ON SOCIAL-TEMPORAL GROUP QUERY WITH ACQUAINTANCE CONSTRAINT

D.N Yang
Y.L Chen
W.C Lee
M.S Chen

Sep 2011

Presented by
Roi Fridburg
AGENDA

- Activity Planning
- Social Graphs
- Proposed Algorithms
  - SGSelect
  - SGTSelect
- Experimental Results
- Efficiency and Reliability
- Conclusions
ACTIVITY PLANNING

- How do we plan an activity?
- Who do we invite?

Available time
Social relations with the members
Familiarity with the initiator

Familiarity with the initiator
Social relations with the members
Available time
The purpose: automatic activity planning

Proposal:

• **Social-Temporal Group Query (STGQ)** to find **activity time and attendees with the minimum total social distance to the initiator incorporating an acquaintance constraint**
ILLUSTRATING EXAMPLE
SOCIAL GRAPH OF CASEY AFFLECK

q = Casey Affleck
ILLUSTRATING EXAMPLE
SOCIAL GRAPH OF MATT DAMON

- $k=0$ each attendee knows everybody else
- $F$: \{M.Damon, K.Costner, R.De Niro\}
- $F$: \{M.Damon, R.De Niro, A.Jolie\}
- $F$: \{M.Damon, N.Cage, J.Roberts\}
- $F$: \{M.Damon, R.De Niro, J.Roberts\}
- Which one should we choose?

**Answer:** The one that has the minimum total distance between $q$ and every vertex $v$. 

$$\min \sum_{v \in F} d_{v,q}$$

$q = \text{Matt Damon}$
k=1 each attendee has at most 1 unfamiliar attendee

For example:
- F: \{M.Damon, K.Costner, R.De Niro, A.Jolie\}

There is no solution for k=0 in this case!
**Basic Terms and Parameters**

- **q** – activity initiator
- **G** – social graph of q
- **p** – activity size
  - number of expected attendees
- **s** – social radius
  - Scope of candidate attendees
  - All acquainted attendees are located no more than \( s \) edges away from the initiator on his social graph
- **k** – acquaintance constraint
  - Social relationship between attendees
  - Each attendee can have at most \( k \) other unacquainted attendees
- **m** – activity length of time (slots)
- **d** – social distance
  - Distance between vertices in the group
  - How “close” the vertices are to each other
ACTIVITY EXAMPLES

- A person has a given number of tickets for a movie and would like to invite some friends
  - Set of mutually close friends and the time that all of them are available
    - small $k \rightarrow$ all attendees know each other very well
  - $s=1 \rightarrow$ all attendees are acquainted with him
ACTIVITY EXAMPLES – CONT.

- A person would like to arrange a party and invite some friends
  - large $s \to$ friends of friends can be also invited
  - large $k \to$ more diverse event
WHERE IS ALL THE DATA TAKEN FROM?

- **Social network websites** – social relations and contact information
  - Facebook
  - LinkedIn

- **Web collaboration tools** – shared available time
  - Google calendar

- **Enterprise tools**
  - MS Outlook
ILLUSTRATING EXAMPLE WITH TIME CONSTRAINT

\[ p = 4; s = 1; k = 0; m = 3 \]

q = Casey Affleck
ILLUSTRATING EXAMPLE WITH TIME CONSTRAINT

\[ C_3^5 = 10 \]

\[ \{v_7, v_2\} \]

\[ \{v_7, v_3\} \]

\[ \{v_7, v_4\} \]

\[ \{v_7, v_2, v_3\} \]

\[ \{v_7, v_2, v_4\} \]

\[ \{v_7, v_2, v_6\} \]

\[ \{v_7, v_3, v_4\} \]

\[ \{v_7, v_3, v_6\} \]

\[ \{v_7, v_3, v_4, v_6\} \]

\[ \{v_7, v_4, v_6\} \]

\[ \{v_7, v_2, v_3, v_4\} \]

\[ \{v_7, v_2, v_3, v_6\} \]

\[ \{v_7, v_2, v_3, v_4, v_6\} \]

\[ \{v_7, v_2, v_6, v_8\} \]

\[ \{v_7, v_3, v_4, v_6\} \]

\[ \{v_7, v_3, v_6, v_8\} \]

\[ \{v_7, v_6, v_8\} \]

\[ \{v_7, v_4, v_6\} \]

\[ \{v_7, v_4, v_8\} \]

\[ \{v_7, v_6, v_8\} \]

\[ \{v_7, v_8\} \]

\[ \{v_7\} \]

\[ d = 64 \]

\[ d = 65 \]

\[ p = 4; s = 1 \]

\[ k = 0; m = 3 \]

\[ \begin{array}{ccccccc}
 t & 1 & 2 & 3 & 4 & 5 & 6 \\
 \hline
 v_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 v_2 & 0 & 0 & 0 & 0 & 0 & 0 \\
 v_3 & 0 & 0 & 0 & 0 & 0 & 0 \\
 v_4 & 0 & 0 & 0 & 0 & 0 & 0 \\
 v_5 & 0 & 0 & 0 & 0 & 0 & 0 \\
 v_6 & 0 & 0 & 0 & 0 & 0 & 0 \\
 v_7 & 0 & 0 & 0 & 0 & 0 & 0 \\
 v_8 & 0 & 0 & 0 & 0 & 0 & c \\
\end{array} \]
COMPLEXITY AND CANDIDATE GROUPS

- Simple approach: consider every possible $p$ attendees and find the total social distance

- Needs to evaluate $\binom{f - 1}{p - 1}$ candidate groups
  - $f$ is the number of candidate attendees

- If an initiator would like to invite 10 attendees out of his 100 friends, the number of candidate groups is in the order of $10^{13}$
  - Scenario considers only the friends of the user

- This problem is NP-hard
SOCIAL GROUP QUERY (SGQ) PROBLEM DEFINITION

- Given an activity initiator $q$ and his social graph $G=(V,E)$
  - $v$ is a candidate attendee
  - $e_{u,v}$ is the distance between $u$ and $v$ representing their social closeness
- Need to find $SGQ(p,s,k)$ which satisfies all the conditions
SOCIAL GROUP QUERY (SGQ) ALGORITHM DESIGN

- Proposed algorithm: SGSelect
- Step 1: derive feasible graph $G_F = (V_F, E_F)$ from $G$ which satisfies:
  - Social radius constraint
  - $d_{v,q}$ is the minimal social distance between $v$ and $q$

$\begin{align*}
  d_{v_{3},q} &= 8 \\
  d_{v_{6},q} &= 16
\end{align*}$

$q = \text{Ryan Giggs}$
How to create the feasible graph?

Define $d_{v,q}^i$ the $i$-edge minimum distance between the vertex $v$ and the vertex $q$

- $1 \leq i \leq s$
**RADIUS GRAPH EXTRACTION – CONT.**

- **Target**
  - \( G(V, E) \rightarrow G_F(V_F, E_F) \)
  - \( d_{v,q} \) satisfying s constraint

1. \( d_{q,q}^0 = 0, d_{u,q}^0 = \infty \) for \( u \neq q \);
2. for \( i = 1 \) to \( s \) do
3. \( d_{q,q}^i = 0; \)
4. for all vertex \( u \neq q \) in \( V \) do
5. \( d_{u,q}^i = d_{u,q}^{i-1}; \)
6. for all vertex \( v \) in \( N_u \) do
7. if \( d_{v,q}^{i-1} + e_{u,v} < d_{u,q}^i \) then
8. \( d_{u,q}^i = d_{v,q}^{i-1} + e_{u,v}; \)
9. Extract all vertices \( w \) in \( V \) with \( d_{w,q}^s < \infty \) and form the set \( V_f \)

\[
\begin{align*}
d_{q,q}^0 &= 0 & d_{q,q}^1 &= 0 & d_{q,q}^2 &= 0 \\
d_{u2,q}^0 &= \infty & d_{u2,q}^1 &= 3 & d_{u2,q}^2 &= 3 \\
d_{u3,q}^0 &= \infty & d_{u3,q}^1 &= 10 & d_{u3,q}^2 &= 8 \\
d_{u4,q}^0 &= \infty & d_{u4,q}^1 &= 4 & d_{u4,q}^2 &= 4 \\
d_{u5,q}^0 &= \infty & d_{u5,q}^1 &= \infty & d_{u5,q}^2 &= 10 \\
d_{u6,q}^0 &= \infty & d_{u6,q}^1 &= \infty & d_{u6,q}^2 &= 16 \\
d_{u7,q}^0 &= \infty & d_{u7,q}^1 &= \infty & d_{u7,q}^2 &= \infty \\
\end{align*}
\]

q = Ryan Giggs
ACCESS ORDERING

- The selection of a vertex at each iteration is critical
  - Avoid selecting a vertex \( v \) that
    - Increases the total social distance
    - Leads to violation of the acquaintance constraint
  - It is preferable to first include
    - Well-connected vertex
  - Early pruning of unqualified solutions is a key factor to the overall performance

- Terminology
  - \( G_F(V_F, E_F) \) : feasible social graph of \( q \)
  - \( V_S \) : intermediate solution obtained so far
  - \( V_A \) : remaining vertices in \( V_F \) (\( V_A = V_F - V_S \))
ACCESS ORDERING – CONT.

- Step 1: construct the feasible graph $G_F$
- Step 2: explore $G_F$ to find the optimal solution
  - Initially $V_S$ includes only $q$
  - The remaining vertex set $V_A$ is $V_F - \{q\}$
  - At each iteration moving a vertex from $V_A$ to $V_S$
  - $V_S$ represents a feasible solution when $|V_S| = p$
Two conditions should be satisfied to exploit the acquaintance constraint:

- **Interior unfamiliarity**
  - The maximum number of unacquainted vertices that a vertex \( v \) in \( V_S \) has

- **Exterior expansibility**
  - The maximum number of vertices that \( V_S \) can be extended from
**INTERIOR UNFAMILIARITY**

- Interior unfamiliarity \((U)\) – definition:
  - \(U(V_S) = \max_{v \in V_S} |V_S - \{v\} - N_v|\)
    - \(N_v\) is a set of neighboring vertices of \(v\) in \(G_F\)
    - \(V_S - \{v\} - N_v\) is a set of non-neighboring vertices of \(v\) in \(V_S\)
  - In other words:
    - The maximum number of unacquainted vertices that a vertex \(v\) in \(V_S\) has
  - **Small value** of \(U\) indicates that every vertex \(v \in V_S\) has plenty of neighboring vertices in \(V_S\)
  - It is preferable first to include a well-connected vertex (low \(U\)) since it will make selections of other vertices in the later iterations easier
INTERIOR UNFAMILIARITY – CONT.

- Interior unfamiliarity ($U$) – condition:
  - $U(V_S \cup \{v\}) \leq k \left[ \frac{|V_S \cup \{v\}|}{p} \right]^\theta$
    - $\theta > 0$
    - $\frac{|V_S \cup \{v\}|}{p}$ is the portion attendees that have been considered
  - With $\theta = 0$ the inequality reaches its maximum ($k$)
  - Large $\theta$ allows choosing a vertex from $V_A$ that connects to more vertices in $V_S$
  - At early stages $\left( \frac{|V_S \cup \{v\}|}{p} \ll 1 \right)$ it is preferable to add well-connected vertices
  - Algorithm SGSelect reduces $\theta$ if there exists no vertex in $V_A$ that can satisfy the above condition
**Exterior Expansibility**

- Exterior expansibility \((A)\) – definition:
  
  \[
  A(V_S) = \min_{v \in V_S} \{|V_A \cap N_v| + (k - |V_S - \{v\} - N_v|)\}
  \]

  - \(V_A \cap N_v\) contains the neighboring vertices of \(v\) in \(V_A\)
  - \(V_S - \{v\} - N_v\) is a set of non-neighboring vertices of \(v\) in \(V_S\)

  - In other words:
    - The maximum number of vertices that \(V_S\) can be extended from

- We can select at most \(k - |V_S - \{v\} - N_v|\) extra non-neighboring vertices \(v\) in \(V_A\)
Exterior Expansibility — Cont.

- Exterior expansibility \((A)\) – condition:
  - \(A(V_S \cup \{v\}) \geq (p - |V_S \cup \{v\}|)\)
    - \(p - |V_S \cup \{v\}|\) is the number of attendees that have not been considered
  - If the inequality doesn’t hold, the new intermediate solution set obtained by adding \(v\) is not expansible
  - There is no degree of freedom in this case
EXAMPLE
SOCIAL GRAPH OF CASEY AFFLECK

$V_S = \{v_7\}, \ V_A = \{v_2, v_3, v_4, v_6, v_8\}$

- Select $v_2$ (since it has smallest social distance)
  - Calculate exterior expansibility $A(V_S \cup \{v_2\})$
    - $|V_A \cap N_{v_2}| + (k - |V_S - \{v_2\} - N_{v_2}|) = 2 + (1 - 0) = 3$
    - $|V_A \cap N_{v_7}| + (k - |V_S - \{v_7\} - N_{v_7}|) = 4 + (1 - 0) = 5$
    - Choose the smallest one -> $A(V_S \cup \{v_2\}) = 3$
  - Validate exterior expansibility condition
    - $p - |V_S \cup \{v_2\}| = 4 - 2 = 2$
  - Exterior expansibility holds for $v_2$

- $G_F$
- $q = v_7$
- $p = 4$
- $s = 1$
- $k = 1$
EXAMPLE – CONT.
SOCIAL GRAPH OF CASEY AFFLECK

\[ V_S = \{v_7\}, \quad V_A = \{v_2, v_3, v_4, v_6, v_8\} \]

- Calculate interior unfamiliarity \( U(V_S \cup \{v_2\}) \)
  - \( |V_S - \{v_2\} - N_{V_2}| = |\emptyset| = 0 \)
  - \( |V_S - \{v_7\} - N_{V_7}| = |\emptyset| = 0 \)
  - Choose the largest one -> \( U(V_S \cup \{v_2\}) = 0 \)
- Validate interior unfamiliarity condition
  - \( k \left[ \frac{|V_S \cup \{v_2\}|}{p} \right]^{\theta} = 1 \times \left( \frac{2}{4} \right)^2 = \frac{1}{4} \) (assuming that \( \theta = 2 \))
  - Interior unfamiliarity holds for \( v_2 \)

- \( G_F \)
- \( q = v_7 \)
- \( p = 4 \)
- \( s = 1 \)
- \( k = 1 \)
EXAMPLE — CONT.

SOCIAL GRAPH OF CASEY AFFLECK

- $V_S = \{v_7, v_2\}$, $V_A = \{v_3, v_4, v_6, v_8\}$
  - Select $v_3$
    - Calculate exterior expansibility $A(V_S \cup \{v_3\}) = 1$
    - Validate exterior expansibility condition $p - |V_S \cup \{v_3\}| = 1$
    - Exterior expansibility holds for $v_3$
    - Calculate interior unfamiliarity $U(V_S \cup \{v_3\}) = 1$
      - Validate interior unfamiliarity condition $k \left[ \frac{|V_S \cup \{v_3\}|}{p} \right]^{\theta} = \frac{9}{16}$
      - Interior unfamiliarity is violated for $v_3$
  - In this stage, no reduction of $\theta$ is done, since there are still more vertices in $V_A$
  - Put $v_3$ in parentheses and temporarily skip it
EXAMPLE — CONT.
SOCIAL GRAPH OF CASEY AFFLECK

- $V_S = \{v_7, v_2\}, V_A = \{(v_3), v_4, v_6, v_8\}$
  - Select $v_6$
    - Both conditions hold
- $V_S = \{v_7, v_2, v_6\}, V_A = \{v_3, v_4, v_8\}$
  - Select $v_3$
    - Interior unfamiliarity is violated for $v_3$
    - Put $v_3$ in parentheses and temporarily skip it
- $V_S = \{v_7, v_2, v_6\}, V_A = \{(v_3), v_4, v_8\}$
  - Select $v_8$
    - Exterior expansibility is violated for $v_8$
    - Remove $v_8$ from $V_A$
- $V_S = \{v_7, v_2, v_6\}, V_A = \{(v_3), v_4\}$
  - Select $v_4$
    - Both conditions hold – $p=4$
    - Obtain first feasible solution $V_S = \{v_7, v_2, v_6, v_4\}$
    - Total social distance = 67
EXAMPLE — CONT.
SOCIAL GRAPH OF CASEY AFFLECK

- $V_S = \{v_7, v_2, v_6, v_4\}$ (Total social distance = 67)
  - Is this the optimal solution?
  - Does it have the lowest social distance?

- By backtracking to steps where interior unfamiliarity constraint was violated and reducing $\theta$ to zero, it is possible to get a better solution
  - $V_S = \{v_7, v_2, v_3, v_4\}$ (Total social distance = 62)
SOCIAL-TEMPORAL GROUP QUERY (STGQ)

- Extension of SGQ by exploring the temporal dimension (m)
- STGQ is more complex than SGQ
- Intuitive approach
  - Find SGQ solution for each individual activity period and then select the one with the minimum total social distance
  - Computationally expensive
SOCIAL-TEMPORAL GROUP QUERY (STGQ) ALGORITHM DESIGN

- Proposed algorithm: STGSelect
- STGSelect constructs the $V_S$ with vertices which have more available times lots in common
- Pivot time slot is chosen
  - It is better to choose time slot which has plenty vertices
  - It is mandatory to choose a time slot which has more than $p$ vertices
TEMPORAL EXTENSIBILITY

Temporal extensibility (X) – definition:

- **X(V_S) = |T_s| − m**
  - $T_s$ set of the **consecutive** time slots available for all vertices in $V_S$
  - In other words:
    - How many extra mutual time slots are available

Temporal extensibility (X) – condition:

- **X(V_S ∪ {v}) ≥ (m − 1) \left(\frac{p−|V_S∪\{v}\}|}{p}\right)^\varphi**
  - $\varphi ≥ 1$
  - $\frac{p−|V_S∪\{v}\}|}{p}$ is the portion of attendees that have **not** been considered
Choose $t_3$ as pivot time slot

$V_S = \{v_7\}, V_A = \{v_2, v_3, v_4, v_6, v_8\}$

- Select $v_2$ (since it has smallest social distance)
- Both exterior expansibility and interior unfamiliarity conditions hold
- Calculate temporal extensibility $X(V_S \cup \{v_2\})$
  - $T_s = \{t_1, t_2, t_3, t_4, t_5\} \rightarrow |T_s| = 5 \rightarrow X(V_S \cup \{v_2\}) = 5 - 3 = 2$
- Validate temporal extensibility condition
  - $(m - 1) \left[ \frac{p - |V_S \cup \{v\}|}{p} \right]^q = 2 \times \left( \frac{2}{4} \right)^2 = \frac{1}{2} \rightarrow \text{condition holds}$
EXPERIMENTAL RESULTS

- Experiment setup – need to find G
  - Scheduling
    - 194 people from various communities (schools, government, business, industry)
    - Google calendar is utilized for scheduling
  - Social graph
    - Synthetic dataset with 12800 people is generated from a social network website
    - The schedule of each person in each day is randomly assigned from the above 194-people real dataset
  - Existing algorithm - PCArrange
    - Imitating the behavior of manual coordination via phone calls, where the initiator q sequentially invites close friends first then finds out the common available time slots
    - Doesn’t include acquaintance constraint
      - They obtain k from each activity manually
RESULT ANALYSIS

Running time with different $p$

Execution time [ns]

SGSelect
Baseline

Running time with different $k$

Execution time [ns]

SGSelect
Baseline

Running time with different $s$

Execution time [ns]

SGSelect
Baseline

Running time with different network sizes

Execution time [ns]

SGSelect
Baseline
RESULT ANALYSIS – CONT.

Relation of $p$ and $k$

Relation of $p$ and the total social distance
ARE WE ALWAYS GETTING WHAT WE WANT?

- Time constraint can sometimes lead to undesirable solution. E.g. Arrange a party at 09:00 – 14:00
- Possible solution:
  - Define also time interval, e.g. m=3 & 18:00<T<02:00
ARE WE ALWAYS GETTING WHAT WE WANT?

- Event may be too diverse for the initiator
  - p=8
  - s=2
  - k=6

- Possible solution
  - Give more “weight” to s=1 – include more direct friends of the initiator
ARE WE ALWAYS GETTING WHAT WE WANT?

- Data is not complete, currently there are no social graphs which includes also timetable
  - Reliability – “inactive” users → wrong social distance calculation
  - Social graph may not contain actual friends
CONCLUSIONS

- **Novel system** for automatic activity planning based on social and temporal relationship
- **Two algorithms proposed:**
  - SGSelect
  - STGSelect
- **Algorithm proposed, significantly outperforms** the existing algorithms
- **Research result can be adopted in social network websites**
THANK YOU