A REFERRAL-REWARD EMBEDDED, BI-PHASE INFORMATION DIFFUSION TECHNIQUE

Sneha Mondal

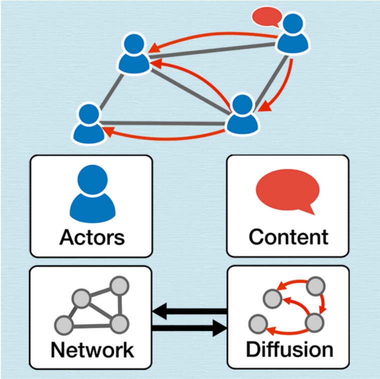
M.Sc. (Engg.) Colloquium
Faculty Advisor : Prof. Y. Narahari

January 30, 2017

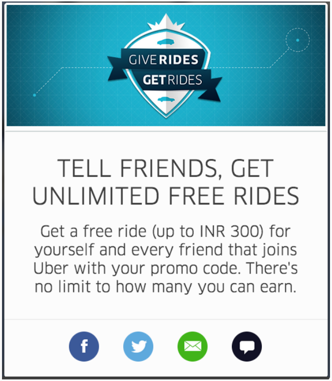
Department of Computer Science and Automation
Indian Institute of Science, Bangalore

INTRODUCTION

- Diffusion via word-of-mouth
  Identify initial adopters.
  Word-of-mouth influence propagation



- Diffusion via word-of-mouth
- Referral Rewards
  Refer product to friends and acquaintances. Get incentives for successful referral



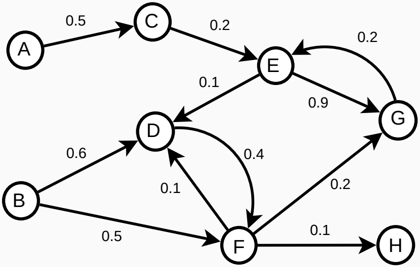
- Given
  - Target consumer base
  - Estimates for “influence” between individuals
  - Budget K as initial endowment

- Given
  - Target consumer base
  - Estimates for “influence” between individuals
  - Budget K as initial endowment
- Goal
  - Trigger cascade of product adoptions
  - Maximize set of eventual customers!

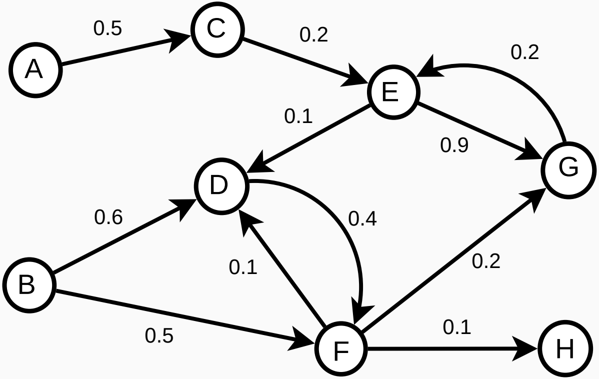
PROBLEM SETTING

- Given
  - Target consumer base
  - Estimates for “influence” between individuals
  - Budget K as initial endowment
- Goal
  - Trigger cascade of product adoptions
  - Maximize set of eventual customers!
- Design Problems
  - Which advertising channels are most effective?
  - How to spread initial budget across advertising channels?

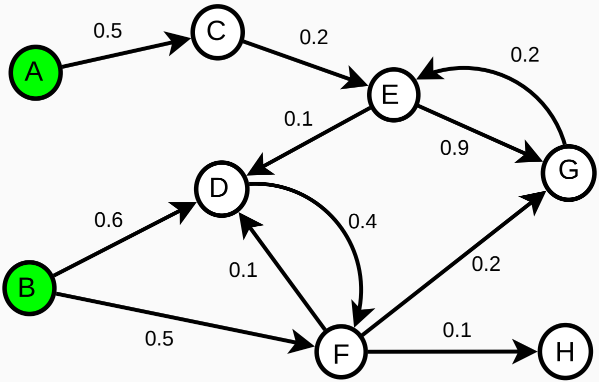
DIFFUSION MODEL - INDEPENDENT CASCADE (IC)



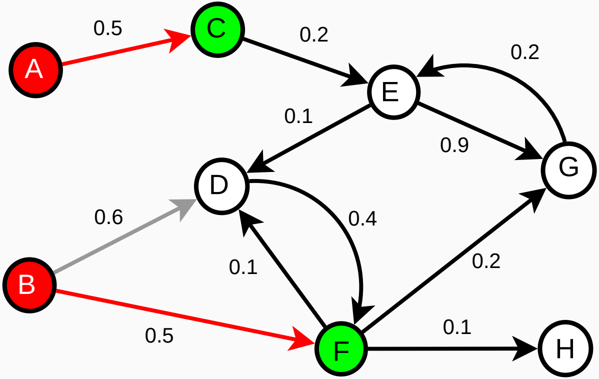
- Social network graph G
- When node u becomes active, it has a single chance of activating each currently inactive neighbour v
- Activation attempt succeeds with probability p_{uv}
- Process terminates when no further nodes can be activated


A
C
E
G
D
B
F
H
0.5
0.2
0.2
0.1
0.9
0.6
0.4
0.1
0.2
0.5
0.1

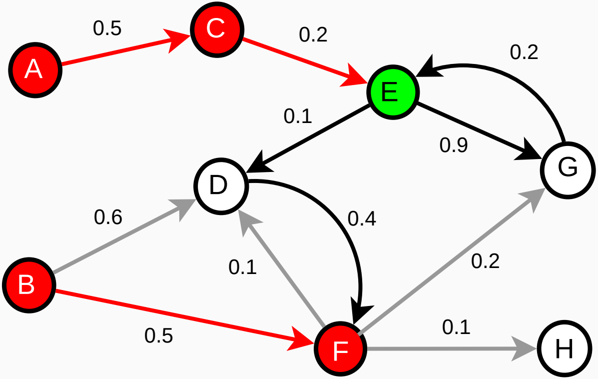
INDEPENDENT CASCADE – EXAMPLE



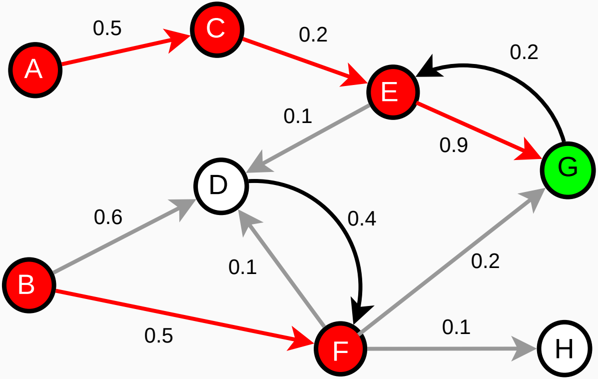
INDEPENDENT CASCADE - EXAMPLE



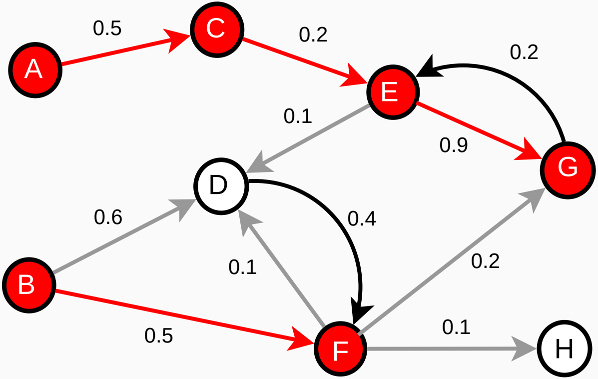
INDEPENDENT CASCADE – EXAMPLE



INDEPENDENT CASCADE - EXAMPLE



INDEPENDENT CASCADE - EXAMPLE



COMPUTING INFLUENCE – LIVE GRAPH

$$A \xrightarrow{0.5} B \xrightarrow{0.5} C$$

Live Graph ($\mathcal{X}$)	$f_{\mathcal{X}}(\{A\})$	$P(\mathcal{X})$
	1	0.25
	2	0.25
	1	0.25
	3	0.25

$$
\begin{aligned}
f(\{A\}) &= \sum_{\mathcal{X}} \mathrm{P}(\mathcal{X}) f_{\mathcal{X}}(\{A\}) \\
&= 0.25 * (1 + 2 + 1 + 3) \\
&= 1.75
\end{aligned}
$$

Live Graph ($\mathcal{X}$)	$f_{\mathcal{X}}(\{A\})$	$P(\mathcal{X})$
	1	0.25
	2	0.25
	1	0.25
	3	0.25

$$\mathrm{P}(\mathcal{X}) = \prod_{e \in \mathcal{X}} p_e \prod_{e \notin \mathcal{X}} (1 - p_e)$$

$$f(S) = \sum_{\mathcal{X}} \mathrm{P}(\mathcal{X}) f_{\mathcal{X}}(S)$$

RELEVANT LITERATURE AND RESEARCH GAP

- Influence maximization in a network in a single phase using seed nodes [1]
  $max_{|S| \leq K} f(S)$
  sub-modular optimization, greedy algorithm

[1] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In ACM SIGKDD, pages 137–146, 2003.

[2] P. Dayama, A. Karnik, and Y. Narahari. Optimal incentive timing strategies for product marketing on social networks. In AAMAS, pages 703–710, 2012.

[3] S. Dhamal, K. J. Prabuchandran, and Y. Narahari. Information diffusion in social networks in two phases. IEEE TNSE, 3(4):197–210, 2016.

- Influence maximization in a network in a single phase using seed nodes [1]
- Influence maximization using referral incentives [2]
  Optimal referral pricing, maximize profit to company

[1] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In ACM SIGKDD, pages 137–146, 2003.
[2] P. Dayama, A. Karnik, and Y. Narahari. Optimal incentive timing strategies for product marketing on social networks. In AAMAS, pages 703–710, 2012.
[3] S. Dhamal, K. J. Prabuchandran, and Y. Narahari. Information diffusion in social networks in two phases. IEEE TNSE, 3(4):197–210, 2016.

- Influence maximization in a network in a single phase using seed nodes [1]
- Influence maximization using referral incentives [2]
- Influence maximization in a network in two phases using seed nodes [3]
  Given $K \rightarrow$ select k_1 seeds for phase 1 $\rightarrow$ observe spread $\rightarrow$ select remaining $K - k_1$ seeds

[1] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In ACM SIGKDD, pages 137–146, 2003.

[2] P. Dayama, A. Karnik, and Y. Narahari. Optimal incentive timing strategies for product marketing on social networks. In AAMAS, pages 703–710, 2012.

[3] S. Dhamal, K. J. Prabuchandran, and Y. Narahari. Information diffusion in social networks in two phases. IEEE TNSE, 3(4):197–210, 2016.

- Influence maximization in a network in a single phase using seed nodes [1]
- Influence maximization using referral incentives [2]
- Influence maximization in a network in two phases using seed nodes [3]

Influence maximization with budget-split in two phases, using seed nodes, followed by referral incentives

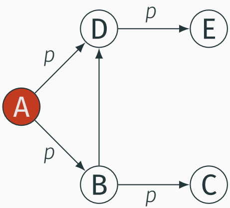
[1] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In ACM SIGKDD, pages 137–146, 2003.

[2] P. Dayama, A. Karnik, and Y. Narahari. Optimal incentive timing strategies for product marketing on social networks. In AAMAS, pages 703–710, 2012.

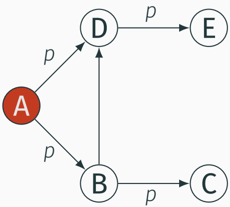
[3] S. Dhamal, K. J. Prabuchandran, and Y. Narahari. Information diffusion in social networks in two phases. IEEE TNSE, 3(4):197–210, 2016.

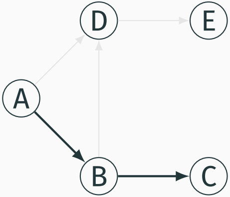
MODEL AND PROBLEM FORMULATION

EXAMPLE (K=2, k= 1, α = 0.5) PHASE 1



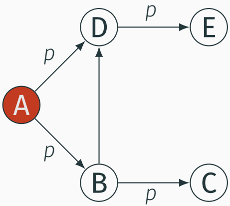
Example (k=2, k= 1, α = 0.5) phase 1

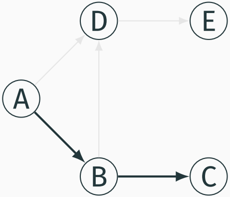




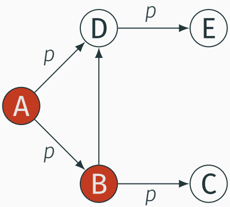
Live Graph

EXAMPLE (K=2, k= 1, α = 0.5) PHASE 1



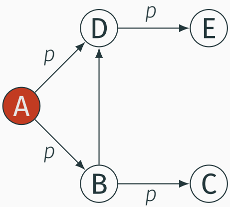


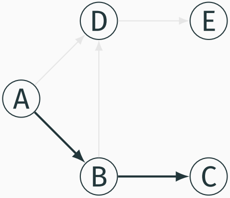
Live Graph



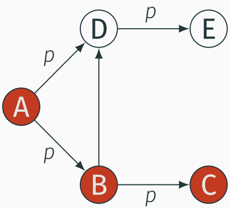
Phase 1

EXAMPLE (K=2, k= 1, α = 0.5) PHASE 1



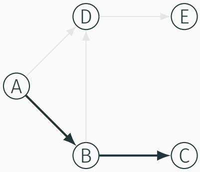


Live Graph

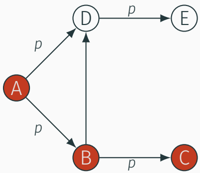


Phase 1

Phase 1 ($k = 1$)



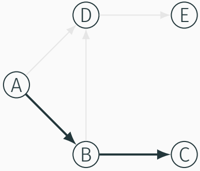
$\mathcal{X}$



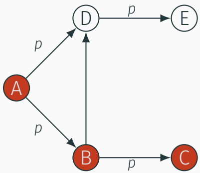
Phase 2 ($\alpha = \frac{1}{2}$)

PHASE 2

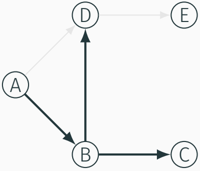
Phase 1 ($k = 1$)



$\mathcal{X}$



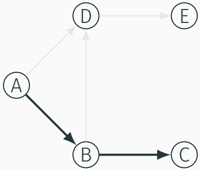
Phase 2 ($\alpha = \frac{1}{2}$)



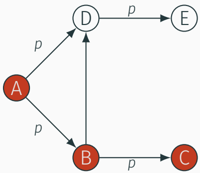
$\mathcal{Y}_1|\mathcal{X}$

PHASE 2

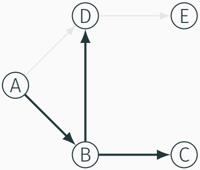
Phase 1 ($k = 1$)



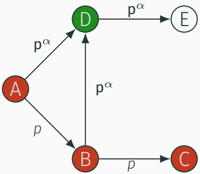
$\mathcal{X}$



Phase 2 ($\alpha = \frac{1}{2}$)

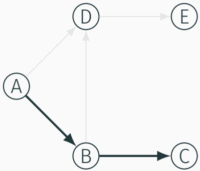


$\mathcal{Y}_1 | \mathcal{X}$

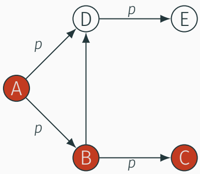


PHASE 2

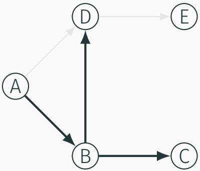
Phase 1 ($k = 1$)



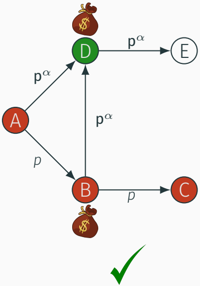
$\mathcal{X}$



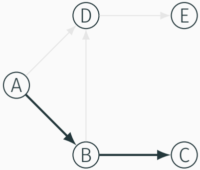
Phase 2 ($\alpha = \frac{1}{2}$)



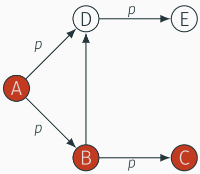
$\mathcal{Y}_1 | \mathcal{X}$



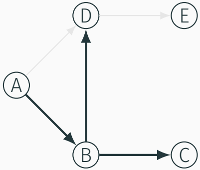
Phase 1 ($k = 1$)



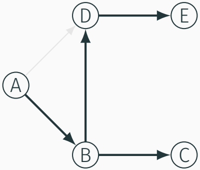
$\mathcal{X}$



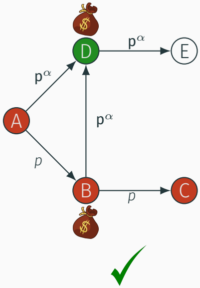
Phase 2 ($\alpha = \frac{1}{2}$)



$\mathcal{Y}_1|\mathcal{X}$

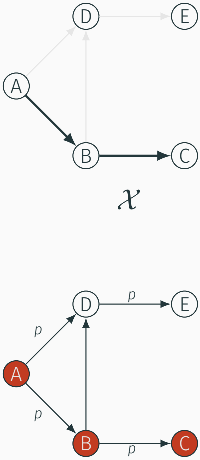


$\mathcal{Y}_2|\mathcal{X}$

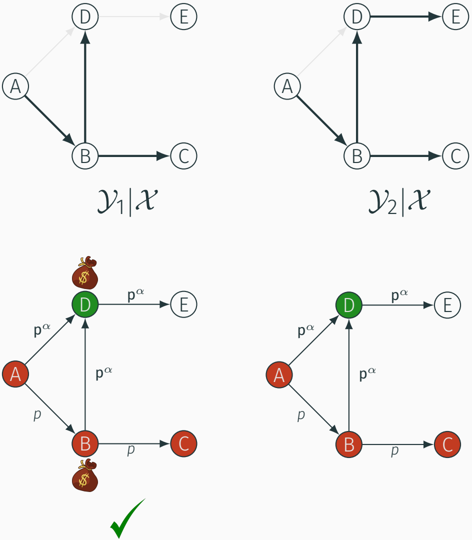


PHASE 2

Phase 1 ($k = 1$)

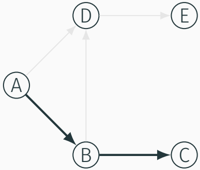


Phase 2 ($\alpha = \frac{1}{2}$)

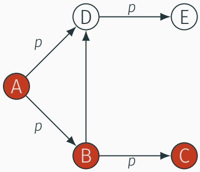


PHASE 2

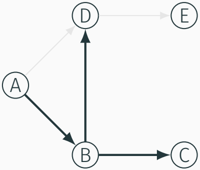
Phase 1 ($k = 1$)



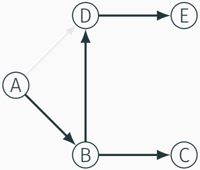
$\mathcal{X}$



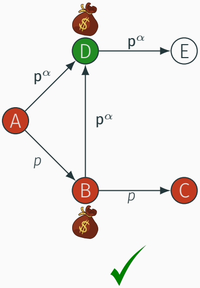
Phase 2 ($\alpha = \frac{1}{2}$)

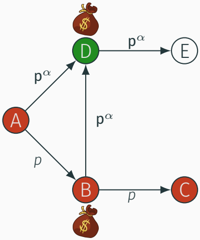


$\mathcal{Y}_1 | \mathcal{X}$



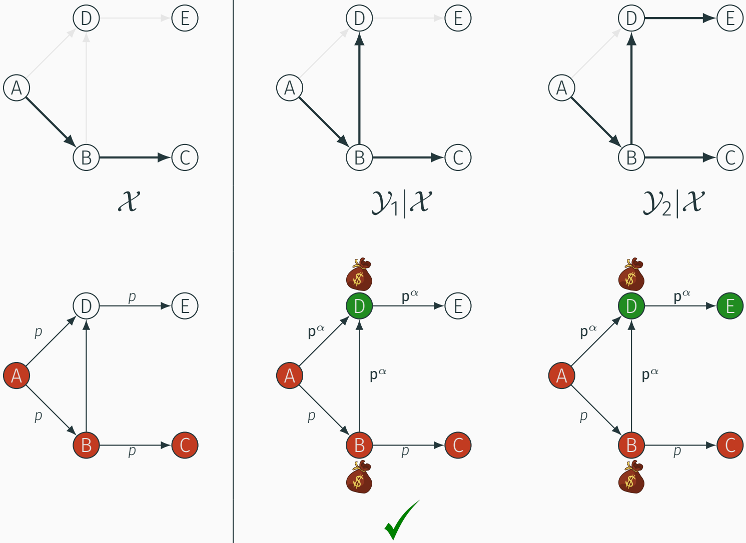
$\mathcal{Y}_2 | \mathcal{X}$





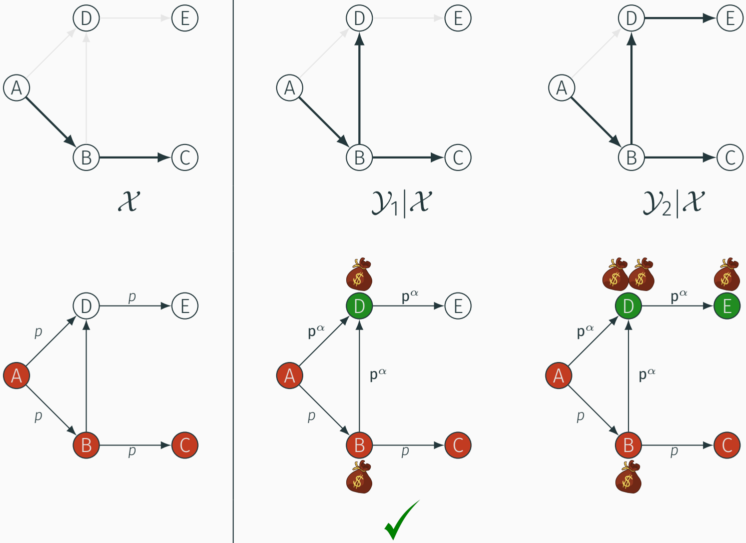
PHASE 2

Phase 1 ($k = 1$) | Phase 2 ($\alpha = \frac{1}{2}$)

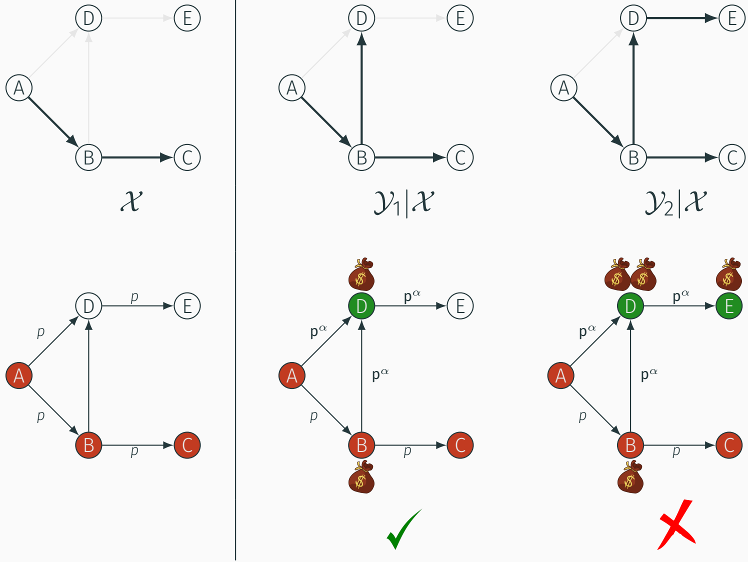


PHASE 2

Phase 1 ($k = 1$) | Phase 2 ($\alpha = \frac{1}{2}$)



Phase 1 ($k = 1$) | Phase 2 ($\alpha = \frac{1}{2}$)



How to capture effect of α-reward?

How to capture effect of α-reward?

- Incentive to individual nodes
- Overall edge influence probabilities expected to increase!

EFFECT OF α

How to capture effect of α-reward?

- Incentive to individual nodes
- Overall edge influence probabilities expected to increase!

h(α) : fractional increase in edge influence probability

How to capture effect of α-reward?

- Incentive to individual nodes
- Overall edge influence probabilities expected to increase!

h(α) : fractional increase in edge influence probability

- non-negative
- no reward $\implies$ no increase in probability; $h(0) = 0$
- non-decreasing in [0,1]

How to capture effect of α-reward?

- Incentive to individual nodes
- Overall edge influence probabilities expected to increase!

h(α) : fractional increase in edge influence probability

- non-negative
- no reward $\implies$ no increase in probability; $h(0) = 0$
- non-decreasing in [0,1]
- $h(\alpha) = log(1 + \alpha)$
  $p^{\alpha} = (1 + log(1 + \alpha))p$

How to capture effect of α-reward?

- Incentive to individual nodes
- Overall edge influence probabilities expected to increase!

h(α) : fractional increase in edge influence probability

- non-negative
- no reward $\implies$ no increase in probability; $h(0) = 0$
- non-decreasing in [0,1]
- $h(\alpha) = log(1 + \alpha)$
  $p^{\alpha} = \min\{1, (1 + log(1 + \alpha))p\}$

- $\mathcal{X}$: Live graph obtained after phase 1

- $\mathcal{X}$: Live graph obtained after phase 1 ; $p(\mathcal{X})$

- $\mathcal{X}$: Live graph obtained after phase 1 ; $p(\mathcal{X})$
- $A^{\mathcal{X}}_{diff}$ = Nodes active after phase 1

$$A^{\mathcal{X}}_{diff} = \{v \mid v \text{ is reachable from } S^k \text{ in } \mathcal{X}\}$$

- $\mathcal{X}$: Live graph obtained after phase 1 ; $p(\mathcal{X})$
- $A_{diff}^{\mathcal{X}}$ = Nodes active after phase 1

$$A_{diff}^{\mathcal{X}} = \{v | \ v \text{ is reachable from } S^k \text{ in } \mathcal{X}\}$$

- $\mathcal{Y}$: Live graph obtained after phase 2

- $\mathcal{X}$: Live graph obtained after phase 1 ; $p(\mathcal{X})$
- $A^{\mathcal{X}}_{diff}$ = Nodes active after phase 1

$$A^{\mathcal{X}}_{diff} = \{v |\ v \text{ is reachable from } S^k \text{ in } \mathcal{X}\}$$

- $\mathcal{Y}$: Live graph obtained after phase 2 ; $p(\mathcal{Y}|\mathcal{X};\alpha)$

- $\mathcal{X}$: Live graph obtained after phase 1 ; $p(\mathcal{X})$
- $A^{\mathcal{X}}_{diff}$ = Nodes active after phase 1

$$A^{\mathcal{X}}_{diff} = \{v |\ v \text{ is reachable from } S^k \text{ in } \mathcal{X}\}$$

- $\mathcal{Y}$: Live graph obtained after phase 2 ; $p(\mathcal{Y}|\mathcal{X};\alpha)$
- $A^{\mathcal{Y}}_{ref}$ = Additional nodes activated in phase 2

$$A^{\mathcal{Y}}_{ref} = \{v |\ v \text{ is reachable from } A^{\mathcal{X}}_{diff} \text{ in } \mathcal{Y}\} \setminus A^{\mathcal{X}}_{diff}$$

- $\mathcal{X}$: Live graph obtained after phase 1 ; $p(\mathcal{X})$
- $A_{diff}^{\mathcal{X}}$ = Nodes active after phase 1

$$A_{diff}^{\mathcal{X}} = \{v |\ v \text{ is reachable from } S^k \text{ in } \mathcal{X}\}$$

- $\mathcal{Y}$: Live graph obtained after phase 2 ; $p(\mathcal{Y}|\mathcal{X};\alpha)$
- $A_{ref}^{\mathcal{Y}}$ = Additional nodes activated in phase 2

$$A_{ref}^{\mathcal{Y}} = \{v |\ v \text{ is reachable from } A_{diff}^{\mathcal{X}} \text{ in } \mathcal{Y}\} \setminus A_{diff}^{\mathcal{X}}$$

- $f(S^k, \alpha)$ = Expected number of influenced nodes

$$f(S^k, \alpha) = \sum_{\mathcal{X}} p(\mathcal{X}) \left\{ |A_{diff}^{\mathcal{X}}| + \sum_{\mathcal{Y}} p(\mathcal{Y}|\mathcal{X};\alpha) |A_{ref}^{\mathcal{Y}}| \right\}$$

- $\mathcal{X}$: Live graph obtained after phase 1 ; $p(\mathcal{X})$
- $A^{\mathcal{X}}_{diff}$ = Nodes active after phase 1

$$A^{\mathcal{X}}_{diff} = \{v|\ v \text{ is reachable from } S^k \text{ in } \mathcal{X}\}$$

- $\mathcal{Y}$: Live graph obtained after phase 2 ; $p(\mathcal{Y}|\mathcal{X};\alpha)$
- $A^{\mathcal{Y}}_{ref}$ = Additional nodes activated in phase 2

$$A^{\mathcal{Y}}_{ref} = \{v|\ v \text{ is reachable from } A^{\mathcal{X}}_{diff} \text{ in } \mathcal{Y}\} \setminus A^{\mathcal{X}}_{diff}$$

- $f(S^k, \alpha)$ = Expected number of influenced nodes

$$f(S^k, \alpha) = \sum_{\mathcal{X}} p(\mathcal{X}) \left\{ |A^{\mathcal{X}}_{diff}| + \sum_{\mathcal{Y}} p(\mathcal{Y}|\mathcal{X};\alpha)|A^{\mathcal{Y}}_{ref}| \right\}$$

- For a fixed α, $f(S, \alpha)$ is non-negative, monotone, and sub-modular in S

Select (S_k, α) to give

$$\max_{\substack{k \leq K, \alpha \in [0,1] \\ S_k \subset V}} f(S_k, \alpha) = \underbrace{\mathbb{E}\left[|A_{diff}(S_k)|\right]}_{\text{depends on } k} + \underbrace{\mathbb{E}\left[|A_{ref}(S_k; \alpha)|\right]}_{\text{depends on } k,\, h(\alpha)}$$

subject to

$$\mathbb{E}\left[|A_{ref}(S^k; \alpha)|\right] \leq \frac{K-k}{2\alpha}$$

Algorithm: A modified greedy algorithm for seed selection

Input: Graph G, budget K, split (k, α)

Output: Optimal seed set S_k such that $|S_k| \leq k$

$S_k \leftarrow \phi$

for $t \leftarrow 1$ *to* k **do**

  for $v \notin S_k$ **do**

    Compute $f(S_k \cup \{v\})$

  $V_{valid} \leftarrow \left\{v \in V \setminus S_k : \mathbb{E}|A_{ref}(S_k \cup \{v\})| \leq \frac{K-k}{2\alpha}\right\}$

  $v_t \leftarrow \arg\max_{v \in V_{valid}} f(S_k \cup \{v\}) - f(S_k)$

  if $\{v_t\} \neq \phi$ **then**

    $S_k \leftarrow S_k \cup \{v_t\}$

  else

    return S_k

Algorithm: A modified greedy algorithm for seed selection

Input: Graph G, budget K, split (k, α)

Output: Optimal seed set S_k such that $|S_k| \leq k$

$S_k \leftarrow \phi$

for $t \leftarrow 1$ ***to*** k **do**

  for $v \notin S_k$ **do**

  $V_{valid} \leftarrow \left\{v \in V \setminus S_k : \mathbb{E}|A_{ref}(S_k \cup \{v\})| \leq \frac{K-k}{2\alpha}\right\}$

  $v_t \leftarrow \arg\max_{v \in V_{valid}} f(S_k \cup \{v\}) - f(S_k)$

  if $\{v_t\} \neq \phi$ **then**

    $S_k \leftarrow S_k \cup \{v_t\}$

  else

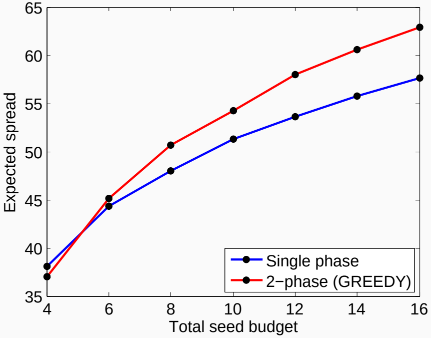
    return S_k

EXPERIMENTAL EVALUATION

- Les Miserables
  - 77 nodes, 254 undirected edges
  - Suitable for running time and memory intensive algorithms

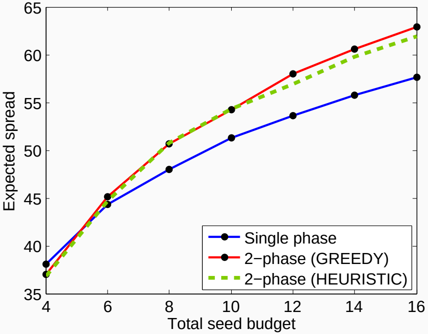
- Les Miserables
  - 77 nodes, 254 undirected edges
  - Suitable for running time and memory intensive algorithms
- NetHEPT
  - 15233 nodes, 31398 undirected edges
  - Exhibits most structural properties of "social-network" graphs

PERFORMANCE OF 2-PHASE VS. SINGLE-PHASE



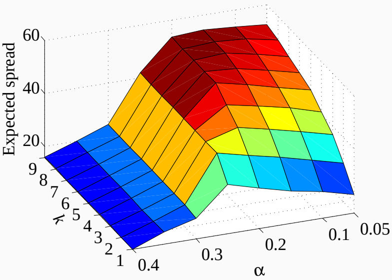
- Budget-split detrimental for small *K*, yields significant gains for moderate-high *K*, relative gain increases with *K*

PERFORMANCE OF 2-PHASE VS. SINGLE-PHASE (LES MISERABLES)

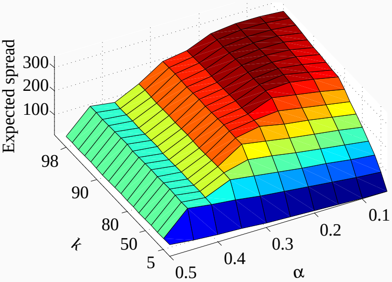


- Budget-split detrimental for small *K*, yields significant gains for moderate-high *K*
- PMIA heuristic performs nearly as well as 2-phase greedy

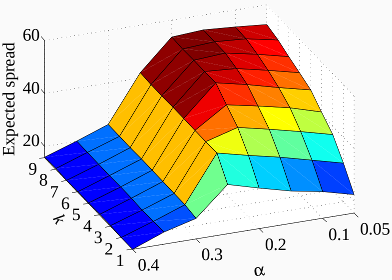
EFFECT OF k,α



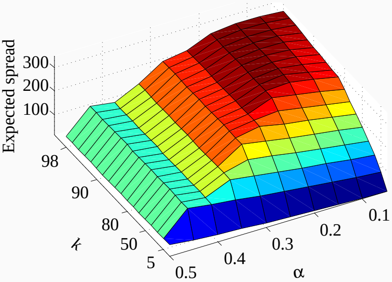
LM ($K = 10$)



NetHEPT ($K = 100$)

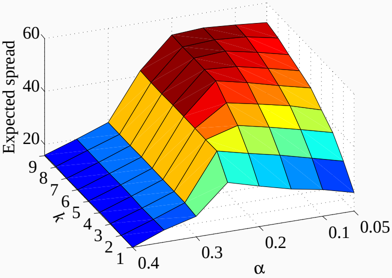


LM ($K = 10$)

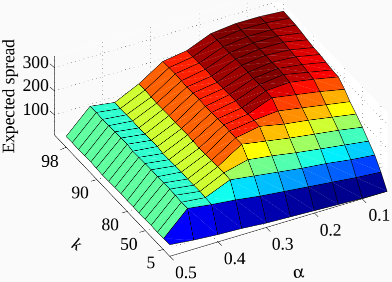


NetHEPT ($K = 100$)

- Maximum spread observed at high k, low α pairs
  Optimal split for LM : (7, 0.15). Gain ≈ 6%
  Optimal split for NetHEPT : (82, 0.15). Gain ≈ 7%



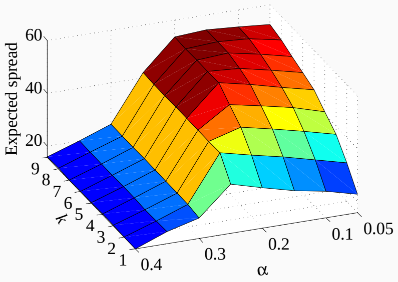
LM ($K = 10$)



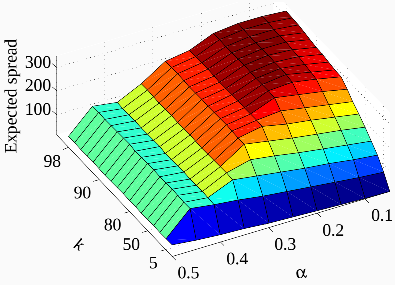
NetHEPT ($K = 100$)

- Maximum spread observed at high k, low α pairs
  Optimal split for LM : (7, 0.15). Gain $\approx$ 6%
  Optimal split for NetHEPT : (82, 0.15). Gain $\approx$ 7%
  WHY?

EFFECT OF k,α



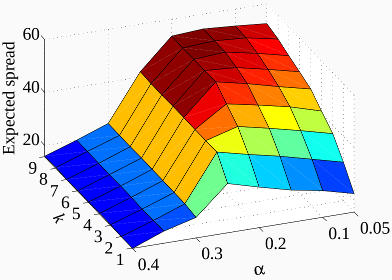
LM ($K = 10$)



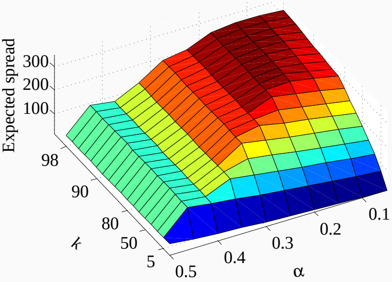
NetHEPT ($K = 100$)

- Maximum spread observed at high k, low α pairs
  Optimal split for LM : (7, 0.15). Gain $\approx$ 6%
  Optimal split for NetHEPT : (82, 0.15). Gain $\approx$ 7%
  WHY?
  Need enough active nodes after phase 1 to act as referring agents for phase 2!

EFFECT OF k,α



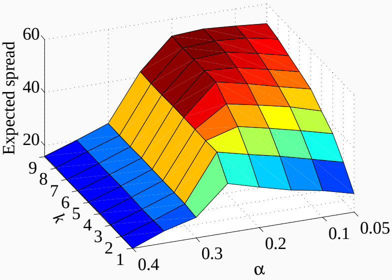
LM ($K = 10$)



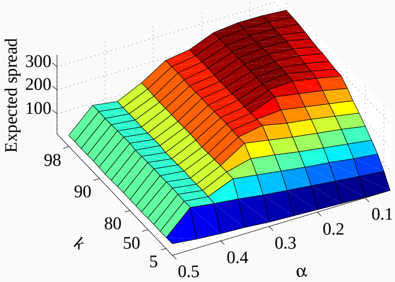
NetHEPT ($K = 100$)

- Maximum spread observed at high k, low α pairs
- Improved spread never attained at very high α

EFFECT OF k,α



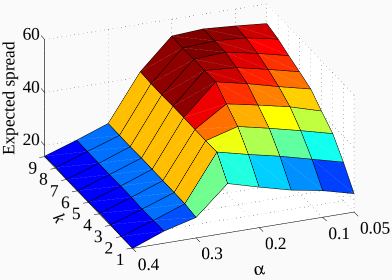
LM ($K = 10$)



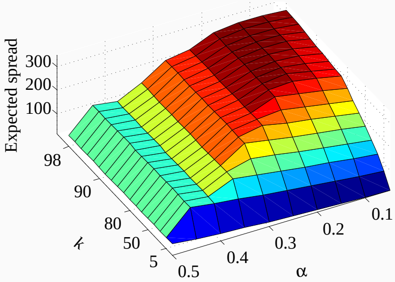
NetHEPT ($K = 100$)

- Maximum spread observed at high k, low α pairs
- Improved spread never attained at very high α
  WHY?

EFFECT OF k,α

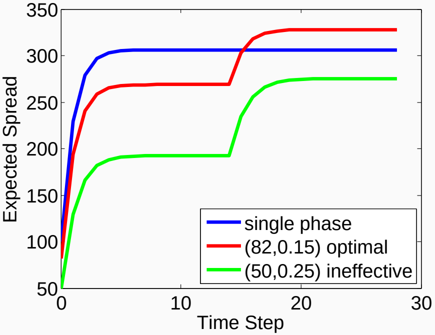


LM ($K = 10$)



NetHEPT ($K = 100$)

- Maximum spread observed at high k, low α pairs
- Improved spread never attained at very high α
  WHY?
  Higher $\alpha \implies$ fewer permissible active nodes in phase 2!



NetHEPT ($K = 100$)

- Single phase saturates earliest
- Two-phase saturates after phase 1, shoots up on initiating phase 2
- Allocating sufficient budget for phase 1 is crucial!

SUMMARY AND FUTURE WORK

- In conclusion, we have:
  - Proposed a referral incentive based model
  - Analysed the mathematical properties of said model
  - Studied efficacy of the model on real-life datasets

- In conclusion, we have:
    - Proposed a referral incentive based model
    - Analysed the mathematical properties of said model
    - Studied efficacy of the model on real-life datasets
- Future Work
    - Use real cascade data to infer an appropriate $h(\alpha)$
    - Analyse the greedy-modified algorithm; establish provable guarantee for unconstrained optimization

Thank You

PUBLICATION BASED ON THIS THESIS

- Sneha Mondal, Swapnil Dhamal, and Y. Narahari.
  Two-Phase Influence Maximization in Social Networks with Seed Nodes and Referral Incentives.

  The 10th International Workshop on Mining Actionable Insights from Social Networks (MAISoN). To appear, Cambridge, UK. ACM, 2017.