O-24

論文引用ネットワークの次数分布と数理モデル*

Degree distribution and mathematical model of citation network

相馬亘[1]，治部眞里[2]
Wataru Souma[1], Mari Jibu[2]

Abstract: We investigate citation network constructed from 24,160,348 papers and 385,692,359 citations. This citation network is regarded as the directed graph. Through numerical analysis, we clarify that the distributions of incoming degree and outgoing degree follow the generalized Beta functions of the second kind. We also construct the mathematical model to explain this distribution.

１．はじめに

企業にとっては、売上高を伸ばすことよりも、生産性を高めることが重要である。生産性を高める方法にはいくつかあるが、その中でもイノベーションは、最も重要な要素である。イノベーションの定義は様々であり、広義には「人々の生活を豊かにすること」である。そのため、様々な分野でイノベーションを起こすことができる。だが、そのような分野の中でも、科学技術に対する人々の期待は大きい。そこで、本稿では、科学技術を基にしたイノベーションを、最も根本的なレベルから議論するために、論文引用の分析を考えることにする。

論文の引用といっても、その定義には 3 種類ある。これらは、直接引用、共引用(Small, 1973)、書誌結合(Kessler, 1963) である。ここで、直接引用は，単に論文が論文を引用することである。たとえば、右図に示すように、論文 C が論文 A と論文 B を引用していて、論文 D と論文 E が論文 C を引用しているような場合、直接引用は、図 1 の矢印で表される。この場合、論文 A と論文 B の被引用数は 1 で、論文 C の被引用数は 2 である。そして、論文 C の引用数は 2 で、論文 D と論文 E の引用数は 1 である。

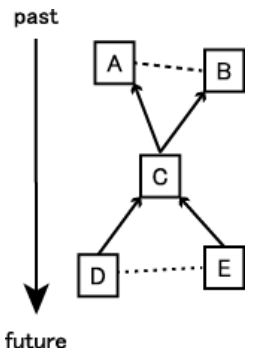

図 1. 論文引用のパターン

共引用とは、2 つの論文が同じ論文によって引用されている場合のつながりである。例えば、論文 A と論文 B が論文 C によって引用されている場合、論文 A と論文 B は共引用の関係にあるといわれる。図 1 の破線が、この関係に相当する。

書誌結合とは、2 つの論文が同じ論文を引用した場合のつながりである。例えば、論文 D と論文 E が論文 C を引用している場合、論文 D と論文 E は書誌結合の関係にあるといわれる。図 1 の点線が、この関係に相当する。

このように、論文の引用は、論文をノードとし、引用関係をリンクとすれば、グラフやネットワークと考えることができる。そのため、計量書誌学や科学計量学ではこれまで、社会ネットワーク分析の手法を用いた解析がなされてきた(Leydesdorff, 2001)。しかし、近年、急速に発展してきたネットワーク科学の立場からも、論文引用ネットワークが研究されている。たとえば、Redner(1998)は、ISI に収録されている論文の中で、1981 年に出版された 783,339 本の論文や、Physical Review D の第 11 巻から第 50 巻に収録された 24,296 本の論文に対し、直接引用によるネットワークの次数分布を解析した。その結果、次数分布が、べき分布にしたがっていることを明らかにした。また、Redner(2005)は、Physical Review を対象とし、1893 年から 2003 年 6 月 30 日までに出版された 353,268 本の論文に対し、直接引用によるネットワークの次数分布を解析し、その分布が対数正規分布にしたがっていることを明らかにした。これらの研究の他にも、直接引用によるネットワークの次数分布を解析した研究がいくつかあるが、その結果は、べき分布か対数正規分布のどちらかを支持している。このように、相矛盾する結果が混在している原因の一つは、解析に用いているデータの網羅性にある。そこで本稿では、網羅性の高いデータを用いて、次数分布を再考する。

＊　本研究は，平成 24 年度日本大学理工学部基礎科学研究助成金（研究助成 B）の助成を受けています.

１：日大理工・教員・一般

２：独立行政法人 科学技術振興機構

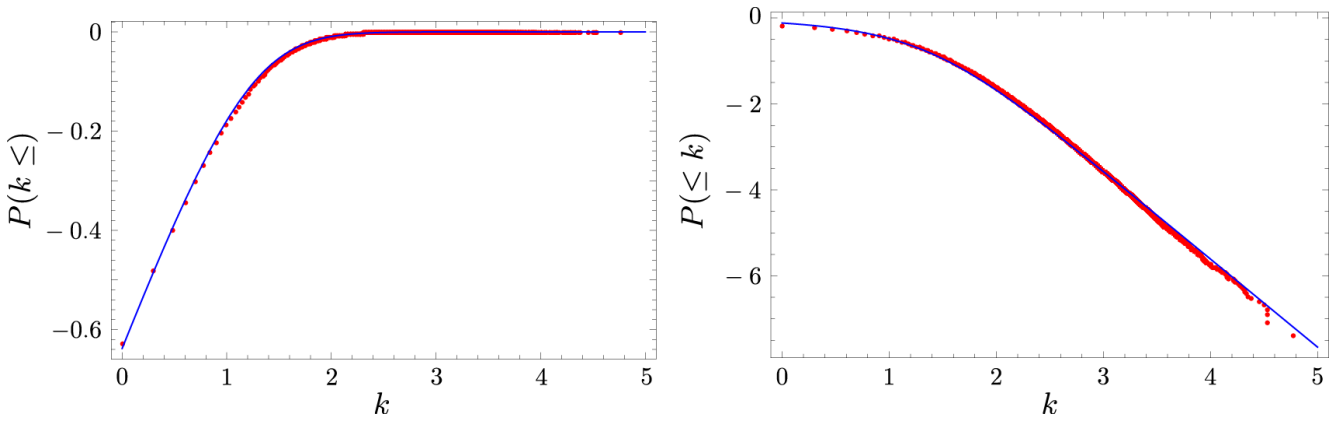

図 2. 被引用数の累積分布（左）と補累積分布（右）。図中の実線は、第 2 種一般化ベータ関数。

２．データ、解析結果、展望

本研究では、Thomson Reuter's Web of Science (WoS) の Science Citation index (SCI) に含まれている論文として、1981 年から 2011 年までに出版された論文を解析の対象とする。このデータは、24,160,348 本の論文を含み、総引用数は 385,692,359 件に及ぶ。そして、これらが形成するネットワークで、各論文の被引用数（図 1 で各論文に入ってくる矢印の数）の分布をプロットすると、図 2 が得られる。双方の図で横軸は被引用数で、図 2（左）の縦軸は累積確率密度、図 2（右）の縦軸は補累積確率密度である。どちらの図においても、両軸ともに対数をとっている。これらの図から、分布の両側は直線的になっていることがわかる。このことは、分布の裾野部分は両側とも、べき分布にしたがっていることを意味している。このような性質は、引用数の分布に対しても当てはまる。

分布の両側でべき分布が再現される、という性質を持つ関数としてよく知られているものとして、第 2 種一般化ベータ関数がある。この関数形は、

$$p(k;\mu,\nu,q,k_0) = \frac{q}{kB(\mu/q,\nu/q)}\left(\frac{k}{k_0}\right)^{\nu}\left[1+\left(\frac{k}{k_0}\right)^{q}\right]^{(\nu+\mu)/q}$$

で与えられる。図中の実線は、この関数で分布をフィットした結果である。このように、被引用数の分布は、第 2 種一般化ベータ関数でよく説明できることがわかる。

次のステップとして、論文引用のダイナミクスを理解するためには、第 2 種一般化ベータ関数を再現する数理モデルの構築が必要である。しかし、第 2 種一般化ベータ関数は、保険数理の分野などでよく用いられるが、それを再現するような確率的数理モデルはまだ存在していないため、モデルを最初から構築する必要がある。そして、その際にヒントとなるモデルが、修正 BA モデルだと考えられる。これは、被引用数の時間変化を

$$\frac{dk_i}{dt} = A_t k_i^{\alpha}$$

で説明するものである。発表では、このモデルの妥当性について議論する。

３．参考文献

[1] Small, H. (1973), Co-citation in the scientific literature: A new measure of the relationship between two documents, Journal of the American Society for Information Science, July-August, 265-269, 1973.

[2] Kessler, M.M. (1963), Bibliographic coupling between scientific papers, American Documentation, 14, 10-25.

[3] Leydesdorff, L. (2001), The Challenge of Scientometrics: The development, measurement, and self-organization of scientific communication, Universal-Publishers.

[4] Rednar, S. (1998), How popular is your paper? An empirical study of the citation distribution, Eur. Phys. J. B4, 131-134.

[5] Redner, S. (2005), Citation statistics from 110 years of Physical Review, Physics Today, Vol. June, 49-54.