ISSN: 2639-7455

Research Article

Earth & Environmental Science Research & Reviews

Social Network Analysis of Ant Group

Runsheng Rong* and **Dr. Dong, Dapeng**

MSc Business Analytics, International Business School Suzhou, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 223000, China,

***Corresponding Author**
Runsheng Rong, *MSc Business Analytics, International Business School Suzhou, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 223000, China.*

Submitted: 2023 Dec 04; **Accepted**: 2024 Jan 10; **Published**: 2024 Jan 22

Citation: Rong, R., Dapeng, D. (2024). Social Network Analysis of Ant Group. *Earth Envi Scie Res & Rev, 7*(1), 01-13.

Abstract
Through constructing animal social network structures based on the relationships between individual ants of two groups of ants, this essay utilizes social network analysis methods to compare the degree of concentration and utilization power on resource and information of these two groups.

Keywords: Animal Social Network Analysis, Ant Group, Comparative Analysis

I. Introduction and Project Aim

Social network analysis (SNA) is a heating method for community research, which is also widely used in animal studies. Social network analysis can provide important insights into the complex social structures of different species and help us better understand how they interact with each other. By identifying key individuals or species in a network, SNA can help conservationists target their efforts more effectively, for example by focusing on protecting important nodes in a network. SNA can also help us understand how environmental and other factors impact social structure and behavior in different species, which could have important implications for conservation and management [1].

The level of social behavior varies not only between different species, but also among groups of the same type of animals, which definitely include ants. Ant is a classical gregarious insect, whose society is known for its highly complex and organized structure. Ant colonies consist of different castes, including the queen, workers, and soldiers, each with specific roles and responsibilities. Ants communicate through chemical signals or pheromones, allowing them to coordinate their activities effectively. They also exhibit division of labor, with different individuals specializing in tasks such as foraging, nest maintenance, or defense. Ant colonies can display impressive levels of cooperation and collective decision-making, often relying on decentralized systems. Additionally, ants exhibit social behavior such as altruism, where individuals may sacrifice their own reproductive potential to benefit the colony as a whole [2]. But the degrees of socialization vary between different groups of ants. Social network analysis (SNA) research in animals has largely focused on mammals, with comparatively little study on insects. Furthermore, there are few analysis indicators available for SNA on ant. To address this, the present study uses multiple analysis methods to compare the sociality of different groups of ants. Two different groups of ants – group 1 and 2 are selected as representative examples of different ant societies.

This paper considers individual ants as nodes and their interactions as links, resulting in two matrices of 113×113 and 131×131 for group 1 and 2, respectively.
The purposes of this paper are summarized as follows:
(1) Compare the similarities and differences in network structure between two ant groups;
(2) Investigate whether there exist only one or several rulers in both ant groups, and the roles individual ants play in both groups;
(3) Determine whether there are distinct clusters present in both ant groups.
Key methods used to address the objectives are described as follows:
(1) Structural analysis: density and centrality;
(2) Community detection: edge betweenness, greedy optimization of modularity, K-means clustering;
(3) Link analysis: PageRank;
(4) Proximity measures: Neumann Kernel, Shared Nearest Neighborhood (SNN).

In terms of the structure of the essay, Section 2 offers a literature review that analyses and critiques existing research. Section 3 presents empirical analysis using social network analysis techniques, comparing the strengths and weaknesses of the methods used. Section 4 offers an interpretation of the results, and finally, Section 5 provides conclusions, contributions, and directions for further research.

II. Literature Review

Animal social network analysis has become an increasingly popular research area in recent years, with studies focusing on a wide range of species including primates, birds, fish, and mammals. Some common research questions in this field include understanding the patterns of social interaction and group dynamics, identifying key individuals or species in the network, and investigating the impact of environmental and other factors on social structure.

One of the earliest and most influential studies in this field was the work of Lusseau (2003) [3], who used network analysis to investigate the social structure of bottlenose dolphins. Since then, numerous studies have applied network analysis to other species, including chimpanzees (Silk et al., 2010) [4], African elephants (Archie et al., 2006) [5].

Different measures of centrality and network structure have been used in these studies, including degree centrality, betweenness centrality, and clustering coefficient. More recent studies have also explored more advanced methods such as the use of community detection algorithms to identify subgroups within the network (Brent et al., 2019) or the use of longitudinal social network analysis to investigate changes in social structure [6].

In summary, many scholars have made some achievements in the research of animal social network. However, the research mentioned above has obvious shortcomings. Obviously, here is already sufficient research on mammals, but is lacking in the ones on insects.

The approaches of related methods will be discussed in the following section, and the software used is RStudio.

III. Implementation and Application Demonstration

In this section, several social network methods are utilized, and their advantages and disadvantages are analyzed subsequently.

A. Social Network Graph

A social network graph is a visual representation of social relationships and interactions between individuals or groups, often displayed as nodes (representing individuals or groups) and edges (representing relationships or connections between them). Social network graphs can be used to analyze and understand social networks, including identifying key actors, communities or groups, and patterns of interaction and communication [7]. The interactive relationships of individuals of both two ant groups can be plotted through the social network graph of sphere and random. Fig. 1 and 2 portray these connections through straight lines that illustrate the relationships of each member. The size of nodes in the figure is represented by degree/4. The position and importance of each subject in the social network can be roughly understood in the social network graphs.

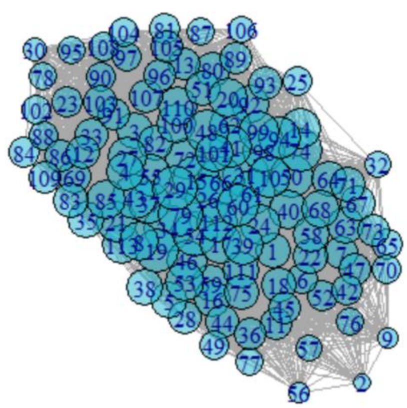

Figure 1. shows social network graph of group 1

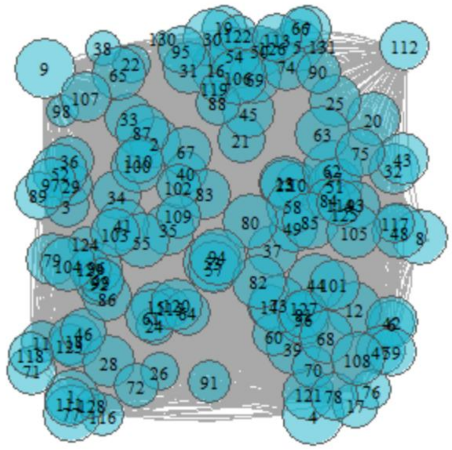

Figure 2. shows social network graph of group 2

B. Structural Analysis

1) Density

The concept of network density pertains to the proximity or relationship between nodes in a network structure [8]. A higher level of closeness among network members leads to a more tightly-knit ant community. A network in which every node is linked to every other node has a density of 1. The densities of the group 1 and 2 are 0.3594343 and 0.3869642, respectively.

2) Centrality analysis

Centrality is a metric that evaluates how central an individual is within a network [9], which is a powerful tool for identifying key individuals or nodes in a network, helping to understand the structure and functioning of the network. Different centrality measures reveal different aspects of a network's importance, and it's important to consider multiple measures to get a more complete picture. This paper examines four types of centrality measures: degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality. These measures describe the node's influence, control over resources, speed of information spread, and overall influence, respectively [10]. Table 1 and 2 below show the descriptive statistics of these centrality indicators for both groups.

However, two important limitations of centrality indices exist [11]. Firstly, what works best for one network may not work well for other networks. Additionally, while identifying the most important nodes in a network is helpful, it may not apply to all other nodes in the network.

	Group 1			
	degree	***betweenness***	***closeness***	***eigenvector***
Mean	80.51	15.39	7.03e-03	0.75
Std.Dev	13.58	12.12	6.67e-04	0.12
Min	41.00	0.00	5.46e-03	0.37
Q1	72.00	7.63	6.58e-03	0.68
Median	79.00	11.79	6.90e-03	0.75
Q3	92.00	20.74	7.58e-03	0.86
Max	109.00	61.20	8.70e-03	1.00
Skewness	-0.20	1.22	2.54e-01	-0.38

Table 1. shows centrality analysis of group 1

	Group 2			
	degree	***betweenness***	***closeness***	***eigenvector***
Mean	100.61	14.42	6.34e-03	0.80
Std.Dev	17.05	9.90	6.01e-04	0.14
Min	25.00	0.00	4.26e-03	0.17
Q1	94.00	7.30	6.02e-03	0.75

Median	104.00	13.47	6.41e-03	0.84
Q3	112.00	20.83	6.76e-03	0.89
Max	128.00	50.22	7.58e-03	1.00
Skewness	-1.66	0.86	-7.89e-01	-1.88

Table 2. shows centrality analysis of group 2

C. Community Detection

1) Community detection based on edge betweenness (Newman-Girvan)

Edge betweenness is a widely used measure for detecting communities in a network. The edge betweenness measure is effective in detecting both small and large communities within a network. Using this approach, edges with high betweenness are identified as bridges between highly connected clusters of nodes. Thus, by iteratively removing the edge with the highest betweenness, we can break down a network into a hierarchy of nested communities. Fig. 3 and 4 display the community detection results based on edge betweenness for both group 1 and 2.

While the method may improve the speed of computer processing, as only intermediate betweenness values are recalculated after edge removal, it only results in a continuous process of breaking down the network into smaller and smaller communities [12]. It does not indicate which partition is best, which means that edge betweenness may not always be the best measure for detecting communities in a network, as other measures such as modularity or conductance may better capture the characteristics of a particular network, which will be introduced in the next section. Despite its limitations, the edge betweenness method remains a useful tool for community detection in social network analysis.

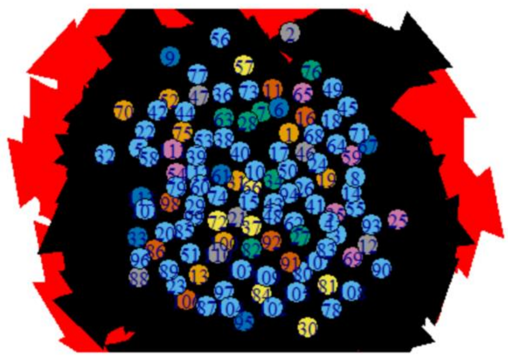

Figure 3. shows edge betweenness of group 1

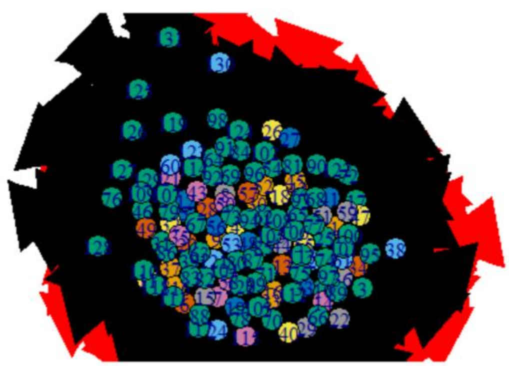

Figure 4. shows edge betweenness of group 2

2) Community detection based on greedy optimization of modularity

Different nodes can belong to different communities, and we aim to enhance the modularity by including nodes that can contribute the most to a single community. We ultimately prefer a partition with a higher modularity score [13]. As it shows in Fig. 5 and 6, the community detection results based on greedy optimization of modularity for both two groups are presented.

The greedy optimization of modularity is a precise and computationally efficient method of detecting communities in networks [14]. However, it has two shortcomings. The method tends to merge communities that are connected through a single link if their size is below a certain threshold. Additionally, as the number of nodes in the network increases, obtaining the best partition becomes increasingly challenging [15].

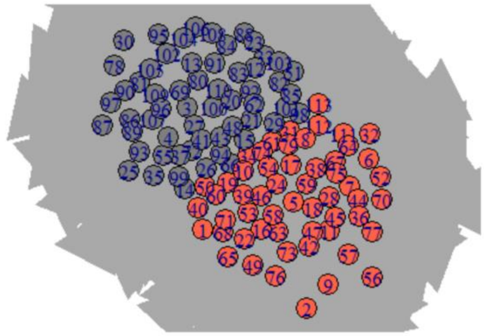

Figure 5. shows greedy optimization of modularity of group 1

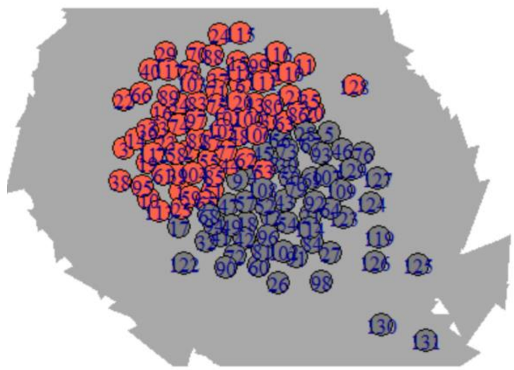

Figure 6. shows greedy optimization of modularity of group 2

3) K-means clustering

The K-means clustering algorithm is commonly used in social network analysis due to its simplicity and efficiency in detecting clusters or communities in the network. K-means clustering divides the data set into K predefined different clusters, so that each node comes into to the cluster with the nearest mean. In addition, the average distance from the center is plotted as a function of K, and the "elbow point" where the rate of descent changes violently can be used to roughly determine K [16]. Fig. 7 and 8 reveal the K-means clustering results for both group 1 and 2.

As for its strengths, K-means clustering is a useful technique for identifying distinct groups or communities within a social network, it can scale to large data sets and easily generalize to clusters of different shapes and sizes [17]. However, K-means clustering is limited in its ability to capture the complex dynamics and interconnections within social networks. The number of clusters needs to be assigned and cannot handle noisy data and outliers [18].

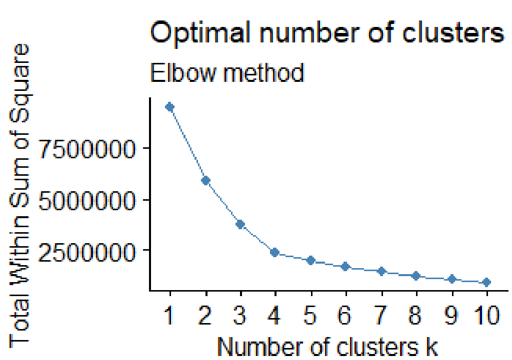

Figure 7. shows K-means clustering of group 1

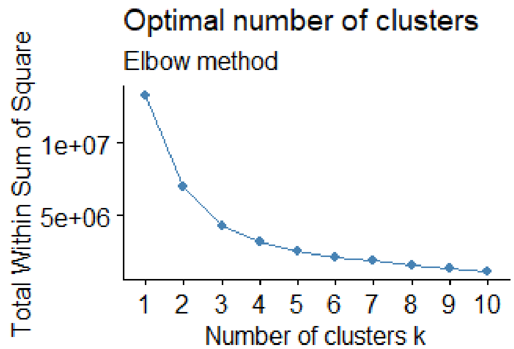

Figure 8. shows K-means clustering of group 2

D. Link Analysis

1) PageRank

PageRank is a metric that measures the importance of a node based on the concept that a node is considered important if other important nodes link to it. The algorithm calculates a score for each node by looking at the incoming links to that node, and then normalizes those scores to add up to 1. The nodes that have higher PageRank scores are viewed as more crucial within the network. This metric can identify key individuals that are highly connected and influential within the social structures of both two ant groups [19]. As it shows in table 3, the descriptive statistics of the PageRank results for both ant groups are listed.

On the positive side, PageRank offers a straightforward and simple way to evaluate the importance of nodes in a network. It is commonly used and well-understood by many researchers in the area of social network analysis. Additionally, it is successful at identifying nodes that serve as important connectors between various groups within the social network. However, PageRank also has certain limitations. It may not differentiate between incoming and outgoing links in the network, giving equal importance to both kinds of links. Therefore, this metric could be susceptible to manipulation by individuals who want to increase their perceived importance within the network. Lastly, it may not consider the impact of other significant features of the network, such as the strength or quality of the connections between nodes [20].

	Group 1	Group 2
Mean	8.85e-03	7.63e-03
Std.Dev	1.60e-02	1.24e-02
Min	2.39e-03	2.14e-03
Q1	2.94e-03	2.69e-03
Median	4.15e-03	3.81e-03
Q3	7.21e-03	6.93e-03
Max	1.41e-01	1.08e-01
Skewness	5.78e+00	5.20e+00

Table 3. shows PageRank of group 1 and 2

E. Proximity Measures

1) Neumann kernel

The Neumann kernel, also known as the Laplacian smoothing function, is a mathematical function that is applied to a network in social network analysis to measure nodal centrality. It is a smoothing function that is based on the average of the neighboring nodes' importance scores. The function assumes that more central nodes will have more high-quality neighbors, and that a node's score should be related to the scores of its neighbors. By applying the Neumann kernel to a social network, researchers can identify important nodes based on the concept of shared influence or centrality. The correlation of nodes in the graph based on the immediate connections and remote connections are modeled through Neumann Kernel. It employs a customizable parameter to adjust the weight assigned to connections that are further away. As a result, two matrices can be generated, K_γ and T_γ. By examining the diagonal values in the K matrix, the relative ranking of nodes can be determined [21]. Here in table 4, the descriptive statistics of the Neumann kernel ranking results of both group 1 and 2 are presented.

One benefit of the Neumann kernel is that it takes into account both relevance and importance. Nevertheless, it might not consider other important network characteristics, such as the quality of the connections between nodes or the existence of dominant individuals who govern social interactions [22].

NK ranking	Group 1	Scores	NK ranking	Group 2	Scores
1	Individual 8	5.66e+04	*1*	Individual 113	2.21e+05
2	Individual 71	5.25e+04	*2*	Individual 114	2.05e+05
3	Individual 68	4.47e+04	*3*	Individual 112	1.90e+05
4	Individual 60	4.07e+04	*4*	Individual 120	1.88e+05
5	Individual 89	3.93e+04	*5*	Individual 108	1.78e+05
...	…	…	...	…	…
112	Individual 3	0	*130*	Individual 3	0
113	Individual 7	0	*131*	Individual 130	0

Table 4. shows Neumann kernel ranking of group 1 and 2

2) SNN

The similarity between nodes is determined by the Shared Nearest Neighbor (SNN) method, which considers the number of common neighbors. Even if there is no direct connection, nodes can be considered similar if they have more than k neighbors in common [23].

The strength of SNN lies in its ability to capture both direct and indirect connections, which facilitates the detection of similarity between non-adjacent vertices. Additionally, in a clustering environment, SNN can handle clusters of various sizes, shapes, and densities. However, it does not take into account the weight of links between nodes [24].

IV. Analysis of Results and Discussion

A. Social Network Graph

According to the analysis results, both group 1 and 2 have multiple nodes with high degrees, no obvious central nodes can be found in both groups.

B. Structural Analysis

1) Density

Group 1 has a density value of 0.3594343 and 4,549 edges, while the density and edge number of group 2 are 03869642 and 6,590, respectively. These values indicate that there is a high level of mutual interaction among individual ants in both groups, which means that the transmission of information and resource in ant society is of a very high efficiency.

2) Centrality analysis

Regarding degree centrality, group 1 has a maximum degree of 109 and a minimum degree of 41, with an average degree of 80.51 and a standard deviation much smaller than the average value at 13.58. On the other hand, group 2 has a maximum degree of 128 and a minimum degree of 25, with an average degree of 100.61 and a standard deviation also quite lower than average at 17.05.
Compare these two ant groups, the degree centrality indices are both high, which means that most of nodes are important in both groups [25]. However, the degree centrality of group 2 is significantly higher than that of group 1, which is further supported by Fig. 9 and 10, which are their respective degree distribution histograms. These histograms show standardized degrees on the x-axis and counts on the y-axis, with the values being standardized across the graphs.

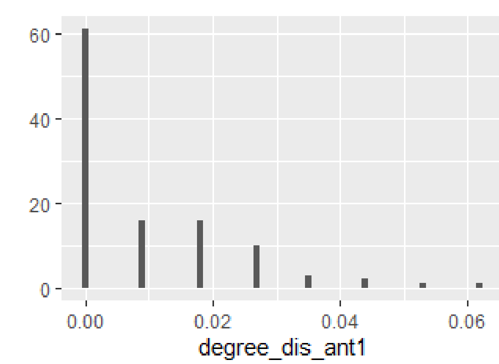

Figure 9. shows degree distribution of group 1

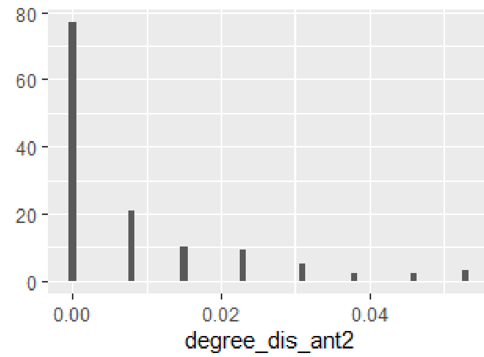

Figure 10. shows degree distribution of group 1

From Fig. 11 and 12, it can be observed that the betweenness centrality results for group 1 show that the maximum value is 61.20 and the minimum is 0. In comparison, group 2 has a maximum value of 50.22 and a minimum of 0. Although their maximum values differ, sorting the nodes by betweenness shows that all the nodes in both groups have low betweenness values, implying that all the nodes play important roles when obtaining resources and information.

```
> sort(betweenness_ant1, decreasing = TRUE)
        15         54         50         61         26         94
61.2027854 50.2846177 49.3707939 40.8728557 37.7494783 37.2804377
        43         27         17         48         74        101
37.0292136 35.9085815 35.6649017 35.3758210 35.0841085 34.7768803
        79         66         55         14         41        110
34.6260637 32.5153192 32.4640490 31.1358825 30.6379480 29.0052175
```

Figure 11. shows betweenness centrality of group 1 (part)

```
> sort(betweenness_ant2, decreasing = TRUE)
        69         52         73         54         79         43
50.2178990 41.4141300 38.6151383 37.9763903 37.0719022 36.0992200
        62         30         64         42         44         63
35.4105583 33.9045313 32.0535851 30.7858518 28.8300795 28.2714093
        96         92         53         12         81         80
27.4426821 26.4639861 25.6121327 24.9107610 24.8568269 24.8110167
```

Figure 12. shows betweenness centrality of group 2 (part)

Closeness centrality measures the average shortest path between a node and other nodes in the network. Nodes with high closeness centrality are closer to other nodes in terms of their information and capabilities. The average, minimum, maximum, and median values of closeness centrality for group 1 are almost identical to the ones of group 2. This suggests that the information dissemination distances of group 1 and 2 are nearly the same, resulting in same level of efficiency in transmitting resources and information between individuals.

Eigenvector centrality measures the efficiency of information transmission and the breadth of information release in the network. From an eigenvector centrality perspective, the average value of group 2 is higher than that of group 1. This implies that the information diffusion and resource acquisition of group 2 are more effective.

C. Community Detection

1). Community detection based on edge betweenness (Newman-Girvan)

Using the “length()” function, we discovered that group 1 is divided into 58 communities, with a modularity of 0.058, while there are 67 communities in group 2, and the modularity is 0.059. This suggests that the partitioning for both group is bad, which can be explained by the special feature of ant colony that individual ants work together to complete different jobs, an individual may be involved in several tasks [26].

2) Community detection based on greedy optimization based on modularity

Using this algorithm, both group 1 and 2 are divided into 2 communities. This algorithm is more effective than the edge betweenness algorithm in terms of the number and division of the communities, for which may be explained by the different social levels that individual ants belong in these communities [2].

3) K-means clustering

The situations of group 1 and 2 are similar. First, the curves of the optimal number of clusters based on the total within sum of squares are drawn. The plots show a significant drop in the sum of squares when the clusters increase from 1 to 2, after which the slope remains relatively constant. As a result, the larger the number of clusters, the smaller the sum of squares. However, because K-means does not provide a clear turning point or “elbow” in the plot, it is difficult to determine the appropriate size of K, or how many clusters to divide into.

D. Link Analysis

1) PageRank

Based on the results listed in Fig. 13 and 14, in regard to PageRank, the scores of top 1 nodes of both groups are extremely high, above 0.1, and there are both approximately ten nodes having scores above 0.02. Moreover, there are 23 individuals in group 1 and 28 individuals in group 2 with PageRank scores above average. From these results, it can be inferred that there exists exactly one leader, which is the queen in both groups.

```
> sort(PR_ant1, decreasing = TRUE)
         71           8          89          68          64 
0.140787237 0.064207603 0.062330151 0.046460761 0.029457673 
         83          60           4          73          29 
0.026223383 0.026059866 0.025768131 0.022690533 0.017004741 
         39         104          49          45          36 
0.016581590 0.016143637 0.013782454 0.013718943 0.013453870 
```

Figure 13. shows PageRank scores of group 1 (part)

```
> sort(PR_ant2, decreasing = TRUE)
        114         113         112         111         110 
0.108316651 0.057952393 0.044838845 0.038694695 0.035985629 
        120         109         108         106         107 
0.033056811 0.030201994 0.025689047 0.021287902 0.020308003 
        105         104          68         101         103 
0.019267010 0.016188018 0.015376046 0.012718033 0.012595437 
```

Figure 14. shows PageRank scores of group 2 (part)

E. Proximity Measures

1) Neumann kernel

The results of the Neumann kernel reveal that the rankings of both groups exhibit a relatively even decrease in scores, while the scores of the top 5 nodes are all extremely high, which confirms the point that there exist strict social level systems in ant colonies.

2) SNN

Regarding the Shared Nearest Neighbor, when the neighborhood index is set as 60 for group 1, among numerous node pairs, there are over 100 pairs of nodes sharing sixty identical neighbors. Subsequently, as the number of neighbors increased, as it shows in Fig. 15, the 75 pairs of nodes in group 1 eventually reaches 18 identical neighbors.


```
> sort(PR_ant1, decreasing = TRUE)
         71           8          89          68          64 
0.140787237 0.064207603 0.062330151 0.046460761 0.029457673 
         83          60           4          73          29 
0.026223383 0.026059866 0.025768131 0.022690533 0.017004741 
         39         104          49          45          36 
0.016581590 0.016143637 0.013782454 0.013718943 0.013453870 
```

Figure 13. shows PageRank scores of group 1 (part)

```
> sort(PR_ant2, decreasing = TRUE)
        114         113         112         111         110 
0.108316651 0.057952393 0.044838845 0.038694695 0.035985629 
        120         109         108         106         107 
0.033056811 0.030201994 0.025689047 0.021287902 0.020308003 
        105         104          68         101         103 
0.019267010 0.016188018 0.015376046 0.012718033 0.012595437 
```

Figure 14. shows PageRank scores of group 2 (part)

E. Proximity Measures

1) Neumann kernel

The results of the Neumann kernel reveal that the rankings of both groups exhibit a relatively even decrease in scores, while the scores of the top 5 nodes are all extremely high, which confirms the point that there exist strict social level systems in ant colonies.

2) SNN

Regarding the Shared Nearest Neighbor, when the neighborhood index is set as 60 for group 1, among numerous node pairs, there are over 100 pairs of nodes sharing sixty identical neighbors. Subsequently, as the number of neighbors increased, as it shows in Fig. 15, the 75 pairs of nodes in group 1 eventually reaches 18 identical neighbors.


```
[1]   31--34   34--66   34--98   34--37   21--34   31--37   31--66
[8]   31--72  31--113   34--72   21--31   31--98  21--113   34--46
[15] 34--113   21--37   21--72   66--98
```

Figure 15. shows Shared Nearest Neighborhood of group 1

When the number of neighbors is small, the result in group 2 is similar to the above situation. However, as the number of neighbors further increased to 100, Fig. 16 demonstrates that there are 12 pairs with the same 100 neighbors. Therefore, the similarity in the follow-up relationships of group 2 is higher than that of group 1.

```
[1]   4--9   1--4   4--8   4--5   8--9   4--12   1--9
[8]   9--12  4--2   5--9   4--13  9--13
```

Figure 16. shows Shared Nearest Neighborhood of group 2

V. Conclusion

A. Summary of Results

After analyzing and comparing multiple indicators between ant group 1 and 2, the following conclusions have been drawn:

1) Similarities between the two groups

Most algorithms used in this essay don't indicate an obvious clustering trend in either ant group 1 or 2. The community detection methods don't provide an effective means of dividing the clusters. This corresponds to the third purpose outlined in the introduction. Almost all nodes in both groups have the ability to transmit information and allocate resources freely, while there exists a ruler in both groups who takes them into control. This confirms to the second purpose mentioned in the introduction. For instance, both group 1 and 2 have nearly distinct node with betweenness centrality above 50, while almost all the rest of nodes in both groups have considerably low values.

Also, the individual ants in both groups are in corporation with other individuals in many tasks, which tend to be a complex system, but the social level systems are also strict in both groups. And the density values of both groups are both high. This can react to the 1st half of the first purpose.

2) Dissimilarities between the two groups

The overall centrality of group 2 is comparatively high, indicating that the connections between the individual ants in group 2 are closer and the resource and information dissemination is more efficient. To sum up, the dissimilarities respond to the 2nd half of the first purpose in the introduction.

B. Contribution and Future Work

This essay makes two primary contributions. Firstly, it adds to the field of animal social network research by comparing the social behavior of two distinct ant groups, which were previously understudied and largely focused on insect research. Secondly, it utilizes RStudio and multiple indicators to conduct a more comprehensive social network analysis that goes beyond the basic indicators examined in existing research.

However, there are some limitations to this study. Firstly, the datasets used only included two samples. Secondly, the datasets used only chose one single type of ant, which may not be representative of the entire ant population. Additionally, as ant groups are dynamic social groups, the relationships between individual ants change over time. This essay, however, only uses static data for network analysis, which does not capture the dynamic change process very well.

To address these issues, future research could expand in the following areas, including collecting datasets from authoritative biological science websites that cover a more extended time period to improve the representativeness of the data.

Data Availability Statement

The data used for research in this paper are available as open data via the Network Data Repository: https://networkrepository.com/. Example from: https://networkrepository.com/insecta-ant-colony1-day01.php

References

1. Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). Analyzing social networks. Sage.
2. Hölldobler, B., & Wilson, E. O. (1990). The ants. Harvard University Press.
3. Lusseau, D. (2003). The emergent properties of a dolphin social network. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(suppl_2), S186-S188.
4. Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., ... & Cheney, D. L. (2010). Strong and consistent social bonds enhance the longevity of female baboons. Current biology, 20(15), 1359-1361.
5. Archie, E. A., Moss, C. J., & Alberts, S. C. (2006). The ties that bind: genetic relatedness predicts the fission and fusion of social groups in wild African elephants. Proceedings of the Royal Society B: Biological Sciences, 273(1586), 513-522.
6. Ellis, S., Snyder-Mackler, N., Ruiz-Lambides, A., Platt, M. L., & Brent, L. J. (2019). Deconstructing sociality: the types of social connections that predict longevity in a group-living primate. Proceedings of the Royal Society B, 286(1917), 20191991.
7. Viégas, F. B., & Donath, J. (2004, November). Social network visualization: Can we go beyond the graph. In Workshop on social networks, CSCW (Vol. 4, pp. 6-10).
8. Faust, K. (2006). Comparing social networks: size, density, and local structure. Advances in Methodology and Statistics, 3(2), 185-216.
9. Freeman, L. C. (2002). Centrality in social networks: Conceptual clarification. Social network: critical concepts in sociology. Londres: Routledge, 1, 238-263.
10. Borgatti, S. P. (2005). Centrality and network flow. Social networks, 27(1), 55-71.
11. Rodrigues, F. A. (2019). Network centrality: an introduction. A mathematical modeling approach from nonlinear dynamics to complex systems, 177-196.
12. Newman, M. E. (2004). Analysis of weighted networks. Physical review E, 70(5), 056131.
13. Newman, M. E. (2006). Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23), 8577-8582.
14. Schuetz, P., & Caflisch, A. (2008). Efficient modularity optimization by multistep greedy algorithm and vertex mover refinement. Physical Review E, 77(4), 046112.
15. Fortunato, S., & Barthelemy, M. (2007). Resolution limit in

community detection. Proceedings of the national academy of sciences, 104(1), 36-41.
16. Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics), 28(1), 100-108.
17. Yang, X., Wang, Y., Wu, D., & Ma, A. (2010, November). K-means based clustering on mobile usage for social network analysis purpose. In 2010 6th international conference on advanced information management and service (IMS) (pp. 223-228). IEEE.
18. Chang, W., Ji, X., Liu, Y., Xiao, Y., Chen, B., Liu, H., & Zhou, S. (2020). Analysis of university students' behavior based on a fusion K-means clustering algorithm. Applied Sciences, 10(18), 6566.
19. Bianchini, M., Gori, M., & Scarselli, F. (2005). Inside pagerank. ACM Transactions on Internet Technology (TOIT), 5(1), 92-128.
20. Suri, S., Gupta, A., & Sharma, K. (2019, August). Comparative study of ranking algorithms. In 2019 International Conference on Computing, Electronics & Communications Engineering (iCCECE) (pp. 73-77). IEEE.
21. Wegmann, R., Murid, A. H. M., & Nasser, M. M. S. (2005). The Riemann–Hilbert problem and the generalized Neumann kernel. Journal of Computational and Applied Mathematics, 182(2), 388-415.
22. Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods.
23. Ertoz, L., Steinbach, M., & Kumar, V. (2002, April). A new shared nearest neighbor clustering algorithm and its applications. In Workshop on clustering high dimensional data and its applications at 2nd SIAM international conference on data mining (Vol. 8).
24. Samatova, N. F., Hendrix, W., Jenkins, J., Padmanabhan, K., & Chakraborty, A. (Eds.). (2013). Practical graph mining with R. CRC Press.
25. Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. Industrial psychiatry journal, 18(2), 127.
26. Pinter-Wollman, N., Hubler, J., Holley, J. A., Franks, N. R., & Dornhaus, A. (2012). How is activity distributed among and within tasks in Temnothorax ants?. Behavioral Ecology and Sociobiology, 66, 1407-1420.

Appendices

```
# Due to the page limit, the code only contains analysis for one community. The analysis for another community is very similar, except that the data is different.
library(readxl)
library(haven)
library(ggplot2)
library(networkR)
library(igraph)
library(summarytools)

# Import data and plot social network graphs
ant1 <- read_excel("D:/Master Study/CAN404/Resit Assessment/insecta-ant-colony1-day01/ant1.xlsx")
ant1_matrix <- as.matrix(ant1)
ant1_adjacency_matrix <-graph_from_data_frame(ant1_matrix)
deg_ant1 <- igraph::degree(ant1_adjacency_matrix, mode = "all")
plot(ant1_adjacency_matrix, vertex.size = deg_ant1 / 4, edge.arrow.size = .0015, vertex.color = rgb(0.1, 0.7, 0.8, 0.5))
# This (layout_randomly) works for another community (dataset) as a comparison:
plot(ant2_adjacency_matrix, vertex.size = deg_ant2 / 4, edge.arrow.size
   = .0015, edge.curved = .1, vertex.color = rgb(0.1, 0.7, 0.8, 0.5),
   vertex.frame.color = "#555555",
   vertex.label.color = "black",
   vertex.label.cex = .7)

#density and number of edges
graph.density(ant1_adjacency_matrix, loops = FALSE)
gsize(ant1_adjacency_matrix)

#degree centrality
degree_ant1 <- igraph::degree(ant1_adjacency_matrix)

# descriptive statistics for degree centrality
summarydegree_ant1 <- summarytools::descr(degree_ant1)
```

```
#degree distribution
degree_dis_ant1 <- degree_distribution(ant1_adjacency_matrix)
degree_dis_ant1 <- as.data.frame(degree_dis_ant1)
qplot(degree_dis_ant1, data = degree_dis_ant1, geom = "histogram", binwidth = .001)

#betweenness centrality
betweenness_ant1 = igraph::betweenness(ant1_adjacency_matrix)
summarybetweenness_ant1 = summarytools::descr(betweenness_ant1)
sort(betweenness_ant1, decreasing = TRUE)

#closeness centrality
closeness_ant1 <- igraph::closeness(ant1_adjacency_matrix, mode = "all")
summarycloseness_ant1 <- summarytools::descr(closeness_ant1)

#eigenvector centrality
EigenCentrality_ant1 <- eigen_centrality(ant1_adjacency_matrix)
EigenCentrality_ant1 <- as.data.frame(EigenCentrality_ant1)
summaryeigen_ant1 <- summarytools::descr(EigenCentrality_ant1$vector)

#community detection based on edge betweenness (NewmanGirvan)
ceb_ant1 <- cluster_edge_betweenness(ant1_adjacency_matrix)
plot(ceb_ant1, ant1_adjacency_matrix)
length(ceb_ant1) # number of communities
modularity(ceb_ant1) # how modular the graph partitioning is

#community detection based on greedy optimization of modularity
cfg_ant1 <- cluster_fast_greedy(as.undirected(ant1_adjacency_matrix))
plot(cfg_ant1, as.undirected(ant1_adjacency_matrix))
V(ant1_adjacency_matrix)$community <- cfg_ant1$membership
colrs <- adjustcolor(c("gray50", "tomato", "gold", "yellowgreen", alpha = 0.6))
plot(ant1_adjacency_matrix, vertex.color = colrs[V(ant1_adjacency_matrix)$community])

#K-means
# Optimal number of clusters and # Add subtitle "Elbow method"
fviz_nbclust(ant1_matrix, kmeans, method = "wss") + labs(subtitle = "Elbow method") # Add subtitle "Elbow method"

#PageRank
PR_ant1 <- page.rank(ant1_adjacency_matrix)$vector
summary_PR_ant1 <- summarytools::descr(PR_ant1)
sort(PR_ant1, decreasing = TRUE)

#Neumann_Kernel
Neumann_Kernel <- function(graph, gamma) {
  X <- as_adjacency_matrix(graph, sparse=F)
  my_K <- crossprod(t(X), X)
  K_inverse <- solve(diag(vcount(graph)) - gamma * my_K)
  K_hat <- crossprod(my_K, K_inverse)
  return (K_hat)
}
gamma_ant1 <- 1 / max(igraph::degree(ant1_adjacency_matrix))
NK_ant1 <- Neumann_Kernel(ant1_adjacency_matrix, gamma_ant1)
K_ant1 <- NK_ant1[ , c(1 : 113)]
d_ant1 <- diag(K_ant1)
sort(d_ant1, decreasing = TRUE)
```

```
#Shared Nearest Neighbor (SNN)
SNN_GRAPH <- function(graph, k) {
  snn_graph <-graph - E(graph)
  vertex_ids <- as.numeric(V(graph))
  for(u in vertex_ids[1:(length(vertex_ids) - 1)]) {
    for(v in vertex_ids[(u+1):length(vertex_ids)]) {
      counter <- 0
      for(m in vertex_ids) {
        if(are_adjacent(graph, v, m) & are_adjacent(graph, u, m)) {
          counter <- counter + 1
        }
      }
      if(counter >= k) {
        if(!are_adjacent(snn_graph, u, v)) {
          snn_graph <- snn_graph + edge(u, v, weight=counter)
        }
      }
    }
  }

  return (snn_graph)
}
SNN_ant1 <- SNN_GRAPH(ant1_graph, 60)
snn_ant1 <- as_data_frame(SNN_ant1)
snn_ant1
SNN_ant1 <- SNN_GRAPH(ant1_graph, 75)
snn_ant1 <- as_data_frame(SNN_ant1)
snn_ant1
SNN_ant2 <- SNN_GRAPH(ant2_graph, 80)
snn_ant2 <- as_data_frame(SNN_ant2)
snn_ant2
SNN_ant2 <- SNN_GRAPH(ant2_graph, 100)
snn_ant2 <- as_data_frame(SNN_ant2)
snn_ant2
```

CONFLICT OF INTEREST STATEMENT

The author declares that he has no conflict of interest.

FUNDING DECLARATION

The author did not receive support from any organization for the submitted work.