

**Figure 1**

Complex Overlap Decomposition. A simplified illustration of the Complex Overlap Decomposition (COD) method. An edge, (3, 4), connecting a pair of weak siblings is added to the graph. A fill-in edge between proteins 5 and 8 is added to eliminate all five 4-cycles in the graph: {5, 6, 8, 7}, {1, 5, 7, 8}, {2, 5, 7, 8}, {1, 5, 6, 8}, and {2, 5, 6, 8}. If the modified graph is chordal, all clique tree representations are computed (cf. Methods). Each clique tree representation results in a Tree of Complexes representation, where the Tree of Complexes is constructed by projecting each maximal clique in the modified graph, G^* , to a functional group in the original graph G . For example, a four node maximal clique, {1, 2, 5, 8}, in G^* is projected to a four node functional group in G , by removing a fill-in edge (5, 8). Each functional group is represented by a Boolean expression, such as $(1 \wedge 2) \wedge (5 \vee 8)$, which means that the functional group contains two variants of a complex, {1, 2, 5} and {1, 2, 8}.

Cographs are another well-studied graph family [20]. A *cograph* can be characterized by an absence of an induced subgraph which is a path of length four (P_4), where the length is the number of nodes in the path. Thus, the diameter of a connected cograph is at most two. Subsequently, connected cographs are dense and cliquish, consistently with the assumption made by algorithms that delineate protein complexes. What makes cographs even more attractive is that for every cograph there exists a Boolean

expression which describes all the maximal cliques in the graph. (In terms of modular decomposition used in [16] it means that a cograph can be decomposed by modular decomposition without leaving non-trivial non-decomposable prime module.) This Boolean expression describes in a compact and hierarchical way all the possible variants of protein complex within a functional group.