Scientometric Mapping of Interpreting - Group 3

Review 1

1. **Quality of data selection, cleaning, preparation, and documentation**

Selection of data from SCOPUS using keywords listed in the document is appropriate. The cleaning process noted that there was a lot of noise in the data. It would be good to clarify with examples what noise means in this case.

2. **Appropriate selection of tools, algorithms, workflows, and parameter values**

Your use of Burst analysis and its associated Temporal graphing techniques will show change of vocabulary (or research interest) over time. It is good to see you use the Abstract field, in addition to the Title field, to perform your analysis. You'll get a larger coverage of keywords in this way, than using just Title.

Did you perform your burst analysis with different parameters? I remember playing with the Gamma parameter during Week 1. Another one that might be useful will be the Density scaling depending on your dataset.

Another word of caution around burst analysis. I ran some tests with multiple bursts of the same frequency over different time periods and SCI2 detected only the first one. You may want to use other toolsets (maybe just excel) to detect multiple bursts.

The addition of a Word Cloud will be a great starting point to show where current research is at too.

3. **Quality of data analysis, visualization results, and discussion of insights gained**

As stated before, use of burst analysis is a good tool. Have you considered just count frequency of keywords and perhaps their correlation to geospatial distribution? Another one would be correlation with temporal distribution.

I love the idea of combining visualization results from different tools like SCI2 and Tableau. I see two temporal charts in your intermediate results, but cannot make out the difference other than starting and ending years. Please consider labeling them outside of SCI2 in your final presentation.

It is good to see how many types of data cleansing you had to go through before the final visualization. Perhaps you can add detail into domain specific insights you gained. Why did certain words/topics burst in specific time frames? Were there regional correlations? This will take the readers past the approach to Visualization to the realm of interpreting them. This initial discussion will lead readers to other insights on their own.

4. **Completeness and quality of validation and redesign**

The Temporal and Word Cloud visualizations are good introduction to this field. Your redesign to include the Term and Document frequency measures is a good add to the interpretation. It should help you find truly relevant words and phrases. Will there be before/after visualization or explanation?

5. **Overall quality of content, including the accuracy and completeness of information, the expressiveness and clarity in communication of ideas using text and**

visualizations, and the appropriateness of references to/attribution(s) for the work of others

The sequence in your methodology is clear and makes sense. The steps you have taken to pre-process the information is also lucid.

There are details in this presentation that are appropriate for the timing, but may not be relevant in the final presentation. For example, the details on how you performed burst analysis or created the Word Cloud are useful, but perhaps in an appendix. You could also simply provide a link to a Wiki or help site. In its current location, it is almost distracting.

Please consider adding titles and details for each of your visualizations, Temporal charts and Word clouds.

You have noted that there isn't much research in this area. You could consider pointing out how further research could be conducted in the area of scientometrics of interpretation. What did you want to achieve, but ran out of time before you could. Could someone create a database that has a lot of the information from SCOPUS, but organized and categorized in multiple ways? Can the process be streamlined so that additions can be plugged in, instead of being all re-evaluated each time? This viewpoint could bring you more citations than you have today.

Review 2

1. The documentation of data was exceptional. They were very clear.
2. Their workflow seemed appropriate and capable of being replicated. The tools seemed appropriate for their desired visualization.
3. They have a specific section for insights gained. The insights match the insights gained from looking over their visualizations and analysis of data.
4. They are clear on their future plans.
5. The references were clearly provided. The content was clear and thorough. They seem to have a handle on the project.

Review 3

1. Quality of data selection, cleaning, preparation, and documentation

The description concerning the data cleaning issues was nice. Overall quality of data selection, cleaning, preparation and documentation was done well.

2. Appropriate selection of tools, algorithms, workflows, and parameter values

The description of your workflow processes is thorough and seem appropriate for your project.

3. Quality of data analysis, visualization results, and discussion if insights gained

The insights discussion was good. Talking about how some words might be skewing overall results was a good insight provided by the visualizations that you generated.

4. Completeness and quality of validation and redesign

The validation and and redesign to account for words like interpret were appropriate. I think it was a good move to try and remove those words, because it doesn't bring much meaning.

5. Overall quality of content, including the accuracy and completeness of information, the expressiveness and clarity in communication of ideas using text and visualizations, and the appropriateness of references to/attribution(s) for the work of others

The overall content of the document was good. Some suggestions would be to add figure numbers and descriptions to the images that are displayed and make reference to them within the document. I like the idea of a world cloud, however, I think that the burst analysis is a stronger visualization to communicate your project goals. So, you might not need to have the word cloud. Have you conveyed your visualization plans to your client? If they like it, definitely continue, however if it is not needed you might reconsider it. Overall though good job.

Review 4

Some suggestions:

- You may want to change the word 'world' to 'word' in the Topic Analysis. I'm sure it must be a typo.
- Step 8: Insights – point no.2, Please correct the sentence to "to try and remove".

When you tried to obtain your data from Scopus database, how did you decide to use the search query terms you used for search? It would be nice to explain a little bit about that.

In the project title, you mentioned the year as **'1789-1861'** and while creating the temporal bar graphs, it shows that most bursting happened in the years 1990-2016. Now after reading this insight, it wasn't clear to me the time-period of the analyzed data Vs the visualized data. Please make this clear in your write up.

Wouldn't it be better to also filter 2-letter words? For example: In section 8, you mentioned the top 10 words for title. The last word is 'la' which I don't think carries any meaning. This can further reduce the noise in your dataset?

Also, is it necessary to have the topic analysis related to both 'title' and 'abstract'? For example: in the top 10 words in section 8, there are lot of words that are common for these 2 fields and the ones that are uniquely present in Abstract, do not necessarily relate to the domain of interpreting. Although it would be interesting to see the results for both fields after you apply TF-IDF, maybe it can prove me wrong. ☺

Other than these points, I think this team did a really great job at collecting the right data, filtering and analyzing it, using the appropriate visualizations and algorithms (TF-IDF great choice), validating and documenting everything step-by-step in a way that can be easily understood. This is a project where I think you have been able to apply all that we have learnt during the entire course. The interactive visualization you created is impressive provided you apply filters and legends to describe it, which you plan to as I understand. It would also be interesting to look at the future work that you have planned to complete in the next few weeks. Really great work!!

Review 5

1 Comprehensive Title: It states the duration for which the analysis is carried.
2 Expressive Title: Clearly demonstrates the intent of the research
3 Team members are mentioned. However, the affiliations are missing.
4 Great section. Each section explains the rationale behind the need of the visualizations. May it

be "Word Cloud", Citation Network, etc. Very nice legible visualization. The verbose that goes with the visualization clearly elaborates the fact. Some sections, clearly states the intent of the final project assets and had a place holder, which gives a reader an impression of what to expect in the final project e.g. Paper Citation Network.

5 Nicely explained the constraints the team ran into, which not many comparative reports are published yet. However, for the reports those were discovered, were nicely cited.

6 Neatly explained which data set to look for, what to look for, and what is the analysis that's being carried over. The high-level stats such as total number of records (2932), gives the reader an idea of how much data is taken into consideration while carrying out the analysis. All that major attributes were identified. The key attribute (Affiliation's ID) relationship with particular graph is clearly mentioned.

7 Well organized and documented on repeatable research steps to arrive at the visualization using the Sci2 tool. The visualization gives different perspectives such as word cloud by Title and by Abstract. However, the section's line spacing has taken up space, which probably could have been avoided which could helped in organizing the page breaks nicely e.g. there is some space between the bulleted points under #1, which if properly squeezed could have been squeezed. Proper page breaks could help in keeping the title (or intro) of the graph with the graph/visualization e.g. in Page 7 where is states "...Below are the visualizations for the same...." the visualization in the next page.

8 Key Insights were clearly articulated in bulleted format, which gives reader to focus one fact at a time.

9 Challenges, learnt while conducting analysis as a part of the project, are captured and elaborated perfectly. As an illustration: the correlation of the "Interpret" with other facts and rejecting that it's an outlier.

10 Well thought challenges and opportunities are captured. Team's approach to overcome some of the challenges are well documented.

Review 6

1. Quality of data selection, cleaning, preparation, and documentation

Nice description of data. This group described their processes for cleaning and creating visualizations at length, so it's safe to say that their documentation is good.

2. Appropriate selection of tools, algorithms, workflows, and parameter values

They seemed really mindful about using relevant data and input, and the development of an interactive map is a big score extending what they learned in the course into new areas.

3. Quality of data analysis, visualization results, and discussion if insights gained

There needs to be a larger discussion on the insights from the visualizations. More discussion can be had on the Co-Authorship Network attributes and how they are significant.

4. Completeness and quality of validation and redesign

They redesigned their word cloud visualization in a sensible way. It makes sense that the very words that they used to query Scopus for articles would be the ones that pop up primarily in the word cloud, so they made it more relevant.

5. Overall quality of content, including the accuracy and completeness of information, the expressiveness and clarity in communication of ideas using text and visualizations, and the appropriateness of references to/attribution(s) for the work of others

The insertion of demo charts at the beginning is unnecessary at this point and confusing for the reader. I also find the subject of 'scientometric interpreting' really vague, I know it isn't the group's fault, but I could use more explanation of the topic. The section "Why this project is important" begins to discuss the topic but could use more explanation and really doesn't say why it is important.

I'm guessing that the Burst detection visualizations were printed using the online converter from PostScript to pdf, that's what I did all semester and the visualizations don't look near as good as others that used Adobe. I highly recommend so that at least the words will be visible. Why did the changes in Burst detection take place?

Review 7

1. **Quality of data selection, cleaning, preparation, and documentation**
   Data selection seems to accurate and appropriate for every visualization. However, there is no documentation on data cleaning process. The document gives details about problems with data and results of dealing with it. However, it does not give details on the process of cleaning the raw data. Apart from this documentation has all necessary points.

	Data Selection	Cleaning	Preparation	Documentation
Visualization 1	The data selection seems to be accurate.	Information about data cleaning is missing.	The visualization is about burst analysis on “Title” and “Abstracts” and all the preparation is done on “Title” and “Abstract” using Sci2 via Lowercase, Tokenize, Stem, and Stopword preprocessing.	Documentation seems to fine as it provides all the necessary information
Visualization 2	The data selection seems to be accurate as the visualization is based on “title”	Information about data cleaning is missing.	Data Preprocessing has been done in Sci2 and visualization in Tableau. **Suggestion:** Could have done both in Tableau using custom functions feature of tableau	Documentation seems to fine as it provides all the necessary information about the process of making the visualization.
Visualization 3	Selection of data is elaborate. Required features with values have been mentioned precisely.	Information about data cleaning is missing.	Used Sci2. Preparation of data is nicely mentioned.	Documentation seems to fine as it provides all the necessary information about the process of making the visualization.
Visualization 4	Selection of data features is mentioned properly.	Information about data cleaning is missing.	Nice usage of R shiny. However, no information on the process of preparing the data	Documentation here is not too strong.

2. **Appropriate selection of tools, algorithms, workflows, and parameter values**

Selection of tools are varied and are nicely used. Algorithms, workflows are parameter values are well documented. However, same is missing for Visualization 4 (Proportional Symbol Map). But as the work is in the process, it is expected that they will add this final report.

3. **Quality of data analysis, visualization results, and discussion if insights gained**

	Data Analysis	Visualization Results	Insights
Visualization 1	The quality of data analysis is good. Have used correct algorithms to gain insight into data.	Visualization results are good but could have been better with the usage of clear and simpler title and colors.	Insights seem to appropriate to the visualization however they are mostly qualitative in nature rather than quantitative. For example, they could have talked about max and min busting terms.
Visualization 2	Data analysis seems to be correct but there is no sufficient documentation to support it.	Visualization seems to be appealing. However, no quantitative conclusion can be reached.	Insights are good. May find a way to embed them in visualization.
Visualization 3	An excellent elaboration of data analysis and algorithm used.	Visualization looks very good and includes all the necessary points	Insights are elaborate too. Nicely done.
Visualization 4	No documentation on data analysis process. Just a few parameters are mentioned on which data analysis is done.	Visualization is very appealing. The layering of data points is nicely done. However, the map does not give any proper insight or information. It looks like simple numbers embedded on the map.	No insights are mentioned.

4. **Completeness and quality of validation and redesign**
   The visualizations seem to fulfill the client requirements as mentioned on the Canvas. They have created all the visualization which addresses the requirements of “Students will create visualizations that focus on mapping the topic areas covered in the journal, drawing a historical evolution of topics in the field of interpretation, and network visualizations showing collaborations between authors.”.
   The work is done is iteratively and almost every detail is well documented. However, I would strong suggest revamping the temporal visualization and add more appropriate items to Proportional symbol map visualization to make it more readable.

5. **Overall quality of content, including the accuracy and completeness of information, the expressiveness, and clarity in communication of ideas using text and visualizations, and the appropriateness of references to/attribution(s) for the work of others**
   Overall quality is very good. Documentation is nicely done. The intent and intuition behind all visualizations are very clear. But, more documentation on Proportional symbol map visualization is required. The document contains all the attribution and references to others’ works. However, documentation has plenty of spelling and grammatical errors. I recommend proofreading the document carefully.