2.5D Visualisation of Overlapping Biological Networks

David CY Fung[1]**, Seok-Hee Hong**[1]**, Dirk Koschützki**[2]**,**
Falk Schreiber[3,4] **and Kai Xu**[5]

[1]The University of Sydney, Sydney (Australia)
[2]Furtwangen University of Applied Sciences, Furtwangen (Germany)
[3]IPK Gatersleben, Gatersleben (Germany)
[4]Martin-Luther University Halle-Wittenberg (Germany)
[5]ICT Centre, CSIRO, Hobart (Australia)

Abstract

Biological data is often structured in the form of complex interconnected networks such as protein interaction and metabolic networks. In this paper, we investigate a new problem of visualising such *overlapping biological networks*. Two networks overlap if they share some nodes and edges. We present an approach for constructing visualisations of two overlapping networks, based on a restricted three dimensional representation. More specifically, we use three parallel two dimensional planes placed in three dimensions to represent overlapping networks: one for each network (the top and the bottom planes) and one for the overlapping part (in the middle plane).
Our method aims to achieve both drawing aesthetics (or conventions) for each individual network, and highlighting the intersection part by them. Using three biological datasets, we evaluate our visualisation design with the aim to test whether overlapping networks can support the visual analysis of heterogeneous and yet interconnected networks.

1 Introduction

In the last few years large and complex network models emerged in the biological sciences. Visualisation can be an effective analysis tool for such networks. Good visualisation reveals the hidden structure of the networks and amplifies human understanding, thus leading to new knowledge and enabling predictions. However, creating good visualisations of large and complex networks is very challenging, due to scalability and complexity issues.

1.1 Interconnected biological networks

Life emerges out of complex molecular interactions and highly orchestrated biological processes. One example of such a process is the transformation of metabolites into other metabolites. Inside living organisms these transformations are usually performed with the help of additional molecules called enzymes. Enzymes are proteins which are products of a process called gene expression. During gene expression, proteins are produced according to the information stored in the genes. Which proteins are constructed and are therefore available at a specific point in time is controlled by a process called gene regulation. Inside cells, proteins interact in various ways, they form, for example, protein complexes which are groups of functionally related proteins.

The combination of all these processes forms one single complex network. Usually this network is divided into specific networks dealing with one particular mode of interaction. Metabolic reactions form the metabolic network. Here the output of one reaction can be the input of several other reactions, and this network captures the transformation of metabolites. Interactions between proteins form the protein interaction network. And the expression of a single gene is usually controlled by other genes, and thus the regulatory mechanisms between genes form the gene-regulatory network.

Studies about biological processes usually focus only on one of these networks such as the metabolic network or the protein interaction network of that organism. This restriction is often based on the availability of the data and motivated by specific methods used to investigate the network. The major problem of such a reduced approach is the failure to capture the interconnection between different networks, which clearly occurs within living organisms. In the light that the same biological process can involve more than one mode of interaction, the reduced approach severely limits the biological insights that can be generated. Therefore, an integrated network which captures important interactions between different networks is required.

One particular important one is the integrated network of metabolism and protein interaction. Some proteins occurring in a protein interaction network are enzymes and catalyse reactions in the metabolic network. The interactions between proteins can have different functional meanings which directly influence the metabolism. Interactions can mean that no single protein can catalyse a reaction, but that the proteins have to form a protein complex. It can give hints about the regulation of the activity of an enzyme, and therefore about regulatory processes within the metabolism. Also, interaction can suggest spatial relationships between proteins, e. g. proteins which are necessary for subsequent steps in a metabolic network and which may be coupled, for example, to a membrane. The proposed method for the visual analysis of two overlapping networks can help in the investigation of this integrated network.

Another interesting example of overlapping networks to consider is the overlapping between protein interaction networks and signal transduction networks. Living cells respond to external stimuli such as availability of nutrients, or cell-to-cell contact via the signal transduction network. Protein interactions within the signal transduction network are a subset of the entire human protein interaction network. Signal transduction proteins typically interact with each other and one protein activates or de-activates the other by phosphorylation within milliseconds. This reaction happens in the presence of special energy-storing molecules, namely, cyclic-ATP and GTP.

Signal transduction network influences certain biological functions such as the cell cycle by phosphorylating proteins in the nuclear protein interaction network. Nuclear proteins are those that interact with each other within the cellular component, the cell nucleus. Therefore, visual analysis of the integrated signal transduction and nuclear protein interaction as two overlapping networks will help to investigate the influence of signal transduction on certain biological functions.

1.2 Visualisation of overlapping networks

In this paper, we study a new problem of visualising a set of *overlapping* networks. Roughly speaking, two networks overlap if they share a subset of nodes and edges. In general, a set of k graphs may overlap each other. Note that unlike temporal networks, evolution networks

or dynamic networks, which consists of a set of similar networks arising from one single network, two overlapping networks may be *heterogeneous*. That is, two networks differ from each other, possibly with only a small fraction of overlaps (i. e. intersections). Our research problem was inspired by a real world application: the visual analysis of two heterogeneous but interconnected biological networks within a single visualisation. In general, a set of overlapping networks appears frequently not only in biology but also in other application domains.

Good visualisations of overlapping networks can enable *integrated* analysis, insight, and comparison, which cannot be easily supported in a separate visualisation of each single network. Furthermore, more complex high level analysis can be supported by relating two or more heterogeneous networks. More specifically, we enable visual analysis by highlighting connections between different networks, and simultaneously exposing their differences.

Complex biological processes in organisms constitute a number of networks which attract visual analysis methods in the biological field. Several authors deal with the visual investigation of protein interaction networks [4, 11, 12, 17] or visualisation methods for metabolic pathways [5, 16, 21, 22]. These studies about biological processes only focus on a single network of an organism. To date, only a few works has been done on the automatic visualisation of integrated networks, for example in the Patika system [10]. However, in such integrated networks, the different modes of interaction such as protein interaction and metabolism are not visually separated anymore and therefore it is not possible to investigate and compare these different interactions.

Recently, there have been methods proposed to visualise a set of similar biological networks [6, 7]. All these methods are based on the 2.5D visualisation approach: each network is drawn in a 2D plane and then these networks are stacked in the third dimension with the same nodes aligned. Note that these methods are designed for a set of *similar* networks that are of the same type and the difference among them is relatively small: a set of phylogenetic trees derived from the same dataset [6] or the changes of a metabolic pathway over time [7].

Novel visualisation methods of related, heterogenous, overlapping networks are therefore required especially in the analysis of biological networks. In this paper, we present an approach for visualising overlapping networks. Our methods highlight the overlapping parts, while displaying the network structure of each network. As the first step towards this new problem, we consider *two* overlapping networks in this paper and evaluate the method with different datasets of biological networks (metabolic pathways, protein interaction networks, gene regulatory networks and signal transduction networks).

2   Representation of Two Overlapping Networks

In order to represent two overlapping networks G_1 and G_2, where $G_3 = G_1 \cap G_2$, we use three parallel two dimensional planes in three dimensions. More specifically, we draw G_1 and G_2 on the top and the bottom planes, and draw the overlapping part G_3 in the middle plane, see Figure 1. Then inter-plane edge sets E_{13} and E_{23} are added to connect those corresponding nodes in G_1, G_2 and G_3.

More formally, we can define our research problem of two overlapping networks visualisation as the following. The inputs of our methods are:
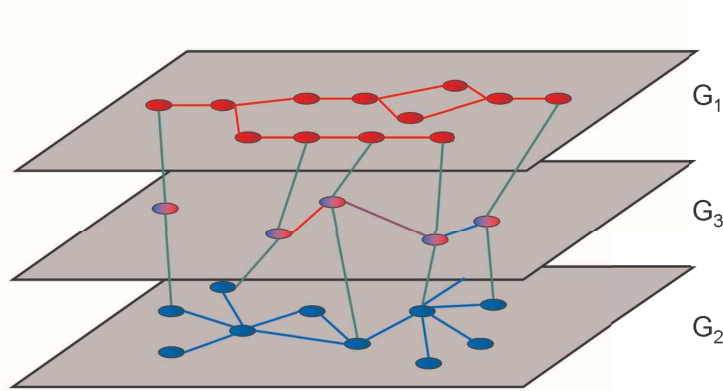
Figure 1: General layout approach. Two networks, G_1 shown in red on the top layer, G_2 shown in blue on the bottom layer and the overlapping nodes (G_3) shown in red-blue on the middle layer. Edges connected to the overlapping nodes are shown in the middle layer. Edges existing only in G_1 are shown in red, edges existing only in G_2 are shown in blue, and those edges existing in both networks are shown in purple.

- Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, where V_1, V_2 are the node sets and E_1, E_2 are the edge sets;

- A 1-to-1 mapping $M_V : V_{11} \leftrightarrow V_{22}$ which defines the overlapping nodes between G_1 and G_2, where V_{11} (respectively, V_{22}) is a subset of V_1 (respectively, V_2), i. e. $V_{11} \subseteq V_1$ (respectively, $V_{22} \subseteq V_2$).

- A 1-to-1 mapping $M_E : E_{11} \leftrightarrow E_{22}$ which defines the overlapping edges between G_1 and G_2, where E_{11} (respectively, E_{22}) is a subset of E_1 (respectively, E_2), i. e. $E_{11} \subseteq E_1$ (respectively, $E_{22} \subseteq E_2$).

The outputs of our methods are:

- Construction of overlapping part $G_3 = (V_3, E_3)$, where the node set V_3 is defined as those common nodes between G_1 and G_2, i. e.

$$V_3 = \{v \mid \exists v_1 \in V_1, v_2 \in V_2, v_1 \leftrightarrow v_2 \in M_V\}$$

and the edge set E_3 is defined as those common edges between G_1 and G_2, i. e.

$$E_3 = \{(v, v') \mid (v_1, v_1') \in E_1 \ and \ (v_2, v_2') \in E_2, (v_1, v_1') \leftrightarrow (v_2, v_2') \in M_E\}$$

- The layouts L_1, L_2 and L_3 of G_1, G_2 and G_3 respectively, including the two edge sets E_{13} and E_{23} that connect the corresponding nodes between G_1 and G_3, and G_2 and G_3 respectively.

3   Visualisation of Two Overlapping Networks

In this section, we present a method for constructing visualisation of two overlapping networks, based on the representation discussed in the previous section. For our overlapping network

visualisation, we aim to achieve both drawing aesthetics and conventions for each individual network G_1 and G_2, and highlighting the intersection part G_3. Furthermore, we try to minimise the total edge length of E_{13} and E_{23} between the parallel planes, in order to minimise occlusion problem in three dimensions.

Note that in many applications, layouts of G_1 or G_2 may be fixed (or given). For example, protein interaction networks are usually drawn with force-directed layout, while metabolic pathways are preferably drawn with hierarchical (layered) layout or using the conventional (fixed) KEGG layout. Based on this, we designed a visualisation method for overlapping networks which takes into account one given layout. Note that the case of two given layouts can be solved by using the fixed co-ordinates for each layout and drawing G_3 using the barycenter of v_1 in L_1 and v_2 in L_2.

Suppose that one of the graph (say, G_1) has a fixed layout (for example, a pre-computed hierarchical layout). In this case, we need to draw G_2 and G_3, considering the fixed layout of G_1. A variation of a force-directed method can be used to produce a reasonably good layout of G_2 and G_3, while reducing the total inter-plane edge length. More specifically, we have the following steps:

1. Draw G_1 with a given layout L_1;

2. Assign the position of node v_1 in L_1 as the initial position of its mapped node v_2 in G_2;

3. Add inter-plane edge sets E_{13} and E_{23} and model each inter-plane edge as a zero-length natural spring (i. e. attraction force only). Note that such spring does not change the distance between the planes;

4. Draw G_2 and the edge sets E_{13} and E_{23} with a force-directed layout with the previous initial positions;

5. Draw G_3 using the barycenter of v_1 in L_1 and v_2 in L_2;

6. Assign three different z-coordinates to the nodes in G_1, G_2 and G_3.

At step 2, by assigning a good initial position based on L_1, it can help the force-directed layout of G_2 at step 4 to converge quickly. Further, the corresponding nodes v_1 in L_1 and v_2 in L_2 have similar x-y coordinates. At step 3, we add the zero-length natural spring for inter-plane edges, in order to reduce the total edge length of inter-plane edges. Note that at step 4, this force competes with other forces of G_2 which try to produce a good layout for G_2. As a result, the corresponding nodes may not always perfectly aligned with a straight-line. At step 5, we use the barycenter of L_1 and L_2 to draw G_3, as in general case, the mapping between G_1 and G_2 is not necessarily a 1-to-1 mapping.

The presented visualisation method has been implemented as a plugin for GEOMI [2]. GEOMI is a visual analysis tool for the visualisation and analysis of large and complex networks. GEOMI provides a collection of network analysis methods, new 2.5D graph layout algorithms and several graph navigation and interaction methods. GEOMI is part of a new generation of visual analysis tools, which tightly integrating graph visualisation techniques with network analysis methods. The two networks under investigation can be loaded into the tool as files, but have to be created individually from databases or other data sources beforehand. The middle layer

(intersection) of each network is thereby computed automatically based on common identifiers between the two networks.

4 Overlapping Biological Networks

We now present the results of our method using three examples of overlapping biological networks: metabolic network and protein interaction network, protein interaction network and gene regulatory network, and protein interaction network and signal transduction network.

4.1 Overlapping metabolic network and protein interaction network

4.1.1 Construction of the overlapping networks

We used two databases for constructing an overlapping biochemical network. The Database of Interacting Proteins (DIP) [20] stores information about protein interactions and the pathway section of the KEGG databases [15] provides information about metabolic pathways. From those databases we extracted information about *Escherichia coli*, one of the best studied bacteria in molecular biology.

Protein interaction data for *E. coli* was downloaded from DIP (release July 2007). Each line of the downloaded file describes a pairwise interaction between two proteins. Each protein is identified by an unique identifier, and several other identifiers such as the UniProt ID [23] are given in the file. As current methods for determining interactions between proteins only report that two proteins interact without detailed information concerning type and directionality of the interaction, they are modelled as undirected graphs. Using all interaction pairs that are currently stored in DIP for *E. coli*, a network with 1846 nodes and 8013 edges was constructed. The network consisted of 351 connected components. The largest connected component consists of 1440 nodes with 7279 edges. All other components contain between 1 and 6 nodes.

Metabolic reactions are organised in groups called pathways. For *E. coli*, 108 pathways are described in the KEGG database Release 43.0 (July 2007). From this database, we extracted the glycolysis pathway. Within the glycolysis, sugars such as fructose and glucose are converted into pyruvate and additional energy in the form of the molecule ATP. The network was extracted from KEGG. In total, the network consists of 52 nodes and 57 edges. In metabolism, enzymes are responsible for the catalysis of reactions, and those enzymes are proteins which might occur in the protein interaction network.

In both databases, proteins and enzymes are annotated with common identifier from the UniProt database [23]. Using this identifier, connections between the protein interaction network and the metabolic pathway can be established. In total, 9 enzymes from the glycolysis pathway have a corresponding protein in the protein interaction network. To construct an integrated network, both networks are combined. Every enzyme in the metabolic network having a corresponding protein in the protein interaction network is connected to a new node representing both, the protein in the protein interaction network and the enzyme in the metabolic network. The protein interaction network was reduced to the 1-neighbourhood for all proteins connected to the metabolic pathways. Finally, labels for the proteins in the interaction network were replaced by the corresponding gene name, if known.

4.1.2   Visual analysis of the overlapping networks

Figure 2 shows a visualisation of the two overlapping biochemical networks. A relevant part of the protein interaction network G_1 is shown in green, and the blue network G_2 represents a metabolic pathway, the glycolysis. The protein interaction network G_1 is laid out using our method described earlier with the visualisation of the metabolic network G_2 computed using a layout approach which takes into consideration the given KEGG coordinates where possible. In this visualisation, nodes occurring in both networks are connected to the corresponding nodes in G_3 by yellow edges.

An analysis of the integrated, overlapping networks reveals several interesting aspects. It can be clearly seen that several, but not all enzymes of the glycolysis are present in the protein interaction network. One would expect every enzyme to be represented in the protein interaction network, as every enzyme is a protein. This is not the case, because protein interaction databases, including DIP, usually only refer proteins for which at least one interaction is known.
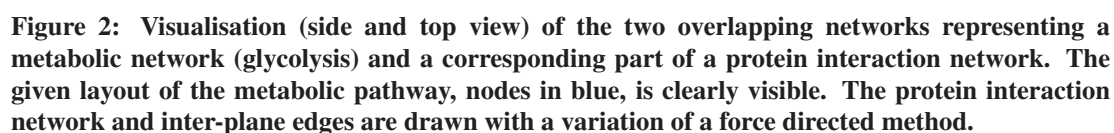
Usually the activity of metabolic reactions is regulated, and metabolic pathways can be regulated at different points. Often this regulation is carried out via protein interactions. There are several enzymes which interact with other proteins, and most are together in the connected component of the protein interaction network. This may be a clue about specific regulation mechanisms of metabolic reactions or a multi-protein complex (metabolon).

For example, DIP:10622N (pyruvate kinase II) is in the protein interaction network connected to DIP:10467N (pyruvate formate-lyase), see Figure 3. Whereas pyruvate kinase II catalyses the reaction from p-enol-pyruvate to pyruvate, the interacting protein pyruvate formate-lyase catalyses the reaction which transforms pyruvate into acetyl-CoA and formate in a subsequent metabolic reaction. Also another aspect, the connection of the glycolysis to cold shock (e. g. DIP:9334N) and heat shock proteins (e. g. DIP:6142N) is not surprising as, for example, the low-temperature stress on the glycolytic activity is known for other bacteria such as *Lactococcus lactis* [24]. Interesting is also the highly connected protein DIP9040N (dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex) which is currently under further investigation.
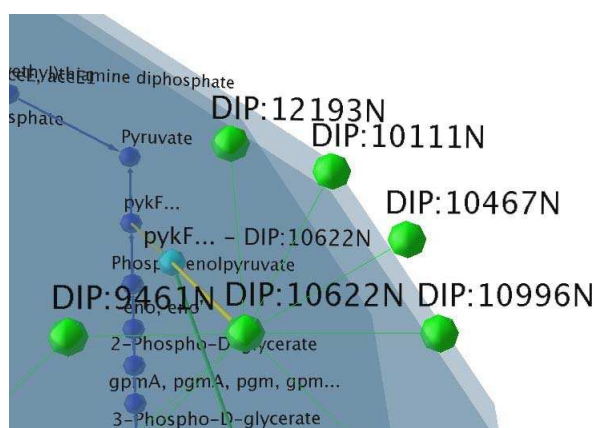
4.2   Overlapping gene regulatory network and protein interaction network

4.2.1   Construction of the overlapping networks

To understanding how the gene regulatory network in *E. coli* influences the physical organisation of its protein interaction network, we used the overlapping networks to visualise the interactions between these two networks. We used two datasets to construct the overlapping networks. The gene regulatory data for *E. coli* was downloaded from the public database RegulonDB [19]. Each row describes the pairwise and directed interaction between two proteins. The source nodes were listed in the right column and end nodes were listed on the left column. From this, a network of 1371 nodes and 2030 edges were constructed. The protein interaction network was constructed from the DIP data as described in Section 4.1.1.
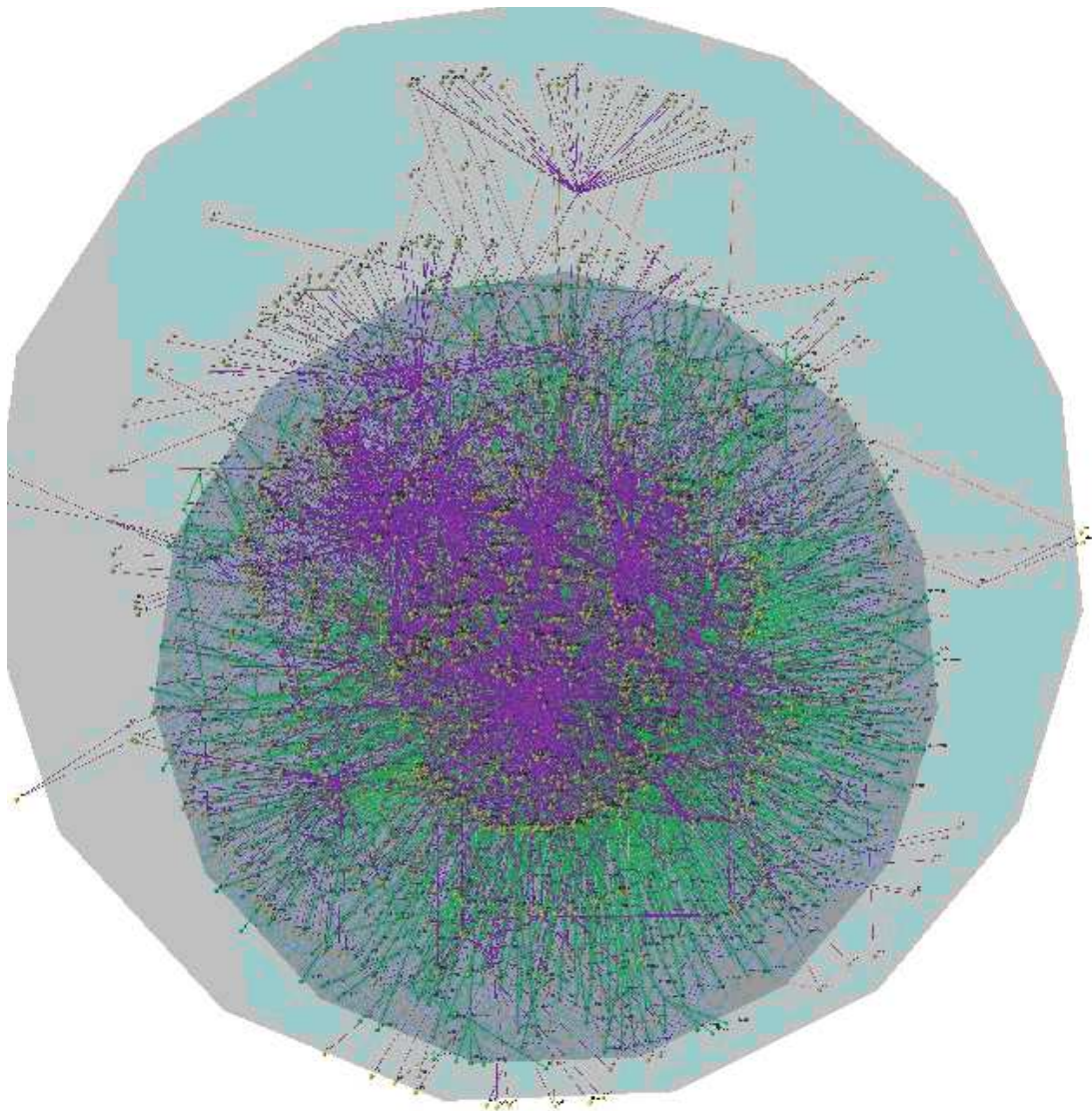
Figure 2: Visualisation (side and top view) of the two overlapping networks representing a metabolic network (glycolysis) and a corresponding part of a protein interaction network. The given layout of the metabolic pathway, nodes in blue, is clearly visible. The protein interaction network and inter-plane edges are drawn with a variation of a force directed method.

Figure 3: The protein DIP:10622N in the protein interaction network is identical to the enzyme pykF in the metabolic network. It interacts with several other proteins. See the text for a discussion.

4.2.2 Visual analysis of the overlapping networks

Figure 4 shows the visualisation of the overlapping networks in 2 layers. Here, G_1 representing the largest connected component of the gene regulatory interaction network is shown in magenta. The green network G_2 represents the largest connected component of the protein interaction network. The overlapping networks were drawn according to our method. The fixed co-ordinates for G_1 was computed according to the Kamada-Kawai layout [14] while G_2 was laid out using the force-directed method. G_1 contains directional interactions that indicate the flow of control on gene expression. Given the interaction protein A $\rightarrow$ protein B, means that protein A induces or represses the expression of protein B. G_2 contains the physical interactions between proteins. In this visualisation, nodes occurring in both networks were connected by yellow edges. Figure 5 shows that there are numerous edges between G_1 and G_2 demonstrated that the two networks are highly interconnected.

When viewing from the top of G_1, one can identify six interconnected hubs. The centre nodes of these hubs are *crp, ihfAB, hns, fis, lrp, arcA,* and *fnr*. This is suggesting that the global state of the gene regulatory network can be determined by the on/off states of a few core transcription factors. However, none of these transcription factors are mapped to the protein interaction network. One possible explanation is the DIP dataset is incomplete. The alternative explanation is that the core transcription factors do not require any protein co-factors to facilitate gene regulation thereby greatly reduces the amount of just-in-time transcription required. Hence, the physical organisation of the protein interaction network is indirectly influenced by the state of the core transcription factors. Such a design allows the bacterium to fine tune its functional organisation rapidly in response to any external environmental challenges but can also be crippled by any loss-of-function mutations carried by either one of the above transcription factors.

Visually, *crp* appears to have the highest node degree which indicates that it is the central transcription factors in *E. coli*. This observation aligns with the current consensus that *crp* is the key hub in the *E. coli* gene regulatory network [3]. Visualisation of G_2 shows that *crp* is directly connected to *fnr*, *hns* and *ihfAB* indicating that the on/off states of the four transcription factors are inter-dependent. To date, *crp* has been known to regulate more than 200 operons which are involved in diverse biological functions, i. e. carbohydrate metabolism, ion transportation,

Figure 4: Visualisation of the two overlapping networks representing a gene regulatory network and and a corresponding part of a protein interaction network. The gene regulatory network is in blue, whereas the protein interaction network is in green. The proteins co-shared by both networks are in red.
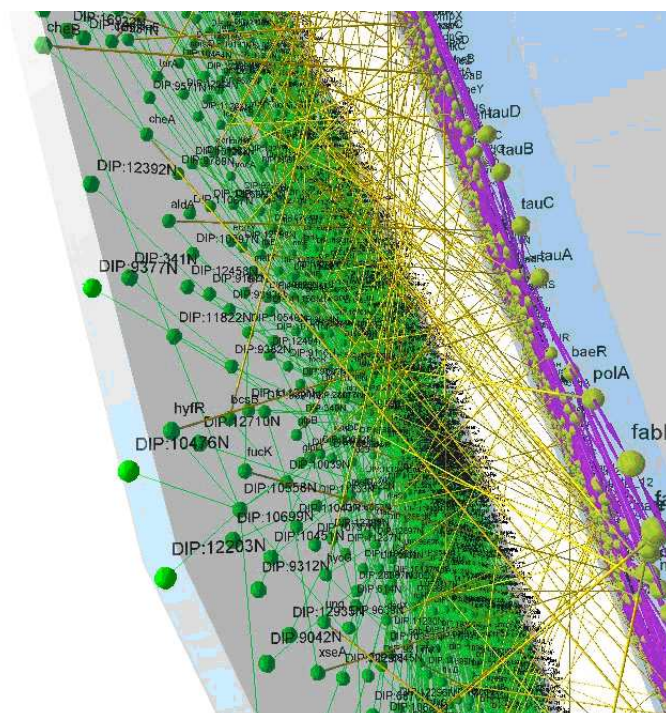
Figure 5: Enlargement of the visualisation in Figure 4 which shows numerous inter-cluster edges (yellow colour) between G_1 and G_2 representing the gene regulatory proteins that were also present in the protein interaction network.
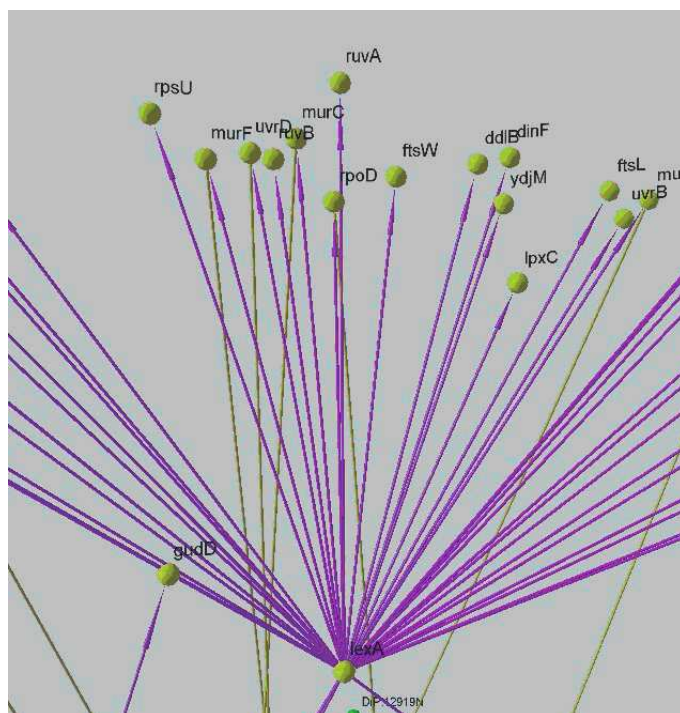
Figure 6: Enlargement of the visualisation in Figure 4, see text for an explanation.

amino acid metabolism, energy production, and downstream gene regulation [26]. Another transcription factor *arcA* has been estimated to regulate 100 to 150 operons. *arcA* has recently been found to be a repressor of anaerobic respiration enzymes [25]. It is a part of the two component *arcA/arcB* signaling protein complex within the cytoplasm.

Figure 6 shows that the proteins occurring in both networks lied at the periphery of G_2, thus shows that they are effector proteins regulated by the gene regulatory hierarchy. In G_1, they can be divided into two groups according to their topology. One group are nodes peripheral to a certain hub, e. g. *aroA, cheB, cysD, cysP, cysI, fliN, flgD, ftsK, phr, polB, recX, rpoD, ruvB, murF, torD, uvrB* and *ygbK*. Another group are nodes connected to two or more hubs, e. g. *cheA, cysJ, ccmF, ftsQ, murE, hycG, speA, torA, topA* and *uvrD*. They are frequently known as the bottleneck proteins.

In *E. coli*, these two groups of proteins exhibit different interaction dynamics. Because many biological processes are operated by protein complexes, there is a need to optimise transcriptional efficiency while ensuring that the right protein complex for a particular set of biological processes is fully operational on demand. The peripheral protein is usually a component specific to a particular complex and is only expressed when a certain biological process is needed. Once the peripheral protein is expressed, the protein complex becomes fully functional. The bottleneck protein,on the other hand, can be a shared component between two or more complexes. Their interaction with each complex depends on the relative molecular abundance among the various targeted complexes. This type of interaction dynamics is known as differential affinity. By regulating the expression of the bottleneck protein, the organism can easily switch between various protein complexes and therefore biological processes when necessary.

In summary, the 2-layer overlapping networks showed that the gene regulatory network controlled the physical organisation of the protein interaction network by regulating the structural integrity of protein complexes. In turn, the functional organisation of *E. coli* at a certain time point is dependent on which protein complexes are in operation and their relative molecular abundance. This design allows the organism to rapidly orchestrate the activities of multiple biological processes in response to environmental challenges.

4.3　Overlapping signal transduction network and protein interaction network

4.3.1　Construction of the overlapping networks

To study the probable effect of signal transduction of the cell cycle of *Homo sapiens*, we used the overlapping networks to visualise the interactions between the transforming growth factor beta (*TGFβ*) network and the nuclear protein interaction network. We used four datasets to construct the overlapping signal transduction networks. The dataset for the *TGFβ* signal transduction network was manually curated from Cui *et al.* [18]. From this, a network of 48 nodes and 52 edges was constructed. The nuclear protein interaction data was extracted from three datasets downloaded from the Gene Ontology Consortium [1], the BioGRID [8] and the ECHO databases [13].

Because many of the protein interactions in the BioGRID dataset have been verified by multiple laboratory techniques, it is more reliable than data generated solely by the yeast two-hybrid method. Each record in the BioGRID dataset describes a pairwise interaction between two
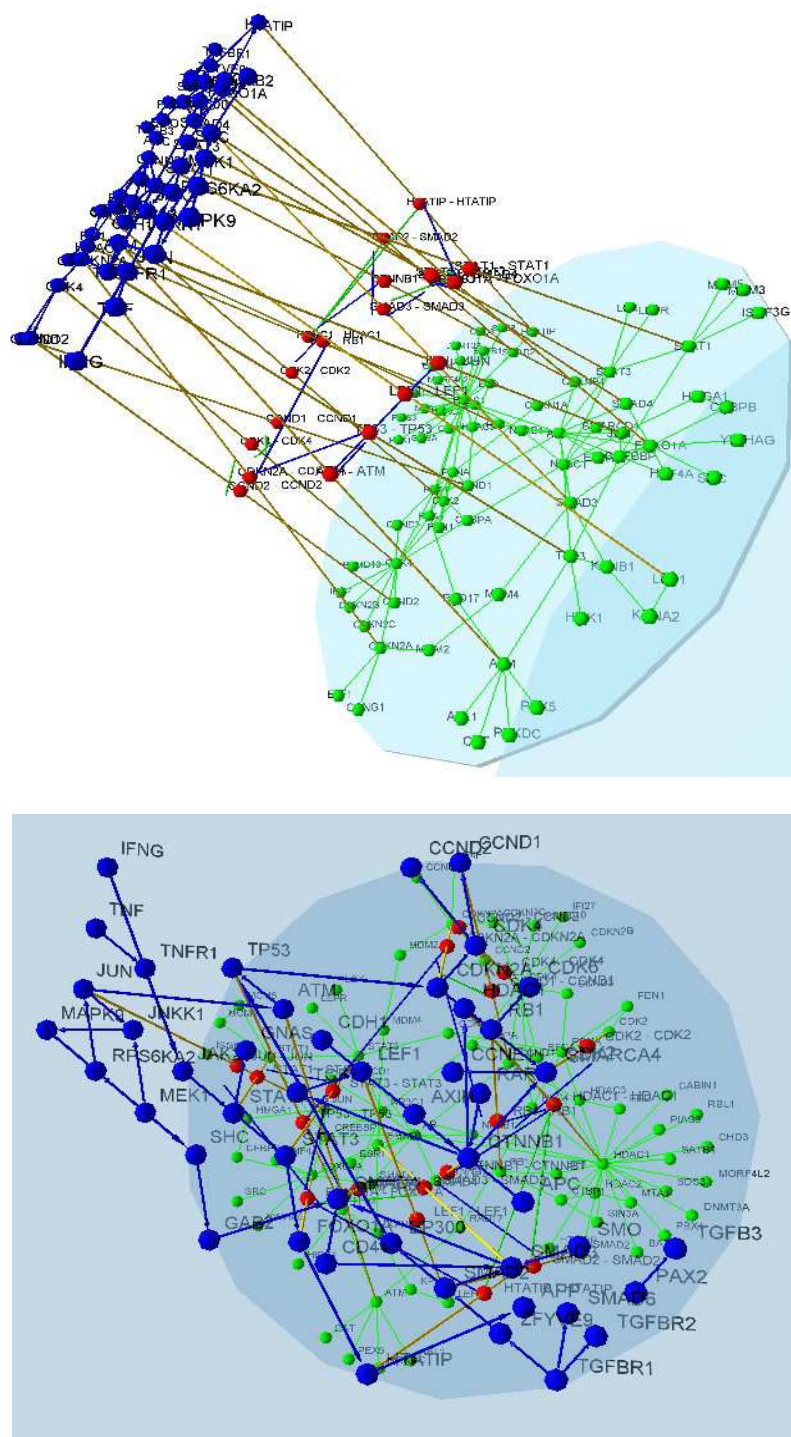
Figure 7: Visualisation of the two overlapping networks representing a signal transduction network (*TGFβ*) and a corresponding part of a nuclear protein interaction network. The signal transduction network is in blue, whereas the protein interaction network is in green. The proteins co-shared by both bietworks are in red.

proteins. The ECHO database provides a list of liver cancer-specific proteins. By querying against the Gene Ontology dataset, BioGRID records that describe protein interactions within the cellular component *nucleus* were first extracted. The resulting dataset which contained 1748 proteins was then queried against the list of liver cancer-specific proteins. The final dataset contains protein interactions that are not only found in the cell nucleus but also liver cancer-specific. From this, a network of 583 nodes and 711 edges was constructed. Protein names assigned by the Human Genome Organization have been used as node labels in the above networks.

4.3.2　Visual analysis of the overlapping networks

Figure 7 shows how the visualisation effectively represents two different types of protein networks in separate layers. Here, G_1 representing the *TGFβ* signal transduction network is shown in blue. The green network G_2 represents the largest connected component of the nuclear protein interaction network. The overlapping networks were drawn according to our algorithm. The fixed co-ordinates for G_1 were manually assigned while G_2 was laid out using the force-directed method.

G_1 contains directional interactions that serve the sole purpose of propagating the phosphorylation signal from the *TGFBR1* receptor protein to the rest of the network. Given the interaction, protein A $\rightarrow$ protein B, means that protein A *phosphorylates* protein B. Some proteins are activated while others are inactivated by phosphorylation. Each interaction in G_1 is transient and fast which occurs within milliseconds. G_2 contains physical interactions between nuclear proteins. Nodes occurring in both networks are shown in red and are connected to the corresponding nodes in G_3 by yellow edges. Edges in G_3 are shown in the same colour as their corresponding edges in either G_1 or G_2.

When viewing G_2 from the top and the overlapping networks on the side, one can clearly recognise that some signaling proteins e. g. *ATM, CDK4, FOXO1A, HDAC1,* and *SMAD3* are themselves highly connected in G_1 (node degree > 10). Viewing G_2 from the top also helps one to identify the signaling order of the above proteins. Of interest, the proteins within G_3 can be divided into four groups according to their functionality. *SMAD2,3,4* are transducers for the *TGFBR1* receptor protein. *HDAC1, CDK2, CDK4, CCND1* and *CCND2* are cell cycle proteins that drive cell division. *JUN, LEF1, STAT1* and *STAT3* are oncogenic proteins that activate the above cell cycle proteins. *CDKN2A, RB1* and *ATM* are tumour suppressing proteins that inactivate the same set of cell cycle proteins. This observation suggests that *TGFβ* is a secondary control point of the cell cycle. Even though *TGFβ* can directly activate cell cycle proteins by phosphorylation, this alone is inadequate for initiating the cell cycle. Rather, it relies on the relative molecular abundance between the oncogenic proteins and the tumour suppressing proteins, which in turn is determined by the activity of other signaling proteins prior to the onset of *TGFβ* signaling. This type of interaction dynamics is known as differential signaling. For example, the authors noticed that *IFNG* is one of the proteins in G_1. This protein activates *STAT1* and *STAT3* by phosphorylation via *JAK2*. If *TGFβ* signal transduction is being activated simultaneously, the molecular abundance of *STAT1* and *STAT3* will out-compete the tumour suppressing proteins leading to the initiation of the cell cycle. *IFNG* is produced by virally-infected cells. It has been known that liver cancer can be induced by the persistent infection of hepatitis A, B, and C viruses [9]. As the cancer progresses, genome instability

often leads to the loss of tumour suppressor genes thereby reinforcing the cancer-promoting impact of *IFNG* and *TGFβ*.

In summary, overlapping networks can be an effective representation of the biological complexity underlying a signal transduction network and is particularly useful for oncology, immunology and endocrinology research.

5 Conclusions

In this paper, we introduce a new problem of visualising overlapping networks, and present a method for visualising two overlapping networks using a 2.5D representation. Our methods can achieve both drawing aesthetics and conventions for each individual network, and simultaneously highlighting the overlapping part between them.

The usability of our visualisation approach has been studied on real world applications from biology. It showed that our visualisation can lead to new insights and analysis into the data. More specifically, it can provide visual analysis of network integration of two heterogeneous biological networks (metabolic pathway and protein interaction network, protein interaction network and gene regulatory network, and protein interaction network and signal transduction pathway). However, it should be noted that data quality, which depends on the chosen data sources, is crucial for the visual analysis step.

The presented method can be used for large networks, by reducing the size of the whole network to a subset of relevant overlapping networks, as in our examples. This paper presents the first step, the visualisation of two overlapping networks. Our current work is to design new methods for a set of overlapping networks.

6 Acknowledgements

References

[1] Gene ontology consortium. The Gene Ontology (GO) project in 2006. *Nucleic Acids Research*, 34(1):D322–D326, 2006.

[2] A. Ahmed, T. Dwyer, M. Forster, X. Fu, J. Ho, S.-H. Hong, D. Koschützki, C. Murray, N. S. Nikolov, R. Taib, A. Tarassov, and K. Xu. GEOMI: GEOmetry for Maximum Insight. In P. Healy and N. S. Nikolov, editors, *Proc. International Symposium on Graph Drawing (GD '05)*, volume 3843 of *LNCS*, pages 468–479. Springer, 2006.

[3] C. L. Barrett, C. D. Herring, J. L. Reed, and B. D. Palsson. The global transcriptional regulatory network for metabolism in *escherichia coli* exhibits few dominant functional states. *Proc. Natl. Acad. Sci, U.S.A. (PNAS)*, 102:19103–19108, 2005.

[4] W. Basalaj and K. Eilbeck. Straight-line drawings of protein interactions. In *Proc. International Symposium on Graph Drawing (GD'99)*, volume 1731 of *LNCS*, pages 259–266. Springer, 1999.

[5] M. Y. Becker and I. Rojas. A graph layout algorithm for drawing metabolic pathways. *Bioinformatics*, 17(5):461–467, 2001.

[6] U. Brandes, T. Dwyer, and F. Schreiber. Visual triangulation of network-based phylogenetic trees. In *VisSym 2004: Proc. Symposium on Visualization*, pages 75–83, 2004.

[7] U. Brandes, T. Dwyer, and F. Schreiber. Visual understanding of metabolic pathways across organisms using layout in two and a half dimensions. *Journal of Integrative Bioinformatics*, 1(1):119–132, 2004.

[8] B. J. Breitkreutz, C. Stark, T. Reguly, L. Boucher, A. Breitkreutz, M. S. Livstone, R. Oughtred, D. H. Lackner, J. Bähler, V. Wood, K. Dolinski, and M. Tyers. The BioGRID interaction database: 2008 update. *Nucleic Acids Research*, 36(1):D637–D640, 2008.

[9] D. Y. But, C. L. Lai, and M. F. Yuen. Natural history of hepatitis-related hepatocellular carcinoma. *World Journal of Gastroenterology*, 14:1652–1656, 2008.

[10] E. Demir, O. Babur, U. Dogrusöz, A. Gürsoy, G. Nisanci, R. Çetin Atalay, and M. Ozturk. PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics*, 18(7):996–1003, 2002.

[11] C. Friedrich and F. Schreiber. Visualization and navigation methods for typed protein-protein interaction networks. *Applied Bioinformatics*, 2(3):19–24, 2003.

[12] K. Han and B.-H. Ju. A fast layout algorithm for protein interaction networks. *Bioinformatics*, 19(15):1882–1888, 2003.

[13] C.-N. Hsu, J.-M. Lai, C.-H. Liu, H.-H. Tseng, C.-Y. Lin, K.-T. Lin, H.-H. Yeh, T.-Y. Sung, W.-L. Hsu, L.-J. Su, S.-A. Lee, C.-H. Chen, G.-C. Lee, D. Lee, Y.-L. Shiue, C.-W. Yeh, C.-H. Chang, C.-Y. Kao, and C.-Y. F. Huang. Detection of the inferred interaction network in hepatocellular carcinoma from ehco (encyclopedia of hepatocellular carcinoma genes online). *BMC Bioinformatics*, 8:66, 2007.

[14] T. Kamada and S. Kawai. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7–15, 1989.

[15] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research*, 34(1):D354–D357, 2006.

[16] P. D. Karp, S. M. Paley, and P. Romero. The pathway tools software. *Bioinformatics*, 18(Suppl. 1):S225–S232, 2002.

[17] R. Mrowka. A java applet for visualizing protein-protein interaction. *Bioinformatics*, 17(7):669–670, 2001.

[18] Q.Cui, Y. Ma, M. Jaramillo, H. Bari, A. Awan, S. Yang, S. Zhang, L. Liu, M. Lu, M. O. Connor-McCourt, E. O. Purisima, and E. Wang. A map of human cancer signaling. *Molecular Systems Biology*, 3:152, 2007.

[19] H. Salgado, S. Gama-Castro, M. Peralta-Gil, E. Daz-Peredo, F. Snchez-Solano, A. Santos-Zavaleta, I. Martnez-Flores, V. Jimnez-Jacinto, C. Bonavides-Martnez, J. Segura-Salazar, A. Martnez-Antonio, and J. Collado-Vides. RegulonDB (version 5.0): *Escherichia coli K-12* transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Research*, 34(1):D394–D397, 2006.

[20] L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie, and D. Eisenberg. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research*, 32(1):D449–D451, 2004.

[21] F. Schreiber. High quality visualization of biochemical pathways in BioPath. *In Silico Biology*, 2(2):59–72, 2002.

[22] M. Sirava, T. Schäfer, M. Eiglsperger, M. Kaufmann, O. Kohlbacher, E. Bornberg-Bauer, and H.-P. Lenhof. BioMiner - modeling, analyzing, and visualizing biochemical pathways and networks. *Bioinformatics*, 18(Suppl. 2):S219–S230, 2002.

[23] The UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Research*, 35(1):D193–D197, 2007.

[24] J. A. Wouters, H. H. Kamphuis, J. Hugenholtz, O. P. Kuipers, W. M. de Vos, and T. Abee. Changes in glycolytic activity of *lactococcus lactis* induced by low temperature. *Applied and Environmental Microbiology*, 66(9):3686–3691, 2000.

[25] C.-H. Yeang and M. Vingron. A joint model of regulatory and metabolic networks. *BMC Bioinformatics*, 7:332, 2006.

[26] D. Zheng, C. Constantiniduo, J. Hobman, and S. Minchin. Identification of the crp regulon using in vitro and in vivo transcriptional profiling. *Nucleic Acids Research*, 32:5874–5893, 2004.