Chapter 8
Using Social Network Analysis to Guide Theoretical Sampling in an Ethnographic Study of a Virtual Community

Enrique Murillo
Instituto Tecnológico Autónomo de México – ITAM, Mexico

ABSTRACT

Social Network Analysis (SNA) provides a range of models particularly well suited for mapping bonds between participants in online communities and thus reveal prominent members or subgroups. This can yield valuable insights for selecting a theoretical sample of participants or participant interactions in qualitative studies of communities. This chapter describes a procedure for collecting data from Usenet newsgroups, deriving the social network created by participant interaction, and importing this relational data into SNA software, where various cohesion models can be applied. The technique is exemplified by performing a longitudinal core periphery analysis of a specific newsgroup, which identified core members and provided clear evidence of a stable online community. Discussions dominated by core members are identified next, to guide theoretical sampling of text-based interactions in an ongoing ethnography of the community.

INTRODUCTION

The Usenet discussion network is a popular area of the Net, attracting participants from every corner of the world. It works as a public bulletin-board, organized into topical discussion groups called *newsgroups*, whose number Turner, Smith, Fisher and Welser (2005) put at 150,000. It thus provides a convenient way for people with very diverse interests to find each other and discuss their passion.

Usenet's large base of participants, neatly organized into distinct knowledge realms, has resulted in the emergence of many topically-focused virtual communities. Most are virtual communities of interest (Blanchard & Markus, 2004), focused on a particular passion or hobby, such as the *Harry Potter* books (alt.fan.harry-potter) or stamp collecting (rec.collecting.stamps.discuss). Others are more specialized communi-

DOI: 10.4018/978-1-60960-040-2.ch008

ties serving specific professional groups; most dealing with computer topics, but many others addressing fields far removed from computers, such as taxation issues (misc.taxes.moderated), medical transcription (sci.med.transcription) or farming (uk.business.agriculture).

This spontaneous forming of online communities explains why Usenet was an early platform for conducting unobtrusive naturalistic research of computer-mediated communications and the social environments that emerge from them (Lee, 2000). Since newsgroup messages are posted online for all to see, they constitute a publicly accessible record of discussions, offering a wealth of research data. A substantial portion of the extant literature about virtual communities is focused on Usenet newsgroups and mailing lists or listservs, a similar technology. As for the newer Web 2.0 platforms, research about virtual communities is fairly recent, with blogs taking the lead. Various approaches have been proposed for identifying communities that form around linked blogs (e.g. Kumar, Novak, Raghavan, & Tomkins, 2005; Chin & Chignell, 2007; Chau & Xu, 2007), and there have been some qualitative studies of specific communities (e.g. Kaiser, Müller-Seitz, Pereira Lopes, & Pina e Cunha, 2007; Silva, Goel, & Mousavidin, 2008). There are fewer studies of wiki-based communities (e.g. Bryant, Forte, & Bruckman, 2005), and fewer still of communities based on social networking sites like Facebook or Tweeter (e.g.; Ellison, Steinfield, & Lampe, 2007), although the number of publications will undoubtedly grow in the coming years.

Since the aim of most studies of virtual communities is to describe a culture, they tend to rely on qualitative methods, which involve reading and analyzing samples of messages exchanged by community members. It is therefore important to provide a theory grounded rationale for sample selection in order to avoid the trap of "anecdotalism" (Silverman, 2000). Howard (2002) proposes using Social Network Analysis (SNA) to identify significant members of the online community and to focus ethnographic analysis on this purposive sample.

This chapter describes the steps taken for theoretical sample selection in an ongoing ethnography of the newsgroup misc.taxes.moderated (henceforth MTM) which hosts a long established community of practice (CoP) of tax professionals. The ethnography aims to provide a rich description of day-to-day interactions in the community. Theoretical sampling of relevant interactions was guided by a specific SNA technique, the continuous core-periphery model (Borgatti & Everett, 1999). The chapter describes the procedure used for data collection, social network analysis, identification of core members and theoretical sampling of discussion threads dominated by core members.

The chapter is organized as follows. Section 1 provides background research on virtual communities and applications of SNA to Usenet. Section 2 explains how participants form social networks in Usenet and how these can be derived. Section 3 demonstrates the power of this technique by performing a longitudinal core-periphery analysis of MTM spanning six years. The last section discusses the rationale and implicit assumptions of this approach and how it generalizes to other Internet platforms.

VIRTUAL COMMUNITIES AND QUALITATIVE RESEARCH

Studies of virtual or Internet-based communities began in the early nineties. Early examples are Rheingold's (1993) book and North's (1994) thesis about Usenet culture. Scholarly research grew rapidly in the following years (e.g. Jones, 1995), painting a broad picture of the characteristics of these communities, and the kinds of social interaction the plain-text medium of newsgroup or listserv messages can support.

Researchers were surprised to find that virtual communities can exhibit rich cultures (Baym, 1995; Tepper, 1997), develop a sense of commu-