

# Mining in Large Networks

## Assignment 3

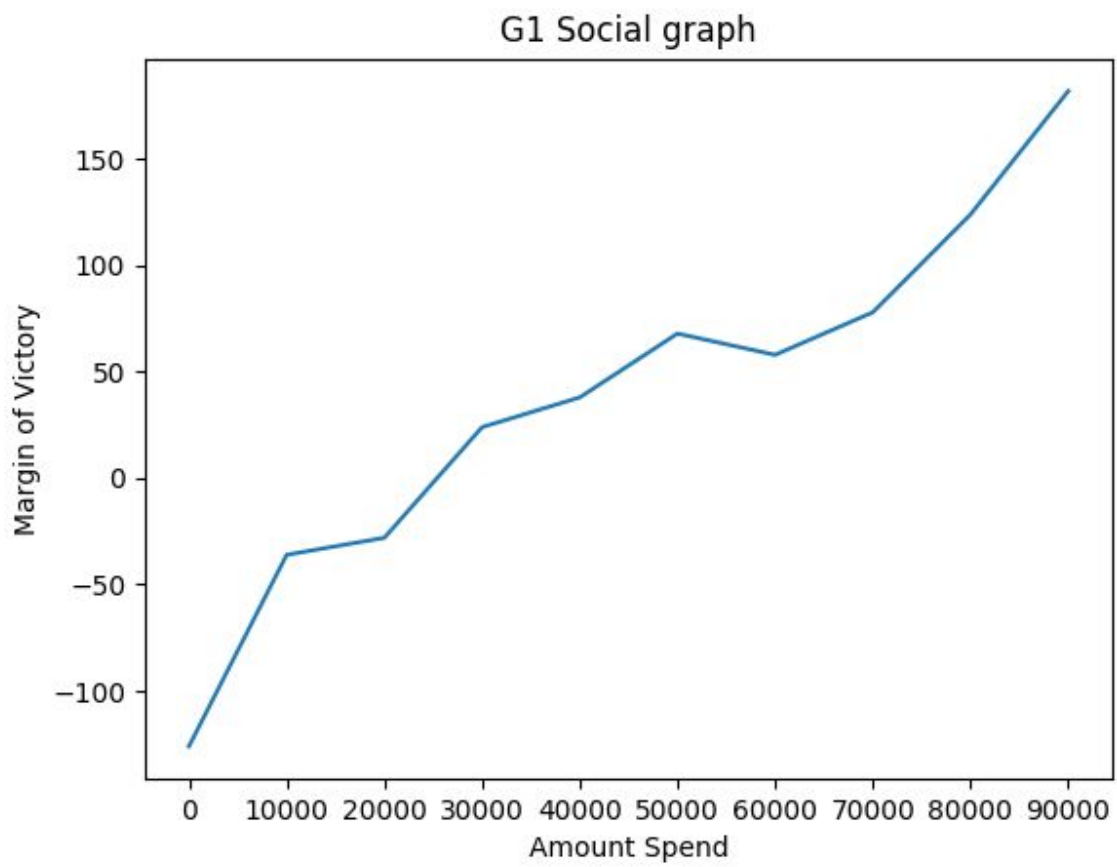
- Nirav Diwan (2017072)

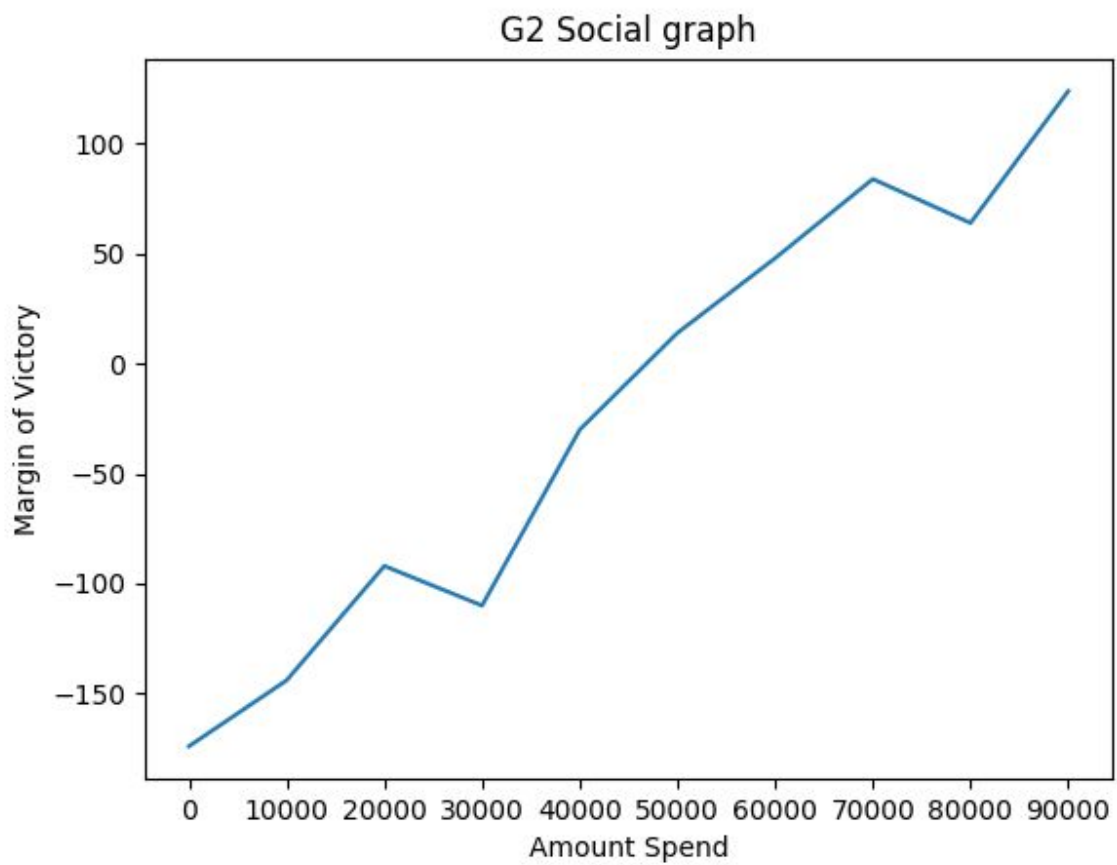
### Q1.a)

For graph G1, The election was won by **B** with a margin of 126

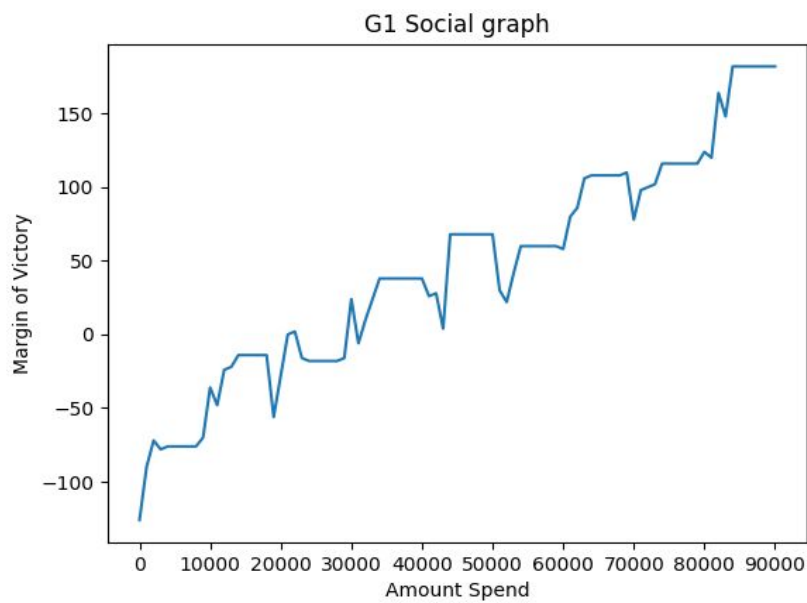
For graph G2, The election was won by **B** with a margin of 174

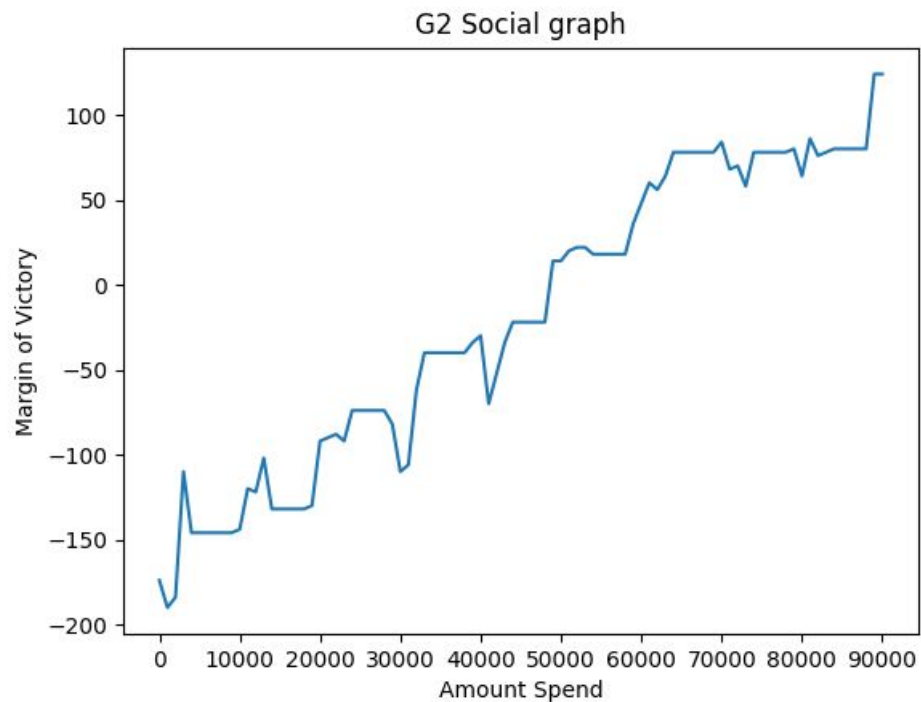
**Q1 b)**





**At the gap of 1k interesting results.....**





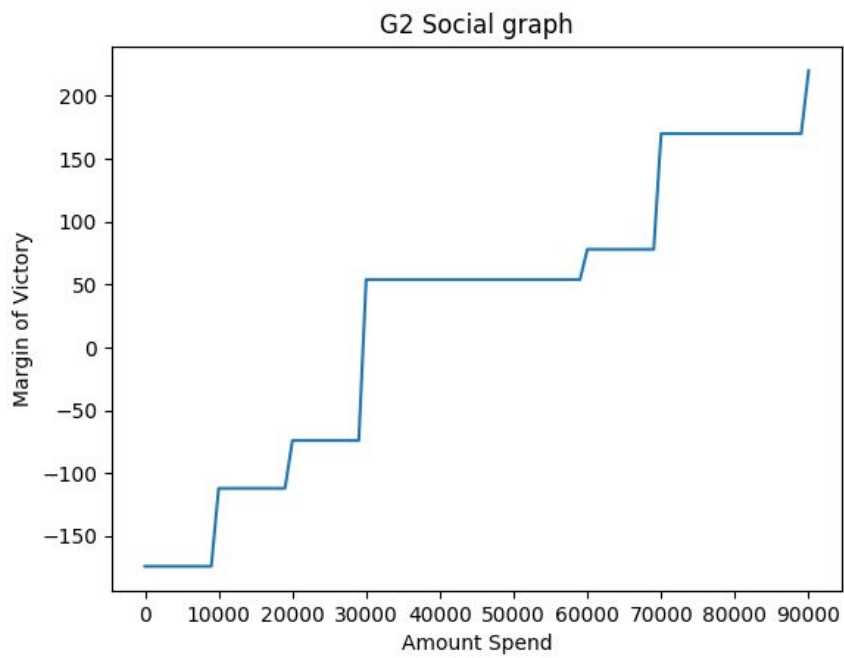
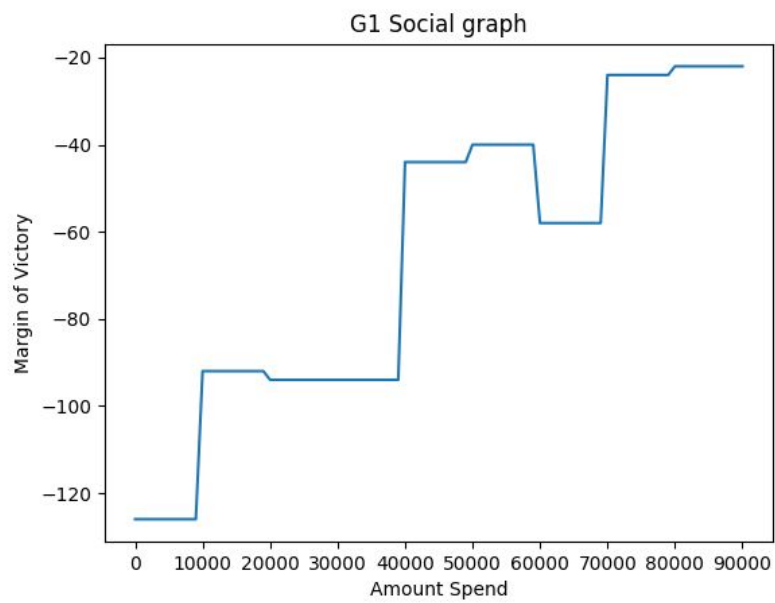
-----SOCIAL GRAPH 1-----

The minimum amount of money that needs to be spent by A is 22000 winning by 2. - > 30k

-----SOCIAL GRAPH 2-----

3The minimum amount of money that needs to be spent by A is 49000 winning by 14. - > 50k

**Q1c)**



-----SOCIAL GRAPH 1-----

A cannot win

-----SOCIAL GRAPH 2-----

The minimum amount of money that needs to be spent by A is 30000 winning by 54.

### **Q1d)**

#### **Q1b) Explanation**

The margin of victory is much higher in G1 than in G2 this may be because the avg degree of the 3000- 3099 id group is higher in G1 than in G2.

#### **Q1c) Explanation**

The implications of G2 having a higher local clustering coefficient would mean that on if we pick a node with a highest degree in both the graphs, then in G2 nodes is able to influence more undecided voters than G1 on average.

### **Best Strategy**

The best strategy would be the following:

“Procure the the best candidates which provide the largest individual jump in polls from the advertising group and the degree distribution.”

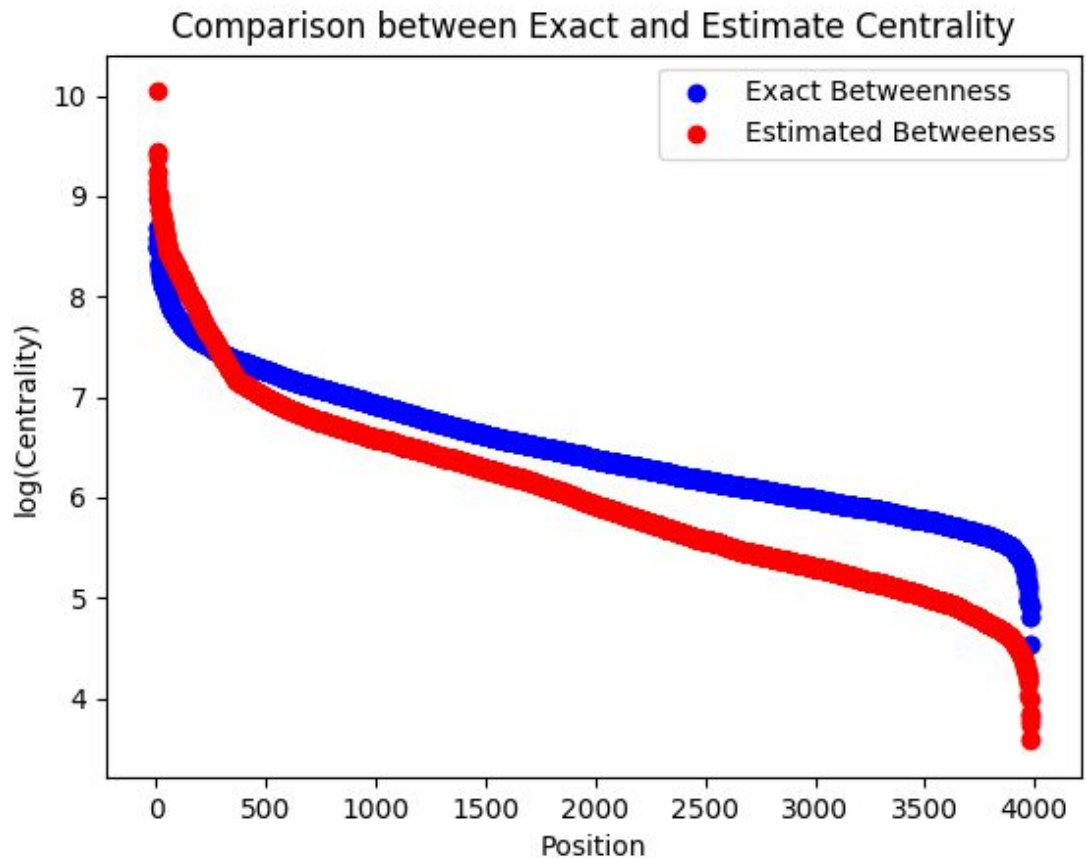
This can be done by taking into account the following

- 1) Study the influence of the polls on the on each individual node.
- 2) The nodes which already have a majority of candidates who will vote for “A” may not increase polls by much
- 3) Nodes which have most “U” voters and a high degrees may be the most important nodes.

### **Other Strategies**

Use a Weighted combination of different metrics such as pagerank and coreness. Order the vertices according to this new metric and evaluate based on this.

**Q2)**



The following observations are noted -

1. The higher valued estimated centralites are better estimated. The metric that was used to identify was per 50 size step was -  

$$\frac{\text{abs}(e1 - e2)}{\max(e1, e2)}$$
It first decreases till the 200 - 250 elements and then rises. This is verified by the baove graph.  
Values given below
2. Secondly, It is also important to note if we increase the #random set then the estimated



centrality more closely approximates the exact centrality.

Starting,Ending, Normalized Metric value

0 50 26.16067132981203  
50 100 22.16689926202688  
100 150 17.502487910541618  
**150 200 9.41057553375187**  
**200 250 1.8681038787171065**  
**250 300 6.3968444648835545**  
**300 350 10.122692346348645**  
350 400 11.426287684554197  
400 450 12.414988108015109  
450 500 12.938193419432803  
500 550 13.333625477877947  
550 600 13.5223422019435  
600 650 13.379796557910227  
650 700 13.626469154595306  
700 750 13.80965289542903  
750 800 14.04915131094047  
800 850 13.74682450326083  
850 900 13.781174981240667  
900 950 13.86328124537315  
950 1000 14.139307433528867  
1000 1050 14.330864780645632  
1050 1100 14.355930120682595  
1100 1150 14.307447803252998  
1150 1200 14.215728937218483  
1200 1250 14.433553831576146  
1250 1300 14.604334265198293  
1300 1350 14.883501251375108  
1350 1400 15.407504683773963  
1400 1450 15.838132198785257  
1450 1500 16.187965635988636  
1500 1550 16.360836861861507  
1550 1600 16.56624188258288  
1600 1650 16.999672077916525  
1650 1700 17.51632359136941  
1700 1750 18.46499481464469

1750 1800 19.023562203229737  
1800 1850 19.300866601617056  
1850 1900 19.69269312319733  
1900 1950 19.94189217359531  
1950 2000 20.468648591320434  
2000 2050 20.90786166782976  
2050 2100 21.122373174486487  
2100 2150 21.379858510073827  
2150 2200 21.75607426888963  
2200 2250 21.826987277465516  
2250 2300 21.9240677611024  
2300 2350 22.177822445689266  
2350 2400 22.303443523016732  
2400 2450 22.602791411295765  
2450 2500 23.013999391044386  
2500 2550 23.375001226395497  
2550 2600 23.680591839401053  
2600 2650 23.8834768290086  
2650 2700 24.086063786087323  
2700 2750 24.26896228758495  
2750 2800 24.280183680127525  
2800 2850 24.45464446106666  
2850 2900 24.78076939320505  
2900 2950 24.983213372774788  
2950 3000 25.33749026858178  
3000 3050 25.510828177508916  
3050 3100 25.681867462601492  
3100 3150 25.838975130504377  
3150 3200 25.98647457340207  
3200 3250 26.286686362042303  
3250 3300 26.622451144195136  
3300 3350 26.82968819886186  
3350 3400 26.90165213716215  
3400 3450 27.05996545977703  
3450 3500 27.34378229978729  
3500 3550 27.57526685194183  
3550 3600 27.739362909354437

