

Rocky Mountain Bioinformatics Conference

De Novo Signaling Pathway Reconstruction From Multiple Data Sources

Dongxiao Zhu
Stowers Institute for Medical Research

Existing Approaches

Genetic Epistasis Analysis (Avery and Wasserman, 1992)

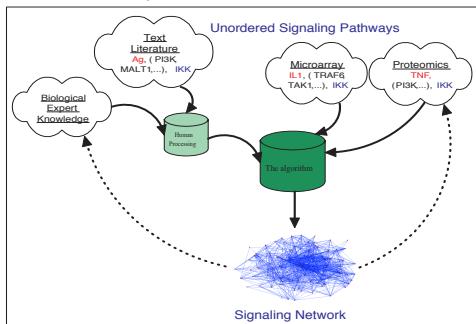
- What kind of phenotype should be measured
- How to quantify this phenotype
- Need exponentially increasing mutants

New Epistasis Analysis (Van Driessche et al, 2005)

Computational Approach (Liu and Zhao, 2004)

- Score pathway permutations
- Gene expression data: test if the correlation is significantly higher than a random pair
- Protein-protein interaction data: binomial distribution parameter (false negative) is estimated from DIP

Problem Formulation and Assumptions



• Formulation

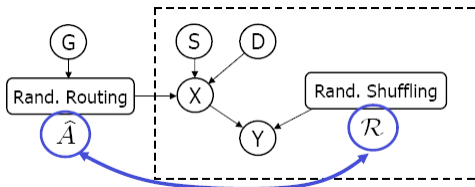
- $\mathbf{X} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(T)})$ are the *incomplete data*: observed T pathways composition.
- $\mathcal{R} = (\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, \dots, \mathbf{r}^{(T)})$ are the *augmented data*: T hidden permutation matrixes.
- $\mathcal{Y} = (\mathbf{X}, \mathcal{R})$ are the *complete data*.
- Markov chain parameters: initial state distribution π and transition matrix \mathbf{A} .

• Assumptions

- Pathways are independently generated by the Markov chain.
- Uniform priors on random permutations

Maximum likelihood estimation via the EM algorithm

- **E-step:** Compute $Q(\mathbf{A}|\hat{\mathbf{A}}) = E[\log P[\mathcal{Y}, \mathcal{R}|\mathbf{A}, \boldsymbol{\pi}]|\mathcal{Y}, \hat{\mathbf{A}}, \boldsymbol{\pi}]$, i.e. average over permutations weighted by fitness with current estimate of \mathbf{A} .
- **M-step:** Solve $\hat{\mathbf{A}} = \operatorname{argmax}_{\mathbf{A}} Q(\mathbf{A}|\hat{\mathbf{A}})$, i.e. update estimate of \mathbf{A} based on expected complete data log-likelihood.



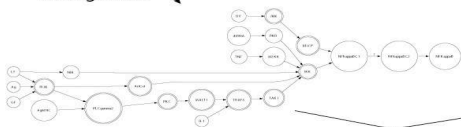
(Source: Nettomo presentation by Michael Rabbat, University of Wisconsin, Madison)

Example: Pathway Components Order Reconstruction

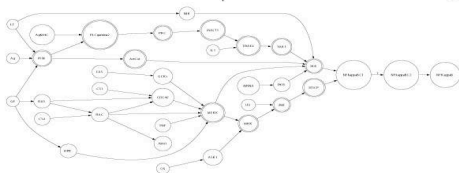
NFκB Pathway

Ag (PKC, PI3K, PLCγ2, MALTI, TAK, TAB1/2, IKK, TRAF6, NFκBC2, NFκBC1), NFκB
 Ag MHC, (TRAF6, PLCγ2, MALTI, TAK, TAB1/2, PKC, IKK, NFκBC1, NFκBC2), NFκB
 IL-1, (IKK, TRAF6, TAK, TAB1/2, NFκBC1, NFκBC2), NFκB
 dRNA, (NFκBC1, IκB, IKK, NFκBC2), NFκB
 TNF, (IKK, MEKK, NFκBC1, NFκBC2), NFκB
 GF, (AKT, COT, IKK, PI3K, NFκBC2, NFκBC1), NFκB
 LT, (PI3K, IKK, NFκBC1, NFκBC2, AKT, COT), NFκB
 LT, (IKK, NIK, NFκBC1, NFκBC2), NFκB
 UV, (bTLCP, NFκBC1, NFκBC2, JNK), NFκB

EM Algorithm



Assemble



SAPK/JNK Pathway

GF1, (HPK, MKK, MEKK), JNK
 GF2, (HPK, EKK, MKK), JNK
 GF3, (RAC, RAS, MEKK, MKK), JNK
 GF4, (RAS, CDC42, RAC, MEKK, MKK), JNK
 GF5, (RAS, RAC), RHO
 CS1, (CDC42, RAC, MEKK, MKK), JNK
 CS2, (MEKK, RAC, MKK), JNK
 FASL, (GSK3, MEKK, MKK), JNK
 OS, (ASK1, MEKK, MKK), JNK

EM Algorithm

