Report Homework 5

Data Mining (ID2222)
Report by Henny Selig (hselig@kth.se) and Ignacio Amaya de la Peña (iadlp@kth.se)
08.01.2017

Comparison of the graphs and algorithm

Evaluating input parameters and effects

We have implemented the required changes of task 1. We compare the 3 graphs 3elt, add20 and facebook. The values from this execution will be the reference for future comparison and are shown in the overview below.

	3elt Standard	**Add20 Standard**	**Facebook Standard**
Edge cut	2604	2095	134246
Swaps	1580209	1090263	21200364
Migrations	3328	1751	47763
runtime	33 sec.	16 sec.	37 min. 1 sec.

The plots can be found in the annex.

We can see in all 3 plots that the edge cut doesn't much decrease after a while (3elt and add20 very clear around 340).

We have then changed different parameters (one at a time) and recorded the changes. First we changed the temperature T to start with. We see that swaps go on for longer, if we increase T. The same applies when decreasing delta. This is due to the fact that the function to calculate the new T-value is linear. This changed when we implemented a new function. However, changing these parameters will have an impact on convergence of the algorithm. With these parameters we can influence the point when no bad swaps are allowed and thereby the point when the edge cut hardly decreases anymore. For example, for the 3elt graph, the point when hardly any swaps happen anymore changed from around round 350 to around round 650, when we increased T to 3. However, we can also see that this can lead to a lower edge cut, see table below.

Edge cut with:	T=1.5	**T=2 (standard)**	T=3
3elt	2566	**2604**	2449
Add20	2249	**2095**	2100
facebook	140011	**134246**	120800

In general, the configuration of T and delta should depend on the graph and its properties, and should fit each other as they influence the result in their combination.

Secondly, we changed the value of alpha. You can find an overview in the table below.

Edge cut with:	Alpha = 1	**Alpha = 2 (standard)**	Alpha = 3
3elt	2788	**2604**	2437
Add20	1797	**2095**	2184
facebook	168294	**134246**	131208[1]

We can't really see a trend for the edge cut in this comparison. However, if we compare the evolution of the edge cut, we see major differences in the course of the graph. Alpha is the parameter used in comparing whether a swap of colour makes sense or not. It therefore, in

[1] Alpha = 4 for facebook graph due to mistake when running it.

combination with the development of T, defines how much impact a change must make to be a good swap candidate. If alpha = 1, one can say it is ignored.

One thing to note is that the number of swaps doesn't necessarily correlate with a better edge cut. In general, there are many parameters to influence the behaviour of the algorithm. It is important to understand how they interact and influence the outcome, but they need to be adapted and tried depending on the graph. They might also impact the runtime. For the facebook graph, we have experienced runtimes between 32 minutes (with T=3), 2.5 hours (when changing alpha) and 4 hours (with T=1).

New function for T calculation

We have implemented a new function to calculate T, as instructed. As expected, the edge cut decreases and the number of swaps increases much steeper. An overview can be found below in Fig. 1, the larger graphs can be found in the annex. We see the same effect for the other graphs as well. When running the new function, T must be set to 1 (or lower) and delta is still given as a small value (e.g. 0.1). The lowest value of T is hardcoded, currently 0.001. It might be required to change this. Also, different examples should be run with different parameters. The effect should be the same as discussed above (of course in a different scale).

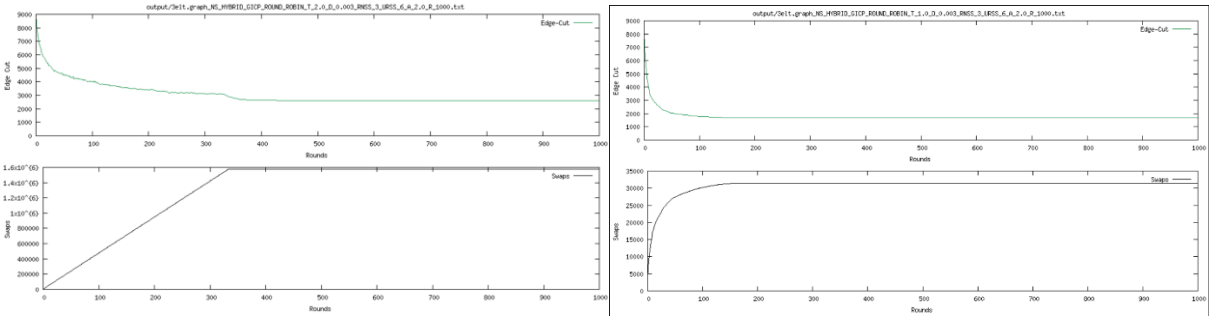

Figure 1: Comparison of old (left) and new (right) T calculation for 3elt

Resetting T after 400 rounds

We changed the code to reset T to its original value after round 400. We ran this with the old and the new way of calculating the new T value. In both cases, the resulting edge cut was not improved. If we allow bad swaps at a larger scale again by resetting T, the result can even be worse, as shown in the table below. As expected, the number of swaps increases, but these don't improve much of the edge cut. It makes more sense to control the behaviour of swaps with the T and delta parameters.

Edge cut	**standard**	**With T-restart**
3elt	2604	2622
Add20	2095	2432

Additional improvements

In our case, the runtime of the facebook graph was a serious problem. As the edge cut only increases marginally after a while, we introduced a check that stops the algorithm if hardly any changes occur anymore. In our case, we hardcoded that the algorithm stops if the edgecut hasn't improved for more than 10 rounds. This, however, could be implemented as a margin (e.g. increased x% over time) or set the 10 as a parameter, as this may depend on the graph.

With this improvement, we see the following for graph facebook:

facebook	**standard**	**Cut if no improvements**
Edge cut	134246	134265

swaps	21200364	21200345
Migration	47763	47764
runtime	37 min. 1 sec.	20 min 52 sec.

As we see, the runtime is roughly 16 minutes faster (43%), while the edge cut is only 19 higher (0.01%). In our opinion, this is a helpful improvement if runtime matters.

Annex

The Standard graphs

These are the graphs used for comparison.

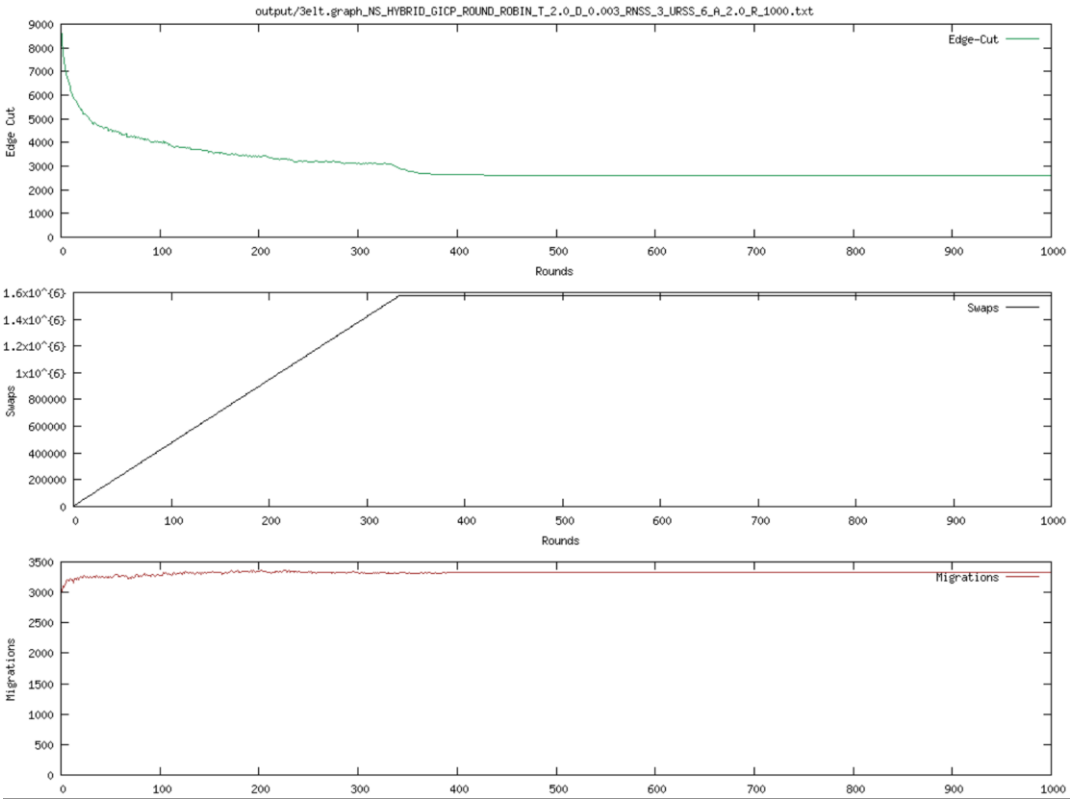

Figure 2: 3elt standard execution

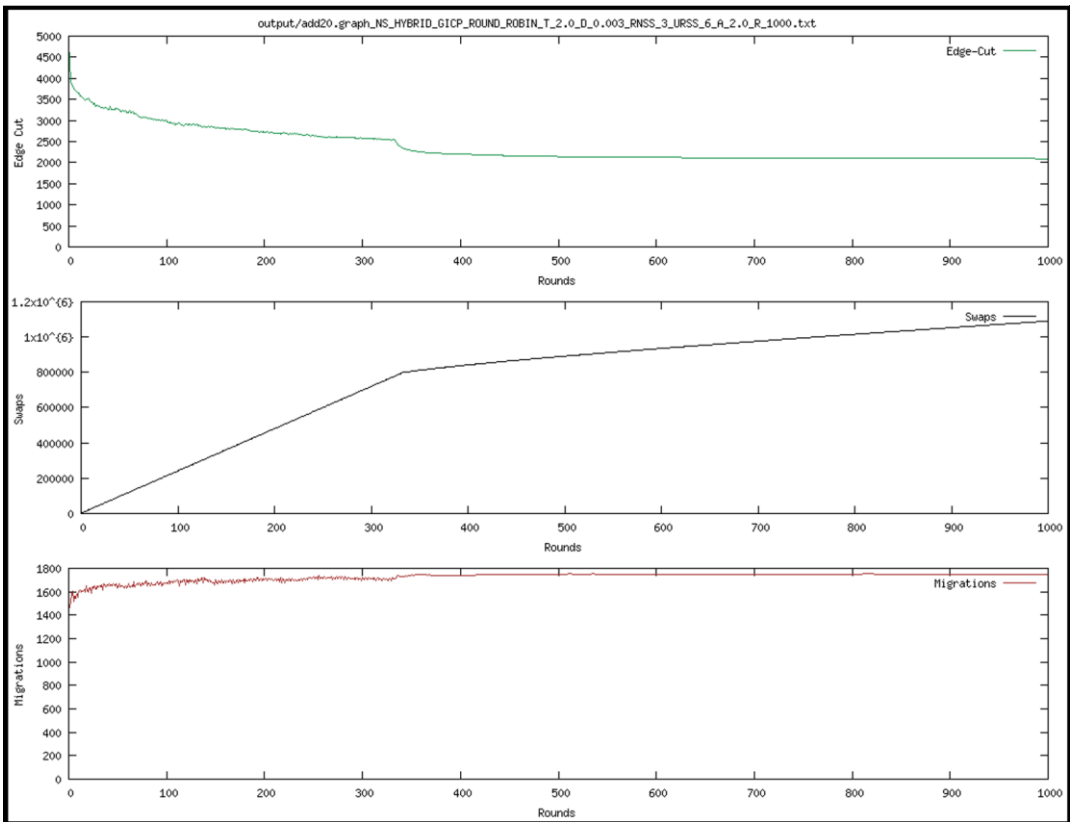

Figure 3: add20 standard execution

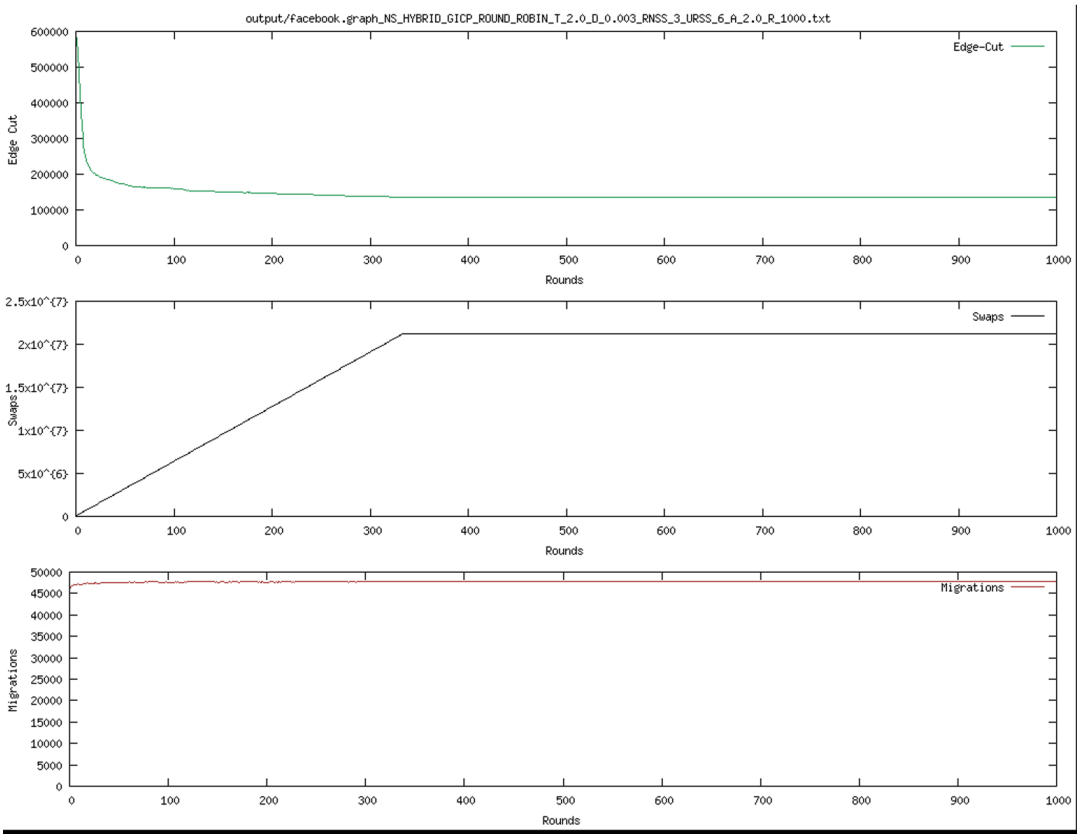

Figure 4: Facebook standard exection

Alpha comparison

The following shows the edge cut for different alpha values.

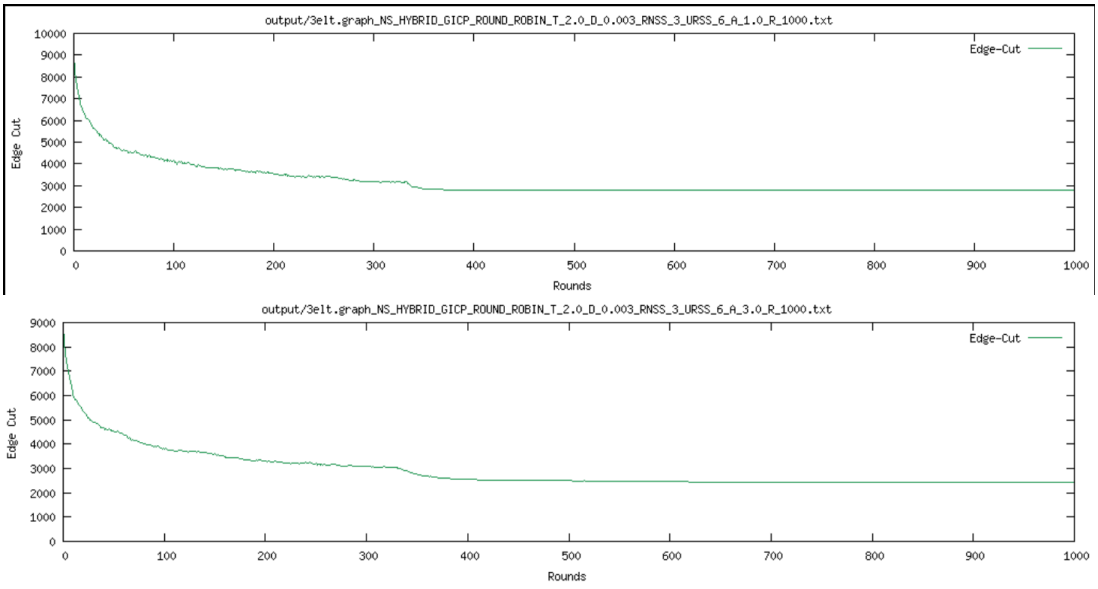

Figure 5: 3elt with alpha=1 (upper) and alpha=3 (lower)

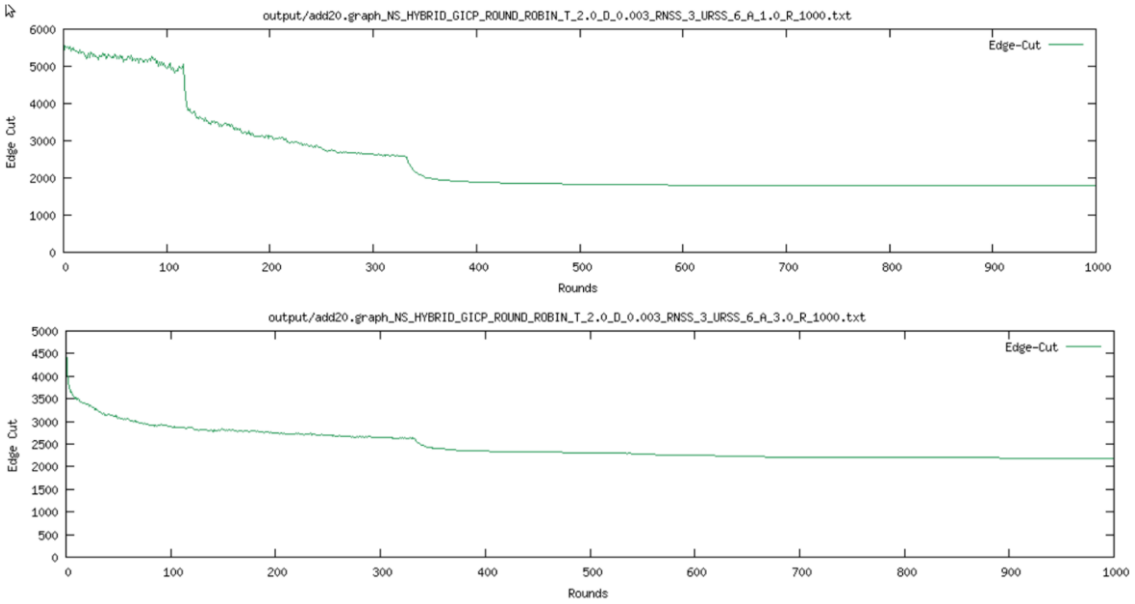

Figure 6: add20 with alpha=1 (upper) and alpha=3 (lower)

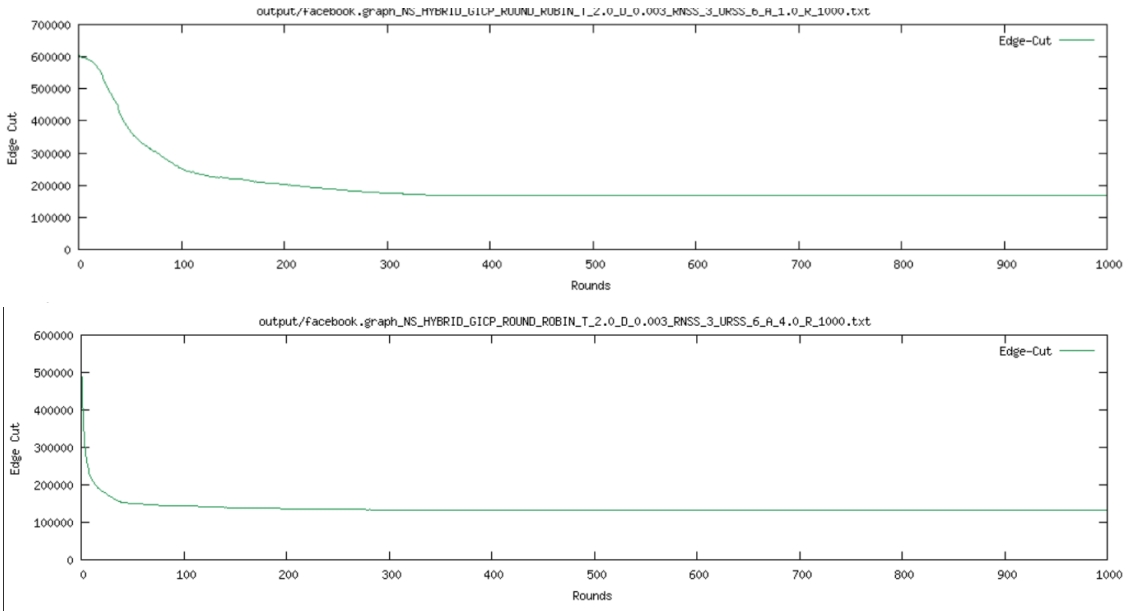

Figure 7: facebook with alpha=1 (upper) and alpha=4 (lower)

Change of Temperature calculation

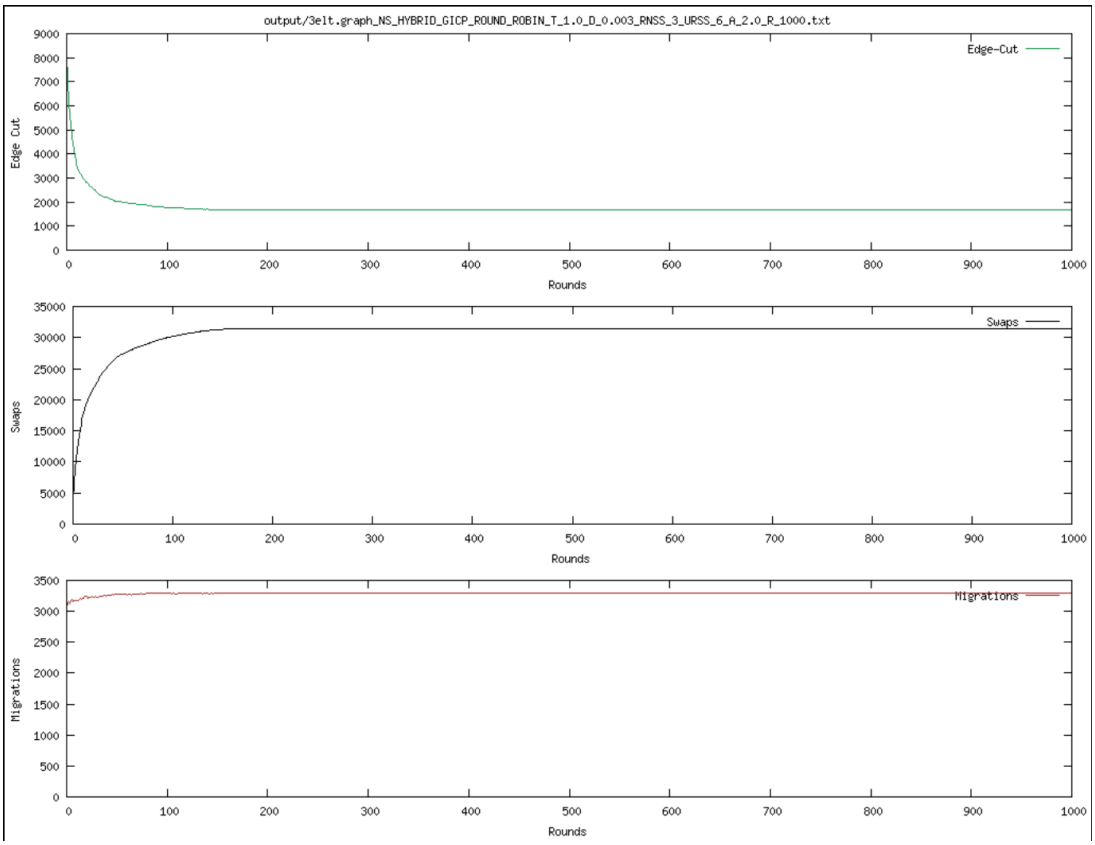

Figure 8: 3elt with new temperature calculation

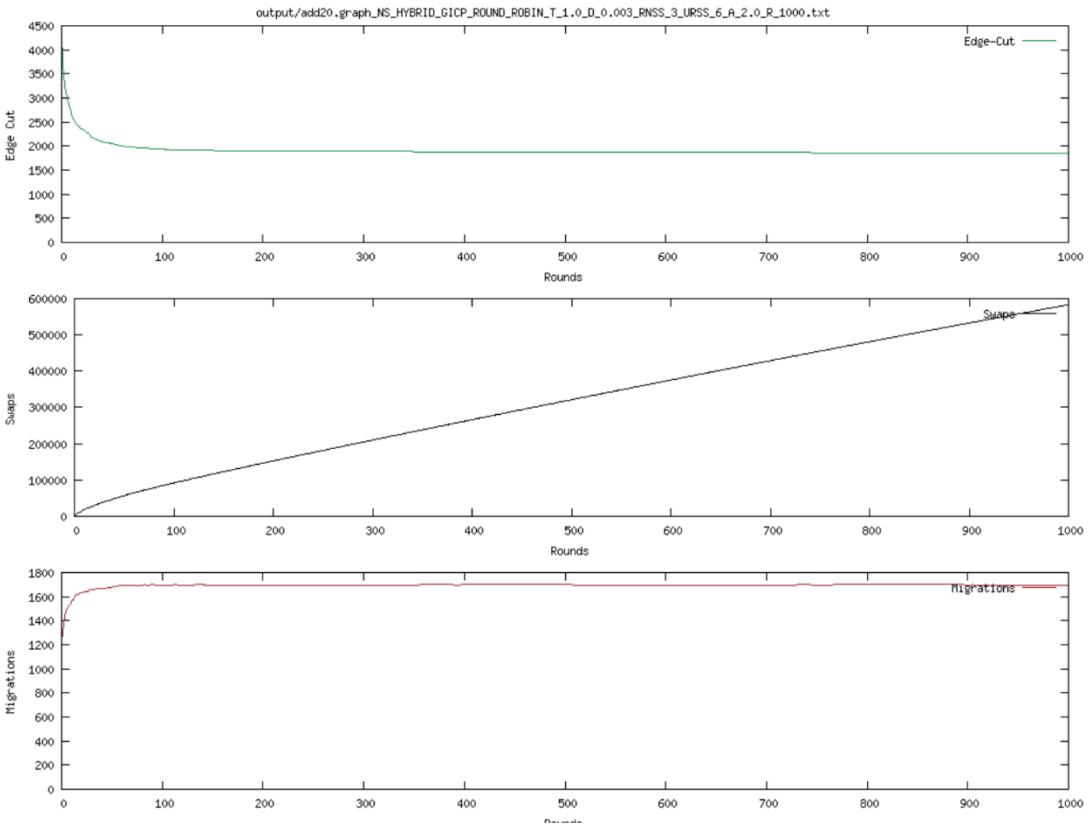

Figure 9: add20 with new temperature calculation

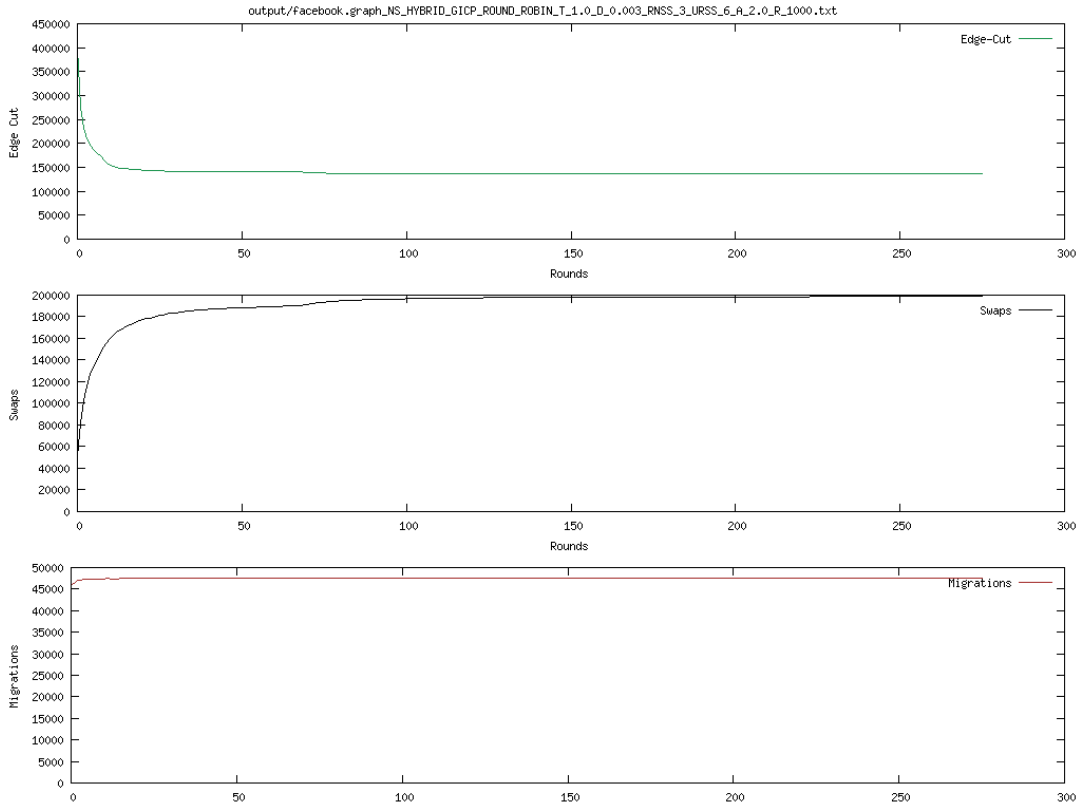

Figure 10: with new temperature calculation (incomplete)

Plots with resetting T

3elt and add20 with resetting T to its original value after round 400.

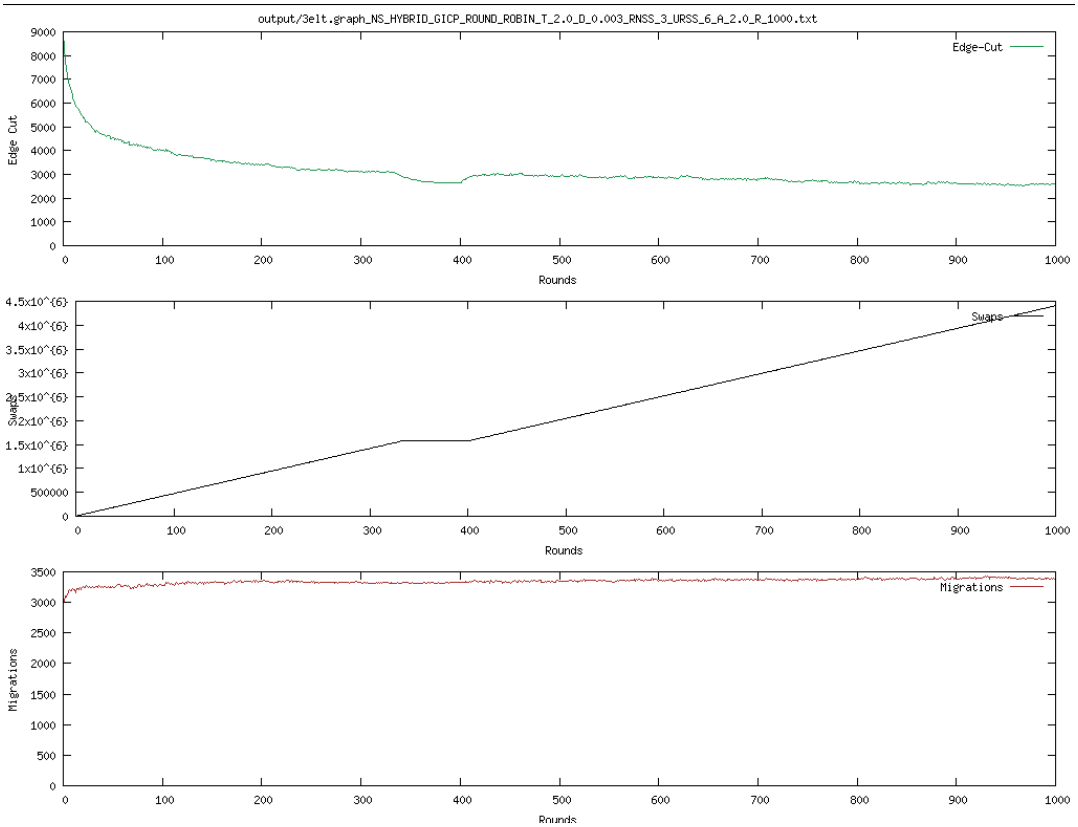

Figure 11: 3elt with restarting T

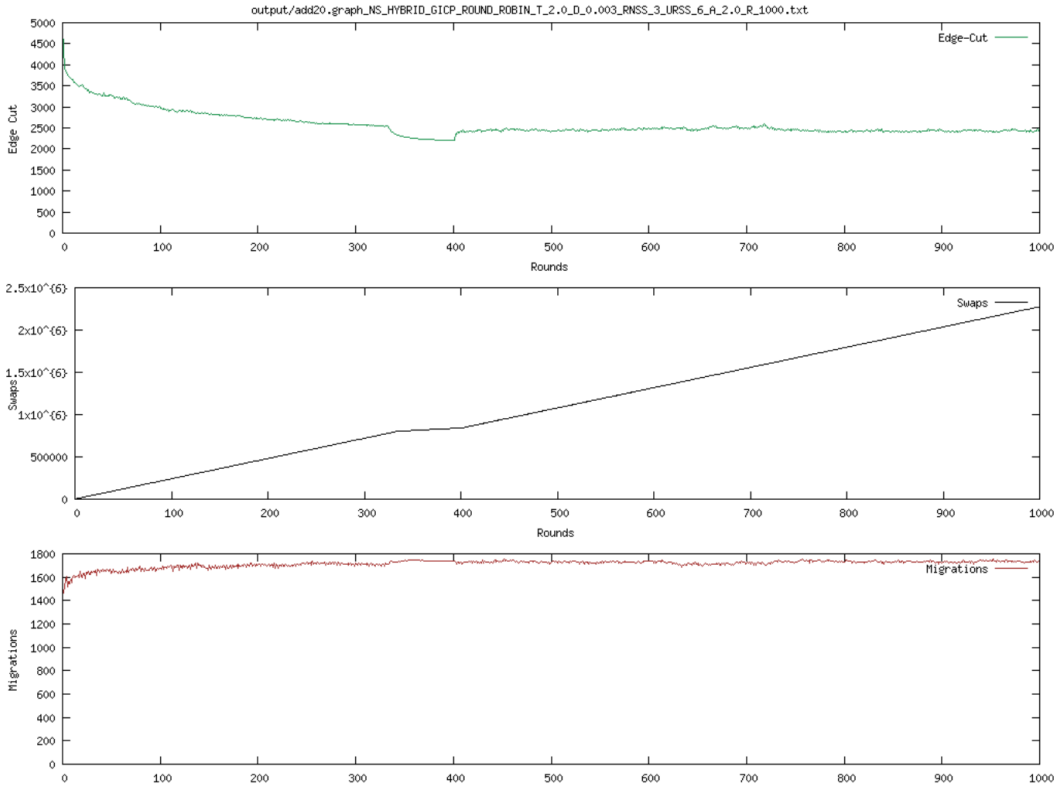

Figure 12: add20 with resetting T