Non-Submodularity and Approximability: Influence Maximization in Online Social Networks

Huanyang Zheng, Ning Wang, and Jie Wu

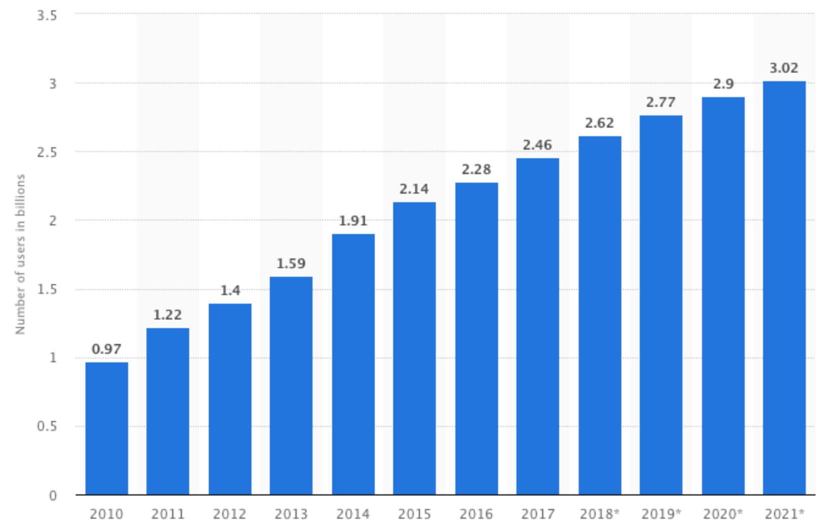
Google

Rowan University

Center for Networked Computing, Temple University

Influence Maximization in OSNs

- Online Social Networks (OSNs)
  - Facebook, Twitter, and so on



https://makeawebsitehub.com/social-media-sites/
https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Influence Maximization in OSNs

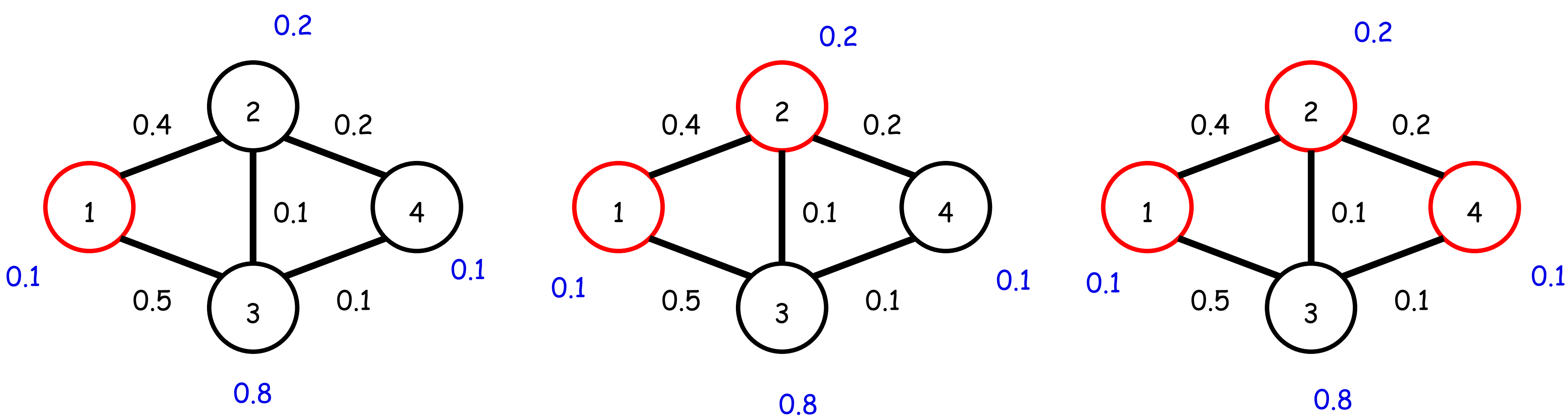
- How does influence propagate in OSNs?

Influence Maximization in OSNs

- Social influence maximization
  - Virtual marketing, personalized recommendations, feeds/news ranking …
  - Influence propagation models (NP-hard for both)
    - Linear threshold
    - Independent cascade

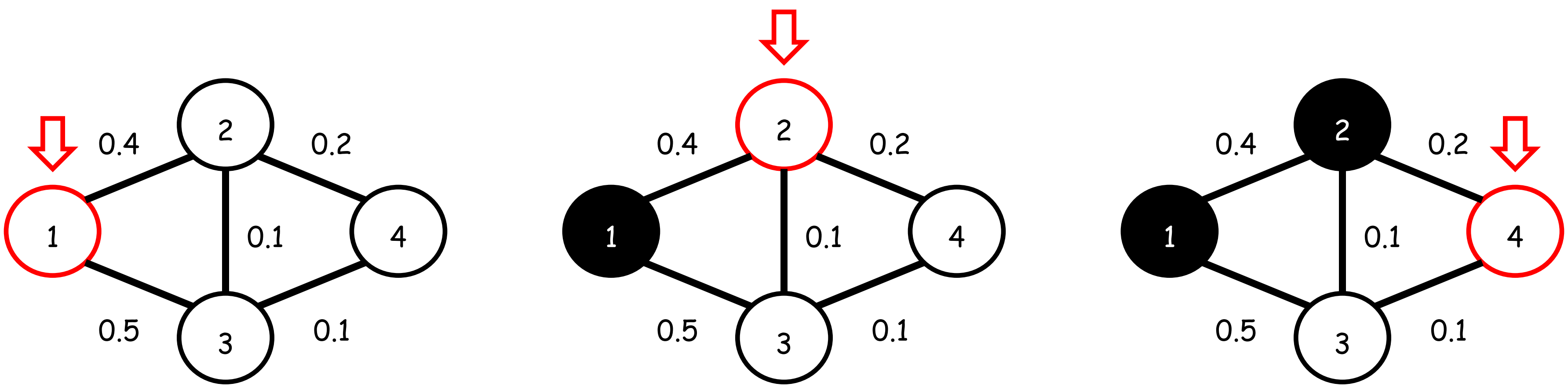
Influence Maximization in OSNs

- Linear threshold model
  - Some nodes are initially active
  - Influence spread unfolds in discrete steps
  - Each node has a threshold in the interval [0,1]
  - A node becomes active if the sum of weights of active neighbors exceeds its threshold.



Influence Maximization in OSNs

- Independent cascade model (ICM)
  - Some nodes are initially active
  - Influence spread unfolds in discrete steps
  - A active node has a single chance to activate its neighbors
  - Activation probability depends on the edge weight



Social Influence Max Problem (SIMP)

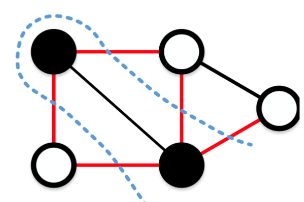
- Objective
  - Let S be the set of seed nodes (initially active)
  - Let σ(S) be the number of eventually active nodes under ICM
  - Maximize σ(S) s.t. |S|=k
- Properties
  - Monotone
    - $\sigma(S') \le \sigma(S)$ if S is a subset of S'
  - Submodular
    - $\sigma(S' \cup \{v\}) - \sigma(S') \le \sigma(S \cup \{v\}) - \sigma(S)$, i.e., diminishing return

Social Influence Max Problem (SIMP)

- Monotone & Submodular -> diminishing return

Meal 1	Meal 2
	

- Greedy leads to a 1-1/e approximation ratio.
- Some problems are not monotone or submodular
  - Submodular, but non-monotone, e.g., max-cut problem
  - Monotone, but non-submodular, e.g., welfare max problem (single shoe vs. a pair of shoes)



Two SIMP Variations

- Profit-maximization SIMP*
  - Profit maximization of a seed set, S, with the cost of the seed set

$$\sigma'(S) = \sigma(S) - c(S)$$

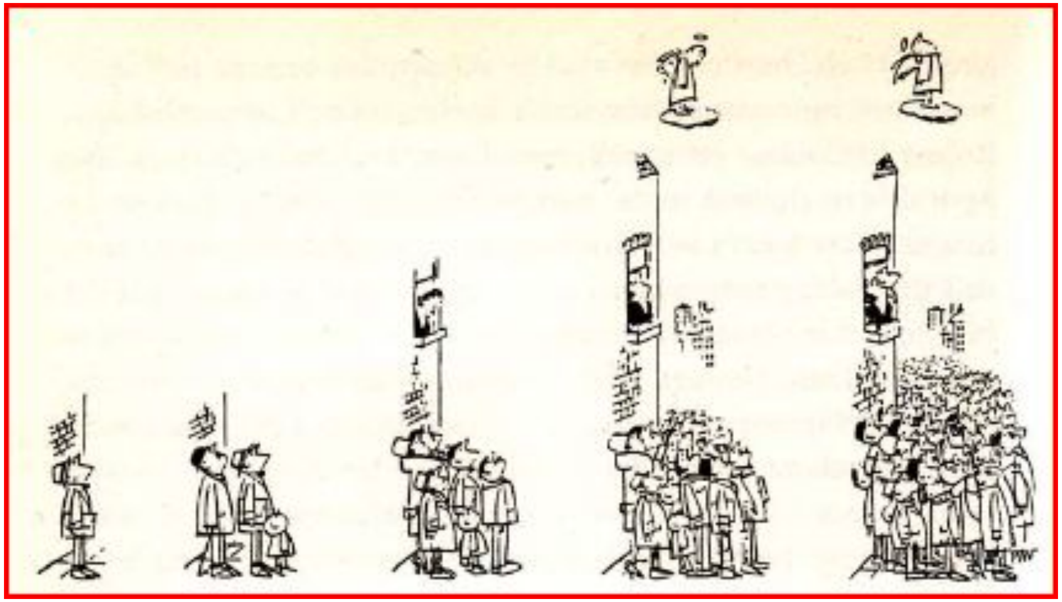
  - Submodular, but non-monotone

- Crowd-influence SIMP (this paper)
  - Crowd influence in addition to individual single seed influence
  - Monotone, but non-submodular

*Tang et al, "Profit maximization for viral marketing in online social networks," IEEE ICNP'16.

Crowd-influence

- Justification
  - Most people tend to follow the crowd

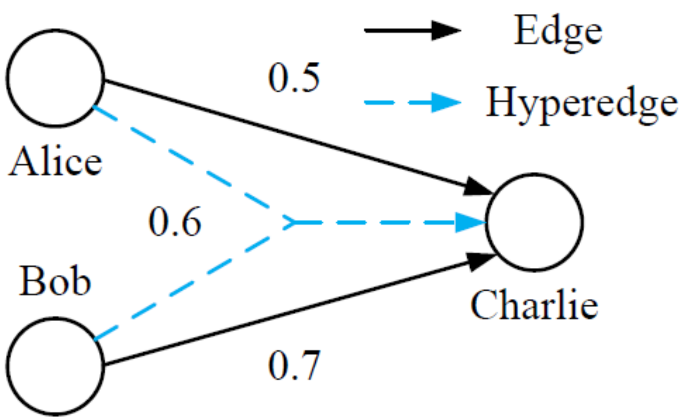


http://tutkimu.blogspot.com/2014/08/the-power-of-crowd.html

- SIMP with crowd influence poses unique challenges
  - Influence is many-to-many through hyperedges

Hyperedge

- Model
  - Influence is no longer one-to-one

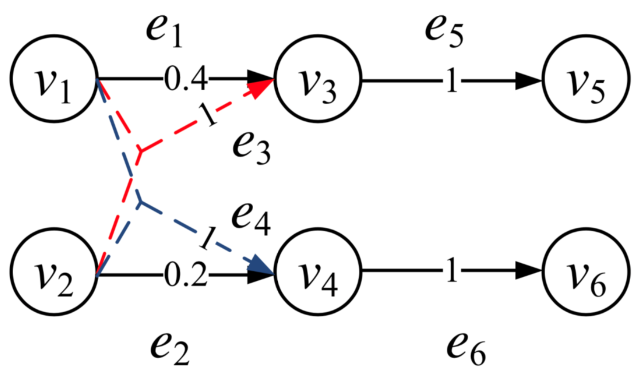


Influenced?		Probability to propagate the influence to Charlie
Alice	Bob	
Yes	No	0.5
No	Yes	0.7
Yes	Yes	1-(1-0.5)(1-0.6)(1-0.7)

- SIMP with crowd influence poses unique challenges
  - Influence is many-to-many through hyperedges
  - Monotone, but non-submodular

Supermodular degree

- Modularity set*
  - The modularity set of node v
    $M_v = \{v' \mid \sigma(S \cup \{v,v'\}) - \sigma(S \cup \{v'\}) > \sigma(S \cup \{v\}) - \sigma(S)\}$
    i.e., all nodes that might increase the marginal gain of v



  - The supermodular degree is $\Delta = \max_v |M_v|$
  - If k = 2, choosing v_1 and v_2 is the optimal.
    However, the Naïve Greedy picks v_3 and v_4.

* Feldman et al, "Constrained monotone function maximization and the supermodular degree," ACM-SIAM SODA'14.

Naïve Greedy

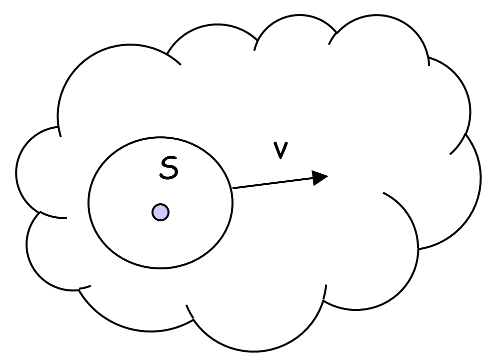
- Naïve Greedy is unbounded
  - Select the seed node that brings the largest marginal gain

Algorithm 1 Naive Greedy (NG)

Input: a hypergraph, G, and a constant, k.
Output: a set of seed nodes, S, initiated $\emptyset$.

```
while |S| < k do
    Find v = arg max_{v∈V} σ(S ∪ {v}) − σ(S).
    Update S = S ∪ {v}.
```

  - Time complexity: O(k|V||E|)



Improved Greedy

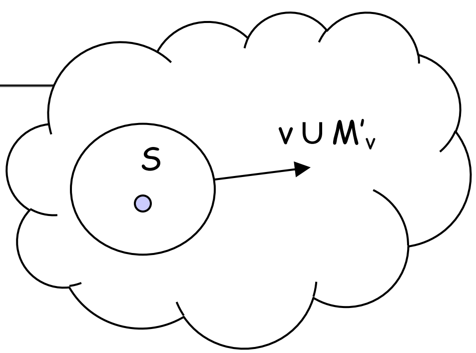
- Improved Greedy guarantees $1/(\Delta+2)$
  - Select the seed node and its modularity set that brings the largest marginal gain

---

Algorithm 2 Improved Greedy (IG)

Input: a hypergraph, G, and a constant, k.
Output: a set of seed nodes, S, initiated $\emptyset$.

while $|S| < k$ **do**
    Find $\arg\max_{v \in V, M'_v \subseteq M_v} \sigma(S \cup \{v\} \cup M'_v) - \sigma(S)$ s.t. $|S \cup \{v\} \cup M'_v| \leq k$.
    Update $S = S \cup \{v\} \cup M'_v$.



  - Time complexity: $O(2^{\Delta}k|V||E|)$

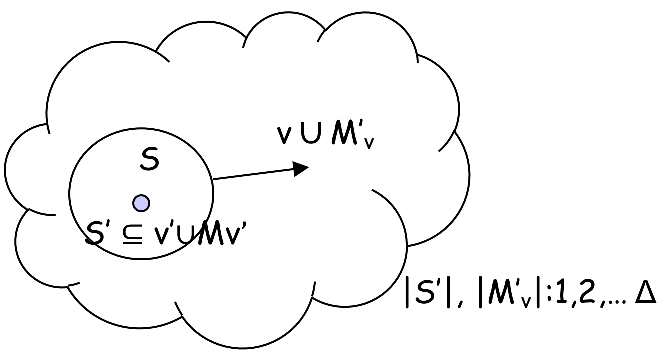
Capped Greedy

- Capped Greedy guarantees $1 - e^{-1/(\Delta+1)}$
  - Select the seed node and a capped modularity set that brings the largest marginal gain
  - Try all possible initializations on a seed node and its capped modularity set.

Algorithm 3 Capped Greedy (CG)

Input: a hypergraph, G, and a constant, k.
Output: a set of seed nodes, S, initiated $\emptyset$.

```
for each v' ∈ V do
    for each Δ' from 1 to Δ do
        for each S' ⊆ {v'} ∪ M_{v'} s.t. |S'| ≤ min{k, Δ'} do
            while |S'| < k do
                Find arg max_{v∈V, M'_v⊆M_v} σ(S' ∪ {v} ∪ M'_v) −
                σ(S') s.t. |S' ∪ {v} ∪ M'_v| ≤ k and M'_v ≤ Δ'.
                Update S' = S' ∪ {v} ∪ M'_v.
            if σ(S') > σ(S) then
                Update S = S'.
```



- Time complexity: $O(4^{\Delta}\Delta k|V|^2|E|)$

Theoretical Results

Hung et al, "When Social Influence Meets Item Inference" KDD'16:

For the SIMP with crowd influence in general graphs, no algorithm can guarantee a ratio of $|V|^{\varepsilon-1}$ for any $1 > \varepsilon > 0$

- |V| is the number of nodes in a general graph
- Inapproximable in general graphs

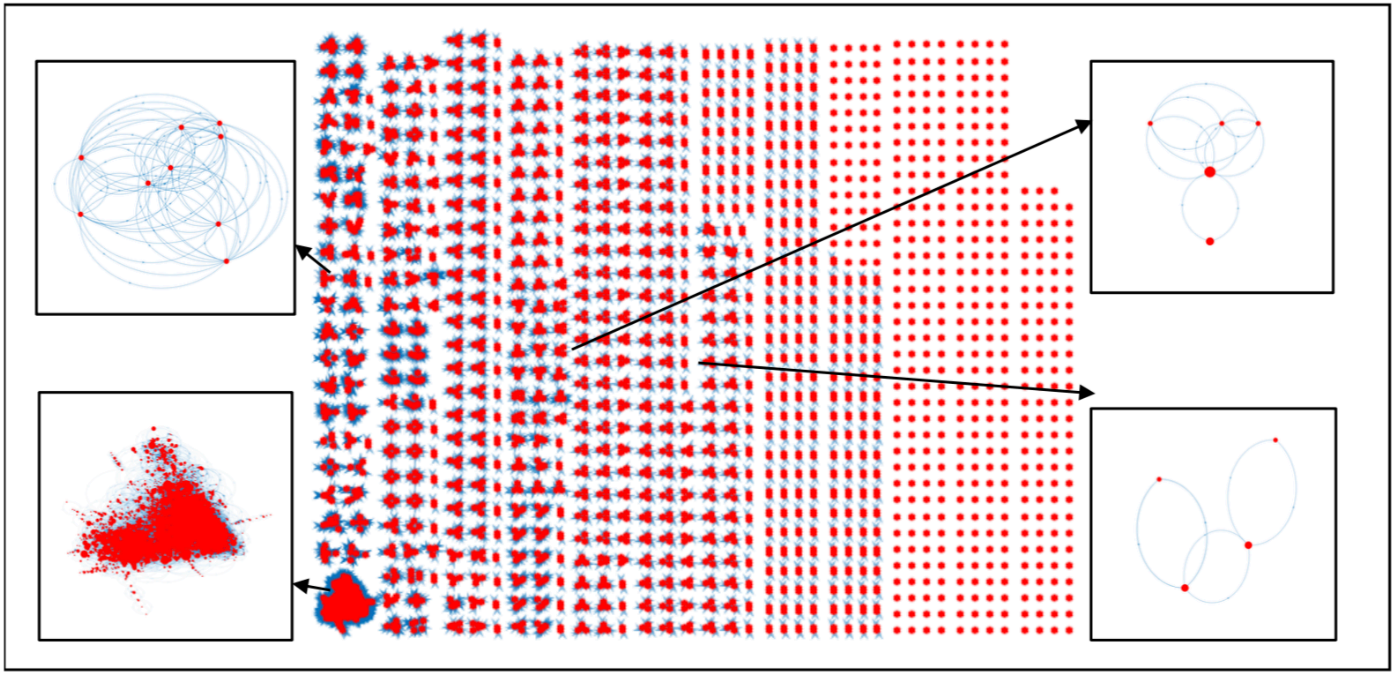
Theorem (our main contribution): The supermodular degree of most scale-free OSNs with $\overline{w}$ and γ has the following:

$$\lim_{|V|\to\infty} \frac{\Delta}{O(|V|)} = 0, \text{ when } 4 + 6\overline{w}\frac{\gamma-1}{\gamma-2} \leq 3\left(\frac{\gamma-1}{\overline{w}^{\gamma+1}}\right)^2$$

- Leveraging the structural properties of OSNs
- $\overline{w}$: average weight of the hyperedge ($w < 0.7$)
- γ: parameter in power-law distribution, $p_d = (\gamma - 1)d^{-\gamma}$ ($\gamma > 2.1$)
- $\overline{w}$ for propagation capability and γ for component size

Experiment Data

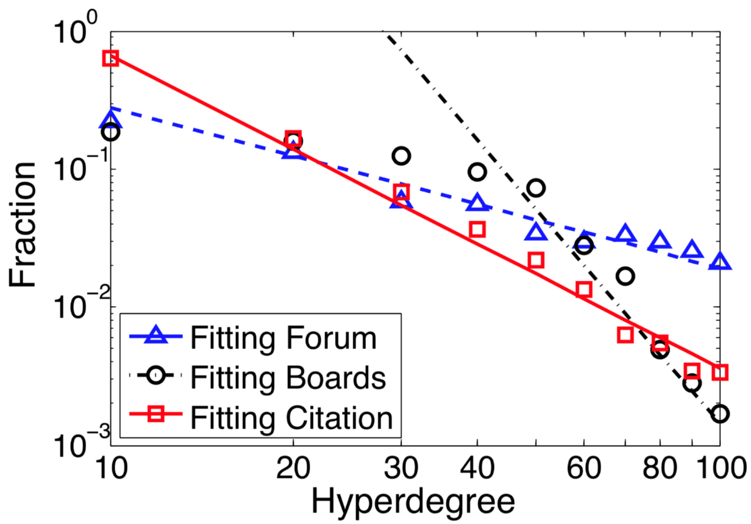
- Three datasets
  - Forum: user activates in a forum with different topics (899)
  - Board: directors belonging to the boards of companies (355)
  - Citation: collaborations among paper authors (16,726)

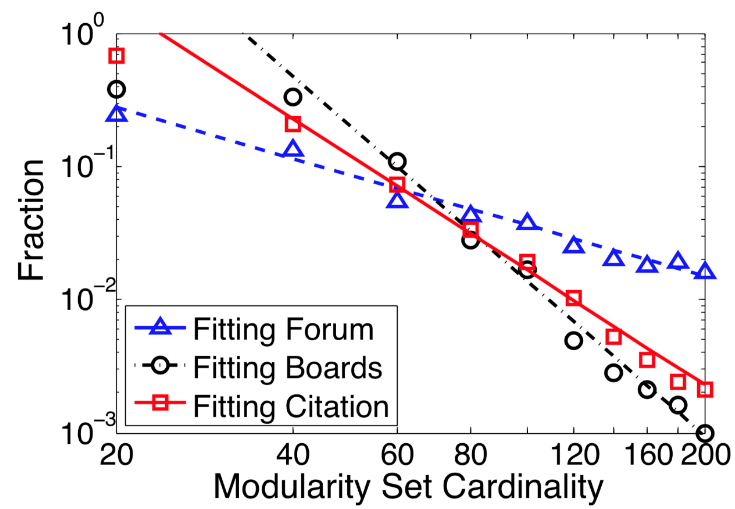


Citation Network Topology

Experiments: Modularity Set

- Trace Validation (through Monte Carlo)
  - Flags represent the real distributions by statistics,
  - Lines are the fitting curves.

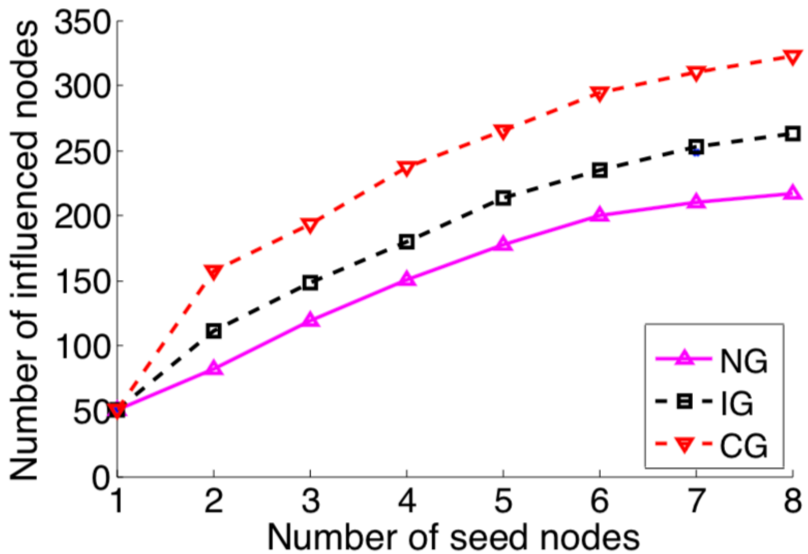




- Result
  - A small fraction of nodes have modularity sets with cardinalities larger than 100.

Experiments: Influence Maximization

- Performance evaluation in Citation dataset
  - Naïve greedy (NG) has the worst performance
  - Improved Greedy (IG) and Capped Greedy (CG) achieve better performance.



Conclusions

- Submodular function is an important property
  - Solving combinatorial problems with bounded results
- However, many real applications are not modeled as submodular and monotone functions
  - Submodular, but non-monotone
  - Monotone, but non-submodular
- The optimization using supermodular degree
  - Applied to OSN-related problems with relatively low complexity

Questions