

Computer Lab 5

Bioinformatics

Linköpings Universitet, IDA, Statistik

2022 XII 09

Kurskod och namn:	732A51 Bioinformatics
Datum:	2022 XII 06—2022 XII 14 (lab session 9 XII 2022)
Delmomentsansvarig:	Krzysztof Bartoszek, Hao Chi Kiang
Instruktioner:	<p>This computer laboratory is part of the examination for the Bioinformatics course</p> <p>Create a group report, on the solutions to the lab as a .PDF file.</p> <p>Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.</p> <p>All R code should be included as an appendix into your report.</p> <p>In the report reference ALL consulted sources and disclose ALL collaborations.</p> <p>The report should be handed in via LISAM</p> <p>(or alternatively in case of problems e-mailed to hao.chi.kiang@liu.se), by 23:59 14 December 2022 at latest.</p> <p>Notice there is a deadline for corrections 23:59 22 January 2023 and a final deadline of 23:59 12 February 2023 after which no submissions nor corrections will be considered and you will have to redo the missing labs at the next course opportunity.</p> <p>The report has to be written in English.</p>

Question 1

Go to the webpage <http://snap.stanford.edu/biodata/> and choose one of the provided datasets. Download it and reproduce the statistics concerning the graph. If you obtain different values, then discuss this in your report. Visualize the graph. The next step is to try to identify some clusters (communities in the graph). You can follow the tutorial at <https://psych-networks.com/r-tutorial-identify-communities-items-networks/> to achieve this. Once you have found some clusters, identify the elements in it and try to find information on this cluster. Is it related to some known biological phenomena? If you do not find anything, then document your search attempts. If it will not be possible to do this question on the whole downloaded graph, then you may take some sub-graph of it.

Question 2

Recreate one of the three analyses that can be found on <https://strimmerlab.github.io/software/genenet/index.html>. Document and discuss all your steps. In the analyses there is the step where you select the edges to keep. There a particular criterion is chosen for edge inclusion. Vary this criterion and explore how the resulting clusters will differ with the changes. Take one found cluster, identify the elements in it and try to find information on this cluster. Is it related to some known biological phenomena? If you do not find anything, then document your search attempts.