The Flexible Socio Spatial Group Queries

Bishwamittra Ghosh\textsuperscript{1}, Mohammed Eunus Ali\textsuperscript{2}, Farhana M. Choudhury\textsuperscript{3}, Sajid Hasan Apon\textsuperscript{2}, Timos Sellis\textsuperscript{4}, Jianxin Li\textsuperscript{5}

VLDB 2019

\textsuperscript{1}National University of Singapore
\textsuperscript{2}Bangladesh University of Engineering and Technology
\textsuperscript{3}RMIT University and University of Melbourne, Australia
\textsuperscript{4}Swinburne University of Technology, Australia
\textsuperscript{5}The University of Western Australia
Socio-spatial Graph

The Flexible Socio Spatial Group Queries

VLDB 2019
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 $d_{max}$
  - $\text{dist}(v, q) \leq d_{\text{max}}, \forall v \in V$

- minimum group size $n_{\text{min}}$, maximum group size $n_{\text{max}}$
  - $n_{\text{min}} \leq |V| \leq n_{\text{max}}$
Score of group $G_i = (V, E)$ w.r.t. meeting point $q$

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

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

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

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

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

- specific group size vs variable group size
- finding only the best group vs top \( k \) groups
- fixed meeting point vs multiple meeting points
- average social connectivity constraint vs minimum social connectivity constraint
- ranking function combining social and spatial factors 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
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 cannot increase the score of the intermediate group

- Greedy approach
  - avoid backtracking
Simulation

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

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

- 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\}
Simulation

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

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

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 
  $\{a, b, c, d \ldots\}$

- 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
  $\{b, a, c, \ldots \}$

$\deg(c, \{a\}) < \text{lower bound on social connectivity}$
Simulation

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

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

\[
\text{degree}(c, \{b\}) \geq \text{lower bound on social connectivity}
\]
Meeting point $q_1$
- Distance ordered members
\{a, b, c, d \ldots \}

Meeting point $q_2$
- Distance ordered members
\{b, a, c, \ldots \}
Approximation ratio of fast approximate algorithm

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

<table>
<thead>
<tr>
<th>Emphasis</th>
<th>Weights</th>
<th>Approximation ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social score</td>
<td>$\alpha = 1$, $\beta = \gamma = 0$</td>
<td>$\frac{c}{n_{max} - 1}$</td>
</tr>
<tr>
<td>Spatial score</td>
<td>$\beta = 1$, $\alpha = \gamma = 0$</td>
<td>1</td>
</tr>
<tr>
<td>Size score</td>
<td>$\gamma = 1$, $\alpha = \gamma = 0$</td>
<td>$\frac{n_{min}}{n_{max}}$</td>
</tr>
</tbody>
</table>
Experimental Results

B = Baseline\(^\text{11}\), E = Exact, A = Approximate, FA = Fast approximate, GA = Greedy approximate

Figure: Computation time of different algorithm
Experimental Results

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

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