Indian Journal of Science and Technology, *Vol 8(16), 63571, July 2015*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

An Enhanced Travel Package Recommendation System based on Location Dependent Social Data

C. Abilash Reddy* and V. Subramaniyaswamy

School of Computing, SASTRA University, Thanjavur - 613 401, Tamil Nadu, India;
ashreddy@outlook.com, swamy@cse.sastra.edu

Abstract

Location based social networking systems add location as main dimension to the social connections which gives necessities for personalized location recommendations. Travel packages can be personalized by obtaining user preferences, POI attractions and patterns between them from LBSNs. Existing recommendation systems concentrate mostly either recommending locations , travel packages to a single user or not precise enough, just recommending a list of possibly suitable packages to select by a user group. In our system travel packages are personalized to a user group by considering their common interests, social connections among them along with their individual interests, constraints. Recommendations made precise by considering multiple metrics which varies in degree of personalization and time period of evaluation. We built a prototype system and evaluated results based on data obtained from foursquare site. Experimental results prove system recommends effectively in single user scenario and also adapted well to user group scenario.

Keywords: Enhanced Travel Package Recommendation System, Interest Ratio, Personalized, User Group, Regular Sequences

1. Introduction

Location Based Social Networking Sites like Foursquare, Trip Advisor, Time city are going into mainstream and becoming common with the increasing popularity of GPS enabled mobile devices. A user not only gets location information from LBSNs but also expresses opinion by doing a check-in, like, writes a comment, review, uploads related photos and also follows known, similar users etc. User check-ins, likes, whom following gives the better chance for personalization.

A travel package contains POIs of various categories (e.g., food, attractions, and entertainment), but also the visiting sequence of the selected POIs. These POIs can belong to multiple categories and their attraction may vary with time.

In this paper we focus mainly on recommending travel packages to a user group exploring a particular region, where existing systems lacking. It shows similar personalization, if not better, on single user case also. It can be expanded to long distance travel scenarios by considering whole travel as connecting a series of travel regions. Our system works well even when data availability about a region is less due to variety of metrics used.

In our system we first construct a user profile and location model from the location based social data and find regular sequences among locations. User group and tour details are taken as input from the user. We then find nearby POIs in region and find their likeness by using user interests. Overall group preferences are obtained by giving equal weight to all users but due to considering social connections gives more priority to popular ones opinion .Variety of used metrics provides importance to even less popular but strong user opinions. Popular POIs of each time period are short listed and all possible routes among them which not contradicting with

** Author for correspondence*

regular sequences are considered. Each route is evaluated by finding transition probability from one location to another and ranked accordingly. Finally top ranked packages are recommended to the user.

2. Literature Survey

Personalized location recommendation approaches are hot topic in location-based services. Li et al.[1] utilized the travel behavior sequences and the hierarchical property of locations to find similarity between users from their location history. Zheng et al.[2] proposed a user-centric CF model based on user similarity for a personalized friend and location recommendation. Y. Zheng and X. Xie.[3] mined GPS trajectories of multiple users and User behaviors, experience of a region are taken into account to find hot locations. Leung et al.[4] formed a tripartite graph with User-Activity-Location relations to recommend locations to users. Agrawal et al.[5] first item sets mined through association rules to find frequent ones. Later on, Agrawal and Srikant[6] presented Apriori algorithm to find frequent item sets depending on a user-defined minimum support.

There has been a number of travel route recommendation approaches proposed in recent years based on GPS data, check-in data and geo-tagged photos. Yoon et al.[7] mined multiple user-generated GPS trajectories, given a start, end points and travel duration, to generate an itinerary recommendation. Lu et al.[8] proposed a framework which mined user's check-in behaviors to efficiently recommend the personalized trips meeting multiple user constraints. It estimated the POI attraction values based on both user preferences and temporal-based properties. Hsieh et al.[9] proposed time sensitive route recommendations by considering the popularity, proper visiting time of each place, and visiting order of places, the transit time from one place to another. Chen et al.[10] designed a general itinerary planning service, based on users' preference, to generate multiday itineraries for the users. Zhuang et al.[11] presented a probabilistic approach which understands the user intent on the go without any explicit voice or text query from user. On similar notes, Sang et al.[12] mined check-in data and used a probabilistic approach to recommend a sequence of POIs which are both personalized and also context relevant. Choudhury et al.[13] generated travel itineraries automatically from user online activity like picture uploads, where number of times a POI mentioned by users defines its attraction value. T. Kurashima[14] mined geo tagged photos to estimate the probability of a landmark visited by a photographer.

Existing recommendation systems mainly provide users with a list of recommendations, but each recommendation only consists of a single item or a single type. However, users intend to receive more specific and composite recommendations. For instance, it would be useful to provide users with a travel plan that contains a sequence of different types of locations along with feasible travel routes. Liu et al.[15] provideda personalized travel package recommendation but the target recommended travel packages are fixed ones, collected from travel companies. Xie et al.[16] considered composite recommendations, each recommendation comprises a set of items. This work mainly focuses on generating packages which respects user constraints like limited budget etc. Yu, Zhiwen et al.[17] recommended a personalized travel package which consists of different location types and also integrated components like mapping but it doesn't consider multiple user interests and also metrics are not precise enough for recommendations to a user group.

Our proposed system provides an enhancement to existing recommender systems by considering multiple user interests, improving preciseness and better utilization of POI hotness and regular sequences followed by users.

3. System Architecture

We aim to build a prototype system which is shown in Figure 1, recommends personalized travel packages by taking user input and recommends packages to the user using GUI. System contains mainly three modules modeling system, recommendation system, GUI.

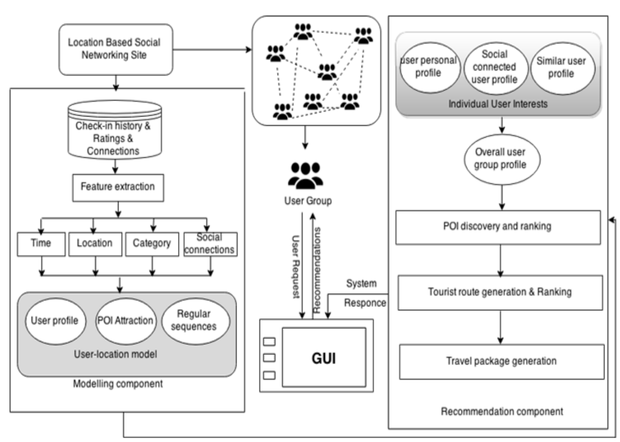

Figure 1. Proposed System Architecture.

Modeling system will do user profiling, POI modeling and finally regular POI sequences detection. User profiling, POI modeling usedto find spatial temporal characteristics of users, POIs by leveraging user ratings, check-in history. User profiling finds user interest ratio of a category, instead of ranking, which needed in multiple user scenarios. We label the POIs with multiple predefined categories where POI may belong to multiple categories and have different attraction values for each category. For example, a location "IMAX" belongs to shopping, entertainment, food categories and it have high attraction for entertainment, average for food, poor for shopping. We find regular sequences among user check-in sequences which later used to eliminate routes to be considered and also to provide more credibility to the recommended packages. This module operates offline as not depends on user input.

Recommendation system recommends personalized travel packages to a user group by considering multiple user interests, constraints and their connections. This module operates online. This module contains mainly three steps nearby POI discovery and ranking, travel route generation and ranking, travel package generation. Nearby POIs are discovered in a region of interest and ranked according to likeness obtained by evaluating a user personal, social connected, similar user's interests and also POI hotness. All possible route combinations which are following regular sequences are considered. Routes are ranked by evaluating transition probability to move from one location to another. Travel package generation will generate complete travel package for the selected route. This module operates online as recommendations personalized to specific user group.

GUI module takes user group details, region of interest, start, and end times of travel as input from the user. Recommends top 3 ranked routes as options to the user. Finally provides complete package for the selected route to the user.

3.1 Modeling System

3.1.1 User Profiling

User interests generally vary with time period of a day. For example a user in the early morning prefers sightseeing, visiting popular places in morning, eating at famous restaurants in afternoon, participating in events in evening, hangout places in late night. To accommodate those user behaviors a day is divided into 6 time periods of 4 hours each.

We will calculate user interest ratio for each category in each time period. UIR (u,c,t) represents the user u interest ratio for category c in time periodt, which is VT(u,c,t) the ratio of number of times a user visited a POI of category c to VT(u,t) all categories in a time period t.

We find similar(u,t), a fixed size set of users having similar interests to the user in LBSN using similarity formula sim(u, su,t)[18], where su is other user. We also find social (u), set of users whom a user u following in LBSN .A user profile is formed with user interest ratios and similar user, follows sets.

$$\mathbf{UIR}(\mathbf{u},\mathbf{c},\mathbf{t}) = \frac{\mathbf{VT}(\mathbf{u},\mathbf{c},\mathbf{t})}{\mathbf{VT}(\mathbf{u},\mathbf{t})} \tag{1}$$

$$\mathbf{sim}(\mathbf{u},\mathbf{su},\mathbf{t}) = \sum_{c=1}^{nc}\left(\frac{\mathbf{UIR}(\mathbf{u},\mathbf{c},\mathbf{t})}{\mathbf{UIR}(\mathbf{su},\mathbf{c},\mathbf{t})}\right) \tag{2}$$

3.1.2 Location Modeling

In general, popularity of POI may vary during different periods of time in a year. For example a nearby hill station is popular during summer but not so in winter. Granularity of these periods of time can be a season, a month, a week, a day etc. We take a month on average as granularity to contain the number of calculations to do per POI.

We measure POI hotness for each month. VT (p,m) represents number of times a POI p visited in a month m. rating (p) is average of user ratings given to a POI p. max Rating is maximum of ratings allowed on that LBSN. *α*, *β* are constants.

$$\mathbf{Hot}(\mathbf{p},\mathbf{m}) = \alpha * \frac{\mathbf{VT}(\mathbf{p},\mathbf{m})}{\sum_{m=1}^{12}\mathbf{VT}(\mathbf{p},\mathbf{m})} + \beta * \frac{\mathbf{Rating}(\mathbf{p})}{\mathbf{max\,Rating}} \tag{3}$$

3.1.3 Regular Sequences

In general most users follow a particular sequence of visiting POIs. These sequences are directional, A to B is different from B to A, and there may be multiple possibilities from POI, for example both A to B, A to C are possible.

We use Apriori algorithm[6] to detect regular sequences followed by users. Apriori transactions are obtained from each user check-in history where time gap between visiting POIs are within threshold. We set gap threshold as 2 days, more than that we truncate that history into parts, and a transaction should contain at least two POIs. Regular sequences of length not more than 3 obtained using apriori algorithm, as relation after 3 not likely hold in most cases. We find next (p) which contains all possibilities of next POI, if any, for a poi p.

3.2 Enhanced Travel Package Recommendation System

3.2.1 Nearby POI Discovery and Ranking

We divide the whole travel into time periods. This gives the list of time periods to consider. We calculate overall user group interest ratios of each category for each time period in list. We find all nearby POIs in circular area where region of interest is the center of it.

We find user group interests for each category in a time period as ratio of sum of all users personal interest ratio UIR (u, c, t)of that category in that time period to nu, total number of users in group . UGPIR (c,t) represents user group personal interest ratio for category c in time period t.

$$\mathbf{UGPIR}\,(\mathbf{c},\mathbf{t}) = \frac{\sum_{\mathbf{u}=1}^{\mathbf{nu}} \mathbf{UIR}\,(\mathbf{u},\mathbf{c},\mathbf{t})}{\mathbf{nu}} \tag{4}$$

We find social (u,p) each user's social connected users set who did check-in or rating at a POIp for each POI and their similarity with user u, likeness of POI. Likeness[2] is obtained by using ratio of number of check-ins at POI p, if any, to total number of check-ins in time period t and ratio of rating given by social connected user su to poi p to maximum rating allowed. Sim (u,su) represents similarity between user and social connected user. Similarly we calculate for similar users set similar (u,p).Then we obtain each user's USIR(p,t) as weighted sum of UIR $_{similar}$ and UIR$_{social}$ .we set equal weights for both. Overall user group UGSIR (p,t) obtained as sum of all USIR (p,t).

$$\mathrm{UIR}_{\mathrm{similar}}\,(\mathrm{u},\mathrm{p},\mathrm{t}) = \frac{\sum_{\mathrm{similar(u)} \cap \mathrm{u(p)}} \left(\frac{\mathrm{sim}\,(\mathrm{u},\mathrm{su},\mathrm{t}) * \mathrm{Like}\,(\mathrm{su},\mathrm{t},\mathrm{p})}{\mathrm{sim}\,(\mathrm{u},\mathrm{su},\mathrm{t})} \right)}{\mathrm{n}\,(\mathrm{similar}(\mathrm{u}) \cap \mathrm{u}\,(\mathrm{p}))} \tag{5}$$

$$UIR_{social}\,(u,p,t) = \frac{\sum_{social(u) \cap u(p)} \left(\frac{sim\,(u,su,t) * Like\,(su,t,p)}{sim\,(u,su,t)} \right)}{n\,(social(u) \cap u\,(p))} \tag{6}$$

$$UGSIR\,(p,t) = \frac{\sum_{u=1}^{nu} \left(UIR_{similar}(u,p,t) + UIR_{social}\,(u,p,t) \right)}{2 * (nu)} \tag{7}$$

$$UGIR\,(p,t) = 3 * UGPIR\,(c,t) + 2 * UGSIR\,(p,t) + Hot\,(p,m) \tag{8}$$

All nearby POIs are ranked after ordered by their respective UGPIR(c,t), UGSIR(p,t), Hot(p,m) values for each time period t in list. UGIR specifies user group interested category for time period and those of same category POIs ordered by their UGSIR (p,t) values and then by hotness of POI in that month. Hotness value useful especially when data availability is less. Variety of these metrics used for ranking provides preciseness in finding best nearby POIs for a time period. POI list formed for each time period by taking top 5 rank POIs.

3.2.2 Travel Route Generation and Ranking

Travel route generation aims to find all possible routes which are not contradicting regular sequences. We propose an enhanced travel route generation algorithm which combines general tree traversal algorithm with regular sequences to eliminate routes which are not likely to be best. It uses time periods in a travel and their POI list along with next set of each POI. Algorithm constructs a travel route as a path tree where each POI in start time's time period POI list acts once as starting point. We consider POI in the POI list of each time period as tree nodes of different levels. A user can traverse multiple nodes in same level if a traversal ends again in same time period. Then all possible travel routes generated by traversing tree nodes and if a POI has a non-empty next set then only those child POIs are considered during traversal.

We rank the all generated routes by using transition probability to move from one location to another. TP (u,x,y) represents the transition probability for user u to move from POI x to POI y. This depends on values of remaining travel time, distance between x and y etc. Te represents end time of travel, Tc represents current time, move(x,y) specifies time required to travel from x to y, stay(y) represents stay time at POI y. We empirically defined stay time values 3 hours for movie, 12 hrs for cricket ground etc.

$$\mathbf{TP}\,(\mathbf{v},\mathbf{x},\mathbf{y}) = \frac{\mathbf{T_e} - \mathbf{T_c} - \mathbf{Move}\,(\mathbf{x},\mathbf{y}) - \mathbf{stay}\,(\mathbf{y})}{\mathbf{1} + \mathbf{log}\,(\mathbf{Dis}\,(\mathbf{x},\mathbf{y}))} \tag{9}$$

We find overall transition probability for each route generated and rank those accordingly. Regular sequences are used to find the best routes if needed. The best 3 routes are given to the user group as options. Travel package generator integrates all necessary ones like maps to selected route and gives complete package to the user group.

4. Implementation and Evaluation

4.1 Dataset

We used UMN/Sarwat Foursquare dataset[19,20]. This Foursquare data set contains 2153471 users, 1143092 venues, 1021970 check-ins, 27098490 social connections, and 2809581 user ratings given to venues. All data extracted[21] from the Foursquare through the public API. All users' data have been anonymized. Each user, venue is represented by an id, geospatial coordinates. This dataset lacks necessary POI categories and their properties. We used Google places API to collect related POI category and its properties, given geospatial coordinates. We focused mainly on region of Minnesota State, United States. We did preprocessing[22,23] and removed users, venues, check-ins lacking sufficient information, not related. Number of check-ins per venue is less which useful to prove credibility of our method in sufficient data lacking scenarios.

4.2 Graphical User Interface

We built a prototype system and developed a desktop java application as the graphical user interface. Travel details, such as user group details, starting time, end time, region of interest specified usually by a landmark POI in a region, taken as input through graphical user interface. Each recommended route is shown to user using Google Static Maps API V2. Selected route package, consists of full map with details, transit details etc generated using Google Maps API, displayed to user, a sample venues, time list and recommended map is shown in Figure 2, Figure 3 and Figure 4.

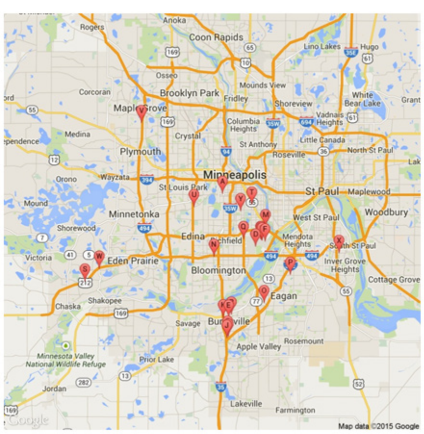

Figure 2. A Sample map of Nearby POIs around a Given Region of Interest.

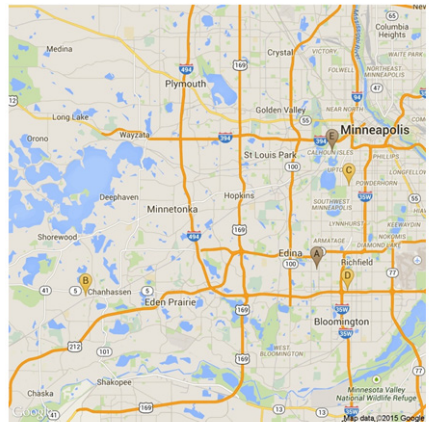

Figure 3. A Sample of Generated Time List.

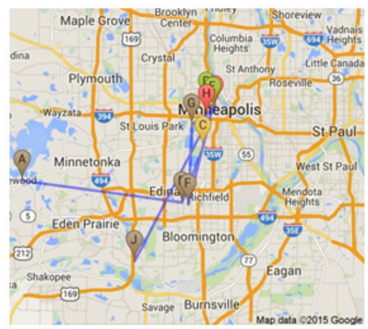

Figure 4. A Sample of Recommended Packages.

4.3 Experimental Results

We analyze performance of the system in both single users, multi user group scenarios. We compared with existing system[17], in single user scenario and results are shown in Figure 5. We got almost similar recommendations. Our system considers multiple likeness metrics which causes extra overhead but compensated by reduction in travel routes to evaluate. Performance of system depends mainly on travel time. Our system has comparatively less recommendation time as travel time increases.

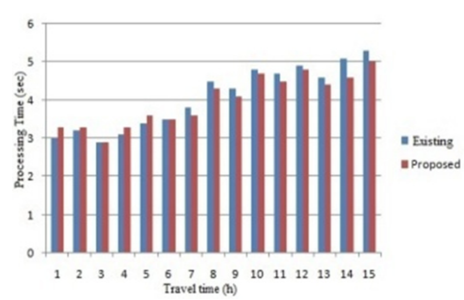

Figure 5. Performance Results in Comparison with Existing System.

We used a new metric, overall user satisfaction ratio, to test the effectiveness of results in multi user scenario. This metric represents sum of each user satisfaction with the recommended package. It depends mainly on similarity between individual user interests, connections among them and also user group size. We evaluated system for 50 test cases and results are shown in Table 1. Our system shows overall satisfaction ratios between 72 to 95 percent. Best case is when similarity of interests high. Worst case is similarity of interests is low and group length is high. Overall system shows best possible satisfaction ratios. Recommendation time is around 7s which is acceptable to users.

Table 1. Overall user satisfaction ratio values for 50 sample test cases

Range	72 - 75	75 - 80	80 - 85	85 - 90	90 - 95
Count	1	3	10	11	5

5. Conclusion

In this paper, we tackle the problem of personalized travel package recommendations to a user group based on location dependent social data. We considered overall user group interests, evaluated from individual interests, while recommending packages to user group. This system works not only for single user but also for multi user group. As a future work, we plan to recommend packages consist of multiple routes between intermediate POIs, to better fulfill individual interests, whenever needed.

6. References

1. Li Q, Zheng Y, Xie X, Chen Y, Liu W, Ma W-Y. Mining user similarity based on location history. Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems; 2008.
2. Zheng VW, Zheng Y, Xie X, Yang Q. Collaborative location and activity recommendations with gps history data. Proceedings of the 19th ACM International Conference on World Wide Web; 2010.
3. Zheng Y, Xie X. Learning travel recommendations from user-generated GPS traces. ACM Transactions on Intelligent Systems and Technology (TIST). 2011; 2(1):2.
4. Leung KW-T, Lee DL, Lee W-C. CLR: a collaborative location recommendation framework based on co-clustering. Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2011.
5. Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. ACM SIGMOD Record; 1993.
6. Agrawal R, Srikant R. Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB); 1994.
7. Yoon H, Zheng Y, Xie X, Woo W. Smart itinerary recommendation based on user-generated GPS trajectories. Ubiquitous Intelligence and Computing. Springer; 2010. p. 19–34.
8. Lu EH-C, Chen C-Y, Tseng VS. Personalized trip recommendation with multiple constraints by mining user check-in behaviors. Proceedings of the 20th International Conference on Advances in Geographic Information Systems; 2012.
9. Hsieh H-P, Li C-T, Lin S-D. Exploiting large-scale check-in data to recommend time-sensitive routes. Proceedings of the ACM SIGKDD International Workshop on Urban Computing; 2012.
10. Chen G, Wu S, Zhou J, Tung AK. Automatic itinerary planning for traveling services. IEEE Transactions on Knowledge and Data Engineering. 2014; 26(3):514–27.
11. Zhuang J, Mei T, Hoi SC, Xu Y-Q, Li S. When recommendation meets mobile: contextual and personalized recommendation on the go. Proceedings of the 13th ACM International Conference on Ubiquitous Computing; 2011.
12. Sang J, Mei T, Sun J-T, Xu C, Li S. Probabilistic sequential POIs recommendation via check-in data. Proceedings of the 20th ACM International Conference on Advances in Geographic Information Systems; 2012.
13. De Choudhury M, Feldman M, Amer-Yahia S, Golbandi N, Lempel R, Yu C. Automatic construction of travel itineraries using social breadcrumbs. Proceedings of the 21st ACM Conference on Hypertext and Hypermedia; 2010. p. 35–44.
14. Kurashima T, Iwata T, Irie G, Fujimura K. Travel route recommendation using geotags in photo sharing sites. Proceedings of the 19th ACM International Conference on Information and Knowledge Management; 2010. p. 579–88.
15. Liu Q, Ge Y, Li Z, Chen E, Xiong H. Personalized travel package recommendation. IEEE 11th International Conference on Data Mining (ICDM); 2011.
16. Xie M, Lakshmanan LV, Wood PT. Breaking out of the box of recommendations from items to packages. Proceedings of the 4thACM conference on Recommender systems; 2010. p. 151–58.
17. Yu Z, Feng Y, Xu H, Zhou X. Recommending travel packages based on mobile crowdsourced data. Communications Magazine, IEEE. 2014; 52(8):56–62.
18. Zheng Y, Xie X, Ma W-Y. GeoLife: a collaborative social networking service among user, location and trajectory. IEEE Data Eng Bull. 2010; 33(2):32–9.
19. Sarwat M, Levandoski JJ, Eldawy A, Mokbel MF. LARS*: a scalable and efficient location-aware recommender system. IEEE Transactions on Knowledge and Data Engineering (TKDE); 2013.
20. Levandoski JJ, Sarwat M, Eldawy A, Mokbel MF. Lars: a lo-

cation-aware recommender system. IEEE 28th International Conference on Data Engineering (ICDE); 2012.
21. Abraham RM, Rani BA, Manjula R. Data mining: building social network. sayali nishikant chakradeo. Indian Journal of Science and Technology. 2015 Jan; 8(S2):212–16.
22. Gharehchopogh FS, Khaze SR, Maleki I. A new approach in bloggers classification with hybrid of k-nearest neighbor and artificial neural network algorithms. Indian Journal of Science and Technology. 2015 Feb; 8(3):237–46.
23. Vedanayaki M. A study of data mining and social network analysis. Indian Journal of Science and Technology. 2014 Nov; 7(S7):185–87.