Curve Estimation among Citation and Centrality Measures in Article-level Citation Networks

문헌 단위 인용 네트워크 내 인용과 중심성 지수 간 관계 추정에 관한 연구

So Young Yu(유소영)*

ABSTRACT

The characteristics of citation and centrality measures in citation networks can be identified using multiple linear regression analyses. In this study, we examine the relationships between bibliometric indices and centrality measures in an article-level co-citation network to determine whether the linear model is the best fitting model and to suggest the necessity of data transformation in the analysis. 703 highly cited articles in Physics published in 2004 were sampled, and four indicators were developed as variables in this study: citation counts, degree centrality, closeness centrality, and betweenness centrality in the co-citation network. As a result, the relationship pattern between citation counts and degree centrality in a co-citation network fits a non-linear rather than linear model. Also, the relationship between degree and closeness centrality measures, or that between degree and betweenness centrality measures, can be better explained by non-linear models than by a linear model. It may be controversial, however, to choose non-linear models as the best-fitting for the relationship between closeness and betweenness centrality measures, as this result implies that data transformation may be a necessary step for inferential statistics.

초 록

이 연구에서는 인용 및 동시인용 문헌 네트워크에서의 중심성 지수를 사용한 추론 통계 적용의 첫 번째 단계로써 이들 간 관계의 선형성을 살펴보고자 하였다. 703개의 문헌 동시인용 네트워크를 활용하여 인용 빈도, 연결정도 중심성, 인접 중심성, 매개 중심성 간의 4가지 주요 관계의 패턴을 살펴본 결과, 모든 인용 및 중심성 간 관계가 선형모델보다는 비선형적 모델로 더 잘 설명될 수 있음을 통계적으로 확인되었다. 따라서 이들 간의 인과관계에 대한 다중회귀분석과 같은 추론 통계 분석의 기반이 되는 선형성을 확보하기 위해서는 논리적인 기준에 근거한 데이터 변환이나 실제값을 구간값으로 변환하는 과정이 필요하다고 할 수 있다.

Keywords: centrality, co-citation, citation network, curve estimation, linearity, non-linearity, normal distribution
중심성, 동시인용, 인용 네트워크, 곡선 추정, 선형성, 비선형성, 정규 분포

* 연세대학교 문헌정보학과 시간강사(soyoung.yu21@gmail.com)

■ 논문접수일자 : 2012년 5월 30일 ■ 최초심사일자 : 2012년 5월 30일 ■ 게재확정일자 : 2012년 6월 19일
■ 정보관리학회지, 29(2), 193-204, 2012. 〔http://dx.doi.org/10.3743/KOSIM.2012.29.2.193〕

1. Introduction

1.1 Background and Purpose

The question on what the characteristics of quality research are or how they are measured is one of the fundamental questions in scientometrics. The application of centrality measures in citation networks could be regarded as one of approaches to solve these questions.

Statistical analysis with centrality measures and bibliometric indices has been studied to examine the characteristics of centrality measures in the context of bibliometrics. The strong relationship between bibliometric indices and centrality measures has been revealed by correlation analysis or factor analysis. The strong relationship among centrality measures has also been identified by the same methods. These methods are useful for showing the strong relationships among them, especially at journal level (Bollen, Van de Sompel, Hagberg, & Chute, 2009; Bornmann, Mutz, Hug, & Daniel, 2011; Costas, van Leeuwen, & Bordons, 2010; Leydesdorff, 2009; Yu & Lee, 2008), and the results could show the usability of centrality measures as a new bibliometric indicator (Leydesdorff, 2009).

To see the characteristics of citation or centrality measures in citation networks, it is possible to apply inferential statistics, such as multiple linear regression, to follow correlation analysis. The first step of the multiple linear regression analysis is to estimate the curve, that is, the relationship pattern between one indicator (independent variable) and another indicator (dependent variable). If the relationship pattern is revealed as non-linear, the raw data should be transformed or binned to assure the linearity.

In addition to applying the linear regression analysis, the relationship patterns should be identified to use the centrality measures as new indicators properly, because the strong relationship between a bibliometric index and a centrality measure cannot assure that a node with a high value of the bibliometric index could have the high value of the centrality measure, unless the relationship is revealed to be linear. Therefore, the curve estimation process could be necessary to see the relational patterns between centrality measures and bibliometric indices, even though the strong relationship between the two is already found. Curve estimation is also necessary to clarify the strong relationships among centrality measures, because each centrality measure shows a different kind of central node, such as local/global center or brokerage. Interpretation on centrality measures in citation networks could be enriched by by-identifying the relational distribution between two different centrality measures, because the clear understanding on the relationship can enhance the understanding on the centrality measures.

There has been, however, few studies that examine the relationship patterns among bibliometric indicators or centrality measures. In addition to the necessity for finding the relationship patterns, only a few studies attempted to explain the centrality of a citation network with article-level bibliometric indices. It is necessary to determine whether the linear relationships could be found at the article level to

generalize the strong relationships among indicators and to apply the inferential statistics.

Therefore, in this study the relationships among bibliometric index and centrality measures in co-citation network at the article level were examined to determine whether the linear model is the best fitting model as a first step of applying regression.

1.2 Research Questions

There were four research questions in this study. The first research question was to identify the best fitting relationship pattern between the number of citations of a paper and the number of co-cited articles of the paper. The assumption was that a paper with high citation counts could be highly co-cited with others.

RQ 1. Is the citation count of a paper linearly related to the degree centrality of the paper in a co-citation network?

Three research questions were posed to determine the relationship patterns among centrality measures in co-citation network. The degree centrality in the network is for local center and the closeness centrality is for global center. The betweenness centrality is for brokerage role (Freeman, 1978). If they are linearly related, then a paper with high degree centrality tends to have a high closeness or betweenness centrality, and therefore multiple linear regression could be applied for modeling their relationships.

RQ 2. Is the closeness centrality of a paper linearly relate to degree centrality of the paper in a co-citation network?

RQ 3. Is the betweenness centrality of a paper linearly relate to degree centrality of the paper in a co-citation network?

RQ 4. Is the betweenness centrality of a paper linearly relate to closeness centrality of the paper in a co-citation network?

1.3 Previous Research

The studies that compare various indicators to find an effective indicator have increased (Bollen, Van de Sompel, Hagberg, & Chute, 2009; Bornmann, Mutz, Hug, & Daniel, 2011; Costas, van Leeuwen, & Bordons, 2010; Kim & Lee, 2010; Lee, 2011; Leydesdorff, 2009; Yu & Lee, 2008). The analysis methods are correlation analysis, factor analysis, or MDS. The analysis unit was mainly a journal (Bollen et al., 2009; Leydesdorff, 2009; Yu & Lee, 2008).

The previous studies regarded centrality measures as new indicators and attempted to reveal the characteristics of centrality measures. These studies, however, could not find what centrality can measure in citation network or research network, and only showed the relation between centrality measures and other indicators. The centrality measures are related to other indicators, however, sometimes the centrality measures are grouped in separate factors. Also the analysis unit has been a journal rather than an article due to the difficulty of obtaining proper data resources.

2. Methodology

2.1 Data Collection and Indicators

In this study, 703 highly cited physics articles published in 2004 were the sampled data. The articles were indexed in SCIE/SSCI/A&HCI databases (denoted as WoS database) and were identified as top 1% articles by its citation frequency. The threshold of citation frequency for physics articles published in 2004 was over 90. The 78,892 citing articles were also collected for the co-citation network of 703 articles. Co-citation network of 703 articles was made by using 703 articles and 78,892 citing articles.

There are four indicators as variables in this study: citation counts, degree centrality, closeness centrality, and betweenness centrality in co-citation network. The citation counts were collected from WoS database and three centrality measures were calculated using PAJEK 1.28. The values of degree centrality were calculated using the equation of Nieminen (Nieminen, 1974; Freeman, 1978), and the values of closeness centrality were calculated using the equation of Sabidussi (Sabidussi, 1966). The values of betweenness centrality were determined by the equation of Freeman (Freeman, 1978).

2.2 Curve Estimations

Curve estimation is a process to compare a linear model and a non-linear models to determine the best fitting curve on the regressional relationship between one independent variable and one dependent variable (Garson, 2012). DV (Y axis) and IV (X axis) were not clearly differentiated in this study, because the purpose of this test was not to find cause-effect relationship between centrality measures, but rather to check whether or not the linear model is the best-fitting model of the relationship.

There are, however, two considerations to set each axis for curve estimation.

This study set citation count (denoted as 'cites') as the X axis and degree centrality in co-citation network (denoted as 'cc_degree') as Y axis, because co-citation relation can be inferred after cites, i.e., direct citation were identified.

We also set closeness centrality (denoted as 'cc_close') or betweenness centrality (denoted as 'cc_betw') as each Y axis with cc_degree as the X axis based on the history of their conceptualization. Degree centrality was devised for examining local center and closeness or betweenness centrality measures which were expanded from the degree centrality. Therefore, closeness centrality was coined for finding a global center, and betweeness centrality was for

〈Table 1〉 Variables of the curve estimations

axis	indicators			
Y axis	cc_degree	cc_close	cc_betw	cc_betw
X axis	cites	cc_degree	cc_degree	cc_close

a brokerage center.

11 regressional models were estimated for each pair of indicators; Linear, Logarithmic, Inverse, Quadratic, Cubic, Compound, Power, S, Growth, Exponential, and Logistic models. After model estimation process, the paired t-test of residual differences method was performed to test the significance of the difference of the R Square's between linear model and selected non-linear models (Garson, 2012). All curve estimations and pair-wise t-tests were performed using SPSS 18 and the results were described in APA format (Green, S. & Salkind, 2007).

3. Results

3.1 Distributional Properties of the Indicators

If two variables are normally distributed, then it is easy to estimate that they have a linear relationship (Tabachnick & Fidell, 2007). To see the linearity, skewness and kurtosis were checked.

As the distributions of four indicators were skewed with high kurtosis, it was assumed that the relationship between indicators could not be linear, but rather non-linear. As shown in <Table 2>, the distributions of cites and cc_betw were extremely positively skewed with extreme values of kurtosis. The cc_degree was also positively skewed and closeness centrality was negatively skewed with high kurtosis.

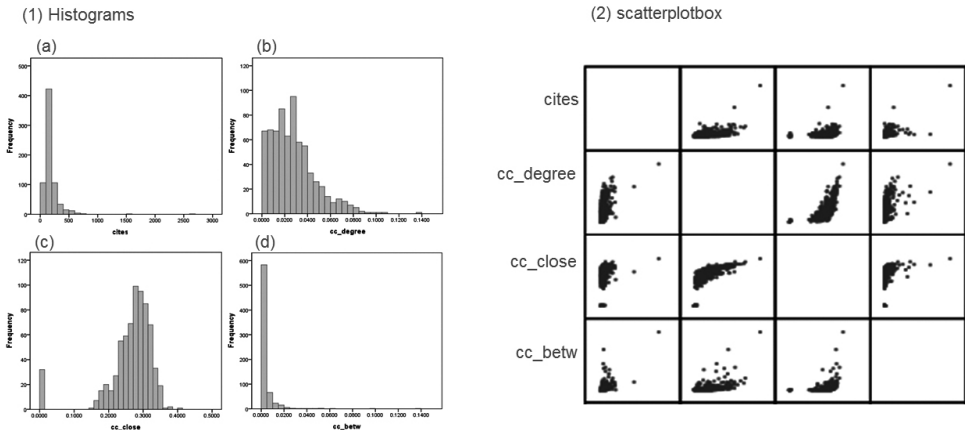
In addition to skewness and kurtosis, the number of cases with the value of "0" was examined, because the distribution shape is expected to be peaked or skewed around the value of '0', if the ratio of cases with the value of "0" is higher. For cites, there were no papers with no citation. For example, betweenness centrality had many papers with the value of "0" so that the ratio of papers with '0' for this centrality was 18% (127/703). This implies that the distribution of betweenness centrality in the network was highly skewed and peaked around '0' (see <Figure 1-(1)>) and this was verified by the high value of skewness and kurtosis in <Table 2>.

The scatterplots for the relationships among indicators showed the high probability of non-linearity (<Figure 1-(2)>). As Shown in <Figure 1-(2)>, the most of relationships between two indicators are hardly identified as linear.

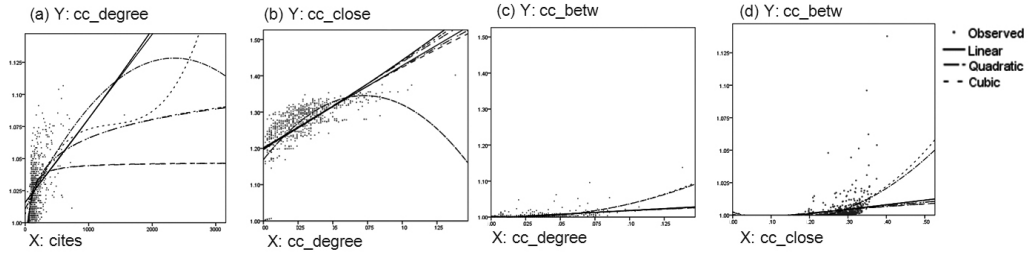
The results of curve estimations are shown in <Figure 2> and the best-fitting curve for each relationship are discussed as follows. The scatterplots for

〈Table 2〉 Descriptive statistics of variables

Indicators	Mean	SD	Range	Skewness	Kurtosis	number of paper with the values of "0"
cites	178.53	149.72	2634.00	8.48	120.53	0
cc_degree	0.03	0.02	0.14	1.16	2.09	11
cc_close	0.26	0.07	0.40	-2.25	6.08	11
cc_betw	0.00	0.01	0.14	8.65	105.60	127



〈Figure 1〉 (1) Distributions of indicators and (2) relationship patterns among indicators



〈Figure 2〉 Estimated Curves for the relationships (Co-citation Network is denoted as 'CCN')

Note: (a) the relationship between citation count and local centrality in CCN, (b) the relationship between local centrality and global centrality in CCN, (c) the relationship between local centrality and global centrality in CCN, and (d) the relationship between global centrality and brokerage centrality in CCN

the relationships between two indicators in <Figure 2> showed the high probability of non-linearity. The results of Quadratic and Cubic fits in comparison to a linear model will be only reported and discussed, because other nonlinear models were not significantly different from linear line in terms of effect size even though 10 non-linear curves were all estimated.

3.2 Curve Estimations and Identifying Linearity

3.2.1 Relationship Pattern between Citation Count and Local Centrality in Co-Citation Network

This result answers RQ 1 as it was revealed that the linear model was not the most proper to explain the relationship pattern between citation counts and degree centrality in CC network. Curve estimation

evaluated the linearity of two centrality measures. The most accountable models was Cubic curve, R^2 = .29, F (3, 699) = 93.61, p < .01, as shown in <Table 3>.

A paired-sample t-test was conducted to evaluate whether the linear relationship between citation count and degree centrality was statistically the best-fitting model. The results indicated that the relationship between citation counts and degree centrality in co-citation network was significantly non-linear (i.e., Quadratic or Cubic), $t(702)$ = 2.31 or 2.20, respectively, p < .05. It is clear that a highly cited article would not be highly co-cited with other articles, because the relationship between citation count and degree centrality in co-citation network could not be statistically linear. The nonlinear model was more effective than linear model as the effect size of the Qubic or Quadratic was more than 5% (d = .07 or .08, respectively).

3.2.2 Relationship Pattern between Local Centrality and Global Centrality in Co-Citation Network

This result answers RQ 2 as the linear model was not the most proper to explain the relationship pattern between degree centrality and closeness centrality in co-citation network. After curve estimation, all 3 estimated curves were statistically significant and the most effective model was revealed as Quadratic or Cubic curve, R^2 = .49, F (2, 700) = 335.74, respectively, p < .01, as shown in <Table 5>.

As the result of curve estimation only showing

〈Table 3〉 Curve estimations for Citation count and degree centrality

No.	Equation	R^2	F	df1	df2	Sig.	b1	b2	b3	Constant
1	Linear	0.25	231.74	1	701	0.00	0.00			1.02
2	Quadratic	0.28	134.77	2	700	0.00	0.00	0.00		1.01
3	Cubic	0.29	93.61	3	699	0.00	0.00	0.00	0.00	1.01

〈Table 4〉 Pair-wise t-test for linearity of citation counts and degree centrality

Pair	Mean	SD	t	df	Sig. (2-tailed)	Cohen's d
1-2	0.0002	0.0027	2.31	702	0.02*	0.07
1-3	0.0003	0.0036	2.20	702	0.03*	0.08

* p 〈 .05

〈Table 5〉 Curve estimations for degree centrality and closeness centrality

No.	Equation	R^2	F	df1	df2	Sig.	b1	b2	b3	Constant
4	Linear	0.40	463.19	1	701	0.00	2.26			-1.06
5	Quadratic	0.49	335.74	2	700	0.00	71.56	-33.36		-37.04
6	Cubic	0.49	335.74	2	700	0.00	71.56	-33.36	0.00	-37.04

the most effective relational line, the appropriateness of selecting non-linear curve on the relationship between degree centrality and closeness centrality was also verified by the paired sample t-test with the linear model and each of non-linear models (see <Table 6>). The relationship degree centrality and closeness centrality was significantly non-linear, as the mean difference between linear model and each non-linear model was statistically significant, $t(702) = 2.89$, respectively, $p < .05$. Also, as the effect size of each test was over 0.5, so that either the Qubic or Quadratic model is significantly different from linear model and more effective to estimate the relationship pattern between two centralities.

3.2.3 Relationship Pattern between Local Centrality and Bridge Centrality in Co-Citation Network

The relationship pattern between degree centrality and betweenness centrality in co-citation network was explained more effectively by non-linear models. While all of estimated models were statistically significant, the most accountable models was non-linear model, that is, Quadratic or Cubic curves, $R^2 = .40$, respectively, $F(2, 700) = 228.71$, or 233.99, respectively, $p < .01$.

The appropriateness of non-linear curve for the relationship between degree centrality and betweenness centrality was also verified by the significance test between the linear model and other non-linear models. The results of paired-samples t-test showed that the relationship between degree centrality and betweenness centrality was estimated better by a non-linear curve, $t(702) = 4.26$, respectively, $p < .05$ (see <Table 8>). Also, the effect size of each test was sufficient ($d = 0.17$, respectively).

3.2.4 Relationship Pattern between Global Centrality and Bridge Centrality in Co-Citation Network

This result answers RQ 4 as the linear model was revealed as not the best to explain the relationship pattern between closeness centrality and betweenness centrality. All 3 estimated models were statistically

〈Table 6〉 Pair-wise t-test for linearity of degree centrality and closeness centrality

Pair	Mean	SD	t	df	Sig. (2-tailed)	Cohen's d
4-5	0.0016	0.0150	2.89	702	0.00*	0.11
4-6	0.0016	0.0150	2.89	702	0.00*	0.11

* $p < .05$

〈Table 7〉 Curve estimations for degree centrality and betweenness centrality

N.	Equation	R^2	F	df1	df2	Sig.	b1	b2	b3	Constant
7	Linear	0.23	205.52	1	701	0.00	0.21			0.79
8	Quadratic	0.40	228.71	2	700	0.00	-11.40	5.59		6.82
9	Cubic	0.40	233.99	2	700	0.00	0.00	-5.36	3.50	2.86

〈Table 8〉 Pair-wise t-test for linearity of degree centrality and betweenness centrality

Pair	Mean	SD	t	df	Sig. (2-tailed)	cohen's d
7-8	0.0005	0.0029	4.26	702	0.00*	0.17
7-9	0.0005	0.0030	4.23	702	0.00*	0.17

* $p < .05$

〈Table 9〉 Curve estimations for closeness centrality and betweenness centrality

N	Equation	R^2	F	df1	df2	Sig.	b1	b2	b3	Constant
10	Linear	0.08	61.76	1	701	0.00	0.04			0.96
11	Quadratic	0.21	90.61	2	700	0.00	-0.81	0.36		1.46
12	Cubic	0.22	96.61	2	700	0.00	0.00	-0.35	0.20	1.15

〈Table 10〉 Pair-wise t-test for linearity of closeness centrality and betweenness centrality

Pair	Mean	SD	t	df	Sig. (2-tailed)	Cohen's d
10-11	0.0001	0.0023	1.52	702	0.13	0.04
10-12	0.0001	0.0023	1.41	702	0.16	0.04

significant and the most accountable model was Cubic curve, $R^2 = .22$, F (2, 700) = 96.61, $p < .01$, as shown in <Table 9>. Therefore, it is assumed that a node with a high value of global centrality could not always have the high value of bridge centrality in co-citation network.

The non-linearity between the centralities, however, was not statistically significant. Two paired-sample t-tests were conducted to evaluate whether the non-linear curve was significantly different from linear line (see <Table 10>). The results indicated that the relationship between the centralities can be estimated as being linear, because the non-linear models were not significantly more effective than the linear model, $t(702) = 1.52$ or 1.41, respectively, $p > .05$. Also the effect size was less than 5%, so that the difference between the linear model and each non-linear model was regarded as occurring by chance.

3.3 Application

The results consistently show that non-linear model such as Quadratic or Cubic is more suitable to estimate the relationship between two indicators in this study. This is a critical consideration when we building a prediction model for citation counts of centrality measures with other indicators, because sophisticated inferential statistics assumes linearity between variables. If the relationship is not linear, the process to build a prediction modeling could be more complicated. Therefore, data transformation or eliminating outliers were conducted to assure linearity and normality.

There are, however, no articles regarded as outliers

in an article-level citation network, because an article with an extremely high number of citation is regarded as 'research excellent' which should be included in the analysis. With this assumption, we cannot eliminate cases with extreme values, but rather transform or bin the raw values to assure linearity and to apply inferential statistics. As the identified non-linear models were quadratic or cubic, to make three or four bins with raw values could be one of the appropriate approaches, because there are three or four phases with different relationship slopes. Alternatively, the raw values may be transformed with proper functions, such as log or square root, to avoid the false-positive effect in the analysis caused by the binning.

4. Conclusion

The relationship patterns between the citation counts and centrality measures in co-citation network were examined as the first step to apply the inferential statistics in prediction modeling of citation or centrality measures. The purpose was to check linearity between two indicators, because this determines the proper statistical approach for modeling.

The statistical model with centrality measures is to identify the characteristics of the centrality measures in article-level citation networks and to check the usability of the measures as new indicators. A new indicator based on the centrality measures could be properly developed with understanding the relationship pattern between the centrality measures and traditional bibliometric indices, because the pattern could be evidence of what bibliometric traits are involved for the characteristics of the new indicators.

The results of this study revealed that the relationship between citation counts and degree centrality in co-citation network were not be best fitted by linear model. Also, the relationships among centrality measures in co-citation network could be best fitted to other non-linear models. The results could be inferred from their highly skewed and peaked distribution. Therefore, it was identified that the relationships among citation counts and centrality measures could be non-linear when raw values of each indicator were used. This leads to the issue of data transformation or binning with the assumption that linear relationship between them was hidden by noisy or extreme cases.

Due to this non-linearity, data transformation or making bins with raw data could be a feasible approach with consideration of citation or co-citation network as a scale-free network which does not show normal distribution in connection degree. Constructing bins the size of which is exponentially increased with degree in drawing degree distribution of scale-free network has been used in complex network analysis, because the distribution of citation counts can fit to a power law and co-citation network was also identified as a one of scale-free networks (Burrell, 2002; de Solla Price, 1965, 1976; Glanzel, 1992; Gupta et al., 2005; Nadarajah & Kotz, 2007; Redner, 1998, 2004; Rousseau, 1994; Simkin & Vwani, 2007). That is, when the degree distribution is not a normal distribution and noisy, researchers have already been used binning or data transformation to describe the

traits of the network (Newman, 2003).

In conclusion, this study identified that transforming raw data is required to proceed with inferential statistics with citation counts and centrality measures in citation networks as their relationship patterns were non-linear. Further studies to examine the multiple linear regression with other bibliometric indices and centrality measures could advance this study to inferential modeling to examine the centrality measure as a new indicator.

References

Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A Principal Component Analysis of 39 scientific impact measures. PLoS ONE, 4(6), e6022. http://dx.doi.org/10.1371/journal.pone.0006022

Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H. D. (2011). A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. Journal of Informetrics, 5(3), 346-359. http://dx.doi.org/10.1016/j.joi.2011.01.006

Burrell, Q. L. (2002). The nth-citation distribution and obsolescence. Scientometrics, 53, 309-323. http://dx.doi.org/10.1023/A:1014816911511

Costas, R., van Leeuwen, T. N., & Bordons, M. (2010). A bibliometric classificatory approach for the study and assessment of research performance at the individual level: The effects of age on productivity and impact. Journal of the American Society for Information Science and Technology, 61(8), 1564-1581. http://dx.doi.org/10.1002/asi.21348

de Solla Price, D. J. (1965). Networks of scientific papers. Science, 149, 510-515.

de Solla Price, D. J. (1976). A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science, 27, 292-306.

Garson, G.D. (2012). Curve Estimation. Quantitative Research in Public Administration, NC State University. Available from http://faculty.chass.ncsu.edu/garson/PA765/curve.htm

Glanzel, W. (1992). On some stopping times of citation processes. Information Processing and Management, 28, 53-60. http://dx.doi.org/10.1016/0306-4573(92)90092-E

Green, S., & Salkind, N. J. (2007). Using SPSS for Windows and Macintosh: Analyzing and Understanding Data(5th ed). New York, NY: Prentice Hall.

Gupta, H. M., Campanha, J. R., & Pesce, R.A.G. (2005). Power-law distributions for the citation index of scientific publications and scientists. Brazilian Journal of Physics, 35(4A, December), 981-986. http://dx.doi.org/10.1590/S0103-97332005000600012

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. Social Networks, 1(3), 215-239.

Kim, Pan Jun, & Lee, Jae Yun (2010). A study on journal impact measurement with Hirsch-type indices. Journal of the Korean Society for Information Management, 27(1), 269-287. http://dx.doi.org/10.3743/KOSIM.2010.27.1.269

Lee, Jae Yun (2011). A study on document citation indicators based on citation network analysis. Journal of the Korean Society for Library and Information Science, 45(2), 119-143. http://dx.doi.org/10.4275/KSLIS.2011.45.2.119

Leydesdorff, L. (2009). How are new citation-based journal indicators adding to the bibliometric toolbox? Journal of the American Society for Information Science and Technology, 60(7), 1327-1336. http://dx.doi.org/10.1002/asi.v60:7

Nadarajah, S., & Kotz, S. (2007). Models for citation behavior. Scientometrics, 72, 291-305. http://dx.doi.org/10.1007/s11192-007-1717-9

Newman, M. E. J. (2003). The structure and function of complex networks. SIAM Review, 45(2), 167-256. http://dx.doi.org/10.1137/S003614450342480

Nieminen, J. (1974). On the centrality in a graph. Scandinavian Journal of Psychology, 15(1), 332-336. http://dx.doi.org/10.1111/j.1467-9450.1974.tb00598.x

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. Eur Phys J B, 2, 131-134.

Redner, S. (2004). Citation statistics from more than a century of Physical Review. arXiv:physics/0407137.

Rousseau, R. (1994). Double exponential models for first-citation processes. Scientometrics, 30, 213-227. http://dx.doi.org/10.1007/BF02017224

Sabidussi, G. (1966). The centrality index of a graph. Psychometrika, 31(4), 581-603.

Simkin, M. V., & Vwani, P. R. (2007). A mathematical theory of citing. Journal of the American Society for Information Science and Technology, 58, 1661-1673. http://dx.doi.org/10.1002/asi.20653

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics. Boston: Pearson/Allyn & Bacon.

Yu, So Young (2012). Ascertaining the Characteristics of Centrality Measures in Citation Networks as Revealed by Their Relationship with Bibliometric Indices. (Doctoral dissertation). Yonsei University.

Yu, So Young, & Lee, Jae Yun (2008). Journal citation analysis for library services on interdisciplinary domains: A case study of department of biotechnology, Y University. Journal of the Korean Society for Information Management, 25(4), 283-308. http://dx.doi.org/10.3743/KOSIM.2008.25.4.283 [Korean]