CHAPTER 4

The Connectivity of Large Genetic Networks: Design, History, or Mere Chemistry?

Andreas Wagner*

Abstract

I review evolutionary explanations of broad-tailed connectivity or degree distributions observed in metabolic networks and protein interaction networks. Self-assembled chemical reaction networks show degree distributions similar to those observed for metabolic networks, which argues against the postulated role of natural selection in maintaining this degree distribution. In addition, metabolic networks contain traces of their ancient history in the form of highly connected metabolites. Similarly to the degree distribution of metabolic networks, that of protein interaction networks can be explained without resorting to natural selection on the network level. I present data suggesting that highly connected proteins are not distinguishably older than other proteins, and explain this finding with a simple model of how a protein’s degree changes in evolutionary time.

Introduction

Graph representations of biological networks have become popular with the recent accumulation of functional genomic data on such networks. Graphs are mathematical objects consisting of nodes and edges connecting these nodes. The degree or connectivity $d$ of a node is the number of edges emanating from it, or, equivalently, the number of its neighbors in the graph. Multiple biological networks show a connectivity or degree distribution that is broad-tailed and often consistent with a power-law. That is, when choosing a node from such a network at random, the probability $P(d)$ that it has $d$ interaction partners is proportional to $d^{-\gamma}$, $\gamma$ being some constant that is characteristic of the network. Most prominently, this holds for metabolic networks, whose nodes can be substrates, reactions, or both, depending on the network representation one chooses, and protein interaction networks, where two nodes (proteins) are connected if they interact physically inside the cell. Broad-tailed degree distributions have also been demonstrated for other cellular networks.1,3

The degree distribution of a genetic network can be viewed as a feature of an organism like any other feature. It raises the same basic question: Why this and not some other degree distribution? There are three possible answers. First, a network’s degree distribution could be a mere consequence of chemistry, the chemistry of DNA, RNA, and proteins, and the patterns of molecular interactions this chemistry allows. This possibility may seem far-fetched, given that

*Andreas Wagner—University of New Mexico and The Santa Fe Institute, Department of Biology, 167A Castetter Hall, Albuquerque, New Mexico 87131-1091, U.S.A.
Email: wagnera@unm.edu

molecular networks have many biological functions which may constrain their structure. However, this possibility is not without precedent. An illustrative example exists on a lower level of biological organization, protein structure. The thousands of currently known protein structures have a highly skewed distribution. There is a small number of 'frequent' tertiary structures, such as the TIM-barrel or the Rossmann fold, found in nucleotide-binding proteins. While these folds are small in number, many proteins adopt them. Conversely, the majority of tertiary structures are 'unifolds' that may have originated only once in evolution, and are adopted by few proteins. Does this skewed distribution of protein structures contain important information about design principles of proteins? For instance, do frequent structures have superior properties that lead to their frequent occurrence in proteins? The likely answer is no. Similarly skewed distributions of structures—a small number of excessively frequent structures and a vast majority of rare structures—occur in simple models of protein folding, models where polymers composed of parts with properties similar to amino acids fold into three-dimensional structures. The distribution of protein structures may be a mere consequence of polymer chemistry.

The second possibility is that the degree distribution of genetic networks might somehow reflect their history, much like the jumble of streets in a medieval city reflects the city's growth over centuries. An important class of mathematical models, originally devised to explain power-law degree distributions in growing networks like the internet, do indeed link a network's history to its degree distribution. In their original and simplest incarnation, such models involve only two simple rules that change the structure of a network. First, the network grows through addition of nodes. Second, newly added nodes connect to previously existing nodes, such that already highly connected nodes are more likely to receive a new connection than nodes of lesser connectivity. Over many cycles of node addition and linking to existing nodes, a power law degree distribution emerges. A great variety of variations to this model have been proposed (reviewed in ref 10). They differ greatly in detail but retain in some way or another the rule that new connections preferably involve highly connected nodes. Importantly, most such models make a key prediction: Highly connected nodes are old nodes, nodes having been added very early in a network's history. In this sense, they link a network's degree distribution to its history.

The third possibility is that molecular networks have their degree distribution, because this structure is somehow best suited to the network's biological function. From an 'organismal design' perspective, this is the most interesting possibility. It means that natural selection has shaped the global connectivity pattern of a network, and that network structure reveals something about the design principles of biological networks.

A recent hypothesis postulates that the observed broad-tailed degree distribution of biological networks is indeed a product of natural selection. This 'selectionist' hypothesis is based on the following observation. In networks with a broad-tailed degree distribution, the mean distance between network nodes that can be reached from each other (via a path of edges) is very small and it increases only very little upon random removal of nodes. (In contrast, this mean distance or mean path length increases drastically when highly connected nodes are removed.) A network's mean path length can be thought of as a measure of how 'compact' the network is. In graphs with other degree distributions, mean path length increases more substantially upon random node removal, and the network becomes more easily fragmented into disconnected components. These observations have led to the proposition that robustly compact networks confer some advantages on cells, and that a broad-tailed degree distribution reflects the action of natural selection on the degree distribution itself. The nature of this advantage is unknown, except in the case of metabolic networks, where one can venture an informed guess. A possible advantage of small mean path lengths in metabolic networks stems from the importance of minimizing transition times between