

RESEARCH

Open Access



# Bayesian graphical models for computational network biology

Yang Ni<sup>1\*</sup> , Peter Müller<sup>2</sup>, Lin Wei<sup>3</sup> and Yuan Ji<sup>3,4</sup>

From The Fourth International Workshop on Computational Network Biology: Modeling, Analysis, and Control (CNB-MAC 2017) Boston, MA, USA. 20 August 2017

## Abstract

**Background:** Computational network biology is an emerging interdisciplinary research area. Among many other network approaches, probabilistic graphical models provide a comprehensive probabilistic characterization of interaction patterns between molecules and the associated uncertainties.

**Results:** In this article, we first review graphical models, including directed, undirected, and reciprocal graphs (RG), with an emphasis on the RG models that are curiously under-utilized in biostatistics and bioinformatics literature. RG's strictly contain chain graphs as a special case and are suitable to model reciprocal causality such as feedback mechanism in molecular networks. We then extend the RG approach to modeling molecular networks by integrating DNA-, RNA- and protein-level data. We apply the extended RG method to The Cancer Genome Atlas multi-platform ovarian cancer data and reveal several interesting findings.

**Conclusions:** This study aims to review the basics of different probabilistic graphical models as well as recent development in RG approaches for network modeling. The extension presented in this paper provides a principled and efficient way of integrating DNA copy number, DNA methylation, mRNA gene expression and protein expression.

**Keywords:** Directed graph, Undirected graph, Chain graph, Reciprocal graph, Causality

## Background

This article starts with a comprehensive review of graphical models and recent development in constructing biological networks using reciprocal graphs (RG's). In order to integrate multi-omics data including proteomics, we extend the work in [1] utilizing fundamental biological knowledge and the factorization of the joint distribution.

Computational network biology (CNB) is an emerging research field that encompasses theory and applications of network models to systematically study different molecules (DNA, RNA, proteins, metabolites and small molecules) and their complex interactions in living cells. CNB provides new insights in diverse research areas including system biology, molecular biology, genetics, pharmacology and precision medicine. The development

of high-throughput profiling technologies allows for the interrogation of the status of many molecules in a cell at the same time, which makes it possible to jointly model these cell components with network modeling approaches. CNB is an interdisciplinary research area and has received contributions from researchers with various background. In this paper, we focus on network modeling approaches from a statistical perspective, namely, probabilistic graphical models (PGM) which allow a complete probabilistic description of interaction patterns between molecules and the associated uncertainties when estimating such interactions from noisy genomic and/or proteomic data.

PGM are probabilistic models for multivariate random variables whose conditional independence (also known as Markov property) structure is characterized by an underlying graph. PGM provides a concise, complete and explicit representation of joint distribution and allows for convenient Gibbs factorization of the density (if it exists) and hence local computations. The Markov property can

\*Correspondence: [yangni87@gmail.com](mailto:yangni87@gmail.com)

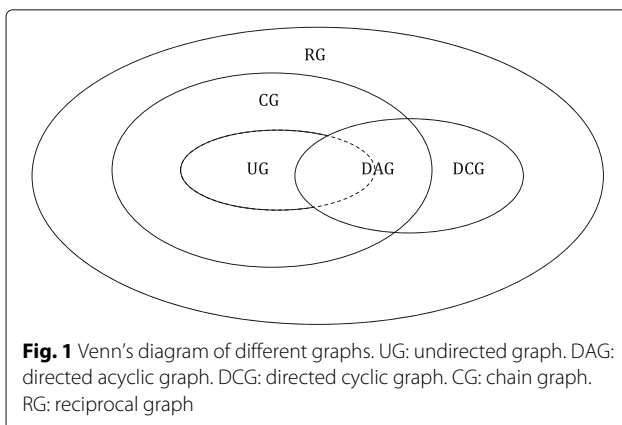
<sup>1</sup>Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, TX 78712, USA

Full list of author information is available at the end of the article

be directly read off the graph through the notion of graph separation. PGM is closely related to causal inference, path analysis and expert system and finds a wide range of applications in political sciences, economics, genetics, biology, physics and psychology.

The most commonly used graphs are undirected graphs (UG) and directed acyclic graphs (DAG). UG represents symmetric interactions between variables whereas DAG allows for asymmetric ones that can be potentially interpreted as cause-effect relationships. Both UG and DAG are a special case of a more general graphical model known as chain graphs (CG) [2, 3] which allow for both symmetric and asymmetric conditional independence relationships. Despite of the flexibility of CG, the directed relationships are still constrained to be (block-) recursive, that is, reciprocal causality is prohibited. However, cyclic causal relationships are fundamental and ubiquitous in science, including, for example, the law of supply and demand in economics or feedback mechanism in gene regulatory networks. Spirtes [4] and Koster [5], in their respective seminal work, independently proposed coherent non-recursive graphical models including directed cycles of which the joint distribution was previously thought ill-defined ([6], p. 72). Spirtes [4] developed directed cyclic graphical (DCG) models which include DAG's as a special case, and [5] proposed an even more general class of graphical models, reciprocal graphical (RG) models, which include CG (hence UG and DAG) and DCG as special cases. The relationships between the aforementioned graphs are depicted as a Venn's diagram in Fig. 1.

In this paper, we focus on RG's due to their statistical generality, the capability of modeling feedback loops and their still underappreciated popularity in the bio-statistics and bioinformatics literature. The rest of this paper is structured as follows. We first review the basics of UG, DAG and RG in this section. In the next section, we provide a comprehensive review of novel Bayesian approaches in modeling molecular networks using RG.



Then, we extend the RG approach in [1] to modeling The Cancer Genome Atlas (TCGA) ovarian cancer data that are measured on three different levels: DNA, RNA and proteins. The extension is non-trivial: we factorize an RG into two separate RG's (one for DNA and RNA and the other for RNA and protein) and make coherent joint inference by exploiting two known biochemical processes: DNA transcribes to RNA and RNA in turn translates to protein, the latter also allowing us to estimate the direction of the relationship between RNA and protein which informs the directionality of the regulatory relationships between proteins. To the best of our knowledge, we are among the first to reconstruct multi-omics functional networks that consider mass spectrometry proteomics data generated by the CPTAC.

*Overview of directed, undirected and reciprocal graphical models.* Graphical models are a class of statistical models for a set of random variables  $\mathbf{Z} = (Z_1, \dots, Z_p)^T$  that provide a graphic representation of conditional independence relationships among the variables. Using graphical models allow practitioners to simplify inferences, obtain parsimonious solutions via variable selection, enable local computations, and define a framework for causal inference.

In a graphical model, variables are represented by a set of nodes and their associated interactions are represented by edges. A missing edge usually represents the conditional independence between the corresponding pair of random variables conditional on a certain set of other variables. The conditioning set depends on the chosen type of graph. Here, we provide a minimal set of terminologies required to describe the Markov properties in later sections.

A graph  $\mathcal{G} = (V, E)$  is defined by a set of nodes  $V = \{1, \dots, p\}$  and a set of directed and undirected edges  $E = E^d \cup E^u$ . For a pair of nodes  $i, j \in V$ , we denote an undirected edge by  $i-j$  or  $\{i, j\} \in E^u$  and a directed edge by  $i \rightarrow j$  or  $(i, j) \in E^d$ . We write  $i \mapsto j$  if there exists a path from  $i$  to  $j$ . A *path* is an ordered sequence  $(i_0, \dots, i_K)$  of distinct nodes except possibly  $i_0 = i_K$  such that  $\{i_{k-1}, i_k\} \in E$  or  $(i_{k-1}, i_k) \in E$  for  $k = 1, \dots, K$ . A *cycle* is then a path with  $i_0 = i_K$ . A *path component* is a set of nodes that are all connected by an undirected path. The *boundary* of a node  $i$  is  $bd(i) = \{j \mid \{j, i\} \in E \text{ or } (j, i) \in E\}$  and the *boundary* of a subset  $A \subseteq V$  is  $bd(A) = \bigcup_{i \in A} bd(i) \setminus A$ . A node  $i$  is said to be a *parent* of node  $j$ , if  $(i, j) \in E$ . The set of parents of  $j$  is denoted by  $pa(j)$ . The *ancestors* of node  $i$  are  $an(i) = \{i\} \cup \{j \mid j \mapsto i\}$ . The *descendants* of node  $i$ , denoted by  $de(i)$ , are the nodes  $i$  such that there is a path from  $i$  to  $j$ . The *non-descendants* of  $i$  are  $nd(i) = V \setminus (de(i) \cup \{i\})$ . For  $A \subseteq V$ , the induced *subgraph* by  $A$  is  $\mathcal{G}_A = (A, E_A)$  where  $E_A = E \cap (A \times A)$ . A subset  $A \subseteq V$  is *anterior* if  $bd(A) = \emptyset$  and the minimal anterior set of  $A$  is denoted by  $an(A)$ . Note that we use  $an()$  to denote both ancestors

and minimal anterior set; it should not cause any confusion because in fact  $an(\{i\}) = an(i)$ . Notice that (minimal) anterior set [5] is equivalent to (minimal) ancestral set in [7]. A graph is *complete* if all nodes are joined by an edge. A complete subgraph that is maximal with respect to the subset relation is called *clique*. A useful procedure in defining Markov properties of graphs with directed edges is *moralization*. The moralization is performed in two steps. Make the boundary of each path component complete by adding undirected edges and then replace all directed edges by undirected ones. We call the resulting undirected graph *moral* graph, denoted by  $\mathcal{G}^m = (V, E^m)$ .

We say that graph  $\mathcal{G}$  is a undirected graph (UG) if  $E = E^u$  contains only undirected edges. And graph  $\mathcal{G}$  is said to be a directed acyclic graph (DAG) if  $E = E^d$  contains only directed edges and no cycles. A *reciprocal graph* (RG) is a graph  $\mathcal{G}$  that can have both directed and undirected edges and cycles, with the only restriction being no directed edges between nodes in the same path component [5]. In the following, we provide a concise overview of each type of graph. More comprehensive and technical details of the contents can be found in [7] and [6].

**Undirected graphical models**

Let  $P$  denote the joint distribution of  $\mathbf{Z} = \mathbf{Z}_V = (Z_1, \dots, Z_p)^T$ . We say  $P$  has a *Gibbs factorization* (or simply factorization) with respect to a UG  $\mathcal{G}$  if its density  $f(\mathbf{Z})$  can be written as

$$f(\mathbf{Z}) = \prod_{C \in \mathcal{C}} \psi_C(\mathbf{Z}_C) \tag{1}$$

where  $\mathcal{C}$  is the collection of cliques of  $\mathcal{G}$  and  $\psi_C$  is an arbitrary nonnegative function on the domain of  $\mathbf{Z}_C$  for  $C \in \mathcal{C}$ .

Not surprisingly, the factorization property of  $P$  is closely related to its Markov properties; in fact, the former implies the latter. The Markov property of a UG relies on a notion of graph *separation*. For three disjoint subsets  $A, B, C \in V$  of nodes,  $A$  is said to be *separated* from  $B$  by  $C$  if every path between  $A$  and  $B$  includes a node in  $C$ . The global Markov property of a UG is given in the following definition.

**Definition 1** (Global Markov property) *Given a UG  $\mathcal{G}$ , the joint distribution  $P$  of  $\mathbf{Z}_V$  is global Markov with respect to  $\mathcal{G}$  if for any three disjoint subsets  $A, B$  and  $C$  of  $V$ ,  $\mathbf{Z}_A$  is independent of  $\mathbf{Z}_B$  given  $\mathbf{Z}_C$ , written as  $\mathbf{Z}_A \perp\!\!\!\perp \mathbf{Z}_B \mid \mathbf{Z}_C$  whenever  $C$  separates  $A$  and  $B$ .*

The global Markov property always implies the local Markov property.

**Definition 2** (Local Markov property) *Given a UG  $\mathcal{G}$ , the joint distribution  $P$  of  $\mathbf{Z}_V$  is local Markov with respect to  $\mathcal{G}$  if for any node  $i \in V$ ,  $\mathbf{Z}_i \perp\!\!\!\perp \mathbf{Z}_{V \setminus \{bd(i), i\}} \mid \mathbf{Z}_{bd(i)}$ .*

The local Markov property in turn always implies the pairwise Markov property.

**Definition 3** (Pairwise Markov property) *Given a UG  $\mathcal{G}$ , the joint distribution  $P$  of  $\mathbf{Z}_V$  is pairwise Markov with respect to  $\mathcal{G}$  if for any pair of nodes  $i, j \in V$ ,  $\mathbf{Z}_i \perp\!\!\!\perp \mathbf{Z}_j \mid \mathbf{Z}_{V \setminus \{i, j\}}$ .*

In summary, factorization in (1)  $\Rightarrow$  global Markov property  $\Rightarrow$  local Markov property  $\Rightarrow$  pairwise Markov property. A natural question to ask is: is the relation also true backwards? The answer is yes when  $P$  has a positive and continuous density, which is summarized in the following theorem, often referred to as Hammersley-Clifford theorem [8].

**Theorem 1** (Hammersley-Clifford) *Suppose a probability distribution  $P$  has a positive and continuous density  $f$ . Then  $P$  satisfies the pairwise Markov property with respect to a UG  $\mathcal{G}$  if and only if  $f$  factorizes according to  $\mathcal{G}$ .*

When  $P = N(0, \Lambda^{-1})$  is multivariate Gaussian with precision matrix  $\Lambda$ , undirected graphical models are defined by zero constraints on  $\Lambda$ . There is a one-to-one correspondence between the graph structure  $\mathcal{G}$  and the zero patterns in  $\Lambda$ . Specifically, the precision matrix  $\Lambda$  is constrained to the cone of positive definite matrices with off-diagonal entry  $\Lambda_{ij} = 0$  if and only if there is a missing edge in  $\mathcal{G}$  between nodes  $i$  and  $j$ . Therefore, in the Gaussian case, learning the graph structure is equivalent to recovering the zero patterns in  $\Lambda$ . Interested readers are referred to recent Bayesian approaches [9–11] and non-Bayesian approaches [12, 13].

**Directed graphical models**

We say the distribution  $P$  factorizes with respect to a DAG  $\mathcal{G}$  if its density  $f(\mathbf{Z})$  can be written as

$$f(\mathbf{Z}) = \prod_{i=1}^p f(Z_i \mid \mathbf{Z}_{pa(i)}) \tag{2}$$

Similar to UG’s, the Markov property of a DAG can be also defined by the concept of graph separation. However, unlike UG’s, the separation is more complicated in DAG’s. There are two different but equivalent approaches in describing separation in a DAG. Pearl [14] introduced the notion of d-separation characterized by several conditions and then defined the global Markov property of a DAG in the same way as that of an UG except that separation is replaced by d-separation. Another approach [3] is

based on the moralization and graph separation in UG's. We elaborate more on the latter because it will be used again in defining Markov properties of RG's. Similar to the previous section, we introduce global, local and pairwise Markov properties of DAG and then state the relationships between them. Recall that  $\mathcal{G}^m$  denotes the moralized graph.

**Definition 4** (Directed global Markov property) *Given a DAG  $\mathcal{G}$ , the joint distribution  $P$  of  $\mathbf{Z}_V$  is global Markov with respect to  $\mathcal{G}$  if for any three disjoint subsets  $A, B$  and  $C$  of  $V$ ,  $\mathbf{Z}_A$  is independent of  $\mathbf{Z}_B$  given  $\mathbf{Z}_C$ , written as  $\mathbf{Z}_A \perp\!\!\!\perp \mathbf{Z}_B \mid \mathbf{Z}_C$  whenever  $C$  separates  $A$  and  $B$  in  $\mathcal{G}_{an(A \cup B \cup C)}^m$ .*

**Definition 5** (Directed local Markov property) *Given a DAG  $\mathcal{G}$ , the joint distribution  $P$  of  $\mathbf{Z}_V$  is local Markov with respect to  $\mathcal{G}$  if for any node  $i \in V$ ,  $\mathbf{Z}_i \perp\!\!\!\perp \mathbf{Z}_{nd(i)} \mid \mathbf{Z}_{pa(i)}$ .*

**Definition 6** (Directed pairwise Markov property) *Given a DAG  $\mathcal{G}$ , the joint distribution  $P$  of  $\mathbf{Z}_V$  is pairwise Markov with respect to  $\mathcal{G}$  if for any pair of nodes  $i, j \in V$ ,  $\mathbf{Z}_i \perp\!\!\!\perp \mathbf{Z}_j \mid \mathbf{Z}_{nd(i) \setminus \{j\}}$ .*

Similar to UG's, it is easy to show that factorization in (2)  $\Rightarrow$  directed global Markov property  $\Rightarrow$  directed local Markov property  $\Rightarrow$  directed pairwise Markov property. In fact, directed global and local Markov properties are equivalent [15]. Moreover, a theorem similar to Hammersley-Clifford's theorem holds for DAG's as well.

**Theorem 2** ([7]) *Suppose a probability distribution  $P$  has a continuous density  $f$ . Then  $P$  satisfies the local Markov property with respect to a DAG  $\mathcal{G}$  if and only if  $f$  factorizes according to  $\mathcal{G}$ . Furthermore, if the density  $f$  is positive, local and pairwise Markov properties are equivalent.*

In a Gaussian DAG, each conditional distribution in (2) is a linear model

$$f(\mathbf{Z}_i \mid \mathbf{Z}_{pa(i)}) = N(\mathbf{Z}_i \mid \sum_{j \in pa(i)} \beta_{ij} \mathbf{Z}_j, \sigma_i^2),$$

where  $N(\cdot \mid \mu, \sigma^2)$  is a Gaussian density with mean  $\mu$  and variance  $\sigma^2$ . Essentially, a DAG is decomposed as a system of recursive and independent linear regressions for which the graph structure inference can be performed using standard model selection technique.

However, there is an important caveat in using DAG's for statistical inference. That is the the notion of *Markov equivalence* – different DAG's may induce the same set of conditional independence relationships. For example, the graphs  $i \leftarrow j \rightarrow k$  and  $i \leftarrow j \leftarrow k$  induce the same conditional independence relationship,  $i \perp\!\!\!\perp k \mid j$ . Both are equivalent to the UG  $i - j - k$ . In fact, any perfect DAG

(i.e. a DAG without configuration  $i \rightarrow j \leftarrow k$ , that is, all parents are married) is Markov equivalent to a decomposable UG (i.e. UG in which all cycles of length four or more have a chord, which is an edge that is not part of the cycle but connects two nodes of the cycle). Conversely, given an undirected decomposable graph, there exists a Markov-equivalent perfect DAG; see [7] or [16]. One approach to avoid repeatedly analyzing Markov equivalent DAG's is adopting a prior ordering of the nodes. The ordering may be obtained from subjective prior knowledge such as a reference network [17] or more objectively from an ordering learning algorithm such as order-MCMC [18]. For other recent DAG approaches, see [19–21] or [22].

**Reciprocal graphical models, simultaneous equation models and path diagrams**

We introduce the global Markov property for RG's in the following definition and omit the local and pairwise Markov properties since they are not as practically useful as in UG and DAG.

**Definition 7** (Reciprocal global Markov property) *Given a RG  $\mathcal{G}$ , the joint distribution  $P$  of  $\mathbf{Z}_V$  is global Markov with respect to  $\mathcal{G}$  if for any three disjoint subsets  $A, B$  and  $C$  of  $V$ ,  $\mathbf{Z}_A$  is independent of  $\mathbf{Z}_B$  given  $\mathbf{Z}_C$ , written as  $\mathbf{Z}_A \perp\!\!\!\perp \mathbf{Z}_B \mid \mathbf{Z}_C$  whenever  $C$  separates  $A$  and  $B$  in  $\mathcal{G}_{an(A \cup B \cup C)}^m$ .*

Similarly to UG and DAG, there is also a Hammersley-Clifford-type theorem for RG's ([5], Theorem 3.4).

However, while it is a beautiful theoretical result on the equivalence between the global Markov property and the Gibbs factorization in RG, the latter is difficult to work with in practice due to its complex form. Fortunately, in the Gaussian case, there is an easy connection between RG and simultaneous equation models (SEMs). To describe this connection, let  $\mathbf{Z} = (\mathbf{Y}, \mathbf{X})$  be divided into two sets of variables  $\mathbf{Y}$  and  $\mathbf{X}$  and consider an SEM,

$$\mathbf{Y} = \mathbf{A}\mathbf{Y} + \mathbf{B}\mathbf{X} + \mathbf{E} \tag{3}$$

where we assume the following five conditions:

- C1.  $(\mathbf{Y}, \mathbf{X}) \sim N(0, \Sigma)$ ;
- C2.  $\mathbf{E} \perp\!\!\!\perp \mathbf{X}$ ;
- C3.  $\mathbf{I} - \mathbf{A}$  is invertible;
- C4.  $Cov(\mathbf{E})$  is diagonal;
- C5.  $\Psi = Cov(\mathbf{X})$  is block diagonal with full diagonal blocks.

To link an SEM to an RG, we draw a path diagram  $\mathcal{G} = (V, E)$  of SEM by the following rules:

- (i) define nodes  $V = \{1, \dots, p, p + 1, \dots, p + q\}$  which represent  $(\mathbf{Y}, \mathbf{X}) = (Y_1, \dots, Y_p, X_1, \dots, X_q)$ ;

- (ii) draw directed edges  $E^d = \{(j, i) \mid a_{ij} \neq 0 \text{ or } b_{i,j-p} \neq 0\}$ ; and
- (iii) draw undirected edges  $E^u = \{(i, j) \mid \Psi_{i-p,j-p} \neq 0\}$ .

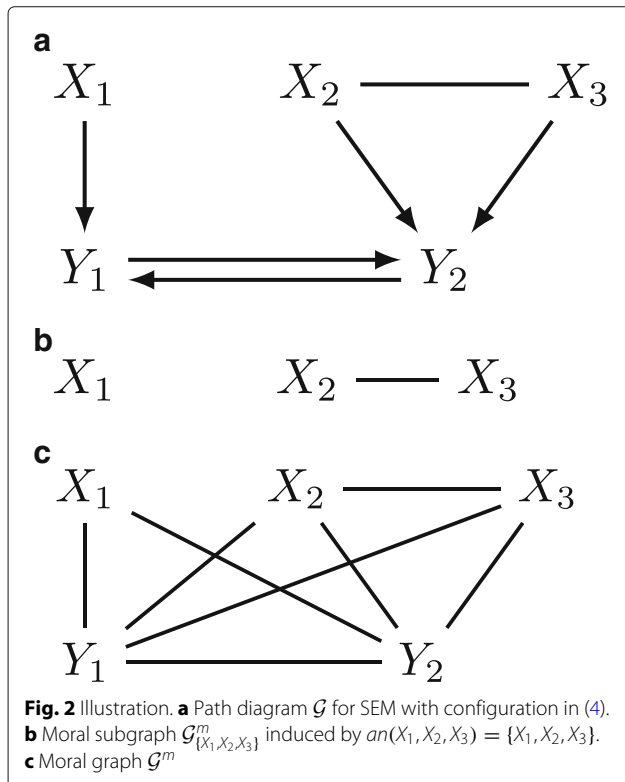
In words, (i) we introduce a node for each variable in  $(Y, X)$  with nodes  $j = 1, \dots, p$  corresponding to  $Y_j$  and  $p + j$  corresponding to  $X_j$  for  $j = 1, \dots, q$ ; (ii) nodes  $i = 1, \dots, p$  (i.e.  $Y_i$  nodes) become targets of directed edges from node  $j$  if the corresponding  $a_{ij} \neq 0$  or  $b_{i,j-p} \neq 0$ ; (iii) we introduce undirected edges between  $X_i$  and  $X_j$  (i.e. nodes  $i + p$  and  $j + p$ ) if  $\Psi_{ij} \neq 0$ . Figure 2a shows an example of an RG with  $p = 2$  and  $q = 3$ ,

$$A = \begin{bmatrix} 0 & * \\ * & 0 \end{bmatrix}, B = \begin{bmatrix} * & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{bmatrix} \text{ and } \Psi = \begin{bmatrix} * & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{bmatrix}, \quad (4)$$

with  $*$  indicating non-zero elements. With conditions C1-C5, the Markov properties of an SEM can be read off its path diagram.

**Theorem 3** ([5]) *Let  $Y = AY + BX + E$  be an SEM satisfying conditions C1-C5. The joint distribution  $P$  of  $(Y, X)$  is global Markov with respect to the path diagram  $\mathcal{G}$ .*

For example, the SEM with configuration in (4) implies  $X_1 \perp\!\!\!\perp X_2, X_3$  which is evident by looking at the moral subgraph  $\mathcal{G}_{an(X_1, X_2, X_3)}^m = \mathcal{G}_{\{X_1, X_2, X_3\}}^m$  in Fig. 2b. Clearly, the empty set  $\emptyset$  separates  $X_1$  and  $X_2, X_3$ . We can also find



$X_1 \perp\!\!\!\perp X_2, X_3 \mid Y_1, Y_2$  by moralizing the entire path diagram (Fig. 2c) since set  $\{Y_1, Y_2\}$  separates  $X_1$  and  $X_2, X_3$  in  $\mathcal{G}^m$ . However, we remark that the class of SEMs satisfying conditions C1-C5 is a subclass of RG. For instance,  $i \rightarrow j - k \leftarrow l$  is a CG (hence an RG) but does not correspond to an SEM. Similarly to DAG's, DCG's (or more generally RG's) can also be divided into Markov equivalence classes; see [23–28] for characterization and learning of Markov equivalence classes of DCG's.

**Methods**

In this section, we will review recent RG approaches in modeling molecular networks including [29, 30] and [1].

**Modeling signaling pathways through latent variables**

Telesca et al. [29] considered gene expressions  $y_{ij}$  for  $p$  genes measured over  $n$  samples,  $i = 1, \dots, n$  and  $j = 1, \dots, p$ . They follow [31] and introduce trinary latent variables  $e_{ij} \in \{-1, 0, 1\}$  defining three possible categories of each expression  $y_{ij}$ ,

$$e_{ij} = \begin{cases} -1 & \text{underexpressed gene } j \text{ in sample } i \\ 1 & \text{overexpressed gene } j \text{ in sample } i \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

The  $e_{ij}$  can be interpreted as an underlying biologic signal of under-, over- or normal expression. The motivation of the proposed inference is that only  $e_{ij}$  should be modeled without over-interpreting the additional noise in  $y_{ij}$  as biologically meaningful. Given  $e_{ij}$ , a Gaussian-uniform mixture model is assumed for gene expression  $y_{ij}$ ,

$$\tilde{y}_{ij} | e_{ij} \sim f_{e_{ij}, j}(\tilde{y}_{ij}), \quad (6)$$

where  $f_{-1, j} = U(-\kappa_j^-, 0), f_{1, j} = U(0, \kappa_j^+), f_{0, j} = N(0, \sigma_j^2)$  and  $\tilde{y}_{ij} = y_{ij} - (\alpha_i + \mu_j)$  is the normalized relative abundance, corrected for a sample-specific effect  $\alpha_i$  and a gene-specific effect  $\mu_j$ . Conditional conjugate priors are assigned to  $\alpha_i, \mu_j, \kappa_j^-, \kappa_j^+$ :  $\alpha_i \sim N(0, \tau_\alpha^2), \mu_j \sim N(m_\mu, \tau_\mu^2), \kappa_j^- \sim IG(a_\kappa^-, b_\kappa^-), \kappa_j^+ \sim IG(a_\kappa^+, b_\kappa^+), \sigma_j^2 \sim IG(a_\sigma, b_\sigma)$ . Note that  $\alpha_i$  and  $\mu_j$  are not identifiable because  $\alpha_i + \mu_j = (\alpha_i - c) + (\mu_j + c)$  for any constant  $c$ . If desired, the identifiability problem can be resolved by setting  $\sum_{i=1}^n \alpha_i = 0$ . For posterior simulation, the constraint can be implemented by setting  $\alpha_i \leftarrow \alpha_i - c$  and  $\mu_j \leftarrow \mu_j + c$  with  $c = \sum_{i=1}^n \alpha_i$  at each iteration.

The dependence of gene expressions is modeled through the prior model for the latent variables  $e_{ij}$ . In [29], they further introduced another set of latent Gaussian variables  $z_{ij}$  as probit scores for each trinary  $e_{ij}$ ,

$$e_{ij} = \begin{cases} -1 & \text{if } z_{ij} > 1 \\ 1 & \text{if } z_{ij} < -1 \\ 0 & \text{otherwise.} \end{cases}$$

The continuous latent probit scores are then modeled by an SEM

$$z_{ij} = m_{ij} + \sum_{k \neq j} a_{jk}(z_{ik} - m_{ik}) + \epsilon_{ij}$$

with  $\epsilon_{ij} \sim N(0, s_j^2)$  and  $m_{ij} = \mathbf{x}_i^T \beta_j$  for covariates  $\mathbf{x}_i$ . Conjugate normal priors are assumed for regression coefficients  $\beta_j$  and  $a_{jk}$ . By (3) this defines an RGM. Let  $\mathbf{A}$  be the  $p \times p$  matrix of which the diagonal elements are zeros and the off-diagonal  $(j, k)$  entries are  $a_{jk}$ . Then the joint distribution of  $\mathbf{Z}_i = (z_{i1}, \dots, z_{ip})^T$  is given by

$$p(\mathbf{Z}_i | \mathbf{m}_i, \mathbf{A}) = N(\mathbf{m}_i, (\mathbf{I} - \mathbf{A})^{-1} \mathbf{S} (\mathbf{I} - \mathbf{A})^{-T}),$$

where  $\mathbf{m}_i = (m_{i1}, \dots, m_{ip})^T$  and  $\mathbf{S} = \text{diag}(s_1, \dots, s_p)$ . The structure of the path diagram  $\mathcal{G}$  for this SEM is determined by the zero patterns in  $\mathbf{A}$ , i.e.  $a_{jk} = 0$  if and only if there is a missing edge from  $k$  to  $j$ . They define the prior over graph  $\mathcal{G}$  to be a subgraph of some fixed graph  $\mathcal{G}_0 = (V, E_0)$ . The idea is to specify  $\mathcal{G}_0$  as some known maximum pathway. That is,  $p(\mathcal{G} | \psi) = \psi^{|E|} (1 - \psi)^{|E_0| - |E|}$  where  $\psi \sim \text{Beta}(a_\psi, b_\psi)$ . Restricting  $\mathcal{G}$  to subgraphs of  $\mathcal{G}_0$  very effectively incorporates the prior information that might be available about an established biological pathway, using, for example, a public database such as KEGG. This restriction to subgraphs of  $\mathcal{G}_0$  is important to mitigate nonidentifiability due to Markov equivalence between distinct graphs. Moreover, it also greatly reduces the graph search space, which allows for efficient posterior simulation through trans-dimensional Markov chain Monte Carlo (MCMC). Further discussion and detailed treatment of this topic can be found in [29].

The network structure in [29] is built on the latent Gaussian variables which are linked to the observed gene expressions through another layer of latent trinary variables. Therefore it defines an indirect dependence structure on the actual gene expression. [30] extend this work to allow for a more direct characterization of the dependence structure of the observed protein expressions. Particularly, they proposed a fully general dependence prior between latent binary variables and developed a parsimonious framework for Bayesian model determination that allows for local computation.

Telesca et al. [30] start with the same sampling model as in (5) and (6), but then deviate substantially from the earlier model by reducing the trinary variable  $e_{ij}$  to a binary indicator variable  $z_{ij} = 2I(e_{ij} = 1) - 1$ . That is,

$$z_{ij} = \begin{cases} -1 & \text{if protein } j \text{ is inactivated or neutral in sample } i \\ 1 & \text{if protein } j \text{ is activated in sample } i \end{cases}$$

The motivation of this reduction is the simplicity of the prior model on the resulting binary vector  $\mathbf{Z}_i = (z_{ij})$ . In fact, it is possible [8] to describe all possible probability

models  $p(\mathbf{Z}_i)$  that comply with a given conditional independence structure. See below. Also, investigators usually focus on protein activations in RPPA studies. The joint distribution of  $\mathbf{Z}_i = (z_{i1}, \dots, z_{ip})^T$  conditioned on the graph  $\mathcal{G}$  is by an Ising model,

$$p(\mathbf{Z}_i | \mathcal{G}, \alpha, \beta) \propto \exp \left( \sum_{j=1}^p \alpha_j z_{ij} + \sum_{(j,k) \in E^m} \beta_{jk} z_{ij} z_{ik} \right).$$

The model is completed with normal priors for  $\alpha_j$  and  $\beta_{jk}$ . In essence, the approach first maps an RG  $\mathcal{G} = (V, E)$  to its moral graph  $\mathcal{G}^m = (V, E^m)$  and then defines the joint probability model for  $\mathbf{Z}_i$  based on  $\mathcal{G}^m$ . The prior model for the graph  $\mathcal{G}$  is again hinged upon a known reference network  $\mathcal{G}_0 = (V, E_0)$ . However, it is not constrained to be a subgraph of the reference network as before. Instead, they assume  $p(\mathcal{G} | \psi) \propto \psi^{d(\mathcal{G}, \mathcal{G}_0)}$  with  $\psi \sim \text{Beta}(a_\psi, b_\psi)$  where  $d(\cdot, \cdot)$  is a discrepancy measure. They choose  $d(\cdot, \cdot) = |E^c \cap E_0| + \delta |E \cap E_0^c|$  with  $\delta > 1$  which is a weighted sum of edges dropped from and added to the reference graph  $\mathcal{G}_0$ . With  $\delta > 1$ , the discrepancy measure allows for parsimonious inference by imposing a heavier penalty for adding edges than removing edges. A distinctive feature of this construction is that the prior decreases to zero exponentially fast as discrepancy increases, which allows the posterior simulation to be concentrated around the reference network. See [30] for more modeling and inference details of this topic.

### Integrative network analysis from multi-platform genomic data

Different types of molecules in a living cell do not act in isolation; in fact, DNA, RNA and proteins work closely together to carry out various functions for each cell. Studies that consider only one specific type of molecules remain unnecessarily restricted. In [1] they integrate DNA and RNA molecules for a more robust and biologically interpretable estimate of a gene network. In particular, they consider  $p$  gene expressions  $\mathbf{Y} = (Y_1, \dots, Y_p)^T$ , together with corresponding copy number and methylation, collectively denoted by  $\mathbf{X} = (X_1, \dots, X_{2p})^T$ . They then exploit the central dogma of molecular biology that gene expression is produced by transcription from segments of DNA on which the copy number and methylation are measured, but the reverse processes are rare and biologically uninterpretable. The importance of this restriction is that in a network of  $\mathbf{Z} = (\mathbf{Y}, \mathbf{X})$  some edges have fixed pre-determined directions corresponding this observation. This in turn allows us to report meaningful inference on the directions of the remaining edges. Similar idea but in a different context has been proposed by

[32–34]. With this assumption, they then model the regulatory relationships between copy number, methylation and gene expression with the following SEM

$$Y = AY + BX + E \tag{7}$$

where  $A = (a_{ij}) \in \mathbb{R}^{p \times p}$  with zeros on the diagonal,  $B = (b_{ij}) \in \mathbb{R}^{p \times 2p}$  and  $E = (\epsilon_1, \dots, \epsilon_p)^T \sim N_p(0, \Sigma)$ . The SEM in (7) immediately prohibits gene expression from regulating copy number or methylation since there is no  $X$  on the right-hand side of the equation. It also explicitly allows for feedback loops between the genes which are quite common motifs in molecular networks and have key functional roles in many cellular processes such as regulating gene expressions and acting as bistable switches [35]. Such feature is missing in DAG's, UG's and CG's.

Next they assign thresholded priors [36] for  $A$  and  $B$ ,

$$a_{ij} = \tilde{a}_{ij}I(|\tilde{a}_{ij}| > t_i) \text{ and } b_{ik} = \tilde{b}_{ik}I(|\tilde{b}_{ik}| > t_i),$$

for  $i = 1, \dots, p, j \neq i$  and  $k = 2i - 1, 2i$ . The threshold parameter  $t_i$  controls the minimum effect sizes of  $a_{ij}$  and  $b_{ik}$  and is assigned a uniform hyperprior  $t_i \sim \text{Uniform}(0, t_0)$ . Normal priors are given to  $\tilde{a}_{ij}$  and  $\tilde{b}_{ik}$ ,  $\tilde{a}_{ij} \sim N(0, \tau_{ij}^{(a)})$  and  $\tilde{b}_{ik} \sim N(0, \tau_{ik}^{(b)})$  and conjugate hyperpriors  $\tau_{ij}^{(a)} \sim \text{IG}(\alpha_\tau, \beta_\tau)$  and  $\tau_{ik}^{(b)} \sim \text{IG}(\alpha_\tau, \beta_\tau)$ . The thresholded priors enjoy nice theoretical properties as they are closely related to spike-and-slab prior and non-local prior as shown in [36–38]. We refer readers to [1] for a more comprehensive description of the work.

### Reverse engineering gene networks with heterogeneous samples

Genomic data are often heterogeneous in the sense that the samples can be naturally divided into *observed* or *hidden* groups. For example, in a pan-cancer study, patients are normally grouped by the known cancer types whereas when studying a specific cancer, patients can be split into homogeneous latent subtypes. The inferential goals of these two cases are typically quite different. In the case with observed groups, much of the statistical work has been devoted to building models that allow information to be shared across groups. Such approaches improve the statistical power in graph estimation especially for groups with very limited samples by borrowing strength from other groups. For hidden groups, the underlying task becomes a clustering problem which aims to partition subjects into homogeneous clusters. In this section, we introduce two hierarchical RG models that are suitable for heterogeneous datasets. Interested readers can find more details in [38].

Suppose there are  $C$  known groups in the data. Let  $\{s_i = c\}$  be the group indicator for patient  $i$  in group  $c \in \{1, \dots, C\}$ . Denote  $Y_c = \{Y_i; s_i = c\}$  and  $X_c = \{X_i; s_i = c\}$

the set of observations for patients in group  $c$ . Then [38] entertain a group-specific SEM

$$Y_c = A_c Y_c + B_c X_c + E_c$$

where  $A_c = (a_{c,ij}), B_c = (b_{c,ij})$  and  $E_c = (\epsilon_{c,1}, \dots, \epsilon_{c,p})^T \sim N_p(0, \Sigma_c)$  with  $a_{c,ij} = \tilde{a}_{c,ij}I(|\tilde{a}_{c,ij}| > t_i)$  and  $b_{c,ik} = \tilde{b}_{c,ik}I(|\tilde{b}_{c,ik}| > t_i)$ . To link the graphs across groups, they impose multivariate normal priors on the edge strength

$$\tilde{a}_{ij} = (\tilde{a}_{1,ij}, \dots, \tilde{a}_{C,ij})^T \sim N_C(0, \tau_{ij}^{(a)} \Omega),$$

$$\tilde{b}_{ik} = (\tilde{b}_{1,ik}, \dots, \tilde{b}_{C,ik})^T \sim N_C(0, \tau_{ik}^{(b)} \Omega).$$

The covariance matrix  $\Omega = (\omega_{cc'})$  connects edge strength and graph structure across groups. When  $\omega_{cc'}$  is close to  $\pm 1$ ,  $\tilde{a}_{c,ij}$  and  $\tilde{a}_{c',ij}$  tend to have similar absolute values, which gives rise to a high probability that the edge  $i \leftarrow j$  is included or excluded in groups  $c$  and  $c'$  simultaneously since the threshold  $t_i$  is shared across groups. Therefore, graphs from group  $c$  and  $c'$  are more likely to share common edges. On the other hand, when  $\omega_{cc'}$  is close to 0, the association between groups  $c$  and  $c'$  is negligible. They do not fix  $\Omega$ . Instead, they use an inverse-Wishart prior,  $\Omega \sim IW(v, \Psi)$  and let the data dictate how strong the associations should be.

The Bayesian hierarchical model above allows for easy extension to the case where the groups are unknown. Ni et al. [38] simply augment the model with a Dirichlet-multinomial (DM) allocation model on the group indicator  $c|\pi, C \sim \text{Multinomial}(1, \pi_1, \dots, \pi_C)$  and  $\pi|C \sim \text{Dir}(\eta, \dots, \eta)$ . The number of clusters  $C$  is usually unknown; they assume a geometric prior for  $C \sim \text{Geo}(\rho)$ , which eliminates the need to fix  $C$  a priori. They set the covariance matrix  $\Omega$  to be diagonal since the goal of clustering is to split samples into groups with disparate networks rather than encourage similarity across the clusters. Alternatively, one can also set  $\Omega$  to be a Stieltjes matrix (i.e. a positive definite matrix with nonpositive off-diagonal entries) to induce repulsive graphs.

### Results

In this section, we generalize the RG approach of [1] by incorporating protein expressions in modeling the molecular network. Proteins provide important and orthogonal information in reverse engineering the network because they represent the downstream cumulative effect of changes that happen at the DNA and RNA levels and are more directly related to the phenotypical changes in cancer cells. Clinical utilization of genomic data alone show limited benefits, which is partly due to the poor concordance between gene and protein expressions [39, 40]. Many factors such as complex interactions between different types of molecules are responsible for the discrepancy between gene and protein abundance.

The goal of this analysis is to explicitly model such complex interactions. We choose RG because it allows for efficient computation and great interpretability and is suitable to model reciprocal causality such as feedback mechanism.

We use the same notations  $Y = (Y_1, \dots, Y_p)^T$  and  $X = (X_1, \dots, X_{2p})^T$  to denote gene expressions and the matched copy number and methylation. Importantly, additionally, for each gene  $Y_j$ , there is also the associated protein expression, denoted by  $Z_j$  for  $j = 1, \dots, p$ . We extend the SEM in (7) to

$$\begin{aligned} Y &= AY + BX + E_Y \\ Z &= CZ + DY + E_Z \end{aligned} \quad (8)$$

where  $A = (a_{ij}) \in \mathbb{R}^{p \times p}$  and  $C = (c_{ij}) \in \mathbb{R}^{p \times p}$  have zeros on the diagonal,  $B = (b_{ij}) \in \mathbb{R}^{p \times 2p}$ ,  $D = (d_{ij}) \in \mathbb{R}^{p \times p}$  and independent errors  $E_Y \sim N_p(0, \Sigma_Y)$  and  $E_Z \sim N_p(0, \Sigma_Z)$  with diagonal covariance matrices  $\Sigma_Y = \text{diag}(\sigma_1, \dots, \sigma_p)$  and  $\Sigma_Z = \text{diag}(\lambda_1, \dots, \lambda_p)$ . In essence, model (8) defines the following joint probability model for  $(Y, Z)$  given  $X$ ,

$$\begin{aligned} p(Y, Z|X) &= p(Z|Y)p(Y|X), \\ p(Y|X) &= N_p \left\{ (I - A)^{-1}BX, (I - A)^{-1}\Sigma_Y(I - A)^{-T} \right\}, \\ p(Z|Y) &= N_p \left\{ (I - C)^{-1}DY, (I - C)^{-1}\Sigma_Z(I - C)^{-T} \right\}. \end{aligned} \quad (9)$$

The network structure is embedded in the parameters  $A, B, C$  and  $D$ . In this analysis, we restrict our attention to *cis*-regulatory relationships between DNAs and RNAs by constraining  $b_{ij}$  to 0 for  $j \neq 2i - 1$  or  $2i$  for  $i = 1, \dots, p$ . That is, copy number and methylation of gene  $i$  can only affect gene expression  $i$ . The interpretation of model (9) is given as follows:  $b_{i,2i-1} \neq 0$  if and only if copy number  $i$  is associated with its gene expression;  $b_{i,2i} \neq 0$  if and only if methylation  $i$  is associated with its gene expression;  $a_{ij} \neq 0$  if and only if gene  $j$  regulates gene  $i$ ;  $c_{ij} \neq 0$  if and only if protein  $j$  regulates protein  $i$ ; and  $d_{ij} \neq 0$  if and only if gene  $j$  regulates protein  $i$ . We use the thresholded prior for  $A, B, C$  and  $D$ . For error variances, we assume inverse-gamma priors,  $\sigma_j, \lambda_j \sim IG(a, b)$  for  $j = 1, \dots, p$  which imply an inverse gamma conditional posterior distribution which in turn allows the use of a Gibbs sampling transition probability by sampling from the respective inverse gamma. The complete MCMC algorithm is provided in Algorithm 1.

Using TCGA-Assembler ([41], version 2), we acquired, processed and combined TCGA ovarian cancer data, including copy number, methylation and gene expression from the Genomic Data Commons (GDC) and mass spectrometry proteomic samples generated by the Clinical Proteomic Tumor Analysis Consortium (CPTAC). The resulting dataset contains 104 samples. In this case

---

### Algorithm 1

---

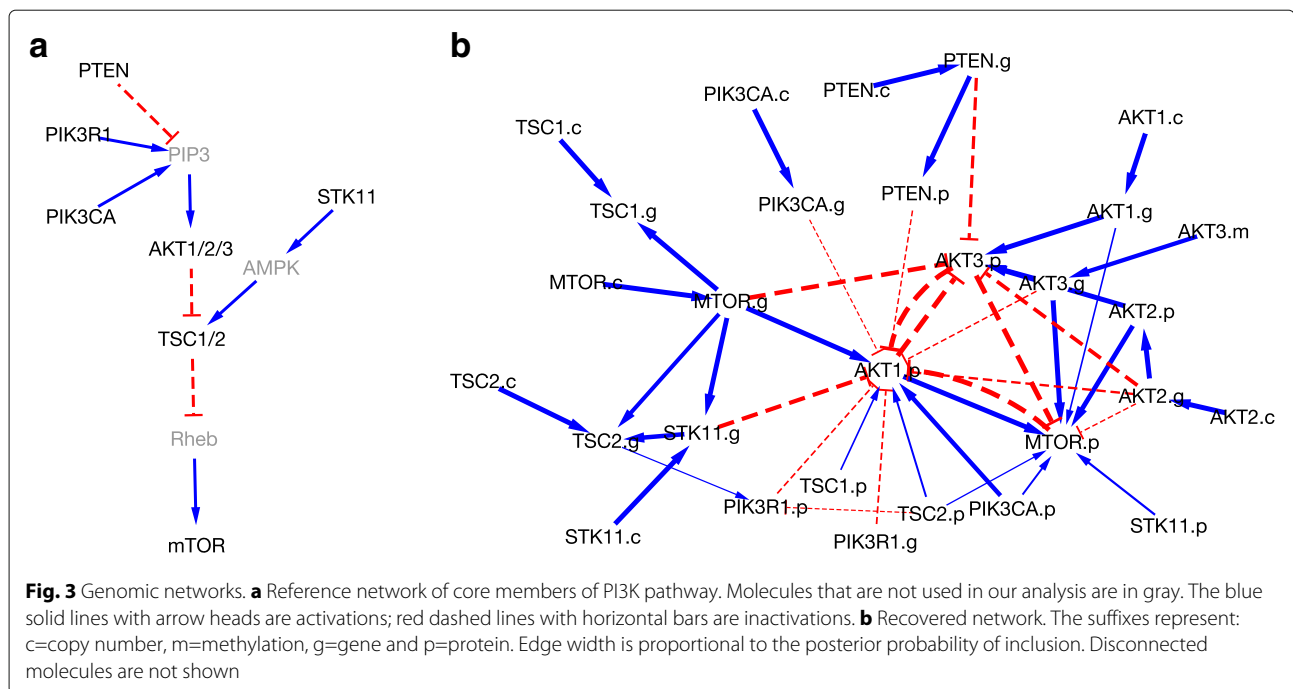
- 1: Initialize  $A = (a_{ij}), B = (b_{ij}), C = (c_{ij}), D = (d_{ij}), t = (t_i)$
- 2: **for**  $i = 1, \dots, p$  **do**
- 3:   update  $\sigma_i, \lambda_i$  by a Gibbs step
- 4:   **for**  $j = 1, \dots, i - 1, i + 1, \dots, p$  **do**
- 5:     update  $\tau_{ij}^{(a)}$  by a Gibbs step
- 6:     update  $\tau_{ij}^{(c)}$  by a Gibbs step
- 7:     update  $a_{ij}$  by a Metropolis-Hastings step
- 8:     update  $c_{ij}$  by a Metropolis-Hastings step
- 9:   **end for**
- 10:   **for**  $j = 2i - 1, 2i$  **do**
- 11:     update  $\tau_{ij}^{(b)}$  by a Gibbs step
- 12:     update  $b_{ij}$  by a Metropolis-Hastings step
- 13:   **end for**
- 14:   **for**  $j = 1, \dots, p$  **do**
- 15:     update  $\tau_{ij}^{(d)}$  by a Gibbs step
- 16:     update  $d_{ij}$  by a Metropolis-Hastings step
- 17:   **end for**
- 18:   update  $t_i$  by a Metropolis-Hastings step
- 19: **end for**
- 20: Repeat steps 2-20 until convergence
- 21: Output posterior samples of  $A, B, C, D, t, \Sigma_Y, \Sigma_Z, \tau$

---

study, we focus on  $p = 10$  genes that are core members of the PI3K pathway [42], as shown in Fig. 3a, by playing a critical role in cell cycle progression, survival, motility, angiogenesis and immune surveillance [43]. It is one of the most frequently altered pathways in ovarian cancer [44, 45]. For each gene, we have matched copy number, methylation, gene expression and protein expression (i.e. 40 nodes in total). We report inference based on 100,000 MCMC samples (thinned out to every 10th iteration) after discarding the first 50,000 iterations as burn-in.

The recovered network is shown in Fig. 3b where the blue solid lines represent activation whereas the red dashed lines are inactivations. The edge width is proportional to the posterior probabilities of inclusion  $p(a_{ij} \neq 0|X, Y, Z)$ ,  $p(b_{ij} \neq 0|X, Y, Z)$ ,  $p(c_{ij} \neq 0|X, Y, Z)$  and  $p(d_{ij} \neq 0|X, Y, Z)$ , which can be approximated by MCMC sample averages, for example,  $p(a_{ij} \neq 0|X, Y, Z) = \frac{1}{L} \sum_{l=1}^L I(a_{ij}^{(l)} \neq 0)$  where the superscript ( $l$ ) denotes the  $l$ th MCMC sample and  $L$  is the total number of samples. The suffixes represent four different type of molecules: c=copy number, m=methylation, g=gene and p=protein. For clarity we do not show disconnected molecules. The degree of each node in Fig. 3b is reported in Table 1. The top three highly connected nodes are the protein AKT1 which is involved in 15 regulatory activities, the protein mTOR with 10 activities and the protein AKT3 with





9 activities. AKT1 and AKT3 belong to the same AKT family which is known to have shown strong oncogenic function and is a key mediator of PI3K pathway function [46]. AKT isoforms are often phosphorylated in ovarian cancers and may play a role in mediating the progression of late-stage serous ovarian carcinomas [47]. mTOR plays a critical role in regulating cell growth and proliferation by integrating signals including growth factors, nutrients, energy and stress [48, 49]. We also capture the well-known positive regulatory relationship between AKT1 and its downstream target mTOR [45, 49]. Other findings that are consistent with the biological literature include the inhibitory relationship between proteins PTEN and AKT1 and stimulatory relationship between proteins PIK3CA and AKT1 [50, 51]. Interestingly, the recovered network includes a negative feedback from proteins of mTOR to AKT1. Recent studies [45, 52–54] confirmed that such feedback mechanism can be either positive or negative depending on many factors including cell types and conditions.

Inference includes some surprising findings. For example, PIK3R1 is expected to activate rather than inactivate AKT1. Although the unexpected result may simply be a false positive, it deserves further experimental validation.

**Discussion and conclusions**

We have reviewed the basics of UG’s, DAG’s and RG’s and discussed recent RG-based approaches in modeling molecular networks. The extension of the approach in [1] allows to integrate multi-platform data including copy number, methylation, gene expression and protein

expression. Our approach can be further be generalized to incorporating other data types such as single nucleotide polymorphisms (SNPs), mutation status and microRNAs. SEM-based RG approaches, however, cannot be directly applied to SNPs and mutation status because they have discrete support and the condition C1 requires the data to be multivariate normal. This limitation can be addressed by imputing latent normal variables. Although we do not find serious violation of normality assumption in our data, if desired, one could

**Table 1** Degrees of molecules from recovered network in Fig. 3b

Molecule	Degree	Molecule	Degree
AKT1.p	15	PTEN.p	2
mTOR.p	10	PIK3CA.p	2
AKT3.p	9	AKT1.c	1
mTOR.g	6	AKT2.c	1
AKT2.g	5	AKT3.m	1
AKT3.g	4	mTOR.c	1
STK11.g	4	PIK3CA.c	1
TSC2.g	4	PTEN.c	1
AKT1.g	3	STK11.c	1
PTEN.g	3	TSC1.c	1
AKT2.p	3	TSC2.c	1
PIK3R1.p	3	PIK3R1.g	1
TSC2.p	3	STK11.p	1
PIK3CA.g	2	TSC1.p	1
TSC1.g	2		

adopt the approach of [29] and introduce additional layers of hidden variables when the data are seemingly non-normal.

Another important assumption of the discussed approaches in this paper is homogeneity across samples. It is an unrealistic assumption in many applications, especially in oncology where the consensus is that tumors are extremely heterogeneous. Characterizing tumor heterogeneity can fundamentally improve our understanding of the cancer biology and practically allow us to divide heterogeneous patient population into homogeneous subpopulations so that refined personalized treatments can be developed targeting specific subgroups of cancer patients. Several approaches have been proposed for modeling networks with heterogeneous samples [38, 55–57].

#### Acknowledgements

The authors appreciate the instructive suggestions from the Editor and anonymous reviewers.

#### Funding

This work and the publication costs of this article were supported by NIH/NCI grant RO1 CA 132897.

#### Availability of data and materials

The TCGA ovarian cancer dataset is publicly available in Genomic Data Commons (<https://gdc.cancer.gov/>).

#### About this supplement

This article has been published as part of *BMC Bioinformatics* Volume 19 Supplement 3, 2018: Selected original research articles from the Fourth International Workshop on Computational Network Biology: Modeling, Analysis, and Control (CNB-MAC 2017): bioinformatics. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-19-supplement-3>.

#### Authors' contributions

YN, PM and YJ conceived the idea of this article. YN carried out the literature reviews, extended the model and drafted the manuscript. PM and YJ made critical revisions. LW and YJ provided the data. All authors read and approved the final manuscript.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, TX 78712, USA. <sup>2</sup>Department of Mathematics, The University of Texas at Austin, Austin, TX 78712, USA. <sup>3</sup>NorthShore University HealthSystem, Evanston, IL 60201, USA. <sup>4</sup>Department of Public Health Sciences, The University of Chicago, Chicago, IL 60637, USA.

Published: 21 March 2018

#### References

- Ni Y, Ji Y, Müller P. Reciprocal Graphical Models for Integrative Gene Regulatory Network Analysis. *Bayesian Anal.* 2017.
- Lauritzen SL, Wermuth N. Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann Stat.* 1989;17(1):31–57.
- Frydenberg M. The chain graph markov property. *Scand J Stat.* 1990;17(4): 333–53.
- Spirtes P. Directed Cyclic Graphical Representations of Feedback Models. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann Publishers Inc.; 1995. p. 491–8.
- Koster JT. Markov properties of nonrecursive causal models. *Ann Stat.* 1996;24(5):2148–77.
- Whittaker J. *Graphical models in applied multivariate statistics*. New York: Wiley Publishing; 2009.
- Lauritzen SL, Vol. 17. *Graphical models*. Oxford: Clarendon Press; 1996.
- Besag J. Spatial interaction and the statistical analysis of lattice systems. *J R Stat Soc Series B.* 1974;36(2):192–236.
- Dobra A, Lenkoski A, Rodriguez A. Bayesian inference for general gaussian graphical models with application to multivariate lattice data. *J Am Stat Assoc.* 2011;106(496):1418–33.
- Green PJ, Thomas A. Sampling decomposable graphs using a markov chain on junction trees. *Biometrika.* 2013;100(1):91.
- Wang H, Li SZ, et al. Efficient gaussian graphical model determination under g-wishart prior distributions. *Electron J Stat.* 2012;6:168–98.
- Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the lasso. *Ann Stat.* 2006;34(3):1436–62.
- Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics.* 2008;9(3):432–41.
- Pearl J. *Probabilistic inference in intelligent systems*. San Mateo: Morgan Kaufmann; 1988.
- Lauritzen SL, Dawid AP, Larsen BN, Leimer HG. Independence properties of directed markov fields. *Networks.* 1990;20(5):491–505.
- Ni Y, Stingo FC, Baladandayuthapani V. Sparse Multi-Dimensional Graphical Models: A Unified Bayesian Framework. *J Am Stat Assoc.* 2017;112(518):779–93.
- Ni Y, Stingo FC, Baladandayuthapani V. Bayesian nonlinear model selection for gene regulatory networks. *Biometrics.* 2015;71(3):585–95.
- Friedman N, Koller D. Being bayesian about network structure: a bayesian approach to structure discovery in bayesian networks. *Mach Learn.* 2003;50(1–2):95–125.
- Shojaie A, Michailidis G. Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika.* 2010;97:519–38.
- Stingo FC, Chen YA, Vannucci M, Barrier M, Mirkes PE. A bayesian graphical modeling approach to microRNA regulatory network inference. *Ann Appl Stat.* 2010;4(4):2024.
- Altomare D, Consonni G, La Rocca L. Objective bayesian search of gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics.* 2013;69(2):478–87.
- Yajima M, Telesca D, Ji Y, Müller P. Detecting differential patterns of interaction in molecular pathways. *Biostatistics.* 2015;16(2):240–51.
- Richardson T. A Discovery Algorithm for Directed Cyclic Graphs. In: *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann Publishers Inc.; 1996. p. 454–61.
- Richardson T. A Polynomial-time Algorithm for Deciding Markov Equivalence of Directed Cyclic Graphical Models. In: *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann Publishers Inc.; 1996. p. 462–69.
- Richardson T. A characterization of markov equivalence for directed cyclic graphs. *Int J Approx Reason.* 1997;17(2–3):107–62.
- Mooij JM, Janzing D, Heskes T, Schölkopf B. On Causal Discovery with Cyclic Additive Noise Models. In: *Proceedings of the 24th International Conference on Neural Information Processing Systems*. USA: Curran Associates Inc.; 2011. p. 639–47.
- Mooij JM, Heskes T. Cyclic Causal Discovery from Continuous Equilibrium Data. In: *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*. Arlington: AUAI Press; 2013. p. 431–9.

28. Lacerda G, Spirtes P, Ramsey J, Hoyer P. Discovering Cyclic Causal Models by Independent Components Analysis. In: Proceedings of the Twenty-Fourth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-08). Corvallis: AUA Press; 2008. p. 366–74.
29. Telesca D, Müller P, Parmigiani G, Freedman RS. *Ann Appl Stat.* 2012;6(2):542–60.
30. Telesca D, Müller P, Kornblau SM, Suchard MA, Ji Y. Modeling protein expression and protein signaling pathways. *J Am Stat Assoc.* 2012; 107(500):1372–84.
31. Parmigiani G, Garrett ES, Anbazhagan R, Gabrielson E. A statistical framework for expression-based molecular classification in cancer. *J R Stat Soc Series B.* 2002;64(4):717–36.
32. Cai X, Bazerque JA, Giannakis GB. Inference of gene regulatory networks with sparse structural equation models exploiting genetic perturbations. *PLoS Comput Biol.* 2013;9(5):1003068.
33. Ni Y, Stingo FC, Baladandayuthapani V. Integrative bayesian network analysis of genomic data. *Cancer Inf.* 2014;13(Suppl 2):39.
34. Zhang L, Kim S. Learning gene networks under snp perturbations using eqtl datasets. *PLoS Comput Biol.* 2014;10(2):1003420.
35. Shin SY, Rath O, Zebisch A, Choo SM, Kolch W, Cho KH. Functional roles of multiple feedback loops in extracellular signal-regulated kinase and wnt signaling pathways that regulate epithelial-mesenchymal transition. *Cancer Res.* 2010;70(17):6715–24.
36. Ni Y, Stingo F, Baladandayuthapani V. Bayesian Graphical Regression. *J Am Stat Assoc.* 2018.
37. Rossell D, Telesca D. Nonlocal Priors for High-Dimensional Estimation. *J Am Stat Assoc.* 2017;112(517):254–65.
38. Ni Y, Müller P, Zhu Y, Ji Y. Heterogeneous reciprocal graphical models. *Biometrics.* 2018.
39. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mrna abundance in yeast. *Mol Cellular Biol.* 1999;19(3):1720–30.
40. Akbani R, Ng PKS, Werner HM, Shahmoradgoli M, Zhang F, Ju Z, Liu W, Yang J-Y, Yoshihara K, Li J, et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat Commun.* 2014;5:3887.
41. Zhu Y, Qiu P, Ji Y. Tcga-assembler: open-source software for retrieving and processing tcga data. *Nat Methods.* 2014;11(6):599–600.
42. Cantley LC. The phosphoinositide 3-kinase pathway. *Science.* 2002; 296(5573):1655–57.
43. Liliac L, Amalinei C, Balan R, Grigoras A, Caruntu ID. Ovarian cancer: insights into genetics and pathogeny. *Histol Histopathol.* 2012;27(6): 707–19.
44. Kotsopoulos IC, Papanikolaou A, Lambropoulos AF, Papazisis KT, Tsolakidis D, Touplikioti P, Tarlatzis BC. Serous ovarian cancer signaling pathways. *Int J Gynecol Cancer.* 2014;24(3):410–7.
45. Cheaib B, Auguste A, Leary A. The pi3k/akt/mtor pathway in ovarian cancer: therapeutic opportunities and challenges. *Chin J Cancer.* 2015;34(1):4.
46. Bast RC, Hennessy B, Mills GB. The biology of ovarian cancer: new opportunities for translation. *Nat Rev Cancer.* 2009;9(6):415–28.
47. Hanrahan AJ, Schultz N, Westfal ML, Sakr RA, Giri DD, Scarperi S, Janikariman M, Olvera N, Stevens EV, She QB. Genomic complexity and akt dependence in serous ovarian cancer. *Cancer Discov.* 2012;2(1):56–67.
48. Wullschlegel S, Loewith R, Hall MN. Tor signaling in growth and metabolism. *Cell.* 2006;124(3):471–84.
49. Liu P, Cheng H, Roberts TM, Zhao JJ. Targeting the phosphoinositide 3-kinase pathway in cancer. *Nat Rev Drug Discov.* 2009;8(8):627–44.
50. Vara JÁF, Casado E, de Castro j, Cejas P, Belda-Iniesta C, González-Barón M. Pi3k/akt signalling pathway and cancer. *Cancer Treat Rev.* 2004;30(2): 193–204.
51. Song MS, Salmena L, Pandolfi PP. The functions and regulation of the pten tumour suppressor. *Nat Rev Mol Cell Biol.* 2012;13(5):283–96.
52. Hay N. The akt-mtor tango and its relevance to cancer. *Cancer Cell.* 2005;8(3):179–83.
53. Dormond O, Madsen JC, Briscoe DM. The effects of mtor-akt interactions on anti-apoptotic signaling in vascular endothelial cells. *J Biol Chem.* 2007;282(32):23679–86.
54. Mead H, Zeremski M, Guba M. mTOR Signaling in Angiogenesis. In: Polunovsky VA, Houghton PJ, editors. *mTOR Pathway and mTOR Inhibitors in Cancer Therapy.* Totowa: Humana Press; 2010. p. 49–74.
55. Ickstadt K, Bornkamp B, Grzegorzczak M, Wiczorek J, Sheriff MR, Grecco HE, Zamir E. Nonparametric bayesian networks. *Bayesian Stat.* 2011;9:283.
56. Rodriguez A, Lenkoski A, Dobra A. Sparse covariance estimation in heterogeneous samples. *Electron J Stat.* 2011;5:981.
57. Mukherjee C, Rodriguez A. Gpu-powered shotgun stochastic search for dirichlet process mixtures of gaussian graphical models. *J Comput Graph Stat.* 2016;25(3):762–88.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

