



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO



Dipartimento di
INFORMATICA

Improving the Prediction of miRNA:mRNA Interactions by Exploiting Co-Clustering Methods

Gianvito Pio¹, Michelangelo Ceci¹, Domenica D'Elia² and Donato Malerba¹

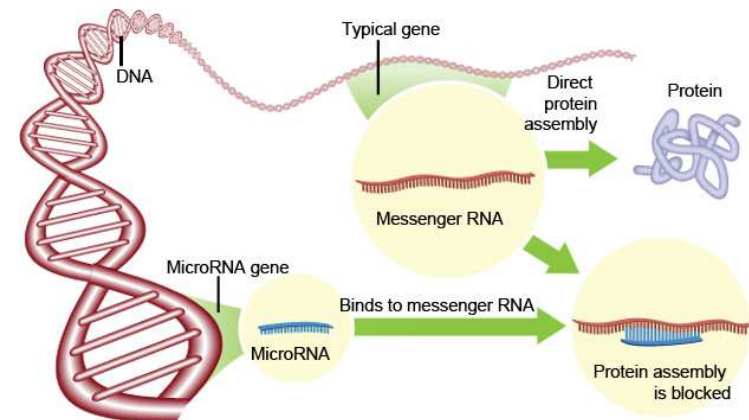
¹University of Bari – Department of Computer Science - Bari, Italy

² CNR-Institute for Biomedical Technologies - Bari, Italy

The considered task

Prediction of **interactions** between microRNAs (miRNAs) and messenger RNAs (mRNAs) for the discovery of putative **miRNA:mRNA interaction networks**.

- a comprehensive analysis of **cooperative targeting of miRNAs** of interest
- the discovery of **unknown miRNA and mRNA functions**
- the discovery of **unknown miRNA targets** which could be worth to be experimentally validated





Network reconstruction

A very small set of validated interactions is available since they derive from **expensive in-vitro experiments**.

Several approaches can be adopted to perform **link prediction**, but they often fail in simultaneously considering all the possible criteria:

- **network topology**
- **nodes properties**
- **autocorrelation among nodes**
- **etc.**



Network reconstruction

- Link Prediction

Learning to combine the output of several prediction approaches to exploit their different peculiarities

G. Pio, M. Ceci, D. D'Elia, D. Malerba, *Integrating microRNA target predictions for the discovery of gene regulatory networks: a semi-supervised ensemble learning approach*, BMC Bioinformatics 15 (Suppl 1), S4, 2014

- Identification of interaction sub-networks

Coclustering for the identification of highly connected subgraphs from known and predicted links

G. Pio, M. Ceci, D. D'Elia, C. Loglisci, D. Malerba, *A Novel Biclustering Algorithm for the Discovery of Meaningful Biological Correlations between microRNAs and their Target Genes*, BMC Bioinformatics 14 (Suppl 7), S8, 2013



Motivations of the work

- Existing methods analyze single interaction **independently each other**.
- **Assumption**: the existence of an interaction can strongly depend on **possible inter-dependencies** among groups of miRNAs and mRNAs



- How can we identify such inter-dependencies?
- How can we exploit them in order to improve the accuracy of predictions?



Motivations of the work



Can we exploit co-clustering methods?

They are able to **identify possible inter-dependencies**, since miRNAs and mRNAs involved in the same sub-networks are somehow (functionally and/or structurally) related/dependent.

We adopt our co-clustering method HOCCLUS2*

*G. Pio, M. Ceci, D. D'Elia, C. Loglisci, D. Malerba, *A Novel Biclustering Algorithm for the Discovery of Meaningful Biological Correlations between microRNAs and their Target Genes*, BMC Bioinformatics 14 (Suppl 7), S8, 2013



Exploiting the identified co-clusters

We **add** a miRNA:mRNA interaction to the dataset:

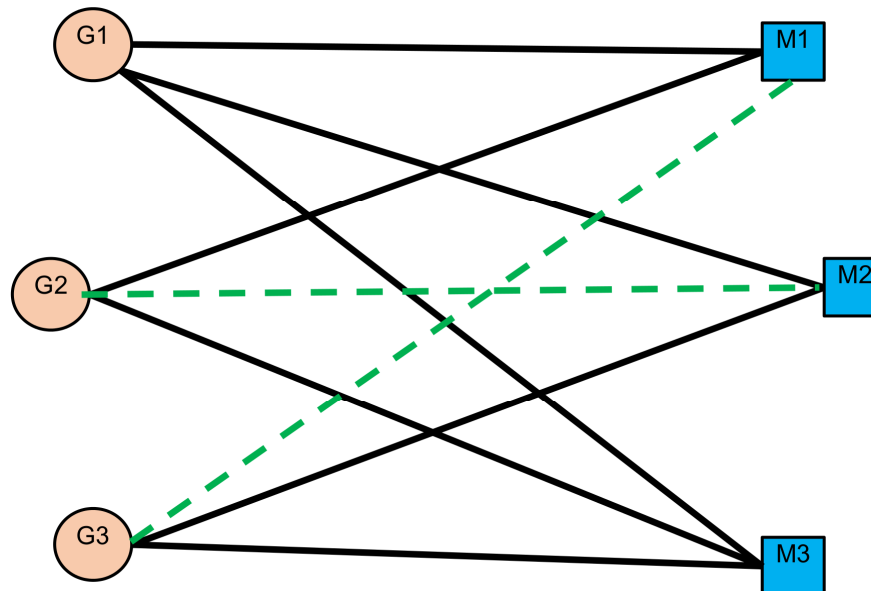
- if it is not already present in the set of interactions and **falls in the same interaction network(s)**
 - ➔ it might be functionally related with other elements in the network because of unknown cooperative activities of the miRNA with other miRNAs and/or because of still unknown functions of the genes in that functional network
- with a score corresponding to the **co-cluster cohesiveness**, which is the average score of all the interactions in the co-cluster
 - ➔ the stronger the known interactions within the co-cluster, the higher the possibility of the existence of the new discovered interaction
(N.B. if it appears in more than one co-cluster, we take the **maximum score**)



Exploiting the identified co-clusters

Input Interactions:

- G1 – M1
- G1 – M2
- G1 – M3
- G2 – M1
- G2 – M3
- G3 – M2
- G3 – M3



Refined Interactions:

- G1 – M1
- G1 – M2
- G1 – M3
- G2 – M1
- G2 – M3
- G3 – M2
- G3 – M3
- G2 – M2
- G3 – M1



Exploiting the identified co-clusters

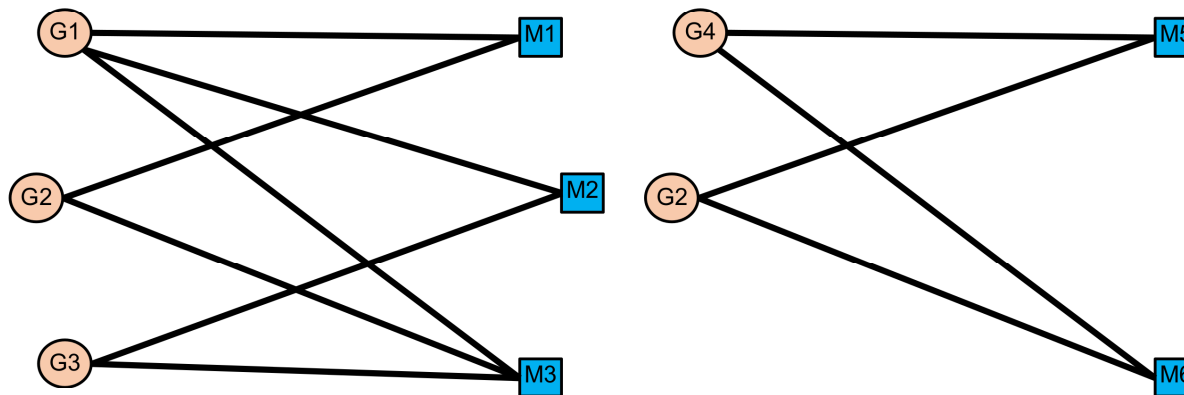
We **remove** a miRNA:mRNA interaction from the dataset:

- if it is already present in the set of interactions and **does not fall in any interaction network(s)**
 - ➔ their interaction is not validated by their possible cooperation and/or similarity in terms of functionality and/or structure.



Exploiting the identified co-clusters

- Input Interactions:**
- G1 – M1
 - G1 – M2
 - G1 – M3
 - G2 – M1
 - G2 – M3
 - G3 – M2
 - G3 – M3
 - G4 – M3
 - G4 – M5
 - G4 – M6
 - G5 – M5
 - G5 – M6



- Refined Interactions:**
- G1 – M1
 - G1 – M2
 - G1 – M3
 - G2 – M1
 - G2 – M3
 - G3 – M2
 - G3 – M3
 - ~~G4 – M3~~
 - G4 – M5
 - G4 – M6
 - G5 – M5
 - G5 – M6



Preliminar Experiments

Initial set of predictions:

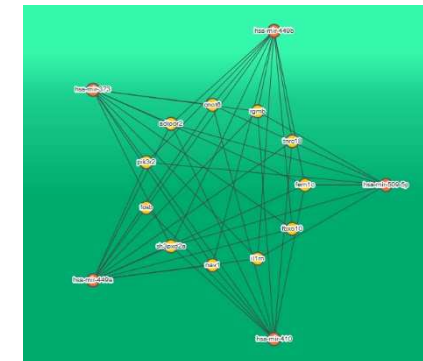
- ~5 millions interactions predicted by our semisupervised method and publicly available at ComiRNet (<http://comirnet.di.uniba.it/>)

HOCCLUS2 paramaters:

- $\alpha \in \{0.2, 0.3\}$, $\beta \in \{0.3, 0.4, 0.5\}$

Settings:

1. Original dataset
2. Original dataset + **new links**, on the basis of HOCCLUS2
3. Original dataset – **removed links**, on the basis of HOCCLUS2
4. Original + **new links**, – **removed links**, on the basis of HOCCLUS2



*G. Pio, M. Ceci, D. Malerba, D. D'Elia, *ComiRNet: a Web-based System for the Analysis of miRNA-gene Regulatory Networks*, BMC Bioinformatics 16 (Suppl 9), S7, 2015



Preliminar Experiments

Evaluation measures:

- **AUC** on the testing set TarBase 6.0 (65,000 validated interactions)
- **Average cohesiveness** μ_q and **biological significance** of the interaction networks

The biological significance is evaluated by means of the **T-Test** which:

- compares **intra-cluster similarity** and **inter-cluster similarity**
- exploits two hierarchies in Gene Ontology (Biological Process and Molecular Function) and outputs **two p-values, p_{BP} and p_{MF}**
- is based on the gene similarity function **simGIC**



Preliminar Experiments

Original dataset AUC: **0.649**

α	β	Setting	AUC
0.2	0.3	+/-	0.502
		+	0.650
		-	0.500
	0.4	+/-	0.503
		+	0.650
		-	0.500
	0.5	+/-	0.502
		+	0.649
		-	0.500
0.3	0.3	+/-	0.502
		+	0.649
		-	0.500
	0.4	+/-	0.501
		+	0.649
		-	0.500
	0.5	+/-	0.500
		+	0.649
		-	0.500



Preliminar Experiments

Original Dataset

α	β	Level 1				Max level					Best level				
		#cc	pMF	pBP	μq	level	#cc	pMF	pBP	μq	level	#cc	pMF	pBP	μq
0.2	0.3	888	1.000	1.000	0.690	7	143	0.000	0.000	0.240	2	444	0.000	0.000	0.520
	0.4	591	0.404	1.000	0.770	8	25	0.175	0.080	0.120	3	148	0.000	0.000	0.390
	0.5	417	0.361	0.244	0.830	7	53	0.000	0.000	0.240	3	105	0.000	0.000	0.430
0.3	0.3	888	1.000	1.000	0.690	6	268	0.002	0.001	0.370	3	309	0.000	0.000	0.420
	0.4	591	0.404	1.000	0.770	6	161	0.001	0.377	0.380	2	298	0.000	0.001	0.570
	0.5	417	0.361	0.244	0.830	7	104	0.000	0.000	0.390	4	110	0.000	0.000	0.420

Original dataset + New Links

α	β	Level 1				Max level					Best level				
		#cc	pMF	pBP	μq	level	#cc	pMF	pBP	μq	level	#cc	pMF	pBP	μq
0.2	0.3	651	0.002	1.000	0.662	7	77	0.000	0.000	0.238	2	326	0.000	0.000	0.548
	0.4	498	0.043	0.051	0.768	7	46	0.019	0.015	0.245	3	126	0.001	0.000	0.481
	0.5	311	0.019	0.006	0.830	6	27	0.001	0.000	0.242	2	156	0.000	0.000	0.674
0.3	0.3	401	0.000	0.000	0.577	6	146	0.000	0.001	0.362	1	401	0.000	0.000	0.577
	0.4	440	0.001	0.024	0.734	6	97	0.000	0.000	0.370	2	220	0.000	0.000	0.616
	0.5	310	0.011	0.009	0.829	6	55	0.000	0.000	0.380	4	62	0.000	0.000	0.420



Conclusions

The idea of exploiting co-clustering for improving the accuracy of miRNA:mRNA prediction appears promising. Preliminary experiments show that:

- **The addition** of interactions according to the discovered sub-networks can **improve the accuracy** and can lead to **more coherent interaction networks**
- **The deletion** of interactions when they do not appear in any interaction network is too much restrictive and can lead to a **degeneration in the results.**



Conclusions

Future works:

- To consider external resources in the co-clustering algorithm in order to avoid strong decisions in the addition and in the deletion
- To consider smarter approaches for the combination of the scores when an interaction appears in more than one co-cluster.



Questions?

Availability:

- Link prediction system: www.di.uniba.it/~ceci/micFiles/systems/semisupervised_HOCCLUS2/
- HOCCLUS2: www.di.uniba.it/~ceci/micFiles/systems/HOCCLUS/
- Biological query system ComiRNet: comirnet.di.uniba.it