Learning the Relation Between Mobile Encounters and Web Traffic Patterns: A Data-driven Study

Babak Alipour, Mimonah Al Qathrady, Ahmed Helmy

University of Florida

Gainesville, FL, USA

{babak.ap,mimonah,helmy}@ufl.edu

ABSTRACT

Mobility and network traffic have been traditionally studied separately. Their interaction is vital for generations of future mobile services and effective caching, but has not been studied in depth with real-world big data. In this paper, we characterize mobility encounters and study the correlation between encounters and web traffic profiles using large-scale datasets (30TB in size) of WiFi and NetFlow traces. The analysis quantifies these correlations for the first time, across spatio-temporal dimensions, for device types grouped into on-the-go *Flutes* and sit-to-use *Cellos*. The results consistently show a clear relation between mobility encounters and traffic across different buildings over multiple days, with encountered pairs showing higher traffic similarity than non-encountered pairs, and long encounters being associated with the highest similarity. We also investigate the feasibility of learning encounters through web traffic profiles, with implications for dissemination protocols, and contact tracing. This provides a compelling case to integrate both mobility and web traffic dimensions in future models, not only at an individual level, but also at *pairwise* and collective levels.

ACM Reference Format:

Babak Alipour, Mimonah Al Qathrady, Ahmed Helmy. 2018. Learning the Relation Between Mobile Encounters and Web Traffic Patterns: A Data-driven Study. In *21st ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '18)*, October 28-November 2, 2018, Montreal, QC, Canada. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3242102.3242137>

1 INTRODUCTION & RELATED WORK

The effect of mobility and network traffic on wireless networks has been established in the literature (e.g. [1]). Several efforts studied models of mobility and network traffic, albeit mostly separately and in isolation. For a summary, we refer the reader

to [2, 3] for surveys of mobility modeling and analysis. Some of the most advanced studies on mobility [4] have identified individual [5], pairwise (encounter), and collective (group) dimensions for mobility modeling. They did not consider traffic however. We hope to bridge that gap by analyzing the interplay of mobility and traffic at the pairwise level.

In this paper, we focus on the *pairwise (encounter)* dimension of mobility and study its interplay with the traffic patterns of mobile users. This has several applications such as realistic modeling and simulation, as well as design of encounter-based services.

Encounters between mobile nodes were studied in previous research (e.g. [6, 7]) to characterize opportunities of inter-user encounters. Others (e.g. [8]) mainly collect encounter traces using mobile devices to analyze and model communication opportunities. None of these studies analyze traffic nor the correlation between encounters and real-world traffic patterns. Several studies analyzed wireless traffic flows [9], and mobile web-visitation patterns [10]. These studies did not investigate the relation with mobility and node encounters.

In addition, many research studies on mobility encounters or traffic patterns did not consider *device type*. Devices' form factor affects mode of usage and varied traffic profiles ([11–13]). But these studies do not study the interplay of traffic with mobility and encounters. Devices are also used during different modes of transportation. Smartphones and e-readers, dubbed *Flutes*, are devices used 'on-the-go'. On the other hand, laptops are 'sit-to-use' devices and are referred to as *Cellos*. In our earlier work [14], we contrast *mobility* and *traffic* features of *Flutes* and *Cellos*, across the *individual* dimension. Here, we investigate the *pairwise* dimension, focusing on encounter-traffic interplay while considering the device types.

We use extensive traces from our collected datasets (76B records, ≈ 30 TB in size) covering ≈ 78 K devices, in 140 buildings on a university campus. The data includes information about WiFi associations, as well as DHCP and NetFlow traces, covering the dimensions of mobility and network traffic. The data is sanitized and categorized based on buildings, days, device types (*Flutes*, *Cellos*), and encounter duration, then the analysis is done across all these dimensions.

The main question addressed in this study is 'How do device encounters affect network traffic patterns, across time, space, device type and encounter duration?' For that purpose, we: i- analyze mobility encounters patterns, ii- define web traffic profiles for users, and iii- look at their interplay. These questions have not been directly studied in-depth before,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MSWiM '18, October 28-November 2, 2018, Montreal, QC, Canada
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-5960-3/18/10...\$15.00
<https://doi.org/10.1145/3242102.3242137>

with our findings showing that for the majority of buildings a consistent correlation exists between traffic profiles and encountered (vs. non-encountered) pairs of users. We also found the correlation to be the strongest for *Cello-Cello* encounters on weekends. Further, we find that such relation strengthens for long encounters, while short encounters are not significantly different from non-encountered pairs. Finally, we utilized a deep learning model to learn encounters of user pairs in a day and building based on their traffic profiles alone. The model achieved a high accuracy (90%+) in many settings, with major implications for encounter-based services, rumor anatomy analysis, and infection tracing.

In addition, mobility modeling and protocol evaluation could benefit from deep integration and interplay of encounters and traffic. We hope for this paper to encourage more studies on mobility and traffic, and their interplay, across individual, *pairwise* and collective dimensions, towards fully *integrated realistic traffic-mobility models*.

In the rest of this document, Sec. 2 describes the datasets. Sec. 3 analyzes mobility encounters. Sec. 4 introduces the web traffic profiles. Sec. 5 presents the *pairwise* encounter-traffic relationship. Next, Sec. 6 introduces our encounter learning methods and results. Finally, Section 7 discusses the findings and concludes the paper.

An extended tech report of this study is available online [15].

2 DATASETS

This study utilizes multi-sourced large-scale datasets we have collected including WLAN, DHCP, NetFlow, and other external sources.

2.1 Wireless LAN (WLAN) & Encounters

The WLAN logs were collected on a university campus during April 2012. Each entry provides a timestamp, an IP address at a corresponding access point (AP) and MAC address of the associated user device, for 1,700 APs and $\approx 78k$ devices. We analyze the device behavior, as identified by its MAC address¹. Pairwise user mobility behavioral patterns are represented through encounters between two mobile nodes. An encounter is defined as when two user devices are associated with the same AP at an overlapping time interval. The **Encounter** traces are generated based on WLAN logs.

2.2 Location Information

To analyze traces in different places, location information of APs is required. Since exact locations of the APs were not available, the APs are assigned approximate locations based on the building where they are installed, i.e. building latitude/longitude from Google Maps API. We used a crowd-sourced service to validate this positioning.

¹MAC address randomization was introduced on popular platforms after our traces were collected, and does not affect our association trace.

2.3 NetFlow

The Netflow traces were collected from the same network and time period². A flow is a unidirectional sequence of packets transferred between two IPs, with each flow record retaining over 30 fields, including flow start/finish times, source/destination IP addresses and ports, transport protocol and traffic volume (packets/bytes). In raw format, dataset size is $\approx 30TB$, providing a vast, high granularity data source. Due to quadratic asymptotic growth of *pairwise* analysis, for this study, we focus on the 10,000 most active users in terms of traffic consumption, to keep computations manageable.

2.4 Device Type Classification

To classify devices into '**on-the-go**' *Flutes* and '**sit-to-use**' *Cellos*, we build upon the same observations and heuristics of our previous work [14], using OUIs (MAC prefix) and analysis of mobile domains in NetFlow. Overall, over 97% of devices in NetFlow traces were labeled ($\approx 50K$ flutes and $\approx 27K$ cellos). This enables classification of pairwise encounters, based on the encounter pair device types: 1) **Flute-Flute (FF)**: encounter event between two flutes. 2) **Cello-Cello (CC)**: the pair are cello devices. 3) **Flute-Cello (FC)**: encounter event between a flute and cello.

3 MOBILITY ENCOUNTERS

Pairwise mobile encounter events provide opportunities for dissemination events such as content dissemination [7, 16] and infection spreading [17]. While encounters have been analyzed in several previous studies (e.g. [6, 7]), here we develop new insights into pairwise events by considering: 1) *Device types*: We distinguish between encounters among the three groups in our analyses (*FF*, *CC*, *FC*). 2) *Large-scale data*: The data is first of its kind in terms of its size, covering 140 buildings with different categories. Also, we analyze mainly indoor (in-building) encounters, unlike most previous studies. 3) *Traffic-encounter analysis*: Daily encounter patterns at buildings are analyzed per device type, then their correlation to traffic patterns are studied for the first time in Sec. 5.

3.1 Daily Encounter Duration at Buildings

The statistical summary of mobility encounters are generated from daily encounters at each building. The total encounter duration, E , of a pair of users $u1, u2$ during day d at building B , $E_{Bd}u1, u2$ is: $E_{Bd}u1, u2 = \text{Sum}_{i=1}^{i=n} E_{Bd}u1, u2_i$, where n is the number of encounters, and $E_{Bd}u1, u2_i$ is the duration of encounter i between $u1$ and $u2$ on day d at building B . If a pair encounters again on a different day or in a different building, that encounter is considered separately. Overall, $\approx 20\%$ have encountered at least twice in any building. The pairs are then separated based on device types.

CC pairs have longer encounter duration than others. For example, the mean *CC* daily encounter duration is **290% longer** than the *FF* pairs. This result is beneficial when

²The last five days of NetFlow cover exam dates and are omitted, since those dates do not represent a typical campus environment.

Figure 1: Encounter duration CDF based on encounter pair device type.

modeling the encounters based on device type, or with applications that use the encounter duration. Figure 1 shows the CDF of the encounter duration for 95% of pairs (the highest 5% omitted for clarity). Note that 80% of *FF* encounters have daily encounter duration ≤ 8 minutes, while only 40% of *CC* encounters are ≤ 8 minutes. For all encounters, 33% are ≤ 38 s, dubbed *short*, the next 33% are ≤ 317 s, called *medium*, and > 317 s are *long* encounters. This definition will be used in Sec. 5 for pairwise analysis of the correlation between encounter duration and traffic profile similarity.

3.2 Encounter Duration Statistical Distributions

Eleven distributions are fit to the total daily encounter duration using maximum likelihood, and goodness-of-fit test methods. The Kolmogorov-Smirnov (*KS*) statistic is used to evaluate the distributions fitness. The three best-fit distributions are presented in table 1. For example, 74% of the buildings have **power-law** (*PL*) distribution as the best fit for their *FF* pair daily encounter duration, while only 39% of buildings have *CC* daily encounter distribution following power-law. Also, table 1 shows the percentage of buildings with *KS*-statistic $\leq 5\%$ and $\leq 10\%$. This is calculated to see if there is a distribution that can be a good fit for the majority of buildings, even if it is not the first-best fit. Power-law and log-logistic distributions usually have *KS*-test with $\leq 10\%$ for 92% for *FF* and *FC* pairs.

4 WEB TRAFFIC PROFILE

We use *NetFlow* to analyze the traffic behavior [14]. To analyze traffic patterns for all buildings and days, we define a **Traffic Profile (TP)** for each user based on *NetFlow*, that is efficient to calculate and granular enough for our analysis:

- First, we select a set of popular websites based on total bytes sent and received, filtering out websites with little usage. The IPs of selected websites form the dimensions of the traffic vector, denoted as κ . There are $\approx 10,000$ IPs in κ , with average *daily* traffic from few *MBs* up to *690GBs*.
- Next, for each IP_j , we calculate $B_{i,j}$, the natural logarithm of total traffic user i , U_i , sent/received with IP_j . This forms the *traffic vector* for U_i , consisting of $B_{i,j}, \forall j \in \kappa$.
- Finally, we apply *term frequency-inverse document frequency* (**TF-IDF** [18]) to the collection of traffic vectors of all users. This reduces the effect of wildly popular websites,

and identifies websites that can distinguish between users' online behavior, enabling us to study the richness in the access patterns. In this context, each IP_j is a *term* and each user *traffic vector* is a *document*. **TF-IDF** is calculated as the product of term frequency (the number of times a term appears in a document, corresponding to $B_{i,j}$ in our context, which reflects the bytes IP_j exchanged with U_i), and inverse document frequency (the inverse of number of documents (*users*) the term (*IP*) occurs in) [19]. Each row of the resulting matrix is a *traffic profile*, TP_i , of user U_i .

This process is applied for *NetFlow* data of *every* building on *each day*, to enable spatial (across buildings) and temporal (across days) analysis of user traffic profiles³. For pairwise comparison of traffic profiles, we use Cosine similarity which computes the cosine of the angle between two user profiles.

5 PAIRWISE ENCOUNTER-TRAFFIC RELATIONSHIP

With mobility encounters and traffic profiles as the pillars, we investigate "*whether physical encounters are correlated with the similarity of traffic profiles*". This analysis outlines our initial findings in the *pairwise* (*encounter*) dimension of mobility-traffic analysis, following our work in [14] on the *individual* aspect, and providing the foundation for *collective* (*group*) analysis in the future. We start with simple steps, and increase the complexity of methods gradually.

As a first step, we seek to establish whether the traffic profiles (TP) of encountered pairs are more similar compared to TPs of non-encountered pairs. For this, we calculate $enc = \text{Sum}_{i,j \in Z} \frac{\text{cosine} TP_i, TP_j}{|Z|}$. Here, Z denotes the set of all encountered pairs. Similarly, for all non-encountered pairs, Z' , $nonenc = \text{Sum}_{i,j \in Z'} \frac{\text{cosine} TP_i, TP_j}{|Z'|}$, calculated for each building every day. Overall, we observe that **$\approx 93\%$ of the time $enc > nonenc$** , with the main exceptions being buildings close to bus stop hubs on campus, with a high pass-by rate of users; resulting in many short encounters that do not show traffic similarity. With that simple observation, next, we asked whether the difference between the similarities of encountered and non-encountered is statistically significant. We find that **in most cases, the traffic profiles of encountered pairs are more similar and the difference between the two groups is statistically significant**. Logistic regression analysis shows that similarity of traffic profiles is significantly associated with the probability of encountering for all days in several buildings, such as the computer department with a big user base, for $\approx 90\%$ of days in libraries, and only for $\approx 48\%$ of days in gym and recreation centers where users normally do not use networks as much. The next question is how consistent these differences are across:

- (1) Device type categories: As discussed earlier, usage patterns of devices differ based on form factor (e.g., on-the-go *flutes* vs. sit-to-use *cellos*). We compare flute-flute (*FF*), cello-cello (*CC*), and flute-cello (*FC*) encounters.

³If a building on a specific day has less than 20 encountered pairs, that (building, day) pair is omitted, to maintain statistical significance.

Table 1: Best fit distributions for total daily encounter pairs duration based on pairs classifications. Percentage of buildings shown in brackets. (Pl: Power Law, Ll: Log Logistic, W: Weibull, G: Gamma, Pr: Pareto, Ln: Log Normal, B: Beta).

Pair-Types	1st best fit	2nd best fit	3rd best fit	KS-test $\leq 5\%$	KS-test $\leq 10\%$
Flute-Flute (FF)	Pl[74%], Ll[10%], Pr[8%]	Ll[29%], Ln[26%], Pr[19%], W[11%]	Ll[39%], Pr[24%], Ln[21%]	Pl[72%], Ll[19%] Ln[15%], Pr[15%]	Pl[92%], Ll[92%] Pr[89%], Ln[85%]
Cello-Cello (CC)	Pl[39%], W[21%], B[14%]	B[22%], Ll[18%] Pl[16%], W[14%]	Ll[25%], W[24%], Pr[13%], Pl[12%]	Pl[34%], W[32%] B[30%], G[16%]	Ll[87%], Pl[86%] W[78%], Ln[71%]
Flute-Cello (FC)	Pl[67%], Ll[13%], Pr[9%]	Ln[28%], Ll[20%], Pr[19%], W[18%], Pl[10%]	Ll[46%], Ln[16%] Pr[14%]	Pl[61%], Pr[12%] Ll[11%], W[9%]	Pl[92%], Ll[92%] Pr[80%], Ln[78%]

- (2) Weekday vs. Weekend: We established significant differences between mobility and traffic patterns of weekdays and weekends in [14]. Here we analyze the encounter-traffic interplay across the weekdays and weekends.
- (3) Encounter duration: We define three encounter duration categories using bins of equal frequency: short ($< 0.6min$), medium ($0.6 - 5min$) and long ($> 5min$).

5.1 Device type categories

We analyze how similarity of traffic profiles for encountered pairs varies: *Flute-Flute* (FF), *Cello-Cello* (CC) or *Flute-Cello* (FC). The results, as presented in Figure 2, show that the similarity of CC is slightly higher than the other groups, while the FF and FC groups show similar trends. Notably, **all three encountered groups are significantly different from the non-encountered group** ($p < 0.05$). This is consistent across most buildings. Given the context of the traces, we suspect heavy use of laptops for educational content on campus. Further analysis website content may shed light on the shared interests among encountered users with various forms of devices. We leave this for future work.

Figure 2: CDF of pairwise cosine similarity of traffic profiles across device types (vertical lines denote medians).

5.2 Weekday vs. Weekend

There are significant differences between weekdays and weekends in user behavior. In [14], we found that numbers of user devices drops notably on weekends, but the remaining devices do not show significant differences in network traffic. Here we *identify and quantify the encounter-traffic correlation over weekdays/weekends* for the first time. As shown in Fig 3, we find the **pairwise similarities of weekend pairs to be overall higher than their weekday counterparts** regardless

of an encounter, with weekend *non-encountered* pairs being more similar than weekday *encountered* pairs. This is partly explained by significantly reduced mobility of devices on weekends. For example, median radius of gyration for *cellos* drops by 66%, and by 15% for *flutes* [14]. In addition to decreased mobility, most activity is clustered around several academic buildings with research labs (on weekends, 33% and 56% of APs handle no *flute* and *cello* traffic, respectively). Thus the increase in similarity during weekends might be explained by the presence of researchers collaborating on related fields of interest and accessing similar content.

Figure 3: CDF of pairwise cosine similarity of traffic profiles in weekdays and weekends (vertical lines denote medians).

5.3 Encounter duration

Using encounter durations (Section 3.1), the categories are 3 bins of equal frequency: short ($< 0.6min$), medium ($0.6 - 5min$), long ($> 5min$). Depicted in Fig. 4, the short encounters are not significantly different from the non-encountered pairs ($p > 0.05$). However, the differences between the other groups are statistically significant, with the **long encounter group showing the highest similarity of traffic profiles**, hinting at a correlation between the encounter durations and similarity. We found insignificant correlation for short and medium encounters. There is a *small positive* correlation between encounter duration and traffic profile similarity for long encounters ($\rho = 0.21$). Breaking down the correlations into device types and weekday/weekend, Fig. 5 shows the highest correlation for *Cello-Cello* (CC) group on weekends, supporting our earlier observation.

Overall, the correlation between encounter duration and similarity is dynamic, changing across space and time. Surprisingly, a few buildings show significant *negative* correlations

Figure 4: CDF of pairwise cosine similarity of traffic profiles with different encounter durations (vertical lines denote medians).

Figure 5: Pearson correlation coefficient between encounter duration and traffic profile similarity. (Cello-Cello (CC) encounters on weekends show the highest positive correlation.)

on weekends (e.g. music and theater), while others show significant *positive* correlations (mostly academic). Further spatio-temporal analysis is left for future work.

6 LEARNING ENCOUNTERS

Given the relationships shown, there is great potential in training a *machine learning model* that can predict an encounter given two traffic profiles (TP). Such a model has several *practical applications*. Given two TPs, if it is possible to **predict** an encounter with good accuracy, then there is useful information in the relationship of encounters and traffic profile similarity, which can be used in design of encounter-based *dissemination services even if mobility traces are not reliable for each user* (e.g. due to MAC address *randomization*), but traffic profiles of users are accessible (via *authentication mechanisms* identifying users at a higher OSI layer).

To investigate the feasibility of this task, for every pair, their *traffic profiles* in each building and day are coupled as input and the binary target label is based on whether the pair has encountered. Since most pairs do not typically encounter on a day, predicting a negative label is trivial in this case. To prevent this bias, we sample the dataset to ensure each label is represented by an equal number of samples for our models.

6.1 Random Forest

We first used Random Forests [20, 21] for this classification task, which is a well-established algorithm for supervised learning problems. Our work showed that on a building (the computer department) the algorithm achieved a promising $\approx 70\%$ accuracy on average across all days. Next, we applied

a dimensionality reduction algorithm, using *Singular-Value Decomposition (SVD)* to preprocess the input vector, adapted from Latent semantic analysis of natural language processing. Its application improved the accuracy, in the same settings, to $\approx 73\%$. This led to the idea of using stacked auto-encoders (SAE) to retain information and connect the SAE to a **deep, fully-connected neural network (DNN)** for classification.

6.2 Deep learning

We utilized several recent ideas from the field of artificial intelligence to improve our learning of encounters significantly. Auto-encoders are a class of artificial neural networks that are trained in an unsupervised fashion to learn an efficient representation of their input, with applications in feature extraction and dimensionality reduction.

We use a *Stacked De-noising Auto-Encoder (SDAE)* (detailed in [22]). We saw a significant **increase** in accuracy to a mean of **92%** for the same building and days. Comparing *device type categories* of encounters, *cello-cello* encounters are the most distinguishable, followed by *flute-cello* and *flute-flute*. However, the difference between accuracies for different device types is $<5\%$ in most locations and dates, a testimony to the *robustness* of the model. This accuracy is also stable across time, with the median, for *weekdays* at 93.25%, and *weekends* at 90.75%, for the computer department samples. The much higher accuracy comes at the cost of higher compute costs and complexity of the model. Details of the neural network architecture and results are in our tech report [15].

7 CONCLUSIONS AND FUTURE WORK

In this study, we presented the first steps to quantify the relation between mobility and traffic. Focusing on the pairwise (encounter) dimension of mobility, its interplay with the traffic patterns was studied, with major implications for realistic modeling, predictive caching, rumor anatomy and content sharing. We used extensive, granular datasets (30TB in size), in more than 140 buildings on a university campus, including information about WiFi events and NetFlow, covering the dimensions of mobility and network traffic. To answer our main question of ‘How do device encounters affect network traffic patterns, across time, space, device type, and encounter duration?’, We analyzed mobility encounters and presented their statistical characteristics. We defined traffic profiles and utilized numerical statistics and machine learning techniques.

The findings are not currently captured by any of the existing mobility or traffic models, while having important implications in many contexts. This is a compelling case for integrated traffic-mobility models that consider multiple dimensions of social context (individual, pairwise, and group). We plan to further investigate the causality between mobility and traffic for pairwise and group dimensions. Exploration of applications of our learning methods, privacy implications, and potential improvements are left for future work.

Acknowledgments: This work was partially funded by NSF 1320694, and Najran University, Saudi Arabia.

REFERENCES

- [1] F. Bai, N. Sadagopan, and A. Helmy, "The important framework for analyzing the impact of mobility on performance of routing protocols for adhoc networks," *Ad hoc networks*, 2003.
- [2] J. Treurniet, "A Taxonomy and Survey of Microscopic Mobility Models from the Mobile Networking Domain," *ACM CSUR*, 2014.
- [3] A. Hess, K. A. Hummel, W. N. Gansterer, and G. Haring, "Data-driven Human Mobility Modeling: A Survey and Engineering Guidance for Mobile Networking," *ACM CSUR*, 2016.
- [4] G. S. Thakur and A. Helmy, "Cobra: A framework for the analysis of realistic mobility models," in *INFOCOM'13*.
- [5] W.-J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling spatial and temporal dependencies of user mobility in wireless mobile networks," *ToN*, 2009.
- [6] W.-j. Hsu and A. Helmy, "On nodal encounter patterns in wireless lan traces," *TMC*, 2010.
- [7] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *TMC*, 2007.
- [8] N. Eagle and A. S. Pentland, "Reality mining: sensing complex social systems," *Personal and ubiquitous computing*, 2006.
- [9] X. G. Meng, S. H. Wong, Y. Yuan, and S. Lu, "Characterizing flows in large wireless data networks," *MobiCom '04*.
- [10] S. Moghaddam, A. Helmy, S. Ranka, and M. Somaiya, "Data-driven co-clustering model of internet usage in large mobile societies," *ACM MSWiM'10*.
- [11] G. Maier, F. Schneider, and A. Feldmann, "A first look at mobile hand-held device traffic," in *PAM '10*. Springer.
- [12] X. Chen, R. Jin, K. Suh, B. Wang, and W. Wei, "Network performance of smart mobile handhelds in a university campus wifi network," *ACM IMC'12*.
- [13] U. Kumar, J. Kim, and A. Helmy, "Changing patterns of mobile network (WLAN) usage: Smart-phones vs. laptops," *IWCMC '13*.
- [14] B. Alipour, L. Tonetto, A. Yi Ding, R. Ketabi, J. Ott, and A. Helmy, "Flutes vs. cellos: Analyzing mobility-traffic correlations in large wlan traces," in *IEEE INFOCOM*, 2018.
- [15] "On the relation between mobile encounters and web traffic patterns: A data-driven study." [Online]. Available: <https://arxiv.org/abs/1808.03842>
- [16] M. ElSherief, B. Alipour, M. Al Qathrady, T. ElBatt, A. Zahran, and A. Helmy, "A novel mathematical framework for similarity-based opportunistic social networks," *PMC*, 2017.
- [17] M. Al Qathrady, A. Helmy, and K. Almuzaini, "Infection tracing in smart hospitals," in *IEEE WiMob '16*.
- [18] G. Salton and M. J. McGill, "Introduction to modern information retrieval," 1986.
- [19] K. S. JONES, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, 1972.
- [20] T. K. Ho, "Random decision forests," in *ICDAR*, 1995.
- [21] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [22] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, 2010.
- [23] P. Pons and M. Latapy, "Computing communities in large networks using random walks," in *ISCIS*, 2005.
- [24] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The annals of mathematical statistics*, 1947.