Using WiFi Technology to Identify Student Activities Within a Bounded Environment

Philip Scanlon and Alan F. Smeaton

Insight Centre for Data Analytics
Dublin City University
Glasnevin, Dublin, 9, Ireland
Email: philip.scanlon@insight-centre.org

Abstract. We use the unique digital footprints created by student interactions with online systems within a University environment to measure student behaviour and correlate it with exam performance. The specific digital footprint we use is student use of the Eduroam WiFi platform within our campus from smartphones, tablets and laptops. The advantage of this data-set is that it captures the personal interactions each student has with the IT systems. Data-sets of this type are usually structured, complete and traceable. We will present findings that illustrate that the behaviour of students can be contextualised within the academic environment by mining this data-set. We achieve this through identifying student location and those who share that location with them and cross-referencing this with the scheduled University timetable.

1 Introduction

The ability of researchers to identify the type of activities and levels of interaction among students on campus is important to research in Learning Analytics and in particular, anthropological studies which explore interactions among students. Historically the collection of base data in such studies has in the main been through observation, questionnaires or a combination of both. In an era where smartphones and WiFi use are widespread, this paper will examine another data source, the use of WiFi-enabled devices within a bounded domain, i.e within a campus.

Our work uses the digital footprint that WiFi-enabled devices leave to identify student location and thus co-location of students. From this co-location analyses we infer peer groupings and levels of interaction. This can be used for identifying peers in a University community and for the identifying popular locations for different students and their peer groups. This paper examines the data collection process we followed. We use spatio-temporal data derived from WiFi system logs to determine on-campus location as a component of student digital footprints. Once gathered, learning analytics can mine this to produce actionable knowledge for use in the learning process. All data has been anonomised and the work is approved by the University's Research Ethics Committee.

2 Related Work

Central to research into peer influence is the ability to identify those who spend time with others, and for what purposes. Previous work by Celant [1] asked students directly to recall who they spend time with, socially, academically, jointly working on homework, or as part of formal study groups. Celant found that there was some *blurring* as "different students may have had a different concept of preparing for an exam with a course mate". Other students who, for different reasons, do not interact at a level they would prefer, may say they study with others with whom they have had little interaction with.

In an early example of the use of technology to collect geospatial data from student activities, data was collect over a two year period using a hand-held GPS device as part of Project Lachesis [6]. The aim was *"... extracting stays and destinations from location histories in a pure, data-driven manner"*. Technology has advanced in the intervening years research which have used technologies to collect data on the interaction between parties could be categorised as:

1. Using geospatial data collected through the use of GPS and GMS location data. Examples of this are [7] and [6].
2. Specifically adapted smartphones utilising bespoke data collection applications. Examples of this are [2], [4] and [5].
3. Badges that collect data relevant to the wearer, including [8] and [11].
4. Smartphones used to collect WiFi base station data [9] and [3].

The research data gathered in these papers ranged in scope from 2 users over a 1 year period, to 48 users for a 10-week period and through to 100 users for a period of 9 months. Our research is based on 3 academic years and a cohort of 174 students and we believe that this compares favourably with similar research.

Nathan Eagle's [2] longitudinal research study used data collected over a 9-month period from 100 mobile phones to demonstrate the ability to use Bluetooth technology to log user behaviour and activity information. The intention was to recognize social interaction patterns and to cluster locations, therefore modelling activities and inferring relationships through the monitoring of temporal and geolocation data. This is an approach used in many research projects and will be used in our research. The interaction between WiFi-enabled devices and the WiFi network which we explore here, allows us to identify information previously impossible to gather on both ad-hoc and formal groupings of people.

3 Co-location Datasets Used

Co-location in the arena of our research can be roughly interpreted as the location of two or more individuals in the same physical place and at the same time. Individual incidence of a co-located pair cannot be interpreted as the individuals as having a relationship and does not infer personal contact between 2 people. Quannan Li [7] having mined subject GPS logs, used an hierarchical clustering algorithm to develop a trajectory model that determines a semantic meaning to stay-points (points where time is spent) and inferred similarity between subjects based on this. In that work, the results could be used as the basis of a recommender system, but in our work we will interpret the *context* of the co-location. In the context of a University campus, meetings can be either formal or informal i.e they can formally scheduled meetings with others through joint attendance at a class or lab or they can be meetings in locations with their peers with whom they have a social relationship with.

The DCU campus is a modern facility contained on a 50-acre campus in North Dublin city. It compromises 27 separate buildings providing an approximate floor space of 180,000m^2. On-campus facilities include 1,400 residential apartments, 7 restaurants/cafes and numerous shops including convenience, book and a pharmacy. The WiFi coverage for the campus is provided by eduroam (Educational Roaming), a cross-site infrastructure which allows users gain access to the WiFi platforms at other eduroam sites where access is provided following authentication by Radius servers at their home institution. Network access at member sites is via 802.1X protocols and at DCU comprises 1,000 individual Network Access Servers (NAS). These NASs are distributed across the campus ensuring continuous WiFi coverage to users.

Our research cohort is drawn from the Faculty of Engineering and Computing and specifically from the School of Computing. Within this School our research will focus on students in two undergraduate programs namely, Computer Applications (CA) and Enterprise Computing (EC). These two programs have been chosen as they share some modules and a degree format which attracts students with similar interests in the IT domain.

Our research data relates to the academic years 2014/5, 2015/6 and 2016/7 and contains c.220 million log file entries. As with any data, in its raw form it comprises multivariate data requiring pre-processing to ensure usability. To augment our WiFi log access data, we have an additional data-set of basic student demographics and exam performance. Students who first registered in 2015 were identified as the initial cohort whose activities would to be analysed during this research. Using the 2015 intake of students, this paper will illustrate how identifying the use of WiFi access as a part of students' digital footprints, can be used to identify the activities of students on a semester, weekly and daily basis.

Our research relies on the premise that students on campus are there for both formal and informal reasons. They are there to attend classes, scheduled labs, additional study and to interact with friends i.e. they will be engaged in either academic and social activities. Based on the on-campus location of subjects, we will infer activities and therefore infer their purpose. Using a similar approach to that of Rui Wang [10] we sub-divided the campus into academic and social areas. While some locations have a dual purpose for example the on-campus residences which could be considered a social area, some students will also use such locations for study and thus we will classify each area based on the majority use. This means that classrooms, laboratories (labs) and libraries are categorised as *academic* while on-campus residences, cafe, restaurants and bars (hang-out areas) as well as public areas such as transit spaces are classified as *social*.

4 Preliminary Profiling of Student Activity

Using the definition of a "meeting" between students as being two devices from two individual students, both connected to the same eduroam WiFi NAS for an overlapping period of 20 minutes or longer, we can compare the activities of the cohorts of students from two different degree programmes, Computer Applications (CA) and Enterprise Computing (EC).

Program	No. Students	No. Dyads	No. Meetings	Avg. Meetings per Dyad
CA	114	5,523	335,465	60.7
EC	60	1,230	106,500	86.6

Table 1. Student numbers, Dyads and Meetings

Table 1 lists the number of students and dyads who had meetings, i.e. student pairs from the same degree program that interact or "met" during the semester and the number of meetings during the semester among those dyads. To achieve a greater understanding of student activity while on campus we divided each location into two categories and a number of sub-categories. The premise is that *friends* spend a lot of time together in the same location at the same time. It was thus necessary to identify the degree to which students collocate and use this in our analysis. Table 2 outlines the number of meetings in social locations and Table 3 the Academic meetings by sub-category. It can be seen that 70% of the the CA students met in *Hang-out* locations compared to 78% of EC students.

Program	Social	Hang Out	Transit	Residence
CA	36,543	25,386 (70%)	8,409	2,865
EC	16,962	13,268 (78%)	3,282	261

Table 2. Student numbers and Meeting Locations

Program	Academic	Class	Lab's	Library
CA	298,922	199,968 (67%)	97,803 (33%)	1,027
EC	89,538	32,284(36%)	55,619 (62%)	1,786

Table 3. Student numbers and Meeting Location Details

When comparing programs there is a large variance between the percentage of CA student *class* meetings i.e. 67% and EC *class* meetings at 36%. At this time we are not sure why this differential is present. In the Academic domain, Table 3 shows the largest number of meetings took place in the *class*. A large portion of *class* meeting occur within the formal environment of lectures with a smaller amount where students have study groups at class locations. It is common for students to congregate in the Labs to study or work on group projects.

As part of our demographic data we examined student *Precision score*, an aggregated score compiled from the exam marks achieved in all program modules during the year. Our analysis compares the *Precision* score of students and the number of times they met in Academic and Social Locations.

Table 4 groups the number of meetings into groupings of 200 and lists the Average delta between students whose meetings fall into that group and also include the max delta between student pairs in the group. In this, a delta is interpreted as the difference between two student Precision marks for an Academic year. We can interpret from this table that as the number of meetings increases between a dyad in Academic settings, there is a decrease in the Delta score. That is, the more meetings a pair of students have, the closer their exam grades. Whether this is an indication of peer influence or that students of similar ability naturally group together will require further study.

	CA		EC	
Academic Meetings	Avg. Delta	Max	Avg Delta	Max
0 : 200	**23.82**	54.84	**6.85**	25.00
200 : 400	11.90	44.59	5.79	19.17
400 : 600	12.22	37.84	7.76	19 .00
600 : 800	11.02	32.75	5.58	15.17
800 : 1000	10.72	28.92	5.22	12.17
1000 : 1200	9.79	15.25	3.71	5.09
1200 : 1400	**1.59**	1.59	**0.84**	0.84

Table 4. Average Delta for Academic Meetings, Grouped

Similarly in Table 5, the Social meetings analysis, the average delta decreases as the number of meetings between pairs increases. However it can be seen that the range of difference in deltas between groups, varies considerably. We see that in the CA (Computer Applications) category there is a wide variance in the Academic deltas.

5 Conclusion

This research has determined that it is possible to identify student locations on a campus through the digital footprint provided by their WiFi activity. We find that, based on co-location data, the degree of each student's activity within the cohort can also be determined. We have illustrated that a longitudinal study of this nature can identify relationships between students and their academic timetables and a difference in group behaviour between students in different programs and variations in their exam performance.

Acknowledgement: This paper is based on research conducted with the support of Science Foundation Ireland under grant SFI/12/RC/2289.

	CA		EC	
Social Meeting	Avg. Delta	Max	Avg Delta	Max
0 : 30	**13.75**	54.84	**6.86**	25
30 : 60	12.39	42.34	6.27	19.17
60 : 90	11.51	32.09	6.67	18.17
90 : 120	9.95	30.42	5.56	11.33
120 : 150	14.56	37.84	5.84	15.17
150 : 180	9.38	15.34	3.02	5.83
180 : 210	6.05	10.25	5.23	10.09
210 : 240	9.38	17.42	5.30	11.84
240 : 270	**10.96**	11.25	**2.39**	4.25

Table 5. Average Delta for Social Meetings, Grouped

References

1. S. Celant. The analysis of students academic achievement: the evaluation of peer effects through relational links. *Quality & Quantity*, pages 1–17, 2013.
2. N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2006.
3. M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
4. A. Gupta, S. Paul, Q. Jones, and C. Borcea. Automatic identification of informal social groups and places for geo-social recommendations. *International Journal of Mobile Network Design and Innovation*, 2(3-4):159–171, 2007.
5. G. M. Harari, S. D. Gosling, R. Wang, F. Chen, Z. Chen, and A. T. Campbell. Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. *Computers in Human Behavior*, 67:129–138, 2017.
6. R. Hariharan and K. Toyama. Project lachesis: parsing and modeling location histories. In *International Conference on Geographic Information Science*, pages 106–124. Springer, 2004.
7. Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma. Mining user similarity based on location history. In *Proc. of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, page 34. ACM, 2008.
8. A. Pentland. The new science of building great teams. *Harvard Business Review*, 90(4):60–69, 2012.
9. J. Rekimoto, T. Miyaki, and T. Ishizawa. Lifetag: Wifi-based continuous location logging for life pattern analysis. In *LoCA*, volume 2007, pages 35–49, 2007.
10. R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell. SmartGPA: How Smartphones Can Assess and Predict Academic Performance of College Students. In *Proc 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 295–306, New York, NY, USA, 2015. ACM.
11. J.-I. Watanabe, S. Matsuda, and K. Yano. Using wearable sensor badges to improve scholastic performance. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 139–142. ACM, 2013.