Judges' Commentary: Spread of News Through the Ages

Fuping Bian
Tianjin University
Tianjin, P.R. China

Jessica Libertini
Dept. of Applied Mathematics
Virginia Military Institute
Lexington, VA
libertinijm@vmi.edu

Robert Ulman
Network Sciences Division
Army Research Office
Research Triangle Park, NC

Introduction

In 2016, the Interdisciplinary Contest in Modeling (ICM) offered a choice of three problems to student teams. Problem D, the 2016 Network Science / Operations Research problem, asked teams to explore the historical evolution of media and its role in spreading information, defining news, and ultimately shaping public opinion. The problem statement is given in the contest report earlier in this issue.

Judges' Criteria

This year's judging panel included representatives from both industry and academia with areas of expertise that included applied mathematics, mathematical modeling, network science, operations research, and engineering. Given the broad scope of the question, only a handful of papers were able to answer all of the elements of the question thoroughly; so even in the final rounds of judging, there were a few papers that did not address all elements. Ideally, we sought papers that addressed the following elements and communicated these elements clearly:

The UMAP Journal 37 (2) (2016) 145–154.

- a predictive model to capture the evolution of the flow of information throughout the ages,
- a way of defining or filtering what qualifies as news,
- validation of the predictive model using past data to compare with today's reality,
- projection of the state of societal media communication in the year 2050 using the predictive model, and
- a means of modeling and measuring the influence of media and society's networks on public and/or individual opinions.

Each paper was evaluated using a common assessment guide. We offer commentary on the components of the problem and discuss strong examples from this year's submissions.

Executive Summary

As always, the clarity of the executive summary is an important factor. The purpose of the executive summary is to prepare the reader by providing an overview of the report; therefore, it *should include a brief description of the problem, the methods used, and the overall findings or results*. Weaker executive summaries often take the form of an abstract, failing to include a description of the problem and/or a summary of the findings. The complete omission of an executive summary generally results in the elimination of an otherwise strong paper from consideration.

Developing the Model(s)

As judges read the papers, they were looking to make sure that teams presented a reasonable set of well-justified assumptions, and that those assumptions were actually used in the model. Some teams unfortunately listed reasonable considerations as assumptions, but then these assumptions were not reflected in the model. Conversely, some teams presented rather simplistic models with simplifying assumptions that were not well-explained or were not justified. The judges note that for some models, assumptions could be best explained in a dedicated section of the paper, but for other models—particularly those in which the assumptions change—the assumptions and their justifications could be clearly folded into the discussion of the development of each model.

The judges also looked for clear exposition in the development of the models. The diversity of the solutions presented this year is a testament to the fact that there is no one right way to answer these questions. Therefore, the judges were not looking for the use of any one specific modeling approach; rather, they were looking for teams that proposed a reasonable model and explained how their model could capture the dynamics of the real world.

The problem asked that teams present a way to define news and identify or filter information that qualifies as news; unfortunately, some teams failed to address this consideration, causing their papers to be eliminated from further consideration. However, for those teams who answered it, this part of the problem offered an opportunity to be very creative in developing their own metric to identify news. Some teams defined news using a temporal threshold, by tracking how long a story stayed "alive" in their time-dependent models. Other teams defined news based on the number or percentage of the population that heard the news, using a news propagation model. Some teams used a combination of time and population to set a threshold. Another approach was not only to look at the number of people who were exposed to the story, but also to use an influence model, based on repeated exposure and source credibility, to determine what percentage of the exposed population believed the story and/or were willing to spread it further.

The problem implicitly asked teams to explore the evolution of information spread as a function of time. However, some teams focused on separate explorations of how information spread within set periods of history, and the judges did not penalize this interpretation of the question. The judges looked holistically at each model presented and how that model was able to answer the question as restated by the team.

Testing and Using the Model

The judges looked for papers in which the teams validated their models before using them to address the questions. Since there was such large diversity in the models used, it is not surprising that the submissions offered a wide variety of approaches to validation. The judges were particularly impressed by papers that found reliable historical data and used it to validate their models; strong examples are offered by some of the Outstanding papers, outlined at the end of this article.

Following validation, teams then implemented their models. It was only through interpretation of their results that teams were able to present meaningful and actionable conclusions that could be used to better understand news spread throughout the ages, including the prediction of the state of the communications system in the year 2050.

In addition to validating the model before use, it was important for teams to assess the quality of their model by identifying its strengths and weaknesses. The judges often use this aspect of the paper, as well as any sensitivity analyses, as a discriminating factor, since it illustrates how well the team really understands the connections between their model and the real world that it represents.

Presenting the Results

Last but certainly not least, the judges were, as always, looking for papers that provided a very clear explanation of the work and the findings. The judges were particularly impressed with papers that were well organized and included smooth transitions between sections of the paper. Additionally, they were impressed with papers that offered a balance between verbal descriptions, mathematical equations / symbolic representation, and graphics to communicate methodologies and results.

Discussion of Outstanding Papers

This year five papers received the distinction of Outstanding. Although several of these teams started with the idea of an epidemiology model (e.g., SIR), the implementations were all different. Additionally, despite the diverse approaches and implementations, each team offered clear expositions of their process as well as very keen insights relating their model and its answers back to the real issue of the spread of news throughout the ages. Summaries of the five Outstanding papers follow.

Chongqing University, China: "Abridge the Distance between Human Minds—Research on Social Information Circulation"

As with many teams, this team started with an epidemiology model to study the spread of information. Specifically, they divide the population into four types of nodes:

- those ignorant of the news;
- those who know the news and spread it;
- those who know the news but do not spread it; and
- super-spreaders, nodes with an ability to deliver the information to a large number of people

Early in the paper, the team offers a table illustrating their interpretation of each form of mass communication through the ages, the topology of that communication network, and an itemized list explaining how properties of the model network reflect characteristics of that form of media in the real world. This team also develops a fuzzy evaluation model to filter what is news; as part of this model, they introduce the idea of an audience awareness index designed to uncover the inherent value of information.

The team also builds two different prediction models and compares their results, ultimately combining their epidemiology model with a neural network model, and using real data to validate their efforts and form the baseline for their prediction of the state of society's communication network in the year 2050.

Ultimately, the team explores public opinion and influence by starting with their existing models and adding another layer of complexity that accounts for attenuation (the idea that interest in a story wanes over time) and social strengthening (the idea that our opinions are shaped by those of our neighbors).

Rensselaer Polytechnic Institute, United States: "Characterizing Information Importance and the Effect on the Spread in Various Graph Topologies"

This team's paper uses a diffusion model over a network to model the propagation of news and a simulation is built from this model. News is divided into categories, and news events in different eras and categories are utilized to investigate the model and the underlying phenomena.

The paper defines newsworthiness as overcoming thresholds of both penetration and retention. There are several examples of newsworthy events from the 1860s to the present (e.g., President Lincoln's assassination and Osama Bin Laden's death). Media are categorized as active (such as newspapers, which a user must actively read) or passive (such as television and radio, which streams to the user little to no effort on the user's part). The paper also delineates types of stories, with categories including politics, trade, religion, consumption, and entertainment. One interesting observation is the change in emphasis of the types of news stories from politics, which was a large majority 1880s, to entertainment being the most populous in 2015.

The authors develop a regression model that takes into account the different diffusion rates within different communities of interest and apply it to the various topics. Communities of interest are created, that are interested in one topic, but some major events are of interest to all communities. A total relevance factor is given to topics to measure the interest to all communities that includes penetration and retention, mentioned before, as well as velocity. Differing effects of a news item as it crosses communities are added to the diffusion model. The use of artificial or "ghost" nodes between two real nodes is an innovative way create delays of more than one step.

To validate further and experiment with the model, the authors developed a simulation. The simulations are run for 1880 and 2015, and the distribution of nodes and stories follows the distribution found earlier, but with only politics, entertainment, and consumption represented. The model has one centralized source for all news. The model is compared to Marilyn Monroe during the period that included her death, which matched very well, but another actor's death recently (Alan Rickman) does not appear to match as well. One thing that is unclear is how relevancy is determined.

There are several sensitivity analyses. Varying the importance level gives the intuitive result that the higher importance dominates. Another varies the location of the source and shows that releasing the information from a more-connected node spreads it faster. The simulation is also used to investigate how information spreads in a more-or-less-connected portion of the network. The authors discovered for themselves the network science principle that the node with the highest connectivity is not necessarily the one that propagates news the faster, but the node's neighborhood must also be considered.

Northwestern Polytechnical University, China: "Analysis of Society's Information Networks"

The team presents a comprehensive review of information network value and influence, and then they develop a double-layer network of information flow to demonstrate the relationship between speed / flow of information vs. inherent value of information. This hierarchical framework closes the gap of the heterogeneous nature of the information network, leveraging five different time periods, with inner layers modeling the information within a region and outer layers modeling information flow from region to region.

On the graph of the inner layer, nodes and edges are attributed by the type of media, value of information, the media effect and personal subjective emotions, the flow of information is simulated with an optimization model. The outer layer combines single layer networks with the same nodes. The Reaction-Diffusion Model is applied to establish a global network to adjust the propagation probability and delay by the factor of distance.

As part of validation process, this paper applies their model to large-scale case simulations, with the support of sensitivity analysis, graphs of results, and minimized prediction error. It also validates with today's predictions, then provides the forecast in 2050 under reasonable assumptions. It's notable that the parameters for the suggested models are calculated; along with the corresponding thresholds also provided, those can be useful in terms of managing public interest and opinion, controlling the value of thresholds, and guaranteeing the quality of news.

For predicting influence in public interest and opinion, the team assumes that individual opinion is not static over time but is influenced by the spread information. A "bounded confidence model" is used to model the public interest and opinion change. With regard to the task of demonstrating factors that influence the spread of information and public opinion, the team uses interesting examples to show the geographical difference in the modeling of propagation probability by information value.

The judges appreciated the model updates and insight into the regional difference with constructing the double-layer network framework. At the end, the team recommends a course of action for the ICM and other relevant entities to adopt, supported by strong analysis and discussion.

Communication University of China, China: "How to Understand the Information"

This team uses a wide variety of methods to address the question, including an SIR model for news spread, an SN–SIR model for the spread of public opinion, a classification method for identifying news, and a hybrid learning approach for networks. The report includes plenty of good ideas that excited the judges, but at times the paper seems a bit segmented. However, the clarity of the exposition for the development of each model helped propel this paper to its final standing as Outstanding. In addition to strong exposition, the team's models are soundly based in real-world assumptions that are well-justified, and a thorough sensitivity analysis offers further insight into how their findings could be related back to the real world. This team also uses a variety of graphics to help readers visualize both their methodology and their findings.

In addition to answering the questions, the team further applies their work to social media, introducing the idea of both "regular nodes" and "hot nodes," the latter representing particularly influential members of an online community. This idea parallels the team's idea of a "hot transmit node" in a more traditional news network.

Overall, the judges felt this team used a diverse yet appropriate set of mathematical modeling tools to address this year's problem.

Huazhong University of Science and Technology, China: "Who Moved My Opinion?"

This team's paper opens with a very clear executive summary that provides an overview of their work while orienting the reader to the structure of the paper. The team starts by defining two types of communication, person-to-person vs. mass media, and then they describe how these two types of communication result in different network topologies. Further, for each type of media communication, the team develops a different network and offers a balance of mathematical rationale and real-world observations to justify each network. To examine the overall network in any one time period, the team creates a multilayered network by layering the single-layered networks for the types of media that would have been prevalent at that time.

The team also uses three states for each individual, similar to an SIR model from epidemiology. However, rather than just throw the SIR model at the problem, the team thought through the meaning of the analogous states and presents the following conditions: known, unknown, and tired, where the tired nodes have exhausted their interest in sharing the news further. The communication model is tested against real data tracking the exposure of a particular story released by the BBC. The team also attempts to predict the introduction of new forms of communication and the decline of existing modes of media by identifying and analyzing the usage trends of current and past modes, such as the telegraph and newspapers.

With a few well-justified modifications, the team then uses their information flow model to explore the networks influence on opinion, introducing and defining a range of clever elements such as powerful media nodes and stubborn minority nodes.

Although some of the writing is a bit unclear, the team does an excellent job of using graphics such as **Figure 1** to help communicate their ideas with clarity and efficacy.

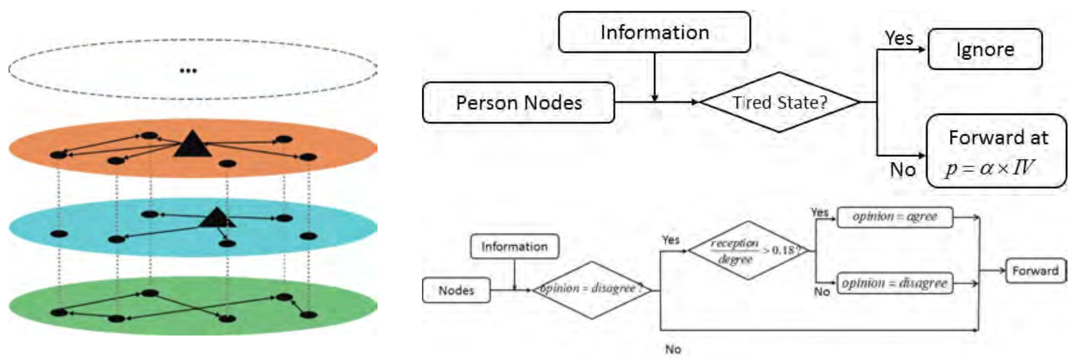

Figure 1. Three figures from the Huazhong University of Science and Technology report illustrating the layered network, the personal communication model, and the opinion-changing model.

Conclusion

Overall, the judges were excited to see the diverse approaches that teams took to address this problem, leveraging tools in a variety of fields, including network science, machine learning, and dynamical systems. There were many teams that did an excellent job of modeling at least one aspect of the problem, even if other aspects were not addressed as thoroughly or as thoughtfully. Also many teams presented a variety of models that each addressed part of the question, but only the strongest teams were able to shape all of these into a single story that flowed throughout the paper. Ultimately, the teams that advanced to the top were those that offered compelling and seamless presentations of well-reasoned solutions to most, if not all, of the questions posed.

Recommendations for Future Participants

- **Do your own work, use strong references, and cite them!** This recommendation is critical. Sadly, a record number of plagiarized papers were submitted this year. In some cases, the violation was blatant and intentional; copying whole papers or sections of papers is directly contrary to the spirit of the competition, which is about providing the opportunity to showcase creative ideas. Other cases were less egregious and perhaps less intentional, such as failing to include a citation for a downloaded image. It is important to give credit where it is due, so as you work, keep track of the resources you use, and be sure that you include both citations (inline, footnotes, endnotes) and a full bibliography.
- **Manage your time**. It can help to develop a timeline for the 96 hours, including interim deadlines and contingency plans. If something goes awry, you can always list it as a weakness of your model and discuss how you might fix the issue in the future work section.
- **Build a strong and diversified team**. Due to the interdisciplinary nature of the ICM questions and the contest format, it is beneficial to have a team whose members have knowledge and skills that complement one another.
- **Answer the problem** (or at least as many parts as you are able), and connect the ideas into a single story that flows seamlessly throughout the paper. Often the strongest papers include a unified model that can be used to address each part of the problem. Even if separate models are developed to answer various elements of the question, your team should discuss how these models are related or connected to one another.
- **Understand the context**. Depending on the question and your team's areas of expertise, you may need to spend time reading and learning about the relevant topic(s).
- **Set aside time for writing**. Your final report should be a clear paper that includes good writing, and when beneficial, thoughtfully designed figures that make the process and/or the results more easily absorbed by your audience. The most clever solution cannot be effective (in the competition or in real life) if it is not understood by others.

About the Authors

Robert Ulman received his B.S. in electrical engineering from Virginia Tech in 1984, his M.S. from Ohio State University in 1986, and his Ph.D. from the University of Maryland in 1998. He worked as a communications system engineer and research engineer at the National Security Agency 1987–2000. Since then, he has worked as the program manager for the Wireless Communications and Human Networks at the Army Research Office (ARO). At ARO, he has built a research program in wireless multihop communications networking. More recently, he has included social networking in his program, emphasizing the application of information theory and other mathematical and engineering techniques to analysis of the vast amount of social data created by the internet revolution. He is also investigating the interaction and interdependence of social and communications networks.

Jessica Libertini holds advanced degrees in both Applied Mathematics and Mechanical Engineering. Jessica is on the faculty at Virginia Military Institute, where she actively engages students in a variety of applied mathematical and educationalresearch topics, both in the classroom and beyond. In order to contextualize mathematical concepts for her students, Jessica draws heavily on her industry experiences working with General Dynamics, the Missile Defense Agency, the National Research Council, and the Army Research Laboratories. She is involved in the development of classroom materials to support the teaching and learning of mathematical modeling, and her two most active research interests are K-16 STEM education and multi-scale mathematical and network-based modeling of food and health systems.

Fuping Bian is a professor of Mathematics at Tianjin University China, where she has served as Chair of the Mathematics Department. She also was a visiting professor 1993–1994 at Florida State University and 1999–2000 at Oxford University. She has been teaching and publishing in the field of mathematical modeling since 1983, with several papers in the journal *Progress in Natural Science*, an international multidisciplinary academic journal co-sponsored by the National Natural Science Foundation of China and the Chinese Academy of Sciences.