Chapter 11
An Introduction to Computational Social Science for Organizational Communication

Andrew N. Pilny
University of Kentucky, USA

Marshall Scott Poole
University of Illinois at Urbana-Champaign, USA

ABSTRACT

The exponential growth of "Big Data" has given rise to a field known as computational social science (CSS). The authors view CSS as the interdisciplinary investigation of society that takes advantage of the massive amount of data generated by individuals in a way that allows for abductive research designs. Moreover, CSS complicates the relationship between data and theory by opening the door for a more data-driven approach to social science. This chapter will demonstrate the utility of a CSS approach using examples from dynamic interaction modeling, machine learning, and network analysis to investigate organizational communication (OC). The chapter concludes by suggesting that lessons learned from OC's history can help deal with addressing several current issues related to CSS, including an audit culture, data collection ethics, transparency, and Big Data hubris.

INTRODUCTION

Many have said we live in an information age, one that emphasizes an ever-increasing ability to retrieve and acquire a plethora of data from a vast amount of resources. Much of this discussion highlights individuals' abilities to act as consumers of information or at least as beneficiaries of mostly large organizations such as Google or the U.S. government to acquire and process data. However, there has been less emphasis on how the information age allows individuals to act as producers, whether conscious or not, of information. Indeed, from every email, text message, financial transaction, Facebook status update,

DOI: 10.4018/978-1-5225-2823-4.ch011

Tweet, and even keystroke, individuals are constantly producing information about themselves and others, and it is increasingly the case that they can analyze this data themselves.

The result is the accumulation of what is more commonly known as Big Data. The term refers to not only the size of data, but also its ability to take many different forms and speed at which it is generated (Zikopoulos & Eaton, 2011). Computational social science (CSS) emerged because of Big Data's overall size and temporal nature and the frequent inability of traditional quantitative methods (e.g., general linear model) to analyze the complexity of Big Data (Lazer et al., 2009).

The purpose of this chapter is to define CSS and describe the potential of CSS for organizational communication (OC). The next section focuses on description by defining three features that underlie CSS and their philosophical assumptions. Then, there will be three brief examples of research related to OC involving methods such as relational event modeling, machine-learning classification, and descriptive network analysis. Finally, the conclusion describes current issues confronting CSS and how insights from OC can help address them.

WHAT IS COMPUTATIONAL SOCIAL SCIENCE?

The Problem of Focusing on Tools

Despite the abundance of new papers aspiring to a CSS approach, defining CSS has been challenging. That is, what do researchers mean when they are doing CSS research? Perhaps a good place to begin is Lazer and colleagues (2009), who made the case that CSS represents a distinct field of research spearheaded by advances in computerization and availability of trace data from numerous sources ranging from the Internet to cell phones. CSS is about how social science can be fundamentally transformed by a newer "capacity to collect and analyze massive amounts of data" (Lazer et al., 2009, p. 2). Researchers may also generate data through simulation methods like agent based modeling and cellular automation (Gilbert, 2007).

A good example of a CSS approach is Shaw and Hill's (2014) examination of Robert Michels' (1911) iron law of oligarchy. Michels argued that as groups, organizations, and societies initially develop democratically, a small handful of early adopters will inevitably dominate because of the complex necessities to run large collectives. This ruling class will then monopolize and consolidate power and tend to look more like an oligarchy than a democracy.

Shaw and Hill (2014) examined Wikia, an online commons-based production firm. Wikia hosts a plethora of different wikis that rely on democratic peer production where users add, modify, and remove all content. Using a sample of 683 of the largest wikis from the site, the authors were able to collect trace data about user interaction, including edits, page creations, reverted edits, and number of editors and administrators added. The dataset eventually included millions of time-stamped data points over a five-year time period (over 264 GB!).

Their results found empirical support for Michels' theory. More specifically, they found the addition of new administrators did not grow at the same rate as new members (i.e., leading to concentration), administrators tended to contribute more to page edits over time, and finally, as memberships grew, administrators tended to remove more old content (i.e., perform more 'reverts'). These three findings represent characteristics consistent with increases in oligarchy (Shaw and Hill, 2014, p. 225).