Determining miRNA-disease associations using bipartite graph modelling

Joseph Nalluri, Bhanu Kamapantula, Preetam Ghosh
Virginia Commonwealth University
Richmond, Virginia
[nallurijj, kamapantulbk, pghosh] @vcu.edu

Debmalya Barh
Institute of Integrative Omics and Applied Biotechnology (IIOAB), WB, India
dr.barh@gmail.com

Neha Jain, Lucky Juneja, Neha Barve
Devi Ahilya University University
Indore, India
[mymailtoneha, ljpreet88, nehabarve2006] @gmail.com

ABSTRACT

Exploring miRNA-disease interactions is critical to identify the impact of a disease on other diseases. Mapping this problem to a graph theoretical concept offers a unique perspective to study unseen relationships among diseases. In our work, maximum weighted matching has been used after mapping the miRNA-disease associations as a bipartite graph. We also address the limitation of this approach using disease ranking scheme and the results are presented.

Categories and Subject Descriptors

E.1 [**Data Structures**]: *Graphs and networks*; G.2.2 [**Discrete Mathematics**]: Graph Theory—*Graph algorithms*

General Terms

Graph Theory, Optimization

Keywords

Bipartite graph, miRNA-disease interaction

1. INTRODUCTION

Complex networks offer a unique perspective to explore relationships among homogenous and heterogeneous entities. These entities can be biological molecules, diseases, genes etc. Single or multiple miRNA(s) up- or down- regulate one or a set of disease(s). The interactions of miRNAs and diseases can be mapped as a complex network such that miRNAs and diseases are nodes in the network. This mapping is critical to explore the associations and depends heavily on the type of interactions. A graph theory concept such as bipartite graph [1] can be used to model this problem. This work is a proof-of-concept to determine top affected diseases by miRNA(s) using bipartite graph analysis.

BCB '13, September 22 - 25, 2013, Washington, DC, USA
ACM 978-1-4503-2434-2/13/09

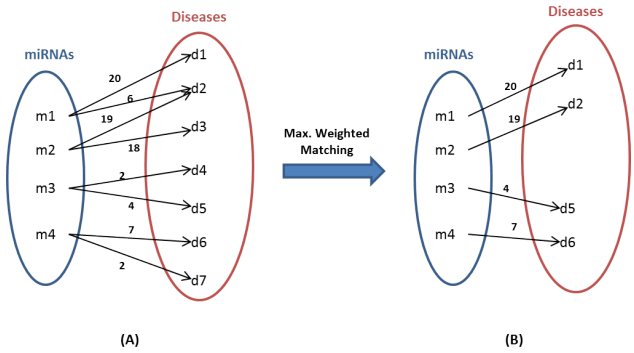

Figure 1: miRNA-disease association

1.1 Database

Exploration and analysis of regulatory networks of miRNA is crucial in understanding the precise mechanisms through which a miRNA exerts its function so that the network can be engineered for various biological aspects. An in-house database hosting all the information related to miRNA such as its regulated pathways, upstream regulators, downstream targets, functions and diseases has been developed to explore the associations among miRNAs and diseases.

1.2 miRNA-Disease Interaction

miRNA(s) up-regulate or down-regulate disease(s). Researchers will be interested to know how a particular miRNA influences a disease. Our work filters the top affected diseases for a given set of miRNAs. This is determined by the number of interactions for any miRNA-disease pair. Using this strength of association, we predict which diseases would be highly influenced upon the activation of certain miRNAs. For this, we have modelled the miRNA-disease interaction as a bipartite graph which is shown in Figure 1. The bipartite graph analysis is explained in Section 2.1.

2. METHODOLOGY

2.1 Bipartite analysis

A bipartite graph is a graph $G(V, E)$ in which the set of vertices V can be partitioned into two disjoint sets V_1 and V_2 such that every edge connects a vertex in V_1 to the one

in V_2 [1]. In our model, miRNAs and diseases have been categorized as two disjoint sets and an edge represents an association between them. The edges are weighted, wherein the weight represents the strength of a pair including up- and down- regulations.

2.1.1 *Maximum weighted matching (MWM)*

In the graph $G(V, E)$, if there is a set of edges such that no two edges share a common end vertex, it is known as a matching. Maximum matching is a matching with largest possible set of edges. A maximum weighted matching is a maximum matching such that the sum of the weights of the edges is maximum. This is explained below.

Consider a miRNA-disease interaction network as in Figure 1, where $m1$ to $m4$ are miRNAs and $d1$ to $d7$ are diseases. The weight on the edge represents the strength of the association of the miRNA-disease pair, in terms of up-regulating and down-regulating a disease. As part (B) shows after the application of the MWM algorithm, the resultant sum of edges is the maximum. Hence, the MWM helps in determining the strongest miRNA-disease pairs among a set of activated miRNAs. The results give the cumulative impact of a set of activated miRNAs on the set of associated diseases, which would be most certainly impacted. The solution to the MWM algorithm in a given G(V,E) can be solved as an optimization problem as described below [2]:

Maximize

$$\sum_{i,j} Weight_{i,j} X_{i,j}$$

subject to $$\sum_{j} X_{i,j} \leq 1 (i = 1, 2, ...m) \text{ and}$$

$$\sum_{i} X_{i,j} \leq 1 (i = 1, 2, ...d)$$

2.2 Limitations and Enhancement

Although, the application of MWM algorithm gives the most prominent miRNA-disease pairs, it has a limitation. Because of the constraint that no two edges can share a common vertex, a strongly associated miRNA-disease pair can get ignored. For example, in Figure 1, for miRNA $m2$, $m2$-$d2$ pair weight is 19 and $m2$-$d3$ pair weight is 18. However, in the resultant matching only $m2$-$d2$ pair is selected (see Figure 2-A). The pairs $m3$-$d4$ and $m4$-$d5$ are selected in the matching but their pair weights are 4 and 7 respectively, which is less than the non-selected pair, $m2$-$d3$. In order to overcome this limitation, a disease ranking scheme has been adopted. Here, miRNA-disease pairs are ranked as per their highest weight (see Figure 2). This helps in obtaining the highest weighted pairs. After the MWM algorithm is applied and the set of diseases are obtained, the rank for the disease which is the least in the set is checked. If there are other diseases which have a higher rank than the selected disease and they are not in the resultant matching set, those diseases are included in the resultant output set of diseases (see Figure 2-C). This method makes sure that a disease which is highly influenced is not missing after the MWM algorithm is applied. MWM algorithm helps in giving a definite and concise set of affected diseases. Disease ranking scheme enhances the result set by overcoming the limitation of the MWM algorithm for this scenario.

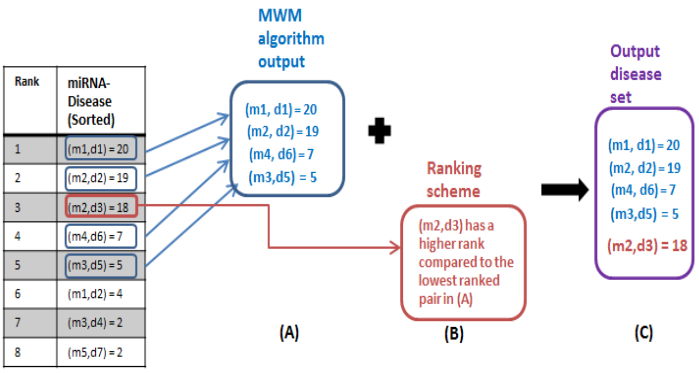

Figure 2: MWM algorithm and disease ranking scheme

Table 1: MWM Algorithm Results

S.No.	PubMed for miRNAs	miRNAs	Diseases
1	23617747	hsa-mir-9-1, hsa-mir-9-2, hsa-mir-200c	Breast cancer, Colorectal cancer, Kidney cancer, Ovarian cancer
2	23272653	hsa-mir-182, hsa-mir-200a, hsa-mir-200b, hsa-mir-200c	Lung cancer, Ovarian cancer (OC), Hepatocellular carcinoma (HCC), Breast cancer, Kidney cancer, Colorectal cancer, Oral squamous cell carcinoma

3. TESTING

GLPK package [3] has been used in Ubuntu operating system to solve the optimization problem of MWM algorithm in the considered bipartite graph. The results of which are represented in Table 1 after the applying disease ranking scheme as explained in Section 2.2 and Figure 2.

4. CONCLUSION AND FUTURE WORK

An online interface including an easy-to-use visualization tool to access these miRNA-disease associations will be developed. This concept will be tested for different data sets such as *miREnvironment Database* [4].

5. REFERENCES

[1] A. S. Asratian. *Bipartite graphs and their applications*. Number 131. Cambridge University Press, 1998.

[2] S. C. Fang. ISE/OR/MA 766 network flows - lecture notes. 2012.

[3] A. Makhorin. GNU linear programming kit. *Moscow Aviation Institute, Moscow, Russia*, 38, 2001.

[4] Q. Yang, C. Qiu, J. Yang, Q. Wu, and Q. Cui. mirenvironment database: providing a bridge for micrornas, environmental factors and phenotypes. *Bioinformatics*, 27(23):3329–3330, 2011.