Nicolas Etessami

APPLYING BAYESIAN NETWORKS TO BRAIN DATA

Supervised by Dr V. Anne Smith and Dr Anastasia Klimovich-Gray

University of St Andrews

Bayesian Network (BN)

A Bayesian network (BN) is a directed graph in which nodes represent variables, and directed edges (arrows) represent probabilistic dependence relationships between these variables.

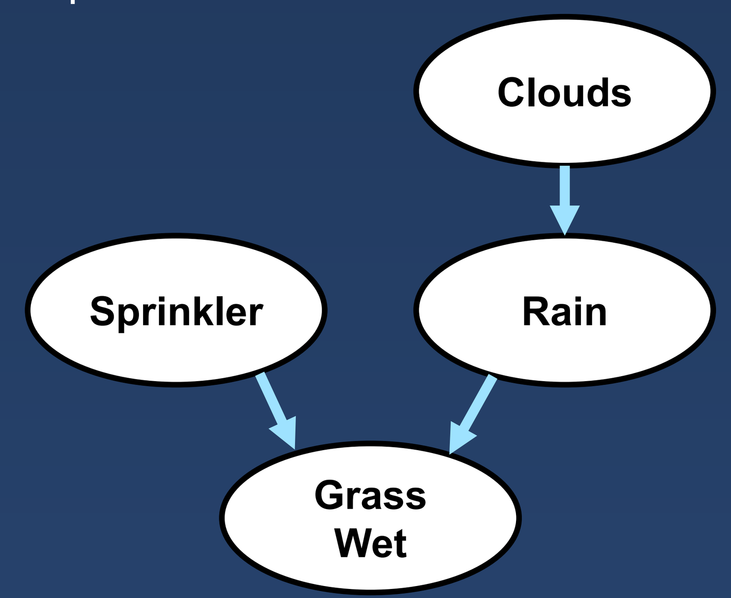

Figure 1: An example 5 node BN. The grass being wet is dependent on rain and sprinklers. It is only dependent on clouds if we don't know if it is raining

Finding the Best BN

We often want to find what the BN of a set of variables is. If we collect data on observations of these variables, we can give each possible BN a score based on how well it fits the data. Then we simply need to use a search algorithm to find the BN with the highest score.

Figure 2: Simulated annealing on a search space where height represents score [1]

This work used a search algorithm called simulated annealing, which repeatedly makes single changes to the BN. If the change increases the score, it moves to the new BN. If the change decreases the score, it moves to the BN with a probability that decreases during the search.

Brain Data (EEG)

This work explores the mostly unresearched area of applying BNs to electroencephalography (EEG) data [1]. This is a low-cost technique used to record electrical activity in the brain using 64 electrodes places on the scalp [2]. If we treat each electrode as a node in a BN, then edges between them represent the flow of neural information over time.

EEG recordings were taken three times from 33 human subjects, while they listened to an audiobook. Each time they performed a different task:

- *No task* – relax and listen casually
- *Semantic task* – listen carefully and answer question afterwards
- *Auditory task* – focus on the speaker's accent

Aims

This work aims to show that **BNs** are effective in learning **meaningful structures** from **EEG data.** If they are, then the BNs would show **task-specific patterns** across subjects, and we would be able to determine which of the three tasks a subject was doing from their learned BN with relatively **high accuracy**.

Setup

The first steps in this work involved pre-processing the data and fine-tuning the search settings of the simulated annealing. One such setting is the **discretisation level**: how many levels (e.g. high, medium, low) the continuous EEG data is split into. This was found to be optimal at **four levels**.

Another setting is the **sampling rate** of the data. This refers to how many electrode readings per second the EEG data contains and determines the time frame of the flow of neural information that the edges in the BNs encode. This was set to **200 Hz**, meaning edges represent neural flow that needed **5 milliseconds** to travel.

Networks

There are many **aspects** we need to consider with the design process. Edges between neighbouring electrodes may not represent neural flow, but rather the electrodes detecting the **same signal**. Also, subjects have **varying skull shapes**, so electrode positions are slightly different for each.

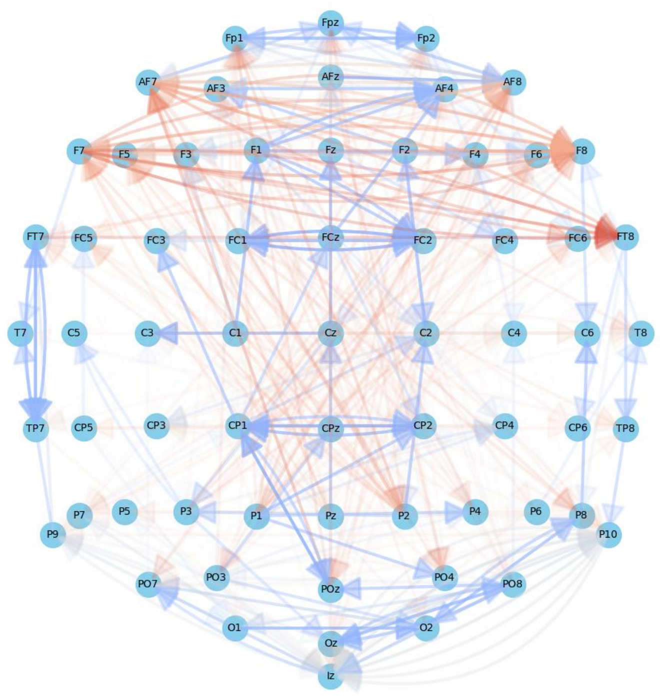

*Figure 3: Soft overlay of the **no task** BN across all subjects. Edges show neural flow. Edges are less transparent if they, or similar, are more common across subjects. Short edges are hidden.*

Node Degrees

The **in-degree** of a node in a BN refers to how many edges enter it, and the **out-degree** how many leave it. In this case, a high in/out degree implies a brain region **receiving/emitting** a lot of neural information, respectively. Since this is very useful information, the heatmaps below were created, however they do not greatly differentiate between tasks.

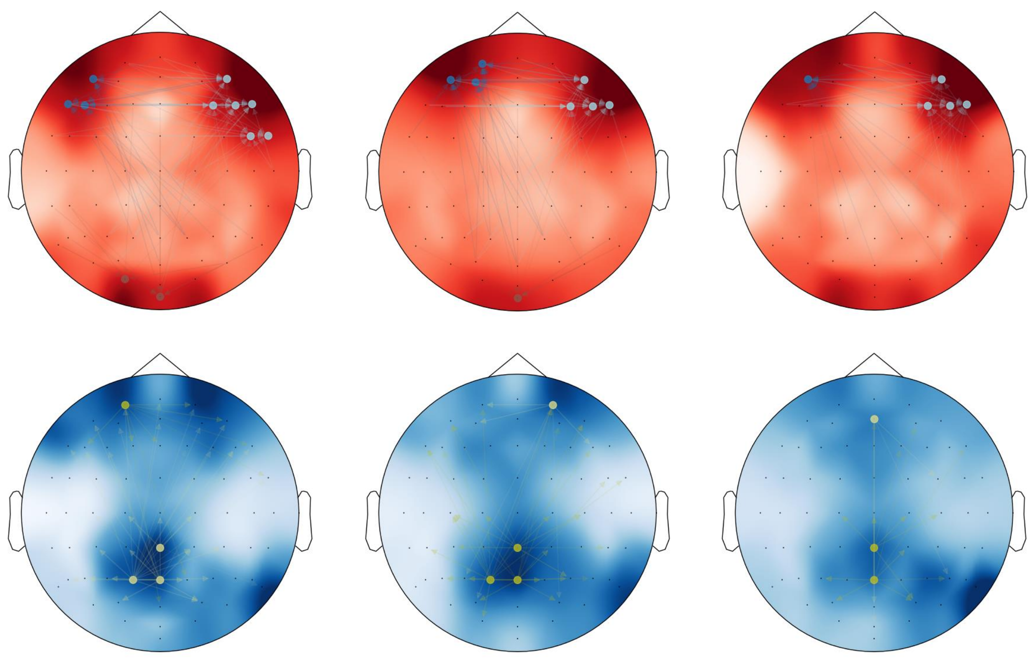

*Figure 4: In (red) and out (blue) degree heat maps. From left to right: **no task**, **semantic task**, **auditory task**. Darker regions indicate higher average degree in that region. Highlighted nodes indicate clusters of high average degrees in that region.*

Clustering

To handle the issue that similar edges convey similar flow but may not be counted as the same, we can also cluster nodes together. This was done using the **Louvain method** for community clustering.

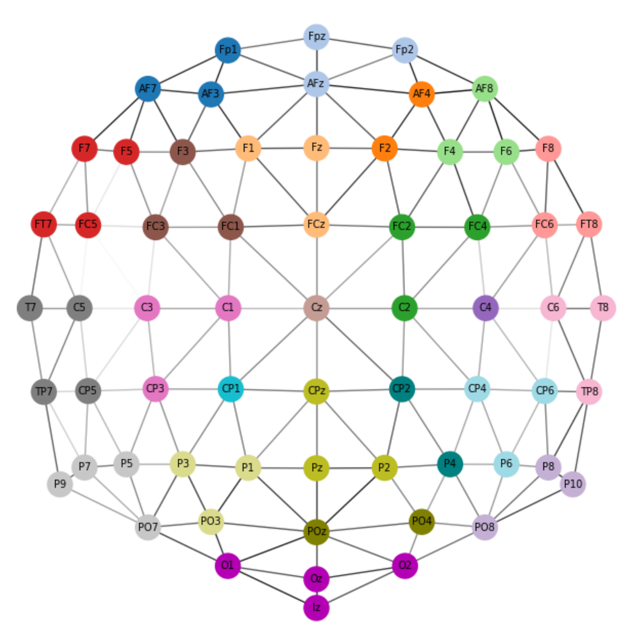

Figure 5: Meta network showing which of the 23 clusters (found to be optimal) each node was mapped to. Edge weights come from how similar nodes are to each other, so clusters are groups of similar nodes.

We can create new **23-node** BNs by treating each cluster as a node. An edge is drawn between two clusters if there was an edge going from a node within one cluster to a node within the other.

Classifier

To effectively differentiate between tasks, we can use a machine learning classifier. A **logistic regression classifier** was trained on edge existence in the 23-node BNs. This was refined to only use 26 edges, giving an accuracy of **0.728**.

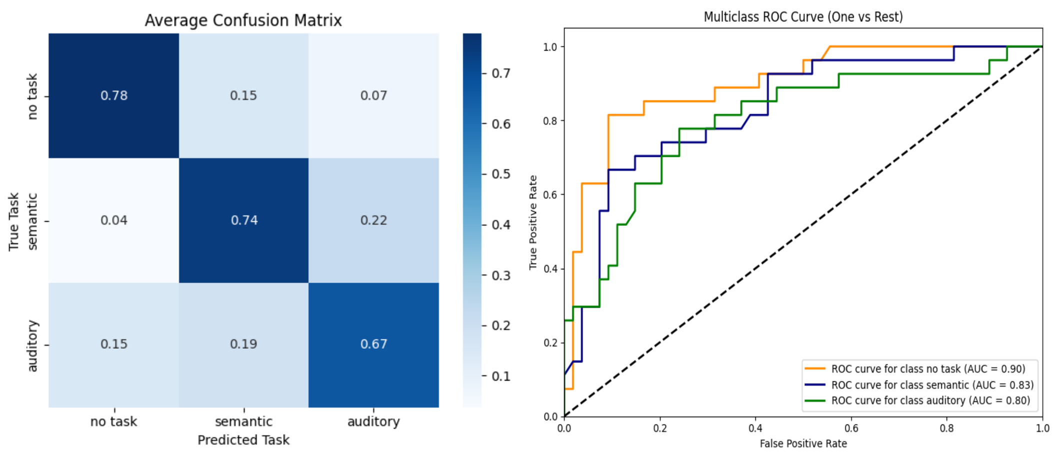

*Figure 6: On the left, a table indicating the **proportions** of classifications for each true task. On the right, a receiver operating characteristic (ROC) indicating the trade-off between **true positive** and **false positive** rates for each task.*

We can use the **weights** that the classifier gives to each edge to find out which edges help to differentiate BNs between tasks, and how.

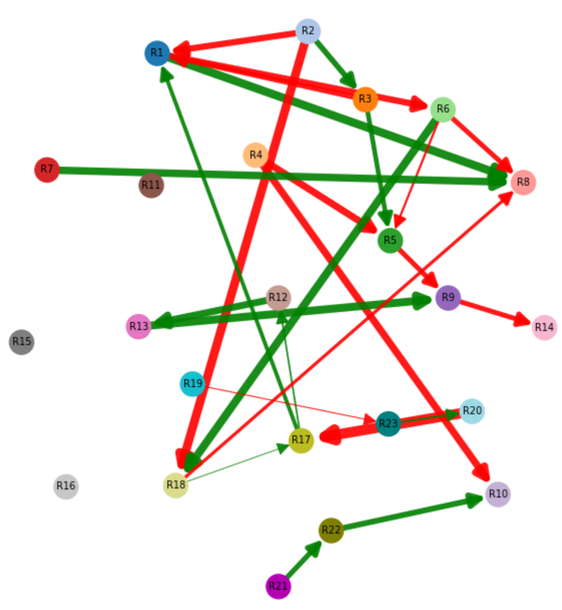

*Figure 7: Classification edge weights for **no task**. **Green** indicates edge is associated with no task. **Red** the opposite. The width of the edge indicates the extent of this. Nodes colours match the clusters in Figure 5.*

Conclusion

There is a lot of **noise** captured by BNs; however, we have shown that we can use certain methods to **filter** some of this out while retaining information. Task-specific structures in the BNs exist, justified by the **high-accuracy** classifier obtained.

Having found what these task-specific structures are, they are **supported** by fundamental knowledge in the field of neuroscience. This clearly suggests that BNs are, to some extent, a **useful** tool that can be applied to EEG data. This work helps toward laying the foundations for **novel discoveries** to be made through future applications of BNs to EEG.

Acknowledgements

I would like to thank the Laidlaw Foundation for their generous support in the Laidlaw Scholars Leadership and Research Programme, without which this work would not be possible. I would also like to thank my supervisor, Dr V. Anne Smith (School of Biology), whose knowledge and guidance has been essential. Finally, I would like to extend my gratitude to the team that gathered and pre-processed the data used in this work at the University of Aberdeen; this was led by Dr Anastasia Klimovich-Gray, who additionally provided very valuable advice and expertise.

References

[1] James Hammond and V. Anne Smith. Bayesian networks for network inference in biology. Journal of the Royal Society Interface, 22(226):20240893–, 2025.
[2] Giovanni Chiarion, Laura Sparacino, Yuri Antonacci, Luca Faes, and Luca Mesin. Connectivity analysis in eeg data: A tutorial review of the state of the art and emerging trends. Bioengineering, 10(3), 2023.

There's a lot more! Follow my profile here

LAIDLAW SCHOLARS