EduMap: Navigating a Learning Adventure

DANIEL PALAMARCHUK*, Virginia Tech, USA
XIAOHANG TANG†, Virginia Tech, USA
SONAL SATHE‡, Virginia Tech, USA
TONG WU§, Virginia Tech, USA
REXIME ABULIKEMU¶, Virginia Tech, USA

Directed study maps are often used in order to create an overlay of knowledge relations. However, these connections often have to manually be created by subject matter and IT experts, which, although are a ground truth, limit these user to the topics of the maps at hand. To address this, we propose EduMap, which takes advantage of GPT 3.5 in a few-shot setting to create a study map for any user-provided topic. We conducted two user studies to address various aspects of the map: (1) A quantitative study to test out the effectiveness of this map medium in the context of navigating information concerning COVID-19 in a expert-curated graph, and (2) A quantitative study to test out the accuracy and usefulness of GPT-generated graphs. Our results show that the map medium is effective, and, though the graphs produced are sensical, there is work to be done before it can be considered ground truth, with issues of vague topics and improper connections between topics arising. Altogether, our results establish a baseline for AI-generated study maps to be improved upon in the future with ground truth to create a more trustworthy software.

ACM Reference Format:
Daniel Palamarchuk, Xiaohang Tang, Sonal Sathe, Tong Wu, and Rexime Abulikemu. 2023. EduMap: Navigating a Learning Adventure. 1, 1 (December 2023), 11 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In the 21st century, self-study on various topics is increasingly common, often facilitated by internet searches. Whereas before students would take into account library sizes when picking a college, nowadays search engines like Google and Bing provide a wealth of information, processing this information can be cumbersome. Relevant data is scattered across different web pages, requiring users to juggle multiple tabs. Such a setup increases cognitive load and potentially hinders comprehension, with users potentially not knowing if they have the knowledge required to learn about a subject.

EduMap, our project, aims to address this challenge by consolidating knowledge into a user-friendly system, simplifying the learning process. By using a LLM to generate personalized study maps, EduMap creates a unique

*Contributed to the background and literature review, and system design and demo
†Contributed to the system design and demo, and the development of the user study
‡Contributed to the background and literature review, and the development of the user study
§Contributed to the background and literature review, and system design and demo
¶Contributed to system design and demo, the development of the user study, and analyzing the study results

Authors' addresses: Daniel Palamarchuk, d4n1elp@vt.edu, Virginia Tech, USA; Xiaohang Tang, xiaohangtang@vt.edu, Virginia Tech, USA; Sonal Sathe, sss20a@vt.edu, Virginia Tech, USA; Tong Wu, tongw@vt.edu, Virginia Tech, USA; Rexime Abulikemu, rexime@vt.edu, Virginia Tech, USA.

Manuscript submitted to ACM

educational journey for each user, linking all necessary resources and information logically. This approach addresses critical issues such as information overload and the difficulty in linking theoretical knowledge with practical applications, with the goal of providing an efficient, personalized, and comprehensive learning experience. Initial feedback has shown that EduMap improves learning efficiency and user engagement, demonstrating its potential impact on educational practices.

2 RELATED WORK

In our development of EduMap, we have drawn inspiration from foundational works in the field of HCI and personalized learning. Nelson [6]'s visionary work in "Computer Lib/Dream Machines" advocates for educational reforms and user empowerment, paralleling our goals in EduMap. This seminal text emphasizes the importance of designing computer experiences as creative media with the user in mind, an ethos central to our project. Nelson's vision of a radical educational approach in which the user is in control over how they learn is shared by us as well. We also take inspiration from educational methodology that prioritizes individual initiative

Complementing this, studies by Lee and Segev [4] and others have explored knowledge maps and individualized learning paths in e-learning, reinforcing the need for adaptive, personalized educational tools. These prior works collectively underscore the significance of our endeavor in creating EduMap, aiming to revolutionize the learning experience by aligning with these pioneering concepts in HCI and education. Prior work in this field has demonstrated previous attempts for use of maps in online learning, in particular by Liu et al. [5], Lee and Segev [4], Sumner et al. [8], and Chen and Fu [2]. This shows that maps have been conceptualized as a learning tool by computer scientists with interests in the use of education and technology. Many of these systems are set up with a ground truth by subject matter experts who can link together the topics in the most logical manner that a user can follow.

Other guiding principles for our work involve drawing from knowledge and methods in the social sciences and natural sciences. More specifically, for the spread and linkage of information, we consider the use of social network analysis as a guide to demonstrating how concepts are interconnected. Nodes in such an analysis can be information points or individuals, actors, etc. in a given network. These nodes have linkages that represent their interconnections, giving learners a framework to understand how information—or disease—can be spread. For example, in the vein of spread of disease, the SEIR model (which stands for susceptible, exposed, infected, and recovered) was also used as a guide for flow of information—in much the same way the disease itself flows via modes of infection, which was discussed by Carcione and colleagues. This is particularly useful for the COVID-19 arm of our EduMap.

We intend to use the GPT 3.5 API from OpenAI [1] in order to generate our custom maps. GPT 3.5 is a type of transformer model [9] which is the Large Language Model (LLM) behind Chat GPT. A LLM is a machine learning language model characterized by its wide variety of use cases. This is possible due to the massive amount of training data that is required to train such models in order to tweak billions (and now trillions) of parameters to be able to operate on text of most sizes to achieve an outcome. The transformer is a type of LLM that is characterized by its use of an attention mechanism in order to figure out which tokens (words) are most related to each other, which allows for a more accurate prediction and outcome to be made. GPT 3.5 is a type of decoder-only transformer, which means that it repeatedly tries to predict the next token until a full piece of text is created. In our case, we provide a few example knowledge graphs to GPT 3.5 of the knowledge graphs we want the model to mimic the structure of (called few-shot prompting) in order to encourage GPT 3.5 to generate knowledge maps of the type we want. This allows us to prompt GPT 3.5 to generate knowledge maps that we can then use to send to a graphical frontend that then creates the actual knowledge graph.

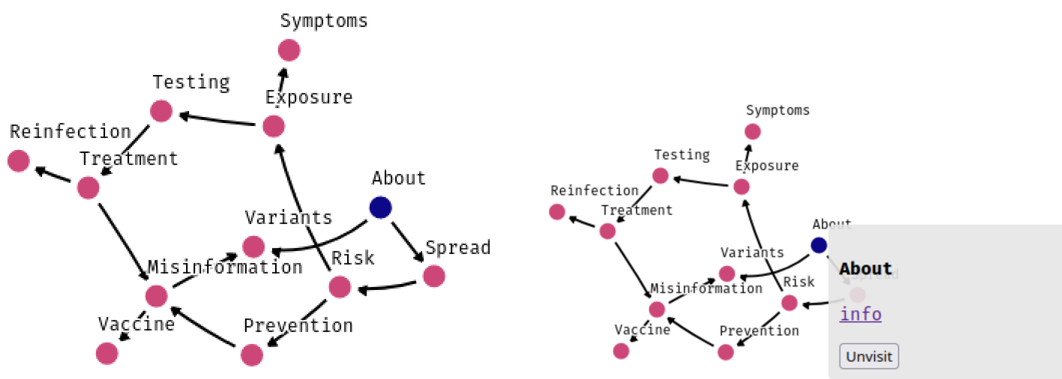

Fig. 1. Left: An EduMap linking together topics about Covid-19. Right: The same map hovering over the "About" node

3 DESIGN AND RATIONALE

3.1 Frontend graph

The graph provides a frontend through which to view the graph json objects that describe topics and relationships between topics (see: Figure 1 and 3.2). It was built using the d3.js package and is a force-directed directed graph. The graph is composed of nodes that can be marked blue for read or magenta for unread. Each node is labeled by a topic. Hovering over a node creates a window that allows you to select to visit or unvisit the node, which will change its color. If provided beforehand, there will also be a link that leads to another page about the current topic. Directed nodes are meant to signify prerequisites in logic acquisition (i.e. "addition → multiplication" means that addition is a prerequisite to learning multiplication). There is also the ability to feed an array of strings that then highlights certain nodes. This is meant to be used to be used with the Key Term Scraper (3.3) so that it can supply the Map with terms that an article is about. These key terms will then be highlighted in the graph, allowing a user to easily see what their current article is about as far as the graph is concerned. Overall this creates a flexible graph structure that can be fed nodes and links from the backend.

3.2 Graph Generation

There are a handful of pre-generated graphs that the user can use. The chosen topics are "Covid Information," "Simple Linear Regression," and "K-Pop" and these have been filled out with connections, nodes, and links to external websites. The user also has the option to create their own knowledge map by selecting "Generate own map" (see: Figure 2), entering their topic into the provided text field, and pressing "Generate Map." This topic will then be sent with a custom input prompt and a context of two pre-generated graph json objects to the GPT-3.5 API from OpenAI [1], which will then return a graph json object that will be fed into the frontend graph interface. This is done using the API's `response_format={"type": "json_object"}` option to guarantee json object creation and the provided context aims to inform the model of the general structure of the returned object. For the purposes of our project we opted to use the GPT 3.5 API due to its greatly reduced cost, but using GPT 4 API is also an option for potentially improving the graph creation.

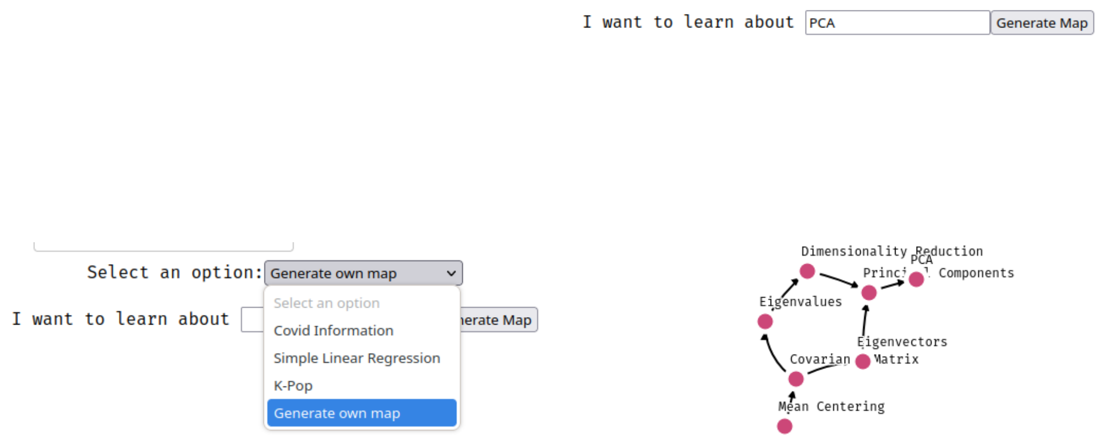

Fig. 2. Left: Menu of possible graphs. Right: Knowledge graph of Principle Component Analysis (PCA)

3.3 Key Term Scraper

To allow users to inspect the information containing in the given URL associated with a node, we designed and developed the feature *Key Term Scraper*. There are mainly two parts in the *key Term Scraper* process. By sending HTTP requests, the system first obtained the text content in the given URL. Then, instead of browsing through the entire content along the website to catch the key meanings of a particular node and its connections with others, users will be able to quickly make sense of the node by examining the extracted top keywords/key phrases. We achieved the keyword extraction by utilizing the contextualized word embeddings in a Large Language Model(LLM)[1] [3]. By computing and comparing every words' word embedding, we selected the top n words which have the most similar embeddings with the whole document.

4 STUDY PROCESS

We conducted two user study to evaluate our system. The primary aim of the first study was to assess the accuracy and relevance of AI-generated maps, particularly in areas where participants possess expertise. The second study aimed to measure the effectiveness of a specific Pre-generated map, the "COVID-19 map", in enhancing user understanding and knowledge.

Participants were recruited from a convenience sample of students in a given computer science course. The first study included a total of six participants, while the second study comprised four individuals.

4.1 Study 1: Evaluation of Custom Maps

In this study, participants were invited to evaluate maps related to their field of knowledge. First, they were asked to create a map using our system, focusing on a topic in which they are knowledgeable. After creating the map, participants were required to complete a survey. The survey aimed to gather their insights on several key aspects of the map: **Relevance of Nodes**, **Correctness of Connections**, and **Topic Concreteness and Learnability**. Additional comments and feedback about the overall utility, user experience, and suggestions for improvements were also collected.

[1]Particularly, we used the KeyBERT.

A copy of the maps and user evaluation of them can be found attached to the end of this report (see: Table 1 and Figure 2)

4.2 Study 2: Assessing the Effectiveness of the COVID Map

As part of our study process, we developed user surveys to be administered in a pre-post test format in order to determine the knowledge outcomes for our users after using the system we developed on the topics we delineated: namely, COVID-19, simple linear regression, and KPOP. Due to the time limitation, we only used COVID-19 map in our first study. Participants were required to answer all questions on the pre and post-survey–that is, before and after use of our EduMap system. Additionally, the post-test survey encompassed mandatory open-ended questions, which delved into the system's helpfulness, suggestions for improvements, and any concerns regarding accessibility. A copy of the surveys, both pre and post, can be found attached to the end of this report (see: Table 2).

5 RESULTS AND DISCUSSION

5.1 Quantitative Results

5.1.1 Study 1: Evaluation of Custom Maps. The quantitative analysis of the first study focused on evaluating the custom maps using a Likert scale ranging from 1 to 5, assessing Relevance, Correctness, and Concreteness & Learnability of the topics.

- Relevance: Average relevance score across all topics was 3.83, indicating a generally high level of topic relevance in the maps.
- Correctness: The correctness of connections showed an average score of 4.0, suggesting logical and accurate linkages between nodes.
- Concreteness: The concreteness, reflecting the learnability of nodes, averaged at 4.33, indicating that most topics were presented in a clear and digestible manner.

Overall, the quantitative data suggests that the custom maps were effective in representing relevant and accurate information in a concrete, learnable format.

5.1.2 Study 2: Assessing the Effectiveness of the COVID Map. The second study assessed the effectiveness of the predefined COVID-19 map. This study involved both an experimental group (using the COVID map) and a control group (using general internet access).

- Pre-Survey Correctness: The average correctness score in the pre-survey was 8.25/10 for the experimental group and 8.0/10 for the control group.
- Post-Survey Correctness: In the post-survey, the experimental group showed an increased average score of 9.5/10, whereas the control group had a score of 8.67/10. This suggests a significant improvement in understanding for the experimental group.

5.2 Qualitative Results

5.2.1 Study 1: Evaluation of Custom Maps. The survey elicited insightful responses regarding the content of the maps, highlighting areas of strength and opportunities for improvement:

Comprehensive Coverage with Room for Improvement in Connections: P1 appreciated the avoidance of overly broad topics like "intro to linear algebra" but noted the odd absence of connections between key concepts such as eigenvalues and eigenvectors. They also suggested that while topics like orthogonality were well-covered, the map could benefit from more foundational linear algebra concepts to better scaffold the learning process.

Need for Better Organization and Connectivity: P2 highlighted a desire for more structured organization, suggesting that different branches could represent distinct topics. This participant observed a division in the map, leading to a lack of cohesion. P3 also pointed out the necessity of creating connections between different branches of the map to enhance the coherence and flow of information.

Clarity in Logic and Explanation of Connections: P4 found all nodes relevant to Python but expressed confusion over the logic and connections within the map, indicating a need for clearer explanations of what the edges and arrows signify. P6 also mentioned that "The logic need to be clearer".

Positive Feedback on Content Comprehensiveness: P5 expressed satisfaction with the map, noting that it encompassed nearly all aspects of the topic they wanted to learn, indicating a high level of content completeness.

These responses collectively suggest that while the maps were largely successful in covering relevant topics, there is a clear need for improvements in organization, clarity of connections, and logical flow to enhance the overall effectiveness and user experience.

5.2.2 Study 2: Assessing the Effectiveness of the COVID Map. In the part, we gathered participant feedback to evaluate the effectiveness of our EduMap system compared to traditional internet-based research for learning about COVID-19. Responses from both the experimental group (users of our system) and the control group (users relying on internet access) were analyzed. The analysis focuses on three key aspects: the perceived helpfulness of the system, suggestions for improvement, and concerns regarding accessibility.

System Helpfulness: In the control group, P2 noted that their pre-existing knowledge was neither significantly challenged nor expanded, implying a lack of depth or novelty in the information available via general internet searches. P3 echoed this sentiment, pointing out the absence of systematic and structured knowledge when using conventional internet resources. As for the experimental group, P1 indicated that our system was not as helpful as they expected, suggesting a potential gap in the system's comprehensiveness or user interface.Conversely, P4 acknowledged the system's efficacy in organizing content, yet expressed a preference for an all-inclusive learning experience within the platform, minimizing the need to refer to external websites.

Suggested improvments: P1 suggested incorporating a feature to allow users to pose questions, indicating a desire for interactive and responsive learning. P4 proposed the utilization of Language Learning Models (LLM) to succinctly summarize key concepts in the nodes, emphasizing the need for clarity and conciseness in information presentation.

Accessibility Concerns: Our participants did not identify any specific concerns regarding accessibility.

5.3 Future Work

We want to publish this work, but before that happens there remains a bit of work to fully flesh out the EduMap. We will describe future work, most of which we will aim to implement before submitting the paper.

First there are some elementary things that can be done as far as prompt tuning. Currently we have created a lengthy prompt and generated a few-shot context using two of our handmade maps. We believe it can be improved with either more examples, more detailed examples, or a more detailed prompt. We also have not explored in-depth the difference between GPT-3.5 and GPT-4 generated maps, though our current system can easily switch between the two.

There are also a few things left to be desired in terms of user interface. General formatting aside, there are three user interface features that would be useful. The first would be relating the keyword finding feature to the map and actually highlighting the nodes rather than just printing the most significant words in an article. The second would be to allow the user to interact with the generated JSON object. The user should be able to download their map and reupload it later in case they come back to it. The third, and biggest, UI feature that we look to implement is to turn the EduMap into a web plugin that can pop out in the corner. That way the user would not have to go to another web page to interact with the map. We hope this change in medium will also allow for the automation of keyword extraction, allowing the important words and subjects in the web page to be highlighted on the map automatically.

Finally, there is the issue of ground truth. Right now we are just hoping that the generated map is correct. There is no direct human interference, just some few-shot prompting to guide the map along the process. We also have no way of connecting any AI generated map to any web pages, let alone to any that have been verified to be trustworthy. This is a pretty big gap that we currently have not thought of an idea plan for surpassing. Another component or two would most likely have to be added to our system to accommodate the storage of, the entry of, and the rating of information. This most likely would have to be addressed in a different work.

There is also a bit to be done in terms of user studies, particularly in relation to disabled user studies. We did not test the compatibility of our EduMap with accessibility softwares yet. There are specific reported barriers to receipt of information in users with disabilities who must use assistive technology such as a screen reader Sharif et al. [7]. Further work in the form of a user study should assess whether our system is compatible with traditional screen readers such as JAWS or NVDA. This will be a useful tool if such tools are available to the blind or visually impaired community for ensuring equitable access and outcomes for all people.

6 CONCLUSION

Large language models can be used in order to generate directed study maps. This will allow users to quickly be able to learn about a topic and figure out what topics different websites address. On the baseline level, there are several loose ends to tie together, and a looming question of how to establish a ground truth in the context of AI-generated study maps. Our user studies provide an elementary check of whether such a path is even worth pursuing both in the context of the usefulness of the study map medium and whether the current state of the art in LLM's is capable of creating sensible graphs and realize the requisite-relationship between topics. These relationships can be formed to a satisfactory degree, but additional work needs to be performed to ground this generation in ground truth to provide more accurate and descriptive graphs. Further work should also involve beta testing for screen reader users, in order to gain insight into accessibility of the EduMap.

ACKNOWLEDGMENTS

Thank you to Dr Yan Chen for instructing and guiding us through this process.

REFERENCES

[1] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner,

Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. *CoRR* abs/2005.14165 (2020). arXiv:2005.14165 https://arxiv.org/abs/2005.14165

[2] Tianying Chen and Hongguang Fu. 2009. Learning Navigation Map on Ontology. *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing* (2009), 169–174. https://api.semanticscholar.org/CorpusID:13986503

[3] Maarten Grootendorst. 2020. KeyBERT: Minimal keyword extraction with BERT. https://doi.org/10.5281/zenodo.4461265

[4] Jae Hwa Lee and Aviv Segev. 2012. Knowledge maps for e-learning. *Computers & Education* 59, 2 (2012), 353–364. https://doi.org/10.1016/j.compedu.2012.01.017

[5] Sannyuya Liu, Guo Dongpo, Jianwen Sun, Jie Yu, and Dongbo Zhou. 2020. MapOnLearn: The Use of Maps in Online Learning Systems for Education Sustainability. *Sustainability* 12 (08 2020), 7018. https://doi.org/10.3390/su12177018

[6] Theodor H. Nelson. 1974. Computer Lib. (1974), 301–338. https://www.newmediareader.com/book_samples/nmr-21-nelson.pdf

[7] Ather Sharif, Sanjana Shivani Chintalapati, Jacob O. Wobbrock, and Katharina Reinecke. 2021. Understanding Screen-Reader Users' Experiences with Online Data Visualizations. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) *(ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 14, 16 pages. https://doi.org/10.1145/3441852.3471202

[8] Tamara Sumner, Faisal Ahmad, Sonal Bhushan, Qianyi Gu, Francis Molina, Stedman Willard, Michael Wright, Lynne Davis, and Greg Janée. 2005. Linking learning goals and educational resources through interactive concept map visualizations. *Int. J. on Digital Libraries* 5 (03 2005), 18–24. https://doi.org/10.1007/s00799-004-0112-x

[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. , 11 pages.

A CUSTOM USER STUDY RESULTS

Custom topic generation, using likert scale ranging from 1 to 5:

Topics	Relevance	Correctness	Concrete
Principle Component Analysis	3	4	5
Cooking	4	4	4
Music	3	4	4
Python	4	3	4
NBA	5	5	5
BERT for Classification	4	4	4

Table 1. Table of all topics generated by the user study, as well as their accuracy metrics based on a Likert Scale of 1-5. Relevance is a measure of how relevant the generated nodes are. Correctness measures whether the connections made are made so in a logical manner. Concrete is a measure of how learnable a node is (ie if it's vague, such as "Introduction to Linear Algebra")

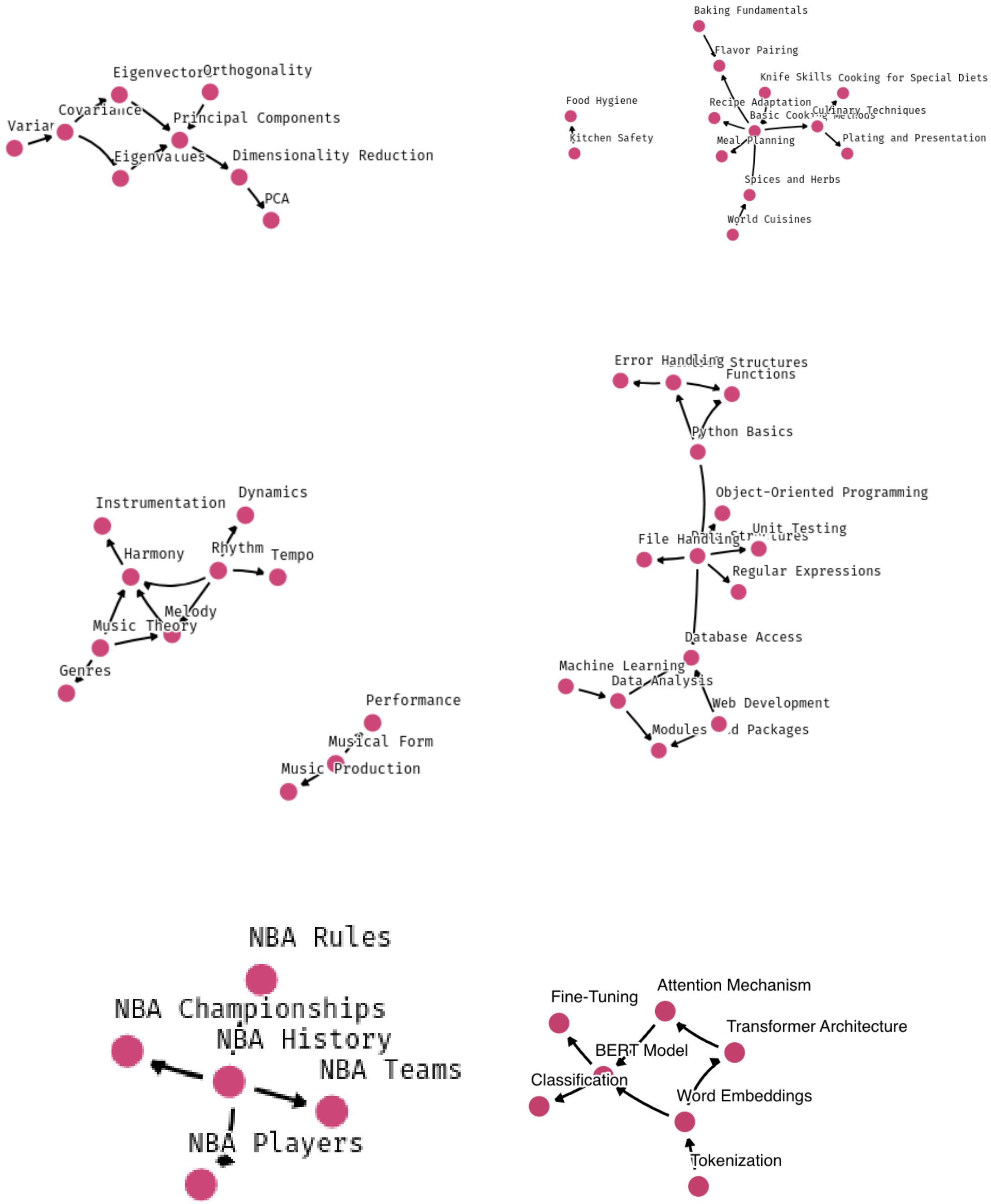

Fig. 3. Maps generated by the user study participants

B COVID STUDY RESULTS

Participant ID	Condition	Pre-Survey Correctness	Post-Survey Correctness
P1	Experiment	7/10	9/10
P2	Control	9/10	10/10
P3	Control	7/10	7/10
P4	Experiment	10/10	10/10

Table 2. Results of the second user study, showing how effective using a map is in figuring out information. The Experiment condition represents a user with access to a predefined COVID mind map. The Control condition represents a user with general internet access.