Team 43 Retrospective

**Part 1 Client Demo and Feedback:**

**Backend:**

**Client’s thoughts:**

Client clarified that lack of ID listed for certain team members was not an error, they most likely don’t have those members SCOPUS Ids.

Due to ballooning, we suggested a limit to the number of authors a publication can have and still be included. Limiting publications to those with 30 authors or less is suggested, which is will potentially miss out on some larger clinical trial papers. We will endeavour to have as large a database as practical, but client is understanding of the possible growth rates.

The four senior researchers in the core team are also causing large scale ballooning, with three of them each individually having more co-authors than the entire rest of the core team combined. We will try and incorporate first degree links for the big four but will likely be unable go beyond at this scale.

There is still value in including the senior researchers, Bruce and Janette could likely be largely overlayed as they publish a lot of collaborative papers. Client noted that if we are treating these large cases differently, to also include the fourth member of the senior researcher group.

No specific maximum size in mind, as there was no specific plan for how many links should be included. Our plan is to produce an instance of the whole network that can be updated once a year along with individual maps.

Changing names and multiple Id’s is a massive problem in all of research and will need to attempt to resolve errors on the backend. Ensure that unresolved errors prompt user with the steps to ensure a fix is implemented. Errors that SCOPUS discovers can be handled relatively easily, but undiscovered duplicates are an issue beyond the scope of this project to fix.

The value of the network is within finding researchers that the core research team does not know and are well connected, potentially build extension to highlight these researchers.

**Teams’ thoughts:**

We are tracking well overall, currently at the point where the biggest rate limiting factor is SCOPUS and their limiting 5000 queries per week. We did not anticipate the exponential nature of networks and the massive amounts of publications and coauthors on different publications resulting in often many more than 5000 queries for links.

We are currently at the point where we have the layout of the database and the script to populate the database done and its now working with SCOPUS to try and populate the database over the next few weeks with the belief that the maintenance and updating queries will be far less than the startup cost.

We are putting a lot of time and effort into ensuring the database is future proof with extra redundancy and checks and balances to ensure that erroneous researchers and conditions are resolved on the backend without user input. The future goals are to build a complete database of first and second degree links and setup Dr Ward’s API key so that the database can be updated.

We did have issues with the scale of the project. Whilst the average researchers and papers on scopus have relatively small impacts on the database as a whole, we unfortunately burned through a few weeks of queries very quickly due to the unforeseen impact of the Senior Researchers, with over 1500 co-authors each. This, combined with a number of 700+ author publications by core researchers, had unfortunately caused our database to be far larger than originally estimated.

Due to the time taken to acquire an API key, and a secure token allowing us to use the API key off campus, we were delayed in discovering this limitation. We would have been well served to check the core research team ahead of time using the scopus website, which would have forewarned us, allowing us to ask for cutoff estimates earlier. We also may have been more able to provide direct demonstrations if we had used a smaller core team for testing, preventing excessive ballooning, whilst showing proof of concept.

**Frontend:**

**Client’s thoughts:**

The lines demonstrating degrees of linkage and relations are ambiguous and don’t demonstrate the relationships properly. Need to update the groups that topics belong to so that there is a discernible key with the randomly assigned colours otherwise the network is ambiguous. Need to have numerous lines to demonstrate multiple publications between researchers. Happy with thickness to display the number of relations rather than count. Is topic based on keywords which are assigned by authors or the SCOPUS assigned topic which are inaccurate. Attempt to use author keywords instead of SCOPUS assigned, try and group these keywords to make meaningful topics. Happy with input form to update list.

**Teams’ thoughts:**

So far, in terms of front-end development things have been tracking along smoothly. Currently we have established a colour palette and theme for the overall design of the app. We have implemented a graph visualising tool which utilises neovis and the local neo4j database to populate nodes as well as having a page to add or remove researchers and publications. There are some tasks remaining such as populating the database with researchers as well as certain changes to the styling of the network graphing tool itself as well as displaying certain data regarding researchers and publications. Overall, progress has been smooth and we expect the outstanding tasks to be completed within the timeline set out for sprint 3.

Our current implementation of graphing software (Neo4J) has had unforeseen limitations restricting the customisation of the network graph, including but not limited to colour allocation of nodes and edges, font, and node shape, as well as the effect of interaction with the nodes. 'Node Physics' or the rules governing the way the nodes behave when moved across the screen are preset in Neo4J, nodes can sometimes overlap, and there is no way to currently prevent edges crossing. In order to introduce further aesthetics and customisability to really make the program stand out, we believe it might be worth researching other graphing tools as we move forward with sprint three.