An Investigation Study on the Mental Disorder Related Topics in the Subject Directory of MedlinePlus Portal

Yifan Zhu[1][0000-0002-3622-2858] and Jin Zhang[1][0000-0002-6665-6606]

[1] University of Wisconsin Milwaukee, Milwaukee, WI 53201, USA
yifanzhu@uwm.edu, jzhang@uwm.edu

Abstract. We examined a subject directory system related to Mental Disorder on the MedlinePlus portal. According to the comparison between the link connection network and the semantic connection network among the 99 collected health topics, 55 bi-directional as well as 23 unidirectional connections were identified and proposed to be added to the corresponding health topic pages. In addition, Mental Disorder related topics were found to be linked to Youth & Child related topics and Daily Health related topics in the subject directory. A mixed research method combining social network analysis and inferential analysis was applied. The recommended connections were evaluated by domain experts and visualized from various perspectives. Suggestions for optimizing and enhancing the current link network among Mental Disorder and related groups of health topics were provided. The findings in this study offered insights to public health portal creators for designing subject directory-based navigation system. Future research directions in other fields were also discussed.

Keywords: Mental Disorder, Mental Health, Subject Directory, MedlinePlus, Social Network Analysis.

1 Introduction

Currently, the trend of searching health information online is considered to be indicated by the rapidly growing amount of online public health portals [1, 2]. Among those comprehensive sites that are normally regarded as online health portals, MedlinePlus was launched in 1998 and was the first primary initiative for providing online health information to the public from the National Library of Medicine [3]. Meanwhile, MedlinePlus applies a group of information retrieval tools that consist of search bar (engine), subject directory, sitemap, etc., to assist users in their information searching and site navigation. Among these tools, the health topic based subject directory system enables users to browse information from general level to specific level through its hierarchical structure. In addition, related subjects (health topics) are also listed in Web pages for individual topics so that users can "jump" to relevant sections when necessary.

Among hundreds of health conditions and diseases, mental disorders are generally characterized by the World Health Organization as "some combination of abnormal thoughts, emotions, behavior and relationships with others" [4]. Common mental diseases such as Depression (300 million), Bipolar affective disorder (60 million), and Schizophrenia (23 million), etc., have had a wide range of consumers globally. WHO also emphasizes the importance of proving access to health care as well as social services that are able to offer treatment and support. Therefore, investigating mental health related information from MedlinePlus portal is necessary and will be supportive for related health consumer groups.

The research problem for this study is to examine if the current subject directories applied by MedlinePlus regarding mental disorder related health topics are consistent with the semantic relationships shared among those health topics. If not, optimizations could be identified and employed in order to improve the navigation system of public health portals like MedlinePlus.

2 Method

MedlinePlus website used five broad sections to divide its over 1000 health related issues in its subject directory system. Each of these five sections included a list of categories, and each of the category had an introduction page that contained a group of related subcategories in an alphabetical order. Every subcategory referred to a specific health-related issue and had an individual Web page. The Web page had a "related health topics" column located at the right side to help end-users to be navigated to other relevant health topics.

The connections among these health topics were extracted and generated from two levels – structural and semantic. The structural connections refer to the fact that one topic is listed under the "related health topics" section of another topic's Web page. The structural connections can form a structural network. The structural network was set up by the website creators and enabled end-users to navigate the website. The semantic connections refer to the semantic strength between two topics' Web page content. The semantic connections form a semantic network. In other words, the semantic network reflects how closely two health topics are related based on the similarity of their textual information.

The selected starting health topic was *Mental Disorders*. This topic and its related health topics were collected to form the first group of health topics as Level 1. After that, related health topics were generated from each of the health topic at Level 1 to form the second group as Level 2. Such data collection process repeated until the fourth group at Level 4 was reached. After the process, 99 health topics were included in total. Meanwhile, Web pages of all the involved four levels of health topics were gathered and the text on these web pages was extracted to form a word list.

This word list was cleansed. First, a stop-word list was applied to filter the word list to remove useless words. These stop-words mainly included those which only function from the grammar aspect, such as "a", "an", "the", "with", "of", "to" etc. Second, synonyms were combined; for instance, anorexia nervosa, binge eating, and bulimia were combined into eating disorders. Third, all the words on the list were kept as their regular form. Different forms were normalized.

After filtering the original dataset, a topic-topic link matrix was built to represent the structural link network (Equation (1)).

$$\text{Topic} - \text{topic link matrix} = \begin{pmatrix} t_{11} & \cdots & & t_{1n} \\ & \cdots & t_{ij} & \\ & \cdots & & \\ t_{n1} & \cdots & & t_{nn} \end{pmatrix} \quad (1)$$

Then, a topic-semantic matrix (TSM) (Equation (2)) was built to represent the semantic network among the selected health topics according to the results calculated through similarity measure based on the term frequency data. In terms of the similarity measure, the cosine-similarity measure was used in this study is shown in Equation (3).

$$\text{Topic semantic matrix} = \begin{pmatrix} h_{11} & \cdots & & h_{1m} \\ & \cdots & h_{ij} & \\ & \cdots & & \\ h_{n1} & \cdots & & h_{nm} \end{pmatrix} \quad (2)$$

$$s_{ij} = \frac{\sum_{k=1}^{n} h_{ik} \times h_{jk}}{\left(\sum_{k=1}^{n} h_{ik}^2 \times \sum_{k=1}^{n} h_{jk}^2\right)^{\frac{1}{2}}} \quad (3)$$

After collecting the value of similarities among all the health topics, the average similarity among those topics that have had structural link connections set by MedlinePlus website was calculated and set as the threshold. Recommended connections for optimization were identified based on the comparison between the average similarity value among the rest of health topics that have not had structural link connections and the threshold number. For those topics that were sharing a similarity value that was larger than the threshold but had no structural link connections built by the portal, it indicated that the semantic relationship was not able to match the structural relationship.

3 Result

260 ties were found in the link network and 9700 ties were found in the semantic network. A visualized figure presenting the structural link network is displayed in Figure 1. The 99 health topics could be briefly classified into three groups – Specific Disease (Mental Disorder and related topics), Consumer Group (Teenager and Child involved topics), and Daily Health (Nutrition related topics). Health topics of same group were found to gather together more closely in the visualized network. The averaged cosine similarity value for these 260 structural links was 0.383677.

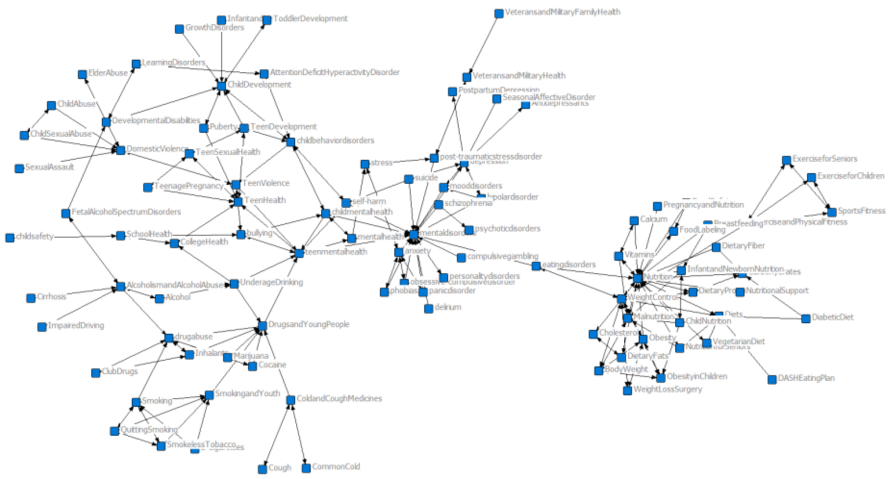

Fig. 1. The display of structural link social network.

Among the 9442 ties that did not have a structural link set by the portal, 133 pairs of categories were found to have a similarity value larger than the average number among the connected topics. To be more specific, there were 110 pairs of topics that were recommended to set up bidirectional connections and 23 pairs of topics that were recommended to set up unidirectional connections (Fig.2 and Fig. 3).

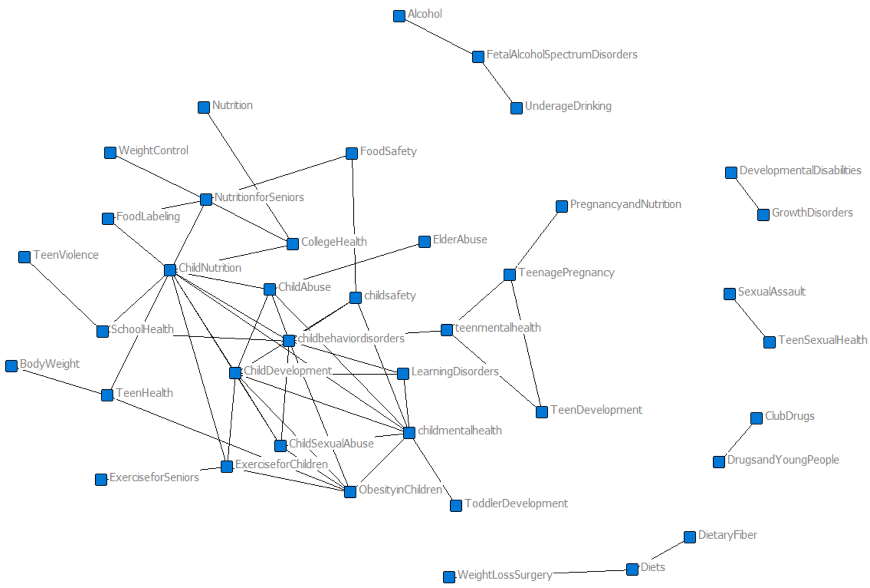

Fig. 2. The display of bidirectional link connections suggested for health topics.

Among these suggested bidirectional edges displayed on the table and the visualized network, some of them were quite straightforward to understand. For instance, *Child Development* and *Child Mental Health* shared a 0.735 similarity value in their textual content, which indicated that these two health topics had a lot of information in common from the semantic perspective. Similar cases also included *Child Sexual Abuse* and *Child Mental Health*, *Teen Development* and *Teen Mental Health*, as well as *Child Nutrition* and *Child Development*, etc. However, a few edges suggested based on high similarity values might not be that clear compared with the previous examples – an instance would be *Exercise for Seniors* and *Exercise for Children* – these two topics had a similarity value of 0.669, which might be caused by their common sections of text introducing about "exercise," while they were actually targeting on different health consumer groups. However, there could be another circumstance worth noticing when considering the fact that users checking on information from MedlinePlus might be the family member that is responsible to take care of the whole family (e.g. a mother). If that is the case, such structural linkages set for the same daily health topic among various health consumer groups might be of great help. Therefore, such bidirectional linkages were still recommended.

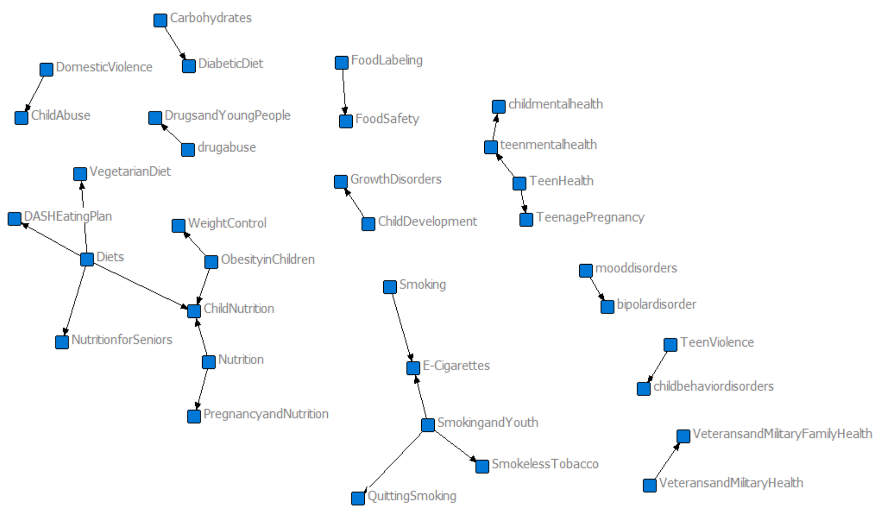

Fig. 3. The display of unidirectional link connections suggested for health topics

From the visualized network, the 23 pairs of unidirectional edges could be divided into two types: one type indicated a hierarchy relationship that was from general to specific or vice versa, such as *Diets* and *Vegetarian Diet*, *Nutrition* and *Child Nutrition*, etc.; the other type referred to a relevance relationship, for example, *Child Mental Health* and *Teen Mental Health*, *Food Labeling* and *Food Safety*, etc. Both types of the unidirectional edges were recommended to be adjusted to bidirectional edges. The reason was that, in a hierarchy relationship, when health consumers browsed from a broad topic to a narrow topic, if the connection between the two topics was unidirectional, the health consumers could not return to its parent topic. The situation would get worse if the narrow topic had no related topics. For instance, when health consumers browsed from *Veteran and Military Health* to *Veteran and Military Family Health*, *Veteran and Military Family Health* had no related health topics listed in its page. As a result, the health consumers had no more topics to follow. Through the creation of an edge pointing back to its parent health topics, the health consumers could go back to any level and find broad topics they previously had browsed and then start another navigation throughout the network. Similarly, for unidirectional edges indicating a relevance relationship, adding the arrow on the other end would also benefit health consumers for navigating purposes.

Moreover, among the 23 pairs of edges suggested and displayed in the figure, most of them simply involved two health topics. However, there were also two small networks containing multiple topics – one of them was about *Diets* and *Nutrition* related topics, while the other was about *Smoking*. Interestingly, these two small networks seemed to form a star structure. Central topics like *Diets* and *Smoking and Youth* were having outbound topics while central topics such as *Child Nutrition* and *E-cigarettes* were having inbound topics.

The suggested 55 pairs of bi-directional and 23 pairs of unidirectional connections were combined with other 66 pairs of health topics that were suggested not to build connection to each other and assessed by two evaluators. Evaluators both hold Master or higher degrees in Preventive Medicine. A kappa test was applied between the combined evaluation list from the two evaluators and the corresponding recommended results proposed by this study and the value was 0.819 ($p < 0.001$), which achieved a nearly perfect agreement. Next, a chi-square test was employed to com- pare the combined evaluation list from the two evaluators and the corresponding recommended results proposed by this study. The Pearson Chi-square value, df, and p- value were 0.125, 1, and 0.723, respectively. The recommendations stated in this study were consistent with those from the expert evaluators.

4 Conclusion

In this poster, we have reported some preliminary findings about the subject directory system in the portal of MedlinePlus regarding mental disorder related health topics. We found that the structural and semantic linkages among collected health topics were not consistent. Plenty of pairs of health topics that were found to share high semantic connections were not linked in the structural network by the portal creators, and that might lead health consumers not able to find enough helpful health information. Our investigation results were consistent with the evaluation results of medical experts. In addition, we identified two health topic groups that were having close relationship to mental disorder related topics: one of them was teenager and child group, which referred to a specific health consumer group related to mental disorder; and the other was daily health group, which reflected the importance of some daily health factors, such as nutrition, to mental disorder.

This study has investigated the subject directories applied by the MedlinePlus portal regarding mental disorder related health topics. Several optimizations have been proposed to improve the current subject directory system. Moreover, for other public health portals that are using subject directories, this study might provide some insights to the portal creators. Finally, the mixed research methods applied in this study might also be utilized when facing with similar studies focusing on various health topics or public health portals.

References

1. Benigeri, M.: Shortcomings of health information on the Internet. Health Promot. Int. 18, 381–386 (2003). doi:10.1093/heapro/dag409
2. Evers, K.E.: eHealth Promotion: The Use of the Internet for Health Promotion. Am. J. Heal. Promot. 20, 1–14 (2006). doi:10.4278/0890-1171-20.4.1
3. Miller, N., Lacroix, E.M., Backus, J.E.: MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. Bull. Med. Libr. Assoc. 88, 11–7 (2000).
4. WHO Mental Disorders Fact Sheet. https://www.who.int/en/news-room/fact-sheets/detail/mental-disorders