Scientific Journal of Impact Factor (SJIF): 4.14

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 3, Issue 11, November -2016

Knn Search on Road Networks by Incorporating Social Influence

Priyanka S Nambiar[1], Smita V Thakur[2], Pooja S Kadam[3], Geeta S Nagargoje[4], Supriya S Sonawane[5]

[1,2,3,4]*Department of Computer Engineering, P.K. Technical Campus, Chakan, Pune.*
[5]*Assistant Professor, Department of Computer Engineering, P.K. Technical Campus, Chakan, Pune.*

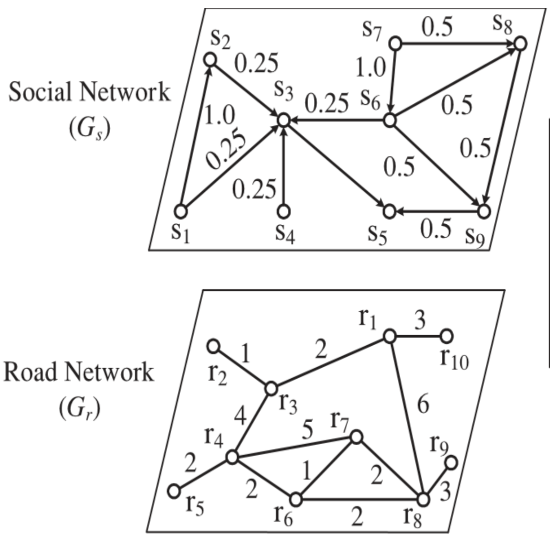
Abstract *- The existing system incorporates road network and social network. Independent Cascade (IC) model in social network is applied to define social influence. One of the challenge was to speed up the computation of the social influence over large road and social networks. To address this challenge, three efficient index-based search algorithms was proposed, i.e. road network-based (RN-based), social network-based (SN-based) and hybrid indexing algorithms. In the RN-based algorithm, employs a filtering-and-verification framework for dealing with the hard problem of computing social influence. SN-based algorithm, embed social cuts into the index, so to speed up the query. In the hybrid algorithm, index was proposed, summarizing the road and social networks, based on which query answers can be obtained efficiently. In proposed system recommendation is given based on the reviews of trusted users.*

Keywords *- KNN query, Social influence, Road Network, Social network*

I. INTRODUCTION

With the ever-growing quality of mobile devices (e.g., smartphones), location-based service (LBS) systems (e.g., Google Maps for Mobile) are wide deployed and accepted by mobile users. The k-nearest neighbor (kNN) search on road networks could be a basic drawback in LBS. Given a question location and a group of static objects (e.g., restaurant) on the road network, the kNN search drawback finds k nearest objects to the question location. Alone with the favored usage of LBS, the past few years have witnessed an enormous boom in location-based social networking services like Foursquare, Yelp, Loopt, Geomium and Facebook Places. All told these services, social network users are usually related to some locations (e.g., home/office addresses and visiting places). Such location info, bridging the gap between the physical world and also the virtual world of social networks, presents new opportunities for the kNN search on road networks.

The said example motivates U.S. to think about the social influence to a user once process the kNN search on road networks. Specifically, alphabetic a question user q would really like not solely retrieving k geographically nearest objects , however get an outsized social influence from q's friends UN agency are to . Therefore, during this paper, we have a tendency to study a completely unique query: kNN search on a road-social network (RSkNN), and propose economical question process algorithms. Specifically, given Gs, Gr and q, the RSkNN search finds k nearest objects (Aq =) to question q's location on Gr, specified the social influence SI(or) to Q through q's friends, UN agency are to or, is a minimum of a threshold.



user	location
s_1	r_4,r_5,r_6
s_2	r_2,r_3,r_4
s_4	r_5,r_6
s_5	r_7
s_7	r_1,r_8
s_8	r_9,r_{10}

Visited Relation

Figure 1. Example of road-social network. The visited relation gives a mapping that users have visited the locations[1].

II. RELATED WORK

In this paper [1] a kNN search on road networks by incorporating social influence (RSkNN). Independent Cascade(IC) model in social network is applied to dene social influence. One critical challenge of the problem is to speed up the computation of the social influence over large road and social networks. In Supercomputing [2]. Under certain reasonable assumptions that even if no revetment is used in the un-coarsening phase, a good bisection of the coarser graph is worse than a good bisection of then graph by at most a small factor. Models for the processes by which ideas and influence propagate through a social network have been studied in [3] a number of domains, Including the diffusion of medical and technological innovations, the sudden and widespread adoption of various strategies in game-theoretic settings, and the effects of word of mouth in the promotion of new products.

In this paper [4], The System can provide novel ranking methods that are different from the ICM, typical methods of social network analysis, and Page Rank method. Moreover, It experimentally demonstrate that when the propagation probabilities through links are small, they can give good approximations to the ICM for sets of influential nodes. A novel hybrid genetic algorithm (GA) [5] that globally optimal partition of a given data into a specie number of clusters. To circumvent these expensive operations, it hybridize GA with a classical gradient descent algorithm used in clustering, viz. K-means algorithm. Hence, the name genetic K-means algorithm (GKA).The location-aware influence maximization problem [6]. One big challenge in location-aware influence maximization is to develop an efficient scheme that offers wide influence spread. To address this challenge, it propose two greedy algorithms with 1 1/e approximation ratio. To meet the instant speed requirement, it propose two efficient algorithms with (1 1/e) approximation ratio for any (0,1).

In this paper [7]. User's interests are modeled by check-inactions. Here a Spatial-aware Interest Group (SIG) query that retrieves a user group of size k where each user is interested in the query keywords and they are close to each other in the Euclidean space. It prove that the SIG query problem is NP-complete. Approximation algorithms have developed in response to the impossibility of solving a great variety of important optimization problems [8]. Too frequently, when attempting to get a solution for a problem, one is confronted with the fact that the problem is NP-hard. While this is a significant theoretical step, it hardly qualifies as a cheering piece of news. Three simple efficient algorithms with good probabilistic behavior [9] two algorithms with run times of O(n(log n)2) which almost certainly nd directed (undirected) Hamiltonian circuits in random graph so fat least cn log n edges, and an algorithm with a runtime of O(n log n) which almost certainly a perfect matching in a random graph of at least cn log n edges.

In this paper [10] the systematic work on GeoSN query processing. Proposed a general frame work that offers exile data management and algorithmic design. Architecture segregates the social, geographical and query processing modules. Each GeoSN query is processed via a transparent combination of primitive queries issued to the social and geographical modules. Design a heuristic algorithm that is easily scalable to millions of nodes and edges in our experiments [11]. The study influence maximization in the linear threshold model, one of the important models formalizing the behavior of influence propagation in social networks [12]. It show that computing exact influence in general networks in the linear threshold model is P-hard, which closes an open problem left in the seminal work on influence maximization.

III. SYSTEM ARCHITECTURE

A. PROCESS BLOCK DIAGRAM

In architecture design, the first block is user here user will login to the system by providing password and username. In RNIndex Search block I_{RN} is road network index, G_s is social network and q is query is provided as input to the block and upper bound and lower bound is calculated. Next block is Filtering step G_s is road network and q_s is query set is offered as input and tight upper bound and lower bound is analyzed that is $SI(O_r)$. Next block is Verification block where G_r is road network, G_s is social network and q is query is delivered as input and true value of $SI(O_r)$ is estimated. It confirms if O_r is a valid answer. In SN Based Search block G_s is social network and SG is delivered as input and $SI(O_r)$ is estimated. Next block is Min Cut Cover in which V_r and V_s are accepted is used to calculate G^r_s is graph constructed for G_s and G_s and k is size of cutmark is provided as input and D the optimal cutmark set is analyzed. Combination of both RNIndex Search and SN Based Search is done in Hybrid Indexing. There is use of Database. Nearest location is given by using these algorithm and recommendation is provided. In road network according to latitude and longitude the distance are measured. By using filtering and sampling we get fine result through which we can get nearest location as per the query.

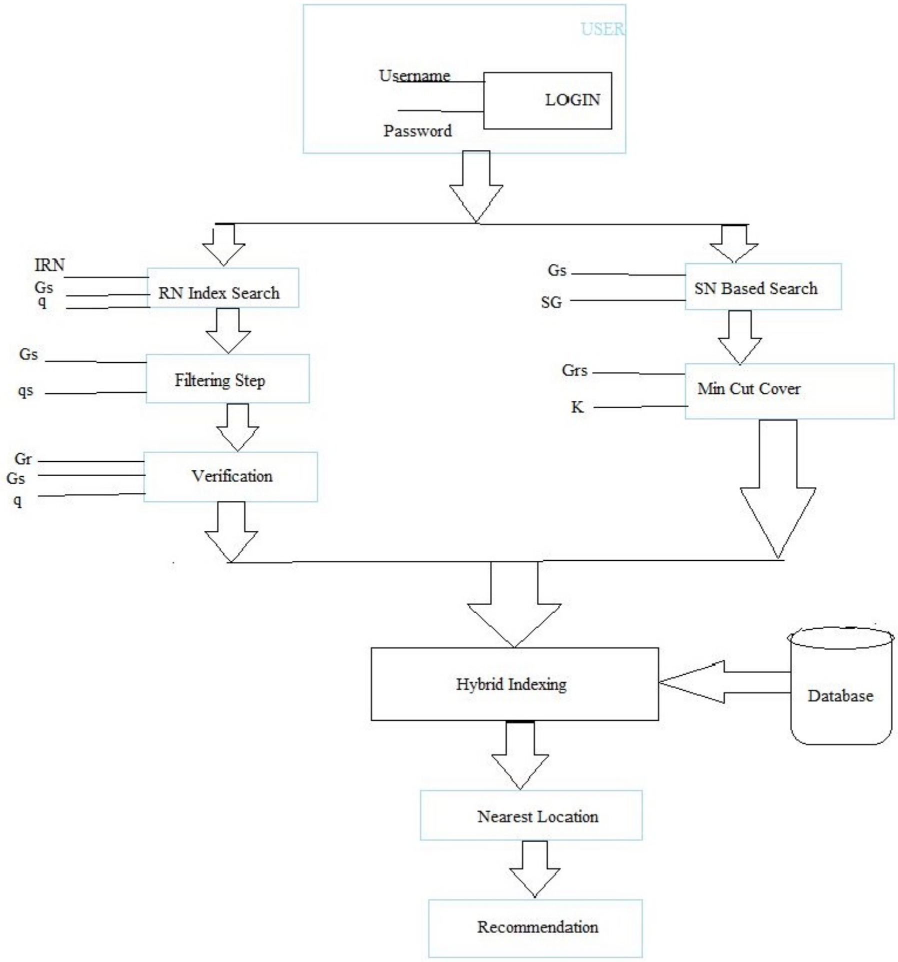


Figure 2. Architecture Diagram

Algorithm 1 RNIndex Search (*IRN*, *Gs*, *q* = <*qr*, *Cr*, *k*, >)
Require: The road network index *IRN*, social network *Gs* and query *q*;
Ensure: Query answer set *Aq*
1: $Aq = \varphi$;
2: **for** each returned object *or* ∈ *Cr* by the shortest-path algorithm from *qr* (in an increasing order of distance) by traversing *IRN* **do**
3: **if** *UpperBound* (*SI*(*or*)) < _ **then**
4: Prune object *or*;
5: **else if** *LowerBound* (*SI*(*or*)) > _ **then**
6: $Aq \leftarrow Aq \cup or$;
7: **else**
8: *SI*(*or*)=Sample(*Gr*, *Gs*, *q*);
9: **end if**
10: **if** *SI*(*or*) ≥ _ **then**
11: $Aq \leftarrow Aq \cup or$;
12: **end if**
13: **if** |*Aq*| == *k* **then**
14: **return** *Aq*;
15: **end if**

Algorithm 2 Sampling (G_s^o, *M*)
Require: Graph G_s^o, the sample size *M*;
Ensure: θ: the estimation of $SI(O_r)$
1: **for** *i* from 1 to *M* **do**
2: Initiate a flag *yi* = 0;

3: Sample edges of G_s^o according to the edge probabilities;
4: **if** (the current sampled graph contains an edge cut of G_s^o) **then**
5: Continue;
6: **end if**
7: **if** (the current sampled graph contains a path from *or* to *qs*) **then**
8: *y i*++;
9: Continue;
10: **end if**
11: **end for**
12: $\theta^{\sim}=\frac{\sum_{i=1}^{M} yi}{M}$

Algorithm 3 SNIndex_Prune (G_r, G_s, q= < q_r, C_r , k, ∈ >)

Require: A road network G_r, social network G_s and query q;

Ensure: The set of candidate objects C_q

1: $C_q = \Phi$
2: **for** (each o_r in C_r) **do**
3:Based on T find cutmarks {cj}⊂ Φ(or), such that c_j is an o_r;q_s-cut in $G^r{}_s$.
4: **if** (cutmarks fc$_j$g exist) then
Compute an upper bound for each c_j as the method in
5:Theorem 4, and then obtain the tightest upper bound UpperB(sI(or))
6: **if** (UpperB(SI(or)< ∈)**then**
7: $C_r = C_r \backslash o_r$;
8: **end if**
9: **end if**
10: **end for**
11: **return** $C_q = C_r$;

Algorithm 4 MinCutCover ($G^r{}_s$; k)

Require: Graph $G^r{}_s$ constructed from G_r and G_s; the size of cutmark set k
Ensure: The optimal cutmark set D
1: **if** (jDj < k) **then**
2: Run the all vertex-pair shortest path algorithm;
3: Choose two vertices s; d 2 $G^r{}_s$ with the largest shortest path distance;
4: Determine the minimum vertex cut c of ðs ;dÞ and remained sub graphs g $_1$;...;g$_f$ after removing c from $G^r{}_s$;
5: D ¼ D [c;
6: **for** (i from 1 to f) **do**
7: MinCutCover(g_i, k)
8: **end for**
9: **end if**
10: **return** D;

B. MATHEMATICAL MODEL

System Description:

- Let S is the Whole System Consist of
- S = {I, O, fme, DD, NP-Complete}
- I = Input.

- I = {U, Q, I_{RN}, G_s, M, $G^o{}_s$, G_r, K, $G^r s$, P, GPS, reviews}
  Where,
  U =Username
  P = Password
  Q = Query Entered by user
  Q = {q1, q2, q3.....qn}
  IRN = Road network index
  G_s = Social network
  M = Sample size
  $G^o{}_s$= Graph
  G_r = Road network
  K = Size of cut mark set
  $G^r{}_s$ = Graph constructed for G_s and G_r
- O = Number of output
- O ={y1,y2}
  Where,
  y1 = Display nearest places that user entered.
  y2 = Most visited places according to the reviews.
- fme ={f_1, f_2, f_3, f_4, f_5, f_6, f_7 }
  Where,
  f_1 is used for system login.
  f_1 (username, password)–> login
  f_2 is used for nearest location using KNN algorithm.
  f_2 (GPS)–> nearest location
  f_3 is used for providing recommendation by using reviews.
  F_3 (reviews)–>recommendation
  f_4 is used to calculate RNIndex by RNIndex search algorithm.
  f_4 (I_{RN}, G_s, q)–> RNIndex
  f_5 is used to calculate sample graph by using sampling algorithm.
  f_5 ($G^o{}_s$, M)–> Sample graph
  f_6 is used to calculate candidate set by using pruning algorithm.
  f_6 (G_r, G_s, q)–>Candidate set
  f_7 is used to calculate optical cut mark by using Hybrid Indexing algorithm.
  f_7 ($G^r{}_s$, k)–> Optimal cutmark
- DD = It is deterministic.
- NDD = If project contains large database, it is hard to determine.
- NP-Complete = this project is NP Complete.
- Successcondition=Whenuserwillgethisnearestlocationwhichhasbeenvisited by their friends, and the reviews enter by the friends will be appropriate.
- Failure condition = It will fail when a friend fails to enter reviews.

IV. PERFORMANCE AND EVALUATION

The search query is entered by user and outcome that is venue recommendation is provided based on friends reviews and user will get nearest venues.
Some websites are mentioned below. These are the datasets used.

1 .http://snap.stanford.edu/data/loc-gowalla.html
2. http://www.cs.fsu.edu/lifeifei/SpatialDataset.html
3. http://www.dis.uniroma1.it/challenge9/index.shtml

Dataset information
Gowalla is a location-based social networking website where users share their locations by checking-in. The friendship network is undirected and was collected using their public API, and consists of 196,591 nodes and 950,327 edges.

Table 1. Real Datasets

Datasets	# Vertices	# Edges
San Francisco (SF)	174,828	223,476
Florida (FL)	1,068,615	1,357,204
Gowalla (GO)	196,591	950,327
Foursquare (FO)	4,520,675	47,191,155

V. CONCLUSION AND FUTURE SCOPE

In this paper there is a feasible solution and query is answered within a specific time. There is a joint social and road processing on networks stored in a distributed manner. There is a use of datasets for getting the past records of most popular visited places. There is a use of GPS for finding the road network and there will be a dummy network for implementing this work. Future work can be adding that location where user have not been visited and that recommendation can be provided by number of person visiting that particular place.

VI. REFERENCES

[1] Ye Yuan, Xiang Lian, Lei Chen, Member, IEEE, Yongjiao Sun, and Guoren Wang, Member, IEEE,"*RSkNN: kNN Search on Road Networks by Incorporating Social Influence*" ,IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 28, NO. 6, JUNE 2016

[2] G. Karypis and V. Kumar,"*Analysis of multilevel graph partitioning*," in Proc. ACM/IEEE Conf. Supercomputer, 1995, p. 29

[3] D. Kempe, J. Kleinberg, and E. Tardos,"*Maximizing the spread of influence through a social network*," in Proc. 9th ACM SIGKDD Int. Conf. Knowledge. Discovery Data Mining, 2003, pp. 137146

[4] M. Kimura and K. Saito,"*Tractable models for information diffusion in social networks*," in Proc. 10thEur. Conf. Principle Practice Knowledge. Discovery Databases, 2006, pp. 259271.

[5] K. Krishna and M. N. Murty,"*Genetic k-means algorithm,"* IEEE Trans. Syst., Man, Cybern. B, Cybern, vol. 29, no. 3, pp. 433439, Jun. 1999.

[6] G. Li, S. Chen, J. Feng, K.-l. Tan, and W.-S. Li,"*Efficient location aware influence maximization*," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2014, pp. 8798.

[7]Y.Li, D.Wu, J.Xu, B.Choi, andW.Su,"*Spatial-aware interest group queries in location-based social networks*," Data Knowl. Eng., vol. 92, pp. 2038, 2014.

[8] D. Hochbaum Ed.,"*Approximation Algorithms for NP-Hard Problems,"* PWS Publishing Company, 1997.

[9] D. Angluin and L. G. Valiant,"*Fast probabilistic algorithms for hamiltonian circuits and matching's*," in Proc. 9th Annu. ACM Symp. Theory Computation, 1977, pp. 3041.

[10] N. Armenatzoglou, S. Papadopoulos, and D. Papadias, "*A general framework for geo-social query processing*," in Proc. VLDB Endowment, vol. 6, no. 10, pp. 913924, 2013.

[11]W.Chen, C.Wang, and Y.Wang,"*Scalable influence maximization for prevalent viral marketing in large-scale social networks*," in Proc. 16th ACM SIGKDD Int. Conf. Knowledge. Discovery Data Mining, 2010, pp. 10291038.

[12]W.Chen, Y.Yuan, and L.Zhang,"*Scalable influence maximization in social networks under the linear threshold model,*" in Proc. IEEE Int. Conf. Data Mining, 2010, pp. 8897.

