INFLUENCE IN NETWORKS

M. Vazirgiannis
École Polytechnique
January 2021

Outline

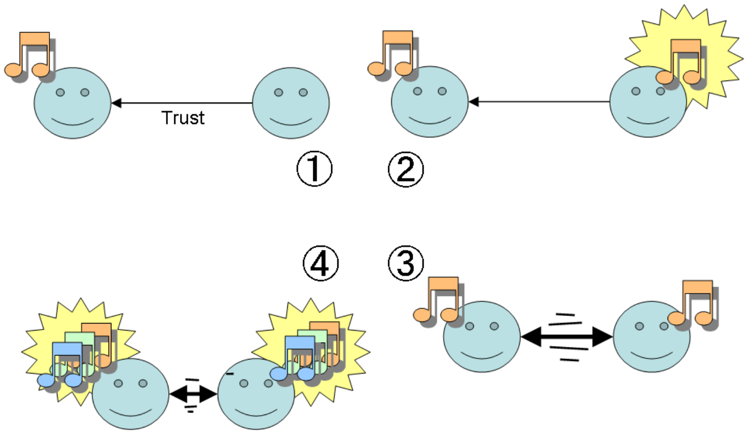
- Introduction
- Identifying Influential Spreaders
- Simulating Information Diffusion
- Influence Maximization

Outline

- Introduction
- Identifying Influential Spreaders
- Simulating Information Diffusion
- Influence Maximization

Influence and Information Diffusion

Social Influence



[Matsuo et al. WWW'09]

Example: Large Scale Behavioral Change

- Can someone use online social networks to shape a massive behavioral change?
- 61M users on Facebook [RM. Bond et al, 2012]

- Users split in three randomized groups
- A simple post was used to motivate voting
- Users could claim they voted by pressing “I voted” and access further information about the elections

Example: Large Scale Behavioral Change

- Simple message :the message included the count of FB users who reported they voted.

- Social message : The message included a list of the users’ friends that have also voted.

- Control group : no message.

Example: Large Scale Behavioral Change

- Validated real world impact using public voting records
- Users who received the social message were more likely to click “I voted” then users with the simple message
- The “closer” the friend, the higher the difference in voting
- Friends of users who received the social message were more likely to vote then friends of users with no message

Influence and Information Diffusion

Applications:

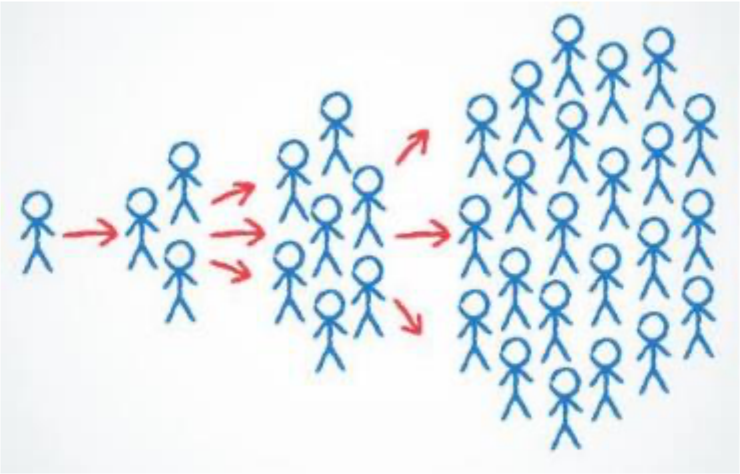
- Viral Marketing
- Political Campaigns
- News and rumor spreading
- Opinion shaping
- Epidemiology and virus propagation
- “Friends” recommendation

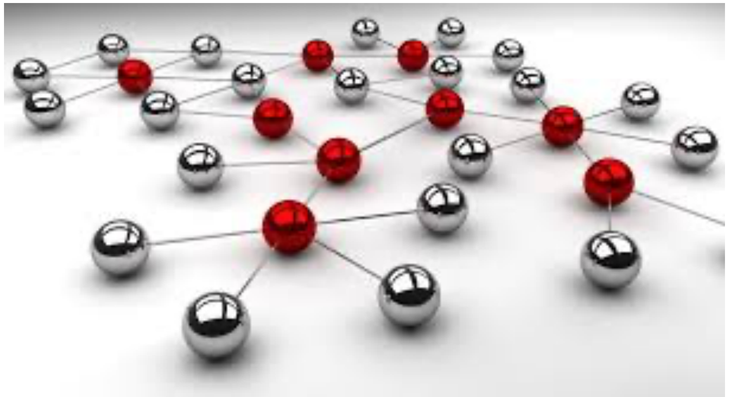
Outline

- Introduction
- Identifying Influential Spreaders
- Simulating Information Diffusion
- Influence Maximization

Identifying Influential Spreaders

- Locate “influencers” that broadcast information in a network
- Each of them can effect a substantial amount of other nodes
  - The effect can extend far from the immediate neighbors
  - We need methods that take into account higher order connectivity

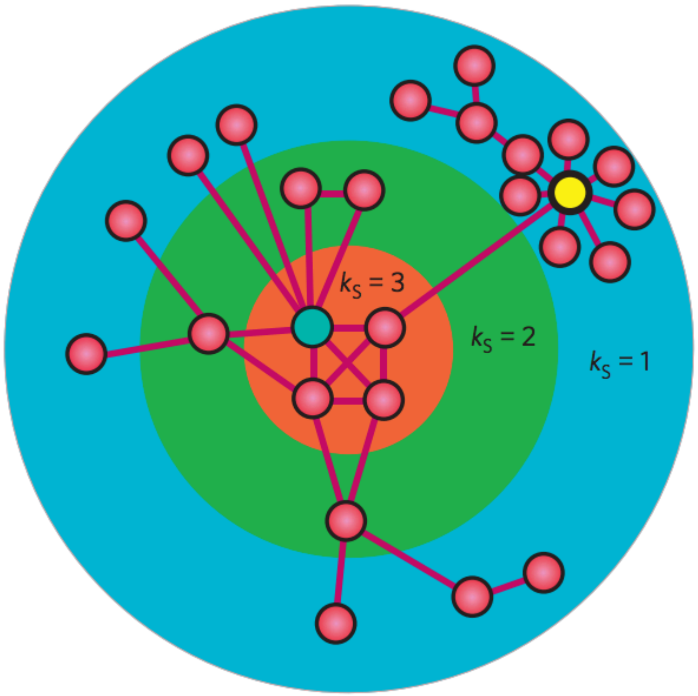




Identifying Influential Spreaders

- Degree
- Indegree
- Pagerank
- Closeness

- What if the node is well connected but in the periphery of the network?
  - Use K-core decomposition



Outline

- Introduction
- Identifying Influential Spreaders
- Simulating Information Diffusion
- Influence Maximization

Diffusion Models

- Stochastic processes run over a directed graph where each node can be activated or not
- Once activated, a node cannot deactivate (does not apply to the epidemiological models)
- The diffusion proceeds iteratively in a synchronous way along a discrete time axis starting from a set of initially activated nodes

Linear Threshold

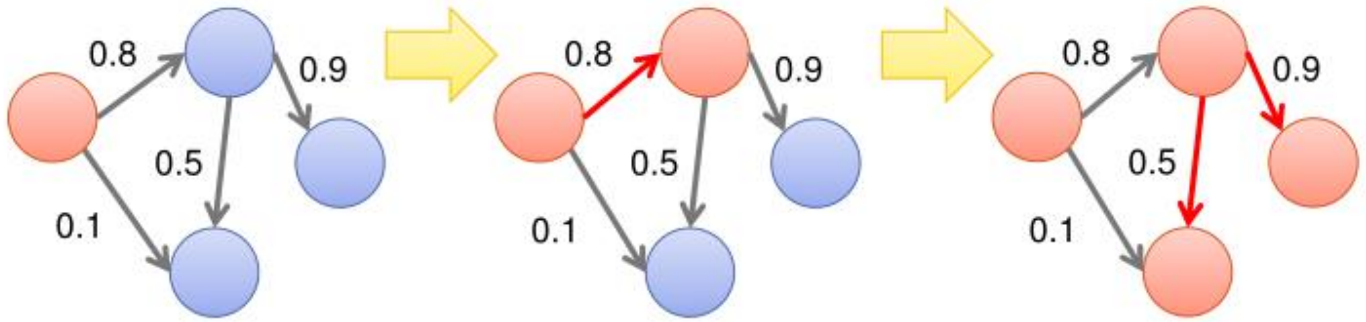
- Every edge $e=(u,v)$ has a weight $w_{u,v}$
- $\Sigma_u w_{u,v} \leq 1$
- Each node v has a threshold $\Theta_v \in [0,1]$
- At each step, for each node v we check

$$\sum_{u \in N(v)} w_u > \theta_v$$

and if so, it is activated

Independent Cascade

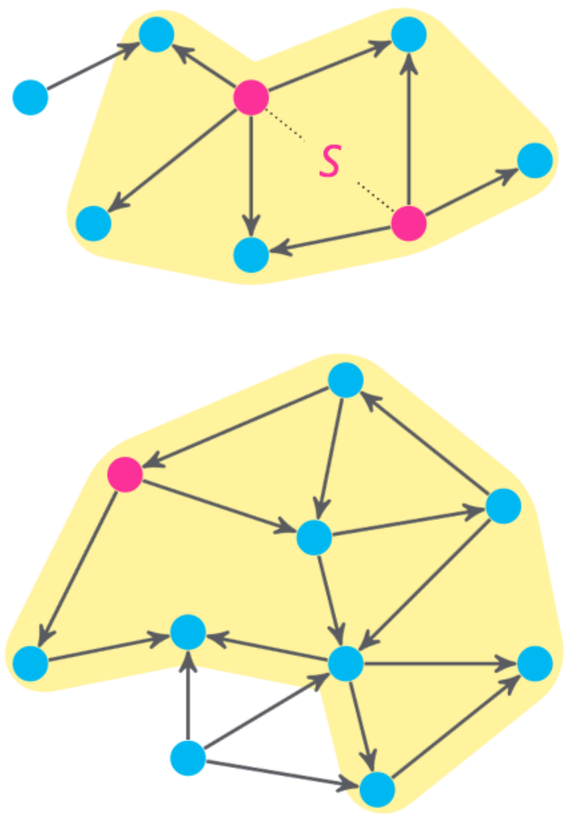
- Every edge $e=(u,v)$ has a propagation probability $p_{u,v}$
- At each step t, all the nodes u activated at $t\text{-}1$ activate their neighbor v with a probability $p_{u,v}$
- Each node has only <u>one chance</u> to activate its neighbor



David Kempe, Jon Kleinberg, and Éva Tardos. "Maximizing the spread of influence through a social network.” ACM SIGKDD,2003.

Influence Spread

- Influence Spread of a seed set: σ(S) = The expected number of nodes activated at the end of a diffusion process started from the nodes in S

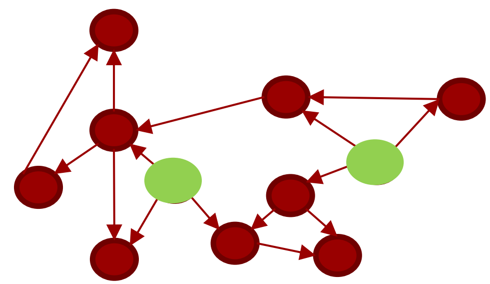


Outline

- Introduction
- Identifying Influential Spreaders
- Simulating Information Diffusion
- Influence Maximization

Influence Maximization

- Given
  - A directed network G(V,E)
  - A diffusion model (IC/LT)



- Find
  - S: A set of *K* nodes that would maximize spreading if diffusion started from them : $S^* = argmax_{|S| \leq k}\ \sigma(S)$
  - σ(S): Estimated total spreading

- The diffusion models have randomly assigned parameters

NP-Complete ☹

Monotonicity and Submodularity

Definition (Monotonicity)

The set function f is **monotone** if adding an element to a set cannot cause f to decrease :

$$\forall v \in V,\ \forall A \subseteq V,\ f(A \cup \{v\}) \geq f(A)$$

Definition (Submodularity)

The set function f is **submodular** if, for any $A \subseteq B \subseteq V$ and for any $v \in V \setminus B$, we have :

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B)$$

Greedy

- Iteratively, add in the seed set the node with the best marginal gain [Kempe D., ,2003]

Algorithm 1 Greedy

Input: G, k, σ_m

Output: seed set S

1: $S \leftarrow \emptyset$
2: **while** $|S| < k$ **do**
3: select $u = \arg\max_{w \in V \setminus S}(\sigma_m(S \cup \{w\}) - \sigma_m(S))$
4: $S \leftarrow S \cup \{u\}$

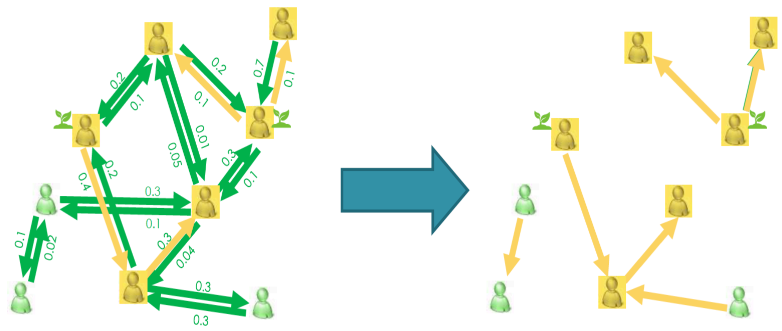
- $\left(1 - \frac{1}{e}\right)$ theoretical guarantee
- Estimating the expected influence spread in each round is #P-hard
  - Because you simulate a diffusion process for each candidate seed
  - Use the live-edge model and Monte Carlo estimation

Live Edge Model

- The influence spread under the live edge model can be computed as:

$$\sigma(S) = \sum_{g} P(g)\sigma_g(S)$$

- During each simulation, a random realization of the network g is sampled and the influence spread is defined as the nodes that lie in a path that is connected to one of the seeds



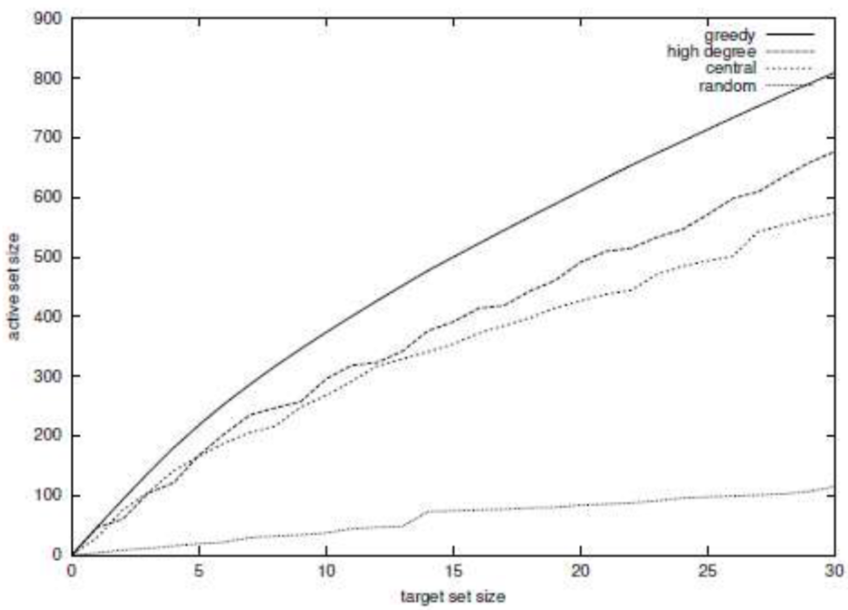
- The probability weighs the influence spread to indicate the possibility of this type of spread happening for this realization and this seed set, and is

$$P(g) = \prod_{e \in g} P_e \prod_{e\prime \in G/g} (1 - P_{e'})$$

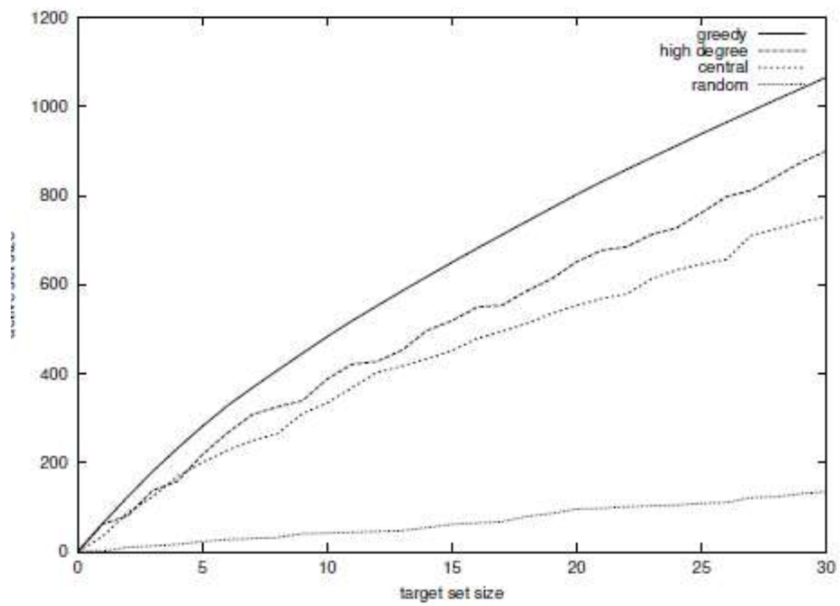
Performance of Greedy Algorithm

- Compare greedy with top nodes based on centralities
- Evaluate centralities seeds by taking the average influence spread from 10000 simulations of LT and IC

weighted cascade model



linear threshold model



References

1. Kempe, D., Kleinberg, J., & Tardos, É. (2003, August). Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 137-146). ACM.

2. Newman, M., Barabasi, A. L., & Watts, D. J. (2006). The structure and dynamics of networks. Princeton University Press.

3. Maksim Kitsak, Lazaros Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Herman Makse. Identication of influential spreaders in complex networks. Nature Physics, 6(11):888{893, Aug 2010.

4. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., & Glance, N. (2007, August). Cost-effective outbreak detection in networks. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 420-429). ACM.

5. Chen, W., Wang, C., & Wang, Y. (2010, July). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1029-1038). ACM.

References

6. Goyal, A., Lu, W., & Lakshmanan, L. V. (2011, December). Simpath: An efficient algorithm for influence maximization under the linear threshold model. In Data Mining (ICDM), 2011 IEEE 11th International Conference on (pp. 211-220). IEEE.

7. Chen, D. B., Gao, H., Lü, L., & Zhou, T. (2013). Identifying influential nodes in large-scale directed networks: the role of clustering.

8. Bond, Robert M., et al. "A 61-million-person experiment in social influence and political mobilization." Nature 489.7415 (2012): 295.

9. Goyal, Amit, Francesco Bonchi, and Laks VS Lakshmanan. "Learning influence probabilities in social networks." Proceedings of the third ACM international conference on Web search and data mining. ACM, 2010.

10. Rodriguez, Manuel Gomez, David Balduzzi, and Bernhard Schölkopf. "Uncovering the temporal dynamics of diffusion networks." arXiv preprint arXiv:1105.0697 (2011).

References

11. Tang, Youze, Yanchen Shi, and Xiaokui Xiao. "Influence maximization in near-linear time: A martingale approach." Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. ACM, 2015.

12. Feng, S., Cong, G., Khan, A., Li, X., Liu, Y., & Chee, Y. M. (2018). Inf2vec: Latent Representation Model for Social Influence Embedding. ICDE.

13. Qiu, Jiezhong, et al. "DeepInf: Social Influence Prediction with Deep Learning." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018.

14. Du, Nan, et al. "Recurrent marked temporal point processes: Embedding event history to vector." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.

15. Li, C., Ma, J., Guo, X., & Mei, Q. (2017, April). DeepCas: An end-to-end predictor of information cascades. In Proceedings of the 26th International Conference on World Wide Web (pp. 577-586).