

# If You've Got the Money, I've Got the Time: Spatio-Temporal Footprints of Spending at Sports Events on Foursquare

Desislava Hristova,<sup>1</sup> David Liben-Nowell,<sup>2</sup> Anastasios Noulas,<sup>3</sup> and Cecilia Mascolo<sup>1</sup>

<sup>1</sup> Computer Laboratory, University of Cambridge, UK

<sup>2</sup> Department of Computer Science, Carleton College, USA

<sup>3</sup> Data Science Institute, Lancaster University, UK

## Abstract

Notable events such as sports games have a highly visible impact both in the media and at the event location itself, but the pre- and post-game activities of attendees—and their spending around such events—are much less evident. In this work, we use social media footprints to analyze and interpret these broader ripples of activity: specifically, we examine the temporal, spatial, and microeconomic patterns of baseball and basketball game attendees' activity around game events for 21 North American cities and 32 teams during 2010–2014, using mobility and price data from the location-based social network Foursquare. We identify some universal spatio-temporal patterns that hold across both sports and cities: exploiting additional information about venues' price levels, we find that, in aggregate, users who check in later at the game come from pricier restaurants and bars. We use these insights to predict the costliness of a user's choice of restaurant/bar, based only on a single transition to or from the stadium/arena, using the temporal properties of the transition such as the time of check-in at the game, distance traveled, and speed of transition, with the last being the most important factor. Our analysis carries broad implications for human mobility studies, mobile applications, and sports economics.

## Introduction

The impact of event spectatorship, which can be felt in traffic congestion, the local economy (Siegfried and Zimbalist 2000), and social media (Kwak et al. 2010), is all a result of the collective behavior patterns of individual spectators around event times. These pre- and post-game activities carry valuable insight about the spatio-temporal dynamics of events and their microeconomic footprints, which are the result of thousands of human choices about how to spend time and money. When Benjamin Franklin wrote in his essay *Advice to a Young Tradesman* “remember time is money,” he was referring precisely to these trade-offs, which define our priorities, values, and ultimately our socioeconomic context.

In this work, we use social media data to explore the temporal, spatial, and economic patterns of the participants at and around a particular class of events—namely, sports events in 21 North American cities during 2010–2014, covering a total of 32 baseball and basketball teams. Sports

teams have a strong following, wide cultural impact, and multi-billion dollar market valuations. Despite the importance of sports events, however, there is little data-driven insight on the behavior of event attendees, such as their spending before and after the game. By complementing user mobility patterns with place–price data from the location-based social network Foursquare, we discover universal temporal and spatial patterns of behavior across both sports and many cities. These observations allow us to predict the price level of a restaurant/bar that a user visits before or after the game solely from the mobility information from that user's single check-in at the stadium.

Sports games, of course, form only a narrow slice of the kinds of events with accompanying ripples of social and economic activity, but sports are a particularly appealing domain for our analysis, for three important reasons. First, a game has a clear start time, a clear finish time, and a precise geographic location. Second, games are big events, typically bringing tens of thousands of fans to the stadium. Finally, and most importantly, baseball and basketball games are largely symmetric; we are able to *aggregate* an entire season's games at a stadium (41 or 81 games, depending on the sport) by shifting each game's start time and overlaying fan behavior from all games at a single stadium.

Similar to work in behavioral economics and marketing psychology that has shown a relationship between higher spending levels and time spent at retail shops (Duncan Herrington and Capella 1995; Iyer 1989; Milliman 1982; Park, Iyer, and Smith 1989, for example), we discover that users who transition at slower speeds to and from the sports stadium tend to visit pricier Foursquare venues before/after the game, as do those who check in at the stadium later into the game. We further formulate a user-level classification task, where we show that, across cities and for both basketball and baseball, speed is a better predictor than check-in time or distance alone when predicting the spending of a user. In short, those who seem to place lower importance on time at the stadium (by moving more slowly from a pre-game venue or by being willing to miss more of the game) also seem to spend more money outside the stadium. Understanding how human mobility patterns around such events relate to microeconomic behavior is of consequence to location-based advertising and to measuring the economic impact of sports events to the local community.

## Related Work

**Sports Economics** Decades of research have been dedicated to the prediction of the outcomes of sports games (Glickman and Stern 1998; Stefani 1980; Zuber, Gandar, and Bowers 1985), with not as much work focusing on attendees’ behavior and experiences at a large scale. It has been suggested that sports stadiums and teams have the ability to boost the local economy and benefit the community through attracting visitors, but this claim has been met with much criticism from researchers (Crompton 1995; Siegfried and Zimbalist 2000). Undoubtedly, however, there is a strong connection between sports and the communities that support them, particularly in the urban context (Riess 1991). In terms of a temporal budget, attending sporting events is time allocated to discretionary activities, which may vary from individual to individual and can be valued by the opportunity cost of spending time on the next best thing (Okada and Hoch 2004). In terms of financial budget, sport event spectatorship has been associated with socioeconomic status (Wilson 2002). While the present work does not focus on the demographics or income levels of event attendees, it builds on a complementary thread where the aim is to understand how the dynamics of spatio-temporal user behavior during sport events reflects spending behavior.

**Events and Social Media** Event detection has been a popular application of the copious amounts of data generated collectively by social media users, where real-world events such as earthquakes and local gatherings can be identified (Sakaki, Okazaki, and Matsuo 2010; Becker, Naaman, and Gravano 2011). Location-based services such as Foursquare have been exploited as a medium to track real world events spatio-temporally; for instance, Sklar, Shaw, and Hogue (2012) demonstrate a framework for online event detection. Georgiev, Noulas, and Mascolo (2014a) investigate the task of event-to-user recommendation in Foursquare, continuing on the lineage of related work in the field where cellular mobility data has been exploited (Quercia et al. 2010; Xavier et al. 2013; Calabrese et al. 2010). Even closer to the spirit of the present paper is the work of Georgiev, Noulas, and Mascolo (2014b), who analyze implicitly the financial impact of a large sports event, the London 2012 Olympics, by tracking foot traffic at nearby food establishments by means of Foursquare check-ins. In this work we exploit the inherent advantage of location-based services, where semantic geographical information is made available by user activity and therefore aligning this data with sporting event schedules gives us a nearly noiseless view of the digital footprints of these events.

## Dataset

Our analysis is based on a four-year longitudinal dataset (January 2010–December 2014) for 21 cities in North America. The data was provided directly by Foursquare for research purposes. This data includes anonymized user transitions either from a venue A to the city’s baseball or basketball stadium (*incoming transitions*), or from the stadium to a venue B (*outgoing transitions*). Each transition lists (i) the venue identifier, (ii) the venue check-in time, and (iii) the

stadium check-in time; for privacy reasons, the resolution of the check-in times was artificially decreased to decouple records, and the dataset includes only transitions whose inter-check-in times are  $\leq 3$ hr. Thus each transition is made by a single user, but users’ transitions cannot be traced beyond one hop, and therefore our analysis reflects crowd behavior and not individual patterns.

The set of public Foursquare venues in a city (i.e., not residences) consist of semantic locations that have more than 100 check-ins (visits). We use this threshold to avoid the use of private venues or of venues that have been added either mistakenly or maliciously. For each venue, we know its position in the city (its geographic coordinates) and its type (Food, Nightlife Spot, Travel & Transport, Hotel, etc.). Additionally, we used the Foursquare API (<https://developer.foursquare.com/>) to collect pricing information about all relevant venues, if such information is available. Most bar/restaurant venues in Foursquare have been associated with a price band ranging from 1 (least expensive) to 4 (most expensive).

To perform temporal analysis on sporting events, we crawled the schedules of basketball and baseball games from ESPN ([espn.com](http://espn.com)) to extract information about the start times of each game during the 2010–2014 seasons. We collected information about 32 teams—16 baseball teams, 16 basketball teams—and all their home games in the time period. For each game, we take a three-hour temporal window around the start of the game, where we consider transitions whose origin or destination was the stadium in question during that time. In this way, we filter out transitions that are unrelated to the given game.

In total, we have 413,603 transitions, of which 288,810 are transitions to the stadium (incoming), and 124,793 are from the stadium (outgoing). We have significantly fewer check-ins for basketball games (29K total), than baseball games (384K total), which is in part due to the shorter basketball season (fewer games) and smaller basketball arenas (fewer attendees per game).

## Spatial and Temporal Properties of Sports Events on Foursquare

The first aspect of our work involves quantifying the temporal and spatial dynamics of attendees at basketball and baseball game events. First, we consider check-in behavior on Foursquare around baseball and basketball games looking at check-in times relative to the game start time. We take into account all check-ins that are made at a team’s stadium within three hours of the start of any home game at that stadium, collapsing together on the temporal axis all games played at that stadium. We also consider all incoming and outgoing transitions at the stadium, examining the check-in times at venues outside the stadium.

In Fig. 1 we plot the probability distributions of check-ins relative to game time, where 0 indicates the start of the game. Although stadium check-in times do not directly correspond to stadium arrival times, as users might wait for some unspecified period after their arrival before they choose to check in, stadium check-ins predominantly occur at the start

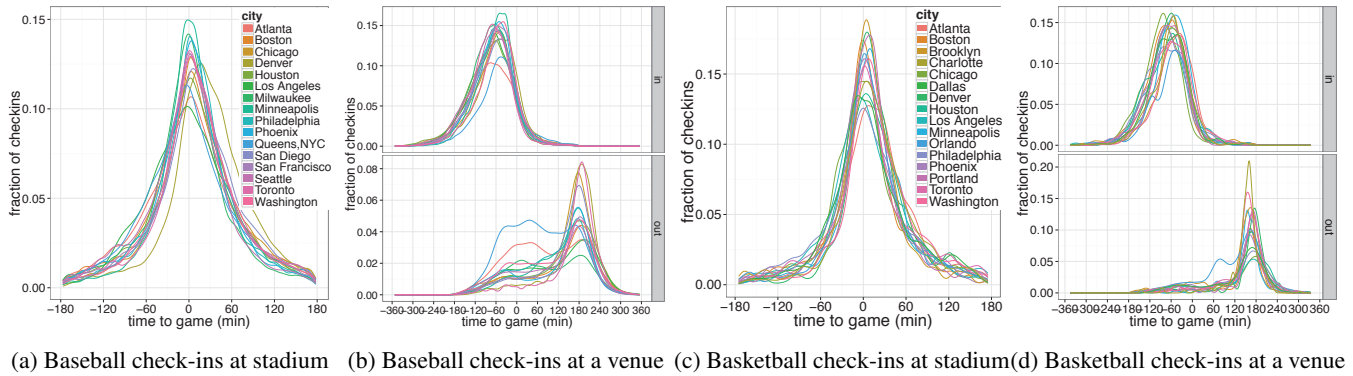


Figure 1: Probability distributions of check-ins relative to game time, where 0 indicates the start of the game. Plots **at the stadium** show the fraction of check-ins at the respective stadium peaking at the start of the game, while the plots **at a venue** show the fraction of check-ins made at venues before (in) and after (out) the stadium. All plots use a kernel density estimator over all cities for 20min intervals. All times are standardized to UTC.

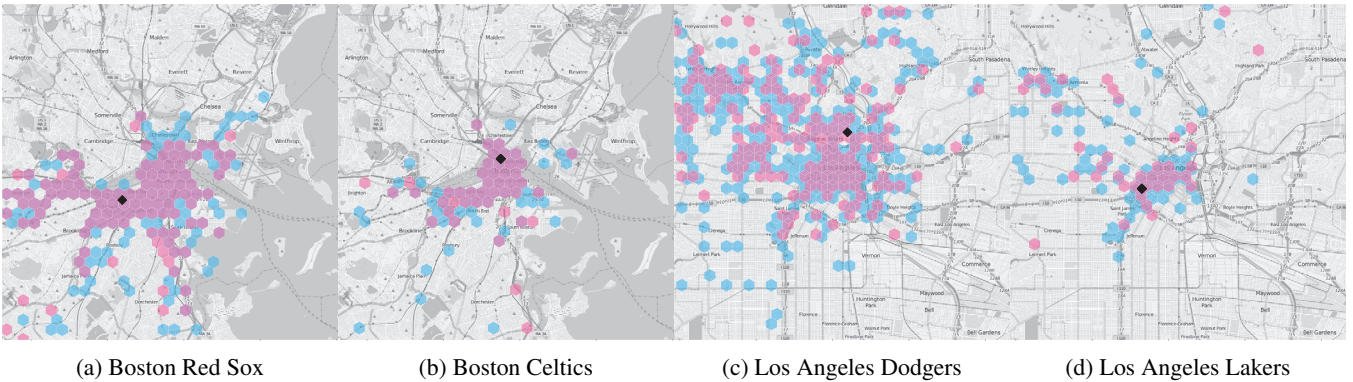


Figure 2: Dispersion of incoming and outgoing transitions per city. The color of the hexagon indicates substantial incoming (blue) or outgoing (pink) transitions; purple indicates that both are represented. The location of the stadium is indicated in black. Each shaded hexagon represents 40 or more check-ins. For space considerations, we only plot two of the cities that have both a basketball and baseball team.

of a game. As Fig. 1a and 1c demonstrate, 15–20% of attendees check in within 10 minutes of the game start. However, the rest of the attendees check in with approximately symmetrical probability up to 3 hours before and after the game start. Next, in Fig. 1b and 1d, we observe the probability distributions of check-in times at venues visited prior to (in) and after (out) the game, respectively. Incoming transitions (top panel) peak about one hour prior to game start, whereas outgoing transitions peak around three hours after the game start. Note that the distribution for baseball games is broader in comparison to the basketball case, as games are about 30–60 minutes longer. What is most striking about Fig. 1 as a whole is the universal temporal patterns that persist regardless of city and sport. These patterns are indicative of event-driven synchronicities influencing the behavior of mobile users resembling observations previously made in other social contexts (Yun, Watanabe, and Shimojo 2012).

The second fundamental property of pre- and post-game transitions we consider, in addition to time, is the distance from the stadium. Fig. 2 shows the geographic dispersion of incoming and outgoing check-ins in two cities for all baseball and basketball games in aggregate, as well as the loca-

tion of the stadium itself. Distances across both sports tend to be quite short: all median distances traveled were computed to be 2.5km or less, which is reflected in the concentration of activities close to the stadium in Fig. 2. We note some variation between the geographic dispersion of incoming and outgoing transitions for both sports. In the case of L.A., we can visibly see on the map that baseball transitions tend to be more spread out than basketball transitions. This pattern could be an artifact of the lower volume of data for basketball events (noted in the dataset section); however, in the case of Boston we can see that distances are very similar for both sports (median distances were computed to be below 1km for both). What seems to emerge as a factor, then, is the morphological character of the city itself and not simply differences between basketball and baseball events. (Of course, differences between the sports certainly do exist, so we consider each sport separately in our analysis.) Overall, we note that there is a lot of variation at short distances between cities, which differ qualitatively in this respect, contrary to the shared patterns of temporal properties in Fig. 1. We investigate the role of time and distance, along with transition speed in relation to spending, in the following section.

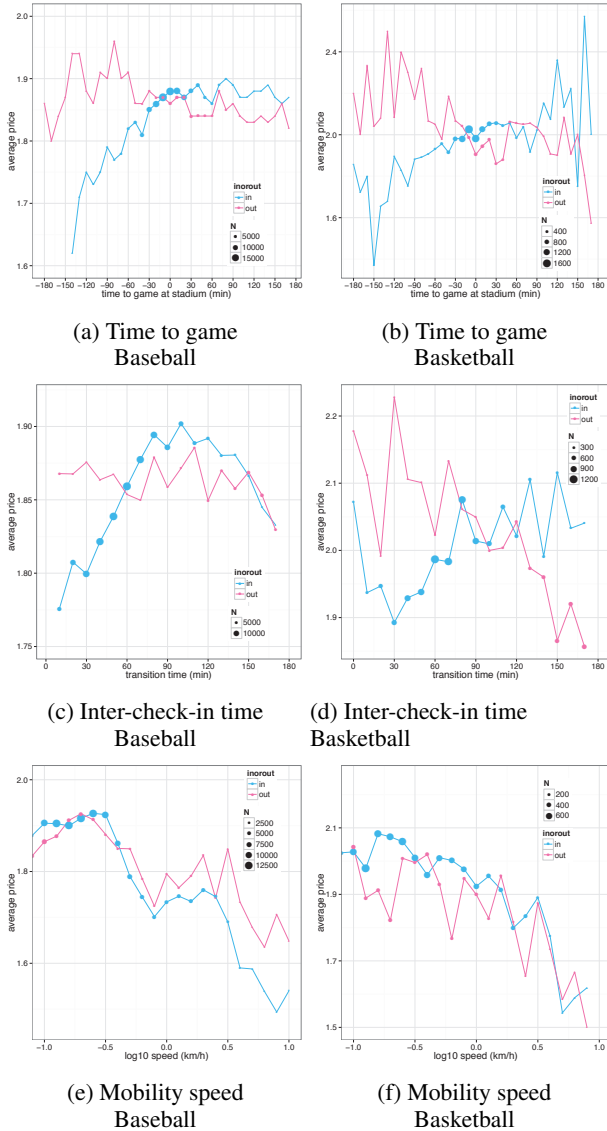


Figure 3: Average price aggregated over spatio-temporal properties of transitions. The size of a node indicates the number of observations it represents.

### Time, Space and Money

In this part of our analysis, we introduce the price property of venues to and from which users transition. We download price information from the Foursquare API to associate each incoming and outgoing transition with a relative monetary value. Only Food and Nightlife venues carry price information, but these venues represent the majority of transitions in our dataset (60% of all transitions in both baseball and basketball). Prices are discrete values on a scale from 1 (cheapest) to 4 (most expensive). In Fig. 3 we plot the relationship between the average price of a venue that a user has visited before or after the stadium against the spatio-temporal properties of their transitions.

First we look at the stadium check-in time for a user and

the average price tier for users who have checked in at that time (Fig. 3a for baseball and Fig. 3b for basketball). We note that although the signal for basketball is weaker, those users who check in well before the start of the game tend to spend much less than those who check in after the start of the game. The most distinct differences can be noticed within 30 minutes of game start, where those who check in about 30 minutes before the game start have spent less on average than those who check in up to 30 minutes after. This relationship is not as strong for outgoing transitions, where price is consistently in the higher range. We note that across all figures, spending is always more on average at basketball games than baseball games, perhaps due to the position of stadiums or the difference in demographics of fans.

We further explore the relationship between time and spending in Fig. 3c and 3d, where the average price tier of visited venues for users who made a transition in  $t$  time are shown for both sports. In particular for baseball we can observe a strong positive relationship between the inter-check-in time and the average price for incoming transitions, which roughly translates to the generalization that *the longer you stay, the more you pay*. In other words, those who take longer to transition from a venue to the stadium visit more expensive venues than those who transition faster to the stadium (less than 30 minutes). The same trend can be observed for basketball games but again it is more noisy. In both sports, outgoing transitions exhibit much less clear behavior, but the trend tends to the reverse—the longer you have stayed at the stadium the less expensive bar/restaurant you are likely to visit afterwards.

Finally, we look at transition speed in Fig. 3e and 3f, which relate both time and distance against the average price of a venue from which users transition. The *speed of a transition* is the distance between two consecutive locations divided by the difference in the check-in times at those locations (the inter-check-in time). Consequently, the speed of human mobility is not equivalent to the travel speed alone. It additionally includes implicit information about the duration of stay at a place, assuming that users check in while at the venue and not exactly when they depart from one location and arrive at the next. We can observe a relationship between average price of a visited venue and spatio-temporal properties of user behavior at events, where *the faster you transition, the less money you spend*. This pattern holds for both incoming and outgoing transitions, where the biggest drop in spending seems to happen after a speed of  $10^{-0.5} \text{ km/h}$  for both sports. With this analysis we have shown that there is a clear relationship among time, distance, and spending that has previously not been observed in this context. We leverage these insights next to predict the amount of money a sports fan spends before or after a game based on information from a single transition.

### Predicting Spending at Sports Events

Our goal thus far has been to understand how spending around sports events can be captured by the spatial and temporal behavior of attendees. We have observed that indeed there is a relationship between the way that people spend time and money, generally—the *more expensive a venue you*

visit is, the more time you spend there, the slower you transition, and the later you check in at the stadium for the game. To formalize this idea, we frame a binary classification task, where we have two classes of spending, class 1—venues with the lowest spending tier—and class 0—venues which are pricier. We use the five variables we have introduced in the previous sections as predictors: time of check-in at stadium, time of check-in at venue, the transition time, distance from stadium, and transition speed. We then rank the list of candidate venues based on the values of each predictor variable. Subsequently, for each variable we vary the prediction list threshold that discriminates the two classes and then we compute the true-positive-to-false-positive ratio at each step. This allows us to measure the area under the curve score (AUC) for these ROC curves which provides indication of how powerful a predictor is, independent of class distribution. A random baseline predictor would always produce a AUC score of 0.5, and therefore AUC scores higher than 0.5 are considered to be better than random. We do not use a multivariate model or supervised learning methods but rather the raw variables only, in order to capture easy-to-describe patterns of human behavior with a focus on interpretability, acknowledging that more sophisticated methods would likely yield higher scores.

In Fig. 4 we report the results of the binary prediction task by plotting the distribution of AUC scores across cities for the different spatio-temporal predictor variables. We compute the area under the curve for each city and sport as well as for incoming and outgoing transitions separately based on the cross-group differences we observed in the previous analysis. The highest AUC scores come from the speed and distance predictors across cities, where higher speeds and distances can be associated with lower spending (class 1), whereas lower speeds and shorter distances can be associated with higher spending (class 0). About 25% of cities in both sports (top quartile) have AUC scores higher than 0.75 when using speed as a predictor. Results for incoming transitions are higher than outgoing transitions for both sports, and the reason is the less conclusive behavior of users of their way out of the stadium which we have seen in the previous sections. The next best performing indicator is the transition time (tt), which has comparable AUC median score with speed and distance for incoming transitions, and in fact has lower variability, but is much less reliable for outgoing transitions. The same can be seen for the time at stadium (ts) but only in the case of baseball games; for basketball events spending in outgoing transitions is better represented by time at stadium. Time at venue (tv) as an indicator performs equally for incoming and outgoing transitions based on the median result for both sports, which could be a good option if spending prediction needs to be reliable in both directions using a single indicator, although with lower AUC scores than if using different indicators.

Overall, with the above prediction task we have formalized our observations from previous sections, namely the relationship between spending time and spending money around sports events, using a binary prediction task with simple threshold varying. We have shown that using basic prediction methods, we can confirm that speed, which is de-

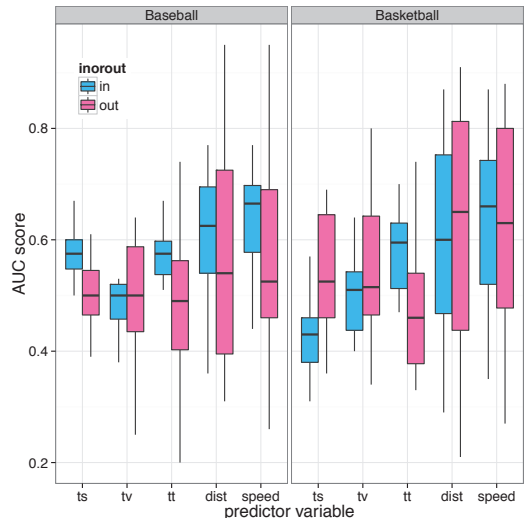


Figure 4: AUC scores distribution of spending predictors in baseball and basketball: time at stadium (ts), time at venue (tv), transition time (tt), distance (dist) and speed. We vary the threshold over each predictor and compute the AUC score for the True to False Positive Rate for each city and plot the distribution of those AUC scores here for incoming (in) and outgoing (out) transitions.

rived from two other mobility indicators (transition time and distance), is a good indicator of spending of sports fans. Additionally, we have shown how other indicators such as transition time alone and time at venue can be useful in different scenarios.

## Discussion & Conclusions

With this work, we have presented a novel case study of individuals’ spending around sports events. Its implications are wide for mobile applications, sports economics, and the advancement of our understanding of large-scale human behavior. We have used real-world sporting events in basketball and baseball in the 2010–14 period and have intersected them with human mobility data from the location-based application Foursquare. This has allowed us to demonstrate the similarities in temporal dynamics across cities and sports in contrast with spatial properties such as distance that seem more dependent on the geographic context. The speed of human transitions between the stadium and other venues, derived from the spatial and temporal properties of the data, emerged as a strong indicator of individual spending leading to the conclusion that *the more money one spends at a given venue, the more time one spends at that venue and transitioning between it and the stadium*. Similar relationships between time and spending have been observed in retail settings (Duncan Herrington and Capella 1995; Iyer 1989; Milliman 1982; Park, Iyer, and Smith 1989), where it has been shown that more time spent in a shop is associated with higher spending. This positive relationship between time and money implies that “time is money” in a quite literal sense.



The scale of our observations, as well as the minimal human mobility data we use for prediction, points to a big opportunity gap for sporting event management despite limitations we have noted throughout this work. With the continuous criticism of the investment of public money into the construction of new stadiums and often questioned benefits to the local economy (Crompton 1995; Siegfried and Zimbalist 2000), it is increasingly important to understand mobility and estimate spending around such events. We have outlined a simple and clear methodology for doing so, using as little as a single data point to predict the potential amount spent by a visitor. Although we only have transitions of up to one hop, we have observed that, in fact, in many cities those transitions (especially outgoing) are local and at a short distance, which suggests that there is spending flowing into local bars and restaurants. Nevertheless, we have also observed that this pattern is greatly dependent on geographical factors unique to cities.

As a result of the above implications, our analysis is also highly relevant to the improvement of location-based mobile applications, in particular for location-based advertisements and recommendations. Being able to segment users into high and low spenders based on simple spatio-temporal behavior with minimal data requirements can greatly aid the recommendation of venues and budget relevant location-based advertisements from marketers. In addition, the spatio-temporal dynamics of users can help predict congestion and support the advancement of “smart stadiums” (Ranadivé 2016), where users can experience the game in new ways through pervasive technology. Ultimately, our work introduces new routes for event-based microeconomic analysis using social media data, and we hope that it will serve as inspiration to the community for future research into this exciting new direction.

**Acknowledgements.** This work was supported by the EP-SRC through Grant GALE (EP/K019392), Carleton College, and Churchill College.

## References

Becker, H.; Naaman, M.; and Gravano, L. 2011. Beyond trending topics: Real-world event identification on Twitter. In *Proc. 6th International Conference on Weblogs and Social Media (ICWSM)*.

Calabrese, F.; Pereira, F. C.; Di Lorenzo, G.; Liu, L.; and Ratti, C. 2010. The geography of taste: Analyzing cell-phone mobility and social events. In *Proc. 8th International Conference on Pervasive Computing*, 22–37. Springer.

Crompton, J. L. 1995. Economic impact analysis of sports facilities and events: Eleven sources of misapplication. *Journal of Sport Management* 9(1):14–35.

Duncan Herrington, J., and Capella, L. M. 1995. Shopper reactions to perceived time pressure. *International Journal of Retail & Distribution Management* 23(12):13–20.

Georgiev, P.; Noulas, A.; and Mascolo, C. 2014a. The call of the crowd: Event participation in location-based social services. In *Proc. 8th International Conference on Weblogs and Social Media (ICWSM)*.

Georgiev, P.; Noulas, A.; and Mascolo, C. 2014b. Where businesses thrive: Predicting the impact of the Olympic Games on lo-

cal retailers through location-based services data. In *Proc. 8th International Conference on Weblogs and Social Media (ICWSM)*.

Glickman, M. E., and Stern, H. S. 1998. A state-space model for National Football League scores. *Journal of the American Statistical Association* 93(441):25–35.

Iyer, E. S. 1989. Unplanned purchasing: Knowledge of shopping environment and time pressure. *Journal of Retailing* 65(1):40.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is Twitter, a social network or a news media? In *Proc. 19th International Conference on the World Wide Web*, 591–600. ACM.

Milliman, R. E. 1982. Using background music to affect the behavior of supermarket shoppers. *Journal of Marketing* 86–91.

Okada, E. M., and Hoch, S. J. 2004. Spending time versus spending money. *Journal of Consumer Research* 31(2):313–323.

Park, C. W.; Iyer, E. S.; and Smith, D. C. 1989. The effects of situational factors on in-store grocery shopping behavior: The role of store environment and time available for shopping. *Journal of Consumer Research* 422–433.

Quercia, D.; Lathia, N.; Calabrese, F.; Di Lorenzo, G.; and Crowcroft, J. 2010. Recommending social events from mobile phone location data. In *Proc. 10th International Conference on Data Mining (ICDM)*, 971–976. IEEE.

Ranadivé, V. 2016. How the Sacramento Kings Are Creating The 21st Century Colosseum. <https://thecauldron.si.com/how-the-sacramento-kings-are-creating-the-21st-century-colosseum-665dbaa14efb>. Accessed: 2016-01-03.

Riess, S. 1991. *City Games: The Evolution of American Urban Society and the Rise of Sports*. An Illini book. University of Illinois Press.

Sakaki, T.; Okazaki, M.; and Matsuo, Y. 2010. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proc. 19th International Conference on the World Wide Web*, 851–860. ACM.

Siegfried, J., and Zimbalist, A. 2000. The economics of sports facilities and their communities. *Journal of Economic Perspectives* 95–114.

Sklar, M.; Shaw, B.; and Hogue, A. 2012. Recommending interesting events in real-time with Foursquare check-ins. In *Proc. 6th Conference on Recommender Systems*, 311–312. ACM.

Stefani, R. T. 1980. Improved least squares football, basketball, and soccer predictions. *IEEE Transactions on Systems, Man, and Cybernetics* 10(2):116–123.

Wilson, T. C. 2002. The paradox of social class and sports involvement the roles of cultural and economic capital. *International Review for the Sociology of Sport* 37(1):5–16.

Xavier, F. H. Z.; Malab, C.; Silveira, L.; Ziviani, A.; Almeida, J.; and Marques-Neto, H. 2013. Understanding human mobility due to large-scale events. In *Proc. 3rd International Conference on the Analysis of Mobile Phone Datasets (NetMob)*.

Yun, K.; Watanabe, K.; and Shimojo, S. 2012. Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific reports* 2.

Zuber, R. A.; Gandar, J. M.; and Bowers, B. D. 1985. Beating the spread: Testing the efficiency of the gambling market for National Football League games. *Journal of Political Economy* 800–806.