Chapter 119
Data Mining Location-Based Social Networks for Geospatial Discovery

Edward Pultar
University of California – Santa Barbara, USA

ABSTRACT
Modern, Internet-based social networks contain a wealth of information about each member. An integral part of an individual’s online profile is their Volunteered Geographic Information (VGI) such as a user’s current geographical location. Social network members in different cities, countries, or continents engage in different activities due to accessibility, economy, culture, or other factors. The work here focuses on data mining separate groups of social network profiles according to their geography in order to discover information about a place. This results in keywords associated with a specific location and provides an automated way to describe a place in an up to date fashion based upon its current local residents. Location-Based Social Network (LBSN) profiles from four different places are analyzed here and the results are presented as they relate to space, time, and activities.

INTRODUCTION
Digital social networks have seen tremendous growth in contemporary times. Seemingly ubiquitous availability of the Internet has led to billions of social network profiles that each present a virtual representation of a real human being. Some online social networks focus on allowing users to communicate with existing real-world ties while others afford new friendships based upon particular interests. The geographic breadth is intense as individuals from all of the world’s continents partake in membership of one or more online social network(s). Although the popularity of these social network services has risen only in the previous decade, their influence is vast. Online communities allow for easy, efficient communication between many people, which
results in emergent events such as flash mobs of
pillow fights in San Francisco or political upris-
ings and the restructuring of governments in the
Middle East.

In this chapter the following research questions
are addressed:

• How can web crawling techniques be used
with location-based social networks to cre-
ate different data repositories according to
geographic area?

• What role can data mining play in search-
ing for useful information within social
network profiles?

• How can social network data be visualized
in order to demonstrate the importance lev-
el of concepts related to space, time, and
activity?

The work here aims to address the challenge
of data mining location-based social networks in
order to discover useful information about differ-
ent geographic areas. The next section provides a
background of relevant literature. The third section
presents the approach and methodology used in
this chapter. Section 4 contains a discussion of the
results and potential future work. The last section
completes the chapter with final conclusions.

BACKGROUND

Social networks have been around since long be-
fore the Internet but some of the first Internet-based
social networks date back to 1997. One example
of these older digital social networks is sixdegrees.
com, but many of the older social networks are
as active as they once were. Networks have been
created for various groups and purposes over the
years with many such as Friendster and MySpace
having distinct high points of success followed
by low levels of membership activity. For years
now there have been academic studies of social
networks such as to how different roles shape an
individual’s activity in the network (Kumar, et al.,
2006; Pultar, et al., 2010a). For a general history
of social networks with in-depth definitions of
the various types the reader is directed to boyd
and Ellison (2007).

With years of frequently updated information
from all parts of the world, social networks are
ripe for data mining and discovery. Data mining
(Han & Kamber, 2006; Miller, 2007) is a tool
from computer science by which patterns are
found in large sets of data that can otherwise
be overwhelming. Available through social net-
works is a wide variety of data about individuals
pertaining to topics such as: eating or traveling
preferences, philanthropy, identity, and education.
See Liu, Maes, and Davenport (2006) for more on
how connections between individuals in a social
network can be structured by tastes.

One critical component of the majority of social
networks is the current geographical location for
each member. This has become a standard feature
in modern mobile phones thanks to the integration
of a Global Positioning System (GPS). A mobile
user’s location changes over time and previous
locations may also be provided in forms such as
historical tracks. Location is a valuable piece of
Volunteered Geographic Information (VGI; Good-
child, 2007) or user-generated content whereby
humans around the world act as sensors of their
environment. In many cases locals are able to
report more accurate and/or timelier information
about their surroundings. Twitter is one Internet-
based medium that allows individuals to instantly
give an account of their situation. Additionally this
brings about the notion of Location-Based Social
Networks (LBSN), which describes those social
networks that explicitly involve both the physical
and virtual worlds and their interactions. This ac-
tive research area hosts an annual workshop at the
ACM SIGSPATIAL GIS (http://www.sigspatial.
org/) conference.

Through the availability of this socio-spatial in-
formation combined with data mining techniques
researchers can examine the differences that exist

2007
Related Content

BIM-FM Implementation: An Exploratory Invesigation
[www.igi-global.com/article/bim-fm-implementation/89440?camid=4v1a](www.igi-global.com/article/bim-fm-implementation/89440?camid=4v1a)

Modern Navigation Systems and Related Spatial Query
[www.igi-global.com/chapter/modern-navigation-systems-related-spatial/70466?camid=4v1a](www.igi-global.com/chapter/modern-navigation-systems-related-spatial/70466?camid=4v1a)

The Columbia Regional Geospatial Service Center System: A Proven Model for the United States
[www.igi-global.com/chapter/columbia-regional-geospatial-service-center/70485?camid=4v1a](www.igi-global.com/chapter/columbia-regional-geospatial-service-center/70485?camid=4v1a)

Methodical Spatial Database Design with Topological Polygon Structures
[www.igi-global.com/chapter/methodical-spatial-database-design-topological/70447?camid=4v1a](www.igi-global.com/chapter/methodical-spatial-database-design-topological/70447?camid=4v1a)