Temporal Social Network:

Group Query Processing

Xiaoying Chen, Chong Zhang, Yanli Hu,
Bin Ge, Weidong Xiao

Contact Us
Science and Technology on Information Systems Engineering Laboratory
National University of Defense Technology, Changsha 410073, P.R.China
chenxiaoying1991@yahoo.com

Contents

Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

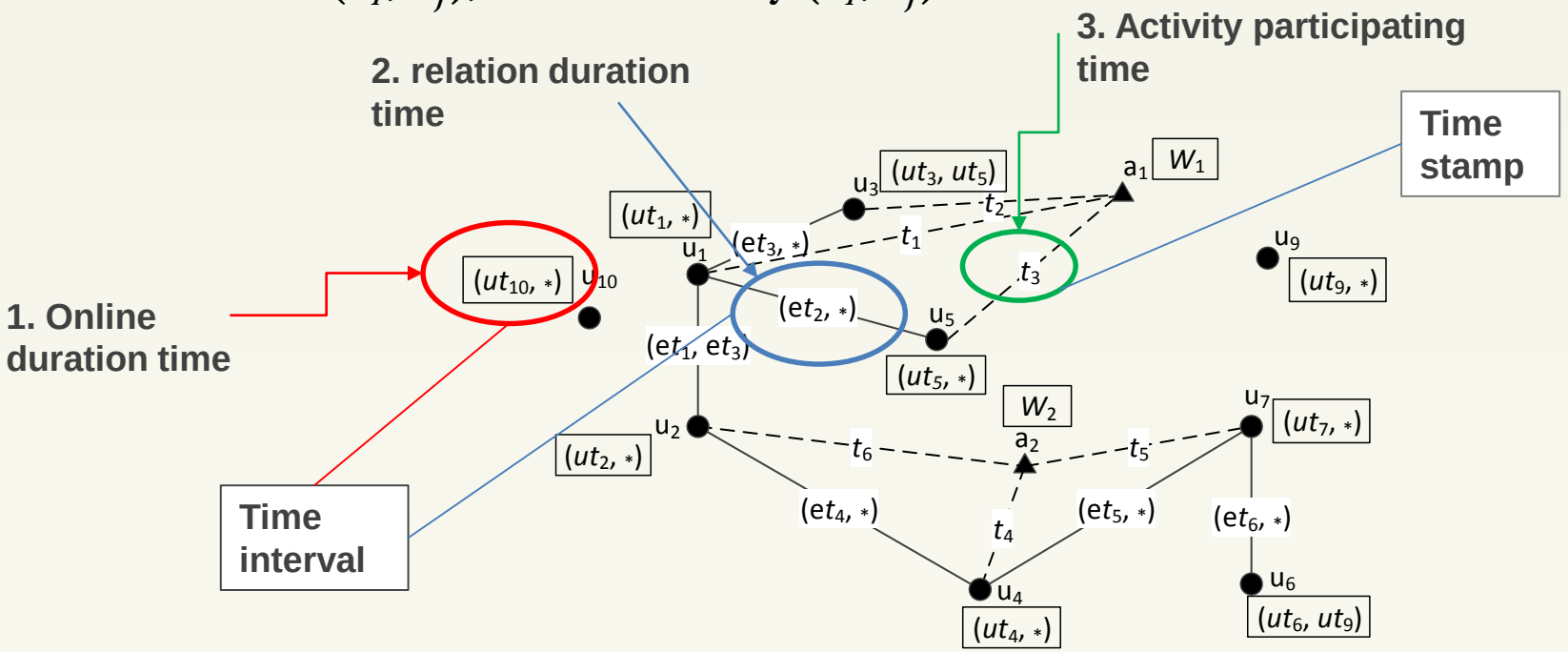
Problem definition

1. Temporal Social Network (TSN)

Given an undirected graph $G=(V, E)$, $V=U \cup A$

U: users A: activities $< aid, W_a >$

E: user-to-user (u_i, u_j); user-to-activity (u_i, a_j)



Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Problem definition

2. Temporal social network (TSN) query

This work is a follow-up to our previous work : "Temporal Social Network: Storage, Indexing and Query Processing". In this previous work, we proposed three TSN query: Friends of Interesting Activities (FIA), Users of Time Filter (UTF), Group of Users with Relationship Duration (GURD) query.

group analytics can provides insights into common interests with various relationships.

3. Temporal group query (TGQ)

✉ Example: we aim to find a set of groups, each group with all member participating in the activity labeled with 'pizza' during last two weeks, and average on-line duration is not less than 24 hours during last two weeks. The result of this query should be appropriately in the form: $\{< g_1=\{u_1, u_2, \ldots\}, t_1, t_2>, < \ldots >, \ldots\}$, where g_1 is a satisfying group valid during $[t_1, t_2]$.

3. Connected graph

1. Being interested in "Pizza"

2.Being active in this community

Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Problem definition

GURD query (m, t_d, W_q) :

- aims to find groups of users, in each group of which users form a connected graph, and the number of users is m
- for each user in the group, at least one activity he/she participated in overlaps W_q
- average relationship duration (ARD) is not less than t_d
- Average relationship duration (ARD) is defined as: $ARD = \frac{\sum_{i \neq j}^{n} d(i,j)}{n(n-1)}$

TGQ query (W,[t_s, t_e], t_{ol}) :

- aims to find a set of groups, in which, each group is a connected graph
- all members have participated in the activities with keywords in W during period [t_s, t_e]
- average on-line duration (AOD) (during [t_s, t_e]) of the members is not less than t_{ol}
- Average on-line duration (AOD) is defined as: $AOD = \frac{\sum_{i=1}^{n} onlineduration(u_i)}{n}$
- The result should be in the form:

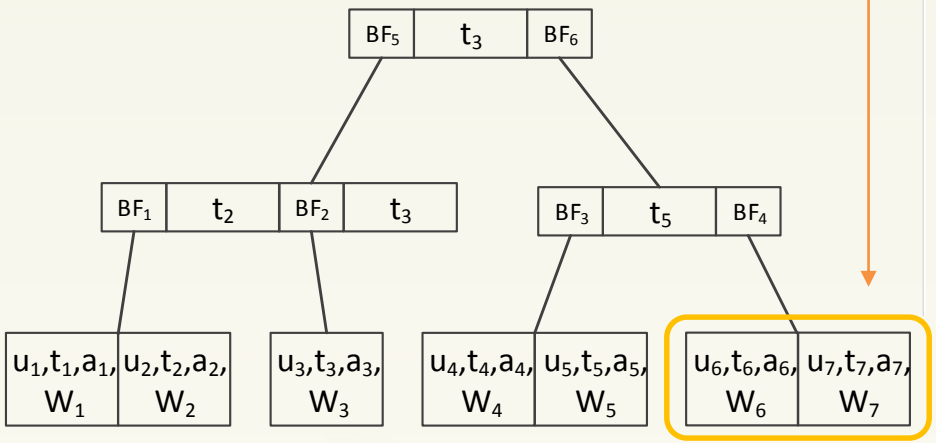
{< g_1={u_1, u_2, . . . }, t_1, t_2>, g_2=< . . . >, . . .}

Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Indexes design

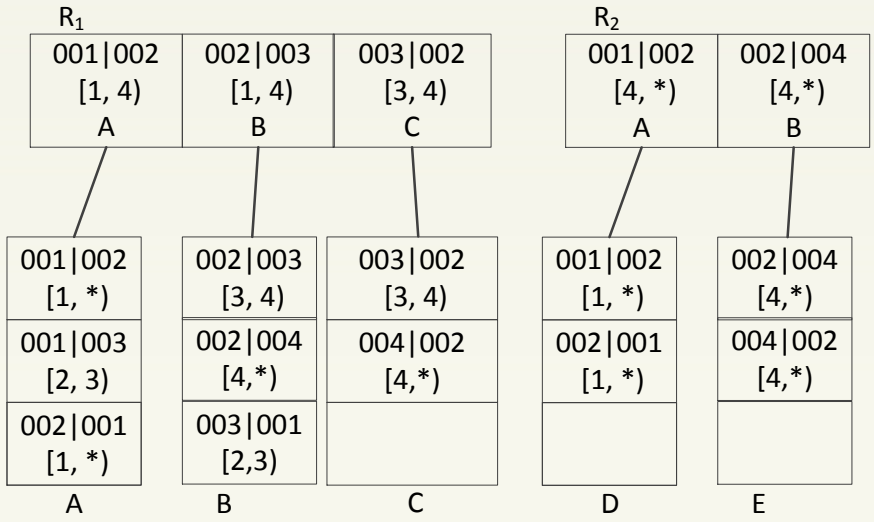
TA-tree (Temporal Activity tree)

- B$^+$-tree injected with Bloom Filter
- data item (entry in leaf node) to be indexed is in the form $<u_i, t_p, a_k, W_{ak}>$
- B$^+$-tree is built according to the key t_p
- keyword sets of all entries in each leaf node constitute a Bloom Filter BF, and BF serves as filter with pointers in internal nodes



TF-tree (Temporal Friendship tree)

- a kind of MVB-tree
- data item to be indexed is in the form $<u_i | u_j, [t_f, t_u]>$

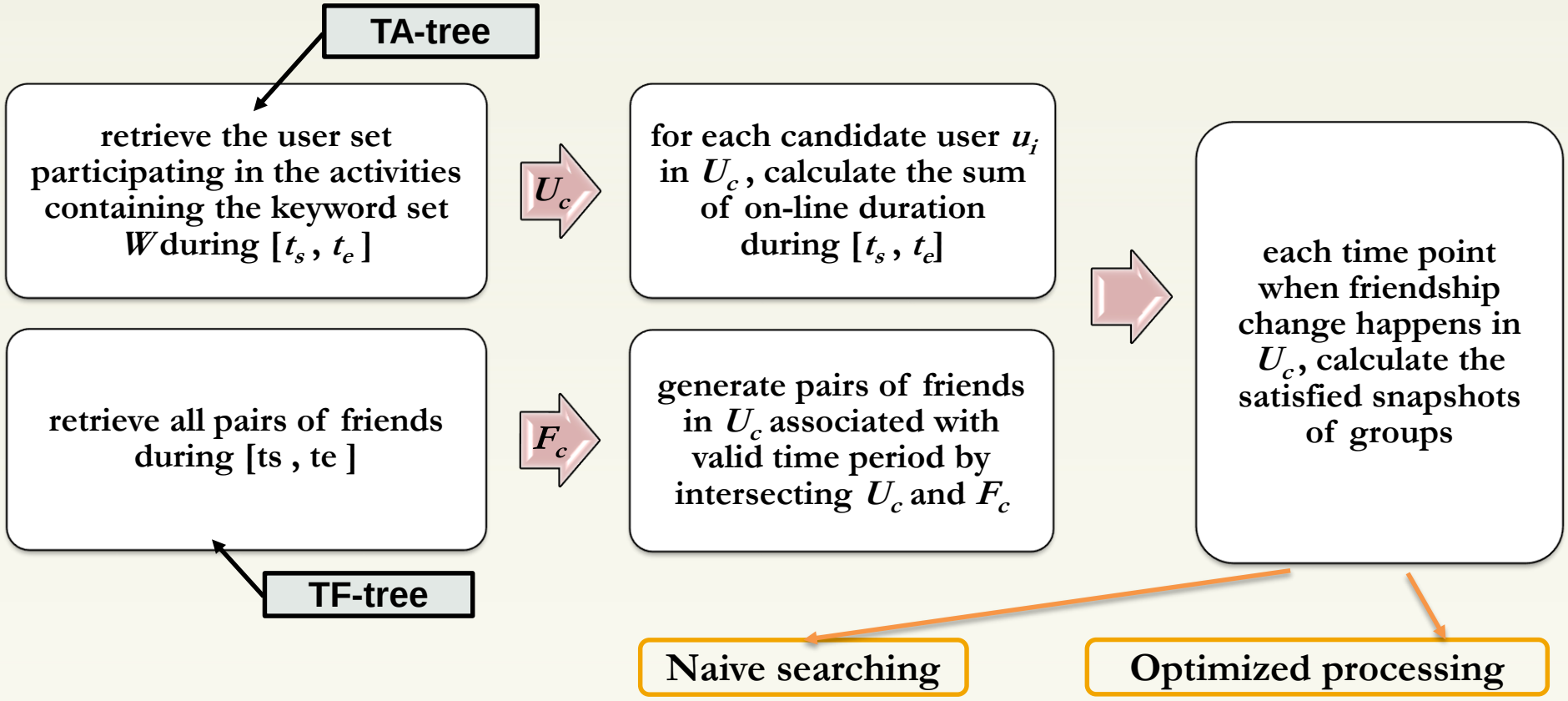


Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Query processing

Basic workflow:

TGQ query (W,[t_s, t_e], t_{ol})

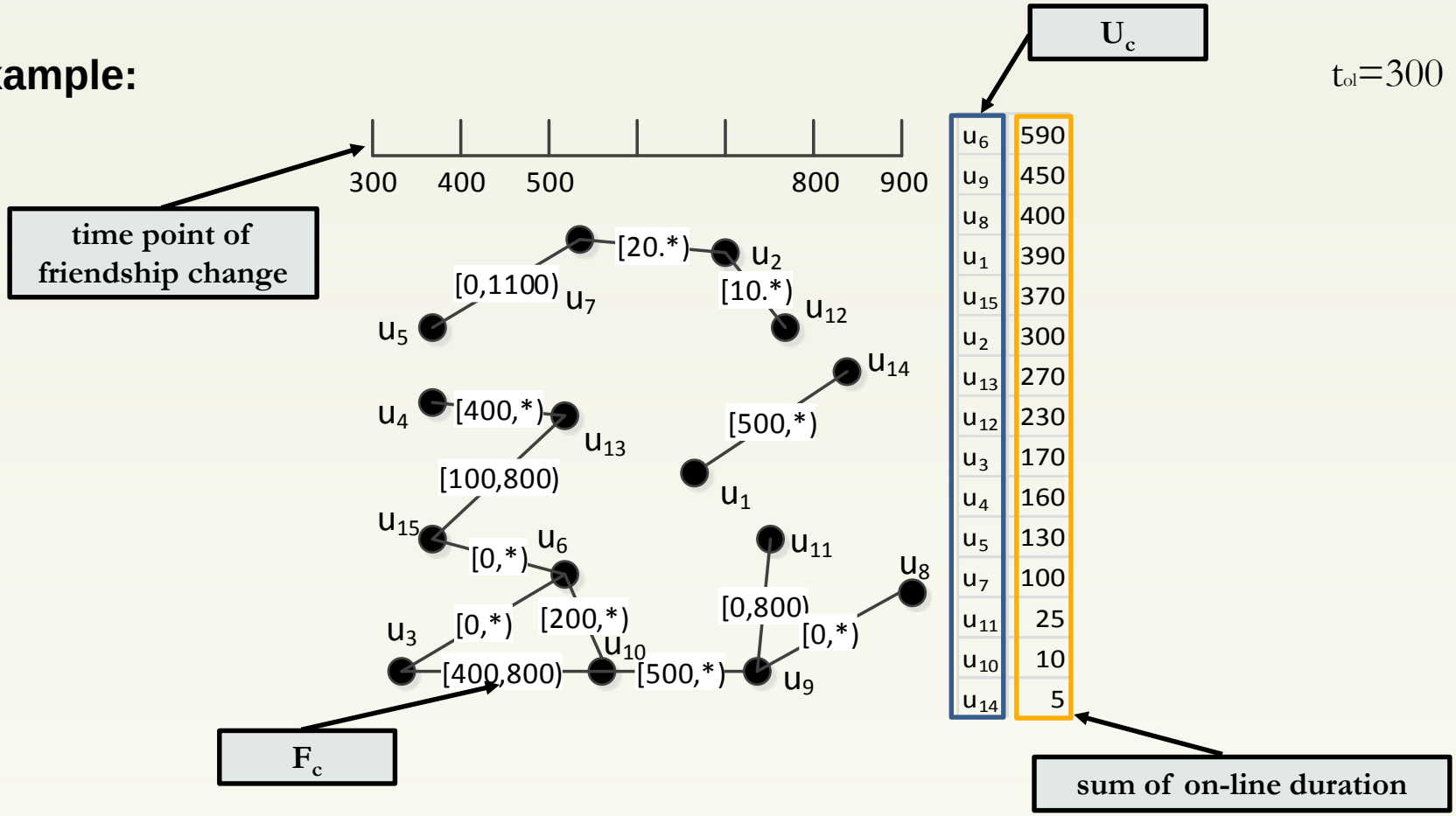


Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Query processing

Example:

$t_{ol}=300$

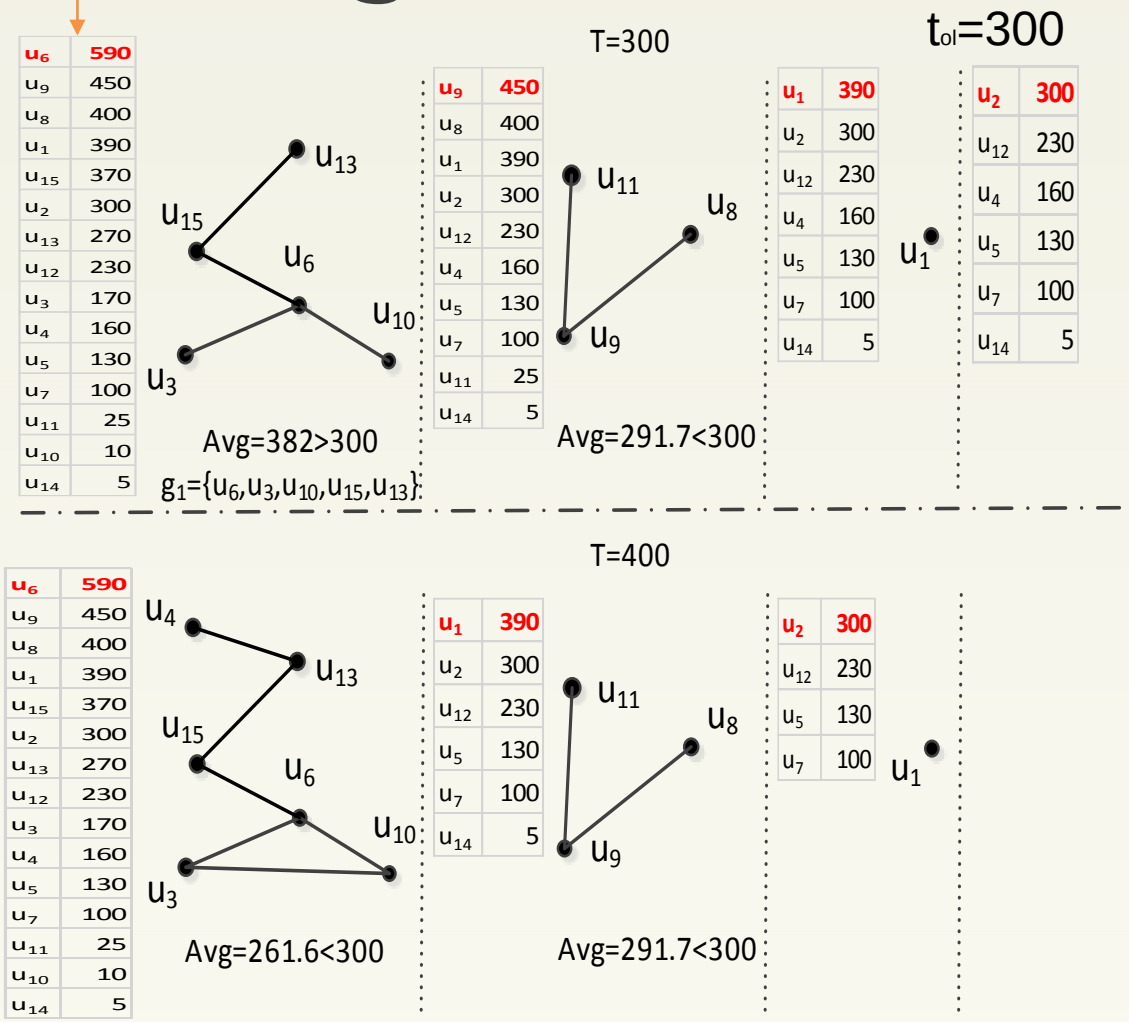


Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Query processing

1. Naive Searching

1. use a max-queue ~~storing~~ users in U_c sorted by on-line duration
2. for each time point, the top element of the max- queue is popped and checked whether on-line duration is larger than t_{ol} , if so, we find its connected graph and check whether the AOD is not less than t_{ol} , then, all the user in this graph is deleted from the queue; otherwise, if the top element is not larger than t_{ol} ,we change to the next time point.



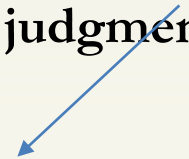
Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Query processing

2.Optimized Processing

❶ it is not necessary to construct connected graph at each time point.

❷ For each time point, according to the changes of relationship, we update the corresponding graph and make judgment whether it is a satisfied result.



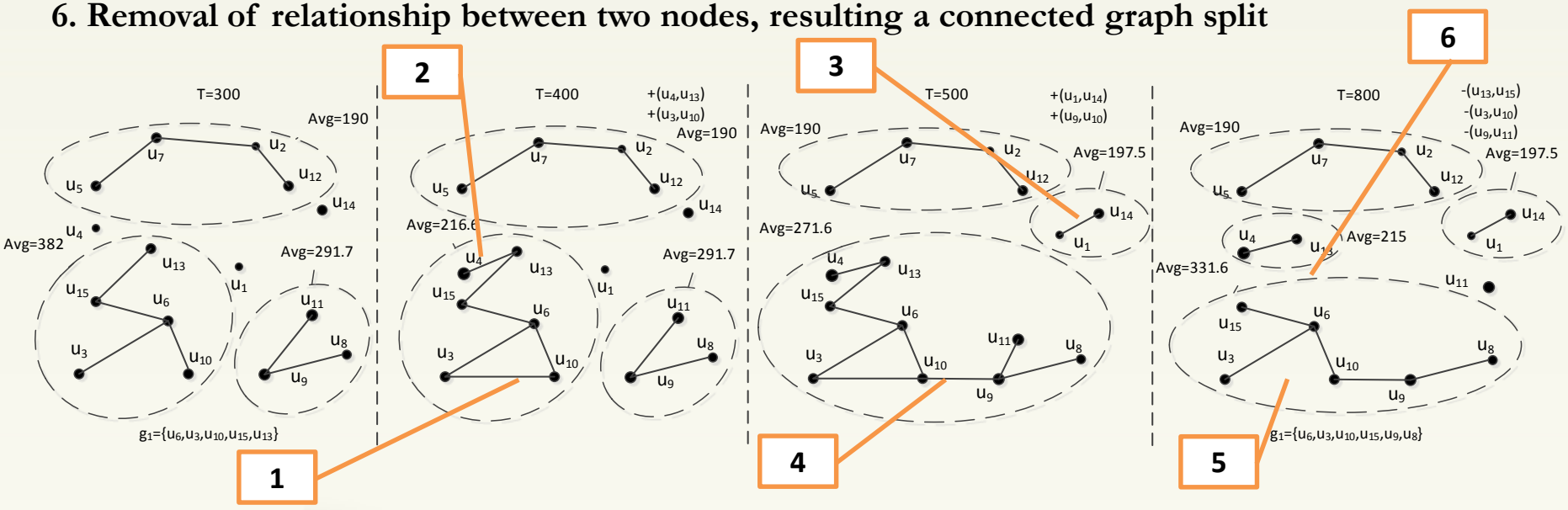
1. Addition of relationship between nodes both in a connected graph
2. Addition of relationship between a node in a connected graph and an isolating node
3. Addition of relationship between two isolating nodes
4. Addition of relationship between one node in a connected graph and the other node in another connected graph.
5. Removal of relationship between two nodes, resulting no split in a connected graph
6. Removal of relationship between two nodes, resulting a connected graph split

Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Query processing

2.Optimized Processing

1. Addition of relationship between nodes in one connected graph
2. Addition of relationship between a node in a connected graph and an isolating node
3. Addition of relationship between two isolating nodes
4. Addition of relationship between one node in a connected graph and the other node in another connected graph.
5. Removal of relationship between two nodes, resulting no split in a connected graph
6. Removal of relationship between two nodes, resulting a connected graph split



Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Experiment

1. *Dataset description*

We have to synthetically generate dataset based on some real datasets which contain partial temporal attributes.

Dataset	Record number	Data size
User	3,223,589	287M
Relation	9,375,374	5.53G
Login	3,223,589	2.45G
Activity	6,980,465	2.28G

http://konect.uni-koblenz.de/downloads/tsv/youtube-u-growth.tar.bz2

1. Data about food theme on GCZX server
2. 25 classes of Amazon commodities
3. 1759 review on the hotel and some web crawling data, etc.

2. *Procedure of experiment*

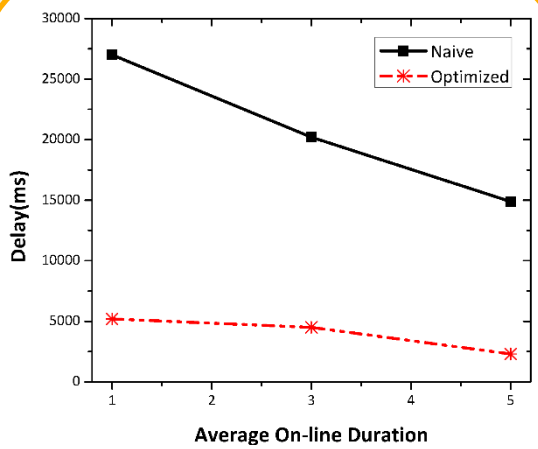
To make comparison, the naive searching approach is used as a baseline. We vary the query parameters and at each testing point, 10 queries are issued to collect the average results. The experiments are conducted on a DELL server with Intel(R) Xeon(R) 2.40GHz processor, 8GB memory and 500GB disk.

queries	parameters	domain	default
TGQ	t_{ol}	1 – 5	3
	W_q `s cardinality	1 - 5	3
	$length_{[ts,te]}$ /temporal extent	0.1% - 3%	1%

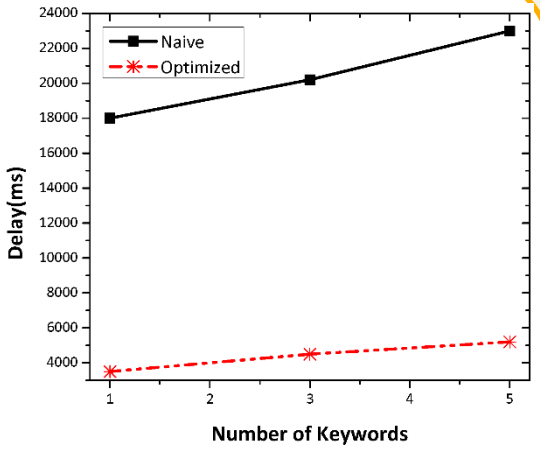
Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Experiment

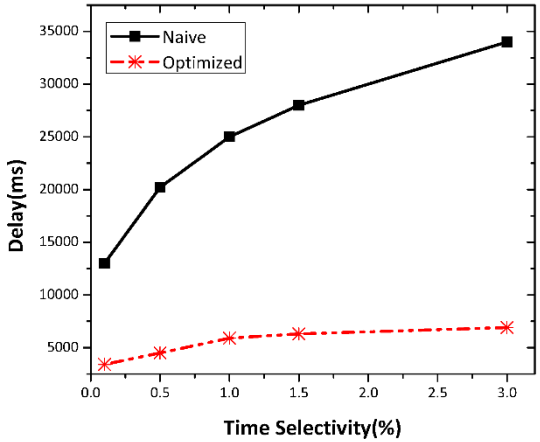
3. Result



- ✓ A larger t_{ol} will terminate loop earlier, thus the processing delay is reduced.
- ✓ The optimized approach utilizes the updates to the graphs, which reduces the simple repeated group forming procedure.



- ✓ The response time also increases with the number of keywords. This can be explained that a larger number of keywords would involve more tree nodes in the TA-tree to be traversed.



- ✓ A larger value causes more candidates to be inspected, thus the response time increases correspondingly.
- ✓ A series of update operations is more efficient than exhausted method.

Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

Conclusion

With applications continuously developing, more and more temporal social network group queries will be paid attention. In this paper, we focus on temporal analytics on social group query, and argue Temporal Group Query (TGQ) is useful and applicable. To efficiently address the query, we design two indexing structures to accelerate the query processing, and two processing algorithms, one is simple, the other is optimized, to accomplish the processing. We conduct experiments on real-synthetic hybrid dataset, and the results show our methods are capable and optimized method is efficient. In the future, we would like to study on social group query with geographic attributes.

Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com

THANK YOU

Contact Us
National University of Defense Technology,
chenxiaoying1991@yahoo.com