



Unraveling the complexity of gene networks: CRS4-Linkalab team among best performers at DREAM5 conference, Columbia University

November 16th, 2010: the team composed by Andrea Pinna, Nicola Soranzo, Vincenzo De Leo and Alberto de la Fuente has qualified as one of the best performers in the "Network inference challenge" in the DREAM5 competition, whose results were presented at the prestigious venue of Columbia University, NYC, during the annual "3rd Joint Conference on systems biology, regulatory genomics and reverse engineering challenges". The team called "ALF" and led by Alberto de la Fuente from the Center for Advanced Studies, Research and Development in Sardinia (CRS4), received the award also thanks to his collaboration with the Laboratory for the study of complex systems Linkalab.

The "Joint Conference" at Columbia University provided computational and experimental scientists, working in the field of regulatory genomics and systems biology, with the chance to compare their study approaches, aiming at a shared understanding of gene regulation mechanisms. "DREAM" stands for "Dialogue for Reverse Engineering Assessments and Methods". The goal of the challenge, at its 5th edition, is quite interesting: "Biological network inference and quantitative modeling of biological systems", explains Alberto de la Fuente, "are only at a starting point: nobody can tell how adequate their methods are. As a consequence of that, DREAM aims at unbiased evaluation of the proposed methods".

"The approach taken by DREAM", adds Andrea Pinna, "is to organize a yearly competition that teams from all over the world can enter to present their algorithms". "The teams taking part in the challenge receive datasets from experiments and need to predict results that are known only by the challenge organizers", explains Nicola Soranzo: "by checking how well the predictions achieved by the participants coincide with the actual results, we can better understand which algorithms are most useful to obtain reliable models from genomic datasets".

29 teams from all over the world, from top universities such as University of Chicago, Vanderbilt University, New York University and ParisTech, entered the DREAM5 network inference challenge. The challenge organizers gave the teams four datasets: three were real biological experiments (from bacteria and yeast) and one consisted of computer-simulated experiments. Each dataset consisted of measurements obtained from several experimental situations: i.e., underexpressing or overexpressing a gene, addition of drugs or introducing changes during a time series. The aim of the research was to identify the links between transcription factors and their target genes. "The challenge organizers", explains Vincenzo De Leo, the researcher from Linkalab who joined ALF team, "gave the teams two months to solve this task: we used this time to define the techniques to obtain the best results to answer the challenge questions".

Alberto de la Fuente, ALF team's scientific head, participates in DREAM competition from its second edition and already had been a best performer in DREAM challenges in 2007 and 2009. De la Fuente develops new methodologies for integrative analysis of gene expression and genotyping data ("systems genetics") to elucidate gene networks. He also focuses on the identification of dysfunctional gene networks from disease gene expression data ("differential networking"). Very often complex diseases are due to the altered interaction between sets of genes, rather than from changes occurred to a single gene. Understanding how these dynamics work using network inference can mean a progress in the research against these diseases. With the words of de la Fuente, network inference from large-scale genomics data works like that: "A cause-effect relation between the genes is established, for example applying specific gene perturbations and measuring which genes change their expression level; these genes are causally affected by the perturbed genes. Then you have to distinguish between direct and indirect cause-effect relations. The resulting network model, a gene network, only consists of connections with direct cause-effect relations between the genes".

Vincenzo De Leo works on a project in co-operation with CRS4 and funded by the Sardinian Region, applying techniques of differential networking analysis to data about diabetes in Sardinia. "Reconstructing complex gene networks is a very important step towards the comprehension of cellular behavior", explains De Leo: "the amounts of data biologists gathered in the last few years typically consist of lists of biological components and information about their interactions: therefore, they are very suitable to be analyzed by Complex networks Theory".

Nicola Soranzo is developing and evaluating computational methods to detect disease related changes in regulatory RNA networks. Andrea Pinna is developing a software tool, SysGenSIM, to simulate gene expression and genotyping data with large gene networks. This software has been used to create the dataset for another challenge at the DREAM5, the Systems genetics challenge, and was crucially important to the good performance of the ALF team in the network inference challenge.

"Our simulator enabled us to test many different methods on data simulated according to the experiments of the DREAM5 Network Inference challenge," explains Pinna, "so that we were able to select the most effective ones to apply to the challenge data".

"Interestingly", remarks Soranzo, "we trained our methods on our own simulated data and while we had a good performance on the real data, we didn't have such a good performance on the simulated data in the challenge: this implies that our simulator generates data which better reflect real data than those simulated by the challenge organizers".

For next year's DREAM6 competition, the members of the ALF team will contribute again to a Systems Genetics challenge that will be based on data produced with SysGenSIM, and are planning to participate in one or more other DREAM6 challenges.



CRS4 Bioinformatics Laboratory

CRS4 Bioinformatics is the computational biology laboratory of the Center for Advanced Studies, Research and Development in Sardinia. The research is directed towards the development of new computational models for the identification of disease-associated markers using functional genomic and systems biology approaches.

Relevant examples of the current research include: data management and integration of high-throughput datasets for the development of models for personalized-medicine; identification of biomarker candidates associated with sample stratification, disease susceptibility or clinical outcome and systems biology: inference of gene and protein networks in complex biological systems of biomedical interest and design and implementation of mathematical algorithm for the modeling of biological processes.

Laboratory for the study of complex systems Linkalab

During the last 10 years, Complex Network Theory has proved to be a valid approach to the analysis of the most diverse phenomena: scientific, social and technological. Linkalab mainly deals with the application aspects of Complex Network Theory, with a special attention on the computational side, the optimization of calculation systems and the scientific cloud computing. With international quality standards, Linkalab developed theoretical and calculation tools in order to work as a research service for P.A.s and companies showing Complex Systems Analysis problems. Linkalab also develops web-groupware systems and last generation e-learning systems for knowledge management activities on web 2.0 platforms.

Further reading:

The website of the conference (link to <http://recombsat.c2b2.columbia.edu/>)

The website of the DREAM project (link to http://wiki.c2b2.columbia.edu/dream/index.php/The_DREAM_Project)

The website of CRS4 Bioinformatica (link to <http://www.crs4.it/bioinformatics>)

The website of Linkalab (link to <http://www.linkalab.it/>)

Dr. Alberto de la Fuente's profile (link to <http://biowiki.crs4.it/biowiki/AlbertodelaFuente>)

The poster describing the ALF team's approach (link to http://www.linkalab.it/files/u77/DREAM5_poster.pdf)

The abstract of the work (link to http://www.linkalab.it/files/u77/DREAM5_abstract.pdf)

The official certificate (link to http://www.linkalab.it/files/u77/Press_release_DREAM5_certificate.pdf)

Figure 1 caption: The 'bow-tie' layout of a yeast gene network obtained by the group at CRS4 Bioinformatica by integrative analysis of gene expression and genotype data. Layout created by Gianmaria Mancosu using the software Pajek.

Dentro la complessità delle reti geniche: il team CRS4-Linkalab fra i migliori classificati nel DREAM5 alla Columbia University

Il team composto da Andrea Pinna, Nicola Soranzo, Vincenzo De Leo e Alberto de la Fuente si è classificato lo scorso 16 novembre fra i primi posti nella competizione sulla "Network Inference" durante il DREAM5, i cui risultati sono stati presentati nella prestigiosa sede della Columbia University di New York durante la terza Conferenza su biologia sistemica, genomica regolatoria e sull'inferenza delle reti biologiche. Il team, chiamato "ALF" e guidato da Alberto de la Fuente del CRS4, ha ottenuto questo premio anche grazie alla sua collaborazione con il Laboratorio per lo studio dei sistemi complessi Linkalab.

La "Joint Conference" alla Columbia ha fornito l'occasione a ricercatori computazionali e sperimentali attivi nei campi della genomica regolatoria e della biologia sistemica di confrontare i rispettivi approcci di studio, con l'auspicio di ottenere, un giorno, una comprensione condivisa dei meccanismi di regolazione dei geni. "DREAM" sta per "Dialogue for Reverse Engineering Assessments and Methods". L'obiettivo di questo appuntamento, giunto alla quinta edizione, è particolarmente interessante: "L'inferenza di reti biologiche e la costruzione di modelli quantitativi dei sistemi biologici", spiega Alberto de la Fuente, "sono ancora ai primi passi: nessuno può ancora dire quanto siano adeguati i loro strumenti. Di conseguenza," prosegue de la Fuente, "DREAM ambisce a una valutazione imparziale dei metodi proposti".

"L'approccio adottato da DREAM", aggiunge Andrea Pinna, "è di organizzare una competizione annuale alla quale gruppi da tutto il mondo possano partecipare per presentare i propri algoritmi". "Si forniscono ai partecipanti alla competizione i dati di un esperimento del quale devono predire i risultati, che sono noti solo agli organizzatori", spiega Nicola Soranzo. "Controllando fino a che punto le previsioni ottenute dai partecipanti coincidono con gli effettivi risultati", conclude, "possiamo capire meglio quali siano gli algoritmi più utili a ottenere modelli affidabili da dataset genomici".

Sono 29 i team, provenienti da tutto il mondo e da atenei quali Università di Chicago, Università di Vanderbilt, Università di New York e ParisTech, che hanno partecipato alla DREAM5 sull'inferenza di reti. Gli organizzatori hanno fornito ai team quattro dataset: tre erano reali esperimenti biologici (da batteri e lieviti) e uno era un esperimento simulato al computer. Ogni dataset consisteva di misurazioni ottenute da diverse situazioni sperimentali: ad esempio limitando o amplificando l'espressione di un gene, aggiungendo un farmaco oppure introducendo modifiche in una serie temporale. Lo scopo della ricerca era identificare le connessioni tra dei fattori di trascrizione e i loro geni bersaglio. "Gli organizzatori", spiega Vincenzo De Leo, il ricercatore che ha rappresentato Linkalab nel team ALF, "hanno concesso ai team due mesi di tempo, che abbiamo impiegato per definire le tecniche per ottenere il miglior risultato possibile e rispondere ai quesiti della competizione".

Alberto de la Fuente, responsabile scientifico del team, partecipa a questa competizione dalla sua seconda edizione e si è già classificato tra i primi posti nel 2007 e nel 2009. De la Fuente si occupa proprio di sviluppare nuove metodologie per l'analisi integrata dei dati di espressione genica e di genotipizzazione ("systems genetics") per l'interpretazione delle reti di geni. Lavora inoltre per identificare reti geniche disfunzionali, partendo dai dati di espressione genica di soggetti malati ("differential networking"). Molto spesso le malattie complesse dipendono, infatti, dall'alterata interazione tra set di geni, e non semplicemente da cambiamenti occorsi in un singolo gene. Comprendere le dinamiche di queste interazioni tramite l'inferenza di reti può rappresentare un progresso consistente per la ricerca su queste malattie. Il principio alla base dell'inferenza di reti dalle osservazioni su larga scala è, spiega de la Fuente, "stabilire un relazione di causa-effetto tra i geni, usando, per esempio, le perturbazioni di specifici geni e misurando quali di questi cambiano il loro livello di espressione; questi geni sono legati a quelli perturbati da un rapporto causale. Si distingue poi tra relazioni di causa-effetto dirette e indirette. Il modello di rete risultante, una rete genica, consiste solo di connessioni che corrispondono a relazioni di causa-effetto di tipo diretto tra i geni".

Vincenzo De Leo lavora a un progetto, in collaborazione con il CRS4 e finanziato dalla Regione Sardegna, che applica tecniche di "Differential networking Analysis" a dati relativi al diabete in Sardegna. "La ricostruzione delle complesse reti di interazioni fra i geni è un passo molto importante nella comprensione del funzionamento delle cellule", spiega De Leo. "Le moli di dati che, negli ultimi anni, i biologi hanno accumulato consistono tipicamente di liste di componenti biologici e di informazioni sulle loro interazioni: sono dunque molto adatte a essere analizzate con la Teoria delle Reti complesse", conclude De Leo.

Nicola Soranzo è impegnato nello sviluppo e nella valutazione di metodi computazionali per l'identificazione di cambiamenti legati a malattie nell'RNA regolatorio. Andrea Pinna, invece, sta lavorando allo sviluppo del software SysGenSIM per la simulazione di dati di espressione genica e di genotipizzazione con grandi reti geniche. Questo software è stato impiegato per creare i dataset per un'altra competizione del DREAM5, la "Systems genetics challenge", ed è stato essenziale per la performance del team ALF.

"Il simulatore ci ha permesso di testare molti metodi diversi sui dati simulati, proprio come nella competizione DREAM sull'inferenza di reti," spiega Pinna: "in questo modo abbiamo potuto scegliere il metodo più efficace per analizzare i dati della competizione".

"Curiosamente", sottolinea Soranzo, "abbiamo testato i nostri metodi sui dati simulati, ma durante la competizione abbiamo ottenuto i risultati migliori dai dati provenienti da esperimenti reali. Ciò significa che il nostro simulatore genera dati più realistici di quelli forniti dagli organizzatori della competizione".

Per la DREAM6 del prossimo anno, i membri del team ALF team contribuiranno nuovamente a una "Systems Genetics challenge", che si baserà proprio sui dati prodotti con SysGenSIM, e hanno in programma di partecipare a una o più delle competizioni del DREAM6.

Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna (CRS4)

CRS4 Bioinformatica è il laboratorio di biologia computazionale del Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna. Le aree di studio di interesse corrente riguardano la gestione e integrazione di dati generati con tecnologie high-throughput per lo sviluppo di modelli per la medicina personalizzata; l'identificazione di marcatori molecolari diagnostici e prognostici per la stratificazione dei campioni biologici; la biologia dei sistemi per l'inferenza di reti di interazione geniche e proteiche in sistemi biologici complessi di rilevanza biomedica e lo sviluppo e implementazione di modelli numerici per la simulazione di processi biologici.

Laboratorio per lo studio dei sistemi complessi Linkalab

Negli ultimi dieci anni la Teoria delle Reti complesse si è affermata come un nuovo approccio all'analisi dei più diversi fenomeni scientifici, sociali e tecnologici. Linkalab si occupa principalmente degli aspetti applicativi della Teoria delle Reti complesse, con un particolare accento sugli aspetti computazionali, l'ottimizzazione dei sistemi di calcolo e il cloud computing scientifico. Con standard qualitativi di livello internazionale, ha sviluppato gli strumenti teorici e di calcolo necessari per operare come un service di ricerca rivolto a istituzioni e aziende che presentano problematiche di Complex Systems analysis. Oltre a ciò, Linkalab sviluppa sistemi di groupware su web e di e-learning di ultima generazione per attività di diffusione della conoscenza sul Web 2.0.

Riferimenti:

Il sito della Conferenza alla Columbia University (link to <http://recombsat.c2b2.columbia.edu/>)

Il sito del DREAM project (link to http://wiki.c2b2.columbia.edu/dream/index.php/The_DREAM_Project)

Il sito del laboratorio di Bioinformatica del CRS4 (link to <http://www.crs4.it/bioinformatics>)

Il sito di Linkalab (link to <http://www.linkalab.it/>)

Il profilo del Dr. Alberto de la Fuente (link to <http://biowiki.crs4.it/biowiki/AlbertodelaFuente>)

Il poster che illustra il lavoro del team ALF (link to http://www.linkalab.it/files/u77/DREAM5_poster.pdf)

L'abstract del lavoro (link to http://www.linkalab.it/files/u77/DREAM5_abstract.pdf)

Il certificato ufficiale (link to http://www.linkalab.it/files/u77/Press_release_DREAM5_certificate.pdf)

Didascalia figura 1: La configurazione 'bow-tie' della rete genica del lievito ottenuta dal gruppo di Bioinformatica del CRS4 dall'integrazione delle analisi di espressione genetica e dai dati genotipici. Figura creata da Gianmaria Mancosu utilizzando il software Pajek.