Aristotle University of Thessaloniki

Thesis Dissertation

A Comparative Study of Recommendation Algorithms in Location-based Social Networks

Author: Dimitrios Ntempos

Supervisor: Dr. Panagiotis Symeonidis

October 22, 2012
## Contents

I Introduction and motivation for the thesis 7

1 Introduction 8

II Social Networks, definitions and concepts 11

2 Online Social Network 12
   2.1 Dynamics 12
      2.1.1 Social Network physically 13
      2.1.2 Social Graph 13
      2.1.3 Gaining popularity 14
      2.1.4 Using embedded recommenders: 16
   2.2 Definitions 18
   2.3 Paradigms (Commercial OSNs) 20
      2.3.1 Facebook 20
      2.3.2 Twitter 22
   2.4 Step towards location 25
      2.4.1 Technology 25
      2.4.2 Culture 26
   2.5 The new structure 30
6 Comparison and Conclusions

6.1 Comparison

6.1.1 Framework and features

6.1.2 Experimental Evaluation

6.1.3 Aggregated Comparison

6.2 Conclusions
## List of Tables

<table>
<thead>
<tr>
<th>Number</th>
<th>Table Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Approved solutions</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>Users with different characteristics</td>
<td>17</td>
</tr>
<tr>
<td>2.3</td>
<td>Generations segmentation</td>
<td>27</td>
</tr>
<tr>
<td>2.4</td>
<td>Facebook VS Foursquare</td>
<td>30</td>
</tr>
<tr>
<td>3.1</td>
<td>Similarity extraction based on Location history</td>
<td>53</td>
</tr>
<tr>
<td>3.2</td>
<td>Similarity extraction based on Location semantics</td>
<td>54</td>
</tr>
<tr>
<td>6.1</td>
<td>Generic/Personalized categorization</td>
<td>83</td>
</tr>
<tr>
<td>6.2</td>
<td>Recommendation type</td>
<td>84</td>
</tr>
<tr>
<td>6.3</td>
<td>Data representation</td>
<td>85</td>
</tr>
<tr>
<td>6.4</td>
<td>Algorithmic type</td>
<td>85</td>
</tr>
<tr>
<td>6.5</td>
<td>Evaluation of Location Recommendation over different datasets</td>
<td>90</td>
</tr>
<tr>
<td>6.6</td>
<td>Evaluation of Activity Recommendation over different datasets</td>
<td>90</td>
</tr>
<tr>
<td>6.7</td>
<td>Optimal Algorithm based on framework and features</td>
<td>93</td>
</tr>
<tr>
<td>6.8</td>
<td>Optimal Algorithm - Using the 164 Dataset</td>
<td>93</td>
</tr>
<tr>
<td>6.9</td>
<td>Optimal Algorithm - Using the 164 Dataset</td>
<td>94</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>2.1</td>
<td>A simple Social Graph</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>Recommendation Process</td>
<td>17</td>
</tr>
<tr>
<td>2.3</td>
<td>Facebook’s GUI</td>
<td>23</td>
</tr>
<tr>
<td>2.4</td>
<td>Twitter’s GUI</td>
<td>24</td>
</tr>
<tr>
<td>3.1</td>
<td>Client/Server model representation</td>
<td>37</td>
</tr>
<tr>
<td>3.2</td>
<td>Mobile Foursquare</td>
<td>43</td>
</tr>
<tr>
<td>3.3</td>
<td>Newbie Badge on Foursquare</td>
<td>45</td>
</tr>
<tr>
<td>3.4</td>
<td>Photogenic Badge on Foursquare</td>
<td>45</td>
</tr>
<tr>
<td>3.5</td>
<td>Gowalla</td>
<td>46</td>
</tr>
<tr>
<td>3.6</td>
<td>Shopper’s eye GUI</td>
<td>47</td>
</tr>
<tr>
<td>3.7</td>
<td>Ubiquito’s GUI</td>
<td>48</td>
</tr>
<tr>
<td>3.8</td>
<td>LoCo Interface</td>
<td>48</td>
</tr>
<tr>
<td>3.9</td>
<td>Screenshots of CityVoyager</td>
<td>49</td>
</tr>
<tr>
<td>3.10</td>
<td>Foursquare users Dec’09-Mar’12</td>
<td>52</td>
</tr>
<tr>
<td>3.11</td>
<td>Users visited different places</td>
<td>53</td>
</tr>
<tr>
<td>4.1</td>
<td>Friend recommendation</td>
<td>67</td>
</tr>
<tr>
<td>4.2</td>
<td>Location recommendation</td>
<td>68</td>
</tr>
<tr>
<td>4.3</td>
<td>Activity recommendation</td>
<td>69</td>
</tr>
<tr>
<td>4.4</td>
<td>Event recommendation</td>
<td>69</td>
</tr>
<tr>
<td>5.1</td>
<td>Matrix</td>
<td>72</td>
</tr>
<tr>
<td>5.2</td>
<td>Tensor</td>
<td>73</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.3</td>
<td>Graph approach</td>
<td>74</td>
</tr>
<tr>
<td>5.4</td>
<td>HGSM system architecture</td>
<td>75</td>
</tr>
<tr>
<td>5.5</td>
<td>TBHG in a recommender</td>
<td>76</td>
</tr>
<tr>
<td>5.6</td>
<td>SLH-MTM</td>
<td>77</td>
</tr>
<tr>
<td>6.1</td>
<td>Data representation and Algorithmic type</td>
<td>86</td>
</tr>
<tr>
<td>6.2</td>
<td>Personalization and Types of Recommendation</td>
<td>86</td>
</tr>
<tr>
<td>6.3</td>
<td>Optimal Classification</td>
<td>95</td>
</tr>
</tbody>
</table>
Part I

Introduction and motivation for the thesis
Chapter 1

Introduction

It might be about 10 years since Social Networking over the Internet came into broader view. Since then, millions of people have been consuming great portions of time and energy on that. Apparently, this extraordinary amount of users led researchers to investigate several folds and contribute remarkable work concerning the field.

At the same time, technological progressions enabled the incorporation of Recommenders [1] and Geo-Location data in the traditional Social Networking structure. That is, the rise of Location-based Social Networks became reality. Hence, the research community started to work, now, on the direction of Location-based recommendation and there is also remarkable work concerning this discipline.

Although researchers have proposed several approaches to achieve the objective, there is not a single contribution that evaluates them in total. Even if there is Collaborative Filtering and Clustering algorithms to extract similarity or Tensor and Matrix-based approaches for the data representation, there is not a clear state that qualifies the optimal solution. Eventually, this is the greatest motivation for the following thesis. That is, this work tries to answer the upcoming question.
“Which is the optimal solution for the recommendation process in a Location-based Social Network?”

It might be true that the tremendous amounts of Data, moving in the web, cannot be handled, easily, by every-day people. In addition, Geo-Location data seems to serve as the physical dimension that web lacks, up to present. In fact, the youngest generations have never lived without the internet.

Hence, it is considered a challenge to study a field which (1) Handles Data, (2) Incorporates physical information where there is lack, (3) It’s probably accepted by the youngest generations of people and (4) It can reap the fruits that Online Social Networks have generated over the recent years.

Eventually, the future seems to litter great expectations for the field. Might Location-based Social Networks become the next best thing of the Internet industry? Time will tell. In the following, there is a brief introduction to the modules of the present thesis.

Chapter 2 - Online Social Networks

The chapter records an overall literature review concerning Social Networks and presents commercial Social networks i.e. Facebook and Twitter. Later on, there is an introduction of Location as an additional dimension in the traditional Online Social Networking structure.

Chapter 3 - Location-based Social Networks

Having presented the broader concept of Social Networks, this chapter deepens in the field of Location-based Social Networks. Author argues that Location serves reality, bridging cyber world with real life. In addition, there is an in-depth exhibition of technological tools for the development of an LBSN system. Foursquare as

1 http://www.facebook.com/
2 https://twitter.com/
3 https://foursquare.com/
well as other projects presentation aims the visualization of LBSN services and the social and economic report investigates the field through a different prism.

**Chapter 4 - Recommendation Framework**

The chapter initiates reader to the recommendation process addressing dynamics for the LBSN field. In particular, the thesis presents studied demographics of a real network and considers approaches that use Location in order to provide recommendation, as well as the different types of recommendation in Location-based social networks.

**Chapter 5 - Recommendation Algorithms**

Following the presentation of frameworks, as well as the use of Location for the recommendation process, this chapter deepens in the Algorithmic area. In particular, it records data modeling frameworks and the State of the Art algorithms, presented in research work from 2008 to 2012.

**Chapter 6 - Comparison and Conclusions**

Having introduced the State of the Art approaches, in the following, this thesis presents comparisons using tables to categorize the algorithms, as well as figures to visualise this categorization. Eventually, a comparison, between already cited approaches is giving rise to discussions and prospective work apropos of the LBSN field.
Part II

Social Networks, definitions and concepts
Chapter 2

Online Social Network

This chapter records a brief literature review, definitions for Social networks, commercial OSN presentations and concludes presenting the transition towards location as an additional OSN dimension. The questions to be answered are "what is a social network?" and "why is the number of Online Social Network users increasing in that rapid pace?". The analysis of commercial OSNs helps reader to realize the huge potential that LBSNs have, based on the concepts that Online Social Networks introduced in recent years.

In section 2.4 the thesis presents the transition, from Online Social Network to Location-based. In the latter, a cultural annotation answers the second question properly. The last section the following chapter is used to initiate the reader in the core subject of the present thesis i.e. Location-based Social Networks.

2.1 Dynamics

In the following sections, we review some useful subjects in order to introduce Online Social Networks smoothly. The reader is transported from the physical to the digital level where the importance of recommenders for the field of Social Networks is introduced.
2.1.1 Social Network physically

In the physical world, a Social Network is a structure consisted of entities (i.e. people, companies etc.) and the ties between them. To visualize it “The aggregation of a person’s social relationships yield his social network”. In recent years, this structure has been delivered to the Internet, creating the core for social networking websites. Specifically, existing social networking services in the web, reenact and in some cases substitute a physical social network structure. To achieve that, they allow users to build their own profiles containing personal info such as age, hobbies or political and religious views. After the user has created a profile, she is able to connect with other entities that have also a profile in the same social networking service.

2.1.2 Social Graph

More than 50 years ago, researchers from different scientific fields were trying to visualize the ties between people in real world. In 1967 Milgram [39] coined the term small-world after the results of his experiment have pointed out that American society is characterized by short path lengths. More recently, in 1998, D.J.W. and S.H.S. [58] arrived to a conclusion that:

In many real-world networks the probability of a tie between two actors is much greater if the two actors in question have another mutual acquaintance, or several.

A personal Social Graph is a graphic representation of all social links that a person has. Overall, Social Graph is a drawing that plots the structure of interpersonal relations in a group situation and depicts all personal relations. This has been referred to as ”The global mapping of everybody and how they are related”.

Figure 2.1 represents a single social graph containing six nodes and the ties between them. Every node represents a person or or-
ganization. Virtually, it can be observed that user 6 is directly connected to user 4 but there are also users 3, 5, 2, 1 in his social graph. Based on the assumption confirmed by [58], that people tend to connect with others, similar to them, non-directly connected nodes are likely to build a connection. Thus, OSN providers are up and coming to perform actions such as recommendations, within a Social Network, based on the analysis of a user’s social-graph plot.

However, it is important to mention that before Online Social Networks, a graph plot like that was impossible to be captured. For the time being, the entire world social graph is still impossible to be captured, though OSN may be considered as a useful tool to achieve great results on that direction.

2.1.3 Gaining popularity

The great boom of social networking services observed in the end of 00’s, enabled to present several sources, reffering to the number of social network users at present and in the future. For instance, according to Radicati research group, social network users will reach 2.3 billion by 2016 [57]. In a more detailed approach, Wikipedia offers a list of all the Social networking sites and the number of their registered users nowadays [1]. The previous references, are highly valuable in terms of realizing the great power that social networks possess in modern societies.

But what pushes users to utilize such a service? In general, the

\[\text{http://en.wikipedia.org/wiki/List_of_social_networking_websites}\]
main reason behind an individual’s choice to use a service or product is to utilize it in order to solve her problem. To argue on that, we can observe some cases through history and get some useful information. The following table resumes some highly useful products/services that really changed human history giving solutions to unsolved problems of that time and thus a great motivation for people to use them.

<table>
<thead>
<tr>
<th>Product/Service</th>
<th>Problem</th>
<th>Date</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>Movement</td>
<td>1908</td>
<td>Approved</td>
</tr>
<tr>
<td>Telephone</td>
<td>Direct Communication</td>
<td>1876</td>
<td>Approved</td>
</tr>
<tr>
<td>Internet</td>
<td>Worldwidenetwork</td>
<td>1980</td>
<td>Approved</td>
</tr>
<tr>
<td>Social Networks</td>
<td>UnifiedCommunication</td>
<td>Late '90s</td>
<td>Approved</td>
</tr>
</tbody>
</table>

Table 2.1: Approved solutions

According to the previous table, social networks first appeared in late ’90s and solved the problem of unified communication. But what does unified communications mean? To answer this question we will quote a real life example.

"Imagine a man who has 300 peers and his connections provide a link to 30 other places throughout the whole planet".

How could he take advantage of using an OSN? It is a fact that, using one, the specific user can reach all of his 300 connections directly or indirectly through his PC monitor at a glance. This might be true, but what makes an OSN more competitive compared to a telephone or internet in general? Before answering, it should be determined that OSNs are based on the internet. Actually, OSN use the internet to provide people with a unified communications platform.

To the point, using an OSN users have online access in the multimedia content generated by their connections, any time. At the same time they use additional services such as live chat, messages,
video call etc. This is the key! Users inform their peers with a simple update procedure that enables friends to see the message anytime and anywhere. In conclusion, online social networks introduced a new way of communications integrating several factors such as illustration, sounds and instant messaging in the same graphical user interface. This new application inspired users and thus OSN gained a great amount of constant users until today.

2.1.4 Using embedded recommenders:

Recommenders are widely used by OSN providers in order to stimulate users to add more entities i.e. people or organizations, within their personal Social Graph. Specifically, recommender systems act like filters, trying to provide the right information to users and reduce the noise. Using specific algorithmic frameworks[1], providers are able to offer useful recommendations for every user based on her profile’s self-stated characteristics. Thereby, providers equip users with a strong motivation to augment their social ties and as a consequence to use this type of service.

Typically, a recommendation algorithm used by services like that, takes as input certain information about the user from her profile and conscript them to yield recommendations for new ties such as friends, companies, places etc. The following case represents deductively the action that an embedded recommender performs in order to yield a recommendation for user 1.

Figure 2.2 illustrates how the algorithm takes an input of three specific user characteristics i.e. 1.Studies, 2.Music preferences and 3.Travelling habits and tries to diagnose the similarity of these three users. Similarity is inferred based on specific user’s characteristics and is used in order to provide users with a friend recommendation. The following similarity table represents deductively the algorithmic process in order to output a friend recommendation for user 1.
The ✓ symbol depicts similarity match and thus qualifies the user to be recommended while ✗ indicates none similar characteristics (mismatch).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>User 1</th>
<th>User 2</th>
<th>User 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies</td>
<td>Auth</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Music preferences</td>
<td>Rock</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Travelling habits</td>
<td>Yes</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>

Table 2.2: Users with different characteristics

Based on table 2.2 data, the algorithm will infer the similarity in terms of the number of matches. The more matches a user has the more similar she tends to be with User 1 (The user to whom the recommendation will be addressed). In this case, the recommendation algorithm outputs User 2 as the most appropriate user to be recommended to User 1.
In Conclusion, the role of recommenders is to motivate people enhance their social graph with new connections, by providing them with efficient suggestions in terms of recommendation. On the other hand, the goal is to augment the use of social networking service giving users a motivation to connect with new people or organizations similar to them.

2.2 Definitions

After the short presentation of some objects regarding OSN field, in this section, the thesis annotates given definitions for Social Networks. The following definitions, derive from different scientific disciplines. In fact, that reflects the multidimensional scientific physis of Social Networks.

There are several definitions for Social networks, originating from different scientific fields, for instance Danah Boyd at [6] reported: "We define social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site".

On the other hand Zheng et al. [69] stated for social networks that "A social network is a social structure made up of individuals connected by one or more specific types of interdependency, such as friendship, common interests, and shared knowledge. Generally, a social networking service builds on and reflects the real-life social networks among people through online platforms such as a website, providing ways for users to share ideas, activities, events, and interests over the Internet".

Back in 1997, Laura Garton et al [21] defined a social network as follows, "When a computer network connects people or organizations, it is a social network. Just as a computer network is a set of machines connected by a set of cables, a social network is a set of
people (or organizations or other social entities) connected by a set of social relationships, such as friendship, co-working or information exchange”.

According to Wikipedia[^1], a social networking service is an online service, platform, or site that focuses on facilitating the building of social networks or social relations among people who for example, share interests, activities, backgrounds or real-life connections. A social network service consists of a representation of each user (often a profile), his/her social links, and a variety of additional services. Most social network services are web-based and provide means for users to interact over the Internet, such as e-mail and instant messaging. Online community services are sometimes considered as a social network service, though in a broader sense, social network service usually means an individual-centered service whereas online community services are group-centered. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks.

Taking into account all the previous definitions combined with further research in the field of social networks, the following definition is proposed

An Online Social Network is an internet platform where people and organizations can create their own profiles containing several attributes such as age, hobbies, education and more. Profiles represent different entities in a social graph and could be considered as snapshots of real life delivered to the internet. Moreover entities interact to each other using several features such as photo sharing or creation of friendships, provided by social network administrators.

service is consisted of two basic elements i.e. people and ties. But how people can built those ties in a digital environment? To answer that, we are going to present two commercial OSN. The presentation aims to bridge the gap between academic definitions and real market implementation.

2.3 Paradigms (Commercial OSNs)

This section presents and briefly analyzes two commercial social networks. Except from visualizing real market, this section aims to depict the future potential of Location-based Social Networks, based on the results of their ancestors.

2.3.1 Facebook

Facebook, launched in February 2004, is the biggest social network nowadays. Although Facebook is not the core object in this thesis, it is presented as a commercial social networking service, in order to have a clear reconstruction of the biggest Online Social Network of 2012 and its features.

To the point, such as any other web platform, Facebook provides some features. It is clear that the combination of these features along with other ideas brought Facebook to the top.

The following description introduces some of the proven successful Facebook features.

Basic features of Facebook:

Profile creation This is the first and basic step in every social network, in order to access provided services by Social networks’ administrators. After a user fulfills her profile information, she has the option to add other attributes such as music preferences, favorite sports, education and more. The previous attributes reflect user’s tastes.
Connect This option enables user to search, find and send friend requests to other users or organizations in order to watch their updates in news feeds.

Update status This feature enables user to generate content which will be posted in the news feeds panel of her friends.

Chatting The instant messaging option enables Facebook users to interact with friends, directly, through a chat box.

Like Enables user to express her countenance for friends’ or organizations’ status updates. Like, is also used by people to follow organizations through liking their fun page. The latter feature enables user to get feeds provided by the organizations in the news feeds panel.

Event Creation Facebook users can create event pages and invite their friends. Moreover they can have a forecast of their events attendance, monitoring who will attend the event, who is about to attend and who is not attending at all. Facebook gives guests the option, either to attend, maybe attend or decline the event invitation.

Groups Every Facebook user is able to create a group and invite people in that. Groups could be open or closed for other users.

Facebook apps Facebook enables other applications to be integrated in it. For instance, a user could easily integrate his Twitter into Facebook. The previous integration enables Facebook and Twitter to communicate in a way that user generated content in Twitter could be automatically posted in Facebook.

Check-in This is a procedure related to the core subject of our thesis. However, check-in enables user to search nearby places and declare his presence there. This check-in action is visible to user’s friends.

Tag This is a feature that enables user to tag others, including
people and organisations, in his personal generated content. For instance, when a user uploads a picture, he can tag his friends in the picture.

**Photo Albums** It’s a feature helping people organize their uploaded photos.

**Other Facebook features**

Except from these Facebook provides users also with other types of features to boost interaction between them, such as poke, birthday calendar, friend or fun page recommendations. Moreover it offers users the possibility to create their own fun page and invite friends to like it.

**Graphical User Interface:** Analyzing the graphical user interface of Facebook, it contains 3 basic columns. In the right one user can watch groups, apps, pages messages and events. In the center column there is the news feeds provided by friends or liked fun pages. User has the option to either to like or share a connections news feed.

Furthermore, in the right column there is the birthday calendar, recommendations, upcoming events, app requests and sponsored ads for users. Finally at the right edge of Facebook’s GUI we can find the chat box, where users can monitor which connection is online or offline and send an instant message to it.

The following figure depicts Facebook’s graphical user interface inside a profile. The figure is cited in order to reflect the environment in which the user navigates during the use of his Facebook account.

**2.3.2 Twitter**

Twitter, created in March 2006, is another type of social networking service. It is actually classified into micro blogging category. By the time twitter is the second OSN in terms of users. Compared to
Facebook is less noisy. It provides users with only 3 basic features. The following description presents Twitter’s features.

**Profile creation** This is the basic step in every social network, in order to access provided services by Social networks’ administrators.

**Tweet action** Tweet is the micro blogging approach of user’s update in Facebook. Based on that, Twitter users can update their contribution to Twitter.com network using tweet function. This function enables user to update her status in a micro-blogging way, utilizing a maximum of 140 characters. Moreover she can add photos, links or locations in her tweets.

**Connect (follow)** This option enables user to search, find and follow other users or organizations in order to watch their updates in news feeds.

In the main page of Twitter, user can watch the tweets of people or organizations that she follows. Moreover she can perform interactive actions such as retweet, reply and #.

**Retweet action** This action enables user to duplicate tweets from other users. The duplicated tweet shows up in user’s tweets,
keeping the original generator as the one who tweeted and the
user, as the one who retweeted i.e. duplicated the tweet.

**Reply** On the other hand, "reply" is an action used to reply a tweet
by another user. The replied tweet is shown up in both users
news feeds so as other followers can see the source user and the
one who retweeted.

**#/Hashtag** The # symbol, called a hashtag, is used to mark key-
words or topics in a Tweet. It was created organically by Twit-
ter users as a way to categorize messages.

**Graphical User Interface:** Unlike Facebook, Twitter’s GUI is
composed by only two columns. In the left column Twitter presents
friend recommendations and trends in terms of hashtagged or simple
worlds. Analyzing trends they are outputted based on the whole
Twitter’s userbase. Moreover, in the right column user can monitor
the tweets generated by his peers. The following figure depicts the
GUI of Twitter inside a profile.

![Twitter’s GUI](image)

**Figure 2.4: Twitter’s GUI**
2.4 Step towards location

The fast growth of devices able to acquire geolocation data, using tools like GPS and/or WiFi networks, led traditional OSN providers like Facebook and Twitter to add Location as a new dimension in the existing OSN structure. Commercial, widely used Social Networks such as Facebook and Twitter added into their existing structure Facebook places Facebook places and location feature respectively.

On the contrary, research labs started to work on that new dimension and there is worth mentioning research work which is presented and analyzed later in Chapters 3, 4 and 5. In the following subsection, an insight to technological progressions that led the transition from Online Social Networks to Location-based is introduced. Later on, author quotes cultural characteristics, regarding people generations. The latter reference is cited in order to argue that new generations are concretely based on Social Networks. This clearly means, that the use of such services is going to grow even more popular in the future.

2.4.1 Technology

Technological progress leads people to new solutions concerning several fields. In the case of Online Social Networks, the rapid evolution of technologies able to acquire geolocation data lead the transition from traditional social networking services to Location-based ones. Along with the smartphone evolution and the integration of technologies such as 3G or 4G, Wi-Fi and GPS in it, geo-position data got easy to be acquired. As a result, the use of Location-based services got accessible by almost every individual owning a smartphone.

In this part we are going to briefly describe the previously men-
tioned technologies in order to have a representative list of the tools that are used for the geo-positioning procedure.

**3G-4G** Is a generation of standards for mobile phones and mobile telecommunication services. Application services include wide-area wireless voice telephone, mobile Internet access, video calls and mobile TV, all in a mobile environment.

**Wi-Fi** Is a popular technology that allows an electronic device to exchange data wirelessly (using radio waves) over a computer network, including high-speed Internet connections.

**GPS** Is a space-based satellite navigation system that provides location and time information in all weather, anywhere on or near the Earth, where there is an unobstructed line of sight to four or more GPS satellites. It is maintained by the United States government and is freely accessible to anyone with a GPS receiver.

In conclusion, these three technologies, enable user to have network access almost everywhere. Hence, Geo-Location data is easy to be acquired using one or a combination of them.

### 2.4.2 Culture

As time passes by, human habits change. This is guided mainly by the dynamic environment in which we exist as human beings. Technological progressions, environmental changes etc. lead people to change their beliefs and habits in order to adopt the evolution. As a logical sequence, young people act different compared to their forefathers. A strong argument on the latter, is the distribution of internet use between different age groups.

However, researchers have clustered people into categories, using geographic, demographic and behavioralistic criteria. Each category represents a generation in terms of age intervals. Moreover,
each generation has unique expectations, experiences, generational history, lifestyles, values, and demographics that influence their behaviors. The following table presents the generation segmentation according to Evans, Ahmad and Foxall [19] in terms of ages. The table is cited in order to present characteristics of the youngest generations and the interdependency that they have with the internet and thus tools, such as Online and Location-based Social Networks.

<table>
<thead>
<tr>
<th>Generation Distinctive Name</th>
<th>Born</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matures</td>
<td>Before 1946</td>
</tr>
<tr>
<td>Baby Boomers</td>
<td>1946-1964</td>
</tr>
<tr>
<td>Generation X</td>
<td>1965-1977</td>
</tr>
<tr>
<td>Generation Y</td>
<td>1977-1994</td>
</tr>
<tr>
<td>Generation Z</td>
<td>After 1994</td>
</tr>
</tbody>
</table>

Table 2.3: Generations segmentation

The following presentation aims to highlight the characteristics of generations Y and Z. It is cited in order to depict the tremendous potential that social networks present, based on the logical assumption that those people will shape the use of these services, over the years.
Generation Y According to Evans, Jamal, and Foxall [19] Generation Y has the following characteristics. They are children of the original Baby Boomers and their numbers rival those of the Baby Boomers. They grew up in a time of immense and fast-paced change including virtually full employment opportunities for women, dual-income households as the standard, wide array of family types seen as normal, significant respect for ethnic and cultural diversity including a heightened social awareness, and computers in the home and schools.

Gen Y individuals are well grounded and wise for their age. They were born into a technological, electronic and wireless society with global boundaries becoming more transparent. They are self-absorbed and self-reliant with a strong sense of independence and autonomy. They want results and are not as concerned with the why of it. They are image-driven and make personal statements with their image. They have a greater need for peer acceptance, connecting with their peers, fitting in, and social networking.

Gen Y individuals are open-minded, optimistic, goal oriented, and highly motivated toward their perceptions of success. Eight key values have been described for Gen Y: choice, customization, scrutiny, integrity, collaboration, speed, entertainment, and innovation. Efficient multi-tasking helps them be successful.

Generation Z Generation Z is the newest generation and for the time being, these individuals are in their early formative years. Their parents marry later and are less likely to get divorced. They face global terrorism, the aftermath of 9/11, school violence, economic uncertainty, recession, and the mortgage crisis. They continue to experience the spread of "tweendom" including commercial exploitation of young girls (and to a lesser ex-
tent boys), that is, pushing a Tween lifestyle heavy on teen aspiration to the cost of the loss of childhood.

In terms of characteristics, lifestyles, and attitudes, Generation Z individuals are the new conservatives embracing traditional beliefs, valuing the family unit, self-controlled, and more responsible. They are accustomed to high-tech and multiple information sources, with messages bombarding them from all sides. They have never lived without the Internet [19].

Generation Z values authenticity and "realness." Peer acceptance is very important to Generation Z, they need to belong. Their self-concept is partially determined by the group to which the Tween belongs. They are a global and diverse generation who come from a wider mix of backgrounds with different experiences and ideas. Generation Z values security more than ever. They are ready to be on mission, confident, and very optimistic. They believe that they can impact the world and can visualize changing places with someone else and can project possible behaviors. They quite possibly are the most imaginative generation and they think more laterally.

Resuming the above mentioned about Generations Y and Z, they are categorized as internet-based generations. They also tend to be social, thus online social networks are a great, digital, place for them to perform their actions.

Summarizing the cultural annotation it is worth mentioning that Gens Y and Z will actually shape the development of such services. Conversely, OSNs will follow the development of those Gens in order to conserve the most multitudinous target group that have been accumulated in recent years.
2.5 The new structure

Although traditional OSNs are successful, an LBSN goes a step ahead. Adding location to existing dimensions, such as hobbies, work, and education not only stimulates users to deal with that, but also helps providers in terms of recommendation, making them more precise and efficient.

On the other hand traditional GUIs considered aged enough for LBSNs’s structure, thus a renegotiation of traditional interfaces takes place in the new structure. The need of a map presence in the Graphical User Interface of the LBSN is considered as a necessity, in order for users to visualize their movements in the geo-locational dimension. In General, geo-social networking co-opts Internet mapping services to organize user participation around geographic features and their attributes.

Before entering the next chapter, an overview of the deductive comparison between two commercial social networks will be given. The following comparison is introduced, in order for the reader to have a representation of the differences between Traditional(Online) and LocationBased Social Networks. The comparison is based on certain features in the biggest commercial Social Networks of each category. Namely, Facebook(OSN) and Foursquare in (LBSN).

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Facebook</th>
<th>Foursquare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check-in</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Photos</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Albums</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Chat</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Recommendation</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Multimedia</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Gamification</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Apps</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>

Table 2.4: Facebook VS Foursquare
• Check in: A feature that enables user to indicate her presence in a specific geo-place.
• Photos: The option for users to upload a simple picture.
• Albums: A way to organize uploaded pictures.
• Chat: An option that enables user to send instant messages through a chat box.
• Recommendation: Suggestions of other users, organizations or places, provided by the social networks administrators.
• Multimedia: The option that enables user to upload other multimedia content, except from pictures.
• Gamification: A technique that uses game concepts and elements. That motivates user and thus, enhances the use of the social network.
• Apps: A feature that enables other applications to be integrated within the social network.
Chapter 3

Location-based Social Networks

Having presented the broader concept of Online Social Networks, this chapter introduces the main idea of Location-based Social Networks. To begin with, it is important to point out that Location-based Social Networks are grouped in the greater field of OSN. Actually, an LBSN uses OSN’s concepts, but qualifies Location as the core object of its structure.

In detail, the chapter initially answers the ”why location?” question. In section 3.1 there are definitions and an in depth exhibition of technology used in LBSN field. In 3.2 a brief report about two commercial LBSNs and some Location-based projects is presented. In 3.4 the economic and social annotation regarding LBSN aims to investigate the field under a different, more market oriented prism. Finally, section 3.5 introduces the recommendation concept in LBSN and that’s actually a prerequisite for part III of the thesis.
Why LOCATION? It is clear that the new structure, officially called Location-based Social Network is able to engage dwellers and companies, yielding key results and changing the way of how people act in their everyday life. Location, the newly brought dimension, is able to serve specific needs that are required to reduce web noise. Using location as INPUT to recommendation algorithms, providers could achieve highly effective recommendations resulting to higher engagement and save of time.

But, why location? One could easily come with plenty of other characteristics. These are gender, hobbies, political views, events an individual joined etc. That might be true! The fact is that all previous characteristics, including location, are being considered by recommendation algorithms as input. The more realistic they are, the most efficient the output will be.

Moreover, recalling the previous discussions about cyber-world where an individual has to update his profile every time a change is happening in real life, it is worth-mentioning how location bridges the gap between this and the real world. The following example illustrates a case where Location exceeds other characteristics regarding as main parameter, real life changes.

Example Imagine a man who made a profile on “Loca-Tio” LBSN. At first he provided all his personal info such as age, marital status, hobbies and favorite music. At that time, “Loca-Tio” could provide him with useful recommendations without the use of Location. That might be acceptable. The fact is that it does not take under consideration the instability of human life.

Imagine the same man, three months later when he is divorced, he stopped playing softball and he is bored of Jazz music. At that situation, the only stable characteristic to take
as INPUT and finally provide him with an effective recommendation should be location. Shouldn’t it?

To sum up, Location is a new dimension added in traditional OSN structure that serves reality in cyber-world terms. It is worth mentioning that, in most of the cases, location exceeds other characteristics provided by an individual when she first made her profile.

Indeed, Location is not a user generated attribute. It could thus considered as the most realistic input. Each time an individual checks-in a place, she reveals main interests. It is clear that someone who checks-in every day in a basketball court, she is a basketball lover. Thus, she gives the provider an un-debatable reason to recommend her other basketball lovers for friends or basketball courts as Points of Interest.

3.1 Definitions

Several definitions concerning Location-based Social Networks or Geo-location Social Networks (GSNs, as they are reported in some cases) are found in bibliography.

Li reported “Location-based Social Networks (LSNs) allow users to see where their friends are, to search location-tagged content within their social graph, and to meet others nearby.”

Zheng who is one of the most evolved researchers in LSNs field worldwide defines in his book a Location Based Social Networks as following: ”A location-based social network (LBSN) does not only mean adding a location to an existing social network so that people in the social structure can share location embedded information, but also consists of the new social structure made up of individuals connected by the interdependency derived from their locations in the physical world as well as their location-tagged media content, such as photos, video, and texts. Here, the physical location con-
sists of the instant location of an individual at a given timestamp and the location history that an individual has accumulated in a certain period. Further, the interdependency includes not only that two persons co-occur in the same physical location or share similar location histories but also the knowledge, e.g., common interests, behavior, and activities, inferred from an individual’s location (history) and location-tagged data.”

Another definition springs from Symeonidis et al.\cite{42} reports the following: ”GSNs allow users to use their GPS-enabled device, to ”check in” at various locations and record their experience. In particular, users submit ratings or personal comments for the location/activity they visited/ performed. That is, they ”check-in” at various places, to publish their location online and see where their friends are. Moreover, they can either comment on a friend’s location or comment on their own”.

According to Wikipedia\footnote{http://en.wikipedia.org/wiki/Geosocial_networking}, Geosocial Networking is a type of social networking in which geographic services and capabilities such as geocoding and geo-tagging are used to enable additional social dynamics. User-submitted location data or geo-location techniques can allow social networks to connect and coordinate users with local people or events that match their interests. Geo-location on web-based social network services can be IP-based or use hotspot trilateration. For mobile social networks, texted location information or mobile phone tracking can enable location-based services to enrich social networking.

In a more general basis, Koeppel\footnote{31} reported about location based services: Location services enable customized information to be delivered or made available based on the specific location of the user. Knowing where the user is at any given time adds a valuable dimension to the kinds of services that can be offered.
Finally, based on all cited definitions and on the knowledge accumulated during the composition of this thesis author quotes the following definition for Location-based Social Networks:

A Location-based Social Network is an internet based social structure. An LBSN exploits geolocation information and qualifies the interdependency of users based on it. That is clearly a step towards the connection between the physical and the cyber world.

3.2 Technology and Tools

Location based social networks are services. These services are clicked on a server in order for users to reach the service and interact not only with it, but also with other users, using the same service. This part presents the models and tools that are used for the development of a location based service, starting from the architectural model and concluding to the database system which is used to store information.

3.2.1 Architecture

Client/Server model: The client/server model is a computing model that acts as distributed application which partitions tasks or workloads between the providers of a resource or service, called servers, and service requesters, called clients.

Clients and servers communicate over a computer network on separate hardware, but both client and server may reside in the same system. A server machine is a host that is running one or more server programs, such as Apache server which share their resources with clients. A client does not share any of its resources, but requests a server’s content or service function. Clients therefore

\footnotesize{urlhttp://httpd.apache.org/}
initiate communication sessions with servers which await incoming requests.

The following figure illustrates the client/server model. In our case, LBSN applications run on the client side.

![Client/Server model representation](image)

**Figure 3.1: Client/Server model representation**

### 3.2.2 Programming tools

The tools that could be used in order to develop a Location-based Social Network application are various. The following part is divided into different sub-parts in order to present them. Specifically, we separate web from mobile development are separated. The introduction of representative tools, aims to illustrate how an LBSN could be developed both in web and mobile environments.

**Web programming tools**

In this part, technologies that can be used for the development of an LBSN in a web environment are listed.

**HTML**

It is used for the interface of Location-based social networks. It is the main markup language for displaying web pages in a web browser. HTML \[16\] is written in the form of HTML elements consisting of tags enclosed in angle brackets. The tags most commonly come in pairs like `<h1>` and `</h1>`.

Between these tags developers can add text as well as other text-based content. Regarding web browsers, the purpose of such an
application is to read the HTML code and compose it into visible web page. Technically, browsers do not display HTML tags, but use them to interpret their content.

As far as design is concerning, CSS is a style sheet language that is used to style HTML pages. CSS can be used to make the appearance and format of an LBSN application more organized.

**PHP**

This technology is used to develop specific features in web environments. In particular, it is a general-purpose server-side scripting language originally designed for Web development to produce dynamic Web pages. It is one of the first developed server-side scripting languages to be embedded into HTML source code.

Ultimately, the code is interpreted by a Web server with a PHP processor module which generates the resulting Web page. It also has evolved to include a command-line interface capability and can be used in standalone graphical applications.

**Javascript**

Sometimes abbreviated as JS, is a prototype-based, dynamic scripting language. It is also weakly typed and has first-class functions. Being multi-paradigm language, it supports several programming styles.

JavaScript is used in order to deliver a GUI, that matches the needs of an LBSN system. Javascript is also used along with HTML, PHP and map application APIs, such as Google maps API for Location-based applications’ interfaces.

[https://developers.google.com/maps/location-based-apps](https://developers.google.com/maps/location-based-apps)
Mobile programming tools

The appropriate technologies for mobile development are presented below briefly. In the attempt to cover almost 100% of the market, this part consists of the three dominant mobile OS in 2012 i.e. Android, Apple IOS and Windows Phone.

Android Android is a Linux-based operating system, currently owned by Google, primarily designed for mobile devices such as smartphones and tablet computers utilizing ARM processors. Instead of its previous use, the light weight OS targets embedded systems such as networking equipment and various other devices such as house hold appliances and wrist watches. The following technologies can be used in order to develop a native Location-based Android application.

XML

It is used for the GUI development in Android OS. XML stands for Extensible Markup Language and it is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. It is defined by the XML 1.0 specification, produced by the W3C and several other related specifications.

The design goals of XML is to emphasize simplicity, generality, and usability over the Internet. It also provides strong support via Unicode for the languages of the world. Although the design of XML focuses on documents, it is widely used for the representation of arbitrary data structures, for example in web services.
Java

It is used to develop application’s functions in Android OS. Java is a general-purpose, concurrent, class-based, object-oriented language that is specifically designed to have as few implementation dependencies as possible [12]. It is intended to let application developers "write once, run anywhere".

Java is currently one of the most popular programming languages in use, particularly for client-server web applications, with a reported 10 million users. Google provides Android developers with SDK[7] which is used to generate code for several Android devices. Android currently provides 4.1 version.

Apple IOS [8] is a mobile operating system developed and distributed by Apple Inc.[9]. Originally released in 2007 for the iPhone and iPod Touch. Later on, IOS has been extended to support other Apple devices such as the iPad and Apple TV. Unlike Microsoft’s Windows and Google, Apple does not license iOS for installation on non-Apple hardware.

Objective-C

It is the core technology used for the development of IOS applications. Objective-C [17] is a general-purpose, high-level, object-oriented programming language that adds Smalltalk-style messaging to the C programming language. Apple provides Cocoa touch API, for the IOS development.

Windows Phone Launched in the second half of 2010, Windows Phone is a mobile operating system developed by Microsoft. It is the successor to its Windows Mobile platform. Windows phone OS is primarily aimed at the B2C rather than B2B market. In the following part, we record the technologies for the development of WP applications.

XAML

It is used for the Graphical User Interface in Windows Phone OS. Moreover, XAML is a declarative XML-based language created by Microsoft and is used for initializing structured values and objects. It is available under Microsoft’s Open Specification Promise. The acronym originally stood for Extensible Avalon Markup Language where Avalon is the code name for Windows Presentation Foundation (WPF).

C#

It is used to develop application’s functions in Windows Phone OS. Furthermore, C# is a multi-paradigm programming language, encompassing strong typing, imperative, declarative, functional, generic, object-oriented (class-based), and component-oriented programming disciplines. It was developed by Microsoft within its .NET initiative and later approved as a standard by Ecma (ECMA-334) and ISO (ISO/IEC 23270:2006).

C# is one of the programming languages designed for the Common Language Infrastructure. C# is intended to be a simple, modern, general-purpose, object-oriented programming language. Its development team is led by Anders Hejlsberg. The most recent version is C# 4.0, which was released on April 12, 2010.
3.2.3 Database system

A database is an organized collection of data, today typically in digital form. In LBSNs’ case, databases are used in order to store users, locations and other contextual information. The data are typically organized to model relevant aspects of reality i.e. the number of check-ins in a place, in a way that supports processes requiring this information i.e. finding the place with the most check-ins.

The term “database system” implies that the data is managed to a certain level of quality (measured in terms of accuracy, availability, usability, and resilience) and this in turn often implies the use of a general-purpose database management system (DBMS). A general-purpose DBMS is typically a complex software system that meets many usage requirements. The databases that it maintains are often large and complex.

The utilization of databases is now so widespread that virtually every technology and product relies on it, for its development and commercialization. Actually, the vast majority of organizations and companies, from small to large, depend heavily on databases for their operations.

Well known DBMSs include Oracle, IBM DB2, Microsoft SQL Server, Microsoft Access, PostgreSQL, MySQL, and SQLite. A database is not generally portable across different DBMS. However, different DBMSs can inter-operate to some degree by using standards like SQL and ODBC together to support a single application. A DBMS is also expected to provide effective runtime execution to properly support (e.g., in terms of performance, availability, and security) as many end-users as needed.

15 http://www-01.ibm.com/software/data/db2/
17 http://office.microsoft.com/en-us/access/
18 http://www.postgresql.org/
19 http://www.mysql.com/
20 http://www.sqlite.org/
3.3 Paradigms

The following section aims to bridge the gap between research definitions and real market presenting real world paradigms and relative projects.

3.3.1 Foursquare

Foursquare[^21] is a location-based social networking website for mobile devices. Users "check-in" at venues using a mobile application by selecting from a list of main venues that are as close as possible to their actual location. Location is based on GPS hardware in the mobile device or network location provided by the application. Each check-in awards the user with points and sometimes with special "badges". This is a part of gamification process[^14] in order to motivate users to augment their check-in behavior.

The service was created in 2009. By April 2012, the company reported it had 20 million registered users and was expected to pass 750 millions of check-ins before the end of June 2011, with an average of about 3 millions of check-ins per day.

[^21]: http://foursquare.com/
Regarding Foursquare’s demographics, male and female users are equally represented and around 50% of its users are outside the US. Support for French, Italian, German, Spanish, and Japanese languages was added in February 2011 in order to motivate non-US users to use the service. In the same direction, support for Indonesian, Korean, Portuguese, Russian, and Thai was added in September 2011.

**Basic features of Foursquare**

In the following part, we present Foursquare’s features and concepts. Compared to traditional commercial Online Social Networks such as Facebook and Twitter, Foursquare qualifies fewer options for the user. Specifically it provides the following basic options:

**Profile creation** This is the basic step in every social network, in order to access provided services by Social networks’ administrators.

**Check-in** Is the process of announcing the arrival in a certain place in Location Based Social Networks. Users can also leave tips for the places that they have checked-in.

**Connect** This option enables user to search, find and connect with other users in order to watch their profiles, check-ins and tips.

**Recommendation** This is the process of recommending venues and friends.

**Gamification in Foursquare**

Except from the previously mentioned features, Foursquare uses gamification methods to motivate users to check-in more and thus increase the use of the service. In particular, it uses the following game elements in order to offer users, a unified game experience.
Paradigms  

Location-based Social Networks

**Ranking** When a user checks-in in a place she is awarded with points. The number of points depends on several factors such as mayorships or "first of friends that checked in the place". In addition, there is a relative ranking in Foursquare’s application, depending on the points that friends of the user have accumulated.

**Badges** It’s a kind of prize given to users under specific requirements. For example:

![Newbie Badge](image)

**Newbie Badge:**
Every user is awarded with the newbie badge after her first Check-in.

![Photogenic Badge](image)

**Photogenic Badge:**
Every user that checks-in more than two times in a venue that has a photo booth is being awarded with the photogenic badge.

**Mayorships** A user is awarded as the mayor of a place, when she is the user with the highest number of check-ins in the specific place.
3.3.2 Gowalla (Closed in 2012)

Gowalla was a primer mobile web application that allowed users to check in to locations that they visited using their mobile device. This was done either through the use of dedicated applications available on Google Android, iPhone, Palm Web OS and BlackBerry, or via the service’s mobile website. Check-ins could be forwarded via Notifications to iPhones, and by linking accounts, to Twitter and Facebook.

Despite the big expectations that Gowalla brought after its launch and the acquisition by Facebook during 2011, it was closed by 2012. Figure 4.5 illustrates a screen of Gowalla’s mobile GUI. There are several opinions about the reasons that led Facebook to close Gowalla. It is possible for instance that Gowalla was not user-friendly. That’s clearly an important reason for an internet service to be closed nowadays.

3.3.3 Projects

Apart from commercialized LBSN many other academic or market projects have emerged over the years. Back in 1998 Shoppers eye[20] appears one of the first Location Based projects. In particular Shopper’s eye was a PDA based, GPS-enabled agent that relies on the knowledge of a shopper’s physical location to support...
her shopping activity while shopping at a mall. Shoppers indicate
their shopping goals to Shopper’s Eye. Then, as shoppers stroll
through a mall, Shopper’s Eye informs them of the availability of
items of their interest available in stores around them. It is even
able to indicate alternative items with lower price. The following
figure illustrates Shopper’s eye GUI.

![Shopper’s eye GUI](image)

Figure 3.6: Shopper’s eye GUI

Another worth mentioning project is iCITY[9] which is a descen-
dant of Ubiquito[2]. iCITY is a social adaptive mobile guide that
uses semantic web technologies in a web 2.0 vision in order to pro-
vide cultural events of the city of Torino. In fact, the recommender
of iCITY provides personalized recommendations using data stored
in the database system, parsed from TorinoCultura[22], a web-site for
social events of Torino city and the current location of the user.

[22] http://www.torinocultura.it/
Furthermore, an approach named I’m Feeling LoCo\cite{53} presented by Savage et al, proposes a system that learns user preferences by mining a person’s social network profile. The physical constraints are delimited by the user’s location and mean of transport, which are automatically detected through the use of a decision tree followed by a discrete Hidden Markov Mode. The novelty of this approach relies on the fusion of information derived from a user’s social network profile and her mobile phone’s sensors for place discovery.
In contrast, CityVoyager by Yuichiro Takeuchi and Masanori Sugimoto \(^{[65]}\), introduces a novel real-world recommendation system, which makes recommendations of shops based on users’ past location data history. This system uses a place learning algorithm, which can efficiently find users’ most visited places, complete with their proper names. Users’ most visited shops are used as input to the item-based collaborative filtering algorithm so as to make recommendations. In addition, researchers provide a method for further narrowing down shops based on prediction of user movement and geographical conditions of the city.

As the LBSN sector accumulates more users several notable proposals, targeted to the real market appear. A good example, Closeklik which was presented at MOBIP\(^{[23]}\) by a Greek startup named ”Brainbow Development”\(^{[24]}\). Closeklik provides users with real-

\(^{[23]}\) http://www.europe-innova.eu/web/guest/innovation-in-services/kis-innovation-platform/mobip/about;jsessionid=97B83EE7C9D91D840000FD4832D96456

\(^{[24]}\) http://www.brainbowdevelopment.com/index_en.html
time information via photos about places and events around their current locations (varying from few blocks away to alongside regions). This offers them a notable mobile live guide and giving them an “as if i was there myself” feeling. The user, whose location is automatically detected, either uploads the photos that he instantly takes online or automatically views other real-time photos, uploaded by other users, regarding his nearby locations.

Closeklik is designed to contain outdoor photos related to weather conditions, urban events (e.g. street art, demonstrations), traffic conditions, public emergencies (e.g. fire), landscapes, entertainment guide services (restaurants, cafes, bars, clubs) and more. Closeklik developers, estimate that ”In combination with location-based services’ boost, Closeklik will be a breakthrough in mobile users’ decision-making attitude and information sharing”.

3.4 Social and economic report

In recent times, Online and Location-based Social Networks have increased massively their number of users. This increase was even powerful to influence several social episodes around the globe, such as the Egyptian Revolution against Mubarak’s government[25]. This social phenomenon is basically caused by the great power that Social Networks hold on unified communication levels. However, they exist significant paradigms that should be considered, in the future, as revolutionary for the humanity as well. For instance Sakaki et al.[52] proposed an approach, able to inform users for an earthquake. The novelty of that approach stands on the speed that a new earthquake is being spread, using Twitter platform. In the same direction De Longueville et al.[11] exploited info from an LBSN to acquire spatio-temporal data on forest fires.

Regarding the economical aspect of the subject, monetization for Social Networks can be achieved through B2B or B2C channels. On the one hand, the most representative model of B2B monetization remains Facebook adds. On the other hand, Linkedin premium accounts work on a B2C background, raising money from users able to get premium features on Linkedin network.

For the time being, however a big revenue stream for commercial Location-based Social Networks has not occured. Despite the big number of users evolved in Location-based services, a clear way to monetize this has not yet been found, thus providers concentrate in the processing of user’s experience. In fact, Foursquare, the biggest Location-based service nowadays has accumulated 25 million users according to Business Insider, but it has not still a strong enough revenue stream. Lately, Foursquare concluded to an agreement with American Express in order to let users access exclusive lists, tips, offers and experiences concerning the things they are interested in.

Apart from cooperations with other organizations, there also other ides for monetizing a Location-based experience, for instance Hsiehl et al. proposed a leisure guide that composes travelling paths from Location traces of an LBSN. The commercialization of these paths it is probably a feasible way to increase income for Location-based services.

Back to the social point of view, users of Foursquare ceratinly gain more and more, enforcing the opinion that Location-based Social Networks such as Foursquare will lead the internet over the upcoming years. To support the latter, it is likely that new generations i.e. Y and Z will be extremely based on social media for socializing.
Additionally, the emerging work on Location-based recommendation systems increases the efficiency of personalized marketing and as a result, the monetization sources. Finally, the increasing number of Foursquare users over the recent years, displayed by the following schema, clearly demonstrates the trend towards Location-based services.

![Figure 3.10: Foursquare users Dec’09-Mar’12](image)

### 3.5 Embedded recommenders

In order to be prepared to deepen, it is important to set some baselines for Location-based recommendation. Generally, embedded recommenders in Location-based Social Networks operate under the same framework as in traditional Online Social Networks. The main difference is that in Location-based services the basic algorithmic input in order to generate a recommendation is user’s Location or
Location history. Specifically, the recommendation algorithm is trying to extract similarity using either the Locations that users have been visited or the current user’s Location.

The next case points out deductively the process of recommendation in a Location-based Social Network. This recommendation algorithm get as input the Locations that a user visited, illustrated on the following map. In point of fact, the tables extract the similarity based on 1. The physical Location and 2. The Semantic meaning of the locations, in terms of matches. Each ✓ is counted as one match. The final recommendation concerning Red User includes Friend and Location recommendation.

![Figure 3.11: Users visited different places](image)

<table>
<thead>
<tr>
<th>Red User</th>
<th>Blue User</th>
<th>Black User</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antarctic</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>North America</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Australia</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

Table 3.1: Similarity extraction based on Location history

Regarding Location data as input, the number of matchings is eliminated to one. This match is acquired from the check-ins that
both the Red and the Blue user performed in the North-America. Thus, a simple Location-based recommender will recommend a friend connection between the Red user and the Blue one.

<table>
<thead>
<tr>
<th>Red User</th>
<th>Blue User</th>
<th>Black User</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pole</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>Near Alaska</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>Oceania</td>
<td>✔️</td>
<td>✗</td>
</tr>
</tbody>
</table>

Table 3.2: Similarity extraction based on Location semantics

Regarding semantic Location data as input, it is likely for the Red User to get a recommendation of Arctic and New Zealand as Points of Interest. That happens because of the semantic meanings assigned to the Locations that both Red user and Blue user checked-in. The semantic meaning inferes three matches i.e. Pole (Antarctic-Arctic), Oceania (Australia-New Zealand) and near Alaska (North America). It is important to be highlighted that semantics have been assigned randomly. Any other semantic assignment could have resulted different outputs.
Part III

Recommendation in Location-based Social Networks
Chapter 4

Recommendation Framework

This chapter introduces an approach of recommendation frameworks in Location-based Social Networks. In the beginning, reader meets dynamics concerning LBSN field. In 4.2, thesis presents demographics based on researches for the LBSN field. 4.3 considers approaches that use Location in order to provide a recommendation. Finally, the different types of recommendation in Location-based social networks are presented.

4.1 Dynamics

Dynamics section is cited in order to highlight several folds concerning LBSN field. In particular, the following presents the algorithmic types used for the recommendation process as well as the requirements for the Location-based recommendation algorithms. Eventually, the thesis considers important subjects concerning LBSN field i.e. Privacy and the use of semantics.
4.1.1 Algorithmic types

Recommendation, as previously revealed, is an important feature for a Social Network, in order to motivate users to increase the ties between them. To achieve that, researchers have employed several algorithmic types. The most common types, found in the bibliography, are listed in the following.

**Collaborative Filtering** The underlying assumption of the Collaborative Filtering is that if a person A has the same opinion as a person B on an issue, A is more likely to have B’s opinion on a different issue x than to have the opinion on x of a person chosen randomly. Applying the core hypothesis of Collaborative Filtering for an LBSN, researchers have proposed several approaches. We meet this approach in several frameworks. A representative work utilizing this type of algorithm for Location-based recommendation can be found in [66].

**Bayesian Networks** Is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph. A representative work uses Bayesian network theory in LBSN in order to provide recommendations in LBSN field is [43].

**Markov Chain** This is a mathematical system that undergoes transitions from one state to another, between a finite or countable number of possible states. It is a random process usually characterized as memoryless, the next state depends only on the current state and not on the sequence of events that preceded it. This specific kind of "memorylessness" is called the Markov property. Quercia at al. [47] proposed an approach where Markov chain algorithms can be used to yield recommendations for users.
**Clustering**
This is the task of assigning a set of objects into groups, called clusters. The main concern of these algorithms is to group the more similar objects in the same cluster. A representative paradigm of an algorithm belongs to this category is CADC algorithm [34].

**Context**
Context-based algorithms are exploiting mathematical models in order to use several dimensions in the recommendation process. Representative approaches can be found in [28] and [22].

**SVD**
The singular value decomposition (SVD) is a factorization of a simple or more complex matrix i.e Tensor. Traditionally, SVD has been used for several useful applications in signal processing and statistics. A prototype application concerning LBSN field can be found in the work of Symeonidis et al. [56] that proposes the Higher Order SVD for tensor dimensionality reduction.

**Ranking**
The Ranking approach is used either before or after the operation of some algorithmic frameworks in order to delimitate the proposed input for the recommendation algorithm [4] or filter its output after the operation. Unlike the traditional use of Ranking-based algorithms, Zheng et al. [67] lately, proposed an approach that incorporates Ranking in the the recommendation process. As regards the traditional use, it can be found also in the generic recommendation framework of Zheng et al. [66].

### 4.1.2 Algorithm Requirements

Researchers have proposed several algorithms, in order to provide efficient recommendations based on current location or location history of the user. Regarding requirements for these algorithms, Horozov et al. [25] recorded some logical inferences about specific characteristics of the recommenders. Based on Horozov at al. and logical
assumptions, this thesis presents the algorithmic requirements for a large scale LBSN.

**Scalability** LBSN community recently accumulates more users and thus more information. Hence, an algorithm used in such a system must scale to user and item populations, potentially numbering in the multi-millions.

**Latency** Typically, LBSN users are on the go. That means they carry a resource-limited device (less storage and real-estate for display). Hence recommendation results need to be chosen well and also need to be delivered quickly.

**Convenience** Location recommendation is addressed, mainly, to mobile users. Based on that it is more convenient for Location-based recommenders to qualify location rather than other dimensions considering as input. For instance, when a user hangs out and searches a restaurant, it is more convenient to recommend him a close one rather than the one which best matches his gastronomical preferences.

### 4.1.3 Privacy

Privacy, indeed, is an important object concerning LBSN field. The physis of a Location-based Social Network, imposes strong privacy barriers. The basic reason is that the imprinting of such a real information i.e. Location to the internet, raises questions for the safety of user’s in several dimensions\[55\] [45].

For that reason, there are several researches concerning Privacy. According to Puttaswamy and Zhao [46] the basic reason that makes LBS applications vulnerable, is the interdependency of such applications with untrusted third party servers. In their work, researchers argue that Location-based services should adapt an approach where the untrusted third-party servers are treated simply as encrypted...
data stores and the application functionality should be moved to the client devices.

Another approach presented by Quercia and Hailes[48], proposed a novel decentralized defence for mobile devices. The name of it is MobID. The idea is that a device manages two small networks, the friends’ and the foes’ one. The device stores information about the other devices it meets. Reasoning on that, the device can determine whether an unknown individual is carrying out a sybil attack or not.

In a different direction, Lee et al. [33] exploit Web 2.0 objects to protect users privacy. In this scheme, location semantics are first learned from Location data. Then, a trusted anonymization server performs the anonymization using the Location semantic information by cloaking with semantically heterogeneous locations. Thus, the location semantic information is kept secure as the cloaking is done with semantically heterogeneous Locations and the true Location information is not delivered to the LBSN applications.

4.1.4 Semantics

As well as for privacy concerns, semantics can be used also in other folds of an LBSN. Indeed, there have been several studies concerning this direction. The following, introduces semantic approaches that can be used within an LBSN service.

A representative approach that exploits Location semantics for an LBSN application can be found in the work of Xin Cao et al. [8]. This proposal, introduces a general framework for the mining of semantically meaningful, significant locations. In particular, researchers captured the relationships between locations and between locations and users with a graph. Significance is then assigned to locations using random walks over the graph. Using these technique they where able to mine semantically meaningful places in a Location-based service.

In similar direction, Ye et al. [63] proposed a semantic annota-
tion technique to automatically annotate all places with category tags which are crucial prerequisite for location search, recommendation services or data cleaning. The annotation algorithm learns a binary support vector machine (SVM) classifier for each tag in the tag space to support multi-label classification. Based on the check-in behavior of users, they extract features of places from i) explicit patterns (EP) of individual places and ii) implicit relatedness (IR) among similar places. The features extracted from EP are summarized from all check-ins at a specific place. The features from IR are derived by building a novel network of related places (NRP) where similar places are linked by virtual edges. Upon NRP, they determine the probability of a category tag for each place by exploring the relatedness of places.

4.2 Demographics

Back in 2009, Nan Li and Guanling Chen\cite{35} studied a real world/commercial Location-based Social Network in order to present demographics of LSN users. Results of that large quantitative research revealed that 73\% of the users were male and only 17\% were female. The latter is opposed to many other studies\cite{24,37}, as well as to real market data of Foursquare where male and female users are represented equally.

Regarding mobility characteristics Li and Chen\cite{35} have clustered users of the studied network in four categories, taking into consideration the location cluster from which they sent their location updates. The mobility characteristics of each cluster are listed in the following.
**Home User** No mobile users. They were 29.31% of whole active users.

**Home-Vacation User** Based on the observation that users sent at least 50% of their updates from one location cluster and no other cluster generated more than 20%. Results shown that 47.43% of users were in Home-Vacation category.

**Home-Work User** Users belong to these cluster sent most of their updates about equally from two location clusters. 11.73% of all active users belong to this group.

**Other User** Users that sent location updates from several location clusters. They were 11.53% of the whole and they cannot be grouped in one of the before mentioned categories.

Regarding future work in LBSN field, Noulas at al. [40] conducted a large empirical study, aiming to inform researchers on how to use data generated in an LBSN to attain better understanding of human mobility in such a network.

### 4.3 Using Location

As the size of data transferred from mobile devices augments rapidly, it is getting more complicated for researchers to deal with these massive amounts of data. Dealing with this issue, recommender systems act like filters, by trying to provide the right information to users and reduce the noise. For this purpose, recommendation services perform their action in 3 basic steps: First, they gather information. Then, they model it, in a way that correlations between entities can be easily inferred and finally, they provide the recommendations. Nowadays, the increasing availability of mobile location acquisition technologies, such as GPS and/or WiFi, adds a new dimension in recommender systems; Location is a new feature, which is acquired automatically through the location aware/enabled
devices. As a new dimension, location can easily bridge the gap between real and cyber/mobile world. Thus, implementations in sectors such as tourism or geo-located marketing can be absolutely useful and highly profitable [18] [50].

There are several approaches that provide location or activities recommendations based on users mobile usage data. For example, there are systems that use Collaborative Filtering solutions, Bayesian Networks (BNs), tensor-based methods, clustering methods etc. In particular, there are methods that build a matrix or tensor data representation to provide location/activity recommendations of high quality. On the other hand, there are Bayesian or Markov chain approaches, to deal with the massive amounts of data, generated by users. In the following, we will investigate different methods for providing geo-social recommendations.

Collaborative Filtering

Geolife 2.0 is a Location-based Social Network introduced by Microsoft Research Asia[70]. It provides users with two types of location/activity recommendations (i.e. generic and personalized). Regarding the generic part, Geolife 2.0 uses a Hits-based algorithm to infer the popularity of the recommended locations/activities. Experimental results have shown that HITS-based [30] algorithm outperforms baseline methods, such as ranking location/activity recommendations by frequency. As far as the personalized recommendations are concerned, Geolife 2.0 uses a hierarchical graph-based similarity measure to model the individuals location history. Moreover, Zheng et al. [66], exploits GPS trajectories to provide both location and activity recommendations. Their approach, denoted as Collaborative Location-Activity Recommendation (CLAR), performs a factorization of the Locations-Activities matrix. They also propagate information from two additional sources (i.e. Location-Feature matrix and Activity-Activity similarity matrix) to reduce
the data sparsity of the Location-Activity matrix.

Furthermore, Mao Ye et al. [64] developed a collaborative recommendation algorithm, which exploits geographical influence and incorporates it in a naive Bayesian model. CLR [34] builds a Collaborative Location Recommendation framework based on co-clustering. It applies a community location model to capture the relations between users, activities and locations. This model aims to mine similarities between the above mentioned entities and divides them into clusters of similar users, locations and activities. To perform clustering, authors use a hierarchical agglomerative clustering algorithm.

**Bayesian and Markov approaches**

There are also other methods that handle data as a Bayesian Network (BN). For example, [43] pre-processes geo-social information and trains the parameters of a BN. They obtain a conditional probability table (CPT) by performing the Expectation-Maximization (EM) model. For every new location/activity request by a user, the highest probability parameter, learned by the BN model, is selected. Quercia et al. [47] suggests measuring friendship probability as a function of two variables (i.e. how many times two users met each other and how long their meetings last). They apply a Markov chain algorithm [47] to provide a location/activity recommendation.

**Cold-start problem**

Regarding the cold-start problem, which refers to users who are not reluctant to provide adequate information (i.e. they provide only their home location), Quercia et al. [49] carried out a research, investigating which is the most effective algorithm (comparing six baseline algorithms) to solve this problem. Results have shown that the most effective algorithm recommends events that are popular among residents of an area. On the contrary, the least effective algorithm recommends events that are geographically close to the area of residents.
Context-based Recommendation

Multiverse Recommendation [28] is a method, which is based on n-dimensional tensor factorization. This approach enables a generic and flexible integration of contextual information in a multi-dimensional matrix (i.e. tensor), instead of performing the traditional 2-dimensional matrix factorization. To achieve this, data is being modelled with an n-dimensional tensor, having dimensions such as user, location and context (i.e. time, weather etc.). Moreover, authors of [28] proposed an algorithm that factorizes the n-dimensional tensor. Results have shown that the proposed method outperforms non-contextual matrix factorization by 30% in terms of Mean Average Error (MAE). Additionally, tensor factorization outperforms state-of-the-art methods by 2.5% to more than 12% depending on the data set.

Tensor-based approaches

In the same direction with the n-dimensional tensor factorization, other tensor-based approaches provide geo-social recommendations as well. For example, Biancalana et al. [5] implemented a social recommender system based on a tensor that is able to identify user preferences and information needs and suggests personalized recommendations for possible Points of Interests (POIs). Moreover, Zheng et al. [68] proposed a method, where geographical data is combined with social data to provide location and activity recommendations. The authors used GPS location data, user ratings and user activities to propose location and activity recommendations to interested users and explain them accordingly. In particular, they proposed a method, known as UCLAF (user-centered collaborative location and activity filtering), which is based on Tensor decomposition and addresses efficiently the data sparsity problem. In particular, Zheng et al. [68] use a 3-dimensional tensor to represent relations between users, locations and activities. Based on this representation, they subsequently apply tensor decomposition.
In contrast to the aforementioned tensor-based methods, Symeonidis et al. [56] [42] proposed a system that both provides (i) location and activity recommendations and (ii) friend recommendations by combining FriendLink algorithm [41] with the geographical distance between users. Furthermore, their tensor reduction method includes an incremental stage, in which newly created data is inserted into the tensor by incremental solutions. Authors have also built a Geo-Social Recommender System prototype, which is an online recommender system that relies on user check-ins to provide friend, location and activity recommendations. Every registered user is presented with the option of checking in. The procedure involves selecting the location he is currently at, the activity he is performing there, and finally rating that activity. Based on the users’ check-in history and friendship network, Geo-Social provides friend, location and activity recommendations. Friends are recommended based on the FriendLink algorithm [41] and the average geographical distances between users’ check-ins, which are used as link weights. Users, locations and activities are also inserted into a 3-order tensor, which is then used to provide location and activity recommendations.

4.4 Recommendation types

Based on location data provided by users, Location-based Social Networks’ providers can yield four different types of recommendations i.e. Friend, POI, Activity and Event recommendation. To achieve these recommendations, providers use some of the algorithmic approaches and frameworks presented in section (Dynamics). In the following part, this thesis introduces paradigms for each type, in order to illustrate recommendation concepts in each of these categories.

[http://delab.csd.auth.gr/geosocial]
4.4.1 Friend recommendation

Regarding friend recommendation, Symeonidis et al. proposed Friendlink algorithm\cite{41}. The novelty of this approach lays on the fact that it provides recommendations by traversing all paths of bounded length within a user’s candidate social graph. Traditionally, Social Networks do not exploit all different length paths of the network. Instead, they consider only pathways of maximum length 2 between a user and his candidate friends.

On the other hand, there are global approaches, which detect the overall path structure in a network, being computationally prohibitive for huge-size social networks. Based on the latter, they provided friend recommendations, by traversing all paths of a bounded length, based on the “algorithmic small world hypothesis”. As a result, they were able to provide more accurate and faster friend recommendations compared to other friend recommendation methods.

![Figure 4.1: Friend recommendation](image)

In the same direction, Scellato et al.\cite{54} studied the link prediction space, finding that about 30% of new links are added among “place-friends”, i.e. among users who visit the same places. They also show how this prediction space can be made 15 times smaller, while still 66% of future connections can be discovered. Thus, they define new prediction features based on the properties of the places visited by users which are able to discriminate potential future links among them. Building on these findings, researchers described a supervised learning framework which exploits these prediction features to predict new links among friends-of-friends and place-friends.
4.4.2 Point Of Interest recommendation

As far as POI recommendation concerns, there are also several approaches. For instance, Ye et al.\cite{64} argues that the geographical influence among POIs plays an important role in user’s check-in behavior and models it by power law distribution. Furthermore he proposes a unified POI recommendation framework which fuses user’s preference to a POI with social and geographical influence.

Another very recent study made by Saez-Trumper et al.\cite{51} has shown that recommending places which are closest to a users geographic center of interest, produces recommendations that are as accurate as, if not more accurate than, item-based recommenders.

4.4.3 Activity recommendation

In most of the cases, activity in Location Based Services is consistent with the POI that the user checks-in. This depends also in the existing categorization of each different Location-based service. For example, when a user checks in a basketball court she may performs different activities such as playing, watching or other. Nevertheless the recommender will suggest to the user other activities related to basketball no matter if she played, watched or just checked-in for other reason.

On the other hand, there are approaches that exploit other contextual information in order to yield activity recommendations for users. For instance, Belloti et al.\cite{3} proposed the Magitti leisure guide. The proposed approach infers user activity from context
and patterns of user behavior. Using those information, without any query generated by user, Magitti leisure guide automatically yields recommendations for different activities. Specific activities that Magitti can recommend are Eating, Shopping, Seeing, Doing, or Reading.

### 4.4.4 Event recommendation

This type of recommendation yields results for social events taking place in a specific location. Compared with the before mentioned types, event recommendation is relatively new. Apart from other works like [10], Daniele Quercia et al.[49] carried out a research about social event recommendations in greater Boston area.

Specifically, researchers extracted social events that took part during a certain period and combined it with obtained estimates for one million mobile phone users. After the analysis of the combined data set they were able to know which social event residents of an area have attended. Thus they could recommend social events to users. Moreover they applied six different algorithms in order
to measure the efficiency of each one for event recommendations. Results of that study shown that the most effective algorithm, is the one that recommends events that are popular among residents of an area. On the contrary the least effective, recommends events that are geographically close to the area.

In a more general context, Kayaalp et al. [29] proposed a mash-up application that utilizes hybridized filtering techniques to recommend events at Facebook website. To evaluate their approach, researchers performed an experiment which revealed very promising results.
Chapter 5

Recommendation Algorithms

Following the presentation of dynamics and the introduction of recommendation process in Location-based Social Networks, this chapter deepens in the recommendation area. In particular, the thesis records data modeling frameworks and state of the art algorithms, presented in research work. Initially, an acquaintance with the data representation is established. Thereafter, complete algorithmic solutions are documented, the State of the Art algorithms, in a chronological order.

5.1 Data representation

Based on the assumption that an LBSN is developed to host millions of users, there is a proper need of efficient data modeling. In that direction, researchers have proposed several frameworks for data modeling as parts of personalized and generic recommendations frameworks. This section advances the concept of data handling and aims to illustrate how data is represented, in order to be processed by the recommendation algorithm and finally outcome
an efficient recommendation for the user. For that reason we are going to present some specific paradigms of how data is accumulated and being modeled for the previously mentioned purposes.

Data acquisition is achieved through GPS or Wi-Fi enabled devices. The accumulated data is stored in databases. However, Location-based Social Networks cache an increasing userbase, thus providers might model the increasing entries efficiently, in order to provide users with efficient recommendations. But how do this massive amount of data is being organized in order to be used for features such as personalized recommendations? The review of the academic work outcomes three different types of data modeling i.e. Matrix, Tensor and Graph based representation.

Matrix-based

A matrix is a 2-dimensional array arranged in rows and columns. The individual items in a matrix are called elements or entries. In the case of Location-based Social Networks, matrices are used to represent relations between the entities of the system i.e. users, locations, activities and events.

A representative example of matrix-based approaches is the work of Zheng et al. \[66\]. In particular, Zheng et al. proposed a data modeling approach in order to obtain location-activity, location-feature and activity-activity matrices which can be used as inputs to train the recommender system of an LBSN. To achieve that, at first they applied a Grid-based clustering to extract the stay regions of a user. Cutting the map into grids, researchers were finally able to employ a strategy to cluster them. In addition they extracted the GPS data, being attached to each stay region by users and used the comments to infer the

```
1 0 0 0 0
0 0 0 1 0
0 0 1 0 0
1 0 0 0 0
0 0 0 0 0
```

Figure 5.1: Matrix
activity which user performed in each region. Finally, they used a POI category database to get counts of different POIs in a stay region and the web to extract activity-activity correlations in order to infer the activities that can be performed in a specific region. After all that the recommendation process was able to start.

**Tensor-based**

A tensor is a mathematical structure consisted of multiple (more than 2) dimensions. Actually, tensors are considered as a multidimensional matrices. The specific structure is used for the representation of interdependent information. Regarding LBSN structure, the representation of information can be ideally modeled in a tensor structure. The main point is that in Location-based Social Networks there are three core dimensions i.e. Location, User and Activity. The interdependency of all those dimensions is crucial in order to yield a highly efficient recommendation for users. Thus, researchers have developed tensor-based approaches to tackle the challenge of adding except from the previously mentioned dimensions, also other contextual information such as user preferences, place rankings etc.

For instance Symeonidis at al. [42] proposed the representation of friends, locations and activities by a 3-order tensor. Furthermore, they apply on it dimensionality reduction and a latent semantic analysis in order to yield recommendations for the users.

In the same direction, Yang ang Wang [62] proposed a novel 3-D Tag-Cloud module for a scalable personalized mobile information pushing platform, which can provide user-friendly and flexible location-based services. Using this module, researchers were able to retrieve useful information and use it in recommendation process.
In particular, they designed a multi-dimensional collaborative filtering algorithm which exploits the GPS location-based services and the latest Web2.0 technologies, in order to achieve either dynamic personalized recommendation or mobile information sharing. The proposed system incorporates efficiently Location-based data using a Service-Oriented Architecture.

On a multidimensional direction, Karatzoglou et al. [28] proposed a model which considers different types of context as additional dimensions in the representation of data as a tensor. The factorization of this tensor leads to a compact model of the data, which can be used to provide context aware recommendations in a Location-based environment.

**Graph-based**

A graph data structure consists of a finite set of ordered pairs. These pairs are edges of certain entities called nodes. As in mathematics, an edge (x,y) is said to point or go from x to y. As in mathematics, an edge (x,y) is said to point or go from x to y. The nodes may be part of the graph structure, or may be external entities represented indices or references. The social graph presented in 2.1.1 could be considered a representative instance of a graph-based structure. In the case of LBSN, graphs are used to model data and infer similarity between the entities of the network. In the following, this thesis introduces three different Graph-based models namely HGSM, TBHG and SLM-MTM.
**HGSM:** In 2008 Li et al. [36] proposed the HGSM framework for geographic information systems, to consistently model each individual’s location history and effectively measure the similarity among users. In this proposal, researchers took into account both the sequence of people’s movement behaviors and the hierarchical property of geographic spaces. The following figure depicts the position of HGSM within a recommender system. As it is illustrated, the HGSM is the final step of Location History representation and it is positioned right before the first step of similarity extraction process.

**TBHG:** Based on HGSM Zheng et al. [72], in 2009, proposed the TBHG framework. Based on TBHG, which stands for Tree Based Hierarchical Graph, researchers were able to model multiple individuals location histories and use HITS-based inference model [30], which regards an individuals access on a location as a directed link from the user to that location, to infer interest for a specific location.
In particular, this model infers the interest of a location by taking into account the following three factors. 1. The interest of a location depends on not only the number of users visiting this location but also these users travel experiences, 2. Users travel experiences and location interests have a mutual reinforcement relationship and 3. The interest of a location and the travel experience of a user are relative values and are region-related. Eventually, based on those three factors, researchers were able to handle massive data and thus, provide efficient recommendations. The following figure represents the TBHG’s position in the recommendation system. As it is illustrated, TBHG initializes the data for the location interest and sequence mining.

**Figure 5.5: TBHG in a recommender**

**SLH-MTM:** In contrast to aforementioned models, Xiao et al. proposed SLH-MTM. An approach which first models a user’s GPS trajectories with a semantic location history (SLH). This model measures the similarity between different users’ Semantic Location Histories by using a Maximal Travel Match algorithm.
The novelty of SLM-MTM approach lies on the semantic meaning that it carries. In addition, the specific approach can estimate the similarity between users without overlaps in the geographic spaces. The following figure illustrates the position of SLM-MTM as a component in a Location-based recommender. The SLM-MTM exploits modeled data in a semantic approach, in order to initialize it for Location history matching.

Figure 5.6: SLH-MTM
5.2 State of the Art Algorithms

This section presents academic work proposed for the field of Location-based recommendation algorithms in a chronological order. The presentation aims to illustrate the most complete solutions that can be implemented in a Location-based social network.

This starts from 2010 and the introduction of CLAF and PCLAF algorithms. Both of them provide Location and Activity recommendation within a Location-based Social Network. Regarding the algorithmic framework, both CLAF and PCLAF are categorized in Collaborative Filtering group. On the other hand, CLAF provides generic recommendations using matrices to represent data, unlike PCLAF which is built to provide personalized recommendations based on tensor data representation.

In 2011, research community proposed three new algorithmic schemes for recommendation in Location-based Social Networks. RMF, CADC and HOSVD algorithms introduced optimized approaches for the field. In particular, CADC and HOSVD incorporated incremental solutions to update data when new data inserted in the system. This approach elevated the recommendation process in a higher accuracy level.

Based on their previous works i.e. CLAF and PCLAF, Zheng et al. [67] proposed a new algorithm. In particular, RPCLAF is the only algorithm using Ranking-based modules within the recommendation process. In the following, there is a brief introduction to each of the previously mentioned approaches.
CLAF Proposed by Zheng et al. in [66]. It stands for Collaborative Location and Activity Filtering. The specific approach merges all the users’ data together and uses a collective matrix factorization model to provide general recommendations within Geolife [71], a prototype Location-Based Networking Service proposed by Microsoft Research Asia.

PCLAF Proposed, also, by Zheng et al. in [68]. PCLAF stands for Personalized Collaborative Location and Activity Filtering. Unlike CLAF, PCLAF treats each user differently and uses a collective tensor and matrix factorization to provide personalized recommendations.

The novelty that PCLAF approach brought, lays on the utilization of User-Location-Activity tensor along with the User-User and User-Location matrices. In particular, combining the 3-D tensor with matrices, researchers were able to tackle the generic recommendation problem and thus yield personalized recommendations for each user. Regarding the algorithmic type, PCLAF is categorized in the Collaborative Filtering group.

RMF RMF is a scheme for personalized place recommendations in Location-Based social networks [4]. Authors proposed an approach that interprets available data to SN providers using a User/Spot(Location) matrix. The recommendation scheme is based on regularized matrix factorization. In contrast to other proposals RMF Recommender does not load data into main memory.

Instead it processes one training data item at a time. The factor vectors are of much lower dimensionality compared to the ratings matrix and hence it can more easily be kept in main memory. Additionally, RMF qualifies fast prediction time. Once the model is learned the prediction can be made in \( O(1) \)

**CADC** CADC algorithm is part of a Collaborative Location Recommendation Framework based on Co-Clustering [31]. The specific framework, considers not only activities (i.e. temporal preferences) but also different user classes (i.e. Patern Users, Normal Users and Travelers) to yield a recommendation for Locations within a Location-based SN. Using this aproach CLR framework is capable of generating precise and refined recommendations to the users.

Unfolding CLR framework, it collects users’ GPS trajectory data and represents it in a Graph-based structure. Researchers proposed this Graph-based representation in order to capture the relations between Users, Activities and Locations. The novelty of CLM graph lays on the ability to from GPS trajectories users’ temporal preferences through their activities.

Moreover, the role of CADC is to cluster the trajectory data, generated by users into groups of similar users, similar locations and similar activities efficiently. In addition, CADC supports incremental updates of the groups when new GPS trajectory data arrives. Eventually, an important observation derived from this work is that, the activity time frame affects significantly the accuracy of the recommendation outputs.
HOSVD  HOSVD is used by Symeonidis et al. \cite{Symeonidis2013} in order to provide location/activity recommendations within the online prototype system, called Geo-Social 2.0 recommender system.\footnote{http://delab.csd.auth.gr/geosocial2/} To achieve recommendation, researchers represent data by a 3-order tensor, on which latent semantic analysis and dimensionality reduction is performed using the Higher Order Singular Value Decomposition (HOSVD) technique. Moreover, they use incremental solutions to update the tensor, as more data is accumulated to the system.

Compared to other solutions, HOSVD is the only algorithm, found in the bibliography, that provides Location, Activity and Friend recommendation. The latter is achieved with the incorporation of Friendlink algorithm \cite{Friendlink2010}. The complete recommendation package can be found in Geo-Social 2.0 recommender system. In addition, HOSVD is the first approach that uses incremental solutions to update data, each time new data inserted to a tensor.

RPCLAF  Proposed by Zheng et al. in \cite{Zheng2014}. It stands for Ranking-based Personalized Collaborative Location and Activity Filtering. The proposed algorithm formulates each user’s pairwise preferences on the location/activities and uses a ranking-based collective tensor and matrix factorization techniques to provide personalized recommendations. The contribution of RPCLAF approach is that it incorporates ranking within the objective function of the recommender. Experimental results revealed that this approach outperforms the before cited CLAF and PCLAF in terms of accuracy.
Chapter 6

Comparison and Conclusions

Having introduced the State of the Art approaches, in the following, the thesis presents comparisons using tables to categorize the algorithms, as well as figures to visualise this categorization. Eventually, a comparison between the State of the Art algorithms, is giving rise to discussions and prospective work apropos of the LBSN field.

6.1 Comparison

Initially, algorithms are being categorized in terms of 1) Personalization 2) Recommendation types 3) Data representation 4) Algorithmic type. Apart from this, there are tables that illustrate the experimental evaluation of the algorithms, as it is cited by their authors. Eventually, there is the aggregated comparison, in order to infer the optimal solution concerning recommendation process in Location-based Social Networks. To achieve results, the thesis uses the previously mentioned categorization tables in conjunction with the experiments performed, over different datasets, by the algorithms’ producers.
6.1.1 Framework and features

In order to highlight the different approaches that State of the Art algorithms present, the thesis introduces tables illustrating either the features that an algorithm qualifies, or the framework that is exploited in order to provide Location-based recommendations. At the end of the part, there is a visualized approach of the following categorizations.

Recommendation Direction

Generic recommendation process takes as input the whole number of check-ins revealed by users for a specific area. Actually, a generic recommender infers the most popular places of an area without considering personal user’s preferences. According to the following table, the vast majority of presented algorithms perform personalized recommendations. That might be logical, this kind of recommendation might be more efficient when addressed to people. Regarding aggregated comparison’s purposes, one point is assigned to each algorithm that affords personalized recommendations.

<table>
<thead>
<tr>
<th>Algorithm/Type</th>
<th>Generic</th>
<th>Personalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAF</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>PCLAF</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>RMF</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>CADC</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>HOSVD</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>RPCLAF</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 6.1: Generic/Personalized categorization
Recommendation Type

Regarding the type of recommendation which was introduced in section 4.4, it can be observed that all algorithms recommend locations. That might be a prerequisite for an algorithm to be qualified for utilization in Location-based services. But what about the other types? Recalling Part II, it might be true for the case of recommendation that, the more features a provider offers the more users is likely to accumulate for the network. As far as aggregated comparison concerns, one point is assigned for each type of recommendation an algorithm qualifies.

<table>
<thead>
<tr>
<th>Algorithm/Type</th>
<th>Friend</th>
<th>Location</th>
<th>Activity</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAF</td>
<td>✔</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>PCLAF</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>RMF</td>
<td>✔</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>CADC</td>
<td>✔</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>HOSVD</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>RPCLAF</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>

Table 6.2: Recommendation type

Data Representation

The data representation approach plays an important role in the entrie modelization. The more efficiently data is being organized the more high performance the recommendation algorithm might be. According to the table 6.3, tensor-based approaches are widely used. The main reason for that should be the flexibility that a tensor qualifies for the data handling.

Recalling that aggregated comparison is addressed to infer the optimal recommender and not the ideal components to achieve it, data representation will not be considered as a criterio for the objective of this thesis.
Comparison and Conclusions

<table>
<thead>
<tr>
<th>Algorithm/Type</th>
<th>Matrix</th>
<th>Tensor</th>
<th>Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAF</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCLAF</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>RMF</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CADC</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>HOSVD</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPCLAF</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3: Data representation

**Algorithmic Type**

In respect of Algorithmic type, researchers use several approaches in order to achieve the recommendation. It is important to be referenced that some of the cited frameworks use hybrid solutions to achieve their objective. For this reason table 6.4 categorizes the algorithms based on the Algorithmic type that is highlighted by their authors. Regarding the objective, standalone hybrid solutions do not provide an advantage for the recommendation process. As a result, this type of categorization will not be considered as input for the aggregated comparison.

<table>
<thead>
<tr>
<th>Algorithm/Type</th>
<th>Collaborative filtering</th>
<th>Ranking</th>
<th>Clustering</th>
<th>SVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAF</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCLAF</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMF</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CADC</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOSVD</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>RPCLAF</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 6.4: Algorithmic type
Visualization

The following figures are visualizing the categorization of the State of the Art Algorithms, as it is performed in the previous pages.

Figure 6.1: Data representation and Algorithmic type.

Figure 6.2: Personalization and Types of Recommendation.
6.1.2 Experimental Evaluation

Apart from frameworks and features, algorithms are being compared based on the results of the experiments operated by their authors. In the following, this thesis presents the metrics that researchers have used to evaluate their proposals, as well as the datasets for those evaluation processes.

Metrics

In the following, there is an introduction to several metrics that have been used for the evaluation purpose. Note that researchers use different metrics to evaluate their proposals.

MAE  The Mean Absolute Error is a quantity used to measure how close forecasts or predictions are to the eventual outcomes.

Precision  Is the ratio of the number of relevant objects in the top-N list relative to N.

Recall  Is the ratio of the number of relevant objects in the top-N list relative to the total number of relevant objects, respectively.

nDCG  The normalized Discounted Cumulative Gain, is a metric for algorithmic effectiveness. It measures the usefulness, or gain, of a document, based on its position in the result list.

RMSE  Root mean square error or simply standard deviation. RMSE, quantifies the difference between values implied by an estimator and the true values of the quantity being estimated.

AUC  The Area Under Curve is equal to the probability that an Algorithm will rank a randomly chosen positive instance higher than a randomly chosen negative one assuming positive ranks higher than negative.
Datasets

Experimental algorithmic evaluation is performed over data. The more illustrative the data is the more precise might be the evaluation result. Thus, researchers have performed the evaluation of their approaches based on real world data accumulated from real users in different places. In the following, thesis introduces the datasets that have been used for the evaluation purpose. Note that where there is not an assigned name to the dataset, the thesis uses the number of the users as a name.

**ATX** This is a dataset used by Berjani et al.\[4\] to evaluate RMF recommender. ATX is comprised of 11896 users, 9525 locations, 249317 check-ins.

**NYC** This is a dataset used also by Berjani et al.\[4\] to evaluate RMF recommender. NYC is a subpart of the Gowalla dataset. It is comprised of 10132 users, 9290 places and 114256 check-ins.

**102** This a dataset used by Papadimitriou et al.\[42\] for the evaluation of HOSVD Algorithm. It is comprised of 102 users, 46 locations and 18 different types of activities.

**50** This a dataset contains 50 users. It is recorded from September 2009 to May 2010 and it is used by \[34\] for the performance evaluation of CLR recommendation framework.

**119** This dataset is collected from 119 users from April 2007 to October 2009. It is used by Zheng et al.\[67\] in order to evaluate CLAF, PCLAF and RPCLAF Algorithms. It contains 12765 GPS trajectories.
This dataset is collected from 164 users from April 2007 to October 2009. It is used by Zheng et al. for the evaluation of PCLAF algorithm [68], as well as by Symeonidis et al. [56] for the assessment of HOSVD algorithm. It contains 12765 GPS trajectories.

This dataset is collected from 162 users from April 2007 to October 2009. It is used by Zheng et al. [66] for the evaluation of CLAF algorithm. It contains 12765 GPS trajectories.

Aggregation of the Experimental Evaluation

Having presented the Metrics as well as the Datasets that researchers proposed to evaluate their approaches, in the following, the thesis assigns a positioning comparison in order to infer the optimal solution in terms of experimental evaluation results.

The academic community qualifies experiments for location and activity recommendation. For that reason, the thesis uses separate comparative tables to aggregate the results for Location and Activity recommendation, respectively. Notice that the overall performance metrics, such as nDCG appear in both the tables.

As there is lack of holistic experiments, the outcome of the following tables will be quite sparse. To the point, the most holistic experiment have been performed incorporating three algorithms, proposed by the same team, while other approaches compete alone. In fact, a global comparison over the experimental results cannot be addressed. However, the outcome of the following tables will be used as input in the aggregated comparison of 6.1.3.
### Table 6.5: Evaluation of Location Recommendation over different datasets

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Metric</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ATX</td>
</tr>
<tr>
<td>CLAF</td>
<td>AUC</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>nDCG</td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>3</td>
</tr>
<tr>
<td>PCLAF</td>
<td>AUC</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>nDCG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>RMF</td>
<td>MAE</td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>(1)</td>
</tr>
<tr>
<td>PCLAF</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CADC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOSVD</td>
<td>Precision</td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>nDCG</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>1</td>
</tr>
<tr>
<td>RPCLAF</td>
<td>AUC</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>2</td>
</tr>
</tbody>
</table>

### Table 6.6: Evaluation of Activity Recommendation over different datasets

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Metric</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ATX</td>
</tr>
<tr>
<td>CLAF</td>
<td>AUC</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>nDCG</td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>3</td>
</tr>
<tr>
<td>PCLAF</td>
<td>AUC</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>nDCG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>HOSVD</td>
<td>Precision</td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>nDCG</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>1</td>
</tr>
<tr>
<td>RPCLAF</td>
<td>AUC</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>2</td>
</tr>
</tbody>
</table>

90
Unfolding the previous comparison, the most global overview is revealed over the 119 and 164 datasets where PCLAF, HOSVD and RPCLAF compete for the first position i.e. The most efficient recommendation. Notice (1) which denotes that the algorithm has been experimentaly tested in a specific dataset, but there is no other evaluation experiment that uses the same dataset as well as the same metrics. Hence, in this case there is not the basis for a comparison using the experimental performance as a criterio. Thus, (1) delivers 0 points as input in the aggregated comparison of 1.6.3.

Regarding further pointing, taking into consideration the different datasets and metrics, 1 point is assigned to each approach revealed as the optimal (i.e. 1) compared to the others over the same dataset. Additionally, 0.5 points are assigned to each algorithm classified in the second position (i.e. 2) when the evaluation incorporates more than two approaches i.e. Dataset 119.
6.1.3 Aggregated Comparison

Having presented experimental evaluations, features and the operative frameworks of the State of the Art Algorithms, in the following, the thesis addresses an aggregated comparison in order to result the optimal Algorithm concerning recommendation in a Location-based Social Network.

The criteria for this comparison spring from different sections of Part III. Notice that the comparison is being addressed three times due to the lack of previous works concerning this direction. In particular, table 6.7 compares the algorithms without taking into consideration the experimental evaluation. This approach qualifies the righteous comparison. However, tables 6.8 and 6.9 incorporate experimental evaluation results only for the algorithms evaluated under the same evaluation framework.

In the following, the thesis unfolds the criteria for the needs of the aggregated comparisons and introduces the comparative tables.

**Direction** This criterion qualifies 1 point for algorithms able to address personalized recommendations.

**Recommendation Type** 1 point is assigned for each different type of recommendation an algorithm provides.

**Incremental Solutions** It favors 1 point for algorithms using incremental techniques for the renewal of data when new entries are inserted in the system.

**164** This is the evaluation of algorithms over the 164 Dataset. The points are assigned based on section’s 6.1.2 tables.

**119** This is the evaluation of algorithms over the 164 Dataset. The points are assigned based on section’s 6.1.2 tables.
Comparison

Comparison and Conclusions

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Direction</th>
<th>Recommendation type</th>
<th>Incremental Solution</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAF</td>
<td>0</td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>PCLAF</td>
<td>1</td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>RMF</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CADC</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>HOSVD</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>RPCLAF</td>
<td>1</td>
<td>2</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Optimal</td>
<td>119</td>
<td>HOSVD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.7: Optimal Algorithm based on framework and features

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Direction</th>
<th>Recommendation type</th>
<th>Incremental Solution</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAF</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>PCLAF</td>
<td>1.5</td>
<td>1</td>
<td>2</td>
<td>4.5</td>
</tr>
<tr>
<td>RMF</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CADC</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>HOSVD</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>RPCLAF</td>
<td>1.5</td>
<td>1</td>
<td>2</td>
<td>4.5</td>
</tr>
<tr>
<td>Optimal</td>
<td>HOSVD</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.8: Optimal Algorithm - Using the 164 Dataset
As the previously performed comparison revealed, the optimal solution for the recommendation process in Location-based Social Networks is the HOSVD algorithm. This approach is used by Symeonidis et al.\textsuperscript{[56]} in order to provide location/activity recommendations within the online prototype system, called Geo-Social 2.0 recommender system\textsuperscript{[1]} \footnote{\url{http://delab.csd.auth.gr/geosocial2/}}. In addition, this approach incorporates the Friendlink algorithm \textsuperscript{[41]} in order to provide friend recommendation as well.

To achieve recommendation, researchers represent data by a 3-order tensor, on which latent semantic analysis and dimensionality reduction is performed using the Higher Order Singular Value Decomposition (HOSVD) technique. Moreover, they use incremental solutions to update the tensor, as more data is accumulated to the system.
6.2 Conclusions

Having covered all the way, right from the wide concept of Online Social Networks, till the comparison of State of the Art Algorithms for Location-based Social Networks, the time had come to address an answer to the inquiry void of this thesis. Definitely, the optimal solution for the recommendation process in Location-based Social Networks, as the comparisons revealed, is the HOSVD Algorithm proposed by Symeonidis et al. in 2011 [56].

Shortly, this approach provides 3 different types of personalized recommendation, using location as input. In addition it qualifies an innovative technique for the renewal of data when it is inserted in the system. As far as an incontrovertible comparison concerns, it might be reality once all the proposals have be evaluated over the same dataset, using the same metrics. Until then, this work reveals the optimal aggregation of the research activity concerning this direction.

![Figure 6.3: Optimal Classification](image)

Apart from the optimal Algorithmic solution, the previous thesis infers much more. The upcoming generations of people are inextricably linked to the internet [19]. The internet is starting to reach every corner of planet earth and the Wi-Fi metropolitan concept is approaching [44]. Hence, Location-based services seem to be a nice
place for users bored of the noise and fictitiousness that Traditional Online Social Networks exhibit. In the same direction, Location-based recommenders suggest innovative types of recommendation (4.4) enabling users to spare time and energy.

Eventually, the use of Geo-Location data qualifies the enhancement of a user’s Social Graph in an alternative way. As an additional feature in this direction, event recommendation should be considered as a great contribution. In fact, events are social meetings in physical level. Thus an efficient personalized recommendations lead, once again, to the junction between cyber world and real life. Although this recommendation type appears in the bibliography [49] [29] [10], there is not a complete solution to tackle the objective. Consequently, event recommendation might be a great challenge for the researchers of the field, over the upcoming years.

Regarding further enhancements, the incorporation of Privacy awareness is considered, for the moment, as the highest priority. The physis of Geo-Location data that is being delivered may rise undesirable side effects during the user experience and thus make any network vulnerable. Approaches like Building Blocks [46] and MobID [48] could be ideally incorporated into Location-based Social Networks and shield them against any unwelcome consequence. Furthermore, the exploitation of semantic meanings may kick start higher recommendation quality, and thus the increased user satisfaction.

Concluding this thesis, as well as in the beginning, the same query appears. Will Location-based Social Network become the next best thing of the Internet industry. Author assesses that “Yes, it will.” but, once again, time will tell.
Bibliography


