Network Biology- part IV

Jun Zhu, Ph. D.
Professor of Genomics and Genetic Sciences
Icahn Institute of Genomics and Multi-scale Biology
The Tisch Cancer Institute
Icahn Medical School at Mount Sinai
New York, NY

http://research.mssm.edu/integrative-network-biology/
Email: jun.zhu@mssm.edu

@IcahnInstitute
Biological networks/pathways

Data required to train models

Association networks

Probabilistic causal networks

Biological details revealed

4267 top genes in BxH liver female rescan qtl overlap 
(num(p(GGC)<1e-15)>100  ~abs(cor)>0.5886)

1. How do genes in the same module interact?
2. How do genes in different modules interact?
3. Can we make causal inferences to elucidate signaling pathway for disease targets?
When informed by his doctor of the correlation between fat dogs and their masters, Brian set out immediately to rectify his weight problem.

From Stephen Friend
A simple biological question: are there causal/reactive relationships?
A Bayesian network approach:

Best model
What are Bayesian networks?

- A Bayesian network is an expert system that captures all existing knowledge;
- They are also called belief networks, Bayesian belief networks, causal probabilistic networks;

A Bayesian network consists of
- a directed acyclic graph (a set of nodes and directed edges connecting nodes)—DAG
- A set of conditional probability tables (for discrete data) or probability density functions (for continuous data)
A Bayesian network

DAG

Conditional probability tables

\[ p(C \mid A, B) \]

\[ p(D \mid B) \]

\[ p(E \mid B) \]

\[ p(F \mid C) \]
Bayesian network

- A tree is a Bayesian network
Bayesian network

- A Bayesian network is not a tree
Bayesian network

- **Conventional Notations**

\[ p(A) = \prod_{i} p(A_i \mid pa(A_i)) \]

\( A = \{A_1, A_2, \ldots, A_n\} \) are nodes.

\( p(A) \) is the joint probability of nodes \( A \).

\( pa(A_i) \) are parent nodes of \( A_i \).
Bayesian network

- A diverging structure

out-degree = 4
Bayesian network

- A converging structure

![Diagram of a Bayesian network with nodes A, B, C, and D, and an in-degree of 3.](image)
Bayesian network

Why a DAG is required?

\[ p(A) = \prod_{i} p(A_i \mid pa(A_i)) \]

It is guaranteed that there is a node \( A_j \) in a DAG that has no child.

\[ p(A) = p(A \setminus \{A_j\}) p(A_j \mid A \setminus \{A_j\}) = p(A \setminus \{A_j\}) p(A_j \mid pa(A_j)) = (\prod_{i \neq j} p(A_i \mid pa(A_i))) \ast p(A_j \mid pa(A_j)) \]
Bayesian network: usages

• Bayesian networks can be used to predict outcomes or diagnose causal effects (if structures are known)

• Bayesian networks can be used to discover causal relationships (if structures are not known)
Bayesian network: an example

- A burglar alarm system
Bayesian network: a classifier

• What is a naïve Bayes net

\[
p(A, B, C, D, E) = p(B \mid A)p(C \mid A)p(D \mid A)p(E \mid A)p(A)
\]

\[
p(A \mid B, C, D) = \frac{p(A, B, C, D, E)}{p(B, C, D, E)}
\]
Bayesian network

- How to train a Bayesian network

\[
\begin{array}{c|ccc}
 & A=a_1 & A=a_2 & A=a_3 \\
B=b_1 & 7 & 12 & 25 \\
B=b_2 & 20 & 30 & 28 \\
B=b_3 & 25 & 20 & 6 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
 & A=a_1 & A=a_2 & A=a_3 \\
C=c_1 & 15 & 8 & 20 \\
C=c_2 & 11 & 25 & 18 \\
C=c_3 & 27 & 10 & 16 \\
\end{array}
\]
Bayesian network

- How to construct a Bayesian network? Enumerating possible structures
Bayesian network

• How to construct a Bayesian network? Enumerating all possible structures is impossible

\[ \sim N^N, \quad N \text{ is the number of nodes} \]
Bayesian network

- How to construct a Bayesian network? Heuristic approach
Bayesian network

• How to construct a Bayesian network? Heuristic approach

Parameters to estimate=3x3x3

Parameters to estimate=3x3x3x3
Bayesian network

- How to construct a Bayesian network? Heuristic approach

\[
p(M \mid D) = \frac{p(D \mid M) p(M)}{p(D)}
\]

\[
\text{BIC} = -2 \ln p(D \mid \hat{M}) + k \ln(n)
\]

\(n\) : number of samples

\(k\) : number of parameters to estimate
Bayesian network

- How to construct a Bayesian network? averaging

Zhu et al., PLoS CompBio, 2007
Zhu et al., Nature Genetics, 2008
Bayesian network

- How to construct a Bayesian network? Enforcing DAG after averaging
  1. Calculate shortest distance
  2. Identify loops
  3. Remove the weakest link in a loop
  4. Go to step 1

Zhu et al., PLoS CompBio, 2007
Bayesian network

- How to construct a Bayesian network? Upper limit on in-degree

Parameters to estimate = $3^{n+1}$
Bayesian network

- Continuous vs discrete models

- Discrete model is faster, easier to capture high order interactions

- Any discretization lost information
Bayesian network

- Missing information
Bayesian network

- Biological network is context specific
- Bayesian network is just a snapshot under a specific condition
Other ways to infer causal networks

- Boolean network, Graphic Gaussian model
- Conditional Mutual Information
- ODE model
- Structural equation
modeling by differential equations

\[ \frac{dx}{dt} = f(x) + u \]

- Response function
- Observed states
- Perturbation

modeling by ordinary differential equations (ODE)

- Assume static state \( \frac{dx}{dt} = 0 \)
- Assume linear relationships \( f(x) = Ax \)

\[
Ax + u = 0
\]

↑

regulatory matrix

\[
x + A^{-1}u = 0
\]

↑

response matrix
ODE: advantages and disadvantages

- **Advantages:**
  - Simple
  - Can model feedback loops

- **Disadvantages:**
  - Need large amount of data
  - Need even more data to capture non-linear relationships
Aknowledgements

Zhu lab
Seungyeul Yoo
Eunjee Lee
Li Wang
Luan Lin
Quan Long

Mount Sinai Genomics Institute
Eric Schadt
Bin Zhang
Zhidong Tu
Charles Powell
Patrizia Casaccia

Supported by:

• Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai
• Janssen
• Canary Foundation
• Prostate Cancer Foundation
• NIH
• NCI

Boston University
Avrum Spira
Joshua Campbell

U Washington
Roger Baumgarner

Berkerley
Rachel Brem

Princeton
Lenoid Kruglyak