A Middleware for Targeted Marketing in Spontaneous Social Communities

by

Zhao Tian

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the M.Sc. degree in
Electronic Business Technologies

Supervisor
Dr. Nancy Samaan

Electronic Business Technologies
Telfer School of Management
University of Ottawa

Abstract

With the proliferation of mobile devices and wireless connectivity technologies, mobile social communities offer novel opportunities for targeted marketing by service or product providers. Unfortunately, marketers are still unable to realize the full potential of these markets due to their inability to effectively target right audiences. This thesis presents a novel middleware for identifying spontaneous social communities (SSCs) of mobile users in ad hoc networks in order to facilitate marketers' advertisements. The contributions of the presented work are two fold; the first is a novel model for SSCs that captures their unique dynamic nature, in terms of community structure and interest in different *hot-topics* over time. These time-varying interests are represented through an inferred *community profile prototype* that reflects dominant characteristics of community members. This prototype is then employed to facilitate the identification of new potential members. The selected community prototypes are also used by marketers to identify the right communities for their services or products promotions. The second contribution of this paper is novel distributed techniques for efficient calculation of the community prototypes and identification of potential community links. In contrast to traditional models of detecting fixed and mobile social networks that rely on pre-existing friendships among its members to predict new ones, the proposed model focuses on measuring the degree of similarity between the new user's profile and the profiles of members of each community in order to predict new users' relationships in the community. The adopted model of SSCs can foster many existing and new socially-aware applications such as recommender systems for social events and tools for collaborative work. It is also an ideal target for business-oriented applications such as short-message-service (SMS) advertisement messages, podcasting news feeds in addition to location/context-aware services. The performance of the proposed work was evaluated using the NetLogo platform where obtained experimental results demonstrate the achieved high degree of stability in the resulting communities in addition to the effectiveness of the proposed middleware in terms of the reduction in the number of routing messages required for advertisements.

Acknowledgements

I would like to thank my supervisor, Dr. Nancy Samaan for her guidance and encouragement throughout the whole period of my study. And I also want to thank my parents and sister, their support has always been found during the research.

Publications

Zhao Tian and Nancy Samaan, **"A Novel Distributed Scheme for Building Spontaneous Social Communities over MANETs"**, the Ninth IEEE International Conference on Mobile Ad-hoc and Sensor Systems (IEEE MASS 2012), October, in Las Vegas, Nevada, USA, 2012

Nomenclature

Acronym	**Definition**
SCs	Social Communities
MSNs	Mobile Social Networks
SSCs	Spontaneous Social Communities
MANETs	Mobile Ad Hoc Networks
SMS	Short Message Service
SNs	Social Networks
SNA	Social Network Analysis
OSNs	Online Social Networks
API	Application Programming Interface
GPRS	General Packet Radio Service
HSPDA	High-Speed Downlink Packet Access
PSN	Pocket Switched Networks
PM	User-profile Manager
CM	Community Manager
FM	Friendship Manager
RDF	Resource Description Framework
OWL	Web Ontology Language
P2P	Peer to Peer
FOAF	Friends of a Friend
SQL	Structured Query Language
ARAS	Average Received Advertising Score

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 An Overview

The unprecedented development and swift convergence of telecommunication technologies over the past decades accompanied by the revolution of Internet web-services has resulted in developing new unforeseen ways of interpersonal communications and social interactions. In particular, numerous participatory sites and social networking services utilized advanced software technologies to provide new forms of collaboration, cooperation and communication platforms to web surfers all over the world. The boom of these social networking services crossed out the limitation of one simple communication platform, was the case in the area of web 1.0, to provide a robust platform where users can engage in more integrated forms of collaboration and communication. The surge in smart mobile devices has pushed social networking platform developers to extend their services to mobile users. Figure 1.1 demonstrates the unprecedent increasing number of smartphone users while Figure 1.2 depicts the percentage of users that access social networking sites using their smartphones. As a result, a natural trend to integrate social networking services with mobile devices has been continuously growing. It has also been a hot topic for current research efforts. Since more and more people are using mobile devices to enjoy social networking services, these social communities hosting mobile users represent an ideal opportunity for marketing activities by advertising businesses. Unfortunately, the full potential of these markets has not been fully explored yet [4]. The first obstacle for marketers is the efficient identification of potential audiences in such dynamic environments.

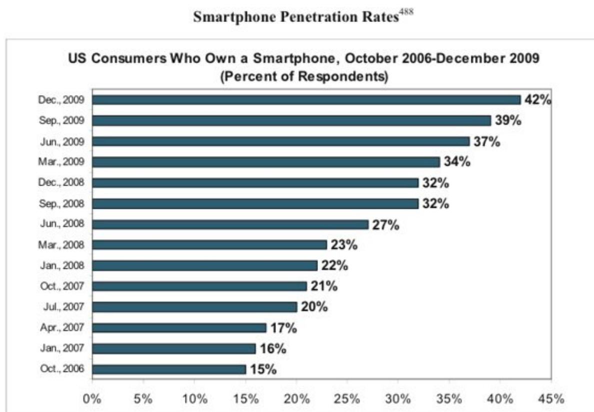

Figure 1.1: Smartphone Penetration Rate in USA [1]

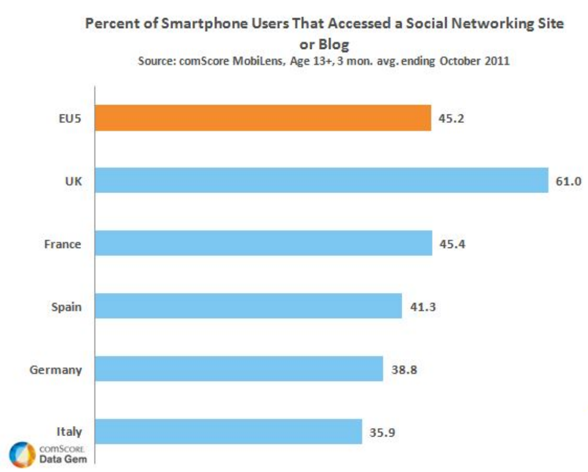

Figure 1.2: Percentage of Smartphone Users Accessing Social Networking Sites or Blogs

1.2 Motivation

Online social communities (SCs) [5] have become an integral part of our daily life. They provide a common virtual place for people who share the same interests or goals to interact and exchange their views as well as capture media contents. These communities also provide the means for people to keep in touch with friends and/or publish thoughts and opinions on current events. They also represent ideal channels for targeted marketing. In fact, they represent exceptional, low cost marketing opportunities where millions of customers can be reached within a very short period. Yet, in order for marketers to take advantage of these markets, they need an efficient software tool that can be employed to identify potential customers.

Due to the demands of integrating online social communities with mobile devices, some social networking sites provided additional versions of their products suitable for mobile devices. Example of these communities are Facebook, Twitter, MySpace and LinkedIn. For example, the statistics from Facebook proved that mobile versions are also growing fast as their computer-supported siblings: more than 350 million active users currently access Facebook through their mobile devices and more than 475 mobile operators globally work to deploy and promote Facebook mobile products [6]. But these services were limited to complement the user experience in mobile circumstances by allowing mobile users to access the same SCs using a limited version of their application that is suitable for mobile devices. In other words, these services worked under the assumption that mobile users have continuous access to a centralized server that manages the users' profiles and coordinates their interactions. To address this limitation, several research efforts have proposed a number of solutions; for example, some approaches relaxed the assumption of continuous access by allowing mobile users to carry a list of their social networking contacts to be updated whenever they have access to the Internet (e.g., the CenceMe architecture [7]). Other approaches such as that of Veneta [8] employ knowledge about existing friends to recommend new friendships. Nonetheless, the majority of existing solutions maintained the same assumption of static SCs where users only interact with friends or people they probably know or at most friends of their friends. They also assume that communication links between two users in a SC can only be established if they share similar long-time interests or preferences together.

The thesis proposes a more general model of mobile social networks (MSNs), which is referred to as spontaneous social communities (SSCs) where short- and long-term communities can be formed by mobile users on-the-go without necessarily imposing any assumptions about any pre-existing relations among them. Unlike the Facebook mobile versions, the proposed middleware framework can be deployed over mobile ad hoc networks (MANETs) environment without the need for additional centralized servers. This middleware facilitates the activities of marketers such as product advertising or service promotion by serving as an interface between potential customers and marketers.

1.3 Objectives and Challenges

In order to build a new model to facilitate the marketing activities over SSCs, the main objectives of the proposed thesis are:

- Find an approach to detect and model dynamic social communities formed by mobile users.
- Facilitate marketing activities over these formed social communities.
- Find a proper approach to evaluate the proposed model in terms of advertising performance.

Meanwhile, there exist some challenges to achieve these objectives. The agile topology of MANETs contributes to the complexity of online social communities detection. The "ad hoc" attribute of MANETs indicates that each user has limited global knowledge of the whole network of other users. Unfortunately, most of current community detection algorithms were designed for stable and static network topologies. So the first mission is to design a proper community detection algorithm which suits the characteristics of SSCs over MANETs and capture the dynamic nature of formed social communities such as community structure, different interest over time. In contrast to web-based environments of SCs, the main features of MANETs can be listed as follows [9] [10]:

- MANETs can be defined as self-configuring networks of mobile devices connected by wireless links. Two users can communicate directly if and only if they are close enough for their devices to detect each other's communication signals.

- As users frequently join and leave the network, the MANETs topology is very dynamic.

- MANETs lack the existence of centralized entities where information about users can be collected. Hence, no global knowledge of MANETs is available which is a big challenge for marketing activities.

- The limited capabilities of current mobile devices make it difficult for marketers to ensure stable and sufficient advertisements spread among targeted audiences.

1.4 Problem Statements

Based on the mentioned challenges above, some problems have to be addressed during the research:

- Find an appropriate approach to model users and decide whether there exist potential friendships between these users.

- Since there is no central server or central station involved in the MANETs, a distributed community detection approach is needed to capture the dynamic nature of mobile users.

- How advertisers can take advantage of the proposed middleware to target right audience at right time?

- Find a platform to model the proposed middleware and evaluate its performance.

1.5 Research Methodology

Design science research was used as the main methodology in this thesis. As defined in [11], Design Science is a fundamentally problem-solving paradigm. It seeks to create innovations that define the ideas, practices, technical capabilities, and products through which the analysis, design, implementation, management, and use of information systems can be effectively and efficiently accomplished [12] [13].

In our research, we followed the main five steps of the cyclical approach of the design science proposed by Takeda, et al [14]: (1) Awareness of problem, (2) Suggestion, (3) Development, (4) Evaluation and (5) Conclusion. Figure 1.3 displays the loop of this methodology.

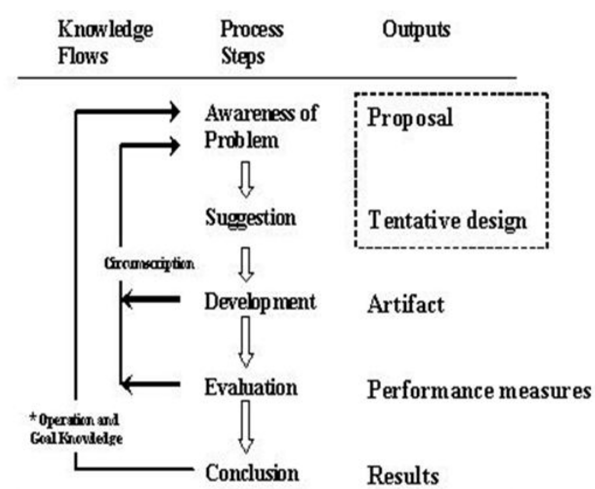

Figure 1.3: The General Methodology of Design of Science [2]

In the first stage of the design science approach I made an effort to find the shortcomings and limitations of the mobile social networks in some literatures. Identified problems have been found such as capturing the dynamic nature of the online communities formed by mobile users, management of the agile MANETs and lack of the advertising approaches over dynamic social communities. In terms of the definition of online communities, author [15] used this term to mean any virtual communities where people come together to get and give information or support, to learn, or to find company.

In the second stage, a tentative design will be made to develop a framework SSCs that models users and networks. Based on the modeling networks, a middleware will be hosted between user interface and transport layer protocol. This proposed middleware will aid the mobile users to identify potential friendships and form the social communities that are of common interest to them. Meanwhile, marketers can target potential audience from these formed social communities to promote their services or products. The term "Middleware" can be defined as "a computer software that provides services to software applications beyond those available from the operating system" [16].

In the development and evaluation stage, I will adopt experimental simulation to develop an IT artifact of the proposed middleware and evaluate its performance. Experimental simulation as its context definition is a simulation method which is conducted in the experimental environment. To build the simulation prototype of the proposed middleware, I will adopt discrete event simulation approach, which is the operation of a system represented as a chronological sequence of events and each event occurs at an instant in time and marks a change of state in the system [17].

Finally, in the fifth stage of the design science, the main outcomes will be illustrated with simulation results. The advertising effectiveness and efficiency of the proposed middleware was validated to show how marketers can take advantage of the proposed middleware to deliver the proper advertisements to the right audience at right time.

1.6 Contributions

A new middleware for constructing social communities over MANETs has been designed. I consider mobile social communities from a broader perspective where spontaneous social communities (SSCs) can be formed by mobile users on-the-go. This perspective captures the unique dynamic nature of these communities, in terms of community structure, member roles as well as community interest in *hot-topics* over time. In a SSC, member users try to infer *predominant interests and properties* of their community based on their interactions with other members. This knowledge is then stored in the form of a distributed community profile prototype. The user then employs this inferred profile in order to identify candidate friends or new members among his/her neighbors. By relating community construction to the most recent interests of the users, I can easily eliminate redundant and expired relations in the SSCs. This, in turn, results in more efficient community structures with a stable number of users. Such communities can provide an ideal media for efficient content sharing and advertisement. To realize these goals, I propose a novel middleware that can be deployed on the users' devices to facilitate the construction and management of users' memberships and interactions in SSCs.

While I focus on simple advertisement message delivery, I envision that the adopted model of SSCs can foster many existing and new socially-aware applications such as recommender systems for social events and tools for collaborative work. It is also an ideal target for business-oriented applications such as short-message-service (SMS) advertisement messages, podcasting news feeds in addition to location/context-aware services. In summary, the main contributions of the presented work can be shown as [18]:

- A novel model for SSCs that captures their unique dynamic nature in MANETs, in terms of community structure and interest in different *hot-topics* over time.
- The second contribution of this paper is novel distributed techniques for efficient calculation of the community prototypes and identification of potential community links. These prototypes help marketers to facilitate targeted marketing to the right communities.

Building on the contribution, I evaluate the proposed framework using the well-known NetLogo platform, where obtained experimental results demonstrate the achieved high degree of stability in the resulting communities in addition to the effectiveness of the advertisements delivery and reduction in the number of routing messages required for content sharing.

1.7 Organization of the Thesis

The reminder of the thesis is organized as follows:

- Chapter 2 provides a detailed overview of mobile ad hoc networks and mobile social networks. Then a literature review of related existing approaches and their characteristics will be added as also.
- Chapter 3 describes the proposed middleware and its main components.
- Chapter 4 introduces the ontologies applied to describe the similarity between users profile feature.
- Chapter 5 introduces the simulation platform and parameters of implementation, then obtained results are provided to demonstrate the performance of the proposed middleware.
- Chapter 6 concludes the thesis with expected improvements and further research directions.

Chapter 2

Literature Review

This chapter provides a basic overview of existing related research work. The reminder of this chapter is organized as follows: Section 2.1 describes the historical development of online social networks. Section 2.2 discusses existing approaches for marketing over these online social networks. Next, in Section 2.3, I discuss the characteristics of mobile social networks (MSNs) and mobile ad hoc networks (MANETs). A comprehensive analysis of existing community detection approaches will be presented in section 2.4. I discuss approaches for developing MSNs in Section 2.5 and marketing approaches in Ad hoc MSNs in Section 2.6. Finally, I provide a summary of this chapter in Section 2.7.

2.1 A Histroical Overview of Social Networks (SNs)

Before the emergence of the well-known social networking sites or services in the recent decades, the concept of “social networks” has been proposed and analyzed as early as 1930s. This concept was developed by researchers investigating the means by which people are connected in societies [19]. Initially, a social network was described in the shape of a social structure, which is made of individuals or organizations denoted as nodes, tied by one or more specific types of interdependencies (or social links) [20]. Social network analysis (SNA) was first conducted to investigate the relationships between feelings of well-being and the structure of people's social lives by some socio-metric analyst in USA in 1930s [20]. During the same period, other researchers [21] focused on examining the existence of fully connected graphs (cliques) in the structures of various social groups in order to identify cohesive subgroups within these social systems. Influenced by previous research work, some Harvard researchers led by Harrison White [22] further explored the mathematical framework for analyzing the structure of social networks in 1960s and 70's. In particular, they introduced algebra knowledge to design models of groups using set theory and multidimensional scaling to establish concepts such as density and distance of connections between people. This new analytical approach of describing the social networks provided a fundamental researching basis for further studies.

In general, researchers relied on the case of graphs to represent existing social networks where nodes represented the social entities interacting in the networks while links represented social interdependencies between these entities. Information transmitted between these nodes are represented by flows. The structure of these networks can be employed to categorize the social networks under examination along several dimensions [20]:

- Centralized, decentralized (that is, multi-centered), or distributed (center-less): this dimension is used to describe the structure of information flow over social network.

- Hierarchical or horizontal social networks: this dimension describes the connection shape of network. A social network structure can be either tree-shaped or flat.

- Finite: this dimension indicates whether or not the social network has fixed limits on the number of nodes and ties.

- Inclusive or exclusive: this dimension represents the social networks' ability to merge with or disconnect from other social networks.

- Intensive (that is, few nodes linked by a multiplicity of dense, strong ties) or expansive (many nodes enabling reciprocal, multidirectional flows); or non-interactive

(enabling only unidirectional flows)

SNA gradually moved from being a suggestive metaphor to an analytic approach to a paradigm, with its own theoretical statements, methods and researchers. The advent of Internet communication technologies have rekindled the interest in social networks analysis [23].

Over the past few decades, The Internet provided a new basis to increase the opportunities for people to communicate and interact beyond the physical location boundaries. Nowadays, people are increasingly available for online communication, often with others in ways they would never have encountered prior to the Internet's emergence. Connections are no longer bounded by propinquity; rather, people can seek out or "bump into" others from all over the globe [20]. Along with the surge in harnessing sociability among friends and strangers, some computer-based social network sites provided a platform for users to create their own social networks to communicate with their friends and to maintain their social relations.

Unlike the simple Internet-based interactions between users through a web or an online forum, which is largely organized around content, online social networks (OSNs) are organized around users. Normally, the participating users join a network, publish their profiles and some related content, and create links to any other users with associated properties. The resulting OSNs provide a basis for maintaining the created social relationships, for finding users with similar interests, and for locating content and knowledge that has been contributed or endorsed by other users [24].

In general, the main components of OSNs are discussed as follows:

- Nodes: to participate fully in an online social network site, nodes represented by users must register with their social profiles, which may include individual identifiers (such as account name, age, education background) and interests.

- Links: the social network is usually composed of nodes and links. Some sites (e.g., Flickr [25], Twitter [26]) allow users to be linked to any other user without the consent from the linking targets. Other sites (such as Facebook [27] and LinkedIN [28]) require consent from both the linking initiator and target; users form links for one of several reasons. The nodes connected by a link can be real-world acquaintances, online acquaintances, or business contacts; they can share an interest or common affiliation; or they can be interested in each other's contributed content [24]. Most of these links can be maintained unless users terminate the relationships. In particular, the links will only be removed by users' own wish. So the complete network topology of online social network is stable or mostly static.

- Communities: most of OSNs enable users to form and join a given community. Members within same community can freely transmit content or information to other members in that community. Like the links, community members are mostly static in OSNs. The structure of the community is stable whether users are online or offline. The admission and dismissal from a certain community are controlled by a user designated as the community's moderator.

2.2 Marketing over Online Social Networks

The majority of OSNs sites allow companies to collect information about the users and their social relationships, which make it very easy for marketers to find the adequate audiences for their marketing messages. For example Facebook allows users to determine how frequently they interact with friends or acquaintances online, what interests they have in common, etc. The large number of surveys indicated that Facebook is a big market for companies to target potential users [6]. Among the various marketing strategies over OSNs, advertisement messages are the most adopted strategy [29]. In [4], the authors discuss the use of OSNs in implementing viral marketing strategies. They identify a bound of strategies called influence-and-exploit strategies to target potential users from influential users' social networks. In [29], authors found that online brand advertising is a huge opportunity for companies to foster potential customers and conduct targeted marketing. Their experimental results proved that the impact of marketing over OSNs is even greater than what was expected.

2.3 Mobile Social Networks

Lately, mobile social networks (MSNs) have become popular [30]. An MSN is a social networking service where one or more individuals of similar interest or affiliation, connect and interact with one other using their mobile devices (e.g., mobile smart phone [31]). Similar to all OSNs, MSNs represent virtual communities. In generally, there are two types of MSNs: the first, referred to as web-based MSNs, represents an extended version of the Internet-based OSNs where the mobile users access the same content of these SNs through a mobile browser. The second type is the ad hoc mobile social networks (ad hoc MSNs) where users don't have access to the Internet and rely on each other to form their own OSNs [32].

In web-based MSNs, a central server is needed to exchange, share data between content providers and mobile users. All of provided services are designed to run on the central server providing simple application programming interface (API) [33] for mobile users to interact with the services from Internet. The central server collects users information such as social profiles and social activities. These information can be utilized to simplify various applications such as marketing or content sharing.

In contrast to the web-based MSNs, ad hoc MSNs are developed in a decentralized manner. Mobile users of these ad hoc MSNs can directly communicate and share content without having to communicate to a centralized server. Since a central server is no longer existing, the mobile users' profiles and activities are stored on their mobile devices. The limited connectivity of the mobile users in the ad hoc MSNs (without Internet access) results in the lack of a global topology knowledge of the whole network by the users. Moreover, unlike the information integration from multiple platforms of web-based MSNs such as OpenSocial API [34] (same users share information over different social network sites), there are seldom ad hoc MSNs providing integrated information aggregation services due to its dynamic and distributed feature.

To overcome some of the challenges facing the creation of ad hoc MSNs, some researchers proposed hybrid architectures of MSNs which combine of the characteristics of web-based MSNs and ad hoc MSNs. In this case, mobile users can access data from the content provider via a fixed central server and utilize the ad hoc MSN as a communication channel to share content with other mobile users. In these hybrid architectures, since network selection is open to mobile users (centralized or decentralized, web or ad hoc connectivity), how to use them in different scenarios is an important issue. A simple comparison between them is shown in Table 2.1.

2.3.1 An Overview of MANETs

Before discussing different types of ad hoc MSNs, I provide an overview of the characteristics of Mobile Ad Hoc Networks (MANETs) first, which is the physical environment for ad hoc MSNs and for the proposed framework. The word “ad hoc” means that the communication between users does not rely on a pre-existing infrastructure, such as routers in wired networks or access points in managed (infrastructure-bound) wireless networks [35]; instead, each user within the network participates in the routing process by forwarding data for other users, and so the determination of which users forward the data is made dynamically based on the network connectivity [36] [37]. MANETs are formed by a network of mobile and self-configuring nodes that act as routers for communication between other nodes. The mobile nodes in MANETS that are dynamically and arbitrarily located in such a manner that the interconnections between users changes on a continuous basis. In MANETs, two nodes can directly communicate (two way communication) if the distance that separates them is less than the minimum of their broadcast ranges [38]. The users of MANETs form a self-organized topology which is adapted to the movement of these users. These characteristics of MANETs make them rapidly deployable in any environments [38].

Table 2.1: A Comparison of Types of MSNs

Types of Architecture	Web-based	Ad Hoc	Hybrid API deployment
Connectivity	1. Internet Access 2. GPRS 3. HSPDA 4. WiFi 5. Bluetooth	1. WiFi 2. Bluetooth	1. Internet Access 2. GPRS 3. HSPDA 4. WiFi 5. Bluetooth
Third-party service	Required	Not required	Depend on the access network
Need for a Central Server	Yes	No	Required but not at all times
Knowledge about the Community	Global knowledge	Limited local knowledge	Hybrid depending on current access technology
An Example Middleware	RoadSpeak [39]	MobileClique [40]	JSA [41]

2.3.2 Characteristics of MANETs

Detailed characteristics of MANETs are summarized as follows:

- The connectivity of MANETs depends directly on the radius of transmission of the nodes in the network. Nodes in MANETs need to dynamically establish communications with neighbors through limited network connectivity management. Nodes in MANETs may be highly mobile and vary significantly in terms of the capacities of the used devices.

- The management of MANETs must be performed in a decentralized or distributed manner since there are no base stations (or a server) that can collect users' information. Information can only be stored on the mobile nodes. MANETs usually depend on anchor of nodes with high density to transmit information to other nodes.

2.3.3 Components of MSNs

Table 2.2 provides a comparison between these two types of MSNs:

Table 2.2: Comparison Between Different Types of MSNs

Types	Properties			
	Topology			Marketing
	Nodes	Social Links	Community	
Web-based MSNs	1. Maintain user profile and location information 2. Continuous connection to a server 3. Users are mobile but communicate through a backbone access (e.g., access point)	1. Inherit from other social networks 2. Usually static and long lived 3. Physical location information about other users is not always known 4. Physical location imposes no influence	1. Inherit from web-based social networks 2. Stable and usually grows slowly	1. Marketers have continuous access to all users' information through a centralized server 2. Long-term marketing plans addressing long term users' interests
Ad hoc MSNs	1. Consist of profile and location information 2. Limited connectivity with nearby nodes and no Internet Access 3. Mobile 4. Base Station not required	1. No pre-existing links to inherit 2. Dynamic and agile, real time established or break down 3. Limited physical proximity 4. Change of physical location may influence link existence	1. Real time detected community 2. May form or change by nodes' mobility 3. More dynamic than Web-based MSNs	1. No centralized access to users' information 2. Rely on other users to spread advertisements

- Nodes: similar to Internet-based online social networks, users (nodes) of MSNs must create their own profiles consisting of identifiers and social profiles. However, users in MSNs maintain knowledge about their physical locations. This location attribute provides dynamic nature of users when considering application development and usage exploration.
- Links: unlike web-based MSNs inheriting link distribution patterns from online social networks, the ad hoc MSNs consists of more dynamic and agile links. Once the geographic location changes, links may break or re-tie. Most of social links are formed under the dynamic conditions such as sharing similar short-term interests or affiliation with real-time geographic neighbors.
- Communities: due to the continuous Internet access and stable topology of web-based MSNs, communities are stable and are formed beyond physical location boundaries. However, the dynamic topology of ad hoc MSNs results in frequent changes in the communities. For example, in a certain time period, a set of nodes form a community according to their links topology. With the short time passing, in contrast to stable topology of web-based MSN, the ad hoc MSNs's topology changes significantly, which impacts its communities' composition.

In conclusion, ad hoc MSNs can be seen as MSNs implemented over MANETs.

2.4 Community Detection Approaches

2.4.1 Communities Defined as Social Graphs

A social networks community can be defined based on its contexts. In social graph perspective [42], natural human social networks have revealed a heterogeneous nature with respect to their links density. Inside those networks, there usually exists groups of vertices, called communities (or clusters) that are defined by high internal connectivity (high inter-link density) and much sparser links to the outside of the group [42].

For online social networks, Herbiet [43] indicated that: connection graphs of vertex communication networks like the Internet or wireless ad hoc networks, have either stable or dynamic structures. So the vertex degree and edge distribution of community graph shows local or global in-homogeneities. And the groups of vertices sharing high density of edges can be envisioned as communities because it is intuitive to regard them as playing similar roles within the graph.

Once we have a huge network, the next step is to divide the nodes within the networks into various communities. This process is referred to as community detection or community partitioning. In practice, community detection mechanisms are used to discover unknown clusters or groups of users sharing the defined interdependency (such as same social behavior or similar interest). The detected groups or communities can contribute to improving the performance of advertising over these networks [33]. In general, there are two types of community detection techniques [44] as shown in Table 2.3.

Table 2.3: A Comparison Between Two Types of Community Detection Methods

Types of Community Detection Methods	Properties		
	Edges Definition	Common Metrics	Social Metrics
Degree-based	The edges between vertices are represented as social behavior connections	1. Encountering Frequency 2. Mobility Pattern 3. Contact Frequency 4. Link Duration	1. Betweenness Centrality [45] 2. Closeness Centrality [45] 3. Degree Centrality [45] 4. Edges Density [45]
Context-based	The edges between vertices are represented as similarity of context-described vertices	Similar vertices described as sharing similar interest and affinity	1. Syntactic Similarity 2. Semantic Similarity

- Degree-based Community Detection: the connected edges between vertices can be represented as high frequency of vertices' social behavior. For example, the probability of node A and node B partitioned in same community of connection graph depends on the frequency of node A and node B encountering. Quercia et al. [46] introduced an algorithm combining the frequency of vertices encountering and the duration of vertices encountering to determine the density of edges between vertices. Since the degree-based approaches are more focusing on the social behavior of vertices in the whole network, the edges of the graph were formed according to the behavior of each vertex's interaction. The social metrics such as closeness or degree centrality are proposed as the measurement of community detection in this type.

- Context-based Community Detection: the connected edges between vertices can be represented as close social distance between these vertices. And this distance can also be defined based on various contexts. For example, Li and Khan [47] introduced ontology method to devise nodes' social profiles and map them to all vertices of network. Based on the similarity of vertices' social profile, the distance of vertices can be measured. Syntactic and semantic similarity measurements are proposed for designing the community detection methods in this type.

- In some special situations, researchers combine different social metrics to design a community detection approach. In [48], authors proposed a hybrid method to conduct a friendship modeling and community detection, which involves measurements of shared interest (context-based), social networks belonging, location factor and social behavior (degree-based).

2.4.2 Community Detection in Static Complex Networks

Due to the complexity of community detection in huge networks (which usually contain over thousands of vertices and edges in topology), centralized approaches have been usually presented. Centralized solutions aim to detect communities through a greedy algorithm, which requires the global knowledge of whole topology. Briefly, centralized algorithms iteratively operate until a measure reflecting the community assignment is optimized, which usually takes a lot of time and cost.

In terms of assessing detected communities, the modularity, Q, introduced by Newman and Girvan in [49] is one of the first metrics used to assess the quality of community detection. The measure for them to discover the communities in networks is to use a metric called "betweenness". Thus the general form of their community structure finding algorithm is listed as follows:

1. Calculate betweenness scores for all edges in the network.
2. Find the edge with the highest score and remove it from the network.
3. Recalculate betweenness for all remaining edges.
4. Repeat step 2 until all edges with the high betweenness score have been removed.

And then they design a modularity algorithm to test the detection performance [49]:

$$Q_g = \sum_r (e_{rr} - a_r^2)$$

where e_{rr} is the fraction of links that connect two vertices inside the community r, a_r is the fraction of links that have one or both vertices inside the community r, and the sum is taken over all communities r in the given network, with subscript g indicating its global nature. The higher value of modularity Q, the better community structure or community detection process was conducted. Based on this approach, an alternative algorithm was also proposed with the principal of extremal optimization by Duch [50]. To address the time-costing problem of centralized approach from Duch [50], Wang, et al. [51] utilized a cached table representing the network structure and significantly reduce the information processing time of each vertex. However, these centralized algorithms have some common disadvantages:

- High complexity for computation, which seriously limits its application for practical application for large size of network.
- Require global knowledge and information of the whole network, which prevents these techniques from being implemented in more decentralized and dynamic network environment.

So they are not suitable to be implemented in dynamic ad hoc environments.

2.4.3 Community Detection in MSNs

To address the unique characteristics of the dynamic ad hoc environment, some decentralized solutions have been proposed. Raghavan et al. [52] introduced the epidemic label propagation principle to allow vertices using current community identifier to interact with neighbor nodes and join the most appropriate community with their adjacent nodes. However, there are some problems in Raghavan's algorithm: label oscillation, formation of monster community and efficient information processing is non-guaranteed.

To address these issues, Herbiet and Bouvry [43] introduced a SHARC model utilizing epidemic propagation of a community label to allow each vertex to independently decide its community assignment and make the assignment sharper to prevent monster community. These asynchronous members from SHARC-based community address issue as label oscillation for Raghavan's algorithm.

2.5 Middleware Approaches for Ad Hoc Mobile Social Networks

In general the critical issue in the management of ad hoc MSNs is the development of a middleware that facilitates building social applications (e.g., content sharing and twitting services) on top of these communities by managing various functionalities related to community construction and membership management. I outline the major research efforts in this area as follows.

Recently, there has been a large number of research efforts that aimed at building efficient middleware for the management of MSNs. For example, MobiCLique [40] eliminates the need for a central server to conduct content exchanges between users by leveraging existing social networks to bootstrap the system. It also takes advantage of the constructed social network overlay to disseminate content. Veneta [8] is also a decentralized approach to explore the social neighborhood of a user by detecting friends of friends. PeopleTones [53] provides techniques for reducing sensor noise and power consumption for users in MSNs. MobilisGroups [54] allows for location-based group formation and links the physical and the virtual world by creating incentives to be at a certain place at a certain time. On the other hand, MobiSoC [55] is a centralized middleware that incorporates several algorithms to discover previously unknown emergent geo-social patterns to augment the social state of physical communities. Hui et al. [56] proposed the Bubble-Rap grouping model with the objective of employing social groups to forward contents over delay tolerant networks. In their model, users are classified into social groups based on their direct and indirect contacts. They first search for fully connected groups of k members and then merge these groups together. The scheme relies on the global knowledge of the users' contacts in order to execute a users' grouping algorithm. In this mode, users' interests are assumed to be fixed for the duration of interactions with other users. SocialCast [57] is another middleware developed by Costa et al. where users are assigned a utility value for different interests. These values are then employed to calculate the probability of the user joining a certain community. In [58], a semi-distributed social networking middleware service that dynamically combines both social and physical proximity relations between mobile users is introduced. The middleware allows a set of mobile users to act as brokers to other users. These brokers collect and exchange other users' interests as well as histories of encounters with other users in order to infer missing social links in the network. In contrast to the static nature of the relations in these social networks, users interactions and participation in ad hoc MSNs are highly dynamic and are usually focused on achieving a specific short-term tasks such as discussing a work related-project, latest news or sharing comments or updates on a specific topic (e.g., an marketing promotion or a goal picture in a stadium).

2.6 Marketing over Ad hoc MSNs

In ad hoc MSNs, due to the frequent changes in the network topology by users' mobility, it is NP-hard for an information propagator or marketers to deliver some content or advertisements to all desirable receivers in the network [59]. Several research efforts tried to examine how social behavior can be used to improve the performance of data access in various types of ad hoc networks. For example, Wang et al. [60] presented a community-based greedy algorithm for mining top-K influential nodes. They then targeted these influential users as good potential carriers to the proposed content. Intuitively, instead of finding influential nodes in all social networks, they divided the whole social networks into several complete communities and find influential nodes of each [42]. The aggregation of these influential nodes represented each community and formed a subset of influential nodes of social networks.

For efficiently optimizing the adaptability of content sharing, Qureshi and Min [61] has proposed an adaptive content sharing protocol to facilitate sharing content in p2p mobile social networks. This proposed protocol utilized content-based routing protocol [62] under delay tolerant environment to conduct content sharing. In terms of directing content transmission, the new protocol is more based on content rather than destination, which achieved object as deliver right content to right receivers. Furthermore, it improved the delivery rate of content sharing over information multi-hop forwarding scenario (which is the characteristics of ad hoc MSNs) and its simulation results provided a robust justification of scalability of content sharing over complicated ad hoc MSNs. Their work explored the possibility of optimization of content applications over MANETs, which can help marketers to efficiently spread promotion messages in ad hoc MSNs.

Usually the content sharing applications over MANETs are designed under PSN (pocket switched networks) [63]. To optimize its processing performance, Hui [56] proposed a new BUBBLE Rap framework, which enhances the efficiency and accuracy of content delivery in delay tolerant networks. The BUBBLE model was created under two main hypotheses:

- Community Affection: human society is structured in the shape of communities and people are more willing to exchange their information and opinion within their communities.
- Centrality: within communities, some people are more popular and interact with more people than others, which means they enjoyed great influence on people within same communities.

Based on two hypotheses, the content forwarding strategies of BUBBLE model can be seen as:

- The first part of the forwarding strategy is to forward messages to nodes which are more popular than the current node (with higher centrality rank).
- The second part of the forwarding strategy is to identify the members of destination communities and to use them as relays.

Finally, the authors proved that the Bubble model outperformed others in terms of the balance between delivery rate and delivery cost.

Another interesting work is that of Gao et al. [64] who studied several issues related to socially-aware multicasting techniques in delay tolerant networks, proposed a model that contributes the performance of information delivery in ad hoc MSNs. In the proposed work, I target general purpose ad hoc networks rather than delay tolerant ones, where content dissemination and interactions between users are more immediate in their nature and are not delay tolerant.

2.7 Summary

In this chapter, I presented some knowledge about social networks as well as various marketing approaches over these networks. I provided an overview of MSNs and discussed details pertaining to a special class of these networks, namely MANETs. Furthermore, I discussed some community detection approaches either in static networks or MSNs. Finally, related middleware for MSNs as well as examples of marketing schemes in ad hoc MSNs have been presented. Therein, to clearly distinguish these types of ad hoc MSNs, I referred to them as spontaneous social communities (SSCs).

Chapter 3

Proposed Middleware for SSCs

This chapter introduces the proposed middleware for the detection and management of SSCs and its main components, namely, user-profile manager, community manager and friendship manager. Section 3.1 gives an overview of proposed middleware SSCs; Section 3.2 discusses the functionalities of the user profile manager component while Section 3.3 describes the role of community manager. In Section 3.5, the advertising function over SSCs will be presented. Section 3.4 describes friendship manager of SSCs. Section 3.6 discusses one of the limitations for the proposed middleware and prospective method to address the issue. In Section 3.7, the main features of the proposed middleware have been outlined. Finally, I summarize this chapter in Section 3.8.

3.1 The Proposed Model of SSCs

The basic premise behind the proposed SSC model is that as users move from one place to the other, they can engage in forming different social communities to achieve a specific task or share some contents that are of common interest to members within these communities. To illustrate the idea behind this premise, consider a user in her work place who forms along with some of her colleagues an SSC to discuss particular issues related to a recent project. As the same user commutes back home, she participates in an SSC with other commuters to exchange latest news updates and opinions and share related media contents. As she leaves the bus, her membership to this SSC expires, but she may still cache some important contents she obtained from that community that might be shared later on with others. At home, she can participate in an ongoing SSC with some of her neighbors or family members to discuss matters related to her neighborhood. When the same user arrives at a shopping center later on that night, she might be interested in participating in a SSC that has some acquaintances she usually sees there and in another SSC to discover the latest marketing promotions or products in the shopping center that are interesting to other people sharing similar interests to hers. Clearly, as the user moves from one place to the other, she changes her focus and interest. Advertisers can take advantage of this fact to give the right promotions to her.

The focus in this section is to build an efficient model for SSCs that can capture the dynamic changes in each of the SSCs that our users participated in.

3.1.1 The Network and User Models

As depicted by the lower layer of Figure 3.1, the departing point of the proposed work is an ad hoc network comprised of mobile users equipped with communication devices with different capabilities (e.g., smart phones or laptops). I assume that these devices can communicate using wireless interfaces such as bluetooth or WiFi. In addition, I rely on the existence of reliable routing and transport protocols that can facilitate the communication between pairs in the network. In this setting, I say that devices of two users are physically connected if they are in the communication range of each other. This physical communication network can be modeled by a time-varying graph $G(t) = (U(t), E(t), \mathcal{P}(t))$, where $U(t)$ represents the set of users moving in a given geographical location, such as a university campus, a railway station or a shopping center at given time t. The set $E(t) \subset U(t) \times U(t)$ represents the physical communication links between pairs of users, where a link $e_{ij} \in E(t)$ indicates that users $u_i, u_j \in U(t)$ can directly communicate at t. The set $\mathcal{P}(t) = \{P_1(t), \cdots, P_{|U(t)|}(t)\}$ is the set of the users' profiles

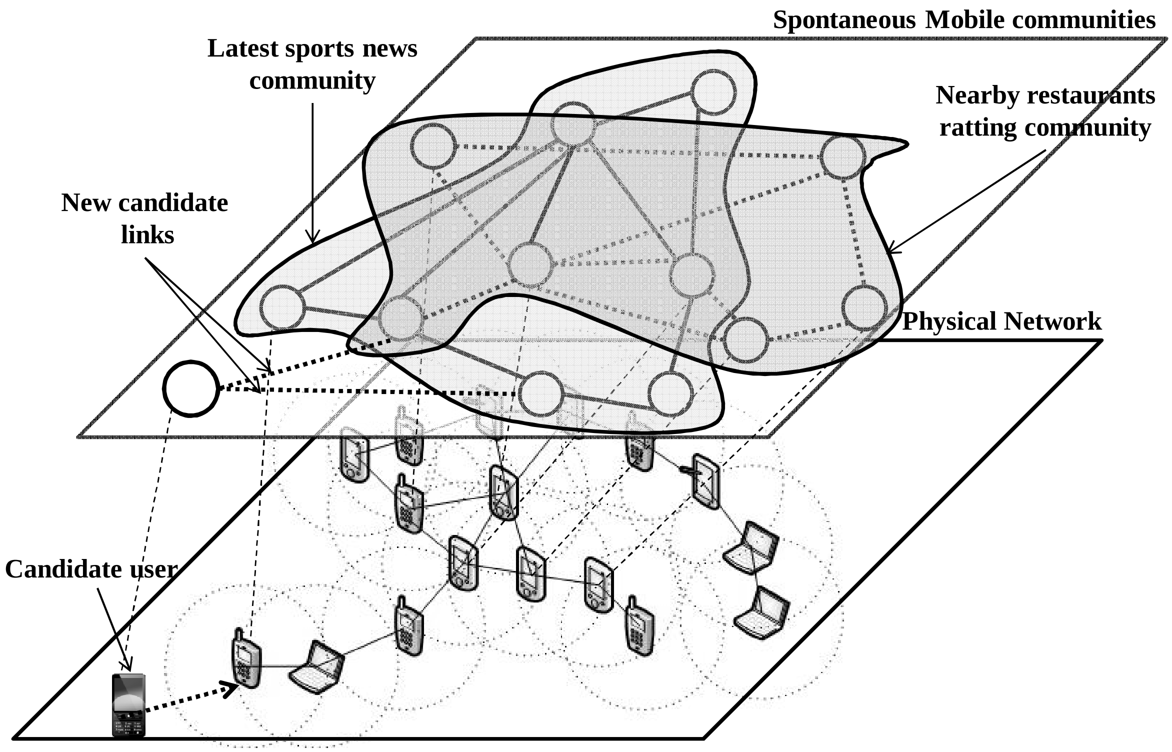

Figure 3.1: An Example of Two SSCs Built on Top of Ad Hoc Network

where a profile $P_i(t) \in \mathcal{P}(t)$ of user $u_i \in U(t)$ is stored in the user' device or in a proxy if the user has an access to a backbone network.

To this end, define $\mathcal{F} = \{f_1, \cdots, f_{|\mathcal{F}|}\}$ to be the set of user features taken into consideration in our user model. Each feature $f_m \in \mathcal{F}$ has a domain $D(f_m)$ that represents the finite possible values that f_m can take. For example, if f_m represents the user gender then $D(f_m)$={female, male}. Then a user profile, $P_i(t) = (p_{i1}(t), \cdots, p_{i|\mathcal{F}|}(t))$ is a vector of feature values that describe the user at time t, where $p_{im}(t) \in D(f_m)$ represents the value of feature f_m, $m = 1, \cdots, |\mathcal{F}|$, for user u_i. The user features may belong to one of three categories:

- The user's personal attributes and demographics such as his/her age, gender, nationality, address and education. The values of features in this category are relatively static and can be usually obtained directly from the user.

- The second category comprises features describing the current context of the user which can be collected through sensory devices [65]. This work focuses only on three attributes, namely, the current user's location, velocity and mode which can be obtained using techniques followed in [65].

- Finally, features in the third category refer to the users' interests or hobbies (e.g., news, cooking, sports and swimming). I adopt a hierarchical relation between these features, which is described in more details in the next chapter. The domain of each of these features is a normalized numeric value in the range $[0, 1]$, where a value $p_{im}(t) = 0$ ($p_{im}(t) = 1$) means that the user has no interest (respectively is highly interested) in f_m.

It is worth noting here that, user features in the first two categories, namely, the user demographics and context, can be seen as global features of the user that do not differ within the context of different SSCs. On the other hand, the user's interests are dependent on the community he/she is currently participating in. For example, a user can concurrently participate in two communities, one for discussing latest political news and the other discussing weekend promotions. In the first community, the user clearly has a high interest in news, which is reflected by the messages he is exchanging in that community. Similarly, for the second community, the user is mainly interested in specific promotions relevant, for example, to his/her interest in fashion or cooking. This distinction of specific users' interests within each community will be critical in the development of our SSC model in the next section.

Finally, to measure the similarities/dissimilarities between two users' profiles, I also associate with each feature f_m a function $s_m : D(f_m) \times D(f_m) \to [0, 1]$ that measures the distance, or normalized degree of similarity between two users with respect to a given feature f_m. Clearly, a larger value $s_m(p_{im}(t), p_{jm}(t))$ indicates that users u_i and u_j are almost similar with respect to feature f_m and a value $s_m(p_{im}(t), p_{jm}(t)) = 0$ means that the two users are totally dissimilar with respect to f_m.

In Chapter 4, more details on how to measure similarity s_m between these features and similarity between users' profiles will be provided.

3.1.2 Proposed Community Model

Given an ad hoc network $G(t)(V(t), E(t))$, at any time t, I can model communities of mobile users built over $G(t)$ by the set $\mathcal{C}(t)=\{C_1(t), \cdots, C_k(t), \cdots\}$. Each community $C_k(t)$ can be also represented as graph $C_k(t)=(U_k(t), E_k(t), P_k^{\text{comm}}(t), W_k^{\text{comm}}(t))$, where $U_k(t) \subseteq U(t)$ is the set of members in community C_k at time t and $E_k(t)$ is the set of friendship or acquaintance relationships among the members. A link $e_k(u_i, u_j)$ indicates that an interaction between two members $u_i, u_j \in U_k(t)$ relevant to $C_k(t)$ has taken place, and $W_k^{\text{comm}}(t)$ represents the importance of features of users profiles in certain community. Examples of interactions among members are exchanges of content, and sent and received requests to access each other's community walls where they post their comments.

Each community $C_k(t)$ is associated with a *community profile prototype* $P_k^{\text{comm}}(t)=\{p_{k1}^{\text{comm}}(t), \cdots, p_{k|\mathcal{F}|}^{\text{comm}}(t)\}$ that describes an estimation of the profile of a typical user in $C_k(t)$. One simple way to estimate $P_k^{\text{comm}}(t)$ for each community is to take the average of the feature values for all members in the community, i.e.,

$$p_{km}^{\text{comm}}(t) = \frac{\sum_{u_i \in U_k(t)} p_{im}(t)}{|U_k(t)|}, \qquad m = 1, \cdots, |\mathcal{F}| \tag{3.1}$$

When f_m is a feature with nominal values, then the mode can be used instead of the mean [66] to obtain $p_{km}^{\text{comm}}(t)$. Here, $p_{km}^{\text{comm}}(t)$ reflects the expected value of feature f_m for a typical member in community $C_k(t)$ at time t. In the following section, I will show how this prototype can be calculated in a distributed manner by members of each community $C_k(t)$.

Since our basic premise is that communities are usually focused on a limited number of interests and tasks, it is clear that in each community users will only share a high degree of similarity in a number of features, while the remaining features should be irrelevant to that community. For example, in a news focused community, one may expect a large variation in the participant's ages but high similarity with respect to the feature describing the users' interest in news. On the other hand, in a community discussing promotions for fashion for teenagers' clothing, then community members are expected to be mostly of similar ages, and so on. In order to identify how critical a feature is in each community I associate with each community $C_k(t)$ a vector of weights $W_k^{\text{comm}}(t)$=$\{w_{k1}^{\text{comm}}(t), \cdots, w_{k|\mathcal{F}|}^{\text{comm}}(t)\}$, that reflects the importance of each of the features (e.g., user demographics such as his age or gender, user context and interest features) in the community. A higher weight associated with a feature f_m should indicate that it is critical to identify community members. Clearly, $w_{km}^{\text{comm}}(t)$ should be inversely proportional to the degree of variability in the value of feature f_m in the profiles of the users in the community. To capture this property, I can employ the ratios between the inverse of the coefficient of variation for values of each feature in the community. More precisely, let $\sigma_{km}(t)$ be the coefficient of variation of feature f_m in profiles of users in community $C_k(t)$ [66], where

$$\sigma_{km}(t) = \frac{\sqrt{\frac{\sum_{u_i \in U_k(t)} d_m(p_{im}(t), p_{km}^{\text{comm}}(t))^2}{|U_k(t)-1|}}}{p_{km}^{\text{comm}}(t)} \tag{3.2}$$

d_m is the dissimilarity or distance between profiles or features. Further details regarding the calculation of d_m will be presented in Chapter 4. Then, the normalized weight for each community feature can be calculated as follows,

$$w_{km}^{\text{comm}}(t) = \frac{\frac{1}{\sigma_{km}(t)}}{\sum_{j=1}^{|\mathcal{F}|} \frac{1}{\sigma_{kj}(t)}} \tag{3.3}$$

Since user membership to different communities as well as community interest evolves over time, it is critical to maintain up-to-date values for the community prototype profile $P_k^{\text{comm}}(t)$ and the community feature weights $W_k^{\text{comm}}(t)$. In the next section, I will show how two vectors can be calculated in a distributed manner.

3.1.3 Inferring Community Membership

Now, if we are given the prototype community profile $P_k^{\text{comm}}(t)$ and the community feature weights $W_k^{\text{comm}}(t)$, then it is possible to determine whether a new user u_i, with profile $P_i(t)$ would be interested in joining the community C_k. A similarity score for community membership can be calculated as follows:

$$score(u_i, C_k(t)) = \sum_{m=1}^{|\mathcal{F}|} w_{km}^{\text{comm}}(t) s_m(p_{im}(t), p_{km}^{\text{comm}}(t)) \tag{3.4}$$

s_m is a function that measures the similarity between two profiles with respect to a feature and is further discussed in Chapter 4. If the user membership score, $score(u_i, C_k(t))$, is larger than a certain community membership threshold ζ_k, then u_i represents a potential candidate that is interested in topics discussed in C_k and, hence, an invite message should be sent to him from one of the community members.

Membership Types with A Community

Once the prospective users testing similarity value with community profile, the next thing for us is to setup the threshold to decide whether these prospective users will join communities. Since members in communities may have different levels of similarity value with community profile, from high to low, a classification approach can be designed to diversify the membership types. This diversity will contribute to enhancing the efficiency of the proposed marketing scheme, as will be shown. For instance, the users sharing extremely high similarity value with community profile will be the first candidates to substitute leading position of community once community leaders leave community. So the several types of member are listed as follows:

- Community Leaders $V_k^{leader}(t)$: These members are the most active users in the community. They generate and share new contents with all other community members. Also, they actively look for new members.

- Community Participants $V_k^{participant}(t)$: In contrast to community leaders, these members do not generate new results. However, they participate in discussions about the shared contents through rating and commenting before forwarding the content to other users.

- Passive Members $V_k^{passive}(t)$: Potential members who agreed to share their interest information among users in the community. They also can receive content shared by leaders.

- Potential Members $V_k^{potential}(t)$: These are new users that are interested in participating in spontaneous communities. They are detected either by community leaders or participants as users that have similar interest to the community.

These groups of users are disjoint such that $V_k^{leader}(t) \cap V_k^{participant}(t) \cap V_k^{passive}(t) \cap V_k^{potential}(t) = \Phi$.

Having outlined the main features of the proposed SSC model, in next part I develop a novel middleware that can be hosted on the users' devices to facilitate the management of communities in a distributed manner. The middleware also facilitates hosting different applications targeting advertisements distribution, as will be shown.

3.1.4 Proposed Middleware

The proposed middleware for distributed management of SSCs is depicted in Figure 3.2. The middleware is an application layer software component that is built on top of the transport protocol and resides on the user's device. As shown in the figure, the middleware serves various content sharing applications such as recommender systems for social events, tools for collaborative work as well as business-oriented applications such as SMS advertisement messages, podcasting, news feeds and location/context-aware services. The main functionality of the middleware is to aid the mobile user in detecting new SSCs that might be of interest to him/her. It can also identify potential new neighbors that might be interested in creating a short-term friendship with the user through an existing or a new community. Finally, it facilitates sharing contents and advertisements with other members in the users' communities.

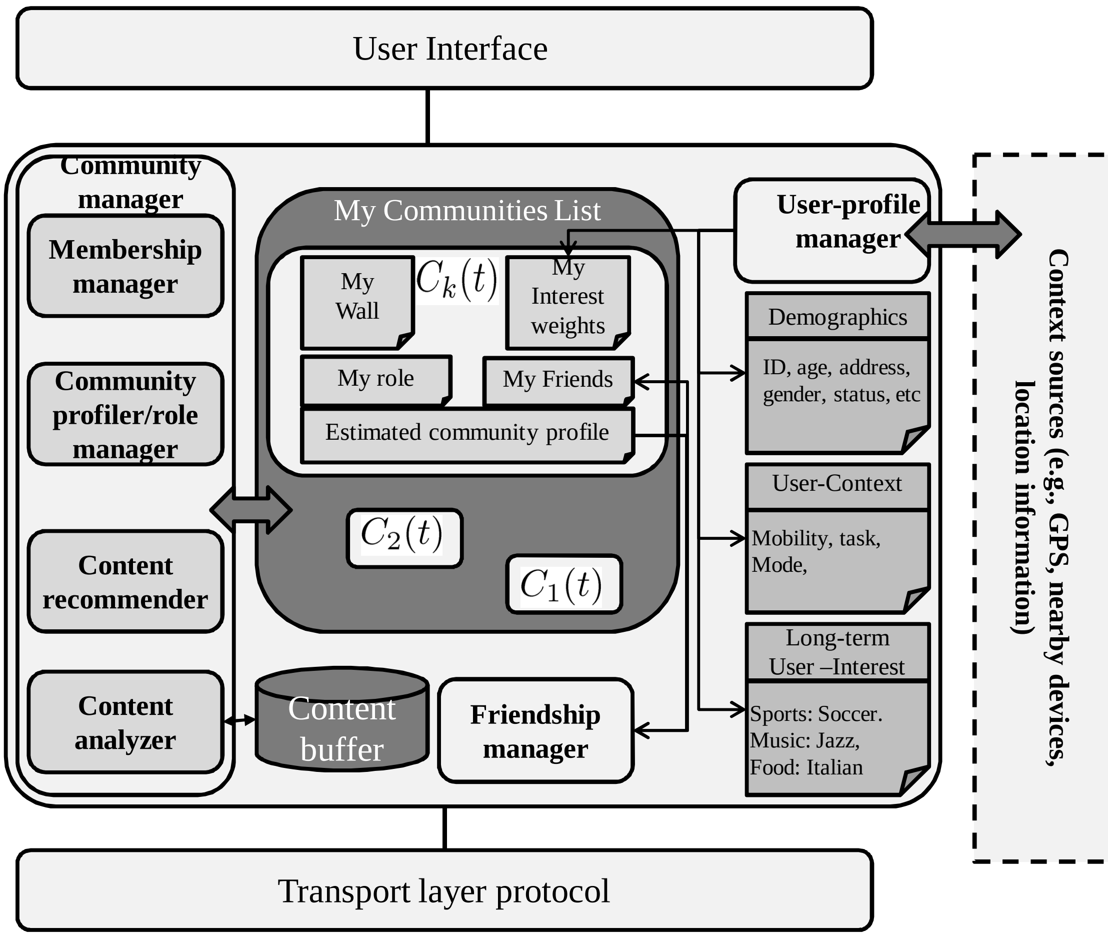

Figure 3.2: Proposed Architecture for distributed Ad hoc community Management

These functionalities are realized through the cooperation of three main components, namely, the user-profile manager (PM), the community manager (CM) and the friendship manager (FM). The former is responsible for the continuous adaptation of the user-profile based on the user's interactions in each of the communities he/she is participating in addition to the retrieved context from sensory devices [65]. The PM is also responsible for inferring and maintaining the relevant interests of the user with respect to each of these communities. On the other hand, the CM is responsible for managing the user membership and role in different communities.

It is also responsible for inferring the suitability of generated or received contents to friends in each of these communities. Finally, the friendship manager maintains the list of the user's friends in each community and continuously scans direct physical neighbors, either to find new communities that might be of interest to the user or to invite other neighbors to join one of its active communities. Finally, depending on the capabilities of the user's device, the user can store recently shared contents (e.g., images or clips and advertisements) to share latter with other communities.

The following sections provide a detailed description of the functionalities of each of these components.

3.2 User-Profile Manager(PM)

As indicated before, the PM is responsible for maintaining up-to-date information related to its user. Figure 3.2 shows that a user u_i's profile $P_i(t)$ is comprised of three components: the user's personal attributes and demographics, her context as well as her long-term interests that are updated over a long period of time. Here, I assume that the users agree on a pre-defined model for user profiling, for instance using ontologies [67], as will be described in details in the next chapter.

However, in order to reflect the short-term interests of the user, which is critical to identify his/her tendency to participate in different communities as well as the scalability of current advertisements, the PM maintains a separate profile $P_i^{(k)}(t)$ for the user in each community that stores short-term values for some of the user features in the interest category while obtaining values for the remaining static features (e.g., age and gender) from the original user profile $P_i(t)$. Hence, when a user joins a new community $C_k(t)$, a new profile $P_i^{(k)}(t)$ is created and the values for all interest features are copied to it. The PM then monitors the behavior of the user within each community that he/she participates in, and for every content consumed or generated within community, the PM updates the value of features related to this content in the user profile in the community. More precisely, assume that $\mathcal{F}_{\text{interests}} \subset \mathcal{F}$ is the set of user's interests (e.g., cooking, news, traveling and books). When the user receives or generates a specific content, and assuming, for simplicity, that the shared content is always associated with a single dominant feature $f_m \in \mathcal{F}_{\text{interests}}$, then the PM updates the user profile $P_i^{(k)}(t+1)$ in community C_k as follows,

$$p_{ij}^{(k)}(t+1) = \begin{cases} p_{ij}^{(k)}(t) + \Delta v & \text{if} \quad j = m \\ p_{ij}^{(k)}(t) - \frac{\Delta v}{|\mathcal{F}_{\text{interests}}|-1} & \text{if} \quad \forall j \neq m \end{cases} \tag{3.5}$$

where Δv represents the update step size for a feature change. This rule can also be extended to include multiple features that are associated with a single content by distributing $\triangle v$ over these features or by using more advanced update rules such as these using reinforcement learning. With this update rule, the shared content contribute in clearly identifying the members' most recent focus or interest. These updates can then be used by the users' community managers to calculate an estimate of the community profile and interest in a given advertisement message as will be shown in the next section.

3.3 Community Manager of SSCs(CM)

The main functionality of the community manager is to calculate an estimate of the profile prototype $P_k^{\text{comm}}(t)$, for each community C_k, the user is participating in. To solve this problem in a centralized manner, community members can send all their profiles to a single member of this community (usually a community leader) who will then use these profiles to calculate the prototype and feature weights in the community using (3.1) and (3.3), respectively. This member can then send back the results to all members. During this process, in addition to the message overhead and the single user bottleneck problem, one lost message will invalidate the obtained values. To address these issues, I adopt simple average consensus techniques, where each user updates its view or estimated values, by adding sums of local differences between its view and views of its neighbors [68]. In the following, I show how to calculate an estimate for one feature f_m in the profile prototype of community C_k, namely, $\hat{p}_{km}^{\text{comm}}(t)$ for $p_{km}^{\text{comm}}(t)$. The remaining feature values and weights can be obtained in a similar manner.

Each user maintains an estimate $\hat{p}_{km}^{\text{comm}(i)}(t)$, which is initialized with the current feature value from his/her profile $P_i(t)$. Then user, u_i, requests from each direct community neighbor $u_j \in N_i(C_k(t))$, $N_i(C_k(t)) = \{u_j : u_j \in U_k(t), e_k(u_i, u_j) \in E_k(t)\}$, his/her current estimated values $\hat{p}_{km}^{\text{comm}(j)}(t)$ and updates his/her view, using the rule:

$$\hat{p}_{km}^{\text{comm}(i)}(t+1) = \hat{p}_{km}^{\text{comm}(i)}(t) + \sum_{u_j \in N_i(C_k(t))} r_{ijk}(\hat{p}_{km}^{\text{comm}(i)}(t) - \hat{p}_{km}^{\text{comm}(j)}(t)) \quad (3.6)$$

where r_{ijk} is an optional weight for neighbor u_j that reflects the strength of the relationship to u_i in community $C_k(t)$. It has been shown in [68] that these estimated values always converge to the centrally calculated value if the graph is connected. The estimated community profile prototype is stored by the community manager for each community the user is participating in. When a new user comes into physical contact with u_i the friendship manager (FM) can then utilize this prototype to decide wether to invite the user to join the community or not as discussed in friendship manager section.

3.3.1 Community Initiation

When a user entering to the network area tries to infer community membership through similarity measurement with all the existing community profile, the measured results indicate that there are no similar existing community profiles. This new user may initiate a new community which contains only one member so far and he/she will automatically become community leader of this new community. Next, the user may invite new members if they share similar social profiles.

3.4 Friendship Manager(FM)

In proposed SSCs, content or information sharing reflects the users' interests in interacting with each other. With the dynamic evolution of these communities users may change their interest over time resulting in some friends on the user's list to become inactive members. To reflect the dynamics of community composition, I allow friendship and membership between users to expire if they do not interact or participate in community communication for a given period of time. Once a user loses all friendship links in a specific community, the community profile is totally erased from his community list.

The FM is also responsible for searching for new friends. When a new user physically approaches u_i, the FM sends a request to obtain his/her profile information and then utilizes the estimated community profiles to determine if there is a community that is suitable for the new user using the score function defined in (3.4). If a suitable community is found, the FM sends an invite message to the new user and adds him/her to the community friendship list once a response is received.

Finally, the FM is also responsible for receiving an invite message from other FMs. It can either announce the invite message to the user to make a decision or automatically accept or reject the request based on the current activity and availability of device resources.

3.5 Advertising over SSCs

In the first chapter of thesis, I stated that one of the contributions is to help marketers to easily find targeted audiences from SSCs. So far I have constructed several communities composed of members sharing similar social profiles. The next step is to find appropriate communities to spread advertisements or promotion messages. Since each community is already represented by their community profile prototypes, I can measure the similarity or distance between the community profile and promotion messages or advertisements to find targeted audiences. For instance, given an advertisement $A = \{a_1, a_2, ...\}$ with a set of feature $\{a_1, a_2, ...\}$ as a_m and community profile prototype $P_k^{\text{comm}}(t)$, I can determine whether this advertisement will be appreciated by a community C_k. The decision score for spreading information in a given community can be calculated as follows:

$$score(A, C_k(t)) = \sum_{a_m \in A} s_m(a_m, p_{km}^{\text{comm}}(t)) \tag{3.7}$$

s_m is defined in chapter 4 as a feature similarity measure. If the decision score, $score(A, C_k(t))$, is larger than a certain community targeted advertising threshold ζ_k, then A represents a proper advertisement which may find potential users or customers. The marketers then forwards this message to as many community leaders as possible. Those leaders in turn efficiently spread this message within communities. Furthermore, this decision score algorithm can also be utilized between each user and proposed advertisements referred to as the average received advertising score in performance evaluation.

3.6 Privacy Concerns

During the construction of communities in SSCs, privacy represents a major concern for the system implementation. Friendships between users and the foundation of communities are based on the results of similarity measurement between users. However during the measurement of this function, users need to provide their personal information to other users.

Community information may be exposed to the public due to the malicious interception. To address the impact of this private information breach (such as age, gender, interest), I can employ some methods to encrypt the message and limit the scope of users who can read this sensitive information, for example, I can employ the public key infrastructure [69], as follows:

- For the announcement containing community profile information, all community members use same community public key to encrypt the message.

- Once a potential member enters a certain area usually campus, he or she will be given the private key to decrypt the message. The receivers have no right to read the plain-text unless they hold the membership of community related to this message. All the similarity measurement is conducted automatically without manual intervention.

- Once the potential member becomes a member in certain community, he will receive the exclusive key to encrypt or decrypt the message spread within community.

- Each community has its own exclusive key, all information spread within community is encrypted and decrypted by certain exclusive key.

In conclusion, the approach to address privacy concern of proposed framework consists of a cryptography method [70] for encrypting message or or advertisements distributed over SSCs:

- Public key and private key methods are adopted to encrypt and decrypt inter-communities message flood.

- Exclusive key are adopted to encrypt and decrypt intra-community message flood.

3.7 Advantages of the Proposed Work

In this section, I outline the main features of the proposed work.

Efficiency in Identifying New Members: since community interest in the proposed middleware summarized by a community profile prototype, it is more efficient to identify the new members of community. Instead of calculating similarity with all members of communities, the proposed middleware helps potential users only measuring profile similarity with community profile. Meanwhile, the distributed manner of inferring community profile prototype addresses centralized calculation issue such as message overhead, lost message causing invalid results.

Simplifying the Targeted Marketing: the proposed middleware simplifies the targeted marketing application with information transmission efficiency: instead of spreading information all over the network or randomly selecting users as starting points of information transmission, the information propagators or marketers can easily transmit information to the targeted community leaders through measuring decision score with community profile. And community leaders will act as sprinklers to spread information or advertisements to members [71].

Applicability to Delay Tolerant Networks: as discussed in the literature review, delay tolerant networks represent a popular environment to host mobile social networks [56] as well as for content sharing. The challenge behind its performance is how to find anchor nodes that will forward content to the right recipients with minimal amount of delay. The proposed community memberships simplify this process of finding anchor node: instead of randomly choosing anchor nodes to forward content, community leaders may represent ideal candidates for anchor nodes since each of them may share certain similarity with their communities' members.

3.8 Summary

In this chapter, I have described the proposed SSCs model and presented a novel middleware that can be hosted on users' devices to facilitate the management of social communities in a distributed manner. I provided details about on how to model users and communities in SSCs and described the main components of proposed middleware. In community manager, I presented a novel distributed manner of community profile prototype to address the centralized problem. To capture the dynamic nature of user interests, I adopted short-term user interest update performed by the user profile manager. The targeted advertisements transmission was also realized in the proposed model and friendship manager keeps dynamic memberships of constructed communities. Finally, the privacy issue have been addressed.

Chapter 4

Users Profiles and Similarity Measurement

In this chapter, I describe the adopted similarity measurement functions and extend the presented user profile defined in Chapter 3 to the use of ontologies. The reminder of this chapter is organized as follows: Section 4.1 gives an introduction to ontologies; Section 4.2 and Section 4.3 provide some examples of ontologies applied to social networks and mobile social networks. In Section 4.4, I describe how to apply ontologies in our model as a similarity measurement function of user profiles. Section 4.5 and Section 4.6 give details about the adopted similarity function for each feature of the user profiles. Finally, I summarize this chapter in Section 4.7.

4.1 An Overview about Ontologies

As mentioned in the literature review chapter, formation of social links in social networks is similar to community detection: both aimed at finding hidden links between nodes (vertices) of social networks (social graph) and the aggregation of social links results in communities. In our middleware SSCs, the common users preferences are adopted to model the friendships between users [72]: in social communities, it is commonly accepted that people who are known to share a specific interest are likely to have additional connected interest. To model the context as "users sharing similar interests", I found that researchers first proposed simple keyword matching approaches [73]. Specifically, if there are matching keywords in users' profiles, the users can connect with each other and form friendships. But this simple approach introduced another problem: the lack of unified understanding of the meaning of a word, the result is that keyword matching may be wrong. For example, user A and B sharing same interest "Soccer" in social networks, but in their profiles, A likes soccer and B prefers football. In some scenarios, the context meaning of football and soccer did not represent same sport activity. Even though A and B share same interest, due to the misunderstanding of the words-context, the keyword matching method can't connect them in the social network. To address this problem, the use of semantic ontologies has been proposed [74]. I utilize ontologies as means for knowledge representation to help us model the user connection process in SSCs. With the semantic knowledge representation by ontologies, the profile distance can be measured in mathematical format and form similarity scores between users. In terms of user profile, our approach is designed over an ontological representation of the domain of discourse where users' profiles are defined. The ontological space for proposed middleware takes the benefits of a semantic hierarchical network of interrelated feature domain. Interests of users in SSCs are defined by their profiles which are composed of extracted features from an ontology vocabulary. So the similarity between users profiles can be mapped as predicted social connections (friendships) in SSCs. In summary, the ontology is intended to visualize the profile and feature similarity measurement to achieve friendship.

Metadata, the data about data, is always the crucial element to find the relationship between data. Effective metadata requires the shared representations of knowledge understanding to establish the basic vocabulary to support metadata assertion. In this case, ontologies, defined as "a shared common understanding of a domain" [75] is precisely intended to achieve this requirement. In terms of construction of machine-readable knowledge representation, ontology aims to provide a shared understanding of a domain both for the computers and humans. Thereby, an ontology describes a domain of features in such a formal way that it can be processed by computers. The outcome is that the computer system knows about this domain. An ontology is also a formal classification schema, which has a hierarchical order to find the relationship between some domains. Furthermore, an ontology comprises the logical components of a "Knowledge Base" [76]. Typically, a knowledge base consists of an ontology, some data and also an inference mechanism. Ontology, comprising the logical components of the knowledge base, defines rules that formally describe how the field of interest looks like. The data can be any data related to this field of interest that is extracted from various resources such as databases, document collections, Web etc. Moreover, the ontology-based representation is more expressive and less ambiguous than a keyword-based or item-based approaches [77]. To deal with the subtleties of users preferences of their profiles, the hierarchical ontology-based method provides an adequate grounding for the representation of relationships between the concepts from specific domain, which is used to define the features in profiles. More importantly, to implement the model in computer environment, due to the characteristics of ontology-based representation: more adaptive for providing further formal, computer-processable meaning on the concepts, ontology-based method is a better choice. In addition, ontology standards, such as RDF [78] and OWL [79], support inference mechanisms that can be used to enhance personalization of the concept relation. For example, a user interested in animals (parent-class of cat) is also recommended items about cats. Inversely, a user interested in lizards and snakes can be inferred to be interested in reptiles [80]. Also, a user interested in China can be assumed to like Shanghai or Beijing, through key-city transitive relation.

4.2 Ontologies Applied to Social Networks

With the emergence of semantic ontologies, in which information is given a well-defined meaning [81], knowledge representation in computer environment is more expressive. As such, semantic ontologies should be a place, where information can be better discovered, can be automatically processed, can be integrated and shared across various applications. Web documents from semantic ontologies can be furnished with information, whose context meaning can be interpreted by software programs and applications [76]. The semantic ontologies enable information to be better understood and processed by computers. Furthermore, social networks applications can be easily developed in computer environment by the semantic ontologies since semantic ontologies bring the applicable environment of unified words-context understanding. Thereby, the extension of mobile social networks such as SSCs also shares the advantages of emergence of semantic ontologies [76]:

- Ontologies is commonly deployed for the specification and explication of concepts and their relationships to a given domain. Social networks have the same purpose but with the focus on social relations and entities, hence domain ontologies related to social entities and relations can be designed and deployed in any social networks environment.

- Through reasoning and inference functions, ontologies do not allow the modeling of contradictory or inconsistent information. Modeling social networks via an ontology ensures the validity and consistency of the encoded information.

- An ontology, together with the inference mechanism, enables information gained through deploying rules to infer new information. Inference mechanisms can be employed over ontology-based social networks to come up with new relations and concepts out of the already existing ones between the social entities.

For the example of ontologies applied to social networks, Peter Mika [82] first presented the Flink system for the extraction, aggregation and visualization of online social networks. Flink employed semantic technologies such as RDF standards of ontologies for reasoning with personal information extracted from a number of electronic information sources including web pages, emails, publication archives and FOAF profiles. The acquired knowledge is used for the purposes of social network analysis and for generating a web-based presentation of the community. The Flink system proposed by Peter Mika proved the semantic technology – ontology as a proper method to conduct knowledge expression in online social networks. Based on the outcome of Flink system, Peter Mika introduced an outperformed model comparing to previous one in [83]: they extended the traditional bipartite model of ontology with the social dimension, leading to tripartite model of actors, concept and instances. They also adopted the folksonomy (from folk and taxonomy) idea to allow a neologism for a practice of collaborative categorization using freely chosen keywords Folksonomies. Finally they demonstrated the applicability of this tripartite model by looking at two different social networks: bookmark **del.icio.us** and synthetic data obtained by using **web mining** techniques. In the results, one of three bipartite models split from tripartite actor-concept network better reflects the conceptualizations of those involved in semantic web research, and this holds especially for those most actively involved in semantic web research. So when introducing ontologies to social networks, it is appropriate to focus on actor-concept bipartite graph when considering user model.

Furthermore, Jung and Euzenat [3] proposed a three layer model which involved networks between people (social network) in Figure 4.1, the network between the ontologies they used (ontology network) and a network between concepts occurring in these ontologies. They described how relationships in one network can be extracted from relationships in another one based on analysis techniques relying on this network's specificity. Similarity in the ontology network can be extracted from a similarity measured on the concept network. They introduced this multi-layered model to facilitate the usage of the structures of knowledge used by people to extract meaningful relationships at the social level for people. Moreover, the extraction of these new relations is used to further improve the collaborative sharing and exploitation of this knowledge. Based on their research, P2P network or P2P systems are adopted to implement this new model to construct social networks. The experimental results of their proposed model also proved that the social profiles composed by similar concepts between users result in prospective friendships. And this concept similarity can be conducted and measured by ontology method.

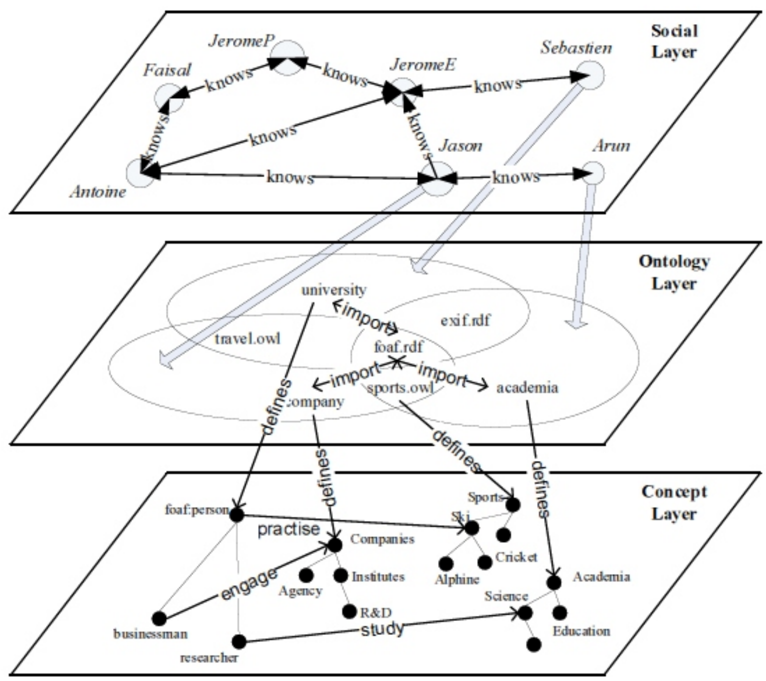

Figure 4.1: Three Layered Social Semantic Networks [3]

Moreover, Aljandal et al. [84] introduced the idea of combination of ontology and association rules to explore the friendship prediction in social networks research. In particular, the nominal and numerical features brought by ontology and AR (association rules) mining of their proposed approaches defined the distance of the sets of users from web-based social networks. Hence, the friendship prediction based on shared interest can be achieved. Finally, a test bed previously developed by using the social network and web-blogging service LiveJounal is extended to measure the approach performance. Evaluation results showed this semantically integrative approach yields a boost in precision, recalls of known friendships and proves the effectiveness of ontology method towards application of knowledge representation in social networks environment by unambiguously representing the users and their contexts.

For further research of ontologies, Hamasaki et al. [85] adopted the main idea from Peter Mika tripartite model [83] and integrated it with another relationship dimension such as friend relations in social networks service and known relations in FOAF (friends of a friend). With an ontology described as a knowledge base shared within community communication by [85], the source of information shared within a community can be used to improve knowledge consistency. Based on that, they provided a so-called "p-concept" which is conceived by individuals before they are shared. This new technique referred to as HAMA model is used to address the polysemy or hononymy problems occurred in inter-communities communication. With experiments of proposed HAMA model, the acquired results shown that HAMA model performed better in terms of information recommendation application both in two datasets, Polyphonet and Blue Dot.

4.3 Ontologies Applied to Mobile Social Networks

In [86], Bottazzi et al. proposed a framework, SAMOA, to develop ubiquitous social networks in context-awareness environment. To construct social networks that reflect the reality of social interactions in ubiquitous environments, they considered the context information, such as user location and reciprocal proximity, user attributes, motivations, attitudes, activities, and social preferences to be accounted for since the expressive meaning of context describing users in social networks is very important for constructing the users social connections. However, the dynamic deployment scenario of ubiquitous social networks complicates context-modeling endeavors. So they adopted emerging ontology standards – Web Ontology Language (OWL), which allowed interoperability between possibly unknown users who might wish to establish a social interaction. In addition, the adaptive matching algorithm for measuring the similarity between users was designed over ontology standards, which consisted of both the profile context information and location context information.

As mentioned before, I found that semantic ontologies are ideal knowledge representation approaches for social networks both in static topology environments [83] [84] [85] and dynamic environments [3] [86]. The major difference between ad hoc MSNs such as SSCs (spontaneous social communities) and Internet-based online social networks is due to the mobility factors, and the topology of SSCs being dynamic and unstable. But this dynamicity won't affect the user information representation when constructing user's profile in SSCs. The only problem for user's profile presentation in SSCs is that personal information is usually stored in user's own device. The different understanding of same thing may lead to polysemy or hononymy problem when two users want to communicate. To address this problem, the unified knowledge vocabulary is necessary to support the consistency of knowledge understanding in SSCs. So I will adopt FOAF [87] as an ontology vocabulary to provide concepts describing user information in the proposed SSCs.

4.3.1 FOAF project

On the official sites of FOAF project, the authors and contributors of FOAF project explain the vocabulary specification of FOAF [88]: *FOAF is a project devoted to linking people and information using the Web. Regardless of whether information is in people's heads, in physical or digital documents, or in the form of factual data, it can be linked. FOAF integrates three kinds of network: social networks of human collaboration, friendship and association; representational networks that describe a simplified view of a cartoon universe in factual terms, and information networks that use Web-based linking to share independently published descriptions of this inter-connected world. FOAF does not compete with socially-oriented Web sites; rather it provides an approach in which different sites can tell different parts of the larger story, and by which users can retain some control over their information in a non-proprietary format.*

So FOAF vocabulary provided us an open and adjustable languages to interpret concepts and knowledge from the physical world to machine-readable style and visualize these concepts and their knowledge relations in computer and web environment. To achieve the demand of describing user information of proposed SSCs, I adopt some concepts from FOAF to construct user profile: ID, name, age, gender, interest, role and current mood. Also since SSCs provide a dynamic environment for users interaction, I also introduce some dynamic concepts to represent some attributes of users: mobility and proximity. All of these concepts will be utilized as features for describing users profiles in SSCs.

4.4 Proposed Similarity Functions

Based on the user profile manager, I map the adopted FOAF concepts as features of users profiles in different categories as shown in Figure 4.2:

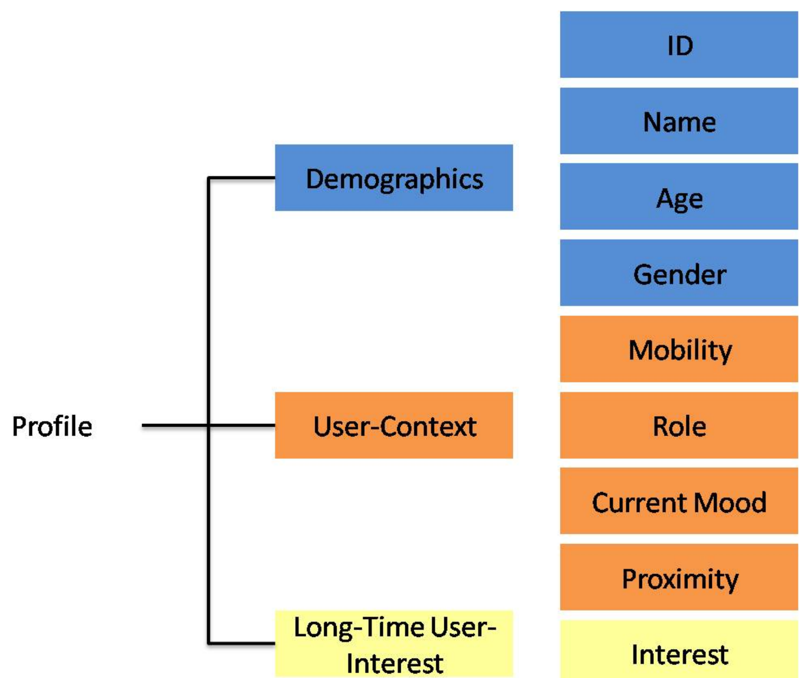

Figure 4.2: Adopted User Profile Feature Classification

Since users in our model of SSCs were represented by several features, the friendships between users come down to the similarity between users' profiles, and measuring two users profiles' similarity comes down to comparing a set of their features. To obtain appropriate results, similarity measurements must be associated with each feature. For instance, comparing the feature "gender" is much easier than comparing feature "interest" since "gender" normally consists of two domain values as "male, female" while "interest" may have hundreds of possible values. To propose proper measurement approaches for each feature, I first consider the meaning and domain values for each feature.

- ID: this feature is used to represent the users' identity in SSCs. So each user may have different id number, and there is no need to compare the id's similarity.

- Name: this feature is used to describe user's name or title in SSCs. In contrast to "id" feature, some users may share same name in SSCs. There is no need to compare its similarity.

- Age: this feature is used to represent users "age" information. The difference of "age" feature may affect whether friendship will be established between users. So to measure user's distance and similarity, the "age" feature must be taken into consideration.

- Gender: this feature is used to define the user's character being male or female. Due to the distinctive characteristics of different sex of users, the "gender" feature must act as one of the factors when considering the similarity between users.

- Interest: this feature is used to define users' interest or hobbies. Due to the diversity of the scope of interest (interest area may be sports, media or etc.), the "interest" feature may be taken as an important factor when considering users' similarity.

- Role: this feature is used to define the "role" of users in certain testing area. The difference of role may lead to different behaviors of users, which may affect users' social connections in SSCs.

- Current Mood: this feature is used to represent the current user's mood. The difference of current mood may be one of the factors to affect the establishment of friendships between users. For instance, even two users share high similarity in other features of profiles, if one user is in the lowest mood, it is hard to persuade them to form friendship or actively interact with each other.

- Mobility: this feature is used to describe the user's velocity in testing area. The difference of velocity may affect the friendship formation. For instance, it is hard for users in high mobility to form friendships with users in low mobility since the difference of mobility make it hard for them to maintain stable friendships.

- Proximity: this feature is only used to setup the users's communication range, only the users in each others' proximity, they can detect others' information.

Based on the description of the above features, I analyze the related features' domain to find appropriate methods of similarity measurements for each of them:

- Age: instead of highlighting the difference between exact age value (such as difference between age 12 and 23), I take the age difference in range. For instance, I measure the age difference as between age range from 10 to 20 and age range from 21 to 50.

- Gender: the possible values for feature "gender" is either male or female.

- Interest: due to the large scope of interest domain, the possible values for this feature should be enormous. And some of values may have relationships with each other. For instance, value "soccer" might be the subset of value "ball games" in sports domain.

- Role: the possible values for this feature might be various when users are in different areas. But the value range for feature "role" is limited in specific area. In a word, the scope of testing area limits the domain of feature "role". For example, the common roles for a place such as campus consist of possible values as “students, professors, administrative employees and guests”.

- Current Mood: the mood is hard to describe. Based on different understanding of context, people might use different words to define. To simplify and unify the context meaning of “mood”, I may choose limited domain for feature “ current mood”.
- Mobility: like the feature “current mood”, I also limit the domain value for the users’ velocity in testing area.

Based on the analysis of domain values for each feature above, I find some of them have limited domains. In this case, I might choose a syntactic measurement approach to define their similarity values since possible values for these features might be simple. But for the feature “interest”, due to its possible inter-relations between all the possible values, I might consider the knowledge of relations between possible values of “interest” when I am measuring similarity of feature extracted from “interest” domain. According to these two scenarios, I classify the measurement approaches into two categories:

- syntactic similarity [89]: it will be used for features other than “interest”.
- semantic similarity [89]: it will be only used for features related to “interest”.

4.5 Syntactic Similarity

In [89], the authors defined syntactic similarity measurement approaches as “the approaches which provide exact or approximate lexicographical matching of two values”. In our understanding, the selection of possible values for these features is simple and limited. The results of similarity measurement are usually taken as “Similar” and “Not Similar”. To specify the difference of the measured results, I may utilize quantified value to represent the measured results to reflect the level of syntactic similarity.

Table 4.1: Similarity Measurement for Feature "Age"

Feature Domain (Similarity)			
Age	Teenage and Youngster	Middle-Aged	Elderly
Teenage and Youngster (10-20)	High	Medium	Low
Middle-Aged (21-60)	Medium	High	Medium
Elderly (61-90)	Low	Medium	High

4.5.1 Age

As seen on the above, age is used to reflect the users natural information as in what stages their lives should be. It is intuitive to take the "age" value in enumerated style. The difference of age might not be mathematical difference between their "age" value. From the "Generation Gap" theory [90], the difference of generation of people reflects their different ways of thinking and behavior towards forming social relations. So in this feature, I compare the age difference in different generations. To simplify the measurement, I select three ranges to reflect users' age domain:

- Teenage and Youngster: this range represents the generation who are in early stage of their lives. They might be more similar with middle-aged group than elder group.
- Middle-Aged: it is used to reflect people in the middle stage of their lives. They should be equally similar with younger group and elder one.
- Elderly: this range represents elder people, where it is assumed that they should be more similar with middle-aged people than youngsters.

So I arrange the sample of similarity between possible values for "Age" feature to be shown in Table 4.1.

Table 4.2: Similarity Measurement for Feature "Gender"

Feature Domain (Similarity)		
Gender	Male	Female
Male	High	Medium
Female	Medium	High

4.5.2 Gender

Gender is the feature to define user's sex. The possible values for this feature is male or female. Unlike the "age" feature, the difference of gender may not reflect the fact as less similarity between these users. Since the object of similarity between feature is to find the friendships or social connections between users in SSCs, precisely, the results of similarity measurement for each kind of feature must reflect their impacts on friendship formation. I adopt the "Sex Difference in Social Behavior" theory [91] to contribute to the measurement for "gender" feature: people in same gender may share a bit higher possibility to become long-term friends than people in different gender.

But there also exists other social theories to prove that people in different gender are more easily to become friends, so the impact of this theory should be limited. Like the age feature, the measured results are shown in Table 4.2.

4.5.3 Role

In a certain area, people may hold different roles. For instance, on campus, people's roles might be classified into student, professor, administrative employee and guest. Each of these roles may hold different behavior or activity. So it is reasonable to think that people in same role are more likely to talk or interact with each other. Then they are more likely to form social connections or friendships. In our proposed middleware, I list three different roles in testing area: Role A, Role B and Role C. Their similarity values

Table 4.3: Similarity Measurement for Feature “Role”

Feature Domain (Similarity)			
Role	Role A	Role B	Role C
Role A	High	Medium	Low
Role B	Medium	High	Low
Role C	Medium	Low	High

Table 4.4: Similarity Measurement for Feature “Current Mood”

Feature Domain (Similarity)			
Current Mood	High	Medium	Low
High	High	Above Medium	Medium
Medium	Above Medium	Medium	Low
Low	Medium	Low	Extreme Low

are also shown in Table 4.3.

4.5.4 Current Mood

The “current mood” feature is used to define the users’ current psychological condition. According to [92], psychological condition of people may affect their cognitive social behavior, especially in friendship formation. So in our case, I classify the current mood of users into three levels: High, Medium and Low. It is reasonable to assume that people in high mood are more likely to make friends than others and vice versa. So the similarity results of this feature are shown in Table 4.4.

Table 4.5: Similarity Measurement for Feature "Mobility"

Feature Domain (Similarity)			
Mobility	High	Medium	Low
High	High	Medium	Low
Medium	Medium	High	Medium
Low	Low	Medium	High

4.5.5 Mobility

Mobility is used to define the velocity of people in a given area. One of the key factors of SSCs is that users have different mobility levels, which leads to a more dynamic users topology. It is reasonable to assume that people with same levels of mobility have more chances to become friends since same mobility might keep them in close area, which may give them more opportunities to interact with each other. In the proposed framework, I classify the possible value of mobility into three simple groups: High, Medium and Low. The measured results of similarity value for mobility is also given in the Table 4.5.

4.6 Semantic Similarity

In our similarity measurement for "interest" features, I comprehensively consider the hierarchical relations and semantic distance relations between these domain values. The degree of matching between concepts extracted from domain values can be determined by the semantic similarity. The whole strategy of measurement takes two concepts as input and compute a semantic similarity as output in four steps:

Step 1: Relation Allocation. Determine the relations between concepts through drawing a whole knowledge base.

Step 2: Distance Table Generation. Record all concepts distance according to knowledge base.

Step 3: Semantic Distance Computation. Compute semantic distance according to the distance table of the knowledge base.

Step 4: Semantic Similarity Computation. The construction of the similarity function and computation of semantic similarity are based on semantic distance computation. The results of the similarity is on mathematical types.

4.6.1 Relation Allocation

In the process of similarity measurement between concepts from the knowledge base as ontology, I allocate the concept relation as depth value to describe the hierarchical distance between the concepts in ontology. First I take the hierarchical ontological tree as ontology structure to represent the semantic distance between these concepts, like the tree shape, this ontology structure composed of various entities: root concept, leaf concept, parent concept and children concept. To simplify the process of similarity measurement, I assume the depth weight between each attached level is equal (which can be modified accordingly in further research). Also this tree-shaped ontology consists of three desirable characteristics:

- The semantic differences (or distance) between upper level concepts are higher than those between lower level concepts, in other words, general concepts are less similar than two specialized ones.

- The distance between sibling concepts is greater than the distance between parent and children concepts. More specifically, the depth of root concept is assumed as 1 (which is convenient for calculation), the depth of other concepts is their paths length to root concept (in our case the distance of each path between two attached concepts is assumed as 1, but it is adjustable according to the demand).

- The longer the semantic distance between two concepts is, the less similarity they share. That is, semantic similarity and semantic distance have an inverse relation [93].

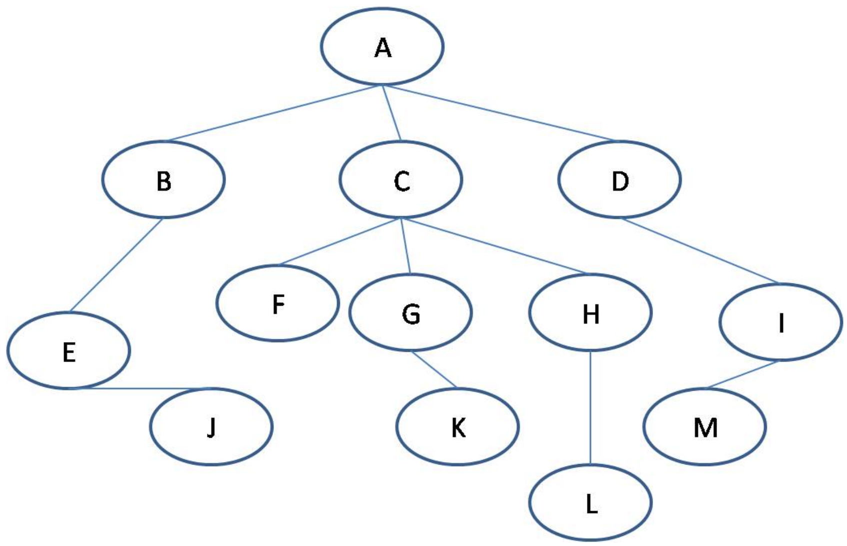

Figure 4.3: Example of Hierarchical Concept Tree

Meanwhile, there exists another structure of knowledge base for ontology method. Besides the similar characteristics with tree-shaped ontology, this graph-shaped ontology consists of not only relations between different levels of concepts (parent concept and child concept), but also direct relations between the concepts in the same level (concept and its siblings or cousins). The characteristics of graph-shaped ontology can be listed as:

- It is possible that there exist multiple inheritance relations between concepts (such as, concept J is subclass of E and F). In this situation, the depth of the concept has multiple values, which means there may exist multiple paths from targeted concept to root concept.

- Since there may exist numerous paths from the targeted concept to the root concept, the distance between the targeted concept to root concept should be the depth of the shortest path between them.

- Like the hierarchical concept tree, the longer semantic distance between two concepts is, the less the similarity between them.

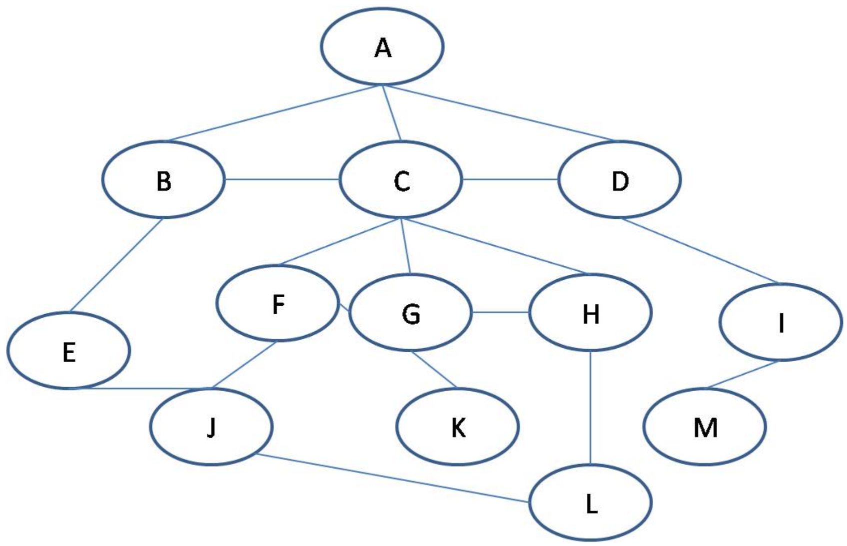

Figure 4.4: Example of Graph Shaped Concept Base

4.6.2 Distance Table Generation

To each concept in the hierarchical ontological concept tree or concept graph, I can get their paths to root concept, and compute related distance values between concepts in paths, and I can use the paths and distance value to create a concept distance table. In this table, I can find the distance and routing path between each concept and root concept. Furthermore, since there may exist multiple paths from one concept to root concept in a concept graph, among them, the shortest path should be the distance from targeted concept to its root concept. And also, for two concepts measuring their distance in next step, I can find their common ancestors and the distance with common ancestors in distance table. The example of distance table according to the two different structures of knowledge bases for ontology can be seen in Table 4.6:

Since the example knowledge base is very simple, the shortest distance from graph-shaped concept base is almost same as tree-shaped counterpart. But when the construc-

Table 4.6: Distance Table of Concept Relation for Two Types of Knowledge Bases

Types of Knowledge Base	Concept	Shortest Routing Path and Distance
Tree-Shaped Concept Base	A	Root Concept
	B	(B,A)(1)
	C	(C,A)(1)
	D	(D,A)(1)
	E	(E,B,A)(1,1)
	F	(F,C,A)(1,1)
	G	(G,C,A)(1,1)
	H	(H,C,A)(1,1)
	I	(I,D,A)(1,1)
	J	(J,E,B,A)(1,1,1)
	K	(K,G,C,A)(1,1,1)
	L	(L,H,C,A)(1,1,1)
	M	(M,I,D,A)(1,1,1)
Graph-Shaped Concept Base	A	Root Concept
	B	(B,A)(1)
	C	(C,A)(1)
	D	(D,A)(1)
	E	(E,B,A)(1,1)
	F	(F,C,A)(1,1)
	G	(G,C,A)(1,1)
	H	(H,C,A)(1,1)
	I	(I,D,A)(1,1)
	J	(J,E,B,A)(1,1,1),(J,F,C,A)(1,1,1)
	K	(K,G,C,A)(1,1,1)
	L	(L,H,C,A)(1,1,1)
	M	(M,I,D,A)(1,1,1)

tion of knowledge base is getting more complex, their difference can also be substantial. Meanwhile the concept routing and distance table can be very different. For example, the distance from one concept to root concept in concept tree is value 17. But in concept graph, there exists a shorter bypass for this concept to root concept and its distance value can be 8.

4.6.3 Semantic Distance Computation

First I take as an example a user i interest profile: $p_{im}(t) \in D(f_m)$ where user i profile feature f_m, m = "*soccer*" and $D(f_{im}) = \{C_i\}$ represents domain value of interest of user i. Then user j interest profile: $p_{jm}(t) \in D(f_m)$ where user j profile feature f_m, m = "*interest*" and $D(f_{jm}) = \{C_j\}$ represents domain value of interest of user j.

A Tree-shaped Knowledge Base

Based on the distance table of ontological knowledge base, I can compute the semantic distance between any two concepts in knowledge base. The only problem is that the two different shapes of the knowledge base may provide different values for the distance between concepts. In tree-shaped knowledge bases, there exists only one path between two concepts comparing to graph-shaped. So for the tree-shaped knowledge bases, the distance between two concepts can be measured: find the common ancestor in this path, and use the distance between concept C_i and its common ancestor with concept C_j to calculate the distance value between concept C_i and C_j. To facilitate this idea, I adopt the measurement algorithm from [47]:

$$d(C_i, C_j) = \frac{1}{2}\left(\frac{d(C_i, C_a)}{d(C_i, C_r)} + \frac{d(C_j, C_a)}{d(C_j, C_r)}\right) \tag{4.1}$$

where,

- C_i and C_j represent concepts from domain values of feature $D(f_{im})$ and $D(f_{jm})$ of user i and user j

- C_a represents common ancestor of C_i and C_j in the hierarchical tree-shaped knowledge base.
- C_r represents the root concept in the hierarchical tree-shaped knowledge base.
- And the value of results is on mathematical type of [0,1], where value "0" represents minimum distance and value "1" represents the maximum distance.
- According to this measured result, the similarity value might be directly calculated.

A Graph-shaped Knowledge Base

For the graph-shaped knowledge base, there exists several paths between concepts. Among them, we should choose the shortest path to calculate concept distance. According to the concept distance table, the semantic distance between concept C_i and concept C_j can be calculated according to different scenarios:

Scenario 1 If concept C_i and concept C_j are the same concept, $d(C_i, C_j) = 0$.

Scenario 2 If the direct path relation between concept C_i and concept C_j, $d(C_i, C_j) = d(R(C_i, C_j))$, where $R(C_i, C_j)$ is the distance between concept C_i and concept C_j mapped from the value of distance routing table.

Scenario 3 If there exists several path between concept C_i and concept C_j,

$$d(C_i, C_j) = \sum_{C \in Spath(C_i, C_j)} R_C(C_i, C_j)$$

where Spath defines the shortest path between concept C_i and concept C_j, C represent the set of common concepts in the shortest path between concept C_i and concept C_j. For instance, concept C_E is the common concept linking concept C_i and concept C_j in their shortest path, $d(C_i, C_j) = d(R(C_i, C_E)) + d(R(C_E, C_j))$.

4.6.4 Semantic Similarity Computation

A Tree-shaped Knowledge Base

According to the measured distance between two concepts in a tree-shaped knowledge base, the similarity of interest feature between user i and user j can be calculated as:

$$s(C_i, C_j) = 1 - d(C_i, C_j) \tag{4.2}$$

$$s_m(p_{im}(t), p_{jm}(t)) = \frac{\sum s(C_i, C_j)}{N_i * N_j} \tag{4.3}$$

where,

- $p_{im}(t)$ and $p_{jm}(t)$ represent two users' interest features in their profile
- N_i represents the number of concepts in user i interest feature
- N_j represents the number of concepts in user j interest feature
- C_i and C_j represent values from "interest" feature's domain $D(f_{im})$ and $D(f_{jm})$

A Graph-shaped Knowledge Base

Based on the theory [93] as "*The longer semantic distance between two concepts, the less similarity they shared. That is, semantic similarity and semantic distance have inverse relation* ", the similarity function is modeled as a linearly decreasing function in the distance. So based on the distance results of concepts from graph-shaped knowledge base, the similarity of interest feature between user i and user j can be measured as various scenarios:

Scenario 1 If $d(C_i, C_j) = 0$, $s(C_i, C_j) = 1$, $s_m(p_{im}(t), p_{jm}(t)) = 1$

Scenario 2,3 $s(C_i, C_j) = \frac{1}{K*d(C_i,C_j)}$, $s_m(p_{im}(t), p_{jm}(t)) = \frac{\sum s(C_i,C_j)}{N_i*N_j}$

where,

- $p_{im}(t)$ and $p_{jm}(t)$ represent two users' interest features in their profile
- N_i represents the number of concepts in user i interest feature
- N_j represents the number of concepts in user j interest feature
- C_i and C_j represent values from "interest" feature's domain $D(f_{im})$ and $D(f_{jm})$
- K ($K \in (0, 1]$) represents the concrete value reflecting the impact degree of semantic distance to semantic similarity.

And for two types of knowledge base of semantic similarity, $d_m(p_{im}(t), p_{jm}(t)) = 1 - s_m(p_{im}(t), p_{jm}(t))$ where d_m represents the dissimilarity between users' profiles.

4.7 Summary

To simplify the similarity measurement process between users in SSCs, I introduced the use of ontologies and proposed an appropriate similarity function for our model. With the combination of syntactic and semantic similarities, I can quantify the similarity between users profiles. Based on experimental results, the threshold for friendship and membership can be determined. In next chapter, I will simulate the advertising over SSCs of the proposed middleware and assess its performance.

Chapter 5

Performance Evaluation

This chapter provides details about assessing the performance of the proposed middleware with respect to efficiently delivering advertisements messages to the right audiences. The operation of proposed middleware of SSCs was simulated through Netlogo. The reminder of this chapter is organized as follows: Section 5.1 describes the platform of simulation implementation; Section 5.2 describes the system implementation to validate the proposed middleware; and in Section 5.3 I propose the configured parameters for system operation. The performance evaluation will be provided in Section 5.4. Finally, I summarize this chapter in Section 5.5.

5.1 Simulation Platform

In order to test the performance of the proposed framework, I use simulation to model the operation of SSCs. By the definition from [94], simulation is "*the process of designing a model of a real or imagined system and conducting experiments with that model. The purpose of simulation experiments is to understand the behavior of the system or evaluate strategies for the operation of the system*". In computer environments, simulation is an attempt to model a real-life or hypothetical situation on a computer so that it can be studied to see how the system works. By changing parameters in the simulation environments, predictions may be made about the behavior of the system. It is defined as a tool to virtually investigate the behavior of the system under study [94]. So based on the above reference, simulation is an appropriate method for us to measure the performance of SSCs model.

5.1.1 Introduction of NetLogo

In our case, I choose NetLogo [95] as the platform to implement the simulated model. NetLogo is a programmable modeling environment for simulating natural and social phenomena. It is an agent-based modeling tool for modeling self-organizing networks or complex networks. One thing that distinguishes NetLogo from other tool is its very strong user community [96]. There exists thousands of third-party extensions to NetLogo. And users can choose any of them to program its own modeling simulation environment. NetLogo is a visual tool and is extremely suitable for interactive simulations. NetLogo world consists of three main properties: turtle, patch and observer. In our case, the three main properties can be represented as:

- Turtle: the dynamic component moves under the simulation world. In our case, I use it to model the users of SSCs.

- Patch: a single place where the turtle resides and composed patches forms the simulation world of NetLogo. In our case, patch is used to model the location of users in the environment.

- Observer: this is a context that can be used in general without relating to either a patch or a turtle. In our case, the observer can be seen as the programming code to control the interaction of users of SSCs. Furthermore, all the logistics of the proposed framework will be visualized by the programming code of observer.

5.1.2 Advantages of NetLogo

The reasons why I choose NetLogo as the simulation platform to model the framework are listed as follows [96]:

- NetLogo is an agent-based modeling tool, which means it is specifically tailored to developing self-organizing or complex system simulations. There exists several strong features for the NetLogo such as direct addressability of nodes, ease of implementation and evaluation of self-organization, and emergence and bio-inspired algorithms as well as the capability of being understandable from the human perspective (having their background rooted in social simulation), all of which make them extremely useful for application in the domain of ad hoc, P2P, and pervasive systems. NetLogo is well suitable for the "Ad Hoc" environment of proposed middleware.

- Another point to note here is that in NetLogo, modeling complex protocols does not have to be limited to the simulation of networks alone; it can easily be used to model human users, intelligent agents interacting with the system, or virtually any concept the framework designer feels worthwhile having in the model. Since the proposed middleware is designed as Ad hoc MSNs under MANETs environment, the main feature of which is mobile users, NetLogo is an very appropriate tool to simulate this scenario.

- The strong user community for NetLogo supports the diversified functionalities of NetLogo. I can choose any of extensive add-on applications of Netlogo to contribute the simulation system.

5.1.3 NetLogo SQL Wrapper

NetLogo SQL Wrapper is an extension to NetLogo which adds primitives to the NetLogo modeling language to support the access to databases using SQL (Structured Query

Language) [97]. In our case, the database is necessary to store the related system information such as user information, knowledge base for constructing users' profiles, the resource of advertisements shared within community and information about community structure. I adopt this extension with NetLogo to link the database "MySQL" with our simulation system.

5.2 Simulation System Implementation

In this section, I present the simulation prototype to validate the proposed approach. The simulation prototype of the proposed middleware is composed of three main components as shown in Figure 5.1:

Component 1 Profile Generator: it is used to generate random social profiles with domain values of features from the FOAF vocabulary and knowledge base. Since the social profiles in our framework consist of multiple features (attributes), the limited domain for some features need to be pre-defined. For example, the possible value for "Gender" feature just contain "Male" and "Female". In terms of the feature "Interest", the knowledge base may be constructed according to two different shapes: tree-shaped and graph-shaped. I may take one of them to define the similarity relationship between feature domain.

Component 2 Community Construction: it is used to construct the communities. In this component, I may have several sub-functions: similarity measurement and membership indication. In the similarity measurement function, with the contribution of information routing protocol from MANETs, users can achieve the remote content distribution or information transmission. Potential users in certain area can measure similarity with community profile stored in neighbors. In membership indication process, based on the community construction threshold and measured similarity score, the community composition can be updated. All the information related to this component must be stored in database for performance evaluation.

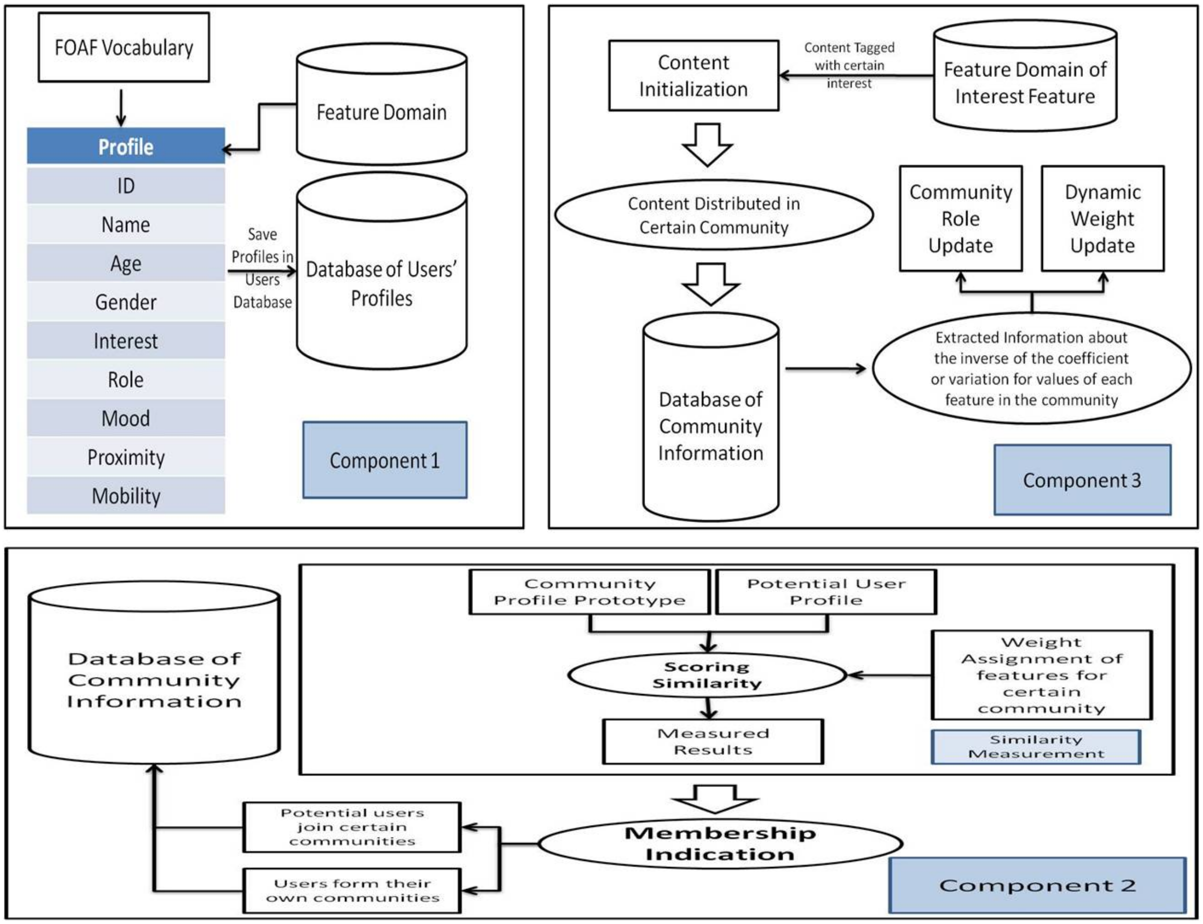

Figure 5.1: Overview of Simulation System

Component 3 Targeted Advertising: it is the component responsible for advertising over targeted communities of SSCs. Based on the constructed SSCs, marketers can spread advertisements or service promotions to the targeted community. An advertisement is tagged with ceratin value from the domain of "Interest" feature. With the frequency of different types of advertisements shared in communities, the short-term interest feature for users in certain communities may also update. And also the friendship manager is added in this component: if users lose all friendships with members in certain community in a period, it is reasonable to assume that they may leave this community. Furthermore, the dynamic feature weights for each community is always updated with the change of community composition. All the information related to this component is stored in database for performance evaluation.

5.3 Parameter Setup

Used parameters can be classified into three categories:

- Environment parameters: the parameter includes network environment for the proposed middleware such as network size, user mobility, user arrival rate and etc..
- Parameters for semantic similarity: is a set of knowledge bases for "Interest" features of users profiles.
- Parameters for syntactic similarity: is a set of knowledge bases for features of syntactic similarity.

5.3.1 Environment Parameters for the Proposed Middleware

To effectively simulate the proposed framework, an enclosed ad hoc network environment was adopted. The enclosed area that contained users of SSCs was off an area of $1000 \times 1000m^2$. Users were randomly distributed over the areas where their arrival rate was set as (0-10) users per minute, which means for every minute it is possible that zero to ten users may enter the enclosed area. The mobility of the users was set to be randomly generated between (0,5) km/hour. The proximity for communication range of user was set to be $50m$. Each user owns a set of randomly generated social profiles. Details of environment parameters, and a snap-shot of the simulated communities are shown in Table 5.1 and Figure 5.2, respectively. And nodes with color besides black are the community leaders.

Table 5.1: Sample of Context Frame Problem Components at Different Layers

parameter	value or range
Environment Area	$1000 \times 1000m^2$
Range of the User Arrival Rate	0-10 users/min
Range of User Speed	0-5 km/hour
Device Communication Range	50 m
Maximum Friendship Distance in a Community	10 hops
Friendship Expiration Rate	10 min
Number of Interest Features $\|\mathcal{F}\|$	50
Targeted Advertising Threshold	0.5
Community Leaders Threshold	0.875
Community Participants Threshold	0.5
Community Passive Members Threshold	0.2
Value for Interest Feature Step Update Δv	0.03

5.3.2 Parameters for Semantic Similarities

In Chapter 4, I found two different types of knowledge bases for semantic similarities: tree-shaped and graph-shaped. In performance evaluation, I adopted tree-shaped semantic similarity as a knowledge bases to simplify the similarity calculation process. The adopted tree-shaped knowledge bases can be seen from Figure 5.3.

5.3.3 Parameters for Syntactic Similarities

Different from feature "Interest", I set up the parameters and results for other features' domain for similarity measurement. The similarity value is on mathematical types of [0,1]. Value "1" represents full similarity, which is contrary to value "0". The measured results are listed in Table 5.2 respectively.

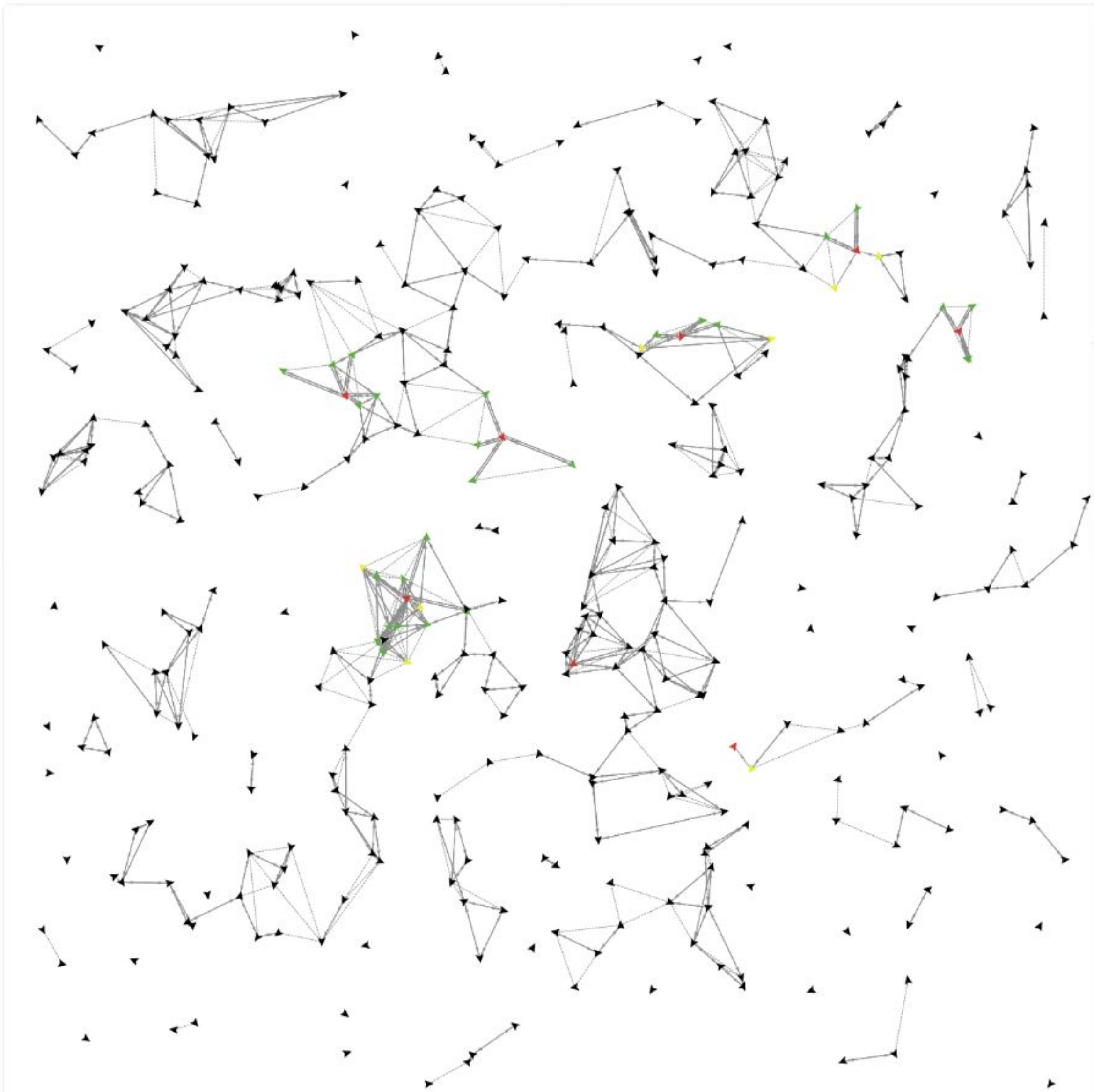

Figure 5.2: Simulation Environments

5.4 Evaluation Results

I compare our scheme against two traditional advertisement propagation models in ad hoc networks, which is referred to as the Influential K model [60] and Model A. In the former, advertisement propagation commences from a community leader of a certain community

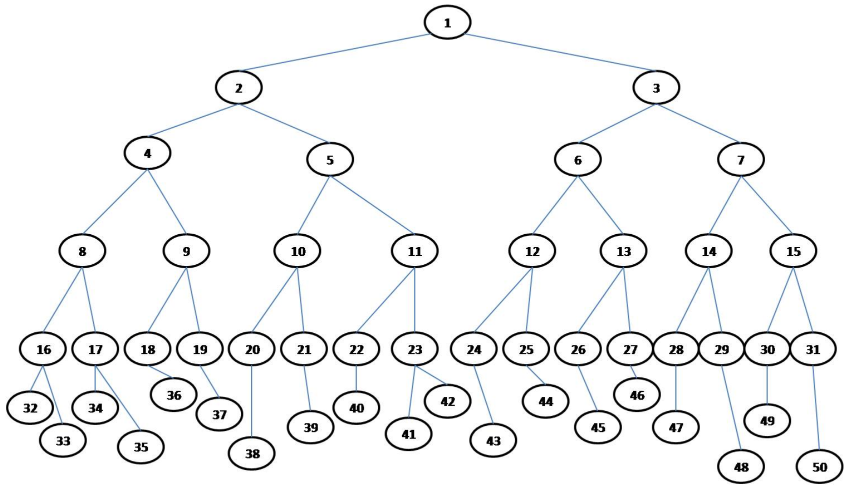

Figure 5.3: Parameter for Tree-shaped Semantic Knowledge Base

Table 5.2: A Sample of the User Features and Their Similarities

Feature Domain (Similarity)			
Age	Teenage and Youngster	Middle-Aged	Elderly
Teenage and Youngster (10-20)	1	0.5	0.3
Middle-Aged (21-60)	0.5	1	0.5
Elderly (61-90)	0.3	0.5	1
Role	Role A	Role B	Role C
Role A	1	0.7	0.5
Role B	0.7	1	0.3
Role C	0.5	0.3	1
Current Mood	High	Medium	Low
High	1	0.8	0.5
Medium	0.8	0.5	0.3
Low	0.5	0.3	0.1
Mobility	High	Medium	Low
High	1	0.7	0.1
Medium	0.7	1	0.4
Low	0.1	0.4	1
Gender	Male	Female	N/A
Male	0.6	0.4	N/A
Female	0.4	0.6	N/A

that shows a high interest in the advertisements and forwards these information to all the users who are in same community. In Model A, I randomly choose a un-targeted first user which again forwards it to all its neighbors until advertising information is expired, which is referred to as flooding [98].

Although there are existing some well-proved advertising metrics such as GRPs, TRPs, CPM, etc. [99], they are mainly developed for more static and stable environments. So I propose an appropriate variable that evaluate the performance of the proposed middleware in terms of the ability to deliver the appropriate advertisements to the interested users, which I term as the *Average Received Advertising Score* (ARAS). I measure the average semantic similarity between shared advertisements and all the received users:

$$score(A, C_k(t)) = \sum_{a_m \in A} s_m(a_m, p_{km}^{\text{comm}}(t)) \tag{5.1}$$

then the ARAS can be calculated as follows:

$$ARAS = \frac{\sum score(A_i, P_{km}(t))}{N_A} \tag{5.2}$$

where an advertisement $A = \{a_1, a_2, ...\}$ with interest features $\{a_1, a_2, ...\}$ as a_m was proposed to the receivers in community K with a profile prototype defined as $P_k^{\text{comm}}(t)$ and N_A is the number of receivers. It is intuitive to state that a higher ARAS means that the advertisements are more appealing to the receivers since their "interest" feature is semantically similar to the advertisements. I also measure the total number of messages needed in order to deliver the shared advertisement to all users that might be interested. Finally, I examine the efficiency of the proposed model with respect to the stability in the size of the constructed communities.

As depicted by Figures 5.4 and 5.5, our proposed work shows a stable and significant improvement in the achieved value of ARAS when compared to the Influential K model [60] and Model A. Meanwhile, the associated cost, in terms of the number of messages needed to deliver shared advertisement is much smaller specially when the network size increases significantly. These results indicate that the proposed work leverages effective advertisement transmission and sharing, i.e., gets the right advertisement to right users, while achieving routing efficiency.

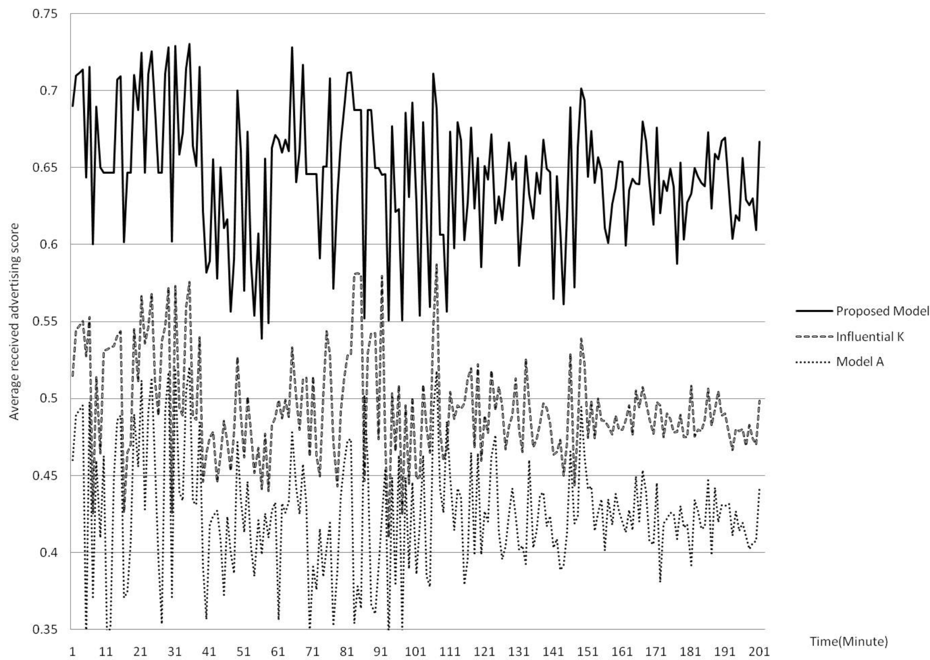

Figure 5.4: Effectiveness for Users Receiving Advertisement

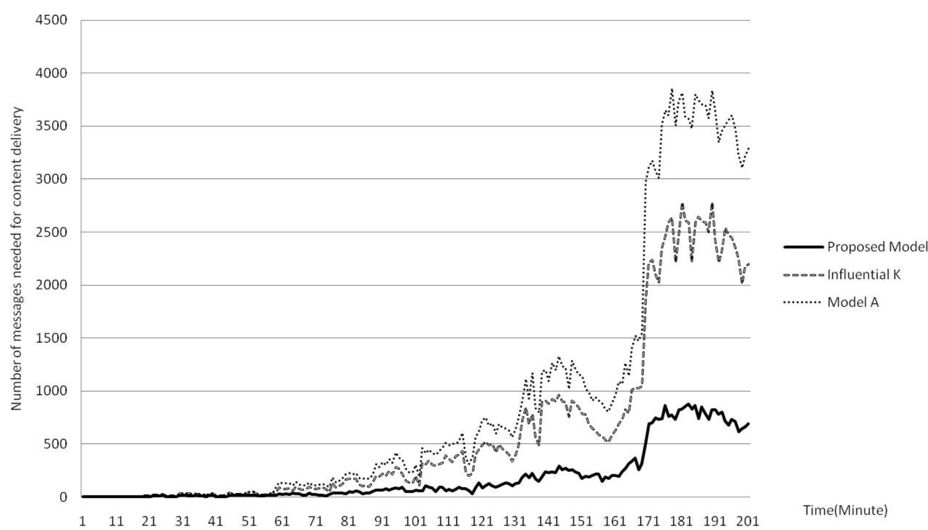

Figure 5.5: The Total Amount of Advertisement Routing

5.4.1 Single Feature vs. Multiple Features for Similarity Measurement

In this section, I measure the difference between single feature-based (just feature "Interest") and multiple features-based approaches with respect to ARAS and amount of routing messages needed. The obtained results are shown in Figure 5.6 and Figure 5.7.

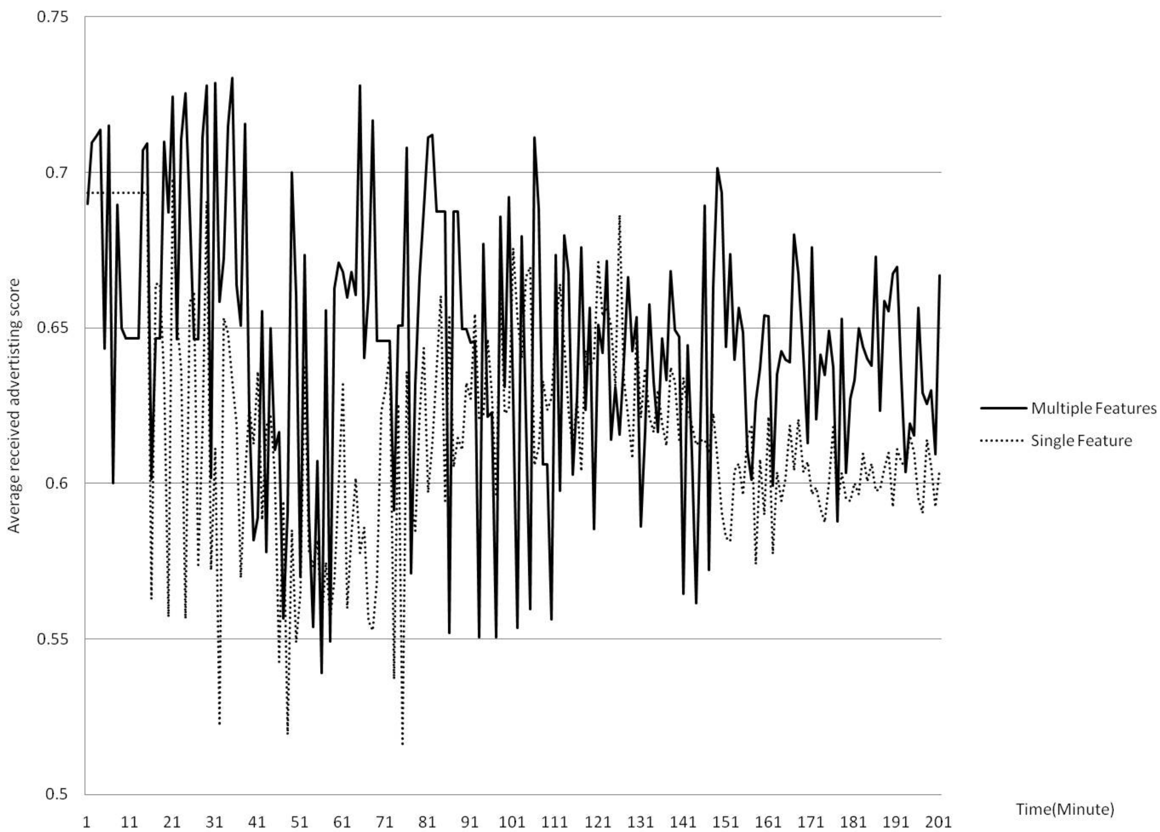

Figure 5.6: Effectiveness for Users Receiving Advertisement

Based on the simulation results, It is notified that the ARAS value for multiple features is less stable than the single feature-based approach although both of them share high value. But with the simulation processing, the ARAS value for multiple features progressed higher than the single feature. Meanwhile, the amount of advertisement routing messages when multiple features used is much less than its single feature counterpart,

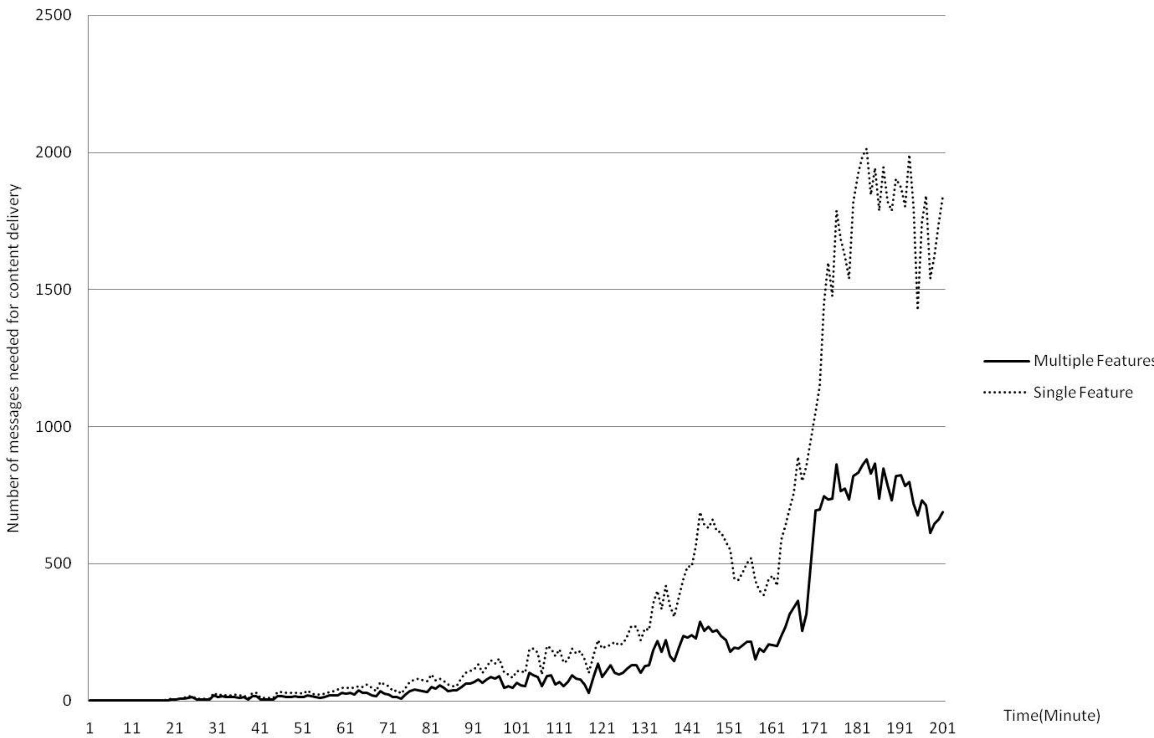

Figure 5.7: The Total Amount of Advertisement Routing

which means the flooding of network environment is controlled more efficiently. Then I can state that multiple features-based approach for our model shares a higher ARAS and better performance in controlling the network traffic than single feature-based approach. So multiple features for similarity measurement in proposed framework is an appropriate choice.

5.4.2 Controlled Hops for Community Construction

With MANETs routing protocols, users in our model can achieve longer information transmission with remote users. For instance in our case, the scalability of community members can be expanded to several hops. The “Hop” is a quantified unit to define the maximum distance for spreading the information from a source user to terminus. In this section, I examine different levels of hops to find out appropriate level of hops for our model. The simulation results are shown in Figure 5.8.

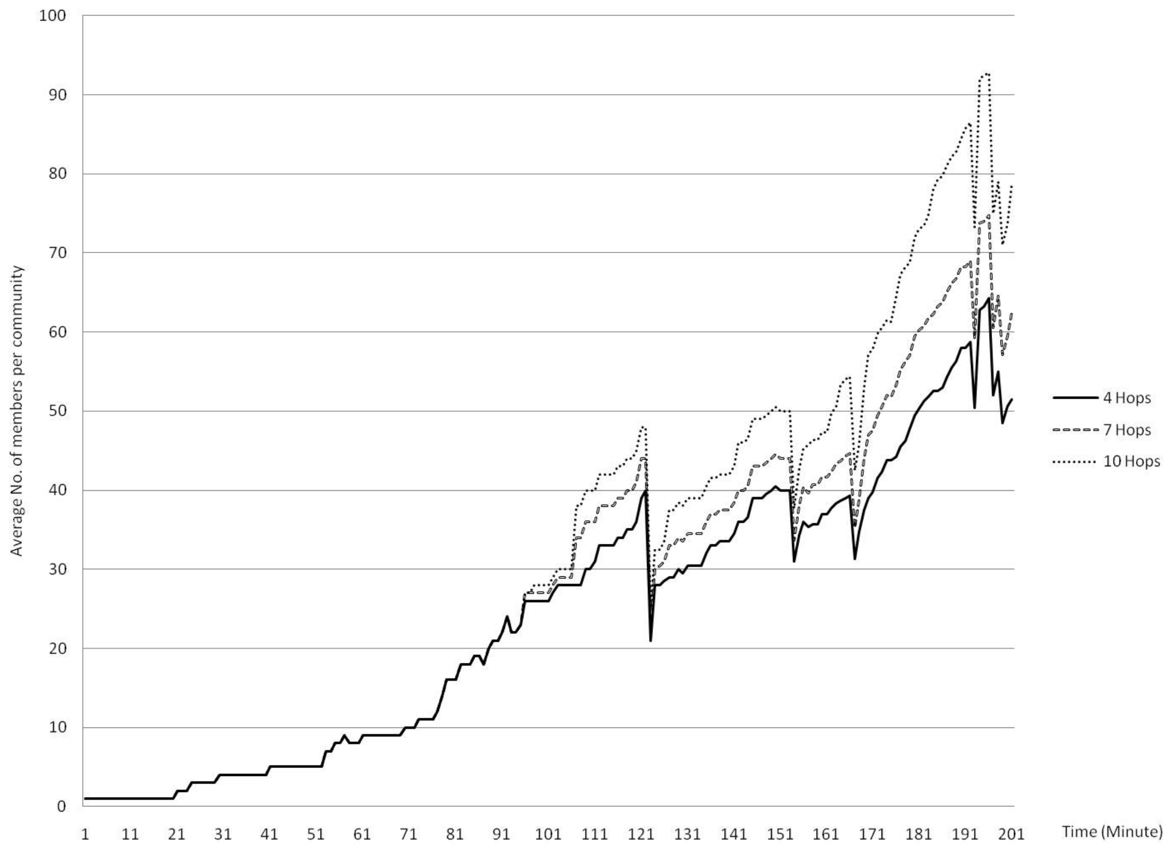

Figure 5.8: Compared Community Size By Different Hops

I find that in the early stages of simulation (limited number of users arriving in testing area). The community size by three different levels of hops is almost the same. With the simulation processing, more and more users entering the testing area, it is obvious that ten hops of routing contributes to a larger size of community than other two since the scope of routing for ten hops is larger. However, with simulation processing longer, the difference of community size by difference levels of hops is increasingly significant. And it is reasonable to state that more hops of routing protocol, more cost for the information flooding since the increased levels of hops cause more information spreading in the network. The number of hops for information routing should be adjusted according to the size of area.

5.4.3 Dynamic Weight Update vs. Fixed Weight Assignment

In the proposed model, I adopt dynamic weight update to find feature in user profile that are more important for certain communities. I have also tested the enhancement obtained from the associated feature weights vector $W_k(t)$, where the simulation was run once without adopting our weight update methodology and then after employing the dynamically updated vector of weights. The compared variables include the amount of advertisement routing and average community size. The results are listed in Figures 5.9, 5.10.

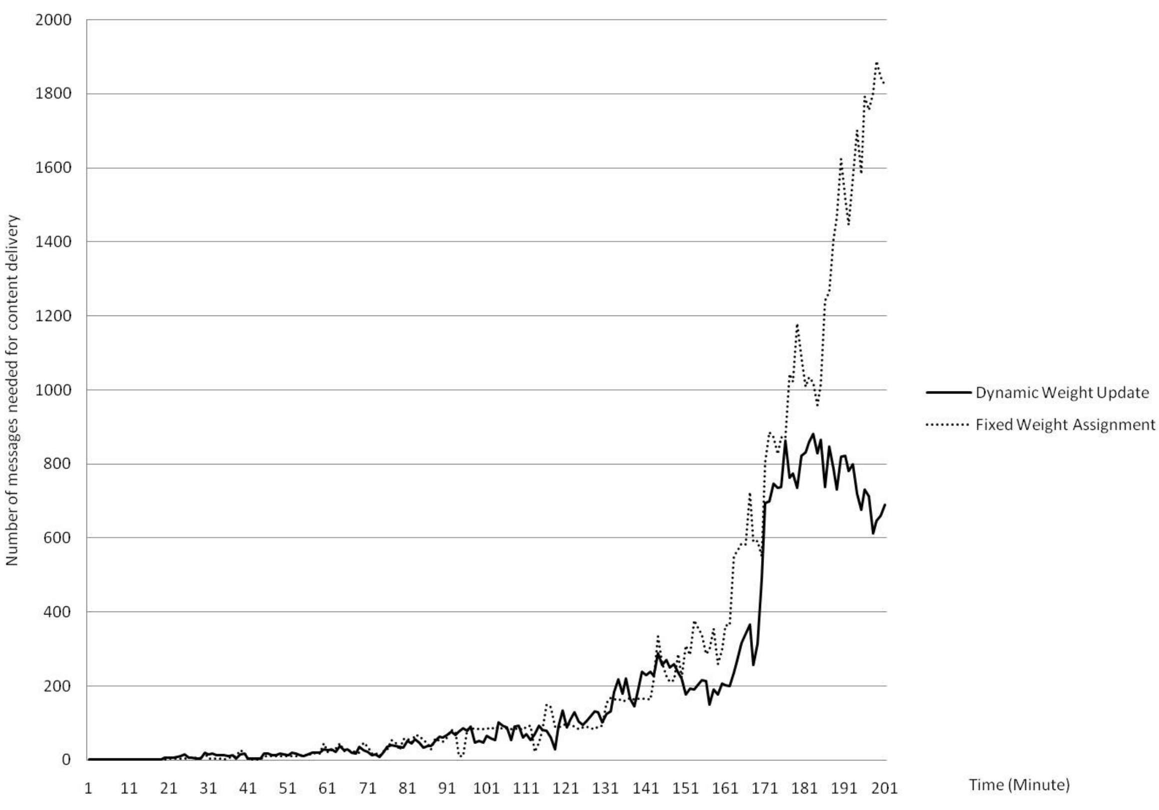

Figure 5.9: The Total Amount of Advertisement Routing

Based on the simulation results, I find that dynamic weight update outperformed

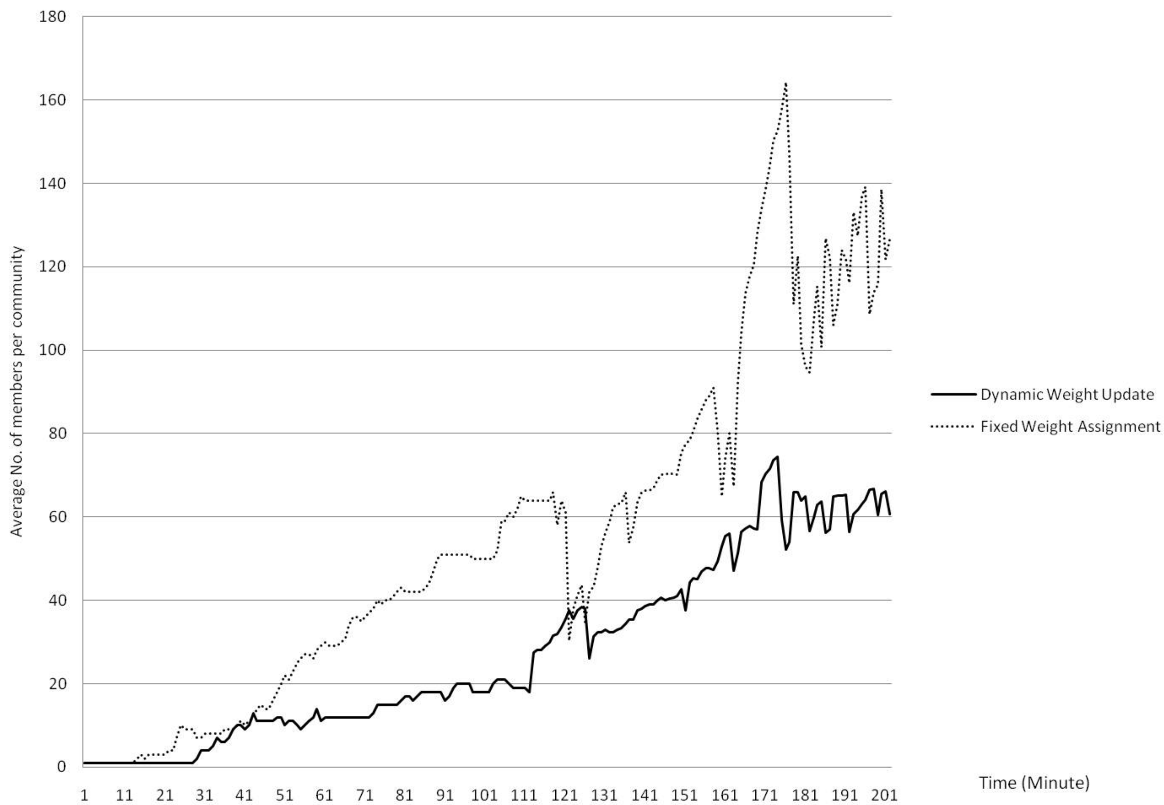

Figure 5.10: The Average Size of Each Community

fixed weight assignment with respect to controlling the number of messages needed to deliver advertisements. As shown in Figure 5.10, the size of communities built using dynamic weights shows a higher degree of stability, i.e., on average the number of users leaving a community is the same as those joining it. This behavior can be attributed to the accurate presentation of the community profile prototypes. As each community's focus changes, the dynamic weights quickly reflect these changes and help the friendship managers to detect friends.

5.5 Summary

In this chapter, I adopted a simulator to assess the performance of the proposed middleware with respect to delivering the advertisement messages to the right audiences. The acquired results showed that the proposed model outperformed the Influential-K model [60] and the Flooding model [98] both in ARAS and message traffic control. And the adopted functions such as multiple-features profile and dynamic weight update contribute to the better performance in ARAS, message traffic control and community size control.

Chapter 6

Conclusion and Futurework

6.1 Summary of Contributions

In this thesis, I have presented a novel model for spontaneous social communities (SSCs) that efficiently captures the dynamic interests and changes in the user's focus in these communities. These social communities are represented through an inferred *community profile prototype*. This prototype is calculated in a distributed manner and employed to facilitate the identification of new potential members.

For modeling users in SSCs, I utilized the FOAF vocabulary to adopt the multiple features to construct users profiles. To achieve user connection, I introduced similarity between users profiles to calculate the users' social distance in SSCs. During this process, according to the characteristics of each feature, I categorized feature measurement functions into two groups: syntactic and semantic similarities. In addition, due to the different structures of knowledge bases for the semantic similarity measurement, I proposed two different algorithms to calculate the distance and similarity between semantic concepts. Moreover, to define the importance of each feature in the profiles, I proposed a dynamic weight update approach correlated with the community composition.

To this end, I evaluated the performance of the proposed middleware of SSCs through the NetLogo simulator [96] [97]. The obtained results illustrated the efficiency in advertisements sharing within SSCs. The use of multiple features for profile construction and dynamic weight updating in the proposed model have shown to contribute to controlling the flooding of advertisements over SSCs. The proposed SSCs can help marketers to conduct targeted marketing in Ad hoc MSNs.

6.2 Limitations and Future Directions

However, some limitations still exist in the proposed middleware:

- With respect to performance evaluation, there is a lack of the metrics that evaluate the detailed operation performance of the proposed middleware.
- In performance evaluation stage, there is no comparison between the proposed middleware and the well-known static online social networks in terms of advertising effectiveness.
- The threshold for proposing advertisements is selected experimentally for simulation. The optimal value of threshold for marketers have not been discussed in the thesis.
- The compatibility issue of the proposed middleware with regards to the device was not addressed.
- The scalability of the proposed middleware has not been measured in the performance evaluation. So whether or not the proposed middleware can accommodate for a large number of users is still an unaddressed question.

To address these limitations, some improvements can be incorporated in the future:

- In the performance evaluation, I adopted a tree-shaped knowledge base for measuring the similarity between "interest" features of users. But in proposed scheme, there exists another graph-shaped knowledge base, I may measure their difference for the network performance of SSCs in further research.

- For the privacy concern section, I considered adopting a public key and exclusive key to encrypt the content spreading in the network of proposed middleware. This may improve the content privacy of our model. But how well this approach affect the network performance is an issue that may be discussed in the future. The details and practical implementation of this approach is still subject to further research.

- Another direction for future research is the incorporation of users' profiles and friendship lists obtained from other more static social networks, and to compare advertising results between static social networks and proposed SSCs.

- Based on the marketing campaign strategy, the optimal value of threshold for proposing advertisements should be measured.

- The practical implementation of the proposed middleware over mobile operating systems should be presented in the future work.

- The last future direction may aim at investigating the scalability of the performance of the proposed work over networks with a large number of heterogeneous users. For example, where there are over dozens of thousands of users in the simulation environment.

Bibliography

[1] D. Frommer, “Smartphone penetration rates,” 2010. [Online]. Available: http://www.businessinsider.com/smartphone-penetration-rates-6-2010-4-charts

[2] V. Vaishnavi and W. Kuechler, “Design science research in information systems,” September 2011. [Online]. Available: http://desrist.org/desrist

[3] J. E. Jason J. Jung, “Towards semantic social networks,” in *ESWC’ 07 Proceedings of the 4th European conference on The Semantic Web: Research and Applications.* Berlin, Heidelberg: Springer-Verlag, 2007.

[4] M. S. Jason Hartline, Vahab S. Mirrokni, “Optimal marketing strategies over social networks,” in *In the proceedings of International World Wide Web Conference 2008*, Beijing, China, April 2008, pp. 189 – 198.

[5] C. Leberknight, H. Inaltekin, M. Chiang, and H. V. Poor, “The Evolution of Online Social Networks: A tutorial survey,” in *Signal Processing Magazine, IEEE*, vol. 29, no. 2. IEEE, March 2012, pp. 41 – 52.

[6] “Facebook statistics,” http://www.facebook.com/press/info.php?statistics.

[7] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, “Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application,” in *Proceedings of the 6th ACM conference on Embedded network sensor systems, SenSys ’08.* ACM, 2008, pp. 337 – 350.

[8] M. K. Marco von Arb, Matthias Bader and R. Wattenhofer, "Veneta: Serverless friend-of-friend detection in mobile social networking," in *In: Networking and Communications, 2008. WIMOB '08. IEEE International Conference on Wireless and Mobile Computing*, 2008, pp. 184 – 189.

[9] H. K. SONI, ""ad hoc network"," in *DoS attack in MANET*, Mar 22nd 2011.

[10] T. Krag and S. Bettrich, "Wireless mesh networking." O'Reilly Wireless Dev Center, Retrieved 2009-01-20.

[11] J. P. Alan R. Hevner, Salvatore T. March, "Design science in information systems research," in *MIS Quarterly*, vol. 28, no. 1, 2004, pp. 75 – 105.

[12] P. J. Denning, "A new social contract for research," in *Communications of the ACM (40:2)*, February 1997, pp. 132 – 134.

[13] D. Tsichritzis, "The dynamics of innovation," in *Beyond Calculation: The Next Fifty Years of Computing, P. J. Denning and R. M. Metcalfe (eds.)*. New York, USA: Copernicus Books, 1998, pp. 259 – 265.

[14] V. P. T. T. Y. H. Takeda, H., "Modeling design processes," in *AI MagazineWinter*, 1990, pp. 37 – 48.

[15] J. Preece, "Sociability and usability: Twenty years of chatting online," in *Behavior and Information Technology Journal*, vol. 20, no. 5, 2001, pp. 347 – 356.

[16] "What is middleware?" http://www.middleware.org/whatis.html/.

[17] S. Robinson, *Simulation - The practice of model development and use.* Wiley, 2004.

[18] Z. Tian and N. Samaan, "A novel distributed scheme for building spontaneous social communities over manets," in *The Ninth IEEE International Conference on Mobile Ad-hoc and Sensor Systems (IEEE MASS 2012)*, Las Vegas, Nevada, USA, October 8th - 11th 2012.

[19] F. K. Wasserman Stanley, "Social network analysis in the social and behavioral sciences," in *Social Network Analysis: Methods and Applications. Cambridge University Press*, 1994, pp. 1 – 27.

[20] C. A. H. Elizabeth F. Churchill, "Social networks and social networking," in *IEEE INTERNET COMPUTING*. IEEE Computer Society, Sep - Oct 2005.

[21] J. P. Scott, *Social Network Analysis: A Handbook (2nd edition)*. Sage Publications, 2000.

[22] R. Azarian, "The general sociology of harrison c. white: Chaos and order in networks." Palgrave USA, 2005, pp. 135 – 140.

[23] E. N. Boyd Danah, "Social network sites: Definition, history, and scholarship," in *Journal of Computer-Mediated Communication*, vol. 13, no. 1, 2007, pp. 211 – 230.

[24] K. P. G. P. D. B. B. Alan Mislove, Massimiliano Marcon, "Measurement and analysis of online social networks," in *IMC'07, proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, ser. ACM, New York, NY, USA, October 2007.

[25] "Flickr," http://www.flickr.com/.

[26] "Twitter," https://twitter.com/.

[27] "Facebook," http://www.facebook.com/.

[28] "Linkin," http://www.linkedin.com/.

[29] R. H. X. Z. A. M. Foster Provost, Brian Dalessandro, "Audience selection for on-line brand advertising: Privacy-friendly social network targeting," in *KDD '09*, Paris, France, June 28-July 1 2009.

[30] N. D. Ziv and B. Mulloth, "An exploration on mobile social networking: Dodgeball as a case in point," in *ICMB'06, IEEE conference proceedings, MBusiness Conference*, Copenhagen, Denmark, June 2006, p. 21.

[31] “Top u.s. smartphone app categories,” July 25th 2011. [Online]. Available: http://www.comscoredatamine.com/2011/07/top-u-s-smartphone-app-categories/

[32] C. Lai, “Understanding the design of mobile social networking,” in *M/C Journal*, vol. 10, no. 1, March 2007. [Online]. Available: http://journal.media-culture.org.au/0703/08-lai.php

[33] P. W. E. H. Nipendra Kayasthan, Dusit Diyato, “Applications, architectures, and protocols designed issues for mobile social networks: A survey,” in *Proceedings of the IEEEE*, September 6th 2011.

[34] M. Hasel, “Opensocial: An enablers for social applications on the web,” in *COMMUNICATIONS OF THE ACM*, vol. 54, no. 1, January 2011, pp. 139 – 144.

[35] C. K. Toh, “Ad hoc mobile wireless networks,” in *Prentice Hall Publishers*, 2002.

[36] M. G. Y. Yi and T. J. Kwon, “Efficient flooding in ad hoc networks using on-demand (passive) cluster formation,” in *Proceedings of the ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, 2002.

[37] G. T. C. Ho, K. Obraczka and K. Viswanath, “Flooding for reliable multicast in multi-hop ad hoc networks,” in *Proceedings of the International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communication (DIALM)*, 1999.

[38] P. Dhakan and R. Menezes, “The role of social structures in mobile ad hoc networks,” in *43rd ACMSoutheast Conference*, vol. 2, Kennesaw, GA, USA, March 2005, pp. 59 – 64.

[39] P. S. S. Smaldone, L. Han and L. Iftode, “Roadspeak: Enabling voice chat on roadways using vehicular social networks,” in *Proc. Workshop Social Netw. Syst*, April 2008, pp. 43 – 48.

[40] A. Pietilinen, E. Oliver, J. LeBrun, G. Varghese, and C. Diot, "Mobiclique: Middleware for mobile social networking," in *In 2nd Workshop on Online Social Networks*, 2009, pp. 49 – 54.

[41] "Sun microsystems standard," in *Jini Technology Surrogate Architecture Specification, Rev. 1.0*, 2001.

[42] M.Girvan and M. E. J. Newman, "Community structure in social and biological networks," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, June 2002, pp. 7821 – 7826.

[43] P. B. Guillaume-Jean Herbiet, "Sharc: community-based partitioning for mobile ad hoc networks using neighbourhood similarity," in *World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium*, Jan 14-17 2010, pp. 1 – 9.

[44] S. Fortunato, "Community detection in graphs." Viale S. Severo 65, 10133, Torino, I-ITALY: ISI Foundation, January 2010, complex Networks and Systems Lagrange Laboratory.

[45] L. Freeman, "A set of measures of centrality based upon betweenness," in *Sociometry 40*, 1977, pp. 35 – 41.

[46] J. E. Daniele Quercia and L. Capra, "Using mobile phones to nurture social networks," in *PERVASIVE computing.* IEEE CS, 2010, pp. 12 – 20.

[47] S. U. K. Juan Li, "Mobilesn: Semantics-based mobile ad hoc social network framework," in *Published in the IEEE "GLOBECOM" 09 proceedings*, 2009.

[48] G. C. Nan Li, "Multi-layered friendship modeling for location-based mobile social networks. mobile and ubiquitous systems: Networking and services," in *MobiQuitous '09. 6th Annual International*, July 2009.

[49] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," in *Physical Review E pp. 0261134*, vol. 69, no. 2, Feb 2004.

[50] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization," in *Physical Review E*, vol. 72, no. 2, August 2005.

[51] G. X. Wang and H. Lu, "A very fast algorithm for detecting community structures in complex networks," in *Physical A Statistical Mechanics and its Applications*, vol. 384, October 2007, pp. 667 – 674.

[52] R. A. U. N. Raghavan and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," in *Physical Review E*, vol. 76, 2007.

[53] S. H. W. G. G. Kevin A. Li, Timothy Y. Sohn, "Peopletones: A system for the detection and notification of buddy proximity on mobile phones," in *In MobiSys '08: Proceedings of the 6th international conference on Mobile systems, applications, and services*. New York, NY, USA: ACM, 2008, pp. 160 – 173.

[54] M. Endler, A. Skyrme, D. Schuster, and T. Springer, "MobilisGroups: Location-based Group Formation in Mobile Social Networks," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference*, 2011, pp. 502 – 507.

[55] A. Gupta, A. Kalra, D. Boston, and C. Borcea, "MobiSoC: a middleware for mobile social computing applications," in *Mob. Netw. Appl.*, 2009.

[56] E. Y. Pan Hui, Jon Crowcroft, "Bubble rap: Social-based forwarding in delay tolerant networks," in *MobiHoc'08*, Hong Kong SAR, China, May 2008.

[57] P. Costa, C. Mascolo, M. Musolesi, and G. Picco, "Socially-Aware Routing for Publish-Subscribe in Delay-Tolerant Mobile Ad Hoc Networks," in *IEEE Booktitle on Selected Areas in Communications*, vol. 26, no. 5, May 2008, pp. 748 – 760.

[58] S. B. Mokhtar, L. McNamara, and L. Capra, "A middleware service for pervasive social networking," in *Proceedings of the International Workshop on Middleware for Pervasive Mobile and Embedded Computing, M-PAC '09*. ACM, 2009, pp. 2:1 – 2:6.

[59] P. Domingos and M. Richardson, "Mining the network value of customers," in *KDD '01 Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM Press, 2001, pp. 57 – 66.

[60] Y. Wang, G. Cong, G. Song, and K. Xie, "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '10. New York, NY, USA: ACM, 2010, pp. 1039 – 1048.

[61] G. M. Basit Qureshi, D. Kouvatsos, and M. Ilyas, "An adaptive content sharing protocol for p2p mobile social networks," in *2010 IEEE 24th International Conference on Advanced Information Networking and Application Workshops*. IEEE Computer Society, December 2010, pp. 413 – 418.

[62] C. M. G. P. P. Paolo Costa, Mirco Musolesi, "Adaptive content based routing for delay tolerant mobile ad hoc networks." University College London, UK, 2006, research Notes RN/06/08.

[63] e. a. P. Hui, A. Chaintreau, "Pocket switched networks and human mobility in conference environments," in *In Proc. WDTN*, 2005.

[64] W. Gao, Q. Li, B. Zhao, and G. Cao, "Social-aware multicast in disruption-tolerant networks," *Networking, IEEE/ACM Transactions on*, to appear 2012.

[65] A. Santos, J. Cardoso, D. Ferreira, P. Diniz, and P. Chaínho, "Providing user context for mobile and social networking applications," in *Pervasive and Mobile Computing*, vol. 6, no. 3, pp. 324 – 341.

[66] M. Kendall and A. Stuart, in *The advanced theory of Statistics, Vol. 1, 4th Edition, London, (1977).*, vol. 1. Charles Grifin and Co., 1977.

[67] T. R. Gruber, "A translation approach to portable ontology specifications," in *Knowledge Acquistion 5(2)*, June 1993, pp. 199 – 220.

[68] J. F. R. Olfati-Saber and R. Murray, "Consensus and cooperation in networked multi-agent systems," in *Proceedings of the IEEE, 95(1)*, 2007, pp. 215 – 233.

[69] M. Benantar, "The internet public key infrastructure," in *IBM SYSTEMS JOURNAL*, vol. 40, no. 3, 2001, pp. 648 – 665.

[70] R. A. Mollin, "An introduction to cryptography." CRC Press LLC, August 2000.

[71] P. Schwartz, "Finding your perfect match." Perigee Trade, January 2006.

[72] M. P. D. G. Liu, H., "Unraveling the taste fabric of social networks," in *International Journal on Semantic Web and Information Systems*, vol. 2, no. 1, 2006, pp. 42 – 71.

[73] V. R. C. Wen-tau Yih, Joshua Goodman, "Finding advertising keywords on web pages," in *Proceedings of the 15th international conference on World Wide Web.* New York, NY, USA: ACM, 2006, pp. 213 – 222.

[74] A. Maedche and S. Staab, "Ontology learning for the semantic web," in *Intelligent Systems, IEEE*, vol. 12, no. 2, Hongkong, China, 2001, pp. 72 – 79.

[75] J. B. W.A. Gruver, "Intelligent manufacturing: Programming environments for cim." London, UK: Springer-Verlag, December 1993.

[76] P. O. Wennerberg, "Ontology based knowledge discovery in social networks," 2005.

[77] S. V. J. Li, "Grid resource discovery based on semantic p2p communities," in *21st ACM Syposium on Applied Computing (SAC)*, 2006, pp. 754 – 759.

[78] R. R. S. O. Lassila, "W3c resource description framework (rdf) model and syntax specification," 1999.

[79] "Owl web ontology language overview," in *W3C Recommendation*, Feb 10th 2004. [Online]. Available: http://www.w3.org/TR/owl-features/

[80] P. C. Ivan Cantador, "Multilayered semantic social network modeling by ontology-based user profiles clustering: Application to collaborative filtering," in *Proceedings of the 15th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2006)*. Podebrady, Czech Republic: Springer Verlag, 2006, lectures Notes in Artificial Intelligence.

[81] J. H. O. Lassila T. Berners-Lee, "The semantic web," in *Scientific American*, vol. 184, no. 5, 2001, pp. 34 – 43.

[82] P. Mika, "Flink: Semantic web technology for the extraction and analysis of social networks," in *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 3, no. 2 - 3, October 2005, pp. 211 – 223.

[83] ——, "Ontologies are us: A unified model of social networks and semantics," in *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 5, no. 1, March 2007, pp. 5 – 15.

[84] D. C. W. H. H. Waleed Aljandal, Vikas Bahirwani, "Ontology-aware classification and association rule mining for interest and link prediction in social networks," in *Proceedings of the AAAI 2009 Spring Symposium on Social Semantic Web: Where Web 2.0 meets Web 3.0*, 2009, pp. 3 – 8.

[85] T. N. H. T. Masahiro Hamasaki, Yutaka Matsuo, "Ontology extraction by collaborative tagging with social networking," in *WWW2008*, Beijing, China, April 2008.

[86] A. T. Dario Bottazzi, Rebecca Montanari, "Context-aware middleware for anytime, anywhere social networks," in *IEEE Computer Society*, September 2007, pp. 23 – 32.

[87] "Foaf project." [Online]. Available: http://www.foaf-project.org/docs

[88] L. M. Dan Brickley, "Foaf vocabulary specification 0.98," August 2010. [Online]. Available: http://xmlns.com/foaf/spec/

[89] A. D. Elie Raad, Richard Chbeir, "User profile matching in social networks," in *IEEE Computer Society, in 2010 13th International Conference on Network-Based Information Systems*, 2010, pp. 297 – 304.

[90] V. L. Bengtson, "Inter-age perceptions and the generation gap," in *The Gerontologist*, vol. 11, no. 4,Pt.2, 1971, pp. 85 – 89.

[91] A. H. Eagly, "Sex differences in social behavior: A social-role interpretation." Hillsdale, NJ, London, UK: Lawrence Erlbaum Associates, December 1987.

[92] J. D. Mayer, "How mood influence cognition," in *Advances in Cognitive Science*, pp. 290 – 294.

[93] A. Budanitsky and G. Hirsta, "Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures," in *In Workshop on WordNet and Other Lexical Resources, Second meeting of the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, 2001.

[94] R. D. Smith, "Simulation article," in *Encyclopedia of Computer Science*, 1998. [Online]. Available: http://www.modelbenders.com/encyclopedia/encyclopedia.html

[95] "Netlogo," http://ccl.northwestern.edu/netlogo/.

[96] A. H. Muaz Niazi, "Agent-based tools for modeling and simulation of self-organization in peer-to-peer, ad hoc, and other complex networks," in *IEEE Communications Magazine*, March 2009, pp. 166 – 173.

[97] M. R. Jan Blom, Rene Quakkelaar, "Netlogo sql wrapper usermanual," July 2011.

[98] A. S. Yoav Sasson, David Cavin, "Probabilistic broadcast for flooding in wireless mobile ad hoc networks," in *In the proceedings of Wireless Communications and Networking Conference in 2003*, vol. 2, Switzerland, March 2003, pp. 1124 – 1130.

[99] J. D. B. Rodney A. Reynolds, Robert Woods, "Handbook of research on electronic surveys and measurements." Igi Global, 2007, pp. 136 – 143.