Using Data Mining for Examining the Coverage of Cellular Infrastructure Companies in Israel

1st Etay Iluz
department of Software and Information Systems Engineering
Ben-Gurion University of the Negev
Be'er Sheva, Israel
etayil@post.bgu.ac.il

***Abstract*—The ability to accurately assess the coverage of cellular companies is essential for ensuring reliable service and customer satisfaction. In this study, we employ data mining techniques in order to examine the coverage of cellular infrastructure companies in Israel. Our approach involves analyzing a comprehensive datasets of cellular base stations location, demographic data, bus stations location data and digital geographic datasets to locate areas with an insufficient cellular coverage. We present the results of our analysis, highlighting regions that experience poor connectivity and offering the Ministry of Communications an independent tool to supervise the infrastructure companies in Israel.**

I. Introduction

Cellular communication infrastructures are essential for maintaining a normal lifestyle, and for the development of the industry in Israel. The deployment of advanced cellular infrastructures allows promoting the development of measures in various fields such as health, transportation and agriculture. The cellular market is regulated through licenses granted by the Ministry of Communications to cellular companies by virtue of the Telecommunications Law. According to the licenses, the cellular companies are obligated to establish cellular transmission facilities and cover areas through networks in order to provide the public a proper service. Cellular communication technology is classified according to generations (generation 2 to 5) and any significant technological innovation is considered a new generation. The transition between generations enables the development of new services and an improvement in the quality of existing services. In July 2024, the State Comptroller's Office published a report [1] concerning the cellular communication infrastructure in Israel. Among the topics examined is the obligations of the cellular companies and the supervision of their activities. According to the licenses of the Ministry of Communications, the cellular infrastructure companies must comply with the obligation to cover with 4th generation technology the area where 99% of the population lives, covering 95% of the area of each settlement and covering 83% - 95% of the roads. The findings of the report indicate that the Ministry of Communications does not require the cellular infrastructure companies to report the extent of the actual cellular coverage and it is content with reports based on the forecasts of the cellular infrastructure companies, which do not reflect the actual cellular reception data. The companies reported compliance with the coverage obligations in 98% - 100% of the settlements. However, according to the data of the auditor's report, considerable gaps were discovered between these reports to the Ministry of Communications and the findings of the audit. Moreover, the State Comptroller's report determines that the Ministry of Communications lacks an independent effective tool to determine the actual cellular coverage in Israel and the quality of the cellular network. In fact, the Ministry of Communications relies solely on the reports provided to him by the companies themselves. In Israel, there are three companies that were licensed to establish cellular base stations: Pelephone, Cellcom and PHI (Partner and Hot mobile).

The purpose of this study is to offer to the Ministry of Communications a supporting tool that will help it independently examine the cellular coverage of the cellular companies. We do that by analyzing rich datasets comprising of cellular base stations locations, population concentrations in statistical areas, geographic information and bus stations location, in order to identify areas with poor coverage.

This paper is structured as follows: We begin by reviewing the relevant literature on cellular network coverage assessment and the application of data mining in telecommunications. Next, we describe the methodology used in our analysis, including data collection, preprocessing, and mining techniques. We then present the results of our study, highlighting key findings and finally we discuss the limitations of our tool. Through this work, we aim to demonstrate the potential of data-driven approaches for enhancing cellular network coverage and ultimately improving the quality of telecommunications services in Israel.

II. Related work

A. Cellular Network Coverage Assessment

Traditional methods for evaluating cellular network coverage often rely on drive tests, where vehicles equipped with signal measurement tools traverse predefined routes to collect data on signal strength and quality[2]. While effective, drive tests are resource-intensive and provide limited spatial coverage. In order to overcome these limitations, researchers have explored the use of crowdsourced data from mobile devices, which can offer a more comprehensive and real-time view of network performance[3]. Studies have demonstrated that crowdsourced data can accurately reflect user experience and identify areas with poor coverage, enabling more targeted network optimization efforts[4].

B. Geographic Data Mining

Geographic data mining and knowledge discovery (GDK) involve the extraction of spatial patterns and knowledge from large geospatial datasets. GDK techniques are particularly relevant for assessment of cellular network coverage, as they enable one to integrate geographic information with network performance data. Spatial clustering, spatial autocorrelation, and hotspot analysis are some of the GDK methods used to identify areas with poor coverage and point out the potential failures of network performance[5].

III. Methods

In this section, we will explain in detail our analytical approach.

First, we collected three datasets from the Data Gov website - The government databases. The first one contains details of every cellular base station in Israel. The second one contains demographic information concerning the spread of population in Israel to statistical areas (in 2024). The last one specifies the location of every bus station in Israel. At the next stage we removed irrelevant data: inactive LTE base stations and old technology LTE base stations (2G). We also merged the demographic data and the geographical data. We have also converted the coordination system of the LTE base stations from CRS to longitude and latitude.

Israel is divided into about 3,000 statistical regions, which are small geographical units within settlements of over than 10,000 people. This division makes it possible to better reflect the uniqueness of each small unit within these settlements. A statistical area usually contains between 3,000 to 5,000 people. The division of Israel and Jerusalem to statistical areas is demonstrated in Figure 1.

We have conducted two analysis in order to examine two types of coverage: Is cellular service per population is sufficient and is the majority of geographical areas in Israel enjoy full cellular coverage.

A. Population Analysis

The goal is to determine the average number of people served by each LTE base station. This involves:

1) Calculating the area that each LTE base station covers.
2) Calculating the population within the coverage area.
3) Dividing the population among overlapping coverage areas.

Each LTE base station has a potential coverage area represented by a buffer zone. The buffer zone is a circular area around the base station with a specified radius, typically representing the range of the base station.

$$\text{Buffer}(B_i, r) = \{x \in \mathbb{R}^2 : \|x - B_i\| \le r\} \tag{1}$$

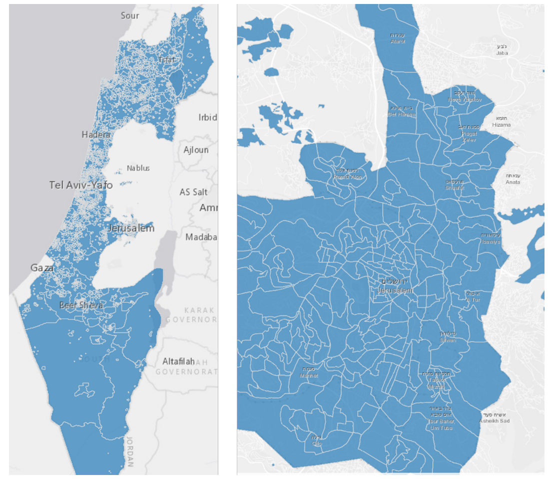

Fig. 1. Left: Statistical areas in Israel, Right: Statistical areas in Jerusalem

Where B_i is the location of the i-th base station and r is the buffer radius. In figure 2, we exemplify the buffer zones of every LTE base station in division to companies.

The population data is divided to various statistical areas. Each statistical area has an associated population amount. The task is to find out which statistical areas fall within the buffer zones of the LTE base stations.

For each statistical area A_j with a population P_j, the overlapping base stations are identified by checking whether the centroid of A_j falls within the buffer of any base station.

$$\text{Covered\_Base\_Stations}(A_j) = \{B_i : \text{Center}(A_j) \in \text{Buffer}(B_i, r)\} \tag{2}$$

If a statistical area is covered by multiple base stations, its population is divided equally among those base stations. This ensures an equal division of load between the overlapping LTE base stations.

For each statistical area A_j covered by n base stations, the population contribution $P_{i,j}$ to each base station B_i is:

$$P_{i,j} = \frac{P_j}{n} \tag{3}$$

The total population served by each LTE base station B_i is the sum of the population contributions from all the statistical areas it covers.

$$P_i = \sum_{A_j \in \text{Covered\_Areas}(B_i)} P_{i,j} \tag{4}$$

The implementation of this analytical approach involves using geospatial data processing libraries such as GeoPandas to handle the spatial operations and Pandas for data manipulation. The key steps include creating buffer zones, performing spatial joins to identify overlapping areas, and aggregating the population data.

By estimating the number of people served by each LTE base station, we can determine the quality of service in this area.

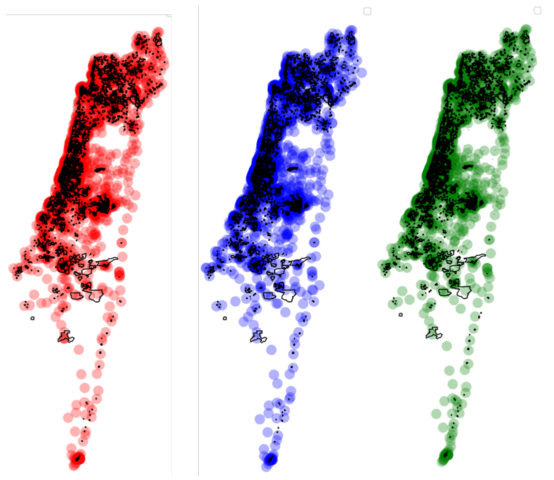

Fig. 2. Buffer zones of each LTE base station divided to companies: Pelephone (red), Cellcom (blue) and PHI (green).

B. Geographical Analysis

Analyzing the LTE base stations coverage of bus stations helps understanding the geographical distribution and service reach of the telecommunications infrastructure.

1) **Buffering Base Stations:** Create buffer zones around each LTE base station to represent their coverage areas. The buffer distance is determined based on the typical range of the LTE base stations.
2) **Spatial Join:** Perform a spatial join to identify which of the bus stations are located within the buffer zones of the LTE base stations.
3) **Counting Covered Stations:** Count the number of bus stations that fall within each coverage area and sum them in order to assess service reach in Israel.

IV. Result Analysis

Our dataset contains the location of 8,000 LTE base stations and around 3,000 statistical areas.

A. Population Analysis

In figure 3, a heat map of the statistical areas in Israel, showing the average number of users per one LTE base station in every statistical area.

We determined the coverage distance of an LTE base station to be 2,000 meters[6]. Having said that, it is important to mention that it is an average number and that the radius coverage is influenced by location, frequency, terrain and environment. Due to the fact that there is a large number of base stations we chose a fixed number for all of them. Moreover, we lack data concerning terrain, environment and network load.

We can note that most of the central district of Israel enjoys a good coverage. However, some areas in the northern district and in the southern district suffer from a poor coverage.

In figure 4, we note that the cities that suffer from the poorest coverage in Israel are Arab towns, ultra-orthodox cities, Beduin towns and Jewish settlements located near Arab towns.

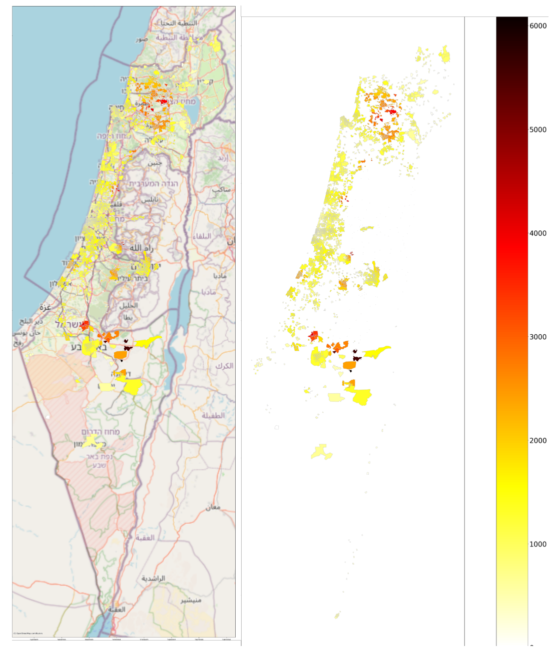

Fig. 3. Left: A heatmap of the number of users per an LTE base station with a map of Israel on the background, Right: The same heatmap without the map

In figure 5 we see that the largest cities in Israel enjoy good coverage more or less equally (except Ashdod).

However figures 6, 7 and 8 demonstrate that the coverage in ultra-orthodox, Beduin and Arab cities is poor, at least 4 times worse than the largest cities. (The coverage in the ultra-orthodox cities is better than in Arab and Beduin cities, but worse than the biggest cities).

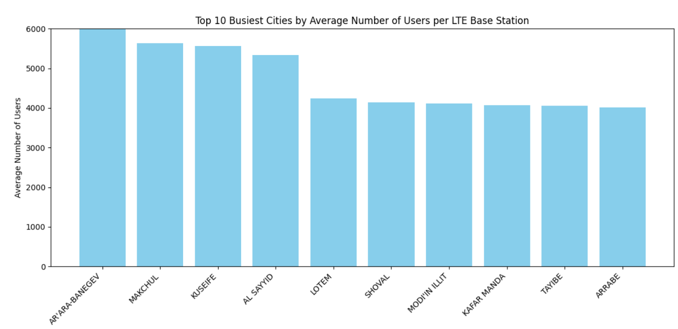

Fig. 4. The busiest areas in Israel in terms of population per LTE base station

B. Geographical Analysis

There are 32,248 bus stations in Israel. We examined the cellular coverage of bus stations as a mean to examine the coverage of roads and streets in Israel.

Once analyzing the coverage of bus stations one can note that most of the bus stations are under coverage. However, there are many bus stations that suffer from poor coverage - especially in southern district.

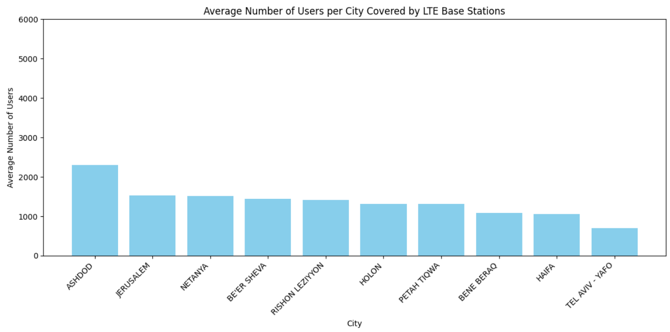

Fig. 5. The largest cities in Israel, population per base station

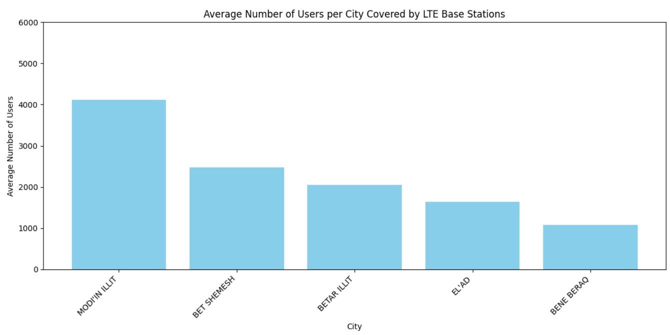

Fig. 6. Ultra-orthodox cities in Israel, population per base station

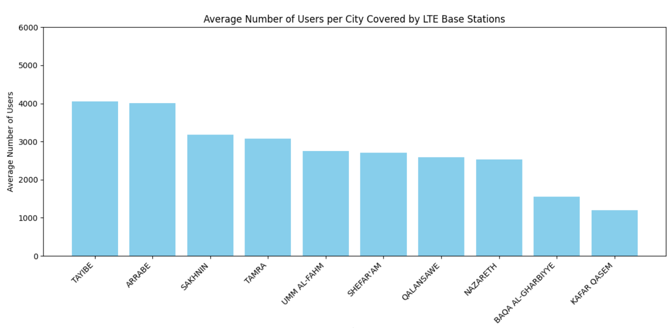

Fig. 7. Arab cities in Israel, population per base station

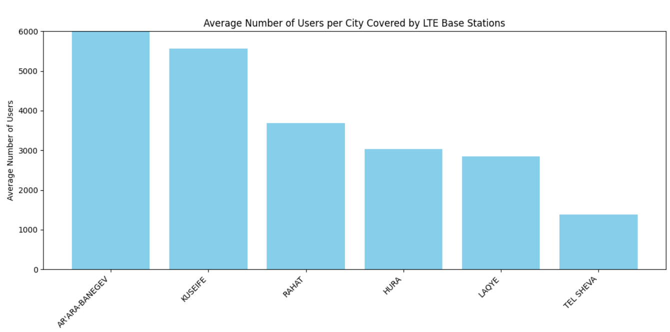

Fig. 8. Beduin cities in Israel, population per base station

There are 170 bus stations that are not covered - 60% of them are in the southern district. This can serve as a tool for the Ministry of Communication to supervise the geographic coverage of the cellular companies, in order to meet 97% geographic coverage of roads and cities as determined in their license.

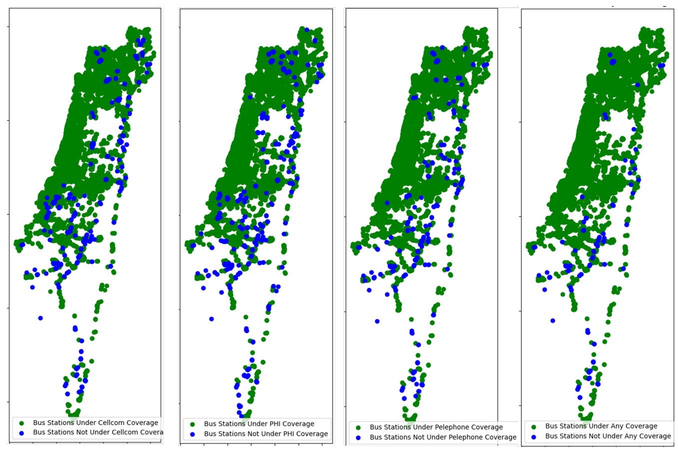

Fig. 9. Coverage of bus stations in Israel. From left to right: Cellcom, Partner (PHI), Pelephone, all companies.

V. Conclusion

In the State Comptroller's report in July 2024, it was determined that the ministry of communications lacks the tools to determine whether or not the LTE infrastructure companies comply with their license given to them by the Ministry. In this study, we offer a supporting tool for the Ministry of Communications to examine the coverage of the cellular companies in Israel. Thanks to the method we presented, We noticed that there is a large gap in terms of cellular coverage between the general population and the Arab, Bedouin and Ultra-orthodox population. Looking into journalist reports, we can offer two explanations to this finding: It was reported that LTE base stations were sabotaged by locals in these areas out of fear from harmful radiation[7]. The second explanation is the rapid growth of population in these cities which does not meet the development of new infrastructures. Looking into the State Comptroller's report, we found further reinforcement for our findings. The Comptroller Office conducted a survey among users and found that in the beduin city of Rahat, the Arab town Baqa al Gharbia and the neighboring Jewish town Harish users reported low satisfaction from the quality of service provided to them by their cellular provider. However, the cellular companies reported to the Ministry of Communication that these towns enjoy an excellent service (99%-100% coverage).

It is worth emphasizing that even settlements neighboring to settlements with poor coverage also suffer from poor cellular coverage due to the load on the LTE base stations, as was exemplified above with the town Harish.

Regarding the geographic coverage, we note that roads in the southern country suffer from poor coverage (for example road 90 and road 40). This finding reinforces our method as a valid one for the Ministry of Communications to determine whether the companies comply with the conditions of their license.

VI. Limitations

Our work also has limitations. One of them is the fact that our method is based on the population of a residential area, even though it is known that during the day the population changes (people go to work or school). In addition, we assumed that the radius covered by each LTE base station is the same. This assumption does not take into consideration further parameters such as the terrain typical to each area and environmental variables. We compensated for this lack of data using an average number since we have a large number of LTE base stations. We can make our tool more accurate, by using the precise radius coverage of each LTE base station. Since only the companies own this data, we developed another method as explained above.

Acknowledgment

I have used the dataset from Data Gov website and I also have used Chat GPT, Google Colab and Excel.

References

[1] State Comptroller's Report - Cellular Communication Infrastructures, July 2024. https://www.mevaker.gov.il/sites/DigitalLibrary/Documents/2024/2024.07-75A/2024.07-75A-104-Cellular.pdf

[2] F. Abed, "Drive Testing for Cellular Network Optimization," IEEE Communications Magazine, vol. 58, no. 5, pp. 70-75, 2020.

[3] M. Lee, "Crowdsourced Data for Mobile Network Performance Assessment," Journal of Network and Computer Applications, vol. 120, pp. 1-10, 2019.

[4] A. Kumar, "Real-Time Cellular Coverage Mapping Using Crowdsourced Data," IEEE Transactions on Mobile Computing, vol. 17, no. 8, pp. 1822-1834, 2018.

[5] C. B. Yang, "Spatial Clustering Methods for Network Coverage Analysis," Journal of Location Based Services, vol. 12, no. 2, pp. 89-101, 2018.

[6] Boulila, W., & Farah, I. R. (2017). A data mining approach to predicting cover crops performance using biophysical variables. In C. G. Ralston & A. C. Martínez (Eds.), Innovative Applications of Big Data in the Management of Complex Organizations (pp. 643-654). Springer

[7] Makorrishon, "The war of the Arab sector against cellular antennas: Yair Kraus," NRG (June 2, 2014), https://www.makorrishon.co.il/nrg/online/54/ART2/547/812.html.