

Community Detection in Bacterial and Viral Networks

Jonas Ziegler

February 20, 2026

Abstract

This paper investigates community structure in bacterial and viral protein similarity networks using modularity-based community detection methods. We compare several algorithms and perform a sensitivity analysis on the resolution parameter using Adjusted Rand Index and Adjusted Mutual Information.

1 Introduction

1.1 Motivation

We live in a highly interconnected world where many physical, social, technological and biological systems consist of agents or entities interacting with each other. Examples include a virus being transmitted over social contact networks, global trade between countries, and the human brain. Any such system can be represented as a network by denoting the entities as nodes and the interactions between them as edges. This makes networks an important type of data spanning a remarkable variety of complex systems. It is therefore very important to have mathematically rigorous and practically useful methods for statistical analysis of networks. In this paper, we will focus on Community Detection in biological networks, in particular in bacterial and viral protein similarity networks.

A protein-protein interaction network (PPI network), also known as the interactome, is a type of biological network where a node represents a protein and an edge represents the physical interaction between each pair of proteins. These interactions are mapped using experimental techniques, such as yeast two-hybrid or mass spectrometry. Newman (2006) argues that community detection is one of the most important problems in the study of networks, and that it is particularly relevant for biological networks.

References

Newman, M. E. J. (2006). “Modularity and community structure in networks”. In: *PNAS* 103.23, pp. 8577–8582.