The Robotic Surgical Network using PubMed Articles

By Kevin Choi, MA 703

Introduction:

Collaborative science has evolved in the past several decades due to the introduction of different types of technology and communication[[1]](#footnote-0). The search engine has been an especially important piece of technology in science; it has allowed scientists to search for new sources of information, and learn from the contents within the search engine. One notable search engine within the medical and life science community is PubMed, which is a free search engine for life science and biomedical topics. Collaboration networks can also help researchers identify topics that are related to a specific research topic and the subfields/communities structured around these topics[[2]](#footnote-1). The project presented attempts to uncover the collaboration between surgeons who use robotic surgery by using network analysis on published articles on PubMed.

The neural network is a convenient method to understand the flow of information between different groups. The construction of the network is non-trivial, but the results are impactful, visual, and insightful. In the medical field surgeons benefit[[3]](#footnote-2) greatly from collaboration as they share different techniques and information with each other, but it’s not always obvious which pair of surgeons can learn the most from each other. This dilemma is mostly due to the fact that surgeons are often times specialists who have a very tight network of collaborators. This paper, and my algorithm attempts to find patterns of collaborations between surgeons in the academic world.

The main surgical fields of interest are general surgery, cardiac surgery, colorectal surgery, gynecologic surgery, head and neck surgery, thoracic surgery, and urologic surgery. The main purpose for choosing robotic surgery is because my lab is focused on robotic surgery outcomes, but moreover they already have some understanding as to which surgeries collaborate a lot with one another. This type of expertise that is present on my team helps me test my algorithm with medical experts.

Methods:

One of the most important tasks in the project is sampling. The sampling in the network follows a simple schema; surgical disciplines which use robotic surgery are being analyzed for collaboration. These surgical disciplines include general surgery, cardiac surgery, colorectal surgery, gynecologic surgery, head and neck surgery, thoracic surgery, and urologic surgery.



Secondly, the construction of the network is one of the most essential tasks involved with the project. It requires efforts in data cleaning, checking, and collecting. The data collection process from the Entrez application programming interface from PubMed is henceforth called the data pipeline. The data pipeline process is shown in the figure below.



The data pipeline requires robust code implementation to consider the volume of data, and extracting specific author information for each article from PubMed. The amount of data cleaning and manipulation meant that the ideal programming language for the project is Python. Once the author information for each article has been retrieved from PubMed the data is organized and structured into an edge list for convenient graph and network implementation. There are several restrictions for the data pipeline including capping the amount of articles collected to 35,000 (despite how large the population of articles for surgery keyword might be), and creating a lower bound of inclusion for the amount of citations for each article to 5.

Once again, the construction of the edge list is non-trivial. The edge list is structured in the following:



Each blue circle represents the nodes in the edge list. The circles are articles based on each surgery field (i.e. cardiac surgery, general surgery, urologic surgery, etc.), and they are connected if they share an author. These article nodes create small communities within each surgery field. For example, the articles related to the surgery keyword cardiac surgery are shown in the graph below.



Each group of connected nodes (i.e. articles) is written by a single author. These nodes all belong to the same surgery keyword for cardiac surgery. In addition, the network is designed in such a way that it is an undirected, weighted graph. However, the motivation behind the connections is to understand how surgeons are involved in other surgical disciplines. For example, we expect surgeons from colorectal surgery to also publish and write in relating fields including urologic surgery. The red graph below shows small network of articles for urologic surgery, and the blue graph shows the small network of articles for colorectal surgery.

 

For a relatively small network of 70 nodes for each graph both urologic and colorectal surgery share three authors who have written and published papers in both surgery fields, which leads to the final methods section the weights. The weighting schema for the edges between each nodes is straightforward. The weights are simply the number of authors existing between surgical fields, and for visual effect the thickness of the edges scale with the weights between each surgical field.

$\sum\_{j=0}^{N}\frac{1}{weight\_{j}}$ (1)

Results:

There are two major findings in the project. One finding in the project are the connections between robotic surgery and the subsequent related surgical fields (i.e. cardiac surgery, general surgery, urologic surgery, etc.). The connections between robotic surgery and the related surgical fields prompts us to understand which surgical fields are intellectually contributing to the field of robotic surgery. The numbers between the nodes represent the number of cross discipline authors (i.e. authors that write in two different surgery fields). For example, gynecologic surgery has 13 authors between itself and robotic surgery.



There is some literature which backs up some of the node connections in the network[[4]](#footnote-3) especially the connection between gynecologic and robotic surgery. Surgery fields which benefit from minimally invasive surgery also seem to adopt robotic surgery easier than other fields. Another noticeable trend in the network is the small amount of connections between cardiac and general surgery with robotic surgery. One explanation is the generality of both cardiac and general surgery. One conclusion about the network is that a smaller proportion of surgeons in cardiac and general surgery are writing about robotic surgery than other surgery fields including gynecologic and urologic.

Another important note about the efficacy of the network is the number of articles and authors being matched between each surgical field. The number is small, approximately 400 articles, however the number of articles per surgical field is evenly distributed throughout the network. In addition, each article has a degree of academic legitimacy because each article must have at least 5 citations to be included in the network. The publication dates have not been considered in the network, and instead the data pipeline extracts a percentage of the total population of articles (i.e. approximately 20%). Since article citations take some time we can assume that many of the articles have not been published recently (i.e. within one year).

The second major finding in the project is the connections between each surgical field. The network looks quite different from the robotic surgery network. One of the first things to notice about the connections between the surgical fields is that general and colorectal surgery are the only nodes that are connected with every other node. The missing edges are cardiac and gynecologic surgery, urologic and head and neck surgery, and gynecologic and thoracic surgery.

|  |  |
| --- | --- |
| cardiac surgery | gynecologic surgery |
| urologic surgery | head and neck surgery |
| gynecologic surgery | thoracic surgery |



There are no collaborative authors between these surgical fields. In addition, unlike the previous robotic network where urologic and gynecologic surgery were the most collaborative fields, the surgery network has cardiac and colorectal surgery as the most collaborative fields. It’s also not surprising to see that cardiac and thoracic surgery sharing 260 authors since thoracic surgery is a subcategory of cardiac surgery. The charts below show the number of authors for each surgical field.



The collaboration skewness for general, thoracic, and cardiac surgery show that the there are several surgical fields that these particular fields collaborate a lot with, but they do not collaborate a lot with some of the other fields.

Another notable feature of the surgery network is the concentration of author collaboration between the four nodes cardiac, general, colorectal, and thoracic surgery. In fact, it’s surprising to see such a large amount of author collaboration between general and colorectal surgery. The highest amount of author collaboration appears to be between the cardiac, general, and thoracic surgery.

Centrality measurements are not the best form of analysis for the robotic and surgery networks. The amount of nodes are small because the project aims to look at the recommended surgeries for robotic surgery. The network does not show deep enough structure to explore centrality measurements, and instead the project relies on visual graph descriptions.

Conclusions:

Collaboration in science has evolved with the introduction of new technologies including the search engine. In addition, popular search engines like PubMed have become widely available for researchers and scientists. By collecting data on PubMed using the Entrez API we can look at how surgeons collaborate between different types of surgical fields. The data can be organized and structured into a network to further understand the collaboration between surgical fields. The main interest in the project is to study how different types of surgical fields interact with robotic surgery, and how these surgical fields interact with one another. One of the main reasons robotic surgery is a topic worth investigating is the fact that it’s a relatively new form of surgery, and different surgery fields are adopting them. The project results show that specific surgery fields have more collaborators with robotic surgery, but they do not collaborate with other surgical fields as much. In addition, broader forms of surgery show higher amounts of collaboration between other surgical fields. In the future, the data collected from PubMed can explore other types of networks including research in certain epidemics are related, and how other medical disciplines are connected with one another. Moreover, the benefit of using neural networks to understand data is the flexibility. The nodes and edges can be constructed fairly easily after the data has been collected. In addition, neural networks allow researchers to uncover patterns in data that are otherwise difficult to identify. Lastly, the project requires a larger network, and a more robust data pipeline.

1. http://www.iseh.org/blogpost/772431/186685/Collaborative-research-the-why-and-how-to-collaborate-in-science [↑](#footnote-ref-0)
2. http://www.casos.cs.cmu.edu/publications/papers/CMU-ECE-2011-017.pdf [↑](#footnote-ref-1)
3. https://www.facs.org/media/press-releases/jacs/tenn-nsqip0112 [↑](#footnote-ref-2)
4. Antoniou, Stavros A., et al. "Past, Present, and Future of Minimally Invasive Abdominal Surgery." *JSLS: Journal of the Society of Laparoendoscopic Surgeons* 19.3 (2015). [↑](#footnote-ref-3)