Measuring the Impact of Co-Author Count on Citation Count of Research Publications

Wael Alharbi
M.B.B.Ch
University of Jeddah Saudi Arabia
College of Computer Science and Engineering

ABSTRACT
Citations of any work is considered as a major trait that leads to the work evaluation and investigation. Citations is one of the major measures to access the quality of the research publication. Citations can have positive or negative impact on any piece of work or publication through many different factors, such as author expertise level, publication venue, topic that is researched etc. This research aims at investigating how co-author count impact the citations of the research publications. There will be a correlation analysis between co-author count and citation of research publications. In this paper, Citation Network Dataset is used. The data set is designed for research purpose. The citation data is extracted from DBLP, ACM, MAG (Microsoft Academic Graph), and other sources. The first version contains 629,814 papers and 632,752 citations. To test the impact of co-author count on citation count of a research publications, two methods are illustrated: (i) Pearson's Correlation Coefficient (PCC), and (ii) Multiple Regression (MR). To test the impact of co-author count on citation count of a research publications, two methods are illustrated: (i) Pearson's Correlation Coefficient Calculation (PCC), and (ii) Multiple Regression (MR). To test the impact of co-author count on citation count of research publications, Pearson's correlation coefficient (ra) between the two variables Number of Authors (NA) and Citation Count (CC) is calculated. Pearson's correlation coefficient between the Citation Count (CC) and the most effective variables to compare between the impact of the number of authors and the impact of the other factors is calculated such as: (i) rc between the two variables Number of Countries (NC) and Citation Count (CC). (ii) rv between Venue Category (VC) and Citation Count (CC). (iii) ry between Year_From (YF) and Citation Count (CC). (iv) rp between the two variables Publisher (P) and Citation Count (CC). (v) rr between the two variables Number_of_references (R) and Citation Count (CC). (vi) rs between the two variables Paper_size (S) and Citation Count (CC). Empirical evidence shows that co-authored publications achieve higher visibility and impact. In order to predict the number of citations from the previous mentioned factors (NA, NC, VC, YF, P, R, S), we use Multiple Linear Regression (MLR). The goal of multiple linear regression (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable. The higher R-square, the tight relationship exists between dependent variables and independent variables. It is observed that the R-square decreases in case of removing NA which means that the NA is the most influential factor.

Keywords
Citation analysis, Research collaboration, co-authorship, Indexing, Publication venues, Academic libraries, Entropy

1. INTRODUCTION
Citation analysis is an important measure for the assessment of quality and impact of academic entities (authors, papers and publication venues) used for ranking of research articles, authors and publication venues. It is a common observation that high-level publication venues, with few exceptions (Nature, Science and PLOS ONE), are usually topic specific. The application of citation data evaluation has attracted a lot of attention in recent years [1]. The idea is to obtain a quantitative measure of the importance of an article using its citation rate. By grouping articles according to researchers (i.e., the authors of the articles), research groups, research institutions, and countries it is possible to conduct evaluations of individuals and groups as well as articles.

There are many criticisms regarding the use of citation data for research evaluation. However, it is undeniable that citation rate provides the most appropriate statistical indicator measuring an aspect of research importance (the degree of impact or utilization of articles) among those presently available. Citation data can be used in research evaluation provided that it is done carefully and that its limitations are considered [1]. It should be noted, of course, that research should be evaluated from various aspects, and citation rates provide valuable data as one of these aspects. The measures based on citations are not objective indicators themselves but complementary to subjective peer review. Even if it is generally accepted that the citation count of an article is an effective measure of its importance, an individual article's count does not always agree with the assessment of the article [2].

The most important thing in a publication is its citation in order to rank it. The measuring impact of a research paper tells us about the importance of the research which is almost has the main author and other have co-authors who hold the same idea of the topic researched. Many studies have been conducted on various factors that may influence the citation rate of an article. This research aims at investigating how co-author count impacts the citations of the research publications. There will be a correlation analysis between co-author count and citation of research publications.

The rest of paper is organized as follows: Section 2 details the literature review of different studies concerned with citation analysis. In section 3, the used dataset, the used methodology and equations are detailed. Section 4 introduces the evaluation results and discussion. Our conclusion is discussed in Section 5.

2. RELATED WORK
Numerous studies have been conducted to examine the amazing rise in research collaboration during the last several decades. The majority of these are concerned with the study of the factors that influence scientific collaboration, in an area of research headed by [5][6]. There are other significant but

fewer research aiming to establish a direct functional link between the number of citations obtained and certain aspects of the authors list for a scientific publication as described in [7][8][9].

The decision to cooperate, particularly with persons with diverse competencies, cultures, and experiences, is motivated, first and foremost, by the complexity and interdisciplinary needed by specific research topics [10]. But there can also be factors to consider that are strictly 'utilitarian,' among the predictors of collaboration. This implies that collaboration is being sought to increase the probability that the manuscript is issued [11], accepted by high-ranking journals [12], or received citations [13][14]. Formal assistance in the form of co-authorship from a well-known and recognized scientist may greatly improve the manuscript's credibility. This occurrence, called the 'Matthew effect' was examined in Robert K. Merton's first study which shows the 'impact of the cumulative activity,' i.e., those with other citations would be mentioned more frequently as scientists who are comparable in the quality of publishing [15].

Social variables such as the author's professional status influence citation selections. Similar factors might also explain why more prominent colleges have more partnerships than others [16], and why more sophisticated nations play a major role in international cooperation networks [16][17].

The phenomena should be taken into account by any main investigator who is tasked with forming a cooperation team and, as a result, co-authorship of their work. Every co-author has a network of connections into which they are more or less inextricably linked, and which will almost certainly result in citations. As a result, more co-authors imply more social networks, and therefore a higher likelihood of co-authored works being cited.

Apart from the large number of studies mentioned in the previous section, a variety of other empirical studies have shown that co-authorships achieve higher visibilities than the average, both in terms of journal significance [18], and in terms of citations [18][19][20] despite rare exceptions in particular areas of resee [21][22][23] have identified a possibility that the citation has increased 7 %, by an increase of one author, using a sample of articles from 17 demographic journals indexed over 1990-1992 inside the Social Science Citation Index. After many years later, the same authors found that the probability of referenced results was increased by 5 % by an increase of one author based on a large sample [24]. In reviewing the publications of leading U.S. research Universities from 1981–1999, [25] suggested that the production and citations were increasing with team size (the number of co-authors) and that influence was increasing with intergovernmental cooperation (measured by quotations).

They think that increasing the number of teams involves an increment in the division of labor, and so infer that with the division of scientific work productivity improves. Research in [26] indicate that co-authorship impacts the potential effect of a paper on the community of reference, as refers to numerous citations obtained, by the scientific literature surveying managers and organizations.

This is confirmed by an earlier analysis of the UK and Irish Accounting and Finance category published by [27], in 1998-1999. Research in [28] analyzes the essential importance of 'diversity in the discipline field within the co-operative team, based on a sample of work in the top of natural journals of science, and identifies the dominance of 'intelligence' over 'social' capital in quotation behavior. Researchers have demonstrated how the cooperation impacts on citations tend to be reduced by analyzes of minor changes in the composition of the networks of co-authors and the papers themselves [29] [30].

Finally, we highlight that the relationship between author numbers and citations received may be due in part to the natural rise in self-citation when works are by more authors' numbers [31] and possibly from different universities [32]. Self-citation, according to [7], contributes to but does not entirely explain, the relationship between impact and cooperation. This appears to be a result of the higher epistemic value associated with collaborative study, rather than a "mechanical" product [33][34].Dataset and indicators.

This section discusses the used dataset and the used methodology.

2.1 Dataset

In this paper, Citation Network Dataset is used from the following link https://www.aminer.cn/citation. The data set is designed for research purpose. The citation data is extracted from DBLP, ACM, MAG (Microsoft Academic Graph), and other sources. The first version contains 629,814 papers and 632,752 citations. Each paper is associated with abstract, authors, year, venue, and title.

The data set can be used for clustering with network and side information, studying influence in the citation network, finding the most influential papers, topic modeling analysis, etc. We use DBLP-Citation-network V13: 5,354,309 papers and 48,227,950 citation relationships (2021-05-14). In each text file, each line represents a paper, which is in JSON schema. The data schema is as shown in Table 1. To make it easier to handle, we convert the dataset into excel sheet.

Table 1. Data Schema (V13)

Field Name	Field Type	Description	Example
Id	string	paper ID	013ea675-bb58-42f8-a423-f5534546b2b1
Id	string	paper ID	43e17f5b20f7dfbc07e8ac6e
title	string	paper title	Data mining: concepts and techniques
authors.name	string	author name	Jiawei Han
author.org	string	author affiliation	Department of Computer Science, University of Illinois at Urbana-Champaign
author.id	string	author ID	53f42f36dabfaedce54dcd0c

venue.id	string	paper venue ID	53e17f5b20f7dfbc07e8ac6e
venue.raw	string	paper venue name	Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial
year	int	published year	2000
keywords	list of strings	Keywords	["data mining", "structured data", "world wide web", "social network", "relational data"]
fos.name	string	paper fields of study	Web mining
fos.w	float	fields of study weight	0.659690857
references	list of strings	paper references	["4909282", "16018031", "16159250", "19838944", ...]
n_citation	int	citation number	40829
page_start	string	page start	11
page_end	string	page end	18
doc_type	string	paper type: journal, book title...	book
lang	string	detected language	en
publisher	string	Publisher	Elsevier
volume	string	Volume	10
issue	string	Issue	29
issn	string	Issn	0020-7136
isbn	string	Isbn	1-55860-489-8
doi	string	Doi	10.4114/ia.v10i29.873
pdf	string	pdf URL	//static.aminer.org/upload/pdf/1254/370/239/53e9ab9eb7602d970354a97e.pdf
url	list	external links	["http://dx.doi.org/10.4114/ia.v10i29.873", "http://polar.lsi.uned.es/revista/index.php/ia/article/view/479"]
abstract	string	abstract	Our ability to generate...
indexed_abstract	dict	indexed abstract	{"IndexLength": 164, "InvertedIndex": {"Our": [0], "ability": [1], "to": [2, 7, ...]}}

2.2 Methodology of Impact Calculation

To test the impact of co-author count on citation count of a research publications, two methods are illustrated: (i) Pearson's Correlation Coefficient (PCC), and (ii) Multiple Regression (MR).

2.2.1 Pearson's Correlation Coefficient (PCC)

Correlation coefficients are used to measure how strong a relationship is between two variables. There are several types of correlation coefficient, but the most popular is Pearson's. Pearson's correlation (also called Pearson's r) is a correlation coefficient commonly used in linear regression. If you're starting out in statistics, you'll probably learn about Pearson's R first. In fact, when anyone refers to the correlation coefficient, they are usually talking about Pearson's.A correlation coefficient of 1 means that for every positive increase in one variable, there is a positive increase of a fixed proportion in the other. For example, shoe sizes go up in (almost) perfect correlation with foot length. A correlation coefficient of -1 means that for every positive increase in one variable, there is a negative decrease of a fixed proportion in the other.

To test the impact of co-author count on citation count of research publications, we calculate Pearson's correlation coefficient (ra) between the most effective variables and Citation Count (CC) as following: (i) we calculate ra between the two variables Number of Authors (NA) and Citation Count (CC) as shown in equation (1). The Number of Authors variable is derived from the **authors.name**.

$$ra = \frac{n(\Sigma NACC) - (\Sigma NA)(\Sigma CC)}{\sqrt{[n\,\Sigma(NA)^2 - (\Sigma NA)^2][n\,\Sigma(CC)^2 - (\Sigma CC)^2]}} \quad (1)$$

(ii) We calculate Pearson's correlation coefficient (**rc**) between the two variables Number of Countries (NC) and Citation Count (CC) as shown in equation (2). The Number of Countries variable is derived from the **author.org**.

$$rc = \frac{n(\Sigma NCCC) - (\Sigma NC)(\Sigma CC)}{\sqrt{[n\,\Sigma(NC)^2 - (\Sigma NC)^2][n\,\Sigma(CC)^2 - (\Sigma CC)^2]}} (2)$$

(iii) We calculate Pearson's correlation coefficient (**rv)** between the two variables Venue Category (VC) and Citation Count (CC) as shown in equation (3). The Venue Category is derived from the **venue.type**.

$$rv = \frac{n(\Sigma VCCC) - (\Sigma VC)(\Sigma CC)}{\sqrt{[n\,\Sigma(VC)^2 - (\Sigma VC)^2][n\,\Sigma(CC)^2 - (\Sigma CC)^2]}} (3)$$

(iv) We calculate Pearson's correlation coefficient (**ry**) between the two variables Year_From (YF) and Citation Count (CC) as shown in equation (4). The Year_From is derived from the year (the published year).

$$ry = \frac{n(\Sigma YFCC) - (\Sigma YF)(\Sigma CC)}{\sqrt{[n\,\Sigma(YF)^2 - (\Sigma YF)^2][n\,\Sigma(CC)^2 - (\Sigma CC)^2]}} (4)$$

(v) We calculate Pearson's correlation coefficient (rp) between the two variables Publisher (P) and Citation Count (CC) as shown in equation (5).

$$rp = \frac{n(\Sigma PCC) - (\Sigma P)(\Sigma CC)}{\sqrt{[n\,\Sigma(P)^2 - (\Sigma P)^2][n\,\Sigma(CC)^2 - (\Sigma CC)^2]}} (5)$$

(vi) We calculate Pearson's correlation coefficient (rr) between the two variables Number_of_references (R) and Citation Count (CC) as shown in equation (6).

$$rr = \frac{n(\Sigma RCC) - (\Sigma R)(\Sigma CC)}{\sqrt{[n\,\Sigma(R)^2 - (\Sigma R)^2][n\,\Sigma(CC)^2 - (\Sigma CC)^2]}} (6)$$

(vii) We calculate Pearson's correlation coefficient (rs) between the two variables Paper_size (S) and Citation Count (CC) as shown in equation (7).

$$rs = \frac{n(\Sigma SCC) - (\Sigma S)(\Sigma CC)}{\sqrt{[n\,\Sigma(S)^2 - (\Sigma S)^2][n\,\Sigma(CC)^2 - (\Sigma CC)^2]}} (7)$$

Where, ra is Pearson's correlation coefficient between the two variables NA and CC, rc is Pearson's correlation coefficient between the two variables NC and CC, rv is Pearson's correlation coefficient between the two variables VC and CC, ry is Pearson's correlation coefficient between the two variables YF and CC. NA is the number of authors. NC is the number of Countries. VC is the Venue Category. CC is the number of citations. YF is the year of publication, n is the number of sample points. P is the publisher. R is the number of references. S is the paper size. The Venue Category can be 1, 2 or 3 (Quarter) depending on the venue where the paper has been published in.

2.2.2 Multiple Regression (MR)

In the second step, we used multiple regression analysis to investigate the extent to which the citation rates of articles are influenced by the potential factors introduced in section 3.1. Multiple regression analysis using four factors as explanatory variables showed that the highly significant explanatory variables were as follows (in decreasing order of partial correlation coefficient): (i) Number of Authors (NA), (ii) Number of Countries (NC), (iii) Venue Category (VC), (iv) Years_From (YF), (v) Publisher (P), (vi) References (R), (vii) Size (S). Formula and Calculation of Multiple Linear Regression represented in (5).

$$NCC = \beta 0 + \beta 1(NA) + \beta 2\,(NC) + \beta 3\,(VC) + \beta 4(YF) + \beta 5\,(P) + \beta 6\,(R) + \beta 7(S) + \epsilon \quad (5)$$

Where:

NCC New Citation Count (The predicted value of Citation Count)

B0 The NCC-intercept (constant value)

β1 The change in NCC each 1 increment change in NA

β2 The change in NCC each 1 increment change in NC

β3 The change in NCC each 1 increment change in VC

β4 The change in NCC each 1 increment change in YF

β5 The change in NCC each 1 increment change in P

β6 The change in NCC each 1 increment change in R

β7 The change in NCC each 1 increment change in S

€ The model error term

The most common attributes that considered as factors potentially affecting the citation frequency of the sample articles are summarized as shown in Table 2 are:

Table 2. Attributes affecting citation frequency

	Attribute	**Description**
1	No of Authors (NA)	Number of authors of the article
2	Number of Countries (NC)	Number of countries where authors from
3	Venue Category (VC)	The category of the venue where the paper has been published
4	Years_From (YF)	Active years for the article from first publication until 2021
5	Publisher Value (P)	The value of the Publisher
6	No of References (R)	No of references used in the paper
7	Paper Size(S)	The size of the paper

3. IMPLICATION AND EVALUATION

This section presents the results of the methodology presented in section 4, which was carried out to investigate how co-author count impact the citations of the research publications.

3.1 Calculating the Pearson's Correlation Coefficient (PCC)

In order to investigate the impact of co-author count on citation count of research publications, we calculate Pearson's correlation coefficient (ra) between the two variables Number of Authors (NA) and Citation Count (CC) as shown in Figure 3.

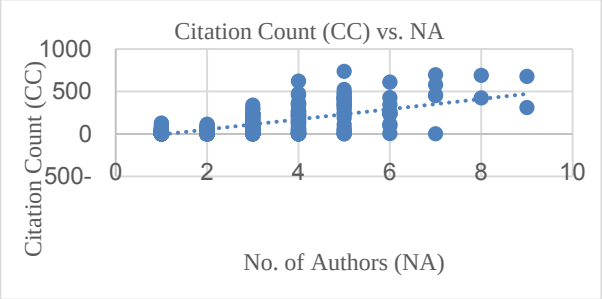

Figure 1. No. of Citation Count VS. No. of authors (NA)

From Figure 1, it is shown that there is a strong relationship between Number of Authors (NA) and Citation Count (CC). The ra=+0.7115 (strong positive correlation) which means that the both variables move in tandem—that is, in the same direction. When NA increases, CC increases. When NA decreases, CC decreases. We also calculate Pearson's correlation coefficient between the Citation Count (CC) and the most effective variables to compare between the impact of the number of authors and the impact of the other factors such as: (i) rc between Number of Countries (NC) and Citation Count (CC) as shown in Figure 2.

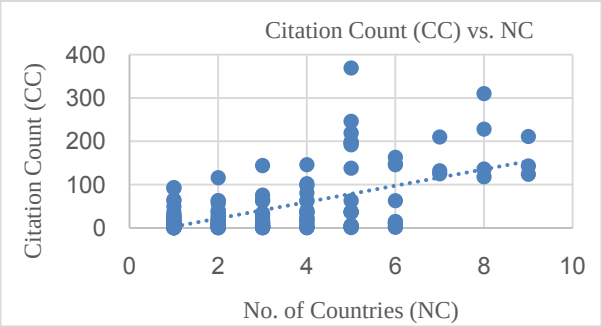

Figure 2. No. of Citation Count VS. No. of Countries (NC)

From Figure 2, it is shown that there is a positive relationship between Number of Countries (NC) and Citation Count (CC). The rc=+0.6102 (strong positive correlation) which means that the both variables move in tandem—that is, in the same direction. When NC increases, CC increases. When NC decreases, CC decreases. (ii) rv between Venue Category (VC) and Citation Count (CC) as shown in Figure 3.

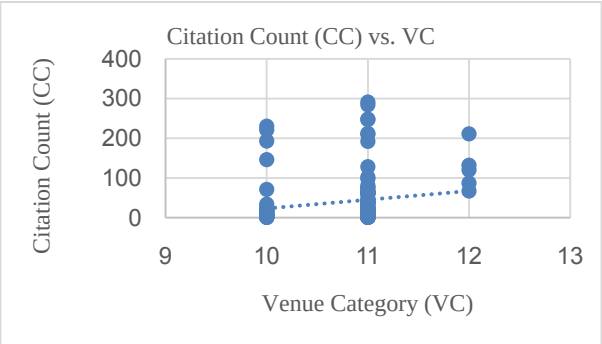

Figure 3. No. of Citation Count VS. No. of authors (NA)

From Figure 3, it is shown that there is a positive relationship between Venue Category (VC) and Citation Count (CC). The rv=+0.1711 (weak positive correlation) which means that the both variables move in tandem—that is, in the same direction. When VC increases, CC increases. When VC decreases, CC decreases. (iii) ry between Year_From (YF) and Citation Count (CC) as shown in Figure 4.

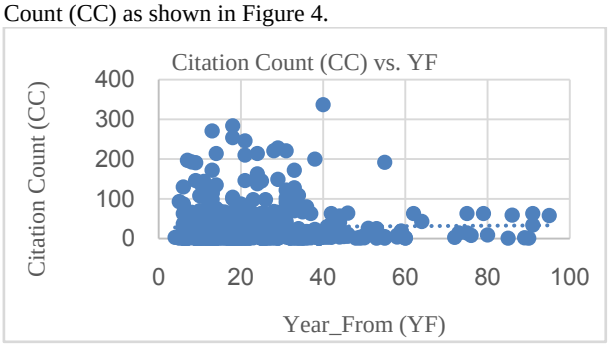

Figure 4. No. of Citation Count VS. Year_From (YF)

From Figure 4, it is shown that there is a positive relationship between Year_From (YF) and Citation Count (CC). The ry=+0.01805 (weak positive correlation) which means that the both variables move in tandem—that is, in the same direction. (iv) rp between Publisher (P) and Citation Count (CC) as shown in Figure 5.

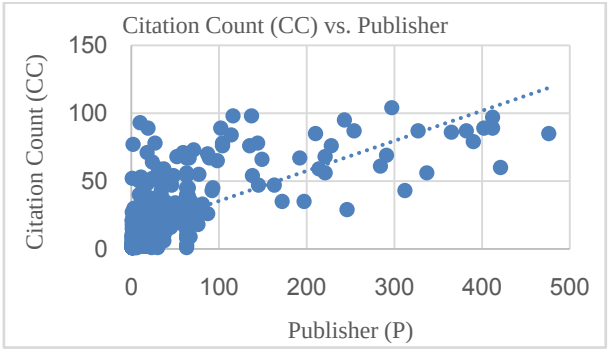

Figure 5. No. of Citation Count VS. Publisher (P)

From Figure 5, it is shown that there is a positive relationship between Publisher (P) and Citation Count (CC). The rp=+0.7018 (strong positive correlation) which means that the both variables move in tandem—that is, in the same direction. (v) rr between No_of _References (RR) and Citation Count (CC) as shown in Figure 6.

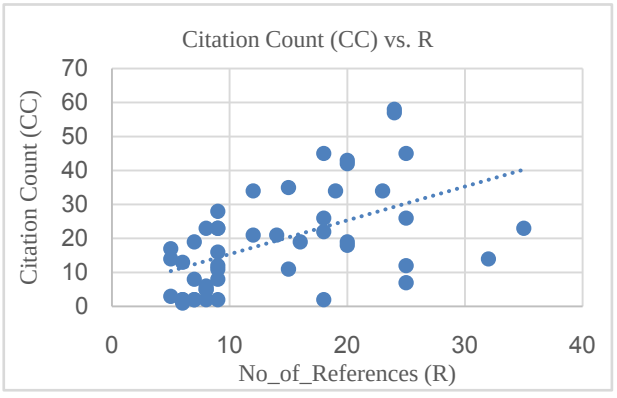

Figure 6. No. of Citation Count VS. No_of_References (R)

From Figure 6, it is shown that there is a positive relationship between No_of_References (R) and Citation Count (CC). The rr=+0.5196 (strong positive correlation) which means that the both variables move in tandem—that is, in the same direction. (vi) rs between Paper_Size (S) and Citation Count (CC) as shown in Figure 7.

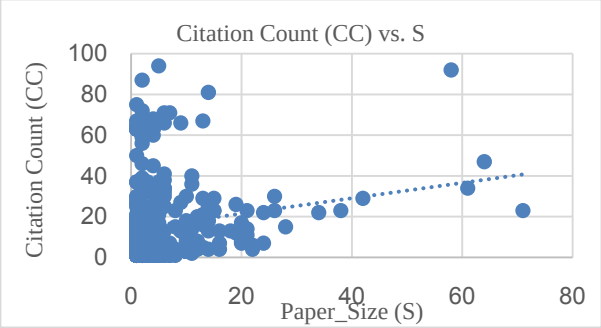

Figure 7. No. of Citation Count VS. Paper_Size (S)

From Figure 7, it is shown that there is a positive relationship between Paper_Size (S) and Citation Count (CC). The rr=+0.1598 (weak positive correlation) which means that the both variables move in tandem—that is, in the same direction. The whole calculated PCC (r) values are shown in Table 3.

Table 3. Impact of selected factors (NA, NC, VC, YF, P, R and S) on CC

	Pearson's Correlation Coefficient (PCC)	Meaning	Value
1	ra	Pearson's correlation coefficient (ra) between the two variables Number of Authors (NA) and Citation Count (CC)	+0.7115
2	rc	Pearson's correlation coefficient (rc) between the two variables Number of Countries (NC) and Citation Count (CC)	+0.6102
3	rv	Pearson's correlation coefficient (rv) between the two variables Venue Category (VC) and Citation Count (CC)	+0.1711
4	ry	Pearson's correlation coefficient (ry) between the two variables Year_From (YF) and Citation Count (CC)	+0.01805
5	rp	Pearson's correlation coefficient (rp) between the two variables Publisher (P) and Citation Count (CC)	+0.70181
6	rr	Pearson's correlation coefficient (rr) between the two variables No_of_References (R) and Citation Count (CC)	+0.5196
7	rs	Pearson's correlation coefficient (rs) between the two variables Paper_Size (S) and Citation Count (CC)	+0.1598

From Table 3, it is shown that the NA is the most influential factor (the relation between NA and CC is the most powerful relation).

3.2 Calculating the Multiple Regression (MR)

In order to predict the number of citations from the previous mentioned factors (NA, NC, VC, YF, P, R, and Size), we use Multiple Linear Regression (MLR). The goal of multiple linear regression (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable. The higher R-square, the tight relationship exists between dependent variables and independent variables. In case of using all the previous factors (NA, NC, VC, YF, P, R, S), the R-square = 0.68. The equation from the data analysis should be NCC = -309.7+ 76.2*NA + 2.96*NC + 19.06*VC - 0.06*YF + 0.21*P - 0.99*S. The Normal Probability Plot is shown in Figure 8.

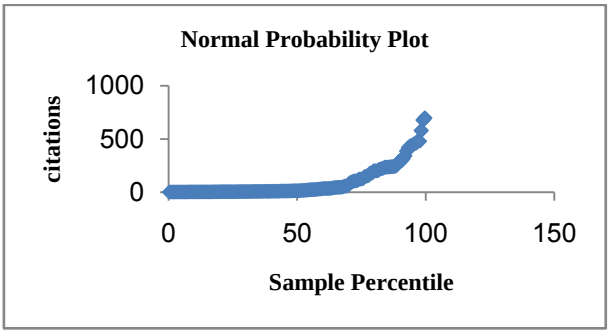

Figure 8. Normal Probability Plot (NA,NC,VC,YF,P,R,S)

From Figure 8, it is shown that the residuals are normally distributed. In case of removing one factor (NA) and using only the three factors (NC, VC, YF,P,R, and S), the R-square = 0.039. The equation from the data analysis should be NCC = -139.2+ 3.2*NC + 18.6*VC + 0. 59*YF+ 0.7*P - 1.9*S. The Normal Probability Plot is shown in Figure 9.

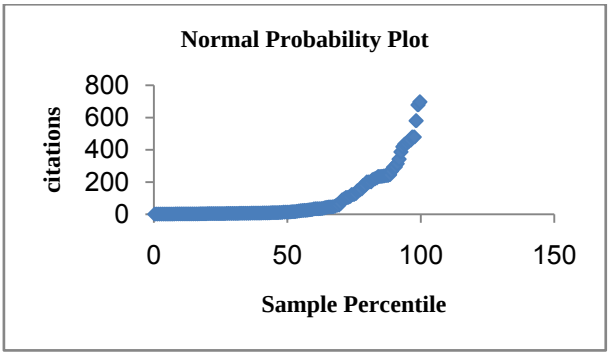

Figure 9. Normal Probability Plot (NC, VC, YF, P, R, S)

It is observed that the R-square decreases in case of removing NA as shown in Table 4 which means that the NA is the most influential factor (the relation between NA and CC is the most powerful relation).

	Used Variables	R-square
1	NA, NC, VC, YF, P, R, S	0.682234001
2	NC, VC, YF, P, R, S	0.03927636

4. CONCLUSION AND FUTURE WORK

The ultimate goal of this study was to investigate how co-author count impact the citations of the research publications. There is a correlation analysis between co-author count and citation of research publications and we saw how co-author count impact getting citations for publications. In future work, we aim to test the relation between the number of co-authors and the number of citations using machine learning and deep learning techniques. The relation between many various factors will also be tested.

5. ACKNOWLEDGMENTS

Our full thanks to all people who have contributed towards this research.

6. REFERENCES

[1] Moed, H.F. 2006. Citation analysis in research evaluation (Vol. 9). Springer Science and Business Media. .

[2] Aksnes, D.W. 2006. Citation rates and perceptions of scientific contribution. Journal of the American Society for Information Science and Technology, 57(2), pp.169-185.

[3] Pal, M. and Bharati, P. 2019. Introduction to correlation and linear regression analysis. In Applications of Regression Techniques (pp. 1-18). Springer, Singapore.

[4] Moed, H.F. 2006. Citation analysis in research evaluation (Vol. 9). Springer Science & Business Media.

[5] Melin, G. and Persson, O. 1996. Studying research collaboration using co-authorships. Scientometrics, 36(3), pp.363-377.

[6] Katz, J.S. and Martin, B.R. 1997. What is research collaboration?. Research policy, 26(1), pp.1-18.

[7] Larivière, V., Gingras, Y., Sugimoto, C.R. and Tsou, A. 2015. Team size matters: Collaboration and scientific impact since 1900. Journal of the Association for Information Science and Technology, 66(7), pp.1323-1332.

[8] Jones, B.F., Wuchty, S. and Uzzi, B. 2008. Multi-university research teams: Shifting impact, geography, and stratification in science. science, 322(5905), pp.1259-1262.

[9] Stewart, J.A. 1983. Achievement and ascriptive processes in the recognition of scientific articles. Social Forces, 62(1), pp.166-189.

[10] He, Z.L., Geng, X.S. and Campbell-Hunt, C. 2009. Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a New Zealand university. Research policy, 38(2), pp.306-317.

[11] Kalwij, J.M. and Smit, C. 2013. How authors can maximise the chance of manuscript acceptance and article visibility. Learned publishing, 26(1), pp.28-31.

[12] Al-Herz, W., Haider, H., Al-Bahhar, M. and Sadeq, A. 2014. Honorary authorship in biomedical journals: how common is it and why does it exist?. Journal of medical ethics, 40(5), pp.346-348.

[13] Sin, S.C.J. 2011. International coauthorship and citation impact: A bibliometric study of six LIS journals, 1980–2008. Journal of the American Society for Information Science and Technology, 62(9), pp.1770-1783.

[14] Leimu, R., Lortie, C.J., Aarssen, L., Budden, A.E., Koricheva, J. and Tregenza, T. 2008. Does it pay to have a "bigwig" as a co- author?. Frontiers in Ecology and the Environment, 6(8), pp.410-411.

[15] Merton, R.K. 1968. The Matthew effect in science: The reward and communication systems of science are considered. Science, 159(3810), pp.56-63.

[16] Piette, M.J. and Ross, K.L. 1992. An analysis of the determinants of co-authorship in economics. The Journal of Economic Education, 23(3), pp.277-283.

[17] Luukkonen, T., Persson, O. and Sivertsen, G. 1992. Understanding patterns of international scientific collaboration. Science, Technology, & Human Values, 17(1), pp.101-126.

[18] Bordons, M., Aparicio, J. and Costas, R. 2013. Heterogeneity of collaboration and its relationship with research impact in a biomedical field. Scientometrics, 96(2), pp.443-466.

[19] Hoekman, J., Frenken, K. and Tijssen, R.J. 2010. Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. Research policy, 39(5), pp.662-673.

[20] Jones, B.F., Wuchty, S. and Uzzi, B. 2008. Multi-university research teams: Shifting impact, geography, and stratification in science. science, 322(5905), pp.1259-1262.

[21] Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J. and Wilson, S. 2008. What makes an article influential? Predicting impact in social and personality psychology. Scientometrics, 76(1), pp.169-185.

[22] Didegah, F. and Thelwall, M. 2013. Determinants of research citation impact in nanoscience and nanotechnology. Journal of the American Society for Information Science and Technology, 64(5), pp.1055-1064.

[23] Van Dalen, H. and Henkens, K. 2001. What makes a scientific article influential? The case of demographers. Scientometrics, 50(3), pp.455-482.

[24] Van Dalen, H.P. and Henkens, K.N. 2005. Signals in science-On the importance of signaling in gaining attention in science. Scientometrics, 64(2), pp.209-233.

[25] Adams, J.D., Black, G.C., Clemmons, J.R. and Stephan, P.E. 2005. Scientific teams and institutional collaborations: Evidence from US universities, 1981–1999. Research policy, 34(3), pp.259-285.

[26] Acedo, F.J., Barroso, C., Casanueva, C. and Galán, J.L. 2006. Co- authorship in management and organizational studies: An empirical and network analysis. Journal of Management Studies, 43(5), pp.957-983.

[27] Beattie, V. and Goodacre, A. 2004. Publishing patterns within the UK accounting and finance academic community. The British Accounting Review, 36(1), pp.7-44.

[28] Skilton, P. 2009. Does the human capital of teams of natural science authors predict citation frequency?. Scientometrics, 78(3), pp.525-542.

[29] Hurley, L.A., Ogier, A.L. and Torvik, V.I. 2013. Deconstructing the collaborative impact: Article and author characteristics that influence citation count.

Proceedings of the American Society for Information Science and Technology, 50(1), pp.1-10.

[30] Didegah, F. and Thelwall, M. 2013. Which factors help authors produce the highest impact research? Collaboration, journal and document properties. Journal of informetrics, 7(4), pp.861-873.

[31] Leimu, R. and Koricheva, J. 2005. What determines the citation frequency of ecological papers?. Trends in ecology & evolution, 20(1), pp.28-32.

[32] Herbertz, H. 1995. Does it pay to cooperate? A bibliometric case study in molecular biology. Scientometrics, 33(1), pp.117-122.

[33] Beaver, D.D. 2004. Does collaborative research have greater epistemic authority?. Scientometrics, 60(3), pp.399-408.

[34] Wray, K.B. 2002. The epistemic significance of collaborative research. Philosophy of Science, 69(1), pp.150-168.