

Wikipedia のハイパーリンク構造を用いた漫画家の影響力の定量化 Quantifying influence of manga artists using the Wikipedia hyperlink structure

高橋 佑樹* 田中 恭平** 芳鐘 冬樹***
Yuki Takahashi Kyohei Tanaka Fuyuki Yoshikane

1. はじめに

近年、漫画は、国内のみならず、世界に通用する日本の優れた文化として広く知られるようになった。日本の文化庁も漫画を「メディア芸術」としており、漫画を文化として高く評価している。それに付随して、漫画に関する研究がいくつか見られるようになった。しかし、漫画に関する研究は質的調査によるものが多く見られ、定量的な観点からのものは少ない。

そこで、本研究では、定量的な観点から漫画家同士の影響関係を明らかにするという目的を設定した。本研究の意義は、実際に影響関係を把握することだけでなく、今後に向けて、質的観点から評価されることが一般的であった漫画家の新たな評価手法を提案することにある。これらにより、今後の漫画界の発展に寄与できるのではないかと考える。また、小説や映画など、他の分野の評価にも応用できると考えられる。

2. 先行研究

2.1 分析対象に関する先行研究

質的な観点から漫画家の影響力を評価したものはいくつか見られる。例えば、堀[2009]や村上ら[1987]は、萩尾望都らをはじめとする「24 年組」(漫画界に強い影響を与えたとされる集団)について質的なアプローチで評価している。一方、量的なデータ分析に基づいて影響力を評価した例はほとんどない。

2.2 分析方法に関する先行研究

高橋ら[2011]は、歴史に関する事柄の重要度について、Wikipedia の記事間の参照関係に着目して計量する手法を提案している。

また、Wikipedia の知識抽出コーパスとしての有用性に言及している研究として、Evgeniy・Shaul[2007]と中山ら[2008]がある。Evgeniy・Shaul は、概念間の意味的関連度の測定において、Wikipedia が有用であると述べている。また、中山らは、密なリンク構造、質の高いリンクテキスト、リアルタイム更新、語彙の一貫性など、Wikipedia が他の Web コーパスと比較して有利な点を多く持つと述べている。本研究は、これらを踏まえ、漫画家の影響力の評価に Wikipedia の記事間の参照情報を用いる。

3. 分析方法

本研究では、高橋らの手法と同様に日本語版 Wikipedia の記事間のハイパーリンク構造に基づいて、漫画家同士の影

響関係について分析し、影響力を計量する手法を提案する。分析には、Wikipedia のメタデータを構造化したデータベースである DBpedia を用い、EndPoint(検索機能)で記事のタイトル、記事間のハイパーリンクの情報を抽出する。

その後、抽出したハイパーリンク情報をもとに、Wikipedia の記事をノード、記事間のハイパーリンクをエッジとした有向グラフの隣接行列を作成し、PageRank アルゴリズムで用いられている計算式を適用して記事の重要度を算出する。

3.1 使用データ

本研究で用いるデータを表 1 に示す。

データセット	日本語版 Wikipedia 2014-01-06 のダンプデータにおける「日本の漫画家」カテゴリに属する記事全て
ノードの総数	5,240
エッジの総数	7,266

表 1 使用データ

3.2 PageRank アルゴリズム

本研究では、漫画家同士の影響力の算出に PageRank アルゴリズムの計算式を用いる。PageRank アルゴリズムとは、Brin・Page[1998]によって提案された、Web ページ間のハイパーリンク構造を利用することでページの重要度を測定して順位付けを行うアルゴリズムである。その計算式は以下の通りである。

$$PR(A) = (1 - d) + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right)$$

ページ A にリンクしているページ T_n の PageRank の値を $PR(T_n)$ 、ページ T_n に含まれる他ページへのリンク総数を $C(T_n)$ とする。また、 d は Damping Factor を表しており、Brin・Page はこの値を 0.85 と定めている。本研究でも同じ値を用いて計算を行った。

3.3 影響力

PageRank アルゴリズムに基づき算出された記事の重要度にスケールファクタとして 100 を乗じた値を、漫画家の影響力とする。漫画家の記事間のリンクは、アシスタントや師弟の関係など、影響関係を示すケースが多いことから、リンク情報をもとに漫画家の影響力を評価できると考える。漫画家のリンク関係の例を図 1 に示す。図 1 は、「尾田栄一郎」の記事からリンクしている記事と、「尾田栄一郎」の記事にリンクしている記事を抽出し、有向グラフとしてプロットしたものである。

* 筑波大学情報学群知識情報・図書館学類

** 筑波大学大学院図書館情報メディア研究科

*** 筑波大学図書館情報メディア系

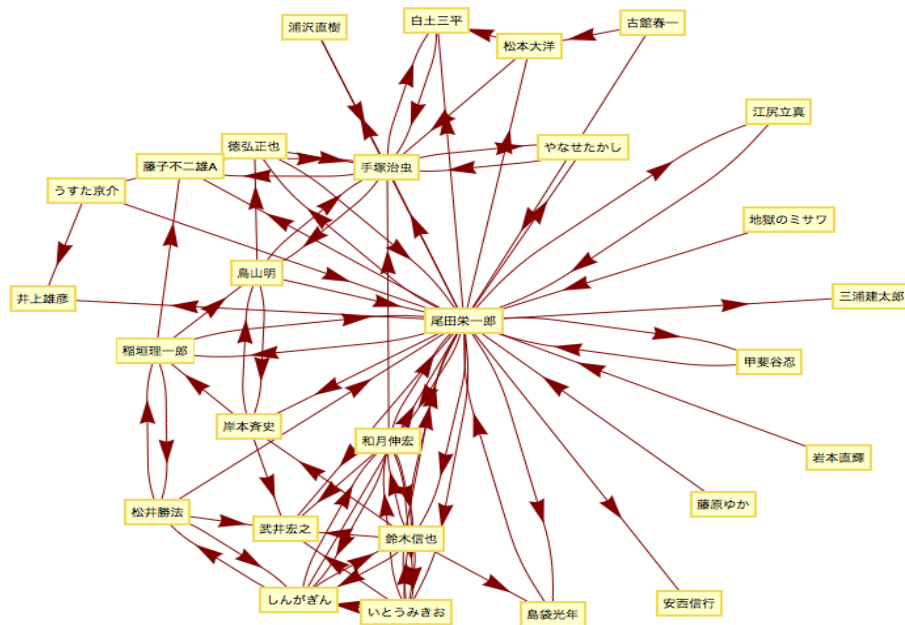


図1 漫画家同士のリンク関係の一例

4. 分析結果

順位	漫画家の名前	影響力	出次数
1	手塚治虫	1.7992	91
2	石ノ森章太郎	1.0229	27
3	ちばてつや	0.5903	8
4	永井豪	0.5471	35
5	赤塚不二夫	0.5443	25
6	大友克洋	0.3475	20
7	松本零士	0.2997	10
8	藤子不二雄	0.2957	9
9	横山光輝	0.2922	10
10	田河水泡	0.2872	10

表2 分析結果

表2は、影響力の数値が高かった上位10人を抽出してまとめたものである。手塚治虫の影響力が1.7992で最も高く、石ノ森章太郎、ちばてつやと続く。また、比較のため、単純に関係の量を示す出次数もあわせて表に記した。

5. 考察

手塚治虫と石ノ森章太郎の影響力が突出している結果が得られた。また、漫画家をジャンルごとにカテゴライズすると、少年漫画・SF漫画を中心に執筆している漫画家の影響力が高く、逆に少女漫画を中心に執筆している漫画家の影響力は低い傾向があった。堀や村上らにより言及されていた24年組に属する漫画家は上位10人には入っておらず、16位の「萩尾望都」がその中で最も順位が高かった。

また、PageRankアルゴリズムに基づく影響力と出次数の相関係数を算出したところ、強い正の相関($r = 0.9184$)が認められた。ただし、表2に示したように、上位10件の中でも、

両者の順位付けには差異があり、より単純な指標(後者)では、本研究が提案する指標(前者)を必ずしも代用できない。

6. おわりに

本研究は、定量的な観点から漫画家同士の影響関係について明らかにすることを目指した。具体的には、Wikipediaのハイパーリンク構造に着目して、漫画家の影響力の強さを定量化する手法を提案し、実データに基づく分析を行った。

今後の課題としては、漫画家を生没年などでカテゴライズし、より細分化した分析を行うことが挙げられる。また、媒介中心性などの指標や、質的な観点から評価した漫画家の影響力と比較して、分析手法・結果の妥当性を検証することにも取り組みたい。

参考文献

- 堀あきこ. 欲望のコード-マンガにみるセクシュアリティの男女差-. 臨川書店, 2009, p.47-48.
- 村上知彦, 高取英, 米沢嘉博. マンガ伝「巨人の星」から「美味しんぼ」まで. 初版第1刷, 東京, 平凡社, 1987, 338p, ISBN4-582-74206-8.
- 高橋侑久, 大島裕明, 山本光穂, 岩崎弘利, 小山聡, 田中克己. インパクトを考慮した歴史エンティティの重要度計算手法. 情報処理学会論文誌, 2011, Vol.52, no.5, p.3542-3557.
- E. Gabrilovich, S. Markovitch. Computing semantic relatedness using wikipedia-based explicit semantic analysis. Proc. of International Joint Conference on Artificial Intelligence (IJCAI), p.1606-1611, (2007).
- 中山浩太郎, 原隆浩, 西尾章治郎. 自然言語処理とリンク構造解析を利用したWikipediaからのWebオントロジ自動構築. 日本データベース学会論文誌. 2008, Vol.7, no.1, p.67-72.
- Brin, S; Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer networks and ISDN Systems, 1998, Vol.30, no.1, p.107-117.