Introducing the Stigmergic Miner, a temporal extension of the Fuzzy Miner based on computational stigmergy

Mario G.C.A. Cimino[1, *] [0000-0002-1031-1959], Francesco Hudema[1],
Sabrina Senatore[2][0000-0002-7127-4290], Francesco Lupi[1][0000-0003-4994-0215]
Pierfrancesco Foglia[1] [0000-0001-6432-4504] and Cosimo A. Prete[1] [0000-0002-8467-8198]

[1] Dept Information Engineering, University of Pisa, Italy
[2] Dept. Information and Electrical Engineering and Applied Math, University of Salerno, Italy
{m.cimino,f.lupi, p.foglia, a.prete}@ing.unipi.it
ssenatore@unisa.it f.hudema@studenti.unipi.it

Abstract. Workflow mining is a data mining research field for discovering graph models from event logs. Particularly, Fuzzy Mining is a well-known approach, suitable for noisy and unstructured logs. It includes in the model the fuzzy notion of significant/correlated edges and nodes, thus enabling the formation of granular graphs. This paper introduces the Stigmergic Miner, a novel workflow miner that incorporates the concept of time, as an extension of the Fuzzy Miner. It takes into consideration the temporal dynamics that exist within event occurrences. The proposed algorithm employs Computational Stigmergy, a bio-inspired paradigm employed in multi-agent systems. Stigmergy creates a representational space where each sample is transformed as a mark in the scalar space, whose intensity is evaporating over a discretized time. Overlapping marks create persistent stigmergic trails, generating scalar-temporal clustering of samples. With respect to the simply scalar fuzzy clusters, stigmergic clusters are then inherently temporal. Thanks to parametric optimization, different agents can be specialized to detect different temporal patterns, by computing similarity between stigmergic trail and stigmergic archetypes. Since each pattern is represented with a different color, the resulting graph is colored, to highlight different temporal patterns associated with event time series for different branches and nodes of the graph. To evaluate the effectiveness of this approach, early experiments with real-world data are reported, allowing for empirical validation of the proposed miner. An open-source web application has been developed and publicly released.

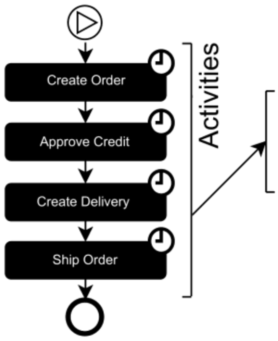
Keywords: Workflow mining, Fuzzy Miner, Computational Stigmergy, Multi-agent systems.

---

* Work partially supported by the Italian Ministry of Education and Research (MUR) in the framework of the FoReLab project (Departments of Excellence), and by the Italian Ministry of Enterprises and Made in Italy, in the framework of the "Agreements for Innovation" Projet "4DDS - 4D Drone Swarms" Ref. no. F/310097/01-04/X56.

1 Introduction and background

The notion of workflow in modeling is used when different entities, such as human or artificial agents, but also passive entities, such as services and devices, cooperate in open networks to achieve some results [1]. Workflows are at the heart of complex technologies in today's networked paradigm of production. In the form of orchestrations or choreographies, workflows coordinate or manage lifecycles of service/product factories, smart cities, cloud systems, multiagent systems, and so on. The challenge of workflow mining, well-known also as process mining, is to exploit event logs, largely available in modern information systems, to streamline workflows, especially those with human or artificial agents in the loop. For this purpose, workflow mining aims to bridge the gap between traditional knowledge-based workflow analysis and advanced data analysis methods like machine learning and data mining [2]. The assumption is that events can be recorded in a sequential manner, where each event corresponds to a specific activity within a defined set, and is associated with a particular case or workflow instance, together with other accompanying information. Fig. 1 shows an example of an essential event log: in the case ID 10, four activities have been carried out: Create order, Approve Credit, Create Delivery, Ship Order.



Event Log

CASE ID	ACTIVITY	TIMESTAMP
10	Create Order	02-23-2019 08:23:34
10	Approve Credit	02-23-2019 09:25:30
10	Create Delivery	02-23-2019 09:29:23
10	Ship Order	02-23-2019 10:12:45
21	Create Order	02-23-2019 18:53:14
22	Change Price	03-12-2019 08:23:34
23	Create Delivery	03-12-2019 08:23:34
24	Ship Order	03-12-2019 08:23:34

⋮

Fig. 1. An essential event log: Case ID, Activity and Timestamp

Different types of analysis, with their objectives and methods, are investigated by Workflow mining: (i) automated process discovery, i.e., mining workflow models from an event log; (ii) conformance checking, i.e., monitoring deviations of the real workflow by comparing a normative model and the log; (iii) model enhancement, i.e., model extension and repairing. Different quality measures are available for a generated workflow model, such as fitness, precision, generalization, simplicity [2]. The complexity of observed behaviors in real-world workflows has exposed a significant drawback in several workflow mining algorithms. When applied to mine logs from workflows characterized by minimal inherent structure, these algorithms tend to produce equally unstructured models that are challenging to interpret. Such models, often referred to as "spaghetti" models, fail to offer any meaningful abstraction from the event logs themselves. Analyzing the interesting aspects of the workflow becomes difficult, as it is challenging to determine the overall directions of workflow execution. To address this issue, the *Fuzzy Miner* algorithm was introduced by Gunther and van der Aalst in 2007 [3]. The Fuzzy Miner applies transformation methods to the process model: edge filtering, by removing edges between activity nodes, as well aggregation and abstraction, by removing clusters of nodes or less significant nodes. As a result, the Fuzzy Miner has the capability to construct hierarchical models in which less significant activities are organized into sub-workflows. When analyzing workflows, time perspective is extremely important to make decisions. In addition to the scalar dimension, i.e. the number of occurrences, the temporal dimension is fundamental in a workflow: the same number of events can be equally distributed for a long period, or be concentrated as a peak or a burst, assuming a completely different significance. In general, fuzzy relevance is not sufficient for considering time perspective. In the literature, Computational Stigmergy is well-known as a bio-inspired coordination mechanism between agents, especially in swarm robotics and swarm optimization. Since 2015, it has been introduced as a temporal mechanism for pattern recognition tasks in physiological/social time series [4][5], as well as in neurocomputing for perceptron weights and working memory [6][7]. Stigmergic clustering can be considered as a temporal extension of Fuzzy clustering, because data points in the same/different cluster are both spatially and temporally close/far. In this paper, computational stigmergy is included in the design and the development of a novel mining algorithm, called Stigmergic Miner, based on the Fuzzy Miner.

2 Materials and methods

Given, an event log L, a workflow mining algorithm is a function that maps L to a process model such that the model is 'representative' of the behavior observed in the event log [2]. Workflow mining is an optimization problem, as it should avoid overfitting (spaghetti or overcomplex models) and underfitting (oversimplified models). To prevent generation of spaghetti models, mining algorithms can employ clustering techniques. Clustering involves matching conditions, such as using syntactic or semantic similarity measures on events or task labels, workflow topology, workflow semantics, and so on. Clustering enables process abstraction, which simplifies a workflow by highlighting its most important aspects, like frequent activities, time spent, or resource requirements and costs. One useful technique is clustering sub-workflow variants, which means grouping similar sub-workflow versions together based on their shared characteristics. The Fuzzy Miner scores edges by combing their significance, based on activity frequency and correlation, which measure how closely activities are related. Using these metrics, it preserves highly significant behavior, aggregates less significant but correlated parts into clusters, and removes non-significant behavior. This filtering is controlled by node and edge cutoff parameters. Fig. 2 shows an example of workflow map generated using the BPI Challenge dataset [8], containing a collection of log events derived from the Volvo Company, where each case includes a series of steps taken by agents to resolve problems. Here, the numbers denote the significance of a node or arc. The Volvo dataset contains cases recorded from 2006 to 2012. The number of new cases has significantly increased in recent years, providing a more accurate representation of how the workflow handles the problems. Therefore, older cases may not be as relevant and could potentially lead to misleading workflow maps. To understand temporal aspects, let us introduce the notion of *token*, i.e., a marker that navigates in the graph. At a given instant of time, tokens represent the current activities being executed. Fig. 3 shows a snapshot of the flow of tokens for the Volvo dataset, made on Thursday 15 Sep 2011 at 19:51:46 with the mining software Disco [9]. Here, tokens are represented as circular shapes flowing into the graph according to the related timestamps. In the animation, when the tokens traffic is high, some "clouds" appear, to highlight that there is a significant number of tokens at the same time window. To categorize different temporal dynamics, different Stigmergic sets can be defined as archetypes [5]. Fig. 4 represents each of them as idealized time series, in a normalized scale of frequency: (a) *Dead* events or transitions, which are no longer active; (b) *Cold* events or transitions, which occur sporadically with spikes or fluctuations; (c) *Falling* events or transitions, with a decreasing trend, denoting a decreasing significance; (d) *Rising* events or transitions, with an increasing trend, denoting an increasing significance; (e) *Hot* events or transitions, which consistently maintain high values. Toassign a continuous degree of similarity to each archetype, a Stigmergic perceptron is used [5]. In brief, a stigmergic perceptron is a neural architecture made of Stigmergic Receptive Fields (SRFs). SRFs are computational agents, each detecting a temporal data pattern in event logs, whose design is based on computational stigmergy. An SRF, for each sample of a given time series, leaves a mark in the scalar space of event occurrences. SRFs are mutually adaptive, thanks to parametric optimization, to enable a kind of spatial-temporal clustering in the stigmergic space. To take the time into consideration, single marks evaporate over time. Therefore, overlapping marks create a more persistent stigmergic trail. A stigmergic trail is then a suitable representative of the time series in a time window. SRFs assess the similarity between the input time series and archetypal patterns. In the proposed Stigmergic Miner, a set of archetypes are selected to observe different dynamics (e.g. rising, falling, intermittent dynamics). The significance of an event or transition is determined by its similarity with respect to the available archetypes. By assigning a different color to each archetype, and by graduating the intensity of such colors according to its significance, a multi-color workflow map is finally generated. Fig. 5 shows the workflow map generated by the Stigmergic Miner on the Volvo dataset for the "Rising" dynamic. Here, events or transitions that do not exhibit a Rising behavior are not represented in the process map. Events or transitions closely matching the archetype are highlighted in red. Choosing a different archetype within the Stigmergic Miner framework highlights different behaviors that are available in the event log.

3 Early experimental studies

A complete web application has been developed as open-source project and publicly released to foster academic and industrial collaborations [10]. This section summarizes an early experiment (refer to [10][11] for additional information). The goal is to generate a behavioral process map for a complex problem where different temporal behaviors are present in the log. The log data are taken from the Blockchain-based Shared Additive Manufacturing (BBSAM) workflow [11]. It is a collaborative protocol for agents distributing resources in additive manufacturing. The experiment highlights behavioral patterns manifested by the workflow when the demand for printing orders fluctuates over time. Fig. 6 shows the workflow map generated by the Stigmergic Miner, assigning a different color to each pattern. For better interpretability, the interested reader is referred to as the BPMN (Business Process Model and Notation) diagram of the workflow [11]. More specifically, the workflow can be divided into three main production cases: orders managed internally by the agent receiving the order, orders managed by the local pool of agents, and orders managed by the external pool of agents. Specifically, events related to order printing and shipping are the most time-consuming, as they are influenced by factors such as product type, quantity, and shipping time. The events leading up to the tasks "check pool order feasibility" and "schedule & print item" manage incoming orders. These events are classified as Falling transitions since they cease to appear in the log once all orders are sent for processing by local or external machines.

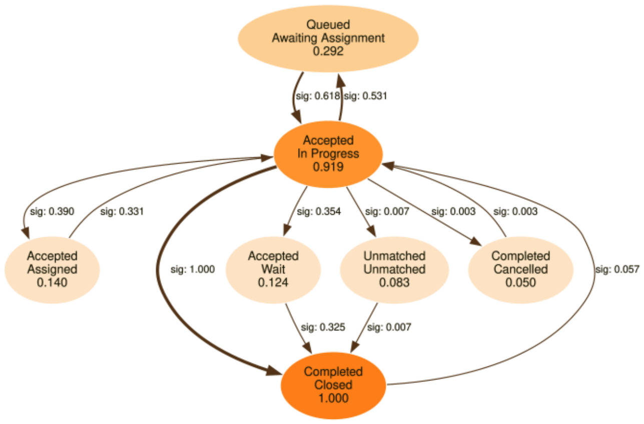


Fig. 2. A workflow map generated by the Fuzzy Miner on the Volvo Dataset [8]

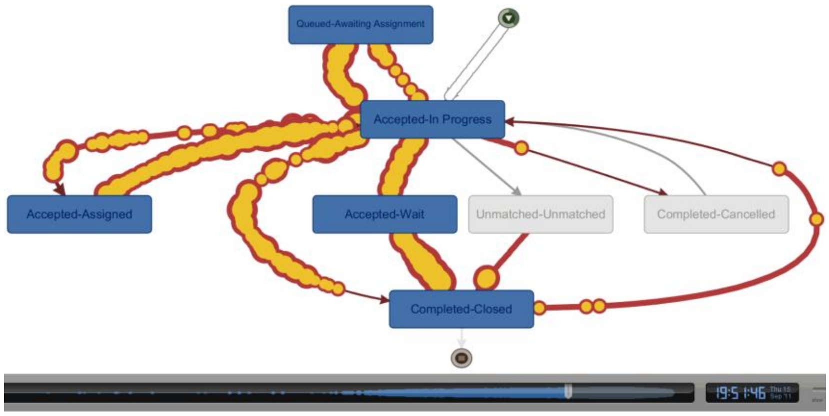


Fig. 3. A snapshot of an animation of the flow of tokens, generated by Disco [9]

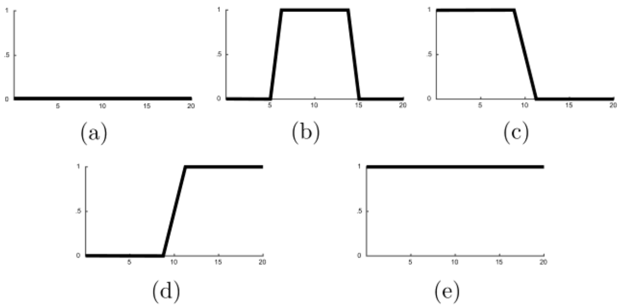


Fig. 4. Stigmergic Sets, representing temporal patterns of events or transitions: Dead (a), Cold (b), Falling (c), Rising (d), and Hot (e).

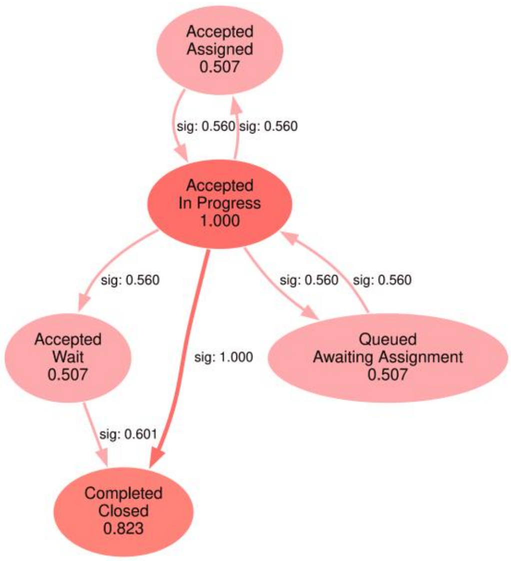


Fig. 5. A workflow map generated by the Stigmergic Miner on the Volvo Dataset [8]

On the bottom side of the diagram, variations in how shipments are handled are visible. *Orders managed internally*: the internal shipment process, in blue color, is associated with a Cold transition because it starts after orders are printed and terminates before the end of the log. O*rders managed by the local pool of agents*: the local pool shipment process, in green color, is associated with a Hot transition since it starts after orders are printed; this event starts almost initially and terminates at the end of the log; it is because some orders remain in the shipment queue and are the last to be processed. *Orders managed by the external pool*, in red color, exhibit Dead transition, because the number of orders rarely necessitated the use of the external pool; indeed, resources in the local pool were sufficient almost all the time to handle the surplus. *Order closing procedure*, in yellow color, are events for both internal shipment and local pool shipment; these events begin after the shipment phase and conclude at the end of the overall process, consequently, they are associated with a Rising transition.

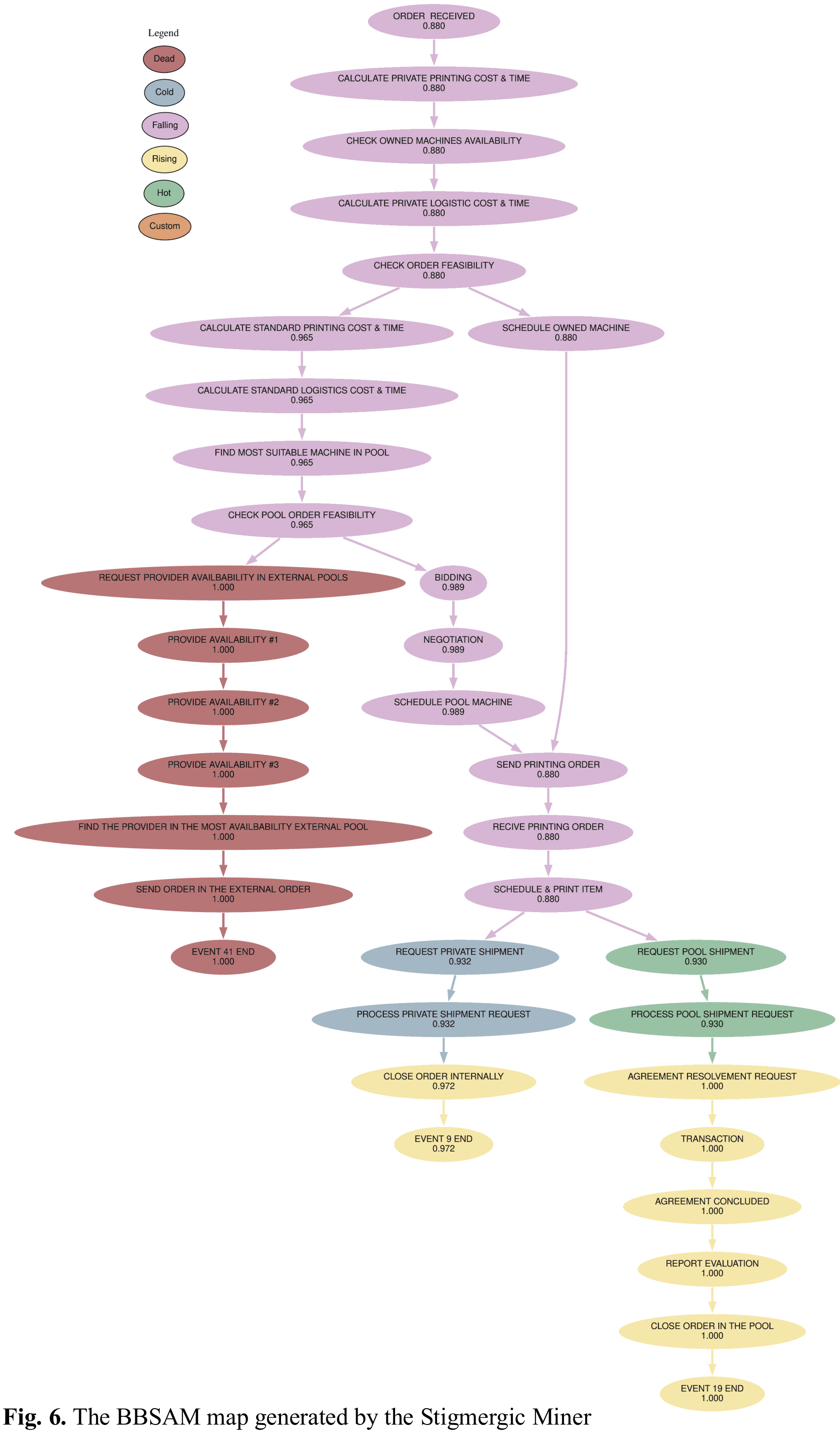


Fig. 6. The BBSAM map generated by the Stigmergic Miner

4 Conclusions

The paper presents the Stigmergic Miner, a novel mining algorithm as an extension of the Fuzzy Miner that integrates temporal data into workflow mining. This approach enhances the representation of complex workflows, making them more understandable and informative. Some examples of generated temporal workflow maps are discussed and interpretated, extracting practical insights from them, showing that they are not just theoretical constructions but valuable tools for decision-making. To the best of our knowledge, this study introduces an original framework to harness the potential of stigmergy in process mining. This area has remained relatively unexplored in prior research. While existing work in computational intelligence explores the use of stigmergy for coordinating swarm and multi-agent systems, the exploration of stigmergy in pattern recognition, data classification, and workflow mining represents a promising frontier in the field.

References

1. M.G.C.A. Cimino, F. Palumbo, G. Vaglini, E. Ferro, N. Celandroni, D. La Rosa, "Evaluating the impact of smart technologies on harbor's logistics via BPMN modeling and simulation", Information Technology & Management, 18:223-239 (2017).
2. W. Van Der Aalst, Data science in action. Springer, 2016.
3. C. W. Gunther and W. M. Van Der Aalst, “Fuzzy mining–adaptive process simplification based on multi-perspective metrics,” Proceedings of the International conference on Business Process Management, pp. 328–343, Springer, 2007.
4. M.G.C.A Cimino, A. Lazzeri, G. Vaglini, "Improving the analysis of context-aware information via marker-based stigmergy and differential evolution", Proc. 14th Int. Conf. ICAISC 2015, Zakopane, Poland, June 14-18, 2015, Part II, LNAI 9120:341-352 (2015).
5. M. Avvenuti, C. Bernardeschi, M.G.C.A. Cimino, G. Cola, A. Domenici, G. Vaglini, "Detecting elderly behavior shift via smart devices and stigmergic receptive fields", Proceedings of The 6th EAI International Conference on Wireless Mobile Communication and Healthcare (MOBIHEALTH 2016), 192:398-405 (2017).
6. F.A. Galatolo, M.G.C.A. Cimino, G. Vaglini, "Using stigmergy to incorporate the time into artificial neural networks", Mining Intelligence and Knowledge Exploration, 11308:248-258 (2018), doi: 10.1007/978-3-030-05918-7_22
7. F.A. Galatolo, M.G.C.A.Cimino, G.Vaglini, "Using stigmergy as a computational memory in the design of recurrent neural networks", Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods, 1:830-836 (2019).
8. B.F., van Dongen, B. Weber, D.R. Ferreira, J. De Weerdt, “Summary of the Business Process Intelligence Challenge 2013, in Proc. of the 3rd BPI Challenge, 1052:1-9(2013).
9. Fluxicon BV, Disco: Process Mining and Automated Process Discovery Software for Professionals, https://fluxicon.com/disco, accessed June 2025.
10. GitHub repository, StigmergicMiner, https://github.com/mariocimino/Stigmergic-Miner, accessed June 2025.
11. F. Lupi, M.G.C.A. Cimino, T. Berlec, F. A. Galatolo, M. Corn, N. Rozman, F. Lupi, M.G.C.A. Cimino, T. Berlec, F. Galatolo, M. Corn, N. Rožman, A. Rossi, M. Lanzetta, “Blockchain-based Shared Additive Manufacturing”, Computers & Industrial Engineering, 183:(2023).