

Donnelly Centre

for Cellular + Biomolecular Research UNIVERSITY OF TORONTO



Donnelly Seminar Series "Data Fusion of Everything"



Dr. Blaž Zupan, PhD

Professor Faculty of Computer and Information Science University of Ljubljana Department of Molecular and Human Genetics Bayor College of Medicine

> Thursday, June 4, 2015 | 11:00 am Donnelly Centre Red Seminar Room

Abstract:

Have you ever been overwhelmed by data---not only by their volume but by their sheer multitude? Since you are a researcher at Donnelly Center, you most certainly have. In life science, the data abounds. One of the grand challenges in bioinformatics is to infer a predictive model by jointly considering, say, gene expression, interactions, functional annotations, phenotype information, various ontologies, disease markers, structural properties of chemicals, and Facebook (ok, I perhaps went too far with the last item). That is, by considering all available information, even if only circumstantially related to the problem at hand. At University of Ljubljana we have developed a computational approach that uses collective matrix tri-factorization and can consider such diverse data sets. Trifactorization infers a joint latent data model. The model can be used for various data mining tasks, such as class prediction and ranking. Our experiments show that through a broad integration of heterogeneous data sets we can substantially increase the accuracy. In the talk, I will present the intuition behind data fusion by matrix tri-factorization, and show how it was used to find new bacterial response genes in social amoeba Dictyostelium.

Host: Dr. Fritz Roth