Homogeneous decomposition of protein interaction networks: refining the description of intra-modular interactions

Géraldine Del Mondo, Damien Eveillard and Irena Rusu*

Computational Biology group (ComBi) - LINA, Université de Nantes, CNRS UMR 6241, 2 rue de la Houssinière, 44300 Nantes, France

Received on March 28, 2008; revised and accepted on February 10, 2009

Advance Access publication February 17, 2009

Associate Editor: Burkhard Rost

ABSTRACT

Motivation: Modules in biology appeared quickly as an accurate way for summarizing complex living systems by simple ones. Therefore, finding an appropriate relationship between modules extracted from a biological graph and protein complexes remains a crucial task. Recent studies successfully proposed various descriptions of protein interaction networks. These approaches succeed in showing modules within the network and how the modules interact. However, describing the interactions within the modules, i.e. intra-modular interactions, remains little analyzed despite its interest for understanding module functions.

Results: We overcome this weakness by adding a complementary description to the already successful approaches: a hierarchical decomposition named homogeneous decomposition. This decomposition represents a natural refinement of previous analyses and details interactions within a module. We propose to illustrate these improvements by three practical cases. Among them, we decompose the yeast protein interaction network and show reachable biological insights that might be extracted from a complex large-scale network.

Availability: A program is at disposal under CeCILL license at: www.lina.univ-nantes.fr/ombi/DH/Home.html

Contact: irena.rusu@univ-nantes.fr

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

By essence, biological systems are not fully explained. They appear as complex systems which emphasize our incapacity to understand the relation between inputs and outputs of the living system (Szallasi et al., 2006). Evidence of modules in biology and utilities of such a concept quickly appeared as an accurate way to summarize complex living systems with simple ones. As an illustration, a molecular complex abstracts numerous and complex interactions of proteins. Using a module description implies to replace some part of the system with an abstraction that maintains a correct property with the given experimental data. This modeling approach introduces the concept of modularity such as it was expressed clearly by Hartwell et al. (1999). Applied on protein interaction networks, these top-down approaches emphasize molecular hubs or functional components within the network (Szallasi et al., 2006). In other words, they find information of interest within a graph structure while describing its modules.

Many studies aim at discovering this kind of information within the structure of biological graphs (Jeong et al., 2000). In particular, Spirin and Mirny (2003) give a strong support in such protein interaction network analyses. They show theoretical modules extracted from the network that correspond to protein complexes (splicing machinery, transcription factors, etc.) or dynamic functional units that can belong to the cell-cycle regulation. However, in silico techniques appear as very sensitive to the completeness of the protein interaction set. Finding biological modules is only efficient for already well-investigated protein–protein interaction graphs.

Due to intensive experimental investigations on Saccharomyces cerevisiae, the yeast protein interaction graph agrees with such a criterion. Thus, this graph quickly appeared as an accurate benchmark for testing protein interaction network analysis techniques (Guimerà et al., 2004; Hart et al., 2007; Ma et al., 2004). Based on tandem affinity purification/mass spectrometry (TAP-MS) experiments (Puig et al., 2001), various techniques aim at characterizing protein complexes [see Gavin et al. (2002, 2006); Ho et al. (2002) and Krogan et al. (2006) for illustration]. Among them, Hart et al. (2007) used the Markov Cluster Algorithm (MCL) technique developed by Enright et al. (2002). This is a statistical scoring-based approach that differentiates direct physical interactions from interactions mediated by other members of the complex. Based on various experiments, the authors emphasize the relevance of combining statistical analyses for inferring biological knowledge. In practice, their study clearly indicates a hierarchical organization of protein complexes in the cell and confirms that the yeast ‘complex-ome’ is almost fully described.

In this context, protein complexes act as biological modules and the method gives a robust overview of how biological modules interact. Nevertheless, it also highlights other interesting questions about the impact of specific interactions in the modular description (He and Zhang, 2006), which is particularly relevant. However, the MCL technique fails to answer these questions intuitively.

To overcome this technique weakness, we follow the assumptions of Hart et al. (2007). We consider the yeast protein interaction network as a graph with an (almost) complete set of protein interactions. Consequently, each unit protein complex (that is, not including smaller protein complexes) appears as a fully connected
part of the network. Note here that the reverse is not true: not every fully connected part of the network necessarily derives from an existing (unit or not) protein complex. Our approach aims at discovering in silico information that is hidden within the protein interaction network. It identifies unit protein complexes and the relations between them. Hierarchical graph decompositions present interesting features for describing complete and large-scale graphs such as the yeast protein interaction network. Therefore, we consider these decompositions as a natural theoretical framework for refining the description of protein complexes (equivalently, biological modules) obtained using the MCL technique. Following a similar assumption, Gagneur et al. (2004) apply the hierarchical decomposition named modular decomposition on protein interaction graphs. Various tests of this method show that theoretical modules obtained this way may correspond to protein complexes. They show as well that modular decomposition is not precise enough to capture several important features of biological systems. One main drawback of the modular decomposition is the existence of too large components when the modular decomposition is finished at all levels. Therefore, as observed using the MCL technique, important relationships between intra-modular components remain hidden in the network analysis.

Notwithstanding, we consider that the assumptions exposed by Gagneur et al. (2004) are convincing. Further investigations using hierarchical graph decompositions might complete the MCL results and show intra-modular interactions. We herein propose to extend the analysis involving modular description by using a natural refinement of the method, called homogeneous decomposition. We precisely explain that this decomposition improves the network partitioning (see Section 2). It hence allows us to go further into our biological purposes by (i) identifying smaller significative components of the network and (ii) showing up their detailed interactions with the other components. We illustrate these theoretical features by an application on various protein interaction networks (see Section 3). We first describe results on a theoretical protein interaction network. It shows various modular insights, emphasized by the homogeneous decomposition (see Section 3.1). Second, we illustrate the improvements obtained with homogeneous decomposition on known complexes, the transcriptional regulator complexes, already analyzed (Gagneur et al., 2004) through modular decomposition (see Section 3.2). Finally, we apply the homogeneous decomposition on the yeast protein interactions (see Section 3.3) that represents an accurate realistic benchmark supporting our method.

2 METHODS

A graph is a data structure used for representing objects and their pairwise relationships. The objects are the vertices of the graph, while the pairwise (undirected) relationships between objects are the edges. We herein assume a protein–protein interaction network as a graph whose vertices and edges are, respectively, the proteins and their pairwise interactions. The structure of this graph provides a lot of information on the groups (or complexes) of proteins that act together to fulfill a biological function. To discover the organization of a graph, one usually uses graph decompositions. To store and analyze this organization, one uses decomposition trees.

Graph theory provides various ways to decompose a graph. Many of them are single-level decompositions because they partition the graph into several components that are not partitionable themselves. In contrast, multi-level (or hierarchical) decompositions have the major advantage of allowing an iterative study of the structure by fitting the components into each other.

The modular decomposition, also known as substitution (Mohring and Radermacher, 1984) or X-join (Habib and Maurer, 1979) decomposition, is probably the most well-known hierarchical decomposition of graphs. It was independently discovered several times [see Mohring and Radermacher (1984) for a review, and various very efficient algorithms [see for instance, McConnell and Spinrad (1994)] exist to compute it. As a natural refinement of the modular decomposition, Jamison and Olaria (1995) propose the homogeneous decomposition, for which Baumann (1996) describes an efficient algorithm. Both modular and homogeneous decompositions build a decomposition tree (see Fig. 1 for illustration), whose leaf nodes are proteins and whose internal nodes (called modules) represent logical rules to combine the child nodes.

Modular and homogeneous decomposition are explained below and illustrated on the graph G in Figure 2a. In this description, the notion of adjacent vertices (or simply neighbors) is paramount. It designates two vertices of G joined by an edge. A graph whose vertices are all neighbors to each other is called a clique.

2.1 Modular decomposition model

Under the modular decomposition model, a module $M$ is a graph inside the given graph $G$ so that all the vertices in $M$ have exactly the same neighbors outside $M$ (let us call that the neighborhood property). The aim of the modular decomposition is to decompose a graph into non-trivial modules (at least two), and then to iterate the decomposition process on the resulting modules until all modules are made of one vertex (such modules are the leaves of the decomposition tree). Thus, the children of each node in the modular decomposition tree (Fig. 1) are its modules, whether they are internal vertices or leaves. Figure 2b shows one decomposition of $G$ in Figure 2a into modules. Module $\beta$ may be furtherly decomposed into modules with vertex sets $\{4, 6\}$ and $\{9\}$, while all the other modules have only trivial decompositions into 1-vertex modules (i.e. leaves).

When a graph is broken up into modules, one must be able to build the graph again using only its modules and a logical rule (otherwise the decomposition looses information). The logical rule is stored in the node of the tree corresponding to the graph as a character with values $0, 1$ or $P$ as

![Fig. 1. Decomposition tree of an arbitrary graph $G$ with modules $\alpha, \beta$ and $\gamma$.](image1)

![Fig. 2. (a) Example graph $G$ and (b) its decomposition into modules $\alpha, 1, \beta, 5$ and $\gamma$.](image2)
The main drawback of the modular decomposition is its inability to further decompose the characteristic graphs associated to the P-modules. The homogeneous decomposition partially solves this problem by identifying P-modules with a specific structure, for which a further decomposition is proposed. It is not meant to replace the decomposition into modules but to refine it, once the modules have been computed and the characteristic graph has been built. The homogeneous decomposition offers, therefore, an obvious qualitative improvement to the modular decomposition, which is described here in an intuitive manner and is illustrated in examples. Notice here that we slightly modified the definition of the homogeneous decomposition tree compared with the original one by Jamison and Olariu (1995), so as to make it easier to handle and to explain.

A graph \( G \) further decomposable by an homogeneous decomposition is called a W-graph (or W-module) in the remainder of the article. It has the wheel structure depicted in Fig. 4a. Such a module has a characteristic graph (Fig. 4b) made of a hub (which is a clique) and of a set of single vertices around the hub that have neighbors in the hub but are not joined to each other. The h-modules of a W-module are the graphs (which are also cliques) made of a single vertex and all its neighbors in the hub. Note that h-modules do not have the neighborhood property as other modules do. These notions are illustrated on the example graph \( G \) in Fig. 4a and b.

The homogeneous decomposition introduces two new logical rules in the decomposition tree (Fig. 5c), described by characters W and H used to label internal nodes:

- W means that the graph \( G \) is obtained from its modules (stored as children) and its h-modules (stored as a specific child which is an H-node) by recovering a wheel structure. This happens when \( G \) is a W-graph (or W-module) and in this case its corresponding node is called a W-node. Graph \( G \) in Fig. 2a is a W-module whose corresponding W-node is shown in Figure 5c.

- H means that the internal node stores the h-modules of its father, which is necessarily a W-module, in the following form: each h-module is stored in a child labeled by a vertex set whose first element is the single vertex identifying the h-module and the other elements are its neighbors in the hub. In this case, the node is called an H-node. Although this is not necessary, in our figures and so as to simplify explanations, the vertex set of the hub labels the H-node. The h-modules of the characteristic graph \( C(G) \) of the example graph \( G \), identified in Fig. 5b, are stored in the H-node in Fig. 5c.

### Homogeneous decomposition

The main drawback of the modular decomposition is its inability to further decompose the characteristic graphs associated to the P-modules. The homogeneous decomposition partially solves this problem by identifying P-modules with a specific structure, for which a further decomposition is proposed. It is not meant to replace the decomposition into modules but to refine it, once the modules have been computed and the characteristic graph has been built. The homogeneous decomposition offers, therefore, an obvious qualitative improvement to the modular decomposition, which is described here in an intuitive manner and is illustrated in examples. Notice here that we slightly modified the definition of the homogeneous decomposition tree compared with the original one by Jamison and Olariu (1995), so as to make it easier to handle and to explain.

A graph \( G \) further decomposable by an homogeneous decomposition is called a W-graph (or W-module) in the remainder of the article. It has the wheel structure depicted in Fig. 4a. Such a module has a characteristic graph (Fig. 4b) made of a hub (which is a clique) and of a set of single vertices around the hub that have neighbors in the hub but are not joined to each other. The h-modules of a W-module are the graphs (which are also cliques) made of a single vertex and all its neighbors in the hub. Note that h-modules do not have the neighborhood property as other modules do. These notions are illustrated on the example graph \( G \) in Fig. 4a and b.

The homogeneous decomposition introduces two new logical rules in the decomposition tree (Fig. 5c), described by characters W and H used to label internal nodes:

- W means that the graph \( G \) is obtained from its modules (stored as children) and its h-modules (stored as a specific child which is an H-node) by recovering a wheel structure. This happens when \( G \) is a W-graph (or W-module) and in this case its corresponding node is called a W-node. Graph \( G \) in Fig. 2a is a W-module whose corresponding W-node is shown in Figure 5c.

- H means that the internal node stores the h-modules of its father, which is necessarily a W-module, in the following form: each h-module is stored in a child labeled by a vertex set whose first element is the single vertex identifying the h-module and the other elements are its neighbors in the hub. In this case, the node is called an H-node. Although this is not necessary, in our figures and so as to simplify explanations, the vertex set of the hub labels the H-node. The h-modules of the characteristic graph \( C(G) \) of the example graph \( G \), identified in Fig. 5b, are stored in the H-node in Fig. 5c.
We propose to illustrate the above notions on the small simplistic network from Wilhelm et al. (2003). (a) Protein complexes elucidated with TAP-MS. (b) The associated interaction network, where all possible interactions within the protein complexes are considered. (c) The corresponding modular decomposition tree. (d) The homogeneous decomposition tree that highlights modules and their intra-modular interactions. The gray area indicates the hub, stored in the H-node, that connects the other h-modules as leaves.

The homogeneous decomposition is described precisely through a careful interpretation of their vertices representing modules. In Figure 5c, the h-module with vertex set \( \{a,1\} \) yields protein complexes \( \{2,1\} \) and \( \{3,1\} \), since module \( a \) is a \( \emptyset \)-module. Therefore, proteins 2 and 3 are alternatives. The h-module with vertex set \( \{y,1,b\} \) contains the hub, thus, the hub itself does not generate specific protein complexes. After the interpretation of modules \( y \) and \( b \), the protein complexes generated by \( \{y,1,b\} \) are \( \{7,8,1,4,6\} \) and \( \{7,8,1,9\} \).

### 3 RESULTS

Like many tools in many fields, modular and homogeneous decompositions might show very useful insights when applied in the appropriate context. The appropriate context, in this case, is an (almost) complete protein interaction network, where each unit protein complex is represented as a clique, due to the method used to infer the protein interaction network. Note here that exceptions to this constraint may either seriously or weakly damage the decomposition, depending on the nature of the exception. Network analysis aims at finding (i) the hierarchical structure of the network, (ii) the relations between complexes (inclusion, disjunction, overlapping), and eventually (iii) the proteins or groups of proteins within the network that play a central role in specific regions of the network. In this purpose, the decomposition tree has to clearly represent the complexes and highlight their relationships, which gives emphasis to the interpretation of each type of node in the decomposition tree.

#### 3.1 Theoretical protein interaction network

We propose to illustrate the above notions on the small simplistic protein complex network shown by Wilhelm et al. (2003) (Fig. 6). This network comes from TAP and HMS-PCI techniques. Once the modules \( \alpha, \beta, \gamma \) and \( \delta \) are identified by modular decomposition, a wheel structure appears, with hub \( \{h, \alpha\} \) and three h-modules. In practice, the hub does not represent a concrete protein complex (since the hub is part of the h-module \( \{h, \alpha, \delta\} \)), but shows the central role of its components in the intra-modular interaction description. Each h-module (which is a clique) yields one or more intra-modular complexes. As an example, the h-module \( \{h, \alpha, \delta\} \) yields the larger clique \( \{h, a, b, i, g\} \) after interpretation of modules \( \alpha \) and \( \delta \). This clique, obtained by our theoretical approach, correctly identifies the complex \( Y \). Similarly, the interpretation of the two other h-modules \( \{\gamma, h\} \) and \( \{\beta, \alpha\} \) conduces to recover cliques with vertex sets \( \{k, j, h\} \) and \( \{c, d, f, e, b, a\} \), that respectively correspond to the complexes \( Z \) and \( X \). Therefore, our theoretical approach perfectly identifies, on this network, the known complexes, thus showing a great accuracy. Moreover, the homogeneous decomposition shows the relationships between these complexes. For instance, complex \( X \) and complex \( Y \) share the module \( \alpha \), that is the pair of proteins \( \{a, b\} \). Despite the fact that they do not build a proper complex, these proteins always have to be considered together from a functional viewpoint (since \( \alpha \) is a \( \emptyset \)-module).

#### 3.2 Transcriptional regulator complexes in yeast

We qualitatively validated our approach by comparing to the results of Gagneur et al. (2004), which used protein interactions from TAP-MS studies to define transcriptional regulatory complexes in yeast. It is composed of five complexes [see Cairns et al. (1994, 1996); Henry et al. (1994) and Kim et al. (1994) for details]: RSC, SWI/SNF (chromatin-remodeling complexes), TFII F, TFII D (general transcription factor complexes) and Mediator (the mediator complex that mediates signals to RNA polymerase II). The modular decomposition tree (Fig. 7a) identifies several modules but fails to identify the relations between the children of the P-node.

In contrast, the homogeneous decomposition tree (Fig. 7b) singularly identifies a W-module that is structured as a wheel, according to the information stored in the H-node. The hub is composed by the protein Anc1 and the module \( \alpha \) formed by Arp7 and Arp9. It plays a central role in the decomposition, since it generates all the interesting cliques in the network. This observation, exclusively based on our decomposition of the network, meets the biological knowledge, since either Anc1 or the module \( \alpha \) belong to the five complexes experimentally known (Fig. 7c). Further analysis argue for the reliability of the homogeneous decomposition and our interpretation of it. The decomposition indicates three children of the H-node that correctly identify all the complexes of the network and their interactions. Each complex is herein associated with the P-value obtained after a Gene Ontology analysis using Go::TermFinder (Boyle et al., 2004):

1. The rightmost child of the H-node represents an interaction between Anc1, as a component of the hub, and the module \( \beta \) that is a \( \emptyset \)-module. Consequently, Anc1 interacts alternatively with the components of \( \beta \). This interpretation emphasizes three alternative cliques: either Anc1, Taf40, Taf19, Taf45, Taf61, Taf90, Taf47, Tsm1, Taf25, Taf67, Taf17, Taf60 (100% cluster frequency; i.e., how many genes from the clique are annotated to the GO term associated with the cluster; P-value = 7.76e−28); or Anc1, Tfg1, Tfg2 (100%, 3.60e−10); or Anc1, Med4, Gal11, Nui2, Nut1, Med2, Med6, Med7, Pgd1, Cse2, Med8, Med11, Dnc1, Srb8, Srb4, Srb7, Srb6, Rgr1, Sin4, Srb2, Rox3 (90.5%, 5.08e−50). These three cliques correspond, respectively, to TFII F, TFII D and Mediator complexes. Based on the TAP-MS experiments,
A module like this one, not further decomposed, represents a subunit of the SWI/SNF complex. Moreover, the homogeneous decomposition achieves to see the relations between their sub-complexes, just by taking a glance at the homogeneous decomposition tree.

- The particular feature of the decomposition is the hub. It is a clique that yields the SWI/SNF complex (100%, 1.50e−20).
- The center child represents an interaction between the module α, as a component of the hub, and γ. The corresponding clique is interpreted as a clique that yields the SWI/SNF complex (90%, 3.94e−23).
- The third leaf indicates interactions between δ, α and Anc1, which is interpreted as a clique that yields the SWI/SNF complex (100%, 1.50e−20).

These complexes are already obtained by using the MCL analysis on TAP-MS data (Hart et al., 2007). However, such a statistical-based analysis does not show a precise description of the complexes and their sub-complexes. The homogeneous decomposition achieves to see the relations between their sub-complexes, just by taking a glance at the homogeneous decomposition tree.

The particular feature of the decomposition is the hub. It is composed by the module Arp7-Arp9 (namely α in Fig. 7b) and Anc1. A module like this one, not further decomposed, represents a sub-complex that is related to RNA polymerase II transcription factor activity (100%, 7.42e−6) and confirmed by experimental studies. Recently, Chen and Shen’s (2007) experiments show that Arp7 and Arp9 compose a crucial subunit of the SWI/SNF complex. Moreover, Szerlong et al. (2003) demonstrate that Arp7 and Arp9 form a stable heterodimer with the properties of a functional module. In particular, they emphasize its impact in both restructuration of chromatin and interactions between transcriptional regulatory complexes, like SWI/SNF and RSC.

Anc1 is the other component of the hub. Anc1 connects three modules to the hub, corresponding to TFIID, TFIIF and the Mediator complex. Following the decomposition tree interpretation, these complexes are alternative. This interpretation implies that Anc1 plays a major regulatory function by interacting with either TFIID, TFIIF or the Mediator complex. Kabani et al. (2005) confirm such an assumption. Experimental evidences indicate indeed that Anc1 is the only non-essential subunit of TFIID. It is also associated with TFIIF, although it is not required for its activity. Anc1 thus appears as not really essential for the proper functioning of complexes despite its overall importance on the whole network [based on its gene-deletion impact (Giaever et al., 2002)].

The three proteins that compose the hub are interacting with another module named δ. It represents the complex SWI/SNF. The interpretation of the decomposition tree indicates that this last complex is functionally independent from other modules, but might be modulated by the combination of Anc1, Arp7 and Arp9. Interestingly, Kabani et al. (2005) indirectly confirm this assumption by not showing a clear interaction of Anc1 alone with the SWI/SNF complex. It intuitively implies the need of another component that

Fig. 7. Hierarchical decompositions of the transcriptional regulator complexes network in yeast. (a) An output of the modular decomposition. (b) An output of the homogeneous decomposition. (c) The experimental knowledge about the known complexes [adapted from Gagneur et al. (2004)]. The hub, i.e. H-node, and its components are highlighted in gray.
we assume, based on the decomposition tree interpretation, being the catalytic subunits of Arp7 and Arp9.

To sum up, the homogeneous decomposition emphasizes complexes that are in accordance with those observed using other techniques on TAP-MS data. These complexes, not depicted by the modular decomposition, present accurate statistical results when compared with biological functions and cellular compounds accessed via Gene Ontology information (see Supplementary Material). As additional features, it indicates how complexes interact using an hub. The function of proteins that belong to the hub is to connect complexes, which provides their regulation. In particular, experiments confirm the role of Anc1 as a regulatory function by modulating the activity of the respective catalytic subunits of the complexes mentioned above. From a topological viewpoint, the homogeneous decomposition identifies, in an optimized manner, proteins that possess the higher degree within the network, like those investigated by independent studies (Zotenko et al., 2008).

3.3 A large-scale network: the yeast protein interactome

Previous examples show the qualitative accuracy of informations extracted by homogeneous decomposition, since the complexes identified in silico correspond to the already known biological ones. We propose an application on a more prospective network where the structure of the protein interaction network [that is, its (sub-)complexes] remains unknown.

Hart et al. (2006) decompose the large-scale network of the yeast using the MCL technique on TAP-MS experiments. They obtain 390 clusters or protein complexes, disjoint from each other. Since several clusters are too complex to be investigated in a precise manner, the homogeneous decomposition appears as a natural complement of the MCL technique. Indeed, the decomposition describes relationships within complexes emphasized by the MCL technique. As an illustration, we investigated the interactions inside each of the 103 complexes that present at least four proteins (see Supplementary Material). Among them, 61 complexes are too simplistic to require a homogeneous decomposition (they have no P-node), 21 complexes show identical modular and homogeneous decompositions and 21 complexes present an homogeneous decomposition that refines the modular one. Among these complexes, the modular decomposition identifies 52 sub-complexes, including 25 that give significant GO results (in average \( P\text{-value} = 3.35 \times 10^{-4} \) and 69.9% of cluster frequency) when investigated with function ontologies (Boyle et al., 2004). As evidences of a gain over the modular decomposition, we further analyze the interactions both between complexes and within a complex is possible. The homogeneous decomposition hence represents a major contribution to infer the overall network. To sum up, the homogeneous decomposition is a theoretical technique that extracts global and local information.

4 CONCLUSIONS

We herein studied the contribution of hierarchical decompositions of graphs to the analysis of protein interaction networks. The modular decomposition has been known for a while and showed great successes for investigating protein–protein interaction networks. This decomposition gives a representation of a graph as a tree of modules. As a fundamental property, all the nodes within a module have the same neighborhood outside of the module. Based on the modular decomposition, Gagneur et al. (2004) interpret these modules as either: (i) \( \bigcap \)-module, (ii) \( \bigcup \)-module or (iii) P-module, the latter one designating undecomposable graphs. Unfortunately, the third case occurs on many protein interaction networks, which makes difficult the analysis of concrete biological networks, despite interesting internal structures of such subgraphs. Whereas recent techniques, like MCL analysis, overcome this problem by using clustering approaches, we here propose to use another hierarchical decomposition that investigates the P-nodes: the homogeneous decomposition. Like the modular decomposition, the homogeneous one aims at finding the maximum number of cliques within a graph. In the protein–protein interaction network context, it gives the maximum number of complexes. As a major improvement, the homogeneous decomposition introduces two supplementary node types, namely W- and H-nodes. They give us the opportunity to identify a wheel structure around a hub, within certain P-modules. Such a structure refines the decomposition, thus allowing further investigation on the protein interaction network.

Compared with the modular decomposition, the homogeneous one efficiently stores the important cliques (i.e. the unit complexes). As a consequence, an easier analysis of the interactions both between complexes and within a complex is possible. The homogeneous decomposition hence represents a major contribution to infer (sub-)complexes and to identify particular features of (groups of) proteins. In particular, it emphasizes the presence/absence of specific proteins within complexes of interest, their impact or interactions on the overall network. To sum up, the homogeneous decomposition is a theoretical technique that extracts global and local information.
It should be used as a complement to experimental techniques that identify the complete set of interactions in a network. Further works should focus on developing new decomposition techniques, refining the current ones.

Conflict of Interest: none declared.

REFERENCES


