Data Mining, ID2222

Short report on homework 5

K-way Graph Partitioning Using JaBeJa

Sanvito Alessandro and Stuart Thuany

December 13, 2021

1 Introduction

The laboratory involved the implementation of the Ja-Be-Ja algorithm, a K-way partitioning algorithm based on the concept of simulated annealing. At first, we implemented the algorithm with a linear acceptance probability function, where:

$p = 1$ if $T * cost_{new} > cost_{old}$ else 0

Then, we implemented a new acceptance probability with the following definition:

$p = 1$ if $cost_{new} > cost_{old}$ else $e^{\frac{cost_{new} - cost_{old}}{T}}$

Finally, we provided our own custom acceptance probability, inspired by the Sigmoid function:

$p = 1$ if $cost_{new} > cost_{old}$ else $\frac{1}{1+e^{\frac{-(cost_{new} - cost_{old})}{T}}}$

Which had the best performance with respect to edge cut in one of our benchmarks.

2 Results

We have experimented running the JaBeJa algorithm with different configurations for the three different annealers: linear, exponential, and custom. We have chosen to vary the parameters that seemed to have bigger effect on the edge cut: temperature, delta, alpha, and the restart of the simulated annealing. The different parameter configurations and the resulting edge cut can be seen in Table 1. The configurations without the restart are indicated with -1 and the

ones with restart contain the number of rounds between the restarts.

The configurations with best results for each graph-annealer combination is presented in bold. Due to computation time, the different configurations were tried on the smaller graph: 3elt, and the best configurations were then applied to the other two graphs. We are aware that this approach does not yield the best results for the add20 and facebook graphs, but we believe that it is enough to observe the behavior of the algorithm.

Table 1: Parameter configurations and resulting edge cut.

Graph	Annealer	Temperature	delta	alpha	Restart	Edge Cut
3elt	**Custom**	**1**	**0.9**	**10**	**-1**	**1223**
3elt	Custom	1	0.9	5	-1	1228
3elt	Custom	1	0.7	5	400	1262
3elt	Custom	1	0.9	3	-1	1308
3elt	Custom	1	0.7	5	-1	1343
3elt	Custom	1	0.9	5	250	1389
3elt	Custom	1	0.9	2	-1	1405
3elt	Custom	1	0.99	5	-1	1537
3elt	**Exponential**	**0.5**	**0.9**	**5**	**100**	**2545**
3elt	Exponential	1	0.9	5	-1	2545
3elt	Exponential	1	0.99	3	-1	2550
3elt	Exponential	1	0.9	2	-1	2573
3elt	Exponential	1	0.9	3	-1	2573
3elt	Exponential	1	0.99	5	-1	2586
3elt	Exponential	1	0.9	5	250	2601
3elt	**Linear**	**2**	**0.03**	**3**	**250**	**1931**
3elt	Linear	2	0.03	3	-1	2151
3elt	Linear	2	0.03	2	-1	2169
3elt	Linear	2	0.0015	2	-1	2332
3elt	Linear	5	0.03	2	-1	2595
3elt	Linear	2	0.03	1	-1	2988
add20	**Custom**	**1**	**0.9**	**5**	**-1**	**2200**
add20	**Exponential**	**1**	**0.9**	**5**	**250**	**2452**
add20	**Linear**	**2**	**0.03**	**3**	**250**	**2181**
facebook	**Custom**	**1**	**0.9**	**5**	**-1**	**137533**
facebook	Custom	1	0.99	5	-1	138492
facebook	Custom	1	0.9	5	200	139614
facebook	**Exponential**	**1**	**0.9**	**5**	**250**	**140412**
facebook	**Linear**	**2**	**0.03**	**3**	**250**	**134310**

In Table 1, we can observe that the linear acceptance function is consistently the best performing in two out of three graph, with our custom acceptance probability reaching a considerable result only in 3elt. However, our custom probability consistently outperforms the exponential one.

In terms of parameters, as both the initial temperature and the delta control

the same exploration/exploitation trade-off, we focused on delta. As expected, the best results are yielded when the exploratory behavior is long enough to skip local minima, but short enough to let the algorithm coverge in the fixed amount of steps.

Alpha on the other hand controls the cost and from the data we can observe that higher alphas lead to better results.

Finally, the restart parameter seems to be particularly relevant only for the linear annealer, which improves greatly, while the exponential and custom annealer can sometimes perform better without any restart at all. The insignificant effect of restart is also confirmed in the following plots and it might be due to the fact that the probability of accepting a "bad" solution in the non-linear annealers never becomes zero, thus always promoting some measure of exploration. On the other hand, once the temperature reaches one in the linear annealer, no more bad solutions are accepted and the exploration stops. Therefore, a restart is required to move from a local minimum.

Figure 1 shows the plots of the edge cut, number of swaps and number of migrations for the best configuration for the 3elt graph. We observe that the number of edge cuts stabilizes (although still decreasing) after around 200 rounds.

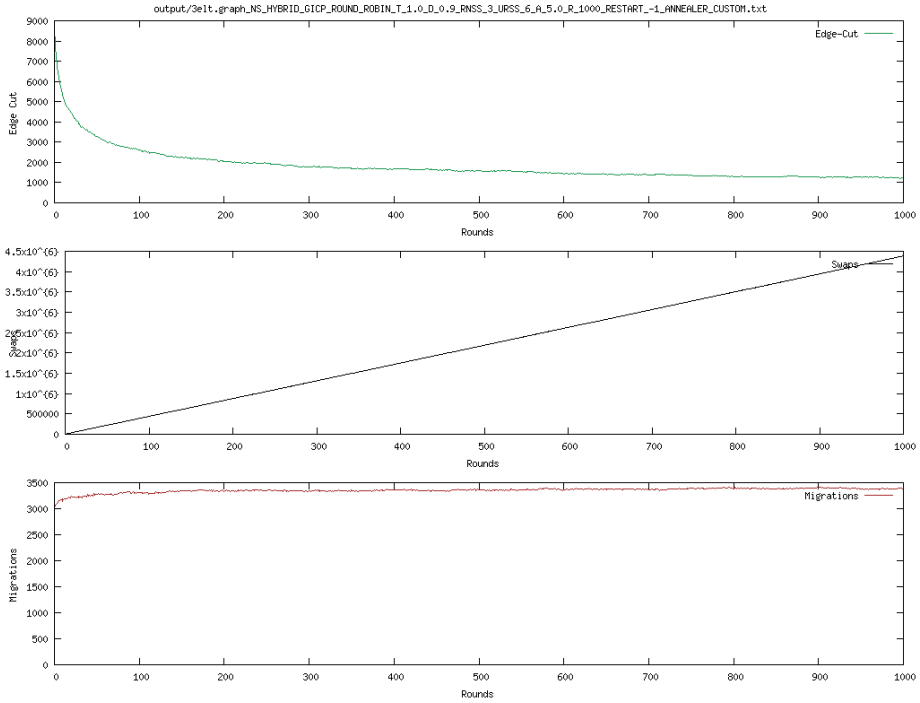

Figure 1: Best configuration for the 3elt graph with Custom annealer, temperature = 1, alpha = 5, delta = 0.9, and without restart.

For the worse pipeline for the 3elt graph, shown in Figure 2 the edge cut converges earlier, at around 100 rounds, after decreasing less from the initial value when compared to the best configuration.

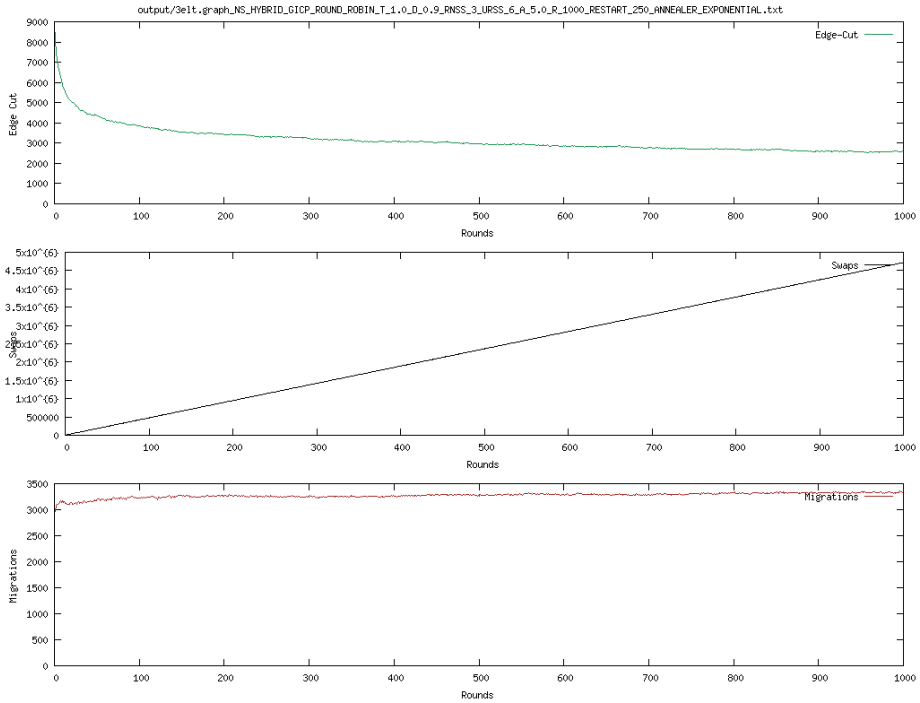

Figure 2: Worse configuration for the 3elt graph with Exponential annealer, temperature = 1, alpha = 5, delta = 0.9, and with restart every 250 rounds.

Figure 3 shows the configuration for the add20 graph with the custom annealer. We observe that the number of edge cuts continuously decreases, although it stabilizes after about 100 rounds. There are no perturbations after the stabilization because there is no restart.

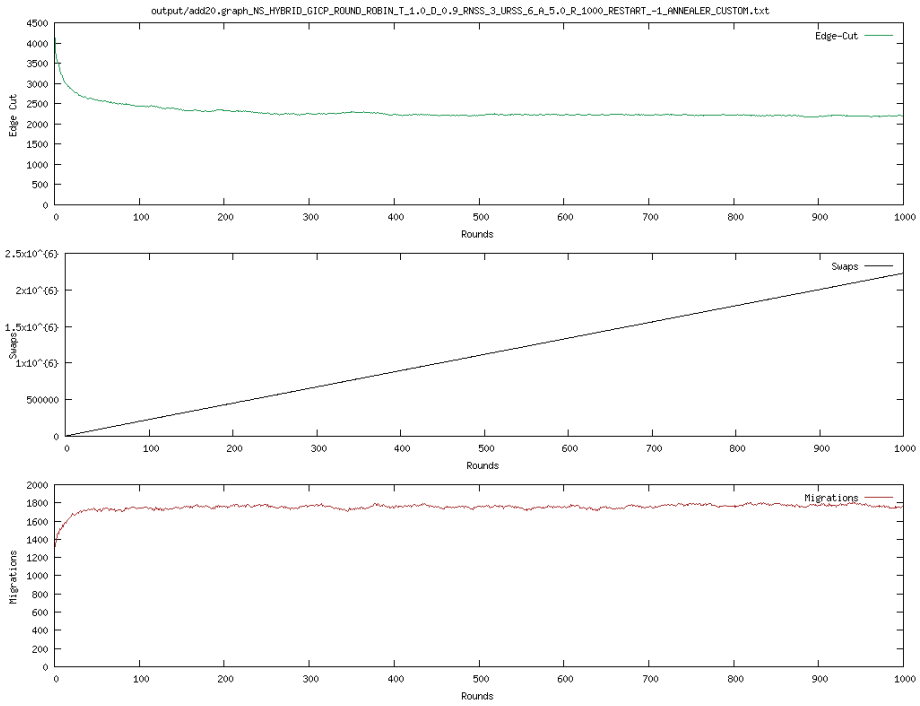

Figure 3: Best configuration for the add20 graph with Custom annealer, temperature = 1, alpha = 5, delta = 0.9, and without restart.

Figure 4 shows the best overall configuration for add20. Compared to the the custom annealer, the linear annealer converges in a slower way, but reaches an overall better solution helped by the restart.

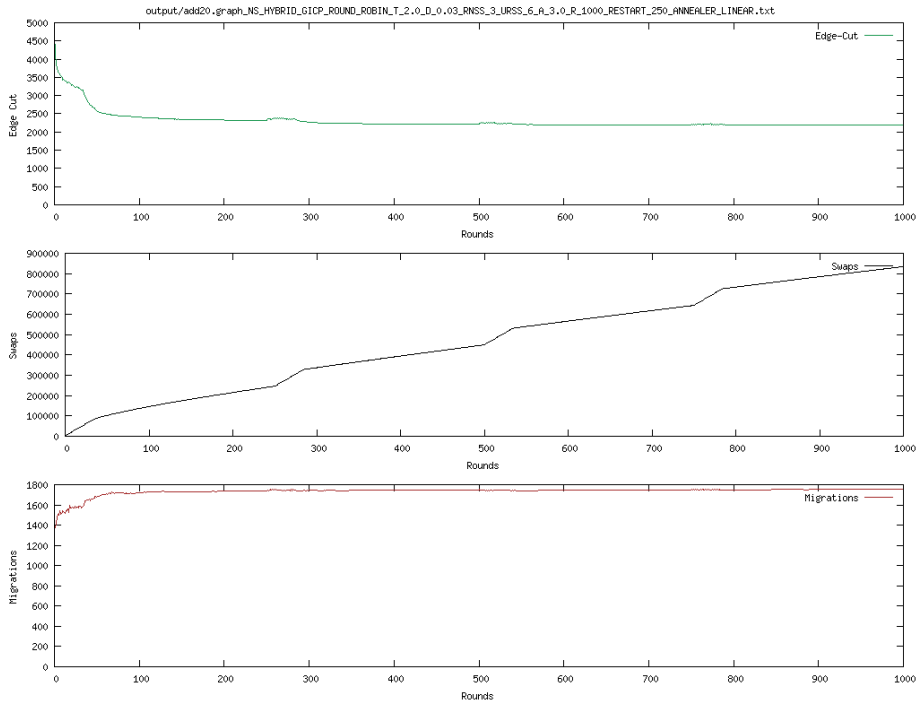

Figure 4: Best configuration for the add20 graph with Linear annealer, temperature = 2, alpha = 3, delta = 0.03, and restart every 250 rounds.

Figure 5 shows the best configuration for the facebook graph. We observe that the restart of the temperature every 250 rounds increases the exploration, therefore, it increases the number of edge cuts and the number of swaps for a a few rounds until it stabilized again. Nevertheless, the impact of the restart in the number of edge cuts seems to be minimal, since the edge cut plot seems to follow the previous trend after the perturbation.

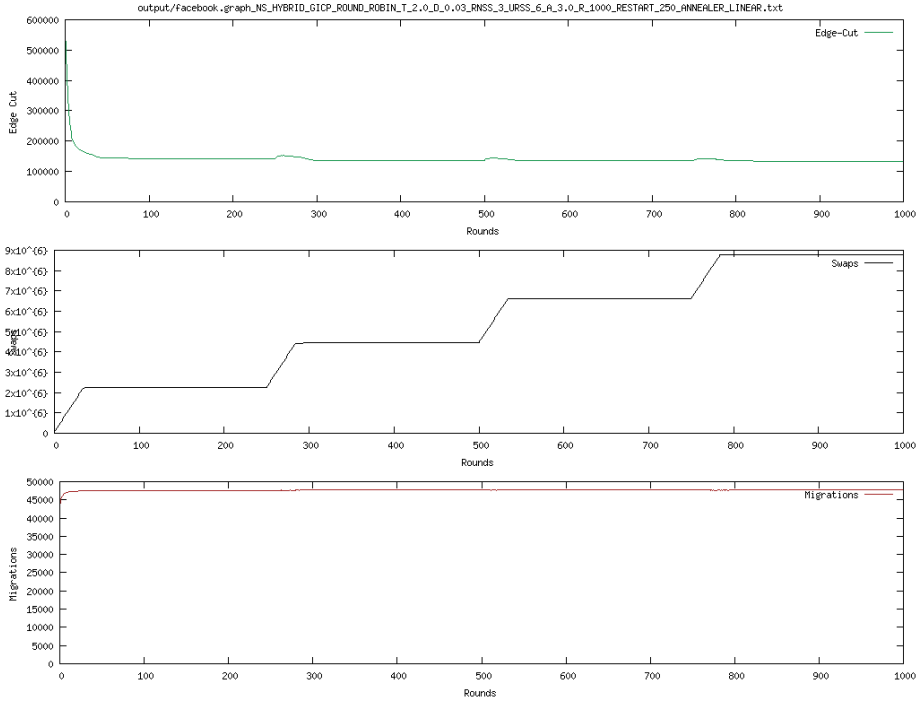

Figure 5: Best configuration for the facebook graph with Linear annealer, temperature = 2, alpha = 3, delta = 0.03, and with restart every 250 rounds.