April 2018

Visualising the Topological Structure of Health-related Message Board User Networks

Maryam Abdollahyan [a,1], Raúl J. Mondragón [a], Conrad Bessant [b] and Fabrizio Smeraldi [a]

[a] *School of Electronic Engineering and Computer Science, Queen Mary University of London, United kingdom*

[b] *School of Biological and Chemical Sciences, Queen Mary University of London, United Kingdom*

Abstract. Health-related message boards are common digital health platforms with great practical and sociological relevance. In this work, we present an investigation of the topological structure of interactions in such forums. We visualise the interactions among users and between users and forum administrators using a simple yet effective network visualisation tool called the Bitmap of Sorted Adjacency Matrix (BOSAM) algorithm. We apply the BOSAM algorithm to six health-related message boards of various scopes and sizes. The outcome reveals major differences between the user interaction networks of these forums. The results of our analysis correlate closely with the characteristics of the respective message boards, including their topic coverage, presence of user communities, commercial or community nature and management style.

Keywords. network models, digital health, visualisation, BOSAM

1. Introduction

Health-related message boards are an important type of digital health platform. These forums, over the years, have evolved from mainly informal support groups to more systematic efforts. These include commercially run forums that support official patient associations, facilitate patient access to medical professionals or aim at gathering information for the benefit of pharmacological and medical research. This has attracted the attention of sociologists, policymakers and computational scientists, with various studies aimed at modelling the evolution of forums [5], characterising the participation styles of patients [2] or optimising patient information extraction [8].

The analysis of the structure of health-related message boards has implications for both forum managers and users. Measuring the engagement and responsiveness levels of user interactions in a forum allows its managers to assess the capacity of their forum and subsequently adjust the level of support in the forum. In addition, it helps patients decide

[1] Corresponding Author: Maryam Abdollahyan, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom; E-mail: m.abdollahyan@qmul.ac.uk

which forums to join. Understanding how users contribute to a forum assists its owners in developing strategies that would promote growth within the forum and facilitate access to the information on the forum [4,10].

In this work, we take a wider-angle view of health-related message boards as social networks, focusing on the topology of interactions among the users and between the users and forum administrators. We use a network visualisation tool called the Bitmap of Sorted Adjacency Matrix (BOSAM) algorithm [6] to represent the structure of these interactions. Many studies require a detailed analysis of user posting behaviour, possibly the semantic annotation of their posts, the information provided by users in their public profiles (e.g. age and gender) or metrics that are only accessible by forum administrators (e.g. total time users are logged in) [1,7]. In contrast, we employ a simple metric that is easily obtainable for any forum. Namely, we define the interactions between users as posting to the same conversation thread. Results of applying our approach to six different health-related message boards indicate that the resulting bitmaps reflect the nature of the forums, in terms of the topics being discussed, participation of their users in communities and their administration style.

2. Overview of Method

The connectivity of a network is characterised by its adjacency matrix. Given an undirected network containing N nodes with indices $1, 2, ..., N$, its adjacency matrix A is defined as an $N \times N$ matrix where a_{ij} is 1 if there is a link connecting nodes i and j, and 0 otherwise. Matrix A can be represented as black and white bitmap image where the pixel at a coordinate (i, j) is black if $a_{ij} = 1$ and white otherwise. This binary representation gives an overall impression of the network connectivity, however, in general it looks like a collection of random points when the indices of nodes are arbitrary. For this representation to reflect the topological properties of the network, the nodes can be sorted so that their indices correspond to a measure of their connectivity [3,9]. One such representation is the Bitmap of Sorted Adjacency Matrix (BOSAM) [6]. The BOSAM of a network is obtained as follows.

Specifically, the nodes are reindexed by sorting them in the ascending order of their degree (i.e. the number of links they have). If two nodes have the same degree, then they are ranked in the ascending order of the largest degree of their neighbours. In the standard algorithm, remaining ties are broken by reordering the nodes in the ascending order of the largest index of their neighbours. Here, we use a slightly different approach: in the case of ties, we reorder the nodes in the ascending order of the largest degree of their second neighbours and so on. The bitmap visualisation of the reordered adjacency matrix is the BOSAM of the network.

The BOSAM often resembles a "flower". The shape of the components of this flower is related to the topological properties of the network. In Section 4, we discuss some of these properties in the context of message board user interaction networks. If a network preserves its macroscopic structure while it grows, then the shape of the components of its BOSAM will scale with respect to the network's size. Hence, looking at the BOSAMs for two networks (which can be of different sizes), one can tell if they have the same macroscopic structure.

3. Datasets

We collected data from six health-related message boards (see Table 1). Among these, three (CROHN, HDA and PSOR) are focused on a specific disease, while the others (HEALTH, INSPIRE, PINFO) cover various health topics. Moreover, the CROHN, HDA and PSOR message boards appear to be community efforts, while the others are managed in a more "commercial" manner. To model the user interaction network for each message board, we created a node for each user and linked two nodes if their corresponding users had posted to the same conversation thread.

Table 1. Names of the message boards (networks) and number of users (nodes) and posts we processed

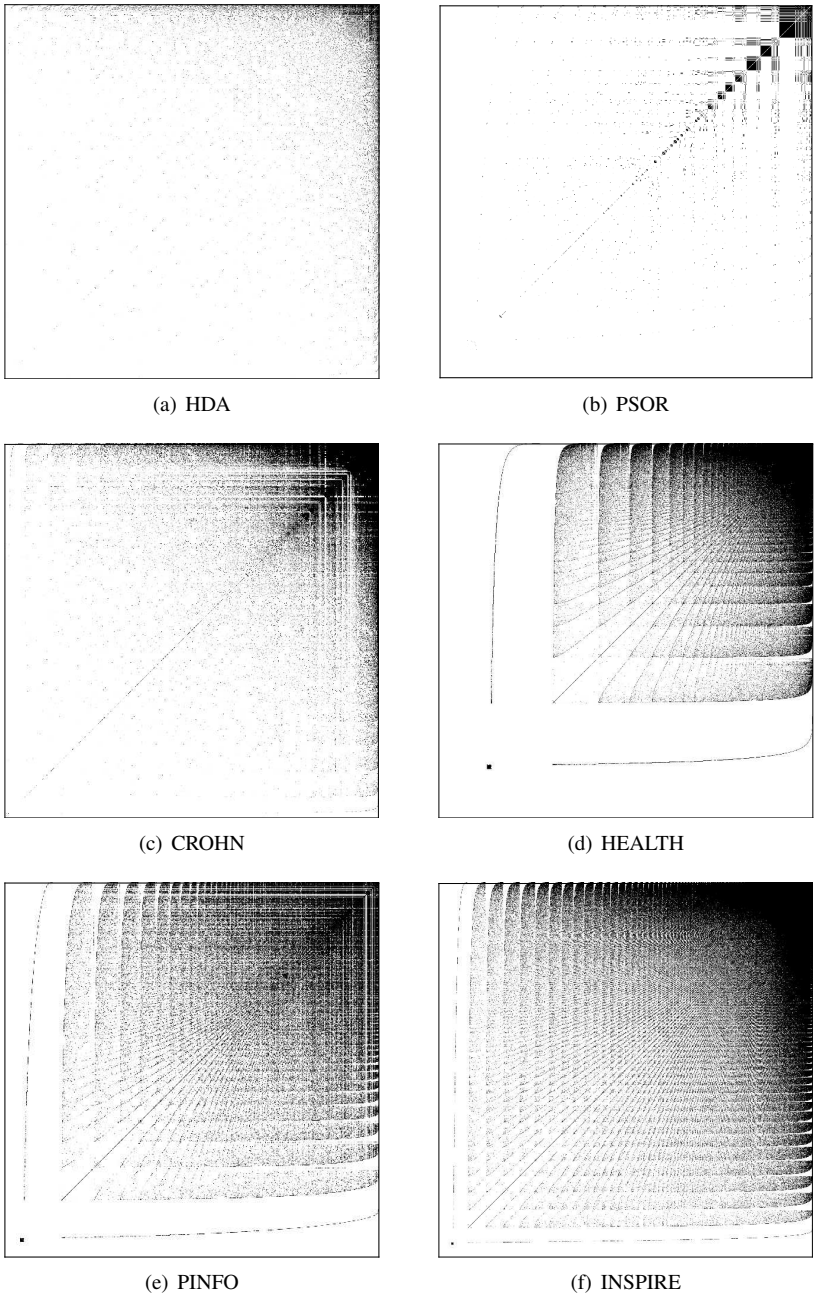
Message Board[1]	Number of Users	Number of Posts
Crohn's Disease Forum (CROHN)	21703	673162
Health Boards (HEALTH)	403682	4929836
Huntington's Disease Association (HDA)	1245	34173
Inspire communities (INSPIRE)	221432	3725555
Patient Info (PINFO)	139567	1699131
Psoriasis Association (PSOR)	1816	4402

4. Results and Discussion

Figure 1 shows the BOSAMs for the networks we modelled. In the HDA network, connectivity is dominated by the nodes with higher degrees as pixels are distributed densely along the upper right corner of its BOSAM, meaning that users of low degree tend to interact with users of high degree and vice versa. However, no other obvious structure is observed. Hence, the HDA message board appears as a single-topic forum that patients use to ask questions from expert users. In contrast, in the PSOR network, users tend to interact with other users of similar degree. This is indicated by the 'squares' near the diagonal of its BOSAM which show the formation of communities of nodes with similar degree, with noticeably larger communities consisting of nodes with high degree that are tightly linked together. Furthermore, these node communities observed in the BOSAM correspond to the patient communities on the PSOR forum, where patients in each community have been diagnosed with a different type of psoriasis. The CROHN network, also a single-topic forum being run by the community, exhibits a mixture of the above two characteristics.

The HEALTH, PINFO and INSPIRE message boards, being large commercially run forums, have much richer structures. Their BOSAMs resemble a flower where each component 'leaf' has a statistical meaning. At a closer look, each leaf has a step-like shape (see Figure 2). The length of the leaf's base is equal to the number of nodes with degree k given by $N\sum_{i=k}^{k+1} P(i)$, where $P(k)$ is the probability of finding a node with degree k in the network. The height of the steps is equal to the number of nodes with degree k', again given by $N\sum_{i=k'}^{k'+1} P(i)$. The tread of the k'-th step is equal to the number of

[1]CROHN: www.crohnsforum.com; HEALTH: www.healthboards.com; HDA: www.hdmessageboard.com; INSPIRE: www.inspire.com; PI: patient.info; PSOR: www.psoriasis-association.org.uk



(a) HDA (b) PSOR

(c) CROHN (d) HEALTH

(e) PINFO (f) INSPIRE

Figure 1. BOSAMs for the modelled networks. Node degree increases along the axes.

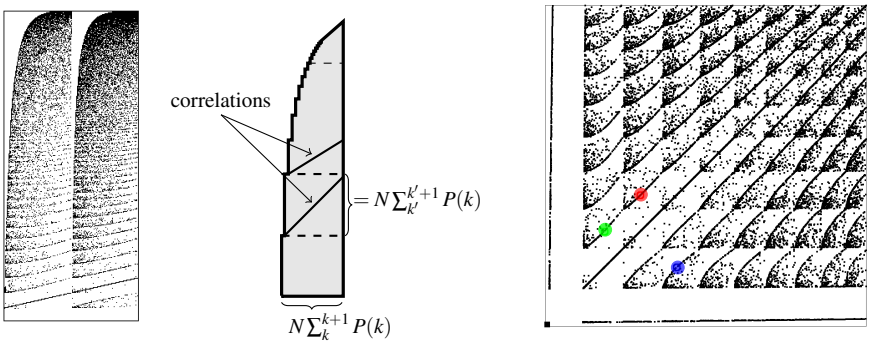


Figure 2. Characteristics of the BOSAM components

links that have degree k at one end and degree less than k' at the other end. The density of points inside a component is related to the conditional probability $P(k'|k)$. The 'lines' radiating from the upper right corner of the bitmap are a feature specific to social networks, and represent correlations. An example of a three point correlation is shown in Figure 2, where the blue, green and red circles have coordinates (x,y), (y,z) and (z,x) that fall on these lines; such points represent a triple of mutually interacting users.

The BOSAMs for the HEALTH and INSPIRE networks exhibit regions of dense shading which are likely related to posts by the site administrators. This is notable since both of these forums are multi-topic and therefore would normally divide into communities with little interaction between them. However, particularly in the case of INSPIRE, site administrators routinely cross-post across the forum and also, to some extent, allow cross-posting by users in the form of journals. This feature tends to obliterate the communities structure. Parts of that structure can still be seen in the BOSAM for the PINFO message board, which suggests that, in comparison, this forum is more lightly managed by its administrators.

5. Conclusions

We presented an analysis of large samples of user interactions from a number of health-related message boards using the BOSAM algorithm. Results based on a simple measure of interaction (i.e. whether two users contribute to the same conversation thread) show an interesting variety of structures that appear to correlate with the features of the corresponding forums such as the topic coverage of the forum (single-topic or multi-topic), the presence of user communities, the commercial or community nature of the forum and consequently its management style. In future work, we plan to include other measures of interaction in our analysis that take into account, for instance, the frequency of interactions and the structure of replies within each thread.

References

[1] E. Børøsund, M. Cvancarova, M. Ekstedt, S. M. Moore, and C. M. Ruland. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. *Journal of Medical Internet Research*, 15(3), 2013.

[2] B. Carron-Arthur, K. Ali, J. A. Cunningham, and K. M. Griffiths. From help-seekers to influential users: a systematic review of participation styles in online health communities. *Journal of Medical Internet Research*, 17(12), 2015.

[3] D. Chakrabarti, C. Faloutsos, and Y. Zhan. Visualization of large networks with min-cut plots, a-plots and r-mat. *International Journal of Human-Computer Studies*, 65(5):434–445, 2007.

[4] T. Chomutare, E. Årsand, and G. Hartvigsen. Characterizing development patterns of health-care social networks. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 2(3):147–157, 2013.

[5] K. T. Durant, A. T. McCray, and C. Safran. Modeling the temporal evolution of an online cancer forum. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pages 356–365. ACM, 2010.

[6] Y. Guo, C. Chen, and S. Zhou. Topology visualisation tool for large-scale communications networks. *Electronics Letters*, 43(10):597–598, 2007.

[7] R. Jones, S. Sharkey, J. Smithson, T. Ford, T. Emmens, E. Hewis, B. Sheaves, and C. Owens. Using metrics to describe the participative stances of members within discussion forums. *Journal of Medical Internet Research*, 13(1), 2011.

[8] Y. Liu and Y. Chen. Patient-centered information extraction for effective search on healthcare forum. In *SBP*, pages 175–183. Springer, 2013.

[9] R. J. Mondragón. Topological modelling of large networks. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 366(1872):1931–1940, 2008.

[10] X. Song, S. Jiang, X. Yan, and H. Chen. Collaborative friendship networks in online healthcare communities: an exponential random graph model analysis. In *International Conference on Smart Health*, pages 75–87. Springer, 2014.