

Sociophysics

Frank Schweitzer

Citation: *Physics Today* **71**, 2, 40 (2018);

View online: <https://doi.org/10.1063/PT.3.3845>

View Table of Contents: <http://physicstoday.scitation.org/toc/pto/71/2>

Published by the *American Institute of Physics*

The *Physics Today* Buyer's Guide

The latest tools,
equipment and
services you need.
**Fast track your
search today!**



PHYSICS
TODAY

Shop with a more powerful search engine now.

Sociophysics



To the extent that individuals interact with each other in prescribed ways, their collective social behavior can be modeled and analyzed.

Frank Schweitzer



Twenty-four years ago Paul Krugman, who went on to receive the 2008 Nobel Prize in Economic Sciences, wrote, “Economics is harder than physics; luckily it is not quite as hard as sociology.”¹ Thirteen years ago Doyne Farmer, Martin Shubik, and Eric Smith posed the question, Is economics the next physical science? (see *PHYSICS TODAY*, September 2005, page 37). If you were skeptical then about sociology as the next physical science, you may be even more skeptical now.

A healthy skepticism regarding those two fields may indeed be better than the unhealthy optimism found in some of today’s physics publications. But between physics and the social sciences there are signs of fruitful encounters, most of which are related to the emerging field of computational social science. The trend is driven by new social data provided by engineers, who build the sensors that increasingly log our everyday lives, and by computer scientists, who build the software that harvests the data. To elucidate those developing relations, it is helpful to start with a historical perspective.

In *A Treatise of Human Nature*, Scottish philosopher David Hume (1711–76) proposed establishing a new science of man in the spirit of mathematics and physics. During the 19th century, new physical theories emerged. Electromagnetism showed that two seemingly different phenomena could be understood from a general perspective. Thermodynamics introduced a new and rather abstract concept of “systems.” French philosopher Auguste Comte (1798–1857) proposed that society follows general laws much as the physical world does. To determine the laws’ empirical basis, Belgian statistician Adolphe Quetelet (1796–1874) applied probability theory to data about humans. In his *Essays on Social Physics* (1835), he derived statistical laws for the average human based on the normal distribution. For example, he defined the body-mass index to quantify obesity. He also analyzed crime and public health. After discovering that Quetelet had appropriated the term “social physics” for his statistical approach, Comte decided to coin, for his new science of man and society, the term “sociology.”

Physics served again as a role model in the 20th century when new fundamental theories were devised. Relativity, with its revision of the concepts of space and time, and quantum mechanics, with its introduction of the uncertainty principle, both shed new light on the role of the observer and the process of observation. Modern physics had a broad impact on philosophy and the social sciences to a degree that can seem surpris-

ing nowadays. By the second half of the 20th century, the impact was no longer through general theories but through generic and abstract modeling approaches. Already during the 1940s, lattice models, later generalized as cellular automata (CA), were being used to study social segregation. The models had tunable parameters, such as migration distance and the ratio of tolerated and intolerated inhabitants in a person’s neighborhood.

The value of CA was readily apparent in its ability to simulate and visualize social dynamics. However, some CA also made it possible to conduct formal analysis. The Ising model, put forward in 1924 by Ernst Ising, was developed as an abstract spin system to explain ferromagnetism. Spins with a value of either +1 or –1 are positioned on a one- or two-dimensional lattice. Depending on the strength of the pairwise coupling constant between neighboring spins, Ising’s solution yielded ferromagnetic phases, in which spins are aligned in the same direction, or antiferromagnetic phases, in which neighboring spins are antiparallel. The generic model later became the paragon for opinion dynamics, with the positive and negative spins representing opinions. But the insights gained with respect to social phenomena were rather limited. In opinion dynamics, one tends to be interested in the conditions under which consensus is obtained (the ferromagnetic phase) or in how a stable coexistence of opinions is reached.² The voter model and other simplified models formalized that type of analysis and extended it to various topologies, including networks. But voters do not vote in those models. Rather, they copy the “opinion” of a randomly chosen spin.

Such models gratified the sociophysicist, but they did not impress the sociologist. Generic modeling approaches that replicate physics insights, such as phase transitions and scaling laws, may reveal a lot about statistical physics but little about social dynamics. Merely using physical metaphors and analogies does not make physics applicable. Noticeable exceptions were obtained only in rare cases in which physicists paid attention to existing social theories. One such example was social impact theory, which was developed by social psychologists in the 1980s to describe how individuals act as sources and targets of social influence. Underlying the theory is the concept of a social force that acts very much like a physical force. Individuals are able to persuade others with opposite opinions and support those with the same opinion, but their influence scales with social distance. When such interactions are simulated, one

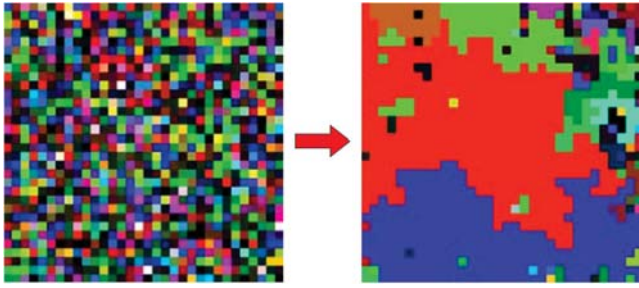


FIGURE 1. CULTURE DYNAMICS. Each agent on a two-dimensional regular lattice is characterized by a vector of features that represents its culture. Features could be cuisine or religion, whose different possibilities—Cantonese, say, or Buddhism—are termed traits. Different cultures are denoted here by different colors. The probability of an agent’s interaction with its neighbors increases with the overlap of traits. Agents are therefore more likely to interact if they already share many traits, and this interaction leads agents to become even more similar. Assigning random traits to agents at the start of a simulation (left) leads in most cases to coexisting domains of agents that share the same culture (right). Other simulations lead to monocultures. (See ref. 4.)

still observes the formation of domains with like-minded individuals, but the phenomena are much richer than in Ising-like models.³

Another example of the fruitful adoption of a social theory in sociophysics is the model of cultural dissemination, which was originally proposed in 1997 by political scientist Robert Axelrod (see figure 1). Its sociophysics version⁴ can be seen as a generalization of opinion dynamics in a Potts model, whose spins can have more than two values. The cultural dissemination model aims to incorporate social mechanisms, such as assimilation (individuals become more similar when they interact) and homophily (individuals interact more often if they are similar).

A different class of sociophysics models came into full swing during the 1970s when concepts of self-organization, the forerunners of today’s theories of complex systems, were formalized. Self-organization was seen as a universal concept: What matters for system dynamics is not the system’s elements but their dynamic interactions. Consequently, insights into the principles of structure formation in, say, the Belousov–Zhabotinsky reaction and other physicochemical systems can be generalized and extended to biological or social systems. Self-organization theory indeed found applications in sociophysics, mostly as a formal approach to social dynamics.⁵ Its applications included migration and opinion dynamics. But, as was typical of its time, it lacked a link to social data.

In the decade 1995–2005, as cheap computing power became available for modest simulations, sociophysics topics burgeoned in the physics community. Then, almost everything was modeled and simulated. Opinion dynamics, marital infidelity, sexual reproduction, the evolution of languages, the emergence of hierarchies—all those phenomena and more received sociophysicists’ attention (see references 6 and 7 for overviews). The advantage and disadvantage of those models was their simplicity. For example, in modeling how children acquire language, generative mechanisms—that is, the processes that cause the effect—were assumed rather than justified. The mechanisms’ influence and the role of certain feedback processes for the system’s dynamics could then be studied without the need to incorporate all details of the problem at hand.

Computational social science

The more recent interest of physicists in socioeconomic problems is driven in part by the availability of so-called Big Data. In the mid 1990s, physicists started to analyze Big Data from

financial markets with the same enthusiasm they had in the mid 1980s for Big Data from experiments in high-energy physics. The development of econophysics was the result. In the mid 2000s, physicists became interested in the Big Data available through the internet in general and through online social networks in particular. Much as was the case in econophysics, the early forays were preoccupied by searches for characteristic patterns in the data and universal statistical laws.

That quest in econophysics nicely echoed Quetelet’s early attempts to identify statistical laws, and it led to several interesting findings. For example, one aspect of human communication, the time interval between two consecutive messages, turns out to be described by a power-law distribution (see figure 2). The exponent seems to be universal across different communication media. Other examples of universal distributions that were uncovered include votes in elections that use proportional representation and citations of scientific publications.⁷

The findings illustrate what British economist Nicholas Kaldor (1908–86) called “stylized facts”—regularities in the social world that are robust across different observations. Physicists identified dynamic mechanisms that could conceivably reproduce such regularities but did not claim that the mechanisms capture the gist of social interactions. Still, the universality emphasized by physicists provoked economists and sociologists and raised questions about its importance and origin. What does it mean to be human if social phenomena fall into physical universality classes? And what does it mean when they don’t?

The current trends in sociophysics are closely related to what is now called computational social science, which denotes a data-driven approach to social phenomena. The data in question manifest what humans do electronically as they use mobile phones, online social networks, search engines, online banking, and so on. Sociology did not ask for that trove of data, which extends the reach of previous empirical analyses by orders of magnitude, nor was it prepared. This generated a void that is now filled by engineers who build and install more sensors and by computer scientists who gather and process ever more massive amounts of data.

Alex Pentland’s book *Social Physics*⁸ and other recent publications about the topic have little to do with physics and more to do with the analysis of Big Data. In that respect, they share the original intention of Comte’s philosophy—to build knowledge on observation and experiment. But instead of understanding the generative mechanisms underlying a phenome-

non, the focus of Big Data analysts is on regulating processes such as traffic flow, on developing apps such as Uber that make use of Big Data, and on solving problems such as predicting what customers will order online.

Despite the lack of emphasis on understanding phenomena, recent trends in Big Data have raised hopes for a new kind of social science based entirely on data processing. In 2008 the former physicist and editor-in-chief of *Wired* magazine, Chris Anderson, wrote that “faced with massive data, this approach to science—hypothesize, model, test—is becoming obsolete.” What his magazine projected instead was a petabyte age: “Sensors everywhere. Infinite storage. Clouds of processors. Our ability to capture, warehouse, and understand massive amounts of data is changing science. . . . As our collection of facts and figures grows, so will the opportunity to find answers to fundamental questions. Because in the era of Big Data, more isn’t just more. More is different.”⁹

There’s nothing wrong with Anderson’s claim that the new science is driven by data and by technology. But the most important ingredients of science are, and always have been, the research questions. Data science may help to answer some fundamental research questions, but it cannot develop such questions by itself. The practice of first collecting data and then seeing what patterns can be extracted will identify new—and mostly spurious—correlations. But it will not lead us to an understanding of causal relationships. In sociology, questions are not just about the how, but also about the why. Thus we need new types of models that embody the “reasoning” that underlies the dynamics of social systems.

Data-driven modeling

Developing such models is not just a technical challenge but also a conceptual one that physics can meet. We physicists can build on the generic understanding of complex systems that we developed in collaboration with researchers in other disciplines. Complex systems consist of a large number of strongly interacting elements, generally denoted as agents. In the tradition of statistical physics, approaches in complex systems aim to predict the collective effects that arise from the agents’ interactions. Physicists have contributed both formal methods—for example, stochastic equations to derive a system’s macroscopic dynamics—and computational approaches to model such systems. In fact, particle-based simulation methods used in computational physics have much in common with agent-based models developed not only in sociology and economics but also in computer science.

As mentioned above, most sociophysics models of the past aimed at revealing generic insights. The limited complexity of the models did not reflect the complexity of any particular social system. For that reason, they could not be calibrated and validated against real data. Big Data cannot cure the validation problem. We need models that are expressly developed with their calibration and validation against real data in mind.

Another problem, also mostly ignored in previous sociophysics models, pertains to the complexity of the agents themselves. Agents that purport to represent humans can barely be captured by up and down spins. Human decisions reflect personal preferences, social norms, and the influence of others. Accommodating those factors is not just a matter of adding degrees of freedom. Agents in socioeconomic systems are also

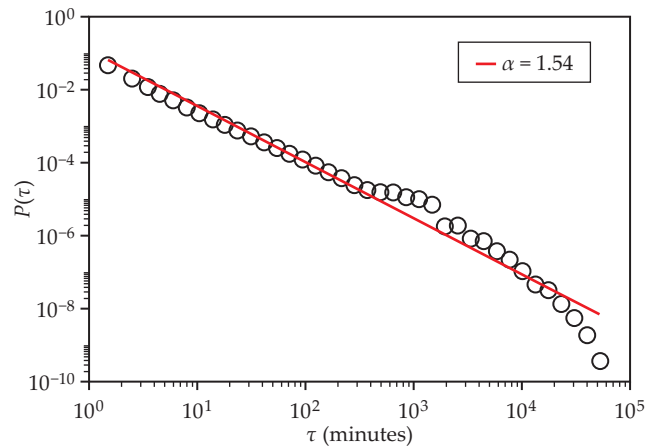


FIGURE 2. HUMAN COMMUNICATION seems to be a scale-free phenomenon. The time lapse between two consecutive messages sent by the same person, also known as the inter-activity time interval, τ , follows a power-law distribution, $P(\tau) \propto \tau^{-\alpha}$ with $\alpha \approx 3/2$.¹³ The finding is quite robust no matter what medium analyzed, whether letters, emails, or online chats (shown in the figure). The slight bulge at 10^3 minutes indicates a daily rhythm. (Adapted from A. Garas et al., *Sci. Rep.* **2**, 402, 2012.)

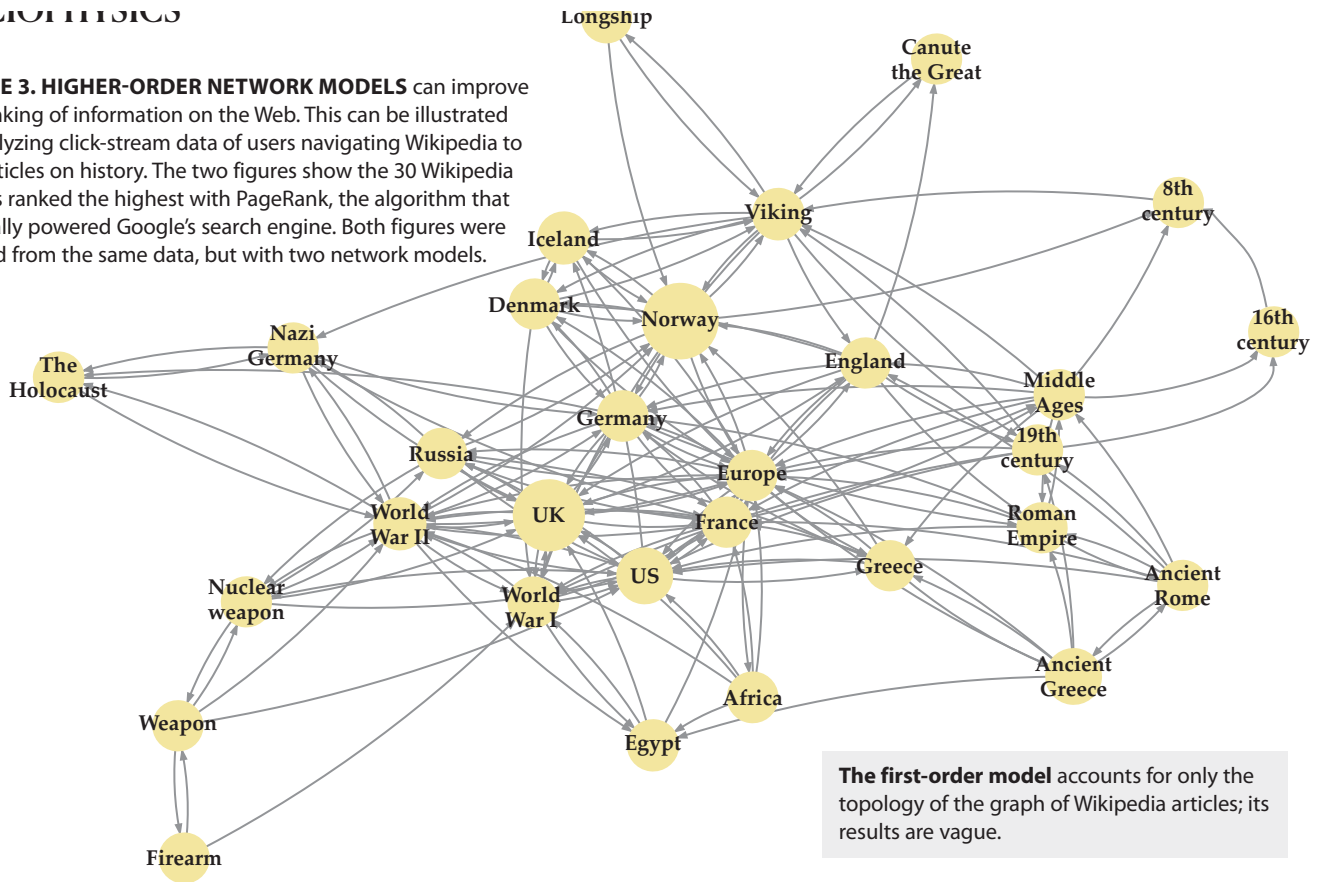
heterogeneous—they vary widely in how they interact under similar situations. They are also adaptive. They respond to incentives and to changes in the system by learning from their experiences. At the same time, they also change the system—for example, by consuming resources or by making innovations. Heterogeneity and adaptivity make the prediction of socioeconomic systems difficult.

Successful sociophysics models tend to have interfaces with both empirical data and social theories. Without the second interface, one may still find interesting phenomena and new results. But how they relate to existing disciplinary knowledge will not be clear, and the findings’ impact may be low. The first interface helps to define the problems that the models are designed to solve, most often in terms of new data that need to be explained or even created. Although machine learning approaches can, by themselves, classify the same data and make predictions, they lack the ability to model the underlying generative mechanisms.

Successful sociophysics models also bridge the micro and the macro. That is, they link interacting agents on small, local scales to dynamics on large, system-wide scales—and they do so in a concrete and testable manner. Ideally, such sociophysics models follow principles of data-driven modeling: Agents are modeled according to the standards in the relevant discipline, such as linguistics or anthropology, and the agent-based model admits the calibration of the interaction mechanisms against empirical data. The model is then validated by a quantitative comparison of the simulated system dynamics with observations.

One application of that approach is to pedestrian dynamics.¹⁰ Models of agents take into account social forces between pedestrians, preferred moving directions, and obstacles. The result is a realistic simulation of pedestrians’ collective dynamics, which can then be used to simulate escape dynamics in case of a terrorist attack or other panic, or to optimize the design of buildings and streets. Similar models describe biological

FIGURE 3. HIGHER-ORDER NETWORK MODELS can improve the ranking of information on the Web. This can be illustrated by analyzing click-stream data of users navigating Wikipedia to find articles on history. The two figures show the 30 Wikipedia articles ranked the highest with PageRank, the algorithm that originally powered Google’s search engine. Both figures were derived from the same data, but with two network models.



swarming phenomena across different branches of the animal kingdom.

Another example of successful data-driven modeling is forecasting the spread of an epidemic through, say, global aviation traffic.¹¹ Based on the calibrated model, control strategies for epidemics have been proposed. A third example is the modeling of collective emotional dynamics (see box), for which hypotheses about the emotional interactions of agents have been tested against data. The calibrated model correctly reproduces large-scale emotional influence in various online platforms.

Complex networks

Models of pedestrians, epidemics, and emotional dynamics might seem distant from electromagnetism, thermodynamics, and other branches of physics. Nevertheless, those models, like traditional physics, lead us closer to understanding real-world phenomena—in our case, social phenomena. Although physics concepts may not be generalizable to other disciplines, physics methods can contribute, in a general manner and with great benefit, to system modeling in the social sciences. The methodological contributions are not restricted to interactive systems, of which agent-based models are prominent examples. Rather, they also extend to so-called statistical models that test assumptions about data-generating processes.

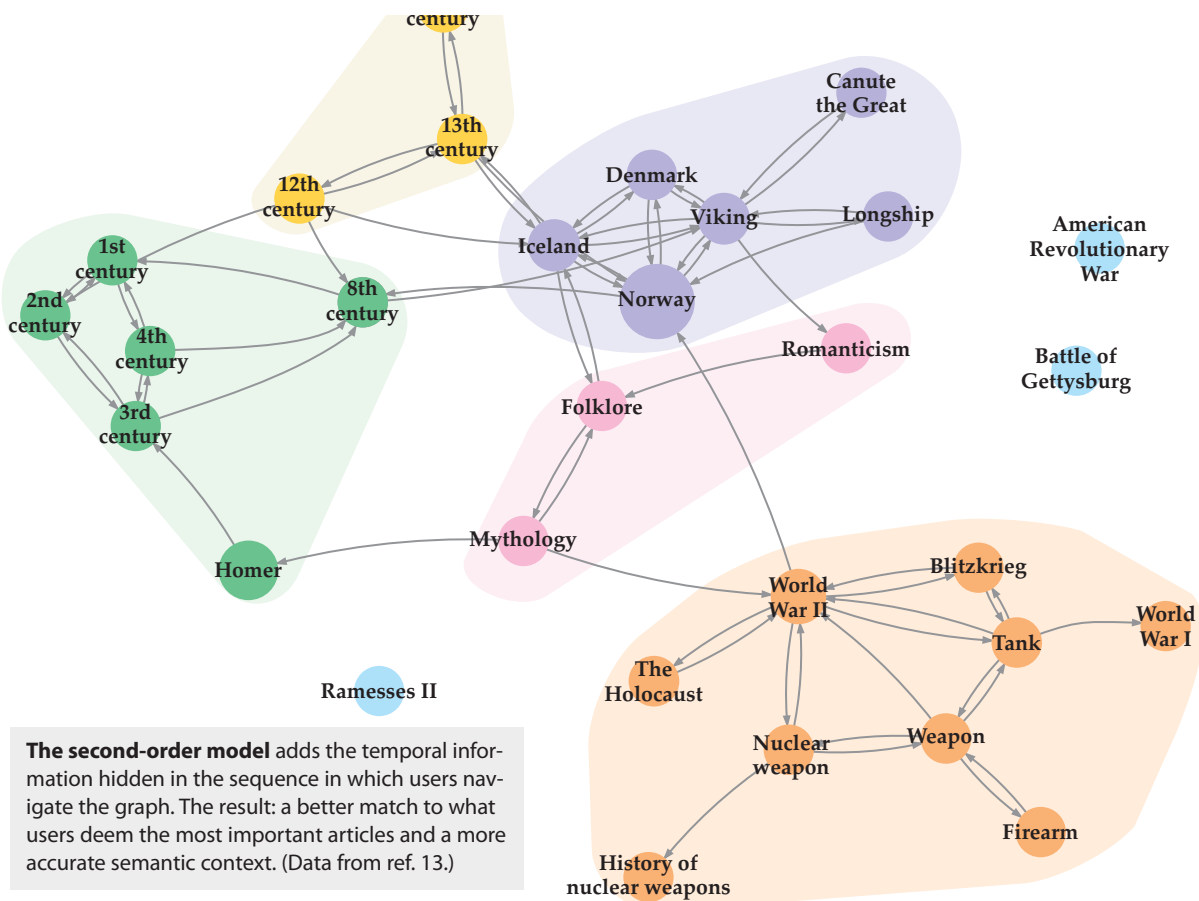
Such models belong in the field of machine learning, which became even more important as massive amounts of data became available. Although handling terabytes of data efficiently is a technical challenge, an additional, scientific challenge arises from handling modestly sized but structurally complex data sets because of the relational information they contain. Examples include online social networks of friends and family

members, citation networks among scientific papers, and navigation patterns through a patent database and other knowledge repositories. Physicists contribute information-extraction methods that go beyond those provided in computer science or the social sciences. The methods belong to another domain of sociophysics, complex networks, which are now discussed in more detail.

Complex networks are one way to represent complex systems. Agents are represented by nodes, and their interactions by links in the network. Systemic properties are then accounted for by the interaction structure—that is, by the network’s topology. Compared with agent-based models, network models have different strengths and weaknesses. The internal dynamics of the network’s nodes, the agents, are not explicitly modeled. What’s more, all types of interaction are decomposed into binary interactions between two agents. If agents act in groups larger than pairs, the applicability of the approach is limited.

On the other hand, using topology to model complex systems has led to applicable, impactful insights in the social sciences. One example is the small world network,¹² which emerges on a regular lattice topology when some links between a node and its local neighbors are reconnected to distant nodes. The rewiring creates short path lengths (connections between any two nodes) and high clustering coefficients (the links between three neighboring nodes form triangles). Because social scientists had independently discussed similar properties, they could relate their theoretical foundations to an explicit generative mechanism, the rewiring.

Another topological example is Google’s PageRank. The algorithm quantifies the importance of a given webpage based on the number and importance of other webpages that link to



it. Formally speaking, the algorithm embodies the solution to an eigenvalue problem, well known in physics, and the importance metric relates to eigenvector centrality. Because of the general nature of the eigenvalue problem, PageRank evaluates websites' relevance based on their interconnections and not on their content.

Such topological analyses require knowledge of the network, which has to be reconstructed from data. By default, the networks are time aggregated. They do not take into account, say, the sequences of other webpages that users visit before they arrive at a given webpage. However, if such temporal correlations are included, the importance ranking changes drastically and context-dependent behavior can be captured (see figure 3). Formally, the temporal conditions are calculated using higher-order Markov models, in which the order represents the persistence of memory in navigation paths. From the Markov models, we can also determine under what conditions temporal correlations can be safely neglected in reconstructing networks. Recent findings in temporal networks have considerably enhanced existing methods to characterize how people navigate Wikipedia and other social-knowledge spaces.

Sociologists have long used social network analysis to characterize the topological position of nodes in static networks. The physics contribution mainly comes with the ensemble approach. As in statistical thermodynamics, such ensembles define what topological configurations are compatible with specific constraints, the likelihood of their occurrence, and expected properties of networks. Using such methods, we can, for instance, identify which node characteristics, such as gender, common friends, and hobbies, influence the formation of links. Such results can be used to form hypotheses about

causal mechanisms that social scientists can test in the field.

Beyond disciplinary borders

What are the challenges and barriers to further advance research in sociophysics and computational social science such that all disciplines involved—physics, the social sciences, computer science, and engineering—can benefit?

Certainly, there are institutional imperatives. University education has to be developed such that curricula and academic degrees reflect the specialized knowledge needed in sociophysics. Existing curricula in the areas of network science and complex systems can serve as starting points. But sociophysics also needs high-quality journals centered around topics and problems rather than methods and disciplines. Such journals would serve as homes for scientific results that would otherwise fall between disciplinary cracks and fail to gain wide recognition. Hiring and tenure committees should also recognize the value of the extra miles that scientists with a multidisciplinary profile have traveled.

Mutual respect for the different scientific contributions each discipline provides has to be encouraged and developed. A starting point could be the admission that at this time no single discipline has all the tools, methods, theories, and knowledge needed to really understand a realm as complex as human society. Data mining, natural language processing, machine learning, and other applications of artificial intelligence are not currently among the core methods of physics. But they should be welcomed, as they give physicists access to data and to analytics that they would not ordinarily have.

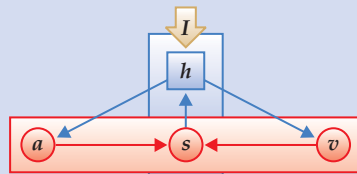
Physicists with a real interest in social phenomena should also acquire a deeper knowledge of the tremendous body of

EMOTIONAL INFLUENCE

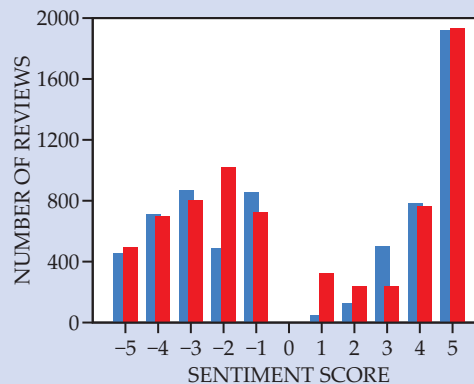
When people read reviews of books and other products on Amazon, they can choose to rate the review as helpful or unhelpful. They might also be inspired to write and submit their own review, which, in addition to carrying a rating of 0 to 5 stars, may range in sentiment from damningly negative to gushingly positive. To what extent do Amazon customers influence each other emotionally?

To address that question, my colleague David Garcia and I analyzed 1.8 million anonymized Amazon reviews of 16 670 products.¹⁴ We used a sentiment detector to automatically rate the reviews on a 10-point scale from -5 (highly negative) to $+5$ (highly positive). Zero was omitted. Then we set ourselves the challenge of reproducing the collective sentiment distributions with a Brownian agent framework.

The framework, depicted schematically on the left, incorporated a well-established



psychological model of emotional influence, the circumplex model. The emotional state of an agent is quantified by valence (v), which represents the pleasure associated with an emotion and ranges from -5 (highly negative) to $+5$ (highly positive). Arousal (a) represents the activity induced by the emotion, such as purchase or rating a review. When a exceeds a threshold, the agents express themselves with a level of sentiment (s). Agents transmit and receive emotional information (h) through social



media and other means, and they are subject to external emotional influences (I) such as coverage of products in mainstream media.

The graph on the right shows the result of running the model on one product, the book *Harry Potter and the Deathly Hallows* (2007). The blue bars are the real sentiment values for the reviews. The red values are the agent-based simulation. Our study reveals, among other things, that individual reviewers are indeed influenced by other people.

work that the social sciences have accumulated. Indeed, the lack of awareness and of understanding of their work is one of the major criticisms raised by social scientists when confronted with the papers of sociophysicists. For their part, sociologists should recognize, much more than they have in the past, their need to collaborate with researchers in other disciplines, in order to make computational science a social one. Their aversion to stylized facts and universal distributions could be overcome by formal models, jointly developed, that explain such findings based on disciplinary theories.

And realistic expectations about multidisciplinary collaboration should be established before a collaboration gets under way. It's naïve to assume that scientists from different disciplines simply fill each others' knowledge gaps to then jointly create results that define the state of the art in their area of collaboration. Success is never guaranteed, and many collaborations ultimately fail because of barriers between their scientific languages, differences in their scientific cultures, and disagreements about where to publish and publicize results. Fostering multidisciplinary collaborations should involve raising awareness of the inevitable hurdles.

Individual scientists should also be realistic in their expectations. Confronted with the challenge of turning from a method-driven to a problem-driven perspective, many sociophysicists eventually find out that their true motivation lies in physics-based methods rather than in social phenomena or data processing. As a result, the potential sociophysicist might withdraw from making the upfront investment to gather the requisite knowledge from social science and computer science. That effort comes with considerable risk of not being rewarded by social scientists, physicists, or institutions. An informed decision is paramount.

Those willing to make the effort, however, can be motivated

and guided by the increasing number of successful applications in sociophysics. They can draw inspiration from fascinating findings, sophisticated methods, and real-world problems. And they can contribute to the foundations of computational social science, which are still being laid.

REFERENCES

1. P. Krugman, *Peddling Prosperity: Economic Sense and Nonsense in the Age of Diminishing Expectations*, W. W. Norton (1994).
2. S. Galam, *Sociophysics: A Physicist's Modeling of Psycho-Political Phenomena*, Springer (2012).
3. M. Lewenstein, A. Nowak, B. Latané, *Phys. Rev. A* **45**, 763 (1992).
4. K. Klemm, V. M. Eguíluz, R. Toral, M. San Miguel, *Phys. Rev. E* **67**, 045101 (2003).
5. W. Weidlich, *Phys. Rep.* **204**, 1 (1991).
6. S. Moss de Oliveira, P. M. C. de Oliveira, D. Stauffer, *Evolution, Money, War, and Computers: Non-traditional Applications of Computational Statistical Physics*, Teubner (1999).
7. C. Castellano, S. Fortunato, V. Loreto, *Rev. Mod. Phys.* **81**, 591 (2009).
8. A. Pentland, *Social Physics: How Good Ideas Spread—The Lessons from a New Science*, Penguin Press (2014).
9. Wired staff, "The petabyte age: Because more isn't just more—more is different," *Wired*, 23 June 2008.
10. D. Helbing, I. Farkas, P. Molnár, T. Vicsek, in *Pedestrian and Evacuation Dynamics*, M. Schreckenberg, S. D. Sharma, eds., Springer (2002), p. 21.
11. L. Hufnagel, D. Brockmann, T. Geisel, *Proc. Natl. Acad. Sci. USA* **101**, 15124 (2004).
12. D. J. Watts, *Am. J. Sociol.* **105**, 493 (1999).
13. I. Scholtes, in *KDD '17—Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM (2017), p. 1037.
14. D. Garcia, F. Schweitzer, in *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust, and IEEE Third International Conference on Social Computing, CPS* (2011), p. 483. PT