Chapter 19
A Beginner's Guide to Geographic Virtual Communities Research

Brent Hecht
Northwestern University, USA

Darren Gergle
Northwestern University, USA

INTRODUCTION

Virtual communities have important geographic components. Community participants live, work, and travel to specific places on the Earth's surface, and communities often reflect the characteristics of these places. In addition, community artifacts are often imbued with geographic information.

Researchers can use these often under-appreciated geographic elements to understand important patterns in virtual communities' interaction with the real world. For instance, one could build and study a shared repository for a biking community's geographic knowledge (Priedhorsky & Terveen, 2008), investigate whether community artifact density is biased towards certain areas of the globe (Hecht & Gergle, 2009), or model the particular characteristics of a community's spatio-social network (Larsen, Axhausen, & Urry, 2006; Larsen, Urry, & Axhausen, 2006).

Geographic analyses can also allow an investigator to answer questions that are *not overtly geographic in nature*. In such cases, these analyses can provide an efficient alternative or supplement to more traditional methods such as large-scale surveys, interviews, or observational techniques. In many ways, it is this capability of geographical analyses that is more powerful for the virtual communities researcher. The number of research topics here are infinite, but could include modeling the relationship between social networking site usage and socioeconomic status, understanding human photo-taking behavior (Hecht & Gergle, 2010; Yanai, Yaegashi, & Qiu, 2009), modeling and sharing dynamic travel behavior based on interaction within social networks (Pultar &

DOI: 10.4018/978-1-60960-040-2.ch019

Raubal, 2009), and identifying self-focus bias in wikis (see the case study at the end of the chapter).

This chapter is targeted at the virtual community researcher who wants to quantitatively examine or employ the geography of a community, but has no training in the methodologies necessary to do so. We take the reader from the data collection stage through the application of several simple techniques, suggesting more advanced literature when space limitations prevent us from delving into details. We also take special care to flag important pitfalls that cause hard-to-notice but critical errors. Finally, we close with a brief but illustrative research project case study.

This chapter is effectively an introductory lesson in Geographic Information Systems (GIS) and Geographic Information Science (GIScience), customized for the virtual communities researcher. A GIS is a "set of tools for performing operations on geographic data that are too tedious or expensive or inaccurate if performed by hand". In doing so, it helps "reveal what is otherwise invisible in geographic information" (Longley, Goodchild, Maguire, & Rhind, 2005b). Another definition many GIS educators find useful describes GIS as a "powerful set of tools for collecting, storing, retrieving at will, transforming, and displaying spatial data from the real world for a particular set of purposes." (Burrough & McDonnell, 1998) GIScience is the science and engineering behind this "set of tools". It can be loosely considered analogous to information science but for the well-defined class of geographic information (Longley, et al., 2005b).

While GIS/GIScience and computer science are closely related, this chapter should be accessible to readers with no programming experience at all. However, programming ability (or access to someone with knowledge of programming) will help the reader more readily leverage the tools we mention for their own research. In particular, experience with web-based application programming interfaces (APIs), Java, and/or statistical programming will be useful.

MINING GEOGRAPHIC INFORMATION FROM VIRTUAL COMMUNITIES

Before engaging in any study involving the geographic component of virtual communities, it is necessary to obtain geographic information or to transform pre-existing geographic information into a "usable" form. Usable forms include latitude/longitude coordinates, bounding boxes around geographic features, and advanced polygonal and polylinear representations (e.g. the shape of the United States and the path of a road), along with the attribute information attached to these data, such as a username, population, etc.

Formally, geographic information is defined as "atomic pairs of the form $<x,z>$ where x is a location in space[1] and z is a set of properties [attributes] of that location; or information that is reducible to such pairs." (M. Goodchild, 2001; M. Goodchild, Yuan, & Cova, 2007). For example, the x in a pair could be a latitude/longitude of a city that is mentioned in a forum posting, and the z could include the average income of the city, the username of the poster, his/her centrality in a social network, and/or the size of the post (Figure 1).

This section discusses important methodologies for obtaining geographic information and making it usable for virtual communities research. We also point the reader to easy-to-use tools for applying these methodologies.

Latitude and Longitude Pairs

A growing number of virtual communities generate community artifacts that contain latitude and longitude coordinates. Assuming this structured information is accurate, it is often immediately "usable" in geographic analyses. Classic examples include the latitude and longitude ("lat/lon[g]") tags that have been manually associated with hundreds of thousands of Wikipedia articles or online photo collections that have been manually or automatically tagged with lat/lon information.