IITM Remote Internship 2018

Project Report

Graph Partitioning

Submitted By
Sahibjot Singh Bedi

Mentored By
Asst. Prof. Rupesh Nasre

Acknowledgement

I am highly indebted to Assistant Prof Rupesh Nasre for his guidance and constant supervision as well as for providing necessary information regarding the project and also for support in completing the project.

Contents

1 Problem

Consider a graph G = (V, E), where V denotes the set of vertices and E the set of edges. For $1 < K < |V|/2$, partition V into K parts (subsets) V_1, $V2_2$, ..., V_K such that the parts are disjoint and have (almost) equal size, and the number of edges with endpoints in different parts is minimized.

2 Kernighan Lin algorithm

The Kernighan Lin algorithm is a heuristic algorithm for finding partitions of graphs. The algorithm has important applications in the layout of digital circuits and components in VLSI. A slight variation of it has been used to find k partitions of a graph.

2.1 Pseudocode

Input : G(V, E) and K
Initially : P :- set of all nodes

```
 1: do
 2: split P into A and B of size N - N/K and N/K
 3: compute D values for all a in A and b in B
 4: let gval, pos1 and pos2 be empty lists
 5:     for (n :- 1 to |B|/2) do
 6:         find a from A and b from B, such that g = D[a] + D[b] − 2*c(a, b) is
    maximal.
 7:         remove a and b from further consideration in this pass.
 8:         add g to gval, position of a in A to pos1 and b in B to pos2.
 9:         update D values for the elements of A = A \ a and B = B \ b
10:     end for
11: find k which maximizes maxg, the sum of gval[1],...,gval[k]
12: if (maxg > 0) then
13:     exchange the elements in A with positions pos1[1], pos1[2],..., pos1[k] and B
    with positions pos2[1], pos2[2],..., pos2[k]
14: permanently remove nodes in B from further consideration
15: K :- K-1, N :- N - N/K and consider A as P
16: while (K > 1)
```

2.2 Time Complexity

Line 3: Initial computation of D: $O(N^2)$
Line 5: The for loop: $O(N/K)$
The body of the for loop: $O(N^2 \log E)$
Lines 2 - 15: Each pass of the do-while loop: $O(N^3 \log E/K)$.

The do-while loop terminates after K-1 passes.
The total running time: O(N^3logE).

2.3 Results

The algorithm was applied on 4 datasets. The results are shown below with the following parameters :
N - No. of nodes
E - No. of edges
K - No. of desired partitions
T - Time taken only for partitioning (in seconds)
M - Mean size of a partition
V - Varience of partition size

1. **Small Graph** : A small undirected and unweighted graph with 8 nodes and 13 edges.

N	E	K	T	M	V
8	13	2	0.001	4	0
8	13	3	0.001	2.6	0.333333
8	13	4	0.001	2	0

2. **Zachary's karate club** : A well-known social network of a university karate club described in the paper "An Information Flow Model for Conflict and Fission in Small Groups" by Wayne W. Zachary.

N	E	K	T	M	V
34	78	2	0.002	17	0
34	78	3	0.002	11.3333	0.333333
34	78	4	0.002	8.5	0.333333
34	78	5	0.003	6.8	0.2
34	78	6	0.002	5.66667	0.266667
34	78	7	0.002	4.85714	0.142857
34	78	8	0.002	4.25	0.214286
34	78	9	0.003	3.77778	0.194444
34	78	10	0.001	3.4	0.266667
34	78	11	0.001	3.09091	0.0909091
34	78	12	0.001	2.83333	0.151515
34	78	13	0.003	2.61538	0.25641
34	78	14	0.004	2.42857	0.263736
34	78	15	0.002	2.26667	0.209524
34	78	16	0.002	2.125	0.116667
34	78	17	0.003	2	0

3. **Caenorhabditis elegans worm's neural network** : The the neural network of the Caenorhabditis elegans worm (C.elegans). It was studied by Watts and Strogatz (1998). The network contains 306 nodes that represent neurons. Two neurons are connected if at least one synapse or gap junction exist between them. The weight is the number of synapses and gap junctions.

N	E	K	T	M	V
306	2345	20	0.269	15.3	0.221053
306	2345	40	0.258	7.65	0.233333
306	2345	60	0.291	5.1	0.0915254
306	2345	80	0.353	3.825	0.146203
306	2345	100	0.394	3.06	0.0569697
306	2345	120	0.48	2.55	0.24958
306	2345	140	0.545	2.18571	0.152312

4. **US power grid** : The network is the high-voltage power grid in the Western States of the United States of America. The nodes are transformers, substations, and generators, and the ties are high-voltage transmission lines. This network was originally used in Watts and Strogatz (1998). Although the transmission lines can be directed and differentiated based on their capacity, this information is not available.

N	E	K	T	M	V
4941	13188	100	220.201	49.41	0.244343
4941	13188	300	165.638	16.47	0.249933
4941	13188	500	205.787	9.882	0.104285
4941	13188	754	286.806	6.55305	0.247514
4941	13188	891	376.032	5.54545	0.248212
4941	13188	1101	404.31	4.48774	0.250077
4941	13188	1333	479.19	3.70668	0.20744
4941	13188	1535	563.054	3.21889	0.17109
4941	13188	1717	582.702	2.87769	0.10741
4941	13188	1999	726.427	2.47174	0.249326
4941	13188	2213	798.938	2.23272	0.17864

2.4 Optimization

The best way to speed performance of this implementation would be to eliminate the descriptive output. Most of the algorithm's running time is bound by I/O, and to eliminate intermediate (descriptive) I/O would be to cut running time dramatically.

3 References

https://goo.gl/8iszXp
https://goo.gl/g2Zs7x
https://goo.gl/6REyFH
https://goo.gl/9vAkYE