Understanding the WiFi Usage of University Students

Alessandro E. Redondi, Matteo Cesana, Daniel M. Weibel
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano
Milan, Italy
Email: {name.surname}@polimi.it

Emma Fitzgerald
Department of Electrical and Information Technology
Lund University
SE-221 00 Lund, Sweden
Email: emma.fitzgerald@eit.lth.se

Abstract—**In this work, we analyze the use of a WiFi network deployed in a large-scale technical university. To this extent, we leverage three weeks of WiFi traffic data logs and characterize the spatio-temporal correlation of the traffic at different granularities (each individual access point, groups of access points, entire network). The spatial correlation of traffic across nearby access points is also assessed. Then, we search for distinctive fingerprints left on the WiFi traffic by different situations/conditions; namely, we answer the following questions:** *Do students attending a lecture use the wireless network in a different way than students not attending a lecture?*, **and** *Is there any difference in the usage of the wireless network during architecture or engineering classes?* **A supervised learning approach based on Quadratic Discriminant Analysis (QDA) is used to classify empty vs. occupied rooms and engineering vs. architecture lectures using only WiFi traffic logs with promising results.**

Index Terms—**WiFi data analysis, user behaviour analysis,**

I. INTRODUCTION

Wireless local area networks (WLANs) based on the IEEE 802.11 standard family (i.e., WiFi) are an essential building block to provide widespread wireless connectivity in diverse indoor/outdoor scenarios. The cities we live in, our workplaces, hospitals and other public and private buildings are equipped with WiFi networks to provide hot spot or capillary wireless connectivity.

The reasons for the success of WiFi range from the use of unlicensed spectrum to their ease of deployment and management [1]. Nowadays, WiFi connectivity is available both in users' terminals (laptops, handheld devices, electronic gadgets) as well as in embedded devices and appliances. To this extent, WiFi plays a key role in the provision of connectivity in urban environments and to fully realize the vision of smart cities.

Besides their primary role of providing connectivity, WiFi networks and devices nowadays come with powerful monitoring systems able to collect and store large quantities of data on the behaviour of the network itself: traffic load, number of users, quality of the wireless signal etc. Such data, which is primarily used for network management, optimization, and fine-tuning, is also a "goldmine" for offering byproduct services; indeed, WiFi logs can be used to provide location-based services by properly localizing the users and/or to estimate flows and spatial distributions of people during events or at shopping malls. Moreover, WiFi logs can be coupled with other types of context data and, more generally, can be used to assess the behaviour of users. Therefore, a careful analysis of such a data provides valuable information that can be used for several purposes. In particular, in the context of smart cities and smart buildings, the availability of techniques for extracting high-level information from network data may help city and building administrators to better understand and react to the citizens' needs.

In this work, we focus on a particular type of building present in many cities in the world, namely a university campus building. We analyze the data coming from the local WiFi network of Politecnico di Milano, a large-scale technical university located in Italy and we provide the following contributions: (i) we propose a temporal and spatial characterization of the WiFi traffic; (ii) we leverage the WiFi traffic logs to answer the following questions: *Do students attending a lecture use the wireless network in a different way than students not attending a lecture?*, and *Is there any difference in the usage of the wireless network during architecture or engineering classes?* To answer these two questions we first propose a set of features which combine different attributes of the WiFi traffic (number of users/devices associated to the network, duration of such connections, etc.). Then we build up a labeled data set by exploiting information from the facility management department database which allows us to know if a room is occupied by a given lecture in a given time slot. Finally we propose a supervised learning approach based on Quadratic Discriminant Analysis (QDA) to classify empty vs. occupied rooms and engineering vs. architecture lectures by observing related WiFi traffic.

This work is organized as follows: Section II reviews the relevant literature on WiFi network data analysis; the reference scenario and spatio-temporal analysis of the WiFi traffic is presented in Section III. Section IV describes the supervised learning approach to classify empty vs occupied rooms and engineering vs architecture lectures. Section V concludes the manuscript and comments on ongoing/future work.

II. RELATED WORK

The analysis of data records extracted from WiFi networks has attracted much attention in the scientific community over

the last decade. The relevant work in the field can be classified by the main target of the data analysis.

The first class focuses on general performance analysis and characterization of WiFi-based wireless networks with the goal of medium/long-term network optimization. In their seminal work [2], Kotz and Essien focus on the analysis of a WiFi campus network composed of 476 Access Points. The reference data set spans a two-month period and is composed of traffic- and association-related information collected through SNMP polling and SYS log messaging. The *active* data collection is further complemented by data collected by passive sniffers to capture back-end traffic. The collected data set is leveraged to perform a rather complete analysis on the traffic load characteristics (per user, per access point traffic, traffic variability over time, per building traffic), the traffic type characteristics (traffic breakdown per application) and user mobility (number of visited Access Points while associated). A similar analysis is performed on the same WiFi network in [3] after two years to assess the changes over time of the aforementioned performance figures. A performance analysis of a WiFi campus network is also targeted in [4].

Calabrese *et al.* use in [5] the wifi data collected within the MIT WiFi network (3000 access points) to perform spatio-temporal analysis of the traffic flowing through the access points. Moreover, the information on the number of connected users per access point over time is used to classify location-dependent network behavior; namely, the authors show that by applying standard clustering techniques, location-dependent "fingerprints" can be determined for the network behavior (number of users, traffic).

The analysis of a corporate WiFi network is addressed in [6]; the authors collect a one-month dataset by polling every 5 minutes via SNMP 177 Access Points over three distinct corporate buildings. Besides analyzing the traffic characteristics, the authors also propose a clustering approach for the users based on two features: the *prevalence* which accounts for the time a user spends at an access point and the *persistence* which captures the total consecutive time a user spends at a AP.

The analysis of outdoor commercial WiFi networks is addressed in [7], [8] and [9]. In [7], Blinn *et al.* consider a Verizon WiFi Hot Spot network composed of 312 access points running IEEE 802.11b which are polled via SNMP every 5 minutes to obtain traffic and load measures. The proposed analysis is mainly targeted to characterize the access point traffic distribution over time, further assessing the spatial correlation existing among adjacent access points. A user mobility analysis is also carried out by leveraging the same clustering framework proposed in [6]. Afanasyev *et al.* [8], [9] address the performance analysis of the Google WiFi network in Mountain View (CA), composed of 500 access points. Different from the previous works, the reference wireless network architecture also has a wireless mesh extension interconnecting the access points. The proposed analysis has three main contributions: (i) the characterization of per-user traffic distributions, (ii) the classification of users in terms of their pattern in network usage and generated traffic (sporadic users, residential users, etc.), (iii) the assessment of users mobility in terms of travelled distance distributions.

Ganji *et al.* leverage in [10] campus WiFi data collected through SNMP polling and SYSlog messaging to evaluate the potential savings of duty cycling management policies of the access points in large-scale campus wireless networks.

A second class of work explicitly targets the assessment of human mobility and the development of data-driven mobility models [11]. Song *et al.* takes as reference the same data logs used in [2] to build up and validate location predictors for WiFi users of campus networks; two approaches are introduced based on Markovian models and Lempel-Ziv predictors. Along the same lines, in [12] and subsequent studies [13], [14], Kim *et al.* introduce and evaluate mobility models for campus WiFi users which are developed and trained on WiFi data logs.

Mobility analysis in hospitals is the main focus in [15] and [16], where movements of people are estimated by leveraging the data coming from a WiFi network deployment composed of 798 access points at Aarhus hospital. Besides characterizing the users mobility, the authors also propose a clustering analysis to classify different types of users (medical staff, visitors, patients, etc.).

The third and last class of works include the literature where WiFi data logs are mainly leveraged to classify the type of users. In [17], a 12-week trace of a wireless LAN is analyzed to analyze the behavior of different users in terms of mobliity, peak throughput rates and application mix. In [18] four months of WiFi data logs are collected in a campus environment and used to characterize the different behavior in terms of mobility between smartphones and laptops. Gember *et al.* [19] extend the comparative analysis between handheld and non-handheld devices by characterizing the different users behaviors in terms of type of generated traffic; namely, full packet data collected in a WiFi campus environment are collected and analyzed to assess the applications/traffic breakdown for handheld and non-handheld devices.

Recently, Wei *et al.* analyzed in [20] the impact of the high penetration of handheld devices in WiFi campus networks. A one year-long DHCP log is used along with a one-month flow level dataset to profile handheld and non-handheld devices in terms of mobility behavior and traffic characteristics.

According to this characterization, this paper stands in the third class.

III. SCENARIO

This work analyses network data logs extracted from the wireless network of the architecture department building of the Politecnico di Milano (PoliMi) university, located in Milan, Italy. The wireless network under study is composed of 28 different access points (APs) located on four different floors of the building and covering rooms devoted to lectures as well as offices, corridors and other public spaces. The APs support the AirWave Management Platform (AMP), that allows to observe every device and user connected to the network. In particular, the AMP allows to sample the bandwidth usage and the number of clients connected for each APs every 5 minutes.
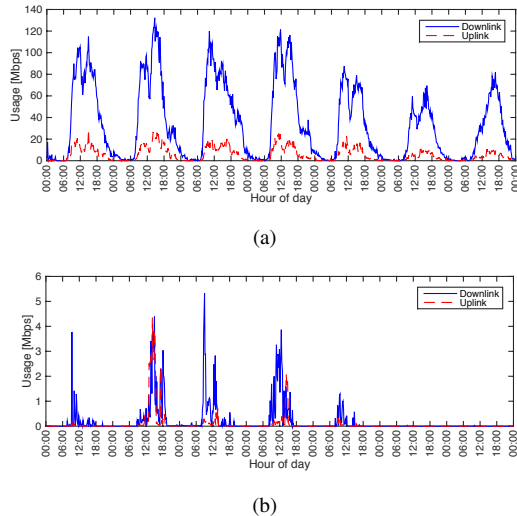
(a)

(b)

Fig. 1. (a) Cumulative downlink/uplink usage over one week; (b) Downlink/uplink usage of one particular AP over the same week

Fig. 2. Temporal correlation averaged over all APs. Local maxima are visible after 7 and 14 days, according to the academic calendar.

In this work, we focus on a period of three weeks, from the 16th of November, 2015 to the 6th of December, 2015. This period does not contain any national holiday, therefore it represents well the "steady-state" behaviour of the wireless network in terms of number of clients connected and average usage. For each access point, the uplink/downlink bandwidth usage and the number of clients connected were sampled and stored every 5 minutes, for a total of $12 \times 24 \times 21 = 6048$ data points per AP. Additionally, the list of devices associated to each AP is downloaded and stored in a database. For each device, the following information is stored and updated every 5 minutes: device MAC address, timestamp of the association with the AP, duration of the connection (i.e., time elapsed from the association event), average and variance of the signal quality of the connection [dB] as well as average and variance of the bandwidth usage [kbps]. Over the three weeks, a total of 27538 unique devices were observed, generating 300681 different association events.

Figure 1(a) shows the cumulative uplink/downlink bandwidth usage summed over all the APs during one of the selected three weeks. As one can see, such cumulative measures show a very nice periodic behaviour: the network usage is very low during the night, increases rapidly in the morning, experiences a short decline during lunch break and fades out during the afternoon with a short constant period of about two hours during the evening. This pattern is clearly associated with the behaviour of students and employees working in the building under consideration, and repeats unchanged every day from Monday to Friday. The behaviour during the weekend is similar to the one during work days, although with fewer clients connected and therefore less bandwidth usage. The same behaviour is obtained for the other two weeks in the dataset, which are not shown here for space reasons. Unfortunately, the nice periodic behaviour observed on the cumulative data is no more visible when analysing the data
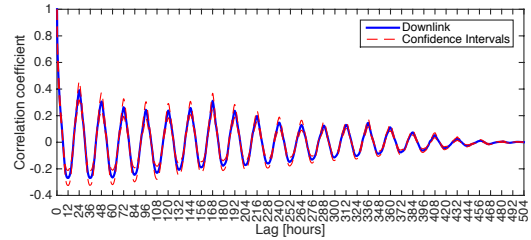
of a particular AP. As an example, Figure 1(b) shows the downlink and uplink usage measured on an AP located in a room used for lectures throughout the week. As one can see, the bandwidth usage is different from day to day and no periodic pattern (except from the trivial night/day differences) may be found. Such consideration, that periodicity is lost when going from cumulative to individual AP data, is similar to what was found in [21] for cellular networks.

In order to provide a more detailed analysis of the behaviour of the network, we investigate how the network load varies both spatially and temporally.

A. Temporal correlation analysis

As a first step, we compute the sample Pearson correlation coefficient $r_i(t)$ of the downlink/uplink usage of the i-th AP with a version of itself delayed by t samples (the so called *lagged* correlation). This is done to determine if some temporal pattern exists in individual AP data. Such information could have tremendous implication in network resource management techniques. Figure 2 shows the average lagged correlation coefficient for the downlink bandwidth usage, where the average is taken over all APs. As one can see, the average temporal correlation is low, with peaks occurring every 12 and 24 hours according to the diurnal human activity pattern. It is interesting to note that the correlation coefficient has local maxima in correspondence to lags 168 and 336, that is after one and two weeks. This can be explained taking into account the schedule of lectures in some of the rooms covered by the APs, which has a periodicity of one week. The 95% confidence interval (red dashed line) is very narrow, meaning that all APs show similar correlation coefficients. We repeated the same test for the uplink traffic (not shown for space reasons), finding weaker temporal correlation although with the same periodic pattern.

B. Spatial correlation analysis

We also analyse the spatial correlation between different APs. Again, such information could be used for mobility modeling, traffic prediction and deployment planning. We compute the Pearson correlation coefficient $r_{i,j}$ for all possible pairs of APs (i, j). Figure 3 shows the computed correlation coefficients for the downlink and uplink traffic, where blue indicates no correlation ($r_{i,j} = 0$) and yellow indicates maximum correlation ($r_{i,j} = 1$). The four red squares indicate correlation values of APs located on the same floor (basement,
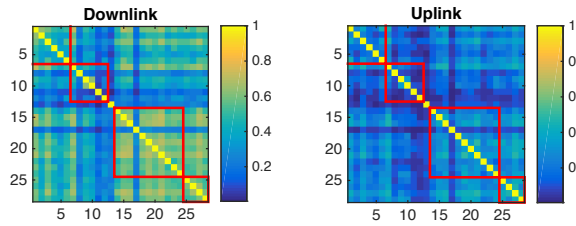
Fig. 3. Spatial correlation among different APs.

Fig. 4. Segmentation of network data into different time slots according to the academic calendar.

ground, first and second floor, from the top left) and the AP with index 13 is located alone on a mezzanine. Some interesting observations can be made from the inspection of Figure 3:

1) The maximum correlation is observed among APs on the first and second floor, which are also those where most of the classrooms are located. Overall, the maximum spatial correlation value for the downlink traffic is 0.7, while it is 0.5 for the uplink case (again indicating a weaker correlation in the uplink usage).
2) Conversely, APs located in the basement and on the ground floor, exhibit weak spatial correlation.
3) Some particular APs (rows 8, 11, 13 and 17) show very low correlation values with all other APs. A more thorough analysis revealed that these APs show the lowest uplink/downlink traffic usage and the lowest average signal quality towards connected devices. This may suggest that either they are malfunctioning or that their location may be changed to improve their performance.

IV. CLASSIFICATION

The wireless network under study is located in a building belonging to the Department of Architecture of PoliMi university. This building contains several classrooms that are used for teaching classes from both the Architecture and Engineering courses of study. Each classroom is different in size (maximum number of students it may host) and is used for lectures only in specific time slots during each day, according to the academic calendar. When a room does not host a lecture, it may be still used by students for studying or just for passing time between a lecture and the following one.

In such a scenario, we rely on the data available AirWave Management Platform and corresponding to those APs located in the classrooms to answer two different questions:

1) Is there a difference between the WiFI usage inside rooms during lecture times and "idle" times of these rooms?
2) Is there a difference between the WiFi usage inside rooms during architecture lectures and engineering lectures in these rooms?

The answer to the first question is somehow expected as positive, as it seems logical to assume that students attending a class will pay more attention to the teacher rather than to their smartphones or laptops. At the same time, a user studying or passing time in an empty room may be using its wireless connection actively. Conversely, the answer to the second 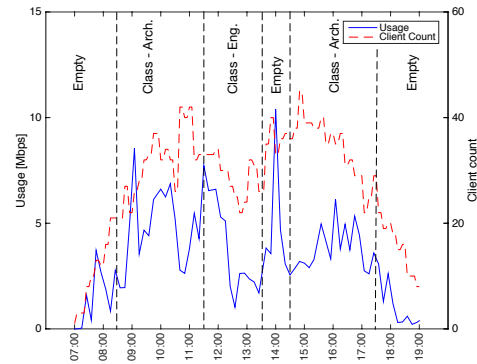question seems more unpredictable: although engineering and architecture students tend to consider themselves as different "species", the way they behave during classes in terms of network usage may be the same.

A. Feature extraction

To answer these questions, we take the following approach. First, we rely on the academic calendar to identify when and in which classroom architecture and engineering lectures are given. This allows to extract several "*time slots*" from the data logs of the APs located in the rooms where lectures are given, where each time slot corresponds to either a particular lecture (from the architecture or engineering courses) or to an empty slot (see Figure 4). Note that we do not consider empty slots occurring during night hours, but only those daytime periods in which a room is not used for a lecture. Therefore, we only consider time slots falling between 7 AM and 7 PM. Over the three weeks under study, we extracted 213 non-empty time slots (156 corresponding to architecture lectures and 57 to engineering lectures) and 101 empty time slots.

For each time slot, we extract the following features:

- *Total number of connections:* the total number of connections whose start time falls inside the time slot.
- *Number of unique devices:* the number of unique MAC addresses that started a connection in the time slot.
- *Number of connections per device:* the ratio between the total number of connections and the number of unique MAC addresses seen in the time slot, or how many times a particular device connected to the network in the time slot.
- *Average normalised duration:* the average duration of each connection, divided by the duration of the time slot.
- *Average and variance of bandwidth usage:* the mean and variance of the bandwidth usage of each connection, averaged over all the connections in the time slots. These features capture how much traffic is produced by students in a particular time slot and how variant this traffic is on average.
- *Average and variance of signal quality:* the mean and variance of the signal quality experienced by each connection, averaged over all the connections in the time slots.
- *Occupancy:* the ratio between the number of unique devices seen during the time slot divided by the seating capacity of
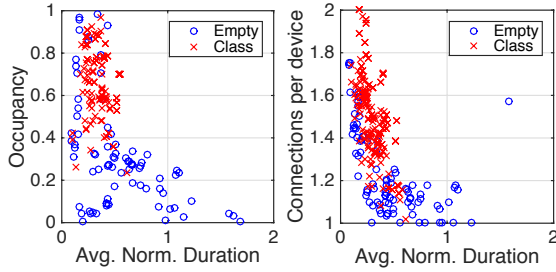
Fig. 5. The set of 213 non-empty time slots and 101 empty time slots plotted versus occupancy, average normalised duration and number of connections per device.

the room in which the particular AP is installed.

- *Normalised occupancy:* the occupancy value normalised with respect to the duration of the time slot.
- *Connection distribution sparsity and peak:* for each time slot, we observe the particular minute at which each connection is started and we build an histogram with bins spaced every five minutes, for a total of 12 bins. We then compute the distribution sparsity and peak as the number of empty bins and as the index of the bin with the highest connection count, respectively. Such features capture the distribution of the connections starting instants. If such a distribution is sparse, it means that many devices started a connection in the same 5 minutes (something that typically happens during lecture breaks). If the distribution is uniform rather than sparse, it means that students tend to start connection randomly during the time slot and may indicate that they are not particularly "interested" or "focused" on the lecture.

Ideally, such features should be computed only for the connections started by devices who are actually located inside a classroom where a lecture is (or is not) given. Therefore, in the computation of the features we consider only those connections whose average signal quality during the selected time slot is greater than the average signal quality of all possible connections to the AP covering that classroom.

B. Rooms with classes VS. empty rooms

To analyse the differences between the behaviour of students during classes and empty slots, we use a classification-based approach. First, we partition the dataset into a training-set (80% of the time slots) and a test-set (the remaining 20%). The training set is used to train different classifiers based on (i) logistic regression, (ii) Linear Discriminant Analysis (LDA) and (iii) Quadratic Discriminant Analysis (QDA), whose performance are evaluated on the test set. The same process is repeated 10 times, according to a stratified k-fold cross validation approach (i.e., maintaining the proportion of the different classes in each fold). Among the three classifiers, QDA exhibits the best performance and only its results are shown in the following for the sake of space.

Table I shows the performance of QDA in terms of a confusion matrix. As one can see, time slots in which a class was given are correctly classified 98% of the time, while this

TABLE I
CONFUSION MATRIX FOR THE EMPTY VS CLASS ROOM CLASSIFICATION

	Predicted: Empty	Predicted: Class
Actual: Empty	83.17%	16.83%
Actual: Class	1.88%	98.12%

accuracy drops to 83% for empty time slots. Overall, the error rate (that is, the percentage of misclassified time slots) of the QDA classifier is as low as 8% and its F1-score is as high as 0.87. To further analyse what are the most discriminative features for such a classification, at each iteration of the k-fold validation we perform forward stepwise feature selection, each time keeping track of the top three selected features. It turns out that the three most discriminative features in this case are the *average normalised duration*, the *occupancy* and the *number of connections per device*. Figure 5 gives a visual explanation of the difference between empty and non-empty time slots: empty time slots are characterised by a lower average occupancy (0.6 vs. 0.9), and exhibit fewer connections per device (1.2 vs. 1.5) but with a longer normalized duration (0.6 vs. 0.3). That means that students occupying empty rooms tend to connect once to the wireless network and to maintain such a connection for a long time, while students attending a class connect more frequently but for shorter periods.

C. Engineering VS. architecture classes

We repeated the same approach for identifying differences between the behaviour of students during engineering or architecture courses. Table II shows the performance of QDA in terms of a confusion matrix. As one can see, even in this case the performance of classification is very good, with an overall error rate as low as 6% (F1-score equal to 0.96). This means that there is indeed a difference in the usage of the network by architecture or engineering students and that the classifier is able to exploit such a difference. To better understand what changes between the behaviour of the two types of students, we again select the top three features obtained with stepwise forward selection, which turn out to be the *occupancy*, the *sparsity of the connection distribution* and the *average normalised duration*. Figure 6 shows the two types of classes when plotted on the planes individuated by such features. We can observe that engineering classes have on average (i) a lower occupancy (0.6 vs 0.8), (ii) a higher sparsity of the connection distribution (2.9 vs 1.9) and (iii) a higher normalised duration with respect to architecture students (0.35 vs 0.2). We do not have access to the exact number of enrolled students per class: assuming that such a number is proportional to the room capacity, our results seem to indicate that engineering students attend classes less frequently, but with more attention than architecture students. As a final remark, note that the methodology used in this work can be applied also in all those scenarios following some kind of schedule e.g. meeting rooms, conferences, train stations etc., not just university classes.
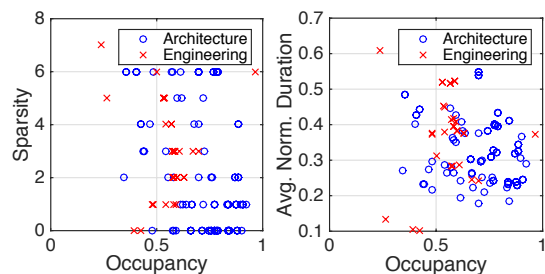
Fig. 6. The set of 156 architecture lectures and 57 engineering lectures plotted versus occupancy, normalised duration and sparsity of the connections distribution.

TABLE II
ARCHITECTURE VS ENGINEERING LECTURES CLASSIFICATION

	Predicted: Architecture	Predicted: Engineering
Actual: Architecture	97.44%	2.56%
Actual: Engineering	14.04%	85.96%

V. CONCLUSIONS

The analysis of WiFi traffic can give fundamental insights on how to optimize and manage the network andit can also reveal patterns on how the end users behave. This high-level semantic information may be used as a basis for implementing future services, especially in the context of smart cities. In this work, we have analyzed traffic logs from a campus WiFi network. First, we have presented a spatio-temporal correlation analysis of the network under consideration. Then, we have proposed a supervised learning approach based on QDA to classify WiFi traffic coming from empty or occupied classrooms and from engineering or architecture lectures. The proposed approach has been tested on a three-weeks WiFi data log with promising classification performance. Future works will address the exploitation of such results to implement high level services.

ACKNOWLEDGEMENTS

REFERENCES

[1] P. Serrano, P. Salvador, V. Mancuso, and Y. Grunenberger, "Experimenting with commodity 802.11 hardware: Overview and future directions," *Communications Surveys Tutorials, IEEE*, vol. 17, no. 2, pp. 671–699, Secondquarter 2015.

[2] D. Kotz and K. Essien, "Analysis of a campus-wide wireless network," in *Proceedings of the Eighth Annual Intl. Conf. on Mobile Computing and Networking (MobiCom)*, September 2002, pp. 107–118. [Online]. Available: http://www.cs.dartmouth.edu/ dfk/papers/kotz-campus.pdf

[3] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Proceedings of the Tenth Annual Intl. Conf. on Mobile Computing and Networking (MobiCom)*. ACM Press, September 2004, pp. 187–201. [Online]. Available: http://www.cs.dartmouth.edu/ dfk/papers/henderson-voice.pdf

[4] E. Zola and F. Barcelo-Arroyo, "A comparative analysis of the user behavior in academic wifi networks," in *Proceedings of the 6th ACM Workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Networks*, ser. PM2HW2N '11. New York, NY, USA: ACM, 2011, pp. 59–66. [Online]. Available: http://doi.acm.org/10.1145/2069087.2069096

[5] F. Calabrese, J. Reades, and C. Ratti, "Eigenplaces: Segmenting space through digital signatures," *Pervasive Computing, IEEE*, vol. 9, no. 1, pp. 78–84, Jan 2010.

[6] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proceedings of the 1st Intl. Conf. on Mobile Systems, Applications and Services*, ser. MobiSys '03. New York, NY, USA: ACM, 2003, pp. 303–316. [Online]. Available: http://doi.acm.org/10.1145/1066116.1066127

[7] D. P. Blinn, T. Henderson, and D. Kotz, "Analysis of a Wi-Fi hotspot network," in *Proceedings of the Intl. Workshop on Wireless Traffic Measurements and Modeling (WiTMeMo '05)*. USENIX Association, June 2005, pp. 1–6. [Online]. Available: http://www.cs.dartmouth.edu/ dfk/papers/blinn-hotspot.pdf

[8] M. Afanasyev, T. Chen, G. Voelker, and A. Snoeren, "Usage patterns in an urban wifi network," *Networking, IEEE/ACM Trans. on*, vol. 18, no. 5, pp. 1359–1372, Oct 2010.

[9] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren, "Analysis of a mixed-use urban wifi network: When metropolitan becomes neapolitan," in *Proceedings of the 8th ACM SIGCOMM Conf. on Internet Measurement*, ser. IMC '08. New York, NY, USA: ACM, 2008, pp. 85–98. [Online]. Available: http://doi.acm.org/10.1145/1452520.1452531

[10] F. Ganji, Ł. Budzisz, F. G. Debele, N. Li, M. Meo, M. Ricca, Y. Zhang, and A. Wolisz, "Greening campus wlans: Energy-relevant usage and mobility patterns," *Computer Networks*, vol. 78, pp. 164 – 181, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128614004241

[11] J. Yoon, B. D. Noble, M. Liu, and M. Kim, "Building realistic mobility models from coarse-grained traces," in *Proceedings of the 4th Intl. Conf. on Mobile Systems, Applications and Services*, ser. MobiSys '06. New York, NY, USA: ACM, 2006, pp. 177–190. [Online]. Available: http://doi.acm.org/10.1145/1134680.1134699

[12] M. Kim and D. Kotz, "Modeling users' mobility among wifi access points," in *Papers Presented at the 2005 Workshop on Wireless Traffic Measurements and Modeling*, ser. WiTMeMo '05. Berkeley, CA, USA: USENIX Association, 2005, pp. 19–24. [Online]. Available: http://dl.acm.org/citation.cfm?id=1072430.1072434

[13] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *INFOCOM 2006. 25th IEEE Intl. Conf. on Computer Communications. Proceedings*, April 2006, pp. 1–13.

[14] J. Kim and A. Helmy, "Analysing the mobility, predictability and evolution of wlan users," *Intl. Journal of Autonomous and Adaptive Communications Systems*, vol. 7, no. 1-2, pp. 169–191, 2014.

[15] A. Ruiz-Ruiz, H. Blunck, T. Prentow, A. Stisen, and M. Kjaergaard, "Analysis methods for extracting knowledge from large-scale wifi monitoring to inform building facility planning," in *Pervasive Computing and Communications (PerCom), 2014 IEEE Intl. Conf. on*, March 2014, pp. 130–138.

[16] T. S. Prentow, A. J. Ruiz-Ruiz, H. Blunck, A. Stisen, and M. B. Kjærgaard, "Spatio-temporal facility utilization analysis from exhaustive wifi monitoring," *Pervasive and Mobile Computing*, vol. 16, Part B, pp. 305 – 316, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1574119214001953

[17] D. Tang and M. Baker, "Analysis of a local-area wireless network," in *Proceedings of the 6th Annual Intl. Conf. on Mobile Computing and Networking*, ser. MobiCom '00. New York, NY, USA: ACM, 2000, pp. 1–10. [Online]. Available: http://doi.acm.org/10.1145/345910.345912

[18] U. Kumar, J. Kim, and A. Helmy, "Changing patterns of mobile network (wlan) usage: Smart-phones vs. laptops," in *Wireless Communications and Mobile Computing Conf. (IWCMC), 2013 9th Intl.*, July 2013, pp. 1584–1589.

[19] A. Gember, A. Anand, and A. Akella, "A comparative study of handheld and non-handheld traffic in campus wi-fi networks," in *Proceedings of the 12th Intl. Conf. on Passive and Active Measurement*, ser. PAM'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 173–183. [Online]. Available: http://dl.acm.org/citation.cfm?id=1987510.1987528

[20] X. Wei, N. Valler, H. Madhyastha, I. Neamtiu, and M. Faloutsos, "A behavior-aware profiling of handheld devices," in *Computer Communications (INFOCOM), 2015 IEEE Conf. on*, April 2015, pp. 846–854.

[21] U. Paul, A. Subramanian, M. Buddhikot, and S. Das, "Understanding traffic dynamics in cellular data networks," in *INFOCOM, 2011 Proceedings IEEE*, April 2011, pp. 882–890.