

Social Network Analysis to Assess the Impact of the CTSA on Biomedical Research Grant Collaboration

Radhakrishnan Nagarajan, Ph.D.^{1,2}, Charlotte A. Peterson, Ph.D.^{2,3}, Jane S. Lowe, Ph.D.², Stephen W. Wyatt, D.M.D.^{2,4}, Timothy S. Tracy, Ph.D.^{2,5}, and Philip A. Kern, M.D.^{2,6}

Abstract

Success of the Clinical Translational Science Award (CTSA) program implicitly demands team science efforts and well-orchestrated collaboration across the translational silos (T1–T4). Networks have proven to be useful abstractions of research collaborations. Networks provide novel system-level insights and exhibit marked changes in response to external interventions, making them potential evaluation tools that complement more traditional approaches. This study is part of our ongoing efforts to assess the impact of the CTSA on Biomedical Research Grant Collaboration (BRGC). Collaborative research grants are a complex undertaking and an outcome of sustained interaction among researchers. In this report, BRGC networks representing collaborations among CTSA-affiliated investigators constructed from grants management system data at the University of Kentucky across a period of six years (2007–2012) corresponding to pre- and post-CTSA are investigated. Overlapping community structure detection algorithms, in conjunction with surrogate testing, revealed the presence of intricate research communities rejecting random graphs as generative mechanisms. The deviation from randomness was especially pronounced post-CTSA, reflecting an increasing trend in collaborations and team-science efforts potentially as a result of CTSA. Intercommunity cross talk was especially pronounced post-CTSA. *Clin Trans Sci* 2015; Volume 8: 150–154

Keywords: social networks, grant collaboration, team science, CTSA, random graphs

Introduction

There has been an increase in team science approaches across a spectrum of disciplines.^{1–5} Solutions to many of the previously intractable health issues facing society especially demand collaborative interactions of researchers from multiple disciplines who can bring to bear a holistic approach to solving these multifaceted problems. Extramural funding agencies, such as the National Institutes of Health, recognize the need for interdisciplinary, multidisciplinary, cross-disciplinary, and team-science initiatives. The Clinical Translational Science Award (CTSA) funding mechanism from the National Center for Advancing Translational Sciences (NCATS) is multidisciplinary in nature and its success demands well-orchestrated collaboration across the translational silos (T1–T4)⁶ within and between CTSA. Since its inception in 2006, the number of CTSA awardees has steadily grown to 62 across 31 states. The University of Kentucky (UK) received NIH/NCATS funding in 2011. Evaluation of the performance and successes of institutions holding these awards is critical in assessing the success of the CTSA funding mechanism. There are several facets to the CTSA, including research, training, and services, encouraging novel quantitative approaches for CTSA evaluation that complement more classical evaluation techniques.⁷

Network analysis has enjoyed considerable attention from both the CTSA Evaluation and Informatics Working Groups under the broad theme social network analysis (SNA). While the former group investigates the choice of network analysis for evaluation, the latter develops tools to enhance collaboration within and between CTSA (e.g., VIVO, Harvard Profiles, CTSAconnect). Networks are composed of nodes and edges representing the entities of interest and their associations. Unlike reductionist representations, networks have the ability to provide novel system-level insights on the impact of interventions as a whole. Interventions are usually accompanied by changes in statistical properties and topology of networks making them useful for

evaluation (pre-, poststudies). Networks can be modeled from distinct data sources such as surveys, publications, and extramural grant funding.^{8–10} Social network analysis using surveys has been used by CTSA to investigate communication patterns among CTSA cores/personnel. A recent study identified publications and grants as metrics of success in translational research and training.¹¹ There have also been instances where grants and publications are investigated together.⁸ Collaborative grants are an outcome of longstanding successful research collaboration often preceded by publications. Publications also serve as preliminary findings in a grant proposal. This study is in line with our ongoing efforts to better understand the temporal evolution of Biomedical Research Grant Collaboration (BRGC) networks^{9,10} in CTSA settings. Nodes in the BRGC network are composed of Principal Investigators and Co-Investigators with an edge representing their collaboration on an extramurally funded research grant. In order to establish the nexus between CTSA (intervention) and changes in BRGC networks (outcome), we restrict the personnel in BRGC networks to only those who have been involved with CTSA. However, it is important to note that BRGC networks are open systems in the sense they are prone to external perturbations such as institutional policies, vision, and economic slowdown. Temporal evolution of BRGC networks is complex, often accompanied by addition and deletion of nodes and edges in a nonuniform manner.^{9,10} Therefore, any conclusion on incremental collaboration with increasing number of nodes is neither trivial nor straightforward. Addition of nodes reflects involvement of new personnel in the CTSA efforts, whereas deletion is due to departure of existing personnel. Deletion of influential nodes in the BRGC networks can result in considerable fragmentation of the network into isolated clusters. Addition of an edge can be attributed to formation of new collaborations whereas deletion corresponds to termination of existing collaboration. Addition

¹Division of Biomedical Informatics, College of Public Health, University of Kentucky, Lexington, Kentucky, USA; ²Center for Clinical and Translational Sciences, University of Kentucky, Lexington, Kentucky, USA; ³College of Health Sciences, University of Kentucky, Lexington, Kentucky, USA; ⁴Preventive Medicine and Environmental Health, College of Public Health, University of Kentucky, Lexington, Kentucky, USA; ⁵Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, Kentucky, USA; ⁶Division of Endocrinology and Molecular Medicine, College of Medicine, University of Kentucky, Lexington, Kentucky, USA.

Correspondence: Radhakrishnan Nagarajan (rnagarajan@uky.edu)

DOI: 10.1111/cts.12247

and deletion of edges has the potential to connect or disconnect research clusters. While social network analysis metrics such as centrality measures are routinely used to quantify the extent of collaborations, this study investigates the presence of communities and their temporal evolution. Communities arise when groups of nodes are more highly connected to each other than to other nodes in a network^{12–19} and are expected across multidisciplinary research projects supported by the CTSA. This study uses overlapping community structure detection algorithms,^{19,20} in conjunction with random graph surrogates, to identify nontrivial communities pre- and post-CTSA. Intercommunity cross talk and its variation pre- and post-CTSA are also investigated in order to evaluate the impact of CTSA on research productivity at UK.

Methods

BRGC network abstraction

Biomedical research collaboration data sets from 2007 to 2012 were obtained from the Offices of the Vice President for Research and Sponsored Projects at UK after obtaining the required consent. Attributes retrieved from the grant database included *Grant ID*, *Personnel ID*, *Role*, *Department*, *College*, and *Year*. Each grant is provided with a unique number represented by the *Grant ID*. A grant can have multiple personnel participating in various roles. These are captured by the attributes *Personnel ID* and *Role*. In this study, we restricted the *Role* to Principal Investigator (PI) and Co-Investigator (Co-I) in a given grant. PI and Co-I represent noteworthy collaborations and have a significant impact on the success of a grant. These roles are usually justified as a part of the grant submission. Restricting the nodes to PI and Co-I also prevents the BRGC network from becoming sparse and disconnected. The attributes *Department* and *College* correspond to Primary Academic Department and College affiliation since faculty can have secondary appointments across multiple departments. The attribute *Year* corresponds to the year in which the grant was awarded. The initial CTSA funding was awarded to UK in 2011. In order to investigate the impact of the CTSA on grant collaborations, we investigated BRGC networks pre- and post-CTSA. The grants considered in this study predominantly correspond to biomedical research grants including those awarded by the National Institutes of Health (NIH). BRGC networks by very definition are directed graphs where the direction is always from the Principal Investigator to the Co-Investigator. Since the Co-I's existence in a given grant is dependent on the PI, edges in the BRGC networks can also be thought of as representing causal associations between the nodes. The degree of a node in the BRGC network represents the extent to which that node is connected to others. Nodes can have high in-degree, out-degree, or both. Nodes with large in-degree usually correspond to dominant Co-I's whereas those with large out-degree corresponds to dominant PIs. Since PI/Co-I can participate in multiple roles across different grants, cycles are unavoidable in BRGC networks. In order to strengthen the nexus between CTSA as a cause and the observed changes in grant collaboration an effect, we restricted the *Personnel ID* to only those that are involved with the CTSA, that is, they received funding, training, or services from the UK CTSA-funded Center for Clinical and Translational Sciences (CCTS). A systematic approach is presented to investigate the evolution of BRGC networks across six years (2007–2012) corresponding to pre-CTSA (2007–2010), the year CTSA was awarded (2011), and post-CTSA (2012).

Community structures in BRGC networks

Recent studies have clearly demonstrated the existence of inherent communities in BRGC networks,^{9,10} where the nodes in the network have a tendency to cluster into groups. Community structure detection algorithms are useful in identifying such inherent clustering and fall under two broad categories: (i) those that identify nonoverlapping communities where nodes belong to only a single community¹² and (ii) those that identify overlapping communities where nodes can belong to multiple communities.^{14,19} While it is tempting to partition the BRGC network into isolated nonoverlapping communities, such a partitioning is unrealistic in translational settings that demand enhanced interaction across groups for research, training, and services. The extent of overlap among communities is an indicator of interdisciplinary collaboration. Constraining the nodes in the BRGC network to be members of nonoverlapping communities undermines the collaborative potential of a node and can adversely affect the conclusions. This study investigates community structure in CTSA BRGC networks using link-community algorithm,^{19,20} where the edges are aggregated hierarchically, to reveal overlap between communities.

Communities of BRGC networks and those of random graphs

Previous studies have compared the evolution of CTSA and non-CTSA groups.⁸ However, such a comparison can be challenging due to widely varying sizes of these two groups. Bootstrapping these groups to control for discrepancies in their size while helpful, may result in elimination of critical nodes and can have a pronounced effect on the overall topology. In this study, communities in the BRGC networks were compared to those generated by certain types of random graphs using a surrogate testing approach^{21–24} to evaluate the presence and evolution of community structures. The discriminant statistic in the surrogate testing approach was chosen as the maximum partition density.¹⁹ Two different null hypotheses (H_0^{ER} , H_0^{ER}) were investigated. H_0^{ER} addressed the null hypothesis that the community structure in the BRGC networks was generated from classical Erdos-Renyi (ER)²⁵ random graphs that retain the number of nodes and edges of the given BRGC network. H_0^{DD} addresses a relatively sophisticated null hypothesis that the community structure in the BRGC networks was generated from random graphs that retains the degree distribution of the given BRGC network. The deviation of the community structure in the BRGC networks from those of its random surrogate counterparts were captured with $S = \frac{|m_{\text{orig}} - \mu_{\text{surr}}|}{\sigma_{\text{surr}}}$ where m_{orig} represents the estimate of the discriminant statistic on the empirical sample, $(\mu_{\text{surr}}, \sigma_{\text{surr}})$ represent the mean and standard deviation of the discriminant statistic estimated across n_s independent surrogate realizations. The number of surrogate realizations was fixed at $n_s = 99$ in this study. While $S > 2$ is traditionally deemed significant,^{22,23} we use S as a relative index to compare the extent of deviation from random graphs pre- and post-CTSA.

Intercommunity and intracommunity cross talk

Individuals in a BRGC network can contribute to facilitating collaboration within (intracommunity cross talk) as well as between (intercommunity cross talk) communities. We are especially interested in intercommunity cross talk where an investigator from a given department facilitates collaboration across at least two distinct communities. Such personnel are deemed critical

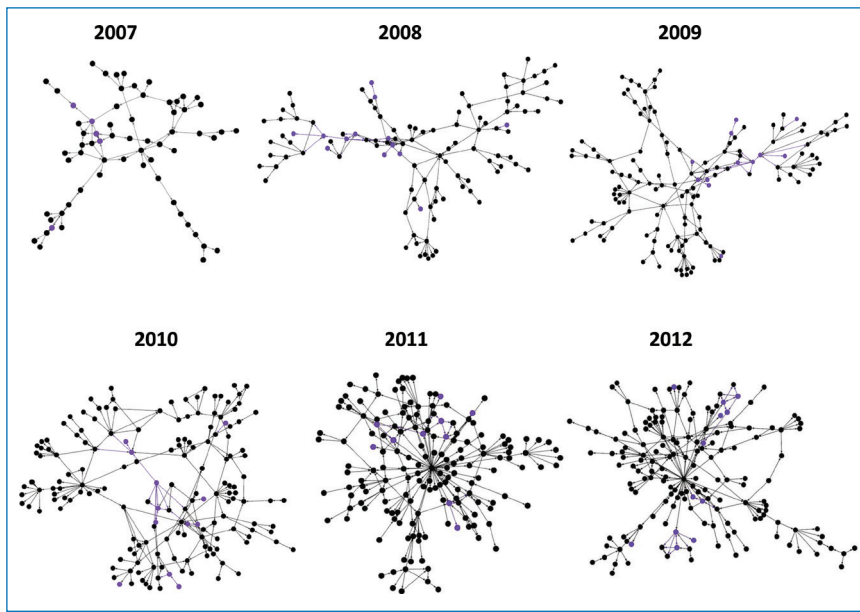


Figure 1. Formation of intricate community structures revealed by Yifan-Hu visualization of the BRGC network connected component from (2007 to 2012) corresponding to the pre-CTSA (2007–2010), the year CTSA was awarded (2011) and post-CTSA (2012). The blue circles represent the evolution of Behavioral Science department in the BRGC network.

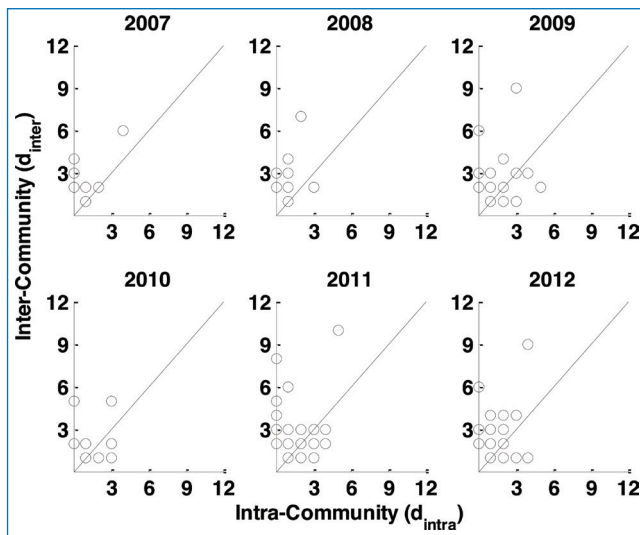


Figure 2. Scatter plot representing intracommunity and intercommunity cross talk from 2007 to 2012. The diagonal represents the line of separation between intracommunity and intercommunity cross talk.

players in cross-fertilization of research themes with the potential to form new campus-wide areas of research excellence. Consider a BRGC network $G(V, E)$ where V represents the number of nodes and E the edges or grant collaborations. Each node in G can belong to only one department (primary department). The nodes V in turn can be mapped onto departments $d_i, i = 1 \dots N$. The total number of occurrences d_i^{tot} of department d_i across the communities consists of occurrences within community d_i^{intra} as well as between community d_i^{intra} , that is, $d_i^{tot} = d_i^{intra} + d_i^{inter}$. Departments with $d_i^{tot} = 1$ are dropped from any further analysis since they did not facilitate cross talk.

Results

The number of nodes and edges for the BRGC networks over the 6 years were (2007: 215 nodes, 136 edges; 2008: 231 nodes, 160 edges; 2009: 277 nodes, 215 edges; 2010: 285 nodes, 246 edges; 2011: 299 nodes, 269 edges; and 2012: 310 nodes, 253 edges), reflecting increased participation in CCTS-supported research with time. However, as noted earlier, increasing the number of nodes does not necessarily translate into increasing collaborations or formation of community structures. The BRGC networks were comprised of a connected component, isolated singleton nodes, and small isolated clusters. To assess collaboration, we focused analysis on the connected component. The nodes and edges of the connected component were (2007: 89 nodes, 123 edges; 2008: 109 nodes, 148 edges; 2009: 141 nodes, 196 edges; 2010: 168 nodes, 230 edges; 2011: 169 nodes, 251 edges; 2012: 166 nodes, 234 edges) also increased across the years, indicating a predominant number of CCTS investigators were connected either directly or indirectly.

Yifan-Hu visualization²⁶ of the giant component generated using Gephi 0.8.2 (<http://gephi.org/>)²⁷ revealed characteristic communities from 2007 to 2012 with more intricate structures post-CTSA (Figure 1). Temporal evolution of nodes corresponding to one of the departments (Behavioral Science) across the six years is also shown in Figure 1. The presence of a connected component and community structures pre-CTSA reflects collaborative efforts that were likely essential for successful execution of the CTSA application. Degree centrality distribution of the BRGC networks were positively skewed indicating the presence of a few highly connected nodes comprising the tail of the distribution. Highly connected nodes consisted of those with high indegree as well as those with high outdegree. While the former consisted of personnel who provide services across multiple grants as Co-I, such as biostatisticians, the latter consisted of PIs of multidisciplinary research grants and Center grants.

Overlapping community structures in BRGC networks

As described in the methods, this study used S estimate as the relative index representing the deviation of the community structure in the BRGC networks from those of its random graph surrogate counterparts. The S estimates obtained from ER surrogates were markedly high across the years ($S_{2007}^{ER} = 9.5$; $S_{2008}^{ER} = 11.6$; $S_{2009}^{ER} = 10.2$; $S_{2010}^{ER} = 10.1$; $S_{2011}^{ER} = 10.4$; $S_{2012}^{ER} = 10.6$), demonstrating inherent community structure significantly different from those of ER random graphs. This is to be expected given the positively skewed degree centrality distributions of the BRGC networks with a few nodes exhibiting large collaborative potential. A similar analysis of BRGC networks using degree preserving type random graph (DD) surrogates across the years resulting in the following S estimates ($S_{2007}^{DD} = 3.6$; $S_{2008}^{DD} = 4.3$; $S_{2009}^{DD} = 3.6$; $S_{2010}^{DD} = 5.1$; $S_{2011}^{DD} = 6.6$; $S_{2012}^{DD} = 6.2$). Of interest is the marked increase in the S estimates in 2011 and 2012 in contrast to pre-CTSA (2007–2010). These results confirm the formation of more intricate, overlapping community structures upon CTSA

funding, shown visually in *Figure 1*. A scatter plot of $(d_{\text{intra}}, d_{\text{inter}})$, shown in *Figure 2*, revealed a shift towards increased inter-community cross talk across years 2011 and 2012 in contrast to pre-CTSA (2007–2010).

Discussion

Beginning in 2005, the University of Kentucky (UK) made a substantial commitment to building/enhancing its infrastructure to support clinical and translational research efforts. This investment resulted in successful competition for a CTSA planning grant in 2006 to support an evolving CCTS. The planning grant propelled campus and state efforts culminating in receipt of a CTSA award in 2011. This analysis investigates the evolution of communities in BRGC networks pre- and post-CTSA at UK. The grant collaboration data were obtained through the UK's Office of Sponsored Projects Administration. The study focused only on investigators associated with the CCTS in order to establish possible cause-effect between the observed changes in community structure and CTSA as an intervention. Our analysis revealed the CTSA-affiliated investigators to be organized into nontrivial community structures different from those that can be generated using random graph models. This was established using statistically sound surrogate testing approaches. The extent of deviation from random graph models was quantified by the parameter sigma. The extent of deviation from the classical ER type random graphs was pronounced pre- and post-CTSA as expected, rejecting ER random graph models as possible generative mechanisms of the observed community structures in the BRGC networks. Subsequently, surrogate testing also rejected the null hypothesis that the community structure in the BRGC networks was similar to those of degree preserving random graphs pre- and post-CTSA. However, the S estimates exhibited a marked increase across 2011 and 2012 (post-CTSA) indicating formation of more intricate communities post-CTSA. Cross talk of the departments within and between communities was subsequently investigated. A scatter plot of the intra- and intercommunity cross talk revealed a shift towards intercommunity cross talk post-CTSA compared to pre-CTSA. The dominant collaborative units across the years were the Markey Cancer Center Core Facility and Biostatistics from the Colleges of Medicine and Public Health, respectively. The College of Medicine's Departments of Behavioral Science and Medicine were also dominant across the years; however, more funded collaborations extended to other departments through interactions with the CCTS. More importantly, faculty collaborating across departments covered T1 through T3 translational areas. Examples of newly formed and overlapping research communities are apparent throughout the network. For example, in 2009, an investigator in the Sanders-Brown, Center on Aging had a funded collaboration with another investigator outside of his department, the Department of Neurology, which had been the case since 2007. The CCTS Clinical Services/Regulatory Core began to support his clinical studies, and he was awarded a CCTS pilot award in 2011 to begin a new clinical trial. This has enabled him to expand his funded collaborations to include the Colleges of Nursing and Public Health, and has resulted in an initiative across academic institutions in the region to assess and address the health disparities related to aging and cognition in Appalachia. The impact of the CTSA funding was also apparent across networks of established investigators. A senior investigator in the Department of Psychology (College of Arts and Sciences) had a well-established collaborative

network with faculty in the Colleges of Medicine (Behavioral Science), Pharmacy and Public Health from 2007 to 2009. Through interaction with the Drug Discovery and Development Core of the CCTS, by 2012, his collaborations increased with those units, and extended his funded collaborations to existing research communities in the College of Medicine's Departments of Radiology, Neurology, and Anatomy. In summary, our analysis revealed the community structures of BRGC networks across CTSA-affiliated investigators to be different from those of random graphs. The strength of deviation from the random graphs and intercommunity cross talk was especially pronounced post-CTSA potentially as a result of increased interactions and environment created through CTSA funding at the University of Kentucky.

Acknowledgment

This work was supported by award number 1UL1TR000117 from the National Center for Advancing Translational Sciences (NCATS/NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.

References

1. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. *Science*. 2007; 316(5827): 1036–1039.
2. Jones BF, Wuchty S, Uzzi B. Multi-university research teams: shifting impact, geography, and stratification in science. *Science*. 2008; 322(5905): 1259–1262.
3. Falk-Krzesinski HJ, Börner K, Contractor N, Fiore SM, Hall KL, Keyton J, Spring B, Stokols D, Trochim W, Uzzi B. Advancing the science of team science. *Clin. Transl. Sci.* 2010; 3(5): 263–266.
4. Scott CS, Nagasawa PR, Abernethy NF, Ramsey BW, Martin PJ, Hacker BM, Schwartz HD, Brock DM, Robins LS, Wolf FM, et al. Expanding Assessments of Translational Research Programs: Supplementing Metrics With Value Judgments. *Eval. Health Professions*. 2013; 37: 83–97. doi: 0163278713500984.
5. Börner K, Contractor N, Falk-Krzesinski HJ, Fiore SM, Hall KL, Keyton J, Spring B, Stokols D, Trochim W, Uzzi B, et al. A multi-level systems perspective for the science of team science. *Sci. Transl. Med.* 2010; 2(49): 49cm24–49cm24.
6. Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet. Med.* 2007; 9(10): 665–674.
7. Rubio DM, Sufian M, Trochim WM. Strategies for a national evaluation of the Clinical and Translational Science Awards. *Clin. Transl. Sci.* 2012; 5(2): 138–139.
8. Hughes ME, Peeler J, Hogenesch JB. Network dynamics to evaluate performance of an academic institution. *Sci. Transl. Med.* 2010; 2(53): 53ps49–53ps49.
9. Nagarajan R, Kalinka AT, Hogan WR. Evidence of community structure in biomedical research grant collaborations. *J. Biomed. Informat.* 2013; 46(1): 40–46.
10. Nagarajan R, Lowery CL, Hogan WR. Temporal Evolution of Biomedical Research Grant Collaborations Across Multiple Scales—A CTSA Baseline Study. Paper presented at: AMIA annual symposium proceedings, Washington, D.C.: 2011.
11. Lee LS, Pusek SN, McCormack WT, Helitzer DL, Martina CA, Dozier AM, Ahluwalia JS, Schwartz LS, McManus LM, Reynolds BD, et al. Clinical and translational scientist career success: metrics for evaluation. *Clin. Transl. Sci.* 2012; 5(5): 400–407.
12. Newman ME, Girvan M. Finding and evaluating community structure in networks. *Phys. Rev. E*. 2004; 69(2): 026113.
13. Clauset A, Newman ME, Moore C. Finding community structure in very large networks. *Phys. Rev. E*. 2004; 70(6): 066111.
14. Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D. Defining and identifying communities in networks. *Proc. Natl. Acad. Sci. U.S.A.* 2004; 101(9): 2658–2663.
15. Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*. 2005; 435(7043): 814–818.
16. Pons P, Latapy M. Computing communities in large networks using random walks. *Computer and Information Sciences-ISCIS Istanbul, Turkey*: Springer; 2005: 284–293.
17. Fortunato S, Barthelemy M. Resolution limit in community detection. *Proc. Natl. Acad. Sci.* 2007; 104(1): 36–41.
18. Evans T, Lambiotte R. Line graphs, link partitions, and overlapping communities. *Phys. Rev. E*. 2009; 80(1): 016105.
19. Ahn Y-Y, Bagrow JP, Lehmann S. Link communities reveal multiscale complexity in networks. *Nature*. 2010; 466(7307): 761–764.

- 20.** Kalinka AT, Tomancak P. Linkcomm: an R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics*. 2011; 27(14): 2011–2012.
- 21.** Weigend AS, Gershenfeld NA. Time series prediction: forecasting the future and understanding the past. Paper presented at: Santa Fe Institute Studies in the Sciences of Complexity, Santa Fe, New Mexico, May 14–17, 1992 (Reading, MA: Addison-Wesley, c1994, edited by Weigend, Andreas S.; Gershenfeld, Neil A.1994), 457–474.
- 22.** Schreiber T, Schmitz A. Discrimination power of measures for nonlinearity in a time series. *Physical Review E*. 1997; 55(5): 5443.
- 23.** Schreiber T, Schmitz A. Surrogate time series. *Phys. D: Nonlinear Phenomena*. 2000; 142(3): 346–382.
- 24.** Nagarajan R. Local analysis of dissipative dynamical systems. *Int. J. Bifurcation Chaos*. 2005; 15(05): 1515–1547.
- 25.** Erdos P, Renyi A. On random graphs I. *Publ. Math. Debrecen*. 1959; 6: 290–297.
- 26.** Hu Y. Efficient, high-quality force-directed graph drawing. *Math. J.* 2006; 10(1): 37–71.
- 27.** Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. 2009.
-