

Analyzing the Scientific Community Focused on Nursing Informatics Using Publication Data

Ulrich SCHRADER^{a,1}

^a*Frankfurt University of Applied Sciences, Frankfurt, Germany*

Abstract. The scientific community focused on nursing informatics can be described as a graph with the authors as vertices and the author-coauthor relationship as the connecting edges. Methods to describe and analyze networks like average path length, diameter, centrality measures, or partitioning into subcommunities are applied to the nursing informatics community. It is shown that the community consists of one large connected subnet with many small disjoint subnets, each representing one or several authors. The interconnectivity of the large subnet is quite high indicating an information flow along several different paths. Using different centrality measures important authors for e.g. the information flow can be identified. While each small disjoint subnet represents a small sub-community, the large central subnet can also be partitioned into subcommunities connected with each other. Some seem to be focused on specific aspects of nursing informatics.

Keywords. Data analytics, publications, Pubmed, scientific community

1. Introduction

A scientific community is traditionally described as a network of scientists working concurrently on a common topic or problem [1]. To gain insight into the structure and development of the network representing a scientific community the linkages between the scientists in the community must be known. These are usually not explicitly observable in its entirety or documented. They can include meetings at conferences, common projects, working at the same place, or to have common publications. Thus, the quality of a linkage differs a lot. Based on [2,3], the coauthor relationship is used to determine the linkages between scientists in the nursing informatics community. This is a rather strong relationship as writing a paper together entails a high level of trust, commitment, and usually a common understanding of the subject matter. The advantage of this approach is that the author-coauthor relationship is usually well documented in literature meta-databases, e.g. Pubmed of the US National Library of Medicine.

This paper assumes, that members of the scientific community focusing on nursing informatics have at least published one paper on the topic of nursing informatics.

¹ Corresponding Author, Ulrich Schrader, Frankfurt University of Applied Sciences, Nibelungen Platz1, 60318 Frankfurt, Germany; E-mail: mail@ulrich-schrader.de.

2. Methods

The literature meta-database Pubmed was used to select papers indexed with the MeSH term “nursing informatics” or that would contain the phrase “nursing informatics” in any text field. The authors of the publications found do form the scientific community as proposed in this paper. The data were downloaded in the pubmed format and then parsed into a database using Python. The Python module NetworkX was used to construct and analyse the co-author relationships as an undirected weighted networks. The graphs were constructed using the Python Matplotlib moduleragraph.

3. Results

A total of 2070 articles ranging from 1978 to November 2019 involving 1872 different authors were found that did satisfy the search criteria.

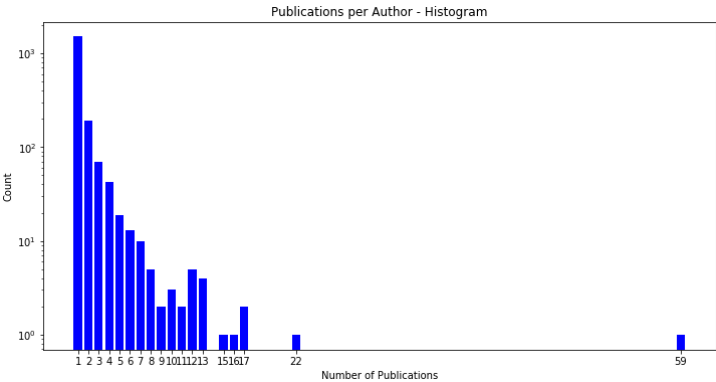


Figure 1. Histogram of the publications per author (logarithmic scale).

The majority of the authors were only involved in one paper (Figure 1) satisfying the search criteria, while one authored 59 such papers. On average each author was involved in 1.1 different papers satisfying the search criteria.

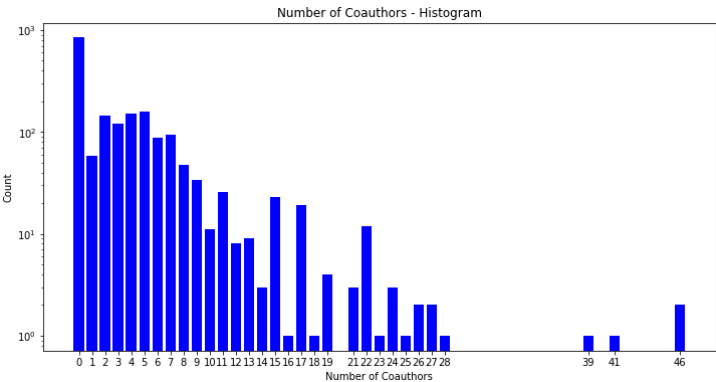


Figure 2. Histogram of number of coauthors for an author (logarithmic scale).

While a large number of papers have been written only by one author, it is more common that papers are written by a group of authors. Figure 2 shows the distribution of the number of different coauthors an author has in all her papers. On average an author has had 3.3 different coauthors in the timespan observed. There are even two authors that have been involved with a total of 46 different coauthors in their papers.

Using the author-coauthor relationship an undirected weighted graph (Figure 3) of the community was created using the number of common publications of an author-coauthor pair as weight. Each single symbol (node) represents an author.

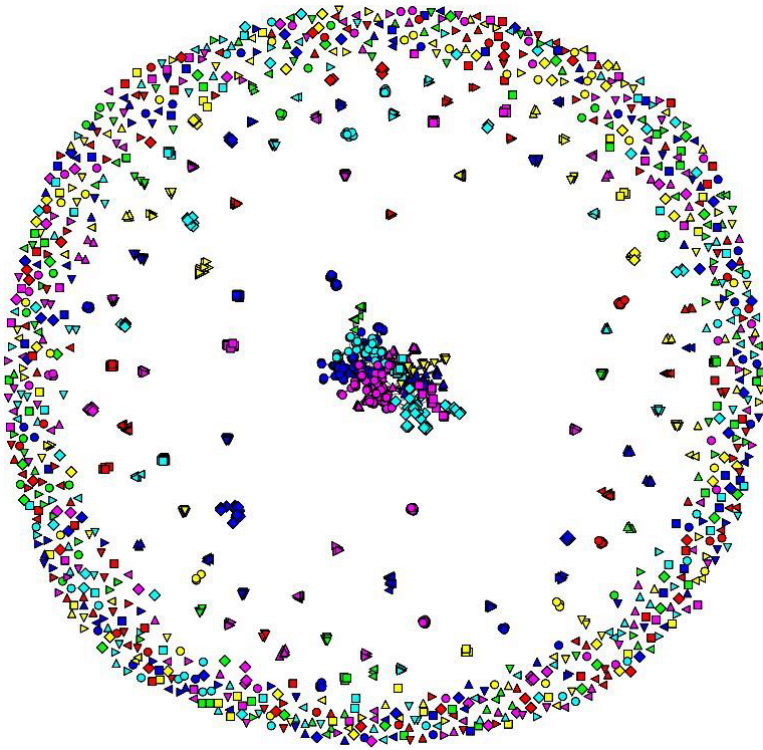


Figure 3. Network of authors in the nursing informatics community.

As can be seen in Figure 3, there are many small disjoint groups of authors of similar symbols (e.g. circle) of the same color representing subcommunities. Typically, these small disjoint subcommunities represent each a small number of authors that have written one or a few articles together. Remarkably, there is a large group of interconnected authors in the center. This large subnetwork is made up of 418 authors that are connected by 1631 edges representing author-coauthor relationships. It is a rather compact network of authors where it takes on average only 4.6 steps to get from one author to the next. The most distant authors can be reached by just 11 steps in the network defining its diameter making it a rather compact network. For the large subnet, it is of interest to identify important (central) authors. Centrality measures provide relative measures of importance of a node in a network. The following measures use different criteria of

importance (Figure 4). Lighter colors and a larger size indicate a higher measure of centrality

- *Degree Centrality (Figure 4a):* Number of connections (edges). An important author has publications with many different coauthors.
- *Closeness Centrality (Figure 4b):* Average length of the shortest paths between an author and all other authors in the graph. An important author is typically close to and can communicate quickly with the other authors in the network.
- *Betweenness Centrality (Figure 4c):* Measures the extent to which a specific author lies on the path between all other authors. An important author will lie on a high proportion of paths between other authors in the network.
- *Eigenvector Centrality (Figure 4d):* An important author is connected to other important authors. The eigenvector centrality measure is based on the idea that an author is important if she is linked to by other important authors. Eigenvector centrality characterizes the “global” (as opposed to “local”) prominence of a node in a graph. Google’s PageRank algorithm is a variation of the eigenvector centrality

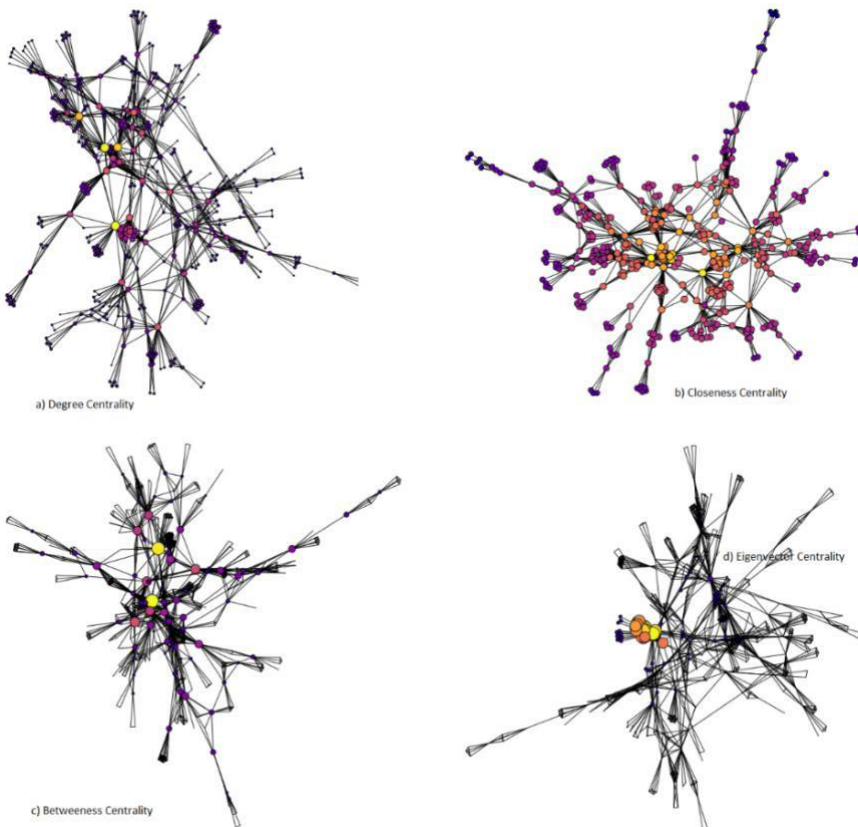


Figure 4. Different measures of centrality for the central subnet.

In the graphs (Figure 4a-d) each measure puts the emphasis on different authors depending on the criteria for importance. Closeness and betweenness centrality have an

important impact on the information flow through the community, while degree and eigenvector centrality can be related more to the influence an author might have.

In a further step to analyze and visualize the central subnet of authors (Figure 3) the Louvain method [4] is used to partition this subnet into smaller subcommunities. 13 different connected subcommunities can be found (Figure 5) that build up this subnet.

4. Discussion

Since only Pubmed has been used to determine the scientific community focused on nursing informatics, this paper only takes into account publications that are referenced in Pubmed, it also ignores other publication types. So only a limited proportion of the community might be observed. The ambiguity of the short author names used by Pubmed can be a further small source of errors. Since the analyzed publications found did span more than 40 years it does not reflect the dynamic of the scientific community with authors appearing or leaving, or their research focus shifting. It would be of interest to visualize the development of the community in time. At this stage, it is not obvious if there is a real difference between the subcommunities found (Figure 5). By looking at the MeSH terms primarily used for papers of each subcommunity, it looks like a subcommunity focused on nursing terminology, and another focused on educational topics can be identified. But this is a very preliminary hypothesis. Other hypotheses might be a regional clustering of authors. Unfortunately, the affiliation field in the Pubmed database is not used consistently, which makes it hard to analyze the affiliation data using automated processes.

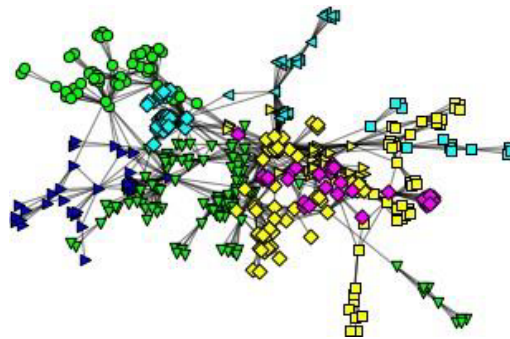


Figure 5. Subcommunities of the central interconnected subnet focused on nursing informatics.

5. Conclusions

Publications can be used as an indicator of the network structure of the nursing informatics community. It is made up of a large subnet of interconnected authors and many disjoint small subnets of one or a few authors. Centrality measures can be an indicator of important authors within the network, each measuring a different type of importance. In what aspect automatically detected subcommunities differ still has to be determined.

References

- [1] Kornfeld WA, Hewit CE. The scientific community metaphor. *IEEE Trans. On Systems, Man, and Cybernetics*. 1981 Jan. SMC-11:1–10.
- [2] Velden T. A new approach to analyzing patterns of collaboration in co-authorship networks: Mesoscopic analysis and interpretation. *Scientometrics*. 2010, 85:219-242.
- [3] Koseoglu MA. Growth and structure of authorship and co-authorship network in the strategic management realm. *BRQ Business Research Quarterly*, 2016. 19(3):153-170.
- [4] Blondel VD, Guillaume JL, Lambiotte R. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008 Oct. P10008.