In Silico Study of Exosomal miR-302b treatment in Mice Liver Cells

Emiliano Marrale
e.marrale@studenti.unipi.it
Student ID: 545552

Abstract

This study explores the in silico effects of miR-302b treatment on mouse liver cells [3]. Starting from the differentially expressed genes identified by the original researchers, the raw RNA-seq data was re-analyzed using the DESeq2 algorithm [10] to obtain gene-level fold changes. A significance threshold was then applied to select a subset of highly differentially expressed genes, which served as "seed" nodes for constructing a protein–protein interaction (PPI) network via the STRING database API [14]. The extended network, including first- and second-order interactors, comprised 2070 additional genes not present in the initial dataset.

The analysis was divided into three main sections. The first involved characterizing the topological properties of the network—such as degree distribution, betweenness centrality, and diameter—and comparing them with Erdős–Rényi [7] and Barabási–Albert [2] models. The second section consisted in performing community detection and evaluating the resulting modules through gene set enrichment analysis (GSEA) [13]. Finally, the potential of fold-changes network propagation on a signed, directed graph was investigated to infer possible new target genes that might be influenced by the treatment. Collectively, these analyses provide an integrated systems-level perspective on the molecular impact of miR-302b treatment in hepatic cells.
[1]

Keywords

Network Analysis, Network Biology, Functional Enrichment, miR-302b, Community Detection, Network Diffusion.

ACM Reference Format:
Emiliano Marrale. 2025. In Silico Study of Exosomal miR-302b treatment in Mice Liver Cells. In . ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

MicroRNAs (miRNAs) are small non-coding RNAs that play a crucial role in the post-transcriptional regulation of gene expression, influencing a wide range of biological processes including development, metabolism, and cellular senescence. Among them, miR-302b has emerged as a key regulator with potential therapeutic applications due to its ability to modulate cell proliferation and rejuvenate senescent cells [3]. While previous studies have demonstrated the phenotypic effects of miR-302b in vivo, the underlying molecular mechanisms and the systemic impact on cellular networks remain mostly unexplored.

High-throughput transcriptomic profiling, such as RNA sequencing (RNA-seq), provides a powerful tool for investigating the global changes induced by miRNA treatments. Analysis of differentially expressed genes can reveal target pathways and molecular interactions that mediate the observed phenotypic effects. However, interpreting these changes in the context of complex cellular networks requires an integrative systems-level approach, combining network analysis, community detection, and functional enrichment methods.

In this study, an in silico investigation was conducted to elucidate the systemic impact of miR-302b treatment on mouse liver cells. Starting from the differentially expressed genes identified, a protein–protein interaction (PPI) network was constructed using the STRING database [14], incorporating both first- and second-order interactors. The network was subsequently analyzed to characterize its topological properties, identify functional modules through community detection, and assess the propagation of gene expression changes across the network. Gene set enrichment analysis (GSEA) [13] was employed to evaluate the biological significance of the identified modules.

This integrative approach provides a comprehensive perspective on the molecular mechanisms influenced by miR-302b, offering insights into the potential pathways through which it exerts its therapeutic effects.

2 Data Collection

As introduced in the abstract, the initial dataset was provided by the researchers and is publicly available at the following link: https://data.mendeley.com/datasets/899tntr9x2/1. The dataset consists of bulk single cell RNA-seq profiles from liver, lung, kidney, and skin tissues collected from young, aging, and aging-302b–treated mice.

For simplicity, only the liver tissue data was considered. Differential expression analysis was performed using the DESeq2 algorithm [10], applying a threshold of adjusted p-value ≤ 0.05 and an absolute $\log_2$ fold change greater than 0.585 (which corresponds to a 1.5-fold change). This resulted in the identification of 835 downregulated genes and 518 upregulated genes.

These initial sets of differentially expressed genes (DEGs) were subsequently expanded using the STRING database API [14]. Mus Musculus was used as reference organism for querying STRING. First- and second-level neighbors of the seed genes were retrieved,

[1] **Project Repositories**
Researcher's Data: https://data.mendeley.com/datasets/899tntr9x2/1
Github Repository: https://github.com/EmilianoMarrale/SocialNetworkAnalysis

Conference'17, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

and the edges among all included nodes were added to construct two extended networks. This expansion yielded a downregulated network with 6568 nodes and an upregulated one with 5871 nodes. Notably, the intersection between the two networks was non-empty, comprising 3934 shared nodes, indicating that many downstream neighbors interact with both up- and downregulated genes.

For the network propagation task, the SIGNOR API [9] was used to map directionality and signed regulatory relationships among the nodes, thereby enabling the construction of a signed, directed interaction network suitable for random-walk–with–restart (RWR) propagation.

2.1 Selected Data Sources

Three main data sources were utilized in this analysis. The first is the reference RNA-seq dataset described above. The remaining two are the STRING and SIGNOR databases, which were used to expand and characterize the gene interaction networks.

2.1.1 STRING. (Search Tool for the Retrieval of Interacting Genes) is a comprehensive database of known and predicted protein–protein interactions [14]. It integrates information from multiple sources, including curated databases, experimental results, computational predictions, and text mining. Each interaction is associated with a confidence score, and interactions may represent direct physical binding or more general functional associations. In this study, STRING was used to expand the DEG lists by retrieving first- and second-level interaction partners, with a confidence score above 0.7, and constructing the corresponding undirected gene interaction networks.

2.1.2 SIGNOR. (SIGnaling Network Open Resource) is a manually curated database of causal relationships among biological entities, primarily focusing on signaling pathways [9]. Unlike STRING, SIGNOR provides directed and signed edges, representing activation, inhibition, or more complex regulatory effects. This causal structure enables analyses that require directionality and polarity of interactions. In this work, SIGNOR was employed to assign direction and sign to edges in the expanded gene networks to improve the interpretability and performance of the RWR propagation algorithm.

2.2 Crawling Methodology and Assumptions

For the STRING-based network expansion, the API was queried for each seed gene. Each query returned the first-level interaction partners, which were added as new nodes to the graph when not already present and connected to the corresponding seed node. Subsequently, for each first-level partner, an additional STRING API request was performed to retrieve its second-level interaction partners. These were incorporated into the graph along with the relevant edges. This two-step expansion resulted in a combined up- and downregulated gene interaction network containing both first- and second-level neighbors of the original DEGs.

For the SIGNOR-based induction of directed and signed interactions, a new directed graph was constructed. Each node in the STRING-expanded network was queried against the SIGNOR API to retrieve potential interaction partners and their corresponding causal relationships. Only interactions with a confidence score greater than 0.5 and with a clearly defined sign (activation or inhibition) were considered. Importantly, no new nodes were added at this stage: only edges among the pre-existing nodes were incorporated, annotated with their directionality and sign. After processing all nodes, isolated nodes (i.e., those lacking any SIGNOR-derived interactions) were removed, resulting in a final directed network comprising 2521 nodes and 6274 edges. The reduced size of this network reflects the manually curated nature of SIGNOR, meaning that many biological interactions have not yet been experimentally characterized or curated yet.

3 Network Characterization

The full extended combined network, as well as the extended upregulated and downregulated networks, were analyzed and compared to their corresponding ER and BA reference networks. For simplicity, the figures showing degree distributions are presented only for the full combined network, while the complete set of comparisons for all network types can be found in Table 1,2.

For each network, the following metrics are reported:

- Table 1: number of nodes and edges, maximum and minimum node's degree, average degree and density.
- Table 2: global clustering coefficient, average betweenness and closeness centralities, the number of connected components (CC), the size of the largest connected component (LCC), the diameter of the LCC and lastly the average shortest path length of the LCC.

Network	Nodes	Edges	Max Deg.	Min Deg.	Avg Deg.	Density
Downreg	6858	21862	69	0	6.38	0.00093
Upreg	5871	17408	55	0	5.93	0.00101
Combined	8795	32313	69	0	7.35	0.00084
ER Downreg	6858	21856	17	0	6.37	0.00093
ER Upreg	5871	17005	17	0	5.79	0.00099
ER Combined	8795	32610	21	0	7.42	0.00084
BA Downreg	6858	20565	188	3	6.00	0.00088
BA Upreg	5871	11738	164	2	4.00	0.00068
BA Combined	8795	26376	274	3	6.00	0.00068

Table 1: Summary statistics for real and synthetic networks.

Network	Clust.	Betwn.	Closn.	Num CC	LCC Size	LCC Diam.	Avg SPL
Downreg	0.25	0.00069	0.157	182	6618	17	6.05
Upreg	0.26	0.00087	0.148	129	5686	14	6.46
Combined	0.28	0.00051	0.157	303	8408	16	5.91
ER Downreg	0.00	0.00058	0.201	14	6845	10	4.98
ER Upreg	0.00	0.00070	0.194	26	5846	11	5.14
ER Combined	0.00	0.00043	0.210	8	8788	9	4.76
BA Downreg	0.01	0.00046	0.241	1	6858	7	4.17
BA Upreg	0.01	0.00066	0.208	1	5871	8	4.86
BA Combined	0.01	0.00037	0.238	1	8795	7	4.23

Table 2: Summary statistics for real and synthetic networks.

Figure 1 displays the degree distribution of the node degrees of the full combined network.

3.1 Comparison with ER

The degree distribution of the combined network was compared against a classical Erdős–Rényi (ER) [7] random graph with the same number of nodes and approximately the same edge density. The NetworkX implementation requires as input the number of

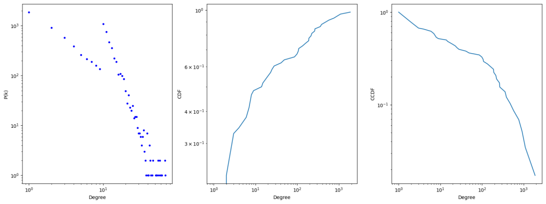

Figure 1: Combined Network Degree Distribution.

nodes and the connection probability p. To match the experimental network, the probability p was computed as

$$p = \frac{2\, n_{\text{edges}}}{n_{\text{nodes}}\,(n_{\text{nodes}} - 1)}.$$

Where, n_{edges} and n_{nodes} denote, respectively, the number of edges and the number of nodes in the experimental network. As expected, the ER model produced a binomial-like degree distribution characterized by a single narrow peak around the average degree. In contrast, the empirical network displayed a highly skewed and heavy-tailed distribution, indicating the presence of hubs and substantial heterogeneity in connectivity. From tables 1 and 2 it is also possible to notice how the number of connected components and the average global clustering coefficient are much higher in the empirical networks.

This discrepancy highlights that the biological interaction network does not behave like a homogeneous random graph: interactions are not uniformly distributed across nodes, and a small subset of nodes possesses disproportionately high connectivity. These structural differences suggest that random connectivity alone cannot explain the topological patterns observed in the integrated biological network.

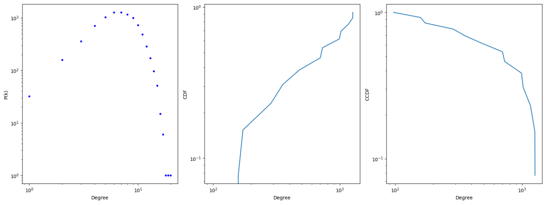

Figure 2: ER Network Degree Distribution.

3.2 Comparison with BA

The combined network was also compared with a Barabási–Albert (BA) [2] scale-free model, which generates networks through preferential attachment and typically yields a power-law degree distribution. The NetworkX implementation requires as input the number of nodes and the the number of edges per node. To match the experimental network, the number of edges per node was computed as the average degree of the experimental network divided by two.

The BA model reproduced several key features seen in the empirical network, particularly the presence of high-degree hubs and a long-tailed degree distribution. However, the empirical data exhibited deviations from a pure power-law behaviour, including a more heterogeneous distribution of medium-degree nodes and a steeper drop-off among the highest-degree nodes.

These differences indicate that while preferential attachment may partially explain the emergence of hub nodes, additional biological and structural constraints—such as pathway specificity, functional modularity, and curated interaction directionality—likely contribute to the network's more complex topology. Therefore, the combined network can be considered "scale-free-like," but it does not fully conform to the idealized BA model.

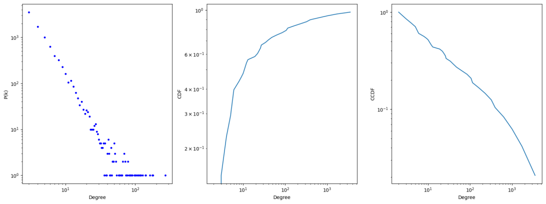

Figure 3: BA Network Degree Distribution.

4 Community Discovery

Community discovery approaches were applied to the largest connected component only, going from 8795 nodes and 32313 edges to 8408 nodes 32178 edges. Three main algorithms were employed:

- **Louvain (Internal Density).** The Louvain [4] method is a fast, hierarchical approach to community detection that optimizes modularity. It iteratively aggregates nodes into communities to maximize internal edge density while minimizing connections between different communities. The algorithm proceeds in phases of local optimization followed by community aggregation. Its efficiency and scalability have made it one of the most widely used community detection methods for large networks.
- **Angel (Percolation).** Angel [6] is a node-centric, bottom–up community detection algorithm based on local-first percolation principles. The underlying assumption is that each node can locally infer its community through exploration of ego-networks, while global community structure may contain complex overlaps. Angel is designed as a faster and more scalable successor to the DEMON algorithm, improving performance while preserving high-quality overlapping community detection.
- **Infomap (Entity Closeness).** Infomap [12] is an information-theoretic community detection method that uses random walks to reveal structural modules. It models information flow on the network and partitions nodes by minimizing the expected description length of a random walker's trajectory. Communities correspond to regions of low conductance, where random walkers tend to remain for long periods before exiting through relatively few inter-community edges. As with modularity-based methods such as Louvain, Infomap effectively identifies well-separated communities, but it is particularly strong in capturing flow-driven structures.

4.1 Enrichment Strategy

For each partition produced by the community detection algorithms, gene set enrichment analysis [13] was performed, selecting KEGG [8] pathways and Gene Ontology [1] (GO) molecular function terms. This was made possible by the endpoint provided by the STRING database API. For simplicity, only one enrichment result was selected from the API response, which was evaluated by taking into account the parameters described below:

- **False Discovery Rate (FDR)**: Lower values indicate higher statistical significance and a lower likelihood that the observed overlap occurred by chance. This serves as an adjusted p-value.
- **Minimum background size of at least two genes**: Categories with very small background sets were excluded to avoid unstable or misleading enrichment ratios.
- **Enrichment Ratio** ($\frac{n_{\text{genes}}}{n_{\text{background}}}$): This quantifies the strength of the enrichment for a given category, where n_{genes} is the number of genes in the partition belonging to that pathway or function, and $n_{\text{background}}$ is the total number of genes associated with that category in the background.

For each partition, the most informative enrichment was identified by selecting the categories with the lowest FDR and the highest enrichment ratio within the two selected annotation domains (KEGG pathways and GO molecular functions).

4.2 Algorithm Evaluation Criteria

To identify the best-performing community detection algorithm, a scoring function was designed in which each evaluation metric is normalized and weighted according to its relevance. The final score for each algorithm is obtained by combining these weighted, normalized metrics.

The evaluation criteria are defined as follows:

- **Number of clusters (n_{clusters})**: Should be balanced. Extremely high or low values suggest suboptimal clustering quality.
- **Average cluster size**: Higher average cluster sizes are preferred, as they indicate more coherent and meaningful communities.
- **Internal edge density, average internal degree, and modularity**: For these metrics, robustness is evaluated using a *best–worst* approach. Specifically, the *maximum of the minimum* values across the three partitions provided by the algorithms was considered, favoring algorithms that maintain strong internal structure even in their weakest communities.
- **Conductance**: In contrast to the above, conductance is evaluated using the *minimum of the maximum* values, since lower conductance indicates better separation between clusters.
- **Number of distinct KEGG and GO terms**: Lower values are preferred, as they suggest that clusters contain genes with more homogeneous biological functions, reflecting biologically meaningful communities.
- **KEGG and GO coverage**: Higher coverage values are desirable, indicating that clusters collectively capture a larger portion of the known biological annotations.

The corresponding weights used in the scoring function are reported in Table 3. These weights reflect the relative importance assigned to each metric.

Metric	Weight
n_{clusters}	0.10
Average cluster size	0.10
Minimum internal edge density	0.05
Minimum average internal degree	0.05
Modularity score	0.05
Maximum conductance	0.05
Number of distinct KEGG terms	0.15
Number of distinct GO terms	0.15
Average KEGG coverage	0.15
Average GO coverage	0.15

Table 3: Weights assigned to each evaluation metric in the scoring function.

4.3 Results

According to the evaluation criteria defined above, the algorithm that achieves the highest overall performance is *Angel* [6]. Table 4 reports the raw metric values for each algorithm, whereas Table 5 presents the corresponding normalized values together with the final computed scores.

A closer manual inspection of the communities produced by Angel reveals the presence of a single very large community containing 5028 nodes, while the remaining communities are substantially smaller, each consisting of fewer than 70 nodes. By contrast, the other two algorithms yield a much larger number of clusters; however, many of these are composed of only a few nodes, indicating potential fragmentation rather than meaningful structure.

A more thorough manual examination of the detected communities would be required to determine whether the identified modules are biologically meaningful and to assess the true interpretability of the results.

Method	Clus.	AvgSz	IED	AID	Mod.	Cond.	KEGG	GO	KCov	GOCov
Angel	50	115.48	0.00	1.50	190.40	0.68	22	37	0.19	0.30
Infomap	740	11.89	0.00	0.00	1154.87	0.75	183	280	0.17	0.26
Louvain	349	25.20	0.00	0.00	302.42	0.25	36	46	0.39	0.26

Table 4: Comparison of community detection metrics.

Method	Clus.	AvgSz	IED	AID	Mod.	Cond.	KEGG	GO	KCov	GOCov	Score
Angel	1.00	1.00	1.00	1.00	0.00	0.13	1.00	1.00	0.09	1.00	0.77
Louvain	0.57	0.13	0.00	0.00	0.12	1.00	0.91	0.96	1.00	0.00	0.56
Infomap	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.05

Table 5: Normalized metrics and resulting scores.

5 Network Propagation

Network propagation is a computational method used to integrate omics data by mapping it onto pre-existing molecular or genetic networks. The core principle involves iteratively smoothing or spreading the initial node scores to neighboring nodes using a weighted average. This process is key because it helps to amplify network regions strongly correlated with a specific phenotype, while simultaneously dampening the signal from areas with minor, less relevant changes (scores near zero). Although different models, such as Random Walk with Restart (RWR) and Heat Diffusion (HD), share a common conceptual structure, they are based on distinct mathematical formulations [5]. In this work, RWR was employed to spread the fold change values of the available seed nodes throughout a sign-directed network to infer the influence of these seed genes over possible novel target genes.

5.1 Sign Directed Network

The sign directed network used for the network propagation task was derived from the largest connected component of the full combined network (the same used for community discovery). The signed network was reconstructed by iterating over the initial nodes and querying the SIGNOR API for their interaction partners. No new nodes were added; only edges were included when both interacting nodes were already present in the initial network. This process yielded a network of 2,521 nodes and 6,274 edges. From this, the strongly connected component (SCC) was extracted, consisting of 492 nodes and 1,853 edges. Figure 4 shows this component, with nodes colored by fold change values (blue: negative; red: positive; white: STRING-derived nodes lacking fold-change data). Red dashed edges represent inhibition and green solid edges indicate activation.

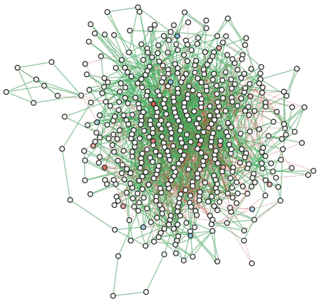

Figure 4: Strongly Connected Component Network. (Cytoscape)

Figure 5 shows the degree distribution of the nodes.

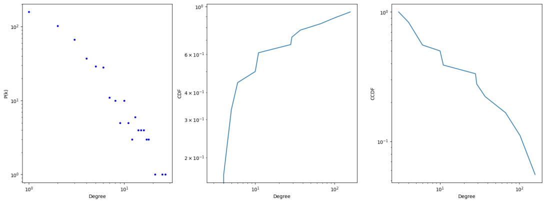

Figure 5: SCC Degree Distribution.

5.2 RWR Parameters Estimation

The parameters of the Random Walk with Restart (RWR) diffusion process were determined by maximizing the *stability* of the diffusion output. Stability quantifies the robustness of the ranked RWR scores to small perturbations in the input fold-change signal. To estimate this quantity, low-amplitude Gaussian noise was repeatedly added to the seed vector across multiple bootstrap replicates, and the diffusion was recomputed for each perturbation. For every replicate, the top-k ranked genes were extracted, and pairwise Jaccard similarities were calculated between all pairs of replicate sets. The final stability metric was defined as the mean of these pairwise similarities, providing a measure of the reproducibility of the RWR ranking under input uncertainty. Optimization of this metric yielded a spreading coefficient of $\alpha = 0.76$ and a total of 100 iterations for numerical convergence.

5.3 Results

Following diffusion, the resulting set of 492 genes was intersected with the initial list of genes provided in the original dataset. This comparison revealed 26 genes that were absent from the initial file. As previously noted, a larger set of 2070 genes was also missing from the original data matrix. Such absences may arise from two major sources: (i) *gene identifier inconsistencies*, in which discrepancies among deprecated gene names, or multiple aliases might lead to unsuccessful matching; and (ii) *single-cell RNA-seq dropout*, in which stochastically missing transcript counts cause genes with low or moderate expression to be undetected in individual cells. This dropout sparsity can obscure biologically relevant genes despite genuine expression [11], resulting in their absence from the original dataset while allowing them to emerge after network-based diffusion.

At the end of the diffusion process, the same fold-change threshold of 0.585 was applied to the set of 26 novel targets, resulting in a final subset of 21 genes. The corresponding predicted fold-change values are presented in Table 6.

In the absence of a ground truth for validation, the interpretation of these predictions remains challenging. Consequently, the predicted values should be viewed primarily as a ranking of the potential influence of each interaction within the network, rather than as precise estimates of fold-change magnitude.

Figure 6 shows the STRING-derived interaction subnetwork associated with these 26 genes. The corresponding KEGG pathway enrichment analysis, summarizing the biological functions in which these genes are most involved, is reported in Table 7.

The same functional enrichment analysis was performed separately for the upregulated and downregulated novel targets. For the downregulated genes, the same KEGG pathways were returned. Among these pathways, *Inflammatory Bowel Disease* was identified, a result consistent with the researchers' observations of reduced inflammation-related responses.

6 Conclusions

The aim of this document was to extend the analysis of the original study on miR-302b treatment [3], providing deeper insights through computational methods. The network characterization served as a tool to demonstrate that the network is neither trivial nor randomly

Node	Value
MC3R	25.7176
MC2R	25.7176
MC4R	25.7176
FOXL2	5.6956
BARD1	4.5353
MAPK10	1.5294
BMPR1B	0.7742
GATA1	0.7392
IL4	-0.7006
IL10	-0.7334
LEF1	-1.1667
IFNB1	-1.2131
PGR	-1.3038
IL6	-1.3331
DCC	-2.9287
IL2	-3.7545
GLI1	-5.4374
SHC3	-5.8882
RET	-6.0807
RIN1	-10.2631
OPRD1	-56.2580

Table 6: Node names and their associated values.

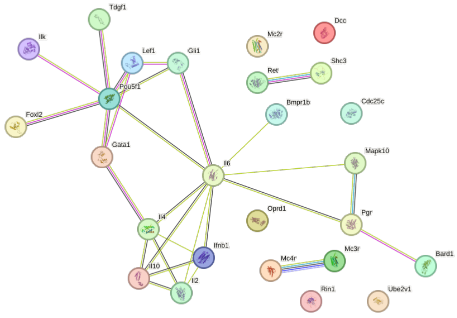

Figure 6: Novel Target Genes Network. (STRING)

Pathway	Description	Count	Strength	Signal	FDR
mmu04672	Intestinal immune network for IgA prod.	4 of 41	1.91	1.69	4.29e-05
mmu05321	Inflammatory bowel disease	4 of 60	1.75	1.58	5.83e-05
mmu05142	Chagas disease	5 of 100	1.62	1.55	4.29e-05
mmu05135	Yersinia infection	5 of 122	1.54	1.50	4.29e-05
mmu04630	JAK-STAT signaling pathway	5 of 165	1.41	1.34	7.32e-05

Table 7: KEGG pathway enrichment results for novel target genes.

generated, but rather the product of complex biological interactions. The community discovery section establishes a foundation for identifying and evaluating suitable partitioning algorithms and sets the stage for obtaining deeper insights into network functionality.

The network diffusion analysis focused on understanding whether, and to what extent, downstream interactions may be identified. Overall, this document only scratches the surface of the extensive analyses and methodologies that can be applied to extract meaningful insights from complex biological networks.

7 Discussion

The initial research questions that motivated this work were:

- What is the impact of the differentially expressed genes initially identified on the entire network?
- Is it possible to identify novel downstream interactors that were not part of the original dataset?
- Are there any disease-related interactors that may need further experimental investigation to assess the safety of the treatment?

Due to time constraints and limited biological expertise, best practices, and methodological resources, this document does not provide definitive answers to these questions. Instead, it outlines a structural framework that could support a more comprehensive analysis in the future.

Potential extensions of this work include:

- More detailed insights into the biological functions of network hubs and bridges.
- Improved biological characterization of the communities identified by the community detection algorithms.
- Deeper interpretation of the downstream interactors identified through diffusion analysis, both in the literature and in experimental settings, to assess their biological relevance.
- Investigation of biological network motifs (e.g., bifans, feed-forward/backward positive/negative loops) within the signed directed network used for the diffusion analysis.

Overall, I am grateful for the challenges, lessons, and experiences encountered during this work. I hope that this document and the accompanying codebase may be useful to others in the future.

References

[1] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* 25, 1 (2000), 25–29. doi:10.1038/75556

[2] A.-L. Barab
’asi and R. Albert. 1999. Emergence of scaling in random networks. *Science* 286 (1999), 509–512. doi:10.1126/science.286.5439.509

[3] Youkun Bi, Xinlong Qiao, Zhaokui Cai, Hailian Zhao, Rong Ye, Qun Liu, Lin Gao, Yingqi Liu, Bo Liang, Yixuan Liu, Yaning Zhang, Zhiguang Yang, Yanyun Wu, Huiwen Wang, Wei Jia, Changqing Zeng, Ce Jia, Hongjin Wu, Yuanchao Xue, and Guangju Ji. 2025. Exosomal miR-302b rejuvenates aging mice by reversing the proliferative arrest of senescent cells. *Cell Metabolism* 37, 2 (2025), 527–541.e6. doi:10.1016/j.cmet.2024.11.013

[4] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (2008), P10008.

[5] Konstantina Charmpi, Manopriya Chokkalingam, Ronja Johnen, and Andreas Beyer. 2021. Optimizing network propagation for multi-omics data integration. *PLoS Computational Biology* 17, 11 (2021), e1009161. doi:10.1371/journal.pcbi.1009161

[6] Michele Coscia, Giulio Rossetti, Fosca Giannotti, and Dino Pedreschi. 2012. DEMON: a local-first discovery method for overlapping communities. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 615–623.

[7] P. Erd
H os and A. R
H enyi. 1959. On Random Graphs I. *Publicationes Mathematicae Debrecen* 6 (1959), 290–297.

[8] Minoru Kanehisa and Susumu Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28, 1 (2000), 27–30. doi:10.1093/nar/28.1.27

[9] Prisca Lo Surdo, Marta Iannuccelli, Silvia Contino, Luisa Castagnoli, Luana Licata, Gianni Cesareni, and Livia Perfetto. 2023. SIGNOR 3.0, the SIGnaling Network Open Resource 3.0: 2022 update. *Nucleic Acids Research* 51, D1 (2023), D631–D637. doi:10.1093/nar/gkac883

[10] Michael I. Love, Wolfgang Huber, and Simon Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 12 (2014), 550. doi:10.1186/s13059-014-0550-8

[11] Peng Qiu. 2020. Embracing the dropouts in single-cell RNA-seq analysis. *Nature Communications* 11, 1 (2020), 1169. doi:10.1038/s41467-020-14976-9

[12] Martin Rosvall and Carl T. Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* 105, 4 (2008), 1118–1123.

[13] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102, 43 (2005), 15545–15550. doi:10.1073/pnas.0506580102

[14] Damian Szklarczyk, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja Hachilif, Annika L. Gable, Tao Fang, Nadezhda T. Doncheva, Sampo Pyysalo, Peer Bork, Lars J. Jensen, and Christian von Mering. 2023. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research* 51, D1 (2023), D638–D646. doi:10.1093/nar/gkac1000