**Reviewer 1 (coordinator)**

**Second Round Recommendation (used only for the second round review)**

*(blank)*

**Second Round Review Text (used only for the second round review)**

*(blank)*

**Summary of all reviews from the Primary Reviewer**

The work presents a interesting and fun visualisation method that is likely to have a contribution. However, in the paper's current state, we don't feel that it is ready to be accepted at PacificVis. In future revisions, we would encourage the authors to revise as follows:  
  
1. For this type of work, a comparison, even in a qualitative sense, with the current state-of-the-art is necessary. This comparison should demonstrate the advantages of the new approach over this state-of-the-art. (R1, R2, R3, R4) Currently, it is not clear what the advantages of the proposed technique are.  
  
2. The contribution should be more clearly emphasised in the paper. It should be clearer what is being claimed as a research contribution (R1, R3)  
  
3. The writing of the paper seems to discuss case studies that are not present in the paper (R3).  
  
We would encourage the authors to further revise the work.

**Recommendation from the Primary Reviewer**

Probably reject

**The Review**

This paper describes a technique for visualising wordstreams interactively. The approach of the paper is to combine wordle with streamgraphs. Metrics are described to place selected words compactly in the streams. Words can be selected based on their frequency in the data set. The approach is tested on a number of data sets ranging from political blogs to contents of IEEE VIS meta data.  
  
I think the paper does have a chance for acceptance, but I'm currently leaning towards reject. My main reason for this decision is that I feel technique papers need to compare against the best of bread systems. There are a number of systems that produce similar sorts of visualisations and, at the very least, I think that the developed technique needs to be compared to those systems. In particular TextFlow and TIARA have strong similarities. I think it is possible that this system brings improvements, but it should be compared, at least qualitatively, to show the improvements over these methods. I don't think that it is sufficient to show only images from their technique without a comparison. I believe that there could be improvements, but at this moment the paper doesn't contain the necessary information to demonstrate these improvements.  
  
For the average normalised frequency, I think TF-IDF may help some of the problems seen by the paper. It may be a good idea to consider this in future to improve the visualisation.  
  
I think that the paper presents a good algorithm. The approach seems relatively easy to implement, and would be efficient. However, I think I need a more explicit comparison to existing techniques to be clear about what benefits the approach brings.  
  
Overall, there is potential here, but I lean towards reject. My main concern is that further evaluation is probably necessary in order to convince the reader of the benefits the approach brings.

**Overall Rating**

Borderline: Overall I would not argue for accepting this paper.

**Expertise**

Knowledgeable

**Reviewer 3 (committee member)**

**The Review**

Summary:  
This paper proposes WordStream, which combines visualization techniques from Wordle and StreamGraph. The authors evaluated four displaying options, which are from Flow and Angle variance features. The paper tries to resolve the problem of text visualization over time, which is previously studied by several researchers. However, it lacks comparison with prior work and its contribution is either not fully emphasized or marginal in its current form.  
  
- The contribution of this paper is not properly emphasized. It seems the contribution is rather marginal when compared to prior work that the authors might want to highlight the strength of the proposed technique. For instance, it is harder to find time points of critical events with WordStream when compared to TextFlow.  
- The authors listed several goals and design decisions in section 3, which could be convincing with additional justification.  
- The authors introduces WordStream as an interactive visual tool for text analysis, but it is difficult to find descriptions regarding interactivity. Fig. 6 shows the result of word selection, which is still not enough--the authors might want to include a video to show a list of interactions supported by this tool.  
- The authors might want to include more detailed case study results with solid usage scenarios for evaluation. Section 6 seems to be too weak to demonstrate the usefulness of the proposed technique.  
- It seems some figures are missing. (e.g., the authors refers to WordStream visualization of Crooks and Liars (in section 5, paragraph 2), which is not included in this paper. In addition, the authors refer to Fig. 2 in section 4.3 to explain random spiral starting points, but Fig. 2 only shows the compactness metric.  
  
Minor comments:  
- Typo in (page 5, section 4.2): StreamGrap   
- Typo in (page 8, reference 23): name of authors are wrong and not properly cited  
- Fig. 5 shows each step to build a WordStream visualization. It would be easier to understand the process when example figures are provided.

**Overall Rating**

Probably reject: I would argue for rejecting this paper.

**Expertise**

Knowledgeable

**Reviewer 2 (reviewer)**

**The Review**

The authors propose a text visualization technique combining the word cloud and streamgraph metaphors for representation of time-varying textual data. The main aim of the technique, as stressed by the authors, is to represent the temporal evolution of text topics in the data. The authors describe the design process and the algorithm in detail. The validation of the approach is based on several use cases and metrics measurements taken for visualization of several text data sets from various data domains. Overall, the paper fits the scope of PacificVis very well.  
  
I think the main strengths of this paper include the clarity of the main idea and the level of detail in the description of the design and implementation process. In contrast to many existing works in visualization, I have no doubts about reproducibility of this contribution. The open-source release of the implementation is also a big plus (if the authors added a video demo to their submission, it would be even better, though).  
  
However, I do have a number of concerns about the paper in its current shape:  
1) While the authors already cover a lot of relevant previous works on word clouds and streamgraphs, I think the comparison of the proposed approach with the following works should be added: Fisheye Word Cloud by Wang et al. (https://doi.org/10.1145/2468356.2468673), which provides another interactive approach for time-varying data visualization with word clouds, and perhaps CiteRivers by Heimerl et al. (https://doi.org/10.1109/TVCG.2015.2467621), which uses a streamgraph representation with embedded tag/word clouds. The authors also mention TIARA in Section 2.3, but do not really argue how their proposed approach is better in supporting the design goals related to temporal data analysis.  
  
2) I would recommend the authors to compare their decisions and findings to the ones described by Felix et al. (https://doi.org/10.1109/TVCG.2017.2746018). According to that study, the effectiveness of various variations of word clouds (including even simple word lists) depends on the user task. The authors of the proposed approach, however, rely on several metrics for optimizing space usage and distance of the resulting visualization rather than evaluating the effectiveness and efficiency with the users. It makes me question the usefulness of the technique. For instance, I would argue that some parts of the visualization depicted in Figure 1 are difficult to perceive due to the rotated word labels.  
  
3) I have further doubts about the metrics introduced in Section 3.2, the corresponding measurements for the test data sets depicted in Figs. 2–4, and the lack of discussion thereafter. I am not sure if any of the results have statistically significant differences, as the authors provide no analysis or discussion; as a reader, I am left to wonder if these results are interesting or not, and whether the introduction of metrics is warranted at all.   
  
4) Related to the previous points, the validation part of the paper is its biggest weakness, in my opinion. I understand that the authors did not conduct a user study, but in this case I would expect to see some interesting case studies or use cases as well as a critical discussion about the usefulness of the proposed approach. The current version of the manuscript offers a rather long description of 11 data sets from various data domains, but then there is basically just a single use case (on political blog data) described in Section 5.1 that provides the reader with an idea of how insights could be obtained with this approach (this use case also mentions filtering/selecting specific terms of interest, but there is no corresponding figure...). I expected at least some minor notes about other data sets and domains.  
  
5) While the manuscript does not exceed the allowed length, the structure and the formatting of the content seem to be off in some regards. For instance, my notes above already include doubt in the usefulness of Figs. 2–4, which collectively occupy a full page right now. Fig. 5 could also be optimized (it also contains an incomplete box between stages 4 and 5). Figs. 6–10 should be positioned before the bibliography, in my opinion. Fig. 6 could be resized to occupy less space, as it contains a lot of whitespace due to filtering (I would also strongly recommend to add a note with a reference to Fig. 1 in the caption of Fig. 6, if it is the same data set as in Fig. 1). Finally, Figs. 7–10 could also be compressed a lot if necessary (I think the researchers from the visualization community would be interested to look at one of these figures in detail, but the comparison of all four variations would make sense even with small subfigures).  
  
6) Finally, with regard to the language, the paper is comprehensible, but there are numerous language issues (see some examples below). The authors should invest much more effort in proofreading the manuscript.  
  
Minor issues:  
- p.1, Abstract: "during a time line" — "over time"  
- p.1, Abstract: "a new... model which emphasize"; "The words is applied"; and a lot of similar coordination issues throughout the manuscript  
- p.1: "font-size" (occurs several times throughout the manuscript)  
- p.1: there is also another formatting issue that starts in the first page — the authors are not consistent with regard to applying the italics for technique titles (e.g., "Wordle" and "WordStream" vs "StreamGraph"). This should be fixed in the complete manuscript in a consistent way.  
- p.1: the first page ends with an orphan line  
  
- p.2: "This approach allows users to convey" — "to perceive" or "to analyse"?  
- p.2: "In section 5" (the same issue occurs several times with other references to sections, equations, etc.)  
- p.2: "the results of the application of WordStream with real data sets" — "to real data sets"?  
- p.2: "give emphases to" — "emphasis" (occurs several times)  
- p.2: "Wordle is aesthetic and visual appealing" — "aesthetically and visually appealing"  
- p.2: "but still reserving the word neighborhoods" — "preserving" (occurs several times)  
- p.2: "placed either in alphabetical or by size" — "alphabetical order"?  
- p.2: "This cannot show the term evolution or topic evolution." — "It ..."  
- p.2: "line-graph"  
- p.2: "It also concerns about" — "It also represents"  
- p.2: "each layer serves a topic of interest" — "serves as" or "represents"  
- p.2: "the aesthetic of"  
- p.2: "All these versions of the stacked graph have common issue is the limitation …" — "… have a common issue: the limitation …"  
- p.2: "are specified as the followings" — "are specified as follows"  
  
- p.3: "Place terms to its corresponding position, regarding to its time step and its topic" (occurs several times)  
- p.3: "The overview for the entire stream. therefore, is clearer"  
- p.3: "an individual words"  
- p.3: "as described as follows"  
- p.3: "for each data sets" (occurs several times)  
- p.3: "due to lacks of"  
- p.3: "the fours displaying options" (occurs several times)  
- p.3: "that one term maybe mentioned several times, in some cases it may not convey additional information" — "that a term mentioned several times might not convey additional information in some cases"  
- p.3: "These include of the followings:"  
- p.3: "This step uses … and visualize them"  
  
- p.5: Fig. 5 issue described above + lack of full stop in the caption  
- p.5: "computed from the pixel data using this formula sprite[i] …" — "using the following formula: sprite[i] …"  
- p.5: "to the equations 3 and 4" — "according to Equations 3 and 4" (occurs several times)  
- p.5: "About scanning for …" — "When scanning for"  
  
- p.6: "with the font-size is relative to" — "with the font size relative to"  
- p.6: "The font size and opacity of each word representing its frequency, …" — "represent"  
- p.6: "while the color encode"  
- p.6: "The next 5 datasets contains" — "The next five datasets contain"  
- p.6: "1990-2016", "1996-2015" — please use en-dashes for time ranges  
- p.6: "Fig. 7 to Fig. 10 shows WordStream application with" — "Figs. 7–10 illustrate WordStream application to"  
- p.6: "This helps provide"  
- p.6: "techniques namely Wordle"  
- p.6: "… and still keeps the presentation-oriented criteria" — "… , while preserving/satisfying the presentation-oriented criteria"  
- p.6: "WordStream is to be able to place"  
  
- pp.6–8, Bibliography: Please fix the capitalization issues in proper names such as "Imdb", "popcha!", "wordle", etc. The capitalization of paper titles is also inconsistent (cf. [3] & [8]; [6] & [9]). I would also fix the titles of some proceedings (e.g., in [6] and [12]) to use the "Proceedings of …" format.  
- p.7, [4]: "I. Rodchenkov, . Babur, N. Anwar, "  
- p.7, [10]: "eyes of experts., pp. 37–58"  
- p.7, [13]: "sep 2017"  
- p.7, [14]: outdated duplicate of [13] !!!  
  
- p.8, [19]: "Dec 2014"  
- p.8, [22]: "F.B.Viegas, M.Wattenberg, J.Feinberg, H.Dwwhqehuj,D.Q.G.Hlqehuj, P. Dxglhqfh, Z. Dv, R. X. U. Zrun, L. V. Ylhzhg, P. Ri, W. Rq, V. Zrug, F. Ru, Y. Dwlrq, P. Wdnh, L. Q. Pdvv, S. Ri, F. W. Srjudsk, D. Q. G. Frpsrvlwlrq, F. Volghv, D. Q. G. Krxvhv, R. I. Zruvkls, and W. K. H. Kdyh." — I seriously doubt the correctness of this author list...  
- p.8, [23]: "H.-c.D.Visualization" — this one, too  
- p.8, [25]: "Weiwei Cui, …" — all author names are abbreviated in the rest of the bibliography  
  
  
In summary:  
+ Nice and clear idea for a text vis technique;  
+ Detailed description of the design process and algorithm;  
+ Open-source implementation;  
  
- Weak evaluation;  
- Lack of critical discussion of some existing approaches and the results;  
- Presentation issues.  
  
I am quite conflicted about this work in its current full paper status; the current amount of content and level of discussion evaluation would make it perfect for a short paper (a VisNote) after some compressions. At the same time, I think it is not impossible for the authors to fix the presentation issues, improve the discussion, and extend the use case demonstration in a time span allocated for the minor revision, if the paper is conditionally accepted. Therefore, I assign a "borderline" score at this time.

**Overall Rating**

Borderline: Overall I would not argue for accepting this paper.

**Expertise**

Expert

**Reviewer 4 (reviewer)**

**The Review**

This paper presents a visualization called WordStream that combines both Wordle and StreamGraph to show topic and term evolution of textual data, such as newspaper articles, scholar papers, and blogs. The authors claim to provide three major contributions related to the benefits of showing Wordle inside StreamGraph, like Tiara, but in a more compact and aesthetically presentable way. They described the algorithm used for processing text and rendering them within given constraints. Also, they compared how different parameters have affected the metrics that they have chosen to evaluate each option. They also explained the characteristics of each data used for the evaluation, followed by conclusion.  
  
I am personally impressed with the aesthetics and fun-aspect of the visualization but was not convinced that the paper had a fair amount of contribution to be presented to the conference. Generally, it requires great amount of revising the paper and adjust the distribution of the space allocated for each section, as well as adding more convinced argument that such visualization can support users’ creative or explorative tasks. The writing seemed rather rushed than polished.  
  
First, although the authors presented design goals and decision made to support them, they do not provide how those goals were selected to improve upon other visualizations and overcome their limitations. For example, G1, G2, and G3 are very common goals found in previous literatures that supports textual visualization for topics that evolve over time, therefore should be supported by default. The authors should have cleared outlined which aspects of G1-G3 that the previous literatures have failed to deliver and focused on how WordStream can do better. While D1-D5 are fairly crafted to support G1-G4, the goals themselves are hardly the reasons that WordStream is better.  
Also, the design decisions (3.2) seem rather technical than goal-oriented. The compactness, display rates, and average normalized frequency were chosen to be metrics. There needs to be more convincing reasons that such what each metric means to the end goal of the visualization. For example, there seems to be an evidence (at least by looking at the chart, not by any statistical means) that the compactness can vary based on the parameters, but what are the implications? Why is it important? Does more compact visualization automatically mean the visualization deliver more message? Wouldn’t too compact visualization clutter the view? The authors need to explain these possible questions and discuss the strengths and weaknesses when appropriate. Also, the authors can explain what tasks the users are trying to achieve and what the scores in each metrics implies for supporting them. The tasks do not necessarily have to be all scientific. They can range, for example, from pure perceived aesthetics to scientific analytics tasks.  
  
In addition, the authors did not fully discuss why they got such results for each metric. The results are shown only through charts, but they do not explain why such parameters made certain metric the way it turned out to be. For example, the chart says “No flow, No angle variance” has higher compactness, but doesn’t say why. Also, Average Normalized Frequency doesn’t seem to be affected as much by the parameters, but why? Some of results can be understood by the natural instinct, but the scientific paper needs to describe the phenomenon in a structured way. There is simply too little analysis of the results, especially when a whole page is dedicated to only show charts for each metric. Instead of providing all results from all dataset, I suggest the authors pick more interesting dataset that clearly demonstrate the strength of WordStream. By doing so, they can reduce the amount of space used for charts and use it to analyze them (perhaps w/ statistical analysis).  
  
The authors can find more space by cutting down on implementation details. For example, placing terms(4.3) can only highlight how WordStream differ from other visualization. The whole ‘wsi’ and ‘bsi’ part including formula (3) and (4) is very generalized description of the word placing algorithm on a board and is not worth a whole column of a page. Yes, everyone loves fast programs. But in my opinion, the main contribution of this paper should come from the visualization itself, not from the programming optimization, because the goal of WordStream is not to render it in more time constrained environment. And there are certainly other types of visualizations that have such goals, for example involving user interaction, in which case such descriptive implementation may be more important. But that’s not the case for WordStream.  
Because the authors dedicated more resource in explaining rather unimportant part, the paper seems to sprint towards the conclusion at the end. Section 5 has too much description on the data, which can be provided as supplemental material. The real insight of this paper is condensed into only one paragraph in 5.1. Section 5.2 only describe the general goal of using textual visualization of timed data but fails to shine the benefits that WordStream brings to the task. The authors should dedicate significant amount of paper for sections like 5.1 across different dataset and findings.  
Section 6 can be removed and merged into Section 4.  
  
Overall, I think the paper can be vastly improved by providing more examples on the scenario and how their visualization benefits when users perform certain tasks. The authors should also allocate spaces based on their importance. I am afraid that I cannot comment further on the study itself, because the authors do not discuss them in detail. I cannot either agree or disagree. In short, I want to see more discussion and less implementational details. And these changes are more substantial than I would consider as a minor revision, thus I do not believe the paper is ready for publication. If the paper is accepted by popular vote, I still suggest the authors carefully review my opinion and improve the paper, because it will greatly benefit the readers. And even if not accepted, I encourage the authors to keep pursuing the research, because I really think that there some great missed opportunity here that they have failed to demonstrate.  
  
Other improvement:  
The authors should carefully review the bibliography. While minor typos and formatting such as inconsistent italicization can be fixed easy for the final version, it is not acceptable to include wrong authors or omit them all-together. For example, [22] and [23] have three authors each, instead of some obfuscated names. I believe this was just a mistake made during the crunch time of submitting the paper. But they still need to be accurate for any publications nonetheless.  
  
Minor typo found  
P3C1P4: Add new line before D5  
P3C1P9: Fours options -> four options  
P3C2P1: Add space before the last sentence  
P3C2p2: Fours displaying options -> four displaying options

**Overall Rating**

Probably reject: I would argue for rejecting this paper.

**Expertise**

Knowledgeable