Paper Summary

David Miller

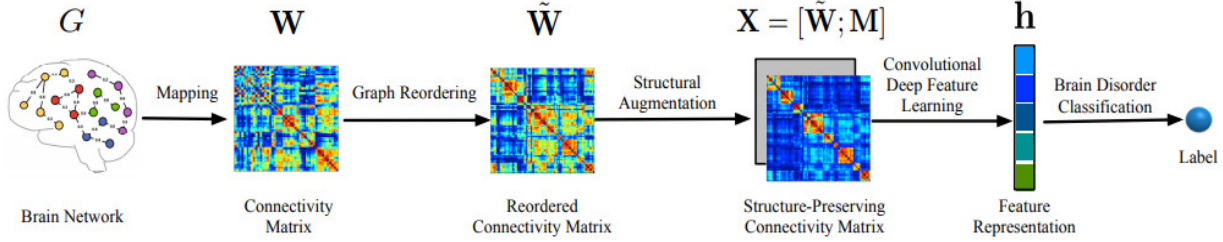CIS 5930: Social Network Mining

February 12, 2018

Figure 1

Mining neurological data is a challenging data. The high non-linearity structure of the brain makes it very difficult for current algorithms to extract accurate representations, or modules, or connected networks. Figure 1 shows modules amongst lobe regions of the brain correctly represented. Amongst all the difficulties the method proposed in this paper, Structural Deep Brain Network (SDBN), attacks four main difficulties: high non-linearity found in underlying brain structure, noise produced from incorrect connection placements, structure preservation to maintain key roles among

Figure 2

modules, and limited quantity and high dimensionality of data. Let $\mathcal{G} = \{G_1, \ldots, G_N\}$ be a set of N brain networks with weighted connectivity matrices $\mathcal{W} = \{W_i, \ldots, W_N\}$, respectively. The goal of the algorithm is the accurately map $f : \mathcal{G} \to \mathcal{H}$, where $\mathcal{H}$ is a set of vector representations of the brain. The three main steps to the algorithm are:
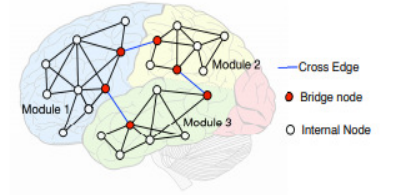
1. **Graph Reordering:** Let $\boldsymbol{L}$ be the Laplacian matrix of the average connectivity matrix $\hat{\boldsymbol{W}}$. K-means clustering is applied to $\boldsymbol{L}$ to achieve K modules $\mathcal{M} = \{M_1, \ldots, M_K\}$ with $V = M_1 \cup \cdots \cup M_K$ and $M_i \cap M_j = \emptyset$ for all i, j such that $i \neq j$. From this the reordered connectivity matrix
$boldsymboltildeW$ for each brain network is produced.

2. **Structural Augmentation:** A module identification $\boldsymbol{M}$ is used to encode the module structure information where

$$m_{i,j} = \begin{cases} k & \text{for} \quad v_i, v_j \in M_k, \ k = 1, \ldots, K \\ 0 & \text{for} \quad i, j \notin M_k, \ k = 1, \ldots, K \end{cases}$$

To matrix $\boldsymbol{M}$ is concatenated with the reordered connectivity matrix $\tilde{\boldsymbol{W}}$.

3. **Unsupervised Learning Augmenting CNN:** The resulting connectivity matrices are fed into a CNN for feature representation.

The obvious future work of this algorithm would be implementation in hospital and physiological structures for early brain disorder detection. Another direction of future work would be generalizing the algorithm for connected networks that need to classify abnormalities, such as crime locations in huge cities. The last direction this paper should head in is scalability. There is no mention of scalability and it is pivotal for today's algorithm to be scalable.

Three strengths I found with the paper are

1. The application of the SBDN network in the medical field can help lower money spent on healthcare through early detection.

2. The algorithm is built upon existing ideas (graph theory, machine learning, neural networks) making it easy to understand and adapt.

3. The SBDN algorithm lends itself well to problems with highly non-linear structure, subgroups, or modules, or networks, that can be used to identify the global network as "regular" or "irregular".

Three weaknesses I found with the paper are

1. There is no mention of scalability. Since there is no mention of the test bench used, I can not assume it can be done on a personal computer.

2. Although the accuracy from the algorithm is higher than existing ones, the overall accuracy does not seem high enough for professional use.

3. The data available for brain imaging networks is limited.

Questions for the reader

1. How exactly does the CNN work? It wasn't exactly clear how it worked as described in the paper.

2. How does the accuracy scale with an increase in samples?

References

[1] Shen Wang, Lifang He, Bokai Cao, Chun-Ta Lu, Philip S. Yu, Ann B. Ragin *Structural Deep Brain Network Mining*, KDD17, August 13-17, 2017, Halifax, NS, Canada.