Modeling and Analytics to Support Emerging International Innovation Partnerships

Jeffrey M. Keisler, Zachary A. Collier, Bilal M. Ayyub, C. Scott Dempwolf, Jacqueline MacDonald Gibson, Alan L. Porter, Vanessa J. Schweizer, Heimir Thorisson, Linbing Wang, Ming Ye, James H. Lambert, Igor Linkov

ABSTRACT

Science and technology innovation is a critical determinant of a nation's competitiveness in the global economy and a key driver of wellbeing for its residents. Building research partnerships with other nations can create and sustain a lasting competitive advantage and facilitate the international transfer of knowledge and skills. While it is tempting to continue to collaborate with established partners, emerging partners may represent a source of untapped benefit. In this paper, we discuss how modeling approaches and techniques may be used to identify innovation partnerships. We describe a number of modeling methods that can be applied in isolation or in combination to inform the partnership identification and selection process. Key findings include that there is a multiplicity of available modeling tools which can be applied, depending on how the problem is scoped with respect to partners, science domains, time frames, and objectives. We develop an objectives hierarchy which can be used to direct future decision making efforts, and we contrast the assumptions and requirements underlying various methodological approaches.

Keywords: Innovation, Predictive Modeling, Decision Modeling, Uncertainty, Problem Framing

1. INTRODUCTION

For countries to be competitive in the global economy, scientific progress and technological innovation constitute an essential national capability [1]. Nations must negotiate the uncertainties of the global economy by developing new tactics to accomplish their development goals and improve standards of living for their residents. One such tactic is finding new partners with whom to collaborate and share knowledge and resources. The ability of the US to leverage international advances in important science and technology areas supports long-term national economic development and human wellbeing goals [2]-[3].

While collaboration with established players in established fields has relatively predictable benefits, it is possible that too conservative an approach will result in missing out on opportunities with potential emerging partners in emerging fields [4]. There may be path dependencies at work when the US identifies priorities for international collaboration, where self-reinforcing mechanisms carry the US along the established trajectory, as opposed to seeking new and disruptive innovation models and partners [5]. Opportunities are potentially being missed for international collaboration with pioneering scientists in countries that differ from where the US has historically partnered.

The National Science Foundation (NSF) supports efforts aimed at international scientific collaboration in various ways, and has numerous interests in their success. In today's global environment of rapid scientific and technological change, the NSF must devise a proactive strategy to identify emerging research partners and leverage untapped intellectual resources. This concern was expressed by Dr. Arden L. Bement, former NSF Director: *"Many nations are accelerating their investments in research and development, education, and infrastructure in order to drive sustained economic growth. To continue being a global leader in S&T, we must ensure that we have access to discoveries being made in every corner of the world"* [6].

With the goal of better understanding the global science landscape, and directing attention toward building research partnerships with new targets of opportunity, a workshop was convened on 13-14 February, 2019 in Washington, DC. Participants discussed models and methods that can use data and knowledge to identify non-obvious opportunities for international collaboration, so that NSF officers can be alert to the possibilities, and monitor and leverage them.

2. UNDERSTANDING THE INTERNATIONAL COLLABORATION LANDSCAPE

We define a target of opportunity to be an identifiable class of activities that can be mapped to funding vehicles available to NSF. Models should be structured to transform obtainable data into an understanding about such targets in terms of relevant organizational and mission-focused objectives. Modeling efforts should also be directed toward generating new understanding of the global research environment in general – the dynamics of how players and fields emerge and interact and what kinds of actions can affect those dynamics. Perhaps most critical, we (and NSF) are concerned with detecting emergence of new, important, and even surprising pools of scientific activity, so that NSF and the nation are well-positioned for rich involvement.

To illustrate, Figure 1 spotlights some findings on nanotechnology research activity. A well-developed search strategy [7]-[8] retrieves over 2 million records from Web of Science. We can analyze those to tabulate traditional measures of national R&D activity (publications and citations) enriched by the addition of emergence scores. "Emergence Scores" reflect the extent of publication on frontier topics within the domain showing recently accelerating attention. Their calculation emphasizes upward trending in publication, while meeting criteria of novelty, perseverance, and community [9]. Figure 1 selects the five nations showing the greatest concentration of nanotechnology research publication that addresses such emergent topics. The surprising "emerging country" is Iran. In the early decade tabulated, Iran shows 49 papers, escalating rapidly to nearly 31,000 in the more recent decade. Taken together, publications, citations, and emergence scores help situate a country's R&D activity. A natural next step is to probe more deeply, say within Iran, about which organizations are publishing on which nanotechnology topics.

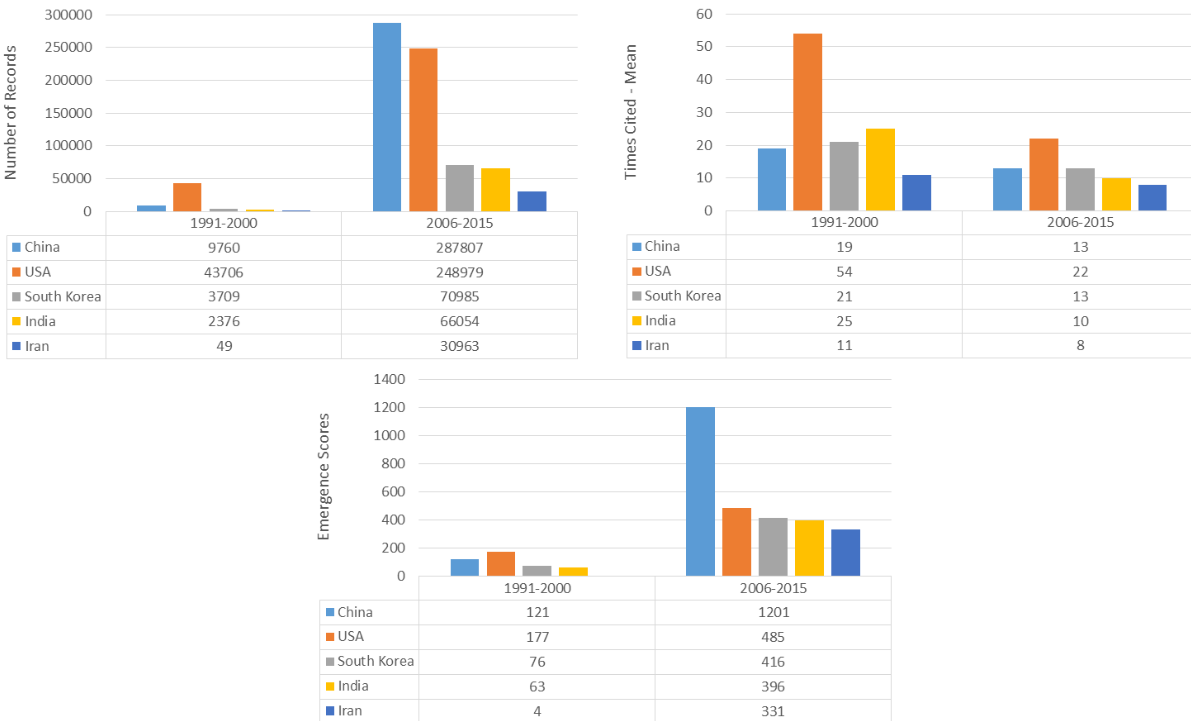

Figure 1: Nanotechnology publication data from five nations

A conceptual difficulty in understanding the international science landscape is that this innovation process can be characterized at numerous different scales. Depending on where we draw the boundaries around the innovation process model, different strategies for approaching the question of how to identify emerging partners present themselves. Namely, NSF can define several distinctions related to the area of scientific collaboration – partner, scientific domain, time horizon, and goals. In other words, with whom should NSF collaborate, on what topic, over what time frame, and why?

The first distinction relates to perhaps the most obvious question – with whom should NSF pursue partnerships? International collaborations take many flexible forms and at different scales including 1) the governmental scale of federal, state and local levels (could be agency-agency level); 2) institution scale; 3) department scale; 4) group scale; and 5) individual scale. At the largest scale, NSF might want to identify broad geographic regions or nations with which to collaborate on scientific ventures. While establishing agreements with various countries can help facilitate the innovation process, a smaller-scale perspective is likely needed to develop any tangible benefits. Identifying particular institutions, and ultimately individual scientists, provides actionable information to NSF and US-based researchers for partnering. However, national-level diplomatic relationships and agreements may be needed to facilitate small-scale partnerships. As a practical matter, individual PIs need to identify other emerging PIs, program managers need to identify emerging institutional partners for their programs at a specific scientific domain level, directorate leaders need to identify emerging institutional or national partners in their broad scientific domain, and NSF can identify emerging national partners. Larger-scale collaborations are built on or supported by smaller-scale collaborations, or are implemented by smaller-scale collaborations, and ultimately are attributed to individual collaborations. Most large-scale collaborations have signed MOUs among two or more parties. Essentially these collaborations can be classified into two major formats: the bottom-up spontaneous collaborations, and top-down diplomatic collaborations. Typically, the former serve as the basis of the latter. Deep and sustaining collaborations often involve multiple sectors including academic, governmental, and industrial sectors.

Similarly, scale becomes a question when deciding which research topics to pursue. When thinking at broad geographic scales, it might be reasonable to consider also broad topics of scientific inquiry – for example, the field of chemistry. As we narrow our scope, we may identify specific sub-fields within a broad scientific field, for example, the field of nanotechnology. Drilling down even deeper, we may then explore open questions and active research niches, such as research on carbon fiber nanotubes. Understanding exactly at what breadth and depth research is to be pursued will impact the strategy for finding research partners.

Another critical consideration is that of timeframe. There are multiple approaches for making sense of possible futures. When it comes to near-term predictions (<4 years), the odds are good that these may be a continuation of near-past trends. This means that near-term predictions could simply project empirical metrics and indicators. In contrast, long-term trends (>4 years) may be subject to discontinuities [10].

These can range from a change in the US administration to major technological breakthroughs. Long-term predictions therefore may make use of more sophisticated modeling approaches and/or scenario analysis. At any given time, most decisions taken up by the NSF are near-term. However, to be well positioned to take best advantage of emergent opportunities for partnerships, NSF should engage in periodic strategic outlook activities, scanning the long-term horizon for trends, disruptions, and emerging opportunities.

Finally, the objectives themselves for pursuing various international collaborations have significant bearing on partners, topics, and time frames. The National Science Foundation Act of 1950 defines the mission of the NSF as *"to promote the progress of science; to advance the national health, prosperity, and welfare; and to secure the national defense…"*. In terms of international collaboration, the NSF Office of International Science & Engineering states as its vision that "*NSF collaborates internationally to advance the U.S. economy, enhance our nation's security, give the U.S. the competitive edge to remain a global leader, and advance knowledge and global understanding*."[1] We note that the latter quoted language parallels the former. Further, *"NSF's international role in science and engineering is guided by its key strategic goals of Discovery, Learning, and Research Infrastructure -- that is, investing in a diverse, internationally competitive and globally engaged workforce of scientists, engineers and well-prepared citizens; investments in discovery across the frontier of science and engineering, connected to learning, innovation, and service to society; and broadly accessible, state-of-the-art and shared research and education tools."*[2]

To identify collaborations that can meet NSF's objectives, it is helpful to articulate a relevant set of goals (noting that there are many equivalent ways of structuring them). We propose here an objectives hierarchy [11]. On the top tier of the hierarchy are elements derived from NSF's *mission*. That is, we assume NSF's collaboration strategy is driven by a desire to maximize the positive economic, political,

[1] https://www.nsf.gov/od/oise/about.jsp
[2] https://www.nsf.gov/od/oise/IntlCollaborations/index.jsp

scientific/technological, and societal impacts of those collaborations. At the lower level of the hierarchy are the *means* by which collaborations can contribute to those higher-level objectives. Through a facilitated brainstorming discussion, we generated the following structured list of lower-level objectives for each main NSF objective (Figure 2). Descriptions can be found in Supplemental Information.

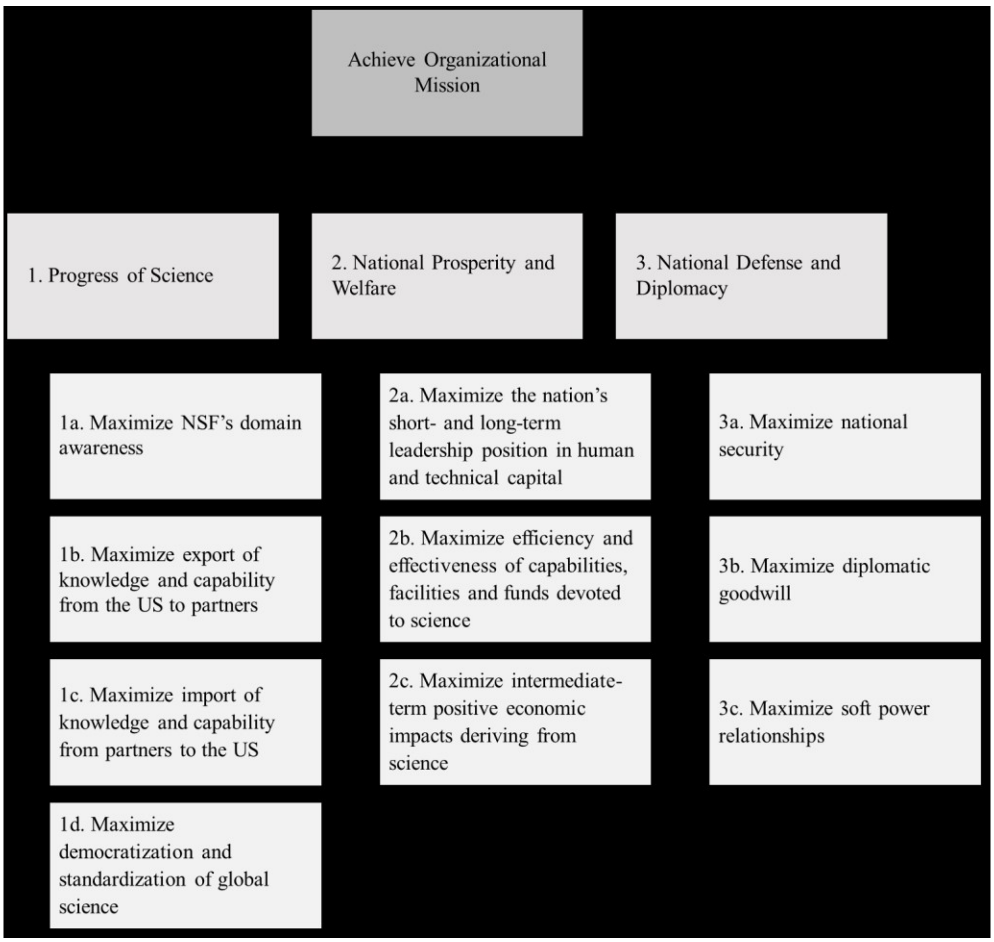

Figure 2: Innovation objectives hierarchy

3. MODELING METHODS

Many modeling approaches have potential to identify international partnership opportunities. These range from qualitative, expert-driven exercises to computationally-complex and resource-intensive techniques. From a data and modeling perspective, many of the activities in individual innovation processes are hidden from view. However, some key activities are publicly documented, and these can be explored via modeling approaches to shed light on the nature and developmental stage of innovation processes. For example, research projects may lead to publications and/or conference presentations, while patent applications document inventions. These and other documents often contain other important data, including dates, the agents involved, their roles, geographic locations, and the specific topics on which the events focused. Topics are specific and generally more detailed than scientific domains. They may be identified by classifications or keywords, but most often through text analysis of abstracts. With such data, partial models of innovation processes and ecosystems can be constructed. While incomplete, these models often represent core processes and networks in the global ecosystem.

Multiple methods may be used with publication and patent data to identify emergent targets of opportunity. Which method is appropriate depends on the nature of the question being asked and the scale and time frame of interest. Therefore, there is no single "best" approach to identify emerging partnership opportunities; rather there exists a plurality of methodologies that can be used individually or in concert to provide a picture of the overall scientific landscape. Of course, every method need not be used in every analysis – analysis ought to be commensurate with the detail required to adequately answer the question and the resources available to conduct the analysis. In the Supplemental Information, we review selected approaches and comment on their strengths and weaknesses.

Different modeling techniques rest on various assumptions and have different requirements regarding data, computational intensity, and so on. Table 1 describes the main differences between the modeling approaches.

Table 1: Comparison of modeling approaches

Modeling method	Application	Data, assumptions, requirements
Expert narrative	Project future developments not represented in available data. Learn lessons that are not suitable for any formal publication. Fill knowledge gap at the individual level.	No data requirements. Personal experience and opinions need to be evaluated. It can be used together with other approaches such as event analysis and game theory. Good elicitation techniques need to be used to ensure unbiased responses.
Cross-impact balance analysis	Macro-level cascading effects of specific partnerships/disciplines	No data requirements, similar to expert narrative. Cross-impact balance analysis provides a de-biasing technique for expert elicitation, where key system variables and their distinct alternative states may be specified by the analyst or through a participatory research approach.
Bibliometric tracking	Near-term trends in disciplines, countries, research groups and individuals	Patent and R&D publication databases, analysis and representation software. Need to coordinate between information providers, information personnel, technology analysts, researchers, and managers/users [12]

Simulation	Estimate number of publications, citations, or patents for levels of investment in particular partnerships	Historical data from which to fit probability distributions, or theoretical assumptions about probability distributions in the absence of empirical data
Decision analysis	Prioritization of opportunities, based on multiple criteria and perspectives	Data is needed on decision-maker objectives and criteria, available alternatives, performance of alternatives along criteria. May include uncertainties/probabilities.
Game theory	Implications of political relationships for science partnerships	Knowledge of player (decision-maker) objectives, which might be different for each player, the available strategies that each player can pursue, and the associated payoffs of pursuing each strategy given what the other player does
Social Network Analytics	Discover patterns in relationships between collaborators (individual, institutional, geographic, etc.)	Data is needed on collaborations such as joint publications, proposals, patents, etc.
Event Analytics	Understanding and visualizing the dynamics and patterns of different innovation trajectories which transform inputs into outcomes.	Requires a conceptual innovation process model. Temporal event data which represent innovation inputs, intermediate products, and outcomes. Data might require substantial cleaning,

		matching, transforming, and linking efforts [13]
Bayesian network models	Assessing role of international research collaborations on US economic impacts derived from science, US leadership in science, and other objectives; identifying which types of collaborations are most likely to benefit US objectives	Historical data on US international scientific collaborations by type and characteristics of collaborations; economic output data by sector; scholarly publications in different fields; international students studying in the US; other data
Structural equation models	Assessing role of international research collaborations on diplomatic goodwill, US scientific output, or other measures	Survey data from scientists, agencies, firms with knowledge of or experience in international collaborations, international relations, and other variables
Agent-based models	Discovering emergent patterns and dynamics of innovation processes	Requires understanding of underlying agent behaviors (between agents and their environment) and ability to represent behaviors mathematically

Given the strengths, weaknesses, and context-specific appropriateness of the various methodological approaches, certain approaches may have more or less utility in answering different types of questions. Figure 3 attempts to organize and recommend methods based on collaboration objectives (why) and timing (when), science domain (what), partners (who), and partner research processes and ecosystem (how). With respect to the last consideration, this is meant to shed light on how innovation processes operate within a partner country (which may be to varying degrees a black box), as well as characterizing the overall science landscape. Whereas the questions of partner and science domain are more prescriptive, the latter consideration attempts to descriptively characterize the process, mechanisms, and overall status of the scientific community.

Objective	1. Progress of Science				2. National Prosperity and Welfare			3. National Defense		
Sub-Objective	1a. Maximize NSF's domain awareness	1b. Maximize export of knowledge and capability from US to partners	1c. Maximize import of knowledge and capability from partners to US	1d. Maximize democratization and standardization of global science	2a. Maximize the nation's short and long term leadership position in human and technical capital	2b.Maximize efficiency and effectiveness of capabilities, facilities and funds devoted to science	2c. Maximize intermediate term positive economic impacts deriving from science	3a. Maximize national security	3b. Maximize diplomatic goodwill	3c. Maximize soft power relationships
Science Domain	Descriptive models (bibliometric counting, other counting methods, statistics)	Descriptive models (bibliometric counting, other counting methods, statistics), judgement or categorization about innovation	Descriptive models (bibliometric counting, other counting methods, statistics)	Cross-impact balances , expert narrative	Machine-learning, Bayesian network models	Expert narrative, cross-impact balances	Machine-learning, Bayesian network models	Game theory	Cross-impact balances	Cross-impact balances
Partners	Cluster and network analysis	Cluster and network analysis	Cluster and network analysis	Expert narrative approaches, game theory	Machine-learning, Bayesian network models	Expert elicitation	Machine-learning, Bayesian network models	Expert narrative approaches, game theory	Expert narrative approaches, game theory	Expert narrative approaches, game theory
Partner research processes and ecosystem	Regression, ELITE Modeling Analysis (mapping)	Regression, ELITE Modeling Analysis (mapping)	Regression, ELITE Modeling Analysis (mapping)	Agent based modeling				Agent based modeling	Agent based modeling	Agent based modeling

Figure 3: Modeling approaches addressing innovation objectives

4. PRACTICAL MODELING CHALLENGES AND RECOMMENDATIONS

One serious modeling challenge is uncertainty and the stochastic nature of the variables that might affect such collaborations. Relevant uncertainties should be incorporated into models wherever practical. Associating important model parameters with probability distributions and running simulations allows analysis and presentation not of a single modeled state, but an ensemble of possible states which reflect the inherent uncertainty. Similarly, some models should incorporate full uncertainty analysis using the following three general steps: (1) characterize uncertainty sources in model inputs, parameters, and structures as well as uncertainty in data used for model development and calibration, (2) run stochastic simulations for propagation of the above uncertainty, and (3) quantify uncertainty of the model outputs using various statistical approaches. When data are unavailable, we can still calculate confidence intervals for model outputs given inputs, examine the dependency of model outputs to single and multiple model inputs, and identify key model inputs and parameters. When data are available, we can evaluate goodness-of-fit between model outputs and the data, calculate biasness of model simulation relative to the data, quantify false positives/negatives and specificity, and estimate value-of-information in the data for model simulations.

Model validation involves determining the correspondence between model results and the behavior of the real world system. It can only be conducted when data are available. Given that the problem at hand is relatively new and interdisciplinary in nature, there are not sufficient data for much model development and calibration. And with such limited data, it is difficult to validate/invalidate a model or build strong evidence on usefulness of a model. Furthermore, models and data may contain errors, and models may be misused. Therefore, over-reliance upon models is itself risky.

A special challenge for identifying newcomers is how to deal with weak signals. In the above discussion of stochastic modeling, it is often the case that only moments (e.g., mean and variance) of model outputs are used for evaluating confidence intervals and calculating biasness. However, newcomers often emerge in a research field in an unnoticeable way. For example, the research leading to the 2018 Nobel Prize of Physics was published in *Optics Communications*, which had an impact factor of only 1.887, as of 2017. Prediction models for scientific discovery must not incentivize or drive policy and funding decisions to the point where they can discourage research where impact is not felt immediately or in disciplines where breakthrough discoveries have been rare in more recent time frames [14]. If development of game-changing methods/technologies is a rare event, special techniques are required to study such events, e.g. rare event sampling can be used to ensure that such events will not be missed in the stochastic modeling.

Finally, once partners are identified, it is still not clear what type of collaboration should occur. Collaborations include joint research and/or education projects, scholar and student exchanges, joint student advising, sharing research facilities, joint organizations of workshops and conferences, joint offering of classes, co-writing books, co-editing proceedings, and joint degree and dual degree programs. As all these collaborations ultimately come down to individual interactions, the importance of face-to-face meetings and social activities can never be overemphasized. Joint and dual degree programs and student exchange programs sustain and deepen relationships and collaborations. Forms and outcomes of collaboration are often tightly linked. With so many possibilities, tools such as mental modeling can be used to elicit capability gaps and potential collaboration solutions [15], and decision analytic strategy tables may be used generate thematic combinations (strategies) of collaboration efforts [16].

5. CONCLUSIONS

In a global marketplace, international collaborations should be top strategic considerations of our governments. While the default path forward is to continue collaborating with established, trusted partners, the opportunity cost of missing out on emerging newcomers can be great. Therefore, methodological approaches to synthesize existing data and generate insight should be developed and implemented. In this paper, we have attempted to identify the "tools for the toolbox," with guidance on what contexts may be more appropriate for certain techniques. The following practical takeaways can be used by analysts and policymakers to effectively develop, implement, and derive actionable insights from models:

- Numerous methodologies exist, but not all of them are appropriate for each decision context, dependent on the appropriate level of analysis, driven by scope and scale considerations. Thus, we don't advocate any single method. Rather, the choice of method is contextual, and we advocate using multiple approaches coherently to address a problem from multiple perspectives and/or at multiple scales to generate the intended insights. A robust computational toolbox, with many available techniques, is better than reliance upon a single approach.
- It is critical to properly frame the scale of the problem, namely in terms of collaboration partner, scientific domain, time horizon, and goals – the *who, what, when,* and *why*.

- As with any modeling endeavor, it is important to not rely simply upon point estimates, but quantify the uncertainty associated with the results. A number of computational methods exist for conducting such analyses.
- No model can pick guaranteed winners for prioritizing international collaboration. Rather, model outputs inform decision makers, e.g., about dependencies within scientific communities or research agendas which may be areas of opportunity for NSF to monitor or leverage. Such insights might also help NSF more critically reflect on the likely impact of research proposals received. Our hope and belief is that models can support successful decision making.

REFERENCES

1. K.H. Hughes, "Facing the global competitiveness challenge," *Issues in Science and Technology*, vol. 21, no. 4, 2005. [Online]. Available: https://issues.org/hughes/
2. J.P. Holdren, "Science and technology for sustainable well-being," *Science*, vol. 319, no. 5862, pp. 424-434, 2008.
3. M. DiChristina, "Why science is important," *Scientific American*, 2014. [Online]. Available: https://www.scientificamerican.com/article/why-science-is-important/
4. I. Linkov, B. Trump, E. Tatham, S. Basu, and M.C. Roco, "Diplomacy for science two generations later," *Science & Diplomacy*, vol. 3, no. 1, 2014. [Online]. Available: http://www.sciencediplomacy.org/perspective/2014/diplomacy-for-science-two-generations-later.
5. S. Thrane, S. Blaabjerg, and R.H. Møller, "Innovative path dependence: making sense of product and service innovation in path dependent innovation processes," *Research Policy*, vol. 39, no. 7, pp. 932–944, 2010.
6. A.L. Bement, "Testimony of Dr. Arden L. Bement, Director National Science Foundation, before the Committee on Science and Technology, Subcommittee on Research and Science Education, United States House of Representatives," 2008. [Online]. Available: http://www.hq.nasa.gov/legislative/hearings/2008%20hearings/4-2-08%20Bement.pdf

7. S.K. Arora, A.L. Porter, J. Youtie, and P. Shapira, "Capturing new developments in an emerging technology: An updated search strategy for identifying nanotechnology research outputs," *Scientometrics*, vol. 95, no. 1, pp. 351-370, 2013.
8. Z. Wang, A.L. Porter, S. Kwon, J. Youtie, P.Shapira, S.F. Carley, and X. Liu, "Updating a search strategy to track emerging nanotechnologies," *Journal of Nanoparticle Research*, (submitted).
9. A.L. Porter, J. Garner, S.F. Carley, and N.C. Newman, "Emergence scoring to identify frontier R&D topics and key players," *Technological Forecasting and Social Change*, vol. 146, no. 1, pp. 628-643, 2019.
10. R.U. Ayres, "On forecasting discontinuities," *Technological Forecasting and Social Change*, vol. 65, no. 1, pp. 81-97, 2000.
11. R.L. Keeney, *Value-Focused Thinking: A Path to Creative Decisionmaking.* Cambridge, MA: Harvard University Press, 1992.
12. A.L. Porter, "Tech mining," *Competitive Intelligence Magazine*, vol. 8, no. 1, pp. 30-36, 2005.
13. C.S. Dempwolf, and B. Shneiderman, "Event analytics for innovation trajectories: understanding inputs and outcomes for entrepreneurial success," *Technology & Innovation*, vol. 19, no. 1, pp. 397-413, 2017.
14. A. Clauset, D.B. Larremore, and R. Sinatra, "Data-driven predictions in the science of science," *Science*, vol. 355, no. 6324, pp. 477-480, 2017.
15. Collier, Z.A., Trump, B.D., Wood, M.D., Chobanova, R., Linkov, I. (2016). "Leveraging stakeholder knowledge in the innovation decision making process." *International Journal of Business Continuity and Risk Management*, 6(3): 163-181.
16. R.A. Howard, "Decision analysis: practice and promise," *Management Science*, vol. 34, no. 6, pp. 679-695, 1988.

Acknowledgements: This work was supported by the National Science Foundation under Grant 1848669 "Assessing International Collaboration Opportunities for Science and Technology Innovation: Methods and Approaches". The opinions expressed herein are those of the authors alone, and not necessarily of their affiliated institutions.