8.592J–HST.452J: Statistical Physics in Biology

Assignment # **6** Due: **5/6/22**

Networks & Cooperativity

1. *Percolation on the Cayley tree:* In simple models of percolation, elements of a lattice (sites or bonds) are independently occupied with a probability p. A cluster is defined as a connected (by neighboring bonds) set of these occupied elements. At small p, only small clusters exist, and the probability that two sites, separated by a distance r, are connected to each other decays as $\exp(-r/\xi)$. The correlation length $\xi(p)$ grows with increasing p, diverging at the *percolation threshold* p_c as $\xi(p) \sim |p_c - p|^{-\nu}$. A so-called *infinite cluster*, spanning the entire system first appears at the percolation threshold p_c, covering more and more sites for $p > p_c$. We can define a probability $P(p)$ that a site belongs to this infinite cluster which, on approaching p_c from above, vanishes as $P(p) \sim |p_c - p|^{\beta}$.

(a) The *Cayley tree* is a hierarchical lattice in which each site at one level is connected to z sites at the level below. Thus the n-th level of the tree has z^n sites. For $z = 2$, obtain a recursion relation for the probability $P_n(p)$ that the top site of a tree of n levels is connected to some site at the bottom level.

(b) Find the limiting behavior of $P_\infty \equiv P(p)$ for infinitely many levels. Give the exponent β characterizing the vanishing of $P(p)$ at p_c.

(c) Show that for starting values close to $P(p)$, the recursion relations admit solutions of the form $P_n = P_\infty e^{-n/\xi}$. Find expressions for ξ for both $p < p_c$ and $p > P_c$, and hence obtain the exponent ν for the divergence of ξ at the percolation threshold (for $z = 2$).

(d) Find the value of p_c for any branching number z. Do the exponents β and ν depend on z?

*****

2. *Preferential network growth with node removal:* Consider the following extension of network growth by preferential attachment, as explored in *C. Moore, G. Ghoshal, and M.E.J. Newman, Phys. Rev. E* **74***, 036121 (2006)*: At each time step a new node is created, its m links attached preferentially proportional to the number of links already present for a pre-existing node. However, before the next node is added, a randomly selected node is removed with probability r.

(a) Show that after many steps t, the average number of nodes and links grow as $N(t) = t(1-r)$ and $L(t) = tm(1-r)/(1+r)$, respectively.

(b) Write down the recursion relation governing the probability $p_k(t)$ of nodes with k links at time t.

(c) Show that in steady state $p_k^* \propto k^{-\gamma}$ for large k, and find the exponent γ.

(d) If instead of nodes, a link is randomly removed at each step with probability s, what happens?

*****

3. (Optional) *'Feed-down' network:* Consider a set of non-negative variables $\{x_n(t)\}$ (e.g.

chemical concentrations), evolving in time according to first order differential equations

$$\frac{dx_n}{dt} = f_n\left(x_{n+1}, x_{n+2}, x_{n+3}, \cdots\right) - g_n\left(x_n\right).$$

While the equations are quite general, we make the following assumptions:
(i) Each component decays at a rate $g_n \geq 0$ which only depends on its value x_n; $g_n(x_n)$ is a monotonically increasing function of x_n, with $g_n(0) = 0$.
(ii) Each component is generated at a rate $0 \leq f_n \leq \max(g_n)$, which can depend only on variables numbered higher than n; i.e. f_n does not depend on $x_1, x_2, \cdots, x_n$.

(a) By considering eigenvalues of the stability matrix show that these equations admit a *stable* fixed point.

(b) If the variables $\{x_n(t)\}$ are made space dependent and allowed to diffuse, such that a term $D_n\nabla^2 x_n$ is added to $\partial x_n/\partial t$, can these equations admit Turing patterns?

(c) If all f_n are monotonically increasing functions of their arguments, show that starting from $\{x_n(t=0) = 0\}$, the variables proceed monotonically to the fixed point values.

*****

4. *Hopfield network with correlated states:* In the paper PLoS Comput Biol. 2014 Aug; 10(8): e1003734 (arXiv:1211.3133) a type of Hopfield Lyapunov function is used to characterize the epigenetic landscape of cells. The expression profiles of transcription factors (simplified to a binary code of off or on for roughly hundred TFs) are specific to each cell type (e.g. liver, skin, heart, $\cdots$), and are modeled as 'associative memories' in the parlance of neural networks. An important subtlety is that unlike typical 'memories' stored in a neural net, the expression profiles are highly correlated. A corresponding variant of the Hopfield model is examined in this problem.

The desired states are characterized by set of binary vectors $\{\vec{\xi^\mu}\}$ for $\mu = 1, 2, \cdots, M$; each vector has components $\xi_i^\mu = \pm 1$ with $i = 1, 2, \cdots, N$. We would like to encode these states into the couplings $\{J_{ij}\}$ of a Hopfield network, composed of variables $\{-1 \leq x_i \leq +1\}$, evolving as

$$\frac{dx_i}{dt} = -\frac{x_i}{\tau} + \tanh\left(h_i + \sum_j J_{ij} x_j\right).$$

(a) Compute a Lyapunov function that is minimized by the above dynamics.

(b) For one state vector $\vec{\xi}$, show that the couplings $J_{ij} = \xi_i \xi_j / N$ (with $h_i = 0$) enable recovery of the pattern, provided $\tau > \tau_c$.

(c) Consider a set of M uncorrelated states, corresponding to orthogonal binary vectors such that $\sum_i \xi_i^\mu \xi_i^\nu = N\delta_{\mu\nu}$. Show that multiple states can be encoded via $J_{ij} = \sum_\mu \xi_i^\mu \xi_j^\mu / N$ (again with $h_i = 0$).

(d) We now relax the condition of orthogonal memories. Show that in this case states can be encoded through the couplings $J_{ij} = \sum_{\mu\nu} [\xi_i^\mu (C^{-1})_{\mu\nu} \xi_j^\nu]/N$, where C^{-1} is the inverse of the correlation matrix, whose elements are $C_{\mu\nu} = \sum_i \xi_i^\mu \xi_i^\nu / N$.

*****

5. *MWC model of hemoglobin:* Consider the Monod, Wyman, Changeux (MWC) model of cooperativity in hemoglobin: $n = 4$ sites that can be in a relaxed (R) or tense (T) state, the latter favored by a factor L, with normalized oxygen fugacities of α and $c\alpha$ per site in the two states, such that the overall fractional occupancy of oxygen is

$$Y = \alpha \frac{(1+\alpha)^{n-1} + Lc(1+c\alpha)^{n-1}}{(1+\alpha)^n + L(1+c\alpha)^n} \, .$$

(a) Estimate the effective Hill coefficient n_h for hemoglobin, either from the maximal slope of $\log[Y/(1-Y)]$ vs $\log \alpha$ in a Hill plot, or from the slope at half-saturation. Comment on the relative importance of parameters $c \ll 1$ and $L \gg 1$ in determining n_h, by considering the limits of $c \to 0$ and $L \to \infty$.

(b) How does the probability of being in the T state change as more oxygen molecules are bound? Calculate the equilibrium probability P_T of being in the T state, and plot it as a function of the normalized oxygen pressure α. Next calculate the conditional probability $P_T(i)$ of being in the T state, given that i molecules of oxygen are bound. Plot $P_T(i)$ as a function of i. What is the effect of oxygen on hemoglobin conformation?

(c) Different species have different mechanisms of adaptation to hypoxia at high altitude. In humans, adaptation to high altitude involves, among other factors, rapid elevation of the level of $2,3-DPG$ (aka $2,3-BPG$) molecule, which is synthesized in red-blood cells and binds preferentially to the T state of hemoglobin. Some birds that fly at high altitude are adapted by having the hemoglobin with a mutation at the interface between its α and β domains, making the transition between the two conformational states of hemoglobin easier, i.e. reducing the free energy difference between the R and T states. How do these two mechanisms of adaptation affect the saturation curve of hemoglobin? Do you expect them to have similar or different effects on oxygen uptake in the lungs, and release in the tissues?

*****

6. (optional) *Binding to non-identical sites:* Consider a variant of the MWC model, with the sites that have different affinities, $K_1, K_2, \cdots, K_n$, for oxygen.

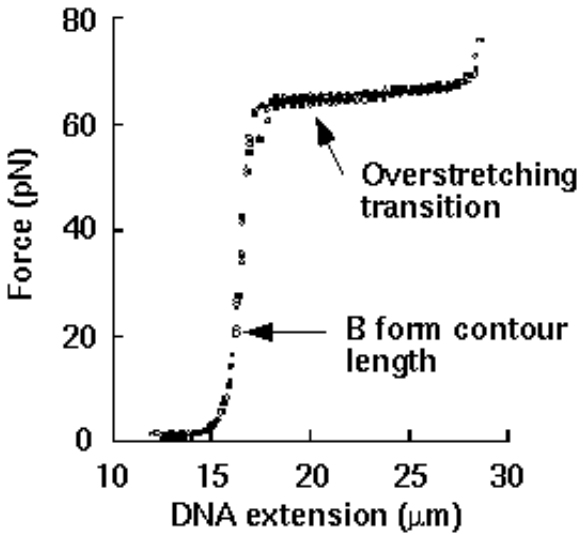
(a) Show that at both very low, and very high pressures, the model can be approximated by the classical MWC model with identical sites, but with an effective binding constant K_{eff}. In either regime, which sites contribute most to K_{eff}, the strongest or the weakest?

(b) Can the above model, in the simplified limit of only one state – say for $L = 0$, lead to a sigmoidal Hill plot with $n_h < n$?

(c) Consider another variant of the model, with different affinities but with extreme cooperativity as considered by Hill, i.e. with either no site occupied or all sites occupied. Can such a model provide a sigmoidal Hill plot with $n_h < n$?

*****

7. *Over-stretching DNA:* In standard (B–from) DNA the basepairs stack in spiral fashion at separation of 3.4Å. As indicated in the following figure [from S. B. Smith, Y. Cui, C. Bustamante, Science **271**, 795 (1996), and *http://alice.berkeley.edu/˜ steve/DNAstr.html*], pulling on DNA with optical tweezers causes it to greatly stretch at forces of around 65±5pN.



(a) One interpretation is that this represents a transition to a new structure of over-stretched DNA, in which the separation of bases has increased to 5.8Å. As a very simple model of this putative state consider DNA as a one dimensional chain in which each unit can either be in the regular form of size 3.4Å, or in the stretched form of size 5.8Å. Assume that an energy U is required to change the regular form to the stretched form. For this part of the problem ignore the three dimensional orientations of each segment, and assume that the state of each element is independent of its neighbors. Calculate the length $L(F,T)$ for this model when pulled by a force F at a temperature T.

(b) Compare the result from part (a) to the experimental figure, and thus estimate the parameter U from the data in the above model from experiments. Is the width of the transition region in F consistent with the assumptions of the model.

(c) Now consider a more realistic model in which neighboring elements tend to be in the same state. Would this lead to a sharpening or widening of the transition region in F?

*****

8. (Optional) *Force–extension curve for stretched DNA:* Configurations of a worm-like chain (WLC) in two dimensions can be characterized by variations of the angle $\theta(s)$ of the tangent at (monomer) position s along the curve. In the presence of a large force $\vec{F}$, variation of the angle (measured with respect to $\vec{F}$) are small, and the energy cost can be approximated as

$$H = \frac{1}{2}\int_0^L ds \frac{\kappa}{R^2(s)} - \int_0^L \vec{ds}\cdot\vec{F} \approx -FL + \frac{1}{2}\int_0^L ds\left[\kappa\left(\frac{d\theta}{ds}\right)^2 + F\theta^2(s)\right].$$

(a) Rewrite the WLC Hamiltonian as a sum over harmonic modes of the DNA "string" $\tilde{\theta}_q$.

(b) Equipartition dictates that each quadratic mode will have an average energy of k_BT. Write the expression for $\langle|\tilde{\theta}_q|^2\rangle$, and hence calculate $\langle\theta^2(s)\rangle$.

(c) Calculate the extension X of the DNA as a function of force F. Invert the relation and plot the function $F(X)$. Use k_BT and persistence length l_p in your answer.

*****