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
- \( \mathbf{X} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(T)}) \) are the incomplete data: observed \( T \) pathways composition.
- \( \mathbf{R} = (r^{(1)}, r^{(2)}, \ldots, r^{(T)}) \) are the augmented data: \( T \) hidden permutation matrixes.
- \( \mathbf{Y} = (\mathbf{X}, \mathbf{R}) \) are the complete data.
- Markov chain parameters: initial state distribution \( \pi \) 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 $Q(A|\hat{A}) = E[\log P[Y, R|A, \pi]|Y, \hat{A}, \pi]$, i.e. average over permutations weighted by fitness with current estimate of $A$.

- **M-step:** Solve $\hat{A} = \arg\max_A Q(A|\hat{A})$, i.e. update estimate of $A$ based on expected complete data log-likelihood.

(Source: Nettomomo presentation by Michael Rabbat, University of Wisconsin, Madison)
Example: Incorporating Prior Information (SAPK/JNK signaling pathway)

\[ \text{GF, RAS, CDC42, MEKK, MKK, JNK, RAC, RHO, HPK, CS1, CS2, FASL, GCKs, OS, ASK1} \]
Example: Pathway Components Order Reconstruction

**NFkB Pathway**
- Ag (PEC, PI3K, PLCγ, MALT1, TAK1, TAB1/2, IKK, TRAF6, NFκB2, NFκB1, NFκB)
- Ag. MHC (TRAF6, PLCγ2, MALT1, TAK1, TAB1/2, PEC, IKK, NFκB1, NFκB2, NFκB)
- IL-1 (IKK, TRAF6, TAK1, TAB1/2, NFκB1, NFκB2, NFκB)
- iNOS (NFκB1, PEC, IKK, NFκB2), NFκB
- TNF (IKK, MEKK, NFκB1, NFκB2, NFκB)
- GF (AKT, COT, IKK, PI3K, NFκB2, NFκB1, NFκB)
- LT (PI3K), IKK, NFκB1, NFκB2, AKT, COT, NFκB
- LT (IKK, NIK, NFκB1, NFκB2), NFκB
- UV, (6TCD, NFκB1, NFκB2, JNK), NFκB

**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
- FAS, (GCK, MEKK, MKK), JNK
- OS, (ASK1, MEKK, MKK), JNK

**EM Algorithm**

![Pathway Diagram](image)

**Assemble**