Masteroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for naturvitenskap
Institutt for bioteknologi og matvitenskap

Vetle Rem

Agent-Based Modelling of Bee Communication Networks

Masteroppgave i Industriell kjemi og bioteknologi
Veileder: Eivind Almaas

Februar 2020

NTNU
Kunnskap for en bedre verden

Vetle Rem

Agent-Based Modelling of Bee Communication Networks

Masteroppgave i Industriell kjemi og bioteknologi
Veileder: Eivind Almaas
Februar 2020

Norges teknisk-naturvitenskapelige universitet
Fakultet for naturvitenskap
Institutt for bioteknologi og matvitenskap

NTNU
Kunnskap for en bedre verden

Acknowledgements

Before anything else, I need to give proper thanks to my supervisor, Professor Eivind Almaas. I'm very grateful that he provided me with the opportunity to work with such a fun and engaging project for my thesis.

I must also give my thanks to my co-supervisors, Emil Karlsen and Pål Røynestad. Throughout this project I've been relying on them for their expertise whenever I've gotten stuck programming scripts or when I've needed constructive and critical input on where I was heading with my ideas.

I would also like to thank my parents. Their support and their interest in my work has been a major driving force in the completion of this thesis.

Lastly, I would like to thank my friends, both old and new. I need to thank Torstein, Håkon, Stine and Alejandra, for being the ones that I could lean on and the ones who would listen to both my woes and my joys. I need to thank all the wonderful people at NTNUi Dans, who helped me discover my passion for dance and who renewed my confidence when it was low. I need to thank André for showing me the *Star Wars*-prequels, an experience I had been missing out on for far too long. There are many more who will go unmentioned here, but I wouldn't be where I am today without my friends.

Abstract

In this thesis an agent based model was constructed in order to recreate the social interaction network of a beehive based on data from an empirical data set of registered communications between honeybees. The data set was analyzed to find traits that the agent based model would have to recreate. An initial model based on random walks was used as a null hypothesis to compare to empirical data. After comparing the initial model with the empirical data, an enhanced model was made. The enhanced model to more accurately describe patterns and dynamics found in the empirical data set. Data from the empirical data sets as well as data generated using synthetic models were analyzed using basic network analysis. The enhanced models were found to more accurately capture the communication dynamics of the bee social network than the initial random walker model.

Sammendrag

I denne avhandlingen ble det laget en agentbasert modell som skulle gjenskape det sosiale interaksjonsnettverket i en bikube basert på et empirisk datasett med registrerte kommunikasjonstilfeller mellom honningbier. Datasettet ble analysert for å finne egenskaper og dynamikker som modelleringen hadde som hensikt å gjenskape. Modelleringen startet ved å anta at biene kunne representeres som tilfeldig vandrende objekter i en matrise. Denne antakelsen ble så sammenlignet med de empiriske dataene, slik at en ny modell kunne programmeres med målsetningen om å gjenskape egenskapene observert i de empiriske dataene mer nøyaktig. Dataene fra det empiriske datasettet og dataene generert i modelleringen ble analysert ved hjelp av nettverksanalyse. Den reviderte modellen viste seg å gjenskape de empiriske kommunikasjonsmønstrene langt bedre enn antakelsen om tilfeldige vandrere gjorde.

Contents

1 Introduction

Humans have domesticated a wide variety of animals across history, but perhaps the most important one is the honeybee, *Apis Mellifera*. Bees are so valuable that the human race has introduced them to all corners of the world.[1] Bees are general pollinators and are vitally important in agriculture, as they ensure the propagation of our food crops.[2] [3] This also gives the honeybees a distinct economic value [1], and as everyone knows, money is very important.[4] In modern times, however, the bees are under a lot of pressure from new threats. Climate change brings perturbations in the bee's environment, and may also give rise to new diseases that threaten the bees, or the spreading of parasites that endanger commercially used bees.[5] [2] Human actions might also be a more direct threat to bees, such as the use of pesticides. [6] Pesticides used to protect plant crops from pests have been found to have detrimental effects on wild honeybee populations, bees that are crucial to the pollination of many of the crops we're trying to protect from pests. Such concerns make it very important to study the bees and how they are affected by such factors. By studying the bees, we can better develop strategies to protect them against threats.

Bees are known to be social insects, and the honeybee society is one of the most advanced social structures in nature. A honeybee society will display traits that humans might recognize from our own social structures of monarchies, dictatorships, but also democracies.[3] As such, much research has gone into studying the honeybee's social structures, their interactions, and their behaviours within the hive.

Tim Gernat has studied honeybees extensively, and recently made a study where the trophallaxis interactions of honeybees were monitored.[7] He found that honeybee social network exhibited traits similar to that of human social networks, in specific the burstiness of the networks was alike. The honeybees displayed bursty interaction patterns, meaning that the bees have periods of frequent pairwise interactions separated by longer periods where the bees won't interact. This is common in human social networks too, as much time passes between our interactions, but our interactions are often in high number within short time spans. This phenomenon has been thought to slow down flow of information in human networks[8], but in the honeybees this did not seem to apply.

Bees have several ways of communicating, but Gernat's experiment specifically charted the trophallactic communication interactions between bees.[7] Trophallaxis involves the exchange of regurgitated liquid food between bees. This allows for oral transfer of bacteria, viruses or ingested pesticides between members of a hive. Mapping the trophallactic interactions of bees allows us to get a better understanding of the bursty interaction patterns that Gernat uncovered. In turn, this will allow us to understand spreading dynamics in beehive communities in greater detail. If we know beforehand how diseases spread in bee communities, we might have better chances at combatting them.

With advanced technology, we're able to create digital models of virtually anything we could imagine. [9] The great thing about computer modelling is the ability to recreate patterns and dynamics seen in both society and nature. We can use these models to make predictions before we make decisions or changes to real life systems. For example, we might be interested in editing a bacterial strain to produce more of a specific compound. Lab work is tedious and time consuming, and bacteria have a lot of moving parts, so we would need to test a whole lot of different edited strains. But, what if we could make a prediction on how the bacteria would behave before editing them? If we model the metabolism of the cell and find what enzymes are needed for different processes in the cell, we can see how the metabolic flux changes when we knock out certain genes. When the bacteria is lacking a specific enzyme, the bacteria might produce more of our desired compound. If we can know beforehand what genes we should edit before moving to the lab, we save a lot of time. Even if we spend a lot of time making the model.

Another aspect of modelling is the ability to scale data to our needs. Gernat's experiment used 1200 bees[7], but a real beehive might contain upwards of 30000 bees, a quite sizeable gap. Provided that the communication dynamics of the bee social networks stay the same regardless of the size of the hive, a computer model will allow us to predict spreading dynamics in a larger hive based on the data from Gernat's experiment.

The aim of this thesis was to model the social interactions of a social bee network through the use of an agent based modelling approach. this was done by using an empirical data set where trophallactic interactions between honeybees in a colony were observed and registered.[7] The empirical data sets as well as synthetic data generated from the computer models were analyzed through network analysis, and the defining traits of the bee social behaviour were investigated. This was done to compare empirical and synthetic data, as to determine to what extent the model was able to capture social bee behaviour. When the differences and similarities between empirical data and model were determined, the findings from the analyses were used to make tweaks and adjustments to the computer modelling. This way, a base model was changed in order to more accurately capture and describe the empirical data.

2 Background and theory

There are three main topics of importance for this thesis. Bee biology and their behaviour, network science and its application, and finally computer modelling and its usefulness. First, the basic biology of honeybees and some observed behaviours that might be of relevance to later modelling will be described. Next, network science and how one may glean useful information from networks through network analysis will be explained. Some specific network attributes that will be relevant for network analysis done later in this thesis will also be elaborated. At last, some information of the capabilities of computer modelling will be provided, and some relevant variants of computer models that will be described.

2.1 Bee biology and behaviour

In both their larval state and their adult lives, bees eat primarily nectar and pollen. They gather these as resources for the hive to feed their young after processing the nectar into honey.[1] Honeybees process nectar into honey by regurgitating gathered nectar stored in their abdomen.[3] The bees use their proboscis, the long tongue-like appendage near the bee's mouth, to suspend regurgitated droplets of nectar. They then stretch their proboscis to elongate the drops into a thin film such that the water in the nectar droplet will evaporate at a higher pace. [3] Repeating this behaviour, the bees can reduce the water content of the nectar, turning it into honey. This technique is also employed by the bees in cases where they need to regulate the internal temperature of their hive, substituting nectar with pure water.[3] The evaporating water helps cool the hive.

As they gather nectar and pollen, bees pollinate flowering plants by transferring pollen from one plant to another. When a bee visits a blooming flower to gather nectar and pollen, some pollen catches onto its body, and rubs off when visiting another flower. This germinates the flower, which is very important in fruit bearing plants, as the pollination is vital for fruit development. [2] Naturally, not all plants alike, as some flowers have evolved in very specific ways to ensure pollination. [10] An example being long spurred flowers, requiring either a long proboscis to reach its nectar or for the pollinating bee to crawl further into the flower. If the bee crawls into the flower, it is more likely to pick up pollen, which ensures pollination of the plant. Many such coevolutionary patterns can be found in various flowers. The honeybee, *Apis Mellifera*, is a generalist pollinator, and will pollinate a variety of flowers. Other bee species might act as specialist pollinators, pollinating only a specific flower or a subset of flowers. In some cases this might require evolutionary specializations, such as the aforementioned elongated proboscis.[2] [10]

2.1.1 Queen, workers and drones

In a beehive, there are three kinds of bees, the queen bee, worker bees and drones. The different bees serve different purposes in the hive.

The Queen is at the center of the colony, and she has given birth to all the workers in her hive.[11] The queen asserts her self as the dominant ruler over the working bees by emitting pheromones that suppresses their reproductive abilities.[12] The queen does not, however, micro manage the worker bees. Hive needs are identified and communicated among worker bees, and in some cases decicions are made in a democratic manner by them.[3] In such cases, the queen is simply obligated to follow along with her subject's desires. A queen bee larva is initially no different from a worker larva, the differentiation between the two comes from the nursing of the larva.[13] Worker bee larvae are fed with honey and royal jelly.[3] Queen bee larvae, however, are fed solely on royal jelly.[13] This changes the gene expression in the larvae, making them mature into another queen. Royal jelly is a special secrete produced from the hypopharyngeal glands, or feeding glands, in young worker bees. [13] [3]

Queens are able to decide what eggs she lays are fertilized and what eggs are not.[11] Fertilized eggs can become workers or queens, while unfertilized eggs become drones. Different cells are constructed for each kind of bee. A Queen is only given sperm once on her mating flight. In some cases, a queen will be unable to fertilize more eggs as she has run out of sperm, and may only produce more drones. Worker bees may in some cases also lay eggs, but these eggs will always result in drones, as they are never given sperm.[14]

Worker bees are the labour force of the hive.[3] They are responsible for the upkeep of the hive, nursing larvae, foraging for food and defending the hive from attackers. The worker bees are all female, but due to the queen's pheromones, their reproductive abilities are suppressed.[12] Should a worker bee end up laying eggs, none of the eggs will be fertilized, giving birth to haploid male drones.[14] A worker bee typically lives between a couple weeks to a few months, depending on local conditions.

The tasks a worker bee performs depend on worker age, but also on hive needs.[3] This is called age polyethism.[15] Younger worker bees will keep mostly inside the hive, performing tasks such as cleaning the hive, constructing cells, and feeding younger brood. Older worker bees are tasked with foraging and exploring outside the hive. However, as there is only a limited amount of tasks available at any time, bees might change tasks independent of age. An unoccupied bee will in some cases go on a search for more tasks by patrolling the hive. If it finds work, it will proceed to perform the found task. The age polyethism of bees is related to feeding glands in the head and wax glands on their abdomen. As the bees age, the glands develop, making them more suited for specific tasks. In most cases, Only the oldest bees will forage, as their feeding glands and their wax glands are diminished. However, in some cases, such as when establishing a new hive after swarming, there will only be older bees in the hive. In these cases the older worker bees are fed such that they regain some of their feeding and wax glands, such that they can spend their rejuvenated youth tending new brood.

Drones do not contribute to the hive upkeep, are all males, and their sole purpose is reproduction.[16] [3] When mating season comes around, drones will leave the hive and search for a queen on her mating flight. Drones are haploid, as they are birthed from unfertilized eggs. Only fertilized eggs will become worker bees or queen bees. Since the drones are haploid, they provide a full copy of their genome when giving sperm to a queen. As a result, sperm cells are never recombined, only egg cells from the queen are recombined. The inheritance of bees is called haplodiploidy, as males are all haploid, while females are diploid. Males will never give their sperm directly to male offspring.

2.1.2 Bee Communication

Bees are social insects.[3] In a similar manner as other social insects such as ants, termites and some wasps, Bees tend to make large colonies where several thousand bees live together. The bees contribute to the hive upkeep by cleaning the hive, constructing more cells for the brood, foraging for food and making honey out of foraged nectar, and so on. To manage all of the tasks the bees have developed severeal means of communication to more effectively delegate tasks and to communicate important information about the hive status and about the outside.

Bees communicate in several different manners. The different communication vectors allow the bees to send a variety of messages through the hive. In the beehive, communication happens through trophallaxis, dance, and through pheromones.[7] [3] [12]

Trophallaxis is the process of exchanging regurgitated food between a pair of bees, usually combined with touching antennae.[17] [3] In trophallaxis one can observe begging behaviour and offering behaviour, bees might ask for food or offer food. One bee might urge another bee to share its food with it. If the other bee has food to share, it will regurgitate some liquid and allow the first bee to partake in it. This behaviour is seen in a variety of social insects, and provides information about hive needs throughout the bee society. Different factors might impact what messages are sent, for example, rejecting an offering bee might send the message that the hive is ”sated”, or that the quality of the offered food is poor. As an example, bees will regulate the temperature in the hive. If the hive gets too warm, the bees need water to cool it down. The bees will reject foragers bringing nectar, only accepting water, communicating the need for water.

Dance is perhaps the most widely recognized way of bee communication.[3] [18] [19] The waggle dance is used to convey a number of different messages, but they usually relate something concerning the environment outside of the hive. Forager bees communicate the location of good food sources using the waggle dance. The dancing bee will angle itself in relation to the sun, such that the angle corresponds to the food source location relative to the hive opening. The bee then performs a series of moves that describe the distance from the hive to the food source and the quality of the food source. A more energetic dance signifies a more plentiful food source. The dance will, if successful, recruit more bees to investigate the food source. Those bees might then perform the same dance, confirming the quality of the food source.[18] In some cases, a hive might become overpopulated, necessitating that some bees move out of the hive along with the queen.[3] Where the bees move, however, is done democratically by the workers. Some bees will scout different possible locations, forgoing foraging. When the scouts find a suitable location, they will attempt to convince the other scouts with the same waggle dance as if scouting for food sources. Once the scouts have decided on a location for a new hive, half the population of the hive including the queen leave, swarming to the new location and building a new home. The remaining workers tend to the old hive, as new queen bees are birthed. The new queens fight to the death over control of the old hive.[20]

2.2 Networks

In 1990, John Guare wrote the award winning play *Six Degrees of Separation.*[21] Towards the end of the play, one of the main characters, Ouisa, delivers a powerful monologue describing how "everybody on this planet is separated by only six other people". Ouisa elaborates on how this amazes her, but also troubles her. The idea of six degrees of separation, was originally proposed by Frigyes Karinthy, a hungarian author.[22] He proposed that Through your friends and acquaintances, you should find someone with a friend or an acquiantance, with a friend or an acquaintance and so on, until the sixth acquaintance, which is a certain predetermined person. Be it the norwegian Prime minister, a famous fashion designer, or someone living in the australian outback. The idea here is that some of your acquaintances have different friends than you and your other friends. Mapping all of these out would make a vast network connecting all people on earth. This social network could, in theory, be so densely connected that between any two given people, at most five people separate them.

In the social network spanning human society, the people are nodes, and the social bonds that ties us together are the edges.[23] Networks, however, can be constructed using a great variety of things using a variety of connecting factor. Transportation networks are easy and very tangible networks, as it is easy to visualize how roads connect cities, it would look akin to a map. But even with this simple network, the nodes could be connected in different ways. For example, we could connect cities using railways. Railways won't go to the same places as roads, and in some cases vice versa. A road map and a railway map would look somewhat similar, but still different. One would for example imagine the railway network to leave some nodes unconnected to the larger network, and the network would likely be less dense.

More complicated and less tangible networks can also be made, such as protein and gene interaction networks. In this kind of network, genes and proteins are the nodes, connected together by what genes code for what proteins, but also by which proteins that impact expression of genes, what proteins interact, and what genes interact.[24] Such networks can be used to see the impact of illnesses, or assist in gene editing of bacteria. Network models make it possible to for example see the impact of a gene knock out before doing such an experiment in real life.

2.2.1 Network attributes

Networks can be described according to a variety of properties. These properties might help in the understanding, explanation or prediction of behaviour in a given network. Categorizing the defining traits of a network may also allow for the recreation of the network.

Degree of connectivity, or simply connectivity or degree, is a node specific measurement that describes how many other nodes a given node is connected to.[23] This number is equal to the number of edges connected to the specific node. Some nodes might have a higher degree of connectivity compared to other nodes in the network, these nodes are called hubs. Average degree is a network wide measurement, that describes the average number of edges connected to each node.

$$\frac{2 \times e}{N} = \overline{k} \tag{1}$$

Average degree, $\overline{k}$ is calculated using equation 1, where e is the number of edges and n is the number of nodes in the network. The number of edges is multiplied by 2 because each edge connects to two nodes.

Degrees may also be made into a degree distribution, by plotting degree against the fraction of total nodes or number of nodes with that degree.[23] [24] The degree distribution can give some insight into whether or not the network is scale free or not. In a scale free network, the degree distribution will follow a power law. This is an important attribute, as scale free networks are often small world networks. Small world networks are, as their name would imply, defined by short distances between nodes, high clustering, and the prescence of highly connected hubs. Moving between any two nodes in such a network requires very few steps.

The clustering of a node describes what fraction of a node's neighbours are interconnected compared to the theoretical maximum of connections between those nodes.[23] Clustering as a measurement provides insight into how densely connected a network is.

$$C_N = \frac{2 \times e}{k(k-1)} \tag{2}$$

The clustering coefficient of a node N, C_N, is calculated using equation 2, where e represents edges among node N's neighbouring nodes, and k is the degree of node N. It is also possible to find the average clustering over a whole network, simply by finding the average of all node specific clustering coefficients. Clustering coefficients found in biological networks have a tendency to be higher than those found in random network models. [24]

2.2.2 Temporal networks

In some networks, it might be useful to investigate how network topology changes over time in order to glean more information from them.[25] When making graphs from networks, it's intuitive to simply make edges between nodes given that the interaction or connection between any given nodes exist. However, not all networks are accurately described by this method. As an example, you could make a network based on social media. Most social media such as Instagram, Twitter and Facebook let users follow other accounts. Since following is not always reciprocated, we could regard them as directed networks. We could make graphs based on this, accounts represented by nodes with directed edges representing their following patterns. Such a network could for example describe the interests of some users, or define groupings within the network. Users with similar content are more likely to follow each other and other accounts with similar content. This network will, however, not accurately describe the interactions between accounts. Users are usually logged onto their accounts at different times, and will therefore interact with each other at different times. We could imagine a network where nodes would turn on and off depending on when users are logged onto their accounts, or edges between nodes only existing in the time where users are interacting, such as through likes or direct messages. The topology of such a network would change with time as users log on and off, thus resulting in a temporal network, a network that changes over time.

Regular network attributes will change over time in temporal networks, but temporal networks will also have inherent attributes of interest.[25] Attributes such as degree of connectivity will obviously change if nodes and edges flicker in and out of existence. This can be used to map activity within the network, as times of higher connectivity would equate to more online accounts. An attribute of particular interest found in temporal networks is burstiness.[26] Burstiness describes a particular pattern in the frequency in interactions between nodes. Bursty interaction patterns are defined by short periods where interactions are frequent, separated by longer periods of infrequent or no interactions. Human social networks can be categorized as bursty interaction networks.

2.3 Computer modelling

The purpose of computer modelling is to capture some phenomenon in the form of computations and algorithms to better understand that phenomenon.[9] Computers, as their name implies, are used to greatly increase the speed of complex computations of both simple and advanced math. The deeper we delve into natural phenomena, the more complex they become, thus demanding the computing power of computers. The aim of computer modelling is to create a replica of patterns or phenomena found in nature. By condensing observed data from natural patterns into algorithms, instructions for a computer, the computer is effectively able to simulate previous observations. A replicated model can be used to great effect in either predicting outcomes of experiments, or to discover new factors that might not match previous expectations.

Several models have been made from bacterial genomes and their metabolic pathways.[9] Such models can be used to predict how the bacteria will react if one or more of their genes were knocked out.[24] Such metabolic models can also allow us to preemptively suggest what genes should be knocked out when we desire a certain result. If maximized cell growth in a bacteria is desired, a computer model may be used use to see what genes can be knocked out to increase cell growth while still having a sustainable bacteria. Accurate prediction through the use of a computer model might save several hours of tedious lab work.

When making a computer model, assumptions must inevitably be made based on the gathered data. The assumptions might not match reality, however. Humans, are exceptionally good at recognizing patterns, even in places where there are none. [27] [28] Replicating observations digitally, however, might not always be easy. Even if data are easily available, precisely why things are the way they are might be unknown. Therefore, assumptions and simplifications must be made. This might result in a model not accurately reflecting the observations. In this case, it's easy to argue that the simplifications were too broad too or that the assumptions were wrong. Or, from a more positive point of view, this might lead to the discovery of a previously unknown factor. Such unknown factors might have greater impacts on the observed data than previously thought. Merely the construction of a computer model could propel a scientific discovery.

2.3.1 Cellular systems

In a cellular automaton, a matrix of cells represent individual nondescript life forms.[9] The matrix is coloured in as to represent alive and dead individuals.

Coloured cells are alive, and uncoloured cells are dead or empty. In the cellular automaton, the cells behave in a particular fashion guided by a specific set of rules. The next "generation" or iteration of the automaton will might be determined solely by the state of the current generation or in a more stochastic manner. In the famous cellular automaton Conway's game of life, the cells die or live based on the state of nearby cells.[29] In Conway's game of life, a "live" cell will die when adjacent to fewer than two or more than three other live cells, as to mimic over population and under population. live cells next to two or three live cells will continue to live. A dead cell next to three live cells will become a live cell, mimicking reproduction. The simple rules in Conway's game of life makes it possible to make schemes of cells that will continuously repeat patterns, or even generate new units that will move independently across the matrix.

2.3.2 Agent-based models

Agent-based models might resemble the cellular automaton, but adds another layer of complexity. In cellular automatons, the cells are binary, but in an agent-based model different entities might occupy the cells.[9] [24] Agent-based modelling can be used to simulate interaction between entities of the same kind, or different kinds. In the model, the entities, or agents, will adhere to a certain set of rules. The rules may dictate how agents move, or how they interact. The rules will usually have a stochastic element, ensuring that any two simulations are never the same.

The agents may represent a variety of different entities, such as cells, molecules or animals.[9] The agents may have several defining characteristics that may be altered during the simulations. Their characteristics could affect how they interact with other agents in the model. As an example, agents in a model could represent a population of people who are susceptible to some disease. If some of the agents are sick and some are healthy, an interaction between these might change a healthy agent into a sick agent. Or perhaps healthy agents could sense the sick agents and avoid them.

3 Method

3.1 Software/materials

3.1.1 Python

The model was implemented in python. Python was chosen for its ease of use, as well as the wide variety of available libraries and packages tailored for specific applications. Pycharm was used as the integrated development environment.

3.1.2 NetworkX

NetworkX is a python package that enables both the construction of network graphs, and the analysis of them. NetworkX was used due to its impressive amount of algorithms covering a wide variety of needs. Data were analysed using python, and in the case of network analysis, NetworkX was instrumental. The following alorithms were used:

- graph() - This function creates an empty network to which nodes or edges can be added. The Graph-function will not add duplicate edges.
- MultiGraph() - This function creates an empty network. Nodes and edges may be added to this network, but unlike graph(), a MultiGraph-network will contain duplicate edges.
- add_edge() - This function adds an edge to a network connecting two nodes. If the nodes are not present, the function will add the missing nodes to the network.
- number_of_nodes() - This function returns the number of nodes in a network.
- number_of_edges() - This function returns the number of edges in a network.
- average_clustering() - This function returns the average clustering coefficient across a network.
- degree_histogram() - This function returns the degree distribution of a network as a histogram.

3.2 Basic model functionality

The construction of the model was heavily inspired by random walkers and cellular automatons. The goal was to make a simplified representation of worker bees moving about in a hive and registering their interactions, and as such digitally recreating Gernat's experiment. The hive would be represented as a matrix, and the bees would be represented as agents in the matrix. First, a simple framework model was constructed, and then the model was tweaked and customised in an attempt to better capture observed features from the empirical data.

The basic hive model was coded with three basic functions in mind; Defining and placing bee agents, moving the bee agents, and recording communication between agents. The design of the different parts of the code was kept as simple as possible to allow for customisation later.

Bee agents were defined as objects with certain attributes. The bees would need names for two reasons. The first is to provide a key that connects its marker in the matrix to its object attributes. The other is to make it possible to differentiate what pairs make up interactions. The bees placement in the hive matrix was also defined as attributes, with each bee having an x and y coordinate in the matrix. The initial placement of the bees were randomly assigned. Any overlapping was dispersed by simply moving the bees for a number of iterations without recording any interactions. The final of the basic attributes provided was a communication cooldown factor. A sort of counter that would increase in value with each iteration. This was implemented to prevent bees from interacting with the same bee too quickly. Interactions would only be recorded if the counter was high enough, resetting the counter after a successful interaction.

Movement of the bees was modelled after random walkers, but initially, two different variants were tested. In the first variant, a bee would make an attempt at moving to a space in its moore neighbourhood[9]. If the targeted cell was empty, the bee would move, and stay in its previous cell if the new cell was occupied. In the second variant, a bee would register all vacant cells in its moore neighbourhood, including its own cell, then choose one of these cells and move into it. After movement, the bee's cooldown counter was increased by one. The bees were all moved sequentially, but their order was shuffled with each iteration. The movement algorithm would also prevent bees from moving outside the hive matrix boundaries, as well as prohibiting them from moving into the same cell.

Interactions were recorded between neighbouring bees, provided both bees were "available". Bees could only be part of one interaction in a given iteration. As mentioned, this was done by only allowing bees with a high enough cooldown counter to interact, and then resetting the counter after registering the interaction. To interact, bees would register all other bees in their moore neighbourhood, and randomly choose one of these bees. If both of these bees were ready for interaction, as in their cooldown counter was high enough, their interaction was registered into a text file. After their interaction, both their cooldown counters were reset to zero. Bees would look for interactions sequentially in the same order as they were moved in any given iteration.

4 Results and analysis

This section contains discussions on various decisions made when the model was programmed, as well as results from analyses of both empirical data and models. Results are presented in the order of their relevance to the development of the model, and they are discussed briefly after. The aim of the analyses was to find similarities and differences between the empirical data and data generated by the model. The results of the analyses was then used to enhance the model.

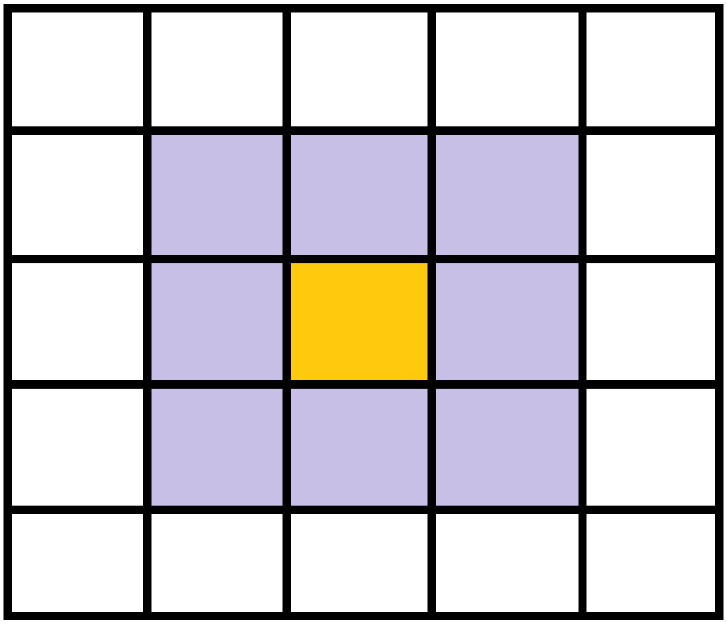

Figure 1: An illustration of a moore neighbourhood. The yellow square represents a bee agent, and the lilac squares represent the agent's Moore neighbourhood.

4.1 Moving the bee agents

The model itself was based on a random walker model. In a random walker, an object or agent moves randomly either in an infinite area or volume, or within boundaries. A random walker may exist in several different dimensions, but for this model a two dimensional matrix was used. This was done to mimic the single layer beehive from Gernat's experiment.[7] A hexagonal grid could have been used for thematic purposes, but a square grid was chosen for simplicity.

The bee agents were set to move as though random walkers in the hive matrix. it was undesirable for bees to move into the same spot in the hive matrix, as the communication function was programmed to function based on adjacency between bees agents. The solution to this required that bee agents were unable to move into the same cell. Since the hive also had boundaries, this would mean the movement of bees would be gated by two factors: The bees could not move outside of the matrix, and the bees could not move into cells already occupied by another bee.

Bees could be moved in two ways. Agents could either attempt to move into a random space, or they could identify vacant spaces before moving. For the first method, the bees could check a random cell in their immediate Moore neighbourhood and move into that cell if it was vacant, standing still if that cell was already occupied. If the chosen space was vacant, the bee would move into that space. If the space was already occupied by another bee, the bee would stay in its current space. Thus, a bee agent with no neighbours had an equal $\frac{1}{9}$ chance of moving into any neighbouring cell or stay put, while a bee with more neighbours would have a higher chance of staying put. Alternatively, the bees could check every cell in its neighbourhood to see what cells were vacant, and then move into one of those. This was done by creating a temporary list of empty cells in an agent's Moore neighbourhood every time it was going to move, including its current space, and then choosing one of those options at random. Using this method, an agent would always have the same possibility of standing still as moving into a nearby vacant cell.

The main difference in these two ways of moving the bees is in how movement of bee agents is affected by other bee agents in their neighbourhood. In the first method, a bee with several occupied neighbouring cells will have a higher chance of not moving. The bee can't move if it checks an occupied cell, and if several cells are occupied, the chances of it checking one increases. This means that at some point, a bee could have a higher chance of not moving compared to the chance of it moving to a given square. Applying this logic to the example provided in figure 2 would mean the yellow agent has a $\frac{4}{9}$ chance of not moving. Three of the possible cells are occupied, and it also has an innate chance of staying put. The other method will always have an equal chance of moving to a given square and standing still, meaning there is a greater chance for surrounded bees to move. In the example provided in figure 2 the yellow agent has an equal

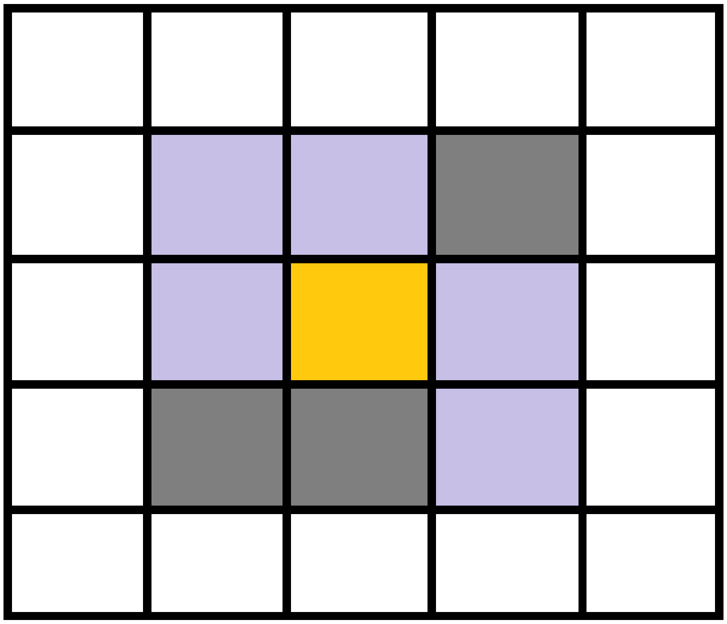

Figure 2: An illustration of a bee agent surrounded with three neighbouring bee agents. The yellow square represents the surrounded bee agent, while the grey squares represent neighbouring agents. The lilac squares represent the yellow agent's possible movement options.

$\frac{1}{6}$ chance of moving into any one of the vacant lilac squares or standing still. The impact of the different movement options is discussed in a later section. Both movement styles were made into separate models for analysis.

4.2 Parameter testing

The initial models needed to generate a number of interactions that would be comparative to the number of interactions in the empirical data set. The trials contained in the empirical data sets contained between 190000 and 330000 interactions, on average about 258000 interactions. The models would have to generate a similar number of interactions in order to compare empirical and synthetic data. At the same time, the model would have to go on for a satisfying amount of iterations. The iterations would be the equivalent of time stamps

An initial concern was striking a balance between the number of interactions and the number of iterations. Getting a high number of interactions was rather simple, but spreading them across a sufficient number of iterations was more complicated. 15000 iterations were chosen as a sufficient basis, then the number of interactions were adjusted to this. Initially the models would generate too many interactions, so several tweaks were made to the models in an attempt to reduce the number of interactions.

In the models, the generating of interactions depended heavily on the closeness of the bee agents. This because the interactions are recorded between neighbouring bees. To reduce So a way of reducing the amount of interactions would be to make more space between bees. This translates to increasing the size of the hive. initially, the hive was 50 by 50 cells, but this would mean just over half the cells were vacant with 1200 bees in the hive. This was changed into a matrix of 100 by 100 cells, quadrupling the available space. The total number of interactions was not significantly impacted by this change, but the change was kept nonetheless. While not analyzed directly, the increased spacing between bees might have had an effect on the frequency of pairwise interactions.

Another way of reducing the number of interactions was done in the form of increasing the threshold for the "communication factor" in the communication function in the models. Initially, the bee agents could communicate after three iterations of no interaction, but this was increased to allow communication only after ten iterations of no interaction. This, as with changing the size of the hive, did not significantly reduce the number of interactions. The change was still kept, as it could be argued that this change, along with the change in matrix size might facilitate bees choosing different partners between interactions.

The most impactful change in the models in regards to reducing interaction numbers was implementing a probability function to determine whether or not interactions would occur. Before communications were registered, a check would be implemented such that only a certain percentage of interactions would proceed to the actual registration of interaction. This would not prevent bees from communicating in subsequent iterations. If a bee agent was in the proper state to interact, but did not pass the stochastic check, it would still be ready to interact in the next iteration. Choosing 10% as the probability of interaction brought the number of interactions down to the desired range. After reaching the desirable amount of interactions, this probability function was changed further, such that it would vary over time. This was done to mimic the circadian rhythm of the bees seen in the preliminary work leading up to this thesis. [30] The circadian rhythm was approximated though the use of a sine function. The implementation would also make it possible to define a number of days the iterations would represent. For the purposes of this thesis, Ten days were set as a baseline. This would imply that a single iteration represented about a minute of real time with the 15000 iteration base.

4.3 Heatmaps of pairwise interactions

After making the initial models, the data they produced were compared to the empirical data by means of heatmaps. First, heatmaps were made for all the empirical data. The heatmaps were made such that the intensity of the heat correlated to how many interactions a given pair of bees made during a trial.The interactions in the empirical data were treated as undirected interactions, and as such the heatmaps were only filled in the top half.

Figures 3 through 7 show the heatmaps made from the empirical data, showing the amount amount of pairwise interactions as heat. It is evident from the heatmaps that no particular pattern seems to form, and no single bee stand out as particularly communicative. If this were the case, one could imagine having brighter lines the heatmaps, but these are not present. The heatmaps show some pairs being especially communicative compared to the rest of the pairs. These communicative pairs reach upwards of 20 interactions, while the more general bee pairs stay under 5 interactions.

Figures 8 and 9 show the heatmaps for the two variants of the initial model. Compared to the heatmaps from the empirical data, these initially seem way busier. They seemingly show a lot more heat, but the scale is different between the heatmaps. The empirical heatmaps have some pairs that communicate a great deal more than the rest of the pairs, but this does not seem to occur in the heatmaps for the models. To better understand the patterns of the heatmaps, the pairwise interactions were plotted as distributions.

Figure 10 shows the distribution of pairwise interactions of both the empirical data sets and the initial models. From this its easier to see where the models and the empirical data stray from each other. For the lower interaction numbers, the models and the empirical data match up rather well. And, seeing as the distribution is on a semilog plot, the majority of the pairs have lower

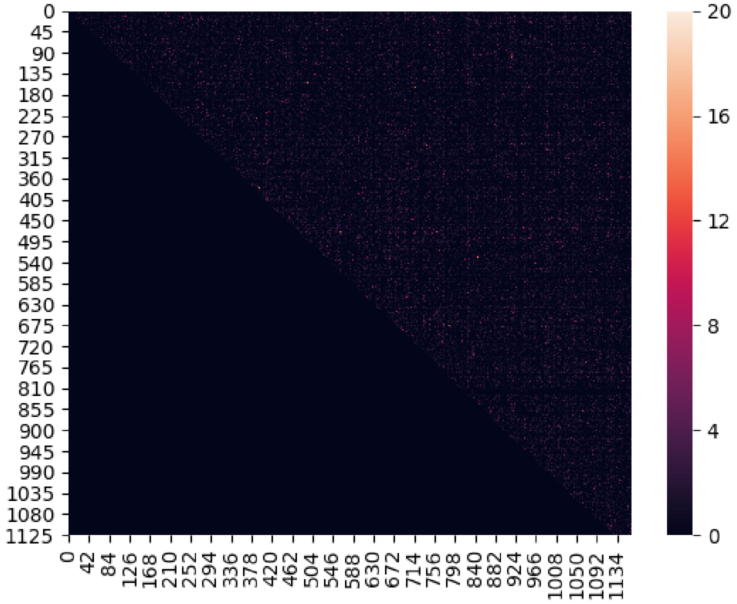

Figure 3: Heatmap for the first set of empirical data. The heatmap was not mirrored, such that the lower half is empty.

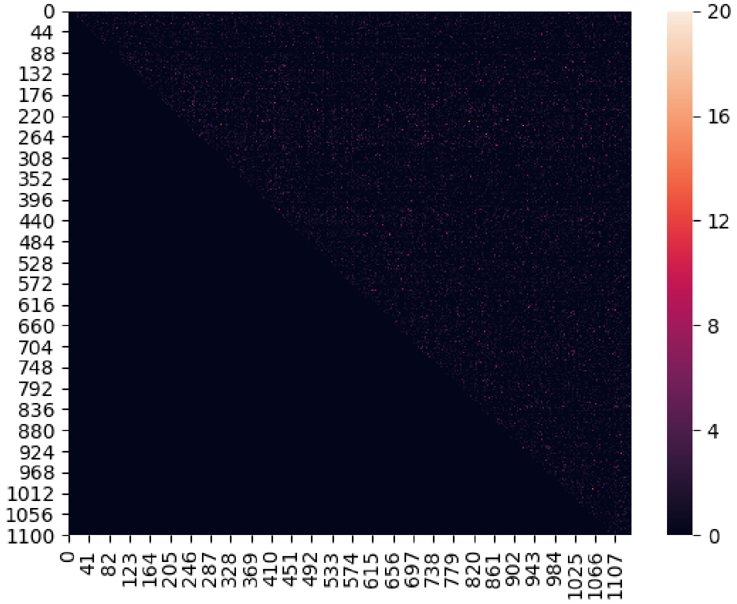

Figure 4: Heatmap for the second set of empirical data. The heatmap was not mirrored, such that the lower half is empty.

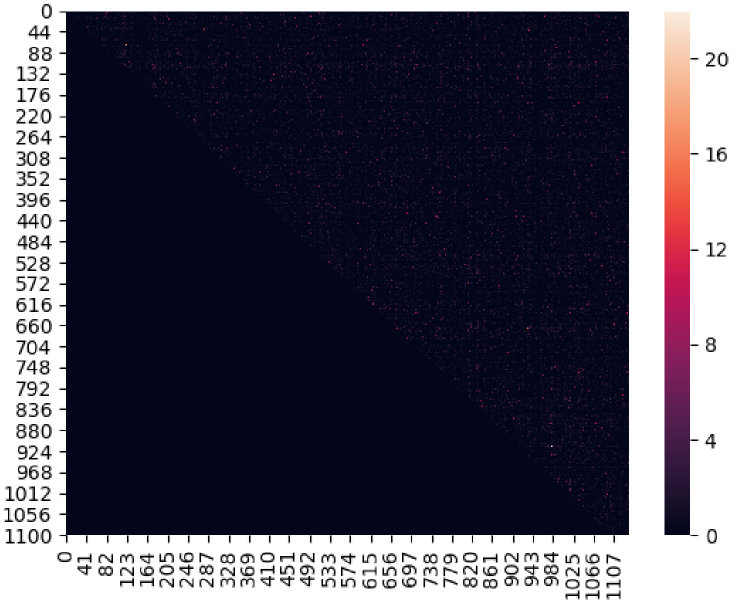

Figure 5: Heatmap for the third set of empirical data. The heatmap was not mirrored, such that the lower half is empty.

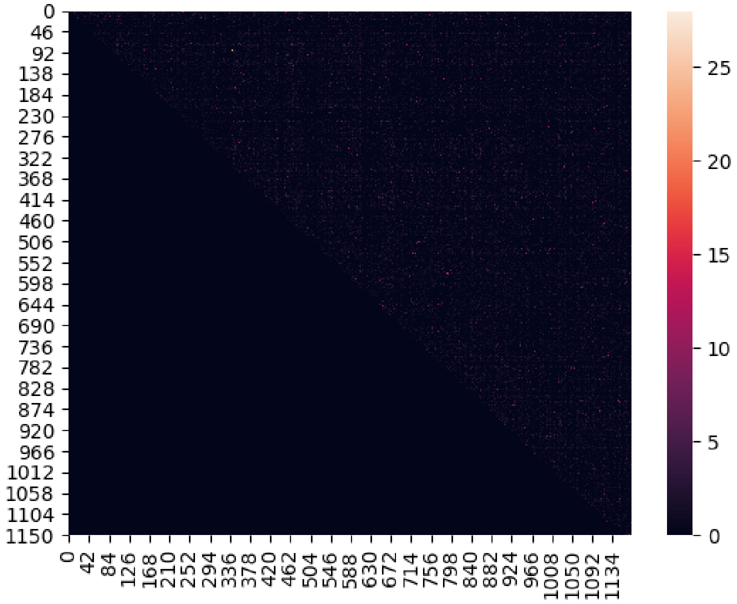

Figure 6: Heatmap for the fourth set of empirical data. The heatmap was not mirrored, such that the lower half is empty.

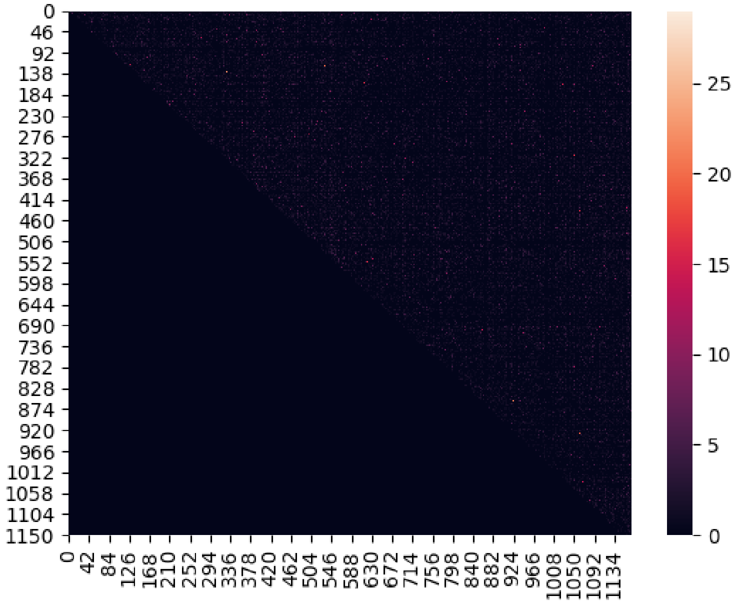

Figure 7: Heatmap for the fifth set of empirical data. The heatmap was not mirrored, such that the lower half is empty.

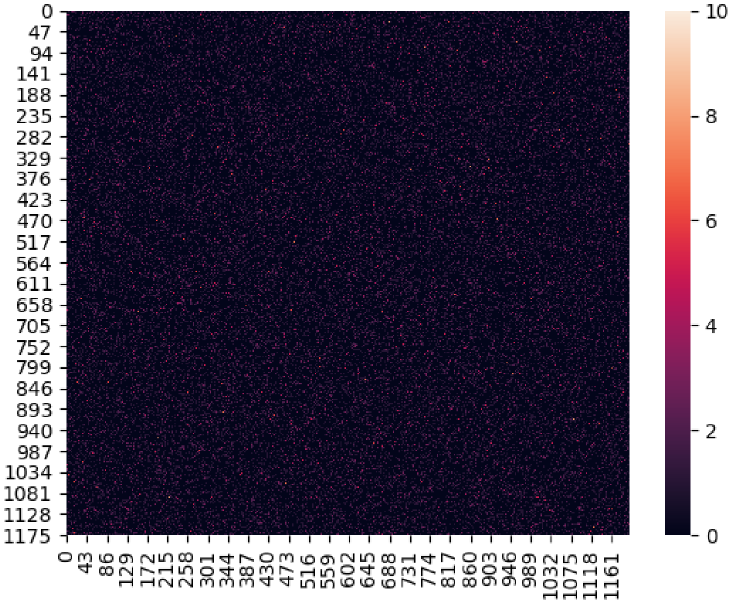

Figure 8: Heatmap for the initial model with the first movement option. The heatmap is mirrored on the diagonal going from the upper left corner down to the lower right corner.

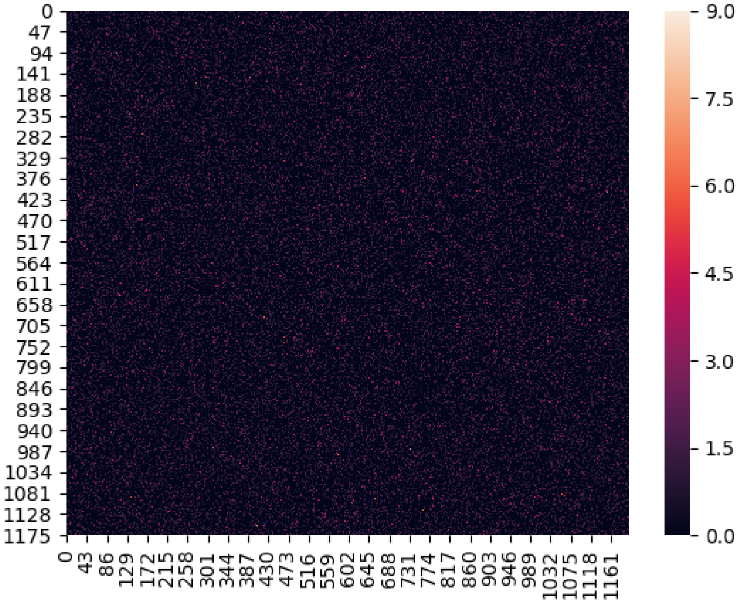

Figure 9: Heatmap for the initial model with the second movement option. The heatmap is mirrored on the diagonal going from the upper left corner down to the lower right corner.

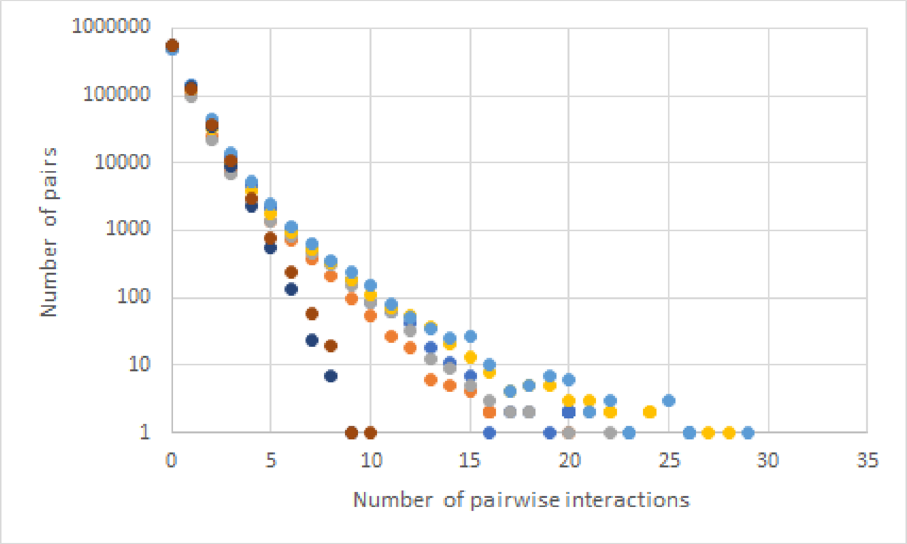

Figure 10: Displays the distribution of pairwise interactions between bees plotted as a semilog plot. The yellow, blue, grey, orange and light blue dots represent the pairwise interactions from the empirical data, while the darkest blue and red dots represent the pairwise interactions from the models. Red dots represent the first movement option, while the blue dots represent the second movement option.

amounts of interactions. From this, the models arguably capture most of the pairwise interaction. However, the long tail seen in the empirical data is absent in the models. From this, it was evident that a further delve into what caused the long tails to appear would be necessary.

4.4 Initial models as a null hypothesis

The goal of the modelling was simplifying the behaviour in the hive, while staying true to the conditions of the experiments. Since all bees registered in the empirical data sets were worker bees, all agents in the model would behave in the same manner with no distinction. This also assumes that worker bees communicate in roughly the same manner regardless of current task, be it nursing, foraging, and so on.

When making the initial models, it was assumed that patterns in the social interactions of the bees could be attributed to random walks. This assumption is only partly correct. Analysis on both the empirical data sets and the synthetic data generated from the models showed that there was a clear discrepancy in the distribution of pairwise interactions. The null hypothesis model captured a big part of the bee behaviour fairly well, but there was clearly some behaviour that could not be explained by a random walker model alone, as seen in figure 10. In particular, the long tail seen in the pairwise interaction distribution showed that some bees in the empirical trial would form more "extreme" pairs. These relatively few pairs communicated with each other many times more than the average bee. These extreme pairs would be investigated further to see if they behaved in a particular manner. If the extreme pairs had some defining traits, those traits could then potentially be implemented into the modelling.

Since the null hypothesis of random walkers was unable to capture all the dynamics of the empirical data, further analyses were carried out with the aim of finding specific traits in the empirical data missing from the models. The end goal was finding traits that could easily be implemented into the base models to better match the empirical data.

4.5 Analysis of extreme bees

The random walker model seemed to capture most of the behaviour seen in the pairwise interactions of the bees, but did not capture the behaviour of some extreme pairs. The extreme pairs made several more pairwise interactions compared to the average bee population. To better understand how this could be implemented into the hive model, further analysis of the extreme pairs were done. This was done in hopes of uncovering some defining traits of the bees in these particular pairs.

An initial thought was to see what amount of interactions the bees in the extreme pairs were part of. The logic behind this was that if the extreme bee pairs made many pairwise interactions, they might be defined by having many interactions overall. This was done by using the degree distribution of the networks generated from the empirical data sets. The total amount of connections to other nodes in the network would be equal to the total number of interactions any given bee made during the experiment. These degree distributions were then made into histograms that would represent the distribution of the total number of interactions for each bee. As the initial degree distributions for the whole networks were rather thinly spread, the numbers were gathered into larger bins to better give an image of how the number of interactions were distributed. After generating these distributions, the individual bees from the extreme pairs were identified, and their total number of interactions were found. These numbers could then be identified in the histograms.

Table 1: Pairwise interactions and total individual interactions for the extreme bee pairs that were investigated in the empirical data sets. The names of the bees is their assigned number identifier.

Trial	Pair	1		2		3		4		5	
1	Names	501	1148	1284	1633	1591	1945	539	772	1270	1923
	Pairwise interactions	20		20		19		18		18	
	Individual total interactions	716	873	340	549	606	636	552	525	810	783
2	Names	812	1757	682	1183	1423	1812				
	Pairwise interactions	20		16		16					
	Individual total interactions	462	456	333	430	305	363				
3	Names	779	1906	239	1673	1830	1796	80	1936		
	Pairwise interactions	22		21		21		20			
	Individual total interactions	493	276	424	787	517	824	534	403		
4	Names	683	1749	213	1596	323	374	874	865		
	Pairwise interactions	28		27		24		24			
	Individual total interactions	386	624	776	475	831	486	958	426		
5	Names	800	1164	717	1014	1019	1659	1385	1403	1715	1913
	Pairwise interactions	29		26		25		25		25	
	Individual total interactions	633	541	1062	693	973	506	673	505	525	695

Figures 11 through 15 show the degree distribution of the empirical data. The degree distributions were compressed into larger bins to show clearer trends.

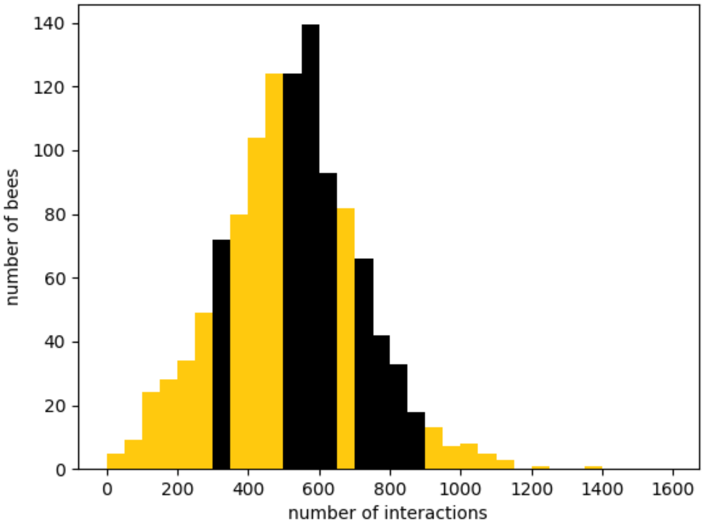

Figure 11: Degree distribution of the first empirical data set compressed into bins with size 50. The black bars signify a bin containing one or more of the bees in the extreme pairs.

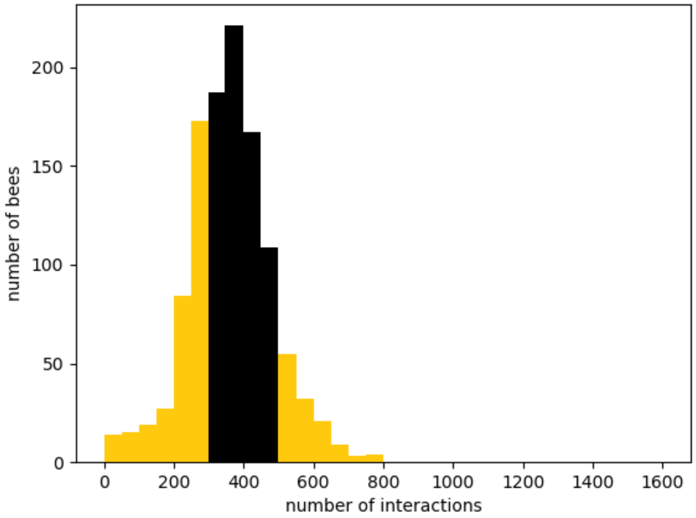

Figure 12: Degree distribution of the second empirical data set compressed into bins with size 50. The black bars signify a bin containing one or more of the bees in the extreme pairs.

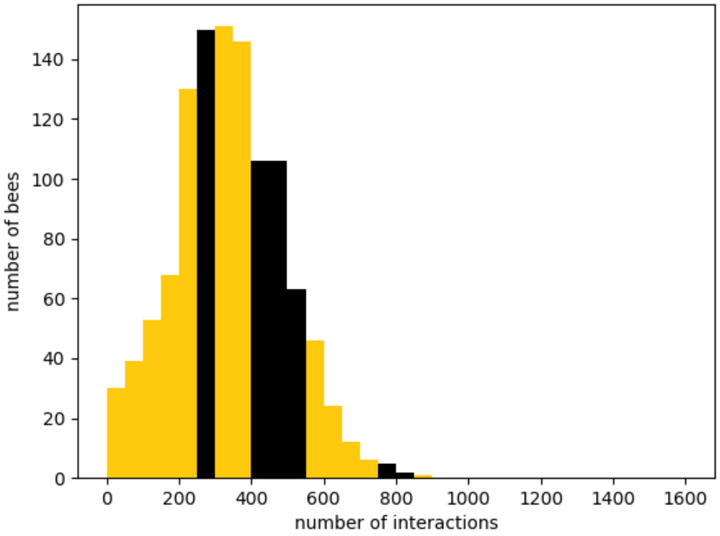

Figure 13: Degree distribution of the third empirical data set compressed into bins with size 50. The black bars signify a bin containing one or more of the bees in the extreme pairs.

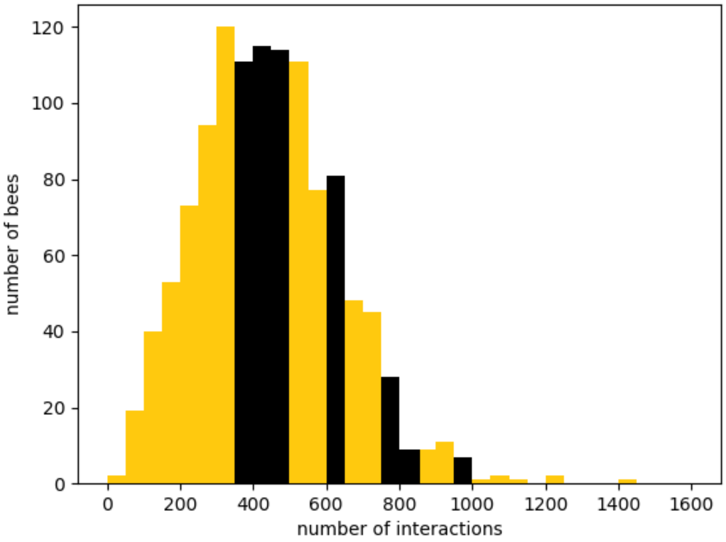

Figure 14: Degree distribution of the fourth empirical data set compressed into bins with size 50. The black bars signify a bin containing one or more of the bees in the extreme pairs.

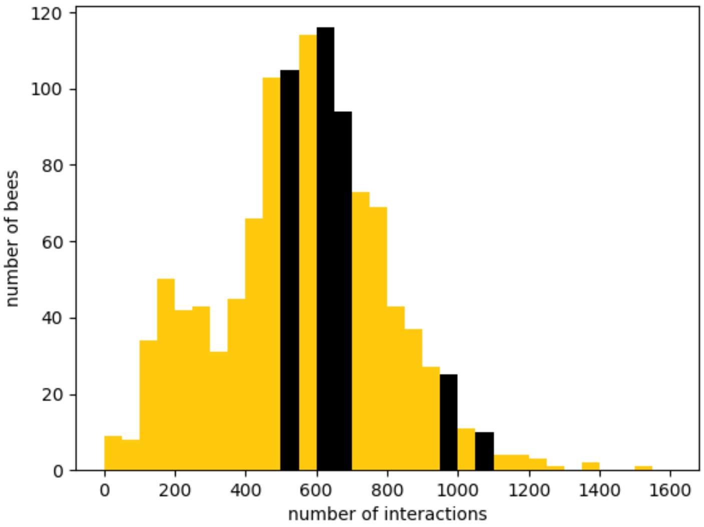

Figure 15: Degree distribution of the fifth empirical data set compressed into bins with size 50. The black bars signify a bin containing one or more of the bees in the extreme pairs.

Bins containing bees that were part of the investigated pairs with notably high pairwise interactions were marked by colouring them black. From the histograms, it was clear that the extreme bee pairs didn't share any specific traits in regards to the total number of interactions. The bees were haphazardly spread about the histogram with no clear pattern, and in particular, they did not seem to gather to any of the extreme ends of the histograms. As such, one could argue that the behaviour of the extreme bee pairs were not defined by the total number of the individual bee's interactions.

Another way of defining the behaviour of the extreme bees was investigated in the frequency of their interactions. The bursty interaction patterns of the bee social network is discussed in Gernat's article.[7] Networks with bursty interaction patterns are defined by their unique temporal spreading of interactions. Interactions between pairs are very frequent within short windows of time, but these interactive phases are separated by long periods of no or little interactions. The article also makes a distinction between long waiting times and short waiting time, with the threshold being 168 seconds. By investigating the waiting times between pairwise interactions in the extreme pairs, the pattern might be further defined. The waiting time was determined by identifying the starting time of interactions in the empirical data, and then finding the difference in starting times for subsequent interactions. This approach does not accurately capture waiting times between interactions, but for the purpose of identifying

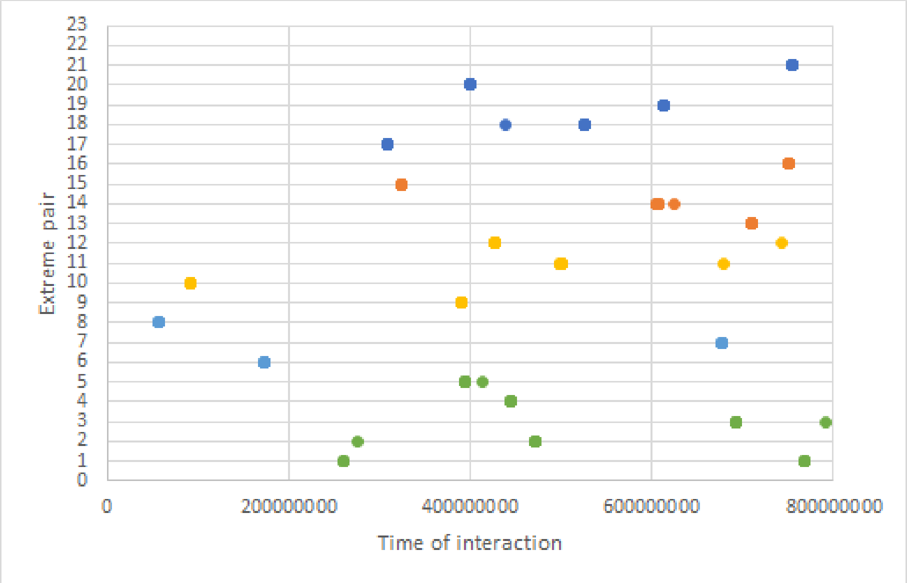

Figure 16: Interactions in extreme pairs of bees plotted at the time of the interaction after experiment start. The green dots represent pairs from the first trial, light blue represent pairs from the second trial, yellow represents pairs from the third trial, orange represents pairs from the fourth trial, and dark blue represents pairs from the fifth trial.

bursts in the extreme pairs, the difference is negligible. In the analysis of these patterns, only the specific extreme bees were investigated.

Figure 16 show that the bursty interaction patterns is apparent from the pairwise interactions in the extreme bee pairs. In several of the pairs, most of, if not all, the interactions are gathered into one burst of sequential interactions. In fact, the interactions are so tightly gathered, they're hard to tell apart when plotting them. Some of these interactions are separated by as little as seconds.

Figure 17 shows a more enhanced visualization of the bursty interaction between the two bees 1591 and 1945 in the first empirical data set. The longest gap between two start times in this figure is a mere 84 seconds. This pattern is common for all the bursts seen in figure 16

In the supplementary information of Gernat's article, it is explained that trophallactic interactions can last for periods of time as long as minutes.[7] Therefore, interactions that were separated by very small amounts of time were combined into larger interactions. Gernat et al. made sure to account for factors that might have separated interactions, such as bees blocking the vision of b-

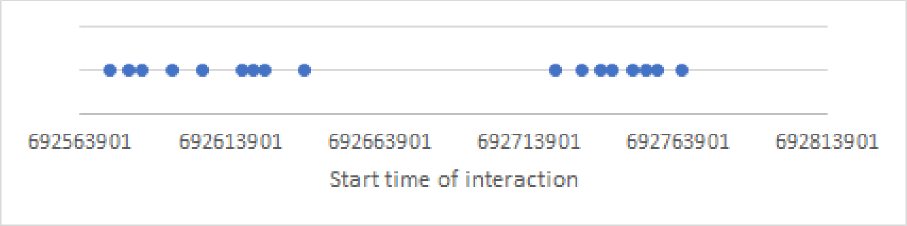

Figure 17: A zoomed in view of interaction start times in the extreme pair 1591 and 1945.

codes. Since these factors were accounted for, the separated interactions must imply that the honeybees make several unique trophallactic interactions with a partner. As in, they are indeed separate interactions, and not a continuous interaction. Now, what separates these interactions could be one of two factors that might have an implication on how a model should behave. The interactions could be separated by silence, meaning that interactions simply come to a stop before starting a new interaction with the same partner. Or, the interactions might be separated because of a change in communication partner. Therefore, this must be investigated. To do so, a further investigation of the five extreme pairs from the first empirical trial was done. The extreme pairs that were investigated further did not swap partners in their bursty phases, implying, but not confirming, that the bees do not swap communication partners during bursty phases. It was assumed from this that bees in bursty phases did not swap partners when the models were enhanced further.

4.6 Network analysis of synthetic data

A great deal of information can be found in seeing how network topology changes in time. The empirical data sets had previously been analysed to find how some basic attributes of the honeybee social network developed as the experimental trials progressed.[30]

An attribute of particular interest is clustering in the network. Clustering describes the tendency of a given node's neighbouring nodes to be connected.[23] Real networks are usually found to have a higher clustering than randomized networks, so comparing the clustering in the empirical networks to the clustering in the synthetic data might give a pointer on how close or far off the model currently is. Clustering also has a tendency to decrease with increasing size in networks. Clustering in the empirical data had previously been analyzed. This was done by using the empirical data as a basis for temporal networks. All interactions within one hour time intervals were made into networks and then analyzed, finding network attributes such as number of nodes present, connectivity and clustering. The same was done with the synthetic data from the model, focusing on the clustering. A number of iterations equivalent to one

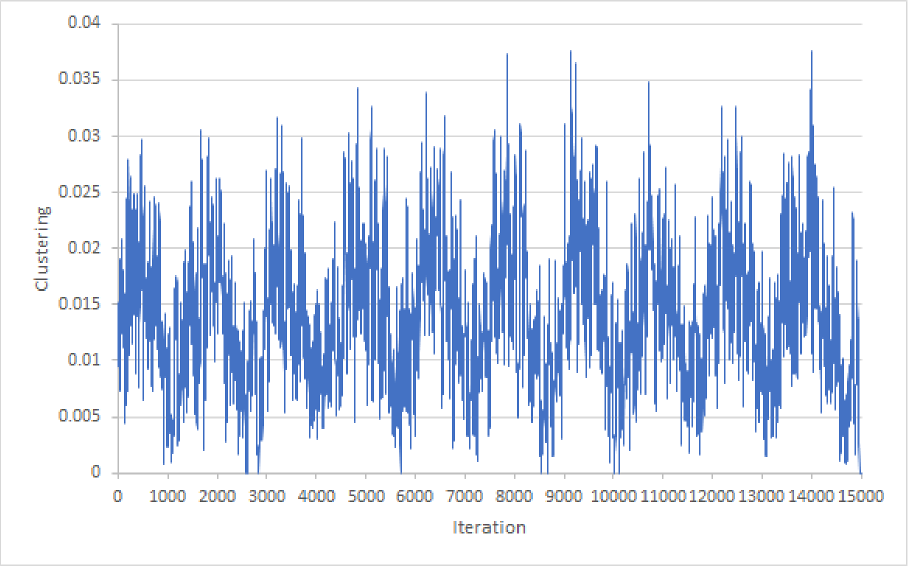

Figure 18: Clustering of networks generated from the initial model using second variant of movement where bee agents registered vacant spots.

hour worth of data were made into networks and analysed, repeating over the data set, creating a temporal network analysis.

From figure 18, it is evident that the clustering of the empirical data and the synthetic model is rather similar. The wave pattern in the figure stems from the day/night cycle that was implemented into the model, but the clustering seems to hover around 0.015. This was strikingly similar to the clustering found in the analyses of the empirical data.[30] This implies that the null hypothesis model captures multiple aspects of the empirical network, even if it consists of only random walkers.

4.7 Burstiness analysis

Burstiness seemed to be a prominent aspect of the bee communication behaviour. Therefore, further analyses of the dynamics of the burstiness in the bee network were performed to get a better understanding of the burstiness with the aim of using the gathered information to improve the accuracy of the hive model.

As described in part 4.5 and 4.6, it was found that the null hypothesis model captured some aspects of the communication dynamics, such as the clustering being similar, and most of the pairwise interactions were similar. The main difference between the model and the empirical data was the long tail seen in the pairwise interactions seen in figure 10. As an experiment, the bursty interactions were removed from the empirical data sets and then these new ”censored”

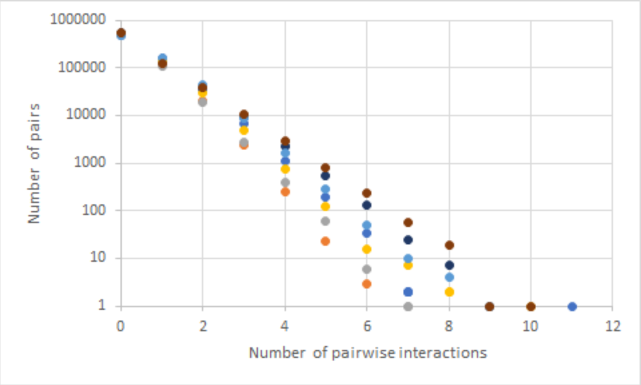

Figure 19: Displays the distribution of pairwise interactions between bees on a semilog plot after removing bursty interactions from the empirical data. The yellow, blue, grey, orange and light blue dots represent pairwise interactions from the empirical data, while the darkest blue and red dots represent the pairwise interactions from the two random walker models.

data were analyzed. The bursty interactions in the empirical data sets were defined by using the threshold of ”short” waiting times that was determined in Gernat’s article. The article defines short waiting times as waiting times between interactions shorter than 168 seconds. Using this, a new data set was made by omitting any data of duplicate interactions between pairs where less than 168 seconds passed from the end time of a given ”first” interaction to the start time of the next interaction. This data set was then sent through the same analysis that was used to find the numbers of duplicate interactions.

Figure 19 when compared to figure 10 makes it evidently clear that the bursty interactions is the main contributing factor to the long tail in the pairwise interactions seen in figure 10. With no burstiness present, the empirical data match up rather well with the null hypothesis model of only random walkers. The random walker models have slightly higher numbers than the empirical data, but the data seem to follow roughly the same linear trend. A possible explanation to why the model data are higher than the empirical data, is that any ”natural” bursts that would occur in the random walker models haven’t been removed. Regardless, what separates the empirical data from the model data is the lack of bursty interactions in the model. To make the models match the empirical data better, some sort of bursty interaction should be implemented into the models.

Since burstiness was so characteristic for the empirical data, the burstiness dynamics were investigated to find some trends that could be implemented into the model. Two main concerns were investigated. Since not all interactions in the bee network could be bursty, something must determine whether or not interactions end up being bursty. It was decided that the frequency at which bursty interactions occur should be determined. The bursts themselves are not equal, some having more interactions that others. A way of investigating this would be making a distribution of the burst lengths to determine how often certain lengths of burst would occur.

While what specifically makes an interaction bursty is not within the scope of this thesis, an argument could be made that there is a chance of any given interaction between bees ending up in a bursty phase. Two concerns need to be cleared up before determining this chance. Are bees in bursty phases only communicating with one specific partner bee, or do they possibly change partners? Also, what is the probability of an interaction being bursty?

First, when the bursty interactions from the empirical data were removed, bees changing partners between interactions were not accounted for. As explained in the supplemental information to Gernat's article, separate trophallactic interactions were registered as one interaction, provided that they were close enough temporally.[7] However, this must mean that all interactions in bursts are truly separate interactions, leaving open the possibility of bees changing partners in the short time frame between repeat interactions. An argument could be made that this is merely a question of how bursty interactions are defined, but it might still be valuable to take a look into this. When the extreme bee pairs were found earlier, the pairs from the first parallel in the empirical data were investigated in detail for the time intervals where these specific pairs displayed bursty behaviour. In these five pairs, the bees did not change partners at any point in their bursty phases, implying, but not conclusively demonstrating, that bees in bursty interactions do not change partners before their bursty phase is over. Thus, It was assumed that this would hold true for all other bursty interactions in the bee network.

Next, What is the probability of any interaction resulting in a burst? To figure this out, some basic calculations of probabilities were necessary. This was done by determining the number of interactions resulting in a burst and dividing this by the number of total interactions without bursts present. For the purposes of this calculation, the number of interactions resulting in bursts would be equal to the number of bursts present.

Table 2: Shows the number of bursts registered for each of the empirical trials, the number of interactions in the trials with no bursts, and the chance of an interaction resulting in a bursty phase.

Trial	1	2	3	4	5
Number of bursts	25900	20184	18707	23452	28655
Number of interactions	255458	169119	155101	216159	278364
Chance of burst	0.1014	0.1193	0.1206	0.1085	0.1029

From table 2 the probability of two bees entering a bursty phase is found, and it varies between empirical trial. An average of these probabilities, and its deviation, can be found.

$$\frac{0.1014 + 0.1193 + 0.1206 + 0.1085 + 0.1029}{5} = 0.1106 \tag{3}$$

$$\frac{\sqrt{(0.1106 - 0.1014)^2 + (0.1106 - 0.1193)^2 + ...}}{5} = 0.0036 \tag{4}$$

$$\sqrt{0.0036} = 0.06 \tag{5}$$

Equations 3 through 5 show the calculation of the average chance of an interaction being bursty and the variance and standard error of this number. From this, the chance of an interaction being bursty is about 11%. Incorporating this into the model would require a simple check, but no substantial changes, which was desirable.

When the bees enter a bursty phase, they interact a number of times that varies with every burst. In order to implement this into a model, the tendency of interactions to last for fewer or more interactions must be defined. When a pair of bees enter a bursty phase, that phase lasts for a certain time, and results in a certain amount of interactions between the two bees. This can be visualized using a histogram showing the number of bursts resulting in every possible number of interactions. This was done using the empirical data sets.

Figures 20 through 24 show the the number of interactions in bursty interactions between pairs of bees as histograms. The histograms display a clear geometric trend where lower numbers are more likely, while larger numbers are comparatively less likely.

4.8 Model enhancement

Analysing the empirical data suggested that burstiness was the missing factor in the initial models. The aim of the investigation of patterns in the burstiness was to find defining traits that could be implemented into a model in order to

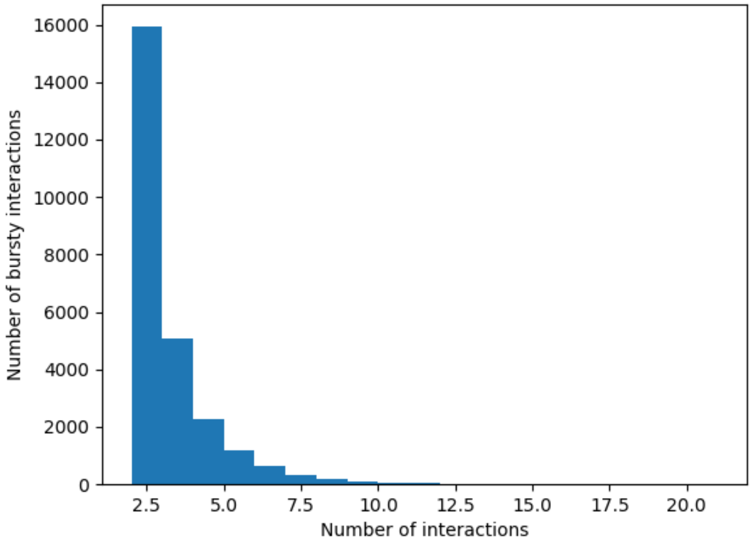

Figure 20: Shows the distribution of number of interactions in bursts found in the first empirical trial as a histogram.

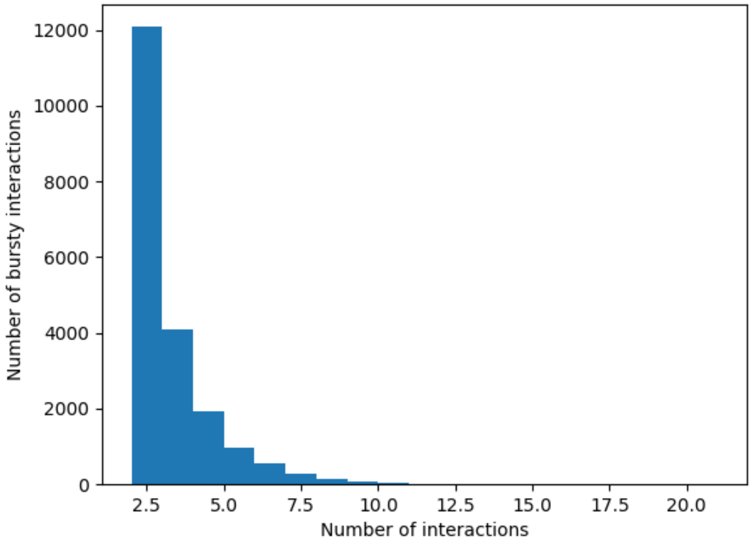

Figure 21: Shows the distribution of number of interactions in bursts found in the second empirical trial as a histogram.

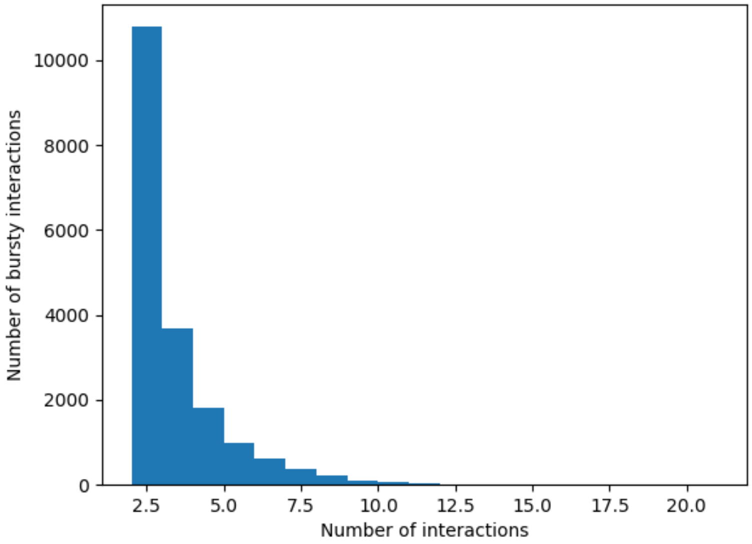

Figure 22: Shows the distribution of number of interactions in bursts found in the third empirical trial as a histogram.

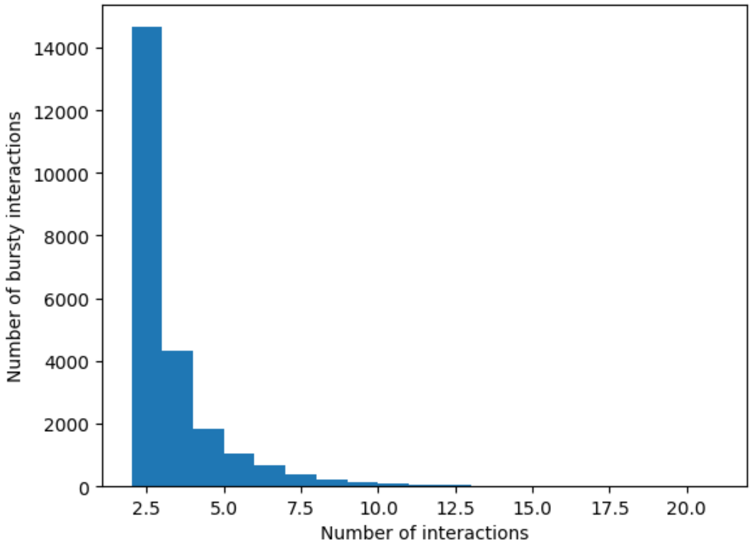

Figure 23: Shows the distribution of number of interactions in bursts found in the fourth empirical trial as a histogram.

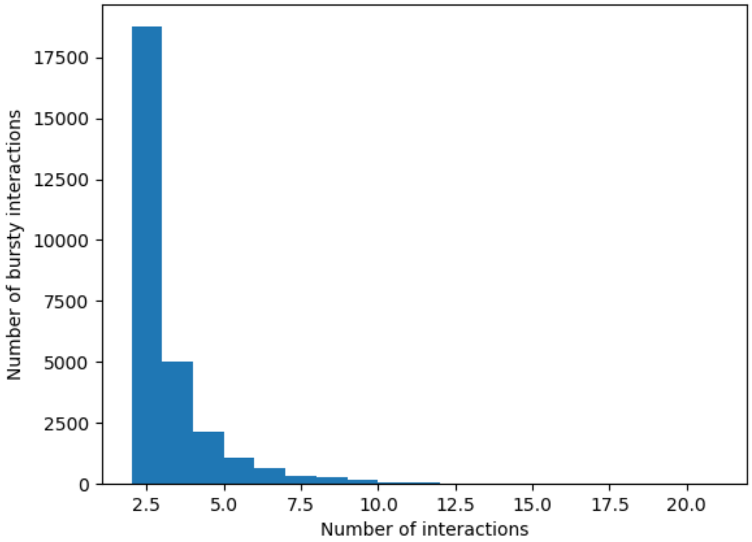

Figure 24: Shows the distribution of number of interactions in bursts found in the fifth empirical trial as a histogram.

enhance it. Ideally, the enhanced model would more accurately describe the interaction patterns in real life beehives.

The model was enhanced to capture the bursty interaction patterns seen in the empirical data by replicating the rates at which bursty interactions would occur, and the length of the bursts. This was accomplished by giving every interaction a chance of ”locking” a bee agent to its communication partner for a certain amount of interactions. The chance of a pair locking together was set to the average probability found in table 2. After locking, the bee agents were assigned a number of interactions to reach. This number was found by pulling a random number from the combined numbers of interactions seen in figures 20 through 24.

For as long as bee agents were locked to a partner, they would cease movement in the hive matrix. Bursty phases in the empirical network were defined by the waiting times between interactions being less than 168 seconds. In the model, every iteration would equate to approximately a minute when using a basis of 15000 iteration over 10 days. Thus, communications in ”bursting” pairs would have to occur at least every three iterations. Since the waiting time could be shorter, paired bee agents were set to have a 1/3 chance of interaction on the next iteration after entering a burst, then a 1/2 chance on the second, and

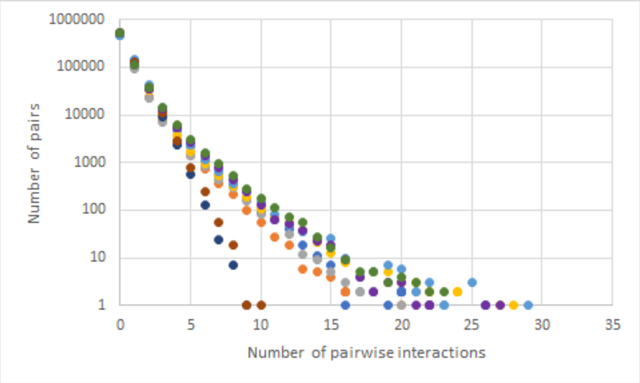

Figure 25: Displays the distribution of pairwise interactions between bees on a semilog plot. The yellow, blue, grey, orange and light blue dots represent pairwise interactions from the empirical data, while the darkest blue and red dots represent the pairwise interactions from the two random walker models. The enhanced models are shown as dark green and purple dots. The purple dots represent the first movement option, while the green dots represent the second movement option.

a guaranteed interaction on the third iteration. Every interaction after pairing was counted in a separate variable assigned to the bee agents. When the bee pairs had made a number of interaction in their bursty phase equal to the predetermined number pulled from the distribution, they would exit the bursty phase. After exiting the bursty phase, the bee agents would behave as regular non-bursting bee agents and resume random walks across the hive matrix. The movement function and communication function of the non-bursting bee agents were kept largely unchanged. The only change to this version of the model was implementing the "locking"-behaviour to make the model generate a more bursty interaction pattern. The data output was also kept in the same format as before.

4.9 Analysing the enhanced model

The data output from the enhanced model was analysed in the same manner as with the empirical data and the null hypothesis model. This was done to compare the models and determine the accuracy of the enhanced model. First, the pairwise interaction for each pair was gathered and plotted in a similar way to that seen in figure 10.

As seen in figure 25, the enhanced models match up with the empirical data a great deal better than the initial models. As previously stated, the null hypothesis models are lacking the long tail of the empirical data. The enhanced models, however, seem to capture the long tail rather accurately. If anything, the enhanced models seem to generate numbers in the upper range of what was seen in the empirical data, but nothing conclusive can be gathered from this without a more statistical approach.

Next, the clustering of the enhanced models were calculated and plotted. The data generated from the enhanced models were made into networks that could be subjected to network analysis. In the same manner as before, 63 iterations at a time were made into networks, which were then analyzed. The average clustering coefficient from each network was then plotted sequentially with each corresponding iteration.

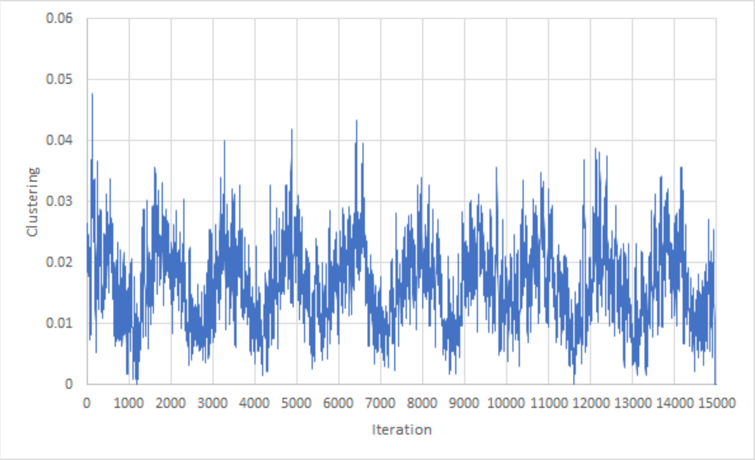

Figure 26: Clustering from the data generated by the enhanced model using the first movement option.

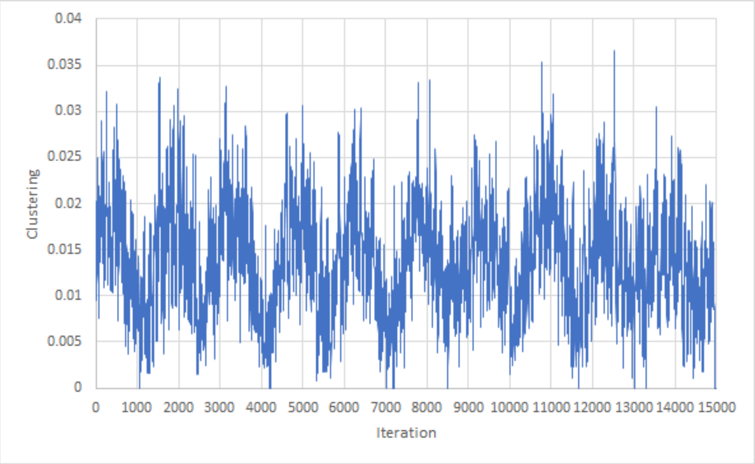

Figure 27: Clustering from the data generated by the enhanced model using the second movement option.

Comparing figures 27 and 26 to figure 18 its evident that the clustering is rather similar. This result is more or less expected, as increased burstiness will add more duplicated edges in a network. Increased clustering would require an increase in edges between different multiple unique pairs. The fluctuations in the data from the models are intentionally made to mimic the day and night cycle of the empirical data, but the clustering coefficients still tend to fluctuate around $0.015 - 0.02$. This is not far from what was previously seen in the empirical data.[30]

4.10 Discussion of enhanced models

The enhanced models captured a great deal more of the patterns seen in the empirical data. This is especially notable in the distribution of pairwise interactions, where the enhanced models were able to rather accurately capture the long-tailed behaviour seen in the empirical data. The clustering patterns in the networks matched up rather well with the empirical data in both the initial models and the enhanced versions. As mentioned before, it is expected for the clustering to remain mostly unchanged after the implementation of the bursty interaction pattern, as the burstiness alone will only add duplicate edged of already established edges. Thus, the clustering remains largely unchanged. An interesting observation here is that the random walker models alone were able to accurately match the clustering previously observed in the empirical data sets.[30]

Several assumptions and simplifications were made when constructing the models. The use of random walkers to represent bees in a hive seems to have worked out rather nicely in the end, but this still remains an approximation of actual bee movement. Bees in a real hive will not be constantly on the move, such as when they are sleeping or if they have no task to tend to. [3] Also, in the models the bees were assumed to behave in the same manner because they were all worker bees. However, due to age polyethism, worker bees have different occupations at different times.[15] It is easy to imagine that the age polyethism might cause different movement patterns in worker bees depending on their current task. As an example, foraging bees are not always present in the hive, meaning they have less time to perform trophallaxis with other bees compared to a nursing bee. Such nuance between subclasses of workers is not implemented in the model, which might lower the accuracy of the model.

The movement of the bee agents in the models seem to have little effect on the overall results. In the end, the first movement option where agents check a random nearby cell for vacancy seems to generate slightly higher numbers than the second movement option where agents locate all vacant nearby cells before moving into one of them. This is observed in both the initial models and the enhanced models. Whether any of the two movement options is better than the other is rather ambiguous. The main difference in the results being the first movement variant giving slightly higher numbers. However, the difference is so small that it is unclear whether higher numbers are undesirable or not. However, in terms of how the models function, the second movement variant might edge out the first variant in terms of customizability. The first variant is rather rigid in its functionality, but the second variant can be tweaked changing the probabilities of moving to certain cells or standing still.

It should be noted that at this point, the enhanced models are still operating with the same base parameters that were set previously. That is, the same number of iterations, the same matrix size, the same number of bee agents and so on. Some of these parameters depend on each other in the performance of the model. For example, the hive matrix needs to accommodate for the number of bee agents. And, as previously discussed, the number of interactions generated per iteration depend on several factors such as the probability check for interactions, hive matrix dimensions, etc. The models work with the current setup, but if the models were to be scaled up, several of these parameters would need to change. This parameter testing is outside the scope of this thesis,but is a definitive concern for any future use and development of the models.

5 Conclusion and outlook

Through the use of agent based modelling, the metrics from the honeybee social network in focus in this project were successfully reproduced in synthetic models. The factors most heavily supporting the notion of the modelling being successful was both the clustering and the distributions of pairwise interactions matching between models and empirical data. Most effort when creating the models was put into having the models recreate the bursty interaction patterns seen in the empirical data sets. As such, the bursty interaction patterns were analyzed in order to create a basis for the models. The distribution of bursty interaction lengths was found from the empirical data set, and it was determined that every interaction has an approximate 11% chance of being followed by a bursty phase. When these factors were implemented into the hive model, the model would match nicely with the empirical data.

The initial modelling was heavily based on the assumption that the bees could be modelled as random walkers on a matrix. While this did not sufficiently capture the long tail in the empirical data as seen in figure 10, it did compare nicely to the empirical data once any bursts in the pairwise interactions were removed as seen in 19. This indicated that a random walker was a good initial step in capturing the interaction dynamics of the honeybees. Tweaking the models such that they would generate bursty interactions of their own made a significant impact. The pairwise distributions for the enhanced models seen in figure 25 match the empirical data a great deal better than the initial random walker models.

The enhanced models do have room for improvement, however. Multiple assumptions were made about the bee behaviour, such as the worker bees all behaving the same and no bees changing partners in bursty phases. These assumptions could be accepted due to the apparent accuracy of the models. Another area of the model that could require more work is the relationship between parameters in the model. A goal of the modelling that was less emphasized in this thesis is the ability to scale the model. The basis for the model was 1200 bees to specifically replicate the empirical data, but scaling this up would require a delve into how the hive matrix should be sized up, and similarly for other parameters in the model.

Bees are important, and there has been reported a worrying trend of bees struggling.[31] Worry alone won't save the bees, and as such we must take action. A wider aim of this project was to get a better understanding of communication patterns between bees, as knowledge is of vital importance when aiding the bees in their struggle. In the end, even if the modelling was successful in capturing the desired metrics, it is also evident that more research is needed to fully capture all aspects of a social bee network. Of notable interest would be the movement and position of bees in the hive. More work would also be needed to make scaling of the model functional.

References

[1] Daniel Simberloff and Marcel Rejmánek. *Encyclopedia of biological invasions*. 3. Univ of California Press, 2011.

[2] Keith S Delaplane, Daniel R Mayer, and Daniel F Mayer. *Crop pollination by bees*. Cabi, 2000.

[3] Martin Lindauer. *Bienes språk*. Gyldendal, 1970.

[4] Charles AE Goodhart. "The importance of money". In: *Monetary Theory and Practice*. Springer, 1984, pp. 21–66.

[5] Yves Le Conte and Maria Navajas. "Climate change: Impact on honey bee populations and diseases". In: *Revue scientifique et technique (International Office of Epizootics)* 27 (Sept. 2008), pp. 485–97, 499.

[6] M. G. Park et al. "Negative effects of pesticides on wild bee communities can be buffered by landscape context". In: *Proc. Biol. Sci.* 282.1809 (June 2015), p. 20150299.

[7] T. Gernat et al. "Automated monitoring of behavior reveals bursty interaction patterns and rapid spreading dynamics in honeybee social networks". In: *Proc. Natl. Acad. Sci. U.S.A.* 115.7 (Feb. 2018), pp. 1433–1438.

[8] Márton Karsai et al. "Small but slow world: How network topology and burstiness slow down spreading". In: *Physical Review E* 83.2 (2011), p. 025102.

[9] Dario Floerano and Claudio Mattiussi. *Bio-inspired artificial intelligence: theories, methods, and technologies*. 2008.

[10] Bruce Anderson and Steven D Johnson. "The geographical mosaic of coevolution in a plant–pollinator mutualism". In: *Evolution: International Journal of Organic Evolution* 62.1 (2008), pp. 220–225.

[11] Francis LW Ratnieks and Laurent Keller. "Queen control of egg fertilization in the honey bee". In: *Behavioral Ecology and Sociobiology* 44.1 (1998), pp. 57–61.

[12] Yaacov Lensky and Yossi Slabezki. "The inhibiting effect of the queen bee (Apis mellifera L.) foot-print pheromone on the construction of swarming queen cups". In: *Journal of Insect Physiology* 27.5 (1981), pp. 313–323.

[13] LA Salazar-Olivo and V Paz-González. "Screening of biological activities present in honeybee (Apis mellifera) royal jelly". In: *Toxicology in vitro* 19.5 (2005), pp. 645–651.

[14] RE Page and EH Erickson. "Reproduction by worker honey bees (Apis mellifera L.)" In: *Behavioral Ecology and Sociobiology* 23.2 (1988), pp. 117–126.

[15] Nicholas W Calderone and Robert E Page. "Genotypic variability in age polyethism and task specialization in the honey bee, Apis mellifera (Hymenoptera: Apidae)". In: *Behavioral Ecology and Sociobiology* 22.1 (1988), pp. 17–25.

[16] KE Boes. “Honeybee colony drone production and maintenance in accordance with environmental factors: an interplay of queen and worker decisions”. In: *Insectes sociaux* 57.1 (2010), pp. 1–9.

[17] Walter M. Farina and Josué A. Núñez. “Trophallaxis in honey bees: transfer delay and daily modulation”. In: *Animal Behaviour* 45.6 (1993), pp. 1227–1231. ISSN: 0003-3472. DOI: `https://doi.org/10.1006/anbe.1993.1144`. URL: `http://www.sciencedirect.com/science/article/pii/S0003347283711449`.

[18] Christoph Grüter and Walter M Farina. “The honeybee waggle dance: can we follow the steps?” In: *Trends in Ecology & Evolution* 24.5 (2009), pp. 242–247.

[19] Joe R Riley et al. “The flight paths of honeybees recruited by the waggle dance”. In: *Nature* 435.7039 (2005), pp. 205–207.

[20] David C Gilley. “The behavior of honey bees (Apis mellifera ligustica) during queen duels”. In: *Ethology* 107.7 (2001), pp. 601–622.

[21] John Guare. *Six degrees of separation: A play*. Vintage, 1990.

[22] Gregory Goth. “Degrees of separation”. In: *Communications of the ACM* 55.7 (2012), pp. 13–15.

[23] A. Barabási. *Network Science*. 2015.

[24] E. O. Voit. *A First Course In Systems Biology*. 2012.

[25] Petter Holme and Jari Saramäki. “Temporal networks”. In: *Physics reports* 519.3 (2012), pp. 97–125.

[26] Mohammad Akbarpour and Matthew O. Jackson. “Diffusion in networks and the virtue of burstiness”. In: *Proceedings of the National Academy of Sciences* 115.30 (2018), E6996–E7004. ISSN: 0027-8424. DOI: `10.1073/pnas.1722089115`. eprint: `https://www.pnas.org/content/115/30/E6996.full.pdf`. URL: `https://www.pnas.org/content/115/30/E6996`.

[27] Manfred Fahle. “Human pattern recognition: parallel processing and perceptual learning”. In: *Perception* 23.4 (1994), pp. 411–427.

[28] Brent Goldfarb and Andrew A King. “Scientific apophenia in strategic management research: Significance tests & mistaken inference”. In: *Strategic Management Journal* 37.1 (2016), pp. 167–176.

[29] Andrew Adamatzky. *Game of life cellular automata*. Vol. 1. Springer, 2010.

[30] Vetle Rem. *Properties and dynamics of a large-scale social bee network*. Jan. 2019.

[31] Myrna E Watanabe. “Pollination worries rise as honey bees decline”. In: *Science* 265.5176 (1994), pp. 1170–1171.

NTNU
Kunnskap for en bedre verden