Supplementary information

This document contains supplementary information to the manuscript entitled "Discovering regulatory and signaling circuits in molecular interaction networks" by Trey Ideker, Owen Ozier, Benno Schwikowski, and Andrew F. Siegel.

It particular, we present a sketch of an NP-hardness proof for a simplified variant of the central search problem. The proof was provided by Richard M. Karp.

The simplified search problem, termed *Maximum-Weight Connected Subgraph Problem (MWCSP)*, for which NP-hardness is proved, has two simplifying characteristics compared to the search problem discussed in the paper.

1. The MWCSP corresponds to the special case of a single experimental condition, where each gene has only a single associated score.

2. Compared to the more complex score function for subnetworks in our manuscript, the MWCSP assumes that the score of a subnetwork is given by the sum of all its node scores.

Maximum-Weight Connected Subgraph Problem Given a graph $G = (V, E)$, and a vertex weight $z_v \in \mathbb{R}$ for each $v \in V$, find a connected subgraph $G' = (V', E')$ of G with maximum weight $z_{G'} = \sum_{v \in V'} z_v$.

Theorem 1 *The Maximum-Weight Connected Subgraph Problem is NP-hard.*

Proof sketch: The optimization version of MINIMUM COVER can be reduced to MWCSP. We will describe how to construct a graph for any instance of MINIMUM COVER. We will show how, in this graph, any optimal solution to MWCSP corresponds to an optimal solution of MINIMUM COVER.

To begin, let $C = \{C_1, \ldots, C_m\}, m \in \mathbb{N}$, be an instance of MINIMUM COVER. Recall that a solution to the optimization version of the problem is a minimum subset $C' \subseteq C$, such that each element of $S = \cup_{i=1}^{m} C_i$ is contained in at least one subset in C'.

Denoting the elements of S by $s_1, \ldots, s_n$, the weighted graph $G = (V, E)$ is constructed as follows. The node set V of G is given by

$$V = \{s_1, \ldots, s_n, C_1, \ldots, C_m, H\}$$

G contains two sets of edges. First, H is connected to any node C_i, $i = 1, \ldots, m$. Secondly, each node s_j is connected to any node C_i, for which $s_j \in C_i$. Weights are -1 on all nodes C_i, and of sufficiently high value (denoted by ∞) for all the other nodes. The figure below gives an example for the MINIMUM COVER input instance $\{\{s_1, s_2\}, \{s_1\}, \{s_2\}, \{s_2, s_3\}\}$.

Observe that any connected subgraph $G' = (V', E')$ of G with maximum weight contains H and all nodes s_i, due to their high weight. Thus, any connected subgraph can be identified by the set of nodes C_j contained in it.

The critical observation that establishes the proof is that any sets of C_j contained in a maximum-weight connected subgraphs of G is a minimum cover

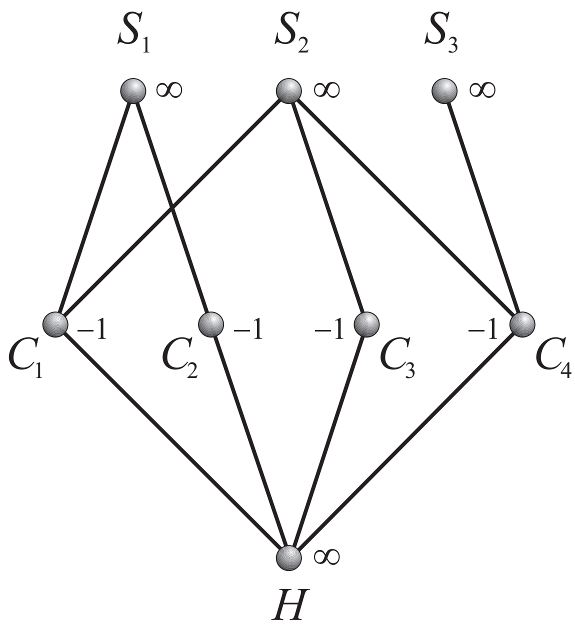

Figure 1: Graph construction for MINIMUM COVER example instance

with respect to the MINIMUM COVER instance, and vice versa. To see why this is the case, note that, since H is included in G', connectedness of G' is equivalent to the requirement that, for each node s_i, there is at least one node $C_j \in V'$ such that $s_i \in C_j$. Maximizing the weight of G' is equivalent to minimizing of the C_j nodes, which, in turn, is equivalent to minimizing the set cover represented by the C_j.

Proof sketch writeup by Benno Schwikowski, benno@systemsbiology.org, Jan. 27, 2002