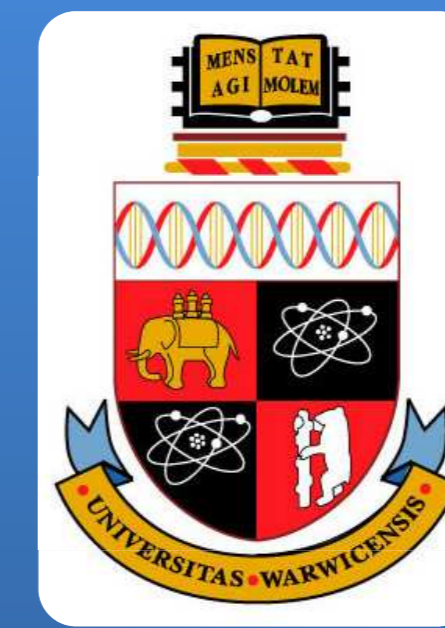


Fast Mixing Monte Carlo For Biological Networks

Dan Barker Supervisors: Sach Mukherjee & Mario Nicodemi

(d.j.barker@warwick.ac.uk, s.n.mukherjee@warwick.ac.uk, m.nicodemi@warwick.ac.uk)



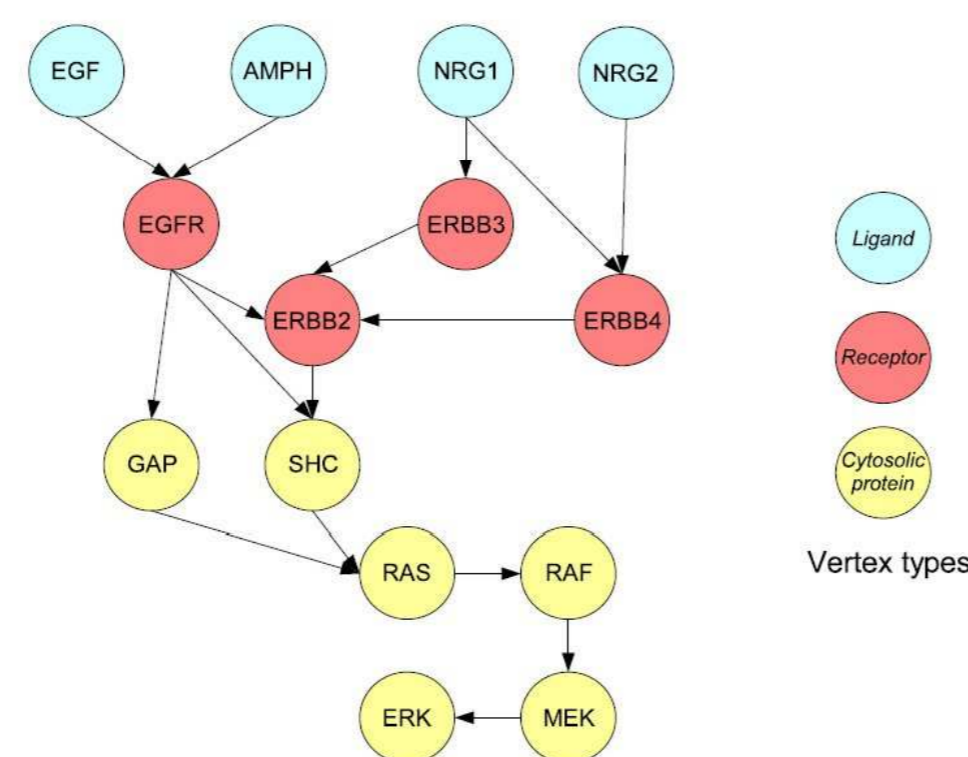
EPSRC

Warwick Complexity Science DTC

Biological Networks

- Biological Networks describe the interactions of a set of molecular constituents, e.g. Proteins or genes.
- They can be represented mathematically using a **graph**.

Figure 1: An example of a biological system and their interactions. Some constituent biological components of the Epidermal Growth Factor Receptor system.



Inference

- We use a stochastic **model** for the biological network called a “graphical model”.
- The graph must be **acyclic** (DAG).
- The **Structure** of the network can be **inferred** from data.

- Use **Bayes’** theorem to relate the “marginal likelihood” to the “posterior distribution”.

$$P(G|X) \propto P(X|G)P(G)$$

posterior
likelihood
prior

$$P(X|G) = \prod_i^N \prod_j^{q_i} \frac{\Gamma(N'_{ij})}{\Gamma(N_{ij} + N'_{ij})} \prod_k^{r_i} \frac{\Gamma(N_{ijk} + N'_{ijk})}{\Gamma(N'_{ijk})}$$

nodes
data
hyper-parameters

parent states
possible values

Monte Carlo

- We want $P(G|X)$ but this requires **enumerating** all DAGs.
- For 11 nodes the number of DAGs is comparable to the number of **stars in the known universe!**
- Enumeration is simply not possible.
- We can still **estimate** the posterior distribution by using Monte Carlo.
- The most common method is the **Metropolis-Hastings** algorithm.
- We move around graph space by adding, deleting and reversing edges.
- Moves are accepted or rejected with certain probabilities to ensure **detailed balance**.

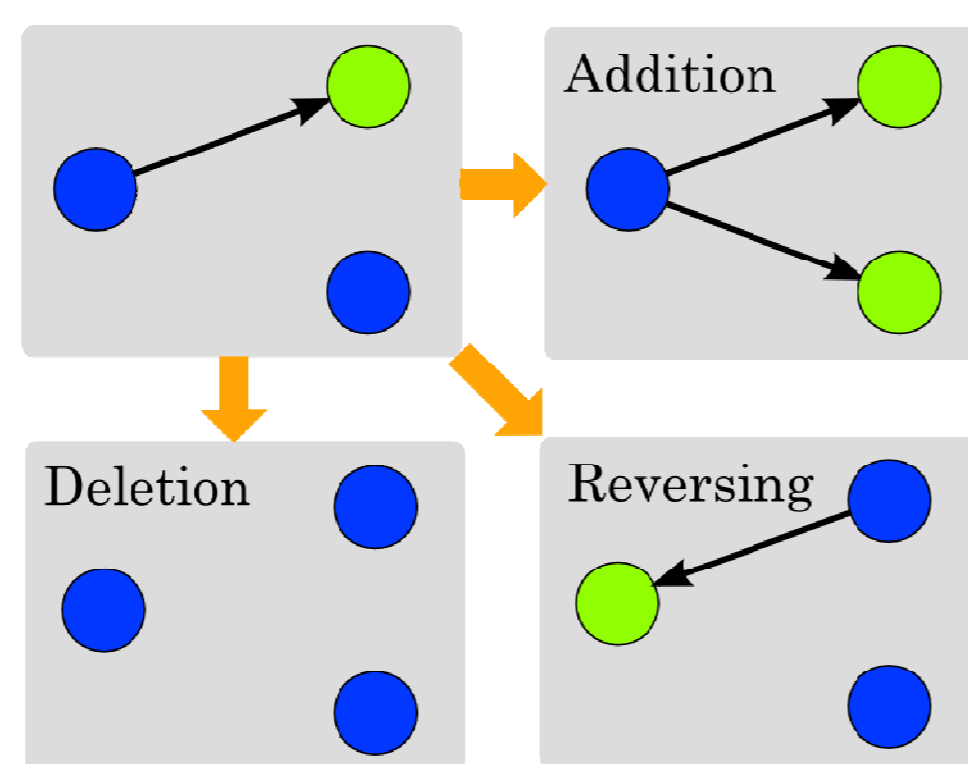


Figure 2: Possible moves in graph space consist of adding, deleting or reversing single edges. All are possible as long as the resulting graph is still acyclic.

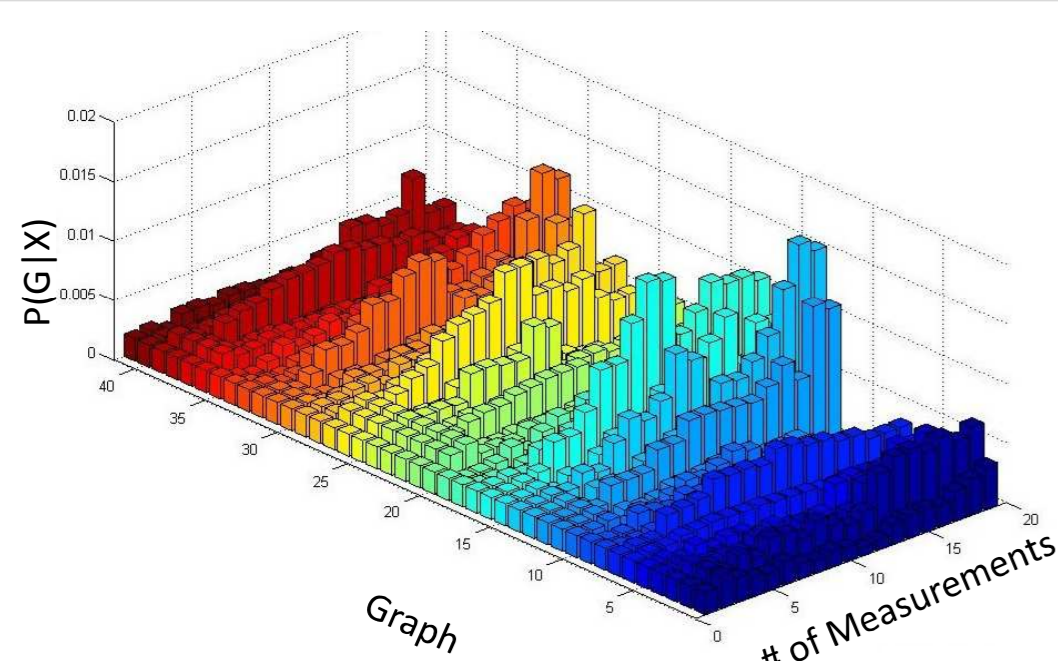


Figure 3: As we increase the number of measurements the probability distribution becomes more peaked. This can cause trouble if the peaks are highly separated.

- Metropolis-Hastings can be slow to converge (mix).
- Also, more data can pose a challenge for Metropolis-Hastings.
- If the most likely graphs are **highly separated** our estimate of the posterior will be poor.
- Schemes such as **simulated tempering** and **tunnelling** Monte Carlo attempt to overcome this.

Alternative Schemes

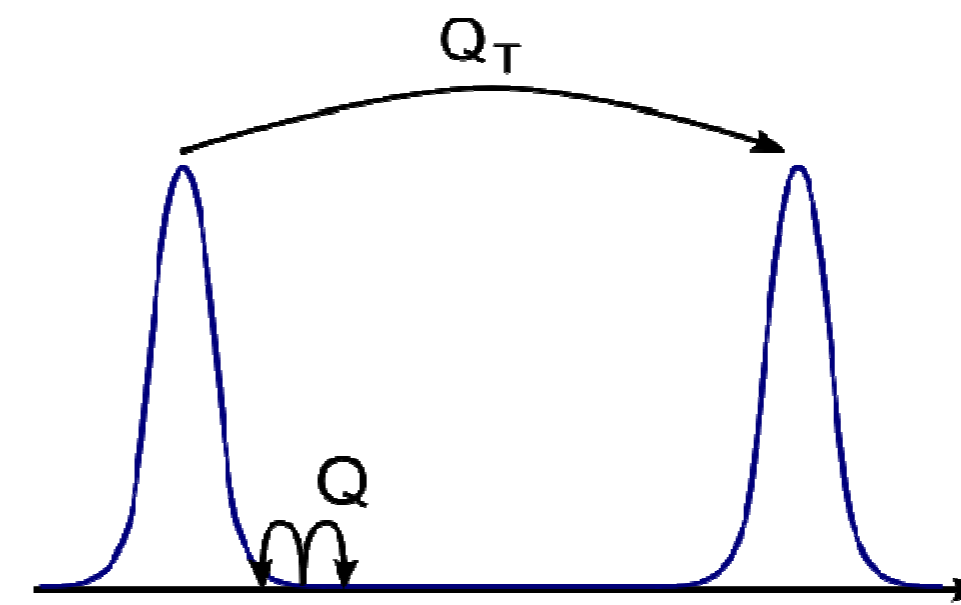


Figure 4: Tunnelling moves avoid getting stuck around local maxima.

- **Simulated Tempering** makes moves at various temperatures.
- The higher temperature moves allow the chain to **move freely** through graph space.
- We can only keep the $\beta=1$ samples so we must generate **many samples** for every one we retain.
- **Tunnelling** Monte Carlo supplements the edge flipping moves of Metropolis-Hastings.
- Long range jumps between **pre-computed** modes are possible.
- These long range jumps stop the Markov chain getting stuck in **local maxima**.

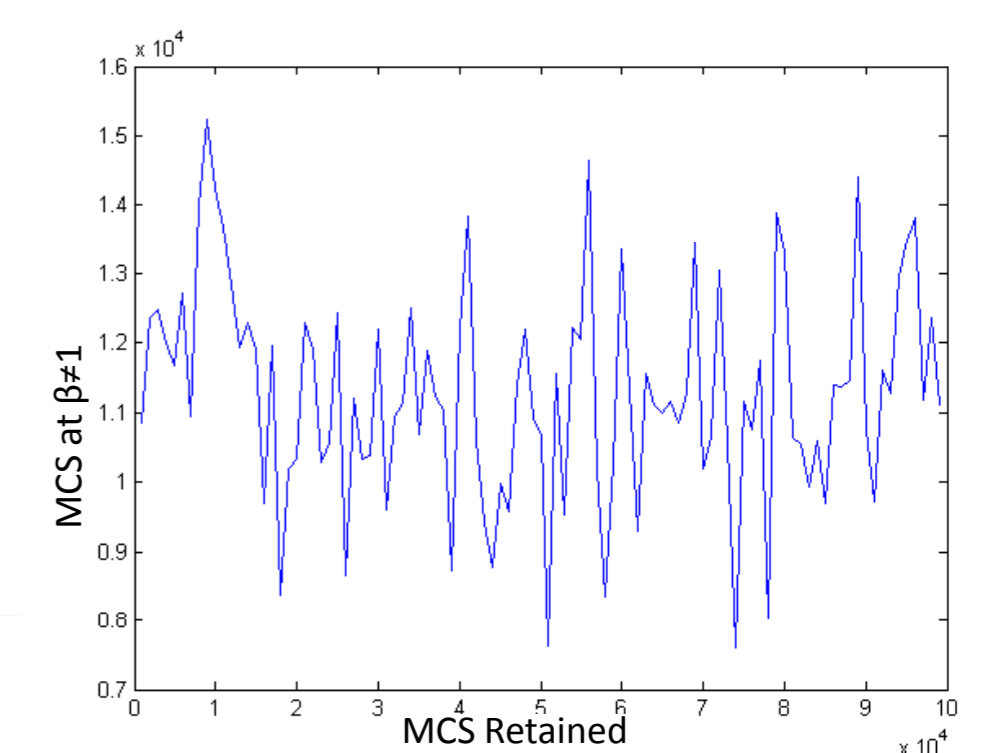


Figure 5: To generate a few correct samples takes many disregarded samples in simulated tempering.

Initial Results

- Convergence to the posterior is measured (for 4 and 5 nodes) using the **total variation norm**.

$$\Delta P = \sum_{i \in \mathcal{G}} |p_i - q_i|$$

- Initial results show that in the **high entropy** regime considered methods provide no advantage over MH.
- In fact ST performs much **worse** since we must disregard many samples.

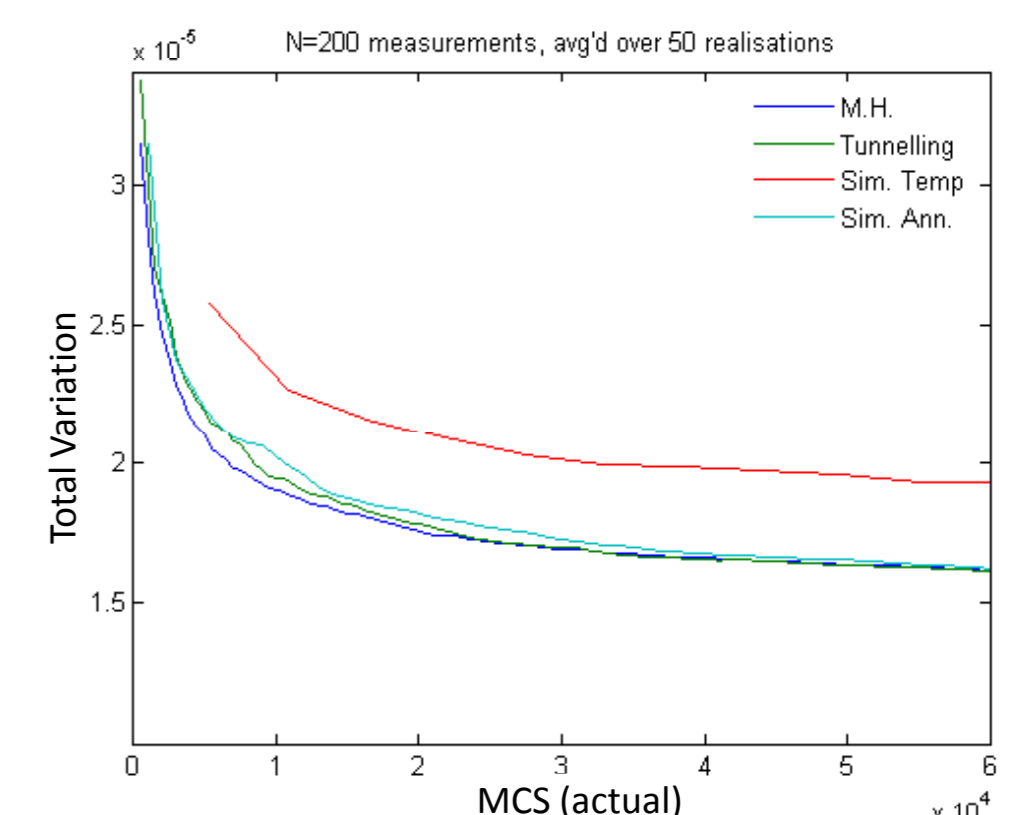


Figure 6: For 5 nodes in the high entropy case all schemes converge relatively quickly (except simulated tempering).

Extensions & Further Work

- In the **low entropy** (many data) regime the tunnelling and simulated tempering schemes may offer some advantage.
 - Run the schemes with **more data**.
 - **Computation time** becomes more of an issue.
- Modification of the **proposal distribution** for use with MH algorithm.
 - Changing more than one edge at a time.
 - Possible application of **cluster algorithms** from statistical physics such as **Swendsen-Wang**.

References

- [1] S. Mukherjee and T. P. Speed. Network inference using informative priors. *PNAS*, 105(38):14313-14318, Sept. 2008.
- [2] E. Marinari and G. Parisi. Simulated tempering: a new monte carlo scheme. *Europhys. Lett.*, 19(6):451 – 458, July 1992
- [3] C. J. Geyer and E. A. Thompson. Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association*, 90(431):909-920, Sept. 1995.
- [4] S. L. Lauritzen. *Graphical Models*. O.U.P. 1996
- [5] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chem. Phys.*, 21(6):1087-1092, 1953.