University of Victoria | Graduate Studies

PROGRAMME

The Final Oral Examination
for the Degree of

DOCTOR OF PHILOSOPHY
(Computer Science)

Tatiana Gutiérrez-Bunster

2008	University of Concepción	MSc
2001	University of Bío Bío	BSc

“Towards a Better Understanding of
Protein-Protein Interaction Networks”

Wednesday, December 17, 2014
9:30a.m.
Engineering Computer Science (ECS), Room 468

Supervisory Committee:
Dr. Ulrike Stege, Department of Computer Science, UVic
(Co-Supervisor)
Dr. Alex Thomo, Department of Computer Science, UVic
(Co-Supervisor)
Dr. John Taylor, Department of Biology, UVic
(Co-Supervisor)
Dr. Christopher Upton, Department of Biochemistry and Microbiology,
UVic (Outside Member)

External Examiner:
Dr. Kishan Mehrotra, Department of Engineering and Computer Science
Syracuse University

Chair of Oral Examination:
Dr. Michael Emme,
Department of Curriculum and Instruction, UVic

Abstract

Proteins participate in the majority of cellular processes that are essential for life. To determine the function of a protein it is not sufficient to solely know its sequence, its structure in isolation, or how it works individually. Additionally, we need to know how the protein interacts with other proteins in biological networks. This is because most of the proteins perform their main function through interactions. This thesis sets out to improve the understanding of protein-protein interaction networks (PPINs). For this, we propose three approaches:

(1) *Studying measures and methods used in social and complex networks.*The methods, measures, and properties of social networks allow us to gain an understanding of PPINs via the comparison of different types of network families. We studied and evaluated models that describe social networks to see which models are useful in describing biological networks. We investigate the similarities and differences in terms of the network community profile and centrality measures.

(2) *Studying PPINs and their role in evolution*. We are interested in the relationship of PPINs and the evolutionary changes between species. We investigate whether the centrality measures are correlated with the variability and similarity in orthologous proteins.

(3) *Studying protein features that are important to evaluate, classify, and predict interactions*. We identify which type of energy values contributes better to predicting protein-protein interactions. We argue that the number of energetic features and their contribution to the interactions can be a key factor in predicting transient and permanent interactions.

Contributions of this thesis include: (1) We identify the best community sizes in PPINs. This finding will help to identify important groups of interacting proteins in order to better understand their particular interactions. We furthermore find that the generative model describing biological networks is very different from the model describing social networks. We showed that the best community size for PPINs is around ten. We revealed differences in terms of the network community profile and correlations of centrality measures; (2) We outline a method to test correlation of centrality measures with the percentage of sequence similarity and evolutionary rate for orthologous proteins. We conjecture that a strong correlation exists. While not obtaining

positive results for our data, we believe that the reason for this is the unreliability of today's data sets; Therefore, (3) we investigate a method to discriminate energetic features of protein interactions that in turn will improve the PPIN data. The use of multiple data sets makes possible to identify the energy values that are useful to classify interactions. For each data set, we performed Random Forest and Support Vector Machine with linear, polynomial, radial, and sigmoid kernels. The accuracy obtained in this analysis reinforces the idea that energetic features in the protein interface help to discriminate between transient and permanent interactions.

Awards, Scholarships, Fellowships

2010-2014 Becas-Chile Award-Fellowship for PhD Studies (Chile).

Presentations

1. Improving energetic feature selection to classify protein-protein interactions. Gutiérrez-Bunster, T.; Poo-Caamano, G. International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics.

2. How do biological networks differ from social networks? (an experimental study). Gutiérrez-Bunster, T.; Stege, U.; Thomo, A.; Taylor, J. International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics 2014.

Conference Proceedings

1. Gutiérrez-Bunster, Tatiana, and German Poo-Caamano. "Improving energetic feature selection to classify protein-protein interactions." *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*. IEEE, 2014.

2. Gutiérrez-Bunster, Tatiana; Stege, Ulrike; Thomo, Alex and Taylor, John. "How do biological networks differ from social networks? (an experimental study)." *Advances in Social*

Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on. IEEE, 2014.