Creative Extension Report — December 2020

Clémence Barsi
clemence.barsi@epfl.ch

Pauline Conti
pauline.conti@epfl.ch

Eloïse Doyard
eloise.doyard@epfl.ch

Abstract

One could expect that on event days like Christmas or New Year's Eve, our mobility patterns would break. By studying mobility on different type of events (religious, cultural and sport events) compared to the rest of the year we find that we don't change much our usual patterns of mobility and neither our presence on social medias.

1 Introduction

Brightkite and Gowalla were two location-based social networks. They allowed the user to "check-in" at certain location, sharing it with their friends. The original paper *Friendship and mobility: user movement in location-based social networks* [1], explores the periodicity of movements in time and location upon friendship. They predicted human movements based on a periodic and social mobility model.

We will study the trends of the distance traveled from home as well as the number of check-ins made on special events and in specific regions of the world. We are also interested in seeing if the distance traveled from one event to the other varies.

Research questions Do certain events specifically influence people's mobility? Is people's mobility due to important events different in each region of the world? How far from their homes do people tend to travel for specific and important events? Lastly, if there is a change in mobility, will it cause an increase or a decrease in the total check-ins count?

Hypotheses We suspect that there will be at least a slight change in the patterns of users' mobility on the events we considered, in terms of number of check-ins made by day as well as in the probability to travel to a certain distance. However, if there is a change, we think it will be faint due to the big amount of checkins overall.

2 Data Collection

Events We chose to construct the datasets ourselves, giving us the freedom of choice over the events studied, in terms of type (ranging from cultural and religious to sports related) and of temporal constraints given by the network data.

Regions of the World After analyzing Figure 1, we chose to study only specific regions, which had the most checkins. To delimit them, we defined abstract rectangles, defined by a top-right and a bottom-left corner (cf. Figure 4), encapsulating the entirety of the land areas we wished to analyse.

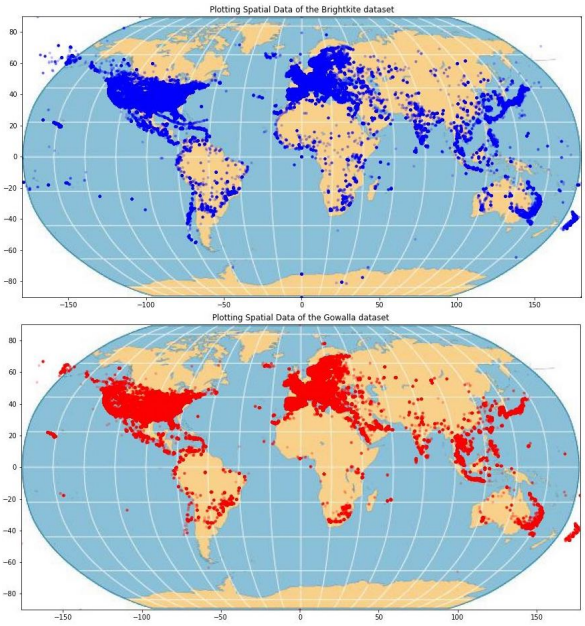

Figure 1: Dispersion of the checkins on a world map for Brightkite **(Above)** and Gowalla **(Below)**

3 Dataset Description

Regions of the world The regions we chose to study are:

- The United State of America (USA), only including mainland.
- Western Europe, composed of: Spain, Portugal, France, United Kingdom, Ireland and multiple other small European countries that were included in our selection when delimiting the land area.
- East Asia, composed of: Japan and South Korea

As the pie plots (cf. Figure 2) show, the USA counts for 70% of the area studied but around 90% of total checkins registered. Hence we know that their impact on the overall distributions will be notably bigger than the ones of the two other regions.

Events We wanted to study the impact of important cultural events on people's mobility. For the feasibility of the project we concentrate our study on only three:

- 4th of July (i.e. the USA National Day): gives us a good example of an event impacting only one of the three chosen regions;

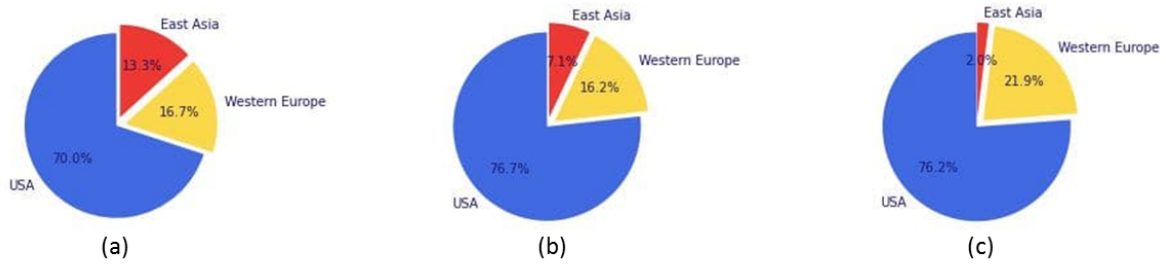

Figure 2: Breakdown of area **(a)** and number of checkins in the chosen regions for Brightkite **(b)** and Gowalla **(c)**.

- Christmas and New Year's Eve: being partially a Christian religious celebration, is a major event in Western Europe and North America, but not necessarily in East Asia;
- Winter Olympic Games of 2010 : involves the whole international community, especially the regions we chose, were winter sports are quite developed.

With this three events we were able to investigate : the end of year celebrations which are mostly family gatherings, while the USA national day and the Winter Olympic games involve different social situations. We chose the dates to span over a dozen of days surrounding the events (in case their were any "travel days"), except for the Olympic Games, where we chose the exact duration of the event.

4 Method

We divided our study, mainly, in two parts : the study of the distance travelled from home and then the number of daily check-ins. We computed different probabilities and performed several statistical tests.

Defining users' homes & travels To define the home and home cell of each user, we followed the method described in the *Friendship and Mobility* [1] paper :

- Discretize the world in cells of 25 km^2
- Define as home cell for a user the one with the most checkins
- Define the home as the average location of all the checkins made by the user in the home cell.

A checkin was considered to be made out of home if it was not located in the user's home cell. The other checkins represented an actual "trip" made by the user. The origin that was assigned to each user was simply the region in which his home was.

$P_{event}(d)$ We named $P_{event}(d)$ the distribution of the probability to travel a distance d on an event day. To calculate this probability we simply computed the ratio of the count of the checkins that were made at a distance on the event days divided by the count of all the checkins that were made at the same distance. We can see this distribution for each event on Figure 7, and for the USA on Figure 8 .

Statistical Tests To see if two samples followed the same distribution, we used the Student T-test, as the Central Limit Theorem was applicable due to the amount of data.

5 Results & Findings

5.1 Checkin Distances From Home

First, we studied the distributions of the distances from their homes at which people checked-in.

All Year Long From Figure 5, we observed that there were no major differences in the distribution of the distance of the checkins travelled from home between the three regions.

We performing statistical tests between the overall distribution and each region's distributions of distance from home. They all rejected the hypothesis that any of the samples had the same distributions. This was surprising for the USA, as the distributions looked very similar.

On Special Events The shape of the distributions of distances in Figure 6 for each region across different events seem to indicate that if there is an impact of special events on people's mobility, it is not pronounced nor marginal. We can again notice the impact of the USA's distribution in the overall one compared to the two other regions. The disparity in number of checkins of each network is not only explained by the region of the world, but also by the variations in popularity of each network across those periods (see 5.2).

Figure 7 visually showcases different distributions of probability of distance travelled from home on the chosen events. We saw that :

- Over Christmas and New Year's Eve, the slight spike between a 100 km and a 1000 km cannot be seen as precisely in any of the other plots. This could be due to the fact that people choose to travel further away to spend the end of the year's celebrations with their relatives.
- USA's national day being an American holiday, most of the data points correspond to American citizens and due to the size of the country, travelling a 100km for the national day seems more feasible for them than in Western Europe or East Asia where countries are smaller. However, this event is only one day-long explaining the shorter distances travelled.

- For the Winter Olympic Games, the overall shape of the probability plot suggests that the impact of the games on people's mobility stays marginal. But, the high peak at very big distances could be a consequence of avid fans travelling abroad to see the competition, as it took place in Vancouver at a distance from the center of Europe and East Asia of the same order of magnitude than the peak.

The statistical analysis confirmed that the distribution of distance travelled from home is statistically different on events periods than in the whole year (with a level of confidence of 0.95).

In The USA As more than 90% of the checkins (cf. Figure 2) were made in the USA, we decided to take a closer look at the distance travelled in this region of the world (cf. Figure 10). We arrived to the same conclusions in the United States than when studying the whole world.

Overall, the study of the checkin distance from home showed that on events' days there were statistical differences. But these conclusions aren't decisive as it seems like that any period of the year is statistically different to the rest of the year regardless of the place or the occurrence of an event.

5.2 Number Of Checkins

All Year Long Unlike in the section 5.1, we chose to only plot the year 2009 as it is the only one where we have data for the entire year on both Gowalla and Brightkite.

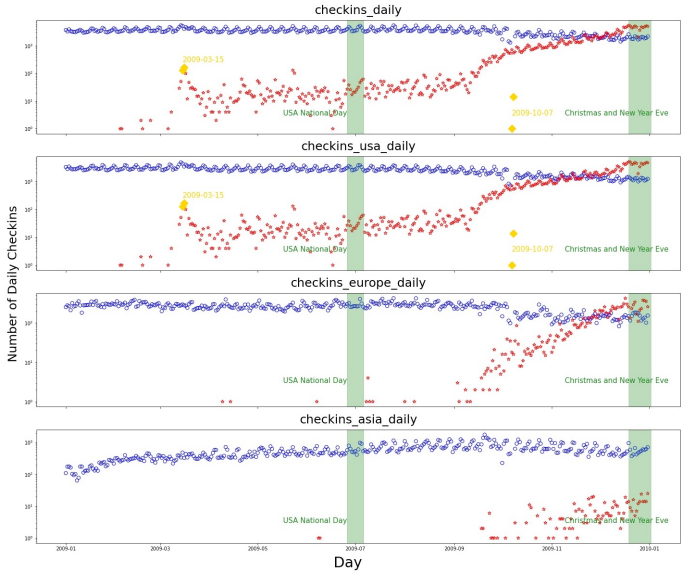

Figure 3: Number of daily checkins of the year 2009 for the world **(1)**, the USA **(2)**, Western Europe **(3)**, and East Asia **(4)**

When plotting the number of checkins across 2009 we saw the rise in popularity of Gowalla especially towards the last trimester of the year. On both networks we noticed a recurring pattern where on the weekends there were more checkins.

We didn't see any disruptions in the patterns due to the events, which suggests that the effect they have is rather faint.

Mystery Days We noticed two major outliers in each one of the dataset in the USA (cf. Figures 11 and 12):

- October 7th, 8th, 2009 for Brightkite (sudden decrease)
- March 15th, 16th, 2009 for Gowalla (sudden increase)

We have linked the decreasing outliers in Brightkite to an operation of maintenance. For Gowalla, we have found that most of the checkins on that date occured in the same city as the headquarters of the network. The increase might be due to an event related to the social network.

On Special Events From visualizing the plots of the Figure 9 we can see that Christmas Eve has an impact (decrease in number of checkins) in the regions of the USA and Western Europe but none on the East Asia. This confirms that depending on the culture some events have an impact and others don't. However the other events didn't point out to any disruptions in the usual patterns (rise in the number of checkins on weekends and decrease at the beginning of the week).

To conclude, even when considering the events in regions where the "usual" patterns were very visible, the events we chose did not seem to alter it noticeably except for Christmas Eve.

Nevertheless, when performing statistical testing to reinforce our hunch, we couldn't accept the hypothesis that the distributions of checkins during the event days were the same as all year long which is surprising when seeing the plot. We believe this is due to the volatility of the number of checkins variable making it difficult to conclude.

In The USA When focusing in the USA, the conclusions we made were exactly the same as in the whole world i.e. that the events didn't alter noticeably the patterns observed all year long except for Christmas Eve.

6 Conclusion

To conclude, while it seemed that except for Christmas Eve the events we chose didn't influence much the mobility patterns, when performing statistical analyses we found the contrary. The distributions of distance from home and number of checkins seem different all year long when considering a group of days regardless of the place and time of the year. Thus studying those variables to investigate the influence of mobility patterns isn't as conclusive as hypothesized.

References

[1] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090, 2011.

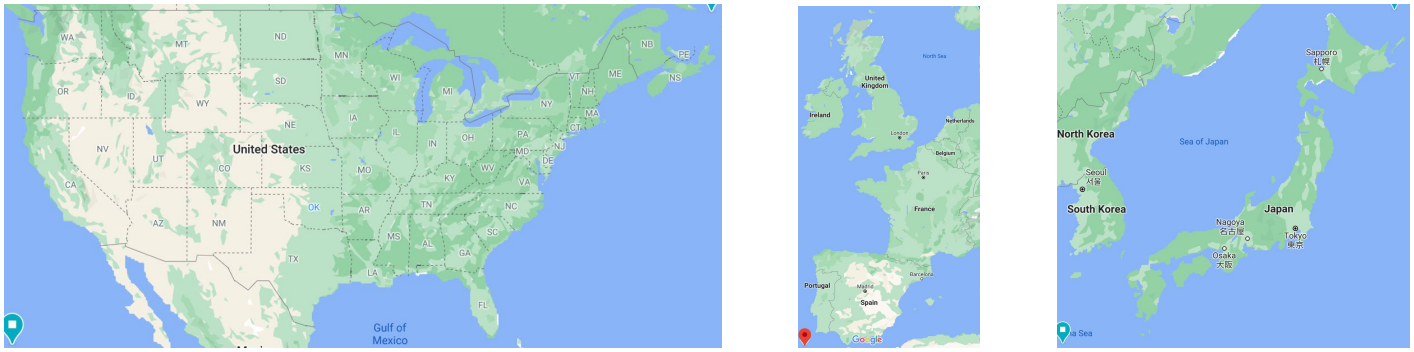

Figure 4: Abstract rectangle for the USA **(Left)**, Western Europe **(Middle)** and East Asia **(Right)**.

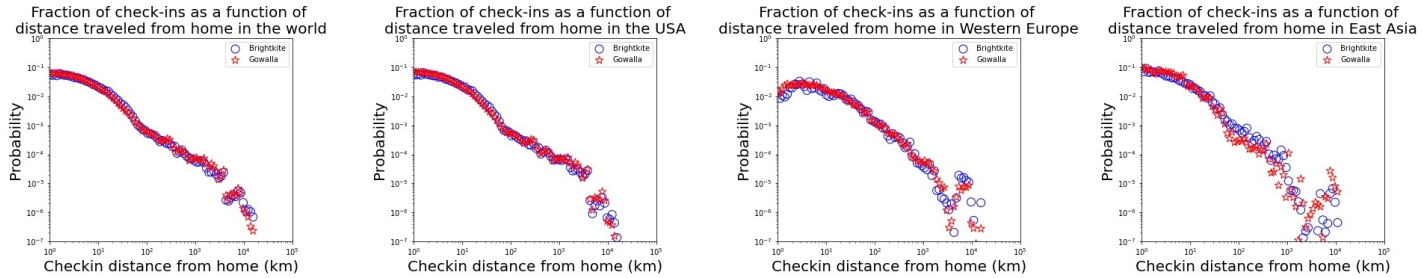

Figure 5: Fraction of check-ins as a function of distance traveled from home for in the world **(a)**, in the USA **(b)**, in Western Europe **(c)** & in East Asia **(d)**

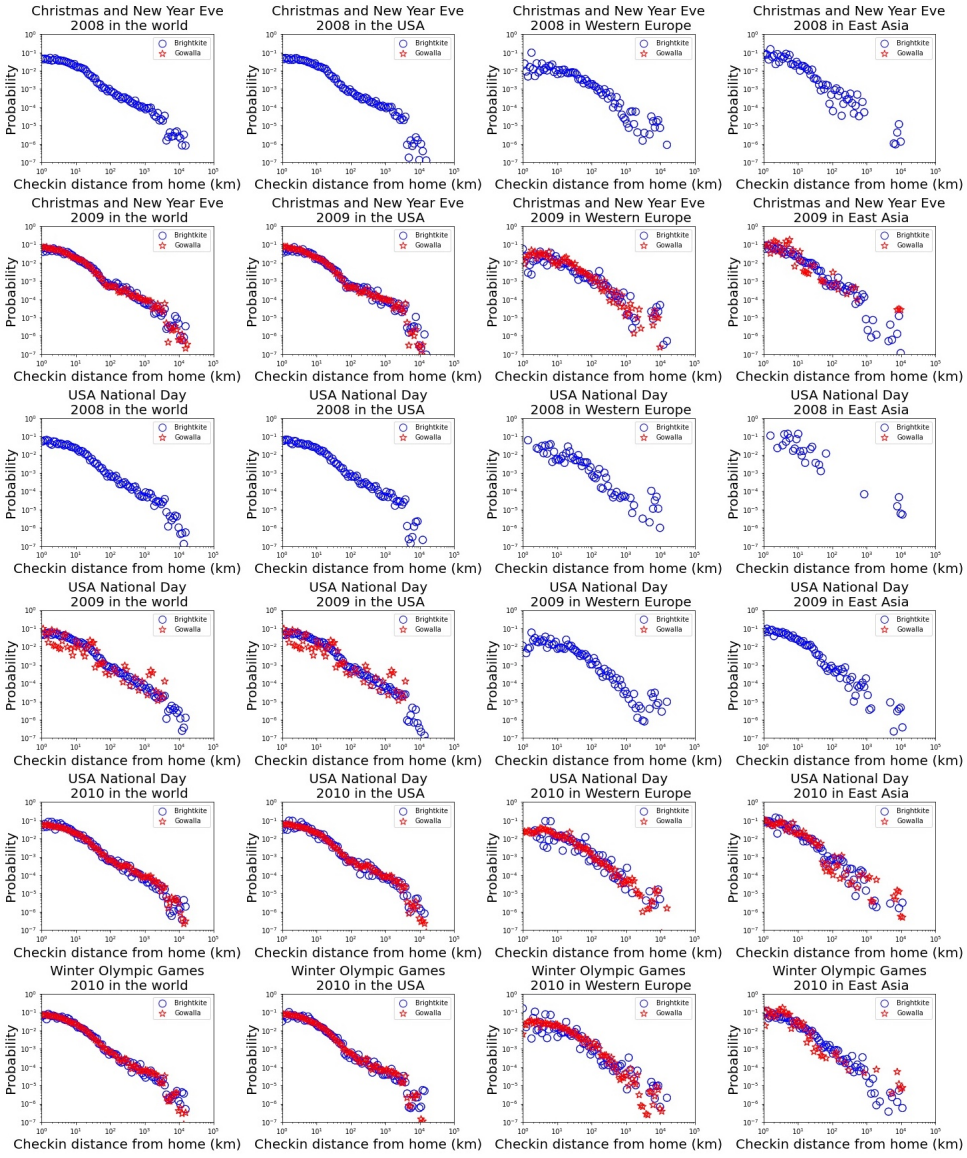

Figure 6: Fraction of check-ins as a function of distance traveled from home on special event worldwide **(Column a)**, in the USA **(Column b)**, in Western Europe **(Column c)**, in East Asia **(Column d)**

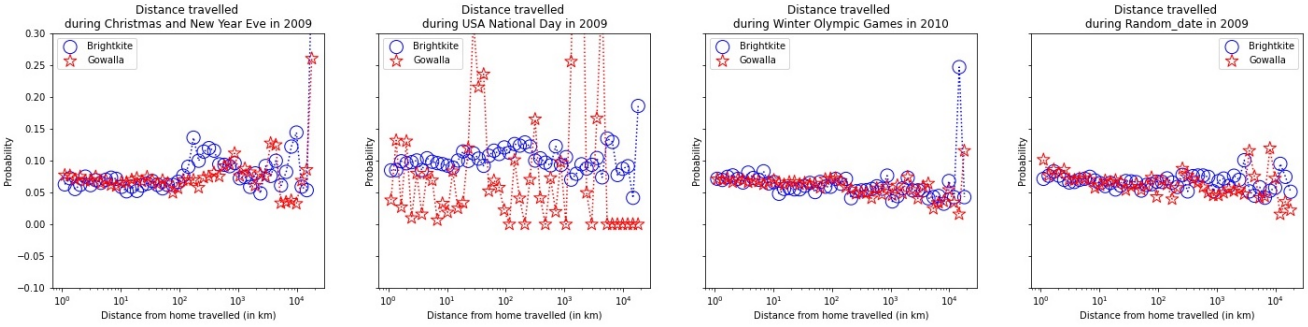

Figure 7: $P_{event}(d)$: Distribution of distance traveled from home across all regions on Christmas and New Year's Eve **(a)**, for the 4th of July **(b)**, for the Winter Olympic Games of 2010 **(c)**, and on random dates **(d)**.

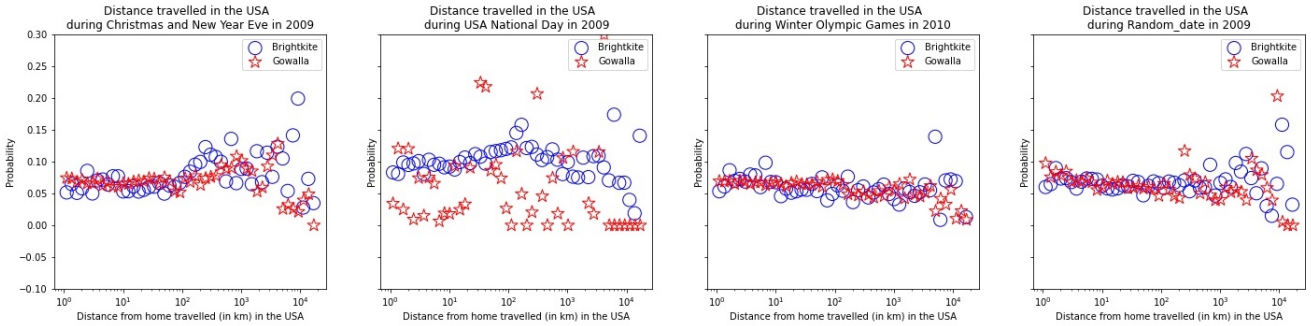

Figure 8: $P_{event}(d)$ specifically in the USA: Distribution of distance traveled from home in the USA on Christmas and New Year's Eve **(a)**, for the 4th of July **(b)**, for the Winter Olympic Games of 2010 **(c)**, and on random dates **(d)**.

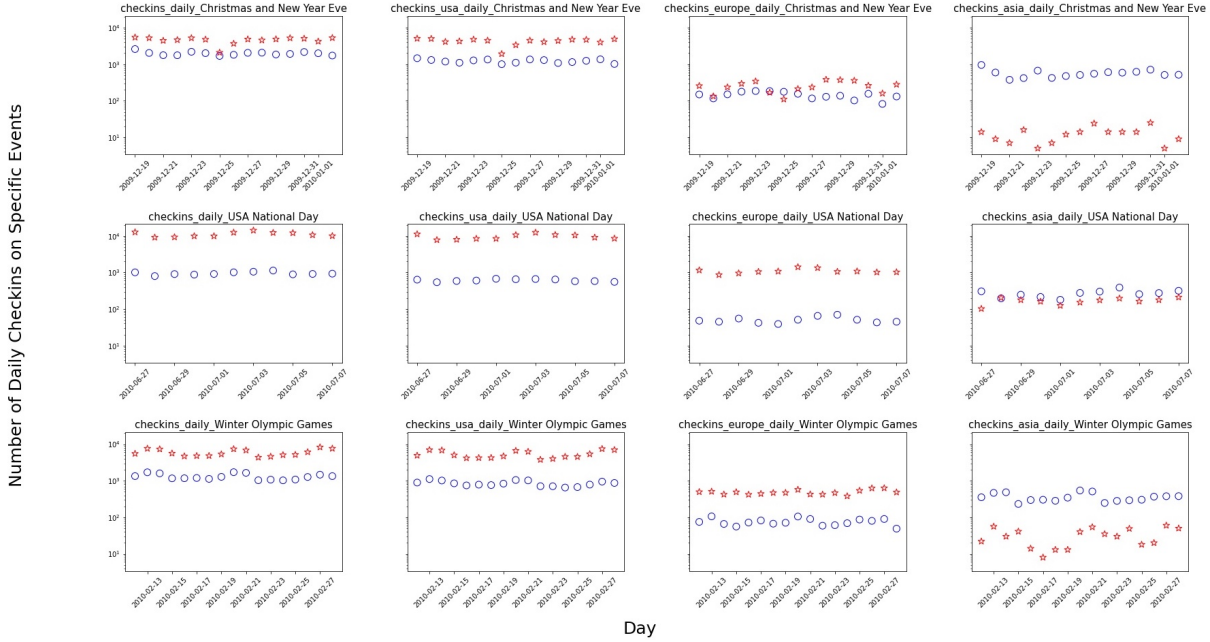

Figure 9: Number of Daily Check-ins on Specific Events in all regions **(Column a)**, the USA **(Column b)**, Western Europe **(Column c)**, and East Asia **(Column d)**

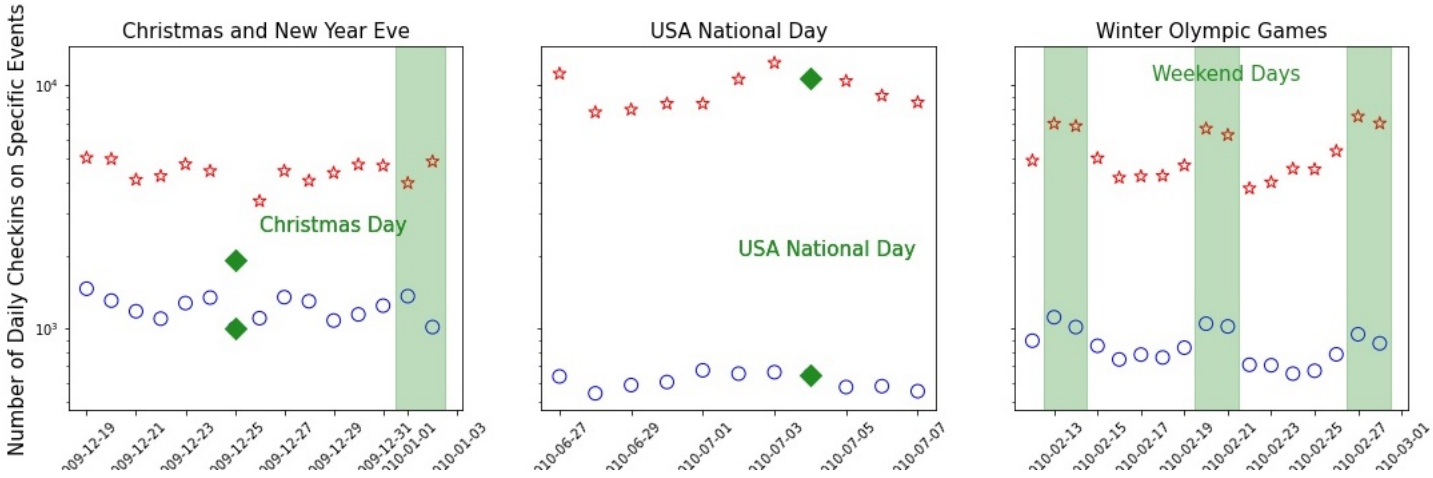

Figure 10: Number of Daily Check-ins on Christmas and New Year's Eve **(Left)**, USA National Day **(Middle)** and the Winter Olympic Games of 2010 **(Right)**

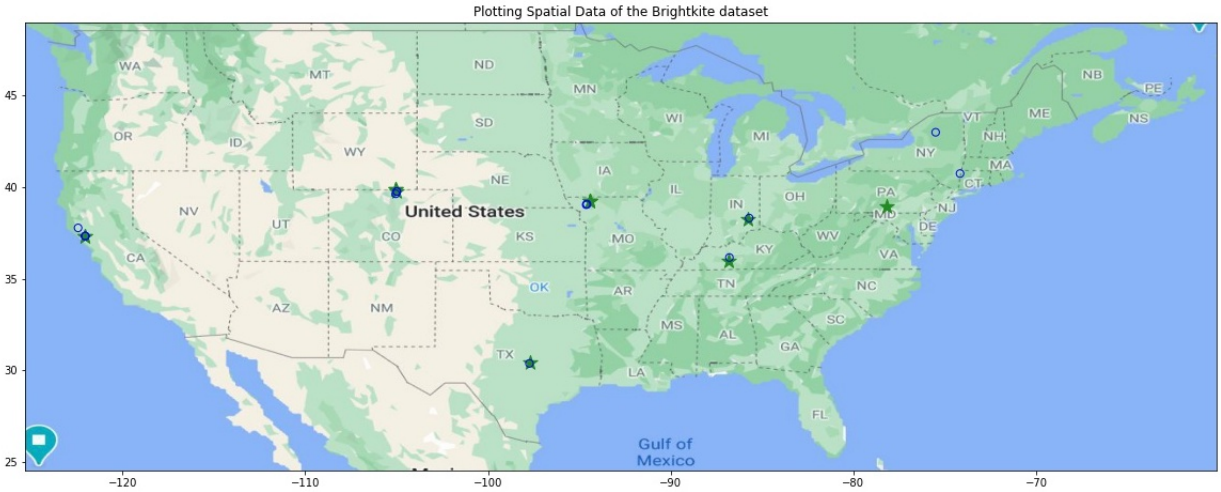

Figure 11: Spacial dispersion of checkins **(blue circles)** and their user's homes **(green stars)** on March 15th and 16th, 2009

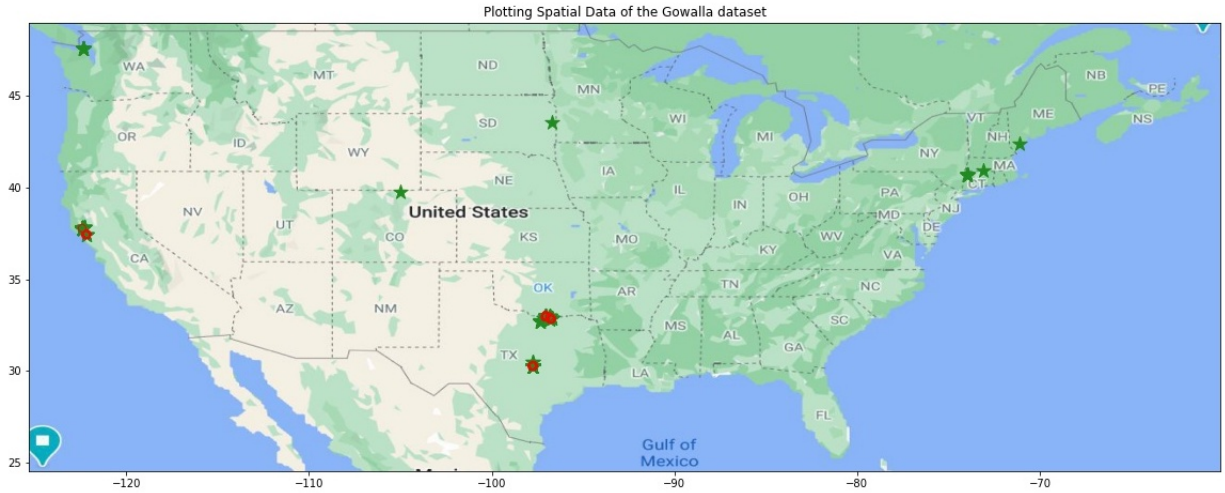

Figure 12: Spacial dispersion of checkins **(red circles)** and their user's homes **(green stars)** on October 5th to 7th, 2009