Mälardalen University

Masters Thesis

Innovation Diffusion in Scale-Free Networks: Signal Analysis in Complex Networks

Author:
Debajyoti Nag

Supervisor:
Baran Cürüklü

Examiner:
Mikael Ekström

A thesis submitted in fulfilment of the requirements for the degree of Master of Science

in the

Division of Intelligent Future Technologies
School of Innovation, Design and Engineering

December 2014

Declaration of Authorship

I, Debajyoti Nag, declare that this thesis titled, 'Innovation Diffusion in Scale-Free Networks: Signal Analysis in Complex Networks' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: __________________________________

Date: __________________________________

"Knowledge is the food of the soul."

Plato.

MÄLARDALEN UNIVERSITY

Abstract

Intelligent Embedded Systems
School of Innovation, Design and Engineering

Master of Science

Innovation Diffusion in Scale-Free Networks: Signal Analysis in Complex Networks

by Debajyoti NAG

With personalisation of computers, specially personalised embedded devices like smartphones, the connections between the users can be seen emerging within the network of these connected devices. Study of these connections can help identify useful patterns or relations within these networks. This thesis focuses on the idea that in such a network, marketing strategies can be optimised for maximising their net profit based on customer involvement policies followed by different companies, for a given campaign or propaganda. Since it is not an easy task to keep records of the connections and concerned interactions of every employee and/or customer, the task of analysis is better performed by a computerised model of the network where the nodes represent humans and the connections represent their interactions. The author follows the Barabási-Albert (BA) model to generate a scale free network, and simulates competitive markets. He tries to gain insight into the methods of interaction in the society, and find a balance in utility and cost incurred for that utility. He simulates an environment with two competing companies trying to attract the customers, and goes on to analyse the results and show that different policies of customer involvement are favoured under different circumstances. In the tested scenarios, the net profit is measured in terms of effective outreach.

Acknowledgements

I would like to thank my supervisor, Baran Cürüklü, for his valuable guidance and patience.

I thank Carina Sjödin, PhD student in Innovation Management, for her guidance regarding the aspects of Innovation in customer relations.

Finally, I would like to thank my loved ones for their support and encouragement.

Contents

List of Figures

List of Tables

For the bewildered spirits who make it their business to know that which other men don't.

Chapter 1

Introduction

1.1 Networks

A network is an arrangement of two or more different entities, mostly linked together for sharing resources. The resources being shared can be either physical or otherwise. These entities being connected are referred to as nodes, and the links between them are called the connections of the network.

1.1.1 Computer Networks

Computer Networks are networks of interconnected devices that can exchange and share information and resources. The nodes in a computer network can by devices such as a laptop computers, smartphones, or even fitness trackers and smart wearables. These networks may be categorized depending on the type of connections being formed, and the rules or protocols governing communications between different nodes within the network. While most nodes may support both incoming and outgoing connections, some nodes may have restrictions and only support one type of connections. Such networks are also known as Data Networks.

1.1.2 Social Networks

Social Networks refer to arrangement of individual agents, most commonly humans or their digital representatives, such that these arrangements are dependent on the actions of the agents and also influence their future actions. This is also referred to as a 'social structure'. With the personalization of technology, computer networks are shaping themselves to get even closer to the social networks in terms of structure, shape, size, and degree distribution, for example. A comprehensive review of the same was done by Newman [1].

This is particularly important and interesting with the emergence of personalized wearable devices. This could lead to a 1:1 man to machine ratio, which would mean, when allowed, the personal devices would network exactly as their users.

1.1.3 Scale-Free Networks

A scale-free network is a network with the property that the fraction P(k) of nodes with 'k' connections is determined as:

$$P(k) \sim k^{-\gamma} \tag{1.1}$$

In general, the degree distribution of the network follows a power law. This means that the network is scalable, and the properties of degree distribution of the nodes is not affected by the size of the network. These networks follow a strategy such that there are a lot of nodes with few connections, and a very few nodes with significantly larger number of connections. These highly-connected nodes can be though of as hubs in the network.

The work in [2] explains how diverse systems are best described as complex topological networks, and shows how prevalent scale-free networks are in natural as well as man-made structures alike. It presents a reproducible model which indicates strong self-governing principles lead to formation of such networks without being affected by the particulars of the individuals.

The author uses this approach successfully in this thesis to develop a model despite the heterogeneous nature of the different agents in the network.

1.1.4 Innovation Diffusion

Diffusion of innovations refers to the theory of understanding the factors governing spread and acceptance of any new idea or product in a social system. It tries to understand and formalise the critical point in the process of spreading the idea after which the idea can self-sustain in the system. In his book [3], Everett Rogers explains how this process relies heavily on human capital and proposes the four main influencing factors, namely :

The innovation itself

Communication channels

Time, and

Social system

Among these, the author decided to focus on the affect of the structure of the social system and time of campaigning for this thesis. He also decided to restrict the communication channel as much as possible to accommodate for the fact of randomness and chance in real world scenarios.

1.2 State of the Art

Humans in present day society must network to sustain and complete tasks. Studies have shown networks to evolve over time to optimize functionalities, and increase reach and longevity. These evolutions may be guided or self-governed. The most coveted type of network, perhaps, is the kind formed naturally among varied entities, i.e., free scaling networks [2, 4].

The first model for free scaling networks was proposed by Barabási–Albert [2], hereinafter referred to as B-A model, and the works of Erdős & Rényi [5], Watts and Strogatz [6], Watts, Strogatz and Newman [7], Saramäki and Kaski [8] , Yang and Jure [9] , and Rycroft [10] also helped greatly in laying the groundwork.

It is prominent that existing social networks tend to scale freely and it is safe to assume that personalized devices, if and when allowed to network, will follow a similar trace.

1.2.1 State of the art in networked devices

Internet of Things [11] - Internet connectivity and accessibility in present day enables direct linking of everyday physical objects, overcoming the spatio-temporal boundaries. This upgrades a part of our lives to the so called cloud. It makes our lives easier, but also raises confusion as the complexity of processes increase manifold.

Cyber-Physical Systems and Machine-to-Machine [12] - Internet of Things is a small part of a larger picture, where computational devices of all size and shape interact with each other and with everyday objects to perform complex tasks, ranging from temperature control in a modern house, to automatic detection of spread of a potentially fatal epidemic. Machine-to-Machine communication makes it possible for sensors spread over a large area to share the load of detecting varied signals, while some entirely different processing entity looks into that raw data and extracts meaningful information from it, and then a strong/special link actuates some physical entity to act based on the information just gathered. However, large scale implementations of Cyber-Physical Systems face a major challenge due to the heterogeneous nature of network elements.

1.2.2 Examples of State of the Art networked devices

1. GremlinMusic [13]

   Gremlin showed a concept of interconnecting embedded devices in a whole new light. It not only allowed users to carry their music along (like every music player), but it allowed friends to connect their Gremlins and legally share music with each other. It allowed for an optimization of storage, bandwidth (in p2p form), and monetary resources for the users. An analogy for the same could be to take an iPod and put Facebook and a free Spotify premium on it.

2. p-Cell technology [14]

   The new technology by Artemis seems promising, and could have crucial impact on the state of networked devices over time.

3. Swarm robots

Swarm robots are a good example of how even heterogeneous entities can work together, like the coordination between the Eye-Feet-Hand bots [15] to achieve their goal.

1.3 Motivation

Since networks are everywhere, it's important we understand their nature and working so as to exploit and utilize them to our benefit. Going back as far as the 18th century, the "Seven Bridges of Königsberg" [16] might be the most famous networking problem. Ranging from the Travelling-Salesman [17] to Graph-Colouring algorithms [18], insight into the working of networks have helped greatly in optimizing several issues.

1.3.1 Why Scale-Free Networks

Scale-Free networks are the most prevalent in nature. Hence, it is paramount to model networking of personal devices used by humans on the same topology. This model aims to gain insight into the workings of such networks, so as to optimize the network elements and make use of their full potential. For e.g.., by finding the optimal positioning for a self-assembling network of satellites, military installations can provide improved security at a lesser cost. Or multiple groups of Swarm networks, like groups of eye-hand-feet bots [15], can work together as a collective with an increased proficiency. Most importantly, due to the autonomous nature of most network elements, these networks can be formed anywhere, water, air, or land, and can accomplish a multitude of tasks.

1.3.2 Human to Human

Human beings are innately complex, yet strive for simplicity [19]. The challenge today as humans is to find, understand and explain the complex in its most simplistic form. This is important not only for businesses, but for overall development of human society to a point where communications may have the desired effect to a greater extent.

1.3.3 Future Sight

Imagine if smart phones were a bit smarter, and they synchronized the user's calender to include events from his/her best friend's public calender automatically, or if it alerted the user that the person they are supposed to meet has now entered the building. Or imagine a world where every child has a personal monitoring device, in form of a wearable device, and as the children go to the playground in the evening, these devices are with them. Now, not only do they serve the purpose of monitoring the child, but these devices communicate with each other and form a bond, in a most likely manner that the children form their bonds, so while the kid is busy playing, his/her device could find out about the studies they missed at school.

Networked devices could make life easier and yet more manageable.

1.4 Aim

This thesis aims to analyse diffusion of innovation in the society. The author simulates a real-world scenario where companies try to influence customers by spreading an idea or publicizing their products. As an effect, the customer agents, hereinafter referred to as 'agents' only, tend to lean towards the company if they are successfully , and enough, influenced. However, this campaigning propaganda incurs some costs to the companies. The author tries to find a balance between gain and cost for the company, in order to formulate a strategy which can be effectively used by the company to optimally attract customers.

Different companies have different policies towards customer involvement and making an effort towards further attracting the customer. The objective here is to analyse and evaluate two of such policies, and to determine which policy is optimal for the given circumstances.

Chapter 2

Approach and Tools

2.1 Approach

2.1.1 Network Model

Naturally occurring topologies like protein chains, cooperation networks and social networks follow scale-free structures, hence it was an obvious choice for modelling the network. For this purpose, the B-A model was followed, which also proposes variants of the scale free structure depending on its characteristics. These characteristics are :

1. Continuous Growth (CG) This characteristic marks that the network grows continuously. i.e., at all times, new nodes are being attached to the the network. At the beginning of the simulation, the network only has a small number of nodes and this is known as the *seed network*, and with time, new nodes are introduced in the network and it grows in size. The author uses a *fully-connected network* as a seed network. Fully-connected network means each node of the network is connected to every other node in the network.

2. Preferential Attachment (PA) This characteristic formulates the rule an incoming node should follow while making new connections within the network. It states that the most connected nodes are the most likely candidates for forming a new connection with. This leads on to a *'rich gets richer'* tendency, and highly connected

hubs are formed. The connections are formed according to a probability value which is calculated as :

$$p_i = \frac{k_i}{\sum_{j=1}^{N} k_j} \tag{2.1}$$

where,

(a) p : probability

(b) k : number of connections

(c) i : agent id

(d) N : total number of agents

A scale free network may exhibit either or both of the above characteristics. However, the author decided to include both attributes in his model to make it as close to real life situations as possible.

2.1.2 Orientation and View of the simulation environment

Initially, an object-oriented view was used in the model to give better control over the agents in the network. But this led to very high time complexity, and the author chose to switch to a connection-oriented view of the model, where more focus was given to the connections being formed and everything was managed according to that view. This resulted in significant decrease in time complexity. Also, focusing on the connections was easier as the whole network could be minimally represented by using the edge list.

Throughout the simulation, the network is represented by an edge list. It's a list, where each entry is a vector, and minimally this vector is a pair of two identification numbers referring to two distinct nodes. The presence of such a pair indicates the presence of a connection between the two nodes. It's notable that the values are always distinct, and this eliminates any chance of self-looping within the network. Optionally, the entries may contain other properties of the particular connection that may be used at different stages throughout the simulation.

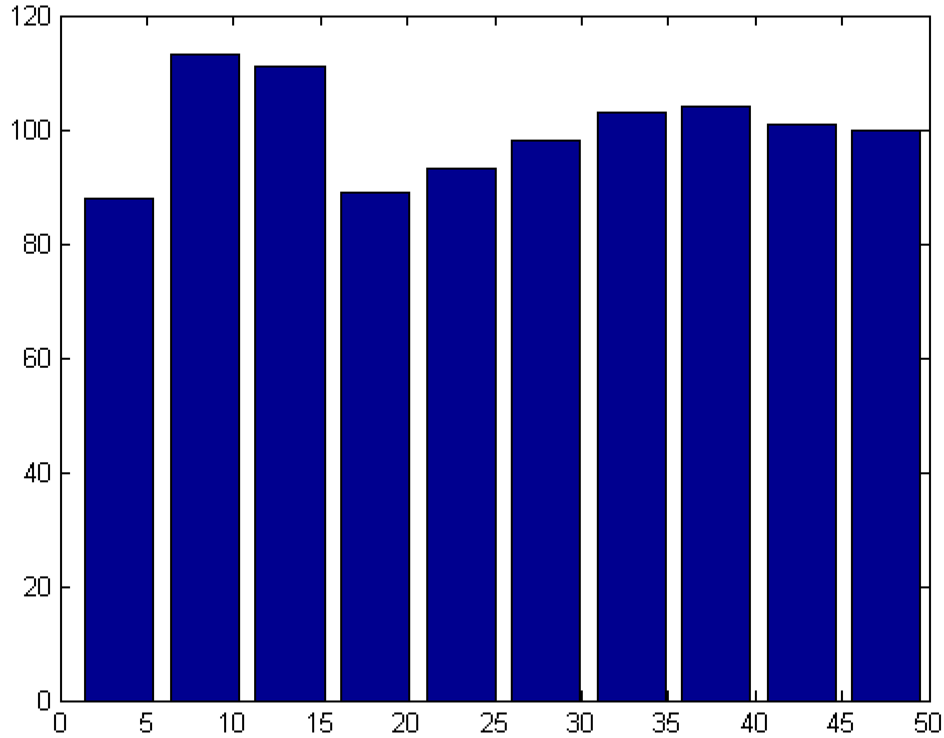


Figure 2.1: Colour Clusters

These connections are treated as a probabilistic possibility for a communication between the nodes. However, presence of such a connection does not ensure that a communication takes place. The actual communication depends on more factors and more conditions need to be satisfied for that. These conditions are explained in detail in the next chapter 3.

2.1.3 Presentation Approach

Due to the large number of agents, a clear and easy way to understand presentation of the simulation was imperative. Initially, the author went with 2-D histograms depicting the membership values of each color group, which accurately presented how the members move around as the height of the bars representing the number of members in a group could be easily seen to rise or fall, and the speed of this descent or ascent would tell about the rate of success or failure for a particular color. An example could be seen in figure 2.1.

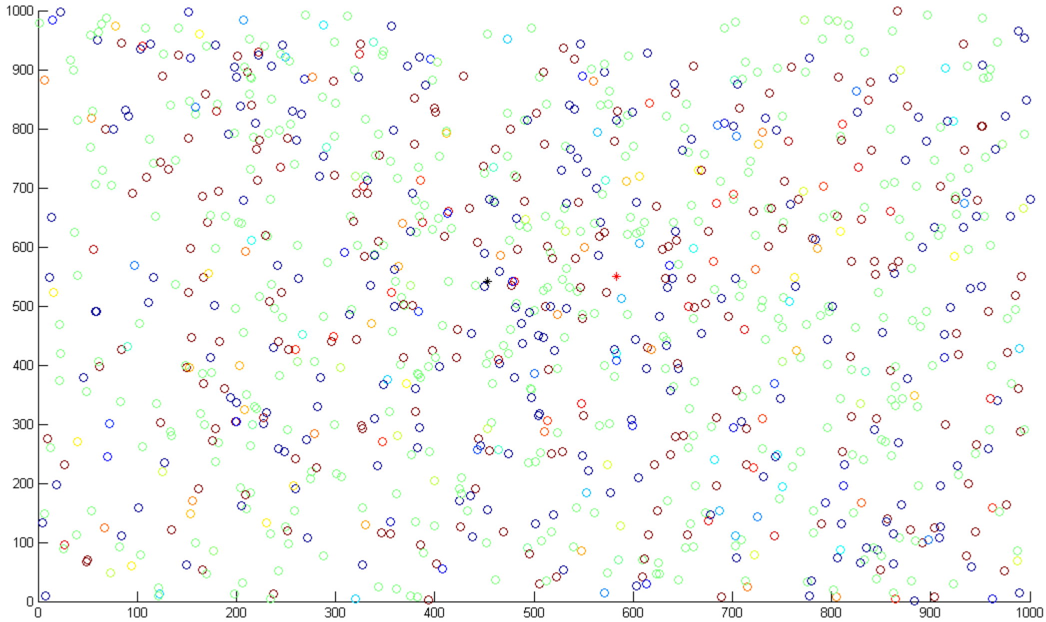


FIGURE 2.2: 3-D Spatial Cluster

However, with the progress of the model this representation became obsolete as it was unable to represent the aspect of "Cost" to a company. To address this necessity, the author decided to switch to a 3-Dimensional presentation, with X- and Y- axes representing generic movement for agents, and the movement or effort as well as incurred cost for the companies, and the color, being the third dimension of the presentation, depicting the value of the attribute. The speed of movement here serves the purpose as the speed of rise/fall in the previous model. An example of this is figure 2.2.

While the X-Y movement in this model, for the agents, is a correlation, the changing colors as they move represent the agents leaning towards or away from a particular company.

2.1.4 Company Policies

Different companies may have different policies towards customer involvement. These decisions are internal matters for them. However, Ives & Olson [20] broadly divide the different approaches in six categories. These categories are :

1 No involvement. Users are unwilling or not invited to participate.

2 Symbolic involvement. User input is requested but ignored.

3 Involvement by advice. User advice is solicited through interviews or questionnaires.

4 Involvement by weak control. Users have sign-off responsibility at each stage of the system development process.

5 Involvement by doing. A user as design team member or as the official liaison with the information system's development group.

6 Involvement by strong control. Users may pay directly for new development output from their own budget or the user's overall organizational performance evaluation is dependent on the outcome of the development effort.

The customer involvement continuum followed by the author follows the above categorisation and is influenced by the works of Bodil Sanden [21], which argues that *interaction is essential to customer involvement, specially in development of new sevices.*

Out of these, this thesis focuses on categories 2 and 5. These two specific categories were chosen because they represent most dissimilar choices while still avoiding the extreme conditions, which allows the model to simulate the more probable scenarios.

2.1.5 Simulation Approach

The author surmised to focus on a competitive diffusion where the companies compete over a single attribute of the agent. The attribute *color* was chosen for the simulation as it is easy to understand and visualise, and it enabled the author to present the effect in a 2-dimensional spatial model where clustering is not directly dependent on the axial values. As a limiting case, the author simulated only two companies. This limits the scope as it does not give much insight into cases where two companies might collaborate for mutual benefit, like to win against a third rival company. This model makes use of random functions wherever possible and applicable, and uses this to simulate the effect of 'chance'. This inherent randomness makes the system non-deterministic and more reliable for the purpose of simulating more naturalistic occurrences.

The simulation focuses on companies attracting agents towards their cause, and the cost incurred in terms of efforts on the part of the company. The model witnesses three types of potential interactions :

1 Company to Agent interaction

These interactions are predefined, in a sense that a certain number of agents are already seeded to believe in a company at the beginning of the simulation. This is explained further in Chapter 3. This helps in creating an established base of users for the company. During the initial seeding, both companies are seeded with equal number of agents randomly chosen from the whole system and it is taken care of that there is no overlap in this initial customer base, i.e., all the agents being seeded are different.

2 Agent to Agent interactions

These interactions are the means by which one agent *may influence* another to lean towards a company. In this model, the agents belonging to either of the companies in question are referred to as *influencing agents*, and the others are referred to as *influenced agents*. This also mimics real life as the connections betweens agents are directional. The *influencing agent* tries to motivate the *influenced agent* to move towards it's parent company, but the success or failure is decided based on a certain threshold value. This value depicts traits of the *influencing agent* in question and the threshold is decided randomly, depicting traits of the *influenced agent* and external factors such as luck and chance.

3 Agent to Company interactions

Whenever an *influenced agent* is successfully influenced, during the simulation, this also causes the company to make some effort, the amount of which depends on its policies. These interactions are what cause the cost to the company.

It is to be noted here that the *influenced agent* in question need not be a new addition to the customer base of the company. It may already be a customer, however, further successful influence means that it's loyalty towards the company is increased, which is denoted by a decrease in the distance between the agent and the company.

The outcome of these interactions are measured by the following two values:

a Gain

This refers to a company successfully influencing an agent to lean towards itself. For the purpose of this thesis, an incremental counter is used for every successful influence by an *influencing agent*.

For better results, a weighted function using the cumulative distance of the agents from the companies could be used. However, the same is beyond the scope of this thesis.

b Cost

This refers to the cost incurred to the company and is a function of the effort made by the company. It depends on the policy being followed by the company.

In this thesis, the cost is represented by the work being done by the company in every step and is an abstract concept, without the notion of any budget. For a deeper analysis, a function of the total displacement of the company in every time step with relation to the initial position may be used.

2.2 Tools

As a requirement from the University, Matlab was chosen as the programming language, and no special toolboxes were used. The curve fitting application was used occasionally to check the output of the simulation.

The formulas, and variants thereof, used by the author were sometimes first tested as a prototype in Python with NetworkX and Mathematica for mathematical validation. However, any of those implementations do not directly contribute to the result.

Chapter 3

Methodology

On an abstract level, the model consists of an environment comprising of numerous agents, and a few companies. This thesis specifically focuses on the various strategies used by the companies to attract the agents, and try to measure its efficiency in terms of the related costs and benefits. Based on these measurements, a company could determine the optimal strategy to maximize its efficiency for the targeted section or sub-subsection of agents.

3.1 The Companies

The companies only exist in a superficial form in the model. In this model, each company plays its role by providing a central point for the agents to cluster about, and the movement of this point itself denotes effort made by the company, which in turn incurs a cost to the company.

The author decided to keep this model limited to two companies competing for the same spot, i.e., any agent cannot totally belong to both companies at the same time. The primary reasons for this were limitations of time and computation power, in that order.

3.2 The Agents

The agents in this model are simpler constructs, each having a "color" attribute, and this attribute changes as the agents tend to believe in either of the companies.

Structurally, each agent has following attributes :

1. Id
2. x-position coordinate
3. y-position coordinate
4. Colour value
5. Influence value

During its lifetime, an agent has constant value for its Id, but the rest of its attributes may change. The influence value of the agent is determined as a function of the number of connections it makes, and remains constant once the whole network is created. All other attributes, however, may change throughout the simulation.

3.3 Connections

The connections between the agents form in such a manner that the model satisfies the requirements of free-scaling. Following in the footsteps of the giants, the implementation is based on algorithm 1, which is a modified form of the B-A algorithm.

3.3.1 The World

The model is first built based on two fixed parameters, and then the simulation runs on it. The author starts building the model for a fixed number of total agents N, and predefined average connectivity d in the environment. The initialization of the model begins with the creation of a small, fully connected network, and then, with time, new agents are

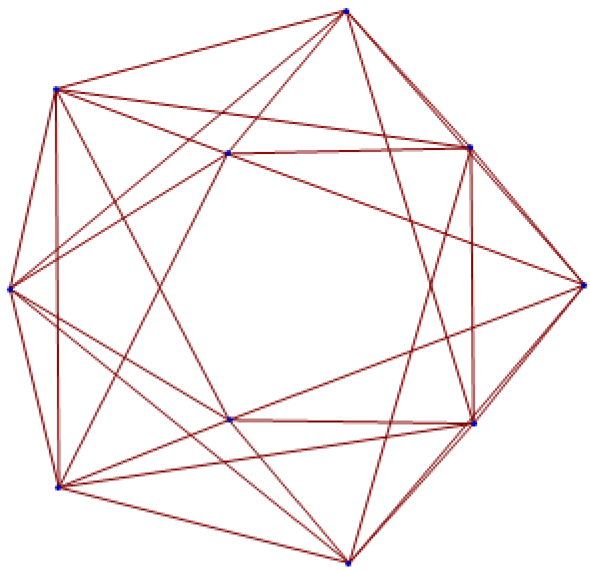


FIGURE 3.1: Sample of Initial network with 10 nodes

added to this network, and these agents form the connections following the rules of PA. The exact method of forming this connection was modified by the author, so as to further trick the environment. This allowed simulation of scenarios where it is possible to form a number of connections, but only a fraction of those connections are actually formed. This was achieved by employing a 'weight' modifier on the probability of an agent to form a connection. This weight, in turn, is defined for every possible pair of agents, and is based on the difference in the value of a special attribute "like-mindedness", defined for each agent in an uniform, pseudo-random manner. This attribute is not considered an essential attribute because it may or may not be used, in various simulations. However, as an effect of using this modifier, the number of connections for an agent comes down considerably, while the nature of the connections still remain same, and the network remains scale free.

The concept of starting with a smaller network and growing it is explained in detail in Chapter 2 in section 2.1.1 and examples are shown in figures 3.1 and 3.2.

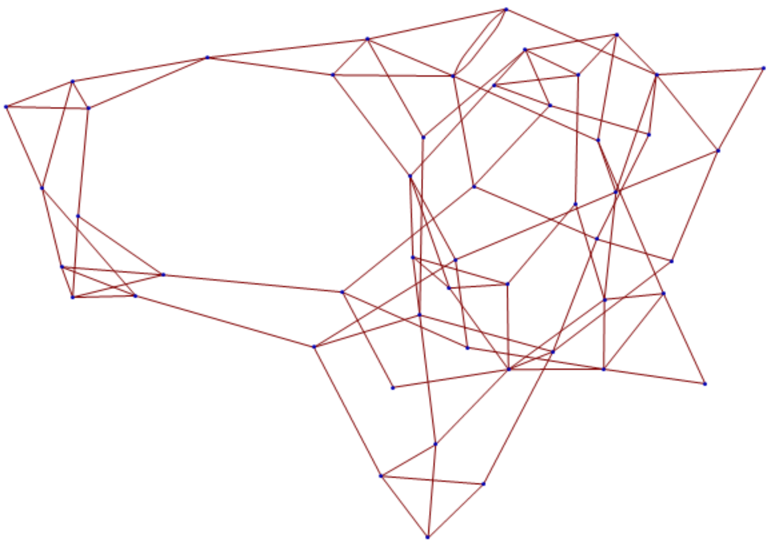


FIGURE 3.2: Sample of Final network with 50 nodes

3.3.2 The Simulation

The simulation begins with allocating "influence" value to each existing agent, as a function of the number of connections it has in the network. This makes sure that the most-connected agents get a chance to be more influential. After this, values are uniformly assigned to all agents for the "color" attribute, and then they are seeded to belong to different companies. 20 % was the value chosen by the author for initial amount of seeding, for the following combination of reasons:

1. It allows for equal amount of seeding in both companies.
2. It allows for double amount of seeding in a neutral company.
3. After above mentioned allocations, it leaves equal amount of distribution randomly, to simulate chance.

The above rules give us the formula

$$
\begin{aligned}
company_1 + company_2 + company_{neutral} + company_{random} &= 100 \\
\Rightarrow x + x + 2x + x &= 100\% \\
\Rightarrow x &= 20\%
\end{aligned} \tag{3.1}
$$

Then, the simulation is run, where in each time step, an edge is selected randomly, and an exchange could happen if the *FromNode* of this edge belongs to one of the two companies. Here, "belonging to company" refers to the value of the "color" attribute for the agent. After this, the influence value of the *FromNode* is subjected to a randomly selected threshold, and upon satisfaction, the actual communication occurs. The threshold is lowered depending on a company's efforts towards attracting the customer.

The selection of the edge depicts existence of a connection, while satisfaction of the threshold depicts the actual communication. This constraint was included to make the simulation model even closer to real-life. Hence, a higher effort would result in more cost, but would give a company higher chances of a successful communication.

For every successful communication, the agents shift their "color" attribute a fraction towards the influencing company. This shift also translates to their x-y coordinate, which is shown in the presentation. The important aspect here is that these changes in color and x-y coordinates, while not always causally related, are always correlated.

For the companies, however, it has different significance. Whenever an agent has moved closer, the companies also make an effort and move some fraction towards the agents. The amount of distance moved by the company represents the effort it makes, and is determined by the company policy. These movements also incurs the cost to the company, and a combined function of the gain and this cost determines the effectiveness of the policy.

Towards the end of the campaigning, a company may increase their effort to maximise chances of gain. This is achieved in the model by doubling the shift amount whenever a certain amount has been spent on the campaigning.

In the end, the gain and cost are calculated for the companies using the equation 3.3. Since cost is already lesser than the gain, this function was kept linear.

$$function = gain - cost \tag{3.2}$$

$$\tag{3.3}$$

3.4 Definitions

1. *Fully Connected*

   It follows the standard definition that each node is connected to every other node. The initial network is formed in such a manner that it is fully connected and satisfies the average number of total connections.

2. *Like-mindedness*

   It is a special attribute of the agents, that is used in this model to modify the probability of formation of a new connection between two agents. It simulates the real life phenomenon that people with similar interests may be more likely to form a connection.

   This subject is very vast in itself, and covers a wide ground [22, 23]. This is also helpful in showing the aspect that some people may be introverts or extroverts. However, that aspect is not simulated in detail in this model.

3. *Weight*

   The *weight* is the difference in *like-mindedness* of two agents. It is calculated as

$$weight_{j,k} = abs(likemindedness_j - likemindedness_k) \tag{3.4}$$

4. *Company Centre*

   The company is not implemented as a structural entity, rather, the company centre is used to keep track of the attitude of the agents towards the company, and also to track the effort and cost on the part of the company.

5. *Influence*

   This attribute determines the chances of an agent to influence another agent. As in real life, we allow more connected agents a chance to be more influential, but ultimately, there is no deterministic end to the model, and whether or not an actual communication takes place and an agent gets influenced is always determined by a randomly determined threshold. This is calculated as a function, described by equation 3.5, of the number of connections the agent makes.

$$influence_i = \frac{connections_i}{connections_{max}} \tag{3.5}$$

6. *Shift*

   This shift represents a change in the state of the overall system. For an agent, a shift in its color attribute represents the agent leaning towards a company. This shift is correlated with the same ratio of change in the agent's x-y coordinate, used for visualization purposes. For a company, the shift only occurs in its x-y coordinate, which is caused by the x-y shift of the agent, and this change represents two things :

   [a] The speed and frequency of movement represent the effort taken by the company

   [b] The distance moved represent the cost incurred to the company

3.5 Algorithms

Algorithm 1 Create Scale-Free Network

N : Input, total number of agents
D : Input, average number of connections
if $D < N + 1$ **then**
  Form a *fully connected* network with D+1 agents
end if
implement CG
$i \leftarrow D + 1$
while $i < N$ **do**
  $Agent_i \leftarrow newAgent$
  assign $likemindedness$
  implement PA
  $k \leftarrow 0$
  while $k \leq \frac{D}{2}$ **do**
    calculate $probability_{basic}$ based on equation 2.1
    compute $weight$ based on equation 3.4
    $probability_{attachment} \leftarrow probability_{basic} + weight$
    $Agent_i$ forms $connection_k$, based on $probability_{attachment}$
  end while
end while

Algorithm 2 Simulation

assign $centres_{company}$
assign $effort_{company}$
calculate $influence$ based on equation 3.5
assign *color* to all agents
seed 20% agents to $color1$
seed 20% agents to $color50$
seed 40% agents to $color25$
seed 2*most influential* agents to different companies
J : number of iterations
$i \leftarrow 0$
while $i \leq J$ **do**
  select $edge_i$
  if $FromNode_{color}$ = *color1* OR *color50* **then**
    select *threshold*
    threshold = *threshold* - $effort_{company}$
    if $FromNode_{influence} \geq$ *threshold* **then**
      $ToNode_{color} \leftarrow ToNode_{color} + shift$
      $ToNode_{x-coordinate} \leftarrow ToNode_{x-coordinate} + shift$
      $ToNode_{y-coordinate} \leftarrow ToNode_{y-coordinate} + shift$
      $Company_{x-coordinate} \leftarrow Company_{x-coordinate} + shift$
      $Company_{y-coordinate} \leftarrow Company_{y-coordinate} + shift$
    end if
  end if
end while

Chapter 4

Experimental Results

4.1 The Model

One of the most common forms of visually representing data is by using a two-dimensional graphical plot. In case of scale-free networks, the horizontal axis represents number of connections, and the vertical axis represents the number of agents having that many connections. This form is used by the author because of it's simplicity. It is to be noted here that the number on either of the axes are not to be confused with the Id of the agents.

Figures 4.1 and 4.2 show the different degrees in the connection in the case of using original BA algorithm and using our modified approach to the algorithm respectively.

It can be observed that using the constraint of 'like-mindedness' results in a drop in the horizontal axis values, which shows that the maximum number of connections being formed for any node is lowered.

The basic model was modified a lot with time, from the initial design, as the author learnt more. The development of the model can be explained in two parts.

4.1.1 Small-World Networks

Figures 4.1 and 4.2 both show discontinuation or a sudden interruption in the tail.

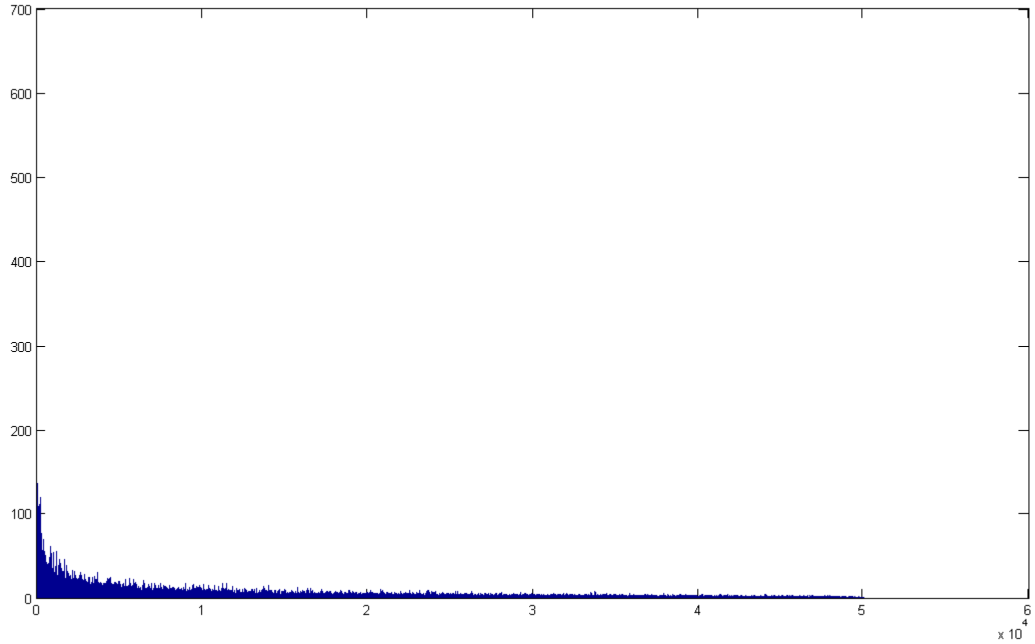


FIGURE 4.1: Small world Network, Original BA algorithm

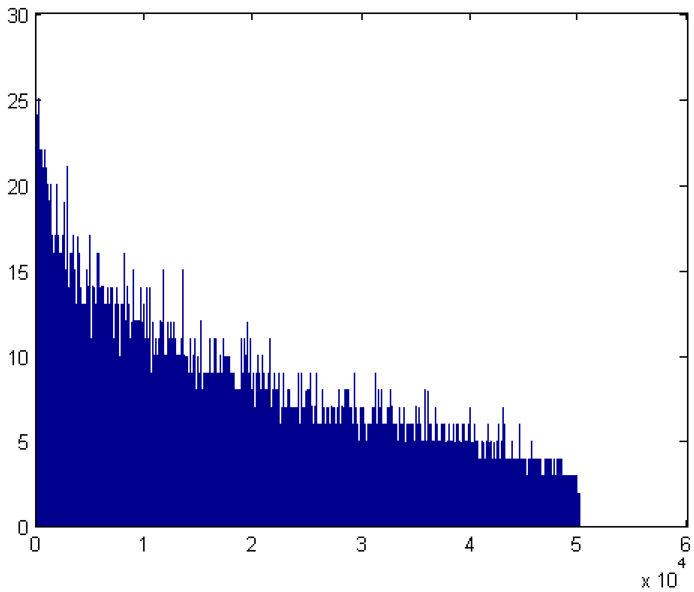


FIGURE 4.2: Small world Network, Modified BA algorithm

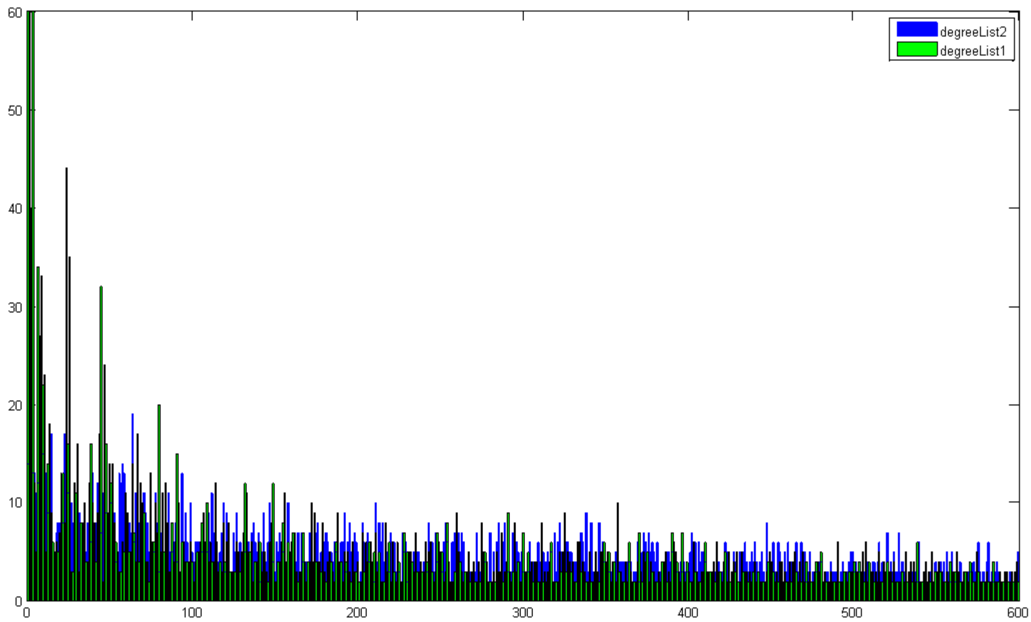


FIGURE 4.3: Comparison of Scale Free Networks generated by the different algorithms for 1000 agents

On further inspection, it was found that the networks were displaying characteristics of Small-World networks [24]. The difference in what was required and what was achieved is made clear by Cohen & Havlin [25].

This led to modifications in the design and the design explained in Chapter 3 was achieved.

4.1.2 Scale-free Networks

An example in comparison of both approaches, i.e., following original BA algorithm and the modified algorithm, can be seen in Figure 4.3, where green color represents the original algorithm and blue represents the modified approach.

These outputs were further tested for fit, using MATLAB's curve fitting toolbox, against a power-law curve, and the results are shown in figure 4.4 and 4.5.

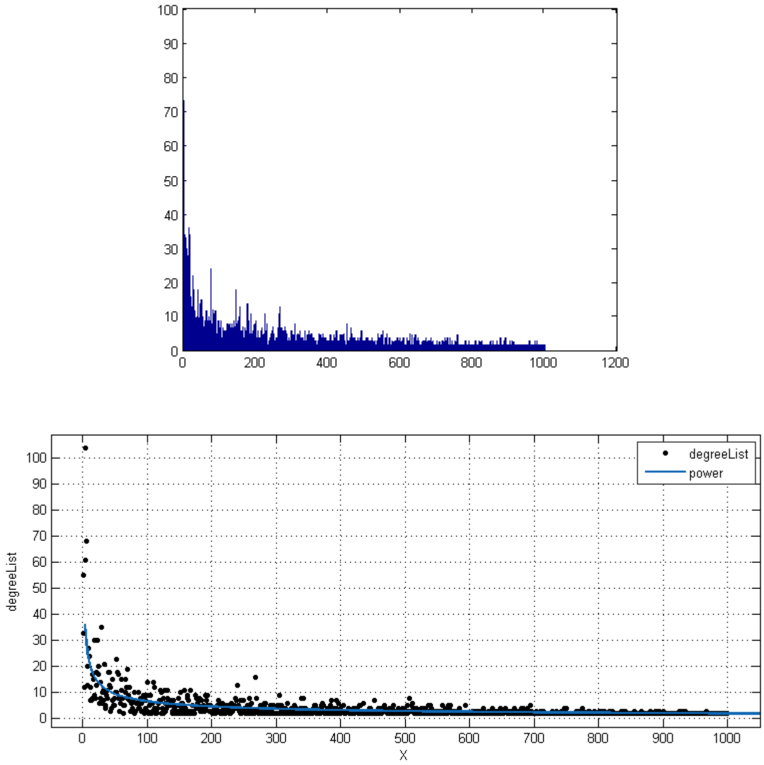


FIGURE 4.4: Scale-Free network using BA algorithm and corresponding power law fit

4.2 The Simulation

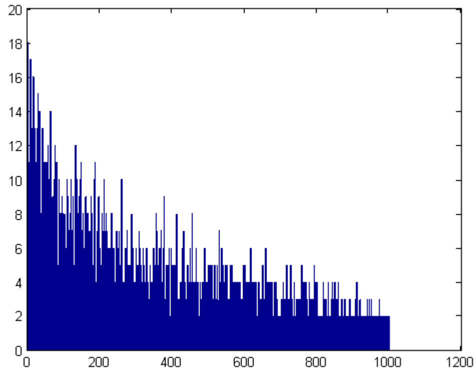
The simulations resulted in three different types of outputs. They were all run for cases where on an average, every agent makes 3 connections. The different types of outputs are as follows:

1 Colour Histograms

2 Spatial Clusters

3 Cost-Gain Functions

For the purposes of this thesis, specifically for the analysis part, a particular set of 'Gain' and 'Cost' values, in accordance with the policies followed by the companies, were used. Keeping the gain for each influence fixed at 'One' unit, and allowing the agents to move one-fifth (0.2 parts) of the distance at a single step, the costs are varied as per table 4.1.



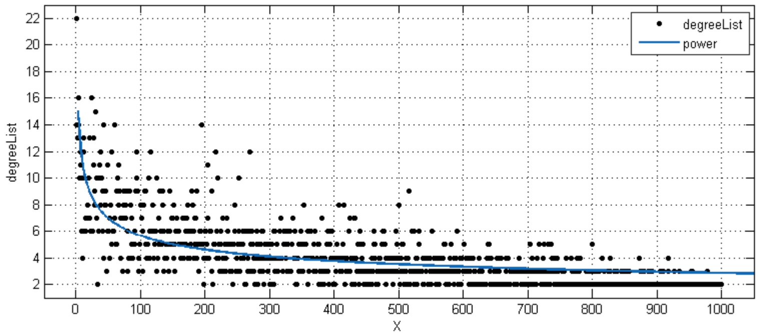


FIGURE 4.5: Scale-Free network using modified BA algorithm and corresponding power law fit

It should be noted that this distance changes with each step, and for the next step, the new distance is used for calculations.

The costs are also equal to the effort, i.e., the amount of distance travelled by the company towards the agent.

It is to be noted here that during the simulations, the costs in the last 6 cases mentioned above were randomly assigned to the companies, i.e., *cost1* and *cost2* may be interchanged. This accommodates for the reverse ordering of the columns without the need of including extra cases.

4.2.1 Colour Histograms

These histograms give the observer a good view of how the agents tend to lean from one company towards another.

caseId	cost1	cost2
Base1	0.001	0.001
Base2	0.002	0.002
Base3	0.01	0.01
Base4	0.02	0.02
Diff1	0.001	0.002
Diff2	0.001	0.01
Diff3	0.001	0.02
Diff4	0.002	0.01
Diff5	0.002	0.02
Diff6	0.01	0.02

TABLE 4.1: Costs for various cases for both the companies

Since the network has a large concentration of agents initially seeded to a neutral color which is equidistant from both *company1* and *company2*, it could be observed that a huge initial spike is in the center, while the two companies in question are at the extreme corners.

The change in the heights of different columns can be seen in fig. 4.6. The horizontal axis represents the color value, and the vertical axis represents the number of agents having that value for the color attribute, or, in other terms, the membership of that company.

Fig. 4.7 shows a similar output for 10000 agents. This case shows that although the agents move away from the central, equidistant clusters, they may not always reach all the way to either of the company clusters. This is evident by the change of height of the intermediate columns.

4.2.2 Spatial Cluster

As mentioned in chapter 2, section **??**, these outputs were necessary to visualise the cost incurred to the companies.

The three dimensions of any element of these plots are its x-y coordinates and its color. The agents are visible as ‘o’ markers, while the companies are shown with ‘*’ markers.

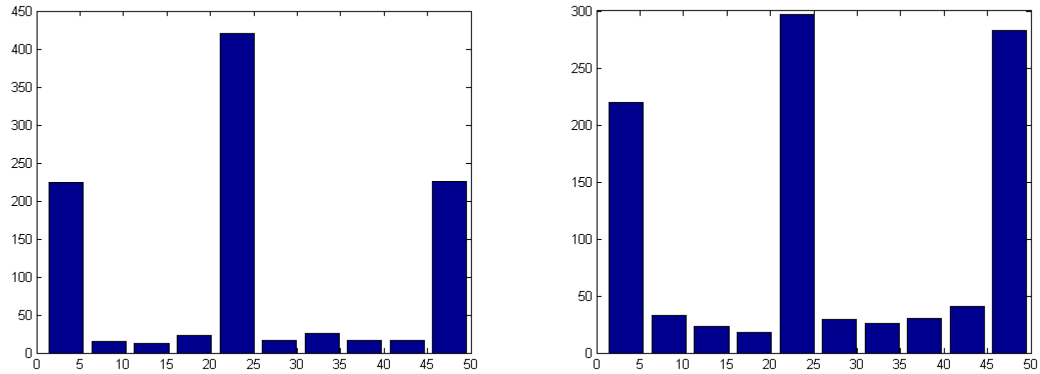


FIGURE 4.6: Colour Clusters for 1000 agents. Initial Cluster on left, and final clusters on right.

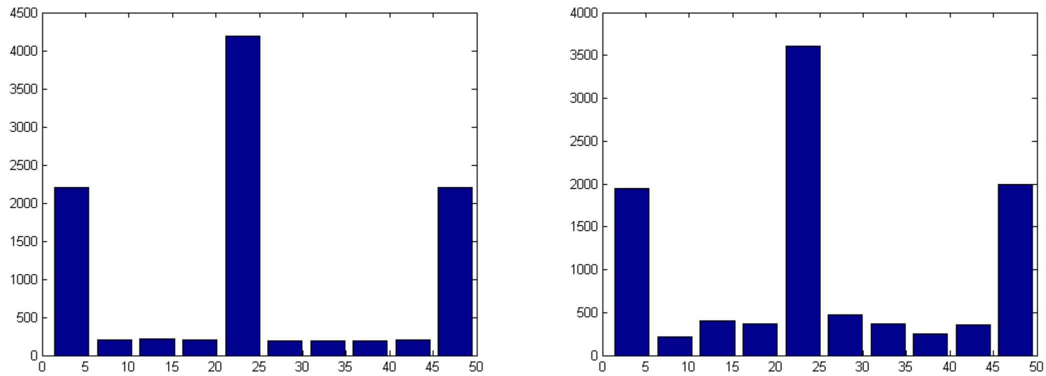


FIGURE 4.7: Colour Clusters for 10000 agents. Initial Cluster on left, and final clusters on right.

The simulation allows the observer to see the agents change their color as they move towards a specific company. Some screen shots of the simulations can be seen in fig. 4.8 and 4.9.

While fig. 4.8 is easier on the human-eye, fig. 4.9 is a result of a lot more iterations and is richer in terms of analysis.

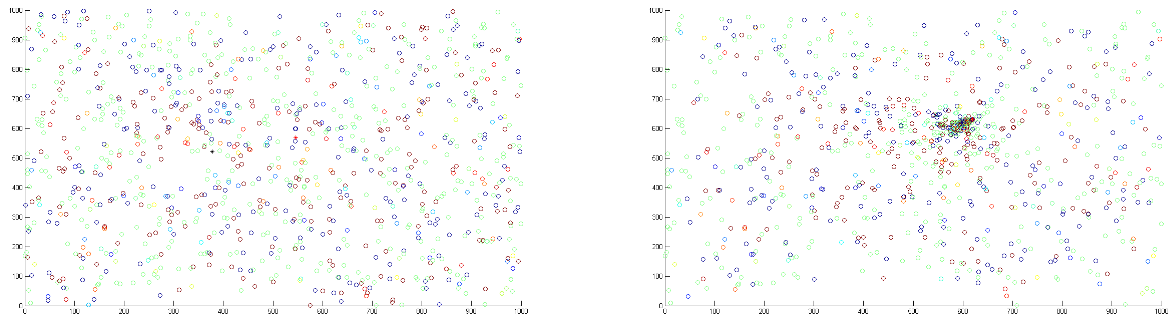


FIGURE 4.8: Spatial Clusters for 1000 agents. Initial Cluster on left, and final clusters on right.

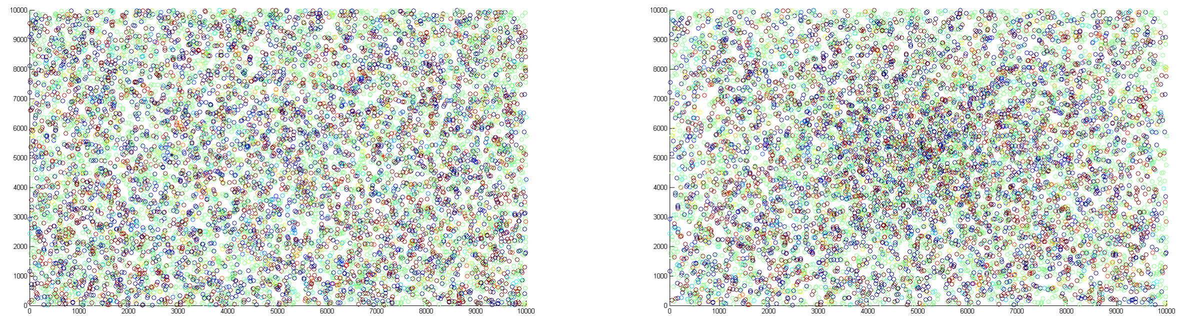


FIGURE 4.9: Spatial Clusters for 10000 agents. Initial Cluster on left, and final clusters on right.

4.2.3 Cost-Gain Function

The simulation keeps track of the cost and gain for each company according to table 4.1. *Company1* follows *cost1* while *company2* follows *cost2*, and the *caseId* tells us about the relation in the policies.

The generic spread and the population count of the companies was monitored, and they tend to show exponential nature of change.

Fig. 4.10 and 4.11 show the increase and decrease of the membership of the companies against exponential fit. The x-axis represents time steps while the y-axis shows the membership value of a cluster, representing the number of agents fully believing in the company.

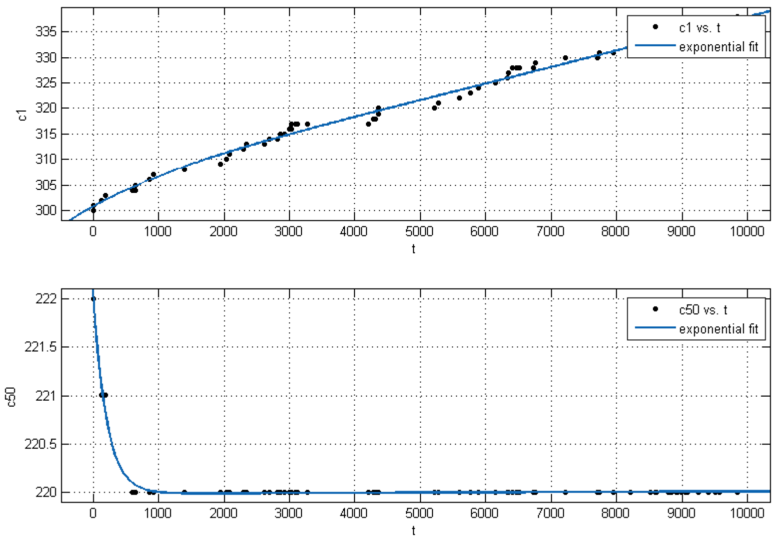


FIGURE 4.10: Changes in membership of company for 1000 agents

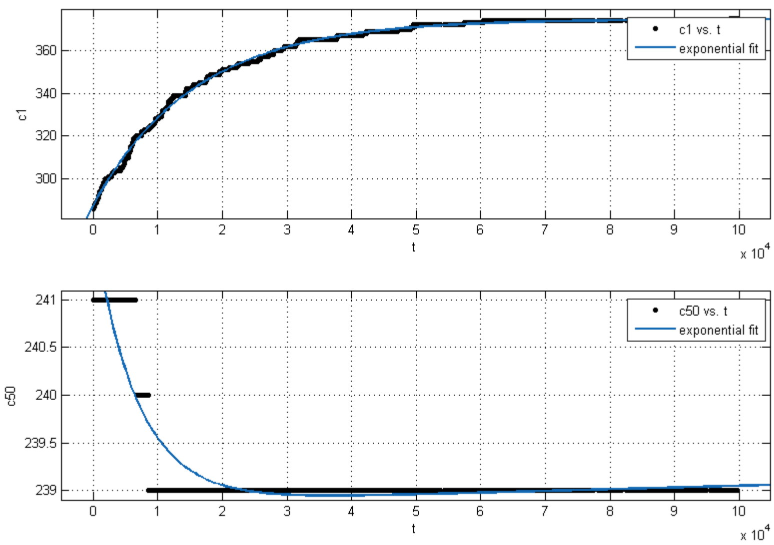


FIGURE 4.11: Changes in membership of company for 10000 agents

The cases from table 4.1 show the functions in the form of equation 3.3. The graphs presented here demonstrate the overall gain for *Company2* relative to *Company1*, changing with time steps along the x-axis. These results can be divided into two categories. These are :

1 *Base* cases

These cases see both companies taking equal cost, and are meant to give the observer an idea about behaviour of both companies under similar policies. The output of these cases can be seen in fig. 4.12, 4.13, 4.14, and 4.15.

2 *Diff* cases These cases witness the companies following different policies, and allows the observer to analyse the policies. In this approach, cost for *Company1* is kept at a constant while the cost for *Company2* is increased, and the observed results are recorded. The output of these cases can be seen in fig. 4.16, 4.17, 4.18, 4.19, 4.20, and 4.21.

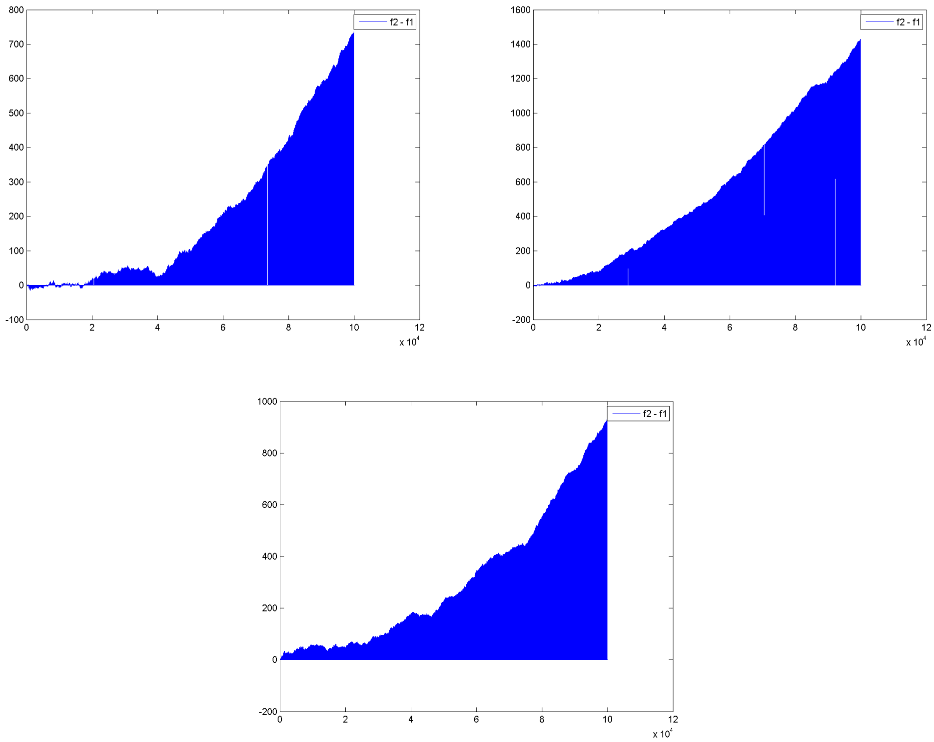


FIGURE 4.12: Results from three simulations for base case 1.

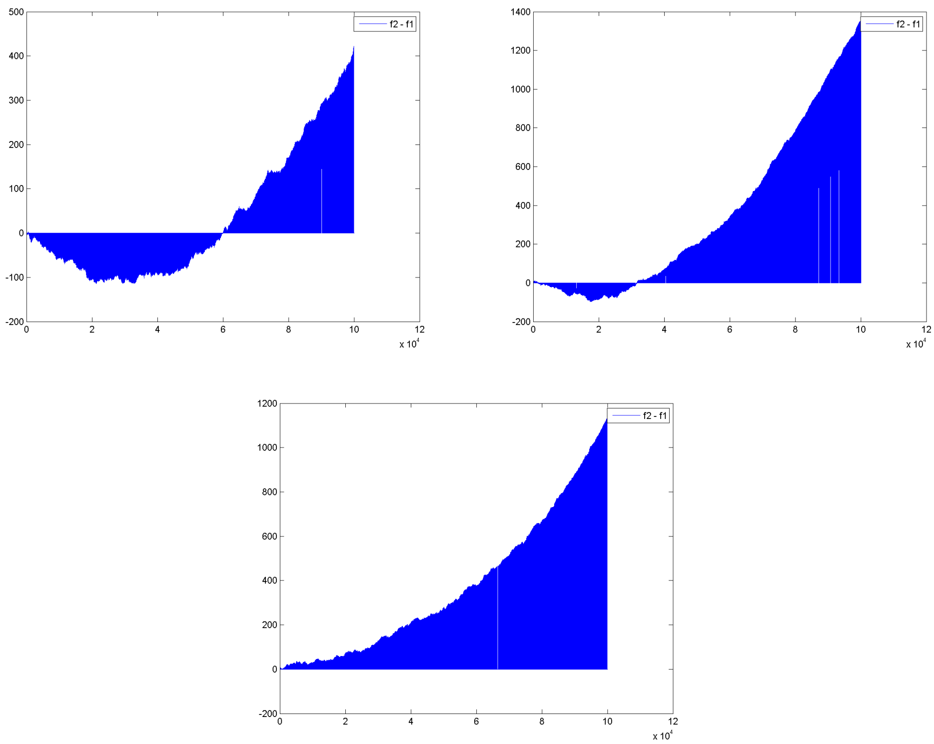


FIGURE 4.13: Results from three simulations for base case 2.

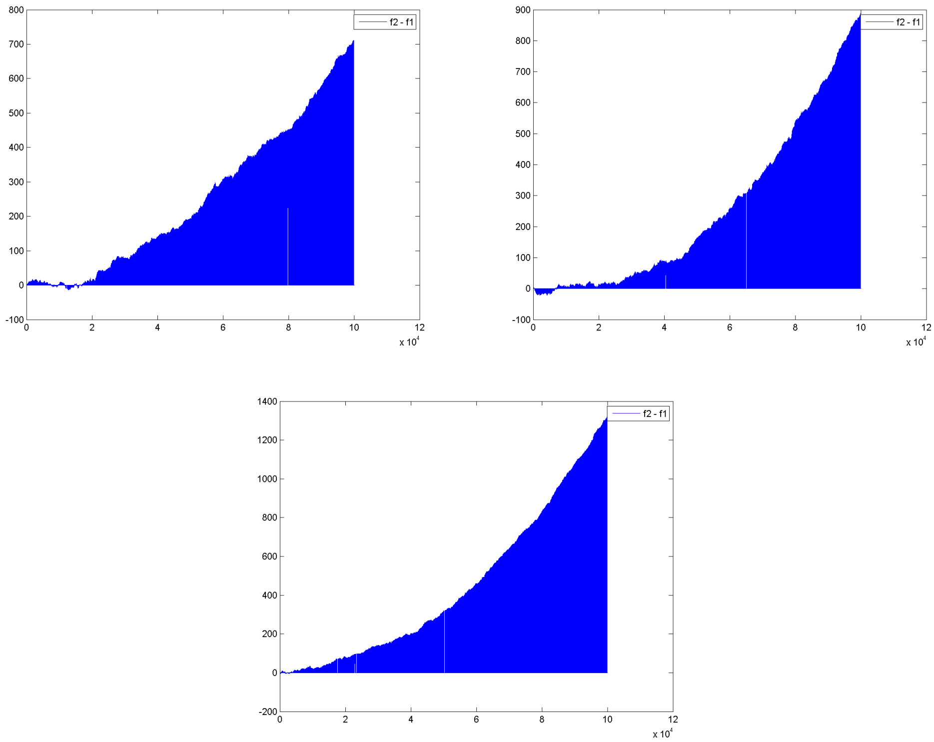


FIGURE 4.14: Results from three simulations for base case 3.

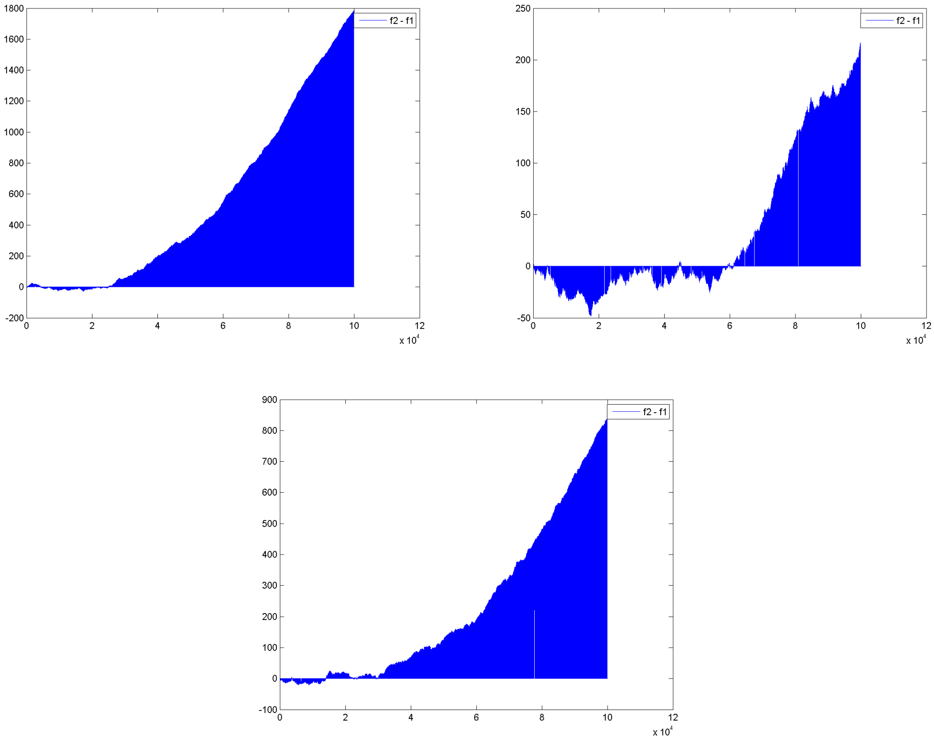


FIGURE 4.15: Results from three simulations for base case 4.

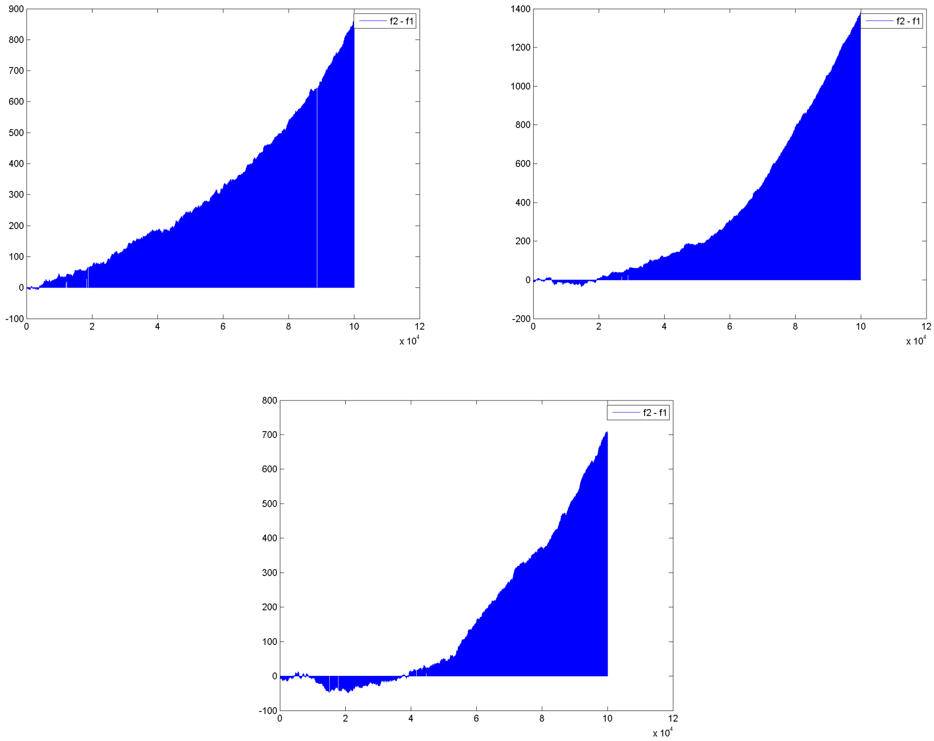


FIGURE 4.16: Results from three simulations for difference case 1.

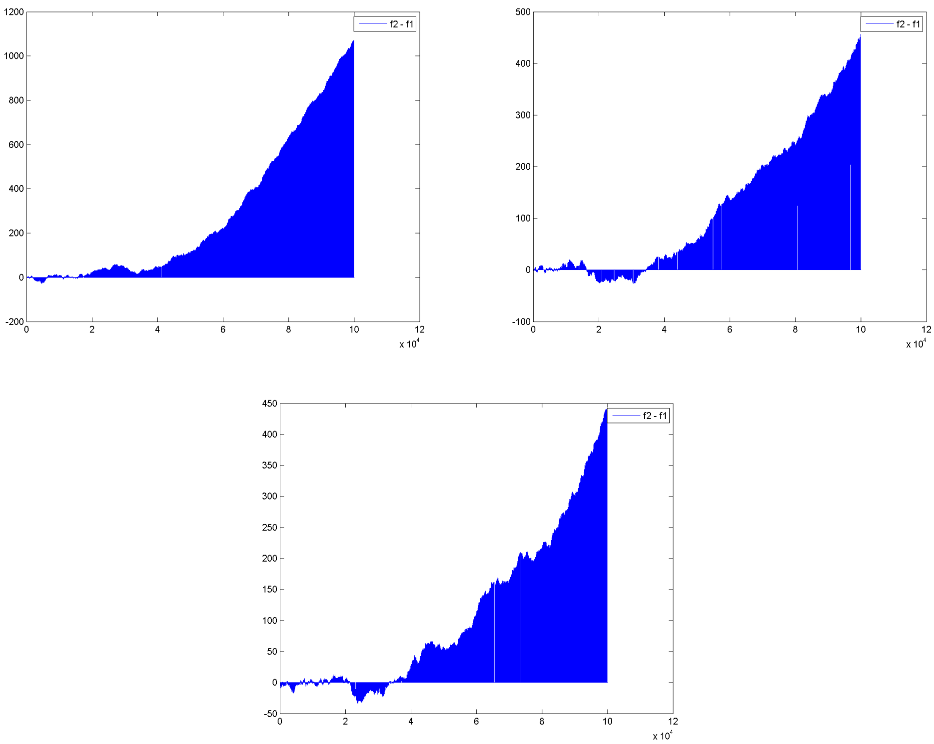


FIGURE 4.17: Results from three simulations for difference case 2.

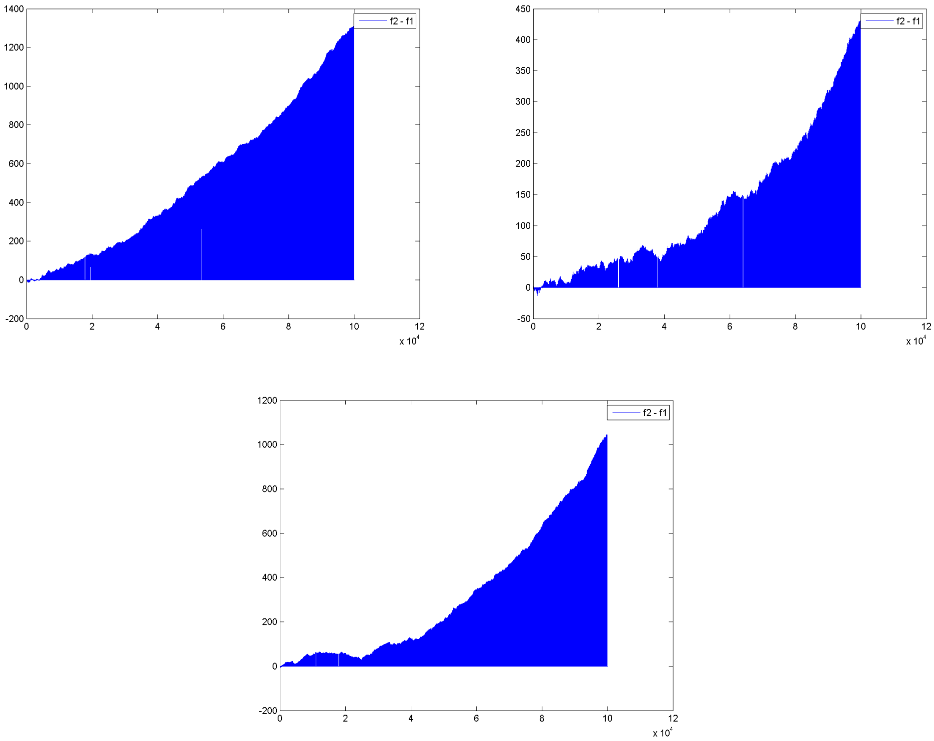


FIGURE 4.18: Results from three simulations for difference case 3.

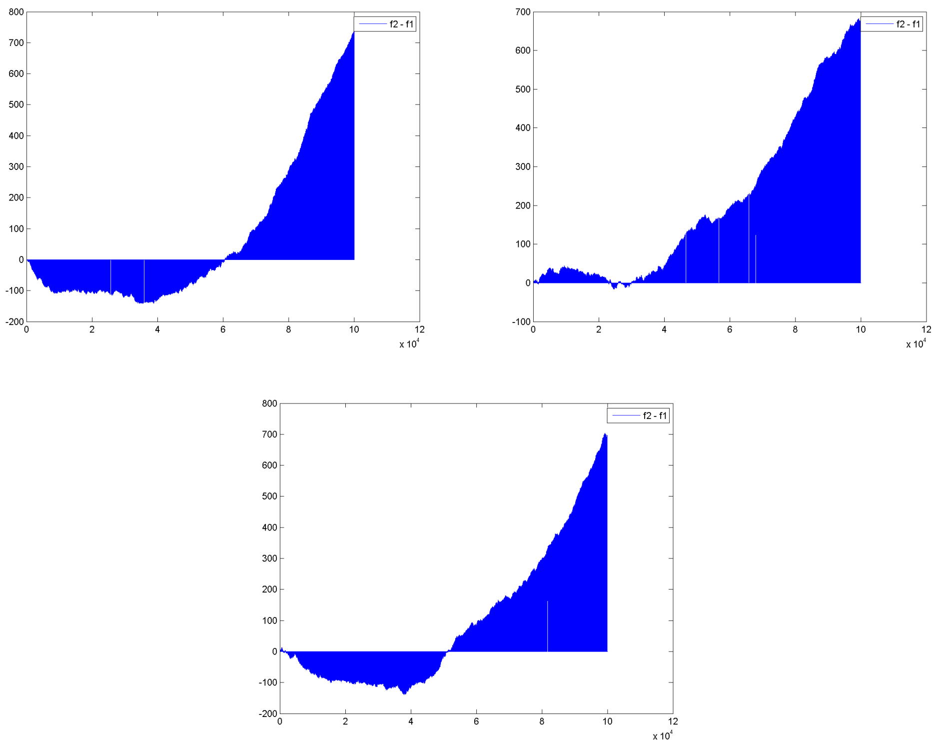
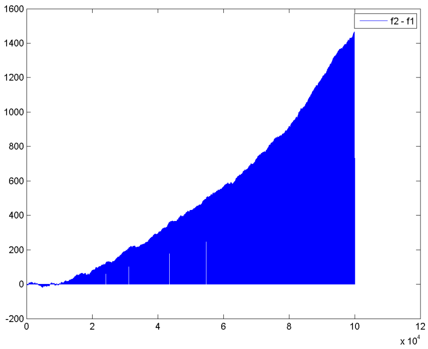
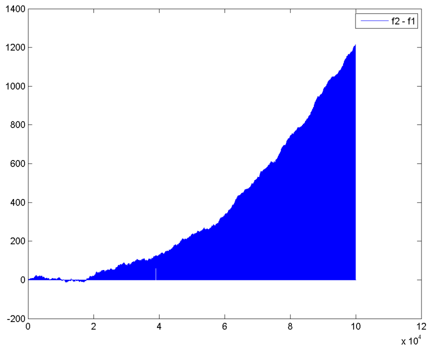


FIGURE 4.19: Results from three simulations for difference case 4.





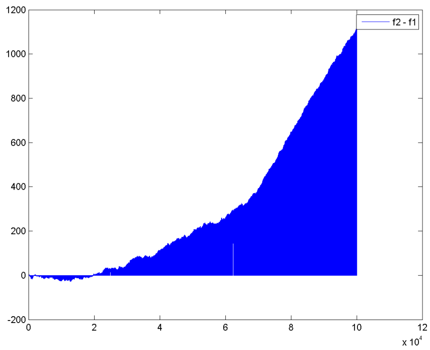


FIGURE 4.20: Results from three simulations for difference case 5.

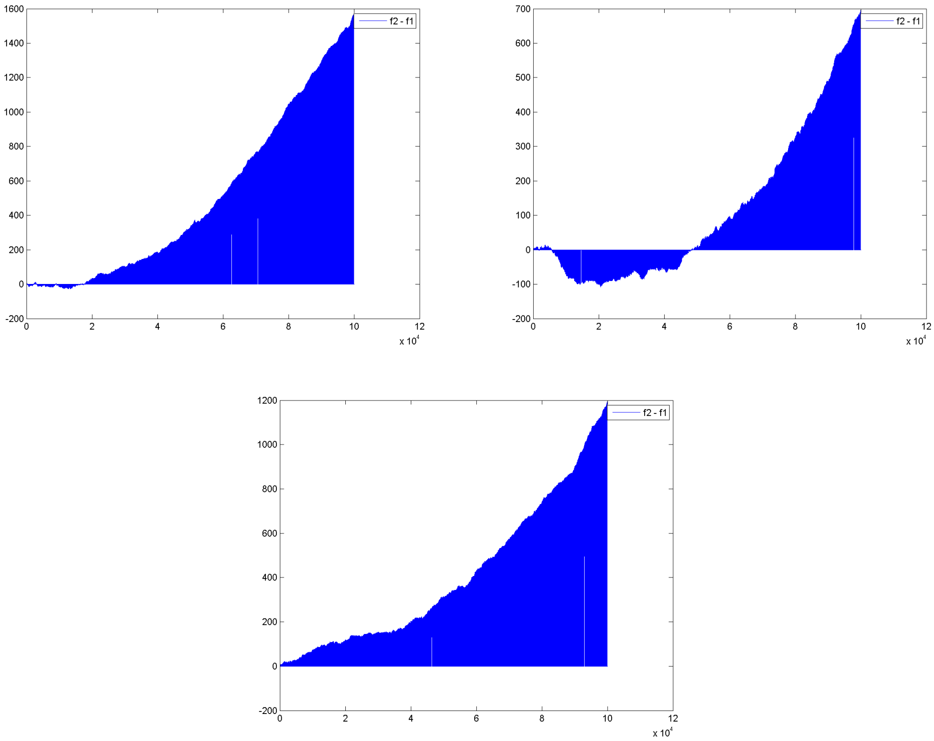


FIGURE 4.21: Results from three simulations for difference case 6.

Chapter 5

Significance of Results

5.1 The Model

The model shows us a fast and easy way to generate a scale-free network, and the effects of using the constraint of 'like-mindedness', defined in chapter 3 section 3.4, between the agents, which reduces the number of connections formed significantly, while the network still follows power law.

Although this thesis is focused on companies, the modularity of the project allows the algorithm 1 to be used for generic simulation of social structures.

5.2 The Simulation

The initial results for diffusion in a scale-free network present a picture where it is clear that the spread, whether it is increase or decrease, tends to be exponential.

The simulation results presented in chapter 4 section 4.2.3 give a varied picture. They are divided into two groups.

5.2.1 Base Group

This group focuses on analysis while cost is same for both companies, and helps the observer to understand the basic nature of the companies.

The base group analyses show that *company2* seems to emerge victor for overall simulation, however, the nature is unstable for shorter runs, i.e., during the initial time steps of the simulations.

The results also suggest that increasing the cost results in more relative profit for *Company1*. It can be seen that the area under the curve suggesting the net relative gain for *Company1* tends to increase as the cost is increased.

5.2.2 Difference Group

This group focuses on analysis while cost is different for both companies, and allows for analysis of the different policies.

The author focused on two types of policies described in chapter 2 section 2.1.4. Specifically, the chosen policies were :

Company1 : Symbolic involvement. User input is requested but ignored.

Company2 : Involvement by doing. A user as design team member or as the official liaison with the information system's development group.

Company1 maintained a minimal involvement of customers in decision making, and made very little effort. This made sure that the company did not take a very high cost.

Company2 had a more open policy where customers were highly involved, and the company made a lot of significant efforts to reach out. This resulted in formation of long-lasting relations which led to profit in the long run.

The analyses show that *company1*, with its minimal inclusion policy, is more effective for short-term projects, when it makes less efforts towards the customers. The policy followed by *Company2* is less profitable during this phase, but it compensates later on and becomes more profitable than *Company1*.

The exact shift point for this change is case dependant, and it varies with the value of gain, and cost. But this model could be further developed to prepare prediction system that helps companies to switch between policies in order to maximize their cost-benefit ratio.

Chapter 6

Extending the Analysis

The present system focuses on only two companies, competing over a single attribute of the agent. It is limited to a specific part of a very wide spectra of possibilities.

In future, there are possibilities to extend this system and analyse a diversity of scenarios.

6.1 Extending the analysis to more companies

The environment could always include more companies competing for the same slot. This might provide valuable insight, specially for cases where two companies might collaborate for mutual benefit against a common competing company.

6.2 Extending the analysis to more attributes

The companies in the present model only compete for a single attribute, but this may be extended to make them compete for multiple attributes of the agents. However, this will significantly increase the complexity if the model.

6.3 Extending the analysis to collaborative diffusion over single attribute or multiple attributes

While the presented system only lets the companies compete against each other, it might be interesting to see how they behave when they are collaborating towards a common goal. This could be done for a single attribute as well as for multiple attributes.

6.4 Extending to Other Systems

The present system is targeted towards company policies, but the generic nature and social structure of the model imparts flexibility to the system, which makes it easy to extend to other systems as well.

Chapter 7

Summary and Conclusion

Technology is ever evolving and the nature of this evolution has an inseparable bond with the users. The need for understanding diffusion in a social structure is very significant in the present day world, where technology is all around and getting ever closer to humans in every aspect. This understanding would allow for the latest technological evolutions to be understood and exploited to their full potential.

In this thesis, the author discusses the impact of various policies followed by companies to involve customers in their decision-making strategies. He discusses significance of the chosen social structure, a scale-free network, and presents methods to create a scale-free network fit for his purpose by following a slightly modified form 1 of the B-A Algorithm. He continues his work by simulating an environment where the agents participate in diffusion of various signals.

The author simulates two companies competing for the attention of the agents, by trying to spread their propaganda. The companies take initiative by making an effort to move towards the agents, while the agents show their inclination by changing their color to get closer to the color representing the company. The effort made by the companies requires some expenditure, in terms of cost to the company and the inclination of the agents result in gain for the specific company. He uses the data gathered from the simulations to analyse the policies followed by the companies while attracting the customers, and presents his findings which suggest a basic model for maximizing gain to cost ratio for the companies.

The author concludes his work by showing that depending on per unit gain and cost, it is more effective to form long-term relations with the customers for stable or continuous assignments while periodically changing assignments, for e.g.. seasonal businesses, might reap more profit by not involving the customers to a high degree.

While the presented analyses are focused on marketing campaigns, the author emphasises on the importance of the work by discussing its relation to the real-world scenario and similarity to real-life. He also suggests that a similar approach could be used in wearable devices to let devices connect and map the social structure of their users, and this model would provide analytical ability over such a network.

Appendix A

Scale Invariance

Scale Invariance refers to the feature that some property of an entity does not change when one of its variable attributes is multiplied by a common factor.

Scale free networks exhibit this feature, often referred to as dilatation, by maintaining the nature of the connections being formed, irrespective of the number of the agents in the network. This makes scale free networks more suitable to be used in a simulation that may or may not be need to be scaled.

Appendix B

Barabási Albert Model

This algorithm for generating random scale-free networks was developed in 1999. This model was developed when researching the structure of the World Wide Web and the initial expectations of the researchers was to find a random network, but instead, they witnessed the existence of Power-Law in the degree distribution of the nodes, and the concept of scale-free networks was born.

This structure explain the presence of a few highly-connected hubs and a larger number of less connected nodes in the networks, and the architecture holds true for a wide range of networked structures, varying from a man-made physical infrastructures like the internet to the biological networks of proteins.

This model also emphasises on the role of hubs, which is very important and have applications in fields ranging from IT security to spread of contagious diseases and more. In words of Albert-László Barabási, the hubs make the scale-free networks “very robust to random failures, but very fragile to targeted attacks”.

Bibliography

[1] M. E. J. Newman. The structure and function of complex networks. *SIAM REVIEW*, 45:167–256, 2003.

[2] A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286:509–512, October 1999. doi: 10.1126/science.286.5439.509.

[3] *Diffusion of Innovations*. Simon and Schuster, 2003.

[4] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, January 2002. doi: 10.1103/RevModPhys.74.47.

[5] P. Erdős and A Rényi. On the evolution of random graphs. In *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, pages 17–61, 1960.

[6] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 1998.

[7] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, Jul 2001. doi: 10.1103/PhysRevE.64.026118. URL http://link.aps.org/doi/10.1103/PhysRevE.64.026118.

[8] Jari Saramäki and Kimmo Kaski. Scale-free networks generated by random walkers. *Physica A: Statistical Mechanics and its Applications*, 341(0):80 – 86, 2004. ISSN 0378-4371. doi: http://dx.doi.org/10.1016/j.physa.2004.04.110. URL http://www.sciencedirect.com/science/article/pii/S0378437104005503.

[9] Jaewon Yang and Jure Leskovec. Modeling information diffusion in implicit networks, 2010.

[10] Robert W. Rycroft. Does cooperation absorb complexity? innovation networks and the speed and spread of complex technological innovation. *Technological Forecasting and Social Change*, 74(5):565 – 578, 2007. ISSN 0040-1625. doi: http://dx.doi.org/10.1016/j.techfore.2006.10.005. URL http://www.sciencedirect.com/science/article/pii/S0040162506001879.

[11] Friedemann Mattern and Christian Floerkemeier. *From the Internet of Computers to the Internet of Things*, volume 6462 of *LNCS*, pages 242–259. Springer, 2010.

[12] I. Stojmenovic. Large scale cyber-physical systems: Distributed actuation, in-network processing and machine-to-machine communications. In *Embedded Computing (MECO), 2013 2nd Mediterranean Conference on*, pages 21–24, June 2013. doi: 10.1109/MECO.2013.6601317.

[13] Peter. Rojas. Musicgremlin: feeding your mp3 player wireless downloads after midnight, 2004. URL http://www.engadget.com/2004/07/13/musicgremlin-feeding-your-mp3-player-wireless-downloads-after/.

[14] Artemis. Steve perlman's pcell, 2014. URL http://www.artemis.com/press.

[15] Swarmanoid. Swarmanoid: Towards humanoid robotic swarms, 2011. URL http://www.swarmanoid.org/.

[16] Eric W. Weisstein. "königsberg bridge problem." from mathworld–a wolfram web resource., . URL http://mathworld.wolfram.com/KoenigsbergBridgeProblem.html.

[17] Eric W. Weisstein. "traveling salesman problem." from mathworld–a wolfram web resource., . URL http://mathworld.wolfram.com/TravelingSalesmanProblem.html.

[18] Eric W. Weisstein. "vertex coloring." from mathworld–a wolfram web resource., . URL http://mathworld.wolfram.com/VertexColoring.html.

[19] Bryan Kramer. There is no more b2b or b2c: There is only human to human (h2h). socialmediatoday, 2014. URL http://socialmediatoday.com/bryan-kramer/2115561/there-no-more-b2b-or-b2c-it-s-human-human-h2h. [online] Accessed: 2014-05-10.

[20] Blake Ives and Margrethe H. Olson. User involvement and mis success: A review of research. *Management Science*, 30(5):pp. 586–603. ISSN 00251909. URL http://www.jstor.org/stable/2631374.

[21] Bodil Sandén. The customer's role in new service development. 2007. URL http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-785.

[22] C.G. Jung. *The question of the therapeutic value of "abreaction"*. 1921. URL http://books.google.ca/books?id=BJo1cgAACAAJ.

[23] Joshua Wilt and William Revelle. Extraversion, 2008.

[24] Ed Jr. Pegg. "small world network." from mathworld–a wolfram web resource. URL http://mathworld.wolfram.com/SmallWorldNetwork.html.

[25] Reuven Cohen and Shlomo Havlin. Scale-free networks are ultrasmall. *Phys. Rev. Lett.*, 90:058701, Feb 2003. doi: 10.1103/PhysRevLett.90.058701. URL http://link.aps.org/doi/10.1103/PhysRevLett.90.058701.