

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

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

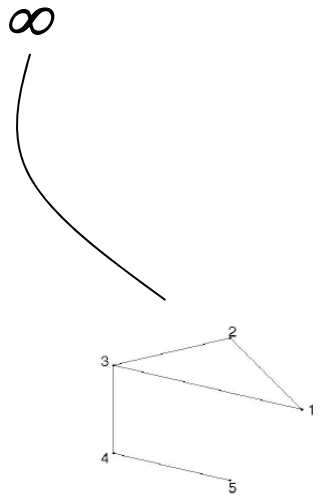


If  $d(N_1, N_2) = k$ , then there are  $2^k$  networks are visitable on shortest paths. If  $2\epsilon$  additional steps are allowed, then  $2^k (L + L(L-1)/2 + (L(L-1)..(L-\epsilon+1)/\epsilon!)$  are visitable.

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

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

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

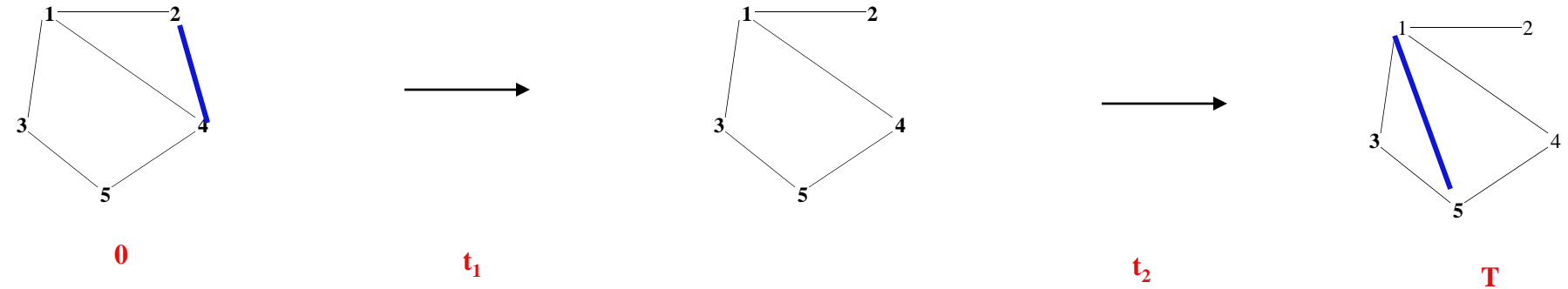


## Simulations

**Forward** with symmetries could be used in specific cases.

**Backward** (coupling from the past)

# Evolving Networks: Integration



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

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

- *The above expression can be shown to be of the form*  
*And recursions  $O(N^2)$  exists to calculate coefficients.*

$$\prod_{n=1}^N \sum_{n=0}^M e^{-q_{i0}T} \sum_{k=0}^{d_n} c_n^k T^k$$

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

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

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

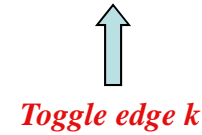
$$P(N \rightarrow N') = \sum_i P(N \rightarrow N'; i \text{ steps})$$

# Evolving Networks: MCMC

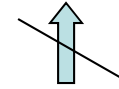
*Present pathway:*



• *Insertion of an edge pair*



• *Deletion of an edge pair*

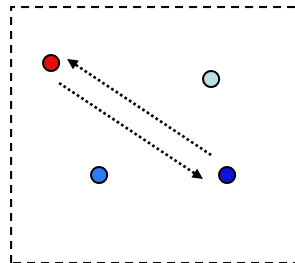


• *Moving of a pair or singles*



• *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*

*Acceptance ratio*

$$\frac{L(\bullet)q(\bullet, \bullet)}{L(\bullet)q(\bullet, \bullet)} J$$

# A Toy Example

(by Aziz Mithani)

## Equilibrium Probability

### Metabolic Universe

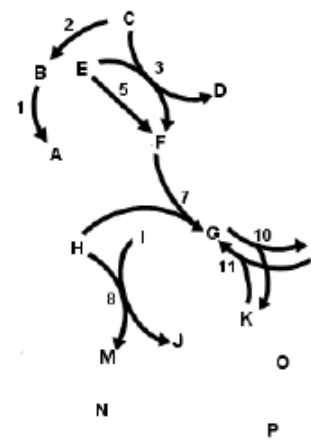
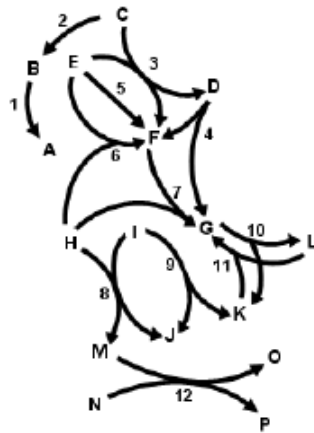
### 12 possible edges

1i 1u 3

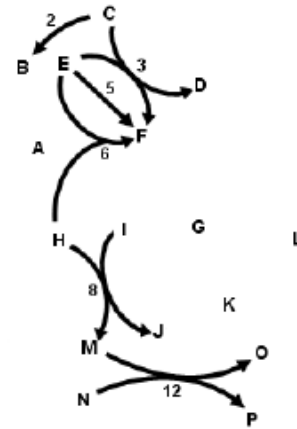
1i 2u 3

2u 1i 3

2i 2u 3



### Transition Probability



dist=6

## Transition Probability:

Full Exponentiation ( $2^{12}$  states 4096)

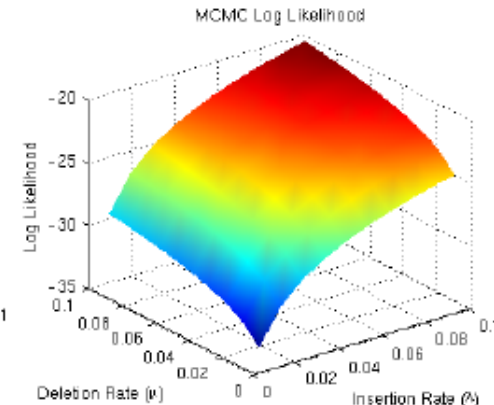
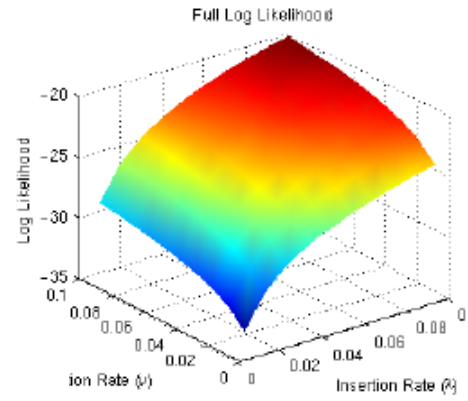
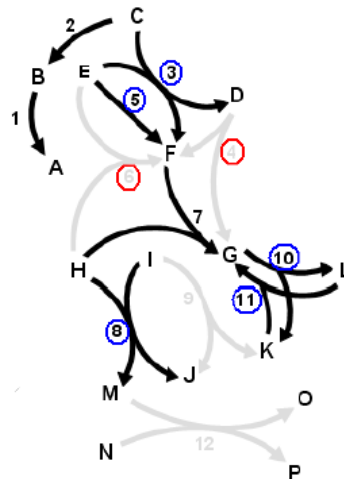
Exponentiation with corner cutting

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

MCMC Integration

## Adding Connectedness

Favouring insertions connecting

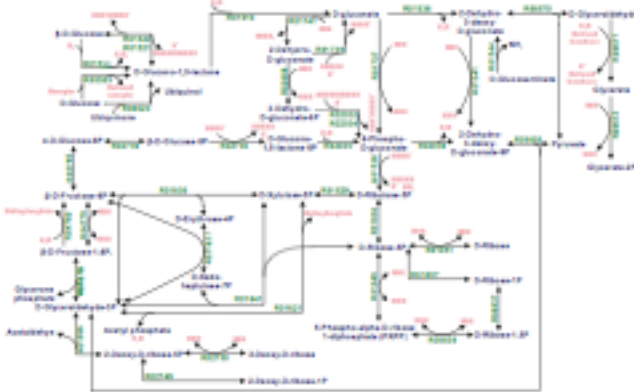


The proportion present:  $\frac{5}{7} = 0.714$

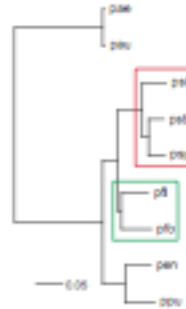
# Application and extension to Phylogeny

## One metabolism

### Pentose Phosphate Pathway



### *Pseudomonas*



## Pairwise analysis

Pathway Map	Start Organism	End Organism	Differences*	CnI <sup>***</sup>	$\lambda_t$	STD	$\mu_t$	STD	$\lambda/\mu$
Pentose phosphate pathway (MAP00030)	<i>P. aeruginosa</i> PAO1	<i>P. syringae</i> DC3000	9 (I:8, D:1)	-	0.31694	0.00853	0.06168	0.00656	5.13849
	<i>P. fluorescens</i> PF-5	<i>P. fluorescens</i> Pfl0-1	2 (I:2, D:0)	+	1.91912	0.26728	0.32402	0.12122	5.92285
				+	0.09724	0.00176	0.01852	0.00186	5.25147
Lysine degradation (MAP00310)	<i>P. aeruginosa</i> PAO1	<i>P. syringae</i> DC3000	1 (I:1, D:0)	-	0.02283	0.00058	0.06976	0.00576	0.32732
	<i>P. fluorescens</i> PF-5	<i>P. fluorescens</i> Pfl0-1	2 (I:0, D:2)	+	0.38110	0.22654	3.25666	3.28194	0.11702
				+	0.00787	0.00074	0.29881	0.00534	0.02635
Phenylalanine metabolism (MAP00360)	<i>P. aeruginosa</i> PAO1	<i>P. syringae</i> DC3000	6 (I:4, D:2)	-	0.07867	0.00759	0.48296	0.08009	0.16289
	<i>P. fluorescens</i> PF-5	<i>P. fluorescens</i> Pfl0-1	7 (I:2, D:5)	+	0.52749	0.13457	1.95439	0.71231	0.26990
				+	0.04486	0.00528	0.78573	0.06140	0.05709
				+	0.34472	0.10428	2.05643	0.46145	0.16763

## Multiple analysis with parameter for neighbor dependence

Pathway Map	Phylogeny	$E(\delta)$	$\text{var}(\delta)$	$E(\lambda)$	$\text{var}(\lambda)$	$E(\mu)$	$\text{var}(\mu)$	$\lambda/\mu$
Pentose phosphate pathway (MAP00030)	( <i>pae</i> ,( <i>pfl</i> , <i>pfo</i> ))	0.3345	0.0072	2.2492	1.8556	1.0969	0.1865	2.0505
	( <i>pae</i> ,(( <i>psb</i> , <i>psp</i> ), <i>pst</i> ))	0.3368	0.0073	1.7887	1.3182	0.8588	0.0998	2.0829
	<i>Pseudomonas</i>	0.2508	0.0044	1.2345	0.0580	0.9431	0.0178	1.3089
Lysine degradation (MAP00310)	( <i>pae</i> ,( <i>pfl</i> , <i>pfo</i> ))	0.0861	0.0037	0.6244	0.2013	2.6961	4.7734	0.2316
	( <i>pae</i> ,(( <i>psb</i> , <i>psp</i> ), <i>pst</i> ))	0.0707	0.0024	0.4732	0.0433	4.0998	6.1328	0.1154
	<i>Pseudomonas</i>	0.0660	0.0022	0.4955	0.0235	2.2045	2.3058	0.2248
Phenylalanine metabolism (MAP00360)	( <i>pae</i> ,( <i>pfl</i> , <i>pfo</i> ))	0.0634	0.0020	0.9337	0.1257	3.2754	2.9057	0.2851
	( <i>pae</i> ,(( <i>psb</i> , <i>psp</i> ), <i>pst</i> ))	0.0552	0.0019	0.9180	0.0825	3.0896	2.6660	0.2971
	<i>Pseudomonas</i>	0.0479	0.0014	0.8943	0.0417	1.9824	0.5205	0.4511

## Main results

### Evolution rates

- Insertion rate higher for pathway maps involved in central metabolism and amino acid biosynthesis than those involved in secondary metabolism and amino acid degradation
- Rates higher in *Pseudomonas syringae* compared to *Pseudomonas fluorescens*
  - Supports experimental findings – High number of deletions in *P. syringae* lineage
- Low insertion to deletion ratio ( $\lambda/\mu$ ) for pathway maps related to amino acids which are poor nutrient sources

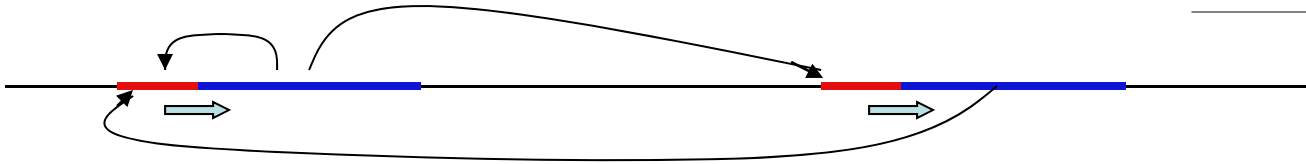
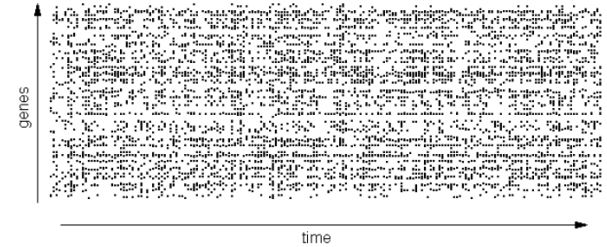
### Neighbourhood structure

- Pathway maps involved in central metabolism and metabolism of essential amino acid have strong neighbourhood structure.

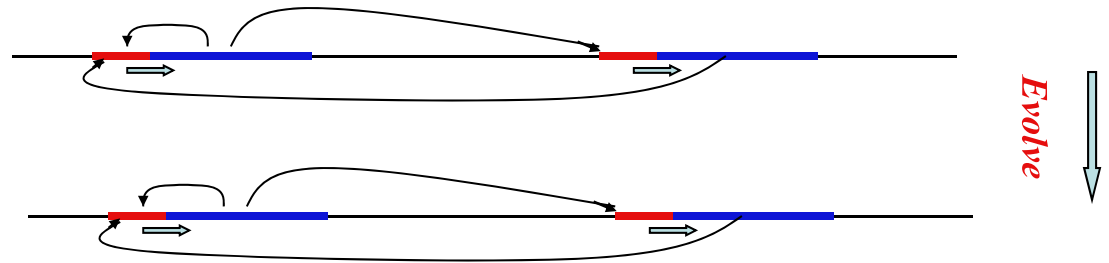
# Regulatory Network Evolution

*Artificial Genome*  
*Riel, 1999:*

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



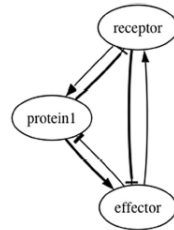
*Evolving Artificial Genome*  
*Quant & Bullocks, 2007:*



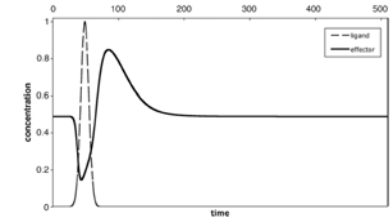
- *Selection will influence final dynamics*

# 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.*

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

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

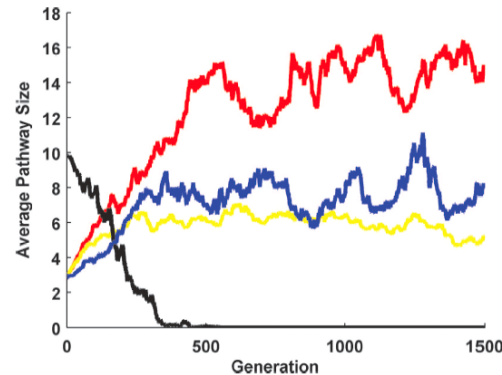
- Fitness**

$$F = 1 - nc \quad \text{if } \alpha = 1$$

$$F = 0 \quad \text{if } \alpha = 0$$

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

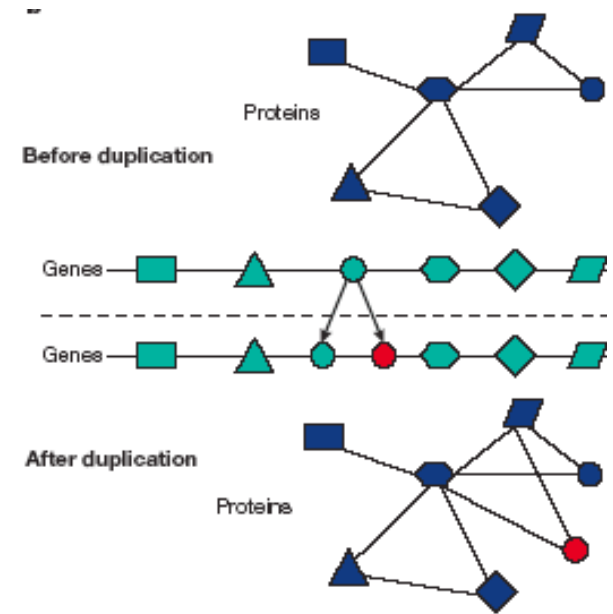
- Evolution**



# Models of Protein Interaction Networks Evolution

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

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



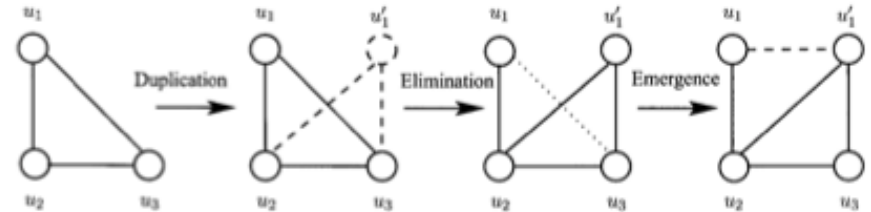
**Berg et al. ,2004:**

- 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.

# Network Alignment

## Basic Operations:

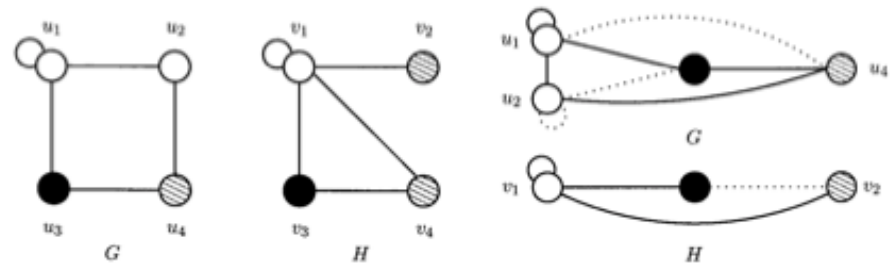
- **Duplication**
- **Elimination** (loss of interaction – edge)
- **Emergence** (gain of interaction – edge)
- **Deletion** (loss of node with edges)



Associated cost function for each operation and cut-off  $\Delta$  (matching disallowed)

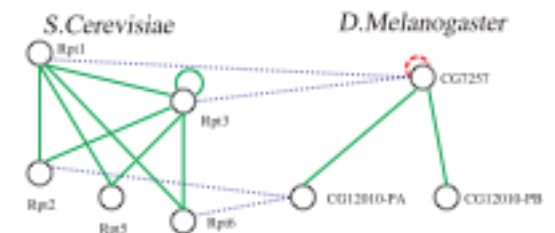
Heuristic Search of optimal matching subnetworks

## Example of small matched networks



## Application to Yeast, C.elegans, D.melanogastor,

Organism pair	# Nodes	# Matched nodes		# Matches		# Mismatches	# Duplications	
		$\bar{\Delta} = 1$	$\bar{\Delta} = 2$	$\bar{\Delta} = 1$	$\bar{\Delta} = 2$	$\bar{\Delta} = 1$	Org. 1	Org. 2
SC vs CE	2746	312	1230	412	3007	40262	6107	6886
SC vs DM	15884	1730	8622	2061	42781	1054241	6107	32670
CE vs DM	11805	491	3391	455	6626	205593	6886	32670

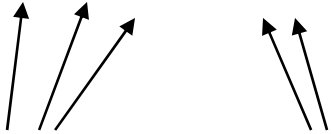


# Likelihood of PINs

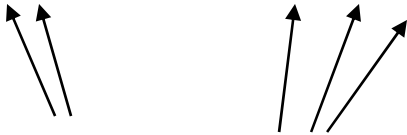
Irreducible (and isomorphic)



735 nodes

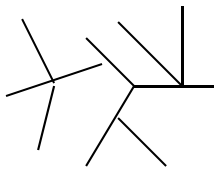


De-connecting



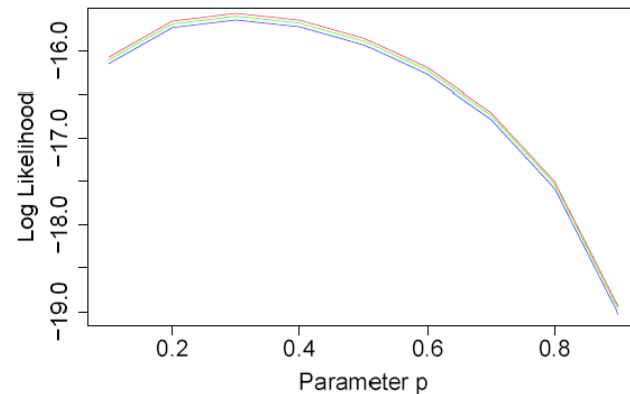
de-DAing

Data



2386 nodes and 7221 links

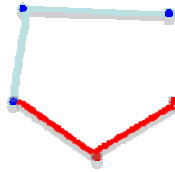
- *Can only handle 1 graph.*
- *Limited Evolution Model*



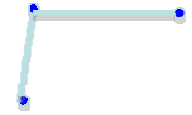
$$\theta_0 = (1, .66, .33, 0)$$

# The size of the human interactome

True network –  $G_N=(V_N, E_N)$



Sampled network  $G_S=(V_S, E_S)$



$$P_{p,\theta}(G_S) = \sum_{G_N \supseteq G_S} P_p(G_S|G_N)P_\theta(G_N)$$

$$P_{p,\theta}(G_S) = Q_p(N_S) \sum_{G_N \supseteq G_S} q(G_S, G_N)P_\theta(G_N)$$

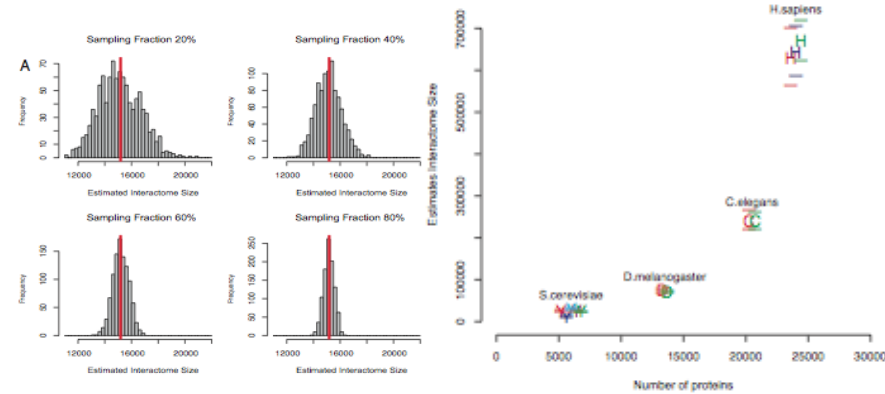
$$\hat{p} = N_S / N_N$$

$$P_\theta(G_N^* | G_S) = q(G_S, G_N^*)P_\theta(G_N^*) / \sum_{G_N \supseteq G_S} q(G_S, G_N)P_\theta(G_N)$$

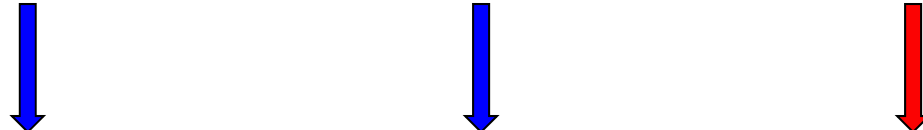
$$\hat{\pi} = \frac{2M_S}{N_S(N_S - 1)}$$

$$\hat{M}_N = M_S \frac{N_N(M_N - 1)}{N_S(N_S - 1)}$$

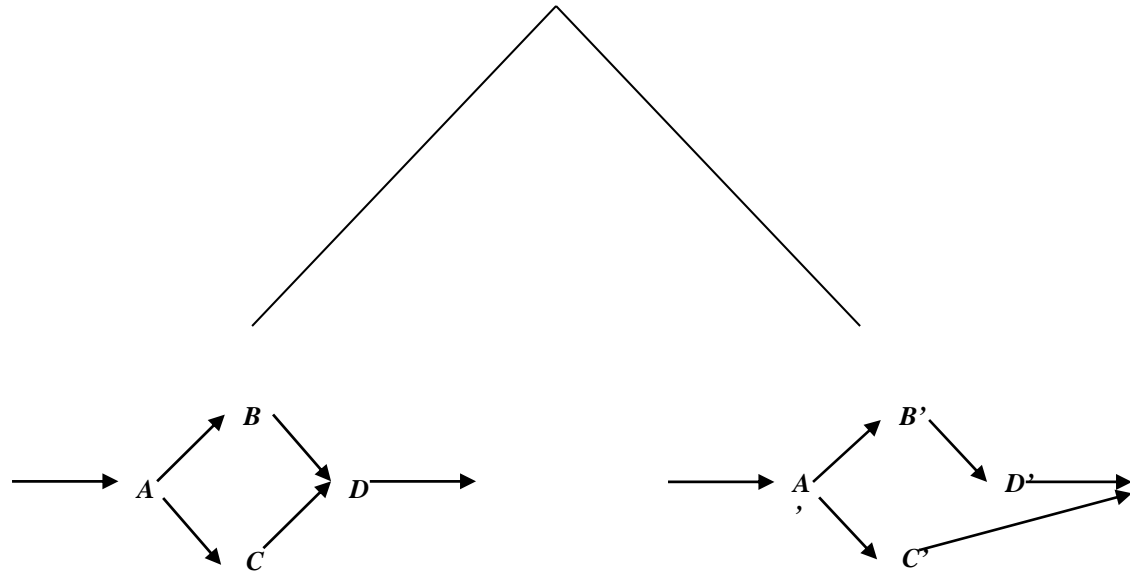
Species	Nodes	Edges	MN	95% CI
Yeast	4,959	17,226	25,229	24,100- 26,440
Drosophila	7,451	22,636	74,336	71,700- 77,100
C.Elegans	2,638	3,970	240,544	220,030-263,270
H.Sapiens	1,085	1,346	672,918	625,170-722,670



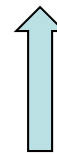
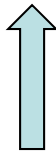
# Inference and Evolution

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


Evolve



Infer network



Observe (data)

Human

Mouse

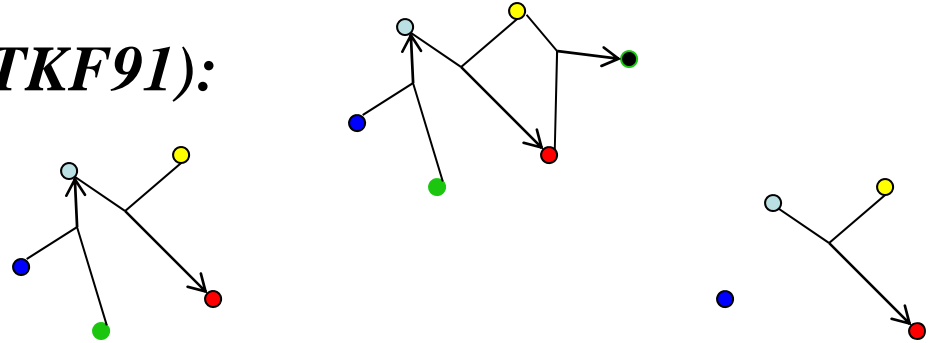
# Suggestion: Evolving Dynamical Systems

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

*Adding/Deleting components (TKF91):*

*Add rate:  $(k+1)\lambda$*

*Delete rate:  $k\mu$*

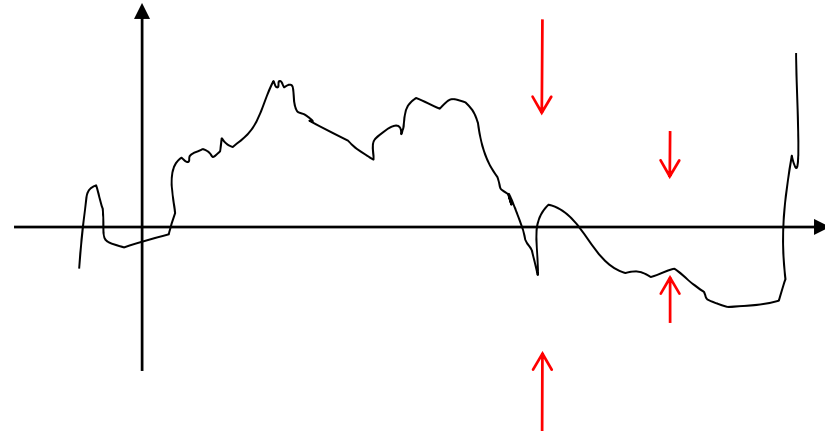


*Adding reactions with birth of component:*

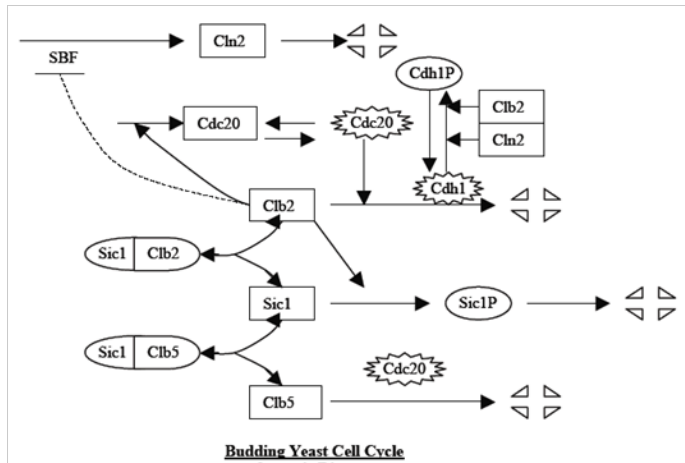
*There are  $3k(k-1)$  possible reactions involving a new-born*

*Reaction Coefficients:*

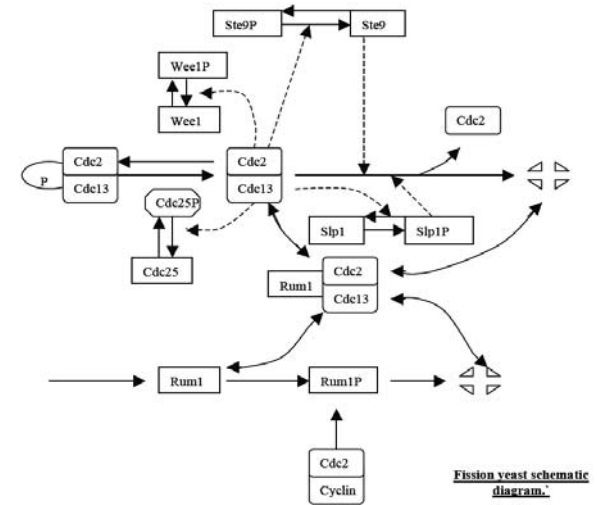
- *Continuous Time Continuous States Markov Process - specifically Diffusion.*
- *For instance Ornstein-Uhlenbeck, which has Gaussian equilibrium distribution*



# Network Example: Cell Cycle



*Evolve!*



Budding (D <sub>1</sub> )	Clb5	Clb2	Cdh1	Cdc20	Sic1	Cln2	SBF	(N/A)	(N/A)
Fission (D <sub>2</sub> )	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?*

# Summary

*The importance of modelling*

*The main classes of networks*

*The development of sequence and networks models*

*Integration of paths*

*Basic Models*

*Modeling Dynamic Systems Evolution*