Predicting process structure after a disruption: An example from the clinical documentation process[1]

Inkyu Kim[1], Kenneth A. Frank[1], Julie Ryan Wolf[2] and Brian T. Pentland[1]

[1] Michigan State University, East Lansing, MI USA
[2] University of Rochester, Rochester, NY USA
kiminky1@broad.msu.edu

Abstract. Using data from the audit trail of an electronic medical record system, we examine the effects of a disruption on the clinical documentation process. We use process mining to construct a network that describes the process and then we use a latent factor selection model to analyze changes to that network. Rather than attempting to discover a particular process model, our goal is to identify theory-based factors that explain change and stability in the overall pattern of actions. We conduct the analysis at two levels of granularity and we compare time periods with and without disruption. The paper contributes to current research on routine dynamics as network dynamics by demonstrating the use of network science to predict the structure of an organizational routine.

Keywords: Routine Dynamics, Network Dynamics, Latent Factor Models

1 Introduction

Organizational routines (defined as recognizable, repetitive patterns of interdependent action carried out by multiple actors) are a foundational element in the science of organization [1]. Over the last 20 years, the field of routine dynamics has focused on the mechanisms of endogenous change: change that occurs in the absence of external influences.

In this paper, we turn our attention to the effects of exogenous disruptions: When routines are disrupted, are how does the overall pattern of action change? Other papers have studied disruptions (e.g., [2]), but this is the first to apply concepts from network science to explain the dynamics of organizational routines. We model routines as directed graphs [3], known in the process management literature as "directly follows graphs" [4]. Using latent factor selection models [5], we study the effects of disruptions

[1] This research was 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 author(s) and does not necessarily represent the official views of the National Institutes of Health.

on the structure of the network. In social networks, mechanisms like reciprocity, homophily, and preferential attachment contribute to formation and dissolution of network ties [6], but analogous network-based mechanisms have never been defined or investigated in the context of organizational routines.

This paper offers a first step towards defining and testing mechanisms that drive routine dynamics. As a sophisticated tool to model and measure process change and theory to explain patterns of actions, process mining and routine dynamics have potential to be mutual complementary, but there has been disconnection between two areas [7]. In this study, we provide implications with respect to intersection of process management and routine dynamics by using a deductive method of process mining to examine dynamics of routines in organizations [8].

2 Background

Routine dynamics concerns understanding the mechanisms that influence stability or change in action patterns [9]. When action patterns persist over time, this persistence can be interpreted in several ways, such as inertia [10], resistance [11], persistence [12], regeneration [13], or resilience [14]. Schulz [15] offers an encyclopedic list of mechanisms that keeps routines "on track", ranging from very macro (institutional norms) to very micro (neuronal priming). Network structure does not appear in the list of mechanisms because at that time, nobody was thinking about modeling routines as directed graphs. But as routines are repeated related actions, they can be understood in terms of the evolution or stability of networks representing the relationships among actions. Theoretical explanations of routine persistence did not consider the structure of the routine itself.

2.1 Routine dynamics as network dynamics

Organizational routines can be seen as patterns of action [1, 16]. These patterns can be represented as a valued, directed graph where the vertices represent categories of action and the edges represent sequential relations between those categories [3]. In process mining, this is called a "directly follows graph" (DFG) [4]. Where a conventional social network represents relations between actors (e.g., people), a DFG represents relations between categories (or clusters) of actions. In research on organizational routines, these graphs are most often referred to as “narrative networks” [17].

To model network dynamics, we need to explain edge formation/dissolution, which is the fundamental mechanism of network dynamics [6]. The structure of the network changes as edges are added or deleted. In social network research, models that predict edge formation or deletion are often referred to as selection models because they predict how people select other people as interaction partners [18]. There are some well-established selection mechanisms in social networks, such as homophily and preferential attachment, that drive network dynamics (edge selection). Our goal here is to begin, for the first time, to identify and test mechanisms that drive routine dynamics.

To address this issue, we use two different levels of granularity; 1) a fine-grained level consisting of screens and clicks and 2) a coarse-grained level composed of touch-point and handoffs. In this study, we use the expanded concept of handoffs as a sequential relationship between not just people, but also between events [3]. As a repetitive patterns of sequentially related events, we can view organizational routines as a series of handoffs [19, 20].

Our main theoretical concern is the extent to which network structure itself has an influence on routine dynamics. If it does, it would provide a new theoretical mechanism to explain and predict the inertia/persistence/resilience of routines, complementary to existing explanations.

2.2 Research setting and data

This analysis is based on data from one outpatient dermatology clinic at the University of Rochester Medical Center (URMC). We compare the pattern of action before and after the start of flu season (September 1, 2016). From prior research on this clinic, we know that flu season causes a measurable change in the EHR audit trail, as shown in Figure 1 [7]. Figure 1 describes changes in the structure of the clinical documentation process over time compared to a fixed reference. In particular, it uses cosine distance to show how the repertoire of actions used in the process changes over time. Each point in Figure 1 represents one day. It shows that the repertoire of actions changes abruptly at the beginning of flu season. This is because a new set of actions is added to the system to track and support seasonal vaccination. Later, at the end of flu season, the repertoire of actions changes back. However, Figure 1 is purely descriptive. Our goal here is to move beyond description to provide a theory-based model that can help explain how the pattern of action responds to this exogenous disruption.

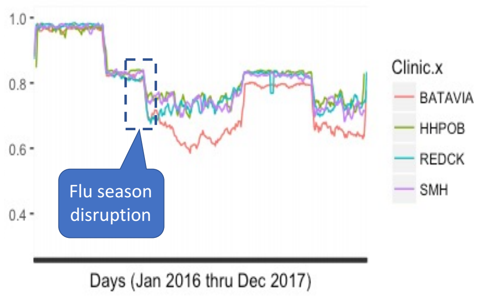

Fig. 1: Flu season disrupts the normal clinical routine (Adapted from Pentland, Vaast and Wolf [7])

In this paper, we focus on three weeks of data from one clinic. We examine the pattern of action two weeks prior to flu season and one week after. This is a small subset of the data collected for the larger project, but it is adequate to demonstrate our approach. Unlike a typical exercise in process mining, our goal is not to discover a process model. Rather, we are trying to identify factors that influence changes in the process.

2.3 Two levels of granularity

We analyze data extracted from the audit trail of the EPIC Electronic Medical Record (EMR) system. The subset of records used here includes detailed, time-stamped records of EMR utilization in 627 patient visits at one clinic from August 18, 2016 (two weeks before the start of flu season) to September 7, 2016 (one week after). The middle column of Figure 2 includes a brief example of the audit trail data.

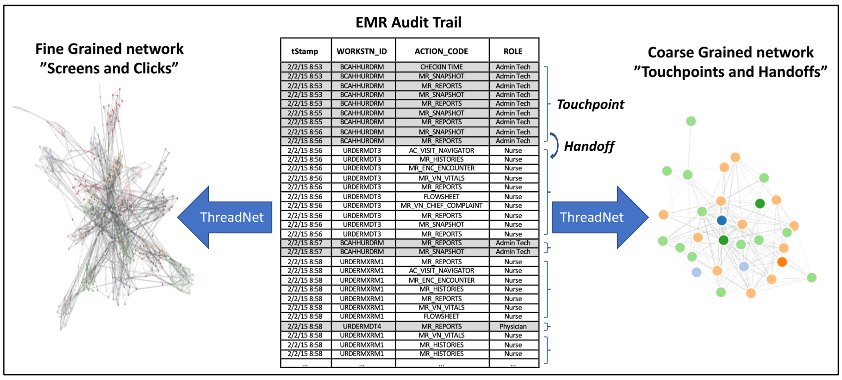

Fig. 2: EMR audit trail can be interpreted at two levels of granularity

Figure 2 shows how *ThreadNet* (Pentland et al, 2020) can be used to convert EMR audit trails into networks at two different levels of granularity. At the center of Figure 2 is a small part of an audit trail where each row is a time-stamped action. On the left side of Figure 2, *ThreadNet* can produce a very fine-grained network by considering each unique row as a node. At the fine-grained level, edges represent pairs of actions within the EMR system. On the right side of Figure 2, *ThreadNet* can produce a coarse grained network where each node consists of a group of actions and each edge can be interpreted as a *handoff [3]*. For example, in Figure 2, there is a handoff between the Admin Tech and the Nurse, and then another handoff back to the Admin Tech. We test our theory of network dynamics at both of these levels.

We examine this disruption at two levels of granularity because we expect the effects to be different. At the fine-grained level, flu season directly changes the repertoire of actions in the network. On September 1, a set of actions was added to the network. Office staff were prompted to ask all incoming patients about their vaccination status and patients were invited to schedule (or receive) a vaccination. This had a visible impact on the clinical documentation process, but the main work of the dermatology clinic was not otherwise affected. This would lead us to expect changes to the pattern of action at the fine-grained level, but not necessarily at the coarse-grained level. Pentland, Vaast and Wolf [7] report that the medical staff (physicians and nurses) were unaware of any particular changes in their work processes for the time period in question. From their point of view, there was no disruption at all.

3 Hypothesis development

Network dynamics can be defined in terms of two basic processes: edge formation and edge dissolution [6]. As with models of social networks, we recognize that formation of new edges is a different process than dissolution of existing edges. In a network model of a routine, an edge represents the sequential relationship of two actions or events. New sequential relationships might form for a variety of reasons that seem difficult to predict and generalize (experimentation, workaround, error, etc.) Therefore, we focus on mechanisms that influence the persistence (or dissolution) of existing edges. For brevity, we discuss and phrase the hypotheses in terms of handoffs, in the general sense of the term used by Pentland, Recker and Wyner [3]. A handoff is simply an edge in a narrative network.

3.1 Frequency of handoffs

Handoffs that occur frequently represent the stereotypical "ruts in the road" [13] that define routinized patterns of action. For example, in Figure 2, the patient checks into the clinic with the Admin Tech, who hands off to the nurse to record vital signs, etc. We operationalize frequency of handoffs by counting the edges between touchpoints in the network. Some handoffs are reinforced more frequently and may also be enabled or constrained by technological, material or organizational structures. We expect frequent handoffs to persist after a disruption to the network.

H_1: More frequent handoffs are more likely to persist after a disruption.

3.2 Speed of handoffs

Handoffs can also be weighted according to how long they take to perform, on average, using time-stamp data from the event log. Faster handoffs represent quicker ways of getting things done. The speed of handoffs is operationalized as how long it takes for an actor to perform handoffs. In other words, we compute mean duration of each handoff in the network. Clinics are busy places, so we hypothesize that faster handoffs (with shorter mean duration), are more likely to persist after a disruption.

H_2: Faster handoffs are more likely to persist after a disruption.

3.3 Paths and betweenness

The edge betweenness is defined as the number of shortest paths going through an edge in the network [21]. In a narrative network, paths represent ways of doing things. An edge that has greater betweenness is on more paths [22]. If a disruption changes how things get done, handoffs with high betweenness are less likely to dissolve (and more likely to persist or "bounce back") than handoffs with low betweenness:

H_3: Handoffs with higher betweenness are more likely to persist after a disruption.

3.4 Coherence

Coherence represents the number of nodal attributes that remain the same across an edge [3]. For example, do both actions occur in the same place? Are they performed by the same actor? Do they require the same tools or technology? This can be easily operationalized in a narrative network, where each node is defined by a number of contextual factors, such as place, actor and technology. When more factors change, the context is less coherent. The logic of this hypothesis is similar to the logic for effects of homophily in social networks ("birds of a feather…"), but the mechanism is different.

H_4: More coherent handoffs are more likely to persist after a disruption.

4 Formal model

Our goal is to predict the frequency of all edges in the DFG that represents the clinical documentation process. To do so, we use the previous state of the process (at time *t-1*) to predict the current state of the process (at time *t*). We can express our four hypotheses in terms of the formal model shown in equation (1):

$$w_{ijt} = \beta_1\left(w_{ijt-1}\right) + \beta_2\left(\overline{speed}_{ijt-1}\right) + \\ \beta_3\left(betweenness_{ijt-1}\right) + \beta_4\left(coherence_{ijt-1}\right) + \theta_i + \theta_j + u_i v_j + e \quad (1)$$

In this model, the time period *t* represents one week. This time scale makes sense because the pattern of action is generally very stable and the disruption from the start of flu season occurs on a specific day (September 1). The dependent variable in this model is w_{ijt}, which represents the frequency of each edge in the network (between actions *i* and *j*) during time period *t*.

The term w_{ijt-1}represents the frequency of edges from the previous time period, as in H_1. $\overline{speed}_{ijt-1}$ reflects the average speed or duration of the edge w_{ij}, as in H_2. $betweenness_{ijct-1}$ represents the number of network paths that pass through w_{ijc}, as in H_3. $coherence_{ijt-1}$ represents the extent to which actions i and j share a coherent material context, as in H_4.

θ_i and θ_j are random effects relating to the base rate of actions *i* and *j*. If *i* and *j* occur more or less often, that will directly influence how often w_{ij} occurs. In terms of this model, they function like control variables. As we apply the model here, θ_i and θ_j reflect the change in the repertoire of actions described in Figure 1.

Lastly, $u_i v_j$ represents the similarity between pairs of nodes on each dimension (action *i* and *j)* of latent space and $e_{i,j}$ is the error term.

5 Results

We estimate this model using a latent factor model including random effects [23, 24] to control for unobserved network effects. This model has the advantage of estimating all of the edges in the network at once, rather than treating them as independent. This model is implemented in the R package *amen* (https://cran.r-project.org/web/packages/amen/amen.pdf) which uses an MCMC (Markov Chain Monte Carlo) procedure for model estimation. This package provides regression coefficients, standard errors and significance levels, but does not provide an R^2 for the model.

We conduct four different versions of the analysis. First, as mentioned above, we examine two different levels of granularity: the fine-grained "screens and clicks" level and the coarse-grained "touchpoints and handoffs" level. At the fine-grained level, the nodes are defined by a single attribute – actions -- so there is no way to compute coherence or test H_4. At the coarse-grained level, nodes are defined by two attributes – actor and workstation – so we can compute coherence and test H_4.

Second, we compare two different time periods: one where there is a disruption and one where there is not. While a more elaborate analysis could be conducted, this provides an indication of how the disruption compares to the normal, steady state operations. The results of these analyses are shown in Table 1.

Table 1: Results of analysis at two level of granularity[2]

	Not Disrupted	Disrupted
Fine-grained (clicks/screens)	t-1: Aug 18 – Aug 24 t: Aug 25 – Aug 31	t-1: Aug 25 – Aug 31 t: Sept 01 – Sept 07
H_1: Frequency	1.912***	0.975***
	(0.060)	(0.039)
H_2: Duration	-0.044	0.014
	(0.024)	(0.025)
H_3: Betweenness	0.213***	0.469***
	(0.037)	(0.039)
Random Effect: θ_i	0.405	0.507
	(0.065)	(0.082)
Random Effect: θ_j	0.389	0.430
	(0.063)	(0.067)
# nodes (unique action)	146	137
# edges (pairs of actions)	1,848	1,800
Coarse-grained (touchpoints/handoffs)		
H_1: Frequency	1.052***	1.063***
	(0.086)	(0.088)
H_2: Duration	0.008	0.064*

[2] The values for the random effects (θ_i, θ_j) indicate variances.

	(0.027)	(0.032)
H_3: Betweenness	0.108	0.141
	(0.071)	(0.075)
H_4: Coherence	0.499***	0.608***
	(0.124)	(0.146)
Random Effect: θ_i	0.684	0.575
	(0.162)	(0.145)
Random Effect: θ_j	0.602	0.571
	(0.109)	(0.148)
# nodes (touchpoints)	86	80
# edges (handoffs)	919	838

5.1 Frequency of handoffs (H_1)

It should be no surprise that the frequency of an edge (handoff) in one time period is a good predictor of its frequency in the next time period. At both levels of granularity, with or without disruption, **H_1** is supported. Metaphorically, the ruts in the road today predict where the traffic will go tomorrow.

However, at the fine-grained level, the disruption appears to change the strength of this effect. Even when we control for other factors (such as the base rate of actions, θ_i and θ_j), the influence of handoff frequency is decreased by more than 50% from the prior week. Perhaps we should say: the ruts in the road today predict where the traffic will go tomorrow *unless the road is blocked*.

5.2 Duration of handoffs (H_2)

Contrary to expectations, the average duration of the handoff does not influence persistence, except perhaps at the coarse-grained level. Faster handoffs are slightly more likely to persist after disruption. Even in a very busy clinic, where time is presumably valuable, the waiting time from one action to the next appears to have a very minor influence on persistence. Conceptually, this could be interpreted as undermining the idea that routines enhance efficiency, but it would be worth testing this hypothesis on a larger, more diverse set of processes.

5.3 Betweenness of handoffs (H_3)

Betweenness provides different results at different levels of granularity. It is significant at the fine-grained level, but not at the coarse-grained level. We do not have an explanation for this difference, so further investigation is needed. However, at the fine-grained level, we can see that the magnitude of this effect is more than doubled when there is a disruption. In a narrative network, betweenness means that an edge (or a node) is on more paths; it is a useful for performing a larger variety of tasks. When there is a disruption, the useful handoffs are more likely to persist.

5.4 Coherence of handoffs (H_4)

Because of the way coherence is defined, we are only able to test its effect at the coarse-grained level. At that level, coherence is a strong predictor of edge persistence. When a pair of actions is more coherent (performed by the same actor, or performed in the same location), it is more likely to persist (with or without a disruption). This finding is interesting because this effect is significant even when we control for the effect of frequency. Coherence appears to influence the persistence of network structure over and above the more familiar effect of repetition.

5.5 Changes in the base rate of actions

In a sense, we are interpreting random effects in equation (1) (θ_i and θ_j) as control variables: Controlling for changes in base rates of the actions, what drives changes in the pairs of actions? The magnitude of these effects is far greater than their standard error in all four conditions (the *amen* package does not provide p-values for these parameters, so we rely on the eyeball test for significance). Thus, we can safely say that in all four conditions, there is considerable variation among actions in the tendency to be part of a chain of actions defining a routine. However, when we compare the coefficients in Table 1, with and without disruption, they are of similar magnitude and sign in all four conditions (with or without the disruption at both levels of granularity).

6 Discussion

This paper represents a first step towards a theory of routine dynamics as network dynamics. The empirical foundation for this theory is generated through process mining, which is usually used to discover a stationary model of a process. Here, we are using process mining to help build theory about stability and change in routines. The contribution here goes beyond the specific findings about a particular dermatology clinic. The main contribution concerns the general idea of using network models to develop new theory about routine dynamics.

This approach to routine dynamics requires the analysis of network time-series data, as discussed in [25]. Here, we have limited our analysis to three time periods because we wanted to focus on the disruption of flu season. In a more elaborate model, we could extend the analysis to a longer time series. While increasingly sophisticated network time series models are available [26], they need to be carefully adapted to narrative networks. As of this writing, we are working on how to apply the *amen* R-package to multi-period time series analysis.

Even with the limited analysis we present here, there still some substantive insights. For example, the strongest predictor of edge persistence from one time period to the next is the frequency (weight) of that edge. In other words, routines tend to be repetitive. It would be disingenuous to count this well-established fact as a contribution, but it does lend face validity to the approach. The more interesting finding is that the magnitude of this effect is reduced by roughly 50% after a disruption. Likewise, the influence of betweenness is more than doubled after a disruption.

While these effects are interesting, the first key contribution here is not the specific magnitude of these effects. It is the overall approach to estimating them in the first place. This approach depends on using process mining to construct narrative networks of the process. While this class of network (a kind of directly follows graph, or DFG) has severe limitations as a process model [4], it provides a sensitive indicator of process stability and change. If the DFG is stationary, so is the process. Given these graphs, we are able to begin testing hypotheses about the factors that drive their stability and change using analytical tools like latent factor models [24].

The second key contribution here is theoretical. Hypotheses 1-4 represent a first attempt at defining formal, generalizable mechanisms that govern the dynamics of narrative networks. The analysis presented here is the first time that the effects of frequency, duration, betweenness or coherence on routine dynamics have been investigated empirically. These mechanisms may seem simplistic, but so do some of the key mechanisms that drive the dynamics of social networks ("birds of feather...", "the rich get richer...", "the friend of my friend..."). Perhaps simplicity is a virtue.

The last contribution of the study concerns the possible implications of this line of inquiry for process management and the BPM life cycle [27]. Routines are hard to change, as they tend to resist intervention and snap back into their old shapes. This study suggests that it may be possible to identify what aspects of routines contribute most to inertia and resistance.

7 Conclusion

Findings from a single case should be regarded as preliminary. We cannot build a generalizable theory from three weeks of data in one dermatology clinic. However, the findings do provide a variety of encouraging mysteries. For example, why does betweenness influence persistence at the fine-grained level but not the coarse-grained level? Where there is mystery, there is an opportunity for learning, so it seems that there may be more to learn from network models of routines.

References

1. Feldman, M.S., Pentland, B.T.: Reconceptualizing Organizational Routines as a Source of Flexibility and Change. Administrative Science Quarterly 48, 94-118 (2003)
2. Edmondson, A.C., Bohmer, R.M., Pisano, G.P.: Disrupted Routines: Team Learning and New Technology Implementation in Hospitals. Administrative Science Quarterly 46, 685-716 (2001)
3. Pentland, B.T., Recker, J., Wyner, G.: Rediscovering Handoffs. Academy of Management Discoveries 3, 284-301 (2017)

4. van der Aalst, W.M.: A Practitioner's Guide to Process Mining: Limitations of the Directly-Follows Graph. Procedia Computer Science 164, 321-328 (2019)
5. Hoff, P.D.: Bilinear Mixed-Effects Models for Dyadic Data. Journal of the American Statistical Association 100, 286-295 (2005)
6. Snijders, T.A.: The Statistical Evaluation of Social Network Dynamics. Sociological Methodology 31, 361-395 (2001)
7. Pentland, B.T., Vaast, E., Wolf, J.R.: Theorizing Process Dynamics with Directed Graphs: A Diachronic Analysis of Digital Trace Data. MIS Quarterly 45, 967-984s (2021)
8. Grisold, T., Wurm, B., Mendling, J., Vom Brocke, J.: Using process mining to support theorizing about change in organizations. In: the 53rd Hawaii International Conference on System Sciences. (Year)
9. Feldman, M.S., Pentland, B.T., D'Adderio, L., Lazaric, N.: Beyond Routines as Things: Introduction to the Special Issue on Routine Dynamics. pp. 505-513. INFORMS (2016)
10. Gilbert, C.G.: Unbundling the Structure of Inertia: Resource versus Routine Rigidity. Academy of Management Journal 48, 741-763 (2005)
11. Becker, M.C., Lazaric, N., Nelson, R.R., Winter, S.G.: Applying Organizational Routines in Understanding Organizational Change. Industrial and Corporate Change 14, 775-791 (2005)
12. Howard-Grenville, J.A.: The Persistence of Flexible Organizational Routines: The Role of Agency and Organizational Context. Organization Science 16, 618-636 (2005)
13. Birnholtz, J.P., Cohen, M.D., Hoch, S.V.: Organizational Character: On the Regeneration of Camp Poplar Grove. Organization Science 18, 315-332 (2007)
14. Grote, G., Weichbrodt, J.C., Günter, H., Zala-Mezö, E., Künzle, B.: Coordination in High-Risk Organizations: the Need for Flexible Routines. Cognition, Technology & Work 11, 17-27 (2009)
15. Schulz, M.: Staying on Track: a Voyage to the Internal Mechanisms of Routine Reproduction. In: Becker, M.C. (ed.) Handbook of Organizational Routines, pp. 228-257. Edward Elgar, Cheltenham (2008)
16. Cohen, M.D., Burkhart, R., Dosi, G., Egidi, M., Marengo, L., Warglien, M., Winter, S.: Routines and Other Recurring Action Patterns of Organizations: Contemporary Research Issues. Industrial and Corporate Change 5, 653-698 (1996)
17. Pentland, B.T., Kim, I.: Narrative Networks in Routine Dynamics. In: Feldman, M.S., Pentland, B.T., D'Adderio, L., Dittrich, D., Rerup, C., Seidl, D. (eds.) Cambridge Handbook of Routine Dynamics. Cambridge University Press (2021)
18. Steglich, C., Snijders, T.A., Pearson, M.: Dynamic Networks and Behavior: Separating Selection from Influence. Sociological Methodology 40, 329-393 (2010)
19. Kremser, W., Schreyögg, G.: The dynamics of interrelated routines: Introducing the cluster level. Organization Science 27, 698-721 (2016)
20. Pentland, B.T., Rueter, H.H.: Organizational Routines as Grammars of Action. Administrative Science Quarterly 39, 484-510 (1994)
21. Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proceedings of the national academy of sciences 99, 7821-7826 (2002)

22. White, D.R., Borgatti, S.P.: Betweenness Centrality Measures for Directed Graphs. Social Networks 16, 335-346 (1994)
23. Hoff, P.D., Raftery, A.E., Handcock, M.S.: Latent Space Approaches to Social Network Analysis. Journal of the American Statistical Association 97, 1090-1098 (2002)
24. Hoff, P.D.: Multiplicative Latent Factor Models for Description and Prediction of Social Networks. Computational and Mathematical Organization Theory 15, 261-272 (2009)
25. Pentland, B.T., Recker, J., Wolf, J.R., Wyner, G.: Bringing Context inside Process Research with Digital Trace Data. Journal of the Association for Information Systems 21, 5 (2020)
26. Rossetti, G., Cazabet, R.: Community Discovery in Dynamic Networks: a Survey. ACM Computing Surveys (CSUR) 51, 1-37 (2018)
27. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer (2018)