

Comparison of Centrality Measures for the midbrain subregion of two populations of prairie voles

Willliam Ofosu Agyapong

21/03/2022

Abstract

In this report I analyzed network data obtained from the brains of Illinois (IL) and Kansas-Illinois (KI) cross-breeds of male prairie voles to determine if any significant differences existed between the two population in terms of the centrality measures computed from the midbrain subnetwork. Degree, closeness, and eigenvector centralities exhibited statistically significant differences, while the differeces seen in betweenness centrality appeared to be insignificant. The reticular formation appeared to be the most central part of the midbrain subregion of the IL and KI populations of prairie volves.

1 Introduction

1.1 Background

Translational research involving animals offers a great opportunity for understanding human behaviors that would otherwise be impossible through a direct study conducted on humans. However, the success of translational research depends largely on the use of appropriate animal models since typical laboratory animals do not exhibit much of the complex social behaviors observed in humans. This is where the prairie vole (*Microtus ochrogaster*) comes in handy as it has been found to provide an extraordinary animal model that can be used to study several aspects of social cognition, because they establish long-term socially monogamous relationships with mates, provide biparental care and in natural habitats, alloparental behavior is commonly observed. The prairie vole model has already contributed to the better understanding of the neurobiology and genetics of social bonding, parental care, social buffering, the effects of early life experience in later social behavior and social related depression (Ortiz et al., 2018).

1.2 Objective

Utilizing functional magnetic resonance imaging (fMRI) data from culturally and behaviorally distinct populations of prairie voles generated by Ortiz et al. (2021), this report conducts a statistical comparison of each of the centrality measures (degree, closeness, betweenness, eigenvector centrality), comparing the two populations of voles (IL and KI). IL and KI denote the prairie male voles from Illinois and male cross-breed off-springs of Kansas dam and Illinois sires, respectively. The goal is to determine if the two populations of prairie voles exhibit any significant differences in prosocial behavior evident from differences in the **midbrain** subregion network centrality measures.

Because Ortiz et. al (2021) found Illinois prairie males to be more sociable than KI males and differential connectivity accounts for differential expression of prosociality and aggression, I hypothesized that differences in some of the centrality measures for the midbrain subregion will definitely be observed.

1.3 Source of data and data description

The data were provided by Richard J. Ordiz, a Research Technician at the Department of Biological Sciences at UTEP, as an excel workbook consisting of three main sheets relevant to our analysis. First sheet contains IL voles brain network data in the form of an adjacency matrix with 111 nodes produced using a vole-specific atlas. The second sheet contains similar data for the KI voles while the third provides a list of all nodes which are the region of interest (ROIs) and their corresponding subregions. Edges in the network data are weighted by the absolute Pearson's correlation coefficients across all node pairs where a **2.3** threshold was used to avoid weak node connections and zero means no connection between any two pairs of nodes.

The various results obtained in this report are highly reproducible not only for the **midbrain** subregion but for all the other subregions as well; first because all the R codes generating the results are placed side-by-side with the report documentation in an accompanying RMarkdown file, and second via an online web application where the user can interact with to see real time results. To replicate the results for any other subregion, simply provide the name of the particular subregion under the **Global settings** at the beginning of the RMarkdown file. Alternatively, just select any subregion on the online web application. Below is a link to the online web application:

<https://william-agyapong.shinyapps.io/vbna/>

All statistical analyses and visualizations were done in the R statistical software.

The rest of the report is organized as follows. Section 2 presents methods covering graph theory concepts and the statistical methods implemented to compare the two populations of voles by centrality measures. We explore the network data and the centrality measures of interest for the midbrain subregion and conduct statistical comparison tests in section 3. We conclude the report with a brief discussion in section 4.

2 Methodology

2.1 Graph Theory Concepts

In graph theory, a network is defined, in its simplest form, as a collection of points joined together in pairs by lines, where a point is referred to as a node or vertex and a line is referred to as an edge. In the mathematical literature, a network is also called a graph. Examples include the internet in which the nodes are computers and the edges are data connections between them; social networks in which the nodes are people and the edges between them are social connections of some kind, such as friendships, communication, or collaboration; and biological networks such as neural networks consisting of connections between neurons in the brain. This report uses data from neural networks in the brains of male prairie voles.

A network could either be directed or undirected, weighted or unweighted. In a directed network each edge has a direction, pointing from one node to another. Edges that connect nodes to themselves are called *self-edges* or *self-loops*. A network that has neither self-edges nor multiedges is called a *simple network*.

2.1.1 The adjacency matrix

Network data can be represented in different ways and the *adjacency matrix* is one of the fundamental mathematical representation of a network. When two nodes are directly connected by an edge they are said to be *adjacent*, hence the name *adjacency matrix*. The adjacency matrix A of an undirected simple network is defined as an $n \times n$ matrix with elements A_{ij} such that

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between nodes } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases}$$

where n represents the number of nodes or vertices.

The networks considered in this report have edges that represent the strength of linear relationship between nodes measured by the absolute value of the Pearson's correlation between pairs of nodes. Such networks are called *weighted networks* where, in our case, the elements A_{ij} equal to the absolute value of the Pearson's correlations.

2.1.2 Centrality Measures

Centrality measures quantify how important or central nodes are in a network, where the definition of importance often relies on the particular context from which the network was derived. This report considers the four popular centrality measures which are described in the next subsections.

2.1.2.1 Degree Centrality Degree is the simplest centrality measure for a node in a network which measures the number of connections (edges) the node has. In directed networks, nodes have both an *in-degree* and an *out-degree*, and both may be useful as measures of centrality in the appropriate circumstances. The out-degree of a node is the number of other nodes to which a vertex has an outgoing edge directed to. The in-degree is the number of edges received from other vertices. A node with the highest degree has the most connections to other nodes in the network. This means degree centrality can help us find very connected individuals, popular individuals, individuals who are likely to hold most information or individuals who can quickly connect with the wider network.

2.1.2.2 Betweenness Centrality Betweenness centrality measures the number of times a node lies on the shortest path between other nodes. A path is a series of adjacent nodes (A series of edges that take us from one node to another node). The shortest path between any two nodes is the least amount of total steps (or edges). If a node C is on a shortest path between A and B, then it means C is important to the efficient flow between A and B. Without C, flows would have to take a longer route to get from A to B. Thus, betweenness effectively counts how many shortest paths each node is on. Nodes with high betweenness are key bridges between different parts of a network. The higher a node's betweenness, the more important they are for the efficient flow in a network.

Betweenness centrality can be very large, so it is often helpful to normalize it by dividing by the maximum and multiplying by some scalar when plotting.

2.1.2.3 Closeness Centrality The closeness centrality also makes use of the shortest paths between nodes. We measure the distance between two nodes as the length of the shortest path between them. Farness, for a given node, is the average distance from that node to all other nodes. Closeness is then the reciprocal of farness ($1/\text{farness}$).

2.1.2.4 Eigenvector Centrality Degree centrality only takes into account the number of edges for each node, but it leaves out information about the relative importance of the neighboring nodes. In many circumstances a node's importance in a network is increased by having connections to other nodes that are themselves important. For instance, If A and B have the same degree centrality, but A is tied to all high degree people and B is tied to all low degree people, then intuitively we want to see A with a higher score than B. This is where the eigenvector centrality comes in as an extension of the degree centrality by also taking into account how well connected a node is, and how many edges their connections have, and so on through the network. In other words, the eigenvector centrality awards each node points proportional to the centrality scores of their neighbors.

2.1.3 Other concepts

Another measure of how interconnected a network is **average path length**. This is computed by determining the mean of the lengths of the shortest paths between all pairs of vertices in the network. The longest path length between any pair of vertices is called the **diameter** of the network graph. **Density** of a network is the proportion of edges that actually exist out of the total possible edges that can be formed. This also indicates how interconnected a network is.

2.2 Statistical test for comparison

I used the Wilcoxon signed rank test to compare whether there is a difference in the centrality measures in the IL and KI populations. It is a non-parametric equivalent to the paired samples t-test which compares the median rather than the mean. The distributions of the various centrality measures are generally skewed with some extreme values (outliers), suggesting that the distributions are not normal, so to be on the safer side, Wilcoxon signed rank test instead of the paired t-test was used. Again, another justification is that normality tests are sensitive to sample size in which case small samples most often pass normality

tests, and considering the fact that the sample size of 12 for the midbrain subregion is not large enough, there is high chance that normality test will suggest the use of paired t.test when in fact the underlying distribution is not normal. In contrast, non-parametric tests do not assume any particular distribution for the dependent variable.

One assumption inherent in the use of the Wilcoxon signed rank test is that the data from the groups of voles are dependent or paired. Here, it is reasonable to assume that the centrality measures for the two populations are paired since the KI voles are offsprings of both Illinois and Kansas parents. Another assumption is independence which requires that the paired samples are randomly and independently drawn from their respective populations. In general this assumption is difficult to assess so we reasonably assume that the voles were randomly selected from their respective populations for the study that yielded the network data. The Wilcoxon test also requires the distributions for the two groups to be identical which is satisfied in our case since the distributions of centrality measures for the two populations were found to be skewed in the same directions.

The null hypothesis for the Wilcoxon signed rank that was performed assumes that the true median difference between the paired samples is zero (0) while the alternative hypothesis assumes that the true median difference between the paired samples is not equal to zero.

3 Analysis and Results

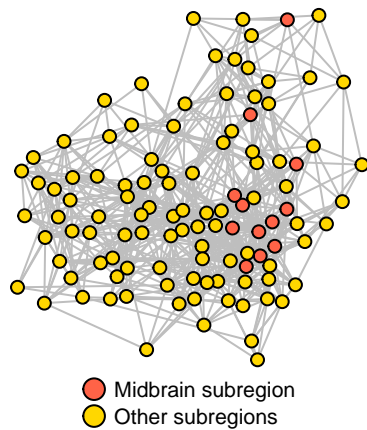
3.1 Network Exploration

Table 1: Characteristics of the two full networks

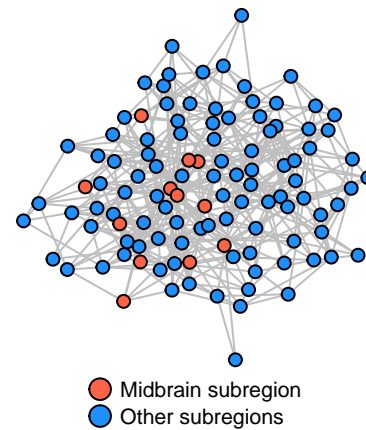
Characteristic	IL Network	KI Network
Number of Vertices	111.0000	111.0000
Number of Edges	680.0000	465.0000
Density (edges)	0.1114	0.0762
Average path length	2.2994	2.5422
Diameter	10.9949	13.3061

3.1.1 Full network with all available edges (connections)

IL Voles Network

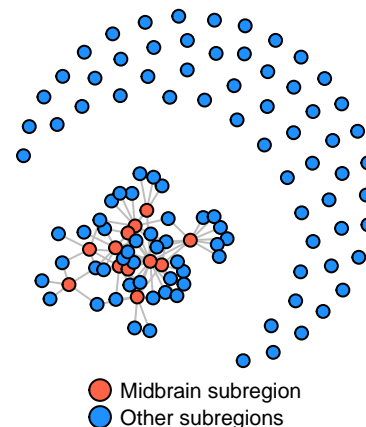
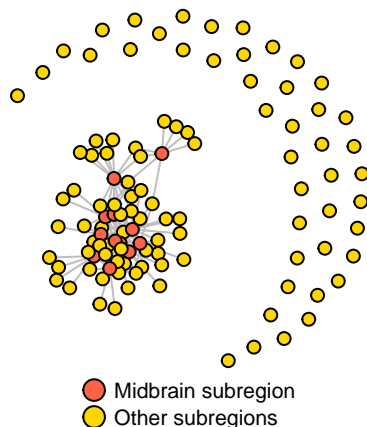


KI Voles Network



The first two graphs (from left to right) display the full networks for the network data for Voles from the two populations including all subregions and all available connections among the 111 vertices. There are a total of 680 and 465 edges, respectively. The twelve vertices related to the `midbrain` subregion, which are the main focus of the report, are earmarked in red on both graphs. **It can be seen that the Illinois network appears densely connected than the Kansas network.**

3.1.2 Full network with edges for only the nodes of interest and their connections



These two graphs are intended to draw attention to the midbrain subnetwork within the full network and the connections among them as well as connections to other vertices part of the whole network. 160 and 95 such connections exist for IL and KI, respectively.

3.1.3 Exploring the midbrain subregion networks

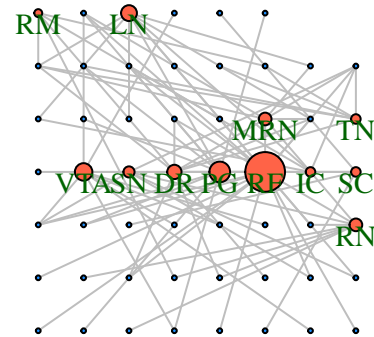
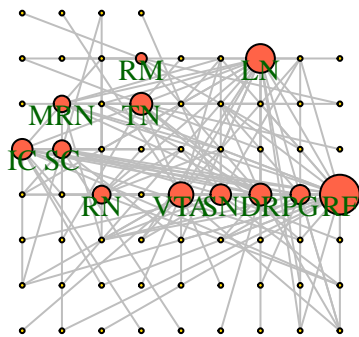
Table 2: Vertices labels lookup table with centrality measures

Subregion	ROI	Abbr	Degree		Closeness		Betweenness		Eigenvector	
			IL	KI	IL	KI	IL	KI	IL	KI
Midbrain	dorsal raphe	DR	16	9	0.0016	0.0014	134	50	0.4963	0.4178
Midbrain	inferior colliculus	IC	15	6	0.0016	0.0013	105	4	0.4245	0.2875
Midbrain	lemniscal nucleus	LN	21	10	0.0017	0.0014	228	23	0.6118	0.5361
Midbrain	median raphe nucleus	MRN	12	8	0.0015	0.0013	32	17	0.3981	0.5006
Midbrain	periaqueductal gray	PG	14	13	0.0015	0.0015	32	147	0.5130	0.5372
Midbrain	raphe magnus	RM	8	5	0.0012	0.0013	54	54	0.0634	0.1283
Midbrain	red nucleus	RN	13	8	0.0016	0.0014	45	116	0.4690	0.2847
Midbrain	reticular formation	RF	29	25	0.0017	0.0017	133	470	1.0000	1.0000
Midbrain	substantia nigra	SN	15	7	0.0015	0.0014	28	22	0.5362	0.3068
Midbrain	superior colliculus	SC	13	6	0.0015	0.0013	51	17	0.3863	0.2891
Midbrain	tegmental nucleus	TN	16	6	0.0015	0.0013	166	44	0.2973	0.2211
Midbrain	ventral tegmental area	VTA	18	11	0.0016	0.0015	117	102	0.6358	0.3780

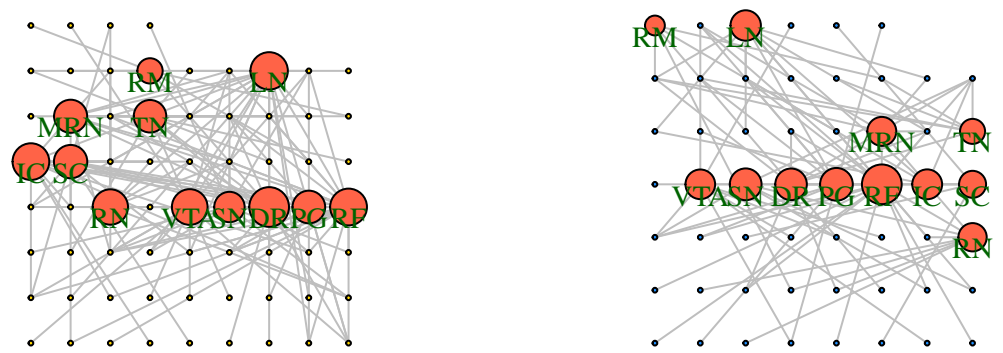
The above table is provided mainly to associate the abbreviated vertices names to their full names for clarity in the subsequent graphs. The centrality measures were simply appended as an additional information.

We explore the **Midbrain** sub-region at length in the network graphs that follow, highlighting the most important vertices by making their sizes proportional to their centrality values. Vertices related to other sub-regions to which the midbrain sub-regions have connections to are less emphasized by minimizing their sizes to a fixed constant. The sizes of the vertices are made proportional to the values of the centrality measure in question for clear visual comparisons. The graphs on the left side correspond to IL prairie voles while those on the right relate to KI prairie voles.

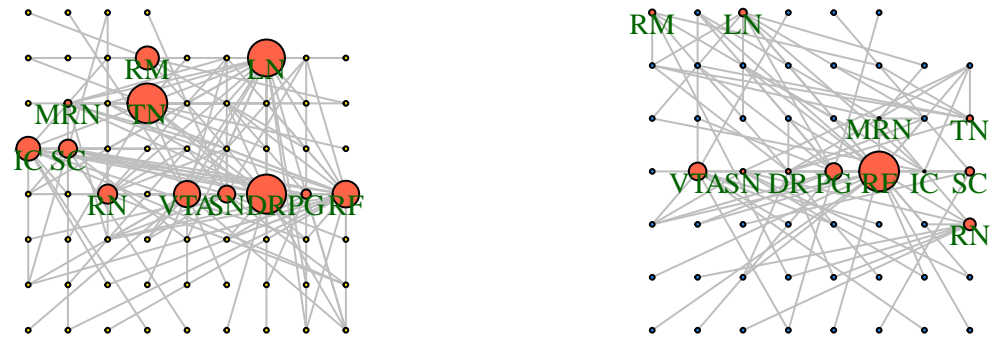
Emphasizing Degree Centrality



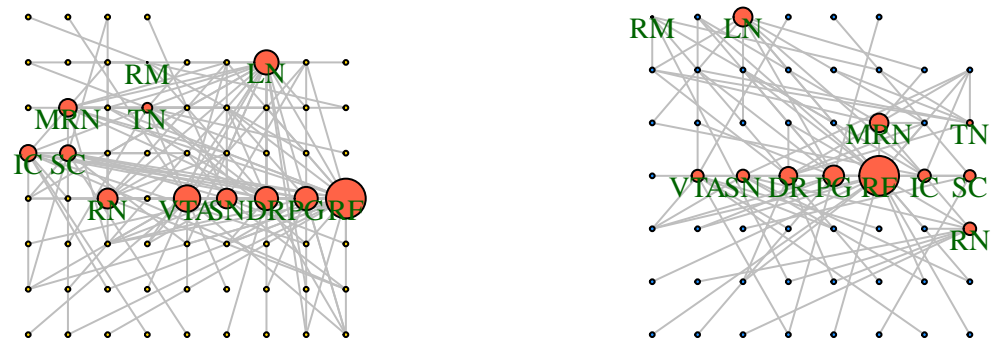
Emphasizing Closeness Centrality



Emphasizing Betweenness Centrality

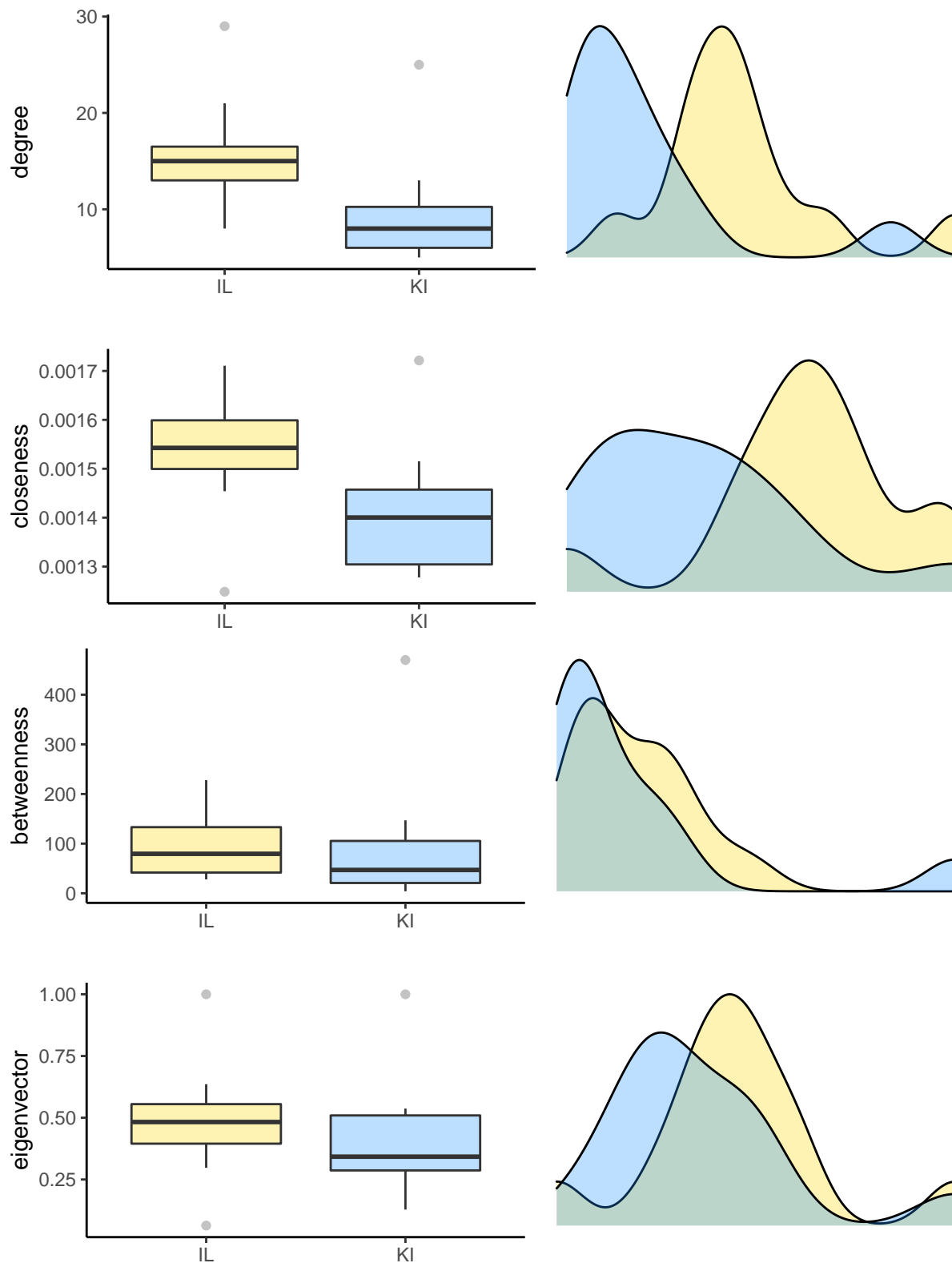


Emphasizing Eigenvector Centrality



Using the information from the table and the above graphs, it is seen that the **reticular formation** is the most important brain part in terms of degree, closeness, and eigenvector centralities for the two populations of voles under consideration. For the betweenness centrality **lemniscal nucleus** happens to be the most important for IL while the **reticular formation** remains the most important node for KI.

3.1.4 Comparison between distributions of centrality Measures



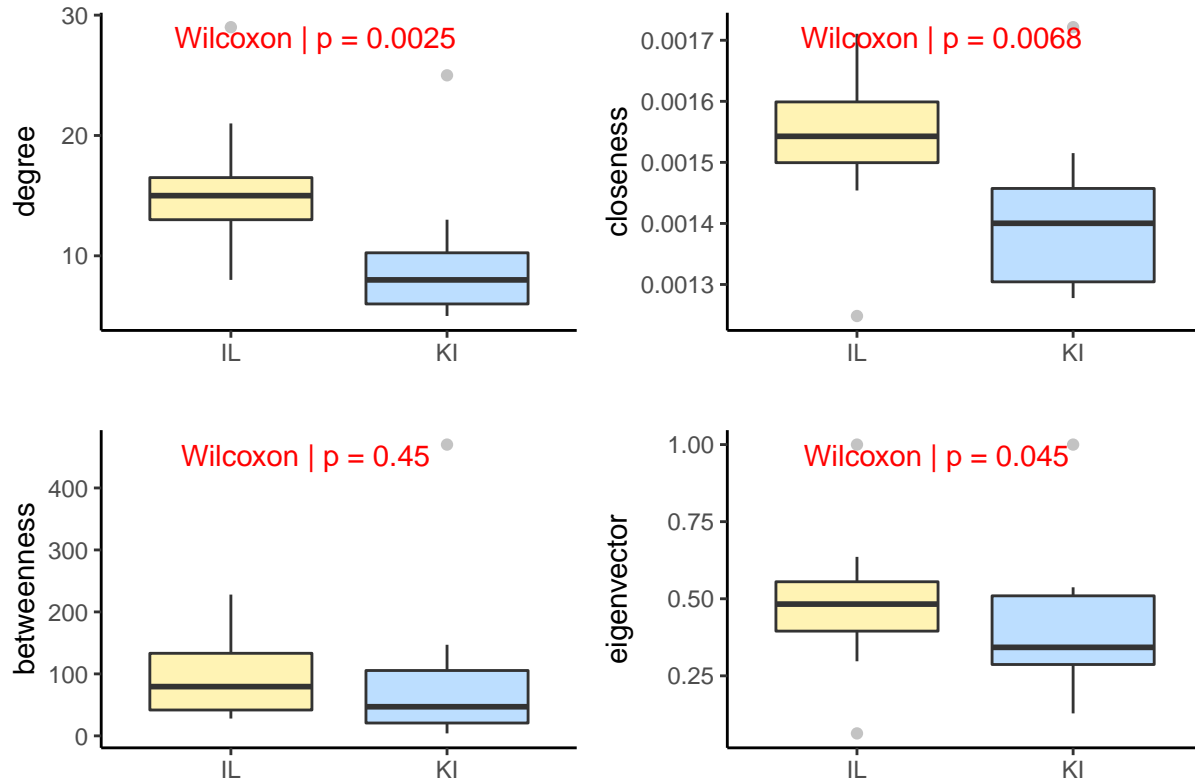
The boxplots provide five-number summaries, median, 1st quartile, 3rd quartile, minimum and maximum values. While the boxplots help us see clearly the median (center), the density plots help reveal the true

shape of the distributions. It is observed that all the distributions are skewed (not symmetric or normal). The points on the graph indicate the presence of extreme values or outliers. For skewed distributions like these ones with outliers, it is best to report the median and interquartile range (or range) as measures of center and spread(variability), respectively.

In general, the medians of the centrality measures for Illinios are considerably larger than those for the KI group. Interestingly, however, the differences are much bigger in terms of degree and closeness. The distributions of betweenness look very similar for the two populations. The sizes of the boxplots or the area under the density curves give us an idea about how the measurements vary by which we see that there is much variability in the distributions for KI than the IL group.

3.2 Statistical test of differences between distributions

From the previous section, noticeable differences were observed in all the distributions for the four centrality measures under discussion. An important question that arises is whether the differences observed is due to chance alone or as a result of the uniqueness in our populations of interest. In other words, is the difference between a given centrality measure for the two populations of Voles statistically significant? To answer this question, we used the Wilcoxon signed rank test, a non-parametric test. The p-values for the tests that were performed are show in red on the plots below. To decide whether to reject or fail to reject the null hypothesis (H_0) that the true median differences between the two populations is zero (0), we used a significance level of 5%, where we reject H_0 if the p-value is less than 0.05, and fail to reject otherwise.



The p-values of the tests for degree, closeness, and eigenvector centrality are significantly less than 0.05. We can conclude that the median degree, closeness, and eigenvector centrality for the Illinois male voles are significantly different from the median degree, closeness, and eigenvector centrality for the KI male voles. On the other hand, the p-value = 0.45 for betweenness is greater than our significance level of 5%, signifying that the betweenness centrality is approximately the same for the two populations of voles.

Tables of statistical tests results including the test statistics, p-values, and confidence intervals can be found in the appendix. The confidence intervals which measure the magnitude of the differences (effect size) also lead us to the same conclusions reached above since the confidence interval for the betweenness centrality is the only one that contains 0.

4 Discussion

The distributions of the centrality measures were found to depart substantially from normality due to the skewness and the presence of some extreme values. This served as grounds for using the Wilcoxon signed rank test instead of the paired two-sample t-test. According to the results for the Wilcoxon signed rank tests we have sufficient evidence supporting the claim that the IL male prairie voles and KI male prairie voles are significantly different in terms of degree centrality, closeness centrality, and eigenvector

centrality when we consider only the **midbrain** subregion. Given that three out of the four centrality measures suggested differences between the midbrain regions of the two population of voles, it is clear that IL prairie voles will exhibit prosocial behaviors and aggression that are different from those exhibited by their KI counterparts, since parts of the midbrain such as *dorsal raphe* and *tegmental nucleus* contribute to the expression of prosocial behaviors.

5 References

- Ashtiani, M., Salehzadeh-Yazdi, A., Razaghi-Moghadam, Z., Hennig, H., Wolkenhauer, O., Mirzaie, M., & Jafari, M. (2017). A systematic survey of centrality measures for protein-protein interaction networks. bioRxiv, 149492.
- Richard J. Ortiz, Amy E. Wagler, Jason R. Yee, Praveen P. Kulkarni, Xuezhu Cai, Craig F. Ferris, Bruce S. Cushing, Functional connectivity differences between two culturally distinct prairie vole populations: insights into the prosocial network, Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2021, ISSN 2451-9022, <https://doi.org/10.1016/j.bpsc.2021.11.007>. (<https://www.sciencedirect.com/science/article/pii/S2451902221003207>)
- Ortiz, J.J., Portillo, W., Paredes, R.G. et al. Resting state brain networks in the prairie vole. Sci Rep 8, 1231 (2018). <https://doi.org/10.1038/s41598-017-17610-9>
- Network Analysis in R by Dai Shizuka: <https://dshizuka.github.io/networkanalysis/tutorials.html>
- Centrality measures: <https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/>
- https://bookdown.org/markhoff/social_network_analysis/your-first-network.html
- <https://kateto.net/netscix2016.html>

6 Appendix

Wilcoxon signed rank test results

Table 3: Wilcoxon signed rank test for differences in degree

Test statistic	P-value	Conf.lower	Conf.upper
78	0.0025	4	8.5

Table 4: Wilcoxon signed rank test for differences in closeness

Test statistic	P-value	Conf.lower	Conf.upper
72	0.0068	0	2e-04

Table 5: Wilcoxon signed rank test for differences in betweenness

Test statistic	P-value	Conf.lower	Conf.upper
42	0.4496	-107.5	103

Table 6: Wilcoxon signed rank test for differences in eigenvector

Test statistic	P-value	Conf.lower	Conf.upper
56	0.0454	0.0054	0.167