Inference of dynamic networks using time-course data

Yongsoo Kim*, Seungmin Han*, Seungjin Choi and Daehee Hwang

Submitted: 22nd January 2013; Received (in revised form): 12th March 2013

Abstract

Cells execute their functions through dynamic operations of biological networks. Dynamic networks delineate the operation of biological networks in terms of temporal changes of abundances or activities of nodes (proteins and RNAs), as well as formation of new edges and disappearance of existing edges over time. Global genomic and proteomic technologies can be used to decode dynamic networks. However, using these experimental methods, it is still challenging to identify temporal transition of nodes and edges. Thus, several computational methods for estimating dynamic topological and functional characteristics of networks have been introduced. In this review, we summarize concepts and applications of these computational methods for inferring dynamic networks and further summarize methods for estimating spatial transition of biological networks.

Keywords: dynamic network; spatiotemporal dynamics; network inference

INTRODUCTION

Dynamic operations of biological networks involve (i) changes in abundances or activities of nodes (e.g. protein, RNA and DNA) and (ii) formation of new edges and disappearance of existing edges over time. To measure the temporal changes of nodes and edges, global assay techniques have been used. Genomic technologies, such as microarrays and mass spectrometry-based methods, permit to measure temporal changes in abundances of nodes. Furthermore, several methods, such as mass spectrometry-based tandem affinity purification (TAP [1]) and chromatin immunoprecipitation-sequencing (ChIP-seq) [2], can be used to measure temporal transitions of protein–protein interactions (PPIs) and protein–DNA interactions (PDIs).

Time-course genomic analysis (e.g. gene expression profiling) to experimentally measure temporal transitions of nodes has been a common practice. Biological networks describing the temporal transitions of nodes measured by the time-course genomic analysis can be reconstructed using PPIs and PDIs deposited in interactome databases, such as Human Protein Reference Database (HPRD) [3], Biological General Repository for interaction datasets (BioGRID) [4] and Biomolecular Interaction Network Database (BIND) [5]. However, these interactomes include false-positive and -negative interactions [6]. Thus, the biological networks reconstructed using these interactomes are limited to describe temporal transitions of interactions (i.e. formation of new edges or disappearance of existing edges over time). Time-course interactome analysis could be performed to experimentally measure the temporal transitions of edges. However, they provide only limited coverage of the interactomes to decode temporal transitions of individual edges at each time point [6, 7].

Corresponding author. Daehee Hwang. School of Interdisciplinary Bioscience and Bioengineering, POSTECH, Pohang, 790-784, Republic of Korea. Tel.: 82-54-279-2393; Fax: 82-54-279-8409; E-mail: dhhwang@postech.ac.kr

*These authors equally contributed to this work.

Yongssoo Kim is a graduate student in School of Interdisciplinary Bioscience and Bioengineering at POSTECH. He has been developing machine-learning approaches for network analyses using various omics data.

Seungmin Han is a graduate student in School of Interdisciplinary Bioscience and Bioengineering at POSTECH. He has been studying design principles of biological networks via mathematical modeling approaches.

Seungjin Choi is a professor in Department of Computer Science and Engineering and Division of IT Convergence Engineering at POSTECH. He leads the research group on machine learning.

Daehee Hwang is an associate professor in School of Interdisciplinary Bioscience and Bioengineering and Department of Chemical Engineering at POSTECH. He leads the research group on systems biology and medicine.
Because of these limitations in the experimental methods and huge amounts of experimental costs, a number of computational methods for inferring dynamic networks using time-course global data have been developed (Table 1). These methods rely on statistical dependence of measured data for inferring active nodes and/or edges: they identify nodes showing expression changes over time and infer edges between pairs of the nodes with statistical dependence in temporal expression changes. Dynamic networks estimated by these methods can be represented as: (i) a single network in which an edge represents a regulatory relationship between a pair of nodes across the whole time span and (ii) a set of networks each of which represents activation of nodes and/or edges in each time point or each segment of the time span, thereby delineating temporal transitions of nodes and/or edges.

In this review, we summarize computational methods for inferring dynamic networks using time-course global data. These methods can be categorized based on characteristics of the resulting networks and the strategies used in these methods: (i) whether the resulting networks delineate temporal transition of nodes and/or edges inferred from time-course data (columns in Figure 1), (ii) whether the edges in the resulting networks are directed or undirected (rows in the left of Figure 1) and (iii) whether the inference methods integrate information other than time-course gene expression data (e.g. GO terms or ChIP-chip; rows in the right of Figure 1). Five types of methods have been often used to reconstruct a dynamic network for time-course data: (i) Bayesian network (BN) [8], (ii) relevance network (RN) [9], (iii) Markov Random Field (MRF) [10], (iv) ordinary differential equations (ODE) [11] and (v) logic-based models [12]. BN, RN and MRF-based methods identify the presence of the interactions between the molecules in the network (i.e. network structure). ODE [11] and logic-based models [12], in addition to the network structure, further identify the parameters (i.e. reaction parameters for ODEs and Boolean logic gates for logic-based models) for the interactions, permitting to simulate network outputs under external perturbations. However, they have a limitation to infer a dynamic network with a large number of nodes. The inference of moderately large dynamic networks is more feasible by BN, RN and MRF-based methods than by ODE and logic-based models. Here, we thus focus on the approaches to infer moderately large dynamic networks from time-course global data.

### Table 1: Publicly available software

<table>
<thead>
<tr>
<th>Name of Software</th>
<th>Description</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minet [82]</td>
<td>R package that implements algorithms based on mutual information, including ARACNE [29] and MRNET [32].</td>
<td><a href="http://bioconductor.org/packages/2.3/bioc/html/minet.html">http://bioconductor.org/packages/2.3/bioc/html/minet.html</a></td>
</tr>
</tbody>
</table>
Many eukaryotic proteins, up to 35%, are localized in multiple compartments [13], leading to interactions with distinct partners in different compartments. Unfolding dynamic networks further into intracellular organelles (e.g. nucleus, cytosol or mitochondria) can provide additional insights into the operation of biological networks. Several experimental approaches and computational methods for inferring spatial transition of nodes and edges in biological networks have been developed. Thus, we also summarize these methods for inferring spatiotemporal operations of biological networks. The abbreviations for technical terminologies used in this review are summarized in Table 2.

**NETWORKS WITH TEMPORAL ASSOCIATIONS**

In the simplest dynamic network, an edge represents a regulatory relationship between a pair of nodes (e.g. genes and proteins) across the whole time span (‘Network with Temporal Associations’ in Figure 1). Several methods have been developed for estimating the regulatory network structure.
using time-course gene expression data. These methods evaluate statistical dependence between the nodes throughout the whole time span using the time-course gene expression data, resulting in a single network with temporal regulatory relationships [14]. These methods include BN, dynamic BN (DBN) and RN methods, which have been applied to identify transcriptional regulatory networks (TRNs) and gene interaction networks (GINs). In this review, we used GINs to indicate the networks, including the links estimated based on the correlation in time-course gene expression data, which are not transcriptional factor (TF)-target interactions in TRNs or genetic relationships in genetic regulatory networks.

BN methods have been developed to infer regulatory relationships between the genes using time-course gene expression data [8]. BN is a probabilistic graphical model that represents conditional dependence among the nodes using directed acyclic graphs (DAGs). BN methods infer static and directed edges that represent active regulatory relationships between the nodes across the whole time span (Bayesian network in Figure 2). For example, Friedman et al. [8] introduced a BN method to reconstruct a GIN for 800 genes with expression changes using time-course gene expression data measured at different phases (M/G1, G1, S, G2 and M) during cell cycle of *Saccharomyces cerevisiae* [15]. The network revealed dynamic activation of diverse modes of transcriptional regulation (e.g. transcriptional cascades or feedforward loops) during the cell cycle. Although the BN methods can infer temporal regulatory associations, they do not take into account the temporal order in the data [i.e. observed gene expression values are assumed to be independent and identically distributed (iid)]. Therefore, temporal associations involving time-delays among the genes are not likely to be detected by these methods. Also, these methods use a DAG as the structure of BN network, and thus cannot detect cyclic regulatory relationships among the nodes, such as feedback loops.

DBN methods [16–18] were developed to use the temporal order of the time-course data in inferring regulatory relationships between the nodes. The DBN method developed by Friedman et al. [16] constructs prior and transition networks (Dynamic Bayesian network in Figure 2). The prior network represents the statistical dependence among the nodes at the initial time point, and the transition network represents the dependence between the nodes at $t$ and $t+1$. Unlike BNs, the transition network can include cyclic regulatory relationships among the nodes (see feedback loops in the inferred network structure of Figure 2). Kim et al. [19] applied BN and DBN methods to the aforementioned time-course gene expression data generated during cell cycle of *S. cerevisiae* [15] and then compared the resulting GINs. They found that the DBN method, compared with the BN method, generated a GIN with less false-positive interactions according to KEGG cell cycle pathways. A major drawback of both BN and DBN methods is that the search space for candidate network structures exponentially increases as the number of the nodes in the networks increases. Thus, BN and DBN methods are used to estimate the networks with only a small number of nodes. Recently, to resolve this problem, efficient algorithms and tools, such as

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full name</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>ARTIVA</td>
<td>Auto regressive time varying models</td>
</tr>
<tr>
<td>BN</td>
<td>Bayesian network</td>
</tr>
<tr>
<td>CC</td>
<td>Correlation coefficient</td>
</tr>
<tr>
<td>DAG</td>
<td>Directed acyclic graph</td>
</tr>
<tr>
<td>DBN</td>
<td>Dynamic Bayesian network</td>
</tr>
<tr>
<td>DEG</td>
<td>Differentially expressed gene</td>
</tr>
<tr>
<td>DPI</td>
<td>Data processing inequality</td>
</tr>
<tr>
<td>GIN</td>
<td>Gene interaction network</td>
</tr>
<tr>
<td>GO</td>
<td>Gene ontology</td>
</tr>
<tr>
<td>GOBP</td>
<td>Gene ontology biological process</td>
</tr>
<tr>
<td>GOCC</td>
<td>Gene ontology cellular compartment</td>
</tr>
<tr>
<td>GOMF</td>
<td>Gene ontology molecular function</td>
</tr>
<tr>
<td>KELLER</td>
<td>Kernel-re-weighted logistic regression</td>
</tr>
<tr>
<td>MI</td>
<td>Mutual information</td>
</tr>
<tr>
<td>MRFS</td>
<td>Markov Random Field</td>
</tr>
<tr>
<td>nhDBN</td>
<td>Non-homogenous dynamic Bayesian network</td>
</tr>
<tr>
<td>nsDBN</td>
<td>Non-stationary dynamic Bayesian network</td>
</tr>
<tr>
<td>ONMF</td>
<td>Orthogonal non-negative matrix factorization</td>
</tr>
<tr>
<td>PDI</td>
<td>Protein–DNA interaction</td>
</tr>
<tr>
<td>PMI</td>
<td>Protein-metabolite interaction</td>
</tr>
<tr>
<td>PNA</td>
<td>Principal network analysis</td>
</tr>
<tr>
<td>PPI</td>
<td>Protein–protein interaction</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>RN</td>
<td>Relevance network</td>
</tr>
<tr>
<td>SANDY</td>
<td>Statistical analysis of network dynamics</td>
</tr>
<tr>
<td>TAP</td>
<td>Tandem affinity purification</td>
</tr>
<tr>
<td>TESLA</td>
<td>Temporally smoothed l1-regularized logistic regression</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TFBS</td>
<td>Transcription factor-binding site</td>
</tr>
<tr>
<td>TRN</td>
<td>Transcriptional regulatory network</td>
</tr>
<tr>
<td>TV-DBN</td>
<td>Time-varying dynamic Bayesian network</td>
</tr>
<tr>
<td>Y2H</td>
<td>Yeast two-hybrid</td>
</tr>
</tbody>
</table>
GlobalMIT [20] and BNFinder [21] have been developed [20–23]. RN methods [9, 24, 25] have been also used to reconstruct GINs [9, 24] and TRNs [25] using time-course gene expression data. For the inference, they use relevance measures for pairs of the nodes (Relevance network in Figure 2), such as correlation coefficients (CC) [9] and mutual information (MI).
For example, Remondini et al. [9] generated time-course gene expression data at five time points (1, 2, 4, 8 and 16 h) from rat fibroblast cells conditionally expressing Myc-estrogen receptor oncoprotein. They estimated temporal associations among genes based on their correlation of time-course gene expression and then reconstructed a GIN using the estimated associations. Using the network, they identified transregulation cascades among c-Myc and its 130 target genes. Unlike the BN and DBN methods, RN methods can effectively reconstruct the networks with large numbers of nodes. However, both MI and CC cannot discriminate indirect links from direct ones, thus leading to a large number of false-positive interactions. To reduce the false positives, several RN methods used post-processing after constructing relevance networks [26–28]. In the algorithm for the reconstruction of accurate cellular networks (ARACNE; [29]), false-positive links are reduced by removing indirect links using data processing inequality (DPI), a criterion that can discriminate direct links from indirect ones. In other RN methods, alternative association measures, such as partial correlation coefficient (ParCorA [30] and GeneNet [31]) and conditional mutual information (MRNET [32]), have been used to sort out direct associations from indirect ones. In particular, Gaussian graphical model (GGM) is an undirected graphical model to estimate the conditional dependence between nodes based on partial correlation coefficient. Efficient algorithms for the inference using GGM have been developed [33–35].

DYNAMIC NETWORKS WITH TEMPORAL TRANSITION OF NODES

All the aforementioned methods estimate the networks with temporal regulatory relationships (first row in Figure 2). However, these networks do not provide how nodes and edges become active at individual time points and how their activities evolve over time. Several approaches (Dynamic networks with temporal transition of nodes in Figure 2) have been introduced to reconstruct dynamic networks to overcome these drawbacks. These approaches have been commonly applied to identify dynamic networks, including PPIs and PDIs. These methods first identify differentially expressed genes (DEGs) over time and then clustered the DEGs based on their differential expression patterns or other criteria, such as functional similarity of the genes. In each time point, a network is then reconstructed for the genes in the clusters by linking the genes with known interactions (e.g. PPIs and PDIs) from public databases, such as HPRD [3], BioGRID [4] and BIND [5]. Fold-changes of the genes are used to represent activation of the nodes in the network at each time point (transitional networks in Figure 2). The edges represent potential interactions among the nodes but not necessarily the statistical dependence between the nodes. The networks at multiple time points can provide temporal activation of the nodes and their transition over time.

Luscombe et al. [36] reconstructed dynamic TRNs in S. cerevisiae using this approach. They identified 455 genes showing expression changes over time using time-course gene expression data generated during the cell cycle and then grouped them into five phase-specific clusters (early and late G1, S, G2 and M phases). Dynamic TRNs were constructed for 70 TFs and their target genes in the five clusters. The TRNs revealed temporal activation of the TFs and their targets at the five cell cycle phases and also their transition along the cell cycle. Also, using this approach, Hwang et al. [37] identified 923 genes showing expression changes along the progression of prion diseases, categorized them into four groups representing PrPSc accumulation, microglial/astrocytic activation, synaptic degeneration and neuronal cell death, and then reconstructed dynamic networks delineating temporal activation of the genes in individual groups during the progression of prion diseases. Moreover, this approach was applied to time-course phosphoproteomic profiles to generate dynamic signaling networks. To monitor temporal dynamics of phosphoproteome on growth factor stimulation, Olsen et al. [38] detected 6600 phosphorylation sites on 2244 proteins and measured the changes in their abundance at 1, 5, 10 and 20 min after treating HeLa cells with epidermal growth factor (EGF). Then, they clustered 1046 phosphopeptides regulated by the EGF signaling into seven functional protein classes based on the temporal transition patterns in their abundance. They then reconstructed dynamic signaling networks that delineate transition of phosphopeptides belonging to the individual protein classes and thus dynamic activation of cellular functions associated with the protein classes upon the EGF stimulation.
In these studies, the procedures for identification and clustering of DEGs and network modeling were performed separately using different tools. To automate these procedures, Kim et al. [39] developed a framework, called principal network analysis (PNA) that can perform these procedures at once using orthogonal non-negative matrix factorization (ONMF). By applying the ONMF to time-course gene expression data, PNA identifies differential expression patterns and then selects not only genes showing the differential expression patterns but also edges linking the genes that show the similar expression patterns among all the known PPIs or PDIs. Using these selected nodes and edges, PNA automatically generates a subnetwork describing both nodes and edges showing each differential expression pattern. Furthermore, PNA combines subnetworks for multiple differential expression patterns. These combined subnetworks provide new insights into temporal activation of the subnetworks. PNA was applied to the time-course gene expression data collected during the progression of prion diseases. The resulting networks revealed that the genes related to PrPSc accumulation and microglial/astrocytic activation were early activated and densely connected, reflecting activation of immune responses in microglia and astrocytes caused by PrPSc accumulation and thus coordinated activation of these cellular processes in the early stage of prion diseases.

**DYNAMIC NETWORKS WITH TEMPORAL TRANSITION OF EDGES**

The aforementioned methods identify only temporal transition of the nodes from the time-course gene expression data and link the nodes using the known PPI and PDI interactomes. Thus, the resulting networks can include false-positive edges that are not active under conditions being investigated. However, on environmental stimuli, new sets of genes and proteins are induced or repressed over time. Proteins can also move from one compartment to another on their activation and change their interacting partners at the two different compartments. These dynamic changes in localization of proteins lead to formation of new edges and disappearance of existing edges. Thus, inference of the transition of the edges reflecting the above situations can provide new insights into dynamic operation of biological networks. Several methods to infer temporal activation of edges and their transition over time have been developed (dynamic network with temporal transition of edges in Figure 2). These methods include variants of DBNs, such as non-stationary BN (nsDBN) and time-varying DBN (TV-DBN) and Markov Random Field (MRF)-based methods, which have been applied to identify GINs and TRNs.

Variants of DBNs have been developed for inferring temporal transition of edges in dynamic GINs from time-course gene expression data [40–43]. For example, non-stationary Bayesian network (nsDBN) developed by Jia and Huan [41] decomposes the whole time span into several segments of the time span and then reconstructs a DBN within each time segment, assuming that the network structure for the time points within the time segment is the same (nsDBN in Figure 2). Using this nsDBN, Jia and Huan [41] reconstructed time-varying GINs from three sets of time-course gene expression data measured from macrophage, *Arabidopsis* and *Drosophila*. Another extension of DBN is time-varying DBN (TV-DBN) [42]. Unlike nsDBN involving segmentation of the whole time span, TV-DBN estimates a network structure for a fixed set of nodes at each time point using a kernel re-weighted auto-regressive method. For continuous transitions of network structures, this method penalizes gene expression data from distant time points more than those from close time points using kernel function. Dimitrakopoulou et al. [44] used TV-DBN to reconstruct dynamic GINs from time-course data measured at every day after the infection of mouse with influenza A virus up to 5 days. They first identified 3500 DEGs and clustered them into 35 groups using k-means clustering. To reduce the complexity, they then reconstructed TV-DBNs for the 35 groups using the mean gene expression profiles in the groups. The resulting dynamic networks revealed that the interactions among the groups including early innate immunity-related genes (*Tlr1* and *Tlr2* and *Tlr6*) were highly activated until 3 days, whereas interactions among the groups including the outputs of the early innate immunity (*Ifnb1* and *Ifnar2*) showed sustained activation up to 5 days. Thus, these examples showed that nsDBN and TV-DBN can be used to effectively model temporal activation and transition of nodes and edges during the course of cellular events. However, these methods require discretization of gene expression data, leading to loss of information. In contrast, auto regressive time varying...
models (ARTIVA) [45] can estimate time-evolving networks without the discretization of gene expression data. ARTIVA was used to reconstruct GINs using two sets of time-course expression data collected during the developmental stages of *D. melanogaster* and the response of *S. cerevisiae* to benomyl poisoning. These applications demonstrated that ARTIVA can recover essential temporal dependencies between the genes.

MRF-based methods [10, 46] have been also developed to reconstruct dynamic GINs using time-course gene expression data. Unlike BNs that use DAGs, MRF methods use an undirected probabilistic graphical model that represents statistical dependence for the nodes, thus permitting to infer cyclic regulatory relationships (MRF network in Figure 2). Among the MRF methods, temporally smoothed l1-regularized logistic regression (TESLA [46]; total variation in Kolar *et al.* [47]) and kernel-reweighted logistic regression (KELLER [10]; smooth in Kolar *et al.* [47]) use binary information (e.g. activation or non-activation) of gene expression data to reconstruct dynamic networks by logistic regression. These methods commonly use l1-regularized logistic regression formulation. However, they use different methods to achieve smooth transition of dynamic networks over time. TESLA minimizes total variation of interactions in the networks over time to generate similar structures in dynamic networks at adjacent time points. On the other hand, KELLER uses kernel weight function for the likelihood being maximized for network inference, thus generating the network structures with smooth transition at adjacent time points. Song *et al.* [10] applied KELLER to time-course gene expression data measured during the developmental course of *D. melanogaster* and generated GINs of 588 genes at 66 time points during the four developmental stages: embryonic, larval, pupal and adult stages. The resulting 66 time-evolving networks provided activation of edges among the genes at the four stages and their transitions along the stages. For example, the interaction between misshapen (*msn*) and dreadlocks (*dock*) was activated during embryonic stage, but then disappeared in the following stages [48]. Moreover, 30 and 27 genes related to embryonic and post-embryonic developments, respectively, have highest interactivities among themselves at the corresponding stages. This result implies that the time-evolving networks well-reflect temporal activation and transition of nodes and edges during the different developmental stages.

These methods to infer dynamic networks with temporal transition of edges use a larger number of parameters to be optimized than those for networks with temporal associations and temporal transition of nodes, leading to an increased search space of network structures and also often to difficulties in identifying the optimal network structures. Thus, these methods suffer from a large number of candidate structures of dynamic networks for a time-course gene expression data set. To overcome this drawback, efficient optimization algorithms or additional constraints on network structure (e.g. integration of biological knowledge from literatures) are needed.

### DIRECTED AND UNDIRECTED DYNAMIC NETWORKS

Dynamic networks inferred from time-course gene expression data can be further categorized into directed and undirected networks, according to whether edges in the network are directed or undirected (rows in the left of Figure 1). The directionality of edges in the network can be interpreted as the causality in interactions between nodes. The causal relationships may produce additional insights into regulatory mechanisms in dynamic networks, such as signaling or transcriptional cascades, and feedback and feedforward loops. The methods based on probabilistic graphical models reconstruct either directed or undirected networks according to their graphical representations: BN methods (BN, DBN, nsDBN, TV-DBN, nhDBN and ARTIVA) generate directed networks, whereas MRF methods (TESLA and KELLER) generate undirected networks.

RN methods generate undirected dynamic networks because of the non-directional nature of the relevance measures (i.e. CC and MI). To infer causal relationships between the nodes, several RN methods, such as a pattern recognition approach [49] and TimeDelay-ARACNE [50], used modified relevance measures to incorporate time-delay [26, 49–51]. For example, time-lagged correlation, a modified Pearson correlation, was used to maximally capture the correlation in the time-course data [51]. The causal relationships inferred using this time-lagged correlation result in a directed GIN that includes cyclic regulatory relationships like the original GIN. Also, Zhao *et al.* [26] used a time-lagged MI to infer a directed dynamic network that describes muscle development during the life cycle of *Drosophila*. In addition, Yuan *et al.* [52]
INTEGRATIVE INFERENCE OF DYNAMIC NETWORKS

Estimation of a dynamic network from time-course gene expression data is an under-determined problem. Thus, there are a number of candidate network structures that can generate observed time-course gene expression data. When there are a large number of candidate network structures (e.g. when temporal transition of edges is estimated), other data complementary to time-course gene expression data can be integrated to reduce the number of possible structures, leading to accurate inference of dynamic networks. Various types of the complementary data have been used (rows in the right of Figure 1). For example, known PPIs and PDIs interactomes are used as prior information in statistical inference of the conditional dependence between the nodes. Furthermore, in silico predicted PDI data, such as TF-binding sites (TFBSs) predicted from motif scan methods, and experimentally measured PDI data, such as ChIP-chip or ChIP-seq data, can be integrated to infer reliable edges between TFs and their target genes in dynamic TRNs.

These data have been integrated to the models used to estimate dynamic networks in various manners to reduce the search space of network structures. BN methods cannot be practically applicable for a large number of genes. Thus, several BN methods (e.g. DBN, nSDBN and TV-DBN) integrated gene-function information to select the genes (nodes) to be included in dynamic networks. For example, Ong et al. [18] selected 15 genes and 9 operons involved in tryptophan metabolism in Escherichia coli and developed a tryptophan metabolism-related dynamic network describing regulatory relationships among these selected operons and genes. Alternatively, the methods for inferring dynamic networks with temporal transition of nodes (i.e. DEGs) integrated known interactions to reduce the edge search space into the known interactomes used [36, 37, 39]. This approach can reduce false-positive interactions that result from the methods to estimate the edges among all possible pairs of DEGs (e.g. TV-DBN and KELLER). PNA selects only an active subset from the known interactomes to reduce false-positive interactions that could not be activated based on time-course gene expression data.

Furthermore, previous knowledge on the system can be integrated at the model construction stage. Ong et al. [18] integrated an operon map that represents co-regulated clusters (operons) of genes, with time-course gene expression data measured from E. coli. They modeled a dynamic network in which operons were used as nodes, edges were inferred to link the operon nodes and the co-regulated genes were used as subnodes that are connected to the corresponding operon node according to the predefined operon map. The integration of the operon map led to reduction in the search space of network structures because of the smaller number of operons than that of genes. Furthermore, the resulting dynamic network correctly captured known interactions between the genes involved in tryptophan metabolism in E. coli, through the links between the operons including the genes, which could not be inferred without integration of the operon map.

Moreover, the data can be integrated into the inference algorithms, in addition to the model as described earlier in the text. For example, a BN method developed by Tamada et al. [53] integrated promoter sequences. This method first estimates an initial network using gene expression profiles obtained after disruption of 100 transcriptional regulators in S. cerevisiae. Based on this initial network, they detected consensus motifs in promoter sequences of target genes regulated by common transcriptional regulators. Using the consensus motifs, the method re-estimated the network. The procedure is repeated until the estimated network structure is converged. The comparison of inferred networks with and
without integration of promoter sequences revealed that the integration of promoter sequences resulted in a more accurate network than that inferred without the integration. For instance, without the integration, GAL2 was inferred as an upstream regulator of GAL11. This erroneous link was corrected in the network inferred with the integration of promoter sequences. Therefore, this example demonstrates that integration of the data to both the model and the inference algorithm can lead to more accurate inference of dynamic networks [53].

FUNCTIONAL INTERPRETATION OF DYNAMIC NETWORKS

Dynamic networks from time-course gene expression data provide the bases for cellular processes sequentially activated and regulatory relationships for their sequential activation (e.g. signaling or transcriptional cascades and feedback or feedforward loops). Thus, interpretation of dynamic networks estimated from the aforementioned methods is essential to understand temporal activation of the cellular processes and also mechanisms underlying their temporal activation. Several strategies have been introduced for functional interpretation of dynamic networks. Most of these strategies are based on GO analysis of the nodes in the dynamic networks. These methods typically identify GO terms enriched by the nodes in a network constructed at each time point and then compare the enriched GO terms at consecutive time points to understand temporal transitions of cellular processes associated with the GO terms (Node-based ontology analysis in Figure 3A). For example, Song et al. [10] identified GO terms related to developmental processes (e.g. muscle development and wing vein morphogenesis) for the nodes in networks at all time points and then identified interactions of these GO terms at individual time points. Each pair of two GO terms is connected when the total number of links between the genes assigned with the two GO terms in the network at each time point exceeds a cut-off value (Figure 3B). The temporal transition of the interactions between the GO terms represents activation of their associated cellular processes (individual GO terms) and their collective activation (a set of GO terms) over time. Temporal transition of cellular processes is defined not only by nodes (genes or proteins) but also by their physical/functional interactions [54, 55]. The node-based GO analysis is difficult to distinguish the difference between cellular functions defined by the networks with the same list of nodes, but with different sets of edges. This raises the necessity to consider both nodes and edges for interpretation of dynamic networks.

To address this issue, a recent study developed an edge-based GO analysis method [56]. This analysis (edge-based ontology analysis in Figure 3A) first assigns link ontologies based on GO terms of the two interacting nodes in the network and then identifies enriched link ontologies at each time point. For example, Wang et al. [56] constructed dynamic networks using gene expression data generated from incipient, moderate and severe stages of Alzheimer’s disease (AD) and applied this method to interpret the dynamic networks. The node-based GO analysis could not result in GO terms that can be used to distinguish different stages of AD progression. In contrast, the edge-based GO analysis correctly identified GO terms specific to each stage of AD: (i) vesicle-mediated transport at the incipient stage; (ii) regulation of kinase activity for the moderate stage; and (iii) sterol transport, apoptosis and proteolysis at the severe stage. Thus, the edge-based GO analysis more effectively distinguishes the difference of cellular functions represented by the networks at individual time points, thereby leading to better understanding of temporal transition of cellular functions delineated by dynamic networks.

In addition to the GO analysis, temporal transitions of characteristics of dynamic networks (e.g. distributions of shortest paths, degrees of centrality and clustering coefficients) have been also analyzed [10]. Furthermore, temporal changes of network motifs in dynamic networks provide useful insights into transition of kinetic behaviors in outputs of dynamic networks (network motif analysis in Figure 3A). Network motifs are recurring building blocks in biological networks, including feedback and feedforward loops [57]. Several tools, such as FANMOD [58] and MAVisto [59], have been developed to identify overrepresented network motifs in a biological network. Comparison of the overrepresented regulatory motifs in dynamic networks across time points can reveal disappearance of existing motifs and formation of new regulatory motifs, providing the knowledge of how dynamic behaviors in response to a certain stimulus can vary over time. For example, Kim et al. [60] selected DEGs from time-course gene expression data generated at multiple stages of Drosophila embryogenesis and reconstructed
active GINs at different stages of the embryogenesis using TRANSFAC database. By comparing the overrepresented network motifs at individual time points, they identified that feedback loops appeared more frequently at later stages than at early stages. Furthermore, based on GO analysis of the nodes in these network motifs, the genes in the feedback loops were related to cell fate determination at the
late stages, implying utilization of the feedback loops for achieving cell fate determination. This is consistent with previous findings that bistability of positive feedbacks serves as a cell fate determination mechanism [61].

**SPATIAL TRANSITION OF NODES AND EDGES**

In eukaryotic cells, the internal space of a cell is compartmentalized, permitting to convert or diversify functions of proteins by modulating their spatial distribution on external stimuli or post-translational modifications (PTMs). Proteins localized at the same subcellular compartment are often functionally associated through their interactions, as shown in phenotypically similar diseases and disease comorbidity [62]. More than 35% of the cellular proteome are localized at multiple compartments [63, 64]. If a protein is localized, for instance, in nucleus and cytosol, it will function differently in the compartments [65, 66] by forming spatially distinct interaction networks. Thus, in addition to temporal transition, spatial transition of proteins and their interactions should be examined to understand dynamic behaviors in biological systems. Abnormal changes of spatial distribution of proteins can contribute to disease pathogenesis. For example, abnormal distributions of proteins between the nucleus and cytosol have been observed in various types of carcinoma cells [67]. Thus, deciphering spatial transition of proteins can be crucial to develop therapeutic means to prevent disease progression.

To understand spatial transition of proteins and their interactions, various experimental and computational approaches for decoding spatial information of proteins have been developed [13]. Proteomic analysis followed by subcellular fractionation is used to analyze the proteomes in subcellular compartments. The fractionation techniques include not only classical methods, such as differential centrifugation, density-gradient centrifugation and differential detergent fractionation, but also emerging methods, such as free-flow electrophoresis, immunoaffinity purification and fluorescent-assisted organelle sorting. Proteomics analysis reveals lists of proteins localized in individual organelles, as well as interacting partners of the proteins and their PTMs in different organelles. The proteomic data are deposited into several databases, such as Organelle DB, MitoCarta, LOCATE and SUBA [68–71]. Also, a number of *in silico* tools for predicting protein localization have been developed. The tools include protein sequence-based methods (e.g. signalP [72], PredSL [73] and Predotar [74]), annotation-based methods (e.g. ProLog-GO [75] and PSLpred [76]), mixtures of two strategies (e.g. MultiLoc [77] and WoLF-PSORT [78]) and methods that systematically integrate prediction results from other tools (e.g. ConLoc [79]). For the dynamic networks with the information of spatial transition of proteins, the aforementioned tools for functional interpretation of dynamic networks (e.g. node- and edge-based GO analyses) can be applied to understand how cellular functions are distinctively activated in different organelles and how they transit over time. Similarly, regulatory motifs defined by spatial segregation of proteins and how they change over time can be identified using the motif detection algorithms aforementioned.

In addition, mathematical modeling using ordinary or partial differential equations can be used to understand functional roles of the spatial transition of proteins and their interactions. Mathematical modeling of biological networks enables us to understand dynamic operation of the networks. Combining mathematical modeling with network motif theory can shed light on understanding the role of dynamic/multiple localization of proteins in the regulation of network dynamics. For example, Santos et al. [80] showed that a positive feedback loop triggering the cell cycle entry is formed by cyclin B1 phosphorylation and nuclear translocation of active Cdk1-cyclin B1. Mathematical modeling of this feedback loop revealed that the spatially coded positive feedback ensures decisiveness and irreversibility of the cell cycle entry. Mathematical modeling of biological networks involving proteins localized in multiple compartments is based on the assumption that molecules within the same compartment are homogeneously distributed. When this assumption does not hold, partial differential equations should be used to incorporate diffusion effect of molecules within the same compartment [81].

Integrating spatial information of proteins into dynamic networks enables to decode molecular networks governing biological processes in both space and time. Specifically, we can unfold a dynamic network at each time point into multiple networks in individual subcellular compartments, compare spatially unfolded networks at consecutive time points and thus understand the transitions of the proteins and
the regulatory motifs in space and time (Figure 4). Protein A activates protein B in cytosol (t = 1); the activated cytosolic protein B then translocates into nucleus and then induces gene C, which induces gene B (t = 2), thereby forming a transcriptional feedback loop between genes B and C; protein C induced by protein B is then localized in cytosol and activates protein D (t = 3); and proteins C and D translocate into mitochondria in which protein C inactivates protein D, but activates protein E, which in turn activates protein D (t = 4). This spatial unfolding of a dynamic network revealed a simple activation of protein D by protein C in cytosol, as well as a feedforward loop among proteins C, D and E in mitochondria, which could not be identified without consideration of spatial unfolding. Furthermore, the unfolding can help clarify conflicting information regarding regulatory relationships between proteins C and D, depending on whether data represent either cytosolic or mitochondrial interaction between them. Thus, the space- and time-evolving nature of the networks provides a comprehensive basis for understanding the roles of key molecules and network motifs in diverse physiological and pathological events. Despite the importance of spatial transition of the nodes, however, the experimental and computational tools are not sufficient to decode transition of the networks in both space and time. Thus, new experimental and computational tools to improve understanding of spatiotemporal transition of the networks are needed.

**DISCUSSION**

In this review, we presented the methods for inferring dynamic networks using time-course global data. We categorized these methods according to the presence and absence of temporal transition of nodes and/or edges. First, BN, DBN and RN methods estimate dynamic networks with temporal regulatory relationships, but with no transition of nodes and edges. Second, the methods developed by

![Figure 4: Spatial unfolding of dynamic networks. A dynamic network (G) (first row) shows temporal activation of nodes and edges at t (t = 1, 2, 3 and 4). A spatially unfolded dynamic network (S) (second row) shows activation and translocation of nodes and their interactions in three subcellular compartments (cytosol, nucleus and mitochondria) at t. Capital letters denote proteins (nodes), and the subscripts represent the subcellular compartments (C: cytosol, M: mitochondria and N: nucleus). The comparisons of G and S at t = 3 and 4 show the use of spatial unfolding of a dynamic network for understanding of spatiotemporal transition of nodes and edges. In S3, D is positively regulated by C in cytosol. In S4, however, D is negatively regulated by C after translocation of C and D to mitochondria. These spatially unfolded relationships can clarify the conflicting data regarding positive and negative regulations of D by C without consideration of spatial unfolding. Furthermore, after their translocation, S4 shows a feedforward loop newly formed in mitochondria at t = 4. These examples demonstrate that spatiotemporally evolving networks can provide a comprehensive basis for understanding transitions of biological systems.](https://academic.oup.com/bib/article-abstract/15/2/212/212014)
In this review, we presented the node- and edge-based GO analyses for functional interpretation of dynamic networks. Despite these analyses, it is still challenging to interpret dynamic networks in appropriate biological contexts. The current network inference methods are mostly designed to model statistical dependence between the nodes for the inference of regulatory relationships. To better extract biological meaning from inferred dynamic networks, however, it would be essential to integrate other data related to functions of the nodes and edges, such as GO biological processes (GOBPs) and/or molecular functions (GOMFs), into the models and the inference algorithms for dynamic networks. In this integrative inference, the nodes should be more likely to interact, when they show similar temporal expression patterns and have the same GOBPs, and further to transition in dynamic networks by forming a functional module (i.e. densely connected nodes) associated with the GOBPs. Multiple functional modules can be formed when the nodes with the same GOBPs show several different patterns of temporal gene expression. Thus, these functional modules can directly guide functional interpretation of the dynamic networks, as well as identification of regulatory motifs and understanding of their transitions.

There has been the lack of tools for inferring spatiotemporal transition of the nodes from global assay data. To decode spatiotemporally evolving networks of genes and proteins, a new method is required that can integrate spatial information (e.g. GO cellular compartments; GOCCs) of proteins with time-course gene expression data during the inference of dynamic networks. In this method, the nodes with the same GOCCs should be more likely to be linked when the nodes show similar temporal expression patterns, thereby forming a functional module for the GOCC. Thus, the functional modules can guide interpretation of spatially unfolded dynamic networks, identification of regulatory motifs essential for spatiotemporal transitions of the networks and understanding of transitions of the key regulatory motifs in both space and time.

Finally, the categorization of the network inference methods in this review can serve as a useful guideline to select suitable tools for one’s own data and study. One can define the type of dynamic networks needed to answer biological questions in the study based on the following aspects: (i) whether dynamic networks should describe transitions of the nodes and/or edges, (ii) whether the networks
should be directed or undirected and (iii) what kinds of data should be integrated into the inference of the dynamic networks to answer biological questions. For each of these aspects of the dynamic networks, one can refer to the corresponding sections in this review to determine an appropriate method for the defined type of the dynamic networks. Furthermore, one should consider the following computational issues: (i) how large dynamic networks should be estimated and (ii) how much information the global time-course data has. To reconstruct dynamic networks involving transition of edges over time, relatively small amounts of data points compared with the number of nodes keep the methods from reliable estimation of temporal transition of edges. For example, one needs to estimate directed dynamic networks from time-course gene expression of 2000 genes measured at 10 time points. In this case, the number of nodes is much larger than that of samples (i.e. time points). It is desirable to use either a method for inferring dynamic networks with temporal association (e.g. RN methods using directed partial CC) or PNA using the known interactomes. However, when one can reduce the number of nodes by focusing on the networks related to a cellular process through integration of GOBPs, directed dynamic networks with temporal transition of edges can be inferred using TV-DBN and nsDBN. When multiple methods can be applied, one can determine the best dynamic networks after comparing the performance of the methods in answering the biological questions through the resulting dynamic networks.

**Key Points**
- We summarized 29 computational methods for estimating dynamic networks using time-course global data after categorizing them according to the presence and absence of temporal transition of nodes and/or edges, directionality of edges and integration of other data than time-course data.
- We presented methods for (i) understanding functional transition represented by dynamic networks, (ii) detecting overrepresented network motifs in dynamic networks, (iii) analyzing temporal transition of network characteristics and (iv) modeling kinetic behaviors represented by dynamic networks.
- We finally discussed methods for unfolding dynamic networks into multiple subcellular compartments. This spatial unfolding of dynamic networks can provide understanding of transitions of biological networks both in time and space.

**FUNDING**

**References**