PREDICTING PROTEIN FUNCTIONS FROM PROTEIN INTERACTION NETWORKS HON NIAN CHUA, LIMSOON WONG pdf

1: - NLM Catalog Result

Predicting Protein Functions from Protein Interaction Networks BACKGROUND Protein Function Prediction using Sequence Information Sequence homology is the classical methodology used to infer the function of a novel protein.

Proteins are essential macromolecules of life and thus understanding their function is of great importance. Methods for determining protein function have shifted th Methods for determining protein function have shifted their focus from targeting specific proteins based solely on sequence homology to analyses of the entire proteome based on protein-protein interaction PPI networks. Since proteins aggregate to perform a certain function, analyzing structural properties of PPI networks may provide useful clues about the biological function of individual proteins, protein complexes they participate in, and even larger subcellular machines. We design a sensitive graph theoretic method for comparing local structures of node neighborhoods that demonstrates that in PPI networks, biological function of a node and its local network structure are closely related. The method groups topologically similar proteins under this measure in a PPI network and shows that these protein groups belong to the same protein complexes, perform the same biological functions, are localized in the same subcellular compartments, and have the same tissue expressions. Moreover, we apply our technique on a proteome-scale network data and infer biological function of yet unclassified proteins demonstrating that our method can provide valuable guidelines for future experimental research. Data is available upon request. Show Context Citation Context Other approaches use the Protein-Protein Interaction PPI networks are believed to be important sources of information related to biological processes and complex metabolic functions of the cell. The presence of biologically relevant functional modules in these networks has been theorized by many researchers. However, the application of traditional clustering algorithms for extracting these modules has not been successful, largely due to the presence of noisy false positive interactions as well as specific topological challenges in the network. In this paper, we propose an ensemble clustering framework to address this problem. For base clustering, we introduce two topology-based distance metrics to counteract the effects of noise. We develop a PCA-based consensus clustering technique, designed to reduce the dimensionality of the consensus problem and yield informative clusters. We also develop a soft consensus clustering variant to assign multifaceted proteins to multiple functional groups. We conduct an empirical evaluation of different consensus techniques using topology-based, information theoretic and domain-specific validation metrics and show that our approaches can provide significant benefits over other state-of-theart approaches. Our analysis of the consensus clusters obtained demonstrates that ensemble clustering can a produce improved biologically significant functional groupings; and b facilitate soft clustering by discovering multiple functional associations for proteins. The neighborhood-based similarity metric is defined as: High-throughput experimental methods, such as yeast-two-hybrid and phage display, have fairly high levels of false positives and false negatives. Thus the list of protein-protein interactions detected by such experiments would need additional wet laboratory validation. It would be useful if the li It would be useful if the list could be prioritized in some way. Advances in computational techniques for assessing the reliability of protein-protein interactions detected by such high-throughput methods are reviewed in this paper, with a focus on techniques that rely only on topological information of the protein interaction network derived from such high-throughput experiments. In particular, we discuss indices that are abstract mathematical characterizations of networks of reliable protein-protein interactionsâ€”e. We also present indices that are based on explicit motifs associated with true-positive protein interactionsâ€”e. Protein-protein interaction network, graph mining, network motif. A paper with Sung the 7th author soon f Function-function correlated multi-label protein function prediction over interaction networks by Hua Wang, Heng Huang, Chris Ding , " Many previous computational methods for protein function prediction make prediction one function at a time, fundamentally, which is equivalent to assume the functional categories of proteins to be isolated. However, biological processes are highly correlated and usu-ally intertwined togeth However, biological processes are

PREDICTING PROTEIN FUNCTIONS FROM PROTEIN INTERACTION NETWORKS HON NIAN CHUA, LIMSOON WONG pdf

highly correlated and usu-ally intertwined together to happen at the same time, therefore it would be beneficial to consider protein function prediction as one indivisible task and treat all the functional categories as an integral and correlated prediction target. By leveraging the function-function correlations, it is expected to achieve improved overall predictive accuracy. To this end, we develop a novel network based protein function prediction approach, under the framework of multi-label classification in machine learning, to utilize the function-function correlations. Besides formulating the function-function correlations in the optimization objective explicitly, we also exploit them as part of the pairwise protein-protein similarities implicitly. We evaluate the proposed approach on Saccharomyces cerevisiae species. The encouraging experimental re-sults demonstrate the effectiveness of the proposed method. Later researchers used global optimization approaches to improve the protein function predictions by taking into account the full topology of ne Computational classification of gene expression profiles into distinct disease phenotypes has been highly successful to date. Still, robustness, accuracy, and biological interpretation of the results have been limited, and it was suggested that use of protein interaction information jointly with the Still, robustness, accuracy, and biological interpretation of the results have been limited, and it was suggested that use of protein interaction information jointly with the expression profiles can improve the results. Here, we study three aspects of this problem. First, we show that interactions are indeed relevant by showing that co-expressed genes tend to be closer in the network of interactions. Second, we show that the improved performance of one extant method utilizing expression and interactions is not really due to the biological information in the network, while in another method this is not the case. Finally, we develop a new kernel methodâ€”called NICKâ€”that integrates network and expression data for SVM classification, and demonstrate that overall it achieves better results than extant methods while running two orders of magnitude faster. One obvious limitation which holds for any PPI-based approach is the network quality. In addition, most network edges originated from in vitro experiments, which may differ from in-vivo conditions. Also, semantically, networks are compiled from pairwise relations, and it is hard to Recent proteome-wide screening efforts have made available genome-wide, high-throughput protein-protein interaction PPI maps for several model organisms. This has enabled the systematic analysis of PPI networks, which has become one of the primary challenges for the system biology community. Here we address the problem of predicting the functional classes of proteins i. We present a maximum likelihood formulation of the problem and the corresponding learning and inference algorithms. The time complexity of both algorithms is linear in the size of the PPI network and our experimental results show that their accuracy in the functional prediction outperforms current existing methods. Because of this, the performance of functional flow is the reference for our algorithm. Experimental results will show that our method achi Wâ€” by Mark N. Wass, Geraint Barton, Michael J. Only a small fraction of known proteins have been functionally characterized, making protein function prediction essential to propose annotations for uncharacterized proteins. In recent years many function prediction methods have been developed using various sources of biological data from protein s In recent years many function prediction methods have been developed using various sources of biological data from protein sequence and structure to gene expression data. CombFunc incorporates ConFunc, our existing function prediction method, with other approaches for function prediction that use protein sequence, gene expression and proteinâ€”protein interaction data. In benchmarking on a set of proteins CombFunc obtains preci-sion and recall of 0. For biological process GO terms precision of 0. In ConFunc we used conserved residues representative of individual GO terms to predict protein function Some methods combine predictions from multiple sources of data 20â€” Estimating support for protein-protein interaction data with applications to function prediction by Erliang Zeng, Chris Ding, Giri Narasimhan, Stephen R. Holbrook - Comput Syst Bioinformatics Conf " Almost every cellular process requires the interactions of pairs or larger complexes of proteins. High throughput protein-protein interaction PPI data have been generated using techniques such as the yeast two-hybrid systems, mass spectrometry method, and many more. Such data provide us with a new Such data provide us with a new perspective to predict protein functions and to generate protein-protein interaction

networks, and many recent algorithms have been developed for this purpose. However, PPI data generated using high throughput techniques contain a large number of false positives. In this paper, we have proposed a novel method to evaluate the support for PPI data based on gene ontology information. Thus semantic similarity between genes serves as a metric of support for PPI data. Taking it one step further, new function prediction approaches are also being proposed with the help of the proposed metric of the support for the PPI data. These new function prediction approaches outperform their conventional counterparts. New evaluation methods are also proposed. Understanding the molecular mechanisms of life requires decoding the functions of the proteins in an organism. Various high-throughput experimental techniques have been developed to characterize biological systems at the genome scale. A fundamental challenge of the post-genomic era is to assign biol A fundamental challenge of the post-genomic era is to assign biological functions to all the proteins encoded by the genome using high-throughput biological data. To address this challenge, we propose a novel Laplacian Network Partitioning incorporating function category Correlations LNPC method to predict protein function on proteinprotein interaction PPI networks by optimizing a Laplacian based quotient objective function that seeks the optimal network configuration to maximize consistent function assignments over edges on the whole graph. Unlike the existing approaches that have no unique optimization solutions, our optimization problem has unique global solution by eigen-decomposition methods. The correlations among protein function categories are quantified and incorporated into a correlated protein affinity graph which is integrated into the PPI graph to significantly improve the protein function prediction accuracy. Functional characterization of genes and their protein products is essential to biological and clinical research. Yet, there is still no reliable way of assigning functional annotations to proteins in a high-throughput manner. In this chapter, the authors provide an introduction to the task of autom In this chapter, the authors provide an introduction to the task of automated protein function prediction. They discuss about the motivation for automated protein function prediction, the challenges faced in this task, as well as some approaches that are currently available. In particular, they take a closer look at methods that use protein-protein interaction for protein function prediction, elaborating on their underlying techniques and assumptions, as well as their strengths and limitations.

PREDICTING PROTEIN FUNCTIONS FROM PROTEIN INTERACTION NETWORKS HON NIAN CHUA, LIMSOON WONG pdf

2: DBLP: Limsoon Wong

Protein Function Prediction Using Protein-Protein Interaction Networks Hon Nian Chua, Guimei Liu, and Limsoon Wong Abstract Proteins perform biological functions by participating in a large number.

Protein-Protein Interaction PPI networks are believed to be important sources of information related to biological processes and complex metabolic functions of the cell. The presence of biologically relevant functional modules in these networks has been theorized by many researchers. However, the application of traditional clustering algorithms for extracting these modules has not been successful, largely due to the presence of noisy false positive interactions as well as specific topological challenges in the network. In this paper, we propose an ensemble clustering framework to address this problem. For base clustering, we introduce two topology-based distance metrics to counteract the effects of noise. We develop a PCA-based consensus clustering technique, designed to reduce the dimensionality of the consensus problem and yield informative clusters. We also develop a soft consensus clustering variant to assign multifaceted proteins to multiple functional groups. We conduct an empirical evaluation of different consensus techniques using topology-based, information theoretic and domain-specific validation metrics and show that our approaches can provide significant benefits over other state-of-theart approaches. Our analysis of the consensus clusters obtained demonstrates that ensemble clustering can a produce improved biologically significant functional groupings; and b facilitate soft clustering by discovering multiple functional associations for proteins. Show Context Citation Context Second, even if the network is assumed to be noise free, partitioning the network using classical graph partitioning or clustering schemes is inherently difficult. A common characteristic of PPI net An assessment of the uses of homologous interactions by Ramazan Saeed - Bioinformatics , " Protein-protein interactions have proved to be a valua-ble starting point for understanding the inner workings of the cell. Computational methodologies have been built which both predict interactions and use interaction datasets in order to predict other pro-tein features. Here we examine and demonstrate the usefulness of homologous interactions in predicting good quality positive and negative interaction datasets. We generate GSP interaction sets as subsets from expe-rimental data using only interaction and sequence information. We can therefore produce sets for several species many of which at present have no identified GSPs. Comprehensive error rate testing demonstrates the power of the method. We also show how the use of our datasets significantly improves the predictive power of algorithms for interaction prediction and function prediction. Furthermore we generate GSN interaction sets for yeast and examine the use of homology along with other protein properties such as localisation, expression and function. Using a novel method to assess the accuracy of a negative interaction set we find that the best single selector for negative interactions is a lack of co-function. However, an integrated method using all the characteristics shows significant improvement over any current method for identifying GSN interactions. The nature of homologous interactions is also examined and we demonstrate that interologs are found more commonly within species than across species. GSP sets built using our homologous verification method are demonstrably better than standard sets in terms of pre-dictive ability. We can build such GSP sets for several species. When generating GSNs we show a combination of protein features and lack of homologous interactions gives the highest quality interaction sets. Alternatively a range of methods exist that only require topological information to ascertain a GSP Saito et al. Functional characterization of genes and their protein products is essential to biological and clinical research. Yet, there is still no reliable way of assigning functional annotations to proteins in a high-throughput manner. In this chapter, the authors provide an introduction to the task of autom In this chapter, the authors provide an introduction to the task of automated protein function prediction. They discuss about the motivation for automated protein function prediction, the challenges faced in this task, as well as some approaches that are currently available. In particular, they take a closer look at methods that use protein-protein interaction for protein function prediction, elaborating on their underlying techniques and assumptions, as well as their

strengths and limitations. The author expresses his immense gratitude to two great mentors who have been nurturing and supporting him in every way possible. And to Professor Marek Sergot for sharing his wisdom, h And to Professor Marek Sergot for sharing his wisdom, his knowledge and insights, and most importantly, for keeping the author in college. He also thanks his colleagues, in particular, Dr Yiehou Lee, for his invaluable advice and help with the biological interpretations. He is also grateful to Professor Maxey Chung for providing the liver cancer dataset that this body of work is mostly based on. Finally, he thanks his friends for remaining in spite of his incessant delusional tirades.

PREDICTING PROTEIN FUNCTIONS FROM PROTEIN INTERACTION NETWORKS HON NIAN CHUA, LIMSOON WONG pdf

3: Using indirect protein interactions for the prediction of Gene Ontology functions - CORE

Proteins perform biological functions by participating in a large number of interactions, ranging from transient interactions in signaling pathways to permanent interactions within stable complexes. Studies have shown that the immediate interaction neighborhood of a protein can be used to infer its.

This article has been cited by other articles in PMC. Abstract Background Protein complexes play an important role in cellular mechanisms. Recently, several methods have been presented to predict protein complexes in a protein interaction network. In these methods, a protein complex is predicted as a dense subgraph of protein interactions. However, interactions data are incomplete and a protein complex does not have to be a complete or dense subgraph. Results We propose a more appropriate protein complex prediction method, CFA, that is based on connectivity number on subgraphs. We evaluate CFA using several protein interaction networks on reference protein complexes in two benchmark data sets MIPS and Aloy , containing and 61 known complexes respectively. We show that CFA predicts more complexes correctly at a competitive level of precision. Conclusions Many real complexes with different connectivity level in protein interaction network can be predicted based on connectivity number. Our CFA program and results are freely available from http: Background Several groups have produced a large amount of data on protein interactions [1 - 9]. It is desirable to use this wealth of data to predict protein complexes. Several methods have been applied to protein inter-actome graphs to detect highly connected subgraphs and predict them as protein complexes [10 - 25]. The main criterion used for protein complex prediction is cliques or dense subgraphs. Spirin and Mirny proposed the clique-finding and super-paramagnetic clustering with Monte Carlo optimization approach to find clusters of proteins [10]. On the other hand, the Markov CLuster algorithm MCL [26 , 27] simulates a flow on the network by using properties of the adjacency matrix. MCL partitions the graph by discriminating strong and weak flows in the graph. It is a cost-based local search algorithm that explores the solution space to minimize a cost function, which is calculated based on the numbers of intra-cluster and inter-cluster edges. However, many biological data sources contain noise and do not contain complete information due to limitations of experiments. Recently, some computational methods have estimated the reliability of individual interaction based on the topology of the protein interaction network PPI network [23 , 28 , 29]. The Protein Complex Prediction method PCP [30] uses indirect interactions and topological weight to augment protein-protein interactions, as well as to remove interactions with weights below a threshold. PCP employs clique finding on the modified PPI network, retaining the benefits of clique-based approaches. Following these past works, we model the PPI network with a graph, where vertices represent proteins and edges represent interactions between proteins. Our algorithm is based on finding maximal k-connected subgraphs. These candidate clusters are then filtered to remove i clusters having less than four proteins and ii clusters having a large diameter. Our algorithm produces results that are comparable or better than these existing algorithms on real complexes of [32 , 33]. Preliminaries Generally, a complete or a dense subgraph of a protein interaction network is proposed to be a protein complex. So we need to define a criterion to predict protein complexes with different topology.

PREDICTING PROTEIN FUNCTIONS FROM PROTEIN INTERACTION NETWORKS HON NIAN CHUA, LIMSOON WONG pdf

4: Biological data mining in protein interaction networks (eBook,) [www.enganchecubano.com]

Protein-protein interactions (PPIs) are important for understanding the cellular mechanisms of biological functions, but the reliability of PPIs extracted by high-throughput assays is known to be low.

Hon Nian Chua - g nus. This is an open access article distributed under the terms of the Creative Commons Attribution License http: Protein-protein interaction has been used to complement traditional sequence homology to elucidate protein function. Most existing approaches only make use of direct interactions to infer function, and some have studied the application of indirect interactions for functional inference but are unable to improve prediction performance. We have previously proposed an approach, FS-Weighted Averaging, which uses topological weighting and level-2 indirect interactions protein pairs connected via two interactions for predicting protein function from protein interactions and have found that it yields predictions with superior precision on yeast proteins over existing approaches. Here we study the use of this technique to predict functional annotations from the Gene Ontology for seven genomes: Saccharomyces cerevisiae, Drosophila melanogaster, Caenorhabditis elegans, Arabidopsis thaliana, Rattus norvegicus, Mus musculus, and Homo sapiens. Our analysis shows that protein-protein interactions provide supplementary coverage over sequence homology in the inference of protein function and is definitely a complement to sequence homology. We also find that FS-Weighted Averaging consistently outperforms two classical approaches, Neighbor Counting and Chi-Square, across the seven genomes for all three categories of the Gene Ontology. By randomly adding and removing interactions from the interactions, we find that Weighted Averaging is also rather robust against noisy interaction data. We have conducted a comprehensive study over seven genomes. We conclude that FS-Weighted Averaging can effectively make use of indirect interactions to make the inference of protein functions from protein interactions more effective. Furthermore, the technique is general enough to work over a variety of genomes. Background fraction of newly discovered sequences have identifiable Although sequence similarity search has proven useful in homologous genes in current databases. Second, the most many cases, it has fundamental limitations. First, only a prominent vertebrate organisms in GenBank have only a Page 1 of 13 page number not for citation purposes BMC Bioinformatics , 8 Suppl 4: Associative properties that have been used to coverage during function inference, they contain too infer function not evident from sequence homology many false positives to be useful. In order to reduce the include: Many approaches have also been functions. FS-Weight improves the precision of function proposed for utilizing protein-protein interaction data for inference, while the inclusion of FS-weighted level-2 functional inference []. A simple but effective neighbors improves both sensitivity and precision. A new approach is to assign a protein with the function that method, FS-Weighted Averaging, which incorporates indi- occurs most frequently in its interaction partners [10]. Some tions for S. Others apply Here we study how FS-Weighted Averaging performs in global optimization techniques, such as Markov random predicting protein functions from the protein-protein fields [15,16] and simulated annealing [17], to propagate interaction maps of seven genomes. We also study how predictions so that the function of proteins without char- the approach performs on noisy interaction data and on acterized neighbors can be predicted. Most of these predicted interactions. Finally, we show some examples of approaches use the observation that a protein often shares novel functions predicted for uncharacterized proteins in functions with proteins that interact with it i. However, proteins that interact with the biologically significant. We observed that there are proteins that do age over sequence homology. We look at two well-studied not share any function with their immediate interaction genomes, S. Out of many more functions can be suggested from interaction 4, annotated yeast proteins, only 1,, or Of the can be suggested from indirect interaction partners on top remaining proteins, , or A tion or annotation, we proposed indirect functional associ- higher E-value will translate to less significant sequence ation as a reasonable explanation for this observation similarity and vice versa. Hence, a higher E-value thresh- [18]. Such an indirect functional association can be con- old will provide better coverage at the expense of lower

sidered as an instance of the "guilt by association" princi- precision. The protein interaction network is then examined to the "association" is precisely the set of proteins that they find out the number of additional known annotations both interact with, namely the level-1 neighbors. It is that can be suggested from direct and indirect interac- plausible that two proteins that interact with a common tions. Figure 2 summarizes the findings of this analysis. Nodes represent proteins, while edges represent interactions. Direct interactions between labelled proteins are indicated by red lines, while indirect interactions between labelled proteins are indicated by blue lines. We also observe that indirect inter- mark, we want to study the prediction performance of FS- actions can provide significant additional coverage over Weighted Averaging over a variety of different genomes. Proteins with known functions are randomly assigning higher weights to direct and indirect interac- divided into 10 groups. In each run, the annotations for tions that involve function sharing. Here we study how one group of proteins will be hidden, and their functions well FS-Weight scores reflect function similarity. All direct predicted based on the annotations for the remaining and indirect interactions are first weighted using FS- groups and the interaction data. The hidden annotations Weight. For each unique score, we compute the fraction of are not available to any preprocessing steps, including interactions with weights higher than or equal to this reliability estimation and the weighting of interactions. This coefficient and receiver operating characteristics, as outlined in the indicates how well the FS-Weight score of an interaction Methods section. Only informative GO terms see Meth- correlates to the likelihood of function being shared ods are used for validation. Only graphs for the biological proc- the likelihood of function sharing. The correlation is ess category are presented to maintain clarity. Graphs lower for molecular function in the M. No results are available for the tional File 1. The fraction of known functional annotations that can be sug- gested through BLAST homology search; and the additional annotations that can be suggested through: BLAST is performed on sequences from the gene ontology database. The top row shows the results from S. The three columns depict results on the biological process left , molecular function center and cellular component right categories of the Gene Ontology. However, note that due genomes. Again, only graphs for GO terms from biologi- to the lack of annotation information, the informative cal process are shown here. We can see that for most of the terms chosen for three genomes, including R. We again observe Methods. For the prediction of informative GO terms that the superiority of FS-Weighted Averaging over the two from the molecular function and cellular component cat- other methods is more significant in the genomes with egories, FS-Weighted Averaging also yields better recall denser interaction networks. Graphs for the two other GO and precision over the two other see Additional File 1. For these categories, no result is available for C. We also Function prediction with predicted protein-protein observe that the superiority of FS-Weighted Averaging interactions over the two other methods is more significant in the One of the main limitations in using protein-protein genomes with denser interaction networks i. If we can harness predicted interactions, such as those from the STRING database [23], the use of Receiver operating characteristics protein-protein interactions for functional inference To compare the receiver operating characteristics ROC becomes potentially more useful. The number of GO knowledge. Here we include the interactions for S. Statistics of interaction data from seven genomes Genome Interactions involving annotated Annotated Proteins Avg. Figure 5 shows the noisy interaction data â€” estimation of the reliability of precision-recall and ROC analysis of the predictions made experimental sources and topological weight. Here we by the three methods using 1 only interactions from want to study how well the method can perform against BioGRID 50, unique interactions ; and 2 a combi- noisy data. To simulate noise in interactions, we take the nation of BioGRID interactions and predicted interactions current interaction data and contaminate it with different from STRING , unique pairs for informative GO forms of synthetic noise. We perform this analysis on the terms from the biological process category. We find that S. To differenti- improvement in both precision-recall and ROC analysis ate the effects of different types of noise, we modify the with the combined interactions, while the performance of interaction data by 1 adding random interactions and 2 FS-Weighted Averaging remains relatively unchanged. Random additions This is due to the fact that the predicted interactions from reflect false positives in experimental

sources, while ran- STRING in fact already include many indirect interactions. The average number of annotated neighbors per anno- Real protein-protein interactions are likely to include tated protein is nearly 65 see Table 1 , which is much both forms of noise. One interesting point to note is that FS-Weighted tions applied. As a comparison, we repeated the predic- Averaging can already achieve outstanding recall and pre- tions using Neighbor Counting. Figure 6 presents graphs cision as well as ROC performance using the much that show the number of informative GO terms that can smaller BioGRID, which is less than one-third the size of be predicted above various ROC thresholds using the two the combined interactions! We find that the performances of both methods are less Robustness of FS-Weighted Averaging against noise and significantly affected by random additions than by ran- missing data dom deletions. Interestingly, we also find that the predic- FS-Weighted Averaging is designed to take into account tion performance of FS-Weighted Averaging actually the fact that interaction data can be rather noisy [25]. As improves with random additions, while the performance mentioned in the Methods section, the FS-Weight meas- of Neighbor Counting deteriorates with added noise. Identifying functions better predicted with indirect neighbors However, with random deletions, the performance of FS- We want to identify functions that can be better predicted Weighted Averaging deteriorates slightly faster than that with FS-Weighted Averaging. These observations indicate that FS- annotated by at least 30 proteins. Due to limited annota- Weighted Averaging is robust to false positives in the tion and interaction data, we study only 4 genomes: Top â€" Graphs showing the number of informa- Effect of noisy interaction data on FS-Weighted tive terms from the Gene Ontology biological process cate- Averaging. Bottom â€" Preci- Counting bottom on synthetically modified interaction data. Each point on the graph represents a Level-4 GO term. For all the four genomes in Figure 4, most points on the We select the top five terms from each GO category based graph lie above the diagonal line, indicating that FS- on their average FL2 in the genomes in which they appear Weighted Averaging performed better than Neighbor and present them in Table 3. Counting for most of these GO terms. By incorporating in at least two of the four genomes. For each term, we interaction reliability, topological weight, and indirect define a score that reflects the relative ROC score of FS- interactions, the method can predict more functions with Weighted Averaging against Neighbor Counting as fol- higher precision in all three categories of the Gene Ontol- lows: It is also reasonably resilient against interaction noise, maintaining consistent prediction performance even when a large number of interactions are randomly added to the interaction data. FL2 score Biological process Cellular biosynthesis 1. When a protein interacts with very few proteins, any form of measure that would assign a high reliability score or high confidence in sharing function without less significant in the genomes with sparse interaction net- additional evidence would be very susceptible to noise works and also less significant when the interaction net- and will not give consistent performance over different work is made sparser by random deletions. This is due to datasets. First, the number of indirect interactions is much lower for sparser networks due to limited connec- Examples of indirect functional association tivity. Indirect interactions can only involve proteins that Here we take a look at two examples that illustrate how interact with at least one other protein, i.

PREDICTING PROTEIN FUNCTIONS FROM PROTEIN INTERACTION NETWORKS HON NIAN CHUA, LIMSOON WONG pdf

5: Protein complex prediction based on k-connected subgraphs in protein interaction network

Predicting Protein Functions from Protein Interaction Networks pred iction te chn iques show ed t hat out of 34 5 genomes list ed in t he K EGG Genome collec tion (K aneh isa et al.

Sorry, we are unable to provide the full text but you may find it at the following location s: Suggested articles Citations Functional classification of proteins for the prediction of cellular function from a protein-protein interaction network. Conservation of gene order: Trends Biochem Sci Detecting protein function and protein-protein interactions from genome sequences. A basic local alignment search tool. Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. The Gene Ontology Consortium. Nat Genet GO molecular function terms are predictive of subcellular localization. BIND â€" the biomolecular interaction network database. Nucleic Acids Res How reliable are experimental protein-protein interaction data? Identification of functional links between genes using phylogenetic profiles. Predicting protein functions from redundancies in large-scale protein interaction networks. The use of gene clusters to infer functional coupling. Use of receiver operating characteristic analysis to evaluate sequence matching. On the number of protein-protein interactions in the yeast proteome. Operons in Escherichia coli: Protein interaction maps for complete genomes based on gene fusion events. Predicting protein function by genomic context: A network of interacting proteins in yeast. Prediction of protein function using protein-protein interaction data. Assessment of prediction accuracy of protein function from protein-protein interaction data. Assigning protein functions by comparative genome analysis: Global protein function prediction from protein-protein interaction networks. Transitive functional annotation by shortest-path analysis of gene expression data. Kernelbased data fusion and its application to protein function prediction in yeast.

6: Biological data mining in protein interaction networks (Book,) [www.enganchecubano.com]

The Protein Complex Prediction method (PCP) uses indirect interactions and topological weight to augment protein-protein interactions, as well as to remove interactions with weights below a threshold. PCP employs clique finding on the modified PPI network, retaining the benefits of clique-based approaches.

PREDICTING PROTEIN FUNCTIONS FROM PROTEIN INTERACTION NETWORKS HON NIAN CHUA, LIMSOON WONG pdf

Britains aircraft industry Post-Conflict Economies in Africa (International Economic Association) Portraits of vision The thirteenth round : hit em hard and hit em fast 1 Brachiopoda and Lamellibranchiata of the Raritan clays and Greensand marls of New Jersey, by R. P. Whit History of henry ford Sports great Jerome Bettis Microcomputers and marketing decisions W. Shakespeare Gent. Goethe and the novel Parents and the death of a child Sangeeta Singg Successful personal selling through TA Foundations of nuclear engineering Daniels story full book Goa travel guide map The Collegeville Bible time-line. Christs counsel to his languishing church of Sardis, or, The dying or decaying Christian British novelists and their styles Making better business decisions steve williams Official Control; Expatriate Dominance; Nigerian Sea of tranquility katja millay bud More seasonal cooking The Debate about Colour Naming in 19th Century German Philology Infrared Holography for Optical Communications (Topics in Applied Physics) Rewriting your personal mythology Imaging inflammation in Parkinsonian syndromes David J. Brooks Human resource development and management Feminizing slavery Complete Illustrated Thorburns Birds Of the Books of the New Testament. Using excel for business analysis danielle stein fairhurst Genetic engineering protects womens reproductive choices George Dvorsky Mechanical engineering for public cleansing Communities, Identities and Crime Wishful thinking : consent, contract and the obligation to die Creating a handbook for band Tax This! An Insiders Guide to Standing Up to the IRS (Self-Counsel Legal Series.) Informatics for Healthcare Professionals V. 4. Studies in landscape design The Stakeholding Society