

Homework 7

Introduction to Big Data Systems

N Y Hendrikx < 2019403077 >

April 28, 2019

1 Intro

This report contains the details of the homework 7. The task is to implement graph partition in a language by choice. My choice is python. The additional libraries are networkx and matplotlib. I use these to visualize the graphs.

2 Implementation

This is an example where the vertices are balanced for edge-cut. The same technique is applied for vertex-cut, but with edges.

```
vertex_list = list(g.nodes())
random.shuffle(vertex_list)

#divide nodes to partitions evenly
while(len(vertex_list) > 0):
    l = len(vertex_list)
    if(l % 10000 == 0):
        print('len: ', str(l))

    for i in partitions:
        if(len(vertex_list) > 0):
            rnd = vertex_list[-1]
            i.add_node(rnd)

            #add edges between masters
            current_nodes = list(i.nodes())
            for u in current_nodes:
                if(g.has_edge(u, rnd)):
                    i.add_edge(u, rnd)

            del vertex_list[-1]
```

2.1 Experiments

The experiments exist out of vertex-cut and edge-cut on the following data-sets:

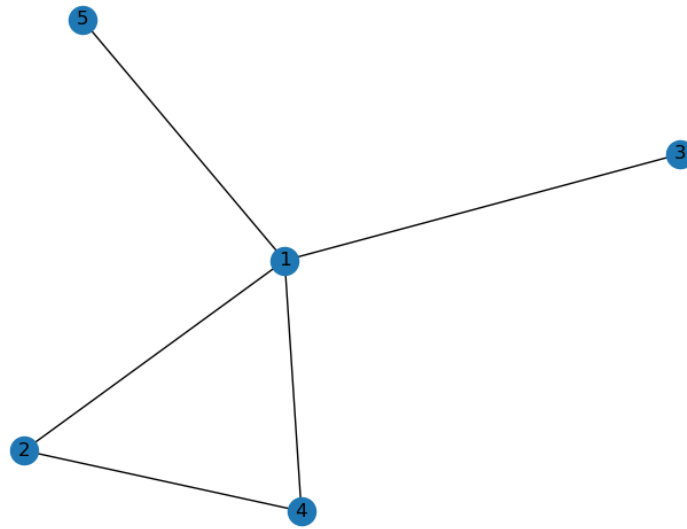
- roadNet-PA.graph.
- synthesized-1b.graph.
- small-5.graph.
- twitter graph.

3 Results

See attached text files for the detailed results of all experiments. They follow the pre-described template. They are separate for each input file. In general edge-cut would be slower because you need to iterate though all the nodes and also look at the edges. I did not run the twitter dataset because my process got being killed and my laptop is not fast enough and did not have enough memory. The greedy heuristic is implemented, but trough lack of time and optimization I did not run it on all data-sets.

3.1 Visuals

Figure 1: original graph



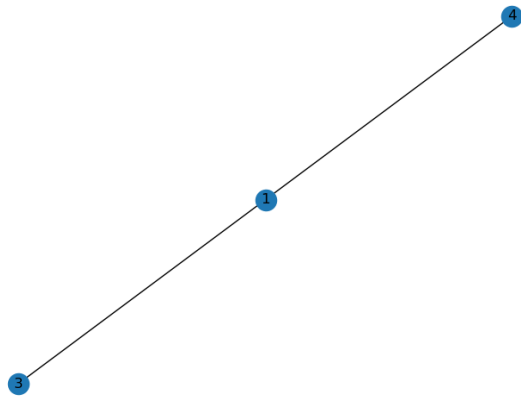


Figure 2: vertex cut 1

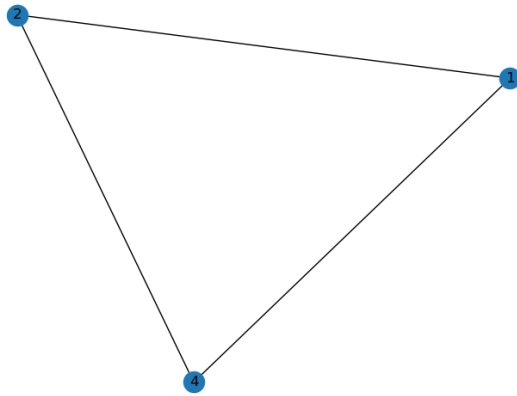


Figure 3: vertex cut 2

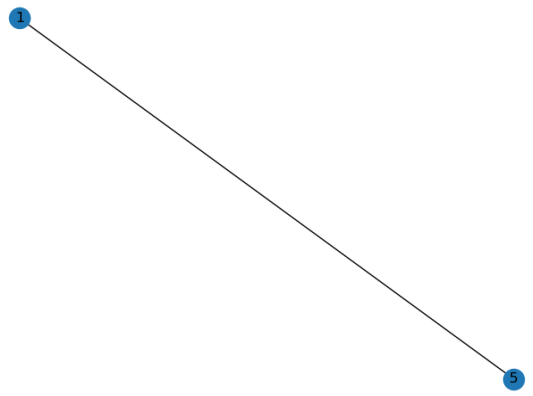


Figure 4: vertex cut 3

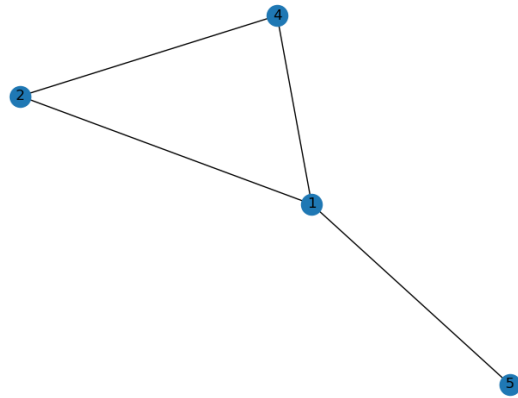


Figure 5: edge cut 1

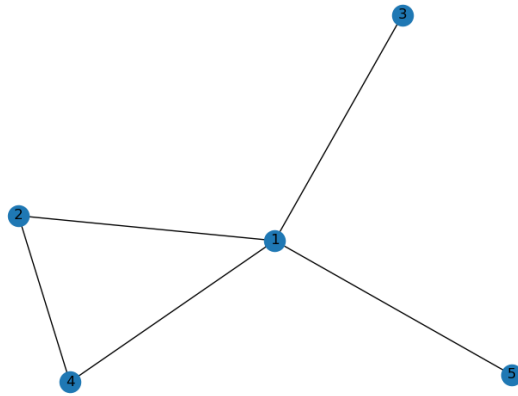


Figure 6: edge cut 2

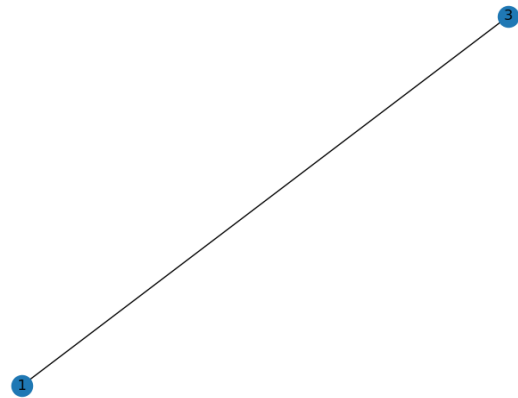


Figure 7: edge cut 3

4 Conclusion

Vertex-cut has way better performance than edge-cut.