Geo-social group queries with minimum acquaintance constraints

Qijun Zhu[1], Haibo Hu[2], Cheng Xu[1], Jianliang Xu[1], and Wang-Chien Lee[3]

[1]Hong Kong Baptist University [2]Hong Kong Polytechnic University [3]Pennsylvania State University

{qjzhu, chengxu, xujl}@comp.hkbu.edu.hk haibo.hu@polyu.edu.hk wlee@cse.psu.edu

Problem Statement

Given an location-based social networking services (LBSN), a *Geo-social group query with minimum acquaintance constraint* (GSGQ) finds a maximal user result set which satisfies both social constraint and spatial constraint.

- **Social Constraint**
  - A subgraph is **c-core** if its minimum degree is larger or equal to c.
  - The user result set and the query user of GSGQ should form a c-*core*, which denotes the social acquaintance constraint.
- **Spatial Constraint**
  - **GSGQ with range constraint** ($GSGQ_{range}$)
    It finds the largest c-core located inside range.
    e.g. find me the largest user group satisfying c-core in 5th Avenue, Manhattan, NYC.
  - **GSGQ with relaxed kNN constraint** ($GSGQ_{rkNN}$)
    It finds a maximal c-core of size no less than $k+1$ with minimum distance.
    e.g. find me the closest (maximal) group of at least nine users satisfying c-core to be eligible for a bulk discount.
  - **GSGQ with strict kNN constraint** ($GSGQ_{kNN}$)
    It finds a c-core of size equal to $k+1$ with minimum distance.
    e.g. find me the closest group of three users satisfying c-core to play tennis doubles with me.

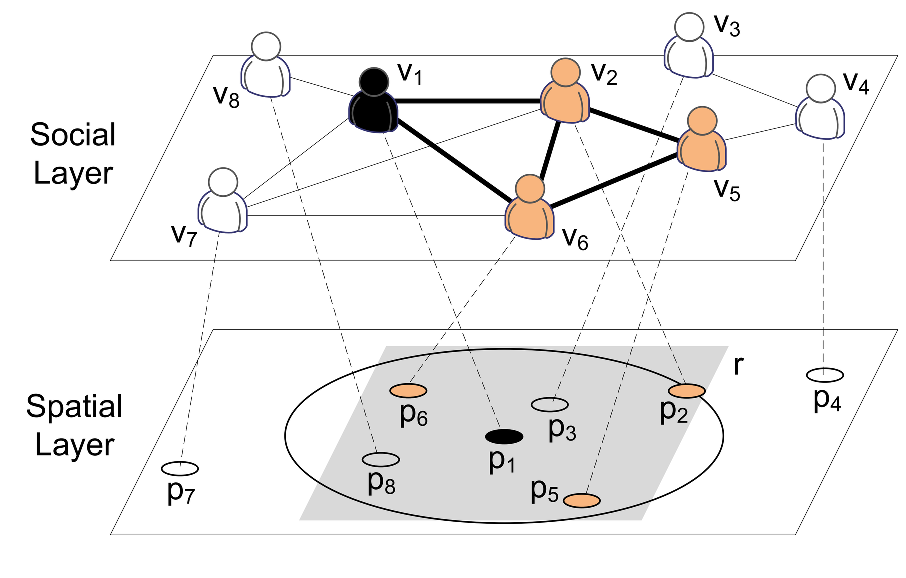


Fig. 1: An example of GSGQ $< v_1, 3NN, 2 >$. Lines between the users represent acquaintance relations and the points on the spatial layer denote the positions of the users.

R-tree-based Query Processing

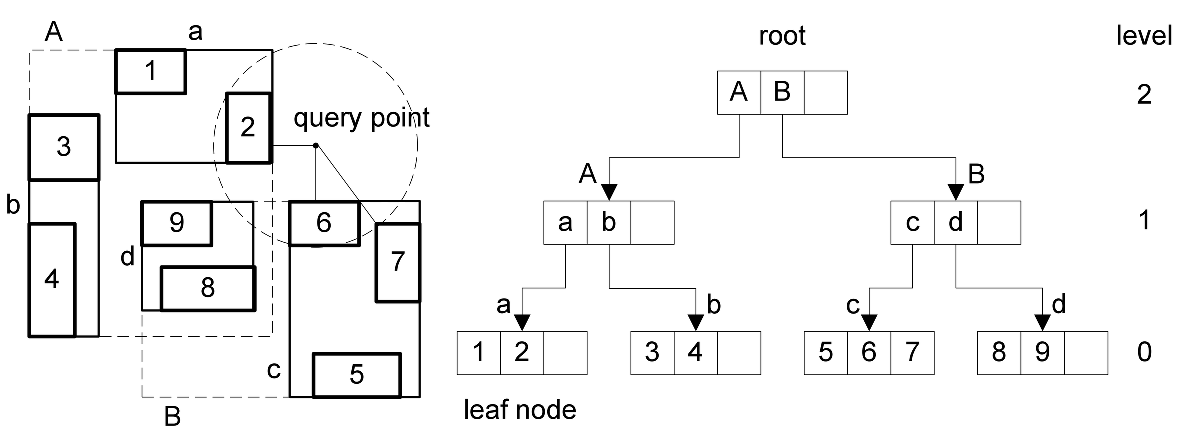


Fig. 2: An example of R-tree

$GSGQ_{range}$ Find all users in the range using R-tree and compute the c-core subgraph W' formed by these users. If $v \in W'$, $W = W' - \{v\}$ is the answer, otherwise there is no result.

$GSGQ_{rkNN}$ Perform kNN search on R-tree. When the size of the candidate set $\tilde{W}$ exceeds k, compute c-core subgraph W' formed by $\tilde{W}$. If $v \in W'$ and $|W'| \geq k+1$, $W = W' - \{v\}$ is the answer, otherwise continue the search.

$GSGQ_{kNN}$ Perform kNN search on R-tree. When the size of the candidate set $\tilde{W}$ exceeds k, enumerate all possible subsets with the size of $k+1$ and containing v. If such a user set W' is a c-core, $W = W' - \{v\}$ is the answer.

Core Bounding Rectangle (CBR)

CBR of a user
Consider a user $v \in G$. Given a minimum degree constraint c, $CBR_{v,c}$ is a rectangle which contains v and inside which any user group with v (excluding the users on the bounding edges) cannot be a c-core.

Computing a CBR

- Build an initial CBR from nearest users which cannot form a c-core.
- Expand the initial CBR outwards to gain the maximal CBR.

CBR of an entry
Consider an entry e with MBR MBR_e and user set V_e. Given a minimum degree constraint c, $CBR_{e,c}$ is a rectangle which intersects MBR_e and inside which any user group containing any user from V_e (not including the users on the bounding edges) cannot be a c-core.

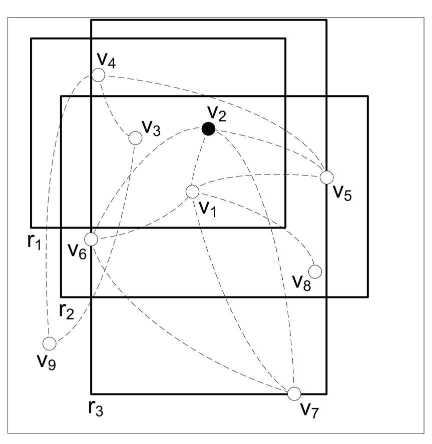


Fig. 3: An example of CBR. r_1 and r_3 are $CBR_{v_2,2}$. r_2 is not, because $\{v_2, v_1, v_6\}$ forms 2-core.

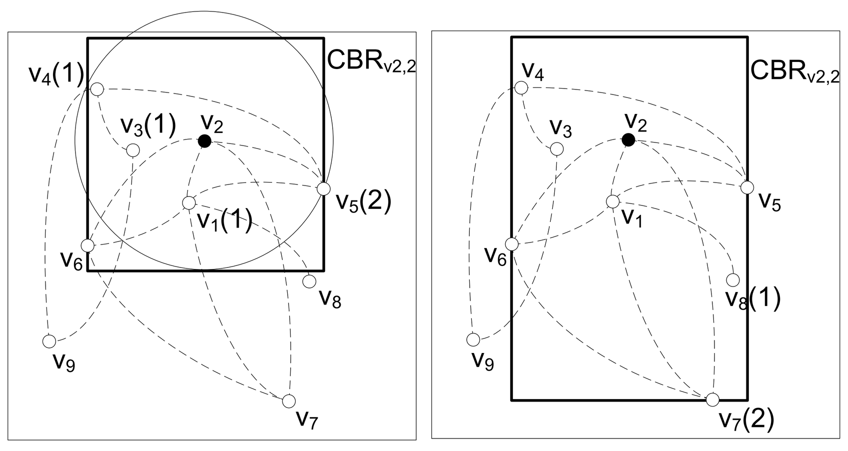


Fig. 4: An example of computing $CBR_{v_2,2}$.

Social-aware R-tree (SaR-tree)

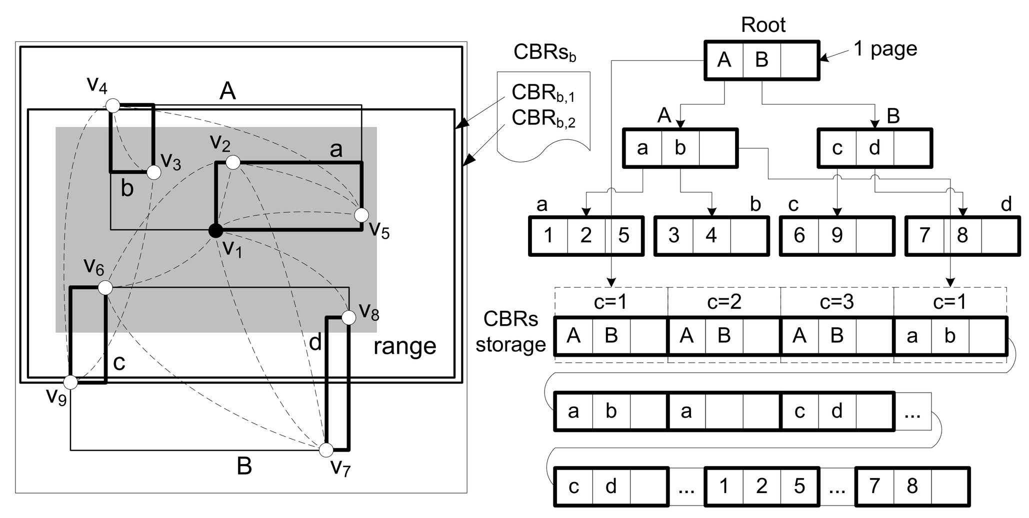


Fig. 5: An example of SaR-tree

- Different from a conventional R-tree, each entry of an SaR-tree refers to two pieces of information, i.e., a set of CBRs and an MBR, to describe the group of users it covers.
- **SaR*-tree**: Inspired by R*-tree, use both CBRs and MBR as closeness metric.

$$I(V) = ||\text{MBR}_V|| \cdot \sum_c (|| \cup_{v \in V} \text{CBR}_{v,c} - \text{CBR}_{V,c}||)$$

GSGQ Processing

$GSGQ_{range}$ If the range is covered by a $CBR_{e,c}$, entry e can be pruned.

$GSGQ_{rkNN}$ The sorting key of the kNN search is $d_e = \max\{d(v, \text{MBR}_e), d_{in}(v, \text{CBR}_{e,c})\}$, where

$$d_{in}(v, \text{CBR}_{e,c}) = \begin{cases} \min_{l \in L_{\text{CBR}_{e,c}}} d(v,l), & v \in \text{CBR}_{e,c} \\ 0, & \text{otherwise.} \end{cases}$$

$GSGQ_{kNN}$ Similar to $GSGQ_{rkNN}$, with additional pruning strategies to find result from candidates.

- Core decomposition-based pruning.
- k-plex-based pruning.

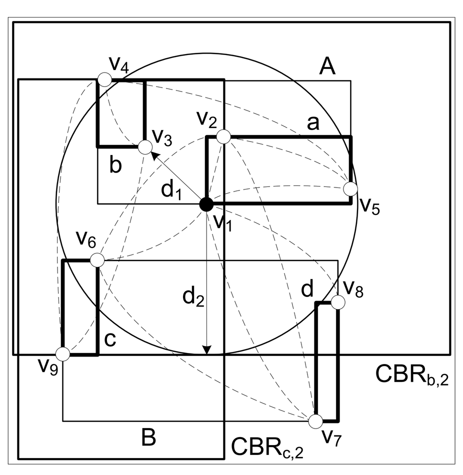
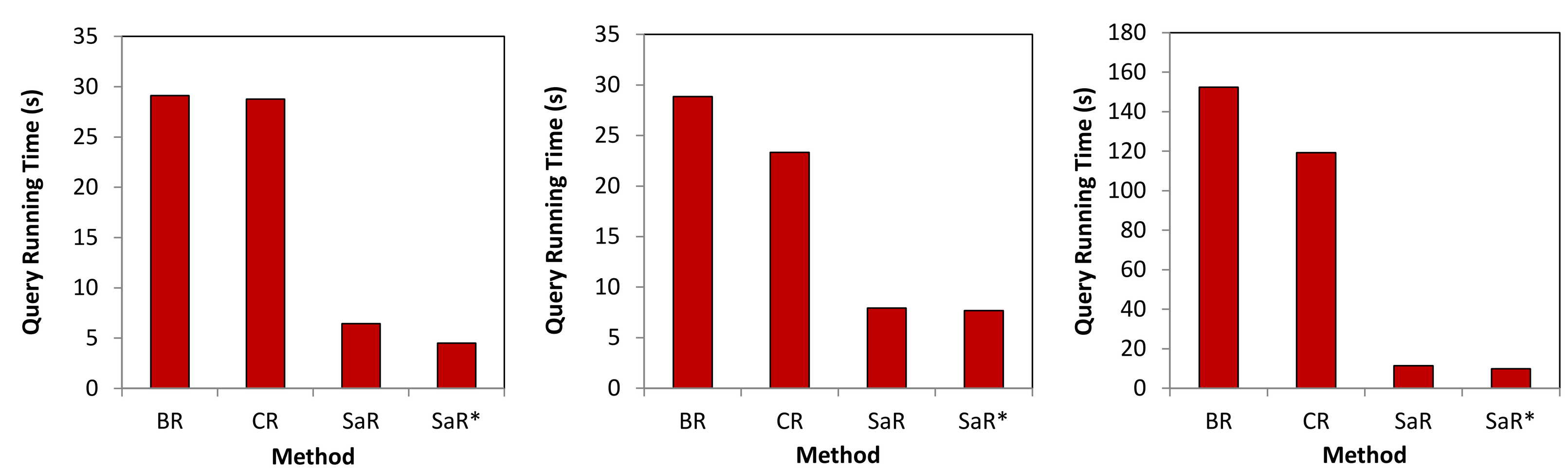


Fig. 6: An example of processing a $GSGQ_{rkNN} < v_1, r3NN, 2 >$. Entry c will be visited before entry b.

Performance Evaluation



(a) $GSGQ_{range}$ ($r = 0.05, c = 2$) (b) $GSGQ_{rkNN}$ ($k = 20, c = 2$) (c) $GSGQ_{kNN}$ ($k = 20, c = 2$)

Fig. 7: Overall performance comparison on Twitter-2010 (41M users, 684M edges).

- Q. Zhu, H. Hu, C. Xu, J. Xu, and W.-C. Lee, "Geo-social group queries with minimum acquaintance constraints," The VLDB Journal (VLDBJ'17), Jul. 2017.