

Personal Social Screen

A Dynamic Privacy Assignment System for Social Sharing in Complex Social Object Networks

Lei Li, Tong Sun and Tao Li

Florida International University
& Xerox Corporation

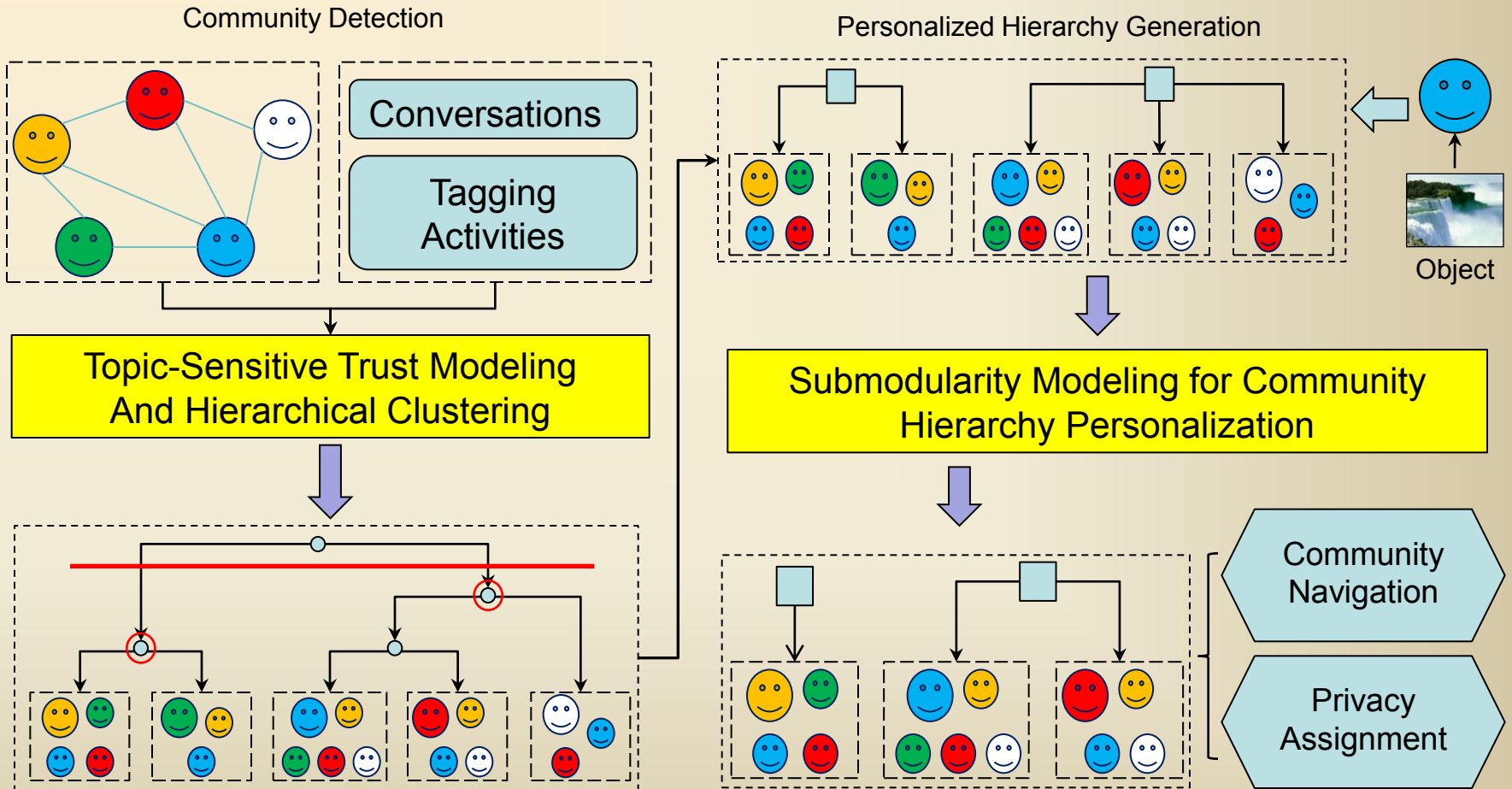
Motivation and Challenges

- *Motivation*
 - Privacy within social networking sites is often not expected or is undefined, which might render potential **privacy risks**.
 - Sometimes people would like to share interesting posts with a specific group of his/her friends who have **similar hobbies or interest**.
- *Challenges*
 - How to find people with similar hobbies?
 - How trust to people that you are going to share with?

Related Work

- *Privacy issues are **unconcerned** by online users.*
 - Gross and Acquisti, Information revelation and privacy in online social networks, 2005.
 - Dwyer et al., Trust and privacy concern within social networking sites: A comparison of facebook and myspace, 2007.
- *Google+ uses the concept of “**social circle**” to allow people explicitly manage their friend lists into friend circles.*
 - **Pros:**
 - People have choice on which circles he/she would like to share.
 - **Cons:**
 - Circles are “**static**”, i.e., managing the contact list.
 - Fail to address the issues of **selective sharing** in user’s **extended** social network.

Framework Overview



Community Hierarchy Generation

- *Atomic Community Generation:*
 - Choose online users' friendships as the base.
 - Calculate their trustiness as

$$t_{ij} = \alpha \cdot \frac{|\hat{c}_i| + |\hat{c}_j|}{|c_i| + |c_j|} + \beta \cdot \frac{|o_i \cap o_j|}{|o_i \cup o_j|} + \gamma \cdot \text{sim}(p_i, p_j)$$

- Dynamically select a subset of friends based on t_{ij} .
- *Hierarchical clustering on atomic communities:*

$$\mathcal{T}_{G_k, G_l} = \sum_{u_i \in G_k, u_j \in G_l} t_{ij}$$

- *Decide the best layer using Dunn's Index:*

$$D = \frac{\min_{1 \leq i < j \leq r} d(c_i, c_j)}{\max_{1 \leq k \leq r} d'_k}$$

Personalized Hierarchy Generation

★ **Submodular Function:** Let E be a finite set and f be a real valued non-decreasing function defined on the subsets of E that satisfies

$$f(T \cup \{\zeta\}) - f(T) \leq f(S \cup \{\zeta\}) - f(S)$$

where $S \subseteq T$, S and T are two subsets of E , and $\zeta \in E \setminus T$.

★ **Budgeted Maximum Coverage:** Given E , where each element is associated with an influence and a cost defined over a domain of these elements and a budget B , the goal is to find out a subset of E which has the largest possible influence while the total cost does not exceed B .

NP-Hard!!!

Greedy

$(1 - 1/e)$ -approximation.
(Khuller et al., 1999)

• Personalized Hierarchy Generation

• Why Submodularity?



Landscape

Canon D50

Hometown

Submodularity!!!

- A picture might belong to different categories. How to make the recommended community list more **relevant** to the picture?
- A user's privacy requirement might be regressive over time. How to capture such **privacy regression**?

$$f(S) = \frac{1}{|S|} \sum_{c_1 \in S} Trust(u, c_1) + Sim(o, S)$$



Trust!



Relevance!

Experimental Evaluation

- Real-world Data Set

- Gather users, photo, comments, tags, groups etc. from Flickr by specifying a group of seed users.
- Preprocessing
 - ✓ Remove comments that are not English.
 - ✓ Remove users without frequent activities, e.g., posting photos, commenting or tagging.
 - ✓ Remove user groups whose group members are less than 5.

- Statistics

User	4,219
Photo	2,292
Comment	29,909
Tag Assignment	17,588
User Relation	7,550
Group	1,071

Experimental Evaluation (Cont.)

- Evaluation Metrics

- *Normalized mutual information (NMI):*

$$NMI = \frac{-2 \sum_{i,j} N_{ij} \cdot \log\left(\frac{N_{ij} \cdot N}{N_{i.} \cdot N_{.j}}\right)}{\sum_i N_{i.} \cdot \log\left(\frac{N_{i.}}{N}\right) + \sum_j N_{.j} \cdot \log\left(\frac{N_{.j}}{N}\right)}$$

N – the confusion matrix of the result and the ground truth;

N_{ij} – the number of nodes in both cluster X_i and Y_j ;

$N_{i.}$ is the sum over row i of N and $N_{.j}$ is the sum over column j of N .

- Jaccard Index (JI):

$$JI(C_1, C_2) = \frac{n_{11}}{n_{11} + n_{10} + n_{01}}$$

n_{11} is the number of sample pairs in the same cluster for both C_1 and C_2 ;

n_{10} and n_{01} are the number of sample pairs belonging to the same cluster in one solution, but not in the other.

Experimental Evaluation (Cont.)

- Community Detection Evaluation

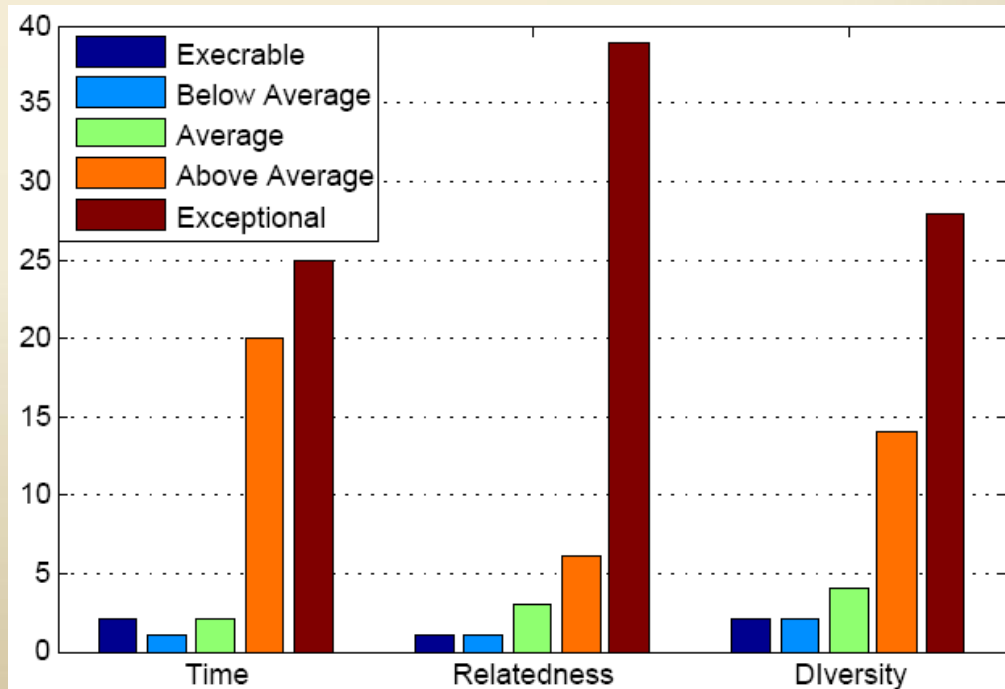
- We compare the NMI and JI of different community detection approaches to demonstrate the efficacy of our proposed method:

Approaches	$\mu = 10$		$\mu = 20$		$\mu = 30$		$\mu = 50$	
	NMI	JI	NMI	JI	NMI	JI	NMI	JI
MODULE	0.5379	0.5230	0.5628	0.5804	0.6207	0.6332	0.7047	0.6938
SHRINK	0.6137	0.5821	0.6277	0.6035	0.6458	0.6217	0.6928	0.6736
CONGA	0.6534	0.6412	0.6803	0.6219	0.6980	0.6434	0.7121	0.7295
EAGLE	0.6607	0.6329	0.6947	0.6608	0.7233	0.6954	0.7332	0.7001
PerCial	0.7235	0.7120	0.7433	0.7395	0.7588	0.7602	0.7625	0.7743

- The modularity-based methods perform poorly when small communities are dominant in the dataset – resolution limit;
- The methods that aim to detect overlapping communities perform better than the hard partition-based ones on our Flickr dataset;
- PerCial is capable to deal with communities with different cardinalities since we not only consider the social graph structure, but also take the topic-sensitive information and activities into account.

Experimental Evaluation (Cont.)

- Community Selection Evaluation (User Study)
 - Hire volunteers to experience *PerCial*
 - Define different experience indices
 - The response time of recommending communities for privacy assignment;
 - The relatedness of the recommended communities to their interest;
 - The diversity of the recommended community list.



Concluding Remarks

- Formalize the *topic-sensitive* trust relations among users by analyzing available resources in multi-modal social networks.
- Propose to *loosely separate* the entire social network and organize it as a *two-level community hierarchy*.
- Propose to recommend communities to users for privacy preference assignment by modeling the community selection problem as a *budgeted maximum coverage* problem.

Thank you!

Questions?

问题?

問題?

Ερώτηση?

Domanda?

문제?

Pregunta?