Lab5: Finding community structures

Jakub Šalagovič
Simon Van den Eynde

November 23, 2016

1 Introduction

This report exists of 2 parts. In the first part we will compare some of igraph's community-finding algorithms on specifically chosen graphs, using the metrics: Triangle Partition Ratio (high is best), expansion (low is best), conductance (low is best) and modularity (high is best).

2 Results

	TPT	expansion	conductance	modularity
edge.betweenness	1.000	4.400	0.524	0.425
fastgreedy	1.000	4.400	0.524	0.425
label.propagation	1.000	4.160	0.482	0.340
leading.eigenvector	1.000	3.200	0.333	0.000
multilevel	1.000	4.400	0.524	0.425
optimal	1.000	4.400	0.524	0.425
spinglass	1.000	4.400	0.524	0.425
walktrap	1.000	4.400	0.524	0.425
infomap	1.000	4.400	0.524	0.425

Table 1: Metrics for HanoiTower(5,2) (HT)

	TPT	expansion	conductance	modularity
edge.betweenness	0.000	1.933	0.476	0.451
fastgreedy	0.000	2.000	0.501	0.411
label.propagation	0.000	1.800	0.430	0.420
leading.eigenvector	0.000	1.667	0.385	0.389
multilevel	0.000	1.933	0.476	0.451
optimal	0.000	1.933	0.476	0.451
spinglass	0.000	2.067	0.527	0.442
walktrap	0.000	1.933	0.476	0.451
infomap	0.000	1.867	0.456	0.416

Table 2: Metrics for Double Star Snark (DSS)

	TPT	expansion	conductance	modularity
edge.betweenness	0.600	2.467	0.529	0.228
fastgreedy	0.400	2.733	0.609	0.217
label.propagation	1.000	1.800	0.333	0.000
leading.eigenvector	1.000	1.800	0.333	0.000
multilevel	0.600	2.600	0.565	0.222
optimal	0.600	2.467	0.530	0.239
spinglass	0.400	2.600	0.577	0.230
walktrap	0.400	2.600	0.575	0.181
infomap	1.000	1.800	0.333	0.000

Table 3: Metrics for the Dorovtsev-Goltsev-Mendes(3) Graph (DGM)

3 Discussion

For more information on the chosen graphs, see section 4.

- In general we notice that the modularity sometimes becomes zero, this happens only if the entire network is one community. In this case we can discard these findings.
- In graphs without triangles (BA, DSS) or with many triangles (HT) we find that that the Triangle Partition Ratio does not convey any useful information.
- We see that for if there are clear communities as in HT, almost every community-finding method can identify them correctly.
- When looking at the metric values for the DSS network, we see that the leading.eigenvalue gives rise a special community structure. Its expansion and conductance are a lot lower than for other graphs, which is good. But its modularity is also a lot lower, which is a bad sign. When comparing network sizes (not included) we note it is the only community structure which consists of only 2 communities. So

	TPT	expansion	conductance	modularity
edge.betweenness	0.000	1.075	0.381	0.680
fastgreedy	0.000	1.100	0.393	0.678
label.propagation	0.000	1.200	0.451	0.614
leading.eigenvector	0.000	1.075	0.381	0.680
multilevel	0.000	1.100	0.393	0.678
optimal	0.000	1.075	0.381	0.680
spinglass	0.000	1.100	0.393	0.678
walktrap	0.000	1.125	0.406	0.652
infomap	0.000	1.125	0.406	0.670

Table 4: Metrics for Barabasi-Albert (BA)

	TPT	expansion	conductance	modularity
edge.betweenness	0.794	3.000	0.497	0.401
fastgreedy	0.824	2.853	0.454	0.381
label.propagation	0.882	2.706	0.427	0.338
leading.eigenvector	0.735	3.059	0.502	0.393
multilevel	0.794	2.912	0.468	0.419
optimal	0.794	2.912	0.468	0.420
spinglass	0.794	2.912	0.468	0.420
walktrap	0.588	3.235	0.545	0.353
infomap	0.882	2.706	0.419	0.402

Table 5: Metrics for Zachary's Karate network (ZK)

depending on how many communities you want, you might consider using different metrics our different community-finding methods.

- The values for the conductance and expansion of the BA network are lower than that of other networks. Also the modularity is higher. This means that this power-law delivers stronger communities than many other graphs.

4 Methods

4.1 Graphs

We chose 5 different graphs. As a real network we chose Zachary's karate (ZK) network. We also analysed a Barabási-Albert (BA) graph on 40 vertices. Furthermore we considered 3 special graphs:

HanoiTower(5,2) (HT) This graph has as nodes the game states of Hanoi Tower game with 5 pegs and 2 disks, there is an edge if you can go from one game state to another in one move. This graph basically consists of 5 copies of K_5 which are sparsely connected. See figure 1.

Double Star Snark (DSS) This is graph is snark, which means it doesn't contain any triangle (this graph does only contain cycles with length > 5), it is also cubic and has no bridges. It consists of 30 vertices. See figure 2.

Dorovstev-Goltsev-Mendes graph (DGM) This is a graph which can be constructed, starting with K_2, as follows: for every edge a triangle (add a vertex and 2 edges). If we do this 3 times we get a graph with 15 vertices, 27 edges and a lot of triangles. See figure 3.

5 Appendix

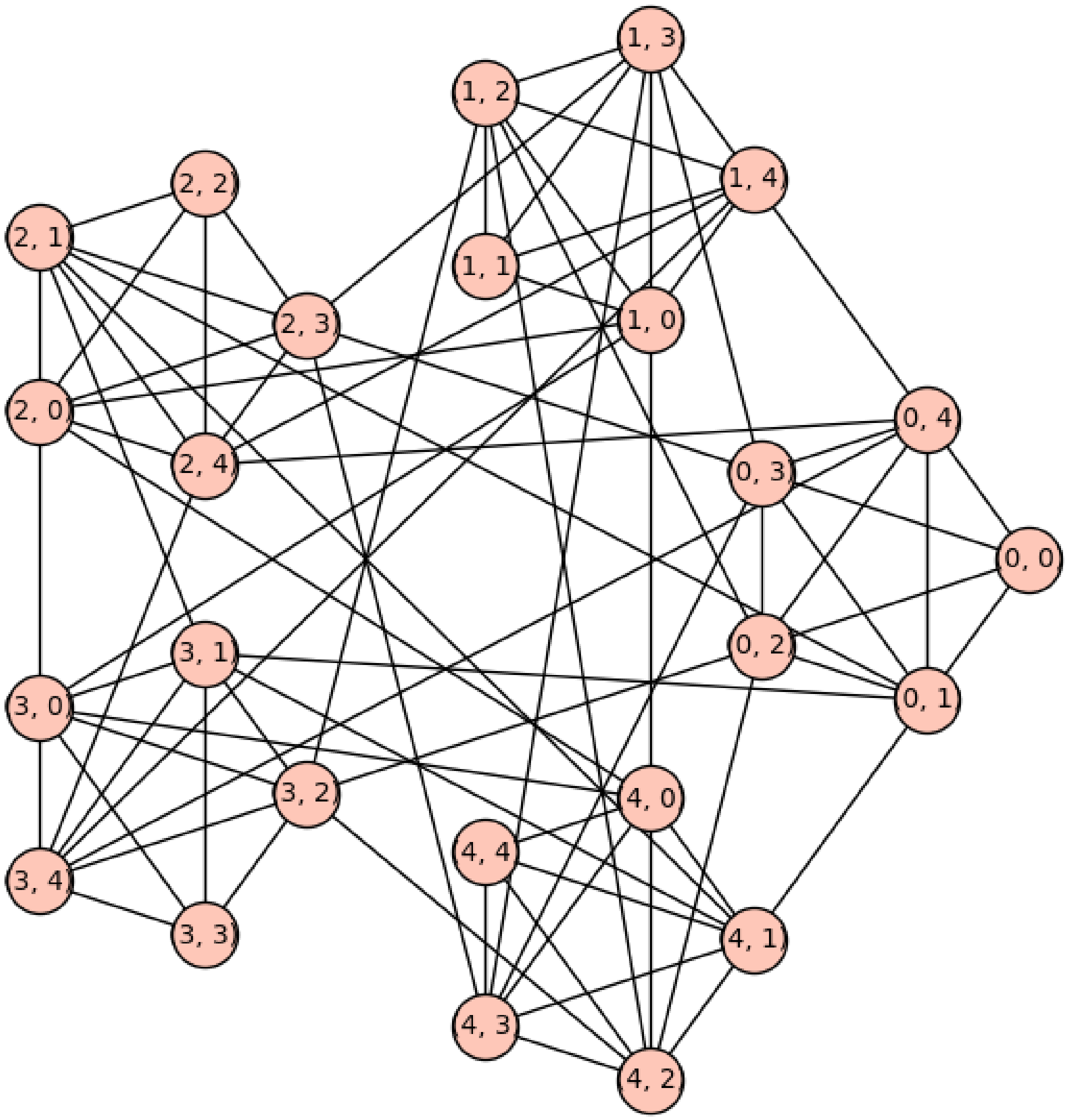

Figure 1: The Hanoi Tower graph with 5 pegs and 2 discs. The labels on the vertices indicate the positions of the two discs ($(1,3)$ means: peg 1 on disk 1, peg 2 on disc 2)

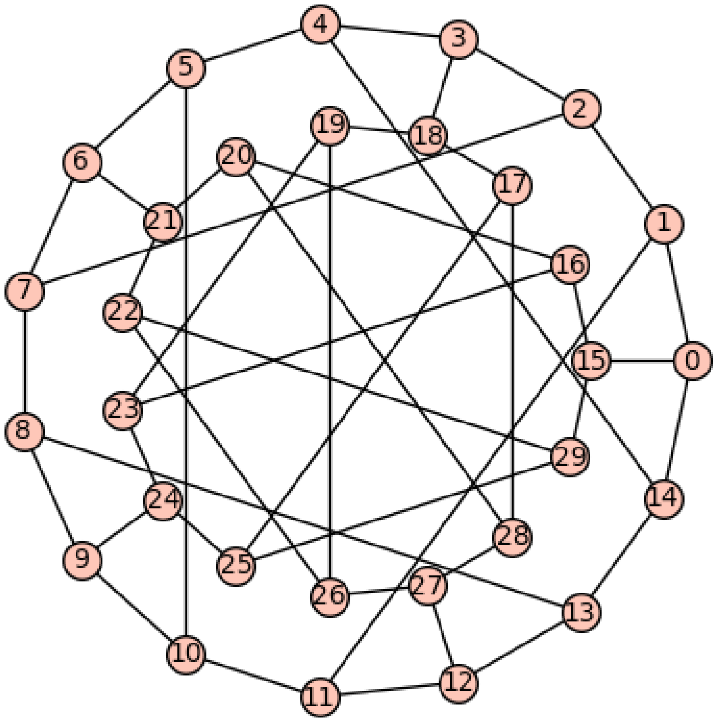

Figure 2: The double star snark

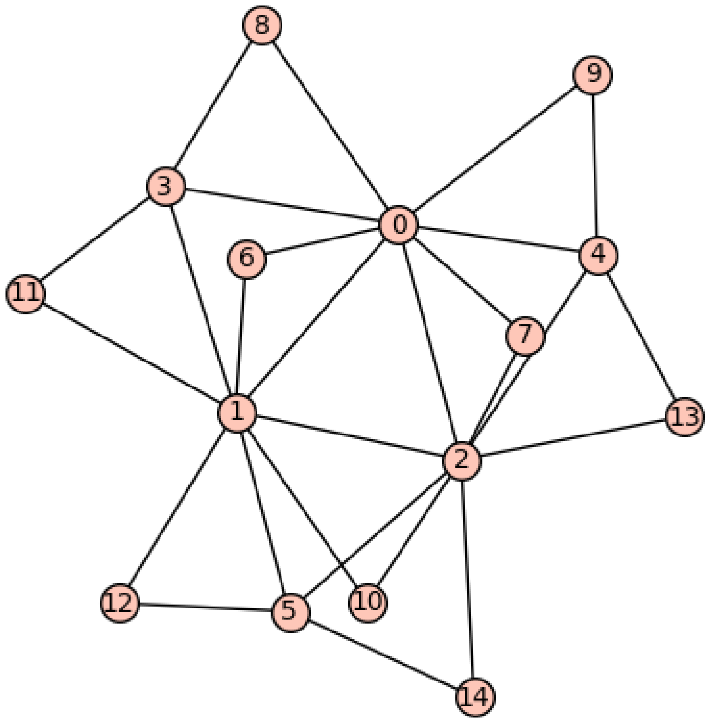

Figure 3: The Dorovstev-Goltsev-Mendes graph after 3 iterations