Understanding Biomedical Research Collaborations through Social Network Analysis: A Case Study

Jiang Bian*, Mengjun Xie†, Umit Topaloglu*, Teresa Hudson‡ and William Hogan*
*Division of Biomedical Informatics
University of Arkansas for Medical Sciences
Little Rock, Arkansas 72205
Email: {jbian, utopaloglu, wrhogan}@uams.edu
†Computer Science
Universit of Arkansas at Little Rock
Little Rock, Arkansas 72204
Email: mxxie@ualr.edu
‡Pharmacoloy Research
University of Arkansas for Medical Sciences
Little Rock, Arkansas 72205
Email: hudsonteresaj@uams.edu

***Abstract*—A recent surge of research on social networks and their characteristics has attracted an increasing amount of interests from the community of biomedicine and biomedical informatics. Social network analysis (SNA) methods have been regarded as an effective tool to assess inter- and intra-institution research collaborations in the Clinical Translational Science Award (CTSA) community. In this paper, we present a case study of SNA on the research collaboration networks (RCNs) at the University of Arkansas for Medical Sciences (UAMS) - a CTSA institution. We have applied graph theoretical analyses to the RCNs prior to and after the CTSA award at UAMS. By virtue of quantitative measures, we have obtained valuable insights into the network dynamics and topological characteristics of the research environment. Moreover, through observing the temporal evolution of the RCNs at UAMS, we are able to demonstrate the effectiveness of the CTSA program and its important role in promoting trans-disciplinary collaborative research within an institution.**

Keywords*—Social Network Analysis; Biomedical Research Collaboration Network; Small-world Network; Clinical and Translational Science Award (CTSA) Evaluation;

I. Introduction

The importance of collaborative research activities across different disciplines and even different geographical locations [1], [2] has gained increasing attention. One of the key objectives of the Clinical Translational Science Award (CTSA), funded by the National Center for Advancing Translational Sciences (NCATS, NIH) (formerly through the National Center for Research Resources (NCRR, NIH)), is to promote cross-disciplinary collaborations that can accelerate the translation and application of biomedical research discoveries into clinical settings. It is essential to quantitatively assess the quality and efficiency of research collaborations, so that we can promptly identify those potential collaborations that are more likely to be productive and make significant impacts.

In the CTSA evaluation community, social network analysis (SNA) methods have been regarded as an effective tool to study and understand inter- and intra-institution research collaborations [3]. Graph theoretical analyses of research collaboration networks (RCNs) would offer us valuable insights, and allow us to understand the network dynamics of collaborative relationships.

Many complex systems can be abstracted as networks (or graphs) where the $nodes$ (vertices) represent entities of interest (e.g., persons, organizations, or objects) and the $edges$ (links) indicate certain relationships or interactions (e.g., friendships in social networks, neurological connections in brain connectivity networks) between nodes [4]. Studies on collaboration networks based on co-authorship in scientific publications [5], [6] have provided insights into the networks' topological characteristics and dynamics [7], [8]. Although being important and informative, the studies on publication co-authorship may not be able to effectively reveal the underlying driving force of collaborations due to their microscopic perspective.

In this paper, rather than based on publication co-authorship, we study research collaborations formed in collaborative biomedical research grants at the University of Arkansas for Medical Sciences (UAMS) – a CTSA institution. One of the key goals of this study is to assess the effectiveness of the CTSA award (i.e., after 2009) at UAMS and its impact on the research collaboration environment within an institution. To achieve this goal, it is necessary to observe the temporal evolution of the RCNs through comparing various network characteristics.

Nagarajan *et al.* presented a baseline study [9] on the collaboration networks prior to the UAMS's CTSA award (i.e., from 2006 to 2009) . Their study suggests that the RCNs at UAMS have "unique characteristics different from those of the established real-world networks". For example, the networks were disconnected with mutually exclusive groups and few weakly connected clusters of staff within the same department. Nevertheless, the baseline study models research collaboration networks as binary (i.e., unweighted) graphs, where it only considers the existence of a collaboration. However, in real-

978-1-4799-1310-7/13/$31.00 ©2013 IEEE

world examples, collaborative research relationships among different investigators may vary. For example, one tends to feel more comfortable to work with existing collaborators rather than finding new peers. Therefore, certain connections in one's collaboration network are considered to be "stronger" than others. It is our tenet that such nature of collaboration relationships should be respected if possible. In this study, we model the RCN as a weighted network, where the weight of an edge is the number of collaborations between the two investigators.

We believe that our network model provides a more precise abstraction of RCNs and helps us study the network dynamics of research collaborations. The fundamental understanding of RCNs will assist in shaping organizational policies, structures, leaderships and resource allocation strategies to make a positive and significant impact on research collaborations and their outcomes.

The contributions of this paper are summarized as follows:

- Our network model assigns non-binary weights to edges to reflect degree of collaboration. Previous studies on scientific collaboration networks [5]–[7], [9]–[11] focus on unweighted or binary networks, where the edge weight is either one or the edge does not exist. However, in real-world examples, collaborative research relationships among different investigators may vary.
- We have studied a number of network characteristic measures such as clustering coefficient, characteristic path length, and number of disjointed components, pertaining to research collaboration networks. Moreover, through observing the temporal evolution of these measures prior to and after the CTSA program at UAMS, we are able to demonstrate the effectiveness of the CTSA program and its important role in promoting collaborative research within an institution. [1]
- We have applied rank aggregation techniques to consolidate four types of network centrality measures (i.e., degree centrality, betweenness, closeness, and eigenvector centrality) [4] to identify important "leader" nodes in RCNs. Network centrality measures express the relative importance of a vertex within the graph. However, different types of centrality measures favor different properties of the vertex. Therefore, they sometimes can make contrary rankings and ambiguous conclusions. Through rank aggregation, we can provide a more concise and representative ranking.

The rest of the paper is organized as follows. We first describe the background of the source data retrieved from an in-house developed research grant management system, ARIA, and how the RCNs are constructed from this dataset in Section II. In Section III, we introduce our weighted network model of research collaboration networks, set forth the concept of network characteristics and introduce some of the basic measures pertain to this study. Subsequently, we discuss our methods of using centrality measures and rank aggregation to identify centrality leaders. In Section IV, we present the analysis results, our interpretations and a few important observations on the trend of collaborative research in a CTSA institution similar to UAMS. In summary, we believe the CTSA award has a positive impact on the RCNs at UAMS through our experiments and observations.

[1]Note that this is an retrospective study such that the administration of the CTSA grant is not aware of these network metrics before this study; thereofre, there is no bias that affects our observations.

II. Materials

In this study, the data for building research collaboration networks is obtained from two sources: ARIA and TRI. The Automated Research Information Administrator (ARIA) is an integrated platform developed internally at the UAMS (partially supported by the National Center for Research Resources (NCRR, NIH)). The ARIA system is used by the Office for Research and Sponsored Programs (ORSP), the Institutional Review Board (IRB), and the Research Support Center (RSC) for managing both research grants and clinical trails. Since 2002, investigators at UAMS have been required to first submit their grant applications to the ARIA system for review before submitting them to the funding agencies. The ORSP at UAMS uses ARIA to keep track of the detailed information regarding research grants such as the requested budget, the budget start/end date, the funding agencies, as well as all the investigators and their roles on the grants.

Besides the ORSP and ARIA, the Translational Research Institute (TRI, UAMS) supports all the CTSA activities at UAMS. With a team of dedicated professionals and many new resources, the TRI helps basic and clinician scientists at both UAMS and external institutions to translate their findings more quickly to clinical practice and into the community. The TRI at UAMS uses the CTSA to provide pilot awards for supporting promising translational studies, and is partnering with multiple research communities across Arkansas to help guide development of meaningful research projects and implementation of research findings. As the TRI tracks all CTSA related activities such as publications, pilot awards and so on, we use the TRI's CTSA reports to obtain the information of whether an investigator on a grant is supported by the CTSA or using TRI services.

Table I shows the statistics of the research grant data we have obtained from both ARIA and TRI. We use the meta-data of those grants to construct an RCN for each budget year from 2006 to 2012. The CTSA at UAMS started on July 14th, 2009. Therefore, in this analysis, the "# of CTSA Investigators" (i.e., investigators who are listed on the original CTSA grant) and "# of CTSA supported investigators" (i.e., investigators who received support from the CTSA) columns in Table I are not applicable for budget years from 2006 to 2009, We only consider the "Principle Investigator", "Co-Investigator", and "Sub-Investigator" roles on the grants, and exclude other personnel such as "Support Staff" and "Laboratory Staff". In addition, we only take into account the grants that have been "Awarded" by the funding agencies for two main reasons: 1) the awarded collaborative research grants indicate successful executions of team science; and 2) a grant might have to go through a few review and revision cycles to get funded. Moreover, resubmitted grant applications normally have the same collaborators. By considering only the final awarded version, we can effectively eliminate some of the noises in

Budget Year	Awarded Grants	# of Investigators	# of CTSA Investigators	# of CTSA Supported Investigators
2006	477	356	N/A	N/A
2007	479	418	N/A	N/A
2008	601	472	N/A	N/A
2009	518	414	N/A	N/A
2010	602	431	34	114
2011	538	463	26	115
2012	548	450	23	322

TABLE I: Statistics of the research grants dataset at UAMS.

**The number of CTSA supported investigators is significantly higher in 2012 than previous years. We think it is because more investigators become aware of and start utilizing the CTSA services as we advertised more to the campus.*

the constructed networks. As each revision is tracked individually as a separate grant application in ARIA, for a research grant spanning multiple years, the grant is considered as an individual one for each fiscal year. For example, a two-year grant whose budget year starts on July 1st, 2009 is counted as an awarded grant for both 2010 and 2011 fiscal years.

III. Graph theoretical analysis of research collaboration networks

A. Network model of research collaborations

An RCN can be modeled as an undirected graph, $G = (V, E)$, where each vertex (v_i) represents an investigator and each edge (e_{ij}) indicates that the two investigators (v_i and v_j) have collaborated in an awarded research grant during the time period of interest. If three investigators (v_i, v_j, and v_k) have worked together on the same research grant, we draw an edge between every two of them (e_{ij}, e_{ik}, and e_{jk}). Previous studies on research collaboration networks [5]–[7], [9]–[11] only model a collaboration network as a binary (unweighted) graph, where the edge is either existent or not existent. However, intuitively, collaborative research relationships between different investigators may vary. For example, an investigator often has many collaborations with the same group of people but much less collaborations with peers outside of that group. Hence, it is crucial to take into account the strength of the collaborative research relationships among investigators.

In this study, we represent each RCN as an *undirected weighted* graph, where the weight (w_{ij}) of an edge (e_{ij}) is the number of research grants the two investigators (v_i and v_j) have collaborated on during the time period of interest. Figure 1 depicts two RCNs, where graph (a) is the RCN at UAMS prior to the CTSA award from 2006 to 2009, and graph (b) is the RCN after the CTSA award from 2010 to 2012. For visualization purpose, we only pick the largest strongly connected components of the two RCNs. Both the original RCNs contain isolated small clusters (i.e., groups that have strong collaborations internally but no connections to other parts of the RCN) and isolated individual nodes (i.e., investigators carried out the research independently). Note that most of these non-collaborative grants are training grants.

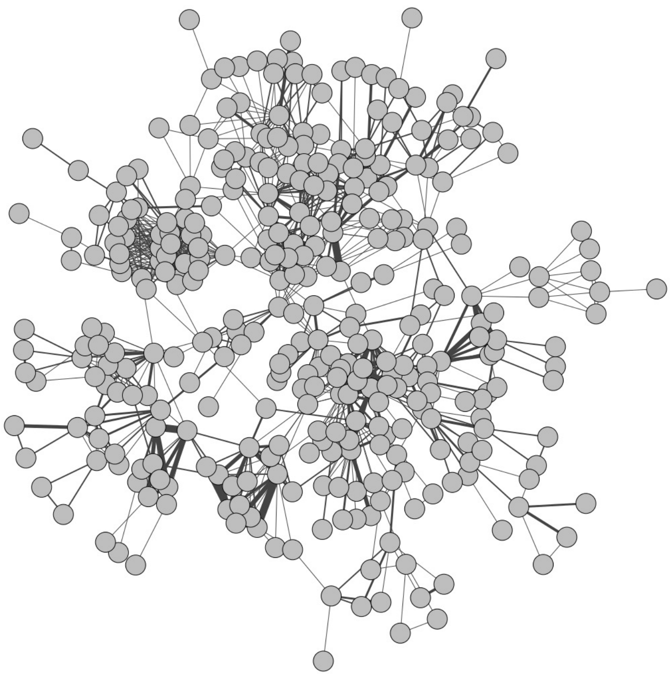
B. Research collaboration network characteristics

In network analysis, the topological features of a network can be quantitatively measured as network characteristics (also called network metrics, network measures, or network indices, which are used interchangeably in this paper). Researchers are constantly seeking metrics to characterize complex networks in a compact and convenient way. In turn, these structural metrics are often used to benchmark or infer functional aspects of a network. For example, the mean path length (characteristic path length) L of a network is often employed to measure the efficiency of information flow on a network.

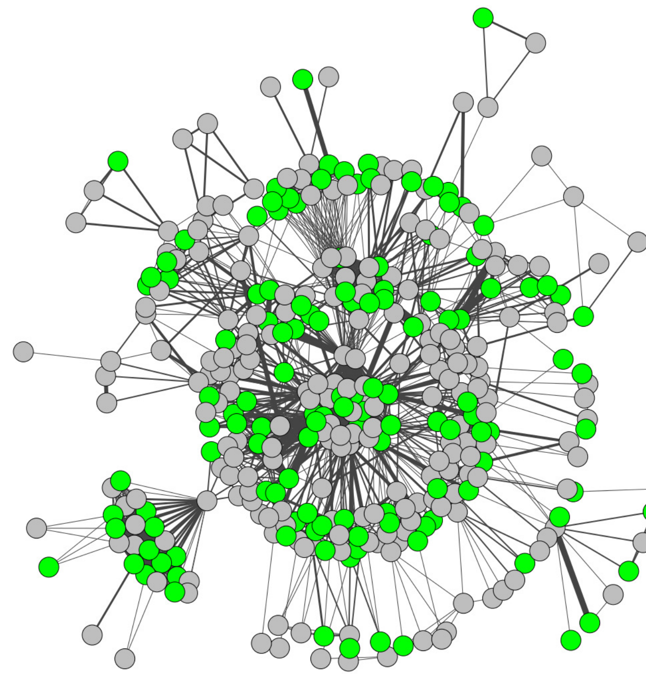
In analyzing research collaboration networks, we are mostly interested in a network's *clustering coefficient* (C) and its *characteristic path length* (L) as they can often be used to categorize a network, e.g., whether the network of interest is a small-world network or a random graph. Network categorization is important since many real-world networks exhibit the small-world property (A small-world network exhibits the low degree of separation among nodes in a sparse yet highly clustered network [12]) and a small-world network has often been hypothesized as more robust to perturbations than other networks [12], [13]. Note that our RCNs are *weighted undirected* graph; therefore, we shall respect the edge weights if possible.

The **characteristic path length** (**L**) refers to the average shortest path length in a network [12], i.e., $L = \frac{1}{n}\sum_{i \in V(G)} L_i$, where n is the total number vertexes in the graph and L_i is the average distance between vertex (v_i) and all other vertices in the network. The characteristic path length on *weighted graphs* are computed similarly, provided that path lengths are calculated with respect to the weights of the edges along the paths. Recall that we define the edge weight w_{ij} as the number of collaborations between investigator i and j. However, algorithms of computing shortest paths often treat the weight of an edge as a cost and prefer edges with smaller weights. Therefore, when calculating shortest path lengths in RCNs, the weight of an edge should be seen as a *resistance factor* of the collaboration. The more number of grants two investigators have worked together, the "resistance" for future collaborations should be lower, and the cost of reaching one node from the other should also be smaller. Thus, we define the resistant factor between two investigators (r_{ij}) as the reciprocal of the number of collaborations between the two, and use r_{ij} as the new edge weights ($w^r_{ij} = 1/r_{ij} = 1/w^o_{ij}$). We use the resistant factor as the weighting schema for all shortest path related measures.

The **clustering coefficient** is a measure of how likely nodes in a graph tend to cluster together. The (local) clustering coefficient of a vertex is defined as the fraction of triangles around the vertex [12], [14], where the (global) clustering coefficient of the network is defined as the average clustering coefficient of all vertices in the graph. The clustering coefficient on a



(a)



(b)

Fig. 1: The research collaboration networks (RCNs) at UAMS, where graph (a) is the RCN prior to the CTSA (i.e., 2006 – 2009); and graph (b) shows the RCN after the CTSA from 2010 to 2012. *(*The edge weights are visualized as thickened lines. The thickness represents the degree of collaboration between two investigators. Nodes in green represent investigators who are supported by the CTSA)*

weighted graph can also be calculated by considering the contributio n of each triangle with respect to the weights of its edges [15]. To model the trend of cross-disciplinary collaborations, which is particularly interested by the CTSA community, we propose a quantitatively "**diversity**" measure for RCNs. Thus, the diversity of a network D_g is defined as follows:

$$D_g = (\frac{1}{n}\sum_{k=1}^{n} L_{S_k->\neg S_k})^{-1},$$

where $L_{S->\neg S}$ is the average shortest path length from nodes in set S to all other nodes in the network, n is the number of distinct groups (a collection of nodes having the same property of certain kind) in the network. If we define each group as a discipline in the RCNs, the diversity measures how easy (or difficult) an investigator from one discipline reaches another investigator in a different research field. Therefore, the higher the diversity value, the more diversified the collaborations are in the RCN, as the average distance is shorter for an investigator to travel from one group to another.

C. Identify centrality leaders in the RCNs

In social network analysis, the *centrality* measures of a node are often used to determine the relative importance of the node in the network. Within the context of social network, a centrality measure can be interpreted as how influential or important a person is in the social network of interest. There are various network centrality measures, where each measure defines the meaning of importance from a different perspective [4]. In this paper, we investigate four widely-used network centrality measures: degree centrality, betweenness, closeness, and eigenvector centrality [16]. We briefly describe them below.

The **Degree centrality** is simply the degree of a vertex (v_i), which is the number of edges incident to the vertex. A node with a high degree can be seen as a highly connected "hub" in the RCN. Since the RCNs are *weighted* graphs, we measure the *weighted degree* (a.k.a., *strength*) of each vertex and define the strength of a vertex (s_i) as the summation of the weights of all edges incident to that vertex. Many real-world complex networks follow a power-law degree distribution, where the majority of the nodes have a small number connections, while there exist a few well-linked hubs. Such networks are often called scale-free.

The **Betweenness centrality** of a vertex (v_i) is defined as the fraction of all shortest paths in the network that pass through that vertex (i.e., the number of shortest paths that go through vertex i over the total number of shortest paths in the network). Betweenness centrality of node i in a *weighted graph* can be defined similarly, given the shortest path length is the weighted shortest path length. Betweenness centrality measures a node's control of the communication between other nodes in the network [17]. Conceptually, in the RCNs, a node with a high betweenness centrality value can be interpreted as an influential (control of communication) investigator in the research community.

RCN	# of nodes	# of edges	density	# of isolated components	clustering coefficient	characteristic path length	diversity	average # of new edges
2006	184	279	0.02	51	0.84	1.03	0.75	N/A
2007	275	678	0.02	44	0.87	2.31	0.39	1.58
2008	276	532	0.01	48	0.73	2.10	0.40	-0.10
2009	262	590	0.02	41	0.79	2.98	0.29	0.47
2006 – 2009	**487**	**1318**	**0.01**	**55**	**0.66**	**2.65**	**0.37**	**N/A**
2010	292	1412	0.03	31	0.77	2.16	0.44	10.79
2011	308	1082	0.02	34	0.76	2.54	0.39	-9.89
2012	280	1083	0.03	31	0.82	2.09	0.48	0.59
2010 – 2012	**425**	**2006**	**0.02**	**36**	**0.70**	**1.68**	**0.56**	**16.47**

TABLE II: Network characteristics of the RCNs at UAMS from 2006 to 2012.

The **Closeness centrality** of a vertex (v_i) (i.e., local closeness centrality) is the inverse of the local characteristic path length (i.e., the sum of its distances to all other nodes) of the vertex [17]. Closeness centrality of a vertex (v_i) on a *weighted graph* can be computed similarly, considering that the path lengths are calculated using the weighted definition. In a connected graph, the closeness centrality of a node reflects how close the node is to all the other nodes. The closeness centrality value can be seen as how fast information can flow from a node to all other nodes [18]. Therefore, a node is more "central" if its total distance to all other nodes is smaller, and its closeness centrality value is higher.

The **Eigenvector centrality** measures the influence score of a vertex (v_i) in the network [19]. Similar to a preferential attachment process, the calculation of the eigenvector centrality score makes the assumption that connecting to a high-score node generally gains more "reputations" (i.e., scores) than connecting to a lower score node. The random walk with restart (RWR) process that will be described in the next subsection essentially calculates the personalized pagerank score of each vertex, which is a variant of the eigenvector centrality measure.

Using these centrality measures, we can rank an investigator's influence (or importance, contribution) in the research community. However, the centrality measures can rarely make a consensus regarding the ranking orders of the nodes in the same network. Therefore, we propose to use rank aggregation techniques [20]–[22] that can combine multiple rankings of nodes (investigators) to generate a better and more concise ranking. There are basically two classes of rank aggregation methods: 1) score-based rank aggregation, where each object in the input ranking is associated with a score and the goal is to combine different scoring systems to produce one set of scores; and 2) order-based rank aggregation, where only the orders of objects produced by individual ranking method is considered. Since the scores given by different centrality measures are diverse and it is difficult to choose a meaningful normalization process, we decide to use the simple Borda count [21] system, which is an order-based voting system. The Borda count system gives each candidate certain points based on her position on each ballot. The candidate with the most points is the winner. If we consider each centrality measure as a voter that gives a preference ranking of all investigators in the RCNs, the final ranking can be easily computed using the Borda count of each investigator.

IV. Results and Discussion

A. Temporal evolution of research collaboration networks at UAMS

We constructed a number of RCNs with varying time periods of interest, in which each RCN is composed with only grants that are awarded during that time period. We constructed seven snapshot RCNs each for one budget year from 2006 to 2012 ($RCN_{2006} \sim RCN_{2012}$). We also constructed two aggregated RCNs, one spanning from 2006 to 2009 ($RCN_{2006-2009}$) (i.e., prior to the CTSA award) and the other spanning from 2010 to 2012 ($RCN_{2010-2012}$). By doing so, we study the structure of RCN from both short-term and medium-term perspectives.

Table II shows the network characteristics we observed for each of the snapshot RCN we have constructed. We eliminated all isolated single nodes that do not connect to any other nodes in the network. These isolated nodes indicate that the investigators carried out the research activities independently. It is reasonable to remove these nodes as they do not contribute any information to the study of collaborations. Besides the above-mentioned clustering coefficient, characteristic path length and the diversity measure, we have also included a few auxiliary metrics that can help us understand the network structures. We briefly introduce these measures as follows.

The **density** of a network $G = (V, E)$ is defined as the ratio of the number of edges in set E over the maximum possible number of edges (for undirected graph, $d = 2 \times |E|/(|V| \times (|V| - 1))$). An **isolated component** in a network is a small community that has no links to any other parts of the network. The number of isolated component can be seen as a measure of the degree of segregation in the collaboration environment. The **average number of new edges** is measured as follows. We compare each year's RCN with that of the previous year, and identify all the nodes exist in both years (i.e., investigators who have collaborative funding grants in both fiscal years). We count the number of newly created edges for each of the identified node, and take the average over all the nodes. We use the FY2006 data as our baseline for this measure. Therefore, there is no result for RCN_{2006} and $RCN_{2006-2009}$. For $RCN_{2010-2012}$, we compare it with the aggregated RCN prior to the CTSA award ($RCN_{2006-2009}$).

As shown in Table II, the RCN at UAMS is moving towards a positive direction, i.e., not only more collaborations, but also more cross-disciplinary teamwork. The RCNs after the CTSA award (i.e., 2010 – 2012) have significantly more edges (i.e.,

	Average Strength		Average Shortest Path Length				
RCN	$\bar{S}^{-}$	$\bar{S}^{+}$	$\bar{L}^{(-)}$	$\bar{L}^{(+)}$	$\bar{L}^{(+\Rightarrow -)}$	$\bar{L}^{(-\Rightarrow \pm)}$	$\bar{L}^{(+\Rightarrow \pm)}$
2006 – 2009	**9.97**	**10.30**	**2.66**	**2.46**	**2.79**	**2.68**	**2.57**
2010 – 2012	**19.88**	**23.60**	**1.65**	**1.61**	**1.81**	**1.67**	**1.69**

TABLE III: Comparing network metrics between CTSA $(+)$ supported investigators and non-CTSA $(-)$ investigators.

collaborations), although the number of grants in each year is relatively stable (see Table I). Moreover, the number of isolated components ($RCN_{2006-2009} = 55$ v.s. $RCN_{2010-2012} = 36$) has decreased dramatically, which is another sign of growing collaborations in the research community.

The **clustering coefficient** and the **characteristic path length** of a network are two important measures that help us understand network topology, stability, characteristics and efficiency. Our results confirm the findings made by the previous study [9] that the RCN at UAMS is moving towards a small-world topology, as the *clustering coefficient* ($C \in [0, 1]$) has increased from 0.66 to 0.70 while the *characteristic path length* (L) has decreased from 2.65 and 1.68 between $RCN_{2006-2009}$ and $RCN_{2010-2012}$. The clustering coefficient measures the degree of the herding effect in a network (i.e., the probability that the immediate neighbors of a node are also connected), where a high cluster coefficient value indicates that the nodes tend to create more tightly knit groups as shown in Figure 1. The dramatic decrease of characteristic path length (i.e., the degree of separation) after the CTSA award suggests that it becomes faster for an investigator to reach another investigator in the RCN and easier to foster new collaborative research projects. However, we have yet to assert that the RCNs at UAMS are actual small-world networks as we need to compare the C and L of the RCNs with those of the random graphs constructed from the same vertex set.

The proposed **diversity** measures of the two aggregated RCNs (i.e., $RCN_{2006-2009}$ and $RCN_{2010-2012}$) have revealed that the research community at UAMS is shifting towards more interdisciplinary collaborations. As the goal of CTSA award is to incubate new multidisciplinary collaborations and high impact research across the spectrum of translational science, the shifting suggests the impact of CTSA.

The **average number of new edges** is also interesting. Compared to the result in the previous year, the average number of new edges (i.e., new collaborations) is rather stable from 2007 to 2009. Surprisingly, there was a surge of new collaborations in 2010, that is, 10.79 new collaborations on average compared to the 2009 dataset. We believe it is a mixed effect of the CTSA award and the American Recovery and Reinvestment Act of 2009, which resulted in a large number of new grants funded that year. However, there was a significant decrease from 2010 to 2011 (i.e., -9.89) possibly due to the economic recession. Without network analysis of the RCNs, these novel observations would not be uncovered. We note that the these findings are important in that they can help an organization to understand the driving forces of research collaborations and adjust its policies to promptly protect the collaboration environment.

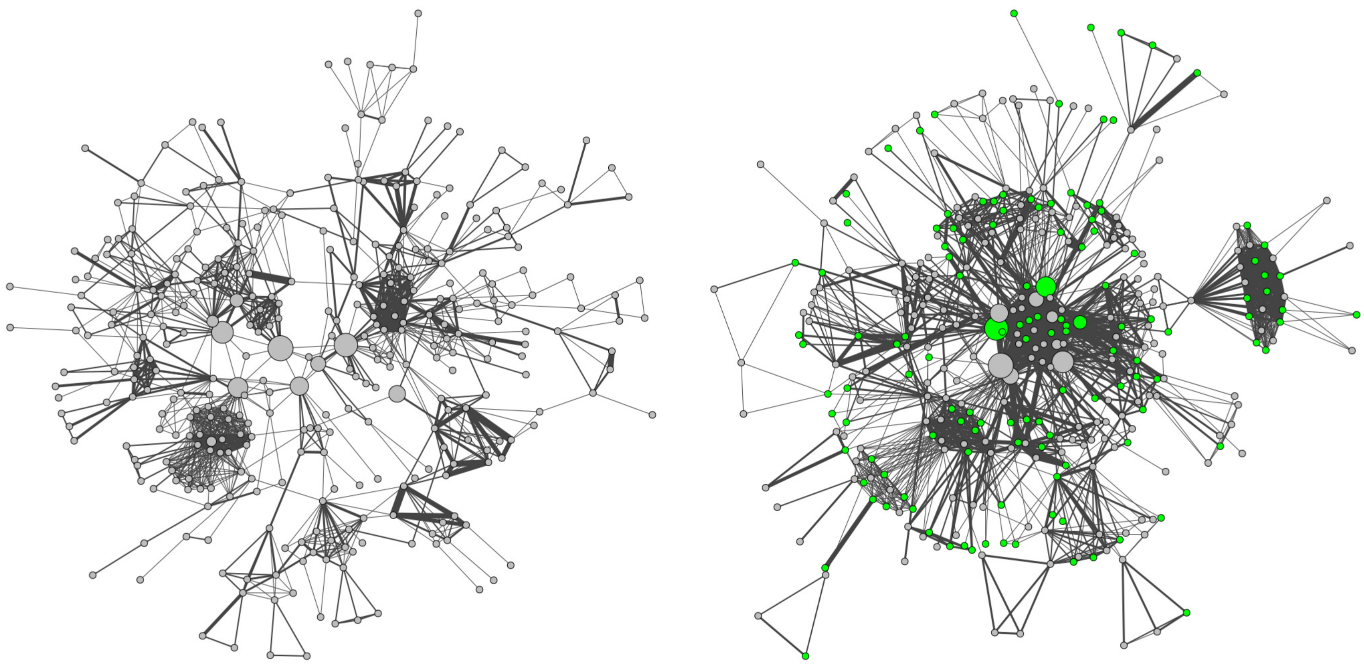
B. The impact of the CTSA on RCNs at UAMS.

To examine the effect of the CTSA award, we also compared network metrics between CTSA $(+)$ supported and non-CTSA $(-)$ supported investigators. We split the nodes into two groups, i.e., CTSA and non-CTSA. For each group, we measure the average strength $(\bar{S})$, the average shortest path length within a group ($\bar{L}^{(+)}$ and $\bar{L}^{(-)}$) (i.e., the average shortest path length from any nodes in a group to any other nodes in the $same$ group), and the average shortest path length across groups ($\bar{L}^{(+\Rightarrow -)}$, $\bar{L}^{(+\Rightarrow \pm)}$, $\bar{L}^{(-\Rightarrow +)}$ and $\bar{L}^{(-\Rightarrow \pm)}$) (i.e., the average shortest path length from any nodes in a group to any other nodes in the $other$ group). Since RCNs are undirected graphs, the average shortest path length across groups is therefore the same for both directions (i.e., $\bar{L}^{(+\Rightarrow -)} = \bar{L}^{(-\Rightarrow +)}$). As we are interested in seeing how these investigators interact prior to the CTSA award, we identify the grouping of investigators in the $RCN_{2006-2009}$ using the $RCN_{2010-2012}$ data, and compute the same set of measures for $RCN_{2006-2009}$.

As shown in Table III, comparing network metrics between CTSA supported and non-CTSA supported investigators reveals valuable insights. First, the average strength (i.e., weighted degree) of nodes within the CTSA group is larger than that of the non-CTSA group for both $RCN_{2006-2009}$ and $RCN_{2009-2012}$. More interestingly, the difference in the average strength between the CTSA group and non-CTSA group increased dramatically from $RCN_{2006-2009}$ to $RCN_{2010-2012}$, which indicates the CTSA investigators became more active in collaborative research grants than the non-CTSA group (i.e., for $RCN_{2006-2009}$, $\bar{S}^{+} = 10.30 > \bar{S}^{-} = 9.97$; and for $RCN_{2010-2012}$, $\bar{S}^{+} = 23.60 > \bar{S}^{-} = 19.88$). Second, as discussed above, the average shortest path length, i.e., the degree of separation, tells us how easy or difficult for an investigator to reach another researcher in an RCN. Comparing $RCN_{2010-2012}$ with $RCN_{2006-2009}$, the degree of separation decreases for both CTSA and non-CTSA groups. This decrease indicates that with the CTSA grant, it is much “easier” for investigators to find other researchers (with or without CTSA funds) and form new collaborations. Another important observation of the path length measures is that the difference between $\bar{L}^{(-\Rightarrow \pm)}$ and $\bar{L}^{(+\Rightarrow \pm)}$ decreases from 0.11 to 0.02 (i.e., between $RCN_{2006-2009}$ and $RCN_{2010-2012}$). Therefore, the CTSA award not only encourages more collaborative research activities among the CTSA supported investigators but also shortens the collaboration “distance” for non-CTSA supported investigators.

C. Centrality leaders in the research collaboration networks at UAMS

We use the proposed method to identify *centrality leaders* (i.e., social leaders) in the two aggregated RCNs ($RCN_{2006-2009}$ and $RCN_{2010-2012}$). As shown in Figure 2,



(a) (b)

Fig. 2: The centrality (i.e., social) leaders identified in the RCNs at UAMS, where graph (a) is the RCN prior the CTSA award (i.e., 2006 – 2009); and graph (b) shows the RCN after the CTSA from 2010 to 2012. **The relative sizes of the nodes illustrate the consented centrality rankings.*

the size of a node is set proportional to its ranking (i.e., the larger the node in size, the higher the investigator's ranking) to depict the leader nodes.

Combining various network centrality measures through rank aggregation leads to discovering key components in a network in a more concise and representative manner. Visually, as demonstrated in Figure 2, the identified centrality leaders are mostly the bridging nodes that connect different parts of the network.

However, the discovered centrality "leaders" are rather different from what we normally perceive as "leaders" in the context of academic institutions. For example, we found that the identified "leaders" of the RCNs at UAMS are neither the actual leaders of the university nor are the leading investigators. Top ranked investigators instead often are biomedical informatics researchers and biostatisticians. Biomedical informatics and biostatistic investigators provide a common service to other researchers, so that they often appear on many grants collaborating with different PIs as "Co-Investigator". Nevertheless, in the context of collaboration networks, these are the "leader"-nodes as they contribute the most to the structure and efficiency of the network.

V. Conclusion

In this paper, we presented an extensive study of biomedical research collaborations using network analysis methods. Distinct from previous studies on the same topic, we proposed a novel network model that considers the strength of the collaborative relationships. We believe that our weighted graph model provides a more natural representation of research collaborations and helps us understand the characteristics and dynamics of the collaborative research environment. Moreover, through observing the evolution of the research collaboration network prior to and after the CTSA award at UAMS, we obtained quantitative evidences that the research environment at UAMS is moving towards a positive direction in terms of both productivity and efficiency. Furthermore, since the CTSA award, investigators at UAMS have established more cross-disciplinary collaborations, which suggests that the CTSA at UAMS has had a positive impact on creating an effective trans-disciplinary research collaboration environment. Last but not least, we applied a rank aggregation method to consolidate the four widely-used network centrality measures to identify "influential" or "important" investigators in the research collaboration networks. We believe our study can help administration and leaderships in a research organization to strategically allocate resources to protect or nurture more "influential" nodes as losing these nodes may have negative impacts on the efficiency of the network.

Acknowledgment

The work describe in this manuscript is supported by award UL1TR000039 through National Center for Advancing Translational Sciences (i.e., formerly UL1RR029884 through the NIH National Center for Research Resources). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

[1] S. Wuchty, B. F. Jones, and B. Uzzi, “The increasing dominance of teams in production of knowledge,” Science, vol. 316, no. 5827, pp. 1036–1039, May 2007.

[2] R. K. Pan, K. Kaski, and S. Fortunato, “World citation and collaboration networks: uncovering the role of geography in science,” Sci Rep, vol. 2, p. 902, 2012.

[3] CTSA Consortium: Evaluation Key Function Committee, “Evaluation - social network analysis,” [Online]. Available from: https://www.ctsacentral.org/committee/evaluation-social-network-analysis, [Accessed March 3rd, 2013].

[4] M. E. J. Newman, Networks : an introduction. Oxford New York: Oxford University Press, 2010.

[5] M. E. Newman, “The structure of scientific collaboration networks,” Proc. Natl. Acad. Sci. U.S.A., vol. 98, no. 2, pp. 404–409, Jan 2001.

[6] S. Uddin, L. Hossain, and K. Rasmussen, “Network effects on scientific collaborations,” PLoS ONE, vol. 8, no. 2, p. e57546, 2013.

[7] M. E. Newman, “Coauthorship networks and patterns of scientific collaboration,” Proc. Natl. Acad. Sci. U.S.A., vol. 101 Suppl 1, pp. 5200–5205, Apr 2004.

[8] Y.-L. Luo and C.-H. Hsu, “An empirical study of research collaboration using social network analysis,” in Computational Science and Engineering, 2009. CSE ’09. International Conference on, vol. 4, Aug., pp. 921–926.

[9] R. Nagarajan, C. L. Lowery, and W. R. Hogan, “Temporal evolution of biomedical research grant collaborations across multiple scales–a CTSA baseline study,” AMIA Annu Symp Proc, vol. 2011, pp. 987–993, 2011.

[10] R. Nagarajan, A. T. Kalinka, and W. R. Hogan, “Evidence of community structure in Biomedical Research Grant Collaborations,” J Biomed Inform, vol. 46, no. 1, pp. 40–46, Feb 2013.

[11] M. E. Hughes, J. Peeler, and J. B. Hogenesch, “Network dynamics to evaluate performance of an academic institution,” Sci Transl Med, vol. 2, no. 53, p. 53ps49, Oct 2010.

[12] D. J. Watts and S. H. Strogatz, “Collective dynamics of ’small-world’ networks,” Nature, vol. 393, no. 6684, pp. 440–442, Jun 1998.

[13] M. Newman, A.-L. Barabsi, and D. J. Watts, The Structure and Dynamics of Networks: (Princeton Studies in Complexity). Princeton University Press, 2006.

[14] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications, 1st ed., ser. Structural analysis in the social sciences. Cambridge University Press, 1994.

[15] “The architecture of complex weighted networks,” Proceedings of the National Academy of Sciences of the United States of America, vol. 101, no. 11, pp. 3747–3752, Mar. 2004. [Online]. Available: http://dx.doi.org/10.1073/pnas.0400087101

[16] “Node centrality in weighted networks: Generalizing degree and shortest paths,” Social Networks, vol. 32, no. 3, pp. 245–251, Jul. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.socnet.2010.03.006

[17] L. Freeman, “Centrality in social networks: Conceptual clarification,” Social Networks, vol. 1, no. 3, pp. 215–239, 1979. [Online]. Available: http://dx.doi.org/10.1016/0378-8733(78)90021-7

[18] M. Newman, “A measure of betweenness centrality based on random walks,” Social networks, vol. 27, no. 1, pp. 39–54, 2005.

[19] P. Bonacich, “Power and centrality: A family of measures,” American Journal of Sociology, vol. 92, no. 5, pp. 1170–1182, 1987.

[20] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, “Rank aggregation methods for the web,” in Proceedings of the 10th international conference on World Wide Web, ser. WWW ’01. New York, NY, USA: ACM, 2001, pp. 613–622. [Online]. Available: http://doi.acm.org/10.1145/371920.372165

[21] Designing an All-Inclusive Democracy: Consensual Voting Procedures for Use in Parliaments, Councils and Committees. Springer, 2007.

[22] J. A. Aslam and M. Montague, “Models for metasearch,” in Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, ser. SIGIR ’01. New York, NY, USA: ACM, 2001, pp. 276–284. [Online]. Available: http://doi.acm.org/10.1145/383952.384007