"Information Diffusion Analysis in Social Networks with Influence Measuring and Maximizing"初稿 5140309118 贾煜

1. Motivation:

A social network is a social structure made up of a set of social actors (such as individuals or organizations), sets of dyadic ties, and other social interactions between actors. The biggest social networks are those formed by social networking sites and apps, which become more and more indispensable for us people.

It's true that nearly everyone is covered in multiple specific social networks, even he/she is not aware of the fact. Online social networking technologies enable individuals to share information with any number of peers, which means the start of information propagating. Numerous examples have proved the strength of quick and far-ranging information diffusion. Thus we are interested in how to quantify the influence of information diffusion and how to maximize it.

2. Problem Modulation:

The social network is represented by a directed graph G, and we write u → v to denote the existence of a directed edge from u to v, which means the relationship.

Each node v is influenced by some neighbor w based on the weight $b_{vw} \in [0,1]$. Node v has a threshold θ_v. Each currently inactive node v becomes active if and only if the total weight of its active neighbors is at least θ_v. In addition, thresholds θ_v are chosen independently and uniformly at random from the interval $[0,1]$. $\mathrm{A_t}$ denote the set of nodes active at time t. S is the set of neighbors of v that are active.

Threshold function (activation function) $\mathrm{f_v}(S)$ satisfies:

$$\mathrm{f_v}(\mathrm{S}) = \min(1, \sum\nolimits_{\mathrm{u \in S}} \mathrm{b_{vu}})$$

We can prove $\mathrm{f_v}(\mathrm{S})$is submodular. The marginal gain from adding an element to a set S is at least as high as the marginal gain form adding the same element to a superset of S.

$$\mathrm{f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)}$$

Assuming each node v has a weight ω_{v}. The target function (weighted influence function) is

$$\sigma_{\mathrm{w}}(\mathrm{A}) = \mathrm{E}[\sum_{\mathrm{v \in A_t}} \omega_{\mathrm{v}}] = \sum_{\mathrm{v \in A_t}} \omega_{\mathrm{v}} * \mathrm{p(v\ active)}$$

3. Algorithm Analysis

It can be proved this is a NP-hard problem.

(1) Greedy hill-climbing algorithm.

First forming an empty set as A, then repeatedly adds to A the node x to maximize the marginal gain: $\sigma(\mathrm{A} \cup \{\mathrm{x}\}) - \sigma(\mathrm{A})$.

(2) Community detection.

Modularity Q is:

$$\mathrm{Q} = \frac{1}{2m}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})\delta(i,j)$$

4. *Experiments*
5. *Conclusion*
6. *Future Work*