Influence Maximization-based Event Organization on Social Networks

Cheng-Te Li

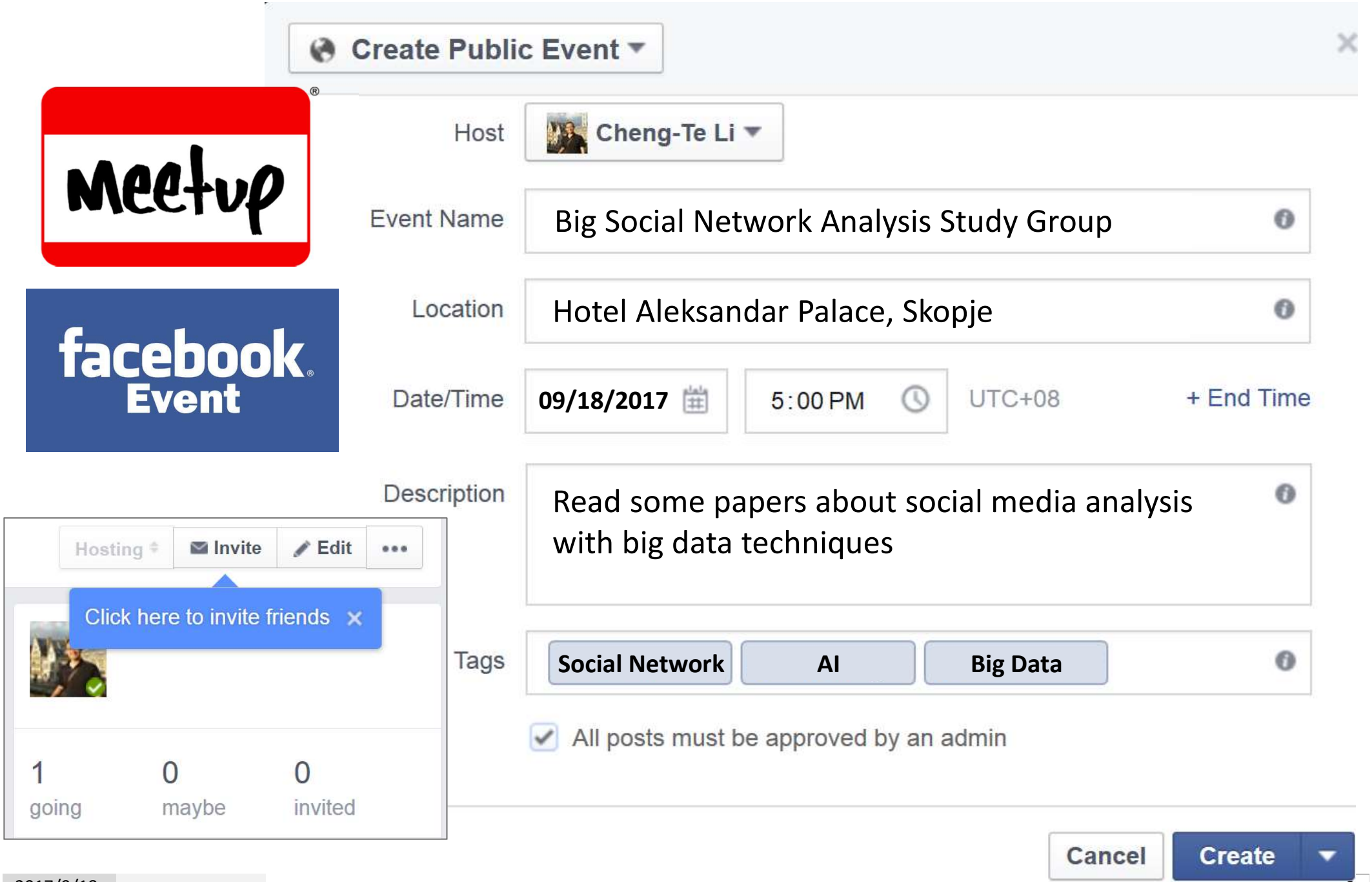
National Cheng Kung University, Taiwan

chengte@mail.ncku.edu.tw

Social Event Organization

- You may want to plan parties or activities that the participated people will enjoy with one another

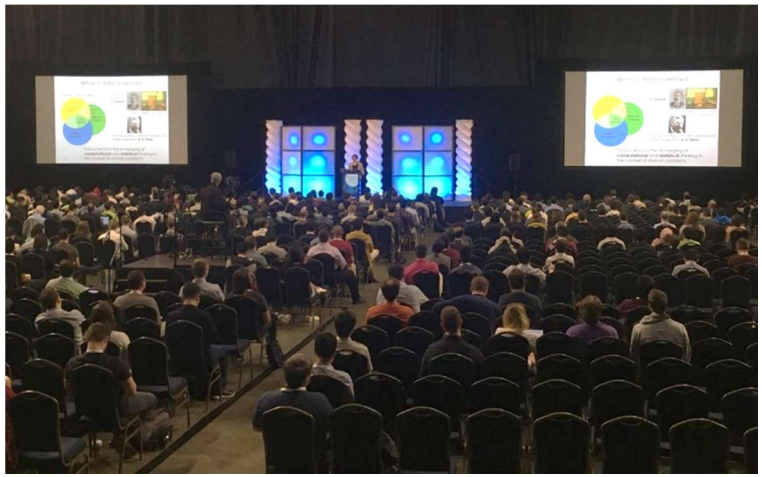
Event-based Social Services



Influential Event Organization

- Find a set of individuals who are interested in organize an event with some theme
  - Have better social interactions
  - Attract more people to participate the event

Technical conference



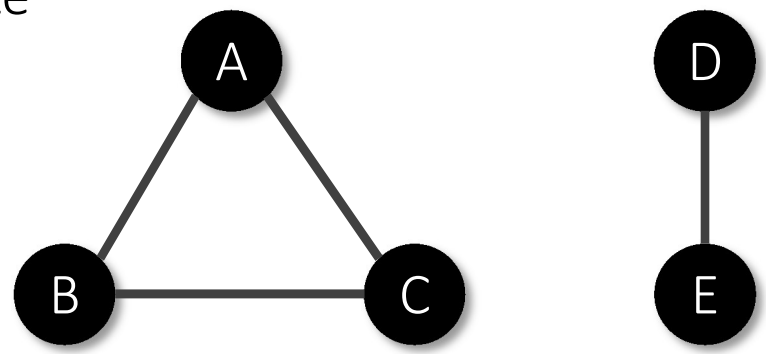
Fundraising for victims

Anti-nuclear campaign

Team Formation

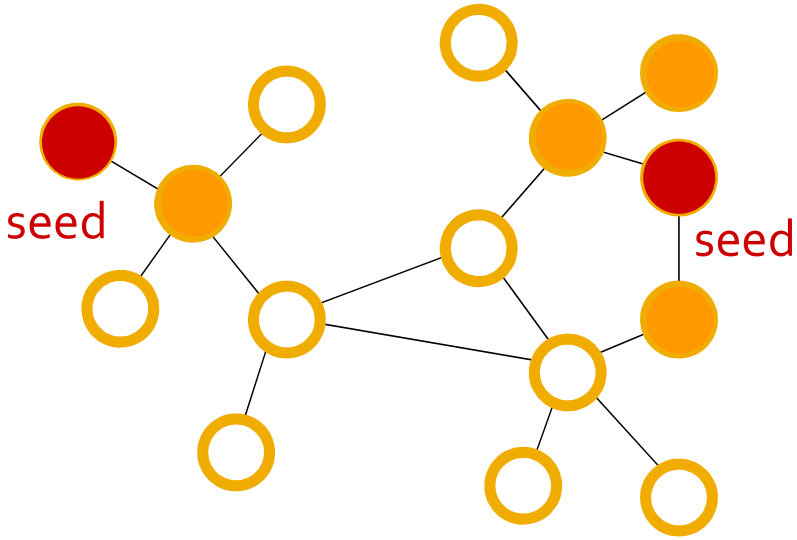
Expert	Skill
A	ML
B	Py
C	Art, Web
D	Art
E	Art, Web, Py

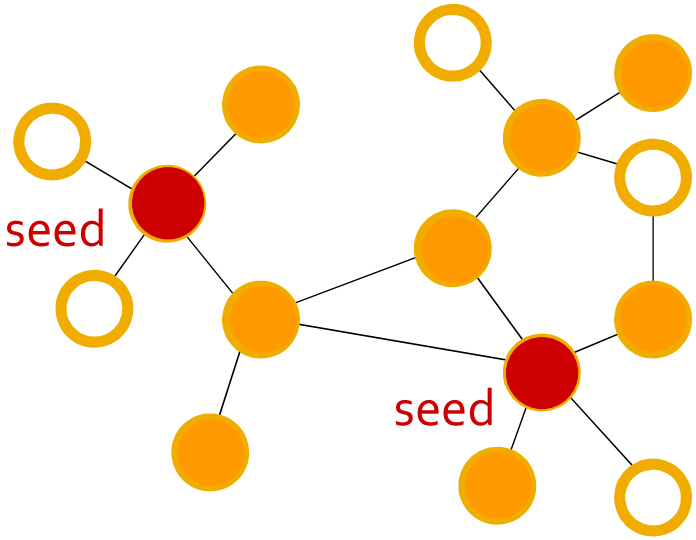
- Task = {ML, Art, Web, Python}
- Without the social network
  - Result 1 = {A, B, C} An effective team
  - Result 2 = {A, E} Hard to communicate
- With the social network



- Given a task consisting of a set of skills, how to find a set of individuals that
  - Cover the required skills
  - Effectively communicate with each other

Social Influence Maximization





- Given a limit budget for initial advertising
- Identify a small set of influential customers (as **seeds**)
- such that by convincing them to adopt the product
- And finally trigger a larger cascade of influence

Influential Team Formation (ITF)

- Given
  - A social network (each node has a set of labels)
  - A set L of required labels depicting the event
  - The size k of the team
- Goal: Find a set S of nodes such that
  - L is covered by S
  - The Influence-Cost Ratio (ICR) of S is maximized

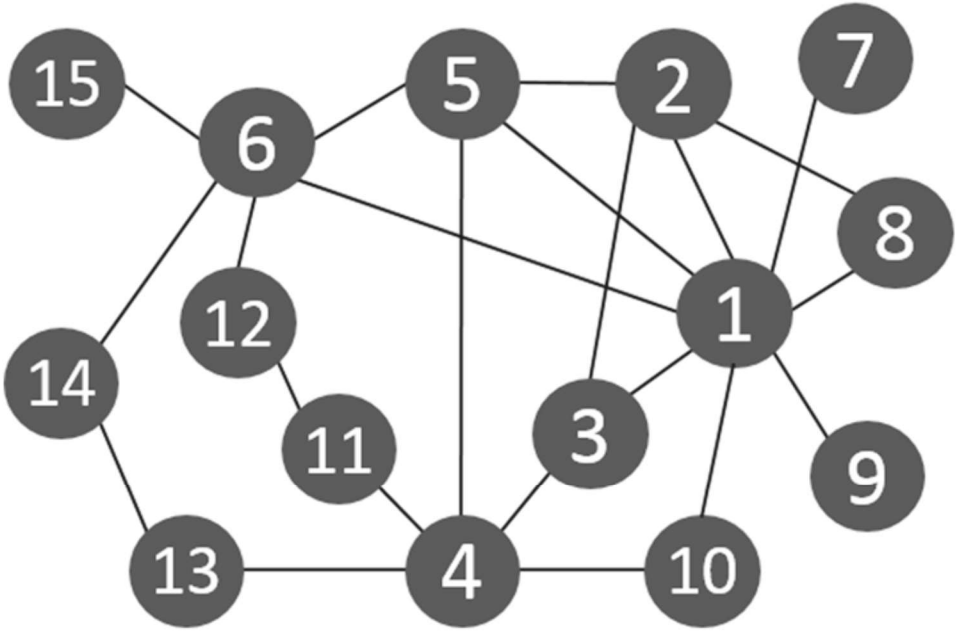
$$ICR(S) = \frac{\sigma(S)}{c(S)}$$

$\sigma(S)$ → **Influence Spread:** Expected number of nodes activated by S

$c(S)$ → **Cost:** Sum of all-pair shortest path length between nodes in S

Maximize $\sigma(S)$ and Minimize $c(S)$ = IM + TF

An ICR Example



Node	Label-Set	Activated-Set
1	{ *a*, *e* }	{ 3, 5, 7, 9 }
2	{ *b*, *c*, *e* }	{ 1, 3, 5, 8 }
3	{ *a*, *b*, *c*, *d* }	{ 1, 2 }
4	{ }	{ 3, 5, 10, 11 }
5	{ *c*, *e* }	{ 2, 4, 6 }
6	{ *b*, *d* }	{ 5, 14, 15 }

L = {a,b,c,d,e}, k = 3	The selected set S	ICR Score
Team Formation (TF)	S_{TF} = {1, 2, 3}	ICR(S_{TF}) = 7/3
Influence Maximization (IM)	S_{IM} = {1, 4, 6}	ICR(S_{IM}) = 10/5
Influential Team Formation (ITF)	S_{ITF} = {1, 5, 6}	ICR(S_{ITF}) = 10/3

Solution 1: ICR-Greedy

- Based on the idea of Greedy influence maximization, we can directly maximize ICR in a greedy manner
- Repeatedly select the next node, which has the maximum marginal gain of ICR, into final set S

For each of k iterations:
Add a node $u \in S \backslash V^L$ to set S that maximizes: $ICR(S \cup \{u\}) - ICR(S)$

$$V^L = \{v | v \in V \ and \ L_v \cap L \neq \emptyset\}$$

The node set containing at least one of required labels

Solution 2: Mixing Influence and Cost Greedy

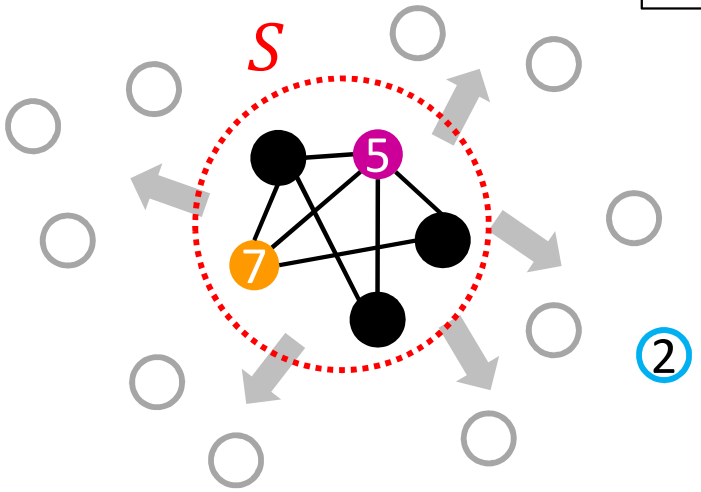
(M-Greedy)

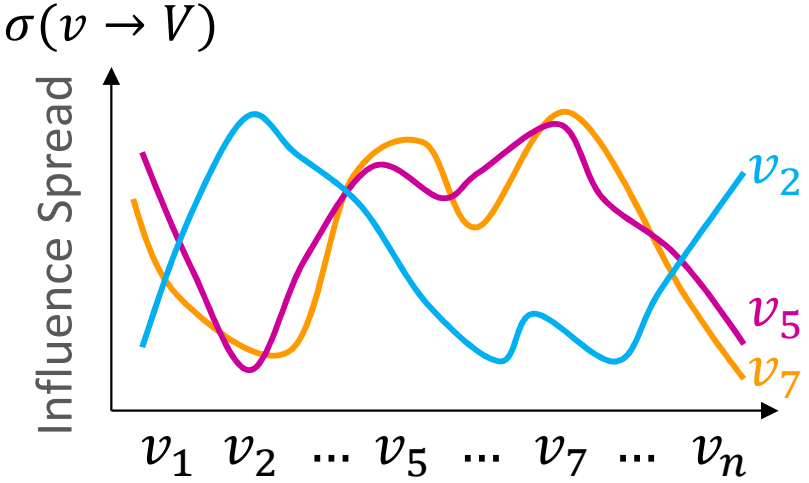
- Directly maximize ICR may destroy the connectivity among team members (i.e., make them disconnected)
- Divide the maximization of $ICR(S) = \frac{\sigma(S)}{c(S)}$ into interweavingly maximize $\sigma(S)$ and minimize $c(S)$
  - Balance the trade-off between $\sigma(S)$ and $c(S)$
- Similar to ICR-Greedy, but
  - In "$i\%2 = 0$" rounds: $\arg\mathbf{max}_{v \in V^L \setminus S} \sigma(S \cup \{v\}) - \sigma(S)$
  - In "$i\%2 = 1$" rounds: $\arg\mathbf{min}_{v \in V^L \setminus S} c(S \cup \{v\}) - c(S)$

(SimIS)

Solution 3: Similar Influence Search Heuristic

- ICR-Greedy and M-Greedy are inefficient to compute $\sigma(S)$
- SimIS Heuristic
  - Use Group-PageRank (GPR) to efficiently approximate the vector $\sigma(S \to V)$: the influence spread from S to all nodes
  - A node set S with lower cost $c(S)$ means its members:
    → Tend to be close to each other in the graph
    → Generate similar distributions of influence to other nodes
  - Select the next node by: $\underset{v \in V^L \setminus S}{\arg\min} \| GPR(\{v\} \to V) - GPR(S \to V) \|_F^2$



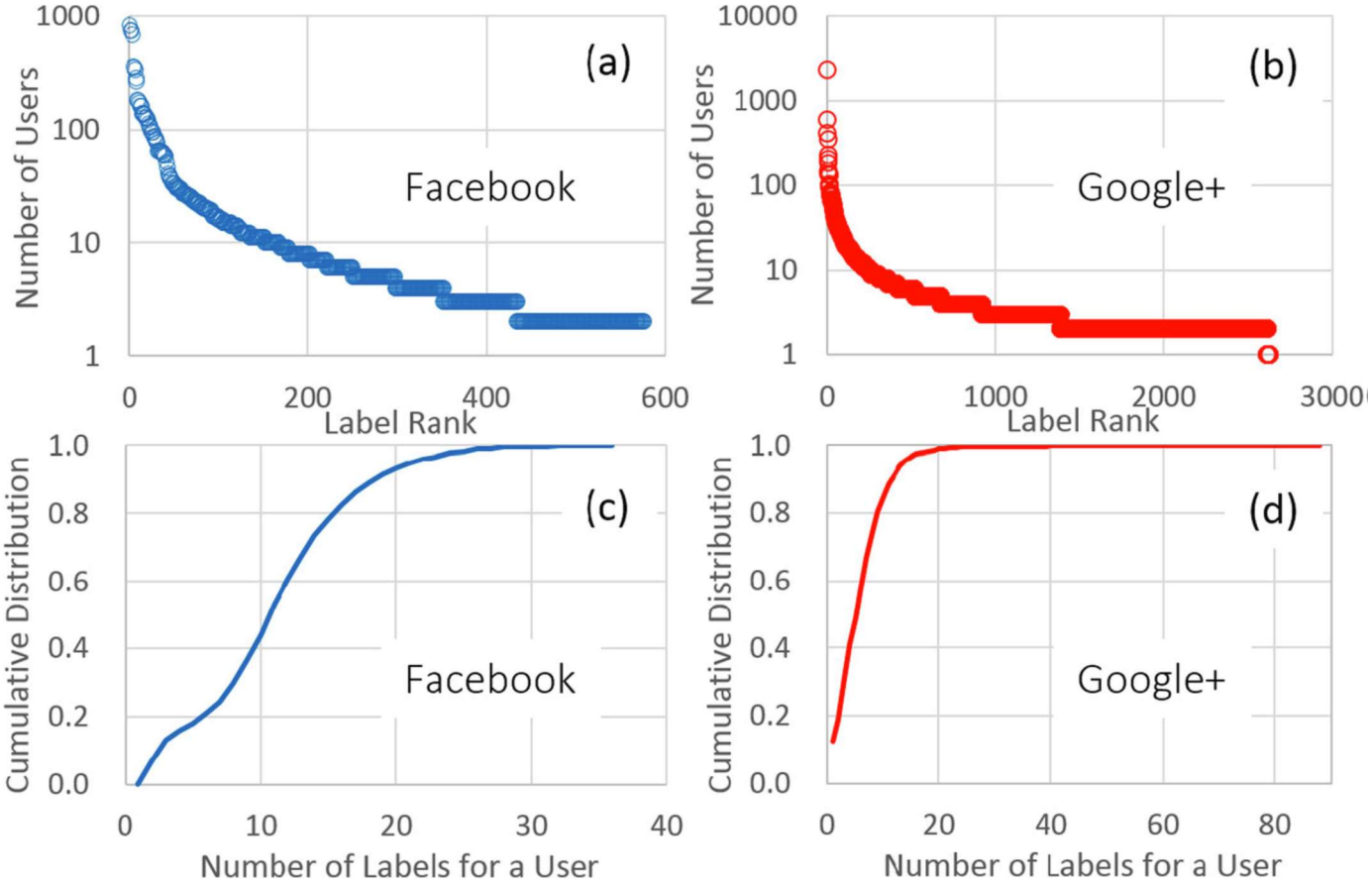


Real Social Network Datasets

- Datasets

	#nodes	#edges	#labels
Facebook	1045	53498	576
Google+	3478	435569	2627

- Labels: user attributes in profiles
  - E.g. gender, interests, skills, hometown, schools



Evaluation Settings

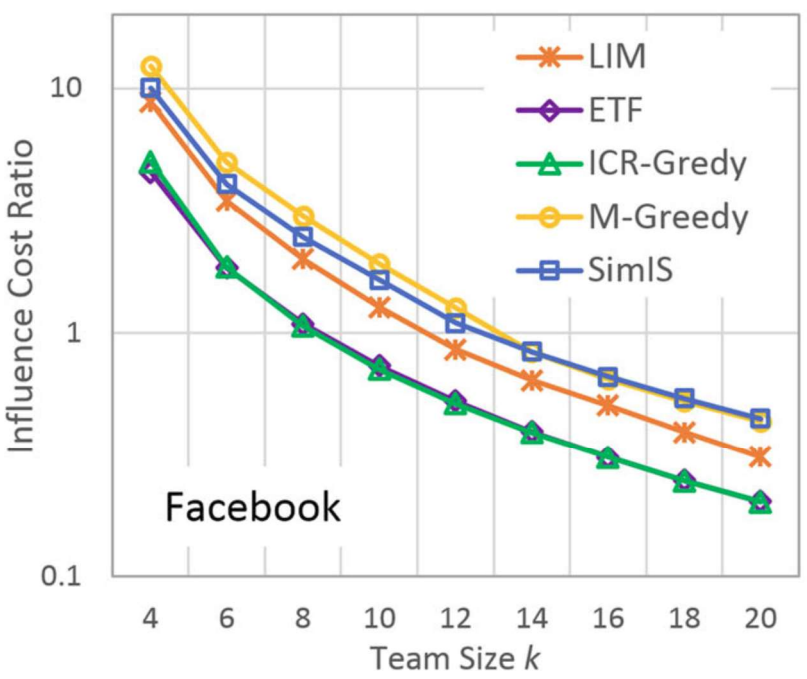
- Each required label set has 5 labels
- Random select 20 sets of required labels
- Independent Cascade model to estimate $\sigma(S)$
- Use the TRIVALENCY model to determine edge probabilities: uniformly selected from {0.1, 0.2, …, 0.9}
- Competing methods
  - Enhanced Team Formation (ETF)

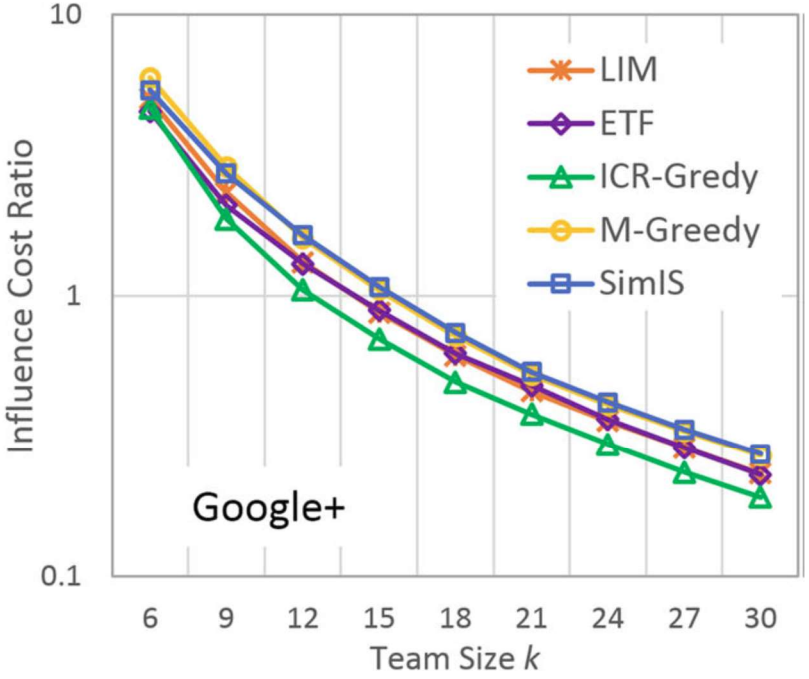
    Lappas, T., et al. Finding a team of experts in social networks. KDD 2009.
  - Linear Influence Maximization (LIM)

    Liu, Q., et al. Influence maximization over large-scale socialnetworks: a bounded linear approach. CIKM 2014.

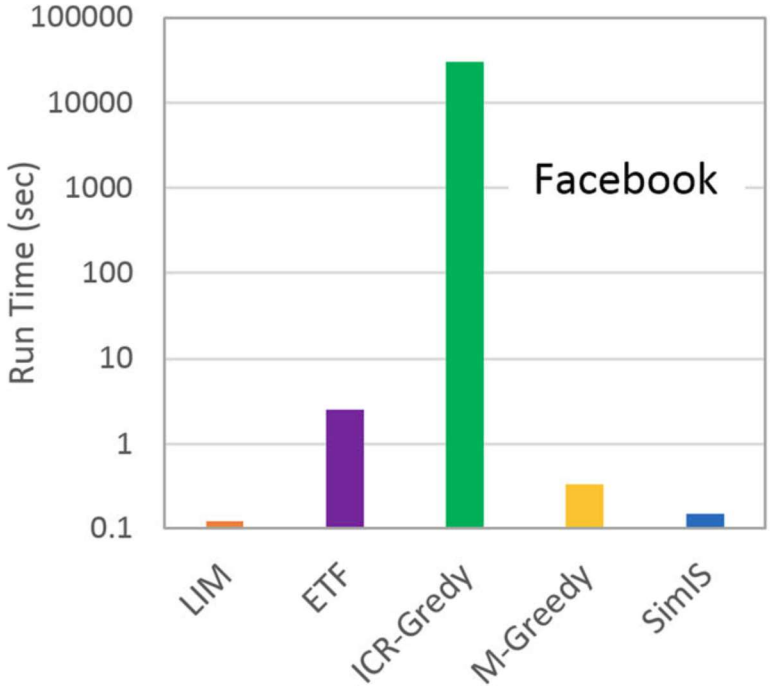
Evaluation on ICR and Time

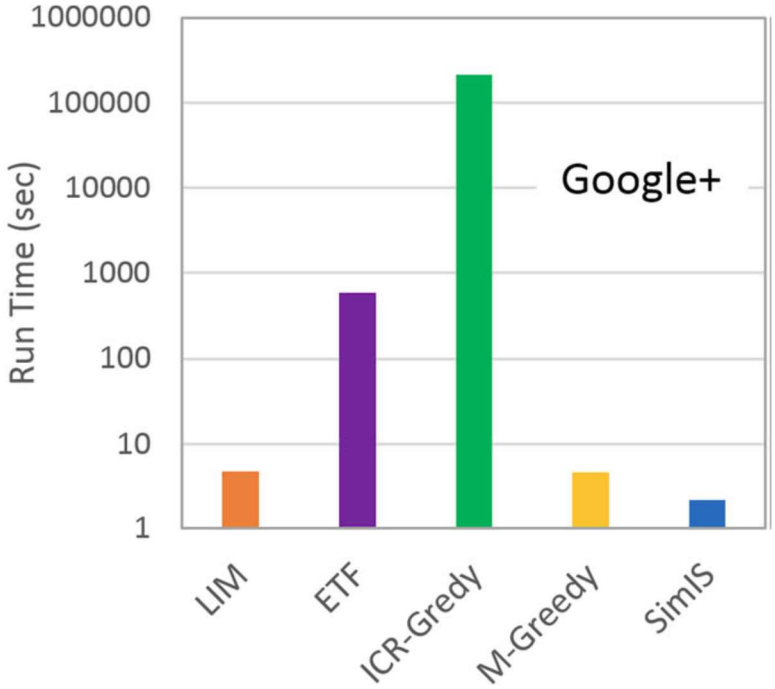
SimIS & M-Greedy are higher than others in ICR





SimS leads to the best time efficiency



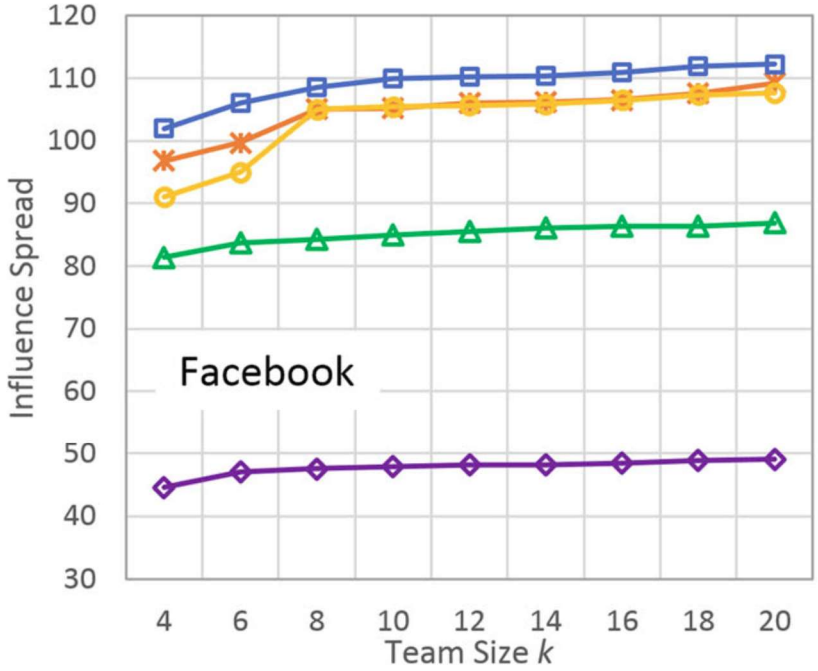


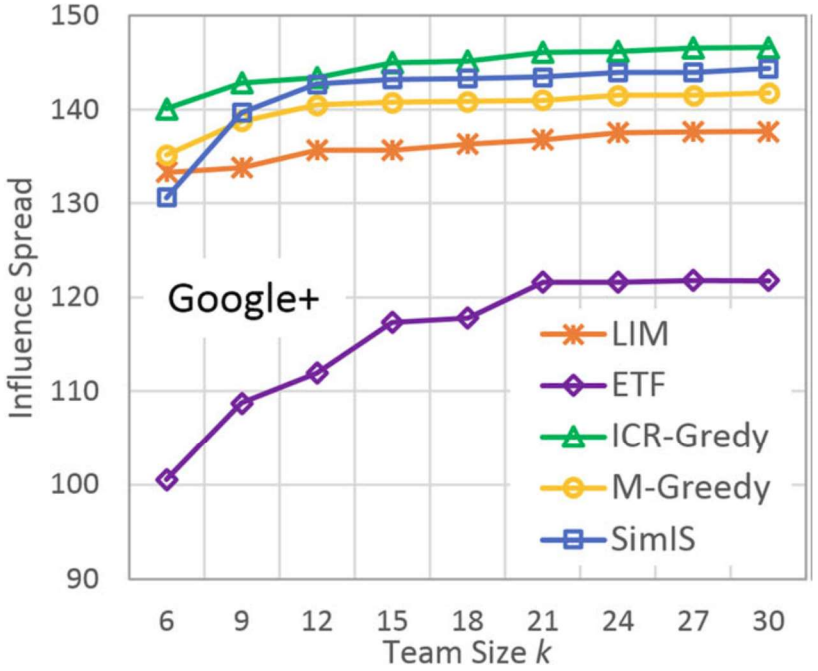
Evaluation on Influence Spread & Cost

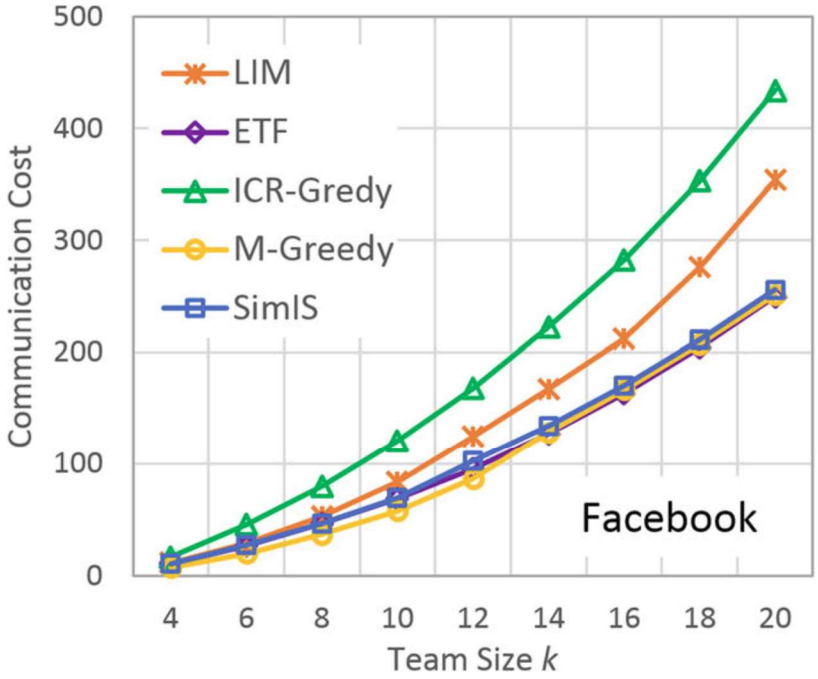
SimIM and ICR-Greedy generate higher σ

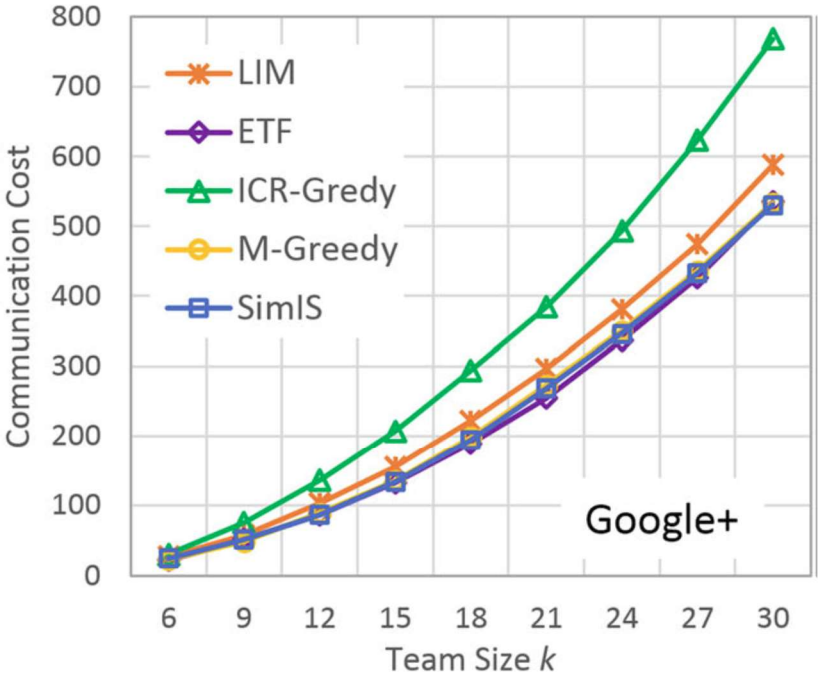
ETF and LIM have a trade-off between σ and c

ICR-Greedy is worst in c , but SimIS tends to be with lower c









Conclusions

- We propose a novel Influential Team Formation (ITF) problem to find a team of users that can best organize influential events
- Three methods are proposed ICR-Greedy, M-Greedy, and SimIS
- Experiments conducted on Facebook and Google+ datasets exhibit the superior of SimIS in terms of ICR and run time

Cheng-Te Li

Email chengte@mail.ncku.edu.tw

Web http://myweb.ncku.edu.tw/~chengte/