Synergies between Network-Based Representation and Probabilistic Graphical Models for Classification, Inference and Optimization Problems in Neuroscience

Roberto Santana, Concha Bielza, and Pedro Larrañaga

Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid
28660 Boadilla del Monte, Madrid, Spain

Abstract. Neural systems network-based representations are useful tools to analyze numerous phenomena in neuroscience. Probabilistic graphical models (PGMs) give a concise and still rich representation of complex systems from different domains, including neural systems. In this paper we analyze the characteristics of a bidirectional relationship between networks-based representations and PGMs. We show the way in which this relationship can be exploited introducing a number of methods for the solution of classification, inference and optimization problems. To illustrate the applicability of the introduced methods, a number of problems from the field of neuroscience, in which ongoing research is conducted, are used.

1 Introduction

Neural systems can often be depicted using networks where each network node represents one constitutive element of the complete represented system. In network representations of neurons, neuronal circuits and brain maps, nodes can respectively represent neuron compartments, neurons, or brain regions. Networks links between the nodes usually represent structural or functional relationships between the constitutive elements of the representation. Sometimes numerical and categorical data is added to the nodes and links, enriching the representation expressivity.

Several authors have stressed the influence that the particular connectivity patterns in neural circuits have in neurophysiological and mental activities [6,16,17]. It is especially relevant to determine how particular connectivity patterns may be related to neuronal diseases or mental disorders [3]. Research on network theory [2,9,18] has shown that network analysis can help to uncover and characterize the patterns of interactions in complex systems. Several applications of network theory have been reported in the field of neuroscience [3,6,29].

A network descriptor is a numerical value that measures a particular structural or topological characteristic of the network. Different network descriptors can offer valuable information about a single network or a family of networks.

However, sometimes a concise representation of a set of networks (e.g. a set of alternative brain region structural connectivity patterns [29], a set of functional connectivity patterns derived from MEG data analysis of different individuals, etc.) is required. In these cases it is not clear to what extent the set of all single network descriptors can serve to represent common characteristics of the networks. An alternative is to construct a more general representation of the set of networks. This can be done using probabilistic modeling of the set of networks.

Probabilistic graphical models (PGMs) [20] are one of the most recurrent machine learning paradigms to specify interactions in complex systems in terms of probabilistic dependencies. They can be used to represent complex relationships between different data sets, including networks. They usually comprise two components: a graphical structure and a quantitative component. The graphical structure displays certain probabilistic conditional independence relationships between variables. The quantitative component, which is a collection of numerical parameters, usually conditional probabilities, gives an idea of the strength of the dependencies. PGMs have been used to represent different neuronal and cortical processes [21,22,24].

A PGM of a set of networks can serve to capture important regularities from the set of networks (e.g. frequent subnetworks) and patterns of interactions between the networks components. It can later be employed to make queries about different structural or topological hypotheses concerning the represented set of networks.

From another perspective, an interesting fact is that the graphical structure of commonly employed PGMs can also be seen as a network. This fact allows the application of results from network theory to this particular domain. The goal is in this case is to extract, from the analysis of the PGMs' derived network descriptors, particular structural characteristics with a possible valuable interpretation in terms of the original system represented by the PGM.

Therefore, we have that, on one hand sets of networks can be conveniently modeled using PGMs, and on the other hand network analysis of the structural component of PGMs can reveal valuable information of the modeled domain. This relationship points to a synergy between network theory and PGMs. In this paper we analyze different aspects of the relationship between these two domains and discuss a number of ways in which research in neuroscience can benefit from this synergy.

2 Networks

We will focus the analysis on undirected, directed and weighted graphs which will also be called networks.

$G = (V, E)$ will represent an *undirected graph*, where $V = \{v_1, \ldots, v_n\}$ is the set of vertices (or nodes) and $E = \{e_1, \ldots, e_m\}$ is the set of edges between the nodes. $G' = (V, E')$ will represent a *directed graph* where $E' = \{a_1, \ldots, a_m\}$ is the set of arcs (directed edges) between the nodes.

In some contexts, in which the distinction between an undirected and directed graph is not relevant, we will use the term *link* to refer to a connection between two nodes, either an edge or an arc.

A *weighted graph* is a directed or undirected graph where parameter w_{ij} represents the weight between nodes v_i and v_j, whenever e_{ij} (respectively a_{ij}) belongs to the set of edges (respectively arcs) of the network.

There are two types of network descriptors: local descriptors that provide information about a node or a link, and global descriptors that contain information about the complete network. The following are examples of network descriptors.

In an undirected graph, the *degree* of a node is the number of adjacent vertices of the node. In a directed graph, the *indegree* (*outdegree*) is the number of incoming (outgoing) arcs of the node.

A path is a sequence of linked nodes that never visit a single node more than once. The *path length* between two vertices is the number of vertices in the path. One node v is *reachable* from another node u if there is a path between them. The *distance* between a node v and a node u is the length of the shortest path between them if u is reachable from v, otherwise it is set to infinity.

The *characteristic path length* of a graph is the average shortest path length between every pair of reachable vertices in the graph. The *betweenness centrality* of a node is the fraction of all shortest paths in the network that traverse a given node. Similarly, *edge betweenness centrality* is the fraction of all shortest paths in the network that traverse a given edge [4]. The *clustering coefficient* of a node is defined as the fraction of the existing number of node links to the total possible number of neighbor-neighbor links [30].

A *(structural) motif* [18,29] is a connected graph or network consisting of M vertices and a set of edges with connectedness ensured forming a subgraph of a larger network. For each M there is a limited set of distinct motif classes. A *module* (also called *community*) is a densely connected subset of nodes that is only sparsely linked to the remaining network.

3 Probabilistic Modeling of Network Sets

We take a set of N directed networks $\mathcal{G} = \{G'_1, \ldots G'_N\}$ as a sample from a wider set of networks that possibly share some type of topological similarity. The number of nodes in each network G' is m. Self-loops are excluded and thus the maximal number of arcs in the network is $\frac{m(m-1)}{2}$. Our goal is to obtain a probabilistic model of networks in $\mathcal{G}$.

For probabilistic modeling, a representation of each solution is needed. We use X_i to represent a discrete random variable. A possible value of X_i is denoted x_i. Similarly, we use $\mathbf{X} = (X_1, \ldots, X_n)$ to represent an n-dimensional random variable and $\mathbf{x} = (x_1, \ldots, x_n)$ to represent one of its possible values.

Although alternative representations are possible, we will represent a network G' using a binary vector $\mathbf{x} = (x_1, \ldots, x_n)$, where $n = \frac{m(m-1)}{2}$. In this representation there is a unique mapping between each variable X_i and a unique arc of

G'. $x_i = 1$ is interpreted as the arc a_i belongs to the network. If $x_i = 0$, arc a_i is absent from the network.

The initial set $\mathcal{G}$ can be mapped to a binary data-set $\mathcal{D} = \{\mathbf{x}^{(1)}, \ldots \mathbf{x}^{(N)}\}$. From this data set it is possible to compute different statistics about the set of original networks. For instance, it is possible to determine the frequencies of all possible subnetwork configurations between a subset of nodes by computing the marginal frequencies of the variables mapping the arcs that depart from or arrive at these nodes. Similarly, we can learn a PGM from $\mathcal{D}$. This will serve as a model of the represented networks. The way in which the initial set $\mathcal{G}$ is selected will influence of the learned probabilistic model [26].

Once a PGM is constructed, it can be used to make queries related to the network class for which the data-set serves as a sample. Applications of this type to a variety of domains exist. We focus here on applications to the domain of network-based problems. Also in this domain, the PGM can be used for tasks involving *classification*, *inference* and *optimization*.

3.1 Classification

Using the PGM it is possible to estimate the probability that a given network G' belongs to the class of networks represented by $\mathcal{G}$. In this case, G' should be transformed into its corresponding binary representation $\mathbf{x}$, and the probability $p(\mathbf{x})$ given by the model can be taken as the class membership probability. A similar strategy can be employed to assign a given network G' to one among a set of k different classes of networks $\mathcal{G}^1 \ldots \mathcal{G}^k$.

3.2 Inference

The PGM can be used to estimate the probability of a particular subnetwork in the class of networks. In this context, partial inference is applied. It is possible, for instance, to estimate how likely is that a given arc is present in the class of networks represented by $\mathcal{G}$. Another approach allows the computation of the network that most likely belongs to the class of networks represented by $\mathcal{G}$. In this case, abductive inference is employed to compute the most probable configuration (or most probable explanation) given by the model, which is then transformed, using the variable to arc one-to-one mapping, to the corresponding network.

3.3 Optimization

Abductive inference is not the only alternative to obtain a given configuration from the PGM. Sampling methods can be also employed with that purpose. Therefore, from a PGM constructed from $\mathcal{G}$ it is possible to generate, using sampling, new solutions that are expected to be similar to those in $\mathcal{G}$. An approach that iteratively applies PGM learning and sampling steps is at the core of estimation of distribution algorithms (EDAs) [14,19], an optimization method based on the use of probabilistic models. EDAs associate a fitness function value

to each possible solution. The fitness function could measure, for instance, how close are the topological characteristics of the candidate network with respect to a given target network [27].

The rationale behind the use of PGMs in EDAs is to capture similar characteristics shared by high quality solutions in order to increase the likelihood of obtaining better solutions, eventually leading to find the optimal solution. Another optimization methods based on PGMs include different applications of loopy belief propagation [31]. However, these methods usually employ a PGM to represent a single network and not a set of networks.

4 Network Analysis of PGMs

Let $P = (G', \Theta)$ be a probabilistic model, where G' and Θ are respectively the graphical and quantitative components of the model. We can assume that G' is a network (e.g. a directed graph). To obtain a weighted network G^w, a matrix of weights could be computed from Θ (e.g. a weight associated to a link can be the mutual information between the related variables in P. In the general case, we have a set of N weighted networks $\mathcal{G} = \{G_1^w, \dots G_N^w\}$ associated to a set of PGMs $\mathcal{P} = \{P_1, \dots P_N\}$.

Our approach considers the computation of networks descriptors for the networks in $\mathcal{G}$ and uses these descriptors to extract information about the models in $\mathcal{P}$. Local and global networks descriptors can be used for different purposes. The former can be applied to reveal characteristics of a single variable or a pair of interacting variables in the PGM. The latter can be used to reveal global characteristics of the PGM.

4.1 Classification

Local network descriptors allow the unsupervised classification of variables (respectively interaction pairs) in different groups according to their role in the system modeled by the PGM. In this case, classification operates in an undirected way: The nodes of a network (respectively, the links) are clustered according to one or more local network descriptors. The variables mapping nodes that are in the same cluster are then considered to belong to the same class. For example, given a threshold, the network links can be classified into two groups according to their betweennes centrality value: Links with low and high betweenness centrality values. We can interpret that links with a high betweenness centrality will play a more important role in PGM-based processes that involve information transmission over the links (e.g. message passing based inference algorithms like loopy belief propagation). Similarly, classification of nodes based on other local network descriptors such as the degree, the clustering coefficient or the reachability values can support additional information about the role played by variables.

Global network descriptors can be directly employed to classify different PGMs according to the topological characteristics of their graphical components. For example, structural network motifs and small cliques, both of which can be seen as a sort of network building blocks, can be used to distinguish PGMs learned from related but different classes of problems [25].

There are two general questions related to the use of network descriptors for classification. The first is the selection of an appropriate classifier. The second is the determination of the (subset of) global network descriptors that better serve to a clear discrimination between the different classes of networks. Another fundamental question is the interpretation that a given network descriptor has in terms of the relationships between the variables (respectively links) in the PGM. It is not clear that every type of network descriptor provides a meaningful interpretation. However, as one of the examples included in the next section shows, some descriptors contain useful information about the PGM.

4.2 Inference

In network theory some research has been devoted [7,8,15] to try to predict, using information about the connections already observed, which vertices are most likely to be connected. Among the network descriptors used to infer if a pair of nodes is linked are the clustering coefficient, the path length, or the vertices degree. Vertices are assumed to have a higher probability of being connected if they have many common neighbors, there are short paths between them or if the product of their degree is large [7]. Inference is not only used to predict the connectivity of missing vertices but also to detect false positive connections, i.e. links that appear in the network but which have a low probability of being connected using the network descriptors.

The application of network descriptor based inference to PGMs seems straightforward. Network descriptors such as the clustering coefficient could be applied to infer interactions between variables in PGMs which are only partially known. They could also be used to detect false interactions which are not rare in PGMs learned from data. Local network descriptors defined for weighted networks could be also applied to predict different measures of interaction strength between variables.

4.3 Optimization

In optimization based on PGMs (e.g. EDAs), the network descriptors extracted from the PGMs learned from the problem candidate solutions, can serve to identify particular characteristics of the optimization problem domain which are captured by the PGM. The network descriptors can also be used to evaluate the impact that the topological characteristics of the PGM graphical component have in the behavior of the optimization algorithm. Network descriptors have been investigated for optimization methods based on belief propagation algorithms [28] and EDAs [25].

For example, in the case of EDAs, each run may produce several PGMs learned from data while evolving. As a result, at the end of the search the user obtains a

set of models which store valuable information about the optimization problem. An analysis of the network descriptors corresponding to the PGMs learned at different generations (e.g. the average vertex degree) provides information about the behavior of the algorithm (e.g. the characteristics of the data sets produced by the EDA). Similarly, we can compare the behavior of the EDA for different optimization problems using the obtained networks descriptors.

5 Applications in Neuroscience

In this section we describe a number of potential application of probabilistic modeling of networks and network analysis of PGMs in the field of neuroscience. In some cases we report ongoing work in this direction.

5.1 Applications to Classification Problems

Different measures of association between neural systems components are usually employed to learn a network from neurobiological data such as MRI or diffusion tensor imaging data [6,23]. Usually, a single network is constructed from the data collected for each individual. In many cases, the networks are a priori classified according to the experimental conditions in which the data has been collected or the characteristics of the individuals. For example, networks can correspond to two sets of healthy and pathological brains. In this case, PGMs representing each set of networks could be constructed and used for classification as described in Section 3.1. Similarly, sets of networks representing inter-neuronal relationships and derived from data corresponding to different single neurons can be used to construct PGMs of neurons and compute the probability of a given neuron configuration, as represented by the corresponding network.

5.2 Applications to Inference Problems

Network-based inference methods are particularly suitable to be applied to inter-neuronal or inter-regional network reconstruction problems. This is also supported by the fact that nodes with similar connection patterns tend to exhibit similar neuronal function.

Let us suppose an undirected (possibly loopy) PGM of a neuronal column is available. In this model, nodes represent neurons and links represent some probabilistic evidence that there is a synapse between two neurons. Due to possible errors in the neuronal column reconstruction, we know that there exist missing (false negative) and false positive links (representing missing and false synapses). In this context, different local network descriptors can be used to classify the links and identify false positives.

Not only inference can be made about links between variables. Algorithms able to determine the modular [11,12] and hierarchical structure [7] of networks have also been proposed and can be used to detect complex structural and functional organization patterns in different types of neuron networks.

5.3 Applications to Optimization Problems

In optimization problems, we have investigated the relationship between the properties of the a priori-known problem structural information and the structural information captured by the PGMs learned by EDAs. To quantify this relationship, network descriptors have been employed.

One of the problems considered has been to investigate the effect that biasing the axonal connection delay values has in the spike-timing dynamics of a class of spiking neural networks. The study of spike-timing dynamics in the brain is of interest for neuroscience since, among other reasons, it is a key issue to investigate the role that the relative timing of spikes of multiple neurons has in the temporal coding in the brain [5,10,13].

In the problem under consideration, we have started from a spiking network whose topological structure is given. Each link between two neurons has an associated numerical value which corresponds to the conductance delay between the neurons. We know that the particular distribution of the spiking network delay values may have an effect in the number of synfire chains [1] and polychronous groups [13] generated by the spiking network. Our ultimate goal is to study the influence of conduction delays in the polychronization process and in particular, to investigate whether the conduction delays can be biased to maximize the number of coexisting polychronous groups. To find networks with optimal conduction delays, an EDA that uses as PGM a tree is applied. The initial spiking network structure is kept fixed and only the delay values are modified by the EDA. Therefore the sample solutions are network delay assignments.

At the end of the EDA optimization process, the trees are processed and different average network descriptors are computed (e.g. degree of each vertex). These descriptors are compared with the network descriptors corresponding to the original spiking network topology. From this comparison we are able to determine to which extent the structural characteristics of the original problem are captured by the PGMs learned by the EDA, supporting information about the accuracy and effectiveness of the learning methods used by the optimization algorithm.

6 Conclusion and Future Work

In this paper we have presented different alternatives for the application of probabilistic modeling of network-based representations and network-based analysis of PGMs to classification, inference and optimization problems. Our proposal relies on two important properties: 1) The representational capabilities of PGMs to describe complex interactions between the components of a given system. 2) The amount of structural information that can be captured by network descriptors.

Although considerable work has been devoted to the use of PGMs to represent the behavior of neural systems and the application of network theory to study neuronal and mental processes, the combination of both approaches has not been treated in detail. The results presented in this paper about the potential application of the synergy between PGMs and network-based representations

are a first, still preliminary, step. We expect more results could be obtained in this domain from the combined approach of both types of representations.

Acknowledgements

This research is part of the CajalBlueBrain project. It has been partially supported by TIN-2008-06815-C02-02, TIN2007-62626 and Consolider Ingenio 2010 - CSD2007-00018 projects (Spanish Ministry of Science and Innovation).

References

1. Abeles, M.: Local Cortical Circuits: An Electrophysiological Study. Springer, Heidelberg (1982)
2. Amaral, L.A.N., Scala, A., Barthélémy, M., Stanley, H.E.: Classes of small-world networks. Proceedings of the National Academy of Sciences (PNAS) 97(21), 11149–11152 (2000)
3. Bassett, D.S., Bullmore, E., Verchinski, B.A., Mattay, V.S., Weinberger, D.R., Meyer-Lindenberg, A.: Hierarchical organization of human cortical networks in health and schizophrenia. Journal of Neuroscience 28(37), 9239–9248 (2008)
4. Brandes, U.: A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25, 163–177 (2001)
5. Brown, E.N., Kass, R.E., Mitra, P.P.: Multiple neural spike train data analysis: State-of-the-art and future challenges. Nature Neuroscience 7, 456–461 (2004)
6. Bullmore, E., Sporns, O.: Complex brain networks: Graph theoretical analysis of structural and functional systems. Nature Reviews: Neuroscience 10, 1–13 (2009)
7. Clauset, A., Moore, C., Newman, M.E.J.: Hierarchical structure and the prediction of missing links in networks. Nature 453, 98–101 (2008)
8. Clauset, A., Moore, C., Newman, M.E.J.: Structural Inference of Hierarchies in Networks. In: Statistical Network Analysis: Models, Issues, and New Directions, pp. 1–13. Springer, Heidelberg (2008)
9. Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: Critical phenomena in complex networks. Journal of Modern Physics 80(4), 1275–1335 (2008)
10. Durstewitz, D., Gabriel, T.: Dynamical basis of irregular spiking in NMDA-driven prefrontal cortex neurons. Cerebral Cortex 17, 894–908 (2007)
11. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science 315, 972–976 (2007)
12. Guimera, R., Amaral, L.A.N.: Functional cartography of complex metabolic networks. Nature 433, 895–900 (2005)
13. Izhikevich, E.M.: Polychronization: Computation with spikes. Neural Computation 18(2), 245–282 (2006)
14. Larrañaga, P., Lozano, J.A. (eds.): Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation. Kluwer Academic Publishers, Dordrecht (2002)
15. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. Journal of American Society for Information Science and Technology 58(7), 1019–1031 (2007)
16. Lichtman, J.W., Livet, J., Sanes, J.R.: A technicolor approach to the connectome. Nature Reviews: Neuroscience 9, 417–422 (2008)

17. McIntosh, A.R.: Towards a theory of cognition. Neural Networks 13, 861–870 (2000)
18. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: Simple building blocks of complex networks. Science 298, 824–827 (2002)
19. Mühlenbein, H., Paaß, G.: From recombination of genes to the estimation of distributions I. Binary parameters. In: Ebeling, W., Rechenberg, I., Voigt, H.-M., Schwefel, H.-P. (eds.) PPSN 1996. LNCS, vol. 1141, pp. 178–187. Springer, Heidelberg (1996)
20. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo (1988)
21. Rao, R.P.N.: Neural models of Bayesian belief propagation. In: Bayesian Brain. Probabilistic Approaches to Neural Coding, pp. 239–267. MIT Press, Cambridge (2007)
22. Rolfe, J.T.: The cortex as a graphical model. Master's thesis, Computation and Neural Systems, California Institute of Technology (2006)
23. Rykhlevskaia, E., Gratton, G., Fabiani, M.: Combining structural and functional neuroimaging data for studying brain connectivity: A review. Psychophysiology 45, 173–187 (2008)
24. Sajda, P., Baek, K., Finkel, L.: Bayesian networks for modeling cortical integration. In: Akay, M. (ed.) Handbook of Neural Engineering, pp. 585–600. Wiley Press, Chichester (2007)
25. Santana, R., Bielza, C., Lozano, J.A., Larrañaga, P.: Mining probabilistic models learned by EDAs in the optimization of multi-objective problems. In: Proceedings of the 11th Annual Genetic and Evolutionary Computation Conference GECCO 2009, pp. 445–452. ACM, New York (2009)
26. Santana, R., Larrañaga, P., Lozano, J.A.: Interactions and dependencies in estimation of distribution algorithms. In: Proceedings of the 2005 Congress on Evolutionary Computation CEC 2005, Edinburgh, UK, pp. 1418–1425. IEEE Press, Los Alamitos (2005)
27. Santana, R., Larrañaga, P., Lozano, J.A.: Evolving optimized brain networks topologies using multi-objective evolutionary computation (submmitted for publication 2010)
28. Santana, R., Mendiburu, A., Lozano, J.A.: An empirical analysis of loopy belief propagation in three topologies: Grids, small-world networks and random graphs. In: Jaeger, M., Nielsen, T.D. (eds.) Proceedings of the Fourth European Workshop on Probabilistic Graphical Models (PGM 2008), pp. 249–256 (2008)
29. Sporns, O., Kötter, R.: Motifs in brain networks. PLoS Biology 2(11), e369 (2004)
30. Watts, D.J., Strogatz, S.: Collective dynamics of small-world networks. Nature 393(6684), 440–442 (1998)
31. Yanover, C., Weiss, Y.: Approximate inference and protein-folding. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems, vol. 15, pp. 1457–1464. MIT Press, Cambridge (2003)