BITS :: Call for Abstracts 2019 - Oral communication

Type	Oral communication
Session	Computational RNA biology
Title	Tripartite graph clustering for the prediction of lncRNA-disease associations
All Authors	Aurora Paternostro (1), Simona E. Rombo (1)
Affiliation	

(1) Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, Palermo

Motivation

The discovery of novel associations between long non-coding RNAs (lncRNAs) and diseases may provide valuable input to the understanding of disease mechanisms at lncRNA level, as well as to the detection of disease biomarkers for disease diagnosis, treatment, prognosis and prevention [1]. Unfortunately, due to costs and time complexity, the number of possible disease-related lncRNAs that can be verified by traditional biological experiments is very limited. Computational approaches for the prediction of potential disease-lncRNA associations can effectively decrease the time and cost of biological experiments, allowing for the identification of the most promising lncRNA-disease pairs to be further verified in laboratory. Such predictive approaches often rely on the analysis of lncRNAs related information stored in public databases, e.g., their interaction with other types of molecules [2, 3, 4]. As an example, large amounts of lncRNA-miRNA interactions have been collected in public databases, and plenty of experimentally confirmed miRNA-disease associations are available as well.

Methods

We propose a novel computational approach for the prediction of lncRNA-disease associations (LDAs) based on known lncRNA-miRNA interactions (LMIs) and miRNA-disease associations (MDAs). In particular, taking inspiration from previous studies on social media [5], we model the problem of LDAs prediction as the clustering of tripartite graphs in which the three sets of vertices represent lncRNAs, miRNAs and diseases, respectively, and vertices are linked according to LMIs and MDAs, as shown in Figure 1(a). Based on the assumption that similar lncRNAs interact with similar diseases [4], we search for “communities” in the so obtained tripartite graphs (see Figure 1(a)) by applying edge clustering as follows. Two hyperedges of the tripartite graph are adjacent if they share at least a common vertex. The goal is to detect communities made of vertices such that the number of adjacent hyperedges involving them is maximized inside the same community and minimized between different communities. An adjacent edge-graph is generated that represents the adjacency of hyperedges in the tripartite graph. Then, hyperedges are represented as vertices, and their adjacencies are represented as edges in the so obtained adjacent edge-graph. Unipartite graph clustering is applied to the edge-graph, and communities are generated by taking into the same partition vertices involved in the same edge-graph cluster. A score is finally assigned to each predicted LDA by considering both its contribution to the community and the internal structure of the community (i.e., its compactness) in its turn.

Results

We have validated the proposed approach on both synthetic and real datasets. In more detail, experimental verified data have been downloaded from starBase [6] for the LMIs and from HMDD [7] for the MDAs. A golden-standard dataset of LDAs has been obtained from the LncRNADisease database [8]. Before proceeding with our discussion, some considerations are needed.
Among the approaches for LDAs prediction presented in the Literature, including machine-learning-based models, only a few of them do not use directly known lncRNA-diseases relationships during the prediction task. However, so far, the experimentally identified known lncRNA-disease associations are still very limited, therefore using them during prediction could bias the final result, as confirmed by the low performance of such approaches on de novo LDAs prediction [4]. This enforces the idea behind our clustering approach, which does not require any positive examples in advance.
The experimental tests show that our approach is able to detect specific situations not captured by its competitors which do not use LDAs during prediction (e.g., [2]). An intuitive example is shown in Figure 1(b-c): the association (L1, D3) would not be captured based on the MDAs “support” (e.g., the p-value [2]), whereas our approach is able to detect it.

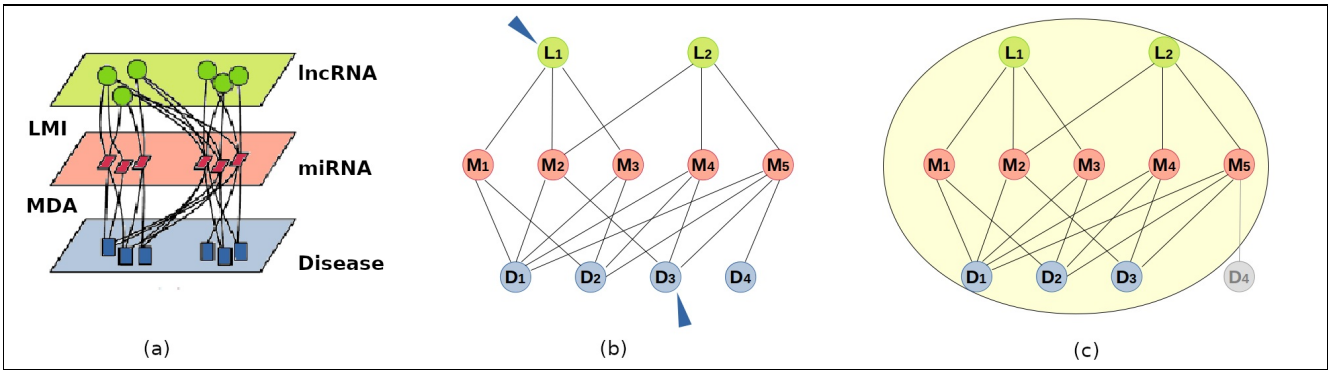
Info

References

[1] Liao, Q. et al. Large-scale prediction of long non-coding RNA functions in a coding–non-coding gene co-expression network. Nucleic Acids Res. 2011, 39, 3864-3878.

[2] Xing Chen. Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA. Scientific Reports, 5:13186 (2015).
[3] Xuan Z., et al. A Probabilistic Matrix Factorization Method for Identifying lncRNA-disease Associations. Genes, 10(2), 2019.
[4] Lu C., et al. Prediction of lncRNA-disease associations based on inductive matrix completion. Bioinformatics, 34(19):3357-3364, 2018.
[5] Ikematsu, K., Murata, T. A Fast Method for Detecting Communities from Tripartite Networks. In Proc. of SocInfo 2013: 192-205.
[6] Li, J.-H., Liu, S., Zhou, H., Qu, L.-H. & Yang, J.-H. starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res 42, D92-D97 (2013).
[7] Li, Y., et al. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. Nucleic Acids Res 42, D1070-D1074 (2014).
[8] Chen, G. et al. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res 41, D983–D986 (2013).

Figure



Availability	http://math.unipa.it/rombo/lncRNA/abstract_rombo.pdf

Corresponding Author

Name, Surname	Simona E., Rombo
Email	simona.rombo@unipa.it
Submitted on	06.04.2019