The COMAP competition is an annual international mathematical modeling competition held in February. Every year there are three problems contestants can choose from with one of them being an interdisciplinary problem. This year, the interdisciplinary problem concerned networking science, charging the participants with the task of developing an algorithm to apply to a network of authors to determine which one is most influential within that network (see attached prompt).

In the digital world we live in today, with the constant flow of information between people, businesses, countries, etc. made possible by the leaps and bounds we have made in technology, analyzing and visualizing these webs of interactions has recently become a hot topic in mathematics. The ability to study and quantify relationships, whether in biological processes, social media, or the publishing network of mathematicians, can reveal many exciting things, and the advancement in technology has also aided us in this area. Combining mathematics with this new technology, a preliminary algorithm has been developed during the COMAP competition to determine the most influential mathematician in what we have labeled as the "Erdos1" network (see attached data sample). The goal is to create an algorithm to predict who the most influential member of a network is or will be, and to optimize its generality, accuracy, and ease of application. The end goal of this project is to apply the final algorithm to two to three different manageable networks to test its generality; one network of authors, given that much data is already gathered for this network, one in some potentially commercial area, or one network concerning a biological system.

This project will contain elements of computer science, particularly data manipulation, and both applied and pure mathematics when looking at the theory of how to construct a relevant mathematical model and consequently solving it for a prediction. From this point forward, the elements of a network will be referred to as nodes, the connections between them edges, and peripheral data relating to each node as attributes. For example, in the network of coauthors, the coauthors themselves are the nodes; the coauthor relationship between to authors the edges, and the importance of each author’s research is an attribute. With this knowledge, we can outline how the algorithm will work as follows:

* Assuming some sort of data had already been collected or is in some accessible database, the relevant data would be extracted and formatted in a way the algorithm can “understand”. This is the data mining and data processing step. In the case of the COMAP solution, the attached data sample is the final result of this step, and should be done for all data concerning nodes and attributes.
* Once in a workable format, each attribute can be formed into its own network and consequently ranked by influence. These resulting rankings will be weights attributed to each corresponding node. In the case of multiple attributes, the resulting weights for each one will be combined to create a final weight to be applied to each node. This ranking will be performed by the first algorithm, which is nested within the main algorithm.
* Using these weights and the structure of the given network, the algorithm will then calculate the ranking of each node from most influential to least.

An idea of what this has looks like can be seen in the COMAP example. Given the network of coauthors, a network was compiled of their research papers and each paper given a rank according to the number of citations it received. In this way, each author received a weight according to how important his or her research was and, armed with this information, the algorithm predicted the influence of each author within this network.

As mentioned previously, this project involves elements of computer science and both ends of the mathematical spectrum (pure vs. applied). One collaborator brings to the table experience in computer programming in both the R and MATLAB languages with over a year of experience using MATLAB and a course taken in R, while the other brings more experience in the applied mathematics branch, given the history in the physics major. Both collaborators come with a strong background in pure mathematics, but with their history of collaboration, it has become noticeable that each approaches problems in different ways, offering fresh perspectives to a problem the other may be stuck on. Lastly, the breadth of their experience combined covers all subjects covered in Asher and others. In this way, especially with the interdisciplinary nature of the given problem, much more progress can be gained from this collaboration relative to just one author.

While both contributors have areas of strength in differing fields, each is proficient and well equipped all around to be able to do what would need to be done in the unforeseen event that one would have to drop the project. In fact, no step in this thesis will be done without both authors fully understanding each step in the process. The benefit comes not exactly from division of labor, with two bodies of work being sewn together at into one at the end, but from each one contributing their strengths to have the project moving forward on all fronts. If this project were to unfortunately be reduced to one student, that student would have the necessary tools to be able to finish but obviously without the same gains.