FourSquare App Checkin at UCLA
A Spatial-Temporal Analysis

Tim Chen

1 Introduction

In the contemporary landscape of mobile technology, Foursquare City Guide emerges as a pivotal tool for local exploration. Despite its potential under-recognition, notable digital and consumer brands such as Apple (for Maps), Uber, Snapchat, Spotify, and Coca-Cola rely on Foursquare for its robust location services technology and anonymized, aggregated datasets. Initially developed to offer tailored recommendations for nearby venues, Foursquare's dataset provides invaluable insights into human mobility patterns and spatial interactions. In this study, we embark on an investigation to unveil the temporal dynamics and spatial intricacies of user check-ins within the vibrant UCLA community. The rationale behind this analysis lies in the potential for leveraging spatial analysis of check-in data to enhance customer satisfaction, optimize operational efficiency, and ultimately bolster competitiveness and profitability within this specific locality.

2 Data

Our Research is made possible through a rich dataset provided by Dingqi Yang, obtained through the systematic scraping of Foursquare's data site. Spanning an extensive 18-month period from April 2012 to September 2013, this dataset encompasses over 33 million check-ins by nearly 267,000 users across 3.6 million venues, collected in 415 cities across 77 countries. All observations in the dataset is recorded when users access the company's platform, and voluntarily sharing their location information. The company's system captures data including the geographic coordinates, the timestamp, and additional metadata such as search history and user demographics. Our focus centers on the dynamic region surrounding the University of California, Los Angeles (UCLA) spanning the 18-month period. Within this localized context, we have a subset comprising 689 check-in points, enabling a detailed exploration of Foursquare check-ins within and around the UCLA campus((Lat,Lon)$\in [34.067089, 34.075976] \times [-118.445003, -118.439485]$) with google map view in Fig.1.

3 Cluster Detection

We first examine the spatial distribution of check-in locations by plotting their geographic coordinates. A scatterplot of latitude and longitude coordinates reveals the spatial layout of check-ins across the UCLA campus vicinity shown in Fig.2 . Each point on the scatterplot represents a unique check-in location, providing a visual depiction of user density and clustering patterns. Marked points are where multiple checkins are excluded, providing extra insights for subsequent spatial analysis.

3.1 Kernel Smoothing

Given the inherent complexity of the spatial dataset, characterized by non-simple features, a subsequent refinement process became necessary. To address this, we implemented kernel smoothing with quartic kernel after removing redundant points, thus enhancing the accuracy of our spatial representations. The resulting density map (Fig.3) revealed clusters of venues, highlighting areas of high activity and urban vibrancy within walking distance of the campus, especially path connecting campus to Westwood and to housing on the hills, where most students live and the path correspondingly reflects the route of commute.

3.2 Spatial Point Pattern Summary Functions

Further analysis using K and L function analysis confirmed the presence of significant spatial clustering of check-ins (Fig.4) . The discernible clustering patterns suggest that check-ins tend to concentrate in close proximity to each other. This phenomenon may be influenced by factors such as accessibility and the margin of error in GPS precision, given that Latitude/Longitude data is recorded with precision up to six decimal places. In both Marked G (Fig.5) and Marked J (Fig.6) function plots, the observed check-in data line deviates significantly from the theoretical line expected under a random Poisson process. This deviation indicates that the check-in locations are not randomly distributed, providing evidence for spatial clustering. To take a closer look at the campus map (Fig.7), we observed significant clustering of check-in locations around key landmarks within the UCLA campus, such as bus stops, student centers, and dining facilities. These areas exhibited higher than expected densities of check-ins, suggesting that they serve as focal points for campus activity and social interaction. Furthermore, spatial anomalies were detected, denoting locations characterized by notably elevated check-in counts. Upon closer examination, it was discerned that these anomalies stemmed from one single user repeatedly logging in over the 18-month period.

4 Model Fitting

4.1 Poisson Process Fitting

By estimating the parameters of the Poisson model, we aimed to identify areas of elevated or suppressed check-in activity and assess the significance of spatial covariates in explaining these patterns. The linear Poisson model was specified as $\lambda(x, y) = \alpha + \beta x + \gamma y$. After fitting the model to the check-in data, parameter estimates were obtained as follows: $\alpha =$ 4.552, $\beta = 0.300$, $\gamma =$ -0.475. The statistically significant coefficients indicate that both the x and y spatial covariates contribute to the intensity of check-ins, with higher values on the bottom right corner of the campus and decreasing along the SE-NW diagonal.

	μ	K	α	β
Estimates	1.32	0.924	0.945	1.45
Standard Error	0.839	0.214	0.813	0.309

Table 1: Hawkes Model Parameter fitting

The quadratic Poisson model was specified as $\lambda(s) = \alpha + \beta x + \gamma y + \delta x^2 + \epsilon xy + \zeta y^2$, with estimates: $\alpha = 3.916, \beta = 4.611, \gamma = -0.088, \delta = -5.965, \epsilon = 3.942, \zeta = -2.565$. The nonlinear relationships is captured in this model by highlighting the high intensity on the central ellipsoid area of the campus, with the major axis along the NE-SW diagonal. The fitted intensity of these two models are shown in Fig.8, and the corresponding super thinned plots using the 2nd order polynomial Poisson shown in Fig.9 appears to be random and shows the model seems to be a good fit.

4.2 Hawkes Model Fitting

In addition to the inhomogeneous Poisson model, we explored the applicability of the Hawkes triggering model, to account for possible self-exciting and mutually-exciting point process dynamics, which in our case may indicate friends checkin or multiple checkin while walking. This model is particularly well-suited for capturing the influence of past check-in events on future activity, as well as identifying spatial clusters of heightened event occurrence. Our model is represented by

$$\lambda(t, x, y) = \mu(x, y) + K \sum g_1(t - t_i) g_2(x - x_i, y - y_i)$$

where background rate is simply expressed as $\mu \frac{1}{xy}$, $g_1(t) = \beta e^{-\beta t}$, and $g_2(x, y) = \frac{\alpha}{\pi} e^{-\alpha(x^2+y^2)}$. The model fitting results presented in Table.1 reveal that large values of parameter K signify heightened triggering events, bolstered by corresponding standard errors that enhance our confidence in the model. Such a finding warrants scrutiny, implying potential unforeseen patterns or biases in the data collection process. Despite these concerns, the model underwent rigorous validation through superthinning analysis. Fig.10 visualizes the original and superthinned points, with the latter demonstrating adherence to a homogeneous Poisson process, affirming the appropriateness of the Hawkes model's fit.

5 Conclusion

In conclusion, our analysis of check-in data within the UCLA community has shed light on several key insights while also highlighting areas for improvement. Firstly, the temporal variability inherent in human activity patterns poses a challenge to our models' assumptions of stationarity. Secondly, the definition of spatial boundaries warrants further refinement to capture the full extent of user interactions. Additionally, the identification and incorporation of relevant covariates impacting the intensity parameter λ remain crucial for enhancing model accuracy. Future work can incorporate academic calendars into analyses, accounting for temporal fluctuations in check-in behavior, thus refining the temporal dimension of our models. Secondly, expanding the analysis regions beyond the immediate UCLA community could offer broader insights into spatial dynamics and interactions. Additionally, leveraging supplementary data sources such as WiFi heatmaps, building locations, and venue categorizations could enhance the granularity and contextuality of our spatial analyses. By addressing these limitations and embracing related improvements, our modeling framework's plausibility and utility in capturing and interpreting spatial clustering and temporal dynamics will be further strengthened, paving the way for more robust spatial analyses in the future.

References

[1] M. N. M. van Lieshout. *Theory of spatial statistics.* Routledge Cavendish, London, England, Mar. 2019.

[2] D. Yang, D. Zhang, L. Chen, and B. Qu. Nationtelescope: Monitoring and visualizing large-scale collective behavior in lbsns. *Journal of Network and Computer Applications*, 55:170–180, Sept. 2015.

[3] D. Yang, D. Zhang, and B. Qu. Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Transactions on Intelligent Systems and Technology*, 7(3):1–23, Jan. 2016.

Appendix

Figure 1: GoogleMap Area of Interest at UCLA

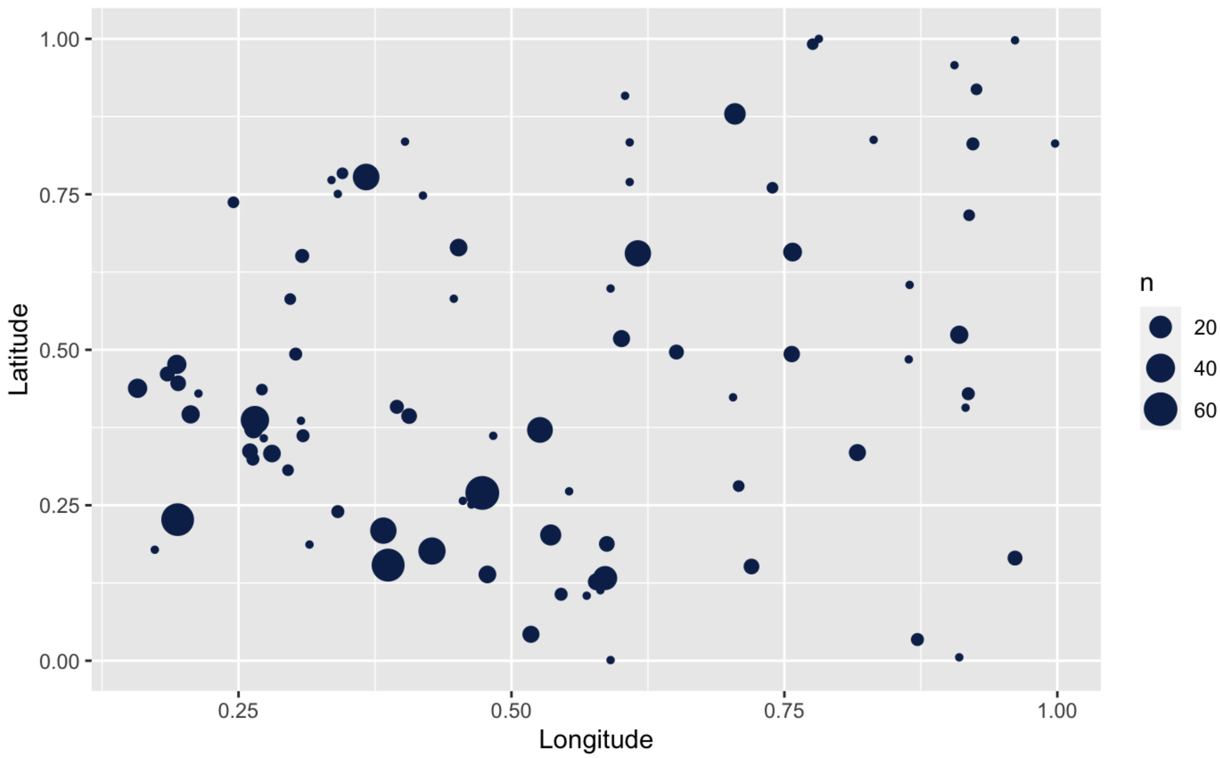

Figure 2: Marked Scatter Plot

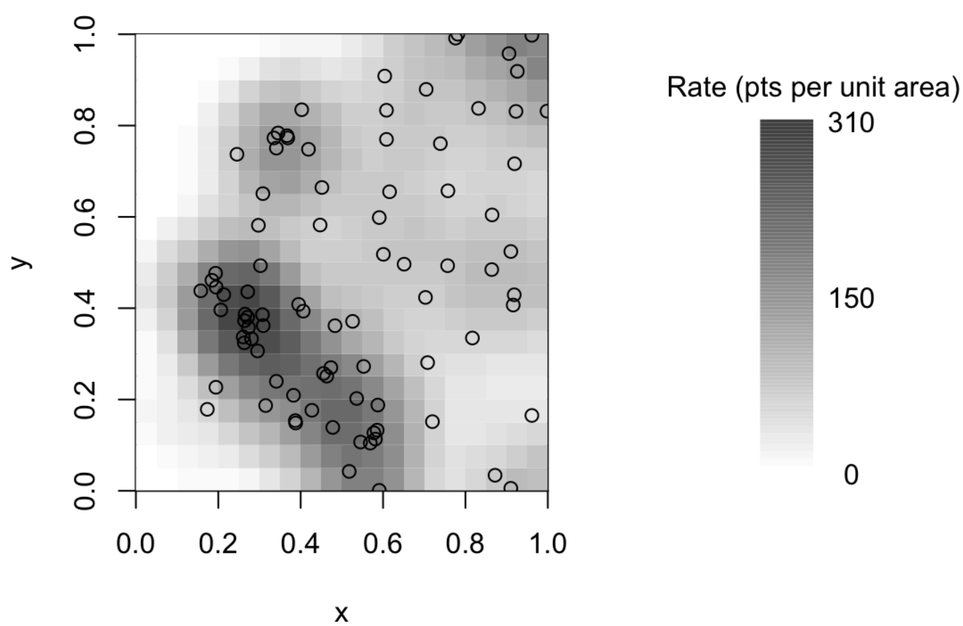

Figure 3: Kernel Smoothing Result

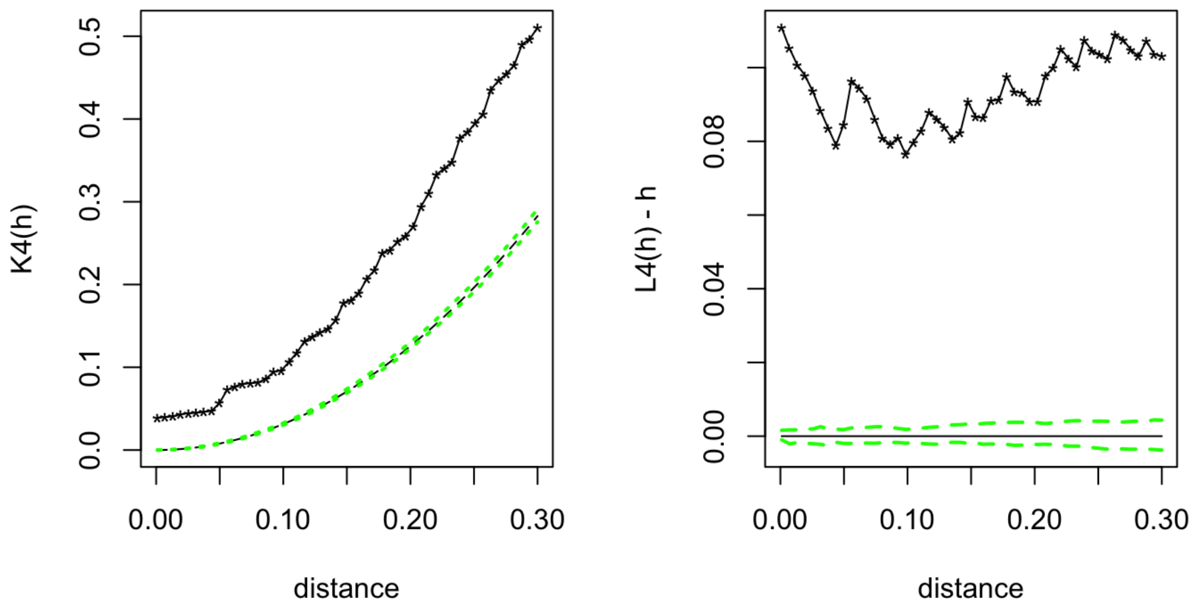

Figure 4: K, L function

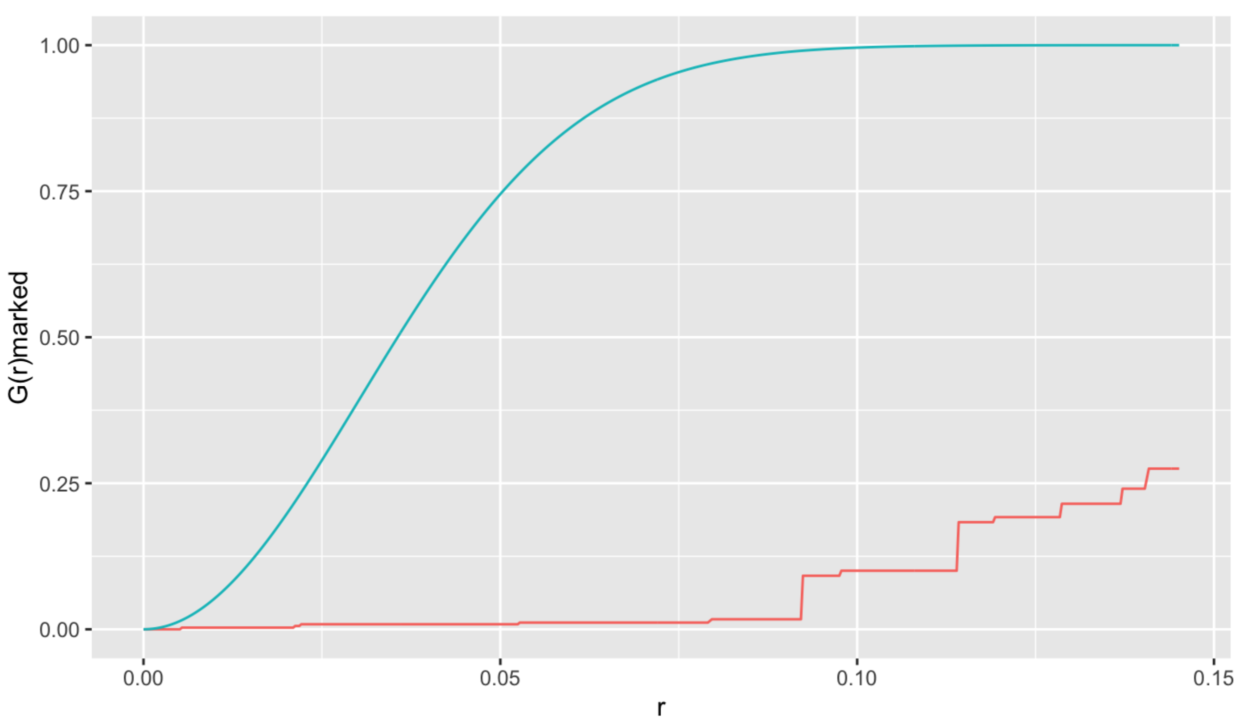

Figure 5: Marked G function

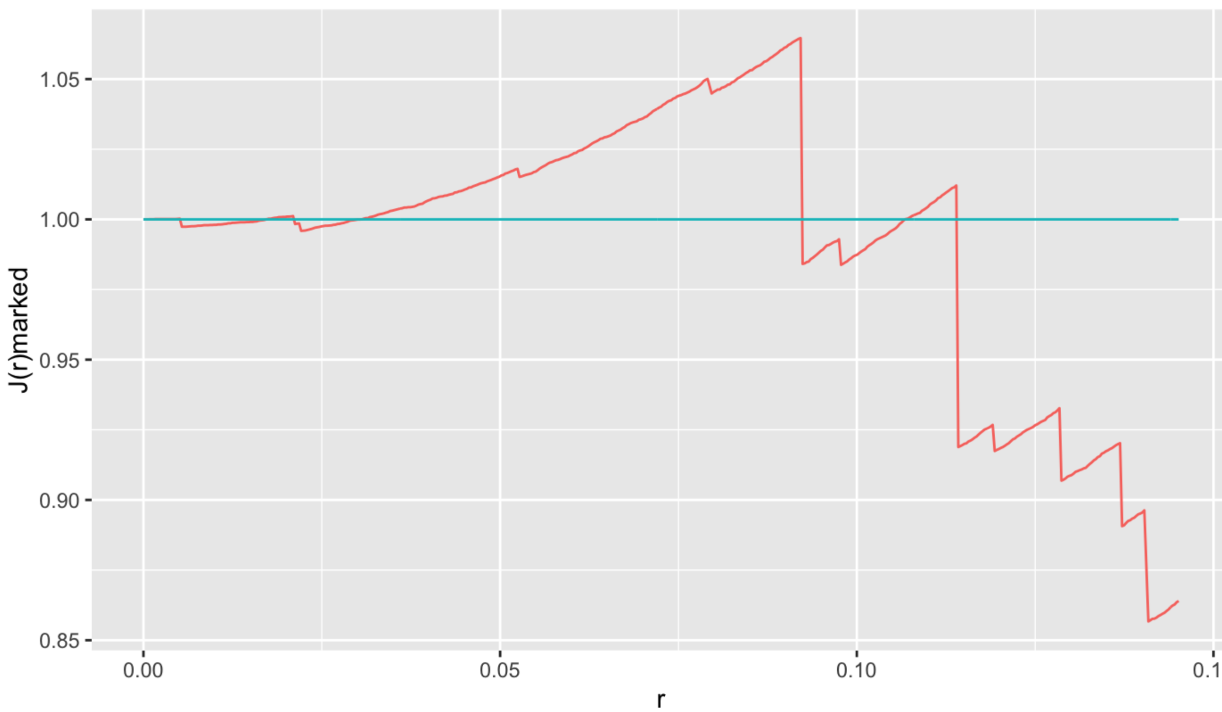

Figure 6: Marked J function

Figure 7: Top 3 Check-in Location

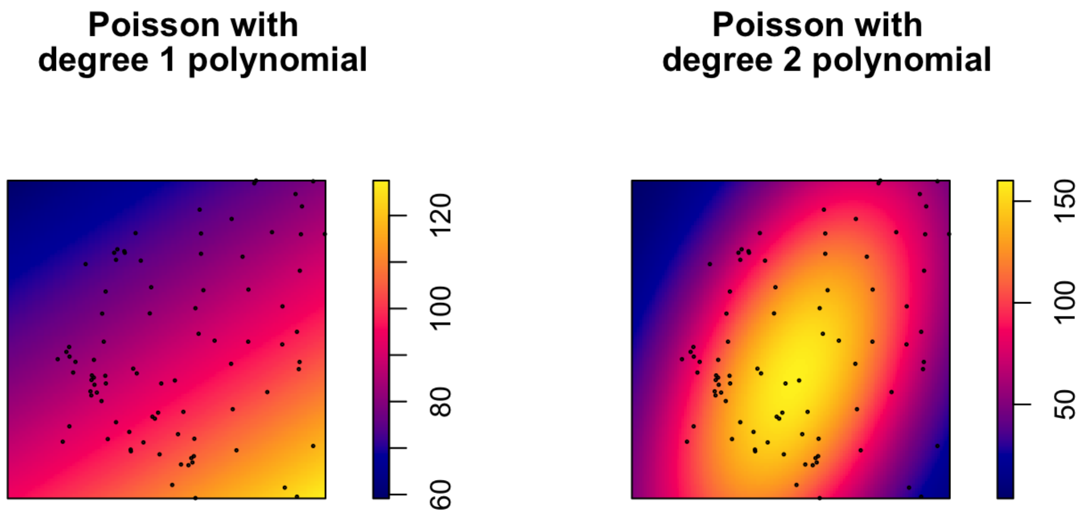

Figure 8: Fitted Intensity of Poisson

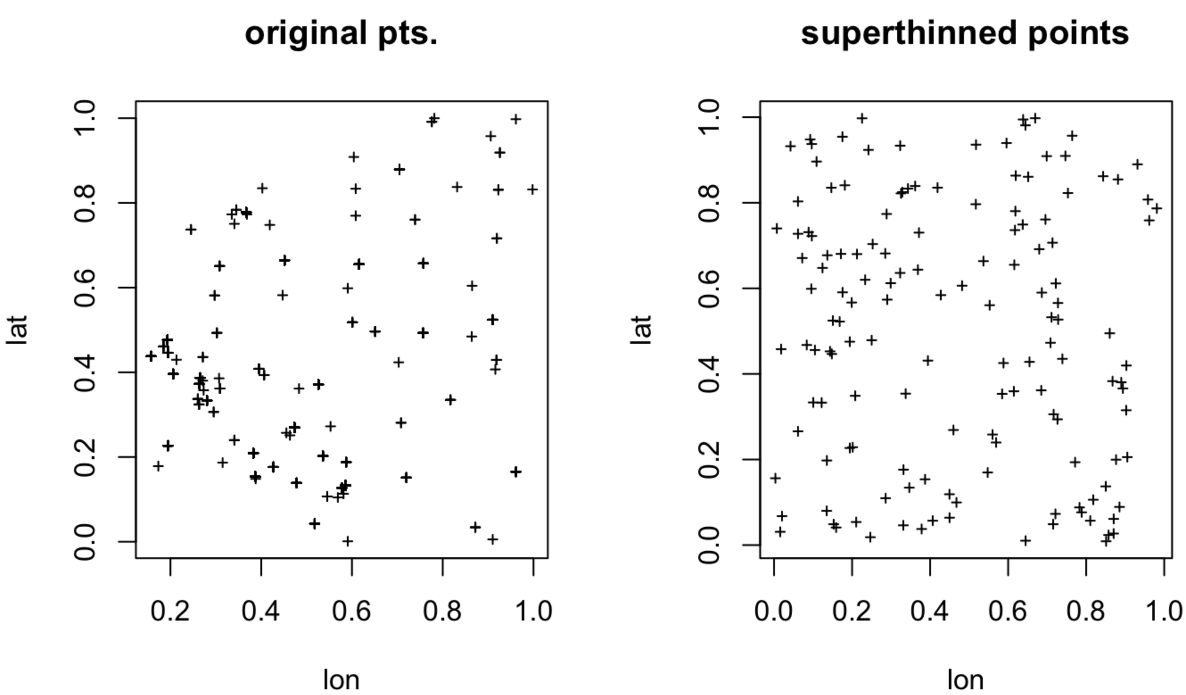

Figure 9: Superthinning after 2nd order Poisson Fitting

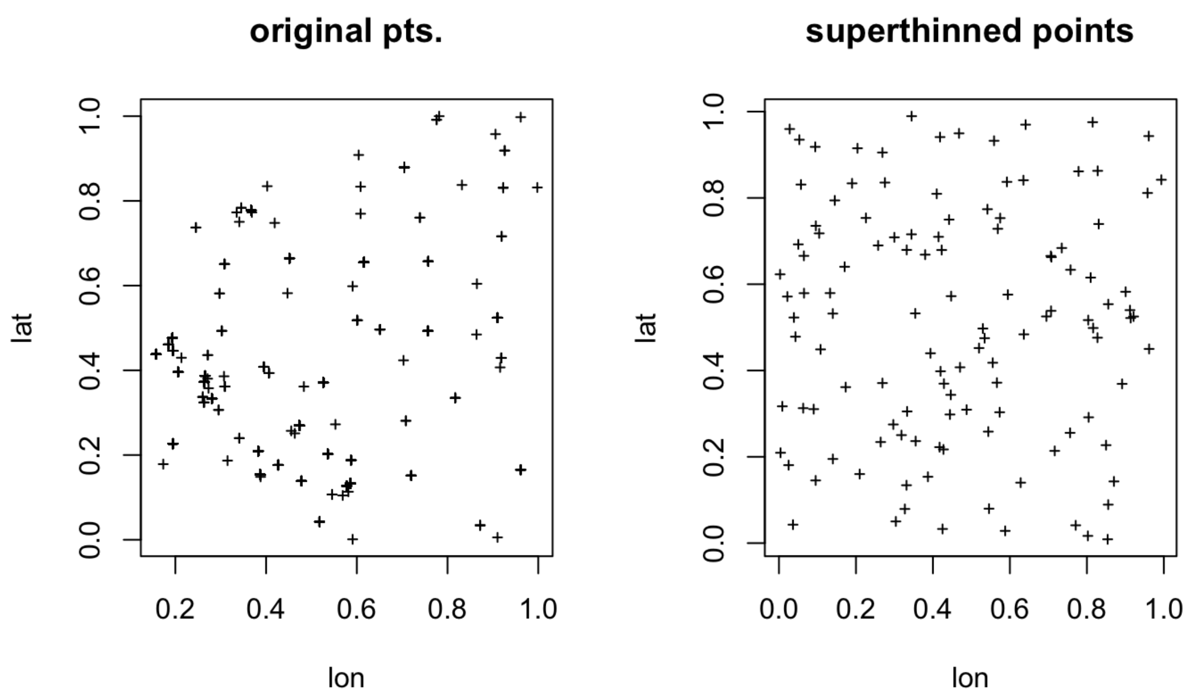

Figure 10: Superthinning after Hawkes model fitting