

Prof Kaplan,

I am giving the basic outline of the work I did:

1. I was given four files (one for each year – 2013 to 2016), consisting of medical records for patients along with the doctors that were involved in the procedure. The objective was to calculate the number of dyadic interactions between each doctor (or a subset of doctors, depending on what Prof. Kaplan wanted).
2. The first step was to merge all the datasets into one single file. Some years had more columns and some had less. (each column here corresponds to a role. Ex: surgeon1, scrub1 , scrub2 etc) . I merged all the four datasets and made sure all the roles were accounted for. I arranged all the entries chronologically since we have to calculate dyadic interactions over time.
3. Then, I removed the columns for which we did not need to calculate the dyadic interactions (Ex: radiologist , chemist etc) .
4. There was a bit of data cleaning after this step (removing stop words, changing the character encoding to UTF-8 etc). At this stage, I also consulted with Prof. Kaplan and removed entries in the dataset which were potential red flags (Ex : There were ~500 entries with the doctor name as Unknown).
5. I wrote the Java code for calculating the dyadic interactions. There were a few issues with memory allocating (owing to the large nature of the dataset, my machine was sucking up all of the ram, and the program kept crashing).
6. I improved the code over ~6 iterations. In the final iteration, I made sure Prof. Kaplan understood how Java worked, and how he could run the program for any subset that he needed for (for example , he can calculate the interactions for patients > 60 , surgeries performed in a pre-determined time delta , for ASA class = 3 , gender = female etc) .
7. I also mentioned the different visualizations that he could work with, especially regarding visualizing the dataset as a social network to understand collaboration clusters.
8. Files submitted to Prof. Kaplan: Final merged dataset, Java code , sample image of network visualization.

Thank you,
Saurabh Rao