Studying Dynamics of Multi-Agent Learning in Networks

دراسة التغييرات الديناميكية في الشبكات الصناعية متعددة الكائنات

by

SIMA IBRAHIM ABDEL KARIM SADLAH

Dissertation submitted in fulfilment

of the requirements for the degree of

MSc INFORMATICS

at

The British University in Dubai

May 2012

DECLARATION

I warrant that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.

I understand that a copy of my research will be deposited in the University Library for permanent retention.

I hereby agree that the material mentioned above for which I am author and copyright holder may be copied and distributed by The British University in Dubai for the purposes of research, private study or education and that The British University in Dubai may recover from purchasers the costs incurred in such copying and distribution, where appropriate.

I understand that The British University in Dubai may make a digital copy available in the institutional repository.

I understand that I may apply to the University to retain the right to withhold or to restrict access to my thesis for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred, the length of the period to be specified in the application, together with the precise reasons for making that application.

Signature of the student

COPYRIGHT AND INFORMATION TO USERS

Abstract

Artificial networks mining and analysis is one of the most recent and interesting area of research due to the explosion growth of this type of networks which is seen everywhere in our modern life. In the last few years few researchers' attempts have emerged to explain the behaviors of agents in artificial networks and to optimize agent's performance using policies learning and adaption methodology or optimizing the overall network's performance using self-organization techniques. In most of the previous researchers' attempts, studies focuses on two agents only and conducting an in-depth analysis was difficult due to lack of enhanced and improved visualization and analysis tools.

In this thesis, the main purpose of the study was to investigate the dynamics of learning in scale free artificial networks (multi-agent system in particular) of different sizes, over different periods of time and using different game theory models. Using Dynamic Network Visualization and Analysis tool (DNVA); three case studies were studied. Observations post to each case study have been made and key results have been formulated based on them. The result of the study showed that different parameters used in a network for example, (size of the network, time and interaction period between agents) have significant effect on the behavior of the artificial network. Also, the study discovered the reasons behind two observations that were reported in previous research. Finally, the study of this thesis revealed issues of interest to the researchers in the filed which is the similarity in factors affecting the cooperation behavior in social and artificial networks.

ملخص الرسالة
تعد دراسة و تحليل الشبكات الصناعية إحدى ابرز الأبحاث المثيرة للإهتمام حديثا و يرجع ذلك لكونها واحدة من أكثر الشبكات استخداما حاليا، حيث نلحظ تزايدا مستمرا في تواجدها و استخدامها بمختلف مناحي الحياة في الوقت الحاضر .و لقد شهدت السنوات الأخيرة القليل من محاولات الباحثين لدراسة وتحسين أداء هذا النوع من الشبكات باستخدام تقنيات التعلم و تغيير السياسات الخاصة بكل كائن أوعنصر (agent) من عناصر الشبكة أو باستخدام تقنيات التنظيم الذاتي و ذلك على نطاق الشبكة بأكملها .الجدير ذكره أن معظم المحاولات السابقة للباحثين اقتصرت على دراسة سلوك التعلم لدى شبكات تحوي كائنين agents) (فقط ذلك أن هناك قلة في البرامج المطورة اللازمة لتصوير و تحليل هذا النوع من الشبكات للحصول على دراسة جيدة و معمقة.
الهدف الرئيس من هذه الرسالة هو دراسة و استكشاف ديناميكية التعلم في الشبكات الصناعية من
مختلف الأحجام و على فترات مختلفة من الزمن و باستخدام أنواع مختلفة من طرق التفاعل و اللعب .
في هذه الرسالة قمنا بتطوير و استخدام برنامج يحوي العديد من الخصائص التصويرية و التحليلية المحدثة و قد قمنا باستخدام هذا البرنامج المطور لإجراء ثلاثة دراسات تحليلية حول الشبكات الصناعية و تدوين الملاحظات التي من شأنها ان تسهم في استخراج مجموعة من النتائج التي تخص هذا النوع من الشبكات.

و قد أظهرت النتائج أن هناك الكثير من العوامل التي تؤثر بشكل كبير في تغيير سلوك التعلم لدى الشبكات الصناعية من مثل :
حجم الشبكة، الوقت النهائي لدراسة سلوك الشبكة و مدة التفاعل ما بين الكائنات في الشبكة .كما شمل البحث في هذه الرسالة دراسة بحثين سابقين في هذا المجال و ذلك لمعرفة أسباب حدوث ظاهرتين غير مفسرتين سابقا .و أخيرا، كشفت الدراسة في هذه الرسالة عن وجود ظاهرة جديدة من شأنها أن تثير اهتمام الباحثين و هي وجود تشابه في العوامل التي تؤثر على حدوث سلوك التعاون ما بين الكائنات في الشبكات الإجتماعية و الصناعية.

Acknowledgement

This thesis would never have been completed without the guidance and devotion of several individuals who in one way or another contributed to accomplish my greatest academic achievement.

First, all praises be to Allah for accepting my pray and give me the patient and strength to complete this thesis. On the academic level - my ultimate gratitude to the thesis supervisor, Dr. Sherief Abdallah for the support, encouragement and inspiration to melt all the obstacles I faced toward the completion of this thesis.

I always say, where would I be without my lovely family? Words can't express my feeling and appreciation for my great father and compassionate mother for their love, care, unlimited support and prays. My deepest gratitude and love for my beloved sisters (Kawther, Tamader, Raba'a, Alaa and Tasneem) and my beloved brothers (Anas, Mohammed and Ahmad) for the continuous support and care.

Last and not least, I would like express my appreciation and thanks for my lovely and sincere friends for prays and support.

Contents

List of Figures

List of Tables

Chapter 1

Overview

1.1 Introduction

Motivated by the widespread of multi-agent system in the real life and due to the increasing demand of artificial and intelligent multi-agent system in particular, mining and visualizing networks having artificial and adaptive agents (artificial networks) has attracted the scientists from different fields. Recently, optimizing performance of artificial networks and studying factors affecting its behavior are considered as active areas of research in the world. Optimizing performance of artificial networks follows two directions (1) using reinforcement learning techniques to optimize local agents' policies[1, 3] (2) using self-organization to optimize overall network structure [1]. To our knowledge, most of the previous researchers' attempts have focused on studying and optimizing the performance of only two agents playing a game using multi-agent reinforcement learning techniques and the network doesn't change over time (no addition or removal in agents or edges) [2, 6, 7]. Understanding the optimization done on large size network structure by applying self-organization technique is not simple due to the complexity of the networks' collective dynamics behavior. Evaluating the effectiveness of performance optimization technique requires enhanced visualization and analysis techniques to help in investigation and explanation of the adaptation in the networks' agents policies.

In this thesis, we provide an in-depth analysis for different types of artificial networks changing over time and explain the behavior of un-justified observations of pervious researchers' studies on artificial networks [4] [5]. The analysis has been done through three different case studies. In case study 1 and 2, we analyze scale free artificial networks of different size (up to 200 agents) with agents playing coordination game in case study 1, prisoner's dilemma game in case study 2 and Q-learning algorithm has been used for agents' learning. We present list of observations and results to reveal the main factors affecting the tested types of networks. In case study 2, we focus on studying the main factor affecting the cooperation behavior of artificial network and search for any similarity on the cooperation behavior between social and artificial networks. In case study 3, we study previous research database generated by Alshamsi, A. Abdallah, S. and Rahwan, I. 2009 [5] to evaluate the optimization in the performance of taxi dispatch system using multi-agent self-organization technique has been made to confirm the researcher findings and provide more explanation of the findings.

1.2 Problem Statement

Understanding artificial networks and its dynamics is a recent research and a serious issue due to the complication in this kind of networks since it has different system parameters changing concurrency over time and resulting on significant effect on the behavior of overall agents in the system. Studying artificial networks becomes increasingly important because the changes in networks' dynamics lead to either improvement or drop in the performance of the system. The main objectives of this thesis are to enrich the research of artificial network analysis by in-depth studies of the behavior agents in artificial networks and providing explanation of un-justified findings of previous researchers [4] [5]

To accomplish the main objective, an open source visualization and analysis tool has been used to run different experiments on three case studies conducted with more than 10 generated datasets in each for reliable outcomes. We focus on multi-agent systems over scale free network using Q-learning algorithm for learning and policy adaption and playing prisoner's dilemma or coordination game. In the last case study, one of the previous researchers study on taxi dispatch system has been analyzed and explained.

1.3 Research Question

This thesis aims to address the following research questions

- Can we explain the behavior of the isolated nodes that keeps looping between different policies in Abdallah, S. 2012 [4]?
- Can we provide explanation for the occurrence of improvement in the performance of taxi Dispatch system using multi-agent self-organization technique as described in Alshamsi, A. Abdallah, S. and Rahwan, I. 2009 [5]?

- Can we provide in-depth analysis of a multi-agent system that has the ability to learn and adapt their policy over time?
- Is there any similarity between the behavior of artificial evolving network learning over time and human or social network?

1.4 Contribution

In this thesis, the contributions are as follow

- In-depth analysis of the behavior of (1) artificial networks playing prisoner's dilemma game over scale free network and using Q- learning algorithm to adapt its policy over time (2) artificial networks playing coordination game over scale free network and using stateless Q-learning algorithm to adapt its policy over time.
- Providing an explanation for two issues raised (1) the unstable behavior of a node playing coordination game over scale free network and using Q-learning algorithm. This node keeps looping and never converged as described in Abdal- lah, S. 2012 [4] (2) the improvement in a taxi dispatch system using multi-agent self-organization technique as described in Alshamsi, A. Abdallah, S. and Rahwan, I. 2009 [5]
- Identifying the importance of interaction period as a factor in a learning and adapted multi-agent system playing prisoner's dilemma game to maintain co- operation.

1.5 Scope

In this thesis, we focus on mining, visualization and analysis of multi-agent system which has the ability to learn and adapt to optimize the final reward over time. The analysis in this thesis has been restricted to three cases including the behavior of (1) multi-agent network over scale free network playing coordination game and using stateless Q-learning algorithm over time (2) multi-agent network over scale free net- work playing prisoner's dilemma game and Q-learning algorithm over time and finally
(3) using multi-agent self-organization technique on real exist Taxi Dispatch system.

1.6 Organization of Thesis

The remaining chapters of this thesis are organized as follows. Chapter 2 provides literature review about networks in general and artificial networks in particular, survey of some available networks visualization and analysis tools and finally review of re- searchers' attempts to artificial networks or multi-agent systems. Chapter 3 presents background about the features of the implement and used Dynamic Network Visualization and Analysis tool, Q-learning algorithm, Game theory and finally games over network simulator which is used to generate the studied datasets. Chapter 4 provides details about the methodology we use to study this thesis case study. Chapter 5 demonstrates detailed discussion and analysis of the studied case studies 1, 2 and 3. At the end of chapter 5, detailed discussion of the research questions and illustration of how the analyses are used to answer the questions are provided. Finally, Chapter 6 concludes the findings of this thesis and provides suggestions and ideas for future research and enhancements.

Chapter 2

Literature Review

In this chapter, a brief explanation about networks, used measures and model in general and artificial networks in particular will be provided. Then details about dynamics network visualization and analysis concept will be provided along with survey of three tools implemented to support the analysis and visualization of networks. Finally, survey of attempts to analyze the behavior of scale free artificial network using prisoner's dilemma or coordination game will be evaluated.

2.1 Networks

A network is a set of agents (nodes or vertices) connected by a set of edges (links). A network in real world can be internet, peer to peer file sharing or a social network in which agents interact with each other in different way [8]. Each node could have one or more neighborhood agent for example in the network shown in figure 2.1, agent A is a neighbor to agent B.

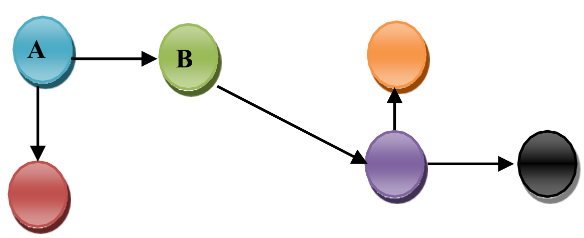


Figure 2. 1: Simple network

Networks have different measures, types and models. In this thesis I will mainly use degree as a network measure as described in section 2.1.1. Also will focus on artificial networks' type as describes in section 2.1.2 and scale free network model as described in section 2.1.3

2.1.1 Networks Measures

The most traditional networks measures are nodes' degree and clustering coefficient can be used to visualizing, mining and analyzing networks of interest.

Node's degree is a characteristic of a node measured by counting number of nodes connected to it. While node's clustering coefficient is the ratio of links of the node's neighbors to the number of all possible links exist between them [9].

2.1.2 Artificial Networks

The focus in this dissertation is on mining, visualizing and analyzing artificial networks containing artificial adaptive agents (nodes). Artificial adaptive agent is non-human agent which can improve its performance and learn to change its policy in a simulated environment in a way similar to the interaction of a human to its real environment [10]. Artificial network can be computers' network in which each computer represents an agent, so when an agent receives an inquiry from its environment it can either executes it or forwards it to other agent in the network. In section 2.3, description of attempt to studying, mining and analyzing artificial networks will be provided.

2.1.3 Scale Free Networks

In real life there are many networks' types like random networks, scale free networks and complex networks. In this thesis I will focus on studying artificial networks (network with nodes learning over time and adapting their policies) over scale free network model. Scale free network is a network that its degree follows power law distribution, which means that with time the probability of new added edges being connected to a node increases proportional to the nodes degree. The higher the node's degree is the greater the probability to connect [11]. In scale free network a node with high degree is called a hub [12].

2.2 Artificial Network Visualization and Analysis Tools

Recently, researchers in multi-agent system and artificial intelligence are working actively to study and improve the performance of artificial networks by optimizing the performance of its adaptive agent local policy and, therefore, the overall network structure.

To ease the studying process of artificial network, visualization methodologies are used to evaluate different performance optimization techniques applied on simple networks over time on specific domain. These simple visualization methodologies depends on the human perceptual ability to extract the network features and it poorly supports the need to generalize the outcomes of the tested performance optimization techniques

Network analysis is an approach summarizes the quantitative characteristics of a network in an understandable way regarding the nature of its domain [13] such as degree distribution. This emphasizes the need to scalable and user-friendly real time visualization and analysis tools to facilitate the data exploratory mining, visualization and analysis. Many researchers and software developers work intensively in the field of social network analysis and as a result useful tools are implemented like Social Network Analysis (SNA) tool [14] and the open source tool Java Universal Network/Graph Framework (JUNG) [15]. Both SNA and JUNG are used as powerful tools for mining, visualizing and analyzing a network or graph; however none of them facilitates real time visualization and analysis.

Gephi is a Dynamic network analysis open source tool which support real time visualization and analysis for large scale size simple and complex networks. It enables advance visualization and facilitates studying basic network data structure features like average degree and average clustering coefficient [16]. However, none of the three mentioned tools (SNA, JUNG and Gephi) implemented mainly to study the behavior of multi-agent system where agent learn to adapt their policy for better performance purposes, therefore, challenges faced in conducting in-depth analysis of the dynamics (like policy and action value) in data structure involved in learning using different performance optimization mechanisms due to the lack of proper software tool. In this dissertation, Dynamic Network Visualization and Analysis (DNVA) open source software (described in section 3.1) is mainly implemented and developed to target multi-agent networks and to study evolution of a network of agents by offering advanced visualization and analysis data mining techniques.

2.3 Related work

In this section, I will provide brief about previous researchers attempts on optimizing agents' local policies in a multi-agent system and studying artificial networks behaviors using different learning algorithms and playing different games.

The majority of the previous research in MARL focused on the simple case of two agents playing a simple game against each other. For example Abdallah & Lesser 2006 [2] studied the behavior of two agents using new proposed gradient ascent learning algorithm and tested on three different games which are coordination, matching pennies and tricky game. Also, Conitzer & Sandholm 2007 [7] who evaluated the behavior of two players using "*Adapt When Everybody is Stationary, Otherwise Move to Equilibrium*" (AWESOM) algorithm in general repeated games.

In this thesis, case studies have been conducted to elaborate the behavior of learning multi-agent system over scale free networks and playing either prisoner's dilemma or coordination games. Focusing on researchers attempt to study similar cases we found that, Mao-Bin Hu, Yong-Hong Wu, Rui Jiang & Qing-Song Wua 2008 [17] studied the behavior of agents in scale free network who are playing prisoner's dilemma game over time with equal percentage of distribution of cooperators and defectors initially. On each time step active agent plays with its active neighbor and parameter b is proposed to measure the advantage of defectors over cooperators. As a result they found that (1) in order to improve the cooperation, increase participation of hub agents (high degree node) and decrease the participation of low degree agent (2) to obtain cooperation optimal value of parameter α should be slightly greater than zero (3) percentage of cooperation depends not only on b parameter but also on parameter α and at different b values we have different optimal values for α.

Further investigation by CHENG Hong-Yan, DAI Qiong-Lin, LI Hai-Hong & YANG Jun-Zhong 2010 [18] who studied the effect of degree-degree correlation in obtaining cooperation in evolutionary scale free network with agents playing prisoner's dilemma game. Degree – degree correlation [19] is a characteristic of scale free network measured by parameter r (the Pearson coefficient). In terms of parameter (r), assortative network has positive r where high degree nodes tend to connect to another high degree nodes and low degree nodes tend to connect to other low degree nodes. On the other hand, disassortative network has negative r where low degree nodes tend to connect to high degree node and high degree node tend to connect to low degree node. Uncorrelated network has $r = 0$. The result of this study was (1) under high value of r, agents with high degree (hub nodes) tend to cooperate while low degree agents work to formulate a compact group and lower the percentage of cooperation (2) the percentage of cooperation increases when we have heterogeneity of degree in the underlying Scale free network.

Also, Szolnoki, Perc & Danku 2007 [20] tried to investigate the ability of scale free network to promote cooperation in prisoner's dilemma game. In this study the normalization parameter α has been used to study the transition toward effective reward as a result they found that cooperators nodes prefer to connect to a node of high degree "hub" regardless level of payoff efficiency.

It is important to mention that none of the surveyed researchers' attempts use learning algorithm in the study of a multi-agent system and as a result of investigation in this field we notice that there is lack on this area of research.

Chapter 3

Background

In this chapter, a basic background about the implemented open source tool for visualizing and analyzing dynamic network evolving over time is provided. Then brief about the used Q-learning algorithm and normal form games are described in section 3.2 and section 3.3. Basic background about the used gamed over network simulator is described in section 3.4. At the end, detailed brief about the case studies involved in the thesis is provided.

3.1 Dynamic Network Visualization Analysis (DNVA) Tool

Dynamic Network visualizing and Analyzing Network (DNVA) is an open source tool built as an extension of Social Network Image animator (SoNIA) open source tool [21].

SoNIA is a java based package implemented mainly to help in exploring dynamic network evolving over time and facilitate studying the relations of networks' entities (nodes). SoNIA implements different algorithmic layouts to facilitate the comparison of network over different layouts, the most commonly used layout for the purpose of the thesis is Kamada Kawai [22]. Thus SoNIA visualizes a network by representing the different nodes and edges of a network over different slices of time. In the below sections we will focus on DNVA tool by describing the enhancement features implemented to improve the visualization features and add the analysis features to SoNIA tool. The target of the enhanced implemented tool is supporting the analysts to make hypothesis, discover patterns and analyze datasets from different domains.

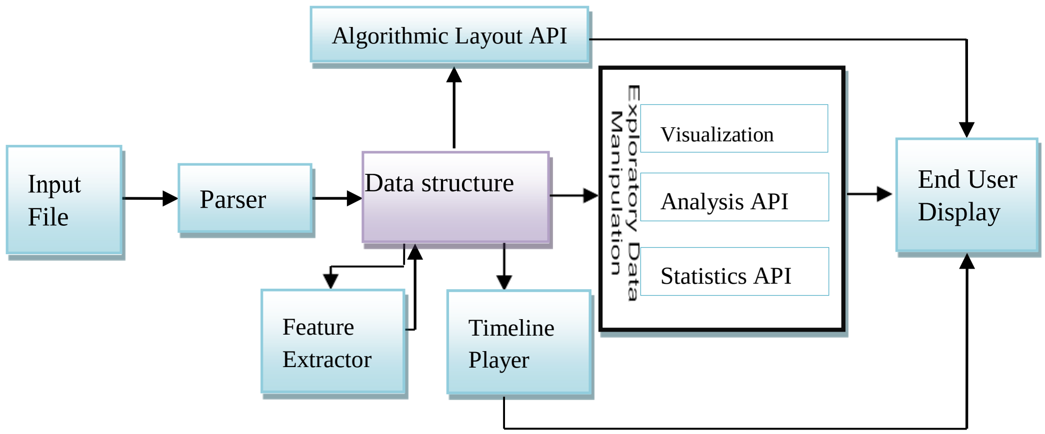


Figure 3. 1: DNVA Tool Block Diagram

Block diagram of DNVA tool is shown in figure 3.1, description about the new proposed input file, the implemented parser and network data structure will be provided in section 3.1.1. In sections 3.1.2, 3.1.3 and 3.1.4 description of the main elements of exploratory data manipulation block which are network visualization, dynamic network analysis and network statistics respectively will be provided. Finally, brief description of timeline player which is considered as big improvement of the base tool (Sonia) will be provided in section 3.1.5.

2.1.4 XML input file

Sonia (base tool) uses different input files including xml, however the format of the existing file capture changes on time by keeping records of all previous changes in the nodes which is in practical and might raise scalability issues. The new proposed xml input file format supports file increment to capture change once it occurs. The main features of the new input file format are as follow

1. XML based file format which is a well-known language
2. Represent network elements (nodes, edges and data associated with them) over different slices of times.
3. Nodes and edges' additional and removal is accepted.
4. Modification in nodes and edges attributes over time is allowed.
5. Default values provided for nodes with missing edges/nodes attributes.

The structure of the proposed DNVA input file is illustrated in table 3.1

```xml
<?xml version='1.0' encoding='UTF-8'?>
<DynamicNetwork>
<!-- Nodes Attributes definition -->
<Nodeattributes>
<attribute name='Node_Attr1' type='Double'/>
<attribute name='Node_Attr2' type='String'/>
</Nodeattributes>
<!-- Edges Attributes definition -->
<Edgeattributes>
<attribute name='Edge_Attr1' type='Double'/>
<attribute name='Edge_Attr2' type='String'/>
</Edgeattributes>
<!--Initial nodes, edges and attributes at time 0 -->
<change time='0'>
<!-- List of nodes -->
<node id='0'>
<attributeValue name='Node_Attr1' value='0.0'/>
<attributeValue name='Node_Attr2' value='value'/>
</node>
<!-- List of edges -->
<edge edgeID='0-1' source='0' target='1' type='add'>
<attributeValue name='Edge_attr1' value='0.0'/>
<attributeValue name='Edge_attr2' value='value'/>
</edge>
</change>
<!-- Once any change occurs it will be added (node/edges addition, removal, at- tributes' values modification-->
<change time='5'>
<!—Same as previous slice of time -->
</change>
</DynamicNetwork>
```

Table 3. 1: Sample XML input file

A parser is implemented to extract the network data structure including all nodes, edges and their attributes. Using the data structure, the tool enables nodes and edges features extraction like degree and it enables future extension.

2.1.5 Network Visualization

It is part of **exploratory data manipulation** block which is considered as a big enhancement and contribution of this tool. The main advanced visualization features are as follow

1. Using nodes' attributes as color, size, x-coord, y-coord or labels visualization attributes.

2. Using edges' attributes as color, size or labels visualization attributes, notice that in both nodes and edges visualization the maximum and minimum attributes' values are calculated and used to show the variance in attributes values.
3. Enable missing data value assignment.
4. Nodes – edges shape visualization.

2.1.6 Dynamic Network Analysis

Dynamic network visualization is the second part of data exploratory block, which is considered as a complementary of the tool and both dynamic network visualization and network statistics features (section 3.1.4) are the core component of the tool.

Analysis features includes dynamic filtering and dynamic highlighting. The dynamic filtering feature works by dynamically generating a list of nodes attributes with the corresponding range of values for each attributes in addition to one calculated at- tribute which is nodes' degree. This feature enables the user to formulate simple query (node's attribute, AND/OR operator, selected value) or complex queries (merge of simple queries). Based on the formulated query the network representation will be up- dated to show nodes that satisfy the query.

The mechanism of work in the dynamic highlighting feature is similar to dynamic filtering feature; the only differences is in the added coloring feature which allows the user to formulate a query and select color to highlight the nodes that satisfy the selected query. Notice that many highlighter and different colors can be formulated and selected.

2.1.7 Network Statistics

The last very interesting feature is the statistics feature which mainly enables the user to generate different graphs on either network or slice time level. On the network level different nodes' attribute based graphs can be generated including selected node's at- tribute heatmap over time which shows gradient color representation of selected node's attribution distribution (frequency of value), selected node's attribute timeline (parallel coordinate) which shows the different value of a node over time, 3-dimensional representation of network's edges based on the selected node's attribute and finally (x, y line) chart that shows the average value of the selected nod’s attribute over time.

It is important to mention that statistics features have been implemented in a way that facilitates development extension. So, it can work on slice level and it allows adding new calculated attributes.

The 2-dimensional graphs generated using statistics features have been implemented based on JFreeChart library[1] and the 3-dimensional graphs has been implemented using Jzy3D library[2]. The generated graphs can be saved in the user directory in different pictures format.

One more feature that facilitates the network analysis process is node Inspector feature which enable the user to show all attributes details of the selected node in specific slice of time, which is automatically changes over time.

2.1.8 Dynamic Network Timeline

To capture the changes in a network over time, play/stop and resume timeline is provided as a main feature of SoNIA tool which is used in the enhanced DNVA tool. The appropriate number of slices and their duration depends on the network of interest. For the purpose of software efficiency and speed, the overall time is scaled to be from 0 to 100 if the maximum time is greater than 100.

3.2 Q - Learning Algorithms

Q-learning is one of the reinforcement learning techniques (learning how to interact with the environment and maximizing the total reward by selecting best action starting from current state [23]).

[1] JFreeChart User guide is available at: http://ktipsntricks.com/data/ebooks/java/jfreechart-0.9.1-US-v1.pdf
[2] Jzy3D User guide is available at: http://www.jzy3d.org/guide.php

Q-learning algorithm working by expecting the future rewards choosing action (a) at state (s) and following the optimal policy $Q(s, a)$, assuming that we have a table of all possible combinations of states and actions, then to calculate the action value for each $Q(s, a)$ combination assuming that the immediate reward of executing action (a) starting from state (s) is (r) which can be calculated using the following general equation[24]

$$Q(s,a) \leftarrow Q(s,a) + \alpha\,[r + \gamma max_b(Q(s',b) - Q(s,a))]$$

Where the left $Q(s, a)$ is the new updated Q and α is the learning rate. Agents are working toward highest Q-value but with very small probability ε choose random action, this method represents e-greedy action selection.

3.3 Game Theory

The focus on this thesis is studying and analyzing the learning of multi-agent system, therefore, game theory is used since it provides a powerful framework for this purpose. Game theory is an approach to study the decision making strategies of set of agents to produce outcomes according to the agents preferences [02aij-wolf.pdf] , [Lit1.pdf], [Game Theory]

A game is a tuple (n, $A^{1\ldots n}$, $R^{1\ldots n}$), where n is the number of players (agents) A^j is the available actions for player j and R^i: $A^1 \ldots\ A^n \rightarrow \mathfrak{R}$ is the reward gained by player *i* when the joint actions played [25].

In conducted case studies of this thesis we will focus on using two games which are coordination game (described in section 3.3.1) and Prisoner's dilemma game (described in section 3.3.2).

3.3.1 Coordination Game

The general coordination game of two players is shown in table 3.2 where (A) represents action 1 and (B) represents action 2. Each cell (i, j) represents the reward received by row player 1 playing action *i* and column player 2 playing action *j* [26].

		Player 1	
Player 2		A	B
	A	a, a	c, d
	B	d, c	b, b

Table 3. 2: 2- actions general coordination game

Assuming that $a > d$ *and* $b > c$ actions (A, A) and (B, B) considered as Nash equilibrium

3.3.2 Prisoner's Dilemma Game

In prisoner’s dilemma game, each player select one of the two actions either “cooperate” or “defect” [27]. The fair or optimal reward for both of the players gained when they both cooperate, but if only one of them cooperates, the defector player will gain more reward. If two players are defector they will either lose or receive low reward. Table 3.3 shows the different actions of row and column players and their expected reward based on the selected action, noticed that the numbers which represent the reward are not fixed and just plotted to help in quantifying the expected reward.

		Player 1	
Player 2		Cooperate	Defect
	Cooperate	2,2	4,1
	Defect	1,4	1,1

Table 3. 3: 2- actions general prisoner's dilemma game

3.4 Games over Network Simulator

It is the tool used to generate data set used in case study 1 and case study 2 (sections 5.1 and 5.2). This simulator generates different topology’s networks including scale free and random networks, over different games including coordination game and prisoner’s dilemma games. Also the simulator enables the user to adjust predefined parameters to value of interest to facilitate the testing and investigation purposes. In addition, learning algorithm is incorporated to support generation of adaptive net- works.

For the purpose of the analysis in the thesis adaptive networks use stateless (case study 1, section 5.1) and previous state (case study 2, section 5.2) Q-learning algorithm is used and setting of configured parameters include network size (number of nodes), maximum time (entire time for testing the designed network), statistics period (time in which attributes values calculated based on the status of the network), interaction period (described in table 5.1) and logger type (output file format).

Chapter 4

Case Study Methodology

In this thesis three case studies will be explored, studied and analyzed to capture interesting observations and conclusions. In case study 1 and 2, scale free adaptive networks with different nodes' attributes playing different games (coordination game and prisoner's dilemma) have been generated and used, while only one dataset about taxi dispatch system [5] has been used in case study 3. Therefore, in this chapter will focus on describing methodology to generate, explore, visualize and analyze the different datasets used in case studies 1 and 2 along with the comparative process and criteria used. Also, in this section I will give glance highlight on how the different visualization and analysis features of DNVA package are used to accomplish the analysis process of the studied case studies

The analysis methodology for case study 1 and 2 starts by setting the testing parameter including network size, statistics period, interaction period and maximum time. Then the below steps have been repeated for reliable outcomes.

- **Step 1**: start by generating a small network dataset for short period of time using the proposed game and learning algorithm.

- **Step 2**: load and visualize the generated dataset using 'Kamada Kawai' algorithmic layout.

- **Step 3**: customize the visualization of the network to analyze the effect of the dataset attributes on the behavior of the network over different slice of time.

- **Step 4**: use different network analysis technique to study the behavior of the network. For example use Filtering feature to display or highlight nodes according to the applied attributes (Policy, reward) filter query/s. In addition to using different statistics charts. For example generate specific network - attribute columnar heat map over all time slices or produce the specific attribute time series of each node over all time slices.

- **Step 5**: compare and study the observation and result of the used visualization and analysis techniques.
- **Step 6**: adjust the tested parameters and repeat steps (2-5)
- **Step 7:** Steps 1 to 6 are repeated for a minimum of 10 generated datasets (networks) to generalize the observation.

Chapter 5

Analysis and Discussion

In this chapter, the analysis of the three conducted cases study for scale free adaptive network is discussed to generalize its behavior over time. To further motivate and to put our result into context, each case study has been intensively analyzed and discussed in section 5.1, 5.2 and 5.3. This chapter is concluded by an-in depth discussion on how the collected findings and results illustrate answers to chapter 1 research questions.

In the Sections 5.1 and 5.2, new terms will be used to describe or analyze the datasets. Table 5.1 illustrates the definitions of some used terms:

Term	**Description**
Statistics period	Periodic time to calculate the updated values of the nodes' attributes
Maximum time	The entire time of studying the network
Interaction period	Parameter used in prisoner's dilemma game and it represents how long two nodes play with each other before moving to other node
Hub node	Node with high degree
Policy	Indicator for the probability of choosing an actions (a) at stat (s) [24].
Reward	The immediate payoff received by the node (agent) after executing action (a)
Action value	It is the expected future reward knowing the current state and using specific policy [28].

Table 5. 1: Terms used with description

For better understanding of the analysis in this thesis, consider the following two notes: (1) for all tested networks, the overall time is scaled to be from 0 to 99 (as described in section 2.1.8) so the term time 99 represents the actual last (max) time in the network. (2) The term adaptive network and artificial network are used interchangeably.

5.1 Case Study 1 – Scale Free with Coordination Game

In this section, we discuss a case study focuses on studying the behavior of scale free adaptive network playing coordination game and using stateless Q-learning reinforcement learning algorithm

$$Q(a) \leftarrow Q(a) + \alpha(r - Q(a))$$

Where r is the received reward sample, and α is the learning rate which has value between 0 and 1. The network use e-greedy as action selection method so most of the time nodes behave greedily and use action with highest Q (action) value, but with small probability e=0.04, nodes choose action that doesn't have highest Q value.

The case study has been conducted and studied on phases described on sections 5.1.1 and 5.1.2.

5.1.1 Dataset Generation and Exploration

Using games over network simulator (described in section 3.4), 4 different types of adaptive networks (with different combination of size and max time) have been generated to be studied using the analysis and visualization features of DNVA tool. It is important to mention that for each 4 types' adaptive networks, minimum 10 sample datasets have been tested and studied.

Table 5.2 illustrates some statistics about the 4 adaptive networks, including the type of the network, size of the network (number of nodes and edges), nodes attributes, maximum time and statistics period.

Adaptive Net-work	Network Size		Max time	Statistics Period
	Nodes	Edges		
DN 1	30 to 50	60 to 100	1000-10000	100
DN 2	30 to 50	60 to 100	50000-100000	100
DN 3	150 to 200	300 to 400	1000-10000	100
DN 4	150 to 200	300 to 400	50000-100000	100

Table 5. 2: Statistics about 4 adaptive networks used in case study 1

Notice that all networks start from time 0, use stateless Q-learning reinforcement learning algorithm having policy (the percentage of choosing first action over time) and reward as nodes' attributes and play coordination game with initial policy=0 and reward=0.

5.1.2 Experimental Setup and Dataset Analysis

Several Datasets of adaptive networks' types mentioned in table 5.2 have been generated and experiments conducted, starting by a small network of type - DN1 which has 30 nodes and 58 edges. This network starts at time 0 and ends at time 10000 with statistics period = 100.

Setting Policy attribute as a color attribute and using Kamada Kawai algorithmic layout, the visualization of the network in slice 0 is shown in figure 5.1

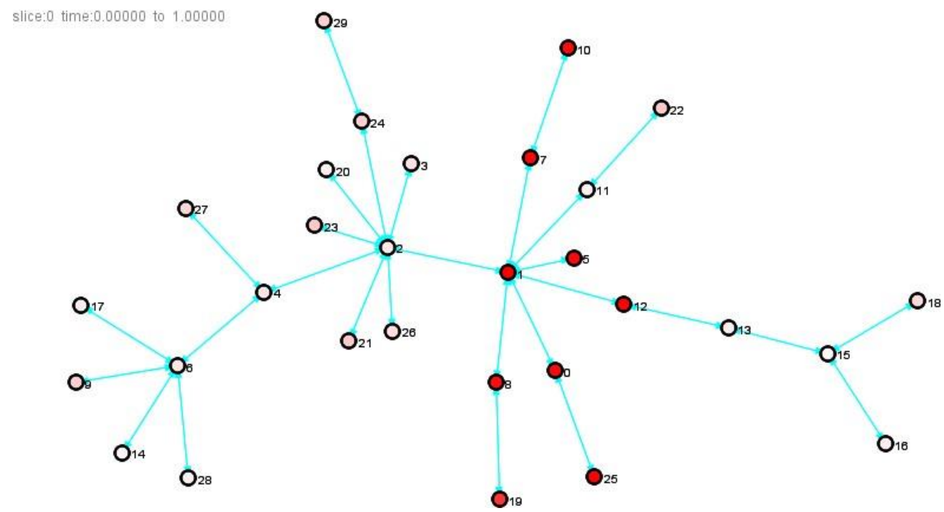


Figure 5. 1: Visualization of DN1 network – policy is visualization attribute at time 0

Notice that nodes’ colors follow gradient color from white to red representing policy =0 to policy=1 respectively and the number indicated the node id.

To study the behavior of each node all over time slices, the nodes’ policy changing timeline has been generated as shown in figures 5.2

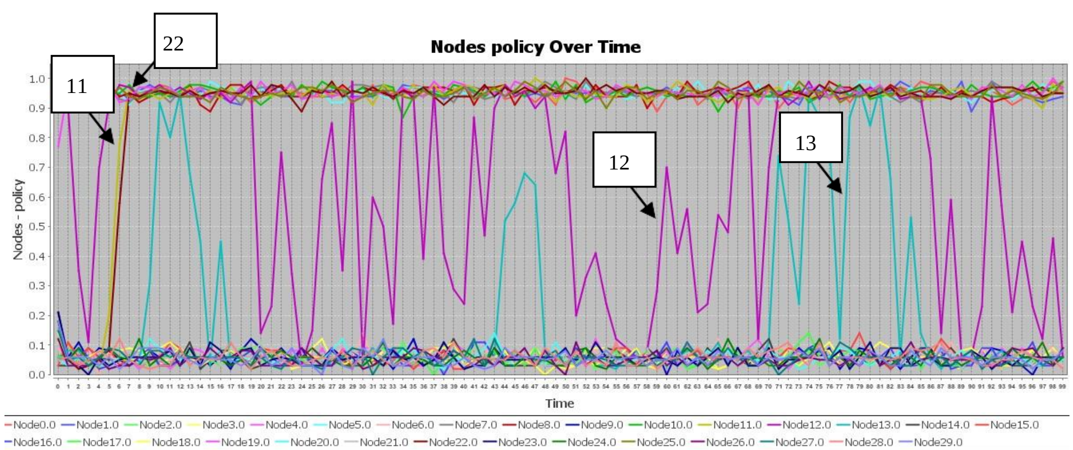


Figure 5. 2: Nodes’ policy changes over time - Timeline

Figure 5.2 show that most of the nodes use either policy 1 or policy 0. Two nodes (12 and 13) keep looping between the two policies and two nodes (11 and 22) switch their policies from 0 to 1.

To identify the looping nodes (12 and 13) over the network and to explain their behaviors, the studied network has been visualized over different time slices. Figure 5.3

shows the representation of the dataset over times 5, 10, 40 and 99.

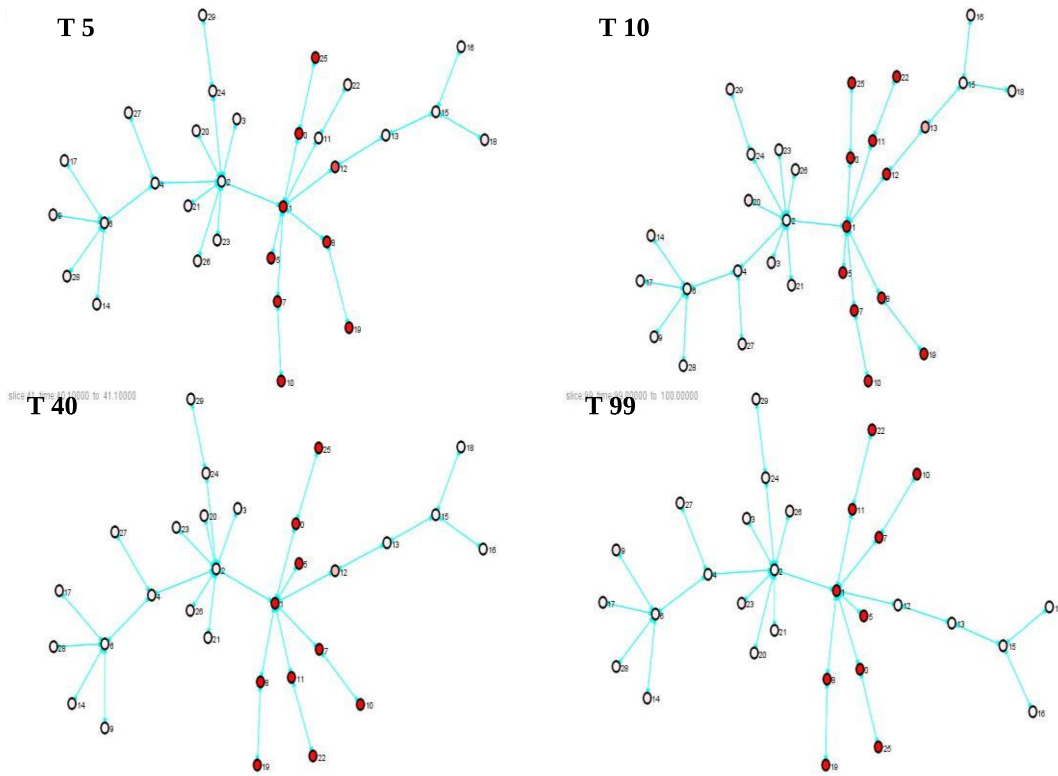


Figure 5. 3: Visualization of DN1 network – policy is visualization attribute at times 5, 10, 40 and 99

Looking at the visualized network in figures 5.1 and 5.3 over different time slices, we observe the following:

Observation 1 - Nodes 1 and 2, which are consider as hub nodes and they are <u>connecting a cluster of nodes of same policy</u> never change their policy.

Observation 2 - Nodes 11 and 22 (which are connected to a hub node 1 and have in-out degree = 1 or 0) switch their policies from 1 to 0 according to the used policy in the nub node connected to them.

Observation 3 - The policies of nodes 12 and 13 (which lie between two central -hub - nodes) never converge and keep alternating between 0 and 1 all over time slices.

Observation 4- Most of the nodes converges in early time and follows either policy 0 or policy 1.

Figure 5.4 displays the 3-dimensional representation of the agents policies over time where each point represents the policy of an edge (x is the source node policy and y is the target node policy).

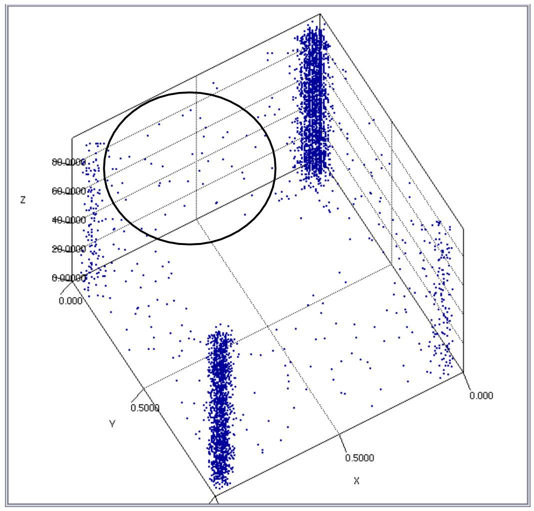


Figure 5. 4: 3-D Visualization of edges over time based on policy attributes

Observations 1- 4 are confirmed using figures 5.3 and 5.4, more observations can be concluded using these figures as follow,

Observation 5 – Most of the edges are connected between nodes that follow same policy (0, 0) or (1, 1) which are most likely nodes of the same cluster other than the hub node, also we notice that few edges connected between nodes that follow opposite policies (0, 1) or (1, 0) which are most likely hub nodes connected between nodes of same policy in one cluster.

Observation 6 – Also, we notice that few edges are connected between nodes that keep looping all over the time and never converge on specific policy. Sample of these edges are circled in figure 5.4.

Observations 5 and 6 are confirmed by the result found by Abdallah, S. 2012 work [4] who studied the behavior of two players in multi-agent system playing coordination game and using Q-learning algorithm over time.

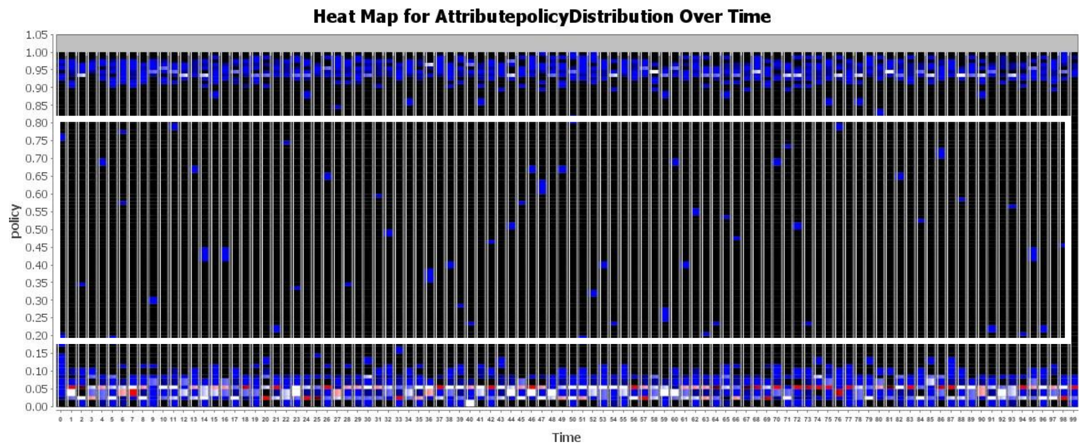


Figure 5. 5: Heatmap for frequency of policy over time

Figure 5.5 displays the distribution of policies of all nodes all over the time following a gradient color vary from dark blue to dark red, where dark red represents most frequent policy, dark blue represents the least frequent policy and black color represent zero. Using figure 5.5, the following observations are concluded,

Observation 7 – Policy 0 and policy 1 are the most frequent policies.

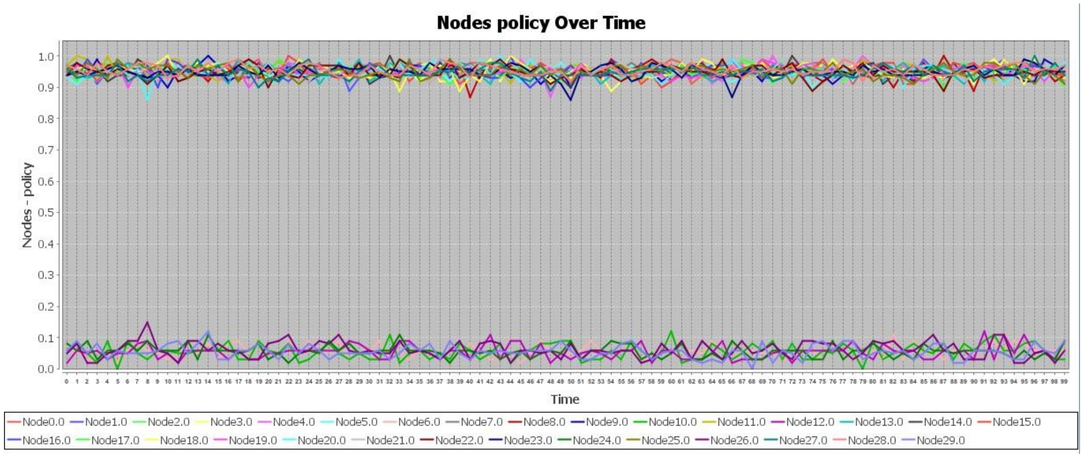
Observation 8 – Few in between policies (> 0 and < 1) indicated by white rectangle, are not stable.

To verify and generalize the observations we have repeated the experiment on more than 10 new small size sample adaptive network datasets of type – (DN1), using similar criteria of the first tested dataset and as a result we notice that observations 1 to 8 applied.

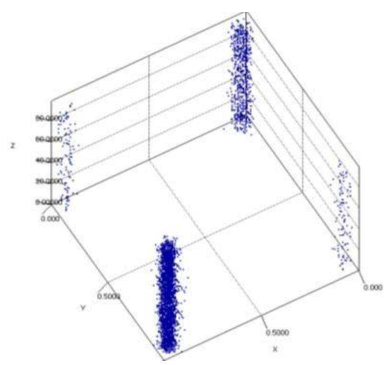
Further to the analysis done in small size adaptive network type – (DN1) and for a short period of time, experiments have been conducted on Small size adaptive network type - DN2 which has 30 nodes and 58 edges but over long period of time, this network starts at time 0 and ends at time 100000 with statistics period = 100.

After repeating the steps applied in adaptive network of type – (DN1) and investigating the different visualization and analyzing techniques using customized attribute(policy) visualization, the nodes' policy changing timeline, 3-dimensional visualization of all edges over time in terms of source-target nodes' policy and Policy distribution over time heatmap charts we conclude that observations (1 to 7) are applied in Network of type – (DN2), however and after testing 10 datasets of type - DN2, a new interesting observation can be concluded as follow

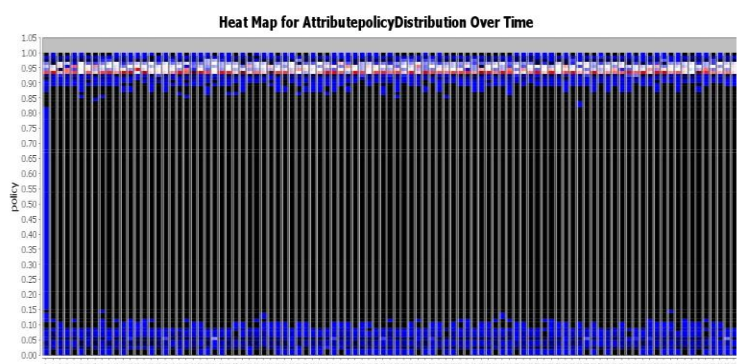
Observation 9- Out of 10 different tested datasets of type – (DN2), **<u>all</u>** nodes in 7 datasets converged to either policy 1 or policy 0 and **<u>none</u>** of the nodes keep looping between the two policies (see figures 5.6 (a), (b) and (c)). Out of the seven converged datasets, all nodes in two datasets have policy 1.



(a)



(b)



(c)

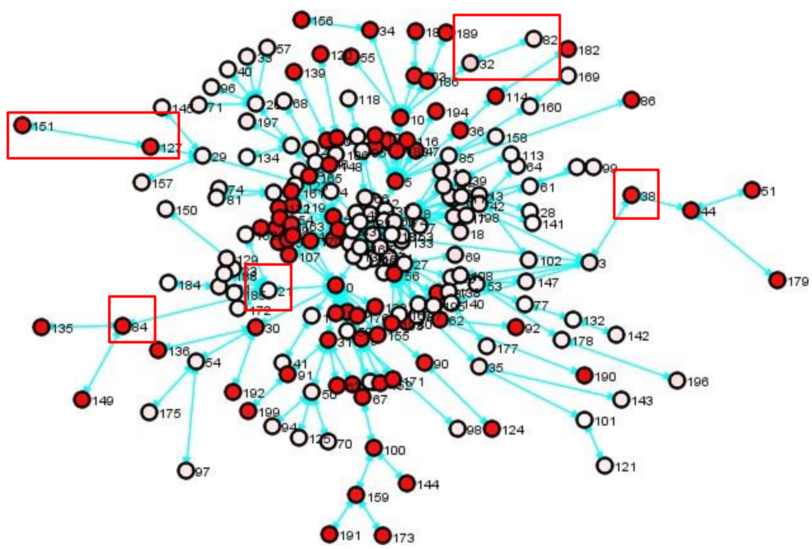
Figure 5. 6 : Small size network over long time nodes' (a) policy timeline, (b) 3-D for edges and (c) Heatmap for frequency of policy

After testing the behavior of small adaptive network over short and long period of time, we then apply the same techniques in a larger adaptive network, starting by adaptive network type - DN3 which has 200 nodes and 398 edges. This network starts at time 0 and ends at time 10000 (short time) with statistics period = 100.

Setting Policy attribute as the color attribute and using Kamada Kawai algorithmic layout the visualizations of the network in the first and last time are shown in figure 5.7.

Comparing the two slice of visualization, we notice that, observation 1 applied on the two hub nodes (21 and 34) indicated by a red rectangle. Also by investigating nodes (32, 82, 127 and 121) we notice that observation 2 applied to them each according to the policy of the connected hub node. Moving to the node of interest which is node 38 that lies between two hub nodes (44 and 3) of opposite policies, it is clear that it remains at its initial policy, however looking at the generated nodes' policy changing timeline chart shown in figure 5.8, we observe that node (38) keeps looping and never converge as mentioned in observation 3.

T 0



T 99

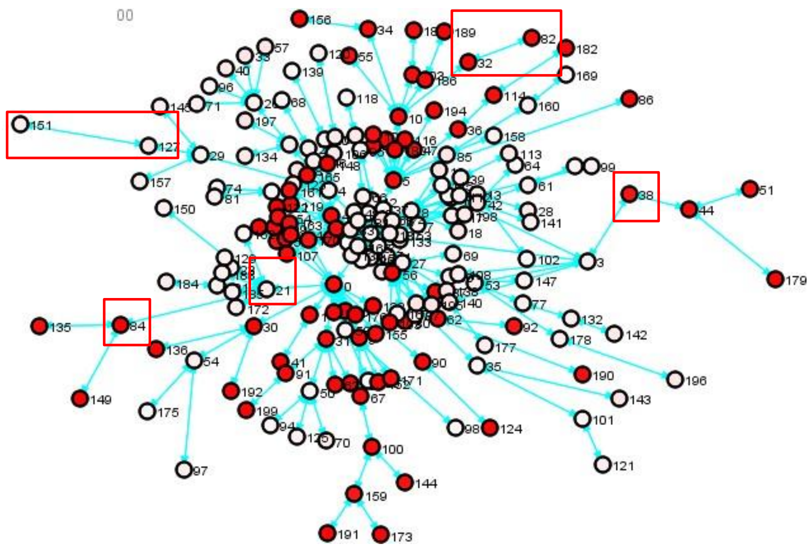


Figure 5. 7: Visualization of type DN3 network at times 0 and 99

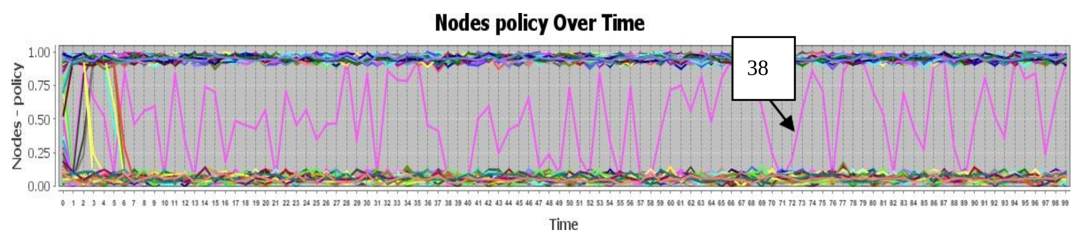


Figure 5. 8: Type DN3 network - Nodes policy changes timeline

Also, referring to figure 5.8, we notice that all nodes converge early to use either policy 0 or policy 1 as in observation 4.

Similar to the technique used to generalize and deeply investigate the behavior of small size adaptive network in scale free network, more than 10 datasets type – DN3 have been generated and tested to confirm the above observations.

The last step in this case study is to test a bit large size adaptive network over long period of time. Experiments have been conducted on large size adaptive network Type – DN4 which has 150 nodes and 298 edges but over long period of time, this network starts at time 0 and ends at time 100000 with statistics period = 100.

Setting policy attribute as a color attribute and using Kamada Kawai algorithmic layout, the visualizations of the network in the first and last time are shown in figures 5.9 and 5.10.

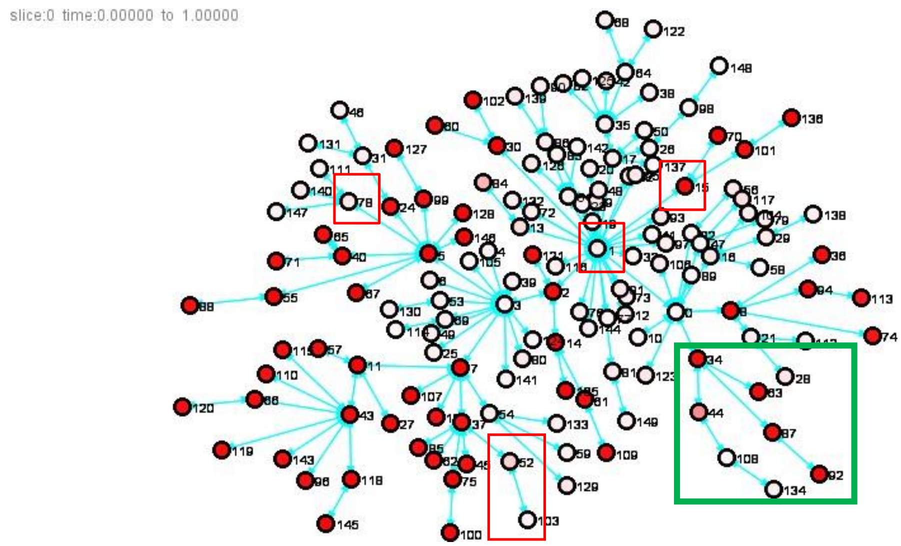


Figure 5. 9: Visualization of DN network over long time at time 0

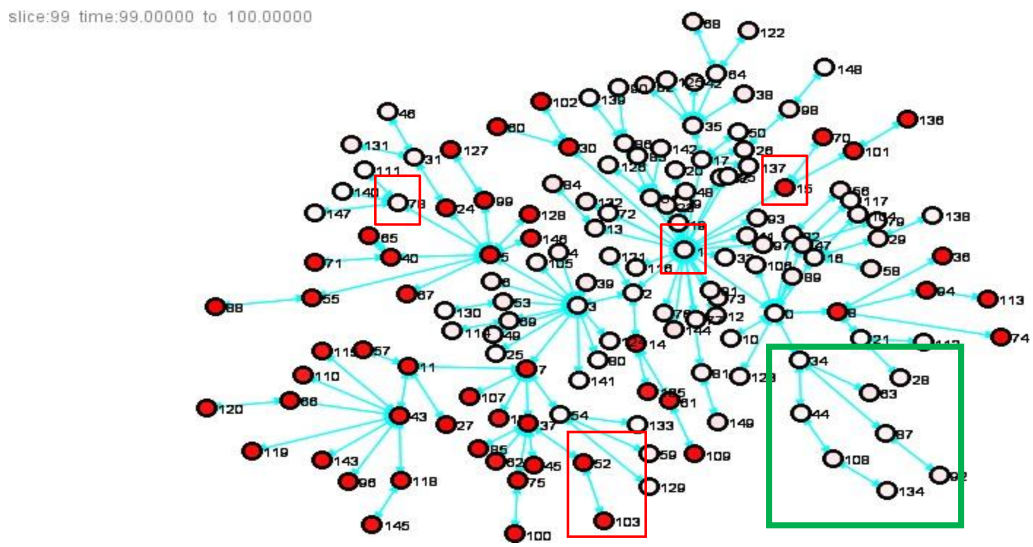


Figure 5. 10: Visualization of DN4 network over long time at time 99

By investigating the network we notice that, observation 1 applies on nodes (1 and 15), observation 2 applies on nodes (52 and 103), and observation 3 applies on node (24).

Figure 5. 11: DN4 network - Nodes policy changes timeline

The new interesting observation on the tested dataset can be concluded by investigating the cluster of nodes indicated by green rectangle in figures 5.9 and 5.10. Looking at node (34) which is considered as hub node, we notice that its policy has been switched from 1 to 0 and it follows policy of the hub node (0) connected to it. A deeper analysis of this hub node which is connected to nodes characterized by having mixed policies and in-out degree =1, and after exploring and investigating similar cases in other adaptive networks of type - DN4, a new interesting observation can be formulated as follows

Observation 10 – Having a hub node connecting between cluster of nodes characterized by having mixed policies and in-out degree =1, gradually the hub node and its connected cluster of nodes will follow a policy of another hub node connected to the cluster hub node.

The behavior of more than 10 datasets of type - DN4 adaptive networks have been studied and analyzed to confirm the observations, as a result two new observations can be concluded as follow

Observation 11 – On the contrary of observation 9, using large size adaptive network (150-200 node) over either short or long time period, will still have nodes keep looping between the different policies, logically this happens due to the increment in the number of nodes which increases the probability of having the case of loop node.

After generating, visualizing and analyzing the 4 different types of proposed adaptive networks (DN1, DN2, DN3 and DN4). Using observation 1 to 6 which are applied to the 4 different tested types of adaptive network, results 1, 2 and 3 can be formulated as follow

Result 1 - Using coordination game and Q learning algorithm in **Small** (30 -50 nodes) **or large** (150-200 nodes) **size** scale free Adaptive network over **Short** (10-100 statistics period) **or long** (500 to 1000 statistics period) **time**, a node or series of nodes with in-out degree =1, connected to a hub node of opposite policy, will switch their policy to hub node's policy.

Result 2 - Using coordination game and Q learning algorithm in **Small** (30 -50 nodes) **or large** (150-200 nodes) **size** scale free Adaptive network over **Short** (10-100 statistics period) **or long** (500 to 1000 statistics period) **time**, a node or series of nodes lies between two hub nodes of different policies will keep isolating and looping between the different values of policies.

Result 3 - Using coordination game and Q learning algorithm in **Small** (30 -50 nodes) **or large** (150-200 nodes) **size** scale free Adaptive network over **Short** (10-100 statistics period) **or long** (500 to 1000 statistics period) **time**, most of the nodes will con- verge in early time and use either policy 1 or policy 0

Using observations 9 and 11, result 4 can be formulated as follow

Result 4 - Using coordination game and Q learning algorithm in **Small** (30 -50 nodes) **size** scale free Adaptive network over **long** (500 to 1000 statistics period) **time**, <u>all</u> nodes in <u>most of the networks</u> will converge to use either policy o or 1 and none of the nodes keeps looping between the different policies.

Using observation 10, result 5 can be formulated as follow

Result 5 - Using coordination game and Q learning algorithm in **Small** (30 -50 nodes) **or large** (150-200 nodes) **size** scale free Adaptive network over **Short** (10-100 statistics period) **or long** (500 to 1000 statistics period) **time,** if we have hub node connected to <u>other</u> hub node and connecting between cluster of nodes characterized by having mixed policies and in-out degree =1, therefore the hub node and all nodes in the cluster will follow the policy of the <u>*other*</u> hub.

5.2 Case Study 2 – Scale Free with Prisoner's Dilemma Game

In this section, we discuss a case study focuses on studying the behavior of scale free adaptive network using prisoner’s dilemma and Q-learning reinforcement learning algorithm.

The initial Q-Action values for the different states and actions are illustrated in table 5.3, where (s) represents the current state, (a’) represents next action and the state or action (0) represents cooperation policy while (1) represents defection policy.

s_{node}	$s_{other\ node}$	a'_{node}	Action value
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

Table 5. 3: Initial Q values used

The case study has been conducted and studied on phases described on sections 5.2.1 and 5.2.2.

5.2.1 Dataset Generation and Exploration

Using games over network simulator (described in section 3.4), 6 different types of adaptive networks (with different combination of size, max time and interaction period) have been generated to be studied using the analysis and visualization features of DNVA tool. It is important to mention that minimum 10 sample datasets have been tested of each of the 4 adaptive networks' types.

Table 5.4 illustrates some statistics about the 4 adaptive networks, including the type of the network, size of the network (number of nodes and edges), maximum time and interaction period

Adaptive Network	Network Size		Max time	Interaction Period
	Nodes	Edges		
DNP 1	10	18	1000-10000	1000
DNP 2	10	18	1000000 - 5000000	1000
DNP 3	10	18	1000000 - 5000000	50000-1000000
DNP 4	30-50	58 - 98	1000-10000	1000
DNP 5	30-50	58 - 98	1000000 - 5000000	1000
DNP 6	30-50	58 - 98	1000000 - 5000000	100000-1000000

Table 5. 4: Statistics about 6 adaptive networks used in case study 2

Notice that all networks start from time 0 with fixed statistics period = 1000, having policy, reward in addition to 8 action values (Q[0-0,1], Q[0-1,1] Q[0-1,0], Q[1-1,0], Q[0- 0,0], Q[1-1,1], Q[1-0,0], Q[1-0,1]) as nodes' attributes. The initial policy of all nodes is (defect) which represented by policy 1 in this case study and the initial reward is 0 for all nodes.

5.2.2 Experimental Setup and Dataset Analysis

Several Datasets of adaptive networks' types mentioned in table 5.4 have been generated and experiments conducted, starting by a network of type – DNP1 which has 10 nodes and 18 edges. This network starts at time 0 and ends at time 10000 with statistics period = 1000 and interaction period=1000.

Setting Policy attribute as a color attribute and using Kamada Kawai algorithmic layout the visualization of the network in slice 0 is shown in figure 5.12

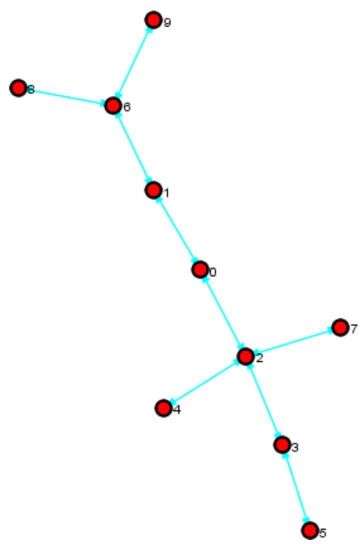


Figure 5. 12: Visualization of DNP1 type network at time 0

As mentioned in section 3.2.1 all nodes defect at the beginning (red color represents defect policy (1)). By visualizing the network over different time slices (17, 60 and 99) as depicted in figure 5.13

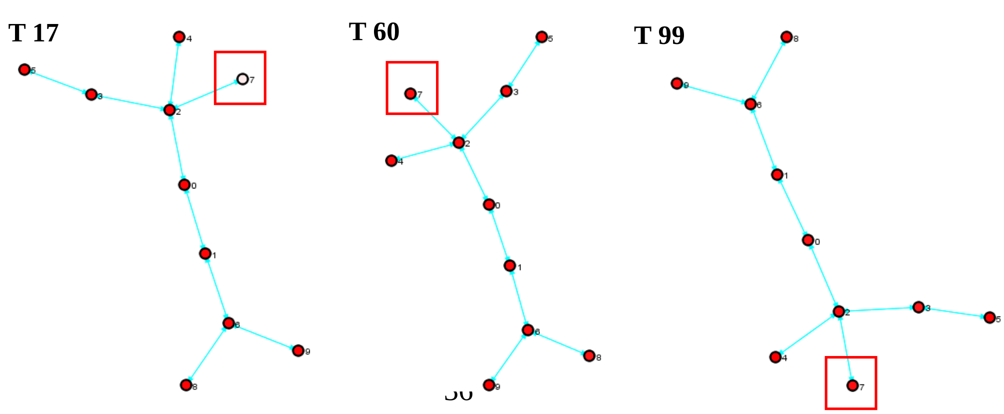


Figure 5. 13: Visualization of DNP1 type network at time 17, 60 and 99 Looking at the visualized network in figure (a, b and c) over different time slices, we observe the following

Observation 12 – Only node 7 seems to follow cooperation policy at early time (17) then it switches to defect.

Observation 13 – The 10 nodes follow defection policy over short period of time (1000).

For in-depth study of the behavior of the network all over time slices, the nodes' attribute (policy, reward and Q-action value) timeline graphs are shown in figures 5.14, 5.15 and 5.16 respectively and the 3-dimensional visualization for all edges of the network over time based on policy attribute have been generated and shown in figure 5.17.

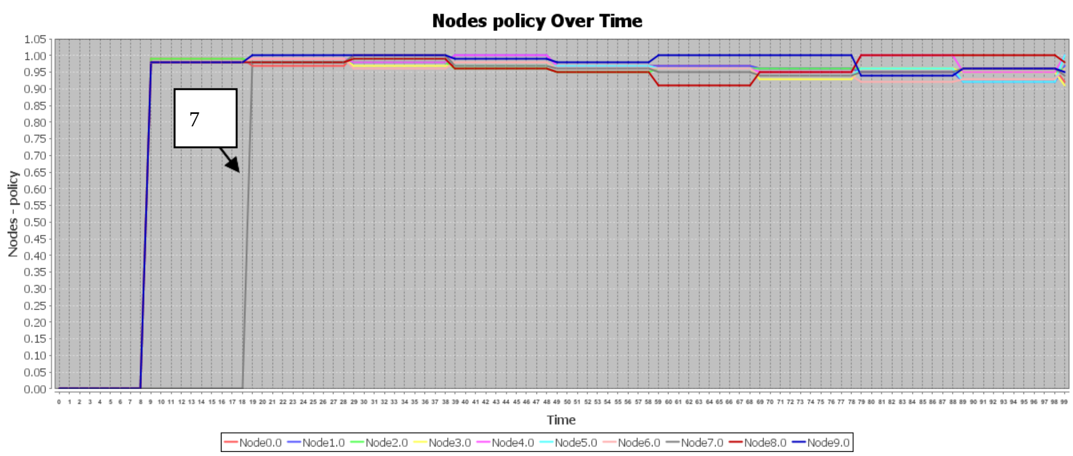


Figure 5. 14: Nodes policy changes timeline - DNP1

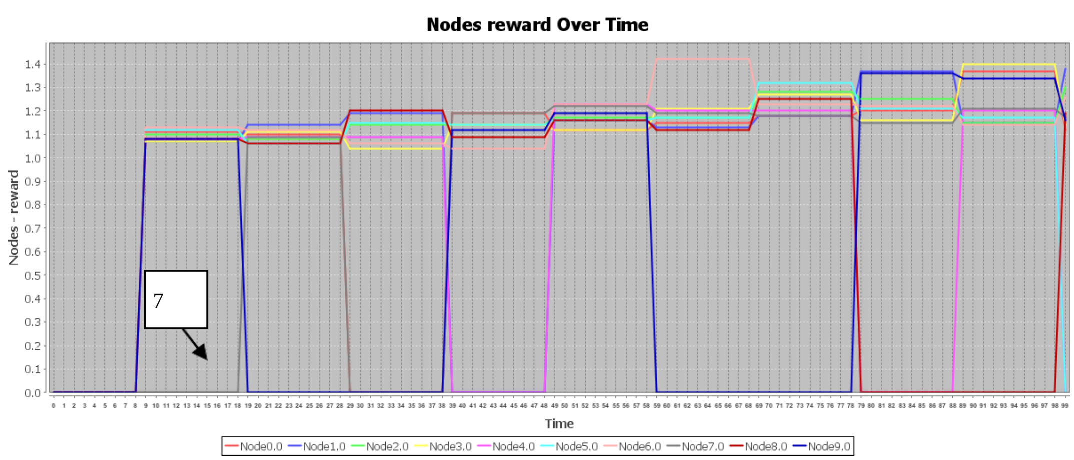


Figure 5. 15: Nodes reward changes timeline - DNP1

Referring to policy and reward graphs in figures 5.14 and 5.15 respectively, we notice that most of the nodes follow defection policy (1) on early time and the reward increase based on the followed policy (observations 1 and 2 confirmed). Focusing on node 7, we notice that within the early time node 7 doesn't have the chance to play that's why it looks cooperate (policy = 0).

To intensively study the changes in the action values for different states over time, timeline graph of the 8 different action values is depicted in figure 5.16.

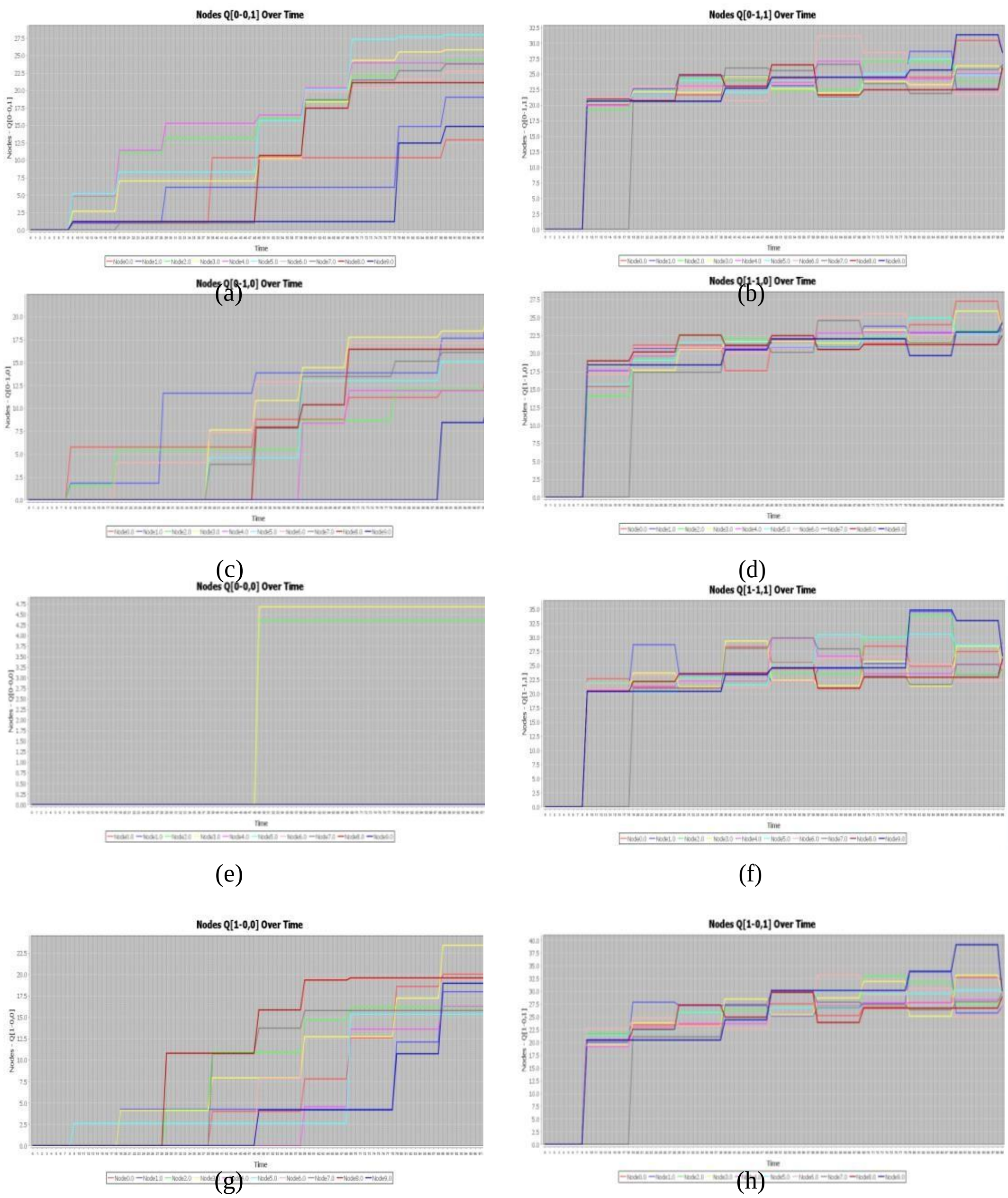


Figure 5. 16: Nodes action values timeline - DNP1

Interesting observation can be concluded from figure 5.16 and formulated as follow

Observation 14 – Action value (reward) for Q[1,1,1] shown in figure. 5.16-f is the highest which means if two nodes defect and keep defecting the reward will increase while the action value(reward) for Q[0,0,0] is lowest which means if two nodes cooperating and keep cooperating the reward will decrease or at least will not increase. Also, we notice that the action value (reward) for Q[0,0,1] is much larger than action value(reward) for Q[0,0,0] which means if two nodes cooperate and one of the nodes switch its policy to defect, its action value(reward) will increase.

Finally, figure 5.17 shows the 3-dimensional visualization of networks edges, where each point represents an edge of the network in specific time slice (x: is the source's node policy and y: is the target node's policy), this figure confirms observations 1,2 and 3 since all points clustered around value 1.

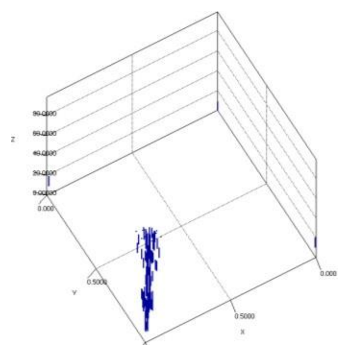


Figure 5. 17: 3-D Visualization of edges over time based on policy attribute

To verify and generalize the observations we have repeat the previously conducted experiment on more than 10 similar adaptive network datasets of type – (DNP1), using similar criteria of the first tested dataset and as a result we notice that observations 1,2 and 3 applied.

Further to the analysis done in adaptive network of type – (DNP1) and to study and realize the parameters that might make significant change and effect on this type of networks, experiments have been conducted on adaptive network type – (DNP2) which has same number of nodes (10), edges (18) with same interaction (1000) and statistics (1000) period but this time over long period of time (1000000).

The network has been visualized over different time slices. Figure 5.18 shows the representation of the dataset over time 55 and 99.

Looking at the visualized network in figure 5.18 over different time slices, we observe the following

Observation 15 – Over long period of time with the same interaction period (1000) and network size (10 nodes and 18 edges), nodes start to cooperate but never converge to specific policy (Cooperate or defect), at time 99 two nodes (1 and 0) out of 10 modes seem to cooperate.

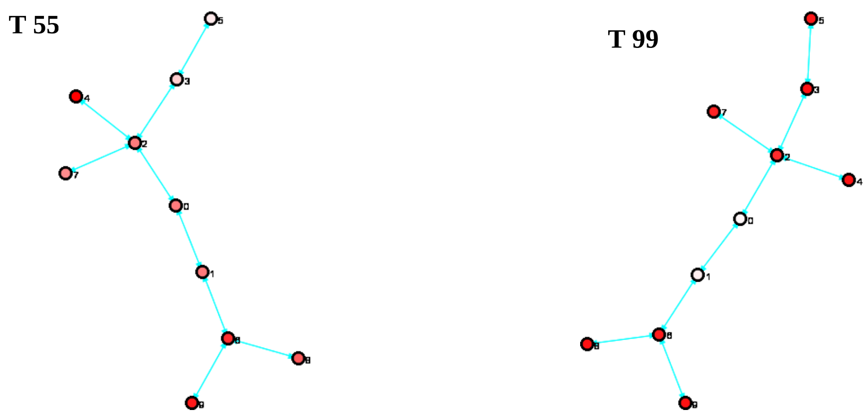


Figure 5. 18: Visualization of DNP2 type network at time 55 and 99

To explain and understand the behavior of the network, nodes' policy/reward changing timeline have been generated in figures 5.19 and 5.20 respectively.

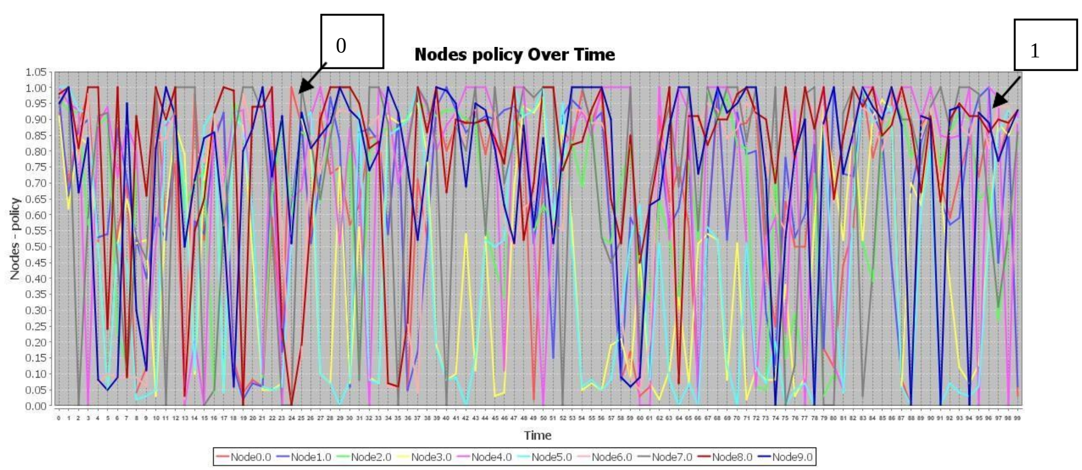


Figure 5. 19: Nodes policy changes timeline - DNP2

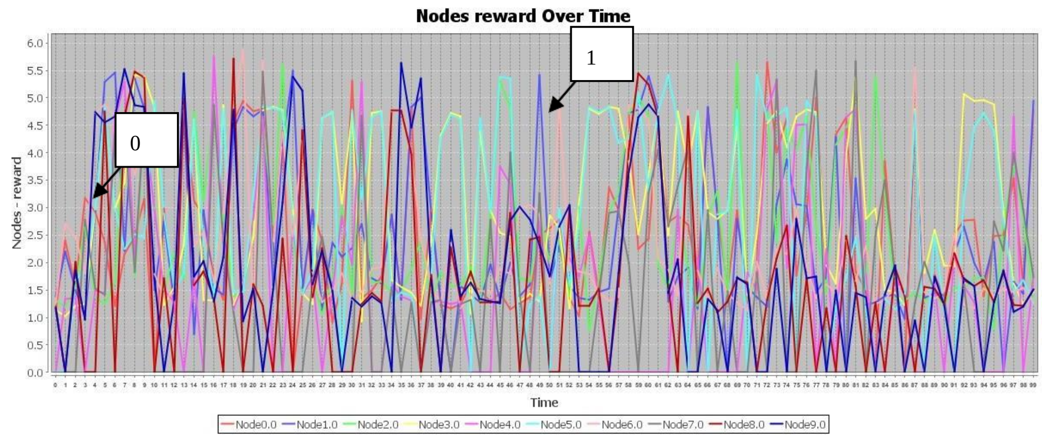


Figure 5. 20: Nodes reward changes timeline – DNP2

Investigating figure 5.19 and 5.20 , we notice that all nodes never converge to use specific policy and keep looping between the cooperating and defecting policies, even node 0 and 1 never converge to cooperate and this confirm observation 4.

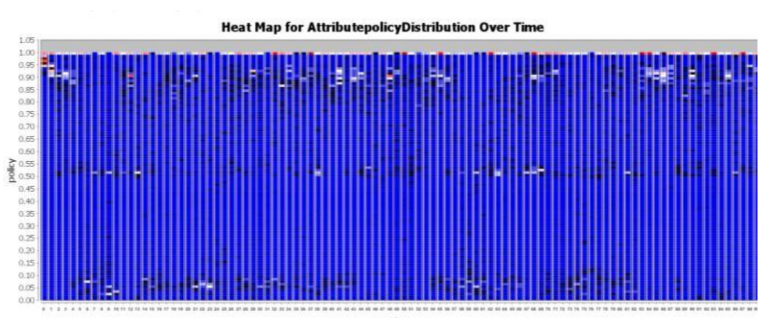


Figure 5. 21: Heatmap represents policy frequency over time - DNP2

Figure 5.21 displays the distribution of policies of all nodes all over the time following a gradient color vary from dark blue to dark red, where dark red represents most frequent policy, dark blue represents the least frequent policy and black color represent zero frequency, by deeply investigating the graph we notice that,

Observation 16 – node’s policy vary from 0 (cooperate) to 1 (defect), but it is clear that policy 1 is more frequent than policy 0 so some of the nodes tend to defect rather than cooperate.

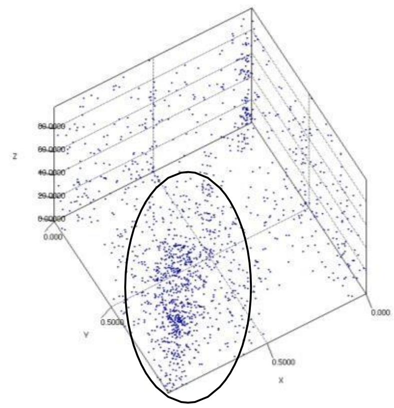


Figure 5. 22: 3-D Visualization of edges over time based on policy attribute

Finally, figure 5.22 shows the 3-dimensional visualization of networks edges, where each point represent an edge (x: is the source's node policy and y: is the target node's policy), this figure confirm observations 5 since the points scattered over policy 0 to 1 over the 100 time period and it is clear that there is intensive amount of points around value 1.

Then, we try to study the effect of interaction period parameter on the same network over long time period, but this time with larger interaction period (100000).

The network of type – (DNP3) has been visualized over different slice of time. Figure 5.23 shows the representation of the dataset over times 25, 50 and 99.

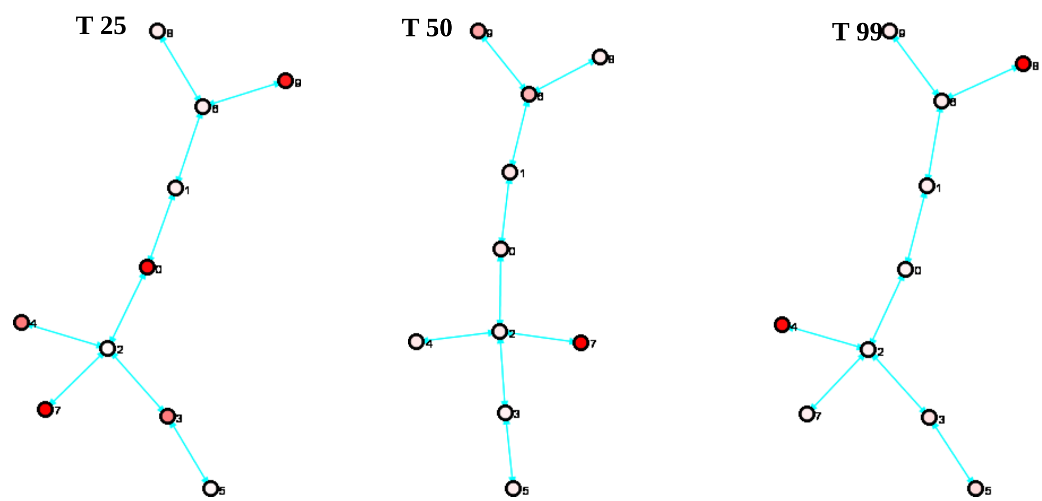


Figure 5. 23: Visualization of DNP3 type network at time 25, 50 and 99

Looking at the visualized network in figure 5.23 over different slices of time, we observe the following

Observation 17 – Number of cooperated nodes increases with time (5 nodes at time 25, 7 nodes at time 50 and 8 nodes at time 99).

Observation 18 – Hub nodes (2 and 6) and inner nodes (1, 0 and 3) cooperate faster than edge nodes.

To explain and understand the behavior of the network, nodes' policy/reward changing timeline have been generated in figures 5.24 and 5.25

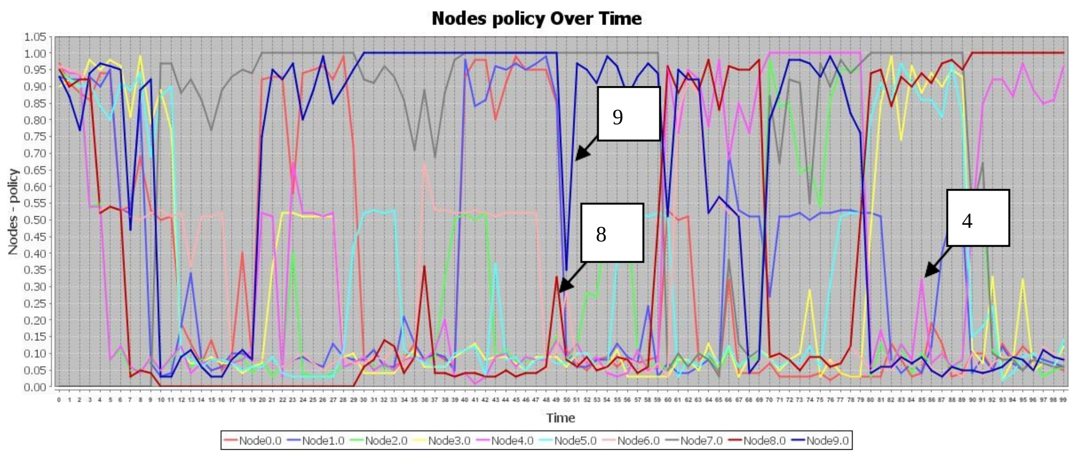


Figure 5. 24: Nodes policy changes timeline – DNP 3

Investigating figure 5.24, we notice that most of the nodes start to converge by following cooperation policy (0), but still we have nodes (4, 8 and 9) that keep looping between the cooperation and defection policies.

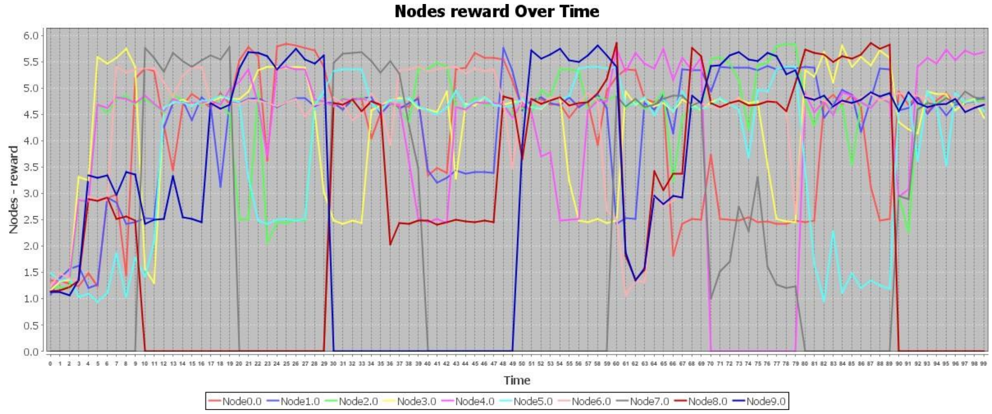


Figure 5. 25: Nodes reward changes timeline – DNP 3

Investigating figures 5.24 and 5.25, we notice that reward for most of the nodes will start gradually increasing within time as nodes starts cooperating.

To study the network and understand its behavior, changes in the key action values Q[0,0,0], Q[0,0,1] and Q[1,1,1] over time are shown in figures 5.26, 5.27 and 5.28.

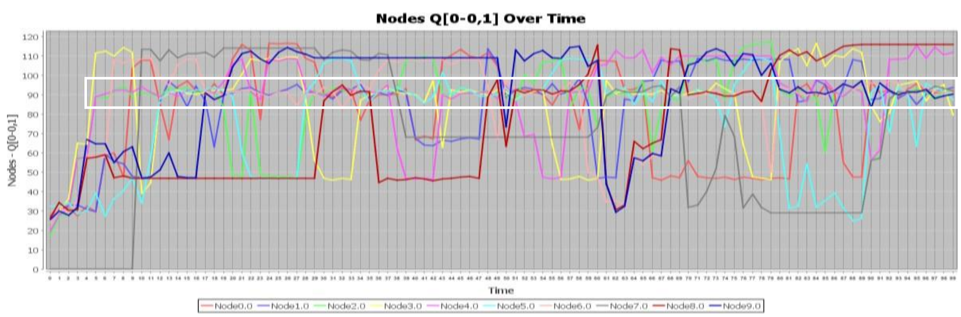


Figure 5. 26: Action value of two nodes cooperated and one defect in next action

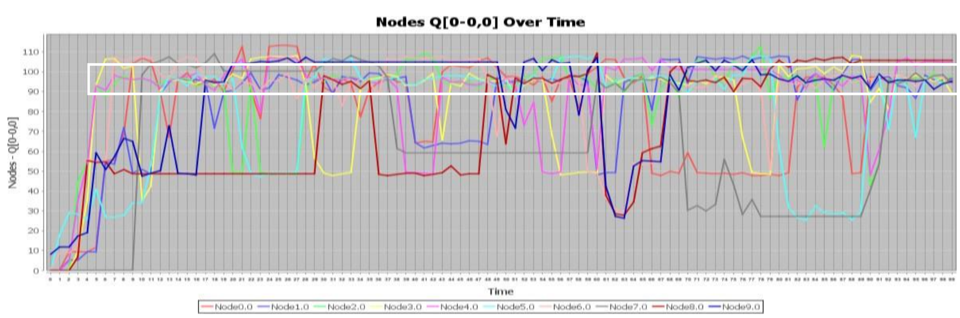


Figure 5. 27: Action value of two nodes cooperated and one cooperate in next action

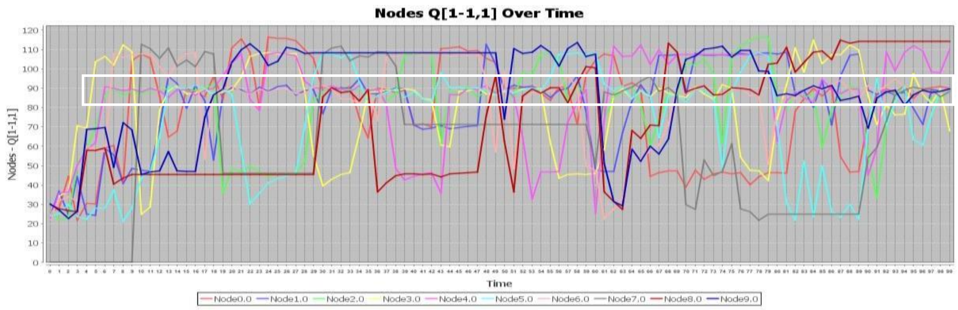


Figure 5. 28: Action value of two nodes defect and one defect in next action

Comparing the action value in figures 5.26, 5.27 and 5.28, we notice that,

Observation 19 – For most of the nodes (indicated by white rectangle), action value of Q[0,0,0] (almost 95) is greater than action value of Q[0,0,1] (almost 90) and action value of Q[1,1,1] (less than 90). Using this observation and since the nodes try to optimize their reward, most of the nodes will tend to cooperate.

To try a larger network, same experiments conducted on datasets of type (DNP1, DNP2 and DNP3) are repeated on adaptive networks of size 30 nodes, 58 edges and over short (1000) and long (1000000) time period using small (1000) interaction period (Network's type DNP 4 and DNP5), following observations are concluded,

Observation 20 – Over short period of time (max time), the observations concluded on adaptive network of size 10 are applied on a network of size 30 and in both cases all nodes defect. However, with larger size network (30-50) some of the nodes might not have the chance to play and ended up with zero reward. See figures 5.29 and 5.30

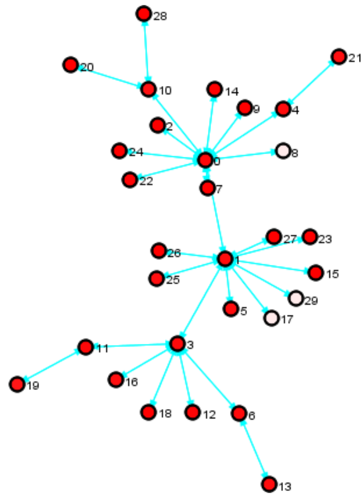


Figure 5. 29: Visualization of DNP4 type network at last time

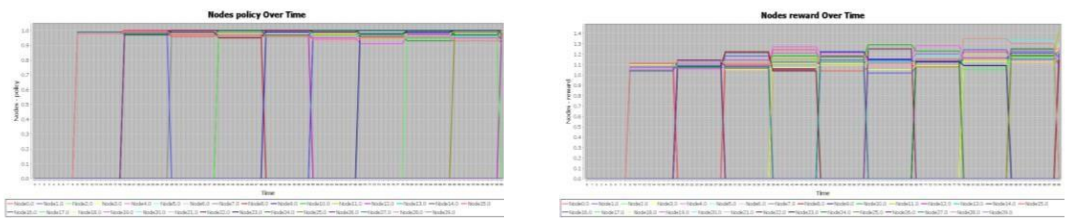


Figure 5. 30: Nodes policy and reward changes timeline – DNP 4

Figure 5.29, represents visualization in the last time for adaptive network with size = 30 nodes and 58 edges, max time =10000 and interaction period=1000, we notice that all nodes converge to defect other than nodes (8, 17 and 29). For more understanding of the network's behavior, nodes changes on policy and reward timeline has been depicted in figure 5.30, studying this figure confirm that most of the nodes defect and it shows that some of the nodes (8, 17 and 29) got no reward because they didn't get the chance to play in the assigned short time period since they are connected to hub nodes of high degree.

Therefore, the following observation concluded,

Observation 21 – Regardless size of the network over a long or short period of time (max time) with short interaction period, observations 1 to 5 applied.

As a result to the concluded findings and observations on adaptive networks of type (DNP1, DNP2, DNP3, DNP4 and DNP5) and to study in depth the effect of interaction period parameter, in the remaining part of this analysis a comprehensives study will focus on network of size 30 nodes and 58 edges, with max time=5000000 and interaction period vary from 50000 to 5000000.

A dataset with max time = 5000000 has been generated and tested over different interaction periods (50000, 200000, 500000, 1000000, 2000000, 2500000, 3000000, 4000000
and 5000000), the analysis focuses on capturing information about the behavior of the network (cooperating or defecting) using the adjusted parameters.

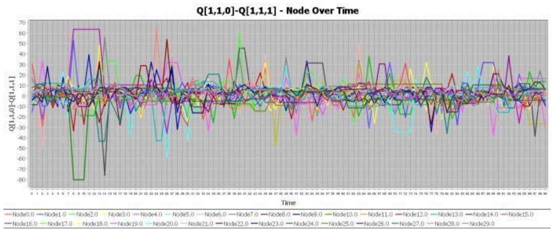
Using nodes' attributes, two main equations will give us better understanding of the studied network's behavior. The first equation can be formulated as,

Action net reward = Action value Q[1,1,0] - Action value Q[1,1,1]

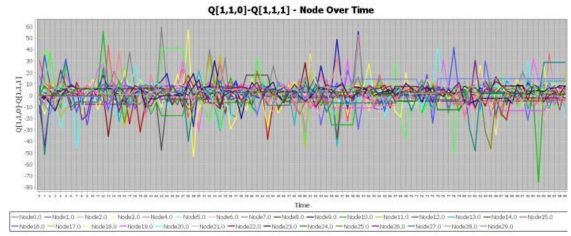
Using equation 1, if the action net reward is greater than 0 this will give us implication that when the current stat for two nodes is defecting, selecting cooperation as next action will increase nods' reward.

Using equation 1, to study the effect of interaction period parameter in scale free adaptive network playing prisoner's dilemma, the nodes attribute timeline graph is generated to visualize the action net reward for each node over time as shown in figure 5.31

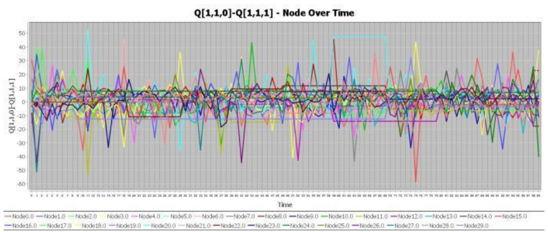
Interaction time = 50000



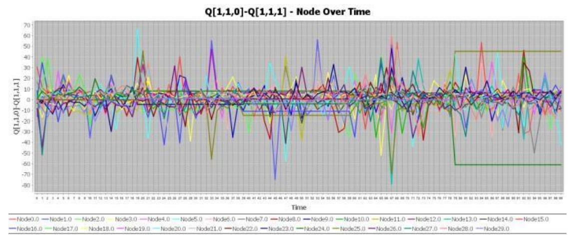
Interaction time = 200000



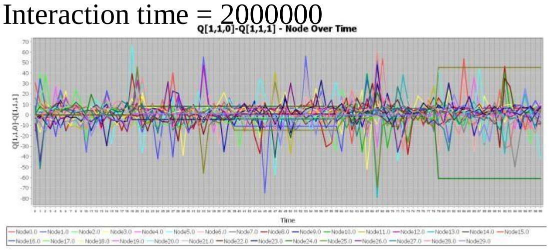
Interaction time = 500000



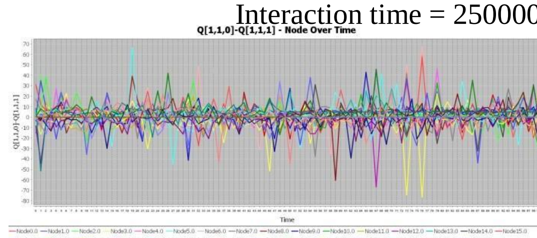
Interaction time = 1000000



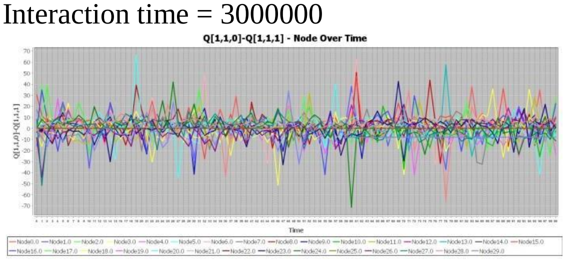
Interaction time = 2000000



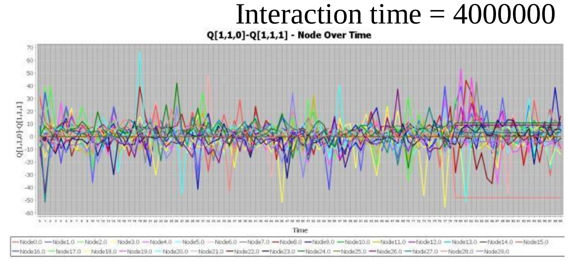
Interaction time = 2500000



Interaction time = 3000000



Interaction time = 4000000



Interaction time = 5000000

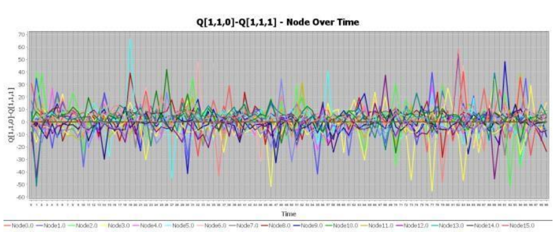
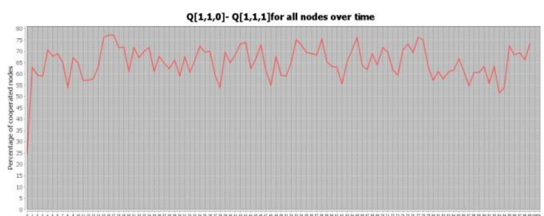


Figure 5. 31: Using Eq. 1 - nodes net action reward timeline

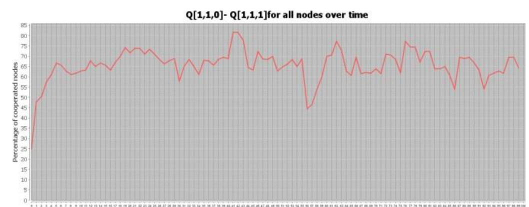
Using figure 5.31, it is clear that over different slices of time some of the nodes have action net reward greater than 0, other nodes have action net reward less than 0 and for some nodes the action net reward is 0. Also, it observable that for some nodes action net reward vary between (–) value, 0 and (+) value, however using figure 5.31 doesn't give us clear pattern about the percentage of cooperated/defected nodes over different slices of time. Therefore, using equation 1 new charts are generated to capture a pattern that makes significant enhancement on studying and analyzing the effect of interaction period parameter.

Figure 5. 32, represents the percentage of nodes tend to cooperate over time using different interaction periods

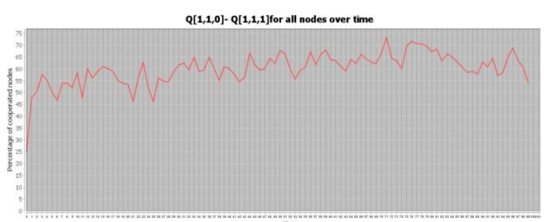
Interaction Period = 50000



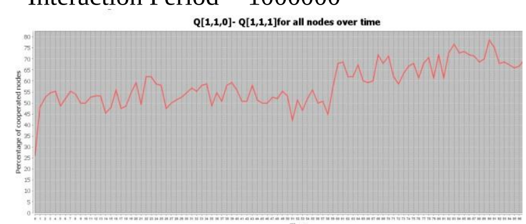
Interaction Period = 200000



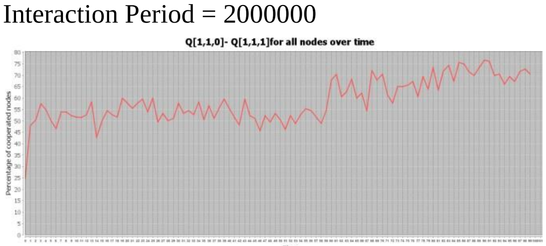
Interaction Period = 500000



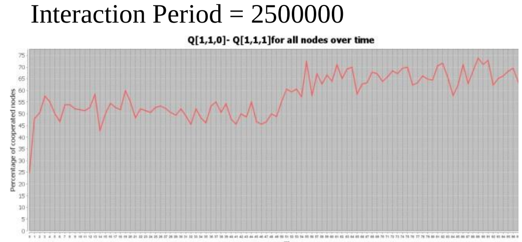
Interaction Period = 1000000



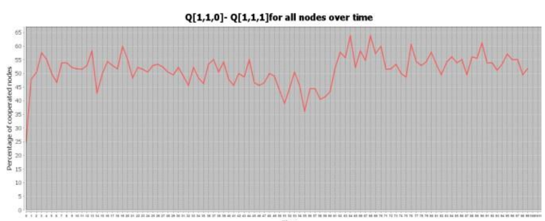
Interaction Period = 2000000



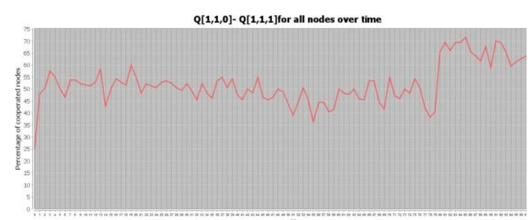
Interaction Period = 2500000



Interaction Period = 3000000



Interaction Period = 4000000



Interaction Period = 5000000

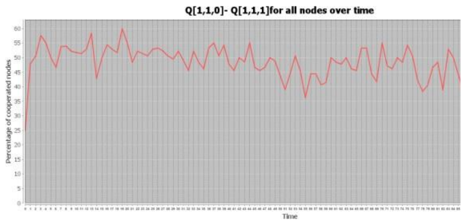


Figure 5. 32: Using Eq. 1 - Percentage of nodes tend to cooperate over time

Studying figure 5.32 in depth, table 5.6 illustrates summary about the network behavior over different interaction period, notice that all provided numbers are approximate numbers estimated from the visualized graphs in figure 5.32 and abbreviation of the used measurements are described in table 5.5.

Measurement	**Abbreviation**
Interaction Period	ITP
Maximum percentage of cooperated nodes	MPCN
Slice of time in which MPCN occurs	TMPCN
Minimum percentage of cooperated nodes	MnPCN
Slice of time in which MnPCN occurs	TMnPCN
Percentage of cooperated nodes at time 5000000	Tlast
Is Percentage of cooperated nodes increases by time	Yes/No
Percentage	%

Table 5. 5: Used measurment and abbreviations

ITP	**MPCN**	**TMPCN**	**MnPCN**	**TMnPCN**	**Tlast**	**Yes/No**
50000	75%	14 to 16, 65, 77 to 78	52%	8, 38, 48, 62, 84 to 85	73%	No, loop between max and min %
200000	82%	41 to 42	45%	56	65%	No, loop between max and min %

500000	73%	71, 75 to 79	45%	21, 25	55%	Almost yes
1000000	77%	90 to 91	43%	14	70%	Yes
2000000	77%	90 to 91	43%	14	70%	Yes
2500000	74%	89 to 91	43%	16	62%	Yes
3000000	64%	64, 68	37%	55	52%	Almost yes
4000000	75%	81 to 85 91 to 92	35%	55	70%	Starts increasing after time 74
5000000	60%	19	36%	55	51%	No

Table 5. 6: Using Eq.1 over different interaction periods results

Table 5.6 illustrates the results of using different interaction periods over a 30 nodes size adaptive network with max time = 5000000, referring to this table, it is observable that,

Observation 22 - Using interaction period starting from 50000 to 200000, maximum percentage of cooperated nodes occurs at different time period (beginning, middle and end), similarly the minimum percentage of cooperated nodes. Thus, the nodes keep looping between cooperating and defecting policy.

Observation 23 - Starting from interaction period = 500000, percentage of cooperated nodes starts increasing with time reaching time 79 with percentage = 73%, but it then drops sharply to 55% at time 100.

Observation 24 - Using interaction period vary from 1000000 to 2000000, percentage of cooperated nodes almost increases proportionally with time and reach maximum value of 77% at time 90. Thus, almost more than 70% of the nodes converged to cooperate using interaction period vary in a range from 1000000 to 2000000.

Observation 25 - Similarly, using interaction period = 2500000, percentage of cooperated nodes almost increases proportionally with time and reach maximum value of 74% at time 90 but it then drops to 62%.

Observation 26 - Starting from interaction period = 3000000 to 4000000, maximum percentage of cooperated nodes increases with time but its value decreases comparing it to maximum value in the previously tested interaction period.

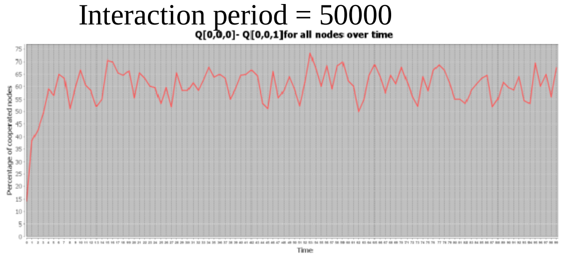
Observation 27 - Using interaction period equivalent to max time, the value of the maximum percentage (60%) of cooperated node decreases comparing it to maximum value in the previously tested interaction period.

Similar to equation 1, intensive analysis will be conducted to equation 2 which can be formulated as, *Action net reward = Action value Q[0,0,0] - Action value Q[0,0,1]*

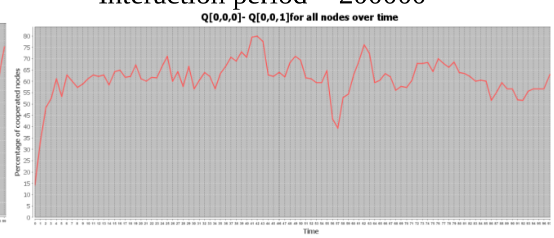
Using equation 2, if the action net reward is greater than 0 this will give us implication that when the current stat for two nodes is cooperating, selecting cooperation as next action will increase nods' action net reward.

Similar to the charts generated using equation 1, figure 5.33 represents the percentage of nodes tend to cooperate over time using different interaction periods

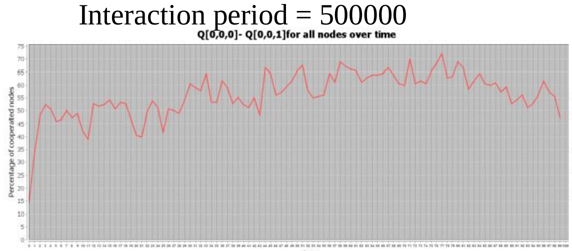
Interaction period = 50000



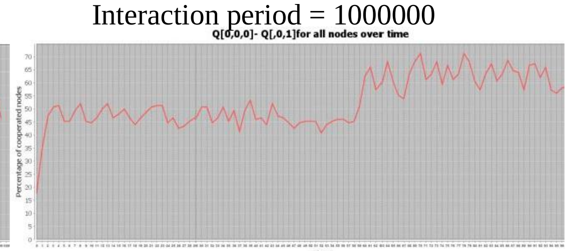
Interaction period = 200000



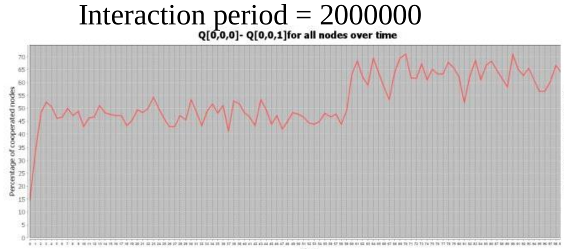
Interaction period = 500000



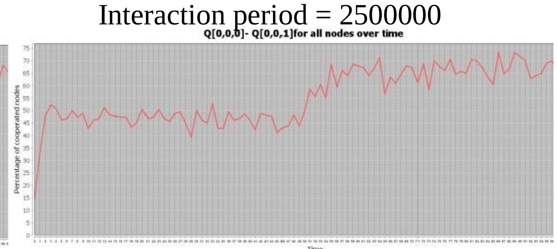
Interaction period = 1000000



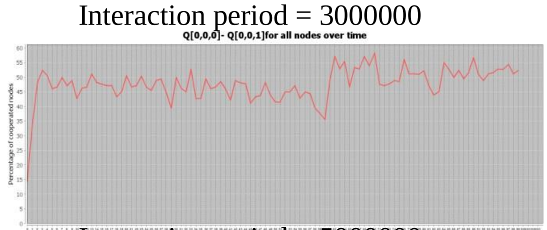
Interaction period = 2000000



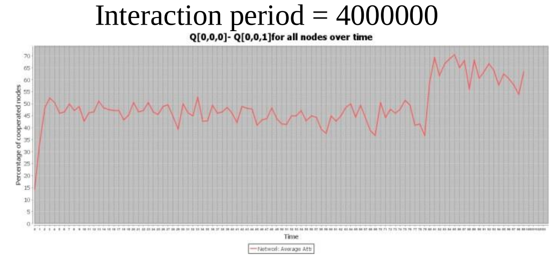
Interaction period = 2500000



Interaction period = 3000000



Interaction period = 4000000



Interaction period = 5000000

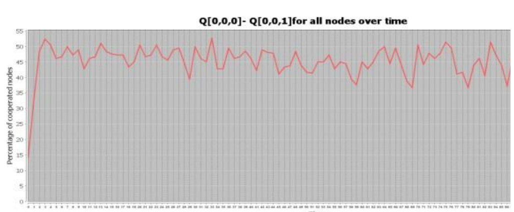


Figure 5. 33: Using Eq. 2 - Percentage of nodes tend to cooperate over time

Studying figure 5.33 in depth, table 5.7 illustrates summary about the network behavior over different interaction period, notice that all provided numbers are approximate numbers estimated from the visualized graphs in figure 5.33 and abbreviation of the used measurements are described in table 5.5

ITP	MPCN	TMPCN	MnPCN	TMnPCN	Tlast	Yes/No
50000	73%	53	50%	8, 62, 87	68%	No, loop between min and max %
200000	80%	42 to 43	39%	57	62%	No, loop between min and max %
500000	72%	77	40%	11	47%	Yes up to time 77, it then starts decreasing
1000000	72%	90	41%	37	63%	Yes
2000000	72%	90	41%	37	64%	Yes
2500000	73%	90	39%	39	60%	Yes
3000000	57%	70	35%	58	52%	Almost yes
4000000	70%	85	35%	79	63%	Starts increasing after time 79
5000000	53%	35, 85	37%	69	41%	No

Table 5. 7: Using Eq.1 over different interaction periods results

Table 5.7 illustrates the results of using different interaction periods over a 30 nodes size network with max time = 5000000, referring to the table, it is observable that,

Observation 28 - Using interaction period starting from 50000 to 200000, maximum and minimum percentage of cooperated nodes occurs at different time period (beginning, middle and end). Thus, the nodes keep looping between cooperating and defecting policy.

Observation 29 - Starting from interaction period = 500000, percentage of cooperated nodes starts increasing with time reaching time 77 with percentage = 72%, but it then drops sharply to 47% at time 100.

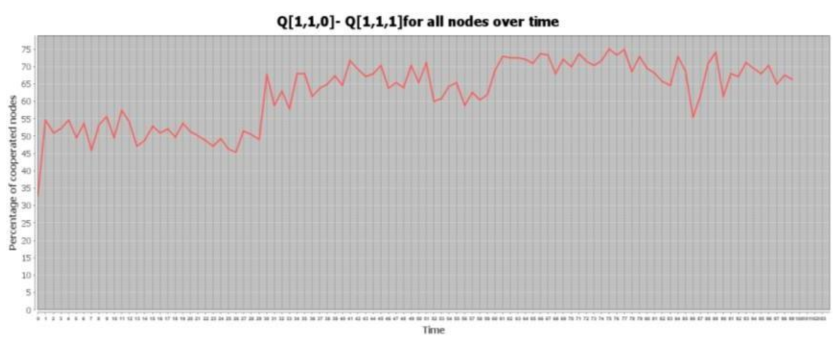
Observation 30 - Using interaction period vary from 1000000 to 2500000, percentage of cooperated nodes increases directly with time and reach maximum value of 72-73% at time 90. Thus, almost more than 70% of the nodes converged to cooperate using interaction period vary in a range from 1000000 to 2500000.

Observation 31 - Starting from interaction period = 3000000 to 4000000, maximum percentage of cooperated nodes increases with time but its value decreases comparing it to maximum value in the previously tested interaction periods.

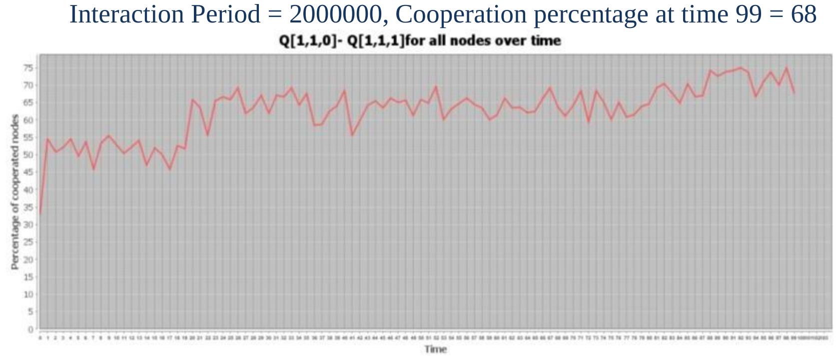
Observation 32 - Using interaction period equivalent to max time, the value of the maximum percentage (53%) of cooperated node decreases comparing it to maximum value in the previously tested interaction periods.

For reliable outcome, the experiment repeated on the same dataset but over longer max time = 10000000 and using different interaction periods, figure 5.34 is depicted to show the action net reward using equation 1 and the following observations found,

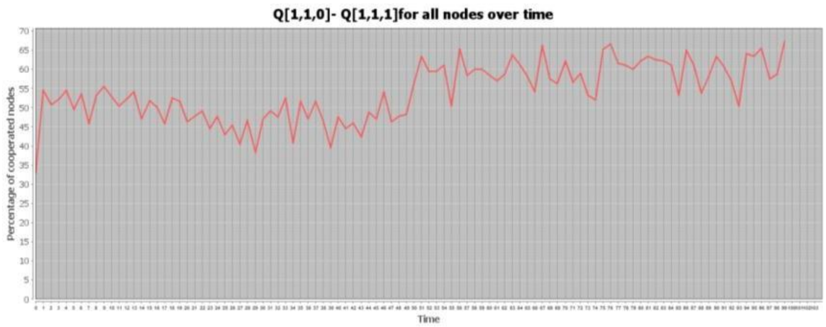
Interaction Period = 1000000, Cooperation percentage at time 99 = 66.3



Interaction Period = 2000000, Cooperation percentage at time 99 = 68



Interaction Period = 5000000, Cooperation percentage at time 99 = 67



Interaction Period = 5000000, Cooperation percentage at time 99 = 67

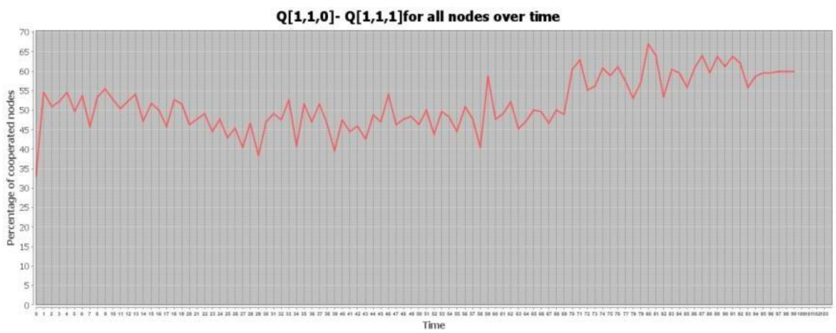


Figure 5. 34: Using Eq. 1 percentage of cooperation over time

Observation 33 – Over longer max time, the preferred interaction period is almost the same which is 1000000 to 2000000 since the percentage of cooperated nodes increases proportionally with time.

After generating, visualizing and analyzing the 6 different types of proposed adaptive networks (DNP1, DNP2, DNP3 and DNP4, DNP5 and DNP6).

Using observation **12 to 16** and observations **20 to 21**, result 6 can be formulated as follow,

Result 6 - Using prisoner's dilemma game and Q learning algorithm, using short interaction period regardless **size** of the network (Small or large size**)** and **max time** (short or long period of time**)**, most of the nodes will either follow defecting policy or never converge and keep looping between defecting and cooperating policy.

Using observation **17 and 19**, results 7 can be formulated as follow

Result 7 - Using prisoner's dilemma game and Q learning algorithm in **Small size** (10 nodes) over **long max time** (starting from 1000000) and **long interaction period** (starting from 100000), number of cooperated nodes increase proportionally with time.

Using observation **18** and observations **22 to 30**, result 8 can be formulated as follow

Result 8 - Using prisoner's dilemma game and Q learning algorithm, under proper adjustment of interaction period, max time and network size parameters **hub** and **inner** nodes of the adaptive network converge to cooperation faster than **edge** nodes.

Using observations **27 and 32,** result 9 can be formulated as follow

Result 9 - Using prisoner's dilemma game and Q learning algorithm, using long max time regardless size of the network and having interaction period equivalent to max time, the maximum percentage of cooperated nodes will decrease.

Using observations **22 to 33,** result 10 can be formulated as follow

Result 10 - Using prisoner's dilemma game and Q learning algorithm in a network of size 30 nodes over long max time starting from 5000000, the best value of inter- action period is vary from 1000000 and 2500000.

Using observations **22 to 33,** result 11 can be formulated as follow

Result 11 - Using prisoner's dilemma game and Q learning algorithm in any adaptive network and over long max time, interaction period is the main parameter that makes significant change in the network's cooperation's behavior.

Researchers on Biology field have plenty of attempts to study the cooperation behavior in human and Organisms. Melis, A. and Semmann, D. 2010 [29] studied approaches in which human and animals interact to evolve cooperation and as a result they found that cooperation in human maintains by keeping track of previous collaborative inter- actions with other human over time and transferring the information to others which leads to maintain cooperation.

Also, Stevens, R., Cushman, F. and Hauser, M. [30] shows that maintaining cooperation in human has many constrains and interaction time is one of the key factors to obtain cooperation. The researchers' findings confirm **result 11** of this thesis.

Using **result 11** which is concluded from observations **22** to **33** and referring to the biology researchers findings as described in Melis, A. and Semmann, D. 2010 [29] and Stevens, R., Cushman, F. and Hauser, M. [30] result 12 can be formulated as follow

Result 12 – Studying the behavior of learning to maintain cooperation in artificial networks and social networks, interaction period parameter consider as one of the main parameters that has significant effect on cooperation behavior of human and non-human agents.

5.3 Case Study 3 –Taxi Dispatch System

In this section, real dataset for Taxi dispatch system using multi-agent self-organization technique [5] will be discussed, studied and analyzed using DNVA tool.

Description about the used dataset, the applied technique to improve taxi dispatch system performance, the objectives and findings are described in section 5.3.1. Visualization and analysis of the dataset before and after applying the proposed technique will be discussed in section 5.3.2

5.3.1 Dataset Exploration

Taxi dispatch system is used to assign vacant taxi to a customer in different locations. The mechanism of work in the current taxi dispatch system is to divide the city into regional dispatch area and adjacent areas are assigned to each dispatch area manually by human expert. The assigned dispatch area are used when there is no vacant cabs in the local dispatch area however using fixed adjacent area approach doesn't take into consideration any modification in road and traffic system. Multi-agent self-organization technique is applied to the taxi dispatch system to enable dynamic allocation of adjacent areas and it mainly aims to (1) maximize number of served calls, (2) maximize number of dispatched calls (2). In the tested dataset, 141 dispatch areas are represented by nodes and adjacent connections are represented by edges.

5.3.2 Dataset Analysis

The analysis started by setting nodes' degree as color visualization, figure 5.35 represents the current dispatch system without enhancement by applying the proposed technique. As it is observable, more than 27 isolated areas (indicated by red rectangles) have no adjacent area so this might leads to either overload in the local area due to lake of vacant cabs and disability of using vacant cabs from another area or improper utilization of vacant cabs from other areas.

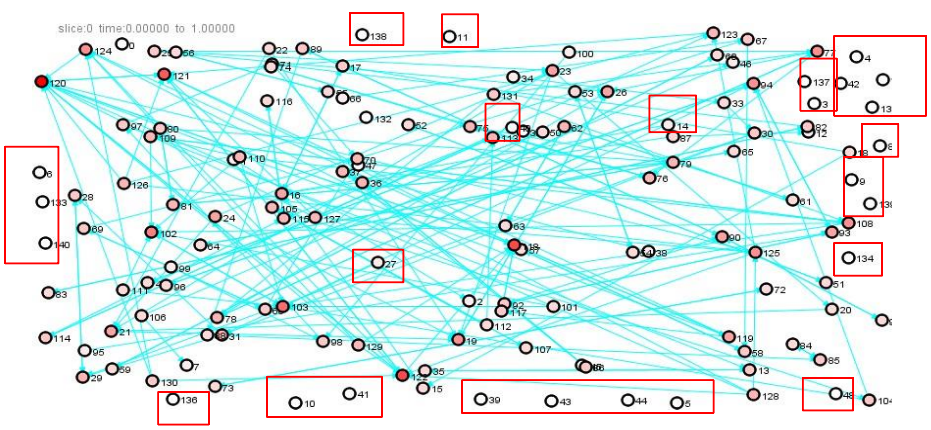


Figure 5. 35: Visualization of Taxi Dispatch system at time 0

Figure 5.36 represents visualization of the network after applying multi-agent self-organization technique for performance optimization purposes at the last time.

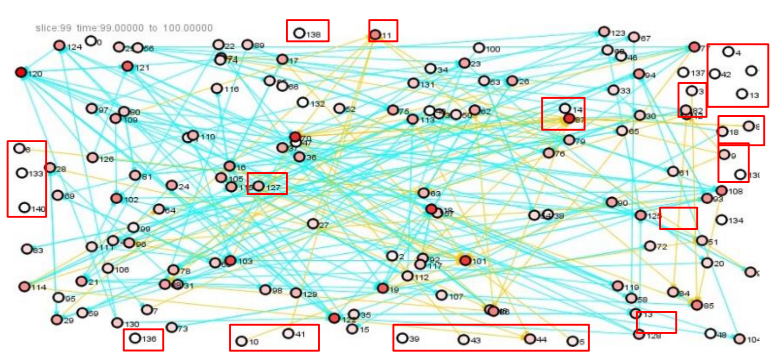


Figure 5. 36: Visualization of Taxi Dispatch system at time 99

In figure 5.36, orange edges represent new adjacent connection added dynamically with time due to the caller (customer) need for service. Investigating the figure, it is observable that,

Observation 34 - many isolated areas (nodes 11, 139, 41, 10, 44, 5, 43, 9, 8, 3, 6 and 27) are connected to one or more area as adjacent, also many areas with low degree (has less number of adjacent area) like nodes 87 and 101 connect to new area as adjacent.

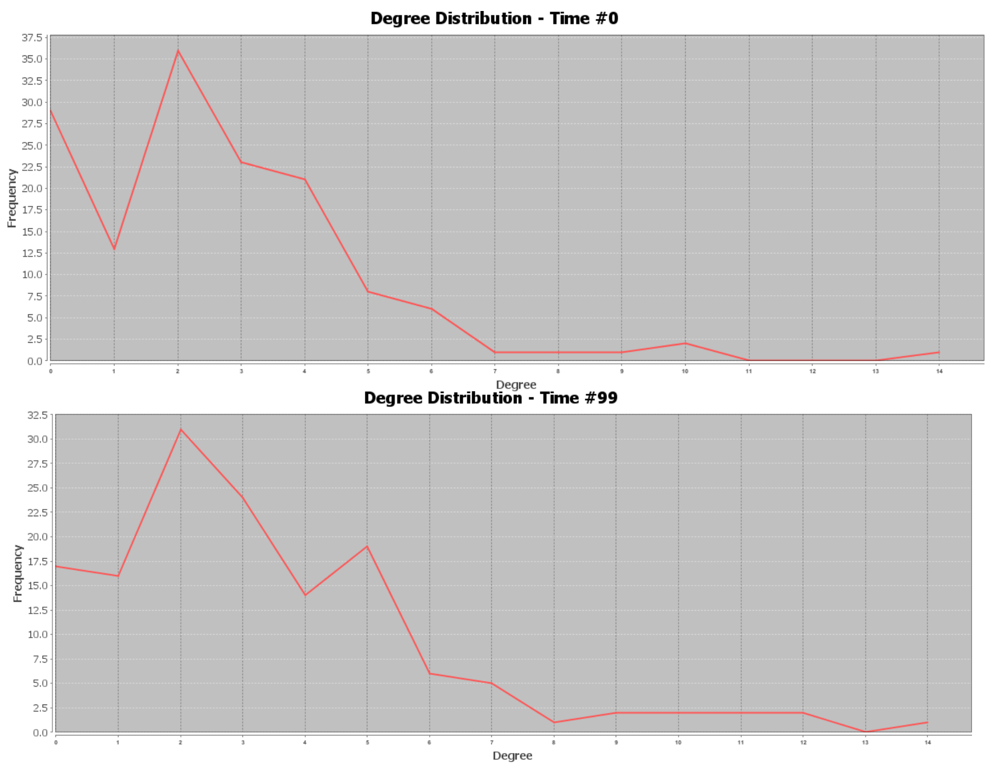


Figure 5. 37: Taxi Dispatch System degree distribution at time 0 and 99 Figure 5.37 represents the degree distribution for all nodes (areas) at time 0 and the corresponding at time 99, it is observable that,

Observation 35 – With time frequency of high degrees (degree 11 to 14) increases at the last time but it is almost zero at time 0 before applying the multi-agent self-organization technique. This observation confirms observation 34.

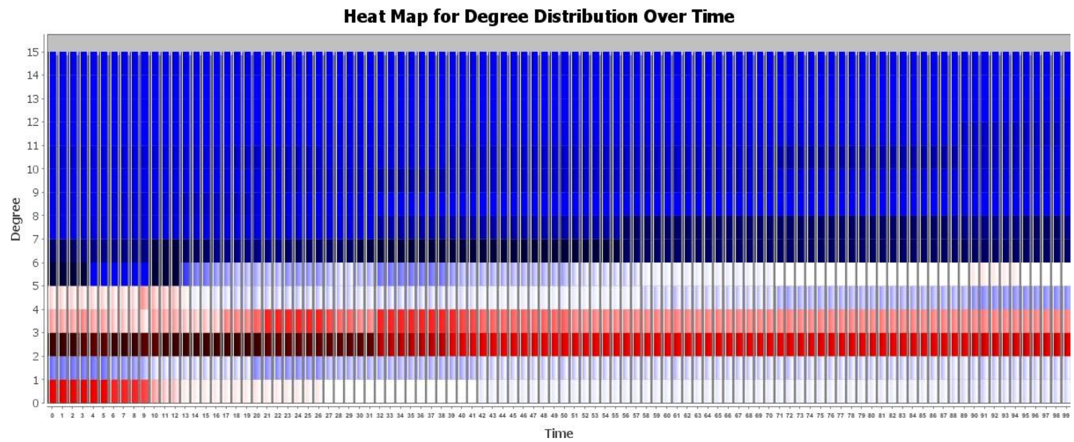


Figure 5. 38: Heatmap for degree frequencies over time

The degree distribution heat map is depicted in figure 5.38, where gradient color vary from dark red to dark blue represents high degree frequency and low degree frequency respectively and black color represent zero frequency. It is observable that,

Observation 36 – frequency of high degree increases with time. This is confirming observation 34 and 35.

Using observations 34, 35 and 36, result 12 can be formulated as follow

Result 13 – As time increases and using multi-agent self-organization technique, adjacent areas for each local area increase, this will lead to **better utilization of** vacant cabs in most of areas, thus **serving larger number** of callers **and increasing** number of **dispatch calls.**

The findings of this section analysis confirmed by the findings in "Multiagent Self-Organization for a Taxi Dispatch System" by Aamena Alshamsi, Sherief Abdallah, Iyad Rahwan [5], which states that “The proposed technique decreases the total waiting time by up to 25% in comparison with the real system and increases taxi utilization by 20% in comparison with results of the simulation without self-organization”.

5.4 Research Questions Discussion

In this section, the analysis and experiments conducted in section 5.3 will be used to show how we answer the research questions raised in section 1.3.

Can we explain the behavior of the isolated nodes that keeps looping between

different policies in Abdallah, S. 2012 [4]?

Referring to the **results** of the analysis in case study 1 in section 5.1, which focuses on studying the behavior of multi-agent system over scale free network and using Q-learning algorithm, we notice that visualization and analyzing a network similar to tested network in Abdalla, S. 2012 [4] and under similar conditions will result in a 3-D graph (see figure 5.4) similar to the visualized 3-D graph in Abdalla, S. 2012 [4] (see figure 5). Using DNVA open source tool features to visualize all nodes' policy time-line as in figure 5.2 and using different analysis techniques, the isolating nodes which keeps looping are identified as formulated in section 5.1, **Result 2** of this thesis.

Can we provide explanation for the occurrence of improvement in the performance of taxi Dispatch system using multi-agent self-organization technique as de- scribed in Alshamsi, A. Abdallah, S. and Rahwan, 2009 [5]?

Referring to the **result** of the analysis in case study 3 in section 5.3, which focuses on studying the dataset of a taxi-dispatch system uses multi-agent self-organization technique as described in Alshamsi, A. Abdallah, S. and Rahwan, I. 2009 [5], we notice that visualizing the network using degree measurement as visualization color's attribute in time 0 shows the existence of many isolating areas (more than 27 areas) which represents the network before applying the proposed improvement technique called multi- agent self-organization technique (see figure 5.35). By visualizing the network at different time reaching the last time (see figure 5.36) an interesting observation shows that number of isolating areas decreases using the proposed improvement technique with time. The observations on the visualization are confirmed by graphs generated using other DNVA open source tool and the final result is formulated in **section 5.3**, **Result 12** of this thesis.

Can we provide in-depth analysis of a multi-agent system that has the ability to learn and adapt their policy over time?

Referring to **result 1** to **result 12** illustrated and concluded from analysis and discussion of case studies 1, 2 and 3 in sections 5.1, 5.2 and 5.3 respectively, we can say that many types of multi-agent systems (see table 5.2 and table 5.4) using coordination or prisoner's dilemma games over scale free network have been studied and analyzed. As a result many findings, observations and conclusions have been derived to enrich the recent area of research focusing on studying and analyzing artificial networks (multi-agent system learn to adapt its agents' policies over time).

Is there any similarity between the behavior of artificial evolving network learning over time and human or social network?

Referring to the intensive study done on section 5.2 and focusing particularly on capturing information about defection and cooperation behavior of 30 nodes (agents) dataset playing coordination game, using Q-learning algorithm and running with max time = 5,000,000 over different interaction periods =[50000, 200000, 500000, 1000000,
2000000, 2500000, 3000000, 4000000 and 5000000]. Generating x, y line charts representing the percentage of cooperation of all nodes in each time using different interaction periods, statistics are illustrated in table 5.6 and 5.7. Studying the findings on the table we notice that percentage of cooperation increases and decreases according to the adjusted interaction period (time). Referring to Biology's researchers' findings to study the cooperation behavior of human in life [29, 30] who elaborated that interaction time between human is an essential factor maintain cooperation between human under different conditions. Thus there is similarity between artificial networks and social networks in maintaining cooperation behavior.

Chapter 6

Further Work and Conclusion

Stimulated by the recent intensive use of multi-agent system, this thesis has investigated the dynamics behavior of learning multi-agent system (artificial networks) evolving over time. The purpose of the current study was to provide an-in depth analysis of scale – free artificial networks having adaptive agents (nodes) learn to improve and adapt their policies over time using Q-learning algorithm and playing either prisoner's dilemma or coordination games. Also, the study was designed to explain the behavior of previous researchers' unjustified observations. To accomplish this thesis's aims, an open source tool for visualizing and analyzing artificial networks has been implemented, tested and used over three conducted case studies. In each case study and for the purpose of reliable outcomes, more than 10 datasets of different tested artificial networks' types have been tested and observations on the main factors affecting the dynamics behavior of the tested networks have been made to formulate the final results of this thesis.

One of the more significant findings of this thesis' study is that interaction period (time) between agents in artificial network is the major factor affecting maintaining co-operation in the network. Similarly and comparing this finding with biology researchers' studies on social network (human), we found that interaction period consider as one of the main factors affecting the cooperation behavior in human. Also, this thesis has provided explanation of previous researcher's observation on having isolated nodes that never converge [4], the study has shown that this behavior is a result of having a node between two opposite policies hub (high degree) nodes. The present study provides additional evidence with respect to the findings of a previous research done by Alshamsi, A. Abdallah, S. and Rahwan, 2009 [5] and confirm that applying multi-agent self-organization technique helps in optimizing the performance on taxi dispatch system. Finally, the study on this thesis reveals conditions under which global norm or convention in multi-agent system will emerge in scale free networks of different sizes, over different maximum time, playing prisoner's dilemma or coordination games and using state or stateless Q-learning algorithm.

As a future work, a number of possible studies using the same experimental set up are apparent. In this thesis, the research is limited to study scale-free artificial network playing prisoner’s dilemma or coordination game and using Q- learning algorithm. Therefore, it would be interesting studying other types of networks like random or regular artificial networks and compare the findings of the different types of networks. Also, future research might explore and investing the behavior of artificial networks use multi-agent reinforcement learning algorithms other than Q-learning algorithms proposed in this thesis. Further experimental investigations are needed to conclude the prime factors affecting the behavior of artificial networks. Finally, this research has thrown up many questions in need of further investigation and deep comparison to study the similarity in the factors affecting the behavior of social and artificial network.

Bibliography

[1] Abdallah, S. and Lesser, V., 2007. *Multiagent Reinforcement Learning and Self-Organization in a Network of Agents.* Honolulu, AAMAS.

[2] Abdallah, S. and Lesser, V., 2006. *Learning the Task Allocation Game. in the proceedings of the Fifth International Joint Conference on Autonomous Agents and Multi-agent Systems,* AAMAS, Hokkaido, Japan, pp. 1-8.

[3] Abdallah, S., 2010. *Using Graph Analysis to Study Networks of Adaptive Agent. in the Proceedings of 9th International Conference on Autonomous Agents and Multiagent Systems*, AAMAS, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), Toronto, Canada, pp. 1-7.

[4] Abdallah, S., 2012. *Using a hierarchy of coordinators to overcome the frontier effect in social learning. in the proceedings of the Eleventh International Conference on Autonomous Agents and Multiagent Systems.* Valencia, Spain, s.n., pp. 1-7.

[5] Alshamsi, A., Abdallah, S. and Rahwan, I., 2009. *Multiagent Self-organization for a Taxi Dispatch System. in the Proceedings of 8th International Conference on Autonomous Agents and Multiagent Systems*, AAMAS, Budapest, Hungary, pp. 1-8.

[6] Tao, N., Baxter, J. and Weaver, L., 2001. *A Multi-Agent, Policy-Gradient approach to Network Routing. in the proceedings of the 18th Int. Conf. on Machine Learning,* Morgan Kaufmann, San Francisco, CA, USA, pp. 553--560.

[7] Vincent, C. and Sandholm, T., 2007. AWESOME: A general multiagent learning algorithm that converges in self-play and learns a best response against stationary opponents. *Mach Learn*, Vol. 67:23–43.

URL: http://dx.doi.org/10.1007/s10994-006-0143-1

[8] Tolba, A., 2007. *Scale Free Networks; A Literature Review. In the Online Proceeding of Seventh International Conference on Complex Systems,* Boston, MA, pp. 1-5.

[9] Abdallah, S., 2009. *Generalizing Unweighted Network Measures Using the Effective Cardinality. in the 3rd SNA-KDD Workshop* (SNA-KDD'09), Paris, France: ACM.

[10] Phelan, S. E., 1997. *Using artificial adaptive agents to explore strategic landscapes* [online]. Ph. D. Thesis. La Trobe University. [Accessed 15 March 2012]. Available at: http://faculty.unlv.edu/wpmu/

[11] Hein, O., Schwind, M. and Ko¨nig, W., 2006. Scale-Free Networks. The Impact of Fat Tailed Degree Distribution on Diffusion and Communication Processes. *Wirtschaftsinformatik*, Vol. 48(4), pp. 267-275.

[12] Barabási, A. and Bonabeau, E., 2003. Scale-Free Networks. *Scientific American Magazine*, 14 April, pp. 50-59.

[13] Bellomi, F. and Bonato, R., 2005. *Network Analysis for Wikipedia. In the proceedings of Wikimania 2005, The First International Wikimedia Conference.* Frankfurt, Germany, Wikimania.

[14] Bohn, A., Feinerer, I., Hornik, K. and Mair, P., 2011. Content-Based Social Network Analysis of Mailing Lists. *The R Journal*, Vol. 3(1), pp. 11-18.

[15] O'Madadhain, J., Fisher, D. and Smyth, P., 2005. Analysis and Visualization of Network Data using JUNG. *Journal of Statistical Software*, Vol.10, pp. 1-35

[16] Bastian M., Heymann S. and Jacomy M., 2009. *Gephi: an open source software for exploring and manipulating networks. In the Proceedings of the Third International ICWSM Conference*, San Jose, California, pp. 1-2.

[17] Mao-Bin Hu, Yong-Hong Wu, Rui Jiang and Qing-Song Wu, 2008. *Dynamic prisoner's dilemma on scale-free network.* Proc. SPIE 6802, 680217

URL: http://dx.doi.org/10.1117/12.769770

[18] CHENG Hong-Yan, DAI Qiong-Lin and LI Hai-Hong et al, 2010. Evolutionary Prisoners' Dilemma Game on Scale-Free Networks with Degree-Degree Correlation. *Chin. Phys. Lett*, Vol. 27.(10), pp. 1-4

URL: http://159.226.36.45/Jwk_cpl/CN/Y2010/V27/I10/100201

[19] Menche, J., Valleriani, A. and Lipowsky, R., 2010. Asymptotic properties of degree-correlated scale-free networks. *Phys. Rev. E.* Vol. 81.(4), pp. 046103: American Physical Society

URL: http://dx.doi.org/10.1103/PhysRevE.81.046103

[20] Szolnokia, A., Perc, M. and Danku, Z., 2007. Towards effective payoffs in the prisoner's dilemma game on scale-free networks. *Physica A.* Vol. 382.(8-9), pp. 2075-2082

URL: http://dx.doi.org/10.1016/j.physa.2007.11.021

[21] Bender-deMoll, S. and McFarland, D., 2006. The Art and Science of Dynamic Network Visualization. *Journal of Social Structure.* Vol. 7(2).

[22] Kamada, K. and Kawai, S., 1989. "An Algorithm for Drawing General Undirected Graphs." *Information Processing Letters* Vol. 31, pp. 7-15.

[23] Sutton, R. and Barto, A., 1998. *Reinforcement Learning: An Introduction.* London, England: The MIT Press.

[24] Kimbrough, S. and Lu, M., 2005. Simple reinforcement learning agents: Pareto beats Nash in an algorithmic game theory study. *Information Systems and e-Buisness Management(ISeB)*. Vol. 3.(1), pp. 1-19

URL: http://dx.doi.org/10.1007/s10257-003-0024-0

[25] Burkov, A., Boularias, A. and Chaib-draa, B., 2007. *Competition and Coordination in Stochastic Games. In the roceedings of the 20th conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence,* Montreal, Quebec, Canada, pp. 26-37: Springer-Verlag.

URL: http://dx.doi.org/10.1007/978-3-540-72665-4_3

[26] Kimbrough, S. and Lu, M., 2005. Coordination Games on Dynamical Networks. *Games 1*. Vol. 3, pp.242-261

URL: http://dx.doi.org/10.3390/g1030242

[27] Calais, G., 2010. Game Theory: Employing the Prisoner's Dilemma to enhance Interdisciplinary Learning. *National Forum for Teacher Education Journal*. Vol. 20.(3), pp. 1-9

[28] Dayan, P. and Watkins, C. J. 2006. Reinforcement Learning: A Computational Perspective. *Encyclopedia of Cognitive Science*. pp. 1-5

[29] Melis, A. and Semmann, D., 2010. How is human cooperation different? *Phil. Trans. R. Soc.* Vol. 365.(1553), pp. 2663–2674.

URL: http://dx.doi.org/10.1098/rstb.2010.0157

[30] Stevens, J., Cushman, F. and Hauser, M., 2005. Evolving the psychological mechanisms for cooperation. *Annual Review of Ecology, Evolution, and Systematics,* Vol. 36, pp. 499-518.

URL: http://dx.doi.org/10.1146/annurev.ecolsys.36.113004.083814