Influence Maximum Problem

Name: 周朦

ID: 11510695

1. Preliminaries

Python Version:2.7

Software: Pycharm2.3

Algorithm to find seed nodes: greedy, CELF

Model to calculate diffusion influence: LT model ,IC model

2. Methodology

First, to begin the project, i use 2-D adjacency list to represent the graph, then i need to use two models to calculate the influence of the seeds that has been chosen as well as the best seeds that two models give.

In LT model, the data structure i use is “in_table”, which means a dict that which nodes point to the node. You can think of it as the inverse list of the in_degree.

In_degree and out_degree respectively means the in degree and the out degree of the nodes.

Then through two models, i can calculate the number of nodes that one node can influence.

The greedy algorithm can act as an iteration to calculate the total influence nodes number. But it is extremely slow, though it can get the best solution.So i introduce CELF to you.

I run IMP when k = 2,3,4, and you can find the comparision of two algorithm. For convinence , i only use IC model and explain it in **part 3.**

Also if you use the same seed set as an input to LT and IC, the conclusion is the influence calculated from LT model is always larger than IC model. **This is because LT model use random_list to record every nodes, and the seeds which is activated can be used to calculate total weight many times, but IC model can only let activated node activate other nodes for once.**

3. Empirical Verification

Describe the experiments that you conducted to test/verify the quality of your

program. This may include (but not limit to) the following:

- How were the experiments designed?

  I divide it into two parts as the requirement , the first part is IMP and second part is ISE

- What data did you use?

  Direct_dict: is the representation of the graph.

  In_degree: the in degree of each node

  Out_degree: the out degree of each node

  In_table: inverse of the in_degree

  Etc....

- How did you measure the performance?

运行IMP时间（ICmodel）+CELF	运行IMP时间（ICmodel）+贪心	ISE(IC model)	种子集合	种子集合数量
9.161000013	45.23900008	19.6815	56,58	2
11.08100009	97.13000011	24.0863	56,58,53	3
13.47000003	159.6919999	27.7605	56,58,53,48	4

  We can find that **CELF is much faster than greedy**. What's more, the accuracy is the same.

- Experimental results

```
[56, 58, 53, 48]    [56, 58, 53, 48]
27.6921             32.8143
```

- Did the results meet your expectation about the program? Why or why not?

  Maybe not, because the LT model is less faster than IC model, i find that it is because the time complexity of LT model is **$O(n^3)$,** compare to IC model, which is only **$O(n^2)$,** is slower. So maybe i can use in_table to optimize it, but something went wrong, but i still can't find why. So i have to use former algorithm to get the correct answer.

4. References

List the references, please follow the IEEE format to prepare your references. The IEEE format can be found at:

```
参考文献:http://www.doc88.com/p-3867998343720.html
```