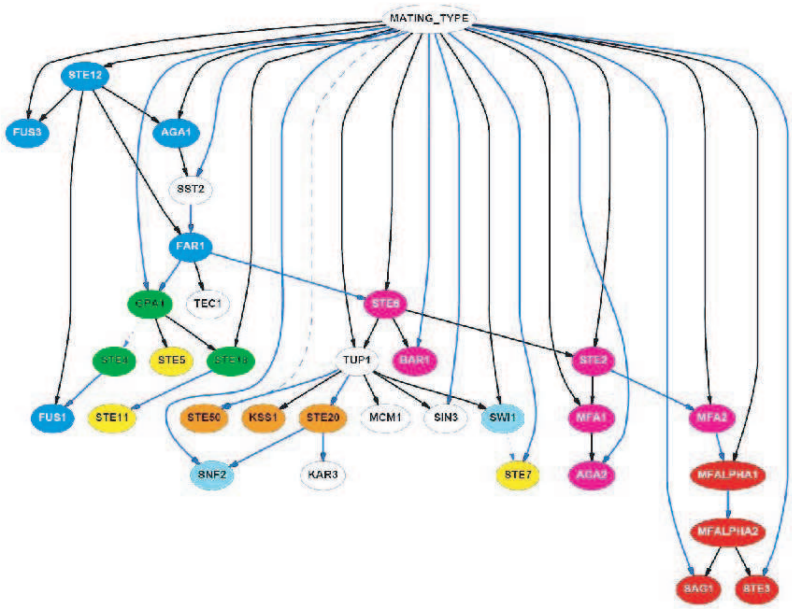
Chapter 5

Networks

Such systems also include biological systems on various levels of organization, like genetic networks, metabolic networks, immune system, neural networks, organisms, ecosystems, biosphere etc.
A Boolean network G(V,F) is defined by a set of nodes corresponding to genes V = x1, . . . , xn and a list of Boolean functions F = (f1, . . . , fn).

5.1 Cytoskeletal Networks

One of the principle components of the cytoskeleton, and in fact one of the most prevalent proteins in the cell is actin



Subnetwork of yeat cell regulation
(taken from A.J. Hartemink, D.K. Gifford, T.S. Jaakkola and R.A. Young Pac Smp. Biocomput. 437-449 (2002)

Figure 5.1: Example of a cell regulation network

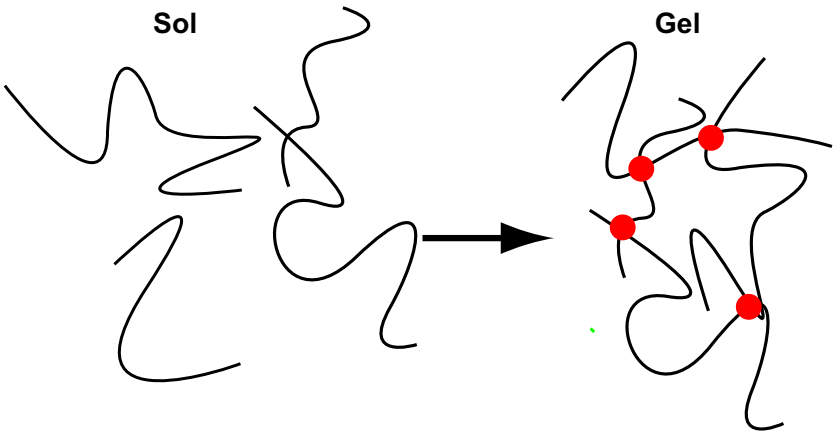


Figure 5.2: Crosslinking in the sol and gel phase

5.2 Gels

Many systems exhibit a network structure (see figure 5.2). The mechanical properties of such networks are often characteristic of a solid even though they being disordered and may even be mostly liquid like. The structure of such a network is described by the following structure parameters. First there is an elastically active network chain between two crosslinks. Then there are dangling chains which are attached to the network by a single point. The gel fraction includes all the material attached to the network.

Let us look at the classical model of gelation proposed by Stockmayer and Flory [236, 237, 238] for the sol gel transition. The model assumes monomers having f valences which can bind (simple) to other monomers to form a network. Not all valences are bound. We assume that each valence has probability of $p(T, c_i, ...)$ of being saturated. Here p is considered to be function of temperature T, concentration c etc.

Take the generating function

$$F_0 = \sum_n \omega_n^{(0)}(p)\theta^n \tag{5.1}$$

where $\omega_n(p)$ is the probability that a randomly picked monomer (A) is part of a cluster of $n+1$ monomers. θ is a factor for every monomer in the cluster except for the considered monomer.

Now

$$F_0 = [1 - p + p\theta F_1(p, \theta]^f = [(1-p)u(p, \theta)]^f \tag{5.2}$$

where F_1 (with $\omega_m^{(1)}$ being the function corresponding to the situation that monomer B, attached to monomer A at the ν-th valence of A, is attached to m other monomers.

To make progress on the computation we assume that the f bonds of A are uncorrelated and that the network has no loops (Cayley-tree, see figure 5.3) (thus a mean-field approximation). Hence

$$F_1(p, \theta) = [1 - p + p\theta F_1(p, \theta)]^{f-1} = [(1-p)u(p, \theta)]^{f-1} \tag{5.3}$$

Solving this we find F_0 and find all $\omega_n^{(0)}$

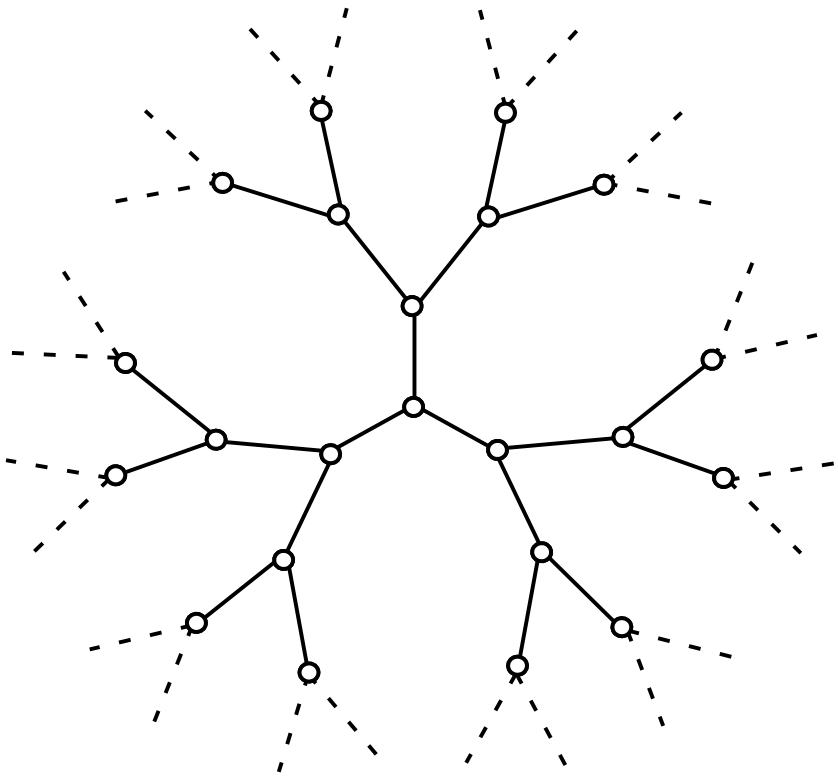


Figure 5.3: The Cayley tree is a tree with all nodes having the same connectivity, here $q = 3$ and without loops

$$\frac{1}{p\theta(1-p)^{f-2}}\left(\frac{p\theta}{1-p}F_1\right) = \left(1+\frac{p\theta}{1-p}\right)^{f-1} \tag{5.4}$$

or

$$u - 1 = xu^{f-1} \tag{5.5}$$

Thus the function $u = 1 + F_1 p\theta/(1-p)$ is a function of $x = p\theta(1-p)^{f-2}$ alone. One can calculate the Taylor-expansion of u

$$u = \sum_{n=0}^{\infty} \frac{[(f-1)n]!x^n}{[(f-2)n+1]!n!} \tag{5.6}$$

from which we find

$$F_1 = (1-p)^{f-1}\sum_{n=1}^{\infty}\frac{[(f-1)n]!}{[(f-2)n+1]!n!}x^{n-1} \tag{5.7}$$

$$F_0 = (1-p)^{f} f\sum_{n=1}^{\infty}\frac{[(f-1)n]!}{[(f-2)n+2]![n-1]!}x^{n-1} \tag{5.8}$$

$$\tag{5.9}$$

To understand F_1 we go back to equation 5.3 and write

$$F_1 = w^{f-1} \tag{5.10}$$

$$\frac{w-1}{p\theta} = w^{f-1} - 1/\theta \tag{5.11}$$

Since w is a monotonically increasing function of θ we find that only one intersection is possible. Assume $\theta = 1$, then there are two cases. Case one

$$f - 1 < 1/p \qquad \text{oder} \qquad p < \frac{1}{f-1} \tag{5.12}$$

This mean that at monomer B the average number of bonds is

$$p(f-1) < 1 \tag{5.13}$$

The chain forming dies out

$$\sum \omega_n(p) = 1 \tag{5.14}$$

and we only find finite connectivity. The system is in the sol phase.
If

$$f - 1 > 1/p \qquad \text{or} \qquad p(f-1) > 1 \tag{5.15}$$

the from generation to generation the number of bonds increase. The *gel point* in this theory is thus

$$p_c = \frac{1}{f-1} \tag{5.16}$$

We shall now investigate the behaviour of the system in the neighbourhood of the gel point. To do so, we will expand various property function around

$$\Delta = p - p_c = p - \frac{1}{f-1} \qquad . \tag{5.17}$$

The gel fraction is given by

$$G = 1 - F_0 = 2f\frac{f-1}{f-2}\Delta \sim \Delta^\beta \quad , \quad \beta = 1 \tag{5.18}$$

Next we look at the *polymerization index* N_ω

$$N_\omega = \frac{\sum_n n\omega_n(p)}{\sum_n \omega_n(p)} = \frac{\partial}{\partial\theta} \ln F_0(\theta = 1) \tag{5.19}$$

We have

$$F_0' = pf(1 + F_1')F_1 \tag{5.20}$$
$$F_1' = p(f-1)(1 + F_1')F_1^{(f-2)/(f-1)} \tag{5.21}$$

and for $\theta = 1$, $p < p_c$, i.e.,$F_0 = F_1 = 1$

$$N_\omega = \frac{pf}{1 - p(f-1)} \approx \frac{f}{|\Delta|} \sim |\Delta|^{-\gamma} \tag{5.22}$$

The next observable we investigate is the cluster size distribution. With the help of the Stirling formula we find from equation 5.9

$$\omega_n \sim e^{-nq(\Delta} n^{-3/2} \sim n^{3/2} e^{-n\Delta^{1/\sigma}} \tag{5.23}$$

with

$$\begin{aligned} q(\Delta) &= -(f-2)\ln\left(1-\frac{f-1}{f-2}\Delta\right) - \ln[1+(f-1)\Delta] && (5.24)\\ &\approx \frac{1}{2}\frac{(f-1)^3}{f-2}\Delta^2 && (5.25) \end{aligned}$$

All in all we find the typical signature for a second-order phase transition with critical exponents $\beta, \gamma, \sigma, \tau$ given by

$$\begin{aligned} G &\sim \Delta^\beta \quad , \quad \beta = 1 && (5.26)\\ N &\sim |\Delta|^{-\gamma} \quad , \quad \gamma = 1 && (5.27)\\ \omega_n &\sim n^{-\tau} e^{-n|\Delta|^{1/\sigma}} \quad , \quad \tau = 3/2, \sigma = 1/2 && (5.28) \end{aligned}$$

Example 5.2.1 (Percolation)
Consider a lattice, which we take, for simplicity, as a two-dimensional square lattice. Each lattice site can be either occupied or unoccupied. A site is occupied with a probability $p \in [0,1]$ and unoccupied with a probability $1-p$. If p is small then only a tiny fraction of the sites is occupied and only small isolated patches of occupied sites that are close to each other exist. On the other hand, if p is large, then large patches of near lying occupied sites exist. We may at this point speculate that for p less than a certain probability p_c only finite clusters exist on the lattice. We define more precisely what we mean by patches or clusters by saying: A cluster is a collection of occupied sites connected by nearest-neighbour distances (see Figure 5.4). For p larger than or equal to p_c, there exists a large cluster (for an infinite lattice, i.e., in the thermodynamic limit) such that for an infinite lattice the fraction of sites belonging to the largest cluster is zero below p_c, and non zero above p_c. For $d = 1$ the situation is straight-forward: $p_c = 1$. For $d > 1$ the computation of p_c is non-trivial, and analytic results for the percolation threshold p_c are only available for certain dimensions or special lattices.

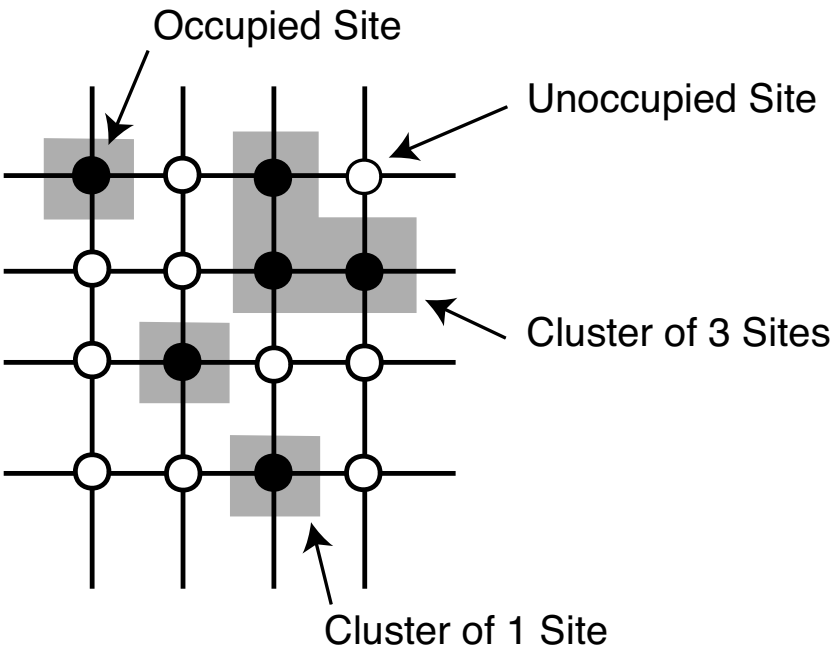


Figure 5.4: Definition of the percolation problem

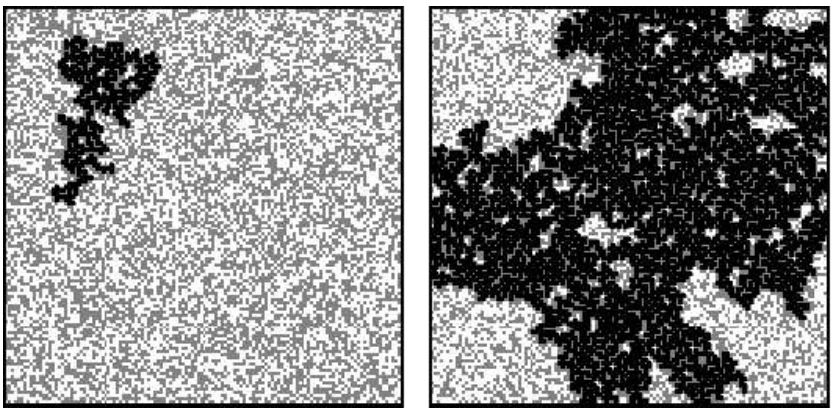


Figure 5.5: Snapshot

5.2.1 Gel-Electrophoresis

Important Concepts: repton model, master equation approach

A widely used and simple method for the separation of DNA fragments of different length is DC *gel-electrophoresis*. The DNA mixture to be separated is introduced into a gel matrix. Since the DNA is charged, it will move in a constant electric field E with velocity $v(E, N)$ where N is the length of the fragment. After some time fragments of different length will have travelled a distance in the gel depending on their length and can therefore be separated.
A simple model to study motion of the polymer chain is to map the chain onto a lattice in which reptation dynamics is mapped onto a lattice [239]. The figure 5.6 shows a configuration of the polymer as represented in the *repton model*. The polymer occupies a connected series of neighboring cells of a lattice. Multiple occupancy by monomers (or reptons) of a lattice site is allowed. Whenever two or more reptons occupy the same site some stored length accumulates along the polymer.
The stored length diffuses along the chain. When a "unit" of stored length formed at one edge of the chain moves out of the polymer at the opposite edge the chain center of mass is slightly shifted.
This can be formulated as follows. An inner monomer in the repton model is in one of the following states:

1. the monomer is in the same pore as both nearest neighbours
2. both nearest neighbours are in adjacent pores
3. the monomer has one nearest neighbour in the same pore and the other one is in an adjacent pore

In states (1) and (2) the monomer is not allowd to move. In state (3) the only allowed move is the move where the monomer joins its neighbour in the adjacent pore. The end monomers may be in one of two states:

- The nearest neighbour is either in an adjacent pore or
- in the same pore

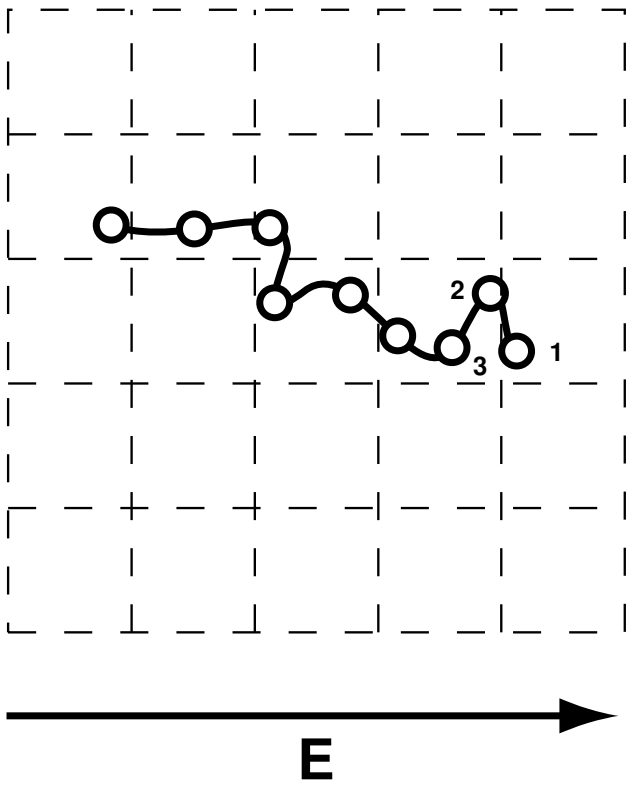


Figure 5.6: Repton model

In the first state, the monomer may join its neighbour in the adjacent pore. In the second state, the monomer is free to move to any of the six adjacent pores (in 3 D). Thus, when neighbouring reptons on the chain lie within the same gel pore a unit of stored length separates them. Otherwise the reptons must occupy neighboring gel pores, because the bond between reptons prevents further excursions. We label the occupied gel pores with index $i = 0, 2, ..., L$. The bond from one pore $i?1$ to the neighboring pore i is denoted by $s_i = \pm 1$, where $i = 1, 2, ...L$. The direction of pore labeling i is chosen as follows: the "head´´ of the polymer is the end that is furthest advanced in the direction of applied force. The other end is the "tail". Labeling begins at the tail and runs to the head. As the polymer drifts, head and tail may occasionally interchange. The set of all bonds s defines the conformation of the polymer's tube.

We shall denote the stored length iby the variable n_i, i.e. for the units of stored

length in gel pore i. The set of stored length occupancy data for all gel pores is denoted $\mathbf{n}$.
The potential energy of the polymer changes by Es_i when a unit of stored length moves from pore $i?1$ to i. Allowed moves occur with frequency ωB^{s_i} where the Boltzmann factor $B = e^{E/2}$.
We start off writing a Master equation for the process ([240])

$$\sum_{\mathbf{n}',\mathbf{s}'} W(\mathbf{n},\mathbf{s}|\mathbf{n}',\mathbf{s}')P(\mathbf{n}';\mathbf{s}') = \sum_{\mathbf{n}',\mathbf{s}'} W(\mathbf{n}',\mathbf{s}'|\mathbf{n},\mathbf{s})P(\mathbf{n};\mathbf{s}) \tag{5.29}$$

from which we can obtain the (tube dependent, i.e., on the conformation dependent) velocity

$$v(\mathbf{s}) = \sum_{\mathbf{n}} v(\mathbf{n};\mathbf{s})P(\mathbf{n};\mathbf{s}) \sum_{\mathbf{n}} P(\mathbf{n};\mathbf{s}) \tag{5.30}$$

To make progress, we replace the full Master equation by assuming that the conformation $\mathbf{s}$ is kept fixed. Further, we assume periodic boundary conditions (see later)

$$\sum_{i=0}^{L} \theta(n_i) \left[B^{-s_{i+1}} P(..,n_i-1,n_{i+1}+1,...,\mathbf{s}) + B^{s_i} P(..,n_{i-1}+1,n_i-1,...,\mathbf{s})\right] = \sum_{i=0}^{L} \theta(n_i) \left[B^{s_{i+1}} + B^{-s_i}\right] P(\mathbf{n},\mathbf{s}) \tag{5.31}$$

θ is the Heaviside function ensuring that n_i is populated. To solve this equation we can try a product ansatz

$$P(\mathbf{n},\mathbf{s}) = \prod_{i=0}^{L} p_i^{n_i} \tag{5.32}$$

assuming a fixed conformation $\mathbf{s}$ and boundary conditions

$$p_{-1} \equiv p_{L+1} \equiv 1 \quad . \tag{5.33}$$

If we plug this into eq (5.31) we obtain

$$B^{-s_{i+1}} p_{i+1} + B^{s_i} p_{i-1} = \left[B^{s_{i+1}} + B^{-s_i}\right] p_i \tag{5.34}$$

or

$$B^{-s_{i+1}}p_{i+1} - B^{s_{i+1}}p_i + B^{s_i}p_{i-1} - B^{-s_i}p_i = 0 \tag{5.35}$$

where can introduce the flux c

$$c = B^{-s_i}p_{i-1} - B^{-s_i}p_i \tag{5.36}$$

which again is linear recurrence relation readily being solved by the following MAPLE code

MAPLE 5.2.1

```
restart;
eqn1 := c = B^(s(i))*p(i-1) - B^(-s(i))*p(i);
init := p(-1) = d;
sol := rsolve({eqn1,init},p);
```

The solution is

$$sol := B^{-2\,s(i+1)} \prod_{i0=-1}^{i} B^{2\,s(i0+1)} \left(-c \sum_{i1=-1}^{i-1} \left(\frac{B^{s(i1+1)}}{\prod_{i0=-1}^{i1} B^{2\,s(i0+1)}} \right) + d \right) \tag{5.37}$$

Using the abbreviation

$$h_i \equiv \sum_{j=1}^{i} s_j \tag{5.38}$$

we arrive at an expression for p_i

$$p_i = p_{-1}B^{2h_i} - c\sum_{k=0}^{i} B^{2(h_i - h_k)+s_k} \tag{5.39}$$

Since now have a result for the p_i we can solve for c

MAPLE 5.2.2

```
restart;
eqn1 := c = B^(s(i))*p(i-1) - B^(-s(i))*p(i);
init := p(-1) = d;
p(i) := rsolve({eqn1,init},p);
```

```
eqn2 := c - B^(s(i+1))*p(i) + B^(-s(i+1))*p(i+1) = 0;
sol := solve (eqn2,c);
```

to find

$$\frac{c(\mathbf{s}) = B^{2R_e^{||}} - 1}{\sum_{k=0}^{L+1} B^{2(R_e^{||} - h_k) + s_k}} \tag{5.40}$$

where we have introduced the end-to-end distance S parallel to the electric field E

$$R_e^{||} = \sum_{i=1}^{L} s_i \tag{5.41}$$

For the velocity we write, using the imbalance due to the electric field

$$v(\mathbf{n};\mathbf{s}) = \frac{\alpha\omega}{N} \sum_{i=0}^{L} \theta(n_i) \left[B^{s_{i+1}} s_{i+1} - B^{-s_i} s_i \right] \tag{5.42}$$

Let $\bar{\theta}$ be the average of $\theta(n_i)$ over the stored length

$$\bar{\theta} = \sum_{\mathbf{n}} \theta(n_i) P(\mathbf{n};\mathbf{s}) \sum_{\mathbf{n}} P(\mathbf{n};\mathbf{s}) \quad . \tag{5.43}$$

which yields with the prodcut ansatz (5.32)

$$\bar{\theta}_i = \sum_{\mathbf{n}} \theta(n_i) \prod_{j=0}^{L} p_j^{n_j} / \sum_{\mathbf{n}} \prod_{j=0}^{L} p_j^{n_i)} \tag{5.44}$$

If we are allowed to interchange the summation over $\mathbf{n}$ with the over i we find for the velocity

$$v(\mathbf{s}) = \frac{\alpha\omega}{N} \sum_{i=0}^{L} \bar{\theta}_i \left[B^{s_{i+1}} s_{i+1} - B^{-s_i} s_i \right] \tag{5.45}$$

Let

$$Q(N_{\mathbf{s}}) = \sum_{\mathbf{n}} \prod_{j=0}^{L} p_j^{n_i} \tag{5.46}$$

with

$$\sum_{j=0}^{L} n_j = N_{\mathbf{s}} \tag{5.47}$$

the partition functions for the stored length distribution. The sums in eq 5.44 allocates $N_{\mathbf{s}} = N_0$ unit oif stored length in the denominator but allocates only $N_{\mathbf{s}} = N_0 - 1$ units of length in the numerator. Define

$$\bar{z}(\mathbf{s}) \equiv Q(N_0 - 1)/Q(N_0) \tag{5.48}$$

as the fugacity for the stored length. Then

$$\bar{\theta}_i = p_i \bar{z}(\mathbf{s}) \quad . \tag{5.49}$$

Upon substitution this into the equation for the velocity we obtain

$$v(\mathbf{s}) = \frac{\alpha\omega\bar{z}(\mathbf{s})}{N} \sum_{i=0}^{L} \left[B^{s_i} p_{i-1} - B^{-s_i} p_i \right] s_i \tag{5.50}$$

Using the above defined flux we arrive at

$$\frac{v(\mathbf{s}) = \alpha\omega R_e^{||} c() \bar{z}(\mathbf{s})}{N} \tag{5.51}$$

5.3 Metabolic Networks

5.4 Bayesian Networks

Let $G = (V, E)$ be a directed acyclic graph. We assume that the vertices $i \in V$ $(1 \leq i \leq n)$ represent for example genes and correspond to random variables x_i. For each y_i we define a conditioned probability

$$P(x_i|\text{parent}(x_i)) = P(x_i|P_a(x_i)) \tag{5.52}$$

and the joint probability distribution

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i|P_a(x_i)) \tag{5.53}$$

We define a parameter θ to be the set of conditioned probabilities