Uncovering Temporal Patterns in *Saccharomyces cerevisiae* Gene Co-expression Networks through Dynamic Community Detection

Carmelo Ellezandro Atienza[1] and Calvin James Maximo[2]

[1] Algorithms and Complexity Lab,
University of the Philippines Diliman
cratienza1@up.edu.ph
[2] Algorithms and Complexity Lab,
University of the Philippines Diliman
ctmaximo1@up.edu.ph

1 Background of the Study

1.1 Network Biology and Biological Networks

Biological systems are inherently complex—molecules such as genes, proteins, and metabolites intricately interact with each other to perform essential functions that regulate cellular processes. Barabási and Oltvai (2004) highlight that traditional reductionist approaches in examining biological functions is insufficient, since such methods only focus on individual cellular components. In reality, biological characteristics are seen as a product from complex interactions between multiple molecules, such as proteins, DNA, and small molecules (Barabási and Oltvai, 2004). This shift in perspective gave way for network biology: an interdisciplinary domain which integrates the computational field of network theory with biological sciences.

Girvan and Newman (2002) emphasize that many networks in different domains of study exhibit community structures, where nodes cluster into tightly connected groups. In particular, this phenomenon is prevalent in biological systems, where such community structures may pertain to functional modules—a group of genes, proteins or other biological entities that work in conjunction to perform a specific biological role. Biological networks, such as gene regulatory networks (GRNs), protein-protein interaction networks (PPIs) and metabolic pathways, are compelling topics for research as it can uncover how certain molecules are grouped within cells (Yu et al., 2013).

Dorogovtsev and Mendes (2002) explained that a crucial aspect of biological networks is that they are not static—they evolve over time, adapting to changing internal and external conditions. This dynamic nature is especially evident in gene co-expression networks, where gene interactions fluctuate throughout different stages of biological processes. Thus, studying the dynamic community structure within these networks can provide insights into how biological functions are regulated, and how brief interactions between genes may arise and disappear during key cellular events. This evolving nature of biological networks makes

dynamic community detection a critical tool for revealing how these molecular interactions change over time.

1.2 Dynamic Community Detection in Networks

Building on this understanding of dynamic biological networks, dynamic community detection has emerged as a key approach for analyzing networks across different domains, including social networks and biological systems. Much of the existing research regarding community detection begins with a strict assumption that real-world phenomena are represented using static networks which are fixed and unchanging over time. However, this assumption does not align with the evolutionary world around us.

Complex networks such as gene regulatory networks are generated at a blistering pace, which spawns the problem of the difficulty of modeling these networks as static relations. Time is an important component in studying the evolution of connectivity and patterns. The dynamic nature of these real-world networks further complicates the task of community detection as the structure of these communities evolve along with the networks themselves. The need to integrate the knowledge from the temporal dimension has led to the emergence of a new field of study: dynamic network analysis (Rossetti and Cazabet, 2018; Sattar et al., 2023).

Dynamic community detection is one of the key problems under this field. Over time, nodes and edges in a dynamic network can appear and disappear, causing significant shifts in its structure. Communities, which are the fundamental building blocks of this complex network, are heavily influenced by these local changes. While the objective of traditional community detection algorithms is to reveal hidden substructures, the goal of dynamic approaches is to monitor these local topological changes and their transformations over time (Rossetti and Cazabet, 2018). Understanding how community structures in complex networks evolve over time are vital in gaining valuable insights. The insights can be used to understand the structural dynamicity of the network. There is also great potential in the medical domain for analyzing disease progression (Redekar and Varma, 2022).

1.3 Gene Co-expression Networks and their Dynamic Nature

Gene co-expression networks (GCNs) are widely used to model interaction between genes based on their expression levels across different biological conditions or datasets, and across varying time points (Ovens et al., 2021; Lau et al., 2020). In GCNs, nodes represent genes, and edges signify co-expression relationships, where the strength of an edge correlates with the similarity in gene expression.

That said, GCNs will be the focus of this paper because of their temporal and dynamic properties. A study done by Lau et al. (2020) show the need to capture the temporal dynamics of GCNs, as gene co-expression patterns fluctuate significantly over time. Consequently, the community structure of GCNs will

also fluctuate, with different clusters of genes forming or dissipating at different timepoints. These changes highlight the importance of dynamic community detection techniques in capturing the evolving structure of GCNs (Ovens et al., 2021), which can reveal crucial insights on their functional grouping.

2 Preliminaries

2.1 Gene Co-expression Networks (GCNs)

A graph-based representation of genes, where nodes represent individual genes and edges represent co-expression relationships based on their expression patterns across multiple conditions or time points. Co-expression relationship refers to the correlation between the expression levels of two or more genes across different conditions (Stuart et al., 2003). Expression levels refer to the amount of mRNA produced by a gene in a cell at a given time. To put it simply, co-expression relationship is when two or more genes show similar patterns of activity (expression levels) across different conditions or samples. If their activity goes up or down together, they are said to be co-expressed, which often means they might work together or be controlled by the same regulatory processes (Ovens et al., 2021).

2.2 Community Detection

The process of identifying clusters of groups of nodes in a network that are more densely connected internally than with nodes outside the group. There are several community detection techniques (Redekar and Varma, 2022), depending on the nature of the network, whether static or dynamic.

2.2.1 Static Community Detection. Static community detection assumes that the network is fixed across different time points. From there, communities are identified based on a single snapshot of the network. This method is well-suited for networks where relationships between nodes are stable and predictable, if not unchanging. Techniques such as Louvain or Girvan-Newman are prime examples of static community detection algorithms, where it can reveal hierarchical structures in the network (Redekar and Varma, 2022).

2.2.2 Dynamic Community Detection. Dynamic community detection is designed for networks that evolve over time. It is able to detect communities and how they vary at each time step, tracking the evolution, merging and splitting of nodes and edges as the network changes (Dorogovtsev and Mendes, 2002). Capturing these temporal changes are crucial for understanding time-dependent phenomena, such as gene interactions that change during biological processes.

For the purposes of this paper, dynamic community detection is employed, as the study focuses on the evolving structure of gene co-expression networks across different phases of the cell cycle in *Saccharomyces cerevisiae*.

2.3 *Saccharomyces cerevisiae*

According to Parapouli et al. (2020), *Saccharomyces cerevisiae* is a unicellular fungus that is hailed as model organism—extensively used in biological research due to its well-annotated genome. In the field of bioinformatics, *Saccharomyces cerevisiae* contains around 6,000 documented genes which allows for in-depth research on cellular processes that is shared with higher-level eukaryotes. This includes gene regulation, cell cycle control, and metabolic functions. Due to its accessibility and applications, the gene expression data of this fungus is an ideal candidate for constructing a GCN through Dynamic Community Detection techniques. In particular, there is preponderance of data from Cho et al. (1998), where they conducted a time-series study on *Saccharomyces cerevisiae* across different the different phases of the mitotic cell cycle. This dataset captures the fluctuating expression of genes throughout the yeast's cell cycle, such as during the phases G1, S, and G2.

3 Review of Related Literature

3.1 Hierarchical Community Detection on GCNs

The study conducted by Calderon and Ventures (2024) provides a foundational exploration of hierarchical community detection on gene co-expression networks. Their research paper sought to explore the functional organization of gene communities within the *Saccharomyces cerevisiae* cell-cycle dataset using hierarchical clustering techniques to uncover the existing substructures of co-expressed genes. In particular, they conducted a performance analysis on three hierarchical community detection algorithms: Girvan-Newman (GN), Paris, and Local Optimization Function Model (LFM).

Hierarchical community detection is appropriate for GCNs due to their hierarchical nature; genes may form smaller functional modules that integrate into larger biological processes. Calderon and Ventures (2024) applied the three community detection algorithms which can reveal modular and nested structures that govern gene regulation in *Saccharomyces cerevisiae*.

However, while their hierarchical approach provided significant insights into the static structure of the GCN, it did not account for the dynamic nature of gene interactions, as mentioned by Calderon and Ventures (2024) in their recommendations—it is an open problem left by their research. This static focus on hierarchical organization is what this present study aims to address; by shifting to a dynamic community detection approach, this paper will explore how gene communities evolve across different cell cycle phases, thereby capturing temporal changes.

3.2 Dynamic Community Detection Algorithms

Dynamic community detection (DCD) algorithms not only uncover communities in complex networks but also the evolution of these networks over time. Several

studies have probed into the performance of various DCD algorithms on both real-world and synthetic datasets (Ma and Dong, 2017; Seifikar et al., 2020; Redekar and Varma, 2022; Sattar et al., 2023; Singh et al., 2020). We will be taking a look on a selection of noteworthy algorithms from previous studies, which all offer unique methods in discovering relevant communities in dynamic networks.

3.2.1 C-Blondel Algorithm. The C-Blondel algorithm is a DCD algorithm proposed by Seifikar et al. (2020) that is built upon the Louvain algorithm. It is an algorithm that leverages the historical information of network which improves its execution time. It applies the Louvain method over a compressed graph generated from the previous network snapshot. It runs faster than the Louvain algorithm by working with a smaller and more efficient compressed graph. This graph consists of supernodes that represent the previously detected communities, and superedges which connect these supernodes. It is important to note that there are no destructive nodes inside of these communities. Destructive nodes are defined to be the nodes in a community wherein the lack of them causes a corresponding community to break up into smaller subcommunities. To recognize these nodes, a degree centrality heuristic is employed. A node is considered destructive if its degree centrality is higher than the community average. The C-Blondel algorithm accounts for all changes in the network which are handled by the removal of a node. The remaining communities and subcommunities are transformed into supernodes, and their links become superedges, forming the new compressed graph.

The study compared C-Blondel with other DCD algorithms, particularly S-Blondel (Greene et al., 2010) and D-Blondel (He et al., 2017), on three real-world datasets. They concluded that it outperformed the selected algorithms, while remaining comparable in terms of the modularity metric (Seifikar et al., 2020).

3.2.2 Semi-supervised ENMF (sE-NMF) Algorithm. The sE-NMF algorithm is an algorithm proposed by Ma and Dong (2017). This algorithm combines evolutionary nonnegative matrix factorization (ENMF) frameworks for dynamic community detection, and spectral clustering. The aim of ENMF is to balance the trade-off between maximizing clustering accuracy at a given time step and maintaining the consistency of the clusters across time steps. A key difference of sE-NMF from classical semi-supervised algorithms is the integration of a priori information to its primary function. It is advantageous because it enables the algorithm to avoid local optima without affecting the run time complexity (Ma and Dong, 2017).

The group evaluated the performance of the sE-NMF algorithm against three well-known algorithms (FacetNet, Kim-Han, DYNMOGA) on three artificial and two real-world dynamic networks. One of the real-world networks is of particular interest, which modeled breast cancer progression based on the gene-expression data. It was found out that sE-NMF performs better than the other algorithms in both the specificity and sensitivity metrics, which is a great indicator of its

ability to discover biological dynamic communities. It is also a great tool for assessing temporal networks of disease progression (Ma and Dong, 2017).

References

Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113. https://doi.org/10.1038/nrg1272

Calderon, J. K., & Ventures, P. A. (2024). Hierarchical community detection on co-expression networks for functional classification of cell-cycle regulated genes of saccharomyces cerevisiae. *Algorithms and Complexity Lab, University of the Philippines Diliman*.

Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J., & Davis, R. W. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular Cell*, *2*(1), 65–73. https://doi.org/10.1016/s1097-2765(00)80114-8

Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in Physics*, *51*(4), 1079–1187. https://doi.org/10.1080/00018730110112519

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799

Greene, D., Doyle, D., & Cunningham, P. (2010). Tracking the evolution of communities in dynamic social networks. *2010 International Conference on Advances in Social Networks Analysis and Mining*, 176–183. https://doi.org/10.1109/ASONAM.2010.17

He, J., Chen, D., Sun, C., Fu, Y., & Li, W. (2017). Efficient stepwise detection of communities in temporal networks. *Physica A: Statistical Mechanics and its Applications*, *469*, 438–446. https://doi.org/https://doi.org/10.1016/j.physa.2016.11.019

Lau, L. Y., Reverter, A., Hudson, N. J., Naval-Sanchez, M., Fortes, M. R. S., & Alexandre, P. A. (2020). Dynamics of gene co-expression networks in time-series data: A case study in drosophila melanogaster embryogenesis. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00517

Ma, X., & Dong, D. (2017). Evolutionary nonnegative matrix factorization algorithms for community detection in dynamic networks. *IEEE Transactions on Knowledge and Data Engineering*, *29*(5), 1045–1058. https://doi.org/10.1109/TKDE.2017.2657752

Ovens, K., Eames, B. F., & McQuillan, I. (2021). Comparative analyses of gene co-expression networks: Implementations and applications in the study of evolution. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.695399

Parapouli, M., Vasileiadi, A., Afendra, A.-S., & Hatziloukas, E. (2020). Saccharomyces cerevisiae and its industrial applications. *AIMS Microbiology*, *6*(1), 1–32. https://doi.org/10.3934/microbiol.2020001

Redekar, S. S., & Varma, S. L. (2022). A survey on community detection methods and its application in biological network, 1030–1037. https://doi.org/10.1109/ICAAIC53929.2022.9792913

Rossetti, G., & Cazabet, R. (2018). Community discovery in dynamic networks: A survey. *ACM Comput. Surv.*, *51*(2). https://doi.org/10.1145/3172867

Sattar, N. S., Buluc, A., Ibrahim, K. Z., & Arifuzzaman, S. (2023). Exploring temporal community evolution: Algorithmic approaches and parallel optimization for dynamic community detection. *Applied Network Science*, *8*(1), 64. https://doi.org/10.1007/s41109-023-00592-1

Seifikar, M., Farzi, S., & Barati, M. (2020). C-blondel: An efficient louvain-based dynamic community detection algorithm. *IEEE Transactions on Computational Social Systems*, *7*(2), 308–318. https://doi.org/10.1109/TCSS.2020.2964197

Singh, D. K., Haraty, R. A., Debnath, N. C., & Choudhury, P. (2020). An analysis of the dynamic community detection algorithms in complex networks. *2020 IEEE International Conference on Industrial Technology (ICIT)*, 989–994. https://doi.org/10.1109/ICIT45562.2020.9067224

Stuart, J. M., Segal, E., Koller, D., & Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science*, *302*(5643), 249–255. https://doi.org/10.1126/science.1087447

Yu, D., Kim, M., Xiao, G., & Hwang, T. H. (2013). Review of biological network data and its applications. *Genomics Inform*, *11*(4), 200–210. https://doi.org/10.5808/GI.2013.11.4.200