Group Informatics: A Methodological Approach and Ontology for Sociotechnical Group Research

Sean P. Goggins, Christopher Mascaro and Giuseppe Valetto
Drexel University 3141 Chestnut Ave Philadelphia, PA 19104. E-mail: s@goggins.com; cmascaro@gmail.com; valetto@cs.drexel.edu

We present a methodological approach, called Group Informatics, for understanding the social connections that are created between members of technologically mediated groups. Our methodological approach supports focused thinking about how online groups differ from each other, and diverge from their face-to-face counterparts. Group Informatics is grounded in 5 years of empirical studies of technologically mediated groups in online learning, software engineering, online political discourse, crisis informatics, and other domains. We describe the Group Informatics model and the related, 2-phase methodological approach in detail. Phase one of the methodological approach centers on a set of guiding research questions aimed at directing the application of Group Informatics to new corpora of integrated electronic trace data and qualitative research data. Phase 2 of the methodological approach is a systematic set of steps for transforming electronic trace data into weighted social networks.

Introduction

The introduction of new technologies that enable peer-to-peer interaction changes how individuals communicate and form groups. Our analysis of dozens of online small groups over the past 5 years recognizes key differences between online groups and face-to-face groups. Online groups are more easily formed than groups in the physical world (Yuan, Gay, & Hembrooke, 2006), multitask more (Goggins, Laffey, & Tsai, 2007), and experience lower social presence than face-to-face groups. Further, individuals who are part of an online group describe the process of becoming a group as incommensurate with the experience of becoming a "real group" in the physical world (Goggins, Laffey, & Gallagher, 2011). Among other differences, we know that online groups represent a looser type of affiliation and that the construct of "online community" is a relatively aspirational notion, compared with conceptualizations of community in the physical world (Kling & Courtright, 2004). This article proposes an ontology for the study of technologically mediated groups and presents a methodological approach for the systematic examination of the new, multivalent types of group emergence and development that occur through information and communication technology (ICT).

Users, managers, and designers of systems to support small work groups, learning groups, emerging civic organizations, governments, and nongovernment service organizations are not consistently able to build engagement through ICTs. Organizational change is influenced by ICT uptake and use (Kling & Scacchi, 1982; Kling, 1979, 1980; Kiesler, Boh, Ren, & Weisband, 2005), but the shift in ICT's focus on work use to wide, diffuse use in daily life (Grudin, 2010; Sawyer, 2009) motivates a reconsideration of the vital role that small groups play in adoption of ICT and organizational change. The small group unit of analysis in ICT research facilitates study of the diffuse information behaviors people exhibit in daily life, leading to a renewed focus on interactions between individuals and their larger cultural, social and technical contexts.

McGrath (1984) was the first to point out the importance of group interaction processes and tasks to the functioning of small groups. The task circumplex he describes classifies all task work along two dimensions. The first, conflict and cooperation, focuses our attention on the nature of interactions between members. The second, conceptual and behavioral, focuses attention on the extent to which group behavior is observable. Ideally, conceptual group actions in traditional small group research are studied phenomenologically, but when observations are not possible, instruments, tests, and interviews must be developed to measure what people are thinking and observing, and otherwise find motivating.

Further consideration of the task circumplex, and especially the conceptual work of groups, leads Arrow, McGrath, and Berdahl (2000) to frame small groups as complex systems. In the complex systems view of small groups, the task maintains an important, central role, but different types

Received December 13, 2011; revised August 7, 2012; accepted August 8, 2012

 • Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.22802

of groups, including social groups, are accounted for by making sense of how groups and group members interact, often in ways that are not predictable. Understanding the complex systems that small groups embody remains important because of the central role they play in organizational change (Cakir, 2007; Goggins, Valetto, Mascaro, & Blincoe, 2012; Healy, White, Eshghi, Reeves, & Light, 2007; Howley, Mayfield, & Rosé, 2012), societal change (Fine & Harrington, 2004), and ICT adoption and use (Goggins, Laffey, & Gallagher, 2011; Mead, 1934, 1958; Stahl, 2006).

What we mean when we talk about groups is the subject of disagreement among scholars who examine groups in technologically mediated contexts. Schmidt and Bannon (1992) called for the abandonment of "group" as a construct for collaborative computing research because of disagreement about the definition. What do we mean? What kind of group? What size? What duration? What purpose? These questions make the complexities and nuances of technologically mediated group work stand out, and make the importance of wrestling with challenges of ICT-mediated group research important for contemporary information science.

The ontology and methodological approach of Group Informatics aims to bring greater specificity to discourse around small, ICT-mediated groups, which may exist for any of the traditional reasons groups exist. ICT-mediated small groups are identifiable through analysis of interactions and do not, in our research, naturally occur with more than a dozen individuals. Our prior work outlines how online groups are different in many respects from face-to-face groups and other emergent groups (Goggins & Mascaro, 2012; Goggins, Laffey, & Gallagher, 2011; Goggins, Mascaro, & Mascaro, 2012; Mascaro & Goggins, 2012). Returning to McGrath's dimensions, Group Informatics recognizes that the technical aspects of small groups that emerge through ICTs have some research advantages. First, online small groups interact through a technical system that captures the behavioral and conceptual dimensions of the task circumplex. Second, as a result of technological mediation, these groups are more observable than face-to-face groups; their behavior may be observed in retrospect (Goggins, Laffey, & Gallagher, 2011).

Small Groups as Sociotechnical Systems

Retrospective, comprehensive observation of emergent, ICT-mediated small groups is possible because common technologies like Facebook and Twitter, along with bespoke discussion forums, all produce logs. These logs, which we refer to more generally as "electronic trace data," contain varying degrees of detail with regards to user profile information, post activity, and read activity. Three main factors limit the integration and utility of many studies for integrative research. First, the focus on specifics of the technical system limit the possibilities for interpreting electronic trace data through the lens of a particular social science research construct (Goggins, Galyen, & Laffey, 2010; Golbeck, Grimes, & Rogers, 2010; Heverin & Zach, 2011; Mascaro & Goggins, 2011a; Thelwall, Buckley, & Paltoglou, 2011a, 2011b). Second, most of these studies are focused on smaller (individual) or larger (entire community or subcommunity) units of analysis, with relatively few focused on the small group. Third, the absence of a shared ontology for discussion of ICT-mediated small groups limits the effect of context-specific empirical studies in the information science discourse.

We address each of these limitations in order, first the limitations associated with interpreting technical log data with respect to social phenomena. The crux of this limitation is that taking electronic trace data at face value, without using or describing a methodological approach for systematically connecting electronic traces with social phenomena, often leads to incoherent, invalid, and unreliable conclusions about the relations between participants (Howison, Wiggins, & Crowston, 2012). The methodological approach of Group Informatics (2012) is aimed at solving this problem. This use of "Group Informatics" contrasts with Travica (2005), who uses the term "Groupomatics" interchangeably with "Group Informatics" in reference to a construct related to an information view of organizations (IVOs). Group Informatics, in our use, is focused on small groups and informed by social informatics, which studies "the design, uses and consequences of information technologies that takes into account their interaction with [sociotechnical] and cultural contexts" (Kling, 2007, 205).

To address the first limitation, the Group Informatics methodological approach views small groups as sociotechnical systems. This is similar to the social informatics frame of focusing on interactions between participants in large social contexts like organizations and society. Group Informatics hones in on the explicit interactions between people that are revealed in electronic trace data in specific sociotechnical contexts. The resulting contextualized interactions are applied to generate social network visualizations and statistics that are empirically grounded and representative of the underlying social phenomena. Part of what a discussion of Group Informatics contributes, then, is the framing of electronic trace data as raw evidence of the social, task, and group context of ICT-mediated behavior. An approach for connecting this raw evidence with specific research constructs, as we will describe here, enables future ICT developers to make ICT context more visible, and possibly more adaptive for users.

Second, the identifiable nature of small groups framed as sociotechnical systems enables the study of group phenomena in ICTs on a scale not possible for the study of face-to-face groups. Some studies have identified the importance of small groups at the core of a larger enterprise before. For example, within open source software projects, a small core of individuals is usually responsible for shepherding the work, and a number of small groups emerge to complete work in a modular organizational structure (Crowston, Wiggins, & Howison, 2010; Crowston & Howison, 2005). In the case of citizen science projects such as eBird.com, the natural interests of hundreds of specialized, local

DOI: 10.1002/asi

hobbyist groups (bird watchers) contribute small chunks of raw data, which scientists then aggregate to track the evolving migratory patterns of birds (Wiggins, 2011). Wikipedia's small groups include topically focused experts who construct information, following a highly regimented organizational hierarchy (Kittur, Chi, Pendelton, Suh, & Mytkowicz, 2007). Although these studies identify the important role of ICT-mediated small groups, in each case the principle unit of analysis is the larger community. Group Informatics, in contrast, begins with a premise that the small group unit of analysis is important for understanding ICT-mediated organizations.

The third gap is ontological. Approaches prior to Group Informatics face challenges in the development of a systematic and clear path for addressing differences in social and technical contexts because many of the details of interaction differ from study to study, including domain, technical system, and social practices. The Group Informatics ontology addresses this gap by providing a language to accompany its methodological approach. Working toward a shared ontology of small, ICT-mediated group research enables comparison across studies and a foundation for discussing broader insights in contemporary information science research.

Closing Ontological and Methodological Approach Gaps

Building on the framing of small groups as sociotechnical systems, there are two specific opportunities for bridging ontological and methodological gaps with Group Informatics. First, the Group Informatics methodological approach and ontology provides a framework for drawing comparisons between studies and synthesizing findings about group activity across technical platforms. Second, the role of group work for building engagement and performing productive work on a larger scale is foregrounded in Group Informatics, which is principally concerned with the emergence and development of groups through ICT.

Big Social Data

The availability of electronic trace data from ICT, including social media, where people interact and form groups is proving to be both an opportunity and a challenge for researchers. One side effect of this "Big Data" (King, 2011) problem for social science researchers is that research to understand group behavior in ICTs using electronic trace data is diffuse, and often fails to integrate electronic trace data with data gathering and analysis methods like ethnography, grounded theory, content analysis, and other qualitative social science methods.

If the trace data alone are a "big data problem," incorporation of other, less structured qualitative data constitutes a crisis of meaning, beyond scale and focused on connecting the pieces to construct an integrated, reflexive stance toward the relationship between data captured through technology and the underlying social phenomena. Some researchers argue that this challenge is intractable. For example, Cox et al. (2011) rationalize a focus away from qualitative data integration because, in their study of Flickr, they found integration of qualitative data with electronic trace data incongruent. Prior attempts have been made to systematically integrate trace data with other methods, but those efforts focus on a single, controlled research environment (Suthers, Dwyer, Medina, & Vatrapu, 2010).

Toward Scalable Methods

Group Informatics recognizes that to integrate qualitative analysis of ICT-mediated small groups with network analysis of electronic trace data from ICT, a road map (methodological approach) and common language (ontology) are both required. Research that integrates these two types of data is most often performed today using small corpora, such as a single hour of discourse. For example, Geiger and Ribes (2011) propose trace ethnography as one possible methodological approach, but like, Stahl's extensive ethnomethodologically informed analysis of electronic trace data (Stahl, 2002, 2006, 2009a, 2009b, 2009c), the approach does not scale to large conversations or longitudinal studies. Previous methodological approaches demand significant investments of time per unit of data (Sacks, 1972).

Alternately, analysis of electronic trace data is often approached quantitatively using data mining and text mining to identify clusters of interaction or keywords from large corpora (Backstrom, Kumar, Marlow, Novak, & Tomkins, 2008; Cronin, 2011; Diani, 2003; Falkowski, Bartelheimer, & Spiliopoulou, 2006; Kittur et al., 2007; Kittur, Lee, & Kraut, 2009; Kraut & Streeter, 1995; Teasley, Covi, Krishnan, & Olson, 2002). Sentiment analysis is another method for data analysis of large corpora that strive to derive meaning using computation (Jansen, Zhang, Sobel, & Chowdury, 2009; Naaman, Becker, & Gravano, 2011; Thelwall et al., 2011a, 2011b). Without grounding in the underlying social phenomena, however, these studies are prone to many of the issues Howison et al. (2012) describe. In studies performing fine-grained social science analysis on small corpora, we see limitations in scope. Studies of large corpora are often limited in their depth of social scientific analysis. At the core of the Group Informatics approach is the idea of integrating qualitative, social science analysis methods with quantitative methods to build a deeper understanding of technologically mediated small groups.

Group Informatics which focuses on analysis of small group phenomena emerges from work on 16 different corpora from ICT-mediated small groups, which we have analyzed individually and across groups and platforms (e.g., Blincoe, Valetto, & Goggins, 2012; Goggins & Mascaro, 2012; Goggins & Valetto, 2010; Goggins, Galyen, & Laffey, 2010; Goggins, Valetto, Mascaro, & Blincoe, 2012; Goggins, 2006; Goggins, Laffey, & Amelung, 2011; Goggins, Laffey, Amelung, & Gallagher, 2010; Goggins, Schmidt, Moore, & Guajardo, 2011; Laffey et al., 2008;

DOI: 10.1002/asi

Mascaro & Goggins, 2011a; Mascaro & Goggins, 2011b; Mascaro & Goggins, 2012; Mascaro, Novak, & Goggins, 2012; Tsai et al., 2008).

Group Informatics provides ontology and an integrated methodological approach for analysis of online groups, where electronic trace type data are one of several types of data. This enables the researcher to transform technical log records of interaction into sociotechnical interaction records across a range of sociotechnical contexts. Simply examining electronic trace data without grounding the analysis in constructs from social science leads to issues of theoretical coherence, validity, and reliability (Howison et al., 2012). Yet even when they do, Howison et al. (2012) points out that the social theories that underlie social network analysis frequently rely on assumptions of ties that cannot be inferred solely from records of interaction captured in electronic trace data. Thus, simply applying social network theories without understanding the nature of behavior and the nature of the platform (and how it varies from group to group) leads to errors in analysis and, thus, interpretation. Too often, research of online phenomena conflates the technical artifacts with the social experience of participants or assumes that the technical artifacts are the units of analysis with no insight into how a group is using these artifacts to participate.

Application of the Group Informatics methodological approach systematically builds an understanding of technologically mediated groups across a range of contexts. We use this approach to analyze and integrate multiple data types, including electronic trace data, interview, survey, and ethnographic data using methods from ethnography and grounded theory. We further expand the richness of our understanding with adductive content analysis methods described by Krippendorff (2004) to identify theoretically grounded constructs like political discourse, learning, and coordination. The triangulated analyses are then used to drive focused theoretically and empirically grounded analysis of electronic trace data.

Constructs, Empirical Refinement, and Theory

In the three sections that follow, we describe the constructs, phases, and theories that Group Informatics builds on. We define two new constructs for the purpose of being explicit about the nature of the groups we study. Our description of the four phases of research that led us to develop the Group Informatics model and methodological approach help the reader to understand our motivations for preparing this ambitious article. Finally, we situate group informatics as a methodological approach in the service of activity theory and embodied interaction perspectives on sociotechnical research.

Constructs For Group Informatics

To aid the reader's conceptualization of how our methodological approach is differentiated from existing approaches, we introduce two new ontological constructs that distinguish technologically mediated social phenomena from more widely studied, less dynamic social phenomena. First, the groups that form in these asynchronous environments are referred to as *small, naturally asynchronous groups* (SNAGs) to distinguish them from previous conceptualizations of physical groups, such as distributed teams, virtual organizations, distance work, and computer-supported cooperative work, broadly defined. The term SNAG reflects the unmet challenge of integrating qualitative and quantitative modeling to understand how interaction, leadership, and social structure are represented in electronic trace data (Ahuja & Carley, 1999; Bansler & Havn, 2006; Blay-Fornario, Pinna-Dery, Schmidt, & Zarate, 2002; Bos, Shami, Olson, Cheshin, & Nan, 2004; Chudoba, Wynn, Lu, & Watson-Manheim, 2005; Convertino, Moran, & Smith, 2007; Edwards, 2005; Ehrlich & Cash, 1999; Fuller, Hardin, & Davison, 2007; Gutwin & Greenberg, 2004; Gutwin, Penner, & Schneider, 2004; Harrison & Tatar, 2007; Hinds & McGrath, 2006; Leinonen, Jarvela, & Hakkinen, 2005; Liu, Laffey, & Cox, 2008; Mascaro & Goggins, 2011a; Nardi & Harris, 2006; Nardi, Whittaker, & Schwarz, 2002; Neale et al., 2004; Neale, Carroll, & Rosson, 2004; Ocker & Fjermestad, 2008; Olson & Olson, 2000; Olson et al., 1998; Olson, Herbsleb, & Rueter, 1994; Olson, Malone, & Smith, 2001; Olson, Olson, & Venolia, 2009; Olson, Olson, Storrosten, & Carter, 1992; Powell, Piccoli, & Ives, 2004; Schmidt & Wagner, 2004; Sonnenwald, Lassi, Olson, Ponti, & Axelsson, 2009; Star & Strauss, 1999; Teasley et al., 2002; Turner et al., 2006; Twidale & Nichols, 1998; Whittaker, 1996; Carroll, Neale, Isenhour, Rosson, & McCrickard, 2003; Crabtree, O'Neill, Tolmie, Colmbino, & Grasso, 2006; Crowston & Howison, 2005; Fuller, Hardin, & Scott, 2007; Grudin, 1994; Lampe, Ellison, & Steinfield, 2008; Roberts, Lowry, & Sweeney, 2006; Saunders & Ahuja, 2006).

Second, we refer to the online contexts in which SNAGs interact as *sociotechnical interaction places* (STIPs). A STIP is any system in which people interact as groups, for a specific purpose, and mediate consistent and meaningful aspects of their activity through technology that generates electronic trace data. Many STIPs create electronic trace data without reference to how these logs might be applied to represent group leadership, emergence, or development, but for keeping track of basic notions of interactivity. Such logs can be conceptualized as log files, with records of interaction that include at least an actor, an artifact, and a timestamp. Our Group Informatics methodological approach systematically allows for the analysis of raw log data from a STIP. The Group Informatics model is a component of the overall Group Informatics methodological approach.

Foundational Empirical Work Leading to Group Informatics

Phase one of our work originates with analysis of how users in an online course used daily digests of course

DOI: 10.1002/asi

activity, something we refer to as a "Context Aware Notification System" (CANS; Amelung, 2005). Our first paper (Goggins, 2006) reports on a qualitative study of the relationship between information foraging and group size in completely online graduate-level courses. This analysis inspired questions at the small group unit of analysis, which we further developed by considering how Stahl's work on Group Cognition (Stahl, 2006) could be applied to asynchronous small groups, which we now conceptualize as SNAGs. In the second phase, we performed an in-depth qualitative research study of a single small group in a completely online course (Goggins, Laffey, & Tsai, 2007) and measured a construct called social ability (Laffey, Lin, & Lin, 2006), which measures a group's overall capacity to behave socially online (Goggins, Laffey, & Galyen, 2009). This informs the positive design (Carroll, Rosson, Farooq, & Xiao, 2009) work of phase three, which treats the ICT-mediated group as a unique and, at times, advantaged new type of social structure. Like Miksa, Burnett, Bonnici, and Kim (2007), we recognize the need for more systematic mechanisms for coding and quantifying online courses across a range of institutional and pedagogical boundaries.

The third phase of development focused on the implementation of a data warehouse and analytical system, integrating electronic trace data with qualitative data. We then apply our system to a variety of contexts and technologies, including online political discourse on Facebook (Mascaro & Goggins, 2011a, 2011b; Mascaro et al., 2012), open source and industrial software engineering practice (Blincoe et al., 2012), and disaster relief on government sponsored discussion forums (Goggins, Valetto, Mascaro, & Blincoe, 2012). The fourth phase is under way.[1] We are expanding our data collection and management capabilities (Black, Mascaro, Gallagher, & Goggins, 2012) and our analytical frameworks for understanding groups across contexts more systematically. The resulting methods and tools are being shared with other researchers to enable comparisons across sociotechnical contexts. This work embodies the Group Informatics methodological approach and it is generating new measures, including proximity (Blincoe et al., 2012) and lurking knowledge construction (Goggins, Galyen, & Laffey, 2010) that are based on a combination of qualitative data analysis and established social network analysis (SNA) measures (Goggins, Laffey, & Gallagher, 2011). Group Informatics is framed for information scientists as a methodological approach for inquiry into SNAGs and STIPs.

Group Informatics works toward models and representations of technologically mediated interaction that reflect the new, emergent forms of social organization embodied by SNAGs. We use models to represent how SNAGs are *experienced* through technology. We illustrate that identifying SNAGs from electronic trace data requires the development of new conceptual models that can then be implemented through a systematic, methodological approach that synthesizes existing qualitative and quantitative methods with modeling.

Activity Theory, Embodied Interaction, and Group Informatics

Group informatics is a methodological approach with a foundation in prominent theories of social computing. Specifically, Group Informatics is informed by the constructs of *artifact*, from activity theory (Kaptelinin & Nardi, 2006; Nardi, 1996), and *interaction*, from embodied interaction (Dourish, 2001) and group cognition (Stahl, 2006).

The principal unit of analysis in the Group Informatics model is the interaction, as called for by Stahl (2009a). Interactions in sociotechnical systems occur between people, or between people and artifacts, depending on the context. Interactions may occur around pieces of code, discussion boards, a medical chart, a map, or innumerable other chunks of information viewable on public and private spaces on the Internet (Blincoe et al., 2012; Goggins, Galyen, & Laffey, 2010; Goggins et al., 2011; Goggins, Laffey, & Gallagher, 2011; Goggins, Valetto, Mascaro, & Blincoe, 2012; Mascaro & Goggins, 2011a). This focus on interactions extends from existing social network theory, which emphasizes that social groupings are conceptually aligned with a cluster of people who interact one on one with a collection of others.

Interactions, which are representations of relations between individuals described in social theory, are captured at a finer grain in Group Informatics than in traditional social network analysis. Each electronic trace of interaction is a concrete record of some contact between individuals. These interactions occur around activities of some kind. Group Informatics is differentiated from prominent theories of human computer interaction and human information behavior, like activity theory (AT; Kaptelinin & Nardi, 2006) or information horizons (Sonnenwald, 1999, 2005), by its focus on this finer unit of analysis—the interaction. For example, activity theory describes the activity as the unit of analysis and incorporates the subject (a person) and object (an artifact) as the principle nouns around which activity occurs. Group Informatics recognizes that in the case of technologically mediated groups, artifacts are ephemeral sites for interaction and work is often split across multiple artifacts. Therefore, analysis centered on interactions represents a more coherent focus in a technologically mediated social context.

Although employing a different unit of analysis that is appropriate to the context, Group Informatics builds on AT by recognizing its implications for technology design, as articulated by Nardi (1996). Artifacts remain an important construct in our model, which incorporates a focus on the interaction of groups around artifacts. In this incorporation, we maintain a commitment to the user's point of view and are working to extend activity theory to include the construct of an interaction.

[1]http://ssrn.com/abstract=2120428

DOI: 10.1002/asi

The Group Informatics methodological approach is informed by this relationship between electronic trace data and the dynamic context. Interaction is broadly theorized about in Dourish's discussion of embodied interaction (Dourish, 2001) and further elaboration of context (Dourish, 2004). We view context as a construct within which users experience technology. Fundamental to Dourish's (2001, 2004) notion of context is the idea that context is dynamically constructed through human behavior in a sociotechnical environment; context is not a static, technical place. SNAG context is embodied mostly through technology, as this is usually the only mechanism of mediation that participants have. Even in cases where there is a relationship between identity in the physical world and the SNAG, the context and experiences of SNAG members are constructed differently through technology. For example, in our study of online recreational discourse, we identify how the online identity that emerges through participation in discussion boards leads to behavior where people refer to each other in person by the monikers chosen in the forum (Novak & Mascaro, 2012). Events in systems like this traverse both the physical and virtual, but the virtual interactions help to further define the context of the physical.

The remainder of this article first situates the methodological approach in recent information science research and decades-old research analyzing social networks. Second, it defines and describes the model in detail. Third, we describe our methodological approach and provide a sample set of questions to help researchers frame analysis of electronic trace data in the social experiences of group members. Fourth, we explain context adaptivity and identify how other researchers can utilize Group Informatics in their own specific domains. We conclude with a discussion of the implications of Group Informatics for the study of electronic trace data.

Group Informatics: Foundations in Social Network Analysis

Network analysis of technologically mediated groups leverages knowledge from decades of social science research focused on understanding how social interactions between individuals evolve into social networks in the physical world, and how these networks influence individual and group behavior (Freeman, 2003, 2004; Straus, 1993). From this research, network researchers have built a set of validated measures to help identify important actors in these social networks. Well-known statistical measures of individual influence and network position include (a) betweenness, which identifies bridging individuals who connect two clusters in a network, (b) closeness, which describes the ability of a person to reach information within the network through a set of ties, and (c) degree centrality, which is a measure of overall connectivity to other actors in the network. These measures have different meanings when viewed through different theoretical lenses and care must be taken to interpret their meaning in each application (Freeman, 1979; Friedkin, 1991).

Researchers who study groups using electronic trace data frequently conflate notions of social connection in SNAGs with constructs understood in the physical world. There are, however, important differences to attend to when examining SNAGs. For example, online connection is not experienced in the same way or influential through the same mechanisms as face-to-face connection. Well-known social network constructs, like leadership (Balkundi & Kilduff, 2006; Fletcher & Kaufer, 2003; Wasserman, 1994), brokerage (Burt, 2005; Diani, 2003; Fleming & Waguespack, 2007), and information diffusion (Valente, 1996) are manifested differently in technologically mediated environments because of how they are experienced, and how they are observed are different.

Our study of online courses shows that high betweenness centrality corresponds with both influence, as originally described, and lurking behavior (Goggins, Galyen, & Laffey, 2010; Goggins, Laffey, & Gallagher, 2011). Betweenness centrality in traditional SNA is a measure that has been used to identify those people who are in a middle position. Through the analysis of electronic trace data, we have demonstrated that the "middle position" could indicate both influence and surveillance depending on if the measure uses post data or read data. To apply the statistic of betweenness in this new context, we need to consider the theories salient to what we observe, and inform the development of new theories based on what we learn.

Degree of connectedness and type of connectedness between people vary by context in both the physical world and the SNAG. Weighted and directed versions of SNA measures embody this understanding (Scott, 2000). Weighted network analysis is especially useful in the case of electronic trace data because strong ties, weak ties, and ephemeral ties are implied by differences in strength of connection between individuals (Granovetter, 1973). Weighting strategies should be grounded in theory or empirical data from the phenomenon under study. Such grounding is fundamental in classic, physical world studies of social networks (Davis, Gardner, Gardner, & Silver, 1965; Roethlisberger, Dickson, & Wright, 1939), but usually overlooked in studies focused on network analysis of electronic trace data. (See Howison et al., 2012, for an inventory of studies on e-mail systems, online discussion boards, and other electronic media where grounding in the real-world phenomena is not addressed).

One positive difference between network analysis using traditional data-gathering methods from sociology, and network analysis derived from electronic trace data, is that electronic trace data reflect a more complete log of interactions (Lazer et al., 2009). This greater completeness helps to overcome sampling issues that occur in network analysis that depends on periodic observation or self-reporting of perceived connections between actors in the physical world (Bernard, Killworth, Kronenfeld, & Sailer, 1984; Freeman, Romney, & Freeman, 1987; Parigi & Bearman, 2005). Analysis of physical social networks through traditional methods is also vulnerable to boundary specification issues (Laumann & L.P., 1997; Laumann, Marsden, & Prensky, 1989), while analysis of technologically mediated groups establishes a clear boundary of participation in the system with some limitations described below.

One caution when analyzing social networks derived from electronic trace data is that not all interactions are necessarily logged in a system the researcher or analyst can see. Awareness of the potential (likely) existence of nonlogged interactions among or between users and other systems is therefore crucial; if you are looking only at log data, there is a good possibility you have not accounted for the full story. For example, software engineers using a bug tracking system, MyLyn, and a source code repository may also e-mail each other. In other instances where face-to-face relationships also exist, the electronic interactions may extend and augment interactions that have occurred in the physical space. As Laumann (2006) notes, it is up to the researcher to determine the representativeness of the sample used for network analysis. In our studies of software engineering, online courses, disaster relief, and other venues, we, like Laumann (2006), find the traces we examine representative of connections overall. We determine representativeness by triangulating electronic trace data and network analysis with other data collection and analysis methods. The Group Informatics model helps to systematically assess representativeness of trace data and provide empirical justification and a clear rationale for connection weighting decisions in different sociotechnical environments.

Defining the Components of the Group Informatics Model

The Group Informatics model comprises four core components: (a) artifacts, (b) interactions, (c) context, and (d) people. Each component has dimensions and relationships with the other components. The four components contribute to the resulting contextualized interaction, which is a weighted network of electronic trace data interactions that reflect social phenomena. This enables the application of our methodological approach to the study of SNAGs from a wide range of STIPs. We provide a visual representation of the model and the relationships between components in Figure 1. In the two sections that follow, we describe the Group Informatics method, which we use to understand the phenomena experienced by SNAGs in a particular STIP through qualitative and network analysis of artifacts, interactions, context, and people.

In Table 1, we provide an overview of the components of the model. The Appendix presents a detailed description of the model components, which are part of the Group Informatics methodological approach.

A Short Example of Results From the Group Informatics Model

Group Informatics is a model and methodological approach that fills ontological gaps in existing discussions of technologically mediated groups by making the components of the model explicit. Applying any methodological approach is a more dynamic activity. Figure 2 describes the application of Group Informatics in one case, an online course, illustrating the reflexive nature of interactions between different types of qualitative and electronic trace data.

To illustrate the power of the Group Informatics methodological approach, in Figure 3 we provide two illustrations of the social networks that result from analysis of electronic trace data from the same group of people during the same time period in an online course. The figure on the left shows that a core group of team members are identifiable as the leaders of their respective groups when the group informatics methodological approach and weighting are applied to the trace data. Each group member on the left is represented with a name that includes their group number (g1-g6) and an identifier. One group is represented with actual names. Each group has a distinct shape and shading combination, such that it is visually clear from the figure that one member of each group is in the core for the overall network. The figure on the right shows the same data without this type of Group Informatics methodological approach applied. These results are replicated across 11 courses and a dozen other sociotechnical contexts.

Operationalizing the Group Informatics Methodological Approach: Artifact, Interaction, Group, and Person

To apply the Group Informatics methodological approach to answer research questions, we provide guidance for researchers, to address two major challenges. First, this approach facilitates study design decisions related to data gathering and analysis, leading to interpretation of electronic trace data that is grounded in social and information science research questions. Second, we provide an explicit strategy and examples for weighting connections between people or between people and artifacts, which are the two types of connection contained in electronic trace data.

Table 2 provides a guide for applying the Group Informatics methodological approach to electronic trace data, in the form of questions organized around the four main components of the model and the overall component of context adaptivity. To fully understand the electronic trace data being examined, it is important to first develop a high level understanding of the SNAGs operating within the STIP being studied. One way may be to take part in the sociotechnical environment as a participant observer, to get a sense of what is happening. This participant observation through technological familiarization helps researchers develop a greater understanding of the technological affordances. The

DOI: 10.1002/asi

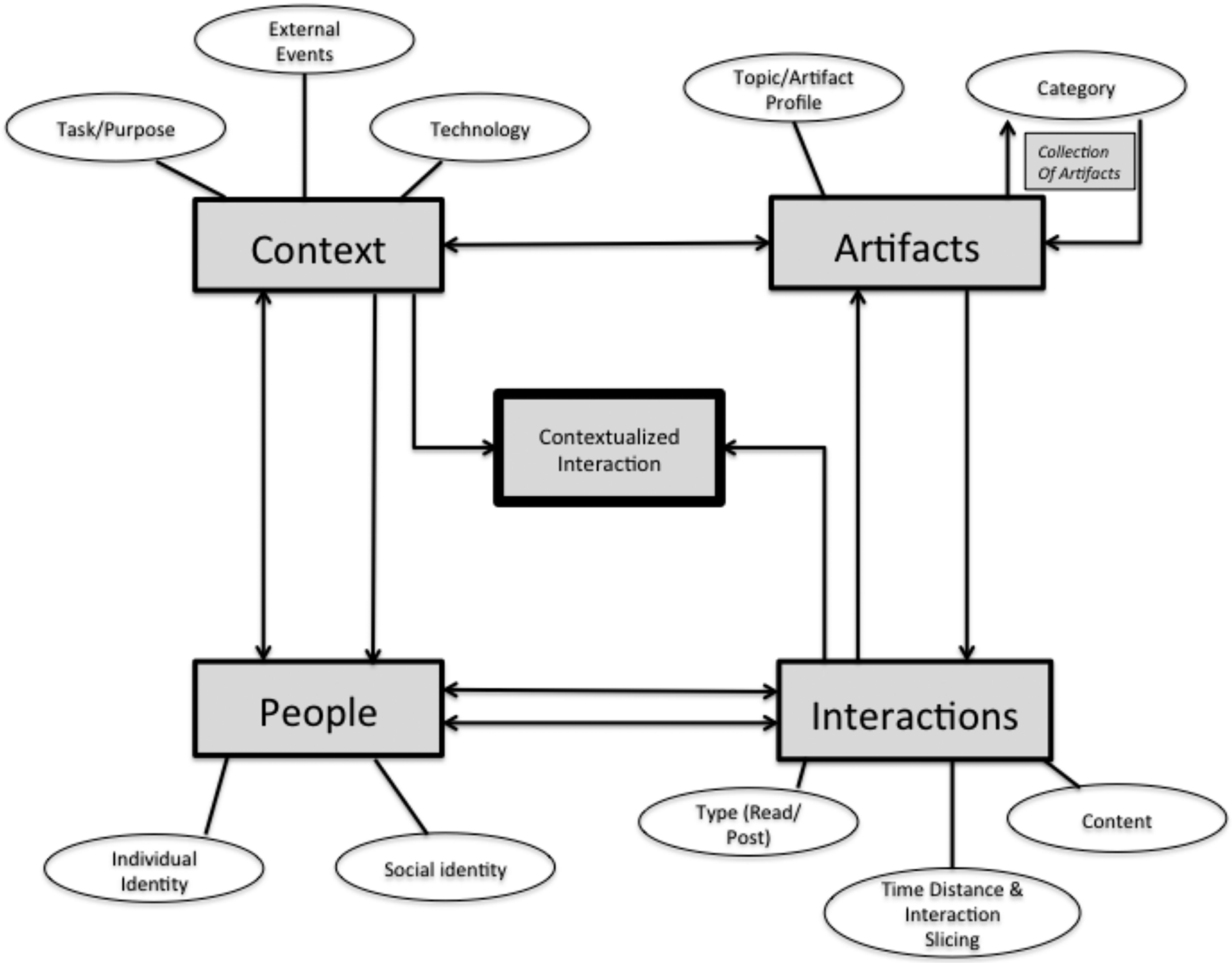

FIG. 1. Model overview of group informatics.

following questions can be best answered during the participation in the environment, or through a combination of qualitative and quantitative methods. The questions are intended to help researchers develop an understanding of each STIP that is grounded in the social, task, or other phenomena of interest; these are not prescriptive research questions, but questions to aid researchers in their operationalization of the Group Informatics methodological approach.

Operationalizing the Group Informatics Model: Context Adaptivity

The remaining component of the model is context adaptivity, shown in the middle of Figure 1. First, context adaptivity is how we carry forward the fundamental premise of early and classic papers on network analysis, which argue strongly that knowledge of the context of study needs to be deep, so that the analyst has triangulated their understanding of what is being measured and can make study specific judgments about the validity of analysis. Second, context adaptivity recognizes that existing measures of network analysis represent conceptual social structures that help researchers and practitioners understand SNAGs, but that how the electronic trace data are created in each STIP is quite different. Even though traditional social network measures were created to understand relationships, they can be applied to electronic trace data if the STIP and the output are understood *in context*.

Third, the context adaptivity component of the model enables the construction of static views that represent the dynamic nature of technologically mediated online groups. We refer to this in our model (Figure 1) as a *contextualized interaction*. This is the locus in the Group Informatics methodological approach, where qualitative research that takes place around the four main components of the model–context, artifact, person, and interaction–is made operational. User profile information, interaction types, interaction frequency, time, and the role of artifacts are examples of

DOI: 10.1002/asi

TABLE 1. Ontology of model components used in the Group Informatics methodological approach.

Component	Dimension	Description (see Appendix for additional description)
Interaction		The primary unit of analysis in Group Informatics. Interactions occur between two people, or between people and artifacts. Groups are discerned from analysis of individuals with similar others in their network, as described in social network theory.
	Time distance	Except in originating interactions, like the first post of a discussion board, each interaction occurs at some time distance from a previous interaction. Interpretation of the time distance statistic for interactions within a STIP is a specific example of how qualitative data is used to inform analysis of log data.
	Type (mode)	There are two modes: Active interaction, like posting or editing and passive interaction, like reading or referencing. Not all STIPs provide both, though many of the ones we have analyzed do. Each of these interaction modes will have a different meaning in each STIP.
	Content	This is what is contained in an interaction record. It could be the text of a post, or dimensions of a piece of code that have been changed. Content is referenced, but not created in passive mode interactions.
Context		The environmental factors influencing SNAG formation, leadership, and development contribute to the dynamic construction of context. The context component of the model exists at the intersection of interactions and artifacts; analysis and visualization of which can be used to make context more visible through applications of the model.
	Technology	The technical aspects of the STIP, and how users adapt these aspects, construct the technology dimension of the context component.
	External events	In some STIPs, external events are central. Disaster relief is a good example. In other cases, they are a nonfactor. Open source software engineering is an example. The key is to account for external events in any specific implementation of the model.
	Task or purpose	Individual motivations, group tasks, and the purpose of a STIP's existence constitute this component.
Artifact		An object around which groups and individuals collaborate. The artifact component of the model can be a piece of source code, a discussion topic, a wiki page, or any other technologically mediated evidence of work.
	Topic and artifact profile	Person-to-person interactions have topics, which is the stated focus of the communication. In discussion boards, this is the subject line. Person-to-artifact interactions have artifact profiles. Artifact profiles reflect either the role of the artifact (e.g., a source file in a module), or a type of work the artifact serves to coordinate (e.g., a boundary negotiating artifact).
	Category and collection of artifacts	Categories and collections are qualitatively or quantitatively grouped sets of artifacts. These can include "all source code updated in a software release," or "all GIS information requests in a disaster relief scenario."
People		The individuals participating in SNAGs and STIPs
	Individual identity	This dimension examines the ways that people express their individual identities within a particular STIP and its dynamic context.
	Social identity	Social identity focuses on the small group unit of analysis in Group Informatics. It is the ways and extent to which individuals identify themselves as members of a particular SNAG.

Note. STIP = sociotechnical interaction places; SNAG = small, naturally asynchronous groups.

contextual attributes that can be accreted from electronic trace data from the sociotechnical environment, attached to those interactions, and then contribute to the construction of a representation of group and task context. The questions in Table 2 aid the researcher in this operationalization.

The transformation of logs, based on analysis of qualitative components, leads to the contextualized interaction record, which is the input to the three steps highlighted in the context adaptivity component of the Group Informatics model (Figure 4): weighting, aggregation, and analysis of the resulting social network.

Contextual Understanding and Domain Specific Theory

In this section, we provide a set of representative transformation techniques that will aid researchers as they address questions about domain specific theory and context. The weighting and slicing techniques are not prescriptive, context, and domain independent solutions. Rather they are presented to help researchers to address how to weight and transform contextualized interactions in their own domains.

Conceptualizing Time Distance and Contextual Factors to Weight Interactions

Weighted network analysis includes an initiating node, a target node, and weight associated with each interaction. Interaction weights are then aggregated for each set of directed pairs, for example, time, topic or some other frame defined by the research questions. Connection weight depends on a number of factors when the connections are derived from electronic trace data, as we have argued from the start. The contextualized interaction incorporates each aspect of the model. The result is a weighted interaction between people, or between people and artifacts in a specific sociotechnical context. This is where the Group Informatics methodological approach is materially different from prior approaches to network analysis of electronic trace data.

DOI: 10.1002/asi

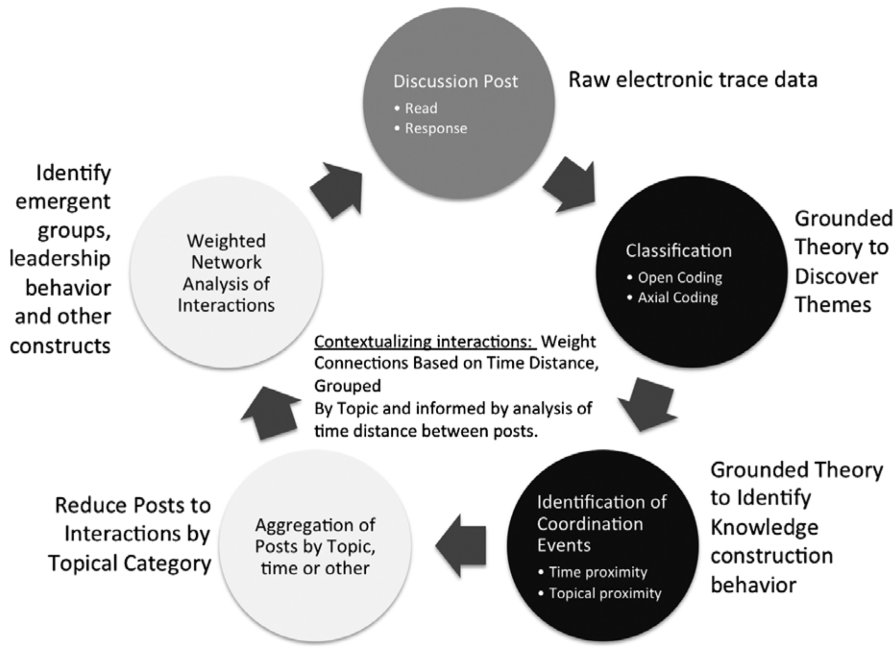

FIG. 2. Dynamic view of one application of Group Informatics methodological approach.

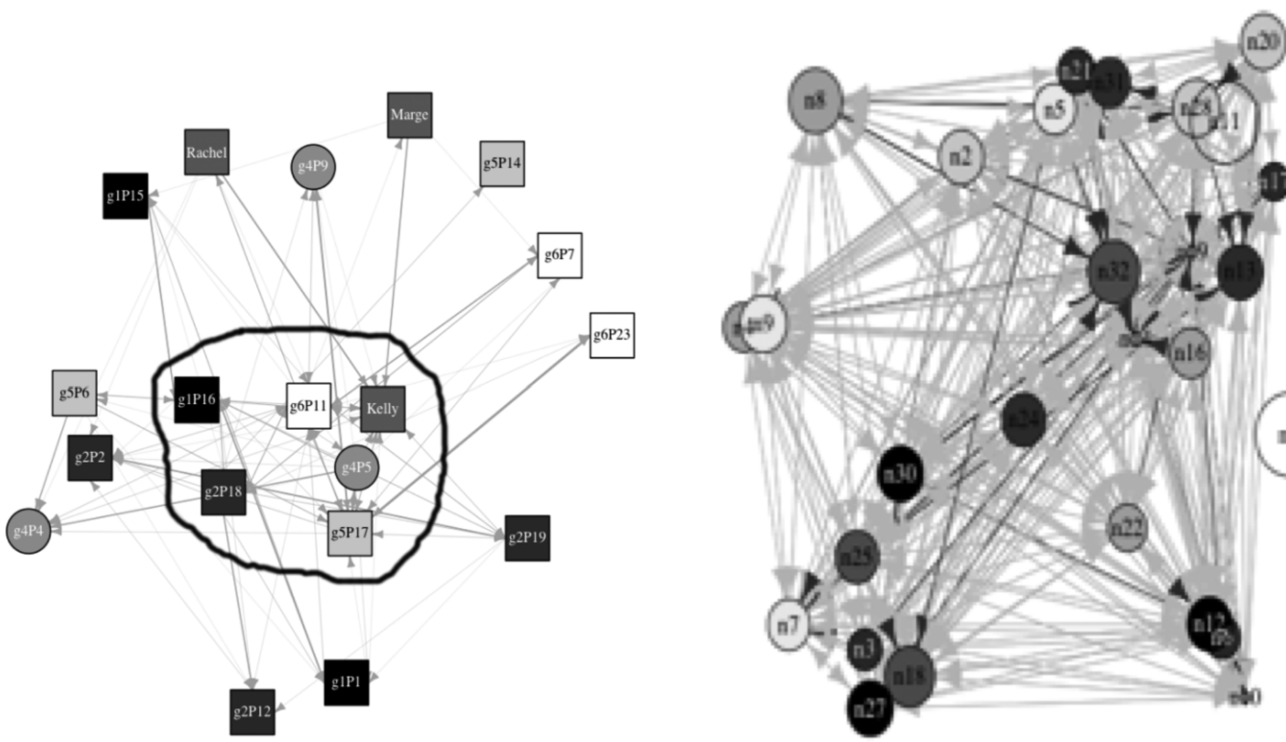

FIG. 3. Example social network after weighting calculations applied (left) and before weighting calculations applied (right).

An example of operationalizing the group informatics methodological approach from our work. An example from our empirical work will help to illustrate how interactions are contextualized in our methodological approach and how this contextualization can be represented through weighting. In the case of our analysis of interactions in online courses, we defined different types of interactions, including discussion board reads, discussion board posts, artifact edits, wiki edits, and archived chat interactions. We followed the course each day, taking field notes, interviewed 14 of 25 participants on three different occasions each, and analyzed discussion board content to determine the level of knowledge

DOI: 10.1002/asi

TABLE 2. Guiding research questions for Group Informatics cases.

Artifacts: Researchers should attempt to understand the range of artifacts within a system

- How can interactions be (re-)constructed through each category of artifact in the STIP? For example, are there categories that represent person-to-person interactions and other artifact categories that represent person-to-artifact interactions?
- Do the artifacts represent a collaborative work product for participants or do they serve a stigmergic purpose (Christensen, 2007, 2008)? Are the artifacts a type of discourse?
- Can the artifacts be created or eliminated by everyone or a small subset of administrators?

People: Researchers should attempt to develop an understanding of the demographics, educational levels and individual traits of users in a STIP.

- How many participants are there in the examined STIP?
- How many are active (the definition of *active* is determined by interactions and can vary depending on the context)?
- Do individuals have to validate their identity before participation in the STIP?
- Are there user profiles associated with the STIP to allow participants to develop an understanding of other participants?

Interactions: It is imperative to understand the ICT's affordances and how interactions differ in each technological system

- How frequent are interactions in the STIP? Between individuals? Between groups?
- What are the different modes of participation in the STIP?
- Which modes are captured in electronic trace data?
- Do individuals have to explicitly identify with a group or are the interactions and group formation more implicit and constructed around specific artifacts or discourse?
- Do individuals receive notification of interactions outside of the STIP such as through email or alerts generated by the technology?
- What are the statistical distributions of time distance between interactions and is there a commonly understood acceptable response time? For example, is the average response within a day? Do people interact daily on average?
- Is it common to have days, weeks, or months of downtime between interactions?

Context: The context must be identified as being constructed from events and interactions occurring within and around a STIP.

- What is the stated purpose of the STIP and do all individuals understand what the purpose is? Does qualitative analysis of interactions reveal divergent understanding of the STIP's purpose?
- How often does the technology change and what effect does that have on the STIP? (This is important in some newer technologies that frequently change how individuals interact.)
- What technological changes are reflected in the electronic trace data?
- What specific technological affordances exist to facilitate interaction?

Context adaptivity (How understanding the components can allow them to be operationalized within the model):

- Do the number of interactions between individuals represent a qualitatively significant indication of the context being constructed?
- Does the amount of time between interactions represent a qualitatively significant indication of the context being constructed?
- What are the mean and standard deviation of time between interactions? To what extent does length of time between interactions correspond with qualitative analysis of the interactions?

Note. STIP = sociotechnical interaction places; ICT = information and communication technology.

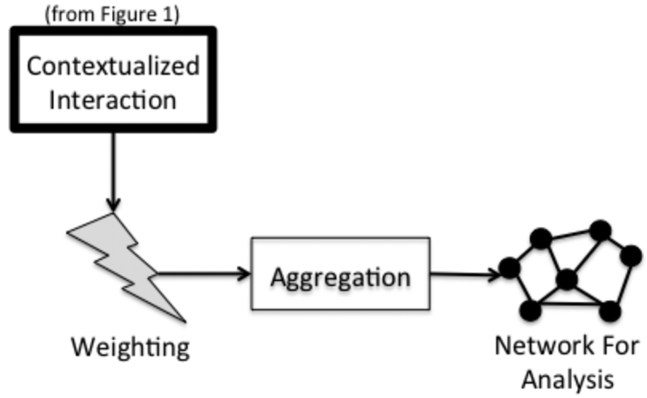

FIG. 4. Model for context adaptivity derived from electronic trace data of interactions.

construction and expression of social identity in each post (Goggins, Laffey, & Galyen, 2009; Goggins, Schmidt, Moore, & Guajardo, 2011). The qualitative data analysis revealed that discussion board interactions–both reads and posts–reflected the strongest social connection type, and only archived chat also represented a social connection. Therefore, we counted these two types of interactions, and not the other three as social interaction.

In the case of discussion board interactions, we observed three factors associated with time distance between both read and post interactions that influenced our weighting strategy. First, interactions within 30 minutes were more likely to include acts that built group identity, so they were weighted more heavily. Second, no posts that occurred greater than 4 days after another post were evaluated as including knowledge construction behavior. Our coded field notes and content analysis were triangulated to show that posts at a time distance greater than 4 days were often "catch up" posts made by students. Finally, read data did not time decay as quickly. People who read older posts but did not respond reported in our interviews that these activities were usually related to information retrieval; they remembered a classmate had a relevant idea, so they went back to look for it.

We can see from this one example that the process of contextualizing interactions involves many of the deep analysis strategies common in qualitative research, combined with a quantitative expression of the resulting social network. The Group Informatics methodological approach is in many ways embodied with this short description. Weighting connections enables context adaptivity and the static visualization of dynamic phenomena; it leverages contextualized interactions. Here, we describe the two types weighting equations we have already used–one for person-person interactions and one for person-artifact interactions.

Person-person interactions. Person-to-person interactions are weighted to account for the raw time distance (RTD) between the first person's activity and the second person's activity in the STIP, including potential "time distance

DOI: 10.1002/asi

TABLE 3. Example calculations (each row is a single interaction).

Raw time distance	Raw time weight
30	93
60	66
5000	0

cliffing" (CF), which is the idea that in some STIPs, interactions after a certain amount of time has passed are substantially less significant. Both the artifact type and the interaction type are taken into consideration, on a per-interaction basis, by another weighting variable, namely, the "context variable" (CV).

Raw time weight (RTW): This factor creates a context-specific distance for each interaction, based on RTD, expressed in minutes, between two interactions and in some cases a "time distance CF." *For each of the equations, we exclude interactions that occur beyond this cliffing factor.* In our studies of online learning, we have found cliffing factors of 3 to 4 days; in the study of disaster relief, cliffing occurs after 12 hours, and in online political discourse in the Facebook Group we studied, cliffing occurred after 72 hours. All of these values for CF were determined using the qualitative research methods described in the Group Informatics model's other components, and are specific to the time period and group studied.

Baseline distance (BD): Baseline distance is determined from an analysis of the distribution of raw distance data, combined with analysis of qualitative data. The goal for this factor in the equation is to set the value of "1" for the ratio between BD and RTD found in the equation below, so that exceptionally close ties are especially emphasized, and ties close to what is typical are weighted more evenly.

Equation 1—Distance

$$\begin{cases} RTW = \sqrt{\dfrac{BD}{RTD}} \times CV \;\; if \;\; RTD \le CF \\ RTW = 0 \; if \;\; RTD > CF \end{cases}$$

For example, if, in a particular context, we determined that an hour (60 min) was an appropriate value for BD, and that interactions older than 3 days (4,320 minutes) were of little utility, the specific equation for RTW in that context would be as shown in Equation 2. Where the RTD exceeds the CF, interactions are not represented in the analysis.

Equation 2—Distance Example

$$RTW = \sqrt{\frac{60}{RTD} \times 4320}$$

Table 3 shows how different values for RTD would be treated as measures of RTW in this case. The use of the square root serves a common mathematical "smoothing" function.

The RTW variable accounts for a good deal of the context-specific variability because the closeness of interactions, relatively, within a context is a measure of the significance of that interaction in some networks; hence a major contribution to connection weight. Different types of interactions and interactions with different artifacts are also weighted differently, and this is accounted for with our CV.

CV: The context variable accounts for differences in weight among artifact type (such as, wikis, discussion boards, and file sharing areas) and interaction type (for example creating, reading and editing and artifact). The context variable includes an artifact factor (AF) and an interaction type factor (ITF). As with the BD and CF variables above, AF and ITF are established for each context through qualitative data analysis. In the case of online courses, a post in response to a user is given an AF of "1" and an ITF of "1." Reads are given an AF of "1" and an ITF of "0.5" or, in some cases (again, depending on qualitative data analysis), the ITF can be calculated based on an analysis of the number of average reads in a discussion board, compared with the reads in a particular discussion board, with less active boards having lower ITF's. Equation 3 illustrates an abstract version of the CV calculation.

Equation 3—Basic CV Calculation

$$CV = ITF \times AF$$

How CV is used in conjunction with the RTW, calculated above, is illustrated in Equation 4.

Equation 4—Full Weight Calculation

$$Weight\,(W) = RTW \times CV$$

Table 4 illustrates how the weights of a set of calculations like those in Table 1 would be calculated, assuming the factors and types of interactions noted in the table. In this example, we assume three types of interactions—read, post and edit; and two types of artifacts—discussion boards and wikis. In our section focused on operationalizing the model, we address the research questions related to each context that will help the users of the Group Informatics methodological approach integrate their qualitative data analysis to produce meaningful numbers for each of the factors outlined here.

The Aggregation of Weighted Interactions (Slicing)

Aggregation provides a summarized view of the interactions described in the first step, and incorporates two additional factors: total interaction count (IC) and a factor for determining how much to influence the interaction count has, relative to the weight of each interaction, represented by W. To do the aggregation, we will build a big table of weights. Each row in the table represents an instance of an interaction between two people, which we then aggregate by pairs.

Interaction count (IC): This is the total interaction count between two nodes in the aggregation unit. It corresponds to a count of the rows that exist between these two nodes, as exemplified in Table 2 above.

DOI: 10.1002/asi

TABLE 4. Example calculations for full weighting, with each row representing a single interaction.

Node 1	Node 2	RTD	RTW	AF	ITF	Weight (W)
Alice	Bob	30	93	Database—1.0	Read—0.5	47.5
Alice	Bob	60	66	Database—1.0	Read—0.5	33
Bob	Alice	5000	0	Database—1.0	Post—1.0	0
Bob	Alice	60	66	Wiki—0.75	Post—1.0	49.5

Note. RTD = raw time distance; RTW = raw time weight; AF = artifact factor; ITF = interaction type factor.

Interaction count weight (ICW): This is a factor that is used to determine the relative importance of interaction count in relation to W. The sum of ICW and interaction weight (IWW) is 1.

IWW: This is the weight to be given to the weight, W, on each row, in the aggregation. The sum of ICW and IWW is 1.

Weight (W): Carried from each row, as examples show in Table 4. Aggregated weight (AW): This is the aggregated weight to be assigned to the connection between to persons, which is the result of the formula.

The Weighting Calculation

The averages represented in Equation 5, then, are averages across the way in which the data is "sliced." Interaction data may be aggregated by standard unit of time, a discrete, context-dependent unit of time like a software release, a qualitative or quantitative category, or by a priori topics. Two of these mechanisms for slicing data rely on time, and the other two rely on categorization of the artifacts and interactions. Although the latter two slices rely on categorization of artifacts, time still plays an integral role in the model as interaction distance is measured according to time distance in all four types of slicing. The four ways we have sliced interaction data and calculated time distance in interactions so far are as follows:

- Standard unit of time (day, week, month, year)
- Discrete, context-dependent unit of time (software release, course module)
- Qualitative or quantitative artifact category
- A priori topic (discussion board threads, software bugs)

The term nW is used to represent a normalized value of W, such that the range of W is between 0 and 1. The term nIC is used to represent a normalized value of IC, such that the range of IC is between 0 and 1.

Equation 5—Aggregated Weight Formula

$$nW = \frac{(W - \min(W))}{(W - \max(W))}$$

$$nIC = \frac{(IC - \min(IC))}{(IC - \max(IC))}$$

$$AW = \left(\overline{nW} \times IWW\right) + (nIC \times ICW)$$

Standard Units of Time

We use standard units of time in two specific ways, both of which treat interactions as raw data for the identification of important relationships between nodes in a network. First, we use standard units of time to explore electronic trace data. This helps to uncover group behavior when there are unidentified, a priori structures to the online groups (disaster relief, for example) or when there are external events that influence the data, but the role of the technical system used by the groups is not known (adult recreational sports, for example). This method of slicing time can be exploratory: a place to begin analysis when there are not a priori, context-dependent units of time, categories, or topics to focus on and a researcher wants to conduct analysis without changing the meaning of the data. Standard units of time are particularly useful for spotting connection between events in a virtual environment and events in the physical world (i.e., disaster relief coordination) during exploratory analysis.

Context-dependent Units of Time

Our previous work utilizes discrete, context-dependent units of time. Most contexts we study lend themselves to analysis of slices determined by the work at hand: examples include software releases and course modules. When we slice the data according to these contextual "time buckets," our analysis recognizes how the purpose of the group, motivations of members, and definition of the tasks to be accomplished frame group experience in different ways within each "time bucket." In an online course or disaster relief scenario, such buckets represent a progression of members through phases with a known end. In the case of software engineering or political discourse, the buckets are more cyclical.

A Priori Topical Categories

Finally, we slice data according to topic. Topics can relate to discourse (e.g., subjects in an online forum) or work (e.g., assignments in a collaborative work environment). These are the natural groupings that are visible in a sociotechnical system as a result of the a priori structure of the STIP. When groups decide to use particular threads for their work, this type of analysis is a rapid way of understanding the group's interaction patterns and is a natural parallel analysis stream to time. This is a slice of analytical convenience for researchers, and should be used with caution, since most online group formation and development does not occur in such neat administrative groupings of interactions. It is also possible that interactions can bleed between topics as individuals become more involved in multiple postings within a STIP.

Qualitatively and Quantitatively Categorized Artifacts

In contrast with the first two approaches to slicing interaction data for analysis, qualitatively determined artifact categories emerge from analysis and coding of discussion threads (or other type of interaction), and the subsequent grouping of all interactions of specific types within a bucket for each code (Glaser & Strass, 1967). Interaction analysis is still dependent on the time distance of the interactions, but in these analytical approaches the interactions are partitioned by categories. This approach allows participant roles and interests to become visible. In our analysis of Facebook political groups, this type of analysis has revealed the existence of "issue entrepreneurs" (Agre, 2004), individuals who participate only in certain topical threads of discourse. In other instances, this approach has identified people who were not among the most active, but who played information-brokering roles across multiple different topical discussions. The model explicitly represents these slices as *artifact categories and collections*.

Quantitative categorizations of interactions emerge from a person-artifact examination of trace data. In software engineering, for example, we identified sets of code artifacts, frequently operated on together, and the sets of developers who operated on those sets. Through this analysis, we developed a measure of developer proximity, which indicates the extent to which groups of developers are virtually near each other. Proximity suggests a need for coordinating communication, and the correspondence between proximity in work and actual, observed coordination behavior is described as sociotechnical congruence (Cataldo, Herbsleb, & Carley, 2008; Cataldo, Mockus, Roberts, & Herbsleb, 2009). Quantitative categorization strategies emerge from qualitative analysis of data, which leads to the identification of the measurable proxies that are available for qualitative phenomena.

The Context Adapted Social Network for Analysis

The result of the first two steps is a weighted social network, aggregated for the analyst's purposes, so that class SNA measures may be applied in a way informed by the context.

Person–artifact interactions. Person-to-artifact interactions must be weighted to reflect the significance of the interaction from the point of view of the actor, that is, the person carrying out some work on a given artifact.

A main component of such a weight is the type of interaction (editing, consultation, transformation, and navigation are among the common interaction types, and they can all be given different weights); we denote that as ITF, as discussed above. If the sociotechnical system supports a fine temporal granularity of interactions recording, and if those interactions can be grouped within the same contextual unit (e.g., as work by the actor in fulfilling a *task*), several of those fine-grained interactions concur in an additive way to the weight, which we call CW. Our computation of CW also includes a cliffing factor CF. As is the case for person-to-person interaction weighing, the determination of the value of CF must be done through qualitative analysis of the domain at hand and the corresponding data; as an example, CF could be set at the median value of all the CW. The final formula for CW, therefore, is shown in Equation 6:

Equation 6—Person-Artifact Interaction Weighting Formula

$$\begin{cases} CW = \sum_{i=1}^{n} ITF_i \ if \ CF < CW' \\ CW = 0 \ if \ CF \geq CW' \end{cases}$$

Where n is the number of interactions with the same artifact being considered as part of the same contextual unit.

The Aggregation of Weighted Interactions (Slicing)

Also person-to-artifact interactions—weighted as described above—may go through a process of aggregation that takes into consideration the time period adopted for interaction slicing and any separate contextual units (e.g., tasks) that occurred within that horizon. In that case—to reflect the accumulated interest of a person for a given artifact—the AW of a person-to-artifact interaction is summative of the nCW values, where nCW are normalized values of CW expressed as in Equation 5. The formula for AW is illustrated in Equation 7:

$$AW = \sum_{i=1}^{m} \sqrt{nCW_i}$$

The Context Adapted Social Network for Analysis

The person-to-artifact interactions described construct a weighted bipartite network, which is one step removed from a social network that enables the analysis of groups, since it captures the interaction between individuals *as mediated by the artifacts upon which those individuals focus their activity within the sociotechnical environment*. Several techniques to transform such a bipartite weighted network into a person-to-person network are, however, well known, and thoroughly described in the literature on social network analysis (see, for example, bicliques; Borgatti & Everett, 1997). When such a transformation is applied, the resulting weighted social network represents the degree of affinity between individuals, measured by means of their interest in the same (or similar) sets of artifacts, and the intensity of that interest. Such a social network is amenable to the same kind of analysis as a social network derived from person-to-person interaction traces. What the network statistics tell us about social phenomena is different in each case, however. We would not, for example, interpret betweenness centralization the same way in person-artifact interactions because the process represented is coordination around an artifact, not a person-to-person transmission process.

Conclusion

SNAGs (small, naturally asynchronous groups) and STIPs (sociotechnical interaction places) are the social and technical constructs where Group Informatics research is focused. Together, they constitute a simple but vital contribution to the ontology needed for describing studies of technologically mediated communities, networks and groups. There are, no doubt, other kinds of groups, like those that exist in physical and virtual space, or perhaps primarily in physical spaces. Our Facebook friends are often an example of a different type of technologically mediated group, since we usually know them from interactions in the physical world. In some respects, Facebook friends and Twitter followers constitute networks of practice (Nardi et al., 2002; Rohde, Klamma, Jarke, & Wulf, 2007) more than small groups or communities (Brown & Duguid, 1991; Wenger, 1998). As social constructions, they differ from SNAGs and STIPs, and likely have distinguishing characteristics. In the act of making our methodological approach explicit, we do not preclude its application or adaptation for studying these other kinds of technologically mediated groups. Instead, through our ontological distinctions of SNAGs and STIPs, we enable the researcher to be clear about the social constructs to which Group Informatics is applied. Our guiding research questions help to make this explicit.

In future studies, we suspect that common patterns we discover across SNAGs and STIPs may apply less directly, or not at all to technologically mediated forms of social organization that include substantial interaction in the physical world, or are focused more on social engagement than task performance. Information science is an interdisciplinary field that is increasingly framing the work within it as sociotechnical. Social media, libraries, scientific collaboration, technologically mediated learning, cooperative work, and archives research inevitably require an understanding of social, technical, and informational components. One cross cutting dimension for sociotechnical information science research is the small group. As a practical matter, small groups are where collaborative work is accomplished within any type of organization. We have argued previously those small, technologically mediated groups, which we now call "SNAGs" are sociotechnical systems (Goggins, Laffey, & Gallagher, 2011).

Models are not right or wrong, but more or less useful (Miller & Page, 2007). Examining technologically mediated (STIPs) small groups (SNAGs) through the lens of a Group Informatics Model and methodological approach enables researchers to systematically ask questions and coordinate research findings across many different systems that leave electronic traces behind. Our specific model is useful to us, and to our collaborators, because it emerges from 5 years of research utilizing electronic trace data, in combination with qualitative research data, to understand small groups as sociotechnical systems. Explaining our model to the information science community enables discourse, revision, challenging, and evolution of this methodological approach and others like it. This article intends to begin a conversation that is vital to the development of scalable information science research in the sociotechnical era. The Group Informatics methodological approach frames the study of SNAGs working inside of STIPs for the research community in a tractable way, and its usefulness is demonstrated by a series of empirical studies. Our model focuses on the interaction as the principle unit of analysis and explicitly recognizes the dynamic nature of context as originally explicated by Dourish (2001, 2004) and the centrality of context for determining how people are connected in sociotechnical systems. Data analysis complementing analysis of raw interaction trace data is necessary to determine theoretically coherent strategies for how relations should be constructed (people-artifacts or people-people) and for measuring the strength of those connections. This basic understanding leads to application of the Group Informatics methodological approach.

We think the conversation about Group Informatics in information science exposes a number of salient, practical concerns for sociotechnical scholars. This article focuses on how to systematically analyze electronic trace data using a methodological approach, and advocates the development of a simple ontology for communicating the results consistently. Other dimensions not addressed in this article include considerations of data collection and management (Butler & Crowston, 2012). We have met this challenge in our work, but there remain few examples where social scientists collaborate around electronic trace data across labs, and many impediments to sharing results–some social, many technical. Projects like Flossmole[2] in open source software provide an example of the kind of electronic trace data sharing infrastructure that is needed to overcome these barriers, but much work remains to be done.

Contributing to Social Informatics

Group Informatics contributes to the social informatics lens for the examination of the social impacts of computing. Kling and Scacchi (1982) were among the first to clearly explicate the reflexive relationships among computing, work, and social organization. Kling, Crawford, Rosenbaum, Sawyer, and Weisband (2000) more recently clarified the focus of social informatics on understanding how personal experience, journalistic reporting, punditry, dystopian

[2]http://www.flossmole.org

opinions, utopian opinions, and public policy among other influences shape ICT use. Lamb and Sawyer (2005) explicated the sociotechnical, social informatics perspective as including notions of interdependent social and technical links between people. Analysis of these links informs our attempts to focus attention on the direct analysis of those links as explicit interactions in a sociotechnical system through Group Informatics.

A methodological approach like Group Informatics provides a mechanism through which social informatics research, emboldened by concrete, identifiable artifact a like electronic trace data, can help to scale social informatics research. We do not claim to extend the tradition of social informatics as a discipline within information science directly, but it is our view that, as Sawyer and Tapia (2007) articulated, scaling social informatics research is an important next step. Like Kling, our work is largely a series of focused examinations of the social effect of computing, and in our case, with a specific focus on the small group unit of analysis. With Group Informatics, we have systematically constructed a methodological approach for studying SNAGs and STIPs. Articulating Group Informatics to the information science community for critique, discourse, and development is an important step for the realization of the scalability of social informatics research and a vital step in the advancement of more general theories of technologically mediated group behavior.

Methodological Reflexivity

Group Informatics calls for explicit methodological reflexivity in the use of electronic trace data to make sense of social science phenomena in technologically mediated small groups. We think the ontology and methodology together provide one candidate approach, built from dozens of empirical studies over the course of 5 years. We do not argue that we have arrived at a single, best approach. Instead, we first point out qualitative research methods like ethnography and ethnomethodology limit the scalability of social computing research focused on small groups. Second, we argue that when computational methods alone are used there are questions of validity and theoretical coherence associated with the results. Without a reflexive stance toward both data and method, the relationship between salient social science constructs and electronic trace data is, at best, unclear. In the worst case, as Howison et al. (2012) point out, using computational methods to analyze trace data in isolation can lead to misleading results.

Finally, we present a systematic methodological approach and ontology aimed at addressing these concerns. Though our approach is aimed at the small group unit of analysis, we think our methodologically reflexive stance may be applied to other units of analysis when electronic trace data are important for making sense of phenomena in social media, participatory media and other similar contexts.

References

Agre, P.E. (2004). The practical republic: Social skills and the progress of citizenship. In A. Feenberg & D. Barney (Eds.), Community in the digital age: Philosophy and practice (pp. 201–223). New York, NY: Rowman & Littlefield.

Ahuja, M.K., & Carley, K.M. (1999). Network structure in virtual organizations. Organization Science, 741–757.

Amelung, C. (2005). A context-aware notification framework for developers of computer supported collaborative environments (Unpublished doctoral dissertation). University of Missouri, Columbia, MO.

Arrow, H., McGrath, J.E., & Berdahl, J.L. (2000). Small groups as complex systems. Thousand Oaks, CA: Sage Publications.

Backstrom, L., Kumar, R., Marlow, C., Novak, J., & Tomkins, A. (2008). Preferential behavior in online groups. Proceedings of WSDM '08 Proceedings of the international conference on Web search and web data mining (pp. 117–128). New York, NY: ACM.

Balkundi, P., & Kilduff, M. (2006). The ties that lead: A social network approach to leadership. The Leadership Quarterly, 17(4), 419–439. Retrieved from http://www.sciencedirect.com/science/article/pii/S1048984306000038

Bandura, A. (1977). Social learning theory. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). Self efficacy: The Exercise of Control. New York: Macmillan.

Bansler, J., & Havn, E. (2006). Sensemaking in technology-use mediation: adapting groupware technology in organizations. Computer Supported Cooperative Work, 15, 55–91.

Bernard, H.R., Killworth, P., Kronenfeld, D., & Sailer, L. (1984). The problem of informant accuracy: The validity of retrospective data. Annual Review of Anthropology, 13, 495–517. Retrieved from http://www.jstor.org/stable/2155679

Black, A., Mascaro, C., Gallagher, M., & Goggins, S.P. (2012). Twitter Zombie: Architecture for capturing, socially transforming and analyzing the Twittersphere. In Proceedings of the 2012 ACM Group Conference, (pp. 229–238). Presented at the ACM Group, Sanibel Island, FL.

Blay-Fornario, M., Pinna-Dery, A.M., Schmidt, K., & Zarate, P. (Eds.). (2002). Cooperative systems design: a challenge of the mobility age. Amsterdam, The Netherlands: IOS Press.

Blincoe, K., Valetto, G., & Goggins, S.P. (2012). Leveraging task contexts for managing developers' coordination. In Proceedings of the 2012 ACM CSCW Conference (pp. 1351-1360). Presented at the Computer Supported Cooperative Work, Seattle, WA: ACM.

Borgatti, S.P., & Everett, M.G. (1997). Network Analysis of Two Mode Data. Social Networks, 19(3), 243–269

Bos, N., Shami, N.S., Olson, J.S., Cheshin, A., & Nan, N. (2004). In-group/out-group effects in distributed teams: an experimental simulation. In Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, Chicago, IL, (pp. 429–436).

Boyd, D., Golder, S., & Lotan, G. (2011). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. Retrieved from http://www.computer.org/portal/web/csdl/doi/10.1109/HICSS.2010.412

Brown, J.S., & Duguid, P. (1991). Organizational learning and communities-of-practice: toward a unified view of working, learning and innovation. Organizational Science, 2(1), 40–57.

Burt, R.S. (2005). Brokerage and closure: An introduction to social capital. Retrieved from http://books.google.com/books?hl=en&lr=&id=ArhGch9zmqwC&oi=fnd&pg=PR9&dq=brokerage+leadership+social+network&ots=ihdZJt_xBC&sig=fidyOMDut6s3MtMjyUw_H0HOJFU

Butler, B., & Crowston, K. (2012, July 30). Data collection and management issues in socio-technical scholarship personal communication.

Cakir, M.P. (2007). How online small groups create and use mathematical objects to do collaborative problem solving. In Proceedings of the 2007 Conference of the International Society of the Learning Sciences, (pp. 839–841).

Carroll, J.M., Neale, D.C., Isenhour, P.L., Rosson, M.B., & McCrickard, D.S. (2003). Notification and awareness: synchronizing task-oriented

DOI: 10.1002/asi

collaborative activity. International Journal of Human-Computer Studies, 58, 605–632.

Carroll, J.M., Rosson, M.B., Farooq, U., & Xiao, L. (2009). Beyond being aware. Information and Organization, 162–185. doi:10.1016/j.infoandorg.2009.04.004

Cataldo, M., Bass, M., Herbsleb, J., & Bass, L. (2007). On coordination mechanisms in global software development. global software engineering, doi:http://doi.ieeecomputersociety.org/10.1109/ICGSE.2007.33

Cataldo, M., Herbsleb, J.D., &Carley, K.M. (2008). Socio-technical congruence: a framework for assessing the impact of technical and work dependencies on software development productivity. In Proceedings of the Second ACM—IEEE International Symposium on Empirical Software Engineering and Measurement, Kaiserslauter, Germany, (pp. 2–11).

Cataldo, M., Mockus, A., Roberts, J.A., & Herbsleb, J.D. (2009). Software dependencies, work dependencies, and their impact on failures. doi:10.1109/TSE.2009.42

Christensen, L.R. (2007). Practices of stigmergy in architectural work. In Proceedings of the 2007 ACM Conference Conference, Sanibel, Island, FL, (pp. 11–20).

Christensen, L.R. (2008). The logic of practices of stigmergy: representational artifacts in architectural design. In Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, San Diego, CA, (pp. 559–568).

Chudoba, K.M., Wynn, E., Lu, M., & Watson-Manheim, M.B. (2005). How virtual are we? Measuring virtuality and understanding its impact in a global organization. Journal of Information Systems, 15, 279–306.

Convertino, G., Moran, T.P., & Smith, B.A. (2007). Studying activity patterns in CSCW. In COmputer Human Interaction 2007 Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, (pp. 2339–2344).

Cox, A., Clough, P., & Siersdorfer, S. (2011). Developing metrics to characterize Flickr groups. Journal of the American Society for Information Science and Technology 62(3), 493–506.

Crabtree, A., O'Neill, J., Tolmie, P., Castellani, S., Colombino, T., & Grasso, A. (2006). The practical indispensability of articulation work to immediate and remote help-giving. In Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (pp. 219–228). Retrieved from http://dl.acm.org/citation.cfm?id=1180910

Cronin, B. (2011). The intelligence disconnect. Journal of the American Society for Information Science and Technology, 62(10), 1867–1868. doi:10.1002/asi.v62.10

Crowston, K., & Howison, J. (2005). The social structure of free and open source software development. First Monday, 10(2). Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1207/1127

Crowston, K., Wiggins, A., & Howison, J. (2010). Analyzing leadership dynamics in distributed group communication. In System Sciences (HICSS), 2010 43rd Hawaii International Conference on System Sciences, Manoa, HI, (pp. 1–10).

Davis, A., Gardner, B.B., Gardner, M.R., & Silver, J.W. (1965). Deep south. Retrieved from http://link.libris.kb.se/sfxumub?sid=google&auinit=A&aulast=Davis&title=Deep+south&genre=book&date=1965

Diani, M. (2003). 5.Leaders' or brokers? Positions and influence in social movement networks. Social Movements and Networks, 1(9), 105–123. Retrieved from http://www.ingentaconnect.com/content/oso/648574/2003/00000001/00000001/art00007

Dourish, P. (2001). Where the action is: Foundations of embodied interaction. Cambridge, MA: MIT Press.

Dourish, P. (2004). What we talk about when we talk about context. Personal and Ubiquitous Computing, 8, 19–30.

Edwards, K.W. (2005). Putting computing in context: An infrastructure to support extensible context-enhanced collaborative applications. ACM Transactions on Computer-Human Interaction, 12(4), 446–474.

Ehrlich, K., & Cash, D. (1999). The invisible world of intermediaries: A cautionary tale. Computer Supported Cooperative Work, 8, 147–167.

Falkowski, T., Bartelheimer, J., & Spiliopoulou, M. (2006). Mining and visualizing the evolution of subgroups in social networks. In Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (pp. 52–58). Retrieved from http://dl.acm.org/citation.cfm?id=1248823.1249048

Fine, G.A., & Harrington, B. (2004). Tiny publics: Small groups and civil society. Sociological Theory, 22(3), 341–356. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.0735-2751.2004.00223.x/abstract

Fleming, L., & Waguespack, D.M. (2007). Brokerage, boundary spanning, and leadership in open innovation communities. Organization Science, 18(2), 165. Retrieved from http://orgsci.highwire.org/cgi/content/abstract/18/2/165

Fletcher, J.K., & Kaufer, K. (2003). Shared leadership. Shared leadership: Reframing the hows and whys of leadership, 21–47.

Freeman, L.C. (1979). Centrality in social networks conceptual clarification. Social Networks, 1(3), 215–239. Retrieved from http://www.sciencedirect.com/science/article/pii/0378873378900217

Freeman, L.C. (2003). Finding social groups: A meta-analysis of the southern women data. Dynamic social Network modeling and analysis, 39–97.

Freeman, L.C. (2004). The development of social network analysis. Vancouver, BC: Empirical Press.

Freeman, L.C., Romney, A.K., & Freeman, S.C. (1987). Cognitive structure and informant accuracy. American anthropologist, 89(2), 310–325. Retrieved from http://onlinelibrary.wiley.com/doi/10.1525/aa.1987.89.2.02a00020/abstract

Friedkin, N.E. (1991). Theoretical foundations for centrality measures. American journal of Sociology, 1478–1504. Retrieved from http://www.jstor.org/stable/2781908

Fuller, M.A., Hardin, A.M., & Davison, R.M. (2007). Efficacy in technology-mediated distributed teams. Journal of Management Information Systems, 23(3), 209–235. doi:10.2753/MIS0742-1222230308

Fuller, M.A., Hardin, A.M., & Scott, C.L. (2007). Diffusion of virtual innovation. The DATA Base for Advances in Information Systems, 38(4), 40–45.

Geiger, R.S., & Ribes, D. (2011). Trace Ethnography: Following Coordination through Documentary Practices. In Proceedings of the 44th Hawaii International Conference on Systems Science, Koloa, Kauai, HI, (pp. 1–10).

Geister, S., & Hertel, G. (2006). Effects of process feedback on motivation, satisfaction, and performance in virtual teams. Small Group Research, 37, 459–489.

Glaser, B., & Strass, A. (1967). The discovery of grounded theory: Strategies for qualitative research. Chicago, IL: Aldine Publishing Company.

Goggins, S.P. (2006). Implications of Group Size on Activty Awareness Tools and Consideration of Information Foraging Theory. Presented at the Workshop on Activity Awareness at CSCW 2006, Banff, Alberta, Canada: ACM. Retrieved from http://www.seangoggins.net

Goggins, S.P., Gallagher, M., Laffey, J., & Amelung, C. (2010). Social Intelligence in Completely Online Groups—Toward Social Prosthetics from Log Data Analysis and Transformation. In Proceedings of the IEEE International Conference on Social Computing, Minneapolis, MN, (pp. 500–507).

Goggins, S.P., Galyen, K., & Laffey, J. (2010). Network Analysis of Trace Data for the Support of Group Work: Activity Patterns in a Completely Online Course. In Proceedings of the 2010 ACM Group Conference (pp. 107–116). Sanibel Island, FL: ACM.

Goggins, S.P., Laffey, J., & Amelung, C. (2011). Context Aware CSCL: Moving Toward Contextualized Analysis. In Proceddings of CSCL 2011, Hong Kong, (pp. 591–596).

Goggins, S.P., Laffey, J., & Gallagher, M. (2011). Completely online group formation and development: small groups as socio-technical systems. Information Technology & People, 24(2), 104–133.

Goggins, S.P., Laffey, J., &Galyen, K. (2009). Social Ability in Online Groups: Representing the Quality of Interactions in Social Computing Environments. In Proceedings of the 1st IEEE Conference on Social Computing,Vancouver, Canada, (pp. 667–674).

Goggins, S.P., Laffey, J., & Tsai, I. (2007). Cooperation and groupness: Community formation in small online collaborative groups. In Proceedings of the 2007 International ACM Conference on Supporting Group Work, Sanibel Island, FL, (pp. 207–216).

Goggins, S., & Mascaro, C. (2013). Context Matters: The Experience of Physical, Informational and Cultural Distance in a Rural IT Firm. The Information Society, in press.
Goggins, S.P., Mascaro, C., & Mascaro, S. (2012). Relief after the 2010 Haiti Earthquake: Participation and Leadership in an Online Resource Coordination Network. In Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work, Seattle, WA, (pp. 57–66).
Goggins, S.P., Schmidt, M., Moore, J.L., & Guajardo, J. (2011). 3D virtual worlds: Assessing the experience and informing design. International Journal of Social and Organizational Dynamics in IT, 1(1), 30–48.
Goggins, S.P., & Valetto, G. (2010). An Interaction Warehouse for Workgroup Development Pattern Identification. Presented at the IFIP 9.1 Workshop, St. Louis, MO. Retrieved from http://www.jpedia.org/ifipwg91
Goggins, S.P., Valetto, P., Mascaro, C., & Blincoe, K. (2012). Creating a model of the dynamics of socio-technical groups using electronic trace data. doi:10.1108/09593841111137322
Golbeck, J., Grimes, J.M., & Rogers, A. (2010). Twitter use by the U.S. Congress. Journal of the American Society for Information Science and Technology, 61(8), 1612–1621. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/asi.21344/full
Granovetter, M.S. (1973). The strength of weak ties. American journal of sociology, 1360–1380. Retrieved from http://www.jstor.org/stable/2776392
Grudin, J. (1994). Eight challenges for developers. Communications of the ACM, 37, 92–116.
Grudin, J. (2010). Timelines: CSCW: Time passed, tempest, and time past. Interactions, 17(4), 38–40.
Gutwin, C., & Greenberg, S. (2004). The importance of awareness for team cognition in distributed collaboration. In E. Salas and S.M. Fiore (Eds.), Team cognition: Understanding the factors that drive process and performance (pp. 177–201). Washington, DC: APA Press.
Gutwin, C., Penner, R., & Schneider, K. (2004). Group awareness in distributed software development. In Proceedings of the 2004 ACM Conference On Computer Supported Cooperative Work, Chicago, IL (pp. 72–81).
Harrison, S., & Tatar, D. (2007). Places: people, events, loci–the relation of semantic frames in the construction of place. Computer Supported Coorperative Work (CSCW), 17(2), 97–133.
Healy, P.G.T., White, G., Eshghi, A., Reeves, A.J., & Light, A. (2007). Communication spaces. Computer Supported Cooperative Work, 16.
Heverin, T., & Zach, L. (2011). Use of microblogging for collective sensemaking during violent crises: A study of three campus shootings. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/asi.21685/full
Hinds, P., & McGrath, C. (2006). Structures that work: social structure, work structure and coordination ease in geographically distributed teams. In Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, Banff, Alberta, Canada, (pp. 343–352).
Howison, J., Wiggins, A., & Crowston, K. (2012). Validity issues in the use of social network analysis with digital trace data. Journal of the Association of Information Systems, 12(2)(2).
Howley, I., Mayfield, E., & Rosé, C.P. (2012). Linguistic analysis methods for studying small groups. Retrieved from http://learnlab.org/research/wiki/images/5/58/Chapter-Methods-Revised-Final.pdf
Jansen, B.J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 60(11), 2169–2188. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/asi.21149/full
Kaptelinin, V., & Nardi, B.A. (2006). Acting with technology: Activity theory and interaction design (acting with technology). Retrieved from http://www.amazon.com/gp/redirect.html%3FASIN=0262112981%26tag=ws%26lcode=xm2%26cID=2025%26ccmID=165953%26location=/o/ASIN/0262112981%253FSubscriptionId=0NM5T5X751JWT17C4GG2
Kiesler, S., Boh, W.F., Ren, Y., & Weisband, S. (2005). Virtual teams and the geographically dispersed professional organization. In Proceedings from The Workshop on Extending the Contributions of Professor Rob Kling to the Analysis of Computerization Movements. Presented at the ASIS&T, Charlotte, NC.
King, G. (2011). Ensuring the data rich future of the social sciences. doi:10.1126/science.1197448
Kittur, A., Chi, E., Pendleton, B.A., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. World Wide Web, 1(2), 19.
Kittur, A., Lee, B., & Kraut, R. (2009). Coordination in collective intelligence: the role of team structure and task interdependence. In Proceedings of the 27th International Conference on Human Factors in Computing Systems, Boston, MA, (pp.1495–1504).
Kling, R. (1979). Social issues and impacts of computing: From arena to discipline. Human Choice and Computers, 2, 146-155.
Kling, R. (1980). Social analyses of computing: Theoretical perspectives in recent empirical research. ACM Computing Surveys (CSUR), 12(1), 61–110.
Kling, R. (2007). What is social informatics and why does it matter? The Information Society, 23(4), 205–220. doi:10.1080/01972240701441556
Kling, R., & Courtright, C. (2004). Group behavior and learning in electronic forums: A sociotechnical approach. In S. Barab, R. Kling, & J.H. Gray (Eds.), Designing for virtual communities in the service of learning (pp. 91–119). New York, NY: Cambridge University Press.
Kling, R., Crawford, H., Rosenbaum, H., Sawyer, S., & Weisband, S. (2000). Learning from social informatics: Information and communication technologies in human contexts. Center for Social Informatics, Indiana University.
Kling, R., & Scacchi, W. (1982). The web of computing: Computer technology as social organization. Advances in computers, 21, 1–90. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0065245808605677
Kraut, R., & Streeter, L. (1995). Coordination in software development. Communications of the ACM, 38(3), 69–81.
Krippendorff, K. (2004). Content analysis: An introduction to its methodology. Thousand Oaks, CA: Sage Publications.
Laffey, J., Amelung, C., & Goggins, S.P. (2009). A context awareness system for online learning: Design based research. International Journal on E-Learning, 8(3), 313–330.
Laffey, J., Amelung, C., Tsai, I.-C., Hong, R.-Y., Goggins, S.P., Reid, D., & Turner, P. (2008). Design research for context awareness. Proceedings of ACM CHI 2008 (in review), Florence, Italy.
Laffey, J., Galyen, K., Reid, D., & Goggins, S.P. (2012). Measurements of Group Social Interaction in an Online Medical Software Support Community, in development.
Laffey, J., Lin, G.Y., & Lin, Y. (2006). Assessing social ability in online learning environments. Journal of Interactive Learning Research, 17(2), 163–177.
Lamb, R., & Sawyer, S. (2005). On extending social informatics from a rich legacy of networks and conceptual resources. Information Technology & People, 18(1), 9–20. doi:10.1108/09593840510584595
Lampe, C., Ellison, Ni. B., & Steinfield, C. (2008). Changes in User Perception of Facebook. In Proceedings of the Computer Supported Cooperative Work, San Diego, CA, (pp. 721–730).
Laumann, E.O. (2006). A 45-Year Retrospective on Doing Networks. Connections, 27(1), 65–90.
Laumann, E.O., & L.P., S. (1997). Measuring social networks using samples, Is network analysis relevant to survey research. In J. Bancroft (Ed.), Researching sexual behavior: Methodological issues (pp. 390–416). Bloomington, IN: Indiana University Press.
Laumann, E.O., Marsden, P.V., & Prensky, D. (1989). The boundary specification problem in network analysis. Research methods in social network analysis, 61, 87.
Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., . . . Van Alstyne, M. (2009). Social science. Computational social science. Science, 323(5915), 721–723. doi:10.1126/science.1167742
Leinonen, P., Jarvela, S., & Hakkinen, P. (2005). Conceptualizing the awareness of collaboration: A qualitative study of a global virtual team. Computer Supported Cooperative Work, 14, 301–322.

Liu, P.-J., Laffey, J., & Cox, K. (2008). Operationallization of Technology Use and Cooperation in CSCW. In Proceedings of the CSCW, San Diego, CA, (pp. 505–514).

Mascaro, C., & Goggins, S.P. (2011a). Brewing Up Citizen Engagement: The Coffee Party on Facebook. In Proceedings of Communities and Technologies 2011, Brisbane, Australia, (pp. 11–20).

Mascaro, C., & Goggins, S.P. (2011b). The Daily Brew: The Structural Evolution of the Coffee Party on Facebook During the 2010 United States Midterm Election Season. In Proceedings of the 4th Annual Political Networks Conference, Ann Arbor, MI.

Mascaro, C., & Goggins, S.P. (2013). Coffee or Tea: The Emergence of Networks of Discourse in Two Online Political Groups. Journal of Information Technology and Politics, accepted.

Mascaro, C.M., Novak, A., & Goggins, S.P. (2012). Shepherding and Censorship: Discourse Management in the Tea Party Patriots Facebook Group. In Proceedings of the 45th Hawaii International Conference on System Sciences, Wailea, Maui, HI, (pp. 2563–2572).

McGrath, J.E. (1984). Groups: Interaction and performance. Retrieved from http://users.ece.utexas.edu/~perry/education/382v-s08/papers/mcgrath84.pdf

McGrath, J.E. (1997). Small group research, that once and future field: An interpretation of the past with an eye to the future. Group dynamics: Theory, Research, and Practice, 1(1), 7. Retrieved from http://psycnet.apa.org/journals/gdn/1/1/7/

Mead, M. (1934). Kinship in the admiralty islands. Retrieved from http://books.google.com/books?hl=en&lr=&id=8Fyj0h0G_OUC&oi=fnd&pg=PR9&dq=mead+"small+group"+"M+mead"&ots=-7aKCAORBO&sig=IzsWex2tPvK2vmGGWBpfm-6OFXs

Mead, M. (1958). Cultural determinants of behavior. Behavior and evolution, 480–503.

Miksa, S.D., Burnett, K., Bonnici, L.J., & Kim, J. (2007). The development of a facet analysis system to identify and measure the dimensions of interaction in online learning. Journal of the American Society for Information Science and Technology, 58(11), 1569–1577. doi:10.1002/(ISSN)1532-2890

Miller, J.H., & Page, S.E. (2007). Complex adaptive systems: An introduction to computational models of social life. Princeton, NJ: Princeton University Press.

Naaman, M., Becker, H., & Gravano, L. (2011). Hip and trendy: Characterizing emerging trends on Twitter. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/asi.21489/full

Nardi, B. (1996). Context and consciousness. Boston, MA: Massachusetts Institute of Technology.

Nardi, B., & Harris, J. (2006). Strangers and friends: Collaborative play in World of Warcraft. In Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, Banff, Alberta, Canada, (pp. 149–158).

Nardi, B., Whittaker, S., & Schwarz, H. (2002). NetWORKers and their activity in Intensional Networks. Computer Supported Cooperative Work, 11, 205–242.

Neale, D.C., Carroll, J.M., &Rosson, M.B. (2004). Evaluating computer-supported cooperative work: models and frameworks. In Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, Chicago, IL, (pp. 112–121).

Novak, A., & Mascaro, C. (2012). Ti-Squaring around: An analysis of e-jargon and group roles in an online community. In Proceedings of the American Association for Applied Linguistics, Boston, MA.

Ocker, R.J., & Fjermestad, J. (2008). Communication differences in virtual teams: Findings from a multi-method analysis of high and low performing experimental teams. The Data Base for Advances in Information Systems, 39(1), 51–68.

Olson, G.M., Atkins, D.E., Clauer, R., Finholt, T.A., Jahanian, F., Killeen, T.L., . . . Weymouth, T. (1998). The upper atmospheric research collaboratory (UARC). interactions, 5(3), 48–55. Retrieved from http://portal.acm.org/citation.cfm?id=275269.275276&dl=portal&dl=ACM&idx=J373&part=periodical&WantType=periodical&title=interactions

Olson, G.M., Herbsleb, J.D., & Rueter, H.H. (1994). Characterizing the sequential structure of interactive behaviors through statistical and grammatical techniques. Human-Computer Interaction, 9(4), 427–472.

Olson, G.M., Malone, T.W., & Smith, J.B. (Eds.). (2001). Coordination theory and collaboration technology. Mahwah, NJ: Lawrence Erlbaum.

Olson, G.M., & Olson, J.S. (2000). Distance matters. Human Computer Interaction, 139–178.

Olson, G., Olson, J., & Venolia, G. (2009). What still matters about distance. In Proceedings of the HCIC. Retrieved from http://research.microsoft.com/pubs/78697/olson9370.pdf

Olson, J.S., Olson, G.M., Storrøsten, M., & Carter, M. (1992). How a group-editor changes the character of a design meeting as well as its outcome. In Proceedings of the 1992 ACM Conference on Computer Supported Cooperative Work, Toronto, Canada, (pp. 91–98).

Parigi, P., & Bearman, P.S. (2005). Cloning headless frogs and other important matters: Conversation topics and network structure. Social Forces, 83(2), 535–557. Retrieved from http://muse.jhu.edu/journals/sof/summary/v083/83.2bearman.html

Powell, A., Piccoli, G., & Ives, B. (2004). Virtual teams: A review of current literature and directions for future research. The Data Base for Advances in Information Systems, 35(1), 6–36.

Roberts, T.L., Lowry, P.B., & Sweeney, P.D. (2006). An evaluation of the impact of social presence through group size and the use of collaborative software on group member "voice" in face-to-face and computer-mediated task groups. IEEE Transactions on Professional Communication, 49(1), 28–43.

Roethlisberger, F.J., Dickson, W.J., & Wright, H.A. (1939). Management and the worker: An account of a research program conducted by the Western Electric Company, Hawthorne Works, Chicago. Retrieved from http://www.getcited.org/pub/101170564

Rohde, M., Klamma, R., Jarke, M., & Wulf, V. (2007). Reality is our laboratory: communities of practice in applied computer science. Behaviour and Information Technology, 26(1), 81–94.

Sacks, H. (1972). An initial investigation of the usability of conversational data for doing sociology. Studies in social interaction, 3, 74.

Saunders, C.S., & Ahuja, M.K. (2006). Are all distributed teams the same. Small Group Research, 37(6), 662–700.

Sawyer, S. (2009). Advancing social informatics. Proceedings of the 5th Annual Social Informatics Research Symposium on "People, Information, Technology: The Social Analysis of Computing in a Diverse and Pluralistic World," ASIST Special interest group on social informatics, Vancouver, BC.

Sawyer, S., & Tapia, A. (2007). From findings to theories: institutionalizing social informatics. The Information Society, 23(4), 263–275. doi:10.1080/01972240701444196

Schmidt, K., & Bannon, L. (1992). Taking CSCW seriously. Computer Supported Cooperative Work (CSCW), 1(1), 7–40.

Schmidt, K., & Wagner, I. (2004). Ordering systems: Coordinative practices and artifacts in architectural design and planning. Computer Supported Cooperative Work (CSCW), 13(5), 349–408.

Scott, J. (2000). Social network analysis: A handbook (2nd ed.). London, UK: Sage.

Semar, W. (2005). Development of a benchmark system for analyzing collaborative group performance as part of an educational online knowledge management system. In Proceedings of the 2005 International Conference on Information and Knowledge Engineering—IKE, Las Vegas, NV, (Vol. 5, pp. 53–59).

Sonnenwald, D.H. (1999). Evolving perspectives of human information behavior: Contexts, situations, social networks and information horizons. In T. Wilson & D. Allen (Eds.), Exploring the contexts of information behavior (pp. 176–190). London, UK: Taylor Graham.

Sonnenwald, D.H. (2005). Information Horizons. In K. Fischer, S. Erdelez, & L.E.F. McKechnie (Eds.), Theories of information behavior (pp. 191–197). Medford, NJ: ASIS&T.

Sonnenwald, D.H., Lassi, M., Olson, N., Ponti, M., & Axelsson, A.-S. (2009). Exploring new ways of working using virtual research environments in library and information science. Library Hi Tech, 27(2), 191–204. doi:10.1108/07378830910968155

Stahl, G. (2002). Contributions to a theoretical framework for CSCL. In Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community, Boulder, CO, (pp. 62–71).
Stahl, G. (2006). Group cognition: Computer support for building collaborative knowledge. Boston, MA: MIT Press.
Stahl, G. (2009a). For a science of group interaction. In Proceedings of the ACM 2009 International Conference on Supporting Group Work, Sanibel Island, FL (pp. 129–138).
Stahl, G. (2009b). Meaning making in VMT. Retrieved from http://www.ischool.drexel.edu/faculty/gerry/vmt/book/26.pdf
Stahl, G. (2009c). Toward a science of group cognition. Retrieved from http://www.ischool.drexel.edu/faculty/gerry/vmt/book/28.pdf
Star, S.L., & Griesemer, J.R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. Social Studies of Science, 19, 387–420.
Star, S.L., & Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. Computer Supported Cooperative Work, 8, 9–30.
Straus, A. (1993). Continual permutations of action. Hawthorne, NY: Aldine De Gruyter.
Suthers, D.D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. International Journal of Computer-Supported Collaborative Learning, 5(1), 5–42. doi:10.1007/s11412-009-9081-9
Tajfel, H. (1978). Differentiation between social groups. New York, NY: Academic Press.
Tajfel, H. (1982). Social identity and intergroup relations. London, UK: Cambridge University Press.
Teasley, S.D., Covi, L.A., Krishnan, M.S., & Olson, J.S. (2002). Rapid software development through team collocation. IEEE Transactions on Software Engineering, 671–683.
Thelwall, M., Buckley, K., & Paltoglou, G. (2011a). Sentiment in Twitter events. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/asi.21462/full
Thelwall, M., Buckley, K., & Paltoglou, G. (2011b). Sentiment strength detection for the social web. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/asi.21662/full
Travica, B. (2005). Information view of organization. Journal of International Technology and Information Management, 14(3), 1–20.
Tsai, I.C., Kim, B., Liu, P.J., Goggins, S.P., Kumalasari, C., & Laffey, J.M. (2008). Building a model explaining the social nature of online learning. Educational Technology & Society, 11(3), 198–215.
Turner, J.C., Brown, R.J., & Tajfel, H. (1979). Social comparison and group interest in ingroup favoritism. European Journal of Social Psychology, 9, 187–204.
Turner, W., Bowker, G.C., Gasser, L., & Zacklad, M. (2006). Information infrastructures for distributed collective practices. Computer Supported Cooperative Work, 15, 93–110.
Twidale, M., & Nichols, D. (1998). A survey of applications of CSCW for digital libraries: Technical Report CSEG/4/1998. Retrieved from http://eprints.lancs.ac.uk/53451/
Valente, T.W. (1996). Social network thresholds in the diffusion of innovations. Social Networks, 18(1), 69–89. Retrieved from http://www.sciencedirect.com/science/article/pii/0378873395002561
Wasserman, S. (1994). Social network analysis: Methods and applications. books.google.com,.
Wenger, E. (1998). Communities of Practice: Learning, Meaning and Identity. New York, NY: Cambridge University Press.
Whittaker, S. (1996). Talking to strangers: An evaluation of the factors affecting electronic collaboration. In Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work, Cambridge, MA, (pp. 409–418).
Wiggins, A. (2011). eBirding: technology adoption and the transformation of leisure into science. In Proceedings of the 2011 iConference, Seattle, WA, (pp. 798–799).
Yuan, C.Y., Gay, G., & Hembrooke, H. (2006). Focused activities and the development of social capital in a distributed learning "community." The Information Society, 22, 25–39.

Appendix

Detailed Description of Group Informatics Model Components

Interactions

The interaction component is the most important aspect of the model. Interactions are a fine-grained representation of the human affiliations and human-artifact connections embodied by task work that occurs in a STIP. Each interaction represents a person performing some action, often reading, posting, or editing an artifact, following another individual or group of individuals. Order is discerned from a sequencing of interactions according to their timestamps. Since our interactions are derived from electronic trace data, there are subtleties of precision, completeness, and time to the record of trace data. Our model allows us to compensate for distinctly different environments in which interactions occur. For example, we find in discussion boards associated with recreational sports leagues that interactions can occur over days or weeks. In environments such as political discourse on Facebook, we identify rapid, pseudo-synchronous interaction that closely resembles a chat room.

Other contexts such as online learning and software engineering differ depending on the artifact and the type of interaction that occurs around it. Through context-specific operationalizations of the Group Informatics method and model, explained throughout the article, we identify how these contexts vary and describe how we account for them in our computation and analysis.

Time Distance

Ordinarily, interactions in a technical system have a time stamp. The model we present relies on the timestamp information as a measure of the distance between two interactions, between either people and people or people and artifacts in the raw data. We use that timestamp to calculate the time distance between an interaction and one or more interactions prior to it. Since interactions occur around artifacts and people, the way time distances are calculated is a function of how we choose to slice the data. The published studies from which the Group Informatics model has evolved incorporate data sliced in one of four ways, though we imagine other slices are possible as the model is applied in new domains.

Type (Mode)

Our model supports two types of interactions: (a) individual posts or contributions to the discourse in an identifiable

DOI: 10.1002/asi

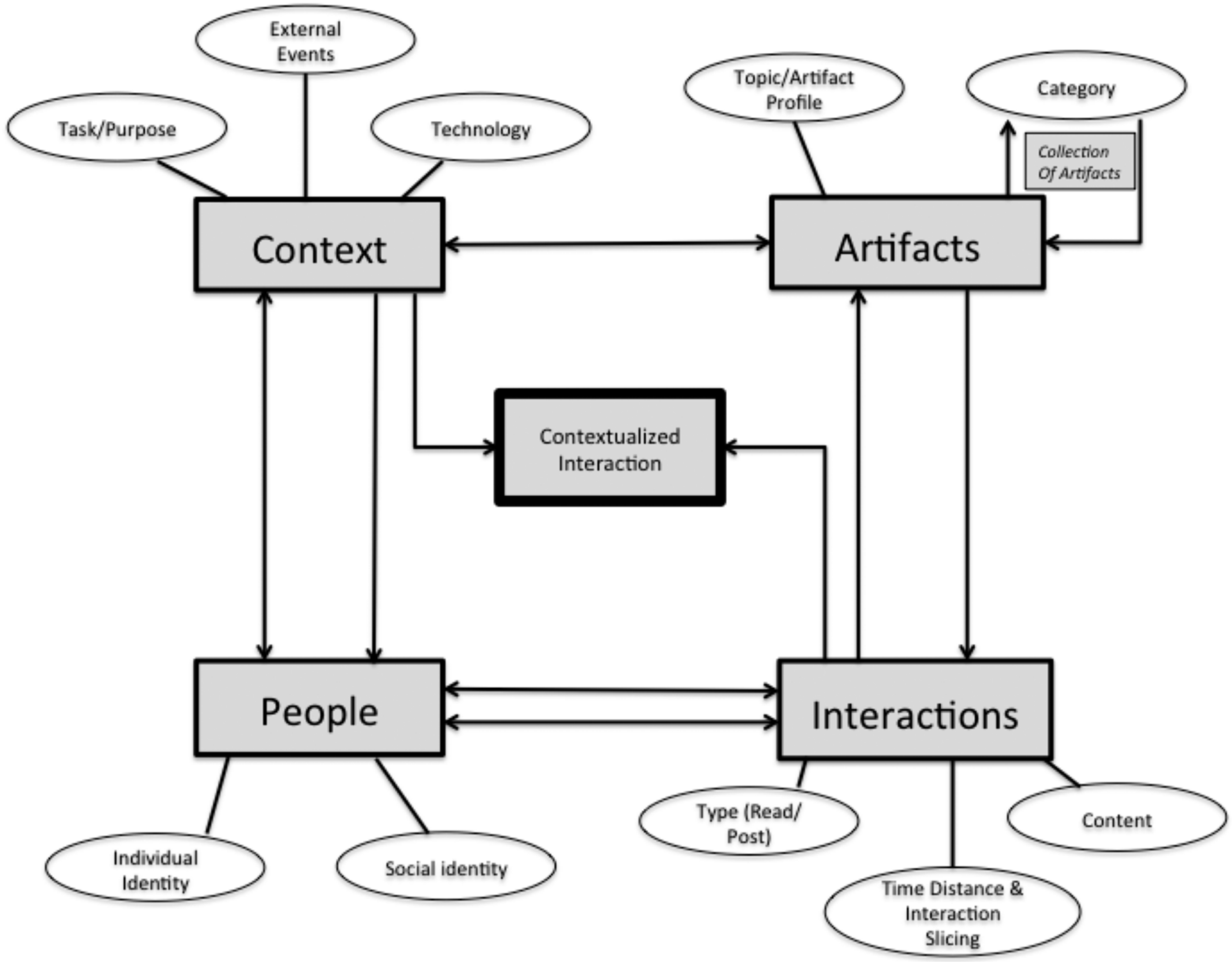

FIG. A1.

manner and (b) passive interaction with an artifact such as reading it without posting a response. The latter type of interaction is less commonly recorded in electronic trace date of most sociotechnical systems. In our model, we give each type of interaction a different weight because the implications of each mode are different. The inclusion of both types of interactions allows for the identification of individuals who play distinctly different but important roles in information behavior. The existence of individuals who play important roles in posting data does not always correlate to those who play important roles in lurking behavior. In our analysis of disaster relief scenarios, we find that key brokers of information do not emerge from easily visible descriptive statistics such as those actors who were most active. Instead, we find that when we take into account read post data together utilizing the lens of social network analysis measures of betweenness that important actors emerged, indicating brokerage between two groups of actors in the system (Goggins, Valetto, Mascaro, & Blincoe, 2012).

Content

Interaction content is sometimes analyzed in conjunction with Group Informatics modeling. Specific content within a comment may indicate a specific set of interactions. In cases involving political discourse and online recreational sports leagues, we look for syntax referencing a direct addressal between individuals or argumentative words (Boyd, Golder, & Lotan, 2011; Mascaro, Novak, & Goggins, 2012). This allows for an explicit identification of conversational networks between individuals that may contribute to a deeper understanding of the activity within the forum.

In certain discussion forums and social networks, specific technological affordances are made to facilitate such interactions. Understanding what these are and how they are used is an integral part of the Group Informatics model that we specifically address in the technology dimension of the context component. For example, in the case of Twitter, individuals are able to re-tweet each other's tweets or

DOI: 10.1002/asi

mention another user. These actions may have distinctly different meanings depending on the content (Boyd et al., 2011). The analyses are complementary, and can be used to classify patterns of interaction and explain how these patterns each contribute to understanding the group through analysis in the Group Informatics model. Qualitative analysis that informs the development of weighting equations represents a researcher's application of the Group Informatics model. How the weighting equations and qualitative analysis reflexively inform each other is different in each context of study.

Context

Context exists at the intersection of an artifact and interaction around that artifact. It is co-constructed through these components of the model. This dynamically constructed context, when embodied in a technical system that provides electronic trace data in the form of read and post activity of users (conceptualized as interaction types) around artifacts, becomes a construct in Group Informatics which can be weighted, measured, and compared. In sociotechnical research, generally, and the Group Informatics model, specifically, context is a dynamic, technologically mediated, and socially constructed setting (Dourish, 2004). The misconception of context as a wrapper around technologically mediated activity is common, as Dourish points out.

In the case of online learning, we show how context can be more visible to users through tools that integrate knowledge of activity, people, and the artifacts around which they operate (Goggins, Galyen, & Laffey 2010; Goggins, Schmidt, Moore, & Guajardo, 2011; Goggins, Laffey, Amelung, 2011; Goggins, Laffey, & Gallagher, 2011; Laffey, Amelung, & Goggins, 2009). Our recent work developing measures for software engineering shows similar results for making software developers aware of with whom they are interacting– knowingly or otherwise–within a distributed software project and through the mediation of the software artifacts they work on (Blincoe et al., 2012).

Technology

Technology frames context in the Group Informatics model. Technology is often adapted and appropriated by people in ways not intended by the designers (Harrison & Tatar, 2007), placing both constraints and new freedoms on users who interact through STIPs. This consideration of technology design as a qualitative data analysis factor in the construction of models for specific STIPs is at the core of what distinguishes application of the Group Informatics model from most analyses of electronic trace data. Specific technological affordances enable specific types of interactions and this may lead to adopted patterns of interaction within STIPs.

As noted in the content dimension of the interactions component, certain technological affordances can dramatically shift the manner in which individuals interact with each other through technology. It is important to take these technological affordances and patterns of interaction into account to understand the technological affordances of the STIP and to make sense of the electronic trace data. In previous research on political discourse, we identify how Facebook enables users to directly address each other in discourse using the "@" symbol, but that adoption of this affordance is in fact low (Mascaro, Novak, & Goggins, 2012). Understanding the reasons for this failure of adoption and the existence of other user-invented mechanisms for marking dialogue is integral to understanding interactions within a technological context.

External Events

Online interactions do not exist in a bubble. Interactions and experiences among groups, even technologically mediated groups who maintain electronic trace data of all of their interactions, are influenced by external events. In some of our studies, the technologically mediated interactions analyzed through our methods and for which we have constructed specific instantiations of the model served to support activities in the physical world. Our study of disaster relief (Goggins, Valetto, Mascaro, & Blincoe, 2012), online political discourse (Mascaro & Goggins, 2011a, 2011b), and forthcoming work on adult recreational sports leagues are all influenced to varying degrees by external events.

The context of the interaction trace data we analyze is influenced by this external context. In the case of disaster relief, online discussion forums provide coordination outlets in response to specific events happening on the ground. In online political discourse groups, the administrators of the group incorporate specific discussion topics from the physical world into the virtual context to tie the discourse occurring online with real-world events. This coupling along with the existing membership of the group influences the discourse that occurs. In other cases, such as online recreational sports leagues, the online environment is an extension of the physical environment in which individuals interact when not in the presence of each other. Therefore, certain external events such as new seasons, social events, and the formation and dissolution of teams are reflected in the electronic trace data. Researchers must be aware of these events and triangulate the occurrences to effectively make sense of the electronic trace data.

By constructing a model, we are implicitly arguing that external events can be explicitly accounted for in each STIP where we apply our model. Developing an understanding of how external events affect the interactions within the STIP helps to determine the appropriate interaction time distance to be used and how analysis needs to be structured. In some cases, we examine external events through interviews and observations with participants. In the case of disaster relief, we created a timeline of the events on the ground following the 2010 Earthquake in Haiti and correlated those with events in the disaster relief coordination system we studied. The role of external events is influenced by the task and purpose of the group, which we discuss next.

Task or Purpose

Others have noted the significance of individual motivations in group performance (Geister & Hertel, 2006; Semar, 2005). Even at the group level, satisfaction of some member need is one of three well-established purposes for group existence (McGrath, 1997); the other two being accomplishment of some task and maintenance of the group. We place this consideration as a dimension of the context component that exists between the artifact and the individual. Understanding the purpose of the STIP allows for researchers to develop a further understanding of the activity and work towards identifying successful undertakings.

Our understanding of any group's purpose emerges from qualitative methods like content analysis, interviews, and open coding of content. Some groups, like open source software engineering teams or those in online courses, have a purpose that is relatively transparent and conceptually tight. The work is task-bounded and time-bounded in a clear way. Participation and membership are relatively stable in any reasonable slice of the data. Other contexts have emergent, multivalent purposes that only surface through in depth qualitative analysis methods. Disaster relief and online political discourse are two examples where our understanding of the different motivations and interests of members emerged only after in-depth analysis of the data. In these cases, our initial network analysis is exploratory and iterative. As we learn more about the member experiences and intentions, the weighting of connections changes and the time slices evolve to allow us to make better sense of the data.

Artifacts

An artifact is a nonhuman object around which people collaborate. In activity theory, artifacts are sometimes more generally referred to as objects. As in activity theory, the artifact component of the Group Informatics model is expected to evolve through use. Artifacts progress forward and do not devolve; they are ever-evolving work products like software code, although in some cases artifacts may stagnate such as a discussion topic that no longer has any discourse. Examples of discussion stagnation can be found in most any online courses. We have also observed that Facebook political discussions in specific political groups have lifespans that tend to last no longer than 72 hours (Mascaro & Goggins, 2011a).

Topic and Artifact Profile

Artifacts can comprise a topic or artifact profiles. By profile, we simply mean "category." These two constructs are related to each other, and we have elected to represent them as a single dimension in the model because they tend to be mutually exclusive. Discussion-oriented artifacts typically have a topic. This may be captured in a subject line or at the discussion forum or website level of analysis. For example, if you are participating in a public discussion for people who have cancer, the topic level analysis may be "skin cancer," and all discussion of skin cancer survival or treatment might be considered part of the same topic in our analysis. Another possibility would be as in the case of an online course, where the subject line of the first post in a thread becomes the topic in our model.

Artifact focused interactions, like those found in software engineering data sets (Blincoe et al., 2012), are more apt to have profiles than topics. The profile of an artifact in this context is twofold. First, it could be source code or comments related to source code in a bug tracking system. In both cases, there is a relationship between a work object and the activities of the participant. Unlike discourse focused contexts, the artifact is a specific "thing" that mediates interaction.

Second, in some cases these artifacts may be thought of as boundary objects (Star & Griesemer, 1989; Turner, Bowker, Gasser, & Zacklad 2006) that aid in communication and coordination about the work at hand. In other cases, the artifacts are opportunities (not always acted upon) for coordination between people (Blincoe et al., 2012; Cataldo et al., 2008; Cataldo et al., 2009; Cataldo, Bass, Herbsleb, & Bass, 2007). The boundaries in these cases are unrecognized.

Category and Collection of Artifacts

Categories and collections of artifacts are conceptually related to time distance and interaction slicing. The principle difference is that categories and collections of artifacts slice the data by artifact, while time distance and interaction slicing pulls out parts of the data based on characteristics of the interaction. The difference is where the "pivoting" takes place in data analysis.

Artifacts can exist in sets, although these sets are often not defined a priori or evident to the individuals participating in the STIP. One of the important qualitative aspects of the model is the open coding of categories of topics to identify individual participation across topic categories in a larger data set. This is often done through a process open and axial coded from grounded theory (Glaser & Strass, 1967) to identify influential individuals based on topical artifacts. These topics then form the basis of a collection of artifacts.

For example, if there are 25 discussion board threads that are open coded to be on the topic of "GIS Information Request," as was the case in our paper examining the coordination between Government Agencies and NGOs following the January 10, 2010 Haiti Earthquake (Goggins, Mascaro, & Mascaro, 2012), they would be considered as a category of artifacts. In these types of studies, the categories emerge from qualitative data analysis. In other studies we have done, we automatically group artifacts together based on information about team members working on multiple sets of the same artifacts at the same

DOI: 10.1002/asi

time. We performed this type of classification in software engineering (Blincoe et al., 2012), defining a measure of developer proximity around artifacts that relies on finding sets of code worked on during the same release cycle and in medical software support communities (Laffey, Galyen, Reid, & Goggins, 2012), where we have identified categories of customer organization through analysis of discussion board participation. In the medical software community support case, we showed how the type of organization influences interaction among members of a single organization with members of other organizations. This approach has also been applied to political discourse communities to identify the emergence of issue entrepreneurs and identify conversational leaders within a subset of topics (Mascaro & Goggins, 2011a; Mascaro et al., 2012).

People

Technologically mediated small groups are made of people. We point this out because modeling requires a complete enumeration of all the constructs it represents: Who the people are, what their identities are, and the roles they play are each factors in how an interaction takes place and how sociotechnical context is constructed and reconstructed. These individual traits influence interactions between people, as well. Given the substantial literature on psychology, sociology, and small groups, we note here that our work is distinguished by its focus on building a model of Group Informatics for technologically mediated small groups in a range of organizational and social contexts.

Although we focus on the interaction as the unit of analysis, personal characteristics influence behavior within STIPs. Therefore, we believe it prudent to understand attributes that may effect interactions by establishing a separate component entitled "people." This component was established after revisiting the dozen studies that led us to this conceptualization of Group Informatics. By revisiting the people component, we have come across two recurring, individual dimensions in the qualitative data that help us to build specific applications of the model regarding the component of people: individual identity (Bandura, 1977, 1997) and social identity (Tajfel, 1978, 1982; Turner, Brown, & Tajfel, 1979). We describe how these dimensions of the people whose groups are modeled using Group Informatics are factored into analysis.

Individual Identity

Participants in technologically mediated groups use them as an expression of their own identity. These expressions of individual identity are manifest in various classes of individual participation. People vary in the extent to which they derive individual identity from participation in a technologically mediated group, but also vary according to how they express identity and this shapes how they interact within the STIP.

Participants express identity explicitly and implicitly. For example, in our study of the U.S. Navy's disaster relief efforts following the January 12, 2010, earthquake in Haiti, one board participant identified herself explicitly: "I am the RFI manager." As a result of this explicit role identification, behavior was easily traceable and identifiable in analysis and participants had an explicit understanding of who was the leader in the forum. In the same forum, another participant implicitly described his role as a GIS specialist by providing information expected of a person in that role. The differences in the explanation of identity in technologically mediated interactions influence context and interactions and we attempt to discern these influences in our model.

When the individual does not provide explicit representations of identity, it is assessed using content analysis or grounded theory. This is especially the case in discourse-focused communities such as those affiliated with political groups, where both supporters and dissenters may participate for different purposes (Mascaro et al., 2012). Previous studies of online political discourse have found that some of the most active individuals in online political groups are those who do not agree with the stated purpose of the group (Mascaro & Goggins, 2012). The findings that result from the grounded theory and content analysis are used in conjunction with the other parts of the model to inform our understanding of the relationship between the evolution of context and the nature of interactions and how individual identity influences these constructions.

Social Identity

Social identity is focused at the small group unit of analysis in the Group Informatics model. Our main consideration here is to discern whether or not participants in a network are assigned to specific groups or roles a priori, or if such membership emerges through the interaction. Social identity emerges through some combination of these two factors. In each of our studies where a priori group assignment is made, most such groups become visible in our trace data as a result of patterned interactions or other explicit identifications. Therefore, in some cases knowledge of a priori social identity guides our research questions.

In other cases, small groups emerge and develop an identity as a small group through interaction or discourse. In some cases, social identity begins with individuals self-selecting into larger groups of others who are like themselves or construct a social identity in contrast to some "other." Both phenomena are explicated by Tajfel (1978, 1982), who describes the process of a person selecting into a group based on sharing some observable component of individual identity with other members of the group. In Tajfel (1982), he describes the way groups relate to each other. Together, selection into a group and differentiation between social groups is a dynamic process. We all construct our social identity with reference to people we are with, and people with whom we identify.

DOI: 10.1002/asi