# Overall Graph Structure

**Outstanding items:**

* Define terms better, including explaining the significance of the metrics I’m using

A graph is a set of vertices and the links between them. Graph theory provides a generalized framework for representing the relationship between entities of various types, including, for example, ant foraging.

Ants operate without central control; the collective behavior of the colony is the result of the decisions of individual ants.

### Efficiency

Buhl *et. al.* concluded based on measurements of the foraging networks of *Formica aquilonia* ants that their transport networks show significant efficiency in terms of two key metrics: the total amount of trail (total length) and the average distance between the nest and foraging sites (route factor).

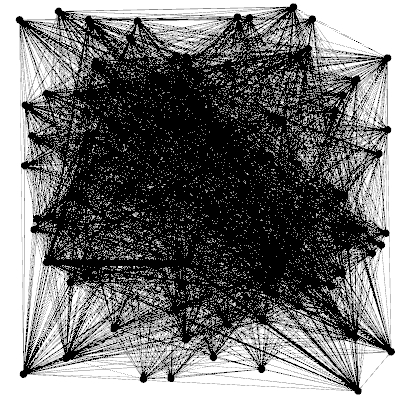
1. Buhl, J., Hicks, K., Miller, E. R., Persey, S., Alinvi, O., & Sumpter, D. J. (2008). Shape and efficiency of wood ant foraging networks. *Behavioral Ecology and Sociobiology,* *63*(3), 451-460. doi:10.1007/s00265-008-0680-7

## Behavioral Repertoire of ants

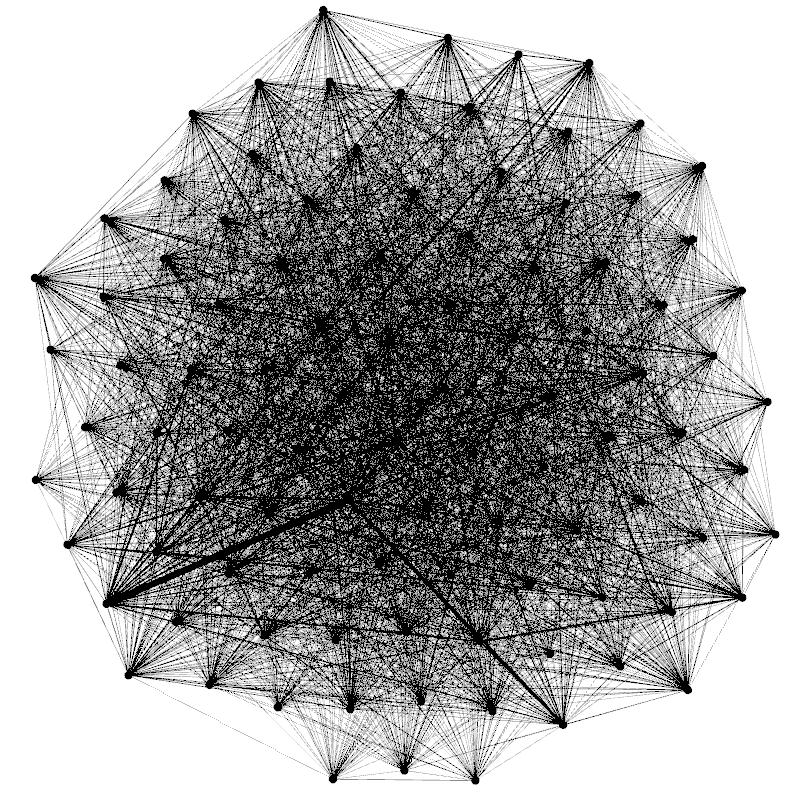
In previous sections, we have determined how the foraging behavior is influenced by local interactions.

## Overview of Network Topology

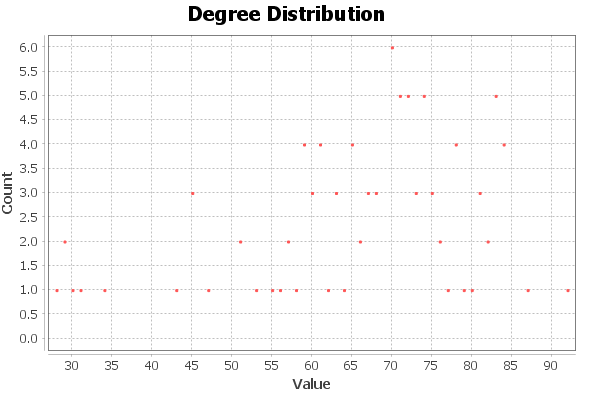
We have already established in previous sections that the number of interactions between ants, even if we restrict the interactions just to foragers or just to ants in the entrance chamber, is not normally distributed. We will look again at the Mersch data set, Colony 1, Day 15. This is a connected undirected graph with 99 nodes and 3301 edges. Gephi gives the following visualization of the graph prior to any layout algorithms being applied:



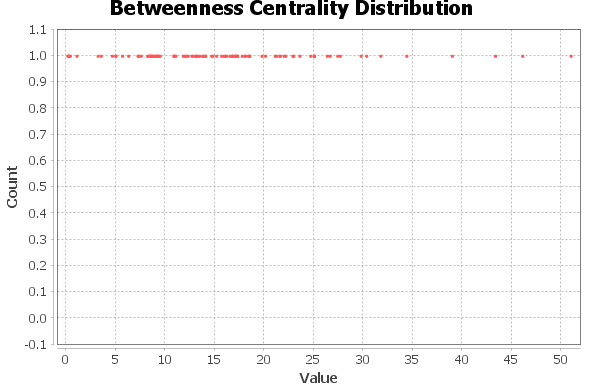
Using the Noverlap layout with a speed of 3, a margin of 150, and a ratio of 1.2 followed by 3 contractions with a scale factor of 0.8:

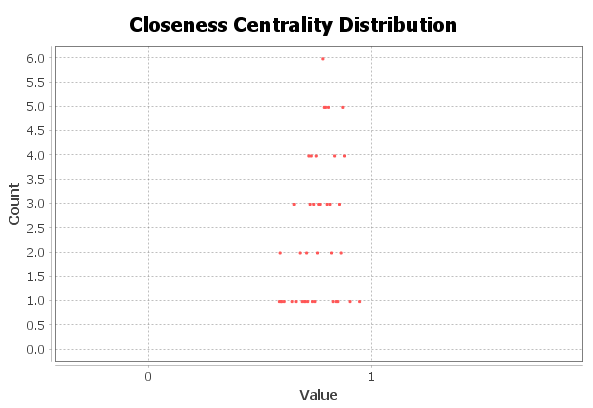


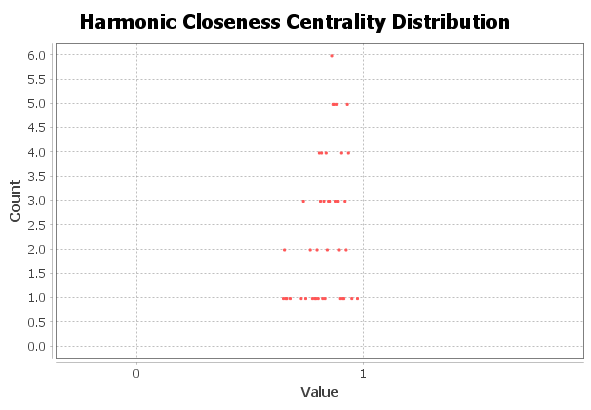
Gephi gives us the following degree distribution chart:

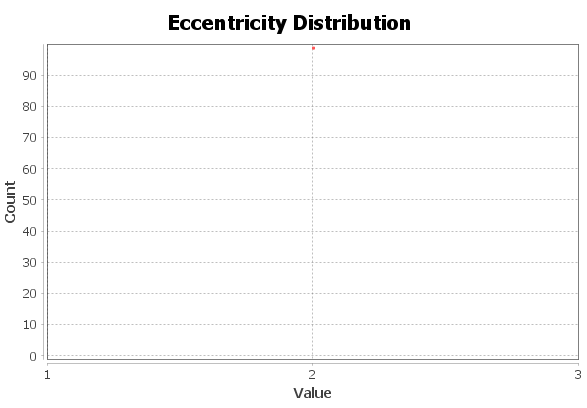


It has an average degree of 66.687, a network diameter of 2, an average path length of 1.3195217480931767, and the following distributions for centrality:

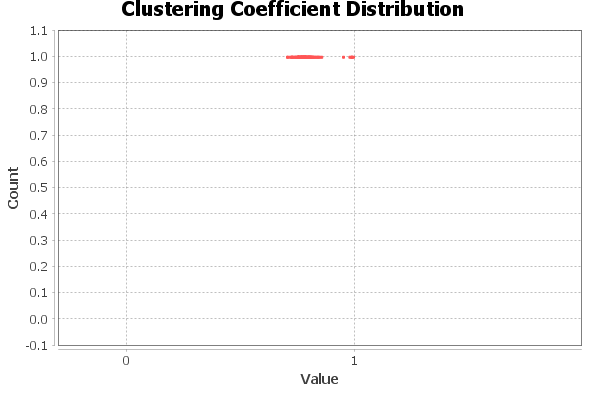








The graph density is 0.680, and the average clustering coefficient is 0.786. The clustering coefficient distribution is:



There are two algorithms of particular importance: PageRank and Hyperlink-Induced Topic Search (HITS).

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1-7), 107–117. https://doi.org/10.1016/s0169-7552(98)00110-x

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*(5), 604–632. https://doi.org/10.1145/324133.324140

HITS identifies two major kinds of web pages: authoritative pages that many other pages link to, and hubs that are not themselves authoritative but link to a large number of other sites. Interestingly enough, J. M. Kleinberg (the creator of the HITS algorithm) pointed out that HITS is not solely applicable to the World Wide Web; however, he identified that as a “particularly compelling” use case due the “problem of search and structural analysis”. In particular, he notes that “while individuals can impose order at an extremely local level, its global organization is utterly unplanned – high-level structure can emerge only through *a posteriori* analysis.” This suggests that this may also work for ants, which share the feature of individuals not planning the overall structure.

Prior research has demonstrated that the Hubness and Authority score, and in particular the ratio of Hubness to Authority (Sometimes called the H/A Ratio), can be very revealing to items’ overall role. Naïve Bayes and Random Forest classifiers based on these scores are particularly effective for this purpose.

1. Szczurek, P., & Horeni, M. (2018). Using link analysis algorithms to study the role of neurons in the worm connectome. *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*. doi:10.1109/aina.2018.00100

PageRank began as an algorithm to aid in ranking search results by estimating the popularity or importance of a page by estimating the probability that someone randomly surfing the web would end up at that particular page. As explained in *Dynamical Processes on Complex Networks*, PageRank “… defines the importance of each document by…. the probability that a random walker surfing the web will visit that document… The PageRank algorithm just gages the importance of each web page by the PageRank value which is the probability that a random walker surfing the web graph will visit the page .” The equation defining PageRank is as follows:

is the adjacency matrix of the graph. (The adjacency matrix for a simple graph – i.e. a graph with no self-loops – is simply a square matrix representing which vertices are adjacent to each other; for adjacency matrix , indicates that vertex and vertex are neighbors on the graph – i.e. that there’s an edge connecting them – and indicates that there is no edge connecting them).

We will again use data from Mersch *et. al*. (described in prior sections). In this instance, we will utilize data from all colonies on all days, calculating the Hub and Authority score, PageRank, and Betweenness Centrality.

We do encounter a problem in that our ant interaction data is undirected, meaning that hubness and authority are equal by definition (so the H/A ratio will always be 1).

1. Mersch, D. P., Crespi, A., & Keller, L. (2013). Tracking individuals shows spatial fidelity is a key regulator of ant social organization. *Science,* *340*(6136), 1090-1093. doi:10.1126/science.1234316

The fundamental idea behind

If we create an Orange package

