Judges' Commentary: Modeling for Crime Busting

Chris Arney
Dept. of Mathematical Sciences
U.S. Military Academy
West Point, NY 10996
david.arney@usma.edu

Kathryn Coronges
Dept. of Behavioral Sciences and Leadership
U.S. Military Academy
West Point, NY 10996

Introduction

The new topic area for this year's Interdisciplinary Contest in Modeling (ICM) was network science. The shift was popular with the student teams, since a record 1,329 teams submitted papers in solution to a "crime-busting" problem. Network science and / or social network analysis will continue to be the topic area for next year's problem as well. So, for teams that enjoyed this year's problem or want to prepare early for next year's contest, prepare by studying network modeling and assemble a team with that subject in mind.

The ICM continues to be an opportunity for teams of students to tackle challenging, real-world problems that require a wide breadth of understanding in multiple academic subjects. These elements are practically part of the definition of network science—an emerging subject that blends concepts, theories, structures, processes, and applications from mathematics, computer science, operations research, sociology, information science, and several other fields. ICM problems are often open-ended and challenging. Some, like the one this year, could be termed "wicked," in that there is not one correct answer nor a set or established method to model such a problem.

The UMAP Journal 33 (3) (2012) 293–303.

The complex nature of the ICM problems and the short time limit require effective communication and coordination of effort among team members. One of the most challenging issues for the team is how to best organize and collaborate to use each team member's skills and talents. Teams that solve this organizational challenge often submit solutions that can rise to the final rounds of judging.

The Criminal Network Analysis Problem

The Information Age, along with its information-laden and highly-linked Internet, has brought us many amazing capabilities, along with new ways to commit crimes. This year's problem focused on potential conspirators within a company's communication network plotting to commit a crime. Some people were already identified either as known conspirators or as known non-conspirators. The goal of the model was to identify the most likely conspirators from the remaining people in the network through the analysis of confiscated and categorized message traffic. The many connections and links between the people and the messages made this an especially appropriate topic for network modeling. The main tasks expected of the students were to:

- **Requirement 1:** Build a model to prioritize the 83 people by likelihood of being part of the conspiracy and explain your model and metrics. Are any senior managers of the company involved in the conspiracy?
- **Requirement 2:** As new information comes to light, use your model to analyze this changing situation. A good network model is flexible and able to handle the changing nature, structure and information in a dynamic network setting.
- **Requirement 3:** If you could obtain the original messages, explain how semantic and text analyses of the message traffic could help you develop even better models.
- **Requirement 4:** Explain the network modeling techniques you developed and how they can be used to identify, prioritize, and categorize nodes in a network involving other kinds of data sources, not just crime and message data. Does your model generalize to other important problems in society? Again, this is the mark of strong models within network science and their potential to impact society.

Judges' Criteria

The panel of expert judges were impressed both by the strength of many of the submissions of individual teams, and fascinated by the variety of

innovative approaches that students used to address the issues, challenges, and questions that were posed by the problem. The papers were rich in modeling methodology and creativity. In order to ensure that the individual judges assessed submissions on the same criteria, a rubric was developed. The framework used to evaluate submissions is described below.

Executive Summary

It was important that students succinctly and clearly explained the highlights of their submissions. The executive summary should contain brief descriptions of both the modeling approach and the bottom-line results. The remaining report provides a more detailed statement of the contents of the executive summary. One mark of an Outstanding paper is a summary with a well-connected and concise description of the approach used, the results obtained and any recommendations.

Modeling

Models and measures were needed to classify the people in the organization to identify conspirators. Many teams used probability or likelihood measures for criminal-like behavior of the people within the context of the known data. Other used decision-making criteria as their basic modeling framework. Some teams used the explicit structures of networks or graphs to determine classic local or global network metrics, properties, node clusters, or performance outcomes. For such a structure, critical assumptions, such as the directionality of influence and connection within the graph, lead to viable network models. Other teams ignored some of the aspects of the network structure and performed data mining, element classification, and discrimination. Those teams often found prioritization and ordering easier than discrimination.

Where to draw the line and commit to predict a conspirator was sometimes difficult. No matter the modeling framework, the assumptions needed for these models and the careful and appropriate development of these models were important in evaluating the quality of the solutions. The better submissions explicitly discussed why key assumptions were made and how assumptions affected the model development. Stronger submissions presented a balanced mix of mathematics and prose rather than a series of equations and parameter values without explanation. One major discriminator was the use or misuse of arbitrary parameters without any explanation or analysis. Establishing and explaining parameter values in models are at least as significant as making and validating assumptions.

Science

Semantic and text analysis are elements of the science of computational linguistics or natural language processing involving many challenging scientific and technological issues related to the nature, value and understanding of information and the production of knowledge or intelligence. Currently, many information-rich systems and organizations are facing data deluge and overload. Vast amounts of unstructured textual data are often collected and held for practically impossible human analysis. The magnitude of data makes this potentially valuable information at best a worthless distraction. Through natural language processing using semantic and text analysis the potentially valuable but hidden information can become visible, understandable, organized, and useful.

The ultimate goals of semantic and text analysis are to identify context, meaning, categorization, and entity attributes, and thereby produce human-ready synopses and standardized, interconnected, structured data (information networks). These highly sophisticated and complex processes are exactly what would be needed to model and solve this network conspiracy problem. Some teams did effective research and insightful analysis that tackled the complexity of the problem and included elements of text or semantic analysis in their model or described how their model could accommodate such capability had the raw message data been available. No matter what modeling was performed by the teams, the interdisciplinary nature of this problem was fully revealed in this requirement. These areas of information science and analytics will experience significant scientific and technological improvements in the future, and the ICM teams were exposed to this developing field in the context of their interdisciplinary science research.

Data/Validity/Sensitivity

Once the model was created, the use of test data and checks on the accuracy and robustness of the solution help to build confidence in the modeling approach. Sensitivity analysis of models to determine the effects of changing data and errors can often be more meaningful than specific output values. This is especially true for highly-structured and powerful data-rich models like networks. Some network structures are highly robust and flexible while others are fragile and highly sensitive to data. While this is a challenging element of network modeling, it was important to address this issue in the report.

Strengths/Weaknesses

A discussion of the strengths and weaknesses of the models is often where students demonstrate their understanding of what they have created.

The ability of a team to make useful recommendations fades quickly if team members do not understand the limitations or constraints of their assumptions or the implications of their modeling methodology. Networks are complex structures and, therefore, the strengths and weaknesses are often hidden from direct view or control of the modeler. Again, the better teams were able to discuss these elements despite these challenges.

Communication/Visuals/Charts

To clearly explain solutions, teams must use multiple modes of expression including diagrams and graphs, and, in the case of this competition, English. A solution that could not be understood did not progress to the final rounds of judging. The judges were delighted by the amazing array of powerful charts and graphs that explained both models and results. **Figures 1–3** on the this page and the next are intended as samples to show the richness of this kind of graphical analysis and reporting.

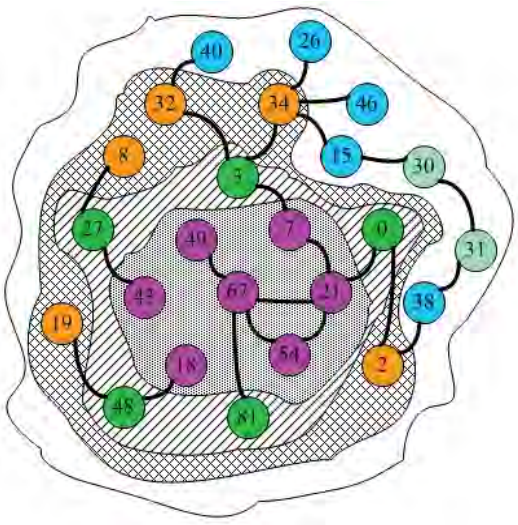

Figure 1. Teams provided informative graphic schematics to show the relationship and connections uncovered by their models. This graphic is from Team 12460 from Harbin Institute of Technology in Harbin, Heilongjiang, China.

Recommendations

Teams were specifically asked to discuss their conspiratorial priorities and the potential involvement of senior managers in their report that would be read by the district attorney. The ability of teams to evaluate the results of their analysis and make recommendations was important in identifying strong submissions.

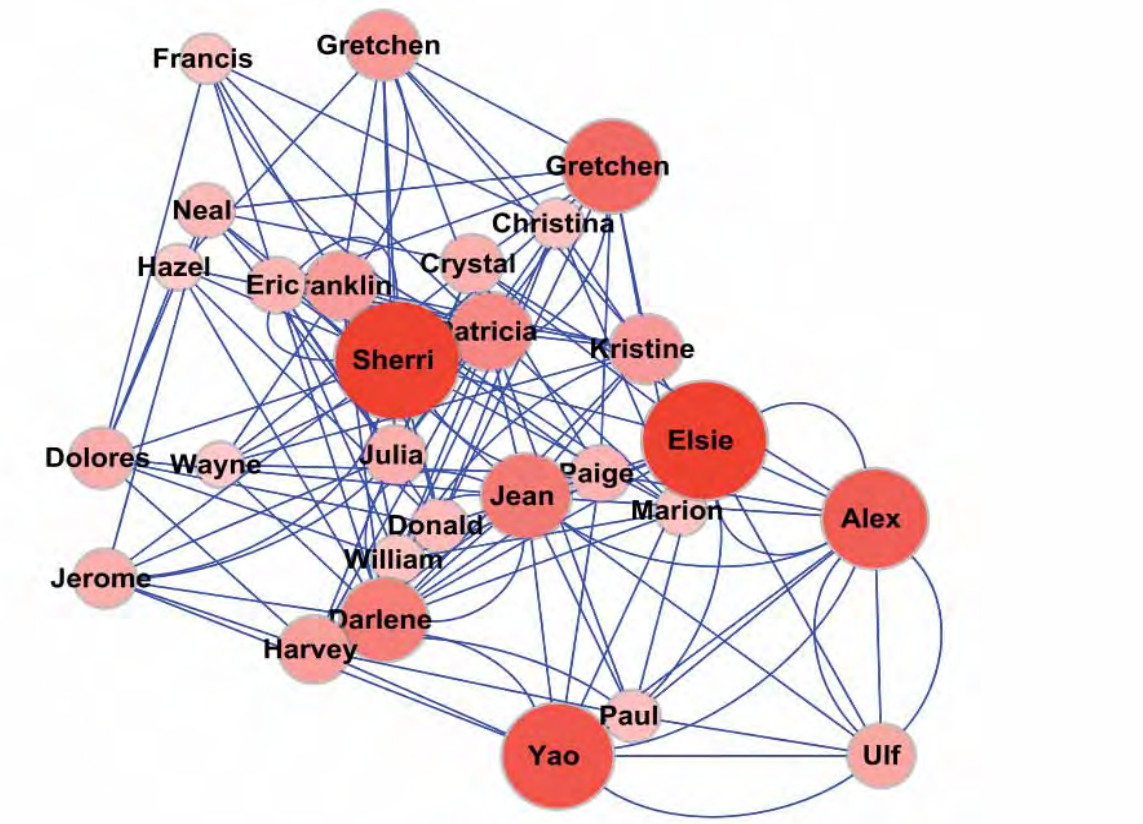

Figure 2. This network portrayal vividly showing the likelihood of conspirators is from Team 16075 from Huazhong University of Science and Technology in Wuhan, Hubei, China.

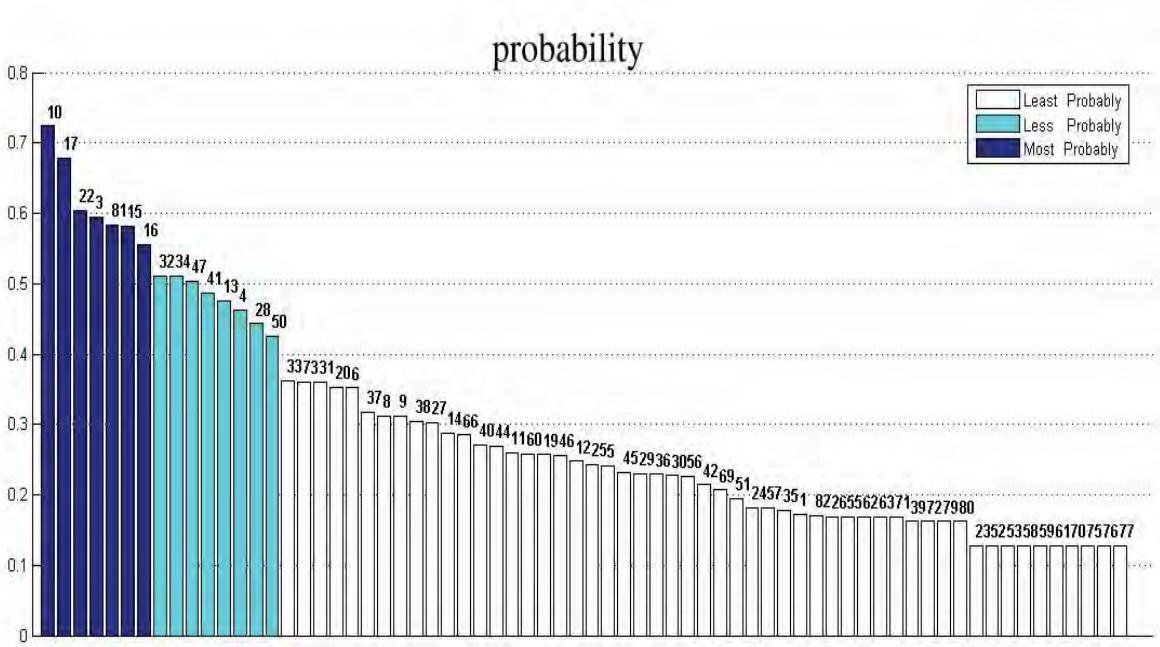

Figure 3. Teams that performed data analysis often used probability charts like this one from Team 13104 from Southeast University, Jiulonghu Campus, Nanjin, Jiangsu, China, to demonstrate their results.

Discussion of the Outstanding Papers

As you will discover in this section, many different approaches were used by ICM teams to model various aspects of the problem. Some teams used the basic structure of networks and their properties and computed classic centrality measures to tackle the issues. Some chose to model using a data mining framework. The Analytic Hierarchy Process (AHP) was a common method for addressing discrimination in the identification of a potential conspirator. As a result, the submissions this year were diverse and interesting to read. Overall, the basic modeling was often sound, creative, and sometimes quite powerful. Those that did not reach final judging generally suffered from two shortcomings. Some lacked clear explanation or evidence to support their conclusions and recommendations. They seemed to jump from their modeling directly to the results without sufficient analysis. Others failed to connect their mathematical models to the aspects and basic elements of information science. In general, poor communication was the most significant discriminator in determining which papers reached the final judging stage. Although the outstanding papers used different methodologies, they all addressed the problem in a comprehensive way by embracing the complexity of the issues, data, questions, and team objectives. These papers were generally well written and presented explanations of their modeling procedures. In several outstanding papers, a unique or innovative approach distinguished them from the rest of the finalists. Others were noteworthy for either the thoroughness of their modeling or the power of their communicated results.

Huazhong University of Science and Technology

The ICM team from Huazhong University of Science and Technology, Wuhan, China performed a thorough network analysis of the information flow and relationships of employees in the organization. In their paper, "Extended Criminal Network Analysis Model Allows Conspirators Nowhere to Hide," they provided an in-depth analysis of the relationships between people and the way the criminal network operated and expanded. This report presented their framework, models, analysis, and results in powerful visual formats that enabled readers to understand their work and feel confident in their results. In many ways, this paper is an excellent example of the potential of network modeling and the power of social network analysis to sort out nodal, edge, and data attributes through use of network measures and data analysis.

Mathematical Modeling Innovative Practice Base

The report entitled "iRank Model: A New Approach to Criminal Network Detection" was submitted by a team from the Mathematical Modeling

Innovative Practice Base, China. The Mathematical Modeling Innovative Practice Base, China, established in 2008, is an institute that promotes interdisciplinary research and educational activities, integrating mathematical modeling and computational approaches to address problems arising in various areas of science and engineering. Their report contained creative analysis of the available data from several perspectives, starting with basic analysis as shown by:

> Carefully examining into the patterns of information exchanges and social connections in the network, we can see that only 24% messages carry conspiratorial information, which seems not systematically significant given that 20% of all the topics are conspiratorial. Therefore, two patterns can be inferred from the statistical results:
>
> - Although conspirators are generally more active than the known innocent people, they exchange irrelevant information with each other. Conspiratorial messages only take a small portion in their message traffic.
> - Since the existing 7 conspirators have already involved in spreading about 40% of the total conspiratorial messages, it is very likely that the total number of conspirators is less than 20.

They also performed a very thorough social network analysis of the message network. This report contained excellent visualizations to explain their algorithm, analysis and results.

Nanjing University of Information Science and Technology

The ICM team from Nanjing University of Information Science and Technology, Nanjing, China, built three different models for finding and separating conspirators and then merged these for their best-case solution. A fourth model was used to identify the conspiracy leaders. Their paper, "Message Network Modeling for Crime Busting," was an excellent synopsis of the diverse methods one could use to approach this problem. Their emphasis was in classical network analysis and data mining algorithms. Once again, this team did a thorough job analyzing semantic analysis and its utility for information and network modeling.

Northwestern Polytechnical University

Finding the hidden features of a network was the theme of the paper entitled "Social Network Analysis in Crime Busting," by the ICM team from Northwestern Polytechnical University, Shaanxi, China. This paper started with the foundations of graphs and networks and built the concept of cooperation within the network. This concept was a fundamentally sound and deeper approach than those of many of the other models. The resulting model was a powerful one for understanding a conspiracy and

the team did an excellent job in their creative modeling and analysis. Their discussion on semantics and text analysis was thorough and insightful in finding ways for possible inclusion of these more powerful methodologies in their models.

Shanghai Jiaotong University

"Crime Busting by an Iterative 2-phase Propagation Method," was submitted by a team from Shanghai Jiaotong University, Shanghai, China. Their classic propagation model of performing iterative and alternating computation of person suspiciousness and topic suspiciousness from each other was creative and powerful. Upon convergence of their model, they produced a priority list of conspirators and performed a thorough analysis. This team's model was both mathematically and scientifically simple yet elegant.

University of Electronic Science and Technology of China

The report and work entitled "Finding Conspirators in the Network: Machine learning with Resource-allocation Dynamics" from the University of Electronic Science and Technology of China, Chengdu, China, was strong from start to finish. This team made careful and thorough assumptions:

> (i) Two classes, conspirators and non-conspirators, are linearly separable in the space spanned by local features of a node, which is necessary to machine learning. (ii) A conspirator is reluctant to mention topics related to crime when talking with an outsider. (iii) Conspirators tend not to talk about irrelevant topics frequently with each other. (iv) The leader of conspiracy tries to minimize risk by restricting direct contacts. (v) A non-conspirator has no idea of who are conspirators, thus treating conspirator and non-conspirators equally.

Then they used machine learning and logistic regression to build their model. They were careful to show their analysis of leader selection and other problem requirements. They followed up their modeling and analysis with sensitivity analysis and a careful discussion of the strengths and weaknesses of their model and its approach. Most impressive was their ability to discuss the incorporation of semantic analysis into their model and the discussion of the power of information modeling to the future.

Cornell University

The team from Cornell University, Ithaca, NY, took a very different approach than the other Outstanding papers. Their paper "Crime Ring Analysis with Electric Networks" presented a model using an electrical circuit analogy for the conspiracy where the interactions between people, represented as circuit nodes, were considered a conductance term. This model

was creative in its structure and enabled the team to perform an interesting analysis of the conspiracy factors. This team was selected as the INFORMS winner.

Conclusion

Among the 1,329 papers, there were many strong submissions, which made judging difficult. However, it was gratifying to see so many students with the ability to combine modeling, science and effective communication skills in order to understand such a complex problem and recommend solutions. We look forward to next year's competition, which will involve another problem in network science and hopefully, the participation of many teams of competent and passionate interdisciplinary modelers.

Recommendations for Future Participants

- **Answer the problem.** Weak papers sometimes do not address a significant part of the problem. Outstanding teams often cover all the bases and then go beyond.
- **Time management is critical.** Every year there are submissions that do an outstanding job on one aspect of the problem, then "run out of gas" and are unable to complete their solution. Outstanding teams have a plan and adjust as needed to submit a complete solution.
- **Coordinate your plan.** It is obvious in many weak papers how the work and writing was split between group members, then pieced together into the final report. For example, the output from one model doesn't match the input for the next model or a section appears in the paper that does not fit with the rest of the report. The more your team can coordinate the efforts of its members, the stronger your final submission will be.
- **The model is not the solution.** Some weak papers present a strong model, and then stop. Outstanding teams use their models to understand the problem and recommend or produce a solution.
- **Explain what you are doing and why.** Weak teams tend to use too many equations and too few words. Problem approaches appear out of nowhere. Outstanding teams explain what they are doing and why.

About the Authors

Chris Arney graduated from West Point and served as an intelligence officer in the U.S. Army. His academic studies resumed at Rensselaer Polytechnic Institute with an M.S. (computer science) and a Ph.D. (mathematics). He spent most of his 30-year military career as a mathematics professor at West Point, before becoming Dean of the School of Mathematics and Sciences and Interim Vice President for Academic Affairs at the College of Saint Rose in Albany, NY. Chris then moved to RTP (Research Triangle Park), NC, where he served for various durations as chair of the Mathematical Sciences Division, of the Network Sciences Division, and of the Information Sciences Directorate of the Army Research Office. Chris has authored 22 books, written more than 120 technical articles, and given more than 250 presentations and 40 workshops. His technical interests include mathematical modeling, cooperative systems, pursuit-evasion modeling, robotics, artificial intelligence, military operations modeling, and network science; his teaching interests include using technology and interdisciplinary problems to improve undergraduate teaching and curricula. He is the founding director of COMAP's Interdisciplinary Contest in Modeling (ICM)®. In August 2009, he rejoined the faculty at West Point as the Network Science Chair and Professor of Mathematics.

Kate Coronges is an Assistant Professor in the Department of Behavioral Sciences and Leadership and a research fellow in the Network Science Center at the U.S. Military Academy. She has a Master's in Public Health and a Ph.D. in Health Behavior Research from the University of Southern California. Kate teaches courses in social network analysis and public policy, working with cadets to apply analytic tools to understand and model complex systems, particularly as they relate to public policy issues such as energy, education, information security, and health care. Her primary research effort involves a social network study of leadership and organizational performance. She also is working on an analysis of social acceptability of automatic biometric authentication tools, social determinants of phishing security vigilance, and modeling social media data to understand how protests turn to riots. Her publications in network science include the study of education, drug addiction, DADT ("Don't ask, don't tell") policy, coalition building, and security.