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 "Erdos1dv" network (see attached data sample for difference between raw data and processed data). 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.

The repository service GitHub will be used to store relevant files in a way that will allow each student to have access while keeping track of any changes. Each of their contributions can be tracked and progress on the project observed by their thesis advisor and second reader to maintain suitable progress throughout the semester. In addition to keeping track of changes and contributions, this will allow the algorithm(s) to be available to the public for eventual use or study, offering the opportunity for this research to directly advance the field of network science.

(Prompt for COMAP Problem)

**2014 ICM Problem  
Using Networks to Measure Influence and Impact**

One of the techniques to determine influence of academic research is to build and measure properties of citation or co-author networks. Co-authoring a manuscript usually connotes a strong influential connection between researchers. One of the most famous academic co-authors was the 20-century mathematician Paul Erdös who had over 500 co-authors and published over 1400 technical research papers. It is ironic, or perhaps not, that Erdös is also one of the influencers in building the foundation for the emerging interdisciplinary science of networks, particularly, through his publication with Alfred Rényi of the paper “On Random Graphs” in 1959. Erdös’s role as a collaborator was so significant in the field of mathematics that mathematicians often measure their closeness to Erdös through analysis of Erdös’s amazingly large and robust co-author network (see the website http://www.oakland.edu/enp/ ). The unusual and fascinating story of Paul Erdös as a gifted mathematician, talented problem solver, and master collaborator is provided in many books and on-line websites  
(e.g., http://www-history.mcs.st-and.ac.uk/Biographies/Erdos.html). Perhaps his itinerant lifestyle, frequently staying with or residing with his collaborators, and giving much of his money to students as prizes for solving problems, enabled his co-authorships to flourish and helped build his astounding network of influence in several areas of mathematics. In order to measure such influence as Erdös produced, there are network-based evaluation tools that use co-author and citation data to determine impact factor of researchers, publications, and journals. Some of these are Science Citation Index, H- factor, Impact factor, Eigenfactor, etc. Google Scholar is also a good data tool to use for network influence or impact data collection and analysis. Your team’s goal for ICM 2014 is to analyze influence and impact in research networks and other areas of society. Your tasks to do this include:

1) Build the co-author network of the Erdos1 authors (you can use the file from the website https://files.oakland.edu/users/grossman/enp/Erdos1.html or the one we include at **Erdos1.htm** ). You should build a co-author network of the approximately 510 researchers from the file Erdos1, who coauthored a paper with Erdös, but do not include Erdös. This will take some skilled data extraction and modeling efforts to obtain the correct set of nodes (the Erdös coauthors) and their links (connections with one another as coauthors). There are over 18,000 lines of raw data in Erdos1 file, but many of them will not be used since they are links to people outside the Erdos1 network. If necessary, you can limit the size of your network to analyze in order to calibrate your influence measurement algorithm. Once built, analyze the properties of this network. (Again, do not include Erdös --- he is the most influential and would be connected to all nodes in the network. In this case, it’s co-authorship with him that builds the network, but he is not part of the network or the analysis.)

1. 2)  Develop influence measure(s) to determine who in this Erdos1 network has significant influence within the network. Consider who has published important works or connects important researchers within Erdos1. Again, assume Erdös is not there to play these roles.
2. 3)  Another type of influence measure might be to compare the significance of a research paper by analyzing the important works that follow from its publication. Choose some set of foundational papers in the emerging field of network science either from the attached list (**NetSciFoundation.pdf**) or papers you discover. Use these papers to analyze and develop a model to determine their relative influence. Build the influence (coauthor or citation) networks and calculate appropriate measures for your analysis. Which of the papers in your set do you consider is the most influential in network science and why? Is there a similar way to determine the role or influence measure of an individual network researcher? Consider how you would measure the role, influence, or impact of a specific university, department, or a journal in network science? Discuss methodology to develop such measures and the data that would need to be collected.
3. 4)  Implement your algorithm on a completely different set of network influence data --- for instance, influential songwriters, music bands, performers, movie actors, directors, movies, TV shows, columnists, journalists, newspapers, magazines, novelists, novels, bloggers, tweeters, or any data set you care to analyze. You may wish to restrict the network to a specific genre or geographic location or predetermined size.
4. 5)  Finally, discuss the science, understanding and utility of modeling influence and impact within networks. Could individuals, organizations, nations, and society use influence methodology to improve relationships, conduct business, and make wise decisions? For instance, at the individual level, describe how you could use your measures and algorithms to choose who to try to co-author with in order to boost your mathematical influence as rapidly as possible. Or how can you use your models and results to help decide on a graduate school or thesis advisor to select for your future academic work?
5. 6)  Write a report explaining your modeling methodology, your network-based influence and impact measures, and your progress and results for the previous five tasks. The report must not exceed 20 pages (not including your summary sheet) and should present solid analysis of your network data; strengths, weaknesses, and sensitivity of your methodology; and the power of modeling these phenomena using network science.

\*Your submission should consist of a 1 page Summary Sheet and your solution cannot exceed 20 pages for a maximum of 21 pages. Raw Data

ABBOTT, HARVEY LESLIE 1974

Aull, Charles E.

Brown, Ezra A.

Dierker, Paul F.

Exoo, Geoffrey

Gardner, Ben

HANSON, DENIS

Hare, Donovan R.

Katchalski, Meir

Liu, Andy C. F.

MEIR, AMRAM

MOON, JOHN W.

MOSER, LEO\*

Pareek, Chandra Mohan

Riddell, James

SAUER, NORBERT W.

SIMMONS, GUSTAVUS J.

Smuga-Otto, M. J.

SUBBARAO, MATUKUMALLI VENKATA\*

Suryanarayana, D.

Toft, Bjarne

Wang, Edward Tzu Hsia

Williams, Emlyn R.

Zhou, Bing

ACZEL, JANOS D. 1965

Abbas, Ali E.

Aczel, S.

Alsina Catala, Claudi

…

Processed Data, Matching Author to Coauthor (Authors with Erdos Number 2 Excluded)

"ABBOTT, HARVEY LESLIE" "HANSON, DENIS"

"ABBOTT, HARVEY LESLIE" "MEIR, AMRAM"

"ABBOTT, HARVEY LESLIE" "MOON, JOHN W."

"ABBOTT, HARVEY LESLIE" "MOSER, LEO\*"

"ABBOTT, HARVEY LESLIE" "SAUER, NORBERT W."

"ABBOTT, HARVEY LESLIE" "SIMMONS, GUSTAVUS J."

"ABBOTT, HARVEY LESLIE" "SUBBARAO, MATUKUMALLI VENKATA\*"

"ACZEL, JANOS D." "RENYI, ALFRED A.\*"

"ACZEL, JANOS D." "STRAUS, ERNST GABOR\*"

"AGOH, TAKASHI" "GRANVILLE, ANDREW JAMES"

…