

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/20051> holds various files of this Leiden University dissertation.

Author: Rahmani, Hossein

Title: Analysis of protein-protein interaction networks by means of annotated graph mining algorithms

Issue Date: 2012-10-30

Analysis of Protein-Protein Interaction Networks

by Means of Annotated Graph Mining Algorithms

Hossein Rahmani

Analysis of Protein-Protein Interaction Networks

by Means of Annotated Graph Mining Algorithms

PROEFSCHRIFT

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus prof.mr. P.F. van der Heijden,

volgens besluit van het College voor Promoties

te verdedigen op dinsdag 30 oktober 2012

klokke 15.00 uur

door

Hossein Rahmani

geboren te Tehran, Iran
in 1983

PhD committee

Promotor:	Prof. dr. J.N. Kok	Leiden University
Co-promotor:	Dr. H. Blockeel	Leiden University

Other members:

Prof. dr. T. Baeck	Leiden University
Dr. Ir. F.J. Verbeek	Leiden University
Prof. dr. K. Marchal	Universiteit Gent



The work reported in this thesis has been carried out at the Leiden Institute of Advanced Computer Science at Leiden University, under the auspices of the research school IPA (Institute for Programming research and Algorithmics). This research is supported by the Dutch Science Foundation (NWO) through a VIDI grant.

Copyright © 2012 by Hossein Rahmani. All rights reserved.

The cover illustration depicts a part of Human Disease Network discussed in this thesis.

Cover design by: Hossein Rahmani.

Published by: Uitgeverij BOXPress, 's-Hertogenbosch

ISBN: 978-90-8891-488-1

Contents

1	Introduction	1
1.1	General Introduction	2
1.2	Knowledge Discovery Process	2
1.2.1	Input Data	2
1.2.2	Data Pre-Processing	4
1.2.3	Machine Learning	6
1.2.4	Data Post-Processing	9
1.2.5	Knowledge Discovery Tool: WEKA	10
1.3	PPI Network and Its Open Problems	10
1.3.1	Data Set	11
1.3.2	Predicting the Functions of Proteins in PPI Networks	13
1.3.3	Predicting Cancer-related Proteins in PPI networks	16
1.3.4	Predicting Disease-related Proteins in PPI network	19
1.3.5	Conclusions	21
2	Predicting Proteins Functions Using Network Global Information	23
2.1	abstract	24
2.2	Introduction	24
2.3	Problem Statement	25
2.4	Related work	26
2.5	A global description of proteins	27
2.6	Experiments	28
2.6.1	Datasets	29
2.6.2	Comparison of Learners	30
2.6.3	Comparison with a transductive method	31
2.6.4	Different Number of Important Proteins	33
2.6.5	Function Levels	36
2.7	Conclusions	37

3	Predicting Proteins Functions Using Collaborative Functions	43
3.1	abstract	44
3.2	Introduction	44
3.3	Related work	45
3.4	Two collaboration-based methods	46
3.4.1	A Reinforcement Based Function Predictor	46
3.4.2	SOM Based Function Predictor	47
3.5	Experiments	48
3.5.1	Dataset and Annotation Data	48
3.5.2	Parameter Tuning	49
3.5.3	Comparison to previous methods	50
3.5.4	Extending Majority Rule	51
3.6	Conclusion	51
4	Predicting Cancer-Related Proteins Using Network Contextual Information	55
4.1	Abstract	56
4.2	Introduction	56
4.3	Methods	58
4.3.1	Formal Definition	58
4.3.2	Protein Description Based on Functional Context	58
4.3.3	Protein Description Based on Structural Context	59
4.3.4	Protein Description Based on Functional and Structural Context	61
4.4	Results	61
4.4.1	Dataset	61
4.4.2	Biological Interpretation of the Most Relevant Functions	62
4.4.3	Biological Interpretation of the Most Discriminative Proteins	65
4.4.4	Comparing Different Contextual Methods	66
4.4.5	Comparing with Previous Methods	69
4.4.6	Random Feature Selection	72
4.4.7	Capacity Identification of New Cancer-Related Proteins	75
4.5	Discussion	76
4.6	Acknowledgments	78
5	Predicting Cancer-Related Proteins using Interaction-based Features	79
5.1	Introduction	80
5.2	Interaction-based feature selection	80
5.3	Results	81
5.4	Conclusions	83

6	Predicting Disease-Related Proteins Using Human Disease Network	85
6.1	abstract	86
6.2	Introduction	86
6.3	Methods	87
6.3.1	Formal Definition	87
6.3.2	Human Disease Network	87
6.3.3	Recommended Prediction Methods	88
6.4	Empirical Results	91
6.4.1	Dataset	91
6.4.2	Comparing Recommended Prediction Methods	91
6.4.3	Informative Human Disease Network	93
6.4.4	Biological Interpretation of the Pruned HDN	95
6.4.5	Predicting Disease-Related Proteins using the Pruned HDN	96
6.4.6	Case Study: Long QT Syndrome	98
6.5	Compare individual and network based prediction for LQTS	102
6.6	Conclusions	103
6.7	Acknowledgment	104
7	Conclusions	105
7.1	Introduction	106
7.2	Shortest-Path Distance and Anova-based Feature Selection	106
7.3	Collaborative Functions	106
7.4	Network Contextual Information	107
7.5	Interaction-based Chi-square	108
7.6	Informative Human Disease Network	108
7.7	Future Works	109
	Bibliography	111
	Summary	133
	Samenvatting	135
	List Of Publications	137
	Curriculum Vitae	139
	Acknowledgments	141

