Quantifying the impacts of social infrastructure on human networks

Yalin Yang*, Yanan Wu and May Yuan

School of Economic, Political and Policy Sciences, the University of Texas at Dallas
* yxy180050@utdallas.edu

Keywords: Place, Social Events, A priori, Monte Carlo Simulation

Introduction

A rich sociology literature claims that social capital plays a more valuable role than physical infrastructure in improving community resilience (Kirmayer & Whitley, 2009; Metaxa-Kakavouli et al., 2018); Community resilience refers to a concerted ability of residents to defy or efficiently recover through cooperation during or after a disaster strikes (*Wiley*, n.d., 2017). Norris et al. (2008) proposed four fundamental adaptive capacities: - Economic Development, Social Capital, Information and Communication, and Community Competence, to flourish community resilience and the need to engage social activities to strengthen the social linkages in a community. The U.S. Federal Emergency Management Agency's National Disaster Recovery Framework recommends that responders build and maintain partnerships with each other by increasing social interaction and activities.

While extensive research contrasts the impact from physical infrastructure and social interactions, this study assumed the built environment provides the context and foundation for social activities. For instance, the corner of a crowded street cannot host a formal business negotiation properly. Same, the fitness room is not an ideal place for academic conference hosting. The so-called "inappropriate match" between the built environment and social activities implies their association in the real world.

Empirical evidence is needed to quantitatively assess the association between the built environment and human activities and changes in human-environment interactions across space and time. A clear understanding of how the built environment facilitates social events can help inform the best investment for community recovery and promote planning strategies for social connections in a community.

This study asks the following research question: How is the built environment spatially associated with various social events? How to quantitatively measure the spatial association? To address these questions, we proposed a framework that employs points of interest (POI) data from Maptitude to describe the site characteristics and utilize the social events from event-listing online platforms, such as Eventful, Eventbrite, Meetup, EventInn, etc. We aim to discover the spatial association between human activities and surrounding physical facilities. This study demonstrated this framework using the DFW area as a case study and discovered 31 significant associations between POI types and social event types. Considering social events bring people to places in coproducing shared experiences, this work drew insights into the influences of the spatial locales on our site perception and on shaping our social behaviours. In addition to new findings of spatial associations between POI types and social events, the study proposed a new approach of collocation analysis for spatial association mining.

Method

Data pre-processing

First, we acquired records of geotagged social events from Meetup, an online platform for users to host local in-person or online events. To the end of 2020, Meetup had more than 52 million registered members across 330,000 groups in 193 countries and 10,000 cities around the world (*One Million Meetup Events Hosted Online*, n.d.). The social event data include events descriptions, attendees' lists, event dates, and locations (Figure 1). These social events are categorized into 24 event types. A total of 9708 social events were collected from 26 February 2020 to 30 January 2021 across Dallas/Fort Worth (DFW) area. After removing records for online events or events without geographic locations, the remaining data include 9,445 in-person events at 1,537 unique locations. We reclassified the 24 types of social events into seven (7) types: *Hobbies & Passions* (Sci-Fi & Games, Dance, Music, Food & Drink, Hobbies & Crafts, Arts, Photography, Film, Fashion & Beauty), *Science & Education* (Tech, Learning, Writing, Book Clubs), *Social Activities* (Social, Language & Culture, Pets, Family), *Movements* (Movements), *Religion and Identity* (Beliefs, LGBTQ), *Career & Business* (Career & Business), *Sports & Health* (Health & Wellness, Sports & Fitness, Outdoors & Adventure). Figure 2 A

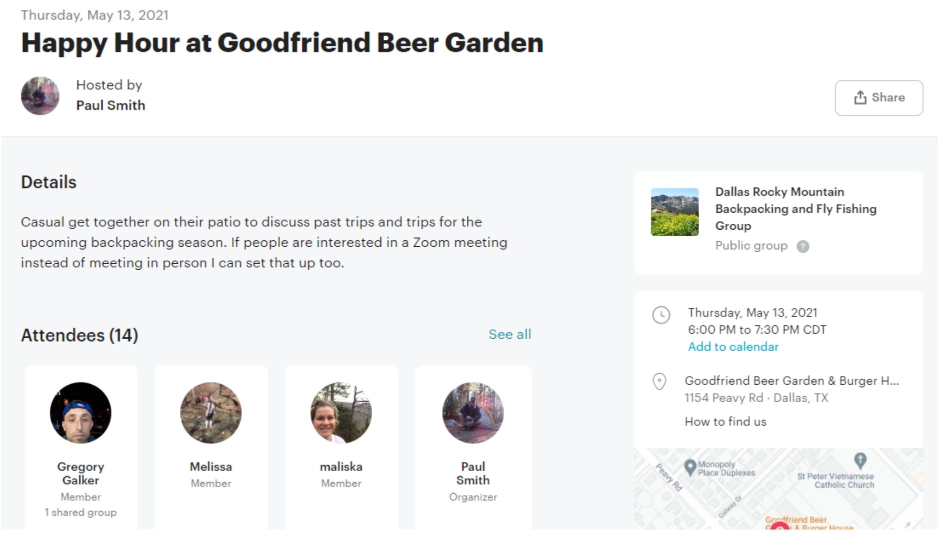

Fig.1. Example of Social events records from meetup website

shows the spatial distribution of seven types of social events within the DFW area. We used kernel density estimation to model the spatial distribution of event intensity. For instance, most types of events were concentrated in north Dallas and the downtown area. Figure 2D shows the counts of social events within 100 meters buffer of correspondent streets. Arlington was considered a secondary place with vigorous human activities.

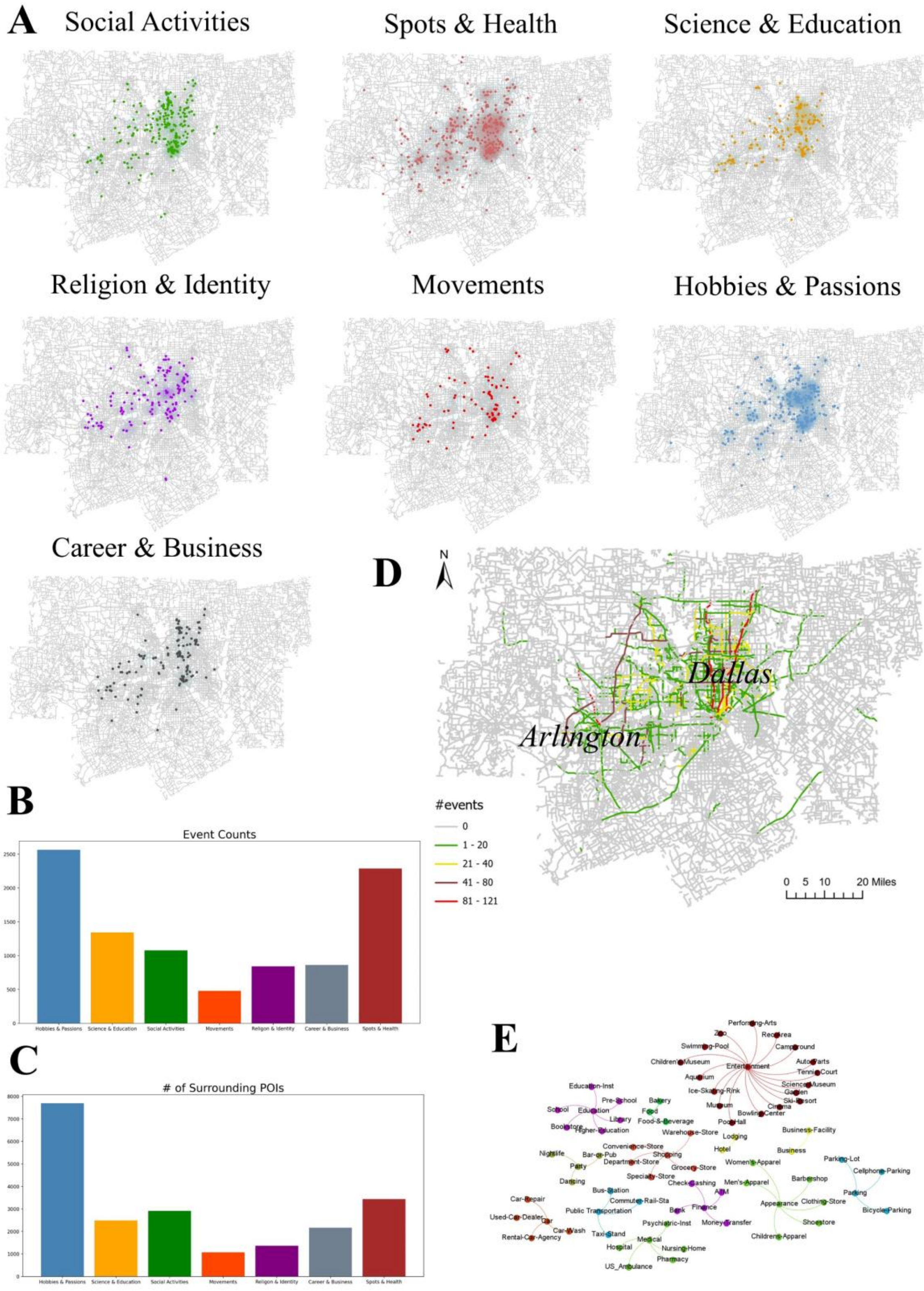

Fig.2. Social event types and POIs in DFW area in 2020. (A) Spatial distribution of seven event types. (B) Number of events of seven types. (C) Number of POIs within 50 meters from social events (D) street-level events number (sum of all types) (E) POIs reclassify rules

We obtained 115,877 POIs in 62 categories from Maptitude in June 2020. We simplified them into 13 categories (Figure 2E). Figures 2B and 2C summarized the numbers of seven types of social events versus the numbers of POI types within a 50-meter buffer from events, respectively. *Hobbies & Passions* events exhibited a strong dependency on selective POI types. People may prefer marketplaces or shopping centres with intense clusters of POIs to host this kind of events. Additionally, the summaries imply that the spatial associations among POIs and events vary from type to type.

A-priori Algorithm

A-priori (or apriori) algorithm is popular for association rules mining in the form of X(antecedent) $\rightarrow$ Y(consequent): if X is included in the transaction, there is a high probability that Y (an item or a list of items) is also present in that transaction (Agrawal et al., 1993). The algorithm has been widely used in many problem domains, including commercial behaviour forecasting, recommendation systems for purchasing or tourism, Flood area estimating, traffic accident prediction and crime analysis (Guo et al., 2017; Harun et al., 2017; Kong et al., 2021; Panjaitan et al., 2019; Ramasubbareddy et al., 2020) Apriori algorithm treats data as a set of items with attribute-value pairs, and identifies association rules that satisfy two requirements: minimum support and minimum confidence. *Support* is the indicator that describes how often the selected item appears in the dataset (Agrawal, Imieliński, et al., 1993), which is defined as

$$Support(X) = \frac{Number\ of\ transactions\ in\ which\ X\ appears}{Total\ number\ of\ transactions} = P(X)$$

And *confidence* represents when transactions have item X, what is the proportion of them also contain Y (Agrawal, Imieliński, et al., 1993)? It could also be interpreted as a conditional probability

$$Confidence(X \rightarrow Y) = \frac{Support(X \cap Y)}{Support(X)} = P(Y|X)$$

Researchers have proposed numerous interestness metrics from different considerations to assert the strength of association rules. *Lift*, one of the most popular metrics in recent publications, is designed to test the relationship between items against the independence assumption. If its value is equal to 1, there is no association between the two items; greater than 1, positively dependent; otherwise negatively dependent (Brin et al., 1997). *Lift* is defined as

$$Lift(X \rightarrow Y) = \frac{Support(X \cap Y)}{Support(X) * Support(Y)} = \frac{P(AB)}{P(A)P(B)}$$

Moreover, many studies used Lift equal to 1.2 as the breakpoint value for strong association rules.

Spatial Association Mining

Spatial statisticians have applied apriori algorithm for spatial association rules mining for years. However, some studies overlook some intrinsic properties of the apriori algorithm.

Critical to applying the apriori algorithm for spatial association mining is how to generate the *transactions*. *Transactions* are sets of records, and each *record* consists of item sets.

In our research, a *record* could be {*Appearance*, *Medical*, *Hobbies & Passions*}, which indicates an example of collocation of *Hobbies & Passions* event instances co-located with *Appearance* and *Medical* POI instances. How to define collocation is detrimental to spatial association mining. A popular way is to build a buffer from the interesting events and count the elements within this buffer to generate a record (Figure 3 left). This process

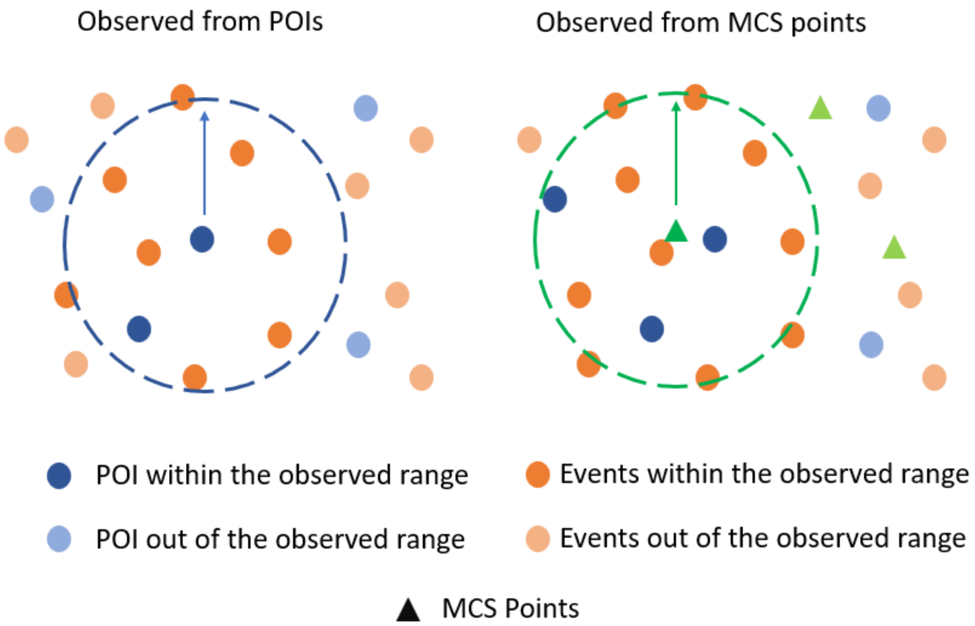

Fig.3. Transaction generating with Monte Carlo Simulation

may distort the intrinsic property of metrics and lead to unaware bias. For example, $Support(S_j)$ should represent the probability of transactions having the specific type of event, which calculated by $\#\, of\ trasactions\ contain\ S_j\ /\ \#\, total\ transactions$. However, since all transactions were simulated from POI, $Support(S_j) = \frac{\#\, having\ S_j}{\#\, having\ POI} = Confidence(POI \rightarrow S_j)$, which is inappropriate.

To avoid this problem, we employed Monte Carlo Simulation (MCS) to generate random points and observe the distribution of POIs and social events from them. Also, since all MCS points are randomly generated, is no need to worry about autocorrelation effects as measurements are taken at random locations (Figure 3 RIGHT).

Preliminary Results and Conclusion

We generated 15,000 MCS points and built 200 meters buffer (the average length of the road segment in the DFW area) to determine collocation in forming spatial *transactions*. We set the $Support_{min} = 0.0005$, specifying that the minimum number of observations contains both specific types of POIs P_i and social events S_j was 8. In our dataset, the background probability of observing a social event from a 200-meter random buffer was 0.09% (i.e., $P(S) = 0.09\%$). Therefore, $Confidence_{min}$ were set to 1% ($Prob(S_j|P_i) \geq 1\%$).

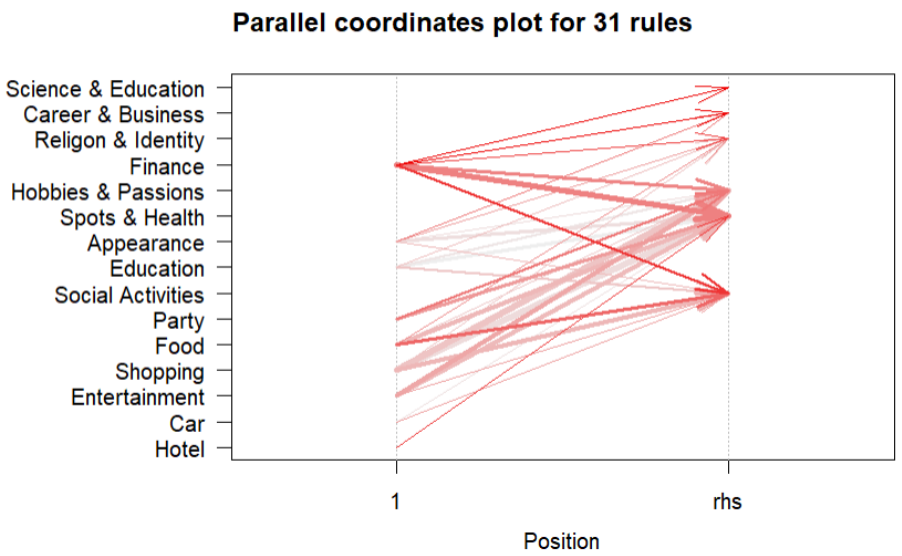

Fig.4. Graph-based parallel coordinate for 31 rules, width of arrow: *support* (0.05%-0.07%), colour: *confidence* (1.07%-4.26%), *lift*: 6.8-24.45

Figure 4 shows the qualified associations between POI types and social events. The left side of the arrow represents the antecedents, and the right side is consequences. Many POIs hosted many event types, suggesting multiple social functions of places. For example, Food (Bakery, Food & Beverage) places supported social activities (Social, Language & Culture, Pets, Family). Finance (Check-Cashing, ATM, Bank, Money-Transfer) places service multiple event types, including Science & Education, Career & Business, Religion & Identity, Hobbies & Recreations, and social activities. Also, one event type takes support from multiple POI types. For instance, social activities are associated with Food, Finance, Car (car repair, car dealer, rental car agency, car wash), and Entertainment.

This study explores the answer to "where social events happened" from the built environment perspective. The findings, while preliminary, suggest the social-spatial associations in the human-environment interactions. A better understanding of the geographic triad of social events, people, and places can inform community planning for improved quality of life. In the future, we will use street image views of POIs to investigate how the physical environment influences humans' perception and lead to different social behaviours.

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association in Large Databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data - SIGMOD '93*, 207–216.

Brin, S., Motwani, R., & Silverstein, C. (1997). *Beyond market baskets*. 265–276. https://doi.org/10.1145/253260.253327

Disaster Risk Reduction for the Built Environment | Wiley. (n.d.). Retrieved April 22, 2021, from https://www.wiley.com/en-us/Disaster+Risk+Reduction+for+the+Built+Environment-p-9781118921494

Guo, Y., Wang, M., & Li, X. (2017). Application of an improved Apriori algorithm in a mobile e-commerce recommendation system. *Industrial Management and Data Systems*, *117*(2), 287–303. https://doi.org/10.1108/IMDS-03-2016-0094

Harun, N. A., Makhtar, M., Aziz, A. A., Zakaria, Z. A., Abdullah, F. S., & Jusoh, J. A. (2017). The application of Apriori algorithm in predicting flood areas. *International Journal on Advanced Science, Engineering and Information Technology*, *7*(3), 763–769. https://doi.org/10.18517/ijaseit.7.3.1463

Kirmayer, L. J., & Whitley, R. (2009). Community Resilience: Models, Metaphors and Measures. In *Journal of Aboriginal Health*.

Kong, X., Das, S., "Tracy" Zhou, H., & Zhang, Y. (2021). Patterns of near-crash events in a naturalistic driving dataset: Applying rules mining. *Accident Analysis and Prevention*, *161*(August), 106346. https://doi.org/10.1016/j.aap.2021.106346

Metaxa-Kakavouli, D., Maas, P., & Aldrich, D. P. (2018). How social ties influence hurricane evacuation behavior. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–16. https://doi.org/10.1145/3274391

Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche, K. F., & Pfefferbaum, R. L. (2008). Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. *American Journal of Community Psychology*, *41*(1–2), 127–150. https://doi.org/10.1007/s10464-007-9156-6

One Million Meetup Events Hosted Online. (n.d.). Retrieved May 5, 2021, from https://www.prnewswire.com/news-releases/one-million-meetup-events-hosted-online-301160106.html

Panjaitan, S., Sulindawaty, Amin, M., Lindawati, S., Watrianthos, R., Sihotang, H. T., & Sinaga, B. (2019). Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns. *Journal of Physics: Conference Series*, *1255*(1). https://doi.org/10.1088/1742-6596/1255/1/012057

Ramasubbareddy, S., Aditya Sai Srinivas, T., Govinda, K., & Manivannan, S. S. (2020). Crime Prediction System. *Lecture Notes in Networks and Systems*, *103*, 127–134. https://doi.org/10.1007/978-981-15-2043-3_16