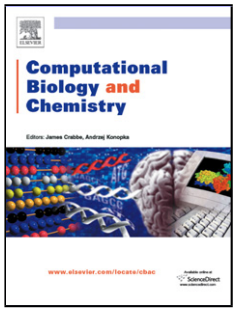

Title: Laplacian Normalization and Random Walk on Heterogeneous Networks for Disease-gene Prioritization

Author: Zhi-Qin Zhao Guo-Sheng Han Zu-Guo Yu Jinyan Li

PII: S1476-9271(15)00025-0
DOI: http://dx.doi.org/doi:10.1016/j.compbiolchem.2015.02.008
Reference: CBAC 6400

To appear in: *Computational Biology and Chemistry*

Received date: 31-12-2014
Accepted date: 3-2-2015

Please cite this article as: Zhi-Qin Zhao, Guo-Sheng Han, Zu-Guo Yu, Jinyan Li, Laplacian Normalization and Random Walk on Heterogeneous Networks for Disease-gene Prioritization, *Computational Biology and Chemistry* (2015), http://dx.doi.org/10.1016/j.compbiolchem.2015.02.008

Laplacian Normalization and Random Walk on Heterogeneous Networks for Disease-gene Prioritization

Zhi-Qin Zhao [1], Guo-Sheng Han [1,§], Zu-Guo Yu [1,2,*], Jinyan Li [3,*]

[1] *Hunan Key Laboratory for Computation and Simulation in Science and Engineering and Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, Xiangtan University, Xiangtan, Hunan 411105, China.*
[2] *School of Mathematical Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, Q4001, Australia.*
[3] *Advanced Analytics Institute & Centre for Health Technologies, University of Technology Sydney, Broadway, NSW 2007, Australia.*

[§] *Joint first author.*

Abstract

Random walk on heterogeneous networks is a recently emerging approach to effective disease gene prioritization. Laplacian normalization is a technique capable of normalizing the weight of edges in a network. We use this technique to normalize the gene matrix and the phenotype matrix before the construction of the heterogeneous network, and also use this idea to define the transition matrices of the heterogeneous network. Our method has remarkably better performance than the existing methods for recovering known gene-phenotype relationships. The Shannon information entropy of the distribution of the transition probabilities in our networks is found to be smaller than the networks constructed by the existing methods, implying that a higher number of top-ranked genes can be verified as disease genes. In fact, the most probable gene-phenotype relationships ranked within top 3 or top 5 in our gene lists can be confirmed by the OMIM database for many cases. Our algorithms have shown remarkably superior performance over the state-of-the-art algorithms for recovering gene-phenotype relationships. All Matlab codes can be available upon

*Corresponding author, Email: yuzg1970@yahoo.com and jinyan.li@uts.edu.au