TheATLAS

The Structure of Distributed Scientific Research Teams Affects Collaboration and Research Output

Sarah Gehlert, PhD[1,2*]**, Jung Ae Lee, PhD**[3]**, Jeff Gill, PhD**[2,4]**, Graham Colditz, MD**[2,3]**, Ruth Patterson, PhD**[5]**, Kathryn Schmitz, PhD**[6]**, Linda Nebeling, PhD, RD**[7]**, Frank Hu, MD**[8]**, Dale McLerran**[9]**, Diana Lowry, MPH**[9]**, the TREC Collaboration and Outcomes Working Group, Mark Thornquist, PhD**[9]**,** [1] *The George Warren Brown School, Washington University, Saint Louis, MO;* [2] *Siteman Cancer Center, Washington University, Saint Louis, MO;* [3] *Agriculture Statistics Lab, University of Arkansas, Fayetteville, AR;* [4] *Department of Political Science, Washington University,St. Louis, MO;* [5]*Moores Cancer Center, University of California at San Diego, La Jolla, CA;* [6]*Clinical Epidemiology and Biostatistics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA;* [7] *National Cancer Institute, Bethesda, MD;* [8] *Harvard T.H. Chan School of Public Health, Harvard University, Boston MA;* [9] *Fred Hutchinson Cancer Research Center, Seattle, WA; Sarah Gehlert, College of Social Work, The University of South Carolina, Pendleton, Columbia, SC, Email: sgehlert@mailbox.sc.edu.*

To understand how the nature of scientific collaboration between individuals and sites in team-based research initiatives affect collaboration and research output, we examined four waves of prospective survey data to measure collaboration across investigators, disciplines, and sites to measure structural determinants of research success. 116 investigators in the five sites of the NIH-funded U54 Transdisciplinary Research on Energetics and Cancer (TREC) initiative were surveyed about their research ties with a 2011 baseline measure and followed by three additional iterations and augmented by bibliometric data. Social network analysis describes the changing structure of contact and cooperation. We found that the network structure of a team science project affects the nature and rate of publications, implying that funded projects vary in research output based on how investigators interact with each other and that the design of scientific research projects affects research output by determining levels of contact between actual and potential collaborators.

Keywords: cancer; research; transdisciplinarity; team science; network models.

1 Introduction

Current approaches to biomedical research rely increasingly on cross-disciplinary collaboration. In addition, recent funding announcements emphasize team science and call for the development of multi-institutional collaborations or hubs that promote team science. Yet despite this emphasis, questions remain about whether the benefits of a team science based approach to research outweigh the operational and transactional costs.

Beginning in the late 1990s, the National Cancer Institute (NCI) incentivized team science through three multi-site initiatives, Transdisciplinary Tobacco Use Research Centers (TTURC), Centers for Population Health and Health Disparities (CPHHD), and Transdisciplinary Research in Energetics and Cancer (TREC).

In the present paper, we perform a social network analysis of four waves of internal survey data from TREC investigators augmented by bibliometric data to determine the extent to which collaboration across investigators, disciplines, and sites affects the timing, rate, and type of publications. This approach is unique as the first longitudinal examination of an NIH transdisciplinary initiative and as the first study allowing comparison between investigators' subjective views of collaboration and objective counts and types of shared publication. Our research question is whether active efforts by a funded research initiative to foster investigator ties (network "edges") within and across sites will affect the density, centrality, and homophily of those ties at the level of the initiative, in turn speeding the onset of publication and the diversity of disciplines and institutions represented in the authorship of those publications. Our comparison is with available data from two other transdisciplinary sites funded by NIH.

NIH Transdisciplinary Initiatives and their Evaluation

In 1999, the NIH established the first of three multi-site transdisciplinary research initiatives, the Transdisciplinary Tobacco Research Use Centers (TTURC) [1]. The Centers for Population Health and Health Disparities (CPHHD) [2] followed in 2003, and the Transdisciplinary Research on Energetics and Cancer (TREC) Initiative (http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-10-006.html) was funded in 2005.

TTURC was the first to undergo evaluation of its impact on research output, as measured by the timing of publications. A retrospectively-conducted bibliographic evaluation of TTURC in comparison with R01s found that after an initial two to three-year lag period, TTURC had higher overall publication rates over a 10-year comparison period [3]. In a companion article to the bibliographic evaluation, Rimer and Abrams cautioned that the ultimate impact and value-added nature of transdisciplinary frameworks may be decades away due to the time that it takes to establish effective team functioning [4].

CPHHD, the second NIH transdisciplinary initiative to be funded, focuses on the determinants of health disparities and translation of this knowledge into solutions. CPHHD is unique among NIH's transdisciplinary initiatives in its focus on local communities with high rates of disparities and its mandate to partner with these communities at all stages of the research process. Okamoto conducted an analysis of one wave of survey data from the ten-year initiative in 2015 [5] and found a similar lag period as seen in TTURC.

The TREC Initiative (2010-2015)

In 2005, TREC was funded by NCI as NIH's third transdisciplinary initiative, in response to increasing evidence of the contributions of nutrition, energetics, and physical activity to cancer incidence, morbidity, and mortality. The NCI Request for Applications (RFA) (http://grants.nih.gov/grants/rfa-files/RFA-CA-10-006.html) describes the TREC initiative's purpose as "to foster collaboration across multiple disciplines and encompasses projects that cover the biology, genomics, and genetics of energy balance to behavioral, socio-cultural, and environmental influences on nutrition, physical activity, weight, energetics, and cancer risk."(p. 1).

The TREC Coordination Center and Dr. Gehlert began prospectively gathering data to evaluate the transdisciplinary efforts of the TREC at the beginning of its second cycle of funding, in 2011. These data were supplemented data with an analysis of prospectively-obtained bibliometric data.

The TREC sites (the University of Pennsylvania, Washington University in St. Louis, the University of California-San Diego, Harvard University, and the Coordination Center at Fred Hutchinson Cancer Research Center in Seattle) comprise more than 120 investigators and over 30 clinical, biologi-

cal, behavioral, and social science disciplines, with inevitable fluctuation over time due to investigator moves. Their scope extends from the biology, genetics, and genomics of energy balance to social and behavioral influences on physical activity and nutrition, weight, energetics, and cancer risk [6]. In part because of the emphasis on the specialized technology inherent in research on energetics and a reliance on animal models, systematic efforts were made to maximize resources by fostering ties across sites. This was incentivized by developmental awards requiring participation by junior and senior investigators across sites that were managed by the TREC Coordination Center. TREC was the only one of the three NIH transdisciplinary initiatives to have a Coordination Center as part of its structure.

The TREC initiative aims to accelerate scientific discovery through a *transdisciplinary* approach to team science, which Rosenfield defines as exchanging information, altering discipline-specific approaches, sharing resources, and integrating disciplines to achieve a common scientific goal [7]. Transdisciplinary research differs from multidisciplinary and interdisciplinary research in the extent to which investigators operate outside the boundaries of their own disciplines to share language, pool knowledge and theories, and develop new methods of analysis. It is generally considered to represent the highest degree of disciplinary collaboration [8].

In a 2008 article introducing the TREC initiative in a special issue of the Journal of Preventive Medicine, Robert Croyle, who heads NCI's Division of Cancer Control and Population Sciences, wrote that "one important assumption underlying these efforts was that the speed of scientific progress and its effective application to public health problems would depend on the integration of discipline-specific efforts and increased support for collaboration, evidence synthesis, and the science of dissemination" [9]. Thus, integration among collaborators is seen as a measure of the success of transdisciplinary team science initiatives like TREC. In another article on team science in that same year, Karen Emmons (now VP, Director of the Kaiser Foundation Research Institute) was quoted as saying "*among* the most important indicators of success is rich team communication" [10].

2 Methods

Study Sample

Survey participants were investigators involved in the five TREC sites from 2010 to 2015. A list of investigators was developed by the TREC Steering Committee and after receiving appropriate IRB approvals, each was sent a letter inviting them to participate along with a copy of the social network survey. The 2011 survey established a baseline measure of ties after the first year of funding. The survey was re-sent yearly to assess the degree to which the density of social network ties changed over time during the height of the grant activity. Because it was sent late in 2011, the next survey was received early in 2013, 13 months after the previous year's survey. Thus, no data are available for 2012, and the time between waves was 13 months rather than 12 months. In 2013, we limit the invitations only to those who responded in 2011 and remained active in TREC. Fewer invitations were sent in 2014 and 2015 due to turnover of faculty over time, which is a typical phenomenon in such projects. The TREC Coordinating Center collected publication data on a regular basis and examined authorship in terms of the disciplines and sites represented. In the current paper, we report on data from the first four waves of data collection. The response rate of the survey is approximately 80 percent in each year, although the absolute number of respondents decreased slightly over the period of study. The number of respondents by academic position and site are summarized in Table 1.

Measures

Collaboration Network. The survey listed the names of all TREC investigators and asked individual respondents if they currently worked with or had worked with prior to TREC each investigator on the list: (1) on a study or grant; (2) on a co-authored publication; (3) on a co-authored presentation; (4) in mentoring or training; (5) on a committee or work group; or, (6) in any other activity. These are the conditions for interacting and thus forming an edge (tie) in the researcher network. Thus, the *behavior* that we study here is two researchers deciding to collaborate in one of these six ways. Note that we are not measuring the quality of these ties, nor the resulting groups from collections of ties, except that a grant award or publication is clearly an indication of edge effectiveness.

Discipline. From a list of 37 academic disciplines, investigators were asked to choose the one that best characterized the disciplinary perspective of their work. For purposes of analysis, responses were collapsed into eight disciplines (Table 2). The subgroups of disciplines and the distribution of disciplines by site are described in Table 3.

Social Network. We consider a social network to be a relational structure among actors. We define individual TREC investigators who participated in our survey to be a set of actors (also called nodes), and assign a tie (as in "tie-together") if there are collaborations between those investigators. These ties between two actors are called arcs or edges. Considering the group of TREC investigators as a social network, or a social entity made up of a number of actors, allows the group's structure to be analyzed in its entirety as well as the dyadic relationships between its members. The underlying assumption is that more frequent communication within and across sites will better foster advances in science, specifically in this case energy balance and cancer. For the purposes of analysis, we define communication as the quality of dyadic and group interactions beyond what is measured in six survey criteria for an edge. We also specify an undirected network, so that if actor 1 is tied to actor 2, the reverse is also true, because we consider peer-level scientific collaborations to be overwhelmingly mutual rather than hierarchical in this context.

Network Size. The network size is the number of actors in the network (i.e., the total number of members in the network).

Network Density. More ties between investigators imply greater network interaction, as defined above. The principal way of evaluating this quality is by measuring the density of social network ties, defined as the number of actual ties between network members compared to the number of potential ties (equal to $(n^2 - n)/2$ for n individuals in an undirected network). Denser networks suggest faster propagation of information and greater group cohesion [11]. Also, individuals who conduct more information tend to be more active in terms of research goals and objectives [12].

Triads. Triangle relationships occur when two individuals with a tie have a tie with the same third individual, and the number of triangles can be greater than the number of direct ties [13], especially in network with multiple attributes and diverse individual backgrounds like TREC. Four types of triadic relations occur in undirected networks: no ties (three actors are isolated without ties/edges), a single tie between two actors while the third actor is isolated, two ties among three actors, or all three ties forming a triangle. Counting the number of these types across all possible triples, a so-called triad census [14], allows us to better understand the local social structure [15], which may not be captured by global measures or dyad density. Relative prevalence of these triangles implies that interpersonal choices tend to be mutual and transitive [15].

Centrality. Another traditional way to evaluate a network is through the node centrality [16]. While the density of a network is a global measure to understand the overall network function, centrality evaluates the power and influence of each node on social relations. For example, some actors in a network are highly central while others are not widely connected. Large differences in centrality for a given network tend to produce hierarchical structures with isolated individual actors at the periphery: a core-periphery structure [17]. We first measure *degree centrality*, which shows how many actors are tied directly to each actor. The measure is computed by counting the number of adjacencies for an actor in a network of size n, that is,

$$degree(i) = \Sigma_{j=1}^n I_{ij} \qquad (1)$$

where I_{ij} is 1 if the actor i and j are connected, 0 otherwise. Our second measure of centrality is *closeness*. This measure provides an index of independence or efficiency [16], meaning the speed with which an actor reaches other actors in the network. The computation of closeness is based on summing the shortest paths (called *geodesics*) from an actor to all other actors in the network, that is,

$$closeness(i) = \Sigma_{j=1}^n (1/d_{ij})(n-1) \qquad (2)$$

where d_{ij} is the geodesic distance between i and j. This closeness score is rescaled between 0 and 1. It is 0 if an actor is an isolate, and 1 if an actor is directly connected to all others.

We also measure centrality by the location of actors of high academic rank (full professors), which can be seen as a measure of the effect of *seniority*. Our third centrality measure is *betweenness*, which quantifies the number of times of an actor being situated as a bridge along the shortest path between

two other actors, to assess how likely this actor is to be a direct route between two actors that are not linked otherwise. This measure is computed by

$$Betweenness(k) = \Sigma\Sigma_{i<j}^{n} g_{ij}(k)/g_{ij} \qquad (3)$$

where $g - ij$ is the number of the shortest paths linking i and j, and $g_{ij}(k)$ is the number of the shortest paths linking i and j that contain k.

Homophily. Similar actors tend to bond with each other for many reasons such as opportunity, affinity, ease of communication, reduced transaction costs, and organizational foci [18-21]. We assume that a tendency toward homophily that exists in the TREC network consists of eight subgroups by discipline and five subgroups by site, within which network members are easily accessible to each other. An effective way to view homophily is through the exponential-family random graph model (ERGM) [22, 23]. The model suggests the probability to be connected between i and j individuals in the network. A tendency toward homophily indicates a higher probability of ties being formed between actors within the same site or within the same discipline.

Number of Publications and Authorship. As per the initial NCI Request for Proposals (http://grants.nih.gov/grants/rfa-files/RFA-CA-10-006.html), we consider the establishment of collaborations among disciplines (i.e., biology, genomics, and genetics of energy balance and behavioral, socio-cultural, and environmental factors that determine cancer risk) as a key measure of TREC's success. In addition to measuring these collaborations subjectively, we measured collaborations within and between TREC sites in terms of the authorship of those papers across disciplines and sites. Examining the production of papers across time allowed us to investigate whether the lag time seen in TTURC and CPHHD was also true of TREC.

Although others have measured impact of publications using citation counts [24], we did not consider the four years of our study sufficient time for citations to accrue in a way that would reflect impact. We thus considered citation data to be a less reliable measure of research output. It will be the subject of future papers as the opportunity for TREC work to be cited increases over time. Data were collected from the Coordination Center at Fred Hutchinson Cancer Research Center through the Annual Progress Report to the National Cancer Institute and the citations of the TREC Center grant number in PubMed. We explored the authorship of each paper reported retrospectively in terms of the disciplines and sites involved.

3 Statistical Analysis

Social network analysis was used as the principal mode of analysis in the present study. All statistical results and graphs were analyzed using R version 3.0.3 [25], and the *network, sna,* and *ergm* R packages [26-28] were specifically used.

4 Results

The network properties of the TREC research sites over years are summarized in Table 2 Note that the network size shown in Table 2 is different from the number of respondents for the survey reported in Table 1. The difference between the number of respondents and the total number of network actors is due to the secondary actors who are included in the network because, although they did not themselves respond to the survey, they were designated by respondents as a link. Thus, we do not know about the relationship among the secondary actors. We take this into account in our analyses because the network density can be sensitive to the number of respondents rather than the network size, which often makes it hard to fairly compare the longitudinal social network data due to different sets of respondents each time point.

The density of TREC network is 0.086 in 2011 and remains similar or slightly decreases over time (Table 2). The relatively lower density 0.082 in 2015 does not necessarily mean less collaboration, because we observe more ties in 2015 than in 2011 (547 vs. 415). The lower density means that there are fewer social ties among actors relative to the chances of those ties occurring, which can occur simply because a network is large. As seen in Table 1, the network size was noticeably smaller in the first year of the grant (n=99), when researchers were building new relationships, than in subsequent years when it stabilized in terms of total participants (n=116, 114, 116). Network ties expanded rapidly in the early period, going from 414 in 2011 to 577 in 2013. However, network ties then stabilized in the mid-500s (t=577, 525, 547). Note that these are not

equilibrium values in the traditional long-run sense since the study contains only four waves.

Triad census is reported in Table 2. Triangle relationships increased over time, from 592 in 2011 to 1175 in 2015. These numbers are translated to the *Transitivity Index 1*, by dividing by all possible triads, and *Transitivity Index 2*, by dividing by the number of one-tie triads. Both transitivity indexes increased in 2015. The second index implies that the characteristics of the small group relationships have been changing from "couple only" to "triangle" in general.

Transitivity of interpersonal choices forms a more clustered structure, as evident in Figure 4. In a transitivity network, the existence of the two ties, a↔b and b↔c, will increase the probability of another tie a↔c, which represents the closure of the triangle. Transitivity therefore leads to the larger connected groups by particular clustering members [14, 15], which are a small number of highly active members surrounded many much less active members creating more triangles than a network with ties distributed uniformly randomly.

The density by each site is displayed in Table 2. A decreasing trend in density is most noticeable in Sites 1 and 2. This occurs because of the lower number of survey respondents in later years and because the total TREC social network has built up more cross-site relationships over time as researchers from different institutions collaborated more closely.

In Table 2, we report the mean and standard deviation (sd) of three types of centrality measures. In terms of *degree centrality*, we see an average of eight to nine edges for an actor per year, as shown in the measure of the degree 8.38, 9.95, 9.21, 9.43 in each year. In terms of *seniority*, we find that the degree of the group of investigators of high academic rank is greater than the network average for all members. We likewise see that it increased over the period of the study, which indicates that senior faculty play a central role as coordinators or gatekeepers amongst the total investigators [29]. As can be seen in Table 2, the scores do not differ markedly among actors, as indicated by low standard deviations in all years. Each member thus has a similar level of dependency (or efficiency) to connect to every network member. In Table 2, we see a large standard deviation of the *betweenness* among actors. This indicates that a small set of actors perform key brokerage functions within the network [29]. Individual differences can

be seen in Figs. 1-4, where the size of the node indicates the degree of the ties for each investigator. Although this refers to the degree, it indirectly shows the betweenness distribution among the actors, since these two centrality measures are highly correlated.

In Fig. 1, each node indicates an individual investigator, and is colored according to discipline, with sites labeled by the letters A to E. We see that the network is clustered by site, and there are key players in each site with large sized circles for greater relative connectedness. For instance, in site A, some investigators in epidemiology and biochemistry/genetics are shown to be key actors in the local network. However, a member in exercise and physiology, located in the center of site B, is similarly highly connected. In site C, the epidemiology and social behavior science fields are at the center of the diagrams. In site D and E, each actor shows widely spread, but more diffuse connectedness in the local network. Therefore, some disciplines are more influential in some sites than others. Combined, these observations imply the collaborations *across* the sites implies cross-disciplinary interactions between geographically distinct leaders.

Homophily. From Table 2, the lines of the homophily, we report the results from two separate ERGM models across both site and discipline. Not surprisingly, there is a homophily effect by both site and discipline over four years, as shown by a small p-value (<0.01), meaning that actors are more likely to make ties with those in their own site than those in other sites. A similar interpretation is possible for disciplines. Also, the probability of collaboration within site was 0.311 in 2011, which is higher than the overall density 0.086: the density can be interpreted as the probability of ties of two actors in general. However, we notice that the probability of ties within sites monotonically decreases over time to 0.219 in 2015. Within-site homophily tendency decreases as cross-site collaborations increase. Likewise, the probability of collaboration with someone from the same discipline was 0.145 in 2011, which is higher than overall density 0.086, after which the number decreases over time. The homophily by discipline is relatively less strong than homophily by site, suggesting that the "transdisciplinary" team mission has been well performed since the beginning of the TREC.

Number of publications. Publication numbers by year are summarized in Fig. 5. The number of

publications increased in later years of TREC as we expected from the expanding collaborations in TREC centers. Furthermore, we calculated the number of papers in which the coauthors were from multiple disciplines and from multiple sites. These numbers are also reported in Fig. 5. Note that information about discipline of authors of papers was not universally accessible. We therefore used information collected in our longitudinal survey and our central TREC directory of TREC investigators, for which only 40% of disciplines are known. This may have resulted in an underestimate of the number of disciplines per publication. In counting multi-site publications, we included both TREC sites and non-TREC sites because the coauthors from non-TREC sites indicate extended collaborations with TREC and spinoff from TREC projects. Having either multi-site or multi-disciplinary authors on a publication is common across all years.

Nature of Collaboration

Perhaps most striking is the rapid development of cross-site ties. In Table 4, we reported the proportion of cross-site versus within-site ties. In 2011, the cross-site ties only account for 15.66% of total ties, increasing over time to 39.67% in 2015. This phenomenon is shown graphically through Figs. 1-4. The network in Fig. 1 is clustered by site, whereas the network in Fig. 4 exhibits a well-mixed cluster of sites. This occurs because cross-site collaborations increased over the years. The number of cross-site ties after the first year was twice as high for TREC than for CPHHD [5], the only other NCI-funded initiative for which they were measured, during the first year of its second round of funding, the only year in which CPHHD network ties were measured. In terms of type of collaboration, we see that in 2011, grant and committee collaborations were most prevalent, while co-authored publications were most prevalent from 2013 to 2015.

5 Discussion

Several things are notable in terms of the collaborations among TREC investigators over time and the research output that accrued. In terms of subjective reports of collaborations, we see that the number of network ties increased over time. These network ties are reflected in an objective count of the number of peer-reviewed publications with authors from multiple disciplines and publications that cross sites. This suggests that TREC was successful as defined in the original NCI Request for Applications for TREC (http://grants.nih.gov/grants/rfa-files/RFA-CA-10-006.html).

Our comparison of network ties is with the one year of data collected from CPHHD. We found many more cross-site ties after the first year of TREC than were observed after the first year of CPHHD. In retrospect, we anticipate that this likely was due to the use of the TREC Coordination Center as a natural network hub for TREC, an unanticipated consequence of the use of a cross-site coordinating mechanism. Neither TTURC nor CPHHD had such a center as part of its structure. The Coordination Center fostered cross-site ties through establishing and maintaining mechanisms such active monthly steering committee teleconference calls with representatives from each site and NCI, cross-site working groups representing multiple disciplines on topics such as cancer disparities, biomarkers, and measures of physical activity, and designated cross-site developmental awards that privileged junior and senior investigators working together, mandating that more than one site be represented.

Our results show two main differences between TREC and the other two NIH transdisciplinary initiatives. In comparison to the CPHHD, which also compared within-site and cross-site network ties after its first year of functioning, TREC had many more cross-site than within-site ties. We expect that this is due to CPHHD's mandate to partner with communities, which drew investigators inward, while TREC investigators' emphasis on technology caused us to look across sites for resources to share. At the same time, the Coordination Center fostered and facilitated the sharing of resources and expertise across sites. Secondly, because TREC was able to promote and support work across disciplines and site, in part through the Coordination Center, TREC did not experience the two to three-year lag in publications to the same degree as TTURC and CPHHD [2-4].

Both our survey and the bibliometric results consistently show that TREC collaborations have been growing and emblemize successful team science in public health and medical research. Both cross-discipline and cross-site collaborations contributed to this growth. Importantly, the analysis here suggests that multi-site team science initiatives are more

likely to foster greater collaboration and cooperation when they are designed to be transdisciplinary from the start, whereas one previous study [24] found that across all types of biomedical studies increasing physical distance between investigators is a major deterrent to scientific impact. In the current environment of shrinking grant support, funding agencies may want to consider focusing on transdisciplinary team science as a way to increase research success for a given level of funding.

In analyzing the network, our study has some limitations since it is a longitudinal study. Losing or adding members in a network in successive years makes some measures less comparable. Lower density can occur simply because of a different size of network or a heterogeneous set of actors of each year. Also note that the density measure is based on dyad relationships, ties between two actors, which may not capture the local social structure like a triangle. Rather than relying solely on results of density, we also looked at the nature of relationships through a triad census and the exponential random graph model. The TREC network is a good example of change in that the overall density is similar over a given year, but the nature of local relationships can change. Another issue may come from the inclusion of the secondary actors who are designated by the survey respondent but did not participate in survey. The number of secondary actors were 21 out of 99 in 2011, 40 out of 116 in 2013, 55 out of 114 in 2014 and 59 out of 116 in 2015. More secondary actors that we have more missing, and therefore the overall edges are underestimated by the survey. Fortunately, we could locate more defined groups, notably triangles, in the network with the most secondary actors [29], buffering our general conclusions.

Our analyses of TREC data are ongoing. The number, pace, and nature of publications are important and visible academic outcomes for any scientific team, as are citations of those publications. As stated earlier, while we considered it premature to analyze article citation data, we intend to do so retrospectively after publications have had time to collect citations in the normal fashion. In the meantime, our network survey data are highly nuanced and provide extensive insight into dyadic and group collaborations as well as network dynamics, improving upon previous investigations from NIH-funded projects.

References

[1] Abrams, D. B., Leslie, F., Mermelstein, R., Kobus, K., & Clayton, R. R. (2003). Transdisciplinary tobacco use research. *Nicotine and Tobacco Research*, 5(Suppl 1), S5S10.

[2] Warnecke, R. B., Oh, A., Breen, N., Gehlert, S., Paskett, E., Tucker, K. L, Lurie, N., Rebbeck, T., Goodwin, J., Flack, J., Srinivasan, S., Kerner, J., Heurtin-Roberts, S., Abeles, R., Tyson, F. L., Patmios, G., & Hiatt, R. A. (2008). Approaching health disparities from a population perspective: The National Institutes of Health Centers for population health and health disparities. *American Journal of Public Health,* 98(9), 16081615.

[3] Hall, K. L., Stipelman, B. A., Morgan, G., Moser, R. P., Berrigan, D., Marcus, S. E., Vogel, A. L., Masimore, B., Stokols, D., & Feng, A. (2012). Assessing the value of team science: A study comparing center- and investigator-initiated grants.*American Journal of Preventive Medicine,* 42(2), 157163.

[4] Rimer, B. K., & Abrams, D. B. (2012). Present and future horizons for transdisciplinary research. *American Journal of Preventive Medicine,* 42(2), 200201.

[5] Okamoto, J. (2015). The Centers for Population Health and Health Disparities Working Group, Scientific collaboration and team science: A social network analysis of the centers for population health and health disparities. *Translational Behavioral Medicine*, 5(1), 12-23. doi: 10.1007/s13142-014-0280-1.

[6] Patterson, R. E., Colditz, G.A., Hu, F. B., Schmitz, K. H., Ahima, R. S., Brownson, R. C., Caron, K., Chavarro, J., Chodosh, L., Gehlert, S., Gill, J., Glanz, K., Haire-Joshu, D., Herbst, K. L., Hoehner, C. M., Hovmand, P. S., Irwin, M. L., Jacobs, L. A., James, A. S., Jones, L. W., Kerr, J., Kibel, A. S., King, I. B., Ligibel, J. A., Meyerhardt, J. A., Natarajan, L., Neuhouser, M. L., Olefsky, J. M., Proctor, E. K., Redline, S., Rock, C. L., Rosner, B., Sarwer, D. B., Schwartz, J. S., Sears, D. D., Sesso, H. D., Stampfer, M. J., Subramanian, S. V., Taveras, E. M., Tchou, J., Thompson, B., Troxel, A. B., Wessling-Resnick, M., Wolin, K. Y., & Thornquist, M. D. (2013). The 2011-2016 Transdisciplinary Research in Energetics and Cancer (TREC) initiative:

Rationale and design. *Cancer Causes and Control,* 24(4), 695-704. doi: 10.1007/s10552-013-0150-z.

[7] Rosenfield, P.L. (1992). The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Social Science and Medicine,* 35(11), 1343-1357.

[8] Gehlert, S. J., Hall, K., Vogel, A., Hohl, S., Hartman, S., Nebeling, L., Redline, S., Schmitz, K., Thornquist, M., Patterson, R., & Thompson, B. (2014). Advancing transdisciplinary research: The Transdisciplinary Research on Energetics and Cancer Initiative. *Journal of Translational Medicine and Epidemiology*, 2(2), 1032.

[9] Croyle, R. T. (2008). The National Cancer Institute's transdisciplinary centers initiative and the need for building a science of team science. *American Journal of Preventive Medicine,* 35(2), S90.

[10] Dingfelder, S. F. (2007). Team science. *Monitor on Psychology*, 38(1), 20-21.

[11] Aboelela, S. W., Merrill, J. A., Carley, K. M., & Larsen, E. (2007). Social network analysis to identify an interdisciplinary research center. *Journal of Research Administration*, 38(1), 61.

[12] Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036-1039. doi: 10.1126/science.1136099

[13] Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography,* 46(1), 103125.

[14] Davis, J. A. & Leinhardt, S. (1972). The Structure of positive interpersonal relations in small groups. *Sociological Theories in Progress*, 2: 218-251.

[15] Holland, P. W., & Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology,* 7, 1-45.

[16] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215-239. doi:10.1016/0378-8733(78)90021-7.

[17] Borgatti, S. P., & Everett, M. G. (1999). Models of core/periphery structures. *Social Networks*, 21, 375-395.

[18] Feld, S. L. (1982). Structural determinants of similarity among associates. *American Sociological Review*, 47, 797-801.

[19] McPherson, J. M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.*, 27: 415-444.

[20] Centola, D., Gonzalez-Avella, J. C., Eguiluz, V. M., & San Miguel, M. (2007). Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution,* 51(6), 905-929. doi: 10.1177/0022002707307632.

[21] Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology,* 37, 131-153. doi: 10.1146/annurev.soc.012809.102709.

[22] Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association,* 81(395), 832-842. doi: 10.1080/01621459.1986.10478342.

[23] Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2017). An introduction to exponential random graph models for social networks. *Social Networks*, 29(2), 173-191. doi: 10.1016/j.socnet.2006.08.002.

[24] Lee, K., Brownstein, J. S., Mills, R. G., & Kohane, I.S. (2010). Does collocation inform the impact of collaboration? PLoS ONE, 5(12), e14279.

[25] R Core Team. (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical, Computing. Retrieved from http://www.R-project.org/.

[26] Butts, C. T. (2008). network: A Package for Managing Relational Data in R. *Journal of Statistical Software*, 24(2), 1-36.

[27] Butts, C. T. (2008). Social network analysis with sna. *Journal of Statistical Software*, 24(6), 1-51.

[28] Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software,* 24(3), 1-29.

[29] Gehlert, S., Carothers, B. J., Le, J. A., Gill, J., Luke, D, & Colditz, G. A. (2015). A social network analysis approach to diagnosing and improving the functioning of transdisciplinary teams in public health. *Transdisciplinary Journal of Engineering and Science,* 6, 11-22.

About the Authors

Sarah Gehlert, PhD is Dean of the College of Social Work at the University of South Carolina and University of South Carolina Educational Foundation Distinguished Professor. She is the E. Desmond Lee Professor of Racial and Ethnic Diversity Emerita at Washington University

(WU) in St. Louis. Prior to her move, Dr. Gehlert was the Co-Program Leader of the Prevention and Control Program of the Alvin J. Siteman Cancer Center, Co-Director of the Transdisciplinary Center on Energetics and Cancer (TREC), and Training Program Director of the Program for the Elimination of Cancer Disparities (PECaD). Dr. Gehlert served as Senior Advisor of the Center for Community-Engaged Research of the Institutes of Clinical and Translational Sciences. Dr. Gehlert was a member of the Board of Scientific Counselors of the National Human Genome Research Institute at NIH until January 2016. She is a member of the Council for Extramural Grants at the American Cancer Society and of the Steering Committee of the California Breast Cancer Research Program. She serves as President of the American Association of Social Work and Social Welfare.

Jung Ae Lee, PhD is Assistant Professor of Agricultural Statistics Laboratory at the University of Arkansas Fayetteville. Dr. Lee received a Ph.D. in Statistics from University of Georgia and has several years of research experience in applied statistics in the medical sciences, including post-doctoral work at the Washington University School of Medicine. Dr. Lees collaborative work includes medical experiments or behavioral studies for disease prevention and control. Dr. Lee has expertise in analyzing social network data and assessing the research performance in transdisciplinary research team science.

Jeff Gill, PhD is Distinguished Professor of Government, Professor of Mathematics & Statistics, and a member of the Center for Behavioral Neuroscience, all at American University. He has done extensive work in the development of Bayesian hierarchical models, nonparametric Bayesian models, elicited prior development from expert interviews, as well in fundamental issues in statistical inference. He has extensive expertise in statistical computing, Markov chain Monte Carlo (MCMC) tools that he uses contribute to empirical knowledge in the biomedical and social sciences. Current theoretical work focuses on developing hybrid algorithms for statistical estimation with multilevel specifications and complex time-series and spatial relationships.

Graham Colditz, MD, Associate Director for Prevention and Control at Siteman Cancer Centerr , led the TREC at Washington University in St Louis. He has longstanding interests in etiology and prevention of obesity and cancers caused by obesity. In 2016, he Chaired the IARC update of the Handbook on Cancer Prevention; volume 16, weight control.

Ruth Patterson, PhD is Professor Emeritus of University of California, San Diego (UCSD). Previously, she served as Associate Director of Population Sciences and Program Leader of the Cancer Prevention Program of the Moores Cancer Center. She was PI of the NCI designated TREC at UCSD focused on the insulin resistance, inflammation, obesity and the risk of breast cancer. She was leader of a TREC project to investigate the degree to which metformin, a lifestyle intervention, or both, can reduce the risk of breast cancer mortality among overweight/obese, postmenopausal breast cancer survivors.

Kathryn Schmitz, PhD is a Professor of Public Health Sciences and Associate Director of Population Sciences at the Pennsylvania State Cancer Institute. She served as Pi of the TREC at the University of Pennsylvania. She is an exercise interventionist who has led many large randomized trials. Dr. Schmitz has published over 190 peer reviewed scientific papers (scopus h-index of 50) and has had continuous NIH funding for her research since 2002. In spring 2017, she was elected president-elect of the American College of Sports Medicine.

Linda Nebeling, PhD, RD is the Deputy Associate Director of the Behavioral Research Program (BRP), Division of Cancer Control and Population Sciences (DC-CPS), National Cancer Institute (NCI). She completed a Dietetic Internship Program at Memorial Sloan Kettering Cancer Center, New York. Research interests are in transdisciplinary science, energy balance, obesity and cancer. She has received the Conrad A. Elvehjem Award for Public Service in Nutrition by the American Society for Nutrition; eight NCI Merit Awards and one NIH Directors Award. In 2001, she was made a Fellow of the Academy of Nutrition and Dietetics.

Frank Hu, MD is Chair of Department of Nutrition, Fredrick J. Stare Professor of Nutrition and Epidemiology at Harvard T.H. Chan School of Public Health and Professor of Medicine, Harvard Medical School and Brigham and Womens Hospital. He received his MD from Tongji Medical College in China and a PhD in Epidemiology from the University of Illinois at Chicago. His research is focused on nutritional and lifestyle epidemiology and prevention of obesity and cardiometabolic diseases, gene-environment interactions, and nutritional metabolomics. Dr. Hu is a member of the U.S. National Academy of Medicine.

Dale McLerran has an M.S in Biomathematics and over 25 years of experience providing statistical and data management support at the Dartmouth Hitchcock Medical Center and Fred Hutchinson Cancer Research Center. Mr. McLerran has expertise in the analysis of data with correlated error structures and in providing innovative analyses for nontraditional statistical problems. He consults with scientific investigators on study design issues and provides statistical analyses for publications and reports. His current primary responsibilities include support for the Validation and Comparison of Biomarkers for the Early Detection of Colorectal Adenocarcinoma (GLNE 010) study as well as the Population-based Research Optimizing Screening through Personalized Regimens (PROSPR) studies.

Diana Lowry, MPH now serves as Program Director for the TREC Training Course, an NCI-funded 5-year grant (R25CA203650, PI Melinda Irwin), and as consultant and special volunteer for the NCI where supports obesity and

cancer research. Previously, Diana managed the TREC
Coordination Center where she integrated efforts and
facilitated transdisciplinary research across the network of
multiple TREC Research Centers. TREC was a multi-site
consortium involving scientists from multiple disciplines
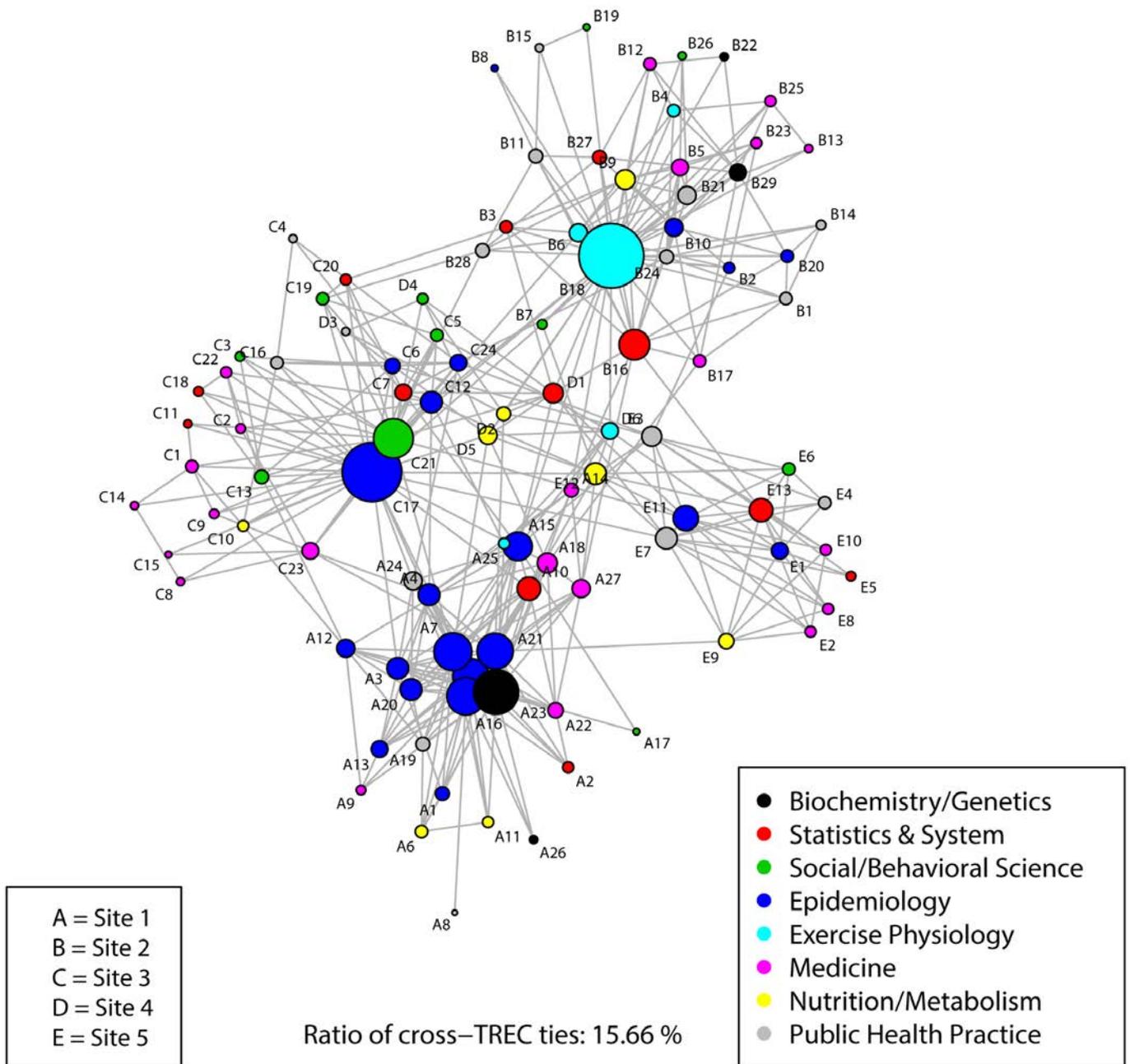

Figure 1: Network plot of the five TREC sites in 2011. Each node indicates individual investigator, which is colored by different disciplines and labeled by the letters A to E for different sites. The size of node indicates the degree of the ties that each investigator has. It is clearly seen that the network is clustered by site.
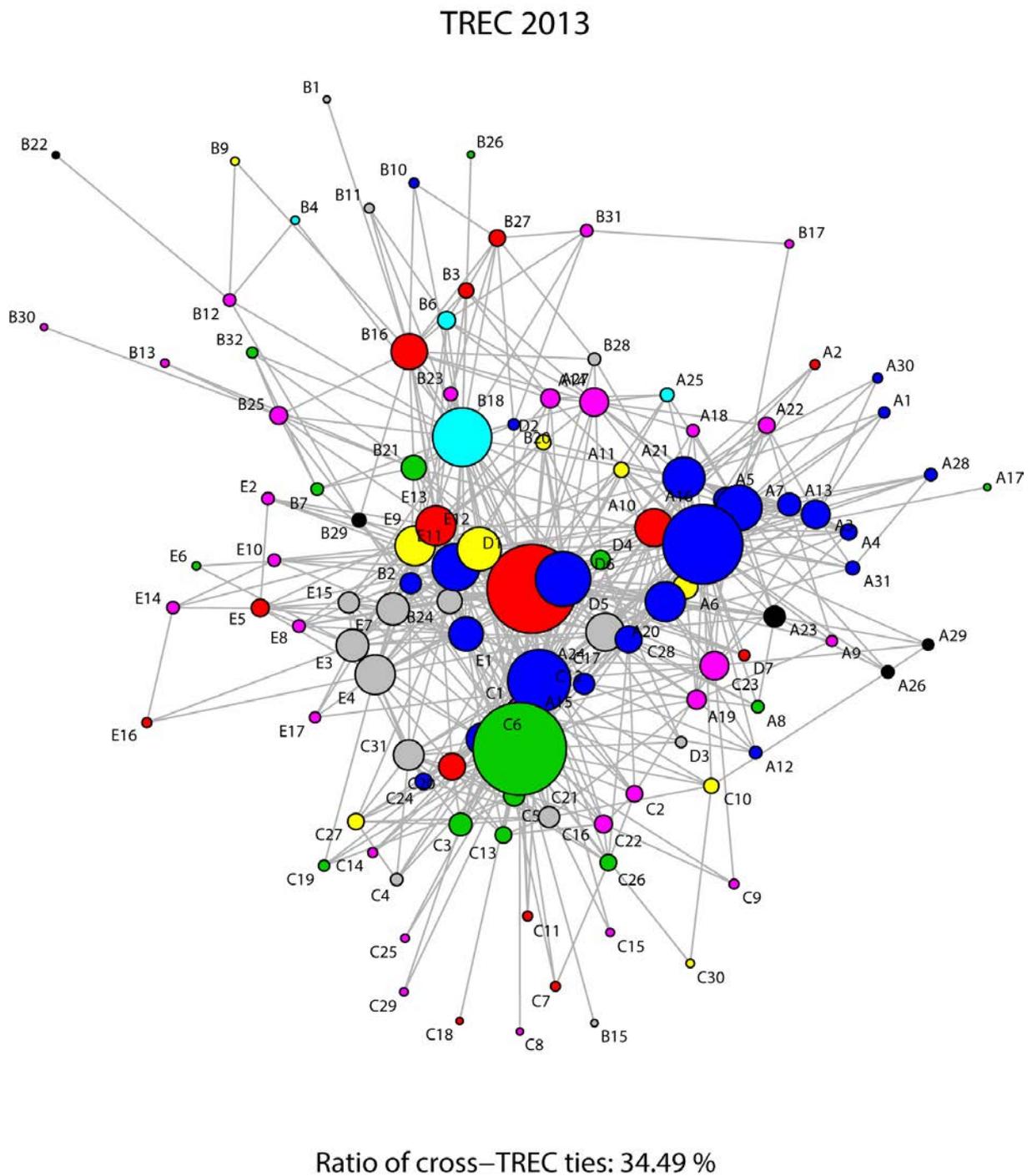
Figure 2: Network plot of the five TREC sites in 2013. This network exhibits a well-mixed cluster of sites compared to Fig. 1. This occurs because cross-TREC collaborations have increased to 34.49% in 2013 from 15.66% in 2011.
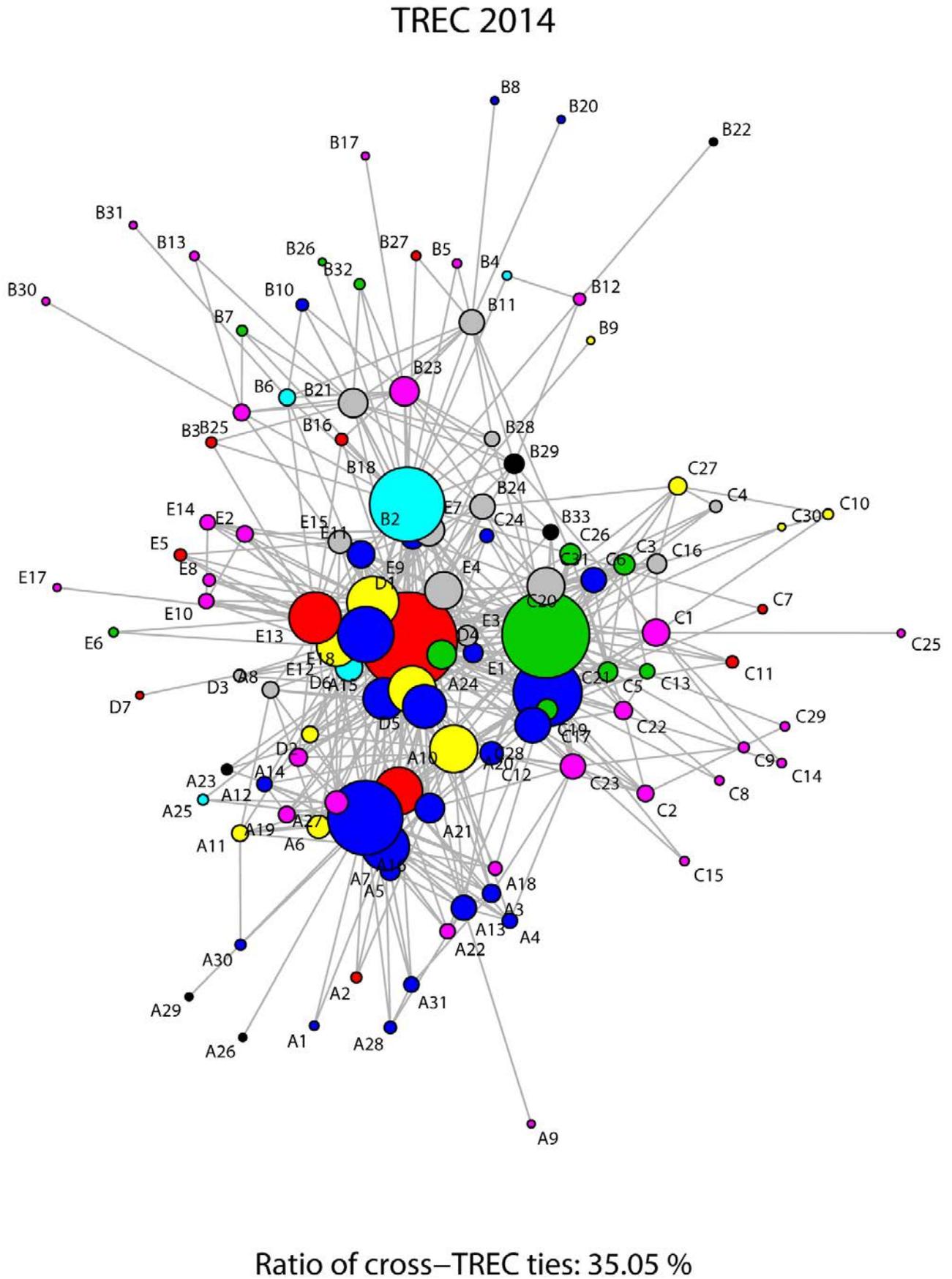
Figure 3: Network plot of the five TREC sites in 2014. The cross-TREC collaborations have slightly increased to 35.05% in 2014 from 34.49% in 2013.
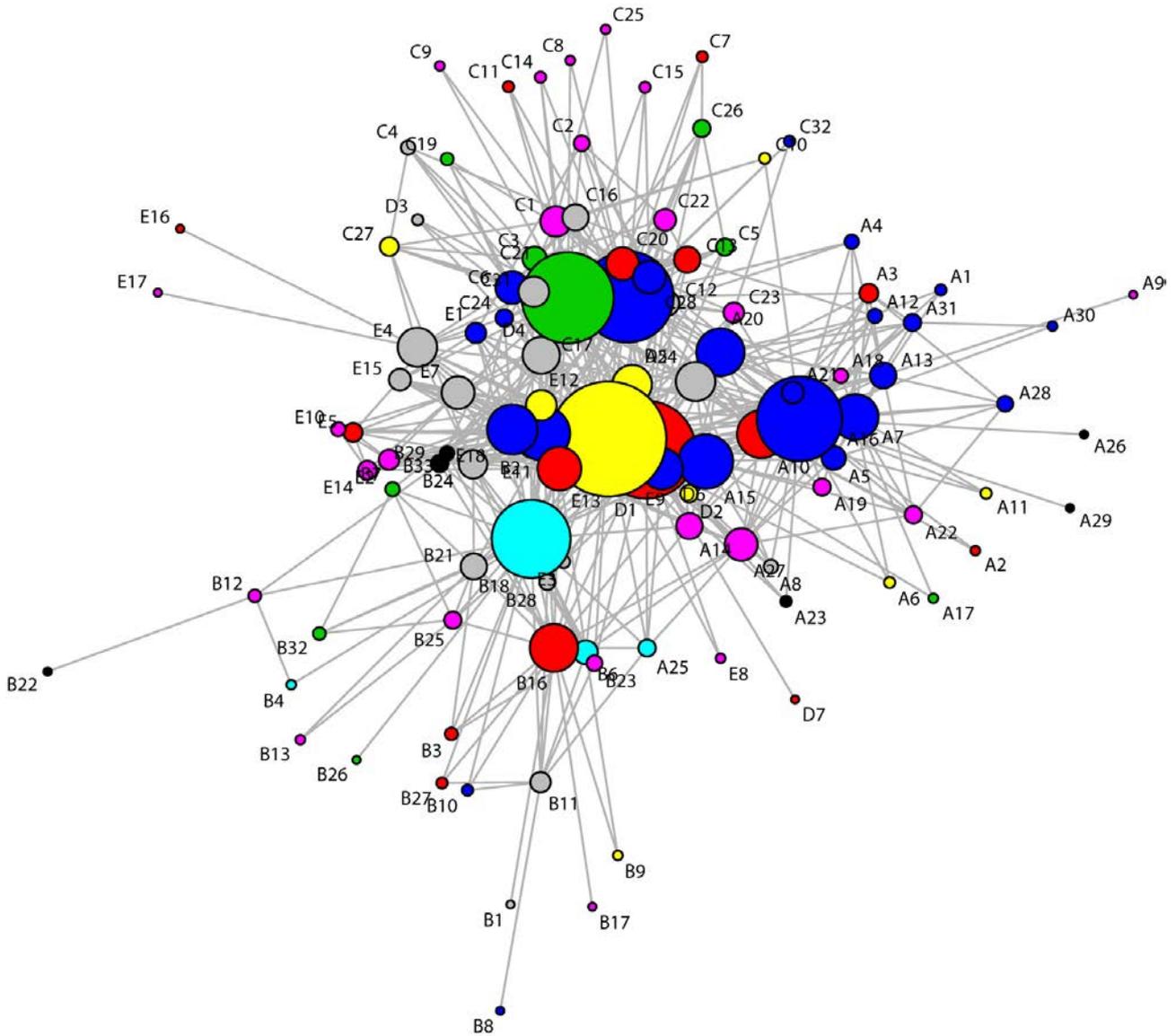
Figure 4: Network plot of the five TREC sites in 2015. The cross-TREC collaborations have increased to 39.67 %, from 35.05% in 2014, which is also more than two times from 15.66% in 2011.
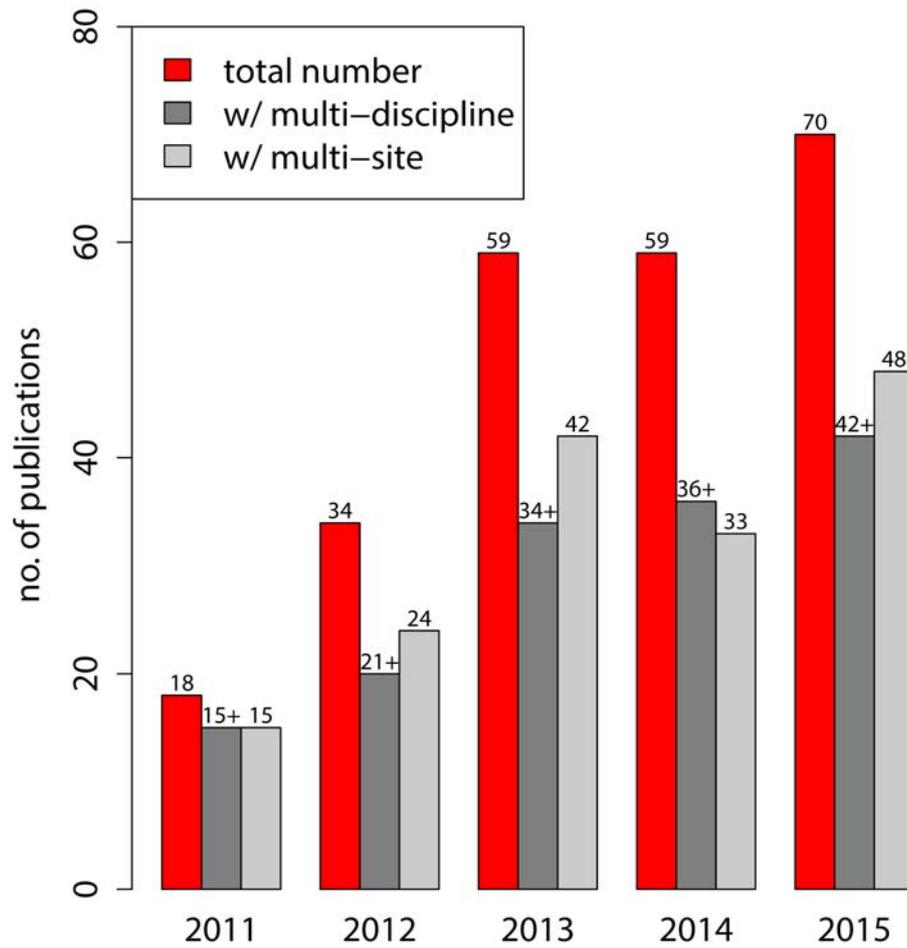
Figure 5: The number of TREC publications per year.

Table 1: Summary of TREC Social Network Survey Response for 2011-2015

		2011	2013	2014	2015
Number invited for survey		100	92	72	70
Number of respondents (rate)		78 (78%)	76 (83%)	59 (82%)	57 (81.4%)
Number of respondents by academic position					
	Professor	30	30	21	24
	Associate Professor	21	17	15	14
	Assistant Professor	12	17	10	9
	Research Associate/Fellow	8	4	4	1
	Other	7	8	9	9
Number of respondents by TREC site (No. invited)					
	Site 1	21 (27)	24 (25)	15 (15)	14 (20)
	Site 2	20 (29)	13 (22)	10 (20)	10 (15)
	Site 3	20 (25)	20 (25)	19 (22)	18 (20)
	Site 4	5 (6)	6 (6)	5 (5)	5 (5)
	Site 5	12 (13)	13 (14)	10 (10)	10 (10)

Table 2: Network Properties of TREC sites in 2011-2015

	2011	2013	2014	2015
Network size (N)	99	116	114	116
Network size by disciplines				
Biochemistry/Genetics	4	5	6	6
Statistics & System Science	12	14	13	16
Social/Behavioral Science	11	14	11	11
Epidemiology	21	25	26	26
Exercise Physiology	5	4	5	4
Medicine	22	29	29	27
Nutrition/Metabolism	8	10	11	9
Public Health Practice	16	15	13	17
Number of ties (collaborations)	415	577	525	547
Density	0.086	0.087	0.082	0.082
Density within each site				
Site	0.356	0.273	0.246	0.213
Site 2	0.236	0.138	0.148	0.131
Site 3	0.261	0.234	0.206	0.225
Site 4	0.733	0.762	0.667	0.619
Site 5	0.590	0.485	0.471	0.412
Triad Census				
No ties	120,893	196,698	189,688	200,402
One tie	32,249	48,851	43,695	44,933
Two ties	3,115	6,806	6,138	6,950
Triangle	592	1,105	943	1,175
Transitivity Index 1 (%)	0.38	0.44	0.39	0.46
Transitivity Index 2 (%)	1.84	2.26	2.16	2.62
Centrality of actors				
Degree (mean (sd))	8.38 (6.11)	9.95 (9.29)	9.21 (9.08)	9.43 (10.1)
Degree (Professor only)	10.69 (7.63)	17.07 (10.52)	18.10 (10.85)	19.13 (10.98)
Closeness (mean (sd))	0.68 (0.09)	0.65 (0.14)	0.65 (0.12)	0.62 (0.14)
Betweenness (mean (sd))	76.36 (189.93)	76.40 (206.86)	81.21 (208.35)	71.35 (185.47)
Correlation between Degree and Betweenness	0.835	0.830	0.835	0.878
Homophily (Probability of ties between two actors within site and discipline				
Site	0.311	0.248	0.233	0.219
(p-value)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Discipline	0.145	0.131	0.124	0.120
(p-value)	(<0.001)	(<0.001)	(<0.001)	(<0.001)

Table 3: Distribution of Disciplines by Site Appeared in the Network through 2011-2015

Disciplines		Site 1	Site 2	Site 3	Site 4	Site 5	Description of disciplines
Biochemistry/Genetics		11	10	0	0	0	
	2011	2	2	0	0	0	
	2013	3	2	0	0	0	Biochemistry, Genetics
	2014	3	3	0	0	0	
	2015	3	2	0	0	0	
Statistics & System		9	12	17	7	10	
	2011	2	3	4	1	2	Biostatistics, Statistics,
	2013	2	3	4	2	3	Informatics, Computer
	2014	2	3	4	2	2	Sciences, Systems
	2015	3	3	5	2	3	Dynamics, Engineering
Social/Behavioral Science		4	14	22	3	4	Social Work, Psychology
	2011	1	3	5	1	1	Communication,
	2013	2	4	6	1	1	Economics, Sociology,
	2014	0	3	6	1	1	Urban Planning,
	2015	1	4	5	0	1	Anthropology
Epidemiology		52	15	19	2	10	
	2011	11	4	4	0	2	
	2013	14	3	5	1	2	Epidemiology
	2014	14	4	5	0	3	
	2015	13	4	5	1	3	
Exercise Physiology		4	12	0	2	0	
	2011	1	3	0	1	0	
	2013	1	3	0	0	0	Exercise Physiology
	2014	1	3	0	1	0	
	2015	1	3	0	0	0	
Medicine		23	29	37	0	18	Oncology,
							Endocrinology, Nursing,
	2011	5	6	8	0	3	OB/GYN Pathology,
	2013	6	8	10	0	5	Geriatrics, Cardiology,
	2013	6	8	10	0	5	Urology, Pediatrics,
	2015	6	7	9	0	5	Immunology
Nutrition/Metabolism		9	4	9	8	8	
	2011	2	1	1	2	2	
	2013	2	1	3	2	2	Nutrition, Metabolism
	2013	3	1	3	2	2	
	2015	2	1	2	2	2	
Public Health Practice		7	22	12	5	15	Behavioral Health, Health
	2011	3	7	2	1	3	Services, Population
	2013	1	6	3	1	4	Health, Health policy,
	2013	1	4	3	1	4	Environment Health,
	2015	2	5	4	2	4	Public Health education

Sarah Gehlert, et al.
The Structure of Distributed Scientific Research Teams Affects Collaboration and Research Output

19

Table 4: Nature of Collaboration for TREC sites in 2011-2015

	2011	2013	2014	2015
Proportion of cross- and within-TREC ties	15.66% vs. 84.34%	34.49% vs. 65.51%	35.05% vs. 64.95%	39.67% vs. 60.33%
Collaboration type (unit: % of the total ties in each year)				
Grant	43.38	42.77	42.00	49.08
Co-authored publication	24.67	35.29	42.94	53.69
Co-authored presentation	14.82	15.67	23.88	28.23
Mentorship	18.10	21.64	23.18	25.92
Committee	34.82	36.10	33.41	33.29
Others	7.65	7.68	7.65	8.06