A Data Analysis Pipeline for Identifying Periodic Processes during Drosophila Development

Mohammad Shafiqul Islam
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
mdshafiqul.islam@northsouth.edu

Mohammad Rafsun Jany Mahin
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
jany.mahin@northsouth.edu

Ahsanur Rahman
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
ahsanur.rahman@northsouth.edu

***Abstract* — In this paper, we propose a computational pipeline that can be used to unearth periodic processes and their regulators in any organism along with the networks governing such periodicity. Our approach is based on mining periodic subgraphs from temporal gene networks. Specifically, we collected 30 time varying gene networks inferred from temporal expression profiles of 588 Drosophila genes, computed periodic subgraphs in those networks, and analyzed them in a number of ways in order to discover their biological significance. Our results show that the largest connected component in the periodic subgraphs as well as hub genes and dense subgraphs in that component are highly enriched in periodically active gene-functions. We also devised a way to find the regulators of these periodic functions. We show the superiority of our approach as compared to a baseline method that computes aperiodic subgraphs by showing that similar analysis on aperiodic subgraphs fails to find any periodic or specific gene function. To the best of our knowledge, this work is the first of its kind in the field of network biology.**

Keywords—periodic subgraph, graph mining, time-varying networks, periodic process, Drosophila development

I. Introduction

In recent years, graph mining has attracted the attention of a lot of researchers. One of the main goals of graph mining is to find interesting patterns from one or more graphs. Such patterns include hubs [1], dense subgraphs [2], frequent dense subgraphs [3], periodic subgraphs [4], and so on. Different patterns are interesting for different purposes. For example, hubs tend to indicate leaders in a social network and essential genes in a protein-protein interaction (PPI) network [5], dense subgraphs reveal communities in a social network [6], frequent dense subgraphs tend to indicate communities in a sequence of communication networks [7], periodic subgraphs often indicate periodically formed communities in dynamic social networks [4], and so on.

Among them, periodic subgraph mining is one of the least studied problems. Especially, to the best of our knowledge, no study attempted to identify periodic subgraphs in time-varying gene/protein networks. It is not only important to fill up this gap but also periodic subgraphs in such networks may give us interesting insights of the underlying biology of the cell. Our intuition come from the following literature review.

Multiple processes get activated periodically to maintain cell-cycle [8] and circadian rhythm [9]. Periodic oscillation was observed in the abundance of cellular molecules such as cAMP secreted by starved amoeba (period ≈ 5 min), cytosolic Ca^{++} ion (period depends on cell type), NF-KB (period ≈ 100 min), tumor suppressor p53 (period is a few hours), etc [9].

Biological functions and regulators exhibit periodicity during organism development, as well. For example, cell cycle occurs every 30 min during amphibian embryonic development and is driven by the periodic activation of a cyclin/CDK complex [9]. In vertebrate embryos, Notch signaling pathway periodically activates cellular processes involved in somitogenesis [9]. DNA synthesis happens periodically during the early development of Xenopus [10]. Finally, a transient neural network causes the generation of periodic waves of activities in the retina (known as *retinal waves*) during the development of certain vertebrates [11].

Clearly, periodic activation of processes and regulators is quite common in biology. Traditionally, it has been assumed that such periodicity results from static regulatory networks [9] of cellular molecules (genes, proteins, mRNAs, etc.). However, recent research suggests that molecular interactions within a cell vary spatiotemporally, thereby resulting in a dynamic network rather than a static one [12]. Finding periodic subgraphs within such a dynamic network may shade light on both the periodic processes and their regulators.

These observations motivated us to compute periodic subgraphs within time-varying gene networks. Specifically, we made the following contributions in this work.

- We did extensive literature review to demonstrate the importance of mining periodic subgraphs in gene networks.
- We propose a computational pipeline that can be used to unearth periodic processes and their regulators in any organism along with the networks governing such periodicity. Our pipeline consists of inferring temporal networks from gene expression profiles, computing periodic subgraphs in those networks, and analyzing those subgraphs to understand their biological significance.
- We show that our approach is effective for discovering cellular processes (along with their regulators) that are periodically active during the development of Drosophila.
- We show that our approach is superior to a baseline method by showing that similar analysis on a random sequence of subgraphs fails to discover any periodic gene function.

Rest of the paper is organized as follows. Section II demonstrates the concept of periodic subgraphs. Section III discusses relevant works. Section IV sheds light on the input networks. Section V elaborates our computational pipeline. We discuss our results in Section VI and compare our results with the baseline method in section VII. We conclude in Section VIII and provide future directions in Section IX.

II. Periodic Subgraph Illustration

We demonstrate the concept of periodic subgraphs in the temporal networks shown in Fig. 1. This figure shows first six snapshots of a time-varying network over eight nodes. Each node is a gene involved in Drosophila's development and an edge exist between two genes if they are likely to be co-regulated (see Section IV for details). These genes are chosen because they are biologically interesting. Specifically, *Nrt, dock, brm, Src64b*, and *msn* are involved in *neuron projection guidance*[1] whereas the master regulator *Notch (N)* regulates this process [13]. Thus this network dataset serves as a motivating example for periodic subgraphs.

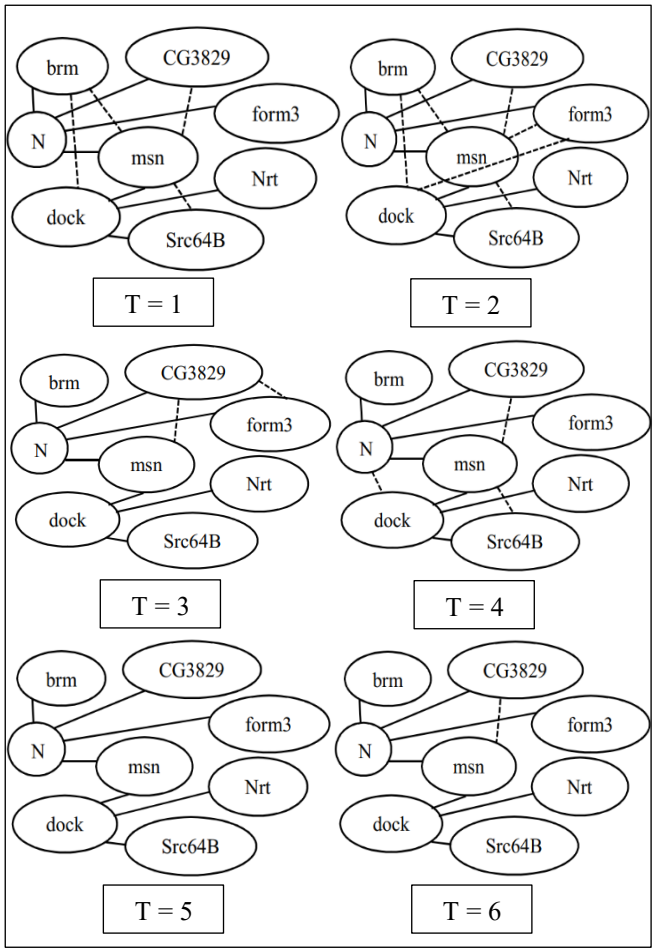

Fig. 1. First six snapshots of Drosophila developmental network over eight genes. Here the graph at timestep 6, namely, G_6 is a periodic subgraph having period 2, support *3,* and support sequence *[2, 4, 6]*. Straight lines represent edges which are present at all timesteps and dotted lines indicate other edges.

Consider the last graph G_6 in this figure. The sequence of timesteps in which G_6 occurs is *[1, 2, 3, 4, 6]*. We call this sequence the *support sequence* of G_6. As it occurs at every timestep from *1* to *4*, we call it a periodic subgraph with *period 1* and *support 4*, or a *(1, 4)-periodic subgraph*, in short. To be precise, we call it a *(1, 4)-periodic subgraph* having support sequence *[1, 2, 3, 4]*. G_6 can also be called a *(2, 3)*-periodic subgraph with support sequence *[2, 4, 6]* because it occurs at *3* timesteps where consecutive steps are always *2* steps apart, *i.e.,* it has period *2* and support *3*.

With this illustration, we are now ready to present the formal definition of periodic subgraphs.

Definition 1. Periodic Subgraph[2]: Given a dynamic network $\mathcal{G}$, a subgraph F is called a *(p, s)-periodic subgraph with support sequence S in* $\mathcal{G}$ if it satisfies the followings.

- **Periodicity:** Consecutive timesteps in S are p steps apart.
- **Minimum Support:** The support, $s = |S| \geq \sigma$, where σ is a support threshold given by the user.
- **Structural Maximality:** No other nodes or edges can be added into subgraph F because if we do so then the modified F does not have period p or support s anymore.
- **Temporal Maximality:** No other timesteps can be added to the support sequence S because F is absent in $\mathcal{G}$ at timestep $S_1 - p$ as well as at timestep $S_{|S|} + p$.

We now demonstrate that G_6 is indeed a *(1, 4)*-periodic subgraph with support sequence *[1, 2, 3, 4]* according to these conditions. For example, if the edge (*brm, dock*) is added to G_6, then the modified graph does not have support 4. Similarly adding any other node or edge changes period or support. So G_6 is structurally maximal. It is also temporally maximal since we cannot add any timestep to its support sequence (as G_6 is absent at timestep, $T = 5$). Similarly, it can be shown that G_6 is a *(2, 3)*-periodic subgraph with support sequence *[2, 4, 6]*.

III. Related Works

Here we discuss other works that are closely related to us. We classify these works into the following two categories.

Mining Periodically Co-expressed Genes: Some work applied mathematical techniques (e.g., Fourier analysis) upon temporal gene expression profiles to find clusters of genes that are periodically co-expressed [14], [15]. These works are neither applicable to temporal gene networks nor can they reveal the underlying networks governing such periodicity. Since temporal networks are slowly becoming available with the recent development of novel experimental techniques [12] computing periodic subgraphs within such networks may help us to achieve these goals.

Mining Periodic Subgraphs in Social Networks: Lahiri and Berger-Wolf [4] developed an algorithm called *PSEMiner* for mining periodic subgraphs from time-varying networks. It stores the actual and potential periodic subgraphs found so far in a tree. This tree is traversed and updated at each timestep. Such whole tree traversals renders it inefficient having a time complexity of $O((V+E)T^3 ln(T/\sigma))$ where σ is the support threshold and *V, E, T* are the numbers of vertices, edges, and timesteps, respectively. Later, Apostolico, Barbares and Pizzi [16] devised a method called *ListMiner* to mine periodic subgraphs more efficiently. For each timestep t and each period p, it maintains a list of networks where consecutive networks are p timesteps apart, the first one occurs at timestep x and the last one occurs at t. Thus a periodic subgraph can easily be computed by taking the intersection of networks in such a list. *ListMiner* gains efficiency by updating intersection of each list incrementally at each timestep. It also relies on the fact that, at a certain timestep t, only lists with $x = t \bmod p$ are required to be updated. These tricks enables *ListMiner* to

[1] A process related to development of neurons via which migration of neuron projection is guided in response to various cues

[2] This definition is a simplified version of the definition of "periodic subgraph embedding" presented in [4]

achieve a time complexity of $O((V+E)T^2 ln(T/\sigma))$, *i.e.,* it is T time faster than *PSEMiner*. Since *ListMiner's* code is available online and it is fast enough for our dataset, we used it to find the periodic subgraphs in our work. Recently, Halder, Samiullah and Lee [17] proposed a more space efficient algorithm called *SPPMiner*. It stores all periodic and aperiodic entities (vertex and edges) in a data structure called supergraph and updates it at each timestep. Its time complexity is $O(V^2(T/\sigma)^2)$.

IV. Dataset

To be able to identify periodic subgraphs, we needed a sufficiently large time-varying network dataset. We collected a dataset [18] consisting of *66* undirected networks for this purpose. These networks were inferred from the temporal expression profile of *588* genes involved in the development related processes in Drosophila. Each network represents the binary associations among these genes at a certain timestep during Drosophila development and an edge between two genes in it indicates that these genes are probably co-regulated at that timestep.

The procedure to infer these networks is briefly described here in order to explain our rational for choosing this dataset. Firstly, a microarray dataset was collected which contains temporal expressions of *4028* genes sampled at *66* distinct time-periods over the first *40* days Drosophila's life [19]. Then a logistic regression method (called kernel reweighed *l1*-regularized logistic regression, or KELLER, in short) [18] was applied on the expression profiles of development-related genes (*588* of total *4028* genes) in such a way that – (i) there may be an edge between two genes at timestep t if their expressions vary in a correlated manner near t and (ii) temporally adjacent networks share many more interactions than the distant ones[3]. This second constraint tries to ensure that the topology of the inferred networks changes slowly over time. Therefore periodic subgraphs are less likely to arise in such networks due to a spurious correlation at any timestep.

V. Methodology

Our pipeline for mining and analyzing periodic subgraphs in Drosophila developmental networks consists of several stages. We wrote Matlab and Python scripts for processing inputs/outputs in different stages of our pipeline. We used a Windows *64*- bit machine with *8* GB RAM and *2.3* GHz Intel *i5* CPU cores for all our operations. We discuss the stages of our pipeline in details below.

A. Preprocessing

We collected *66* networks derived [18] from the author's website[4]. These networks are available in Matlab format. We converted these files to a format amenable to ListMiner (see supplement[5] for details). We kept only the first *30* networks for mining periodic subgraphs because these networks have uniform time-gap between them. These networks represent temporal co-regulatee relations among genes during the early embryonic development of Drosophila. There were total *249382* edges in these networks after removing duplicates and the number of edges in a timestep varied from *3434* to *4124*. ListMiner took about 1 sec to compute all periodic subgraphs with period ≥ 1 and support ≥ 3 from these networks.

B. Choosing Parameters

We applied ListMiner on the pre-processed input using a range of (minimum period, minimum support) parameter combinations. Specifically, we varied minimum period from *1* to *N/3* and minimum support from *3* to *N*/(minimum period), where $N = 30$ is the number of timesteps. Table I summarizes the outputs of ListMiner for these parameter combinations. In this table, a cell at p-th row and s-th column indicates the number of unique genes present in the *(p,s)-periodic subgraph*. As expected, the number of genes in these subgraphs decreases with the increase of period or support.

TABLE I. Number of Unique Genes present in the periodic Subgraphs for Each Parameter combination. Empty cells indicate that no periodic subgraph was found for its corresponding period and support.

Periods	Supports						
	3	*4*	*5*	*6*	*7*	*8*	*9*
1	587	586	571	511	450	360	275
2	585	485	355	232	149	103	10
3	498	333	147	92	15		
4	394	160	12				
5	279	93					
6	171	21	2				
7	113	7	2				
8	28	2					
9	21	2					

Based on these observations we decided to choose parameters in such a way that (i) both period and support are not too low and (ii) number of genes is not too low. Our reasons for choosing parameters in such a way are as follows. Firstly, periodic subgraphs with very low period and support covers almost all of the *588* genes which implies that these subgraphs exhibit pseudo-periodicity rather than true periodicity caused by biological cues because it is highly improbable that all genes in our network are involved in periodic intereactions. Secondly, one of our goals was to understand why a set of genes show periodic inter-connections. Are they involved in any periodic process in the cell (such as cell cycle)? Are they regulated by a periodically active transcription factor? Definite answers to such questions cannot be obtained if the number of genes in the concerned gene-set is too low. Because small sets of genes often share multiple processes (due to multi-functionality of genes [20]) and as such it is difficult to pinpoint the function(s) because of which these genes exhibit periodicity. That's why we stipulate that the number of genes in a periodic subgraph cannot be too low. Specifically, we suggest that the number of genes shouldn't be less than a hundred.

Applying these heuristic rules, we selected two parameter combinations: (period *2*, support *8*) and (period *7*, support *3*). These parameter choices are highlighted in Table I (green cells). These choices allows us to investigate both types of periodic subgraphs: the ones with small period but large support as well as the ones with large period but small support.

[3] This is achieved by assigning more weights to the observations near timestep t and gradually decreasing weights of further observations

[4] URL: http://www.sailing.cs.cmu.edu/main/?page_id=431

[5] Detailed results are available in: https://sites.google.com/view/ppipeline

C. Analyzing Periodic Subgraphs

There are *2* (respectively, *7*) *(2,8)-periodic subgraphs* (respectively, *(7,3)-periodic subgraphs*). Many of them have multiple connected components. Among all the components in each of these periodic subgraphs, the largest component was chosen for biological significance analysis. The reason for choosing the largest component is: (i) connected genes share regulators and therefore should be more biologically meaningful than disconnected ones and (ii) large sets of genes tend to yield more reliable results (than smaller ones) in enrichment analysis[6]. The largest components are shown in Fig. 2 and their properties are described in Table II.

TABLE II. PROPERTIES OF PERIODIC SUBGRAPHS

Parameters	Support Sequence	# Genes	# Edges
Period 2, Support 8	1, 3, 5, 7, 9, 11, 13, 15	33	56
Period 7, Support 3	1, 8, 15	39	64

We analyzed these components in different ways in order to discern relevant periodically active cellular processes and their regulators. Firstly, we analyzed the components themselves. Specifically, we used FlyEnrichr tool [21] to find GO [7] biological process (BP) terms having statistically significant overlap with the genes in each component. FlyEnrichr outputs different measures for each GO term including adjusted p-value and a score (called combined score, henceforth, simply score) calculated by combining simple p-value with an empirical z-score. FlyEnrichr authors suggested that this score is a better measure of the statistical significance of gene-functions than the simple p-values. So we sorted the terms in decreasing order of their scores after filtering out insignificant terms (specifically terms having adjusted p-value > 0.05 or those sharing only one gene with the largest component). We deem remaining terms to have statistically significant overlap (*a.k.a., high enrichment*) with the concerned component.

Secondly, we wanted to understand the roles of the so called "hub genes" (nodes having high degrees) in our largest components. Our motivation for this analysis comes from the fact that hubs in gene networks tend to be essential genes [5]. We wanted to investigate whether hubs in our largest component have any such interesting role. So we sorted the genes in the largest component in the descending order of their degrees and used FuncAssociate tool [22] to check if this ranked list is highly enriched in any GO BP term.

Thirdly, since each edge in our network indicates a co-regulated relationship, genes in a dense subgraph in it are expected to be regulated by the same regulator. If this dense subgraph is periodic then the corresponding regulator should be recurrently/consistently active. To test this hypothesis, we applied ClusterOne algorithm [23] (using default parameters) on each largest component to identify dense subgraphs (*a.k.a., clusters*) in it (Table III). We got the same set of clusters for both parameters. Then we applied FlyEnrichr on these clusters to compute their enrichment in the genes co-regulated by Drosophila transcription factors (*TFs*). Finally, we sorted these *TFs* in the ascending order of their combined scores and applied FuncAssociate on this ranked list to compute its enrichment in GO biological process (BP) terms.

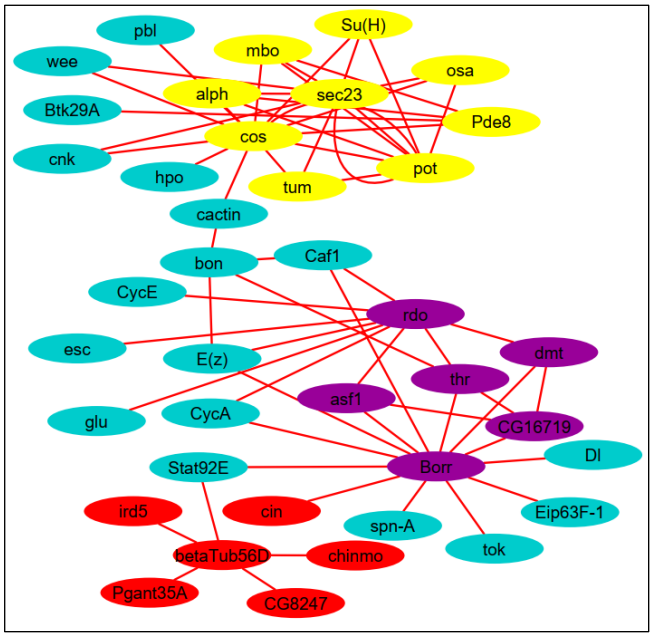

Fig. 2. Largest connected component of periodic subgraphs for our chosen parameters. Red nodes are the genes which are present only in the largest component of *(7,3)-periodic subgraphs*. The subgraphs with yellow and purple nodes represent the dense subgraphs (respectively, cluster-1 and cluster-2) we found in the largest component.

TABLE III. PROPERTIES OF CLUSTERS IN LARGEST COMPONENTS

Clusters	ClusterOne p-value	# Genes	Density
Cluster-1	0.001	9	0.61
Cluster-2	0.183	6	0.67

We used Cytoscape network visualization and analysis platform [24] for analyzing networks. We summarized all our enrichment results using ReviGO [25] and present at most top five GO terms with the highest FlyEnrichr (respectively, FuncAssociate) scores (respectively, LOD ratio) which have ReviGO dispensability[8] less than *0.35* (this cut off is stricter than the *0.5* cut off used in ReviGO paper [25]). We discuss the results got from these analyses in the next section.

VI. RESULTS

A. Largest component is involved in periodic processes

The GO BP term with the highest enrichment in each of our largest components is *positive regulation of cell cycle G1/S phase transition*. Three genes of our largest component, namely, *cycA*, *cycE*, and *su(H)* annotate this term (Table IV), *i.e.,* their protein products activate *G1/S phase transition*. This is a very interesting result because cell cycle transitions, including this one, are controlled by periodic synthesis and degradation of cyclins, which includes *cycA* and *cycE* [26]. We suspect that *su(H)* may have a similar role in cell cycle – which can be validated via wet-lab experiments.

Our third most enriched term (GO:0031062) is annotated by *E(z)* and *esc*: two subunits of *ESC/E(Z) complex* (other

[6] http://software.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html
[7] http://geneontology.org/docs/ontology-documentation/
[8] Low dispensability indicates that a GO term is a good representative of other enriched GO terms that are semantically similar to this term

subunits are *Caf1* and *Su(z)12)*, *a.k.a., PRC2 complex*. This complex activates methylation of certain histone (specifically, *H3K27*) proteins and is regulated by *Cdk1* and *Cdk2*. Since *Cdk1/CycB* and *Cdk2/CycE* are known to be periodically expressed [26], [9]; it is likely that their target, *ESC/E(Z) complex*, is also periodically active – which explains why GO:0031062 is enriched with periodically co-regulated genes.

The other three highly enriched terms are related to Drosophila development. Among them *imaginal disc-derived wing morphogenesis* may be a periodic function because many periodically co-expressed genes are involved in wing disk development [27]. Based on these positive results, we postulate that remaining two processes are also periodic - which can be tested experimentally. Together, these results suggest that the genes in the largest component of a periodic subgraph often participate in periodic processes.

TABLE IV. FLYENRICHR ENRICHMENT OF THE LARGEST COMPONENT OF PERIODIC SUBGRAPHS IN GO BIOLOGICAL PROCESSES. *P2S8* MEANS (PERIOD 2, SUPPORT 8), AND *P7S3* MEANS (PERIOD 7, SUPPORT 3.

GO Term ID	GO Term Name	Adj. p- value		Score	
		p2s8	p7s3	p2s8	p7s3
GO:1902808	positive regulation of cell cycle G1/S phase transition	5E-05	8E-05	40.5	39
GO:0007422	peripheral nervous system development	2E-10	1E-09	39.7	37.7
GO:0031062	positive regulation of histone methylation	2E-03	3E-03	34.1	32.8
GO:0007476	imaginal disc-derived wing morphogenesis	7E-10	2E-07	30.2	28.8
GO:0060541	respiratory system development	2E-06	4E-09	25	28.1

B. Hub genes participate in periodic processes

Four GO biological processes are enriched in the degree-sorted list of genes in our largest component. Only one of them, namely, *Protein-DNA Complex Assembly*, remains indispensable (dispensability value is 0) after ReviGO summarization. This term got high enrichment scores in FuncAssociate (LOD *1.29*, adj. p-value *0.001*). This result is expected because protein-DNA interactions exhibit periodicity during the development of early embryo [28].

C. Densely connected genes participate in periodic processes

We found that dense subgraphs of the largest component are enriched in known periodic processes (Tables V and VI). Some of these processes (italicized in Tables V and VI) were also enriched in the largest components (Table IV), as expected. One of them (*GO:0007476*) was suggested to be periodic (Section A). Among the rest, *regulation of ERK1 and ERK2 cascade* may be a periodic process because *ERK pathway* exhibits periodic oscillation [29]. Also, *Mitotic sister chromatid segregation* happens during the M phase of mitotic cell cycle [30] which explains why it is likely to be periodic.

TABLE V. FLYENRICHR ENRICHMENT OF CLUSTER-1 IN GO BIOLOGICAL PROCESS TERMS

GO Term ID	GO Term Name	Adj. p-value	Score
GO:0070372	regulation of ERK1 and ERK2 cascade	0.0029	17.69
GO:0007476	*imaginal disc-derived wing morphogenesis*	0.0004	14.88
GO:1903508	positive regulation of nucleic acid-templated transcription	0.0218	9.26

TABLE VI. FLYENRICHR ENRICHMENT OF CLUSTER-2 IN GO BIOLOGICAL PROCESS TERMS

GO Term ID	GO Term Name	Adj. p-value	Score
GO:0007422	*peripheral nervous system development*	0.004	11.8
GO:0000070	mitotic sister chromatid segregation	0.004	11.4
GO:0060541	*respiratory system development*	0.006	11.0
GO:0007498	mesoderm development	0.004	10.1

D. Periodic processes are controlled by generic regulators

Table VII shows the TFs whose set of target genes have statistically significant (*i.e.,* adj. p-value ≤ 0.05) overlaps with cluster-1 (cluster-2 wasn't enriched). Among them *TFIIB* is a general transcription factor which is required for expressing any gene; so it is expected to be enriched. This finding is consistent with that of Eser et al. who reported that expressions of periodic genes tend to have significant correlation with *TFIIB* occupancy levels [31]. All of these *TFs* are the highest level regulators in the regulatory network hierarchy [32] - which suggests that these *TFs* perform non-specific regulatory roles, rather than specific ones.

TABLE VII. FLYENRICHR ENRICHMENT OF CLUSTER-1 IN SETS OF GENES CO-REGULATED BY THE SAME TRANSCRIPTION FACTOR (TF)

TF	Adj. p-value	Score	# Genes
TfIIB	0.02	13.4	6
BEAF-32	0.02	11.0	8
dl	0.02	10.1	8
Med	0.02	9.9	7
Cp190	0.02	9.4	8
twi	0.05	7.4	6
h	0.05	6.9	5

To validate this hypothesis, we sorted these *TFs* by their adj. p-values and computed the FuncAssociate enrichment of this list (Table VIII). Among the two enriched terms, *regulation of transcription, DNA-templated* is a generic process which regulates synthesis of RNA from DNA. All *7* TFs annotates this term. This result supports our hypothesis.

TABLE VIII. FUNCASSOCIATE (FA) ENRICHMENT OF SORTED (IN DESCENDING ORDER OF FR SCORE) LIST OF GENES IN THE GO BP TERMS

GO Term ID	GO Term Name	FA Adj. p-Value	FA LOD Ratio
GO:0007501	mesodermal cell fate specification	0.011	2.93
GO:0006355	regulation of transcription, DNA-templated	0	2.34

VII. Comparison

We compared our results with that of almost aperiodic subgraphs. To identify such subgraphs, we applied ListMiner on a random sequence of *90* networks among which *30* are our input networks and the rests are empty graphs (see supplement for details). We found two *(7,3)-periodic subgraphs* in these networks whereas there was no *(2,8)-periodic subgraph*. The largest component of *(7,3)-periodic subgraphs* is an edge between two genes: *cbt* and *CG2678*. This edge is present in all *30* input networks. So this result is expected but uninformative. Yet for the sake of comparison, we computed FlyEnrichr enrichment of these two genes and filtered the resulting GO terms via ReviGO (just like we did before). This analysis yields only one enriched term: *regulation of transcription, DNA-templated* (adj. p-value *0.008*). This is a broad term (annotated by > *7000* Drosophila genes) and therefore uninteresting. We didn't find any evidence in the literature for this function to be a periodic one.

VIII. Conclusion

This work is the first of its kind which shows the importance of mining periodic subgraphs in dynamic gene networks and demonstrates how to analyze those subgraphs to discover periodic processes and their regulators. It can be used for predicting potentially periodic processes and as such may act as a hypothesis generator for biological experiments thereby reducing the time and cost that would incur if all processes were tested in an attempt to find the periodic ones.

IX. Future Directions

This work may beget both computation-savvy and biology-savvy research. On the computation side, algorithms can be developed for computing periodic subgraphs in node- and/or edge-weighted networks. Computing periodic subgraphs in attributed networks is another interesting problem. On the biology side, experiments can be designed to test our predictions. Also, our pipeline can be applied to other networks (for e.g. networks of other organisms or brain networks) to find associated periodic processes.

Acknowledgment

We thank our colleague Muktadir Shahid Hossain for pointing out interesting articles related to periodic processes and for having fruitful discussions with us about this work.

References

[1] Y.-R. Cho and A. Zhang, "Identification of functional hubs and modules by converting interactome networks into hierarchical ordering of proteins," *BMC Bioinformatics*, vol. 11, no. SP. 3, 2010.

[2] H. Hu, X. Yan, Y. Huang, J. Han, and X. J. Zhou, "Mining coherent dense subgraphs across massive biological networks for functional discovery," *Bioinformatics*, vol. 21, no. S. 1, pp. 213–221, 2005.

[3] W. Li *et al.*, "Integrative analysis of many weighted Co-Expression networks using tensor computation," *PLoS Comput. Biol.*, vol. 7, no. 6, 2011.

[4] M. Lahiri and T. Y. Berger-Wolf, "Periodic subgraph mining in dynamic networks," *Knowl. Inf. Syst.*, vol. 24, no. 3, pp. 467–97, 2010.

[5] Y. Lin, X. Yuan, and B. Shen, "Network-Based Biomedical Data Analysis," Springer, Singapore, 2016, pp. 309–332.

[6] J. Chen and Y. Saad, "Dense Subgraph Extraction with Application to Community Detection," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 7, pp. 1216–1230, Jul. 2012.

[7] C. C. Aggarwal, Y. Li, and P. S. Yu, and R. Jin, "On Dense Pattern Mining in Graph Streams," *Vldb*, pp. 975–984, 2010.

[8] S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church, "Systematic determination of genetic network architecture," *Nat. Genet.*, vol. 22, no. 3, pp. 281–285, Jul. 1999.

[9] A. Goldbeter, C. Gérard, D. Gonze, J.-C. Leloup, and G. Dupont, "Systems biology of cellular rhythms," *FEBS Lett.*, vol. 586, no. 18, pp. 2955–2965, Aug. 2012.

[10] C. J. Hutchison, R. Cox, R. S. Drepaul, M. Gomperts, and C. C. Ford, "Periodic DNA synthesis in cell-free extracts of Xenopus eggs," *EMBO J.*, vol. 6, no. 7, pp. 2003–10, Jul. 1987.

[11] K. J. Ford, A. L. Felix, and M. B. Feller, "Cellular Mechanisms Underlying Spatiotemporal Features of Cholinergic Retinal Waves," *J. Neurosci.*, vol. 32, no. 3, pp. 850–863, Jan. 2012.

[12] B. T. Lobingier *et al.*, "An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells," *Cell*, vol. 169, no. 2, pp. 350-360.e12, Apr. 2017.

[13] T. Pierfelice, L. Alberi, and N. Gaiano, "Notch in the Vertebrate Nervous System: An Old Dog with New Tricks," *Neuron*, vol. 69, no. 5, pp. 840–855, Mar. 2011.

[14] B.-R. Kim, L. Zhang, A. Berg, J. Fan, and R. Wu, "A computational approach to the functional clustering of periodic gene-expression profiles," *Genetics*, vol. 180, no. 2, pp. 821–34, Oct. 2008.

[15] G. Rustici *et al.*, "Periodic gene expression program of the fission yeast cell cycle," *Nat. Genet.*, vol. 36, no. 8, pp. 809–817, Aug. 2004.

[16] A. Apostolico, M. Barbares, and C. Pizzi, "Speedup for a periodic subgraph miner," *Inf. Process. Lett.*, vol. 111, no. 11, pp. 521–523, May 2011.

[17] S. Halder, M. Samiullah, and Y.-K. Lee, "Supergraph based periodic pattern mining in dynamic social networks," *Expert Syst. Appl.*, vol. 72, no. C, pp. 430–442, Apr. 2017.

[18] L. Song, M. Kolar, and E. P. Xing, "KELLER: Estimating time-varying interactions between genes," *Bioinformatics*, vol. 25, no. 12, pp. 128–136, 2009.

[19] M. N. Arbeitman *et al.*, "Gene expression during the life cycle of Drosophila melanogaster," *Science*, vol. 297, no. 5590, pp. 2270–5, 2002.

[20] S. Ballouz, P. Pavlidis, and J. Gillis, "Using predictive specificity to determine when gene set analysis is biologically meaningful," *Nucleic Acids Res.*, vol. 45, no. 4, p. e20, 2017.

[21] M. V. Kuleshov *et al.*, "modEnrichr: a suite of gene set enrichment analysis tools for model organisms," *Nucleic Acids Res.*, vol. 47, no. W1, pp. W183–W190, 2019.

[22] G. F. Berriz, J. E. Beaver, C. Cenik, M. Tasan, and F. P. Roth, "Next generation software for functional trend analysis," *Bioinformatics*, vol. 25, no. 22, pp. 3043–3044, 2009.

[23] T. Nepusz, H. Yu, and A. Paccanaro, "Detecting overlapping protein complexes in protein-protein interaction networks," *Nat. Methods*, vol. 9, no. 5, pp. 471–472, May 2012.

[24] P. Shannon *et al.*, "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks," *Genome Res.*, vol. 13, no. 22, p. 6, 2003.

[25] F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc, "Revigo summarizes and visualizes long lists of gene ontology terms," *PLoS One*, vol. 6, no. 7, 2011.

[26] S. Lim and P. Kaldis, "Cdks, cyclins and CKIs: roles beyond cell cycle regulation," *Development*, vol. 140, no. 15, pp. 3079–93, Aug. 2013.

[27] L. Liang, J. S. Haug, C. W. Seidel, and M. C. Gibson, "Functional Genomic Analysis of the Periodic Transcriptome in the Developing Drosophila Wing," *Dev. Cell*, vol. 29, no. 1, pp. 112–127, Apr. 2014.

[28] L. L. Breeden, "Periodic Transcription: A Cycle within a Cycle," *Curr. Biol.*, vol. 13, no. 1, pp. R31–R38, Jan. 2003.

[29] H. Shankaran *et al.*, "Rapid and sustained nuclear-cytoplasmic ERK oscillations induced by epidermal growth factor," *Mol. Syst. Biol.*, vol. 5, p. 332, 2009.

[30] M. Yanagida, "Basic mechanism of eukaryotic chromosome segregation," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 360, no. 1455, pp. 609–21, Mar. 2005.

[31] P. Eser *et al.*, "Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression," *Mol. Syst. Biol.*, vol. 10, no. 1, pp. 717, Jan. 2014.

[32] S. Roy *et al.*, "Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE," *Science*, vol. 330, no. 6012, pp. 1787–1797, Dec. 2010.