

Can Network Visualization and Analysis Accelerate Medical Discoveries? Theoretical, Applied, and Funding Perspectives

Suresh K. Bhavnani^{1,2,3} PhD, Kevin Bassler⁴ PhD, Indra Neil Sarkar^{5,6,7} PhD MLIS,
Adi V. Gundlapalli^{8,9} MD PhD MS, Abdul R. Shaikh¹⁰ PhD MHSc

¹Institute for Translational Sciences, ²Preventive Medicine & Community Health, University of Texas Medical Branch, Galveston TX; ³School of BMI, University of Texas, Houston, TX; ⁴Physics Department, University of Houston, Houston, TX; ⁵Center for Clinical and Translational Science, ⁶Department of Microbiology and Molecular Genetics, ⁷Department of Computer Science, University of Vermont, Burlington, VT; ⁸Internal Medicine and Biomedical Informatics, University of Utah, UT; ⁹VA Salt Lake City Health Care System, UT; ¹⁰National Cancer Institute, National Institutes of Health, Rockville, MD

Abstract

The exponential increase of biomedical data provides new opportunities for the prevention, diagnosis, and treatment of diseases¹. Several fields like bioinformatics and social networking have exploited this opportunity through the use of networks to discover, verify, and validate patterns and their underlying mechanisms in molecular and social data. However, network-based approaches have not yet been fully exploited to analyze the wide range of medical information that is becoming available through electronic health records. This panel brings together theoretical, applied, and funding perspectives to discuss the opportunities and challenges of using networks to analyze medical data. Learning objectives include: research questions that are amenable to network analysis; strengths and limitations of different network representations; domain-dependent and -independent network concepts; complementary relationship between network analysis and biostatistics; skills required for effective network analysis; and funding opportunities for exploratory network analysis. This multi-perspective discussion will be aimed towards defining an agenda for the application of networks to accelerate discoveries in medical data.

Introduction

The rapid growth of biomedical data such as electronic health records has created new opportunities for the prevention, diagnosis, and treatment of diseases¹. However, the cognitive limitations of humans in comprehending large amounts of complex data pose a challenge for researchers. One approach to deal with such complexity has been the use of networks to help researchers discover patterns in complex data through visualizations, verify those patterns through graph-based and biostatistical methods, validate the patterns using random permutations of the data, and infer the mechanisms involved in generating the patterns. Such approaches have been used extensively in the bioinformatics and social networking communities resulting in discoveries ranging from molecular-based classification of diseases², and growth models to explain the topology of the web³.

However, besides a few notable examples, relatively few studies have used networks to analyze medical information such as in medical records. For example, networks have been used to analyze how obesity spreads through social networks⁴, and to analyze co-morbidity patterns in Medicare records⁵. This relative dearth of studies leveraging networks methodologies to analyze medical information could be due to hurdles such as the difficulty of transferring network concepts from social networking and bioinformatics to the medical domain, the complexity of current network analysis tools, the relative slow adoption of techniques to the medical domain, the evolution of health research training and policy, and difficulty in identifying appropriate funding opportunities. To explore and address such hurdles, this panel brings together theoretical, applied, and funding perspectives with the goal of accelerating medical discoveries through the use of current and future network analytical methods.

Position Statements of Panel Members

1. Network Goals: Discovery, Verification, Validation, and Inference (Suresh K. Bhavnani, PhD, Moderator)

Dr. Bhavnani, is associate professor of biomedical informatics at the Institute of Translational Science in the University of Texas Medical Branch, and holds an adjunct appointment at the School of Biomedical Informatics at the University of Texas in Houston. He specializes in network visualization and analysis of biomedical data, with application to the design of decision-support systems. His research in the use of networks has received several awards including a distinguished paper award⁶ at the AMIA fall symposium, and an outstanding paper award⁷ at the AMIA Summit of Translational Bioinformatics.

Dr. Bhavnani will introduce the fundamental elements of networks (requiring no prior background of networks), and discuss four goals of network analysis that have proved useful in the analysis of medical data: (1) **discovery of**

complex relationships through network visualizations and transformations; (2) **verification of patterns** through the use of graph-based and biostatistical measures; (3) **validation of patterns** through comparison with random permutations of the network; and (4) **inference of mechanisms** underlying the emergent patterns. Each of these goals will be illustrated through examples of completed network analysis projects involving medical data. He will conclude with a discussion on how networks complement methods from biostatistics and data mining, and the critical technical and human interaction skills required for conducting effective network analysis in collaboration with domain experts.

2. Network Theory: Inferring Structure and Function in Complex Phenomena (Kevin E. Bassler, PhD)

Dr. Bassler is professor of physics at the University of Houston, and specializes in the structural and dynamical properties of complex networks, with a focus on finding community structure in complex networks. He has received numerous awards for his research [e.g., 8] including a Sloan Research Fellowship, and for his teaching including an Australian Research Council International Fellowship.

Dr. Bassler will present an overview of the ongoing research within the network science community to infer the structure and function of complex networks from empirical data. He will then discuss existing methods of network inference such as the detection of communities or functional modules and their hierarchical relationship within networks. He will also discuss the limitations of existing network analytical methods, and the ongoing challenges to these problems, including their application to weighted and bipartite networks. Throughout his presentation, he will use networks of real data to illustrate the difference between domain-dependent and domain-independent network concepts.

3. Network Representation: Phylogenetic Trees and their Role in Phenomics (Indra Neil Sarkar, PhD, MLIS)

Dr. Sarkar is assistant professor of Microbiology and Molecular Genetics as well as Computer Science at the University of Vermont, where he is also the Director of Biomedical Informatics in the Center for Clinical and Translational Science. He has formal training in microbiology, biomedical informatics, and library science. An internationally recognized phylogeneticist and leader in the biodiversity informatics community, he was a research scientist at the American Museum of Natural History in New York City, and the Marine Biological Laboratory in Woods Hole. Much of his research [e.g., 9] is focused on the development of informatics approaches within the context of understanding organism speciation and genetic evolution of diseases.

Dr. Sarkar will introduce basic directed network representations, with a focus on the role of phylogenetic trees for exploring clinical phenotype data such as those recorded in electronic health records. Using existing electronic health record data, he will describe: (1) **knowledge representation** techniques for encoding clinical phenotypes so they can be subjected to phylogenetic inference algorithms; (2) **phylogenetic tree generation** methods such as *maximum parsimony* for exploring relationships between data that are extracted from electronic health records; and (3) **genotype-phenotype identification** based on correlating genomic data with putative disease patterns ascertained from electronic health record containing “phenotypes.” Dr. Sarkar’s presentation will conclude with practical considerations for using phylogenetic inference to analyze clinical and molecular data in the emerging field of phenomics.

4. Network Application: Spread of Infectious Diseases through Social Networks (Adi Gundlapalli, MD,PhD,MS)

Dr. Gundlapalli is assistant professor of internal medicine, pathology and biomedical informatics at the University of Utah School of Medicine, and staff physician-investigator at the VA Salt Lake City Health Care System in Salt Lake City, Utah. He specializes in treating infectious diseases, and modeling their spread in hospital settings such as emergency departments, and in community settings¹⁰.

Dr. Gundlapalli will discuss the application of social network analytical methods to the medical domain by describing: (1) **current literature** on models for the spread of infectious diseases through social networks; (2) **challenges** faced by clinicians when conducting network analysis, and the role of collaboration with network scientists; and (3) **future applications** of networks at the patient and population levels with the goal of modeling and controlling outbreaks of infectious diseases. He will conclude by discussing the emerging role of social networking applications as a means to affect health behavior change.

5. Network Funding: Opportunities and Strategies (Abdul R. Shaikh, PhD, MHSc)

Dr. Shaikh is a program director and behavioral scientist at the National Cancer Institute’s (NCI) Health Communication and Informatics Research Branch, where he directs research on behavioral informatics and

cyberinfrastructure for cancer prevention and control. He is an active participant in cross-agency initiatives in data harmonization, advanced computation, and cyber-enabled research, helps coordinate NCI's Small Business Innovations Research (SBIR) portfolio, and is lead guest editor of a recent special supplement on Cyberinfrastructure for Consumer Health¹¹.

Dr. Shaikh will discuss: (1) **NCI initiatives** in network and visual analytical methods including the PopSciGrid Community Health Data Portal (<http://cancercontrol.cancer.gov/hcirb/cyberinfrastructure/popsoci.html>); and related federal efforts in data harmonization and advanced computation for population health; (2) **research priorities and funding opportunities** including the recent NCI Developer Challenge tied to the Federal Open Government Plan and Community Health Data Initiative; and (3) **effective strategies** for applying to NIH funding mechanisms related to network and other visual analytical methods.

6. Open Panel Discussion

The position statements will help to frame a broader discussion with the attendees on concrete steps for overcoming theoretical, applied, and funding hurdles. We expect this discussion to include misconceptions related to network analysis, best practices for the acquisition of the technical skills to conduct network analysis, conceptual differences between data-driven versus theory-driven hypotheses generation, and funding strategies for exploratory network analysis. The goal of this discussion is arrive at a deeper understanding of how to ensure more widespread use of network analytics with the goal of accelerating medical discoveries. The panel discussion should be of benefit to: (1) **medical informaticians** and **domain experts** who have no prior background in networks but need to acquire an intuition about current and future methods; (2) **network researchers** who wish to understand the hurdles of using networks to analyze medical data in addition to funding opportunities, (3) **program managers** from funding agencies who wish to understand how best to support network analytics that have the potential of wide impact in making medical discoveries.

Acknowledgements

This research is funded by NIH grants 1U54RR02614 UTMB CTSA(ARB), 1R21OH009441-01A2, R01LM009725 (INS), NIAID U01 AI074419-01, 5P01HK000069.

References

1. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington DC: National Academies of Science Press, 2001.
2. Chuang H, Lee E, Liu Y, Lee D and Ideker T. Network-based classification of breast cancer metastasis. *Molecular Systems Biology* 2007, 3:141.
3. Barabási, A-L, Albert, R. Emergence of scaling in random networks. *Science* 286, 509–512 (1999).
4. Christakis, NA, Fowler, JH. The Spread of Obesity in a Large Social Network Over 32 Years. *New England Journal of Medicine* 2007, 357 (4): 370–379.
5. Hidalgo CA, Blumm N, Barabási A-L, Christakis NA. A Dynamic Network Approach for the Study of Human Phenotypes. *PLoS Comput Biol*, 2009, 5(4).
6. Bhavnani, S.K., Abraham, A., Demeniuk, C., Gebrekristos, M., Gong, A., Nainwal, S., Vallabha, G.K., and Richardson, R.J. Network Analysis of Toxic Chemicals and Symptoms: Implications for Designing First-Responder Systems. *AMIA Annu Symp Proc.* 2007:51-55.
7. Bhavnani, SK, Eichinger, F, Martini, S, Saxman, P, H.V. Jagdish, and Kretzler, M. Network Analysis of Genes Regulated in Renal Diseases: Implications for a Molecular-Based Classification. *BMC Bioinformatics* 10(Suppl 9):S3. (Journal version of paper at conference).
8. Y. Sun, B. Danila, K. Josic, and K.E. Bassler. Improved Community Structure Detection using a Modified Fine-Tuning Strategy, *EPL* 86, 28004 (2009).
9. Sarkar IN, Egan MG, Coruzzi G, Lee EK, DeSalle R. Automated simultaneous analysis phylogenetics (ASAP): an enabling tool for phylogenomics. *BMC Bioinformatics.* 2008 Feb 19;9:103.
10. Gundlapalli A, Ma X, Benuzillo J, Pettey W, Greenberg R, Hales J, Leecaster M, Samore M. Social network analyses of patient-healthcare worker interactions: implications for disease transmission. *AMIA Annu Symp Proc.* 2009 Nov 14;2009:213-7.
11. Shaikh, A.R., Prabhu Das, I., Vinson, C., Spring, B. Cyberinfrastructure for Consumer Health. Special Supplement, *American Journal of Preventive Medicine*, April 2011.

The first author affirms that all panel members have agreed to participate, and have contributed to the preparation of this document.