Uncertain interactions affect degree distribution of biological networks

Andrei Todor, Alin Dobra, Tamer Kahveci
Department of Computer and Information Science and Engineering, University of Florida, Gainesville FL 32611, USA
Email: {atodor,adobra,tamer}@cise.ufl.edu

***Abstract*—Biological interactions are often uncertain events, that may or may not take place under different scenarios. Existing studies analyze the degree distribution of biological networks by assuming that all the given interactions take place under all circumstances. This strong and often incorrect assumption can have misleading results. Here, we address this problem and develop sound mathematical basis to analyze degree distribution of biological networks in the presence of uncertain interactions. We present a comparative study of node degree distributions in two types of biological networks: the classical deterministic networks and the more flexible probabilistic networks. We extend this comparison to joint degree distributions of nodes connected by edges. The number of possible network topologies grows exponentially with the number of uncertain interactions. However, the mathematical apparatus we develop allows us to compute these degree distributions quickly even for entire protein protein interaction networks. It also helps us find an adequate mathematical model using maximum likelihood estimation. [1] Our results confirm that power law and log-normal models best describe degree distributions for both probabilistic and deterministic networks. Moreover, the inverse correlation of degrees of neighboring nodes shows that, in probabilistic networks, nodes with large number of interactions prefer to interact with those with small number of interactions more frequently than expected.**

I. Introduction

Biological networks, such as protein-protein interaction, metabolic and gene regulatory networks, are essential in describing the complex mechanisms by which cells carry out numerous functions. Studying those networks has been very effective in tackling many problems such as understanding the genetic factors that impact various diseases [11], [9], drug discovery [15], [19] and investigating the relationships among organisms and species [14], [8], [9].

Biological networks characterize the interactions between biological molecules within the cell, such as genes, proteins and enzymes. Like many processes in the biological realm, interactions are probabilistic events. An interaction may or may not happen with some probability, depending on a variety of factors such as the size, abundance or proximity of the interacting molecules [1]. Thus, we have less than 100% confidence in such interactions [4].

In the rest of this paper, if a biological network contains at least one probabilistic interaction, we call it a *probabilistic network*. Should all interactions be certain, we name it a *deterministic network*. We represent probabilistic networks using graphs with proteins as nodes and interactions as the edges. The weight on each edge is the probability of the corresponding interaction. An important observation is that a probabilistic network is actually a summary of all possible deterministic networks that are determined by the subset of interactions that take place. This means that a probabilistic network represented as a graph with $|E|$ edges will in fact describe the $2^{|E|}$ deterministic networks that could arise as instances of the probabilistic network, each with some probability. Interaction probability data are becoming increasingly available in popular biological databases, such as MINT [4] or STRING [17]. The probabilities of interactions are crucial in understanding biological networks. However, our survey revealed that they are often ignored in the computational analysis of biological networks. Certainly one deterring factor is that these networks introduce exponential number of alternative deterministic topologies.

The topological properties of biological networks, like degree distributions, have been extensively studied in the literature, noticing that they are scale free and assortative, and proposing various parametric degree distributions. We refer the reader to the extended version of this paper for a review of the literature [18].

Despite the existence of extensive studies of the degree distribution, all the existing methods, to the best of our knowledge, ignore the fact that interactions are probabilistic events. By assuming the networks to be deterministic, these methods implicitly enforce the topology of a specific instantiation of the probabilistic network over the alignment. This has a big risk of yielding biased and thus inaccurate distributions. Thus, accurate mathematical modeling of the network topology that incorporates interaction probabilities is needed. This task, however, is non-trivial, as the number of deterministic instances of the probabilistic network will lead to different network topologies exponential in the number of probabilistic interactions. For instance, the dataset we use in our experiments which is downloaded from MINT [4] contains networks with up to 21,909 probabilistic interactions. Even a medium sized network with 1,000 probabilistic interactions leads to 2^{1000} (i.e., more than 10^{300}) alternative topologies. Studying so many topologies one by one is clearly impossible even for simple computational problems on biological networks. Approximating the degree distribution by sampling graphs from probabilistic interactions is also infeasible. This is

[1]This work was supported partially by NSF under grants CCF-0829867 and IIS-0845439.

978-1-4673-2560-8/12/$31.00 ©2012 IEEE

because it is impossible to generate even a tiny fraction (such as 0.1%) of the total number of network topologies for large networks.

Analytical solutions in the context of *random graphs* have been studied in various contexts under many simplifying assumptions. Erdős and Rényi modeled a random graph using a random process that generates each edge with a given fixed probability [6]. More recent studies have introduced other random graph models that produce a given distribution of node degrees from a small set, such as Poisson, power law and exponential [13], [12]. These studies also use probability generating functions to represent the degree distribution of the whole graph.

It is important to emphasize that the probabilistic networks considered in this paper differ greatly from the random graph models analyzed in the studies discussed above. This is because in our model each edge has its own probability of being present in the graph, rather than having one probability for all edges or a small number of parameters that govern the structure of the whole graph. The research on random graphs has focused primarily on the asymptotic behavior of graphs obtained through these types of random processes, when the number of nodes tends to infinity. This is a consequence of the limited descriptive power of these models. They can provide bounds on certain characteristics, such as vertex degree or graph connectivity, but they cannot precisely characterize the degree distributions, particularly when the interaction probabilities are not all identical.

Contributions of this paper.

We focus on the problem of computing the degree distribution in probabilistic networks. Particularly, we focus on protein-protein interaction (PPI) networks in our experiments. However, all of our technical discussion in this paper applies to other types of biological networks. To the best of our knowledge, this is the first work that allows probabilistic interactions, which are inherent in many biological networks including PPI networks. As we explained above, probabilistic interactions increase the number of alternative network topologies exponentially. In order to tackle this challenge, we compute the expected degree distribution that is derived from all possible deterministic networks induced by the given probabilistic network. We employ a mathematical model that can precisely compute the distribution of all degrees in polynomial time and space. Our algorithm derives its efficiency from performing computations on a polynomial representation of the interactions, called *Probability Generating Function*, rather than directly on probability distributions. We fit six different models that are often used in the literature to the resulting distributions using maximum likelihood. Our results on the PPI networks from MINT demonstrate that power law and log-normal distributions fit the best to the degree distributions of the PPI networks. However, we observe that the model parameters for the probabilistic networks deviate significantly from those when the probabilities are ignored, as it was done by Stumpf et al. [16]. Furthermore, we observed differences in the ranking of the remaining models according to their likelihood using probabilistic as opposed to deterministic networks.

We expand the probabilistic network topology analysis by considering how the node degrees are correlated with the degrees of neighboring nodes. We do this by computing joint degree distributions of nodes connected by edges. In our experiments, we show that degrees of neighboring nodes in biological networks exhibit inverse correlation in the case of probabilistic networks, as reported previously for deterministic networks [10], [3], [12].

Finally, our method scales to large probabilistic networks easily. Most of our experiments took only a few minutes to complete. In summary, the technical contributions of this paper are:

1) We develop the mathematical theory that allows us to model the degree distributions in probabilistic networks
2) We extend this theory to joint distributions and demonstrate the inverse correlation of node degrees for graph neighbors

The rest of the paper is organized as follows. In Section 2, we describe our method in detail. In Section 3, we present our experimental results. Finally, we conclude the paper with a summary in Section 4.

II. Methods

In this section we develop the theoretical models that will allow us to analyze the degree distribution in probabilistic biological networks. First we introduce the concept of *random histogram* (Section II-A). Then we present the probability generating functions, the mathematical tool that underlies our method (Section II-B). Finally, we present a principled approach to maximum likelihood estimation (MLE) for degree distributions, both for individual nodes and joint distributions of pairs of nodes connected by edges (Sections II-C and II-D).

A. Random histograms

In order to approach the problem of degree distribution in probabilistic graphs, we first define this term mathematically.

Definition 1. *A probabilistic network is a triple $\mathcal{G} = (V, E, P)$ where V is a set of nodes, E is a relation on V and P is a function defined on the elements of E, such that for each edge $e \in E$, $P(e)$ is the probability of e being present in the network.*

We consider only undirected graphs, that is, if $(u, v) \in E$, then $(v, u) \in E$. It is, however, trivial to extend the discussion in this paper to directed networks.

Next, we introduce an extension to the classical histogram that is suited for probabilistic data, the *random histogram*

Definition 2. *A random histogram is a histogram in which the value for each bin is a random variable*

Let $\mathcal{G} = (V, E, P)$ be a probabilistic network, N_u the degree of node $u \in V$ and $M = \max\{N_u\}_{u \in V}$. We define the random histogram of the degrees of the nodes in V:

$\mathbf{H} = [H_0, H_1, , ..., H_M]$ where H_i is the number of nodes with i neighbors

We compute statistics of a random histogram defining indicator variables X_{ui} such that X_{ui} is 1 if $N_u = i$ and 0 otherwise. Then $H_i = \sum_{u \in V} X_{ui}$ and the expected value of H_i is $E[H_i] = \sum_{u \in V} E[X_{ui}] = \sum_{u \in V} P(N_u = i)$

Similarly, we define a bidimensional random histogram $\mathbf{B}$ of degrees of neighboring nodes, where B_{ij} is the number of edges such that the endpoints have degrees i and j, respectively. The expected value of an entry in this histogram is

$$E[B_{ij}] = \sum_{(u,v) \in E} P(N_u = i, N_v = j)$$

B. Degree distributions in probabilistic graphs

We denote the set of edges incident on node u by $\mathcal{N}_u$. Let Y_{uv} be an indicator random variable that takes the value 1 if edge (u, v) is present in the graph and 0 otherwise. We have: $N_u = \sum_{(u,v) \in \mathcal{N}_u} Y_{uv}$. In other words, N_u is the sum of Bernoulli random variables, each with a different probability of success. This type of random variable is called Poisson binomial [5]. Next, we define a special function called the *probability generating function (PGF)* and we show how this distribution can be easily computed using this function.

Definition 3 (Probability Generating Function [7])**.** *The probability generating function of a discrete random variable X is defined as the function in z $Q_X(z) = E[z^X] = \sum_{k=0}^{N} P(X = k) z^k$*

The following theorem shows that the distribution of a sum of random variables can be easily computed from the product of the PGFs of the components.

Theorem 1. *[7] Let $X_1, \ldots, X_N$ be independent random variables with corresponding PGFs $Q_1(z), \ldots, Q_N(z)$, respectively. The PGF of $X = \sum_{i=1}^{N} X_i$ is $Q(z) = \prod_{i=1}^{N} Q_i(z)$*

We apply Theorem 1 to the Poisson binomial distribution of N_u. Let $p_v = P(Y_{uv} = 1)$. Then $Q_v(z) = (1 - p_v) + p_v z$ and $Q(z) = \prod_{v \in \mathcal{N}_u} (1 - p_v + p_v z)$. The values of the distribution can be easily extracted as the coefficients of the PGF polynomial. More specifically, for any non-negative integer k. the coefficient of z^k in the PGF of N_u is equal to the probability that the degree of node u is k.

To compute $E[\mathbf{B}]$ we extend the PGF to the bidimensional case.

Definition 4. *The Probability Generating Function (PGF) of a discrete random vector (X, Y) is defined as the function in z_1 and z_2*

$$Q_{XY}(z_1, z_2) = E[z_1^X z_2^Y] = \sum_{i=0}^{M} \sum_{j=0}^{N} P(X = i, Y = j) z_1^i z_2^j$$

Next, we present a result that allow us to compute the PGF of joint degree distributions in probabilistic networks. The proof is given in [18].

Theorem 2. *Let $G = (V, E, P)$ be a probabilistic graph and $e = (u, v) \in E$ an edge in this graph. Let Q_u and Q_v be the PGFs of N_u and N_v, respectively, and $Q_e(z) = 1 - P(e) + P(e)z$. The PGF of the random vector (N_u, N_v) is*

$$Q_{uv}(z_1, z_2) = \frac{Q_u(z_1)}{Q_e(z_1)} \frac{Q_v(z_2)}{Q_e(z_2)} Q_e(z_1 z_2)$$

C. MLE for degree distribution

The PGFs computed for each node of the probabilistic network give us a very detailed and precise model of the node degree distribution. It is customary that the detailed distribution representation be approximated by one or several probabilistic models with a reduced number of parameters (one or two). One standard method for inferring the model parameters from the exact representation is maximum likelihood estimation (MLE) [2].

MLE requires independent and identically distributed samples from the probabilistic model we are trying to fit. Therefore, for obtaining N samples, we consider the sampling process consisting of the following steps repeated N times: 1) select a deterministic graph from the probabilistic graph model; 2) select a node $v_i \in V$ uniformly at random; 3) observe the degree of v_i.

The optimal model parameters for this process are given by

$$\hat{\boldsymbol{\theta}} = arg \max_{\theta} \sum_{k=0}^{M} \frac{E[H_k]}{|V|} \ln p(k|\boldsymbol{\theta})$$

For the deterministic case, the first step of the sampling process is vacuous; the same graph is selected each time. At the same time, H_k is a deterministic value, so that $E[H_k] = H_k$. Therefore, our deterministic solution is congruent with previous work [16].

D. MLE for joint degree distribution in probabilistic graphs

We follow the same steps as for the univariate case to derive an MLE solution for the joint distribution of the degrees of any two nodes connected by edges. This is justified by the following random process: 1) select a deterministic graph from the probabilistic graph model; 2) select an edge $e_i = (u_i, v_i)$; 3) observe the degrees of u_i and v_i.

The optimal parameters are given by

$$\hat{\boldsymbol{\theta}} = arg \max_{\theta} \sum_{k=0}^{M} \sum_{l=0}^{M} \frac{E[B_{kl}]}{|E|} \ln p(k, l|\boldsymbol{\theta})$$

For the joint distribution we fit a bivariate log-normal model $p(k, l|\boldsymbol{\theta})$ [20]. We consider a bivariate distribution with equal means and equal variances of the marginal log-normal distributions, $\mu_1 = \mu_2 = \mu, \sigma_1 = \sigma_2 = \sigma$. The third parameter, ρ, is the correlation between the two random variables of the joint distribution. Let us define the variables y_1 and y_2 as $y_1 = \ln x_1 - \mu$ and $y_2 = \ln x_2 - \mu$. The bivariate log-normal PDF is given by

$$p(x_1, x_2|\mu, \sigma, \rho) = \frac{1}{2\pi x_1 x_2 \sigma^2 \sqrt{1 - \rho^2}} e^{-\frac{1}{2\sigma^2(1-\rho^2)}[y_1^2 - 2\rho y_1 y_2 + y_2^2]}$$

TABLE I

NEGATIVE LOG-LIKELIHOODS OF SIX DIFFERENT MODELS FOR THE DEGREE DISTRIBUTION OF THE PPI NETWORKS OF FIVE ORGANISMS. THE THREE LETTER ORGANISM CODE DENOTES THE FOLLOWING: *dme = D. melanogaster, sce = S. cerevisiae, cel = C. elegans, hpy=H. pylori, eco = E. coli* THE MODELS ARE M_1 = POISSON, M_2 = EXPONENTIAL, M_3 = GAMMA, M_4 = POWER LAW, M_5 = LOG-NORMAL, M_6 = STRETCHED EXPONENTIAL. THE RESULTS ARE REPORTED FOR TWO CASES OF EACH NETWORK, DETERMINISTIC (D) AND PROBABILISTIC (P). THE NUMBERS IN **bold** SHOW THE LIKELIHOOD OF THE BEST MODEL FOR THE CORRESPONDING DATASET AND CASE.

Org. Code	Case	Model					
		M_1	M_2	M_3	M_4	M_5	M_6
dme	D	42039	20340	20264	19054	**18971**	20105
	P	17345	11151	8741	**7971**	8046	8873
sce	D	39798	17013	16962	17464	**16348**	16893
	P	14997	9645	7822	7507	**7289**	7915
cel	D	12612	7073	7013	**5552**	6057	6960
	P	5407	2999	2424	**1871**	2031	2526
hpy	D	2695	1756	1736	1633	**1616**	1741
	P	1251	831	646	**580**	582	674
eco	D	3723	1771	1738	**1410**	1555	1710
	P	1141	627	488	**419**	440	503

III. RESULTS

IMPLEMENTATION DETAILS: We implemented our methods in C++ and R. Computing the degree distribution in deterministic networks is a special case of the probabilistic network method. We set all the interaction probabilities to 1.0 to obtain results for the deterministic networks. We implemented our algorithm to fit the same set of six distribution models to the degree distributions as those used by Stumpf *et al.* [16]. These are Poisson, exponential, gamma, power law, log-normal and stretched exponential distribution. We ran our experiments on a standard desktop computer with 4 GB of RAM and a 2 GHz processor, except for the computation of the joint distributions on probabilistic networks, which was performed on a multiprocessor machine with 256 GB RAM.

DATASET: Our primary data source was the MINT database [4], which contains protein-protein interactions in 30 organisms along with interaction probabilities. We used all the organisms from this dataset. Due to space limitation, we only present the results for five of them; these are the same organisms studied by Stumpf *et al.* for the degree distributions of their networks [16].

A. Evaluation of the degree distribution

In this section, we evaluate the degree distributions of the PPI networks of five organisms . We demonstrate that probabilistic interactions yield significantly different degree distributions than deterministic ones and show how various known distribution models explain the probabilistic PPI of each organism.

1) Agreement between probabilistic and deterministic networks: The first question we need to answer is: Do the degree distributions of probabilistic networks agree with those of deterministic ones? This experiment seeks the answer to this question as follows. For each organism and each of the six distribution models (i.e., totally $5 \times 6 = 30$ combinations), we compute the model parameters that maximize the likelihood of that network. We do this for both the deterministic and probabilistic versions of the data. We then calculate the likelihood of the network using these parameters. Table I lists the negative log-likelihood of these fittings.

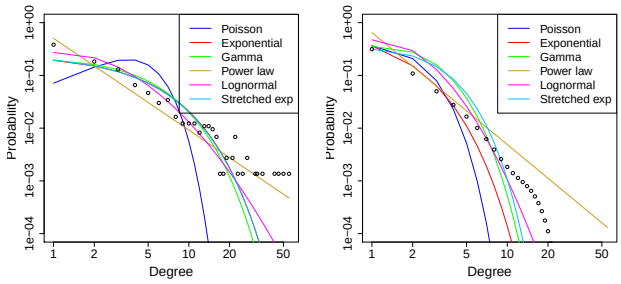

Fig. 1. **Left:** Degree distribution of the deterministic PPI network of *H. pylori* taken from MINT (shown with small circles) and the fitting distributions for six different models. **Right:** The same results obtained by taking the interaction probabilities into account.

Our results suggest several important points. First, the degree distribution is closer to the power law distribution than to log-normal distribution in most of the datasets. More specifically, log-normal distribution fits the best in majority of the deterministic networks (three out of five) while power law distribution fits best in most of the probabilistic ones. Second, power law and log-normal distributions are the top two distributions with the highest likelihood in all datasets with or without probabilistic interactions. Third, regardless of the distribution model, considering probabilistic interactions always yields higher likelihood than the deterministic ones. This is because they produce significantly smoother degree distribution which can be described with these parametric models. We elaborate on this further in the next section.

2) Degree distribution of probabilistic networks: So far, we have demonstrated that the degree distribution of the probabilistic networks deviate from that of deterministic ones significantly. In this section, we take a closer look at the actual degree distributions and show how well each distribution model fits to the actual degree distribution for all the organisms under consideration .

Figure 1 plots the results for both the deterministic and probabilistic PPI networks of *H. pylori*. The results for the deterministic network is very similar to those observed by Stumpf et al. [16]. However, the figure clearly shows that the degree distributions of probabilistic and deterministic networks are significantly different from each other. More specifically, the deterministic network has significantly more nodes with large degrees than what would be expected from the probabilistic network. This is because the deterministic network is nothing but a specific instance of the probabilistic network that assumes all interaction probabilities are equal to one. *Thus, its degree distribution is biased towards large degrees.*

Log-normal and the power law distributions are the top two distributions with the highest likelihood for both networks. However, deterministic (probabilistic) network slightly favors

the log-normal (power law) distribution (see Table I). Interestingly, the degree distribution of the probabilistic network has a knee around degree = 10. More evidence for this is presented in the long version of this paper [18]. As a result, it resembles a mixture of distributions rather than a single distribution. This is because of the dependency of the degrees of the nodes; the existence/absence of a probabilistic interaction affects the degrees of the two nodes that is it adjacent to. We will look into this property in detail in Section III-B. The results on the other four organisms are similar.

This suggests that it worths exploring mixtures of distributions to explain the degree distributions of probabilistic networks as well as joint degree distributions to capture the dependency between the degrees of the interacting nodes.

B. Evaluation of the joint degree distribution

In this section, we analyze the joint degree distribution of the nodes that are connected through an edge. Our results in this section demonstrate that the degree distribution alone is insufficient to describe the dependency of the degrees among the connected nodes. In our experiments, we use the bivariate log-normal distribution [20] to fit the actual distribution. This is because, in Section III-A, we observed that the log-normal distribution has the highest likelihood or has a likelihood that is very close to that of the best fitting model.

Table II lists the parameters of the distribution fitted to the degrees of the adjacent node pairs. Recall that a negative correlation indicates that nodes with high degrees tend to be connected to nodes with low degrees. Thus it shows an anti-correlation. As the value of the correlation approaches to zero, it indicates that the nodes are connected to each other independent of their degrees. We observe that all the networks have a negative correlation. The probabilistic networks produce correlation with slightly smaller magnitude. Among all the networks, *C. elegans*, *H. pylori* and *S. cerevisiae* have correlation with the largest magnitude. These are the organisms with the large knee in the degree distributions. *D. melanogaster* and *E. coli* have the correlation with the lowest magnitude. These results suggest that the degrees of neighboring nodes are not independent. Proteins that have many interactions prefer not to interact with other proteins of the same kind. They rather tend to interact with very specific proteins that can only make a very limited number of interactions.

IV. Conclusion

In this work we introduced a new model for biological networks, the probabilistic network. We used it to compute degree distributions and compare the results with the deterministic solution. We fit probabilistic models to the empirical data via MLE. We performed these experiments both for single node degree distribution and joint degree distribution for nodes connected by edges, finding that the degrees of neighboring nodes are inversely correlated. In future work we plan to assess the robustness of probabilistic biological networks to errors in probability values.

TABLE II
CORRELATION, MEAN AND VARIANCE PARAMETERS OF THE BIVARIATE LOG-NORMAL DISTRIBUTION FITTED TO THE DEGREES OF ADJACENT NODE PAIRS FOR THE PPI NETWORKS OF FIVE ORGANISMS. THE NETWORKS ARE OBTAINED FROM MINT. THE RESULTS ARE REPORTED FOR TWO CASES FOR EACH NETWORK: D = DETERMINISTIC NETWORK IS OBTAINED BY IGNORING INTERACTION PROBABILITIES. P = PROBABILISTIC NETWORK IS OBTAINED BY USING THE INTERACTION PROBABILITIES PROVIDED BY MINT.

Organism	Case	correlation	mean	variance
D. melanogaster	D	-0.024	2.502	1.335
	P	-0.021	1.495	0.902
S. cerevisiae	D	-0.097	2.839	1.205
	P	-0.109	1.589	0.963
C. elegans	D	-0.310	1.979	2.054
	P	-0.231	1.197	1.189
H. pylori	D	-0.269	1.956	1.160
	P	-0.205	1.023	0.703
E. coli	D	0.106	2.385	1.647
	P	-0.085	1.273	0.700

References

[1] J. S. Bader, A. Chaudhuri, et al. Gaining confidence in high-throughput protein interaction networks. *Nature Biotechnology*, 2003.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] D. S. Callaway, J. E. Hopcroft, et al. Are randomly grown graphs really random? *Phys. Rev.*, 2001.

[4] A. Chatraryamontri, A. Ceol, et al. MINT: the Molecular INTeraction database. *Nucleic Acids Research*, 2007.

[5] H. Cramer. *Mathematical Methods of Statistics*. Princeton University Press, 1946.

[6] P. Erdős and A. Rényi. On Random Graphs I. *Publicationes Mathematicae*, 6:290297, 1959.

[7] M. A. Goldberg. *An Introduction to Probability Theory with Statistical Applications*. Plenum Press, 1984.

[8] M. Green and P. Karp. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinformatics*, 2004.

[9] Y. Hu, I. Flockhart, et al. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics*, 2011.

[10] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Phys. Rev.*, 2001.

[11] D.-S. Lee, J. Park, et al. The implications of human metabolic network topology for disease comorbidity. *PNAS*, 2008.

[12] M. E. J Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 2002.

[13] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev.*, 2001.

[14] H. Ogata, W. Fujibuchi, et al. A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. *Nucleic Acids Research*, 2000.

[15] P. Sridhar, T. Kahveci, and S Ranka. An iterative algorithm for metabolic network-based drug target identification. *PSB*, 2007.

[16] M. P. H. Stumpf and P. J. Ingram. Probability models for degree distributions of protein interactions networks. *Europhysics Letters*, 2005.

[17] D. Szklarczyk, A. Franceschini, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research*, 2011.

[18] A. Todor, A. Dobra, and T. Kahveci. Uncertain interactions affect degree distribution of biological networks. *University of Florida CISE Departmental Report REP2012-549*, 2012.

[19] N. Watanabe, M. M. Cherney, et al. Crystal structure of LL-diaminopimelate aminotransferase from Arabidopsis thaliana: a recently discovered enzyme in the biosynthesis of L-Lysine by plants and Chlamydia. *Journal of Molecular Biology*, 2007.

[20] S. Yerel and A. Konuk. Bivariate lognormal distribution model of cutoff grade impurities: A case study of magnesite ore deposits. *Scientific Research and Essay*, 2009.