Modelling students' social network structure from spatial-temporal network data

Quan Nguyen
School of Information, University of Michigan
quanngu@umich.edu

Warren Li
School of Information, University of Michigan
liwarren@umich.edu

Christopher Brooks
School of Information, University of Michigan
brooksch@umich.edu

ABSTRACT: Network analysis in educational research has primarily relied on self-report and/or data generated from online learning environments (e.g. discussion forums). However, a large part of students' social connections occurs through day-to-day interactions on campus. This paper describes an on-going work exploring the application of WiFi network data to model social network structure amongst students on campus at scale. Links between individuals were inferred based on their spatial co-occurrences along a similar temporal dimension (i.e. two individuals connected to the same WiFi access point at the same time). We discussed a potential approach to test the statistical significance of these connections against a null model, in which two individuals might randomly be at the same place at the same time.

Keywords: Network analysis, spatial-temporal data, WiFi log data

1 WIFI NETWORK DATA IN EDUCATION

Wireless local area networks (WLANs) have increasingly become ubiquitous in modern education as they provide seamless internet access to students, teachers, and staffs through a large number of WiFi access points on campus. Log-files generated from WLANs are rich in both temporal and spatial features. They recorded the timestamp of each user's devices being connected to a particular WiFi access point (Table 1).

Table 1: Anonymized sample WiFi data.

User ID	Timestamp	Access point	MAC address
A1234	2018-09-24 08:00:00	TWC-1023NW	Android- A1234
A1234	2018-09-24 08:02:03	TWC-2013NW	Android- A1234
B2314	2018-09-24 08:00:03	BAHR-1210-N	Apple- B2314
C2153	2018-09-24 08:00:05	CQTB-3734	Ubuntu- C2153

While there has been extensive research using WiFi data focusing on signal processing, only a limited number of studies has explored the application of WiFi data for educational purposes. For example, the iSpots project at MIT collected WiFi data and visualized the dynamics changes in wireless traffic on the wireless network and showed how people move around campus in real-time (Sevtsuk, 2009). WiFi data has also been used in predictive modelling. Sarkar, Carpenter, Bader-El-Den, and Knight (2016) estimated the correlations between students' time spent on campus based on WiFi log data and academic performance. In another study, Hang, Pytlarz, and Neville (2018) combined WiFi log data with building location profiles at Purdue University to extrapolate the temporal dynamics of user's location preferences throughout the day, and to predict *Point of Interest* (POI) (e.g. where an user will be on Monday at 9:00 am) using graph embeddings. For instance, Zhou et al. (2016) utilized WLAN data at Tsinghua University to estimate students' punctuality (attendances, late arrivals, and early departures) for lectures as well as to assess the lecture's engagement using mobile phone's interactive states at minute-scale granularity. However, due to the sensitive nature of WiFi data, researchers should be cautious and transparent about their purposes.

Another promising application of WiFi data in educational research, which has yet to be explored, is to understand the social network structure of students. WiFi network data can help researchers capture the dynamic changes in social interactions on campus, which in turns, can be combined with discussion forum data and self-report social network surveys. Nonetheless, the spatial-temporal nature of WiFi data presents unique conceptual and methodological challenges for network analysis, which will be discussed below.

2 NETWORK INFERENCE FROM SPATIAL-TEMPORAL DATA

Data in this study were collected from 3,915 students enrolled in five large undergraduate STEM courses at a public university in the U.S. in the Fall semester of 2018. All students' identifiers were anonymized. To get a sense of the data, Figure 1 visualizes the temporal changes in WiFi access points of two users on a particular date from 08:00 to 20:00. These two users spent a large amount of time in the morning at a fixed WiFi access point, possibly attending a lecture. In the afternoon, these two users shared the same location for 2 hours. After that, each user went on about their day to different areas on campus.

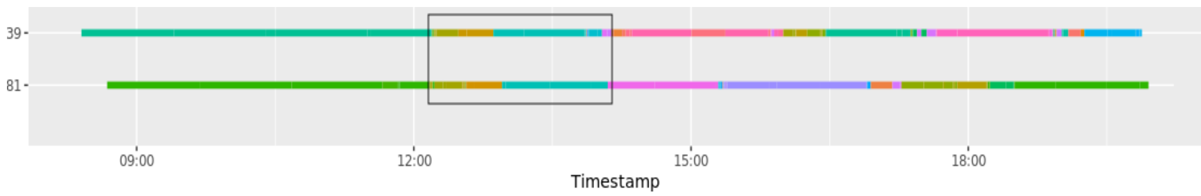

Figure 1: Temporal changes in WiFi access points of two users throughout a day. The boxed area indicates a two-hour period where these users shared the same access point

WiFi log data can be treated as a bipartite network (i.e. user-access point) along a temporal dimension (Figure 2). This can be projected into an undirected weighted one-mode network (i.e. user-user) under the following assumptions:

- Two nodes are linked if they connected to the same WiFi access point within the same time window (i.e. to be at the same place at the same time) (Figure 1).
- Tie's weight is determined as the shared duration for the same WiFi access point
- Tie's weight is discounted for the number of nodes sharing the same access point (i.e. the more people in the room, the weaker the tie between two particular individuals)
- Tie's connections could occur by random chance
- Tie's connections could occur due to shared events (e.g. attend the same lecture)

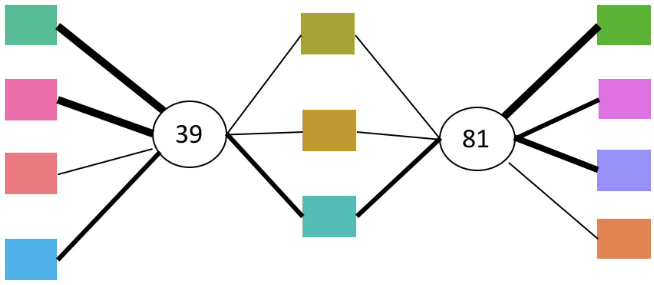

Figure 2: Bipartite network of two users (circles) and WiFi access points (squares). Tie's weight/thickness represents connected duration.

Based on this, we can calculate the total amount of shared duration between two particular users on a given day and aggregate them across a week/semester/year to create a weighted adjacency matrix amongst users. To test the statistical significance of ties, we propose drawing a random observation for each user (i.e. random amount of time spent at a random WiFi access point at a random time of day). This permutation can be repeated to generate a null distribution of shared duration between two particular users, which allow us to test the statistical significance of a given tie (Psorakis, Roberts, Rezek, & Sheldon, 2012). Due to the limited space and the early stage of this work, more concrete results will be presented and discussed at the conference. Nonetheless, we believe the unique opportunities, as well as challenges of using spatial-temporal data for network analysis in education, would bring out a lot of interesting discussions.

REFERENCES

Hang, M., Pytlarz, I., & Neville, J. (2018). *Exploring Student Check-In Behavior for Improved Point-of-Interest Prediction*. Paper presented at the Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, United Kingdom.

Psorakis, I., Roberts, S. J., Rezek, I., & Sheldon, B. C. (2012). Inferring social network structure in ecological systems from spatio-temporal data streams. *Journal of The Royal Society Interface, 9*(76), 3055-3066.

Sarkar, S., Carpenter, B., Bader-El-Den, M., & Knight, A. (2016, 6-8 July 2016). *Where students go and how they do: Wi-Fi location data versus academic performance.* Paper presented at the 2016 9th International Conference on Human System Interactions (HSI).

Sevtsuk, A. (2009). Mapping the MIT campus in real time using WiFi *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City* (pp. 326-338): IGI Global.

Zhou, M., Ma, M., Zhang, Y., SuiA, K., Pei, D., & Moscibroda, T. (2016). *EDUM: classroom education measurements via large-scale WiFi networks*. Paper presented at the Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany.