### 5.2.2 Implementation Result

The algorithm is applied to Geolife dataset. The user files are read in an online manner to simulate the GPS location points received on a mobile device. The stay-points found for user 1 for November 2008 as shown in the Figure 12. The trajectory is shown with the green line and the red arrows indicate the stay-points. These stay-points represent the locations with semantic meaning behind it like “home”, “work”, “restaurant”. This clearly depicts that a lot of noise in the trajectory data is removed at this step and only the significant stays are extracted.

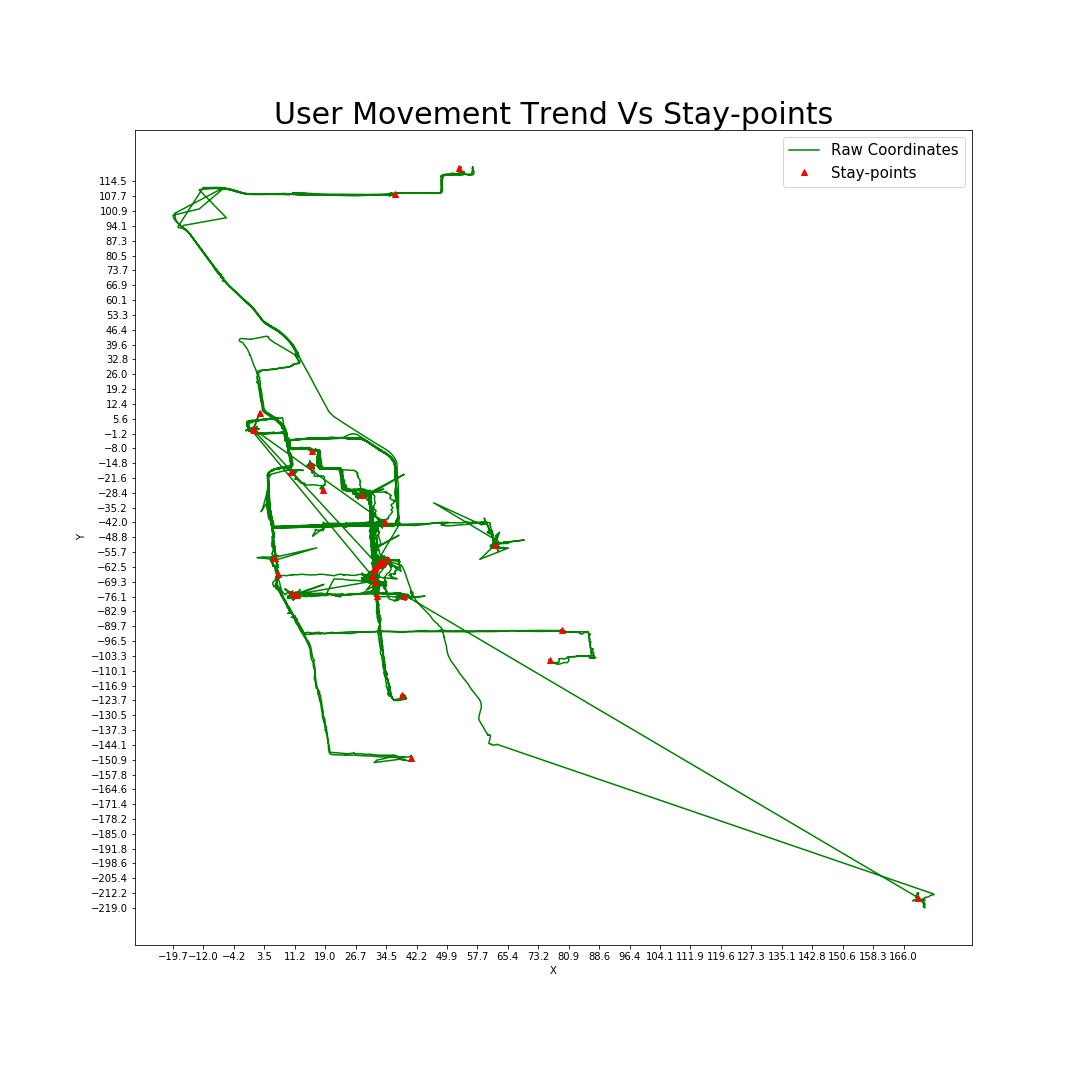


Figure 12 User 1 raw trajectory data vs stay-points extracted

### 5.4.2 Implementation Result

The hourly weights calculate the time-slotted data which is used as the base for markov chain model.

The Geolife dataset date and time are represented in GMT, hence to have the correct visualization, the date and time must be adjusted to the local time in the trajectory data. The Figure 16 below depicts the hourly weighted state data for user 1 for November 2008. The x-axis represents each hour of the day from 0 to 24 and the y-axis represent the days. Each rectangle depicts a state where the width of the rectangle represents the weight of the state in the corresponding hour. For example, state 1 is the first state between 9 am and 10 am on day 0. The distribution of the states over the hours and days makes some hints about the semantic meaning behind the locations. For instance, state 3 is most likely user’s home location as on most days’ user is at this location from 9 pm till next day 7 am and state 6 is most likely user’s work location as on most days’ user is at this location from 8 am till 8 or 9 pm. There are many other locations like 13, 22, 2, 16 and so on which could represent the supermarket, shopping mall, fitness club and so on.

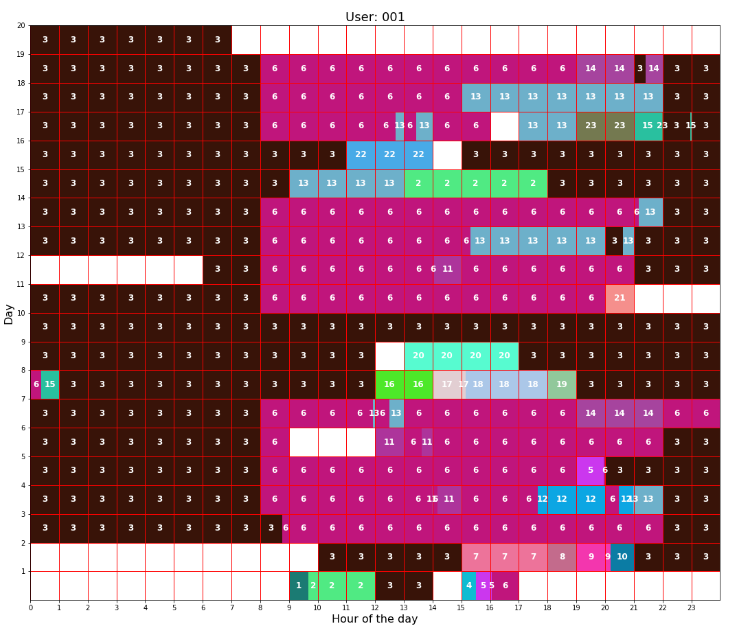


Figure 16 User 1 time-slotted data for the November 2008

### 5.5.2 Implementation Result

The markov chain model is implemented on Geolife dataset as explained above. The Figure 22 depicts the markov chain of user 1 for November 2008 for time-slot hour 8. The columns headings and the row headings indicate several state ids from 1 to 23. Each cell represents a probability of transitioning from one state to another. Red indicates the very low probability. For example, the row one indicating the transition from state 1 to all the other 23 states, are equally probable. The only two exceptions are the transition from state 3 to state 3 and transition from state 3 to state 6. The transition from state 3 to state 6 is more probable (0.66) than the transition from state 3 to state 3 (0.34).

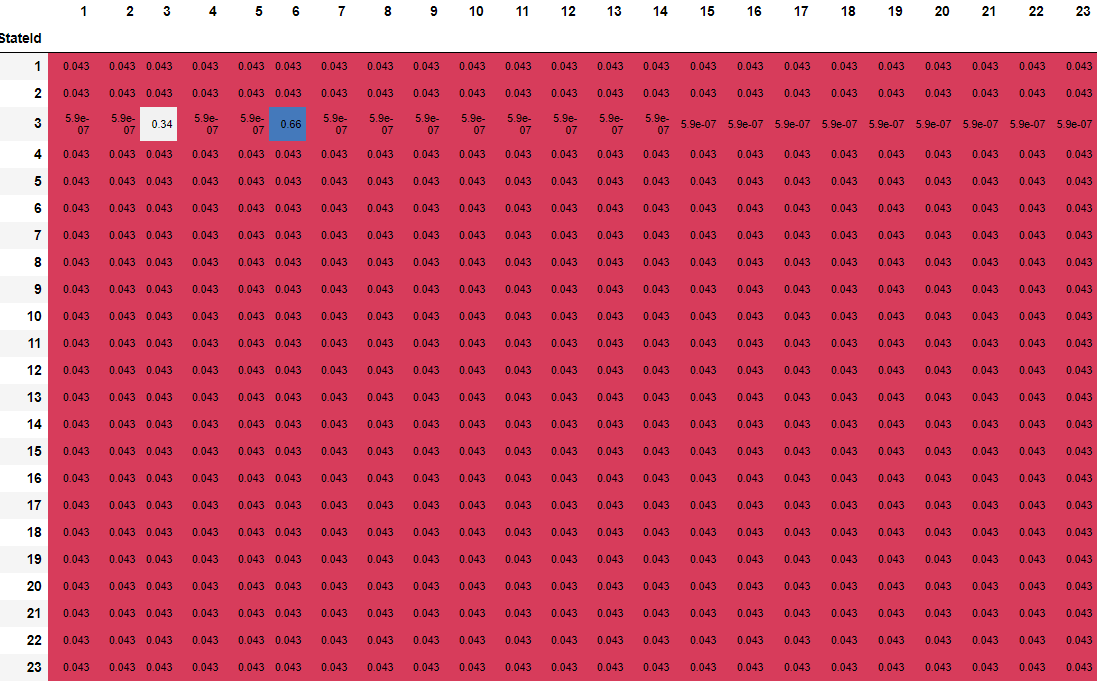


Figure 22 Markov chain model for user 1 for November 2011 time-slot hour 8

Similarly, there exists a matrix for each time-slot. Since we have divided the time-slots in hourly manner, there exists 24 time slots from 0-23 Hours. Hence there exists 24 different matrices, consisting transition probabilities from states-to-states. These Matrices together forms markov chain. This can be used for future prediction. In the Figure 22, the transition probability from state 1 to state 2 from time slot 7 to time slot 8, is 0.043. In other words, if user is at location state 1 at 7 am, there are 4.3% chances that he will move to location state 2 at 8 am. For instance, from state sti in time slot t, the first step predictions can be done in time slot t+1. For this, the matrix for time-slot t+1 and row with state sti is used. The entire row gives the probability to reach all the other states. The highest probable state is the predicted next state. The process continues for time-slot t+2 and onwards.

### 5.6.1 Implementation and Result

The algorithm takes the current hour, location and minimum threshold as input and uses the markov model to predict the several paths representing the several locations in consecutive time-slots with their confidence percentages. The Figure 23 shows the paths predicted for user 1 for a known state 3 at hour 7 with minimum confidence 0.1. The x-axis denotes the hour of the day and each path starts in a new line. There exist two paths. The circle with the number denotes the states predicted and the color of each circle is to indicate the confidence of the prediction. The first predicted state at hour 8 in path 1 is 6. Similarly, the first predicted states at hour 8 in path 2 is state 3. The states with darker color indicate a high confidence and the states with lighter color indicates the lower confidence. The confidence reduces as we go further away from the starting hour i.e. 7. The path continues until the confidence drops below the threshold confidence or the day has ended. There is an additional state shown after the drop of confidence below threshold. This additional state is shown to see the drop in the confidence of the next state after the minimum confidence.

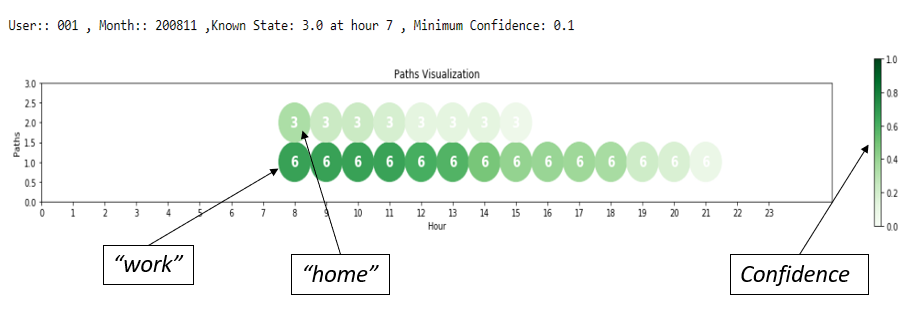


Figure 23 Path prediction for user 1

The path prediction is done to aware the user about the different paths which are predictable based on his/her location data collected. The known location was only state 3 at hour 7, which is then used to predict the paths for several hours. This is a privacy threat for the user.