Tao Jin
Louisiana State University, Baton Rouge, LA, USA

Austin Ward
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Kwan Yi
Eastern Kentucky University, Richmond, KY, USA

The Key to Fostering Transdisciplinary Research Collaboration: Finding the Connections (Paper)

Abstract or Résumé:

How to foster collaboration among researchers is an enduring issue facing university administrators. Homophily, the tendency that people like to work with those similar to themselves, has been found as a fundamental mechanism behind scientific collaboration. When similarity breeds connections, collaboration becomes conceivable. In a multidisciplinary context, like a university, how can a homophily mechanism be applied to facilitate research collaboration? This paper proposes a co-word analysis approach which can be used to process a large volumes of data. The aim is to identify and reveal hidden connections between university faculty members and contribute towards cultivating transdisciplinary research collaboration.

1. Introduction

Universities are places where many talents mingle, viewpoint diversity flourishes, and scientific discoveries occur every day. Despite this, university campuses are also places where silos develop between departments, disciplines, and individual researchers (Hara, Solomon, Kim, & Sonnenwald, 2003). Scientific research has a social heterogeneous nature and science itself has been regarded as "an inherently collaborative enterprise" (Finholt, 2002; Katz & Martin, 1997; Star & Griesemer, 1989). More importantly, it has become a general trend that institutions and funding agencies have increasingly been encouraging and supporting research collaborations that cross disciplinary, cultural, and organizational boundaries, in order to address complex scientific problems that might not be able to solve by a single entity or knowledge domain (Chung, Kwon, & Lee, 2016; Cummings & Kiesler, 2005). Therefore, how to foster collaboration among researchers is an enduring and central question that must be answered by university administrators when they develop relevant theories and policies.

Sociology is one of the disciplines which may assist in addressing this issue. Sociology has a term, *homophily*, meaning that people like to work together with those who share similar backgrounds, interest, and characteristics: as described in the proverb "birds of a feather flock together" (McPherson, Smith-Lovin, & Cook, 2001). Over the years, the principle of homophily has been found as one of the fundamental mechanisms behind scientific collaboration (Freeman

& Huang, 2015; Sie, Drachsler, Bitter-Rijpkema, & Sloep, 2012; Zhang, Bu, Ding, & Xu, 2018). The rationale is that similarity breeds connections, and those connections make collaboration conceivable. However, in a multidisciplinary university context, how can the principle of homophily be applied to facilitate research collaboration?

This proposed paper will present some preliminary results of an on-going research project, which aims to ultimately build an internal web-based knowledge management portal to facilitate research collaborations among 78 full-time faculty members within a college at a research university. Specifically, this proposed paper will report a portion of the larger project, focusing on how we can reveal hidden connections between the faculty members, make the connections searchable on a portal, and eventually nurture and facilitate potential collaborations. One contribution of this submission will be an original *co-word analysis* approach that can be applied to process and analyze large volumes of data in order to identify, measure, and disclose a faculty member's interconnectivity with his or her colleagues within a given unit.

2. Background

The term *research collaboration* can be loosely defined as the social interaction among two or more researchers who work together to achieve a common vision of producing new scientific knowledge (Katz & Martin, 1997; Sonnenwald, 2007). It is often interchangeably used with such terms as *scientific collaboration*, *scholarly collaboration*, or *academic collaboration*.

Different from the concepts of *teamwork* or *research partnership/coalition*, research collaboration emphasizes a mutually beneficial, capacity-building experience, as well as joint effort and ownership rather than shared data and resources, joint operation, or polite cooperation (Byrne & Hansberry, 2007; Waruszynski, 2017).

Considerable research has been done on when and why researchers collaborate. Wagner and Leydesdorff (2005) find that how researchers select partners to initiate a collaboration is highly contingent upon their personal choices. A wide array of internal factors that may influence collaboration propensity have been identified (Bukvova, 2010), such as: *agreement on quality* (Birnholtz, 2007), *reward and credit* (Stokols, Misra, Moser, Hall, & Taylor, 2008), *engagement in coordination activities* (Cummings & Kiesler, 2007), *awareness* (Carroll, Rosson, Farooq, & Xiao, 2009), and *personal characteristics* (Stokols et al., 2008), as well as many other personal factors. Bozeman and Corley (2004) surveyed 997 university faculty members to investigate their collaboration strategies. Based on their results, they developed a typology of collaborators: (1) the *taskmasters*: who tend to select partners based on criteria of reliability and work ethics; (2) the *nationalists*: who tend to choose partners that speak the same language or from the same nationality with their own; (3) the *mentors*: who tend to work together with junior faculty members and their students; (4) the *followers*: who tend to select collaborators with strong reputation; (5) the *buddies*, who like to work with people that they have collaborated before; and (6) the *tacticians*, who tend to choose partners with similar or compatible skills with their own (Bozeman & Corley, 2004). It is notable that four categories out of the six apply to the principle of homophily.

However, while the literature abounds with studies on existent scientific collaborations, few endeavors have been made to explore those emerging and potential collaborations that are still in their infancy. What has been missing in the literature are systematic attempts to reveal those latent similarities between researchers and provide them with a tool to make the connections.

The purpose of this portion of the project is to propose and test an analytical approach to using publicly available data to examine, identify, measure, and reveal hidden interconnectivity between faculty members within a college at a research university.

One research question guided this endeavor: How can we use data from public domains to project and reveal a university faculty member's connectedness with his or her colleagues?

3. The Co-Word Analysis Approach

Co-word analysis is a quantitative technique. Its basic assumption is that co-occurrence frequency of pairs of words or phrases would indicate some linkages or connectedness between those two parent entities that contain the words or phrases (He, 1999). Therefore, it has been widely used to discover association strengths between papers, themes, subjects, or authors (Ding, Chowdhury, & Foo, 2001).

To address the research question and attain the research goal aforementioned, we designed and implemented a co-word analysis approach. This approach includes the following elements: (1) using faculty members' published curriculum vitae (CV) as data sources; (2) extracting personal and publications information from each CV to identify the faculty member's key attributes in terms of his or her research; (3) accumulating, compiling, and normalizing those identified attributes systematically; (4) based on those processed attributes data, running co-word analysis via calculating each pair of faculty members' *Jaccard's* similarity coefficient and *Pearson's* correlation coefficient to measure their interconnectivity; and (5) mounting the results data onto the web-based knowledge management portal, so that faculty members can search their colleagues' expertise, see how much similarity or estimated interconnectivity they share, and communicate with any one whom they would like to collaborate.

Our case study was a college at a university in the southern part of the United States. The university is a Carnegie-R1 comprehensive public research university hosting 18 colleges and schools. The college selected to conduct this research concentrates on human sciences and education. It consists of five academic departments, majoring in education, kinesiology, human resources development, library and information science, and social work, respectively. The college was recently created after a university-wide restructuring initiative. Those five departments are physically located in three different buildings with a walking distance of about 300 to 800 meters to each other. While there are some intra-collaborations observed within each department, few inter-collaborations were seen across departments. There were 78 full-time tenure-track and tenured faculty members when the project was carried out.

Each faculty member produces publications, which demonstrate their research interests, preferable methodologies, and expertise. All of these can be represented by a series of keywords,

terms, or phrases. If we can identify and collect these attributes, and use them to label each faculty member, then we will be able to compare similarities between any pair of faculty members via comparing these attributes.

Below are the procedures that we followed. First, we collected each faculty member's CV from their departmental website: which are publicly available. Then, we extracted their publication citations. The citations were then copied over to a spreadsheet (one spreadsheet per academic department; one tab per faculty member) and the peer-reviewed publication citations were separated from the book and presentation citations.

Next, we used the bibliographical database system, EBSCOhost, to perform a search for the corresponding publications through the collected citations. An effort was made to document the result of the searches, including when the publications could not be found or accessed by the research group. When a publication was found, the following information was recorded: the journal the publication was in, the abstract, keywords found in the publication, the subject-terms used by the database that hosted the publication, and the author-supplied keywords that were culled from the publication if the keywords or subject terms were not present. There were instances where a publication could not be found itself, but the metadata associated with the publication was stored by the database.

At this point, there were hundreds of lists of keywords. For each faculty member, a concatenated list of keywords was created. To remove duplicates from the list, all the text of the list was made lower case, split into individual terms by splitting the string on every comma, and compared to each other to remove any term that already appeared in the list. This process to create a list of distinct keywords for each faculty member was repeated for the subject terms and author-supplied keyword columns. To get this information into a useful form for co-word analysis, the text-to-column function was used to associate a faculty member with their list of distinct keywords and terms in the form of columns.

Next, the data of the keywords was processed by a co-word analysis software to calculate Jaccard's similarity coefficient and Pearson's correlation coefficient for each pair of the 78 faculty members.

We created a database and a website interface so that the desired audience could gain access to the database. The database was created using PostgreSQL (9.4.4). In its current configuration, it utilizes a symmetrical, many-to-many relationship using two tables: Faculty and Faculty Coefficients. The data returned by the co-word analysis needed to be formatted for entry into the database. A distinct list of faculty members was inserted into the database. The coefficient results were arranged such that faculty were matched to each other and their respective numbers. Duplicate relationships in the results were removed. PostgreSQL was used to insert the information in the spreadsheet to the database.

The website was written in PHP 5.3.3 with PostgreSQL. In its current form, the website allows the user to select a faculty member, choose a coefficient, select the number of results, and in what direction the user would like the sorting to take. When submitted, the results are viewed in

a table that lists the faculty member searched and their related faculty member via the selected coefficient.

4. Concluding Remark

This proposed paper presents an approach that was used in an on-going project to identify and reveal latent connections between 78 university faculty members, which could be the key to facilitate transdisciplinary research collaboration in a multidisciplinary academic environment. Numerous tables, visualizations, and screenshots are ready to present (see Appendices as examples). Next step, we will ask the faculty members to use the portal website and get their perceptions about this approach toward fostering collaboration.

Reference List:

Birnholtz, J. P. (2007). When do researchers collaborate? Toward a model of collaboration propensity. *Journal of the American Society for Information Science and Technology, 58*(14), 2226-2239. doi:10.1002/asi.20684

Bozeman, B., & Corley, E. (2004). Scientists' collaboration strategies: implications for scientific and technical human capital. *Research Policy, 33*(4), 599-616.

Bukvova, H. (2010). Studying research collaboration: A literature review. *Sprouts: Working Papers on Information Systems, 10*(3), http://sprouts.aisnet.org/10-13.

Byrne, A., & Hansberry, J. (2007). Collaboration: Leveraging resources and expertise. *New Directions for Student Leadership, 2007*(114), 75-84.

Carroll, J. M., Rosson, M. B., Farooq, U., & Xiao, L. (2009). Beyond being aware. *Information and Organization, 19*(3), 162-185.

Chung, E., Kwon, N., & Lee, J. (2016). Understanding scientific collaboration in the research life cycle: Bio- and nanoscientists' motivations, information-sharing and communication practices, and barriers to collaboration. *Journal of the Association for Information Science and Technology, 67*(8), 1836-1848. doi:10.1002/asi.23520

Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science, 35*(5), 703-722.

Cummings, J. N., & Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy, 36*, 1620-1634.

Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management, 37*(6), 817-842.

Finholt, T. A. (2002). Collaboratories. *Annual Review of Information Science and Technology, 36*, 73-108.

Freeman, R. B., & Huang, W. (2015). Collaborating with people like me: Ethnic coauthorship within the United States. *Journal of Labor Economics, 33*(S1), S289-S318.

Hara, N., Solomon, P., Kim, S.-L., & Sonnenwald, D. H. (2003). An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. *Journal of the American Society for Information Science and Technology, 54*(10), 952-965. doi:10.1002/asi.10291
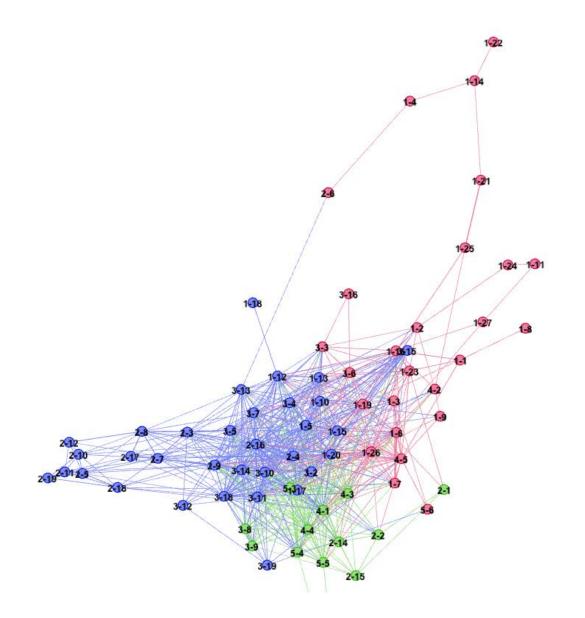
He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends, 48*(1), 133-159.

Katz, J., & Martin, B. (1997). What is research collaboration? *Research Policy, 26*(1), 1–18.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology, 27*(1), 415-444.

Sie, R. L., Drachsler, H., Bitter-Rijpkema, M., & Sloep, P. (2012). To whom and why should I connect? Co-author recommendation based on powerful and similar peers. *International Journal of Technology Enhanced Learning, 4*(1-2), 121-137.

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology, 41*, 643-681.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*(3), 387-420.

Stokols, D., Misra, S., Moser, R. P., Hall, K. L., & Taylor, B. K. (2008). The ecology of team science: understanding contextual influences on transdisciplinary collaboration. *American journal of preventive medicine, 35*(2), S96-S115.

Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy, 34*(10), 1608-1618.

Waruszynski, B. T. (2017). *Collaboration in scientific research: Factors that influence effective collaboration during a period of transformational change.* Royal Road University

Zhang, C., Bu, Y., Ding, Y., & Xu, J. (2018). Understanding scientific collaboration: Homophily, transitivity, and preferential attachment. *Journal of the Association for Information Science and Technology, 69*(1), 72-86.

ID	3-1	2-1	3-2	1-1	2-2	1-2	1-3	2-3	1-4	5-1	3-3	4-1	1-5	5-2	1-6	1-7	1-8	1-9	1-10	4-2	1-11	3-4	1-12	3-5	1-13	1-14	1-15	1-16
3-1	1	0	0.005319	0	0.039474	0.019231	0	0	0	0	0.034483	0.025424	0	0	0	0	0	0	0.036145	0.016393	0.05	0.005	0.008403	0.007407	0.007874	0	0.014599	0
2-1	0	1	0.082988	0.035088	0.126866	0.068376	0.046667	0.061111	0	0.010753	0.03937	0.071672	0.059172	0.019417	0.067901	0.071942	0.051546	0.023077	0.025974	0	0	0.062257	0.060773	0.055838	0.081081	0.010101	0.044554	0.095652
3-2	0.005319	0.082988	1	0.052885	0.100437	0.07109	0.061728	0.118774	0.021053	0.015625	0.114833	0.152778	0.147541	0.024876	0.109312	0.077586	0.041026	0.073394	0.116379	0.082949	0.021053	0.136778	0.118321	0.148699	0.157692	0.02551	0.147059	0.107317
1-1	0	0.035088	0.052885	1	0.048077	0.092105	0.084906	0.056738	0.02	0.019608	0.071429	0.058366	0.087302	0.032787	0.082645	0.080808	0.111111	0.045977	0.045045	0.057471	0.0625	0.089623	0.079137	0.064103	0.07483	0.035714	0.062893	0.166667
2-2	0.039474	0.126866	0.100437	0.048077	1	0.054545	0.064748	0.109756	0.02439	0.024096	0.078947	0.121324	0.055901	0.032258	0.086093	0.085271	0.068966	0.059829	0.049296	0.02459	0.037037	0.090909	0.076471	0.105556	0.079096	0.011111	0.068783	0.054054
1-2	0.019231	0.068376	0.07109	0.092105	0.054545	1	0.089286	0.026316	0.017544	0	0.102273	0.064885	0.058824	0	0.095238	0.055556	0.030769	0.042553	0.088496	0.1	0.09434	0.081818	0.068027	0.094937	0.107383	0.031746	0.035294	0.109756
1-3	0	0.046667	0.061728	0.084906	0.064748	0.089286	1	0.05618	0.071429	0.011111	0.075	0.087413	0.128205	0.030303	0.125828	0.123077	0.042105	0.083333	0.139706	0.091667	0.046512	0.084337	0.118343	0.062176	0.106742	0.054348	0.130435	0.108108
2-3	0	0.061111	0.118774	0.056738	0.109756	0.026316	0.05618	1	0.050847	0.008065	0.051613	0.068111	0.147541	0.022556	0.035533	0.052632	0.031008	0.025	0.067797	0.037736	0.00813	0.085714	0.115	0.106481	0.094787	0.015504	0.1	0.075342
1-4	0	0	0.021053	0.02	0.02439	0.017544	0.071429	0.050847	1	0.038462	0.048387	0.020661	0.046729	0.027027	0.029126	0.093333	0.060606	0.047619	0.034091	0.063492	0.083333	0.009804	0.05042	0.014388	0.023077	0.1	0.035971	0.035088
5-1	0	0.010753	0.015625	0.019608	0.024096	0	0.011111	0.008065	0.038462	1	0.015385	0.029046	0.018018	0.258065	0.009434	0.012195	0.058824	0.046875	0	0.0625	0	0.004854	0.016129	0.014286	0.015152	0	0.028369	0.016949
3-3	0.034483	0.03937	0.114833	0.071429	0.078947	0.102273	0.075	0.051613	0.048387	0.015385	1	0.079245	0.086331	0.026667	0.08209	0.061404	0.042254	0.09375	0.130435	0.049505	0.065574	0.113636	0.108108	0.118012	0.102564	0.014085	0.08284	0.139535
4-1	0.025424	0.071672	0.152778	0.058366	0.121324	0.064885	0.087413	0.068111	0.020661	0.029046	0.079245	1	0.106312	0.027778	0.127586	0.10989	0.044898	0.055147	0.106383	0.099237	0.042194	0.123684	0.094937	0.141956	0.10625	0.02834	0.137072	0.064639
1-5	0	0.059172	0.147541	0.087302	0.055901	0.058824	0.128205	0.147541	0.046729	0.018018	0.086331	0.106312	1	0.042017	0.142857	0.105263	0.061404	0.078014	0.282609	0.077465	0.046729	0.140625	0.146739	0.112745	0.152632	0.053097	0.263736	0.132813
5-2	0	0.019417	0.024876	0.032787	0.032258	0	0.030303	0.022556	0.027027	0.258065	0.026667	0.027778	0.042017	1	0.035088	0.021739	0.044444	0.054054	0.019802	0.025974	0	0.018692	0.037879	0.040816	0.028369	0	0.04	0.028986
1-6	0	0.067901	0.109312	0.082645	0.086093	0.095238	0.125828	0.035533	0.029126	0.009434	0.08209	0.127586	0.142857	0.035088	1	0.5	0.045455	0.073529	0.096154	0.113636	0.049505	0.087452	0.102151	0.0625	0.103627	0.056075	0.131313	0.112
1-7	0	0.071942	0.077586	0.080808	0.085271	0.055556	0.123077	0.052632	0.093333	0.012195	0.061404	0.10989	0.105263	0.021739	0.5	1	0.05814	0.042735	0.057554	0.088496	0.037975	0.069388	0.090361	0.05914	0.086207	0.059524	0.111111	0.095238
1-8	0	0.051546	0.041026	0.111111	0.068966	0.030769	0.042105	0.031008	0.060606	0.058824	0.042254	0.044898	0.061404	0.044444	0.045455	0.05814	1	0.041667	0.020408	0.041096	0.029412	0.033654	0.046875	0.020408	0.044118	0.02439	0.047945	0.079365
1-9	0	0.023077	0.073394	0.045977	0.059829	0.042553	0.083333	0.025	0.047619	0.046875	0.09375	0.055147	0.078014	0.054054	0.073529	0.042735	0.041667	1	0.13913	0.103093	0.081967	0.042373	0.057692	0.022599	0.054878	0.028169	0.101796	0.064516
1-10	0.036145	0.025974	0.116379	0.045045	0.049296	0.088496	0.139706	0.067797	0.034091	0	0.130435	0.106383	0.282609	0.019802	0.096154	0.057554	0.020408	0.13913	1	0.073171	0.070588	0.101626	0.137725	0.101604	0.125	0.042553	0.174157	0.087719
4-2	0.016393	0	0.082949	0.057471	0.02459	0.1	0.091667	0.037736	0.063492	0.0625	0.049505	0.099237	0.077465	0.025974	0.113636	0.088496	0.041096	0.103093	0.073171	1	0.080645	0.042194	0.0375	0.04	0.054545	0.072464	0.10119	0.123596
1-11	0.05	0	0.021053	0.0625	0.037037	0.09434	0.046512	0.00813	0.083333	0	0.065574	0.042194	0.046729	0	0.049505	0.037975	0.029412	0.081967	0.070588	0.080645	1	0.019802	0.02459	0.014388	0.023077	0.064516	0.058824	0.053571
3-4	0.005	0.062257	0.136778	0.089623	0.090909	0.081818	0.084337	0.085714	0.009804	0.004854	0.113636	0.123684	0.140625	0.018692	0.087452	0.069388	0.033654	0.042373	0.101626	0.042194	0.019802	1	0.125461	0.158845	0.125899	0.009479	0.117241	0.127358
1-12	0.008403	0.060773	0.118321	0.079137	0.076471	0.068027	0.118343	0.115	0.05042	0.016129	0.108108	0.094937	0.146739	0.037879	0.102151	0.090361	0.046875	0.057692	0.137725	0.0375	0.02459	0.125461	1	0.116279	0.208333	0.03937	0.140845	0.128571
3-5	0.007407	0.055838	0.148699	0.064103	0.105556	0.094937	0.062176	0.106481	0.014388	0.014286	0.118012	0.141956	0.112745	0.040816	0.0625	0.05914	0.020408	0.022599	0.101604	0.04	0.014388	0.158845	0.116279	1	0.112108	0	0.106838	0.0875
1-13	0.007874	0.081081	0.157692	0.07483	0.079096	0.107383	0.106742	0.094787	0.023077	0.015152	0.102564	0.10625	0.152632	0.028369	0.103627	0.086207	0.044118	0.054878	0.125	0.054545	0.023077	0.125899	0.208333	0.112108	1	0.021898	0.140909	0.121622
1-14	0	0.010101	0.02551	0.035714	0.011111	0.031746	0.054348	0.015504	0.1	0	0.014085	0.02834	0.053097	0	0.056075	0.059524	0.02439	0.028169	0.042553	0.072464	0.064516	0.009479	0.03937	0	0.021898	1	0.041379	0.047619
1-15	0.014599	0.044554	0.147059	0.062893	0.068783	0.035294	0.130435	0.1	0.035971	0.028369	0.08284	0.137072	0.263736	0.04	0.131313	0.111111	0.047945	0.101796	0.174157	0.10119	0.058824	0.117241	0.140845	0.106838	0.140909	0.041379	1	0.10625
1-16	0	0.095652	0.107317	0.166667	0.054054	0.109756	0.108108	0.075342	0.035088	0.016949	0.139535	0.064639	0.132813	0.028986	0.112	0.095238	0.079365	0.064516	0.087719	0.123596	0.053571	0.127358	0.128571	0.0875	0.121622	0.047619	0.10625	1
1-17	0.009346	0.035971	0.164671	0.069565	0.085603	0.072034	0.117647	0.09622	0.037559	0.023041	0.106383	0.20436	0.149813	0.040179	0.140152	0.135246	0.026786	0.069672	0.11284	0.082645	0.057416	0.123249	0.122807	0.135135	0.119454	0.031674	0.189474	0.076271
2-4	0.015152	0.079051	0.147239	0.069444	0.118644	0.101852	0.144068	0.138577	0.035176	0.014706	0.084071	0.179558	0.140625	0.023474	0.130435	0.10084	0.033654	0.064935	0.119835	0.102679	0.040404	0.122093	0.155303	0.180147	0.138182	0.039024	0.16129	0.086364
1-18	0.041667	0	0.025907	0.037736	0.023256	0.033333	0.021739	0.024	0	0	0.029851	0.020325	0.026549	0	0.018519	0.011765	0	0	0.043956	0	0.071429	0.009615	0.151786	0.028369	0.045802	0	0.02069	0.032787

Appendix 1. Jaccard's similarity coefficients (screenshot of a part of the results spreadsheet, and to conceal identity of the faculty members, their names have been coded in a way.)

Appendix 2. Estimated Interconnectivity (To conceal identity of the faculty members, their names have been coded in a way.)