ANT-INSPIRED INNOVATION RESEARCH STRATEGIES

A Thesis

by

TEJAS RAJRATNA ADSUL

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Astrid Layton
Co-Chair of Committee,	M. Cynthia Hipwell
Committee Member,	Robert Puckett
Head of Department,	Bryan Rasmussen

August 2021

Major Subject: Mechanical Engineering

ABSTRACT

Innovations are fundamental for a sustainable growth of science. Studying the process of innovation and understanding how rate of discovering creative and new ideas can be optimized is essential in order to ensure the continued propagation of novelty in science. However, the lack of a foolproof method of finding novel concepts makes finding efficient innovation research strategies difficult. Past solutions in the form of innovation models have been able to simplify aspects of this process and achieved some success related to the ideation of optimal strategies, but have failed to address key aspects of research such as innovation communication, heterogeneity of researchers and effect of failures, among other things.

This thesis focuses on generating new tools for designing strategies of finding innovations. It uses bio-inspiration in the form of ant foraging by building a comprehensive analogy with the innovation process. A simulation model is constructed that uses a bibliometric network built from the metadata of articles within a particular scientific field as the environment, and specifically configured ants as agents interacting with the environment. Multiple aspects of ant foraging that improve efficiency of finding food were shown to have a direct correlation with research aspects that, if followed, can boost innovative discoveries. Some of these aspects include efficient communication of innovation, publishing failures and inconclusive research, and a balance between conservative and risky research.

Finally, the simulations showed that configuring the parameters so that the agents behaved exactly as some of the exemplary species of ants, such as Sahara Desert ants, Argentine ants, etc., results in highly efficient foraging. These parameters correspond to strategies that the ants employ during the foraging process, and the strategies translate to actions that researchers can take to improve the innovation finding rate. Thus, the thesis generates new tools for improved novelty discovery and addresses the limitations of other innovation models.

DEDICATION

This work is dedicated to my parents Rajratna and Shobha Adsul for instilling in me an admiration for science, encouraging me in all my academic pursuits, and ensuring I never overlooked my health. Without their sustained support and numerous sacrifices, this wouldn't have been possible.

ACKNOWLEDGMENTS

I would like to acknowledge my committee chair, Dr. Astrid Layton, my committee co-chair Dr. M. Cynthia Hipwell and my committee member, Dr. Robert Puckett for their guidance throughout the course of this research.

A special thanks to Dr. Robert Vander Meer for contributing to our understanding of ant behavior during foraging and designing parameters incorporating such behavior.

I would also like to thank Yanneliz Nava for her contribution on literature review of different ant foraging strategies, detailed in Section 2.2.3 of this thesis. A foundation of this work was laid by Kristina Viro and Pratyusa Kar, so my thanks go out to them as well.

Finally, a special thanks to my advisor, Dr. Astrid Layton for the time spent improving my work and the encouragement to try new ideas.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Dr. Astrid Layton and Dr. M. Cynthia Hipwell of the Department of Mechanical Engineering and Dr. Robert Puckett of the Department of Entomology.

A section of the thesis (Section 2.2.3) regarding an overview of ant foraging strategies was written by Yanneliz Nava.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

Graduate study was partially supported by the following departmental seed grant:

2019 MEEN Seed Grant, PIs: Drs. Astrid Layton and Cynthia Hipwell **"Enhancing the Impact, Occurrence, and Frequency of Innovation via Network Modeling and Analysis"**

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. MOTIVATION

1.1 Innovation Research

1.1.1 Search for Efficient Strategies

Novelty and creativity are some of the fundamental concepts that contribute towards a consistent growth of science. New impactful ideas in science and technology, often referred to as 'innovations', are instrumental in the sustenance of our society and economy. The term 'innovation' can be defined in various ways pertaining to the subject it is used in [1, 2]. For the purposes of this study, 'innovation' is defined as research that either gives rise to a completely novel idea, or that provides a connection between two disconnected but well-studied ideas. Both these kinds of research attempt to push the bounds of scientific knowledge by adding new entities and thus can be categorized as innovations. As much as novelty is sought in science, it is not easy to achieve. The choice of a research problem that is influential but solvable and allows for further research opportunities in the field dictates whether a project will be innovative. This choice is itself influenced by the researcher's aversion to risk, availability of resources and practicality of the work [3], commercial opportunities [4], and even the high uncertainty in short-term impact [5]. A conservative approach to solving problems is often enough to garner publications and sustain a scientific career, but it does not lead to original research. On the other hand, a risky approach can require years of work and can incur multiple failures, but ultimately leads to insightful contributions and even recognition in the form of prestigious awards [3]. History is ripe with examples of risky strategies that took a very long time to succeed and be accepted by the scientific community. Copernicus' heliocentric model of the solar system took almost a century to be accepted as the correct description of the solar system. Sir Andrew Wiles spent seven years solving Fermat's Last Theorem that had been unsolved for more than 300 years. Even Albert Einstein faced considerable opposition to his Theory of Relativity when it was first published. However, X-rays, penicillin, and cosmic microwave background that led to the theory of Big Bang, are some of the many discoveries that were accidental but highly influential. Thus, there is no foolproof way to achieve innovation. The aforementioned risks associated with making an impactful discovery are the reason that researchers are discouraged from pursuing it, even if the rewards are significant. This conflict of desire for innovation but inability to guarantee its achievement has given rise to a search for strategies that can facilitate the process. While studies have attempted to find efficient strategies for rapid scientific discovery, they often fail to consider some of the most important aspects of research. Time and efforts needed for information dissemination, failures, inconclusive research and effects of publishing them, and individuality of researchers are factors which especially decide the probability of success of a project, but go unaddressed in the search for efficient strategies.

1.1.2 Modeling and Simulation Techniques

Prior works in finding strategies for efficient discovery of innovations have made the use of knowledge networks in order to simplify the process. These networks provide a framework for scientific fields and enable understanding of knowledge production and diffusion through communities [6]. The growth of the network through the addition of new nodes and links mimics the evolution of science over time. Innovation networks form a part of knowledge-related networks that were studied in fields like economics [7], biotechnology [8] and biomedicine [9]. Furthermore, quantitative relationships between aspects of scientific activity such as publication of papers, citation dynamics and growth of specializations were reproduced using simulation techniques [10]. Combining the two approaches led to the rise of simulation models that used innovation networks and agent interactions to show the evolution of a scientific field [11]. Such models have been used to depict the dynamic process of innovation resulting from interactions of various agents. As documents became the primary source of analysis of scientific behavior [12], bibliometric quantities were commonly used for creating a scientific landscape and observing its evolution through the years [13], as well as identifying transformative scientific discoveries [14, 15]. The use of publications to create knowledge networks and simulation of agent-network interactions to find strategies leading to emergence of new ideas has since been seen in the fields of social sciences [16], information [17, 18], economics [7, 11, 19], and biomedicine [3, 20, 21]. Random walks are stochastic processes that have been commonly used to model the searching process in a simulation model. This idea brings the analogy of agents searching for new nodes in a network closer to researchers searching for innovations. The addition of bias to these walks allows modeling the preference towards specific topics that researchers may show based on their interests. An agent walking on a knowledge network thus has a higher probability of visiting nodes it has already visited - a concept called 'preferential attachment' [22]. Studies have shown the importance of such biased random walks with respect to mobility [23, 24], and have applied them to understand innovation process dynamics [25].

1.1.3 Successes and Limitations

All of these models have been successful in addressing important points that were missing in the study of innovation growth. Some have simplified the heterogeneity of actors and their interactions in an innovation process [11], others have modeled these interactions as cognitive processes and have shown the emergence of innovation through them [25], while still others have proven that research has been conservative in specific fields and have suggested how it can be made innovative [3, 20]. However, as all models have some limitations, earlier works on innovation models have not been able to incorporate some of the fundamental parts of research process mentioned before. Efficiently communicating about an innovation to the scientific community is an essential part of the research process [26, 27]. Innovation diffusion studies the process of spread of new ideas, and communicating the innovation forms an important part of it [28]. Thus, the time and efforts required from discovery of the innovation to its dissemination among others are absent in the innovation models constructed so far. The changes in the innovation landscape [11] or the knowledge network [20] after the discovery of an innovation are immediately visible to the other actors in the model, which is not the case in reality. Writing an article describing the ideas and methods used to reach the innovation, accounting for publication delay and rejections [29], scheduling conferences and seminars – all require time as well as efforts before the innovation is accessible to others. Without purposeful dissemination, new findings may struggle to create the impact required of innovations. The use of agent-network interactions has also led to a departure of heterogeneity among the agents due to simplification, even though its significance in this context has been shown before [11, 30]. Finally, the failure rate experienced by researchers that go for high-risk strategies is a factor missing in these models. This failure can either be in the form of inability to reach the goals, inconclusive results or a dead-end. In the case where the research fails to show the promise of innovation, recent models [20] have shown that publishing such failures leads to a drastic increase in the rate of innovation discovery.

1.1.4 Biological Inspirations

The agent-network models thus far discussed consider researchers to be a singular species. While each researcher may have an individuality, they work towards a common goal, which is the development of science. This provides a motivation to look at biological systems which work together towards a common goal as well. Many insects like ants, bees and wasps show complex social behaviors involving cooperation, like division of labor, collecting food, protection of the nest and rearing the young [31]. The foraging strategies of such insects, especially ants, show a lot of similarities with the innovation research process and thus show potential for innovation modeling. This thesis looks at such bio-inspirations for constructing an innovation model that can address the aspects other models have overlooked. Building on the success of learning from the designs generated through millions of years of evolution in nature, this work seeks to extract strategies that evolution has deemed optimal in the search process and characterize them into implementable tactics that those interested (e.g., researchers, funding agencies, companies) can use to increase their rate of generating impactful innovations. These tools would provide a new insight into innovation modeling using bio-inspiration techniques, find strategies that are already efficient in their respective ecologies, and translate them into actions that can improve innovation rate in science.

1.2 Research Questions and Goals

1.2.1 Primary Research Question and Goal

The overall objective of this dissertation is to answer the following research question and achieve the following research goal.

Main Research Question: What insight can biological search strategies provide for improving success rate of finding innovation?

Main Research Goal: Generate tools that will improve the rate of scientific innovation discovery

This is a broad question to answer and an equally broad goal to reach. In the following subsection, a break down into simpler research questions and goals has been provided in order to improve clarity and develop a methodology for answering and achieving them, respectively.

1.2.2 Secondary Research Questions and Goals

1. **Secondary Research Question:** Are there any biological systems that show an analogous behavior to innovation process, and if so which?

   Secondary Research Goal: Identify and show how specific behavioral strategies of some biological systems can be correlated to the research process of finding innovations.

2. **Secondary Research Question:** How can network theory be utilized for understanding the landscape of a scientific field and finding best strategies for discovering novel ideas?

   Secondary Research Goal: Develop a methodology to use innovation/knowledge networks in conjunction with bio-inspired agent-network interactions, and use this to achieve the above secondary goal.

3. **Secondary Research Question:** What strategies can be extracted from the biological systems, and how can they suggest routes towards an improved and consistent rate of innovation generation?

   Secondary Research Goal: Establish various strategies observed in the biological inspiration, which translate to an enhanced rate of innovation discovery.

1.3 Contributions

This work strives to add to the methodologies of finding strategies to improve innovation rate. Benefits include a fresh insight into innovation modeling, new tactics implementable in innovation research, and tools that innovators can utilize to increase their chances of finding scientific discoveries in their respective fields.

1.3.1 Primary Research Contributions

- **Novel strategies that improve the innovation discovery rate**

  The project builds upon the innovation-finding strategies proposed by other models. It serves as a tool to analyze the scientific landscape of any field and evaluate what kind of research would give the best chance of making innovative discoveries. It is hoped that such a simulation model would prove beneficial to innovators for understanding the chances of success in different situations. It would also allow a more efficient allocation of resources in innovation projects.

- **Innovation model that addresses communication of innovation, failure during research and other aspects**

  This work attempts to incorporate the aspects of research that have been shown to be important but go overlooked in other innovation models. Since research is a complex process with innumerable parameters, its modeling requires simplification. Earlier works have addressed most of the major parts of an innovation process. This thesis adds to it other factors like innovation diffusion, failure in risky-research, effects of publishing failures and heterogeneity of researchers. It hopes to bring the simulation model a bit closer to the original research process.

1.3.2 Secondary Research Contributions

- **Validation of an analogy between the innovation research process and biological search strategies**

Bio-inspired Design is an area which borrows inspiration from designs, strategies or methodologies found in nature and uses it to solve problems faced by humans. This thesis builds an analogy between the innovation research process and biological search strategies, such as foraging for food. By finding commonalities between the two processes, the strategies of the biological systems, which are deemed efficient by evolution, can be translated to a set of actions that apply to similar situations in research. Since there are numerous species of animals that apply a variety of strategies corresponding to their environments, such an analogy allows access to multitude of ideas that can be implemented. This can push the rate of innovation discovery significantly and lead to a consistent growth of society and economy.

1.4 Methodology

The proposed research questions and goals are answered in this dissertation by performing the following tasks. They set up a natural order for the progress of the thesis and work towards achieving the main research goal.

Research Task 1: Identify the biological analogue for innovation research process.

1a: Review the literature for biological systems that work together towards a common goal, especially in the context of search strategies. This goal could be foraging, sustenance, existence or survival.

1b: Build an analogy between the candidate biological system and the innovation process, assessing for similarities and differences.

Research Task 2: Establish an innovation network case study.

2a: Build a representative dataset for the innovation process. This may be composed of journal articles from a specific field during a given year range.

2b: Use this dataset to create a bibliometric network of nodes and links. Validate how the growth of the network through addition of nodes and links conforms to the definition of innovation within this work.

2c: Translate the network into an environment that a simulation of the biological system identified in **Task 1** can interact with. Verify that the analogy between the network and the natural environment of the bio-inspiration is sound.

Research Task 3: Construct a simulation model of agents interacting with the network.

3a: Initialize agents within the network that behave similar to the bio-inspiration, and let them employ their strategies to achieve a common goal.

3b: Validate that the changes in network correspond at least qualitatively to changes observed in the dataset.

3c: Parameterize the model and tweak them to observe changes in the model

Research Task 4: Translate the biological strategies into tools for improving successful innovation rates.

4a: Find the parameter sets that correspond to the biological strategies seen in specific species, and calculate model performance.

4b: Mix the parameter values for different species and find if there's an optimum multi-species combination.

4c: Translate these optimized parameter values into a set of actions that researchers can take to improve the rate of innovation discovery.

1.5 Thesis Outline

The dissertation covers a range of topics that involve innovation processes as well as theories that motivate use of bio-inspiration for innovation modeling. The introduction of this chapter on motivation addresses the problem that the thesis tries to solve, the current solutions that have been put forth, and their limitations. Following this, a chapter on background provides a detailed review of innovation research modeling, specifically the use of network theory, bibliometric mapping and a concept called 'Team of Teams' that relates to the collective intelligence of people in a community. These ideas provide reasonable strategies for constructing a simulation model based on bibliometric data of a scientific field. This is followed by different bio-inspirational theories that are relevant to the topic of innovation finding. Two foraging theories - Optimal Foraging Theory and Information Foraging Theory - are discussed. They portray the behavior of human beings in search of knowledge or information as being similar to animals in search of food. The idea of 'Emergence', which can be broadly described as the self-organizing behavior of a complex system, draws a close correlation to the 'Team of Teams' idea mentioned before. All these topics furnish a sound basis for considering the foraging strategies of social animals as the bio-inspiration for innovation modeling. The chapter ends with a detailed account of ant foraging strategies, followed by an analogy between various aspects of ant foraging and the innovation research process. The chapter contributes to many different ideas that warrant the use of foraging behavior of ants as a method to find efficient innovation searching strategies.

The construction of the simulation model is defined in the chapter on method. This starts with building the dataset using articles from the Web of Science Category 'Robotics'. A software called VOSViewer is described that provides bibliometric networks from metadata of journal articles. A keyword co-occurrence network is acquired from this software that shows a network of keywords used in articles. The data corresponding to this network is refined by merging similar keywords. It is then imported into MATLAB and used to create a graph based on the connections in the keyword co-occurrence network. MATLAB allows further analysis of the data in terms of cumulative sum of connections built among nodes per year, which leads to some insights into the evolution of the field. Finally, a detailed characterization of the simulation model is provided. The constructed graph is used as an environment, and ants are considered to be agents that interact with the environment. Different parameters corresponding to the environment and the agents allow for various strategies that can be implemented.

Later, a chapter on some results and discussions follows. The model is validated by showing the rate of food collected by ants in the simulation is qualitatively similar to the rate of addition of connections in the data. Various parameter sets are constructed to depict specific strategies that are followed during foraging by different ant species. The performance of these strategies is tested against each other. All these sets are then mixed to determine whether a combination strategy performs better than the individual ones. The chapter also details a discussion on the translation of simulation strategies into actions that the researchers can take and expect a similar performance.

The utility of the model in terms of providing a fresh insight into innovation discovery strategies is discussed in the chapter on summary and conclusion. The model is summarized and its usefulness as a novel way of searching for efficient research actions that can boost the scientific growth is mentioned. A chapter on future ideas outlines the different versions of the model tried and tested, and lists some improvements that can enhance the model.

2. BACKGROUND

2.1 Innovation Research Modeling

2.1.1 Use of Network Theory

Network theory became an important part of research since the beginning of 2000s. It was re-examined, a theoretical and a practical framework was introduced for it, and the many applications that it could be applied to were put forth [32]. Network modeling has since become a method to visualize structures such as the internet [33], traffic flow [34], etc. The growth of a structure such as the internet can be seen as the evolution of a network by addition of nodes and links. This gave rise to different approaches such as static networks, evolving networks, agent-based modeling, and dynamics on networks [32].

The production and distribution of knowledge was analyzed with the help of two properties: knowledge is a correlational structure, and knowledge is a retrieval or interpretative structure. This allowed knowledge to be seen as something that forms correlations between different concepts, thereby giving it the representation of a network [35]. The evolution of science fits this representation, since every new invention can be seen as an addition of new nodes or links, while research seeks to increase the connectivity among the areas of science. As an example similar to the one given in [8], the work of Sir Andrew Wiles in proving the Fermat's Last Theorem, mentioned in the introduction, involved such an attempt to connect two fields in mathematics. The Taniyama-Shimura conjecture proposed a connection between two very distinct fields of mathematics – topology (concerned with study of shapes) and number theory (concerned with study of whole numbers) [36]. Sir A. Wiles was able to establish the existence of this connection, thereby increasing the connectivity in mathematics and giving rise to completely new ways of solving problems.

Thus, it became quickly established that network science had many uses in understanding the evolution of science. The interactions among people, ideas and organizations to create new innovations have been modeled into network dynamics, giving rise to innovation networks [37]. Such networks have since then been used to model the process of innovation, and find strategies that could lead to an improved rate of novelty in science.

2.1.2 Bibliometric Mapping

The process of finding an impactful innovation begins with a choice regarding the specific research area to start from and the direction in which to continue. An information seeker often wants to gauge maximum profitability from this choice, since the resources that need to be employed are limited. Thus, it is beneficial to study the dynamics of the scientific landscape in order to examine the most potential areas for innovation. As mentioned in the introduction, documents, mainly journal publications, books, etc., have become to be commonly used to analyze the landscape of scientific fields [12, 13]. Bibliometric mapping is a technique where research publications are used to quantitatively analyze this phenomenon. While not exactly new [38, 39], a variety of methods have been developed in the recent years to parse millions of documents and construct easy-to-understand networks [40]. Two common methodological approaches are seen in these techniques [41]. The first one works on a scalar dimension and uses the document as a single independent entity. The growth of science can then be depicted by counting the number of publications in that field, the number of journals, the number of active authors, or the frequency with which the related keywords are used. Google Ngram Viewer (https://books.google.com/ngrams) uses this approach to generate frequency graphs of the search words by parsing through millions of digitized books. Fig. 2.1 shows such a graph for the term 'bibliometric mapping'. It shows an increase in the usage of the term in recent years. The second method attempts to build relational links among attributes of a publication by counting the number of times they were mentioned together, also called co-occurrence. The resulting two-dimensional map consists of nodes being the attributes and links being the co-occurrences. This approach has been used in a majority of studies involving bibliometric mapping [3, 20, 25]. Depending on what the nodes and the links represent, the mapping can be a citation network, a co-citation network, a bibliographic coupling, a co-authorship network, or a co-occurrence network [42, 43]. Fig. 3.2, explained later, is an example of a word co-occurrence

Figure 2.1: Google Ngram result for 'bibliometric mapping'

network. Such a map correlates to an innovation network and thus provides a way to build a dataset for testing different research strategies in order to estimate their efficiency in finding innovations.

2.1.3 Team of Teams

The concept of 'team of teams' was put forth [44] to find the factors which lead to the most innovative creations from a group of people within an organization or a community. Research suggests that the collective intelligence of a community seldom has much to do with the individual intelligence of its members, but rather depends on the connections among them and how easily new ideas can arise within it. The two major aspects of social dynamics which affect the ease of new thoughts within a community are 'engagement', or give and take of ideas within a small group like a team, and 'exploration', or idea flow among multiple teams. Collective intelligence thus stems from this constant interaction and a team which fosters the two mentioned aspects performs better in finding innovations than any individual member could. This idea provides further motivation to look at social animals, especially animals with a very high level of social organization and division of labor, as bio-inspiration for the researchers and the innovation process.

2.2 Bio-Inspiration for Innovation Research Process

2.2.1 Foraging Theories

Foraging behavior of animals consists of the decisions made by them in procuring food, such as what type of food to eat, when and for how long to search for it, and others. Energy and time are spent in this activity, and as the food gets depleted in the immediate vicinity of their habitat, the animals have to spend more resources in finding food that is further away. However, animals also have the choice to leave their current habitat and move to a patch where there is abundance of food. The Optimal Foraging Theory (OFT) hypothesizes that the foraging behavior of animals can be described as a dynamic process of optimizing a cost/benefit ratio [45, 46]. OFT models incorporate a currency (like energy spent per food morsel) that the animals try to maximize when foraging while dealing with constrains (like predators), which helps predict the decisions that animals should take in specific environments. The theory has received substantial support [47, 48, 49, 50] as well as considerable criticism [51, 52, 53]. OFT forms the basis for the Information Foraging Theory (IFT), which states that just like animals, humans portray certain behavioral patterns while consuming information [54]. The decisions taken by an information seeker, such as a researcher, depend on the information environment, the available resources, and the profitability of those decisions, that is, the benefit obtained and the cost required. The theory is commonly used in Web Design [55, 56]. Studies have shown [57] that such foraging theories can be used to predict the best actions a researcher can take to find innovations. Models based on OFT or IFT can correlate diffusion of knowledge to not only the structural properties of an information space but also the interactions among agents, optimization of currencies, and handling of constraints [14]. This provides a motivation for a biologically inspired model that can be analogized to the process of researchers finding innovations.

2.2.2 Emergence

Emergence [58, 59] can be broadly described as the self-organizing behavior of a complex system, which results from interactions among the system's actors. It especially is very difficult to identify the behavior of the system from the behaviors of its components in isolation. Social insects like bees and ants have been shown to make better decisions from collective intelligence rather than independently [60]. This closely corresponds to the 'Team of Teams' idea mentioned before. The process of research often involves conferences, seminars, presentations, discussions and other ways to increase engagement among the community. This leads to an increased flow of ideas and greatly improves the chances of finding innovations. Moreover, since researchers act independently, the evolution of science can be attributed to their efforts although it cannot be completely predicted. This idea is very similar to the concept of Emergence, and thus proves to be an excellent point for finding a bio-inspiration among social insects.

2.2.3 Foraging Strategies of Ants

Ants are social insects whose foraging behavior has been studied over the years [61]. Ants have continually shown that they are efficient in finding food and sharing information about food sources with other members of their colony. Aforementioned points of motivation form an encouragement to focus on ants for inspiration, and determine if ants' foraging characteristics, patterns, or behavior could be used to model innovation networks. In doing so, it allows an understanding of how information can be shared more efficiently between researchers as well as knowing if there is a particular innovation process strategy that can help to increase the number of innovations. Ants have various foraging strategies that are influenced by different factors such as species, colony structure, and food resources. The strategies that have been observed include solitary foraging, social carrying, tandem running, group raids, pheromone trails, and trunk trails [62]. While ants may use different ways to communicate the presence of a food source or a predator, this work focuses on the use of pheromone trails used to recruit other members of their colony.

During foraging, scout ants leave the nest in search of food and once an ant locates a food source it begins to deposit trail pheromones leading to the nest. Once the trail pheromones have been deposited, recruits follow and reinforce the path by depositing pheromones to attract more of their nestmates if they consider the food of high quality [63]. Foragers use these pheromones as a way of communicating and relay information about the food source to worker ants. Some species use various pheromones when foraging including no entry pheromones, long-lasting pheromones, and short-term pheromones [64]. No entry pheromones are used as repellents to keep ants from taking certain routes that do not have any profitable food sources or to keep ants from foraging areas that have been previously explored [65]. Some ants use long-lasting pheromones as a means of home-range markings or external memory. As ants travel further from the nest this trail becomes weaker making it more likely for scouts to randomly explore the surrounding area and leads to more efficient exploration as well as rapid adaptation to changing environments. Finally, short-lived pheromones are used as a guide to the feeding sites and help to increase the foraging range while decreasing recruitment delay. Some ant species have pheromones that can last up to 2 days and are able to rapidly exploit newly discovered food as well as rapidly abandoning trails that don't lead to food through the use of volatile negative pheromones, no entry signals [66].

Simplifying the process of foraging leads to the following steps that an ant might take: leave the nest, start in a direction, try to sense the presence of food, explore, find food, return to the nest while laying a pheromone trail from the food to the nest, and allow other ants to follow this trail and further exploit the found food resource. Similarly, a research process may be simplified as: start from a specialty, choose a direction of research or a problem to solve, try to understand the innovating ideas that might emerge from that work, explore, find the innovation, form a coherent train of ideas linking existing knowledge to the innovation, and publish and allow other researchers to follow this work and further exploit the innovation. The similarities visible between these two processes, along with all the motivations mentioned before, encourage construction of a bio-inspired solution to innovation modeling.

2.3 Analogy between Ants and Researchers

This section follows a discussion of the analogy between the foraging process of the ants and the innovation process of the researchers. Table 2.1 shows a T-chart for similarities between the two processes. It can be seen that most of the ideas are same in nature for both the systems.

2.3.1 Nest

Ants start from and end at their nest, that is, they have a fixed point where all the food collection happens. To them, the nest acts as a fixed resource of information. When an ant finds a food morsel after exploration, it communicates the location of the morsel to other ants in the nest. Other ants can then go on to exploit that food.

Researchers mostly start their work from an idea in their own specialty. It could be their own idea or something inspired from other existing work. They journey through the information space exploring, and may end up finding an innovation. In order to communicate this innovation, they must connect it to knowledge that already exists. In the context of the definition of innovation mentioned in this work, the innovation is either a completely new idea or a connection between two known ideas, which implies that it must be linked to existing knowledge. Thus, the researcher's specialty forms a natural starting point. The process of research can be summarized as starting with knowledge that is already present, identifying a problem, working towards a solution, and then connecting the solution to the extant knowledge so that other researchers can follow the connection and develop their own innovative ideas. The to and fro motion of an ant from the nest to the food and back to the nest mimics this process.

The idea of a fixed starting point is somewhat mismatched, since, unlike ants, it is not necessary that researchers end up exactly where they started from. The innovation could be something that is not completely related to their initial ideas. However, for the sake of simplicity, the construction of this model assumes that researchers strive to link ideas they started from to the innovation they found.

2.3.2 Pheromone Trails

Many species of ants use pheromone trails to communicate information about the location of food resources to other ants in the nest. When an ant returns with information about a food resource, it lays a trail that other ants of the nest can follow. If the other ants are successful in finding the food, they add pheromone to the trail in order to reinforce it. As the pheromone strength increases, more and more ants follow the trail and collect the food. If an ant does not find food, the trail is not reinforced and evaporates in time. Thus, once the food is extinguished and the trail is completely evaporated, the foraging on that path stops [67]. While many nuances dictate the variability in these trails, such as specificity to species, volatility and repelling nature, they simply help ants reduce the exploration time required to find food and help others do the same.

Researchers too have a duty to communicate their innovative discoveries to the scientific community. They do so with the help of publications, lectures, seminars, patents and other opportunities. In all these cases, a researcher is essentially trying to lead other researchers towards his/her discovery, starting from common/established knowledge, so that they too can exploit it. A publication involves a coherent chain of ideas starting from the knowledge that exists, the innovation that was discovered, and the method used to go from the existing knowledge to the innovation. Thus, such media mimic the use of pheromone trails in ant foraging dynamics.

Publications and pheromone trails are similar in more ways than one. Primarily, both start from a point of established knowledge and lead to a point of new knowledge. Moreover, their strength determines how impactful the new knowledge is. As more and more ants are able to collect food from a resource, the strength of the trail leading to it increases, thereby attracting more ants. In the case of publications, the number of citations garnered within a short time is indicative of the novelty and the impact of the discovery put forth [3]. Thus, innovative discoveries attract more researchers, they exploit the discovery and may even search for ideas in the vicinity, which in turn attracts further more researchers. Once the food is extinguished, the trail evaporates since ants do not reinforce it. While citations don't disappear, the impact of an innovation is certainly more apparent from the early citation counts [68] rather than the later ones. Repellent pheromones warn ants of the absence of any food or a danger of any kind, thereby discouraging them from following that path. Publications which mention the failure of an endeavor, especially a risky one with a high probability of failure, discourage other researchers from following that path too. Such publications have in fact been shown to improve the efficiency of finding new ideas [20].

2.3.3 Complete Analogy

There are a number of similarities that can be detailed between the process of ant foraging and the process of innovation. The concepts of a nest and pheromone trails form the primary part of this analogy. An extensive list elaborating this analogy has been provided. This details many ideas, most of which have been used in this model while some are potential developments that can be implemented.

1. Nest: Nest acts as a source of existing knowledge for ants, from which they begin their journey into the unknown territories, find food sources (new knowledge), and bring it back so that other ants can exploit it too. For researchers, the nest is the current intellectual structure of a scientific field. It encompasses all the knowledge that already exists. Researchers start from this structure, find innovative ideas that have never been discovered before, and connect it to the existing knowledge via a coherent thought process, so that the rest of the scientific community can exploit the novel ideas as well.

2. Polydomous colony: Many ant species have polydomous colonies, which means a single colony can have multiple nests within short distances of each other. Thus, they are spatially separated but socially connected [69]. Polydomous colonies can carry out foraging in a dispersed way, collect food at respective nests, and distribute it among the nests. This allows shorter foraging trips while maintaining a long foraging distance [70]. The evolution of a scientific field comprises of many subtopics arising as they are discovered and studied. The scientific field of Mechanical Engineering consists of numerous specializations like Thermodynamics, Automotive Industry, Heat Transfer, Robotics, etc. Researchers, like ants, start from one of these specializations and endeavor to connect innovations related to it. Specialized research thus

contributes to the overall growth of the field.

3. New nest: An ant colony might build a new nest for a variety of reasons. One of those reasons is to better exploit a temporally stable food source [71]. Ants build a nest near a newly discovered food source that can sustain them for long, thus reducing the effort required to exploit it. In a similar way, new opportunistic subtopics within a scientific field can turn into specializations once their connection to the field becomes well-established. As more and more ideas from a new knowledge source are studied, it becomes the starting point for finding other innovations in its vicinity. Existence of research subtopics like Computational Fluid Dynamics, Turbulence Modeling, Boundary Layer Theory, etc. within the field of Fluid Dynamics implies that as newer knowledge is incorporated into the current structure of science, it turns into a specialization that researchers can start from and find novel ideas connected to it.

4. Food: A food source, as mentioned before, is a new source of knowledge for the ants. Ants attempt to find as many food sources as they can. This is to sustain their existence, grow their population, and overall be successful as a species. Innovations are important in a similar way to the communities. They contribute towards social as well as economic development of human society. Thus, the scientific community attempts to push their bounds of knowledge as far as they can by finding these innovative ideas, incorporating them into technologies that people can use, and thereby sustain the growth of society.

5. Exploring speed: Not all ants find food in the same amount of time. Every individual ant takes its own time exploring its surrounding environment while foraging. This results in a randomness in the time taken for an ant to find a particular food source after leaving the nest. Similarly, not all innovators can orient themselves in the correct direction while finding novel ideas. The ability to grasp new concepts differs among individual researchers, thus giving rise to a similar randomness in the time taken to find an innovation.

6. Pheromone trails: Many ant species pass on the information about newly found food sources to other ants in the nest by laying a pheromone trail from the food to the nest. This stimulates

other ants to follow the trail, exploit the available food, and on exhaustion of the food, explore further. The trail acts as a physical route of information that guides new ants, thus reducing the time that it would take to actually find the food. Researchers are similarly obligated to communicate the innovations they discovered with the rest of the scientific community. As discussed in Subsection 2.3.2, publications, lectures, seminars, etc., act as routes of information that can help other researchers follow the ideas involved in discovering the innovations. In both cases, trails connect new knowledge to existing knowledge and allow members to find this new knowledge easily.

7. Straight path from food to nest: Ants, especially desert ants, have been shown to follow a straight-line path from the food to their nest, even if their exploring might have been long and tortuous [72]. They use various techniques such as path integration, landmark guidance, surface structure, etc. [73] to remember the relative location of the nest during foraging. This ensures that other ants can find the food resource without spending too much energy exploring. The process of innovation research too can consist of long, tortuous exploring and dead-ends. However, the researchers always try to communicate the link between current ideas and their innovation in the most direct way possible through the communicating media. The process of explaining the methods used in finding innovations involves simplification and connection to the existing knowledge in a way that others can follow. Thus, the linking of the ideas from an innovation back to the starting point of research forms a close analogy with the way ants return to the nest from the food in a straight line.

8. Pheromone reinforcement: Most ants that follow a trail to a food source and collect food reinforce the trail on the way back to the nest. The strength of the trails allows the colony to efficiently allocate its work force among available food sources without central control [74]. In some species, ants modulate the pheromone trail strength in proportion to the quality of the food resource they have found to ensure more ants find and exploit it [75]. Thus, level of pheromone reinforcement on a trail has a direct connection with the abundance of food at the end of it. In a

scientific field, publications can be thought of as trails. The more innovative and opportunistic an idea is, the more people try to exploit it in various ways, leading to numerous publications connecting the innovation to existing ideas. Thus, the number of publications, or even seminars, conferences, etc. pertaining to a novel idea, is representative of how innovative it is, and encourages more people to pursue it.

9. Pheromone evaporation: Since pheromone is essentially a chemical compound, it evaporates over time. Ants keep reinforcing the trail till food is found at the end of it. Once food has been exhausted, no reinforcement occurs and over time, the trail evaporates completely. This ensures ants do not keep going to a spot where food has already been exhausted. Many bibliometric analyses employ different methods to accurately define the current status of a field based on the relevant publications and citations. Although publications don't evaporate like pheromones, the rate of citations is often used to show the novelty or obsolescence of a field [76]. Citation rates follow a quick growth and slow decay [77], which is similar to the way pheromone strength of a trail shoots up when numerous ants collect food at the end of it and reinforce it but evaporates slowly over time once the food has been exhausted and reinforcement stops.

10. Repellent pheromone: Certain species of ants have been shown to use repellent pheromones [65] in order to warn other ants of predators or food depletion. While not a lot of study has been done on the effectiveness of using such pheromones, it has been adopted into a number of search algorithms [78]. Recent studies have shown that publishing failures, such as dead-ends or inconclusive results, can lead to a better evolution of a scientific field [20]. Thus, the use of repellent pheromones acts as an analogy to publishing failures or lack of innovation in a particular direction.

11. Exploring and exploiting ants: The distinction between exploring and exploiting ants is specific to this work, but is inspired by the difference between scouts and recruits among ants. In many species, scouts are focused on finding new food sources while recruits/workers collect food from existing sources. Although this division of labor has been shown to be further categorized [79],

ants do need to explore food sources as well as exploit them simultaneously. This distinction holds among researchers as well. Some focus on exploring unknown territories, taking risky steps, and discovering impactful innovations. Others concern themselves with more traditional conservative research that is more of a repetition of what has already been established than something completely new. Just like ants, there needs to be a balance between exploration and exploitation for optimal growth of science.

12. Failure: Ants can sometimes fail in their foraging attempt to find food. This can either be due to predation or exhaustion. Innovators too can fail due to depletion of resources or running into a dead end.

13. Preference for pheromones: Different ant species might prefer pheromones in different ways. E.g., The black garden ants rely on their memory while finding food rather than choosing the path with the most pheromone [80]. Thus, an ant may either make decisions based on the highest pheromone strength, or on the lowest strength, or remain unbiased by it. This ties to researchers who are dependent on available knowledge connections in deciding which problem to pursue. Some may solely analyze current publications and orient themselves towards a problem that has been facing a lot of attention, some may decide to try and find a solution in an area which hasn't received much attention, while others may focus on their specialization and remain completely neutral on which subtopics are receiving attention.

14. Preference for food weight: Ants have also been shown to have specific food preferences, which has led to extensive bait studies. Some ants go for large food morsels based on their odor. This is particularly beneficial for mass-recruiting ants since other ants can quickly exploit it. On the other hand, solitary foraging ants go for smaller food so that they can carry it alone easily. Thus, ants may prefer choosing a path that leads to a large food source, or they may go for small food source, or they may again be unbiased by the food size. Among researchers, this analogizes to choosing a problem in an area based on its opportunistic value. The more fertile an area is, the more it can be exploited, which in turn attracts innovators. On the other hand, some researchers

may exclusively choose to go towards problems that do not have a lot of exploiting opportunities but nevertheless contribute towards the growth of science. Finally, researchers may also choose based on other preferences and be blind to the opportunities a project may present.

15. Preference food distance: Odor of a food source also depends on its location and how away the ant is. Different species forage for different distances ranging from a few centimeters to more than 10 meters [81]. It is evident that some ant species prefer finding food at a short distance range, some prefer foraging over long distances, while others are unbiased. Treating the physical distance between food as cognitive distance between ideas, some researchers may simply focus their work within the vicinity of well-established ideas while some may choose to explore obscure areas of their field and try to find impactful innovations.

16. Changing strategy over time: Ants have been shown to change their strategies pertaining to exploratory behavior, walking pattern and food preference based on the number of ants or the availability of food in the field [82]. This allows a dynamic evaluation of the availability of food resources and allocation of appropriate tactics that can increase the efficiency of foraging as compared to following a single strategy forever. In innovation process, this is an important point since science progresses at a very rapid rate and what is relevant at one instant may become obsolete at another. Just like ants, researchers must be able to analyze the current scientific landscape and figure out what kind of strategy would work best. Only following a single set of actions wouldn't work over a long time.

Researchers		Ants
Operational Environment		**Operational Environment**
Research area made up of concepts / ideas	Similar	Patchy environment
Moving one concept / idea to another	Same	Sequential movement from point to point
Innovations randomly spread over the area	Same	Food randomly spread over the area
Starting points of research specializations	Similar	Nests
Function		**Function**
Analyze the field to choose a direction of research	Same	Scent of food and pheromones to choose nearest food
Find innovations	Same	Find food
Communicate innovation through publications	Same	Lay pheromones for disseminating information
Use available information to exploit innovations	Same	Use pheromone trails to find food
Specifications		**Specifications**
Gauge amount of time to be spent in research	Similar	Calculate energy required in collecting a food particle
Communicate information in the easiest way	Same	Lay trails and walk mostly in a straight line from food to nest
Follow information completely / partially	Similar	Follow the trail completely to food, unless there's repellent pheromone in which case change direction midway
Communicate failures through research publications	Same	Lay repellent pheromone trails when food is extinguished
Performance Criteria		**Performance Criteria**
Rate of finding innovations	Same	Rate of finding food
Amount of research field covered	Similar	Amount of area foraged

Table 2.1: T-Chart of similarities between researchers and ants

3. METHOD

3.1 Building the Dataset

3.1.1 VOSViewer

VOSViewer is one of the many softwares available for visualization of scientific networks with the use of bibliometric data. This software introduced a new method for creating networks of objects based on their similarity called VOS, which is an abbreviation for *visualization of similarities*. This was an apparent improvement over the then common method MDS, or *multidimensional scaling* [83, 84]. VOS provides a location of each object in agreement with what have been called the ideal coordinates. These coordinates are calculated after considering the similarity among objects, as well as the secondary similarities between two objects arising from their mutually common entities. The software [85] uses text-mining functionalities to analyze documents, employs similarity measures to calculate co-occurrence of selected objects, and outputs a two-dimensional plot of the network generated [86, 87]. It readily parses data collected from Web of Science and provides a multitude of options for construction of bibliometric networks. Thus, it was selected to build the dataset for the model.

3.1.2 Word Co-occurrence Network

The field of robotics was chosen to be the dataset on which the simulation would be tested. Robotics has shown a considerable growth in the recent years [88]. This can be corroborated using the Web of Science database. The word 'robot*' in the search field of topic returns all the publications in which the word appeared. The asterisk ensures that all terms like 'robot', 'robots', 'robotic' and 'robotics' are included. Plotting the number of records per year since 1960 to 2020 in Fig. 3.1 shows the drastic rise of potential in this field. The 'Robotics' Web of Science Category contains records from 1991 to 2020 from assorted journals. After refining these records by 'Article' document type, the metadata for around 26000 documents was collected to form the dataset.

A keyword co-occurrence weighted network was constructed from this dataset using the soft-

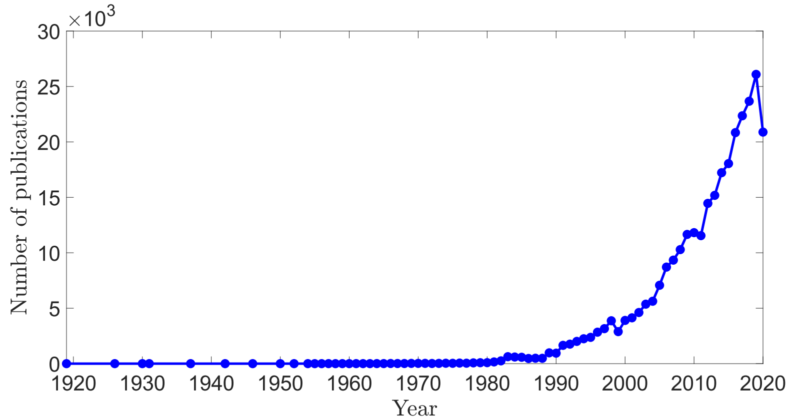


Figure 3.1: Number of records under the topics of ‘robot’ and ‘robotics’ in Web of Science per year

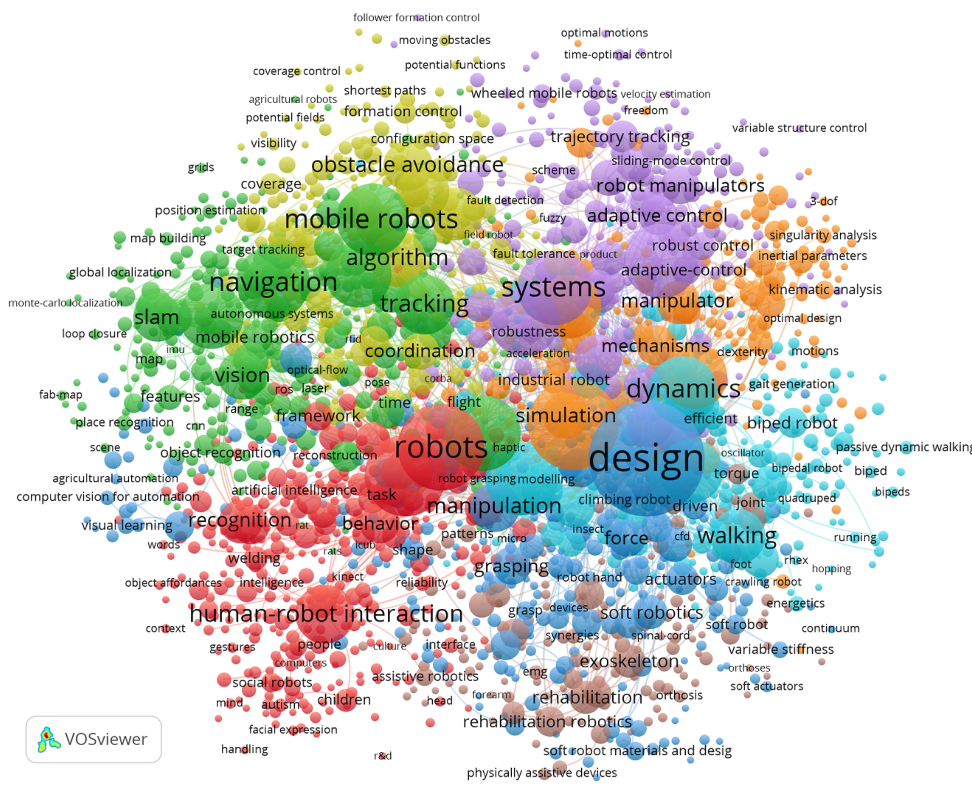


Figure 3.2: Keyword co-occurrence network of Robotics publications from 1991-2010

ware. This uses the keywords by which the paper has been annotated and forms a similarity graph from them. In such co-occurrence networks, the relatedness between two objects is determined by the number of documents in which they have occurred together. Keywords have been commonly used to understand the research areas in a field and the interlinking between them [89], as well as the evolution of the field [42]. Thus, it was deemed reasonable to use a keyword co-occurrence network for this work. Fig. 3.2 shows this network. The following steps outline the process of creating this bibliometric network in VOSViewer.

1. The software provides a link for creating a map based on bibliographic data. This gives an option for inputting database files collected from Web of Science, Scopus, Dimensions of PubMed.

2. The files containing metadata for Robotics from 1991 to 2020 were submitted in this field and were parsed by the software.

3. The next step involves choosing the type of network to construct. A co-occurrence type of analysis was chosen, with 'all keywords' as the unit of analysis. This includes both author keywords and KeyWords Plus. KeyWords Plus is a product of Clarivate Analytics. It is a data of words and phrases that frequently appear in the titles of an article's references, but do not appear in the title of the article itself. Since a comparative analysis of both these types of keywords [90] showed that both are effective in investigating knowledge structure of scientific fields, both were selected in this construction.

4. VOSViewer also allows for an option for the type of counting method to be used – full or fractional. Full counting implies each link in the network has a unit weight. Fractional counting implies the weight of each link is fractionalized. Fig. 3.3 shows an instance where three keywords (K1, K2, K3) co-occur within a document. For full counting (3.3a) the weight of each link is 1, but for fractional counting (3.3b) it is 1/3. The distinction between these two methods has been extensively studied [91]. In this project, full counting is used since it is easier to implement as an analogy with food that ants can collect.

(a) Full Counting (b) Fractional Counting

Figure 3.3: Difference between full and fractional counting for a keyword co-occurrence network

5. Finally, to avoid excessive keywords in the network, a minimum limit of 10 occurrences was placed for the keywords to be considered in the network. All these inputs finally output the network in Fig. 3.2.

The constructed keyword co-occurrence network has 1973 nodes or keywords. There are 106290 unique links among the nodes. This is the number of links in the network disregarding their weight. However, since two keywords can co-occur in multiple documents, the links in the networks have weights corresponding to the number of documents in which they co-occur. Thus, a link with a strength of 5 implies the keywords connected by that link co-occur in 5 different documents. The total link strength for the network is 219594. Each node in the network has an associated coordinate defined by the software. It also provides clusters of nodes based on their relatedness. In this case, eight different clusters have been provided. These clusters are calculated based on a weighted variant of the modularity-based clustering technique, which is commonly used in such bibliometric studies [92]. While they do not accurately represent the specializations within the field, they do provide grouping of similar entities based on the available data. However, within the context of this project, the information about the clusters isn't used. Two other important attributes provided are 'occurrences' and 'average publication year' for each node. The 'occurrences' attribute corresponds to the total number of times for which a keyword has occurred in the entire dataset. The 'average publication year' provides an averaged publication year of all the documents in which the keyword was mentioned.

3.1.3 Refining the Data

The metadata of the network in Fig. 3.2 was imported into MATLAB for refining. The table in Table 3.1 shows the information available for five of the nodes in the network. As explained before, the value in the 'Unique Links' columns essentially signifies the number of nodes a particular node is connected to, while the value in the 'Total Link Strength' column counts the weight of all the links starting from that particular node. Since the weight of all the nodes in this case is 1, the total link strength effectively counts the total number of unweighted links starting from a node. Using the averaged publication year for each keyword, a histogram was constructed to understand the evolution of new keywords in the field of robotics. As shown in Fig. 3.4, new keywords started appearing around the 2000s, with a peak in 2013-2014. Since then, the rate at which new keywords were added has decreased. It must be noted that since only those keywords which occurred more than 10 times in the dataset were selected, this analysis is preliminary and can be improved by a better and comprehensive dataset. However, within the bounds of the data used for the construction of the co-occurrence network, this analysis persists. The evolution of new keywords described here is assumed to be true of the field of robotics, given that a small dataset of around 2000 nodes is used to represent it here. The decline of new keywords over time may be attributed to the difficulty in finding newer concepts consistently. The dynamics of the field assumed true here is used when comparing it to the results from the simulation.

Before the network is used in the simulation, a refining is necessary. The network contains a number of nodes which represent the same concept but differ in spelling, such as 'mobile robots' and 'mobile robotics', or 'adaptive control' and 'adaptive-control'. Thus, a further refining is warranted of this dataset. This was done by using elementary language processing. Concepts which differ only by a plural or a punctuation were merged together. Terms ending with 'robot' and 'robotics' were combined as well. Concepts and their abbreviations such as 'eeg' and 'electroencephalography' were joined too. Other example merges include 'biped locomotion' and 'bipedal locomotion', and 'bioinspiration' and 'biologically inspired'. During the combination of any two nodes, the coordinate and average publication year of the merged node was calculated using a

ID	Label	X	Y	Cluster	Unique Links	Total Link Strength	Occur-rences	Avg Pub Year
7007	config-uration control	0.5910	0.9631	5	36	62	16	1996.5
31177	r&d	-0.3173	-1.0581	1	3	12	11	1996.7
29412	plastics	-0.8582	-0.6655	1	12	35	19	2000.7
7533	control laws	0.4831	1.1264	5	37	57	14	2000.8
42085	variable structure control	0.9754	0.8884	5	33	55	13	2001.1

Table 3.1: First 5 rows of the network nodes with their attributes

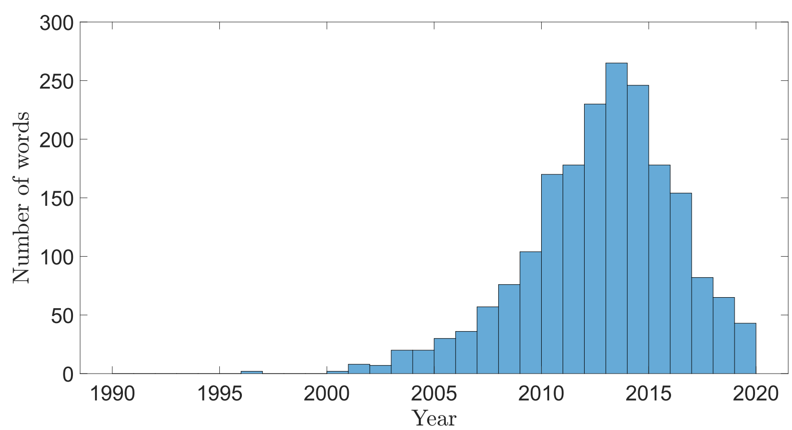


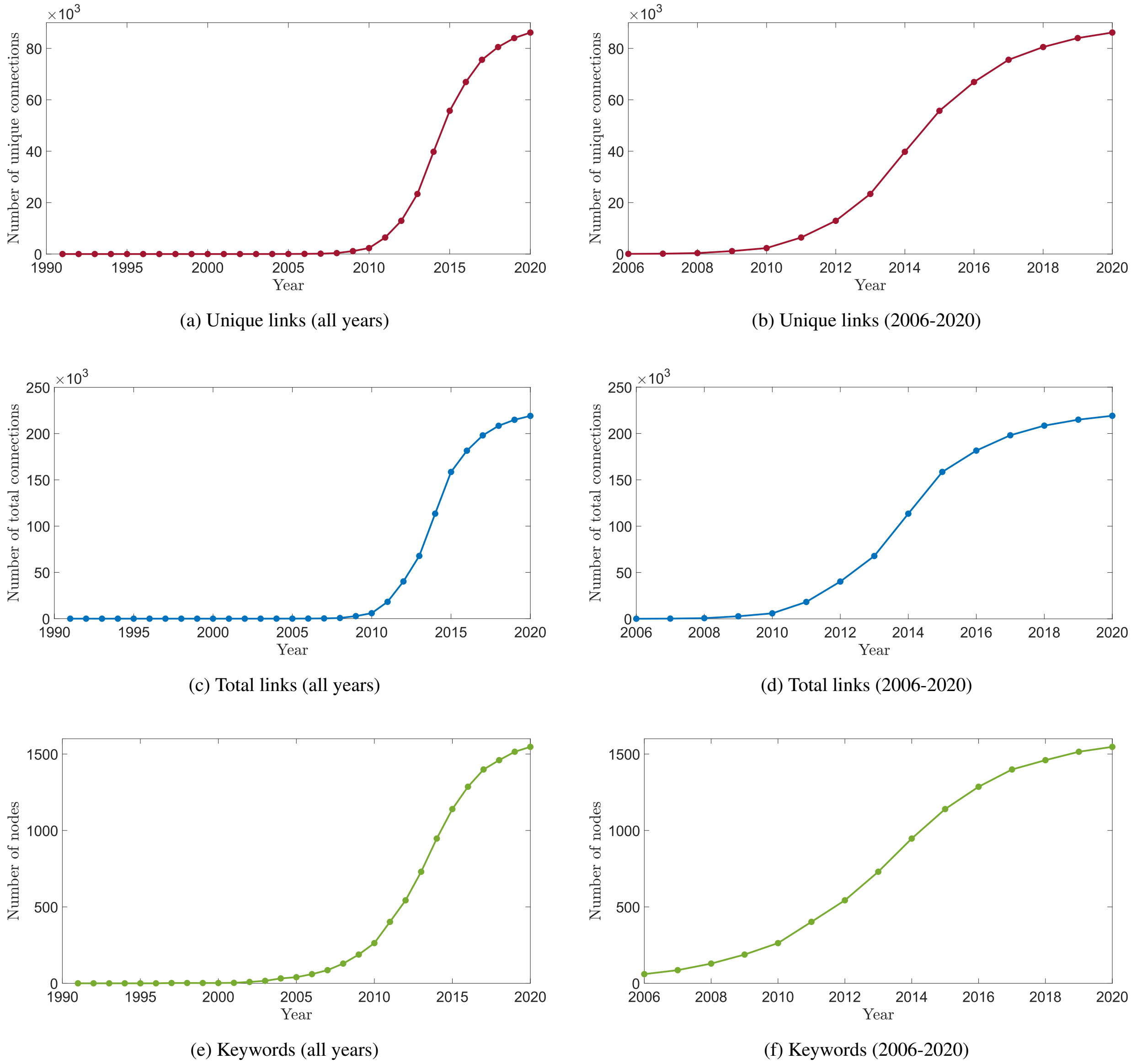
Figure 3.4: Number of new keywords published per year

weighted average of the respective attributes of the two nodes, with the number of occurrences being the weights. This ensures that the resulting attributes are more biased towards the node that had higher number of occurrences. An improved refining can be done by a specialist of the subject. After this processing, the resulting network has 1547 nodes, about 86165 unique links and 219112 total link strength. Thus, while the number of unique links decreased by about 20000, the total link strength decreased only by about 500. This makes sense since the merging of any two nodes involves merging of links connecting these two nodes to the same third node. This was taken as the final network on which simulations would run.

Some further refining included changing node numbers and their coordinates. The node ID shown in Table 3.1 is calculated based on all the keywords parsed by VOSViewer, before setting the limit of number of occurrences. Since there are only about 1500 keywords in the analysis, it makes sense to have node IDs in the same range. Thus, the nodes were sorted in an ascending order of their publication year and were assigned node IDs from 2 to 1548. The node ID 1 was kept reserved for the nest. As mentioned before, the coordinates of the nodes are defined within the software. To simplify the working of the model in terms of calculations of distance between two points, walking speed of ants and number of time steps to go from one point to another, a discrete space was chosen as the environment in which food would be scattered. The coordinates were converted from real numbers to whole numbers. A variable called 'min_node_dist' controls the minimum distance between any two nodes of the network to experiment with different levels of food density. As its value increases, the nodes in the network become farther and farther based on their calculated coordinates. For a constant walking speed of the ants, this implies that ants have to travel more in order to get to the food. A value of 2 was assigned to the variable for initial testing.

3.1.4 Analysis of the Data

The data was further analyzed to determine the number of unique links and total links added to the network per year. Henceforth, the total link strength of the unique edges is referred to as 'total links'. As mentioned before, the link strength can be seen as independent links of unit weight each, since they each correspond to a distinct document that made the connections between



(a) Unique links (all years)

(b) Unique links (2006-2020)

(c) Total links (all years)

(d) Total links (2006-2020)

(e) Keywords (all years)

(f) Keywords (2006-2020)

Figure 3.5: Cumulative number of links and keywords added to the network over the years

the two nodes. Fig. 3.5 shows the cumulative number of links and nodes (or keywords) added to the network over time, with Fig. 3.5a showing the unique links, Fig. 3.5c showing the total links, and Fig. 3.5e showing the keywords for all the years from 1991-2020, and Fig. 3.5b, Fig. 3.5d, Fig. 3.5f showing the same for the years 2006-2020. As seen in Fig. 3.4, a lot of the keywords in the dataset were seen in use after the 2000s. Thus, the dynamics from 2006 onwards is particularly interesting. Table 3.2 provides this data in a tabular form. These plots attempt to depict the evolution of the robotics research over the years, within the context of the dataset. As mentioned before, while a much more rigid method of keyword analysis would give a clearer picture of this evolution, the current truncated data seems to suggest the following points of interest. These are only indicative and are prone to statistical biases, especially because they are based on averaged publication years of these keywords.

Year	Unique Links	Total Links	Nodes	Year	Unique Links	Total Links	Nodes
1991	0	0	0	2006	59	116	60
1992	0	0	0	2007	134	257	86
1993	0	0	0	2008	370	728	129
1994	0	0	0	2009	1138	2759	188
1995	0	0	0	2010	2311	5929	263
1996	0	0	0	2011	6423	18216	402
1997	0	0	2	2012	12923	40225	543
1998	0	0	2	2013	23367	67807	730
1999	0	0	2	2014	39776	113481	947
2000	0	0	2	2015	55716	158537	1140
2001	1	3	3	2016	66917	181530	1286
2002	1	3	9	2017	75567	198070	1399
2003	2	4	16	2018	80525	208461	1460
2004	18	45	32	2019	84019	214914	1515
2005	29	63	40	2020	86165	219112	1547

Table 3.2: Cumulative sum of nodes, unique links and total links added per year

1. The introduction of new keywords points to novel concepts that first appeared in robotics research. This was insignificant at the beginning of the century but began taking off around

2007-2008. A survey of the robotics research published around this time [88] shows the various areas in which robotics was already in full-fledged use.

2. The addition of links to the network represents research being done on the available knowledge. It drastically increased around 2011-2012, probably since the growth of the field attracted many researchers.

3. All the three curves show a saturation towards the year 2020. This might suggest that it became increasingly difficult to find newer concepts, thereby slowing down the research. Fig. 3.4 does show a decline in the number of keywords generated per year since 2014-2015.

Such an analysis would be more accurate from a rigorously constructed keyword network, using natural language processing and field expertise to dictate which keywords refer to the same concept and which ones don't. However, for the purposes of this model, the dynamics depicted by the curves in Fig. 3.5 is assumed to represent the evolution of the robotics field, within the bounds of the dataset.

3.1.5 Translation into a MATLAB graph

After refining, the data was converted into a graph. The advantage of using graphs in MATLAB is that a number of calculations related to nodes and edges become easier due to readily available functions. The graph contains nodes from 2 to 1548 corresponding to the keywords obtained after refining. Node 1 corresponds to the nest. The working of the simulation in terms of this graph is explained in the next section.

Additional attributes were added to the graph to keep a track of the food collected by the ants and the amount of food remaining. At this stage, it is important to clarify what each attribute means and its relation to the original attributes defined in Table 3.1. Tables 3.3 and 3.4 show all the attributes corresponding to the nodes and the links of the graph respectively. Each of these attributes is discussed below.

1. unique_links: The documents in which keywords co-occur form the links or edges in the keyword co-occurrence network. Every document has a unit weight, and multiple documents may

unique_links	total_links	unique_discovered	total_discovered
0	0	0	0
35	62	0	0
3	12	0	0
36	57	0	0
31	55	0	0

Table 3.3: Graph Node Attributes

end_nodes		weight	distance	pheromone	num_links	links_discovered
2	4	Inf	169	0	3	0
2	15	Inf	655	0	1	0
2	19	Inf	113	0	1	0
2	45	Inf	163	0	1	0
2	79	Inf	285	0	6	0

Table 3.4: Graph Edge Attributes

have the same keywords co-occurring. Thus, the attribute 'unique_links' defines the number of non-repeated edges in the network. Every node pair has one unique link. Thus, the number of unique links associated with a node describes the number of other nodes it is connected with. E.g., the node labeled 'configuration control' in Table 3.1 has 36 unique links, implying that it has co-occurred with 36 different nodes in the dataset. Since establishing a unique link implies finding an undiscovered connection between two nodes, the number of unique links represents the number of innovations in the model. In the Fig. 3.6, nodes 'A' and 'B' have multiple links between them, implying multiple documents in which the two keywords have co-occurred, but they have only one unique link between them. In Table 3.3, 'unique_links' counts the number of unique links from that node remaining to be discovered.

2. unique_discovered: This column counts the number of unique links associated with the partic-

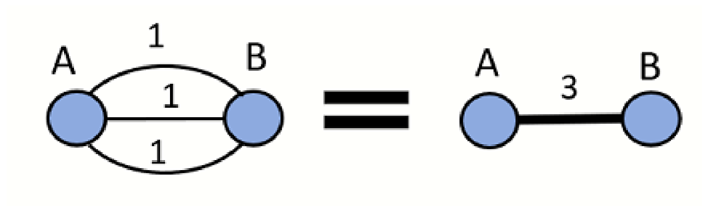


Figure 3.6: Nodes with 1 unique link and 3 total links

ular nodes that have been discovered by the agents during the simulation. The sum of values in 'unique_links' and 'unique_discovered' for any node gives the total number of unique links for that node in the original network.

3. total_links: As seen in Fig. 3.6, two nodes can have multiple links between them. The column 'Link Strength' in Table 3.1 refers to the total number of all links emanating from a particular node. The same is given by the attribute 'total_links' in Table 3.3. The link strength of 3 in Fig. 3.6 simply implies the existence of one unique link and two repetitive links between the two nodes, giving a total of 3. As the simulation progresses, agents discover all the links from a node, till the value in 'total_links' goes to zero. Multiple links imply repetitive establishment of a connection between two ideas. Thus, these are considered to be a result of conservative non-risky research.

4. total_discovered: Similar to 'unique_discovered', this attribute counts the total number of links discovered – unique and repetitive. The sum of values in 'total_links' and 'total_discovered' for any node gives the total number of links for that node in the original network.

5. end_nodes: This column in Table 3.4 simply defines the nodes between which the edge of that particular row lies. E.g., the edge corresponding to the first row in that table is the one between nodes 2 and 14.

6. weight: This is an attribute defined by MATLAB and has nothing to do with the attributes defined so far. This is kept since it allows calculating shortest paths between nodes. Since the graph created is weighted, the algorithm calculating the shortest path considers the weight of

the edges as well. Thus, an edge with 'weight' Inf implies that it has never been discovered by an ant and thus cannot be traveled on while returning to the nest. When it does traverse on an edge, the 'weight' of that edge is changed to 1. This allows it to be considered when calculating the shortest path. Other than this usage, there is no relation of this attribute to link strength or any other properties of the graph.

7. distance: The nodes in the graph created by MATLAB have no exact position unlike the keyword network. Thus, to make use of the coordinates provided in the 'X' and 'Y' columns of Table 3.1, every edge is given a distance calculated using the coordinate information of the nodes it connects. In order to keep the distances whole numbers, Chebyshev distance was used [93].

8. pheromone: This attribute simply keeps track of the amount of pheromone on the edges. It is used by the ants during their foraging, which will be explained in the next section.

9. num_links: This describes the link strength of a particular edge. As shown in Fig. 3.6, 3 different links between the two nodes are represented by a single link in the graph with a link strength of 3. The 'num_links' column for the edge in the figure would contain the value 3. In the Table 3.4, the first row implies that there are 3 links between nodes 2 and 14 (1 unique and 2 repetitive), while the second row implies there is just a single link between the nodes 2 and 15.

10. links_discovered: The 'links_discovered' column keeps track of how many of the multiple links corresponding to an edge have been discovered. This is similar to 'total_discovered' for the nodes in Table 3.3. The sum of values in 'num_links' and 'links_discovered' for any edge gives the total number of links between the two 'end_nodes' connected by that edge. Furthermore, the first link discovered for any edge is considered to be a unique link, while the rest are considered repetitive.

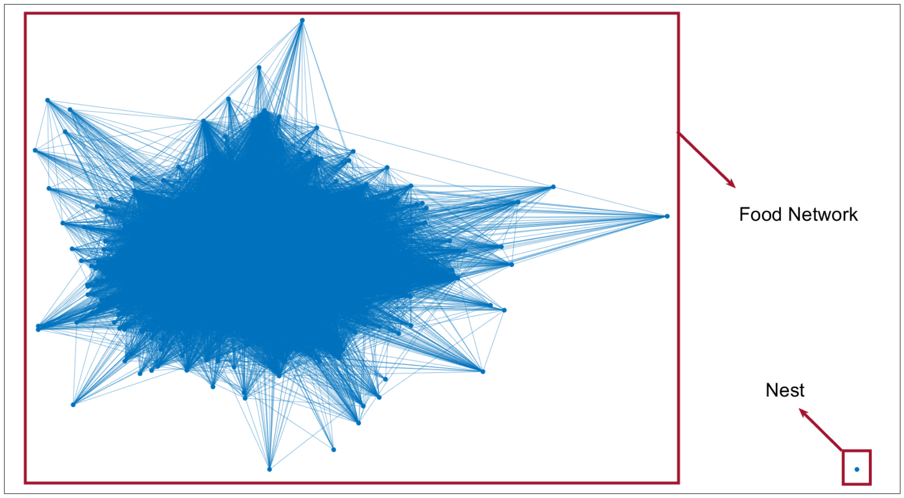
3.2 Simulation Environment and Agents

This section discusses the aspects of the simulation model and provides a pseudocode of the working. All the involved parameters are described in detail.

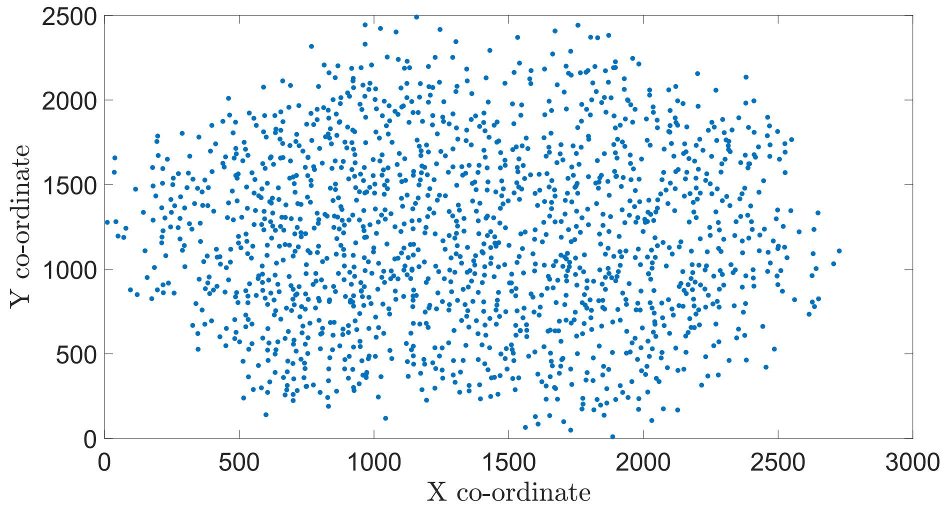
3.2.1 Environment

The keyword network obtained from the VOSViewer software and refined in MATLAB is used as the environment. As described before, the nodes and the links were converted into a weighted graph, with the 'total_links' attribute corresponding to number of times the nodes connected by that link have been mentioned together in the dataset. Fig. 3.7a shows this graph. All the links are undiscovered initially, and ants have to discover them gradually over time. Clearly, the nodes have no specific position as they do in Fig. 3.2, but they have the same connections. Fig. 3.7b shows the same food network spread out in a gridspace. Every food point has an associated co-ordinate as shown in this figure. Additional attributes like distance calculated from these co-ordinates add information to the graph. An additional node corresponding to the 'Nest' has been added to this graph. This node is not a part of the original food network, but is required for the simulation to act as a starting and ending point for the ants.

The nest in the simulation mimics all the existing knowledge in the field of robotics at any given time, also referred to as the 'contemporary intellectual structure' [14]. Initially, three nodes are connected to the nest, corresponding to those with the earliest average publication years in Table 3.2. Ants start from the nest, travel to any of these nodes, and then start foraging for food. Fig. 3.8 shows a schematic of the different types of connections in the graph. The red line between the nest and the food particle 'A' is a nest connection that is added during the simulation. This link represents that node 'A' forms a part of the contemporary knowledge. Thus, 'A' is in fact a part of the nest. It can be seen as one of the openings to a polydomous ant colony. Blue lines show real connections that exist in the food network obtained from the dataset. Ants travel to node 'A' from the nest, then try to discover connections to any of the nodes 'B', 'C' or 'E'. This is decided based on a number of factors discussed later. Suppose the ant ends up going to 'B', then it collects



(a) Graph constructed in MATLAB showing nest and food network



(b) Food network with proper locations and without connections

Figure 3.7: Difference between original food network and graph in MATLAB

one unit of the food particle at 'B' and bring it back to the nest via 'A'. This corresponds to one link between nodes 'A' and 'B' as being discovered. The 'Number of Links', corresponding to the 'num_links' attribute for every graph edge in Table 3.4, shows the number of times the ant can start from a particular node and collect food at the next node. If the ant ends up going to 'E' from 'A', it cannot collect any food since the number of links between these two nodes is zero, implying all the links have already been discovered. Thus, it continues from 'E' and reaches any of 'D' or 'F'

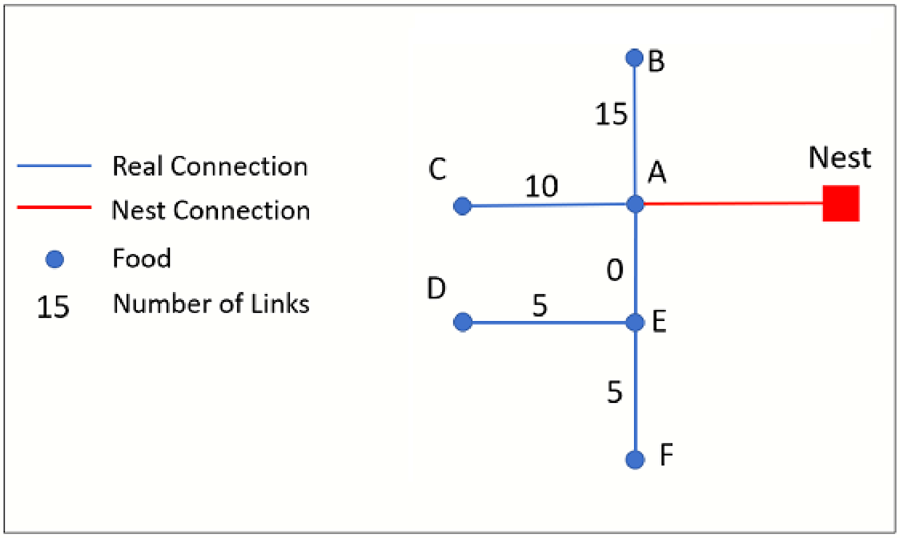


Figure 3.8: A schematic showing the connections between the nest and the food

from where it can collect food and return to the nest via ‘E’ and ‘A’. As time goes on, more nodes are connected to the nest, so that ants can directly start from them and continue foraging. The simulation ends when the ants successfully discover all the connections and collect all the food. In this simulation, the food that the ant collects is dependent on the edge from where the ant reaches it. Thus, even though node ‘E’ has ‘total_links’ 10 (5 links with ‘D’ and 5 with ‘E’), the ant cannot collect any food if it reaches ‘E’ from ‘A’. This is where the working of the simulation departs a bit from the ant behavior. Since the links represent new research connecting nodes, ants have to discover all the links in order to collect all the food. Thus, the ant only sees food at a node as long as there are still connections to be made between that node and the one it came from. If all the connections have been made between those two nodes, the ant will simply continue towards other nodes.

3.2.2 Agents

Ants act as agents in this simulation. They walk around the environment, collect food, and deposit it at the nest. The following parameters define the behavior of ants. Specific parameter sets can thus be constructed to mimic the foraging strategies of different species of ants. Later, these parameter sets can be mixed together to find the best performing set.

1. Exploring and exploiting ants: The simulation initializes two different types of agents - exploring and exploiting. This is to implement researchers who are risky and seek innovative discoveries, and those who are more conservative and go for safe but repetitive work. Exploring ants are incentivized to find food resources which haven't been discovered yet, or which were discovered a long time ago and haven't been completely exploited. Such agents in the simulation can only collect food if there are no pheromones in the vicinity of the food. This covers both the cases – either the food has never been discovered before or there aren't any ants in the process of collecting it. The distinction between these two types of foragers, often referred to as scouts and recruits, has been seen in ant species like *Atta colombica* (leaf-cutting ant) [94], *Solenopsis invicta* (red imported fire ant) [95] and *Lasius niger* (black garden ant) [96]. Thus, the existence of division of labor within foraging in these species provides a reasonable basis for using the two different types of ants in the model.

2. Type of recruitment: Ants have been shown to use different types of recruitment after finding food and returning to the nest [97]. The two types used in the model include solitary foraging and mass recruitment. When pheromone use is set to off, the type of recruitment seen in the model is solitary. Otherwise, mass recruitment is used.

   (a) Solitary foraging: This is a form of zero recruitment, where ants simply find food and return to the nest without any communication with each other outside the nest. Desert ants of the genus *Cataglyphis* are the best known examples for this kind of foraging [62]. The extreme heat in their habitat makes the use of pheromones inefficient. Under this strategy, agents in the model simply find food, return to the nest, and set out to find more.

   (b) Pheromone trails: Use of chemical trails for mass recruitment is considered to be one of the most sophisticated forms of communication among ants [98, 67]. Whenever an ant finds food, it starts laying a pheromone trail while returning to the nest. The scent of this trail stimulates other ants to follow it to the food source. If these ants find food, they enhance the trail on the way back. Otherwise it is not reinforced, and dissipates over time. When the food is completely extinguished and the trail has evaporated, ants no longer follow that path. After returning, other ants follow this trail and reach the food resource. They might either collect the food, or continue foraging. While returning with the food, they reinforce the trail. At every time step, the trail dissipates at a constant evaporation rate.

3. Pheromone properties: When pheromones are set to be used, various other properties related to the pheromones can be added to the model. Evaporation rate is a property of the chemicals present in the pheromone, and can be different for different species. Thus, when this information is available, a proportional value is used. Similarly, the amount by which an ant reinforces a trail when coming back from food varies a lot among the species. These trails can last anywhere from minutes, as in the case of the species *Aphaenogaster albisetosus* [99], to days, as in the case of the leaf-cutting ant *Atta cephalotes* [100]. In fact, it has also been shown that the leaf-cutting ant controls the amount by which it reinforces the pheromone trail while returning from food based on the food quality, the number of ants already at the food source and the current concentration of the trail [101]. Thus, the model can set ants to reinforce pheromones at a constant rate or based on the quality of the food source they are returning from. Moreover, some species like Pharaoh ant *Monomorium pharaonis* have been shown to use a special repellent type of pheromone [65]. This acts as a 'no entry' signal on unrewarding routes, especially paths that no longer lead to any food. The model allows the use of such repellent pheromones as well.

4. Attribute preferences: Whenever an ant is at a node, it is faced with numerous paths to choose from while jumping to the next node. There are different attribute preferences that can be set at the beginning of the simulation that decide what factors the ant will consider in choosing a path. This allows for different kinds of strategies that ants might employ in discovering food. There are three attributes included in this model.

   (a) Pheromones: Every path has some pheromone strength associated with it. Ant can choose to select the path that has the highest pheromone. Of course this is only a valid preference if the species uses pheromone trails.

(b) Food weight: Ants in the simulation can also decide to go after a large food morsel instead of a smaller one. The total link strength of all the links emanating from a node represents the size of that node. Ants can make a choice based on this perceived node size. It has been observed that solitary foragers such as the Sahara Desert ant (*Cataglyphis bicolor*) [102] usually go for small food pieces that can be carried by individual ants. On the other hand, species which implement mass recruitment such as the Argentine ant (*Linepithema humile*) [103] prefer stable food sources so that recruitment is beneficial.

(c) Food distance: Finally, ants in the model can also choose a path based on the distance to the food. Every edge in the graph has an associated distance calculated from the positions of the two nodes. Ants can either select a path that has the shortest distance, or it can go for longer distances. Species such as the black garden ant (*Lasius niger*) [96] have been shown to alter their foraging and recruiting behavior based on the food distance.

The preferences for these attributes are chosen such that they can be positive or negative and their absolute values sum to one. Thus, an ant which prefers strong pheromones, large food weights and long distances will have a preference of 33% for each. Furthermore, since the behavior of exploring and exploiting ants is quite different, these preferences are set differently for both. The absolute value for an attribute determines how strongly the ant prefers it. Thus, an ant preferring strong pheromones by 20% will be less determined to follow the path of the highest pheromone strength than an ant preferring the same by 80%. These preferences allow a level of heterogeneity among the ants. While there are still only two types of ants in the simulation, these preferences permit designing varied behavior of foraging.

3.3 Simulation Algorithm

3.3.1 Main simulation

Once these configurations are set, the simulation can be set to run. The following discussion simplifies the logic of the simulation into a series of steps that a particular ant goes through when completing a trip from the nest to the food and back to the nest. Algorithm 1 provides a general pseudocode for the model.

The formatting of this pseudocode is largely done in the style of the original MATLAB script. During the initialization of the environment, the graph created from the data is imported. As discussed before, the nodes of the graph represent keywords while the edges represent connections among these keywords established by the documents. In the ant analogy, the nodes are food particles and the links are pathways to these food particles. Furthermore, the earliest three nodes are connected to the nest. According to Table 3.2, links started to be established in 2001 and by then, there were three nodes discovered. Thus, these three nodes become part of the existing knowledge. Ants are initialized at node 1, which is the nest. During this process, every ant is provided a table for keeping track of its current location, its previous location, the time at which it will reach its next location, whether it is exploring, whether it is foraging, the total time spent searching for food since leaving the nest and the amount of food collected yet. A specified percentage of the ants are randomly configured to be exploring. The number of exploring an exploiting ants is recorded. All the ants are given a current location corresponding to the nest. This concludes the initialization part and the simulation begins.

The food that an ant can collect is dependent on the link strength of the node that the ant came from and the node that the ant is headed towards. This is given by the 'num_links' column in Table 3.4. Thus, the first node that the ant visits from the nest cannot be collected, since any edges that start from the nest aren't a part of the original keyword network. The following list of functions explains every step of the pseudocode in greater detail.

Algorithm 1: Main Simulation

```
Initialize environment, configure ants and place them all at nest
while 't' <= time limit do
    while 'a' <= number of ants do
        if ant 'a' has reached its destination at time 't' then
            if ant 'a' is at nest and is foraging then
                call function leave_nest( )
            else if ant 'a' is at a node and is foraging then
                call function find_food( )
            else if ant 'a' is at a node and has found food then
                call function find_nest( )
            else if ant 'a' is at nest and has found food then
                call function enter_nest( )
            end
        end
    end
end
if 't' is a multiple of a specified number of time steps then
    call function remove_ants( )
    call function add_ants( )
    call function connect_nodes( )
end
if pheromones are used then
    decrease all pheromones by specified evaporation rate
end
```

3.3.2 Function leave_nest( )

The first step in the ant's journey starts from the nest. Since ants are initialized at the nest and they are foraging or searching for food, this is the first function that is called. The following pseudocode Algorithm 2 explains the details happening in that function.

Algorithm 2: Function leave_nest( )

Get nodes connected to the nest and store in variable *nodes*
Get edges corresponding to the connections of *nodes* and the nest, store in *edges*
Calculate *food_weight_probabilities* based on 'total_links' column of *nodes*
Calculate *food_distance_probabilities* based on 'distance' column of *edges*
if *pheromones are used* **then**
    Calculate *pheromone_probabilities* based on 'pheromone' column of *edges*
else
    Set *pheromone_probabilities* to be equal
end
Choose three path choices, each using one of the three types of probabilities
Select one of the three path choices based on attribute preferences
Update ant's location
Set the ant to arrive at the chosen node after one time step
Return to main

The mention of columns in this algorithm refers to the Tables 3.3 and 3.4. All the attributes of nodes and edges used in the simulation are described in Section 3.1.5. Since the edges from the nest to the nodes aren't present in the original network, they do not have an associated distance. Thus, the ants are set to arrive at one of the nodes neighboring to the nest within one time step. This can be seen as a polydomous ant colony with multiple nest openings. In the researcher analogy, these are just starting points of research. The next step to do is to actually find food or an innovation starting from that point.

3.3.3 Function find_food( )

Once the ant is at one of the nodes neighboring to the nest, it can start finding food. Algorithm 3 provides the pseudocode for this function.

In this function, the lines 2 to 14 check if the ant is already at a node from where food can be collected. This is dependent on the 'num_links' attribute shown in Table 3.4. As mentioned before, this shows the number of times an ant can go to a food node and collect it. Which food the ant can collect is also dependent on where the ant came from. Since initially the ant would arrive at a node from the nest, the edge connecting the node to the nest is checked for any valid connections that can be established. This edge however is not a part of the original network and doesn't have any links strength. Thus, lines 2 to 14 in the function are skipped. If the edge does have some link strength, this implies the ant can collect the food at that node. However, there are certain checks to ensure that exploring ants behave differently from exploiting ants. If the edge has never been traversed before, or that the 'links_discovered' attribute for that edge is equal to zero, then any ant can collect the food. On the other hand, if the food at the end of that edge has been collected before, then the exploring ant can only collect it if there is no pheromone on the edge. This ensures that the exploring ant always tries to collect food that it thinks hasn't been collected before. The absence of pheromone denotes that either the food indeed hasn't been discovered yet or that the ants have failed to exploit it completely. In both cases, it is beneficial for the exploring ant to reconnect the food morsel to the nest. If the food can be collected, the function 'find_nest()' (detailed in the next subsection) is called and the rest of the function is skipped. However, if the lines 2 to 14 are skipped, then the ant starts to consider the next neighboring nodes to the current location. This procedure is the same as in Algorithm 2. The only difference lies in the time that the ant takes to reach the next node.

Algorithm 3: Function find_food( )

```
Find edge between ant's current location and previous location, store in edge
if 'num_links' column of edge is positive then
    if 'links_discovered' column of edge is positive then
        if ant is exploring OR 'pheromone' column of edge is zero then
            Collect food, set ant to not foraging
            Call function find_nest( )
            Return to main
        end
    else
        Collect food, set ant to not foraging
        Call function find_nest( )
        Return to main
    end
end
Get nodes connected to ant's current location, store in nodes
Get edges corresponding to these connections, store in edges
Run same steps as function leave_nest( ) and find the next location, store in dest
Get edge connecting ant's current location to dest, store in path, update ant's location
Calculate min_time to reach based on path distance and ant's walking speed
if 'pheromone' column of path is positive then
    Set ant to arrive at dest within min_time number of time steps
else
    Set ant to arrive at dest anywhere between 1.5 to 2 times the min_time number of time
      steps
end
Return to main
```

Since the edges corresponding to links in the original network have nonzero distances, the ant will take some finite time to reach one of the nodes at the ends of the edges. The minimum time is calculated simply by dividing the distance of the chosen edge by a specified walking speed of the ant. When a pheromone trail is present on the edge, the ant can simply run at maximum speed possible and reach its destination. This, in that case, the ant is set to reach within minimum time. On the other hand, the ant has to explore when there are no trails. This can take longer than the minimum time. Hence in that case, the time required is randomly chosen between 1.5 to 2 times the minimum time. A better estimate may be obtained through some quantitative studies of ant exploration. However, no such conclusive results were found at the time of writing this. This also adds some randomness regarding the efficiency of exploration every ant has. Some ants will take lesser time than others, implying that they are faster in understanding their environment and finding food. It analogizes to the fact that some researchers can reach their goal faster than others.

3.3.4 Function find_nest( )

Once the ant has found food, it needs to return to the nest in the shortest path possible. As mentioned in one of the points in Section 2.3.3, ants follow the most direct path possible to the nest, even if their exploring might have been long and winding. Algorithm 4 describes the pseudocode for the ant once it has found the food and has to get on the way towards the nest.

Since the edges in the graph are set by the original keyword network, it is not possible to add a new direct link from every node that the ant collects food from to the nest. A link connecting the nest to a node implies that the node is part of the existing knowledge for the colony. This cannot be ensured just within a single ant visiting it once. Thus, instead of this direct link, the implication that the ant is taking a shorter path to the nest is given by taking the root sum of squares of the distances corresponding to the edges that make up the shortest path to the nest. This is further explained in Fig. 3.9.

Algorithm 4: Function find_nest( )

Get ant's current location, store in *node*

Get edge connecting *node* to ant's previous location, store in *edge*

Update all relevant columns of *node* and *edge* to show food has been collected

Find shortest path from *node* to nest, store all the edges on this path in *return_path*

Find root sum of squares of all the values in the 'distance' column of the edges in *return_path*

Calculate *time_required* by dividing the root sum of squares value with ant's walking speed

Set ant to arrive at the nest within the *time_required* number of time steps

Return to main

In this figure, the red square is the nest and the blue circles are the nodes. The blue edges denote links that are present in the original keyword network, while the red edges are links that have been added to the network for the purposes of the simulation. Thus, the red edges do not really contain any real attributes such as link strength or distance associated with them, but can be thought of as pathways that ants take to travel. In the figure, an ant starts from nest and immediately ends up at node 'A'. As mentioned before, node 'A' can be considered to be a nest opening of a polydomous colony. Since the link connecting node 'A' to the nest is red, or not in the original network, the ant cannot collect any food at 'A'. Thus, it searches for other nodes and ends up at 'B'. The link 'AB' has a distance of 30 units. If it had non zero link strength, or 'num_links', the ant would be able to collect at 'B'. However, since it doesn't, the ant moves on from 'B' and travels to 'C', traveling another 40 units of distance corresponding to link 'BC'. Now 'BC' has finite link strength, which means the ant can collect food at the node 'C'. The direct path the ant takes from node 'C' to the nest is shown by the red arrow from 'C' to the nest. Since this is not

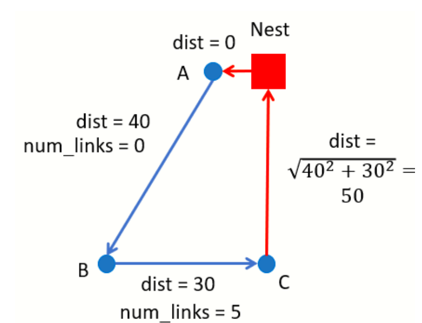


Figure 3.9: Ant taking a shorter path to the nest

present in the original network, the model makes the ant traverse through the shortest path possible using the existing links, but considers a shorter distance to travel. In the figure, the shortest path is 'C-B-A-Nest', which has a total distance of 70 units. Taking a root sum of squares of the distances associated with the two edges 'BC' and 'AB', the distance that the ant has to travel becomes 50 units. Thus, for the same walking speed, the ant reaches the nest in lesser time than it took to reach 'C' while foraging. This enforces the idea of taking the shortest route to the nest without the need to add new links to the network. It must be noted that ants do not physically travel the links in the graph. The links only represent the connections among the nodes and not the actual pathways. Thus, a manipulation of the time taken to reach one node from another takes into account the distance traveled by the ant, which might be lesser than the actual distance between the nodes as in this example.

3.3.5 Function enter_nest( )

Finally, once the ant reaches the nest with the food, it deposits the food and gets ready to forage again. The other ants in the nest are notified of the existence of a new pheromone trail. Based on the set preferences, the pheromone attributes of the graph edges are updated. If repellent pheromones are set to be used, the pheromone at a completely extinguished food node (that is, with zero 'total_links') that the ant's path went through is set to be zero. This allows other foraging ants to focus on paths where pheromone is still present to exploit known sources of food. This function makes all these changes and concludes the journey of an ant in the simulation.

3.3.6 Function remove_ants( )

Every few time steps, some ants are removed. This can be attributed to a number of reasons that cause ants to perish. In the researcher analogy, this implies a dead-end in the project or exhaustion of available resources. A failure probability is set at the starting of the simulation which decides what percentage of ants are removed. The ants to be removed are chosen based on their total experience and their current search time. Ants who have collected a lot of food are considered to be safer when traveling far from the nest as opposed to ants which haven't collected a lot of food.

Thus, higher the time for which a particular ant has been searching and lower the number of food particles it has already collected, higher is the probability that it will be removed. In the innovation analogy, this implies researchers who are experienced in the field will probably be able to navigate unknown research territories, while novice researchers may easily get lost.

3.3.7 Function add_ants( )

This function simply adds a specified number of ants to the total foraging population. A balance between this addition and the removal of ants maintains a constant ant population. Based on the exploration percentage, some of the ants are set to be exploring.

3.3.8 Function connect_nodes_to_nest( )

Once a particular node has been completely exhausted, it is connected directly to the nest. This implies it becomes a nest opening for the ants. Instead of following a long route to that node, the ants can simply start from there as if exiting from an ant opening and forage nearby food. This improves the food collection rate.

3.4 Parameters and Variables

The bio-inspiration in this project comes from the knowledge of ant behavior and their strategies in foraging food efficiently. Different species use different techniques based on their environment, availability of food, seasonal changes as well as presence of predators. An analogy between these techniques and the strategies used by researchers in innovation process implies that the simulation can be fed various ant strategies and the results obtained can be translated into actions that researchers can take to boost their creative output. Thus, this section looks at all the parameters and variables contained in the model.

3.4.1 Constant Parameters

Some of the parameters in the simulation are such that setting their values higher guarantee a higher food-finding rate. This includes total number of ants and walking speed among others, even though these quantities differ species to species. To focus on the efficiency of strategies rather than such parameters, their effects are tested for different values and a constant value is assigned to them. These parameters are discussed below.

1. Time limit: The simulation is run for a constant number of time steps and the amount of food collected is noted. In order to decide what time limit is sufficient for ants to discover most of the food and compare the food-finding rate to the curves in Fig. 3.5, the simulation is run for different time limits and the overall performance is tested. The best time limit is decided such that it does not result in long computation time and provides meaningful data for comparing performance with other simulation runs.

2. Number of iterations: Due to considerable randomness in the model, from choosing which ants are exploring to the path choice that each ant faces at every step, multiple iterations can result in different results. To test the variance in these results, the simulation is run with the exact same parameters a few times. Based on the similarity of results obtained from these iterations, the number of iterations is decided for the next runs.

3. Number of ants: The number of ants plays an integral part in the rate of discovery of food. More the number of ants in the field, higher is this rate. Some ant species have larger foraging population than others, which works to their benefit in finding food and collecting it faster. However, since the primary motive of this model is to test the efficiency of search strategies, simply increasing the number of ants to get a better performance isn't fair. Thus, for all the simulation runs, the number of ants is kept the same. This does not mean all the ants are initialized to find food immediately. As shown in Algorithm 1, some ants are added and some are removed every few steps. Thus, the number of ants is dependent on three parameters:

   (a) Initial number of ants: This is the number of ants that are set out of the nest when the simulation starts.

   (b) Added number of ants: This is the number of ants that are added every few steps.

   (c) Failure probability: This is the percentage of ants that are removed every few steps.

The effect of tweaking these three parameters on the total number of ants is tested. The total number of ants in the field is set based on the values of these parameters. In cases where information about number of ants or failure probability is available for a particular species, an effort is made to use that information while keeping the total colony size same.

4. Walking speed of ants: This is another parameter that differs species to species. Some ants are faster at foraging than others, and thus are able to find food faster. As with number of ants, a faster walking speed obviously results in a faster discovery rate of food. Thus, different walking speeds are tested and a reasonable speed is assigned to all the ants in the simulation. As mentioned in Algorithm 3, the time required for an ant to go from one node to another isn't simply the ratio of distance to walking speed in the absence of pheromones. The lower walking speed while searching for food is represented by higher time taken to reach the next node.

5. Pheromone reinforcement rate: As ants return from a food source, they lay a trail or reinforce an existing one. The strength of this pheromone is different for different species and depends on the chemical composition of the pheromone. This can also vary based on environmental factors. To simplify the use of pheromones, a constant rate of reinforcement is assigned. Since the pronounced use of pheromone trails is largely based on the number of ants reinforcing it rather than the reinforcement strength provided by each ant, it is reasonable to use a constant strength. This strength can be set based on the pheromone evaporation rate.

6. Pheromone evaporation rate: Similar to reinforcement, the evaporation rate of pheromones is also dependent on the chemical composition as well as the environment. Thus, a fixed evaporation rate is set to ensure that pheromones don't last forever.

These parameters define the basic working of the model. Some other parameters such as the number of time steps when ants are added or removed are set randomly.

3.4.2 Variables

Once the constant parameters are set, the variables can be assigned different values to obtain different ant behaviors. These variables are defined below.

1. Exploring percentage of ants: This defines what percentage of the ants outside the nest acts as scouts and what percentage acts as recruits. A balance between the number of exploring and exploiting ants is tested initially. Some early results test the effect of increasing the value of this variable on the overall food-finding rate. In the later runs, this is set according to the information available for different ant species.

2. Use of pheromone: This is a boolean variable that dictates whether or not pheromone trails are used in the simulation. Some ants like Desert ants [62] do not use trails, while most other species do. If the pheromone trails are set to be used, several other variables can be set based on different ways in which pheromones are used.

   (a) Proportionality to food: Some ant species such as Pharaoh ant [75] increase the strength of the pheromone while laying a trail back to the nest if the food is of a good quality. This variable captures that behavior. If set to true, ants while returning lay a trail whose pheromone strength is proportional to the food resource they just found.

   (b) Repellent pheromones: Pharaoh ants have also been shown to use repellent pheromones [65] to warn other ants of exhausted food sources or predators, especially at trail bifurcations. This variable lets ants deposit such a repellent pheromone at one of the branches while returning, if the ant notices that the food at the end of that branch has been completely exhausted. Thus, ants can then focus on the remaining branches where food is available.

3. Attribute preferences: These preferences denote the way in which ants choose one path among many while foraging. The different attributes are explained in Section 3.2.2. The values are set as percentages. These preferences allow modeling a gamut of behavioral strategies.

3.5 Parameterized Ant Species

Some species of ants show exemplary behavior that makes their foraging very efficient. Such species are discussed in this section and their strategies are converted into parameter values so that the model ants can implement them.

3.5.1 *Cataglyphis bicolor* (Sahara Desert ant)

Cataglyphis bicolor, commonly known as Sahara Desert ant, is a well-known ant species that uses solitary foraging methods. These methods have been extensively studied [102, 104, 105] and even applied to searching models [106]. Its simple habitat and well-defined foraging rules make it an easy species to model. Some of these references explain detailed methods that were used to understand the ecology of these ants. Much of the data and observations from experiments mentioned in Ref. [105] can be used in the model. Foragers face a mortality rate of 16%. The species has a polydomous colony, meaning the ants of the same colony occupy multiple nests. The total number of foragers averaged approximately to 290 during the entire experiment. Thus, a large number of foragers are added to compensate the high mortality rate. The efficiency of foragers increases with the number of times they forage. Another study [104] on the individually different foraging methods that these ants employ can provide input regarding ant preferences. Some foragers persistently search an area where food was found before, others don't. This is variable, and isn't dependent on type of food. While an exact percentage is difficult to calculate, it is suggested that most of the ants do not bother thoroughly searching again the place where they previously found food. Due to the high mortality rate, ants show less selectivity towards food. Furthermore, it has also been observed [107] that with more foraging runs, ants travel larger and larger distances. As they collect more food, they adopt one of the two strategies: concentrate on sites where they have previously found food or venture out as far as they can. Using all such information, the ant can be parameterized and its behavior can be implemented in the model.

1. Number of ants: As mentioned, ants have been reported to face a mortality rate of 16%. Furthermore, around 300 ants have been observed to regularly forage. However, since these ants

also have a high worker turnover rate, an equally high number of ants would need to be added. This can cause a faster rate of food discovery due to sheer number of ants. In order to avoid this, the failure probability is kept constant at 1% and 3 ants are added to the foraging population every few time steps.

2. Exploring percentage of ants: Since pheromones aren't available for this species, the only difference between exploring and exploiting ants can be set in terms of their preferences while foraging. In this case, exploring ants can be defined as individuals which forage in new areas every time, while exploiting ants are the ones which persist in areas where they found food previously. Since most of the ants were shown not to persist, the exploring percentage can be kept high – about 80%.

3. Attribute preferences: The absence of pheromones implies there is 0% preference to pheromones. Exploiting ants, as defined above, persist going to locations where food was previously found. Since there is no long-term memory in the model ants, a preference to large food weight can be set for exploiting ants. This allows them to follow the same path most of the time, while the food is still sufficiently large. Exploring ants on the other hands can be given a lesser preference to food weight, since they do not care about going to the same spot. Thus, it is set at 75% for exploiting ants and 20% for exploring ants. The preference for longer distances is present in both cases, albeit in a smaller amount for exploiting ants than for exploring ones. This preference can also be set to increase according to the experience of the ant. Thus, this preference is set between 2% to 25% in exploiting ants, and 20% to 80% in exploring ants, based on the amount of food collected by them.

These values for the variables provide a basis for a simplistic implementation of the foraging methods of *Cataglyphis bicolor*.

3.5.2 *Linepithema humile* (Argentine ant)

This is one of the invasive ant species that has spread throughout the world [108]. Studies have shown that the colonies of Argentine ants are polydomous as well [103]. These ants use persistent trails among the different nests [103]. The process of searching for food in a new territory occurs through collective exploration [109, 110] where ants randomly venture out in different directions and lay trails, even while exploring [111]. Workers of this species have a tendency to orient themselves according to these trails during the exploratory part [110]. Thus, recruitment occurs from these trails rather than from the nest. A study of resource distribution and retrieval rate among Argentine ants [112] showed that discovery times for a large concentrated food resource were the same as for small spread out food resources. However, large but rare food sources were collected at a faster rate than small but common ones. Thus, while it does not exactly dictate food size preference for the ants, the increased odor and visual perceptibility of a large food morsel helps them find these resources. The study also calculated a range of 3.3% to 23.7% of the total forager population to be scouts at different stages of foraging. The ants have been shown to use convoluted searching path when the ant density in the field is high, and straighter paths when the density is on the lower end [109]. This suggests a preference for food distance that is dependent on the number of ants foraging in the area. This information can be used to parameterize the ant species for the model.

1. Number of ants: Argentine ant colonies have a very large number of workers, ranging from 70000 to 374000 according to one report [113]. There is not much information about their mortality rates or the worker turnover rate. Thus, to avoid increase in efficiency due to sheer numbers, the maximum number of ants was kept at a constant 300 by having a mortality rate of 1% and an addition rate of 3 ants (1% of 300).

2. Exploring percentage of ants: The percentage of exploring ants, or scouts in this case, ranges from 3.3% initially to 23.7% at the maximum number of ants. Thus, exploring percentage is dependent on number of ants. It can be set so in the model. At the start of the simulation, 3.3%

of the total ants are exploratory. As the number goes to maximum, the percentage of exploring ants goes to 23.7%.

3. Attribute preferences: One of the most distinct aspects of the foraging method that this ant species employs is the continuous use of pheromones. Trails are laid during exploration as well as while returning to the nest. The exploiting ants are highly inclined towards following trails, with little to no preference for food size or distance. The exploring ants also follow trails, but are perceptive to large food size and choose to go long distances when there are fewer ants in the field. Thus, the preference for pheromones is set at 80% for exploiting ants and 25% for exploring ants. For the exploiting ants, the food size and food distance preferences are set at 10% each. For the exploring ants, food size preference is set at a constant 10%, while the preference for food distance decreases from 65% to 10% as the number of ants increases.

3.5.3 *Lasius niger* (black garden ant)

Lasius niger is another ant species that has been extensively studied for its foraging behavior. This mass-recruiting species is one of the few that tries to use pheromones as well as its memory efficiently in foraging for food [114, 80]. Colonies of this ant average 4000 - 7000 workers. No information could be found about the mortality rate of the workers during foraging. The trail laying behavior of the ant was studied [115]. The foragers do not lay a trail until food is discovered. However, the ants lay trails both to and from a food resource when collecting it, contrary to the idea of mass recruitment where ants only lay trail while returning to the nest. The study also showed that ants showed a predilection for choosing the shorter path, which was tested for a number of mechanisms that the ants could be using. It was concluded that majority of the foragers nearly always chose the shorter path, which was also studied in the context of distance-based changes in foraging and recruitment behavior for this species [96]. The species has also been known to modulate its trail laying strength based on the food source [116]. On another note, pheromones have been shown to take a secondary preference when ants have to decide which path to take, with memory leading the choice determination [80]. The study showed that 82%-100% of the ants choose the path where they had collected food before. They also do not choose a branch based on the proportion of the pheromone present on it. Other work on olfactory conditioning [117] showed that while ants could be trained to forge for food with odor that they knew, untrained ants did not show any preference for food odor. While much has been studied about the food preference of these ants, not a lot is available on the size of food particles and whether this species has any preference for it. Also, majority of the studies distinguish between scouts or recruiters and recruits for this species, indicating a difference corresponding to exploring and exploiting ants. However, none have reported a specific percentage of a natural colony to be scouts. Using all these ideas, the ant can be parameterized in the following way.

1. Number of ants: Black garden ants have average sized colonies, which are still much larger than what is seen in desert ants. similar to the Argentine ants, no statistic could be found regarding their mortality rate or worker turnover rate in their natural environment. Thus, the number of ants is kept the same as Argentine ants – a maximum of 300 ants with 1% removed every few steps and 3 new ones added.

2. Exploring percentage of ants: This is again another statistic which is not available. Different percentages are used to calculate the performance of the colony in the model and an appropriate percentage is selected.

3. Attribute preferences: This species starts using pheromone to and from the nest once a food resource has been found, not until then. Furthermore, the reinforcement of the pheromones is dependent on the quality of the food source that the ants are returning from. Most ants prefer foraging using their memory rather than trails, thus the preference for pheromones can be kept as low as 5% for scouts and a little more at 10% for recruits. Since no study was found which showed that the ants had any preference for food size, it is again kept low at 10% for both exploring as well as exploiting ants. The ants however do have a preference for short path distance when foraging. Thus, the preference for distance is kept at -85% for exploring and -80% for exploiting ants. The negative sign shows that the preference is higher for shorter

distances rather than longer ones.

3.5.4 *Monomorium pharaonis* (Pharaoh ant)

Pharaoh ant is a species commonly known for using different types of pheromones during foraging and recruitment. This species also has a large colony size. The foraging study of these ants [75] showed that they are reliant on pheromone trails for orientation in the foraging environment. In fact, these ants lay trails both while exploring for food as well as while returning to the nest. Some modulation of trail strength has been seen depending on the food quality, but whether the ant is fed or unfed has an effect on it. However, ants were seen to choose the better of the food sources based on the pheromone strength leading to them. The species has also been shown to use repellent pheromones [65] to signal presence of predators or absence of food. Not a lot could be found on food size or food distance preferences, or the division among scouts and recruits for these ants. Thus, the model variables were set in the following way to implement the behavior of this species.

1. Number of ants: This was set the same as the black garden ants and the Argentine ants.

2. Exploring percentage of ants: Since none of the reports on the foraging behavior of this species made a distinction in the behavior of scouts and recruits based on any of the parameters, it is assumed that all the ants are of an exploiting nature.

3. Attribute preferences: A strong preference of 80% is given to pheromones, since studies have shown this species relies on the pheromone trails for orienting themselves. The remaining 20% is equally divided between food size and food distance.

A more thorough analysis of the foraging behavior of this species would be able to give a better estimation of the parameters to be used in this model. However, this species has been considered here primarily due to its use of repellent pheromones.

Table 4.1 summarizes the parameter values set for all the ants discussed in this section.

Ant Species	Pheromone Characteristics	Percentage of Exploring Ants		Preference for Pheromones				Preference for Food Weight				Preference for Path Distance			
				Exploring Ants		Exploiting Ants		Exploring Ants		Exploiting Ants		Exploring Ants		Exploiting Ants	
		Initial	Final	Initial	Final	Initial	Final	Initial	Final	Initial	Final	Initial	Final	Initial	Final
Cataglyphis bicolor (Sahara Desert ant)	No pheromones	80%	80%	0%	0%	0%	0%	50%	20%	97%	75%	50%	80%	3%	25%
Linepithema humile (Argentine ant)	Trails are laid during foraging as well as returning	3%	28%	25%	65%	80%	80%	10%	26%	10%	10%	65%	9%	10%	10%
Lasius niger (black garden ant)	Ants modulate trail strength based on food quality	20%	20%	5%	5%	10%	10%	10%	10%	10%	10%	-85%	-85%	-80%	-80%
Monomorium pharaonis (Pharaoh ant)	Trails are laid during foraging as well as returning. Few ants modulate trail strength based on food quality. Ants use repellent pheromones	100%	100%	80%	80%	80%	80%	10%	10%	10%	10%	10%	10%	10%	10%

Table 3.5: Ant species with their parameter values

3.6 Model Assumptions and Limitations

Throughout the discussion of this chapter, several points have been mentioned where the model diverges from the original data or environment. All the assumptions and limitations are consolidated here.

1. Only documents from 1991 to 2020 available in the ‘Robotics’ Web of Science Category are collected for the dataset.

2. At least 10 citations are required for a keyword to be considered in the keyword co-occurrence network shown in Fig. 3.2.

3. The evolution of the field analyzed in Section 3.1.4 is based on averaged publication years of the keywords, and is assumed to be true within the context of the acquired dataset.

4. Since there is no well-defined exploring speed either for ants or for researchers, it is assumed that an agent, while exploring and foraging for food, takes 1.5 to 2 times the minimum time required to go from one node to another.

5. Similarly, to incorporate the idea that the ants take the shortest route back to the nest, the ants are assumed to travel a shorter distance than the sum of individual path lengths on the way back to the nest, as shown in Fig. 3.9.

6. Parameters such as walking speed, pheromone reinforcement rate, pheromone evaporation rate, maximum foraging population and failure probability are either reported for very few species or are too variable to be given a value. Moreover, changing these parameters has a drastic impact on the model performance. Thus, all these parameters are assumed constant for all the simulation runs.

7. Sahara Desert ant species has been reported to face a mortality rate of 16% and also have a high worker turnover rate. No such information is found for Argentine ant, black garden ant or

Pharaoh ant. To decrease the effect of rapid increase in foraging population due to high worker turnover rate on the efficiency, the mortality rate is assumed constant at 1%.

8. Argentine ants, black garden ants, and Pharaoh ants have been reported to forage in average to massive populations. Again, to reduce the effect of sheer number on foraging efficiency, the population is set constant for all the species.

9. The repellent pheromone in the model removes any attractive pheromone on the selected path.

10. Certain important aspects of ant foraging as well as research are not implemented in the model, one being competition. It is apparent that competition plays a major role in both the processes, and the model can be enhanced to include competition in future, but as of now it is assumed that there is no competition.

4. RESULTS

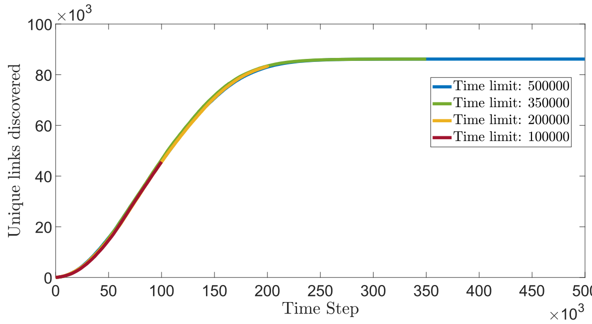
4.1 Preliminary Tests

All the parameters of the model were tested in order to find out their effect on the overall food-collection rate. Each of the following subsections describes the parameter, the results of running the simulation while varying only the specific parameter, an explanation of the results in terms of ant strategies, and what it means in terms of innovation research process.

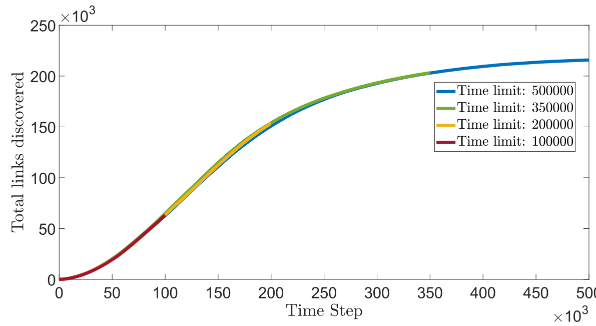
4.1.1 Time Limit

Why is this parameter important? The time limit defines the number of time steps for which the simulation is run. Each time step corresponds to one step movement of all the ants. Fixing a time limit reduces the computational power needed to run the simulation for a long time, while still providing enough information to judge the model performance for a particular scenario. Fig. 4.1 shows the results for varying time limits.

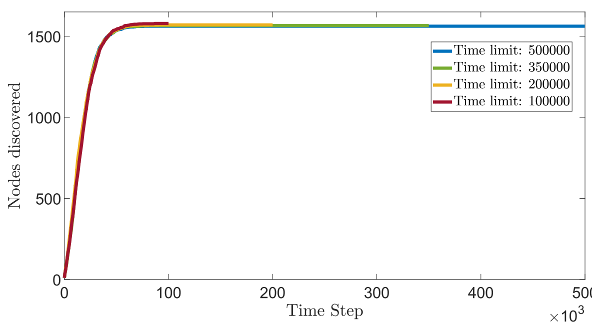
What do the plots imply? The plot lines have been exaggerated to clearly show the overlapping parts. Fig. 4.1a shows the cumulative number of unique links that were discovered at any given time step during the simulation. This counts how many times an ant went from one particular node to another one for the first time ever. As shown in the Table 3.2, there are 86165 such links to be formed. The plot shows that almost all of the links were found by 250000 time steps. Fig. 4.1b shows the cumulative number of total links that were discovered at any given time step during the simulation. This essentially is a count of the total trips made by an ant from the nest to the food and back to the nest. From Table 3.2, there are 219112 total links to be found. The plot shows that about 90% of the links are found by around 300000 time steps. Beyond this point, the food becomes scarce and the rate of finding food starts saturating. Thus, a time limit of 300000 steps is deemed sufficient to understand the performance. In this time, the rate at which ants find food as well as the total food collected indicates their efficiency. It must be noted that the general characteristics of the curves in Fig. 4.1a and Fig. 4.1b are very similar to the ones corresponding



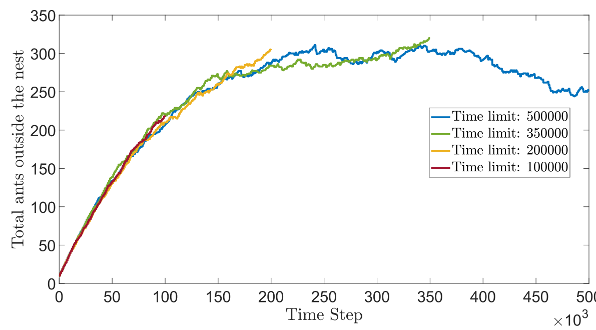
(a) Cumulative number of unique link discoveries over time



(b) Cumulative number of total link discoveries over time



(c) Cumulative number of total node discoveries over time



(d) Total ants outside the nest at any time

Figure 4.1: Testing model performance for four different time limits with same parameters

to the evolution of real data in Fig. 3.5a and Fig. 3.5c respectively. This implies the result of ant behavior is similar to what has been observed in the scientific field. Thus, it validates the use of the model.

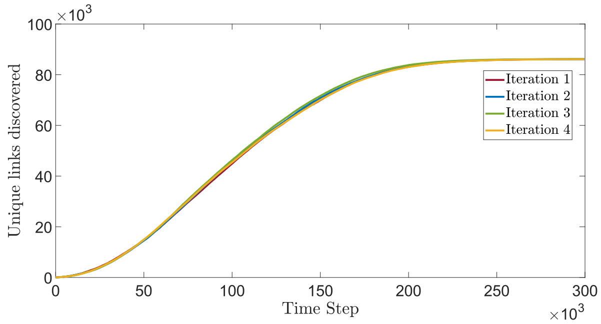
Fig. 4.1c shows the cumulative number of nodes discovered at any given time step. This counts the number of times an ant went from one particular node to another one that had never been visited by any ant ever. There are only 1547 such nodes to be found, as shown in Table 3.2. Almost all of these are discovered by 100000 time steps. Thus, to focus more on the rate of discovery of these nodes, the remaining figures for this statistic will be adjusted for just 100000 time steps. Fig. 4.1d shows the number of ants outside the nest in the field, either searching for food or returning to the nest. The maximum number of ants can be approximately set by using the failure probability and the number of ants added to the foraging population every few time steps. In this case, the failure probability was held at 1% and 3 ants were added every 1000 time steps. Thus, the total number of ants saturates around 300 and keeps oscillating. This ensures that there is a soft upper limit on the number of ants.

How does this translate in the innovation process? In the introduction, innovation was defined in the context of this work as research that either puts forth a completely new concept or idea, or that establishes a novel link between two well-known concepts. Using this definition, the plots corresponding to unique link discoveries and node discoveries correspond to innovations, while the total link discoveries corresponds to all the contribution from the research work, including innovative and repetitive or conservative. Thus, a study of these three plots under different strategies can be helpful in designing a plan of action that improves the rate at which innovations can be found. The dynamics represented in these plots implies that innovations are found rapidly in a newly introduced field, but become difficult to find over time as they become scarce.

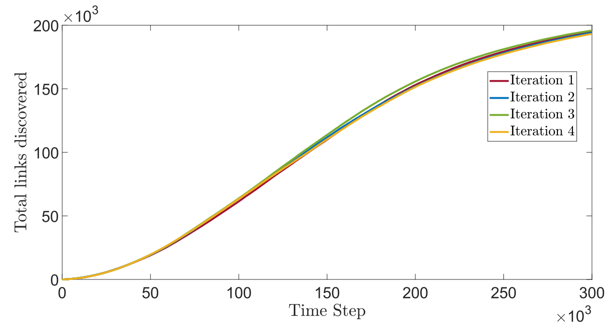
4.1.2 Iterations

Why is this parameter important? In order to test whether different iterations of the simulation resulted in variable performances, four iterations with the same parameter values were run and compared for similarity. This is to ensure that the overall dynamics of the model is not too variable over the entire run, given that there is quite some randomness involved. Fig. 4.2 shows these results.

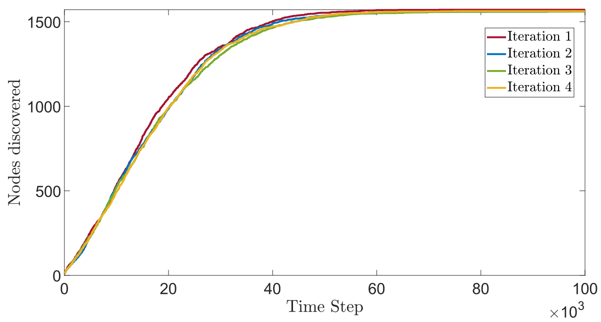
What do the plots imply? The plots in Fig. 4.2a, Fig. 4.2b, Fig. 4.2c show very little variation among all the 4 iterations. Thus, a single iteration should be sufficient to analyze the food discovery rate. Fig. 4.2d shows the number of ants in the field. Since the number of ants added and removed is probabilistic, it is expected that the plots will not be identical. However, they follow a similar trend, and that is a good indication that there is no drastic change in ant population dynamics over the iterations.



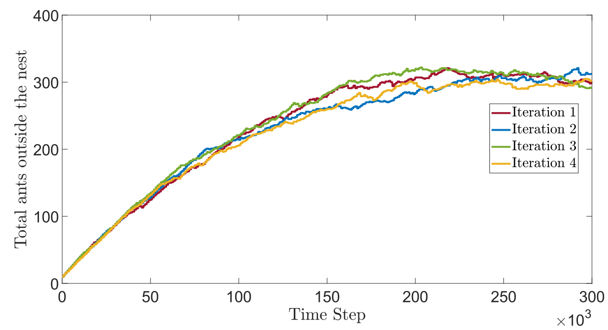
(a) Cumulative number of unique link discoveries over time



(b) Cumulative number of total link discoveries over time



(c) Cumulative number of total node discoveries over time



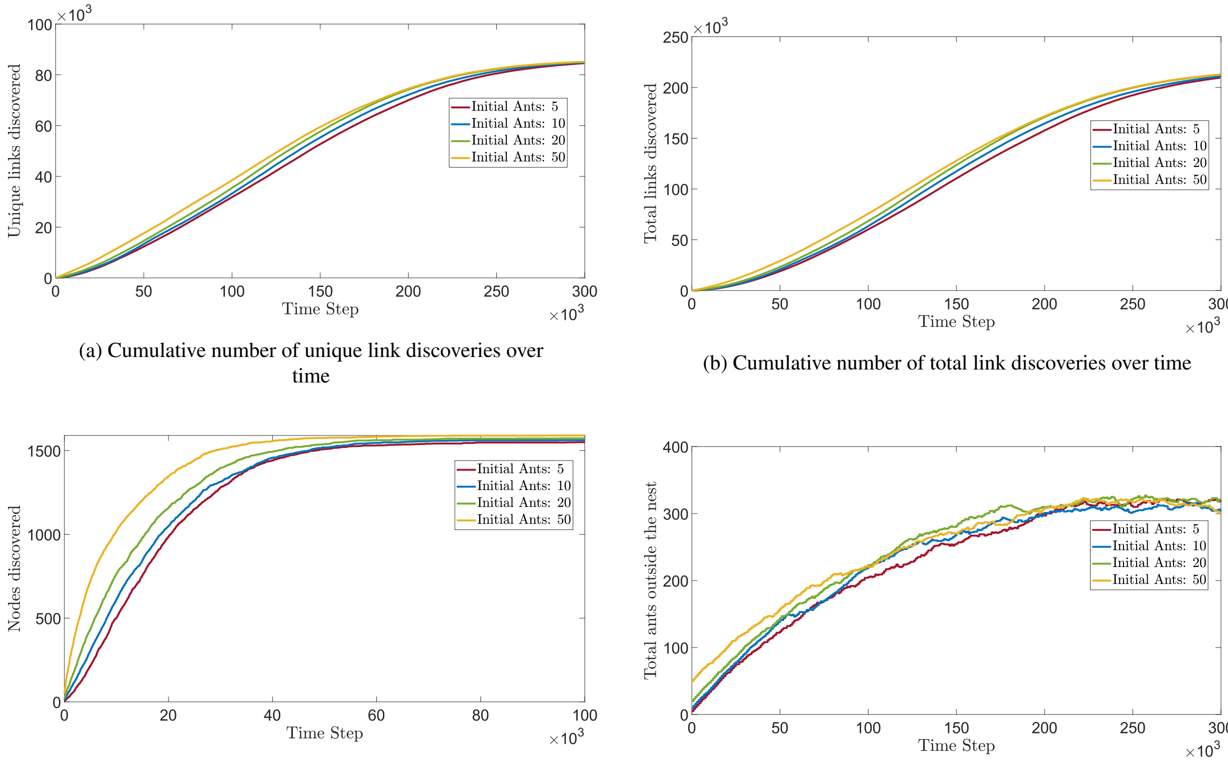
(d) Total ants outside the nest at any time

Figure 4.2: Testing model performance for four iterations with same parameters

4.1.3 Initial Number of Ants

Why is this parameter important? This parameter represents the number of ants that leave the nest foraging for food as the simulation begins. Over time, more ants are added and some of the earlier ones are removed. It is necessary to test whether the initial number of ants affects the dynamics in any way before setting it at a constant value for the rest of the simulation runs. Fig. 4.3 shows the results for four different initial number of ants.

What do the plots imply? All other parameters were kept constant for these four runs. The failure probability was kept constant at 1%. As can be expected, the higher number of initial ants gives an advantage in collecting the food. The plots in Figs. 4.3a, 4.3b, 4.3c show that higher the number of initial ants, faster the discoveries. However, as food becomes scarce, all the plots merge towards the end of the run. The form of the plots is similar for all the runs. Thus, for



(a) Cumulative number of unique link discoveries over time

(b) Cumulative number of total link discoveries over time

(c) Cumulative number of total node discoveries over time

(d) Total ants outside the nest at any time

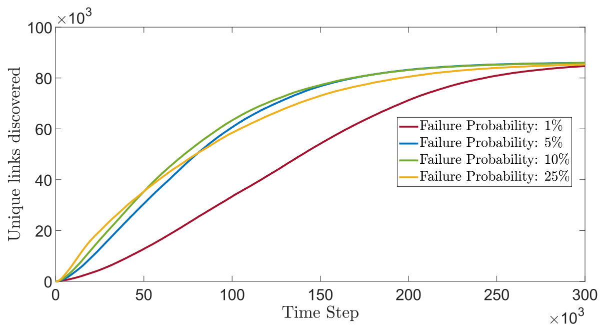
Figure 4.3: Testing model performance for four different initial ant numbers

the remaining simulations, the initial number of ants was fixed at 10. The dynamics of the total foraging population over time is very similar way for all the four runs. While the difference in the initial number of ants can be seen towards the beginning, it becomes unapparent towards the end.

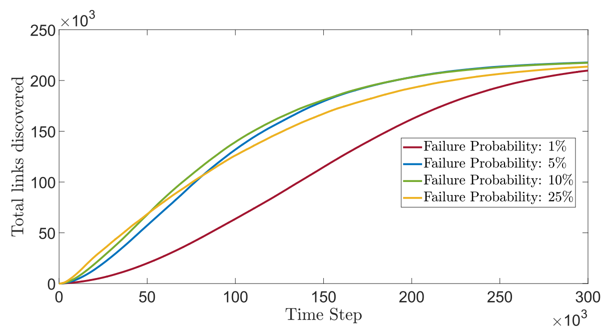
How does this translate in the innovation process? In the innovation process analogy, this implies that if a higher number of researchers venture out into a new field and find novel ideas, the overall discovery rate of innovations is bound to increase. This is because a larger population of innovators can find proportionately large number of innovations, which makes the field opportunistic for further exploration and invites other researchers. The uncertainty of knowing where to go and the risk of failure in a new field is borne by these initial agents. However, once some connections are established, it becomes easier for others to choose a direction for finding new ideas due to an assurance provided by the already discovered novelties.

4.1.4 Ant Failure Probability

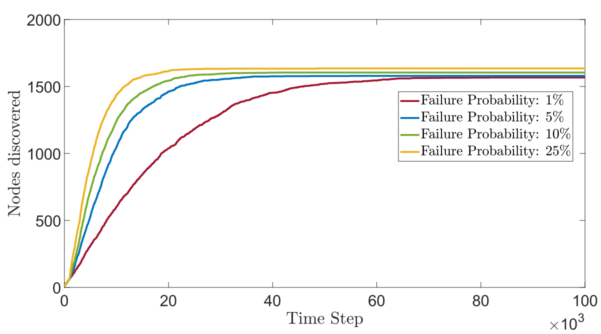
Why is this parameter important? The number of ants added to the foraging population depends on the failure probability of ants. Since the maximum population is to be kept at 300, the number of ants added is set as the failure percentage of 300. E.g., if the failure probability is 10%, then every few time steps, 10% of 300 number of ants are added, which is 30. This ensures that the total foraging population oscillates around the specified number (300 in this case). Fig. 4.4 shows the results for four different failure probabilities of ants.



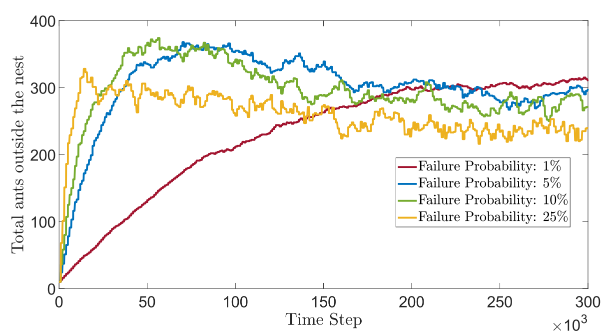
(a) Cumulative number of unique link discoveries over time



(b) Cumulative number of total link discoveries over time



(c) Cumulative number of total node discoveries over time



(d) Total ants outside the nest at any time

Figure 4.4: Testing model performance for four different ant failure probabilities

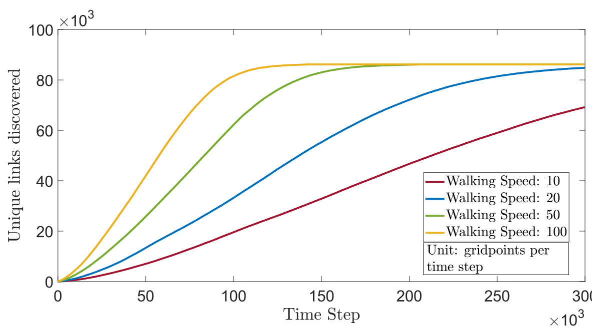
What do the plots imply? The effect of failure probability can be clearly seen on the discovery rates in Figs. 4.4a, 4.4b, 4.4c. Higher the failure probability, higher is the number of ants added.

However, a failure probability of 10% doesn't imply that 10% of the ants will always be removed. Every ant has a 10% chance of failing, and ants with lesser experience and more search times are more susceptible to such failure. Thus, there is always a chance that the number of ants failing might be higher or lower. On the other hand, the number of ants being added is fixed. For failure probabilities of 1%, 5%, 10% and 25%, the number of ants added respectively is 3, 15, 30, 75. Thus, after the first 1000 time steps, at most 1 ant may be removed when failure percentage is 10%, but 30 are immediately added. This shoots up the number of ants drastically, leading to very fast discovery rates. Fig. 4.4d shows the growth in the total number of ants. The rapid spurt can be seen for high failure probabilities. The number shoots to 300, mostly going beyond, and then saturates. In the case of 1% failure probability, the number gradually grows before saturating. Different species of ants have varying levels of mortality and worker turnover rates. Similarly, the number of researchers active in a field can gradually increase or shoot up in a short span of time. The Figs. 3.5a, 3.5c, 3.5e show a rather slow rate of initial discovery, implying that the researcher population gradually increased as the fertility of the field was realized. Clearly, plots corresponding to 1% failure probability show similarities with the real evolution of the field. Hence, in all further runs, the failure probability was set at 1%. This ensures that the foraging ant population increases at a rate expected in the dataset.

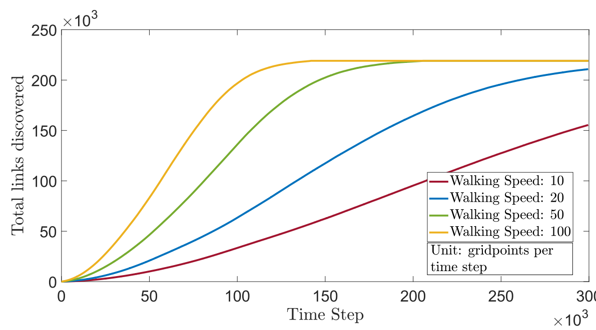
How does this translate in the innovation process? As mentioned earlier, this parameter is useful to control the rate at which the number of ants increases in the field. The number of innovators in a new field too depends on the fertility of the field and the opportunities present. A new field that shows a lot of promise can attract a large population of researchers, whereas a highly complex field with very little novelties possible attracts only a few. As new ideas are established, the fields become more accessible and attract even more innovators. Thus, this parameter ensures that the ants in the field are introduced in a way that best represents the researcher population over time specific to a scientific topic.

4.1.5 Ant Walking Speed

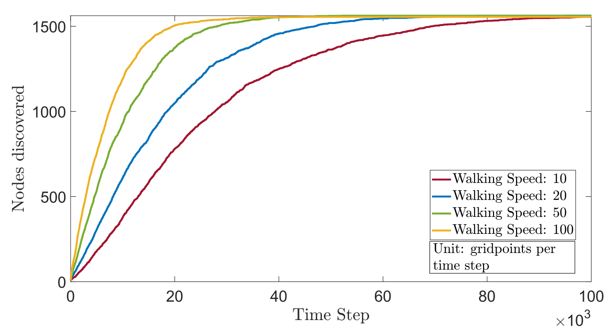
Why is this parameter important? In this case, the walking speed is set as the number of steps taken per time step, where each step corresponds to moving from one gridpoint to another. Since the nodes of the network were assigned integer coordinates, the entire area can be considered to be discrete and made of gridpoints. Thus, ants moving from one node to another essentially move from gridpoint to gridpoint. The number of gridpoints covered in a single time step is defined as the walking speed of the ant. It is necessary to test the effects of varying the walking speed of the ants, since different species walk at different speeds and this speed can be an important factor in the foraging efficiency. Fig. 4.5 shows the results for different walking speeds.



(a) Cumulative number of unique link discoveries over time



(b) Cumulative number of total link discoveries over time



(c) Cumulative number of total node discoveries over time

Figure 4.5: Testing model performance for four different ant walking speeds

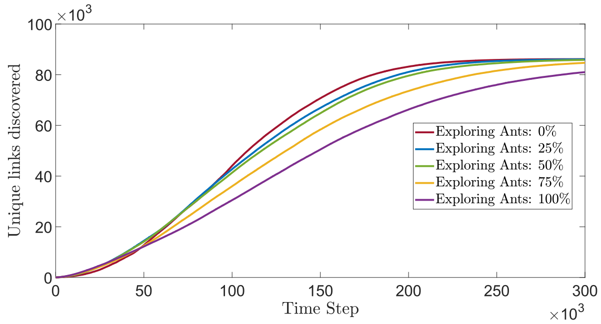
What do the plots imply? As can be expected, ants with a faster walking speed are able to discover nodes and links faster. The plot corresponding to the walking speed of 20 gridpoints per time step in Figs. 4.5a and 4.5b resembles the most to the original data plots in Figs. 3.5a and 3.5c respectively. Thus, this walking speed represents a rate of exploration that is neither too rapid nor too slow as compared to the evolution of the original dataset. Ants with higher walking speeds were able to discover all the links and nodes within half the time. Hence, for further runs, the speed of 10 gridpoints per time step was set.

How does this translate in the innovation process? The results show that ant species which walk faster simply find food faster than slow-walking species. In this case, walking consists of exploring, adjusting to the surroundings, and taking in all the available cues. Innovators who are able to make long cognitive jumps from one concept to another in a short amount of time can perform better exploration with limited resources, thereby increasing the chances of finding an innovation. Such cognitive jumps are possible when the person has an exemplary level of expertise in his/her field. In fact, many species of ants, such as *Cataglyphis bicolor* [107], have been shown to exhibit higher levels of foraging efficiency as their age (or number of foraging trips) increases. Thus, it is helpful in the process of finding an innovation if the researcher has an extensive knowledge of the field and can build connections between ideas rapidly.

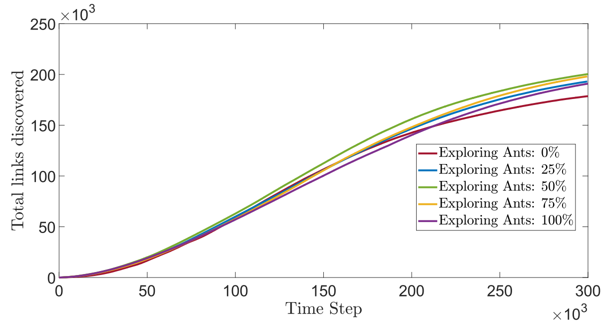
4.1.6 Percentage of Exploring Ants

Why is this parameter important? The number of ants that are exploring affects the performance of the model in the long run. A balance between the exploration and the exploitation parts of the foraging process is important for efficiently collecting all the available food resources. The simulation runs in this subsection test the effects that changing exploring percentage of ants can have on the overall discovery rates. To exemplify the difference in behavior, all exploiting ants were set to prefer only paths with the highest pheromone strength, while all exploring ants were set to choose a path based either on food weight or food distance but not pheromone strength. Fig. 4.6 shows the results.

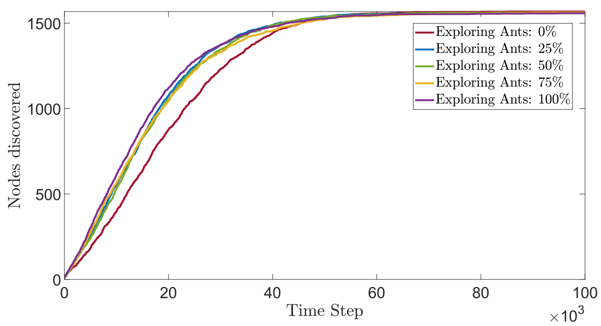
What do the plots imply? Figs. 4.6e and 4.6f show the differences in the number of exploring and exploiting ants based on their percentages. For 50%, there are mostly equal number of exploit-



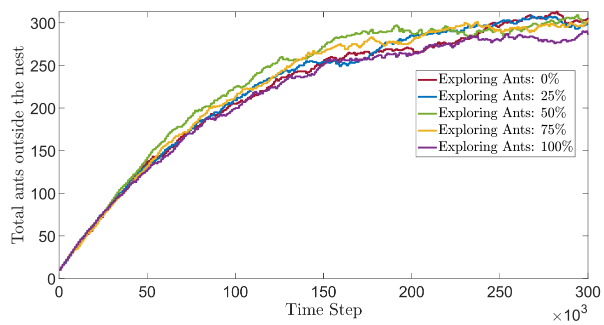
(a) Cumulative number of unique link discoveries over time



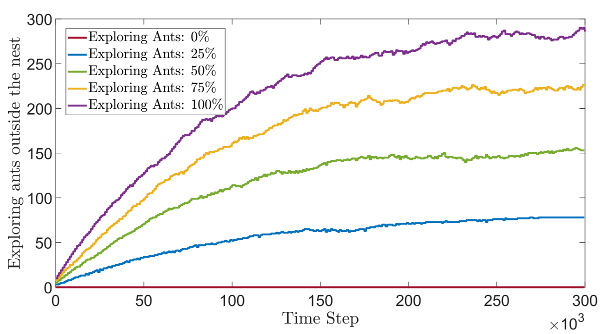
(b) Cumulative number of total link discoveries over time



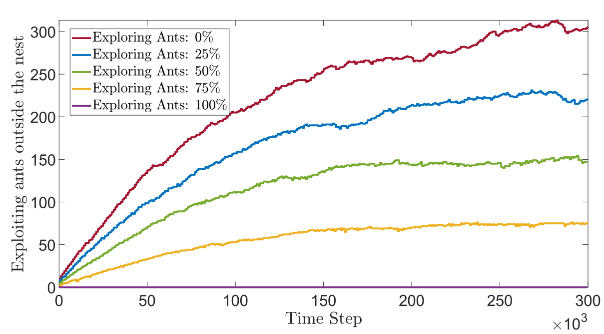
(c) Cumulative number of total node discoveries over time



(d) Total ants outside the nest at any time



(e) Total exploring ants outside the nest at any time



(f) Total exploiting ants outside the nest at any time

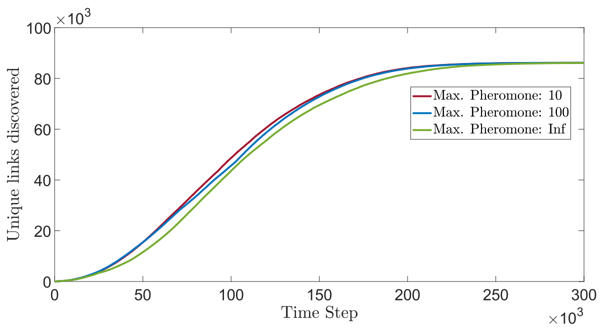
Figure 4.6: Testing model performance for four different exploring percentages of ants

ing and exploring ants in the area. As this percentage increases, exploring ants increase exploiting ants decrease. Fig. 4.6d shows the total number of ants, which does not show any dependence on the exploring percentage. The ants in all the five scenarios, that is ant exploring percentages of 0%, 25%, 50%, 75%, 100%, perform very similarly in total link discoveries, as shown in Fig. 4.6b. However, on closer observation, interesting dynamics can be perceived. The curve corresponding to 50% is always at top from start to finish, implying a balanced number of exploring and exploiting ants outperform other proportions. Curves corresponding to 25% and 75% compete with each other but end up being similar by the end of the simulation. The curve corresponding to 0% starts by being the most suboptimal, reaches some optimum levels towards the middle, but again falls towards the end. Similarly, the curve corresponding to 100% starts almost at the top, but falls by the end. The curves in Fig. 4.6a can be explained by the attribute preferences. As exploiting ants only choose based on pheromone strengths, there is some amount exploration that occurs when pheromones haven't yet been laid, or when the food has just finished. This results in ants finding unique links over time. On the other hand, exploring ants only prefer food weight and short distance. These choices lead most of the exploring ants towards the same food sources, which they do not collect if there is pheromone already present around it. As mentioned in Section 3.2.2, exploring ants only collect food if it has been discovered for the first time or if it isn't being collected by other ants, which is denoted by the absence of pheromones. A lot of time is wasted in exploration, leading to slow rate of unique link discovery. Thus, skewing the balance between exploration and exploitation towards either side changes the overall dynamics of the model.

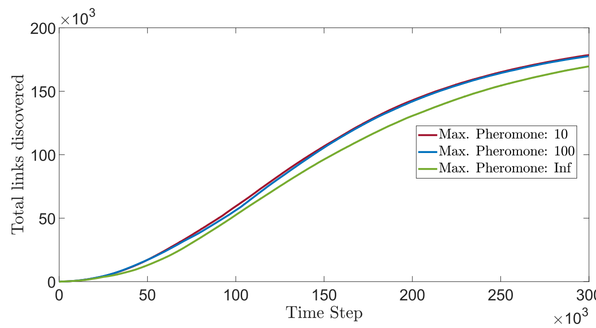
How does this translate in the innovation process? The distinction between exploration and exploitation relates to two kinds of researchers – conservative and innovative. Traditional, conservative, risk-averse researchers are those who pursue finding ideas among well-established concepts, often leading to repetitive contributions that aren't novel. On the other hand, innovative researchers endeavor to find novel ideas by using known knowledge but exploring in obscure areas of the field. Only conservative researchers cannot sustain a fast innovation discovery rate over a long time since they aren't actively taking the risks required. Also, only innovative researchers cannot succeed as well since their strategies are risky, and without sufficient available knowledge, their probability of failure increases. Thus, a balance between these two extreme strategies is necessary for the optimal rate of discovering novelties.

4.1.7 Maximum Pheromone Strength

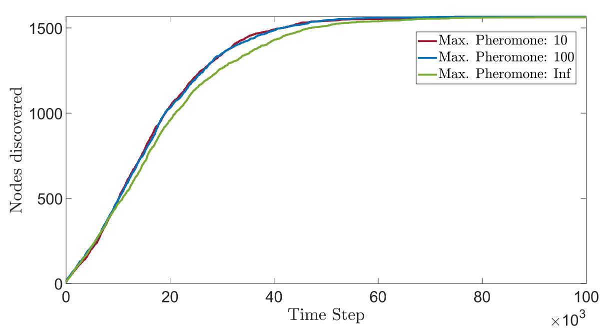
Why is this parameter important? As pheromone trails are reinforced, the amount of pheromone can increase by a lot if it doesn't evaporate fast enough. This is especially true for the edges starting from the nest, or along a node which is being traveled on by numerous ants. Whether this has any effect on the overall discovery rates needs to be tested. Thus, the following simulations were run for different upper limits on the pheromone strength. A limit of 'Inf' implies there was no limit applied. Fig. 4.7 shows the results. A unitless number is used for specifying the pheromone strength.



(a) Cumulative number of unique link discoveries over time



(b) Cumulative number of total link discoveries over time



(c) Cumulative number of total node discoveries over time

Figure 4.7: Testing model performance for four different maximum pheromone strengths

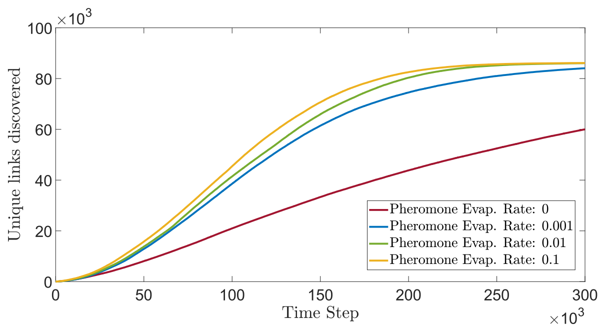
What do the plots imply? In all the three plots in Fig. 4.7, the simulation corresponding to no upper limit on maximum pheromone strength performs poorly as compared to the simulations where an upper limit was set. This is intuitive since the evaporation rate is not fast enough to completely wipe out the pheromone on a trail that has been frequently traversed and reinforced by ants while collecting an abundant food source. Thus, ants dependent on pheromones are led to sites where no food is present, thereby leading to suboptimal efficiency. Not a lot of difference was observed in the performance related to pheromone limits of 10 and 100 units. Thus, an upper limit of 100 was set for future simulations.

How does this translate in the innovation process? The pheromone trails correspond to publications and other science communication media in the innovation analogy. The strength of a pheromone trail signifies the number of ants that have traveled between two nodes. Similarly, the number of publications in which two keywords co-occur implies the number of researchers who found the connection between the two keywords important enough to find a new sort of relation between them. Since documents do not evaporate like pheromones do, it is necessary to consider the dynamics of the number of publications specific to a particular connection in order to understand the current relevance of that idea. A connection between two keywords mentioned by a 100 documents within the span of one year signifies a more contemporary relevance as compared to a connection mentioned by a 100 documents over ten years. Publications form a part of existing knowledge and can inspire a direction of research. Thus, it is necessary to note the dynamics of the number of publications rather than just the total number before deciding how opportunistic a field might be. An upper limit on pheromones (along with evaporation) strives to incorporate this idea into the model, since ants have no way of knowing the dynamics of pheromone strength on a particular trail to judge its usefulness.

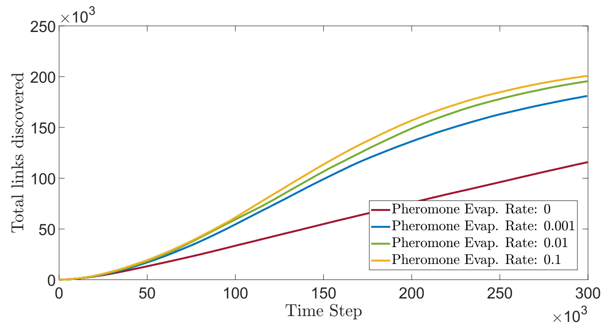
4.1.8 Pheromone Evaporation Rate

Why is this parameter important? To ensure that pheromones do not stay forever and orient ants towards sites where food has exhausted, it is necessary to reduce their strength unless the ants reinforce them. Different evaporation rates can be set for this purpose. The evaporation of pheromones depends a lot on the chemical composition of the pheromones laid by the ant species as well as the environmental conditions. Thus, there is no single value that can work in all the cases.

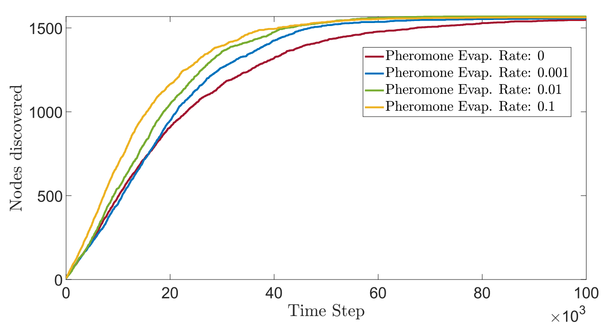
Nevertheless, to understand the effects of evaporation rate on the food discovery, the following simulations were run for different rates. Fig. 4.8 shows the results of these runs. The unit of evaporation rate is just per time step, since pheromone strength is considered to be unitless. Fig. 4.8 shows the results.



(a) Cumulative number of unique link discoveries over time



(b) Cumulative number of total link discoveries over time



(c) Cumulative number of total node discoveries over time

Figure 4.8: Testing model performance for four different maximum pheromone strengths

What do the plots imply? As expected, ants perform very poorly when there is no evaporation of pheromones. This can be explained using a similar reasoning as in the case of no upper limit for pheromone strength. Ants are oriented towards depleted food sources, thus pushing the time to find existing food. Contrary to this, the performance is relatively similar when a certain evaporation rate is given. An evaporation rate of 0.01 units per time step was set for future simulations.

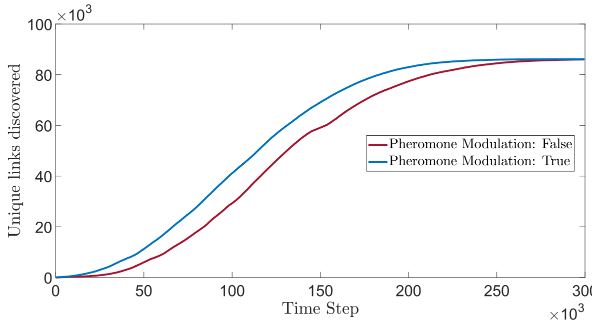
How does this translate in the innovation process? As with maximum pheromone strength, evaporation rate of pheromones helps ants pursue paths that end at available food sources rather than going towards depleted ones. The significance of the number of publications decreases over the years to represent an obsolescence of the field. Various scientometric analyses are used to calculate the relevance or impact of a publication or a journal. Google Scholar uses h-index [118] which calculates the impact of an author or a journal by finding the maximum h number of publications published by the author or the journal, such that each of those publications has been cited at least h times. Journal Citation Reports calculates an impact factor for every journal [119] by using the number of citations received during the year for articles published in the preceding two years. None of these analyses use the number of publications as the sole representation of the impact of the field. The idea of pheromone evaporation is applicable to understanding the most opportunistic contemporary ideas in a field and thus forms an important parameter for the model.

4.1.9 Pheromone Modulation

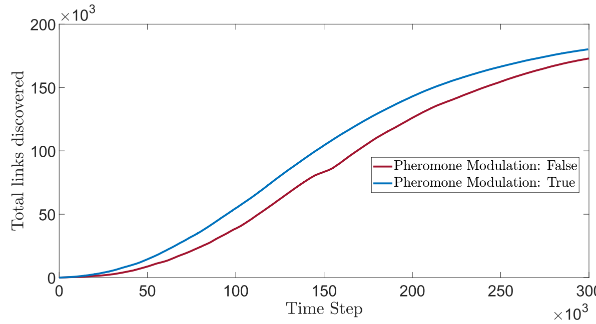
Why is this parameter important? Some ants [75] have been shown to modulate pheromone strength based on the food quality that they return from. The parameter on pheromone modulation incorporates this idea into the model. The parameter can be set to 'true' or 'false'. If set to 'true', the ants lay a trail whose strength is proportional to the food quality. Otherwise, it is a constant equal to the pheromone reinforcement rate. Fig. 4.9 shows the results for the use of pheromone modulation. To amplify the effects, all the ants were set to be exploiting with a preference only for pheromone strength.

What do the plots imply? All the plots in Fig. 4.9 clearly show an improved performance during the use of pheromone modulation. As ants control the pheromone strength based on food quality and since all the ants choose where to go based on pheromone strengths, this strategy ensures that majority of the foraging population goes where abundant food sources are available.

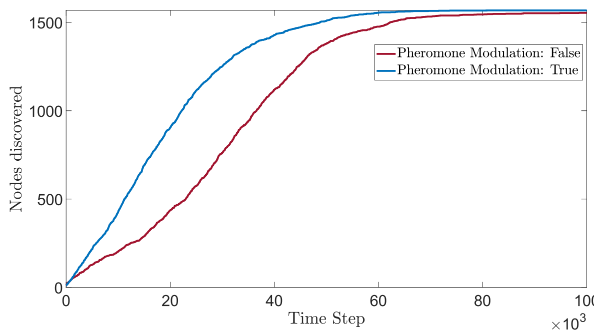
How does this translate in the innovation process? In the innovation analogy, this parameter doesn't have a direct correlation. However, just like the discovery of a rich food source loses its value if the trail is weak enough to evaporate before other ants can find it, the discovery of a novel idea is void if not enough members of the scientific community find out about it. Often it



(a) Cumulative number of unique link discoveries over time



(b) Cumulative number of total link discoveries over time



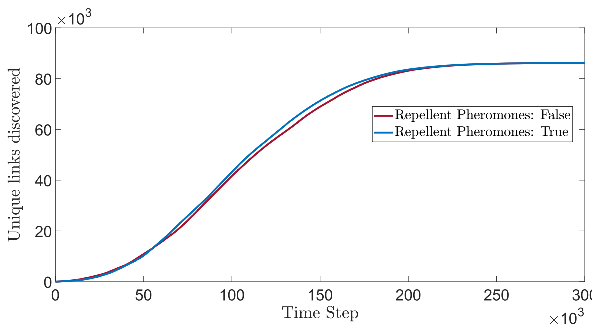
(c) Cumulative number of total node discoveries over time

Figure 4.9: Testing whether pheromone modulation affects model performance

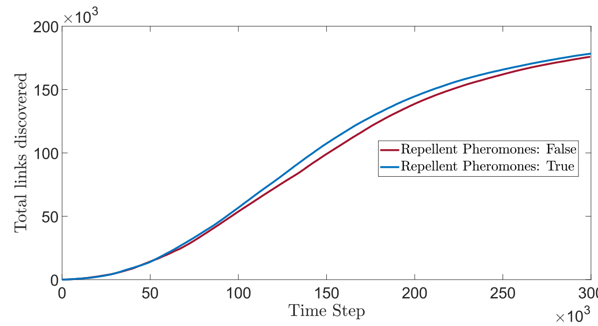
takes longer for novel research to get recognized and have a major impact [5]. Thus, the parameter incorporates the essence of proper communication about a novel scientific discovery and its impact over the community. Ideally, if innovators ensured that enough people knew and understood the novelty of the work they had done, it'd have a far greater impact and would attract more researchers.

4.1.10 Repellent Pheromones

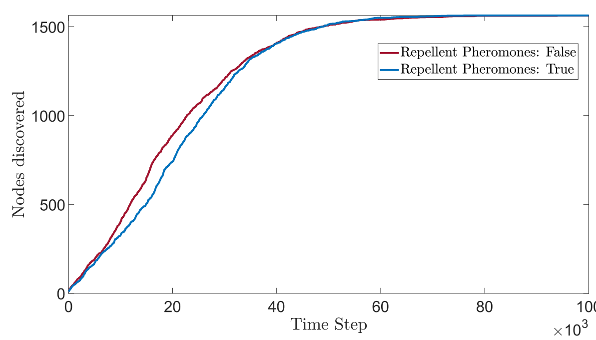
Why is this parameter important? As discussed before, some ants [65] have also been shown to use repellent pheromones to indicate absence of food or presence of predators. These are left at trail bifurcations. This parameter incorporates the idea of repellent pheromones into the model. By setting it as 'true', ants check the food sites while returning to the nest and if one of the sites has no food, they will leave a repellent pheromone at the next trail bifurcation that they come across. The repellent pheromones removes all the pheromone along the specific trail, so that ants with a preference for pheromone strength do not choose that trail and focus on other paths. Fig. 4.10 shows the results for testing the effect of this parameter. Just like for pheromone modulation, all ants were set to be exploiting with a preference for only strong pheromones.



(a) Cumulative number of unique link discoveries over time



(b) Cumulative number of total link discoveries over time



(c) Cumulative number of total node discoveries over time

Figure 4.10: Testing whether repellent pheromones affects model performance

What do the plots imply? While there is not a lot of difference between discovery rates for the two scenarios, the case where repellent pheromones are used shows a slightly higher discovery rate in Fig. 4.10a and Fig. 4.10b, and falls a little short in Fig. 4.10c. Using repellent pheromones ensures that all the food within a particular distance is collected before moving further, which is why the node discovery rate becomes a little slow. However, it ensures that majority of the foraging population focuses on the food that can be collected before going to search for new sites, which leads to higher link discovery rates.

How does this translate in the innovation process? Repellent pheromones in the innovation analogy act as publications that warn researchers about the risk of pursuing a particular path. As mentioned in the introduction, recent innovation models [20] have shown that publishing failures can improve the innovation finding rate, since it saves the efforts of pursuing a path that leads to a dead-end repetitively.

4.1.11 Attribute preferences

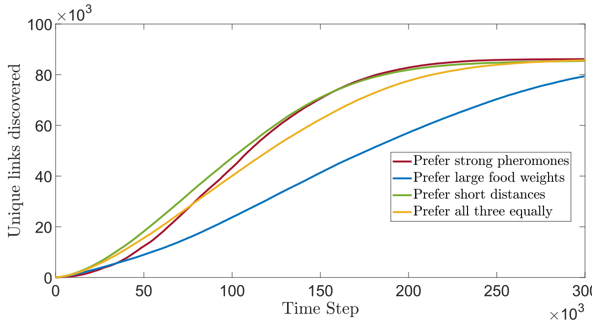
Why is this parameter important? Finally, it is necessary to recognize the effects that attribute preferences can have on food discovery rates. These preferences control the paths that ants choose while foraging, and can lead to variable dynamics over a long time. In the following simulations, different sets of preferences were tested in order to find whether the resulting performance is intuitively correct. All the ants were considered to be exploring, so that the attribute preferences could be applied to the entire population, thereby making their effects prominent. The parameter sets include

1. Only preference for strong pheromones: Here, the ants only consider the pheromone strengths on different paths while choosing the path to travel, and disregard the food weight or distance. Higher the pheromone strength on the path, higher is the probability that the ant will choose it.

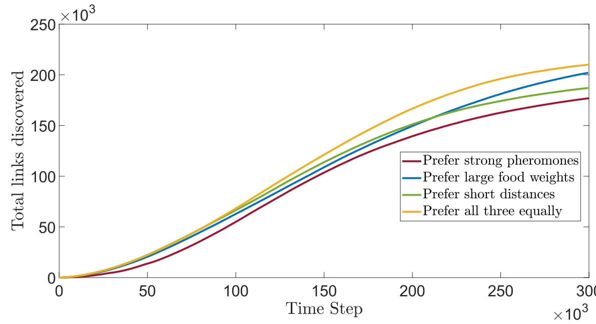
2. Only preference for large food morsels: In this case, ants only consider the food weights at the end of the paths while choosing the path to follow. This implies ants are oriented by the scent of the food and not the pheromones. Larger the food weight at the end of a path, higher is the probability that the ant will choose that path.

3. Only preference for short distances: The ants in this set always prefer choosing the path with the shortest length, that is, they go to the food that is at the least distance from their current location. Smaller the path distance, higher is the probability that the ant will choose it.

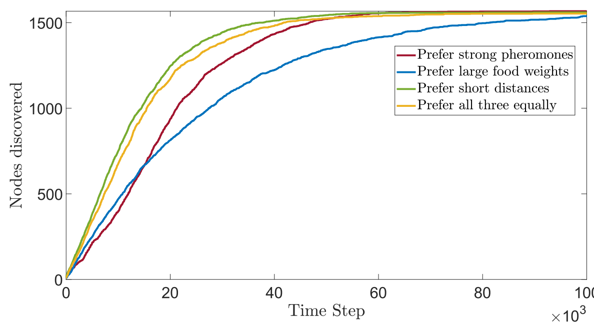
4. Equal preference for strong pheromones, large food morsels, short distances: Here, the ants prefer strong pheromones, large food weights, and short distances equally.



(a) Cumulative number of unique link discoveries over time



(b) Cumulative number of total link discoveries over time



(c) Cumulative number of total node discoveries over time

Figure 4.11: Testing model performance for four different attribute preferences

While most ants use a combination of these attributes while choosing the direction to forage in, these simulation runs provide an understanding of the effect of these attribute preferences on the overall model performance. Fig. 4.11 shows the results of these runs.

What do the plots imply? As expected, different dynamics occur from setting these attribute preferences. Most of the conclusions from these plots are logically reasonable. The preference for large food weights results in an adequate performance when looking at total links discovered in Fig. 4.11b, since it implies most of the ants are focused on exploitation. However, it also means ants aren't doing any exploration, which is why the performance is poor in unique link discoveries and node discoveries in Figs. 4.11a and 4.11c respectively. An equal preference to all three attributes, as one might expect, seems to perform just well enough in all three cases. These plots show that different performance levels can be expected by setting various attribute preferences.

How does this translate in the innovation process? One can make sense of these attributes in the case of innovation processes by considering the aspects that a researcher looks for before deciding the direction of research to pursue. Some study the available literature, find a problem that needs a solution, follow the literature to understand the context, and solve the problem. They are guided by the literature generated by others in the field. Some researchers are inspired by the promise of opportunities presented by an active area of study, and decide to pursue towards it in the hopes of discovering novel ideas. Other researchers focus on new knowledge that can be created with little risk and without needing to go too deep into obscure areas. Some might challenge themselves to take the required risk and venture into little-studied areas in the hopes of discovering something novel. Thus, how an innovator chooses his/her project is dependent, among many things, on his/her attitude towards the process of research. This gives every researcher an individuality. By altering these parameters, the model can incorporate these individualities into the ants. The simulation shows a balance of these attributes among all the researchers can provide the best rate of innovative discoveries.

Ant Species	Pheromone Characteristics	Percentage of Exploring Ants		Preference for Pheromones				Preference for Food Weight				Preference for Path Distance			
				Exploring Ants		Exploiting Ants		Exploring Ants		Exploiting Ants		Exploring Ants		Exploiting Ants	
		Initial	**Final**	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**
Cataglyphis bicolor (Sahara Desert ant)	No pheromones	80%	80%	0%	0%	0%	0%	50%	20%	97%	75%	50%	80%	3%	25%
Linepithema humile (Argentine ant)	Trails are laid during foraging as well as returning	3%	28%	25%	65%	80%	80%	10%	26%	10%	10%	65%	9%	10%	10%
Lasius niger (black garden ant)	Ants modulate trail strength based on food quality	20%	20%	5%	5%	10%	10%	10%	10%	10%	10%	-85%	-85%	-80%	-80%
Monomorium pharaonis (Pharaoh ant)	Trails are laid during foraging as well as returning. Few ants modulate trail strength based on food quality. Ants use repellent pheromones	100%	100%	80%	80%	80%	80%	10%	10%	10%	10%	10%	10%	10%	10%

Table 4.1: Ant species with their parameter values

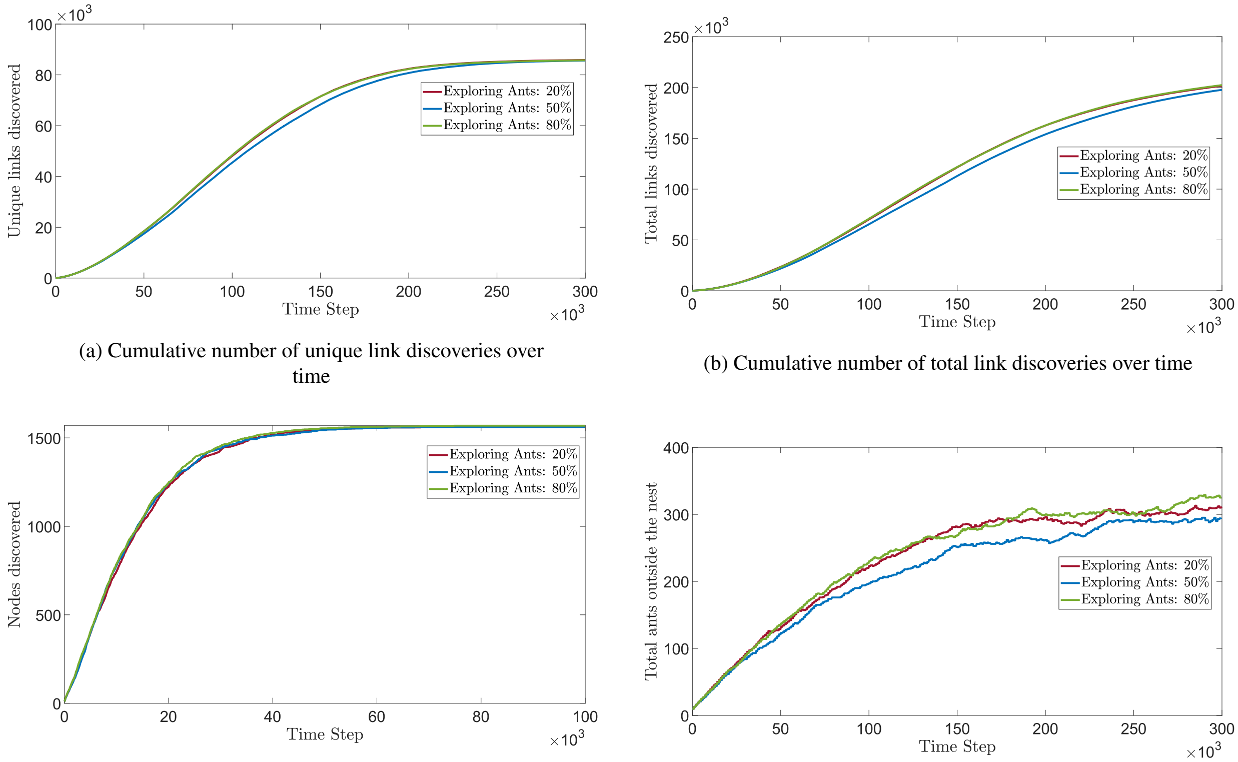
4.2 Parameterized Ant Species Tests

Different ant species described in Section 3.5 were tested for performance. All the constant parameters were defined as mentioned in the previous sections, and the variables were set to best mimic the real foraging strategies that the species implement. These variables are shown in Table 4.1 (reproduced in this section). The idea is to evaluate the efficiency of these different strategies, and whether they can be beneficial when translated into actions that researchers can take to boost their innovation discovery rates.

Initially, the different exploration percentages were tested for black garden ants, since no relevant data could be found on it. Fig. 4.12 shows these results. Since there is not a drastic difference between the behavior of exploring and exploiting ants, the plots show that the performance is very similar in all discovery rates. The scenario with exploring percentage 50% seems to perform a bit suboptimally, but it can be reasoned through Fig. 4.12d. The second scenario ended up having a lower foraging population over time than the other two, which in turn lowered the food discovery rates by a little. For comparison with other ants, an exploring percentage of 20% was chosen. Thereafter, the four species of ants discussed in Section 3.5 were tested and compared with each other.

What do the plots imply? Fig. 4.13 shows these results. The ants show mostly variable dynamics in all discovery rates. The following points attempt to reason why a particular species performs the way it does judging by its relevant parameters.

Figs. 4.13a and 4.13b show that Sahara Desert ants perform inefficiently as compared to the other three species. The Desert ants show a distribution of 80% exploring and 20% exploiting ants. Exploring ants for this species highly prefer going long distances, while exploiting ants highly prefer going for large food weights, so that they can collect it repeatedly till it gets depleted. Due to the high variance in the exploration-exploitation distribution, the efficiency reduces. Majority of the foraging population attempts to go for long distances, which in the absence of pheromones requires a longer time for finding food. A further analysis shows that the average number of discoveries for exploiting ants is 644 while that for exploring ants is 431. Thus, majority of the



(a) Cumulative number of unique link discoveries over time

(b) Cumulative number of total link discoveries over time

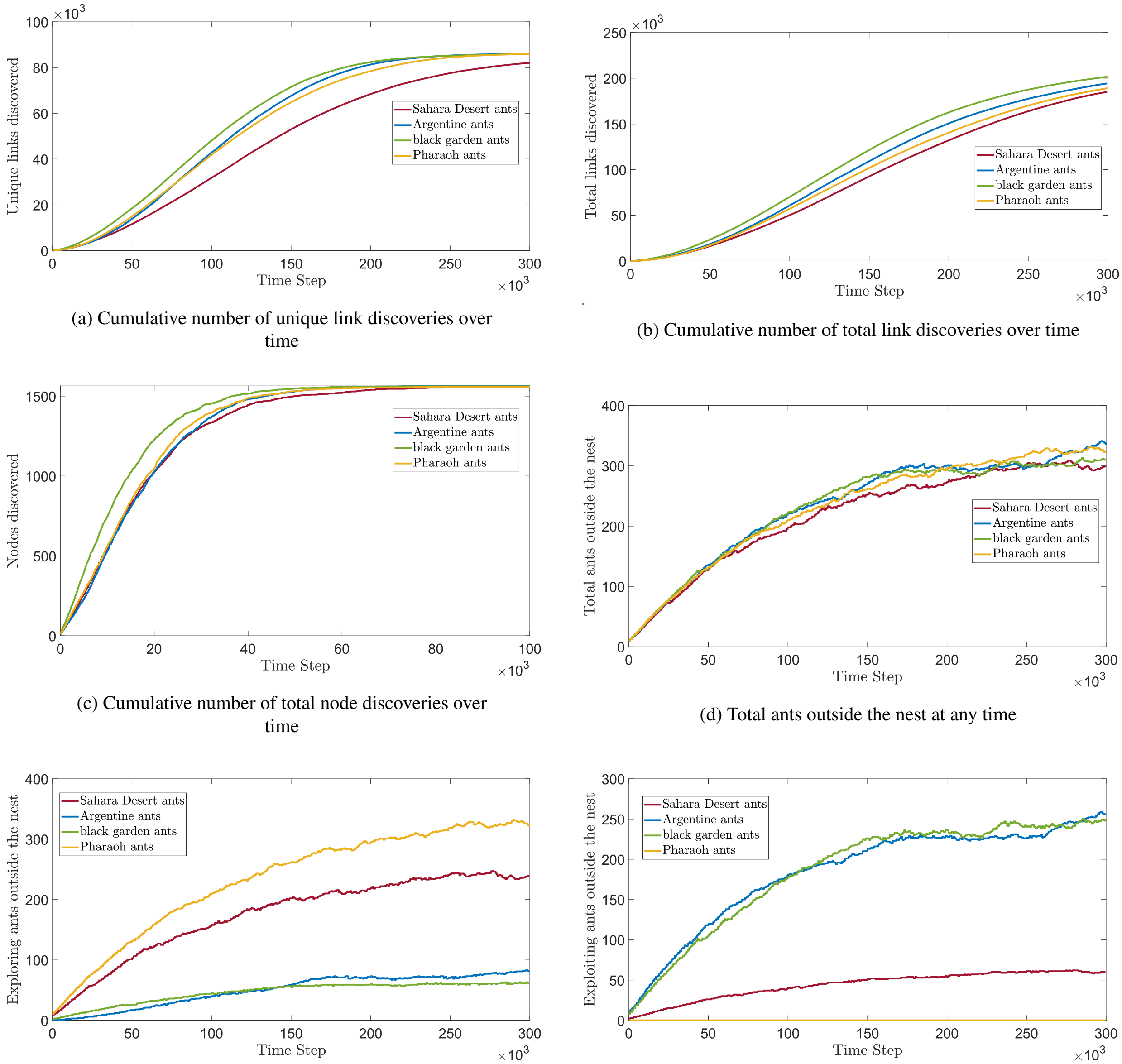
(c) Cumulative number of total node discoveries over time

(d) Total ants outside the nest at any time

Figure 4.12: Performance of black garden ants for three different exploring percentages of ants

population going for long distances does not work out in the long run.

The performance of the other three species, namely Argentine ants, black garden ants and Pharaoh ants, is mostly similar for unique link discoveries as evident from Fig. 4.13a. Pharaoh ants are initially efficient as compared to Argentine ants, but over time, Argentine ants turn out to be faster at discovering unique links. The initial preferences are quite similar for Argentine ants and Pharaoh ants, except that Pharaoh ants use repellent pheromones. This can show an increased efficiency in earlier part of the simulation for Pharaoh ants. However, as the foraging population increases, the exploring percentage of Argentine ants increases and the preference for long distance among these exploring ants decreases. This implies that over time, ants explore significantly early on and then focus on exploiting the found food sources. Such behavior leads to a better performance in both unique link discoveries (Fig. 4.13a) as well as total link discoveries



(a) Cumulative number of unique link discoveries over time

(b) Cumulative number of total link discoveries over time

(c) Cumulative number of total node discoveries over time

(d) Total ants outside the nest at any time

(e) Total exploring ants outside the nest at any time

(f) Total exploiting ants outside the nest at any time

Figure 4.13: Testing model performance for four different ant species

(Fig. 4.13b). Pharaoh ants mostly focus on following pheromone trails with little preference for food weight or distance, thereby reducing the amount of exploration.

Black garden ants outperform all other species in all the discovery rates. They are efficient in finding the unique links, in exploiting the found connections, as well as in discovering new nodes. A primary reason for this is the preference for shorter distances rather than longer ones. The species has little consideration for pheromones or food weights, but extensively prefers choosing paths where food can be reached at quickly. This allows ants to find food at a quicker rate and collect it. An analysis shows that the average number of discoveries (food collected) per ant is 455 for Pharaoh ants, 473 for Argentine ants, and 500 for black garden ants. It must be noted that only preferring short distances does not imply an increase in performance, which can be evidenced from the fact that black garden ants also outperform the scenario where complete preference is given to short distances. In this case, most of the ants going for food at shorter distances effectively reduces the amount of time spent in finding the food, thereby increasing the amount of food found in a given time.

Fig. 4.13b shows that the total number of ants in the field remains more or less similar. Thus, the results are not skewed by the foraging population of a particular species being consistently lower than that of others. Fig. 4.13e also shows an increasing number of exploring ants for the Argentine ant species.

How does this translate in the innovation process? The following points connect the performance of these ant species to the innovation process and attempt to explain what kind of a research strategy every species suggests.

The ant species show that Desert ant performs suboptimally as compared to other mass-recruiting ants. In the case where every innovator has to start from scratch, it is apparent that innovations would take time to be found. It is easier to use available knowledge as guidance in deciding which areas to focus. The efficiency of all mass-recruiting species over the solitary foraging Desert ants shows that use of scientific communication media such as journal articles, conferences, seminars, etc., are valuable in the innovation process as they establish existing knowledge and thus reduce the amount of searching needed to find a new idea.

Pharaoh ants are quite efficient in finding unique links due to their use of repellent pheromones. Repellent pheromones warn other ants of the depletion of food source in the region and directs them towards other sources. Recent models of innovation have shown that publishing failures in the form of dead-ends, inconclusive results or lack of required technology can be useful in letting other researchers focus on other areas. The resources used in exploring such an area are saved by encouraging others to either solve problems that led to the specific failure or pursue other neighboring areas of interest. Thus, the model suggests publishing failures of any kind to improve novelty rate in science.

Argentine ants are an invasive species and are quite efficient foragers, as is evident from Fig. 4.13. While their massive foraging population plays an important role in their success as a species, they also change foraging preferences according to ant density in the area. Improving the number of exploring ants over time while decreasing the preference for food distance as the density of ants increases results in a consistent level of exploration as well as exploitation. It suggests that as a particular field gets perceived as opportunistic and allows numerous discoveries, the proportion of researchers that look for new unique connections must increase as compared to those who invest themselves in conservative repetitive research.

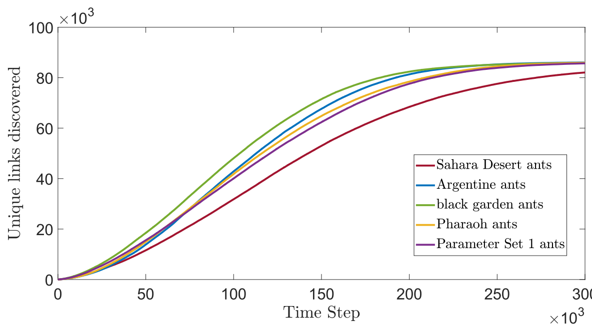
Black Garden ants seem to outperform all the other species according to Fig. 4.13a and Fig. 4.13b. Their strategies involves extensively finding food at short distances and modulating trail strength based on the quality of the food source found. This suggests that searching for innovations that are at a small cognitive distance from existing knowledge results in a fast rate of discovery, which is true. However, it does not imply that is the best strategy to follow. Pursuing an area that is very close to well-established knowledge and repetitively exploiting it will lead to higher number of innovations, but it does not lead to impactful innovations. Impactful innovations requires some amount of risk in obscure areas of study. Argentine ants perform long distance foraging initially and later limit themselves as the ant density increases, and their efficiency is just a little shy of that of black garden ants.

4.3 Mixed Parameter Sets

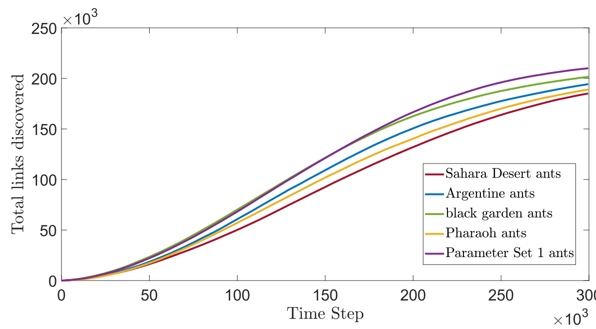
Different ant species show different dynamics owing to differences in the level of exploration and exploitation, and using tactics like modulating pheromone strength and laying repellent pheromones. Based on these ideas, a parameter set can be constructed that mixes the best parts of all the species, uses conclusions drawn from preliminary tests, and attempts to outperform all the scenarios performed so far. This section details the construction of a few of such parameter sets, the results obtained from them, and their implications in terms of innovation research process.

4.3.1 Parameter Set 1

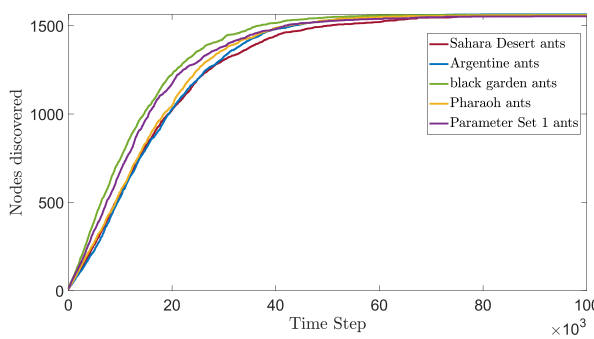
How is the parameter set described? This set consists the same parameters as one of the sets used in Section 1.1.11. The exploring percentage of the ants is kept at 100%, and all attributes are given equal preferences, that is, 33% for strong pheromones, 33% for large food weights and 33% for short distances. Additionally, modulation of pheromones and repellent pheromones both are set to 'true'.



(a) Cumulative number of unique link discoveries over time



(b) Cumulative number of total link discoveries over time



(c) Cumulative number of total node discoveries over time

Figure 4.14: Comparing performance of ant species with that of parameter set 1

What do the plots imply? One of the interesting observations, as seen in Fig. 4.14, is that the parameter set performs better in total link discoveries than all the parameterized ant species tested. The rate of unique link discoveries is slower, but the rate of node discoveries is considerably high. Having a balanced preference across all the attributes makes the ants discover more links in total than any of the tested species. Due to the equal preferences, the ants are randomly distributed within the network. Different food sources are found and collected. The unique link discovery is slow due to exploring ants happening upon the same food repetitively till it is collected. However, because almost all paths are chosen with equal likelihood, the probability that an ant stumbles upon a food source for any given path choice becomes high.

How does this translate in the innovation process? This parameter set shows that preferring research that does not travel a long distance from the current knowledge and is conservative, while leading to a high number of discoveries, is not the most optimal method. Since every researcher has his/her own method of deciding a strategy to pursue a particular problem, this set encompasses all the researchers with their individualities working together. The overall growth of a field can be optimal if all the researchers with their preferences work together. This ensures that while a lot of exploration happens, there is also exploitation happening at a decent rate. The more exploitation there is of a particular field, the easier it is to understand complex topics within that field, thereby inspiring riskier strategies to take farther steps in the hopes of discovering impactful innovations.

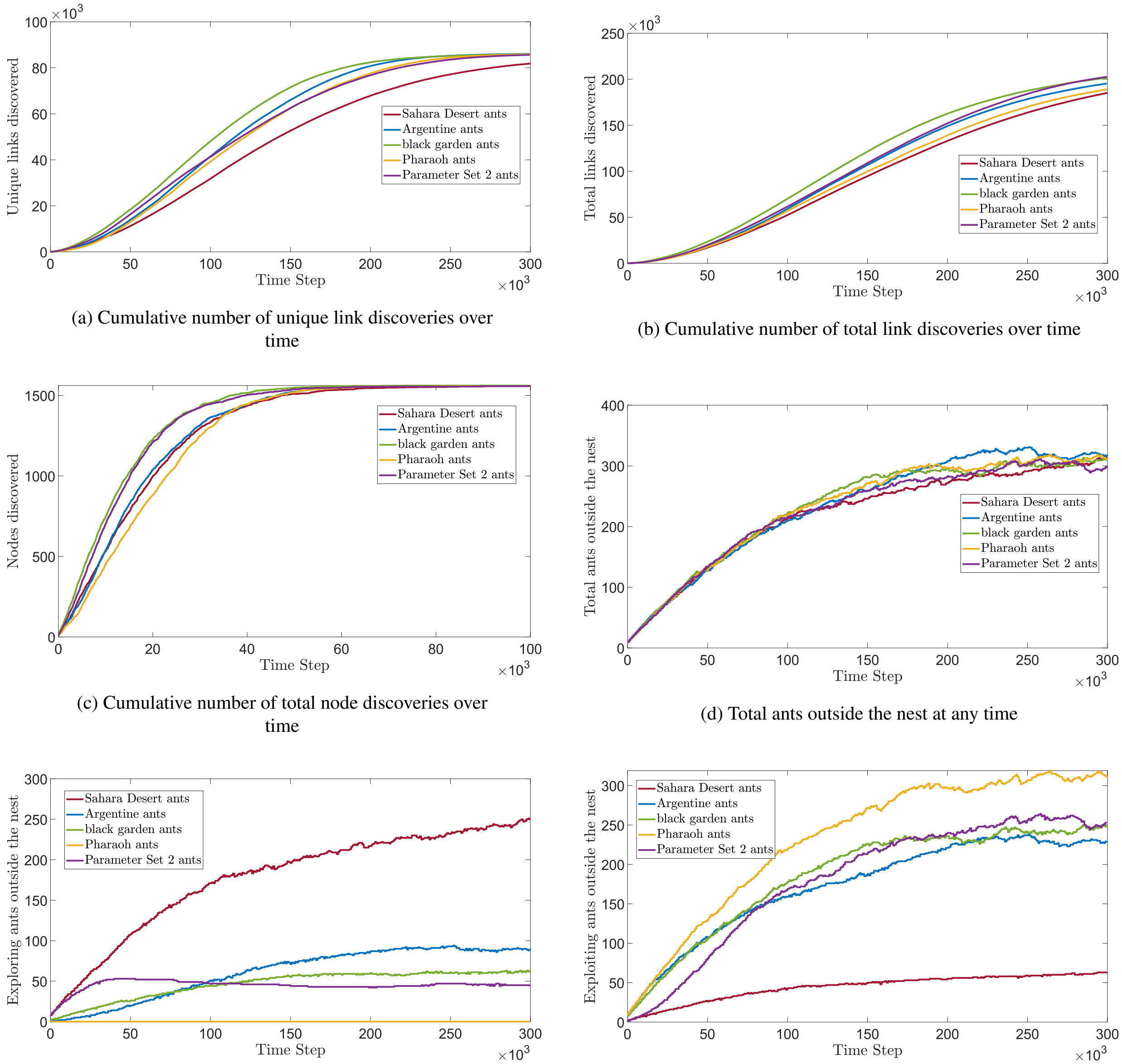
4.3.2 Parameter Set 2

How is the parameter set described? In this set, the percentage of exploring ants is kept at 85%, but as the number of ants in the field increases, this percentage drops to 15%. Initially, the exploring ants prefer finding the food that is at longer distances but do not care about existence of pheromone trails or food weight. Thus, the initial attribute preferences for exploring ants are set at 15% for pheromones, 10% for food weight, and 75% for long distances. On the other hand, exploiting ants initially only focus on finding smaller food morsels. The attribute preferences for exploiting ants are set at 15% for pheromones, 80% for small food weights, and 5% for long distances. This implies initially, exploring ants are primarily searching for food at long distances while exploiting ants are mostly collecting small food morsels as fast as they can.

As time goes on and the foraging population increases, the preferences change. The exploring ants now prefer following the path with the shortest distance or the strongest pheromone each by 25%, but prefer the path with the largest food weight at the end by 50%. On the other hand, exploiting ants make the most of existing trails by preferring trails with the highest pheromone strength by 50%. The prefer large food weight by only 20%, and short distances by 30%. Thus, towards the end, exploring ants are focused on finding large food morsels so that the massive exploiting ant population can collect them. Due to extensive trails established in the later stage of the simulation, exploiting ants focus on following these trails and reinforcing them when they find food at the end of it, while not orienting themselves too much based on food weight or path distance.

What do the plots imply? Fig. 4.15e and Fig. 4.15f show the number of exploring ants decreasing over time along with a proportionate increase in the number of exploiting ants. Fig. 4.15c shows that the node discovery rate for this parameter set is as efficient as the black garden ants. However, Fig. 4.15a and Fig. 4.15b show some interesting dynamics. As a lot of the exploring happens in the earlier stage of the simulation, the unique link discovery is highly efficient initially. Thus, a number of food morsels are discovered in the beginning part. Fig. 4.15b shows an increase in efficiency in the total link discovery during the end part of the simulation. This is caused by the majority of the exploiting ants collecting already discovered food sources. It can be seen that the rate in Fig. 4.15b goes slightly beyond that of black garden ants.

How does this translate in the innovation process? The exploration of a new field at the initial stage allows discovery of numerous novel ideas. These attract more researchers. It is essential to understand these ideas before trying to explore further, since otherwise the failure risk can be high. The parameter set suggests a decrease in the amount of exploration over time should balance out the total network growth in the long run. Further, since the researchers in this parameter set prefer going long distances initially, success implies availability of innovative concepts that are cognitively far from each other. The conservative research helps fill gaps arising from such risky research. An exploratory attitude in the beginning with a preference for cognitively distant con-



(a) Cumulative number of unique link discoveries over time

(b) Cumulative number of total link discoveries over time

(c) Cumulative number of total node discoveries over time

(d) Total ants outside the nest at any time

(e) Total exploring ants outside the nest at any time

(f) Total exploiting ants outside the nest at any time

Figure 4.15: Comparing performance of ant species with that of parameter set 2

cepts, and an exploiting attitude towards the end with a preference for simple concepts can provide a sufficiently optimal knowledge growth rate.

5. SUMMARY, CONCLUSION

5.1 Model Summary

This project began with the study of innovations, the need to find strategies that lead to efficient novelty discoveries, and the methods of finding these strategies. Various innovation models, their construction, their successful findings and their limitations were discussed. Many important theories such as Optimal Foraging Theory, Information Foraging Theory, the concept of Team of Teams, and Emergence provided a fundamental basis for constructing a bio-inspired innovation model. Commonly used concepts such as network theory and bibliometric mapping quickly laid the groundwork for designing a dataset on which to run the model. After an extensive study of foraging strategies of ants and building a comprehensive analogy with the innovation process, a rudimentary model was created.

The model uses different functions corresponding to leaving nest, finding food, returning to nest and depositing the food. Every simulation run configures ants to have specific preferences while searching for food. Moreover, ants can either explore new food sources or exploit available food sources. The simulations attempt to increase efficiency by achieving a balance among all these parameters. Different species are studied and parameterized for the model in Section 4.2, while multiple parameters from all species are combined and tested in 4.3. The performance of every simulation provides a new insight into what works and what doesn't. Furthermore, every test consists of a commentary on how the result translates into the innovation process analogy. Suggestive ideas extracted from these results that can help boost creative research include extensive knowledge of the field, a balance between conservative and exploratory research, proper communication of found innovation, publishing failures, etc. Section 4.3 also shows that a balanced preference among influence by existing publications, influence by fertility of a field and influence by cognitive distance results in an efficient discovery rate over time. Changing the model variables leads to drastically different strategies and makes it possible to discover new tactics implementable in finding innovations.

5.2 Contribution of this Work

This project set out to find new ways of discovering strategies for finding innovations efficiently. Using bio-inspired design, a model implementing foraging strategies of ants was constructed. This model can use various datasets built from bibliometric data of a particular field. The dataset creates an environment in which ants can forage for food. Employing strategies of different species can provide various levels of efficiency. Finally, a comprehensive analogy between foraging ants and innovating researchers allows translating the strategies of ants into actions that researchers can take to further boost the innovation rate. This model encompasses the importance of innovation communication and dissemination among the scientific community as well as the time and efforts required in this process. It also addresses a level of individuality that researchers possess corresponding to their nature of study, which is overlooked in many innovation models due to simplification. Finally, it provides an extensive database of strategies that have been studied among numerous ant species and have been observed to be efficient in nature. The model encourages design of innovation strategies based on the success of such species in nature, which trumps design of optimized parameter sets in often oversimplified model conditions. Furthermore, provided information about similar evolution of two scientific fields, a historic one and a current one, the model can effectively suggest the strategies that might develop the new scientific field in an efficient manner. Thus, it hopes to add a valuable method of studying innovation processes, provides a fresh insight from the point of view of bio-inspiration, and propagates effective solutions towards improving novelty creation in science.

5.3 Conclusion

Sustaining the growth of science and technology is what pushes the society forward. Innovations form the foundation of science, and are the primary reason that knowledge progresses rapidly. In the unimaginable expanse of the scientific landscape that currently exists, it becomes necessary to study and choose which problems, while solvable with allocated resources, can contribute towards the development of the landscape. The model proposed in this work endeavors to address the difficulty in choosing a strategy for finding innovations efficiently. It depends on proven strategies observed in biological species such as ants and translates them into actions that researchers can take in order to help them find new ideas optimally. The work proposes development of such bio-inspired design techniques by proving their aptness in the search for ideas that can improve the rate of discovering novelties, and endeavors to fill in the gaps left by other innovation models.

6. FUTURE WORK

A number of different versions of the model were tried and tested before one was finalized as the best one. This chapter discusses the other ways in which the idea of a bio-inspired innovation model can be used for finding efficient strategies.

6.1 Co-citation Network Model

An earlier version of the model consists of using a co-citation network and ants foraging in a gridspace. It is detailed in the following subsections.

6.1.1 Using Co-citation Network

An additional way in which the evolution of a scientific field can be studied is through co-citation networks. This type of network has references as the nodes and the articles in the dataset as the links. Thus, an article citing two references becomes a link between two nodes. More the number of articles citing two references together, greater the link strength between the nodes corresponding to those two references. Thus, the network measures the relationship between the cited documents. A co-citation analysis maps the landscape of the field as it is perceived by the researchers at present [120]. This is helpful since the different clusters within the field become apparent and the profitability of choosing to seek innovation in a particular direction can be gauged. The development of a co-citation network depicts the efforts taken by researchers to link two or more influential knowledge nodes, thereby establishing a cognitive connection among those concepts. This representation of innovations has been discussed in the context of biomedicine [3, 20], where the nodes are chemicals and the links are connections among these chemicals established by the publications. Innovation can be defined as a new link being formed between any two nodes, whereas repetitive and traditional research comprises of adding to the link strength of an existing link. Co-citation network thus is helpful for tracking the amount of new research that was conducted as opposed to the research that did not add anything novel.

To build this network, the option of creating a map using bibliographic data was selected in

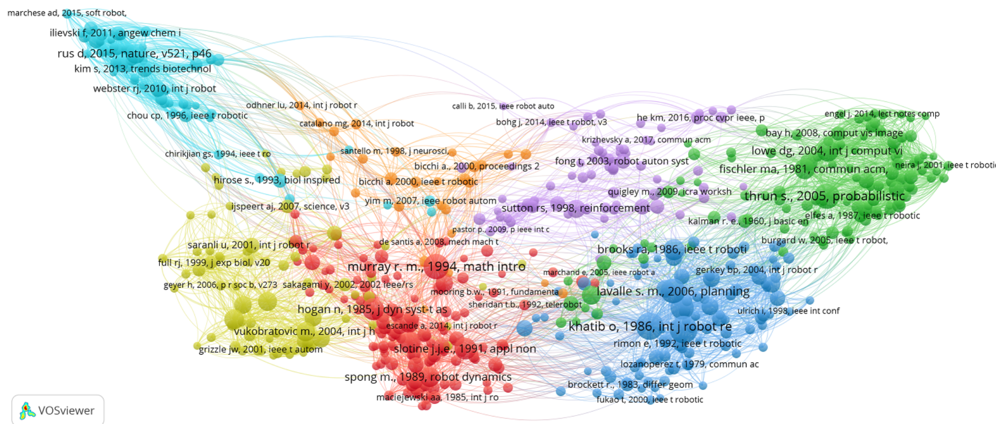


Figure 6.1: Co-citation network generated in VOSViewer

VOSViewer. After selection of files, the co-citation type of analysis with cited references as unit of analysis was chosen. A full counting method was used in order to avoid getting fractional link strengths. The minimum number of citations that a cited reference should have in order to be considered as a node in the network was set at 50. This vaguely represents the influence of the articles selected to be knowledge nodes. The network in Fig. 6.1 consists of 591 nodes, 34050 unique links of which 5000 have been shown, 115313 total links including repeated ones, and 7 clusters constructed by the clustering algorithm of the software. Table 6.1 also shows the first five nodes in the network, similar to Table 3.1 for the keyword co-occurrence network. Here, the reference names correspond to different references cited by the documents in the dataset. As with keywords, two or more documents can cite two references together. Thus, every reference has unique number of links and total number of links. The column on citations refers to the total number of documents in the dataset that have cited the reference.

6.1.2 Evolution of the Network

The dynamics in the case of co-citations differs from that of keywords. A primary reason is that unlike keywords, the number of references does not deplete. Every article that puts forth novel and impactful ideas in the field becomes a reference for the articles that exploit these ideas. As a result, the reference from row 1 of Table 6.1 corresponds to an article from 1998, which falls in the

Node ID	Reference Name	X	Y	Cluster	Unique Links	Total Links	Citations
1210	alami r, 1998, int j robot res, v17, p315, doi 10.1177/027836499801700402	1.0692	-0.1813	1	10	12	5
1352	alexander rm, 1990, int j robot res, v9, p53, doi 10.1177/027836499000900205	-0.0658	-0.0803	6	12	15	5
1620	an ch, 1988, model based control	-0.353	0.0349	2	48	71	35
1719	anderson rj, 1989, ieee t automat contr, v34, p494, doi 10.1109/9.24201	-0.3113	-0.2193	2	11	14	9
2035	arimoto s, 1984, j robotic syst, v1, p123, doi 10.1002/rob.4620010203	-0.4936	-0.487	2	30	53	30

Table 6.1: First five references in the cocitation network of data from 1991-2020

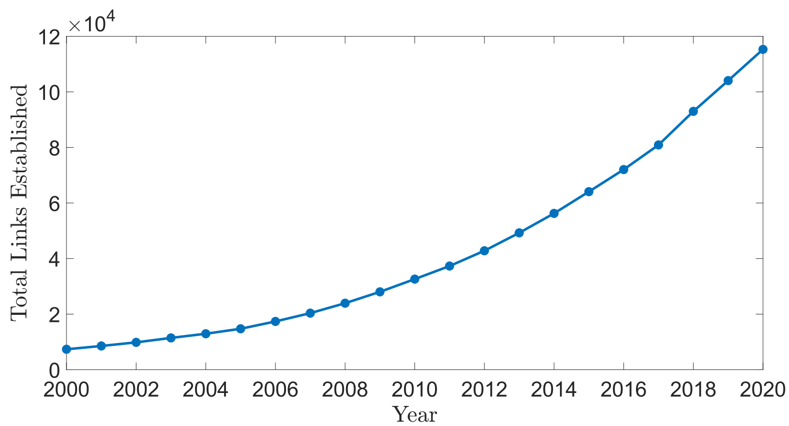


Figure 6.2: Number of connections established in co-citation network from 2000 to 2020

year range of the dataset (1991-2020).

The nodes and links established prior to 2000 were considered to be existing knowledge, and the evolution of the network based on new knowledge was considered from 2000 to 2020. This is shown in Fig. 6.2. About 7500 links were already established by the year 2000. Compared to Fig. 3.5c, it is apparent that the curve doesn't saturate over time. This is simply because the number of references that can be cited by new research increases at a higher rate than the number of keywords. In this case, while a keyword signifies a novel concept or idea, a reference itself could be about anything. Since only references which garner more than 50 citations are considered as nodes in the network, the impact of a publication is defined by the number of citations. Thus, an innovation in this case is research which either results in such an impactful innovation, or which joins the ideas propagated by two major references.

6.1.3 Model Description

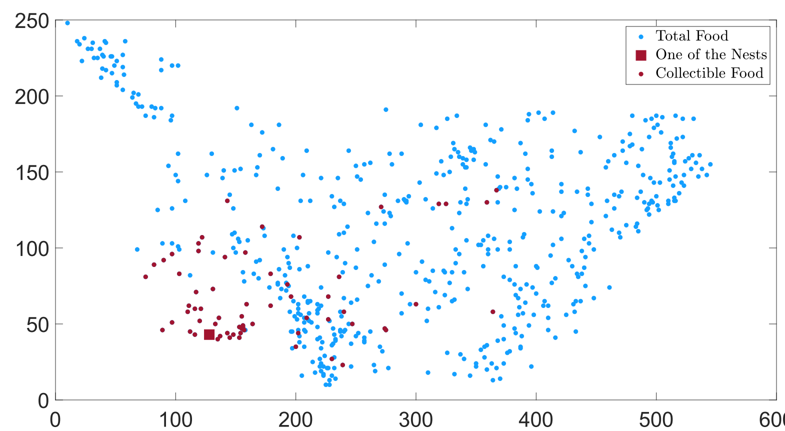


Figure 6.3: Specificity of food particles collectible by ants from a particular nest

In this model, the co-citation network formed the environment, or the food network, with which the ants interacted. Instead of creating a graph, a gridspace was created in which the network is initialized. Every node has a co-ordinate. A distinction is made between food nodes and nest nodes. Nest node forms a starting point for the ants. This is chosen from one of the first references which appear chronologically in the dataset, since it corresponds to the contemporary knowledge when the simulation begins. Further, the nests are chosen so that the corresponding nodes have no connections among them, since ants cannot build connections between two nests. Ants leave the nest nodes, wander according to a correlated random walk [121], and find food at any of the food nodes. Just as in the case of keyword network, which food the ant can collect is dependent on the node it started from. This is because the model only allows building connections that are present in the original network. Fig. 6.3 shows an example of the food particles that an ant can collect if it starts from the shown nest. The initialization of multiple nests, along with addition of new nests as all the collectible food corresponding to existing nests depletes, ensures that the ants build all the connections over time. A snapshot of the simulation is shown in Fig. 6.4. The ants start from different nests, collect food, and return to their nests while laying trails. As new nests are added, new ants start from them and collect food. The clusters provided by VOSViewer act as territories in this case. Nests within a particular cluster are assumed to belong to the same colony. Thus, when one nest becomes obsolete, the ants move on to another nest in the same cluster. The model automatically ensures that new nests are added to each cluster so that ants aren't left with no food to collect.

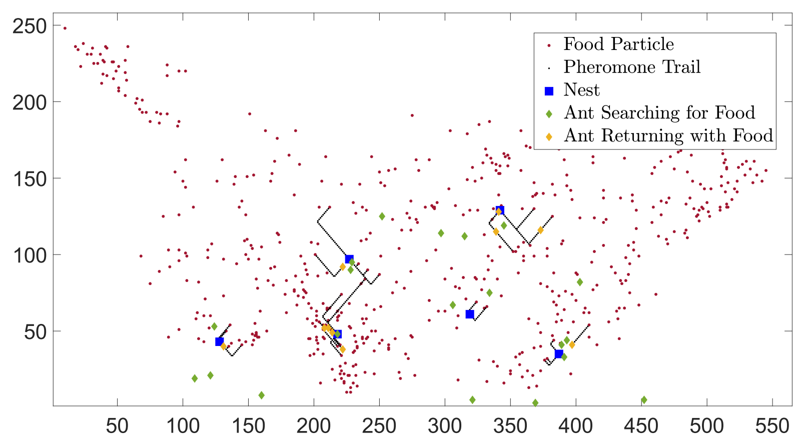


Figure 6.4: Snapshot of simulation showing nest, ants, food and trails

6.1.4 Parameters and Results

The different parameters available in this case include exploring percentage of ants, number of ants and walking parameters. Since ants are free to roam in a gridspace, their walking can be controlled using a correlated random walk, with the help of a standard deviation value. Higher this value, more tortuous and wounded the ant's path. Lower values of standard deviation result in ants talking straighter and longer paths. Further, different parameters related to pheromones such as reinforcement strength, evaporation rate, etc. are available as well.

The results of the model match with the data only within a particular time span and not from start to end. This is because in the case of the simulation, as food becomes scarce, ants have to spend longer to find it, thereby saturating the food-discovery curve. This is not the case with references in the dataset. The number of impactful publications that inspire new research grow in number with time, thus leading to the curve in Fig. 6.2. As a result, the simulation is run till the end but the statistics are checked for a sliced period of time. This ensures that the dynamics matches the real life, where food is available in plenty and is still being collected.

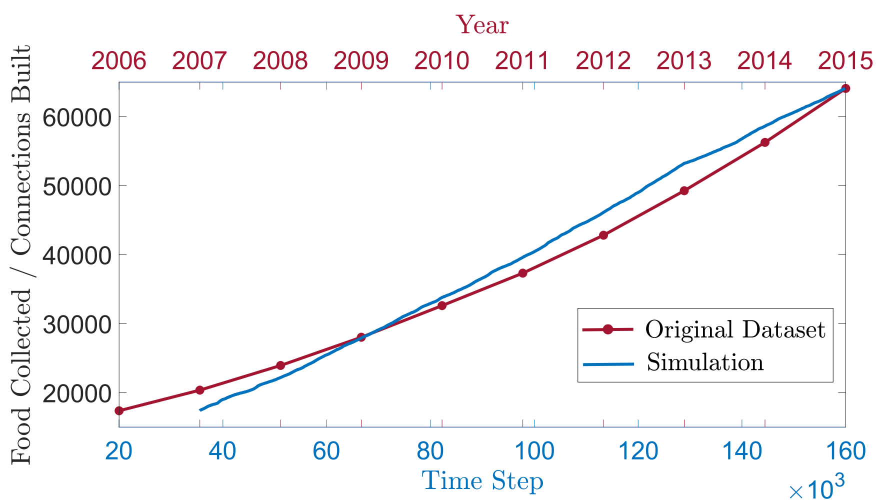


Figure 6.5: Result from co-citation network model

Fig. 6.5 shows the simulation result from one of the runs compared to the evolution of the original dataset. The result is sliced according to the food particles collected (connections established) by 2006 in the dataset, and later by 2015. The two curves show a good correlation, implying that the dynamics of food collection matches the dynamics of innovation discovery, provided none of the factors such as scarcity of food play a role.

Due to local correlation between the simulation and data, the model wasn't pursued further and the keyword network was considered. However, the realistic way of controlling ant movement using correlated random walks, the ability to observe the areas where ants spend the most time exploring, and the importance of co-citation networks in understanding the growth of a scientific field all encourage further work in the construction of this model.

6.2 Competition

Competition is an important aspect of ant foraging as well as innovation processes. This plays a part in the design of strategies. Focusing on a specific part of food network and going for non-distant food may result in suboptimal total network discovery since other competing colonies can find majority of the food and collect it before the original colony has a chance. Thus, since the current model does not have competition incorporated into it, it forms an option for enhancing the model in future.

6.3 Optimization Algorithms and Reinforcement Learning

The use of ant strategies in search algorithms is well known. The model proposed in this work can be modified to implement existing optimization algorithms such as Ant Colony Optimization [122]. The idea of this work can be converted into an optimization problem as 'Collect maximum food particles using minimum energy'. Energy in this case can be defined as the amount of time spent outside the nest foraging. Given different pathways that the ants can take as well as the distribution of pheromones over time, the algorithm can effectively find the best edge that each ant must travel in order to find the best solution. Judging by these selections, a general rule of thumb regarding which path to choose can be extracted in the form of a strategy that the ants seem to follow. Further extensions to the algorithm, such as Ant Colony System [123], Max-Min Ant System [124], etc. can allow a robust design of the model by including multiple parameters and finding best strategies.

Reinforcement learning includes algorithms that train agents how to behave in a given environment and perfect themselves over a number of iterations. Thus, in the context of foraging, this would involve training ants to find the quickest ways around the environment. This also allows ants to spend less time figuring out most optimal strategies in drastically different environments.

The design of reward functions, such as a negative reward for walking at every time step, a positive reward for finding food, a positive reward for depositing the food at the nest, a negative reward for carrying the food away from the nest, etc., can inculcate optimal behavior into the ants. There have been multiple studies based on ant foraging in reinforcement learning and they have either used a modified version of existing algorithms in a multi-agent context, such as Decentralized Q-Learning [125], or have led to the development of new algorithms, such as Pheromone Q-Learning [126]. This model can implement such reinforcement learning and train the ants to find the best foraging strategies by interaction with the environment over time rather than searching the parameter space, bringing the model closer to the way real ants develop and perfect their strategies.

6.4 Further Ideas

Ants can be provided more individuality by adding the concept of memory and a preference towards sites that are stored in memory rather than completely new ones – also known as preferential attachment. An important thing this model lacks is the translation of the evolution of original field into model parameters, meaning figuring out what parameters make the model perform as well as the researchers did in the field. A detailed optimization analysis can add this information, thereby allowing to predict strategies better than the ones originally followed. This can help predict efficient actions to take in the development of a new scientific field, if its current evolution follows similar trend as the past evolution of an old well-studied field. If the evolution of field 'A' and field 'B' shows very similar trends, and it is known how risky or conservative the research was in the development of field 'A', the model can study the evolution of field'A' and suggest strategies that might boost the growth of field 'B'. Thus, the model encourages a number of aspects for improvement.

REFERENCES

[1] A. Baregheh, J. Rowley, and S. Sambrook, "Towards a multidisciplinary definition of innovation," *Management decision*, 2009.

[2] I. L. Popa, G. Preda, and M. Boldea, "A theoretical approach of the concept of innovation," *Managerial Challenges of the Contemporary Society. Proceedings*, p. 151, 2010.

[3] J. G. Foster, A. Rzhetsky, and J. A. Evans, "Tradition and innovation in scientists' research strategies," *American Sociological Review*, vol. 80, no. 5, pp. 875–908, 2015.

[4] J. A. Evans, "Industry induces academic science to know less about more," *American Journal of Sociology*, vol. 116, no. 2, pp. 389–452, 2010.

[5] J. Wang, R. Veugelers, and P. Stephan, "Bias against novelty in science: A cautionary tale for users of bibliometric indicators," *Research Policy*, vol. 46, no. 8, pp. 1416–1436, 2017.

[6] C. Phelps, R. Heidl, and A. Wadhwa, "Knowledge, networks, and knowledge networks: A review and research agenda," *Journal of management*, vol. 38, no. 4, pp. 1115–1166, 2012.

[7] A. Pyka, "Innovation networks in economics: from the incentive-based to the knowledge-based approaches," *European Journal of Innovation Management*, 2002.

[8] P. P. Saviotti, "Knowledge networks: structure and dynamics," in *Innovation Networks*, pp. 19–41, Springer, 2009.

[9] T. van der Valk, M. M. Chappin, and G. W. Gijsbers, "Evaluating innovation networks in emerging technologies," *Technological Forecasting and Social Change*, vol. 78, no. 1, pp. 25–39, 2011.

[10] N. Gilbert, "A simulation of the structure of academic science," *Sociological research online*, vol. 2, no. 2, pp. 91–105, 1997.

[11] N. Gilbert, A. Pyka, P. Ahrweiler, *et al.*, "Innovation networks-a simulation approach," *Journal of Artificial Societies and Social Simulation*, vol. 4, no. 3, pp. 1–13, 2001.

[12] F. Menczer, “Evolution of document networks,” *Proceedings of the National Academy of Sciences*, vol. 101, no. suppl 1, pp. 5261–5265, 2004.

[13] M. Rosvall and C. T. Bergstrom, “Mapping change in large networks,” *PloS one*, vol. 5, no. 1, p. e8694, 2010.

[14] C. Chen, Y. Chen, M. Horowitz, H. Hou, Z. Liu, and D. Pellegrino, “Towards an explanatory and computational theory of scientific discovery,” *Journal of Informetrics*, vol. 3, no. 3, pp. 191–209, 2009.

[15] R. J. Funk and J. Owen-Smith, “A dynamic network approach to breakthrough innovation,” *arXiv preprint arXiv:1212.3559*, 2012.

[16] J. E. Perry-Smith and C. E. Shalley, “The social side of creativity: A static and dynamic social network perspective,” *Academy of management review*, vol. 28, no. 1, pp. 89–106, 2003.

[17] M. Andjelković, B. Tadić, M. Mitrović Dankulov, M. Rajković, and R. Melnik, “Topology of innovation spaces in the knowledge networks emerging through questions-and-answers,” *PloS one*, vol. 11, no. 5, p. e0154655, 2016.

[18] G. C. Rodi, V. Loreto, and F. Tria, “Search strategies of wikipedia readers,” *PloS one*, vol. 12, no. 2, p. e0170746, 2017.

[19] F. Saracco, R. Di Clemente, A. Gabrielli, and L. Pietronero, “From innovation to diversification: a simple competitive model,” *PloS one*, vol. 10, no. 11, p. e0140420, 2015.

[20] A. Rzhetsky, J. G. Foster, I. T. Foster, and J. A. Evans, “Choosing experiments to accelerate collective discovery,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 47, pp. 14569–14574, 2015.

[21] M. Cokol, I. Iossifov, C. Weinreb, and A. Rzhetsky, “Emergent behavior of growing knowledge about molecular interactions,” *Nature biotechnology*, vol. 23, no. 10, pp. 1243–1247, 2005.

[22] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[23] D. Boyer and C. Solis-Salas, "Random walks with preferential relocations to places visited in the past and their application to biology," *Physical review letters*, vol. 112, no. 24, p. 240601, 2014.

[24] J. Choi, J.-I. Sohn, K.-I. Goh, and I.-M. Kim, "Modeling the mobility with memory," *EPL (Europhysics Letters)*, vol. 99, no. 5, p. 50001, 2012.

[25] I. Iacopini, S. Milojević, and V. Latora, "Network dynamics of innovation processes," *Physical review letters*, vol. 120, no. 4, p. 048301, 2018.

[26] A. J. Meadows, *Communicating research*. Academic Press a division of Harcourt Brace & Company, 1998.

[27] C. Udovicich, V. Kasivisvanathan, and C. L. Winchester, "Communicating your research (part 1)–to the scientific community," *Journal of Clinical Urology*, vol. 10, no. 4, pp. 396–399, 2017.

[28] B. H. Hall, "Innovation and diffusion," tech. rep., National Bureau of Economic Research, 2004.

[29] K. Powell, "Does it take too long to publish research?," *Nature News*, vol. 530, no. 7589, p. 148, 2016.

[30] D. Corsaro, C. Cantù, and A. Tunisini, "Actors' heterogeneity in innovation networks," *Industrial Marketing Management*, vol. 41, no. 5, pp. 780–789, 2012.

[31] L. Keller and M. Chapuisat, "Eusociality and cooperation," *e LS*, 2001.

[32] K. Börner, S. Sanyal, and A. Vespignani, "Network science," *Annual review of information science and technology*, vol. 41, no. 1, pp. 537–607, 2007.

[33] S. Zhou and R. J. Mondragón, "Accurately modeling the internet topology," *Physical Review E*, vol. 70, no. 6, p. 066108, 2004.

[34] M. Herty and A. Klar, "Modeling, simulation, and optimization of traffic flow networks," *SIAM Journal on Scientific Computing*, vol. 25, no. 3, pp. 1066–1087, 2003.

[35] P. P. Saviotti, "Considerations about the production and utilization of knowledge," *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift Für Die Gesamte Staatswissenschaft*, pp. 100–121, 2004.

[36] E. W. Weisstein, "Taniyama-shimura conjecture," *https://mathworld. wolfram. com/*, 2009.

[37] A. Pyka *et al.*, "Innovation networks," *Chapters*, 2007.

[38] H. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," *Journal of the American Society for information Science*, vol. 24, no. 4, pp. 265–269, 1973.

[39] M. M. Kessler, "Bibliographic coupling between scientific papers," *American documentation*, vol. 14, no. 1, pp. 10–25, 1963.

[40] K. Börner, C. Chen, and K. W. Boyack, "Visualizing knowledge domains," *Annual review of information science and technology*, vol. 37, no. 1, pp. 179–255, 2003.

[41] B. Gupta and S. Bhattacharya, "Bibliometric approach towards mapping the dynamics of science and technology," *DESIDOC Journal of Library & Information Technology*, vol. 24, no. 1, 2004.

[42] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools," *Journal of the American Society for information Science and Technology*, vol. 62, no. 7, pp. 1382–1402, 2011.

[43] E. Yan and Y. Ding, "Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 7, pp. 1313–1326, 2012.

[44] A. Pentland, *Social physics: How good ideas spread-the lessons from a new science*. Penguin, 2014.

[45] R. H. MacArthur and E. R. Pianka, "On optimal use of a patchy environment," *The American Naturalist*, vol. 100, no. 916, pp. 603–609, 1966.

[46] G. Pyke, "Optimal foraging theory: An introduction," in *Encyclopedia of animal behavior*, pp. 111–117, Elsevier Academic Press, 2019.

[47] P. Meire and A. Ervynck, "Are oystercatchers (haematopus ostralegus) selecting the most profitable mussels (mytilus edulis)?," *Animal Behaviour*, vol. 34, no. 5, pp. 1427–1435, 1986.

[48] L. M. Bautista, J. Tinbergen, and A. Kacelnik, "To walk or to fly? how birds choose among foraging modes," *Proceedings of the National Academy of Sciences*, vol. 98, no. 3, pp. 1089–1094, 2001.

[49] P. Schmid-Hempel, A. Kacelnik, and A. I. Houston, "Honeybees maximize efficiency by not filling their crop," *Behavioral Ecology and Sociobiology*, vol. 17, no. 1, pp. 61–66, 1985.

[50] J. Manatunge and T. Asaeda, "Optimal foraging as the criteria of prey selection by two centrarchid fishes," *Hydrobiologia*, vol. 391, no. 1, pp. 221–238, 1998.

[51] G. H. Pyke, "Optimal foraging theory: a critical review," *Annual review of ecology and systematics*, vol. 15, no. 1, pp. 523–575, 1984.

[52] G. J. Pierce and J. Ollason, "Eight reasons why optimal foraging theory is a complete waste of time," *Oikos*, pp. 111–118, 1987.

[53] R. D. Gray, "Faith and foraging: a critique of the "paradigm argument from design"," in *Foraging behavior*, pp. 69–140, Springer, 1987.

[54] P. Pirolli and S. Card, "Information foraging.," *Psychological review*, vol. 106, no. 4, p. 643, 1999.

[55] W.-T. Fu and P. Pirolli, “Snif-act: A cognitive model of user navigation on the world wide web,” *Human–Computer Interaction*, vol. 22, no. 4, pp. 355–412, 2007.

[56] C. S. Miller and R. W. Remington, “Modeling information navigation: Implications for information architecture,” *Human-computer interaction*, vol. 19, no. 3, pp. 225–271, 2004.

[57] P. E. Sandstrom, “Scholars as subsistence foragers,” *Bulletin of the American Society for Information Science and Technology*, vol. 25, no. 3, pp. 17–20, 1999.

[58] J. H. Holland, *Emergence: From chaos to order*. OUP Oxford, 2000.

[59] S. Johnson, *Emergence: The connected lives of ants, brains, cities, and software*. Simon and Schuster, 2002.

[60] O. Feinerman, A. Korman, J. D. Levine, D. J. Kronauer, and M. H. Dickinson, “Individual versus collective cognition in social insects,” *Journal of Experimental Biology*, vol. 220, no. 1, pp. 73–82, 2017.

[61] J. F. Traniello, “Foraging strategies of ants,” *Annual review of entomology*, vol. 34, no. 1, pp. 191–210, 1989.

[62] M. Lanan, “Spatiotemporal resource distribution and foraging strategies of ants (hymenoptera: Formicidae),” *Myrmecological news/Osterreichische Gesellschaft fur Entomofaunistik*, vol. 20, p. 53, 2014.

[63] J. M. Chalissery, A. Renyard, R. Gries, D. Hoefele, S. K. Alamsetti, and G. Gries, “Ants sense, and follow, trail pheromones of ant community members,” *Insects*, vol. 10, no. 11, p. 383, 2019.

[64] T. J. Czaczkes, C. Grüter, and F. L. Ratnieks, “Trail pheromones: an integrative view of their role in social insect colony organization,” *Annual review of entomology*, vol. 60, 2015.

[65] E. J. Robinson, D. E. Jackson, M. Holcombe, and F. L. Ratnieks, “‘no entry’signal in ant foraging,” *Nature*, vol. 438, no. 7067, pp. 442–442, 2005.

[66] A. Dussutour, S. C. Nicolis, G. Shephard, M. Beekman, and D. J. Sumpter, "The role of multiple pheromones in food recruitment by ants," *Journal of Experimental Biology*, vol. 212, no. 15, pp. 2337–2348, 2009.

[67] E. O. Wilson, "Chemical communication among workers of the fire ant solenopsis saevissima (fr. smith) 1. the organization of mass-foraging," *Animal behaviour*, vol. 10, no. 1-2, pp. 134–147, 1962.

[68] J. Adams, "Early citation counts correlate with accumulated impact," *Scientometrics*, vol. 63, no. 3, pp. 567–581, 2005.

[69] G. Debout, B. Schatz, M. Elias, and D. McKey, "Polydomy in ants: what we know, what we think we know, and what remains to be done," *Biological Journal of the Linnean Society*, vol. 90, no. 2, pp. 319–348, 2007.

[70] J. McIver, "Dispersed central place foraging in australian meat ants," *Insectes Sociaux*, vol. 38, no. 2, pp. 129–137, 1991.

[71] M. C. Lanan, A. Dornhaus, and J. L. Bronstein, "The function of polydomy: the ant crematogaster torosa preferentially forms new nests near food sources and fortifies outstations," *Behavioral Ecology and Sociobiology*, vol. 65, no. 5, pp. 959–968, 2011.

[72] R. Wehner, "Desert ant navigation: how miniature brains solve complex tasks," *Journal of Comparative Physiology A*, vol. 189, no. 8, pp. 579–588, 2003.

[73] T. Merkle and R. Wehner, "Desert ants use foraging distance to adapt the nest search to the uncertainty of the path integrator," *Behavioral Ecology*, vol. 21, no. 2, pp. 349–355, 2010.

[74] D. J. Sumpter and M. Beekman, "From nonlinearity to optimality: pheromone trail foraging by ants," *Animal behaviour*, vol. 66, no. 2, pp. 273–280, 2003.

[75] D. E. Jackson and N. Châline, "Modulation of pheromone trail strength with food quality in pharaoh's ant, monomorium pharaonis," *Animal behaviour*, vol. 74, no. 3, pp. 463–470, 2007.

[76] H. F. Moed, “Citation analysis of scientific journals and journal impact measures,” *Current Science*, pp. 1990–1996, 2005.

[77] N. L. Geller, J. S. de Cani, and R. E. Davies, “Lifetime-citation rates: A mathematical model to compare scientists’ work,” *Journal of the American Society for Information Science*, vol. 32, no. 1, pp. 1–15, 1981.

[78] E. J. Robinson, F. L. Ratnieks, and M. Holcombe, “An agent-based model to investigate the roles of attractive and repellent pheromones in ant decision making during foraging,” *Journal of theoretical Biology*, vol. 255, no. 2, pp. 250–258, 2008.

[79] J. C. Biesmeijer and H. de Vries, “Exploration and exploitation of food sources by social insect colonies: a revision of the scout-recruit concept,” *Behavioral Ecology and Sociobiology*, vol. 49, no. 2, pp. 89–99, 2001.

[80] C. Grüter, T. J. Czaczkes, and F. L. Ratnieks, “Decision making in ant foragers (lasius niger) facing conflicting private and social information,” *Behavioral Ecology and Sociobiology*, vol. 65, no. 2, pp. 141–148, 2011.

[81] K. Eguchi, T. V. Bui, and S. Yamane, “A preliminary study on foraging distance and nesting sites of ants in indo-chinese lowland vegetation,” *Sociobiology*, vol. 43, no. 3, pp. 445–457, 2004.

[82] D. M. Gordon, “The expandable network of ant exploration,” *Animal Behaviour*, vol. 50, no. 4, pp. 995–1007, 1995.

[83] N. J. Van Eck and L. Waltman, “Vos: A new method for visualizing similarities between objects,” in *Advances in data analysis*, pp. 299–306, Springer, 2007.

[84] N. J. Van Eck, L. Waltman, R. Dekker, and J. van den Berg, “A comparison of two techniques for bibliometric mapping: Multidimensional scaling and vos,” *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2405–2416, 2010.

[85] N. J. Van Eck and L. Waltman, “Software survey: Vosviewer, a computer program for bibliometric mapping,” *scientometrics*, vol. 84, no. 2, pp. 523–538, 2010.

[86] N. J. v. Eck and L. Waltman, “How to normalize cooccurrence data? an analysis of some well-known similarity measures,” *Journal of the American society for information science and technology*, vol. 60, no. 8, pp. 1635–1651, 2009.

[87] N. J. Van Eck and L. Waltman, “Text mining and visualization using vosviewer,” *arXiv preprint arXiv:1109.2058*, 2011.

[88] E. Garcia, M. A. Jimenez, P. G. De Santos, and M. Armada, “The evolution of robotics research,” *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 90–103, 2007.

[89] H. Li, H. An, Y. Wang, J. Huang, and X. Gao, “Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network,” *Physica A: Statistical Mechanics and its Applications*, vol. 450, pp. 657–669, 2016.

[90] J. Zhang, Q. Yu, F. Zheng, C. Long, Z. Lu, and Z. Duan, “Comparing keywords plus of wos and author keywords: A case study of patient adherence research,” *Journal of the Association for Information Science and Technology*, vol. 67, no. 4, pp. 967–972, 2016.

[91] A. Perianes-Rodriguez, L. Waltman, and N. J. Van Eck, “Constructing bibliometric networks: A comparison between full and fractional counting,” *Journal of Informetrics*, vol. 10, no. 4, pp. 1178–1195, 2016.

[92] L. Waltman, N. J. Van Eck, and E. C. Noyons, “A unified approach to mapping and clustering of bibliometric networks,” *Journal of informetrics*, vol. 4, no. 4, pp. 629–635, 2010.

[93] B. Lei, G. Xu, M. Feng, Y. Zou, F. Van der Heijden, D. De Ridder, and D. Tax, *Classification, Parameter Estimation and State Estimation*. Wiley Online Library, 2017.

[94] J. J. Howard, L. M. HENNEMAN, G. Cronin, J. A. Fox, and G. Hormiga, “Conditioning of scouts and recruits during foraging by a leaf-cutting ant, atta colombica,” *Animal Behaviour*, vol. 52, no. 2, pp. 299–306, 1996.

[95] W. R. Tschinkel, “The organization of foraging in the fire ant, solenopsis invicta,” *Journal of Insect Science*, vol. 11, no. 1, p. 26, 2011.

[96] C. Devigne and C. Detrain, “How does food distance influence foraging in the ant lasius niger: the importance of home-range marking,” *Insectes sociaux*, vol. 53, no. 1, pp. 46–55, 2006.

[97] R. Planqué, J. B. Van Den Berg, and N. R. Franks, “Recruitment strategies and colony size in ants,” *PLoS One*, vol. 5, no. 8, p. e11664, 2010.

[98] B. Hölldobler, M. Möglich, and U. Maschwitz, “Communication by tandem running in the ant camponotus sericeus,” *Journal of Comparative physiology*, vol. 90, no. 2, pp. 105–127, 1974.

[99] B. Hölldobler, N. Oldham, E. Morgan, and W. König, “Recruitment pheromones in the ants aphaenogaster albisetosus and a. cockerelli (hymenoptera: Formicidae),” *Journal of insect physiology*, vol. 41, no. 9, pp. 739–744, 1995.

[100] J. Billen, W. Beeckman, and E. D. Morgan, “Active trail pheromone compounds and trail following in the ant atta sexdens sexdens (hymenoptera formicidae),” *Ethology ecology & evolution*, vol. 4, no. 2, pp. 197–202, 1992.

[101] K. Jaffe and P. Howse, “The mass recruitment system of the leaf cutting ant, atta cephalotes (l.),” *Animal Behaviour*, vol. 27, pp. 930–939, 1979.

[102] R. Harkness and R. Wehner, “Cataglyphis,” *Endeavour*, vol. 1, no. 3-4, pp. 115–121, 1977.

[103] T. P. Flanagan, N. M. Pinter-Wollman, M. E. Moses, and D. M. Gordon, “Fast and flexible: Argentine ants recruit from nearby trails,” *PloS one*, vol. 8, no. 8, p. e70888, 2013.

[104] P. Schmid-Hempel, “Individually different foraging methods in the desert ant cataglyphis bicolor (hymenoptera, formicidae),” *Behavioral Ecology and Sociobiology*, vol. 14, no. 4, pp. 263–271, 1984.

[105] P. Schmid-Hempel and R. Schmid-Hempel, “Life duration and turnover of foragers in the ant cataglyphis bicolor (hymenoptera, formicidae),” *Insectes sociaux*, vol. 31, no. 4, pp. 345–360, 1984.

[106] R. Harkness and N. Maroudas, “Central place foraging by an ant (cataglyphis bicolor fab.): a model of searching,” *Animal Behaviour*, vol. 33, no. 3, pp. 916–928, 1985.

[107] R. Wehner, C. Meier, and C. Zollikofer, “The ontogeny of foragwehaviour in desert ants, cataglyphis bicolor,” *Ecological Entomology*, vol. 29, no. 2, pp. 240–250, 2004.

[108] J. K. Wetterer, A. L. Wild, A. V. Suarez, N. Roura-Pascual, and X. Espadaler, “Worldwide spread of the argentine ant, linepithema humile (hymenoptera: Formicidae),” *Myrmecological News*, vol. 12, pp. 187–194, 2009.

[109] D. M. Gordon and N. E. Heller, “The invasive argentine ant linepithema humile (hymenoptera: Formicidae) in northern california reserves: from foraging behavior to local spread,” *Myrmecol News*, vol. 19, pp. 103–110, 2014.

[110] S. Aron, J. M. Pasteels, and J. L. Deneubourg, “Trail-laying behaviour during exploratory recruitment in the argentine ant, iridomyrmex humilis (mayr),” *Biology of Behaviour*, vol. 14, no. 3, pp. 207–217, 1989.

[111] J.-L. Deneubourg, S. Aron, S. Goss, and J. M. Pasteels, “The self-organizing exploratory pattern of the argentine ant,” *Journal of insect behavior*, vol. 3, no. 2, pp. 159–168, 1990.

[112] H. R. T’ai and J. Silverman, “The effect of food size and dispersion pattern on retrieval rate by the argentine ant, linepithema humile (hymenoptera: Formicidae),” *Journal of insect behavior*, vol. 15, no. 5, pp. 633–648, 2002.

[113] G. P. Markin, “Foraging behavior of the argentine ant in a california citrus grove,” *Journal of Economic Entomology*, vol. 63, no. 3, pp. 740–744, 1970.

[114] S. E. Evison, O. L. Petchey, A. P. Beckerman, and F. L. Ratnieks, “Combined use of pheromone trails and visual landmarks by the common garden ant lasius niger,” *Behavioral Ecology and Sociobiology*, vol. 63, no. 2, pp. 261–267, 2008.

[115] R. Beckers, J.-L. Deneubourg, and S. Goss, “Trail laying behaviour during food recruitment in the ant lasius niger (l.),” *Insectes Sociaux*, vol. 39, no. 1, pp. 59–72, 1992.

[116] R. Beckers, J.-L. Deneubourg, and S. Goss, "Modulation of trail laying in the antlasius niger (hymenoptera: Formicidae) and its role in the collective selection of a food source," *Journal of Insect Behavior*, vol. 6, no. 6, pp. 751–759, 1993.

[117] R. Beckers, J. P. Lachaud, and D. Fresneau, "The influence of olfactory conditioning on food preference in the ant lasius niger (l.)," *Ethology ecology & evolution*, vol. 6, no. 2, pp. 159–167, 1994.

[118] A.-W. Harzing and R. Van der Wal, "A google scholar h-index for journals: An alternative metric to measure journal impact in economics and business," *Journal of the American Society for Information Science and technology*, vol. 60, no. 1, pp. 41–46, 2009.

[119] E. Garfield *et al.*, "The impact factor," *Current contents*, vol. 25, no. 20, pp. 3–7, 1994.

[120] P. Van den Besselaar and G. Heimeriks, "Mapping research topics using word-reference co-occurrences: A method and an exploratory case study," *Scientometrics*, vol. 68, no. 3, pp. 377–393, 2006.

[121] P. Kareiva and N. Shigesada, "Analyzing insect movement as a correlated random walk," *Oecologia*, vol. 56, no. 2-3, pp. 234–238, 1983.

[122] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE computational intelligence magazine*, vol. 1, no. 4, pp. 28–39, 2006.

[123] M. Dorigo and L. M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *IEEE Transactions on evolutionary computation*, vol. 1, no. 1, pp. 53–66, 1997.

[124] T. Stutzle and H. Hoos, "Max-min ant system and local search for the traveling salesman problem," in *Proceedings of 1997 IEEE international conference on evolutionary computation (ICEC'97)*, pp. 309–314, IEEE, 1997.

[125] G. Arslan and S. Yüksel, "Decentralized q-learning for stochastic teams and games," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1545–1558, 2016.

[126] N. Monekosso and P. Remagnino, “Phe-q: A pheromone based q-learning,” in *Australian Joint Conference on Artificial Intelligence*, pp. 345–355, Springer, 2001.