Supplementary material for the manuscript entitled "Prediction of GO Terms Based on Partitioning PPI Networks into Highly Connected Components", ID: TCBB-2022-11-0938

Milana Grbić, Branislava Gemović, Savka (ex. Janković) Vračević, Radoslav Davidović, Dragan Matić,Aleksandar Kartelj

May 27, 2025

This document contains supplementary material supporting the manuscript titled *"Prediction of GO Terms Based on Partitioning PPI Networks into Highly Connected Components"*. The additional content provided herein includes extended experimental results, supplementary tables and figures, detailed examples of the GO term prediction process, as well as further elaboration on the methodology and dataset descriptions.

1 Biological interpretation of metabolite HCD components

As it is stated in main paper, some of the HCD components obtained by partitioning given metabolic networks are analyzed from the biological point of view. More precisely, the process of fatty acid synthesis is recognized in HCD components of all four metabolic networks. This process consists of three consecutive phases: activation, elongation and termination [1, 2]. In each of the phases, some of the intermediates are repeated, so in some organisms the metabolites participating in these phases are denser, whereas in others they are less frequently related. Based on metabolic networks, a network corresponding to one of the phases of fatty acid synthesis is obtained as an HCD component for all four organisms. Figure 1 shows one of the highly related components of the *Saccharomyces cerevisiae* organism. This HCD component represents the termination phase of the fatty acid synthesis process. On the other hand, the HCD component of the metabolic network *Staphylococcus aureus* is a structure which represents the first two phases of the fatty acid synthesis process, i.e., the phases of activation and elongation (Figure 2). Unlike the organism *Staphylococcus aureus*, where the intermediates of the first two phases of the fatty acid synthesis process are connected into one HCD component, the activation phase for the organism *Tuberculosis – Mycobacterium tuberculosis* is separated into a separate component (Figure 3). For the organism *Escherichia coli*, one of the HCD components represents the elongation phase of the fatty acid synthesis process (Figure 4).

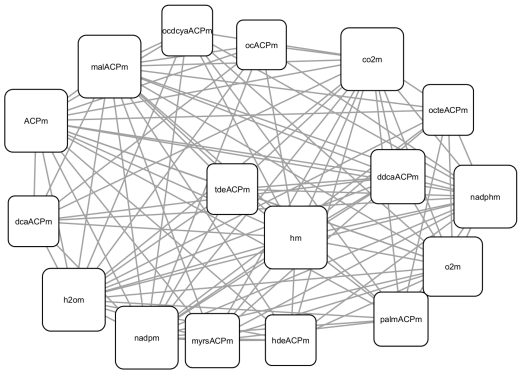

Figure 1: The termination mechanism in fatty acid synthesis in *Saccharomyces cerevisiae* network

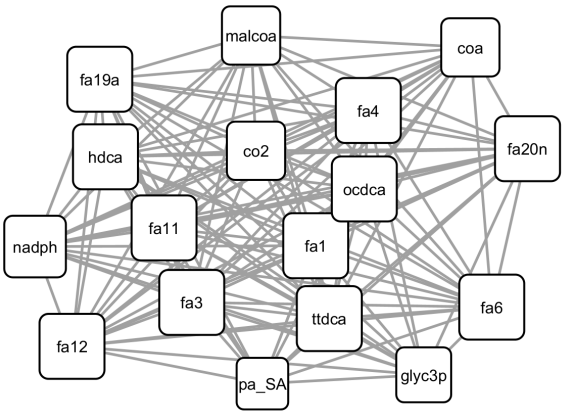

Figure 2: Activation and elongation phases in fatty acid synthesis in *Staphylococcus aureus* network

These findings are in the line with conclusions from paper [3]. More precisely, in paper [3] a network is partitioned into subgraphs, called k-plexes where the degree of each vertex within the k-plex with n vertices is at least $n - k$. In the same study is demonstrated that relaxing the partitioning requirements by varying k can lead to more descriptive and biologically relevant results. Since, the HCD components can be considered as k-plexes with $k = n/2 - 1$, they represent more relaxed partitioning scheme which can lead to more descriptive and biologically relevant results. It is important to mention that in the paper [3] solely to the metabolic network of *Saccharomyces cerevisiae* (yeast) was considered, while in our study, we extend the analysis to include multiple organisms. It is important to note that the k-plex partitioning has not been applied to PPI networks, which prevents us from making a direct comparison with methods of that type.

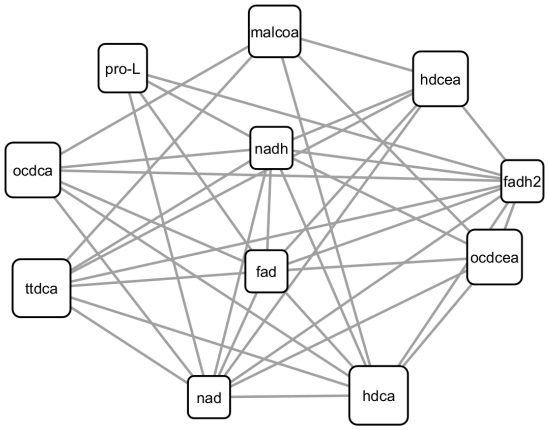

Figure 3: The activation phase in fatty acid synthesis in *Tuberculosis –Mycobacterium tuberculosis* network

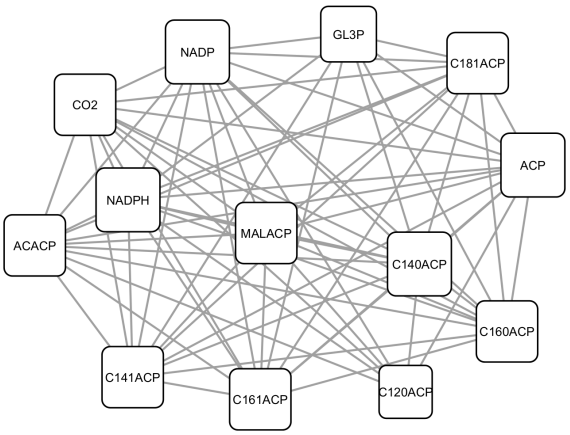

Figure 4: The elongation phase in fatty acid synthesis in *Escherichia coli* network

2 Parameters of VNS and GOTerm predition methods

In this section, we provide an overview of the key parameters used in the algorithms. A detailed description of each parameter, along with its value is shown in Tables 1 and 2, is provided to ensure clarity and to allow for reproducibility of the experiments.

References

[1] V. M. Anoop, U. Basu, M. T. McCammon, L. McAlister-Henn, and G. J. Taylor, "Modulation of citrate metabolism alters aluminum tolerance in yeast and transgenic

Parameter	Description	Value
k_{min}	Minimum size of neighborhood in Shaking	1
k_{max}	Maximum size of neighborhood in Shaking	20
$iter_{total}$	Total iterations performed in VNS	10000
$iter_{withoutImpr}$	Maximum iterations without improvement before stopping	2000
$prob$	Probability switching to a new solution when objective functions are equal	0.5
max_time	Maximum time allowed for the algorithm to run	4 hours

Table 1: VNS Parameters

Parameter	Value
Number of repetitions	100
Number of different values for **Score1**	10
Number of different values for **Score2**	10
Range for **Score1**	0.001 – 0.05
Step for **Score1**	0.0054
Range for **Score2**	0.001 – 0.999
Step for **Score2**	0.11
Metric used for evaluation	F1 measure

Table 2: Parameters for GO Prediction Method

canola overexpressing a mitochondrial citrate synthase," *Plant physiology*, vol. 132, no. 4, pp. 2205–2217, 2003.

[2] O. Tehlivets, K. Scheuringer, and S. D. Kohlwein, "Fatty acid synthesis and elongation in yeast," *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, vol. 1771, no. 3, pp. 255–270, 2007.

[3] M. Grbić, A. Kartelj, S. Janković, D. Matić, and V. Filipović, "Variable neighborhood search for partitioning sparse biological networks into the maximum edge-weighted k k-plexes," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 17, no. 5, pp. 1822–1831, 2019.