

Data-Driven Profiling of Inland Areas: Studying Changes in Mobile Users Presence After COVID-19

Andrea Pimpinella*, Cristina Boniotti[†], Carmelo Ignaccolo[‡], Fabio Martignon*, Andrea Pavon^{||}, Luisa Venturini^{||}

* *Department of Management, Information and Production Engineering (DIGIP), University of Bergamo*

[†] *Department of Architecture, Built Environment and Construction Engineering (DABC), Politecnico di Milano*

[‡] *Department of Urban Studies and Planning (DUSP), Massachusetts Institute of Technology*

^{||} *Vodafone Group, Network Engineering and Delivery*

contact: andrea.pimpinella@unibg.it

Abstract—Cellular networks worldwide are currently experiencing a significant surge in service demand, forcing operators to focus on the accurate modeling of network dynamics as a key task to enhance efficiency. Besides being useful for optimizing network functioning, mobile data analytics have unleashed unforeseen opportunities to address several social and urban issues on a large scale. In this work, we seize such opportunities and propose a framework capable of profiling urban settlements based on the interplay between their attractiveness and the characteristics of the built environment. Focusing on the impact of the COVID-19 pandemic on mobile users' behavior, we conduct a comprehensive case study in Italy. Leveraging real-world mobile radio access data, we investigate the spatial variations in people's visiting patterns, providing insights into how these changes correlate with the social and urban context characterizing the reference area.

Index Terms—Cellular Data Analysis, Data-driven Modeling, Mobile Network Signatures, Clustering.

I. INTRODUCTION

Cellular networks worldwide are witnessing a significant surge in both cellular traffic volumes, projected to exceed 325 exabytes (10^{18} B) per month by 2028, and number of subscriptions, estimated to surpass 9 billion in 2028 [1]. This increase in complexity has raised Mobile Network Operators' (MNOs) awareness that the efficiency of numerous network operations (e.g., planning, dimensioning, monitoring, etc.) relies on how well mobile network dynamics can be analyzed and modeled in both urban and peripheral scenarios.

In addition to optimizing mobile network functionality, mobile data analytics has opened unforeseen opportunities to address various social (e.g., spatial investigations [2], effects of the pandemic on people's behaviours [3], [4], etc.) and urban (e.g., hot-spot detection [5], city design performance assessment [6], etc.) related issues at large scale. Call Detail Records (CDRs) data are often leveraged: in a nutshell, a record is generated each time a connected device interacts with the network for communication (call, text, Internet sessions, etc.) or operational (e.g., handover, etc.) purposes.

Given the ubiquity of mobile radio access, the connection between people's habits and the characteristics of their environment applies to cellular users as well: for example, mobile traffic demand near users' residences differs substantially from demand in business areas or urban transportation hot-spots [7], [8]. To leverage this relationship, *clustering* algorithms are

commonly designed to group radio access sites based on the spatio-temporal characteristics of their network activity. Sites are typically grouped according to the dynamics of the served traffic, an option that offers advantages to service providers for easing management tasks [8].

In this work, we employ mobile data analytics to investigate social life dynamics and study the effects of the COVID-19 pandemic on the presence of cellular users in Italian inland areas, focusing on the Valtellina region. Using a real-world dataset collected from Vodafone's cellular network, we cluster radio access sites based on temporal changes in the relative share of mobile users served before and after the pandemic. Finally, we provide insights into the relationship between the attractiveness of the selected area and the characteristics of the built and natural environment.

Results indicate that: i) mobile users' visiting behavior in Valtellina has changed after the pandemic and ii) this change is spatially heterogeneous, signifying a modification in the attractiveness of urban settlements based on post-pandemic needs. These insights can be valuable for both public entities and urban planners to profile urban settlements according to their attractiveness (quantified here using data related to mobile users' presence) and the characteristics of the built environment: we provide a glimpse of how clustering could be exploited with such a perspective.

This paper is organized as follows. Section II reviews the related literature, while Section III details the dataset and the methodological approach. Results are provided in Section IV, whereas Section V concludes the paper and outlines directions for future works.

II. RELATED WORKS

The study of the correlation between cellular service consumption and the characteristics of the geographical areas where cellular activities occur has gained significant attention in the past decade [2], [5]–[10]. A well-established approach to spatially characterize the utilization of cellular infrastructures is to group radio access sites based on either application-specific [7] or cumulative [8] traffic demand using clustering algorithms. Results consistently demonstrate a high correlation between the urbanization level of an area, the type of services provided to citizens, and cellular service consumption patterns.

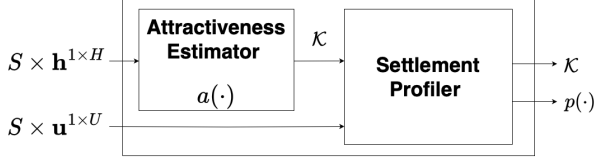


Fig. 1: Profiling Framework: first, AE learns a set of attractiveness profiles \mathcal{K} . Then, SP returns a model $p(\cdot)$ that associates a given environment with a specific profile $k \in \mathcal{K}$.

Indeed, mobile network metadata are often leveraged to address various urban-design related issues, including the modeling and estimation of people mobility [9], the characterization of human hotspots in a city or region [5], and the study of the relationship between people’s habits and specific city design choices [6], [10]. In particular, identifying hot-spots in a geographical area is crucial to understand how and for how long that area attracts citizens, and whether such attractiveness changes in response to modifications of citizens’ needs [5], [10]. In [6], authors perform a multivariate linear regression analysis to identify associations between clusters of citizens’ activity and several urban features, to foster an evidence-based approach to urban design and planning.

Beyond urban design, several recent works have used mobile data to investigate the impact of COVID-19 pandemic on people’s mobility and social life [3], [4]. In [3], the authors study the space-time variability of human presence before and during COVID-19 lockdown in selected inner areas of Italy. They observed a shift in the presence of people from dense urban cities, attributing remote working and near-home tourism as likely drivers of such phenomenon.

Merging the interests in both urban design and social insights, this work leverages experimental data collected at the access of a cellular network to: i) design a profiling framework able to correlate cellular data information with features related to the built and natural environment of a given geographical region; ii) cluster radio access sites based on the spatial and temporal variation of the share of connected cellular users, concerning the COVID-19 pandemic; iii) provide insights about the relationship between the clustering output and the characteristics of the underlying geographical area.

III. METHODOLOGY

This Section outlines the methodology used to cluster mobile radio network sites based on the changes in the distribution of mobile users during the COVID-19 pandemic in the Valtellina region. Firstly, we introduce the focus and the context of this study. Secondly, we describe the dataset, and finally we provide details about the clustering approach.

A. Overview

a) *Profiling Framework*: The framework comprises two building blocks (Figure 1). Firstly, the *Attractiveness Estimator* (AE) learns a set \mathcal{K} of attractiveness profiles using an unsupervised learning algorithm denoted as $a(\cdot)$. It takes S

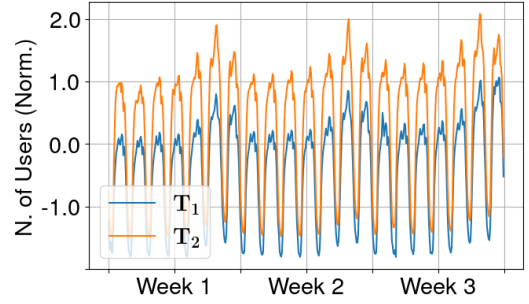


Fig. 2: Normalized hourly number of connected users during T_1 (blue) and T_2 (orange).

vectors $\mathbf{h}^{1 \times H}$ as input, each measuring the attractiveness of the corresponding settlement during a selected time horizon such as a week or a month. Note that H can be either equal to 1 (indicating a vector with only one entry, e.g., the weekly or monthly number of tourists in a settlement) or greater than 1 (indicating a vector with multiple entries, e.g., the hourly number of people visiting the settlement for each hour of the reference time window). Secondly, the *Settlement Profiler* (SP) employs a supervised learning approach to generate a classification model denoted as $p(\cdot)$; this model associates a settlement with given built and natural environment configuration, represented by a vector of U indicator values $\mathbf{u}^{1 \times U}$, with a specific attractiveness profile $k \in \mathcal{K}$. The vector $\mathbf{u}^{1 \times U}$ provides information that include, for example, the extension of the settlement’s urban footprint, the type of available physical and digital infrastructures, the type of landscape, the distribution of cultural heritage sites, and other related aspects.

b) *Focus*: In the next sections, we detail the operations of the AE, designing an algorithm capable of profiling a settlement based on a measure of its attractiveness. To illustrate, we take the Italian region of Valtellina as a case study and leverage k -means clustering, a well-known unsupervised learning approach [8]. However, instead of directly profiling settlements in the area, we cluster the set of eNodeBs deployed there. The input for clustering is derived from the temporal changes in the relative share of connected mobile users served before and after the COVID-19 pandemic. Although we acknowledge that the service area of each eNodeB may cover more than one settlement, we address the issue of mapping a physical location to the most likely serving cell through the use of coverage maps, that are often available to MNOs [11]. This issue is thus not considered in our work.

c) *Case Study*: Geographically situated in northern Italy and administratively ruled by the province of Sondrio, Valtellina is characterized by small, rural villages experiencing a gradual depopulation phenomenon [12]. Given the renaissance opportunities for inland territories and communities prompted by the pandemic, an analysis of people’s habits can contribute to defining strategies for enhancing such areas. The next Section introduces the dataset used in this work.

B. Dataset

We leverage a dataset of Key Performance Indicators (KPIs) measured at the access of the LTE network of Vodafone, a popular European mobile operator¹. The dataset comprises information from 61 eNodeBs serving the Valtellina region. We select as reference KPI the number of users that are Radio Resource Control (RRC) connected to each cell site, available in hourly sampled time series. Consequently, we exclude users employing fixed Internet access infrastructures from the discussion. Given that most connections from mountainous inland areas, such as Valtellina, rely on mobile networks for better overall performance (as fiber optics availability is typically limited), this exclusion has a negligible impact on our analysis.

We focus on two 3-week periods: $T_1 = \{20/01/2020, 10/02/2020\}$, immediately before the pandemic outbreak in Italy² and $T_2 = \{24/01/2022, 13/02/2022\}$, after the national government removed most COVID-19 related restrictions. Let $r(1, e, t)$ and $r(2, e, t)$ represent the number of users connected to the e -th eNodeB at hour t during T_1 and T_2 , respectively. The total hourly number of connections is computed as:

$$r(i, t) = \sum_e r(i, e, t), \quad i = 1, 2. \quad (1)$$

Figure 2 illustrates $r(1, t)$ (blue) and $r(2, t)$ (orange), with data normalised for privacy reasons. In both cases the number of connections increases from workdays to week-ends within each week, and from the first to the third week. This is partly attributed to a seasonality effect, as Valtellina is a popular destination for winter sports practitioners. Additionally, we observe that the overall weekly presence of mobile users in Valtellina has increased after the pandemic: on average, the hourly number of connections in T_2 is 30% higher than in T_1 . This trend contrasts with the long-term depopulation process experienced by most settlements in Valtellina over the last decades [12]. In the following, we describe how we process the data to cluster eNodeBs.

C. Clustering Procedure

We employ the k -means algorithm to cluster radio access sites based on the Euclidean distance among what we refer to as *change* signals. The change signal $c(e, t)$ of eNodeB e is defined as follows:

- 1) Firstly, we compute the Median Weekly Signatures $\mathbf{m}(1, e, t)$ and $\mathbf{m}(2, e, t)$ of $r(1, e, t)$ and $r(2, e, t)$. In essence, a median weekly hourly sample represents the median value of samples at the same hour on the same day (Monday, Tuesday, etc.) measured in the reference period. For more computational details, refer to [8].
- 2) Secondly, we normalize hourly-wise both $\mathbf{m}(1, e, t)$ and $\mathbf{m}(2, e, t)$ to $r(1, t)$ and $r(2, t)$. We refer to them as $\mathbf{m}^n(1, e, t)$ and $\mathbf{m}^n(2, e, t)$, respectively.

¹Vodafone covers more than 25% of the market share in Italy, and distributes uniformly across different users age and economic segments (AGCOM, 2023).

²First recognized COVID-19 case in Italy dates to 9/02/2020.

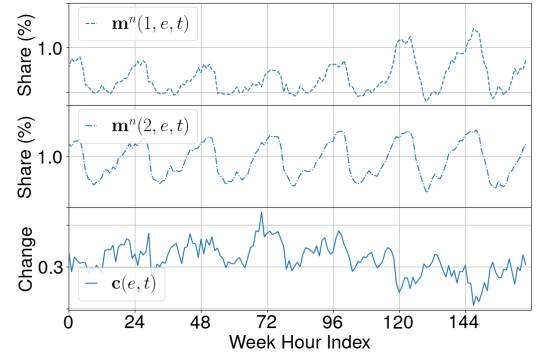


Fig. 3: Median weekly share of served users during T_1 (top) and T_2 (middle), and the change signal (bottom) resulting from their difference for the generic eNodeB e .

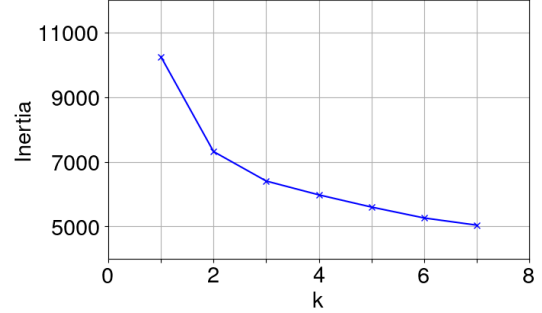


Fig. 4: Values of inertia obtained after the grid search procedure over the set $k = [1, \dots, 8]$.

- 3) Finally, we compute $\mathbf{c}(e, t)$ as the hourly-wise difference between $\mathbf{m}^n(2, e, t)$ and $\mathbf{m}^n(1, e, t)$, i.e.:

$$\mathbf{c}(e, t) = \mathbf{m}^n(2, e, t) - \mathbf{m}^n(1, e, t). \quad (2)$$

Therefore, $\mathbf{c}(e, t)$ represents the median variation from T_1 to T_2 of the share of the number of cellular users taken by eNodeB e at every hour t .

We provide in Figure 3 a representative example of $\mathbf{m}^n(1, e, t)$, $\mathbf{m}^n(2, e, t)$ and $\mathbf{c}(e, t)$ for one selected eNodeB. Repeating this process for each eNodeB in the network permits to perform k -means clustering with the change signals $\mathbf{c}(e, t)$ as input. The only hyper-parameter used by the algorithm, the number k of expected clusters, is determined through a data-driven approach involving a grid search over the set of candidate values $k = [1, \dots, 8]$. For each candidate, the algorithm is run with 1000 different initialization seeds, and the clustering configuration corresponding to the best obtained *inertia* value is saved. There is a trade-off between the inertia of a configuration (lower inertia indicates better robustness of the clustering output) and the number of expected clusters. Since the k -means algorithm is reliable for small numbers of clusters, a good model has low inertia and few clusters. Figure 4 displays the values of the inertia obtained after the grid search procedure. Considering that inertia decreases with

increasing values of k , we set $k = 4$ according to the well-known *Helbow* criterion [8].

In conclusion, we remark that our approach takes into account both spatial and temporal dimensions of the problem. On the one hand, $c(e, t)$ embeds spatial domain information by construction, as it represents the variation of the (spatial) share of the total number of users connected every hour to each eNodeB serving Valtellina. On the other hand, signals $c(e, t)$ are normalised to mean and variance (i.e., in the time domain) before being used, as our interest is to group eNodeBs regardless of the amplitude of the users' presence variation. We discuss the clustering results in the following Section.

IV. EXPERIMENTAL RESULTS

This section analyzes clustering outcomes, investigating cluster centroids (Subsection A). We study the driving factors of the clustering algorithm using Principal Components Analysis (PCA) on change signals (B). Further, we characterize clusters by exploring hourly user distributions during T_1 and T_2 (C). Lastly, we visualize the spatial distribution across Valtellina, using a color-coded map, depicting eNodeB locations and their coverage areas (D).

A. Clustering Output

In Figure 5 we display the centroids of the identified clusters, each one representing the change signals of all the eNodeBs in the corresponding cluster. For each centroid, a positive/negative value at a given hour indicates that the number of users served at that hour is greater/smaller than the centroid's average change in the week, represented by the black horizontal dashed line. Moving from the top to the bottom plot of Figure 5, we interpret the profiles as follows:

- 1) The red cluster encompasses 20 eNodeBs (33% of the total of 61 eNodeBs), and is the largest of our configuration. The centroid samples are positive during the night (from 0:00 a.m. to 8:00 a.m.), while they are slightly below the average during the day, for both workdays and week-ends. Thus, eNodeBs of this cluster have (on average) acquired more users at night than during the day, and we label them as *Nightly Gainers* (NG).
- 2) The blue centroid represents 19 eNodeBs (31%), which we name as *Daily Gainers* (DG). In contrast to the red centroid, this centroid lies above the average line from 8:00 a.m. to 11:00 p.m. both in workdays and week-ends, while negative values are shown during the night.
- 3) The green centroid represents 14 out of 61 eNodeBs (21%) and shows a profile slightly above the average during workdays. Conversely, values turn negative during the week-end, indicating that the eNodeBs of this cluster have more likely lost users from T_1 to T_2 during the week-end than on workdays. Therefore, we label such eNodeBs as *Week-End Losers* (WEL).
- 4) The orange cluster contains 8 eNodeBs (14%), and is the smallest of our configuration. This centroid lies mostly below the average during work-days and shows

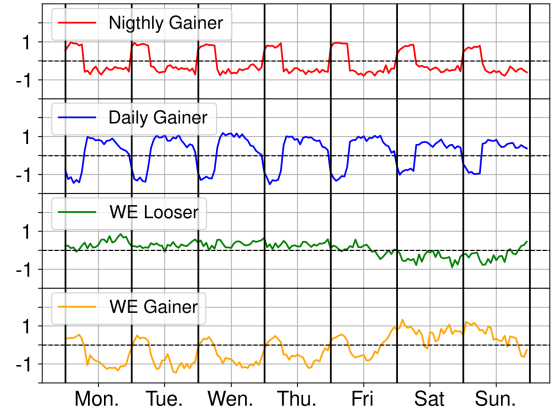


Fig. 5: Weekly profiles of the identified centroids.

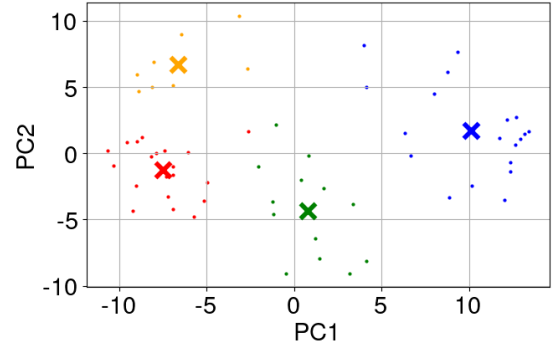


Fig. 6: Scatterplot of the 61 eNodeBs along the first two principal components (PC1 and PC2).

a slightly positive upward trend approaching the week-end, when change values turn positive. Hence, we name the eNodeBs in this cluster as *Week-End Gainers* (WEG).

Overall, DG served almost 43% of the overall users connected during T_2 , making it the most-serving cluster with a 10% higher share than the second-ranked NG. As far as WEL and WEG clusters are concerned, they served during T_2 a similar fraction of users (13% and 11%, respectively) despite the former cluster being almost twice as large as the latter.

B. Drivers of the Clustering Process

To gain a deeper understanding of how the algorithm learnt the clustering configuration, we represent the set of vectors $c(e, t)$ for $e = \{1, \dots, 61\}$ as an $E \times H$ matrix. In this matrix, each row vector corresponds to a given $c(e, t)$ and $H = 168$ corresponds to the number of hours composing a week. We use such matrix to perform PCA³: Figure 6 presents a color-coded scatterplot illustrating the first principal component versus the second one, while Figure 7 reports the time series of the first 3 principal components, that together explain approximately 55% of the data variance. Note that in Figure 6 each point represents a specific eNodeB in the

³PCA is a well-known unsupervised approach capable of reducing the dimensionality of large datasets while minimizing information loss.

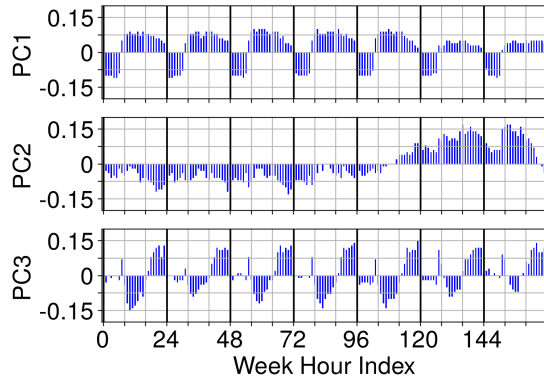


Fig. 7: Time series of the first three principal components.

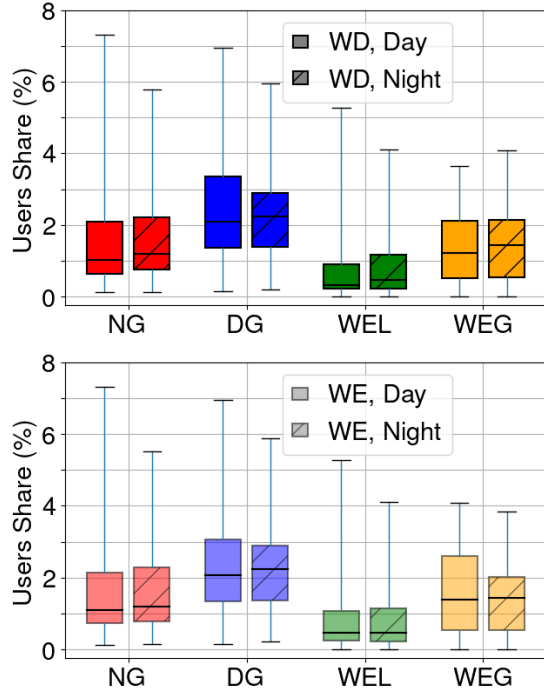


Fig. 8: Distribution of the hourly share of served users, during Work Days (WD, top) and Week-Ends (WE, bottom).

network, while crosses denote the locations of each cluster's centroid in the principal Euclidean space. Figure 6 reveals four different groups of points, accurately identified by the clustering algorithm. Additionally, looking at Figure 7, we observe that:

- the first component (PC1) captures eNodeBs that gained users during daily hours in T_2 , particularly on workdays rather than week-ends;
- In contrast, the second PCA component (PC2) is associated to higher positive variations observed during the week-end;
- Lastly, the third PCA component (PC3) correlates with eNodeBs that acquired users during the evening (after 8.00 pm) compared to T_1 .

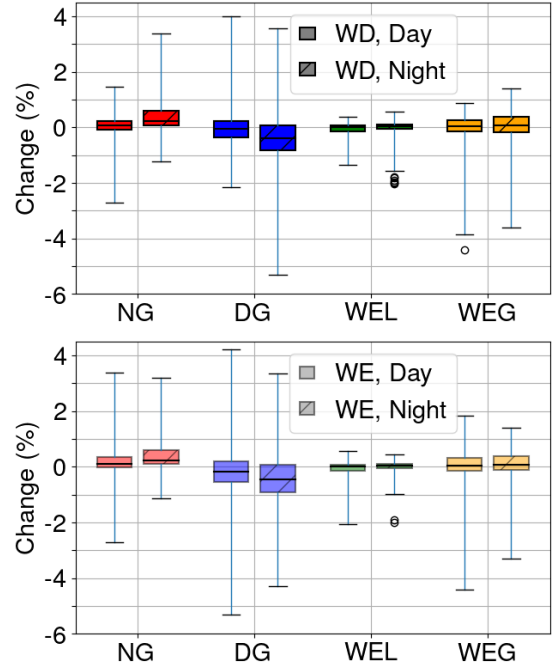


Fig. 9: Box-plot of hourly change of served users from T_1 to T_2 , for Work Days (WD, top) and Week-Ends (WE, bottom).

Examining the identified centroids, it is evident that the algorithm combined information from each PCA component and appropriately clustered eNodeBs accordingly.

C. Characterization of the Generated Clusters

To investigate the characteristics of the clusters, we illustrate in Figure 8 the cluster-wise distributions of the hourly share of served users during the period T_2 . For our analysis, we categorize hours between 8.00 a.m. and 8.00 p.m as daily hours, and the opposite as nightly hours. As depicted, during workdays (top plot), the median share of hourly served users is higher during nightly hours for all clusters. While this holds true for NG and WEL clusters even for the 75-th percentile values, DG and WEG clusters experience an opposite trend, with higher values during daily hours. A similar pattern emerges during the week-ends (bottom plot), except for the WEL cluster where the median share of hourly served users during daily hours is nearly equal to that observed at night.

It is important to note that the designed algorithm does not consider the *amplitude* and *sign* of the change signals when clustering eNodeBs, but solely focus on their *shape*. In other words, a given cluster might group sites that have experienced either an overall decrease or increase in the total number of users from T_1 to T_2 , given that their error signals share similar temporal dynamics. To elaborate on this aspect, we plot in Figure 9 the (non-normalized) hourly change values from T_1 to T_2 . Notably, 75% of the eNodeBs labelled as nightly gainers have not lost users compared to the pre-pandemic scenario, both during workdays (top plot) and week-ends (bottom plot). This contrasts with the DG cluster, where negative median

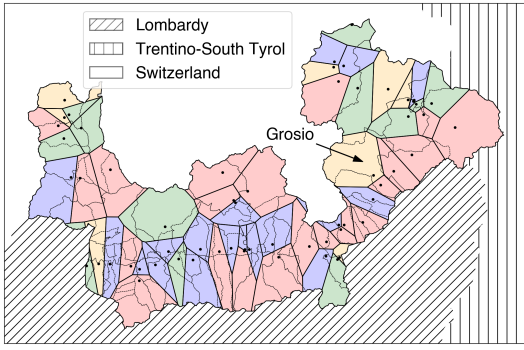


Fig. 10: Map of Valtellina: it is part of Lombardy, and borders on Switzerland and Trentino-South Tyrol to the north and east, respectively. Voronoi polygons are color-coded according to the cluster of the corresponding eNodeB.

changes are observed irrespective of the day and hour type. Regarding the WEL and WEG clusters, median change values are close to 0, but the distributions for workdays and week-ends are displaced towards negative changes (for both day and night), suggesting that, in general, the members of these clusters are more likely to have lost users from 2020 to 2022.

D. How Do Clusters Physically Distribute?

Figure 10 displays a colour-coded map of Valtellina, with black markers indicating the eNodeBs' location and thicker black boundaries outlining their radio coverage areas, approximated here by Voronoi polygons⁴. Each polygon is colored based on the corresponding eNodeB's cluster. Additionally, thinner black lines represent the administrative boundaries of municipalities in the valley. Observing the map, most DG (blue polygons) are concentrated in the middle of Valtellina. In fact, municipalities in this area have recently upgraded their digital infrastructures thanks to public investments, a development we believe significantly enhances users' experiences, especially in facilitating remote working options. It is noteworthy that DG also acquire users during the week-end, possibly influenced by various initiatives undertaken by local public entities to promote tourism in these areas. Turning attention to NG (red polygons), they are predominantly situated in lateral valleys. This distribution may be attributed to the limited availability of digital infrastructures and services, as well as job opportunities, prompting citizens to commute during daily hours and return home at night. Lastly, we emphasize that Grosio, administratively overlapping with the Voronoi polygon pinpointed by an arrow on the map, is categorized as WEG. Since 2020, it has served as a central hub for optical fiber distribution to surrounding districts.

V. CONCLUSIONS AND FUTURE WORKS

This work envisions a framework to profile urban settlements based on their attractiveness and model its relationship

⁴How to represent the coverage area of a radio access point without resorting to coverage maps is still a topic of literature debate [13].

with built and natural environment characteristics. Using the Italian region of Valtellina as a case study, we detail the design of the attractiveness estimator. Leveraging real-world cellular network data, we apply k-means to cluster eNodeBs based on the median variation of connected user shares pre- and post-COVID-19. Interestingly, we observe a 30% increase in cellular users in Valtellina post-pandemic, contrary to the depopulation trend in the region. Results indicate that 33% of eNodeBs gained users at night compared to T_1 , while 31% increased their shares during daytime. Conversely, 14% gained users on week-ends, with 21% experiencing a decline. We present a color-coded map illustrating the clusters, offering insights and rationale for these variations. In conclusion, beyond improving living spaces, modeling cellular user presence based on geographical environment types can guide MNOs in energy-aware management strategies. Future work includes i) designing the Settlement Profiler and ii) extending the approach to other settlements, considering diverse regions, including metropolitan cities.

VI. ACKNOWLEDGMENTS

This work was supported in part by project *SERICS* (PE00000014) under the NRRP MUR program funded by the EU - NGEU. Also, it was partly supported by PRIN Project *NEWTON* (2022ZA8T22). We finally thank Prof.s Brent D. Ryan (MIT) and Giovanna Fossa (Politecnico di Milano) for their supervision within "The Smart Villages of Italy" project.

REFERENCES

- [1] Ericsson, "Ericsson mobility report," 2022.
- [2] M. Ghahramani *et al.*, "Mobile phone data analysis: A spatial exploration toward hotspot detection," *IEEE Trans. on Automat. Science and Eng.*, vol. 16, no. 1, pp. 351–362, 2018.
- [3] G. Lanza *et al.*, "Impacts of the Covid-19 pandemic in inner areas. Remote work and near-home tourism through mobile phone data in Piacenza Apennine," *Journ. of Land Use, Mobility and Environment*, vol. 2, pp. 73–89, 2022.
- [4] A. Legeby *et al.*, "New urban habits in Stockholm following COVID-19," *Urban Studies*, vol. 60, no. 8, pp. 1448–1464, 2023.
- [5] R. P. Lopez and J. Ferreira, "Identifying spatio-temporal hotspots of human activity that are popular non-work destinations," *Env. Plann. B: Urban Analytics and City Science*, vol. 48, no. 3, pp. 433–448, 2021.
- [6] A. Noyman *et al.*, "Reversed urbanism: Inferring urban performance through behavioral patterns in temporal telecom data," *Env. Plann. B: Urban Analytics and City Science*, vol. 46, no. 8, pp. 1480–1498, 2019.
- [7] R. Singh *et al.*, "Urban vibes and rural charms: Analysis of geographic diversity in mobile service usage at national scale," in *The W.W.W. Conf.*, 2019, pp. 1724–1734.
- [8] A. Pimpinella *et al.*, "Forecasting Busy-Hour Downlink Traffic in Cellular Networks," in *IEEE ICC*, 2022, pp. 4336–4341.
- [9] G. Khodabandelou *et al.*, "Estimation of static and dynamic urban populations with mobile network metadata," *IEEE Trans. on Mobile Computing*, vol. 18, no. 9, pp. 2034–2047, 2018.
- [10] T. Louail *et al.*, "From mobile phone data to the spatial structure of cities," *Scientific Reports*, vol. 4, no. 1, p. 5276, 2014.
- [11] A. Pimpinella *et al.*, "Unsatisfied today, satisfied tomorrow: A simulation framework for performance evaluation of crowdsourcing-based network monitoring," *Comp. Commun.*, vol. 182, pp. 184–197, 2022.
- [12] F. Mazza, "Tourism and Marginalization in the Alps. The Case of Media-Alta Valtellina Region," in *Int. Conf. on Tourism Research*, vol. 6, no. 1, 2023, pp. 440–448.
- [13] O. E. Martínez-Durive *et al.*, "VoronoiBoost: Data-driven Probabilistic Spatial Mapping of Mobile Network Metadata," in *19th Annual IEEE Int. Conf. on Sens., Comm. and Netw. (SECON)*. IEEE, 2022, pp. 100–108.