

Expression and Reception: An Analytic Method for Assessing Message Production and Consumption in CMC

Kang Namkoong, Dhavan V. Shah, Bryan McLaughlin, Ming-Yuan Chih, Tae Joon Moon, Shawnika Hull & David H. Gustafson

To cite this article: Kang Namkoong, Dhavan V. Shah, Bryan McLaughlin, Ming-Yuan Chih, Tae Joon Moon, Shawnika Hull & David H. Gustafson (2017): Expression and Reception: An Analytic Method for Assessing Message Production and Consumption in CMC, Communication Methods and Measures, DOI: [10.1080/19312458.2017.1313396](https://doi.org/10.1080/19312458.2017.1313396)

To link to this article: <http://dx.doi.org/10.1080/19312458.2017.1313396>



Published online: 18 Apr 2017.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Expression and Reception: An Analytic Method for Assessing Message Production and Consumption in CMC

Kang Namkoong^a, Dhavan V. Shah^b, Bryan McLaughlin^c, Ming-Yuan Chih^d, Tae Joon Moon^b, Shawnika Hull^e, and David H. Gustafson^f

^aDepartment of Community and Leadership Development, University of Kentucky, Lexington, Kentucky; ^bSchool of Journalism and Mass Communication, University of Wisconsin, Madison, Wisconsin; ^cCollege of Media and Communication, Texas Tech University, Lubbock, Texas; ^dCollege of Health Sciences, University of Kentucky, Lexington, Kentucky; ^eDepartment of Prevention and Community Health, The George Washington University, Washington, DC; ^fCenter for Health Enhancement Systems Studies, University of Wisconsin, Madison, Wisconsin

ABSTRACT

This article presents an innovative methodology to study computer-mediated communication (CMC), which allows analysis of the multi-layered effects of online expression and reception. The methodology is demonstrated by combining the following three data sets collected from a widely tested eHealth system, the Comprehensive Health Enhancement Support System (CHESS): (1) a flexible and precise computer-aided content analysis; (2) a record of individual message posting and reading; and (3) longitudinal survey data. Further, this article discusses how the resulting data can be applied to online social network analysis and demonstrates how to construct two distinct types of online social networks—open and targeted communication networks—for different types of content embedded in social networks.

As the field of communication increasingly turns its attention to online communication, it is important to consider the ways in which we can advance our ability to study communication in these contexts. In many ways, the field has been trying to catch up with the changes occurring in communication patterns as a result of the technological advancements of digital media. Traditionally, for example, a reception-effects paradigm has dominated much of communication research, especially in the field of mass communication, which assumes most influence results from exposure to messages, as the ideas that people encounter either inform or persuade (Fishbein & Cappella, 2006; Lasswell, 1950). With the advent of computer-mediated communication (CMC), however, an expression-effects approach is emerging as a significant alternative to this dominant approach (Pingree, 2007; Shah, 2016). Indeed, an impressive body of research examines the effects of CMC expression on the message sender's psychological outcomes (Han et al., 2008; Shaw, Hawkins, McTavish, Pingree, & Gustafson, 2006).

Much of the empirical research on CMC either does not distinguish expression and reception effects (e.g., Price & Cappella, 2002) or focuses solely on expression effects (e.g., Shaw et al., 2006). In other words, many scholars have seemingly traded one unidirectional effects paradigm for another. Expression and its effects, however, do not occur in isolation in CMC contexts. Simultaneously examining and distinguishing message reception and expression can help scholars understand the “whole picture” of online communication effects. In this article, we explain how message reception and expression in CMC can be distinguished from one another and objectively measured to gain new theoretical insights.

The process through which online message expression and/or reception affect attitudinal, emotional, or psychological outcomes is inherently complex. Take, for example, the context of breast cancer survivors joining a computer-mediated social support (CMSS) group. It has been well established that individuals suffering from a health crisis can attain important psychological and emotional benefits from joining CMSS groups (e.g., Coursaris & Liu, 2009). Yet, numerous mechanisms through which these benefits are achieved have been suggested. Breast cancer patients have been found to benefit from receiving emotional and informational support (e.g., Lieberman & Goldstein, 2006), writing out their thoughts and feelings (e.g., Shaw et al., 2006), expressing support to others (e.g., Han et al., 2011), establishing social bonds (Holland & Holahan, 2003), and acquiring a sense of belonging (e.g., Gottlieb & Bergen, 2010). This complex context calls for methodological approaches that can simultaneously examine these different pathways. Ideally, an analysis of an online social group would include an examination of (1) what type of communication messages were exchanged, (2) how frequently an individual expressed or received each type of message, (3) emotional, attitudinal, or psychological outcomes, and (4) the patterns and strength of social relations that develop among participants.

Analyzed separately, each of these data points could provide interesting, perhaps notable, information. But they would necessarily be incomplete. Content analysis itself does not reveal much about individual experiences unless it is combined with data that allows the researcher to measure who wrote or read each message. Further, reception and expression data cannot provide much understanding of the effects of these messages of communication messages without survey data that allow assessment of outcomes. Finally, without the inclusion of social network analysis, it is hard to provide a more thorough picture of how the exchange of messages can foster social bonds and, subsequently, the effect of these bonds on various outcomes. For example, measuring message reception allows us to study “lurkers,” who would not be captured by the traditional content-based network analysis. Reading support messages from afar may lead to very different effects than being a central node within the network, which may be different still from the effects of having a smaller number of bonds, but ones that are characterized by the exchange of intimate support messages.

The methodology presented here helps establish a way to parse out these diffuse relationships, which may all be occurring simultaneously within one, seemingly homogenous, online social group. Specifically, in this article, we present a novel method that combines three distinct analytic approaches: computer-aided content analysis, action log data analysis, and survey data analysis. Recent studies have successfully used this methodology to examine the nature of message expression and reception occurring within CMC (Han et al., 2011; Kim et al., 2011; Namkoong et al., 2010, 2013; Yoo et al., 2014). In addition, we discuss how the resulting data can be used for social network analysis. This article also provides a more comprehensive explanation of this methodology, and detailed information about it can be extended to other CMC contexts.

Case study: Expression and reception of emotional support in a CMSS group

With the increasing need to understand both expression and reception effects, as well as to analyze large bodies of social media data, it is necessary to continually develop new methodological approaches for textual analysis (Grimmer & Stewart, 2013). The methodological approach outlined below can help address a number of lingering concerns, while being customized to a researcher’s specific needs and goals. Because CMC data can be complex—expansive yet idiosyncratic—we created a methodological approach that is (a) systematic yet flexible, (b) capable of handling a large quantity of data, (c) incorporates several levels of data, and (d) able to make important conceptual distinctions. More specifically, we combined (1) an iterative, mixed-method coding process of CMC messages, (2) with action log data that record who wrote or read which messages, and (3) longitudinal survey data about these users.

We demonstrate the data construction process by providing an example of how this method was applied to examine expression and reception of emotional support in a computer-mediated social

support (CMSS) group embedded in the Comprehensive Health Enhancement Support System (CHESS).¹ CHESS is a web-based interactive communication system that employs data on user health status to help users monitor their condition and guides them to online information, communication and coaching services (Gustafson, McTavish, Hawkins, Pingree, & Arora, 1998; Gustafson et al., 1994, 1999, 2008).

Below we detail the sequential process through which our methodology was applied to produce important insights about the psychosocial effects of expression and reception in the CMC context. This approach could be applied to capture a wide range of theoretical concepts from any computer-mediated social group, so long as the researcher has access to (1) information about who posted and read which message and (2) survey or other outcome data about the participating users.

Pre-coding steps

Data preparation

The first step in conducting this computer-aided content analysis is to acquire and clean the data. It is important at this point to decide what unit of analysis the study will examine. In our case, it would have also been possible to organize the data by paragraph or entire post, but we chose to treat each sentence as a unit of analysis in order to examine the data at a more granular level. The decision of what unit of analysis to use should be informed by the goals of the research project and the nature of the dataset (Neuendorf, 2002).

Literature review

Next, decisions need to be made about how to conceptualize and operationalize the variables of interest. For the purposes of coding emotional support, our starting point was a literature review of prior work in this topic. Prior to creating a coding scheme that would detect emotional support expressions, we performed an extensive review of the relevant literature on the definitions of emotional support and its sub-dimensions. Scholars have defined emotional support as the provision of constructs, including sympathy, understanding and empathy, reassurance, encouragement, concern, physical affection, relationality, confidentiality, prayer, and universality (Braithwaite, Waldron, & Finn, 1999; Coursaris & Liu, 2009; Cutrona & Suhr, 1994; Shaw, McTavish, Hawkins, Gustafson, & Pingree, 2000; Yalom, 1970). Based on previous definitions found in the literature review and through group discussions, we conceptualized emotional support along four main sub-categories: (1) empathy/sympathy, (2) encouragement/reassurance, (3) care/affection, and (4) universality/relationality.

¹The data were collected from two larger randomized clinical trials that examined health benefits of CHESS use. Recruitment was conducted from April 1, 2005, through May 31, 2007. Eligible subjects had been diagnosed with breast cancer or recurrence within 2 months of recruitment. 661 women agreed to participate in the studies and were randomly assigned to one of six conditions for a 6-month study period: (1) Internet-only, (2) human cancer mentor only, (3) CHESS information service only, (4) CHESS information and social support services, (5) CHESS information, social support, and interactive skill services (Full CHESS—having access to all major services provided by CHESS), and (6) human cancer mentor and full CHESS (Baker et al., 2011; McDowell, Kim, Shaw, Han, & Gumieny, 2010). Among the six experimental conditions, participants assigned to the three experimental conditions (4, 5, and 6) had access to the CHESS CMSS group, a text-based asynchronous bulletin board that allows users to anonymously share their thought and experience. The CMSS groups are monitored by a trained facilitator who ensures that discussions are supportive and do not contain inaccurate or harmful information. Any patient who did not have access to a computer with an Internet connection was provided a computer and free Internet access for the 6 months of the study. CHESS CMSS group messages and users' action log data collected during the 32 month study period. Over the course of the study, 504 people posted 18,064 messages in the CHESS CMSS groups. Although about half of the CHESS discussion group users were non-research subjects (e.g., discussion group facilitator and breast cancer survivors who participated in previous CHESS research), everyone who either wrote or read a message in the discussion group were included in this case study, because a social network cannot be accurately constructed if it excludes any person who contributes to the communication within the CMSS group. Mean age of 236 study participants was 51 years old ($SD = 9.19$). Majority of them were Caucasian (89.6%). More than half of participants (55.7%) had bachelor's or higher degree. On average, a study participant posted 21.4 ($SD = 54.5$) and read 466.2 ($SD = 934.7$) messages.

Once these sub-categories have been defined, the researchers should start creating working dictionaries of keywords that are likely to tap into these sub-categories. For example, “sorry” is categorized as an indicator of the empathy/sympathy category and “hope” would often represent for the encouragement category. Table 1 presents some keywords of each sub-category.

Content coding

Much of the expression effect research employs textual analysis programs to analyze the various linguistic dimensions of online messages (Alpers et al., 2005; Han et al., 2008; Lieberman & Goldstein, 2006; Shaw et al., 2007, 2006). However, there is a growing need to develop methodological approaches that are sensitive to the syntactical complexity involved in online discussion (Grimmer & Stewart, 2013). In fact, one of the largest criticisms of computer-aided content analysis is that it typically has low levels of validity (Short, Broberg, Cogliser, & Brigham, 2010) and struggles to capture the deeper meanings embedded in social texts (Lewis, Zamith, & Hermida, 2013). This is often due to an overreliance on “word counting” programs, as opposed to approaches that are attentive to the shifting meaning of words depending on their context and usage.

These problems exist both for studies that utilize basic dictionary approaches (Lieberman & Goldstein, 2006; Owen et al., 2005; Shaw et al., 2007, 2006), as well as the more sophisticated supervised learning methods currently being developed and refined (Quinn, Monroe, Colaresi, Crespin, & Radev, 2010). Regardless of the approach, all computer programs are limited in their ability to analyze the linguistic complexity involved in the construction of communication messages (Grimmer & Stewart, 2013; Lewis et al., 2013). Most of the computer programs are not capable of taking context into account, which is crucial for interpreting the meaning of the written communication (Alpers et al., 2005). As a result, studies that only utilize computer content analysis often struggle with the syntactical complexities of language when attempting to code for the presence of latent content, resulting in errors in the coding process and the resulting estimates of communication influence (Han et al., 2011). Below, we highlight our approach to computer-aided content analysis, which is designed to address these limitations.

Computer-aided content analysis software

Crucial to achieving high levels of reliability as well as internal and external validity is utilizing high-quality software or having the programing capabilities to create a personalized model (Grimmer & Stewart, 2013; Lewis et al., 2013). There are several programming options that allow for more complex computer-aided content analysis, for a detailed discussion of the types of software available see Krippendorff (2013). More recently, scholars have increasingly turned to supervised learning models (Quinn et al., 2010). These approaches have considerable strength, but still lack the sensitivity to appreciate the complexity of language or tap into the latent meanings in text (Grimmer & Stewart, 2013; Lewis et al., 2013).

Our particular approach to computer-aided content analysis is to use the highly customizable software InfoTrend, developed by David Fan (Fan, 1985, 1990, 1994). Other programs allow for similar coding construction (e.g., Wordstat), or those proficient in computer programing can create

Table 1. Sub-categories of emotional support and their keywords.

Emotional Support	Keywords
Empathy/Sympathy	empathy, sympathy, understand, sorry, worry, concern, etc.
Encouragement/ Reassurance	hope, wish, trust, congratulation, cheer, hang in there, keep stay strong, keep marching, don't give up etc.
Care/Affection	take care, hugs, kisses, love, etc.
Universality/Relationality	common, isolated, army of chessling, sisterhood, notalone, together, etc.

their own coding systems. InfoTrend allows the user to code for key ideas and idea combinations in the text through the implementation of a dynamic rule structure (Fan, Wyatt, & Keltner, 2001; Shah, Watts, Domke, & Fan, 2002). More specifically, InfoTrend uses computer language to enter (a) idea categories, (b) words that tap or reveal those idea categories, and (c) coding rules that allow pairs of ideas in the text to be combined to form more complex meaning. With these three components, human coders can create and refine specific coding rules.

As a result, InfoTrend can capture syntactical complexities of language by handling linguistic context, such as homographs (e.g., “shift,” a period at work vs. “shift,” to move quickly), heterophony (e.g., “bass,” a stringed instrument vs. “bass,” a freshwater fish), qualification (e.g., a physical “wound” vs. an emotional “wound”) and negation (e.g., “helping” vs. not “helping”) (Shah et al., 2002). The InfoTrend system has been utilized for analyzing news media coverage across diverse areas, such as information management business, government projects (Bengston, Potts, Fan, & Goetz, 2005), and academic research (Fan et al., 2001; Jaspersen, Shah, Watts, Faber, & Fan, 1998; Shah et al., 2002). Recently, this program has been successfully applied to analyze natural language used in online social support groups (Han et al., 2011; Kim et al., 2011; Namkoong et al., 2010, 2013; Yoo et al., 2014).

Creating a coding scheme

In our study, it was important to develop coding rules that could both accurately reflect the emotional support categories established through previous literature and be abstract enough to apply across a wide range of cases and contexts. At the same time, it was necessary to establish a precise enough set of coding rules that the analysis almost exclusively caught what we intended it to.

Different data sets, research goals, and the software being utilized may all lead to different customizations in a researcher’s coding approach. We detail our specific approach below as a means of demonstrating how others can employ a multi-faceted approach to creating coding rules. To achieve our goals we systematically followed a reoccurring four-step process that relied on a mixed method approach to ensure a thorough and deep understanding of the text as well as a reliable set of coding rules. As shown in Figure 1, the steps were as follows: (1) keyword search; (2) grounded examination; (3) rule creation; and (4) rule testing. Regardless of the specific iteration a researcher uses, we believe it is crucial that they include both inductive and deductive approaches and human and computer readings of the text. This mixed-method approach helps achieve the goals of being systematic, while accounting for nuance and deeper meaning.

Step 1: Keyword search

The first step is to identify a subset of data, which will serve as a starting point for rule creation. Our approach was to look for keywords that were likely to tap into the constructs we were interested in. After data has been loaded into the system, the InfoTrend software allows for keywords searches that produce a random selection of posts. The keywords were established a priori from the emotional support dictionaries. Once a keyword is entered, InfoTrend identifies every document in the provided data set that contains the term at least once. Alternatively, many programs now provide textual mapping and extraction tools, which allow the researcher to examine keywords inductively, based on which terms or concepts appear most frequently. We chose to have 100 randomly selected posts pulled from our data set in order to make navigation through the posts manageable.

Step 2: Grounded examination

Having retrieved a targeted subset of data, the next step is to thoroughly examine the data inductively in order to establish an understanding of how the terms are actually being used. This helps promote a deeper understanding of the meaning being constructed through discourse, as well the syntactical nuances that need to be captured by the coding rules. This should not be confused with the grounded theory method (Glaser & Strauss, 1967), although we adapt and incorporate some

of the principles and practices of grounded theory in our approach (see Charmaz, 2006). Of particular importance, we adhered to the *constant comparative method*, which entails an ongoing process of meticulously checking developing codes, categories, and insights against the data and against each other.

During this process, we discussed the language and phrases that were used to express emotional support and then began to formulate ideas about how we could capture these concepts. Our understanding of the unique terms, phrases, references, and thoughts included an examination of the context surrounding specific words and phrases. To code the data, we employ a system similar to *open coding* (Glaser & Strauss, 1967). We start by going through the data line by line, providing labels that closely represent what is actually being said (e.g., “I’m so sorry you are having such a hard time,” was coded as “sorry”). Initially, we labeled any terms, concepts, or ideas that might serve to represent any form of emotional support. In doing so, we remained open to the myriad of ways that emotional support could be expressed, and learned about various constructs that we could not have anticipated prior to open coding. As the coding process moved forward, we adopt an approach similar to *axial coding* (Strauss & Corbin, 1990), with the goal to construct more inclusive subcategories. Similar open codes were grouped together to construct a smaller set of emotional support subcategories (e.g., “empathy”), which served as “axes” around which open codes are “clustered.” This inductive process led to a deeper understanding of the data. The use of a deep reading early in the process is particularly important, even for machine learning approaches. There is no guarantee that the concepts or terms extracted by a computer-program will be significant or meaningful (Grimmer & Stewart, 2013).

It is important to note that our grounded examination has the same goals as many machine learning procedures. The goal of machine learning and related pattern recognition techniques focused on language processing is *text classification*—the process of assigning language within a text to specific categories (Sebastiani, 2002). The validity of these classification procedures can be improved through several steps. The first is *noise removal*, which is the practice of eliminating words or word combinations that are determined to be of little value, thereby obscuring the more important syntactical properties of the text (Yang, 1995). Next, researchers can test for *precision*, which calculates how frequently text segments are classified as belonging to a category that they do not actually belong to (i.e., Type I error), and *recall*, which calculates how frequently text segments should be categorized to a particular category, but are not (i.e., Type II error). Precision and recall are used to determine how well a machine learning algorithm is performing (Davis & Goadrich, 2006).

The primary difference in our approach is that humans, rather than software programs, learn from the data. Our approach assumes that humans can develop a more accurate understanding of how to classify human communication. For example, words that appear trivial, but are actually important, may be overlooked. Without developing a deeper understanding of the text, a researcher may make assumptions about a program’s performance, when they have a misunderstanding of the idiosyncratic use of certain words, phrases, or references within an online community. For example, inductive analysis discovered that within the CHESS breast cancer support group, ice cream was used as a symbolic representation of group membership and served as a reminder to enjoy life (McLaughlin, Hull, Namkoong, Shah, & Gustafson, 2016). Ice cream would not, typically, be expected to serve as a central mechanism through which emotional support could be delivered. It might, perhaps, be discarded as noise without a deeper familiarity of the specific way group members talked and acted toward each other.

Step 3: Rule creation

Having developed an initial understanding of the ways emotional support is expressed, researchers can then begin to capture the concepts of interest. The process of constructing rules will differ depending upon which software is being used. It is recommended that researchers either use software that gives them similar capabilities to the InfoTrend process described below, such as

Wordstat, or develop their own program. Constructing coding rules that accurately capture emotional support requires several steps.

InfoTrend provides the ability to create fully customizable dictionaries. At the most basic level, InfoTrend allows for the creation of an idea category that alerts the system to acknowledge and classify the identified string of characters. The idea category may contain a single term or several words. For example, an “emotion” idea category such as ‘positive emotion’ includes a wide range of terms such as “happy,” “excited,” “glad,” “pleased,” etc. A personal pronoun category such as “you” would contain a smaller set of word derivatives, such as “you, your, yourself,” etc. While a more abstract idea category such as “presence” would contain phrases such as “here for” or “be there,” etc. InfoTrend provides the option of choosing whether the program should identify the presence of a word in isolation or in all its various forms and derivatives. The exact nature of the derivatives, for example, whether letters can come before or after the chosen word, is highly customizable.

After defining the idea categories, rules are created by establishing a relationship between two terms, phrases, or concepts. The program allowed us to establish a set number of spaces in between terms as well as the direction in which they are related. By determining the order in which words appear, we were capable of establishing directionality. This function allowed us to capture emotional support as it occurs in natural language without erroneously capturing different concepts that used the same term(s). For example, the phrase “I am here for you” would be counted as the provision of emotional support while “I am so glad you are here for me” would not, specifically due to the order of the phrase “here for” in relation to the word “you.”

In addition, once these rules have been created, they can be used as an idea category in a new rule. This allowed us to create several layers for rules that captured complex expression unable to be identified by most traditional coding software.

Step 4: Rule testing

The next step is to check to see how the constructed rules are performing. Once an initial set of rules were established we tested the rules on a random selection of data to assess their performance. This again relied on a systematic inductive process, whereby we closely examined if our rules were actually capturing the concept we intended them to or if there were displays of emotional support that the system did not capture. If the rules were not performing proficiently, we would discuss as a group how the rules could be altered to more accurately measure the concepts. Regardless of the software used, it is imperative that the researcher frequently and thoroughly checks the validity of their coding rules (Short et al., 2010).

Repeat Steps 1–4

This process is repeated multiple times to develop rules around any given set of keywords. Further, once researchers make improvements on their coding rules, they should start the process again at step one with a new key term and examine these rules against a new set of message posts. These iterations test the performance of the coding rules, leading to greater precision in the computer’s application of the content analysis. Once a high level of consistency is achieved between the researchers’ “reading” and the system’s “coding” of a sample of the content, an independent coder should proceed to perform a reliability test, comparing the computer’s coding to the judgment of a human coder, with adjustments made for agreement by chance. Given our in-depth coding process, a high reliability score assures that the computer-coding accurately and consistently measures what we intended it to capture.

Reliability tests and external validity check

There is no hard rule on how large of a sample should be used to perform a reliability test (for a detailed discussion of this matter, see Neuendorf, 2002). For our study, reliability estimates were

conducted on a random subset of 200 discussion posts from the CHESS CMSS breast cancer group. Researchers hand coded this subset of discussion posts and compared their content coding of the data to the computer-aided content analysis, with any disagreement between the human and machine noted.

In our case, reliability tests between human and computer coding produced an estimate of 91.0% agreement across the four different categories (empathy: 90.9%, encouragement: 88.4%, care: 92.0%, universality: 95.0%). Scott's pi was also calculated by comparing the expected agreement by chance across the four coded categories with the actual agreement, which was determined to be 86.2% greater than by chance across these coding categories.

It is also important to check the external validity of the coding rules (Short et al., 2010). Seeing how the rules perform using a different data set can help establish external validity. To examine the external validity of our coding rules, the coding scheme was also tested against 200 discussion posts in a caregiver support group. These support groups messages were collected from different studies, which mainly focused on the effect of CHESS on lung cancer caregivers' psychological health benefits (DuBenske et al., 2008). The CMSS groups consisted of individuals who were caregivers for a lung cancer, breast cancer, or prostate cancer patient. Even though a caregiver's relationship to cancer is different than a cancer patient's, the same rules that were constructed for analysis of breast cancer patient data also worked quite well for caregivers' discussion messages, with an estimate of 87.4% agreement across the four different categories (empathy: 93.3%, encouragement: 88.3%, care: 86.3%, universality: 87.5%). Scott's pi was 78.9%, indicating a reliable coding. This demonstrates how performing a reliability test of coding rules against a related, but different data set can strengthen the confidence that your rules are not just reliable, but also possessed a higher level of external validity.

Combining the content coding with the action log data

While content coding allows researchers to figure out how concepts such as emotional support are expressed online, the content analysis itself does not reveal much about effects unless it is combined with data that allows the researcher to measure who wrote or read each message. It is therefore necessary for researchers to combine multiple datasets in order to make inferences about the consequences of online reception and expression. Below, we demonstrate our method to parse out the effects of engagement with an online social support group.

Specifically, we combined the content analysis of emotional support in the CMSS group with action log data gathered in the CHESS database management system. The action log data contains information about who posted each message and who clicked on any given message to read it, two main technological capabilities within the messaging system, which enable the participants to communicate with one another in CMSS groups. The action log file automatically records usage data at the level of individual post submission, page navigation, or clicks of hyperlinks. This file shows precisely who requested a page or a message by clicking hyperlinks, the date and time of the page call, and the requested Universal Resource Locator (URL) along with any system parameters sent as part the particular page view instance. These data allow us to establish the types of activity in which the user was engaged on the CHESS website.

To get a comprehensive view of all message expression and reception traffic, the reception data in the action log file were appended with the date/time, user ID, and message ID from the message database. An additional marker was added to the resulting table, which designated each log entry as a message read or written. As a result, our data contained records not just of the content of messages that participants read, but also the content of the messages they wrote. Last, using the message ID, this targeted traffic database was joined with the content analysis results described in the previous stage.

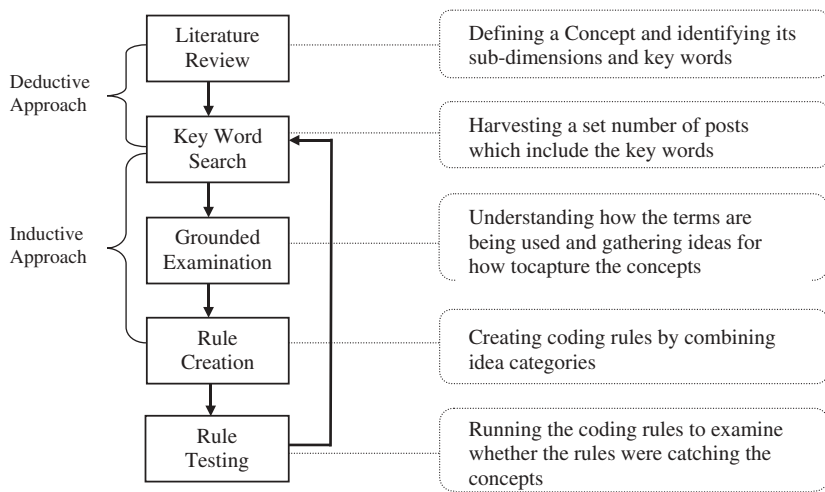


Figure 1. Computer-aided content analysis: Combination of deductive and inductive approaches.

Aggregated for each user, the resulting data contain an individual’s cumulative number of messages expressed and received, as well as her content coding scores across the subcategories of emotional support message expression and reception. In other words, these data provided the information about: (1) who posted a message, (2) when the message was posted, (3) which and how many coding categories were included in the message, and (4) who actually read the message. This enables us to examine actual writing and reading behaviors in the CMSS group and observe the changes in communication behaviors throughout the study period. As shown in Figure 2, message reception had significantly increased in the first three weeks and decreased gradually throughout the next 21 weeks of the clinical trial. Compared to message reception, expressive behavior had been low but steady across the study period.

In addition, the resulting data have a separate count of the number of messages expressed and received by each user as well as a sum of the content analysis scores for expression and reception for each of the emotional support subcategories. Extending what we examined in Figure 2, this merged data allow us to see the changes in the volume of the emotional support expression (Figure 3) and reception (Figure 4) over time. For example, among the four subcategories of emotional support, encouragement, and caring were expressed and received more than empathy and universality throughout the 24 weeks.

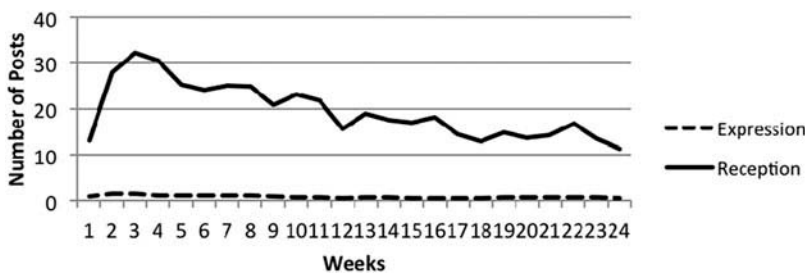


Figure 2. Number of writing and reading posts in a CMSS group over the study period (24 weeks).

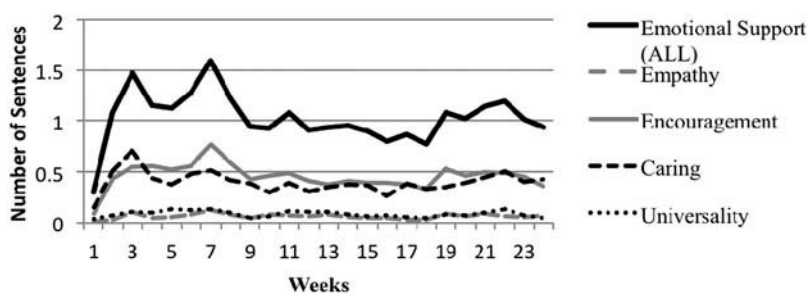


Figure 3. Number of sentences that participants wrote social support expression over the study period (24 weeks).

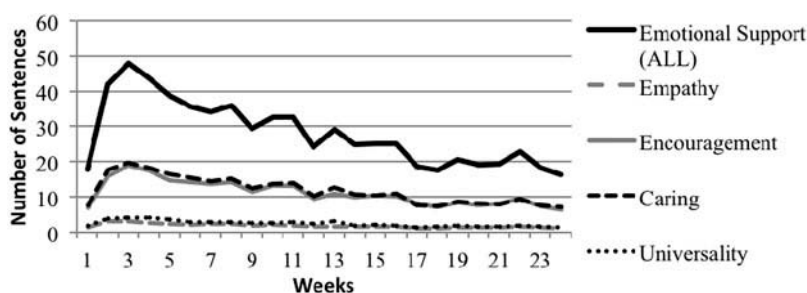


Figure 4. Number of sentences that include social support expression and were read by participants over the study period (24 weeks).

Merging survey data into the coding/log data

On its own, content analysis and action log data analysis cannot provide much more than a descriptive account of communication messages. Combining this descriptive data with survey data creates the possibility of testing the causal effect of communication messages on specific outcome variables (Neuendorf, 2002). The combined data allows researchers to examine the multi-layered effects of online expression and reception.

Recent studies have confirmed the validity and demonstrated the feasibility of this methodology to examine both antecedents (Kim et al., 2011) and consequences of expression and reception (Han et al., 2011; Namkoong et al., 2010, 2013). If there are multiple waves of data collection, message production, and consumption can be calculated based on the intervening time period and integrated into the survey data on that basis (Yoo et al., 2013). Merging survey data into the content coding and action log data sets, more specifically, researchers have investigated the psychological benefits of emotional and informational support expression and reception occurring within CMSS groups for breast cancer patients (Han et al., 2011; Namkoong et al., 2010, 2013; Yoo et al., 2014). For example, Namkoong et al. (2013) found that expression, but not reception, of emotional support increases cancer patients' perceived bonding with other patients in CMSS groups, which in turn mediates the effects on positive coping strategies. Similarly, Han et al. (2011) found that expression, but not reception of empathy, a sub-category of emotional support, has positive impacts on reducing cancer-related concerns. Furthermore, with analysis of interaction effect between empathy expression and reception, they found that a combination of empathy expression and reception is important to attaining optimal benefits from CMSS groups. We presented a list of studies that employed the triangulation of content coding, action log, and survey data sets in Table 2. The studies exemplify

Table 2. Studies that employed the triangulation of three data set analysis (computer-aided content analysis, action log data analysis, and survey data analysis).

Author(s)	IVs	DVs	Note
Namkoong, Shah, & Gustafson (2013)	Network centrality in open and targeted communication network	Perceived bonding; Cancer information competence; Coping behaviors	Social network analysis
Namkoong et al. (2016)	Family environment; Perceived availability of social support	Network centrality in open and targeted communication network	Social network analysis
Han et al. (2011)	Emphatic expression and reception	Cancer-related concerns	
Kim et al. (2011)	Age; Education; Coping strategy; Perceived availability of social support	Emotional support expression and reception	
Kim et al. (2012)	Supportive message expression and reception	Cancer-related concern; Coping strategy	
McLaughlin et al. (2013)	Expression of “deferring control to God”	Cancer-related concern; Quality of life	
McLaughlin et al. (2016)	Religious expression	Perceived bonding; Positive coping behaviors	
Namkoong et al. (2010)	Informational support expression and reception	Emotional wellbeing	
Namkoong et al. (2013)	Emotional support expression and reception	Perceived bonding; Positive coping behaviors	
Yoo et al. (2013)	Age; Living situation; Comfort level with computer and the Internet; Coping strategy	Emotional support expression	Longitudinal analysis (Growth curve modeling)
Yoo et al. (2014)	Emotional support expression and reception	Cancer-related concern; Quality of life	

how our methodology can be used to examine the multilayered effect of the act of expression and reception and demonstrated the roles of each dataset in CMC research.

Online social network analysis with expression and reception data

To produce a more detailed conceptualization of the complex social relationships formed in CMSS groups, we advocate for also including a social network analysis. Social network analysis identifies the patterns and strength of social relations that develop among the CMSS group participants. In our study, because the message production and reception data can be used to reproduce conversational networks, the action log data facilitates performing a social network analysis of group participants without their retrospective answers about social interactions in the CMSS group. This step allows for a more accurate depiction of participants’ roles and positions in their online social network and allows assessment of their social capital, reflecting the conception of social capital as a valued resources embedded in network ties measurable by any member’s access to these resources (Kawachi, Subramanian, & Kim, 2008; Lin, 1999). This data can then be used to examine the relationship between a participants’ social network position and their psychological health outcomes.

Previous work has constructed social networks for online support groups using content analysis. For example, Bambina (2007) examined the messages posted on the SOL-Cancer (Support OnLine) Forum and successfully visualized online communication networks and concept-specific social networks (e.g., emotional and informational support networks). In addition, she identified the forum users’ roles in the online communication networks. While past research provides important advances for the social network analysis of CMC messages, there is room for improvement in two regards. First, online social networks have usually been constructed by the content of message, not group member’s actual use data. However, online social network constructed solely based on posted messages (writing) is inevitably incomplete, because it is possible that a message would not include any cue about the recipient(s) of the message. Even when a message specifies who the intended recipient(s) is, there is no guarantee that the specified recipient actually read the message. More

important, in many online social groups, there are numerous lurkers who do not write any messages, but are still engaged in message reception. In these cases, even though these lurkers may actively read messages and receive social support, they would not be included in the network analysis. Second, previous research has rarely examined the relationship between the group participants' network positions and their psychological outcomes. Thus, it also leaves some unanswered questions about the consequences of exchanging social support through an online social network.

With a public dataset of a social network site, Facebook, Lewis and his colleagues (2008) conducted social network analysis considering several distinctive features of Facebook data, such as *Facebook friends* and *Picture friends*. This study presented how we can construct online social networks based on actual social media user behaviors, such as making Facebook friends or uploading a photograph and identifying people with a "tag" function. However, because it used a specific service-based secondary dataset, it is hard to apply to other online social relationships solely developed from message expression (i.e., writing) and reception (i.e., reading), the most typical CMC behaviors.

Using the analytical methods described here, we seek to improve online social network analysis to reflect the complex relationships that are formed in online communities. Furthermore, we can construct online social networks developed in the CMSS group with two communication approaches—mass or group communication (one-to-many) and interpersonal communication (one-to-one). In reality, both can occur simultaneously, but distinguishing the two forms of communication reflect the complex relationships that are formed online. On one hand, because these posts are present for all group members to read, the social network can be examined as a form of mass communication. Thus, we construct online social networks with this mass or group communication approach, identifying the readers of a message regardless of its intended recipient. Namkoong and his colleagues (2016) named it an *open communication network*. On the other hand, members may direct specific posts to other members by referencing the intended targets in the message. In other words, members can use the CMSS group as an interpersonal communication tools even while they are aware that all group members can access these messages. Thus, we construct online social networks with an interpersonal communication approach, analyzing the intended flow of information. This network was dubbed as a *targeted communication network* (Namkoong et al., 2016). In sum, the two communication networks—open and targeted—allow us to identify and distinguish two prominent communication patterns that often occur simultaneously in online discussion posts.

Open communication network

In a text-based message board type CMSS group, such as the one discussed in this study (similar to comments sections under news articles), even posts that are specifically directed to another individual can function as broadcast messages. Accordingly, CMC in the support group can be regarded as mass communication, assuming that a person is aware of the potential for a written message to be read by some or all of the group members. As a result, the most effective way to analyze the data from a mass communication perspective is to define the social network by message reception. By identifying who read what messages, we are able to track the extent to which participants are engaged with communication, with whom and to what effect. In other words, the social network can be constructed by specifying that a person read others' messages (in-degree) and that their messages were read by other participants (out-degree) (Namkoong et al., 2016).

Figure 5a shows the open communication network. There are three layers in the online social network graphs. Individuals who have high in- and out-degree are located in the central positions in the network. People in the core part of the network (Area A in Figure 5a) are those who actively read others' messages and whose messages were read by many other group members. People in the middle layer are represented as small spheres with a number of connections (Area B in Figure 5a). This indicates that they have low out-degree and high

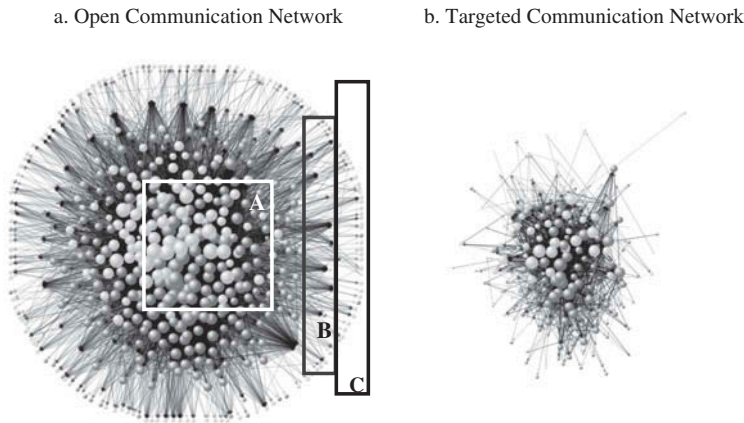


Figure 5. Online social networks (Namkoong et al., 2016). Dark sphere: Participants; Bright sphere: Non participants; Open communication network: Vertices (504), Edges (14358), In-degree ($M = 28$; Med = 11), Out-degree ($M = 28$; Med = 16); Targeted communication network: Vertices (198); Edges (2,253); In-degree ($M = 11$; Med = 7); Out-degree ($M = 11$; Med = 4).

in-degree, which means they read a lot of messages, but produced few messages that were read by others. Thus, the second layer shows the “lurkers” in the CMSS group. Finally, those who did not actively interact with other members, people who seldom read messages or produced messages, are located in the peripheral positions (Area C in Figure 5a).

Targeted communication network

Focusing only on message reception, however, provides an incomplete picture because messages that are explicitly intended for a specific individual represent a different type of social relationship. Therefore by addressing the identified targets in a message, the CMSS group can be conceived of as an interpersonal communication channel. The interpersonal communication network can be established by identifying the referenced persons or targets in a message. In other words, the social network can be constructed by indicating that a person specifies a target in his or her messages (out-degree) and that a person is referred to as a target of the messages (in-degree). The out-degree is counted only when the targeted person actually read the message (Namkoong et al., 2016).

Figure 5b visualizes the targeted communication network. In this model, social network analysis is conducted based on both message expression and reception. For example, if a person writes a message with a specific target(s), her writing activity is represented as her out-degree and as the message readers' in-degree. In the network graph, the arrow starts from the writer and points to the target of the message, indicating the direction the information flows. In this way, the online social networks visually represents group participants' interpersonal communication patterns. As shown in the figures, most of the people who are located in the second layer and peripheral areas of the open communication network disappear. In other words, those who develop personal relationships in the CMSS group are more actively engaged in the group discussion than those who participate in the group without interpersonal connections.

Furthermore, our analytic approach allows us to examine actual flow of “content” during interactions that happen over online social networks. In other words, we can create online social networks based on the content of the message post, such as emotional and informational support, beyond using a post as a unit of network analysis. We can also distinguish the content-based networks based on sub-categories of emotional support expressions, such as empathy, encouragement, and universality. These online social network analyses provide unique opportunities to observe actual flow of content without asking participants' retrospective report of their social

interaction. They also allow assessment of differences in social networking effects in terms of the content of messages exchanged in the CMSS group. In addition, this method permits calculation of a person's network size and centrality as measures of individual-level social capital (Namkoong et al., 2016), and allows examination of the relationship between individuals' social capital and their psychological health benefits (Namkoong et al., 2013).

Discussion

While computer-mediated communication (CMC) has been of significant interest to numerous scholars, the methodological sophistication used to study a variety of CMC phenomenon has not always kept pace with complexity inherent in online contexts. This article provides an in-depth explanation of an innovative methodology that can help scholars examine the multi-layered effects of expression and reception in CMC. This methodology combines: (1) flexible and precise computer-aided content analysis data; (2) action log data of the message posting and reading based on keystrokes and clicks; and (3) longitudinal survey data. In addition, the current study shows how online social network analysis can be utilized for disentangling message expression and reception from the perspective of open and targeted communication, which can then be used to examine the effects of social network ties on psychological outcomes.

First, this article sought to demonstrate the need to further conceptualize the complex way in which social relations are formed and maintained through message expression and reception. As such, we believe this methodology provides a blueprint for thinking about how scholars can disentangle the effects of message expression from reception, as well as conceptualize online discussion posts as more complex relational networks that require several levels of analysis. Whether occurring synchronously or asynchronously, CMC is a process where message expression and reception often interact in complex ways.

As more and more scholars begin to turn their attention to the effects of expression, it is important to distinguish them from the effects arising from message reception and the expression/reception nexus (Han et al., 2011). The action log data collection system used in this study enabled us to identify which participant wrote and/or read each message. In this manner, we could begin to disentangle the effects of both message expression and reception. Further, the structure of the data allows the researcher to construct system use variables (e.g., expression, reception) in a variety of ways, depending upon the theoretical perspective taken or the research question being asked. For example, for some analyses, the researcher may want to utilize a measure that reflects the total number of posts providing emotional support. For other research questions however, the researcher may be interested in the ratio of expressions of emotional support to receptions of emotional support. Coupled with the survey data, the ability to construct variables that reflect the ways in which members of online discussion groups actually use these platforms may provide valuable insight about the potential effects of such use, which may otherwise go unexplored.

Another common limitation is the inability of most computer-assisted content analysis programs to handle the syntactical and logistic complexities involved in the language used in online group communication, thus compromising the internal validity of the analysis. This can result both from using basic, traditional textual analysis programs, as well as from using more sophisticated machine learning approaches (Grimmer & Stewart, 2013). There is, therefore, a need for scholars to develop better mixed-methods approaches to computer-aided content analysis (Lewis et al., 2013). By incorporating qualitative, inductive readings throughout our analysis, we demonstrate how a multi-faceted approach can be systematic while also capturing the deeper meaning embedded in human communication.

Our use of InfoTrend illustrates how scholars can construct coding rules that not only have high levels of reliability, but also high levels of internal and external validity. The system's ability to code for complex syntax allowed us to accurately capture the larger context of emotional support expressions, producing internal validity. At the same time, we were able to keep our rules abstract

enough that they displayed a high level of reliability against a related, but distinct data set. Other computer-aided content analysis programs also allow for highly customized rule creation (see Krippendorff, 2013), and scholars are increasingly becoming more proficient in the creation of their own programs. It is highly recommended that researchers who are serious about achieving high levels of validity with computer-aided content analysis invest in, or develop the programming skills necessary to create, a program that can account for the syntactical complexity involved in online discourse.

Achieving high levels of validity is not just a matter of employing the right program. Communication constructs can take on a wide range of forms in an online context. It is important to be able to deal with the idiosyncrasies that are likely to be present in online social groups. Employing a combination of deduction and induction can help to account for some of the difficulties involved in studying online communication. The coding scheme developed in this study provides an example of how coding rules can be strengthened by using a mixed method approach. Regardless of which program a researcher uses, we hope the illustrations of our process highlight some of the ways that content coding can be improved through the combination of deduction and inductive approaches.

Last, but most notably, this study introduces a new methodological approach for social network analysis. By analyzing action log data, we can examine online social networks based on actual message reception as well as production data. Given that previous research on CMC network analysis has been based on the contents of messages or participants' retrospective answers, this approach allows us to construct more objective, accurate, and comprehensive social relations developed in a CMSS group because the log data analysis includes not only active participants but also lurkers who are not captured by the traditional content-based network analysis. In addition, applying the results of computer-aided content analysis to network analysis, we can produce the concept-specific network in a CMSS group, which allows us to examine the actual flow of content through the online social ties. The inclusion of survey data also allows us to examine the relationship between the participants' network positions and their psychological health outcomes, which has been seldom investigated in much of social network analysis research.

Because computer-mediated communication allows for the possibility of several forms of communication flow, we constructed two different types of CMC networks—open and targeted communication networks—for different types of content embedded in the social networks (Namkoong et al., 2016). Distinguishing the two networks provides important insights that elucidate the different communication patterns that occur in online discussion groups. Thus, it is important for scholars to continue to account for both of these two models and content-based communication network analysis when examining the effects of CMSS group communication.

It should be also noted that the social network analysis presented in this article is only one possibility out of many potential applications of our methodological approach. For example, one can further construct a two-mode network that consists of different *mode* of nodes (e.g., “messages” and “people”) or a longitudinal panel network that allows us to examine the patterns and predictors of emotional and informational support tie formation.

The methodology employed here relied on the ability to access participants' action-log data. Many researchers may not have access to comparable data sets without either developing social networking platforms such as CHESS that they can use for research or gaining access to these networks through cooperative agreements with existing organizations (e.g., the American Cancer Society) or via a social network's API (i.e., application program interface). In addition, there are growing opportunities for scholars to come up with creative ways to pair content analysis with survey data (Niederdeppe, 2016), also allowing for novel insights. An endless stream of digital trace data is being created at any given moment (Freelon, 2014; Howison, Wiggins, & Crowston, 2011), although gaining access to some of this data (e.g., Facebook data) is becoming more restricted as commercial interests and privacy concerns take hold. Nonetheless, we are advocates for the coming age of *computational social science*, and believe this turn holds special promise for the field of

communications, given how well positioned it is for these emergent methods (see Lazer et al., 2009; Shah, Cappella, & Neuman, 2015).

Indeed, we prefer the term computational social science to the term “big data,” because the latter seems to imply that a research project must be performed on a grand scale or involve terabytes of data in order to be meaningful. Some of the most important research questions may be examined within a relatively small social network, however. Micro-communities are formed around a myriad of topics, issues, or identities (e.g., physical and mental health support groups, political discussion groups, brand communities, and so on). Scholars with some resources, and available budgets, have more options, of course. For example, scholars have found innovative ways of tracking mobile phone usage patterns by having participants install a mobile app into their phone (e.g., Boase & Kobayashi, 2012; Wells & Thorson, 2015). These data can be used to examine a wide range of interactions, including the effects of message expression and reception.

There are still numerous possibilities for those with fewer resources, if they are unable to build relationships and gain access to a micro-community. Web scraping tools such as Context Miner (<http://contextminer.org>) or Tubekit (<http://tubekit.org>) allow researchers to capture user interaction data for those who use an open access system. For example, if the researcher is studying a message board, they could analyze the messages contained in a thread in which a participant replied, examining the content of others (reception effects) and the content of the participant (expression effects). A similar approach could be used on Twitter, where any tweet a participant favorited, retweeted, replied to, mentioned, or tweeted would be captured and coded. This coded content could then be paired with responses to longitudinal surveys disseminated to participants. This data, while not ideal, would still provide a pathway to performing the type of analysis presented here. Undoubtedly, there are numerous other innovative approaches that researchers could develop based on the methodological procedures outlined here.

Fast growing social media platforms, like Twitter and Instagram, provide new opportunities to apply and advance our analytic approaches, although this social media work will differ slightly from the discussion forum analyzed in our study. For example, hashtags (i.e., user-defined keywords widely used in social media) can help filter out relevant messages for coding scheme creation. Hashtags may also supplement existing content coding schemes as another way to classify messages (Borge-Holthoefer, Magdy, Darwish, & Weber, 2015). Future studies are needed to employ and advance our proposed method to solve challenges in analyzing new social media. For example, recent research has shown a parsimonious rule-based sentiment assessment model to be effective in analyzing social media messages, considering they are often shorter by design, and given their unique spelling, composition and grammar (Hutto & Gilbert, 2014).

The method presented in this article is quite complex and requires a significant investment by researchers. This methodology, however, provides a sophisticated way to accurately and reliably examine message expression and reception effects on very large data sets in a variety of contexts. For example, while our methodology was largely discussed in the context of health communication, the method presented in this study can be applied to numerous kinds of computer mediated communication research. This methodology could be utilized to understand the ways in which participation in online political exchanges influence civic participation. This approach is particularly well suited to address questions regarding the effects of civil and uncivil online political discourse by exploring the effects of both expression and reception of content coded messages in terms of a variety of psychological and behavioral outcomes. In summary, although the method detailed here is demonstrated through the use of a case study, this method provides a means to examine expression and reception effects in a wide range of CMC contexts. We highlighted a range of methodological issues and concerns, and offered recommendations for how scholars can best appreciate and examine the complex communication patterns that occur online. Appendix 1 provides a summary of some of our recommended “best practices” for this methodology.

In conclusion, we believe this methodological approach offers a clear example of some of the ways researchers can overcome the significant limitations generally found in the use of computer-aided content

analysis and examinations of online discussion effects. This approach allows for an analysis that respects the complicated nature of online discussions and the ways in which participation in discussion groups can influence psychological and behavioral outcomes. Furthermore, the construction of online social networks based on message expression and reception allows for a more detailed conceptualization of the complex social relationships formed in CMSS groups. We hope that our presentation of this methodology provides a blueprint that can help other researchers adopt a similar approach. At the very least, this article raises attention to some critical issues involved in employing methodological approaches that can provide a greater understanding of the effects of computer-mediated communication.

Funding

National Cancer Institute (P50 CA095817-05).

References

- Alpers, G. W., Winzelberg, A. J., Classen, C., Roberts, H., Dev, P., Koopman, C., & Taylor, C. B. (2005). Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior*, 21, 361–376. doi:10.1016/j.chb.2004.02.008
- Baker, T., Hawkins, R. P., Pingree, S., Roberts, L., McDowell, H., Shaw, B. et al., (2011). Optimizing eHealth breast cancer interventions: Which types of eHealth services are effective? *Translational Behavioral Medicine*, 1, 134–145. doi: 10.1007/s13142-010-0004-0
- Bambina, A. (2007). *Online social support: The interplay of social networks and computer-mediated communication*. New York, NY: Cambria Press.
- Bengston, D. N., Potts, R. S., Fan, D. P., & Goetz, E. G. (2005). An analysis of the public discourse about urban sprawl in the united states: Monitoring concern about a major threat to forests. *Forest Policy and Economics*, 7(5), 745–756. doi:10.1016/j.forpol.2005.03.010
- Boase, J., & Kobayashi, T. (2012). Mobile communication networks in Japan and America. *China Media Research*, 8, 90–98.
- Borge-Holthoefer, J., Magdy, W., Darwish, K., & Weber, I. (2015). *Content and network dynamics behind Egyptian Political Polarization on Twitter*. Presented at the Eighth International AAAI Conference on Weblogs and Social Media (pp. 700–711), New York, New York, USA. ACM Press. doi:10.1145/2675133.2675163
- Braithwaite, D., Waldron, V., & Finn, J. (1999). Communication of social support in computer-mediated groups for people with disabilities. *Health Communication*, 11(2), 123–151. doi:10.1207/s15327027hc1102_2
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks, CA: Sage.
- Coursaris, C. K., & Liu, M. (2009). An analysis of social support exchanges in online HIV/AIDS self-help groups. *Computers in Human Behavior*, 25(4), 911–918. doi:10.1016/j.chb.2009.03.006
- Cutrona, C. E., & Suhr, J. A. (1994). Social support communication in the context of marriage: An analysis of couples' supportive interactions. In B. R. Burleson, T. L. Albrecht, & I. G. Sarason (Eds.), *Communication of social support: Messages, interactions, relationships, and community* (pp. 113–135). Thousand Oaks, CA: Sage.
- Davis, J., & Goadrich, M. (2006). *The relationship between precision-recall and ROC curves*. Proceedings of the 23rd International Conference on Machine Learning, Pittsburg, PA.
- DuBenske, L. L., Wen, K., Gustafson, D. H., Guarnaccia, C. A., Cleary, J. F., Dinauer, S. K., & Mctavish, F. M. (2008). Caregivers' differing needs across key experiences of the advanced cancer disease trajectory. *Palliative & Supportive Care*, 6(3), 265–272. doi:10.1017/S1478951508000400
- Fan, D. P. (1985). Ideodynamics: The kinetics of the evolution of ideas. *Journal of Mathematical Sociology*, 11(1), 1–24. doi:10.1080/0022250X.1985.9989978
- Fan, D. P. (1990). Information processing expert system for text analysis and predicting public opinion based on information available to the public. *U.S. Patent 4,930,077*. Washington, DC: U.S. Patent and Trademark Office.
- Fan, D. P. (1994). Information processing analysis system for sorting and scoring text. *U.S. Patent 5,371,673*. Washington, DC: U.S. Patent and Trademark Office.
- Fan, D. P., Wyatt, R. O., & Keltner, K. (2001). The suicidal messenger how press reporting affects public confidence in the press, the military, and organized religion. *Communication Research*, 28(6), 826–852. doi:10.1177/009365001028006005
- Fishbein, M., & Cappella, J. N. (2006). The role of theory in developing effective health communications. *Journal of Communication*, 56, S1–S17. doi:10.1111/j.1460-2466.2006.00280.x
- Freelon, D. (2014). On the interpretation of digital trace data in communication and social computing research. *Journal of Broadcasting & Electronic Media*, 58, 59–75. doi:10.1080/08838151.2013.875018

- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick, NJ: Transaction.
- Gottlieb, B. H., & Bergen, A. E. (2010). Social support concepts and measures. *Journal of Psychosomatic Research*, 69, 511–520. doi:10.1016/j.jpsychores.2009.10.001
- Grimmer, J., & Stewart, B. (2013). Text as data: Promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297. doi:10.1093/pan/mps028
- Gustafson, D., Wise, M., McTavish, F., Taylor, J. O., Wolberg, W., Stewart, J., ... Bosworth, K. (1994). Development and pilot evaluation of a computer-based support system for women with breast cancer. *Journal of Psychosocial Oncology*, 11(4), 69–93. doi:10.1300/J077V11N04_05
- Gustafson, D. H., Hawkins, R., Boberg, E., Pingree, S., Serlin, R. E., Graziano, F., & Chan, C. L. (1999). Impact of a patient-centered, computer-based health information/support system. *American Journal of Preventative Medicine*, 16, 1–9. doi:10.1016/S0749-3797(98)00108-1
- Gustafson, D. H., Hawkins, R., McTavish, F., Pingree, S., Chen, W. C., Volrathongchai, K., ... Serlin, R. C. (2008). Internet-based interactive support for cancer patients: Are integrated systems better? *Journal of Communication*, 58(2), 238–257. doi:10.1111/j.1460-2466.2008.00383.x
- Gustafson, D. H., McTavish, F., Hawkins, R., Pingree, S., & Arora, N. (1998). Computer support for elderly women with breast cancer. *Jama-Journal of the American Medical Association*, 280(15), 1305. doi:10.1001/jama.280.15.1305
- Han, J. Y., Shah, D. V., Kim, E., Namkoong, K., Lee, S. Y., Moon, T. J., ... Gustafson, D. H. (2011). Empathic exchanges in online cancer support groups: Distinguishing message expression and reception effects. *Health Communication*, 26, 185–197. doi:10.1080/10410236.2010.544283
- Han, J. Y., Shaw, B. R., Hawkins, R. P., Pingree, S., McTavish, F., & Gustafson, D. H. (2008). Expressing positive emotions within online support groups by women with breast cancer. *Journal of Health Psychology*, 13(8), 1002–1007. doi:10.1177/1359105308097963
- Holland, K., & Holahan, C. (2003). The relation of social support and coping to positive adaptation to breast cancer. *Psychology & Health*, 18(1), 15–29. doi:10.1080/0887044031000080656
- Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, 12, 767–797.
- Hutto, C. J., & Gilbert, E. (2014). *Vader: A parsimonious rule-based model for sentiment analysis of social media text*. Presented at the Eighth International AAAI Conference on Weblogs and Social Media, Eighth International AAAI Conference on Weblogs and Social Media, Ann Arbor, MI, USA.
- Jasperson, A. E., Shah, D. V., Watts, M., Faber, R. J., & Fan, D. P. (1998). Framing and the public agenda: Media effects on the importance of the federal budget deficit. *Political Communication*, 15(2), 205–224. doi:10.1080/10584609809342366
- Kawachi, I., Subramanian, S. V., & Kim, D. (2008). *Social capital and health*. New York, NY: Springer.
- Kim, E., Han, J. Y., Moon, T. J., Shaw, B., Shah, D. V., McTavish, F. M., & Gustafson, D. H. (2012). The process and effect of supportive message expression and reception in online breast cancer support groups. *Psycho-Oncology*, 21(5), 531–540. doi:10.1002/pon.1942
- Kim, E., Han, J. Y., Shah, D. V., Shaw, B., McTavish, F. M., Gustafson, D. H., & Fan, D. (2011). Predictors of supportive message expression and reception in an interactive cancer communication system. *Journal of Health Communication*, 16(10), 1106–2011. doi:10.1080/10810730.2011.571337
- Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Thousand Oaks, CA: Sage.
- Lasswell, H. D. (1950). *Politics: Who gets what, when, how*. New York, NY: Whittlesey House.
- Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... Van Alstyne, M. (2009). Life in the network: The coming age of computational social science. *Science*, 323(5915), 721–723. doi:10.1126/science.1167742
- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook. com. *Social Networks*, 30(4), 330–342. doi:10.1016/j.socnet.2008.07.002
- Lewis, S., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting and Electronic Media*, 57, 34–52. doi:10.1080/08838151.2012.761702
- Lieberman, M. A., & Goldstein, B. A. (2006). Not all negative emotions are equal: The role of emotional expression in online support groups for women with breast cancer. *Psycho-Oncology*, 15, 160–168. doi:10.1002/pon.932
- Lin, N. (1999). Building a network theory of social capital. *Connections*, 22(1), 28–51.
- McDowell, H., Kim, E., Shaw, B., Han, J. Y., & Gumieny, L. (2010). Predictors and effects of training on an online health education and support system. *Journal of Computer-Mediated Communication*, 15(3), 412–426. doi:10.1111/j.1083-6101.2010.01516.x
- McLaughlin, B., Hull, S., Namkoong, K., Shah, D. V., & Gustafson, D. H. (2016). We all scream for ice cream: Physical desires and positive identity negotiation in the face of cancer. In A. Novak, & I. J. El-Burki (Eds.), *Defining identity and the changing scope of culture in the digital age* (pp. 81–98). Hershey, PA: IGI Global.
- McLaughlin, B., Yang, J., Yoo, W., Shaw, B., Kim, S. Y., Shah, D., & Gustafson, D. (2016). The effects of expressing religious support online for breast cancer patients. *Health Communication*, 31(6), 762–771. doi:10.1080/10410236.2015.1007550

- McLaughlin, B., Yoo, W., D'Angelo, J., Tsang, S., Shaw, B., Shah, D., ... Gustafson, D. (2013). It is out of my hands: How deferring control to God can decrease quality of life for breast cancer patients. *Psycho-Oncology*, 22(12), 2747–2754. doi:10.1002/pon.3356
- Namkoong, K., McLaughlin, B., Yoo, W. H., Hull, S., Shah, D. V., Kim, S. C., ... Gustafson, D. H. (2013). The effects of expression on perceived bonding: How computer mediated social support shapes cancer patients' coping strategies. *Journal of National Cancer Institute Monograph*, 47, 169–174. doi:10.1093/jncimonographs/igt033
- Namkoong, K., Shah, D. V., & Gustafson, D. H. (2013, November). *Social connections in an online cancer community: The effects of individual-level of social capital on psychological health benefits*. Paper presented at the annual meeting of the National Communication Association, Washington, DC.
- Namkoong, K., Shah, D. V., & Gustafson, D. H. (2016). Offline social relationships and online cancer communication: Effects of perceived social and family support on online social relationship building. *Health Communication*, 1–8. doi:10.1080/10410236.2016.1230808
- Namkoong, K., Shah, D. V., Han, J. Y., Kim, S. C., Yoo, W., Fan, D., ... Gustafson, D. H. (2010). Expression and reception of treatment information in breast cancer support groups: How health self-efficacy moderates effects on emotional well-being. *Patient Education and Counseling*, 81(Supplement 1), S41–S47. doi:10.1016/j.pec.2010.09.009
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.
- Niederdeppe, J. (2016). Meeting the challenge of measuring communication exposure in the digital age. *Communication Methods & Measures*, 10, 170–172. doi:10.1080/19312458.2016.1150970
- Owen, J. E., Klapow, J. C., Roth, D. L., Shuster, J. L., Bellis, J., Meredith, R., & Tucker, D. C. (2005). Randomized pilot of a self-guided internet coping group for women with early-stage breast cancer. *Annals of Behavioral Medicine*, 30(1), 54–64. doi:10.1207/s15324796abm3001_7
- Pingree, R. J. (2007). How messages affect their senders: A more general model of message effects and implications for deliberation. *Communication Theory*, 17, 439–461. doi:10.1111/j.1468-2885.2007.00306.x
- Price, V., & Cappella, J. N. (2002). Online deliberation and its influence: The electronic dialogue project in campaign 2000. *IT & Society*, 1(1), 303–329.
- Quinn, K., Monroe, B., Colaresi, M., Crespin, M., & Radev, D. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54, 209–228. doi:10.1111/j.1540-5907.2009.00427.x
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1–47. doi:10.1145/505282.505283
- Shah, D. V. (2016). Conversation is the soul of democracy: Expression effects, communication mediation, and digital media. *Communication and the Public*, 1(1), 12–18. doi:10.1177/2057047316628310
- Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 6–13. doi:10.1177/0002716215572084
- Shah, D. V., Watts, M. D., Domke, D., & Fan, D. P. (2002). News framing and cueing of issue regimes: Explaining Clinton's public approval in spite of scandal. *Public Opinion Quarterly*, 66(3), 339–370. doi:10.1086/341396
- Shaw, B., Han, J. Y., Kim, E., Gustafson, D., Hawkins, R., Cleary, J., ... Lumpkins, C. (2007). Effects of prayer and religious expression within computer support groups on women with breast cancer. *Psycho-Oncology*, 16(7), 676–687. doi:10.1002/pon.1129
- Shaw, B., Hawkins, R., McTavish, F., Pingree, S., & Gustafson, D. (2006). Effects of insightful disclosure within computer mediated support groups on women with breast cancer. *Health Communication*, 19(2), 133–142. doi:10.1207/s15327027hc1902_5
- Shaw, B. R., McTavish, F. M., Hawkins, R. P., Gustafson, D. H., & Pingree, S. (2000). Experiences of women with breast cancer: Exchanging social support over the CHESS computer network. *Journal of Health Communication*, 5(2), 135–159. doi:10.1080/108107300406866
- Short, J. C., Broberg, J. C., Coglisier, C. C., & Brigham, K. H. (2010). Construct validation using Computer-Aided Text Analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods*, 13, 320–347. doi:10.1177/1094428109335949
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Wells, C., & Thorson, K. (2015). Combining big data and survey techniques to model effects of political content flows in Facebook. *Social Science Computer Review*, 1–20. doi:10.1177/0894439315609528
- Yalom, I. D. (1970). *The theory and practice of group psychotherapy*. New York, NY: Basic Books.
- Yang, J. (1995). *Noise reduction in a statistical approach to text categorization*. Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, USA.
- Yoo, W., Chih, M. Y., Kwon, M. W., Yang, J., Cho, E., McLaughlin, B., ... Gustafson, D. H. (2013). Predictors of the change in the expression of emotional support within an online breast cancer support group: A longitudinal study. *Patient Education and Counseling*, 90(1), 88–95. doi:10.1016/j.pec.2012.10.001
- Yoo, W., Namkoong, K., Choi, M., Shah, D. V., Aguilar, M., Tsang, T. J., ... Gustafson, D. H. (2014). Expression and reception of emotional support online: Mediators of social competence on health benefits for breast cancer patients. *Computers in Human Behavior*, 30, 13–22. doi:10.1016/j.chb.2013.07.024

Appendix 1

Best practices for studying computer-mediated communication

Recommendation

Data

Think big and small: The term “big data” implies that a research project must be performed on a grand scale, but some of the richest data is found in small social networks (micro-communities)

Triangulate datasets: Ideally, researchers should combine datasets that (1) capture who posted or read content messages, (2) reflect the content of these messages (content analysis), (3) measure attitudes or psychological responses (e.g., surveys), and (4) conceptualize the complex social relationships formed in online groups (social network analysis).

Capture information about who posted and read which message or utilize innovative approaches for tracking usage patterns and/or interaction data

Computer-aided Content Analysis

Invest in high-quality software or create a program that can account for the syntactical complexity in online discourse

Create a customized dictionary

To improve validity, include both inductive and deductive approaches and human and computer readings of the text

To check the external validity, test how the rules perform using a different data set

Perform reliability tests between human and computer coding

Social Network Analysis

Utilize social network analysis to identify the patterns and strength of social relations among group participants

Distinguish between open (mass/group) and targeted (interpersonal) communication networks.

Examine the relationship between a participants’ social network position and their psychological health outcomes
