Social Media and Social Reality

Theory, Evidence and Validation

William N. Reynolds, Marta S. Weber
Least Squares Software, Inc.
12231 Academy Rd. NE #301-192
Albuquerque, NM 87111

Robert M. Farber, Courtney Corley, Andrew J. Cowell, Michelle Gregory
Pacific Northwest National Laboratories
P.O. Box 999
Richland, WA 99352

***Abstract*— Social Media provide an exciting and novel view into social phenomena. The vast amounts of data that can be gathered from the Internet coupled with massively parallel supercomputers such as the Cray XMT open new vistas for research. Conclusions drawn from such analysis must recognize that social media are distinct from the underlying social reality. Rigorous validation is essential. This paper briefly presents results obtained from computational analysis of social media - utilizing both blog and twitter data. Validation of these results is discussed in the context of a framework of established methodologies from the social sciences. Finally, an outline for a set of supporting studies is proposed.**

Keywords - social media; social science; validation; method; multimethod; blogs; twitter; graph theory; scale-free networks; supercomputing; thought leaders; opinion leaders

I. Introduction

Humans exhibit remarkable cognitive capabilities – demonstrating remarkable recall and flexibility [1]. Nevertheless, human reasoning is constrained by a number of signal shortcomings, one of the most relevant being a lack of working memory – summed up in the phrase "7 plus or minus 2" which represents the number of active elements that an average human can reason with simultaneously [2]. This shortcoming makes *complex systems* – systems with large numbers of elements and or configurations - a particular challenge for human reasoning. To help, researchers search for *simplifying frameworks* that reduce the complexity of the system to a small number of concepts that are cognitively accessible [28]. The most familiar of such frameworks are statistics, such as means and variances, which condense large amounts of data into just a few numbers.

An important question is the extent to which a framework consisting of a cognitively accessible number of elements (that is 5 to 9) can represent a much more complex system. The problems of statistics are well documented (and verbalized in expressions such as "lies, damn lies and statistics" [3]) , but such problems are a general characteristic of any simplification of a complex system [4]. Interdisciplinary studies such as social media exacerbate the challenge in creating a simplifying framework as teams composed of experts from different fields do not share "mental models" and experience communication barriers that degrade group performance [26, 27]. The essence of the argument presented here is that validation of interdisciplinary technical/social research requires the development of interdisciplinary mental frameworks that enable validating comparisons to be made. We illustrate this argument by presenting current results based on technical analysis of social media data and developing approaches for validating these results based in the social sciences methodologies.

II. Social Media As A Measurement Instrument for Social Reality

A. *Measurements on Social Media*

In this paper, we discuss validating findings assessing the impact of participants in an interaction using *Social Media*. Broadly, Social Media are defined as online communications delivered and interacted with, via text, audio and or video (*e.g.* Facebook, YouTube, weblogs (popularly known as blogs) and micro-blogs) [5]. We describe results obtained through computational analysis of social media corpora using commodity computers and the Cray XMT supercomputer [6,7]. These approaches use novel computational techniques that generalize to massively parallel computational architectures, some with near-linear scaling behavior, to handle massive amounts of data [25]. While clearly not a focus of this paper, the computational aspects of data pre-processing and the calculation of essential metrics such as graph betweeness and centrality can become intractable when presented with the volume of data from real sources unless appropriate machines and algorithms are utilized [6].

Central to our work is the concept of a *thought leader*. *Thought-leaders* are actors (nodes) in social media who have a disproportionately large impact on the underlying social collective. Broadly, if we view a post as a stimulus then "impact" is a measure of the amount of response it induces. Natural measures of impact are the number of comments a blog post receives, the number of mentions (or retweets in Twitter) of a user's micro-blog status update, or the quantity of links to and from a blogger.

A number of metrics currently used in graph analysis were investigated. Each of these metrics provided a quantitative measure of a geometrical property of nodes that can be

This research was funded via the CASS-MT program at Pacific Northwest National Laboratory. Additionally, WNR's and MW's material is in part based upon work supported by the Office of Naval Research under Contract No. N03N0001409C0612.
Please address correspondence to rmfarber@pnl.gov.

mapped onto concepts of influence in the system represented by the network [8]. Here we report measurements conducted on Twitter [9] and blog data [10] and then discuss appropriate validation techniques. The measurements explored were applied to networks – represented as graphs - inferred from the implicit or explicit interactions represented in the social media, thereby mapping the data onto the paradigm of the "social network" [11].

B. Climate Bloggers as Thought-Leaders

A set of blogs related to climate change were selected[1] based on their Technorati Authority number, as described in [7]. For each blog, a harvesting process was executed to capture the full page content of all the posts on the site. The blog post content was isolated using the Universal Parsing Agent (UPA), a front end to a regular expression engine and aligned with a previously developed blog schema. The post content was preserved in both original form and plain text in order to capture links to other content. For each blog post, comments (if enabled by the blog and present in the dataset) were also captured, with associated metadata (e.g., author, date, content).

Once the online data has been gathered, the entirety of the community posts can be expressed as links between the nodes of a graph - where post and comment authors are uniquely represented by the nodes in the graph. With this representation, the degree of outgoing edges reflects the overall number of comments received by a post author. The number of incoming links to an author represents the number of comments made by that author. The network constructed from the climate blog corpus is given in Figure 1, which shows the majority of participants in this network are comment authors.

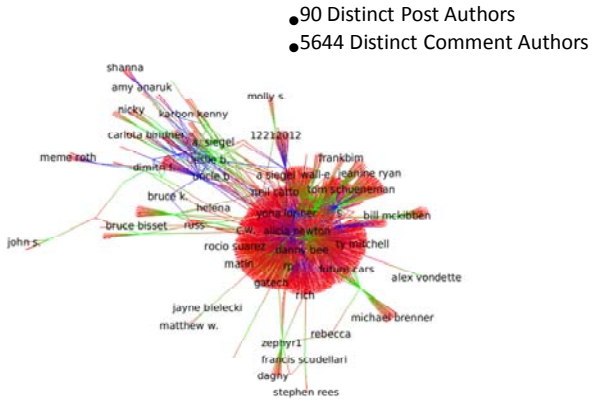

Figure 1: Author-Commenter Directed Graph of Climate Blog Corpus

[1] Those selected included: RealClimate, Climate Progress, DeSmogBlog, Celsias, It's Getting Hot In Here, Climate of our Future, Climate Ark

C. Influenza Tweeters as Information-Leaders

Complementary to thought-leaders, we measure the impact of broadcast communications to identify *information-leaders* in the micro-blogging platform, Twitter. A harvest of all public tweets[2] posted during an arbitrarily chosen timeframe of September 2009, containing the keywords flu, H1N1, influenza and swine flu were aggregated into one data set. A social network was then inferred by adding an edge to the network for every mention of a user to another tweet author (denoted by a @ symbol). Duplicate edges are discarded, leaving a social network of unique user interactions.

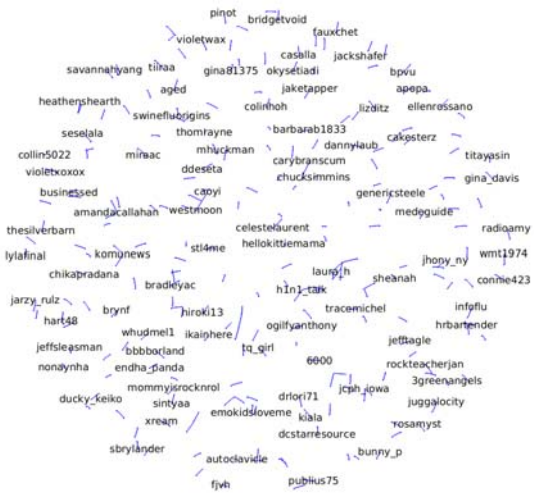

Figure 2: Information-Leaders in September 2009 H1N1 Tweet Graph

The distinction between the *thought-leader* and *information-leader* are as follows. Thought-leaders are high-impact nodes that emerge in point-to-point media such as blogs, whereas information leaders represent high-impact information sources found in broadcast oriented media. We will discuss this distinction further in the context of social science theories of thought leadership in Section IV.

D. Validating Measurements

Greater understanding and confidence can be ascribed to any results through a validation process. Specifically, the validation process should provide precise inferences about the implied underlying social reality and the extent to which these inferences are correct.

Validation is a complicated question, as it must account for the numerous failures in both thinking and measurement. The key question is this: Is the focal concept of the thought- or information-leader valid, which is to say meaningful and useful? Specifically can the focal concept be reproducibly and consistently measured using different measuring instruments? Does the concept generalize across different social systems? Is it reasonable to believe that the measurements are capturing leaders *vs*. something else? Does the concept allow relevant predictions to be made?

[2] Social media crawling provided by Spinn3r [12].

III. Theory And Validation In Social Sciences

In this section the current state of validation in social sciences relevant to this paper is reviewed. The application of these approaches to the validation of our results concerning social media is discussed.

Social scientists have long recognized the problem of developing and validating theories of complex social systems and have devoted much effort to addressing it [13,14]. Here we review a framework for validating measurement of social reality developed by Brewer and Hunter [13]. This approach is based on the idea of *multimethod*, which espouses the use of multiple approaches for representing and capturing the underlying social reality and validation through comparison. The key idea of multimethod is to measure a focal concept in multiple ways and to compare, or *triangulate*, those measurements for validation.

A central idea is that of the "focal concept" that the research is trying to measure. This is a central difficulty in complex social simulation. Validation in natural sciences, *e.g.* DoD systems simulation [15], assumes quantities being modeled are physical properties that can, at least in principle, be measured using physical instruments. In these cases, the focal concepts are self-evident. In contrast, focal concepts in social simulation are much less concrete.

Formalizing these concepts is very relevant. For example multimethod and focal concept identification is a central problem in many validation approaches, such as "Model Docking" for validating complex simulation [16]. In this approach, the results of two simulations are compared. Without a solid understanding of the focal concepts used by the simulations, inexact "apples to oranges" comparisons preclude meaningful validation. One way around this is to use efficient, structured techniques, such as morphological analysis to conduct systematic captures of expert knowledge that are tailored to the focal concepts in the target simulation, and triangulate these data against simulation [17].

There may also be limited domains of the validity of a focal concept or measurement. Using a statistical analogy, some measures such as the mean do not make sense when calculated for certain distributions. [18]. Understanding the bounds of when a focal concept is applicable is a key question. These limitations can be leveraged in a positive way. For example, a NAS report on behavioral modeling proposes a validation criterion, *validation for use*, or *validation for action*, [19] based on whether the application of a social science theory or model is valid within the context of a given application. Although the concept may in general break down, it can be valid within the specific context of an application. Naturally, such a validation scheme requires a good understanding of the failure modes of a concept or measurement.

The task then becomes deciding what to measure, assessing the validity of the measurement (in our case, graph degree and other centrality metrics), and identifying alternative measurements to verify the findings.

The most fundamental validation technique is *face validation* [13, 15], wherein subject matter experts are consulted as to the measures' representation of the focal concept. This is a form of measurement, albeit one that is subject to the vagaries of human cognition (formally known as *cognitive biases)* [20]. Recent developments in structured elicitation techniques may mitigate some of the problems with expert-based validation [17].

A more formal approach is to identify other measures of the focal concept and compare (or *triangulate)* [13] the two measurements. Empirically, social scientists have observed that measurements using similar methodologies and data-sets tend to share biases. Consequently, a recommended procedure is *multimethod* studies that rely on different data collection methodologies. Brewer and Hunter [13] identify four fundamental methods for collecting social science data: 1) fieldwork, wherein a social scientist observes the system first hand over a period of time 2) survey data, wherein subjects are systematically interviewed 3) experiment, wherein responses are elicited in an environment which controls for confounding factors and 4) nonreactive collection using passive observation (*e.g.* official archives or sensor data).

Having collected multiple measurements, the next step is to perform *Convergent Validation* , which seeks to establish that variations in measurements are due to variations in the underlying focal concept, not due to systematic biases introduced by the measurement instrument.

Another technique for validation is *Discriminant Validation* in which we verify that our measures are not correlated with additional, confounding factors. This requires a hypothesis of which factors are related or unrelated to the focal concept.

Finally, when we have identified a number of face valid, convergent, well discriminated measurements, we are said to have *Construct Validity*, which means we believe that we have meaningful, useful focal concepts and that we can measure them.

When we have several such focal concepts, we may start to construct *theories* which relate focal concepts to one another. Having reliable, convergent measures of the concepts allows hypotheses to be tested.

IV. Validating Insights Gained from Social Media Analysis Against the Underlying Social Reality

This section discusses how the preceding concepts of validation relate to the study of thought- and information-leaders in social media. Specifically, what is being measured, and how are these measurements validated?

A. *Focal Concepts*

Our analysis utilizes the focal concept of thought-information- leadership, which we describe as actors in the social network with very high degrees of incoming references from other actors. There are numerous instances of similar concepts in the literature. The business literature [21] defines a *thought leader* as someone who generates innovative ideas and promotes or shares these ideas. In other social theory, an *Opinion Leader* [22] interprets the meaning of media messages or content for "lower-end" media users. Another formulation is Hallin's [23] concept of *spheres of discourse* in media, where content which is "deviant" in the opinion of a group is excluded from the media. In this formulation, the thought leader represents the core of the consensus. Linking is the explicit action of the group to select views that are "consensus" and to exclude views that are "deviant". These definitions differ in subtle ways –which are nevertheless important for designing validation studies.

The concept of "leader" implies that there is an underlying social collective that is being led or at least influenced. In what sense does the "thought leader" influence a collective? Do they, for instance, influence thought, belief or action? Or do they provide a proxy or a representation of the collective's opinion? With the introduction of social media into the social order, traditional social science notions of leadership and social groupings appear to give way to new conceptions or at least, require new approaches and new questions. Fundamentally, what is the significant relationship between thought leaders and their social collective? Once we have made these concepts explicit, we can start to ask how might we validate our results with alternative methods of collection and analysis.

B. *Current Results from Nonreactive Measurements*

Initial observations indicate that both Twitter and blog data appear to follow a power law distribution where most of the posts are generated by a relatively small number of entities.

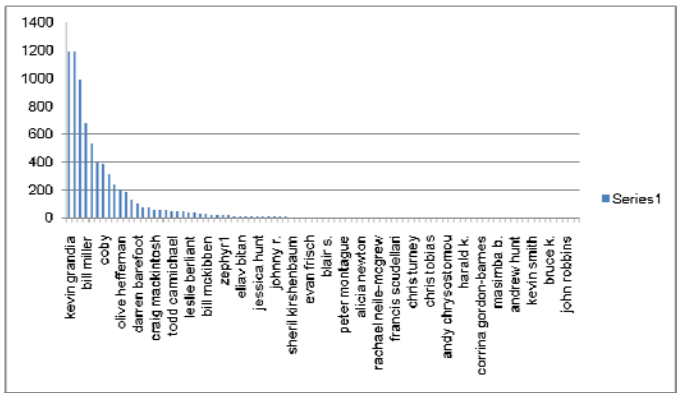

Figure 3: Power law distribution from climate blog data

This naturally suggests that that a smaller number of posting authors tend to dominate discussion and information dissemination. Our analysis concretely utilizes the posting frequency as one key metric in identifying an online leader.

A second observation is that some frequently posting entities do not add value to the online discussion. Examples include entities who express a consistently negative viewpoint (e.g. flamers) or those that essentially reiterate the opinions expressed by others. The lack of added value can be identified by noting that posts of this nature rarely generate significant numbers of responses, which is a measurable quantity. In other words, the comments per post by these entities tend to be consistently low.

A third and final observation is that some online groups and even communities appear to exist solely to reinforce a common set of opinions or group concepts. In effect, these groups act as "echo chambers" because they cooperatively say the same things. In these groups, leaders do not exist in the sense that no single entity, or small set of individual entities, consistently make posts that generate a discussion or controversy. Measuring the number of comments per post author is a way of concretely measuring and finding those entities whose posts consistently generate discussion within the group.

All these measurements appear to be consistent with the *Spheres of Discourse* theory proposed by Hallin [23].

Once the online data has been gathered it is then possible to calculate the three quantities previously discussed and sort them from highest to lowest value. Those entities that rank highly in all three measures are, by our definition, thought- or information-leaders.

Initial work has utilized expert opinion to triangulate the efficacy of this approach on the two data sets described Section II. In these and other studies, noted experts in the field were found. We consider this to be an initial face validation by subject matter experts in the field.

C. *Other Measurements*

A current shortcoming of the existing work is that we have only used a single data source: nonreactive collection using social media data. We would like to introduce other measurements of the *thought leader* concept to validate it. As mentioned, there are four general ways to collect data in social science. We outline two studies we could perform, based on survey and experimental measurement. In addition, we outline other non-reactive studies using the available social media data. By comparing and contrasting data from these studies with our current nonreactive study, we could validate and improve our current results.

1) Survey Methods A classic tool of social science analysis is the survey. Survey data could help us assess the relationship between Hypothesized Thought Leaders (HTLs) and the other

members of the social collective. For example, we could survey a random sample of responders to specific posts and produce an over-all assessment based on:

- Frequency of response to HTL posts compared to other posts.
- Extent to which information from HTL adjudged to be accurate, relevant, timely.
- Extent to which responder regarded an HTL as a primary or a secondary source of information and ideas.
- Extent to which responder's opinions are influenced by HTL.
- Determine the frequency which a responder sought the position of HTL in the course of determining his own view or position.
- Determine the number of times the responder's position or view was changed by that of HTL
- Determine the number of times the responder disseminated post of HTL.
- Determine whether a responder would choose to read an HTL's future posts.
- Compare these variables in responses to non-HTL posts.

2) Experimental Methods Experiment is another method from the social science toolbox focused on isolating cause and effect. This is a promising approach to develop in-depth understanding of how HTL's influence others. The following is one possible experiment:

1. Select a sample of HTL responders plus a control group (adjusting for potential biases such as willingness-to-participate).
2. Introduce double-blind topics of interest to both groups.
3. Establish a baseline by having the groups score selected topics for accuracy, legitimacy, agreement with existing views and/or other variables of interest.
4. Introduce the HTL position in mix of topics. Alternatively, introduce similar positions from sources other than the HTL (if we are interested in discriminating whether the HTL is a primary or secondary source of information).
5. The groups score selected topics, as above.
6. After exposing groups to posts from a collection of HTLs, allow groups to choose which HTL's they would subsequently read.
7. Analyze the group responses for correlation between the HTL/primary position and topic scoring.

3) Additional Studies Based on Available Nonreactive Data

a) Grouping – Provide an explicit notion of what we mean by group, given other associations we have in the micro data.

b) Content Analysis –Ascertain the opinions of the various groups by examining the content generated by the members of the groups and check for shared properties. These same opinions can be used to determine the extent to which ideas originate with thought leaders as opposed to elsewhere.

D. Triangulating The Different Measurements

Assuming the studies described in the section B have been performed, this new information can be used to validate results by:

1) Convergent Validation Ascertaining the extent to which the two sets of measurements agree. The original nonreactive collection indicates that responders are more likely to respond to a post by a thought leader than that of another poster**.** The new information will confirm or falsify this finding.

a) Utilize the *survey* data to corroborate the findings of the nonreactive and experimental measurements. Determine if responders conciously seek out the opinion of thought leaders. Determine if responders opinions are influenced by thought leaders.

b) Determine if the *experimental* data corroborates the findings between the nonreactive and survey measurements. Investigate whether the opinions of thought leaders change the opinions of responders, or do whether they reinforce existing views. Determine if these findings are consistent with the self-reported influence of thought leaders on responders from the survey data.

2) Discriminant Validation Given the several different media theories, we use the new data to perform cross validation discriminate which theory is operational in a given social system. One key question is whether a thought leader is an originator of ideas (per [21]) or is a disseminator of existing ideas (per [22]), or is a gatekeeper of acceptable discourse (per [23]). One way to discriminate this effect in experiment would be to compare the impact that an HTL's formulation of another's idea as compared to the original idea. We could extract similar information from survey data – do responders perceive the value of a thought leader in terms of their restatement of existing ideas or because thought leaders are generating novel ideas of their own? Survey and experiment could resolve whether responders more or less likely to seek out the posts of posters that they agree or disagree with. This could be further compared to additional nonreactive content analysis – can we trace the propagation of ideas and identify where new concepts originate? Can we understand the information foraging dynamics that responders use, and how they are influenced by thought leaders.

3) Construct Validation The proposed procedure elucidates the complexity of the problem of understanding and measuring social systems. A critic could legitimately complain that the different proposed studies each introduce biases of their own, and are consequently suspect – for example, surveys will suffer from bias of self-reporting. How will a responder be able to unpack whether an idea was introduced to him from one source or another? This critique is accurate, and exactly the point. Nevertheless, there is a chance that we will be fortunate, and that the measurments of some attributes of

our thought leaders will converge – which is to say will be confirmed by multiple agreeing measurements. Such attributes will be valid *constructs* of a social theory. If we are fortunate enough to identify such attributes, we can go forward with real confidince that we are capturing social reality.

4) Theory of Social Media and Predictive Validation Having identified valid constructs, we would be in a position to identify (or develop) the appropriate *theories* that describe the causal linkages between the constructs can be created. Such theories are operationally useful, in that they can be used to make predictions about the social collective. Successful prediction is one of the strongest forms of validation [24]. For example, if we identify that the operational theory is "spheres of influence" theory, then we could start to predict the potential of an individual to become a thought leader, or the potential of a given piece of information to be propagated through the social collective.

V. CONCLUSION

Social Media provide an exciting and novel view into social phenomena. Further, the computational capabilities of supercomputers such as the Cray XMT allow researchers to probe amounts of social media data that were previously computationally intractable. In combination, the data and current computational capabilities open new vistas for research.

This enthusiasm must include the realization that social media are distinct from the underlying social reality. Essentially, any conclusions drawn from such analysis must be rigorously validated. Towards this end, we have reviewed results obtained from computational analysis of social media - specifically blog and twitter data. Further, we described the challenges of validating conclusions drawn from computational analysis. Based on a review of established methodologies from the social sciences an outline for a set of supporting studies, based on established research methodologies, has been proposed to further validate computational based results.

REFERENCES

[1] Gigerenzer, Gerd and Todd, Peter M. and ABC Research Group *Simple Heuristics That Make Us Smart*, Oxford University Press, 2000.

[2] Miller, George A. "The Magical Number Seven, Plus or Minus Two", *The Psychological Review* 63, 87, 1956.

[3] Attributed to Benjamin Disraeli

[4] Edmunds, Bruce. 1991. *Syntactic Measures of Complexity*. Doctoral thesis, University of Manchester, Manchester, UK. Available online at http://bruce.edmonds.name/thesis, last accessed 26 March 2010.

[5] Surowiecki, James, *The Wisdom of Crowds*, Anchor, 2005.

[6] Ediger, David *et. al.* "Massive Social Network Analysis: Mining Twitter for Social Good", Proceedings of the 39th International Conference on Parrallel Processing, San Diego, CA September 13-16, 2010.

[7] Reynolds, William N., Farber, Robert M., Corley, Courtney, Cowell, Andrew and Gregory, Michelle "Novel Metrics for Identifying Thought Leaders in Social Media", *in preparation.*

[8] Hanneman, Robert A. and Mark Riddle. 2005. *Introduction To Social Network Methods*. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/ , last accessed 3/23/2010).

[9] http://www.twitter.com last accessed 26 March, 2010.

[10] Corley, C.D.; Cook, D.J.; Mikler, A.R.; Singh, K.P. Text and Structural Data Mining of Influenza Mentions in Web and Social Media. *Int. J. Environ. Res. Public Health*, 7, 596-615, 2010.

[11] Newman, Mark "The Structure and Function of Complex Networks", SIAM Review, Vol. 45,No . 2 pp . 167–256, 2003.

[12] Spinn3r http://spinn3r.com/ last accessed 26 March, 2010.

[13] Brewer, John & Hunter, Albert. *Foundations of Multimethod Research: Synthesizing Styles*. Thousand Oaks, CA: Sage Publications, 2006.

[14] H. Russel Bernard, *Research Methods in Anthropology*, Alta Mira, 2006.

[15] *VV&A Recommended Practices Guide*. http://vva.msco.mil/, last accessed 25 March, 2010.

[16] Axtell, R., R. Axelrod, J.M. Epstein, and Cohen, M.D., "Aligning Simulation Models: A Case Study and Results." Computational and Mathematical Organization Theory, 1(2): 123-142, 1996.

[17] Reynolds, William N."Breadth-Depth Validation of Complex Systems Modeling and Simulation", this volume.

[18] Applebaum, David (December 2004), "Lévy Processes—From Probability to Finance and Quantum Groups", *Notices of the American Mathematical Society* (Providence, RI: American Mathematical Society) **51** (11): 1336–1347, 2004.

[19] Zacharias, Greg L, Jean MacMillan, and Susan B. Van Hemel, eds. 2008. *Behavioral Modeling and Simulation: From Individuals to Society*. Committee on Organizational Modeling from Individuals to Societies, National Research Council of the National Academies. Prepublication copy, uncorrected proofs. Washington, D.C.: National Academies Press.

[20] Kahnemann, Daniel, Slovic, Paul and Tversky, Amos, eds. *Judgment Under Uncertainty*, Cambridge University Press, 1982.

[21] Kurtzman, Joel *Thought Leaders: Insights on the Future of Business*, Jossey-Bass/Simon & Schuster, San Francisco, December 1997.

[22] Katz, Elihu, Lazarsfeld, Paul Felix, *Personal Influence: the Part Played by People in the Flow of Mass Communications*, Transaction Publishers, 2005.

[23] Hallin, Daniel C., *The Uncensored War: The Media and Vietnam*, Oxford University Press, 1986.

[24] Lipton, Peter "Testing Hypotheses: Prediction and Prejudice", *Science*, 14 (307). No 5707 pp. 219-221, 2005.

[25] Madduri, K., Bader, D.A., J.W. Berry, and Crobak, J.R., "An Experimental Study of A Parallel Shortest Path Algorithm for Solving Large-Scale Graph Instances", *Workshop on Algorithm Engineering and Experiments (ALENEX)*, New Orleans, LA, January 6, 2007.

[26] Johnston, Rob, *Analytic Culture in the U.S. Intelligence Community: An Ethnography*, Center for the Study of Intelligence, Central Intelligence Agency, Washington DC. Chapter 5, 2005.

[27] Kerr, Norbert L. and R. Scott Tindale. 2004. Group Performance and Decision-Making, review article in Annual Review of Psychology, vol.55: 623-55, 2004.

[28] Pierce, C. S., "Deduction, Induction and Hypothesis", *Collected Papers*, Hartshorne, Charles and Weiss, Paul (eds.), Harvard University Press, Cambridge, MA, 1931-1935.