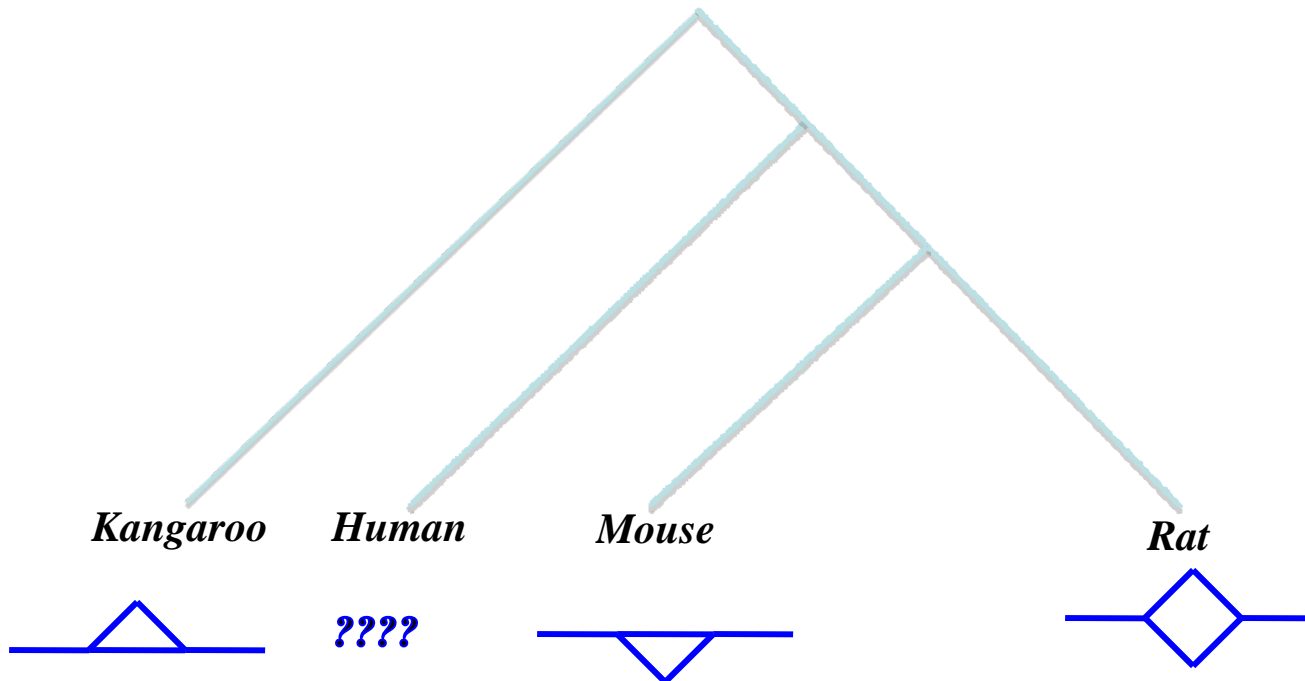
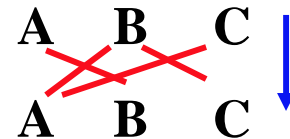
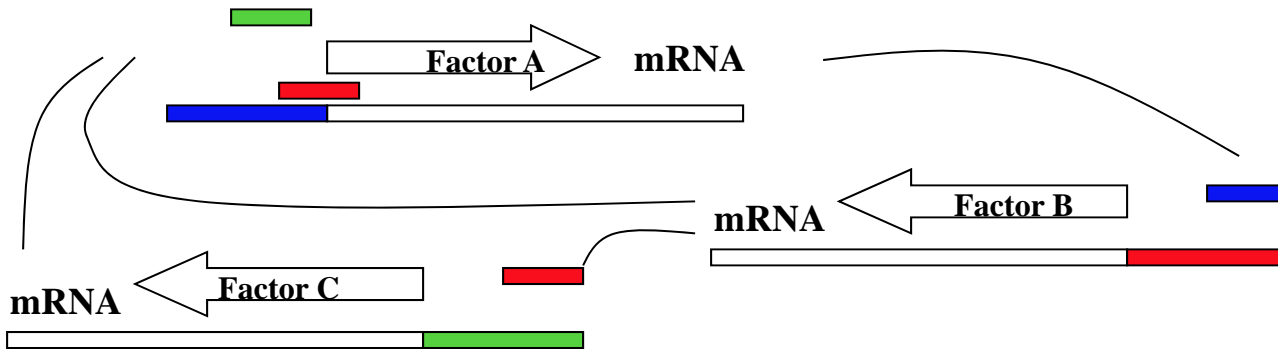
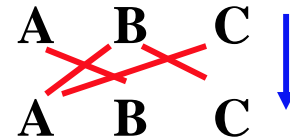
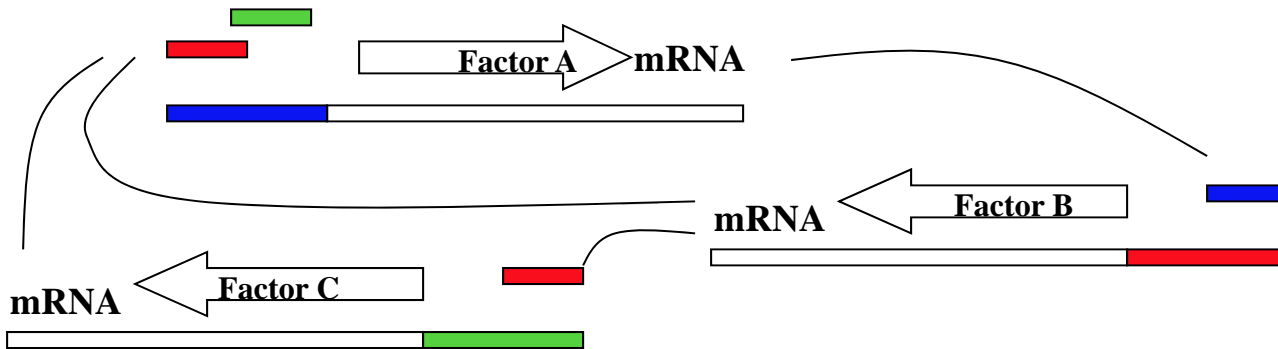
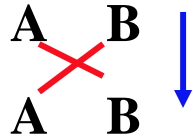
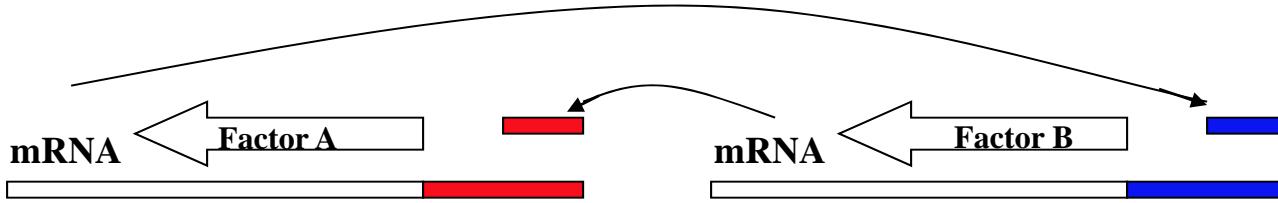


Network Evolution

- *Understanding Evolution*
- *Comparative Annotation*
- *Knowledge and Model Organisms*

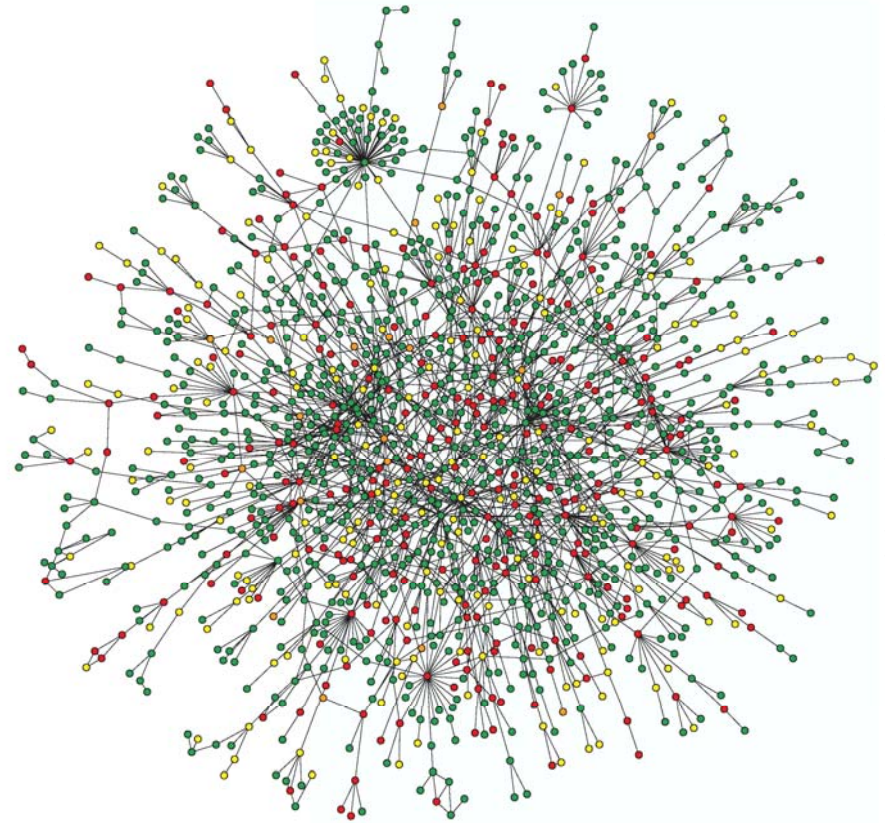


Gene Regulatory Network



Protein Interaction Network

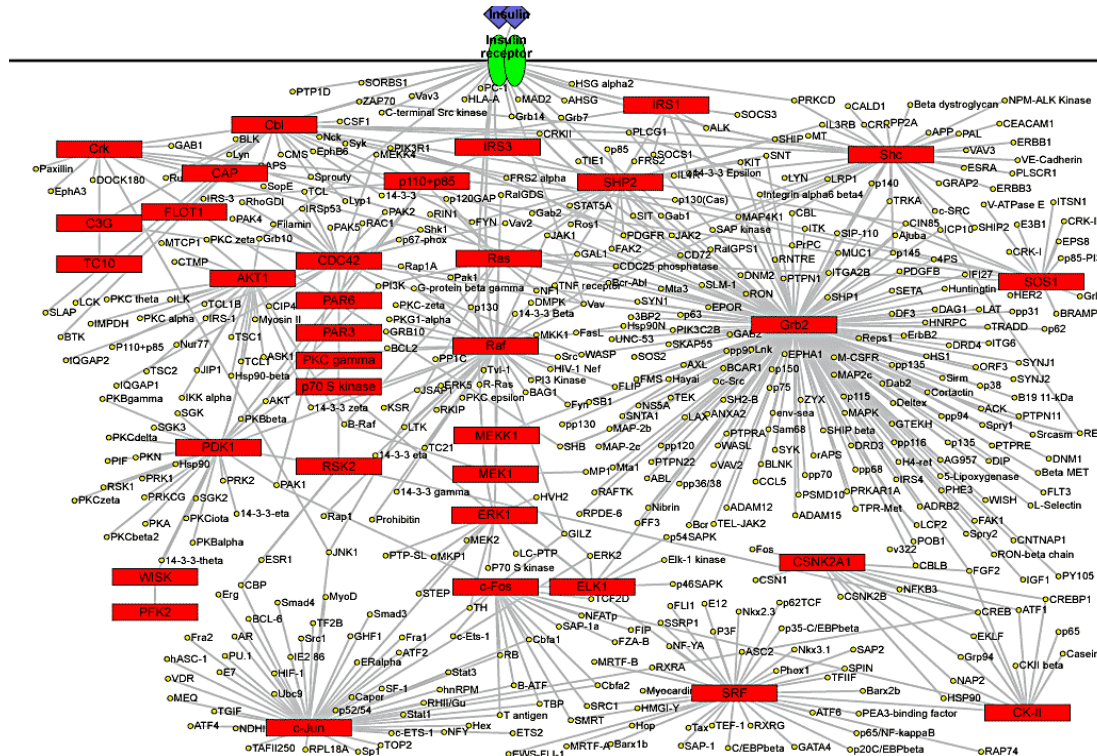
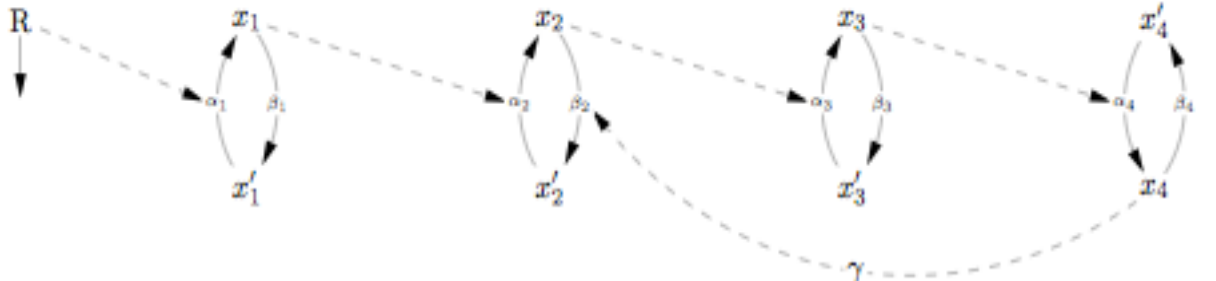
- *The sticking together of different protein is measured by mass spectroscopy.*
- *The nodes will be all known proteins.*
- *Two nodes are connected if they stick together. This can be indicator of being part of a functional protein complex, but can also occur for other reasons.*



Signaling Pathways

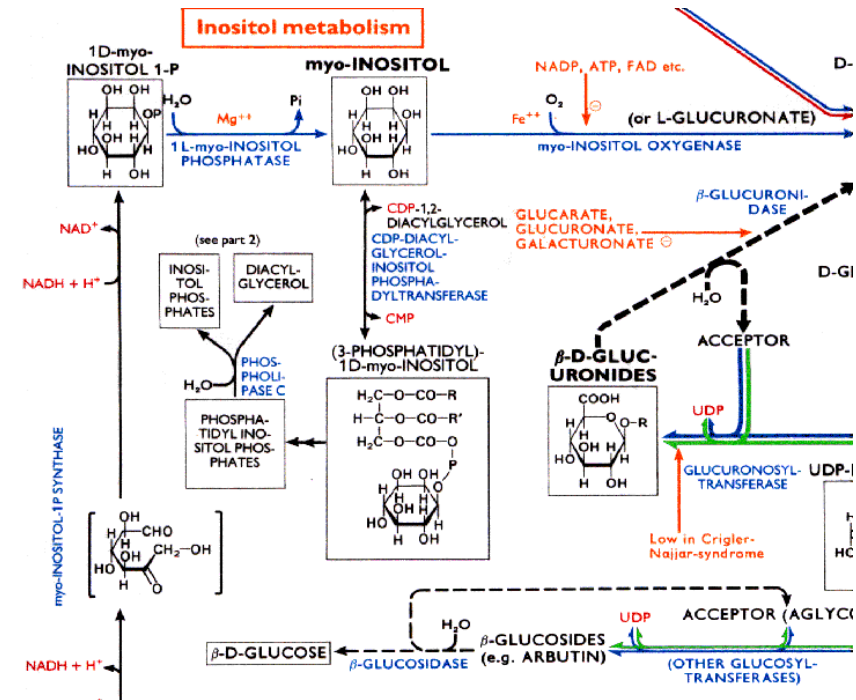
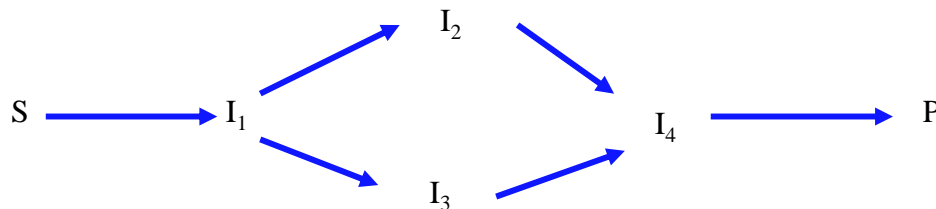
- Transmits signals from membrane to gene regulation.

- Its function is enigmatic as some of the molecules involved are common to different functions and how cross-interaction is avoided is unknown.



Metabolic Pathways

- Flux Analysis
- Metabolic Control Theory
- Biochemical Systems Theory
- Kinetic Modeling



RAFs – Reflexive Autocatalytic Foodsets

Kauffman, 1986; Steel, 2001, Hordijk and Steel, 2004; Mossel and Steel, 2005

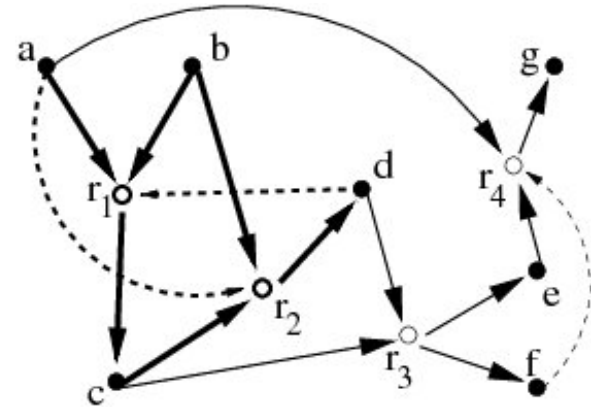
- *a set of molecule types, X ;*
- *a set of reactions where each reaction converts one set of molecules (reactants) into another set (products), R ;*
- *a set of catalysations: molecules that accelerate a reaction (or set of reactions), C ;*
 - *a food set: a small set of molecules assumed to be freely available and constantly replenished, F .*

$X = \{a, b, c, d, e, f, g\}$

$R = \{r_1, r_2, r_3, r_4\}$

$C = \{(d, r_1), (a, r_2), (f, r_4)\}$

$F = \{a, b\}$



RAFs – Reflexive Autocatalytic Foodsets

Kauffman, 1986; Steel, 2001, Hordijk and Steel, 2004; Mossel and Steel, 2005

Key achievements:

- *The probability of existence;*
- *Algorithms to find them*

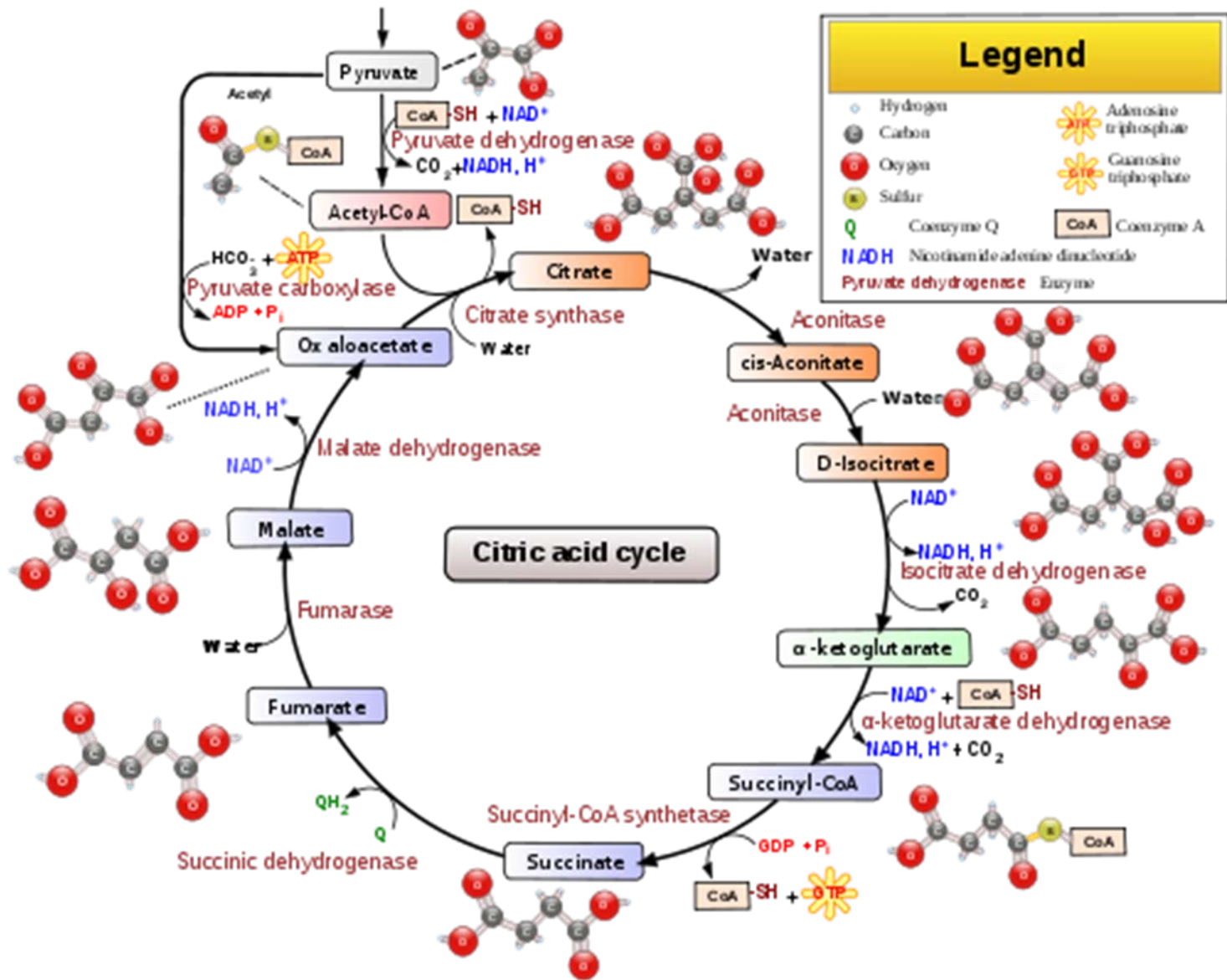
Key problems:

- *Realism;*
- *Predicting catalysis;*

Natural Extensions:

- *Let RNA be the molecules, concatenation by base-pairing*
- *Kinetic version: concentrations and rates*
- *Evolving version*
- *RAFs based on real molecules*
 - *combinatorially defined*
 - *empirically defined – observed molecules (Beilstein)*
 - *empirically defined – observed **life** molecules (Metabolism)*

Real Molecule Example of RAF: Citric Acid Cycles

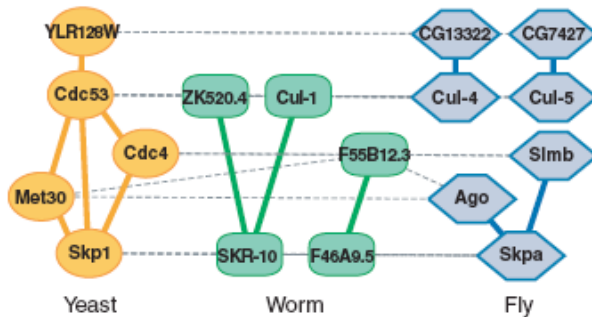


Network Alignment & Motifs

Barabasi & Oltvai, 2004, Sharan & Ideker, 2006

1. Are nodes/edges labelled?
2. Which operations are allowed?
3. Pair/Multiple?

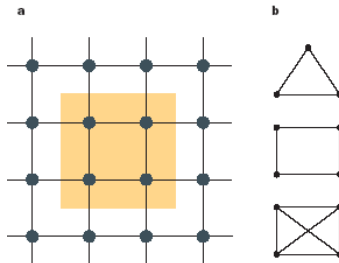
•Network Alignment



•Network Search

Find (approximately) a network within a set of others.

•Motifs



•Network integration

Combine a set of networks to one large network.

Metabolic Random Fields

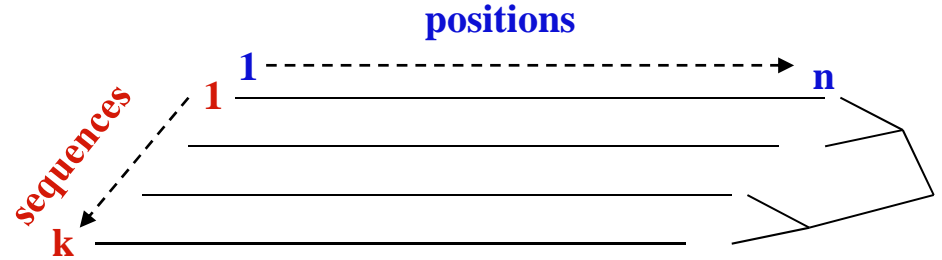
Artemisa Labi, Chris Campbell, Istvan Miklos,

Sequence Analysis:

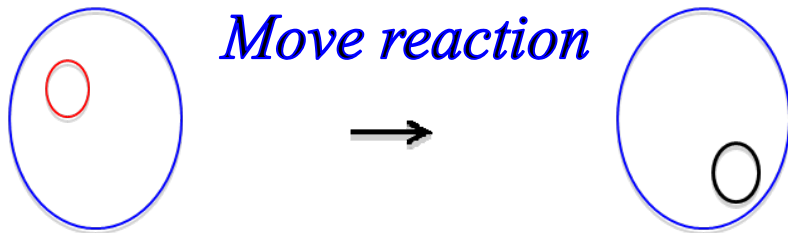
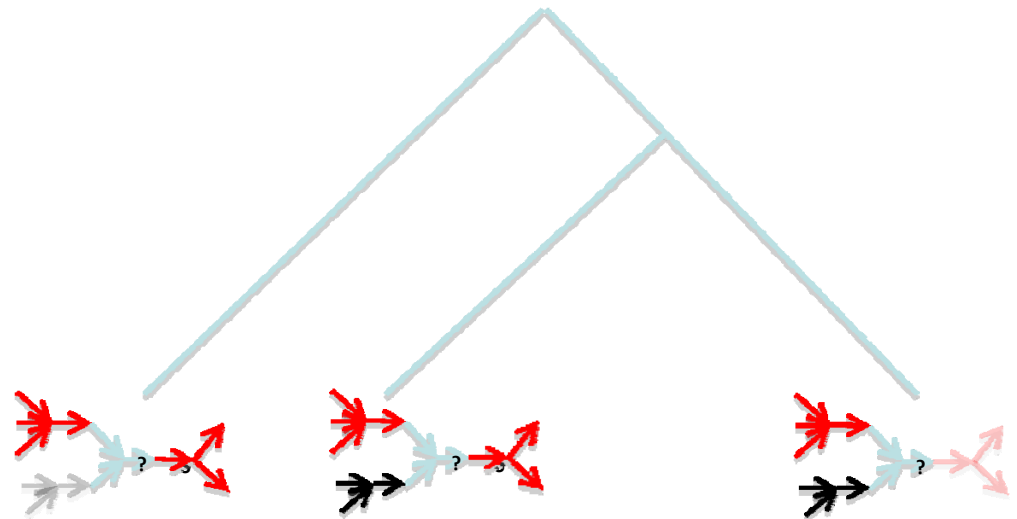
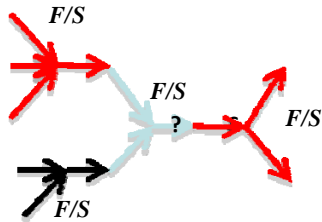


slow - r_s
fast - r_f

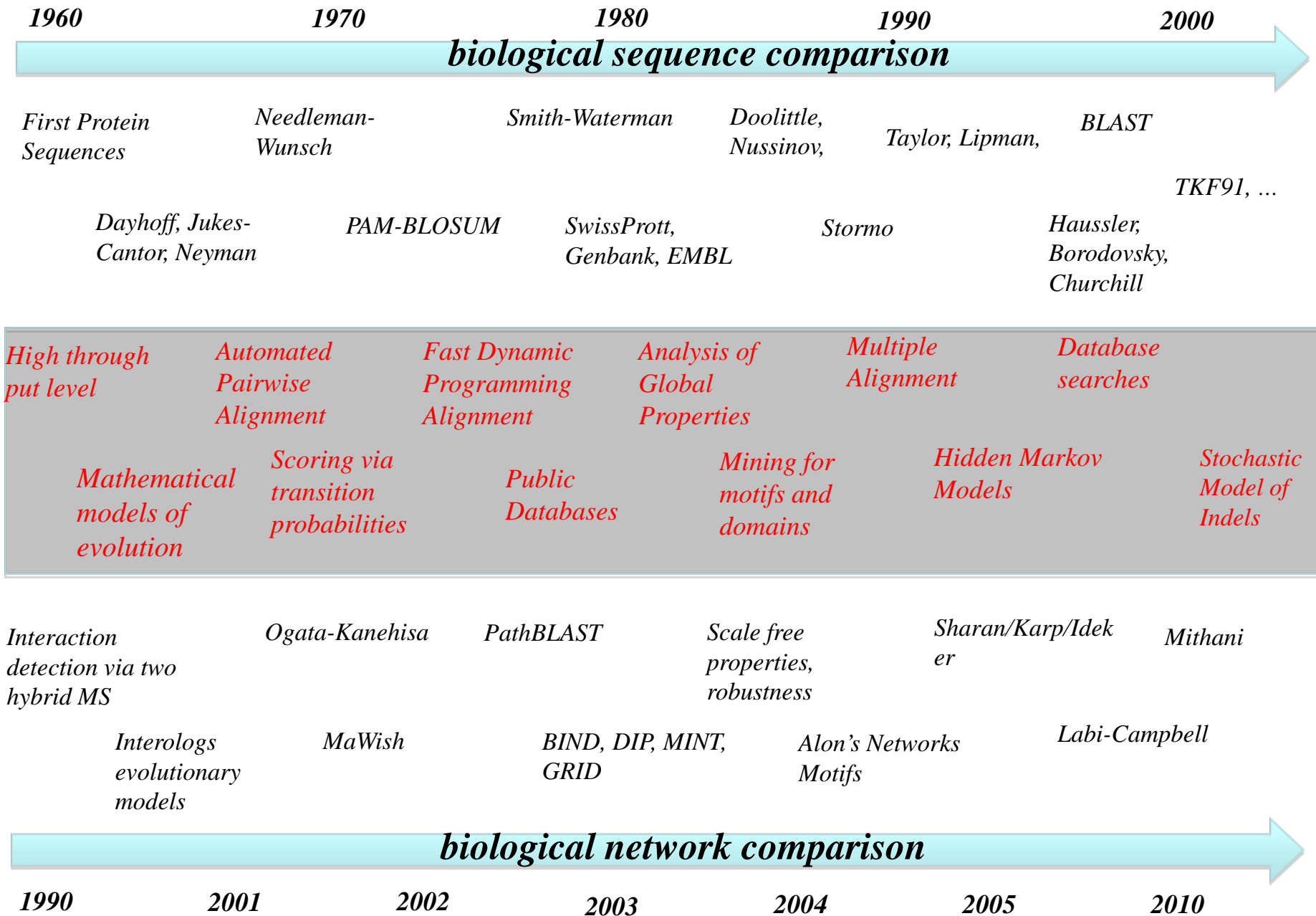
HMM



Network Analysis:



Development of Network/Sequence Analysis



Network Description and Statistics I

Barabasi & Oltvai, 2004

- Degree/Indegree/Outdegree
- Shortest Path $Dist(i, j)$
- Mean Path Length
- Diameter: $Max_{i, j} \{Dist(i, j)\}$
- Clustering Coefficient - $C_I = 2T_I / n_I(n_I - 1)$

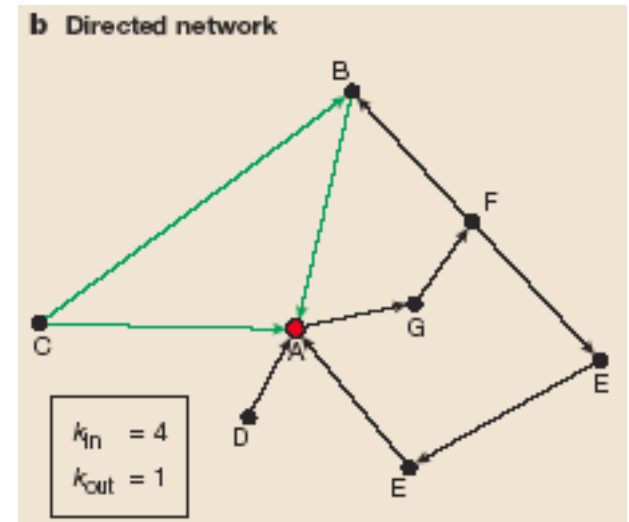
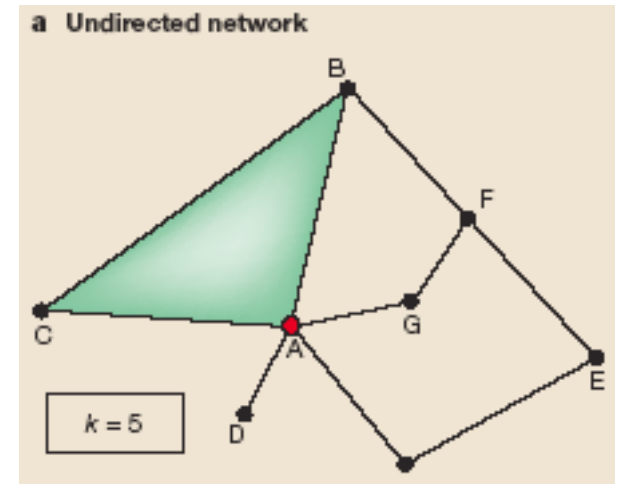
$$C_A = 1/10$$

- Degree Distribution - $P(k)$
- Scale Free Networks $P(k) \sim k^{-\gamma}$ $\gamma > 2$
- Hubs: multiply connected nodes

The lower γ , the more hubs.

Small World Property:

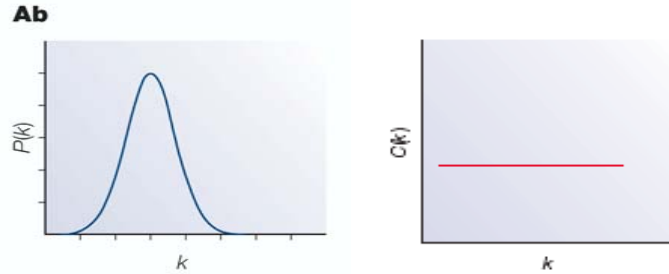
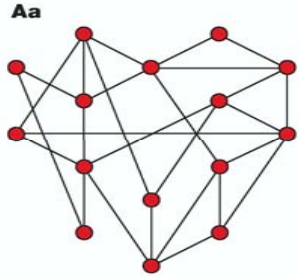
Graph connected and path lengths small



Network Description and Statistics II

Barabasi & Oltvai, 2004

A. Random Networks [Erdos and Rényi (1959, 1960)]



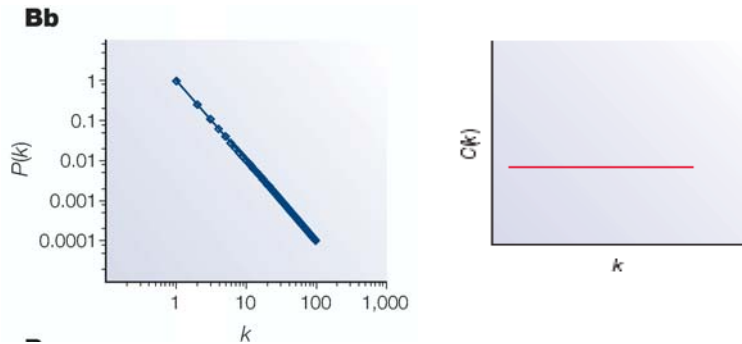
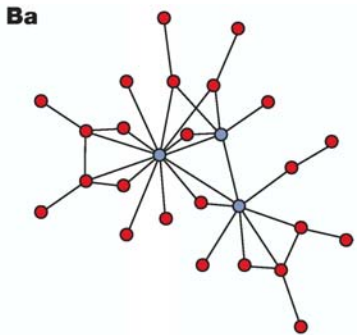
$$P(k) = \frac{e^{-\bar{k}} \bar{k}^k}{k!}$$

Mean path length $\sim \ln(k)$

Phase transition:

Connected if: $p \geq \ln(k)/k$

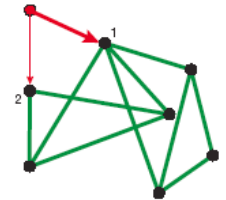
B. Scale Free [Price,1965 & Barabasi,1999]



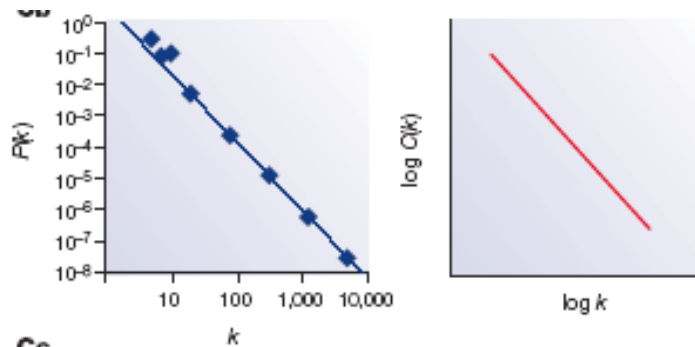
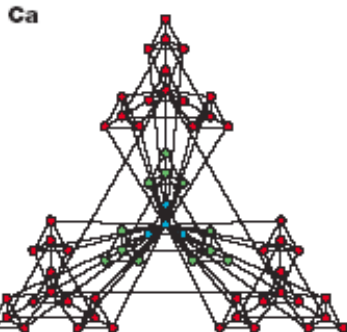
$$P(k) \sim k^{-\gamma}, k \gg 1, 2 < \gamma$$

Mean path length $\sim \ln \ln(k)$

Preferential attachment. Add proportionally to connectedness



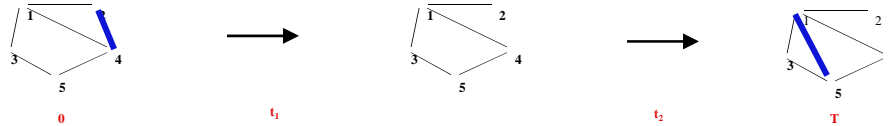
C. Hierarchical



Copy smaller graphs and let them keep their connections.

Stochastic Modeling of Network Evolution

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



What do models of network evolution do?:

Test models

Estimate Parameters in the Evolutionary Process

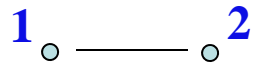
Ancestral Analysis

Framework for Knowledge Transfer

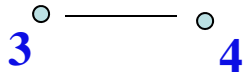
Likelihood of Homologous Pathways

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

Number of Metabolisms:



+ 2 symmetrical versions



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



Approaches:

Continuous Time Markov Chains with computational tricks.

MCMC

Importance Sampling



A Model for the Evolution of Metabolisms

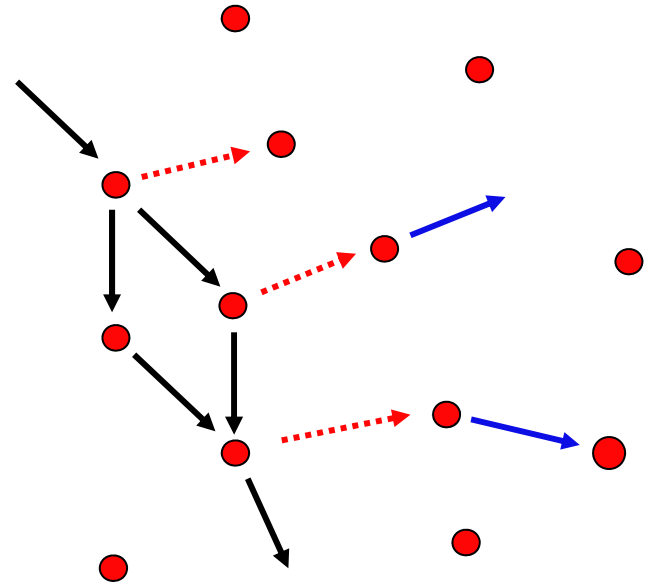
- A given set of metabolites: 
- A given set of possible reactions -
arrows not shown.
- A core metabolism: 
- A set of present reactions - **M**
black and **red** arrows

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

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

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

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

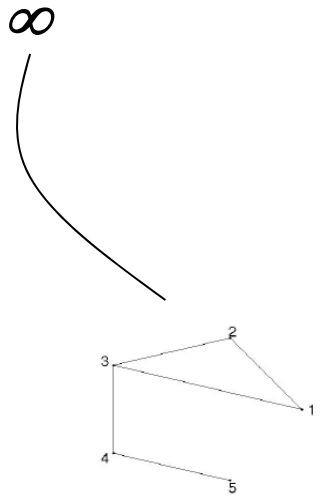


If $d(N_1, N_2) = k$, then there are 2^k networks are visitable on shortest paths. If 2ϵ additional steps are allowed, then $2^k (L + L(L-1)/2 + (L(L-1)..(L-\epsilon+1)/\epsilon!)$ are visitable.

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)

A Toy Example

(by Aziz Mithani)

Equilibrium Probability

Metabolic Universe

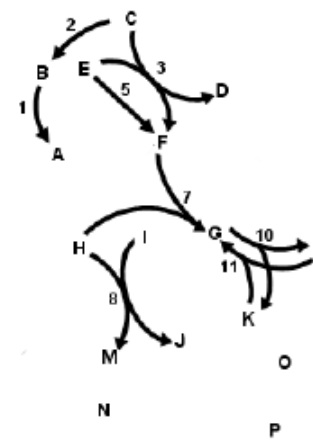
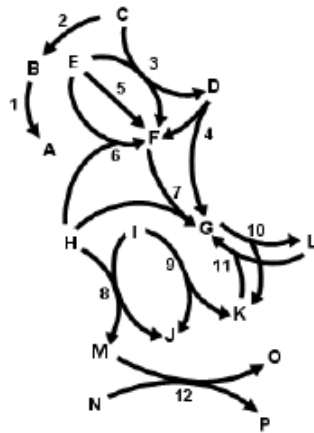
12 possible edges

1i 1u 3

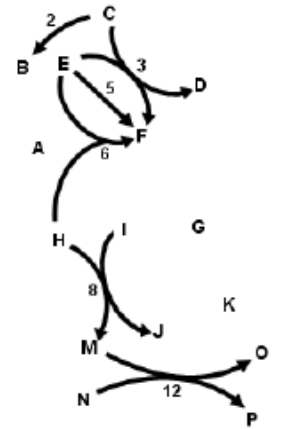
1i 2u 3

2u 1i 3

2i 2u 3



Transition Probability



dist=6

Transition Probability:

Full Exponentiation (2^{12} states 4096)

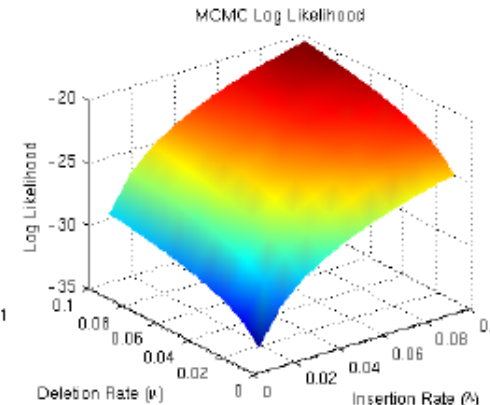
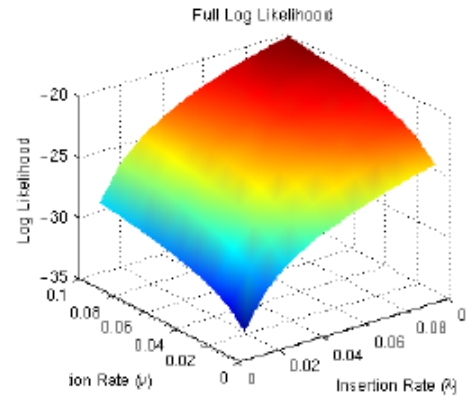
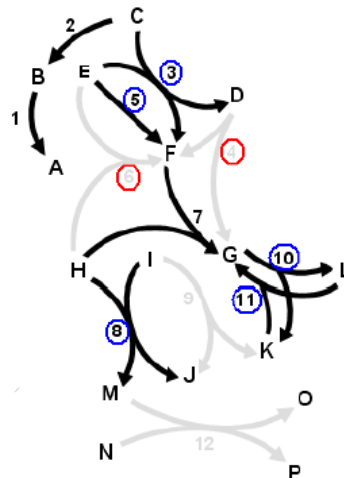
Exponentiation with corner cutting

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

MCMC Integration

Adding Connectedness

Favouring insertions connecting



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