

POINTS OF SIGNIFICANCE

Bayesian networks

For making probabilistic inferences, a graph is worth a thousand words.

Many physical and biological processes can be naturally modeled as a network of causal influences. When the number of influences is large, interactions between causes and effects can be modeled using Bayesian networks, which combine network analysis with Bayesian statistics. Bayesian networks are widely used in genetic analysis, integration of biological data and modeling signaling pathways^{1,2}. We have already seen how Bayes' theorem is used to infer the probability of a cause when its effect is observed³. This month we provide a brief description of Bayesian networks and how Bayes' theorem is used to propagate information in them.

A Bayesian network is a graph in which nodes represent entities such as molecules or genes. Nodes that interact are connected by edges in the direction of influence; the edge A→B implies that A (the parent) has an effect on B (the child). In general, a Bayesian network is a directed acyclic graph—cycles are not allowed. Importantly, each node has attached to it probabilities that define the chance of finding the node in a given state. Conditional probabilities are used if the state of a node depends on the state of another. These dependencies propagate through the network and influence the probabilities of other nodes, which are updated as new information about the nodes becomes available. Thus, Bayesian networks are also called probabilistic causal models.

Nodes with continuous variables are parameterized using probability functions, and those with discrete variables using probability tables. For example, consider the simple two-node network A→B where A and B are binary variables with two states (N or Y). The table at node A would contain the marginal probability $P(A = Y)$. For simplicity, we'll use "A" to mean $A = Y$ and "a" to mean $A = N$ so that $P(A = Y)$ and $P(A = N)$ can be written more briefly as $P(A)$ and $P(a)$, respectively. By complementarity, $P(a) = 1 - P(A)$. At node B we would have the conditional probabilities $P(B|A)$ and $P(B|a)$ that define how the state of B depends on the state of A. The conditional probability table (CPT) can be completed using complementarity: $P(b|A) = 1 - P(B|A)$ and $P(b|a) = 1 - P(B|a)$. Thus, the marginal table

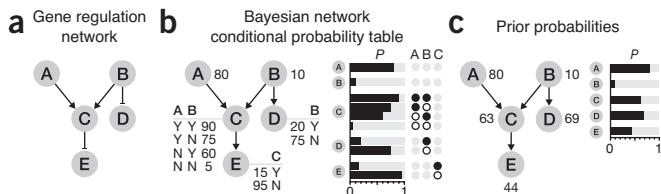


Figure 1 | Bayesian network of regulation between five genes.

(a) A five-gene regulation pathway. A and B activate C. C and B inhibit E, and B inhibits D. (b) Bayesian network representation of the regulation pathway with interactions parameterized as probabilities of the active state. Conditional probabilities for nodes C, D and E describe dependencies on parent nodes. A bar plot of the probability table helps with quantitative comparisons. (c) Prior probabilities calculated from the conditional probabilities in b. All values are probabilities of the active state expressed in percent and rounded to the nearest integer.

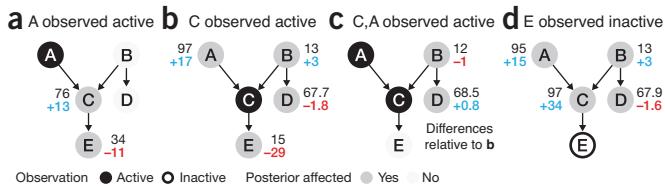


Figure 2 | Observing the state of a node can change the estimate of the states of other nodes. Shown are posterior probabilities updated from the priors from **Figure 1c** (and the difference computed using rounded probabilities except for D, for which these differences are small) using observations about nodes. (a) Observations propagate along serial chains, such as A→C→E. By observing A active, C's posterior $P(C|A) = 76\%$ increases by 13% from the prior $P(C) = 63\%$. B and D are unaffected. (b) Effect of observation can propagate backwards along a path. Observing C affects posteriors of all nodes—its causes and effects. (c) Propagation of information can be altered by observations. Once C is observed, observations about A now influence B and D but no longer influence E. (d) Observing something about an effect changes our estimates of all of its causes.

for A lists all possible states for A, and the CPT lists all possible state combinations of A and B. Once the network is constructed and the probabilities specified, Bayes' theorem is used to propagate probability through the model.

We'll use a hypothetical gene regulation pathway (**Fig. 1a**) to illustrate calculations and inferences in the corresponding Bayesian network, in which genes are modeled as binary variables with a probability of being in an active (Y) or inactive (N) state (**Fig. 1b**). Genes A and B have no incoming edges and their probabilities do not depend on the state of other genes. These genes are therefore characterized by their marginal probabilities $P(A) = 80\%$ and $P(B) = 10\%$. The state of genes C, D and E depends on the state of others, so conditional probabilities are used and reflect that A and B have an activating effect on C (e.g., $P(C|AB) = 90\%$) and that B and C have an inhibitory effect on D and E, respectively (e.g., $P(E|C) = 15\%$). Note that the conditional probabilities for a gene are expressed only in terms of its immediate parent—although A influences E, only the state of C is used in E's CPT.

Using the CPT (**Fig. 1b**), we can compute the prior probabilities for each node³ (**Fig. 1c**). For A and B, these are the observed base rate (80% and 10%, respectively). The prior for C can be calculated by considering the total of all the probabilities of combinations of states of A and B that activate C, which is, $P(C) = P(A)P(B)P(C|AB) + P(a)P(B)P(C|aB) + P(A)P(b)P(C|Ab) + P(a)P(b)P(C|ab) = 63\%$. Similarly, the priors for D and E are $P(D) = 69\%$ and $P(E) = 44\%$.

An important quantity in a Bayesian network is the joint probability distribution, which allows us to calculate the probability of all the nodes being in any given set of states. For example, the probability that all genes in our network are active is very unlikely: $P(ABCDE) = P(A)P(B)P(C|AB)P(D|B)P(E|C) = 0.8 \times 0.1 \times 0.9 \times 0.2 \times 0.15 = 0.2\%$. Because B and C have an inhibitory effect and the chance of B being active is low, a much more likely state is $P(ABCD) = P(A)P(b)P(C|Ab)P(D|b)P(e|C) = 0.8 \times 0.9 \times 0.75 \times 0.75 \times 0.85 = 34\%$.

If we make no additional observations about the gene states and our only source of information about their states is the CPT, the states of A and B are independent because they share no edge or common ancestor. Consequently, knowing the state of A does not change our beliefs about the state of B (**Fig. 2a**). For example, if we observe that A is active, our belief about the state of B being active remains unchanged: $P(B|A) = 0.1$. However, because A influences

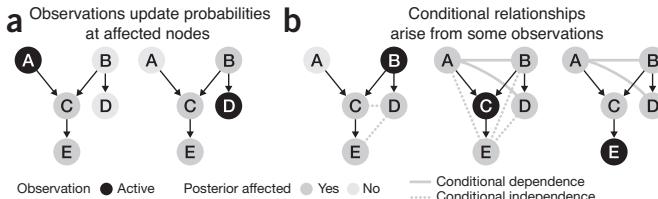


Figure 3 | Observation about child nodes creates conditional dependencies and independencies between its parent nodes on different paths. (a) Observation about A and D does not create new conditional relationships. (b) Observation about B, C and E generates both dependencies and independencies. Observing C or E causes A and B (or D) to influence each other. Observing B or C splits the model and blocks propagation (e.g., D and E become independent).

C, any new knowledge about A requires us to update our estimate of the probability that C is active by calculating the posterior probability. We do so by considering both states of B and obtain $P(C|A) = P(B)P(C|AB) + P(b)P(C|Ab) = 76\%$ (Fig. 2a). Similarly, $P(E|A) = P(C|A)P(E|C) + P(c|A)P(E|c) = 34\%$. In this way Bayes' theorem can integrate the CPT and new observations and propagate probabilities of each node in the direction of the edges, $A \rightarrow C \rightarrow E$.

Influence between nodes can also propagate backwards along an edge. For example, Bayes' theorem for $A \rightarrow C$ is $P(A|C) = P(C|A)P(A)/P(C)$, which tells us that the state of A depends on information about C. By observing C, we can refine our estimation about the state of A—knowing the state of an effect can inform us about the cause. For example, if we observe that C is active, we find $P(A|C) = 0.765 \times 0.8 / 0.63 = 97\%$, an increase of 17% over the prior (Fig. 2b). Here we used $P(C|A) = P(B)P(C|AB) + P(b)P(C|Ab) = 0.1 \times 0.9 + 0.9 \times 0.75 = 0.765$. Similarly, because B is also a parent of C, the posterior for B can be updated to $P(B|C) = 13\%$, which is an increase of 3% (Fig. 2b).

Having information about one node can change how information propagates through other nodes. Above, we saw that A affects C but not B (Fig. 2a). However, if we have information about C, we find that A now affects B, even though they do not share an edge or ancestors (Fig. 2c). This relationship between A and B is called conditional dependence and occurs, for example, between two parent nodes in the presence of information about a common child. In other words, if we know something about the effect (C) and one cause (A), we can say something about the alternative cause (B). Similarly, information about node E induces conditional dependencies in all the nodes.

In the context of our gene network, we can reason about this conditional dependence as follows. If A and B both activate C (Fig. 1a) and we find that C is active, observing A to be active allows us to attribute activation of C to A and thus reduces our belief that B must be active. In other words, $P(B|AC) < P(B|C)$, as seen by the decrease in posteriors of B from 13% to 12% (Fig. 2b,c). We can calculate these posteriors using the conditional variant of Bayes' theorem, $P(B|AC) = P(C|AB)P(B|A)/P(C|A)$. This relationship can be derived from factoring the joint probability $P(ABC) = P(B|AC)P(C|A)P(A) = P(C|AB)P(B|A)P(A)$. Using $P(C|A) = 0.765$ as calculated above, we have $P(B|AC) = 0.9 \times 0.1 / 0.765 = 12\%$ (Fig. 2a). If instead we observe A inactive, then we attribute the activation of C to B and increase our belief that B is active—the posterior $P(B|C) = 13\%$ increases to $P(B|aC) = 57\%$.

New observations can also block the propagation of information down a path. For example, if we make an observation about the state of C, information about the state of A no longer provides information about the state of E (Fig. 2c). In this case, E becomes conditionally independent of A given C and we can write $P(E|AC) = P(E|C)$.

Figure 3 shows cases in which new information creates conditional dependence and independence—relationships not explicitly represented by edges. Information about A and D does not create any such dependencies: in light of new observations, information propagates as before (Fig. 3a). However, observation about B splits the model, and conditional independencies arise—changing D no longer affects C and E (Fig. 3b). Observing C connects A and B as well as A and D by conditional dependencies and disconnects the effects of A, B and D on E (Fig. 3b). Observing E connects A to B and D.

The concept of conditional independence has practical implications when propagating probability in a Bayesian network and gives rise to three types of basic connections: serial, diverging and converging. Depending on the type of connection, new observations about nodes can change the scope of propagation of information. In a serial type of connection (causal chain), propagation can be blocked. As we've seen, information about any of the nodes along $A \rightarrow C \rightarrow E$ updates the others in the chain forward and backward and is limited to the nodes in the chain—altering A affects C and E but not B (Fig. 3a). But if we observe C, E becomes independent of A (Fig. 3b). In a diverging connection ($C \leftarrow B \rightarrow D$), a similar block to propagation can occur. Here, child nodes are related through the parent node, and probability propagates from child to child. However, if evidence is gathered for B, the child nodes become conditionally independent (Fig. 3b). In converging connections such as $A \rightarrow C \leftarrow B$, additional information can actually extend the scope of propagation. If we know something about an intermediate variable (C), the model splits, and each side of the chain turns conditionally dependent given the observed variable—now A has an effect on B (Fig. 3b).

Bayesian networks are statistical tools to model the qualitative and quantitative aspects of complex multivariate problems and can be used for diagnostics, classification and prediction. Time series and feedback loops, common in biological systems, can be modeled by using dynamic Bayesian networks, which allow cycles. One of the most interesting fields where Bayesian networks are used is the identification of 'latent' structures of relations in big databases⁴. Learning a Bayesian network automatically by estimating the nodes, edges and associated probabilities from data is difficult, but it can help to discover unsuspected relations between, for example, genes and diseases.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Jorge López Puga, Martin Krzywinski & Naomi Altman

1. Needham, C.J., Bradford, J.R., Bulpitt, A.J. & Westhead, D.R. *PLoS Comput. Biol.* **3**, e129 (2007).
2. Beaumont, M.A. & Rannala, B. *Nat. Rev. Genet.* **5**, 251–261 (2004).
3. Puga, J.L., Krzywinski, M. & Altman, N. *Nat. Methods* **12**, 277–278 (2015).
4. Nagarajan, R., Scutari, M. & Lèbre, S. *Bayesian Networks in R with Applications in Systems Biology* (Springer, 2013).

Jorge López Puga is a Professor of Research Methodology at Universidad Católica de Murcia (UCAM). Martin Krzywinski is a staff scientist at Canada's Michael Smith Genome Sciences Centre. Naomi Altman is a Professor of Statistics at The Pennsylvania State University.