Extracting and Analysing Social Networks of Physicians

Saujoe Liaw\textsuperscript{1}, Wilson Wong, Wei Liu and David Glance
School of Computer Science and Software Engineering
University of Western Australia
Crawley WA 6009
\textsuperscript{1}saujoe@gmail.com, \{wilson,wei,david\}@csse.uwa.edu.au

Abstract

In this paper, we developed a prototype system to extract social data from online sources, build social structures of physicians using multiple, heterogeneous 2-mode networks, and apply social network analysis methods to analyse the networks. A two-part case study is carried out to demonstrate the effectiveness of this approach, which reveal interesting conjectures concerning the issue of language barrier in the healthcare service in Western Australia.

1 Introduction

Increasingly more researchers are employing established techniques and concepts in \textit{Social Network Analysis (SNA)} to uncover patterns, and to assist interpreting the underlying structures of growing and dynamic online social networks. Unlike the conventional gathering of social research data through surveys, data sets obtained from online sources differ in terms of size and complexity. In this paper, we analyse the relations between physicians in Western Australia (WA) using social data gathered from an online directory known as the \textit{Health Directory Service (HDS)}\textsuperscript{1} managed by the \textit{Great Southern Managed Health Network (GSMHN)}. As \textit{HDS} is not intended to be a social networking site, the data gathered from \textit{HDS} do not provide explicit social relations among the physicians. Therefore, we construct heterogeneous 2-mode networks based on attributes associated with physicians such as their expertise, spoken languages, qualifications and workplaces. As a proof of concept, we investigated a two-part case study based on the issue of language barrier in healthcare services to demonstrate the feasibility of studying multiple, attribute-based online social networks using standard \textit{SNA} techniques and concepts.

2 Related Work

Recent work shows an increasing interest in studying online social data using \textit{SNA} techniques and concepts. Memon & Larsen [5] employed cohesion, role and power analysis to uncover hidden structures within 1-mode terrorist networks. The approach identifies specific terrorists who play critical roles in the network. Cai et al. [3] utilised publication data on \textit{DBLP} to construct multiple 1-mode networks consisting of authors as the sole entities. Each network captures co-authorships of a particular conference. Saltz et al. [6] applied \textit{SNA} on an online learning environment to study students' interactions within a class. Degree centrality is applied to determine the extent of students' online participation, an important factor in the success of online learning. Korfiatis et al. [4] modeled the collaborative editing environment of Wikipedia using 2-mode networks consisting of articles and contributors to determine the latter's authoritativeness. The notion of authority is based on contributor degree centrality and article degree centralisation. Ahn et al. [1], on the other hand, compared the structures of three online social networking sites to study the degree distribution and correlation over time.

3 Methodology

The primary source of our data, \textit{HDS}, is an online directory of physicians and healthcare institutions. The fact that \textit{HDS} is not an online social networking service\textsuperscript{2} makes it an interesting and more representative source of data for studying the role of informal, attributed-based relations in \textit{SNA} in the context of healthcare services. We constructed a web crawler for crawling the directory and parsing the HTML pages. Figure 1 summarises the three basic entity types (i.e. organisation, physician and expertise) and the three derived entity types (i.e. qualification, vernacular and location) ex-

\textsuperscript{1}http://hds.gsmhn.com.au/

\textsuperscript{2}HDS does not contain freely associating physicians and organisations.
Figure 1. The entity types extracted from the HDS for this study.

<table>
<thead>
<tr>
<th>entity types</th>
<th>category</th>
<th>total no. of entities</th>
</tr>
</thead>
<tbody>
<tr>
<td>organisation</td>
<td>basic</td>
<td>1,424</td>
</tr>
<tr>
<td>physician</td>
<td>basic</td>
<td>1,161</td>
</tr>
<tr>
<td>expertise</td>
<td>basic</td>
<td>1,408</td>
</tr>
<tr>
<td>qualification</td>
<td>derived</td>
<td>140</td>
</tr>
<tr>
<td>location</td>
<td>derived</td>
<td>89</td>
</tr>
<tr>
<td>vernacular</td>
<td>derived</td>
<td>56</td>
</tr>
</tbody>
</table>

Figure 2. The actors and their attributes in 2-mode matrices.

(a) The number of attributes with degree $d_i > 0$ and the corresponding 2-mode networks is reduced to in-degree of actors. These attributes and clade as vertices attributes with actors contribute to the vertices and degree $d_i > d_j$ and the corresponding actors.

(b) For visualisation purpose, the degree $d_j > 0$ and the corresponding 2-mode networks is reduced to in-degree of actors. These attributes and clade as vertices attributes with actors contribute to the vertices and degree $d_i > d_j$ and the corresponding actors.

Figure 3. The 2-mode network for the matrix PV shows the “speak” relation between physicians and vernaculars. All vernaculars included in this network have more than 5 physicians associated to them. Only 12 vernaculars with a total of 101 physicians and 172 “speak” relations are shown.

3.1 Network Visualisation

We construct a 2-mode network for each actor-attribute combination. We represent our 2-mode data using $m \times n$ incidence matrix (i.e. attribute matrix), $A = \{a_{ij}\}$. The rows of the attribute matrices represent the actors while the columns are the attributes. In our case, we constructed four attribute matrices $PO, PV, PQ$ and $PE$ with $m = |P|$. The cell $a_{ij}$ is set to 1 if the physician at row $i$ has the attribute at column $j$, or $x_j \in X_i$. These 2-mode networks have directional, boolean relations. After obtaining the matrix representation for our actor-attribute data, we proceed to visualise the 2-mode networks. The attributes and actors summarised in Figure 2(a) contribute to the vertices in the network. We use the degree of the attributes to reduce the network for visualisation. The degree of an attribute $j$ (i.e. the number of actors with that particular attribute) is $d_j = \sum_{i=1}^m a_{ij}$. Only vertices with a degree more than zero are included in the 2-mode networks. For example, only 802 out of 1, 424 organisations in matrix PO in Figure 2(a) have physicians associated to them. To illustrate, we included the reduced 2-mode network for the attribute matrix PV in Figure 3. In these reduced networks, only attributes with $d_j$ larger than the threshold $d_T$, together with the actors associated to them are displayed. Figure 2(b) summarises the threshold values, and the associated actors and attributes.

3.2 Matrix Transformation

We transform each actor-attribute 2-mode matrix into one actor common attribute matrix and one attribute overlap matrix for analysing the potential relations between physicians based on their common attributes using standard techniques. The 1-mode actor common attribute matrix, $B = \{b_{ij}\}$ records the number of attributes common to each pair of physicians, or $b_{ij} = \sum_{k=1}^n a_{ik} a_{jk}$. The attribute overlap matrix, $C = \{c_{ij}\}$ captures the number of physicians each pair of attributes has in common, or $c_{ij} = \sum_{k=1}^m a_{ki} a_{kj}$. Unlike the 2-mode networks, these 1-mode relations are non-directional and valued. We construct the 1-mode matrix (i.e. sociomatrix) as the product of the corresponding 2-mode matrix $A$ with its transpose $A^T$, or $B = AA^T$ and $C = A^T A$. As a result, eight 1-mode matrices were generated. The four actor common attribute
matrices are $P_O$, $P_V$, $P_Q$ and $P_E$. The non-directional and valued relations in these networks are known as common attribute relations. The four attribute overlap matrices are $O_P$, $V_P$, $Q_P$ and $E_P$. The relations in these networks are referred to as attribute overlap relations.

### 3.3 Case Studies

The actor common attribute networks enable us to examine the ties among physicians from different perspectives. Here, we study the vernacular and organisation aspects of the relations between physicians in a two-part case study. The networks relevant to our case study are $P_V$, $V_P$, $P_O$ and $O_P$. The issue of under-representation of physicians from various ethnic groups with different vernacular backgrounds is a major concern in the healthcare service [2]. Instead of performing a full scale study in WA, we use the data available as summarised in Figure 1.

![Figure 4. The components of the network based on matrix $V_P$, where the edges in these components have values larger than or equal to 2. There are 30 edges in total with 18 vertices.](image)

In the first part, we analyse the ties between physicians based on their vernacular skills, and the extent to which such ties reflect the state of physician diversity in WA. The visualisation of the 1-mode matrix $V_P$ is shown in Figure 4. The vertices are labeled with the values of closeness centrality and the edges show the number of physicians each pair of vernaculars has in common. There is a visible pattern corresponding to the vernacular capability amongst the physicians. The two components in Figure 4 show the presence of two groups of common languages spoken by physicians, namely, Indo-European (IE) languages and East Asian (EA) languages. This indicates the polarisation of the vernacular capability of physicians. A physician who speaks an EA language is unlikely to be able to converse in any IE language. To study the cohesion among the languages spoken, we analyse the density of these two components. A high density in this case study reflects a high tendency of speaking another language of the same group if a physician is already well-versed in one. Given that $\ell_C$ and $\vartheta_C$ are the sets of edges and vertices in component $C$, we compute the density of $C$ as $\Delta_C = \frac{\sum_{e \in \ell_C} w_e}{|\ell_C|\cdot|\vartheta_C|\cdot(1/2)}$, where $w_e$ is the weight or value attached to edge $e$, and $r$ is the maximum weight in the larger network. In this case study, $r = 10$ and the density of the component EA, $\Delta_{EA} = 0.39$ is larger than that of IE, $\Delta_{IE} = 0.064$. Since every EA language is more likely to be paired up with others in the same component, we speculate that a physician who speaks an EA language has better chances of being well-versed in other EA languages as compared to the physicians in the IE component. Figure 4 shows that Cantonese and Malay have the highest closeness centrality at 0.88 in the EA component. Both the Cantonese and Malay languages are connected to all other languages with distance 1. In other words, any physician who speaks Cantonese or Malay will also speak at least one other language in the EA group. This implies the commonness of these two vernaculars among the physicians who speak EA languages. This in fact supported our conclusion drawn based on density. The French language, with the highest betweenness centrality at 0.4909 in the IE component plays a crucial role as a bridge to connect two subgroups of IE languages spoken by the physicians. French is also the most common language in the IE component with highest closeness centrality at 0.57. Intuitively, a physician who is already well-versed in an IE language has a lower tendency of speaking French. This section of the case study shows the disparity in the vernacular capability of the physicians in WA, with a tendency towards the EA language group. This indicates the need for physicians with vernacular skills capable of bridging the EA and IE groups.

In the second part, we analyse the diversity of physicians in terms of their vernacular capabilities in different organisations. The inclusion of the organisational aspect of the relations among physicians requires the combination of the two matrices $P_V$ and $P_O$. The conventional approach of compounding relations to generate multi-relational networks in SNA is based on the compositionality of relations across different networks (e.g. the compounding of “is a sister of” and “is a mother of” produces “is an aunt of”). Such lateral combination of primitive relations through matrix multiplication may not always be applicable especially in cases where the compound relations are socially inadmissible (e.g. “mother’s wife”). In our case, the new relation should capture the common attributes represented by all the relations to be combined. In this study, the boolean, undirected edge $e_{ij}$ in the new matrix $C$ should reflect both the common vernacular and common organisation of the physi-
of conversing in the same vernacular in each organisation. A social position (i.e. physicians who speak the same language in the same organisation) with no redundancy will have no substitutes if the only member becomes nonfunctional. Thus, a higher redundancy over a diverse range of vernaculars in each organisation would ensure a better chance of providing quality service to patients of low English proficiency.

4 Conclusion

The rise of user-generated content on the Web has contributed to the renewed interest in SNA for analysing increasingly large online social networks. In this paper, we studied multiple, heterogeneous 2-mode networks derived from an online health directory service using standard SNA techniques. Our case study on the issue of language barrier in the healthcare service using these networks revealed several interesting conjectures. Firstly, our analysis revealed the possible presence of disparity in the vernacular capability of physicians in WA. Secondly, there are inadequate physicians capable of conversing the same language (i.e. low redundancy) over a diverse range of vernaculars in each organisation. Some of our future work will include the study of other aspects of the relations between physicians such as qualification and expertise. We are also interested in combining the various common attributes among physicians to establish a multi-relational network.

References