Digital Business Ecosystem

Contract n^{O} 507953

Workpackage 19

P2P Network Behavior

Deliverable 19.3

Network Complexity and Scaling

Information Society Technologies

Project funded by the European Community under the "Information Society Technology" Programme

Contract Number: 507953
Project Acronym: DBE
Title: Digital Business Ecosystem

Deliverable No: D19.3
Due date: 06/01/2006
Delivery date: 09/02/2006

Short Description:
In the deliverable D19.2 (Network Topology from Combining Graph-theoretic and System Requirements), the performance of the P2P FADA network's dependence on the network topology and dynamics of the network has been studied. In this report, the SME business network is modelled as an information ecosystem and the FADA network is the communication infrastructure that supports business processes in the SME network. The dynamic relationship between evolution-generated communications and the performance of the FADA communication network is studied, along with the performance of the DBE system as a whole.

Partners owning: HWU
Partners contributing: HWU
Made available to: Consortium and EC

Versioning		
Version	**Date**	**Author, organisation**
1	09/02/06	Juan Wang, HWU
1	09/02/06	Philippe De Wilde, HWU

Evolution-generated Communications and System Performance

Juan Wang
Philippe De Wilde

Intelligent Systems
Department of Computer Science
School of Mathematics and Computer Sciences
Heriot-Watt University

February 2006

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the modern world, computers are now interconnected to a higher degree and in a more complicated way than before, along with increasing demands upon them and improving technologies. Interaction and adaptation of entities in such a complex, dynamic system occur continuously and concurrently. Consequently, centralized control of the system is no longer applicable and proved to cause many technical problems, due to the failure of adapting to the changing environment. This is opposite to the situation in natural systems. Although behaviours of natural systems are unpredictable and undefined, they always display a high degree of resilience and flexibility. Therefore, it inspires a lot of researchers to apply many aspects of natural systems to different areas. The concept of "information ecosystem" is one of them, which emphasizes the information being exchanged between networked entities rather than the entities themselves. It is titled ecosystem because the information flow in it is an analogy with the flows of materials and energy in a natural ecosystem. An European project called Decentralized Information Ecosystem Technologies (DIET) which aims to develop software for understanding and managing large-scale distributed the information system is based on the concept of information ecosystem [3]. The concept is also been used for the description of a security model in [4].

An information ecosystem may represent a wide range of complex systems including computer networks, associated equipments, IT telecommunications and many others, as long

as there are interactions between information users and providers in the networks. Thus, one way to model the SME network is to regard it as an information ecosystem. In this ecosystem, SMEs manage and communicate with other SMEs. Services and products provided by SMEs are stored in their habitats. Through selection process, genetic algorithms and exchanges of services between the habitats, individuals of the SME network can benefit by combining their services with those of others, building new products and services so that they can expand their own business easily. In this report, a simplified model of the SME business network is presented as an information ecosystem, where numerous SMEs communicate with others to produce service chains and some services of high fitness values are exchanged between SMEs.

Interactions between different SMEs are supported and executed by the communication infrastructure called FADA, which has been introduced in detail in D19.2. Many researchers have presented a number of approaches of information and communication infrastructures to support business processes between different business enterprizes. CORBA [5], Java RMI[6] and DCOM[7] provide basic information infrastructures, which allow distributed, heterogeneous application systems and data to be shared among consumers, suppliers and business partners. Some approaches have been proposed to improve these basic infrastructures. In [8], a message infrastructure called Business Object Documents (BOD) is built on top of these basic infrastructures. In this way, it separates message transferring from the infrastructure to avoid impacts of changes of the infrastructure on the application systems. Another enhanced infrastructure which consists of several replicable e-business servers for collaborative e-business applications, is presented in [9]. In [10], several adaptive methods and rules for routing messages are applied to construct a self-adaptive infrastructure for business process.

All the infrastructures and methods discussed above emphasize primarily how to support and improve the integration of cooperative business processes among multiple enterprizes. However, in the field of business process reengineering (BPR), most of the researchers suggest that information infrastructure plays an important role in the reengineering process

and they indicate that it should not be seen just as a tool to implement business processes but as an enabler of organizational change [11, 12]. Despite business process and information infrastructure interact in practice, efforts on BPR and information infrastructure are rarely conducted together. But obviously, the study of business processes can not be isolated from information infrastructure. It is important to explore and understand the dynamic relationships between the two and explore the effects of changes of one of these on the other. Some practitioners have investigated this thinking. In [13], the authors present a simulation-based methodology called BPR-II to assess impacts of changes to business process and the infrastructure mechanisms that support the processes. A UK funded research project ASSESS-IT [14] proposes a set of steps to develop simulation models to exchange measurements of performance in order to reflect impacts of changes to IT infrastructure on business process. Based on the knowledge gained from the ASSESS-IT project, an alternative simulation framework, namely BPISS is proposed [15]. It provides guidelines on how to develop simulation models that depict business process and information systems performance. These simulation-based methods and models show the close correlation between business processes and information infrastructures. However, further analysis on this relationship in more detail is needed.

In this report, a two coupled networks model which is more complex than the model presented in D19.2 is proposed. The discrete-event simulation tool is used to model the dynamics and behaviours of a simplified version of the real-world DBE system. In this way, a comprehensive analysis of interactions between the dynamics related to the business process in the SME business network and the FADA network communication infrastructure which supports the business processes can be achieved. The architecture of the two coupled networks model and the dynamic interactions between the SME network and the FADA network are described in chapter 2. More details on the different aspects of the model are represented in the following chapters. In chapter 3, the exchange of services between SMEs and evolution of populations of services in SMEs are illustrated. Section 3.4 describes how solutions to requests, which are from DBE consumers, are generated and optimized through simple genetic algorithm. The dynamics shown in these sections reflect

the SME business network as an information ecosystem is an analogy with the natural ecosystem. Chapter 4 shows the simulation parameters and results, and discussions are also presented. A summary is provided in chapter 5.

Chapter 2

A Geography Based Two Coupled Networks Model

In the two coupled networks model presented in the D19.2, it was considered that all FADA nodes were available to all SMEs. Thus every SME was able to register its services in whichever FADA node. This was for the purpose of making a simple assumption so that the effects of different network topology mechanisms and service registry decisions on the performance of the FADA network could be studied. However, in the real world, capacities and bandwidths of links between nodes located in different geographical regions are various. Moreover, some nodes may be unavailable to some users. These considerations do not affect the findings that were obtained in the D19.2. Nevertheless, this needs to be taken into consideration here as the focus of the work in this report is on the effects of communications and interactions between SMEs in the upper layer (UL) and FADA nodes in the lower layer (LL). How the two networks interact with each other is important. In this model, communications generated by evolutions of populations of services in the evolutionary environment and migrations of copies of services between habitats of SMEs, are considered. Moreover, the registration of services is also driven by the communications to implement the interactions between the UL and LL. Therefore, how SMEs communicate with each other and with FADA nodes is an important factor. A modified model is illustrated in the Fig. 2.1.

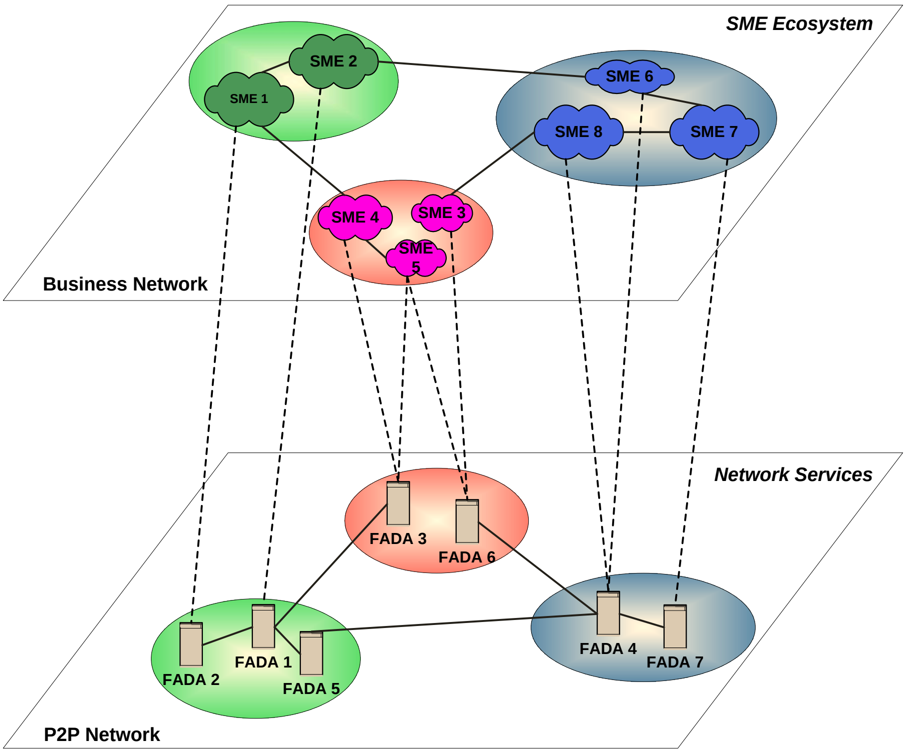


Figure 2.1: Architecture of the two coupled network of the DBE

Business interactions happen more frequently between SMEs within the same geographical region, therefore SMEs are modelled to be virtually grouped according to their geographical locations. As FADA nodes in the communication infrastructure support the business processes of SMEs, they are also virtually grouped depending on the locations of the SMEs that they are serving. SMEs in a region can only communicate with FADA nodes which are in the same region as them.

The details of the model are discussed as following. In the *initialization stage*, the FADA network of LL is *set up* by using ER random graph model (see the details in Appendix A) with a fixed connectivity probability p_{FADA}. As the communications between SMEs are dependent on the topology of the network of SMEs, the network of SMEs are modelled by using the three most popular topologies in the field of complex network, which are ER random graph model, BA scale-free model and small-world network model (see the

details in Appendix A). SMEs initially *register* their services in the FADA nodes which are available to them. The number of FADA nodes in which a service can be registered is denoted by n_{reg}.

In the *evolving stage*, at each time step, each FADA node checks leases of proxies that are registered in it. The lease of a proxy is the amount of time allocated to a proxy. It indicates how long a service proxy can be kept in the FADA. If the lease of a proxy is to expire, the node has to make a decision whether to renew it or not, depending on the remaining space of the node s_{rem}. If $s_{rem} > c_s s_{ini}$, it renews the proxy, otherwise, it deletes the proxy so that some memory can be set free. c_s and s_{ini} here are system parameters set as constants. s_{ini} is the initial memory assigned to each node. The renewed lease of a proxy is allocated as a randomly selected value but bounded by $\frac{s_{rem}}{s_{ini}} \cdot lease_{max}$, where $lease_{max}$ is the parameter of the maximum lease time of a proxy. The procedure of renewing service proxies is illustrated in Fig. 2.2.

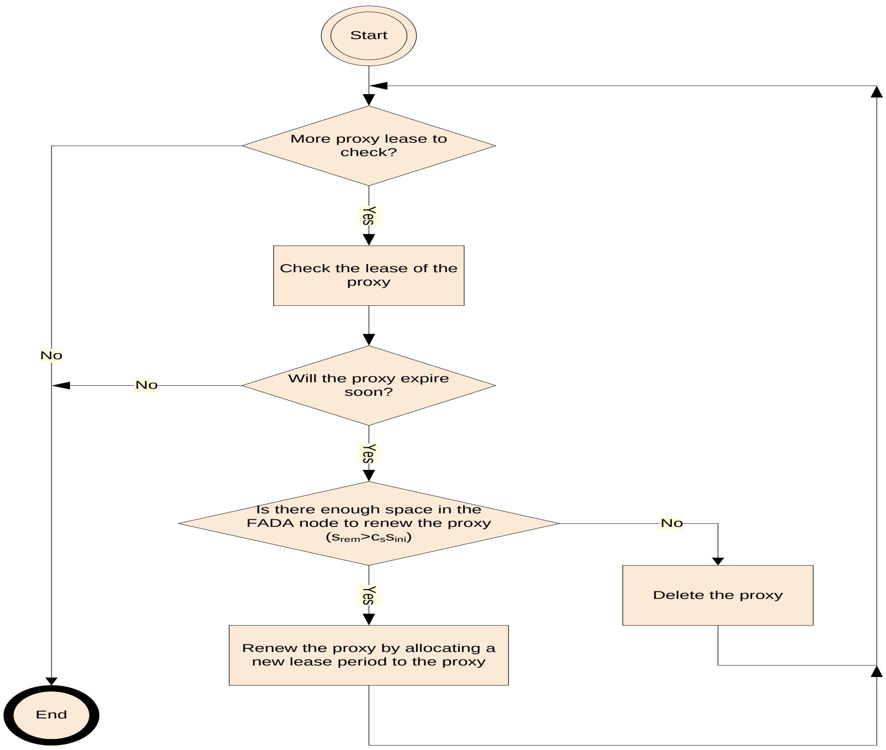


Figure 2.2: The diagram of the procedure of renewing service proxies.

As time evolves, SMEs join and leave the system frequently. The SMEs are modelled *joining* the system at rate r_{join} in the simulation in this research. When a new SME joins in the system, a FADA node is generated to serve it accordingly. In order to maintain the topology of the FADA network as a ER random graph model with the fixed p_{FADA}, n_{edg} links are made connecting the new joining FADA node and some randomly selected FADA nodes, where $n_{edg} = p_{FADA} \cdot\ N$ (N is the size of the FADA network). The new SME registers its services proxies in randomly selected FADA nodes which are available to them. This process is illustrated in Fig. 2.3.

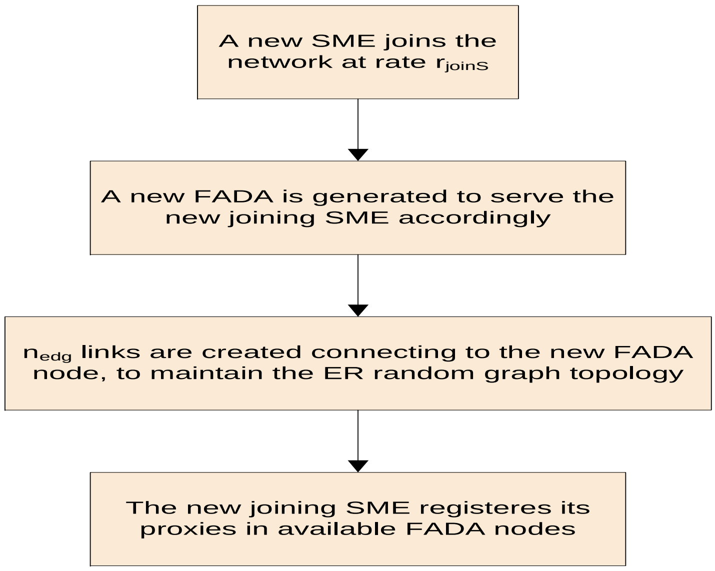


Figure 2.3: The scenario of a new SME joining the system.

In the simulation, *requests* are generated at rate r_{req} per time step (details of the format of requests are described in section 3.4). When a request is sent to, for example SME i, an evolving population in the service habitat of the SME i is activated. Service habitats of SMEs are the places where copies of pointers to services provided by their own SMEs and also other SMEs are stored. According to the requirements of the request, fitness values of different services are computed by using a fitness function (defined in Eq. 3.12). Through a set of simple genetic algorithm, an *optimal solution* is created (details are

shown in section 3.4). For the purpose of exchanging and sharing good services between SMEs, copies of pointers to services in the optimal solution generated from the service habitat of SME i are *migrated* to the service habitats of the SMEs which are neighbors of the SME i. For those habitats of SMEs that the copies of pointers have been migrated to, the migrated copies result in an evolution of populations of services in the habitats. The strategy of the evolution is that popular services will reproduce and less-used services will die out. Accordingly, the registrations of proxies of these services in the FADA network will be changed.

The optimal solution, which contains the IDs of the selected services, is then sent to a randomly selected FADA node available to the SME i. The *lookup* procedure is then performed in the FADA network to look for appropriate proxies for the selected services. The lookup procedure is as the following: the selected FADA node is called the initiator of the lookup procedure. The initiator node makes a lookup in its own data and uses the flooding algorithm to propagate the request in the network. The request is extended to the initiator's neighbours and its neighbours' neighbours and so on. A counter, $numHops$, controls how many hops the request can be sent from the initiator node.

A FADA node, who finds a matched proxy returns the proxy to the initiator FADA node via a direct call to the initiator node, instead of a reverse path so as to allow a fast response. Consequently, a new connection is generated between the FADA node that contains the required proxy and the initiator FADA node. The service consumer SME i downloads the proxies from FADA nodes. By using the methods on the proxies, the service consumer SME i can communicate can invoke actual services from different SMEs as providers by using the Java RMI.

Chapter 3

Evolution-generated Communications in the SME Network

3.1 Evolution of Populations of Services in SMEs

In the SMEs business network of the two coupled networks model, SMEs exchange services with each other in order to achieve better profit. Inspired by evolutions of species in natural ecosystems, where fit species reproduce and weak species die out, populations of different services in SMEs can also evolve according to the fitness values of different services in this way. In order to define the rules for the evolution of populations of services, replicator dynamics in evolutionary game theory has been applied to the model.

3.2 Replicator Dynamics

The replicator dynamics describes the evolution of frequencies of strategies in a population as in [16]. Let $p_i(t)$ represent the number of individuals programmed to strategy $i \in k$, and let $p(t) = \sum_{i \in k} p_i(t) > 0$ represent the total population. A mixed strategy x is defined

as a population state

$$x(t) = x_1(t), \ldots, x_k(t)), \tag{3.1}$$

where each component $x_i(t)$ represents the population share of individuals programmed to pure strategy i at time t. Therefore,

$$x_i(t) = \frac{p_i(t)}{p(t)} \tag{3.2}$$

is the proportion or frequency of the individuals at pure strategy i in the total population. In game theory, what payoff each player will get as the outcome of the game is shown by the so-called payoff function. Let u be the payoff function indicating the effect on an individual's fitness, the populations of individuals change over time according to the payoff function. Let $\beta \geq 0$ be the background fitness of individuals in the population and $\delta \geq 0$ be the death rate which are the same for all individuals, the population dynamics is defined as the following:

$$\dot{p_i} = [\beta + u(e^i, x) - \delta]\ p_i. \tag{3.3}$$

Accordingly, the dynamic for the specific population frequency x_i becomes

$$\dot{x_i} = [u(e^i, x) - u(x, x)]\ x_i. \tag{3.4}$$

Here $u(e^i, x)$ is the payoff of the individual at strategy i, and $u(x, x)$ is the average payoff of the whole population:

$$u(x, x) = \sum_{i=1}^{k} x_i u(e^i, x). \tag{3.5}$$

3.3 Applied Replicator Dynamics in SMEs

In the SMEs business network layer, each SME owns a service habitat, in which services of different business types exist. For instance, for a SME which is a travel agency, its habitat may maintain services such as airline, car rental, hotel, restaurant, cleaning services and so on. The providers of these services can be the SME itself or by other SMEs. When a request is sent into a SME, some services in the SME are selected to construct

a service chain. The selection of services is based on their fitness values, which are depending on the differences between services and the request and also their usage histories. The usage history of a services means the number of times that has been used in the past. The selection process is discussed in detail in section 3.4.3. Therefore, if a SME can be guided by some rules so that it knows how to keep and reproduce services of high fitness values and get rid of services of low fitness values, solutions that the SME can offer may be better, benefiting not only the individual SME but also the whole DBE system.

Let p_i^j be the number of services of type i in population j (SME j) and x_i^j be the proportion of services of type i in population j. Therefore,

$$x_i^j(t) = \frac{p_i^j(t)}{\sum_{i=1}^{n} p_i^j(t)}, \tag{3.6}$$

where n is the number of types of different services in all populations and

$$\overrightarrow{x}^j = (x_1^j, x_2^j, \ldots, x_n^j). \tag{3.7}$$

At each time step, each service in the population of services in a SME is decided to be reproduced, deleted or maintained, depending on an associated value v_e. For instance, ${v_e}_i^j$ is associated with service type i in SME j. When ${v_e}_i^j \geq 0$, the absolute value of it represents the probability of reproducing service i in j, denoted as ${prob_r}_i^j$. On the other hand, when ${v_e}_i^j < 0$, the absolute value of it means the probability of deleting the service i in j, denoted as ${prob_d}_i^j$. It is illustrated as the following:

$$\begin{cases} {prob_r}_i^j = |{v_e}_i^j|, & {v_e}_i^j \geq 0, \\ {prob_d}_i^j = |{v_e}_i^j|, & {v_e}_i^j < 0. \end{cases} \tag{3.8}$$

Accordingly, the population dynamics in the model becomes the following:

$$\dot{p}_i^j = {v_e}_i^j \; [p_i^j + {p_i^j}'], \tag{3.9}$$

where $p_i^{j'}$ is the population of service i migrated from SME j' to SME j.

Comparing Eq.3.9 in the model presented in this research with Eq.3.3 in replicator dynamics theory, it can be noted that the ${v_e}_i^j$ can be regarded as the payoff function, and it plays an important role in the population dynamics. As is in other large complex systems, interactive behaviours between individuals in the model in this research are difficult to observe and quantify. Therefore instead of defining a clear form for the payoff function, a function to compute the value of the ${v_e}_i^j$ for each service is defined as its evolution guidance. In view of both system aspects and the functionality of the SME network of providing services, ${v_e}_i^j$ is defined as a function of N, Δu and r. Here, N is the total number of services in a SME habitat, $\Delta u = u_i - \overline{u}$ is the difference between the usage history of service i and the average usage of all services in a SME habitat including copies of services migrated from other SME habitats, and $r = \frac{n_i}{n_{max}}$, where n_i is the number of service i and n_{max} is a fixed number to limit the amount of a certain service type in a habitat. The considerations are as follows: as the memory of each SME habitat is limited, evolutions of services in SME habitats have to take N into account so that the size of the population will not exceed the memory limit. Moreover, the amount of services that have been used for more times should be larger than other services in the population, which is reflected by the variable Δu. However, in order to maintain the diversity of types of services and also avoid the case that a few strong services occupy the whole population, the number of services of a certain type in a population should be limited, as represented by r. In a word, the function for ${v_e}_i^j$ is defined through a set of comprehensive considerations. It enables services in SMEs habitats to evolve "rationally".

In the model, the range of $v_e(N, \Delta u, r)$ is $[-1, 1]$. It is defined as:

$$v_e(N, \Delta u, r) = \frac{1}{2.2}[\frac{\exp(H) - 1}{\exp(H) + 1} - 1.2(\frac{\exp(\alpha(N - \beta) - 1)}{\exp(\alpha(N - \beta) - 1)})], \tag{3.10}$$

where $\alpha = 0.02$ and $\beta = 300$. H is a function dependent on Δu and r:

$$H(\Delta u, r) = -12.7r + 0.04\Delta u + 4.1. \tag{3.11}$$

The Probability Density Function (PDF) is shown in Fig.3.1.

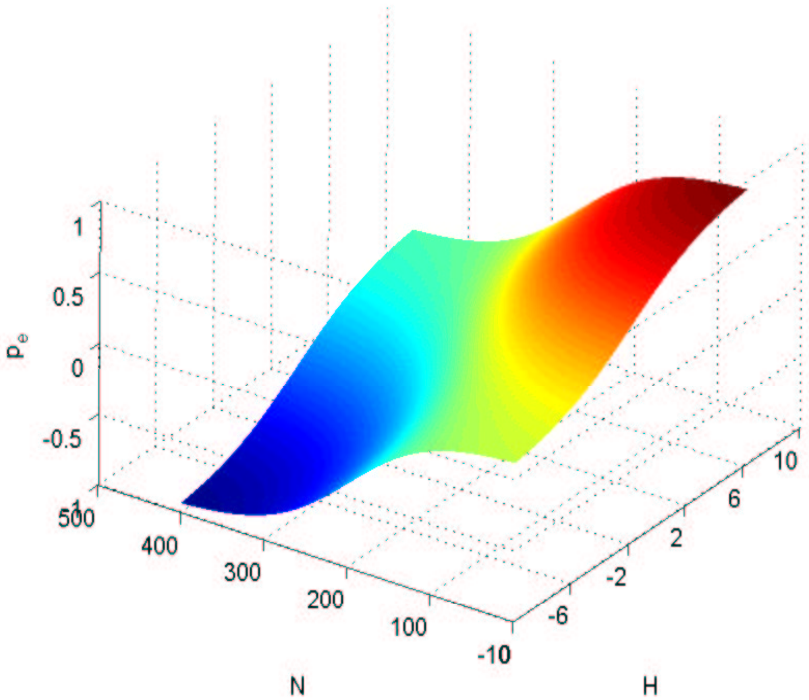


Figure 3.1: The PDF of the probability of generating and deleting a service.

The parameters in Eq.3.10 and Eq.3.11 are set by considering the corresponding values of v_e under different cases of system conditions. These cases include when the number of services in a SME, N, approaches N_{min} and N_{max}; the number of a service of a specific type in a population, r, approaches r_{min} and r_{max}, and the Δu is extremely large and small. These are summarized in table 3.1.

Table 3.1: Corresponding probability v_e from Eq.3.10 in conditions of the model

N	r	Δu	$\|\Delta u\|$	p_e
$\to N_{min}$	$\to r_{min}$	> 0	large	$\to 1$
$\to N_{min}$	$\to r_{min}$	> 0	small	$\to 0.7$
$\to N_{min}$	$\to r_{max}$	> 0	large	$\to 0.3$
$\to N_{min}$	$\to r_{max}$	> 0	small	$\to 0.1$
$\to N_{max}$	$\to r_{min}$	> 0	large	$\to -0.1$
$\to N_{max}$	$\to r_{min}$	> 0	small	$\to -0.3$
$\to N_{max}$	$\to r_{max}$	> 0	large	$\to -0.7$
$\to N_{max}$	$\to r_{max}$	> 0	small	$\to -1$
$\to N_{min}$	$\to r_{min}$	< 0	large	$\to 0.2$
$\to N_{min}$	$\to r_{min}$	< 0	small	$\to 0.7$
$\to N_{min}$	$\to r_{max}$	< 0	large	$\to 0.1$
$\to N_{min}$	$\to r_{max}$	< 0	small	$\to 0.1$
$\to N_{max}$	$\to r_{min}$	< 0	large	$\to -0.9$
$\to N_{max}$	$\to r_{min}$	< 0	small	$\to -0.3$
$\to N_{max}$	$\to r_{max}$	< 0	large	$\to -1$
$\to N_{max}$	$\to r_{max}$	< 0	small	$\to -0.9$

3.4 Optimizations of Solutions to Requests

3.4.1 Description of services

Each service in a SME habitat is primarily described by the following entries:

- busiType: describes the business type of a service. It is presented by a randomly chosen number in the range of $[1, 10]$. Each number represents a specific business type, e.g "1" represents the airline name, "2" represents the flight destination, "3" represents the place of hotel accommodation and so on.

- SM: is the description of a service in Business Modelling Language (BML). In the model, a binary string of length of 4 is used to represent a service of a certain business type. Therefore, there are at most $2^4 - 1 = 31$ different services of a specific business type.

- smID: the unique ID of a service.

- homeHabitatID: the ID of the SME habitat where the service originally exists.
- history: the number of times that a service has been used in the past.

3.4.2 Description of requests

It is modelled that at each time step, one request is created and sent to the system looking for services provided by SMEs. In a request, required services are represented in the form of a service chain. Each request is composed of the following entries:

- requestID: the ID of a request.
- smeID: the ID of a randomly chosen SME which a request is sent to.
- chainLength: the length of a request, which counts the number of services required.
- busiTypes: the business types of services required in a request.
- requiredSM: the service manifests (service descriptions) of services required in a request.
- requestTime: the record of the time step when a request is sent to the system.

An example of a request is shown in Fig. 3.2.

requestID: 1
smeID: 8
chainLength: 6
busiTypes: 3 5 6 8 9 10
requiredSM: 6 5 1 3 11 14
requestTime: 1

Figure 3.2: An example of a request.

3.4.3 Optimizing solutions to requests

When a request is sent to a SME, the SME creates an evolving population, in which corresponding services are aggregated to construct a number of service chains to the request. A

simple optimization process by using crossover genetic algorithm is applied to the service chains generated in the evolving population in order to provide the optimal solution to the request. The procedure of answering a request in the DBE is illustrated in Fig 3.3.

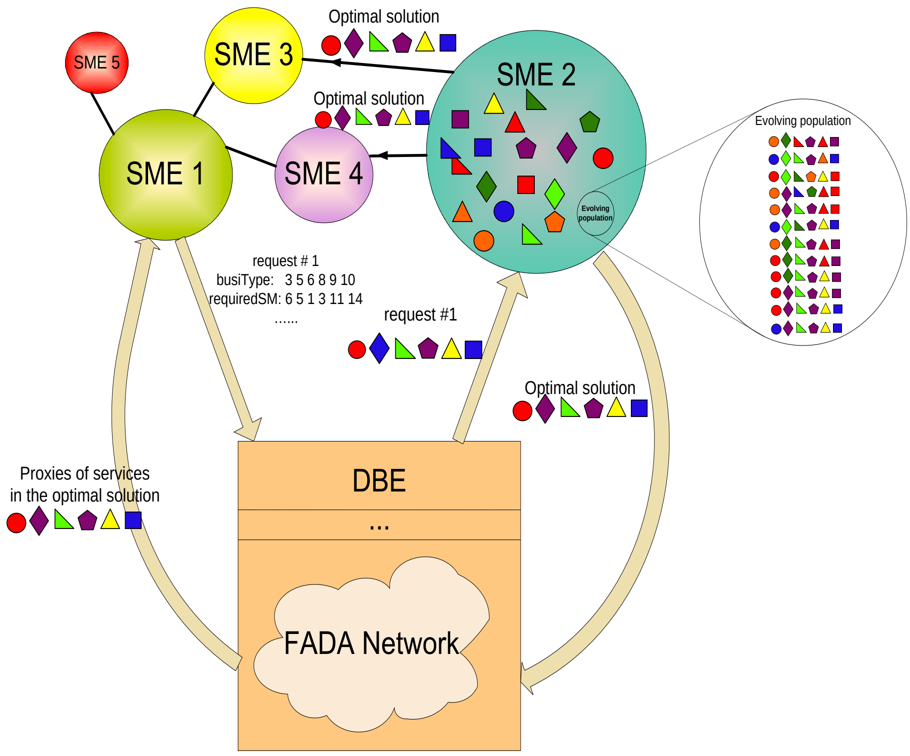


Figure 3.3: The procedure of answering a request in the DBE

For clearer explanation, the example of a request in the previous section is used here to demonstrate how solutions to the request are generated. For instance, SME 1 sends a request 1 to the DBE system. The request is in the form of a service chain, and it contains the information of the types and the service manifest descriptions of the required services, which are represented by "busiTypes: 3 5 6 8 9 10 " and "requiredSM: 6 5 1 3 11 14" respectively in Fig. 3.3. The DBE system sends the request to a suitable SME, e.g. SME 2. In the figure, each shape represents the business type of a specific service, and each colour represents the service manifest descriptions of a specific service. SME 2 receives the request 1 from the DBE system, and it generates an evolving population corresponding to the request 1. In the evolving population, it first collects all the services whose business

types are the same as the required services in the request, but these services are of different service manifest description. Then the SME 2 generates a number of possible solutions by aggregating these services into service chains. Due to the memory limitation of SMEs, the number of solution chains generated in the evolving population in SME 2 is bounded by $S = 1000$. SME 2 computes the fitness value of each solution chain. The solution with the highest fitness value is selected as the optimal solution to the request 1, and is sent to the DBE. The FADA network then provides the proxies of the services in the optimal solution to SME 1, where the request is originally sent from.

The details of how the optimal solution is generated from the evolving population of SME 2 are explained as follows. The format of service chains generated in the evolving population as solutions to the request is presented in Fig. 3.4.

```
solution 1:
              ID:  1
       busiTypes:  3     5     6     8     9     10
              SM:  13    21    13    15    7     2
            smID:  2424  1930  701   983   2625  6548
         history:  1     0     1     2     3     5
  homeHabitatID:  8     8     3     1     8     14
        fitValue:  27.2439

solution 2:
              ID:  2
       busiTypes:  3     5     6     8     9     10
              SM:  13    7     13    2     11    15
            smID:  6634  8902  4930  3566  2607  307
         history:  0     2     1     2     4     5
  homeHabitatID:  8     4     20    8     8     3
        fitValue:  27.6204

solution 3:
     ......
```

Figure 3.4: Examples of solutions to a request.

The "fitValue" in the examples of solutions in Fig. 3.4 means the fitness values of the

solutions. Fitness value of each solution is computed according to Eq. 3.12 below:

$$f(i) = c\ [\exp(-\alpha \cdot |d(i)|) + \frac{\exp(\beta \cdot h(i))}{1 + \exp(\beta \cdot h(i))}], \tag{3.12}$$

where

$$d(i) = \sum |d(i,j)| = \sum |s(i,j) - r(i,j)|, \tag{3.13}$$

and

$$h(i) = \sum h(i,j). \tag{3.14}$$

Here $d(i,j)$ is the measurement of the difference between a service provided in a solution and a service required in a request and $h(i,j)$ is the usage history of a service provided in a solution. α, β and c are system parameters. c is set as $c = 50$ to make the maximum fitness value be 100. Eq. 3.12 implies that a relatively high fitness value is assigned to a solution which contains services that not only are similar to services required in a request but also have good usage records.

In order to produce the optimal solution from the limited number of solutions in an evolving population of a SME, it is necessary to introduce some genetic algorithms to the system. In genetic algorithms, crossover is a genetic operator which combines two parents to produce a new child, which may be better than the parents by possibly taking the best characters from each of the parents [17]. Generally, there are several types of crossover, including one point, two Point, uniform, arithmetic and Heuristic. The simplest type, the one point crossover is used in the model in this research. It is operated by randomly selecting a crossover point within a chromosome and then interchanges the two parents chromosomes at this point to produce two new children. It is illustrated in Fig.3.5, where the symbol "|" represents the crossover point.

Chromosomes used in crossover are usually represented by binary strings, as shown in Fig. 3.5. However, it is not appropriate to represent services in the binary string form in the model here. This is because a SME can only offer certain services, and after applying

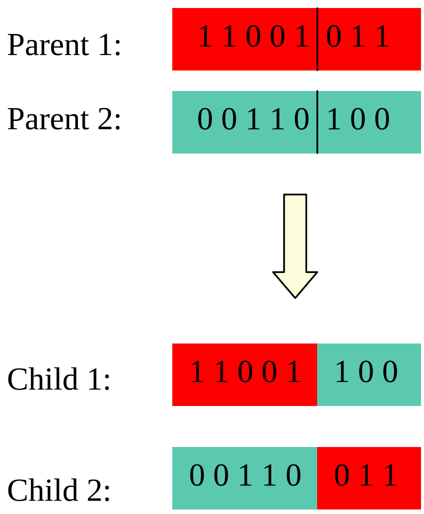


Figure 3.5: An example of one point crossover algorithm.

crossover operations on services represented in the binary string form may generate some services which the SME can not provide. Therefore, chromosomes of two parents in the model presented here are service manifests of services in two randomly selected solutions among all solutions in the evolving population, as shown in Fig. 3.6.

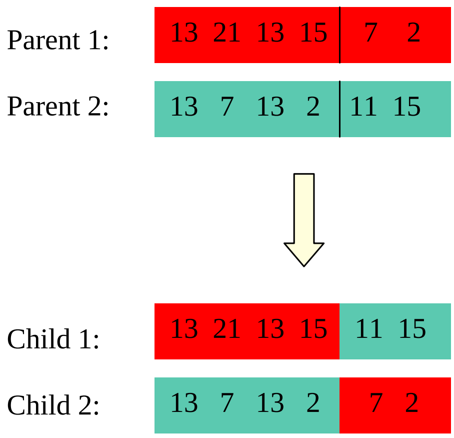


Figure 3.6: An example of one point crossover algorithm applied to the model.

Fitness values of two new solutions represented by two new children generated from crossover operation are computed by using Eq. 3.12 and then are compared to fitness values of other existing solutions. If the fitness value of a new solution is higher than the

lowest fitness value of existing solutions, the new solution will replace the existing solution of the lowest fitness value. The procedure of the crossover operation and replacement needs to be carried out for a number of generations N_g. In order to set an appropriate value of N_g, several experiments were run. It was found that $N_g = 1000$ is a suitable setting, because the optimal solution with the highest fitness value can usually appear after 1000 generations, and it also avoids redundant resource consumptions. Therefore, after N_g generations, the solution of the highest fitness value is selected as the optimal solution to the request.

Chapter 4

Simulations

4.1 Simulation Implementation

The description of the model has been described in detail in chapter 2. Based on the model description, discrete-event simulations have been carried out. Primary simulation parameters are summarized in table. 4.1. Some parameters have different values in order to investigate the effects of topologies of the SME business network, interaction between the SME network and FADA network on the system performance. The simulation results are presented and discussed in the following sections.

Table 4.1: Parameters and the values of them in the simulations.

Simulation parameters	values
connectivity probability of SME network by using ER model, p_{SME}	0.2, 0.4, 0.6, 0.8
connectivity probability of FADA network by using ER model, p_{FADA}	0.2
initial space assigned to each FADA node, s_{ini}	64,000
maximum size of one proxy, s_{proxy}	200
maximum lease time for a proxy, $lease_{max}$	100
maximum number of hops in lookup procedure, $hops_{max}$	1,2,5
the number of FADA nodes for registrations of one proxy, n_{reg}	1,3,5
rate of sending requests, r_{req}	1,2,3,4,5,6

4.2 Effects of Graph Connectivity of the SME Business Network

Communication flows generated due to the migration of copies of pointers to good services and the changes of registrations of services proxies are depending on the topology of the SME business network. Therefore, it is interesting to study how the topology of the SME business network (UL) affects the the FADA network (LL) and also the system performance.

In the simulation, ER random graph model is applied to both the SME network and the FADA network. Due to the objectivity of the study of this section, the connectivity probability of the FADA network is fixed as $p_{FADA} = 0.2$, while the connectivity probability of the SME network p_{SME} is of various values as shown in table 4.1. As described in section 3.3, the idea of replicator dynamics has been applied to the evolution of populations of services in the model. It was firstly shown that although copies of pointers to services are migrated from some SMEs' habitats to others over time and the evolution of populations of services happen every now and then, thanks to the comprehensive considerations for the function of v_e in Eq. 3.10, the total number of services in a habitat can be bounded by a limit after a short period of evolution. The result is shown in Fig. 4.1.

Furthermore, as the migration and the evolution of populations of services happen at all time, services that are kept in any SME habitat are changing accordingly. How does the value of p_{SME} affect the usage records of services maintained in SMEs' habitats? The result is shown in Fig. 4.2.

It can be seen that as time elapses, the average usage record of services in SMEs' habitats increases. It implies that SMEs are able to keep services which have good usage records and delete less used services. It is also shown that the higher the p_{SME}, the larger the average usage history of services in habitats is. This is because that if the p_{SME} is larger, popular services can be migrated to more SMEs' habitats. Furthermore, it is found the

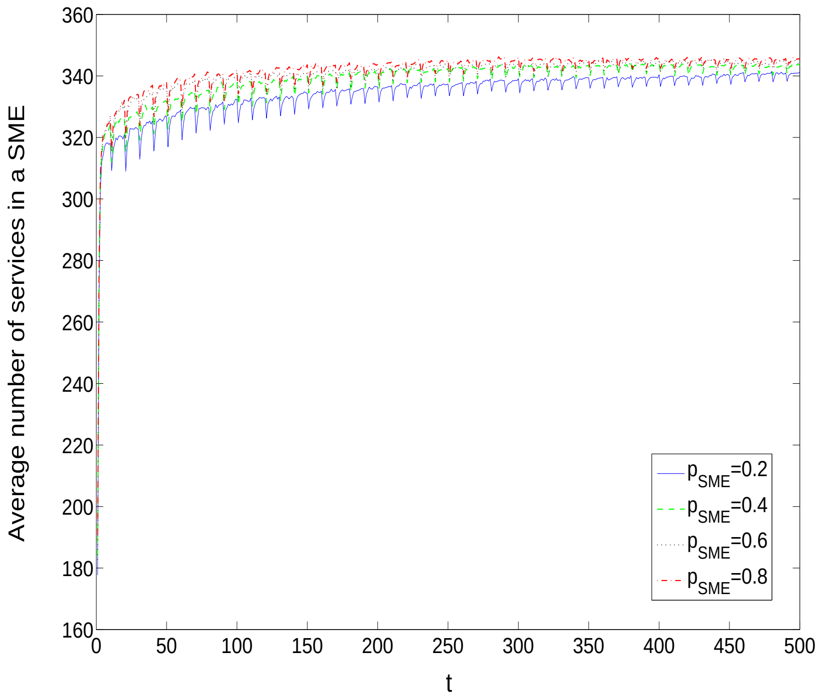


Figure 4.1: The number of services in SMEs' habitats.

scaling law between the p_{SME} and the average usage of services in habitats is linear.

As the p_{SME} shows the impact on the usage records of services in habitats, does it also affect the fitness value of solutions? According to the fitness function defined in Eq. 3.12, fitness values of solutions are dependent on the differences between services provided by SMEs and services required in requests, and also the usage records of services maintained in the habitats. The potential solution whose fitness value is the highest is selected as the optimal solution. Here the fitness values of optimal solutions offered by the SME business network with different p_{SME} are investigated. The result is presented in Fig. 4.3.

For the purpose of clarity, the results of $p_{SME} = 0.2$ and $p_{SME} = 0.8$ are only presented here. It is shown that the solution provided by the SME network of larger p_{SME} is better than of smaller one, benefitting from the higher average usage records of services in SMEs' habitats in the SME network of larger p_{SME}.

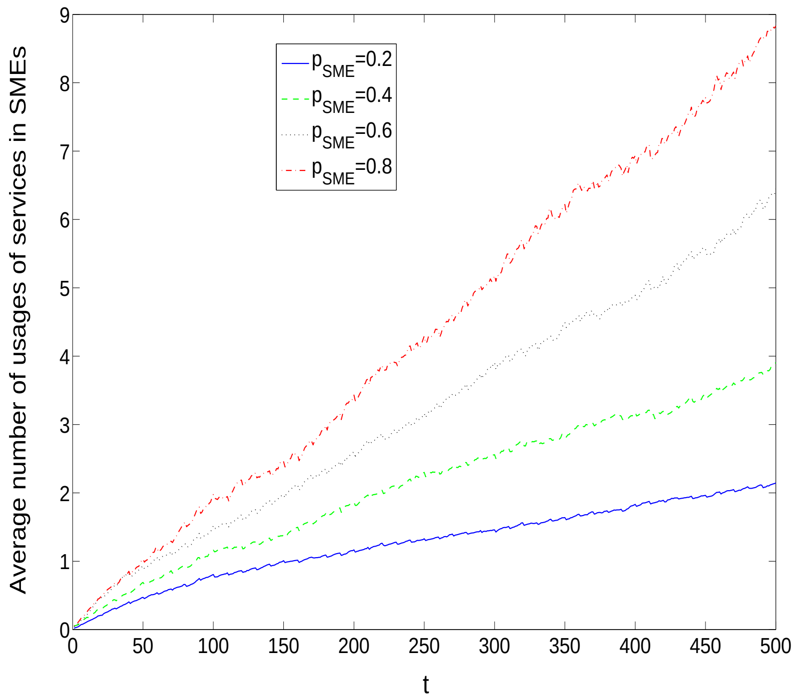


Figure 4.2: The average number of usages of services in SME habitats in a SME business network of different connectivity probabilities.

As described in the model description in chapter 2, the evolution of populations of services in the UL results in changes of registrations of services proxies in the LL. How different values of p_{SME} affect the topology of the FADA network on the average degree, average path length and network diameter was studied. Firstly, the average degree of FADA nodes as time evolves was plotted when p_{SME} are set as various values in Fig. 4.4.

In the simulation, at every 10 time steps, one new FADA node is generated and new connections originated from it are made according to the ER random graph model. Therefore, theoretically, the average degree of FADA network should increase along with the network size grows as presented in the random graph theory [1]: $\langle k \rangle_{RG} = p\ N$. Here p is $p_{FADA} = 0.2$. However, it is shown in Fig. 4.4 that the average degree of FADA nodes increases much faster than k_{RG} when $p = p_{FADA} = 0.2$. This indicates the significant impact of new connections generated during FADA network's lookup procedure. More-

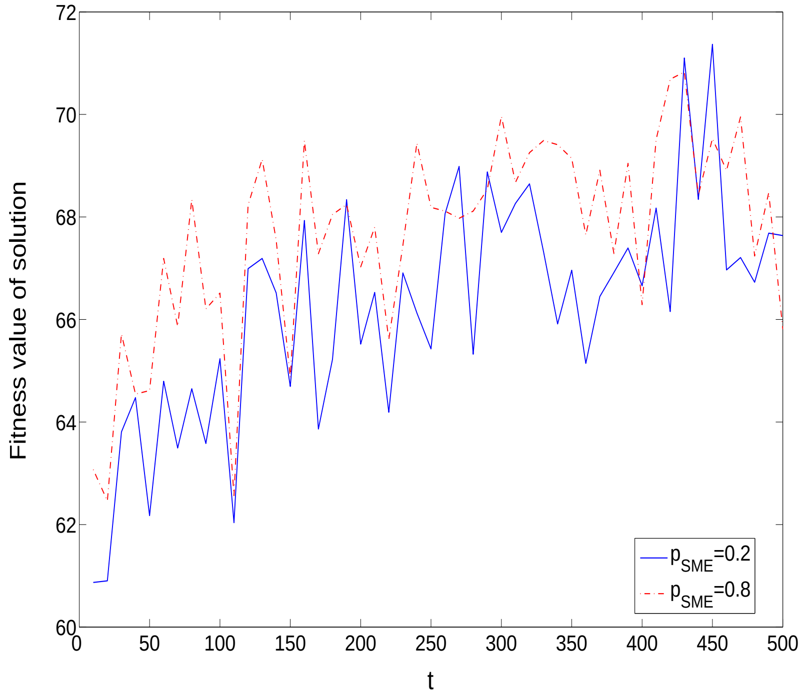


Figure 4.3: The average fitness value of solutions offered by the SME business network of different connectivity probabilities.

over, it can be noticed that as p_{SME} increases, the increase of the average degree of FADA nodes is faster. The explanation of this is as the following: larger p_{SME} results in more communication flows between SMEs and migrations of services in SMEs' habitats from a wider range of SMEs. As described in chapter 2, SMEs can only register their services in several local FADA nodes that are available to them. When an optimal solution is sent to the FADA network to look for actual proxies, if the solution consists of services originally exist in SMEs of a wider range, initiator FADA node which initiates the lookup request may find proxies in registered FADA nodes of a wider range accordingly. Consequently, more new connections may be made to the initiator when nodes return results directly to it. Therefore, the larger the p_{SME}, the higher the average degree of FADA nodes is. As the average degree of FADA network with larger p_{SME} is higher, it can be conjectured that the average path length is shorter in this case. It is confirmed by the simulation result shown in Fig. 4.5.

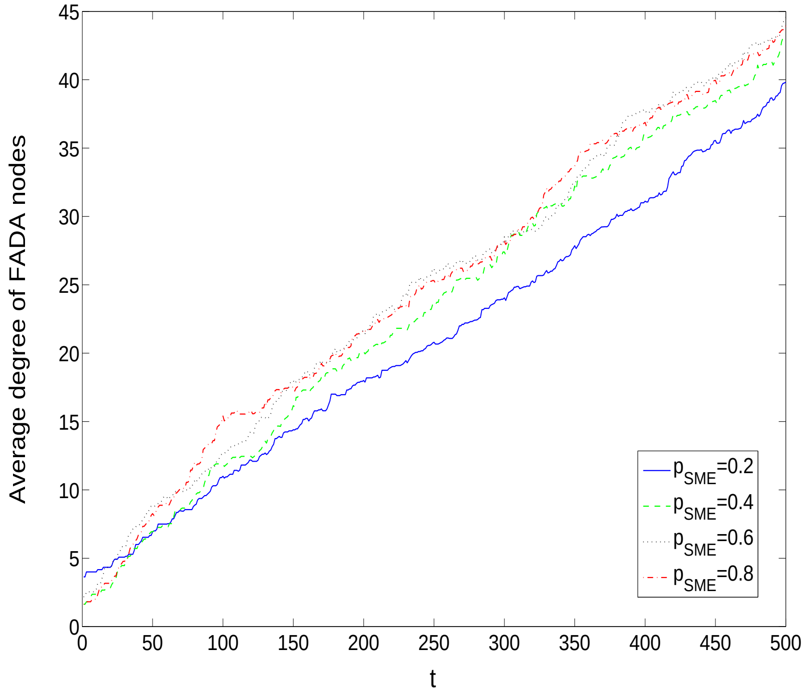


Figure 4.4: The average degrees of FADA nodes in the FADA networks which are coupled with the SME business network of different connectivity probabilities.

Furthermore, in Fig. 4.5, it can also be seen that, the average degree of FADA nodes increases so fast that the average path length drops steeply to some very small values after a short time, which are between 1.2 and 1.6. Another important topological property of a network, the network diameter (described in Appendix A), which is defined as the maximal distance between any pair of nodes has also been studied. It is an important character of a network because it shows the worst case for two nodes to communicate with each other. The network diameter of the FADA network d_{FADA} against time was plotted in Fig. 4.6. For various values of p_{SME}, it is shown that d_{FADA} always equals to 2 approximately, which implies that the FADA network is very efficient for communication in all circumstances.

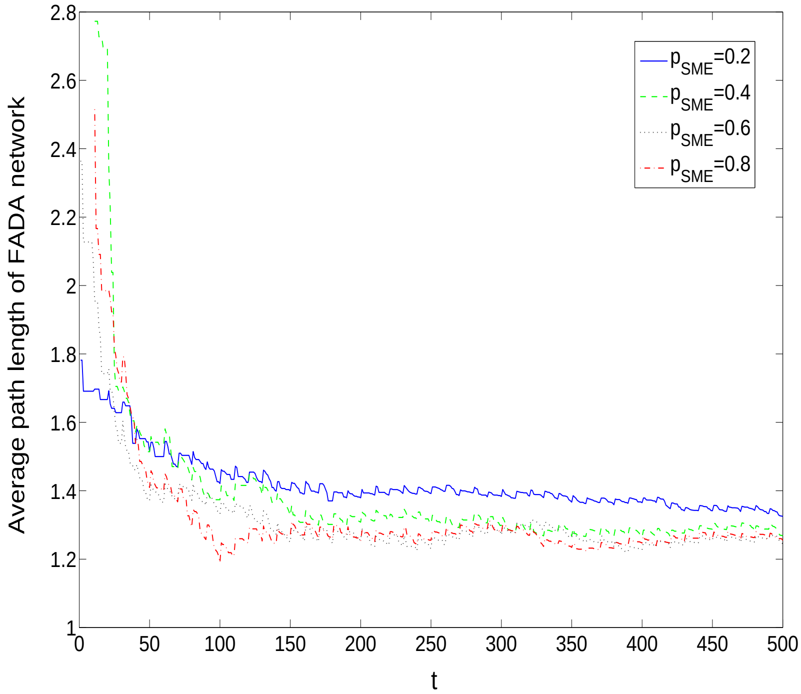


Figure 4.5: The average path lengths of FADA networks which are coupled with the SME business network of different connectivity probabilities.

4.3 Effects of Topologies of the SME Business Network

In the previous section, the effects of using various connectivity probabilities p_{SME} of the ER random graph model for the topology of the SME network on the performance of the DBE have been discussed. Changing the p_{SME} reflects the impact of communication flows between SMEs on the system directly. Besides that, another interesting and important thing investigated was how different topologies of the SME network affect the FADA network and the system performance. Three dominating topologies in the territory of complex networks which are random graph, scale-free network and small-world network have been applied to the SME network. An analysis of the effects of using these three different topologies is provided in this section.

In order to make a fair comparison between the three topologies, the networks are modelled to have the same average degree and the number of edges. The methodology is to make

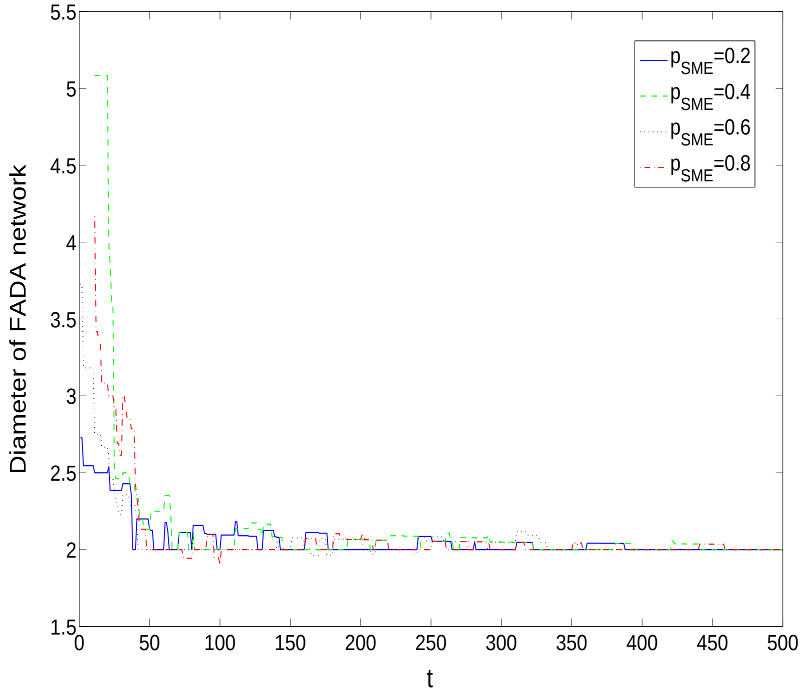


Figure 4.6: The network diameters of FADA networks which are coupled with the SME business network of different connectivity probabilities.

a fixed number of new connections for a new node when it joins in the network for all the three topologies. For (1) random graph network: new connections link to randomly chosen existing nodes, (2) scale-free network: new connections are made according to the "preferential attachment" algorithm, which means the placements of new connections depend on the degrees of existing nodes. We use the evolution principles of the ER random graph model and the BA scale-free network model in this report as in the D19.2. The details of the ER random graph model and the BA scale-free model can be found in Appendix A. Therefore, it is only discussed how to make an appropriate model of the small-world network for the SME network here.

Small-world Network: A brief overview of the small-world network is given in Appendix A. In 1998, Watts and Strogatz firstly proposed a simple model (WS model) to generate small-world networks [2]. It starts with a ring lattice with n vertices and each vertex connects to its k nearest neighbors. Each edge is rewired at random with probability p.

The parameter p is able to control the graph to be between completely regular ($p = 0$) and completely random ($p = 1$).

A much-studied variant of the original WS model was then proposed by Newman and Watts in [18, 19], in which some edges as short-cuts are added between randomly chosen pairs of nodes instead of removing edges. Alternative models have been widely studied as in [20, 21]. In 2000, in order to investigate the small-world effect on delivery time to forward a message, Kleinberg presented a generation of the WS model which is based on a two-dimensional square lattice [22].

All the models described above illustrate the small world character well and are reasonable models to describe how some networks are formed. However, these models are not dynamic because the sizes of the networks of these models do not change. In reality, most of networks grow over time with new nodes joining in continuously and evolves to large size networks. Some models of growing network have been illustrated in [1, 23]. In particular, the models of growing networks with small-world property are of interest here. Ozik, Hunt and Ott proposed a simple mechanism for the evolution of small-world networks in [24], which is, when a new node joins in the network, it connects only to existing nodes that are geographically close to it. It is a very realistic model for the SME business network, as SMEs locate in different geographical regions and the network size of the SME network is supposed to grow gradually. Therefore, this simple mechanism is applied to construct the small-world topology for the SME network. The procedure is as the following: firstly, in the initial stage in the simulation, it is aimed to build a SME network of ten nodes. It starts with three fully connected nodes and a new node is added in the network which is placed in a randomly chosen geographical region. A new node makes two connections to its two nearest neighbors according to their locations. It continues until the network size grows to ten nodes. Then in the evolving stage in the simulation, the network grows according to the following steps: (1) a new node joins in the network at rate r_{join} and it is placed in a randomly chosen location (2) the new node connects to eight nearest neighbors of it. The details of the initial stage and evolving stage of the simulation are described in

chapter 2.

In order to show whether the SME network constructed by using Ozik's simple mechanism exhibits small-world behaviour or not, the average path length ℓ and the clustering coefficient c of the SME network are plotted in Fig. 4.7. The average values of ℓ and c are also summarized in table 4.2.

Table 4.2: Average values of average path lengths and clustering coefficients of SME networks of random graphs, scale-free networks and small-world networks.

	L	C
Random graphs	1.73	0.30
Scale-free networks	1.72	0.34
Small-world networks	2.01	0.59

It can be seen that ℓ of the SME network by using Ozik's model is not much larger than the ER random graph model, and the c of it is much larger than the random graph model (almost two times as the ER random graph model). These satisfy the two requirements of small-world networks, small average path length and large clustering coefficient. Moreover, the result in Fig. 4.7 also shows that ℓ increases and c decreases as time evolves. This is because that although the network size is growing, the number of new generated edges is a constant over time. As a consequence, the average degree of the FADA network is decreasing over time.

Now the effects of different topologies on the system performance are studied, including the fitness values of solutions provided by the SME network and the topological properties of the FADA network.

Firstly, the average usage records of services maintained in SMEs' habitats is plotted in Fig. 4.8 and the fitness values of solutions in Fig. 4.9 respectively. It is found in Fig. 4.8 that the usage records of services in SMEs' habitats when the SME network is a scale-free network are slightly better than when it is a random graph or a small-world network.

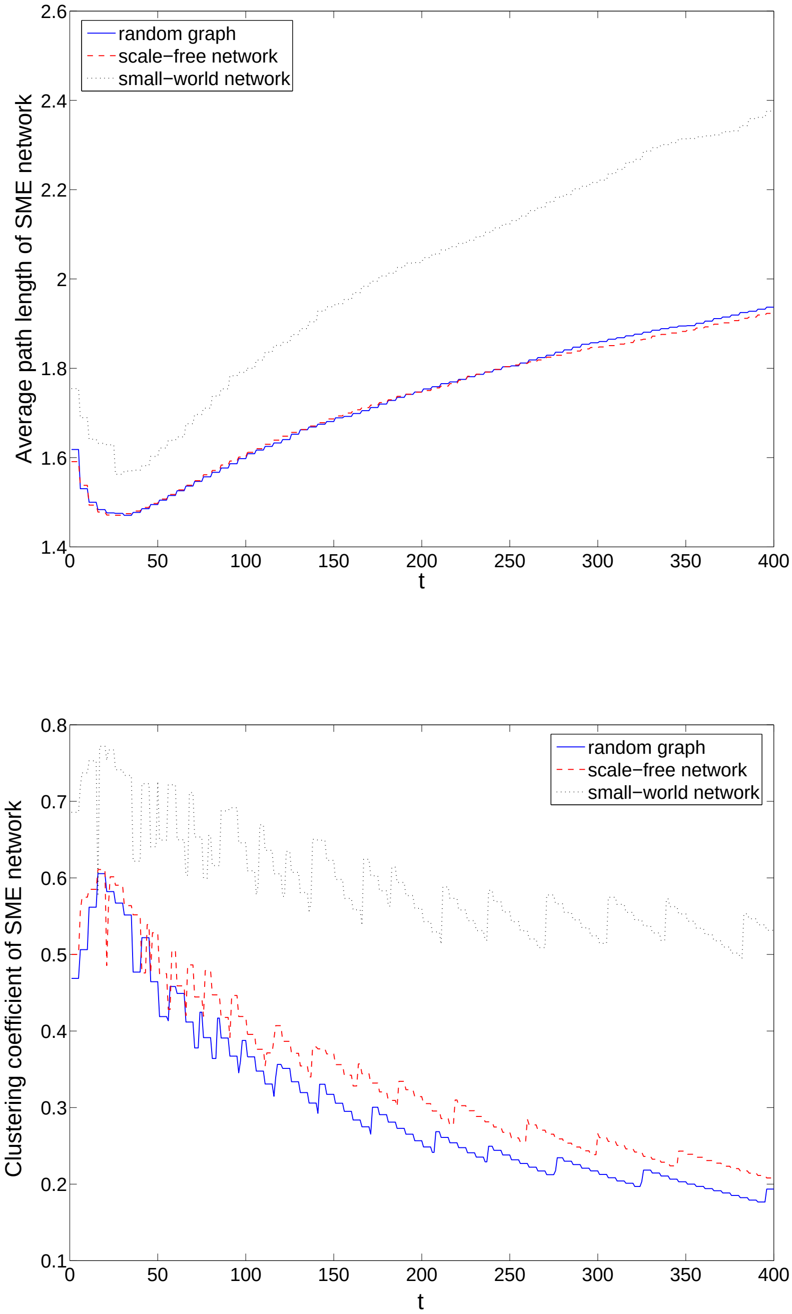


Figure 4.7: The average path lengths and the clustering coefficients of the SME networks by using models of random graphs, scale-free networks and small-world networks.

However, the result in Fig. 4.9 shows that the difference of the usage records of services is not significant enough to improve the fitness values of solutions.

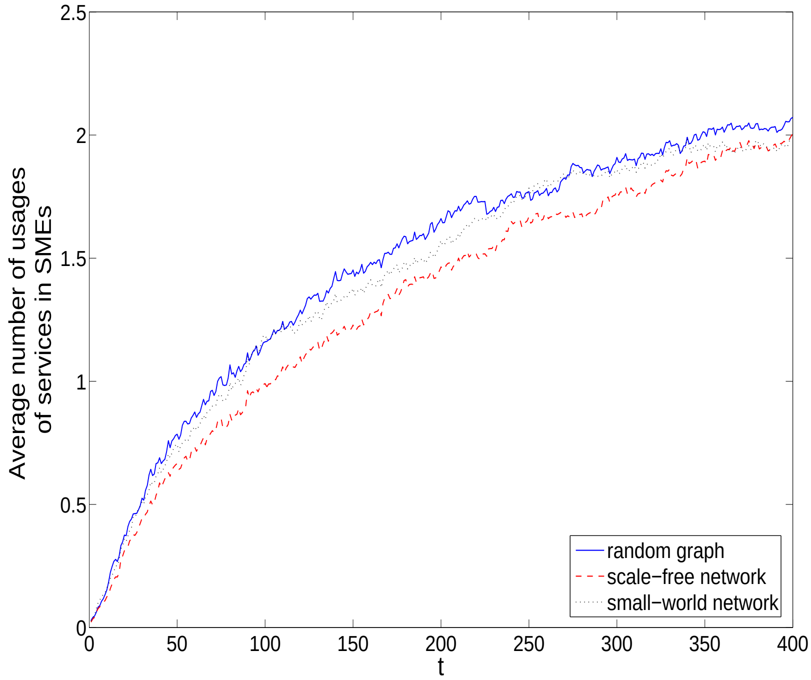


Figure 4.8: The average number of usages of services in SME habitats in a SME business network of different topologies.

Secondly, the impacts of applying different topology models to the SME network on the topological properties of the FADA network were studied, including the average degree, average path length and network diameter. In Fig. 4.10, it is noticeable that when the SME network is constructed by using the small-world network model, the average degree of FADA nodes is smaller than by using the random graph model and the scale-free network model. Consequently, the average path length of the FADA network is larger when the SME network is of small-world network topology, as shown in Fig. 4.11.

Moreover, the result in Fig. 4.12 shows the network diameter of the FADA network drops to 2 after a short while for all the three cases. This implies that the FADA network can provide good communication efficiency no matter which topology model is used for the SME network.

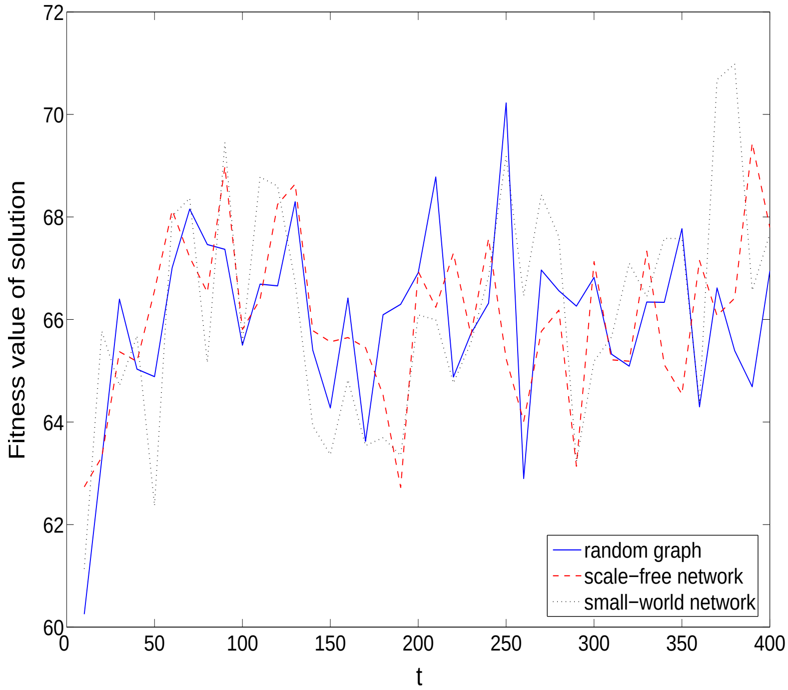


Figure 4.9: The average fitness values of solutions offered by the SME business network of different topologies.

4.4 Scalability of the FADA Network

As a distributed system, the scalability of the FADA network is very important. Whether it can operate efficiently in a wide range of sizes and system configurations is valuable to investigate. In [25], a scalability metric based on productivity is used to examine the scalability of a distributed system as the following: the productivity $F(k)$ is defined as the value delivered per time divided by the cost per time. The value here corresponds to throughput, quality of service or performance. Then the scalability metric at two different scale factors is defined as the ratio of their productivity measures:

$$scalability\ metric\ \psi(k_1, k_2) = F(k_2)/F(k_1). \tag{4.1}$$

Usually, k_1 is fixed at a value and the metric can then be written as $\psi(k_2)$ or $\psi(k)$. In this section, the idea of this scalability metric is applied to evaluate the scalability of the FADA network. k_1 is fixed to 11 which is the number of FADA nodes when the FADA

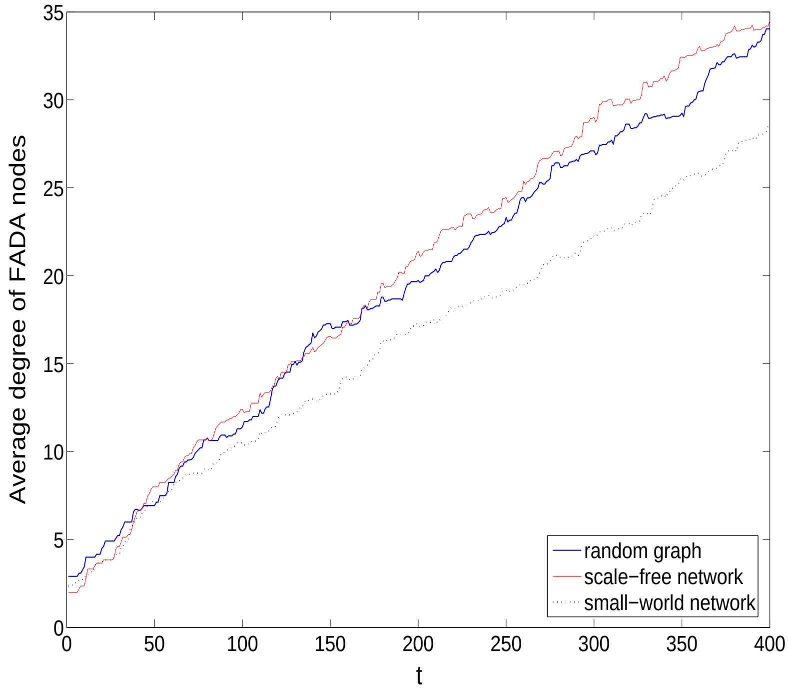


Figure 4.10: The average degrees of FADA nodes in the FADA networks which are coupled with the SME business network of different topologies.

network is at the beginning of the evolving stage in the simulation. A comprehensive discussion on the productivity and the cost is provided. As one of the main functions of the FADA network is to look for and provide appropriate proxies to specific services, three variables are considered as the productivity here, including the number of proxies which are located, the minimum number of hops to locate a proxy and the success rate of finding proxies to services requested. The number of nodes visited and the memory usage of FADA nodes for proxies registration are regarded as the costs. Which variable to be used as productivity and cost to compute the scalability metric is depending on what system configuration is focused on, which are discussed in detail below.

The scalability of the FADA network with different settings of system parameters are of interest in this research. This includes the maximum number of hops in the lookup procedure, the number of FADA nodes that the proxy of a service is registered in and the rate of SME sending requests to the FADA network. They are analyzed over a scale factor

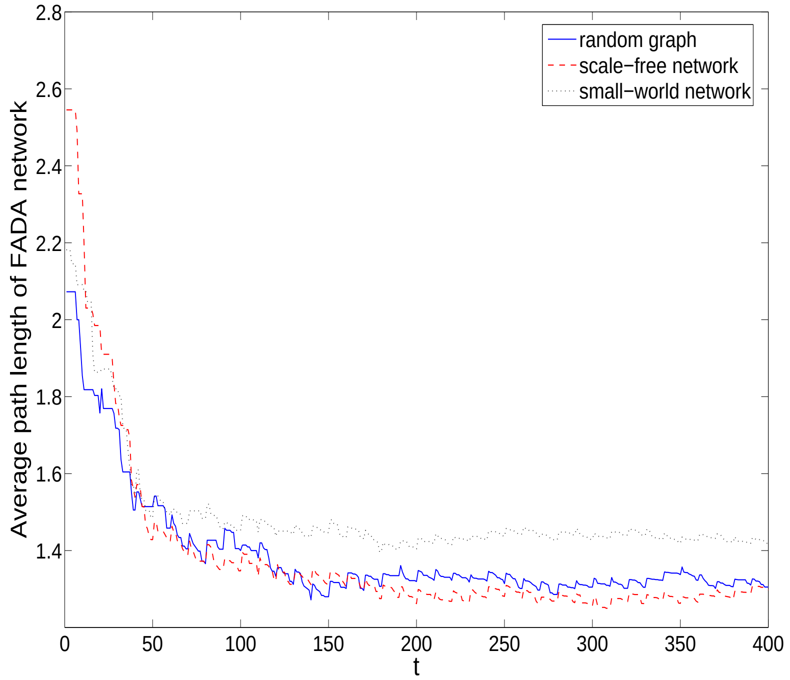


Figure 4.11: The average path lengths of FADA networks which are coupled with the SME business network of different topologies.

interval of 1 to 7. The results and discussion focusing on these three different system parameters are provided respectively in the following sections. As described in the model description in chapter 2, the number of FADA nodes increases as time evolves in the evolving stage in the simulation. The scalability analysis here is also a real-time evaluation of the evolution of the FADA network when the size of the network is growing.

Scalability: the maximum number of hops

Firstly, the effects of different values of the maximum number of hops, $hops_{max}$, on the scalability performance are studied. Both the number of proxies that the FADA network can locate for a service requested, denoted by $n_{proxies}$ and the average success rate of finding proxies to services requested, denoted by r_{succ} are plotted against the scale factor k with different $hops_{max}$. The results are shown in Fig. 4.13. It is shown that when $hops_{max} = 1$, $n_{proxies}$ and r_{succ} drop steeply up to $k = 1$, and then fluctuate within a

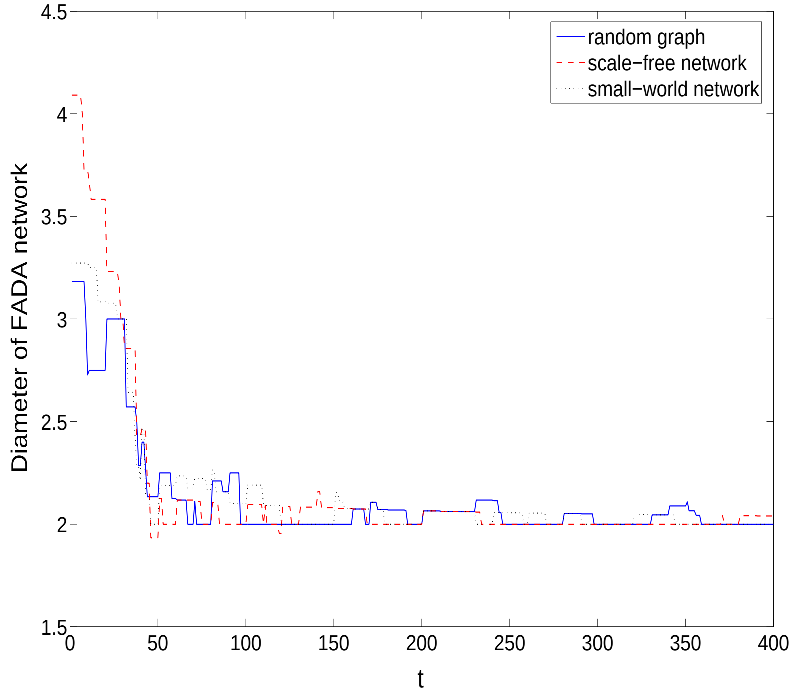
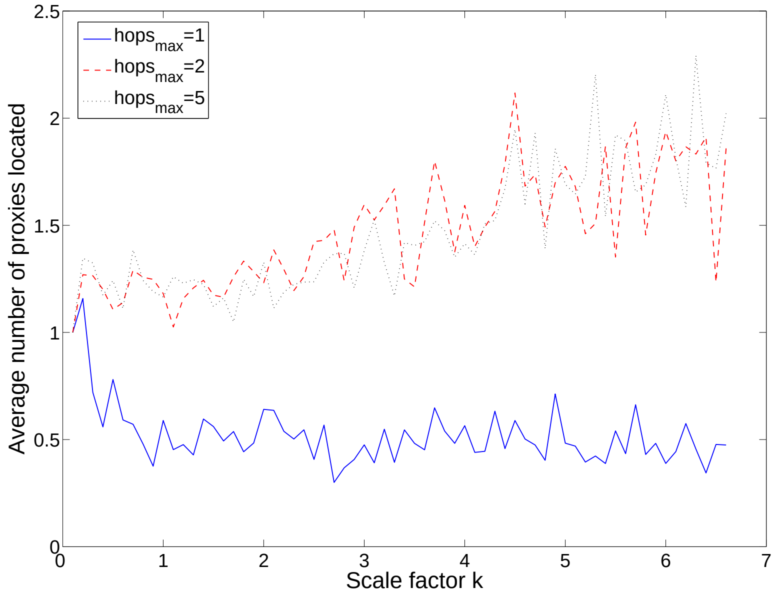


Figure 4.12: The network diameters of FADA networks which are coupled with the SME business network of different topologies.

range. When $hops_{max} = 2$ and 5, $n_{proxies}$ and r_{succ} are much larger than the case of $hops_{max} = 1$. Moreover, as scale factor k increases, $n_{proxies}$ increases gradually which indicates a good availability, and r_{succ} remains 1 which implies a guaranteed performance of the system of finding proxies.

On the other hand, the cost is calculated by the number of nodes visited when propagating a request in the FADA network according to the flooding algorithm. It is denoted as c_{nv} and the result is presented in Fig. 4.14. It is shown that when $hops_{max} = 1$, the cost is much less than when $hops_{max} = 2$ and 5. As k increases, c_{nv} remains at a small value when $hops_{max} = 1$ and when $hops_{max} = 2$ and 5, it increases steeply. As is known, when the maximum number of hops is bigger, the number of nodes visited should be larger. But the result in Fig. 4.14 shows that c_{nv} is larger when $hops_{max} = 2$ than when $hops_{max} = 5$ with respect to the scale factor k. This is because the initial value of c_{nv} at $k_1 = 11$ when $hops_{max} = 5$ is bigger than when $hops_{max} = 5$. Therefore, the average number of nodes



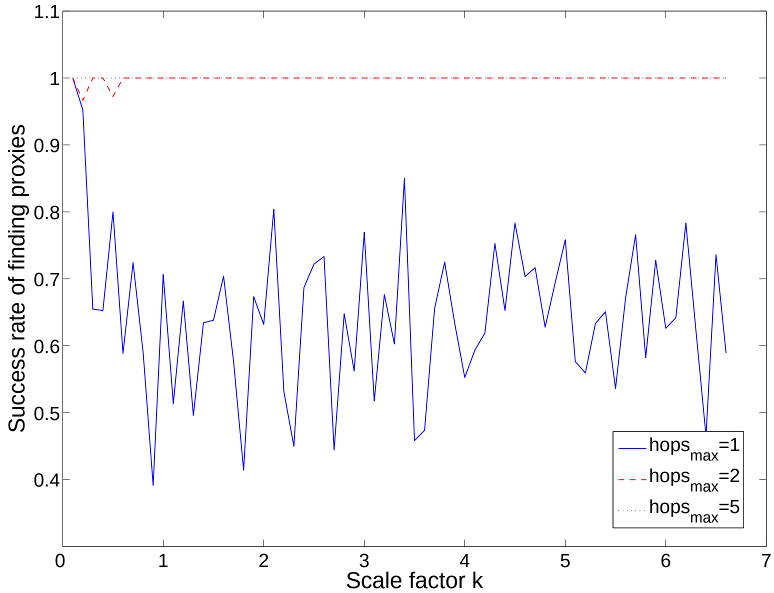


Figure 4.13: Productivity of the system v.s the scale factor with different values of $hops_{max}$**.**

visited is not in order with respect to $hops_{max}$. It also indicates that some performance of such a complex system is not predictable and the studies on it have to rely on simulations.

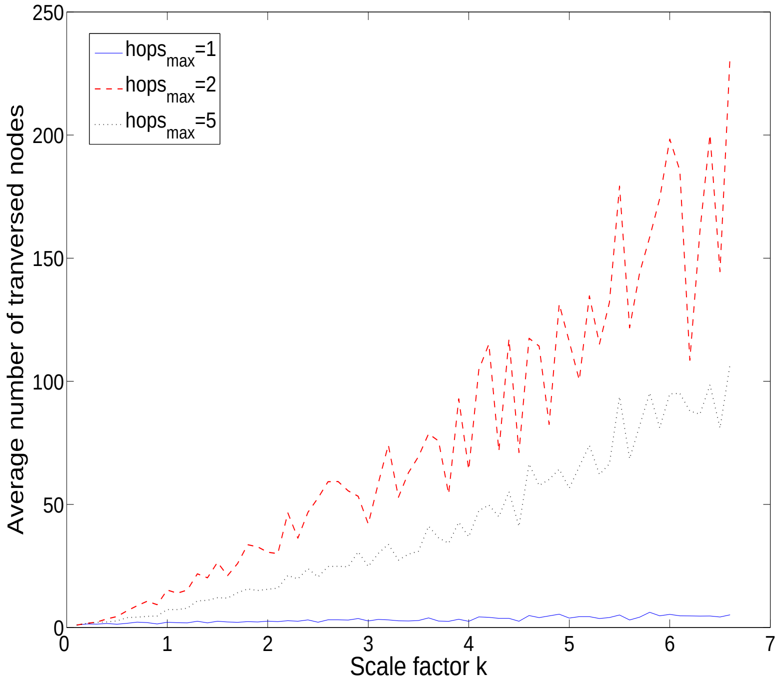


Figure 4.14: Cost with respect to the number of nodes visited when propagating a request in the FADA network v.s the scale factor with different values of $hops_{max}$.

Furthermore, by combining the result shown in Fig. 4.13 and Fig. 4.14, it can be seen that there is a trade-off between the productivity and the cost. For better understanding, two metrics the scalability metric 1 and scalability metric 2 are defined based on $n_{proxies}$, r_{succ} and c_{nv} in the equations below:

$$scalability\ metric\ \psi_1(k) = \frac{n_{proxies}(11)}{n_{proxies}(k)} \cdot \frac{c_{nv}(k)}{c_{nv}(11)} \tag{4.2}$$

$$scalability\ metric\ \psi_2(k) = \frac{r_{succ}(11)}{r_{succ}(k)} \cdot \frac{c_{nv}(k)}{c_{nv}(11)} \tag{4.3}$$

The results on the scalability metric ψ_1 and ψ_2 are presented in Fig. 4.15. In Fig. 4.15, it is shown that the scalability is not very reasonable. The scalability metrics degrade steeply up to $k = 1$ and then more gradually. ψ_1 drops to roughly 0.2 at $k = 1$ and ψ_2 drops to approximately 0.3, 0.2 and 0.1 at $k = 1$ when $hops_{max} = 1$, 5 and 2 respectively.

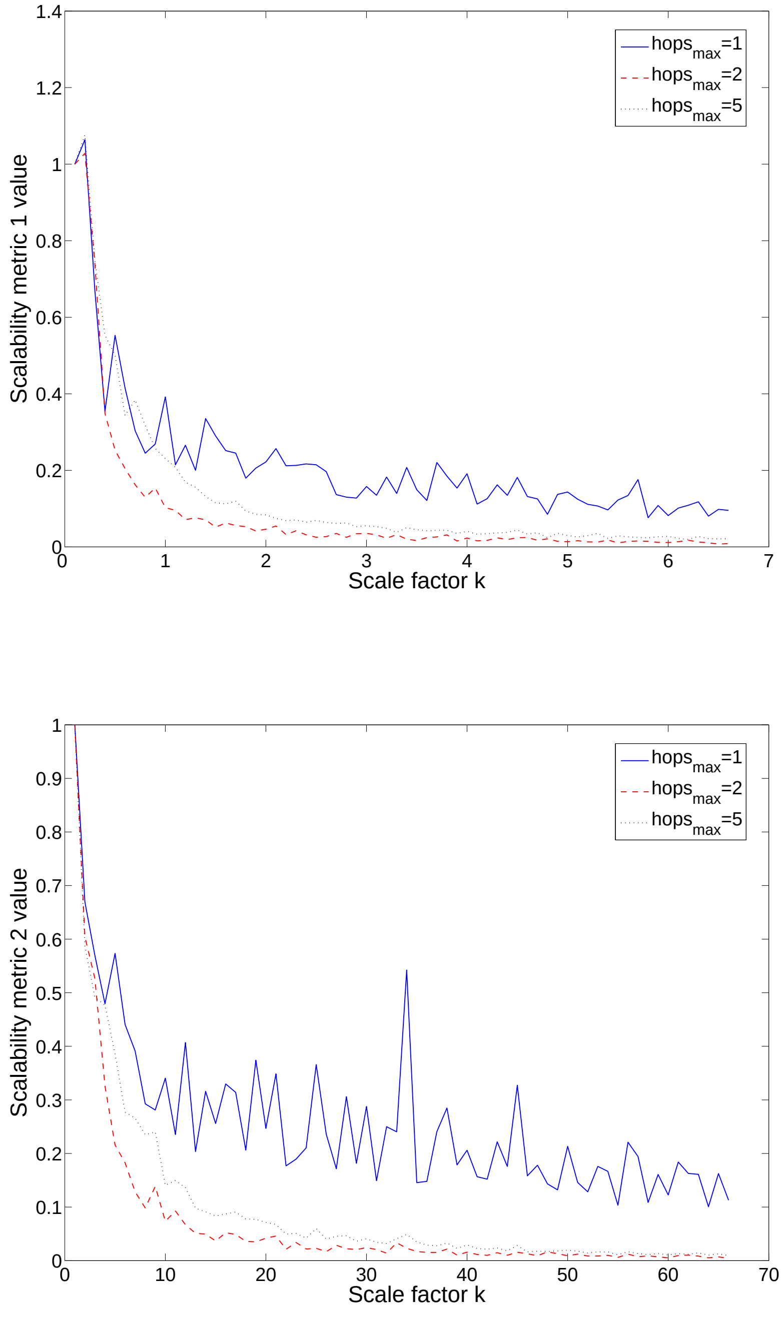


Figure 4.15: Scalability metric ψ_1 and scalability metric ψ_2 v.s the scale factor with different values of $hops_{max}$.

The reason for this is that as the number of FADA nodes increases, the cost to propagate a request in the network increases rapidly due to the flooding algorithm, while the

productivity of the network is not improved. By comparison, it is noticeable that when $hops_{max} = 1$, the values of both the scalability metric ψ_1 and ψ_2 are the best among the cases when $hops_{max} = 1$, 2 and 5.

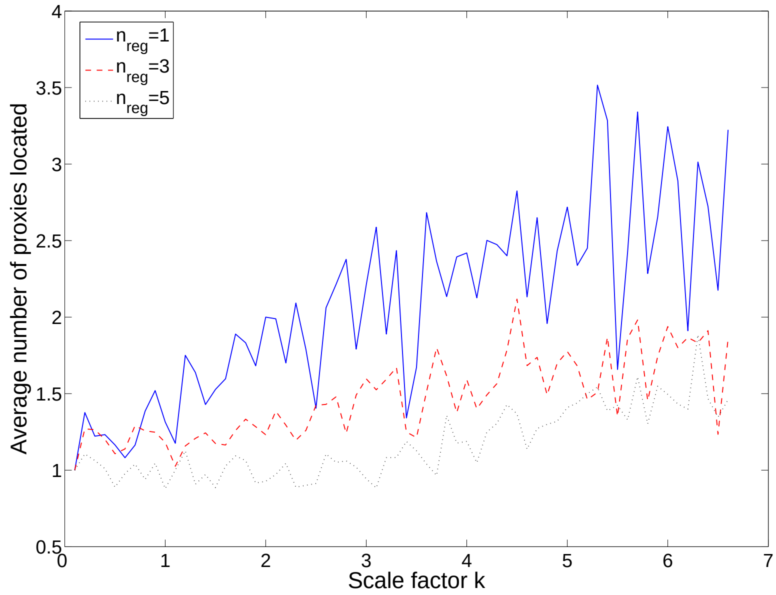
Scalability: the number of FADA nodes for registrations of a proxy

As SMEs register proxies of their services in FADA nodes, it is interesting to investigate how the number of FADA nodes that one proxy is registered in, denoted by n_{reg}, affects the system performance along with the size of the FADA network growing. Here $n_{proxies}$ is regarded as the productivity of the system as in the previous discussion and the average memory usage over all FADA nodes, denoted as c_{mem} is regarded as the cost of registering the proxies. The results are presented in Fig. 4.16.

It is shown in Fig. 4.16(a) that $n_{proxies}$ tends to increase with k. In spite of fluctuation within a range, it is still noticeable that the increase of $n_{proxies}$ when $n_{reg} = 1$ is faster than when $n_{reg} = 3$ and 5. In Fig. 4.16(b), it can be seen that although the number of FADA nodes increases which means that there is more and more memory available for proxies registrations, the average memory used for proxies registrations increases linearly as k increases. This can be explained as the following: as described in the model description in chapter 2, the rate of new FADA nodes being generated is set to be the same as the new SMEs joining in the system. But new joining SMEs register proxies in FADA nodes continuously and proxies of popular services are registered additionally in more FADA nodes. Therefore, although the number of FADA nodes increases to offer more memory and also the leases to proxies are modelled in an intelligent way that the leases time is allocated depending on the remaining memory of FADA nodes, the memory consumption in the FADA nodes is much more.

The scalability metric 3 is defined based on $n_{proxies}$ and c_{mem} as the following:

$$scalability\ metric\ \psi_3(k) = \frac{n_{proxies}(11)}{n_{proxies}(k)} \cdot \frac{c_{rem}(k)}{c_{rem}(11)} \tag{4.4}$$



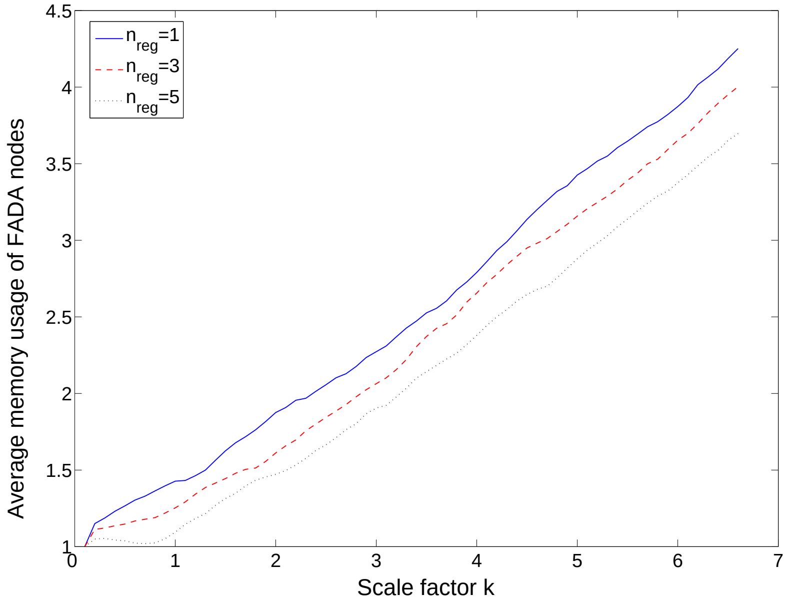


Figure 4.16: (a) Productivity based on the number of proxies being found to requested services (b) Cost based on the memory usage of FADA nodes for registering proxies in the FADA network v.s the scale factor with different values of n_{reg}.

The result on ψ_3 is presented in Fig. 4.17. It is shown that the values of scalability metric 3 increase along with k when $n_{reg} = 1$, 3 and 5, which implies a reasonable scalability performance in terms of the availability of finding proxies and the cost of memory consumption for proxies registrations. Furthermore, the system with $n_{reg} = 1$ displays the best performance on the scalability metric 3.

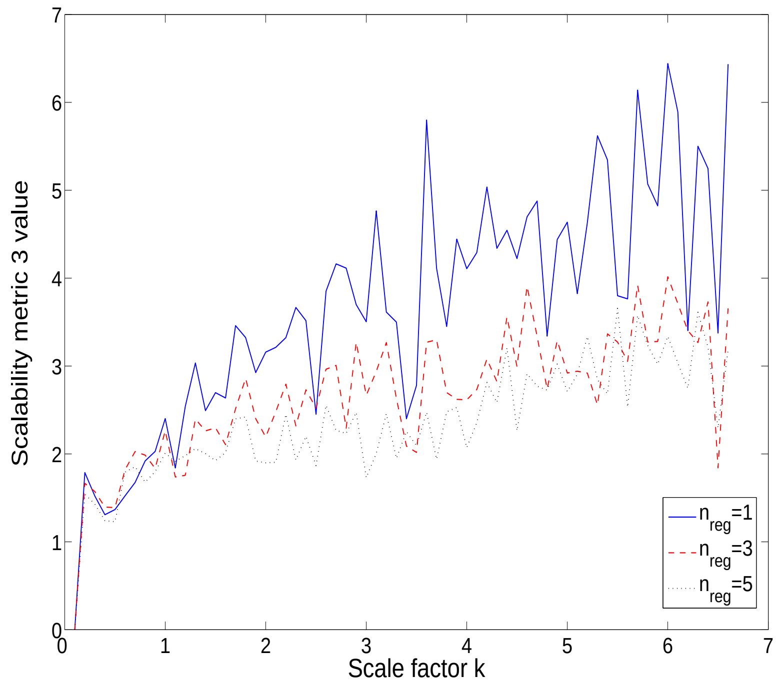


Figure 4.17: Values of the scalability metric ψ_3 v.s the scale factor with different values of n_{reg}.

Scalability: request rate

One major interaction between the SME network and the FADA network is that SMEs send requests to the FADA network to look for proxies of services so that real services required can be executed. How the increase of the rate of sending requests to the FADA network r_{req} affects the performance of the FADA network is of interest here.

The result on the minimum number of hops to find proxies as r_{req} increases is plotted in Fig. 4.18. Each data presented in the figures in this discussion corresponds to the average value over the period of the network evolving from $t = 1$ until $t = 400$. The result in

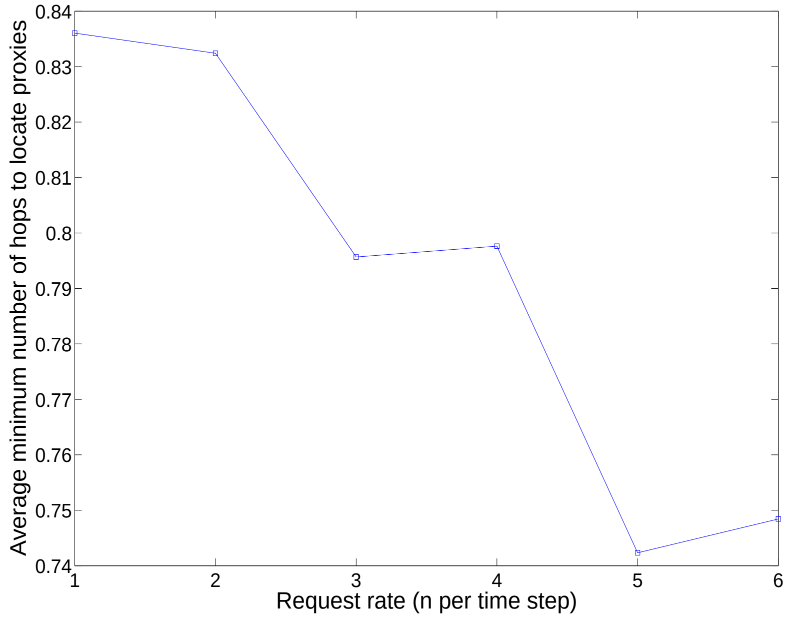


Figure 4.18: The effects of increasing rate of sending requests to the FADA network on the average minimum number of hops to locate proxies to request services.

Fig. 4.18 shows that when r_{req} is larger, proxies can be found within less hops. The explanation is as the following: because of the flooding algorithm used in the lookup procedure in the FADA network, new connections are made between nodes which contain the proxies to services requested and the node which initiates the lookup request. Therefore, the more requests being sent to the FADA network, the more connections are generated in the FADA network which makes the network more connected. As a result, generally it takes less hops to find proxies in some nodes from the initiator node. The result in Fig. 4.19 supports the explanation above. Nevertheless, if the cost of propagating a request in a more connected network is considered, then the number of nodes visited with larger r_{req} is more than with smaller r_{req} which is shown in Fig. 4.20. It can be found that beyond $r_{req} = 3$, the number of nodes visited reaches a limit. It is conjectured that when r_{req} is larger than 3, it is big enough to cause the FADA network to become fully connected in a very short time. Therefore, there is no difference in the effects of different r_{req} on the network when it is beyond $r_{req} = 3$.

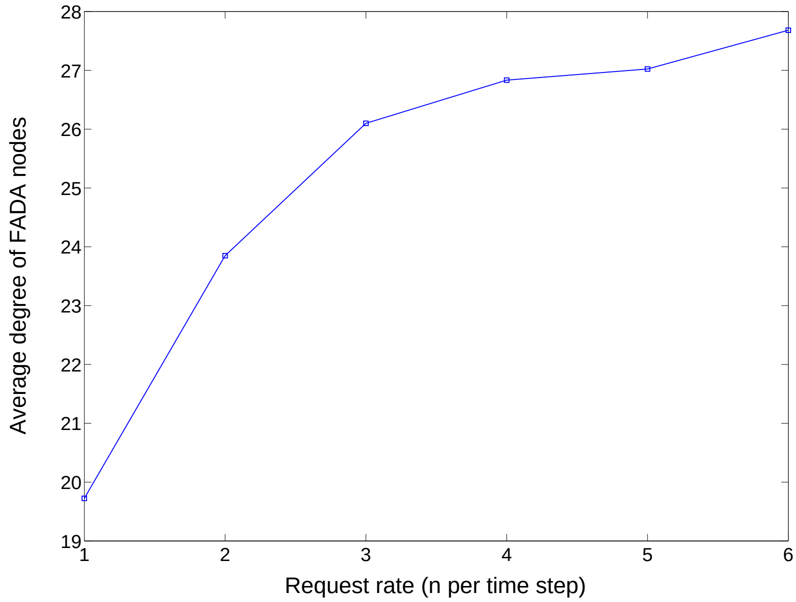


Figure 4.19: The average degree of FADA nodes with the increasing rate of sending requests to the FADA network.

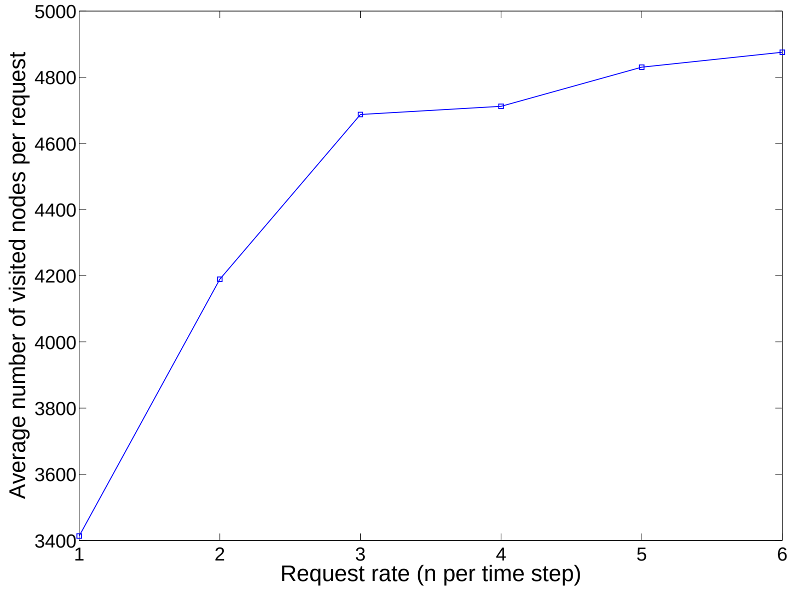


Figure 4.20: The effects of increasing rate of sending requests to the FADA network on the average number of nodes visited for propagating a request in the FADA network.

Chapter 5

Conclusion

In this report, a two coupled networks model of more complexity compared to the model that was introduced in the D19.2 has been presented as a simplified version of the DBE system. The SME business network is regarded as an information ecosystem, where behaviours of business processes such as exchanges of services between SMEs and optimizations of service solutions are modelled. The FADA network is the communication infrastructure which supports the business processes of the SME network. These two networks interact with each other and evolve over time.

Although business processes and communication infrastructure that supports the processes interact in practice, very few research has worked on how the changes on any of the two could have impacts on the other. The discrete-event simulation tool was used to show how to study and understand the dynamic relationship between the two. These studies primarily focused on three aspects as the following: 1) the effects of graph connectivity of the SME network on the system; 2) the effects of different topologies of the SME network on the system; 3) the scalability of the FADA communication network when it supports real-time business processes in the SME network.

The main findings can be summarized as follows:

1. The graph connectivity of the SME network matters: the SME network is con-

structured and evolved based on the ER random graph model with various connectivity probability p_{SME}. With larger p_{SME}, the fitness values of solutions that the DBE system can offer are higher. Furthermore, it was found that the scaling law between the p_{SME} and the usage records of services in SMEs is linear. The results also showed the effects of p_{SME} on the topology of the FADA network. With larger p_{SME}, the average degree of FADA network is higher, which results in smaller average path length. However, different values of p_{SME} have insignificant effects on the diameter of the FADA network.

2. The topologies of the SME network have no effects on the performance of SME network but do have effects on the FADA network: three most popular models of topologies in the filed of complex networks are applied to the SME network, which are the random graph, the scale-free network and the small-world network. The results showed that different topology models have trivial effects on the SME network. Nevertheless, it has been shown that by using small-world model for the SME network, the average degree of the FADA nodes is smaller and the average path length of the FADA network is larger than by using the other two topology models for the SME network. Again, the effect on the diameter of the FADA network is insignificant.

3. The FADA network shows different scalability performances when considering different aspects of the system: due to the complexity of the FADA network and different aspects to be analyzed, three scalability metrics accordingly have been defined. Effects of different system parameters on the scalability have been investigated, including parameters of the maximum number of hops that a request can be propagated $hops_{max}$, the number of FADA nodes for registrations of a proxy n_{reg} and the request rate of sending a request from the SME network to the FADA network r_{req}. The result showed that when the size of the FADA network increases, the scalability is not reasonable when the number of nodes visited during the lookup procedure is regarded as the cost. However, the scalability displays quite reasonable behaviours by considering the memory usages of FADA nodes for registrations of proxies of services

as the cost. Furthermore, when $hops_{max} = 1$ and $n_{reg} = 1$, the system showed the best performance compared to other values for the parameters. On the other hand, it was found that when r_{req} increases, it takes more hops to find proxies for services requested. Moreover, as r_{req} increases, the average degree of the FADA nodes, as well as the number of nodes visited, increase steeply until they reach a limit. The critical point is at $r_{req} = 3$ (requests per time step) approximately. From the results, it was conjectured that when r_{req} is beyond 3, that the FADA network becomes a fully connected network.

Bibliography

[1] Reka Albert and Albert-Làszlò Barabàsi. Statistical mechanics of complex networks. *Review of Modern Physics*, 74:47, January 2002.

[2] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small world' networks. *Nature (London)*, 393:440–442, 1998.

[3] *DIET Web site*. http://www.dfki.uni-kl.de/IVSWEB/DIET/public/index.html.

[4] Bengt Carlsson and Andreas Jacobsson. On contamination in information ecosystems. In *Proceddings of the 38th Hawaii International Conference on System Sciences*, January 2005.

[5] *Object Management Group. The Common Object Request Broker: Architecture and Specification*, 3.0 edition edition, July 2002.

[6] Sun Microsystems, http://java.sun.com.

[7] http://www.microsoft.com/com/tech/dcom.asp.

[8] Stanley Y. W. Su, Youzhong Liu, and Jie Meng. Distributed and concurrent processing of business object documents in support of e-enterprise integration. *4th International Enterprise Distributed Object Computing Conference*, September 2000.

[9] Stanley Y. W. Su, Herman Lam, and Minsoo Lee. An information infrastructure and e-services for supporting internet-based scalable e-business enterprises. *5th International Enterprise Distributed Object Computing Conference*, September 2001.

[10] Yves Caseau. Self-adaptive middleware: supporting business process priorities and service level agreements. *Advanced engineering informatics*, 19:199–211, May 2005.

[11] T. H. Davenport. *Process innovation: reengineering work through information technology*. MA: Havord business school press, Boston, 1993.

[12] Cheri Speier, Michael Harvey, and Jonathan Palmer. Implementing intra-organizational learning: an infrastructure to support business process reengineering. *Knowledge and process management*, 5(2):76–86, 1998.

[13] Michael K. Painter, Ronald Fernandes, Natarajan Padmanaban, and Richard J. Mayer. A methodology for integrating business process and information infrastructure models. *Proceedings of the 1996 winter simulation conference*, pages 1305–1312, December 1996.

[14] Julie Eatock, Ray J. Paul, and Alan Serrano. Developing a theory to explain the insights gained concerning information systems and business process behaviour: The assess-it project. *Information systems frontiers*, 4(3):303–316, 2002.

[15] Ray J. Paul. Collaborative information systems and business process design using simulation. *Proceddings of the 37th Hawaii international conference on system sciences*, January 2004.

[16] Jörgen Weibull. *Evolutionary game theory*. Cambridge, MA: The M.I.T Press, 1995.

[17] John Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, Michigan, 1975.

[18] M. E. J. Newman and D. J. Watts. Scaling and percolation in the small-world network model. *Physical Review E*, 60(6), December 1999.

[19] M. E. J. Newman and D. J. Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263:341–346, December 1999.

[20] Shudong Jin and Azer Bestavors. Small-world characteristics of internet topologies and implications on multicast scaling. *Computer Networks*, April, 2005.

[21] Rajesh Kasturirangan. Multiple scales in small-world graphs. *arXiv:cond-mat*, (9904055), September 2005.

[22] Jon M. Kleinberg. Navigation in a small world. *Nature*, 406:845, August 2000.

[23] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167 – 256, 2003.

[24] Jonathan Ozik, Brian R. Hunt, and Edward Ott. Growing networks with geographical attachment preference: Emergence of small worlds. *Physical Review E*, 69(026108), February 2004.

[25] Prasad Jogalekar and Murray Woodside. Evaluating the scalability of distributed systems. *IEEE Transactions on Paralel and Distributed Systems*, 11(6):589–603, 2002.

[26] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Advanced Physics*, 51:1079–1187, 2002.

[27] Duncan J. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, 1999.

[28] Paul Erdös and Alfrad Rényi. On the evolution of random graphs. *Publicationes Mathematicae (Debrecen)*, 6:290, 1959.

[29] L. SattenspielCarl and P. Simon. The spread and persistence of infectious diseases in structured populations. *The spread and persistence of infectious diseases in structured populations*, 90(1-2):341–366, July-August 1998.

[30] Stanley Milgram. The small world problem. *Psychology Today*, 61, 1967.

[31] Chunguang Li, Hongbing Xu, Xiaofeng Liao, and Juebang Yu. Synchronization in small-world oscillator networks with coupling delays. *Physica A*, 335(3-4):359–364, April 2004.

[32] Damián H. Zanette. Dynamics of rumor propagation on small-world networks. *Physical Review E*, 65(041908), March 2002.

[33] Nouredine Zekri and Jean Pierre Clerc. Statistical and dynamical study of desease propagation in a small-world. *Physical Review E*, 64(056115), November 2001.

[34] Theoden I. Netoff, Robert Clewley, Scott Arno, Tara Keck, and John A. White. Epilepsy in small-world networks. *The Journal of Neuroscience*, 24(37):8075–8083, September 2004.

[35] Albert-Làszlò Barabàsi and Reka Albert. Emergence of scaling in random networks. *Science*, 286:509, October 1999.

[36] S. Redner. How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B*, 4(2):131–134, 1998.

[37] R. Albert, H. Jeong, and A.-L. Barabàsi. Internet: Diameter of the world-wide web. *Nature (London)*, 401:130 – 131, September 1999.

[38] B. Huberman and L. Adamic. Internet: Growth dynamics of the world-wide web. *Nature*, 401:131, September 1999.

[39] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Computer Communication Review*, 29(4):251–262, Ocotober 1999.

Appendix A

Topologies of Networks

Networks of agents play an important modelling role in a variety of fields. In general, networks of agents organize themselves according to some purpose. For instance, the network of neurons in the brain is organized to process information and a country is organized to ensure the survival and well-being life of its habitants. These cases might show that the functionality of a network determines its structure.

However, it is worth pointing out that the function of a network is always affected by its structure. For example, the spread of information and disease is affected by the topology of social and organizational networks, and the robustness and stability of power transmission is affected by the topology of the power grid. Therefore, it is important to find out how networks grow, how the pattern of connections, or the shape, is influenced by the growth process, and what mechanisms lead to each type of topology.

The structure of various networks, including social and computer networks, has been widely studied in the physics literature [1, 26]. There have been findings suggesting that some behaviours and properties can be found in a number of different types of networks. Therefore, there are an increasing number of researchers who believe that there must be some laws describing these networks, and once these laws are found for a certain network, they can be utilized in certain other networks and explain the behaviour more correctly.

In spite of advances in understanding the structures of networks, study on large, complex networks and their dynamic properties, especially for networks that change in time, is relatively minor. Therefore, this dissertation emphasizes on studying topological properties of evolving networks of interacting agents, and how behaviours of the systems are affected by different types of topologies applied on them.

A.1 Topological Properties

The structure of networks has been studied by graph theory. A network is regarded as a graph, where vertices (or nodes) represent entities in the network and arcs (or links) represent interactions between the entities.

Many quantities and measures of complex networks have been proposed and investigated. In this section, the most important graph-theoretic properties that are used to measure a network's topology in this dissertation are introduced. These are degree distribution, average path length, diameter, and clustering coefficient.

A.1.1 Degree and Degree Distribution

The number of edges connected to a vertex is called the *degree* of the vertex. In a network, vertices always have different degree. The degree of a node is characterized by a distribution function $P(k)$, which is the probability that a randomly chosen node has exactly k connections. Another important characteristic of a graph is the average degree (or average number of links), as it gives a relationship between the size of the graph and its sparseness [27]. A high average degree means individual vertices will have a very large number of neighbours, which place a higher burden on communication.

A.1.2 Average Path Length and Diameter

The average path length is the average of the shortest path lengths from each node to every other node. It characterizes the spread of a graph. For example, it affects the number of "hops" that a packet must take to get from one computer to another computer on the Internet, how fast requests are propagated through a network, the time it takes for a disease to spread through a population, and so on. Another important way to characterize the spread of a graph is to calculate the diameter, which is defined as the longest distance between any two nodes.

A.1.3 Clustering Coefficient

A common property of social networks is the form of cliques, which represent circles of friends, where every member knows every other member. In a graph, it is called a cluster, when there is a subset of nodes that are fully connected to each other. It is quantified by the clustering coefficient [2] as the probability that two of a particular node's nearest neighbours are connected. A graph with a high clustering coefficient implies a large number of clusters, while a low clustering coefficient implies disconnected vertices.

In this dissertation, three recently most popular topologies are used as modelling paradigms: random graphs, small-world models, and scale-free models.

A.2 Random Graph

Network topologies such as chains, grids, lattices and fully-connected graphs are regular graphs. These topologies are simple for analysis without being bothered by complexity in the network structure itself. However, networks with a complex topology and unknown organizing principles are often random. Therefore, people start to think about graphs that are completely random.

The theory of random graphs was introduced by Erdös Rényi (ER), as an early effort to model complex networks in [28] about 40 years ago. In an ER random graph, edges are distributed randomly and the presence or absence of any edge between two nodes in the network is dependent on a fixed connection probability p. Random graph theory aims to determine at what p, a particular property of a graph will most likely appear.

The construction (or the evolution process) of a random graph is as the following: starting with N isolated nodes, then connect every pair of nodes with probability p. The graphs obtained at different stages correspond to larger and larger p, until obtaining a fully connected graph for $p \to 1$. An example of the evolution process is shown in Fig. A.1.

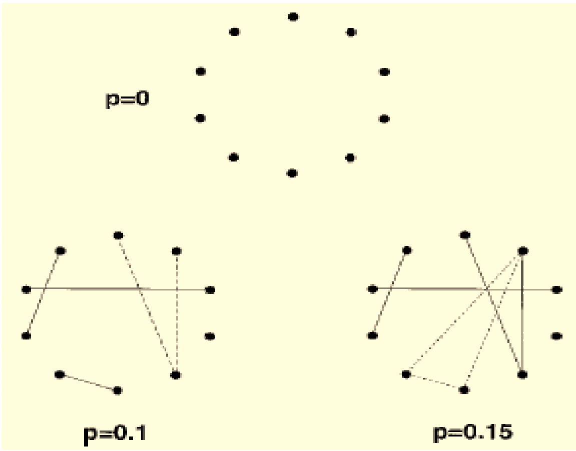


Figure A.1: [Courtesy of [1]]The graph evolution process for Erdös Rényi (ER) model.

The degree distribution of a random graph has been found to follow a Poisson distribution:

$$P(k) = \binom{n}{k} p^k (1-p)^{N-k} \simeq \frac{z^k e^{-z}}{k!}, \quad k \geq 0.$$

Here k represents nodes having exactly k edges, n is the total number of edges and N is the number of nodes in the network. Moreover, the average path length of a random graph has been found to scale with the number of nodes:

$$\ell \sim \frac{\ln N}{\ln \langle k \rangle}. \tag{A.2}$$

Various types of networks in the real world have been modelled as random graphs, particularly in epidemiology studies [29].

A.3 Small-world Networks

Many real-world networks such as the World Wide Web, the Internet, communication networks, electric power grid and networks of movie actors, have been found to display the so-called "small-world network" phenomenon. These small-world networks are characterized by two primary properties, which are small average path length and unusually large clustering coefficient. The most famous manifestation of the small worlds is the concept of "six degree separation", which has been concluded by the psychologist Stanley Milgram in 1967 [30] that two random US citizens were connected by an average of six acquaintances.

In 1998, Watts and Strogatz firstly proposed a simple model (WS small-world model) to generate small-world networks [2]. The inspiration of the WS model is based on the fact that many networks actually lie somewhere between the extremes of order and randomness. Although regular networks and random graphs are both useful idealizations, most of the real-world networks are neither entirely regular nor entirely random. The procedure of the WS model is as the following: it starts with a ring lattice with n vertices and each vertex connects to its k nearest neighbors. Each edge is rewired at random with probability p. The parameter p is able to control the graph to be between completely regular ($p = 0$) and completely random ($p = 1$), as illustrated in Fig. A.2.

The small-world networks model has been widely used for real-world applications, such as oscillator networks[31], disease propagation[32], information propagation[33] and neural networks[34].

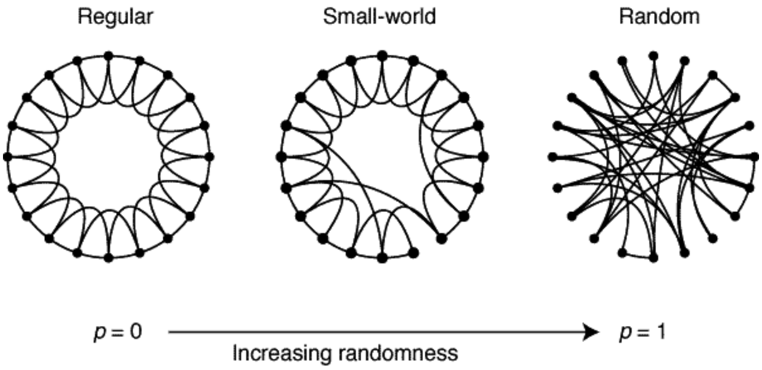


Figure A.2: [Curtersy from [2]]The rewiring procedure of WS small-world network model.

A.4 Scale-free Networks

A common characteristic of the ER random graph and the WS small-world model is that the degree distribution of the network is homogenous, with peak at an average value. However, recently, some real-world networks, such as networks of movie actor collaboration [2, 35], science collaboration [36], WWW [37, 38] and Internet [39], have been found to have a degree distribution that follows a power-law with different exponents: $P(k) \sim k^{-r}$, which deviates from poisson distribution of the random graph and WS small-world model. The power-law distribution means that a few network nodes are far more connected than other nodes. It indicates the networks can self-organize to a scale-free state, where the highly connected "hub" nodes strongly affect the structure and dynamics of the networks. The average path length of a scale-free network has been found to increase approximately logarithmically with the number of nodes, N. Furthermore, it has also been observed that the average path length ℓ of a scale-free network is smaller than a random graph of the same network size and average degree.

In 1999, Barabàsi and Albert (BA) proposed another model to explain the origin of power-law degree distribution. The BA model suggests that the scale-free structure is the consequence of two main ingredients [35]:

1. Growth of nodes: at each time step, a new node is added, with some edges connecting to nodes that already present in the network.

2. Preferential attachment to well connected nodes: the probability of a new node connecting to a node depends on the degree of the node. In other words, a new node is more likely to connect to nodes that are highly connected.

The algorithm of BA scale-free model is as the following: starting with m_0 nodes, and at each time step, add a new node with m edges connecting to pre-existing nodes. The probability $\prod_i$ that a new node will be connected to node i depends on the degree k_i of node i, in such a way that $\prod_i = k_i / \sum_j k_j$.