Future Directions of Network Science

A Workshop Report on the Emerging Science of Networks

September 29–30, 2016

Kate Coronges, *Northeastern University*
Albert-László Barabási, *Northeastern University*
Alessandro Vespignani, *Northeastern University*

Prepared by Kate Klemic Ph.D. and Jeremy Zeigler
Virginia Tech Applied Research Corporation

VT-ARC
Virginia Tech
Applied Research Corporation

PREFACE

Over the past century, science and technology have brought remarkable new capabilities to all sectors of the economy; from telecommunications, energy, and electronics to medicine, transportation, and defense. Technologies that were fantasy decades ago, such as the internet and mobile devices, now inform the way we live, work, and interact with our environment. Key to this technological progress is the capacity of the global basic research community to create new knowledge and to develop new insights in science, technology, and engineering. Understanding the trajectories of this fundamental research, within the context of global challenges, empowers stakeholders to identify and seize potential opportunities.

The Future Directions Workshop series, sponsored by the Basic Research Office of the Office of the Assistant Secretary of Defense for Research and Engineering, seeks to examine emerging research and engineering areas that are most likely to transform future technology capabilities.

These workshops gather distinguished academic and industry researchers from the world's top research institutions to engage in an interactive dialogue about the promises and challenges of these emerging basic research areas and how they could impact future capabilities. Chaired by leaders in the field, these workshops encourage unfettered considerations of the prospects of fundamental science areas from the most talented minds in the research community.

Reports from the Future Directions Workshop series capture these discussions and therefore play a vital role in the discussion of basic research priorities. In each report, participants are challenged to address the following important questions:

- How might the research impact science and technology capabilities of the future?
- What is the possible trajectory of scientific achievement over the next 10–15 years?
- What are the most fundamental challenges to progress?

This report is the product of a workshop held September 29–30, 2016 at the Basic Research Innovation and Collaboration Center in Arlington, VA on the Future Directions of Network Science. It is intended as a resource to the S&T community including the broader federal funding community, federal laboratories, domestic industrial base, and academia.

Innovation is the key to the future, but basic research is the key to future innovation.

—Jerome Isaac Friedman,
Nobel Prize Recipient (1990)

Executive Summary

As the world becomes more globally connected, it is increasingly defined by networks. Through scientific approaches, our ability to quantify underlying factors that drive these networks has vastly improved. With roots in physical, information, and social sciences, network science provides a formal set of methods, tools, and theories to describe, prescribe, and predict dynamics and behaviors of complex systems. Despite their diversity, whether the systems are made up of physical, technological, informational, or social networks—they share many common organizing principles and thus can be studied with similar approaches. With this powerful framework, we have discovered how networks form, grow, transform, dissolve, evolve, learn, coordinate, converge, and behave collectively; how they facilitate the flow and the spread of information, behaviors, resources, and disease; how knowledge transforms a network and how networks transform knowledge; what it means to be resilient, healthy, or optimized; what control mechanisms drive them; and how we can intervene to disrupt or rehabilitate networks.

On September 29–30, 2016, 25 distinguished network science researchers gathered for the Future Directions of Network Science Workshop in Arlington, VA to assess the current state of this emerging field. The diversity of participants reflects the truly interdisciplinary nature of network science—domain expertise included mathematics, physics, computer science, biology, sociology, epidemiology, population health, and communication. The goal of the meeting was to characterize major challenges, identify important application areas, and map the trajectory of the research over the next 5, 10, and 20 year horizon. This report summarizes the major insights and themes that resulted from the two-day workshop. The greatest impacts are expected for five application domains:

- Group decision-making
- Personal and population health
- Biological systems and brain
- Socio-technical infrastructure
- Human-machine partnerships

The workshop participants considered the ways that network science is making or has the potential to make substantial impact on innovations and advances in the future. They discussed the gaps in the foundational basic science of networks, particularly those research areas whose resolution could lead to substantial advances in knowledge and technological excellence over the next two decades. One of the unforeseen contributions of the meeting was to foster discussion on the common challenges across disciplines and across applications. What emerged was a set of methodological, data-related, and theoretical topics that represent the fundamental features and unique technical capabilities of network science. They identified specific technical areas within the following categories:

- **Mathematics and computation:** methods for modeling system-level processes that take place across multiple dimensions and acting at different temporal and spatial scales.
- **Data analysis and processing:** methods for data visualization, network inference techniques, and tools to navigate and synthesize network data.
- **Theory and mechanisms:** formalization of underlying forces and factors that drive network processes, such as diffusion, control, and coordination.

The participants outlined several other important issues for the network sciences community, including the importance of interdisciplinary research efforts, high performance computing infrastructure, and the ethics around data sharing. Finally, they saw the need for solidifying a core set of educational and focus areas for network science that will establish the field as a science of its own, rather than a collection of tools and concepts extracted from other disciplines.

The workshop participants were optimistic that the trajectory for network science will help society meet the challenges of the rapidly evolving world and provide much needed insights for our understanding of its many interconnected systems.

> "With roots in physical, information, and social sciences, network science provides a formal set of methods, tools, and theories to describe, prescribe, and predict dynamics and behavior of complex systems."
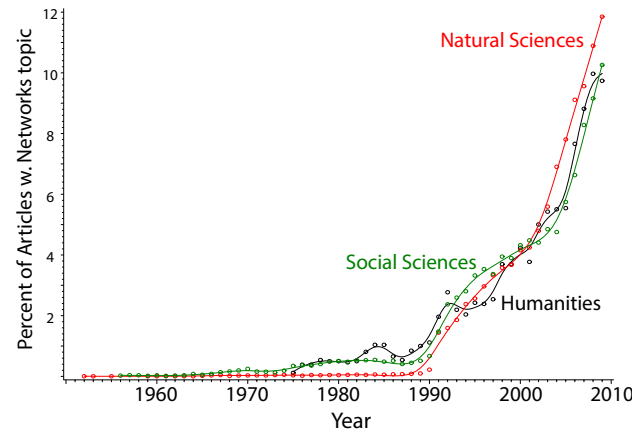
Introduction

The modern world relies on a collection of increasingly interdependent cyber-physical, socio-technical, and socio-ecological networks. Each network is independently complex with emergent properties and a large number of degrees of freedom. Network science offers a powerful conceptual and practical framework for evaluating and modeling these complex, natural, technological, and social systems. The network paradigm also provides a unique opportunity to integrate different types of elements (social, technological, natural, etc.) over time and space.

Network science is built on theories and methods from multiple disciplines, including: graph theory and matrix algebra from mathematics; statistical mechanics from physics; data mining and information visualization from computer science; inferential modeling from statistics; and social structure and group dynamics from sociology. With these computational tools, mathematical methods, and theories, network science enables characterization and prediction of complex networked systems using a common lens, rooted in the insight that despite their diversity of behavior, the structure and dynamics of different kinds of networks share many common organizing principles and thus can be studied using similar approaches.

Advances in network science have emerged due to numerous converging factors: improved mathematical capabilities for multi-layer and dynamic networks; access to human-based data at a massive scale, including health, mobility, and communication; high performance computing technology to capture and manage such big data; and, more generally, the adoption of network thinking across academia, government, and industry. As a result, over the last two decades, network based tools have become essential in the discovery of mechanisms and principles of complex systems, unveiling phenomena that were previously undetectable with other quantitative approaches.
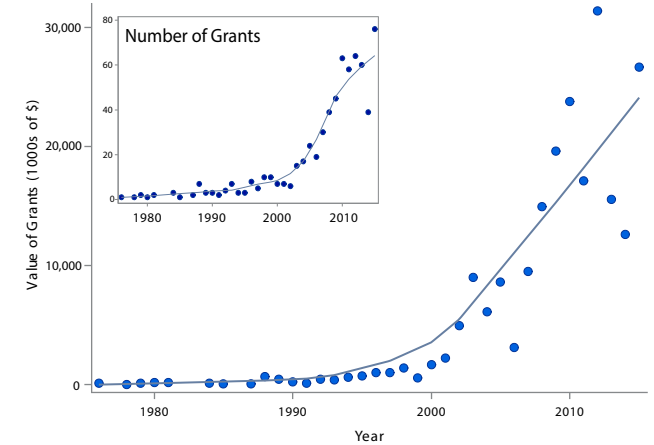

Figure 1. Publications of network studies and NSF grants with "network analysis" has increased dramatically from 1950 to 2016. *(Credit: James Moody, Duke University)*

Growth of network sciences as a field

Historically, network science did not have its own journals and institutions, with the field beginning to crystallize around the term 'network science' only around the turn of the 21st Century. At that point, publications of network research and NSF grants with "network analysis" in the title increased dramatically and has continued to go up every year (Figure 1).

We estimate that during the past two decades over 40,000 publications have used the term 'network science', with another 16,000 papers that cite these core papers (but are not themselves cited by network scientists). Social sciences and humanities first identified the field from a sociological perspective in the 1970's (with the first journals and conferences beginning in the 1980's), and mathematicians have explored the underlying graph theory problem since the 1960's. Network science is increasingly cited by engineering, life sciences, physics, and mathematics, which is visualized in Figure 2 as relationships among journal articles. Journal articles, represented as nodes in the network, are linked to articles by which it was cited. The nodes are color coded by discipline and sized according to the number of citations. Therefore, the most significant publications and relevant disciplines are easy identified. In this case, blue and green dominate the network, revealing the central role of physics and, to a lesser extent, mathematics in the growth of networks in the natural sciences (represented by bigger notes and more prominent color). Life sciences also shows a high number of network science articles while engineering, which only recently became engaged in the field, has only contributed a few high impact publications. The representation is only a partial view of the field, as the data from Web of Science does not capture network scientists working in the social sciences and computer science. However, it does offer a sense of the expanse of network science's impact across disciplines.
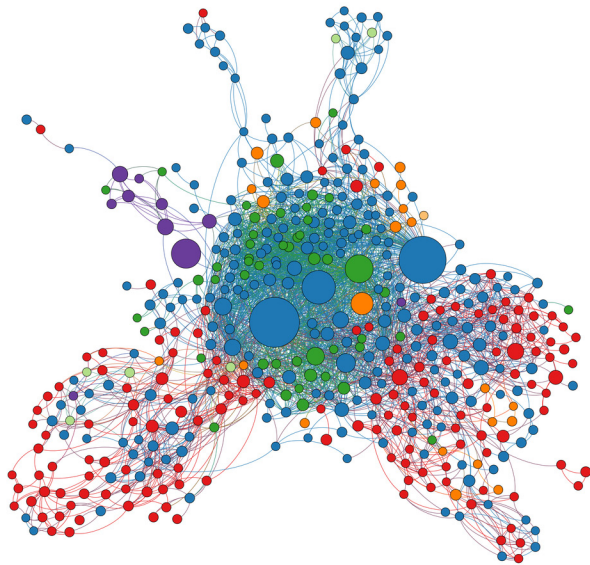
Figure 2. Natural sciences network of top 500 cited 'network science' papers. Nodes represent journal articles and directional links connecting nodes represent citations. Fields are depicted in different colors, with physics and mathematics having a notable number of cited network science publications. *(Credit: Junming Huang, Northeastern U)*

Table 1. Network Science Research Centers and Institutes

Institution	Founded	Center Name
Notre Dame University	2006	Interdisciplinary Center for Network Science (iCeNSA)
Northeastern University	2007	Center for Complex Network Research (CCNR)
Indiana University	2009	Center for Complex Networks and Systems Research (CNetS)
Rensselaer Polytechnic Institute	2009	Network Science and Technology Center (NEST)
US Military Academy	2009	West Point Network Science Center
Duke University	2010	Network Analysis Center (DNAC)
Central European University	2012	Center for Network Science
University of Pennsylvania	2013	Warren Center for Data & Network Sciences
Indiana University	2014	Network Science Institute (IUNI)
University of California, Santa Barbara	2014	NSF Network Science—NSF IGERT
Yale University	2014	Institute for Network Science (YINS)
Northeastern University	2015	Network Science Institute (NetSI)
Harvard Medical School	2015	Channing Division of Network Medicine
University of Maryland	2016	Network Biology (COMBINE) NSF NRT

There has been tremendous growth in journals, conferences, research institutes, and companies specifically dedicated to the field of network science. Just in the last 5 years, six new journals have been established: Journal of Complex Networks (2013), Network Science (2013), IEEE Transactions on Network Science and Engineering (2014), IEEE Transactions on Control of Network Systems (2014), and Applied Network Science (2016). Conferences that serve the research community include NetSci (a biannual conference supported by the Network Science Society), Sunbelt (since 1981, annual conference of the International Network of Social Network Analysts), IEEE Network Science & Engineering (since 2012), and dozens of other related annual conferences (CompleNet, since 2000, Complex Systems, since 2013), special topics conferences (e.g., urban networks; network visualization), and smaller workshops (e.g., SIAM Workshop on Network Science). Academic research centers and institutes (Table 1) have been formed to aggregate and coordinate this increased research activity. Graduate and doctoral programs are now offering training to a new cadre of network scientists (MS at Indiana University; PhD minor at UC Santa Barbara, funded as an NSF IGERT; MSc at Queen Mary University of London; PhD program at Northeastern University; and a PhD program at Central European University).

In 2005, at the request of the Director of the US Army's Basic Research Division, the National Research Council (NRC) examined network science research, both its emergence as a new research field and its importance to national security and global competitiveness. Following these reports, the National Science Foundation established a network science directorate and a series of network science centers were established by all the military services (including the Army Research Lab's 5-year, $40M Collaborative Technological Alliance effort which began in 2009).

The field has truly begun to converge into a formalized discipline with multi- and inter-disciplinary threads that underlie a new paradigm for how to understand information and map the spaces around us. Network science has generated tools and theoretical paradigms that have significantly changed the kinds of solutions we can develop across diverse domains. These approaches have been effective in improving health and disease outcomes, business and learning strategies, predictions of economies and social conflict, design of resilient technical infrastructures, security and access to information, and distribution of natural and technical resources. While network science is no longer a field in its infancy, its continued development requires the creation of strong institutional foundations, funding opportunities, and a rigorous theoretical framework, so future network scientists can tackle the unforeseen opportunities and challenges for this interdisciplinary basic research for the next generation.

It was in this context that 25 prominent network science researchers gathered for the Future Directions of Network Science Workshop on September 29–30, 2016 in Arlington, VA. They reviewed the current state of the field and discussed the opportunities and challenges for network science over the next 20 years. This report summarizes the key research directions and proposed trajectory that arose from those discussions.
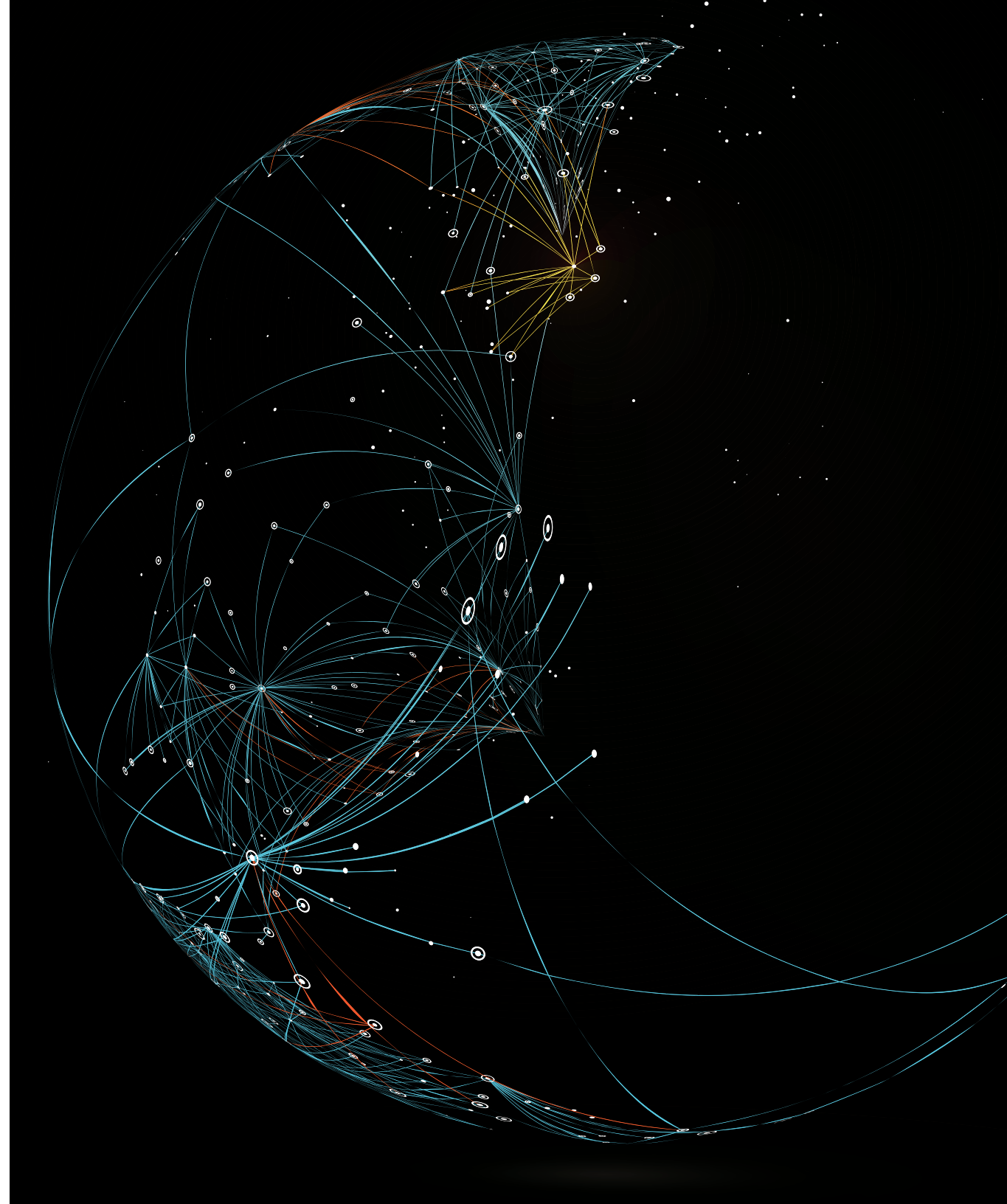
Recent Advances and Opportunities for Network Science

Network science is dedicated to understanding the structure and dynamics of systems with a focus on representing these systems as graphs or networks to discover and capture the inherent dependencies between the system's components. Much like statistics, network science represents a set of tools and theoretical perspectives on data and information. However in network science, the data and information is defined by the relationships between elements within a system. The elements are nodes in the system and the sets of linkages between them are edges. The basic unit of observation is the dyad, or pair. This new formalism opens up entirely new realms of discovery about the structure of a system, the movement of things through a system, and the functional capabilities that result from coordinated action within and between systems. As such, network science is not confined to specific application areas. It can be effectively used to understand any system that contains relationships and whose behavior is driven by these connections.

Advances in network science have dramatically improved our understanding and provided innovative solutions to complex network systems across scientific domains and sectors, such as cybersecurity, disease diagnosis, pharmaceutical development, infectious disease management, implementation of public policies, and entrepreneurial, educational, and governmental organization structures. Network science aims to provide descriptive, prescriptive, and predictive solutions.

The workshop participants discussed these advances and prospects in the context of five application domains:

- Group decision-making
- Personal and population health
- Socio-technical infrastructure
- Biological systems and the brain
- Human-machine partnerships

The following sections describe each of these domains, how network science research currently supports these areas, and the prospects for greater impact over the next 20 years.

Group decision-making

As modern societies become increasingly interconnected, multi-faceted, and complex, more rigorous models and techniques are needed to understand how groups organize, communicate, build knowledge, and make decisions. Understanding collective phenomena of human beings focuses on harnessing shared human effort (collective phenomena, crowd sourced problem-solving). Research in this area seeks to explain how groups build knowledge/expertise, reach consensus, achieve breakthroughs, and perform complex problem solving that would not be attainable through either individual efforts or a sequence of additive contributions. Collective phenomena involve connections between people, patterns of connectivity among them, and performance of the group given particular tasks.

Network science approaches have enabled the rigorous exploration of collectives as emergent systems, where structural and dynamical properties help to explain mechanisms in play that are otherwise hidden by classical statistical approaches. The goals of these efforts are to build theories, databases, and models that contribute to the formalization of methods and constructs to inform the study of group processes and performance.

For example, modern group structures are moving away from siloed and hierarchal to more network, matrix based organizations that are often distributed across fuzzily defined, overlapping groups. Network science methodologies can be applied to better understand the systemic implications of a large set of overlapping groups with overlapping missions, such as formal command structures and informal resource sharing relationships. Network science enables a way to understand this new form of human organization to (i) re-theorize what are the most functional forms of behavior for group tasks, (ii) conceptualize human production as human-to-human-to-computer collaborations ("thought collaborations"), and (iii) understand what patterns of connectivity, across what domains, yield effective outcomes for the collective.

In the not too distant future it will be possible to unobtrusively record group interactions and derive a deep understanding of how that group is organized (who is the leader, informal leader), what are the tasks that make up the work process of the group, who is connected to

"Network science aims to provide descriptive, prescriptive, and predictive solutions."

what task, and the underlying social structure, values, norms, and intentions. In the future, we may be able to automate the processing of face-to-face interactions to better understand the group process. The major breakthrough will be to increase computational capacities for passively collecting human data and, with automated classification schema, integrate audio, visual, image, and informational dimensions. These technologies will enable software that can process a video recording to decode who is looking at whom, what language people are using,

and what kinds of temporal patterns are being exhibited. Integration of network science with other computational techniques will extend these capabilities even further. For example, it should be possible to take digital traces of an organization's interactions to quickly understand the informal structure of that organization and the best HR management and leadership practices (recruitment, promotion, retention). Thus, integration of data fusion technologies within network scientific framework will radically improve human decision-making capacity.

Personal and population health

Networks have been essential for identifying and detecting epidemics, disease spreading, disease surveillance, and for population and community health interventions. Network tools have contributed to tremendous advances in our models of human disease and health related behaviors, and to mitigating or intervening in both personal and population health strategies. As an example, complex models that simulate human mobility patterns and social network interactions with disease transmission, have led to the definition of predictive approaches to infectious disease spreading that are increasingly adopted by national and international public health agencies (Figure 3). In regards to behavior-dependent health outcomes, it is necessary to explore combined diffusion processes across multiple levels (media platforms, spatial and social networks, etc.), while accounting for the interplay with other processes (behavioral feedback effects of intervention strategies, changes in interconnectedness due to nodal changes). Network capabilities that capture these multiplex relationships and account for time varying systems have significantly shifted this field.

The ultimate goal of health research is to provide interventions to protect and promote healthy options. The objectives of any intervention must be considered carefully to determine how to best integrate health and regulation policies at each data level (individual, organizational, and population). To that end, the intervention becomes part of the system, which requires models that not only account for these feedback loops, but also the development of dynamic, adaptive interventions that incorporate the reaction to the previous time step's intervention into the prescription at the following time step. The implication of changing
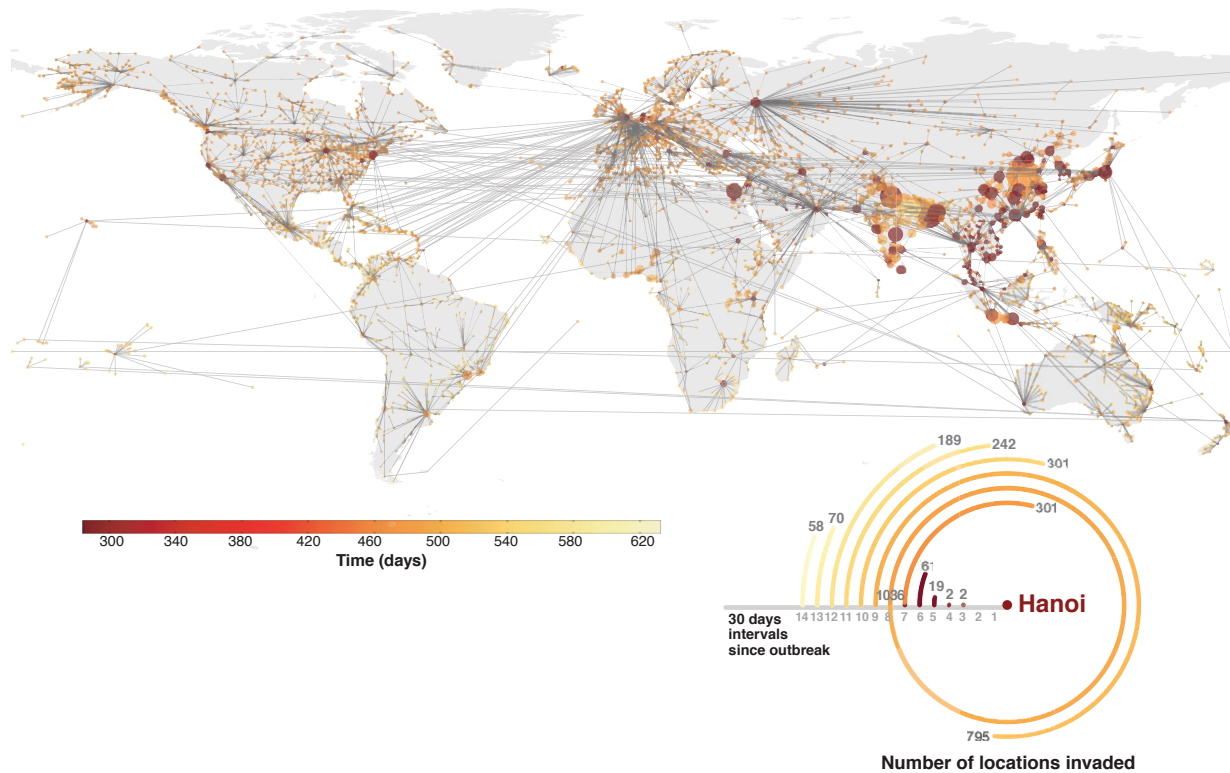

Figure 3. Infection tree for an outbreak from Hanoi, Vietnam. The size of each node is proportional to the population, and the color corresponds to the time of disease arrival at that node from dark red *(earlier)* **to light yellow** *(later)***. Each concentric arc on the circular plot on the bottom right is proportional to the number of locations invaded at 30 day intervals since the outbreak** *(Credit: Pastori-Piontti, A., et al. (2014) Network Science, 2(1): 132)*

a network *on the network* is computationally intensive particularly in multiple time-varying networks (i.e., infection contagion, vaccination, and behavioral change systems have vastly different time scales). Other challenges of interventions include unintended second or third order effects; formally defining the boundary of the target population (in the social world, boundaries may be fuzzy); whether to educate all citizens or target populations; designing evidence-based policy making to mandate behaviors; and ethical, legal, and cultural issues. Finally, given that there are a large number of endogenous and exogenous factors that interact over time and are able to affect a variety of health outcomes (e.g., cancer, diabetes, heart disease) during critical or sensitive periods of development, there is the potential for network science to aid greatly in identifying and understanding such relationships, especially for higher order interactions (e.g., diet, behavior, exposure).

Socio-technical infrastructures

Over the past 15 years, infrastructure has been extensively studied from the point of view of network science. This offers a unified description of the structure of diverse systems as networks of nodes (system components like electrical buses or metro stations) joined by links that represent connections between the nodes (e.g., transmission lines or railways). Significant study has focused on, for example, network models of regional and urban water supplies, specifically their ability to continue to meet user demand while being taxed by changing market conditions, growing population, natural disasters (Figure 4), and climate change. In power systems, this means a "smart grid" capable of self-healing and intelligent response to perturbations, and which must be robust against loss of generator synchrony, and voltage collapse, both key culprits in blackouts. In transportation systems (Figure 5), such a system could avoid congestion and facilitate the smooth flow of people and cargo in the face of natural disasters or terrorist attack.

Existing efforts in this area have focused largely on the structural and technological integrity and connectedness of the system. However, infrastructure is not purely
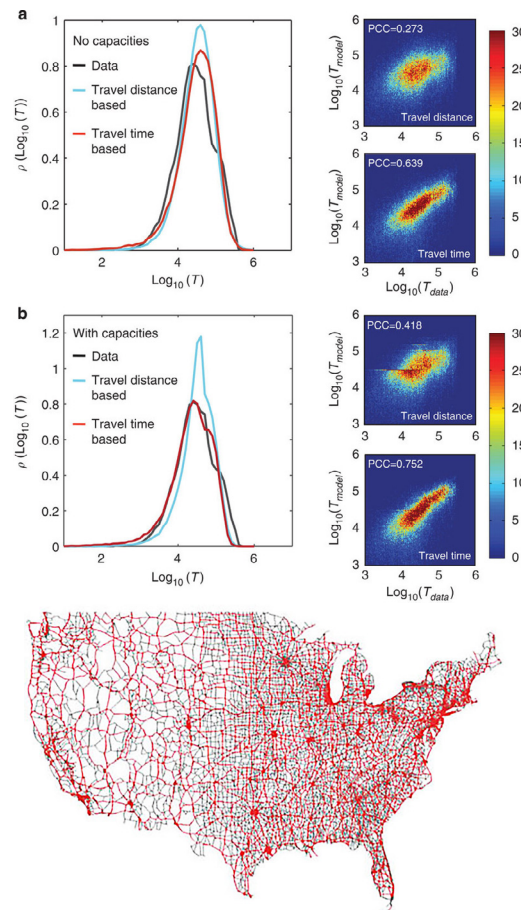

Figure 4. Natural or man-made disaster: Impact, failure, and recovery of the network. Model for the resiliency of the Indian Railways Network from random and intentional attacks *(left)* **and realistic natural and cyber or cyber-physical threat scenarios.** *(Credit: Bhatia, U. et al., (2015)* <u>PLoS ONE,</u> *10(11): e0141890)*

Figure 5. Spatial transportation networks (such as the US highway system) form the vascular system of our society. It has evolved organically, and its properties encode the modalities in which our economy interacts across space and time. Understanding flows in spatial networks driven by human mobility (socio-technical networks) has many important consequences: it enables us to connect throughput properties with demographic factors and network structure; it informs urban planning; helps forecast the spatio-temporal evolution of epidemic patterns; helps assess network vulnerabilities, and allows the prediction of changes in the wake of catastrophic events. The Figure shows the distribution of highway traffic along with the population in the continental US. It can be accurately modeled using a social mobility model coupled with a cost-minimizing algorithm for efficient distribution of the mobility fluxes through the network. *(Credit: Ren, Y., et al., (2014)* <u>Nature Comm</u>*, 5: 5347)*

an engineering problem. For example, traffic can become gridlocked even if a natural disaster leaves the underlying roads, bridges, and railways relatively unscathed. There is mounting recognition that in order to truly address the robustness and readiness of networked infrastructures, we must also measure social factors that drive their utilization. Fundamental questions in this research include: How to govern behavior in the social domain through intervention in the technical structure? How to influence efficiency of the technical domain through intervention in the network structures of the social domain? Frameworks to handle these fundamental questions will also need to model security, privacy, and ethical aspects of such new frameworks in socio-technical systems.

The move towards complex networks for socio-technical systems is driven by two central trends: 1) increased autonomy in technological systems is creating a hybrid system where nodes of the network are both humans and autonomous agents, and 2) the share economy is creating a massive network of peer-to-peer dynamic resource-sharing among people and organizations. Network science can be used to analyze, design, and govern such complex socio-technical systems by modeling the co-evolution of social, sensing, and technological domains and by developing proper theoretical and methodological frameworks for the multi-layered, multi-resource networks. Moreover, the impact on the collective behavior of socio-technical systems can be steered through the structure of the network. Finally, it will be important to define the trajectory of the systems evolution. Understanding dynamics of socio-technical systems requires

understanding intermediate states of agents in the system, and the evolutionary forecasts and confidence bounds when estimating such intermediate states.

When it comes to design of socio-technical systems, or policy analysis for socio-political systems, an equivalent of a scalable, reliable lab will be needed to run large-scale experiments that can be generalized to real-world problems. A new set of methodologies that enable reliable, data-driven simulations is a crucial enabler. These simulation environments will be used to test various interventions through a series of "what-if" scenarios. One of the fundamental issues in a virtual experimentation environment is the challenge of replication and reproducibility. Thus, new validation and verification methods for simulation scenarios are needed. In the next 10-20 years, virtual experimental simulation scenarios will be used as one of the main pillars for design and governance of such infrastructures as we make a transition towards a sharing economy and hybrid systems of human and autonomous agents.

Biological systems and the brain

Applications in biological sciences and the brain are powerfully driven by a steady stream of rich and voluminous data. Increasingly, researchers recognize that biological processes and architectures are best described and modelled as complex systems. Linking scales (micro to macro) has become a critical feature in understanding biological systems, for example in the structure and function of the nervous system or gut microbiome where structural composition, dynamics, and function are intrinsically linked. This generates a strong need for mathematical and statistical tools to model and analyze these systems. A new field of "network neuroscience" has begun to emerge, (Figure 6), drawing on network data across different scales—relationships among molecules revealing gene and protein networks; neuronal connectivity indicating synaptic networks; chemical and electrical circuitry in neuronal networks; and functional brain systems in brain networks (Figure 6).
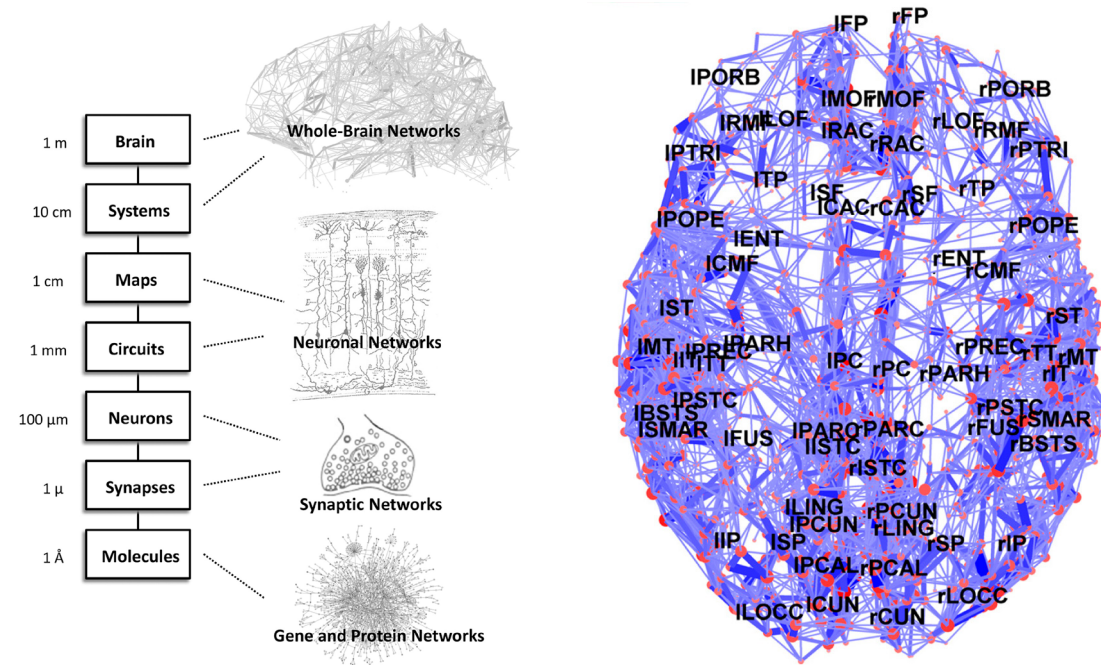
Figure 6. Connectivity of the brain: sub-cellular to whole brain networks. *(Left)* **Brain networks and cognitive architectures** *(Credit: Petersen SE and O Sporns (2015)* Neuron*, 88, 207-219). (Right)* **Mapping the structural core of human cerebral cortex.** *(Credit: Hagmann, P., et al., (2008)* PLoS Biology*, 6: e159)*

Network science has already significantly changed the way we analyze and understand biological systems. In particular, the family of -omics—connectomics, genomics, transcriptomics, proteomics, metabolomics, microbiomics, etc.—all exhibit properties of complex systems, whose network-level functionality and dynamics are still largely unexplored. Only recently, with the advent of network and system based analytic methods, have we been able to study these diverse systems, and begun to develop models to explain how their interactions enable fundamental systemic integration. Even descriptive characterization of properties has provided first glimpses of plausible models of structural and dynamic aspects of interactions among neural, protein, genetic, bacterial, and metabolic units. For example, network medicine has offered novel tools to identify drug targets or to repurpose existing drugs for new diseases. Similarly, community detection algorithms have enabled the discovery of major building blocks in anatomical and functional brain networks with applications in basic and clinical sciences. Community detection algorithms have also been important for detecting functional units in gene regulatory and protein interaction networks, identifying disease modules, and predicting the function of unknown system elements. Network science will continue to play a critical role in brain and biological science in the next decades, particularly as biological systems increasingly reveal essential poly-omic properties, where the various -omic families are deeply and fundamentally integrated, acting as a system of systems.

In spite of all the progress of network science for biological research, there are few common toolsets and approaches to carry out these research activities, even within restricted application areas. Building this infrastructure is crucial for generating consistent and reproducible findings, and for promoting a coherent theoretical framework. Further, the problem of how best to turn raw anatomical and functional brain data into network representations still remains a major barrier for collaborative developments in this field. There is a need for better tools dedicated to comparing features of biological networks, to characterizing individual differences in network architecture, and to developing realistic generative models that can identify factors underpinning network growth and evolution. In parallel, neuroscience data sets will likely expand 3 to 4 orders in magnitude over the next decade, driven by new developments in brain mapping and recording technologies. Further, biological sciences demand scalable, real-time and multi-omic network science tools and methods. Driven by the complexity of nodes and interactions, network mathematics and models of hypergraphs, polyadic relationships, heterogeneity of nodes and edges, and multi-layer networks which often operate on different spatial and temporal scales, will become increasingly essential in this area.

One of the most exciting possibilities for biological sciences is using network science tools for clinical practice. Imagine a scenario where a person presents with a problem that we think is associated with a disruption of brain structure or function. We then work from a causal mechanistic network science approach in real time to diagnose disruptions of brain connectivity - and we use this same mechanistic and causal framework to devise network-based therapeutic strategies and interventions. This "network-centric" approach is directed not just at single neurons, brain regions, and connections, but instead capitalizes on knowledge about the topological and dynamic aspects of connectivity. These kinds of approaches offer the first real possibilities for replacing or correcting entire networks in the brain. Brain replacement or repair may represent one of the most fundamental network challenges. Guided by future knowledge of structural dynamics, neural growth and function could be used to repair focally damaged sections of the brain. A slightly more hypothetical possibility would be to develop a process to build synaptic implants, where parts of the brain that have been damaged could be regrown under network-based structural, physical, and functional constraints, and then reconnected appropriately within the system, *in vivo*. The goal here would not be to replicate the missing part—instead one would aim for restoring the

overall functionality of the network. If we ever achieve this level of brain repair, network scientific advances will undoubtedly play a role in recovering or enhancing functions of the brain and other biological systems in a principled way. Repairing the brain, either through remapping of networks, or even a synaptic implant, may also be building blocks for brain augmentation possibilities in the next hundred years and beyond.

Human-machine partnership

It has become evident over the last decade that new discoveries and tools continue an ever-deepening dependency between computers and humans. In the next 20 years, a successful human-machine symbiosis will make possible the integration of biological, health, infrastructure, and social behavioral systems. As a society, it will become crucial to redraw the boundaries between humans and computers to understand how best to deploy human effort in a way that is functional for the collective. Advances in theory, methods, visualization, and digital platforms are expected to enhance performance, improve health, increase happiness and creativity, and reduce redundancy and inefficiencies. At the same time, this will decrease waste, reduce toxic emissions, and generally improve the wellness and sustainability of the planet. Much of the critical research necessary to achieve these outcomes is identified in the next section of this report.

These include the ability to model multi-layered, time-dependent, scalable networks; new models of diffusion, control, resilience; and applying new theories that address resiliency of population health, decision-making, and socio-spatial infrastructural systems. In this way, we see possibilities in the human-machine domain as the culmination of major technical advances across all of network science areas. Network science will play a critical role in navigating new man-machine capabilities, where network approaches will be able to handle the multi-dimensional, highly interdependent systems; and it will provide strategies for intervening in these systems (through optimization, synchronization, and control).

One of the most powerful advances of the human-computer partnership will be to enable enhanced decision-making with personalized data tailored to the individual, and for policy and decision makers it will aid in population level analyses. Social science research has shown that the challenge to changing behaviors at the individual level is almost never about identifying the "right" decisions, but instead persuading people to make good decisions at the right time. We know the healthy foods to eat, the amount of physical activity to do; we know that smiling makes us happier, and that deep breathing makes us calmer. Yet, all humans (with rare exceptions) must deliberatively expend a great deal

of energy to carry out these "best practices" each day. Poor decision-making is largely due to dysfunctional cognitive processes that are inherently human—e.g., implicit biases, confirmation biases, group think, etc.—forces to which computers are entirely invulnerable.

As machines become intrinsically embedded into human activities, they will have the power to dampen human cognitive interferences by providing customized information, tracking, reminders, nudges—using algorithms to know when and how to intervene, to effectively reduce biases and harness the wisdom of the crowds. New tools would include performance and health optimization strategies, as well as risk assessment and other susceptibilities, such as genetic prediction and epigenetic factors (projected as 3-D or 4-D network-based graphics). For example, the interface could integrate sensors at the individual level with virological/bacteriological sensors to assess susceptibility to sickness or risk of spreading the disease when traveling from one place to another. Human-machine partnerships at this level would fundamentally change the way humans make decisions, and significantly improve health and productivity, while profoundly shifting health and productivity of entire populations.

"The most important contact in your network in ten years may not be a human, but a machine that's tailored to help you overcome your deep and unconscious biases about how you evaluate information and process it, and this will be what makes the individuals more effective decision makers."

– Brian Uzzi, Northwestern University

Technical capabilities and challenges for network science research

Network science is fundamentally a mathematical framework to describe, prescribe, and predict complex systems. As described in the previous sections, network science has the potential to dramatically transform our understanding of complex systems across a range of application domains. It offers unique capabilities to measure, model, predict, and visualize these systems from an analytic vantage that accounts for the composition of systems, the types of nodes, the kinds of interactions between them, and what the nodes do when specific configurations, or sets of nodes and linkages, are activated. Networked systems are typically constrained by a set of forces that determine its structural elements, which often act heterogeneously across the system or between specific motif configurations, but which are driven by many of the same underlying principles. As a result, network science tools can rigorously quantify how structure affects individual entities within the system, and how the individual affects the system's structure. Further, these complex systems can be modeled as multiple network layers with different dynamics at varying temporal and spatial scales. For example, infrastructure models may include energy, communication, and transportation networks; epidemiological models might include mosquito, sexual, and mobility networks; social network models may include friend, coworker, and family networks.

Network approaches require sophisticated mathematical techniques, highly complex and relational data sets, and the continued development of new theories to explore properties and mechanisms of complex systems. Ongoing research in the following technical areas will be necessary to realize the impacts of network sciences on the application domains described in the previous section.

The key technical capabilities and challenges of network science are
1. Mathematical and computational methods
2. Data analysis and processing
3. Theory and mechanism

This section reviews the state of the technical capabilities and outlines the challenges and research aims for the next 20 years.

Mathematics and computation methods

The methodologies used in network science are adapted from traditional mathematics and computer science techniques. To fully realize the potential of network sciences, new methods are needed to account for scalability, dimensionality, and temporal dependencies of complex systems. In many cases, new theories are needed to inform the development of mathematical representations.

Scalability

Challenge: To develop meaningful measurements, methodologies, metrics, models, and theories that can provide actionable insights for processes that occur at multiple scales (micro, meso, and macro levels).

Scalability research explores how well models of the local dynamics of small groups (tens or hundreds) of the system's components can be scaled to populations of thousands or even millions. For example, social network research is concerned with developing theories that can explain relationships between the constitutive elements of social systems (usually individuals, but can also be micro-level cognitive agents) and the emergent phenomena that result from their interactions on larger scales, such as organizations (at the meso level), and whole societies (at the macro level). In brain research, network models capture relationships between the molecular (transcriptome), to anatomical (connectome), functional (effective connectivity), and behavioral (social) levels (Figure 7). Just as our models of humans must characterize their behaviors at various levels, an understanding of the brain requires analysis of the effects of its connectivity at multiple scales. The importance of this work is to develop overarching theories and models that scale from node/tie level to multi-level network levels. Without this, we lack truly fundamental insights into our natural and designed systems. For many network tools, models, and theories, the scalability is not yet understood. For example, the function of a "high centrality" node will be dramatically different in a system with ten agents vs hundreds or millions. It is also crucial to develop theoretical foundations to decide the appropriate meso-level to increase computational efficiency on the one hand and capture underlying group/module level properties on the other.

Research Aims: Define rigorous foundational theory of functional and behavioral emergence in networked systems to determine the point the system shifts from individual agents to a population. Develop theories for the hierarchy of information transfer across scale of the system. Develop meta-networks aggregation processes that provide effective complexity reduction schemes across representation scales.
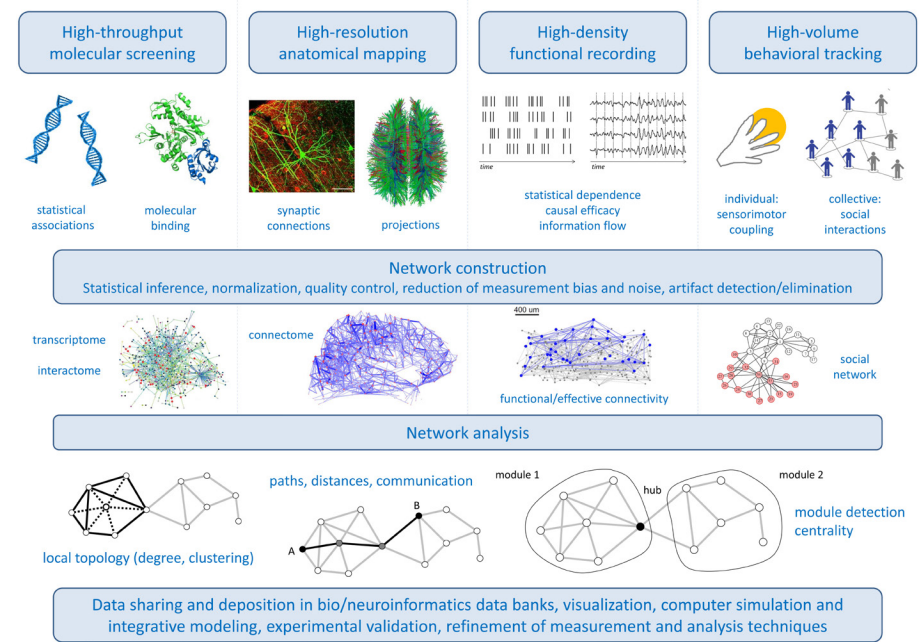
Figure 7. Multiple scale of brain network models. In studying the brain, network models are developed at multiple scales to incorporate processes occurring at the molecular, anatomical, functional, and behavioral levels. *(Credit: modified from Bassett D.S. and O. Sporns (2017) Nature Neuroscience, doi:10.1038/nn.4502)*

Dimensionality (multi-layer and multi-modal systems)

Challenge: To develop techniques that model and analyze interdependencies across different types of networks, each supported, or are driven by processes with different constraints across varying time scales.

Network models are well equipped to account for systems with multiple dimensions, and their interdependencies. Multi-level models investigate how individual agents in networks are nested within larger social, cultural, and technical networks and across physical space (geographic location/physical environmental constraints/policy space). Modeling multiple dimensions of systems can bridge multiple layers, such as cognitive, sociological, geo-spatial, technological, and ecological domains. In such models, structural and behavioral parameters of one layer can act as the constraints and boundary conditions for other layers, thus enabling descriptions of co-evolution and interdependent dynamics of various layers and modes. The scientific objectives of this research identify overarching mechanisms that span dimensions of human, natural, and technological systems that will parameterize, model, and predict the behavior of networked systems. For example, to study population health, it is necessary to explore combined diffusion processes across levels (media platforms, spatial and social networks, environmental dependencies, etc.), while accounting for the interplay with other processes (feedback from intervention strategies and changes in interconnectedness due to node change).

Another example rich in terms of both developing methodology and significance of the domain problem is modeling resilience of interdependent network infrastructures (Figure 8).

Research Aims: Develop a general theory of multiplex networks capable of extending dynamical process and control theory to heterogeneous multi-layer networks (e.g., understand the interference that constrains dynamical processes across multiple levels).
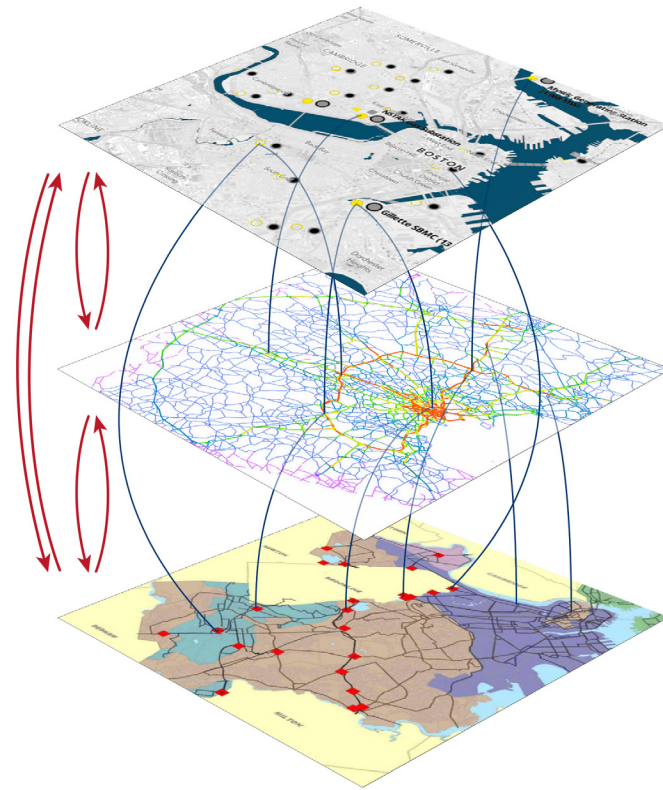
Figure 8. Interdependencies of core infrastructure systems in Greater Boston. The three layers depict the power *(top)***, transport** *(center)***, and water** *(bottom)* **infrastructure layers of Boston.** *(Credit: László Barabási, Sean Cornelius, Kim Albrecht, Northeastern U)*

Temporality

Challenge: To develop mathematical representations of time and changes in rates, such that models of complex interactions and their evolution can be updated in an iterative manner to reflect reality.

Non-linear, complex systems often exhibit emergent and collective phenomena over time. Most application areas must confront dynamic network issues—networks vary in temporal resolution and dyadic relations change over time. Because systems are understood as networks of relationships, dynamical models must specify time as an additional dimension—a temporal network layer, which can account for varying dependencies in time across layers (Figure 9). Temporality can be induced by multiple fundamental mechanisms: by learning, evolution, exogenous shocks, and phase change caused by changes in the level of clustering, activity, weight, etc. In socio-technical/socio-political applications learning and exogenous shocks are key to dynamical shifts, where as in physical systems, phase transitions dominate. In biological systems evolutionary and exogenous effects are critical. Understanding and predicting network dynamics across application areas will be critical in the near future for crisis response and disaster management, and efficient re-structuring of organizations, as well as improved service provision and health care. Current technologies that are used to model system dynamics, such as agent-based modeling, streaming metrics, and deep learning. However, these very data greedy, computationally demanding, and human labor intensive. New tools are needed that are more efficient and can identify fundamental mechanisms of emergence across domains.

Research Aims: Develop new methodologies and tools that link change to function—specifically, by developing a function-based methodology for predicting and measuring time variation in emergent coordination phenomena.

Figure 9. Visualization of longitudinal networks. The evolution of friendship among 17 students (known as Newcomb's Fraternity) is depicted by showing the imbalance of mutually assigned ranks bottom to top within matrix cells. Rows and columns are ordered to block the two emerging groups and a group of outliers with lower-than-expected ranks as indicated by color. *(Credit: Brandes, U and B. Nick (2011) IEEE Trans. on Visualization and Computer Graphics, 17(12): 2283)*

Data analysis & processing

The second technical area examined by the workshop participants is data analysis and processing. Complex systems have tremendous diversity of data sources, structure, and relevance. Ongoing research in data analysis and processing falls into five areas:

1. Causation in networks
2. Missing data & data accuracy
3. Detection of hidden communities
4. Multi-modal data integration
5. Data visualization

Each of these research areas presents unique challenges and opportunities described below.

Causation in networks

Challenge: To develop efficient ways to infer causality in network models assuming dependency across observations and parameters; and to develop new protocols that can accurately account for missing data and error within this network inference framework.

The ability to infer causal relationships is crucial for scientific reasoning. As network tradition assumes that structure is the basis for how systems behave, and given that networks are typically not evenly distributed (they are often scale-free and exhibit heterogeneous degree distribution), random sampling is not appropriate. Due to the inherent interdependencies, network indicators and model parameters may be sensitive to small changes or perturbations in the network. Additional actor attribute measurements and behaviors make the problem of missing data more complex.

There is also a need to systematically investigate how sampling approaches and associated data loss affect network models, particularly those that link behaviors to the social context. We do not know how commonly implemented network sampling mechanisms (which by its nature often leads to missing data) impact network-based statistical models and substantive conclusions drawn from them. Future work must provide rigorous evaluation and create appropriate adjustment factors that account for how incomplete data affects network-based statistical models and network-based inferences.

Work in network causation methodologies has shown evidence that a model's predictive power is negatively correlated with the human interpretability of the model—that is, either they have predictive power or power to explain, but not both. At this point in time, additive models seem to provide the best choice if one cares about both predictive and explanatory power. Thus, the research must guide whether one picks a highly predictive model, say a deep learner, with no explanatory power or a less predictive model, say an additive model with strong explanatory power. In some cases, one may not require explanation from the model—i.e., cases in which robust predictions are sufficient. In other cases, one

requires not only interpretability but a certain type of explanation. More work is needed to evaluate the trade-offs between predictability and interpretability.

Research Aims: Develop new ways and frameworks for inferring causality in network models. Systematically investigate how sampling approaches and associated data loss affect network models, particularly those that link behaviors to the social context. Develop algorithms that formalize the tension between predictability and interpretability.

Missing data & data accuracy

Challenge: To overcome the problems of incomplete data with new techniques that synthesize more accurate data, probe the network to identify the most relevant data, and account for the effects of incomplete data on causal inferences.

Missing and inaccurate data is one of the great road blocks of network science. Not only is it hard to get complete data when deliberately surveying a population, but from a network perspective, all networks are incomplete. There are few scenarios where data collected "in the wild" will capture natural, authentic behavior completely. Further, it is often the case that the hard to reach or hard to find populations are precisely those that we aim to capture (including adversarial or dark networks, or those engaged for example, in illicit drug use, risky sexual behavior). In other cases, deliberate deception can skew a data set. There are numerous techniques to manage these missing data, typically synthesizing or imputing plausible values, or otherwise accounting for the uncertainty. One approach is to ensure we get the right data. A small set of the most relevant data will substantially improve the network model over a larger set of unnecessary data.

Another approach is data imputations and simulations, built from data driven and theoretical approaches. For example, link and subgroup estimations focus on inferring links from incomplete data, particularly by incorporating human factors such as sociological and psychological metrics and theory. Probabilistic models utilize spatial and behavioral factors to build algorithms that capture which set of individuals are most likely to join specific groups and what characteristics will likely drive relevant interactions. New techniques are thus needed to account for inevitable data loss effects by quantifying and accounting for these biases in the resulting inferences derived from network models.

Research Aims: Build standardized rigorous approaches to (a) synthesize more accurate data, (b) probe the network to get the most relevant data, and (c) account for effects of incomplete or inaccurate data on causal inferences.

Detection of hidden communities

Challenge: To detect hidden groups of nodes to better understand the organization of the system and to predict its evolution.

Community detection algorithms use information about the topology of the graph (adjacency matrix) to find densely clustered subgroups within large networks. Community detection is an unsupervised classification problem and, as such, there are no universal protocols on how to validate algorithms or compare their performance. As a result, we are still far from having a reliable set of tools that can be applied on real networks. Many current approaches take an agnostic perspective, applying algorithms to any dataset ignoring the specificity of networks across domains despite clear indications that the clusters identified will vary depending on the motivating set of hypotheses and theories. Thus, there is a pressing need for frameworks that are sufficiently flexible to accommodate various features of networks and community structure, and that can be applied to a variety of different domains and clustering problems.

A recent trend toward developing domain-dependent algorithms aims to exploit as much information and peculiarities of the particular network data sets as possible. Generalist methods could still be used to get first indications about community structure and orient the investigation in promising directions, but domain-specific initiatives must be undertaken in the next years, especially in brain, social, and information networks. One of the promising approaches is generative models of networks. For example, stochastic block models can deliver full network models, recognize various types of group structure (community, multipartite, and core-periphery) when they are simultaneously present in the same network (instead of only targeting a single objective), and handle directed and weighted networks, overlapping communities, and multiple layers. Further, by simultaneously accommodating topology and metadata, this approach enables us to predict both links and annotations. Figure 10 demonstrates this approach using past games amongst college football team to predict future games, given the teams conference.

Research Aims: Develop extensions to community detection algorithms that are able to: a) validate and verify that algorithms are accurate representations of groups; (b) increase speed of computations through parallel implementations, and (c) better capture the dynamics of groups.

(a) Data   (b) Metadata   (c) Node prediction
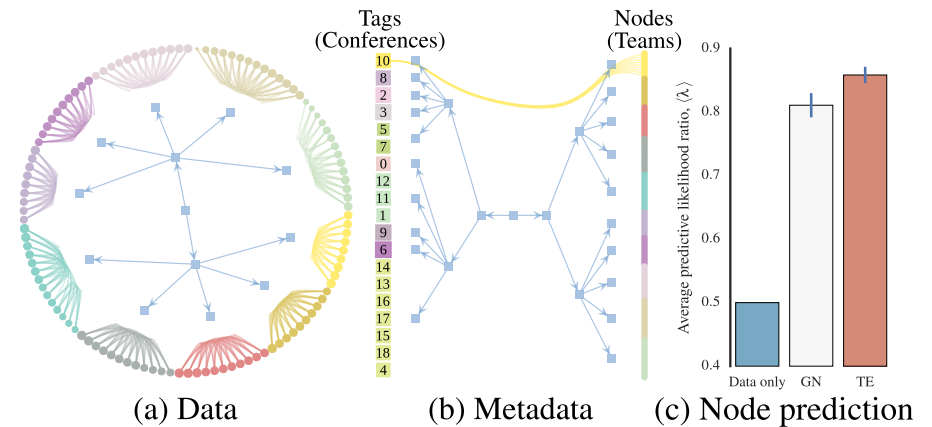
Figure 10. Joint data-metadata stochastic block model inferred on the college football network, where nodes are teams and links connect teams that have played each other during the regular season. a) Hierarchical partitions of the teams, which are indicated by the colors. b) Partition of the bipartite network formed by the teams (right) and their conferences (left). c) Node prediction: probability to assign a node to the correct conference based only on its annotation (here the conference the team belongs to). *(Credit: Hric, D., et al. (2016) Phys. Rev. X, 6: 031038)*

Multi-modal data integration

Challenge: To develop tools and theories that enable meaningful integration of multiple modalities of data sources into single integrated models.

In the next 20 years, we will be faced with increased interconnectedness and dependencies, more high-resolution data, and more integrated data. These will bring efforts for system level engineering and design by researchers, companies, and governments to integrate these data sources and produce incredibly rich datasets. For the first time in history, we have the ability to process recorded information on numerous dimensions of behaviors of humans and human-based systems, including fairly accurate mobility and location information, interpersonal contacts with detailed interactions, biometric data (e.g., individual health sensors), disease & death incidents, as well as individual perceptions and choices. However, we do not yet have the technology or the understanding of these systems to allow us to incorporate data sources into meaningful analytic frameworks.

Networks are the most viable paradigm to carry out such a task. A network based approach offers a relational framework of interconnectivity that enables us to meld social, spatial, technical, and behavioral information by overlaying information theoretic measures onto the network structures. For example, linguistic and structural cues derived from multiple modalities of communication (face-to-face, multi-group chats, texting, phone calls, email, and crowd based decision-making) could be used to generate models and algorithms capable of decoding social signals from dyadic and group conversations. Advances in this area will generate new capabilities that can ultimately be used to build data-rich, interdisciplinary models to detect, interpret, and predict the "social" environment of humans.

Research Aims: Improve the flexibility, speed, and scalability for streaming network metrics and develop rapid construction and re-use technologies for streaming data. In particular, streaming metrics will need to discriminate between change that is real and change due to data corruption or collection bias.

Network visualization

Challenge: To build network visualization tools that capture processes and features of systems in a way that are accurate and meaningful to the user.

Network visualization, machine learning, and data mining are often used to predict, reason, or describe networks. The scientific leaps of embedding network data analytics and visualization across myriad applications are potentially transformative in terms of understanding and designing interventions for the management of socio-technical and physical systems. For example, network based visualizations are by far the most effective way to communicate system processes (such as diffusion, contagion, resiliency, control), system level characteristics (e.g., clustering, centrality) or other hidden features (e.g., emergence, dark or latent subgroups)—critical features in the study of most human based systems (e.g., infectious diseases, critical infrastructure systems, cyberspace), see Figure 11. There are dozens of ways to represent such models—with varying degrees of accuracy and accessibility. For instance, how do we best represent pairwise interactions (link between two nodes), while also representing relationships among three or more people?

Better visualizations that are built faithfully on network data analytics to explore, describe, and communicate essential system features will drive better intuition and understanding of mechanisms in different application domains and lead to better policies and interventions. In order to achieve this, we need to better understand human perception of this kind of data. Who is the person using, perceiving or acting on this visualization? How does it make sense to them? How are we using different features in the visualization? How does it vary with size of image and format (2D v. 3D projection v. 3D hologram)? The answers to these questions surely vary from methods development to applications and for the purpose of the visualization (exploration, description, communication).

Research Aims: Design new network visualization methodologies collaboratively with psychology, cognitive science, and human factors research in the context of human perception.
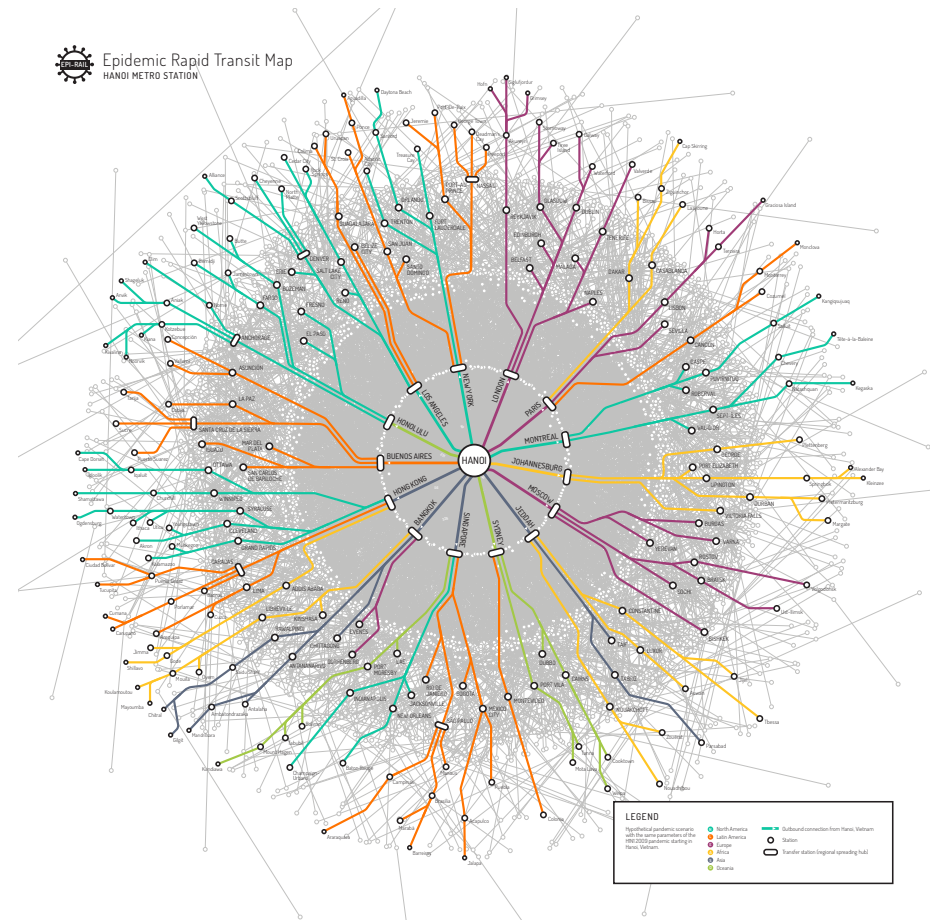
Figure 11. Epidemic rapid transit map. A hypothetical simulation of the H1N1 pandemic starting in Hanoi, Vietnam is visualized using the Global Epidemic and mobility (GLEAM) model. The different transit stops represent the cities reached by the disease. Some of them work as final destinations, while the major cities work as transfer stations, where the people traveling could follow different paths according to their final destinations. This network representation of the spreading disease provides evidence that even very remote regions can be reached by the disease, as long as a connection to the rest of the world exists. It also highlights the relevant role played by some cities with a large number of connections (hubs), such as New York, London, and Paris. *(Credit: Ana Pastore-Piontti, Luca Rossi, Nicole Samay and Alex Vespignani, Northeastern U)*

Theory of network systems

The last technical area for future network science research is the theory of network systems. The challenges here are in developing deeper mechanistic understanding of the following network processes:

1. Diffusion & spreading
2. Intervention and control
3. Coordination and collective phenomena

Diffusion & Spreading

Challenge: To model spreading, contagion, and diffusion processes of small and large scale networks over time, across social, technical, informational, and physical domains.

Diffusion and propagation often occur through systems as a result of exposure or influence between connected or proximal entities, whether those are people, bacteria, cells, or sensors. Spread of diseases, innovations, behaviors, information, and beliefs through systems are constrained by the topology and geometry of its connections. As such, these phenomena can be most effectively understood and analyzed with network-based tools. Network approaches have been used to create forecasting models of spreading (Figure 12) and adoption processes (e.g., new technologies, infectious diseases, medical practices, cultural norms, socio-political movements, radicalization, and fake news), and to develop dissemination and containment strategies (e.g., health behavior interventions, evacuation plans, critical infrastructure disruptions, logistics planning.

Recent developments in technology and media have vastly increased interconnectedness between systems facilitating diffusion and contagion, enabling rapid spreading through social platforms (e.g., Twitter, Facebook), airline alliances, economic interdependencies, etc. The result is that new and more complex diffusion processes constantly challenge our understanding. Fingerprints of this emergent collective behavior are the rapid onset of socio-political movements or the increased risk of pandemics. There is some evidence that these processes are not only just bigger and faster, but that the process of diffusion and spreading is substantively different at these massive scales. Further, it is increasingly recognized that diffusion processes in different kinds of systems (social vs technological vs the brain) have substantively different constraints. In addition, diffusion processes that involve multiple spatial and temporal scales (which we increasingly have the data to support) constitute new challenges that have only been recently identified (for instance, multi-strain disease, cooperative and interacting diseases). Finally, it is widely recognized that diffusion and spreading processes operate on dynamically changing networks, which adds an additional layer of complexity to this research.

Research Aims: Develop new mathematical and computational models of diffusion and spreading to resolve: a) scalability (do diffusion and spreading among 100 individuals differ substantively from that of 1000 or 1 million?); (b) multi-disciplinarity (identify diffusion and spreading properties that are universal and those that differ across social, technological, informational, and physical systems; (c) multiple spatial and temporal scales; and (d) rate and quality (e.g., fidelity of information, virility of virus) of spreading within with dynamically changing systems.
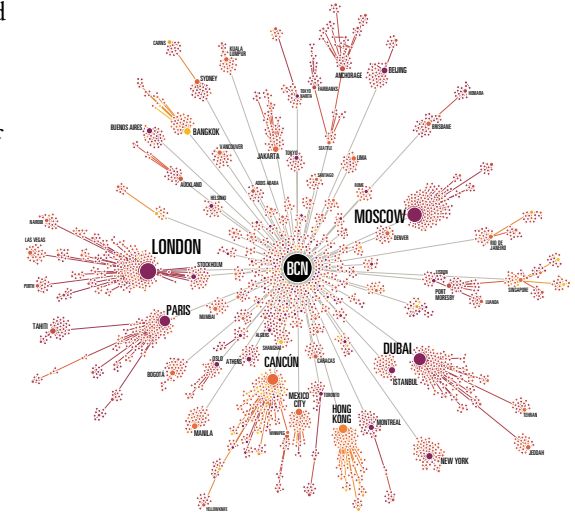
Figure 12. Simulation of spreading scenario in social and physical space. Nodes represent individuals, with darker colors indicating earlier time of adoption/infection. *(Top)* **A simulated scenario showing Barcelona as the first incidence of an infectious disease and spreading to cities connected through transportation routes;** *(Bottom)* **The corresponding map visualization of the same scenario.** *(Credit: A. Pastore-Piontti et al., (2016) in preparation, Northeastern U)*

Interventions and control

Challenge: To determine how best to influence and control networked systems.

Control and influence over a system requires a technical understanding of its underlying mechanics, to the extent that this knowledge informs how to perturb or resolve its functionality. Network systems typically have millions of degrees of freedom, making it impossible to know the state of all the factors in a system, let alone apply the standard control theories and tools, which assume omniscience. Further, it remains to be understood how to represent social systems using control theoretical algorithms. Advances are needed that extend traditional mathematical control theory and that develop a new mathematics to understand social systems. In addition, theoretical advances are needed to better determine the minimal knowledge of the system needed to enable control or influence (how much information, and at what level of granularity). New theory will appropriately account for two critical contextual features of the network (which have often been overlooked in the past): (1) whether we want to control the global (macro) behavior of the system or the local (micro) behavior of every single degree of freedom (e.g., do we care if the whole system is magnetized or just specific portions of it; are we concerned with the overall health of the economy versus optimizing for the target individuals or communities), and (2) what are the objectives of the agents in the system (e.g., are the agents adversarial, aiming to evade attack as in cyber or cancer; or individualistic, seeking to achieve personal gain). Resolving these issues (minimal knowledge needed, control of micro vs macro, agent objective) will lead to new theoretical frameworks to elucidate mechanisms to clarify under what conditions and at what scale, control, and influence can be achieved. Further, models will offer new understanding of network coordination, how elements in a system synchronize, and network resiliency, how to recover a system that has been disrupted. New theories will provide models that account for essential features of the network, that once formalized can be applied across domains.

Research Aims: Develop appropriate frameworks to capture higher-order interactions in networks instead of treating a network as an aggregate of pair-wise interactions. Develop a better underlying theory of nonlinear dynamics and steering trajectories. Apply effective field theory that would help us extend traditional control theory equations to more complicated realms. New applications will involve designing well-timed "nudges" to steer systems and specific interventions (smoking cessation; pharmaceutical drug design). The goal is to understand how small multi-modal interventions of a few nodes have the ability to nudge the entire system towards a desired state.

Coordination and collective phenomena

Challenge: To formalize models of collectives and characterize emergent properties that are meaningfully across domains from social, physical, biological systems.

The study of collectives draws from sociology, psychology, physics, and biology, where the overarching question is: how do the (inter)actions of the components give rise to the behavior of the system as a whole? The underlying mathematical tools have amassed a string of recent successes in addressing collective phenomena in diverse fields. For example, in materials science, one can combine simple components into novel "meta-materials" with properties neither found in nature nor foreseeable from the properties of the constituents in isolation. In physics, modeling collectives involves defining simple, immutable rules for the interactions of the constituents within a system and the emergence of associated macroscopic phenomena (e.g., bulk magnetism, elasticity, and the phase transitions between ordered and disordered states). In biology, within the mammalian immune system, any single cell recognizes only one or a few pathogens, such that the overall immune response is the result of the collective action of cell responses. Studies in the social sciences have focused on collective phenomena and team composition where findings have led to, for example, prescribed strategies to increase resilience and robustness of groups in response to the loss of leaders or team members; match team capabilities with mission objectives; optimize group size, diversity, and expertise. Each of these examples is fundamentally a system-level phenomenon that emerges from the networked interactions of a large number of individual components, be they genes, molecules, or people.

While the social sciences have primarily focused on small teams of 4–20 participants, a "collective" phenomenon in the physical sciences requires hundreds of actors at minimum and a measurable shift in a macroscopic variable (such as density). For yet further contrast, collective processes in biology are often considered to arise out of a global objective, such as fitness. If anything is clear from these examples, it's that much deeper analysis is needed to determine what these different notions of "collective" have in common. Our understanding of systems relies on uniquely network-based analytic approaches to detect, quantify, and predict how groups coordinate, and how and at what point do their behaviors exhibit emergent properties.

Research Aims: Multidisciplinary exploration to define what is an emergent phenomenon; what are the contexts under which various forms of emergence occur; what are the lower and upper boundaries for the population size required to achieve emergent capabilities; and what are structural and attributional parameters that determine how well a system performs, achieves success, or optimizes its fitness. To study social coordination and collective phenomena, we need to design large-scale data experimentation frameworks that can process massive amounts of data from groups, their interactions, and performance. To test hypotheses of optimization of groups, we need intervention strategies that can be embedded into automated delivery systems, as well as matching algorithms to improve how to study groups.

Considerations for an Emerging Network Science community

Throughout the discussions about application domains and technical aspects of network science, important considerations about the emerging network science community were discussed. While not all technical in nature, these topics represent the scaffolding which supports the science of networks.

Data access & ethics

Much of the data used in network science is built and created with specific objectives in mind. Alternatively, the participants envision new tools that collect useful and meaningful data that enable sensing and processing information purposefully. New infrastructure is also envisioned to enable tiered sharing of data that resembles the open-source software community. Data sharing is critical for reproducibility and replicability of research findings. Validation and verification of methods may require the community to undergo a cultural shift whereby findings from both positive and negative results are captured and shared.

In addition, there was a consensus that ethical considerations will need to be addressed more proactively in the development of new data sharing technologies and methodologies. For example, how might sensitive data, such as cell phone call records or Facebook data, be shared in a fashion that facilitates research while preserving privacy interests of consumers and proprietary interests of companies? How might canonical Twitter data be created to allow communal research? How might archives of news be created that could be shared amongst scholars? New strategies are needed to address increased safety (disseminating information to citizens), privacy (avoidance of "Big Brother" appearance) and ethics (clearly marking the line between customization of meta-data and social engineering). Participants discussed possible solutions, such as neutral, central data brokers that could mitigate some of these concerns.

Generalizability of new theories and methodologies

For network science to fulfill the roles outlined in this report, it needs to continually develop its theoretic core, offering new tools and methodologies. Much like statistics, network science is not confined to certain application areas. Instead fundamental methodological challenges in networks include questions about the generalizability of methods from one family or class of networks to another. While in some applications, "the network" can be a meaningful entity, in many others it is not. Bridging the gap between empirical observations and identifying the appropriate distribution(s) that could have generated those observations is essential for assessing the validity of many of the analytical approaches that have been very successful for understanding some families of networks but which may be inappropriate for others. Networks might present an appropriate approach to capture (certain) aspects of (some) systems, but they also serve to conceptualize phenomena such as co-purchasing of products that are not easily thought of as systems. Axiomatic approaches must be developed that guide what properties the analysis and representation of networks should satisfy, and new methodologies need to be developed for better understanding of the consequences of such choices.

A new mathematics for social systems

If we were to describe the world through a collection of theoretical functions and models, there would be so many equations that it would render these descriptions inaccurate or even meaningless. Some of the participants believe that our current mathematics are just not capable of describing the complexity of all the factors and their interdependencies, particularly human-based systems that involve cognition and behavior. It was suggested that we may need a new mathematics to describe social systems—one that involves time dependence, multiple scales, competing objectives, game theory, and others. In the 10–20 year horizon, the participants envision a new mathematics for social systems could be developed that is akin to a 21st century renormalization group theory, where we extract effective interactions across scales. Within this new matematical framework, network science can offer insights that allow us to engineer or design the 'best network' given a set of constraints and objectives.

Network representation learning

New successes in deep learning offer new methods for detecting network causation. In this computer science-based approach, representations are 'learned' by embedding them in lower-dimensional vector spaces, allowing us to reconstruct the original network and support network inference. Inspired by the deep learning methods like word2vec, these vector representations lead to very efficient learning frameworks which also provide intuitive interpretations at unprecedented scales. Examples include computation of node importance, community detection, network distance, link prediction, node classification, and network evolution. Network representation learning can take different forms based on the inference one seeks, such as structure-preserving network embedding, property-preserving network embedding, etc. For example, it is possible to create a joint network by building property sequences according to structural proximities. Depending on the inference task, different levels of network properties, from edge properties to node clusters, can be treated as basic units of the sequence. Similarly, different topological distance measures can be used to better capture the real process the network is representing.

Computing infrastructure

As discussed extensively in this report, the quantity of human data has increased, human processes can now be explored well beyond the classical statistical regression modeling framework to incorporate multi-dimensional, dynamic and simulation-based information. However, rigorous hypothesis testing in human groups are incredibly computationally intensive. A collection of only a dozen individuals can already represent a system with hundreds of parameter combinations, making it difficult to navigate theoretically, let alone experimentally or computationally. For example, while a simulation involving a set of five parameters on a small network of 100 individuals may be feasibly computed within an hour, the timeframe may turn into months when scaled to 10 parameters and a network with 200 individuals.

In the unchartered realm of big data, new norms concerning core infrastructure need to be established—necessities such as PCR machines, imaging systems, and microscopes, are considered common equipment in a biological lab, but analogous infrastructure for data labs are not yet considered requirements for university or government research labs. Data research labs require high performance computing (HPC) power to conduct large-scale experimental and simulation modeling. Integration of large multi-modal data sets require fast and robust processing, as well as large memory and storage. Increased computing power will fuel research capabilities by fostering, for example, large scale human experimental data collection and integration, simulation of social systems, processing digital trace data, and developing visual and data analytics to better understand complex human systems.

Identity of network science as a science

There is still debate over what constitutes the fundamental techniques, methods, and theories of network science. That is, how do we identify ourselves in a field of study that is becoming so pervasive across so many disciplines? Curriculum for network science programs reflects these debates, with existing programs showing quite different approaches—while some emphasize the physical scientific methods, others focus on the computer science and data driven techniques, and still others are based on social network analytic methods. In some cases, mostly out of convenience, programs are collections of courses from other departments while other programs have created new courses to meet the fundamental scientific challenges of the next century. Success for the field demands a connected, coherent framework for training the next generation of network scientists who will accomplish many of the major breakthroughs presented in this report. If the field is unable to do so, "network science" may further embed itself distributively across disciplines. For network science to grow into a truly developed field, it must develop both a core curriculum that establishes itself in academia, as well as a set of basic research initiatives that establish network science divisions in funding agencies.

Research trajectory for network science

This report outlines urgent obstacles which must be overcome to meet the major challenges of the 21st century and strategizes how the practice and application of network science will impact potential solutions. The trajectory for network science research to meet these challenges can be described for the near-, mid-, and long-term (5-, 10- and 20-year) horizons. In the near-term, research will establish the foundations of theoretical and data infrastructure. This will enable needed advances in computational and mathematical approaches and data processing over the mid-term that will result in long-term impacts for the five application domains. This section will describe the specifics for each of these timeframes.

Near-term advances for network science research (5-year)

Much of the excitement of network science in the last two decades has surrounded the remarkable discovery that, across domains, many complex systems exhibit common network properties, like assembly, growth, collapse and recovery. This universality formalizes the fundamental principles of these disparate systems, but many challenges remain for applying generic network models and algorithms to domain-specific phenomena. Thus, some of the greatest developments in the next 5 years are expected to be theoretically driven, domain-specific modeling achievements.

A deeper understanding of both the universalities and the differences found across systems of different types undergoing these processes will provide actionable implications for how to integrate these systems and how to build, control and rehabilitate them. This work will be successful if it identifies universal mechanisms and properties, as well as those that differ across social, technological, informational, and physical systems. The specific near-term achievements expected for **network theory** include:

- Theories of diffusion and spreading to resolve rate and fidelity of transmission, in networks with multiple spatial and temporal scales, and with dynamically changing systems.
- A rigorous foundational theory of functional and behavioral emergence in networked systems.
- Theories of the hierarchy of information transfer across the various scales of the system with an understanding of the interference constrains of dynamical processes across multiple levels.
- General theory of multiplex networks capable of extending dynamical process and control theory to heterogeneous multi-layer networks.
- Control theories to formalize the mechanistic process by which small multi-modal interventions of a few nodes have the ability to shift the entire system towards a desired state.

Near-term research will also focus on building data and computing infrastructure to enable research on large multi-modal data sets. Computing infrastructure requires fast and robust processing, as well as large memory and storage capabilities, while expansion of data infrastructure will include data repositories, data brokers, and systems to manage security, privacy and ethics. The specific near-term advances for **data analysis and processing** include:

- New data repository infrastructure that enables secure, tiered sharing of data that resembles the open-source software community.
- Development of strategies for neutral, central data brokers to increase safety and address privacy concerns and ethical issues.
- Better access to high performance computing (HPC) infrastructure to conduct large-scale experimental and simulation modeling.
- Development of new technology that increases speed of computations through parallel implementations.

Mid-term advances for network science research (10-year)

The near-term advances in domain-specific mechanistic models of network processes will enable development of advanced mathematical and computational techniques that capture the multi-level, multi-scale, and multi-temporal features of complex systems. In particular, the advances for **mathematics and computation** expected over the 10-year timeframe include:

- Methods for meta-network aggregation that effectively reduce complexity across representation scales.
- Development of a function-based methodology for predicting and measuring time variation in emergent coordination phenomena.
- Methods that extend control theory based models, such as to capture higher order interactions; develop nonlinear dynamics theory.

The near-term advances in data infrastructure will enable new research that will result in dramatic improvements for visualizing complex relationships, integrating diverse modes of data, and improving the ability to process and probe incomplete data. The specific **data analysis and data processing** advances expected over the next 10 years include:

- Development of multi-modal data integration capabilities, e.g., integration of data fusion technologies within network scientific framework.
- Improvements in flexibility, speed, and scalability for streaming network metrics that produce rapid, real-time construction, and re-use technologies for streaming data (e.g., streaming metrics that can discriminate between change that is real and change due to data corruption or collection bias).
- Design of new network visualization methodologies that integrate human perception constraints.
- Extensions to community detection algorithms to validate and verify that algorithms are accurate representations of groups.
- Build standardized, robust approaches to accurately synthesize human data, probe the network to get the most relevant data, and estimate effects of incomplete or inaccurate data on causal inferences.
- Use deep learning techniques to reconstruct the original network (from incomplete data) and support network inference.

Long-term advances for network science research (20-year)

Over the long-term, network science research will establish the network mathematics, theories, and tools to drive a number of profound achievements in the five domain areas: group decision-making, personal and population health, biological systems and brain, socio-technical infrastructure, and human-machine partnerships. At the workshop, some participants proposed that we need an entirely new mathematical framework to adequately characterize social systems—one that involves time dependence, multiple scales, competing objectives, games theory, etc. Research advances in methods for modeling multi-level, dynamically changing networks; rapid, real-time data collection and processing; data integration (and data fusion) technologies; theories concerning control and resilience, diffusion and contagion, and collective phenomena; visualization tools (particularly to communicate time-sensitive data to decision-makers) are expected to enable the following advances for the 5 domains:

- Record group interactions through passive technologies (e.g., sensors, micro-behavioral tracking) to decode formal and informal leadership and status structures, to identify the set of tasks being carried out, to determine who is connected to whom, to what task, and to quantify the underlying values and norms and intentions of group members.
- Implementation of dynamic, adaptive interventions that incorporate the effect of the intervention on the system and model reaction to the intervention, e.g., model the previous time step into the prescription at the following time step.
- Utilize network maps of higher-order, health-related interactions such as diet, behavior, and exposure to inform medical procedures and health policies.
- Develop methodologies that enable reliable, data-driven simulations of natural disasters—i.e., simulation environments that can be used to test various intervention scenarios through a series of "what-if" scenarios.
- Develop functional and structural maps of the brain that can be used to repair injured areas in clinical settings, either through re-mapping of networks, or even synaptic implants.
- Dampen human cognitive interferences (and harness the wisdom of the crowds) during decision-making by providing artificial agents to customize information, tracking, reminders, and nudges that will improve performance and health optimization strategies.

These major research initiatives will support the growth and innovation of network science research areas to support significant discovery over the next 20 years. We are optimistic that this work will be carried out and the network science of ten years from now will be dramatically more sophisticated than the one today.

Conclusions

Network scientific advances have enabled a deeper understanding of dynamical network systems across scientific domains. The field has grown rapidly over the past two decades, with dramatic increases in both studies that use network science terminology and grants awarded to network science based projects. A workshop, held in September 2016, brought together two dozen experts in the field to discuss the challenges and research trajectory over the next 5, 10 and 20 years. Representing the diversity in the field, participants included physicists, biologists, mathematicians, psychologists, computer political and behavioral scientists, and public health and communication experts.

The objectives of the workshop were two-fold: 1) to identify the important domain areas where network science capabilities have either led to significant discoveries or where network science tools and capabilities could make groundbreaking advances for future growth. Within the constraints of time, and the almost limitless possibilities, the discussion was confined to five substantive areas, including (a) group decision-making, (b) personal and population health, (c) biological systems and brain, (d) socio-technical and socio-political infrastructure, and (e) human-machine partnerships. 2) to identify challenges across domains, and in doing so, highlighted the specific capabilities that are unique to network scientific approaches. To this end, the discussion focused on the research steps that are required to make significant advances in (a) mathematics and computation, (b) data analysis and processing, and c) theoretical aspects of network scientific approaches.

The participants expect some of the greatest developments in the next 5 years to be theoretically driven, domain-specific modeling achievements, while the next 10 years will be characterized by major breakthroughs in mathematical and computational techniques that capture multi-level, multi-scale, multi-temporal features of complex systems. For data processing advances the important advances in the next 5 years will be largely infrastructural, including new data repositories, data brokers, and high performance computing. In 10 years, the major developments will be in data visualization, network inference techniques, and better tools to navigate and synthesize incomplete data.

Out of the vibrant and lively workshop discussions, participants converged on a number of other important topics that were deemed as critical for the trajectory of network science, including network representation learning, network validation and verification methodologies, and a new mathematics of social processes. For network science to grow into a truly developed field, it must develop both a core curriculum (to establish itself in academia) and a set of basic research initiatives (to establish network science divisions in funding agencies). We expect that in the next 5 years, departments dedicated to network science will become established, followed by increased number of network science divisions within government S&T communities, as well as other funding agencies.

"We expect some of the greatest developments in the next 5 years to be theoretically driven, domain-specific modeling achievements, while the next 10 years will be characterized by major breakthroughs in mathematical and computational techniques that capture multi-level, multi-scale, multi-temporal features of complex systems."

Appendix I
Network Science Workshop Attendees

Chris Arney
http://www.usma.edu/nsc/SitePages/Chris%20Arney.aspx
United States Military Academy, *david.arney@usma.edu*
Department of Mathematics
PhD (1985), Mathematics, Rensselaer Polytechnic Institute

Chirs Arney is Professor of Mathematics at the United States Military Academy. He graduated from the United States Military Academy and served 30 years in the Army before retiring in 2001. His graduate studies led to an MS in computer science and a PhD in mathematics from Rensselaer Polytechnic Institute. Chris spent much of his military career as a mathematics professor at West Point (NY) with other assignments in military intelligence at Schofield Barracks (HI), Fort Bragg (NC), Fort Huachuca (AZ), Norfolk Naval Base (VA), Fort Dix (NJ), and Fort Monmouth (NJ). He also served as the Dean of Mathematics and Sciences and as Interim Vice President for Academic Affairs at the College of Saint Rose in Albany and the Division Chief of the Mathematical Sciences and the newly established Network Sciences Divisions of the Army Research Office (NC). At ARO he managed and performed research in the area of cooperative systems, with particular interest in information networks, pursuit-evasion modeling, intelligence processing, artificial intelligence, and language for robots. Chris has authored 22 books, written over 125 technical articles, made over 250 presentations, and reviewed over 200 books. His technical areas of interest include network science, mathematical modeling, cooperative systems, and the history of mathematics and science. His primary teaching interests are in modeling and inquiry.

Albert-László Barabási
www. Barabásilab.com
Northeastern University, *Barabási@gmail.com*
Department of Physics
PhD (1994), Theoretical Physics, Boston University

Albert-László Barabási is the Robert Gray Dodge Professor of Network Science and a Distinguished University Professor at Northeastern University, where he directs the Center for Complex Network Research and holds appointments in the Departments of Physics and College of Computer and Information Science, as well as in the Department of Medicine at Harvard Medical School and Brigham and Women Hospital in the Channing Division of Network Science. He is also a member of the Center for Cancer Systems Biology at Dana Farber Cancer Institute. A Hungarian born native of Transylvania, Romania, he received his Masters in theoretical physics at the Eotvos University in Budapest, Hungary and was awarded a PhD three years later at Boston University. Barabási's latest book is "Bursts: The Hidden Pattern Behind Everything We Do" (Dutton, 2010) available in five languages. He has also authored "Linked: The New Science of Networks" (Perseus, 2002), currently available in eleven languages, and is the co-editor of "The Structure and Dynamics of Networks" (Princeton, 2005). His work lead to the discovery of scale-free networks in 1999 and proposed the Barabási-Albert model to explain their widespread emergence in natural, technological and social systems, from the cellular telephone to the WWW or online communities.

Ulrik Brandes
http://algo.uni-konstanz.de/brandes/
University of Konstanz, *ulrik.brandes@uni-konstanz.de*
Department of Computer & Information Science
PhD (1999), Computer Science, University of Konstanz

Ulrik Brandes is a Professor of Computer Science at the University of Konstanz since 2003. After receiving a Diploma from RWTH Aachen in 1994, a PhD and the habilitation from University of Konstanz in 1999 and 2002, he became Associate Professor at the University of Passau in the same year. With a background in algorithmics, his main interests are in network analysis and visualization, with application to social networks in particular. He is a member of the board of directors of the International Network for Social Network Analysis (INSNA), associate editor of Social Networks, and area editor of Network Science. He is a co-author of the visone software for network analysis and the GraphML data format. In a Reinhart Koselleck-Project Social Network Algorithmics funded by DFG, he takes a shot at improving the methodological foundations of network science. As a principal investigator in the ERC Synergy Project NEXUS 1492 he is working on reconstructing archaeological networks from fragmentary and heterogeneous observations.

Kathleen Carley
http://www.casos.cs.cmu.edu/carley.html
Carnegie Mellon University, *kathleen.carley@cs.cmu.edu*
Department of Computer Science, Institute for Software Research
PhD (1984), Sociology, Harvard University

Kathleen M. Carley is a Professor at the Institute for Software Research in the School of Computer Science at Carnegie Mellon University. She is the Director of the Center for Computational Analysis of Social and Organizational Systems (CASOS), a university wide interdisciplinary center that brings together network analysis, computer science, and organization science (www.casos.ece.cmu.edu). Professor Carley's research combines cognitive science, social networks and computer science to address complex social and organizational problems. Her specific research areas are dynamic network analysis,

computational social and organization theory, adaptation and evolution, text mining, and the impact of telecommunication technologies and policy on communication, information diffusion, disease contagion and response within and among groups particularly in disaster or crisis situations. She and her lab have developed infrastructure tools for analyzing large scale dynamic networks and various multi-agent simulation systems. She is the founding co-editor of the journal Computational and Mathematical Organization Theory which she now co-edits with Dr. Terrill Frantz. She has co-edited several books in the computational organizations and dynamic network area.

Noshir Contractor
http://sonic.northwestern.edu/people/noshir-contractor/
Northwestern University, *ncontractor@gmail.com*
Department of Engineering and Applied Science
PhD (1987), Communication, University of Southern California

Noshir Contractor is the Jane S. & William J. White Professor of Behavioral Sciences in the McCormick School of Engineering & Applied Science, the School of Communication and the Kellogg School of Management at Northwestern University. He is the Director of the Science of Networks in Communities (SONIC) Research Group at Northwestern University. He is investigating factors that lead to the formation, maintenance, and dissolution of dynamically linked social and knowledge networks in a wide variety of contexts including communities of practice in business, translational science and engineering communities, public health networks and virtual worlds. His research program has been funded continuously for over 15 years by major grants from the U.S. National Science Foundation with additional current funding from the U.S. National Institutes of Health (NIH), NASA, Air Force Research Lab, Army Research Institute, Army Research Laboratory, the Gates Foundation and the MacArthur Foundation. Professor Contractor has published or presented over 250 research papers dealing with communicating and organizing. His book titled Theories of Communication Networks (co-authored with Professor Peter Monge and published by Oxford University Press, and translated into simplified Chinese in 2009) received the 2003 Book of the Year award from the Organizational Communication Division of the National Communication Association. In 2014 he received the National Communication Association Distinguished Scholar Award recognizing a lifetime of scholarly achievement in the study of human communication. In 2015 he was elected as a Fellow of the International Communication Association. He is the co-founder and Chairman of Syndio, which offers organizations products and services based on network analytics. Professor Contractor has a Bachelor's degree in Electrical Engineering from the Indian Institute of Technology, Madras and a PhD from the Annenberg School of Communication at the University of Southern California.

Kate Coronges
http://www.networkscienceinstitute.org/people
Northeastern University, *kcoronges@gmail.com*
Network Sciences Institute
PhD (2009), Health Behavior Research, University of Southern California

Kate Coronges is the Executive Director of the Network Science Institute at Northeastern University. She provides administrative leadership to the Institute by contributing to long-term strategic planning and vision for its role in the larger scientific community. Prior to this, she was a Program Manager at Army Research Office where she ran two portfolios of high-risk, high-impact research to support US Army's basic science investments in Social and Cognitive Networks and Social Informatics; and an Assistant Professor in the Department of Behavioral Sciences and Leadership at the US Military Academy. She received her PhD in Health Behavior Research from the University of Southern California in 2009. Her research has focused on social structures and dynamics of teams and communities and their impacts on communication patterns, behaviors and performance. She has published in social science, computer science, and network science journals, and was the Managing Editor for Connections journal, International Network of Social Network Analysis for 10 years.

Raissa D'Souza
http://mae.engr.ucdavis.edu/dsouza
University of California at Davis, *raissa@cse.ucdavis.edu*
Department of Computer Science/Physics
PhD (1999), Statistical Physics, MIT

Raissa D'Souza is Professor of Computer Science and of Mechanical Engineering at the University of California, Davis and an External Professor at the Santa Fe Institute. She received a PhD in statistical physics from MIT in 1999 then was a postdoctoral fellow, first in Fundamental Mathematics and Theoretical Physics at Bell Laboratories, and then in the Theory Group at Microsoft Research. Her interdisciplinary work on network theory spans the fields of statistical physics, theoretical computer science and applied math, and has appeared in journals such as Science, PNAS, and Physical Review Letters. She serves on the editorial board of numerous international mathematics and physics journals, is a member of the World Economic Forum's Global Agenda Council on Complex Systems, and is currently the President of the Network Science Society.

Tina Eliassi-Rad

http://eliassi.org

Northeastern University, *tina@eliassi.org*

Department of Computer Science, Network Science Institute

PhD (2001), Computer Sciences, University of Wisconsin-Madison

Tina Eliassi-Rad is an Associate Professor of Computer Science at Northeastern University in Boston, MA. She is also on the faculty of Northeastern's Network Science Institute. Prior to joining Northeastern, Tina was an Associate Professor of Computer Science at Rutgers University and before that she was a Member of Technical Staff and Principal Investigator at Lawrence Livermore National Laboratory. Tina earned her Ph.D. in computer sciences (with a minor in mathematical statistics) at the University of Wisconsin-Madison. Her research is rooted in data mining and machine learning and spans theory, algorithms, and applications of massive data from networked representations of physical and social phenomena. Tina's work has been applied to personalized search on the World-Wide Web, statistical indices of large-scale scientific simulation data, fraud detection, mobile ad targeting, and cyber situational awareness. Her algorithms have been incorporated into systems used by the government and industry (e.g., IBM System G Graph Analytics) as well as open-source software (e.g., Stanford Network Analysis Project). In 2010, she received an Outstanding Mentor Award from the Office of Science at the US Department of Energy.

Santo Fortunato

https://sites.google.com/site/santofortunato/

Indiana University, *santo@indiana.edu*

Department of Computer Science

PhD (2000), Theoretical Particle Physics, University of Bielefeld

Santo Fortunato is Professor of Complex Systems at the Department of Computer Science of Aalto University, Finland. Previously he was director of the Sociophysics Laboratory at the Institute for Scientific Interchange in Turin, Italy. Prof. Fortunato received his PhD in Theoretical Particle Physics at the University of Bielefeld in Germany. He then moved to the field of complex systems, via a postdoctoral appointment at the School of Informatics and Computing of Indiana University. His current focus areas are network science, especially community detection in graphs, computational social science and science of science. His research has been published in leading journals, including Nature, PNAS, Physical Review Letters, Reviews of Modern Physics, Physics Reports and has collected about 15,000 citations (Google Scholar). His review article Community detection in graphs (Physics Reports 486: 75–174, 2010) is the most cited paper on networks of the last years. He received the Young Scientist Award for Socio- and Econophysics 2011, a prize given by the German Physical Society, for his outstanding contributions to the physics of social systems.

Michelle Girvan

www.networks.umd.edu

University of Maryland, *girvan@umd.edu*

Department of Physics

PhD (2004), Physics, Cornell University

Michelle Girvan is an Associate Professor in the Department of Physics and the Institute for Physical Science and Technology at the University of Maryland, College Park. She is also a member of the external faculty at the Santa Fe Institute. Her research operates at the intersection of statistical physics, nonlinear dynamics, and computer science and has applications to social, biological, and technological systems. More specifically, her work focuses on complex networks and often falls within the fields of computational biology and sociophysics. While some of the research is purely theoretical, Girvan has become increasingly involved in using empirical data to inform and validate mathematical models.

Kimberly Glasgow

Johns Hopkins University, *kimberly.glasgow@jhuapl.edu*

Applied Physics Laboratory

Kimberly Glasgow is a Researcher at Johns Hopkins University Applied Physics Laboratory, where she leads a social media analytics research group. Her current work and interests include development of social network methods for cyberdefense and insider threat detection, extending social network methods and techniques to new or unique intelligence problems, assessing the impact of social networking and social media in violent protests, civil uprising, and crisis, reflections of social-psychological and mental health-related process in social media, and the role of social networks in influencing cognition and decision-making in individuals. She is the author of the chapter "Big Data and law enforcement: advances, implications and lessons from an active shooter case study," in Big Data and National Security: A Practitioner's Guide to Emerging Technologies for Law Enforcement, 2015, as well as a number of scholarly publications in the areas of social media, machine learning, social networks, and computational social science.

Jesus Gomez-Gardenes

http://complex.unizar.es/~jesus/

University of Zaragoza, *gardenes@gmail.com*

Department of Condensed Matter Physics

PhD (2006), Physics, University of Zaragoza

Jesus Gomez-Gardenes is Associate Professor at the University of Zaragoza, Spain. He obtained his PhD in Physics in 2006 and after postdoctoral periods in different Universities and Research Institutes. In 2006 he received an awarded

from the Royal Physical Society of Spain as the Best Young researcher. He has coauthored more than 90 research articles covering different topics related to Network Science and Dynamical processes such as synchronization, epidemics, evolutionary dynamics and stochastic processes, among others. He is currently involved in different national and international research projects about Network Science, being "Profesor Visitante Especial" of the Brazilian CNPq.

Babak Heydari
www.stevens.edu/cens
Stevens Institute of Technology, *babak.heydari@stevens.edu*
Department of Systems and Enterprises
PhD (2008), Electrical Engineering, University of California at Berkeley

Babak Heydari is a faculty at the School of Systems and Enterprises, at Stevens Institute of Technology and director of Complex Evolving Networked Systems Lab. He holds a PhD in Electrical Engineering from University of California at Berkeley with a minor in management and economics and has three years of industry experience in Silicon Valley. Dr. Heydari has a diverse set of research interests and academic backgrounds and does interdisciplinary research at the intersection of engineering, economics and systems sciences. His current research, is on developing model-driven approach in analysis, design and governance of complex networked systems. His research interests are network resource sharing formation and diffusion of collective behavior, modularity, emergence and evolution of collective behavior and the co-evolution of structure and behavior in complex networks. His research has been funded by NSF, DARPA, INCOSE, SERC and a number of private corporations. Prof. Heydari is the recipient of NSF CAREER Award in 2016.

David Lazer
www.lazerlab.net
Northeastern University, *davelazer@gmail.com*
Department of Political Science
PhD (1996), Political Science, University of Michigan

David Lazer is a Distinguished Professor in the Department of Political Science and College of Computer and Information Science. His research interests include group learning in technology-mediated environments, as well as consensus and opinion formation in groups, particularly in political settings, or pertaining to governance.

Yang-Yu Liu
http://scholar.harvard.edu/yyl/biocv
Harvard Medical School, *yyl@channing.harvard.edu*
Brigham and Women's Hospital, Channing Division of Network Medicine
PhD (2009), Physics, University of Illinois at Urbana-Champaign

Yang-Yu is Assistant Professor at Harvard Medical School and Associate Scientist in the Channing Division of Network Medicine at Brigham and Women's Hospital. He is a statistical physicist by training, with expertise in analytical calculation, modeling and data analysis. His Ph.D. research encompassed a broad range of topics from statistical to condensed matter and biological physics. Currently, he is working on complex networks and systems biology. The primary goal of his current research is to combine tools from control theory, network science and statistical physics to address the challenging questions pertaining to controlling and observing complex biological systems, which could have a major impact in network medicine, a rapidly developing field that applies systems biology and network science methods to human disease.

Patricia Mabry
www.iuni.iu.edu
Indiana University, *pmabry@iu.edu*
Department of Public Health, Indiana University Network Science Institute
PhD (1996), Clinical Psychology, The University of Virginia

Patricia Mabry is the first Executive Director of the new Indiana University Network Science Institute (IUNI) and a Senior Research Scientist in the IU-Bloomington School of Public Health. Through interdisciplinary collaboration, educational offerings, methodological innovation, theoretical development, and provision of supercomputing and IT resources, IUNI incubates 21st century network science among its 150 faculty members. Prior to joining IU, Dr. Mabry had a 15-year career at NIH galvanizing interdisciplinary research in systems science among behavioral and social science researchers. Her expertise spans obesity, tobacco control, diabetes, mood disorders, scientific rigor, and big data. Her work has been published in Science, the American Journal of Public Health, the American Journal of Preventive Medicine, Nicotine & Tobacco Research, and PLoS Computational Biology. Dr. Mabry is a Fellow of the Society of Behavioral Medicine and was a 2008 recipient the Applied Systems Thinking Prize. Dr. Mabry holds a Ph.D. in clinical psychology from the University of Virginia.

Madhav Marathe
https://www.bi.vt.edu/faculty/Madhav-Marathe
Virginia Tech, mmarathe@vbi.vt.edu
Department of Computer Science, Biocomplexity Institute of Virginia Tech
PhD (1994), Computer Science, University of Albany

Madhav Marathe is the Director of the Network Dynamics and Simulation Science Laboratory at the Biocomplexity Institute of Virginia Tech and Professor of Computer Science at Virginia Tech. Before joining Virginia Tech, he was a Team Leader in the Computer and Computational Sciences Division at the Los Alamos National Laboratory where he led the basic research programs in foundations of computing and high performance simulation science for analyzing extremely large socio-technical and critical infrastructure systems. He has published more than 250 research articles in peer reviewed journals, conference proceedings, and books. He has over ten years of experience in project leadership and technology development, specializing in high performance computing algorithms and software environments for simulating and analyzing socio-technical network science. He is a Fellow of the IEEE, ACM and AAAS.

Yamir Moreno
http://cosnet.bifi.es/~yamir
University of Zaragoza, yamir.moreno@gmail.com
Department of Theoretical Physics
PhD (2000), Physics, University of Zaragoza

Yamir Moreno is a Professor in the Department of Theoretical Physics, head of the Complex Systems and Networks Lab (COSNET), and the Deputy Director of the Institute for Bio-computation and Physics of Complex Systems (BIFI) at the University of Zaragoza in Spain. His research interests include the study of nonlinear dynamical systems coupled to complex structures, transport processes and diffusion with applications in communication and technological networks, dynamics of virus and rumors propagation, game theory, systems biology of TB (Tuberculosis), the study of more complex and realistic scenarios for the modeling of infectious diseases, synchronization phenomena, the emergence of collective behavior in biological and social environments, the development of new optimization data algorithms and the structure and dynamics of multilayer complex systems.

Prof. Moreno has published more than 165 scientific papers in international peer-reviewed journals. He is a Divisional Associate Editor of Physical Review Letters, a member of the Editorial Boards of Scientific Reports, Applied Network Science and Journal of Complex Networks, and Academic Editor of PLoS ONE. Prof. Moreno is the President of the Complex Systems Society (CSS), the Vice-President of the Network Science Society, and a member of the Future and Emerging Technology Advisory Group of the European Union's Research Program H2020.

Peter Mucha
www.mucha.web.unc.edu
University of North Carolina, mucha@unc.edu
Department of Mathematics
PhD (1998), Applied and Computational Mathematics, Princeton University

Peter Mucha is the Bowman and Gordon Gray Distinguished Term Professor of Mathematics at the University of North Carolina at Chapel Hill. His research includes a variety of topics in network science, including developments in community detection, network representations of data, modeling network dynamics, model interactions in networked materials, and diffusive processes with applications to disease and health behaviors. His group activities are fundamentally collaborative, with collaborators in departments, including Archaeology, Epidemiology, Finance, Geography, Infectious Diseases, Neuroscience, Pharmacology, Pharmacy, Physics, Political Science, Psychology, Public Policy, Sociology, and Statistics, among others.

Olaf Sporns
www.indiana.edu/~cortex
Indiana University, osporns@indiana.edu
Department of Psychological and Brain Sciences
PhD (1990), Neuroscience, Rockefeller University

Olaf Sporns is the Robert H. Shaffer Chair, a Distinguished Professor, and a Provost Professor in the Department of Psychological and Brain Sciences at Indiana University in Bloomington. He is co-director of the Indiana University Network Science Institute and holds adjunct appointments in the School of Informatics and Computing and the School of Medicine. After receiving an undergraduate degree in biochemistry, Dr. Sporns earned a PhD in Neuroscience at Rockefeller University and then conducted postdoctoral work at The Neurosciences Institute in New York and San Diego. His main research area is theoretical and computational neuroscience, with a focus on complex brain networks. In addition to 200 peer-reviewed publications he is the author of two books, "Networks of the Brain" and "Discovering the Human Connectome". He currently serves as the Founding Editor of "Network Neuroscience", a journal published by MIT Press. Sporns was awarded a John Simon Guggenheim Memorial Fellowship in 2011 and was elected Fellow of the American Association for the Advancement of Science in 2013.

Christoph Stadtfeld
http://www.social-networks.ethz.ch/
ETH Zurich, *c.stadtfeld@ethz.ch*
Department of Social Sciences
PhD (2011), Economics, Karlsruhe Institute of Technology

Christoph Stadtfeld is Assistant Professor of Social Networks at ETH Zurich, Switzerland. He holds a PhD from Karlsruhe Institute of Technology and has been postdoctoral researcher and Marie-Curie fellow at the University of Groningen (with Tom Snijders), the Social Network Analysis Research Center in Lugano (with Alessandro Lomi), and the MIT Media Lab (with Sandy Pentland). His research focuses on the development and application of theories and methods for social network dynamics.

V.S. Subrahmanian
https://www.cs.umd.edu/users/vs/
University of Maryland, *vs@umiacs.umd.edu*
Department of Computer Science, Lab for Computational Cultural Dynamics, Center for Digital International Government
PhD (1989), Computer Science, Syracuse University

V.S. Subrahmanian is Professor of Computer Science and Director of the Lab for Computational Cultural Dynamics and Director of the Center for Digital International Government at the University of Maryland. He previously served a 6.5 year stint as Director of the University of Maryland Institute for Advanced Computer Studies. His work stands squarely at the intersection of big data analytics for increased security, policy, and business needs.

Zoltán Toroczkai
http://obelix.phys.nd.edu/~toro/
University of Notre Dame, *toro@nd.edu*
Department of Physics
PhD (1997), Theoretical Physics, Virginia Tech

Zoltán Toroczkai is a Professor of Physics and a Concurrent Professor of Computer Science and Engineering at University of Notre Dame and a Fellow of the American Physical Society. He obtained his PhD in theoretical physics from Virginia Tech in 1997. He spent his postdoctoral years in the condensed matter group at University of Maryland at College Park and as a Director Funded Fellow at the Center of Nonlinear Studies (CNLS), Los Alamos National Laboratory (LANL). He then joined the complex systems group at LANL as member of the research staff. In 2004, he became the Deputy Director of the CNLS until his joining the department of physics at University of Notre Dame in 2006. His research is in the general areas of statistical physics and nonlinear dynamical systems, with topics including complex networks, fluid flows, population dynamics, epidemics, agent-based systems, game theory, brain neuronal systems and foundations of computing.

Brian Uzzi
http://www.kellogg.northwestern.edu/faculty/uzzi/htm/
Northwestern University, *uzzi@northwestern.edu*
Kellogg School of Management
PhD (1994), Sociology, State University of New York, Stony Brook

Brian Uzzi is the Richard L. Thomas Professor of Leadership and Organizational Change at the Kellogg School of Management, Northwestern University. He also co-directs NICO, the Northwestern Institute on Complex Systems, is the faculty director of the Kellogg Architectures of Collaboration Initiative (KACI), and holds professorships in Sociology at the Weinberg College of Arts of Sciences and in Industrial Engineering and Management Sciences at the McCormick School of Engineering. He is a globally recognized scientist, teacher, consultant and speaker on leadership, social networks, and new media. He has lectured and advised companies and governments around the world and been on the faculties of INSEAD, University of Chicago, and Harvard University. In 2007-2008, he was on the faculty of the University of California at Berkeley where he was the Warren E. and Carol Spieker Professor of Leadership.

Thomas Valente
https://ipr.usc.edu/faculty.php?faculty_id=46
University of Southern California, *tvalente@usc.edu*
Department of Preventive Medicine, Institute for Prevention Research, Keck School of Medicine
PhD (1991), Mass Communication, University of Southern California

Thomas W. Valente is a Professor in the Department of Preventive Medicine, Institute for Prevention Research, Keck School of Medicine, at the University of Southern California. He is author of Social Networks and Health: Models, Methods, and Applications (2010, Oxford University Press); Evaluating Health Promotion Programs (2002, Oxford University Press); Network Models of the Diffusion of Innovations (1995, Hampton Press); and over 165 articles and chapters on social networks, behavior change, and program evaluation. Valente uses social network analysis, health communication, and mathematical models to implement and evaluate health promotion programs designed to prevent tobacco and substance abuse, unintended fertility, and STD/HIV infections. He is currently working on specification for diffusion network models and implementing network interventions. Valente has received the Simmel Award from INSNA and the Rogers award from APHA. Valente received his BS in Mathematics from the University of Mary Washington, his MS in Mass Communication from San Diego State University, and his PhD from the Annenberg School for Communication at USC. From 1991 to 2000 he was at the Bloomberg School of Public Health. In 2008, he was a visiting senior scientist at NIH (NHGRI) for 6 months. In 2010-2011 he was a visiting Professor at the Ìacole des Haute Ìatudes en SantÌ© Publique (Paris/Rennes). Valente is co-editor (with Martin Everett) of Social Networks, and on the editorial boards of Network Science and the Journal of Health Communication.

Alessandro Vespignani
www.mobs-lab.org
Northeastern University, *alexvespi@gmail.com*
Network Science Institute
PhD (1994), Physics, University of Rome—La Sapienza

Dr. Vespignani is currently the Sternberg Family Distinguished University Professor at Northeastern University, where he is the founding director of the Northeastern Network Science Institute. Professor Vespignani received his undergraduate degree and Ph.D., both in physics, from the University of Rome –La Sapienza. He completed his postdoctoral research at Yale University and Leiden University. Professor Vespignani worked at the International Center for Theoretical Physics (UNESCO) in Trieste and at the University of Paris-Sud in France as a member of the National Council for Scientific Research (CNRS). From 2004 to 2011, Vespignani was J.H. Rudy Professor of Informatics and Computing at Indiana University and the founding Director of the Center for Complex Networks and Systems Research and Associate Director of the Pervasive Technology Institute. Vespignani is an elected fellow of the American Physical Society, member of the Academy of Europe, and a fellow of the Institute for Quantitative Social Sciences at Harvard University. Vespignani's current research focuses on the data-driven computational modeling of epidemic and spreading phenomena and the study of biological, social and technological networks.

Appendix II

Workshop Co-chairs

Albert-László Barabási, Northeastern University

Kate Coronges, Northeastern University

Alessandro Vespignani, Northeastern University

Observers

Jiwei Lu, OASDR&E (Basic Research Office)

Bob Bonneau, OASDR&E

Art Conroy, DIA

Purush Iyer, ARO

Edward Palazzolo, ARO

Ananthram Swami, ARL

Sara Rajtmajer, DARPA

Adam Russell, DARPA

David Sweeney, DTRA

Paul Tandy, DTRA

Bruce West, ARO

Ryan Zelnio, OASDR&E

Rapporteurs

Kate Klemic, Research Scientist
Virginia Tech Applied Research Corporation

Tom Hussey, Senior Consultant
Virginia Tech Applied Research Corporation

Matt Bigman, Research Analyst
Virginia Tech Applied Research Corporation

Kyle Doverspike, Project Manager
Virginia Tech Applied Research Corporation