Collaboration Networks: University of Mosul Case Study

Basim M. Mahmood¹, Nagham A. Sultan², Karam H. Thanoon³, Dhiya Sh. Khadhim⁴

E-mail: ¹bmahmood@uomosul.edu.iq, ²naghamajeel@uomosul.edu.iq, ³karamhatim@uomosul.edu.iq, ⁴dkhadhim@uomosul.edu.iq

1,2Computer Science Department, 3Software Engineering Department, 4Mechanics Department
1, 2, 3, 4University of Mosul, Mosul, Iraq

Received on: 06/11/2019 Accepted on: 19/12/2019

ABSTRACT

Scientific research is currently considered as one of the key factors in the development of our life. It plays a significant role in managing our business, study, and work in a more flexible and convenient way. The most important aspect when it comes to scientific research is the level of collaboration among scientific researchers. This level should be maximized as much as possible in order to obtain more reliable solutions for our everyday issues. To this end, it is needed to understand the collaboration patterns among researchers and come up with convenient strategies for strengthening the scientific collaboration. The scientific collaboration among the University of Mosul researchers—which is our case in this study—has not yet been investigated or analyzed. In this work, we aim at revealing the patterns of the scientific collaboration of the scientific colleges in the University of Mosul. We generate a co-authorship network for the university; the generated network is based on the data we collected from each individual researcher. The generated co-authorship network reveals many interesting facts regarding the collaboration patterns among the university researchers.

Keywords: Collaboration Networks, Co-Authorship Networks, Complex Networks, University of Mosul.
1. Introduction

Collaboration networks or co-authorship networks are used to investigate the collaboration patterns that exist among scientific authors/researchers. In a co-authorship network two or more authors are considered to be tied if they have co-authored an article. In such a network, nodes are represented as authors and the links among them indicate that they have co-authored an article or a scientific paper (see Figure 1). Co-authorship is also used to measure the status of an individual researcher in a particular research community. In addition to the role of co-authorship networks in revealing the activities of the author, it can also be used to predict the future potential collaboration. Scientific collaboration also positively contributes in disseminating knowledge through the network.

Our approach in this work is not based on traditional statistical analysis. However, this work is based on the concepts of Complex Networks in the analysis approach.

![Collaboration Pattern](image1)

**Figure 1:** A simple example of how co-authorship networks can be generated. Consider 6 authors R1,...,R6, and 3 articles: Article 1, Article 2, and Article 3. As can be seen, a network of 6 nodes is generated including the ties among them. These ties are generated based on co-authoring in articles. This network reflects The scientific collaboration among the authors mentioned.

The field of complex networks is basically emerged from statistics and graph theory. Furthermore, using this field of study enables us to deeply investigate the relations among network actors (authors). Complex networks analysis is a technique to generate and measure network properties. For example, one can analyze the relationships among people, teams, groups, companies and other entities. Complex networks have been used to understand and study collaborations in co-authorship networks. The characteristics of a network can be described on two levels: at the entire/global network and at the individual actors. For the entire network, we can measure the density of the network (in terms of collaboration), diameter of the network, clusters (research groups), etc. At individual level, we can analyze centrality of nodes (individuals), degree (the number of papers with others), betweenness (how influential an author is in the research community), closeness (how close an author to other authors).
We consider the University of Mosul (UOM)\(^1\) as our case in this study. After generating the UOM co-authorship network, we believe that it is possible to reveal many facts about the actual situation in the research community of the UOM colleges.

2. Related Works

The field of co-authorship networks has attracted the research community due to its role in improving and strengthening the collaborations patterns among researchers. Revealing the collaboration patterns among scientific researchers using co-authorship networks has been widely used in many studies such as the distinguished study of Newman [1], he used three bibliographic datasets for three field of study; biology, mathematics, and physics. The goal of this study was to find the collaboration patterns among researchers and among the authors of the same area of study and among the three mentioned areas. One of the interesting result he obtained was that the biological scientists have strong tendency to co-author papers with authors of the same field and this tendency is significantly less than mathematicians or physicists. In another study by Newman [2], he generated three networks for three areas of research; computer science, physics, and biomedical research. He investigated and studies these networks and found the best connected scientists in terms of the strength of collaboration. Mena- Chalco et. al [3] were also generated co-authorship networks to deeply understand the structures and the dynamics among the researchers of Brazil for all the available areas of research. They evaluated information of eight major of research: biological science, earth science, agricultural sciences, humanities, engineering sciences, social science; health sciences, and linguistics, letters, and arts. The authors analyzed the relations among the researchers and among different areas of research. Moreover, they measured the level of collaboration for each of the aforementioned areas. In [4], the authors investigated the collaboration patterns and citation patterns among the authors of the Association for Computing and Machinery (ACM). The generated two collaboration networks; the first one was based on the citation of articles and the second was based on publications only without considering articles’ citation.

They used several of complex networks metrics such as degree centrality, betweenness centrality, closeness centrality, and the characteristics of community structures in evaluating authors. Then, they compared the results of the two generated networks to rank authors. The characteristics of the social networks of scientific collaborations can also be a useful tool in understanding the collaboration patterns among researchers as presented in the distinguished of Barabasi et. al in [5]. They considered the evolution of the social network of scientific collaboration in deeply understanding what is driving the collaboration patterns that exists among researchers.

3. Data Collection

Since the UOM does not have a central database for the published articles due to the unstable situation in Iraq for a long time (e.g., wars, political, and economic issues), the hardest part of this work was the data collection process. This part represented a challenging

---

\(^1\) University of Mosul is the second largest university in Iraq. It is located in the north part of Iraq. The University of Mosul contains 23 colleges in different fields and specializations.
task for us to be performed since it needed time and efforts especially the manual processes performed for the data collection. Moreover, in Iraq publishing articles can be locally or internationally; in the former, the researchers publish their articles in the Iraqi local journals. Some of these local journals are not available in the World Wide Web and cannot be accessed (local access only). In the latter, researchers publish their articles in international journals, which are accessible on-line. However, the information on the locally published articles was not available on-line and should be collected manually. To this end, we accurately designed a specific form for the purpose of data collection. This form included fields that were filled by the researchers. These fields represent information on the researchers themselves and the articles they have published. These fields were accurately chosen to be collected and further use them in generating the UOM co-authorship network. These fields are; researcher name, age, degree, position, specialization, department, college, number of articles published, articles names, number of co-authors in each articles and co-authors names, journals names for each published article, publishing year, and the affiliation of each co-author (local or international). The co-authorship network we planned to generate was based on the aforementioned information. Then, after defining the information needed for our work, we distributed the form to every single researcher in the scientific colleges in the UOM. The colleges we targeted in this work are; Agriculture and Forestry, Administration and Economics, Computer Science and Mathematics, Science, Engineering, Environment Science and Techniques, Petroleum and Mining, Education for Women and Education for Pure Sciences. The total number of the authors for the aforementioned colleges is about 2210 scientific authors including different scientific positions (Assistant Prof., Associate Prof., and Prof.). The actual number of authors we collected the data from was about 1000, which represents about 45% of the total actual number of authors. It should be mentioned that all the information collected about the papers published holds the name University of Mosul as the affiliation of the UOM author(s).

4. Statistics

According to the data collected from the researchers of the UOM scientific colleges, we extracted statistics related to the published articles. These statistics are important in this study since they lead to find indicators that may help us in understanding the collaboration patterns among the UOM researchers community. The statistics in this work is presented for each college. The statistics are presented for each college in the form of graphs, each of which has two parts A and B. Part A of each graph represents the number of researches published during the period of 1990-2018 for each department in that college as a separated and different colored lines. Part B for each graph depicts the number of internationally and locally published papers. Figures 2, 3, 4, 5, 6, 7, 8, 9, and 10 in both their parts A and B present the number of researches published during the period 1990 to 2018 as well as the locally and internationally published articles.

Based on these statistics, the thriving period of time for the UOM colleges and authors in terms of research publishing is between the year of 2010 and the first part of the year of 2014. The reason behind this peak, the ministry of higher education and scientific research in Iraq was highly supported the scientific research and projects in the Iraqi universities, institutions, and research centers by providing them with the required fund, labs, and tools needed. Moreover, during that period the awareness towards scientific research in the UOM
were very high since there were many of the researchers studied abroad and brought their experience in different field of research, which positively contributed in increasing their productivity. However, these statistics do not provide us with information on the collaboration patterns that exist in these data. For a microscopic view, we adopt the concepts of complex networks aiming at investigating the collected data and extract information on how the UOM authors and colleges connected and collaborated with each other (as we see in the next sections).

5. **Network Measurements**

The analysis of this work is based on many complex networks measurements, each of which has the ability to reveal a particular fact(s) on the UOM network. These measurements can be either used with node level or at network level. Below we list the measurements used in this work.

- **Average clustering coefficient (C):** In a network, it reflects the tendency of network nodes to cluster together [6]. In UOM network, it reveals the collaboration level among the UOM authors. The maximum level of collaboration when $C$ equals 1, while 0 means no collaboration.

- **Average path length ($l$):** For all the possible pairs of nodes in a network, it is defined as the average number of paths (steps) for all the shortest paths among the pairs [6]. In the UOM network, it shows the shortest distance among the authors. In other words, $l$ measures how far that UOM authors to collaborate.

- **Diameter ($O$):** For a network, it is the longest path among all the shortest paths [7]. In our work, it calculates the distance between the farthest authors in the network.

- **Density ($D$):** It is the proportion of the number of a network edges to the number of potential (possible edges) in that network [7]. In UOM, this measurement depicts the collaboration density among authors.

- **Communities ($cu$):** It refers to the groups of nodes in a network that are densely connected with each other. In co-authorship networks, it reveals the research groups that have papers in common (collaborative groups). In our work, we used Girvan-Newman Clustering algorithm [8] to find the number of research communities in the UOM network.

- **Betweenness Centrality ($CB$):** It shows how many times a node appears in the shortest path between network pairs [9]. It reveals the importance of a particular node in the flow of information within a network. In other words, it represents the importance of an author in a research community. In this work, $CB$ shows how influential (importance) an author in the UOM network.

- **Degree Centrality ($CD$):** It reflects the number of connections a particular node has in a network [9]. In UOM, it reflects the actual number papers an author published.
Figure 2: College Of Agricultural And Forestry

Figure 3: College Of Administration And Economics

Figure 4: College Of Computer Science And Mathematics

Figure 5: College Of Environment
6. The Generated Collaboration Networks

6.1 UOM Co-Authorship Network

As mentioned in Section 1, the UOM co-authorship network was generated based on a particular strategy that states when two authors participate in a paper, a link is created between them. This strategy is followed in almost all the works in the literature [2] [4] [3]. Moreover, this strategy was used in generating all the co-authorship networks in this work.

UOM network consists of 3444 nodes (about 1000 of them are UOM authors while the others were collaborators from outside the UOM) and 4240 edges for the whole network. According to Girvan-Newman Clustering algorithm [8], UOM network includes 210 communities representing all the research communities in the university of Mosul colleges. Figure 11 shows the UOM co-authorship network including all the scientific colleges.

Figure 12 depicts the degree distribution of UOM network nodes; the distribution follows a power-law distribution. According to [5] and [1] co-authorship networks follows this kind of distributions since there are a few authors with high degree (authored or co-authored large number of articles), while large number of them with low degree. This phenomenon represents one of the most important characteristics of co-authorship networks.

According to the distinguished work of Barabasi and Bonabeauin [10], when the degree distribution of a network follows a power-law it means this network is called Scale-Free. Therefore, UOM network is considered to be a scale-free network and has all the characteristics of this kind of networks.

We benchmark UOM network against some other co-authorship networks in the literature aiming at showing some facts on the UOM network. Table 1 presents a comparison between our network and 3 other international co-authorship networks [4], namely, ACM, Biology, and Physics networks. According to the aforementioned table, C value reflects a low collaboration level among UOM authors. This means UOM authors do not have strong tendencies to collaborate with each other. Also, l value is higher than the corresponding
values of the other networks. This means the shortest distance among UOM authors is long comparing to other networks, which needs to be shortened more. Table I reveals the low collaboration level and the weak connections among UOM authors.

**Table 1:** Comparison Between UOM Network And Other Different Collaboration Networks.

| Network  | C   | l   |
|----------|-----|-----|
| UOM      | 0.161 | 7.613 |
| ACM      | 0.60  | 4.99 |
| Biology  | 0.088 | 4.92 |
| Physics  | 0.45  | 6.19 |

As mentioned, UOM network contains 3444 authors (nodes), 1000 of these authors are affiliated to the UOM colleges, while the majority (2444) of the collaborators are from universities and institutions outside the UOM. This leads us to state that UOM authors tend to collaborate and participate 2 times more with researchers and authors out of their local academic communities.
Figure 11: Co-Authorship Network For The Scientific Colleges Of The University Of Mosul.
6.2 UOM Colleges Co-Authorship Networks

In this section, we present the visualization of 6 networks each of which represents a college at the University of Mosul. It should be mentioned that these scientific colleges are the top 6 colleges out of the ones taken in this work. In the visualization process, we follow the same strategy used in UOM network. Figures 13, 14, 15, 16, 17, and 18 demonstrate the co-authorship networks of the colleges. For each college network, the colors represent different scientific departments under a particular college. The size of nodes reflects the actual number of the co-authored papers of a particular author (node degree). Based on these networks, it can be observed there are a few number of nodes with high degree and many nodes with low degree. This observation reflects the fact that the degree distribution of each college network follows a power-law distribution, which is necessary for co-authorship networks.

Now, it is needed to perform a comparison among UOM colleges in terms of some measurements mentioned in Section 5. Table 2 presents the scientific colleges with the corresponding values of measurements. Based on of this table, it can be observed that the college of Science has the highest number of communities, which reflects the highest level of collaboration among UOM colleges. A high number of communities in a network reflects the tendency of authors to collaborate with each other in that network. However, when observing the number of communities, the number of authors in that network (nodes) should also be observed and taken into considerations. We see that the performance of a network in terms of scientific collaboration should be measured according to the ratio ($r$) of the number of scientific communities ($cu$) to the actual number of authors (number of nodes). This yields a better evaluation when investigating the scientific collaboration among authors in a particular network. In UOM network, the largest scientific communities exist in the college of Science. This college also reflects an acceptable value of $C$ and $D$, which means the authors in this network tend to cluster together and collaborate since their community is relatively dense. However, the values of $l$ and $O$ are the largest among all the other colleges and this is due to the high number of authors in this college (224).
Figure 13: The Network Of The College Of Engineering Including Its 7 Departments.

Figure 14: The Network Of The College Of Science Including Its 6 Departments.

Figure 15: The Network Of The College Of Computer Science And Mathematics Including Its 5 Departments.

Figure 16: The Network Of The College Of Agriculture And Forestry Including Its 8 Departments.

Figure 17: The Network Of The College Of Administration And Economics Including Its 8 Departments.

Figure 18: The Network Of The College Of Education For Pure Sciences Including Its 5 Departments.
6.3. UOM Best Connected Authors

As mentioned in Section 5, measurements can be used in two levels; network level and nodes level. In the previous sections, we analyze the collaboration patterns among UOM authors using network level measurements. In this section, we analyze our network based on nodes level measurements. We aim at using some centrality measurements mentioned in Section 5. Table 3 ranks the UOM authors based on the value of their betweenness centrality (CB) measurement and the frequency of collaboration with other authors. As mentioned, CB reveals how influential an author in a community since it expresses the number of times an author appears in the shortest paths of network pairs. It is clear that the college of Science has 4 authors out of the top 10 best connected authors list. The frequency of collaborations is not the main factor in determining the best-connected authors as the position of the author in community does. For example, the UOM network has authors with more than 122 published articles but their positions in the network do not make them influential. This means the position of an author plays a significant role in improving the collaboration level insofar as it improves the whole collaboration level of the network. As mentioned in Section 6.1, UOM network is a scale-free it means the concept of preferential attachment [11] can be applied and considered in this work. Increasing the level of collaboration among UOM authors can be obtained when the authors connect (collaborate) with best connected authors within the network. This case leads the clustering coefficient of the network to be increased since the number of triangles will also increase.

6.4 Collaboration Among UOM Colleges

All the aforementioned analysis was about the collaboration among authors and how they are connected in each college. In this section, we present the actual collaboration among UOM colleges that contain different and similar specializations. In fact, the analysis of this section is important insofar as it shows the scientific integration of different colleges and specialization. Figure 19 reveals how the colleges are connected and collaborated with each other in terms of co-authoring papers. In this figure, each college is represented as a node and the edge between each pair of colleges is formed if there is collaboration in co-authoring papers between them. The figure also shows the level of collaboration between each pair of colleges represented by edge weight, while nodes size reflects the actual size of the colleges in terms of the number of authors. It is clear that some pairs of colleges reflect a high level of collaboration such as the college pairs (Computer Science and Mathematics- Education for Pure Sciences, Administration and Economics-Computer Science and Mathematics, Engineering-Administration and Economics, and Science-Agriculture and Forestry).
Table 2: Comparison The Collaboration Networks Of The Colleges Of The University Of Mosul. This Table Is Ordered Based On The Number Of Communities Each College Has In Its Network (From High To Low).

| Network         | # of cu(r) | # of Nodes | # of Edges | C   | L   | D   | O |
|-----------------|------------|------------|------------|-----|-----|-----|---|
| Science         | 118(0.526) | 224        | 133        | 0.319 | 5.330 | 0.005 | 10 |
| Education       | 89(0.597)  | 149        | 69         | 0.309 | 2.575 | 0.006 | 7  |
| CSM             | 81(0.547)  | 148        | 52         | 0.234 | 4.018 | 0.007 | 9  |
| Agriculture     | 67(0.598)  | 112        | 52         | 0.28  | 2.439 | 0.008 | 6  |
| Engineering     | 66(0.550)  | 120        | 78         | 0.347 | 4.056 | 0.011 | 10 |
| Administration  | 65(0.528)  | 123        | 76         | 0.232 | 4.806 | 0.01  | 10 |
| Education W     | 27(0.964)  | 28         | 1          | na   | 1    | 0.003 | 1  |
| Environment     | 21(0.913)  | 23         | 2          | 0    | 1.333 | 0.008 | 2  |
| Petroleum       | 20(0.869)  | 23         | 3          | na   | 1    | 0.012 | 1  |

Table 3: Top 10 Best Connected Faculty Members According To Their Colleges. This Table Is Ordered According To The Value Of CB From The Highest Value To The Lowest One. The Frequency Of Collaborations Is Also Listed In The Table, Which Expresses The Number Of Articles An Authored Is Co-Authored.

| Faculty Member College | Department        | Cb    | Collaboration Frequency |
|------------------------|-------------------|-------|-------------------------|
| Science                | Biophysics        | 476227.25 | 122                    |
| Science                | Chemistry         | 393431.53 | 56                     |
| Education for Pure Sciences | Computer Science | 365928.79 | 34                     |
| Agriculture and Forestry | Animal Production | 325392.82 | 33                     |
| Engineering            | Architectural     | 285042.29 | 31                     |
| Science                | Chemistry         | 245861.18 | 38                     |
| Administration and Economics | Business Management | 242574.54 | 41                     |
| Administration and Economics | Marketing       | 209667.36 | 44                     |
| Science                | Chemistry         | 203790.99 | 46                     |
| Administration and Economics | Industrial Management | 195877.80 | 35                     |

This integration leads the value of $C$ to be 0.583 and $l$ among network pairs of colleges equals 1.58. Moreover, the highest value of $CB$ gained by the college of engineering (9.833), which reflects the strong tendency of the authors in this college to collaborate more with other specializations. Finally, the integration of a particular field of research with other fields opens the horizon to the authors of both fields to come up with new contributions that will significantly improve the quality of research.
7. Conclusion:

In this work, we investigate and analyze the collaboration patterns among the UOM authors considering 9 scientific colleges, namely, Engineering, Computer Science and Mathematics, Education for Pure Sciences, Petroleum, Environmental Science and Techniques, Science, Administration and Economics, Agriculture and Forestry, and Education for Women). Our first step in this work was to perform a statistical analysis for the actual scientific situation of these colleges in terms of the number of publications and the years published in.

We generated and visualized a co-authorship network called UOM network containing all the aforementioned colleges. The dataset used in this work was collected from the UOM authors including their publications for the period of 1990 to 2018. These publications were taken from authors with different scientific positions (Assistant Prof., Associate Pro., and Prof). We generated and visualized co-authorship networks for the 6 biggest of the mentioned colleges. We also generated a collaboration network for the colleges representing how much each two colleges collaborate in coauthoring papers.

This work can be summarized by the following:

- The UOM network reflects a weak performance in terms of scientific collaboration when benchmarking with other international networks such as the ACM co-authorship network that contains authors from worldwide.

- Based on the results obtained, the best connected authors were from the college of Science. Increasing the level of collaboration within the UOM network, authors have to be well-connected to each other. To this end, authors should be connected to the best connected ones that have highest $CB$ value, which in turn leads to increase the $C$ of the network and eventually increase the scientific productivity of the UOM University.
• The authors in the UOM network have strong tendency to collaborate with authors from outside their university. The results showed about 2 times of the collaborations are international comparing to the local collaboration.

• Improving the collaboration level is not about the quantity of researches published, it is about to whom authors connected or with whom collaborated.

• The UOM reflects a high level of collaboration among the colleges. This specific case is important because it improves the quality of research projects when involving theories inspired from different fields of research.
REFERENCES

[1] M. E. Newman, “Coauthorship networks and patterns of scientific collaboration,” Proceedings of the national academy of sciences, vol. 101, no. suppl 1, pp. 5200–5205, 2004.

[2] K. Newman, “Who is the best connected scientist? a study of scientific coauthorship networks,” in Complex networks. Springer, 2004, pp. 337–370.

[3] J. P. Mena-Chalco, L. A. Digiampietri, F. M. Lopes, and R. M. Cesar, “Brazilian bibliometric coauthorship networks,” Journal of the Association for Information Science and Technology, vol. 65, no. 7, pp. 1424–1445, 2014.

[4] P. Divakarmurthy and R. Menezes, “The effect of citations to collaboration networks,” in Complex Networks. Springer, 2013, pp. 177–185.

[5] A.-L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, “Evolution of the social network of scientific collaborations,” Physica A: Statistical mechanics and its applications, vol. 311, no. 3-4, pp. 590–614, 2002.

[6] D. J. Watts and S. H. Strogatz, “Collective dynamics of `small-world' networks,” nature, vol. 393, no. 6684, p. 440, 1998.

[7] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” Reviews of modern physics, vol. 74, no. 1, p. 47, 2002.

[8] M. Girvan and M. E. Newman, “Community structure in social and biological networks,” Proceedings of the national academy of sciences, vol. 99, no. 12, pp. 7821–7826, 2002.

[9] R. Menezes, A. Evsukoff, and M. C. González, Complex Networks. Springer, 2013.

[10] A.-L. Barabási and E. Bonabeau, “Scale-free networks,” Scientific american, vol. 288, no. 5, pp. 60–69, 2003.

[11] M. E. Newman, “Clustering and preferential attachment in growing networks,” Physical review E, vol. 64, no. 2, p. 025102, 2001.