Interaction networks predict optimal harvester ant foraging behavior

by

Ethan Sudman

Dr. Piotr Szczurek, Advisor

Masters project submitted in partial fulfillment of the

requirements for the Master of Science in Data Science degree

in the College of Aviation, Science, and Technology of

Lewis University

*Abstract*

Network Science is an emerging interdisciplinary field that uses the tools of statistics, graph theory, and linear algebra to represent relational data. We will demonstrate that graph metrics like Hubness and Authority and common algorithms like PageRank can predict when ants become foragers. Additionally, it can predict the foraging behavior of ants and help understand baffling points about the optimality of foraging behavior.

*Index Terms*— Optimal Foraging Theory, network science

# INTRODUCTION

Ants are highly social insects that operate without central control or globally available information. Prior research by Gordon *et. al.* demonstrates that foraging behavior among ants and other social insects is heavily driven by interactions between foragers, dictating everything from where to look for food to what kinds of food to prefer (Gordon 2010). Foragers have a particular “smell” that is distinct from other casts of ants; additionally, social insects appear utilize scent cues left by other foragers to recognize which kinds of food to prefer (Jarau and Hrncir 2019).

Optimal Foraging Theory is predicated on the idea that evolutionary pressures will naturally drive organisms towards optimal foraging behavior over time. In an environment where there are multiple colonies (or even multiple species) competing for the same food supply, this will likely be User Equilibrium, whereas in cases where there was limited competition for a particular kind of food it could presumably be closer to Global Optimum. (Even in this case, however, we must consider the possibility that species evolved in such a way as to avoid competition with other species; in this case, what appeared to be a global optimum may, in fact, have evolved due to that being a User Equilibrium).

A major consideration for any model is that the reproductive unit is the colony as a whole, *not* the individual ant, and the reasons is the following: the vast majority of ants are sterile females. A Leaf-Cutter Ant colony can consist of up to 10 million sterile female workers, up to 5 fertile female queens, and some number of males who mate once and die shortly thereafter. Additionally, female ants are diploid (having two sets of chromosomes) and ants are haploid (having only one set of chromosomes); the result of this is that ants inherit 50% of their genes from their father and, on average, 25% of the genes of their mother. This results in ants sharing 75% of genes with their sisters and only 25% of their genes with their daughters (Gordon 2010). Thus, natural selection strongly prefers behaviors that are optimal for the colony as a whole, not just for individual ants.

# Optimal Foraging Behavior

The Red Harvester Ant (*Pogonomyrmex barbatus*) is native to the deserts in the southwestern United States. Their preferred food source is seeds, which is also their primary source of hydration (due to the moisture that occurs in the fat stored in the seeds). The seeds are typically carried to their location by wind or flooding. Foraging behavior is strongly influenced by both the ongoing need for food and water and by the need to minimize moisture loss brought on by the desert heat.

Until the recent introduction of invasive species like the Argentine Ants, the Red Harvester Ants did not face significant competition for food within their foraging range; ranges of individual colonies do not overlap very much (Adler and Gordon 2003), suggesting that ant colony ranges are in a state of User Equilibrium. The total size of a colony’s range is 25 – 30 square meters (Gordon 1995).

Ants tend to start their lives as nurses, progress to being nest cleaners (who tend to be spatially distributed throughout the colony), and foragers (who tend to spend a fair amount of time either outside the nest or in or around the nest chambers) (Mersch *et. al.* 2015). This progression serves the dual purpose of decreasing the risk of the spread of disease (because foragers do not typically perform other jobs within the colony after foraging) and increasing the average lifespan of the colony (because riskier jobs are performed by older ants).

The risk of disease transmission is a concern not only to the individual foragers, but also to the colony as a whole; for this reason, colonies appear to practice a form of “social distancing” with foragers (Quevillon *et. al.* 2015).

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). In fact, they found that they achieved efficiency on these metrics that were comparable to highway networks designed by humans, in spite of the fact that ants lack central planning or control (Buhl *et. al.* 2008).

While ant foraging behavior has numerous dependencies, such as the season (Jarau and Hrncir 2019), weather (Gordon 2013), and probability of encountering other competing ants (Adler and Gordon 2003), the probability of success appears to be a primary concern for how many foragers are active at any given time. For this reason, many species of ants, including Red Harvester Ants, rapidly increase their foraging activities in response to successfully returning foragers (Gordon 2010). This helps colonies to conserve resources because the number of active foragers is directly proportional to their probability of success. While a successful foraging trip’s caloric costs are less than 0.1% of its caloric reward (Ferral1988), unsuccessful foraging trips carry costs in calorie and water loss (which is a major concern in desert environments) as well as risk of predation and exposure to pathogens and other dangerous substances.

For species of ants that rely on transient food sources, the rapidity of the increase in the number of active foragers is itself adaptive. (Food sources can be transient either due to competition with other species or colonies or due to it being from a source that is likely to be moved or removed, such as a human picnic). This is where we encounter our first puzzle in ant foraging behavior: Red Harvester Ants don’t rely on transient food sources. Whether this is adaptive in some other way is an open question in optimal foraging theory (Gordon 2010). This question will be addressed in the sections on Ising Models and SIS models.

# Overview of Ant Interaction Network

Ant behavior is heavily determined by interactions with other ants (Gordon 2010). Danielle Mersch and several collaborators used an experimental apparatus to continuously track workers from six different colonies of the *Camponotus fellah* ant over a period of 41 days. Over nine million interactions were recorded, and four different behavioral repositories (Forager, Cleaner, Nurse, and Queen) were identified with ants progressing between them in a predictable order as they age (with the exception of Queen, which obviously does not change with age). Interaction within groups was much higher than interaction between groups (Mersch *et. al.* 2013).

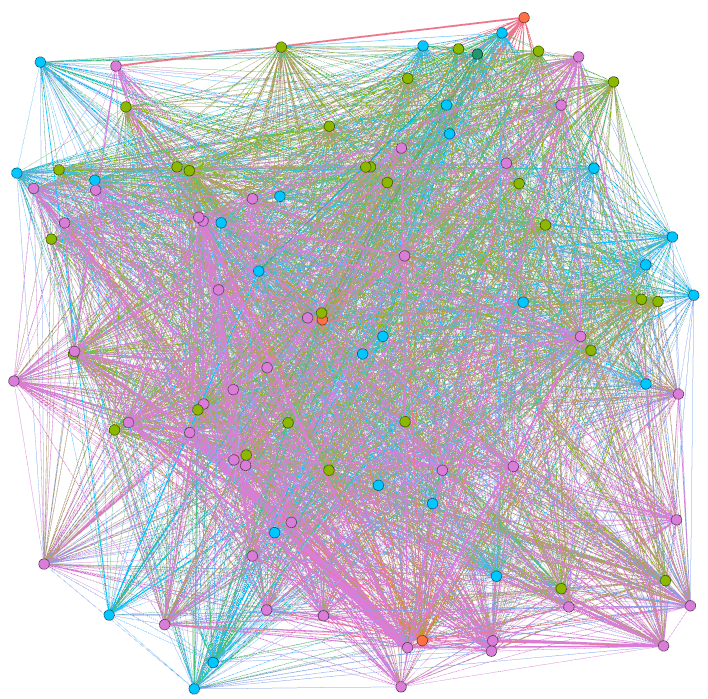
This kind of data readily lends itself to representation by a graph. A graph is simply a way of representing the relationship between different entities; a graph consists of a set of edges and a set of vertices , typically written as . A vertex is an entity, and the presence or absence of an edge between two vertices implies the existence or absence of an interaction or relationship of some kind. In this case, each node represents an ant, and each edge represents an interaction between two ants. The degree of a vertex is the number of times that that particular ant interacted with other ants.

Once ant interaction networks are represented as graphs, their overall interactions can be readily visualized (Figure 1). Graphs do not imply a particular geometry (and there’s not even a guarantee that the entities involved will readily lend themselves to a geometric interpretation), leading to significant flexibility as to how they are visualized. There is no particular mathematical constraint on how a graph can be drawn, as long as all vertices and edges are accurately represented.

In fact, the formal definition of graph equivalence does not refer to any properties of the graph drawing, labels, or even the specific type of entity or relationship involved:

Two graphs are said to be *isomorphic* if there exist between their vertex sets a one-to-one correspondence having the property that whenever two vertices are adjacent in either graph, the corresponding two vertices are adjacent in the other graph. Such a one-to-one correspondence is called an *isomorphism*. If G and H are isomorphic graphs we denote this by writing “”. (Trudeau 2015)

In other words, two graphs are equivalent (isomorphic) if they have the same number of vertices and the vertices are “hooked up” in the same way.



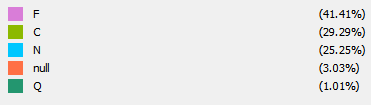


Figure Interaction network for ant colony 1, day 15 (Mersch et. al. 2013). Generated using the Gephi graph visualization package.

We can get a further sense of the interactions by visualizing the interactions of an arbitrary Cleaner (Figure 2).

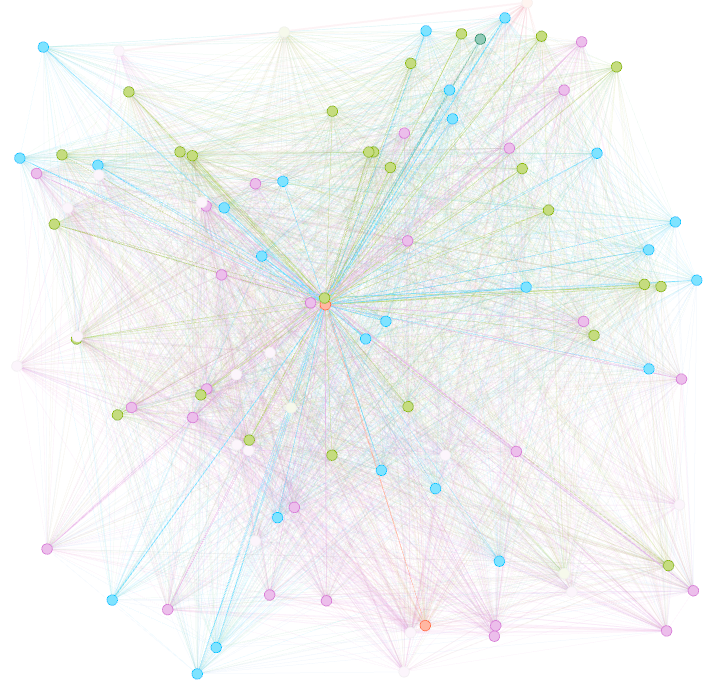


Figure Interactions of an arbitrary Cleaner

An arbitrary Nurse was selected in Figure 3.

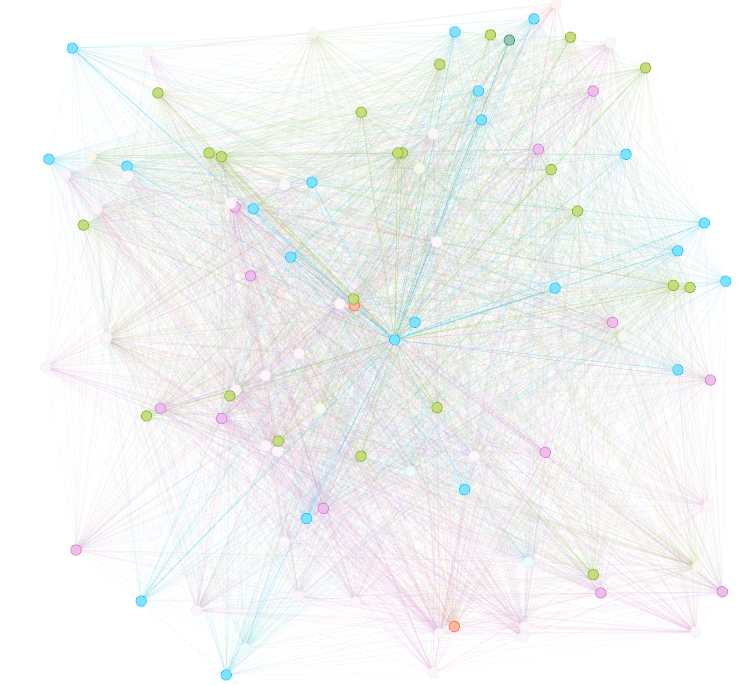


Figure Interactions of an arbitrary Nurse

An arbitrary forager was selected in Figure 4.

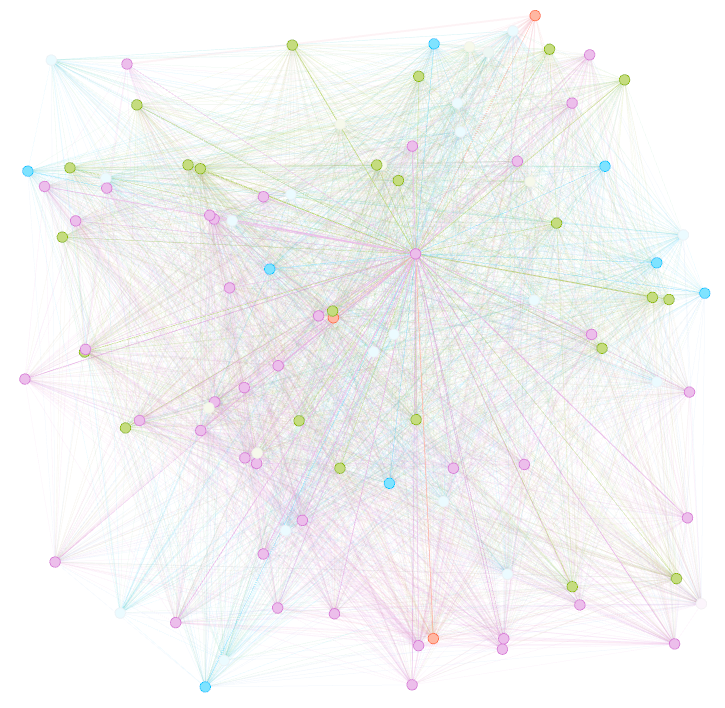


Figure Interactions of an arbitrary Forager

The graph approach allows us to quantify how influential or important each ant is within a colony. The most straightforward measure of an ant’s importance is its degree; the loss of an ant with a high degree would presumably tend to have a disproportionate effect on the way that the colony as a whole was interacting and could even affect the speed of information flow throughout the network. Another key metric is what’s known as Betweenness Centrality. The intuition behind this metric is that, if a vertex lies on many shortest paths between other vertices, then it is likely to be important (because it enables the rapid spread of information through the network). Given a connected graph , The Betweenness Centrality of vertex is given by

where is the set of vertices of graph , , is the set of shortest paths between the vertices and , and is the set of shortest paths between and that include . A connected graph is a graph in which there exists a path between all vertices in the graph (Steen 2010).

A subgraph is merely some graph H such that such that with , , where indicates the presence of an edge between vertices and (Steen 2010).

A connected component H is a is a maximal connected subgraph G – i.e. (where is a path between and ), subject to the constraint that H is not contained in a larger connected subset of G (i.e. ) (Steen 2010).

The number of connected components in the graph is denoted . A graph is connected if, and only if, (i.e. the entire graph consists of a single connected component); otherwise, it must be the case that such that there is not a path between and . NetworkX will give us the number of connected components in a graph by running sum(1 for component in nx.connected\_components(G)). Looping over every file in the Mersch data, we can verify that for every graph , so the graphs are indeed connected.

Betweenness Centrality for our network is distributed as follows:

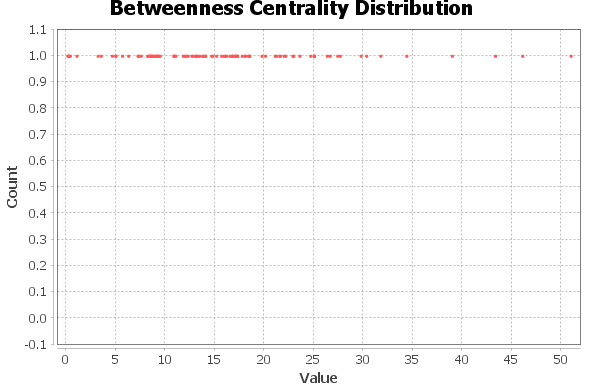


Figure Betweenness Centrality distribution for colony 1, day 15 of Mersch et. al. 2013

Betweenness Centrality is closely related to a network’s community structure. A community is a subset of nodes that are more densely connected to each other than they are to others. Networks can have multiple communities, some of which could be overlapping. Communities are sometimes detected using hierarchical clustering. Another popular algorithm (which is implemented in the Python network science library NetworkX) is the Girvan-Newman Algorithm, which generalizes betweenness centrality to edges and exposes communities by iteratively removing the edges with the highest betweenness centrality (Girvan and Newman 2002).



Figure Normal curve, histogram, and summary statistics for the total number of communities per day according to the Girvan-Newman Method

The Kurtosis gives a strong indication that this data is not normally distributed (since data that was “perfectly” normal would have a Skewness of 0 and a Kurtosis of 3). We can confirm that this data is, indeed, not normally distributed by the Goodness of Fit tests:

| **Goodness-of-Fit Tests for Normal Distribution** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Kolmogorov-Smirnov** | **D** | 0.11205721 | **Pr > D** | <0.010 |
| **Cramer-von Mises** | **W-Sq** | 0.20871995 | **Pr > W-Sq** | <0.005 |
| **Anderson-Darling** | **A-Sq** | 1.58786002 | **Pr > A-Sq** | <0.005 |

Since for all tests, we reject the null hypothesis that the data is normally distributed. Tests for fit of the Gamma, Exponential, Weibull, and Lognormal distributions similarly have , so we conclude that the data does not fit those distributions either.

The degree distribution of a graph is frequently written as (where P is the a probability and is the degree of a node) – in other words, the probability that a randomly selected graph will have (degree) . The closely related average degree of a graph, often written as , can simply be defined as

In this case, we will represent this with a histogram of different degrees. The degree distribution of the graph representing ant interactions is highly heterogeneous. Based on an analysis of the number of interactions of Colony 1, Day 10 from [7], we receive the following histogram:

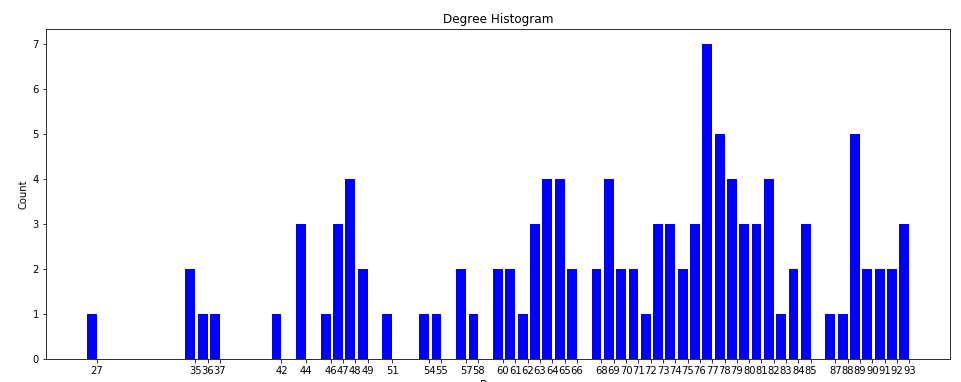


Figure Number of interactions between ants, Colony 1, Day 10 (Mersch et. al. 2013)

For comparison, Colony 1, Day 15 has the following degree distribution:

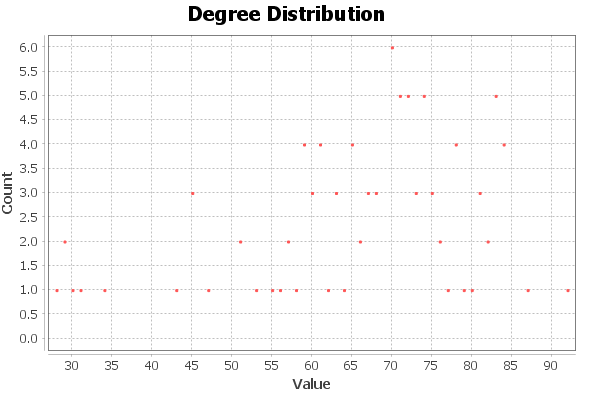


Figure Degree distribution (distribution of the number of interactions) of Colony 1, Day 15 (Mersch 2013)

This is clearly far from a uniform (or even normal) distribution of degrees. Additionally, looking at the betweenness centrality for individual ants (from the same data set):

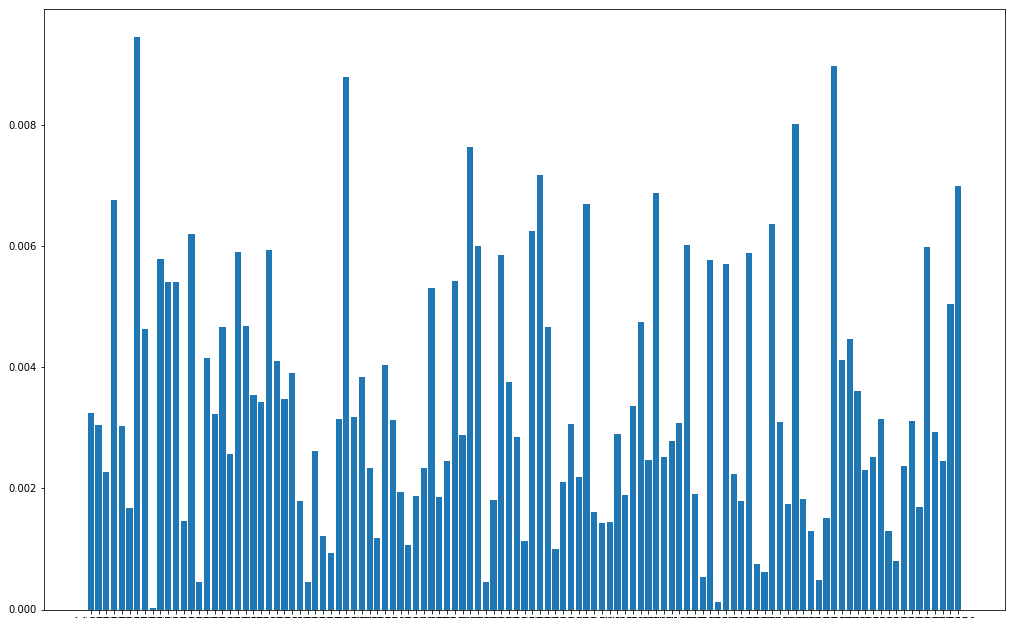


Figure Betweenness Centrality for each individual ant. Note that this is **not** a histogram, but rather represents the actual Betweenness Centrality value for individual ants.

Sorting by value,

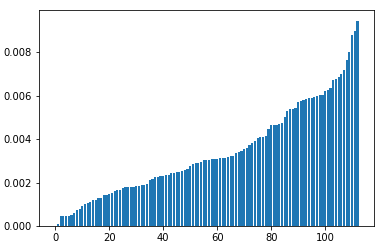


Figure Betweenness Centrality of individual ants, sorted by size, suggesting a power-law distribution

Note that neither of these graphs are specific to interactions with other foragers. In order to get this, we check the attributes of the graph and remove all ants that are not foragers. This allows us to get the following histogram:

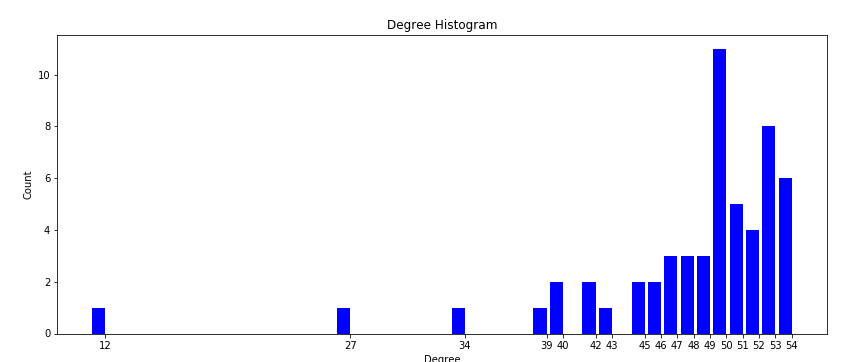


Figure Degree histogram for ants that were foragers in period 1, colony 1, day 15 (Mersch et. al. 2013)

Breaking this down by percent rather than count, we get the following histogram:

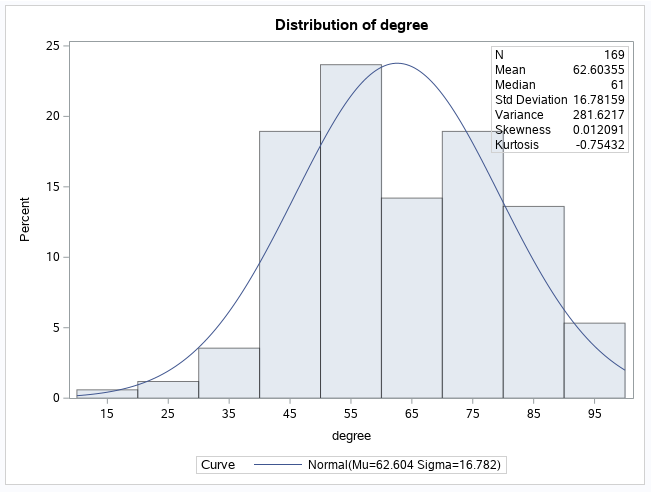


Figure Degree distribution for interactions among ants that were foragers in period 1, colony 1, day 15 (Mersch 2013)

The Probability Plot is as follows:

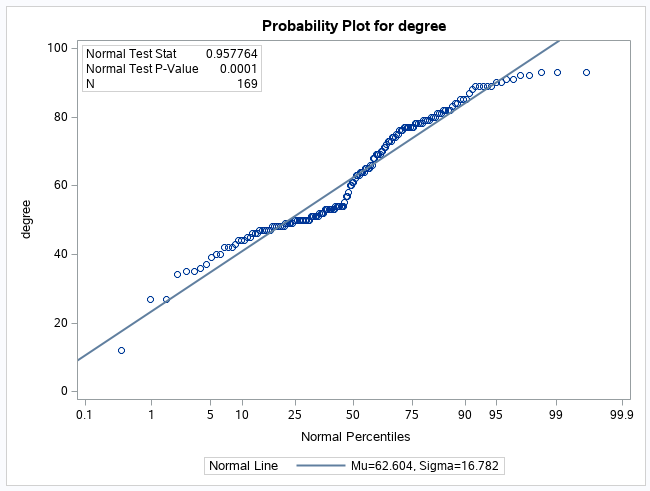


Figure Probability plot for degrees of ants that were foragers in period 1, colony 1, day 15 (Mersch 2013)

SAS gives us the following value for Goodness-of-Fit Tests for the Normal Distribution:

**Fitted Normal Distribution for degree**

| **Goodness-of-Fit Tests for Normal Distribution** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Kolmogorov-Smirnov** | **D** | 0.15153301 | **Pr > D** | <0.010 |
| **Cramer-von Mises** | **W-Sq** | 0.54601769 | **Pr > W-Sq** | <0.005 |
| **Anderson-Darling** | **A-Sq** | 2.95194971 | **Pr > A-Sq** | <0.005 |

This suggests that the data is not, in fact, normal (since p < 0.05 and the null hypothesis is that the data is normally distributed), but it’s not nearly as far off from normal as the previous data set. In particular, in the normal distribution, skewness is approximately 0, which is the case here; however, kurtosis is 3 in the normal distribution (as opposed to -0.75 in this case). Thus, this has more “tailedness” than a strictly normal distribution.

Additionally, the average clustering is approximately 0.9152 and the average shortest path length is approximately 1.1253 with a standard deviation of approximately 0.3249. By way of comparison, a Watts-Strogatz Graph with 163 nodes, mean of 61, and a p of 0.5 has an average shortest path length of approximately 1.62963 (with a standard deviation of 0.4973) and average clustering coefficient of 0.3798. A t-test (calculated using SciKit-Learn) shows a p-value of approximately , showing that there is a statistically significant difference between the average path lengths of the two graphs. (Similar computations were not performed for clustering coefficient). The fact that the clustering coefficient for our graph is higher than the clustering coefficient of a Watts-Strogatz Graph (which has been shown to have the small-world property) and the average path lengths are statistically significantly lower, this suggests that our network also has very strong small-worldedness.

Full code for these calculations can be found in the Jupyter Notebook in the supplemental material.

# Active Forager Rate – Statistical Mechanics Approach

Recently, methods developed for applications within statistical physics have found additional applications for everything from epidemic modeling to protein networks (Barrat *et. al.* 2013). These models consider the macro behavior of the system as a function of the states of the constituent parts (sometimes known as the microscopic behavior of the system).

A phase is a certain “macro state” of the system which behaves in some way that is significantly different than other phases. A familiar example is how water can exist in 3 phases: liquid, solid (ice), and gas (steam). Which phase the system is dictated by the parameters of the system. For example, water’s phase is determined both by the air pressure and the temperature. At 1 ATM, water freezes (transitions to its solid phase) at exactly 0 degrees Celsius, transitions to its gas phase at precisely 100 degrees Celsius, and is in its liquid phrase when the temperature is strictly between those two temperatures.

In the case of the model of ferromagnetic phase transitions, paramagnetic materials like nickel and iron exhibit differing magnetic behavior depending on their temperature. The specific temperature at which the phase transition occur is called the critical temperature or the Curie Temperature and often symbolized as . The atom as a whole carries a “spin” that is the result of the behavior of its electrons. This is typically referred to as either a positive (+) or negative (-) spin.

Furthermore, there is an equilibrium which, given an arbitrarily long amount of time with no external disruption of the system, the system will eventually reach. The stationary distribution is sometimes symbolized as , where

where is the temperature, is the Boltzmann Factor that provides the correct dimensional unit, is the Hamiltonian (which associates each state with its energy, and is the normalization factor supplied by

(i.e. the sum of the Hamiltonian of every possible state), subject to the constraint that .

The Hamiltonian is defined by

Where represents the energy reduction if spins are aligned and is either (if and are neighbors on the graph or lattice) and otherwise.

Each node is arranged in either a graph or a lattice of some type (typically rectangular or cubic) and is allowed to interact with a certain number of neighbors. If all nodes are influenced by all other nodes equally, we can apply Mean-Field Theory to assist in deriving an analytic solution. Mean-Field theory relies on the homogeneous assumption, which means that there are no relevant correlations between microstate variables; no individual element has any “unique” properties, and all interactions “felt” at any given node are the result of the average interaction of the system as a whole (Barrat *et. al.* 2013).

Analogously to Markov processes, we can define a transition rate between different configurations of the system. By way of example, we can determine specific transition rate that drives the system towards the equilibrium distribution by applying the following constraint:

where refers to the transition rate from configuration to configuration and refers to the opposite transition rate. In other words, a system is in equilibrium if it transitions to one state at the same rate that it transitions back. Each pair of terms in the Master Equation (see below) has a null contribution. (This does not imply that the system is completely “still” – it simply means that the system no longer changes its average behavior). Note that this is a sufficient condition for the system to be in a stationary state, and some systems may have other stationary states (Barrat *et. al.* 2013).

The behavior of the system as a whole is determined by the “Master Equation,” which consists of the evolution equation for (which is the probability of finding the system in configuration at time ):

Ising Models (and Ising-inspired models) have been successfully applied to everything from epidemic models (Barrat *et. al.* 2013) to explaining how ants coordinate to move heavy loads (Gelblum *et. al.* 2015).

In our Ising Model, a positive “spin” will correspond to a forager that is actively seeking food and a negative spin will correspond to a quiescent forager.

The vast majority of Red Harvester Ants forage until they are successful. While J. H. Fewell found success rates between 50% and 75% for some species of ants (Fewell 1988), the rate for Red Harvester Ants has been found to be approximately 90% (Gordon 1991, Schafer and Gordon 2006).

Ising Models were originally developed for spins that were arranged in a lattice structure. They have since been adapted for graphs. As previously described, the distribution of degrees of ant interactions is not uniform (or even normal), even if we only consider interactions between foragers. Thus, a lattice is unlikely to be an adequate model. It does, however, show signs of being a small-world network.

Given a temporal network, a quiescent forager will become active when it has a minimum of two neighbors in a 20-second period who have successfully returned from foraging – in other words, a vertex with a negative spin will change to having a positive spin if, and only if, two or more of its neighbors also have a positive spin. (The number of neighbors that have a negative spin makes no difference as to when the transition occurs). We can also assume that successful foragers who recently encountered a quiescent forager will soon put down the food, since quiescent foragers wait inside the nest to minimize water loss. (They may, however, encounter multiple quiescent foragers before they do so).

As previously demonstrated, ant colonies actively take measures to ensure the health of the colony as a whole, including taking measures to minimize energy expenditure. Ants may actively recruit other ants to assist in tasks; for example, foragers will attempt to recruit more foragers to carry heavy loads (Gelblum *et. al.* 2015). It is reasonable, then, to consider overall energy expenditure in our models of ant foraging behavior. A major feature of Ising Models is the fact that when spins “align,” the overall energy state of the system is lower. This corresponds to a larger number of ants “aligning” to perform the same task (i.e. either both being quiescent or both being active). This is highly plausible in light of the fact that ants appear to actively coordinate their activities in at least some cases.

The rate of active foragers transitioning to quiescent foragers will be the total time of the trip outside the nest, however long it takes them to find the food, and some number of seconds inside the nest. As previously stated, the foraging range of Red Harvester Ants is, on average, 25 – 30 square meters (Gordon 1995).

In terms of speed of foragers, two major factors are food size and temperature, with larger food size and temperature both increasing total time. (Body size, however, does *not* have a statistically significant effect on forager speed). On average, though, red harvester ants (*P. barbatus*) trips took an average of 13.1 minutes, with the outbound trip taking 2.3 minutes on average, the inbound trip taking 2.6 minutes on average, and the food search taking 8.4 minutes on average. The distance of the inbound trip was, on average, 5.4 meters (Morehead and Feener 1998).

The energy saving is a function of the following: water loss rate due to sun and energy expenditure during foraging. Consider a simple colony that consists of two foragers. If both foragers have a negative spin (i.e. are quiescent), they have significantly diminished water loss rate and energy expenditure (but that will never be 0 because all organisms consume water and calories merely in the course of the day). However, they will not retrieve any food. This would occur if, for example, they “gave up” on finding food. On the other hand, if one is active and the other is not (i.e. one has a positive spin and the other has a negative spin – their spins are *not* aligned), that presumably means that they have not successfully brought food back and are actively looking for it, which incurs the “full” energy penalty for the duration of their unsuccessful foraging. If both are foraging (i.e. they both have a positive spin), then they will get food at a fraction of the cost of both wondering around aimlessly. It will take merely the cost of the round-trip with food, without any of the “wandering around.” For simplicity, let us assume that the cost of this state is 0.25 the cost of the unaligned state:

assuming that and are both positive.

This introduces a problem: if the state where both spins are positive is higher energy than a state where both are negative, then wouldn’t the system tend towards the negative spins in the thermodynamic limit? In this case, our system would predict that, if all ants temporarily stopped foraging for some reason, there would be no way for them to start foraging again. This is decidedly *not* true.

One solution is that successful introduces energy from outside. Indeed, in an Ising Model in statistical physics, a system can be attached to an external heat bath.

If we wanted to calculate the equilibrium, we would need to find a way to compute the temperature of the system. Recall that the formula for the equilibrium is:

It is not immediately obvious what the temperature of the colony means. However, the fact that the unit of this is joules provides a significant clue: there *are* ways to provide a meaningful energy state in joules for the foragers. The most logical value for this is the total approximate energy expenditure of the colony in joules. (We could also use the energy of their food stores in joules, or the net energy of the colony by subtracting the energy expenditure of the colony from the total energy stores).

The net energy of foragers at any given time is:

subject to the constraint that . We have already calculated that outgoing trips take 2.3 minutes on average, searching takes 8.4 minutes, and inbound trips take 2.6 minutes for a total of 13.3. This is fairly similar () to a finding described below in which a related species of ant foraged for 15.15 minutes on average in an experimental apparatus.

Note that the result for will vary by colony. In particular, foragers appear to be less active in larger colonies than in smaller colonies, which results in energy consumption of colonies scaling sublinearly with size (Ferral *et. al.* 2017).

For the *Pogonomyrmex occidentalis* harvester ant, energetic costs for foraging ants are less than 0.1% of the caloric rewards of harvested seeds. Furthermore, ants appear to prefer to maximize time efficiency rather than energy efficiency due to high food to energy ratio. Fewell 1988 obtained metabolic rate data by measuring ants’ oxygen consumption in a sealed glass environment that was set up to be free of carbon dioxide. Inactive ants had an average metabolic rate of 1.215 ml , vs. 1.846 ml for walking ants, suggesting that inactive ants consume as much energy as foraging ants. Ants walked at an average rate of 1.11 m/min in the experimental apparatus, as opposed to 1.35 m/min in field measurements (Fewell 1988).

Additionally, the study found that 59.4% of the trip distance was travel to the food source and 40.6% was made while returning with the food. The mean weight of the items was 8.07 mg, which was 120% of the mean weight of the ants (6.02 mg). The weight increased metabolic rate by 1.62 times in another study cited by the author, which was not re-tested for this study.

An additional 1.45 minutes was spent handling food items before returning, which was not directly measured but was assumed to be equal to the measured oxygen consumption during seed handling.

The mean energy cost per foraging trip was estimated to be 0.088 joules.

In the field, the number of seeds collected varied by whether the routes they followed were cleared (an average of 18.31 seeds per ant per day, maximum 92) or not (an average of 9.69 seeds per ant per day, maximum 36).

Seeds provide an energy content of 92.9 J (Kelrick and MacMahon 1985).

Ants in the apparatus walked for an average of 15.15 minutes and were inactive for an average of 6.06 minutes, indicating that ants on average spent of their time foraging and of their time inactive.

The success probability in (Fewell 1988) was 75% for ants on lower cover routes and 50% for ants on vegetated rates. However, for Red Harvester Ants, (Gordon 1991) has found that Red Harvester Ant foragers tend to forage until they find food, with 90% of them foraging until they are successful.

Knowing these facts, we can rewrite the previously-stated equation to simply reflect the proportion of the time that they forage. On Sunday, June 6, 2021, sunrise in Phoenix, AZ was 5:18 AM and sunset was 7:35 PM for a total day length of 14:17:43 [32]. Converting to seconds, we have

in a day. A “cycle” of rest and inactivity is

This means that an ant could, on average, complete

trips per day. The total energy expenditure for each forager, then, is:

For a colony of 2000 foragers, this is approximately 6422.41 joules. (Note that I did *not* include the energy expenditure of other types of ants, such as queens and nurses).

Thus, we will find that for the entire day (since the Boltzmann Constant, which is , serves primarily to relate the average kinetic activity of particles of gas to the thermodynamic temperature) [33].

Next, recall that

Next, we need to calculate Z, which is a normalization factor that sums over every possible state:

Note that we have used the energies for the individual ants for an individual trip (rather than the energy values for the entire colony for a day). Thus, for our hypothetical two-ant system, the probability that both of them will be inactive foragers in equilibrium is:

The probability of them both being active foragers is:

We can easily verify that , which is to be expected. This immediately introduces a problem: this appears to be “inverted” – we had observed that foragers are active 71.43% of the time and inactive 28.57% of the time. In fact, these appear to be exactly “flipped”. It turns out that the thermodynamic equilibrium, of course, prefers the *lower* energy solution (whereas we have set the foragers to be the higher energy state). If we simply reverse these numbers, the model becomes fairly close to the actual proportions of foragers and non-foragers that were observed experimentally. This suggests that harvester ant foraging is, in fact, close to a mathematical equilibrium.

To see what this can tell us about why Red Harvester Ants might “swarm” over food, recall that a high rate of successful return signals that foragers are likely to be successful. We have stated above that an ant searches, on average, for 8.4 minutes for food. Suppose that a large number of successful ants signals readily available food – in fact, let us suppose that it’s so readily available that it would only take ants 30 seconds to find food instead of 8.4 minutes. Then the total time is the outbound trip + the lower search time + the inbound trip, which results in a foraging trip having a cost of

We will recompute the total cost for a day of foraging at this rate:

And the Hamilton:

We further note that the total energy consumption for 2000 foragers is 2600 J (2064 of which is while they’re active and the rest of which is while they’re inactive). This is obviously a *significant* cost saving.

We will recompute at

Thus, the equilibrium for state 1 becomes:

We can compute as:

As described above, we need to “reverse” these and find that 78.63% of the foragers are active and 21.37% of the foragers are inactive. Thus, this model predicts that it *is*, in fact, globally optimal for ants to “swarm” over food because it results in an improved energy profile for the colony as a whole.

# Transition Rate between Active and Inactive Foragers

Having established the equilibrium mechanics, we will now turn our attention to the transition rates. There are two rates that we must fix: the rate of transition from inactive forager to active forager, and from active forager to inactive forager.

Based on these two rates, we can establish the rate of change in the probability that our system will be in a given state at a given point. The rate of change of the probability of the system being in a state at time is given by

We have already determined that the probability of a forager being active at any given time is 71.43% and the probability of them being inactive is 28.57%.

The transition rate is a matter of some interest, and the reason is the following: we have already established that this will not be a constant. The two items of particular interest are the following: first, the distribution of ant interactions is not uniform. Secondly, the rate at which harvester ants leave the nest is proportional to the number of ants that returned successfully.

(Prabhakar *et. al.* 2012)suggested that the number of departures in the *n*th time slot can be described as a Poisson random variable of mean , where

with . is the number departing at time , and is the number arriving at time . is the rate of outgoing foragers, which increases by when a successful forager returns. decreases by for each outgoing forager. decays by during each time slot. They set based on field observations (0.15 to 1.2 ants per second) and a ants per second. They set for their simulation but indicated that other studies suggest that this parameter may be important. They swept between 0.01 and 0.25 for *c*.

The supplemental materials contains Matlab code that implements the above equation for several fixed values of . The values are the following:

| **Level of An** | **N** | **Dn** | |
| --- | --- | --- | --- |
| **Mean** | **Std Dev** |
| **4** | 15 | 15.6666667 | 4.5773771 |
| **5** | 15 | 16.1333333 | 4.1034248 |
| **6** | 15 | 21.2666667 | 5.5737480 |
| **7** | 15 | 26.8000000 | 9.8720384 |
| **8** | 15 | 27.9333333 | 7.3627117 |
| **9** | 15 | 29.8666667 | 9.2339952 |
| **10** | 15 | 33.8000000 | 11.1046966 |

Levene’s Test is as follows:

| **Levene's Test for Homogeneity of Dn Variance ANOVA of Squared Deviations from Group Means** | | | | |
| --- | --- | --- | --- | --- |
| **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| 134543 | 22423.8 | 2.88 | 0.0126 |
| 763527 | 7791.1 |  |  |

Welch’s ANOVA is as follows:

| **Welch's ANOVA for Dn** | | | |
| --- | --- | --- | --- |
| **DF** | **F Value** | **Pr > F** |
| 6.0000 | 14.26 | <.0001 |
| 42.9757 |  |  |

The distribution of is as follows:



Figure Distribution of the simulated departure rate of foragers

Perhaps unsurprisingly, then, in the model, is, in fact, having an effect on (which is what we would expect, given that the model was explicitly designed for that purpose). Perhaps more relevantly for our purposes, this gives us a good sense of the relationship between and . This brings us to another issue: the means for are significantly higher than field observations suggest. Fortunately, we now have the confidence to realize that we can target our to the observed . Less fortunately, this exposes a weakness in our Matlab script: the **poissrnd** method will only return natural numbers (because it represents the number of times that the event in question happens; an event could happen, for example, 0 or 2 or 3 times in a particular time period, but it makes little sense to say that it happened 0.3 times). We could solve this in one of several ways:

* Only accept parameters that result in (since
* “Cheat” and accept (since
* While holding the other parameters equal, run the program several times where (i.e. run the simulation several times and select the results where is 1, 2, or 3, subject to the constraint that the average of all of the results for fall between 0.15 and 1.3 (inclusive) when we stop.

Our Matlab script follows the first result: only accept results where . (Note again that because ).

Obviously, there is a degree of randomness to this computation. Here are a few examples of average number of arriving ants for various values of *c* (computed using Excel’s subtotal/average function):

|  |  |  |  |
| --- | --- | --- | --- |
| c | An | an | d |
| 0.24 | 0.2 | 1.05 | 1 |
| **0.24 Average** | 0.2 | 1.05 |  |
| 0.23 | 0.1 | 0.3 | 1 |
| 0.23 | 0.3 | 1.2 | 1 |
| 0.23 | 0.4 | 1.8 | 1 |
| **0.23 Average** | 0.266667 | 1.1 |  |
| 0.22 | 0.2 | 0.85 | 1 |
| 0.22 | 0.6 | 3.1 | 1 |
| **0.22 Average** | 0.4 | 1.975 |  |
| 0.21 | 0.2 | 1 | 1 |
| 0.21 | 0.4 | 1.55 | 1 |
| 0.21 | 0.5 | 2.2 | 1 |
| **0.21 Average** | 0.366667 | 1.583333 |  |
| 0.2 | 0.2 | 0.8 | 1 |
| 0.2 | 0.4 | 1.45 | 1 |
| **0.2 Average** | 0.3 | 1.125 |  |
| 0.19 | 0.2 | 0.75 | 1 |
| 0.19 | 0.3 | 1.25 | 1 |
| 0.19 | 0.4 | 1.6 | 1 |
| 0.19 | 0.6 | 1.95 | 1 |
| **0.19 Average** | 0.375 | 1.3875 |  |

Unfortunately, Gordon *et. al.* did not publish their “raw” data in this instance, rendering it difficult to determine for certain which value of *c* to use. We will, therefore, reproduce our computation values for *c* and the corresponding and :

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0.24 Average** | | | 0.2 | | | 1.05 | | |
| **0.23 Average** | | | 0.266667 | | | 1.1 | | |
| **0.22 Average** | | | 0.4 | | | 1.975 | | |
| **0.21 Average** | | | 0.366667 | | | 1.583333 | | |
| **0.2 Average** | | | 0.3 | | | 1.125 | | |
| **0.19 Average** | | | 0.375 | | | 1.3875 | | |
| **0.18 Average** | | 0.375 | | | 1.3125 | | |  | |
| **0.17 Average** | 0.425 | | | 1.52725 | | |
| **0.16 Average** | 0.65 | | | 2.083333 | | |
| **0.15 Average** | 1 | | | 2.95 | | |
| **0.14 Average** | 0.55 | | | 1.45 | | |
| **0.13 Average** | 1.08 | | | 2.84 | | |
| **0.12 Average** | 1.1 | | | 2.7 | | |
| **0.11 Average** | 0.925 | | | 2.025 | | |
| **0.1 Average** | 1.15 | | | 2.175 | | |
| **Grand Average** | 0.689091 | | | 1.894709 | | |

We will simply utilize the average value for (0.689091). Also, using Excel’s sort functionality, we can determine that the range for the “raw” values of fall between 0.1 and 2.4. However, since the second-largest value is 1.8, we will test to see if this is an outlier. Using **=QUARTILE(A2:A71, 3)** and **=QUARTILE(A2:A71, 1)** we can determine that the third quartile is 1 and the first quartile is 0.4, giving us an interquartile range of 0.6. We will then calculate that and . (Tukey 1977, Hoaglin 1986). This means that 2.4 is, in fact, an outlier but 1.8 and 0.1 are not. Therefore, we will expect the arrival rate to fall between 0.1 and 1.8 with an average of 0.689091.

The values for (which is the average departure rate) unfortunately do not match the data quite as well, with a range of 0.3 to 4.7 (vs. 0.15 to 1.2 in the field). A procedure similar to the above confirmed that there are no outliers in the data. Unfortunately, this suggests that this parameter should not be used. If we modify our Matlab script to allow only , has an average of 0.679592, a range of 0.17 to 1.2, an interquartile range of 0.5, and no outliers. has a range of 0.1 to 0.24, an interquartile range of 0.07, an average of 0.1579, and no outliers.

Thus, we have our final calculations for the rates: active foragers will transition to inactive foragers at a rate of approximately 0.69 with a range of 0.1 to 1.8, and inactive foragers will transition to active foragers at an average rate of 0.679592 with a range of 0.15 to 1.3.

# Epidemic Models

Having discussed how the Ising Model applies, we will now turn to models inspired by epidemic models. Perhaps the best known example of an epidemic model is the so-called SIR model, which consists of three compartments: Susceptible (individuals who have not yet been affected by the disease), Infected (individuals who currently have the disease), and Removed (individuals who are now immune, either by virtue of having recovered from the disease or by receiving a vaccine).

At this point, our previous finding that interaction networks demonstrate small-world properties is rather convenient, since real social networks also show small-world properties. Note, however, that the average path length appears to be considerably larger in human social networks than in ant colony networks (which makes sense given the relatively confined space of an ant colony). Travers and Milgram famously computed an average path length of 5.2 (giving rise to the famous “six degrees of separation”), as opposed to our previous determination that the average path length for an ant social network is approximately 1.62963. (Travers and Milgram 1969).

Obviously, if a single vertex in a graph is “infected,” the entire connected component that it’s part of will eventually become infected. If we relax the restrictions on the model somewhat, this could be prevented by promptly quarantining the affected individuals; a notable example of this was Mary Mallon (better known as “Typhoid Mary”), who was the first person in the United States proven to be an asymptomatic carrier of a disease. She was forcibly quarantined after contact tracing found her to be the cause of a large number of cases. She was eventually released on the condition that she not return to her previous occupation of cooking; after she failed to comply with these terms, she was forcibly quarantined for the final two decades of her life [38].

A SIS model allows infected individuals to recover; however, no immunity is conferred from having the disease. is the recovery rate.

The model for a SIR is (Barrat *et. al.* 2013):

We will initially start with the SIS model described above. This would perhaps be more aptly named an IAI model (Inactive – Active – Inactive) model.

In this case, we can immediately recognize that and will be the same as and from our previous model – they are simply the rate of departing and arriving foragers. (Video analysis confirms that returning foragers often simply drop their load shortly after return for other ants to bring further down into the nest, which supports our assumption that the rates will, in fact, be the same as the number of arriving and departing foragers). (The Regulation of Ant Colony Foraging Activity without Spatial Information, Gordon). We already calculated that , , and . So,

We will also pick the initial condition

This is, of course, a separable differential equation. We can divide both sides by and multiply both sides by to get:

Which is equivalent to:

We will now integrate both sides:

Rather than work through the full math, we will simply use Matlab’s ODE45 solver to solve for a timespan from 0 to 10:

mu = 0.679592;

k = 62.6;

beta = 0.689091;

tspan = [0 10]

a0 = 0.7

[t, y] = ode45(@(t, a) -(mu \* a) + (beta \* k \* a) - (beta \* k \* a^2), tspan, a0)

plot(t, y, '-o')

This results in the following plot:



Figure Rate of departure for k = 62.6

The values fairly quickly move from a value of 0.7 to stabilize at a value of approximately 0.9840 (with a small amount of cyclic behavior). This is, of course, rather problematic in that it implies that approximately 98% of foragers are active at all times, which is definitely *not* the case (especially for larger colonies). It also ignores fluctuations – we know that foragers do, in fact, have multiple possible rates. This suggests that something has gone wrong with our model. We need the “stable” baseline rate to be around 70%, and a second “higher” rate.

This does not appear to be a function of the initial value we choose; selecting, for example, or results in a very similar outcome.

However, this does *not* appear to be a function of the model itself; if we select , and , we get the following plot:



Figure Rate of departure for , , and

In other words, this strongly suggests that the way that the parameters we have chosen and the model interact strongly.

For , , and :



Figure Rate of departure for , , and

For , , and :



Figure Rate of departure for

For , , and :



Figure Rate of departure for

Our problem, then, appears to be twofold: our choice of and our choice of . In fact, since is the total spreading rate, one is initially tempted to simply say that we were mistaken in setting and that we should simply state that

We can simulate this with our original Matlab code simply by changing k = 62.6; to k = 1;

Unfortunately, the situation with this variable doesn’t appear to be much better, as it predicts that the foraging rate will drop from 0.7 to a little over 0.1 within ten time intervals:



Figure Result of setting k to 1 in the Matlab script

Such a conclusion is, of course, roundly contradicted by empirical evidence.

If we continue to utilize , for we get the following plot:



Figure Result of setting

For we get the following:



Figure Result of setting

Both are plausible.

We finally arrive at the source of the trouble: our original data set recorded all interactions by forager ants anywhere. This “worked” for our prior models. However, if is the average number of ants that an active forager encounters and is the rate at which successful foragers are returning, then is the rate at which successful foragers are encountering quiescent foragers.

We will use a data set supplied by Pless *et. al.* In this study, the “roof” of an entrance chamber was removed and replaced with a transparent cover and ant interactions were video recorded and later tagged using a Matlab script. (Prior research demonstrated that the light does not substantially alter the ants’ behavior). This data set identifies two types of ants: ants that leave the entrance chamber as active foragers (type 1) and ants that descend into the nest.

An example of the format that the data set is provided in as follows:

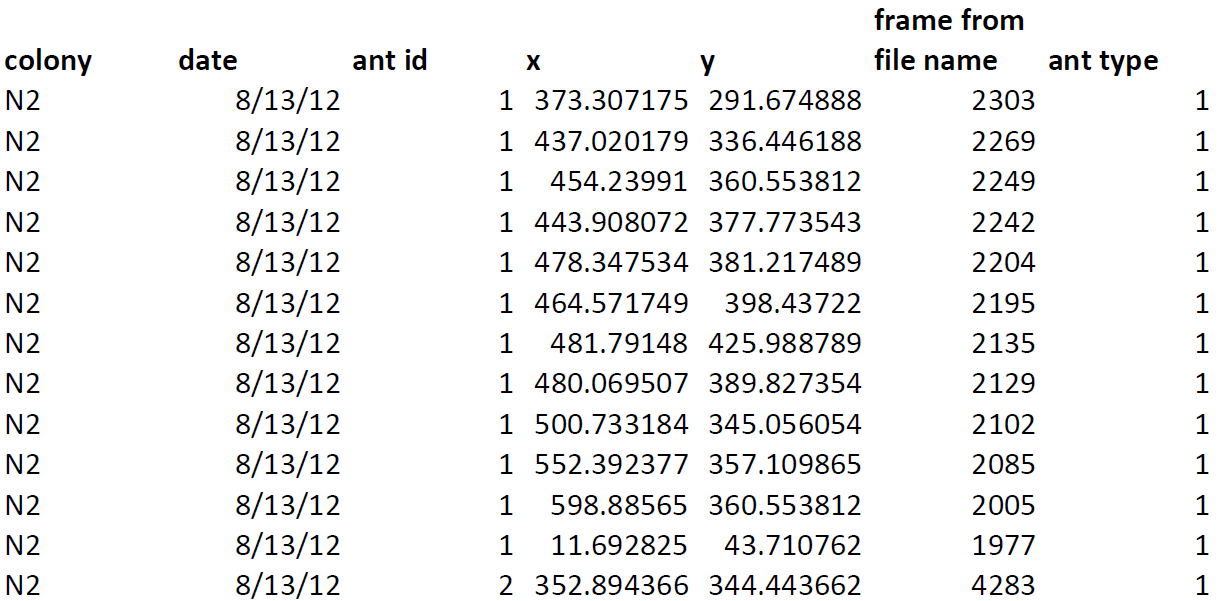


Figure Example of file format for the forager data set data set (Pless et. al. 2015)

Conveniently, if we copy this table Adobe Acrobat Reader into Notepad, it preserves the tabular format. By doing a “Replace All” to replace all whitespace with commas, we can easily convert this to a CSV (which will be used below).

In this data set, the ID is not necessarily unique across days or colonies, and the ID is not necessarily chronological by time. The combination of date, colony, and Ant ID uniquely identifies a particular ant. There are two ant types: 1 indicates an outgoing forager, and 2 indicates a descending ant. **X** and **Y** are special coordinates within the chamber. The first entry marks where an ant entered the entrance chamber, the last entry indicates where they left. All other entries represent interactions. So, in the example shown above, the ant enters the chamber at , interacts 11 times, and finally leaves the entrance hall to forage at . Thus, if we group by the colony, date, and ant ID, get the count, and subtract 2 (to exclude entrances and exits), we get the number of interactions:

import pandas as pd

# Read only the columns we need

df = pd.read\_csv(' NestChamberInteractions.csv', usecols = ['colony', 'date', 'antid', 'antType'])

# Filter out all descending ants - we only want ants that become active foragers

df.where(df['antType'] == 1, inplace = True)

# We don't need that column anymore

df.drop(columns = 'antType', inplace = True)

# Add a Count column that's populated with blanks

df['Count'] = [''] \* len(df)

# Get the count and write it to a new CSV file

countDf = df.groupby(['colony', 'date', 'antid']).count()

# Exclude entrances and exits, so that counts represents the true interaction number

countDf['Count'] = countDf['Count'] - 2

countDf.to\_csv('NestChamberInteractionsCounts.csv')

As mentioned in the code comments, we will only examine ants that become active foragers. We have several options for how to get our value of from this. One is to simply look at how many each ant interacts. That will, however, overestimate the number of interactions that take place because the data set does not distinguish between quiescent foragers interacting with successful foragers, successful foragers interacting with other successful foragers, and quiescent foragers interacting with other quiescent foragers. (In fact, some of the ants in the entrance chambers may not be foragers at all but are rather some other class of ant, such as a cleaner).

A second possibility is to estimate how long they are in the entrance chamber on average and compare that to the rate at which foragers are arriving in order to estimate which percent of the interactions are with successful foragers. (We can get their duration in the entrance chamber by calculating the number of video frames that occurred between the first and last entry of a particular ant being in the chamber; since the frame rate is 30 frames per second, we simply divide by 30 to get the number of seconds that they were in the entrance chamber on average).

Another possibility is to determine which pairs of interactions from the same colony and same day have the smallest Euclidean Distance and assume that those are the interactions that take place. (This can be done with a k-d tree, for example, which is part of SciKit-Learn).

We will examine the first two possibilities in turn. First, we will use SAS to perform a distribution analysis of the count data we generated with the code shown above. We find that the data is not normally distributed:

| **Goodness-of-Fit Tests for Normal Distribution** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Kolmogorov-Smirnov** | **D** | 0.12004132 | **Pr > D** | <0.010 |
| **Cramer-von Mises** | **W-Sq** | 0.26724748 | **Pr > W-Sq** | <0.005 |
| **Anderson-Darling** | **A-Sq** | 1.73778206 | **Pr > A-Sq** | <0.005 |



Figure Histogram, reference normal curve, and descriptive statistics for the count distribution shown above

We will now determine how long ants are in the frames. The code is the following:

import pandas as pd

import numpy as np

# Read only the columns we need

df = pd.read\_csv('NestChamberInteractions.csv', usecols = ['colony', 'date', 'antid', 'antType', 'FrameFromFileName'])

# Filter out all descending ants - we only want foragers

df.where(df['antType'] == 1, inplace = True)

# We don't need that column anymore

df.drop(columns = 'antType', inplace = True)

groups = df.groupby(['colony', 'date', 'antid'])

# The first row in a group is the ant entering

entranceFrames = np.array(groups.nth(0).FrameFromFileName)

# The last row in a group is the ant exiting

exitFrames = np.array(groups.nth(1).FrameFromFileName)

frames = np.subtract(entranceFrames, exitFrames)

seconds = np.divide(frames, 30)

pd.DataFrame({'Frames': frames, 'Seconds': seconds}).to\_csv('Ants time in chamber.csv')

We will now run distribution tests on this in SAS:



Figure Distribution of the number of seconds that ants spend in the entrance chamber

The skewness and kurtosis suggest that this is not normally distributed. Indeed, , indicating that the data is indeed not normal.

| **Goodness-of-Fit Tests for Normal Distribution** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Kolmogorov-Smirnov** | **D** | 0.2847429 | **Pr > D** | <0.010 |
| **Cramer-von Mises** | **W-Sq** | 2.6592751 | **Pr > W-Sq** | <0.005 |
| **Anderson-Darling** | **A-Sq** | 13.6566448 | **Pr > A-Sq** | <0.005 |

The unusually low average number of seconds in the entrance chamber is surprisingly low. If foragers are returning at a rate of per second, that would suggest that

This is as opposed to the estimate from the first suggested method.

This has the rather unfortunate effect of



Figure Result of setting

On the other hand, has the following plot:



Figure Result of setting

Using (based on the median number of interactions) is quite similar.

If we split the difference and set

we end up with the plot:



Figure Results of setting

This appears to be quite satisfactory.

Regardless of the exact value of , then, this model appears to be making reasonable predictions. It also has another extremely important feature: for higher values of , the number of foragers increases (potentially very rapidly). This is exactly consistent with field observations, which indicate that ants very rapidly scale up the number of foragers in response to an increase in the number of successful forager. A key question we have posed previously was whether this behavior is itself adaptive, given that harvester ants are foraging for food that more or less stays in the same place.

Our Ising Model suggested that it’s highly advantageous for the number of active foragers to be based on the probability of success, since this results in lower energy consumption for the colony as a whole; however, the speed at which this change took place may be a byproduct of local interactions of ants aligning in a way to minimize their individual energy consumption. This model gives further support to this conclusion.

Note that these models do *not* conclusively prove that this behavior does not have an adaptive purpose; it would likely be quite difficult or impossible to prove that such an adaptive purpose does not exist. (In fact, for several species of ants, this feature of their behavior is demonstrably adaptive). However, it does suggest that this behavior would be present whether it’s adaptive or not. (The question of whether it would continue to exist if it is somehow maladaptive is entirely separate; however, as the behavior does not appear to cause any specific harm, this point appears to be moot).

# Cast Selection

Larger ant colonies tend to have a high degree of division of labor. Research by Mersch *et. al.* [7]tracked six separate colonies over 41 separate days and found three distinct behavioral repertoires, with much higher within-group interaction than between-group interaction. This graph of interactions is highly informative, and various metrics can be used to predict which ants will fall within which group. Merely by applying Random Forest or Neural Network algorithms to two graph metrics on the data set from the study, Hubness and PageRank, we were able to predict the behavioral repertoire of the ants with 55% accuracy, which is considerably better than random chance. Hubness/Authority Ratio as a graph metric has been studied as a means of classifying, for example, neuron types in the brain (Szczurek and Horeni 2018).

We will now demonstrate what factors lead to ants becoming foragers in the first place. As previously described, ants tend to follow a predictable lifespan, starting as nurses, eventually becoming cleaners, and finally becoming foragers (Mersch *et. al.* 2013), with the transition to forager typically being permanent (Porter and Jorgensen 1981).

This progression is not arbitrary, especially for larger colonies that since there are significant advantages to the oldest ants becoming foragers, given that they’re at higher risk for predation and exposure to harmful pathogens than other ants are. Porter and Jorgensen 1981 found that the average longevity of *Pogonomyrmex owyheei* (Western Harvester Ant) foragers was only 14 days, leading them to conclude that the social structure of this particular type of Harvester Ant (which is subject to unusually high predation) adapted to be able to tolerate high forager loss. While the longevity for Red Harvester Ants is likely considerably less extreme, foragers are still subject to considerable predation risk (especially if they are part of a large colony), given that a predator may regard a large colony as a stable food source.

Figure 27 gives a general sense of the overall interaction network.

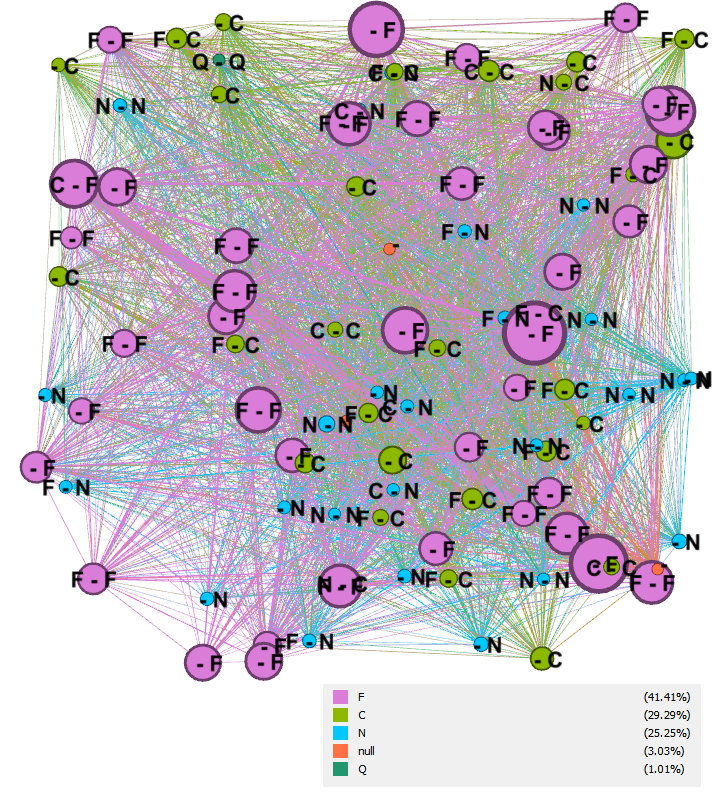


Figure - Evolution of foragers. Size indicates the number of interactions with foragers, color indicates initial behavioral repertoire, and the label indicates initial and final behavioral repertoire. From Mersch et. al. 2013, colony 1, day 15.

We will use three key metrics: Betweenness Centrality, PageRank Score, and Hubness (or Authority).

Hyperlink-Induced Topic Search (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 (Kleinberg 1999). 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 (Szczurek and Horeni 2018).

We encounter an issue at this point in that, since our data set is an undirected network, the in-degree and out-degree are the same by definition. (In a directed network, the in-degree of vertex *v* is the number of “incoming” arcs, and the out degree is the number of “outgoing” arcs; an undirected network has no distinction between an incoming and an outgoing arc). Thus, Hubness and Authority will always be the same, and the H/A Ratio is 1 by definition. For this reason, we will simply use Hubness in our predictions. (The choice of Hubness over Authority is purely arbitrary).

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 (Brin and Page 1998). 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 .” (Barrat *et. al.* 2013) 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).

is the out-degree of node . (Since our data is an undirected network, this will simply be the node degree).

is a *dampening factor* (which is the probability that the web surfer stops following links and merely jumps to a random page; this prevents the surfer from getting “stuck” on nodes with no out-degree).

is simply the number of web pages (or, in our case, ants).

Betweenness Centrality, PageRank, and HITS statistics are all measures of how important a particular ant is in the network.

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.

Next, we will use the Orange data mining package to compare the performance of several Machine Learning models:

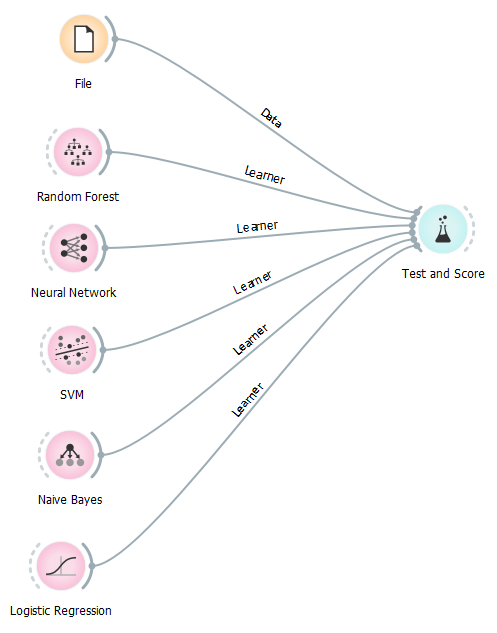


Figure Orange Data Mining package used for training Supervised Learning models

The result of that comparison is as follows:

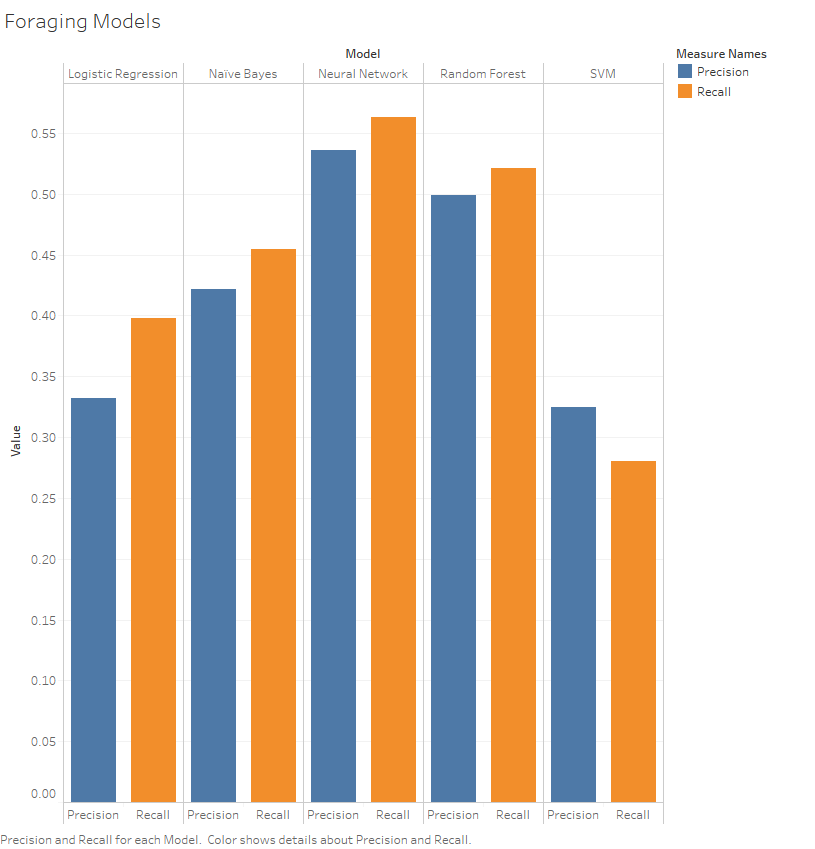
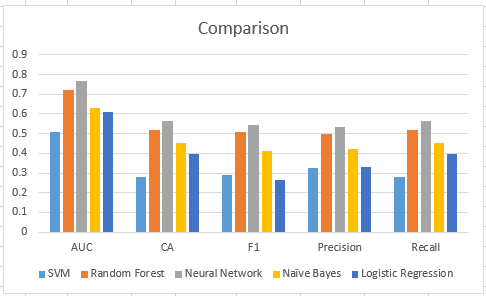


Figure Precision and Recall of each Supervised Learning model

This is further confirmed if we compare all of the metrics that Orange produces against each other. Neural Networks outperform all other approaches, followed by Random Forest classifiers as a close second:



For thoroughness, we will now compare these results to age to ensure that age is not a confounder. In order to do that, we train the same five model types (Logistic Regression, Naïve Bayes, Neural Network, Random Forest, and SVM) using the colony day as the sole feature and class as the target.

We will look first at the effect on precision:

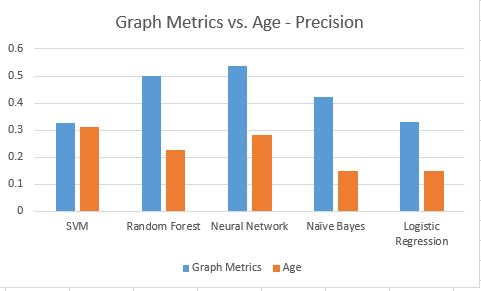


Figure Precision and Recall of Supervised Learning models based on either age alone (colony day) or Graph Metrics

This chart shows that, with the exception of Support Vector Machines (which shows only a minor difference), graph metrics are a far superior predictor than age alone. This is further confirmed by Single-Factor ANOVA in Excel, which shows the following summary statistics:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
| Graph Metrics | 5 | 2.114 | 0.4228 | 0.0091077 |
| Age | 5 | 1.117 | 0.2234 | 0.0056108 |

and the following ANOVA table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.0994009 | 1 | 0.0994009 | 13.50693345 | 0.006262537 | 5.317655072 |
| Within Groups | 0.058874 | 8 | 0.00735925 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 0.1582749 | 9 |  |  |  |  |

Thus we can see that graph metrics are superior predictors than age.

The same thing was tried using both graph metrics and age; while the results initially appear marginally better, surprisingly enough, the improvement was not statistically significant:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.002434 | 1 | 0.002434 | 0.199074 | 0.667299 | 5.317655 |
| Within Groups | 0.097797 | 8 | 0.012225 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 0.10023 | 9 |  |  |  |  |

The low F value and high P-value suggest that age is not actually adding much.

Next, we will compare this to simply using the number of interactions between different ant types.

Gephi’s MDS Layout with a Network Scale of 20, nb\_interaction\_cleaners as dimension 1, and nb\_interaction\_foragers as dimension 2 gives us the following cleanly separated visualization:



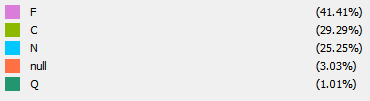


Figure MDS Layout of colony 1, day 15 of ant interactions (Mersch 2013)

That being the case, it is worth checking whether we can use the same features to cluster our data. However, a major problem with many clustering algorithms is that they will find clusters even if clusters don’t “exist” in the data. The Visual Assessment of (Cluster) Tendency (VAT) algorithm creates a heatmap based on reordered dissimilarity data (obtained with some type of norm on the data) which aids in identifying whether clusters exist in the data. Clusters typically appear as a “darker portion” of the heatmap (Bezdek and Hathaway n.d.). Unfortunately, our heat map is as follows (Abuomar 2021):



Figure VAT heat map for the nb\_interaction\_cleaners and nb\_interaction\_foragers features. Matlab VAT implementation from (Abuomar 2021).

The darker portions of the heat map are widely diffuse, which is not strongly supportive of the existence of clusters. The scatter plot is not much more helpful in this regard:



Figure Scatter plot corresponding to Figure 34

It would appear, then, that clustering is not the best approach in this instance. However, supervised learning algorithms perform exceptionally well in this case, with the k-Nearest Neighbor algorithm trained on the number of interactions with foragers, cleaners, nurses, and queens having a precision of 0.928 and a recall of 0.938. If we compare the Precision of seven different Supervised Learning algorithms trained against graph metrics like PageRank and the same supervised learning algorithms trained against the number of interactions with other classes of ants, ANOVA gives us the following:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.723677786 | 1 | 0.723677786 | 190.1972216 | 1.0118E-08 | 4.747225347 |
| Within Groups | 0.045658571 | 12 | 0.003804881 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 0.769336357 | 13 |  |  |  |  |

This leads us to reject the null hypothesis that Supervised Learning algorithms trained against PageRank and other algorithms is just as accurate as Supervised Learning algorithms trained on the number of interactions with other classes of ants. This is also consistent with, and supportive of, the conclusion of Mersch *et. al* that interactions within groups are much higher than interactions outside of groups.

Ultimately, then, the most predictive factor is which class of ant an ant is interacting with. While age is predictive, it is difficult to use age alone to identify exactly when they will make the switch. Importance within the colony is also predictive, although not quite as predictive as which specific ants the ants are interacting with.

# Conclusion

Ants make decentralized decisions based on local interactions. Their patterns of interactions strongly influence both when ant foragers leave the nest to forage and when ants become foragers in the first place.

Additionally, local patterns of ant foraging results in emergent high-level behavior that may be quite different than what may be predicted only based on individual ants. Since the colony as a whole is the reproductive unit of ants, natural selection predicts that the net effect of these emergent behaviors will be to increase the optimality of the ant colony’s behavior (or that, at a minimum, they will not be actively harmful to reproductive fitness).

There is, however, no guarantee that every attribute of every property is itself adaptive, as some are simply side effects of the way that the system as a whole works. In particular, we have identified the speed at which foraging increases as being an incidental byproduct of the way that the system works. While this particular behavior is highly adaptive for some species of ants, it is not necessary to assume that this holds for all species of ants, as this property would likely continue to hold even if it wasn’t adaptive.

# Acknowledgement

A portion of the charts and data analysis were generated using SAS University Edition. Additional data analysis was performed using NetworkX 2.5.1, SciKit-Learn 0.21.2, Pandas 1.5.2, Tableau 2018, Gephi, Pandas 1.2.5, Numpy 1.16.1, Orange 3.23.0, and Excel 2013.

References

1. Gordon, D. (2010). *Ant encounters: Interaction networks and colony behavior*. Princeton, NJ: Princeton University Press.
2. Jarau, S., & Hrncir, M. (2019). *Food exploitation by social insects: Ecological, behavioral, and theoretical approaches*. Boca Raton: CRC Press/Taylor & Francis Group.
3. Adler, F., & Gordon, D. (2003). Optimization, conflict, And Nonoverlapping foraging ranges in ants. *The American Naturalist,* *162*(5), 529-543. doi:10.1086/378856
4. Gordon, D. M. (2013). The rewards of restraint in the collective regulation of foraging by harvester ant colonies. *Nature,* *498*(7452), 91-93. doi:10.1038/nature12137
5. Barrat, A., Barthélemy, M., & Vespignani, A. (2013). *Dynamical processes on complex networks*. Cambridge: Cambridge University Press.
6. Lopes, J. F., Brugger, M. S., Menezes, R. B., Camargo, R. S., Forti, L. C., & Fourcassié, V. (2016). Spatio-Temporal dynamics of Foraging networks in the Grass-Cutting Ant ATTA bisphaerica FOREL, 1908 (Formicidae, Attini). *PLOS ONE,* *11*(1). doi:10.1371/journal.pone.0146613
7. 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
8. 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
9. Kikuta, S., Toriumi, F., Nishiguchi, M., Liu, S., Fukuma, T., Nishida, T., & Usui, S. (2020). Framework for role discovery using transfer learning. *Applied Network Science,* *5*(1). doi:10.1007/s41109-020-00281-3
10. Ribeiro, L. F., Saverese, P. H., & Figueiredo, D. R. (2017). Struc2vec. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. doi:10.1145/3097983.3098061
11. 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
12. Quevillon, L. E., Hanks, E. M., Bansal, S., & Hughes, D. P. (2015). Social, spatial and temporal organization in a complex insect society. *Scientific Reports*, *5*(1). https://doi.org/10.1038/srep13393
13. Fewell, J. H. (1988). Energetic and time costs of foraging in harvester ants, Pogonomyrmex occidentalis. *Behavioral Ecology and Sociobiology*, *22*(6), 401–408. https://doi.org/10.1007/bf00294977
14. Gelblum, A., Pinkoviezky, I., Fonio, E., Ghosh, A., Gov, N., & Feinerman, O. (2015). Ant groups optimally amplify the effect of transiently informed individuals. *Nature communications,* 6, 7729. <https://doi.org/10.1038/ncomms8729>
15. Gordon, D. M. (1991). Behavioral Flexibility and the Foraging Ecology of Seed-Eating Ants. *The American Naturalist*, *138*(2), 379–411. https://doi.org/10.1086/285223
16. Schafer, R. J., Holmes, S., & Gordon, D. M. (2006). Forager activation and food availability in harvester ants. *Animal Behaviour*, *71*(4), 815–822. <https://doi.org/10.1016/j.anbehav.2005.05.024>
17. Kelrick, M. I., & MacMahon, J. A. (1985). Nutritional and Physical Attributes of Seeds of Some Common Sagebrush-Steppe Plants: Some Implications for Ecological Theory and Management. *Journal of Range Management*, *38*(1), 65. https://doi.org/10.2307/3899336
18. Fewell, J. H. (1988). Energetic and time costs of foraging in harvester ants, Pogonomyrmex occidentalis. *Behavioral Ecology and Sociobiology*, *22*(6), 401–408. https://doi.org/10.1007/bf00294977
19. Ferral, N., Holloway, K., Li, M., Yin, Z., & Hou, C. (2017). Heterogeneous activity causes a nonlinear increase in the group energy use of ant workers isolated from queen and brood. *Insect Science*, *25*(3), 487–498. https://doi.org/10.1111/1744-7917.12433
20. Gordon, D. M. 1995. The development of an ant colony’s foraging range. Anim. Behav. 49: 649-659
21. Morehead, S. A., & Feener, D. H., Jr. (1998). Foraging behavior and morphology: seed selection in the harvester ant genus, Pogonomyrmex. *Oecologia*, *114*(4), 548–555. <https://doi.org/10.1007/s004420050479>
22. Gelblum, A., Pinkoviezky, I., Fonio, E., Ghosh, A., Gov, N., & Feinerman, O. (2015). Ant groups optimally amplify the effect of transiently informed individuals. *Nature communications*, *6*, 7729. <https://doi.org/10.1038/ncomms8729>
23. Kelrick, M. I., & MacMahon, J. A. (1985). Nutritional and Physical Attributes of Seeds of Some Common Sagebrush-Steppe Plants: Some Implications for Ecological Theory and Management. *Journal of Range Management*, *38*(1), 65. https://doi.org/10.2307/3899336
24. Steen, M. van. (2010). Graph theory and complex networks: an introduction. Maarten van Steen.
25. 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
26. 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
27. 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
28. Gordon, D. M. 1995. The development of an ant colony’s foraging range. Anim. Behav. 49: 649-659
29. Ferral, N., Holloway, K., Li, M., Yin, Z., & Hou, C. (2017). Heterogeneous activity causes a nonlinear increase in the group energy use of ant workers isolated from queen and brood. *Insect Science*, *25*(3), 487–498. https://doi.org/10.1111/1744-7917.12433
30. Fewell, J. H. (1988). Energetic and time costs of foraging in harvester ants, Pogonomyrmex occidentalis. *Behavioral Ecology and Sociobiology*, *22*(6), 401–408. https://doi.org/10.1007/bf00294977
31. Kelrick, M. I., & MacMahon, J. A. (1985). Nutritional and Physical Attributes of Seeds of Some Common Sagebrush-Steppe Plants: Some Implications for Ecological Theory and Management. *Journal of Range Management*, *38*(1), 65. <https://doi.org/10.2307/3899336>
32. *Sunrise and sunset time, day length in Phoenix, Arizona, USA*. Sunrise and sunset time, day length in Phoenix, Arizona, USA today and for the current month. (n.d.). https://dateandtime.info/citysunrisesunset.php?id=5308655.
33. Wikipedia contributors. (2021, June 2). Boltzmann constant. In *Wikipedia, The Free Encyclopedia*. Retrieved 19:47, June 6, 2021, from <https://en.wikipedia.org/w/index.php?title=Boltzmann_constant&oldid=1026393161>
34. Prabhakar, B., Dektar, K. N., & Gordon, D. M. (2012). The Regulation of Ant Colony Foraging Activity without Spatial Information. *PLoS Computational Biology*, *8*(8). https://doi.org/10.1371/journal.pcbi.1002670
35. Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
36. Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association*, *81*(396), 991–999. https://doi.org/10.1080/01621459.1986.10478363
37. Travers, J., & Milgram, S. (1969). An Experimental Study of the Small World Problem. *Sociometry*, *32*(4), 425. https://doi.org/10.2307/2786545
38. Wikipedia contributors. (2021, May 21). Mary Mallon. In *Wikipedia, The Free Encyclopedia*. Retrieved 03:35, June 8, 2021, from <https://en.wikipedia.org/w/index.php?title=Mary_Mallon&oldid=1024328309>
39. Pless, E., Queirolo, J., Pinter-Wollman, N., Crow, S., Allen, K., Mathur, M. B., & Gordon, D. M. (2015). Interactions Increase Forager Availability and Activity in Harvester Ants. *PLOS ONE*, *10*(11). https://doi.org/10.1371/journal.pone.0141971
40. Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart, [“Exploring network structure, dynamics, and function using NetworkX”](http://conference.scipy.org/proceedings/SciPy2008/paper_2/), in [Proceedings of the 7th Python in Science Conference (SciPy2008)](http://conference.scipy.org/proceedings/SciPy2008/index.html), Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008
41. 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
42. 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>
43. Porter, S. D., & Jorgensen, C. D. (1981). Foragers of the harvester ant, Pogonomyrmex owyheei: a disposable caste? *Behavioral Ecology and Sociobiology*, *9*(4), 247–256. <https://doi.org/10.1007/bf00299879>
44. Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799
45. Bezdek, J. C., & Hathaway, R. J. (n.d.). VAT: a tool for visual assessment of (cluster) tendency. *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*. https://doi.org/10.1109/ijcnn.2002.1007487
46. Abuomar, Osama (2021). VAT Algorithm implementation.
47. Trudeau, R. J. (2015). *Introduction to graph theory*. Dover Publications.