Of intent, citation game, and scale-free networks: A heuristic argument

V. Christianto & F. Smarandache

Math Department, Univ. of New Mexico, USA

A heuristic argument was presented in favor of hypothesis that scientific communication corresponds to a process known as scale-free network. As a result, it is argued that scientific referencing through citation follows the same process, therefore it could be expected that this shall also exhibit fractality as observed in various phenomena associated with scale-free networks. This argument appears conceivable because the process of citation involves a decision-making, coined here as 'citation game.' In this regard, it is recommended to conduct citation analysis to measure the fractality of this process. While at present this heuristic argument cannot be considered as conclusive, further research is recommended to verify or refute this hypothesis.

Keywords: scientific communication, citation analysis, scale-free networks

Introduction

Nowadays, there is a vast amount of scientific publications in this world, with a strong tendency towards more specialized subjects. This phenomenon seems to support Kuhn and Polanyi's viewpoint that scientific progress is advanced via communication within various scientific societies. Furthermore, scientific communication is conducted in various forms, including: i) attending gathering (lectures, seminars, conferences etc.), ii) reading relevant text (periodicals, reports, textbooks etc.), iii) direct meeting (visiting each other), iv) sending emails; v) visiting online homepage (including arXiv.org). With the advancement and wide availability of TCP/IP-based networks, apparently method iv) and v) are growing fast in popularity among scientists prior to committing in a stronger form of participation in the other three methods of scientific communication. And it seems that this was the original intention of the proposal by Berners-Lee some decades ago.

This article was partly motivated by a recent communication with Prof. M. Pitkänen who happened to see that his TGD entry in Wikipedia has been categorized in 'delete list' [1], merely because some visitors argued that this is his own original research. Regardless of the content of TGD itself, this seemingly common practice by Wikipedia (and also other online publications) obviously raises a question whether such a 'silent voting' is actually acceptable, at least from theoretical viewpoint (i.e. science sociology). This author predicts that similar situation also happens to other scientists, who believe that their wholehearted research was 'deleted' from having a chance to be published or referred in various periodicals, because of the similar 'silent voting' happens.

It seems also worth noting that nowadays citation analysis has been widely used to measure the popularity rate of certain articles, known as 'impact factor' analysis. But it is known that popularity does not equal to the experimental verification required by a scientific theory (Popper). It is not surprising therefore that this process induces some critics, for it is quite similar to other types of mass communication, i.e. rating has replaced depthness, and popularity has replaced reality. While such popularity-voting methods are generally acceptable in other popular culture, it seems that science demands more than this. However, according to this ‘popularity’ proponents, scientists’ task does not include merely to spread the ultimate reality of Nature *per se*, but to produce *discourses*, i.e. to tell a conceivable story. There is other argument suggesting that science is merely consensus among the experts in a respective field, and therefore popularity could be a good indication of such consensus. Summarizing, it seems that we could expect that the popularity-discourse proponents will argue in favor of this phrase: ‘*in the land of the blind, the best storyteller will be the king*.’ The storyteller therefore is not required to be not (so) blind, suffice it if he could produce good stories of how wonderful the world looks like.

After discussing this citation analysis from some considerations, some implications and plausible future direction of research is discussed.

Wheeler’s game: to participate or not to participate

While surely there is other method to describe scientific citation process, for instance using imperfect information theory: ”…*scientists may trade ideas to generate citations*,” [2] apparently the present method is not conclusive for analytical purposes. Therefore in this article we use another route: the quantum mind hypothesis.

According to J.A. Wheeler, the Universe comes into reality through the participation of its observers; therefore completion of Quantum Mechanics can be viewed in this regard through integrating the role of observers. While this proposition could lead us to a paradoxical ontological question, i.e. whether there is reality without consciousness (observer), nonetheless this imposes some interesting implications. There are also some recent theories developed around this line of thought, suggesting the role of Mind in Quantum Reality. For instance, Stapp puts forward this hypothesis by arguing that there is distinction between 'Attention', 'Intention' and 'Will' in Quantum Physics [3]. This is a starting premise in the present article.

In this regard, supposed we could accept that scientific progress is determined by dissemination of journals and other kind of scientific publications, then it seems reasonable to expect that science merely consists of accumulation of decision making. This scientific decision making process could be termed as: '*to participate or not to participate'* choice (in Wheeler's sense) of other scientists' viewpoint, which usually was represented in citation. From this reasoning a new term is coined: '*citation game'*, corresponding to quantum-like decision making [4][5][6]. It seems worth therefore to conduct a citation analysis to measure and track backward this process of decision making.

A plausible numerical test: scale-free network hypothesis

Supposed we could accept that various electronic communication methods have gained popularity in recent years among scientists (method iv and v as described above), then it could be expected that the fractality property as observed in scale-free networks of electronic communication [7][8][9][10] could also play a significant role in scientific communication, provided there is no other decisive factor. In other words, it is argued here that provided freedom to publish scientific articles are preserved without restriction, then citation analysis in the respective journals will reveal fractality pattern, because it corresponds to scale-free networks, and *vice versa*. This method will enable us to conduct a precise analysis of the hypothesis as stated here.

In this article, citation analysis was recommended only in order to test a hypothesis of the scientific decision-making behind the citation process:

a. *Hypothesis*: Number of citation received by most popular scientific articles was attributed to natural phenomena (mutually exclusive events), i.e. it follows fractality property similar to other phenomena associated with scale-free networks.

b. *Null Hypothesis*: Number of citation received by most popular scientific articles doesn't follow fractality property similar to other scale-free network phenomena, because there are other factors involved in the decision-making process.

While at first glance this hypothesis offers nothing new, this is intended to provide a formal basis of such statistical data analysis where citation is the focus of attention.

Now the remaining question is how to provide a numerical test of the proposed hypothesis. As first step, it seems worth to mention here that there are some recent suggestions [11][12] that indeed we live in a fractal world (world inside world). This hypothesis subsequently implies that there are various phenomena, which exhibit fractality, from the scale of particle physics up to astrophysics scale. Accordingly, there are some recent reports suggesting that powerlaw function has a neat linkage to fractality property and also scale-invariance in Nature [13][14]. This could be a plausible basis of numerical test. Accordingly, it is hypothesized here that in scale-free network environment, the relationship between number of citation W and year Y could be expressed in a powerlaw function:

$$W = \alpha . Y^{\beta} \qquad (1)$$

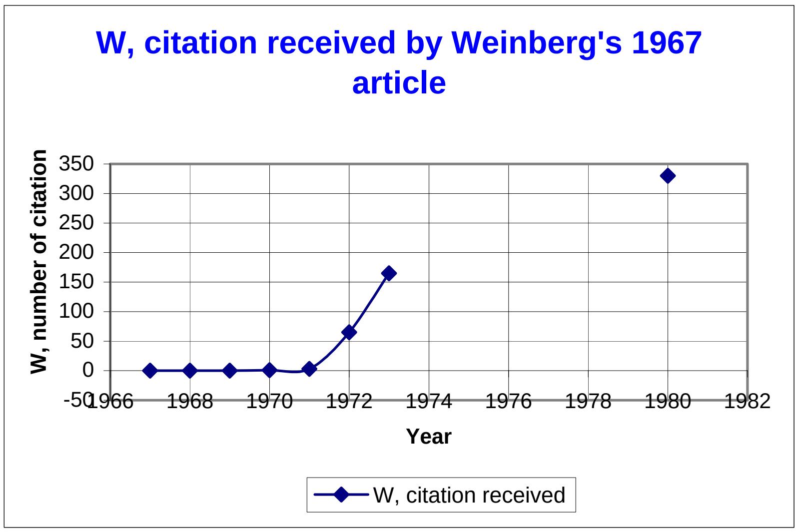
and the power coefficient β is neatly related to fractal dimension d in the form of [13][14]:

$$\beta = f(d) \tag{2}$$

Table 1. Number of citation received by Weinberg's 1967 article

Year	W, number of citation received
1967	0
1968	0
1969	0
1970	1
1971	3
1972	65
1973	165
1980	330

Graph 1. Graph plot of citation received by Weinberg's 1967



Using the similar line of thought, we could make a simple analysis of citation received by a famous article of Weinberg in 1967. The data was obtained from Weinberg's own book [15], and it is shown in Table 1 and Graph 1. This data was selected merely based on some obvious reasons: i) it has been published in a book form, so it is directly accessible; ii) this article was referred to as "the most frequently cited article on elementary particle physics of the previous half century." [15]

It shall be noted here, that because of lack of data for year 1974-1979, then we do simple linear proportional 'filling' for these years (grey data). This 'normalized' data is presented in Table 2. The subsequent curve fitting method shows that linear and logarithmic regression gives a good correlation ratio of R^2=0.939. The graph plot of these methods is shown in Graph 2.

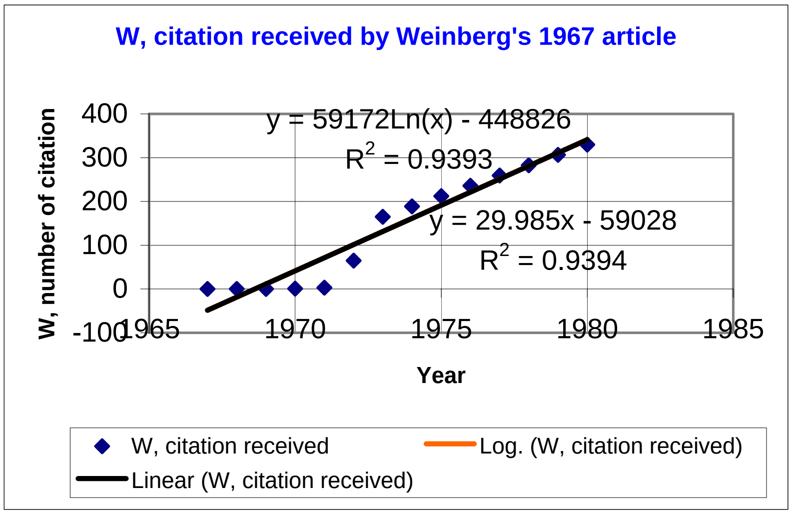
Converting the data to log-natural (ln) scale, we find that log-linear regression gives less correlation ratio R^2=0.6382. The result of this statistical regression is also shown in Table 2. The log-natural scale regression is shown in Graph 3, which obviously shows that parabolic regression at log-scale gives better curve fitting. This subsequently implies that citation received by this article apparently doesn't follow the proposed powerlaw/fractality hypothesis outlined above.

Table 2. Regression analysis of citation

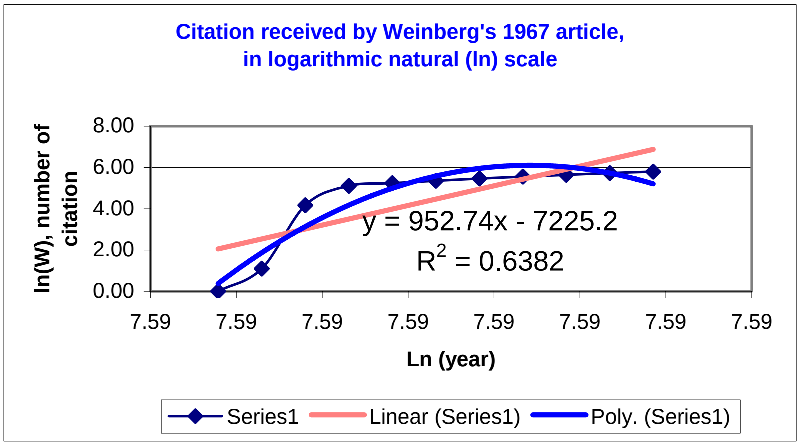
Actual		Logarith.	Linear	Lnscale	Lnscale	Loglinear
Y, Year	W	W'	W''	Ln(Y)	Ln(W)	W'''
1967	0	-49.88	-47.51			
1968	0	-19.81	-17.52			
1969	0	10.25	12.46			
1970	1	40.30	42.45	7.59	0.00	2.08
1971	3	70.33	72.43	7.59	1.10	2.57

1972	65	100.34	102.42	7.59	4.17	3.05
1973	165	130.34	132.40	7.59	5.11	3.53
1974	188.57	160.32	162.39	7.59	5.24	4.02
1975	212.14	190.29	192.38	7.59	5.36	4.50
1976	235.71	220.24	222.36	7.59	5.46	4.98
1977	259.29	250.18	252.35	7.59	5.56	5.46
1978	282.86	280.10	282.33	7.59	5.64	5.95
1979	306.43	310.01	312.32	7.59	5.72	6.43
1980	330	339.90	342.30	7.59	5.80	6.91
R^2		0.9393	0.9394			0.6382
St.dev		8.85	8.84			0.29

Graph 2. Graph plot of linear and logarithmic regression fitting



Graph 3. Graph plot of linear regression fitting in ln scale



The numerical expressions of the above regression lines are as follows:

a) Logarithmic regression at normal scale (W'):

$$W' = 29.985 x Year - 59028 \quad (3)$$

b) Linear regression at normal scale (W''):

$$W'' = 59172 x \ln(Year) - 448826 \quad (4)$$

c) Linear regression at natural logarithmic (ln) scale (W'''):

$$\ln(W''') = 952.74 x \ln(Year) - 7225.2 \quad (5)$$

Because linear regression at log-scale could be directly translated to powerlaw function (1), then we conclude from the data that citation received by Weinberg's article does not follow assumption of scale-free networks. This observation, however, could be attributed to the fact that prior to 1990 the electronic communication was not available to scientists, therefore its scale-free effects were not observed. While surely this statistical analysis is very simple, this method could be used as a preliminary citation analysis. For more extensive data, of course more advanced statistical techniques are recommended. More extensive study of the scaling properties of journals is available elsewhere [16].

Thanks to the presence of citation database in various leading labs (SLAC, CERN, for instance), it seems possible to conduct such an extensive analysis, particularly using special analysis tools [17]. In turn, it could provide a quantitative picture of how good is the 'freedom to publish' principle has been kept in the real world of scientific journals. Even a negative result could be a good sign of the presence of other factors, which could play a role in the publishing policy of the journals. This method could also be plotted spatially or per journal basis to encourage further analysis of why in some countries people could expect less restriction to publish while this perhaps does not happen in other countries. It is also known that Bose-Einstein condensate with Hausdorff dimension D_H~2 could exhibit fractality, so in the near future it could be expected that such scale-free network property could be observed in lab scale [18].

Further research is recommended to verify whether the scale-free network hypothesis of citation data as outlined here is conceivable, corresponding to the observed citation data of scientific articles, particularly in the fields of astrophysics and particle physics (in CERN or SLAC).

Acknowledgment

The writers are grateful to Profs. M. Pitkänen & C. Castro for insightful discussion and remarks. Special thanks goes to Prof. A. Rubćić for sending Ref. [13] & [14].

References

[1] Pitkänen, M., personal communication, Oct. 2004.

[2] Donangelo, R., *et al.,* arXiv:cond-mat/00021385 (2000). Also quant-ph/0404159
[3] Stapp, H., "Quantum Ontology and Mind-Matter synthesis," arXiv:quant-ph/9905053; also quant-ph/9905054 (1999).
[4] Shi, Y., arXiv:quant-ph/0008095 (2000).
[5] Grib, A.A., *et al.,* arXiv:quant-ph/0308074 (2003), quant-ph/0211068.
[6] Schweitzer, F., *et al.,* arXiv:cond-mat/0211605.
[7] Dorogovtsev, S.N., *et al.,* arXiv:cond-mat/0009090 (2000).
[9] Ebel, H., *et al.,* arXiv:cond-mat/0201476 (2002).
[10] Stiger, J., *et al.,* arXiv:cond-mat/0401283 (2004).
[11] Castro, C., *et al.,* arXiv:hep-th/0004152 (2000).
[12] Castro, C., arXiv:hep-th/0203086 (2002).
[13] Paar, V., *et al.,* "Scale invariant powerlaw and fractality of molecular weights," *Chemical Phys. Lett.* **336**, Elsevier (2001) 129-134.
[14] Paar, V., *et al.,* "Powerlaws and fractal behavior in nuclear stability, atomic weights and molecular weights," *Chaos, Solitons, Fractals* **14** (2002) 901-916.
[15] Weinberg, S., *Dreams of a final theory*. Vintage, London (1993) p. 96.
[16] Farkas, I., *et al*., "Networks in life: scaling properties and eigenvalue spectra," *Physica* A **314** (2002) 25.
[17] Indyk, P., "High dimensional computational geometry," Dissertation submitted to Stanford Univ., 70p (2000)
[18] Toral, R., *et al*., cond-mat/0304684.