Aligning protein-protein interaction networks using random neural networks

Hang T.T. Phan, Michael J.E. Sternberg
Division of Molecular Biosciences
Faculty of Natural Sciences
Imperial College London
London SW7 2AZ, United Kingdom
Email: hang.phan07, m.sternberg@imperial.ac.uk

Erol Gelenbe
Intelligent Systems & Networks
Department of Electrical & Electronic Engineering
Imperial College London
London SW7 2AZ, United Kingdom
Email: e.gelenbe@imperial.ac.uk

***Abstract*—We have developed RNNI, a global alignment method for protein-protein interaction networks between species, using a random neural network model (RNN) tailored for the alignment problem. The benchmark of the method in comparison with other available alignment approaches was performed using a range of measurements. The alignment results of the human and yeast pair showed that RNNI is capable of generating alignments with large conserved networks with functionally-related protein pairs while maintaining the closeness to the naïve- sequence homology approach (BLAST).**

***Keywords*-protein interaction network alignment, random neural network**

I. Introduction

Network comparison, particularly network alignment, can identify conserved functional modules, protein complexes or pathways, predict protein functions or and provide insights to evolutionary processes. In a network alignment problem, the aim is to establish the connections between proteins in one species with proteins in one or more other species that reflects a related biological function or a sequence homology. Aligned proteins are expected often to be homologous or functionally orthologous. While local alignment methods try to map proteins locally, global alignment methods aim to establish a global mapping of proteins between two or more PPINs. Græmlin 1.0 [1] and Græmlin 2.0 [2] include evolutionary events such as gene duplications and deletions in the scoring function of the alignment and progressively clusters proteins in the direction that improves the alignment scoring function. The GRAAL family approaches including GRAAL [3], H-GRAAL [4] and MI-GRAAL [5] rely more on aligning the PPINs based on network topology, particularly the graphlet degree vector. Other information such as sequence and functional information can also be used to enforce the alignments. IsoRank [6] and IsoRankN [7] employ Markov random field approaches in which the scores of proteins across species including the neighbourhood similarities are calculated using eigen-value based approaches. The domain-based aligner DOMAIN [8] converts PPINs in to interaction networks of domains and then map interactions in these networks using the domain similarity. CAPPI [9] reconstructs the ancestor's PPINs by estimating the probabilities of ancestral interactions using evolutionary model and then establishing the mappings of proteins of species through the ancestral network. More recently, the global alignment method PINALOG [10] has been introduced to align PPINs using a seed-and-extend approach where the seeds are found by similar community mapping. In this method, protein mapping in the extension step implicitly includes the network similarity via extra scores given to protein pairs that have interactions with one or more seed protein pairs.

In this paper, we introduce an approach to align PPINs based on the random neural network (RNN) which is a stochastic network model first introduced in [11]. RNNs have been used in image processing [12], [13], pattern recognition [14], [15], learning [16] and network routing [17]. Similar models [18] have been used in systems biology to infer a gene regulatory network from multiple data sources (gene expression, molecular interactions and gene ontology)[19], [20]. Other recurrent networks have been used for secondary structure prediction [21], [22], [23] and protein contact map prediction [24]. This is the first time the RNN model or any other recurrent network, is applied to PPIN alignment, a problem which is related to combinatorial and network optimisation [25], [26], [27]. The method we developed, called $RNNI$, is applied here to obtain the alignment of human and yeast PPINs and other pairs of species. The comparison of RNNI alignments with other approaches using different assessment measures showed that RNNI produced alignments with better functional coherence between aligned proteins and with higher number of conserved interactions than other alignment methods such as IsoRank, MI-GRAAL and PINALOG. More importantly, the RNNI alignments have significant overlap with those obtained by reciprocal best BLAST approach. It is noted that the BLAST approach is limited to areas where there is sequence similarity. However, RNNI, in addition to covering most of the BLAST alignments, is able to reach beyond the sequence-based limit to identify the correspondence of proteins across species.

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

II. Methods

A. RNN model

The RNN [28] is an instance of G-Networks [29], [30] that represent neuronal activity with a discrete state-space, continuous time stochastic process, mediated by external inputs and spike exchanges among neurons (cells). The state of each cell i is described by its electrical potential $K_i(t)$ at time $t \geq 0$, a non-negative integer variable. If the cell is excited, $K_i(t) > 0$, it outputs spikes at rate r_i that excite another cell j with probability $p^+(i,j)$, inhibit it with probability $p^-(i,j)$, or depart the network without affecting other cells with probability d_i and $d_i + \sum_{j=1}^{N}[p^+(i,j) + p^-(i,j)] = 1$ for L neurons. Upon arrival of an excitatory spike, the receiving cell's potential increases by one, or decreases by one when it receives an inhibitory spike as long as $K_i(t) > 0$. If an inhibitory spike arrives to a "quiescent" cell with $K_i(t) = 0$ then it has no effect, while an arriving excitatory spike will increase the cell's potential by one. After sending a spike, the sending cell's potential decreases by one. A cell also receives external input as a Poisson flow of excitatory spikes at rate Λ_i, and inhibitory spikes as a Poisson flow of rate λ_i. The excitatory weights are the quantities $w_{ij}^+ \equiv r_i p^+(i,j)$ and the inhibitory weights are $w_{ij}^- \equiv r_i p^-(i,j)$ which are non-negative since they represent rates. Note that $r_i = \sum_{j=1}^{N}[w_{ij}^+ + w_{ij}^-]$ can be interpreted as the total outgoing firing rate of a cell, and r_i^{-1} is the average time between successive spikes sent by a cell which is excited. A useful result concerning this model is that the stationary probability distribution for the state of this class of networks has product-form solution [31]:

$$\lim_{t\to\infty} P(K_1(t), \ldots, K_N(t) = (k_1, \ldots, k_n)) = \prod_{i=1}^{N} q_i^{k_i}(1-q_i), \tag{1}$$

where $k_i \geq 0$ is the value taken by the state of each cell, and:

$$q_i = \min\left(\frac{\Lambda_i + \sum_j q_j w_{ji}^+}{r_i + \lambda_i + \sum_j q_j w_{ji}^-},\ 1\right). \tag{2}$$

where $q_i = lim_{t\to\infty} P(K(t)_i > 0)$. This result is used in this paper because it allows the simple algebraic computation of network state even though the network is recurrent and stochastic.

B. Applying the RNN to PPIN alignment

Unlike other neuronal models which are often used for learning and recognizing patterns, the RNN model in this network alignment framework does not involve parameter learning, because in PPIN network alignment, there is no standard alignment for parameter learning. The RNN here only involves the solution of a model that yields the probability of neural excitatory states, which can be found numerically as proven in [32]. First we describe the detailed alignment approach. Then the similarity metrics for estimating the similarities between proteins for network alignment are introduced.

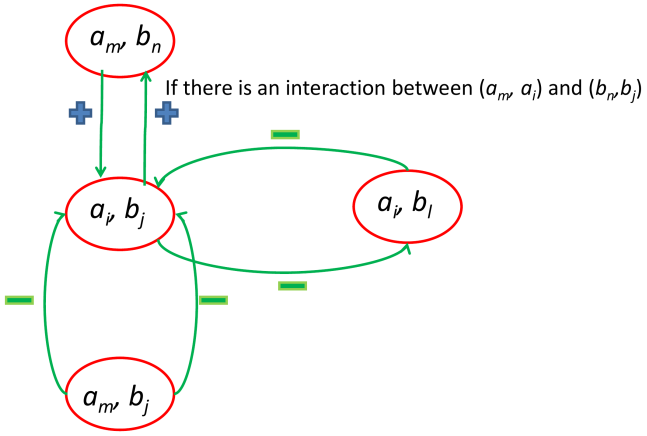

Figure 1. RNNI model for network alignment. Each neuron in the model represents the mapping of two proteins, one from each species. Because there is a competition between mapping a given protein in A with different proteins in B and vice versa, negative signals are exchanged between neurons that share a common protein. Meanwhile, as we are aligning PPINs where interactions between aligned proteins are expected to be conserved, there are positive signals between neuron pairs where interactions exist between proteins from the neuron pairs in each species.

Let A and B be two species with PPINs $G_A = V_A, I_A$ and $G_B = V_B, I_B$ where V_A and V_B are the set of proteins in G_A and G_B respectively. $V_A = \{a_1, a_2, \ldots, a_M\}$ and $V_B = \{b_1, b_2, \ldots, b_N\}$, and I_A and I_B are the adjacency matrices of G_A and G_B. $I_A(i,k) = 1$ if there is an interaction between a_i and a_k, and 0 otherwise, and similarly for $I_B(j,l)$. The similarity between proteins $a_i \in V_A$ and $b_j \in V_B$ is defined as

$$f_{ij} = s_1(a_i, b_j) \bigoplus s_2(a_i, b_j) \tag{3}$$

where $s_1(a_i, b_j)$ and $s_2(a_i, b_j)$ are the sequence and functional similarity between a_i and b_j respectively. The RNNI model for PPIN alignment (see Figure 1) is constructed as follow:

- Each protein pair (a_i, b_j) is encoded in one neuron (i,j) of the RNN.
- An inhibitory signal going from neuron $(i,k) = (a_i, b_k)$ to $(l,j) = (a_l, b_j)$ and vice-versa represents the competition between associating a_i with b_j, as compared to associating a_i with other proteins in G_B, and associating b_j with other proteins in G_A.
- An excitatory signal will link $(i,j) = (a_i, b_j)$ to $(m,n) = (a_m, b_n)$ and vice-versa, when $I_A(i,m) = 1$ and $I_B(j,n) = 1$, i.e. if there is an interaction between a_i and a_m and also between b_j and b_n. The excitatory signals between neuron pairs therefore allow the RNNI to model conserved interactions in the alignment.

As a consequence of RNN theory [28], for each protein pair (a_i, b_j), the probability q_{ij} that any neuron (i, j) is excited, is given by:

$$q_{ij} = \frac{f_{ij} + \beta \sum_{m,n=1}^{M,N} q_{mn} q_{ij} I_A(m,i) I_B(n,j)}{\mu_{ij} + \beta \sum_{m,n=1}^{M,N} I_A(m,i) I_B(n,j) + \alpha \sum_{m \neq i} q_{mj} + \alpha \sum_{l \neq j} q_{il}} \quad (4)$$

From these equations one computes the stationary probability of the state of all of the neurons in this RNN. The numerical solution of the equations (4) is then obtained by a fixed point iteration using the following algorithm that proceeds in discrete steps $t = 0, 1, 2, \ \ldots$ to compute successive values of q_{ij}^t for all (i, j):

1) First set $q_{ij}^0 = 0.5$.
2) Compute all the q_{ij}^{t+1} from all of the q_{ij}^t using (4)
3) If $\forall\ (i, j)$ we have $|q_{ij}^{t+1} - q_{ij}^t| \leq \epsilon$, e.g. for $\epsilon = 0.05$, then stop the iteration and set all the q_{ij} to the values $q_{ij} \equiv q_{ij}^{t+1}$.

The RNNI alignment algorithm then uses the above computational algorithm for two sets of proteins V_A and V_B:

1) First set the network interaction parameters or "weights" to $\alpha = \frac{2}{M'+N'}$ and $\beta = 1 \times 10^{-3}$, where M' and N' are the size of protein sets from two species A (V_A') and B (V_B') in each iteration respectively. V_A' is initially set to V_A and V_B' to V_B. Note that β relates to the excitation between alignments, while α relates to the inhibition between competing possibilities of alignment.
2) Then set $\mu_{ij} = \alpha \times (M' - 1 + N' - 1) = \frac{2 \times (M'+N'-2)}{M'+N'}$ so that the requirement from RNN theory [28] regarding the total outgoing neuron firing rates is satisfied. This will guarantee that the q_{ij} exist and are unique.
3) Calculate all the q_{ij} using the iteration as described above.
4) Select the protein pairs (a_i, b_j) that have the largest q_{ij}, and place them in the mapping list.
5) For any (i, j) selected in Step 4 above, set $q_{ij} = 1$ and all $q_{ik} = 0$ for $k \neq j$, and $q_{kj} = 0$ for $k \neq i$.
6) Repeat the algorithm after removing a_i and b_j from the sets of proteins considered in the next iteration of the algorithm, i.e. by using the new sets $V_A' \Leftarrow \{V_A' \setminus a_i\}$ and $V_B' \Leftarrow \{V_B' \setminus b_j\}$.
7) The algorithm stops when one of the two sets V_A', V_B', or both, become empty, or when the largest q_{ij} is 0.

C. Similarity measures of proteins

In this implementation of RNNI, we adopted the normalised BLAST score as the sequence similarity between two proteins as has been used in other approaches such as PINALOG and [33].

$$s_1(a_i, b_j) = \frac{S(a_i, b_j)}{\sqrt{S(a_i, a_i), S(b_j, b_j)}} \quad (5)$$

where $S(a_i, b_j)$ is the BLAST bit score value between protein a_i and b_j.

We used a similarity measure of protein function that is based on the semantic similarities of GO (Gene Ontology, [34]) terms annotated to the proteins. Many measures have been developed to estimate the functional similarity of proteins based on GO terms (see review in [35]) and the Schlicker's similarity [36] was chosen in this paper as it was shown to outperform other measures in [37].

III. Alignment results

We used the above algorithm to align PPINs of different species pairs including human and yeast. The PPIN data are from IntAct [39], and are the same as those used in PINALOG. RNNI was applied using two different similarity metrics for protein similarity: i) RNNI-1: using only sequence similarity calculated by BLAST scores, $f_{ij} = s_1(a_i, b_j)$, and ii) RNNI-A: using a combined function of sequence and functional similarity ($f_{ij} = \theta s_1(a_i, b_j) + (1 - \theta) s_2(a_i, b_j)$) with θ identified automatically based on the number of reciprocal best hits between two species (see the PINALOG paper, [10], for more details). The performance of RNNI-1 and RNNI-A was compared with IsoRank, MI-GRAAL, PINALOG-1, PINALOG-A and the naïve approach using reciprocal best BLAST hits. Analogous to RNNI-1 and RNNI-A, PINALOG-1 uses only sequence information while PINALOG-A uses a combination of sequence and functional similarity.

The assessment of the quality of PPIN alignments is notoriously difficult due to the lack of a gold-standard alignment and measurements. Therefore, we use a wide range of statistics to analyse the resulting alignments by these approaches to make comparison of the performances of alignment methods. The reported statistics include:

- NA - the number of aligned protein pairs.
- NC - the number of conserved interaction between two species.
- NH - the number of aligned proteins that belong to the same Homologene orthologous group [40].
- NF - the number of protein pairs whose functional similarity are larger than 0.5.
- NI - the number of conserved interactions that are interlogs [41].

NA and NC provide an overview of the size of an alignment. NH and NI are useful in examining whether the alignment of two networks manages to align readily known homologous proteins or not. Homologene [40] is a homologue detection program for eukaryotic genomes using BLAST, phylogenetic trees of species and genomic context. It has been used in IsoRank to assess the quality of aligned proteins. An

Table I
ALIGNMENT RESULTS OF HUMAN-YEAST (H-Y) AND HUMAN-FLY (H-F) ALIGNMENTS USING R1 (RNNI-1), RA (RNNI-A), P1 (PINALOG-1), PA (PINALOG-A), IS (ISORANK) MG (MI-GRAAL) AND BL (BLAST).

Method	R1	RA	P1	PA	IS	MG	BL
NA	4,581	5,268	3,948	5,222	5,674	5,674	1,818
NC	3,680	1,254	3,388	3,319	717	4,107	530
NH	798	798	497	454	165	0	818
NI	423	432	471	460	136	0	465
NF	1,807	3,884	1,333	3,139	734	146	1347

interlog is defined as a conserved interaction between two aligned protein pairs (a_i, b_j) and (a_m, b_n) where $I_A(i, m) = I_B(j, n) = 1$ and the BLAST E-values between a_i and b_j, and between a_m and b_n are less than 10^{-10} [41].

Due to the difficulty in assessing the alignments for the biological relevance of aligned protein pairs, beside the closeness in term of sequence, the functional similarity of the protein pairs is considered. One of the measure we adopted for assessment is the functional similarity measure summarised by NF, the number of protein pairs with functional similarity larger than 0.5. Schlicker *et al.* [42] suggested that two proteins are functionally similar if their Schlicker's similarity score is above 0.5.

A. Human-yeast alignments

Overall, RNNI-1 has a good balance of all the alignment statistics (Table I) with a high number of conserved interactions and Homologene pairs, and a mid-range number of functionally similar protein pairs. RNNI-A, by including functional information in aligning proteins, produced an alignment with twice the number of protein pairs with high functional similarity compared with RNNI-1 (3,884 vs. 1,807). However, the high similarity in term of function of RNNI-A's alignment is not without a lower score in a different metric- the number of conserved interactions (1,254 in RNNI-A vs. 3,680 in RNNI-1). This indicates the dominance of functional information over the interaction information in guiding the alignment process in RNNI-A.

In comparison with other alignment methods, almost all alignment methods obtained a similar number of mappings between two species, with the exception of BLAST (1,818 protein pairs vs. 4,000 to 5,000 by other methods). The alignments by RNNI-1, PINALOG-1, PINALOG-A and MI-GRAAL are the ones with higher number of conserved interactions, from 3 to 6 fold more than those found in the alignments of RNNI-A, IsoRank and BLAST. The largest connected component in the conserved network obtained by RNNI-1 has 2,076 proteins with 2,994 interactions among them and has a higher clustering coefficient (0.070) and network density (0.0014) compared to that of MI-GRAAL (which has a higher number of proteins - 3,773 and interactions - 3,789, but with the clustering coefficient 0.001, and network density 0.0005). Here the two topological properties of networks, clustering coefficient and network density, were calculated using the NetworkAnalyzer plug-in [43] of Cytoscape [44]. to estimate the PINALOG-A has 1,858 proteins connected by 2,774 in its largest conserved connected component (clustering coefficient 0.091, network density 0.002). That of RNNI-A only has 155 proteins linked by 287 interactions, however it has higher clustering coefficient (0.173) and network density (0.024) compared to other methods.

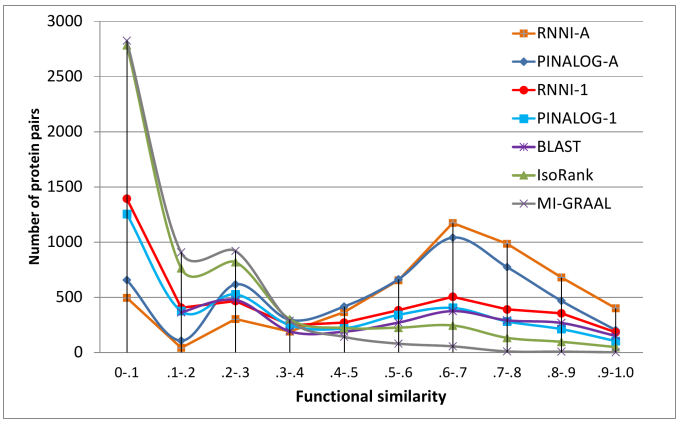

Figure 2. Distribution of functional similarities of aligned protein pairs by different alignment methods. The order of alignment methods in the legend from top to bottom matches with the order of the distributions at the local peaks 0.6-0.7 range.

In terms of sequence similarity indicated by the number of aligned Homologene pairs and interlogs, the performance of the RNN methods (RNNI-1 and RNNI-A) are good, being comparable with those of BLAST (798 and 798 vs. 818 Homologene pairs and 423 and 432 vs. 465 interlogs). PINALOG methods belong to the average performance group with the number of Homologene pairs being 40% less compared to BLAST, although PINALOG methods obtained higher numbers of interlogs.

Figure 2 plots the distributions of functional similarities between aligned protein pairs obtained by different alignment methods. RNNI-A is slightly better than PINALOG-A with more protein pairs with higher similarity. The Kolmogorov-Smirnov test (KS test, [45]) of the two distributions suggested that aligned protein pairs in RNNI-A are significantly more functionally similar than those in PINALOG-A (p-value $< 2.2 \times 10^{-16}$). Similarly, RNNI-1 has more functionally similar protein pairs than PINALOG-1 (KS test p-value = 3.016×10^{-8}), IsoRank (KS test p-value $< 2.2 \times 10^{-16}$) and MI-GRAAL (KS test p-value $< 2.2 \times 10^{-16}$).

In summary, for the human-yeast alignments, RNNI-1 obtained a balanced alignment in all three categories of assessment measures of conserved network, sequence similarity and functional similarity. RNNI-A, on the other hand, although outperformed in functional similarity assessments,

produced an alignment with low conservation in terms of interactions. RNNI alignments are closest to the traditional sequence-based approach BLAST while extending the alignment of BLAST as guided by additional interaction information and functional information.

In comparison with other methods, for the human-yeast alignments, the overlaps of RNNI alignments with BLAST are the highest (83% for RNNI-1 and 79% for RNNI-A). PINALOG-1 and PINALOG-A are in the middle range where the overlaps with BLAST are only 48% and 41% of the BLAST mappings. IsoRank has a lower overlapping rate, with only 15% and MI-GRAAL has no overlap at all with BLAST result. The coverage of BLAST alignment by PINALOG-1 and PINALOG-A in this pair of species is lower than in the human-fly alignments, with only 29% and 26%, although it is twice that of IsoRank (13%).

IV. Discussion

Assessment of the quality of alignments between networks is not straightforward due to the lack of a gold-standard alignment. Therefore several metrics of resulting alignments have been adopted to quantify the similarity of the networks being aligned. However, in the future, it might be beneficial for the PPIN alignment community to have a set of simulated networks from actual interaction data with proper evolutionary models for sequence mutations, gene duplications, deletions and interaction rewiring. Such a set of networks would make bench-marking less problematic and more reliable.

Comparisons of methods were performed for one-to-one alignments between networks. However, many-to-many mapping might resolve the limitation of one-to-one mapping in aligning PPIN highlighted above. We are developing a framework where many-to-many mapping is allowed, and preliminary results indicate that the many-to-many mapping can cover most of the BLAST mapping results and increase the overlap between alignment methods.

One of the goals of global alignment is to provide a framework to understand the evolution of PPINs. Multiple network alignment is the key to such analysis and has been addressed by many algorithms such as NetworkBLAST-M [46] for local alignments, IsoRankN [7] and CAPPI [9] for global alignment. However, evaluating the multiple network alignments will be even more complicated. To date, evaluation for multiple alignment approaches mainly relies on known homologous groups of proteins such as Homologene [40]. However, these criteria are biased on sequence similarities of proteins and do not reflect the conservation of interactions, which otherwise could be obtained simply not using network information at all. We propose to adapt RNNI to align multiple networks simultaneously by extending the proposed RNNI framework and introduce other measures to evaluate the conservation of PPINs in all aspects including sequence, function and interaction, in a multiple alignment problem.

Acknowledgment

The authors would like to thank the EPSRC Bridging the Gaps for funding the research.

References

[1] J. Flannick, A. Novak, B. S. Srinivasan, H. H. McAdams, and S. Batzoglou, "Græmlin: General and robust alignment of multiple large interaction networks," *Genome Res.*, vol. 16, no. 9, pp. 1169–1181, 2006.

[2] J. Flannick, A. Novak, C. B. Do, B. S. Srinivasan, and S. Batzoglou, "Automatic parameter learning for multiple network alignment," in *Proceedings of the 12th annual international conference on Research in computational molecular biology*, ser. RECOMB'08. Springer-Verlag, 2008, pp. 214–231.

[3] O. Kuchaiev, T. Milenkovik, V. Memisevic, W. Hayes, and N. Przulj, "Topological network alignment uncovers biological function and phylogeny," *Journal of The Royal Society Interface*, vol. 7, no. 50, pp. 1341–1354, 2010.

[4] T. Milenković, W. L. L. Ng, W. Hayes, and N. Przulj, "Optimal network alignment with graphlet degree vectors." *Cancer informatics*, vol. 9, pp. 121–137, 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2901631/

[5] O. Kuchaiev and N. Pržulj, "Integrative network alignment reveals large regions of global network similarity in yeast and human," *Bioinformatics*, vol. 27, no. 10, pp. 1390–1396, May 2011.

[6] R. Singh, J. Xu, and B. Berger, "Global alignment of multiple protein interaction networks with application to functional orthology detection," *PNAS*, vol. 105, no. 35, pp. 12 763–12 768, 2008.

[7] C.-S. Liao, K. Lu, M. Baym, R. Singh, and B. Berger, "Isorankn: spectral methods for global alignment of multiple protein networks," *Bioinformatics*, vol. 25, no. 12, pp. i253–i258, 2009.

[8] X. Guo and A. J. Hartemink, "Domain-oriented edge-based alignment of protein interaction networks," *Bioinformatics*, vol. 25, no. 12, pp. i240–1246, 2009.

[9] J. Dutkowski and J. Tiuryn, "Identification of functional modules from conserved ancestral protein protein interactions," *Bioinformatics*, vol. 23, no. 13, pp. i149–158, 2007.

[10] H. T. T. Phan and M. J. E. Sternberg, "PINALOG: a novel approach to align protein interaction networksimplications for complex detection and function prediction," *Bioinformatics*, vol. 28, no. 9, pp. 1239–1245, May 2012.

[11] E. Gelenbe, "Random neural networks with positive and negative signals and product form solution," *Neural Computation*, vol. 1, no. 4, pp. 502–510, 1989.

[12] V. Atalay and E. Gelenbe, "Parallel algorithm for colour texture generation using the random neural network model," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 6, no. 2&3, pp. 437–466, 1992.

[13] C. Cramer and E. Gelenbe, "Video quality and traffic qos in learning-based subsampled and receiver-interpolated video sequences," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 2, pp. 150–167, 2000.

[14] E. Gelenbe, K. Harmanci, and J. Krolik, "Learning neural networks for detection and classification of synchronous recurrent transient signals," *Signal processing*, vol. 64, no. 3, pp. 233–247, 1998.

[15] E. Gelenbe, T. Feng, and K. Krishnan, "Neural network methods for volumnetric magnetic resonance imaging of the human brain," *Proceedings of the IEEE*, vol. 84, no. 10, pp. 1488–1496, 1996.

[16] E. Gelenbe and S. Timotheou, "Synchronized interactions in spiked neuronal networks," *The Computer Journal*, vol. 51, no. 6, pp. 723–730, 2008.

[17] E. Gelenbe, "Steps toward self-aware networks," *Communications ACM*, vol. 52, no. 7, pp. 66–75, 2009.

[18] E. Gelenbe and A. Stafylopatis, "Global behavior of homogeneous random neural systems," *Applied Mathematical Modelling*, vol. 15, no. 10, pp. 534–541, 1991.

[19] E. Gelenbe, "Steady-state solution of probabilistic gene regulatory networks," *Physical Review E*, vol. 76, no. 3, pp. 031 903+, 2007.

[20] H. Kim and E. Gelenbe, "Anomaly detection in gene expression via stochastic models of gene regulatory networks," *BMC Genomics*, vol. 10, no. Suppl 3, p. S26, 2009.

[21] P. Baldi, S. Brunak, P. Frasconi, G. Soda, and G. Pollastri, "Exploiting the past and the future in protein secondary structure prediction," *Bioinformatics*, vol. 15, no. 11, pp. 937–946, Nov. 1999.

[22] G. Pollastri, D. Przybylski, B. Rost, and P. Baldi, "Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles," *Proteins*, vol. 47, no. 2, pp. 228–235, 2002.

[23] R. Adamczak, A. Porollo, and J. Meller, "Combining prediction of secondary structure and solvent accessibility in proteins," *Proteins: Structure, Function, and Bioinformatics*, vol. 59, no. 3, pp. 467–475, May 2005.

[24] A. Vullo, I. Walsh, and G. Pollastri, "A two-stage approach for improved prediction of residue contact maps." *BMC Bioinformatics*, vol. 7, no. 1, Mar. 2006.

[25] E. Gelenbe, V. Koubi, and F. Pekergrin, "Dynamical random neural approach to the traveling salesman problem," *ELEKTRIK*, vol. 2, no. 2, pp. 1–10, 1994.

[26] E. Gelenbe, A. Ghanwani, and V. Srivivasan, "Improved neural heuristics for multicast routing," *IEEE Journal of Selected Areas of Communications*, vol. 15, no. 2, pp. 147–155, 1997.

[27] E. Gelenbe, "Une généralisation probabiliste du problème sat," *Comptes-Rendus Acad. Sci.*, vol. 313, no. II, pp. 339–342, 1992.

[28] E. Gelenbe, "Learning in the recurrent random network," *Neural Computation*, no. 4, pp. 154–164, 1993.

[29] E. Gelenbe, "G-networks: A unifying model for neural and queueing networks," *Annals of Operations Research*, vol. 48, no. 5, pp. 433–461, 1994.

[30] E. Gelenbe, "Cognitive packet network," *U.S. Patent*, no. 6804201, 2004.

[31] E. Gelenbe and J. Fourneau, "Random neural networks with multiple classes of signals," *Neural Computation*, vol. 11, no. 4, pp. 953–963, 1999.

[32] E. Gelenbe, "Learning in the Recurrent Random Neural Network," *Neural Computation*, vol. 5, no. 1, pp. 154–164, Jan. 1993.

[33] L. Chindelevitch, C. Liao, and B. Berger, "Local optimization for global alignment of protein interaction networks," in *Pacific Symposium on Biocomputing*, 2010, pp. 123–132.

[34] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." *Nature Genetics*, vol. 25, no. 1, pp. 25–29, May 2000.

[35] X. Guo, R. Liu, C. D. Shriver, H. Hu, and M. N. Liebman, "Assessing semantic similarity measures for the characterization of human regulatory pathways," *Bioinformatics*, vol. 22, no. 8, pp. 967–973, 2006.

[36] A. Schlicker, F. Domingues, J. Rahnenfuhrer, and T. Lengauer, "A new measure for functional similarity of gene products based on gene ontology," *BMC Bioinformatics*, vol. 7, no. 1, p. 302, 2006.

[37] T. Xu, L. Du, and Y. Zhou, "Evaluation of go-based functional similarity measures using s. cerevisiae protein interaction and expression profile data," *BMC Bioinformatics*, vol. 9, no. 1, p. 472, 2008.

[38] J. J. Jiang and D. W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy," in *Proc Int'l Conf Research in Computational Linguistics, ROCLING X*, 1997.

[39] B. Aranda, P. Achuthan, Y. Alam-Faruque, I. Armean, A. Bridge, C. Derow, M. Feuermann, A. T. Ghanbarian, S. Kerrien, J. Khadake, J. Kerssemakers, C. Leroy, M. Menden, M. Michaut, L. Montecchi-Palazzi, S. N. Neuhauser, S. Orchard, V. Perreau, B. Roechert, K. van Eijk, and H. Hermjakob, "The IntAct molecular interaction database in 2010," *Nucleic Acids Research*, vol. 38, no. Database issue, pp. D525–D531, Oct. 2009.

[40] D. L. Wheeler, T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, L. Y. Geer, W. Helmberg, Y. Kapustin, D. L. Kenton, O. Khovayko, D. J. Lipman, T. L. Madden, D. R. Maglott, J. Ostell, K. D. Pruitt, G. D. Schuler, L. M. Schriml, E. Sequeira, S. T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, T. O. Suzek, R. Tatusov, T. A. Tatusova, L. Wagner, and E. Yaschenko, "Database resources of the National Center for Biotechnology Information." *Nucleic Acids Research*, vol. 33, no. suppl 1, pp. D39–D45, 2006.

[41] H. Yu, N. M. Luscombe, H. X. X. Lu, X. Zhu, Y. Xia, J.-D. D. Han, N. Bertin, S. Chung, M. Vidal, and M. Gerstein, "Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs." *Genome Research*, vol. 14, no. 6, pp. 1107–1118, 2004.

[42] A. Schlicker and M. Albrecht, "FunSimMat: a comprehensive functional similarity database," *Nucleic Acids Research*, vol. 36, no. suppl_1, pp. D434–439, 2008.

[43] Y. Assenov, F. Ramirez, S.-E. Schelhorn, T. Lengauer, and M. Albrecht, "Computing topological parameters of biological networks," *Bioinformatics*, vol. 24, no. 2, pp. 282–284, Jan. 2008.

[44] M. E. Smoot, K. Ono, J. Ruscheinski, P.-L. L. Wang, and T. Ideker, "Cytoscape 2.8: new features for data integration and network visualization." *Bioinformatics*, vol. 27, no. 3, pp. 431–432, Feb. 2011.

[45] F. J. Massey, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[46] M. Kalaev, V. Bafna, and R. Sharan, " Fast and Accurate Alignment of Multiple Protein Networks," *Journal of Computational Biology*, vol. 16, pp. 989–999, 2009.