Mining Crowd Mobility and WiFi Hotspots on a Densely-populated Campus

Mengyu Zhou
Tsinghua University
Beijing, China
mengyu.chou@gmail.com

Kaixin Sui
Tsinghua University
Beijing, China
suikaixin@gmail.com

Dan Pei
Tsinghua University
Beijing, China
peidan@tsinghua.edu.cn

Thomas Moscibroda
Microsoft Research
Beijing, China
moscitho@microsoft.com

ACM.
UbiComp/ISWC'17 Adjunct, September 11–15, 2017, Maui, HI, USA
ACM 978-1-4503-5190-4/17/09.
https://doi.org/10.1145/3123024.3124419

Abstract
Understanding crowd activities at large-scale and diagnosing existing problems of planning on densely-populated campus are fundamentally hard through traditional ways of measurement and management. In this paper, we demonstrate how to collect data from ubiquitous WiFi networks (WLAN), and further to characterize the mobility of campus residents by exploring time-frequency patterns with spatial context. On the campus of Tsinghua University (where everyday nearly $60,000$ mobile devices appear in the public areas of more than 110 buildings), we obtain large-scale observations on physical activities, and provide insights for better diagnosing of WiFi hotspots.

Introduction
The management and planning of densely populated urban areas usually face great challenges through traditional ways with time-consuming process of data collection, outdated analyses, and sometimes subjective and anecdotal arrangements. This is especially true for campuses, where the activities of students and faculties are seldom measured and analyzed at large-scale, let alone further used to guide planning and possible improvements.

In the past decade, the ubiquitous WiFi infrastructure and smart phones offer a great opportunity to study people's physical activities and interactions. On school campuses,

Table 1: Building Basic Information

Bvld.	type	area(m^2)	#AP
A	cafeteria	9696	8
B	gym	12600	19
C	dorm	35668	84
D	classroom	34045	115
E	library	20000	50
F	department	36780	121

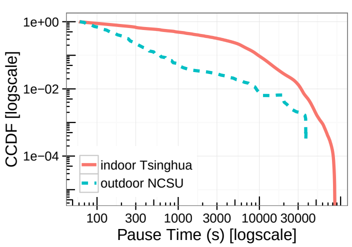

Figure 1: Pause time distributions.

nowadays students and teachers usually study/work/live with good WLAN coverage with their carry-on wireless devices. Their usage of WiFi and mobile devices produces an unprecedented wealth of information, leading to a long line of research that focus on campus [2, 5].

In this paper, we characterize crowd mobility at Tsinghua University, where a deployment of more than $2,800$ APs gives essentially complete coverage across a diverse set of 114 buildings. Based on the data collected from campus WLAN, we are able to explore the following questions:

1. How do residents move around and spend their time on campus?
2. What are frequent and recurrent temporal patterns of crowd activities?
3. How do spatial contexts (buildings) impact physical activities?
4. Are public facilities well utilized? What are their potential problems?

Our large-scale measurement study sheds light on how to analyze mobility traces with building context, and diagnose potential WiFi deployment problems for better planning.

Data Collection

The campus of Tsinghua covers an area of $\sim 4.4km^2$ on which $\sim 45,000$ students and $\sim 12,000$ faculty and staff members are living. By August 2016, there are $2,890$ Cisco enterprise APs (access points) in 116 buildings on the campus (*e.g.*, the basic information of 6 example buildings are listed in Table 1.), providing a dense deployment in most areas. At peak, there are $\sim 20,000$ devices concurrently connected to the campus WLAN. The total number of unique devices surpass $60,000$ each day, which means on average everyone uses at least one wireless device.

The ubiquitous enterprise WLAN on the campus allows us to track the devices of the large resident population. As described in [5], the mobility of each device is derived from records of probe/rogue packet RSSI and connected RSSI in sampled (with $\sim 5min$ interval) SNMP objects with the connection history in full SNMP trap messages. No matter whether a device is connected to the campus WLAN or not, the mobility of the device can be determined as long as its probe scan or data packets can be sensed by APs.

The “mobility" of a device is defined as its pause interval and pause location at room-level. The mobility detection algorithm described in [5] is training-free — The complicated and costly procedure of absolute indoor positioning is avoided. During a week of spring semester at Tsinghua, $69,154$ client devices appeared in the campus WLAN. The pause time (the duration of a pause interval) distribution is shown in Fig. 1.

Mobility Analyses

From the mobility traces of smartphones on the campus, we try to explore the following questions by aggregating traces together and conducting statistical analyses with spatial contexts: Where do students spend their time? And how their activities are different at different places?

The differences of indoor and outdoor activities lead us to focus on activities of indoor scenarios on the campus. Meanwhile, the heterogeneous usages — such as arrival and departure flows — of buildings can be used as indicators of *spatial context*. Then, we build a pause time model based on the building types.

Indoor v.s. Outdoor

In general, human behavior inside buildings is quite different from outside. There are less motions but more pauses at indoor areas. Outdoor human mobility shows statisti-

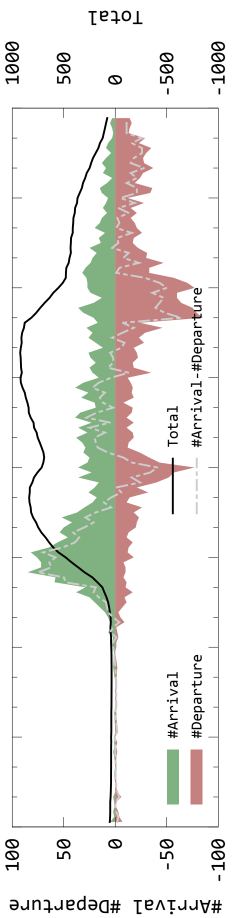

Figure 2: Arrival (upper green field) & departure (lower red field) flows, and total device number (solid line) in the department building FIT during a Wednesday. (Aggregate bin width = $10min$.)

cal resemblances to Lévy walks[1] [3] which is also found in similar mobility behavior of other wild animals like jackals, spider monkeys, *etc.* As Fig. 1 shows, indoor mobility still follows the scale-free (that in any scale human movement has similar patterns) pattern of heavy-tailed pause time distributions, which is a feature of Lévy walks.

However, indoor mobility no longer follows past models of outdoor mobility. In buildings, smaller area no longer leads to shorter pause time as presented in outdoor cases. For example, [3] shows that both flight length and pause time decreases as the outdoor boundary is more confined. But as shown in Fig. 1 (where the pause time distribution from outdoor GPS traces on the Campus-I/NCSU in [3] is also plotted), outdoor pause time on NCSU campus is much shorter than pause time within buildings (much smaller areas than a whole campus) of Tsinghua.

At Tsinghua, mobility still follows the observation that modern humans spend roughly 90% of the time indoor. Meanwhile, 93% of the indoor time is in pause mode at 3.1 different spots all over the campus. As shown in Table 2, on average the time spent in each building is around $210min$, while pausing 2.6 times for about $196min$. We can see the trend that most people are paused in the buildings for most of the time.

Buildings restrict people into bounded areas. In buildings people are less fluid and less sparse than their outdoor states.Thus it is necessary to take a closer look at buildings for better understanding of crowd activities.

[1]Levy walks are defined as random walk trajectories that are composed of self-similar jumps, and consist of many short flights and occasionally long flights. They are more diffusive than Brownian motion while less diffusive than random-waypoint movement.

Building Types and Heterogeneous Activities

Most of the buildings on the campus have their own special regular and significant temporal patterns through a long period of time. In other words, each of them usually has a fixed major usage. Thus depending on the general usages and schedules of the building, we manually categorize the buildings into 8 major types: administrative, cafeteria, classroom, department, dorm, gym, library and others.

Before investigating indoor pauses and outdoor transitions separately, we take an overview at the aggregated arrival and departure flows of each building. *E.g.*, in Fig. 2, flows of arrival, departure and total number of smartphones (derived from mobility traces) on a weekday the example department building FIT are shown.

To further understand similarity and differences of activities among buildings, in [5] we try to cluster buildings based on the similarities between their flows over a long time. To compare two time series of flow, we calculate cross correlation [2] between them with lag $|\tau| \leq 30min$ to derive the similarity of two buildings. Then we apply a hierarchical clustering on all the buildings. The results in [5] show that same type of buildings are well clustered together. Building types can be used as coarse labels for spatial context.

Visitors v.s. Occupants

Finally, we look back at the pause time distributions through the categorization of building types. Dotted "WLAN" line in Fig. 3 demonstrates the mountain range undulating patterns of pause time — most types of buildings have peaks in their tails far away from the power-law head around $<20min$. These heavy peaks in the tails reflect the common

[2]Cross correlation (also known as sliding dot product or sliding inner-product) can be defined as $\rho_{XY}(\tau) = E[(X_t - \mu_X)(Y_{t+\tau} - \mu_Y)]/(\sigma_X \sigma_Y)$ for series X and Y.

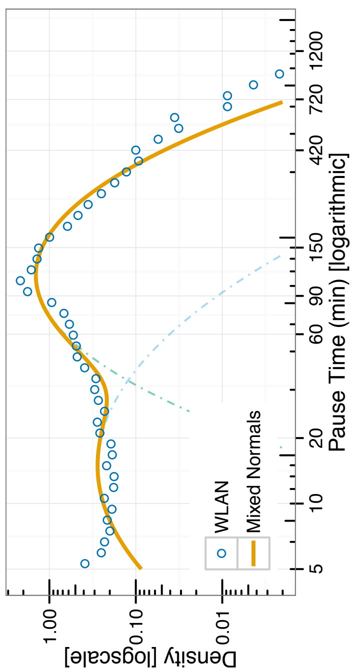

Figure 3: Pause time distributions of building types.

Mixed normal fit of log_{10}(pause time). The x axis is labelled with original pause time, using logarithmic bin size.

Table 2: Pause Statistics (one weekday).

The number of pauses, percentage of pauses to indoor time, total pause time (min), total indoor time (min), number of paused spots.
(Averaged over all observations in buildings of each type.)

type	#pause	%pause	pause(min)	indoor(min)	#spots
admin	2.60	92%	198.68	211.29	1.40
cafeteria	1.08	99%	48.74	49.13	1.29
classroom	1.64	97%	125.49	133.66	1.98
department	4.17	87%	242.92	268.09	2.27
dorm	2.48	91%	387.53	408.89	1.59
gym	3.27	87%	84.44	94.55	1.55
library	1.66	97%	198.47	201.58	1.60
ALL	2.64	93%	196.35	210.76	3.11

schedules, long-time usages, and frequent patterns of occupants/residents of the building as we partly saw in Table 2. The peak at their heads around 10-$20min$, we conjecture, is caused by the short-time usage of visitors, random process in the human walks similar to outdoor mobility, and influenced by the building space properties (design, architecture, location, environment, *etc.*).

The multiple peaks of indoor pause time distribution can hardly be fitted well using same simple heavy tail distributions, *e.g.* Log-normal, Weibull, Truncated Pareto, *etc.* which are widely adopted in the past works on outdoor or general mobility. By our observation that most types of buildings have long-term occupant usages and short-term visitor usages presented by two main peaks in most distributions of $log_{10}(\text{pause time})$ in Fig. 3, we proposed a Log-mixed-normal model for indoor pause time: $P \sim exp(\lambda_v \mathcal{N}(\mu_v, \sigma_v) + \lambda_o \mathcal{N}(\mu_o, \sigma_o))$ where $\mathcal{N}(\mu_v, \sigma_v)$ and $\mathcal{N}(\mu_o, \sigma_o)$ are normal distributions defining the left and right peaks. The model parameters of λ_v, μ_v, σ_v, λ_o, μ_o and σ_o for each building type is fitted using standard EM algorithm. The fitting results are shown in Fig. 3 and Table 3. The pause time distribution in Fig. 3 is well approximated. The model parameters in Table 3 are in the same pattern as shown in Table 2. The mean μ and deviation σ describe

Table 3: Fitted Log-mixed-normal Model

type	λ_v	μ_v	σ_v	λ_o	μ_o	σ_o
admin	0.222	2.722	0.175	0.778	3.722	0.446
cafeteria	0.918	3.042	0.268	0.082	4.135	0.283
classroom	0.214	2.949	0.310	0.785	3.829	0.219
department	0.231	2.730	0.181	0.769	3.708	0.443
dorm	0.802	3.590	0.567	0.198	4.505	0.106
gym	0.222	2.676	0.141	0.778	3.395	0.391
library	0.233	2.901	0.284	0.767	3.927	0.312

our common senses of the building usages. *E.g.* the $\mu_o = 3.829$ and $\sigma_o = 0.219$ for classrooms corresponds with the common $95min$ to $155min$ ($log_{10}(95 * 60) \approx 3.756$ and $log_{10}(155 * 60) \approx 3.968$) length of lectures.

WiFi Hotspots

With the analysis on mobility traces, in this section we further explore how to diagnose for campus planning, especially indoor WiFi hotspots.

Since internet access through WiFi is a basic need on the campus, we take its access points as an example of facilities. Based on our mobility detection results, two major types of hotspots - "**pause**" and "flight" hotspots where a lot of devices pause at or fly by - are identified. "Pause" hotspots can be extensively used by occupants. "Flight" hotspots are the important cases for wireless networks study and practical WLAN administration because they are the most challenging cases for wireless communication and roaming. So we further characterize **entrance/exit** (where most people enter and leave the WLAN of the building), **leaky** (high number of short appearances of non-resident devices) and **roaming** hotspots (high number of quick pass-bys of resident devices) as subclasses of flight hotspots. Under utilized for most of the time. Imbalance of AP (take FIT as an example, 4 types of hotspots: pause, entrance/exit, leaky, roaming).

Table 4: Hotspots Classification with Mobility in the 6 Example Buildings A $\sim$ F.

(#p: number of pause hotspots. #e: number of entrance hotspots. #l: number of leaky hotspots. #r: number of roaming hotspots.)

	A	B	C	D	E	F
#p	4	9	55	106	37	103
#e	4	4	1	3	1	2
#l	7	10	16	29	16	16
#r	7	17	15	26	17	24
$\sum$	8	19	84	115	50	121

Table 4 shows the 4 types of hotspots (with overlap) in the 6 example buildings. The WLAN was deployed based on the signal coverage and the prior estimation of population density. However, we can find that past deployments leave a lot of potential performance problems. *E.g.* in cafeteria A, 8 APs are in short supply and have their power set too high; in classroom D there are several leaky APs due to the hollow structure of the building. In the example building F, APs on floor 1 (ground floor) and 6 (top floor) are classified as very different hotspot types. The entrance/exit hotspots near the gate of floor 1 could be identified. While most of APs on floor 6 are not affiliated with any hotspot type, which means they may be wasteful cold-spots. In one word, by knowing the hotspots in each building, network operators can adjust the current deployment of APs and provide a much better WiFi experience.

Related Work

It is a fundamental problem to understand human mobility and behaviors, with far reaching impact. Examples include: civil planning, transportation improvement, energy optimization, improvement of networking services, healthcare, online *v.s.* offline social networks, group behaviors, event detection [4, 1, 2, 5].

Conclusion

Campus planning and management are usually conducted through time-consuming process of manual data collection, outdated analyses, and sometimes subjective and anecdotal arrangements. In this paper, we demonstrate how to utilize ubiquitous WLAN to help better monitor the activities and diagnose problems on campuses, which sheds light on the more advanced studies about campus and urban life in the future.

REFERENCES

1. Chloë Brown, Christos Efstratiou, Ilias Leontiadis, Daniele Quercia, Cecilia Mascolo, James Scott, and Peter Key. 2014. The architecture of innovation: Tracking face-to-face interactions with ubicomp technologies. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 811–822.
2. Kasthuri Jayarajah, Archan Misra, Xiao-Wen Ruan, and Ee-Peng Lim. 2015. Event Detection: Exploiting Socio-Physical Interactions in Physical Spaces. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 508–513.
3. Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, Seong Joon Kim, and Song Chong. 2011. On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking (TON)* 19, 3 (2011), 630–643.
4. Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. 2011. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1100–1108.
5. Mengyu Zhou, Kaixin Sui, Minghua Ma, Youjian Zhao, Dan Pei, and Thomas Moscibroda. 2016. MobiCamp: a Campus-wide Testbed for Studying Mobile Physical Activities. In *Proceedings of the 3rd Workshop on Physical Analytics*. ACM.