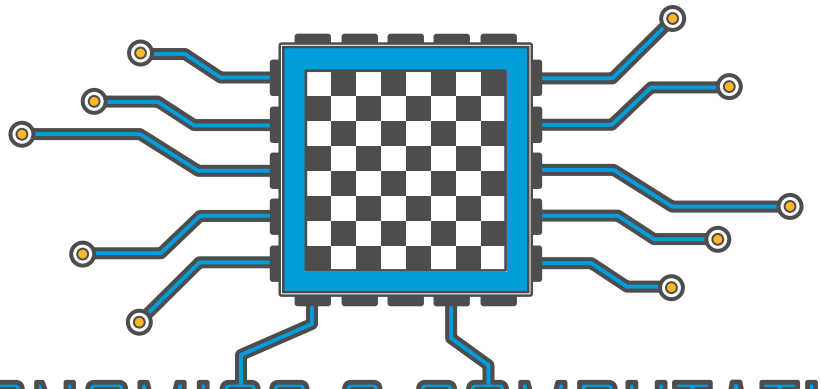


Spring 2025 | Lecture 18

Influence Maximization

Ariel Procaccia | Harvard University

MOTIVATION

- Firm is marketing a new product
- Collect data on the social network
- Choose set S of early adopters and market to them directly
- Customers in S generate a cascade of adoptions
- Question: How to choose S?

INFLUENCE FUNCTIONS

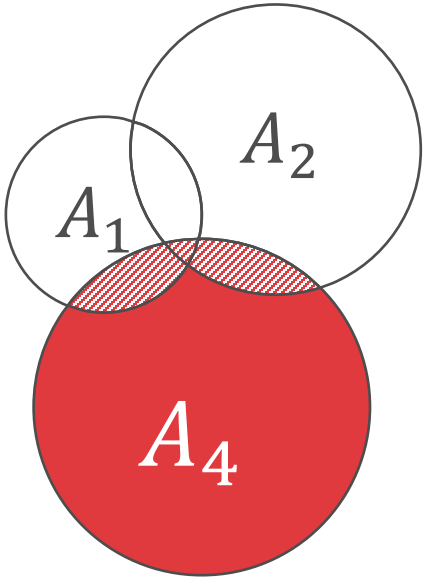
- Assume: finite directed graph, progressive process
- Fixing a cascade model, define influence function
- $f(S) =$ expected #active nodes at the end of the process starting with seed nodes S
- Maximize $f(S)$ over sets S of size k

SUBMODULARITY

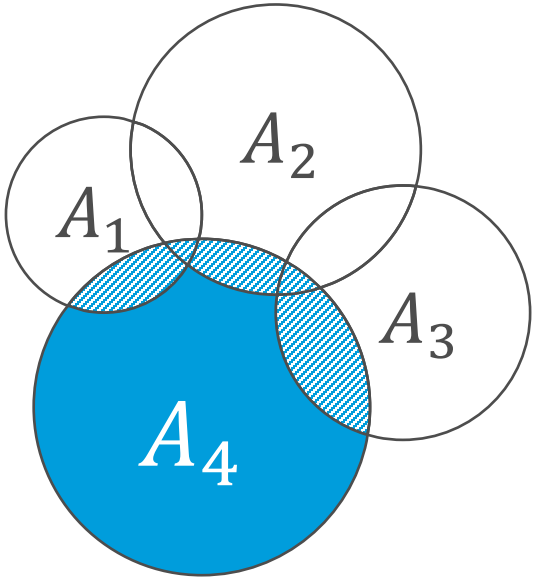
- Try to identify broad subclasses where good approximation is possible
- f is submodular if for $X \subseteq Y, z \notin Y$,
$$f(X \cup \{z\}) - f(X) \geq f(Y \cup \{z\}) - f(Y)$$
- f is monotone if for $X \subseteq Y, f(X) \leq f(Y)$
- Theorem: f monotone and submodular, $S^\star$ optimal k-element subset, S obtained by greedily adding k elements that maximize marginal increase; then
$$f(S) \geq \left(1 - \frac{1}{e}\right) f(S^\star)$$

EXAMPLE: COVERAGE FUNCTIONS

- Let $U, A_1, \dots, A_n \subset U$, and $f: 2^{[n]} \to \mathbb{R}^+$
- The coverage function is $f(S) = |\cup_{i \in S} A_i|$
- This function is monotone submodular



$f(\{1,2\} \cup \{4\}) - f(\{1,2\})$



$f(\{1,2,3\} \cup \{4\}) - f(\{1,2,3\})$

EXAMPLE: COVERAGE FUNCTIONS

- Let $U, A_1, \ldots, A_n \subset U$, and $f: 2^{[n]} \to \mathbb{R}^+$
- The coverage function is $f(S) = |\cup_{i \in S} A_i|$
- This function is monotone submodular
- Consider two more functions:
  - $f_1(S) = \mathbb{I}_{1 \in S} \cdot |\cup_{i \in S} A_i|$
  - $f_2(S) = \mathbb{I}_{1 \in S} \cdot |A_1| + |\cup_{i \in S} A_i|$

Poll 1

Which function is monotone submodular?

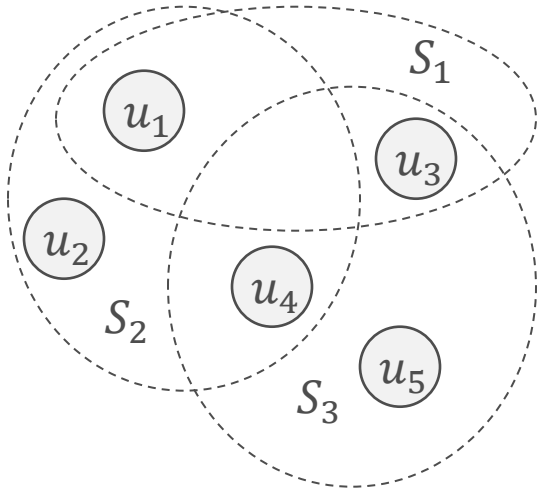
- ○ Only f_1
- ○ Only f_2
- ○ Both functions
- ○ Neither one

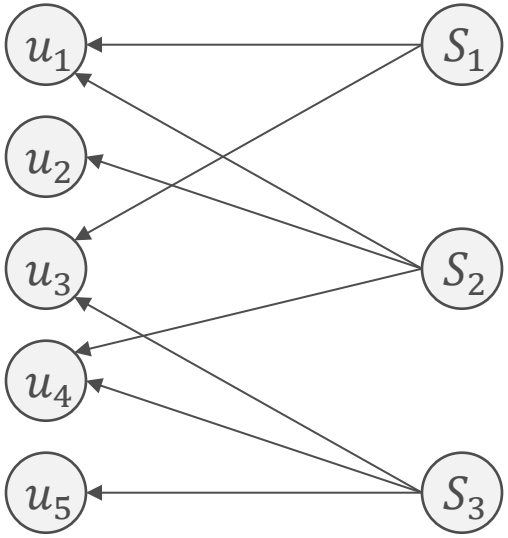
INDEPENDENT CASCADE MODEL

- Assume $\forall (i,j) \in E, w_{ij} \in [0,1]$
- For convenience, for $(i,j) \notin E$, let $w_{ij} = 0$
- When $\exists (i,j) \in E$ s.t. i is active and j is not, i has one chance to activate j with prob. w_{ij}
- Theorem: Under the independent cascade model:
  - Influence maximization is NP-hard
  - The influence function f is submodular

PROOF OF NP-HARDNESS

- SET COVER: subsets $S_1, \dots, S_m$ of $U = \{u_1, \dots, u_t\}$; cover of size k?
- Bipartite graph: $u_1, \dots, u_t$ on one side, $S_1, \dots, S_m$ on the other
- If $u_i \in S_j$ then there is an edge (S_j, u_i) with weight 1
- Min SC of size $k \Rightarrow t + k$ active
- Min SC of size $> k \Rightarrow$ less than $t + k$ active ■





PROOF OF SUBMODULARITY

- Lemma: If $f_1, \dots, f_r$ are submodular functions, $c_1, \dots, c_r \geq 0$, then $f = \sum_{i=1}^{r} c_i f_i$ is a submodular function
- Proof: Let $X \subseteq Y$ and $z \notin Y$, then

$$f(X \cup \{z\}) - f(X) - \big(f(Y \cup \{z\}) - f(Y)\big) =$$

$$\sum_{i=1}^{r} c_i [f_i(X \cup \{z\}) - f_i(X) - (f_i(Y \cup \{z\}) - f_i(Y))]$$

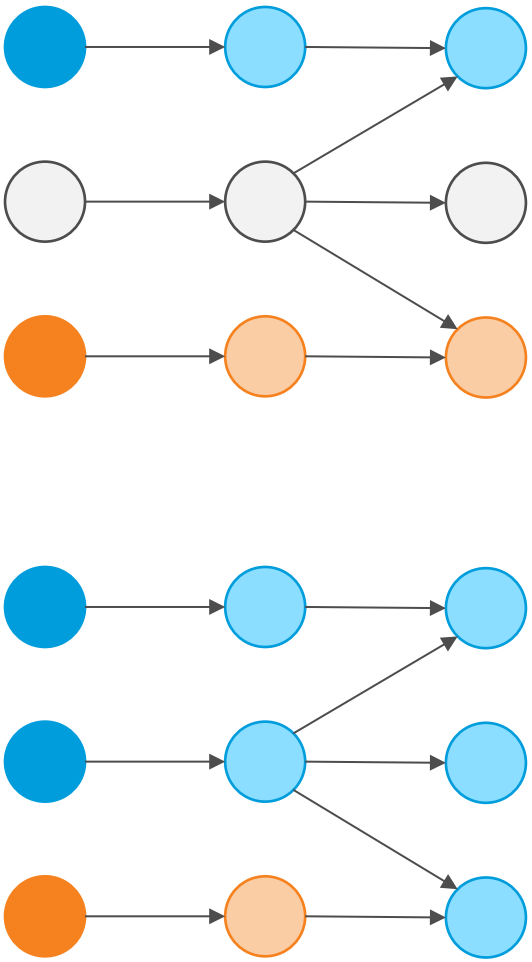
$$\geq 0$$

PROOF OF SUBMODULARITY

- Key idea: for each $(i, j) \in E$ we flip a coin of bias w_{ij} in advance
- Let α denote a particular one of the $2^{|E|}$ possible coin flip combinations
- $f_\alpha(S) =$ activated players with S as seeds and α coin flips
- $i \in f_\alpha(S)$ iff i is reachable from S via live edges

PROOF OF SUBMODULARITY

- f_α is submodular: it's like a coverage function where each seed node is associated with all reachable nodes
- $f(S) = \sum_\alpha \Pr[\alpha] \cdot f_\alpha(S)$, that is, f is a nonnegative weighted sum of submodular functions
- By the lemma, f is submodular ■



LINEAR THRESHOLD MODEL

- Assume $\forall j \in N, \sum_i w_{ij} \leq 1$
- Each $j \in N$ has threshold $\theta_j \in [0,1]$ chosen uniformly at random
- j becomes active if

$$\sum_{\text{active } i} w_{ij} \geq \theta_j$$

LINEAR THRESHOLD MODEL

Poll 2

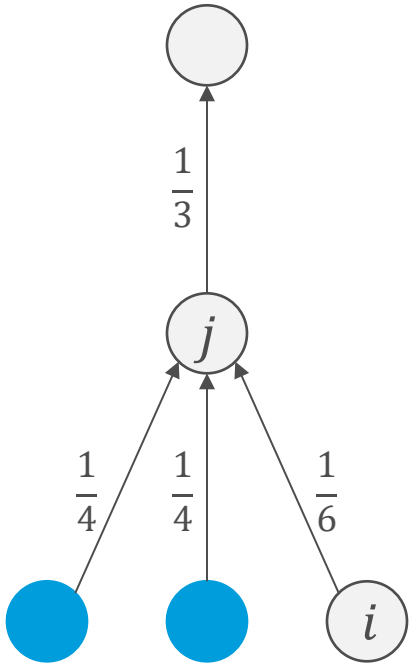
What is $f(S)$?

- ○ 2/3
- ○ 8/3
- ○ 5/2
- ○ 13/4

Poll 3

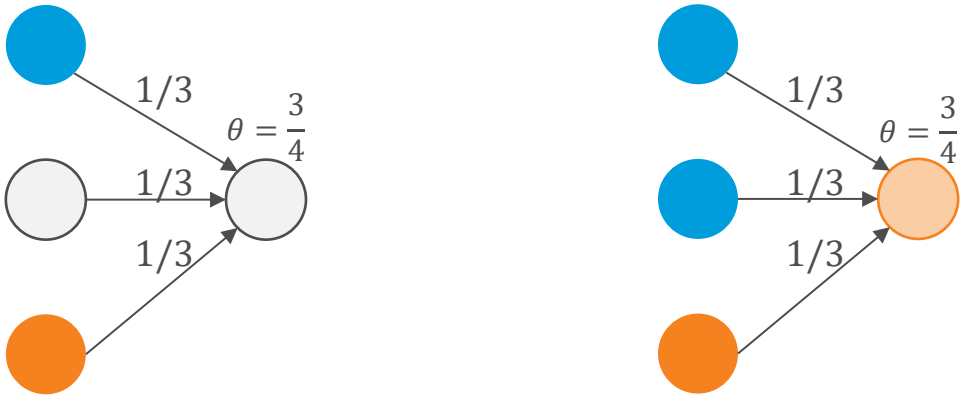
Given that j is inactive, probability it becomes active after i does?

- ○ 1/3
- ○ 2/3
- ○ 1/2
- ○ 1



LINEAR THRESHOLD MODEL

- Theorem: Under the linear threshold model:
  - Influence maximization is NP-hard
  - The influence function f is submodular
- Difficulty: fixing the coin flips α, f_α is not submodular



PROOF OF SUBMODULARITY

- Each j chooses at most one of its incoming edges at random; (i, j) selected with prob. w_{ij}, and none with prob. $1 - \sum_i w_{ij}$
- If we can show that these choices of live edges induce the same influence function as the linear threshold model, then the theorem follows from the same arguments as before

PROOF OF SUBMODULARITY

- We sketch the equivalence of the two models
- Linear threshold:
  - A_t = active players at end of iteration t
  - $\Pr[j \in A_{t+1} \mid j \notin A_t] = \frac{\sum_{i \in A_t \setminus A_{t-1}} w_{ij}}{1 - \sum_{i \in A_{t-1}} w_{ij}}$
- Live edges:
  - At every times step, determine whether j's live edge comes from current active set
  - If not, the source of the live edge remains unknown, subject to being outside the active set
  - Same probability as before ■

APPLICATIONS

RESEARCH ARTICLE SUMMARY

The Diffusion of Microfinance

Abhijit Banerjee,* Arun G. Chandrasekhar,* Esther Duflo,* Matthew O. Jackson*

READ THE FULL ARTICLE ONLINE
http://dx.doi.org/10.1126/science.1236498

Cite this article as A. Banerjee *et al.*, *Science* **341**, 1236498 (2013). DOI: 10.1126/science.1236498

Introduction: How do the network positions of the first individuals in a society to receive information about a new product affect its eventual diffusion? To answer this question, we develop a model of information diffusion through a social network that discriminates between information passing (individuals must be aware of the product before they can adopt it, and they can learn from their friends) and endorsement (the decisions of informed individuals to adopt the product might be influenced by their friends' decisions). We apply it to the diffusion of microfinance loans, in a setting where the set of potentially first-informed individuals is known. We then propose two new measures of how "central" individuals are in their social network with regard to spreading information; the centrality of the first-informed individuals in a village helps significantly in predicting eventual adoption.

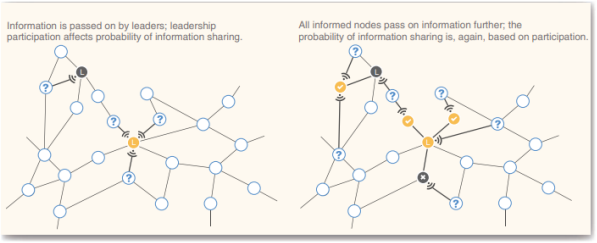
Methods: Six months before a microfinance institution entered 43 villages in India and began offering microfinance loans to villagers, we collected detailed network data by surveying households about a wide range of interactions. The microfinance institution began by inviting "leaders" (e.g., teachers, shopkeepers, savings group leaders) to an informational meeting and then asked them to spread information about the loans. Using the network data, the locations in the network of these first-informed villagers (or injection points), and data regarding the villagers' subsequent participation, we estimate the parameters of our diffusion model using the method of simulated moments. The parameters of the model are validated by showing that the model correctly predicts the evolution of participation in each village over time. The model yields a new measure of the effectiveness of any given node as an injection point, which we call communication centrality. Finally, we develop an easily computed proxy for communication centrality, which we call diffusion centrality.

Results: We find that a microfinance participant is seven times as likely to inform another household as a nonparticipant; nonetheless, information transmitted by nonparticipants is important and accounts for about one-third of the eventual informedness and participation in the village because nonparticipants are much more numerous. Once information passing is accounted for, an informed household's decision to participate is not significantly dependent on how many of its neighbors have participated. Communication centrality, when applied to the set of first-informed individuals in a village, substantially outperforms other standard network measures of centrality in predicting microfinance participation in this context. Finally, the simpler proxy measure—diffusion centrality—is strongly correlated with communication centrality and inherits its predictive properties.

Discussion: Our results suggest that a model of diffusion can distinguish information passing from endorsement effects, and that understanding the nature of transmission may be important in identifying the ideal places to inject information.

SUPPLEMENTARY MATERIALS

Materials and Methods
Supplementary Text
Tables S1 to S7
References



Diffusion of information and participation. (Left) First-informed households have decided whether to participate and stochastically pass on information to their neighbors. **(Right)** Participation may affect the probability of passing information. Newly informed nodes make their decisions, possibly being influenced by the decisions of their neighbors. After newly informed nodes make their participation decisions, all informed nodes engage in another round of stochastic communication.

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: banerjee@mit.edu (A.B.); arungc@stanford.edu (A.G.C.); eduflo@mit.edu (E.D.); jacksonm@stanford.edu (M.O.J.)

Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)

Bridging the Gap Between Theory and Practice in Influence Maximization: Raising Awareness about HIV among Homeless Youth

Amulya Yadav, Bryan Wilder, Eric Rice, Robin Petering, Jaih Craddock, Amanda Yoshioka-Maxwell, Mary Hemler, Laura Onasch-Vera, Milind Tambe, Darlene Woo
Center for Artificial Intelligence in Society, University of Southern California, LA, CA, 90089
{amulyaya, bwilder, ericr, petering, jaih.craddock, abarron, hemler, onaschve, tambe, darlenew}@usc.edu

Abstract

This paper reports on results obtained by deploying HEALER and DOSIM (two AI agents for social influence maximization) in the real-world, which assist service providers in maximizing HIV awareness in real-world homeless-youth social networks. These agents recommend key "seed" nodes in social networks, i.e., homeless youth who would maximize HIV awareness in their real-world social network. While prior research on these agents published promising simulation results from the lab, the usability of these AI agents in the real-world was unknown. This paper presents results from three real-world pilot studies involving 173 homeless youth across two different homeless shelters in Los Angeles. The results from these pilot studies illustrate that HEALER and DOSIM outperform the current modus operandi of service providers by ∼160% in terms of information spread about HIV among homeless youth.

1 Introduction

The nearly two million homeless youth in the United States [Toro *et al.*, 2007] are at high risk of contracting Human Immunodeficiency Virus (HIV) [Pfeifer and Oliver, 1997]. In fact, homeless youth are twenty times more likely to be HIV positive than stably housed youth, due to high-risk behaviors that they engage in (such as unprotected sex, exchange sex, sharing drug needles, etc.) [CDC, 2013; Council, 2012]. Given the important role that peers play in these high-risk behaviors of homeless youth [Rice *et al.*, 2012a; Green *et al.*, 2013], it has been suggested that *peer leader based interventions* for HIV prevention be developed for these youth [Arnold and Rotheram-Borus, 2009; Rice *et al.*, 2012a; Green *et al.*, 2013].

As a result, many homeless youth service providers (henceforth just "service providers") conduct peer-leader based social network interventions [Rice, 2010], where a select group of homeless youth are trained as *peer leaders*. This peer-led approach is particularly desirable because service providers have limited resources and homeless youth tend to distrust adults. The training program of these *peer leaders* includes detailed information about how HIV spreads and what one can do to prevent infection. The *peer leaders* are also taught effective ways of communicating this information to their peers [Rice *et al.*, 2012b].

Because of their limited financial and human resources, service providers can only train a small number of these youth and not the entire population. Thus, the selected peer leaders in these interventions are tasked with spreading messages about HIV prevention to their peers in their social circles, thereby encouraging them to adopt safer practices. Using these interventions, service providers aim to leverage social network effects to spread information about HIV, and induce behavior change (increased HIV testing) among people in the homeless youth social network.

In fact, there are further constraints that service providers face – behavioral struggles of homeless youth means that service providers can only train 3-4 peer leaders in every intervention. This leads us to do sequential training; where groups of 3-4 homeless youth are called one after another for training. They are trained as peer leaders in the intervention, and are asked *information* about friendships that they observe in the real-world social network. This newer *information* about the social network is then used to improve the selection of the peer leaders for the next intervention. As a result, the peer leaders for these limited interventions need to be chosen strategically so that awareness spread about HIV is maximized in the social network of homeless youth.

Previous work proposed HEALER [Yadav *et al.*, 2016] and DOSIM [Wilder *et al.*, 2017], two agents which assist service providers in optimizing their intervention strategies. These agents recommend "good" intervention attendees, i.e., homeless youth who maximize HIV awareness in the real-world social network of youth. In essence, both HEALER and DOSIM reason strategically about the multiagent system of homeless youth to select a sequence of 3-4 youth at a time to maximize HIV awareness. Unfortunately, while earlier research [Yadav *et al.*, 2016; Wilder *et al.*, 2017] published promising simulation results from the lab, neither of these agent based systems have ever been tested in the real world.

[1] Amulya Yadav (amulyaya@usc.edu) is the contact author