**Visualization Project Description: Autoimmune Hepatitis (AIH) Project**

Client: Department of BioHealth Informatics, Indiana University SOIC, Indianapolis

Main contact: Anand Kulanthaivel, M.I.S. (PhD student, BioHealth Informatics)

Email: [akulanth@iu.edu](mailto:akulanth@iu.edu) | Office: Walker Plaza (WK) #121, 719 Indiana Ave, Indianapolis 46202

Responsible faculty: Dr. Josette Jones, R.N., Ph.D. (BioHealth Informatics)

**The problem.** Autoimmune Hepatitis (AIH) is an incurable, non-viral disease affecting the liver. Due to the disease’s rarity (less than 3,000 Americans are affected by it), research on AIH patients is logistically difficult. However, support for patients with the disease is often provided over social media, such as Facebook. AIH patients typically use social media to provide every day social support, and also to ask health and medical questions about their disease and how it affects them.

We are fortunate enough to have the data from an AIH-related Facebook group, and performed computational topic modeling of 280 users in the group, allowing us to determine the 50 most commonly discussed topics across these users, most of whom are patients.

**General goal of the project:** To enhance the insight of everyday issues faced by AIH patients and those associated with them. Our intent is to target students and professionals who do not have medical knowledge but are interested in learning about the medical and everyday issues that patients with AIH face. Your team will be part of our target demographic, so you can easily self-test the quality of your visualization.

**The Data.** The data are structured in a graph (network) format and metrics have already been analyzed for you.

* Edgelist connecting 280 users to their top 5 strongest topics (thus, a bipartite graph containing 330 nodes connected by 1400 edges)
* User description node file
* Topic description node file.

Edgelist: usertopicedges.csv

*UID,TID*

U0,T4

U122,T23

(etc., for 1,400 user-topic combinations)

User Node File: usernodeproperties.csv

*UID,CommentsMade,PostsMade,CommentsReceived,AdjTopicPopularity,TA,TB,TC,TD,TE\*\**

U33,102,4,80,29.5,T3,T19,T21,T30,T48

(etc., for all 280 users)

\*\**Note: TA through TE are the first through fifth ranked topics for that individual user. They can be treated as equally weighted because the first five topics for each user account for 95% of all user topic strength.*

Topic Node File: topicnodeproperties.csv

*TID,UserPopularity,TopicRank,TopicTitle,Keywords*

T22,11,26,”Insurance problems with medication”,”payor anthem aetna not covering Imuran”

(etc., for all 50 topics)

**Desired Deliverables**

1. A general visualization of all nodes and edges from the user-topic list
   1. Note: The software package you use is your choice (and can even be fully custom code). Anything that can satisfy the deliverable specification can be used.
2. The ability for viewers to click a node and retrieve data about the user or topic that was clicked
   1. Our idea of a wireframe for this functionality is in the attached PPT slide.
3. (Ideally) functionality that allows viewers to color or size nodes by the numerical node metrics in the node file.
4. Given that you are part of our target audience, we appreciate a list of insights that you gained of patients with AIH and associated individuals. As per what you gain from your visualization, remark on what kinds of topics frequently-commenting, frequently-posting, non-posting, non-commenting (etc.) users tend to talk about.
5. We will appreciate that any additional write-up that you do for the course about the project also be submitted to us.
6. Challenge: Our research lab knows for sure that one member in the group is definitely not a patient (as he is the physician who leads the group). Just by using your own visualization, try and find out which user the physician is. (Hint: If your product can color user nodes by their different network metrics, you’ll see that one user has all of his metrics stand out!).

The data are fully de-identified and you will only receive user codes (U0 through U279) and the metrics for each of these users – So, there are no privacy/human subject issues and you can handle and represent the data as you wish.