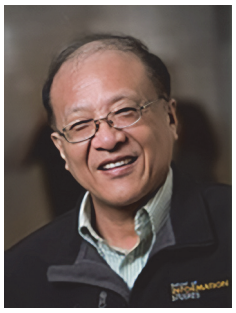
An Optimization Analysis of the Subject Directory System on the MedlinePlus Portal – An Investigation of Children Related Health Topics

Yifan Zhu* and Jin Zhang**

* School of Public Affairs, Zhejiang University, no. 866, Yuhangtang Rd, Hangzhou, Zhejiang 310000, China
** School of Information Studies, University of Wisconsin-Milwaukee, 2025 E Newport Avenue, Milwaukee, WI 53211, USA
*<yifanzhu1105@zju.edu.cn> / **<jzhang@uwm.edu>

Yifan Zhu is a postdoctoral fellow in the School of Public Affairs at Zhejiang University, China. His research interests include health information, knowledge organization, information seeking behavior, and information ethics. He holds a BM in information resources management from Sichuan University, a MLIS from the University of Wisconsin-Madison, and a PhD in information studies from the University of Wisconsin-Milwaukee.



Jin Zhang is a full professor at the School of Information Studies, University of Wisconsin-Milwaukee, U.S.A. His research interests include visualization for information retrieval, information retrieval algorithm, data mining, big data, consumer health informatics, social media, social network analysis, metadata, search engine evaluation, digital libraries, knowledge system evaluation, and human computer interface design. He has published papers in journals such as the Journal of the American Society for Information Science and Technology, Information Processing & Management, Journal of Documentation, Journal of Information Science, etc. His book "Visualization for Information Retrieval" was published in the Information Retrieval Series by Springer in 2008.

Zhu, Yifan, and Jin Zhang. 2023. "An Optimization Analysis of the Subject Directory System on the MedlinePlus Portal – An Investigation of Children Related Health Topics". *Knowledge Organization* 50, no. 4: 272-289. 60 references. DOI:10.5771/0943-7444-2023-4-272.

Abstract : In this study, a mixed-methods research approach was employed, integrating social network analysis, descriptive analysis, and inferential statistical analysis to examine the health topic subject directories and the interconnections among health topics within the subject directory system of the MedlinePlus portal. One hundred and fifty-nine health topics related to children's health as well as 1457 qualified keywords were collected and analyzed. As a result, 184 new connections (140 bidirectional and 44 unidirectional) were proposed to be added to the original subject directory on MedlinePlus. Five new core topics were identified as influential topics in the subject network. This new optimized structural network was proved to be significantly improved from the original one and the importance of the newly identified core topics were verified. The evaluations also included participants containing both medical professionals (2) and medical students (31). The user evaluation results confirmed that the recommendations suggested by this study are solid and effective. The findings of this research would improve the information searching effectiveness for the portal users and offer insights to public health portal creators.

Received: 8 February 2023; Revised: 21 May 2023; Accepted 30 June 2023

Keywords: MedlinePlus, subject directory, children health, subject networks.

1.0 Introduction

Web portals are playing a significant role among the Web resources that are available to the public for seeking health information. Some scholars posited that distinctive structural characteristics of a Web portal can considerably influence user behaviors, including information-seeking activities and decision-making processes (Baird et al. 2012). For instance, some scholars argued that information systems should focus on their interaction features such as website

structure and information organization, webpage display, usability, as well as system performance (Li et al. 2021).

In an examination of numerous public health portals, two primary search functions had been identified: subject directories which facilitate browsing and search engine capabilities that allow for query searches (Zhang et al. 2015). As explained by Ellis and Vasconcelos (2000), subject directories function as indices for diverse subjects, arranged in a hierarchical structure to proffer an organized and easily navigable information landscape. Such directories empower users to sift through information, navigating through primary and subsidiary categories, thereby offering flexibility to explore areas of interest in either a broad, narrow, or related context. Conventionally, the construction of these subject directories involves the expert hand of human indexers, who apply advanced indexing and abstracting methodologies to ensure the precision and exhaustive nature of the subject directories. Since the advent of Web-based information seeking, the focus on subject directories and the method of browsing has been significantly amplified. Typical scholarly efforts were centered on ontology (Naskar and Das 2019) and the discernment of user preferences (Oh et al. 2015).

Among those comprehensive sites that are universally regarded as online health portals, MedlinePlus is normally considered a successful representative. The portal was launched in 1998 by the National Library of Medicine and was the first primary initiative for providing online health information to the public (Miller et al. 2000). The MedlinePlus portal uses an organized subject directory system to enable users to browse information from a general level to a specific level through its hierarchical structure, and vice versa. Meanwhile, related subjects (health topics) are also listed in Web pages as individual topics so that users can "jump" to relevant topics when necessary. In other words, the MedlinePlus platform exemplifies an effective application of subject directories, significantly enhancing public comprehension of health-related issues and bolstering health literacy (Ahmed 2019). As per the findings in Ahmed's research, the MedlinePlus platform is meticulously organized, with the lion's share of its content categorized into more than 1000 health topic pages, aimed at both English and Spanish-speaking audiences. These curated pages provide an extensive range of health information, encapsulating aspects like body systems, various disorders, and conditions, and are customized to suit the needs of different demographic groups. Each topic page is distinct, incorporating resources from related health topics. Such a topic-based subject directory system is described by Ahmed as a "portal" model, a model that has persistently been a dependable reservoir of health information.

Compared with other health consumer groups, the children group is a unique one considering its two outstanding characteristics. First, children usually depend on their parents or caregivers. Hence, when evaluating how children related health information is organized, researchers often investigate mothers (Bernhardt and Felter 2004), parents, and preschool caregivers (Naidu et al. 2015) rather than children because these are the ones that actually search for children related health information. Second, due to the immaturity in both physical and mental aspects, compared with other demographic groups, such as men and older adults that are listed on MedlinePlus, children are faced with a wider range of health conditions that include, but are not limited to, daily health problems such as oral health and fever (Fallis and Frické 2002), growth and nutrition problems (Kuperminc and Stevenson 2008), genetics related issues (Silventoinen et al. 2021), as well as mental disorders (Liu et al. 2020). Such facts indicate that it is necessary to investigate if the current children related subject directory system on MedlinePlus is fulfilling the information needs of children, their parents, as well as caregivers.

This research study focuses on the evaluation and optimization of the children related topic-based subject directory employed by the public health portal MedlinePlus. The primary research problem of this study is to explore, assess, and optimize the connections among the children related health topics in the subject directory of MedlinePlus. The research objects are the health topics related to children on MedlinePlus. These health topics refer to the corresponding topic pages created by MedlinePlus for introducing related information and resources. The connections among the health topics hold significance for this investigation, as they facilitate the construction of networks for later analysis. The connections among the health topics can be divided into two types: structural connections and semantic connections. To be more specific, structural connections represent the physical linkages set by the portal creators for connecting a specific health topic to other topics, while semantic connections refer to the linkages hidden behind the textual information of the topic pages among various health topics. In other words, the structural connection between two health topics is ascertained by the presence of embedded links, whereas the semantic connection is established based on the similarity of the respective topics' Web page content.

Based on the primary research problem, this study attempts to address the following three research questions:

RQ1: How are health topics related to children connected in the subject directory on the MedlinePlus portal: are the structural and semantic connections of the health topics consistent?

RQ2: Are there significant differences between the original children subject directory and the optimized children subject directory in terms of its link structure on the MedlinePlus portal?

RQ3: Are there any significant differences between the optimized subject directories from this study and the evaluation results from the health field experts?

2.0Literature review

Health consumers have shifted from passively accepting orders from their physicians to actively searching for online information that relates to their health conditions (Nie et al. 2014). As Ellis-Danquah (2004) pointed out, the comprehensiveness of online health information has the potential to enable consumers to make important health-related decisions. However, they have to overcome the challenges brought by the myriad of Web information resources and find specific, reliable, and timely health information first.

2.1 Browsing and subject directories

Previous studies have revealed that a substantial proportion (80%) of online users seeking web-based information initiated their search from a general portal site (46%) rather than a search engine like Google (33%) (Horrigan et al. 2003). Therefore, Web portals could be seen as the front door for the information needed by online users (Zhang et al. 2016). More than that, as users always desire easy access to and valid information from Web portals, the interaction between portals and users has been enhanced under the integration of Web 2.0 (Postigo 2011). Public health portals such as MedlinePlus predominantly employed a portal model in which information was organized by subjects or categories (Ahmed 2019). Despite the considerable advancements in search capabilities and search engine-based interfaces over the past decade or so, the portal model and its subject directories continue to play a crucial role as a valuable source of health information. Noticeably, in recent years, online health communities (OHCs) have raised wide attention among scholars on Q&A sites regarding diabetes related issues (Zhang and Zhao 2013), knowledge sharing among physicians (Qiao et al. 2021; Yang et al. 2021), user identities (Zhao et al. 2022), as well as information needs of patients and their family members (Ma et al. 2021).

When searching information in Web portals, compared with search engines, some researchers considered the subject directory services as a better retrieval tool since it provides more satisfying results that have gone through manual review and classification (Chung and Noh 2003). When it comes to the health area, as Yeo et al. (2010, 60) pointed out in their study, "existing disease information systems support the classification of disease data and provide users with the data through the web" while the data formats applied among different databases vary. Previous research considered browsing as an alternative method for acquiring information in an increasingly fragmented digital landscape. It was argued that browsing potentially facilitates a more profound engagement for information seekers in comparison to search engines and social media platforms (Möller et al. 2020). Meanwhile, the utilization of the Internet as a primary source of health information by laypeople continues to grow. Nonetheless, the process of locating and identifying appropriate information remains a challenge as there is an insufficient level of support provided for the exploration of health information within these platforms (Pang et al. 2016).

Subject directories play an important role in organizing information on the Internet, with some specialized systems following standard library schemes like DDC or UDC while others use schemes specifically designed for online resources (Chung and Noh 2003). In the case of public health portals, subject directories are generated from multiple sources such as user interviews, literature reviews, communication with clinical institutions, and information exchange with other health portals (Schilling and McDaniel 2010). However, the way in which subject directories are organized can vary greatly, despite sharing common sources for classification (C. Gray 2005). One of the advantages of subject directories is that they are humanly indexed, which tends to result in more relevant information being retrieved compared to other search tools (Ellis and Vasconcelos 2000). However, Ellis and Vasconcelos also noted that subject directories have the drawback of requiring more time to include new subjects and review them. As a result, the success of subject directories highly depends on professional human experts and the efficiency of adding new subjects, which can be challenging due to the vast amount of online information available.

Given these issues, one of the primary tasks of a subject directory is to effectively and efficiently maintain and revise its categories. To do so, it is important to come up with a practical method to optimize a subject directory based on the semantic connections among the items it describes and organizes. Before automatic means for producing directory systems occurred, most of the subject directories on the Web for either general or specialized sites were classified or categorized by the editors or surfers until some researchers like Chung and Noh (2003) and Yang and Lee (2004) proposed directory systems based on automatic classification, text-mining technology based on Self-Organizing Map, and text categorization techniques to automatically assign Web content to relevant subject categories, thus reducing the manual working hours. Although Chung and Noh highlighted that the sustainability of manual classification efforts became untenable as the volume of Web documents experienced a substantial growth, they did admit that their automated classification scheme based on *k*-nearest neighbors (kNN) could only achieve a precision of 77%, thus requiring human subject specialists to verify and reassign in-

correct classified documents. Consequently, it can be inferred that automated classification and its associated techniques are best employed as ancillary tools, with the aim of streamlining the time and effort dedicated to the creation of subject directories. Moreover, some other prior studies attempted to better collect and understand the vocabulary used by specific groups of consumers from other creative ways such as users' searching query terms through transaction logs (Zhang and Wolfram 2009). Recently, as social network analysis arose with greatly increased online searching, researchers have also proposed to optimize subject directory through a social network analysis of the structural and semantic relationships among various subject terms so that the interconnectedness among categories and subcategories of subject directory could be improved (Zhang et al. 2015; Zhang et al. 2016; Zhu and Zhang 2020).

2.2 Children related health information

According to the literature, children's health seems to be a common topic among the experiences of seeking health information online (Bray et al. 2019; Meppelink et al. 2019; Dalton et al. 2020). Earlier studies used to examine the perspectives of children, parents, and health professionals on providing preparatory information to children undergoing medical procedures. They found that children's health literacy is influenced by adults, often leading to misconceptions about procedures, thus highlighting the need for improved health literacy, and emphasizing the importance of addressing children's concerns to reduce anxiety and ensure a better understanding of medical procedures (Bray et al. 2019). However, the seeking behavior of children has been rarely discussed in the previous literature. Considering the lack of required online searching literacy and other qualifications among children, their needed health information, for instance, a cough condition (Pandolfini et al. 2000), would normally be sought by their parents (Khoo et al. 2008).

Such a fact was discussed by scholars like Meppelink et al. (2019), and they underscored the importance of addressing confirmation bias in health communication, particularly for parents with high health literacy levels seeking early-childhood vaccination information. Meanwhile, prior researchers also pointed out that amid the rapidly changing COVID-19 pandemic, children are exposed to vast information and adult stress. Children, even as young as two years old, notice these changes. It is essential for adults to communicate honest, age-appropriate information to them. Considering the child's developmental stage is crucial for effective communication. Providing accurate and meaningful explanations can benefit the long-term psychological well-being of children and their families (Dalton et al. 2020).

Based on the WHO website (https://www.who.int/health-topics/child-health#tab=tab_1) and MedlinePlus portal (https://medlineplus.gov/childrenandteenagers.html), children are closely linked to a group of health issues including diabetes, violence, nutrition, growth disorders, genetic problems, school health, and mental health. As for the resources containing children related health information, a list of nine websites specifically designed for children to get access to health information used to be identified by Izenberg and Lieberman back to 1998 (Izenberg and Lieberman 1998), and three of those portals, including Children with Diabetes, Dole 5 a Day, and KidsHealth.org, are still available today. However, prior studies had questioned the reliability of health information related to children, such as children with high fevers (Impicciatore et al. 1997), and childhood diarrhea (McClung et al. 1998), offered by some online websites. As a result, the researchers pointed out that even from those major academic medical centers' online portals, health information might be inaccurate and of low quality.

Among all children related conditions, mental illnesses account for a high proportion in the prior literature. Typical mental disorders such as ADHD (attention deficit/hyperactivity disorder) were included in a few earlier studies (Sage et al. 2018; Tandi Lwoga and Florence Mosha 2013). For information about mental illness related to children, not surprisingly, the Internet had been recognized as one of the primary sources (Bouche and Migeot 2008; Sage et al. 2018), or even the most important resources in Tanzania (Tandi Lwoga and Florence Mosha 2013).

2.3 Social network analysis

Tracing its origins to the 1930s, social network analysis was first conceptualized as a novel theory by Barnes (1954), with its foundations grounded in social action theories (Coleman 1986). Consequently, by the 1980s, social network analysis had emerged as a well-established discipline within the realm of social sciences (Borgatti et al. 2009). Social network analysis has been defined as a methodology for exploring social structures through the application of network and graph theories (Otte and Rousseau 2002). It provides a research framework to measure structural relationships between members within a network and intends to reveal reality occurring among the interactions and progress behind the scenes (Borgatti et al. 2009). The goal of social network analysis is to discover and measure structural relationships among entities and nodes within a given network (Zhang et al. 2016; 2015). For Web portals, Zhang et al. (2016, 2168) stated that social network analysis could be employed to "gauge and compare the connection network structure and semantic network structure of the subject directory".

In the context of the healthcare sector, previous scholarly endeavors have explored the applications of social network analysis. A comprehensive examination of 52 related studies within the healthcare domain revealed an emergent theme; the scholars recommended that subsequent research efforts should pivot toward the assessment of interventions underpinned by social network analysis (Chambers et al. 2012). Moreover, health care settings were considered to be possibly understood better if social network analysis was combined with public health communication methods (Luke and Harris 2007). Prior investigations in health informatics have encompassed the examination of information exchanges among health consumers (Mertens et al. 2012), the influence of culture on consumers' health information-seeking behavior (Smith and Christakis 2008), the utilization of feedback from adolescents regarding health information (N. J. Gray et al. 2005), and hospital nursing (Pow et al. 2012).

In recent years, social network analysis was also applied with other proper research methods (e.g., content analysis), especially when the research target shares common characteristics with social network analysis and other specific approaches. For example, a study examining breast cancer-related health information on Twitter was conducted by Kim et al. (2016). Conversely, several researchers proposed the intriguing concept of applying social network analysis to assess the topic or subject-based navigation systems of public health portals, such as the World Health Organization (Zhang et al. 2015) and the government agriculture portal (Zhang et al. 2016).

To summarize, the existing body of literature has primarily focused on social network analysis of sites and search-based navigation approaches. Past research has largely addressed health consumers actively seeking specific health information or emotional support. However, the importance, benefits, and potential growth of well-organized health information for general browsing purposes, as provided by professional health institutions, have not been thoroughly investigated. Moreover, subject directory systems pertaining to children have been infrequently addressed in prior studies. An examination of the children's section within a representative health Web portal's subject directory could yield opportunities for improving the system's efficacy and afford a fresh perspective for health consumers and professionals alike in understanding relevant health information.

3.0 Methodology

3.1 MedlinePlus and its health topics

MedlinePlus delivers trustworthy and contemporary health information, addressing the requirements of a broad spectrum of health professionals and health consumers. The portal uses five broad sections to divide its over 1000 health related issues: Body Location/Systems, Disorders and Conditions, Diagnosis and Therapy, Demographic Groups, and Health and Wellness. Each of these five sections includes a list of subcategories, and each subcategory has an introduction page that contains a group of health topics in an alphabetical order.

A health topic pertains to a particular health-related concern and is represented by a distinct Web page, commencing with an overview table encompassing sections such as “basics,” “learn more,” “see, play, and learn,” and “research,” among others. Concurrently, the Web page features a “related health topics” column situated on the right side, designed to assist end-users in navigating to other relevant health topics. For instance, the topic Children’s Health has 13 related health topics: Child Dental Health, Child Development, Child Safety, etc.

The dataset for this study was extracted from the children subcategory on the MedlinePlus portal. Three steps were followed to collect the topic data. The first step was to determine the health topic that could serve as the starting point for the subcategory group relating to children’s health. It decided the initial numbers as well as specific related health topics to start to construct the structural link network. The second step was to expand the health topic group into a reasonable size through involving more health topics listed under the “related health topics” column from the previously selected health topics’ Web pages. A proper group size enabled this study to generate a meaningful and strong network so that in-depth explorations could be applied to an appropriate range of the subject directory system. The last step was to collect data related to each selected health topic through its individual Web page. The data gathered contained two parts: all the health topics listed as the “related health topics” were obtained as structural connection data and all the textual information in the introductory section was collected as semantic connection analysis data.

The sampling strategy included the first and second steps mentioned above. For these two steps, the general process was that once a specific health topic was selected to serve as the starting point, this initial topic and the health topics listed under the “related health topics” column of the initial topic were then collected to form the first level of health topics. Subsequently, the health topics listed under the “related health topics” column of each health topic included in the first level were then collected to form the second level of health topics. This data collection process was repeated until the attainment of a third level was achieved. An ideal health topic group should contain between 100 and 150 topics to generate a corresponding social network so that the network could possess enough information for later analysis. The determination of the group size in this study was based on the relevance of health topics collected across various levels, as the relevance tends to decrease with

increasing distance from the initial health topic. For instance, within the MedlinePlus subject directory pertaining to children's health, the starting topic Children's Health encompasses 13 closely related health topics in the first level, including Childhood Immunization and Hearing Problems in Children. Upon examining the second level, out of 54 health topics, a portion remains highly relevant to Children's Health, such as Child Behavior Disorders, Growth Disorders, and Developmental Disabilities. However, certain topics, like Hepatitis B, Benefits of Exercise, and Assistive Devices, appear less relevant. In the third level, the relevance of the 91 topics to Children's Health continues to decrease, making it challenging to find strongly related topics, such as Child Nutrition. At the fourth level, topics become even more diffuse, pertaining to different demographic groups or specific conditions, organs, or body systems. Examples include Depression, Miscarriage, Tumors and Pregnancy, Vitamin A, and Calcium. Consequently, the relevance of topics at deeper levels diminishes in comparison to the initial topic. Based on this analysis, a group size of 100-150 topics at three levels was deemed reasonable.

The data were collected in August 2020. All the information collected for this study is publicly accessible online (https://medlineplus.gov/healthtopics.html). Although MedlinePlus has been observed to make minor adjustments to its subject directories regularly, no time patterns were found to exist among health topics.

The Web pages encompassing all three levels of health topics were collected, and the text from these topics' pages was extracted using a coded Python program, resulting in the creation of a word list. Textual information from the overview table as well as side menu text and navigating hyperlinks were ignored because the textual information made no difference among various health topics' pages – it is standardized by the portal creators to keep a consistent format among different Web pages of health topics. Therefore, the complete set of textual information, ranging from the entries in the "basics" column to those in the "for you" column, was comprehensively collected. The automatically collected textual information was further filtered through manual review. After the review process, it was used as semantic data in this study. Subsequently, this word list underwent further refinement. Initially, a stop-word list was utilized to eliminate irrelevant words. Secondly, synonyms were consolidated. Lastly, all the words on the word list were converted to their regular forms. For instance, "psychotherapies" was converted to "psychotherapy".

3.2 The formation of social networks

In this study, each health topic selected from the subject directory of MedlinePlus served as a node while the structural connections among these health topics served as the edges. Unlike the structural connections which could be directly observed through the hyperlinks set by the portal creators, the semantic connections were not directly reflected through any visible links. Therefore, there were no edges in the network that could represent the semantic connections possessed among the health topics. Some of the edges between two health topics are unidirectional, while others are bidirectional. It was important to differentiate unidirectional edges from those bidirectional ones because the former type of edges may cause "dead ends" in the subject directories and prevent users from navigating back to the previous health topic's page.

A group of matrices were built to represent the structural link network and the semantic network among the selected health topics related to children:

The initial matrix, known as the subject link matrix (SLM), illustrates the structural connections between the collected health topics. Within this matrix, l_{ij} signifies whether health topic i is classified as a "related health topic" of another health topic j, and 'n' represents the quantity of health topics chosen from the subject directory, resulting in an $n \times n$ asymmetrical matrix. In the matrix, the cell value of l_{ij} conveys the relationship between two health topics – if topic i is included in the related health topics for topic j, the cell value l_{ij} is allocated a value of 1; otherwise, it receives a value of 0. It is important to note that the matrix is asymmetrical, as the linkage of topic i to topic j does not ensure a reciprocal connection from topic j to topic i (Equation (1))

The equation formulated for the SLM is depicted below, illustrating that a health topic cannot contain itself as a connected health topic (Equation (2)).

Following the SLM, a subject-keyword matrix (SKM) is developed. In this matrix, each row is associated with a

$$SLM = \begin{pmatrix} l_{11} & \cdots & l_{1n} & \cdots & l_{ij} & \cdots \\ & & l_{n1} & \cdots & l_{nn} & \end{pmatrix} \quad (1)$$

Equation 1.

$$l_{ii} = 0, \qquad 1 \le i \le n \quad (2)$$

Equation 2.

health topic, and each column corresponds to a keyword originating from the word list previously generated. The cell value reflects the level of connection between keyword j and topic i. This connection is determined by the weight, or the term frequency (tf), of the keyword presented in the respective topic's Web page. Equation (3), established for the SKM, is shown below. Here, f_{ij} denotes the frequency occurrence of keyword i in a health topic j's Web page. 'n' represents the number of chosen health topics, while 'm' indicates the total number of keywords in the word list that have been extracted from the health topic's Web pages.

Following the creation of SLM and SKM, a subject-semantic matrix (SSM) was developed to depict the semantic connections among the selected health topics by utilizing similarity measures based on the subject-keyword matrix (SKM). Equation (4) presents the subject-semantic matrix. In this equation, s_{ij} represents the similarity value between health topics i and j, while 'n' signifies the number of chosen health topics. For similarity measure, the cosine-similarity measure was used as the major similarity measure in this study along with the Pearson correlation similarity measure and the Euclidean distance similarity measure for comparing purposes. The Pearson correlation similarity measure, as described by Segaran (2007), is employed to calculate the strength of two correlated variables. The Euclidean distance similarity measure serves as the basis of many measures of similarity and dissimilarity, including Manhattan Distance, Minkowski Distance, etc. (Borgatti et al. 2013). Therefore, these two similarity measures were applied as supplemental methods to compare and verify the results generated according to the cosine similarity measure. The cosine-similarity is displayed in Equation (5).

According to Equation (5), the cell value (similarity) between topic i and topic j is maintained as equal to that between topic j and topic i, indicating that the SSM is a symmetrical matrix. Additionally, the SSM is non-directional since both directions exhibit the same similarity between two topics from a semantic standpoint.

3.3 Data analysis and optimization method

Upon determining the similarities among all health topics in the semantic network, each edge was assigned a similarity value based on each of the three similarity measures. These edges were subsequently categorized into one of three groups: Edge Set A, Edge Set B, and Edge Set C. Edge Set A contains edges where a topic is linked to itself, yielding a similarity value of 1. Since these edges do not contribute to subsequent analysis, they were omitted. Edges in Edge Set B had corresponding links in the structural link network, while edges in Edge Set C did not share any links with the structural link network.

The average similarity for edges in Edge Set B was calculated and employed as the threshold for selecting recommended topic edges within the structural link network, a procedure known as optimization of the structural link network. If an edge in Edge Set C exhibited a similarity value exceeding the threshold, the corresponding edge/link was suggested for addition to the structural link network. The recommended edges in Edge Set C had two situations: 1) no connection between two topics, T1 and T2, in the structural link network, and 2) merely a single connection from topic T1 (T2) to topic T2 (T1) within the structural link network. Both instances were considered. As a result, the recommended edges formed a new set, Edge Set D, which is a subset of Edge Set C. It is clear that the optimized or final structural link network encompasses both Edge Set B and Edge Set D.

In this study, the node-level measurements from the social network analysis were applied to investigate the characteristics of each health topic's position within the network. To be more specific, the centrality was selected as the actor

$$SKM = \begin{pmatrix} f_{11} & \cdots & f_{1m} & \cdots & f_{ij} & \cdots \\ f_{n1} & \cdots & f_{nm} & & & \end{pmatrix} \quad (3)$$

Equation 3.

$$SSM = \begin{pmatrix} s_{11} & \cdots & s_{1n} & \cdots & s_{ij} & \cdots \\ s_{n1} & \cdots & s_{nn} & & & \end{pmatrix} \quad (4)$$

Equation 4.

$$s_{ij} = \frac{\sum_{k=1}^{n} f_{ik} \times f_{jk}}{\left(\sum_{k=1}^{n} f_{ik}^2 \times \sum_{k=1}^{n} f_{jk}^2\right)^{\frac{1}{2}}} \quad (5)$$

Equation 5.

feature for comparison. The centrality consists of three measurements to accurately identify crucial actors or nodes and gauge their connections or edge strength within a network. It lays the foundation to specify and evaluate the potential optimized edges for the study. The centrality includes degree, betweenness, and closeness. It can indicate the importance of an actor within a network from multiple perspectives. For instance, the degree refers to the number of connections an actor possesses to other actors within the network. The betweenness represents the number of an actor sitting between pairs of other actors on their shortest paths (Freeman 1978). The closeness measures the extent to which an actor is distant from other actors within the network (Freeman 1978).

Based on the analysis results of centralities, a group of influential health topics were discovered. Compared with other actors, influential actors can have more impact on other actors. They play a more important role in controlling the information flow among the whole network through locating at a comparatively central position. The influential health topics on the portal can be recognized when a topic is found to have many "related health topics" on its Web page and being listed as one of the "related health topics" on many other topics' Web pages. These characteristics quantify the impact or contribution of a health topic to the subject directory and provide quantitative data to investigate the relationships among the health topics in the network.

3.4 Evaluation

After the health topic suggestions were made for optimizing the current children related subject directory system, two groups of evaluators were invited to assess the optimization results. One group contained two medical professionals while the other group involved 31 medical students as general users. The two medical professionals were recruited from a formal research institute in the United States. They either went through the preliminary examination of a M.D. program or obtained at least a master's degree in a medical or health related field. They have had at least five years of experience in the field. For the general user group, the students were recruited from two formal research institutes – one in China and the other in the United States. The users were required to major in a medical or health related discipline, and both undergraduates and graduates were recruited.

In terms of the evaluation lists and procedures, for the medical expert group, a list that contained around 200 paired health topics were generated. Among them, about two thirds of the paired health topics were health topics which were suggested to be added to the structural network in the subject directory of MedlinePlus. Besides those recommended topics, the remaining one third of paired health topics were also evaluated. Those health topics were not linked on the structural link network in the MedlinePlus portal and were found to have low semantic connections according to the similarity results generated in this study. They were used for comparative purposes. Such a mixed list of paired health topics could avoid potential bias from the experts. A screenshot of a health topic page, featuring its "related health topics" list, was presented to the evaluators along with a succinct explanation of the methods employed by the MedlinePlus portal to generate and showcase structural connections among pertinent health topics. Subsequently, evaluators were instructed to discern and indicate pairs of health topics they deemed relevant.

Different from the expert group, the evaluation list for the general user group was shorter – the experts' evaluation list mentioned above was divided into three sub-lists with various focuses. All the sub-lists for the user group maintained the same composition – each sub-list contained around 50 recommended health topic pairs along with 25 irrelevant paired topics. The users were then divided into three sub-groups accordingly, and each user sub-group evaluated a corresponding sub-list. The same set of evaluation materials was subsequently disseminated among the three distinct user sub-groups. In this context, they played the role of regular MedlinePlus subject directory system users in order to scrutinize and assess the recommended results.

3.5 Data analysis instruments

The analyses were processed using UciNet (Version No. 6.669) and SPSS (Version No. 25.0).

4.0 Results and discussion

4.1 The original structural network analysis

The starting health topic was Children's Health, and a total of 159 health topics distributed at three levels were included. From the introductory pages of the 159 health topics, textual content was extracted. A validation process was carried out, and all stop-words were eliminated. Keywords with only a single occurrence were disregarded, revealing 1919 keywords that featured in a minimum of two separate health topics' pages. Lastly, when refining the keyword compilation further at cut-off point 2 to obtain more meaningful similarity assessment results, the total number of keywords amounted to 1457.

On the other hand, 472 structural connections about children's topics were observed in the original network created by the MedlinePlus portal. A visualized figure of the structural link network of children related topics is displayed in Figure 1.

Figure 1. Display of the structural link network of children related topics

In Figure 1, both unidirectional and bidirectional connections were identified in the structural network of the children subcategory. Based on the visual observation in Figure 1, the children related health topics in the structural network were evenly distributed – no topics were found to be apparently clustered within this network.

With respect to the semantic network, the cosine similarity measure was employed to ascertain the similarity between two health topics' pages based on the textual keywords. Consequently, the average similarity value for the 472 structurally connected edges in Edge Set B amounted to 0.444744, whereas the average similarity for the remaining 24,650 pairs of topics (Edge Set C) was 0.086589. The overall average similarity stood at 0.093318.

Subsequently, among the 24,650 semantic connections absent in the structural link network, 184 pairs of topics exhibited a similarity value surpassing the threshold (i.e., the average similarity value of Edge Set B). This indicated an inconsistency between the structural and semantic connections. Out of these 184 pairs of topics, 140 pairs were recommended to establish bidirectional connections, with ten sample pairs of these edges presented in Table 1.

In addition to the aforementioned 140 pairs of bidirectional edges, there were 44 unidirectional pairs of health topics possessing a similarity value exceeding the threshold. Six (6) sample pairs of these health topics are shown in Table 2.

According to Table 2, both hierarchical and associative relationships were identified in the recommended unidirectional connections of children related health topics. For instance, Pregnancy and Pregnancy and Medicines were presenting a hierarchical relationship while Children's Health and Toddler's Health were presenting an associative relationship.

Going forward, the node-level features including the three centrality measurements were analyzed. Those health topics that were playing the most influential roles were identified and listed. In the meantime, same node centrality features were investigated from the revised semantic-based network and key health topics were identified and compared.

As a result, two ranking lists were built in terms of each of the three node centrality measurements – one was based on the original structural link network, and the other was based on the revised semantic-based network. The ranked health topics located on each ranking list were compared regarding their consistency. If a health topic was ranked high in the revised semantic-based network but low in the original structural link network, it indicated that this health topic was an influential one within the subject directory in terms of its semantic connections with other topics. However, such semantic influence was not reflected in the original structural link network. Hence, those health topics with strong semantic connections but which did not appear in the original network should be identified and integrated

Pairs of Topic A & Topic B		Pairs of Topic A & Topic B	
Topic A	Topic B	Topic A	Topic B
Children's Health	Child Behavior Disorders	Child Dental Health	Dentures
Children's Health	Toddler Development	Child Dental Health	Child Mental Health
Children's Health	Baby Health Checkup	Child Dental Health	Child Nutrition
Children's Health	Child Mental Health	Child Development	Medicines and Children
Child Dental Health	Toddler Health	Child Development	Toddler Health

Table 1. 10 sample bidirectional pairs of health topics that require structural linkages in the children related topic group

Topics and their related topic		Topics and their related topic	
Topic	Related topic	Topic	Related topic
Children's Health	Child Nutrition	Teen Mental Health	Teen Health
Child Dental Health	Children's Health	Teenage Pregnancy	Teen Health
Child Dental Health	Cosmetic Dentistry	Toddler Nutrition	Toddler Health

Table 2. 6 sample unidirectional edges recommended for the children topic group

Out_Degree	Revised_Out_Degree	In_Degree	Revised_In_Degree
Childhood Immunization	Child Development	Children's Health	Children's Health
Medicines	Medicines and Children	Childhood Immunization	Medicines and Children
Child Development	Children's Health	Medicines	Toddler Health
Medicines and Children	Toddler Health	Birth Defects	Medicines
Hearing Disorders and Deafness	Child Mental Health	Uncommon Infant and Newborn Problem	Uncommon Infant and Newborn Problems
Birth Defects	Child Dental Health	Hearing Disorders and Deafness	Child Development
Child Dental Health	Childhood Immunization	Teen Mental Health	Childhood Immunization
Dental Health	Teen Health	Sports Fitness	Teen Mental Health
Child Behavior Disorders	Child Behavior Disorders	Assistive Devices	Birth Defects
Sports Fitness	Infant and Newborn Care	Common Infant and Newborn Problems	Child Mental Health

Table 3. Degree centrality of the children subcategory

into the original structural link network to make the subject directory more effective.

Table 3 displays the ranking lists of the degree centrality features between the original structural and revised semantic-based networks of the children topic subcategory. The three health topics that had the largest differences in terms of their rankings on the ranking lists are highlighted for both out-degree and in-degree centralities. Here, the in-degree and out-degree refer to the inbound and outbound edges possessed by a given node, respectively:

From Table 3, four different health topics were identified as key nodes in the network in terms of their in-degree and out-degree centrality. To be more specific, Children's Health, Toddler Health, Medicines and Children, and Child Mental Health were not included in the top 20 rankings of the original structural network. However, these four topics possessed strong impact in terms of their semantic connections with other topics in the network. Hence, the importance of these four health topics, especially the two topics that occurred in both the in-degree and out-degree ranking lists (Toddler Health and Child Mental Health), was underestimated in the original subject directory.

Following the same procedures described above, six health topics were identified as key nodes in the network in

terms of their in-closeness and out-closeness centrality. For the out-closeness feature, Children Development, Children's Health, and Toddler Health were not included in the top 20 rankings of the original structural network. For the in-closeness feature, Medicines and Children, Child Mental Health, and Child Dental Health were not included. Similarly, the importance of these six health topics was underestimated in the original subject directory despite that they served as key nodes in the semantic-based network.

Regarding the betweenness centrality feature between the original structural and revised semantic-based networks, the three health topics that had the largest differences in terms of their rankings on the ranking lists are Medicines and Children, Prenatal Testing, and Child Safety.

4.2 The optimized structural network

Since the structural and semantic connections were not consistent regarding children related health topics in the subject directory of MedlinePlus, recommendations were made for a group of health topics in order to add more relevant connections that had high similarity values in terms of Web page textual information. Moreover, some hidden core health topics were detected through the comparison of the ranking lists between the structural and semantic network. As a result, an optimized structural link network was developed. This optimized subject directory is shown in Figure 2.

In Figure 2, the original structural connections are shown in black lines while the recommended connections are in red lines. It is clear to find that a lot of recommended connections were added to the health consumer groups (i.e., children, teenagers, toddlers, infants, older adults etc.) related health topics. Moreover, newly suggested connections were also found in the health topics relating to pregnancy, hepatitis and daily exercises. After the optimization process, the structured connections in the optimized subject directory combined both Edge Set B and Edge Set D. Hence, there were 656 connections in total as a result, thus leading the average similarity value to increase to 0.476088. The average similarity among the rest of the 24,466 pairs of topics (new Edge Set C) was decreased to 0.083055.

In terms of the key nodes, the five core health topics identified were: Medicines and Children, Children's Health, Toddler Health, Child Development, and Child Mental Health. Combining all connections relating to these five core topics, there were 57 suggested links in total, excluding the overlapped links. These 57 suggested links weighted 30.98% of all the recommended links, hence they could prove the important roles played by the five core topics.

To verify that there are significant differences between the original and optimized structural networks of children related topics in terms of similarity values, a null hypothesis was created:

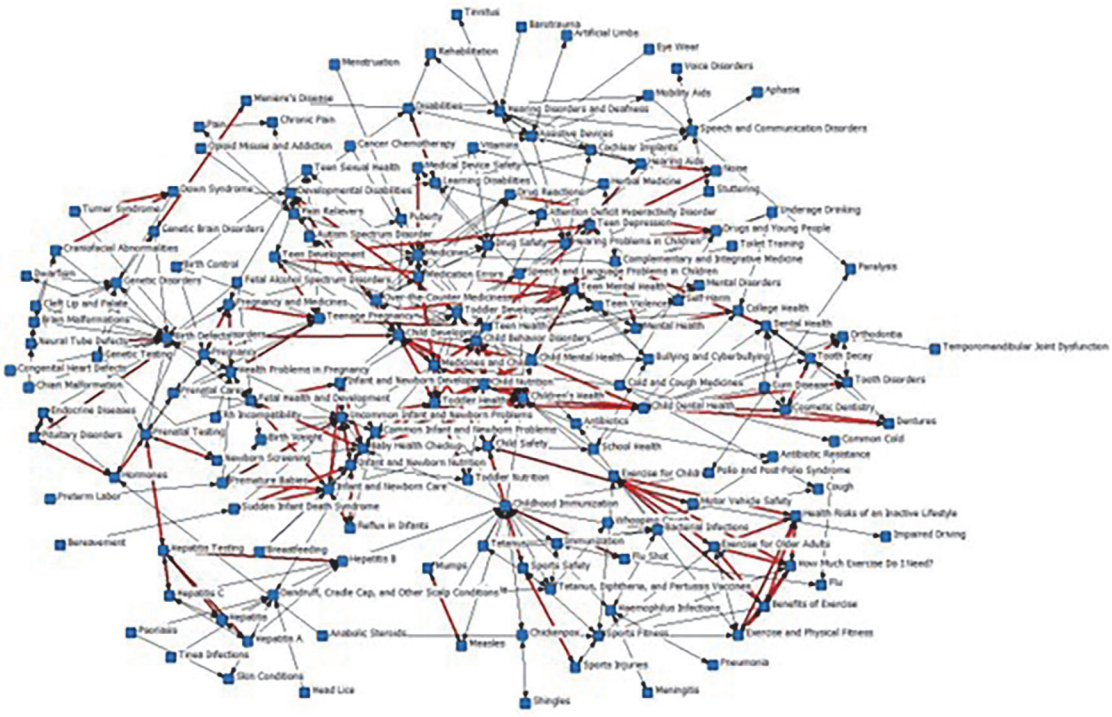


Figure 2. Optimized children structural link network

H01: There are no significant differences between the original and the optimized structural networks in terms of similarity values among the health topics related to children on the MedlinePlus portal.

The similarity values of the selected children related health topics were not able to be analyzed through the standard T-test since they did not follow a normal distribution. Therefore, the Mann-Whitney test was applied to investigate the differences of similarity values between the original and optimized subject directories. Table 4 summarizes the mean rank and sum of ranks of the similarity values from the original structural network (n=472) and the optimized structural network (n=656). The statistical analysis revealed that there was a systematic difference (p=0.002<0.05). Therefore, the hypothesis was rejected. In other words, the similarity value of the connections was significantly improved after the optimization process.

Similarly, node-level centrality data could not be analyzed through standard inferential statistical tests because the data did not follow a normal distribution either. However, UciNet uses a series of permutation tests (also called randomization tests) to modify the standard methods and makes the data suitable for the revised inferential statistical test (Borgatti et al. 2013). Therefore, to verify that there are significant differences between the original and optimized structural networks, a series of the customized T-tests were applied to investigate the differences of the three node centrality features. As a result, the p-values of both out-degree and in-degree centrality were smaller than the significant level (0.05), which indicated that there was a significant difference of degree centrality measures between the original structural network and the optimized structural network.

Similarly, the p-values of both out-closeness and in-closeness centrality were smaller than the significant level (0.05) as well, which indicated that there was a significant difference of closeness centrality measures between the original structural network and the optimized structural network. In other words, both the degree and the closeness centrality measures in the optimized structural network had generated a significantly higher average value than the original structural network.

Unlike the degree and the closeness centrality measures, the p-value of the betweenness centrality was larger than the significant level (0.05), which indicated that there was no significant difference between the original and the optimized structural networks. The reason for this phenomenon is that after the optimization, more connections were added to the structural network, and more pairs of health topics had been linked directly. Betweenness centrality denotes the frequency with which a node functions as a connector between two other nodes via the shortest path within the network. Our findings showed that the betweenness centrality had less impact on the network optimization. It is not surprising because the betweenness of a node measures the extent to which two other nodes are connected through this specific node in a network.

4.3 Evaluation from experts and users

In the last stage of the evaluation, two field experts were invited to evaluate the optimization results generated by this study. In terms of the recommendation results assessed by the health professionals, one expert identified 115 relevant pairs while the second expert identified 113 relevant pairs. Subsequently, a Kappa test was conducted between the two evaluators' lists, revealing a "Measure of Agreement" value of 0.952 ($p < 0.001$), signifying an almost perfect agreement between the two experts. Following this, the evaluation lists were combined, and another Kappa test was employed to determine the consistency between the joint evaluation results from the experts and the corresponding recommendations derived from this study. The "Measure of Agreement" value was 0.951 ($p < 0.001$), also indicating an almost perfect agreement.

In addition to the Kappa test, a Chi-square test was utilized to provide supplementary validation of the agreement achieved between the study's results and the assessments of the two health professionals. The Pearson Chi-square value, degrees of freedom (df), and p-value were 0.049, 1, and 0.825, respectively. The test outcomes demonstrated no significant difference between the findings of this research and the evaluations from the medical professionals.

Similarity value		
The Mann-Whitney test	Original structural network	Optimized structural network
Mean rank	528.76	590.21
Sum of ranks	249575.00	387181.00
z-statistic	-3.125	
p-value	0.002	

Table 4. Statistical analysis result for H01

In addition, the optimization results were also evaluated by 31 students as a general user evaluation. In the user group, 4 users (13%) were recruited from the United States and 27 users (87%) were recruited from China. The Chinese users were recruited from a top 5 university in China, and all of them were ensured to be able to read English texts proficiently. Among them, 22 (71%) are female while nine (29%) are male. Moreover, 19 users (61%) were in the 19-23 age group, 11 users (35%) were in the 24-27 age group, and one user (3%) was in the 28 and over age group. In terms of their educational backgrounds, 16 (52%) were undergraduates while 15 (48%) were graduates.

The 31 users were divided into three sub-groups and each sub-group was required to evaluate one of the three sub-lists of paired health topics. The three sub-lists focused on different topics - the first user sub-group (12), the second user sub-group (12), and the third user sub-group (7) focused on health-related services, child health and development, and nutrition, respectively. Overall, about 89% of the recommended connections proposed by this study were confirmed by the users. The detailed confirmation rates for each sub-list are displayed in Table 5. In addition, within each sub-group, two evaluation lists were randomly selected to perform a Kappa test to examine the agreement level in each sub-group. The "Measure of Agreement" values are shown in Table 5 as well.

In conclusion, the analysis results show that the health topic suggestions proposed by this study were confirmed and agreed by both medical professionals and general users.

4.4 The comparison among the three similarity measures

For comparison purposes, the semantic network was also processed through UciNet using the Pearson correlation similarity measure and the Euclidean distance similarity measure. Interestingly enough, the overlap condition was uncovered between the recommendation results generated by the cosine and the Pearson correlation similarity measures. To be more specific, the cosine similarity measure explored six more bidirectional connections and one additional unidirectional connection than the Pearson correlation similarity measure. On the other side, the results concluded through the Euclidean distance measure only shared four common connections with the cosine similarity measure. The rest of the 115 connections uncovered by the Euclidean distance similarity measure were unique.

Moreover, Table 6 shows the five core health topics generated by the three similarity measures among their recommended connections:

According to this table, it is not surprising to find that the five key nodes in the network identified by the cosine and the Pearson correlation similarity measures were the same. However, the five core health topics identified through the Euclidean distance similarity measure were different as they were mostly concentrating on specific diseases, devices, and medical checking procedures. Moreover, the suggested connections generated through the Euclidean distance similarity measure were highly clustered among the five key health topics. The health topic Dwarfism was suggested to be linked to other 16 topics, including topics like

Confirmation rates & Kappa test results	Health related services related sub-list	Child health and development related sub-list	Nutrition related sub-list
Confirmation rate of recommended connections	87.85%, N=48	92.19%, N=48	86.29%, N=50
"Measure of Agreement" value of Kappa test	0.742 (substantial agreement), N=72	0.724 (substantial agreement), N=72	0.830 (almost perfect agreement), N=75

Table 5. Confirmation rates and Kappa test results of user evaluation results

The core health topic list	Cosine similarity measure	Pearson correlation similarity measure	Euclidean distance similarity measure
Core health topic 1	Medicines and Children	Medicines and Children	Dwarfism
Core health topic 2	Children's Health	Children's Health	Mobility Aids
Core health topic 3	Toddler Health	Toddler Health	Artificial Limbs
Core health topic 4	Child Development	Child Development	Barotrauma
Core health topic 5	Child Mental Health	Child Mental Health	Baby Health Checkup

Table 6. Five core health topics identified through the three similarity measures

Dentures and Cosmetic Dentistry. Given that the outcomes yielded from the Euclidean distance measure did not exhibit comparable efficacy to those produced by the cosine similarity measure and the Pearson correlation similarity measure, the proposed connections derived from the Euclidean distance measure were deemed unacceptable.

Since the recommendation results generated between the cosine similarity measure and the Pearson correlation similarity measure had great overlap, a series of Mann-Whitney tests were applied. The results showed that there was no significant difference in terms of similarity values generated between the cosine and the Pearson correlation similarity measures (p-value=0.429>0.05).

In conclusion, the Pearson correlation similarity measure and the cosine similarity measure may be jointly utilized for similarity analysis, thus optimizing additional subject directories of other online portals. When employing one of them for similarity calculation, the other could serve as the supplemental methodology for verification purposes. This could ensure a more convincing and effective optimization result. On the other side, the Euclidean distance similarity measure was less effective in optimizing health topic-based subject directories.

4.5 Novel strategies towards social networks

In prior studies, researchers attempted to create a semantic network that was built purely based on the semantic relationships possessed by the selected health topics (Zhu and Zhang 2020; Zhang et al. 2016). Several structurally connected health topics in the original structural network were removed in the semantic network due to low similarity values.

Compared with those prior studies, the revised semantic-based network built in this study kept the structurally connected health topics that were included in the original structural network even if they possessed low similarity values. The reason was that these connections were derived from official subject headings such as MeSH and had been manually reviewed by health professionals. In other words, the MedlinePlus portal established these connections for specific reasons. For instance, the health topics connected might be generated from other relationships such as an ontology-based system instead of semantic relationships. Therefore, keeping these original structural connections along with the newly recommended health topics in the revised semantic-based network could better reflect the overall subject directory system of MedlinePlus from a broader view – both semantic relationships and other types of relationships.

4.6 Implications for subject directory optimization

The theoretical implications of this study are manifested in the distinctive approach employed to identify connections possessing weak semantic relationships within a subject directory and enhance the directory by incorporating missing links that exhibit strong semantic associations. In this context, weak semantic relationships pertain to those semantic connections that display a similarity value falling below the prescribed threshold. Conversely, strong semantic relationships denote those semantic connections that display a similarity value surpassing the designated threshold. Compared with prior research (Zhu and Zhang 2020; Zhang et al. 2016; Zhang et al. 2015), the similarity measurements employed by this study for calculating semantic relationships are more diverse and creative.

The methodological implications lie in the mixed research method utilized in this study: social network analysis was applied with descriptive and inferential statistical analysis to examine missing connections in the children related subject directory and evaluate the optimization results; three different similarity measures were employed to calculate the semantic relationships possessed by the selected health topics; last, two levels of evaluations toward the optimization results were performed in both the medical expert group and the general user group. This combined research method could be applied to explore health topics of other health conditions/diseases or health consumer groups.

In light of the practical implications, the findings of this study can contribute to enhancing users' information searching effectiveness when they navigate the subject directory to browse relevant medical information. Meanwhile, the findings would also help MedlinePlus creators to optimize their directories. Compared with prior studies which focused on agriculture, general health, and mental health related topics (Zhu and Zhang 2020; Zhang et al. 2016; Zhang et al. 2015), the children related health topics investigated in this research study are novel.

4.7 Limitations and future research

One limitation of this study was that the optimization process conducted in this study had only focused on the system's side. All the recommendations were identified and presented according to the structural link networks built in the subject directory system on the MedlinePlus portal. No investigation was performed from the users' perspective. Another limitation of this study was that the optimization results were presented based only on part of the subject directory system on the MedlinePlus portal. Furthermore, the semantic analysis of this study relied upon the descriptive text procured from the webpages of health topics within subject directories. It should be noted that optimization efforts based on such textual information may not necessarily encompass or accurately represent all the significant associated health topics. Meanwhile, the term frequency (tf) is only one of the possible measures for semantic analysis. In

addition to the term frequency, the medical relevance of a term might also be considered. For instance, if a term appears in a medical thesaurus, it receives more weight. Also, the result evaluators, including both medical professionals and university students, were not real users of children related health information on MedlinePlus. Using real users of children related health information for evaluation could definitely yield more reliable findings. The last limitation of this study was that the methodologies applied for optimization in this study were limited. Only the similarity values and the three node-level centrality features were employed for proposing suggested connections. In addition, when calculating the similarity values of the selected health topics, only three similarity measures were utilized.

The future research directions include, but are not limited to, expanding the utilization of social network analysis techniques to encompass a broader array of subject directory systems within health portals or health information systems. Concurrently, the incorporation of additional aspects from social network analysis, particularly at the network level—such as the clustering coefficient—may prove beneficial in analogous research endeavors. More similarity measures can also be employed to set up various thresholds in terms of exploring new structural connections. As for the evaluation performed by the health field experts, additional qualitative methods can be applied with quantitative analysis. Moreover, besides evaluating and optimizing subject directories from the system's side, users could also be involved.

5.0 Conclusion

This research study concentrates on the investigation and optimization regarding the topic-based subject directory applied by the representative public health portal MedlinePlus. As a result, the current structural link network relating to children applied on the MedlinePlus portal was found to be not consistent with their semantic relationships among the involved health topics. Hundreds of pairs of health topics possessing similar Web page content were not able to be navigated to each other by the portal users. Furthermore, among these missed connections, a few health topics were found to have great impact within the whole network in terms of linking other topics through semantic connections. The control and responsibilities taken by these hidden core health topics were underestimated.

The recommended new structural connections were mostly added to the health consumer related health topics including children, teenagers, infants, toddlers, etc. Besides the health consumer groups, connections were also found in pregnancy, hepatitis, and daily exercises related topics. The average similarity value of the structural connections in the optimized network was improved after new connections were added. This difference was proved to be significant through the Mann-Whitney test. Key health topics were identified to have huge impact on the whole network. In addition, the node centrality measures were tested through the customized T-test, the results showed that the degree centrality and closeness centrality measures were significantly increased in the optimized structural network while the betweenness centrality measure was not significantly different from the original structural link network.

In the last stage, the optimization recommendations proposed by this study were evaluated by two health field experts and 31 academic users. Their evaluation outcomes confirmed that the suggested connections generated by this study fit into professional assessments and user evaluations. In other words, the recommendation results were supported by both the semantic relationships and human judgements.

References

Ahmed, Terry. 2019. "MedlinePlus at 21: A Website Devoted to Consumer Health Information." *Information Services & Use* 39, no. 1/2: 5-14. https://doi.org/10.3233/ISU-180038

Baird, Aaron, Michael F. Furukawa, and T. S. Raghu. 2012. "Understanding Contingencies Associated with the Early Adoption of Customer-Facing Web Portals." *Journal of Management Information Systems* 29, no. 2: 293–324. http://www.jstor.org/stable/41713891.

Barnes, John Arundel. 1954. "Class and Committees in a Norwegian Island Parish." *Human relations* 7, no. 1: 39-58. https://doi.org/10.1177/001872675400700102

Bernhardt, Jay M., and Elizabeth M. Felter. 2004. "Online Pediatric Information Seeking among Mothers of Young Children: Results from a Qualitative Study using Focus Groups." *Journal of medical Internet research* 6, no.1: e36. https://doi.org/10.2196/jmir.6.1.e7

Borgatti, Stephen P., Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. 2009. "Network Analysis in the Social Sciences." *Science* 323, no. 5916: 892-95. https://doi.org/10.1126/science.1165821

Borgatti, Stephen P., Martin G. Everett, and Jeffrey C. Johnson. 2013. *Analyzing Social Networks*. Los Angeles, CA: SAGE Publications.

Bouche, Gauthier and Virginie Migeot. 2008. "Parental Use of the Internet to Seek Health Information and Primary Care Utilisation for their Child: A Cross-Sectional Study." *BMC Public Health* 8: 1-9. https://doi.org/10.1186/1471-2458-8-300

Bray, Lucy, Victoria Appleton, and Ashley Sharpe. 2019. "'If I Knew What Was Going to Happen, it Wouldn't Worry Me So Much': Children's, Parents' and Health Professionals' Perspectives on Information for Children Undergoing a Procedure." *Journal of Child Health*

Care 23, no. 4: 626-38. https://doi.org 10.1177/1367493519870654

Chambers, Duncan, Paul Wilson, Carl Thompson, and Melissa Harden. 2012. "Social Network Analysis in Healthcare Settings: A Systematic Scoping Review." *PLoS ONE* 7, no. 8: e41911. https://doi.org/10.1371/journal.pone.0041911

Chung, Young Mee, and Young-Hee Noh. 2003. "Developing a Specialized Directory System by Automatically Classifying Web Documents." *Journal of Information Science* 29, no. 2: 117-26. https://doi.org/10.1177/016555150302900204

Coleman, James S. 1986. "Social Theory, Social Research, and a Theory of Action." *American Journal of Sociology* 91, no. 6: 1309-35. https://www.jstor.org/stable/2779798

Dalton, Louise, Elizabeth Rapa, and Alan Stein. 2020. "Protecting the psychological health of children through effective communication about COVID-19." *The Lancet Child & Adolescent Health* 4, no.5: 346-47. https://doi.org/10.1016/S2352-4642(20)30097-3

Ellis-Danquah, La Ventra. 2004. "Addressing Health Disparities: African American Consumer Information Resources on the Web." *Medical Reference Services Quarterly* 23, no. 4: 61-73. https://doi.org/10.1300/J115v23n04_06

Ellis, David, and Ana Vasconcelos. 2000. "The Relevance of Facet Analysis for World Wide Web Subject Organization and Searching." *Journal of Internet Cataloging* 2, nos. 3/4: 96-114. https://doi.org/10.1300/J141v02n03_07

Fallis, Don, and Martin Frické. 2002. "Indicators of Accuracy of Consumer Health Information on the Internet: A Study of Indicators Relating to Information for Managing Fever in Children in the Home." *Journal of the American Medical Informatics Association* 9, no. 1:73-79. https://doi.org/10.1136/jamia.2002.0090073

Freeman, Linton C. 1978. "Centrality in Social Networks Conceptual Clarification." *Social Networks* 1, no. 3: 215–39. https://doi.org/10.1016/0378-8733(78)90021-7

Gray, Caryl. 2005. "Health and Medical Resources: Information for the Consumer." *Journal of Library Administration* 44, nos. 1/2: 395-428. https://doi.org/10.1300/J111v44n01_05

Gray, Nicola J., Jonathan D. Klein, Peter R. Noyce, Tracy S. Sesselberg, and Judith A. Cantrill. 2005. "Health Information-Seeking Behaviour in Adolescence: The Place of the Internet." *Social Science & Medicine* 60, no. 7: 1467-78. https://doi.org/10.1016/j.socscimed.2004.08.010

Horrigan, John B., Lee Rainie, and Michael Cornfield. 2003. "Part Three: The Portals." *Pew Research Center: Internet, Science & Tech. Pew Research Center*, December 31, 2019. https://www.pewresearch.org/internet/2003/03/20/part-three-the-portals/.

Impicciatore, Piero, Chiara Pandolfini, Nicola Casella, and Maurizio Bonati. 1997. "Reliability of Health Information for the Public on the World Wide Web: Systematic Survey of Advice on Managing Fever in Children at Home." *BMJ: British Medical Journal* 314, no. 7098: 1875-79. https://doi.org/10.1136/bmj.314.7098.1875

Izenberg, Neil, and Debra A. Lieberman. 1998. "The Web, Communication Trends, and Children's Health Part 4: How Children Use the Web." *Clinical Pediatrics* 37, no.6: 335-40. https://doi.org/10.1177/000992289803700601

Khoo, Kaylyn, Penny Bolt, Franz E. Babl, Susan Jury, and Ran D. Goldman. 2008. "Health Information Seeking by Parents in the Internet Age." *Journal of Pediatrics and Child Health* 44, nos. 7/8: 419-23. https://doi.org/10.1111/j.1440-1754.2008.01322.x

Kim, Eunkyung, Jiran Hou, Jeong Yeob Han, and Itai Himelboim. 2016. "Predicting Retweeting Behavior on Breast Cancer Social Networks: Network and Content Characteristics." *Journal of Health Communication* 21, no. 4: 479-86. https://doi.org/10.1080/10810730.2015.1103326

Kuperminc, Michelle N., and Richard D. Stevenson. 2008. "Growth and Nutrition Disorders in Children with Cerebral Palsy." *Developmental disabilities research reviews* 14, no. 2: 137-46. https://doi.org/10.1002/ddrr.14

Li, Yuelin, Xiaojun Yuan, and Ruoqi Che. 2021. "An Investigation of Task Characteristics and Users' Evaluation of Interaction Design in Different Online Health Information Systems." *Information Processing & Management* 58, no. 3:102476. https://doi.org/10.1016/j.ipm.2020.102476

Liu, Jia Jia, Yanping Bao, Xiaolin Huang, Jie Shi, and Lin Lu. 2020. "Mental Health Considerations for Children Quarantined Because of COVID-19." *The Lancet Child & Adolescent Health* 4, no.5: 347-49. https://doi.org/10.1016/S2352-4642(20)30096-1

Luke, Douglas A., and Jenine K. Harris. 2007. "Network Analysis in Public Health: History, Methods, and Applications." *Annual Review of Public Health* 28, no. 1: 69–93. https://doi.org/10.1146/annurev.publhealth.28.021406.144132

Ma, Dan, Meiyun Zuo, and Liu Liu. 2021. "The Information Needs of Chinese Family Members of Cancer Patients in the Online Health Community: What and Why?" *Information Processing & Management* 58, no. 3: 102517. https://doi.org/10.1016/j.ipm.2021.102517

McClung, H. Juhling, Robert D. Murray, and Leo A. Heitlinger. 1998. "The Internet as a Source for Current Patient Information." *Pediatrics* 101, no. 6: e2. https://doi.org/10.1542/peds.101.6.e2

Meppelink, Corine S., Edith G. Smit, Marieke L. Fransen, and Nicola Diviani. 2019. "I was Right About Vaccination": Confirmation Bias and Health Literacy in Online Health Information Seeking." *Journal of Health Communication* 24, no.2: 129-40. https://doi.org/10.1080/10810730.2019.1583701

Mertens, Frédéric, Johanne Saint-Charles, and Donna Mergler. 2012. "Social Communication Network Analysis of the Role of Participatory Research in the Adoption of New Fish Consumption Behaviors." *Social Science & Medicine* 75, no.4: 643-50. https://doi.org/10.1016/j.socscimed.2011.10.016

Miller, Naomi, Eve-Marie Lacroix, and Joyce EB Backus. 2000. "MEDLINEplus: Building and Maintaining the National Library of Medicine's Consumer Health Web Service." *Bulletin of the Medical Library Association* 88, no. 1: 11-17. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC35193/

Möller, Judith, Robbert Nicolai van de Velde, Lisa Merten, and Cornelius Puschmann. 2020. "Explaining Online News Engagement based on Browsing Behavior: Creatures of Habit?" *Social Science Computer Review* 38, no. 5: 616-32. https://doi.org/10.1177/0894439319828012

Naidu, Rahul, June Nunn, and Jennifer D. Irwin. 2015. "The Effect of Motivational Interviewing on Oral Healthcare Knowledge, Attitudes and Behaviour of Parents and Caregivers of Preschool Children: An Exploratory Cluster Randomised Controlled Study." *BMC Oral Health* 15: 1-15. https://doi.org/10.1186/s12903-015-0068-9

Naskar, Debashis, and Subhashis Das. 2019. "HNS Ontology Using Faceted Approach." *Knowledge Organization* 46, no.3: 187-98. https://doi.org/10.5771/0943-7444-2019-3-187

Nie, Liqiang, Yiliang Zhao, Mohammad Akbari, Jialie Shen, and Tat-Seng Chua. 2014. "Bridging the Vocabulary Gap between Health Seekers and Healthcare Knowledge." *IEEE Transactions on Knowledge and Data Engineering* 27, no. 2: 396-409. https://doi.org/10.1109/TKDE.2014.2330813.

Oh, Kyong Eun, Soohyung Joo, and Eun-Ja Jeong. 2015. "Online Consumer Health Information Organization: Users' Perspectives on Faceted Navigation." *Knowledge Organization* 42, no.3: 176-86. https://doi.org/10.5771/0943-7444-2015-3-176

Otte, Evelien, and Ronald Rousseau. 2002. "Social Network Analysis: A Powerful Strategy, Also for the Information Sciences." *Journal of information Science* 28, no. 6: 441-53. https://doi.org/10.1177/016555150202800601

Pandolfini, Chiara, Piero Impicciatore, and Maurizio Bonati. 2000. "Parents on the Web: Risks for Quality Management of Cough in Children." *Pediatrics* 105, no. 1: e1. https://doi.org/10.1542/peds.105.1.e1

Pang, Patrick Cheong-Iao, Shanton Chang, Karin Verspoor, and Jon Pearce. 2016. "Designing Health Websites Based on Users' Web-Based Information-seeking Behaviors: A mixed-Method Observational Study." *Journal of Medical Internet Research* 18, no. 6: e145. https://doi:10.2196/jmir.5661

Postigo, Hector. 2011. "Questioning the Web 2.0 Discourse: Social Roles, Production, Values, and the Case of the Human Rights Portal." *The Information Society* 27, no. 3: 181-93. https://doi:10.1080/01972243.2011.566759.

Pow, Janette, Kaberi Gayen, Lawrie Elliott, and Robert Raeside. 2012. "Understanding Complex Interactions Using Social Network Analysis." *Journal of Clinical Nursing* 21, no. 19-20: 2772-79. https://doi.org/10.1111/j.1365-2702.2011.04036.x

Qiao, Wanxin, Zhijun Yan, and Xiaohan Wang. 2021. "Join or not: The Impact of Physicians' Group Joining Behavior on their Online Demand and Reputation in Online Health Communities." *Information Processing & Management* 58, no. 5: 102634. https://doi.org/10.1016/j.ipm.2021.102634

Sage, Adam, Delesha Carpenter, Robyn Sayner, Kathleen Thomas, Larry Mann, Sandy Sulzer, Adrian Sandler, and Betsy Sleath. 2018. "Online Information-Seeking Behaviors of Parents of Children with ADHD." *Clinical Pediatrics* 57, no.1: 52-56. https://doi.org/10.1177/0009922817691821

Schilling, Katherine, and Anna M. McDaniel. 2010. "Development and Evaluation of a Cancer Information Web Portal: The Impact of Design and Presentation on User Engagement." *Journal of Consumer Health on the Internet* 14, no. 3: 242-62. https://doi.org/10.1080/15398285.2010.501736

Segaran, Toby. 2007. *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. Sebastopol, CA: O'Reilly.

Silventoinen, Karri, José Maia, Aline Jelenkovic, Sara Pereira, Élvio Gouveia, António Antunes, Martine Thomis, Johan Lefevre, Jaakko Kaprio, and Duarte Freitas. 2021. "Genetics of Somatotype and Physical Fitness in Children and Adolescents." *American Journal of Human Biology* 33, no. 3: e23470. https://doi.org/10.1002/ajhb.23470.

Smith, Kirsten P., and Nicholas A. Christakis. 2008. "Social Networks and Health." *Annual Review of Sociology* 34: 405-29. https://doi.org/10.1146/annurev.soc.34.040507.134601

Tandi Lwoga, Edda, and Neema Florence Mosha. 2013. "Information Seeking Behaviour of Parents and Caregivers of Children with Mental Illness in Tanzania." *Library*

Review 62, no. 8/9: 567-84. https://doi.org/10.1108/LR-10-2012-0116

Yang, Han, Zhijun Yan, Lin Jia, and Huigang Liang. 2021. "The Impact of Team Diversity on Physician Teams' Performance in Online Health Communities." *Information Processing & Management* 58, no. 1: 102421. https://doi.org/10.1016/j.ipm.2020.102421

Yang, Hsin-Chang, and Chung-Hong Lee. 2004. "A Text Mining Approach on Automatic Generation of Web Directories and Hierarchies." *Expert Systems with Applications* 27, no.4: 645-63. https://doi.org/10.1016/j.eswa.2004.06.009

Yeo, Myung-Ho, Yoon-Kyeong Lee, Kyu-Jong Roh, Hyeong-Soon Park, Hak-Sin Kim, Jun-Ho Park, Tae-Ho Kang, Hak-Yong, and Jae-Soo Yoo. 2010. "Design and Implementation of a Directory System for Disease Services." *International Journal of Contents* 6, no. 1: 59–64. https://doi.org/10.5392/IJoC.2010.6.1.059

Zhang, Jin, and Dietmar Wolfram. 2009. "Visual Analysis of Obesity-related Query Terms on HealthLink." *Online Information Review* 33, no.1: 43-57. https://doi.org/10.1108/14684520910944382

Zhang, Jin, Shanshan Zhai, Hongxia Liu, and Jennifer Ann Stevenson. 2015. "Social Network Analysis on a Topic-based Navigation Guidance System in a Public Health Portal." *Journal of the Association for Information Science and Technology* 67, no.5: 1068-88. https://doi.org/10.1002/asi.23468

Zhang, Jin, Shanshan Zhai, Jennifer Ann Stevenson, and Lixin Xia. 2016. "Optimization of the Subject Directory in a Government Agriculture Department Web Portal." *Journal of the Association for Information Science and Technology* 67, no. 9: 2166-80. https://doi.org/10.1002/asi.23550

Zhang, Jin, and Yiming Zhao. 2013. "A User Term Visualization Analysis based on a Social Question and Answer Log." *Information Processing & Management* 49, no. 5: 1019-48. https://doi.org/ 10.1016/j.ipm.2013.04.003

Zhao, Yuehua, Kejun Chen, Jiaer Peng, Jiaqing Wang, and Ningyuan Song. 2022. "Diverse Needs and Cooperative Deeds: Comprehending Users' Identities in Online Health Communities." *Information Processing & Management* 59, no. 5: 103060. https://doi.org/10.1016/j.ipm.2022.103060

Zhu, Yifan, and Jin Zhang. 2020. "Social Network Analysis of the Mental Health Sub-Topic on the Medlineplus Subject Directory." *Information Research: an International Electronic Journal* 25, no. 4: paper 876. https://doi.org/10.47989/irpaper876