même ligne, sans être forcément des gares consécutives sur cette ligne. En faisant l'hypothèse que les trajets des trains restent sur une ligne, cela revient à dire que deux gares sont reliées par une arête si et seulement si il est possible de les relier sans faire de correspondance. Cela permet d'établir une comparaison (bien qu'imparfaite) avec le graphe du réseau français qui est construit à partir de la liste des trajets TER directs entre les gares. Ce graphe est représenté Fig.16.

Comme pour le réseau "linéaire", certaines caractéristiques du graphe "small-world" sont présentées dans le tableau 4. Il a toujours le même nombre de composantes connexes car cette méthode de création de graphe ne fait qu'ajouter des arêtes entre des gares qui étaient déjà dans la même composante connexe. En revanche, le nombre d'arêtes est logiquement beaucoup plus important, de même que le degré moyen (maintenant de 21.47), la densité et le coefficient de clustering moyen. En effet, la méthode de construction de ce graphe "small-world" crée énormément de triangle le long des lignes d'où un coefficient de clustering proche de 1 pour ce graphe à l'inverse du graphe linéaire.

La répartition des degrés des nœuds est présentée Fig.17. Contrairement au graphe français, la distribution des degrés augmente dans un premier temps puis semble suivre une loi de puissance dans un second temps avec cependant beaucoup de bruit. Le nombre de gares allemandes de faible degré semble en proportion plus faible que pour le réseau français. Cela signifie qu'il y a moins de gares "isolées" en Allemagne et témoigne donc d'une meilleure interconnexion des gares par rapport à la France.

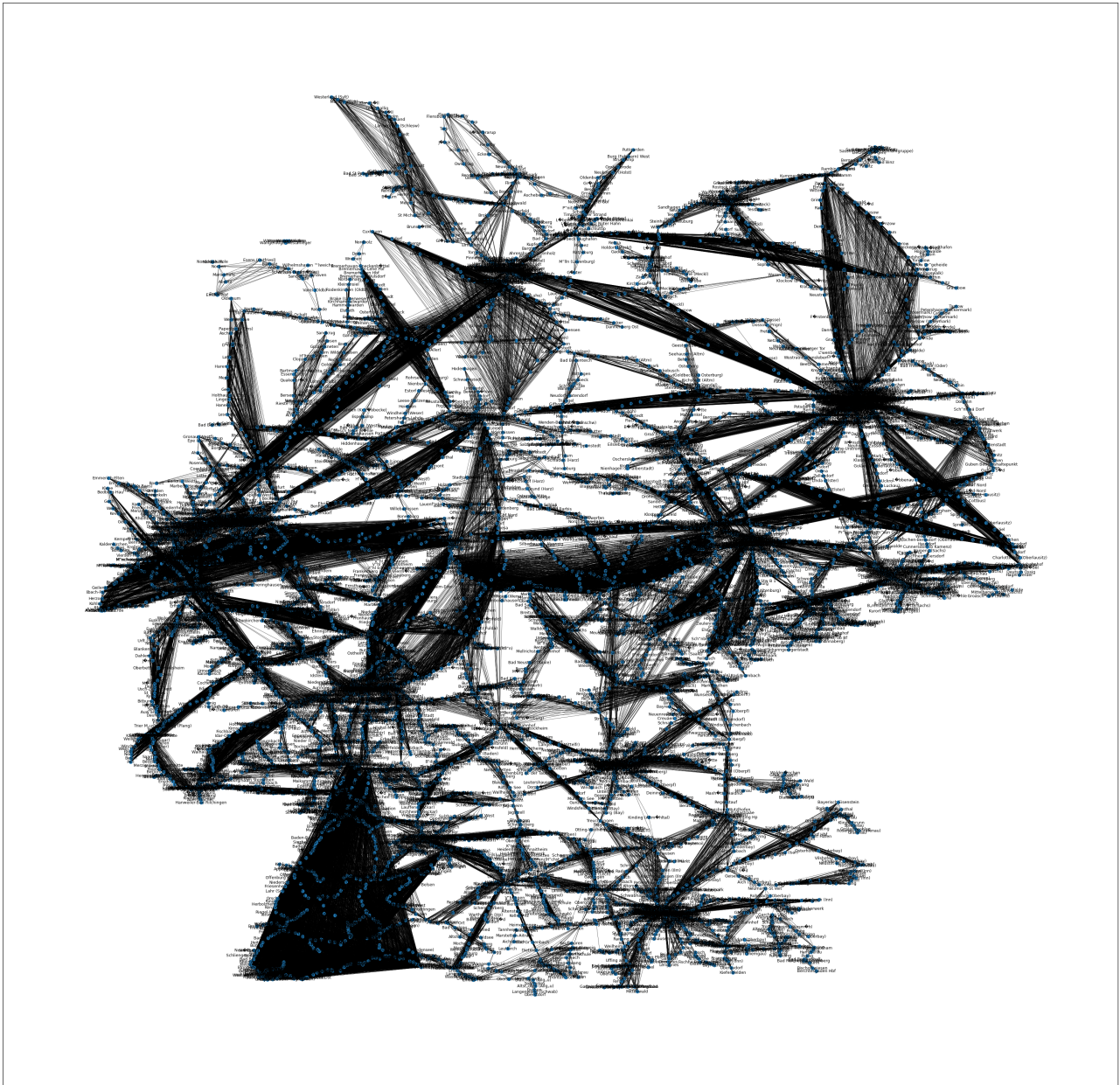


FIGURE 16 – Représentation "small-world" du réseau ferré allemand

	Graphe linéaire	Plus grande composante connexe
Nœuds	6546	5680
Arêtes	70271	67246
Nombre de composantes connexes	194	1
Degré moyen	21,47	23,68
Densité	0,00328	0,00417
Clustering moyen	0,9316	0,9428
Diamètre		17 trajets
Plus court chemin moyen		7,06 trajets

TABLE 4 – Étude préliminaire du réseau ferré allemand représenté de façon "small-world" et de sa plus grand composante connexe.

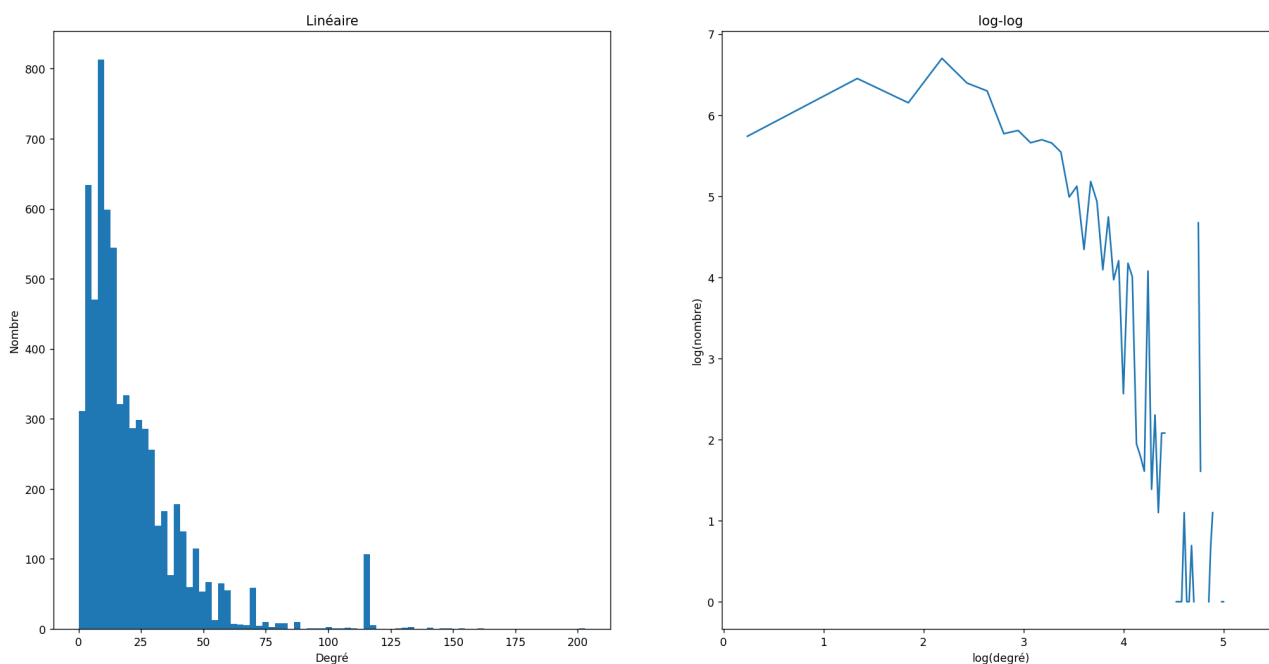


FIGURE 17 – Répartition des degrés des nœuds du réseau "small-world". La distribution est à gauche, sa version log-log à droite.

4.3.2 Étude des communautés

Il peut être intéressant de voir si, à travers l'étude des communautés de ce graphe, il est possible de retrouver la structure fédérale de l'Allemagne avec ses régions. Dans un premier temps, l'algorithme de Louvain a été utilisé sur la plus grande composante connexe. Selon le paramètre de résolution, ce dernier trouvait plus ou moins de communautés, avec un minimum de 36 communautés pour un paramètre de résolution de 0,74. Cela reste bien supérieur aux 16 régions administratives de ce pays. Les résultats de l'algorithme de Louvain sont donnés en Fig.18. Contrairement à la France, il est plus difficile de retrouver les régions administratives à partir du réseau ferré. Ces dernières sont beaucoup plus reliées entre elles. Les lignes maillent mieux le territoire.

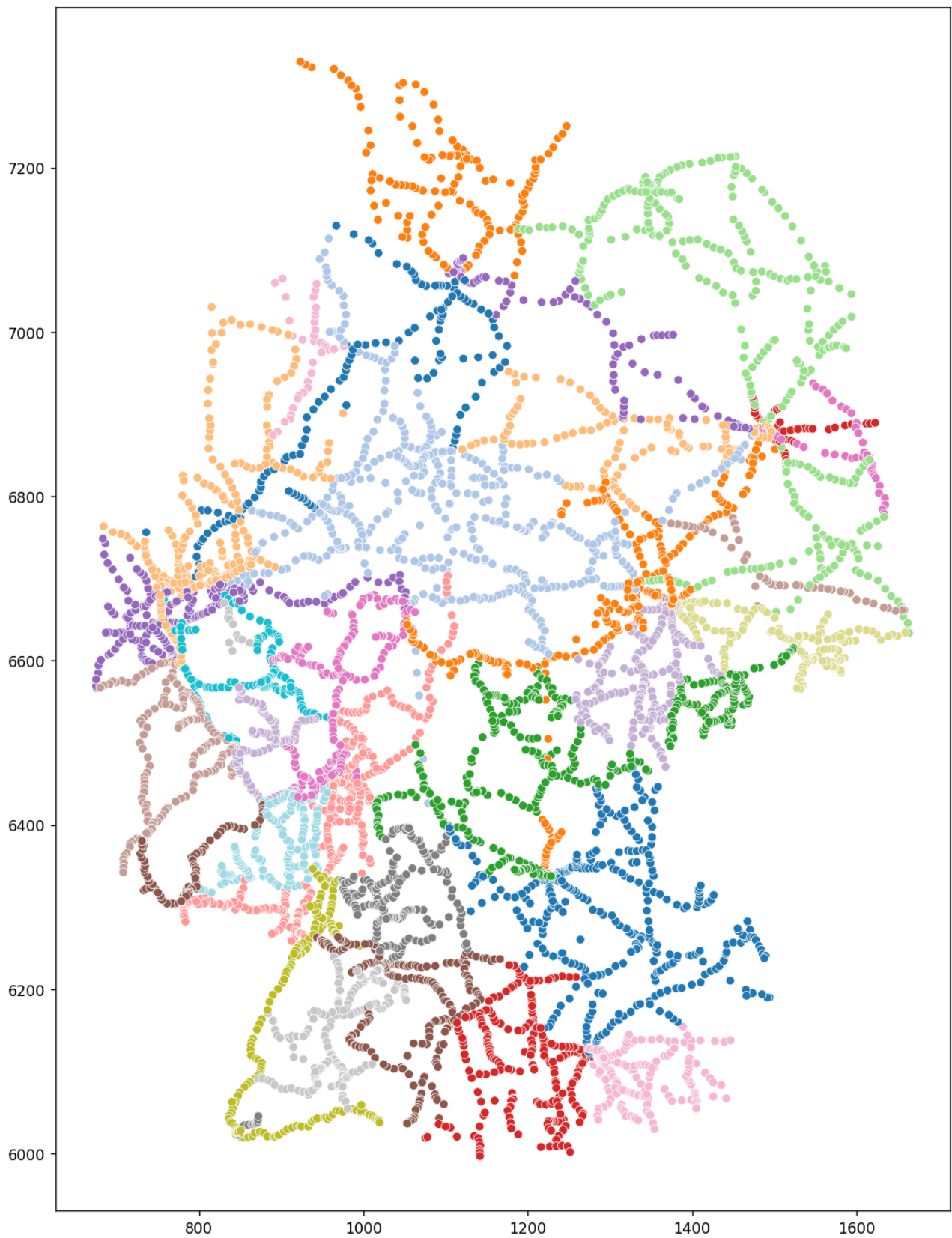


FIGURE 18 – Recherche de communauté sur la plus grande composante connexe du graphe small world pour un paramètre de résolution de 0,74.

4.3.3 Étude de centralités

De la même manière que pour le graphe "linéaire", les gares avec les plus grands degrés et PageRank sont présentés tableau 5. Les mêmes remarques que précédemment peuvent être faites : les gares centrales (Hbf pour Hauptbahnhof signifie "Gare principale") des grandes villes occupent ce classement. Mais l'absence de certaines villes comme Berlin peut sembler étrange. Cela peut s'expliquer si la ville répartit son trafic ferroviaire sur de nombreuses gares.

Graphe "small-world"			
Top 10 Gares Degrés		Top 10 Gares Pagerank	
Halle (Saale) Hbf	203	Nürnberg Hbf	0,000910
Mannheim Hbf	162	Halle (Saale) Hbf	0,000808
Düsseldorf Hbf	154	Hannover Hbf	0,000718
Karlsruhe Hbf	149	Hamm (Westf) Pbf	0,000676
Gieáen	145	Frankfurt (Main) Hbf	0,000670
Singen (Hohentwiel)	141	Düsseldorf Hbf	0,000646
Kassel Hbf	140	Münster (Westf) Hbf	0,000586
Bruchsal	133	Rosenheim	0,000582
Basel Bad Bf	132	Gieáen	0,000568
Frankfurt (Main) Hbf	132	Mannheim Hbf	0,000560

TABLE 5 – Liste des 10 gares de plus hauts degrés et respectivement de plus hauts PageRank.

5 Conclusion

Si vous lisez cette phrase, bravo à vous, nous avons bientôt fini notre escapade dans le monde fascinant des TER .

Plus sérieusement, ce travail a permis d'étudier dans le détail le réseau de TER français et de confronter nos intuitions (structure influencée par les régions, influences de la population et de la fréquentation sur la centralité, ...) aux caractéristiques du graphe construit. Résultat : bien souvent, notre sens commun ne nous trompe pas ! Nous avons également amorcé une étude du réseau ferroviaire allemand en calculant ses propriétés principales mais nous n'avons pas eu le temps d'aller dans le détail. De plus, la comparaison entre les deux pays pourrait être poursuivie et améliorée au regard des éléments apportés dans la dernière partie. Un début de comparaison a pu être amorcé : nous avons pu créer deux nouveaux réseaux français équivalents des deux réseaux allemands. Nous avons manqué de temps pour finir cette étude, mais le début de ce travail est disponible en Annexe [A](#).

Par ailleurs, il serait intéressant de comparer nos graphes à des "null models" de type gravité relaxé. Par exemple, on pourrait construire des modèles aléatoires où la probabilité de relier deux nœuds dépend de la population des deux villes et de la distance entre elles via une fonction de "deterrence" ad hoc. Une autre construction pourrait utiliser le degré des nœuds en lieu et place de la population des villes, ce qui permettrait de mieux caractériser l'influence de la population sur la probabilité de faire des liens. L'utilité de ces null models est aussi de tester la robustesse des propriétés exhibées par les graphes réels : s'expliquent-elles seulement par la distance à vol d'oiseau et la population/le degré ?

Pour finir, nous pensons (espérons en tout cas) que notre travail rend justice à la richesse et à l'utilité des outils propres aux réseaux complexes. Notons cependant que contrairement à ce que nous pouvons faire en cours ou en TD, ce projet met au jour les difficultés non seulement à obtenir et nettoyer les données, mais également à les utiliser pour construire un graphe. Nous avons par exemple construit plusieurs graphes avec des propriétés très différentes en partant pourtant des mêmes bases de données. Il y a donc forcément des partis pris, et avec cela des conséquences sur les résultats et leurs interprétations. C'est sans doute une bonne leçon de morale que de se rappeler des biais introduits dans le choix des données, leur traitement, ainsi que la construction du graphe à partir de celles-ci.

Table des figures

1	Représentation schématique des fusions réalisées entre la BDD gares voyageurs, la BDD fréquentation et la BDD population INSEE. Les nombres en gras correspondent au nombres de lignes (de gares donc) qui apparaissent dans la BDD en question.	5
2	Le prix des trajets en TER donné par la SNCF contient-il le prix de l'abonnement à l'achat ? Il semble que non, vu le cas de ces deux liaisons : Lille - Santes et Lille - Don Sainghin. Ce schéma récapitule le raisonnement. À noter que le prix de l'abonnement mensuel de la région Hauts-de-France est intuité. Pour les curieux.se, le prix de l'abonnement mensuel coûte actuellement 22 euros.	6
3	Schéma récapitulatif de l'analyse de la BDD tarifs. Les flèches indiquent si le trajet est à sens unique ou à double sens (aller-retour). Le symbole euro A/R indique si le prix de l'aller est identique au prix du retour ou non. Enfin, si les prix A/R sont différents, il est précisé si les abonnements pris à l'aller et au retour sont également différents ou non.	7
4	Schéma explicatif d'un cas où la gare de départ (gare A) est sur une ligne ferrée (ligne 1), la gare d'arrivée (gare B) est sur une autre (ligne 3) et le trajet de TER emprunte une troisième ligne ferrée (ligne 2) qui ne peut pas être découverte seulement avec les données (Ligne, PK) des gares A et B.	8
5	Carte géographique des gares de TER retenues pour l'étude du réseau français. Les positions des gares ont été calculées en projection Mercator avec les coordonnées GPS.	11
6	Graphe du réseau de TER français. On rappelle que les nœuds sont les gares de TER et les liens sont les liaisons directes en TER entre ces gares. Les gares sont positionnées par leurs coordonnées GPS avec une projection Mercator.	13
7	Graphe du réseau de TER français faisant apparaître par des couleurs les communautés trouvées par la méthode de Louvain. Comme avant, les nœuds sont les gares de TER qui sont positionnées par leurs coordonnées GPS avec une projection Mercator. Les axes sont exprimés en kilomètres et ont été laissés pour l'échelle.	14
8	Étude des degrés des nœuds du graphe français. En haut à gauche, l'histogramme des degrés des nœuds du graphe. En haut à droite, la distribution des degré en semi-log. En bas à gauche, la distribution des degrés en log-log. Les droites rouges positionnées au dessus de la distribution ont des pentes de -0.75 et -3 respectivement. En bas à droite, le tableau des 20 gares de plus haut degré, rangées par ordre décroissant.	17
9	Tableau des 25 gares de betweenness les plus élevées, classées par ordre décroissant.	18
10	Représentation de la distribution des distances des nœuds. À gauche, l'histogramme linéaire. À droite, le graphe en semi-log.	19
11	Matrice de corrélations des attributs sur les nœuds.	20
12	Structure de la base de données pour le réseau de train allemand	22
13	Liste des abréviations correspondant aux différents types d'infrastructures listés dans la base de données. Dans "Nom allemand", "COMPOSITION" signifie que l'abréviation est une composition d'abréviations détaillées plus haut dans le tableau.	24

14	Carte des nœuds du réseau, les points en vert correspondent aux nœuds gardés pour l'étude et les points rouges à ceux qui ont été supprimés.	25
15	Représentation dite "linéaire" du réseau ferroviaire de l'Allemagne.	27
16	Représentation "small-world" du réseau ferré allemand	29
17	Répartition des degrés des nœuds du réseau "small-world". La distribution est à gauche, sa version log-log à droite.	30
18	Recherche de communauté sur la plus grande composante connexe du graphe small world pour un paramètre de résolution de 0,74.	32
19	Carte géographique des nœuds du réseau français basé sur la BDD de la liste de gares	38
20	Représentation linéaire du réseau ferroviaire pour la France	39
21	Représentation small-world du réseau ferroviaire pour la France	40

Liste des tableaux

1	Table de la moyenne de différentes caractéristiques des communautés créées avec la méthode Louvain	15
2	Étude préliminaire du réseau ferré allemand représenté de façon linéaire et de sa plus grand composante connexe.	26
3	Liste des 10 gares de plus hauts degrés et respectivement de plus hauts PageRank.	28
4	Étude préliminaire du réseau ferré allemand représenté de façon "small-world" et de sa plus grand composante connexe.	30
5	Liste des 10 gares de plus hauts degrés et respectivement de plus hauts PageRank.	33

A Comparaison du réseau allemand avec un autre réseau français

Un des problèmes de notre étude est que l'on ne peut pas directement comparer le réseau français au réseau allemand car ils ont été construits de différentes manières. Le réseau français est basé sur les trajets entre TER : les gares sont reliées s'il existe un trajet TER direct proposé entre ces 2 gares. Le réseau allemand est basé sur les lignes : les gares sont reliées entre elles si elles sont sur la même ligne pour le cas du réseau "small-world", ou si elles sont sur la même ligne et adjacentes pour le cas du réseau "linéaire". Mais dans le cas allemand, nous ne savons pas quels sont les trajets de train direct (sans changement de train mais avec possible changement de ligne) entre deux gares.

Pour remédier à cela, nous avons créé les équivalents des 2 réseaux allemands à partir de la base de données donnant la liste des gares. On peut tirer de cette BDD, pour chaque gare, le code ligne unique ainsi que le point kilométrique (PK). On est donc en capacité de créer 2 réseaux français de la même manière que les 2 réseaux allemands.

A.1 Construction des graphes français sur le modèle des graphes allemands

En utilisant la même méthode de construction que pour les deux graphes allemands, nous obtenons 2 réseaux français de 3 417 nœuds chacun. Une représentation des nœuds des réseaux est disponible Fig.19 et les 2 graphes sont visibles Fig.20 et 21. La représentation des nœuds nous permet d'observer les gares TGV en Île-de-France qui n'étaient pas visibles dans le réseau TER. Visuellement, on peut constater que le réseau "small-world" de la France est bien plus centralisé que le réseau allemand.

A.2 Comparaison avec le réseau allemand

Premièrement, les graphes allemands comportent beaucoup plus de nœuds que leur équivalent français : 5 680 contre 3 417 alors que le territoire français est pourtant 1,5 fois plus grand. Cela est en partie dû aux nombreuses gares S-Bahn présentes dans les grandes villes alors que le RER est seulement développé à Paris, les autres grandes villes de France préférant d'autres alternatives comme le métro ou le tram.

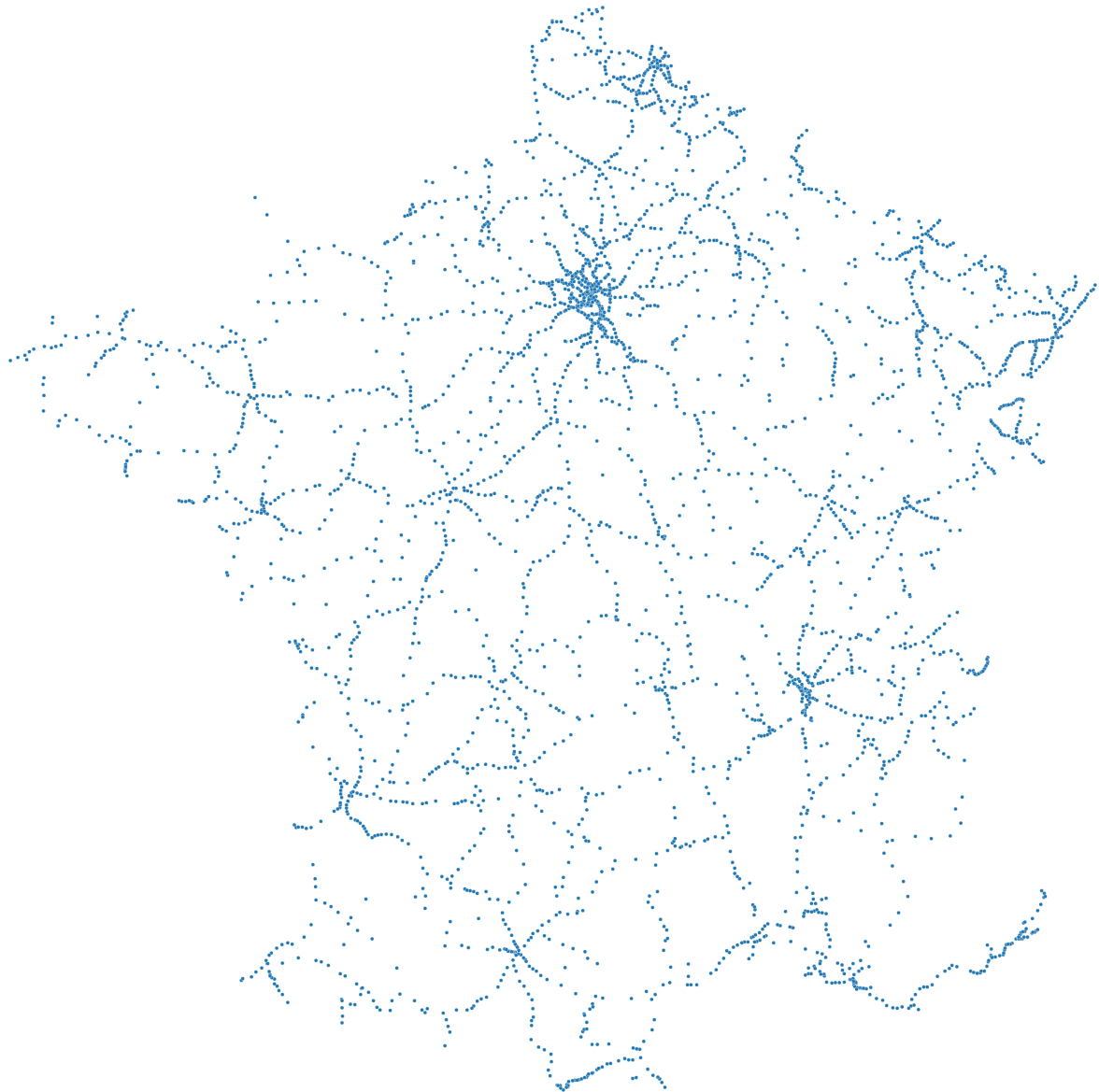


FIGURE 19 – Carte géographique des nœuds du réseau français basé sur la BDD de la liste de gares

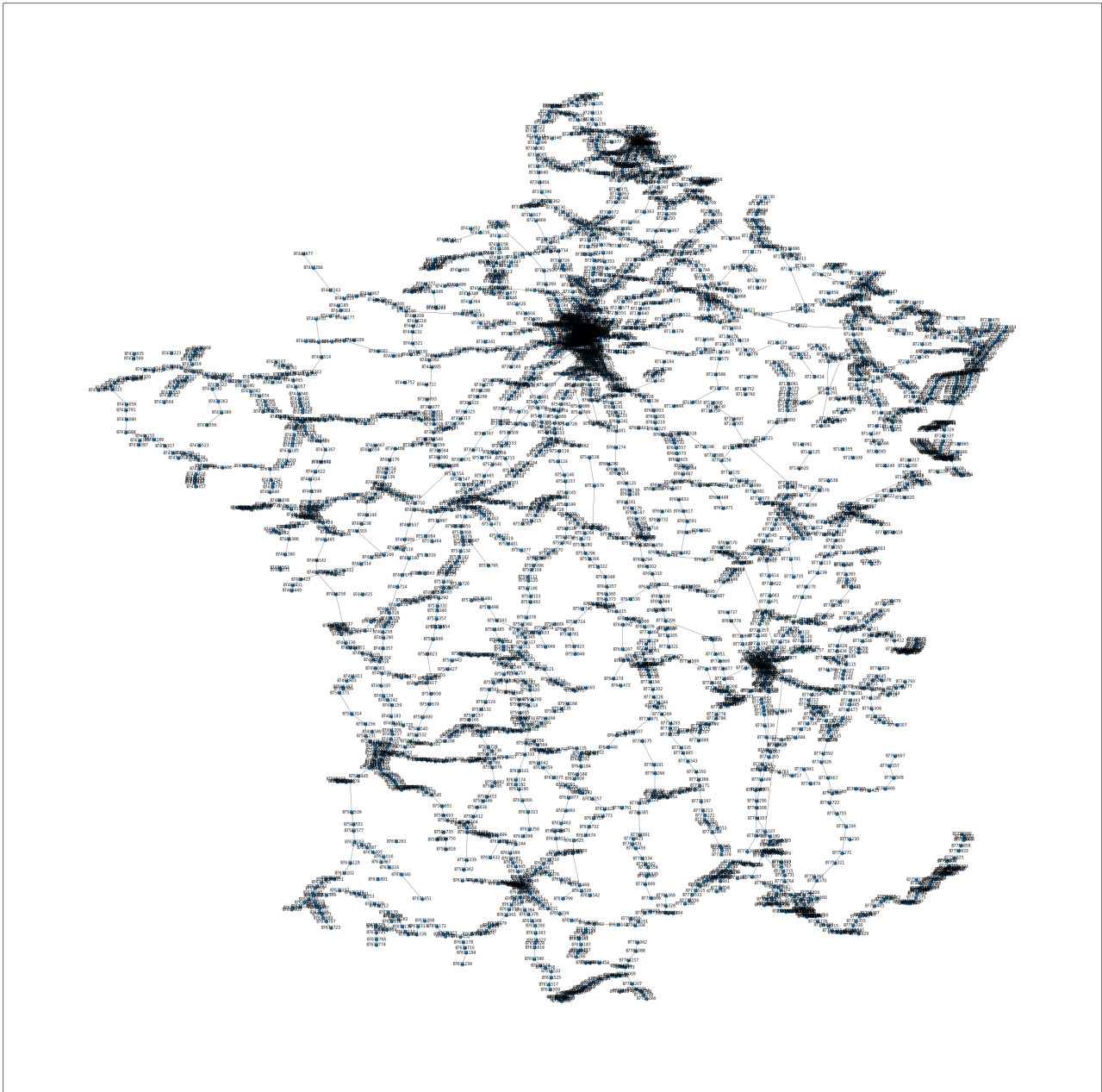


FIGURE 20 – Représentation linéaire du réseau ferroviaire pour la France

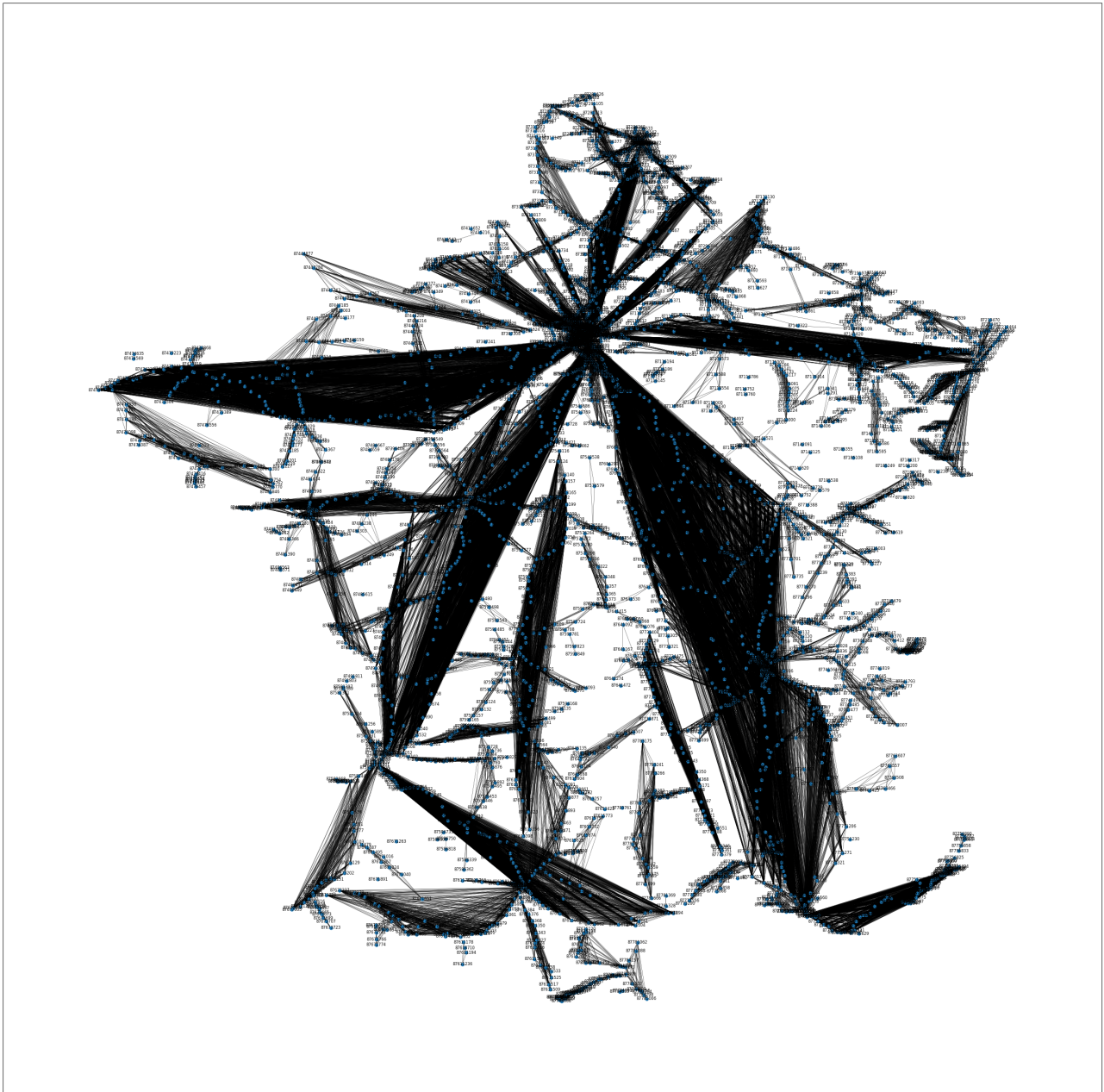


FIGURE 21 – Représentation small-world du réseau ferroviaire pour la France