RESEARCH

Insight-Driven Analysis of Co-authorship Networks in Philippine State Universities and Colleges

Marvee dela Resma[*†], Patrick Guillano La Rosa, Miguel Carlo Pingol, Christian Alfred Soriano, Damian Dailisan, Michael Dorosan and Erika Fille Legara

[*]Correspondence: MarveeDe-laResma.MSDS2022@aim.edu
Aboitiz School of Innovation, Technology and Entrepreneurship, Asian Institute of Management, 123 Paseo de Roxas, Makati City, Philippines
Full list of author information is available at the end of the article
[†]Equal contributor

Abstract

Here, we present insights into the existing collaboration dynamics of state universities and colleges (SUC) in the Philippines. Ultimately, the objective is to contribute to optimizing the R&D productivity of SUCs, a critical component in the generation of new knowledge, fostering innovation, and the long-term economic development of a developing country. In our approach, we identified three progressive profile sets of SUC co-authorship networks (Figure 1) that can be useful in crafting strategies to further collaborations. To wit, networks in Profile 1 were found to be highly disconnected with no prominent components; instead, they consist of dyads, small star-like networks (ego), and small isolated cliques. We also differentiated between co-authorship networks comparable to small-world networks (Profile 2) and Profile 3, which implicates the emergence of large components from the eventual linkage of smaller components found in both Profile 1 and 2. From here, university-level policies are suggested to improve or develop research collaboration systems. We also performed exploratory network analysis to identify critical entities in the networks, exploring various centrality measures focused on insights-extraction. As part of the descriptive analysis, we also extracted network communities (Figure 2) that may be significant in helping funding agencies identify relevant R&D programs for various SUCs.

Keywords: Co-authorship Network; Collaboration Profiling; Community Detection; Degree Centrality; Small-World Networks; Research Policy

1 Introduction

The Commission on Higher Education (CHED) holds a national mandate for improving the standing and quality of research efforts conducted by Philippine State Universities and Colleges (SUCs) through various policies and programs. Through this mandate, the organization aims to prioritize research that is multidisciplinary, policy oriented, and participative which explicitly places emphasis on integrative cooperation. The reason is apparent, being that collaborative research encourages interdisciplinary ideas, develops new skills, and provides access to different sources of funding (Bansal et al., 2019). All scientific progress was built on knowledge that came before it and countless scientific breakthroughs had teams composed of different backgrounds working towards a common purpose.

However, encouraging collaborative research has always been a challenge due to the underlying systemic organizational nuances present within established institutions. Network science presents an approach for analyzing the dynamics behind these individuals and their groups (Chen, 2022). The field offers a systematic interpretation of organizational social dynamics which is applicable to academic research institutions, and potentially yields significant insights for the formation of research policy.

2 Related studies

According to NHERA (2009), networking is an important consideration in identified priority themes so higher education institutions in the Philippines can coordinate and pool their resources together. In research analytics, collaboration between authors is proxied by their co-authorship of a research paper (Kumar, 2015). Patterns and relationships within these networks generate insights on how researchers behave. In a co-authorship network, the nodes are represented by authors and the edges represent the frequency of their shared work. Additional attributes can be attached to any of the nodes or edges. Weights of the edges can also be modified depending on the use case. Literature on co-authorship networks is still growing and this section focuses on the use of these generated insights in policy making.

Network analysis of co-authorships can be used to assess and evaluate metrics for public policy. A co-authorship network analysis of Brazilian tuberculosis research (Vasconcellos and Morel, 2012) revealed that the Law of Innovation passed by the Brazilian government could not overcome systemic deficiencies of collaboration between the public and private sectors and prompted further adjustments in maintaining the continuity of production development and the need for government support for infrastructure to improve research efforts.

In another use case, co-authorship as a proxy for collaboration was used to evaluate a policy promoting inter-programmatic collaboration over time within four research programs at the Markey Cancer Center (MCC) at the University of Kentucky (Fagan, et al., 2018). The goal of their program was to increase collaboration to foster innovation and develop best practices within departments. Co-authorships are social networks therefore policies created from generated insights should be made considering these nodes as people and their behaviors are indicative of that. Funding both formal and informal programs where researchers could network is an example of this people-centric policy adoption. The formal programs included research grants specifically for interdepartmental research. After this, the informal events such as symposia and networking events allowed for authors to get to know each other.

Beginning with departmental cliques of authors as the largest components in 2007, the policies implemented grew their research collaboration network and grew into a giant component crossing different research departments in just 7 years (Fig 1). MCC and their research efforts increased diversity of fields of study and co-authorships based on previous networks as benchmarks.

3 Data Collection

The data was collected from 11 participating State Universities and Colleges and the Commission on Higher Education also provided additional data from other universities in the Philippines. Information about research undertakings from 2017 to 2021 were gathered and used for this project. Data features included the research

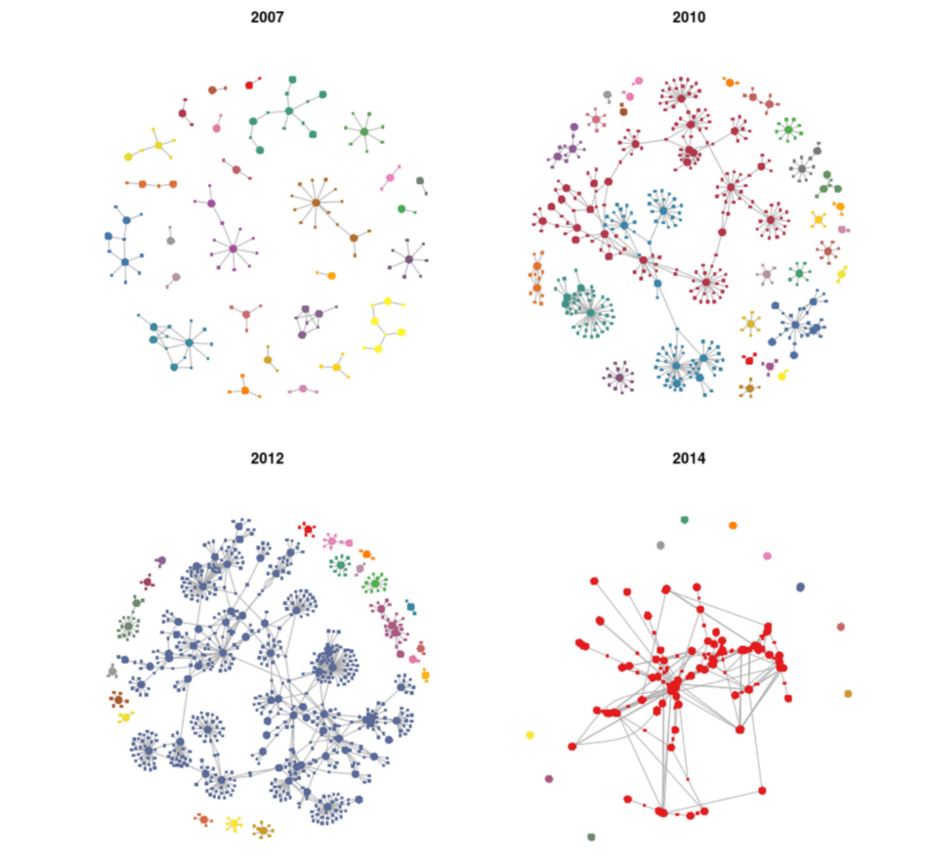

Figure 1 Markey Cancer Center research collaboration network over time. Implementing policies fostering interdepartmental collaboration greatly improved their collaboration network.

title, authors, year, abstract and corresponding keywords. A total of 1264 research works were included in the study.

4 Methodology

4.1 Individual SUC Networks

It is important to analyze co-authorship dynamics at a university level to identify the strong points and areas of improvement in order to create a more robust collaboration network. The network analysis methodology is seen in Figure 2.

Figure 2 Methodology for the SUC Collaboration Network Profiling & Analysis.

4.1.1 Network Creation

The research co-authorship network for the study is an undirected weighted network representing collaborations. The collaboration pairs were determined for each research paper and the corresponding pairs could appear multiple times if they also collaborated in other research works. Repeated pairings were represented as multi-links in the network. In this case multi-links or repeated collaboration are represented by edge weights. Each of the state universities were analyzed separately to discover internal collaboration dynamics.

The authors describe the network using set $A = \{a_1, a_2, \ldots, a_m\}$ which is the set of authors found in the database. The authors then define E as a set of pairs where each $e = (a_i, a_j) \in E$, corresponds to a pair a_i and a_j who have collaborated on at least one research paper. This relationship is then represented in an adjacency matrix of size $m \times m$, where both rows and columns are research authors. Using any pair e, the value in the adjacency matrix will be the number of times a pair of authors collaborated. These matrices were represented, visualized, and analyzed throughout this study.

Analysis for each state university was done and network properties were computed, such as the average degree, degree centrality, node-component ratio, betweenness centrality, average shortest path length and clustering coefficient. These network properties were later used in profiling the SUC (Fig 3).

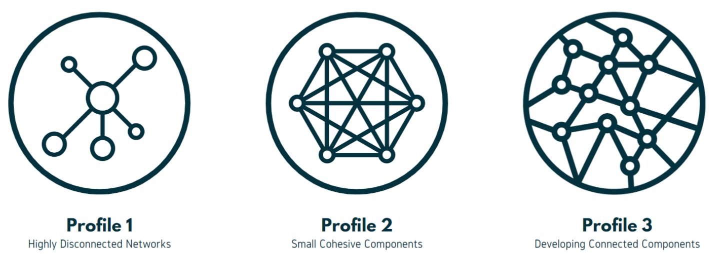

Figure 3 Three (3) collaboration network profiles to be explored in this study.

4.1.2 Connectivity

The degree is the number of links to a node, in this case the number of research collaborations an author is involved in. The average degree describes the mean number of connections a node has in a particular network. Meanwhile, a component is a single connected entity in the network that could stand separately. It could be a single node, a pair, or a multi-node component. In a co-authorship network, a single component could be a solo research project (single node), or a collaboration between two or more authors. A network could be one single connected component, given all nodes are connected, or several isolated components.

The node-component ratio is an assessment of the number of nodes and components found in a network. A low value could be an indication of a network's lack of connectivity, meaning a network could have a lot of nodes but disconnected components. This metric was the basis for the first profile of networks identified in this study. *Profile 1* networks are those with *highly disconnected components.*

They have no prominent components (Fig 4); instead, they consist of dyads, small star-like networks (ego), and small isolated cliques. Due to this, further analysis of its network properties has limited value. The networks not belonging to this profile were further analyzed using other metrics.

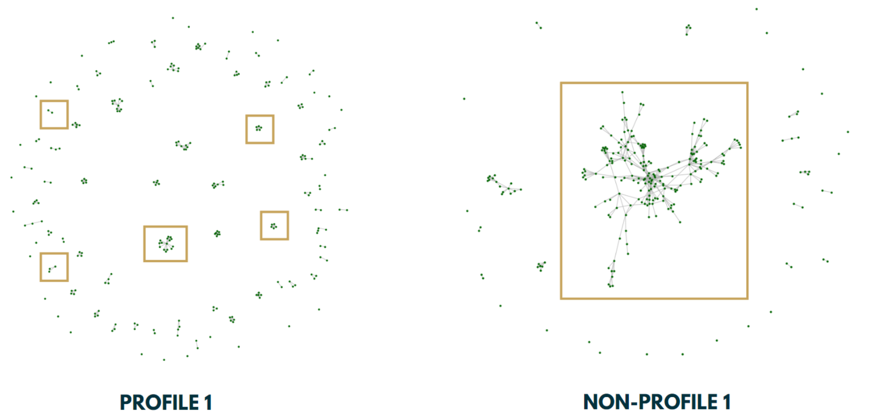

Figure 4 Examples of Profile 1 and non-Profile 1 networks emphasizing the lack of large connected components in Profile 1 networks. Each yellow box contains a single component of the network.

4.1.3 Giant Component Identification

The next part of the analysis for non-profile 1 networks involves the identification of the network's largest component. A giant component is the component containing the greatest number of connected nodes in the network. This will represent the most connected component of the network.

4.1.4 Small World Properties

After extracting each network's giant component, they were analyzed by their average shortest path length and clustering coefficient. The average shortest path is the average number of nodes along the shortest path between a pair of nodes in a network. In this collaboration network, it is the average number of authors you must connect with before eventually reaching your target author. Meanwhile the clustering coefficient is the tendency of nodes to cluster together, or in this case authors to collaborate together. These metrics are often connected to small world properties. A small world network is described as "highly clustered, like regular lattices, yet have small characteristic path lengths" (Watts and Strogatz, 1998). As for this co-authorship social network, a small world phenomenon is characterized by authors linked to each other by a short chain of acquaintances. A small world network is described to have low average short path lengths and high clustering coefficient values.

These properties were used to create the other profiles. *Profile 2* networks are characterized as having giant components with low average path lengths and high clustering coefficients much like small world networks. Profile 2 networks are described as having *small cohesive components*. Meanwhile, *Profile 3* networks exhibit longer average short path lengths for their giant component compared to Profile 2 networks. They are usually made up of smaller components, probably those from Profile 1 and 2 linking together to form *larger and developing connected components*.

4.1.5 Centrality Analysis

Several centrality measures were used to describe and analyze the largest components of Profile 2 and 3 networks. The study included degree centrality, betweenness centrality and closeness centrality to identify important nodes, or authors in the collaboration networks.

Degree centrality is a measure of the connections of one node. High degree centrality means that a certain author is directly connected to the greatest number of authors through collaborations. Due to this they are usually called the hubs of the network.

Betweenness centrality is a measure of the frequency that a node lies on the shortest path from one node to another. It gives a premium to authors which are most traversed when looking for shortest paths between two different authors. Those with high betweenness centralities are mostly identified as bridges in a network as they are vital in connecting two different components of a network. They are elements of vertex cuts that could break a component into smaller ones. For this study we can see them as authors who connect two groups/departments or disciplines to collaborate on a specific research topic.

Finally, closeness centrality is the measure of the average length of a node's shortest path to other nodes found in the network. A node with a high closeness centrality means that it is nearer to the other nodes which does not necessarily imply a direct connection. They are described as optimized hubs or those that have connections to other highly connected nodes. They can be considered as the authors who have the best connections from other departments or research groups.

4.1.6 Modularity Analysis

The final step was based on the giant component's modularity describing the structure of the network in terms of its communities. Modularity measures the strength of division of a network into modules or communities. This analysis was done to Profile 3 networks only because their giant components are large enough to break down further. Communities refer to dense sets of nodes with higher connectivity within each of its nodes compared to others identified outside of the community. Modularity also refers to the potential of division of a network into communities. This part of the methodology identified and analyzed the resulting communities for each SUC co-authorship network.

4.2 Author-Topic Network

Topic modeling was leveraged to create an inter-university network based on research topic classifications due to the lack of co-authorship links between authors from different universities. This author network has two node sets, first for the authors and second for the unique topics or fields of research. The different links between the two node sets were also defined.

4.2.1 Topic Modeling

The topics were determined through a separate unsupervised methodology. Research abstracts were first transformed into document embeddings and clustered with HDBSCAN (cite) to determine the best clusters to represent the fields of research undertaken by the SUCs.

4.2.2 Network Creation

The topic distribution for the research papers were used to create author-topic pairs to connect the two node sets to form a bipartite network. Corresponding projection networks were then created for each node-set. In the resulting projected networks, edge weights were defined by the number of times the common association was repeated.

4.2.3 Centrality and Modularity Analysis

The same method of analysis using centrality and modularity from the individual topic networks was applied to the projected author network.

5 Exploratory Data Analysis

5.0.1 Author Distribution

Several author metrics were measured in terms of collaboration for this dataset. A total of 550 or 43% of the research projects involved only a single author (Fig 5). Extreme cases were found where more than 20 authors were involved in a single research project with a maximum of 32, leaving room for possible collaborations.

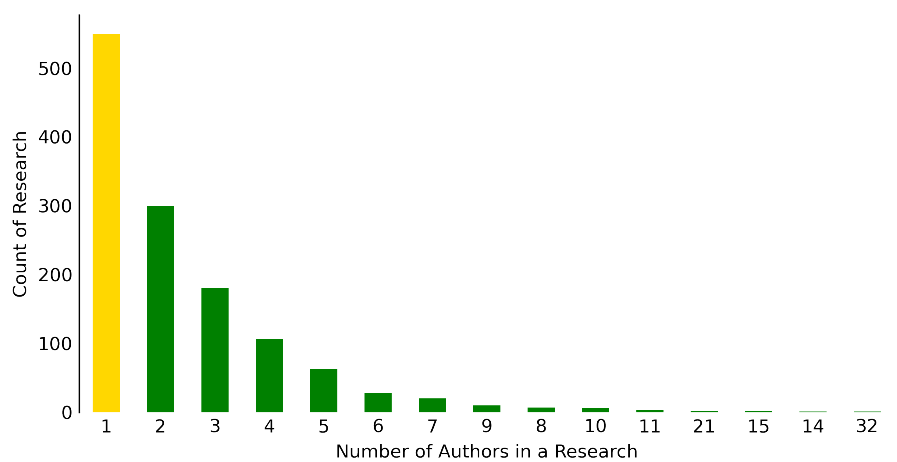

Figure 5 Topic distribution per collaboration type. 43% of research projects had a single author.

5.0.2 Author-Topic Distribution

The ratio of solo research to collaborative research was also determined for each topic to determine any preferences in the field. Most research fields tend to collaborate except for Covid-19 Pandemic with a preference for solo research(Fig 6).

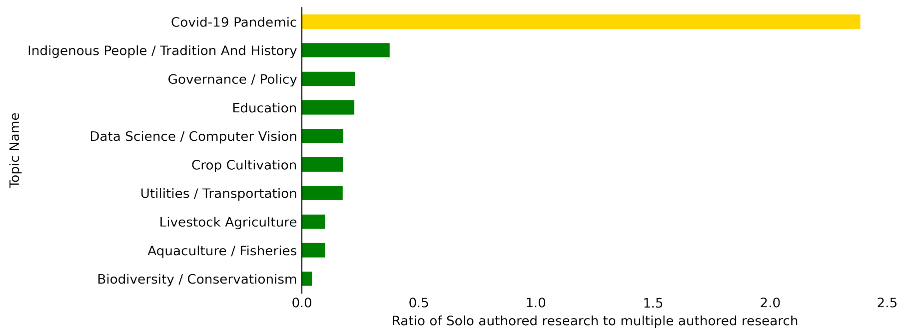

Figure 6 Ratio of Solo Authored Research to Multiple Authored Research.

6 Results

6.1 Individual SUC Networks

Individual networks for each SUC were created to determine the differences between their basic network properties such as the node counts, edge counts, component counts, average degree, and node-component ratios. Comparison of these different properties uncovered 3 levels of collaboration present throughout the different SUCs which the authors categorized into corresponding profiles.

Profile 1 describes highly disconnected networks that lack prominent giant components. These are further characterized by a large presence of individual nodes, dyads, small star-like components, and small cliques quantified through a network's node-component ratio (Table 1). These can be interpreted as collaborations manifested by faculty and student partnerships. This ratio gives importance to the connectivity of each node present to a specific component. By setting a threshold of 4 for the node-component ratio, several universities were classified into Profile 1 such as Agusan del Sur State College of Agriculture and Technology (ASSCAT), Bulacan State University (BulSU), Mindoro State University (MinSU), Tarlac State University (TSU) and University of the Philippines-Baguio (UPB).

Table 1 Network Properties of SUC Co-Authorship Networks.

University	Nodes	Links	Average Degrees	Node-Component Ratio
ASIST	265	483	3.6	7.36
ASCAT	26	30	2.3	3.71
BSU	437	1259	5.7	5.68
BulSU	251	321	2.6	2.91
DNSC	100	208	4.2	11.11
MinSU	109	199	3.7	3.30
PSU	272	686	5.0	5.55
TSU	17	11	1.3	1.70
UPB	261	975	7.5	3.58
RSU	201	404	4.0	4.10
QSU	41	87	4.2	4.10

The general component networks for the remaining SUC networks were isolated and analyzed by their small world properties (Table 2) resulting in the use of average shortest path length and clustering coefficient as the basis of profiling (Fig 7).

Profile 2 SUC networks are characterized by the presence of small tightly knit or cohesive components which may indicate the existence of an imbalance between high research activity and external collaboration, as these groups stand out as isolated

Table 2 Network Properties of Profile 2 and 3 Giant Components (GC).

University	GC Nodes	GC Links	GC Average Degrees	Average Shortest Path Length	Clustering Coefficient
ASIST	193	422	4.37	5.71	0.61
BSU	157	469	6.00	6.18	0.79
DNSC	83	196	4.72	4.09	0.73
PSU	68	299	8.79	2.48	0.91
RSU	16	52	6.50	1.57	0.84
QSU	11	31	5.64	1.44	0.95

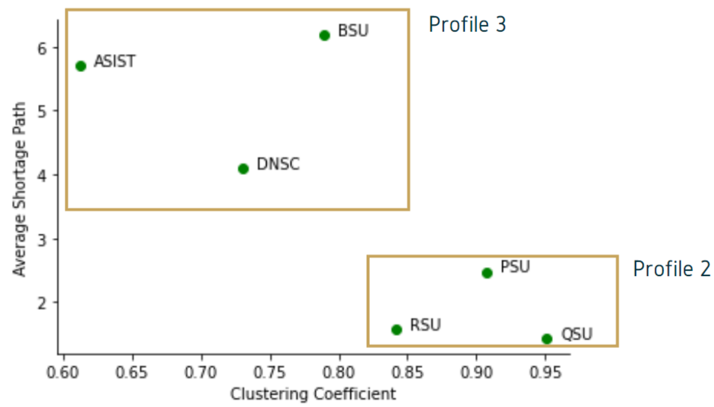

Figure 7 Small world property plot of the remaining networks' giant components Profiling was based on the average shortest path length and the clustering coefficient.

and exclusive research groups among the authors in an SUC. Strengthening of current inter-disciplinary research efforts could become the focus of these universities to build a network with increased collaboration.

In terms of research dynamics, these profiles of low short path values indicate closer connections through collaborations while high clustering coefficients indicate the existence of research groups. Consistent with small world networks, schools such as Romblon State University (RSU), Quezon State University (QSU) and Palawan State University (PSU) exhibited high clustering coefficients (> 0.8) and low shortest path length averages (< 3.0).

Lastly, Profile 3 SUC networks are characterized by giant components which have a large number of nodes. The large number of nodes implies the involvement of more individuals in the collaboration network. These universities are in the process of creating a large network of researchers because they already have existing connections between different groups. These universities exhibit a moderate clustering coefficient (between 0.6 to 0.8) but are mostly characterized with high average shortest path lengths (< 4) which indicates higher distance between nodes which is opposite of Profile 2. The high number of nodes makes it difficult for the components to be more connected to all the other nodes. Visually, they can also be characterized as outstretched networks compared to the more compact nature of profile 2 components (Fig 8).

The average clustering coefficient implies that there are existing research groups in the network. Lastly, the high shortest path values support the idea of interdisciplinary or inter-departmental collaboration through various connections, since it

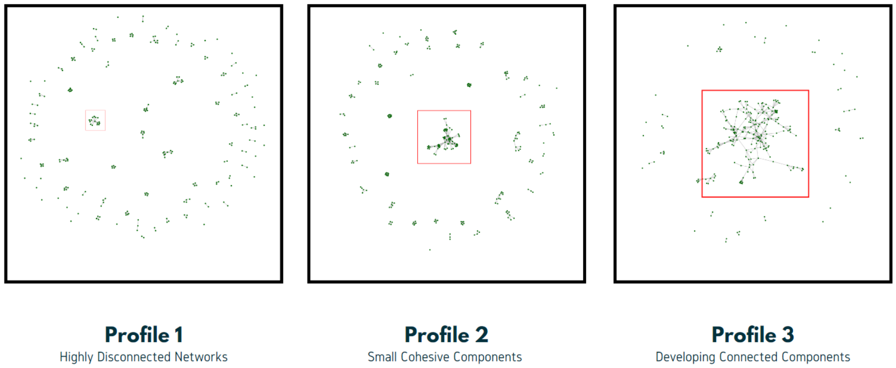

Figure 8 Examples of SUC co-authorship networks for each collaboration profile. The components inside the red boxes are the largest components of the network.

takes longer to get a link between two authors from different fields of research. Profile 3 networks also indicate the emergence of larger components in a network from the eventual linkage of the smaller components found from the other profiles. The following SUCs were classified under Profile 3: Abra State Institute of Science and Technology (ASIST), Benguet State University (BSU), and Davao del Norte State Colleges (DNSC), described as networks with large developing connected components.

6.1.1 Centrality Analysis

Centrality analysis was conducted to identify the essential nodes that keep the giant components connected. This yielded a list of prominent authors that play a major role in their SUC network (Table 3).

Table 3 Top Authors per Centrality Measure.

	Degree Centrality		Betweenness Centrality		Closeness Centrality	
University	**Name**	**Score**	**Name**	**Score**	**Name**	**Score**
	Vasquez	0.104	Gannapao	0.435	Atmosfera	0.282
ASIST	Gannapao	0.094	Vasquez	0.428	Gannapao	0.278
	Abalos	0.083	Atmosfera	0.365	Vazquez	0.269
	Lee	0.154	Dolipas	0.516	Dolipas	0.229
BSU	Doyog	0.147	Abiasen	0.505	Abiasen	0.226
	Lumbres	0.128	Gomez	0.500	Sajise	0.225
	Decano	0.244	Decano	0.421	Decano	0.374
DNSC	Gumanao	0.134	Rodriguez	0.325	Padios	0.343
	Sanatamaria	0.134	Padios	0.222	Duran	0.337
	Mangusad	0.507	Mangusad	0.480	Mangusad	0.657
PSU	Lactuan	0.403	Lactuan	0.229	Lactuan	0.573
	Caabay	0.329	Verturillo	0.166	Caabay	0.558
	Sa	1.000	Sa	0.343	Sa	1.000
RSU	Walitang	0.867	Walitang	0.195	Walitang	0.882
	Choi	0.600	Choi	0.052	Choi	0.714
	Camayang J	1.000	Camayang J	0.533	Camayang J	1.000
QSU	Legaspi	0.600	-	-	Legaspi	0.714
	Garingan	0.600	-	-	Garingan	0.714

ASIST has three prominent members in its author network, namely Mr. Vasquez, Mr. Gannapao and Ms. Atmosfera, who mostly dabble in research involving education, governance, and crop cultivation. Their expertise is on statistics and mathematics. They exhibit high metrics for all centrality measures (degree, closeness and betweenness), which indicate their roles as hubs or bridges and their importance in keeping their network connected. The interdisciplinary nature of their expertise could be key in collaborations between research groups. Through external verification, it was determined that the two authors who exhibited the highest values for degree and betweenness centrality also hold important positions in the university. Prof. Raymond Vasquez is the Assistant Director for Research and Development, while Dr. Jubert Gannapao is the Dean of the College of Agriculture. These explain their roles as authors with the most direct connections and authors who act as bridges between different components or departments. Meanwhile, with a high score for closeness centrality, Prof. Rynheart Atmosfera, who is the Chairwoman for Mathematics and Natural Sciences, is connected to both Dr. Gannapao and Prof. Vasquez. This reinforces her role as an optimized hub, who is connected to other important authors.

For Benguet State University, specific researchers such as Mr. Young Jin Lee and Ms. Nova Doyog exhibited high degree centrality indicating they are directly connected to other researchers through collaborative work and possibly mentoring. Meanwhile, researchers such as Professor Bretel Dolipas and Assistant Professor Jovalson Abiasen who mostly work on education or crop cultivation related initiatives, have high betweenness and closeness centrality indicating they collaborate with different groups or departments.

For Davao del Norte State College, the most notable researcher is Dr. Ronald Decano who has the highest score for all three centrality measures, indicating his vital role in the network. He is also the Dean of the Institute for Advanced Studies which offers several masters and doctorate programs with a wide range of field expertise. His professional expertise is on educational management and mathematics and he is involved in research topics such as education, risk assessment and some data science initiatives. Similarly, Palawan State University's (PSU) network also indicated a particular researcher who had the highest score for all centrality metrics. Ms. Vernaluz Mangusad is a researcher on topics centering mostly on Biology and Environmental Science. She also is involved in large co-authorship groups explaining her importance in the PSU's connectivity.

Identification of these individuals (Fig 9) can be used in analyzing social dynamics between researchers for each university and could also be a framework in analyzing the respective contributions of individual researchers.

6.1.2 Modularity Analysis

Modularity analysis serves as the next step in identifying the communities or modules present through network analysis and to infer reasons why certain groups exist using available data. Using Louvain heuristics, modularity of the network is maximized and, in the process, produces the best possible partition of the network into community divisions or modules present. A higher modularity score is indicative of more definite or distinct partitioning. Giant components of networks (Table 4) that belonged to Profile 2 exhibited a module count of 3-7 with low modularity scores, consistent with their small world properties. Meanwhile, those that are under Profile 3 had 9 or more modules identified with relatively higher modularity scores at 0.7. Profile 3 universities were analyzed, and some insights were inferred regarding the formed modules.

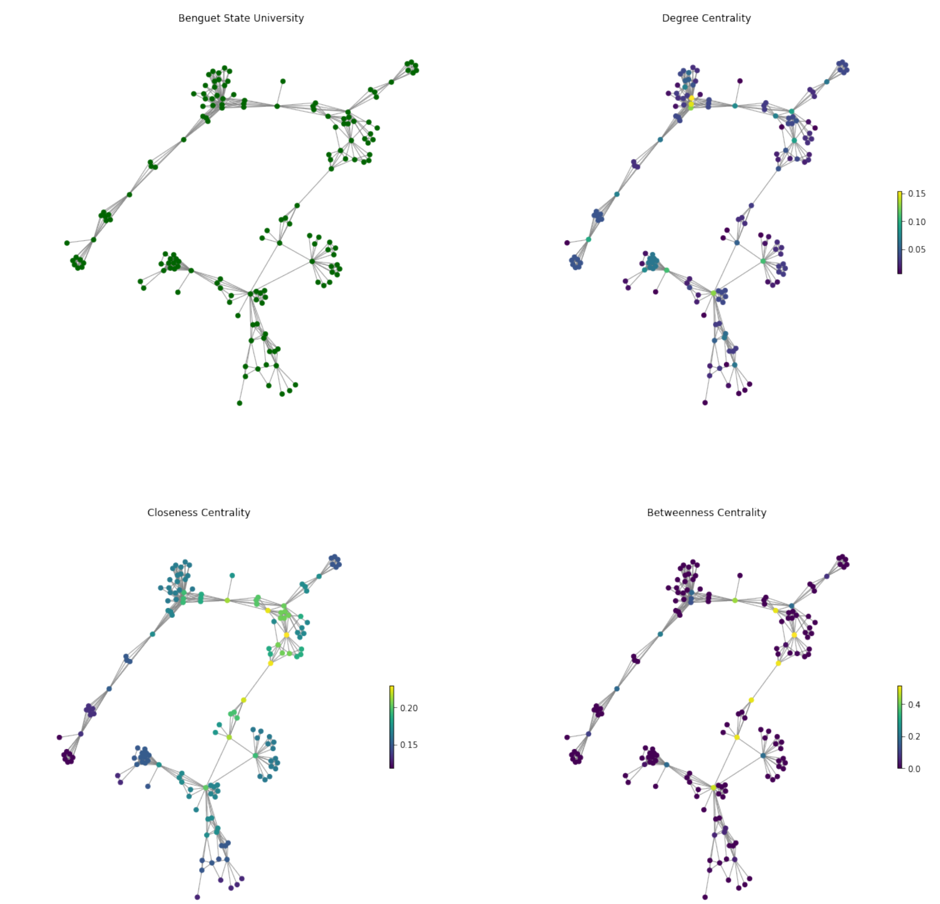

Figure 9 Example of a centrality network visualization for ASIST. Network visualizations for other universities are found in the appendix.

Table 4 Top Authors per Centrality Measure.

University	Modularity	No. of Communities
ASIST	0.786	12
BSU	0.806	9
DNSC	0.697	9
PSU	0.542	7
RSU	0.217	3
QSU	0.293	2

We identified smaller working groups, consisting of a main research group and authors a few of the main members worked with. There are nodes that connect their research group to the rest of the university research network, usually characterized by a high betweenness centrality score. More granular analysis of research dynamics could be done by using divisive methods in community detection wherein nodes with the highest betweenness centralities are removed from the network.

One of the Profile 3 SUC networks, DNSC, was analyzed using external verification to confirm the composition of the identified communities. The model yielded a list of authors per community detected and the connections between these community members was deduced using available information online. There were 9 communities detected for the DNSC network and most of the communities were based on existing departmental groupings and corresponding expertise (Fig 10). DNSC has 5 institutes namely Institute of Computing (IC), Institute of Advanced Studies (IAS), Institute of Teacher Education (ITE), Institute of Applied and Aquatics Sciences (IAAS) and Institute of Leadership, Entrepreneurship and Good Governance (ILEGG). Each of these institutes have their own set of departments as well (Table 5).

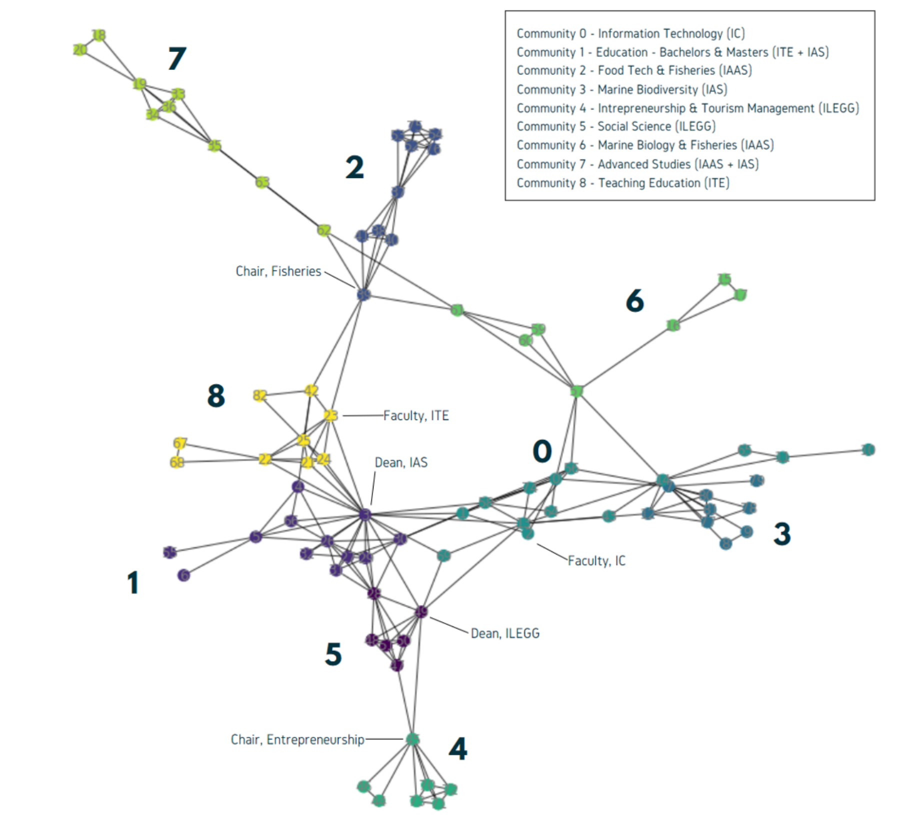

Figure 10 DNSC modularity analysis network plot. Each community is color coded and numbered with the corresponding composition analysis found on the community legend. Authors with high betweenness centralities were also identified and placed in the plot to see how they act as bridges between the communities identified.

Some communities are made up of authors coming from the same institute such as Community 2 who are all from IAAS but are composed of the Food Technology and Fisheries departments, indicating interdisciplinary collaboration fostered by being members of a common institute. The same can be said for Community 6 who are also from IAAS but the collaboration is now between members of the Fisheries and Marine Biology departments. This time the collaboration was most likely fostered by the connection of expertise of both departments. Another example of this is Community 4 which is from ILEGG. Their expertise are Entrepreneurship and Tourism Management, which are also related.

Table 5 Top Authors per Centrality Measure.

Institute	Departments
Institute of Applied and Aquatic Sciences (IAAS)	Agroforestry, Fisheries and Aquatic Sciences, Food Technology, Marine Biology
Institute of Computing (IC)	Information Technology, Information Systems
Institute of Leadership, Entrepreneurship (ILEGG)	Disaster Resiliency Management, Entrepreneurship, Public Administration, Social Science, Social Work, Tourism Management
Institute of Teacher Education (ITE)	Professional Education, General Education, Science, English, Mathematics
Institute of Advanced Studies	Educational Management (PhD & MA), English (MAEd), Fisheries Management, Marine (MM) Biodiversity (MS), Biology (MST), Mathematics (MST)

Other communities involve only one institute and one department such as Community 0 composed of authors mostly from the IC's IT department and Community 5 with authors mostly from the Social Sciences department from ILEGG.

Other dynamics were also discovered like Community 3 where some of its members are from IAS, all involved in the advanced studies of Marine Biodiversity collaborating with foreign authors. Upon external validation, there were several research collaborations on marine life headed by Dr. Girley Gumanao, a faculty of the IAS under the Marine Biodiversity department, which involved foreign authors.

Since communities were already detected, the bridge nodes or linkages between these different communities were also checked (Figure 50). Following our previous analysis on betweenness centrality, the authors with the highest scores were identified in the network and then checked which communities they help bridge or connect. For example, Dr. Decano, the Dean of IAS and a faculty member of ITE is the link between Community 0 (Information Technology), Community 1 (Teacher Education) and Community 3 (Marine Biology). His position as the dean of the Institute of Advanced Studies is why he can collaborate with various disciplines. Next is Ms. Bernadita Rodriguez, the Chairperson for Fisheries from IAAS who serves as the bridge between Communities 2, 6 and 7 which are described as those involved in marine biology and fisheries both at the undergraduate and graduate studies level.

Another example is Ms. Florie Fermil, the Dean of ILEGG who links the two ILEGG departmental communities (4 & 5). Mr. John Duran, a faculty member who connects the two ITE communities (1 & 8) and Mr. Ian Padios, an IT faculty member who connects to non-IT communities such as Community 1 (Education) and 5 (Social Sciences) due to some research where IT expertise was needed.

6.2 Author-Topic Network

A total of 1,182 authors and 11 unique research topics were used which yielded a combined total of 1295 unique links or edges. The bipartite network was then used to create separate projection networks for the authors and the topics in Table 6.

6.2.1 Centrality Analysis

For the projected author network a similar analysis was done wherein centrality measures and modularity were explored. There were 5 authors who consistently had the highest scores for both degree and closeness centrality measures. Mr. Panalba,

Table 6 Topic Distribution of Research.

Topic	Number of Research
Education	239
Crop Cultivation	203
Data Science	51
Government and Policy	48
Indigenous people	39
Covid-19 Pandemic	37
Utilities and Transportation	21
Livestock Agriculture	20
Fisheries	18
Biodiversity and Conservation	15

who is from BulSU with expertise in the social sciences, was involved in various research topics such as Crop Cultivation, Indigenous Peoples, Education and Governance. Included in the list are Mr. Agutaya from MinSU, Ms. Palaoag from BSU, Mr. Gannapao from ASIST and Mr. Padios from DNSC. They are involved in research related to various topics but mostly Crop Cultivation and Education. Some of them even used their expertise to get involved in a variety of research topics. An example is Mr. Padios whose expertise is in Information Technology. He was able to contribute to research projects of various topics which involved specific skills from his expertise such as website development, GIS, and web security. He also exhibited this collaboration dynamic in his own SUC network. Their importance in the researcher network can be narrowed down to the (1) quantity of collaborations they have done and the (2) variety of research topics they are involved in.

6.2.2 Modularity Analysis

Knowing that they already have several topics the authors still explored the ways that this author network could be separated into communities. The network yielded a moderate modularity score of 0.47 and 7 corresponding communities (Fig 11). Compared to the 11 topics, there were only 7 communities that emerged from the modularity analysis. Four of the communities represented a single topic which includes Crop Cultivation, Education, Pandemic and Biodiversity. Meanwhile some topics combined into the 3 remaining communities: (1) Government and Indigenous Peoples, (2) Data Science and Marine Biology and (3) Livestock and Utilities.

7 Conclusion

Information about the collaboration dynamics of the different SUCs were uncovered by using network analysis. The first major insight is the profiling of co-authorship networks based on its component connectivity and its giant components properties. It was seen that different universities are currently in different stages of developing their own co-authorship networks. The transition from small, isolated pairs or groups, to establishing more connections within those small components to form more cohesive units and finally to the eventual linkage of these individual components to form a larger network of authors is the main development pipeline.

The SUCs currently in profile 1 (highly disconnected networks) could investigate and identify these small groups and pool those that have more commonalities like topics or departmental expertise. One way to advance these networks into Profile 2 is by incentivizing the quantity of research output per department which will nudge people with the same expertise to collaborate more with each other leading to more

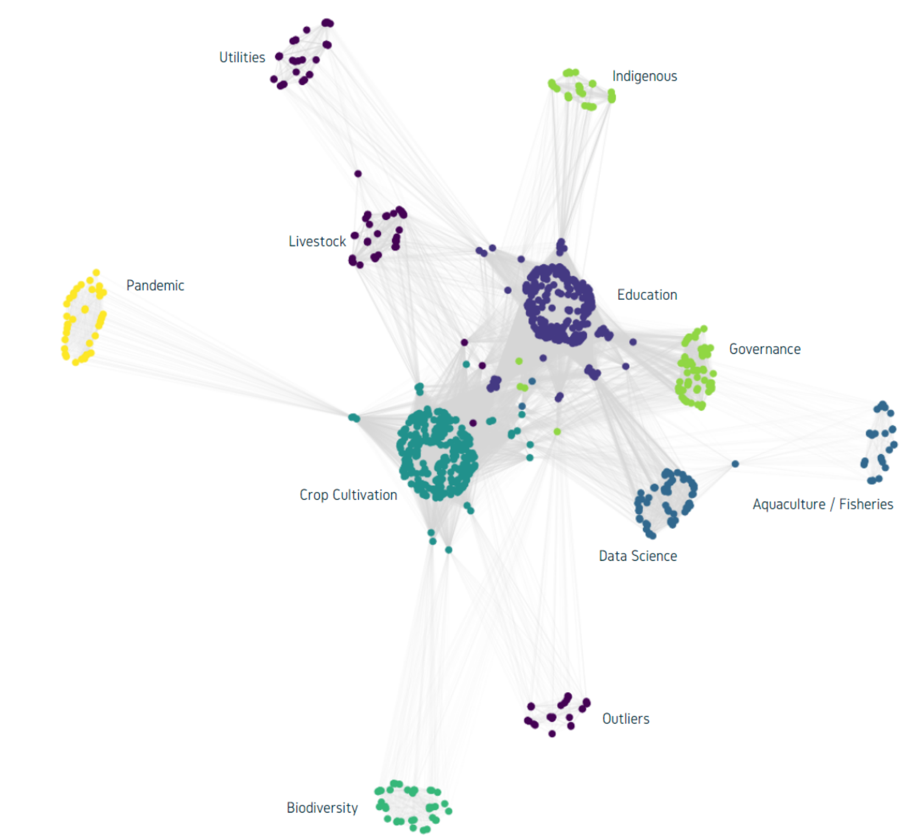

Figure 11 Author Projected Network Modularity Visualization

cohesive networks. This could be implemented by giving directives to the important authors of the departments, in terms of authority and standing in research quality. Increasing output does not ensure an increase in quality, so the universities will also have to set other metrics to ensure that they do not focus purely on quantity. After creating more cohesive units, inter-disciplinary research will play its role in advancing and improving the research network. Connecting departments from the same institutes or across other institutes by identifying multidisciplinary topics is key. Incentivizing these types of research projects by providing more funding for research that involve authors from different fields could help push this initiative. Tasking institute heads to target this could also help cascade this objective of improving the collaboration network. Initiating seminars or consortiums for knowledge sharing could also help foster these types of collaborations. This will eventually lead to Profile 3 networks or those with larger developing connected components. It must be noted that there is no upper limit for this profile. It could involve adding more authors into the network or improving the component's cohesiveness despite the growing number of authors. Both will contribute to creating a good co-authorship social network.

Other insights such as identification of important nodes who are vital to the network's connectivity is also important as they can be assigned as mentors and research leaders in terms of promoting collaboration and interdisciplinary research.

Community detection is also important in identifying the existing research groups within the network. This could identify a certain expertise per community and can be used to identify which connections should be utilized. It could also be used to identify communities who belong to the same collaboration network but are not directly connected. This gap can be reduced by means of knowledge sharing exercises or brainstorming sessions. In a larger scale such as the inter-university network, these communities could become working groups of the same expertise with members coming from various universities.

8 Recommendations

An increase in participation from more SUCs could improve the analysis and profiling of the different co-authorship networks. As it stands, only 11 SUCs submitted data with some only partially submitting. Including all of the data will guarantee that the corresponding interpretations will be appropriate and relevant to the respective SUCs. Granular information about the authors such as basic demographics, department affiliation and relevant positions, will help in interpreting node importance and community formation. Finding legitimate inter-university co-authorship relations would also pave the way into creating an inter-university network based on authorship.

9 Limitations

A limitation of the method is the misrepresentation of the authors depending on their names found in the database. Multiple entries that refer to the same person can occur if author names were misspelled. The authors of this study did their due diligence by verifying the research authors of SUCs manually. Another limitation is that only one topic was assigned to each research work in the author-topic bipartite network. Authors who are related to research works with no assigned topics were excluded from the inter-university network.

10 Assumptions

The list of authors per research were assumed complete with only declared authors in the data collection process included in the network. Duplicate research work entries were identified. The network reflects and assumes equal contribution and does not discriminate between the principal and collaborating authors.

Abbreviations

SUC, state universities and colleges; CHED, Commission on Higher Education; HDBSCAN, Hierarchical Density-Based Spatial Clustering of Applications with Noise.

Acknowledgements

The authors would like to acknowledge the Philippine Commission on Higher Education through Dr. Corazon Abansi and Avereen Tibalao for discussions and valuable domain insights.

Our sincere appreciation goes towards our mentors Professor Erika Legara, Dr. Damian Dailisan, and Sir Michael Dorosan. Their valuable insights served as our guiding light throughout this study.

Appendix

Giant Component Network Plots

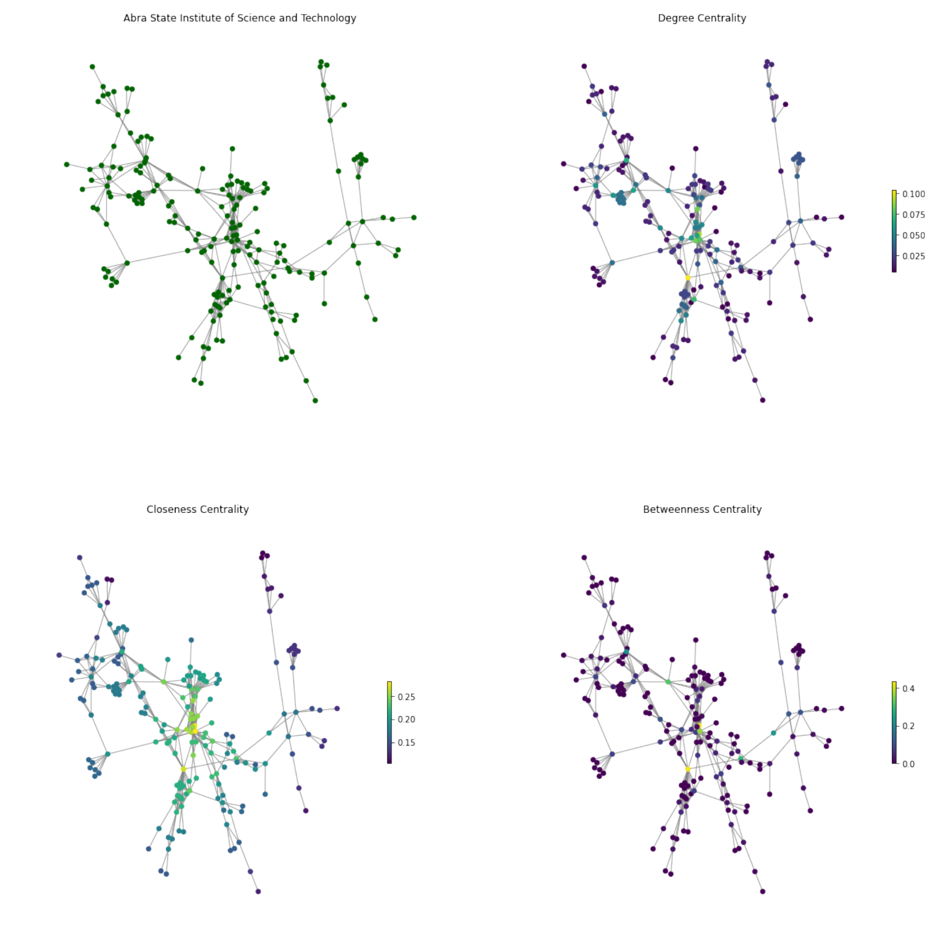

Figure 12 BSU Giant Component Co-Authorship Network Centrality Visualization.

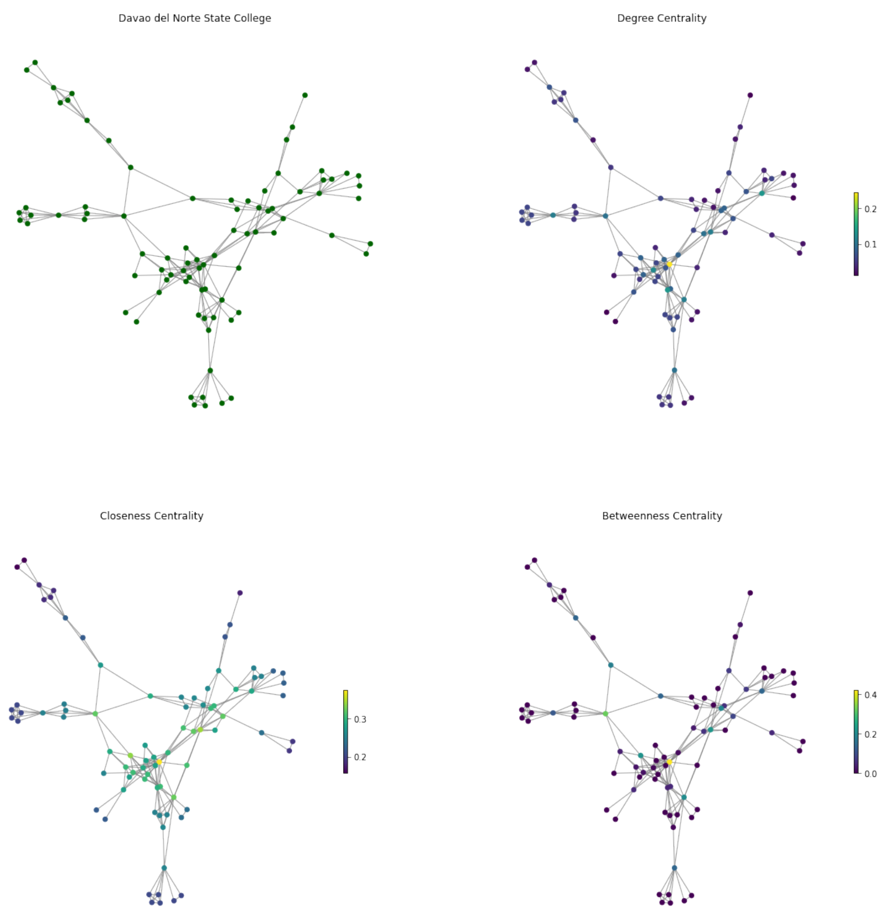

Figure 13 DNSC Giant Component Co-Authorship Network Centrality Visualization.

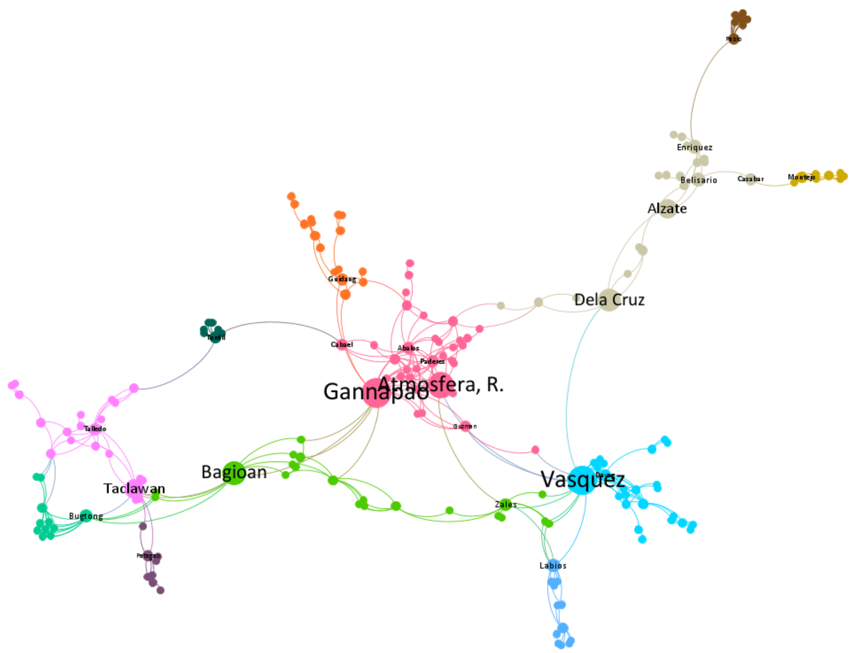

Figure 14 ASIST Giant Component Co-authorship Network Modularity Visualization. Colors indicate different communities identified and the size of the node are representative of the betweenness centrality values.

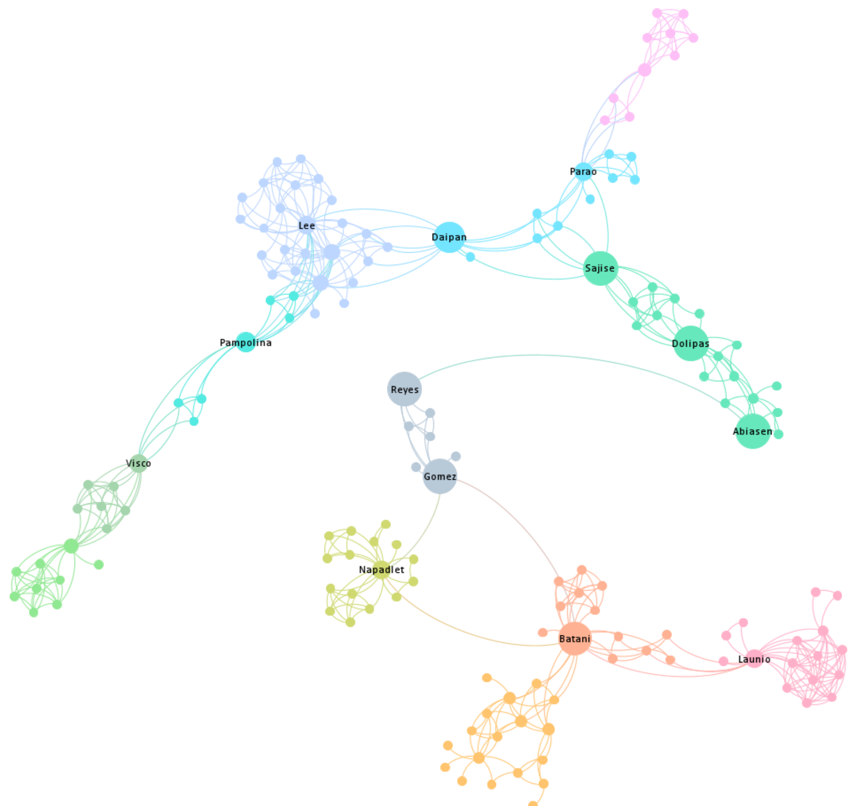

Figure 15 BSU Giant Component Co-authorship Network Modularity Visualization. Colors indicate different communities identified and the size of the node are representative of the betweenness centrality values.

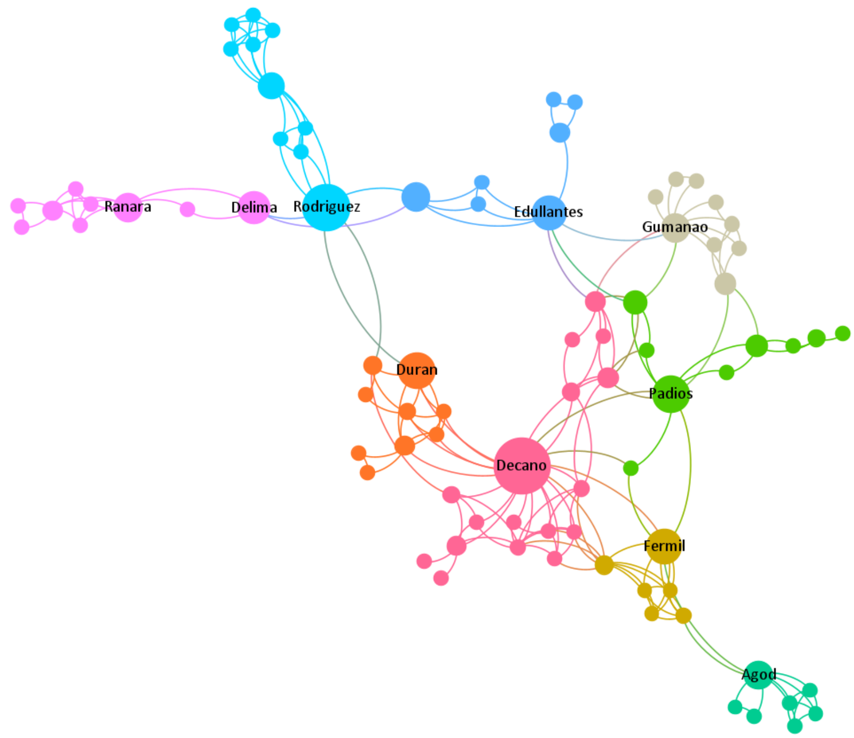

Figure 16 DNSC Giant Component Co-authorship Network Modularity Visualization. Colors indicate different communities identified and the size of the node are representative of the betweenness centrality values.

References

1 Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. Journal of informetrics, 5(4), 594-607.

2 Bansal, S., Mahendiratta, S., Kumar, S., Sarma, P., Prakash, A., & Medhi, B. (2019). Collaborative research in modern era: Need and challenges. Indian journal of pharmacology, 51(3), 137–139. https://doi.org/10.4103/ijp.IJP_394_19

3 Barabási, Albert-László. (n.d.). Network science. Retrieved May 29, 2022, from http://networksciencebook.com/

4 Cavalcanti T., Giannitsarou C., Johnson, C. (2016). Network cohesion. Econ Theory 64, 1–21

5 Chen, H., Mehra, A., Tasselli, S., & Borgatti, S. P. (2022). Network Dynamics and organizations: A review and research agenda. Journal of Management, 48(6), 1602–1660. https://doi.org/10.1177/01492063211063218

6 Commission on Higher Education. (2009). National Higher Education Research Agenda 2 (NHERA) 2009-2018. https://planipolis.iiep.unesco.org/sites/default/files/ressources/philippines_national_higher_education_research_agenda.pdf

7 Fagan, J., Eddens, K. S., Dolly, J., Vanderford, N. L., Weiss, H., & Levens, J. S. (2018). Assessing research collaboration through co-authorship network analysis. The journal of research administration, 49(1), 76.

8 Kalhor, G., Asadi Sarijalou, A., Sharifi Sadr, N., & Bahrak, B. (2022). A new insight to the analysis of co-authorship in Google Scholar. Applied Network Science, 7(1), 1-17.

9 Kumar, S. (2015). Co-authorship networks: a review of the literature. Aslib Journal of Information Management.

10 Marti, J., Bolibar, M., Lozares, C. (2016) Network cohesion and social support. Social Networks 48, 192-201

11 Moody J., Coleman J., (2015) Clustering and Cohesion in Networks: Concepts and Measures. International Encyclopedia of the Social & Behavioral Sciences (Second Edition) 906-912

12 National Economic Development Authority. (2022). 2022 National Priority Plan. https://docs.google.com/document/d/1kJZIctviXnWaaeiYaV4QqWNaEJ0uRECO

13 National Economic Development Authority. (2020). National Economic Development Authority Brochure. https://neda.gov.ph/wp-content/uploads/2020/05/NEDA-Brochure.pdf

14 Paraskevopoulos, P., Boldrini, C., Passarella, A., Conti, M. (2021) The academic wanderer: structure of collaboration network and relation with research performance. Applied Network Science 6, 31

15 Sethasathien, N., & Prasertsom, P. (2020, September). Research Topic Modeling: A Use Case for Data Analytics on Research Project Data. In 2020 1st International Conference on Big Data Analytics and Practices (IBDAP) (pp. 1-6). IEEE.

16 Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., & Laurienti, P. J. (2011). The ubiquity of small-world networks. Brain connectivity, 1(5), 367–375.

17 Vasconcellos, A. G., & Morel, C. M. (2012). Enabling policy planning and innovation management through patent information and co-authorship network analyses: a study of tuberculosis in Brazil.

18 Watts, D. J., & Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. Nature 393, 440–442

19 Williamson, B. (2019). Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education. British Journal of Educational Technology, 50(6), 2794-2809.