Open Personalized Navigation on the Sandbox of Wiki Pages

Chuankai An*
chuankai@cs.dartmouth.edu
Department of Computer Science, Dartmouth College
Hanover, NH, 03755

Daniel N. Rockmore†
Department of Computer Science and Mathematics,
Dartmouth College
Hanover, NH, 03755

ABSTRACT

In this paper we present a proof-of-concept of a visual navigation tool for a personalized "sandbox" of Wiki pages. The navigation tool considers multiple groups of algorithmic parameters and adapts to user activity via graphical user interfaces. The output is a 2D map of a subset of Wikipedia pages network which provides a different and broader visual representation – a map – in the neighborhood (according to some metric) of the pages around the page currently displayed in a browser. The representation schema includes the incorporation of a kind of transparency in the algorithmic parameters affecting the presentation of the landscape visualization, which in turn enables the delivery of a personalized canvas, designed by the user. A case study shows the combination of four different sourcing (i.e., identification and extraction of the neighboring pages) rules and three layouts over the same Wikipedia subnetwork. The basic schema is readily adapted to other search experiences and contexts.

CCS CONCEPTS

• **Information systems** → **Wikis**; • **Human-centered computing** → *Graph drawings*; • **Computing methodologies** → Network science.

KEYWORDS

Wikipedia; Network Visualization; Transparent Navigation

ACM Reference Format:

Chuankai An and Daniel N. Rockmore. 2019. Open Personalized Navigation on the Sandbox of Wiki Pages. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3308560.3316755

1 INTRODUCTION

Wikipedia is an important source of information [30]. For many people, going to Wikipedia is just the first step in an information search task. A standard search trajectory would then take place, realized as a sequence of clicks, effectively something of a constrained, yet still "random" walk from the Wikipedia page starting point, alike at least in spirit to the "random surfer" model that gave birth to PageRank [9] (and Google) whose depth and penetration of the space of relevant webpage resources can and does depend on many contingent characteristics. Regardless of the starting point, in this click-by-click revealing of the relevant (one hopes!) knowledge, it may be easy to miss or get distracted away from the original motivation for inquiry. More broadly, in such a blinded navigation the user is unaware of the way in which the webpages she visits relate to one another. Inadvertently she may be stuck in cul-de-sac of narrowly defined information or strayed very far from her initial search goal. It is with this in mind, that we take on and suggest an alternate option, one that promotes a notion of visual search, that presents a map-like visual summary of the page neighborhood, thereby possibly promoting a broader field of vision in the search process and highlighting different criteria for navigation.

We propose a visual navigation tool for Wikipedia based network visualization, which allows users to select their preferred query target as the root page, and visualizes the local "sandbox" of related pages in the form of a 2D map on a "canvas" (viewing platform). Our motivation arises from the user experience of standard query on Wikipedia. Figure 1a shows a list of related pages when the query input is not exactly matched to a Wikipedia page. More often, Wikipedia will load a new page in the browser or redirect it to a similar one as Figure 1b. We believe a visual navigation tool might be a useful broadening of our verified knowledge boundary during browsing better than such a list of results or unexpected redirection.

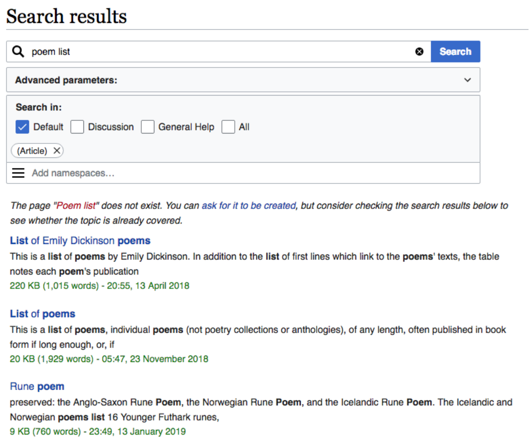
For example, a series automatically updated maps of the surrounding pages in the network space could display a broader view of the information space that both illustrate the distance among those nodes (pages) while also providing some sense of context for the material on the page. A visual navigation tool based on Wiki page networks could also facilitate a user's understanding of the local network structure, and would bring more transparency to the query results. While the network structure articulates the link relationships between pages, the use of other kinds of metadata (from the user and other users as well from the webpages) raises the possibility of creating a non-link distance structure (metric) for the neighborhood, and with that, new possibilities for display and user interaction. User response to the 2D Wiki map might also offer interaction data for user-behavior oriented research projects. While the focus of this article is on the Wikipedia environment, the general framework of user-controlled network navigator is not limited to Wikipedia corpus. For example, the dynamic graph visualization may also work as a recommender for online shopping or the World Wide Web as a whole.

In this paper, we present some initial ideas around the design of a personalized visual navigation system on Wikipedia. Generally, the data flow starts from a seed Wikipedia page. A "sandbox" of related pages is defined by a distance threshold on the Wikipedia

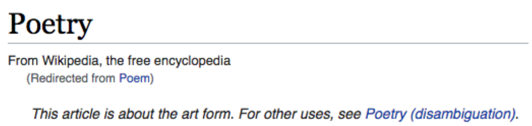
*Corresponding Author
†External faculty of the Santa Fe Institute, Santa Fe, NM, 87501

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3316755



(a) Query result: a list of related Wikipedia pages.



(b) Redirection to a new page.

Figure 1: Interfaces of query results on Wikipedia.

page network, in which a directed edge from page A to page B means page A cites B in the HTML source file. The navigator will interact with users to get the desired algorithmic parameters of the visualization to be personalized. Behind the user interface, diverse algorithms implement the tasks of node filtering, coordinate computation and edge selection due to a limited size of screen. Though researchers have intended to diversify the user experience of Wiki with visual effects [13, 24, 27], our contributions include the possibility of real-time updated visual navigation that responds to users browsing in the open space of Wikipedia pages, and the fact that nodes on the screen are determined by the personalized algorithmic parameters directly set by users. As users are often only aware of the pages that comprise their browsing paths only relatively "blind" to any "surrounding" ones, we hope our design of such an immersive visual navigation would make for a more useful Wikipedia search experience.

Section 2 surveys related work about Wikipedia tools and network navigation. In Section 3, we propose a data pipeline of the visual navigation system and the options of three groups of algorithmic parameters. Section 4 provides the details of implementation. Section 5 illustrates the final output of network visualization of a query word, as well as the comparative analysis of the effects of algorithmic parameters. Section 6 briefly describes the design of user study. Finally, Section 7 concludes the initial findings and directions of further development.

2 RELATED WORKS

Several projects have focused on Wikipedia navigation, as relates to efficient browsing. Lamprecht [19, 20] discussed the influences of Wikipedia navigation policies and the structure of Wikipedia pages network. Odor [24] presented the evolution of Wiki hyperlink networks to aid navigation and understanding. Another automatic tool [27] aimed to generate info-boxes for Wikipedia pages from a Wikipedia knowledge graph. In a different direction, Leibon et al. [22] show how the Wikipedia pages around a given topic – e.g., mathematics – can support a metric and thus structure of a hyperbolic geometry and with that, enables the construction of geodesics (paths in the Wikipedia space) that optimally guide a user's viewing experience (the use case of the paper is the MathWikipedia). Figshare [1] provided content-based embeddings learned from Wiki corpus as the navigation vectors on a 2D plane. Cartograph [29] enables the presentation of a vast map of Wikipedia pages with the embeddings learned from neural networks. The last of these differs from our proposed navigation schema which when fully realized would allow users to change the visualization and any underlying metric supporting the visualization.

Also related is work on semantic annotation and some applications of collaborative filtering applied to Wikipedia data. IkeWiki [28] and SweetWiki [10] made the inherent structure of a Wikipedia page accessible to users and computing machines via annotations derived from semantic methods (e.g., RDF and conceptual graphs). A visual analytics framework [13] illustrated how editors could work together for a public visualization of Wikipedia data.

Researchers also have been working on diverse kinds of Wiki tools to improve knowledge transfer and user experience. Harder et al. [17] designed a new measure to model and display the degree of "verifiability" of a Wikipedia page and implemented a demo in a Chrome browser extension. A visual article development tool [15] explored editor interaction history to deal with disagreement. Balaraman [7] proposed a new metric to describe the relative completeness of Wikipedia data. Gundala [16] reported the initial progress about predicting hyperlinks between pairs of non-connected pages that are helpful for search navigation. WikiTrails [26] provides a tracking system of visited Wikipedia pages to facilitate the understanding of Wikipedia content structure. Omnipedia [8] visualized multi-language editions of the same Wikipedia page via colorful circles in different sizes based on an article alignment algorithm, but it ignored network analysis. SuggestBot [11] proposed a link recommendation framework to match people with suitable editing tasks on Wikipedia.

Lastly, there is now a growing body of experimentation with the digital interfaces for searching and exploring traditional information materials, specifically for the interaction with libraries. An interesting example of this is the Harvard Stacklife project[1], that aims to bring back to online library search the missing – and bemoaned – loss of the serendipity of browsing the stacks that occurs when going to retrieve a book of interest. It is in the spirit of such exploratory serendipity that we present the work in this paper. Another interactive graph [21] allows users to set up the weights of link structure measures and textual similarity for a 2D map of legal documents, from which researchers explore how new opinions influence the search behavior of judges and litigants and thus affect the law.

[1] http://stacklife.harvard.edu

3 DATA PIPELINE

In this section, we present the data pipeline starting from a seed Wikipedia page to the visualization of related pages. The following steps combine user interface and algorithm-based computation for a personalized and transparent visualization. In this case, transparency means that users would know how the thing they are looking at is made, while customizability means that users have the power to directly change the input parameters in our proposed navigation system. We hope the customizability and the idea of network space visualization could make the navigation system more transparent.

Wikipedia Seed Selection. The "seed page" represents the user-defined center or starting point for a neighborhood of Wikipedia pages of interest for a given topic.

Wikipedia pages crawler. If a Wikipedia page cites another one in the main context, they are a pair of linked nodes in the network of Wikipedia pages. Among the billions of Wikipedia pages, in this preliminary proof-of-concept study we work with only a very small subset and limit the range of the crawler with some threshold on the distance between a candidate page to the seed.

Parse sandbox structure. Given a seed page, the downloaded subset of Wikipedia "nearby" pages is our "sandbox". Our processing and analysis do not edit their content. This step aims to build the network of the observed local Wikipedia pages around the seed. There are various options for thresholding the neighborhood (e.g., all pages within some fixed linked distance of the seed). A Wikipedia (sub)-network not only contains the surrounding nodes, but also the edges among them. Here we define the weight of an edge between two nodes as the number Wikipedia pages that cite both of them. For a given node, this weight enables a sorting of its direct neighboring nodes (i.e., with a distance of one) in the network.

Set algorithmic parameters. Different from traditional digital-art based user interface (UI) design, here we propose a framework for algorithmic visualization for a (sub-) network of Wikipedia pages. It contains three groups of parameters, set by users, to make the visualization more transparent:

(1) The rule of nodes sourcing and ranking. Here we apply four different methods: semantic content-based similarity, graph structure, collaborative filtering of users browsing and an overall *PageDist* (cf., [21]) metric derived from link and content similarity. The navigation tool could display a limited number of nodes within a canvas, compute the internode distance matrix and then use that for node placement/visualization. For example, we compute the distance between two Wikipedia page titles according to the word vector representation GloVe [18]. It is also possible to sort the neighboring nodes with some network science features (e.g., node degree, PageRank centrality). To simulate collaborative filtering, we assume that the frequency of concurrence on a third page is proportional to the probability of users preference for the two pages. The PageDist [21] metric considers the commute distance in a transition matrix, which is derived from both in-out link structures and semantic similarity of texts. All the sourcing methods are independent with the link structure of the downloaded pages in the sandbox. In the previous step of crawling, we have applied a distance threshold. If the navigator serves the whole Wikipedia network without a radius in the crawling step, those sourcing methods could narrow down the range of candidates as well.

(2) Definition of "nearby" nodes on the canvas. According to the selected sourcing and ranking method for a small number of candidate nodes to display with the seed on a canvas, we could sort all surrounding nodes according to their feature similarity to the root page. Therefore, for any two nodes appearing on the canvas, their relative proximity to the center node would be in accordance with their rankings in the sorted list of their feature similarity to the root page. Another definition of proximity comes from the result of a user preference predictive model, where the neighboring nodes with a larger probability of preference will be closer to the center node. Those two settings may be in accordance with each other, but sometimes a user might explore some new pages rather than the most similar one. This might be especially true when looking for information about particularly divisive or "charged" subjects.

(3) Layout of nodes. We try to locate the current Wikipedia page at the center of a canvas, except in the case of using the 2D multidimensional scaling [12] (MDS). If the second setting (i.e., closeness to the center) is defined by the feature similarity from the sourcing rule, the surrounding nodes should follow the order of their distances to the root page in the feature vector space. We implement spiral and spectral layouts to adapt to a ranking of the selected nodes. Both layouts point out the "close" neighbors and grant users the access to adjust parameters for their desired neighboring nodes. The assigned node coordinates in MDS match the idea of preserving between-node distances rather than the arbitrary design of spiral/spectral layouts.

Visualization. As a part of computational visualization, several factors might limit the actual effects, such as the size of the available screen (i.e., "canvas") to present the Wikipedia network, the number of pixels in a fixed size canvas (i.e., resolution), and the suitable number of nodes/edges. In addition, the location of a node should follow the general direction defined by the layout. Therefore, except for the MDS option, we first compute coordinates in a polar system, then transform it to the 2D plane coordinates. This step needs the help of an external visualization package which places nodes on a 2D plane at the accurate coordinates, so that users could present a non-standard yet desired layout on a canvas. Though many algorithmic terms are introduced in this tool, non-expert users could compare the differences in visualization and – with a little experience and/or training – adjust parameters for their preferred result.

Update the Sandbox. We hope the navigation system could incorporate user activities, such as hyperlink clicks, revisiting a page or long-time browsing. Once monitoring the above activity, the system should return to the second step to crawl some new Wikipedia pages, and update the Wikipedia network with the following steps, such as a new seed page and new selected neighbors. In this way, the navigation tool could extend to an open Wikipedia space and gradually collect user preference records for other personalized services on Wikipedia.

4 SYSTEM IMPLEMENTATION

In this section we briefly introduce the implementation of a proof-of-concept visual navigation system for Wikipedia pages which follows the data pipeline defined in Section 3. It is developed in Python because of multiple efficient existing programming packages. Selenium [23] enables the detection of the current URL in a browser. Tkinter [4] offers the UI modules (e.g., input frames and radio buttons) for Wikipedia seed confirmation in Figure 2 and algorithmic parameters settings in Figure 3. With the input of a seed Wikipedia page, Urllib [5] downloads all the cited Wikipedia hyperlinks in the seed page with the help of a regression expression matching function. BeautifulSoup [2] facilitates the analysis of hyperlinks in local HTML files so that we could build the network of Wikipedia pages in the “sandbox”. NetworkX [3] could place a node at the given coordinates in a 2D plane, so the navigator displays the same layout as what users choose (see Figure 3). Here we show examples of the visualization.

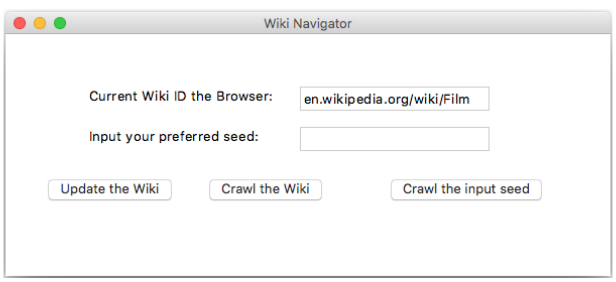


Figure 2: The user interface of Wikipedia seed selection. Users could input a seed or select the current one in a browser.

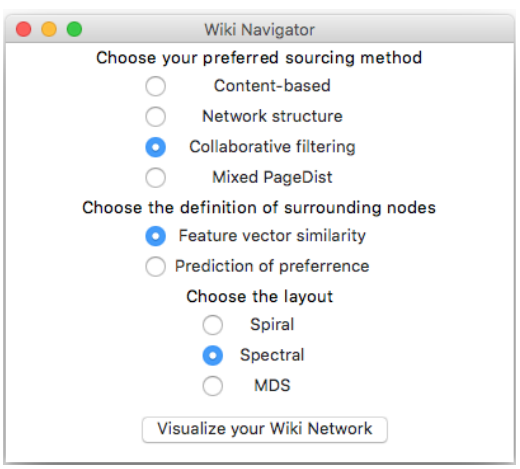


Figure 3: The user interface of algorithmic parameter selection for network visualization.

The implementation we describe above thus assumes an offline deployment to determine a subset of edges and links on the screen.

1	Film	36	Classical_Hollywood_cinema
2	Screenplay	37	Cult_of_personality
3	Documentary_film	38	Public_relations
4	Television	39	Principal_photography
5	Film_production	40	Color_motion_picture_film
6	Film_genre	41	Spectacle_(critical_theory)
7	Short_film	42	Script_breakdown
8	Art_film	43	Videography
9	Movie_studio	44	Main_Page
10	Independent_film	45	Film_industry
11	Sound_film	46	Cinematography
12	Silent_film	47	Special_effect
13	Soundtrack	48	Internet
14	Science_fiction_film	49	Visual_effects
15	Film_history	50	Post-production
16	Film_director	51	Storyboard
17	Film_editor	52	Film_score
18	Feature_film	53	Film_crew
19	Animation	54	Sound_effect
20	Film_release	55	Guerrilla_filmmaking
21	Film_editing	56	Filmmaking
22	Pitch_(filmmaking)	57	Streaming_media
23	Digital_object_identifier	58	American_Dream
24	Concentration_of_media_ ownership	59	Film_treatment
25	News_broadcasting	60	Media_event
26	Shooting_schedule	61	Docufiction
27	Occupation_(protest)	62	Culture_industry
28	Cinema_of_the_United_ States	63	Managing_the_news
29	Crowd_manipulation	64	Strike_action
30	Recuperation_(politics)	65	United_States
31	International_Standard_ Book_Number	66	Daily_progress_report
32	Daily_production_report	67	Mainstream_media
33	Breaking_down_the_script	68	Screenwriting
34	Demonstration_(protest)	69	Political_satire
35	Roadshow_theatrical_release	70	Bollywood

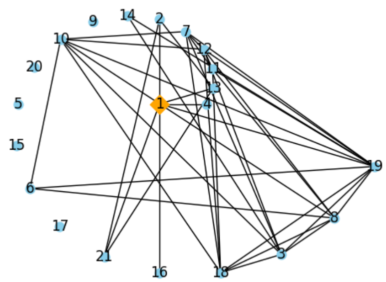
Table 1: Dictionary of the nodes selected by four sourcing methods.

Extensions of this simple approach may include (1) a much wider range of online Wikipedia pages around the seed page (2) a combination of more advanced algorithmic settings without too much time cost.

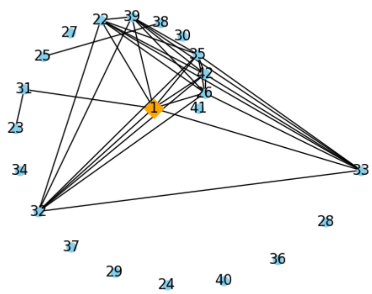
Therefore, an upgraded version of Wikipedia navigation might be an online application deployed on a powerful server to execute the data pipeline fast. As a starting point, the proof-of-concept satisfies the requirements proposed in Section 3, and several packages could be reused in the advanced version, too.

5 CASE STUDY

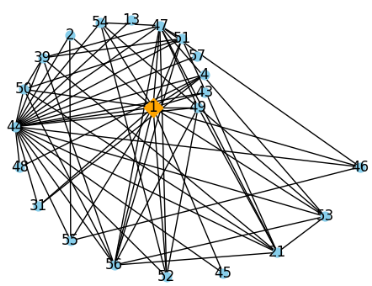
To illustrate the diverse kinds of 2D maps for Wikipedia navigation, we take the Wikipedia page “Film" as the seed, and crawl all its direct neighbors at depth one, all of which are cited on the “Film” page. Following the steps in Section 3, we set a threshold of 100 on the node degree (i.e., the number of links it has) to get a denser network



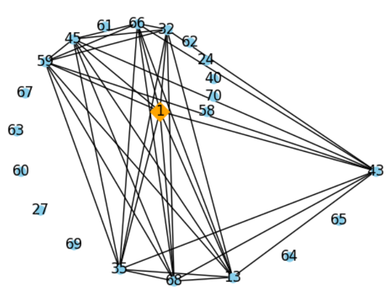
(a) Semantic content.



(b) Network structure.

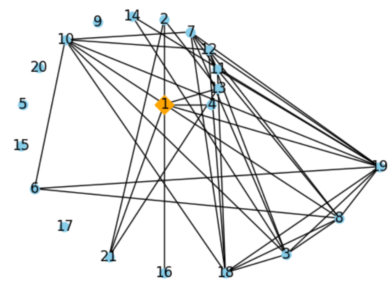


(c) Collaborative filtering.

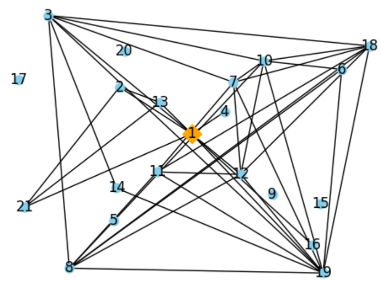


(d) A mixed PageDist [21] metric.

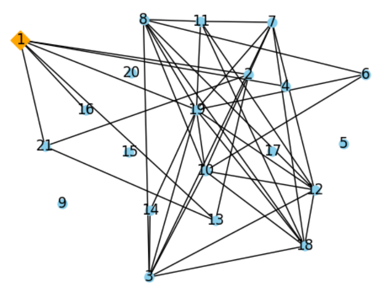
Figure 4: 2D map visualization of different sourcing methods. The distance to Node 1 is derived from feature similarity. The common layout is spiral. An orange diamond represents the root page.



(a) Spiral.



(b) Spectral.



(c) MDS.

Figure 5: 2D map visualization under different layouts. The distance to Node 1 is derived from feature similarity. The common sourcing method is semantic content. An orange diamond represents the root page.

with 8,083 directed edges and 151 nodes (pages). For a clear network visualization we only select top 20 neighboring nodes according to their feature similarity (depends on the choice of sourcing rule) to the seed node ("Film"), and display the top one-third of edges among those selected edges based on the edge weight defined by the times of concurrences on a third page.

Since we apply four kinds of sourcing and ranking methods introduced as the first group of algorithmic parameters in Section 3, in total there are 80 nodes selected for all the maps, but some nodes might be selected by multiple sourcing rules. Table 1 shows a dictionary of them. Since an accurate user preference prediction requires real user behavior data, we choose the second algorithmic parameter introduced in Section 3 as “the distance based on feature vector similarity” instead of a user preference prediction.

As for layout options, we take the coordinates directly generated by MDS and compute spiral and spectral coordinates in a polar system, respectively. MDS exploits a pairwise distance matrix to present a sense of how near or far points are from each other in a low dimensional space (e.g., 2D plane) to users. The spiral and spectral layouts tend to prove that users may choose their personalized layouts beyond the traditional MDS visualization method, and the navigation system is flexible enough to support the function. In total, we exploit the navigation system to generate the enumerations of available sourcing-ranking methods and layouts, some of which are displayed in Figure 4 and Figure 5.

Figure 4 illustrates the spiral 2D maps of Wikipedia nodes according to four different sourcing methods. Their common algorithmic parameters suggest that the distance to the center node (“Film”) corresponds to the ranking of their feature similarity to that of the Node “Film”. For example, in Figure 4a, the semantic content method treats Node 4 (“Television”) as the most similar neighbor to “Film”, and the second one is Node 13 (“Soundtrack”). The farthest neighbor is Node 19 “Animation”. In Figure 4b, according to some network science feature (e.g., degree of a node within the sandbox), the most significant two nodes are “Spectacle_(critical_theory)” and “Shooting_schedule” (a daily plan of film production). For the collaborative filtering map in Figure 4c, “Visual_effects” and “Videography” occupy the nearest two locations to the center. In the PageDist map (Figure 4d), “American_Dream" and “Bollywood" become the nearest neighbors. Users would recognize the obvious differences among the maps and choose their desired method for the following browsing.

With a limit of 20 or so nodes to a canvas in Figure 4, the four derived node sets have almost no intersection. That is, the different metrics produce very different neighborhoods in terms of their underlying node sets. If we use a larger bound of 50 nodes on a given canvas, the semantic-content set and network structure set have 10 nodes in common, the semantic-content set and collaborative filtering set share 16 nodes, while the intersection of collaborative filtering and PageDist contains 13 nodes. Going further, the first three sets (semantic, network and collaborative), have five nodes in common “Film_budgeting", “Cinematography", “Roadshow_theatrical_release", “Film_industry", “Principal_photography". The diverse navigation maps will have varying levels of utility to different user groups.

Figure 5 displays three layouts of the same subset of nodes according to the semantic content sourcing method, with a condition that the similar nodes of “Film" would be placed close to the center. For the MDS one (Figure 5c), the coordinates of all nodes are derived from a similarity matrix so that the node “Film" may not be at the center of the canvas. More importantly, MDS considers mutual similarity between any pair of nodes on the canvas, while for the other layouts, the comparative distance is only meaningful between the root node “Film” and another node.

Since only the edges with a large enough weight could be added to the map, the dense edges suggest several local clusters, such as Nodes (4, 13, 11, 12, 7), or another group (18, 3, 8, 19) in Figure 5a. Besides, the spiral layout clearly shows the similarity-based distance to the center node in an anti-clockwise order. For the second spectral layout in Figure 5b, we allocate the nodes mainly in four directions (upper right, upper left, down right, down left). It might be more difficult to compare the distance to the center for two nodes (e.g., Nodes 9 and 14), but the spectral layout makes it possible to cluster the neighboring nodes into several groups and deploy each group along a “beam”. In Figure 5c, the MDS layout considers the distance matrix of all nodes in terms of the semantic vector of the corresponding Wikipedia page’s title and computes their coordinates with a standard dimensionality reduction algorithm, so the root page “Film” is automatically placed in the upper left corner.

In this way, without the special color/shape, it might not attract users attention at the first glance. MDS is a popular standard visualization method, but when users choose the second algorithmic parameter about closeness in the navigator as “a probability from a predictive model”, it is more difficult to define a complete distance matrix, especially between pairs of surrounding nodes.

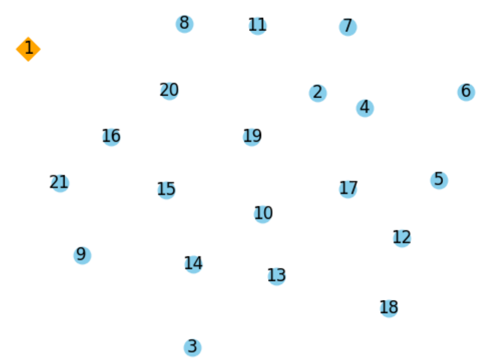


Figure 6: A non-edge version of Figure 5c with MDS.

Beyond the above algorithmic parameters and layout options, other visualization factors may be critical. Figure 6 displays the non-link version of MDS layout, in which the neighborhood is determined by a textual distance instead of link-based distance on the subnetwork. We would anticipate associating such a non-edge map with some kinds of “sliders" that would allow the picture to vary according to user feedback.

6 FUTURE WORK: DESIGN OF USER STUDY

This paper is a proof-of-concept. Before realizing a full-blown application, user study would be necessary. This is among the proposed directions of future work. Herein we outline the plan of user study to evaluate the Wikipedia navigator from two aspects: subjective feeling of transparency and general user experience; the effects of algorithmic parameters in terms of Wikipedia pages recommendation.

Questions about user experience are often an important part of a user-oriented study [6, 14, 25, 31]. We would like to design several questions to determine the degree that users believe the navigation tool improves transparency via customizability and network visualization, as well as their preferred parameter settings. Possible data processing methods/metrics for user feedback include a correlation matrix of multiple numerical ratings [25] and statistical tests [31] to validate some significant difference.

7 CONCLUSIONS

In this paper, we have presented a proof of concept for an open navigation tool of Wikipedia pages to broaden the understanding of the information context of Wikipedia pages to a user, along with a form of algorithmic transparency for the users with the goal of enabling them to better understand why they get the current map of a vast Wikipedia network. We find that the sourcing and ranking method can greatly affect the set of finally selected nodes on the canvas, and different layouts highlight (according to the different underlying metrics) different significant neighboring nodes in the corresponding local cluster on the map.

As per the above, future work may include a more complete user study to quantify the effects of different algorithmic parameters on their preference of Wikipedia pages. In this way, we may be able to explore how the structure of Wikipedia articles influences user navigation, and improve the design of the user interface. User data could also contribute to other related research projects, such as a Wikipedia based recommender. There is also the possibility of an upgraded version of navigation tool merged into a browser (e.g., a Chrome extension) or a back-end deployment on a web server to speed up the Wikipedia page visualization.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Neukom Institute for Computational Science.

REFERENCES

[1] 2016. Research:Wikipedia Navigation Vectors. Retrieved Feb 24, 2019 from https://meta.wikimedia.org/wiki/Research:Wikipedia_Navigation_Vectors
[2] 2018. Beautiful Soup Documentation. Retrieved Jan 23, 2019 from https://www.crummy.com/software/BeautifulSoup/bs4/doc/
[3] 2018. Overview of NetworkX. Retrieved Jan 23, 2019 from https://networkx.github.io/documentation/stable/
[4] 2019. Tkinter, Python interface to Tcl/Tk. Retrieved Jan 23, 2019 from https://docs.python.org/2/library/tkinter.html
[5] 2019. Urllib, Open arbitrary resources by URL. Retrieved Jan 23, 2019 from https://docs.python.org/2/library/urllib.html
[6] Paritosh Bahirat, Yangyang He, Abhilash Menon, and Bart Knijnenburg. 2018. A data-driven approach to developing iot privacy-setting interfaces. In *23rd International Conference on Intelligent User Interfaces*. ACM, 165–176.
[7] Vevake Balaraman, Simon Razniewski, and Werner Nutt. 2018. Recoin: Relative Completeness in Wikidata. In *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 1787–1792.
[8] Patti Bao, Brent Hecht, Samuel Carton, Mahmood Quaderi, Michael Horn, and Darren Gergle. 2012. Omnipedia: bridging the wikipedia language gap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1075–1084.
[9] S Brin and L Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Computer Networks And ISDN Systems*, Melissa Crouch and Tim Lindsey (Eds.). Elsevier, 107–117.
[10] Michel Buffa and Fabien Gandon. 2006. SweetWiki: semantic web enabled technologies in Wiki. In *Proceedings of the 2006 international symposium on Wikis*. ACM, 69–78.
[11] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. 2007. SuggestBot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*. ACM, 32–41.
[12] Trevor F Cox and Michael AA Cox. 2000. *Multidimensional scaling*. Chapman and hall/CRC.
[13] Stefano De Sabbata, Arzu Çöltekin, Kathryn Eccles, Scott Hale, and Ralph Straumann. 2015. Collaborative Visualizations for Wikipedia Critique and Activism. In *Ninth International AAAI Conference on Web and Social Media*.
[14] Ivania Donoso-Guzmán and Denis Parra. 2018. An Interactive Relevance Feedback Interface for Evidence-Based Health Care. In *23rd International Conference on Intelligent User Interfaces*. ACM, 103–114.
[15] Fabian Flöck, David Laniado, Felix Stadthaus, and Maribel Acosta. 2015. Towards Better Visual Tools for Exploring Wikipedia Article Development–The Use Case of âĂIJGamergate ControversyâĂİ. In *Ninth International AAAI Conference on Web and Social Media*. p48–55.
[16] Laxmi Amulya Gundala and Francesca Spezzano. 2018. ReadersâĂŹ Demanded Hyperlink Prediction in Wikipedia. In *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 1805–1807.
[17] Reed H Harder, Alfredo Velasco, Michael Evans, Chuankai An, and Daniel Rockmore. 2017. Wikipedia Verification Check: A Chrome Browser Extension. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 1619–1625.
[18] Christopher D. Manning Jeffrey Pennington, Richard Socher. 2014. GloVe: Global Vectors for Word Representation. Retrieved Jan 25, 2019 from https://nlp.stanford.edu/projects/glove/
[19] Daniel Lamprecht, Denis Helic, and Markus Strohmaier. 2015. Quo vadis? On the effects of WikipediaâĂŹs policies on navigation. *Links* 80 (2015), 100.
[20] Daniel Lamprecht, Kristina Lerman, Denis Helic, and Markus Strohmaier. 2017. How the structure of wikipedia articles influences user navigation. *New Review of Hypermedia and Multimedia* 23, 1 (2017), 29–50.
[21] Greg Leibon, Michael Livermore, Reed Harder, Allen Riddell, and Dan Rockmore. 2018. Bending the law: geometric tools for quantifying influence in the multinetwork of legal opinions. *Artificial Intelligence and Law* 26, 2 (2018), 145–167.
[22] G Leibon and D N Rockmore. 2013. Orienteering in Knowledge Spaces: The Hyperbolic Geometry of Wikipedia Mathematics. *PLoS ONE* (2013). https://doi.org/10.13f71/journal.pone.0067508
[23] Baiju Muthukadan. 2018. Selenium with Python. Retrieved Jan 23, 2019 from https://selenium-python.readthedocs.io/
[24] Gergely Odor, Emanuele Bugliarello, and Robert West. 2018. Poster: How Did Wikipedia Become Navigable?. In *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee.
[25] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 157–164.
[26] Silvan Reinhold. 2006. WikiTrails: Augmenting wiki structure for collaborative, interdisciplinary learning. In *Proceedings of the 2006 international symposium on Wikis*. ACM, 47–58.
[27] Tomás Sáez and Aidan Hogan. 2018. Automatically Generating Wikipedia Infoboxes from Wikidata. In *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 1823–1830.
[28] Sebastian Schaffert. 2006. IkeWiki: A semantic wiki for collaborative knowledge management. In *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE'06. 15th IEEE International Workshops on*. IEEE, 388–396.
[29] Shilad Sen, Anja Beth Swoap, Qisheng Li, Brooke Boatman, Ilse Dippenaar, Rebecca Gold, Monica Ngo, Sarah Pujol, Bret Jackson, and Brent Hecht. 2017. Cartograph: Unlocking Thematic Cartography Through Semantic Enhancement. In *22nd International Conference on Intelligent User Interfaces, IUI 2017*. Association for Computing Machinery, 179–190.
[30] Neil Thompson and Douglas Hanley. 2018. Science Is Shaped by Wikipedia: Evidence From a Randomized Control Trial. Retrieved Jan 23, 2019 from https://ssrn.com/abstract=3039505
[31] Yuan Yao and F Maxwell Harper. 2018. Judging similarity: a user-centric study of related item recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems*. ACM, 288–296.