Gene Network: Model, Dynamics and Simulation

Shiquan Wu and Xun Gu*

Center of Bioinformatics and Biological Statistics
Iowa State University, Ames, IA 50011, USA
sqwu@cs.iastate.edu, xgu@iastate.edu

Abstract. A gene network is modeled as a dynamical random graph whose vertices and edges represent genes and gene-gene interactions, respectively. The network grows through three biological mechanisms: (1) gene duplication and loss; (2) gene-gene interaction adding and removing; and (3) genome duplication. The evolutionary dynamics of gene networks is discussed. It is shown that: (1) the vertex degree distribution (i.e., the distribution of the number of the gene-gene interactions per gene) always follows power laws and the power law exponents may be changed by genome duplications; and (2) the network degree distribution (i.e., the distribution of the total number of the gene-gene interactions in the network) has a complex behavior: If no genome duplication occurs, it follows a power law. If a genome duplication occurs, it may be away from the power law state. However, after a sufficient long evolutionary time, it approaches to a power law tail. The dynamics is confirmed by computer simulations. By allowing genome duplications, our model and dynamics (describing the dynamic behavior of gene networks) are more realistic than other previous ones (containing only static behavior).

1 Introduction

Systems biology on genetic networks is an important area in bioinformatics. Its primary goal is to understand biological organisms at a system-level [1–3], because many cellular functions can only be understood by simultaneously studying a group of genes, proteins, and other biological components that interact each other. Various biological networks have been extensively studied (see [4] for a general review), including cellular functional links, regulatory pathways, signal transductions, protein interactions, and metabolic correlations [5–10].

Mathematically, the growth of a large complex network can be governed by two mechanisms [4, 11]: (1) the network grows continuously by adding new vertices and edges, and (2) new vertices connect preferentially to those already well connected. Under these assumptions, it has been shown that the degree distribution of every vertex in the network follows a power law [4, 12]; the power law exponent only depends on the network growth process, not the initial state of the network (e.g. mechanisms decide dynamics) [13]. It is also possible that the network diameter is small and the cluster coefficients of nodes are large [14–17].

* Corresponding author

L. Wang (Ed.): COCOON 2005, LNCS 3595, pp. 12–21, 2005.

The duplications of both genes and genomes are the major resources for gene network growth and functional divergence [18]. Although several gene network models considered gene/genome duplications [6, 13, 19, 20], the mechanisms people introduced are somewhat artificial, due to mathematical convenience rather than biological reality [4, 10, 12, 13, 15, 16, 19–21]. Therefore, it is very interesting and important to develop gene networks allowing duplications for both genes and genomes under more biological realistic frameworks. This motivates our study.

In this paper, a gene network is formulated as a dynamical random graph whose vertices denote genes and whose edges represent gene-gene interactions among the genes. During evolution, the network grows through three biological mechanisms: (1) gene duplication and loss, (2) edge (i.e., gene-gene interaction) adding and removing, and (3) genome duplication. Moreover, preferential attachment is applied, i.e. a gene network is scale-free.

Based on these mechanisms, we study the network dynamics, focusing on the vertex degree distribution (i.e., the distribution of the gene-gene interactions per gene, which is a local statistical property in the netwotk) and the network degree distribution (i.e., the distribution of the gene-gene interactions in the network, which is a global statistical property in the netwotk). We show that the vertex degree distribution always follows power laws with different exponents, which can be changed by genome duplications. A genome duplication drives the network degree distribution away from its power law state. However, after a sufficient long time, the network degree distribution will somehow gradually recover to its power law state. The dynamics is verified by computer simulations. The genome duplication we allow makes our model and dynamics more general and realistic than the previous ones [2, 4, 13], which contain static behavior, whereas ours describe the dynamic behavior of gene networks. Our model and simulation also provide an efficient way to explore large scale complex biological networks.

2 Mathematical Model of Gene Network

A gene network is defined as a dynamical random graph G whose vertices represent genes and edges stand for gene-gene interactions. The network grows through three mechanisms: (1) gene duplication and loss; (2) gene-gene interaction adding and removing; and (3) genome duplication. A set of rates control gene duplication, gene loss, edge adding, edge removing, and genome duplication. Denote the graph

$$G = (V, E; g_d, g_l; \lambda, \mu; G_d, G_l).$$

The terms of the gene network G are defined as follows:

(1) V is the set of all vertices which represent genes (other biological components can be included, but here we restrict the discussion on genes).

(2) E is the set of all edges which stand for the gene-gene interactions (or influences) among the genes. For any two genes, if there exists any gene-gene interaction between them, the two genes are joined by an edge.