Bringing context inside process research with digital trace data

Brian T. Pentland
Michigan State University
pentland@broad.msu.edu

Jan Recker
University of Cologne
jan.recker@wiso.uni-koeln.de

Julie Ryan Wolf
University of Rochester
Julie_Ryan@URMC.Rochester.edu

George Wyner
Boston College
george.wyner@bc.edu

Abstract. Context is usually conceptualized as "external" to a theory or model and treated as something to be controlled or eliminated in empirical research. We depart from this tradition and conceptualize context as permeating processual phenomena. This move is possible because digital trace data is now increasingly available, providing rich and fine-grained data about processes mediated or enabled by digital technologies. We introduce a novel method for including fine-grained contextual information from digital trace data within the description of process (e.g., who, what, when, where, why…). Adding contextual information can result in a very large number of fine grained categories of events, which is usually considered undesirable. However, we argue that larger numbers of categories can make process data more informative for theorizing. Including contextual detail enriches our understanding of processes as they unfold. We demonstrate this by analyzing electronic medical records (EMR) audit trail data using *ThreadNet*, an open source software application developed for the qualitative visualization and analysis of process data. The distinctive contribution of our approach is the novel way to contextualize events and action in process data. By providing new, usable ways to incorporate context, it can help researchers ask new questions about the dynamics of processual phenomena.

Keywords. Narrative networks; Qualitative methods; Process analysis; Routines; Processual phenomena; Digital trace data; Electronic medical records.

Acknowledgement. We are thankful to the Senior Editor, Patrick Finnegan, and two anonymous reviewers for helpful feedback on this paper. This material is based upon work supported by the National Science Foundation under Grant No. SES-1734237. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This research was also supported in part by University of Rochester CTSA (UL1 TR002001) from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. All faults remain ours.

1 INTRODUCTION

In this paper, we develop an approach to incorporate context in the analysis and visualization of digital trace data to theorize about processual phenomena. In empirical research, context is often seen as an external or situational threat to be controlled or eliminated (Avgerou, 2019). Researchers try to control for context in an effort to increase generalizability (Schofield, 2002; Whetten, 2009), identify causality, or improve robustness (Johns, 2006). This is unfortunate, because one of the deepest and most influential theoretical insights in information systems research has been that *context matters* (Avgerou, 2019; Burton-Jones & Volkoff, 2017; Hong, Chan, Thong, Chasalow, & Dhillon, 2014). From the web model of computing (Kling & Scacchi, 1982), to ensemble view of IT artifacts (Orlikowski & Iacono, 2001), and the sociomaterial view of technology in general (Orlikowski & Scott, 2008), we see that technology is mangled (Pickering, 2010), entangled (Orlikowski, 2007) and imbricated (Leonardi, 2011) with its social and material context. As Swanson (2019) argues, technologies come alive in the world to the extent that they inhabit recognizable, repetitive patterns of actions (Feldman & Pentland, 2003; Leonardi, 2011). This entanglement is processual, in the sense that it unfolds and emerges over time (Emirbayer, 1997; Tsoukas & Chia, 2002). For example, just as work cannot meaningfully be decoupled from technology anymore (Orlikowski & Scott, 2008), so too is everyday life increasingly permeated by digital technology (Lyytinen & Yoo, 2002; Yoo, 2010). The context of IS phenomena is no longer restricted to the organizational container (Winter, Berente, Howison, & Butler, 2014). Context needs to feature much more prominently in our research (Avgerou, 2019).

The question remains how to operationalize this insight in research practice. In research that embraces the importance of context, contextual entanglement has most often been described and analyzed through ethnographic fieldwork (e.g., Burton-Jones & Volkoff, 2017; Orlikowski, 2000). Because fieldworkers are there, in the situation, they are ideally positioned to see and describe situational effects in context (Charmaz, 2006). Fieldwork has been extremely fruitful for theory building, but when the entanglement stretches across time and space, and physical and digital worlds (Baskerville, Myers, & Yoo, 2020), it may be impractical.

To help remedy this limitation, the use of digital trace data has been proposed as a way to extend our reach (Berente, Seidel, & Safadi, 2019; Levina & Vaast, 2015). Digital trace data is evidence of activities and events that is logged and stored digitally (Freelon, 2014, p. 59). Since almost everything people do is now mediated by digital technologies (Yoo, 2010), digital trace data looms as an exciting prospect to qualitative scholars to theorize about the emergence and unfolding of processual phenomena. In response, calls have been made for qualitative scholars to lean on computational approaches (Lazer et al., 2009) involving automated data processing and algorithmic pattern recognition and analysis to help them discover patterns in the vast digital volumes of digital trace data that may otherwise be undiscoverable even to trained qualitative scholars (Lindberg, 2020).

However, most computational tools available to qualitative scholars interested in process are strikingly ignorant of context. For example, many kinds of tools exist for process mining and modeling but they incorporate context to a very limited extent, if at all. Van Der Aalst and Dustdar (2012) advocate for the importance of context, but they identify four levels of context (instance, process, social and external) that are all *outside* the execution of the process. They do not consider contextual factors *within* the execution of a process (e.g., who performs what step with what tool). Process mining (Breuker, Matzner, Delfmann, & Becker, 2016; van der Aalst, 2011b), usually reduces processes to a single dimension (a stream of time-stamped *actions*). Process models (Recker, Rosemann, Indulska, & Green, 2009) show *actions* and *actors* as they are designed, but they fail to incorporate the contextual circumstances under which the dynamics of such a process might change during enactment (Rosemann, Recker, & Flender, 2008). Likewise, computational tools that can handle processual data such as social network analysis (Wassermann & Faust, 1994) or sequence analysis (Abbott, 1995) reduce digital traces to variables such as actors (for social networks) or events (for sequence analysis). Paradoxically, the analysis of digital trace data for developing process theory is mostly devoid of the context that is so paramount to the unfolding and situatedness of the processes that scholars seek to explain.

In this paper, we describe a novel way of representing and visualizing contextual entanglements in processual phenomena that overcomes this limitation. We build on the concept of narrative networks (Pentland & Feldman, 2007), a special kind of directed graph where the nodes are categories of events and the edges represent sequential relations between those events (Pentland, Recker, & Wyner, 2017c). Narrative networks were introduced for the purpose of representing patterns of technology-in-use. They have been applied in field research and simulation (Pentland, Feldman, Becker, & Liu, 2012), but the emphasis has been on representing patterns of action. In these action-only models, as with other techniques for process mining (van der Aalst, Weijters, & Maruster, 2004) and modeling (Breuker et al., 2016; Recker et al., 2009), processes are disembodied and dissociated from their sociomaterial context.

Here, we advance the state of the art by incorporating context in a novel, systematic way. Instead of viewing context as the temporal, geographical, cultural, cognitive, emotional or any other sort of outside "environing" (Avgerou, 2019, p. 978), we locate context *inside* processes by using contextual factors available in digital trace data to define events in the narrative network. This is a key departure from established traditions that view context as something outside a process (e.g., weather, location) (Rosemann et al., 2008; van der Aalst & Dustdar, 2012). Instead of stripping these variables from digital trace data to make the data fit the format of process analysis tools, such as eXtensible Event Stream (Bala, Mendling, Schimak, & Queteschiner, 2018), we let context permeate everything: we use it to define the events that make up the process at it unfolds.

The advantage of bringing context inside the process is that contextual factors can be included at *any* level of granularity, so that context can change as fast as the process itself. Adding contextual factors in this way can result in a combinatoric explosion of fine-grained categories of events (who x what x when x where x how x …). In ethnographic research, the presence of the researcher helps manage the combinatoric explosion but the scope of research is limited to the here and now (Myers, 2009). The abundance of digital trace overcomes this limitation but at the cost of data explosion (Lindberg, 2020). The prevailing wisdom is that this proliferation is undesirable, but as we will demonstrate, it results in two transformative insights:

The first insight is that large numbers of *fine-grained categories can be useful* to theory development. This contradicts prevailing wisdom about the necessity of recoding data into more abstract 2nd or 3rd order categories for rigor, conceptual clarity and theoretical scaling (Gioia, Corley, & Hamilton, 2013; Urquhart, 2013). We discovered that a larger number of fine-grained categories of events in a narrative network resulted in a much clearer, more readily interpretable visualization of the process. By analogy, more pixels make a clearer picture.

The second insight results from the way that *absence highlights presence*. The relationship between absence/presence is a general principle in semiotics (Derrida, 1981; Eco, 1976; Rotman, 2016), but it is overlooked in conventional research methods where only presence is considered relevant. When we visualize hundreds of fine-grained categories as a network of sequentially related events, it is the *absence* of connected events (the white space) that makes the processual structure visible and interpretable. While context is sometimes seen as "muddying the waters" in conventional research (Avgerou, 2019), adding more contextual factors tends to disentangle and clarify process visualizations. To the extent that structural regularities are present in the context (e.g., division of labor), inclusion of more contextual factors will result in more white space (lower density), which will enhance the clarity and interpretability of the visualization.

To make our approach useful to IS research practice, we operationalize our insights with a software application called *ThreadNet. ThreadNet* is an open source R package that we have been developing with support from the National Science Foundation, as part of a larger research program (*Antecedents of Complexity in Healthcare Routines,* NSF SES-1734327). In this paper we introduce *ThreadNet.* We demonstrate how to use *ThreadNet* by visualizing the processes in a dermatology clinic at the University of Rochester Medical Center. *ThreadNet* is a flexible software for process data analysis that allows researchers to freely choose contextual information to be included in the definition of events that make up a process. It manages the combinatorics of context and makes it easy to see and compare how social/material contextual factors are entangled with processual phenomena. Without a convenient tool for visualization, the conceptual insights we describe here would never have emerged.

We begin by defining the essential theoretical concepts that provide a foundation for our work: process, events, context, and digital trace data. Then we describe how narrative networks allow us to define events through context, thereby making the critical conceptual move: bringing context inside the representation of process rather than leaving it on the outside. We demonstrate this approach by visualizing the electronic medical record (EMR) audit trail from a dermatology clinic. This example shows how the recordkeeping process is entangled with its social and material context. The example is typical of healthcare processes (Plsek & Wilson, 2001), and we show it as the complex socio-materially entangled mess that all expect it to be (van der Aalst, 2011a). We then demonstrate that, as we add contextual factors, *ThreadNet* disentangles the visualization in clearer, more comprehensible ways. We then compare our approach to other processual and qualitative data analysis tools, to clarify the conceptual and methodological novelty and transformative potential of our work. Finally, we discuss the implications, possibilities and limits of this approach for process scholarship in information systems and beyond.

2. Essential concepts

This paper builds on concepts and terminology from a diverse set of theoretical traditions, from process mining to structural linguistics. In this section, we present the bare essentials necessary to understand our analysis of the EMR recordkeeping process. After presenting the example, we compare the concepts presented here to related work in information systems and other fields.

2.1 Processual phenomena

The framework we present here puts processual phenomena in the foreground (Emirbayer, 1997; Langley and Tsoukas, 2016). By *processual phenomena*, we mean any progression of events that unfolds over time (Abbott, 2016; Tsoukas & Chia, 2002), such as routines, projects, workflows, or business processes. As these examples suggest, we intend to encompass a broad range of processual phenomena, independent of level of granularity (e.g., sequences of tasks, actions, processes, workflows, life stages), timing (e.g., by seconds, minutes, years) or extent of formalism (e.g., routine, business process, action pattern, algorithm). In what follows, we will use the collective label “process” to refer to any kind of processual phenomena.

2.2 Processes are sequences of events

Processes can be conceptualized as recognizable, repetitive sequences of events that unfold over time (Abbott, 2016; Pentland, Haerem, & Hillison, 2010; Tsoukas & Chia, 2002). Events are the abstract categories that formed from instantaneous observations or occurrences (Pentland & Liu, 2017). This translation from occurrences to events is the essential first step in theorizing about sequential data (Abbott, 1990): we observe occurrences, but we theorize about events. In particular, the sequential relation between events describes how a process unfolds over time (Pentland, 1999).

Of course, events with duration can overlap in time (e.g., in preparing spaghetti, one might cook the pasta while making the sauce). In the data we analyze here, we treat events as instantaneous. To capture duration and overlap, "cook the pasta" would be represented as a series of finer grained occurrences (put the pasta in the water, check doneness, drain the pasta…) that mark the start and stop times of various activities. Overlapping activities could also be modeled as part of the constantly changing context. Thus, it is always possible to regard events as sequentially related.

2.3 Events are defined by context

Within a process, events are defined by "what happens," but also by contextual specifics such as the time (now, later, ...), place (here, there, ...), subject (me, you, ...), and so on (Barnes & Law, 1976; Heritage, 2013). Speech acts (Austin, 1962) serve as an illustration for this idea, because their functional effect depends on the context of the utterance. For example, "I pronounce you husband and wife," has a different effect depending on who says it and who is present when it is said. In short, events are defined by context (the circumstances that form the setting for an event).[1]

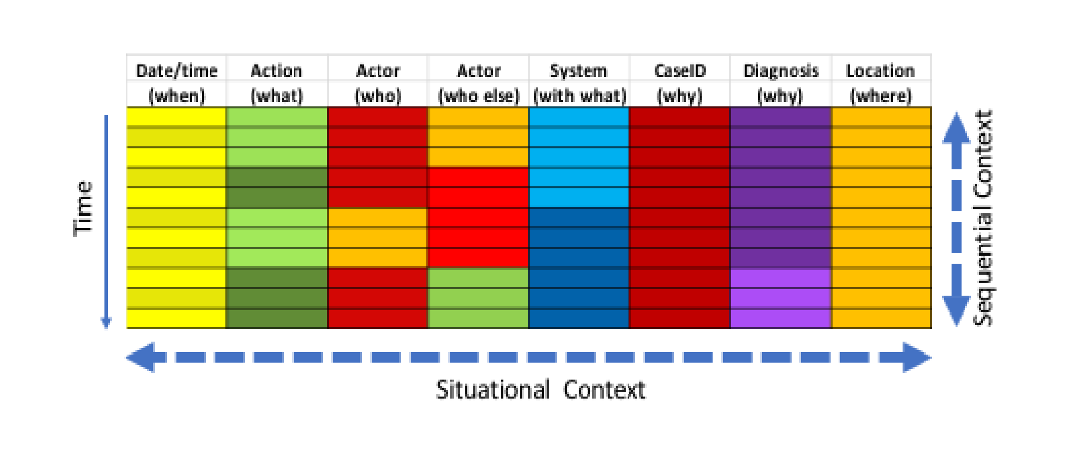
2.4 Situational and sequential context

Events typically occur as part of larger sequences that form a process, such as a business process, a workflow, or a routine (Bose & van der Aalst, 2009). To illustrate, Figure 1 shows a set of typical contextual factors changing over time. Each row represents an event, each column represents a dimension of context. Figure 1 echoes Van de Ven & Poole's (1990) classic framework for research on innovation processes: a set of factors that change over time.

Figure 1 differentiates context into two dimensions: situational and sequential context. In research paradigms derived from classical linguistics, situational and sequential context would be referred to as the paradigmatic and syntagmatic dimensions (de Saussure, 1974). *Situational* context refers to the situational particulars we might use to describe occurrences in any process, routine, or story: who, what, where, why… (Burke, 1962). *Sequential* context arises because, in practice, nothing happens in isolation; events are always located in an on-going sequence of other events (Bose & van der Aalst, 2009; Goh & Pentland, 2019). In traditional process terms, this is the timestamp that signals when an event occurred and which occurrences form part of that event.

Figure 1. Events in processes are defined by situational and sequential context

[1] This definition of context draws on the idea of frames as the definition of the situation as an excerpt of ongoing activity (Goffman, 1974, pp. 10-11): Depending on which circumstances are chosen, the definition of a situation (i.e., an event) changes.



There are several essential concepts here. First, as Figure 1 shows, context is crucial to the idea of process (Abbott, 2016; Pettigrew, 1997). Change implies a baseline or a point of reference from which we discern a difference. Without context, it is impossible to detect or even conceptualize change (Pettigrew, 2012). Second, context has multiple layers (Rosemann et al., 2008). These layers can change on different time scales, as shown in Table 1. Some contextual dimensions may change quickly while others may remain relatively stable, giving them different roles in understanding the process that unfolds. For example, it is natural to define events in terms of the aspects of context that change *fastest* (e.g., specific actions by specific actors). Contextual attributes that change *most slowly*, such as the clinic location, are likely to be external dimensions of context and typical dimensions of comparison (Avgerou, 2019; van der Aalst & Dustdar, 2012). We could compare routines between two or more clinics, for example. All of these layers of context co-exist in each occurrence, as shown in Figure 1, but their different type and role have typically not been incorporated into existing approaches to process analysis and theorizing.

Table 1: Temporal layers of analysis

Event	Thread	Processual phenomenon
Check-in with receptionist (seconds-minutes)	Visit to dermatology clinic (hours)	EMR Record keeping (on-going and constantly changing)

Another essential concept is that the definition of an event is not pre-determined; it depends on what contextual dimensions we choose to include. For example, we can define events in terms of contextual attributes that change *more slowly* than the data in an event log. For example, it would be perfectly natural to define each patient visit as a single event (based on the visit ID), although it might consist of hundreds of fine-grained events from check-in to check-out. As mentioned earlier, this provides a natural way to represent duration and overlap, if so desired.

Finally, we note that contextual dimensions may be correlated or aligned (Kim et al., 2019) in varying degrees. For example, in an idealized world, each actor might perform one task with the same tool in a single location. If so, those contextual dimensions would perfectly be aligned; using an additional dimension to describe the action would not add information.

To summarize, while the importance of context has long been recognized in process mining and modeling (Bose & van der Aalst, 2009; Rosemann et al., 2008), it has been conceptualized and operationalized as something that exists *outside* of processes (Rosemann et al., 2008). Moreover, context has usually been seen as static and a-temporal (Pettigrew, 2012). Our conceptual move is to put context *inside* the definition of process, allowing it to be as dynamic and performative as the process itself. Putting context inside the process mixes the “in-here” and the “out-there” (Hernes, 2007, p. 2). This move sets the stage for a novel approach to conceptualizing and analyzing dynamic, processual phenomena.

2.3 Narrative networks: A framework to incorporate context explicitly

Narrative networks provide a way to incorporate situational and sequential context into the definition of events that are constitutive of processual phenomena. The narrative network was introduced as a method for describing technology-in-use within the repetitive, recognizable patterns of events that characterize organizational routines (Pentland & Feldman, 2007). Formally, a narrative network is a weighted, directed graph where the nodes represent categories of events and the edges represent sequential relationships between those categories (Pentland et al., 2017c).

It is important to be clear about what this class of network does and does not represent. First, the nodes in a narrative network represent categories of events. For example, in a medical clinic, a typical event would be: "The nurse takes your blood pressure." Traditional process models represent the descriptive or constative nature of processes (Recker et al., 2009; van der Aalst, 1998), such as, for example, the declaration “take blood pressure”, then “record blood pressure.” In contrast, narrative networks represent performative trajectories (Hernes, 2017), i.e., accounts of what happens, what is being done. As Hernes (2017, p. 604) argues,

> *Events … are not to be seen as representative of a trajectory, but as performing the trajectory. Every event takes active part in performing the temporal trajectory, by defining the present events in the context of its predecessors and antecedent events.*

This view entails the assumption that the social world is a continually unfolding process and thus the "dynamic, unfolding process becomes the primary unit of analysis rather than the constituent elements themselves" (Emirbayer, 1997, p. 287). So, while each constituent event is performative, when they are sequentially related in a set of threads (or paths, Goh & Pentland, 2019, or in Hernes' (2017, p. 604) terms, trajectories), and incorporated into a network of events, the overall performative effect unfolds.

Second, a narrative network represents sequential relationships between events, defined as actions, activities, or processes (depending on vocabulary and granularity used to define and label the nodes). These networks do not correspond to social networks (the nodes are events, not people), flowcharts or petri nets (they do not model state changes), or Markov models (the nodes do not represent system states). Unlike these more familiar classes of networks, the nodes represent categories of events in a domain. The edges indicate temporally sequential adjacency of those events along a set of observed threads (e.g., "college first, then graduate school"), but they do not necessarily indicate causality. Past events influence future events but they do not determine them (Goh & Pentland, 2019).

Situational context enters via definition of the nodes of the network. Our conceptual move is that we broaden how we define the events in narrative networks. Published examples of narrative networks have included nodes defined by *actions*, such as those in an invoice processing system (Pentland et al., 2010). Occasionally, the events have been defined as both *actions* and *actors*. Goh et al. (2011), for example, use narrative networks to identify where and how the introduction of health information technology changes sequences of actions performed by actors. Yeow and Faraj (2011) use narrative networks to investigate changes to actors and actions resulting from an ERP implementation.

These examples show that there is merit to representing processual phenomena as events in a narrative network, but we also see that current applications tend to limit their inclusion of context to either actions, people or technology, but neither all three simultaneously nor additional context such as location, reason, date and/or time. Event definitions in the literature to date have generally been limited to "action" or "action-actor." We have previously demonstrated how action- or actor-only network graphs skew our view of what is going on, e.g., what constitutes a handoff (Pentland et al., 2017c). We now demonstrate below how a richer, more contextualized definition of events changes the narrative network and thus changes how the processual phenomena are represented – and what we might learn about them.

Sequential context enters via the edges of the network. The idea of tracing associations between actions is based on the idea that actions do not happen in isolation – they occur as part of streams of activity, thus forming an action-centric view of the world (Pentland, Pentland, & Calantone, 2017a). Musicians rarely just play one note; they play tunes. The sequential relationship is determined empirically by tracing the sequence of actions within a thread. These networks can be automatically constructed from “traces” (Bala et al., 2018; De Weerdt, vanden Broucke, Vanthienen, & Baesens, 2013) in computerized event logs, middleware or other forms of digital trace data. Using digital trace data, i.e., digitally recorded and time-stamped logs of sequential events, thus opens new possibilities to expand our view both conceptually and empirically.

2.4 Contextualizing digital trace data

Methodologically, the value of our approach (defining events by combining relevant contextual factors) rests on the assumption that data about different layers and changes of context are available. Ethnographic fieldwork allows researchers to access context through immersion or embeddedness (Feldman, 1995; Lewis & Russell, 2011). However, fieldwork is ill-equipped to handle the large volume of data traces now collected and stored on digital platforms (Floridi, 2012). Just as processes of work and organizing cannot meaningfully be decoupled from technology anymore (Orlikowski, 2007; Orlikowski & Scott, 2008), all aspects of our life are increasingly mediated by digital technology (Alaimo & Kallinikos, 2017; Yoo, 2010).

Digital trace data is inherently processual in nature. As the name suggests, it “traces”, i.e., connects actions and events enabled or mediated by digital technologies as they unfold over time: it captures the sequence of events that constitutes a process because it includes time-stamped logs of activities and events enacted through digital technologies or platforms. This allows more precise and more voluminous data on actions and events than traditional modes of collection such as observations, interviews or archival data (Schensul, Schensul, & LeCompte, 1999).

Digital trace data provides opportunities for qualitative scholars (Sundararajan, Provost, Oestreicher-Singer, & Aral, 2013) but also requires serious methodological adjustment to the particulars of this new type of data (George, Osinga, Lavie, & Scott, 2016). For example, when confronted with digital trace data, qualitative scholars tend to become overwhelmed by the sheer size of these datasets (Lindberg, 2020). Moreover, digital trace data is organic, not designed, so it is inherently susceptible to validity issues (Xu, Zhang, & Zhou, 2019). Also, digital trace data can be both heterogeneous and unstructured (Dhar, 2013), making it difficult to analyze and confront the meaning of digital trace data as a conceptualization of the events and mechanisms it records (Levina & Vaast, 2015).

In response, computational social science has been advocated as a methodological advance (Chang, Kauffman, & Kwon, 2014; Lazer et al., 2009) but it comes with a number of challenges. These include the difficulty of analyzing complex and messy social phenomena, and the tendency of researchers to oversimplify (naturalize) these complex relationships, thereby curtailing the search for meaning (Törnberg & Törnberg, 2018). Scholars have thus been advised to complement and blend computational analyses of digital trace data with deep qualitative inquiry to account for, and understand, the context(s) in which that data is generated (Lindberg, 2020; Whelan, Teigland, Vaast, & Butler, 2016). In essence, scholars are advised to add context to digital trace data “from the outside” through manual data collection or analysis. We propose an alternative: rather than add context to the computational analysis of digital trace, we suggest incorporating it inside.

3 Visualizing Record Keeping Routines at a Dermatology Clinic using *ThreadNet*

3.1 Brief Introduction to Threadnet

Because we aim to contribute to research practice, we want to demonstrate the usefulness of bringing context inside process data for theorizing about the process. To that end, we introduce *ThreadNet*, a software tool for the analysis and interpretation of processes in context based on the idea of a narrative network (Pentland et al., 2017c). We developed *ThreadNet* iteratively and presented a variety of prototypes to the community over the years (Pentland, Recker, & Kim, 2017b; Pentland, Recker, & Wyner, 2015, 2016). The original version of *ThreadNet* was developed in MatLab. With support from the National Science Foundation (NSF SES-1734237), we now made *ThreadNet* available as an open source R package on GitHub (http://www.github.com/ThreadNet), together with source code, instructions, documentation and sample data (http://routines.broad.msu.edu/ThreadNet/).

Much like other computational approaches (e.g., Gaskin, Berente, Lyytinen, & Yoo, 2014; Indulska, Hovorka, & Recker, 2012; Larsen & Monarchi, 2004), *ThreadNet* uses sequential, categorical data to allow analyses and visualizations of processual phenomena. Metaphorically, it weaves threads into fabric. The key feature relevant to this paper is that *ThreadNet* makes it particularly convenient to choose contextual dimensions to define events and visualize the resulting network.

ThreadNet was developed to help make computational tools for the analysis of digital trace data accessible by qualitative researchers. This resulted in some straightforward design criteria. First, it should have a graphical interface that can be used without any coding or programming. We wanted to remove barriers to use and minimize the learning curve. Thus, we used Shiny R to create the user interface. Second, the emphasis should be on visualization, not statistics. *ThreadNet* contains a variety of simple visualizations for narrative networks. Third, it should not duplicate functionality from other network or sequence analysis packages (e.g., UCINet or TraMineR). Rather, ThreadNet provides the capability to export narrative networks for use in other software.

3.2 Research setting: Record Keeping at a Dermatology Clinic

The data we use here stems from a larger research project investigating the antecedents of complexity in healthcare routines (NSF SES-1734237). The data were collected from the dermatology clinics at the University of Rochester Medical Center. Superficially, dermatology clinics would seem to be one of the simplest possible clinical settings. In interviews, clinical staff describe the workflow as a fairly uniform series of steps: (1) check-in; (2) "rooming" (taking the patient to an examination room); (3) taking vital signs and history; (4) examining the patient; (5) administering treatment and/or writing prescriptions; and (6) check-out. However, the data we extracted from the electronic medical record (EMR) system indicates that the process contains a lot of variation and complexity. This setting provides a revelatory and representative exemplar of how processes can be analyzed on the basis of digital trace data – in our case EMR record keeping logs. EMR are a notorious source of digital trace data (Kunzman, 2018; Lee et al., 2017): the traces EMR provides are both rich and noisy, in turn underscoring how being able to recognize everything "in context" is critical to understanding what is going on.

3.2 The Digital Trace Data

The EMR audit trail is useful for this paper because it contains contextual dimensions that are typically not present in most digitized event logs available in standardized formats (van der Aalst, 2016). For example, in addition to the action and the timestamp (what and when), our EMR data contains the role of actor (who) and the workstation they used (where). We interpret the workstations as indicating location, rather than technology, because the EMR user interface is the same for each user at all workstations. Therefore, each individual is always using the "same system," but they are using it in different locations.

We focus on the record keeping process at one clinic on one day in February, 2015. The dataset (available at http://routines.broad.msu.edu/ThreadNet/OneDayOneClinic.csv) includes 24 visits from that day. The data was completely de-identified, with no identifying information about patients or providers. Table 2 shows the first five minutes of one visit to the clinic, as captured in the EMR audit trail. Each row corresponds to an event. Events are described by a timestamp, a visit ID, Workstation_ID, Action_code, Role and Clinic.

When patients arrive at the clinic, they check in with a receptionist whose formal role in the system is “Admin_Tech.” The computer workstation at the reception desk is labeled “BCAHHURDRM”. To complete the check-in, the Admin_Tech visits several screens in the EMR system (e.g., “MR_SNAPSHOT”). After the patient is checked in, a Licensed Nurse obtains the patient history, and enters vital signs and the chief complaint. Every patient visit begins with checking in at reception and ends with printing a visit summary at checkout. But throughout each visit, the situational and sequential context is constantly changing.

Table 2: The first five minutes of one patient visit

tStamp	VISIT	WORKSTN_ID	ACTION_CODE	ROLE	CLINIC
2/2/15 8:53	1	BCAHHURDRM	CHECKIN TIME	Admin Tech	A
2/2/15 8:53	1	BCAHHURDRM	MR_SNAPSHOT	Admin Tech	A
2/2/15 8:53	1	BCAHHURDRM	MR_REPORTS	Admin Tech	A
2/2/15 8:53	1	BCAHHURDRM	MR_SNAPSHOT	Admin Tech	A
2/2/15 8:53	1	BCAHHURDRM	MR_REPORTS	Admin Tech	A
2/2/15 8:55	1	BCAHHURDRM	MR_SNAPSHOT	Admin Tech	A
2/2/15 8:55	1	BCAHHURDRM	MR_REPORTS	Admin Tech	A
2/2/15 8:56	1	BCAHHURDRM	MR_SNAPSHOT	Admin Tech	A
2/2/15 8:56	1	BCAHHURDRM	MR_REPORTS	Admin Tech	A
2/2/15 8:56	1	URDERMDT3	AC_VISIT_NAVIGATOR	Licensed Nurse	A
2/2/15 8:56	1	URDERMDT3	MR_HISTORIES	Licensed Nurse	A
2/2/15 8:56	1	URDERMDT3	MR_ENC_ENCOUNTER	Licensed Nurse	A
2/2/15 8:56	1	URDERMDT3	MR_VN_VITALS	Licensed Nurse	A
2/2/15 8:56	1	URDERMDT3	MR_REPORTS	Licensed Nurse	A
2/2/15 8:56	1	URDERMDT3	FLOWSHEET	Licensed Nurse	A
2/2/15 8:56	1	URDERMDT3	MR_VN_CHIEF_COMPLAINT	Licensed Nurse	A

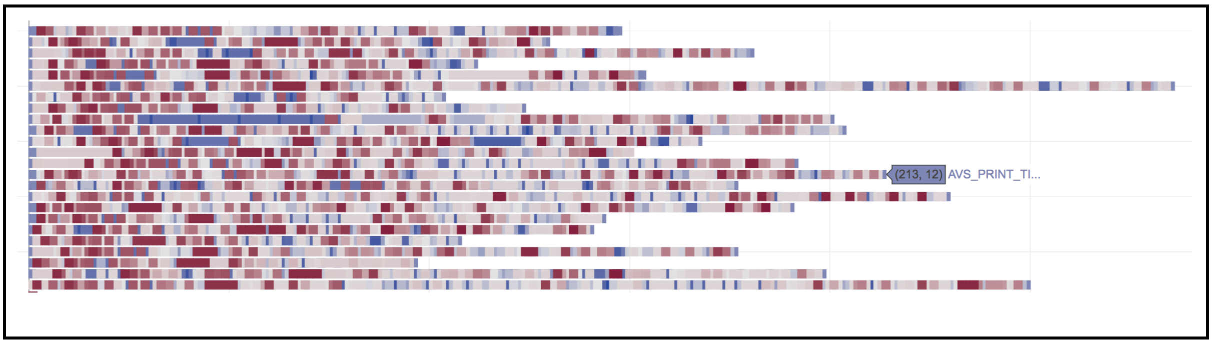
2/2/15 8:56	1	URDERMDT3	MR_REPORTS	Licensed Nurse	A
2/2/15 8:56	1	URDERMDT3	MR_SNAPSHOT	Licensed Nurse	A
2/2/15 8:56	1	URDERMDT3	MR_REPORTS	Licensed Nurse	A
2/2/15 8:57	1	BCAHHURDRM	MR_REPORTS	Admin Tech	A
2/2/15 8:57	1	BCAHHURDRM	MR_SNAPSHOT	Admin Tech	A
2/2/15 8:58	1	URDERMXRM1	MR_REPORTS	Licensed Nurse	A
2/2/15 8:58	1	URDERMXRM1	AC_VISIT_NAVIGATOR	Licensed Nurse	A
2/2/15 8:58	1	URDERMXRM1	MR_ENC_ENCOUNTER	Licensed Nurse	A
2/2/15 8:58	1	URDERMXRM1	MR_HISTORIES	Licensed Nurse	A
2/2/15 8:58	1	URDERMXRM1	MR_REPORTS	Licensed Nurse	A
2/2/15 8:58	1	URDERMXRM1	MR_VN_VITALS	Licensed Nurse	A
2/2/15 8:58	1	URDERMXRM1	FLOWSHEET	Licensed Nurse	A
2/2/15 8:58	1	URDERMDT4	MR_REPORTS	Physician	A
2/2/15 8:58	1	URDERMXRM1	MR_VN_VITALS	Licensed Nurse	A
2/2/15 8:58	1	URDERMXRM1	MR_HISTORIES	Licensed Nurse	A
2/2/15 8:58	1	URDERMXRM1	MR_HISTORIES	Licensed Nurse	A
...	...	...	...	...	...

Note how the structure of the data in Table 2 resembles the conceptual layout of Figure 1. The rows are events, and the columns contain a set of contextual dimensions. Some dimensions change quickly (e.g., ACTION_CODE), others change slowly (e.g., ROLE), some remain constant for a majority of events (e.g., CLINIC). An important conceptual move is to treat all of these dimensions on an equal footing. Rather than privileging the role of the *actor* (as in much of traditional organizational scholarship), or the label on the *action* (as in typical process analysis and process mining), they are both just aspects of context that we can use to define categories of events.

We have shaded sections of Table 2 to show a typical pattern in the record keeping work. In each set of shaded rows, a particular actor (e.g., Licensed Nurse) stands at a particular workstation (e.g., URDERMDT3) to perform a series of actions. They may perform several actions or just one.

We can visualize the whole set of 24 patient visits as a set of threads, as shown in Figure 2. Each dot represents one event (one row from the Table 2) and each row represents the sequence of events in patient visit in event time (Poole, Lambert, Murase, Raquel, & Joseph, 2017). The shading of each dot indicates the corresponding event. Visualizing the threads as straight lines makes it easy to see that they vary in length and in sequence. No two threads look alike. However, it does not show how contextual factors shape the overall pattern of action. To see that, we use *ThreadNet* to visualize how the events within each thread are related.

Figure 2: Twenty-five visits to the dermatology clinic



3.3 The *ThreadNet* algorithm

Conceptually*, ThreadNet* constructs narrative networks by making two passes through the data. In the first pass, it identifies the nodes. In the second pass, it traces and counts the edges between the nodes.

1. **Identify the nodes**. To define the nodes that will be included in the network, *ThreadNet* combines a set of contextual dimensions selected by the user. Nodes are labeled by combining the values in the columns of the data.[2] Only the unique combinations that occur in the data appear in the network.

1. **Trace the edges**. *ThreadNet* follows each thread from one event to the next. Whenever the sequentially adjacent events within a thread are different, it adds an edge between that pair of events, from one event to the next.[3] The strength of the tie between those events can be based on the frequency of each pair of events (how often that transition occurs in the data). The resulting network is a valued, directed graph that is unimodal (one kind of node) and unidimensional (one kind of edge).

3.4 Incorporating context make the network easier to interpret

To demonstrate how context changes our analysis and understanding of process, consider how

---

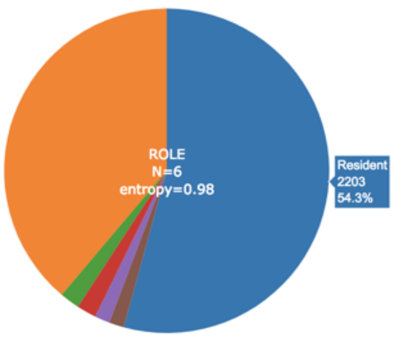
2 The combination of contextual factors is implemented in R using the **unite** function from the **tidyr** package. This function combines columns in a data frame to create a new column from the values in the original columns.

3 This functionality is implemented in R using the **ngram** package to count the 2-grams within the observed threads. This provides an edge list that can be used to construct the network, as well as the weight of each edge.

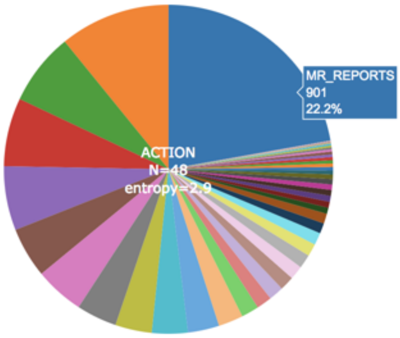
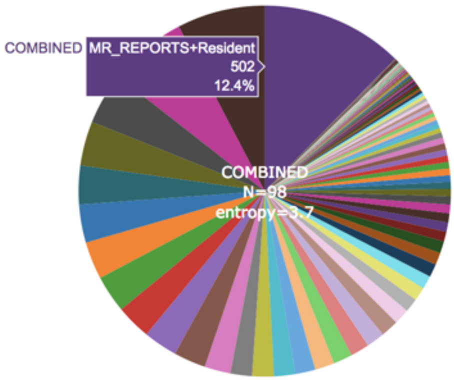
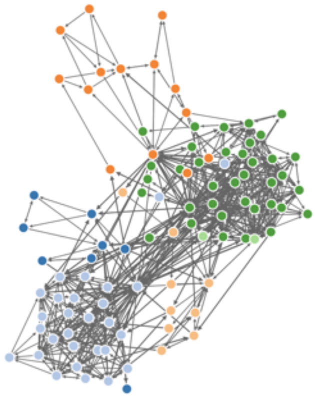
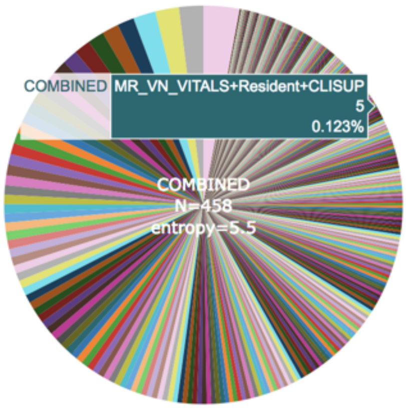
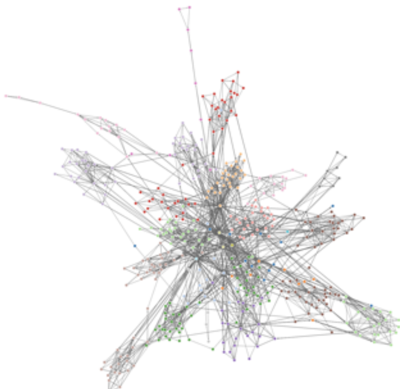
incorporating context in different ways changes the apparent structure of the narrative network that describes the record keeping process. To illustrate, Figure 3 shows how we used *Threadnet* to display the same 25 visits with the nodes defined in four different ways: (1) role only, (2) action only, (3) action + role, (4) action + role + workstation. Figure 3 also displays some quantitative information about each row (the entropy of the data and the density of the graph) that helps explain the visualizations.

In the left-hand column of Figure 3, we see the frequency of categories within each contextual dimension (or combination of dimensions) rendered as pie charts. For example, in the first row (Role Only), we see that there are six roles, and the Resident was involved in 54.3% of the record keeping that day. By hovering over the pie chart with the mouse in *Threadnet*, we would see that the Technician was the second most active, with 38.8%. In the right-hand column of Figure 3, we see the narrative network with the nodes defined by the contextual dimension shown in the left-hand column.[4] It is worth considering the differences between these four visualizations in some detail.

Figure 3: Contextualizing an event in four different ways changes the apparent structure of the process

Contextual dimensions of event definition	**Frequency of occurrences (n=4060)**	**Narrative Network**
Role Only	 N=6, Entropy = 0.98	 Density=0.78

[4] Note that colors are assigned independently in the right-hand v. left-hand column.

Action Only	 N=48, Entropy = 2.9	 Density = 0.212
Action + Role	 N=98, Entropy = 3.7	 Density = 0.083
Action + Role + Workstation	 N = 458, Entropy =5.5	 Density = 0.0098

The first row shows the relation between the six roles in the clinic (Resident, Technician, Admin_tech, Physician, Registered_Nurse, and Staff). As we noted above, the network graph is an event network, not a social network, but it does provide an actor-centric perspective on the hand-offs between the roles in the clinic (Pentland et al., 2017c). The graph is extremely dense (density = 0.78), which obscures any underlying processes (Pentland et al., 2017a). Each role in the clinic handed off recordkeeping to nearly every other role at least once during the day. We can see that all the roles are involved, but we cannot see what they are doing.

In the second row, we see same data with nodes defined by the Action (n=48). This would be the typical way to define nodes in process mining and discovery (van der Aalst, 2011b). Due to the highly variable nature of the work, this graph looks like the classic "hairball" (Dianati, 2016). In principle, process mining tools could be used to refine and filter this representation (usually by frequency of occurrence) to get a "comprehensible" model, which is what most applications of process mining try to achieve (Breuker et al., 2016; van der Aalst, 2011a). Rather than attempting to simplify or reduce the data to reveal an idealized model, we embrace the complexity that is present in the data.

To illustrate, instead of recoding the data into more abstract categories as proposed in traditional inductive qualitative analysis (Gioia et al., 2013; Urquhart, Lehmann, & Myers, 2010), we add more situational context. In the third row, we combine the actions and the roles to define the nodes in the network. This combination results in 98 unique action-role values, each of which becomes a node in the graph. When we construct the graph this way, it begins to reveal regions of activity associated with each clinical role. The two large clusters correspond to the Resident role and the Technician role. We are adding context in the events by including the role that takes each action. Adding context starts to make the graph less dense (density = 0.083). The increase in "white space" (absence) allows us to begin to see structure in the process. As work is carried out, some roles frequently have handoffs with others, while other roles carry out their work without frequent interactions with others.

For example, the group of actions carried out by the Physician (sparse set of orange nodes in the upper left) is mainly connected to the actions carried out by the Resident (the relatively dense set of green nodes in the upper right). The group of actions carried out by the Technician is also relatively dense, but separated from the Physician. We can see that Physicians have frequent handoffs with Residents, but less frequently with the Technicians. For all of these roles, the most frequent handoff occurs via "MR_REPORTS", which is a kind of landing screen in the EPIC user interface (as reflected by its frequent occurrence in Table 1).

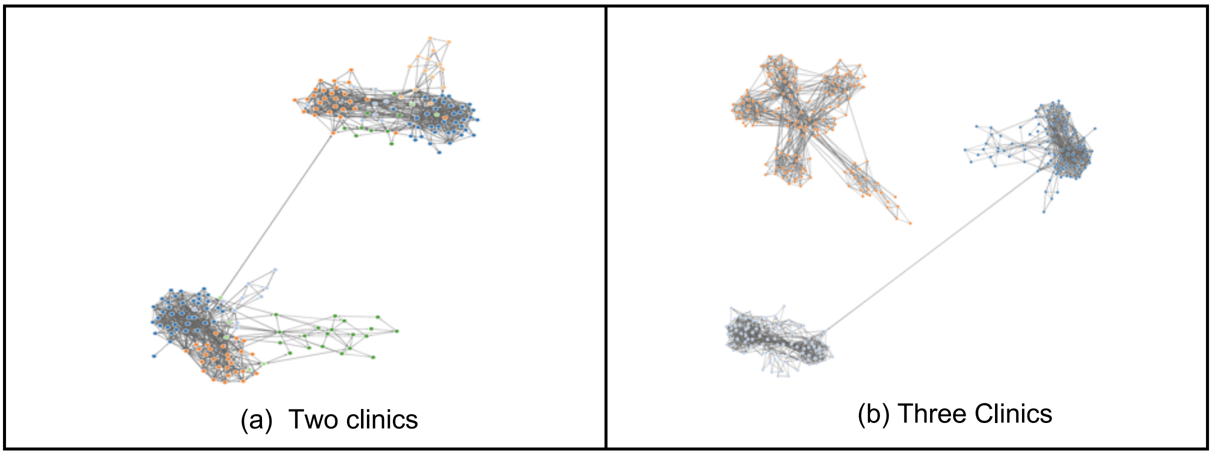
In the fourth row, we add more situational context by adding "workstation" into the definition of the nodes. This combination results in 458 unique values for action-role-workstation events. When we construct the graph with these 458 nodes, the sociomaterial structure of the work process becomes more clearly visible. The clusters in the graph correspond to activity at the workstations around the clinic. The influence of the material technology (specific workstations) is clearly visible in the patterns of action because the clusters in the network correspond to workstations. The visibility increases because adding context adds both presence (the clusters of action-role-workstation that ***do*** occur) and absence (combinations of action-role-workstation that ***never*** occur). By adding context we add both information and white space to the graph, which helps us see what Hernes (2007, p. 1) called the "tangled world."

3.5 Entanglement and Disentanglement through context

Based on this example, we can begin to generalize about the ways that contextual dimensions can influence the visualization. Not all contextual dimensions are equally informative. If a dimension never changes, or if it co-varies with other dimensions, it does not add information. Conceptually, adding more dimensions of context is only valuable if they add information (Kim et al., 2019). Context only matters when it changes. For example, if we added a dimension for *Continent*, it would have the same value for all visits to all clinics at the University Rochester Medical Center.

By contrast, some contextual dimensions may have multiple values, but the observed threads never cross between them. In other words, those dimensions are not visibly entangled by the process. *Threadnet* can help us see such a situation in the dermatology audit trail data quite easily if we expand our view of the process to include *clinic* as a contextual dimension. To illustrate, consider nodes in the graph defined with three dimensions: *action-role-clinic.* So, when a nurse takes an action in one clinic, we treat that as a different event than a nurse taking the same action in a different clinic. The University of Rochester Medical Center operates multiple dermatology clinics, each of which is at a distinct physical site. Visits to one clinic do not usually extend to the other clinics. Figure 4(a) shows one full day at two clinics (51 visits) and Figure 4(b) includes 3 clinics (76 visits). Comparing both network graphs it becomes immediately apparent where context was entangled with the process and where it was not: Two of the clinics had a single interaction that day, the third clinic was completely disconnected. Hence, when we include the situational context *clinic* in the definition of the nodes, it disentangles the graph. This is because patient visits tend to be localized to particular clinical sites; the threads do not cross these contextual boundaries very often. When the analysis shows that they do, it allows "constructing mystery" in theorizing (Alvesson & Kärreman, 2007): the graph clearly shows the path, it allows us to see what happened during these exceptions. Spotting such deviant cases through traditional qualitative fieldwork involves complex methodological and analytical work (Mertens et al., 2016). With our approach, it can be computed.

Figure 4: Context disentangles the graph



In a tangled world, however, threads cross organizational boundaries (Avgerou, 2019; Winter et al., 2014). For example, in-patient medical care may involve multiple clinical specializations, specialized labs, specialized treatment facilities, and so on. These seemingly distinct units become entangled in practice. When this happens, our intuition about the resulting processual phenomena is weak, at best. Digital data sources that trace such processes typically span multiple systems and technologies, making it even more difficult to identify and reason about the events that that unfold. But aside from technological difficulties in constructing the digital trace data (Bala et al., 2018) we also need to explore different conceptualizations of "what happened" to theorize about the process. Our conceptualization of context within an event allows for this conceptual latitude (Burton-Jones, McLean, & Monod, 2015) because it allows constructing views on the process constituted by different conceptualizations of "context" to find structure and meaning in the graph that informs theorizing about what is going on.

4 DISCUSSION AND IMPLICATIONS

Context is clearly important to our understanding and analysis of processes, but it has been difficult to put this idea into practice. Context is usually conceptualized and operationalized as something that exists *outside* of processes (Rosemann et al., 2008); moreover, context is usually seen as static and a-temporal (Pettigrew, 2012). When we conceptualize context as something "out there", in the background, then we might investigate the role of context in process by asking if the workflow is different in one clinic vs. another (Avgerou, 2019; van der Aalst & Dustdar, 2012) or if it changes by season (Rosemann et al., 2008). But when we conceptualize context as constitutive of the events within a process, as we have done here, it takes us on an entirely different scholarly journey. The visualizations in Figures 3 and 4 embody a novel way to incorporate context in the description of processual phenomena. Putting context inside the conceptualization mixes the “in-here” and the “out-there” (Hernes, 2007, p. 2). Adding contextual dimensions into the definition of events in a process changes how a process looks both through presence and absence of new information. *ThreadNet* provides a convenient way for researchers to incorporate and visualize the influence of context on the structure of a process. It allows researchers to bring deep qualitative inquiry of the context of digital trace data (Lindberg, 2020; Whelan et al., 2016) right into the computational analysis rather than adding it as a complement. In the following sections, we discuss the implications of this innovation.

4.1 Comparison to other approaches to processual analysis

We start by explaining how our approach and implementation in *ThreadNet* is different to typical ways for analyzing process data. We sampled a range of leading analysis tools potentially suitable for process scholarship that are based on different underlying conceptual frameworks (specifically, NVivo, ATLAS.ti, TraMineR, ProM and BupaR). Our comparison is by no means comprehensive. For example, we have excluded an enormous set of other sequence analysis tools, many of which are specific to bioinformatics or content analysis, but can be adapted to organizational research (see, e.g., Gaskin et al., 2014). Likewise, we excluded several approaches to social network analysis (Borgatti, Everett, & Johnson, 2013) even though they too can at least to some extent be used to implement similar ideas (e.g., events as nodes in semantic networks). Instead, we sampled the two most prominent types of tools used for process scholarship – those traditionally used in qualitative data analysis (Flick, 2018, pp. 519-536) such as NVivo or ATLAS.ti and those typically used in digital trace data analysis (Gabadinho, Ritschard, Müller, & Studer, 2011; van der Aalst, 2016). We inspected the capacity of both the conceptual frameworks and the concrete features of the chosen tools to allow for our conceptualization of processes as sequences of contextualized events in a narrative network. Table 3 summarizes our insights.

Table 3: Alternative frameworks for processual analysis

Conceptual framework	**Open coding**	**State sequences**	**Process mining**	***Narrative networks***
Example software	NVivo ATLAS.ti	TraMineR	ProM BupaR	*ThreadNet*
Input data structure	Text, fieldnotes, images, etc.	One sequence per row, one state per column	.XES (xml document with specialized structure for timestamped events)	*.CSV with one time-stamped event per row and contextual attributes in columns*
How is context represented?	In text and diagrams	No incorporation of context: states are defined by single-value codes	Limited incorporation of context: actions may be associated with a resource or other attributes such as location	*Any number of contextual dimensions can be included as coded categories*
How are events and states represented?	In text and diagrams	Events are implicit in sequence of states	Events and states explicitly modeled in Petri Net	*Events are nodes. States are implicit in sequence of events*

Open coding. Open coding is the ultimate in flexibility, but it is also labor intensive and thus not well suited to handling digital trace data (Indulska et al., 2012). Most qualitative analysis is based on coding of text or some other kind of document. Tools like NVivo, for example, allow the creation of "nodes" and "node hierarchies." These can be used to code contextual categories, and could be used to code sequential categories, as well. However, working directly with text, even with a tool like NVivo, it would be difficult to keep track of hundreds of categories and their sequential relationships in a large corpus of data. Also, since ethnographic field notes and interviews are often focused on the "ethnographic present" (Sanjek, 1991), time or sequence are often not present in the original data sources. In contrast, *ThreadNet* traces all of the sequential relations between every unique combination of contextual categories.

State-sequence analysis. Career progressions provide the archetypal example of a state sequence analysis in social science research (Abbott & Hrycak, 1990). The state-sequence framework is instantiated in *TraMineR* (Gabadinho et al., 2011), along with a broad array of sophisticated sequence analysis tools. *TraMineR* has mainly been applied in sociological research on life course progression, although the methods are applicable to a broad range of problems (Poole et al., 2017). States are defined by a single attribute (e.g., married or unmarried). Each row is a case (e.g., a career) and each state occupies one column, so here is no way to show multiple states (attributes) at the same time. In contrast, *ThreadNet* can handle events defined by any combination of attributes and it allows users to change the combination at will.

Process mining. We use the term "process mining" to refer to a large, diverse, and evolving set of tools and ideas (van der Aalst, 2016). What we focus on here corresponds largely to what is considered "classical" process mining. The ProM framework (van Dongen, Alves de Medeiros, Verbeek, Weijters, & van der Aalst, 2005) is a platform where researchers can publish new tools and techniques in the form of software plug-ins. Outside of ProM, BupaR (Janssenswillen & Depaire, 2017) is an R package that implements basic process mining capabilities.

In many, but not all, applications of process mining and modeling, the goal is to find a clean model that provides a reference for the execution of a process. Most process mining algorithms assume that the underlying process is stable such that *discovery* of the stable process and *conformance checking* are the primary applications (van der Aalst, 2005, 2011b). To that end, a typical goal of visualization has been to simplify overly cluttered graphs into *comprehensible* models (Breuker et al., 2016; van der Aalst, 2011a). In contrast, our explicit goal is to reveal the extent of the mess and to display processes as they unfold in event time.

Our approach. In contrast to these other frameworks for processual analysis, the narrative network, on which *ThreadNet* is based, provides a way to embrace and internalize context. It can handle a fluid notion of events constituted by changing context. This is important because even in a repetitive process or routine, there is no a priori reason to expect that events will repeat in an exact pattern. Especially on longer time scales, it is reasonable to expect on-going change (Pentland et al., 2012).

As we show in our illustration of the EMR record keeping trace data, our approach is capable of revealing structure and paths of the clinical routine that would be invisible using any of the alternative frameworks in Table 3: open coding would potentially lead to a rich description of certain observed occurrences or the surfacing of particular salient patterns or classifications, but it does not provide an effective method for looking at the patterns that emerge from 458 unique combinations of role, actor, and workstation. Process mining and state-sequence analysis do not provide complex representations that, as we show above, can help disentangle and explain what is really going on. They typically either abstract away such complexity by design or only reveal the "hairball" that is the work on its surface.

4.2 Implications

Using contextual information potentially available in digital trace data to relax our assumptions of what constitutes the context of events in processes has interesting implications for process scholarship.

Abstraction and richness of processual data. The conventional wisdom is that large numbers of first-order categories need to be reduced into a smaller number of higher order categories (Berente et al., 2019; Gioia et al., 2013). We suggest taking the opposite approach: include as much detail as possible. As more context informs the definition of events, the number of categories of events increases, with each event category becoming one node in the narrative network. Yet, the total number of event categories that actually occurs in the data – the actualized combinations – remains relatively small: in the data we use here, 458 actualized event nodes exist, making up about 7.2% of the combinations that could occur. If every role performed every action at every workstation, there would be 6 roles x 48 actions x 22 workstations = 6336 combinations. If we analyzed a larger sample, that fraction would undoubtedly rise, but overall it will remain low because the world is full of structure and specialization: not everyone uses every tool to do every task in every location. In other words, more categories do not simply clutter the graph. They present meaningful new information while also adding clarifying absence into the visualization. When we include more contextual dimensions in the definition of events, we can begin to visualize that structure in a novel and interesting way, which helps inductive theorizing.

This approach is a dramatic departure from standard research practice. In most analyses, we strive to keep the number of codes (categories) to a minimum. Tools like NVivo or ATLAS.ti can be used for any number of first order constructs, but their key feature is to make it easy to reduce the number of second order constructs (Gioia et al., 2013) in order to speed up data analysis (Flick, 2018, p. 520). Second order constructs assist theoretical abstraction and scaling (Urquhart et al., 2010), but they also reduce the richness of the description by reducing entropy. Fewer codes are more manageable and easier to write about (Myers, 2009). The same approach is seen in process mining, as well. In process mining, less frequent paths are often just filtered out to make a cleaner looking process model (van der Aalst, 2009), one that is easier to comprehend (Breuker et al., 2016). In terms of Weick's (1969) trade-off between simplicity, accuracy and generality, a small number of abstract categories favors simplicity and generality at the expense of accuracy.

Entropy, density, and absence as context. While our contribution is primarily of qualitative value, it rests on quantitative properties: as we add contextual dimensions, we get more categories, and the entropy of the data increases.[5] Entropy is a measure of information content, so more entropy means more information. When we go from a lexicon of 48 unique actions to a lexicon of 458 unique action-role-workstation combinations, as shown in Figure 3, we have much more information available. Whether or not one calls these visualizations "richer," they are definitely more informative.

At the same time, the density of the network drops dramatically. This is because a rather small increase in number of edges (sequential relations between the nodes) is spread out over a much larger set of possible nodes.[6] Visually, we are able to see the sequential relationships that occurred in the data because they stand out more clearly against the background of the relationships that do *not* occur. The blank space makes the shapes more visible. And in a world with social and technical division of labor, where *some* actors use *some* tools to do *some* tasks, there is a lot of blank space. Most combinations never occur just as most affordances are never actualized (Strong et al., 2014). By incorporating more contextual information in the description of the process, we not only change the apparent structure, we improve the visualization. This effect is demonstrated vividly by the images in Figure 3.

The effect we are describing here is the result of including sequential context in our analysis. Sequential context is an essential aspect of all processual phenomena (what happened before? What

[5] *Threadnet* computes the entropy values shown in Figure 3 using the standard Shannon formula: $-\sum p \log p$, where *p* is the probability of each category in the data. This is the entropy of the data as shown in the pie charts, which shows the relative frequency of each category in the pie. It does not include any information about sequence.

[6] *Threadnet* computes density using the standard formula for a directed graph: $\text{edges}/\text{nodes}^2$.

will happen next?). When we capture sequential relations between events in a narrative network, it creates an exponential canvas of possibilities. To build on the visual metaphor, this canvas has n^2 pixels, where n is the number of possible events we use to describe the process. Thus, the white space (the absence) grows exponentially faster than the observed events (the presence).

Absence as context is a powerful idea. The things that never occur provide a background for interpreting the things that do. The absent and the invisible are key issues with processual phenomena, especially where safety or any kind of undesirable outcome is at risk. For example, consider medical procedures or flight operations (Gawande, 2009). The work process may include checklists or other actions that are intended to prevent bad things from happening (e.g., infections, crashes). These elements can only be understood in light of what doesn't happen. Checklists can seem like a waste of time because the value they add shows up in what does not happen. Why does a team pause in the operating theatre before proceeding? Why does a doctor write on the patient's left elbow before going in to operate? The value added depends on what is prevented (the absence).

5 Future Research Directions

The approach we outline here has the potential to generate new research directions on the basis of digital trace data and contextual process analysis. It emphasizes processes as units of analysis (rather than objects or actors). It provides a formal way to represent process as a narrative network, as suggested by Pentland and Feldman (2007). Because more and richer digital traces are becoming available, the network approach is not only more feasible. We showed how such a network can include more context, as well.

5.1 The role of context in process dynamics

Our approach provides a foundation for process dynamics as network dynamics (Goh & Pentland, 2019). By dynamics, we mean changes to the structure of a process over time. Digitized processes are a prime candidate for exploring such questions (Pentland, Liu, Kremser, & Hærem, 2019). We can also use this approach for diachronic (or longitudinal) comparison (Barley, 1990; Berente et al., 2019). In diachronic analysis, we are interested in change over time: What is the trajectory of the process? What keeps it on track? What causes it to change? The first step in answering these questions is to represent the process within the context in which it unfolds.

Diachronic analysis further sets the stage for inquiry into why processes take the form they do. Beyond describing and modeling processes, we can begin to explain and predict how processes form and change over time. These are central themes in research on routine dynamics (Feldman, Pentland, D'Adderio, & Lazaric, 2016). What we present here is purely descriptive: one day at one clinic. Echoing Gregor (2006), the opportunity (and challenge) is now to move from level 1 (description) to higher levels (explanation and prediction).

5.2 Processual perspective on heterogeneous ensembles

Because we allow varied definition of context, our approach generalizes easily to non-human actors. While traditional behavioral science assumes that actors are human, recent theories of sociomateriality point towards the increasing importance of technology in the constitution and emergence of agency (D'Adderio, 2008, 2011; Faulkner & Runde, 2009; Leonardi, 2011). With digital technology becoming increasingly malleable, performative, and editable (Ekbia, 2009; Kallinikos, Aaltonen, & Marton, 2013; Yoo, Henfridsson, & Lyytinen, 2010), agency in routines is becoming less pre-defined, more distributed and no longer solely human-centric (Beane & Orlikowski, 2015; Leonardi, 2011; Orlikowski, 2007). As we allow more situational context to enter the definition of an event (e.g., actor and artefact), we now place social and material agents on equal footing: we "de-center" traditional categories because we treat everything equally. Actors, artifacts, actions are all just aspects of context. As artificial intelligence increasingly plays a role in organizational and private processes and routines, our approach allows visualizing, and analyzing how social and material agents interact.

5.3 Process theorizing with contextual digital trace data

The application of narrative networks *with context* is dependent on the availability, richness and quality of digital trace data, and the ability to collect and encode sequential trace data that contains meaningful contextual categories. At present, several issues remain that condition the possible use of our approach and require further research and development:

1. **Granularity.** Granularity is always an issue in processual analysis (Poole et al., 2017). The idea of incorporating situational and sequential context into process analysis benefits from data that includes multiple levels of temporal granularity, as some contextual dimensions can change at different rates. To gain meaningful insights, at least some contextual dimensions must be captured at the time scale of the phenomena being investigated or faster.

2. **Observability.** Observability is another fundamental concern (Poole et al., 2017). Only observable aspects of context can be included. For example, if one wanted to include what people were thinking as a contextual factor, these data would need to be captured somehow. Systems that record event logs, by contrast, have ability to record context data independent of pace of change; however, most process-aware information systems tend to record only a limited amount of observable context.
3. **Sequential coherence.** This framework requires data that have a coherent, narrative structure. To create a meaningful narrative networks, one must start from meaningful narratives -- sequences of events that are related. For example, in our EMR data, events are related because they are part of a patient visit.
4. **Data quantity.** In principle, our approach does not require large amounts of data. For example, the methodology outlined by Pentland and Liu (2017) assumes that data are collected through structured interviews. However, digital trace data makes it possible to compare processes across time and space in ways that would be difficult with interviews or observations. Still, processes unfold across technological and organizational containers, which makes it difficult to trace events at the same level of observability and granularity and which may require imputing their sequential coherence (Bala et al., 2018; Bayomie, Di Ciccio, la Rosa, & Mendling, 2019)
5. **Pre-coding is required.** To incorporate context into processual analysis requires that contextual categories are coded. This is the hard work that qualitative researchers perform using tools like NVivo and ATLAS.ti. For example, a corpus of email messages, where the main body of the data is un-coded text, cannot be directly analyzed. The messages would need to be coded. In many kinds of digital trace data (like the EMR data reported here), some categorization (e.g., into actions, actors or location) is available, but data quality and coding remains an important precondition (Bose, Mans, & van der Aalst, 2013).
6. **Limits of Dimensionality**. While adding contextual dimensions can be helpful, we cannot add dimensions indefinitely. As Bellman (1957) pointed out, as the dimensionality of a feature space increases, the number of configurations grows exponentially. For the reasons we explain above, this can work in our favor when visualizing processes as narrative networks. At the same time, we must be mindful of the corollary challenge: the number of configurations covered by given set of observations can decrease. Thus, we may be seeing only a fraction of possible

process patterns. Fortunately, using 3-4 dimensions to represent a process is safely within normal human experience.

6 CONCLUSION

Bringing context into the description of processual phenomena involves a re-orientation in our methods and our thinking, beyond switching figure and ground, by putting actions in the foreground and actors in the background. It allows re-examining what counts as figure and ground in order to analyze and ultimately theorize about processual phenomena in different ways from digital trace data.

Our contribution is primarily conceptual, not computational, but our working software *Threadnet* provides the computational support necessary to make the concepts useful in practice. Traditionally, with a moderate corpus of field notes, it is feasible to code the data and construct networks by hand. But as we have shown here, restricting our analysis to a modest number of categories (regardless of how defined) tends to suppress the richness of how we see and interpret processual phenomena. Bringing context inside provides a new approach – not for proposing answers, but for asking new and different questions, an ability that will gain prominence as more and more digital trace data becomes available for study.[7] And as calls for computational methods for theorizing are increasing (Berente et al., 2019; Lindberg, 2020), a strong conceptual focus on context and its role in process theorizing allows us to shift our focus away from "what explains the process" (Van de Ven & Poole, 1995) to more nuanced questions around "how does it change, and why?", "how is it different?" and other inquiries of comparison, dynamics and emergence in a digital world.

REFERENCES

Abbott, A. (1990). A primer on sequence methods. *Organization Science, 1*(4), 375-392.

Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology, 21*(1), 93-113.

Abbott, A. (2016). *Processual sociology*. Chicago, Illinois: University of Chicago Press.

Abbott, A., & Hrycak, A. (1990). Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers *American Journal of Sociology, 96*(1), 144-

[7] For example, we are aware that Pachilova and Sailer (2019) use our approach in conjunction with digital trace data from logged activities and locations, and digital hospital layout plans to study how spatial layout designs influence the way staff and patients interact in hospital routines.

185.

Alaimo, C., & Kallinikos, J. (2017). Computing the everyday: Social media as data platforms. *The Information Society, 33*(4), 175-191.

Alvesson, M., & Kärreman, D. (2007). Constructing mystery: Empirical matters in theory development. *Academy of Management Review, 32*(4), 1265-1281.

Austin, J. L. (1962). *How to do things with words*. Oxford, England: Harvard University Press.

Avgerou, C. (2019). Contextual explanation: Alternative approaches and persistent challenges. *MIS Quarterly, 43*(3), 977-1006.

Bala, S., Mendling, J., Schimak, M., & Queteschiner, P. (2018). Case and activity identification for mining process models from middleware. In R. A. Buchmann, D. Karagiannis & M. Kirikova (Eds.), *Poem 2018: The practice of enterprise modeling* (Vol. 335, pp. 86-102). Vienna, Austria: Springer.

Barley, S. R. (1990). Images of imaging: Notes on doing longitudinal field work. *Organization Science, 1*(3), 220-247.

Barnes, B., & Law, J. (1976). Whatever should be done with indexical expressions? *Theory and Society, 3*(2), 223-237.

Baskerville, R., Myers, M. D., & Yoo, Y. (2020). Digital first: The ontological reversal and new challenges for is research. *MIS Quarterly, 44*, forthcoming.

Bayomie, D., Di Ciccio, C., la Rosa, M., & Mendling, J. (2019). A probabilistic approach to event-case correlation for process mining. In A. H. F. Laender, B. Pernici, E.-P. Lim & J. P. M. de Oliveira (Eds.), *Conceptual modeling: Er2019* (Vol. 11788, pp. 136-152). Salvador, Brazil: Springer.

Beane, M., & Orlikowski, W. J. (2015). What difference does a robot make? The material enactment of distributed coordination *Organization Science, 26*(6), 1553-1573.

Bellman, R. (1957). *Dynamic programming*. Princeton, New Jersey: Princeton University Press.

Berente, N., Seidel, S., & Safadi, H. (2019). Data-driven computationally-intensive theory development. *Information Systems Research, 30*(1), 50-64.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. Los Angeles, California: Sage.

Bose, R. P. J. C., Mans, R. S., & van der Aalst, W. M. P. (2013). *Wanna improve process mining results?* Paper presented at the IEEE Symposium on Computational Intelligence and Data Mining, Singapore, Singapore.

Bose, R. P. J. C., & van der Aalst, W. M. P. (2009). *Context aware trace clustering: Towards improving process mining results*. Paper presented at the SIAM International Conference on Data Mining Sparks, Nevada.

Breuker, D., Matzner, M., Delfmann, P., & Becker, J. (2016). Comprehensible predictive models for business processes. *MIS Quarterly, 40*(4), 1009-1034.

Burke, K. (1962). *A grammar of motives, and a rhetoric of motives*. Cleveland, Ohio: World Pub. Co.

Burton-Jones, A., McLean, E. R., & Monod, E. (2015). Theoretical perspectives in is research: From variance and process to conceptual latitude and conceptual fit. *European Journal* *of Information Systems, 24*(6), 664-679.

Burton-Jones, A., & Volkoff, O. (2017). How can we develop contextualized theories of effective use? A demonstration in the context of community-care electronic health records. *Information Systems Research, 28*(3), 468-489.

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems, 63*, 67-80.

Charmaz, K. C. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London, England: Sage.

Ciborra, C. (2000). *From control to drift: The dynamics of corporate information infrastructures*. New York, New York: Oxford University Press.

D'Adderio, L. (2008). The performativity of routines: Theorising the influence of artifacts and distributed agencies on routine dynamics. *Research Policy, 37*(5), 769-789.

D'Adderio, L. (2011). Artifacts at the centre of routines: Performing the material turn in routines theory. *Journal of Institutional Economics, 7*(2), 197-230.

de Saussure, F. (1974). *Course in general linguistics*. London, England: Peter Owen Ltd.

De Weerdt, J., vanden Broucke, S., Vanthienen, J., & Baesens, B. (2013). Active trace clustering for improved process discovery. *IEEE Transactions on Knowledge and Data Engineering, 25*(12), 2708-2720.

Derrida, J. (1981). *Positions*. London, England: Athlone Press.

Dhar, V. (2013). Data science and prediction. *Communications of the ACM, 56*(12), 64-73.

Dianati, N. (2016). Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review, E93*, 012304-012301 - 012304-012308.

Eco, U. (1976). *A theory of semiotics*. London, England: The Macmillan Press.

Ekbia, H. R. (2009). Digital artifacts as quasi-objects: Qualification, mediation, and materiality. *Journal of the American Society for Information Science and Technology, 60*(12), 2554-2566.

Emirbayer, M. (1997). Manifesto for a relational sociology. *American Journal of Sociology, 103*(2), 281-317.

Faulkner, P., & Runde, J. (2009). On the identity of technological objects and user innovations in function. *Academy of Management Review, 34*(3), 442-462.

Feldman, M. S. (1995). *Strategies for interpreting qualitative data* (Vol. 13). Thousand Oaks, California: Sage.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly, 48*(1), 94-118.

Feldman, M. S., Pentland, B. T., D'Adderio, L., & Lazaric, N. (2016). Beyond routines as things: Introduction to the special issue on routine dynamics. *Organization Science, 27*(3), 505-513.

Flick, U. (2018). *An introduction to qualitative research* (6th ed.). London, England: Sage.

Floridi, L. (2012). The road to the philosophy of information. In H. Demir (Ed.), *Luciano floridi's philosophy of technology* (pp. 245-271). Dordrecht, The Netherlands:

Springer.

Freelon, D. (2014). On the interpretation of digital trace data in communication and social computing research. *Journal of Broadcasting & Electronic Media, 58*(1), 59-75.

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in r with traminer. *Journal of Statistical Software, 40*(4), 1-37.

Gaskin, J., Berente, N., Lyytinen, K., & Yoo, Y. (2014). Toward generalizable sociomaterial inquiry: A computational approach for zooming in and out of sociomaterial routines. *MIS Quarterly, 38*(3), 849-871.

Gawande, A. (2009). *The checklist manifesto: How to get things right*. New York, New York: Metropolitan Books.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal, 59*(5), 1493-1507.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research. *Organizational Research Methods, 16*(1), 15-31.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. London, England: Harper and Row.

Goh, J. M., Gao, G., & Agarwal, R. (2011). Evolving work routines: The adaptive routinization of technology in healthcare. *Information Systems Research, 22*(3), 565-585.

Goh, K. T., & Pentland, B. T. (2019). From actions to paths to patterning: Toward a dynamic theory of patterning in routines. *Academy of Management Journal*(In press).

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly, 30*(3), 611-642.

Heritage, J. (2013). *Garfinkel and ethnomethodology*. New York, New York: John Wiley & Sons.

Hernes, T. (2007). *Understanding organization as process: Theory for a tangled world* (Vol. 2). London, England: Routledge.

Hernes, T. (2017). Process as the becoming of temporal trajectory. In A. Langley & H. Tsoukas (Eds.), *The sage handbook of process organization studies* (pp. 601-606). Thousand Oaks, California: Sage.

Hong, W., Chan, F. K. Y., Thong, J. Y. L., Chasalow, L. C., & Dhillon, G. (2014). A framework and guidelines for context-specific theorizing in information systems research. *Information Systems Research, 25*(1), 111-136.

Indulska, M., Hovorka, D. S., & Recker, J. (2012). Quantitative approaches to content analysis: Identifying conceptual drift across publication outlets. *European Journal of Information Systems, 21*(1), 49-69.

Janssenswillen, G., & Depaire, B. (2017). *Bupar: Business process analysis in r*. Paper presented at the 15th International Conference on Business Process Management, Barcelona, Spain.

Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review, 31*(2), 386-408.

Kallinikos, J., Aaltonen, A., & Marton, A. (2013). The ambivalent ontology of digital artifacts. *MIS Quarterly, 37*(2), 357-370.

Kim, I., Pentland, B. T., Ryan Wolf, J., Xie, Y., Frank, K., & Pentland, A. (2019). Effect of attribute alignment on action sequence variability: Evidence from electronic medical records. In T. Hildebrandt, B. F. van Dongen, M. Röglinger & J. Mendling (Eds.), *Business process management forum: Bpm forum 2019* (pp. 183-194). Vienna, Austria: Springer.

Kling, R., & Scacchi, W. (1982). The web of computing: Computer technology as social organization *Advances in computers* (Vol. 21, pp. 1-90). New York, New York: Academic Press.

Kunzman, K. (2018). Why are emrs so terrible? *MD Magazine*.

Larsen, K. R. T., & Monarchi, D. E. (2004). A mathematical approach to categorization and labeling of qualitative data: The latent semantic categorization method. *Sociological Methodology, 34*(1), 349-392.

Lazer, D., Pentland, A. P., Adamic, L. A., Aral, S., Barabási, A.-L., Brewer, D., . . . Van Alstyne, M. (2009). Computational social science. *Science, 323*(5915), 721-723.

Lee, C., Luo, Z., Ngiam, K. Y., Zhang, M., Zheng, K., Chen, G., . . . Yip, W. L. J. (2017). Big healthcare data analytics: Challenges and applications *Handbook of large-scale distributed computing in smart healthcare* (pp. 11-41): Springer.

Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly, 35*(1), 147-167.

Levina, N., & Vaast, E. (2015). Leveraging archival data from online communities for grounded process theorizing. In K. D. Elsbach & R. M. Kramer (Eds.), *Handbook of qualitative organizational research: Innovative pathways and methods* (pp. 215-224). New York, New York: Routledge.

Lewis, S. J., & Russell, A. J. (2011). Being embedded: A way forward for ethnographic research. *Ethnography, 12*(3), 398-416.

Lindberg, A. (2020). Developing theory through integrating human & machine pattern recognition. *Journal of the Association for Information Systems, 21*, forthcoming.

Lyytinen, K., & Yoo, Y. (2002). Research commentary: The next wave of nomadic computing. *Information Systems Research, 13*(4), 377-388.

Mertens, W., Recker, J., Kohlborn, T., & Kummer, T.-F. (2016). A framework for the study of positive deviance in organizations. *Deviant Behavior, 37*(11), 1288-1307.

Myers, M. D. (2009). *Qualitative research in business and management*. Thousand Oaks, California: Sage.

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science, 11*(4), 404-428.

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies, 28*(9), 1435-1148.

Orlikowski, W. J., & Iacono, C. S. (2001). Research commentary: Desperately seeking the 'it' in it research-a call to theorizing the it artifact. *Information Systems Research, 12*(2), 121-134.

Orlikowski, W. J., & Scott, S. V. (2008). Sociomateriality: Challenging the separation of technology, work and organization. *The Academy of Management Annals, 2*(1), 433-

474.

Pachilova, R., & Sailer, K. (2019). *Ward layout, communication and care quality: Spatial intelligibility as a key component of hospital design*. Paper presented at the 12th Space Syntax Symposium, Beijing, China.

Pentland, B. T. (1999). Building process theory with narrative: From description to explanation. *Academy of Management Review, 24*(4), 711-725.

Pentland, B. T., & Feldman, M. S. (2007). Narrative networks: Patterns of technology and organization. *Organization Science, 18*(5), 781-795.

Pentland, B. T., Feldman, M. S., Becker, M. C., & Liu, P. (2012). Dynamics of organizational routines: A generative model. *Journal of Management Studies, 49*(8), 1484-1508.

Pentland, B. T., Haerem, T., & Hillison, D. (2010). Comparing organizational routines as recurrent patterns of action. *Organization Studies, 31*(7), 917-940.

Pentland, B. T., & Liu, P. (2017). Network models of organizational routines: Tracing associations between actions. In S. Jain & R. Mir (Eds.), *Routledge companion to qualitative research in organization studies* (pp. 422-438). New York, New York: Routledge.

Pentland, B. T., Liu, P., Kremser, W., & Hærem, T. (2019). The dynamics of drift in digitized processes. *MIS Quarterly*, forthcoming.

Pentland, B. T., Pentland, A. P., & Calantone, R. J. (2017a). Bracketing off the actors: Towards an action-centric research agenda. *Information and Organization, 27*(3), 137-143.

Pentland, B. T., Recker, J., & Kim, I. (2017b). *Capturing reality in flight? Empirical tools for strong process theory*. Paper presented at the 38th International Conference on Information Systems, Seoul, Republic of Korea. http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1208&context=icis2017

Pentland, B. T., Recker, J., & Wyner, G. (2015). *A thermometer for interdependence: Exploring patterns of interdependence using networks of affordances*. Paper presented at the 36th International Conference on Information Systems, Fort Worth, Texas.

Pentland, B. T., Recker, J., & Wyner, G. (2016). *Conceptualizing and measuring interdependence between organizational routines*. Paper presented at the 37th International Conference on Information Systems, Dublin, Ireland.

Pentland, B. T., Recker, J., & Wyner, G. (2017c). Rediscovering handoffs. *Academy of Management Discoveries, 3*(3), 284-301.

Pettigrew, A. M. (1997). What is a processual analysis. *Scandinavian Journal of Management, 13*(4), 337-348.

Pettigrew, A. M. (2012). Context and action in the transformation of the firm: A reprise. *Journal of Management Studies, 49*(7), 1304-1328.

Pickering, A. (2010). *The mangle of practice: Time, agency, and science*. Chicago, Illinois: University of Chicago Press.

Plsek, P. E., & Wilson, T. (2001). Complexity, leadership, and management in healthcare organisations. *British Medical Journal, 323*(7315), 746-749.

Poole, M. S., Lambert, N., Murase, T., Raquel, A., & Joseph, M. (2017). Sequential analysis of processes. In A. langley & H. Tsoukas (Eds.), *The sage handbook of process organization studies* (pp. 254-270). Thousand Oaks, California: Sage.

Recker, J., Rosemann, M., Indulska, M., & Green, P. (2009). Business process modeling: A comparative analysis. *Journal of the Association for Information Systems, 10*(4), 333-363.

Rosemann, M., Recker, J., & Flender, C. (2008). Contextualization of business processes. *International Journal of Business Process Integration and Management, 3*(1), 47-60.

Rotman, B. (2016). *Signifying nothing: The semiotics of zero* (2nd ed.). London, England: Palgrave Macmillan Limited.

Sanjek, R. (1991). The ethnographic present. *Man, 26*(4), 609-628.

Schensul, S. L., Schensul, J. J., & LeCompte, M. D. (1999). *Essential ethnographic methods: Observations, interviews, and questionnaires*. Walnut Creek, California: AltaMira Press.

Schofield, J. W. (2002). Increasing the generalizability of qualitative research. In M. Huberman & M. B. Miles (Eds.), *The qualitative researcher's companion* (pp. 171-203). Thousand Oaks, California: Sage.

Strong, D. M., Volkoff, O., Johnson, S. A., Pelletier, L. R., Bar-On, I., Tulu, B., . . . Garber, L. (2014). A theory of clinic-ehr affordance actualization. *Journal of the Association for Information Systems, 15*(2), 53-85.

Sundararajan, A., Provost, F., Oestreicher-Singer, G., & Aral, S. (2013). Research commentary: Information in digital, economic, and social networks. *Information Systems Research, 24*(4), 883-905.

Swanson, E. B. (2019). Technology as routine capability. *MIS Quarterly, 43*(3), 1007-1024.

Törnberg, P., & Törnberg, A. (2018). The limits of computation: A philosophical critique of contemporary big data research. *Big Data & Society, 5*(2), 2053951718811843. doi: https://doi.org/10.1177/2053951718811843

Tsoukas, H., & Chia, R. (2002). On organizational becoming: Rethinking organizational change. *Organization Science, 13*(5), 567-582.

Urquhart, C. (2013). *Grounded theory for qualitative research: A practical guide*. Thousand Oaks, California: Sage.

Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the theory back into grounded theory: Guidelines for grounded theory studies in information systems. *Information Systems Journal, 20*(4), 357-381.

Van de Ven, A. H., & Poole, M. S. (1990). Methods for studying innovation development in the minnesota innovation research program. *Organization Science, 1*(3), 313-335.

Van de Ven, A. H., & Poole, M. S. (1995). Explaining development and change in organizations *Academy of Management Review, 20*(3), 510-540.

van der Aalst, W. M. P. (1998). The application of petri nets to workflow management. *The Journal of Circuits, Systems and Computers, 8*(1), 21-66.

van der Aalst, W. M. P. (2005). Business alignment: Using process mining as a tool for delta analysis and conformance testing. *Requirements Engineering, 10*(3), 198-211.

van der Aalst, W. M. P. (2009). Tomtom for business process management (tomtom4bpm) In P. van Eck, J. Gordijn & R. Wieringa (Eds.), *Advanced information systems engineering - caise 2009* (Vol. 5565, pp. 2-5). Amsterdam, The Netherlands: Springer.

van der Aalst, W. M. P. (2011a). *Process mining: Discovering and improving spaghetti and lasagna processes* Paper presented at the 2011 IEEE Symposium on Computational Intelligence and Data Mining, Paris, France.

van der Aalst, W. M. P. (2011b). *Process mining: Discovery, conformance and enhancement of business processes*. Heidelberg, Germany: Springer.

van der Aalst, W. M. P. (2016). *Process mining: Data science in action*. Berlin Heidelberg: Springer.

van der Aalst, W. M. P., & Dustdar, S. (2012). Process mining put into context. *IEEE Internet Computing, 16*(1), 82-86.

van der Aalst, W. M. P., Weijters, A. J. M. M., & Maruster, L. (2004). Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering, 16*(9), 1128-1142.

van Dongen, B. F., Alves de Medeiros, A. K., Verbeek, H. M. V., Weijters, A. J. M. M., & van der Aalst, W. M. P. (2005). The prom framework: A new era in process mining tool support. In G. Ciardo & P. Darondeau (Eds.), *Applications and theory of petri nets 2005* (Vol. 3536, pp. 444-454). Miami, Florida: Springer.

Wassermann, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, England: Cambridge University Press.

Weick, K. E. (1969). *The social psychology of organizing* (2nd ed.). New York, New York: McGraw-Hill.

Whelan, E., Teigland, R., Vaast, E., & Butler, B. S. (2016). Expanding the horizons of digital social networks: Mixing big trace datasets with qualitative approaches. *Information and Organization, 26*(1-2), 1-12.

Whetten, D. A. (2009). An examination of the interface between context and theory applied to the study of chinese organizations. *Management and Organization Review, 5*(1), 29-55.

Winter, S., Berente, N., Howison, J., & Butler, B. S. (2014). Beyond the organizational 'container': Conceptualizing 21st century sociotechnical work. *Information and Organization, 24*(4), 250-269.

Xu, H., Zhang, N., & Zhou, L. (2019). Validity concerns in research using organic data. *Journal of Management*, In Press.

Yeow, A., & Faraj, S. (2011). Using narrative networks to study enterprise systems and organizational change. *International Journal of Accounting Information Systems, 12*(2), 116-125.

Yoo, Y. (2010). Computing in everyday life: A call for research on experiential computing. *MIS Quarterly, 34*(2), 213-231.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research, 21*(4), 724-735.

Author Biographies

Brian T. Pentland is the Main Street Capital Partners Endowed Professor in the Department of Accounting and Information Systems at Michigan State University. His research is focused on the analysis of repetitive patterns of action, such as organizational routines. His creative work has appeared in *Academy of Management Review, Academy of Management Journal, Accounting, Organizations and Society, Administrative Science Quarterly, JAIS, Journal of Management Studies, Management Science, MIS Quarterly, Organization Science, Organization Studies, YouTube, Soundcloud,* and elsewhere. He received his PhD in Management from the Massachusetts Institute of Technology in 1991 and SB in Mechanical Engineering from the Massachusetts Institute of Technology in 1981.

Jan Recker is an AIS fellow, Alexander-von-Humboldt fellow, chaired professor of information systems and systems development at the University of Cologne, and adjunct professor at Queensland University of Technology. He is presently editor in chief of the *Communications of the Association for Information Systems*. His research focuses on systems analysis and design, digital innovation and entrepreneurship, and digital solutions for sustainability challenges.

Julie Ryan Wolf is an Associate Professor of Dermatology and Radiation Oncology, as well as a member of the University of Rochester NCI Community Oncology Research Program (NCORP) Research Base. She received her BA from University of Chicago, her PhD in Pathology from UNC-Chapel Hill, and her MPH from University of Rochester. Her research combines the fields of dermatology, radiation biology, oncology, and health services/outcomes research. Dr. Wolf strives to improve healthcare by facilitating access to evidence-based, patient-centered care and to optimize metrics for assessing quality of care. Her research is currently or has been funded by the National Science Foundation (NSF), National Cancer Institute (NCI), Pfizer, Wilmot Cancer Institute, Hope Foundation, Biomedical Advanced Research and Development Authority (BARDA), National Institutes of Allergy and Infectious Disease (NIAID), Dermatology Foundation, and University of Rochester CTSA.

George Wyner is an Associate Professor of the Practice of Information Systems at the Carroll School of Management at Boston College. His research interests include the analysis and design of organizational processes and the information systems that support them. He has published in *Management Science, MIS Quarterly, Academy of Management Discoveries, Information Systems,* and elsewhere.