

The K-Clustering Project

FAQ

Working enviorement	
Can I download cplex to my personal computer?	The formal answer is no. Cplex is very expensive and the only way to use is to connect to the linux servers. However, at your on risk, you can try the trial version of IBM http://www.ibm.com/developerworks/university/software/get_software.html But we do not know to support the installation and usage.
Working from home	http://www.cs.tau.ac.il/system/faq/services/servers/NX_CLIENT Fill yourself like you are in Schreiber lab while you are at home
Nova is slow	You can use other servers. From nova's terminal: <ol style="list-style-type: none"> 1. Type "ssh fermat-01" 2. enter password 3. type "top" to see if this server is busy. Shift+M will show you your processes at the top. If you see a process you forgot was running, go to the next comment. 4. Ctrl + c will take you out of the processes' list 5. go to your home directory (~) and continue as usual 6. If fermat-01 is busy try one of the others: fermat-01 to fermat-25 euler-01 to euler-25
When I typed "top" I saw that my process had been running for days, and is taking so much memory. How embarrassing... Can I kill it?	Yes please. Type "kill" and then your process' PID
General Instructions	
Error handling	The guidance is to handle errors with a wish to never stop. Regarding errors that you cannot overcome – for example one of the input files does not exists – exit the program with an appropriate message and memory release. Regarding errors that you can overcome – double edges for example – you inform the user, ignore the second edge (the one that caused the problem), and carry on with the program. Of course, you never ever crash or exit the program in an unorderly fashion.
How to inform the user?	"Error: <your message>"
Parameters	
The parameters are passed as	argv
What they contain?	relative path + slash, relative path + slash, C value
Project evaluation	
Can we change the declaration of the cluster	You most definitely can. Part of the project's checks are the design, and how you build the communication between the

function	different parts. Fill free to change whatever you want.
Sometimes the clusters' numbers are not unique, for example if I have two clusters with the same size and the same score. Or when there is more than one optimal solution	The automatic tests take that into consideration and both clustering would be counted as equal.
Part 1 (Cplex and statistics)	
Which arrays should contain some variables with coefficient of zero?	The array obj, correlated to the optimization function, should contain all variables. If a variable has no effect it will have a coefficient of zero. The rest of the arrays (matval for example) should not contain a variable with a coefficient of zero. To those arrays we would insert only the variables participating in the current constraint only.
What is the diameter of a cluster with isolated components (When at least one vertex in the cluster is unreachable from another vertex)?	The diameter is infinite. The output should be: Cluster <cluster_no>: score - <cluster_score> diameter - inf
Have you noticed that the objective function contains a part that does not change with the different assignment? And that in fact no variable is connected to it?	©Yes, I have. There are many ways to solve this, some very very simple, and some only simple. I trust you will figure this out.
Part 2 (cytoscape and xgmml)	
Can I use cytoscape on windows?	Yes, you can download it for free from www.cytoscape.org/
Cytoscape crashed when I try to upload any xgmml file (on linux)	Some versions of cytoscape have a problem with the 'label' attribute of the graph. Try, just for the sake of uploading to cytoscape, to change its value to "". (label="")
When I open an xgmml file I cannot see some of the attributes	Try opening the file using notepad++ and not a browser
I hate cytoscape	Close it, and open again. And maybe try to import only one network at a time.
What do I need to do in order to free all of the xml elements	Free any memory you allocated yourself xmlFreeDoc(file); xmlCleanupParser(); xmlMemoryDump();
While we are on the subject, is there a tool we can use to validate	valgrind is considered nice and easy. Try http://valgrind.org/ . It is not taught nor supported for the class, but you can try it anyway if you like. It should be helpful down the road.

memory release?	
Bonus	
Some of the network take forever to run	See comment about free servers and killing old processes. But also – the network that are not real usually take more time (because they are hard to cluster). So you can start with the one that ran fast. If you don't fill the scores for the other ones you will still get part of the bonus grade.