Ising models of strongly coupled biological networks

Lina Merchan[1] and Ilya Nemenman[2]

***Short Abstract* — Biological networks consist of a large number of variables that can be coupled by complex multivariate interactions. However, several neuroscience, cell biology, and genetics investigations have reported that observed statistics of network states can be approximated surprisingly well by pairwise factor models, also known as maximum entropy models that constrain correlations within pairs of variables. Here we argue that this approximation is expected to work for a large class of interaction networks provided such distributions are sufficiently constrained, so that only a handful of well-defined states are highly probable. This observation is related to the Hopfield approximation of the landscape over network states. Success of pairwise models in applications to real systems can be explained by a recent argument that maps network states onto random satisfiability problems and to constrained optimization performed by evolution.**

***Keywords* — pairwise model, biological network, maximum entropy, satisfiability, coarse-graining.**

I. Introduction

BIOLOGICAL systems are complex networks with highly coupled elements. Experimentally, it would be very costly to study all possible interactions among the elements of the network, whether it is a cellular regulatory network or a neural network. Phenomenological coarse-grained models able to approximate statistical dependences in the data without accounting for all microscopic interactions are needed. Several recent studies have shown that pairwise factorial models, known as maximum entropy models, can serve this purpose. In neural [1], cell biology [2,3], and genetics [4] applications, such approximate models provide remarkably accurate fits for dependences in data, where true biological interactions should be a lot more complex. Here we investigate computationally under which conditions the quality of the approximation holds.

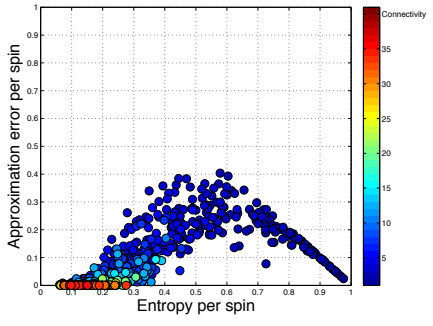
II. Methodology

We describe the state of every variable in the network as a binary variable σ_i (expressed or not expressed), also known as an Ising spin. Then the state of the network is defined by long binary words. The question of validity of the pairwise maximum entropy model [5] is then equivalent to the question of whether the distribution of states of the network can be approximated by a Boltzmann distribution with an effective pairwise Ising Hamiltonian $H_{eff} = -\sum J_{ij}^{eff}\sigma_i\sigma_j - \sum h_i^{eff}\sigma_i$. We investigate networks of N~20 spins coupled by M randomly chosen interactions with random strength, where each interaction couples three or four spins; for example $H_{true} = -\sum J_{ijkl}\sigma_i\sigma_j\sigma_k\sigma_l$. We then use the Blahut-Arimoto algorithm to find an effective pairwise model that best approximates the true distribution.

III. Results and Conclusion

The figure illustrates that pairwise maximum entropy models provide highly accurate approximations to data whenever the overall network connectivity (measured by M and the scale of true couplings J) is high, and entropy per spin is low. We emphasize that this holds even though there are *no* true second order interactions in networks being approximated. In this regime, the number of highly probable network states (which we call phenotypes) is low.

This success of pairwise models can be explained by an observation that a Hopfield network [6] that uses phenotypes as memories is guaranteed to approximate the data well in this "nearly-frozen" regime. As was argued recently [7], evolution drives biological systems to satisfy as many constraints as possible, and hence many systems will find themselves near the satisfiability threshold [8]. The states that satisfy most constraints and are highly probable are few and far in between, which is the regime where pairwise models succeed. Therefore, our findings should be relevant for a wide class of biological systems, suggesting a way of constructing coarse-grained models of biological networks, focused on phenotypes rather than on network variables.



Acknowledgements: This research was partially supported by the James S McDonnell Foundation grant No.220020321.

[1]Department of Physics, Emory University E-mail: lmerchan@moses.physics.emory.edu

[2]Departments of Physics and Biology, Computational and Life Sciences Initiative, Emory University. E-mail: ilya.nemenman@emory.edu

References

[1] Schneidman E, et al. (2006) *Nature* 440, 1007.
[2] Margolin A, et al. (2006) *BMC Bioinformatics*
[3] Mora T, et al. (2010) *PNAS 107 (12)*
[4] Otwinowski J and Nemenman I (2013) *PLoS ONE*
[5] Schneidman, et al. (2003) *PRL 91*
[6] Hopfield JJ, (1982) *Proc. Nat. Acad. Sci.* 79, 2554
[7] A Lander (2012), *Beyond q-bio?* Closing talk at q-bio 2012
[8] Krzakala F, Zdeborova L, et al. (2007) *PNAS* 104 10318–10323