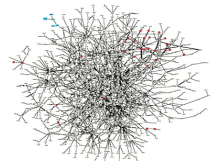


Network Evolution

Only topology of networks will be considered. I.e. dynamics and continuous parameters often ignored.

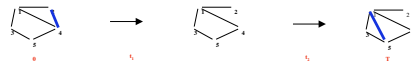


Protein Interaction Networks (PINs) are graphs where the nodes are labeled with protein names. Two nodes are connected if the proteins stick to each other.

(PINs) do not have a temporal dynamic

What do models of network evolution do?:

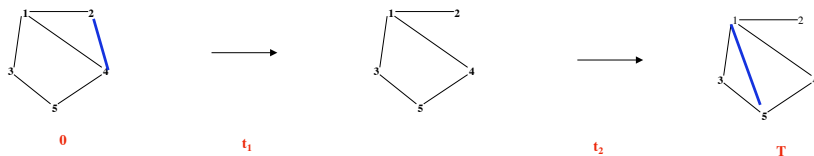
- Test models
- Estimate Parameters in the Evolutionary Process
- Ancestral Analysis
- Framework for Knowledge Transfer



Overview of today's lecture: General considerations in transforming one network into another
Facts and Models for the major networks

- Metabolism**
- Regulatory
- Signal Transduction
- Protein Interaction**
- Combining Inference and Evolution

Evolving Networks: Integration



Integrate of all waiting times (t_1, \dots, t_i) and state assignments of length i gives probability of specific trajectory

$$P(N \rightarrow \dots N_{i-1} \rightarrow N^i) = \int_{t_1 \rightarrow t_i} P(N \rightarrow \dots N_{i-1} \rightarrow N^i; t_1, \dots, t_i) d\vec{t}$$

The above expression can be shown to be of the form $\prod_{n=1}^N \sum_{m=0}^M e^{-q_{10}t} \sum_{k=0}^{d_n} C_n^k T^k$
And recursions $O(N^2)$ exists to calculate coefficients.

Sum over i state assignments gives probability of paths of length i .

$$P(N \rightarrow N^i; i \text{ steps}) = \sum_{N_1, \dots, N_{i-1}} P(N \rightarrow \dots N_{i-1} \rightarrow N)$$

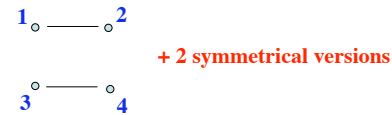
Sum over all path lengths gives probability of N turning into N'

$$P(N \rightarrow N') = \sum_i P(N \rightarrow N'; i \text{ steps})$$

Kulkarni, D. (2005) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 92, 711-732. doi:10.1017/S0007102405000344

Likelihood of Homologous Pathways

Number of Metabolisms:



n	Number of all graphs with n nodes	Number of states
1	1	1
2	2	2
3	8	8
4	64	61
5	1024	969
6	32768	31738
7	2097152	2069964
8	268435456	267270033
9	68719476736	68629753641
10	35184372088832	35171000942698

$$P_{\Theta}(\text{graph}_1, \text{graph}_2) = P_{\Theta}(\text{graph}_1) P_{\Theta}(\text{graph}_1 \rightarrow \text{graph}_2)$$

Approaches:

- Continuous Time Markov Chains with computational tricks.
- MCMC
- Importance Sampling

Eleni Giannoulitou

Evolving Networks: MCMC

Present pathway:



Insertion of an edge pair



Deletion of an edge pair

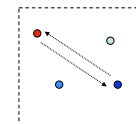


Moving of a pair or singles



Metropolis-Hasting integrating of all paths - Green (1995) version:

Set of paths:



Likelihood - $L(\bullet)$
Probability of going from \bullet to \bullet - $q(\bullet, \bullet)$
 J - Jacobian

$$\text{Acceptance ratio} = \frac{L(\bullet)q(\bullet, \bullet)}{L(\bullet)q(\bullet, \bullet)} J$$

$P(N_1 \rightarrow N_2)$ and Corner Cutting

- How many networks could be visited on "almost shortest" paths?



If $d(N_1, N_2) = k$, then there are 2^k networks are visitable on shortest paths. If 2ϵ additional steps are allowed, then $2^k (L + L(L-1)/2 + (L(L-1) \dots (L-\epsilon+1)/\epsilon!)$ are visitable.

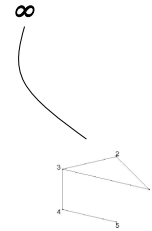
Example. 15 nodes, $L=105$, $\lambda t = \mu t = 0.05$, $\epsilon = 2$, $d=4$. $P(4) = e^{-5} \cdot 5^4 / 4! \approx .003$ $P(6) = e^{-5} \cdot 5^6 / 6! < 10^{-4}$

How can $P(\infty)$ be evaluated?

Can be found in $P(\infty)$ at appropriate rows.
In general not very useful (number of metabolisms).

Simulations

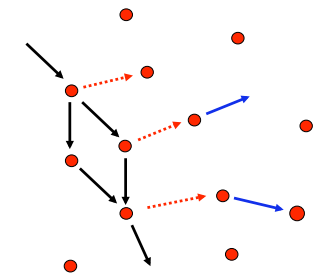
- Forward with symmetries could be used in specific cases.
- Backward (coupling from the past)



Old Hagmann (2002) Finite Markov Chains and Algorithms: Application. Cambridge University Press. Lyapunov, R. V.S. Song and J.J.Hsin (2000) "Accurate Computation of Likelihoods in the Coalescent with Recombination via Parameters" In press Research

A Model for the Evolution of Metabolisms

- A given set of metabolites: ●
- A given set of possible reactions - arrows not shown.
- A core metabolism: →
- A set of present reactions - **M**
black and red arrows



Restriction R:

- A metabolism must define a connected graph
- M + R** defines
- 1. a set of deletable (dashed) edges $D(M)$: - - - - -
- 2. and a set of addable edges $A(M)$: →

Let μ be the rate of deletion
 λ the rate of insertion

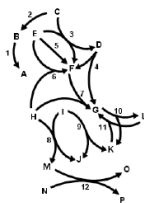
Then

$$\frac{dP(M)}{dt} = \lambda \sum_{M' \in D(M)} P(M') + \mu \sum_{M'' \in A(M)} P(M'') - P(M)[\lambda|D(M)| + \mu|A(M)|]$$

A Toy Example

(by Aziz Mithani)

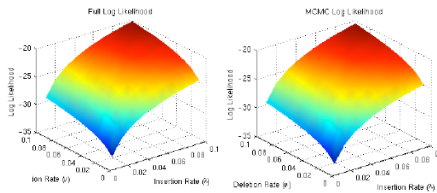
- Metabolic Universe
- 12 possible edges
- 1i 1u 3
- 1i 2u 3
- 2u 1i 3
- 2i 2u 3



Equilibrium Probability

Transition Probability

dist=6



Transition Probability:

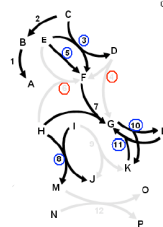
Full Exponentiation (2^{12} states 4096)

Exponentiation with corner cutting
 $2^6 - 64, 384, 960, 1280, 960, 384, 64$

MCMC Integration

Adding Connectedness

Favouring insertions connecting

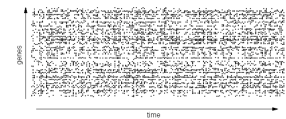


The proportion present: $\frac{5}{7} = 0.714$

Regulatory Network Evolution

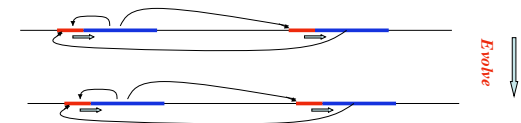
Artificial Genome
Riel, 1999:

- Regulatory control according to rules
- Proteins can bind the regulatory regions



Evolving Artificial Genome
Quant & Bullocks, 2007:

- Selection will influence final dynamics



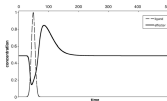
R. Senoguz & CA Szepeski (1996) Modelling the Complexity of Genetic Networks. Complexity 1, 45-64. Metabolic stability and epigenesis in randomly constructed genetic networks. Journal of Theoretical Biology 176, 467-478. A. Kaufman, T. Barkai, & R. Milojevic (2007) Dynamics of Gene Expression in an Artificial Genome: Implications for Biological and Artificial Systems. PLoS ONE 2(12): e1200. Quant & Bullock, 2007. Modeling the evolution of genetic regulatory networks. Journal of Theoretical Biology 238 (4), pp 717-737. Bullock, M.M., Lacombe, N.M., Azevedo, L., Garçon, M. & Kauffman, A. (2004) Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14, 20-27. The evolution of genetic networks by nonadaptive processes. Michael Lynch

Networks: Signal Transduction Pathways

Dynamics



	receptor	protein 1	effector
receptor	0.000	-0.986	0.007
protein 1	0.020	0.000	-0.040
effector	-0.733	0.726	0.000



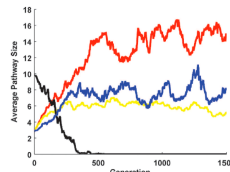
One protein is receptor, one effector.
Activating receptor creates cascade effect described by simple equation system.

$$\frac{d[P_i]}{dt} = [P_i^* \sum_j l_{ij} [P_j^*] - [P_i] (\delta_{in} [L] \sum_j k_{ij} [P_j^*])]$$

Mutational Process: recruitment/loss + change of interactions

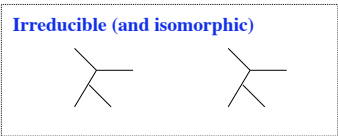
Fitness $F = 1 - nc$ if $\alpha = 1$ n - number of proteins, c - fitness cost per protein,
 $F = 0$ if $\alpha = 0$ a - functionality criteria

Evolution



Reverse Engineering Biological Networks: Opportunities and Challenges. In Computational Methods for Pathway Inference. Am. N.Y. Acad. Sci. 11151: 22-50 (2007).
 Simulating the evolution of signal transduction pathways. BIOLOGICAL CYBERNETICS 39(4): 247-253 (2005).
 A novel method for signal transduction network inference from indirect experimental evidence. Journal of Computational Biology 14:927-949, 2007.
 Evolution of complexity in signaling pathways. Proc Natl Acad Sci USA 2006;103:10372-10374

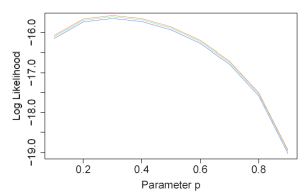
Likelihood of PINs



- Can only handle 1 graph.
- Limited Evolution Model



2386 nodes and 7221 links

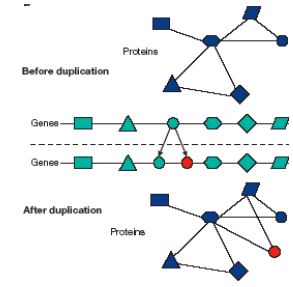


$\theta_0 = (1, .66, .33, 0)$

Models of Protein Interaction Networks Evolution

Barabasi & Oltvai, 2004 & Berg et al., 2004; Winf et al., 2006

- A gene duplicates
- Inherits its connections
- The connections can change



Berg et al., 2004:

- Gene duplication slow $\sim 10^{-9}$ /year
- Connection evolution fast $\sim 10^{-6}$ /year
- Observed networks can be modeled as if node number was fixed.

Inference and Evolution

$$P(D_{mouse}, D_{human}) = \sum_{N_1, N_2} P(D_{human} | N_{human}) P(D_{mouse} | N_{mouse}) P(N_{human}, N_{mouse})$$

