Recommendations of Research Articles by Experts: Visualizing Relationships and Expertise

Peiling Wang[1][a], Scott Shumate[2][b], Pinghao Ye[3] and Chad Mitchell[4]
[1]*School of Information Sciences, University of Tennessee-Knoxville, Knoxville, TN 37996, U.S.A.*
[2]*Felix G. Woodward Library, Austin Peay State University, Clarksville, TN 37044, U.S.A.*
[3]*Wuhan Business School, Wuhan, P.R.C.*
[4]*Polinode.com*

Keywords: Recommender System, Biomedical Publications, Expert Evaluations, Visualizing Experts, Visualizing Research Frontiers.

Abstract: The paper applied data analytics and network visualization to show the potentials of employing Faculty Opinions beyond literature recommendations by domain experts. Based on a set of highly recommended articles by at least four experts with a sum of 10 or more stars (A recommended article is assigned a score between one to three stars by the recommender.), this study tests the new ideas and methods of identifying and visualizing relationships between scientific papers, experts, and categories. Despite of the available dataset in the study is small, the findings show that a platform designed for recommending and retrieving publications has the potential as a knowledge base for seeking experts. The results are indicative rather than conclusive; further study should apply AI methodology to include multiple data sources to corroborate findings and to enhance the applicability of data visualization towards knowledge graphs.

1 INTRODUCTION

1.1 Literature in Big Science Era

Scientists and researchers today face the sheer volumes of available publications (Björk et al., 2009; Bornmann & Mutz, 2015; White, 2021). Additional challenges include the rising threats of predatory journals and retractions (Wang, 2023). Senior researchers do not have enough time to read what have been published; junior researchers lack the necessary expertise or knowledge to select important publications on their own (Pontis et al., 2017). After finding one or two relevant publications, a user typically asks "*what should I read next*?" (Bruns et al., 2015). Recommendations of important publications by domain experts can help budding researchers as well as seasoned researchers who develop new areas. Researchers distinct between expertise retrieval and expertise seeking. Expertise retrieval is about the relevant documents authored by experts while expertise seeking is about seeking experts or experts' knowledge.

To identify publication venues, Forrester, Björk, and Tenopir (2017) advises how to choose of appropriate journals for submission and the useful tools (but mostly not free). Integrating social network analysis and contextual similarity, a personalized recommender of papers, named DISCOVER (Diversified Integrated Social network analysis and COntextural similarity-based scholarly VEnue Recommender) based the citation relationships (e.g., reference, bibliographic coupling and co-citation) and content similarity to model, which has limitation in scalability.

There are various approaches to scientific paper recommendation systems (e.g., user profile, popularity measures (author, journal, citation), key terms, topic model, meta-path etc.). The critiques of the current body of work on publication recommendations point to the lack of incorporating users' level (i.e., junior or senior researchers) and use scenarios. Alinani et al. (2018) also propose a

[a] http://orcid.org/0000-0003-4202-7570
[b] http://orcid.org/0009-0007-7893-6692

Wang, P., Shumate, S., Ye, P. and Mitchell, C.
Recommendations of Research Articles by Experts: Visualizing Relationships and Expertise.
DOI: 10.5220/0012720700003756

In *Proceedings of the 13th International Conference on Data Science, Technology and Applications (DATA 2024)*, pages 269-276
ISBN: 978-989-758-707-8; ISSN: 2184-285X

recommender system for suggesting research collaborators for a specific topic area, based on several factors including published articles, the journals (impact), author's keywords, and co-authorship. Transparent Interactive Graph Recommender System (TIGRS) (Bruns et al., 2015) was a personalized visual publication recommender system proposed for an interactive graph of publication networks.

A better understanding of usage context, users, and recommendation scenarios has focused on a data-driven approach (Tran et al., 2021). User interface designed for scholarly paper recommender systems is often a part of bibliographic databases with list-like output of items. What and how to present recommendations, improve serendipity and exploration, are important for user experience. As an interactive recommender system to visualize items for exploration and serendipity, (Calero Valdez et al., 2016) applied the algorithms relying on the individual ratings by all (user-based); and content-based filtering that uses meta-data or features from individual items (IR and text-mining tools). The techniques to find experts by Al-Taie, Seifedine, and Obasa (2018) also use social network-based graphs (nodes and edges) and ranking of retrieved documents. The work by Everton et al. (2022) demonstrates that tie strengths between nodes (e.g., 1, 2, …, or 5) affected a network's structure. The underlying assumptions are important for choosing centrality measures (degree, closeness, betweenness or Eigenvector).

1.2 Faculty Opinions: Recommending Publications by Domain Experts

A well-developed article recommendation system in biology and medicine is H1Connect, previously the Faculty Opinions until July 11, 2023. Faculty Opinions was a rebrand of F1000Prime on April 12, 2020 while F1000Prime was incepted in 2009. Despite of the changes in names, URL domains, and record displays over the three milestones (Figures 1-3), the database provides a platform for experts to recommend the articles and for users to browse and search these recommended articles. Established and recognized domain experts (e.g., Nobel Laureates) are selected as Experts who can recommend alone or co-recommend with their associate experts (Figures 2 and 3). Figure 3 shows a recommendation in F1000Prime and H1Connect. Faculty Members (FMs) and Associate Faculty Members (AFMs) in F1000Prime are Experts and Associate Experts in H1Connect. The information about the recommenders: in F1000Prime, (A) shows pictures of the faculty members, their domains, and affiliation; in H1Connect, (B) shows smaller pictures with only names with a hyperlink to each name (C) and (D) to the faculty members' webpages with profiles. From each expert's profile, all the recommended articles by this expert can be accessed. Researchers have studied the potential of F1000Prime as a new data source for research evaluation (Waltman & Costas, 2014).

1.3 Hidden Knowledge

As the first well-developed scientific article recommender system the F1000Prime, then Faculty Opinions and now H1Connect, is unique because the rating, classification, and evaluative comment are by domain experts rather than average readers from the community. However, the system interactions with the users are similar to a typical information retrieval (IR) system. For example, the results of searching or browsing are a list of bibliographic records, each of which can be further viewed. However, the system has hidden knowledge and relationships that IR systems do not have.

This study aims to explore the potential to leverage the power of the article recommendation system beyond publications. We applied several strategies and used network analysis to visualize:

a. potential reading and publishing venues (journals)
b. new development or research frontiers in the field
c. hidden intellectual relationships among domain experts (to find collaborators, mentors, peer reviewers, etc.)

2 MATERIALS AND METHODS

2.1 Dataset

The dataset includes 1,452 articles recommended by 3,660 Experts. Each article has a sum of stars equal to or greater than 10 stars (we do not use the weighted sum of stars in Figure 2 because it is updated periodically). Each article has been recommended by at least 4 Experts (not including co-recommenders). Experts who dissented the recommendation(s) by other Experts are included as Experts (Figure 2) because technically, the system handles dissents as recommendations without a star rating. The total of 8,543 recommendations includes 33 dissents.

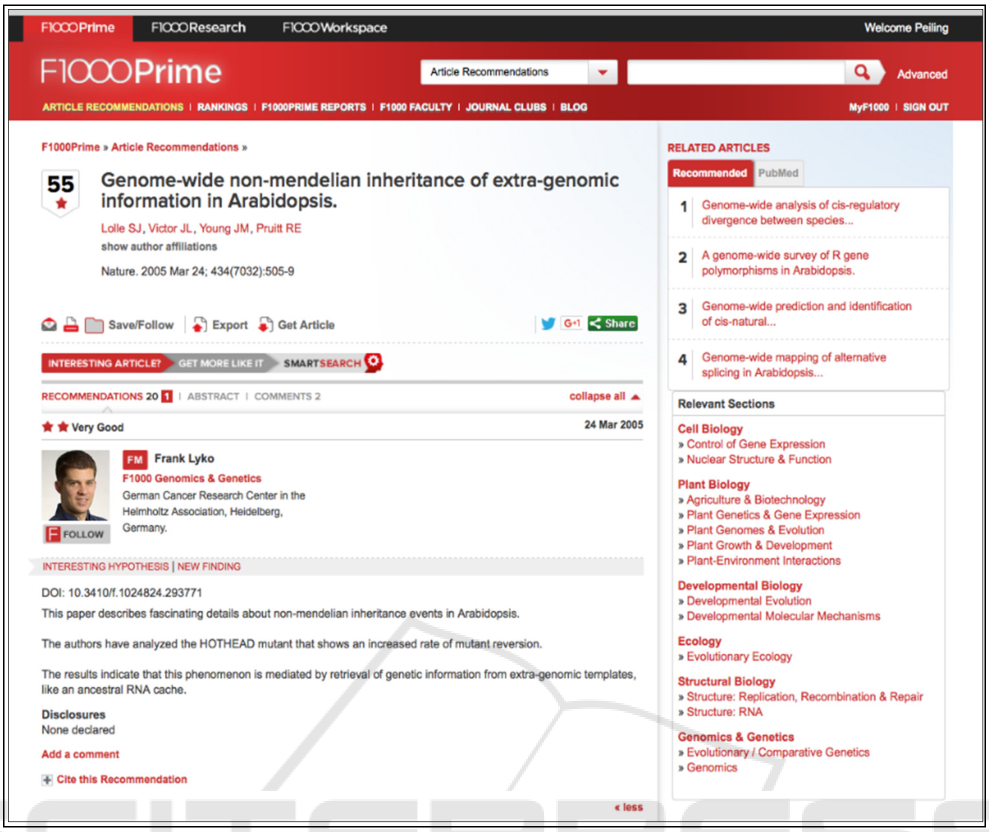

Figure 1: An example of F1000Prime article recommendation.

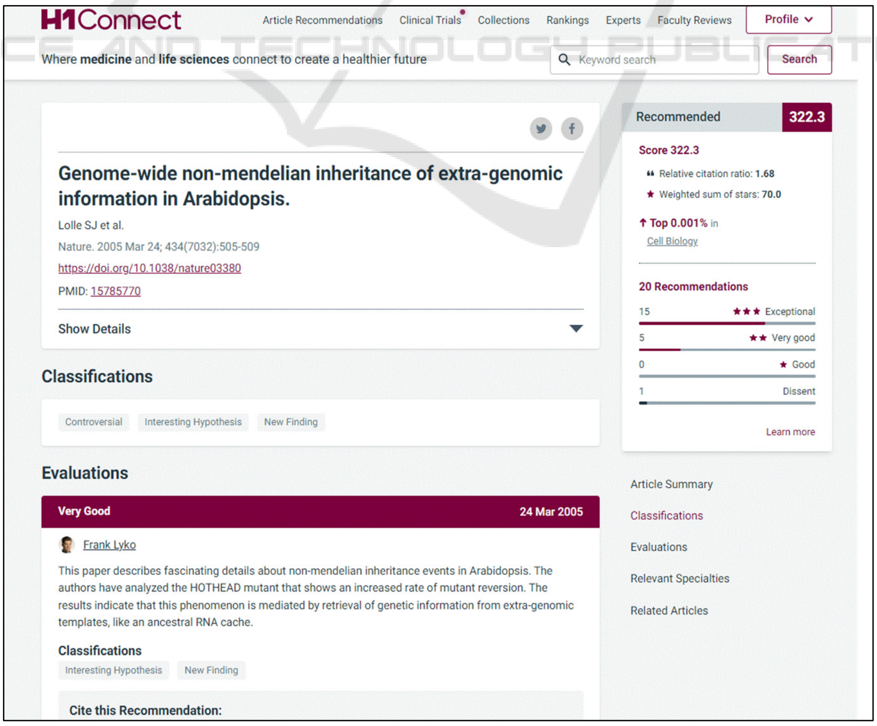

Figure 2: The same recommended article in H1Connect (as Figure 1).

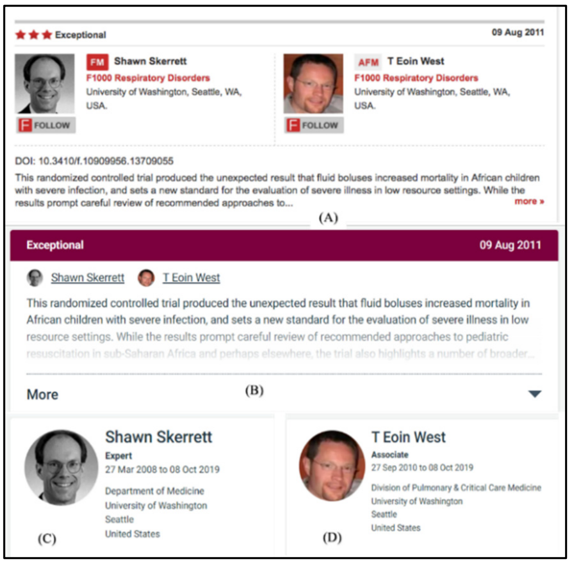

Figure 3: Experts in F1000Prime vs. H1Connect.

The collected data were used only for research in accordance with fair use principle. The database for data analysis was password protected and was never shared with others or used for other purposes. The data will not be uploaded to repository despite of our strong belief of open science. Researchers interested in this line of research need to collect data and be aware of changes in the system.

2.2 Data Collections

For the purpose of this study, the focus on the articles received at least four commendations with a total of 10 or more stars was based on the long-tail distributions of recommended articles (Bornmann, 2015; Wang & Su, 2021). The majority of the articles (81%) were recommended by only one expert; only about 5% of the articles scored 7 and more total stars. These studies also reported the low level of agreement in ratings among experts. We collected data at two times: 1) 1178 from about 134,333 articles (0.88%) from F1000Prime on April 25, 2016; and 2) updated the dataset from Faculty Opinions on July 15 2022 adding 274 articles (i.e., 1452 out of approximate 192,826 articles (0.75%). As an observational research project, the steps to collect data include:

Step 1. Using browsing mode and sort by Score to include all the articles with recommendation total Score equal or great than 9 from F1000Prime, which resulted in 1669 hits as of June 2016.

Step 2. Using browsing mode and filter publication year 2016 to 2022 and sort by Score as in *Step 1* from H1Connect (Figure 2).

Step 3. Develop a Python program to download the articles' bibliographic data and the link to the Experts' recommendations.

Step 4. The two datasets were merged after all the 2016 collected 1669 articles were updated by searching the H1Connect. The articles less than less than 10 total stars were removed resulted in 1452 articles.

Step 5. The records were imported to a database for analysis after data wrangling.

2.3 Data Structure

Figure 4 is the basic data structure to show that each article has one to many recommendations; each recommendation is by one Expert (FM) who assigns one to many categories. The entities' identifiers from the system were used, which made direct access to the records efficient and responsive to domain changes. For example, each article has a unique identifier taking from the last part of the URL as PK of the article entity. Article_ID = 1024824 which is from https://connect.h1.co/article/1024824 and previously the URL https://www.f1000.com/prime/1024824. Similarly, from the FM's profile webpage's URL https://connect.h1.co/member/1988460240298681 the PK for the Faculty entity is as FM_ID = 1988460240298681 (the previous URL was https://facultyopinions.com/member/1988460240298681) Both old and new URLs have the same digits.

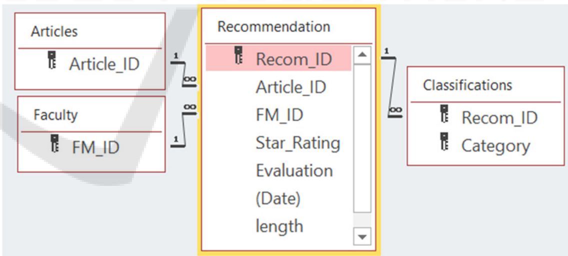

Figure 4: Data structure for analysis.

2.4 Data Analysis

1. Visualizing experts' recommendations:
   - Which journals published the most recommended articles?
   - Which experts recommended these journals?
2. Visualizing experts' recommendations and their assigned categories:
   - Which categories were recommended by experts?
   - How are experts assigned categories linked?

3. Visualizing associations between journals and categories:
   - o Which types of articles in journals were recommended?
   - o How are categories of the recommended articles linking the journals?
4. Visualizing experts and their networks:
   - o Which experts recommended the same articles and how many?
   - o Who are the other experts recommended articles with top recommenders?

For data analysis, queries developed in the database (Figure 4) output data to statistical tools (SPSS) to generate measurements of nodes and edges. The tools for social network analytics and visualization include NetDraw (Borgatti, 2005) and NodeXL Pro (https://www.smrfoundation.org/nodexl/).

3 RESULTS

The highly recommended articles scored between 10 and 55 total stars and their recommenders (including dissenters) ranged between 4 and 22 Experts. Of the 3,660 Experts, the majority recommended one or two articles showing a long-tail trend (Figure 5).

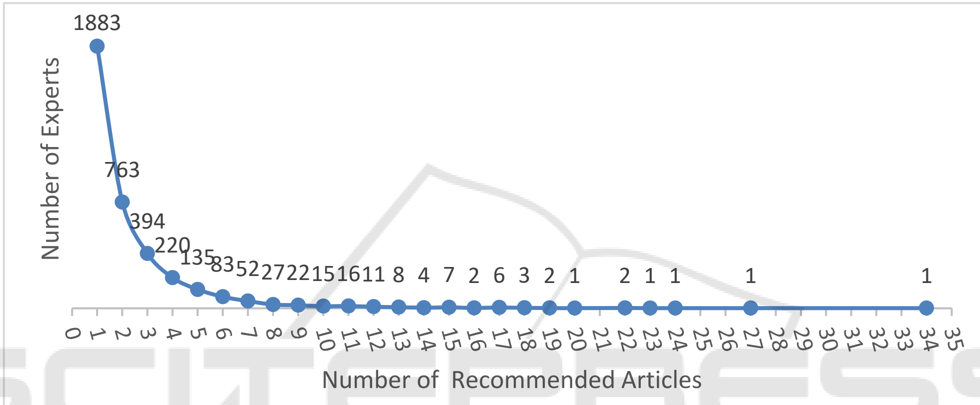

Figure 5: Number of Experts recommending the same articles.

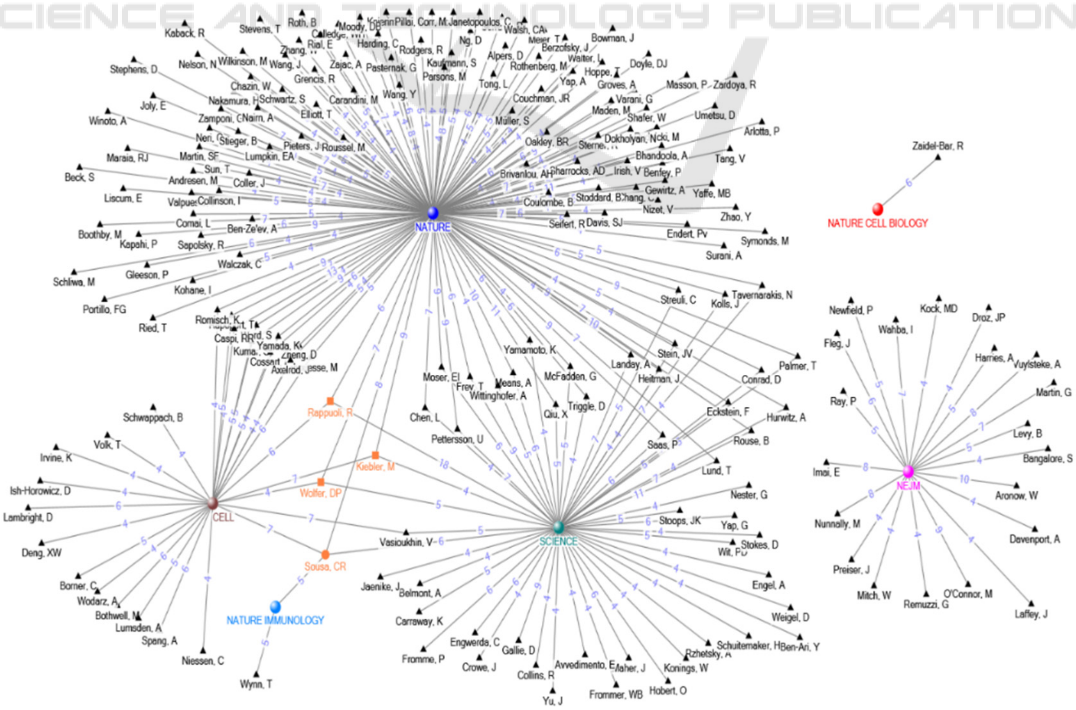

Figure 6: Journals published the most recommended articles.

3.1 Experts' Journals

Setting the threshold of minimum four articles published in the journal, a total of 199 Experts were associated with six journals and three journals were highly connected by the experts (Figure 6): *Nature* (140), *Science* (52), and *NEJM* (*20*). Figure 6 shows the clusters of six journals.

For each of the clusters, a second threshold included the experts whose two-third of the recommended articles were published in the journal. Experts recommended articles from one journal more than other journals (Figure 7). Of the 140 *Nature* recommenders*,* 55 Experts recommended articles mostly in *Nature* (2/3 or more); 12 of the 55 Experts only recommended articles in *Nature.* For the 20 Experts recommended articles in *NEJM*, 9 recommended only *NEJM* articles and 5 recommended mostly *NEJM* articles (Figures 6 and 7).

3.2 Experts' Categories

Each recommendation (or dissent) was assigned between 1 and 7 categories from the predefined scheme (10 categories). Figure 8 depicted the categories recommended by the experts and the number of articles for the category vs. total recommended articles. The two most used categories are New Finding and Interesting Hypothesis. The left-lower part of Figure 8 shows the Experts that bridged different categories. For example, Saas classified all 17 recommended articles as New Finding, 14 of which also classified as Interesting Hypothesis, and 9 as Novel Drug Target. Similarly, Kiebler applied three categories: New Finding, Interesting Hypothesis, and Technology Advance. Controversial was classified by three Experts Boero, Nunnally, and O'Connor in most of their recommendations.

3.3 Journals and Categories

Figure 9 depicts an interactive graph for the associations between the recommended articles' venues (journals) and the categories. For example, the user can zoom in the journal *Nature Medicine* (see the red arrow)*,* the eight categories assigned to this journal's recommended articles were highlighted to red color too. From any of these categories, the journals are linked to provide access to the recommended articles. The most recommended articles in *Nature Medicine* were classified as New Finding (181 out of 351 articles or 52%).

To zoom in New Finding, the linked journals will be highlighted. The two highly associated journals are *Cell* (1,114 articles) and *Nature* (2,660 articles). The zoom-in feature can hide the other categories to simplify the graph; or be displayed as a popup table with a preferred order. From the category Controversial, *Nature*, *New England Journal of Medicine* and *Science* contributed the most recommended articles. The dense Figure 9 as the first display provides a big picture for further interactions.

3.4 Visualizing Experts

Out of the 3,660 experts, 162 experts recommended at least three *same* articles (Figure 10). This network has several subnetworks or ego networks that can be identified by the most connected Experts. Figure 11 illustrates three experts' ego networks. The two by Lund and Rappuoli are connected by two paths: 1) they recommended the same two articles; and 2) Caspi recommended with each of the two as a bridge

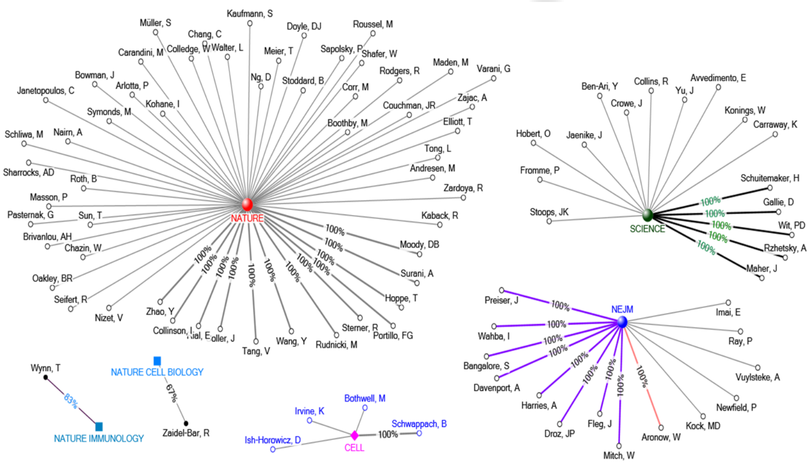

Figure 7: Journals have the most recommended articles by Experts.

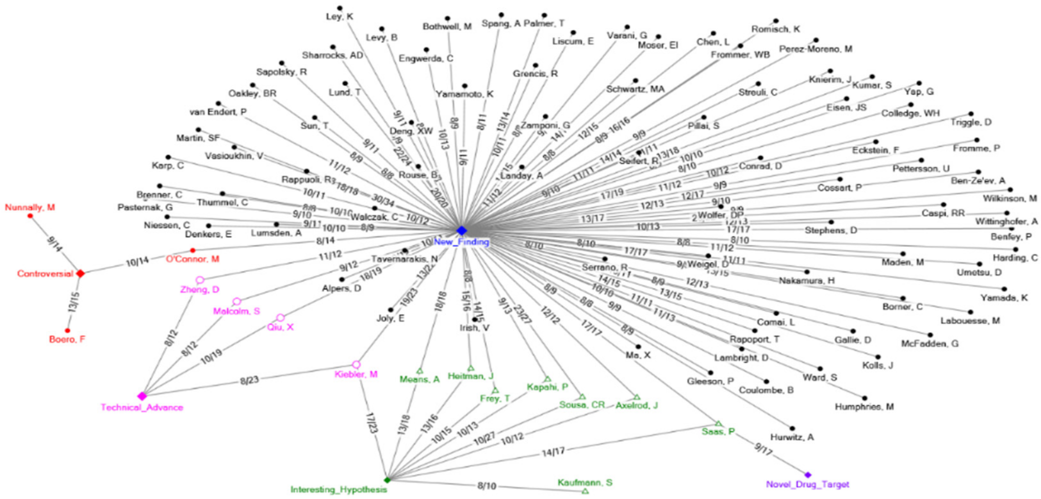

Figure 8: Recommended articles classified by Experts.

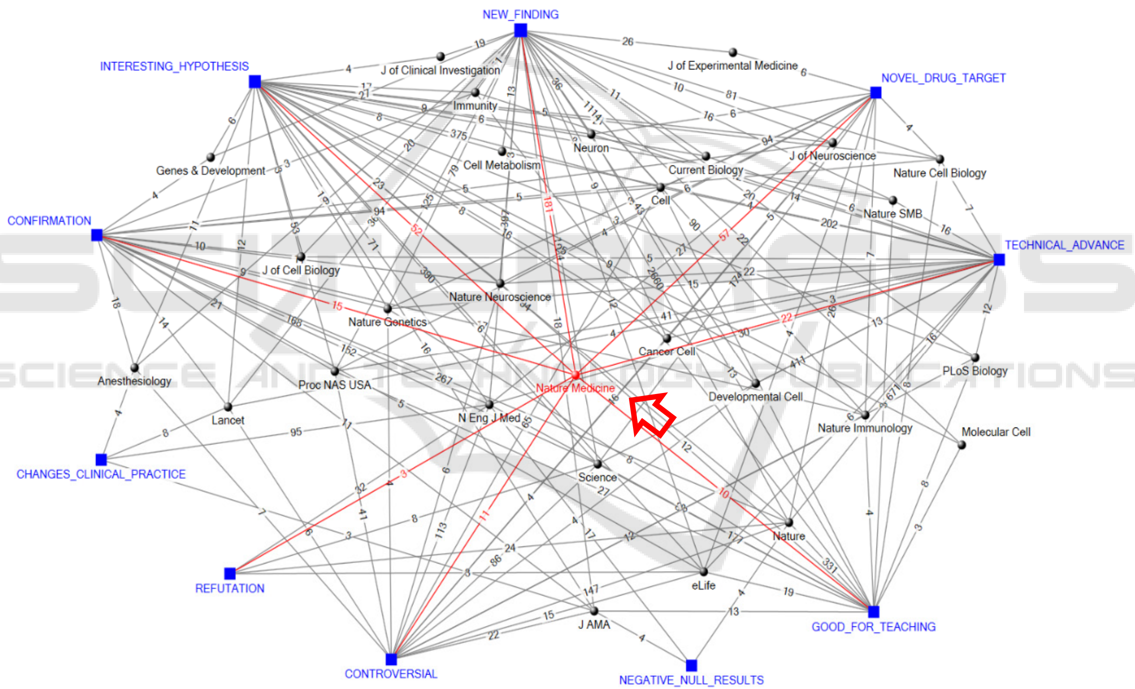

Figure 9: Journals and categories of recommended articles.

Expert. Benfey's ego network does not connect with either Lund or Rapouli. From the links to these Experts' profiles, their respective specialties can explain the results: Lund (immunology) and Rapouli (medical microbiology) are more likely to share research interests than Benfey (plant biology).

4 CONCLUSIONS

Data analytics and network visualization of Experts' recommendations of articles in H1Connect show that the value and potentials to discover hidden knowledge go beyond finding important articles (the system was design for). The application should complement the bibliometric-based approach to find successful publication venues (Klemiński, et al., 2021). Further, the interactive design can help users quickly zoom in research hot topics and identify experts for peer reviews or mentors they do not already know.

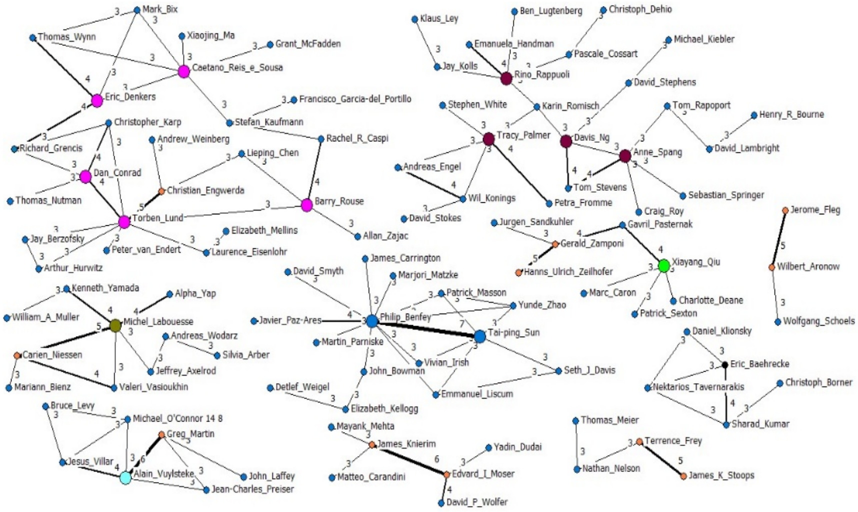

Figure 10: Network of Experts based on recommending the *same* articles (>=3).

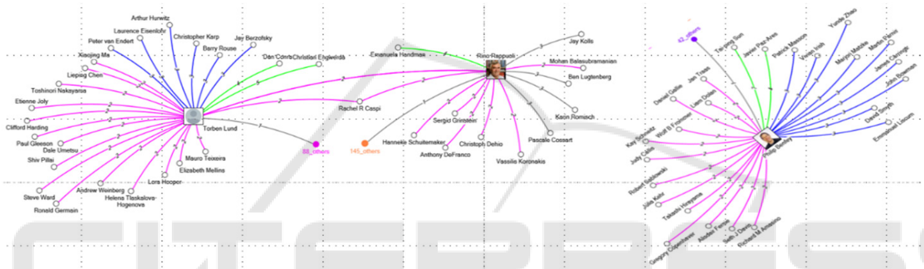

Figure 11: Ego-networks of Experts.

However, as the first step to explore the potentials, this study has some limitations. The dataset is relatively a small percentage of the collection in H1 Connect (a.k.a., Faculty Opinions or F1000Prime). As the system grows, more data can be curated to generate broader results. The analysis and visualization are based on one source and the recommendations are by relatively a small percentage of experts in the biomedical fields whose reading may be limited to their research in scope and coverage.

Further research will need to curate data from multiple sources such as citations in context using AI approach to broaden coverage of publications.

REFERENCES

Alinani, K., et al. (2018). Aggregating author profiles from multiple publisher networks to build a list of potential collaborators. IEEE Access, vol. 6, pp. 20298-20308

Al-Taie, MZ., Seifedine, K., Obasa, AI. (2018). Understanding expert finding systems. Social Network Analysis & Mining 8, Article 57

Björk, B-C., Roos, A., Lauri, M. (2009). Scientific journal publishing: Information Research, 14(1) paper 391.

Borgatti, SP. (2005). Centrality and network flow. Social Networks. 27(1): 55-71

Bornmann, L. (2015). Interrater reliability and convergent validity of F1000Prime peer review. JASIST. 66 (12).

Bornmann, L., Mutz, R. (2015). Growth rates of modern science. JASIST 66(11)

Bruns, S., et al. (2015). What should I read next? Lecture Notes in Computer Science, vol 9173

Calero Valdez, A. et al. (2016). Recommender Systems for Health Informatics. LNAI 9605, pp. 391–414

Everton, S., et al. (2022). Strong ties and where to find them … Social Network Analysis & Mining, 12 Article 112

Klemiński, R., et al. (2021). Where should I publish? Journal of Informetrics, 15 (3), pp 1-20

Pontis, S., et al. (2017). Keeping up to date. JASIST, 68

Tran, TNT., et al. (2021). Recommender systems in the healthcare domain, J Intell Inf Sys 57: 171–201

Waltman, L., Costas, R. (2014). F1000 recommendations as a potential new data, JASIST 65(3), pp 433–445

Wang, P. (2023). Rising of retracted research works and challenges in information systems, *CHIIR23*

Wang, P., et al. (2020). F1000Prime recommended articles and their citations, Scientometrics, 122: 933-955.

Wang, P., Su, J. (2021). Post-publication expert recommendations. doi: 10.1016/j.joi.2021.101174

Wang, P., Su, J. (2022) Expert recommended biomedical journal articles ... doi: 10.1177/01655515221074329