Étude théorique : stations de base du réseau téléphonique français

Progression du stage

Paul MÉHAUD, Brendan SÉVELLEC

České Vysoké Učení Technické v Praze

mai - août 2024

Déroulé de la présentation

Introduction

Contexte général

Objectifs

- Déterminer si les stations de base sont en zone urbaine ou rurale ;
- Chercher les stations de bases voisines les unes des autres pour aider à déterminer si les utilisateurs sont en mouvement.

Méthodes

- Approche par la théorie des graphes ;
- Approche par le machine learning.

Données

Données

Arcep

Autorité de régulation des communications électroniques, des postes et de la distribution de la presse.

Jeu de données

Le jeu de données `2023_T4_sites_Metropole.csv` [a] représente les stations de bases au trimestre 4 de 2023 avec leur position géographique (taille : $16,7\,\mathrm{Mo}$).

a. `https://data.arcep.fr/mobile/sites/`

A retenir :

- $108\,838$ sites ;
- 29 attributs.

A quoi ressemble notre base ?

code_op	nom_op	num_site	id_site_partage	id_station_anfr	x	y	latitude	longitude	nom_reg
20801	Orange	00000001A1	nan	0802290015	687035	6985761	49,97028	2,81944	Hauts-de-France
20801	Orange	00000001B1	nan	0642290151	422853	6249263	43,28861	-0,41389	Nouvelle-Aquitaine
20801	Orange	00000001B2	nan	0332290026	416932	6422196	44,84112	-0,58333	Nouvelle-Aquitaine
20801	Orange	00000001B3	nan	0472290005	511106	6349234	44,21666	0,63556	Nouvelle-Aquitaine
20801	Orange	00000001C1	nan	0512290147	836824	6889450	49,09028	4,87333	Grand Est

nom_dep	insee_dep	nom_com	insee_com	site_2g	site_3g	site_4g	site_5g	mes_4g_trim	site_ZB
Somme	80	Curlu	80231	1	1	1	0	0	0
Pyrénées-Atlantiques	64	Jurançon	64284	1	1	1	1	0	0
Gironde	33	Bordeaux	33063	1	1	1	1	0	0
Lot-et-Garonne	47	Agen	47001	1	1	1	0	0	0
Marne	51	Sainte-Menehould	51507	1	1	1	0	0	0

site_DCC	site_strategique	site_capa_240mbps	date_ouverturecommerciale_5g	site_5g_700_m_hz	site_5g_800_m_hz
0	0	0	nan	0	0
0	0	1	2020-12-14	0	0
0	0	1	2021-02-22	0	0
0	0	1	nan	0	0
0	0	1	nan	0	0

site_5g_1800_m_hz	site_5g_2100_m_hz	site_5g_3500_m_hz
0	0	0
0	1	0
0	0	1
0	0	0
0	0	0

Table 1 – Premières valeurs de la base

ČVUT
ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Description (1/2)

Ce qui nous intéresse

1. *longitude*, *latitude* : coordonnées de chaque site ;
2. *nom_op* : nom commercial de l'opérateur ;
3. *nom_reg*, *nom_dep* et *nom_com* : nom de la région, du département et de la commune d'implantation du site ;
4. *site_xg* : équipement du site en technologie xG ($x \in \{2, \ldots, 5\}$) ;
5. *num_site* : identifiant du site issu du SI de l'opérateur.

Description (2/2)

Ce qu'il faut retenir

1. Répartition équitable du nombre de sites en fonction de l'opérateur ($\simeq 27\,000$) ;
2. $99,6\%$ des sites équipés en 4G ;
3. 6 stations en moyenne par commune.

La construction de cette base ne nous permet pas de faire de statistiques descriptives intéressantes.

From 01/07/24 to 05/07/24

Introduction to the New Dataset

- The dataset is provided by the French government and includes detailed information on radioelectrical installations with power greater than 5 watts.
- It is publicly accessible through the portal: `https://www.data.gouv.fr/fr/datasets/donnees-sur-les-installations-radioelectriques-de-plus-de-5-watts-1/`.

Components of the Dataset

- **SUP-ANTENNE**: Information about antennas, including dimensions, azimuths, and altitude.
- **SUP-BANDE**: Frequency bands used by the installations.
- **SUP-EMETTEUR**: Details about the emitters, including system names and service dates.
- **SUP-STATION**: General information about the stations, including implementation and service dates.
- **SUP-SUPPORT**: Support information including types and identifiers of the supports.

Detailed Look at SUP-ANTENNE

Fields Included

- STA-NM-ANFR: Station identifier.
- AER-ID: Antenna identifier.
- TAE-ID: Type of antenna.
- AER-NB-DIMENSION: Antenna dimension.
- AER-FG-RAYON: Ray type.
- AER-NB-AZIMUT: Azimuth angle.
- AER-NB-ALT-BAS: Base altitude.
- SUP-ID: Support identifier.

Potential Usefulness

- Helps in understanding the physical characteristics of the antennas.
- Essential for modeling and simulation purposes, particularly in determining the coverage areas.

Detailed Look at SUP-BANDE

Fields Included

- STA-NM-ANFR: Station identifier.
- BAN-ID: Band identifier.
- EMR-ID: Emitter identifier.
- BAN-NB-F-DEB: Start frequency.
- BAN-NB-F-FIN: End frequency.
- BAN-FG-UNITE: Frequency unit.

Potential Usefulness

- Provides information on the frequency spectrum used by the installations.
- Crucial for interference analysis and frequency planning.

Detailed Look at SUP-EMETTEUR

Fields Included

- EMR-ID: Emitter identifier.
- EMR-LB-SYSTEME: System label.
- STA-NM-ANFR: Station identifier.
- AER-ID: Antenna identifier.
- EMR-DT-SERVICE: Service date.

Potential Usefulness

- Provides operational details about the emitters.
- Useful for tracking the deployment and operational status of different emitters.

Detailed Look at SUP-STATION

Fields Included

- STA-NM-ANFR: Station identifier.
- ADM-ID: Administrator identifier.
- DEM-NM-COMSIS: Commune code.
- DTE-IMPLANTATION: Installation date.
- DTE-MODIF: Modification date.
- DTE-EN-SERVICE: Service date.

Potential Usefulness

- Offers a comprehensive view of the station's administrative details.
- Important for understanding the history and modifications of each station.

Conclusion

The most useful file for our research is the **SUP-ANTENNE** file. This file contains critical information about each antenna, including:

Useful information

- Station identifier (STA-NM-ANFR)
- Antenna identifier (AER-ID)
- Type of antenna (TAE-ID)
- Antenna dimensions (AER-NB-DIMENSION)
- Azimuth angles (AER-NB-AZIMUT)
- Base altitude (AER-NB-ALT-BAS)

Methodology

- From this data, we can extract information about the azimuth angles of antennas.
- This allows us to accurately calculate adjacent cells (neighbors) for each station.

Introduction to New Methodology

- Given the valuable data from the antenne dataset, now we can take a look on a new methodology for calculating base station coverage.
- This methodology leverages geometric techniques and detailed information about antenna directions to improve accuracy.
- The approach starts with partitioning the area using Voronoi tessellation.

Voronoi Tessellation

Voronoi Tessellation

- The method starts by partitioning the area using Voronoi tessellation.
- Each base station is a node, and its Voronoi cell contains all points closer to it than to any other base station.
- The computational complexity is improved using Fortune's algorithm.

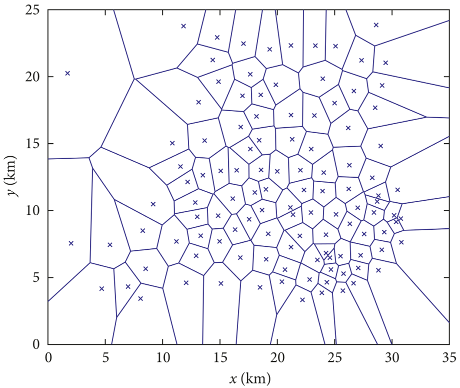

Figure 1 – Voronoi Diagram for Cellular Network Sites

Cell Border Definition

Cell Border Definition

- After generating Voronoi cells, the next step is to define cell borders based on antenna directions.
- The border is defined by intersecting the Voronoi edges with the antenna beamwidth.
- This process accounts for the directional nature of antennas.

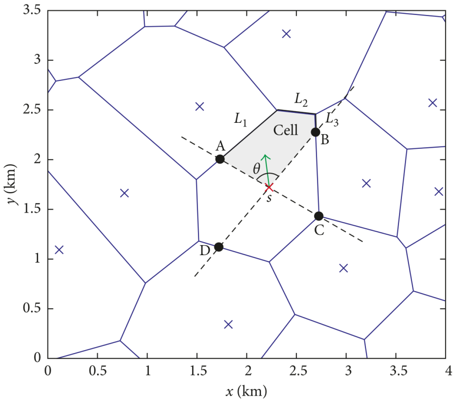

Figure 2 – Defining Cell Borders Using Antenna Directions

Average Distance to Cell Border

Average Distance to Cell Border

- The final step is to compute the average distance from the base station to the cell border.
- This involves calculating the average distance to each segment of the cell border.
- The formula used is:

$$\mathrm{dist}(P, L) = \frac{1}{l_L} \int_0^{l_L} \sqrt{(x' - x'_P)^2 + y'^2_P}\, dx' \tag{1}$$

- This ensures a weighted average distance, dominated by longer segments.

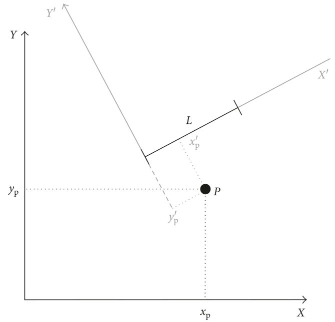

Figure 3 – Point Translation for Distance Calculation

Cell Range Calculation

- The cell range (CR) for each cell is computed using:

$$\mathrm{CR}_{\mathrm{cell}}(c) = \frac{\sum_{i=1}^{n_{\mathrm{seg}}(c)} \mathrm{dist}(s, L_i) \cdot l_{L_i}}{\sum_{i=1}^{n_{\mathrm{seg}}(c)} l_{L_i}} \tag{2}$$

- This method accounts for local variations and provides a more accurate estimate than site-level averages.

Conclusion

- The geometric method provides a detailed and computationally efficient way to estimate cell ranges.
- It addresses limitations of traditional methods by considering individual cell characteristics.
- This approach can be integrated into network planning and optimization processes.

From 01/07/24 to 05/07/24

How to describe departments?

Methodology

We wanted to have an indicator about every departement to find out a more precise approximation of coverage cells' diameters. Here is what we did :

Mean distance to neighbours

The basic idea is to compute the sum of distances between every base station in a department, divided by the number of base stations. Let n be the number of base stations in a departement, we will name each base station in $\{0, \ldots, n\}$. Let $d(\cdot, \cdot)$ be the function that calculates the distance bewteen two base stations.
So, we obtain this formula (let γ be the number we look for):

$$\gamma = \frac{\sum_{i=1}^{n} \sum_{j \neq i} d(i, j)}{n}$$

The idea that motivated the implementation of this method is that if a departement is mainly constituated of mountains, the mean distance between base stations is shorter. We will see that this distance as other influential parameters.

Fine tuning of this new method

We now have to think about the things that can influence our results, to normalize them.

Things that influence γ

We have identified several influencing factors such as:

- The amout of population;
- The topography of the departement (is it flat or not?);
- The highways or railways;
- The size of the departement.

However, we will, for now only use the number of inhabitants and the size of the department.

Normalization of the data

Basically, here is the formula we use :

$$\gamma_{\text{norm}} = \frac{\gamma}{\text{number of inhabitants} \times \text{surface area}}$$

Conclusion - results

nom_dep	city	countryside	total	normalized
Gironde	5304.87	19134.2	26241.4	19.0708
Rhône	5253.74	5258.18	12435.4	22.9532
Bouches-du-Rhône	12387.8	7876.89	22356.1	23.0575
Loire-Atlantique	6210.51	12295.7	19599.4	23.7876
Isère	5022.42	16067.9	21875	25.1173
		...		
Meuse	700.178	10018	10831.2	90.7332
Hauts-de-Seine	2429.9	-1	2429.9	91.0103
Hautes-Alpes	845.008	5664.28	6674.97	91.1299
Haute-Corse	376.022	5495.8	6415.91	92.9077
Haute-Marne	1301.65	9773.92	11333	97.837
Corse-du-Sud	759.224	4241.56	5271.69	110.742
Lozère	54.238	5703.01	5835.86	146.682
Paris	3437.78	-1	3437.78	151.144

Table 2 – Results

Conclusion

We cannot use this results to quantify the topography of each department, but we could still find some use to approximate the coverage cells' size.

From 01/07/24 to 05/07/24

New road detection method

Building city connection graph (1/4)

City detection

A clustering method (here DBScan) is used only on the base station detected as a city center by the 3NN method. Thus, these base stations are now separated in different cities.

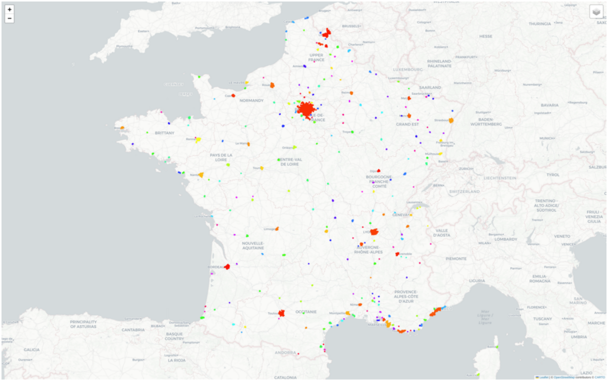

Figure 4 – city clustering

Building city connection graph (2/4)

Center computation

The center of each cluster is computed.

city separation according to size

The cities that contain more than a certain number of stations are separated from the ones that contain less.

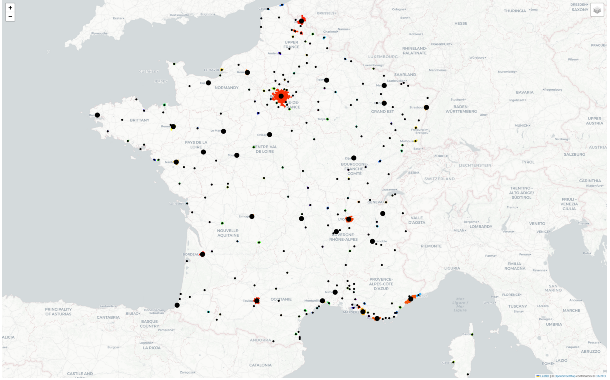

Figure 5 – cities centers and size classification

Building city connection graph (3/4)

Delaunay triangulation

The Delaunay traingulation is then applied to the big cities to create a graph.

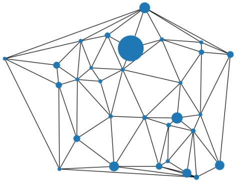

Figure 6 – Big cities connection graph unfiltered

Building city connection graph (4/4)

Graph filtration

The angle and distance criteria are then applied to filter the connections between the cities.

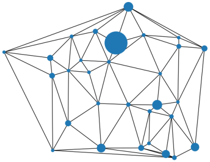

Figure 7 – Unfiltered

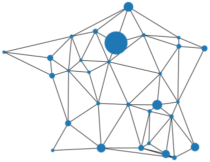

Figure 8 – Distance filtered

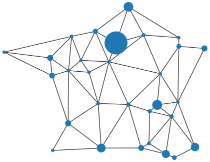

Figure 9 – Distance and angle filtered

From 08/07/24 to 12/07/24

From 08/07/24 to 12/07/24

further advancement on road detection

Cities and big cities connection graph

Figure 10 – cities and big cities connection graph

Little cities linking method

Steps

- For each edge of the connection graph ;
- Detect all little cities close enough to this edge ;
- Create multiple edges to link all these cities together by a path with the same start and ending city than the big edge ;

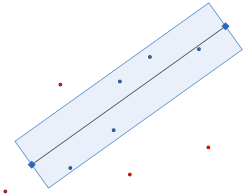

Figure 11 – Little cities selection

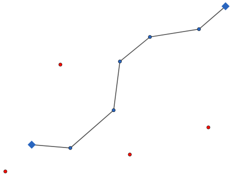

Figure 12 – Cities linkage

New graph after little city linkage

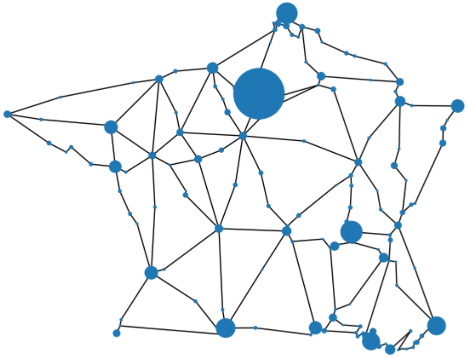

Figure 13 – Final graph (width parameter = 0.2)

New graph on a map

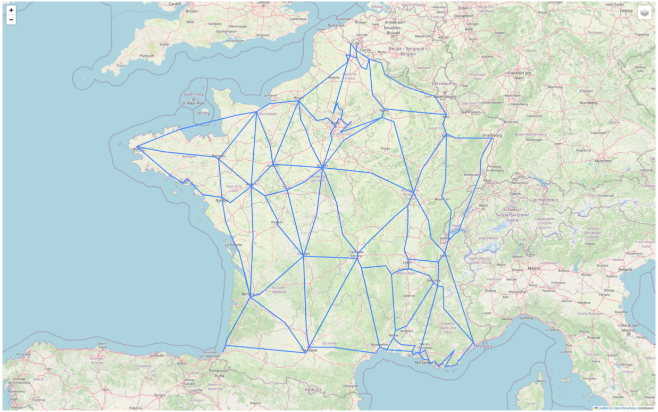

Figure 14 – Final graph on a map (width parameter $= 0.2$)

Reminder : pre-road detection thanks to the clustering methods

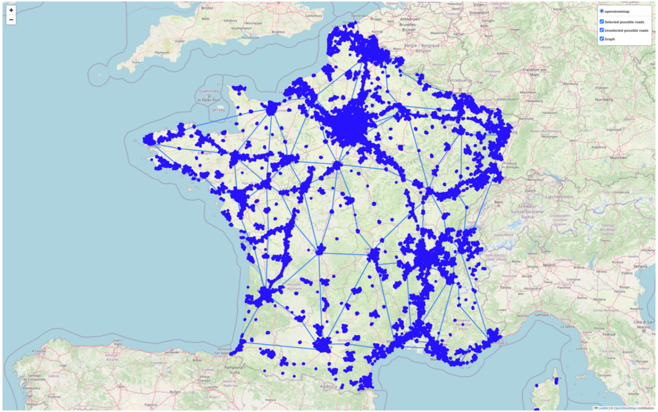

Figure 15 – Road pre-detection and graph

Final selected road base stations

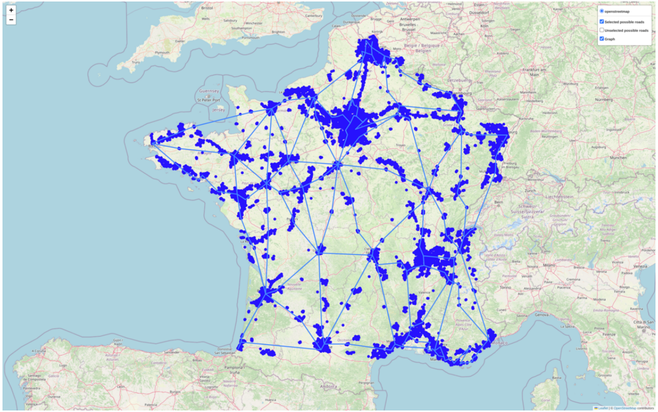

Figure 16 – Nodes detected as roads after the method

From 08/07/24 to 12/07/24

Further advancement on Coverage Calculation

Creation of New Database

- We have created a new database containing detailed information about base stations and their antennas in Normandy, France.

Database Fields

- Station ID (`id_station_anfr`)
- Latitude and Longitude
- Department and Commune Names
- Presence of 2G, 3G, 4G, and 5G technologies
- Antenna Dimensions and Azimuths

Program for Visualization

- We implemented a Python program to visualize the base stations on a map and analyze their coverage using Voronoi diagrams.

Key Steps in the Program

- Load and clean the data.
- Add markers and pop-ups for each base station.
- Calculate and visualize Voronoi cells and antenna sectors.

Visualization Results

- The map is centered on Normandy and shows base stations as markers.
- Each marker provides detailed information about the station and its antennas.
- Voronoi tessellation is used to visualize the coverage area of each base station.
- Antenna azimuths are displayed to indicate coverage directions.

Example Visualization

- The green lines represent Voronoi boundaries.
- The markers indicate base station locations with detailed popup information.
- Direction lines show the azimuths of the antennas.

From 15/07/24 to 18/07/24

From 15/07/24 to 18/07/24

further advancement on road detection

ČVUT
ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Little advancement

City name detection

For each city detected, we find the closest base station to the center of the said city. We then declare that the name of the city is the value contained in the field "nom_com" of the base station.

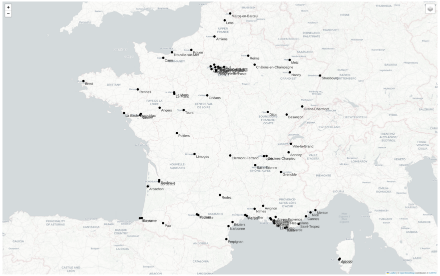

Figure 17 – Cities centers with name

New method (1/2)

Edges weight

For each couple of cities, we compute a weight using this formula :

$$w_{\text{city1, city2}} = \frac{\min(\text{size(city1)}, \text{size(city2)})}{\text{dist(city1, city2)}}$$

With the size of a city refering to the number of base stations detected inside that city.
We the keep all the edges whose weight is superior to a certain value.

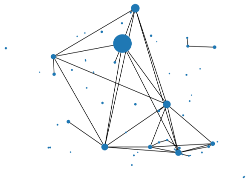

Figure 18 – weight filtration, cap value $= 0.1$

New method (2/2)

Edges weight

We then apply the angle criterion to the resulting graph

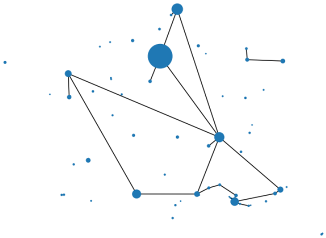

Figure 19 – weight and angle filtration, cap value $= 0.1$, angle value $= 15$ degrees

From 15/07/24 to 18/07/24

Method for Finding Neighboring Base Stations Using Voronoi Diagrams

Voronoi Diagrams for Neighboring Base Stations

- The method uses Voronoi diagrams to determine the neighborhood of base stations.
- If the cells share a common boundary, the base stations are considered neighbors.

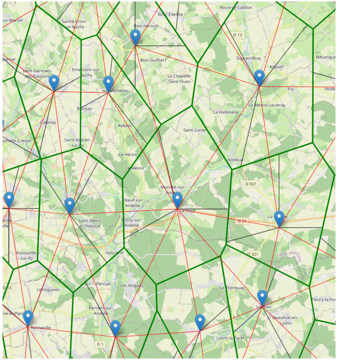

Figure 20 – Voronoi Diagram Showing Neighboring Base Stations

Analysis of Results

Explanation of inaccuracy

- The method considers only geometric proximity without accounting for actual network topology or physical barriers.
- It's analogous to using Delaunay triangulation and assuming all neighboring stations in the triangulation are true neighbors.

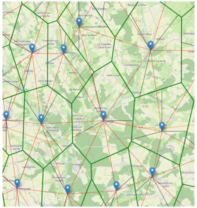

Figure 21 – Voronoi Diagram Showing Neighboring Base Stations

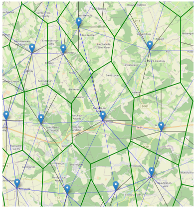

Figure 22 – Delaunay Triangulation of Base Stations

From 15/07/24 to 18/07/24

Introduction to the New Method for Determining Neighboring Base Stations

New Method for Determining Neighboring Base Stations

- The method uses Delaunay triangulation to identify potential neighboring base stations.
- Verification of neighboring status is done by checking antenna coverage angles.
- Stations are considered neighbors if their antenna directions align within their coverage angles.

Determining Antenna Coverage Angles

- Calculate the coverage angle for each antenna by dividing 360 degrees by the number of unique azimuths.
- This gives the angular sector covered by each antenna on the station.

Validating Neighboring Status

- For each pair of stations connected by an edge, check the direction of the connecting line.
- Compare this direction with the antenna coverage angles of both stations.
- If the direction falls within the coverage angles, the stations are real neighbors.

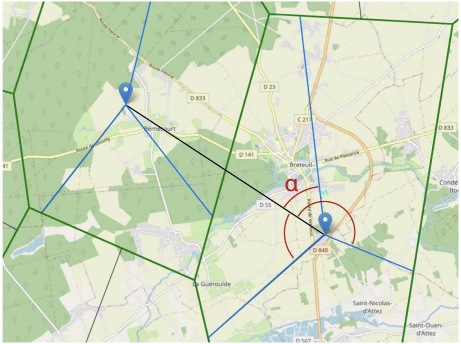

Figure 23 – Validation of Neighboring Status

Detailed Steps with Formulas

- Calculate Direction of the Connecting Line:

$$\theta = \arctan 2(y_2 - y_1, x_2 - x_1)$$

  where (x_1, y_1) and (x_2, y_2) are the coordinates of the two base stations.
- Calculate the Antenna Coverage Angles:

$$\text{Coverage Start Angle} = A - \frac{B}{2}$$

$$\text{Coverage End Angle} = A + \frac{B}{2}$$

  For an antenna with azimuth A and beamwidth B.
- Validate if the Connecting Line Falls Within Coverage Angles: Normalize angles to be within $[0^\circ, 360^\circ]$. Check if the direction θ falls within $[A - \frac{B}{2}, A + \frac{B}{2}]$.
- Bi-Directional Check: Repeat the validation from both base stations' perspectives. If $\Delta\theta$ for both stations is within the beamwidth, they are considered true neighbors.

Updating the Dataset

- Add validated neighboring stations to the dataset.
- Update the data frame with pairs of stations identified as real neighbors.

Result

- Improved accuracy in identifying neighboring base stations.
- More reliable data for network and coverage analysis.

Summary and Conclusion

- The new method combines geometric triangulation with antenna direction validation.
- This approach enhances the accuracy of identifying neighboring base stations.
- This method can be further refined and applied to past methods as an additional criterion.

From 22/07/24 to 26/07/24

From 22/07/24 to 26/07/24

Further advancement on road detection

Adjustements in the method

Stop using the angle criterion

After analysing last week's results, I decided to stop using the angle criterion because it doesn't delete the right edges. Therefore, I fine-tuned the weight calculation method, so that it doesn't need it anymore

New edge weight calculation

For each couple of cities, we compute a weight using this formula :

$$w_{\text{city1, city2}} = \frac{(\min(\text{size(city1)}, \text{size(city2)}))^{1.3}}{(\text{dist(city1, city2)})^{1.2}}$$

With the size of a city refering to the number of base stations detected inside that city.

Results

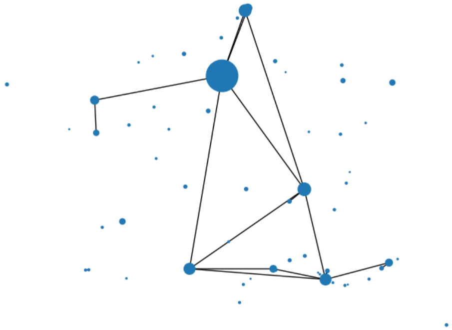

Figure 24 – weight filtration with new calculation, cap value $= 0.2$

Figure 25 – TGV lines covered in 4G by Orange

From 22/07/24 to 26/07/24

State of the art of our different neighbouring methods

ČVUT
ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Delaunay and simple criteria

Principle of the method

At first we will apply a Delaunay triangulation, to have a potential neighbours graph, and, afterwards, we apply the simple criteria : distance, angle and quadrant, to filter the bad neighbouring connexions.

Criteria

- Distance: we filter every edge longer than $15\,\mathrm{km}$;
- Angle : we filter every edge separated from the nearest one by an angle more narrow than $20°$;
- Quadrant : we filter according to the simple quadrant criterion.

Delaunay and simple criteria - pros and cons

Pros

- Every edge is treated equally, wether it in countryside or in a city;
- This method is really simple to put in place.

Cons

- Every edge is treated equally, wether it in countryside or in a city. This leads to non precise results.

Delaunay and enhanced criteria

Principle of the method

At first we will apply a Delaunay triangulation, to have a potential neighbours graph, and, afterwards, we apply the enhanced criteria : distance, angle and quadrant, to filter the bad neighbouring connexions. Same principle as before but we will take in account if a city is in or not in a city.

3NN cityness classification method

To classify each base station we need a method. Here, we will take the mean distance to the 3 nearest neighbours. Let's call this value γ, in km. So, we have 4 different categories. For each category, we will apply a different distance and angle criterion :

- $\gamma \in]0, 1]$: city center (max_distance $= 2\,\mathrm{km}$, min_angle $= 5°$);
- $\gamma \in]1, 2]$: urban area (max_distance $= 5\,\mathrm{km}$, min_angle $= 15°$);
- $\gamma \in]2, 4]$: extra-urban area (max_distance $= 10\,\mathrm{km}$, min_angle $= 25°$);
- $\gamma \in]4, \infty[$: courntryside (max_distance $= 15\,\mathrm{km}$, min_angle $= 30°$).

And then, of course, we apply the enhanced quadrant criterion.

Delaunay and enhanced criteria - pros and cons

Pros

- We take in account the “cityness” of each base station to compute its neighbours.

Cons

- We have no method to be sure that we choose the correct distances and angles;
- The filtering does not seems good sometimes. On one hand we filter too much and on the other hand, we filter too less.

Those problems could come from the Delaunay triangulation, which is maybe not the base for this use case.

Gabriel graph and enhanced criteria

Principle of the method

At first we will create a potential neighbours graph : the Gabriel graph and, afterwards, we apply the enhanced criteria : distance, angle and quadrant, to filter the bad neighbouring connexions. Same principle as before but we apply a pre-filtering with the Gabriel graph.

Gabriel graph [1]

A Gabriel graph is a subgraph of a Delaunay triangulation. Formally, it is the graph G with vertex set S in which any two distinct points $p \in S$ and $q \in S$ are adjacent precisely when the closed disc having pq as a diameter contains no other points.

And then, of course, we apply the enhanced criteria, as listed before.

1. https://en.wikipedia.org/wiki/Gabriel_graph

Gabriel graph and enhanced criteria - pros and cons

Pros

- We add another level of filtering.

Cons

- We add another level of complexity.

This is not perfect, maybe we can get rid of the Delaunay triangulation.

k-NN and enhanced criteria

Principle of the method

At first we will create a potential neighbours graph with the k-NN method and, afterwards, we apply the enhanced criteria : distance, angle and quadrant, to filter the bad neighbouring connexions.

k-NN [2]

To create this graph we will, for each base station, connect it to its k nearest neighbours.

And then, of course, we apply the enhanced criteria, as listed before.

2. `https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.NearestNeighbors.html`

Gabriel graph and enhanced criteria - pros and cons

Pros

- We get rid of the Delaunay traingulation, so we could connect more base stations.

Cons

- A lot of edges to filter;
- The filtering methods is maybe not adapted.

This is not perfect, but got rid of the Delaunay triangulation.