Homework 1 (chapters 3-4 of Alon *"An Introduction to Systems Biology"*).

Date delivered: April 11, 2014
Date due: April 24, 2014, 24.00
Mode of delivery: By e-mail, in PDF format.
Whereto: Mr Nicolas Innocenti <njain@kth.se>

This homework uses the YEASTRACT data base of transcriptional regulation in *Saccharomyces cerevisiae* and tools as described in the first tutorial. Start by retrieving the table of genetic interactions at `http://www.yeastract.com/` Mention in your report which version of the data base you have used as well as the date of the download.

1. Divide the genes in YEASTRACT into *regulated genes* which do not directly regulate the transcription of any gene (themselves included) and *transcription factors* (all the other genes).

   Consider the regulated genes as *output nodes*, the transcription factors as *internal nodes*, and regulation as *links*.

   How many output nodes, how many internal nodes, and how many directed links of different types between internal nodes and output nodes are there in this data set? Call the number of internal nodes N^*, the number of output nodes N^{out}, the number of directed links between internal nodes M^* and the number of links from internal nodes to output nodes M^{out}.

2. (analytical) Consider a network with N nodes, where every directed link is present with probability p (independence assumed). What is the expected number of directed links expressed in N and p? Call this function $E(p, N)$.

3. What is the value of p such that $E(p, N^*) = M^*$? Call this value p^*.

4. (analytical) In a network with N^* nodes, where every directed link is present with probability p^* what is the expected number of

   (a) auto-regulatory motifs ?

   (b) feed-forward loop (FFL) motifs ?

   (c) feed-back loop motifs ?

5. Consider the YEASTRACT data of transcription factors only. What is the actual number of...

   (a) auto-regulatory motifs ?

   (b) feed-forward loop (FFL) motifs ?

   (c) feed-back loop motifs ?

6. (Partial conclusion) Discuss the discrepancies between the answers to points 4 and 5 above (if any).

7. Redo points 4, 5 and 6 above for the FFL with the condition that node Z in the FFLs is an output node. Define your random graph null model for this situation and compute the relevant probabilities in terms of the counts N^*, N^{out}, M^*, M^{out}. Describe clearly what you do.

8. Based on the structure of the network described in point 1, explain why the null model as defined in point 4 is sub-optimal for estimating the number of feedback loops. Propose a better graph null model and redo points 4 and 6 for the feedback loop.

9. Redo point 4 for the three 4-node generalizations of the feed-forward loop defined in Alon Fig 5.6b page 83. Describe clearly what you do. (Optional) Redo points 5 and 6 for those three cases.

Warning note

This assignment is the same as when the course was last given (and earlier years). But the correct answers are not the same. The underlying databases are regularly updated.