Judges' Commentary: Cost of Privacy

Chris Arney
COMAP
Bedford, MA
arneyicm@gmail.com

Kathryn Coronges
Network Science Institute
Northeastern University
Boston, MA

Introduction

This year's policy problem asked teams to investigate an issue that has become pervasive in our society and is frequently in the news: How can we utilize personalized digital platforms while maintaining control of our privacy? [Roose 2018; Acquisti 2016].

Our reliance on social media and digital services has become interwoven in our daily lives. In many cases, access to these services has become essential for many transactions including finances, housing applications, memberships, purchases, entertainment, job searches, medical guidance, voting information, work assignments, and communication with friends, family, and colleagues.

It is understood by most digital users that the more that we reveal about ourselves, the better targeted and more personalized these services become. Despite their knowing that information about browsing habits, search topics, and physical location can provide private details about an individual, such as sexual orientation, political and religious views, race, substance use, health issues, intelligence, and personality, people are still quite willing to share their private information to gain access to these digital resources. An individual's private information (PI) and online activity clearly has utility, and this utility vastly increases when metadata is compiled from groups who are linked by friendship, common beliefs, shared location or socioeconomic status. This compilation further complicates matters of privacy, because a person's decision to share their own data will

The UMAP Journal 39 (2) (2018) 223–237.

often reveal information about others to whom they are linked. As a result, data sharing from one citizen can expose private information of groups of people and provide incentive for information leakage and hacking.

Recent revelations about Facebook data-sharing issues highlight the danger and sensitivity of data leaks and the power that comes from networked data structures [Porter 2018]. Companies and individuals are already engaged in a commodity exchange market, where individuals provide their personal information for improved services, and companies use this information for increased profit or persuasion.

However, when ICM teams were asked to explore this problem explicitly, they worked to consider how society might go about monetizing personal data. The questions they considered include: Is it important to give individuals direct control over their data? How can we estimate a dollar value to specific types of data? For specific types of people? What are the incentives for companies to keep data protected? What role does the government have, if any, in regulating this market? Given these complex privacy issues and questions, ICM teams were asked to design and model an economic system to determine the price or value of private information (PI), while recognizing the benefit and harm that exposing PI can cause.

This year's ICM policy problem can be summarized by the following: With the steady increase in electronic communications and social media, considerable PI and online activity data can be disseminated across the internet. Through these processes, PI has gradually become a commodity that garners close attention from individuals, organizations, communities, and countries. ICM teams were asked to build models to better understand the economics and privacy issues associated with personal information. Teams sought to model how tradeoffs and risks of maintaining data protection affect the price and security of these data. With the rapid development of the value of social information, the commercialization of on-line information has become an important issue. The ICM problem statement asked teams to establish a privacy pricing system to develop a stable and healthy system for personal information sales market.

The basic tasks were as follows:

- **Task 1:** Develop a price point for protecting one's privacy and PI in various applications. Consider the level of risk, the various domains of information, and the different characteristics of information for different groups of people to find the best balance for the model parameters.
- **Task 2:** Model the cost of privacy across at least three domains (social media, financial transactions, and health/medical records) and consider how different basic elements of the data (e.g., name, date of birth, gender, social security or citizenship number) contribute to your model. Your model should design a pricing structure for PI.
- **Task 3:** Establish a pricing system for individuals, groups, and entire nations. With private information becoming a commodity, is it appropriate

to consider forces of supply and demand for PI.

- **Task 4:** Consider the assumptions and constraints of the pricing model, judge the relationship between information privacy and human rights, and adjust the model to make it universal in a dynamic environment.
- **Task 5:** Consider the influence of age on the risk-benefit ratio of the population, and consider the similarities and differences between private information, personal property, and intellectual property.
- **Task 6:** Consider how the information network effect will affect the model.
- **Task 7:** Consider the impact of large-scale PI leakage. And consider whether the responsibility for data disclosure should be responsible for information disclosure.
- **Task 8:** Write a two-page policy memo to the decision maker on the utility, results, and recommendations based your modeling of this issue.

Judges' Criteria

The judges used various perspectives, including the following criteria, to evaluate and determine each paper's merits. The following are the basic judging criteria:

- Does the team have well-thought-out measures for the value or cost of privacy? Do they identify the key factors that are important as they build their model and conduct their analyses?
- Does the team build a meaningful model to determine value of personal data. Do they, for example, consider the increased value when data types can be linked—e.g., purchasing patterns and voter data is much more powerful than either one alone?
- Does the team present a market process that explains how data is exchanged and valued? Given that personal data does not have an intrinsic value, does the team address some of the important considerations—for example, a person cannot know the market value of their data if they do not know how it will be aggregated. Should individuals negotiate in the open market or is there a broker who mediates and advocates for them? Are there important time constraints that are relevant to market value? (e.g., voter data is relevant only until election day and some medical data are probably valuable long after the person is dead).
- Does the team suggest how to categorize individuals into subgroups based on a set of shared characteristics that gives them reasonably similar levels of risk across domains?

- Does the team consider shared privacy risks and if communities can influence or protect citizens' PI? Does the team consider how the pricing would change for individuals versus metadata of groups, communities, or entire nations?
- Does the team consider the role of industry or government to regulate the market? Are governments responsible to protect its citizens from exposure, making some data off limits; or are citizens responsible to become digital literate to ensure they are protected? Should the data's contribution to possible benefits to society play in how prices should be regulated?
- Does the team find, create, or use data to test their measures and models?
- Does the team develop a price point for protecting one's privacy across at least three domains (social media, financial transactions, and health/ medical records)?
- Does the team consider PI as a basic human right and consider policy recommendations based on their assumption?
- Does the team consider generational differences in perceptions of the risk-to-benefit ratio of PI and data privacy?
- Does the team model a massive data breach?
- Does the team discuss the basic elements of modeling: assumptions, measures, strengths, weaknesses, and sensitivity?
- Does the team's one-page introduction summary clearly convey their modeling work and results?
- Does the team write an insightful two-page policy memo based their policy modeling? Are they able to use their model to inform policy recommendations? Are they able to translate their modeling outcomes into policy evaluation?

Modeling Methodologies Used

Many papers used a traditional economic data approach to build their measure for the value of PI. This required the team to find and use either national economic data or company information. Some teams simulated these data, because finding reliable or complete data was a challenge within the time constraints of the contest. Once the data were analyzed, many teams used a supply-and-demand model to test and verify the functionality of their model, while many others simulated data for game-theory-type approaches. To further refine the model to make distinctions

for the various groups, subgroups, and informational domains (e.g., social, financial, medical), the teams used a diverse set of skills, including population demographics and information science in creative ways. The best teams included elements in their model that accounted for the inherent needs of humans for privacy and information security. Given the scope of the tasks and the complex nature of personal data, this problem may have been the most challenging of the six problems in this year's modeling contests, and perhaps the most demanding of all 27 ICM problems offered since the contest's inception in 1999.

One of the critical steps in this year's problem solution, and in modeling problems in general, was to make good assumptions and explain the rationale for the assumptions that were made. A team from the University of Colorado–Boulder with students Brendan Palmer, Aparajithan Venkateswaran, and Johann Kailey-Steiner, did an outstanding job on this part of the problem in their Meritorious paper "Value of Identity: Measuring the Cost of Privacy." Their assumptions went to the root of the issues for this problem:

- PI exists in a free-market environment.
- The team considered and weighed the fitness of the variables outlined in the problem.
- The force of law and legality on all PI compounds risk and value: The more PI considered, the more the value of each component of PI increases (this in essence is a non-linearity assumption for PI).
- PI technology is relatively mature but will undergo a drastic change in the near future.

By providing these assumptions, this team built a model that was both innovative and robust.

A unique approach from a team at the same school, the University of Colorado–Boulder, with students David Bloom, Lucas Laird, and William Shand, was that rather than looking at how data impacts an individual, the team viewed PI and its price strictly from a societal perspective. Their network-based model was a large random graph, with people as nodes and links between people defined through social media or other shared affiliations. With the network framing, they were able to account for how an individual's data value influences their neighbors' value, particularly in the event of a data leak. They then calculated the value of PI as the difference between benefits and damage associated with the release of the data. Their solution, entitled "Privacy Informatics," was theoretically strong and offered a powerful model. However, the complexity of the model also created challenges to determine parameters and appropriate values for distributions within the network. These kinds of tradeoffs were common in the models for this problem.

Discussion of Outstanding Papers

The best papers focused on the potential improvements in structures and processes associated with PI to develop policy recommendations. The four teams judged to be Outstanding developed an array of creative and relevant modeling techniques and analytic methods. These teams used and sometimes integrated tools in economics, network science, systems science, complex systems, and data science, to refine their models to consider all the tasks and the deeper concepts of PI. Almost all teams created a separate network model to analyze the effects of communities on the cost and value of privacy. Other modeling techniques, such as dynamical systems and game theory, also provided the capabilities to deal with issues of projecting changes in PI value over time. Summaries of the reports rendered by the four teams ranked as Outstanding follow.

Peking University: "Private Information Pricing System Based on Game Theory and Graph Theory"

This team devised a pricing system for PI based on game theory. Their innovative approach enabled them to incorporate required parameters in the payoff representation of the game. By solving for the equilibrium, they established a method for the government to set a price for PI, which can be sold if the individual agrees and the government offers adequate supervision and regulation.

The team applied a form of decay for different types of information, and then incorporated differential rates of value decline in their model, to explain the impact of a massive data breach. Like most teams, they generated a scale-free network to simulate the process of PI disclosure through the network. They did an excellent job refining their model by relaxing their initial assumptions. Some of those limiting game-theoretic assumptions were:

- people care only about their own benefit;
- there is only one step in the process (private information sold to organizations cannot be sold to other entities);
- all participants in the system have perfect information; and
- the government's sole goal is to maximize social utility.

The innovative feature in this team's model was to present the disclosure of PI and the protection of PI as two sides of the same coin—protected PI has benefits and costs that are inversely related to the opportunity cost and benefits of PI disclosure. The team developed three parameters to model the tradeoffs between costs and benefits associated with keeping PI private. Their game-theory model tracked and calculated the appropriate

price for each transaction. The demand curve of the organizations represents willingness to pay, so that the final price for PI is determined by an information-based supply-and-demand model.

The team also presented an innovative treatment of dynamics, where they considered how personal beliefs about the worth of private information change over time. While some information is perishable, with its value depreciating overtime, other information becomes invalid or worthless due to the changing nature of the world. The team used this framing to estimate the speed at which personal information declines in value based on the type of information. In modeling a massive data breach, they tailored their model to account for the different kinds of information being leaked. Their model predicted that once data breaches occur, cascades of breaches will follow, making the personal data market uncontrollable and no longer viable.

To understand the commodity effects, the team conducted simulations on an undirected scale-free graph, assuming that the relationships between people are bidirectional. Using a Barabási-Albert model to randomly generate a scale-free social network, their network used the Pigouvian tax system with subsidies and taxes to control the price behavior [Barabási and Albert 1999; Pigou 1932]. This approach produces a government tax system that compensates for the shared risks through payoffs to every individual. The negative effects are therefore eliminated using a Pigouvian tax. The team opines that private information pricing may become important in the future and recommended further modeling and study of PI.

Ludong University: "How Much is your Privacy Information Worth?"

This ICM team used the definition and classifications of privacy information to help determine the value contained in the information. They developed 26 indicators based on the principles of PI classification, including personal attributes, financial transactions, social networks, assets, and health care. They classified personal privacy into five domains:

- Citizenship: Objective information that can be used to identify a specific individual.
- Social contact: Information generated through communication within networks.
- Finance: Individual's financial information.
- Health/medical: Reflects the condition and behaviors of an individual's body.
- Personal assets: A person's economic status.

Table 1.

The Ludong University team shows how personal informational elements were quantified.

Orginal Parameter		Quantized Parameters
Personal Assets	Total Assets	Average Deposit
	Income	Disposable Income
	Expenditure	Disposable Expenditure
	Intellectual Property	Education Expenditure
Social Contact	Social Way	Social Consumption Profit
	Friends	Social Web Site Profit
	Social Signal	Profit of Communication
Finance	Financial Credit	Average Credit
	Trading Information	Average Stock Investment
	Transaction on Amount	Average Transaction Amount
	Debt	Average Liabilities
Health/Medical	Physical Health Status	Health Expenses
	Medical Insurance	Insurance Cost
	Medical History	Medical Records
	Medical Expense	Medical Expenditure
	Genetic Map	Medical Research Funds

The team used analytic hierarchy processing and machine learning on census data from China from 1997 to 2016 to determine the risk coefficient, value coefficient, and correlation degree. Their innovative approach simulated 500 transactions of individuals selling their private data, establishing prices for the information and expected return value, risk value, and relative value for each individual. They determined that the privacy information for a person is worth \$5,578 (per year).

To show the network influence on the pricing model, the team modified their initial assumptions and modeling approach:

- Use an economic-man hypothesis, where each person considers only his or her own interest.
- Use a modified small-world network model to simulate group information leaks, where members who sell information also leak personal information about other members in their neighborhood.
- Any member's sale of PI is immediately known to other members.
- Each member has the same relative impact on information disclosure.
- Information privacy is not considered a human right.

Considering dynamics, the team built a three-phase model: Development period, Mature period, and Decline period.

- In the Development period (labeled Evolutionary in **Figure 1**), the relationship between the value of privacy and the amount of demand is flexible. When the price of privacy rises, most people are willing to sell,

and when the price falls, people will likely keep their information privacy.

- In the Mature period, there is a deepening of a commodity, where institutions have established a fixed demand and the price steadies to a fixed value.
- During the Decline (labeled Winter in the figure) period, the market for PI is so saturated to the point that there is no need for a market.

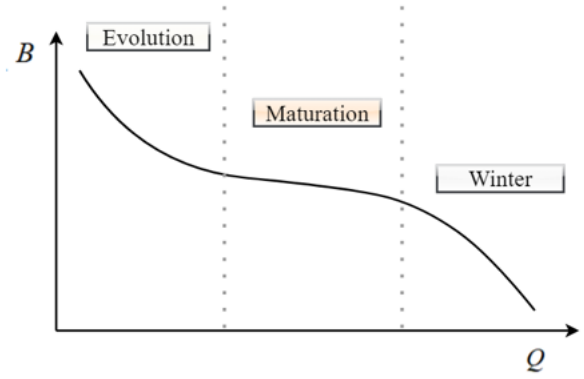


Figure 1. The team from Ludong University shows the decaying dynamics of the value of PI in the information market over time.

The highlights from the team's policy memo include:

- considering the value of information from both the information owner and the information user,
- measuring the value of information,
- considering the influence of dynamic factors,
- considering the additional risks to privacy and security when selling information, and
- advocating for disciplinary punishment for illegal behavior.

The policy memo was successful in translating components of the model into actionable policy recommendations.

University of Electronic Science and Technology of China: (INFORMS Award) "A Smart Privacy Commodity: A Quantitative Model of Dynamic Pricing Strategy"

This team developed PI price and privacy risk as two separate elements that they combined to build their model. Their PI measure considered basic characteristics from four domains: social media, financial activity, medical care, and e-commerce. They analyzed these domains from the perspective of the individual, from the community, and as a nation. Using the

analytical hierarchy process and their own rankings, they determined an appropriate weight for each of the four domains.

- They used TOPSIS (technique for ordered preference by similarity to an ideal solution) [Wikipedia 2018] to determine the value of PI for individuals.
- For the community perspective, they quantified the risk perception using generational differences, network effects, and community differences.
- For the national scope, they built the price model based on three factors: the market, policies, and culture.

Their analysis determined the average value for an individual's PI to be \$783. Finally, they tested their model through error analysis and robustness analysis.

Their approach included the following steps:

- Categorize individuals into subgroups and develop a price point for each subgroup by grouping the weights of the four domains (social media, financial transactions, health/medical records, electronic commerce).
- Account for information supply and demand processes, and personal preference.
- Capture the network effects of data sharing, to determine roles and responsibilities involving PI that exist within the community.
- Determine the liability of institutions responsible for data breaches.
- Analyze the national level from a market perspective, policy perspective, cultural perspective using a Bayesian Nash equilibrium model.

The team used a Mugglestone's approach to show social network effects [Augustin 1996]. They showed that as time passes and data are shared, people's social intersections gradually increase, and the clustering of the subgroups becomes more obvious (see **Figure 2**). Therefore, the team concluded that with the sharing of data, social network effects increase over time.

Using an auction market dynamic, the team modeled an information commodity market as shown in **Figure 3**.

These strengths of their model were that they:

- considered many factors in building their model;
- used sound mathematical methods and algorithms;
- presented an innovative approach; and
- produced adaptable models.

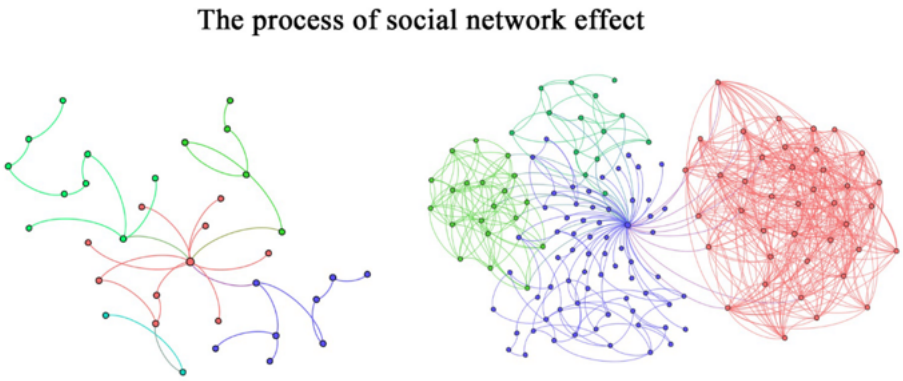


Figure 2. The team from the University of Electronic Science and Technology of China built notional networks to demonstrate the effects of communities on individual PI. The older network is on the left and over time is transformed into the more distinctive and better-connected network on the right.

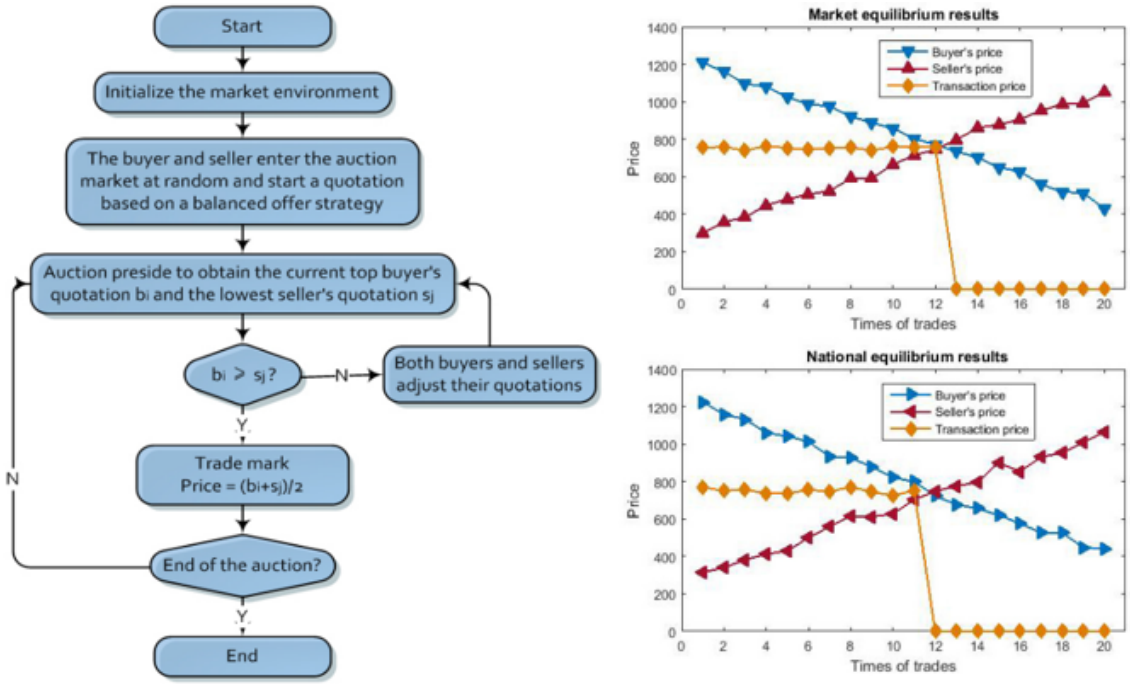


Figure 3. This figure from the University of Electronic Science and Technology of China shows their commodity auction algorithm to determine the pricing dynamics for PI.

Shanghai Jiaotang University: (Vilfredo Pareto Award) (International COMAP Scholarship Award) "PIPE: Estimate the Value of Private Information"

This team took an approach different from the other solutions. They first established that, for the most part, private data systems are currently under poor management and therefore the value of data cannot be fully extracted to benefit either its provider or its new owner. The team argued the need for a market system that not only prices and rewards data sharing, but also regulates and protects private information. They performed analysis and built a model based on a dataset they call PI-DATA. Their model, Private Information Price Estimation (PIPE), estimated the price of PI across different data domains and social subgroups.

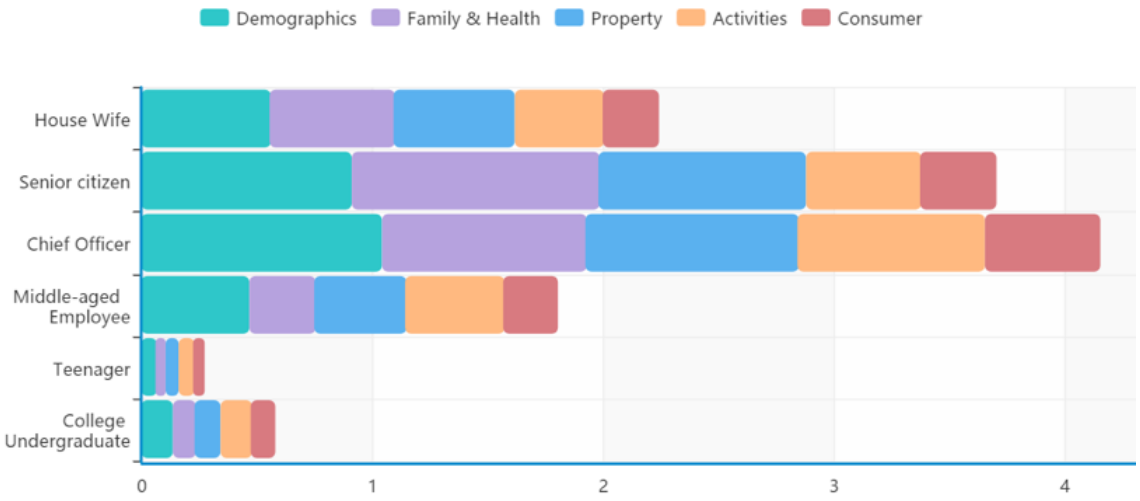


Figure 4. The Shanghai Jiaotang University team showed the value of different domains of personal data across different subgroups.

They constructed a measure for keeping PI protected and the fees that it would cost for others to possess or utilize this PI. They broadly defined PI as the record of "everything a person makes and does." The team developed a thorough list of personal data types to include:

- digital identity (e.g., names, addresses, phone numbers, demographic information, social network profile information, etc.);
- relationships to other people and organization (social media, contact list and profiles);
- communication data and logs (emails, SMS, phone calls, IM, and social network posts);
- media produced, consumed and shared (in-text, audio, photo, video, and other forms of media);

- financial data for transactions, accounts, credit scores, physical assets, and virtual goods);
- health data (health/medical records, medical history, medical device logs, prescriptions, and health insurance coverage); and
- institutional data (government, academic, and employment data).

The risks that they considered involve loss of safety, money, valuable items, intellectual property (IP), the person's electronic identity, professional embarrassment, loss of a position or job, social loss (friendships), social stigmatization, and marginalization.

In their policy recommendation, they include several important elements:

- Define the role of the government to shape commercially available products and solutions that the private sector can then leverage.
- Provide specific mechanisms for enhancing trust in all players within PI transactions.
- Integrate principles to develop new services and platforms that include PI.
- Launch an international dialogue to include nations, international organizations such as the World Trade Organization, privacy rights groups, and groups from the private sector on the issues associated with PI policies.

Conclusion

This problem presented many complex challenges that required teams to build a viable model from a very *qualitative* concept of information privacy and data security to form *quantitative* measures to represent the value of PI. In addition, translating model results to policy recommendations was an extremely difficult task.

Many teams had innovative and useful ideas for some parts of the problem, especially the tools to organize and analyze their data sets, but were unable to satisfy all the tasks required in the problem within the time constraints.

The four teams rated Outstanding seemed to fulfill those tasks the best, providing explanations for their strategies and assumptions in developing and testing their models. The judges see benefit to this kind of ICM problem because time-constrained policy modeling is a demanding task performed by public and private analysts throughout the world.

The judges congratulate all the teams who selected this problem and wish well to those ICM modelers who will become expert policy modelers providing creative analyses and making valuable recommendations to decision makers.

References

Acquisti, Alessandro, Curtis R. Taylor, and Liad Wagman. 2016. The economics of privacy (8 March 2016). *Journal of Economic Literature* 52 (2) 2016. Sloan Foundation Economics Research Paper No. 2580411. `https://ssrn.com/abstract=2580411` or `http://dx.doi.org/10.2139/ssrn.2580411` .

Augustin, Nicole, M.A. Mugglestone, and Steve Buckland. 1996. An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* 33 (2): 339–347.

Barabási, Albert-László and Ráek Albert. 1999. Emergence of scaling in random networks. *Science* 286 (15 October 1999): 509–512.

Pigou, Arthur C. 1932. *The Economics of Welfare.* 4th ed. London: Macmillan and Co.

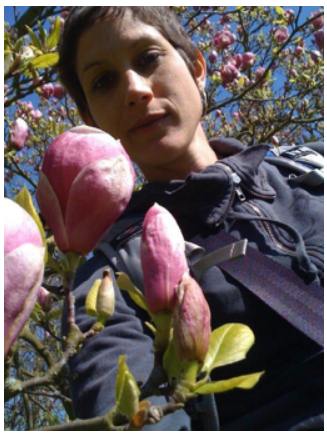
Porter, Eduardo. 2018. The Facebook fallacy: Privacy is up to you. *New York Times* (24 April 2018).

Roose, Kevin. 2018. Can social media be saved? *New York Times* (28 March 2018).

Wikipedia. 2018. TOPSIS. `https://en.wikipedia.org/wiki/TOPSIS` .

About the Authors

Chris Arney is Professor Emeritus of the United States Military Academy (USMA). His Ph.D. in mathematics is from Rensselaer Polytechnic Institute. He taught mathematics and network science at USMA, served as a Dean and acting Vice President for Academic Affairs at the College of Saint Rose in Albany, NY, and as division chief and program manager at the Army Research Office in Research Triangle Park, NC. His research has been interdisciplinary, mostly centered in dynamic modeling, cooperative systems, information networks, and artificial intelligence. Chris is the founding director of the ICM.



Kathryn Coronges received a Masters in Public Health and a Ph.D. from the University of Southern California in Human Health Behavior. She serves as the Executive Director of the Network Sciences Institute at Northeastern University. Previously, she held positions as an Assistant Professor in the Department of Behavioral Sciences and Leadership, US Military Academy, a Research Fellow in the Network Science Center at West Point, and a Program Manager at the Army Research Office. Her research interests focus on the role of social and organizational network structures, and the dynamics of these networks, in communication patterns and performance of teams, groups, and societies. She has served as the Head Judge for the ICM Policy Modeling problem for the past three years.