Dear editor and reviewers,

We are submitting the original manuscript “From Data to Knowledge Graphs: A Multi-Layered Method to Model User’s Visual Analytics Workflow for Analytical Purposes” to the IEEE Transactions on Visualization and Computer Graphics for your consideration. This manuscript has not been published before, and it is not being considered for publication elsewhere. Yet, we would like to request to retain reviewer continuity from our rejected submission to IEEE VIS 2022 area 6.

The proposed conceptual framework, called Visual Analytics Knowledge Graph (VAKG), is a conceptual framework that models and structures the knowledge gathering process of a VA workflow into a set of linked knowledge graphs with which users can store and analyze user behavior. More specifically, through a novel Set-Theory formalization of knowledge modeling and, we define a multi-layer ontology structure by unfolding the VA knowledge model into its temporal and ownership aspects. By using this unified modeling and structuring process of a VA workflow, we present VAKG as a process with which to model and structure a given VA workflow. We also propose the use of a Temporal Knowledge Graph (TKG) in a novel multi-layer configuration to store relevant data, such as user insights and interactions, and as a means for VAKG users to analyze their own users’ knowledge-gathering process, providing a unified theoretical approach to bridge VA theory and practice. We demonstrate VAKG usage through a running example and a comparison to other theoretical works, showing that VAKG follows the needs and limitations of relevant literature. VAKG is unique in what it proposes by allowing for modeling, structuring, storing, and providing analysis guidelines for user behavior and knowledge generation, enabling comparison of users and VA tools.

We would like to thank you for your hard work in these uncertain and stressful times. May the community emerge stronger from these hardships.

Kind regards,
The Authors
Common Points

Reviewer #1

1) “What is the motivation? While authors are arguing the lack of connection between temporal aspects (e.g. sequences of events) and the workflow state space they are not necessarily pointing out what such a connection will yield in terms of benefits.”

Answer: Thanks for the concern with the paper’s readability. We understand from the reviewers’ feedback that the way our motivation was presented was unclear and confusing due to the high amount of introduced technical aspects. In general lines, VAKG is a conceptual framework that VA developers and researchers can use to model, store, and analyze their (user’s) workflow. This was motivated by several points. First, the usage of theoretical work to model and structure VA workflows has significant effort, but it’s usage in non-theoretical work, such as VA tools and systems, are considered only an inspiration or baseline. This causes the underlying theory to be used differently by each application paper. We also identified a need for a conceptual framework which can be directly used in practice due to the lack of reproducibility of user behavioral analysis. By collecting and comparing application and theory literature, we identified many similarities of needs, but no standardization where one tool could be compared to the other. We were not able to reproduce their analysis with our own users since the approach used does not directly follow VA knowledge model theory. VAKG was conceived to bridge this gap between theory and practice by providing a well-grounded modeling process of the VA workflow which, although still theoretical in nature, can be directly applicable in practice through knowledge graphs. The VAKG itself can be shared, and so can the results of its analysis. We have changed the introduction, related work, and the first subsection of the methodology to expose these motivations better.

2) “Is there a goal to inform the design of visual analytic systems to better support knowledge generation? Authors seem to argue that previous research only produced conceptual models but in the paper authors don’t make concrete steps, how the proposed architecture can operationalized, and as such stays conceptual. For example, authors talk about supporting both intra-dataset and inter-dataset operations but how this is done is not described. There are many questions that remain, for example, even at a conceptual level what constitutes a dataset? What about entities and relations and properties defined in such datasets? How is knowledge that is derived from a particular slice/dice of the dataset represented?”
Answer: We agree that we fail to exemplify how to model and construct a VAKG. In the revised version of the paper, we included a running example demonstrating the basics of concrete steps to use VAKG in practice. We also included better exemplification of inter and intra-dataset, as well as examples of analysis of two users using one VA tool and two users using two different tools as a running example of Sec. 4 and 5.

Regarding dataset definition, VAKG can hold whatever metadata pertained to a dataset’s origin and changes so long as they are included within VAKG’s property maps or sub-components. It is important to note that VAKG is still a conceptual framework, which means that it is not a system or library to be used by an application, but a formalized modeling process with which to perform user behavior tracking (e.g., provenance) for analytical purposes. This means that whoever wishes to use VAKG must decide what constitutes the property maps of its nodes, meaning that any dataset information, such as the slice/dice applied to the datasets, depends on whether the VAKG user includes it during the VAKG construction. We have better explained this concept, as well as the guidelines of what to include in VAKG’s property-maps in Sec. 4.3.

3) “Paper is really hard to understand with many abstract terms. Writing can be much improved, by using shorter sentences, defining concepts, giving examples, using figures, etc. I think having a running example throughout the paper to accompany the concepts, discussion, etc. would improve utility and readability of the paper substantially. I would also suggest authors to rethink the organization of the paper.”

Answer: Thank you for your feedback and suggestions. We reorganized the entire paper to expose the concepts more clearly. We included a simple running example throughout the paper to illustrate each part of VAKG. We also concentrated all term definitions in a section, especially those pertaining to theoretical research of VA workflows. We hope the revised paper is easier to follow and delivers the correct message.

4) “There are many statements in the paper which are not substantiated, or positioned as possibilities, yet they are used to suggest the utility beyond the framework. For example, by applying graph mining it is claimed it is possible to find patterns of analytical behavior. For example, by applying GNNs (Graph Neural Networks) we can find embeddings to represent classes and relations in the KG. For example, applying shortest path algorithms to reach an insight/knowledge faster. Unfortunately, there is a significant gap between how these approaches work, what kind of data they need (incl. labeled data), whether sufficient data can be collected from VA tools, whether
the collected data has sufficient signals/features to achieve good models, etc. that at the end of the day might not match claims/expectations. As such a significant part of the paper reads as position/vision paper. I would suggest authors to focus on one aspect/problem and make concrete progress towards the goal.”

**Answer:** Thank you for the feedback. We take our failure from not properly conveying the difference between VAKG’s technical guidelines and contribution. As the reviewer noted, we described that if one follows the process presented in this work, they will model and create a VAKG that structures and stores the information about some VA tool’s workflow. This VAKG graph can then be analyzed using many techniques, including GNN. Of course, to use VAKG in any of such techniques, the information within it must follow the conditions required by the technique. Therefore the examples used in our work are for exemplification purposes, where users can utilize as a starting point to model their own VAKG in a way to fulfill their requirements. For instance, such user would need adapt their property-map design so that GNN can be applied.

We have revised the text to differentiate between VAKG technical guidelines and contributions, especially the Sec. Limitation and Future Work to explain the potential of such future works.

5) “While it is great to see interviews conducted the linkage between findings from the study to the design of the proposed framework is missing. They are connected to the goals put forward but not necessarily inform the design of the framework (in fact the goals are applicable pretty much in any data science workflow, provenance scenarios)”

**Answer:** Thank you for your advice. Due to the feedback, we have removed the interview sections and focused on providing better examples and motivations through analysis of related work.

6) “Paper is lacking concrete contributions”

**Answer:** We appreciate your feedback. Although we have modified the paper to better expose and explain our contributions, our core contribution is a conceptual framework that retains all dynamics of a visual analytics workflow. Therefore, we can argue that our contribution is not concrete (or an application) but theoretical. In short, VAKG is a novel conceptual framework that brings VA theory into practice by a novel formalization of the VA theory through Set Theory. Also, we propose a novel ontology and method to construct temporal knowledge graphs based on the proposed formalization. Unlike existing works, we use VA theory instead of reinterpreting
for domain-specific use cases and simultaneously provide a formal method to model, store, and analyze user behavior. Together with the included running example, we demonstrate how to use VAKG in practice, showing its novelty compared to prior works. VAKG is guided by related work’s previous limitations or domain-specific goals and generalizes for any VA workflow. To clarify this, in the revised paper, we have significantly modified our introduction, related works, and the first section of the methodology.

7) “Typos”

**Answer:** We have also corrected the indicated typos and proofread the paper. We appreciate the feedback.

**Reviewer #2**

8) “The text lacks a more formal explanation of the mapping of VAKG into the basic VA models. It can be improved by moving some pieces of information from supplementary materials into the paper.”

**Answer:** Thank you for the feedback. We agree and have included and polished our presentation of the VA models used as a basis for VAKG. We also better formalized VAKG in the first half of our methodology to demonstrate one of the novelties of VAKG: the formalization of VA knowledge models through Set Theory.

9) “Unclear relevance of interviews with domain experts to the results of the paper. Instead, a “proof of concept” should be presented, e.g., by providing examples of the mapping of the two business domains onto the VAGS.”

**Answer:** Thank you for the feedback. After careful consideration, we concluded that the interviews did not convey VAKG’s motivation and potential as we wished, so we have removed them from the paper following your and other reviewers’ feedback. Also, we have included a running example and further examples of how to use VAKG in practice.

**Reviewer #3**

10) “I had a lot of difficulties reading the paper because the terminology is dense and not always introduced. Are the VAKG a “multi-layer method to model ...” or a “conceptual framework”?”
Answer: Indeed, we realize that, given space constraints, the reduced number of examples and the large number of technical terms affected the paper’s readability. We appreciate your feedback, and based on that and other reviewers’ comments, we have restructured the paper to reduce the number of technical terms to clarify what is VAKG. To directly answer the question, VAKG is a conceptual framework, a somewhat generic term that includes multiple underlying pieces. VAKG uses existing VA models, formalizes it with Set Theory as a theoretical system, defines an ontology, and finally exposes this ontology as a multi-layer knowledge graph. Definitions for these concepts were included in our revised background definitions and related works, which now describe each concept and compare them to related literature. Also, the revised manuscript includes the complete formalization and process where the theory of VAKG is defined, and then examples of how to apply it in practice are described.

11) “The abstract starts with: ”the importance of knowledge generation drives much of VA” but I would instead think that VA drives knowledge generation.”

Answer: We agree that VA drives knowledge generation. In our previous abstract we tried to emphasise that the VA research efforts are driven by the concept of knowledge generation. Therefore, we do not see our affirmation contradicting yours but reinforcing it. To better show what we mean, we have changed the text.

12) “[The abstract] also mentions the workflow "state space” but does not explain what is that state space, taking it for granted.”

Answer: We agree that we fail to explain all terms used. We have changed the abstract to better convey our meaning while attempting to simplify our wording. Thank you for your feedback.

13) “Then VAKG is introduced as a "provenance model and a state space graph”, in addition to being a conceptual framework and a multi-layer method to model …”

Answer: Similar to the previous answers, we have revised our work to define VAKG more clearly and convey its concepts. Once again, VAKG is a conceptual framework that includes, among others, an ontology that is used to design a temporal knowledge graph (TKG). This TKG has four layers, one for each of the concepts described in our revised methodology: Human temporal sequence, Human space-state, Machine space-state, and Machine temporal sequence. Later
in the text, we argue that the structure defined within VAKG is well suited to model and store provenance-related data. For instance, behavior provenance is the act of modeling a data structure for collecting a user’s behavior in time and, after collecting it, analyzing it for some goal. Therefore, the data stored within Human Update nodes can be used for behavior provenance. Similarly, data provenance can be applied to the Machine Update sequence. Hence, such concepts are more related to the usage of VAKG than the definition of VAKG itself. We have revised much of our work to clarify VAKG’s definition.

14) “Then later: ”VAKG introduces the concept of a multi-lane TKG model, which unfolds the VA typology states over time while keeping the separation between Human and Computer concepts and between the sequential and the state space aspects of a VA workflow, resulting in a knowledge graph which provides an architecture for VA systems to structure, store and link users’ knowledge generation workflows, all computer-side changes over time and the relationship between users’ reasoning and the executed workflow.” This sentence is long and very difficult to parse. It continues with ”As our core contribution, VAKG introduces the concept of a multi-lane TKG model”, but this concept becomes ”4-lane” later... To add to my confusion, section 7 starts with ”we argued that VAKG is more capable than other existing ontologies” so now, VAKG is an ontology.”

Answer: Thank you for the feedback. We understand the confusion. As discussed, VAKG is a novel conceptual framework that brings VA theory into practice by a novel formalization of the VA theory through Set Theory, with which we propose a novel ontology and method to construct temporal knowledge graphs based on such a formalization. Unlike existing works, we directly use VA theory instead of reinterpreting for domain-specific use cases and simultaneously provide a formal method to model, store, and analyze user behavior. Therefore, an ontology is part, but not the whole, of VAKG. We agree that the text is confusing, so we have changed it completely to better convey how VAKG uses a newly proposed ontology to achieve its goals.

15) “Therefore, at the end of the introduction, I am still unsure what the contribution of this paper is.”

Answer: Our introduction was revamped significantly to avoid complex terms at all costs, using them only in the subsequent sections. With this, we introduced VAKG with a single goal instead of a single goal from multiple angles (given the impression we had multiple goals).
16) “In the end, what I understand is that the authors are trying to model a VA session as a temporal graph of elements/fragments of knowledge gathered/constructed through interactions. Yet, as written in the Limitation section “it does not solve the issue of how to perform user-tracking”, so we don’t know what are these fragments of knowledge, which seems critical for the model to be applicable.”

**Answer:** Our goal was not to present a model nor to provide a technical solution to user-tracking and behavior analysis but a conceptual framework firmly founded and formalized from VA’s knowledge model wherein VAKG users may model, store, and analyze user behavior and knowledge gathering using any of the presented existing works. We also argue that different from existing works, the knowledge graph generated using VAKG provides a structure where the two temporal sequences and two space state graphs coexist and are interlinked. VAKG not just follows and solves the limitations of the related works but enables a single structure to solve their diverse goals, such as analyzing user behavior, user interaction, data modification, and knowledge gathering. Therefore, VAKG’s aim is not to solve the technical aspect of user-tracking but to bridge the gap between theory and application-based works concerning the application of VA’s knowledge model. We have revised most of our introduction to clarify this motivation and our limitations and future work section to describe, among other things, that technical solution such as user-tracking, though important and relevant to VAKG is better solved in domain-specific or application-based contributions.

17) “Overall, I think the work is important and can be useful, but the paper in its current state is too difficult to read, uses inconsistent terms, introduces too many concepts, and does not convince me that VAKG is more useful than the state of the art in knowledge acquisition modeling. I would advise the authors to explain better their goals, simplify the description of their model and better show the benefits of using VAKG given its complexity.”

**Answer:** We appreciate the feedback. We have modified our motivation and comparisons with existing work to show that no other existing framework can provide the complete analysis capabilities of VAKG. In summary, the changes explain that every existing work does not consider the full picture, only focusing on its domain-specific goals. For instance, certain works only track user knowledge gained over time without considering how the knowledge came to the user (e.g., which visualization it came from), or others focus on storing user’s interaction while ignoring what users were attempting to do with the interactions (e.g., user intentions). They also do not use the VA knowledge model directly. Instead, each reinterprets the theory in a non-standardized way. VAKG is unique in its way
of storing provenance (e.g., the timeline of events), the state-space (e.g., what are all the possible visualizations that the user can analyze), and how the two are connected (e.g., on average, which user intention among all users caused which interactions) in a way which solves all aforementioned gaps.

18) “I believe the work has value but would need a major rewriting to clearly explain the goals, clarify the concepts used in the paper, and describe a few examples convincing the reader that previous models are not sufficient and that VAKG can be used to answer more questions.”

Answer: Thank you for your feedback. We have modified our manuscript following the reviewers’ comments, including multiple examples of the usage of VAKG in practice.

19) “Commenting on the Appendix of the paper, VAKG becomes much more mathematical and so less confusing. Still, it is quite complex, and the appendix is referencing an unknown figure (it mentions square nodes in the text but I cannot find any square node in the figures). The notation seems wrong too, the text mentions $K_t$ and $K_e$ whereas the mathematical notation has $K^T$ and $K^\epsilon$. Still, the VAKG model is more explicit. My main issue is that modeling a human as a state machine is very simplistic, although maybe useful. The authors seem to prefer the mathematical notation they use in the appendix to the informal descriptions they use in the main paper. It might be better to introduce VAKG as a mathematical formalism in the main paper instead.”

Answer: Indeed, VAKG’s first submission attempted to explain it by focusing on its mathematical formalism, but we moved this approach to an appendix based on reviewer’s feedback. Since now the space constraint issue is somewhat lessened, we have reinstated the mathematical formalism and its reasoning into the main paper as part of our modification and restructuring process. Also, regarding the simplicity of modeling a human as a state machine, we agree that if we were to model the whole user as a state machine, it would miss much of the user’s inherent complexity. Therefore, to be more clear in our examples, we have removed the state machine example but instead focused on the terms from the existing VA knowledge model works. With this, we can say that VAKG tries to model the user’s atemporal data (e.g., tacit knowledge or intentions/goals) and the machine’s atemporal data (e.g., the data being analyzed or the structure, schema, or state of the VA tool) as an atemporal space-states. This is separate from time-related information of the user (e.g., perception or exploration) or machine (e.g., visualization updates). Therefore, VAKG is, for instance, able to show what were
the actions taken by the user while they were in the state of "Investigating United States History" while the user browses Wikipedia. That said, the question of how granular the state definition should be is not part of the conceptual framework we have proposed but part of the domain-specific application of VAKG. We modified the text to better explain this reasoning and the example explanations to show what we expect from the theoretical nature of VAKG and what will be up to each individual application of VAKG.

Reviewer #4

20) “1. More illustrative figures would be helpful for the reader s to follow the workflow. 2. The texts in Section 1, 2, and 3.1 could be more compact. 3. While there are comparisons in Section 5, these are not led to elaborate case studies. Some case studies would be good for understanding the workflow. 4. arXiv papers must be either removed or changed to the proper citation. 5. The work by Sacha et al. and Federico et al. could be introduced in more detail so that the readers catch the main points directly.”

Answer: Thank you for your feedback. We have modified our submission in a way that follows the reviewer’s feedback.

21) “However, the authors confuse me due to the low quality of the narrative style despite the good logical development. In addition, it is necessary to reconsider the paper structure. Case studies would also be necessary to increase the understanding and trust in the proposed workflow. I had to mentally imagine the VAKG proposed in this paper only through the explanations and examples provided by the authors. In addition, it is expected that intuitive practicality would be difficult in utilizing the proposed VAKG.”

Answer: Thank you for your feedback. After extensive review and restructuring of the entire text, and with the additions of several examples, including a running example, we believe our new submission is now following your advice.

22) “The authors visualize knowledge following a complex workflow with multiple symbols and colors, which depends on the user’s memory and makes complex knowledge maps more complex. Therefore, it is necessary to devise intuitive visualization.”

Answer: We realize that the differentiation of VAKG’s ontology structure and resulting VAKG graphs when it is applied in practice was not well done. We have modified our language throughout the paper to better convey the need for the formalization of VAKG’s
internal ontology, which includes the different symbols and colors used in our figures, and finally, how this affects the resulting VAKG when it is applied in practice.

That said, in case your comment is meant to ask about the visualization of graphs, VAKG is not something to be used for visualization of the VA workflow, but instead a conceptual framework with which you can model, structure, store and analyse the VA workflow. The images used throughout the paper are only for exemplification purposes. Future work regarding the visualization of VAKG is planned, as we have now included in our revised future work section.

23) “1. Authors need to improve the overall narrative style of the paper. A. It is difficult to find the message the authors want to convey in paragraph 1 of the introduction. The current introduction makes it difficult to understand the motivation of the research. B. Overall, the authors use too many examples to support their arguments, which makes me difficult to understand what the authors wanted to tell. C. The authors need to be more specific. For example, i. In the first sentence of Section 1, “To model user knowledge and its relationship to VA system, researchers have been using from the very beginning the knowledge generation model of Sacha et al. [39] as a guide.”, it is difficult to understand what the model of Sacha et al. [39] is and why people use it. To understand this sentence, we must read Sacha et al. [39] or other related studies. ii. the Last paragraph of Section 3 “Although this limits certain aspects of VAKG, as we have discussed before, specific use-cases can upgrade the property-map of a summary Update into an inner ontology which enables further information to be modeled within the TKG.”, this is difficult to understand whether “we have discussed before” in the sentence refers to the previous paragraph or another section. ”

Answer: Thank you for your feedback. Following your advice, we have rewritten our introduction to better convey the motivation of VAKG and to better separate our contributions from related works. We also have provided a single running example throughout our methodology, only discussing other examples in the Sec 5. Our methodology (now Sec 4) also was vastly modified to better structure the contributions of VAKG, including the advice you have given regarding the last paragraph. While we revised our introduction, we have also revised our wording in our methodology to better explain that the model of Sacha et al. is in Fig. 1(A).

24) “2. Authors need to rethink the section structure. A. The roles of the background of Sections 1, 2, and 3.1 overlap. It is necessary to separate the critical scope in each section and revise it to the scope that the authors want to convey clearly. B. I think the design process of VAKG, visual encoding, and examples for understanding VAKG are essential. However, I also think that
VAKG and various case studies generated in the actual analysis process are more important than that. A larger portion of the paper should be allocated for the VAKG and various case studies generated by the analysis process. C. Section 5 is more suitable for Related work. Section 5 now seems inadequate to play any role in the case study or evaluation. ”

Answer: Thank you for your feedback. We have significantly modified the paper structure to follow your advice. We have included a running example to better help with the paper’s understandability and concentrated other examples in Sec. 4. In order to better separate the roles of Sections 1, 2 and 3.1, we also included a separate background section (now Section 2), and removed such role from the other sections. Therefore, Sec. 2 is the only one that exposes the complexity of the theoretical background used by VAKG, which simplified the wording and roles of the adjacent sections.

25) “3. Case studies on VAKG presented in this paper are necessary. The authors conducted domain expert interviews without case studies. Therefore, even after reading this paper, I had to imagine real cases by relying on the theories and examples of the new workflow that the authors propose.”

Answer: VAKG is a conceptual framework. Since each potential use-case would incur its own ways of performing visualization and analysis of VAKG and its own requirements in provenance, we opted to follow other theoretical papers by comparing our goals and limitations to other existing works and through multi-faceted examples. Therefore, we have updated our introduction and limitations sections to include a better explanation of how VAKG fits within the VA literature, especially concerning its theoretical contributions. We also now discuss in our limitation and conclusion sections that although VAKG is a conceptual framework, we expect as future work a VAKG implementation for certain domain-specific problems, including user-evaluation of such implementation. We also updated the first half of our methodology and removed the interview section altogether to better convey that although VAKG is unique in its ability to be directly used in practice, VAKG’s novelty is in its theoretical contribution and therefore follows the evaluation process of similar theoretical work, such as Vis4ML.

26) “4. It is difficult to understand how the descriptions of the Human States and the Computer States in Fig. 3 are generated. Authors need to specify whether they are manual or automatic and, if so, how they are generated.”

Answer: Thank you for your feedback. We have significantly restructured our explanations to better convey the four lanes, including Human/Computer States, now called Human/Computer Space-State, and Human/Computer Updates, now called Human/Computer
temporal sequence. We also gave a running example to better differentiate and explain involved concepts throughout the methodology. That said, it is important to note that VAKG is a conceptual framework that does not attempt to perform the collection of data. Therefore, we do not attempt to implement or impose the methods of how an individual domain-specific application would collect the information to fill VAKG. In our running example, for instance, the data was collected by watching the user’s videos and filling in the required data. In another example, an EDA tool could automatically track user mouse clicks and ask for text feedback through any of the existing user-tracking libraries. Therefore, each use-case requires a domain-specific solution, and the contribution of VAKG is on how to structure, store, and analyze data. We modified the paper for this goal to be more clearly understood.