SIGBio Record

Newsletter of the SIGBio
ACM Special Interest Group

Volume 9, Issue 1, May 2019 ISSN 2159-1210

Editor's Notes:

As discussed in the SIGBio meeting, our community is growing and the research interests are broadening. Consequently, the SIGBio Record will contain in future more articles and perspective covering novel arguments.

The issue presents;

A contributed article:

NITR-DHH: A T2-weighted Brain Magnetic Resonance Image Dataset by Swagatika Devi, Sambit Bakshi, and Manmath Narayan Sahoo

The abstract of the PhD Thesis by Marianna Milano at University of Catanzaro.

We thank contributors for this issue and hope that readers will find interesting references to their work in Bioinformatics and Health Informatics area.

Pietro Hiram Guzzi
Pierangelo Veltri

NITR-DHH: A T2-weighted Brain Magnetic Resonance Image Dataset

Swagatika Devi, Sambit Bakshi, and Manmath Narayan Sahoo
Department of Computer Science and Engineering
National Institute of Technology Rourkela
Odisha – 769008, India
swagatiikaa@gmail.com, sambitbaksi@gmail.com, sahoom@nitrkl.ac.in

ABSTRACT
This report describes the dataset NITR-DHH that bears images of T2-weighted brain from magnetic resonance imaging system. The dataset provides magnetic resonance images of brain for different age groups. These images can be utilized by the researchers for investigation in the domain of magnetic resonance image segmentation, privacy preservation, or classification of pathological brain.

Categories and Subject Descriptors
H.1.2 [**Models and Principles**]: User/Machine Systems – *human factors, human information processing*.

General Terms
Medical imaging, Medical datbases.

Keywords
Magnetic Resonance image classification, Pathological brain detection, Medical data, Privacy Preservation.

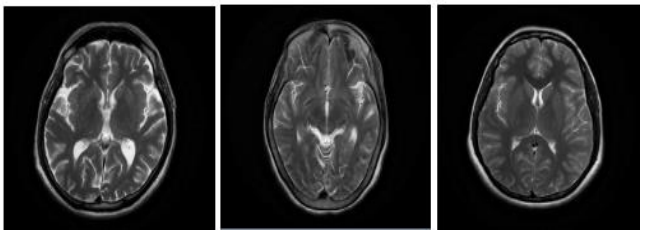
1. INTRODUCTION
Magnetic Resonance Imaging (MRI) has several benefits over other medical imaging modalities as it is based on the magnetization properties of atomic nuclei. Hence, it has an important role in analysis of neurotic abnormalities and neuroscience research. This technique incorporates imaging in various ways to provide the functional activity of the brain. These brain MR images can be utilized by the researchers who are working on image segmentation, privacy preservation, or classification of pathological brain.

A stringent need of a dataset is felt for these researches, as very few standard datasets are publicly available for brain MR images. The research groups working in this domain are testing the result of their proposed mechanisms on small ad-hoc datasets. Hence, the results obtained by different research groups are not directly comparable, and sometimes suffers from imbalanced class problems as there may be very less samples of a particular class. Furthermore, such small datasets cannot cater to the need of large (and variety of) training data needed for deep learning environment. We have collected a dataset from District Headquarter Hospital Puri and announcing it to be publicly available so that global researchers can have a common platform for performance evaluation of their works in this domain.

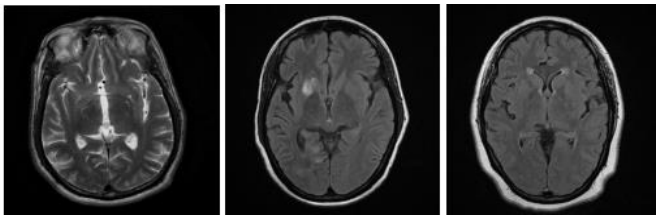
2. DETAIL OF THE DATABASE
The dataset comprises of T2-weighted MR images of the brain along axial plane. This is a purely *anonymous* dataset, i.e. no information of patients will be made public. It can be used mainly for pathological brain detection. Besides the images of normal brain, NITR-DHH also includes some diseased brain images. The detail on these dataset is described in Table 1. Some sample brain MR images (normal and abnormal) are shown in Figure 1.

Table 1. Description of the dataset used

Dataset	#images	Image Format	Types of samples
NITR-DHH	356	JPEG	Normal:209 Abnormal:147



(a) Normal Brain MR images



(b**)** Abnormal Brain MR images

Figure 1. Some sample images from the dataset

3. SUMMARY
This dataset is intended for research towards (i) neurotic abnormalities analysis, (ii) MR image classification, and (iii) impact of privacy preservation on classification. The dataset can solely be used for research purpose with proper acknowledgement. The database can be downloaded from [1].

References
[1] NITR-DHH: A T2-weighted brain Magnetic Resonance image dataset. Available online at: http://nitrkl.ac.in/CS/NITR-DHH.htm.

Alignment Algorithms for Biological and Biomedical Networks Comparison

Marianna Milano
Department of Surgical and Medical Sciences University of Catanzaro
Catanzaro, Italy
m.milano@unicz.it

ABSTRACT

This abstract present a Ph.D Thesis. The experimental work has been developed at University of Catanzaro, Italy.

CCS Concepts

•**Networks** → Alignment Algorithms;

Keywords

Network alignment; Graph theory; PPI networks; Brain Network; Heterogeneous Network

1. PH.D. DISSERTATION

The modeling of biological systems as networks has become a de-facto standard and the application fields span from molecular biology to connectome analysis [8]. According to graph-theory, nodes of the graph represent biological entities and edges represent the associations among them. For instance, biological networks also referred to as Protein-Protein Interaction Networks (PINs), model biochemical interactions among proteins [10]. Nodes represent the proteins from a given organism, and the edges represent the protein-protein interactions. Similarly, the whole system of the brain elements and their relations, so called brain connectome [9], can be modelled as graph where the nodes represent the region of interest (ROI), and the edges represent the functional or anatomical connections.

In both cases, the use of graph theory enables the straightforward analysis of biological properties through the investigation of graph properties.

For example, the comparison of PINs has evidenced the conservation of patterns of interactions among the evolution [2]. Instead, the graph analysis applied to brain connectome has led to the identification of changes in the structure of the networks related to aging and diseases [3, 1]. Comparison of graphs is based on comparing their global properties, such as clustering coefficient, or node degree distribution, or on the analysis of their internal structure, formally known as network alignment (NA).

NA algorithms fall into two major classes [5]. The local network alignment searches relatively small similar subnetworks in the two input networks, which are likely to represent conserved functional structures [6]. This algorithms exploit both biological and topological information of input network to build the alignment. Instead, the global network alignment looks for the best superimposition of the whole two input networks [4] and they rely on the use of only the topology of input networks to build an alignment. More recently, the trend is to integrate both approaches to improve performances of the alignments [7].

In this context, the first goal of Ph.D. work regarded the possibility to improve the performance of the network alignment algorithm by mixing the two approaches, global and local. For improving local alignment algorithms and especially to increase their robustness, since the impact of the bias is strictly related to the building of the alignment graph, we considered the possibility to use topological information extracted from global alignment to guide the steps of the local alignment. Therefore, such considerations have been used to design GLAlign (Global Local Aligner), a novel local network alignment algorithm. GLAlign combines the topology information from global alignment with biological information (e.g. homology relationships) by applying a linear combination of weights that combines topological and biological information. Finally, it uses this global mapping as input for a local network aligner. We tested it on PPINs from different species and we evaluated the quality of results by computing the semantic similarity among the modules. Results confirm that GLAlign outperforms other state of the art methods. Then, we extended GLAlign by implementing SL-GLAlign. To assess our methodology, we tested SL-GLAlign on biological networks. According to these analyses, we demonstrated the SL-GLAlign methodology, i.e. the use of topological and biological information for the local alignment building, improves the local alignments results regardless the local aligner applied.

The second topic of Ph.D. thesis was developed starting from the consideration that many of al algorithms are able to deal with networks with a single class of nodes and edges, also referred to as homogeneous networks. Recently, many different approaches tried to integrate into a single model the interplay of different molecules, such as genes, transcription factors and microRNAs. A possible formalism to model such scenario comes from node/edge coloured net-

WOODSTOCK '97 El Paso, Texas USA

DOI: 10.475/123_4

works (or heterogeneous networks) implemented as node/ edge-coloured graphs. Therefore, the need for the introduction of alignment algorithms able to analyse heterogeneous networks arises. We focus on the local comparison of heterogeneous networks that may be formulated as a network alignment problem. To the best of our knowledge, the local alignment of heterogeneous networks has not been explored in the past. For this purpose, the second goal of my research activity regarded the development of L-HetNetAligner new algorithm for alignment of heterogeneous networks (node-coloured graphs). This approach is conceptually different from the work proposed in literature since we propose a local alignment algorithm. Considering, other classical local alignment algorithms it should be noted that they are unable to consider heterogeneous networks. The novelty of our method is to take into account of the difference among nodes into the building of the alignment graph. We tested it on some selected heterogeneous biological networks. We find that the use of colours (i.e. taking into account of the difference among nodes) produce better alignments. Moreover, the improvement is also remarkable when considering more than two colours. Considering drawbacks and possible future directions it should be noted that we should work on developing faster methods for the building of the alignment graph as well as refining the building of the alignment graph to discard low quality edges. In parallel, in a similar way of GLAlign, we may consider the use of global network alignment to make the process more robust to noise.

The third topic regards the application of the alignment algorithms on brain network domain. This research topic is focused on the representation of connectomes by using graph theory formalisms, where the nodes are region of interest (ROI) , and the functional or anatomical connections are edges. Main research questions in the field of MRI connectomics fall into two major categories: (i) identification of the structure of the networks representing the connectome, (ii) identification of relevant modules (e.g. a sub-graph) that may be interpreted as biomarkers. These research questions greatly depend on the ability to compare brain networks across subjects and groups of subjects. In this context, the aim of my research activity is based on the the possibility to apply the NA methods for the analysis of to MRI connectomics. NA approaches are widely applied in molecular biology analysis, but they can not be easily applied in the connectome alignment problem. The reason is related to the strategy underlying methodology of alignment. For example, the local network aligners, widely used to build the alignment of PINs [10], as discussed above, require biological information such as homology relationships between nodes of PINs. Since the nodes of the brain networks represent ROIs, the homology information cannot be obtained in the case of connectome networks and then, the local alignment cannot be applied. For this reason, the third goal of the experimental work included the application of global alignment algorithms in order to identify which one builds the best alignment. For this purpose, we selected different global alignment algorithms in order to align structural brain networks, then we analyzed the alignment results in terms of topological quality measures and performance, individuating a possible global alignment algorithm suitable for connectome analysis. However, recent studies have demonstrated in an independent way that the multiple alignment algorithms can exact deeper information than pairwise alignment algorithm when these one are applied to molecular biology analysis. According to these studies, we choose to apply multiple alignment algorithms on MRI connectomics. According to this, we selected different existing state of the art multiple alignment algorithms to build the alignment of diffusion MRI-derived brain networks. The algorithms are applied to build the multiple alignments among the diffusion MRI-derived brain networks. After the alignments were built, we compared the performance of these algorithms individuating a possible multiple alignment algorithm suitable for connectome analysis. Our ongoing study is focused on the implementation of an ad hoc algorithm for connectome alignment. Since there are many conditions in which the classical parcellation is not useful, we retain that this work may open the way for the use of network alignment in atlas-free parcellation.

Acknowledgments

Prof. Mario Cannataro and Prof. Pietro Hiram Guzzi supervised the Ph.D research.

2. REFERENCES

[1] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

[2] M. Cannataro, P. H. Guzzi, and P. Veltri. Impreco: Distributed prediction of protein complexes. *Future Generation Computer Systems*, 26(3):434–440, 2010.

[3] Y.-R. Cho, M. Mina, Y. Lu, N. Kwon, and P. H. Guzzi. M-finder: Uncovering functionally associated proteins from interactome data integrated with go annotations. *Proteome science*, 11(1):S3, 2013.

[4] F. E. Faisal, H. Zhao, and T. Milenković. Global network alignment in the context of aging. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 12(1):40–52, 2015.

[5] P. H. Guzzi and T. Milenković. Survey of local and global biological network alignment: the need to reconcile the two sides of the same coin. *Briefings in bioinformatics*, page bbw132, 2017.

[6] E. Hirsh and R. Sharan. Identification of conserved protein complexes based on a model of protein network evolution. *Bioinformatics*, 23(2):e170–176, 2007.

[7] N. Malod-Dognin, K. Ban, and N. Pržulj. Unified alignment of protein-protein interaction networks. *Scientific Reports*, 7(1):953, 2017.

[8] T. Nepusz and A. Paccanaro. Structural pattern discovery in protein–protein interaction networks. In *Springer Handbook of Bio-/Neuroinformatics*, pages 375–398. Springer, 2014.

[9] O. Sporns, G. Tononi, and R. Kötter. The human connectome: a structural description of the human brain. *PLoS Comput Biol*, 1(4):e42, 2005.

[10] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 122(6):957–968, 2005.