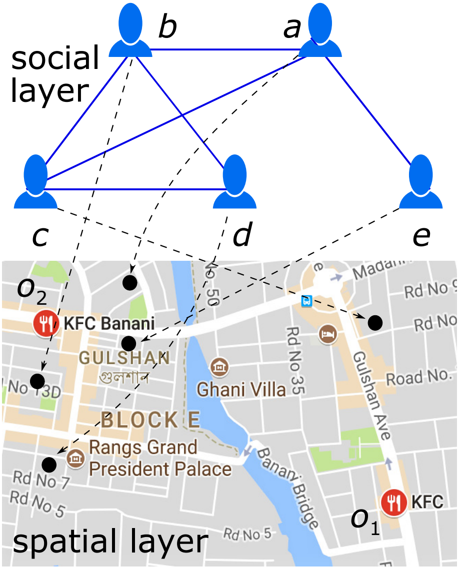
The Flexible Socio Spatial Group Queries

Bishwamittra Ghosh[1], Mohammed Eunus Ali[2], Farhana M. Choudhury[3], Sajid Hasan Apon[2], Timos Sellis[4], Jianxin Li[5]

VLDB 2019

[1]National University of Singapore
[2]Bangladesh University of Engineering and Technology
[3]RMIT University and University of Melbourne, Australia
[4]Swinburne University of Technology, Australia
[5]The University of Western Australia

Socio-spatial Graph



Problem Formulation

Given

- Set of meeting points Q
- Socio-spatial graph $G = (V, E)$

Find top k groups such that

$$\text{score}(G_i, q_i) \geq \text{score}(G_{i+1}, q_{i+1})$$

where G_i is a subgraph of G, $q_i \in Q$ and $1 \leq i \leq k-1$

Constraints for a feasible group $G_i = (V, E)$

- minimum social connectivity constraint c
  - $\text{degree}(v) \geq c, \forall v \in V$
- maximum distance d_{max}
  - $\text{dist}(v, q) \leq d_{max}, \forall v \in V$
- minimum group size n_{min}, maximum group size n_{max}
  - $n_{min} \leq |V| \leq n_{max}$

Score of group $G_i = (V, E)$ w.r.t. meeting point q

$$\text{score}_{\text{social}} = \frac{2|E|}{|V|(|V|-1)}$$

$$\text{score}_{\text{spatial}} = 1 - \frac{\sum_{v \in V} \text{dist}(v, q)}{d_{max}|V|}$$

$$\text{score}_{\text{size}} = \frac{|V|}{n_{max}}$$

$$\text{score} = \alpha \cdot \text{score}_{\text{social}} + \beta \cdot \text{score}_{\text{spatial}} + \gamma \cdot \text{score}_{\text{size}}$$

Literature review

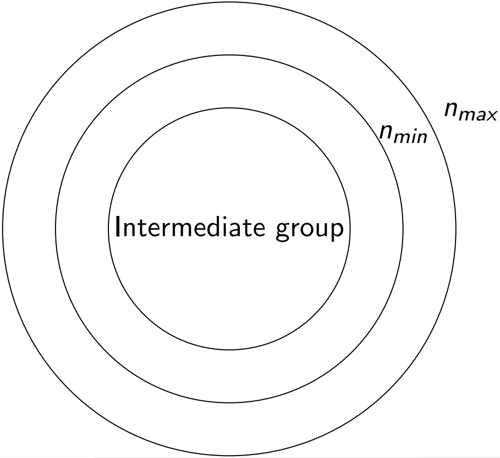
There are existing works that address socio spatial group queries. The major gaps are

- specific group size[6] vs variable group size
- finding only the best group[6] vs top *k* groups
- fixed meeting point vs multiple meeting points[7]
- average social connectivity constraint[8] vs minimum social connectivity constraint[9]
- ranking function combining social and spatial factors[10] vs ranking function combining social, spatial and group size factors

[6] [Fang17], [Shen16], [Zhu14],[Yang12]
[7] [Shen16]
[8] [Shen16], [Yang12]
[9] [Fang17],[Zhu14]
[10] [Armenatzoglou15]

Contribution

- Exact algorithm
  - member ordering based on spatial distance
  - optimistic assumption (maximum) on social connectivity of including members
  - early termination based on upper bound on spatial distance



Continued...

- Heuristic approximate approach
  - member ordering based on spatial distance
  - lower bound on social connectivity while including a member in the intermediate group

Continued. . .

- A fast approximate approach
  - a tighter lower bound on social connectivity while including a member in the intermediate group
  - upper bound on spatial distance and lower bound on social connectivity that improves the rank of current exploring group
  - prune when including a member can not increase the score of intermediate group
- Greedy approach
  - avoid backtracking

Simulation

- meeting point q_1
- distance ordered members $\{a, b, c, d \dots\}$

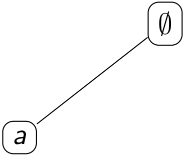
$\boxed{\emptyset}$

- meeting point q_2
- distance ordered members $\{b, a, c, \dots\}$

$\boxed{\emptyset}$

Simulation

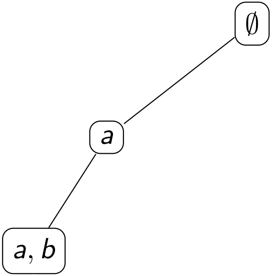
- meeting point q_1
- distance ordered members $\{a, b, c, d \ldots\}$



- meeting point q_2
- distance ordered members $\{b, a, c, \ldots\}$

Simulation

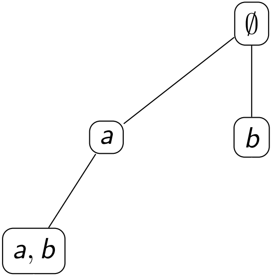
- meeting point q_1
- distance ordered members $\{a, b, c, d \dots\}$



- meeting point q_2
- distance ordered members $\{b, a, c, \dots\}$

Simulation

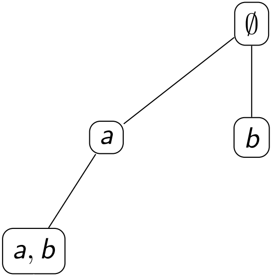
- meeting point q_1
- distance ordered members $\{a, b, c, d \dots\}$



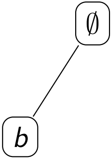
- meeting point q_2
- distance ordered members $\{b, a, c, \dots\}$

Simulation

- meeting point q_1
- distance ordered members $\{a, b, c, d \dots\}$



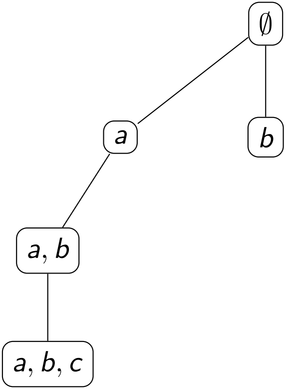
- meeting point q_2
- distance ordered members $\{b, a, c, \dots\}$



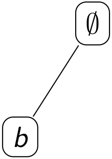
select meeting point that has minimum spatial distance to first unexplored member

Simulation

- meeting point q_1
- distance ordered members $\{a, b, c, d \ldots\}$



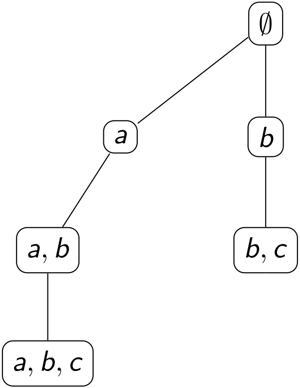
- meeting point q_2
- distance ordered members $\{b, a, c, \ldots\}$



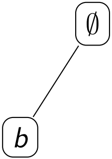
$\{a, b, c\}$ is a result group

Simulation

- meeting point q_1
- distance ordered members $\{a, b, c, d \ldots\}$

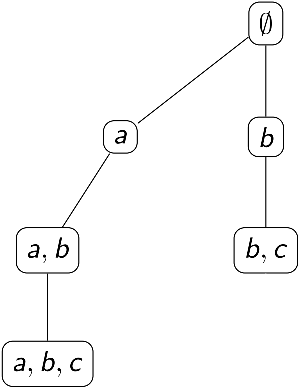


- meeting point q_2
- distance ordered members $\{b, a, c, \ldots\}$

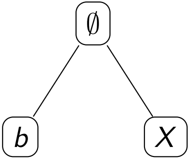


Simulation

- meeting point q_1
- distance ordered members $\{a, b, c, d \ldots\}$



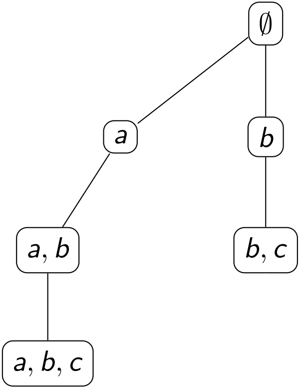
- meeting point q_2
- distance ordered members $\{b, a, c, \ldots\}$



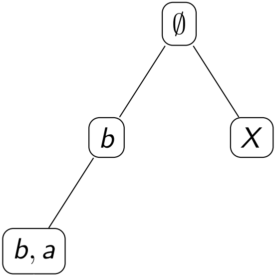
Advance termination based on upper bound on spatial distance

Simulation

- meeting point q_1
- distance ordered members $\{a, b, c, d \dots\}$

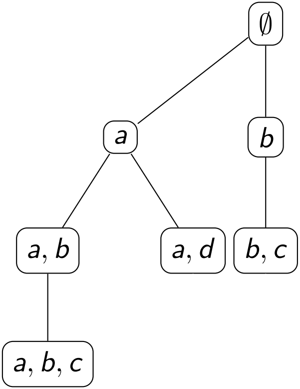


- meeting point q_2
- distance ordered members $\{b, a, c, \dots\}$

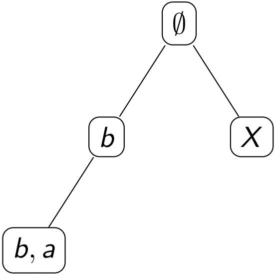


Simulation

- meeting point q_1
- distance ordered members $\{a, b, c, d \ldots\}$



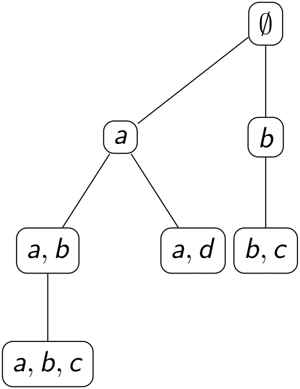
- meeting point q_2
- distance ordered members $\{b, a, c, \ldots\}$



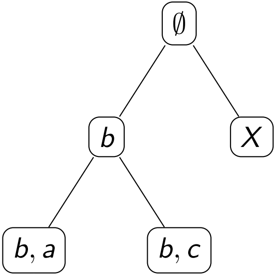
degree$(c, \{a\}) <$ lower bound on social connectivity

Simulation

- meeting point q_1
- distance ordered members $\{a, b, c, d \dots\}$



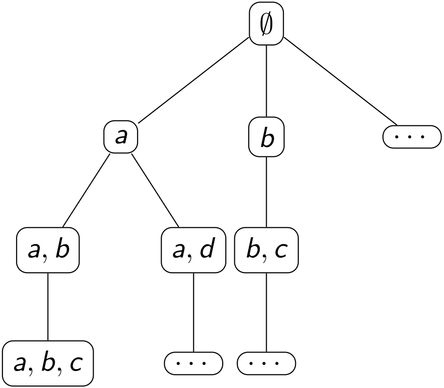
- meeting point q_2
- distance ordered members $\{b, a, c, \dots\}$



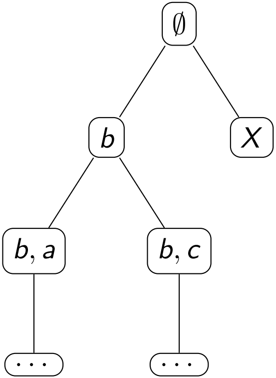
degree$(c, \{b\}) \geq$ lower bound on social connectivity

Simulation

- meeting point q_1
- distance ordered members $\{a, b, c, d \dots\}$



- meeting point q_2
- distance ordered members $\{b, a, c, \dots\}$



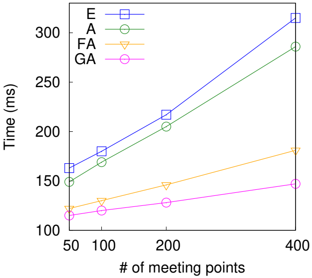
Approximation ratio of fast approximate algorithm

$$\text{approximation ratio} = \frac{\text{lowest scoring retrieved group}}{\text{best scoring group that may not be retrived}}$$

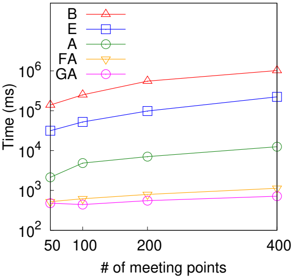
Emphasis	Weights	Approximation ratio
Social score	$\alpha = 1, \beta = \gamma = 0$	$\frac{c}{n_{max}-1}$
Spatial score	$\beta = 1, \alpha = \gamma = 0$	1
Size score	$\gamma = 1, \alpha = \gamma = 0$	$\frac{n_{min}}{n_{max}}$

Experimental Results

B = Baseline[11], E = Exact, A = Approximate, FA = Fast approximate, GA = Greedy approximate



(a) Brightkite



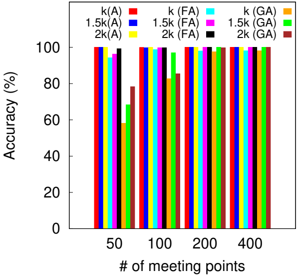
(b) Gowalla

Figure: Computation time of different algorithm

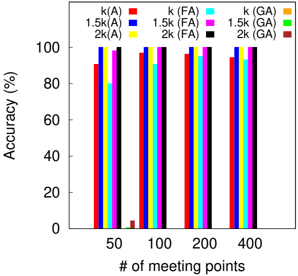
[11] [YANG12]

Experimental Results

A = Approximate, FA = Fast approximate, GA = Greedy approximate



(a) Brightkite



(b) Gowalla

Figure: Percentage of groups in top k of approximate algorithm that also appear in top k, top $1.5k$, and top $2k$ of the exact algorithm

Conclusion

- we propose novel top k flexible social spatial group queries
- we devise a ranking function combining social closeness, spatial distance, and group size
- we propose exact algorithm and efficient approximate algorithms
- Exact algorithm runs up to $10\times$ faster than the baseline
- Fast approximate algorithm runs up to $100\times$ faster than exact algorithm and returns the same set of results in most cases

Thank You