



Community Finding in Phylogenetic Networks using PHYLOViZ Online

Luís Rita¹

Supervision Team: Bruno Gonçalves³, Alexandre Francisco^{1,2}, João Carriço³

(1) Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

(2) INESC-ID, R. Alves Redol 9, 1000-029 Lisboa, Portugal

(3) Instituto de Microbiologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

Master Thesis

The aim of the thesis was to develop an improved version of PHYLOViZ Online by implementing **community finding algorithms**^[1], as well as, **adding new tools for data visualization**. Similar algorithms are being used in other domains. In phylogenetics, it is a step forward for handling infectious diseases.

This tool is intended to assist distinct health professionals, including doctors and bioinformaticians, and is designed to enable medical and research purposes.

Introduction

Microbial typing allows us to determine how similar multiple strains are and to infer a possible evolutionary path. We may trace the source of an outbreak, infer about the pathogenicity/virulence of a recently discovered strain or, based on serotype data or other phenotypic data, to determine the most suitable strains to be targeted for vaccination.

With the advent of high-throughput sequencing methods, new ways for visualizing and analysing increasingly amounts of data are needed. Although some software already exist, they do not scale well for big data or require advanced technical knowledge to be useful in phylogenetics. PHYLOViZ, along with the tools under development, will provide a brand-new insight into the field of microbiology.

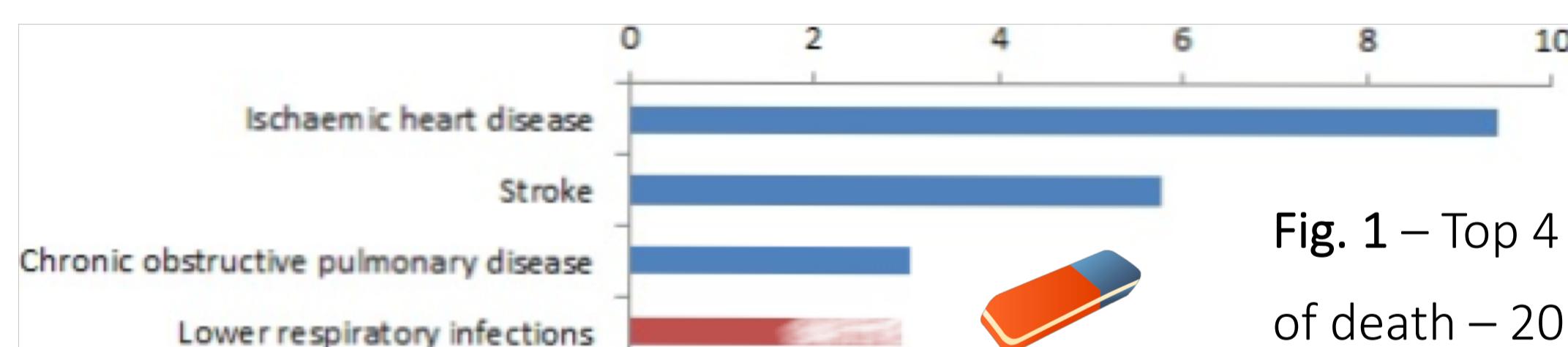


Fig. 1 – Top 4 global causes of death – 2016.^[2]

Implementation

PHYLOViZ Online is a Node.js application widely supported by most browsers. It uses goeBURST to trace Minimum Spanning Trees for different data type inputs.

Along with this user-friendly application, a RESTful API was also created. It allows to retrieve and upload data, and run available tree algorithms on it.^[3]

Visualization

Currently, the user may display data in several ways: Minimum Spanning Trees (MSTs), tables, sequences, charts or distance matrices. Next, it is presented how PHYLOViZ can be used to analyse a set of *Streptococcus pneumoniae* strains.

Streptococcus pneumoniae

Gram-positive bacteria, usually found in pairs.

It asymptotically resides in healthy carriers, typically, colonizing respiratory system.

It is the major cause of pneumonia and meningitis in children and the elderly.^[4]



Fig. 2 – *S. pneumoniae*.^[5]

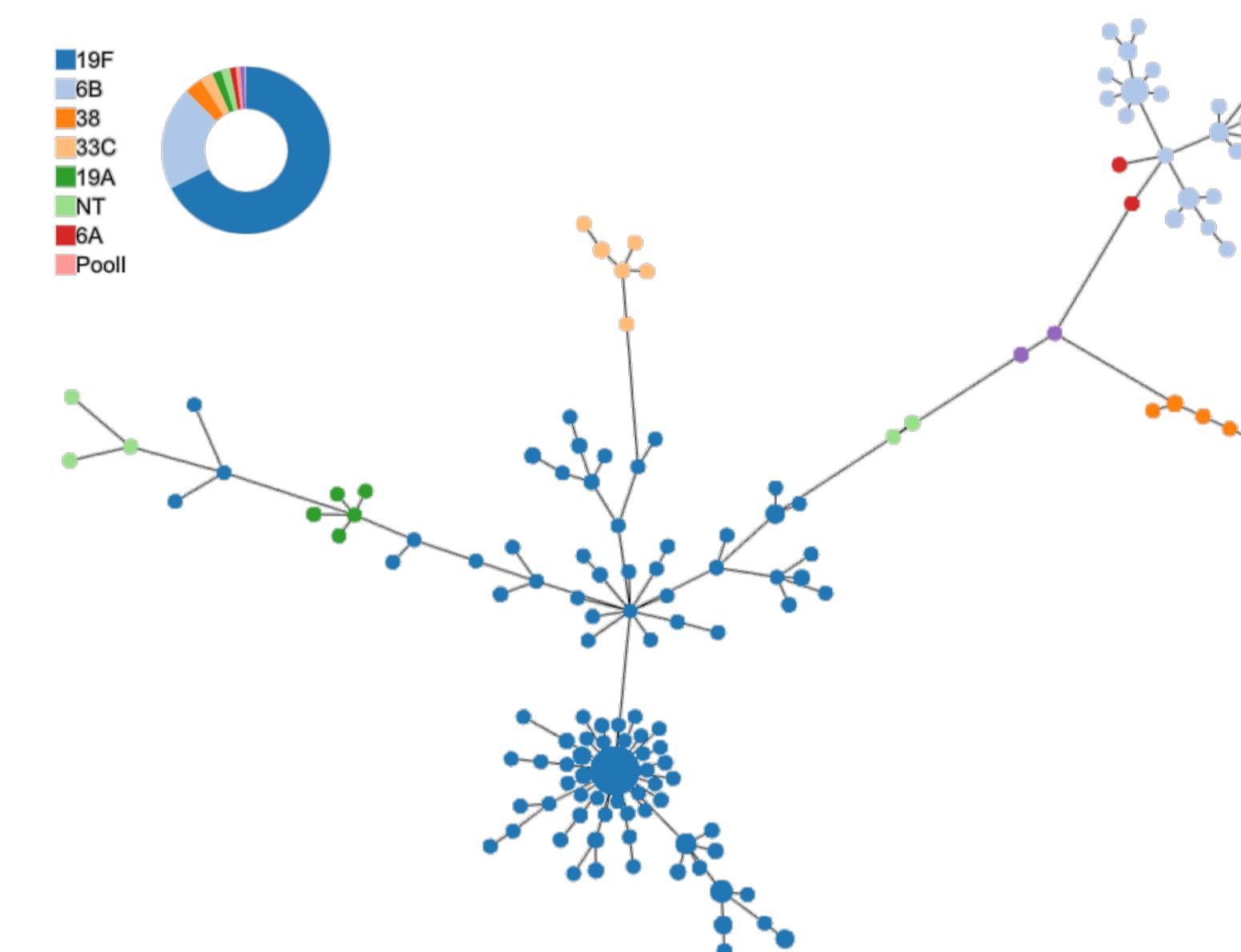


Fig. 3 – 145 strains of *S. pneumoniae*. Profile data composed of 237 loci based on a cgMLST scheme. Nodes coloured according to serotype. Length of each link proportional to strain dissimilarity.

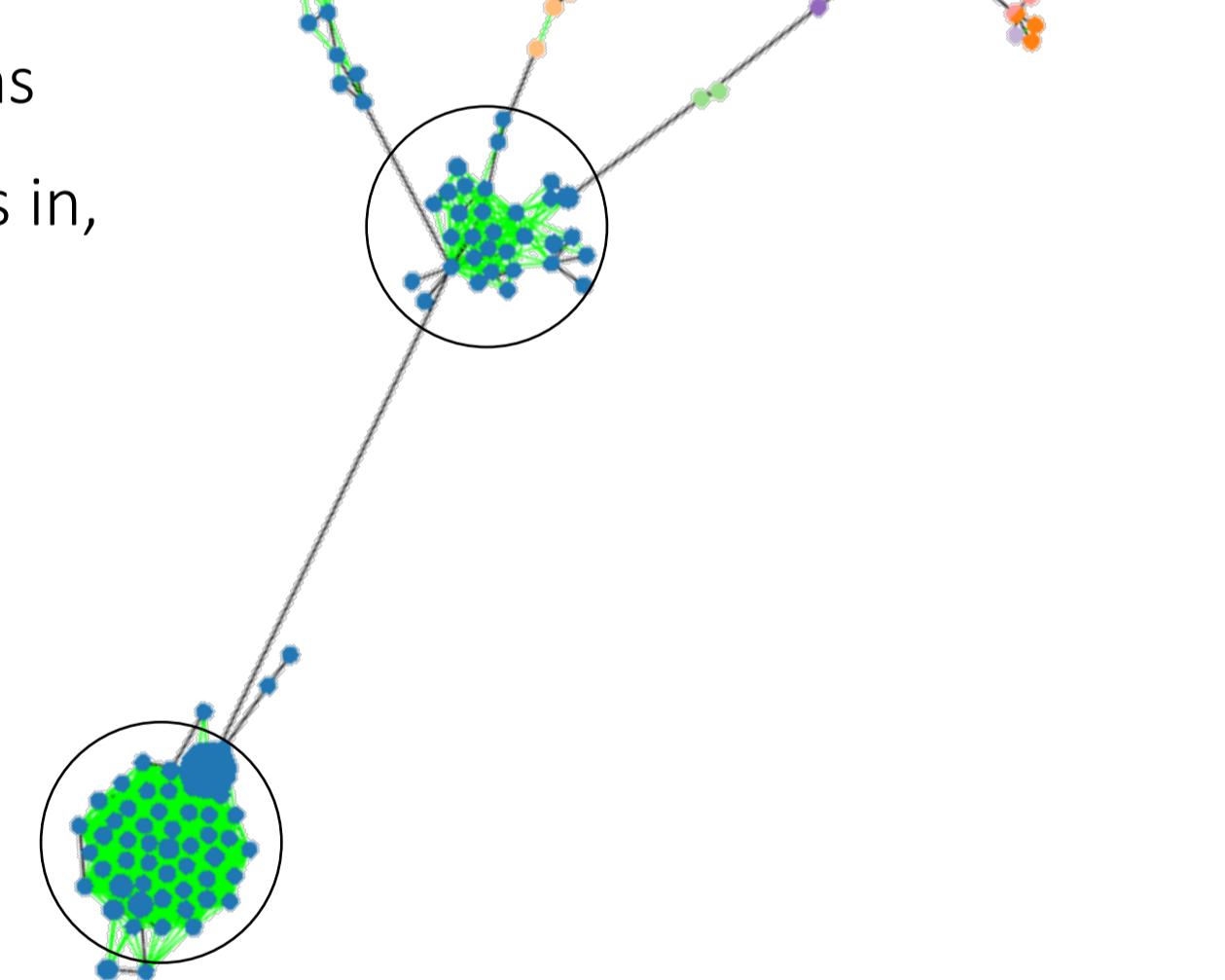


Fig. 4 – n Locus Variant option connects all strains until an input distance. Aim of the thesis consists in, programmatically, identifying each of these represented partitions, guiding the choice of threshold and to provide new ways to attribute nomenclature to clusters of related strains.

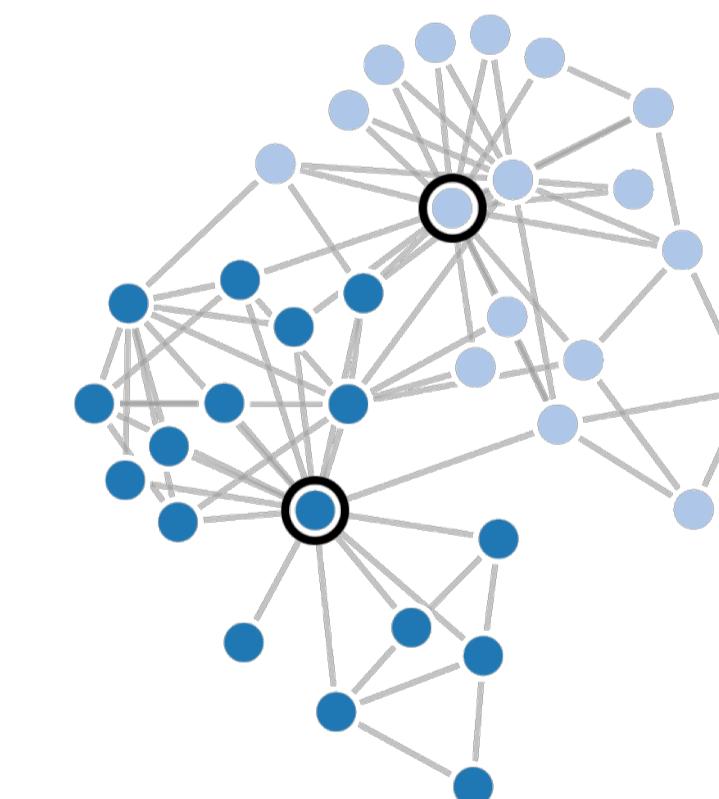


Fig. 5 – Zachary's karate club. An analogy that relates to the problem we pretend to solve in a microbiological context (i.e., data partitioning).

References

- Albert-László Barabási Books. Network Science. <http://networksciencebook.com>. Accessed in October 2018;
- WHO. The top 10 causes of death. <http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. Accessed in October 2018;
- Bruno Ribeiro-Gonçalves, Alexandre P. Francisco, Cátia Vaz, Mário Ramirez, João André Carriço; PHYLOViZ Online: web-based tool for visualization, phylogenetic inference, analysis and sharing of minimum spanning trees, *Nucleic Acids Research*, Volume 44, Issue W1, 8 July 2016, Pages W246-W251, <https://doi.org/10.1093/nar/gkw359>;
- Wikipedia. *Streptococcus pneumoniae*. https://en.wikipedia.org/wiki/Streptococcus_pneumoniae. Accessed in October 2018;
- CDC. Pneumococcal Disease. <https://www.cdc.gov/pneumococcal>. Accessed in October 2018.

