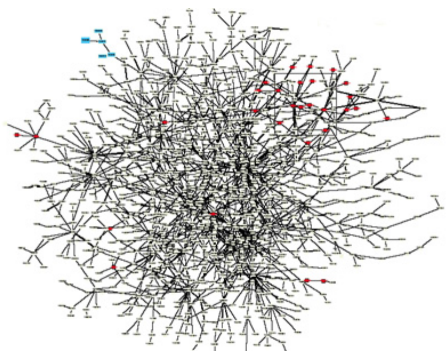
Network Evolution

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

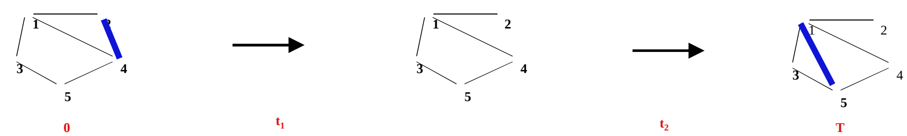


Protein Interaction Networks (PINs) are graphs where the nodes are labeled with protein names. Two nodes are connected if the proteins stick to each other.

(PINs) do not have a temporal dynamic

Yeast Protein Interaction Network from http://www.visualcomplexity.com

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

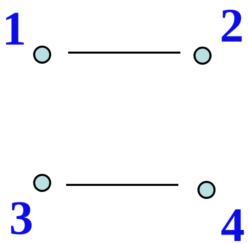
Facts and Models for the major networks

- ***Metabolism***
- *Regulatory*
- *Signal Transduction*
- ***Protein Interaction***

Combining Inference and Evolution

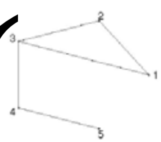
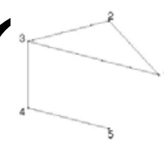
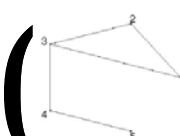
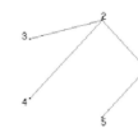
Likelihood of Homologous Pathways

Number of Metabolisms:



+ 2 symmetrical versions

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

$P_\Theta($  ,  $) = P_\Theta($  $) P_\Theta($  $->$  $)$

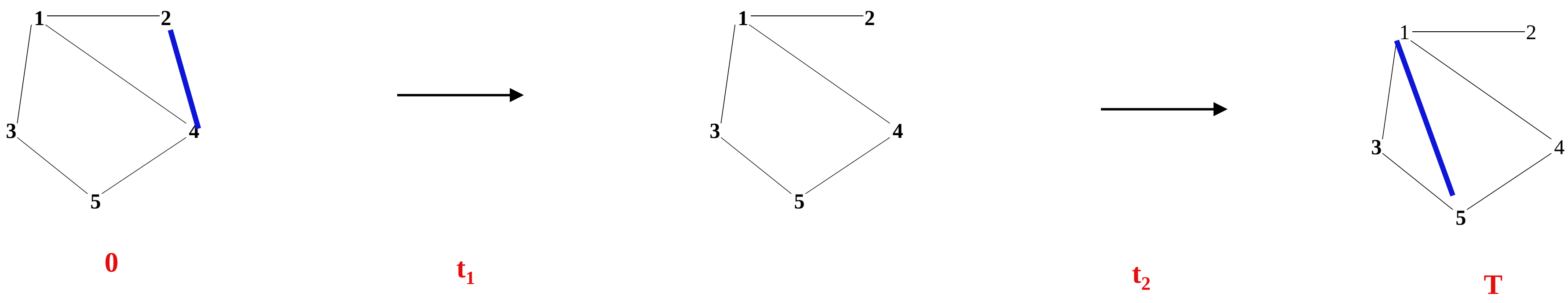
Approaches:

Continuous Time Markov Chains with computational tricks.

MCMC

Importance Sampling

Evolving Networks: Integration



- ***Integrate of all waiting times (t_1,..,t_i) and state assignments of length i gives probability of specific trajectory***

$$P(N \to ..N_{i-1} \to N') = \iiint_{t_1,..t_i} P(N \to ..N_{i-1} \to N; t_1,..t_i) d\bar{t}$$

- ***The above expression can be shown to be of the form*** $\prod_{n=1}^{N} \sum_{n=0}^{M} e^{-q_{i0}T} \sum_{k=0}^{d_n} c_n^k T^k$
  And recursions $O(N^2)$ exists to calculate coefficients.

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

$$P(N \to .N'; i\ steps) = \sum_{N_1, N_2,..N_i} P(N \to ..N_{i-1} \to N)$$

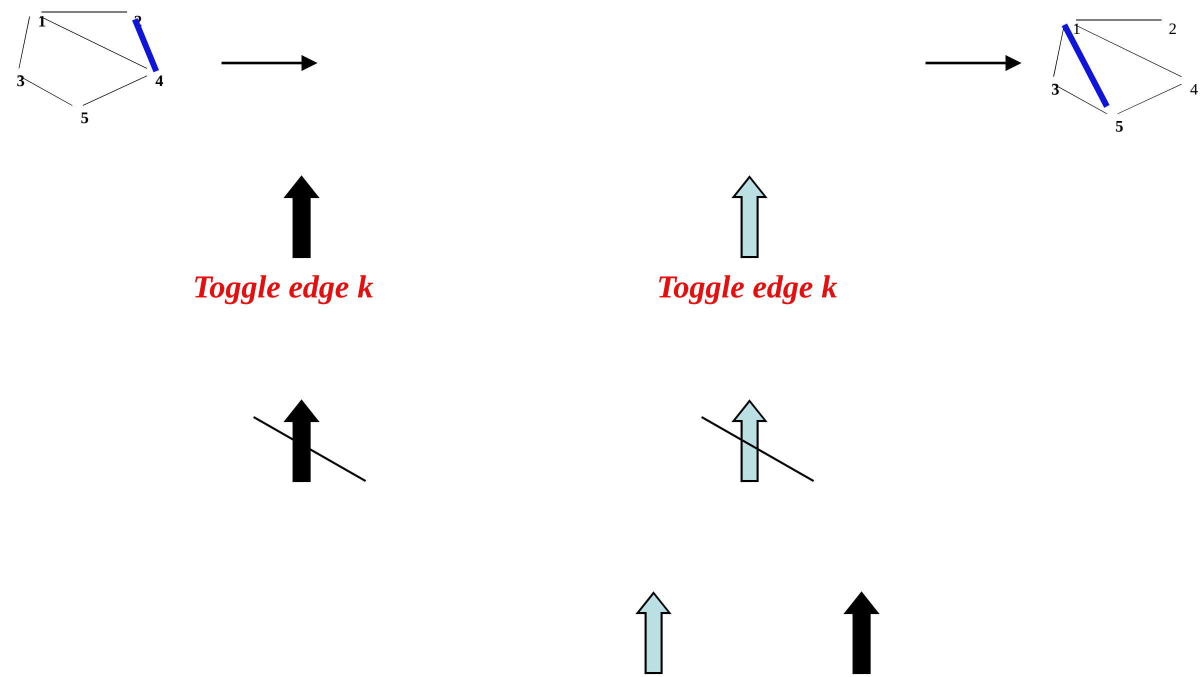
- ***Sum over all path lengths gives probability of N turning into N'***

$$P(N \to .N') = \sum_{i} P(N \to .N'; i\ steps)$$

Koskinen,J. (2004) Bayesian Inference for Longitudinal Social Networks. Research Report, number 2004:4, Stockholm University, Department of Statistics. Koskinen,J. and Snijders,T. (2007) Bayesian inference for dynamic social network data, Journal of Statistical Planning and Inference, 137, 3930--3938. R. Sharan, T. Ideker, *Modeling cellular machinery through biological network comparison*, Nature Biotechnology, 24, 427 (2006). Snijders, T. (2001) "Statistical evaluation of social networks dynamics" in Sociological Methodology By Michael Sobel Snijders, T. et al. (2008) "Maximum Likelihood Evaluation for Social Network Dynamics" In press I. Miklos, G.A. **Lunter** and I. Holmes (2004) A "**long indel**" model for evolutionary sequence alignment. Mol. Biol. Evol. 21(3):529-540. Appendix A

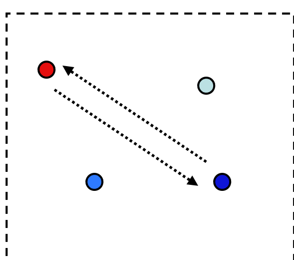
Evolving Networks: MCMC

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:

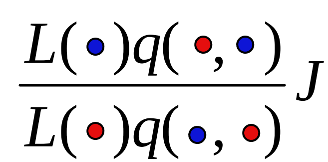


Likelihood - L(●)

Probability of going from ● to ● - q(● ,●)

J - Jacobian

Acceptance ratio



Green, P. J. (**1995**) Reversible jump Markov chain Monte Carlo computation and Bayesian model **determination**, **Biometrika**, **82**, 711-732

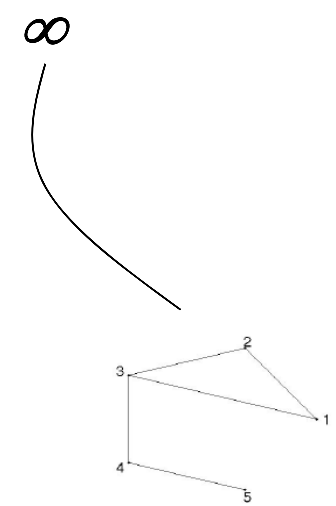
$P(N_1 \text{-->} N_2)$ and Corner Cutting

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

If d(N1,N2) = k, then there are 2^k networks are visitable on shortest paths. If 2ε additional steps are allowed, then 2^k $(L + L(L-1)/2 + (L(L-1)..(L-\varepsilon+1)/\varepsilon!)$ are visitable.

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

How can P() 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)

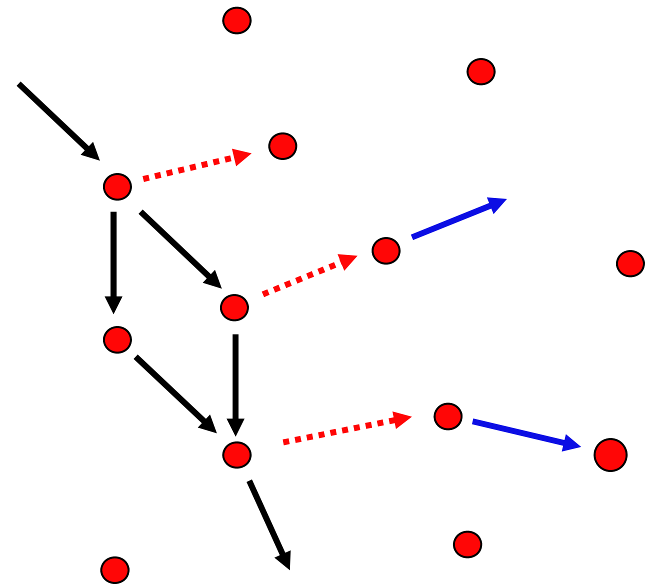
Olle Haegstroem (2002) *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press Lyngsø R, Y.S.Song and J.J.Hein (2008) "Accurate Computation of Likelihoods in the Coalescent with Recombination via Parsimony " In press Recomb

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)

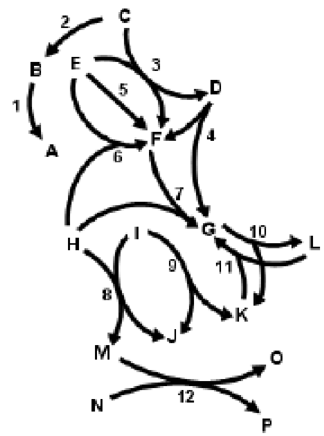
- ***Metabolic Universe***
- ***12 possible edges***

1i 1u 3

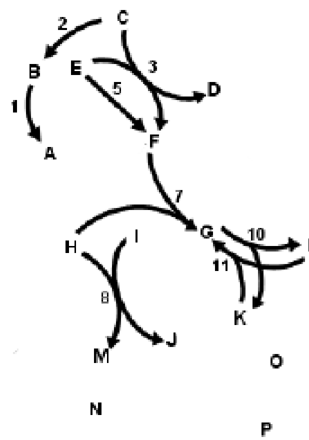
1i 2u 3

2u 1i 3

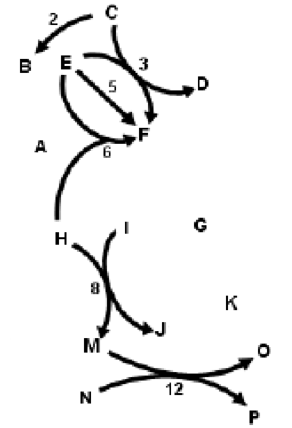
2i 2u 3



Equilibrium Probability



Transition Probability



dist=6

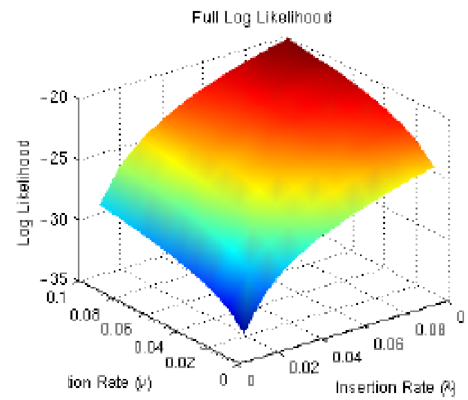
Transition Probability:

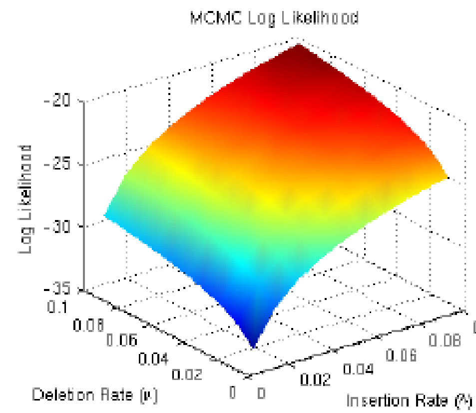
Full Exponentiation (2^{12} states 4096)

Exponentiation with corner cutting

2^6 - 64, 384, 960, 1280 ,960, 384, 64

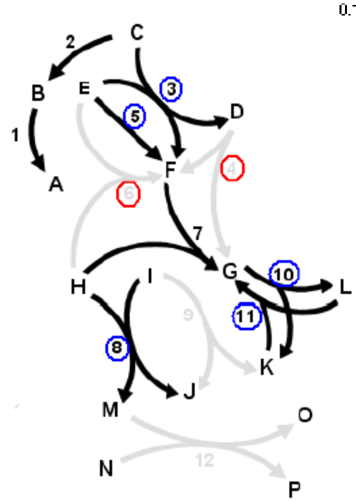
MCMC Integration





Adding Connectedness

Favouring insertions connecting

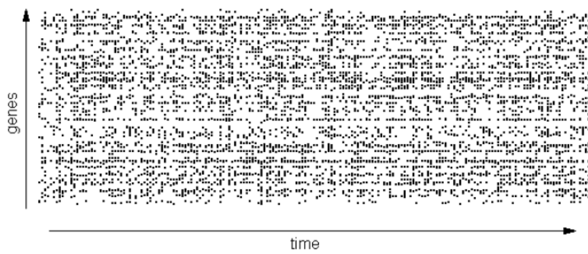


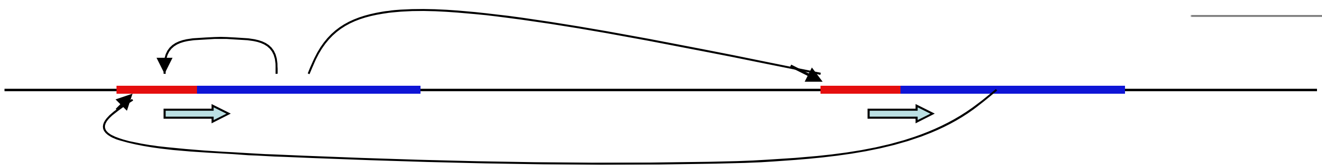
The proportion present: $\frac{5}{7} = 0.714$

Regulatory Network Evolution

Artificial Genome Riel, 1999:

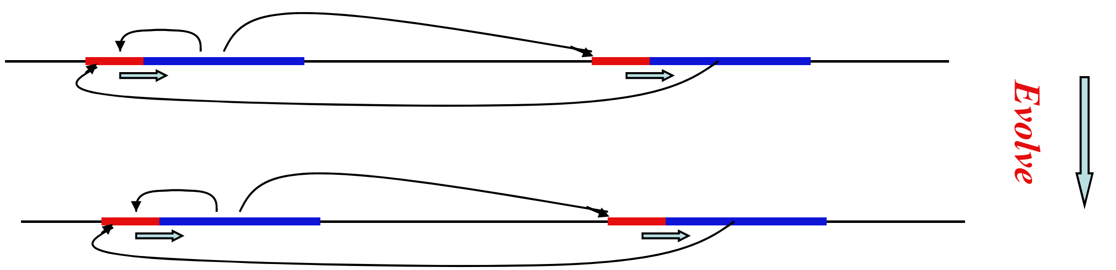
- ***Regulatory control according to rules***
- ***Proteins can bind the regulatory regions***





Evolving Artificial Genome Quant & Bullocks, 2007:

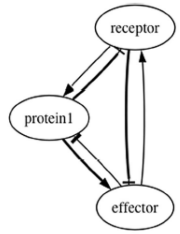
- ***Selection will influence final dynamics***



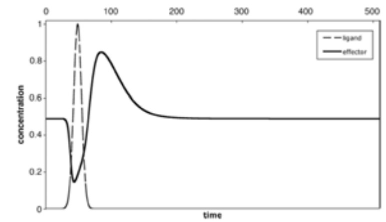
R.Somogyi & CA Sniegoski (1996) Modelling the Complexity of Genetic Networks Complexity 1.6.45-64. **Metabolic stability and epigenesis in randomly constructed genetic nets***Journal of Theoretical Biology,/\Volume 22, Issue 3,/\March 1969, Pages 437-467*S. A. **Kauffman****Torsten Reil**: Dynamics of Gene Expression in an Artificial Genome - Implications for Biological and Artificial Ontogeny. ECAL **1999**: 457-466 **Quayle**, A. and **Bullock**, S. (**2006**) Modelling the evolution of genetic regulatory networks. Journal of Theoretical Biology, 238 (4). pp. 737 Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. & Teichmann, S.A. (2004) Structure and evolution of transcriptional regulatory networks. *Curr. Op. Struc. Biol.*, **14**, 283-291753. Nat Rev Genet. 2007 Oct ;8 (10):803-13 17878896. The **evolution** of genetic networks by non-adaptive 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.

$$\frac{d[P_i]}{dt} = [P_i^* \sum_j l_{ij}[P_j^*] - [[P_j](\delta_{i1}[L] \sum_j k_{ij}[P_j^*])]$$

- ***Mutational Process:*** *recruitment/loss + change of interactions*

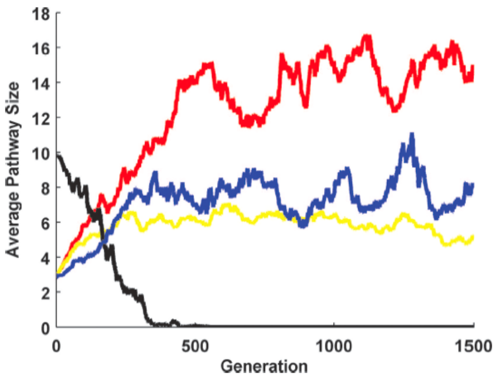
- ***Fitness***

$F = 1 - nc$ if $\alpha = 1$

$F = 0$ if $\alpha = 0$

n - number of proteins, c - fitness cost per protein,
a - functionality criteria

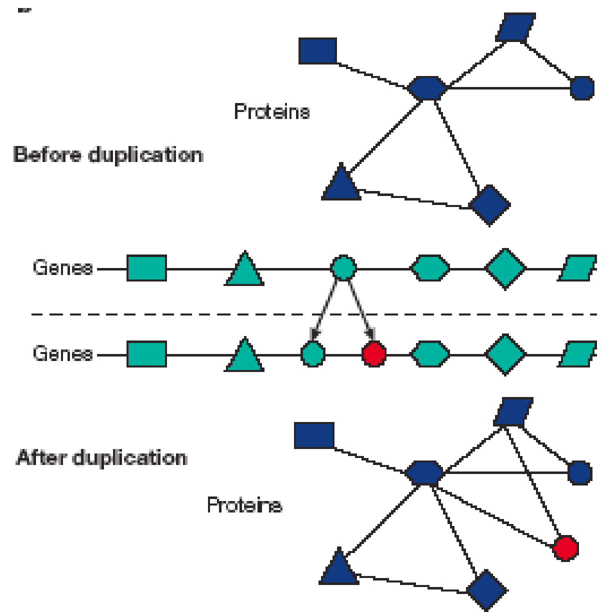
- ***Evolution***



Reverse Engineering Biological Networks: Opportunities and Challenges in Computational Methods for Pathway Inference Ann. N.Y. Acad. Sci. 1115: 32-50 (2007). ARNOLD J. LEVINE, WENWEI , ZHAOHUI FENG AND GERMAN GIL Soyer, OS, Pfeiffer, T, **Bonhoeffer**, S Simulating the **evolution** of signal **transduction** pathways JOURNAL OF THEORETICAL BIOLOGY (2006) 241:223-232. R. Albert, B. DasGupta, R. Dondi, S. Kachalo, E.D. Sontag, A. Zelikovsky, and K. Westbrooks. **A novel method for signal transduction network inference from indirect experimental evidence**. *Journal of Computational Biology*, 14:927-949, 200 7 **Soyer** OS, **Bonhoeffer** S. Evolution of complexity in signaling pathways. Proc Natl Acad Sci USA. 2006;103:16337–16342

Models of Protein Interaction Networks Evolution

Barabasi & Oltvai, 2004 & Berg et al. ,2004; Wiuf etal., 2006

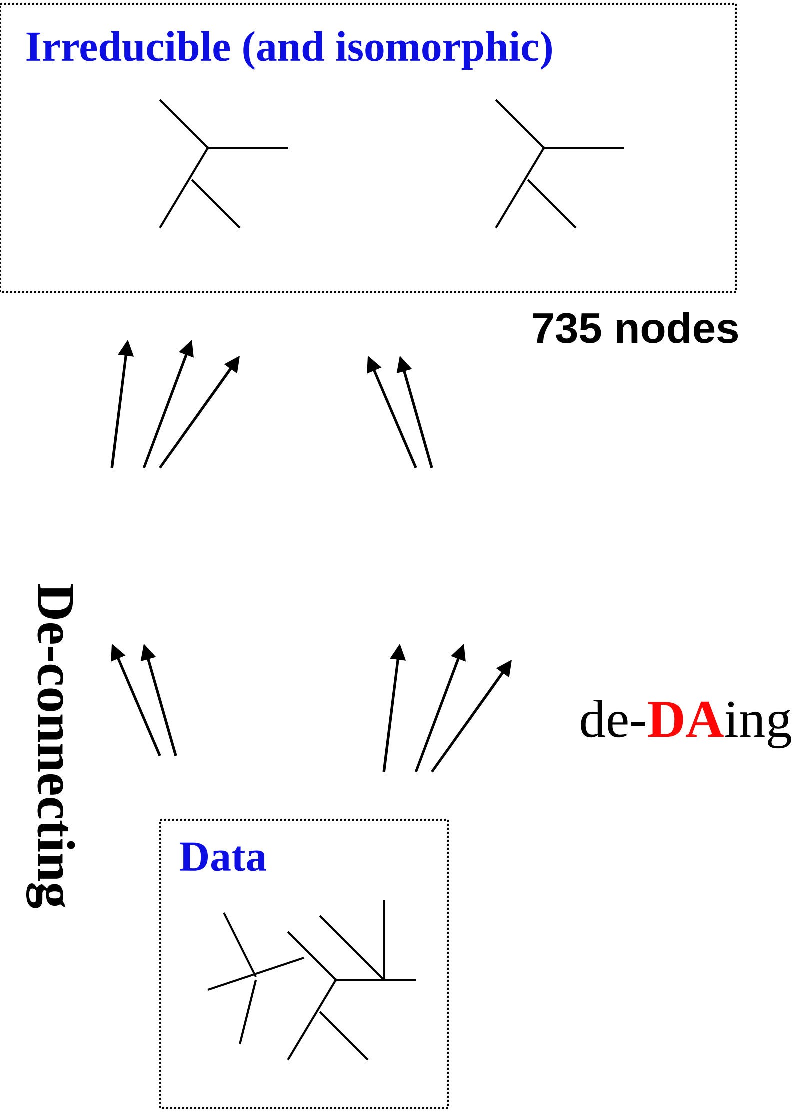


- A gene duplicates
- Inherits it connections
- The connections can change

Berg et al. ,2004:

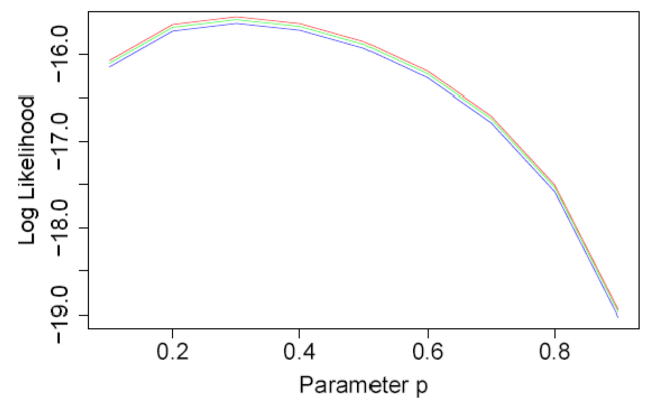
- Gene duplication slow ~10^{-9}/year
- Connection evolution fast ~10^{-6}/year
- Observed networks can be modeled as if node number was fixed.

Likelihood of PINs



2386 nodes and 7221 links

- ***Can only handle 1 graph.***
- ***Limited Evolution Model***

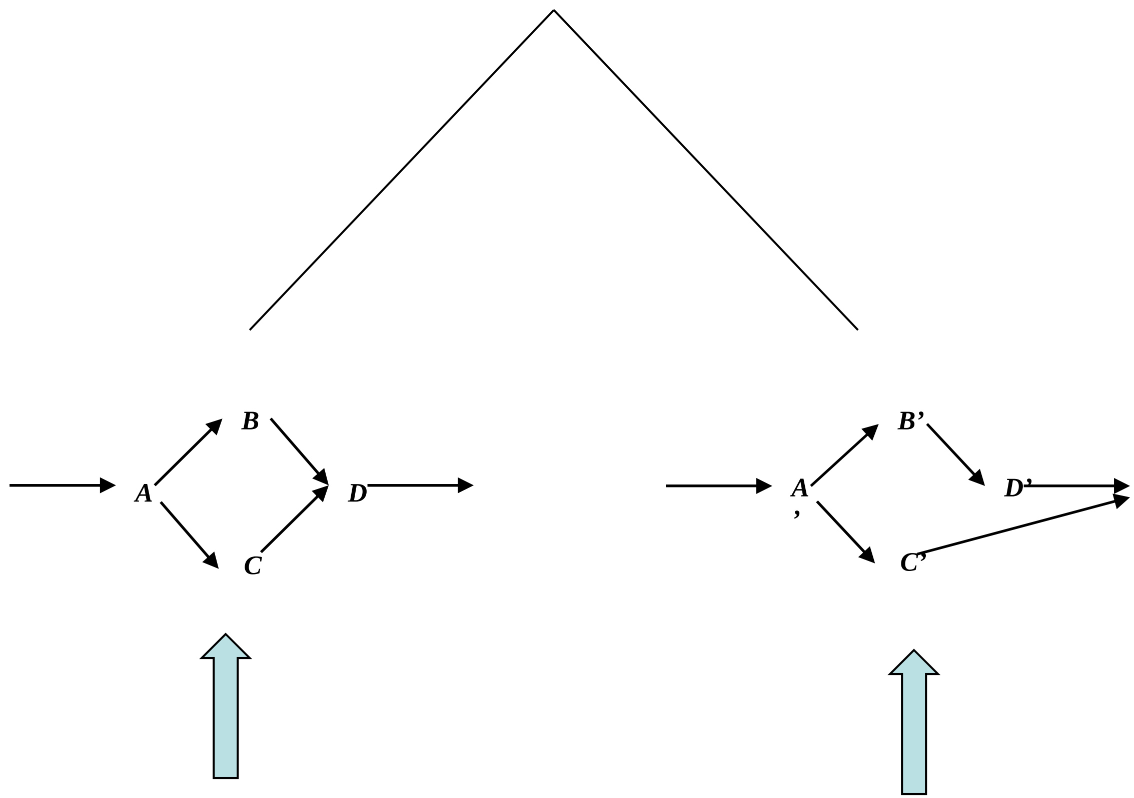


$$\theta_0 = (1,.66,.33,0)$$

Inference and Evolution

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

Evolve



Infer network

Observe (data)

Human

Mouse

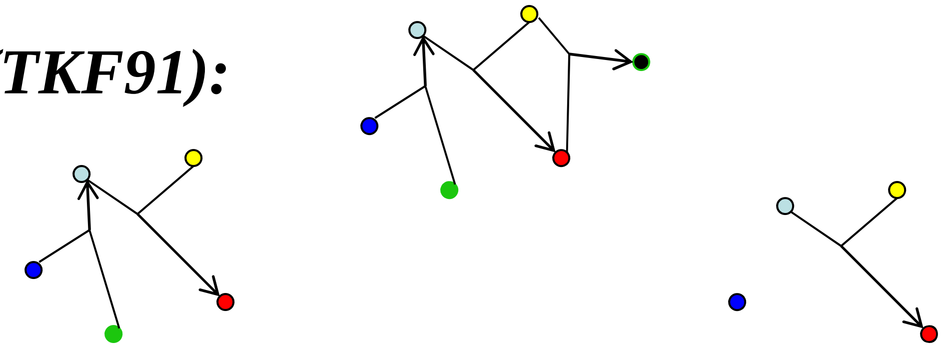
Suggestion: Evolving Dynamical Systems

- ***Goal: a time reversible model with sparse mass action system of order three!!***

Adding/Deleting components (TKF91):

Add rate: (k+1)λ

Delete rate: kμ



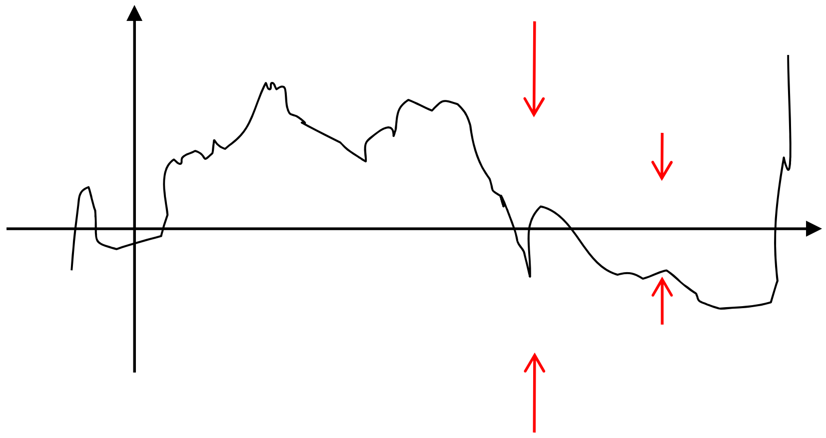
Adding reactions with birth of component:

There are 3k(k-1) possible reactions involving a new-born

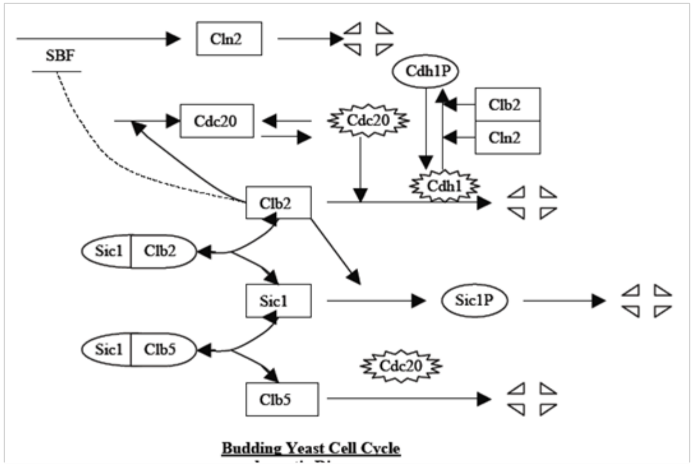
Reaction Coefficients:

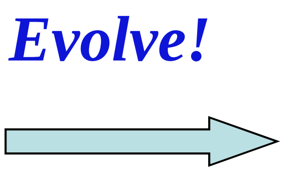
- *Continuous Time Continuous States Markov Process - specifically Diffusion.*

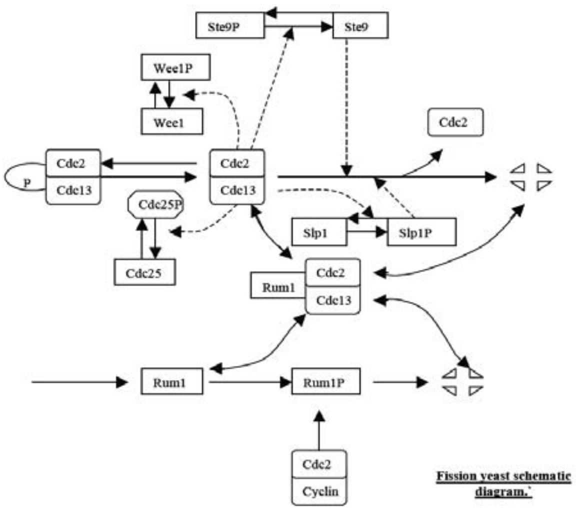
- *For instance Ornstein-Uhlenbeck, which has Gausssian equilibrium distribution*



Network Example: Cell Cycle







Budding (D_1)	Clb5	Clb2	Cdh1	Cdc20	Sic1	Cln2	SBF	(N/A)	(N/A)
Fission (D_2)	Cig2	Cdc13	Ste9	Slp1	Rum1	(N/A)	(N/A)	Cdc25	Wee1

- ***What is the edit distance?***
- ***Which properties are conserved?***
- ***If you only knew Budding Yeast, how much would you know about Fission Yeast?***
- ***As N1 starts to evolve, you can only add reactions. Isn't that strange?***
- ***On a path from N1 to N2 how close to the minimal has evolution travelled?***
- ***What is the number of equation systems possible for N1?***