Workpackage 9

Link: http://www.rsc.org/Publishing/Journals/MB/article.asp?doi=b922400a

The analysis of complex networks with some causal relation among elements is established, as it occurs in gene regulatory networks, can be done by the detection of the so called dynamical backbone and dynamical modules of the net as explained in our publication Rodriguez-Caso et al 2010. The article and and software for analysis is freely available at Molecular Biosystems wetsite.

WP 9. Multiple Sclerosis network (Dr. Ricard Solé. Universitat Pompeu Fabra)

Periodic activity report. (01/09/2008 - 28/02/2010)

Objectives: Modeling and analysis of network topology and dynamics in multiple sclerosis (MS)

WP responsible and involved partners

The second period (01/09/2008 - 28/02/2010) for the COMPLEXDIS project related to the WP9 has been conducted by UPF team and coordinated by Prof. Ricard Solé in a strength collaboration with the Dr. Villoslada's group (WP7). As it was addressed in the previous report, the original proposal considered the UNAV and UPF teams to work in a single WP with a common budget. According with the Consortium Agreement, the WP9 person/month for the total period achieves to 20 person/moth whereas 16 person/month are for UNAV (in total the 36 person-month of the old WP7). As we have a full dedicated person for this project, our estimated person/months necessarily exceed from the expected in the Consortium Agreement for WP9. In any case objectives of the WP7 and WP9 are focused on the study of the Multiple Sclerosis (MS) disease from a multidisciplinary and systemic perspective. The development of the WP9 (from month 19 to 32 month) has been carried out by Dr. Carlos Rodríguez-Caso with a full-time dedication to this project. Personal costs of the WP9 budget has been used to pay him through a post-doctoral contract with a little part required for the costs of congress and meeting attendance for result's dissemination. The estimated person/month for Dr. Rodriguez-Caso is 18 and 3 person/month for Dr Solé at the time of this report in total 21 person/month for this 18 months. During this period, it is worth to mention the collaboration with Bernat Corominas Murtra, from Prof. Sole's lab, Dr Joaquin Goñi (former member of Dr. Villoslada's group), Dr. Francisco Esteban from the University of Jaén and Dr. Raul Montañez from University of Málaga.

Progress towards objectives.

The objective for WP9 is the modeling and analysis of network topology and dynamics in multiple sclerosis (MS). At the end of the first period of the project a number of lines were developed. A major part is related to D26 "New mathematical and computational tools for analyzing the dynamics and robustness of evolving networks". At the end of the 18th month we developed two research lines about 1) "study and implementation of gene-gene network inference through the use of virtual microarrays performance" and 2)"the development of new tools for the study and organization of biological networks (with especial attention to genetic and metabolic networks)". As it is described in this report this last point has been the main direction for the second period and where a major scientific productivity has been obtained. This research orientation was motivated by the technical problems encountered in the experimental data obtaining from Villoslada's group. This part was the preliminary step for MS gene network inference and this delay motivated a change of strategy in coordination with Villoslada's group for tackling our D28: Topological analysis of MS network. In order to overcome this limitation an alternative in parallel based on a computational network inference of a Multiple Sclerosis and immune network using the available knowledge compiled in databases was leaded by Dr. Villoslada in collaboration with Dr Francisco Esteban from the University of Jaen and with our expertise in network reconstruction and analysis.

The work carried out from the second period regarding deliverables is resumed in the following points:

1. Exploring the organization of biological networks:

_____ *1.1. Tackling the dynamical modularity of gene regulatory networks*

_____ *1.2. Towards a proper definition of hierarchy for complex networks.*

_____ *1.2.1. Elaboration of Null models for the evaluation of feedforward networks*

_____ *1.2.2. Entropy as a measure of hierarchy in feedforward networks*

_____ *1.2.3. Towards a proper definition of hierarchy for directed networks*

_____ *1.3. The impact of network definition on result's interpretation using metabolism as a case study.*

2. A preliminary construction of the Multiple sclerosis gene network through the use of databases

1. Exploring the organization of biological networks:

1.1. Tackling the dynamical modularity of gene regulatory networks

This work has been carried in a large extent during the first period of this project. However, the final process of publication required a significantly effort during the beginning of the second period. This work deals with the definition of a dynamical modularity in hierarchy by using the concepts of strongly connected component detection and graph condensation - borrowed from directed graph theory. By using these measures, the dynamical regulatory cores of the gene regulatory networks of the three best described organisms up to day, i.e, yeast, *E. coli* and *B. subtilis*, were identified and explored in a comparative study. This work has given rise a article entitled: On the basic computational structure of gene regulatory networks. It is cover journal of the special issue of computational systems biology in the Molecular Biosystems journal (Impact factor: 4.23). December 2009 *Mol. BioSyst.*, 2009, 5, 1617 – 1629 (see figure 1). This work was also presented in BCNet workshop December 10-12, 2008 as a contributed talk with the title of Dynamical hierarchy and modularity in gene regulatory networks.

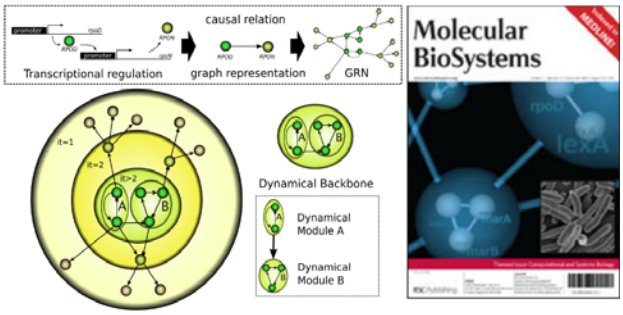
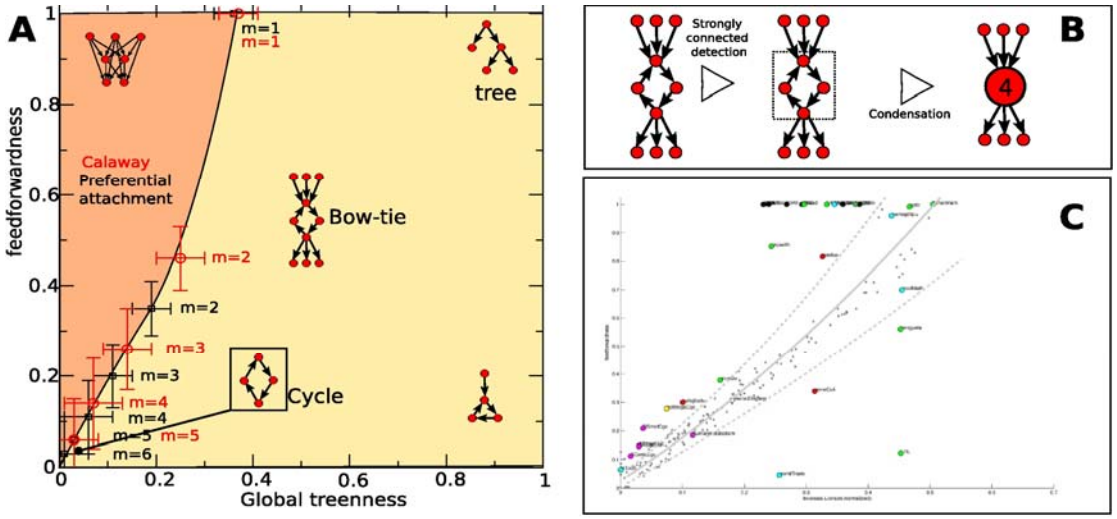


Figure 1. Graphical abstract representing the methodology applied in Rodriguez-Caso et al 2009 (left). Cover journal of our article (right).

Our results reveal a marked top-down hierarchy containing several small dynamical modules for *E. coli* and *B. subtilis*. Conversely, the yeast network displays a single but large dynamical module in the middle of a bow-tie structure. We found that these dynamical modules capture the relevant wiring among both common and organism-specific biological functions such as transcription initiation, metabolic control, signal transduction, response to stress, sporulation and cell cycle.

1.2. Towards a proper definition of hierarchy for complex networks.

A conclusion derived from this work in relation to the project is that, although it has been repeatedly claimed about the hierarchical organisation of gene regulatory networks, little has been said from in a rigorous quantitative perspective. The main problem is the lack of a



proper measure of hierarchy. By this reason the next step of our research regarding D26, was the development of a theoretical framework considering the essential trait when we talk about hierarchy, the directed relation among elements. The resulting work is illustrated in figure 2.

Figure 2.*:* The coordinates of hierarchy. Morphospace indicating the location of extreme graphs. Preferential attachment model and Callaway uncorrelated networks exhibit a similar behavior defining two regions (A). Process of treatment of the network for the measure of treenness and feedforwardness (B). Morphospace for a set of real directed graphs including: gene regulatory networks, metabolism ,electronic circuits, food webs, social and technological networks (C).

However, it is worth to stress that contrasting many other works based in an heuristic approaches, our research is aimed to provide an estimator rigorously derived from the principles of uncertainty -borrowing concepts of Information Theory- and set and order theory by the application of the principle of causality. The ongoing research is split in three branches that will produce three research publications:

1.2.1. Elaboration of Null models for the evaluation of feedforward networks

Real networks are often studied by comparison of null random models in which some characteristics are preserved, e.g. number of node, links, degree distribution, the distribution of connected component are example of topological properties. However, very few works concerning the construction of null models for feedforward networks -also known as directed acyclic graphs (DAGs)- has been done up to these days. Within the perspective of causality, a

DAG represent a map of causal relations. Following this argument, two non connected DAGs represent two different systems in terms of causality. By this reason, the connected component preservation in the null model construction of DAG is required but it has not been previously reported in the literature. By this reason, we have developed four methods of randomization preserving the degree distribution and the connected components. In addition we had to develop a measure based on the mutual information to estimate the effect of randomization in DAGs. This effort has given rise to a work called Exploring the randomness of directed acyclic networks. Goñi J, Corominas-Murtra B, Sole R and Rodriguez-Caso C. that will be submitted in Physical Rev E, in the next days. In figure 3 we can observe the effect of the proposed algorithm applied to the *c. elegans* cell lineage networks. In panel e we observe the effect of randomization using a dissimilarity measure. In this work, to the seek of generality two additional different systems were evaluated in two DAGs: a citation network and another based of the map of relations between PhD students an its advisors.

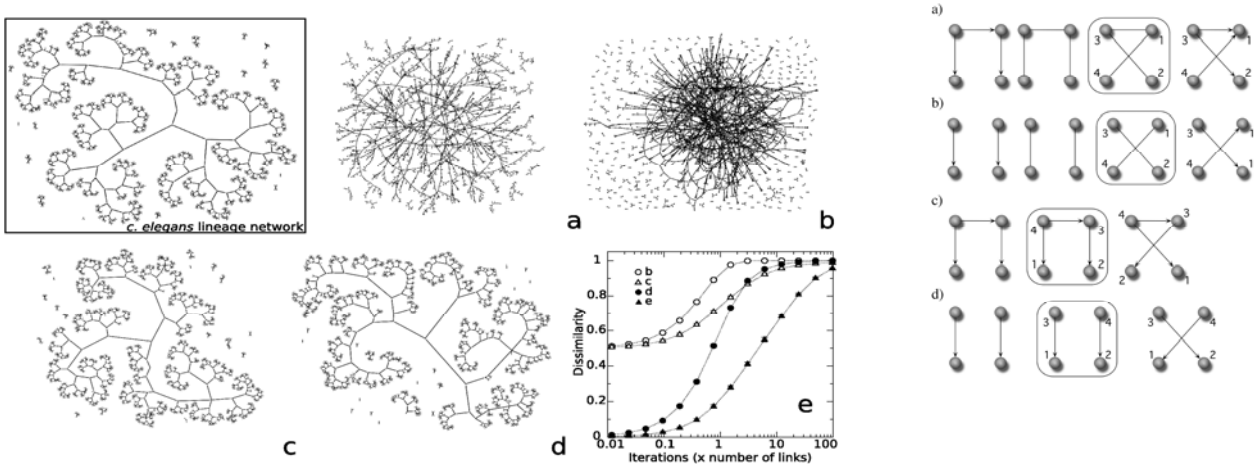


Figure 3. A prototypic randomized *c. elegans* network after a number of iterations equal to100 times the number of links are present for the 4 randomizations proposed: randomization preserving undirected degree sequence and component size distribution (a), randomization only preserving the undirected degree sequence (b), randomization preserving directed degree sequence and component size distribution (c) and randomization preserving only directed degree sequence (d). Panel e represents the dissimilarity respecting the original network along the process of randomization for every randomization type. The mean and the standard deviation of one hundred graph randomizations are shown for each point.

1.2.2. Entropy as a measure of hierarchy in feedforward networks

Once defined a DAGs and its uncorrelated null model, we can quantify how similar are these networks to a tree-like structure, i.e. the ideal hierarchical structure. The result of our investigation is the formalization of three entropy measures.

Local entropy: It measures the average uncertainty associated to any node of the DAG.

Weigthed global entropy: it measures the average uncertainty associated to any node to chose a given pathway resècting the all possible decisions to achieve, climbing up, a given maximal.

Unweigthed global entropy: it measures the average uncertainty associated to any node to chose a given pathway to achieve, climbing up, a given maximal but assuming that any pathway is equally probable.

Maximal entropy were analytically found for these three measures. It is worth stressing that, although the formal details are far from the aim of this report, this mathematical work is a key piece for the definition of a hierarchy measure, and constitutes a step-forward to develop a still non-existent theoretical framework for hierarchy quantification. In the context of our objective within the ComplexDis project, this result essential for the understanding of how genetic regulatory networks are structured beyond using the standard measures from the graph theory. Now we are writing a manuscript that it going to send after the article about DAG null model methodology is sent.

1.2.3. Towards a proper definition of hierarchy for directed networks

This issue constitutes the last part of this line of research. In a preliminar analysis we have reconstruct a morphospace in 2D see figure 2, where a number of very diverse real networks are placed according to the *feedforwardness* (the impact of cycles in the feedforward structure) and the *treenness* (how tree-like structure is the resulting feedforward graph after condensation process) based on entrpy measures addressed in the 1.2.1 section. We remain that condensation process gives the feedforward structure of any directed graph by merging all cycles in individual nodes. In this part of the work we have formally defined the feedforwadness as the fraction of pathways that can be ordered in a feed forward manner. In extreme case we can observe that cycle, after condensation give rise a single node. By definition the feedforwardness and the treenness of an individual node is zero. By contrast an ideal tree, for example a linear pathway, will exhibit a treenness equal to one and; as it is devoid of cycles its feedforwardness is also one. Preliminary results of this work is presented in figure 2.

1.3. The impact of network definition on result's interpretation using metabolism as a case study.

The search for a systems-level picture of metabolism as a web of molecular interactions provides a paradigmatic example of how the methods used to characterize a system can bias the interpretation of its functional meaning. Metabolic maps have been analyzed using novel

techniques from network theory, revealing some non-trivial, functionally relevant properties. These include a small-world structure and hierarchical modularity. However, some of these properties might actually result from an inappropriate way of defining network interactions. Starting from the so-called bipartite organization of metabolism, where the two meaningful subsets (namely reactions and metabolites) are considered, most current works use only one of the subsets by means of so called graph projections. Unfortunately, projected graphs often ignore relevant biological and chemical constraints, thus leading to statistical artifacts. Some of these drawbacks and alternative approaches need to properly addressed.

This idea has been recently published in Bioessays: Montañez R, Medina MA, Sole R and Rodriguez-Caso C. When metabolism meets topology: reconciling metabolite and reaction networks. Bioessays. 2010 Mar;32(3):246-56 and disseminated as oral presentation in:

Carlos Rodriguez-Caso. When metabolism meets topology. BIFI2010: Networks a framework for cross disciplinary applications. Zaragoza Spain 3-6 February 2010.

2. A preliminary construction of the Multiple sclerosis gene network through the use of databases

In relation to deliverable D28, technical problems on data acquisition has impeded the development of this part. As an alternative way to carry out the research on this topic, our group has participated in a initiative of Dr. Villoslada to reconstruct different the Multiple sclerosis and different immune system networks from protein-protein interaction information obtained from different biological databases: in particular Ingenuity and String databases. Our expertise has contributed to the elaboration of preliminary analysis performed by Dr. Esteban and Dr Villoslada. Figure 8 shows a preliminary network for MS obtained using this approach. Networks are going to be validated using null models in order to observe that the resulting networks obtained by specific filters are far from the expected by a blind screening.

3. Other scientific contributions in relation to ComplexDis project.

Collaboration in SYNLET EU project, giving rise the reasearch article:Human synthetic lethal inference as potential anti-cancer target gene detection. Conde-Pueyo N, Munteanu A, Solé RV, Rodriguez-Caso C BMC Syst. Biol. 2009. 3:116.

Metabolism study presented in work 1.2. is part of a Doctoral Thesis defended by Raul Montañez from University of Málaga.

Deviations from the project. Work programme.

Some technical problems have delayed the obtaining of the preliminary data for MS network

elucidation. This problem has been partially overcame by the construction of a computational inferred MS network. However, this is in a too much preliminary stage to produce a meaningful topological analysis. We have to stress that a number of tools developed during the duration of are ready for their application. It is expected that in the next weeks data from Dr. Villoslada's group will ready for a preliminary analysis of its topology.

List of deliverables.

Shared with WP8:

D25. A set of biological pathways involved in MS pathogenesis:
1. Description of a set of gene expression patterns and modules associated with MS. Submission to public databases and journal publication (month 18):
We have identified a set of immune and non-immune pathways differentially overrepresented in MS and we have validated one of then (neurotrophins pathway) in the animal model of MS
2. Description of new dynamics associated with the development of MS. journal publication and workshop (month 24):
We have found that the negative feed-back provided by regulatory T-cells is critical for establishing the prototypical relapsing-remitting dynamics in MS (Velez et al. Under review) workshop on systems biology Barcelona October 2010

D26. New mathematical and computational tools for analyzing the dynamics and robustness of evolving networks. Workshops and consortia web diffusion (month 30):
ASPI, COMPUTEGTOM, switching-random walker algorithm. New mathematical and computational tools for analyzing the dynamics and robustness of evolving networks. (due date of submission: month 18) Status - Research done, most of the publications and the results dissemination done. Hierarchy measures and entropy in directed graphs are articles in preparation. Submission in the next months. Workshop on systems biology Barcelona 2010

D27. A set of biological markers to monitor response to immunomodulatory therapy (IFNB) and candidate therapeutic pathways (which will require further validation). Journal publication and workshop (month 36): Jagged-Notch pathway, neurotrophin pathway Workshop on systems biology Barcelona 2010

D28: Topological analysis of MS network:
Analysis of the gene expression patterns and IP modules associated with MS. Submission to public databases and journal publication (due date of submission: month 24) Status - Computation implementation done. Preliminary version of MS network and immune system network was obtained in a computational approach in collaboration with Dr. Villoslada's group. Preliminary results in WP7 scientific report.
Analysis of new dynamics associated with the development of MS. Workshops and consortia web diffusion. (due date of submission: month 30) Status - Preliminary analysis of network topology for MS network and immune system network done. Details in WP7 scientific report. Pending of experimental validation

List of publications related to this research:

1. Goñi J, Rodríguez-Caso C, Solé RV, Villoslada P and Munteanu A (2008) Goals and pitfalls of the gene network inference methods: A comparative study from virtual microarrays and network dynamics. Proceedings of the II International Congress: Modeling and Computation on Complex Networks and related topics. Net-Works 2008 (Pamplona, June 9-11, 2008) pp91-97.
2. Rodriguez-Caso C, Valverde S and Conde-Pueyo N. K-scaffold subgraph: a measure to uncover the backbone of biological networks. Manuscript in preparation.
3. Rodriguez-Caso C, Corominas-Murta B and Solé RV. On the basic computational structure of gene regulatory networks. Mol. Biosyst 2009 *Mol. BioSyst.*, 2009, 5, 1617 - 1629,
4. *Rodriguez-Caso C and Corominas-Murta B. "Defining the dynamical backbone of gene regulatory networks".*9th International conference on Systems Biology ICSB2008 (Göteborg 22-28 august 2008). Poster presentation.
5. Montañez R, Medina MA, Sole R and Rodriguez-Caso. When metabolism meets topology: reconciling metabolite and reaction networks. Bioessays. 2010 Mar;32(3):246-56.
6. Carlos Rodriguez-Caso. When metabolism meets topology. BIFI2010: Networks a framework for cross disciplinary applications. Zaragoza Spain 3-6 February 2010.
7. *Bernat Corominas- Murtra*. Randomization of Directed Acyclic Networks BIFI2010: Networks a framework for cross disciplinary applications. Zaragoza Spain 3-6 February 2010.
8. Carlos Rodriguez-Caso Dynamical hierarchy and modularity in gene regulatory networks. BCNet workshop December 10-12, 2008
9. Rodriguez-Caso C and Conde-Pueyo N 2008. Topological analysis of cellular networks in *"Data Mining in Medical and Biological Research" ed* Kordic V, I-Tech Education and Publishing, Austria. In press.
10. Goñi J, Corominas-Murtra B, Sole R and Rodriguez-Caso C. Exploring the randomness of directed acyclic networks. To be submitted to Physical Rev E.