On Maximizing Diffusion Speed in Social Networks

- Impact of Random Seeding and Clustering-

Jungseul Ok

Joint work with Youngmi Jin, Jinwoo Shin and Yung Yi

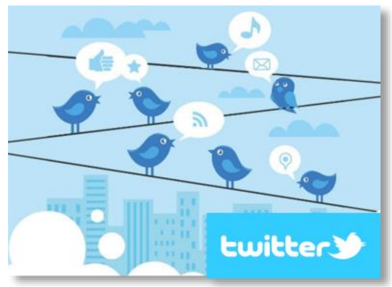
Electrical Engineering

Korea **A**dvanced **I**nstitute of **S**cience **T**echnology

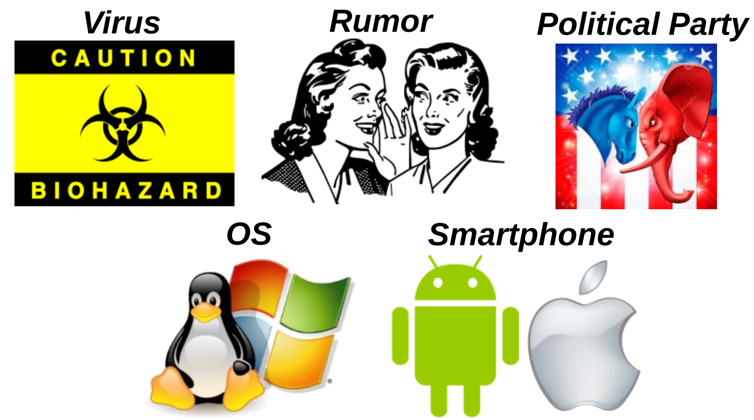
18 June 2014, ACM Sigmetrics

Diffusion over Social Network

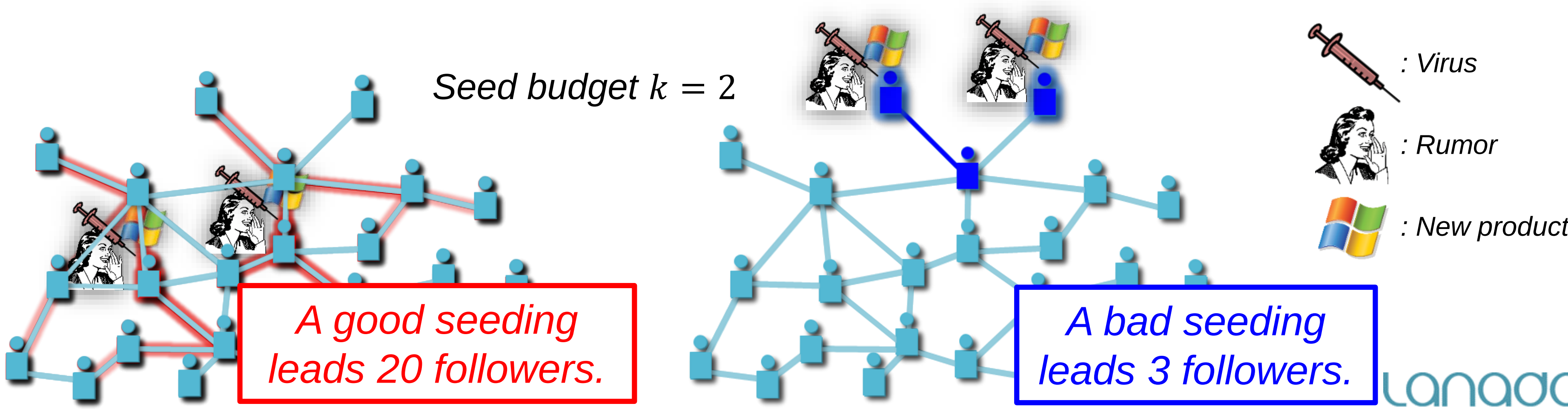
- Diffusion by social interaction



Online Social Network Services



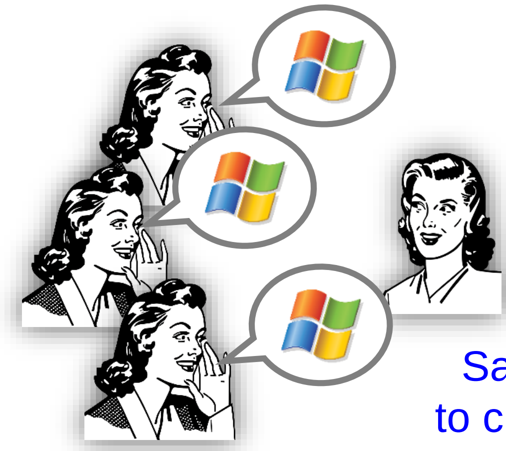
- Influence maximization via seeding
  - For given budget *k*, selecting *k* people as seeds to maximize number of influenced people.
  - Seeding: injecting virus, telling rumor, or supporting new product in advance.



Understanding of Diffusion

Epidemic-based model

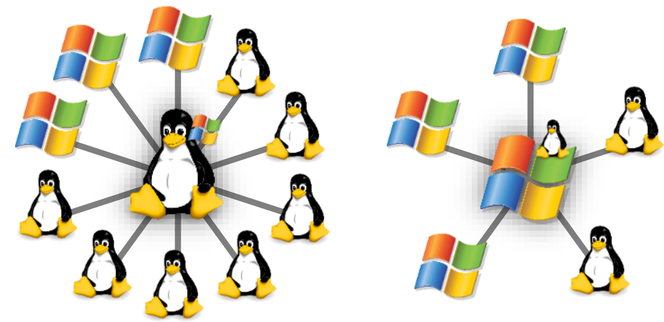
- A user is passively infected by preceding followers.
  - Number of followers is important.



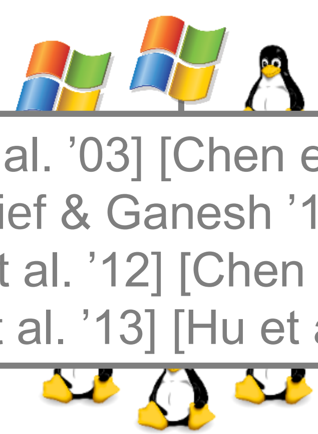
Same probability to choose Windows

Game-based model

- A user strategically selects his choice against his neighbors.
  - Fraction of followers is important.



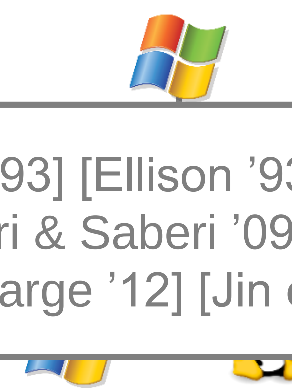
User A has higher probability to choose Windows than user B.



[Kempe et al. ’03] [Chen et al. ’10] [Draief & Ganesh ’11] [Banerjee et al. ’12] [Chen et al. ’12] [Goyal et al. ’13] [Hu et al. ’13]

=

≠



[Blume ’93] [Ellison ’93] [Morris ’00] [Montanari & Saberi ’09] [Liu et al. ’10] [Lelarge ’12] [Jin et al. ’13]

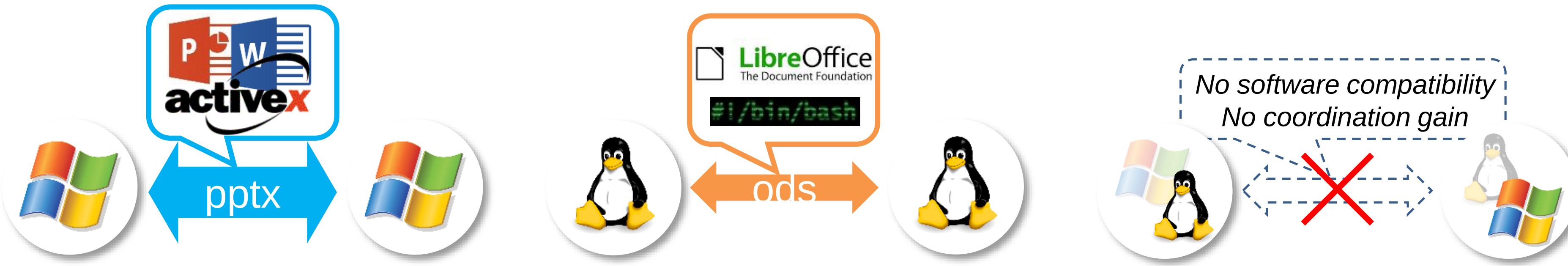
Seeding in Diffusion Models

- **Epidemic-based model**
  - Theoretic studies
    - [Kempe et al. ’03] [Chen et al. ’12] [Goyal et al. ’13]
  - Practical studies
    - [Chen et al. ’10] [Hu et al. ’13]
- **Game-based model?**

Our Main Goal

Seeding algorithms for maximizing diffusion over game-based model.

Pairwise Coordination Game



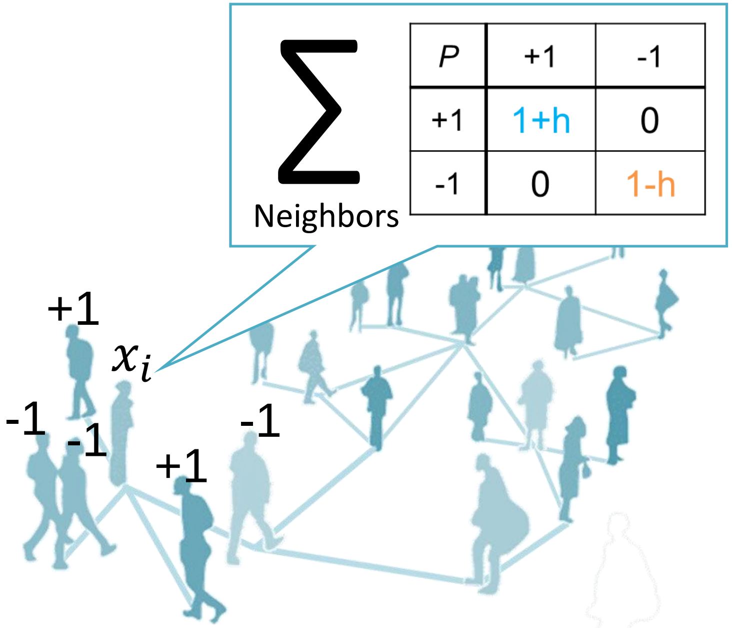
- Coordination gain
  - Gain from same choices, e.g., *software compatibility*
- *(Only in this talk)*
  Suppose that Windows provides more coordination gain.
- Payoff matrix P with game parameter $0 < h < 1$
  - +1: new one with better coordination gain
  - -1: old one

  - Game parameter h: difference of coordination gains of Windows (+1) and Linux (-1).

P	+1	-1
+1	1+h	0
-1	0	1-h

Networked Coordination Game

- Social network: graph $G = (V, E)$
  - V: vertex set, E: edge set
  - Vertex: user, edge: social relationship
- Total payoff: sum of the payoffs with each neighbor
- Update based on total payoffs with +1 and -1
  - A user updates his choice at each tick of unit Poisson clock.

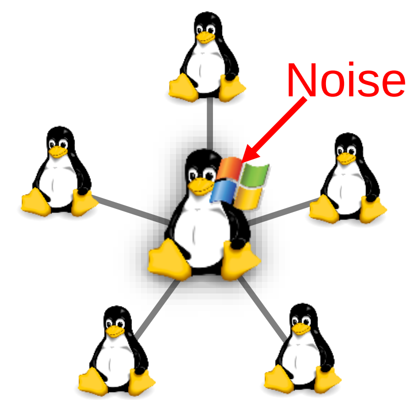


P	+1	-1
+1	1+h	0
-1	0	1-h

Update Rule

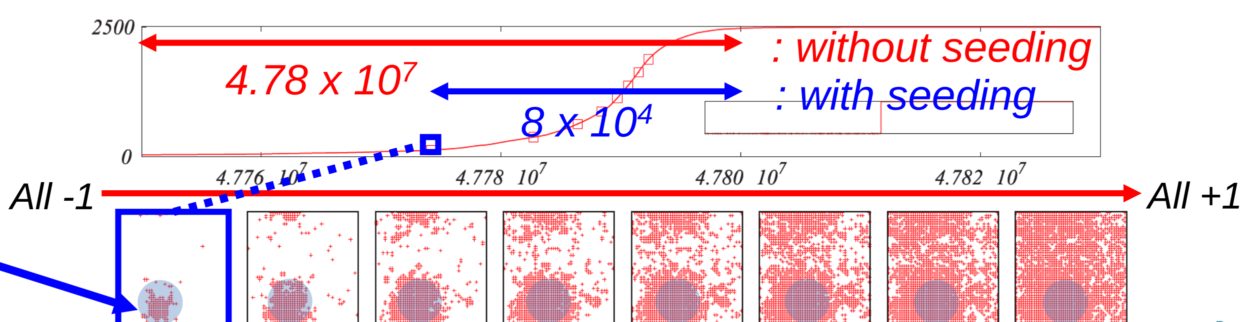
- Noisy best response
  - Best response: a rational user tries to maximize his payoff at each update.
  - Noise: people are sometimes irrational.
- Logit dynamics with noise parameter β
  - probability of being rational $\propto \exp(\beta K_i)$
  - probability of being irrational $\propto \exp(-\beta K_i)$

  where K_i is absolute value of difference between payoffs with best response and its reverse.



- Slow diffusion time
  - Every user finally adopts +1 and mostly keeps +1.
  - However, diffusion time to all +1 is often impractically long.
    - In small world network: $\approx \exp(\beta\Omega(\text{number of users}))$ [Montanari & Saberi '09]

Diffusion speed maximization via seeding



Problem and Challenge

- Diffusion time $\tau(C)$: median of diffusion time with seed set C
  - Users in seed set C are persuaded to adopt +1 forever.
  - Just estimating $\tau(C)$ takes exponential time.
  - Minimization of $\tau(C)$ is neither algebraic nor combinatorial.
- Using meta-stability theory [Montanari & Saberi '09],

$$\tau(C) \approx \exp(\beta\Gamma(C))$$

$$\Gamma(C) = \max_{C \subset S_0 \subset V} \min_{S: S_0 \to V} \max_{t < |S|} [H(S_t) - H(S_0)] \quad H(S) = \mathrm{cut}(S, V \setminus S) - \sum_{i \in S} h|N(i)|$$

≈ (tilted) cut-width, a combinatorial value

  - Now, we have a combinatorial optimization but it is still challenging.
  - There is no sub/super-modularity (diminishing return or synergy effect).

P1.

$$\min_{C \subset V} \quad \tau(C) \qquad \text{subject to} \quad |C| \le k$$

→

P2.

$$\min_{C \subset V} \quad \Gamma(C) \qquad \text{subject to} \quad |C| \le k$$

→ **Approximate Solution**

Approximate Solution

Definition. (γ, δ)-approximate solution of *P2*

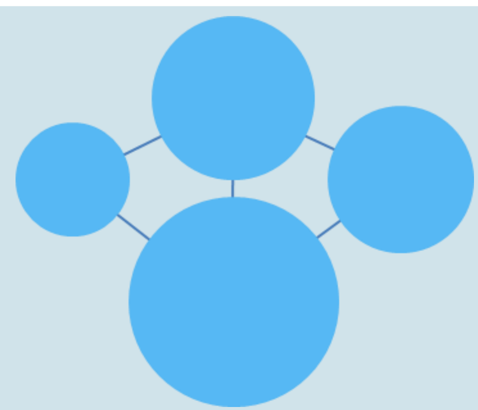
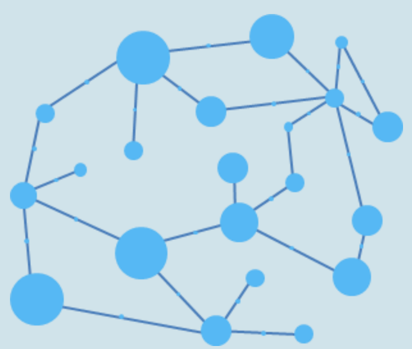
γ: degree of sub-optimality in *objective value*

- A seed set $C \subset V$ with $|C| \leq k$ is called a (γ, δ)-approximate solution of the seeding problem if

$$\Gamma(C) \leq \gamma \cdot \min_{C' : |C'| \leq \delta k} \Gamma(C')$$

where $\gamma \geq 1$ and $\delta \leq 1$

δ: degree of sub-optimality in *budget*

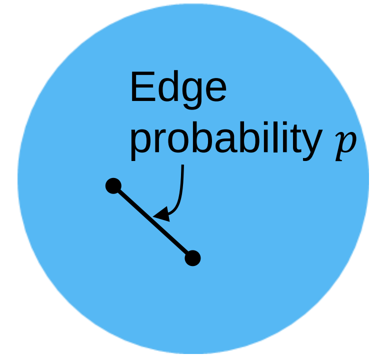
(i) Erdős–Rényi graph *Globally well-connected*	**(ii) Planted partition graph** *Locally well-connected with big clusters*	**(iii) Geometrically structured graph** *Locally well-connected with small clusters:*
		

(i) Erdős–Rényi Graph

• y p.

Theorem 1. ER Graph

- For an ER graph with $\omega(1)$ average degree and given seed budget k, an arbitrary seed set C with $|C| = k$ is almost surely a $(1+\varepsilon, 1)$-approximate solution for large enough n and any given $\varepsilon > 0$.

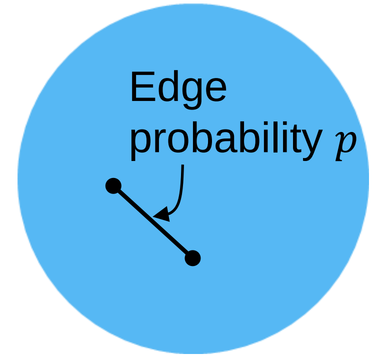


→ No careful seeding mechanism is necessary.

(ii) Planted Partition Graph

Theorem 1. ER Graph

- For an ER graph with $\omega(1)$ average degree and given seed budget k, an arbitrary seed set C with $|C| = k$ is almost surely a $(1+\varepsilon, 1)$-approximate solution for large enough n and any given $\varepsilon > 0$.



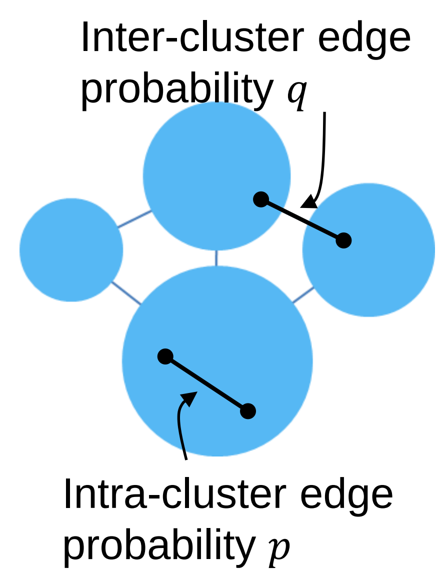
Theorem 2. Planted Partition Graph

- For a planted partition graph with $p/q = \omega(1)$ and the seed budget k, every seed set C such that

$$C \in \arg \min_{\{C' : |C'| \leq k\}} \max_{1 \leq l \leq m} \left(\frac{1-h}{2} |V_l| - |C' \cap V_l| \right)$$

is almost surely a $(1+\varepsilon, 1)$-approximate solution for large enough n and any given $\varepsilon > 0$.

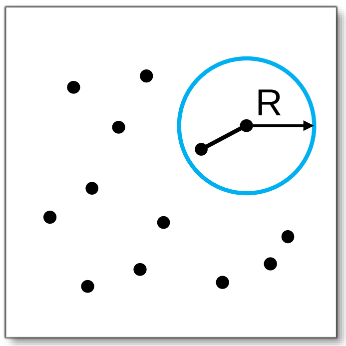
Roughly speaking, it is budget allocation proportional to cluster size.



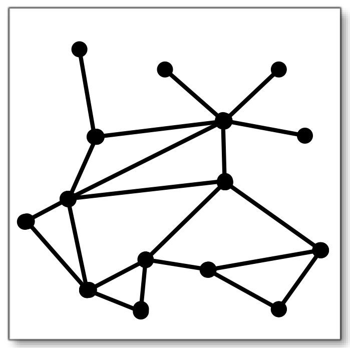
→ Only carefully allocated number of seeds in each cluster (min-max) is required.

(iii) Geometrically Structured Graph

- *d*-dimensional graph and planar graph



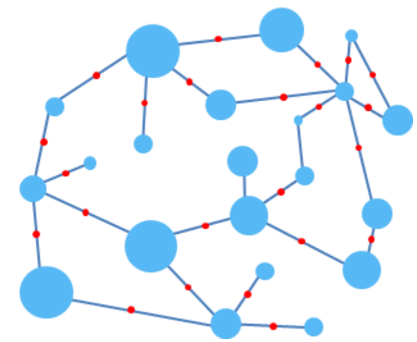
2-dimensional graph



planar graph

[Jung & Shah '07]

***Red**: inter-cluster connectors*
***Blue**: local users*



Locally well-connected with small clusters ($O(1)$ size)

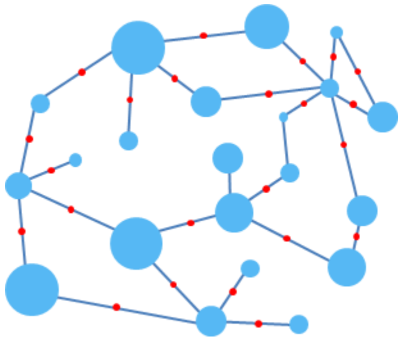
Theorem 3. Geometrically Structured Graph

- For a d-dimensional graph or planar graph and given budget k, we design a polynomial-time algorithm that outputs a $(1, 1-\varepsilon)$-approximate solution for any given $\varepsilon \in (0,1)$.

Partitioning and Seeding

- **Partitioning phase**
  - Partition the graph by inter-cluster connectors.
  - And select them as default seed set.

***Red**: inter-cluster connectors*
***Blue**: local users*



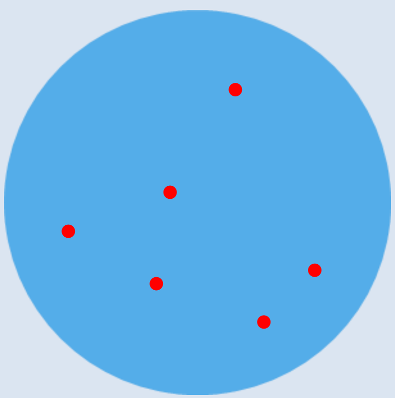
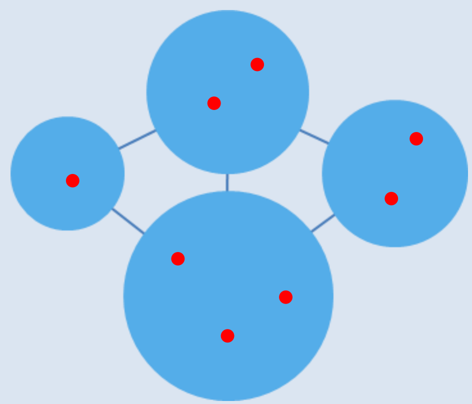
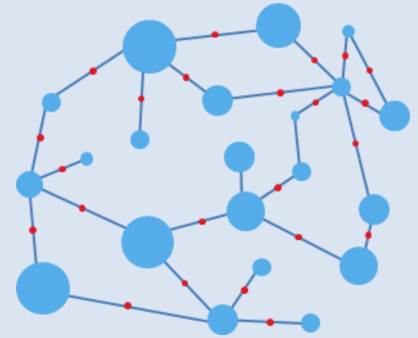
- **Seeding phase**
  - For remaining seed budget: "greedy" allocation
    - Step1. Find slowest cluster
      and add one more budget to the slowest cluster.
    - Step2. Optimize seed set for the cluster with the increased budget.

→ For this locally well-connected graph with small clusters,
we should give higher priority to seed inter-cluster connectors.

lanada

Result at a Glance

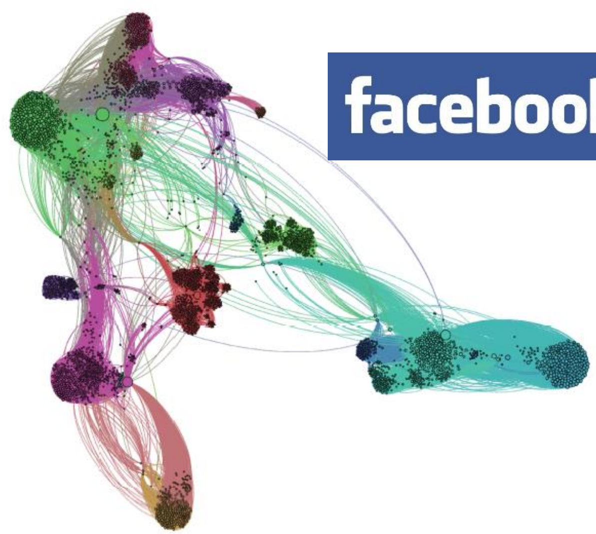
- Nearly optimal seeding strategies:

(i) Erdős–Rényi graph *Globally well-connected*	**(ii) Planted partition graph** *Locally well-connected with big clusters*	**(iii) Geometrically structured graph** *Locally well-connected with small clusters:*
		

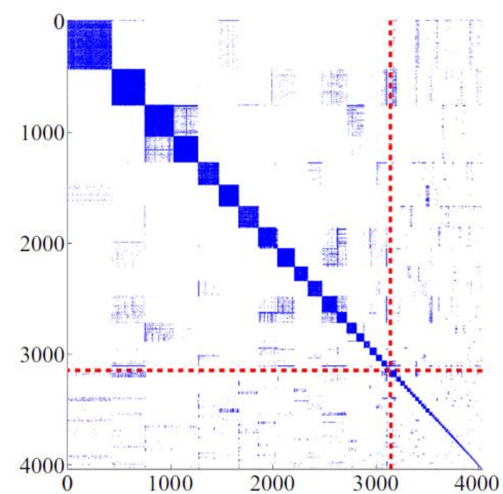
Practical Seeding Algorithm

- **PrPaS:** Practical PaS
  - Scalable & non-parametric
  - (i) Partitioning phase
    - A partitioning algorithm
    - We select [Rosvall & Bergstrom '08].
  - (ii) Seeding phase
    - Proportional budget allocation
    - Random seeding in each cluster

- Comparison group
  - **Degree** selects *k* nodes with highest degrees.
  - **GreedyCut** greedily maximizes $\mathrm{cut}(C, V \setminus C)$.
  - **Random** randomly chooses k nodes.



A blueprint of Facebook network
consisting of 4039 nodes and 88234 edges

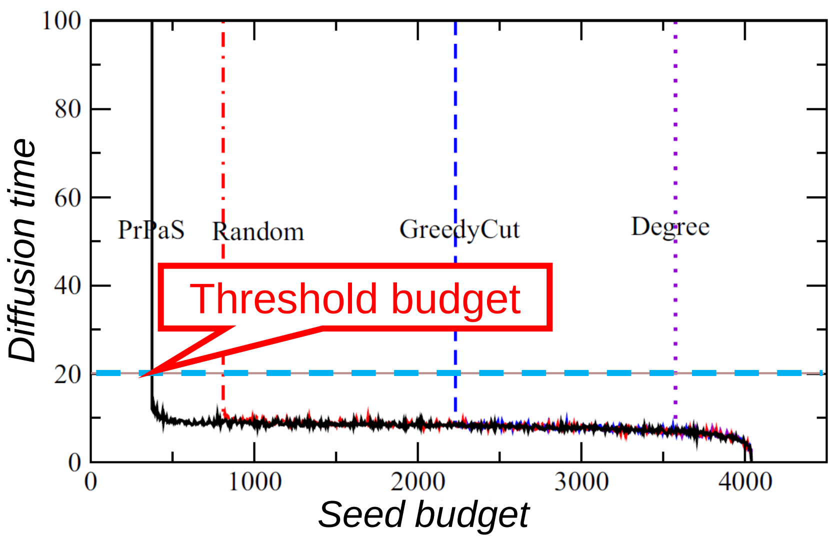


The adjacency matrix of Facebook network

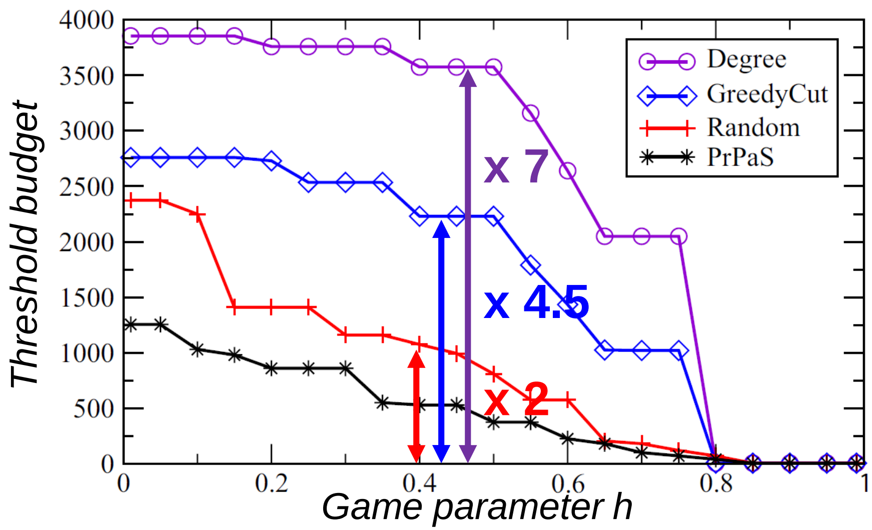
* Facebook network data is from http://snap.stanford.edu/data/egonets-Facebook.html

Simulation Results

- The rationality parameter β=10.



Diffusion times with varying seed budgets (h = 0.5)

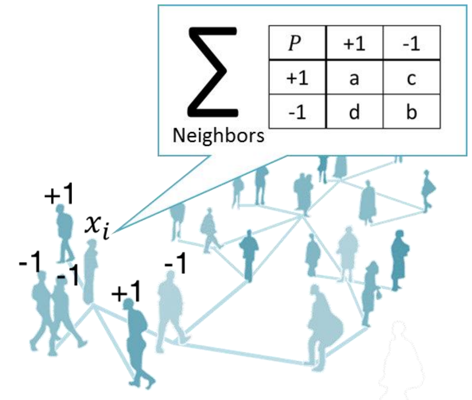


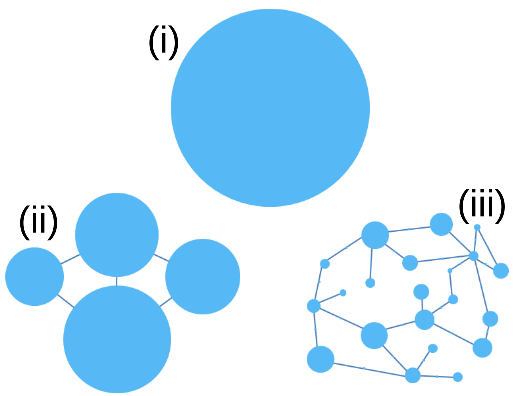
Threshold (=20) with varying h for different algorithms

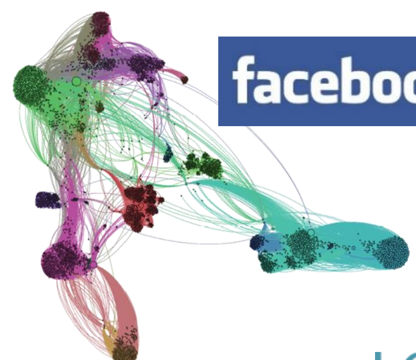
- Threshold budget for target diffusion time x
  - Minimum seed budget required for $\tau(C) < x = 20$.
  - Less threshold implies better seeding.

Summary

- **Model**
  : Networked coordination game among users with bounded rationality
- **Formulation**
  : Diffusion time minimization problem
- **Main Result**
  : Impact of **random seeding** and **clustering**
  - (i) globally well-connected graph
  - (ii) locally well-connected graph with big clusters ($\Theta(n)$ size)
  - (iii) locally well-connected graph with small clusters ($O(1)$ size)
- **PrPaS Algorithm**
  : **Pr**actical **Pa**rtitioning **a**nd **S**eeding algorithm
  - Extensive simulation on Facebook network







facebook

Thank you

Jungseul Ok
ockjs@kaist.ac.kr