# Cell-to-Cell Activity Prediction for Smart Cities

Blerim Cici, Emmanouil Alimpertis, Alexander Ihler, Athina Markopoulou
University of California, Irvine
*{bcici, ealimper, aihler, athina}@uci.edu*

***Abstract*—In this paper, we analyze data from a large mobile phone provider in Europe, pertaining to time series of aggregate communication volume $A_{i,j}(t) > 0$ between cells $i$ and $j$, for all pairs of cells in a city over a month. We develop a methodology for predicting the future (in particular whether two cells will talk to each other $A_{i,j}(t) > 0$) based on past activity. Our data set is sparse, with 80% of the values being zero, which makes prediction challenging. We formulate the problem as binary classification and, using decision trees and random forests, we are able to achieve $85\%$ accuracy. By giving higher weight to false positives, which cost more to network operators, than false negatives, we improved recall from 40% to 94%. We briefly outline potential applications of this prediction capability to improve network planning, green small cells, and understanding urban ecology, all of which can inform policies and urban planning.**

## I. Introduction

Cellular penetration has increased dramatically over the past decades and the number of unique mobile subscribers is estimated around 3.4 billion users [17]. At the same time there is an even greater growth in demand for wireless access bandwidth worldwide, due to the fast adoption of smartphones. The traffic volume generated by mobile phones will increase approximately by 8 times in 2020 (30.6 exabytes/month) compared to 2015 (3.7 exabytes/month), according to traffic trends forecasts [6].

To address this demand, mobile phone providers and the 3GPP are currently working on improvements to the current 4G standards as well as on future 5G networks [15]. More specifically, a mixture of macro-cells and small cells (i.e. heterogeneous nets) is currently being considered for increasing 4G capacity. Small cells are feasible by utilizing femtocells [3], i.e. low power base stations with limited range, typically designed for use in a home or business, covering the spectrum holes of the larger cells. This shift (towards smaller cells and denser networks) is closely connected with a shift towards virtualization of computational resources, that follows software defined networking (SDN) and self-organizing networks (SON) principles[1].

However, a dense infrastructure is complicated and costly to maintain. The energy consumption of base stations is one of the largest costs for mobile phone providers [7]. Hence, they try to make their infrastructure more energy efficient, e.g. by switching femtocells on or off or by lowering the transmission power. The aforementioned technologies will incorporate the necessary logic for smart decisions and network configuration based on network events, to automate resource allocation. For instance, [8] describes the architecture of SDN - SON where traffic prediction algorithms will be utilized in the control plane for the assignment of virtualized radio resources. Thus, being able to predict cellular traffic patterns city-wide, can inform and enable network provisioning and control.

In addition, cellular activity reflects information about human activity patterns in a city. In our prior work [5], we showed the connection between cellphone activity (in particular, aggregate cellular activity per cell) and urban ecology. Using time series analysis, we showed that the seasonal component captures regular patterns of socio-economic activity within an area and can be used to segment a city into distinct clusters (such as business, residential, etc), while the residual component enables the detection of regions that are subject to mutual social influence or in direct communication contact. Such intra-urban structure, often referred to as urban ecology, is difficult to obtain using traditional methods (e.g., informant interviews, ethnographic observation, etc.) especially today that urban growth is rapid, but can be invaluable for urban governance (e.g., urban planning, infrastructure management, administration, and law enforcement). Understanding city-wide human activity patterns as manifested in cellular activity, is an opportunity to achieve that goal in an automated and inexpensive way that also covers a large part of the population.

In all the aforementioned cases, accurate prediction of mobile phone traffic in a city is necessary for enabling urban planning and a number of smart city applications. In this paper, we develop a building block in that direction: machine learning techniques for cell-to-cell mobile traffic prediction based on past cellular records,.

We analyzed a data set provided by Telecom Italia as part of the Big Data Challenge [18] competition, and more specifically the part of the dataset that describes the intra-city activity in Milan: time series $A_{i,j}(t)$ describe the communication volume between two areas of the city, $i$ and $j$, for $t = 1..N$. We formulate the traffic prediction problem as a classification problem. Based on past activity our goal is to predict whether two cells will talk to each other during at time $t$, i.e., $A_{i,j}(t) > 0$. First, we visualize important aspects of our data using SVD to better understand the data. We use the insights gained from the data analysis for feature selection; for example, we found that neighbors tend to talk more to each other and are more correlated. Second, we used decision tree classifiers and random forest in order to do prediction. We were able to achieve accuracy $85\%$, which outperforms the naive max-class predictor $(80\%)$ that predicts the most

This work has been supported by NSF awards 1028394, 1254071, and 1526736 . Emmanouil Alimpertis was supported by an Broadcom Fellowship.

[1]SON is an example of this trend [11, 2, 10]. SON is a software module responsible for planning, configuring, and managing the cellular infrastructure. For example, SON could use cognitive radio techniques to exploit under-utilized spectrum in the unlicensed bands, during high load hours [15].

frequent class. A key insight and challenge was the sparsity of the dataset: most cell pairs have zero communication activity with each other. This leads to high skewness of our classes and low recall rate (lower than 40%). Since, $F_p$ (false positive) and $F_n$ (false negative) errors don't have the same cost for providers, we show how to improve recall up to 94%, by giving higher weight on $F_p$.

The rest of this paper as organized as follows. Section II discusses related work, and Section III formulates the problem. Section IV presents the data and the analysis. Section V describes the methodology and Section VI the results. Finally, Section VII concludes the paper.

## II. Related Work

Related work can be roughly classified in three categories: (a) traffic volume prediction from a mobile telephony cell tower, (b) link prediction in telecommunication or or social networks and (c) analysis and assessment of network operators' data sets which reveals the spatio-temporal characteristics and the dynamics of the cellular network infrastructure.

In traffic volume prediction, the goal is to forecast the volume (voice or data) generated by a specific base station (i.e. cell tower) for a future time window, given historic traffic traces. Methodologies include, but not limited to, moving average [7], Holt-Winters's exponential smoothing [7], [19], [12], hybrid prediction models [7], temporal compressive sensing [8] and Kalman filtering [7]. For instance, work in [8] utilizes entropy for assessing the predictability of the traffic and quantify what time window (temporal dimension) and how many adjacent cells (spatial dimension) would actually help. Furthermore, [7] proposes a framework for optimizing power consumption by switching off a portion of the base stations in low network traffic condition. Interestingly, [7] distinguishes the base stations between the typical traffic profiles and the opportunistic profiles (e.g. stadiums where the traffic is present only in weekends), which was a key observation made also by our prior work [5].

However, the traffic volume prediction problem differs significantly in 4G and 5G due to the small cells deployment [4]. A femtocell covers a much smaller area with less users, therefore, bursty traffic is more likely rather than a periodic volume activity which is usually generated by a macro-cell. More interestingly, burstiness and sparsity of the traffic were observed in our data set analysis. Thus, [4] proposes a solution which combines Gaussian processes (GPs) and kernel based methods for the periodic component and tolerance intervals for the bursty component. The data used are a combination of synthetic and real data sets. In contrast, our work studies a real world data set and we predict *if* a cell $i$ communicate with a cell $j$, i.e. a binary classification problem considering directed communication, which has not been studied by any of the previous works.

Link prediction in networks (social networks, IP subnetworks, mobile phones etc) is also related to our problem. The goal is to predict if two nodes of the network (e.g. two persons in an social network or two mobiles) will form a link and communicate at time $t$. For instance, work in [16] tries to predict a network attack, given historic data and by considering properties that network attackers and regular users share. The authors use the recommendation systems framework and utilize SVD for principal component analysis, an idea explored in this paper as well. Work in [9] considers the problem of link prediction in time $t$ for a data set containing phone calls between users. It investigates several factors such as the class imbalance problem, the sparsity of the links, time and statistical features , the strong neighborhoods and other topological features of the phone calls graph. Then, it assesses several supervised learning approaches and proposes a novel flow-based predictor.

Analysis of network operators' data sets, such as Call Detail Records (CDRs), have also been studied [14], [13]. The casual influence from a base station to neighboring base stations load is studied in [14] to assess Granger Causality for traffic prediction. In [14], the time granularity for traffic aggregation is studied showing higher cross correlation between pairs of base stations for time interval of one hour vs 10min intervals. Our prior work in [5] also looked at the data set from the city of Milan and used the aggregate activity per cell as a signature of human activity in that cell, in order to cluster similar areas of the city together for urban ecology. In contrast, this paper (i) studies not a single cell time series but cell-to-cell communication series and (ii) its goal is to predict future based on past activity.

In summary, this paper focuses on traffic prediction between two different areas of the city/cellular network and not a call prediction between two independent users, and has the following main differences compared to prior work. First, we consider aggregated CDRs in the spatial dimension (i.e., total volume of calls between all users in the two cells), while [9] considers phone calls between individual users and do not not consider the problem from the perspective of cellular providers: false negatives can be really costly. Last but not least, with all the concerns regarding privacy [20], aggregated CDRs are more likely to be available from the providers.

## III. Formulation

Let $S$ be the set of cells; in this paper, the term cell refers to a small area unit (in the city of Milan), which we will describe in the next section[2]. And, let $T$ be the time axis as a set of timestamps $T = \{t_1, t_2, ..., t_m\}$. We denote as $A_{i,j}(t)$ the volume of mobile activity from cell $i$ to cell $j$, where $t \in T$, and $i, j \in S$.

We formulate the traffic prediction problem as a *binary* classification problem. Given activity series $A_{i,j}(t)$, which shows the activity from $i$ to $j$ at time $t$, we build a set of features using the past ($< t$) records, and our goal is to predict if $A_{i,j}(t) > 0$. In other words, our goal is to predict if cells $i$ and $j$ communicate at time $t$ (class 1), or not (class 0).

Finally, we partition time $T$ into two subsets: (i) a training set called $T_{train}$ and (ii) a testing set called $T_{test}$. This is done via a random 70/30 split of the data, where 70% of the data is used for training and the remaining 30% for testing.

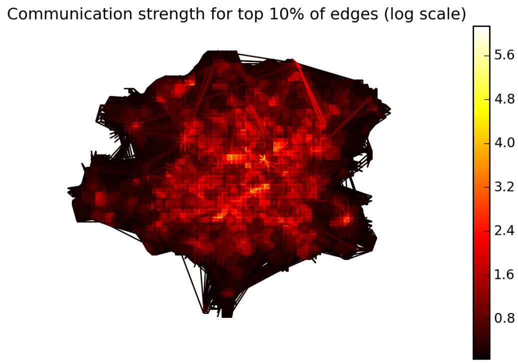



Fig. 1. The figure shows the communication strength between cells. The communication strength have been calculated by aggregating the interaction during the $1^{st}$ of Nov. 2013. We can observe that there is strong communications between neighboring cells. For clarity we show only the top 10% of the edges.

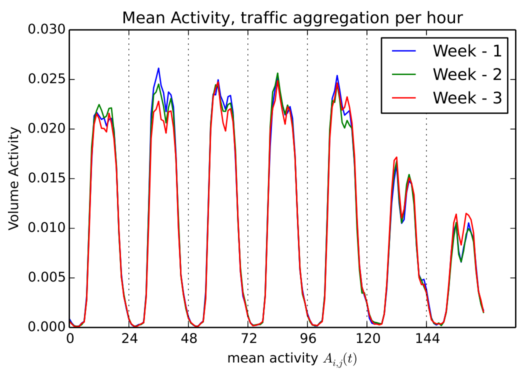



Fig. 2. Average activity $A_{i,j}(t)$. We see that when mobile phone activity, when averaged across all cell-to-cell traffic, is predictable and follows expected daily and weekly patterns. Also, these patterns are similar across various weeks.

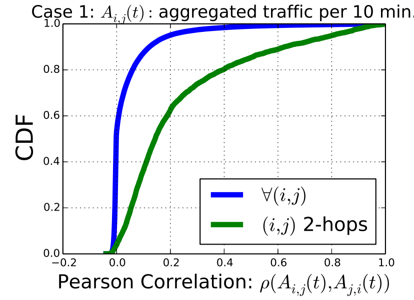
## IV. Data and Key Observations

Our data set consists of time series of aggregate cell phone traffic sent or received by users within small areal units in the city of Milan, made available for the *Big Data Challenge* [18] competition. In this paper, we focused on a 4-week period of November 2013. The city of Milan, an area of 550 $km^2$, was divided into a $100 \times 100$ grid. Each square of the grid has the same dimensions: a side length of 0.235 $km$ and an area of 0.055 $km^2$. This is the areal unit we use throughout the paper, and we refer to it as a "cell". The temporal unit is the 10-minute interval[3]. The data set contains information regarding the *directional interaction strength* (as per terminology in [5]) between two cells, based on the calls exchanged between them. Each activity record consists of the following fields: ID of cell $i$ where call was initiated from, ID of cell $j$ where the call was made to, time slot, and value of directional strength from $i$ to $j$. Fig. 1 visualizes the 10% strongest connections.[4]

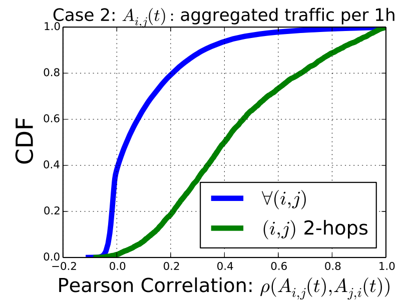
[2]This is the best approximation for a cell tower that is publicly available.

[3]Telecom Italia, which is the data set provider, decided the spatial and the temporal granularity of the data before making it available for the competition.

[4]We focus on the prediction of voice traffic since dropped or bad quality voice calls are noticed immediately as a"poor service" by the customers. However, the methodology should apply to any data set of cell-to-cell communication activity.




(a) 10-minute time interval




(b) 1-hour time interval

Fig. 3. The above figures show the distribution of activity correlations for all pairs of cells ($\forall(i,j)$), as well as pairs that are within a maximum distance of 2-hops; the correlation for a pair of cells, $i$ and $j$, is calculate by this formula: $\rho\left(A_{i,j}(t), A_{j,i}(t)\right)$, where $\rho$ denotes the Pearson correlation. We observe that there is much higher correlation when $i$ and $j$ are neighbors. Moreover, this picture shows that by aggregating cell phone activity into hourly time reports then traffic becomes more structured ($A_{i,j}(t)$ and $A_{j,i}(t)$ are more correlated).

### A. Aggregation of Traffic per Hour

The initial data consist of 10-minute traffic reports. However, we aggregated traffic per 1 hour because (1) traffic in such short intervals is very dynamic and fluctuates heavily, (2) allocation of resources in the cellular infrastructure (e.g. by SON) is not an easy task and planning of resource allocation in 10-min. intervals may lead to unstable networks or excessive overhead, (3) 95% of the activity is zero in such short time intervals, making traffic very sparse and (4) work in [14] faced the same dilemma regarding the time granularity for traffic aggregation (10-min. vs 1 hour) and concluded in 1-hour aggregation since it had higher cross correlation between pairs of base stations.

Traffic aggregation leads to more predictable series. As you can you see from Fig. 3(a), in the case of 10-minute time intervals, traffic between pairs of cells seems to be random and uncorrelated; 60% of the random pairs have zero correlation and more than 95% of them have a correlation lower than 0.2. However, when we aggregate traffic into one-hour time intervals, the time series become more similar and correlated (See Fig. 3(b)).

### B. Challenges and Key Insights

In this section we present some of the intuitions that we obtained from our exploratory analysis, and we highlight the challenges for the traffic prediction.

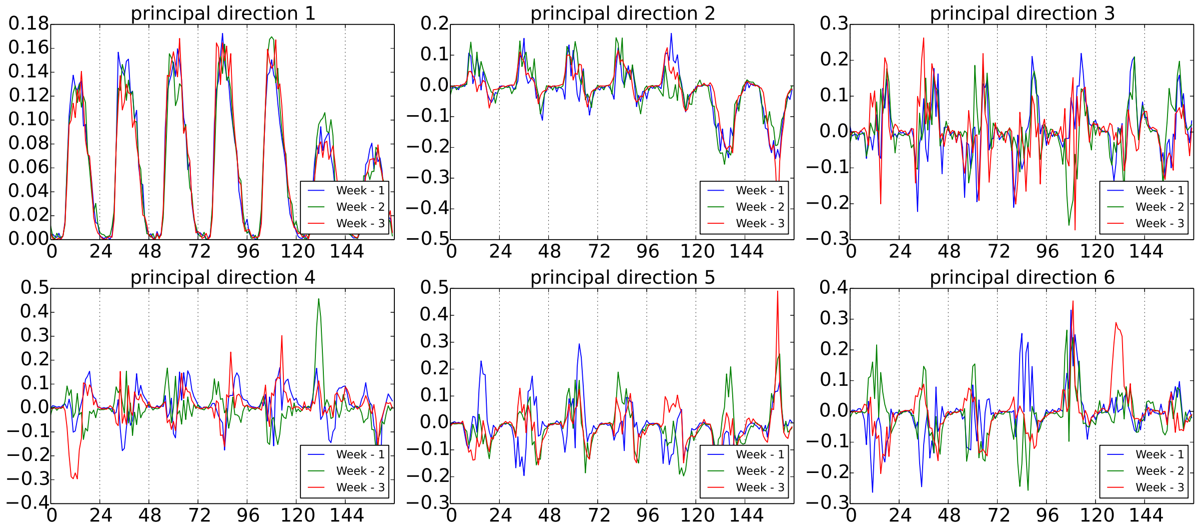
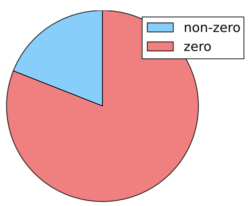
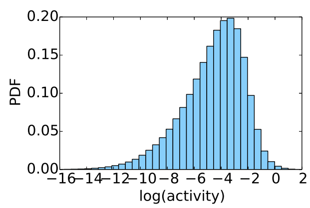



Fig. 4. SVD of $A_{(i,j)(t)}$ for 1week for aggregated data. Top-6 Principal Components.




(a) sparsity




(b) non-zero distribution

Fig. 5. Distribution of communication values. These figures describe a zero-inflated and skewed distribution.

(1) *Zero–Inflated Distribution:* One of the main challenges with cell-to-cell communication data is their sparsity. The distribution of the cell phone activity is a zero–inflated distribution – almost 80% the activity is zero (See Fig. 5) – while the remaining non-zero activity follows a skewed distribution. This makes prediction very hard since we cannot fit traditional time series models. Also, on aggregate mobile activity exhibits well-understood seasonal patterns and is easy to predict (See Fig. 2), but cell-to-cell traffic is dynamic, it fluctuates and prediction is hard.

(2) *Strong communication between Neighboring Cells:* Traffic between neighboring cells is much stronger (see Fig. 1), in comparison to the rest of the city. Also, traffic between neighboring cells is more structured, e.g. $A_{i,j}(t)$ and $A_{j,i}(t)$ are more likely to be correlated when cells $i$ and $j$ are within a 2-hop distance (See Fig. 3(b)).

(3) *We observe seasonal patterns in data:* In order to get a better understanding of the data, we decompose the activity series into their first six principal components, which we achieve via singular value decomposition (SVD). Fig. 4 demonstrates the components of the traffic from three different weeks. We observe that the first two principal components are structured and tend to be similar across weeks, but the remaining principal components look more spiky and dissimilar across weeks. For example, the second principal component shows the areas that exchange traffic *only* on weekdays and that do not communicate on weekend (observe the negative direction in weekends). This happens for example in universities or in business areas. In addition, there are spikes in the traffic around noon (5-th principal component) which models another "direction" of cellular traffic. This is also demonstrated by Fig. 6, where we use the principal components for a week (training week) as a model for another (testing week). The parameter $k$ (number of principal components used) denotes the complexity of the model. We observe that as the complexity increases the model is a better fit for the training week – the mean squared error (MSE) decreases. However, MSE for the testing week MSE decreases until $k = 2$, and the it start increasing. This shows that we cannot expect much gain in prediction from the seasonal patterns of our data.

(4) *Unexpected events affect communication:* Finally, traffic can be affected by unpredictable event. For example, in Fig 4, for the second principal component we observed a significant difference at the traffic level for the Friday ($t = 96 \cdots 120$) of the $1^{st}$ and the $2^{nd}$ week. This principal component encapsulates the traffic during the week days as we discussed. The traffic in the 2nd week is significant lower *only* for Friday. This day was the 15-th of November of 2013. We were intrigued from this difference and we searched for potential causes. After a short search in Google, we found that the 15-th of November was the first day of the big social protests and strikes in Italy in 2013 [1]. Apart from the fact that mobile traffic can be affected by unexpected events, this also shows that cell phone activity series can enable many other types of Smart City application, i.e. they can be used to reveal abnormal activities in a city.

## V. Methodology

Since we are dealing with a zero-inflated distribution – and our data are skewed towards zero – it is difficult to apply classical time series prediction methodologies, such as ARIMA models. Instead we will treat the prediction task as

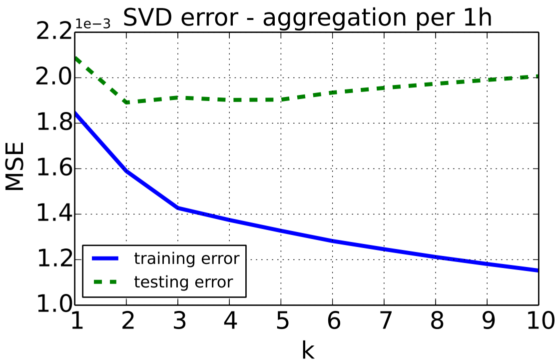



Fig. 6. Error for $A_{ij}(t)$ reconstruction by $k$ principal components of the data utilizing SVD (aggregation per hour). For this case, the utilization of the first principal components can improve the error on the testing data.

a classification problem; we will generate a set of features for each prediction we want to make, and we will apply standard but powerful classifiers. More specifically we will use a Random Forest Classifier. Next, we describe the features we used for the prediction. This decision was inspired from key insights (1) and (4) from the previous section.

### A. Feature Selection

For each communication series from cell $i$ to cell $j$ we use the following features:

*1) Static features:* We denote as static features, those features that are constant across weeks. These are:

- Geographic distance between cell $i$ and cell $j$. This was inspired by our earlier analysis that show that neighboring cells tend to talk more.
- Hour of the day. Human activity and communication is heavily influenced by the hour of the day.
- Day of the week. Human activity and communication may change depending on the day of the week (e.g. Monday vs. Saturday).

The geographic distance feature was inspired from key insights (2) of the previous section, while the other two static features were inspired by key insight (3).

*2) Dynamic features:* These are features that change from one week – or even day – to the other.

- Past traffic (3 previous hours) of $A_{i,j}$. For example, for target value $A_{i,j}(t)$, the features are $A_{i,j}(t-1)$, $A_{i,j}(t-2)$, $A_{i,j}(t-3)$.
- Past traffic (3 previous hours) of reverse series, i.e. from cell $j$ to cell $i$. E.g. for target value $A_{i,j}(t)$, the features are $A_{j,i}(t-1)$, $A_{j,i}(t-2)$, $A_{j,i}(t-3)$.
- Average traffic of neighbors (3 previous hours). For target value $A_{i,j}(t)$:
  - $E[A_{i,k}(t-1)], E[A_{i,k}(t-2)], E[A_{i,k}(t-3)]$, where $k$ is a neighbor of $j$.
  - $E[A_{k,j}(t-1)], E[A_{k,j}(t-2)], E[A_{k,j}(t-3)]$, where $k$ is a neighbor of $i$.
- Standard deviation of neighboring traffic (3 previous hours). For target value $A_{i,j}(t)$:
  - $\sqrt{Var[A_{i,k}(t-1)]}, \sqrt{Var[A_{i,k}(t-2)]}, \sqrt{Var[A_{i,k}(t-3)]}$, where $k \in$ neighborhood($j$).
  - $\sqrt{Var[A_{k,j}(t-1)]}, \sqrt{Var[A_{k,j}(t-2)]}, \sqrt{Var[A_{k,j}(t-3)]}$, where $k \in$ neighborhood($i$).

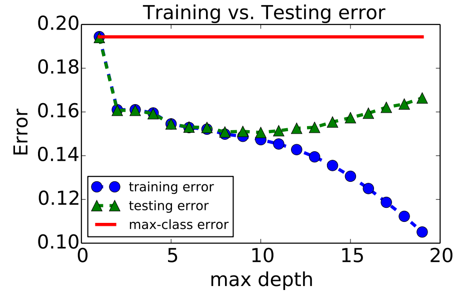
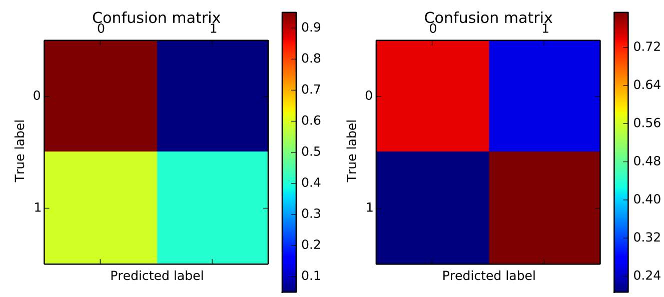



Fig. 7. Error on training vs. testing data for a decision tree.

| Scores | Max-class *predictor* | Decision Tree | Random Forest |
|---|---|---|---|
| Accuracy | 80% | 84% | 85% |
| Precision | - | 72% | 68% |
| Recall | - | 37% | 40 % |
| F1-score | - | 48% | 51% |

TABLE I
SCORES FOR MAX-CLASS PREDICTOR, DECISION TREE CLASSIFIER AND RANDOM FOREST.




(a) Classes equally important (b) Class 1 more important

Fig. 8. Normalized Confusion Matrix. Each square of the matrix has been normalized based on the true class, e.g. the square at (0,0) and (0,1) have been divided with the size of class 0, and the square at (1,0) and (1,1) have been divided by the size of class 1.In Fig. 8(a) when both classes have the same importance, class 0 is accurately predicted 96% of the times, but class 1 is predicted correctly only 41% of the times. In Fig. 8(b), where class 1 is considered more important than class 0, then the accuracy for class 0 dropped down to 74%, but accuracy for class 1 increased to 79% of the times.

The latest set of features (average traffic of neighbors, and standard deviation of neighboring traffic) were inspired by our analysis that showed that there is higher correlation among 2-hop neighbors (key insight (2) from the previous section).

## VI. RESULTS

We elected to use a tree classifier, since tree classifiers are powerful tools that can learn complex functions. We tuned the classifier's parameters and made sure that we don't overfit by applying standard complexity control techniques (see Fig. 7).

After tuning the classifier we apply it on our testing data and we analyze the initial results (see Tab. I). Since we are dealing with a highly skewed distribution (class 0 dominates our data set) we will compare against the majority predictor –

| Scores | weight = 3 | weight = 4 | weight = 10 |
|---|---|---|---|
| Accuracy | 79% | 75% | 61% |
| Precision | 47% | 42% | 33% |
| **Recall** | **71%** | **79%** | **94%** |
| F1-score | 57% | 55% | 49% |

TABLE II
RESULTS FOR DECISION TREE WITH WEIGHTED SAMPLES. WE CAN IMPROVE THE RECALL BY GIVING HIGHER WEIGHT TO POSITIVE SAMPLES, IN EXPENSE OF PRECISION.

a naive predictor that always predicts the most frequent class. Based on the accuracy, we see that our model outperforms the majority class predictor. However, when we dive into more details, namely we look at precision[5], recall[6], and F1-score – the harmonic mean of precision and recall – we see that due to the highly skewed distribution the recall is very low (37%). This is also confirmed by the confusion matrix in Fig. 8(a).

We applied a Random Forest of 200 ensembles (larger numbers did not show improvement). Because the ensemble averaging will avoid overfitting, we also increased the maximum depth of each tree to 30. The random forests classifier improved the results, e.g. recall increased to 40%, and the F1-sore reached 51%.

### A. Improving Recall

Up to now we have made the assumption that $F_p$ and $F_n$ have the same cost. However, this may not be the case. Providers would prefer having a high recall than high accuracy or precision ($F_n$ will have a higher cost ).

The same as when detecting patients with cancer, $F_n$ has a much higher cost. In this case a naive classifier that predicts always zero – a patient doesn't have cancer – will be very accurate, due to the skewness of the two classes, but that's not necessary the best classifier.

Therefore, in this last section we will investigate how to improve recall, even if that means sacrificing accuracy. This is achieved by increasing the weight of positive samples. A higher weight on positive samples forces the decision tree to pay more attention to class 1 than class 0. Tab. II show the improvement of recall given different weights, e.g. for weight of 3 – positive samples are 3 times more important than negative ones – recall rises to 79% (from 37%), and for a weight of 10 – positive samples are 10 times more important than negative ones – recall is 94%. This change is also reflected in the confusion matrix (see Fig. 8(b)).

## VII. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we applied machine learning techniques to predict cell-to-cell activity, based solely on past cellular activity records. We were able to achieve 85% accuracy and 94% recall, for the voice call data set provided by Telecom Italia for the city of Milan.

In future work, we will further improve the prediction by exploiting information outside the cellular activity data set, such as similarities between cells based on the socio-economic activity occurring in the surrounding areas. We could also extend the problem formulation (for example, instead of binary traffic prediction, we could predict multiple classes of traffic, such as no traffic, low traffic and high traffic between cells) or apply the methodology to other data sets (e.g., other cities or data instead of voice activity). Finally, we will investigate the use of this prediction methodology as a building block for network planning and control, urban ecology and smart city applications.

[5] $\frac{T_p}{T_p+F_p}$, where $T_p$ is the true positive rate, and $F_p$ the false positive rate.

[6] $\frac{T_p}{T_p+F_n}$, where $T_p$ is the true positive rate, and $F_n$ the false negative rate.

## REFERENCES


[1] 2013 Italian social protests. http://en.wikipedia.org/wiki/2013_Italian_social_protests, 2013.

[2] H. Ali-Ahmad, C. Cicconetti, A. La Oliva, V. Mancuso, M. Reddy Sama, P. Seite, and S. Shanmugalingam. An SDN-Based Network Architecture for Extremely Dense Wireless Networks. In *IEEE Conf. for Future Networks and Services (SDN4FNS)*, pages 1–7, Nov. 2013.

[3] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed. Femtocells: Past, Present, and Future. *IEEE Journal on Selected Areas in Communications*, 30(2):497–508, 2012.

[4] R. Cavalcante, S. Stanczak, M. Schubert, A. Eisenblaetter, and U. Tuerke. Toward Energy-Efficient 5G Wireless Communications Technologies: Tools for decoupling the scaling of networks from the growth of operating power. *IEEE Signal Processing Magazine*, 31(6):24–34, 2014.

[5] B. Cici, M. Gjoka, A. Markopoulou, and C. T. Butts. On the Decomposition of Cell Phone Activity Patterns and their Connection with Urban Ecology. In *Proc. ACM MobiHoc 2015*, Hangzhou, China, June 2015.

[6] Cisco. Visual Networking Index: Global Mobile Data Traffic Forecast Update 2015–2020 White Paper. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html, Feb. 2016.

[7] S. Dawoud, A. Uzun, S. Gondor, and A. Kupper. Optimizing the Power Consumption of Mobile Networks Based on Traffic Prediction. In *IEEE 38th Annual Computer Software and Applications Conf. (COMPSAC)*, pages 279–288. IEEE, 2014.

[8] R. Li, Z. Zhao, X. Zhou, J. Palicot, and H. Zhang. The Prediction Analysis of Cellular Radio Access Network Traffic: From Entropy Theory to Networking Practice. *Communications Magazine, IEEE*, 52(6):234–240, 2014.

[9] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla. New Perspectives and Methods in Link Prediction. In *Proc. of the 16th ACM Int. Conf. on Knowledge discovery and data mining (SIGKDD)*, pages 243–252, 2010.

[10] D. López-Pérez, A. Ladányi, A. Jüttner, H. Rivano, and J. Zhang. Optimization Method for the Joint Allocation of Modulation Schemes, Coding Rates, Resource Blocks and Power in Self-Organizing LTE Networks. In *Proc. IEEE INFOCOM.*, pages 111–115, Apr.

[11] G. Magdalena Nohrborg. Self-Organizing Networks. http://www.3gpp.org/technologies/keywords-acronyms/105-son, 2015.

[12] S. Morosi, P. Piunti, and E. Del Re. A Forecasting Driven Technique Enabling Power Saving in LTE Cellular Networks. In *Wireless and Mobile Computing, Networking and Communications (WiMob), 2013 IEEE 9th Int. Conf. on*, pages 217–222, 2013.

[13] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das. Understanding Traffic Dynamics in Cellular Data Networks. In *Proc. IEEE INFOCOM*, pages 882–890, 2011.

[14] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das. Understanding Spatial Relationships in Resource Usage in Cellular Data Networks. In *Computer Communications Workshops (INFOCOM WKSHPS), 2012 IEEE Conference on*, pages 244–249. IEEE, 2012.

[15] Qualcomm Technologies. 1000x Data Challenge. https://www.qualcomm.com/invention/technologies/1000x, 2015.

[16] F. Soldo, A. Le, and A. Markopoulou. Predictive Blacklisting as an Implicit Recommendation System. In *Proc. IEEE INFOCOM*, 2010.

[17] Statista. Number of mobile phone users worldwide from 2013 to 2019 (in billions). http://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/.

[18] Telecom Italia. Big Data Challenge. http://www.telecomitalia.com/tit/en/bigdatachallenge.html, 2014.

[19] D. Tikunov and T. Nishimura. Traffic Prediction for Mobile Network Using Holt-Winter's Exponential Smoothing. In *2007 15th Int. Conf. on Software, Telecommunications and Computer Networks*.

[20] H. Zang and J. Bolot. Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study. In *Proc. of the 17th ACM Int. Conf. on Mobile computing and networking (MobiCom)*, pages 145–156. ACM, Apr. 2011.