Identification of Co-evolving Temporal Networks

Rasha Elhesha, Aisharjya Sarkar, Christina Boucher, and Tamer Kahveci
University of Florida, CISE Department, Gainesville, Florida 32611
{relhesha,aisharjya,cboucher,tamer}@ufl.edu

ABSTRACT
Biological networks describe the mechanisms which govern cellular functions. Temporal networks show how these networks evolve over time. Studying the temporal progression of network topologies is of utmost importance since it uncovers how a network evolves and how it resists to external stimuli and internal variations. Two temporal networks have co-evolving subnetworks if the topologies of these subnetworks remain similar to each other as the network topology evolves over a period of time. In this paper, we consider the problem of identifying co-evolving pair of temporal networks, which aim to capture the evolution of molecules and their interactions over time. Although this problem shares some characteristics of the well-known network alignment problems, it differs from existing network alignment formulations as it seeks a mapping of the two network topologies that is invariant to temporal evolution of the given networks. This is a computationally challenging problem as it requires capturing not only similar topologies between two networks but also their similar evolution patterns.

We present an efficient algorithm, Tempo, for solving identifying co-evolving subnetworks with two given temporal networks. We formally prove the correctness of our method. We experimentally demonstrate that Tempo scales efficiently with the size of network as well as the number of time points, and generates statistically significant alignments—even when evolution rates of given networks are high. Our results on a human aging dataset demonstrate that Tempo identifies novel genes contributing to the progression of Alzheimer’s, Huntington’s and Type II diabetes, while existing methods fail to do so.

CCS CONCEPTS
• Applied computing → Biological networks; Bioinformatics; Molecular evolution;

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
ACM-BCB’18, August 29-September 1, 2018, Washington, DC, USA © 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-5794-4/18/08. https://doi.org/10.1145/3233547.3233686

1 INTRODUCTION
Biological networks describe the interaction between molecules. They are frequently represented as graphs, where the nodes correspond to the molecules (e.g., proteins or genes) and the edges correspond to their interactions [12]. Formally, we denote a biological network as $G = (V, E)$ where $V$ and $E$ represent the set of nodes and the set of edges, respectively. Network alignment is one of the most challenging problem as it has a profound set of applications ranging from the detection of conserved motifs to the prediction of protein functions [3]. This problem aims to find a mapping of the nodes of two given networks in which nodes that are similar in terms of content (i.e. homology) and interaction structure (i.e. topology) are mapped to each other. Here, we represent the alignment between two given networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ as a bijection function $\psi : V_1 \rightarrow V_2$, and the score resulting from alignment $\psi$ as $\text{score}(G_1, G_2)$. The network alignment problem seeks the function $\psi$ that maximizes this score.

Biological networks have dynamic topologies [6]. There are various reasons behind this dynamic behavior. For example, genetic and epigenetic mutations can alter molecular interactions. Due to this dynamic behavior, the topology of the network that models the molecular interaction evolve over time [6]. We define a biological network using a model that accounts for the evolution of the underlying network at consecutive time points. We refer to this model as a temporal network. We view this model as containing a single snapshot of the network at each time point in a sequence of time points and thus, as a time series network. More formally, we denote a temporal network with $t$ consecutive time points as $G_t = [G_1, G_2, \ldots, G_t]$, where $G_t = (V, E_t)$ represents the topology of the network at the $t$th time point.

In this paper, we consider the problem of identifying co-evolving subnetworks in a given pair of temporal networks. We say that two subnetworks are co-evolving if their topologies remain similar even though their topologies evolve over time. We define this more formally as follows. We consider two input temporal networks $G_1 = \{G_1^1, G_1^2, \ldots, G_1^t\}$ and $G_2 = \{G_2^1, G_2^2, \ldots, G_2^t\}$, where $G_i^t \in \{G_1^1, G_1^2, \ldots, G_1^t\}$ and represent $G_2^t$ respectively at the time point $i$. Without losing generality, let $G_1^1$ be the smaller network and $G_2^t$ be the target network, i.e., $|V_1^1| \leq |V_2^t|$. An alignment of $G_1^1$ and $G_2^t$ maps $G_1^1$ to $G_2^t$ across all time points $i$. Thus, we represent the alignment of the two temporal networks $G_1^t$ and $G_2^t$ as a bijection of their nodes and denote it as a function $\psi : V_1^1 \rightarrow V_2^t$. We compute the score of the alignment $\psi$ of $G_1^1$ and $G_2^t$, denoted with $\text{score}(G_1^1, G_2^t; \psi)$, as the sum of the scores of the alignment at all time points. Hence, $\text{score}(G_1^1, G_2^t; \psi) = \sum_{i=1}^{t} \text{score}(G_1^i, G_2^i; \psi)$.

Contributions in this paper. We develop an efficient algorithm, Tempo, to identify co-evolving subnetworks in a given pair of the temporal networks. Briefly, our algorithm first finds an initial alignment between the input networks $G_1^1$ and $G_2^t$ using the similarity score between pairs of aligned nodes across all time points. It then performs a dynamic programming strategy that maximizes the alignment quality (i.e. score) by repeatedly altering the aligned nodes in the target network. We demonstrate the accuracy of Tempo using real data. We compare quality of the alignments found by Tempo against those of three existing alignment algorithms MAGNA++ [11] and GHOST [10]. Note that all these networks are tailored towards optimizing alignment at a single time point. To have a fair comparison, we allow each of these methods to consider each time point independently then apply the resulting alignments to all other time points and took the average.
2 RESULTS

We obtain our real dataset from two sources. The first one is the human brain aging dataset [2]. Recall that this dataset contains gene expressions of 173 samples obtained from 55 individuals spanning 37 ages from 20 to 99 years. The ages in this dataset are not uniformly spaced. In order to bring consecutive time gaps to a more uniform values, we remove two data points which have an age gap of more than 5 years from their successive age values, leading to 35 ages. We select five temporal networks each having seven time points. Next, we explain how we do that for the first temporal network. We start with the first (i.e., youngest) time point in the aging data. We then skip the next four time points and take the sixth time point in aging data iteratively until we have seven time points. Similarly, for $1 < j \leq 5$, we select the $j$th temporal network starting from the $j$th time point. In this manner, we form five non-overlapping and interleaved temporal networks. In order to integrate static PPI network with gene expression data to form age-specific PPI networks, we set a cut-off on the gene-expression value. All the interactions that have a lower transcription value for either or both the proteins are removed from the corresponding age-specific network. For the second source, we select phenotype specific query temporal networks from this dataset. We use two neurodegenerative disorders which are conjectured to be age-related (Alzheimer’s and Huntington’s) and a third one which we expect to be less prone to aging (Type II diabetes). We retrieve the gene sets specific to these three diseases from KEGG database [9]. We form three query PPI temporal networks by keeping only the interactions where both the interactors are from each of the three phenotype-specific (Alzheimer’s, Huntington’s or Type II Diabetes) gene set.

We evaluate Tempo based on the quality of its alignment as follows. For each combination of disease and mapped gene pairs identified by Tempo, we first search PubMed for publication evidence specific to that disease. For instance, in case of Alzheimer’s disease, the gene DAB1 that was selected by Tempo and was identified as a potential gene that encode proteins related to functions in biological pathways relevant to the disease [5]. We also found similar results for type II diabetes.

In order to determine the biological processes of the aligned genes found by Tempo in gene aging dataset compared to other methods, we perform the gene ontology analysis of the aligned genes in target network using Gene Ontology Consortium [4]. We identify the biological processes or signaling pathways that play significant roles in the disorder. We calculate how many related pathways found by our method (Tempo) against MAGNA and GHOST and their significance. We also counted the frequency of those pathways when used different range of time points. Table 1 present the results. We find references of certain pathways that are related to specific neurodegenerative disorders (Alzheimer’s and Huntington’s diseases). For genes we identify when we use Alzheimer’s disease as a query network, we find two pathways, namely Alzheimer disease amyloid secretase and Alzheimer disease presenilin are related to Alzheimer’s disease [8]. Various growth factors alter the brain development process at younger age, that manifest as a variety of risk factors at an older age and eventually results in aging-related diseases such as Alzheimer’s and Huntington’s diseases [1]. For the genes we identify when we use type II diabetes phenotype as a query, we find two pathways that they are commonly associated with type II diabetes [7] namely Insulin/IGF pathway-protein kinase B signaling cascade and Insulin/IGF pathway-mitogen activated protein kinase kinase/ MAP kinase cascade. On the other hand, MAGNA or GHOST found at most one pathway with very low significance

| Disease       | Tempo | MAGNA++ | GHOST |
|---------------|-------|---------|-------|
| Alzheimer     | 2 / 4 | 2.29E-14| 1 / 2 | 2.14E-03 | 1 / 2 | 3.32E-04 |
| Huntington’s  | 1 / 4 | 1.15E-22| 0     | 0       |       |
| Diabetes      | 2 / 4 | 2.29E-09| 1 / 2 | 2.2E-01 | 0     |

Table 1: Number and significance of functional pathways associated with the underlying disease observed among the aligned genes of target network. Each cell lists the results in the form x/y/z. Here, x represents number of pathways identified, y denotes the number of time points at which these pathways are observed, and z is the statistical significance (p-value) of the least significant of these pathways. The cell with the value 0 implies that no pathways were found.

and did not appear through all tested target networks (see Table 1). In conclusion, studying temporal networks in general and human aging specifically using Tempo enables us to identify age related genes from non age related genes successfully. More importantly, Tempo takes the network alignment problem one huge step forward by moving beyond the classical static network models.

3 CONCLUSION
In this paper, we modeled the problem of network alignment between two given temporal networks. We developed a novel method to solve this problem by optimizing the alignment score and generating a persist alignment through all time points. Our algorithm incorporates a dynamic programming approach which iteratively refines the alignment to monotonically increase the alignment score. We adapted MAGNA++ and GHOST which are used for pairwise static network alignment, to align two temporal networks by aligning snapshots at each time point independently. We compare the quality of the resulting alignment of both our method and other methods. Our experimental results on human aging dataset using gene ontology analysis on aligned gene pairs suggest that our method could successfully align genes from target network that are significantly related to genes of the underlying query phenotype unlike other methods which failed to do so.

REFERENCES

[1] G. Bartkoski. 2004. Age-related myelin breakdown: a developmental model of cognitive decline and Alzheimer’s disease. Neurobiology of Aging 25, 1 (2004), 5–18.
[2] N. C. Berchtold, D. H. Crihbs, P. D. Coleman, J. Rogers, E. Head, R. Kim, T. Beach, C. Miller, J. Troncoso, J. Q. Trojanowski, et al. 2008. Gene expression changes in the course of normal brain aging are sexually dimorphic. Proceedings of the National Academy of Sciences 105, 40 (2008), 15605–15610.
[3] J. F. Clemente, K. Satou, and G. Valiente. 2006. Finding conserved and non-conserved reactions using a metabolic pathway alignment algorithm. Genome Informatics 17, 2 (2006), 46–56.
[4] Gene Ontology Consortium et al. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research 32, suppl 1 (2004), D258–D262.
[5] H. Gao, Y. Tao, Q. He, F. Song, and D. Saffen. 2015. Functional enrichment analysis of three Alzheimer’s disease genome-wide association studies identifies DAB1 as a novel candidate liability/protective gene. Biochemical and Biophysical Research Communications 463, 4 (2015), 490–495.
[6] P. Holme and J. Saramäki. 2012. Temporal networks. Physics Reports 519, 3 (2012), 97–125.
[7] M. Liu, A. Libenzon, S. Kong, W. R. Lai, P. J. Park, I. S. Kohane, and S. Kasif. 2007. Network-based analysis of affected biological processes in type 2 diabetes models. PLoS Genetics 3, 6 (2007), e96.
[8] M. P. Mattsson. 2004. Pathways towards and away from Alzheimer’s disease. Nature 430, 7000 (2004), 631–639.
[9] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research 27, 1 (1999), 29–34.
[10] R. Patro and C. Kingsford. 2012. Global network alignment using multiscale spectral signatures. Bioinformatics 28, 23 (2012), 3105–3114.
[11] V. Vijayan, V. Saraph, and T. Mälenkovic. 2015. MAGNA++: Maximizing Accuracy in Global Network Alignment via both node and edge conservation. Bioinformatics 31, 14 (2015), 2409–2411.
[12] X. Zhu, M. Gerstein, and M. Snyder. 2007. Getting connected: analysis and principles of biological networks. Genes & Development 21, 9 (2007), 1010–1024.