SCANNET

User Manual

Version 1.0
2015

ProjectSN

SUMMARY

ProjectSN

1. INTRODUCTION

The Network Similarity Analysis (NSA) method is a powerful tool for identifying communities in networks whose links depending on the similarity of proteins [1]. Its original implementation consisted of separated algorithms coded in different programming languages, making it difficult to use. To solve this problem was created the SCANNET, a package that integrates all algorithms of the NSA into a unique tool to perform all the steps needed to identify communities in protein networks as a single workflow.

The experiments have shown that SCANNET is a simple, efficient, reliable, and intuitive implementation of the NSA method to identify communities in protein networks.

2. GENERATING THE EXECUTABLE FILE

To compile and generate the executable file you must have a C compiler installed on your computer.

We suggest installing the compiler MinGW (Minimalist GNU for Windows) for Windows OS, available at: http://www.mingw.org/ or GCC (GNU Compiler Collection) for Linux or Mac OS, available at: https://gcc.gnu.org/.

After installation compiler, is required run the command line below to compile the scannet.c and generate the executable file.

gcc -o <EXECUTABLE NAME> scannet.c

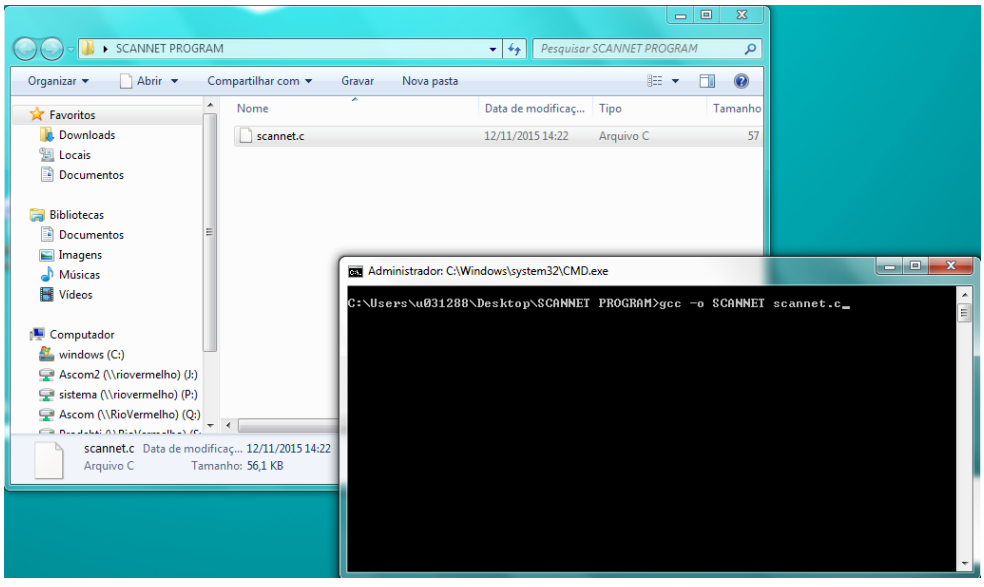

Figure 1 - Running command line

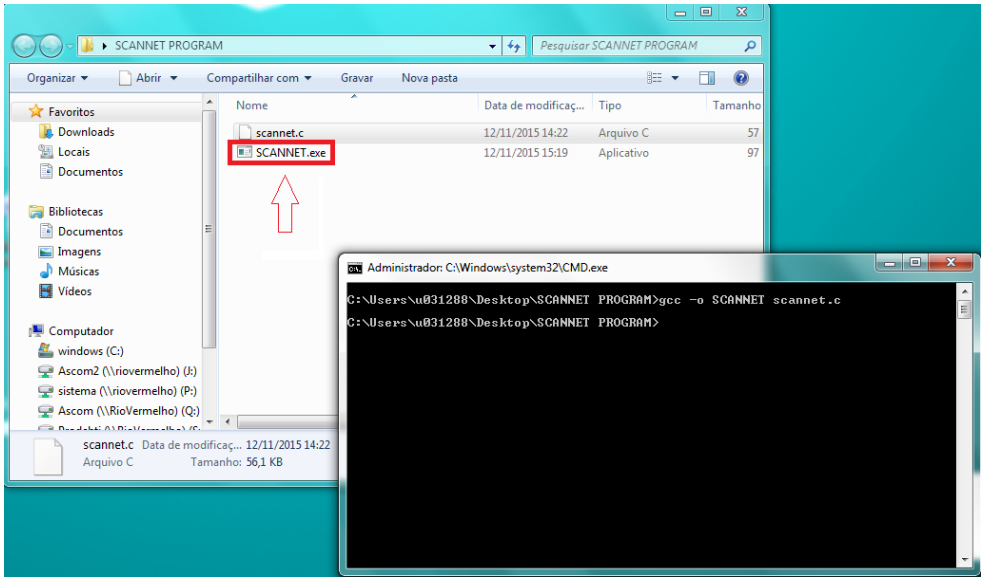

Figure 2 - Generated executable file

3. RUNNING PROGRAM

3.1. SETTING INPUT FILE

SCANNET use as input, a file (.txt or .dat) containing the similarity matrix between the proteins that will be analyzed. So you must provide the name of the input file. The used input file must be in the same SCANNET run directory.

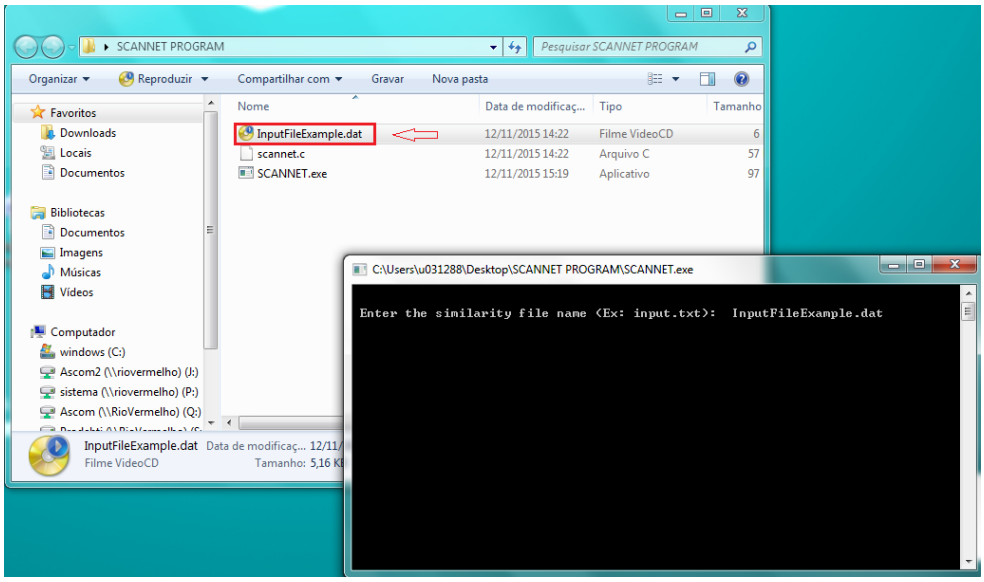

Figure 3 - Setting input file

3.2. SETTING PROTEIN ORDER FILE

During the communities identification process, the SCANNET reassembles the proteins of the similarity matrix, changing the initial order these proteins. To facilitate the identification of proteins in the generated output (Dendrogram, Complex Network and Color Matrix), the SCANNET generates an output file with the final order of proteins. If the user wanted has this feature, it's necessary set the file (.csv or .dat or .txt) with the initial protein order. The protein order file is optional, but if used must be in the same SCANNET run directory.

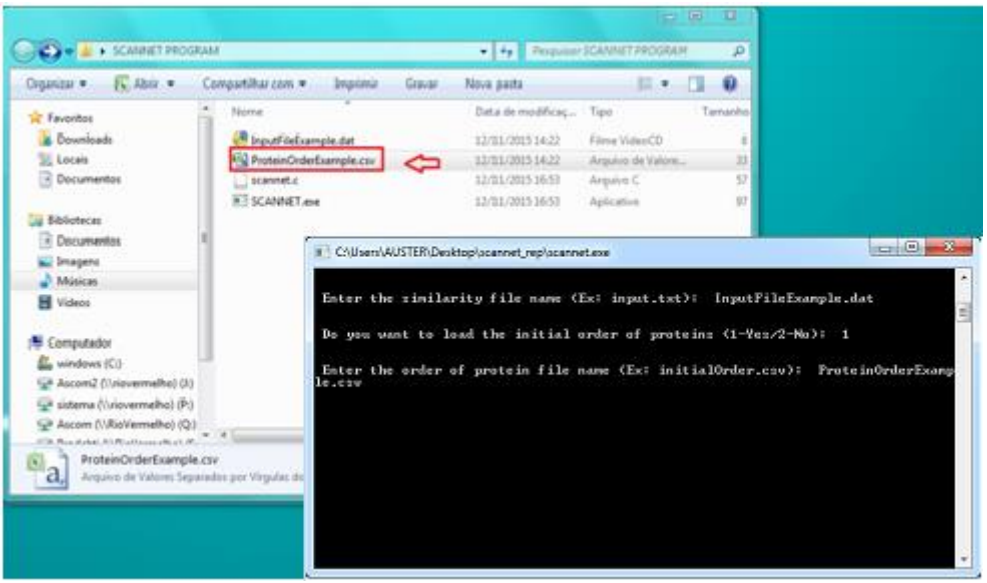

Figure 4 - Setting protein order file

3.3. SELECT SIMILARITY VALUE

The NSA method assumes that a set of 101 networks is generated from the dataset, each one for a specific threshold of similarity S in the range [0,100]. Afterwards, it calculates the distance between all pairs of subsequent networks in the network set, in order to select the similarity value that best represents the communities in the dataset. Such network is generally the one with the largest distance from its immediately subsequent network. In some cases, more than one similarity value could be selected. To facilitate the work of selecting the best similarity value, SCANNET displays a chart with all distances between pairs of neighboring networks and prompts the user to select the similarity value on which he will continue the analytical process.

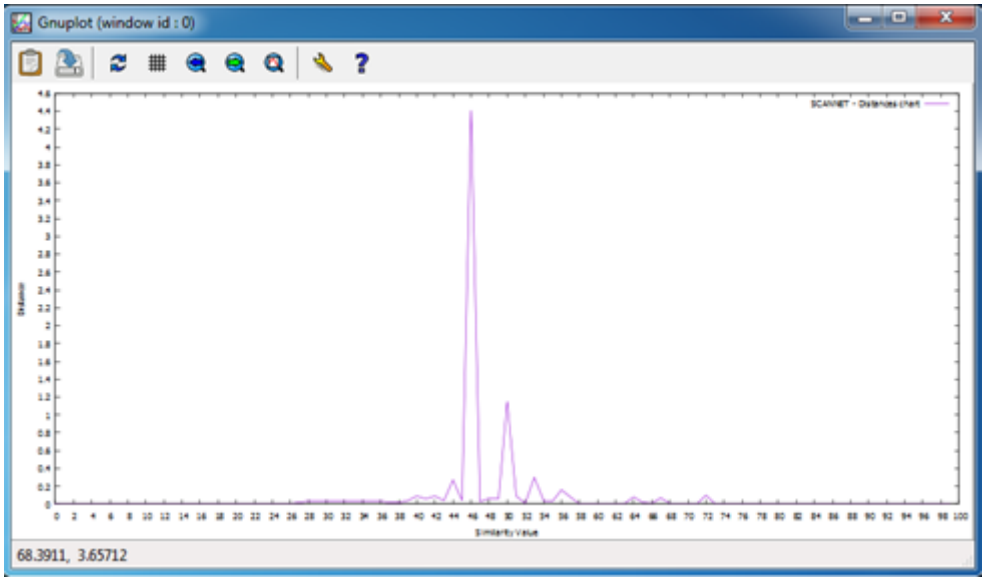

Figure 5 - Chart with all distances between pairs of neighboring networks

The chart is characterized by a series of peaks, where each peak represents a point at which the obtained networks has greater chances of suffer important alterations topological [1], or to be divided in new communities. After the analysis of the graph the user can select the similarity value to be analyzed.

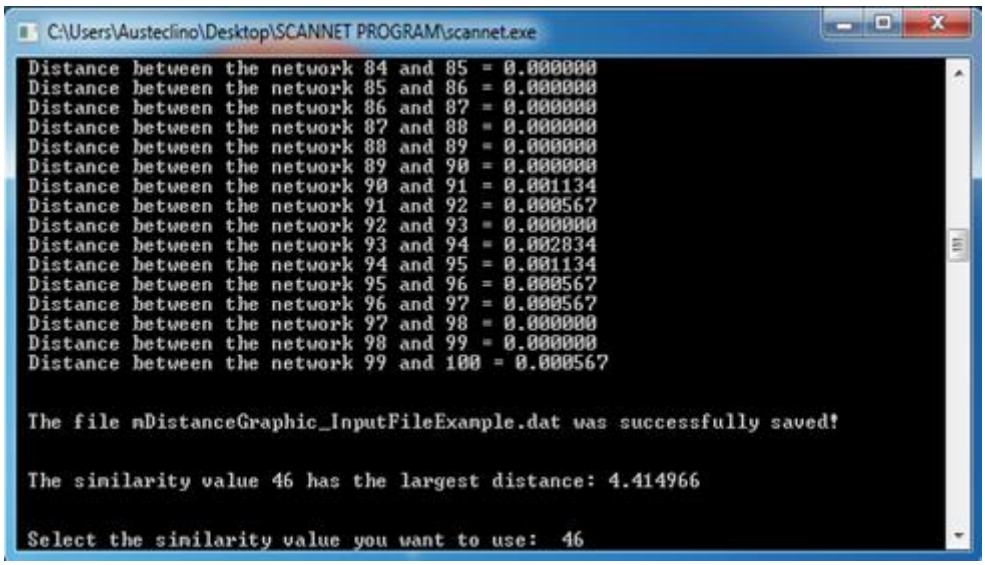

Figure 6 - Selecting the similarity value

* For the SCANNET displays the chart with all distances between pairs of neighboring networks is necessary you have installed on your computer GNUPLOT software, available at: http://www.gnuplot.info/.

3.4. END OF EXECUTION

At the end of the run, the SCANNET generates the critical network, the dendrogram, the color representation of the neighborhood matrix and the ordered proteins, each one in a different output file. It is also generated a log file with all the information of program execution. The output files can be read by several graphical analysis softwares such as Origin Lab®, Matlab®, GNU Plot, etc.

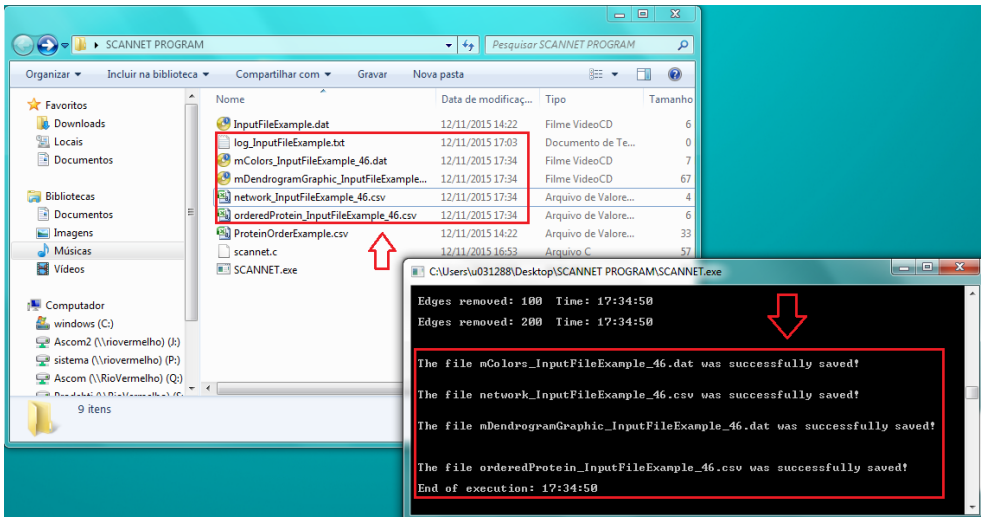

Figure 7 - Generating output files

4. REFERENCES

1. Andrade, RFS. Rocha-Neto, IC. Santos, LBL. Santana, CN. Diniz, MVC. Lobão, TP. Goés-Neto, A. Pinho, STR. El-Hani, CN. **Detecting Network Communities: An application to Phylogenetic Analysis.***PLoS Computational Biology*. 2011;**7**: e1001131 doi:10.1371/journal.pcbi.1001131.