Understanding Process Dynamics through Trace Variant Rank-Frequency Curves

Lennart Ebert[1], Kate Revoredo[1], Jan Mendling[1,2,3], and Benoît Depaire[4]

[1] Humboldt University of Berlin, Berlin, Germany
{lennart.ebert,kate.revoredo,jan.mendling}@hu-berlin.de
[2] Vienna University of Economics and Business, Vienna, Austria
[3] Weizenbaum Institute, Berlin, Germany
[4] Digital Future Lab, Hasselt University, Hasselt, Belgium
benoit.depaire@uhasselt.be

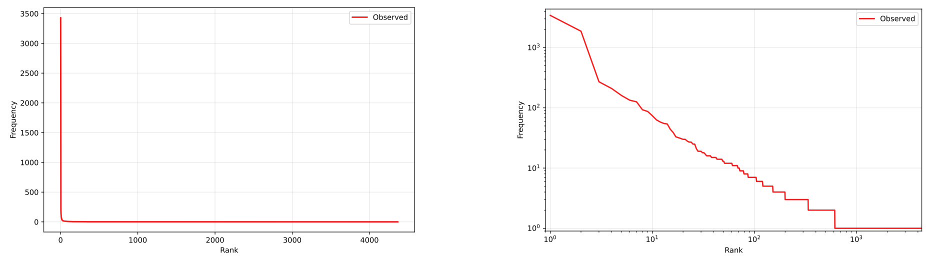
Abstract. Business processes typically follow a behavioral pattern in which a small number of variants occur very frequently, while many others occur only rarely. In this paper, we investigate variant-frequency distributions of real-world business processes and explore how their structural patterns can contribute to understanding process dynamics. We do so by analyzing trace variant rank-frequency curves (RFCs), which order variants by decreasing occurrence frequency. Using real-world event logs, we observe recurring patterns, including happy paths, power-law-like decay, and long tails of single-occurrence variants. To investigate potential causes of these patterns, we complement the empirical analysis with controlled simulations that test for the influence of process structure, process composition, and process change on RFCs. We find that even simple static process models may result in power-law-like behavior and that process change influences RFC shapes. Based on these findings, we propose potential mechanisms underlying observed RFC patterns. Our results suggest that trace variant RFCs provide a promising perspective for studying process dynamics.

Keywords: Process Dynamics · Network Theory · Process Mining

1 Introduction

Process dynamics research aims to explain stability and change in business processes [7]. Event logs, which record as-is process executions [11], form a data-centric basis for studying these phenomena. Analyses of these logs can identify stable periods as well as manifestations of change, such as concept drift [14] and workarounds [3].

Event logs show that process traces are often concentrated in a small number of frequently occurring trace variants (i.e., unique sequences of activities), while many variants form a long tail of infrequent or exceptional behavior [12]. Existing process mining approaches analyze and compare individual variants [10],



(a) Linear (b) Log-log

Fig. 1: Trace variant RFCs for BPIC12 (loan application process) log in linear and log-log scales

compute complexity scores based on variant distributions [1], and use rare variants to estimate unseen process behavior [5]. Yet, the overall distribution of trace variant frequencies has received limited attention, even though it reflects recurrence and variation, which are central to understanding process dynamics.

Research in statistical linguistics, network science and other fields has shown that frequency distributions can point to underlying mechanisms [15,2,6]. In linguistics, for example, the distribution of word frequencies follows a power law, a pattern that has been associated with the principle of least effort [15]. In network science, power-law distributions can result from preferential attachment during network growth [2]. Similarly, frequency distributions of trace variants may provide insights into underlying process mechanisms related to process structure, composition, and change.

Motivated by the possibility that frequency distributions point to underlying mechanisms, we explore trace variant rank-frequency curves (RFCs) of real-world event logs and investigate the patterns generated by potential underlying mechanisms. Trace variant RFCs plot the occurrence frequency of each variant against its rank after ordering variants by decreasing frequency. While established process dynamics analysis approaches, such as change point detection and analyses of complexity over time, focus on temporal changes, RFCs provide a graphical representation of how behavioral variety is distributed across an entire log. Figure 1 illustrates an RFC of a real-world event log using linear and log-log scales. The linear plot shows strong concentration of behavior in a small number of variants, while the log-log plot visualizes the rate of frequency decay for lower-frequency variants. After the two most frequent variants, the curve is approximately linear in log-log space, a pattern indicative of a power-law distribution. This observation raises the question of why such a distribution may have emerged. To explore this perspective further, we pose the following research questions:

RQ1: What recurring patterns characterize empirically observed trace variant-frequency distributions?

RQ2: Which underlying process mechanisms may contribute to the emergence of these patterns?

We contribute to process dynamics research in three ways. First, we analyze RFCs from a broad set of empirical event logs and identify recurring patterns. Second, through process simulation, we demonstrate how selected process mechanisms can generate or modify these patterns. Third, drawing on the empirical and simulation results, we develop an interpretation framework for RFCs and propose explanations for the emergence of power-law-like behavior. Overall, we introduce RFCs as an analytical lens for characterizing process behavior and reasoning about the mechanisms that may shape it.

The paper is structured as follows. Section 2 formalizes trace variant RFCs and reviews related work. Section 3 describes the methodology used to analyze real-world RFCs and conduct simulations, with the corresponding results presented in Section 4. Section 5 discusses the implications of the findings, and Section 6 concludes the paper.

2 Background and related work

2.1 Trace variant rank-frequency curves

Trace variant RFCs can be directly computed from simple event logs. Following van der Aalst [11], we define simple event logs as follows.

Definition 1 (Simple event log). *Let $\mathcal{A}$ be a set of activity names and let $\mathcal{A}^*$ denote the set of all finite activity sequences over $\mathcal{A}$. A trace is an element $\sigma \in \mathcal{A}^*$. A simple event log is defined as a multiset of traces, i.e., $L \in \mathcal{B}(\mathcal{A}^*)$.*

Consider the activities a, b, and c. A possible simple event log over these activities is $[\langle a,b,c\rangle^2,\ \langle b,c\rangle, \langle a,a,c\rangle^4,\ \langle a,a\rangle]$. This multiset representation already aggregates traces with identical activity sequences into trace variants. Moreover, it directly captures the multiplicity of each variant, e.g., the variant $\langle a,a,c\rangle$ occurs four times. Based on a simple event log, a trace variant rank-frequency curve can be defined as follows.

Definition 2 (Trace variant rank-frequency curve). *Let $L \in \mathcal{B}(\mathcal{A}^*)$ be a simple event log. The trace variant rank-frequency curve of L is a function f_L defined for ranks between 1 and the number of trace variants in L. For each such rank r, $f_L(r) \in \mathbb{N}$ denotes the frequency of the trace variant at rank r after sorting all trace variants according to decreasing multiplicity in L. Equal frequencies occupy consecutive ranks in arbitrary order.*

The most frequent trace variant occupies rank one in the rank-frequency curve. Continuing the example from before, $f_L(1)$ corresponds to the frequency of the trace variant $\langle a,a,c\rangle$, which is four. Likewise, $f_L(2) = 2$, as the second most frequent variant occurs twice. Finally, both $f_L(3)$ and $f_L(4)$ evaluate to 1, since the two remaining variants each occur once.

Various distributions across economics, statistical linguistics and network science frequently appear to follow power laws [6]. Power-law functions have the form

$$f(x) = Cx^{-\alpha}, \tag{1}$$

where α is the power-law exponent and C is a normalization constant. In linear-scale visualizations, these functions are difficult to interpret due to their rapid rate of decline. However, when both axes are represented on logarithmic scales, power-law functions become linear with slope $-\alpha$. Nevertheless, an approximately linear appearance in log-log scale graphs does not necessarily imply that an RFC follows a power-law distribution [6].

2.2 Related work

This paper primarily relates to three research streams: (i) process variant analysis, (ii) process complexity, and (iii) process species estimation.

In the first stream, process instances are grouped into variants based on shared characteristics, such as organizational units or control-flow behavior [9,10]. In practice, variants are most commonly defined by their control flow [10]. Analysts then compare variants to identify process improvement opportunities and gain insights into process behavior [10]. However, this stream of research generally focuses on the identification and comparison of variants rather than on the overall distribution of trace variant frequencies within a process.

In the second stream, the focus has been on measures that summarize event logs as scalar values [1,13]. Examples include the *trace variant count* [4] and *variant entropy* [1]. Variant entropy quantifies how evenly process executions are distributed across variants by computing the entropy of the trace variant-frequency distribution. While these measures capture aspects of behavioral variability, they aggregate the trace variant-frequency distribution into a single value and therefore lose distributional detail.

In the third stream, the focus is on estimating the amount of unseen process behavior from observed behavior represented through event logs [5]. To estimate the number of unseen variants, these methods rely on the frequencies of rare variants, particularly single-occurring and twice-occurring variants. Consequently, they focus primarily on the tail of the trace variant-frequency distribution rather than its overall shape.

In summary, existing research has investigated process variants, process complexity, and rare process behavior. However, comparatively little attention has been paid to the overall shape of trace variant-frequency distributions. This paper addresses this gap through trace variant rank-frequency curves (RFCs), which provide a compact representation of how process executions are distributed across variants and enable the analysis of the overall shape of trace variant-frequency distributions.

3 Method

To address the research questions, we perform exploratory analyses of real-world and simulated event logs. The analysis of real-world event logs addresses RQ1 by identifying recurring patterns in empirical trace variant RFCs. The simulation experiments support answering RQ2 through controlled tests of how process

Table 1: Overview of analyzed event logs.

Log(s)	Description	|C|	|E|	|TV|	|A|
ACCRE	Academic credentials	398	1.9k	65	16
BPIC11	Hospital treatment (gynaecology)	1.1k	150.3k	981	624
BPIC12	Loan applications	13.1k	262.2k	4.4k	36
BPIC13 (2 logs)	IT incidents and problems	1.5k, 7.6k	6.7k, 65.5k	327, 2.3k	7, 13
BPIC14	IT incidents	46.6k	466.7k	23.3k	39
BPIC15 (5 logs)	Municipality building permits	832-1.2k	44.4k-59.7k	828-1.3k	356-410
BPIC17	Loan applications	31.5k	1.2M	15.9k	26
BPIC18	Agriculture subsidies	43.8k	2.5M	28.5k	41
BPIC19	Purchase-to-pay	251.7k	1.6M	12.0k	42
BPIC20 (5 logs)	Travel permits	2.1k-10.5k	18.2-72.2k	89-1.5k	17-51
CALL	Call center	260.9k	445.6k	1.7k	19
ITHD	IT help desk	4.6k	21.3k	226	14
MOBIS	Travel management	6.6k	83.3k	1.1k	26
RTFMP	Road traffic fine management	150.4k	561.5k	231	11
SEPSIS	Hospital treatment (sepsis)	1.1k	15.2k	846	16

$|C|$: # Cases; $|E|$: # Events; $|TV|$: # Trace variants; $|A|$: # Activities.

characteristics affect RFC shapes. The full results and analysis scripts are publicly available under a CC-BY-4.0 license.[5] We further provide an interactive notebook to perform additional experiments directly from a web browser.[6]

For the empirical analysis, we selected 24 publicly available event logs covering a diverse range of domains, process sizes, and levels of behavioral variability (see Table 1). For each log, we constructed the RFC and plotted it on a log-log scale. We then examined the resulting curves visually and quantitatively compared them to several simple statistical models. Specifically, we fitted linear, exponential, power-law, and bounded power-law models using negative log-likelihood minimization. The bounded power-law model is a power-law model only fitted to frequencies greater than one to reduce the influence of single-occurrence variants. The objective of this analysis is not to provide mathematical proof of the underlying distribution, but to characterize recurring patterns and assess whether power-law-like behavior frequently occurs in practice.

To explore potential explanations for the observed patterns, we conducted a set of simulation experiments. The simulations were designed to isolate three classes of process mechanisms, i.e., process structure, process composition, and process change. Rather than reproducing specific real-world processes, the simulations use a parameterized process model that systematically varies key control-flow characteristics such as exclusive choices, concurrency, loops, and branching probabilities. This design allows mechanisms to be studied in isolation while maintaining comparability across experiments. Table 2 defines the list of simulation parameters, Figure 2 depicts an example process model, and Table 3

[5] `https://doi.org/10.5281/zenodo.21456320`

[6] `https://mybinder.org/v2/gh/lennartebert/process-mining-rfcs/HEAD?urlpath=%2Fdoc%2Ftree%2Finteractive_simulation_experiments.ipynb`

Table 2: Simulation parameters

Category	Parameter	Description
General	Num. cases (N)	Number of simulated process instances
Control flow	Num. parallel branches (m)	Number of branches at AND gate
	Num. XOR gates (n)	Number of XOR gates per parallel branch
	Choices per XOR gate (o)	Number of branches per XOR gate
	Loop probability (p)	Probability of loop to start
	XOR probability distribution (q)	Distribution of XOR branch probabilities
	Power-law exponent (α)	Power-law exponent when $q = power\_law$
Drift	Overlapping activity sets (s)	Whether the activity sets for processes a and b are *overlapping* or *non-overlapping*.

Options for q: *equal* (uniform), *majority* (one option 80%, remainder equally distributed), or *power_law* (power law with exponent α)

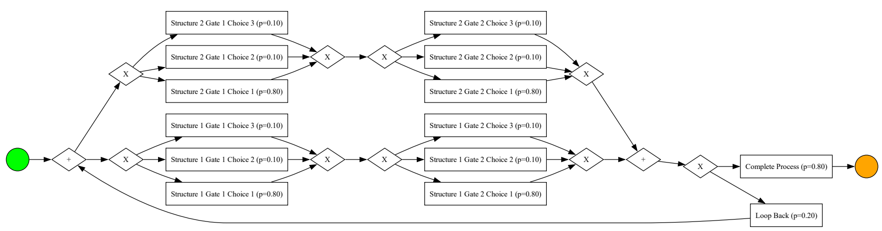


Fig. 2: Example business process model generated from simulation parameters

summarizes the 16 experiments. These simulation experiments do not aim to reproduce empirical process structure directly, but to isolate different mechanisms.

4 Results

4.1 Empirical RFCs from real-world logs

Figure 3 presents the empirical RFCs and model fits for the power-law and bounded power-law models. Four recurring patterns can be observed, namely (1) *power-law-like slopes*, (2) *head deviations*, (3) *long tails of rare variants*, and (4) *variation in slope*. Most RFCs exhibit approximately linear decay in log-log space over multiple orders of magnitude and closely follow the fitted power-law or bounded power-law models. This observation is supported by the model fitting results, for which one of the two power-law-based models achieves the lowest NLL for all logs (see repository for details). At the same time, several RFCs deviate from ideal power-law behavior in the head of the distribution, i.e., the leftmost ranks. Examples include CALL, BPIC12, and RTFMP, whose RFCs exhibit flattened heads, indicating the presence of multiple similarly frequent variants and potentially coexisting happy paths.

Table 3: Overview of simulation experiments.

				Process a							Process b					
Scenario	ID	Experiment name	N	m	n	o	p	q	α	s	m	n	o	p	q	α
Simple	a1	Single equal XOR	10,000	1	1	10	0.0	E								
	a2	Single majority XOR	10,000	1	1	10	0.0	M								
	a3	Single power law XOR	10,000	1	1	10	0.0	PL	2.0							
	a4	Multi-equal XORs	10,000	1	6	10	0.0	E								
	a5	Multi-majority XORs	10,000	1	6	10	0.0	M								
	a6	Multi-power law XORs	10,000	1	6	10	0.0	PL	2.0							
	a7	Single majority XOR, loop	10,000	1	1	10	0.2	M								
	a8	Single power law XOR, loop	10,000	1	1	10	0.2	PL	2.0							
	a9	Parallel, majority XOR	10,000	2	1	10	0.0	M								
	a10	Parallel, power law XOR	10,000	2	1	10	0.0	PL	2.0							
	a11	Complex majority XORs	10,000	2	6	10	0.2	M								
Composition	b1	Composition, α constant	10,000	1	1	100	0.0	PL	1.0		1	1	100	0.0	PL	1.0
	b2	Composition, α change	10,000	1	1	100	0.0	PL	1.0		1	1	100	0.0	PL	3.0
Change	c1	Non-overlapping change	10,000	1	1	100	0.0	PL	2.0	NO	1	1	100	0.0	PL	2.0
	c2	α change, overlapping	10,000	1	1	100	0.0	PL	1.0	O	1	1	100	0.0	PL	3.0
	c3	XOR change, overlapping	10,000	1	2	10	0.0	M		O	1	3	10	0.0	M	

q: E = equal, M = majority, PL = power-law; s: NO = non-overlapping, O = overlapping.

The remaining two patterns concern the tails and slopes of the RFCs. Several processes, including BPIC11, BPIC15, and SEPSIS, contain large numbers of single-occurrence variants that form plateauing tails in the lower-right part of the RFCs. These rare variants also explain why the bounded power-law model often provides a better fit than the unbounded power-law model. In addition, power-law exponents vary considerably across logs, with bounded power-law exponents ranging from 0.00 (BPIC15_2 and BPIC15_5) to 2.33 (RTFMP), indicating substantial differences in the concentration of process behavior across domains.

4.2 Simulation experiments

Figure 4 presents the RFCs and fitted power-law models for the simulation experiments. The results provide initial observations regarding how process characteristics influence RFC shapes. Notably, RFCs most closely resemble power-law distributions when XOR choice probabilities are themselves power-law distributed. In this case, a single XOR gateway is sufficient to generate a power-law distribution of variant frequencies (see Figure 4 (a3)). By contrast, equal and majority-distributed routing probabilities produce characteristic plateaus, which appear as distinct steps when multiple branches receive identical probabilities (e.g., Figure 4 (a5)). Furthermore, process loops increase the number of possible variants and reduce the fitted power-law exponent α (see Figure 4 (a8)).

Additional observations concern the effects of process composition and process change. Combining processes may alter the overall RFC slope (see Figure 4 (b1)), and introduce visible kinks if the α exponent changes (see Figure 4 (b2)).

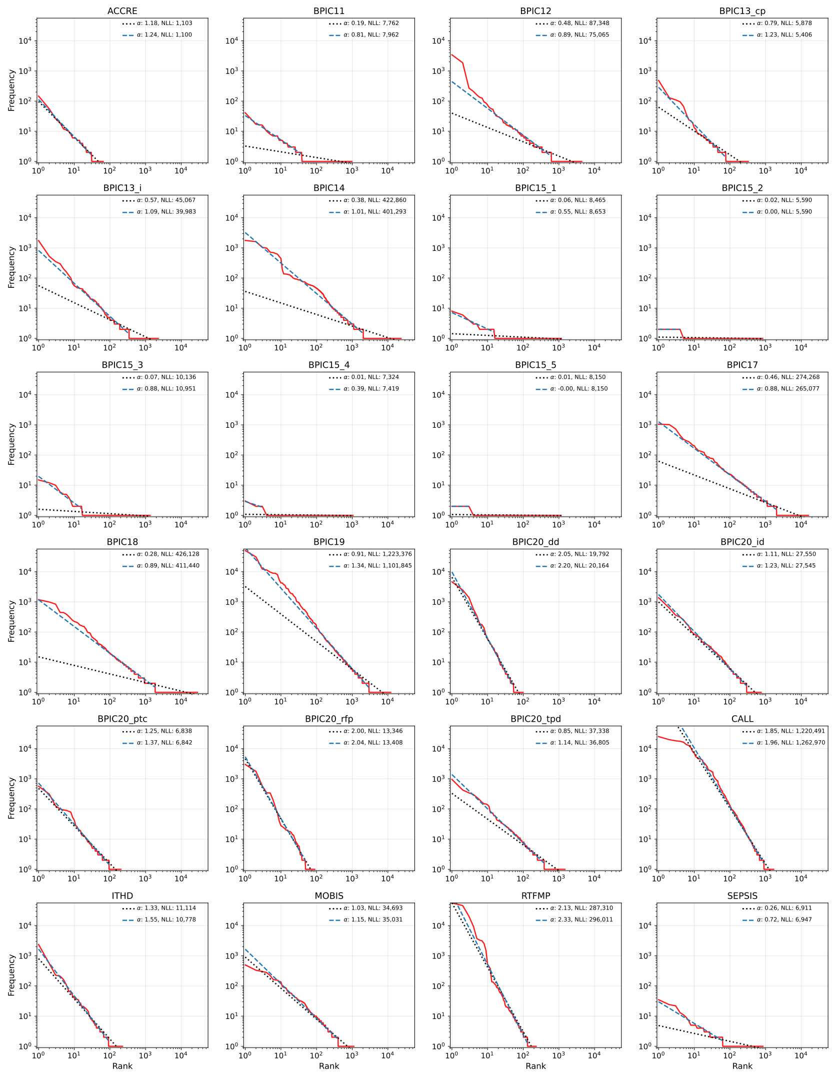
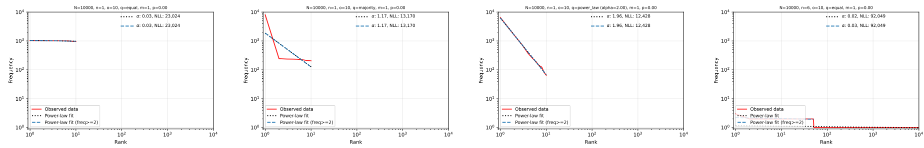
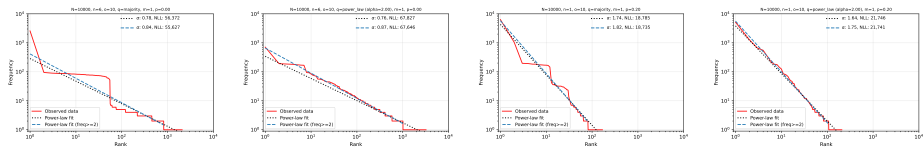


Fig. 3: Trace variant rank-frequency RFCs of 24 real-world event logs

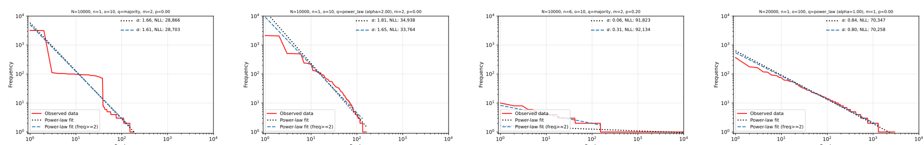
Process drift, in turn, may affect the RFC slope (see Figure 4 (c1) and (c2)), increase the frequency of the highest-ranked variants relative to a fitted power-law distribution (see Figure 4 (c2)), and affect plateaus (see Figure 4 (c3)). Together, these observations suggest that both static and dynamic process mechanisms are reflected in RFC shapes.



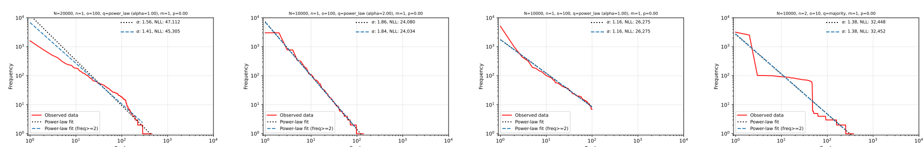
(a1) Single equal XOR (a2) Single majority XOR (a3) Single power law XOR (a4) Multi-equal XORs



(a5) Multi-majority XORs (a6) Multi-power law XORs (a7) Single majority XOR, loop (a8) Single power law XOR, loop



(a9) Parallel, majority XOR (a10) Parallel, power law XOR (a11) Complex majority XORs (b1) Composition, α constant



(b2) Composition, α change (c1) Non-overlapping change (c2) α change, overlapping (c3) XOR change, overlapping

Fig. 4: Simulation experiment RFCs: simple process structure (a1–a11), process composition (b1–b2), and process change (c1–c3).

5 Discussion

Empirical RFCs and simulation results suggest that process dynamics leave identifiable imprints on trace variant RFCs. Consequently, RFCs may provide a useful representation for studying process dynamics. To support such analyses, Table 4 proposes a mapping between RFC characteristics and potential underlying mechanisms. RFCs can be characterized by their slope, the head of the distribution (i.e., low-rank, high-frequency behavior), the tail of the distribution (i.e., high-rank, low-frequency behavior), and deviations from power-law shapes (i.e., non-linear decay in log-log space). These characteristics may indicate mechanisms related to process standardization, flexibility, and exceptional behavior.

Table 4: Interpreting rank-frequency curves (RFCs) in process analysis.

Characteristic	Observation rationale	Potential interpretation
Slope	What is the exponent α of a fitted power-law distribution?	Higher α values may reflect greater process standardization, whereas lower values may indicate more diverse process behavior and flexibility.
Head	Does the head of the RFC deviate from the slope observed in the remainder of the distribution?	Low-rank variants whose frequency exceeds that of the fitted power-law may indicate highly standardized happy paths.
Tail	Is there a long tail of single-occurring trace variants?	Long tails may reflect exceptions, flexibility, or incomplete observations of the process.
Divergence from power law	Does the RFC bend downward or upward relative to a fitted power-law distribution?	Deviations may be associated with process change, process composition or concurrency.

For example, the RFC of the BPIC12 log (see Figure 3) exhibits a comparatively moderate power-law exponent ($\alpha = 0.89$), suggesting a moderate degree of process standardization. Its flattened head, where the two most frequent variants exceed the fitted power law, may indicate two dominant happy paths, while the tail of single-occurrence variants may reflect exceptional executions or process flexibility. Finally, apart from the head and tail, the RFC closely follows the fitted bounded power law, providing no clear indication of process composition or change based on the RFC alone. Together, these observations provide an initial characterization of the process, while complementary analyses are required to validate these interpretations.

Despite the ability of the simulation experiments to reproduce power-law-like RFCs, the mechanisms underlying such behavior require further investigation. The experiments showed that even simple process models containing a single XOR gateway can generate power-law-like RFCs, provided that the branching probabilities themselves follow a power-law distribution. However, such a process model is unlikely to reflect real-world processes and does not explain why gateway probabilities should be power-law distributed. This experiment should therefore be interpreted as a proof of concept rather than a realistic generative model. Nevertheless, it provides initial evidence that power-law-like RFCs may emerge from static process structures.

In addition to process structure, several other mechanisms may contribute to the emergence of power-law-like RFCs, including *case heterogeneity*, *endogenous heterogeneity*, and *reinforcement*. Case heterogeneity implies that the shape of an RFC may be inherited from the distribution of process inputs. Consider a car manufacturing process in which vehicles differ by color. If the process invokes different coating activities depending on a vehicle's color and the distribution of colors itself follows a power-law distribution, the resulting trace variant-frequency distribution may also exhibit power-law-like behavior. More generally, case context may induce different activity sequences, thereby shaping the resulting RFC. Endogenous heterogeneity refers to the influence of organizational factors on variant frequencies. Resource availability, workload, personal preferences, or organizational policies may systematically favor specific execution paths, contributing to increasingly skewed RFCs. Finally, reinforcement describes that the frequent enactment of process paths makes them more likely to be enacted again in the future [8]. This is comparable to preferential attachment in social networks, where new nodes are more likely to connect with already well-connected nodes [2]. Consequently, reinforcement may also contribute to power-law-shaped RFCs.

Our study has several limitations. First, because the objective was to identify recurring RFC patterns across a broad set of real-world processes, the analysis prioritized breadth over depth. Future work could examine specific RFC shapes in greater detail and empirically evaluate the interpretation guide proposed in this paper. Second, the simulation experiments were based on a single parameterized process model. While this design enabled the isolation of individual mechanisms such as XOR probability distributions, loops, process composition, and process change, additional simulations based on alternative process structures would strengthen the generalizability of the findings. For example, future studies could investigate process models containing multiple interacting loops or more complex control-flow constructs. Third, further research is required to investigate alternative explanations for the emergence of power-law-like RFCs. Case heterogeneity, endogenous heterogeneity, and reinforcement may represent additional mechanisms underlying the observed power-law-like shapes. Exploring these mechanisms presents a promising direction for future research and may contribute to a better understanding of process dynamics.

6 Conclusion

This study investigated patterns in the trace variant RFCs of real-world processes and conducted a simulation study to investigate potential underlying mechanisms. We found that RFCs of real-world processes frequently exhibit power-law-like behavior, alongside other recurring patterns. The simulation results demonstrated that power-law-like RFCs can emerge even from simple process models in which decision probabilities follow a power-law distribution. Furthermore, loops, concurrency, process composition, and process change were found to influence RFC shapes.

These findings suggest that RFCs offer a promising perspective for analyzing process dynamics. Characteristics such as the RFC's slope, head, tail, and divergence from a power-law distribution may indicate process standardization, flexibility, exceptions, and process change. However, the reasons why RFCs frequently exhibit power-law-like behavior require further investigation. By identifying recurring RFC patterns and proposing potential underlying mechanisms, we hope this study will encourage further inquiry into the relationship between trace variant-frequency distributions and process dynamics.

References

1. Augusto, A., Mendling, J., Vidgof, M., Wurm, B.: The connection between process complexity of event sequences and models discovered by process mining. Inf. Sci. **598**, 196–215 (2022)
2. Barabási, A.L., Albert, R.: Emergence of Scaling in Random Networks. Science **286**(5439), 509–512 (1999)
3. Beerepoot, I., van de Weerd, I., Reijers, H.A.: The Potential of Workarounds for Improving Processes. In: Di Francescomarino, C., Dijkman, R., Zdun, U. (eds.) Business Process Management Workshops: BPM 2019 International Workshops. LNBIP, vol. 362, pp. 338–350. Springer, Cham (2019)
4. Günther, C.: Process Mining in Flexible Environments. PhD thesis, Technische Universiteit Eindhoven, Eindhoven (2009)
5. Kabierski, M., Richter, M., Weidlich, M.: Quantifying and relating the completeness and diversity of process representations using species estimation. Inf. Syst. **130**, 102512 (2025)
6. Newman, M.: Power laws, Pareto distributions and Zipf's law. Contemporary Physics **46**(5), 323–351 (2005)
7. Pentland, B., Vaast, E., Wolf, J.R.: Theorizing Process Dynamics with Directed Graphs: A Diachronic Analysis of Digital Trace Data. MIS Quarterly **45**(2), 967–984 (2021)
8. Pentland, B.T., Liu, P., Kremser, W., Hærem, T.: The Dynamics of Drift in Digitized Processes. MIS Quarterly **44**(1), 19–47 (2020)
9. Rubensson, C., Mendling, J., Weidlich, M.: Variants of Variants: Context-Based Variant Analysis for Process Mining. In: Guizzardi, G., Santoro, F., Mouratidis, H., Soffer, P. (eds.) Advanced Information Systems Engineering, LNCS, vol. 14663, pp. 387–402. Springer, Cham (2024)
10. Taymouri, F., Rosa, M.L., Dumas, M., Maggi, F.M.: Business process variant analysis: Survey and classification. Knowledge-Based Systems **211**, 106557 (2021)
11. van der Aalst, W.M.P.: Process Mining: Data Science in Action. Springer, Berlin, Heidelberg, 2nd edn. (2016)
12. van der Aalst, W.: A practitioner's guide to process mining: Limitations of the directly-follows graph. Procedia Computer Science **164**, 321–328 (2019)
13. Vidgof, M., Wurm, B., Mendling, J.: The impact of process complexity on process performance: A study using event log data. In: Di Francescomarino, C., Burattin, A., Janiesch, C., Sadiq, S. (eds.) BPM 2023. LNCS, vol. 14159, pp. 413–429. Springer, Cham (2023)
14. Yeshchenko, A., Di Ciccio, C., Mendling, J., Polyvyanyy, A.: Visual Drift Detection for Event Sequence Data of Business Processes. IEEE Transactions on Visualization and Computer Graphics **28**(8), 3050–3068 (2022)
15. Zipf, G.K.: Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Addison-Wesley Press, Inc., Cambridge, Massachusetts (1949)

Acknowledgments. The research of the authors was supported by the Einstein Foundation Berlin under grant EPP-2019-524, by the German Federal Ministry of Research, Technology and Space under grant 16DII143, and by Deutsche Forschungsgemeinschaft under grants 496119880 (VisualMine), 531115272 (ProImpact), and SFB 1404/2 (FONDA). Kate Revoredo is funded by the Berliner Chancengleichheitsprogramm (BCP) as part of the DiGiTal Graduate Program.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.