

ウィキペディア編集コラボレーションにおける 編集者間の社会ネットワークの効果

根本 啓一^{1,a)} ピーター グロア² ロバート ローバッカー² 岡田 謙一¹

受付日 2014年4月13日, 採録日 2014年10月8日

概要: 本稿では, 日本語版と英語版のウィキペディアを対象とし, ウィキペディアの記事を編集するユーザ間に存在する社会ネットワークが, 記事編集という協調作業においてどのような影響があるかを分析した. 記事編集に関わるユーザ間の社会ネットワークを計測するために, 個々のユーザが持つユーザ会話ページへの書き込みによるインタラクションに着目した. 英語版のウィキペディアでは, 記事の質が最も高い 3085 の秀逸な記事と, 比較的質の高い良質な記事を含む 80154 の記事での記事編集に関わるユーザのインタラクションを取得した. 日本語版ウィキペディアでは, 69 の秀逸な記事と 873 の良質な記事における記事編集に関わるユーザのインタラクションを取得した. 協調作業のパフォーマンス指標として, 記事の質があるレベルから 1 段階向上し, 秀逸な記事や良質な記事と評価されるまでに要した時間を利用した. その結果, 記事編集に関わるユーザ間に社会ネットワークが事前に構築されていると, 記事の質の向上に要する時間が短くなることが示された. さらに, ユーザ間のインタラクション関係を見ると, ユーザ間の関係構造がより密なネットワークを形成しており, 中心性の高いネットワークであると, 記事の質を高める協調作業のパフォーマンスが高いことが分かった. これらの結果からユーザ間の社会ネットワークが編集コラボレーションのパフォーマンスに寄与することが示唆された.

キーワード: 社会ネットワーク, コラボレーション, ウィキペディア, インタラクション, パフォーマンス

Social Network Increases Efficiency of Collaboration among Wikipedia Editors

KEIICHI NEMOTO^{1,a)} PETER A. GLOOR² ROBERT LAUBACHER² KENICHI OKADA¹

Received: April 13, 2014, Accepted: October 8, 2014

Abstract: In this study we measure the impact of pre-existing social network on the efficiency of collaboration among Wikipedia editors in the Japanese and English Wikipedias. To construct a social network among Wikipedians we look to mutual interaction on the user talk pages of Wikipedia editors. As our data set, we analyze the communication networks associated with 3085 and 69 featured articles — the articles of highest quality in the English and Japanese Wikipedia, comparing it to the networks of 80154 and 873 articles of lower quality from the English and Japanese Wikipedia, respectively. As the metric to assess the quality of collaboration, we measure the time of quality promotion from when an article is previously promoted until it is promoted to the next level. The study finds that the higher pre-existing social network of editors working on an article is, the faster the articles they work on reach higher quality status, such as featured articles. The more cohesive and more centralized the collaboration network, and the more network members were already collaborating before starting to work together on an article, the faster the article they work on will be promoted or featured.

Keywords: social network, collaboration, Wikipedia, interaction, performance

1. はじめに

ウィキペディアはインターネット上のオープンなコラボレーションにより作成されている百科事典である。その運営方法は一般の企業などとは異なり、階層的な組織は存在せず、基本的にオンライン上のだれでもが編集可能とすることで、人類の英知を結集し、既存の百科事典に匹敵する質の記事を無料で公開している [1]。オープンソースソフトウェアの開発とならびインターネット上の集合知として、最も成功した事例の1つとして注目されている。現在 287 の言語^{*1}で開設されており、その記事編集プロセスには文化的特徴と関連することが知られている [2]。しかし、公式な組織を有しないオープンなコラボレーションがどのようにして質の高い記事を作成しているかについてはまだ十分に明らかになっていない。

今後増加していく多様で複雑な社会的課題に対し、ウィキペディアに見られるようなオンライン・分散環境のオープンなコラボレーションのメカニズムを明らかにすることにより、コラボレーションによる英知を結集して、複雑な課題を解決する分散コラボレーションの支援につなげていけると考える。そこで、文化的な差を考慮しウィキペディアのコラボレーションのメカニズムを明らかにするために、最も記事数が多い英語版のウィキペディアと日本語版のウィキペディアを対象とした分析を行うこととした。

コラボレーションのパフォーマンスに対して、参加者間の社会関係資本は、その他の様々な資本と同様に、様々な利益をもたらすことが調査されてきた [3], [4]。CSCW の文脈では、分散コラボレーションであるオフショア開発のプロジェクトを事例とした研究が行われており、信頼に着目した社会関係資本はプロジェクト遂行に対し正と負の効果があることが示されている [5]。

これらの社会的関係資本は信頼、規範、ネットワークという概念を特徴としており [6]、社会関係資本が指し示しているのは個人間のつながり、すなわち社会的ネットワーク、およびそこから生じる互酬性と信頼の規範であるといわれている [7]。バートは社会ネットワークにおける構造的空隙と社会関係資本の関係を明らかにしており [8]、Borgattiらは社会ネットワークの構造と社会関係資本の関係を整理している [9]。リンはこれらの理論を整理し、社会関係資本を、人々が何らかの行為を行うためにアクセスし活用する社会ネットワークに埋め込まれた資源と定義している [10]。これらの既存研究から、社会関係資本は社会ネットワークの存在が前提にあるという一致が見られる。そこで本稿では、社会関係資本の前提と考えられる社会ネットワークに

着目し、社会関係資本を社会ネットワークの側面からとらえることとした。

しかし、個人やグループが形成してきた社会ネットワークを計測することは、調査票などの社会調査手法を利用した既存の手法では計測することが困難であった [11], [12]。そこで本稿では、人々のインタラクション履歴がログとして残り、計量可能な環境であるウィキペディアというオンラインのコラボレーション環境を対象とし、ウィキペディアの編集者グループが持つ社会ネットワークがどのようにグループのコラボレーションに影響しているかを調査した。特に、編集者間で構築された社会ネットワークが、記事の質を一定以上とするための編集作業にかかる時間に対し、どのように影響しているかを調査した。

既存の研究では、企業内の組織構造とは異なる非公式な社会ネットワークが組織のパフォーマンスに影響があることが研究されてきた [13]。しかし、このような非公式な社会ネットワークは、調査票などの社会調査手法によって同僚との関係性を主観的に聞くことで計測するにとどまっていた [14]。近年では、このような関係性を電子メールの交換の履歴から計測する研究などが行われており、電子メールの交換によって形成された社会ネットワークにおける媒介性 [38] や構造的空隙 [8] といったネットワーク的な位置とその個人のパフォーマンスに高い相関があることが示されている [15]。そこで我々はウィキペディアの編集に関わるユーザ間での社会ネットワークを抽出し、計測することで、ウィキペディアというオンラインコラボレーション環境においても、社会ネットワークが記事編集のパフォーマンスに影響するのではないかと考える。

2章ではコラボレーションにおけるパフォーマンスやウィキペディアに関する関連研究について述べる。3章では、ウィキペディアでの編集作業とユーザ間のインタラクションについて説明し、本稿で着目したユーザ間の社会ネットワークの取得方法、記事編集のパフォーマンスの計測方法、そしてこれらをモデル化する分析方法について述べる。4章で結果を示し、5章でコラボレーションのパフォーマンスの視点から考察を行い、6章で結論と今後の課題について述べる。

2. 関連研究

ウィキペディアは、インターネット上でのグループコラボレーションの例として、オープンソースプロダクション [16]、ピアプロダクション [17]、クラウドソーシング [18]、協調イノベーション [19]、インターネット上での集合知 [20] といった、様々な観点から注目されている。このようなオープンな協調作業環境に関する関連研究として、ユーザ間のインタラクションパターンとグループパフォーマンス

¹ 慶應義塾大学

Keio University, Yokohama, Kanagawa 223-8522, Japan

² MIT Center for Collective Intelligence, Cambridge Center, Cambridge, MA 02142, USA

a) nemoto@a7.keio.jp

*1 2014年7月1日現在

の研究, オープンソースソフトウェア開発に関する研究, そしてウィキペディアそのものの研究について記述する。

(1) インタラクションとグループパフォーマンス

企業などの組織において, 社会ネットワーク構造のグループパフォーマンスへの影響に関する研究は多く存在する [15], [21], [22], [23], [24]. これらの研究は, 協調作業のタスクの違いによって高いパフォーマンスを出す社会ネットワークの構造が異なることを示している. Uzzi らによる研究 [24] ではブロードウェイでの舞台制作というタスクでは, 過去の協業経験や, タスク参加者の多様性, たとえば新参者の参加, といった様々なタイプのネットワークの特徴が, 舞台の興行収入と関係があることを示した. すなわち過去に構築した社会ネットワークがその後の協業におけるパフォーマンスへ影響していることが示唆されている. しかし, このような既存の関係性がウィキペディアのようなインターネット上のコミュニティ活動でも成り立つかを明らかにされていない。

(2) オープンソースソフトウェア (OSS) 開発

インターネットによるボランティアベースの協調作業の事例として, オープンソースソフトウェア (OSS) の開発に関して多くの研究がなされている. さらに, OSS は既存のソフトウェア開発プロジェクトによるものよりも, 質が高いことが示されている [25]. 最も典型的な研究は, どのようにして OSS 開発チームが協調し, どのような動機づけによって参加しているかといった点を調査したものである [26], [27].

OSS 開発をより効果的に行うための方法に関する研究も多く実施されている [28]. これらの研究は, 組織的な実践手法, ソフトウェアのアーキテクチャ, プロジェクトマネジメントや開発プロセスに関する点に着目されている. これらの研究の 1 つとして, オニオンモデルと呼ばれる, 少数のコアとなる開発者と, その周辺に存在するバグの報告や, 実際の試用を通じてフィードバックをする周辺的な参加者によるコミュニティ構造の発見があげられる [27]. さらに, OSS 開発における参加者間の社会ネットワーク構造分析から, 中心性を持つネットワーク構造は, バグを修正するなど主体とする組織において効果的であり, 中心性が変動するような組織では, 新たなコードを生成するといったタスクにおいて効果的であることが明らかになっている [29]. しかし, このようなコラボレーションにおける社会ネットワーク構造が, ウィキペディアでの記事作成や記事の質向上というタスクに参加するユーザ間に存在しているかは明らかにされていない。

(3) ウィキペディアに関する研究

ウィキペディアにおいて, 記事の質ではなく記事編集をするユーザの質に着目した初期の研究として, Anthony ら [30] の研究があげられる. この研究では, ユーザが過去に編集した箇所のうち, 現時点でどの程度割合が残ってい

るかによってユーザの質を計測している. ウィキペディアのユーザには, 登録されているユーザと, 登録されていないユーザの大きく 2 種類が存在している. 質の高い記事編集は, 登録ユーザによる多くの編集と, 未登録ユーザによる少数の編集によるものであることが明らかにされている. このことから, 質の高い記事は, コアとなる経験豊かなユーザによる編集と, 特定の専門性がある周辺的なユーザによる修正によるものであることが示唆される。

ユーザの編集の質を推定する手法として, Adler ら [31] はユーザの編集の質を, その編集がウィキペディア上に存在している時間とし, Priedhorsky ら [32] は, その編集された部分がどの程度閲覧されたかによって計測した。

ウィキペディアの各記事全体の質の評価は, ウィキペディアのコミュニティによって進められており^{*2}, 2010 年 5 月の段階では, 英語版のウィキペディアは 2,714,054 の記事に 7 段階の評価が付与されている. また, これらのコミュニティによる評価は外部評価者の結果との有意な相関を示していることが分かっている [33]. Wilkinson ら [34] は記事の質変化の時間的な変遷に対して体系的な調査を実施している. 記事の作成から調査時点までの時間と, 記事の長さをそろえた秀逸な記事とランダムな記事の比較の両方から, 秀逸な記事では, より多くの編集と, 多くのユーザが関与していることを示している。

Kittur ら [33] は, ウィキペディアの記事編集の質に関して, 記事の方向性や範囲を定める初期の部分では特定の少数のユーザにより編集される「潜在的コーディネーション」方が, 多数のユーザによるものよりパフォーマンスが高いことを示している. 一方, 記事の方向性が定まり, 骨格ができた後は多数のユーザに編集される方がよりパフォーマンスが高いことを示した。

Liu ら [35] はクラスタリングの手法から, ウィキペディアのユーザを, その編集履歴から 6 つのタイプに分類している. さらに, 記事自体を 5 つのタイプに分類しており, 各分類のユーザがどのタイプの記事を編集しているかを分析している. その結果, 最も質の高い記事は様々な記事を編集している all-around editor と分類されるユーザによって作成されていることが示されている。

これらの研究から, 経験の豊かなコアの編集者と, ときどきに参加する周辺的な参加者による少量の編集によって質の高い記事が作成されていることが示唆される. すなわち, ウィキペディアの記事編集を行うユーザ間の社会ネットワークは, コアな編集者による中心性ととも周辺的な参加者による凝集性の高い関係構造となることが考えられる。

^{*2} http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Council/Guide/WikiProject#Assessments_in_practice

3. 手法

本章では、ウィキペディアを編集するユーザ間の社会ネットワークを推定するための手法について説明する。3.1節では、記事編集作業の過程において、ユーザ間でどのようなインタラクションが行われているかについて説明し、3.2節でそこからユーザ間の社会ネットワークを抽出した方法について記述し、3.3節で社会ネットワークの記事編集への影響度を推定するため、ユーザが取り組んでいる記事編集の難易度を規定する方法について説明する。3.4節で協業ネットワークの記事編集タスクへの影響度を分析する生存時間分析の方法と、モデルに取り入れた変数について説明する。

3.1 ウィキペディアの編集作業で発生するインタラクション

多くのウィキペディアを利用するユーザは、記事自体のページのみを閲覧しており、場合によっては記事に付随する議論（ノート）ページを見る程度である。一方、記事編集の視点から見ると、ウィキペディアには大きく3つの種類のページ、記事ページ、ユーザページ、ウィキプロジェクトページが存在する。

記事ノートページは、記事編集に関する議論する場所として、すべての記事ページと対となって存在している。記事ノートページでは、たとえば記事編集にともない、記載されている内容の確認や編集での方針、削除すべき項目や追加すべき項目などが議論されている。

一方、ウィキペディアには、記事ページ以外にも、登録ユーザごとのユーザページが存在する。ユーザページは個人のホームページのように利用されており、個々人の関心がある領域の表明や、他のウィキペディアユーザから得たバッジなどを掲載しているユーザも存在している。さらに、このユーザページは、記事ページと同様にノートページが対で存在している。記事編集の過程では、ユーザが行った記事編集に対する感謝のメッセージの送付や、意見を求めること、ユーザの行った編集内容に関する個別の議論など、ユーザノートページを利用した個と個のインタラクションが行われている。

このほかに、個別の記事を横断して、あるテーマに対する編集方針などを議論するウィキプロジェクトに関するページが存在する。ウィキプロジェクトに関する詳細は後述するが、ウィキプロジェクトではテーマ共通に必要な記事構造の作成や、編集の指針などが議論されている。記事編集過程では、このウィキプロジェクトの指針などを参照に記事編集を行うことが考えられる。

このように、ウィキペディアでの編集作業で発生するユーザ間のインタラクションは様々な場所で発生するが、個別の記事に関してのインタラクションは、主に記事ノ

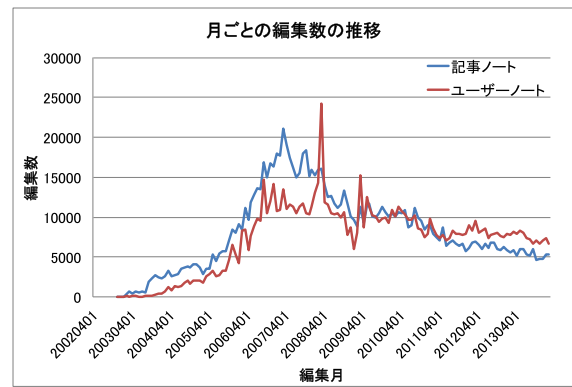


図 1 日本語版ウィキペディアでのノートページへの編集数の推移
Fig. 1 Temporal change of the number of edits on article talk and user talk pages in the Japanese Wikipedia.

トページとユーザノートページで行われている。

ウィキペディアでの編集数は日本語版、英語版ともに2007年にピークを迎えている*3。記事ノートページとユーザノートページに目を向けると、英語版のウィキペディアでは2003年から2005年にかけての増加率が、記事ノートページが11倍なのに対してユーザノートページへの書き込みは78倍であることが示されている[36]。日本語版のウィキペディアでは、ユーザノートへの編集数は2011年には記事ノートへの書き込み数を上回っていることが分かる(図1)。これらから、記事編集にともなうウィキペディアでのユーザのインタラクションは、記事に付随する記事ノートページの編集からユーザ同士の直接のやりとりへ移行してきているといえる。さらに、記事ノートでの議論は、特定の相手を指定しない議論であり、ユーザ同士の個と個の関係をとらえることが困難である。一方、ユーザノートへの書き込みは、明示的な個と個の関係であると考えられる。

そこで本稿では、社会関係資本の前提となる社会ネットワークとして、このようなユーザのノートページを通じた個と個のインタラクションに着目した。そして、特定の記事に関わらないユーザ間の直接的なやりとりが、ユーザ間の個と個のつながりである社会ネットワークを代替すると考えた。本稿では、このユーザノートページでのやりとりを協業ネットワークと呼び、協業ネットワークによって計測された社会ネットワークの記事編集というタスクに対する影響度の分析を行った。

3.2 記事編集者の協業ネットワーク

ユーザ間の協業ネットワークを抽出するため、Crandallら[37]の手法を参照し、同一の記事の編集に関わるユーザを対象とし、それらのユーザがお互いのノートページへの書き込みを行うことをユーザ間の関係と定義した。たとえば、ユーザA、B、Cが記事編集に関わっていた場合、Aが

*3 <http://stats.wikimedia.org/EN/TablesDatabaseEdits.htm>

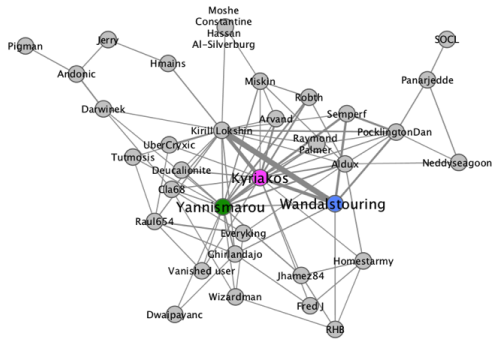


図 2 英語版ウィキペディア「War against Nabis」記事編集者の協業ネットワーク

Fig. 2 Collaboration network constructed by comments on user talk pages about the article “War against Nabis”.

B のユーザノートページへ書き込みをし、B が C のユーザノートページへ書き込みをしていた場合、A と B の間、B と C の間に関係があるものとした。

ウィキペディアを編集するユーザには、大きく分けて匿名の IP アドレスユーザ、bot と呼ばれる機械プログラム、そしてユーザ登録をしている登録ユーザが存在する。英語版ウィキペディアでの全体の編集に占めるユーザの割合は、IP ユーザが 23%、bot が 3%、登録ユーザが 74%であった。日本語版ウィキペディアでは、IP ユーザが 9%、bot が 12%、登録ユーザが 79%であった。協業ネットワークの構成対象としては、ユーザのノートページを保持している登録されているユーザのみとした。匿名の IP アドレスユーザは、ログインをしていないユーザであり、アクセス元の IP アドレスが便宜的にふられている。そのため、ユーザと IP アドレスの関係は一意である保証がなく、協業ネットワークを抽出することができない。このような特定できない相手とのネットワークは、相手が特定できるネットワークと異なる可能性もあるが、以上の理由から匿名の IP アドレスユーザを対象外とした。また、bot は機械プログラムが利用しているアカウントであり、bot の編集は自動化されている。記事編集に関わる人と人の社会ネットワークという観点から、bot を対象外とした。

図 2 では、“War against Nabis” という記事の編集者間での協業ネットワークである。ノードはこの記事を編集したユーザ、エッジはユーザのノートページへの書き込みの関係を表している。この記事は、2006 年 11 月 18 日に作成され、2007 年 2 月 17 日の 3 カ月で FA (Featured Article) に昇格した記事である。この記事での最も活発なユーザは Kyriakos であり、全編集の 41% の記事編集 (651 回に対して 266 回の編集) を実施していた。Wandalstouring は 35% (651 回に対して 228 回) の編集を行っていた。

図 3 には、実際の Kyriakos から Yannismarou へのコメントを記載している。Yannismarou は全編集の 3% を行っているユーザである。図 3 に示したコメントから、ユーザ

Roman-Spartan War

[edit]

Hi Yanni, I withdrew the Roman-Spartan War from FAC and I put it on Peer Review at WPMILHIST. So if you have the time and don't mind could you please go and leave some suggestions because they always seem to be of good value and worth while. Thanks in advance and also I forgot to congratulate you on El Greco becoming a FA. It's great that you took the article of the greatest Greek painter to FA level. Kyriakos 05:49, 4 January 2007 (UTC)

図 3 Yannismarou のユーザノートページへの Kyriakos の書き込み

Fig. 3 A friendly comment on the user talk page of Yannismarou.

間が記事に関する連絡や個人的な友好的なコメントを行っていることが分かる。

協業ネットワークで関係性を表す紐帯は重さや方向のない無向グラフとして扱った。これは、ユーザ同士のユーザノートページへ書き込みや、その返信を行う際に、返信を相手のノートページに行う場合や、自身のノートページに返信を書く場合が存在するためである。

このような記事編集に関わるユーザ間の協業ネットワークを測る指標として、グループ中心性 [38] (Group Degree Centrality: GDC), クラスタリング係数 [39] (Clustering Coefficient: CC) を用いた。GDC は各ノードが持つ次数 (エッジの数) の分布を表す。あるノード n_i の次数を $d(n_i)$ 、ネットワークのノードの数を g としたとき、ネットワーク全体のグループ次数中心性は以下の式から算出することが可能である。

$$C_D = \frac{\sum_{i=1}^g [C_D(n^*) - C_D(n_i)]}{[(g-1)(g-2)]}$$

$C_D(n^*)$ はネットワークの全ノードの中での最大次数

GDC の最大値は 1 であり、このとき 1 人のユーザのみが他のすべてのユーザへのエッジを持つ状態であり、すべてのユーザが同数の次数を持つ場合に最小値の 0 となる。GDC の値が高いとき、そのネットワークは中心的な影響力の強いユーザの存在を示唆する。

クラスタリング係数はネットワークの凝集性を表す。最大値は 1 となり、任意の n_i とエッジを持つ n_j と n_k の間にエッジが存在する状態である。逆に任意の n_i とエッジを持つ n_j と n_k の間にエッジが存在しない場合に最小値の 0 をとる。高いクラスタリング係数はネットワークが凝集性の高い、相互に密な関係構造を築いていることを表す。

3.3 記事編集難易度の推定

ウィキペディアでの記事編集作業は多岐にわたるため、ユーザグループが取り組む個々の編集作業を比較することは困難である。そこで、記事の質の変化から、記事編集の難易度を規定する方法を用いた。それにより、特定の難易度の記事編集タスクのみを抽出し、タスクに取り組むユーザグループ間のパフォーマンスの違いを比較できる。

ウィキペディアの記事の質はウィキプロジェクト (3.4.3 項に詳述) によって評価されるものと、プロジェクトに限らずコミュニティ全体で評価されるものとに分かれている*4。英語版のウィキペディアでは、記事の質は低いものから Stub, Start, C, B, GA (Good Article), A, FA (Featured Article) の順に7段階のクラスで評価できるようになっている。記事が最も高い評価である FA クラス (日本語版では秀逸な記事) となるまでの記事の変遷の例として、Atom という記事では、2001年10月8日に Stub クラスとなり、その後、2002年9月20日に Start クラス、2004年9月18日に C クラス、2006年8月19日に B クラス、2008年2月10日に GA クラス、そして2008年2月12日 FA に昇格するといった変遷を経ている*5。日本語版のウィキペディアでは、良質な記事と秀逸な記事の2つの評価がなされている*6。以降では、良質な記事は GA、秀逸な記事は FA と呼ぶ。

英語版のウィキペディアから、2010年11月17日現在での3,085件のFAを、2010年11月20日に10,058件のGAを、2010年11月29日に70,096件のBの記事を収集した。日本語版のウィキペディアからは2014年3月30日に69件のFAと873件のGAの取得を行った。

それぞれの記事がいつFAやGAへ昇格したかを示す昇格日の取得を行った。昇格日を得ることができたのは英語版のウィキペディアではFAが3,080件、GAが10,051件、そしてBクラスが69,627件であった。日本語版のウィキペディアではFAが69件、GAが873件であった。

英語版のウィキペディアでは記事の昇格プロセスには典型的なパターンが存在しており、47%のFAはGAから昇格をしており、AクラスからFAへと昇格するものはわずか6%である。同様に41%のGAはBクラスから昇格したものである。そこで、本稿では収集した記事の中から、現時点での記事クラスへの昇格日に加えて、それ以前での記事クラスへの昇格の履歴をすべて抽出し、収集した記事がどのような記事クラスの遷移をしていたかを取得した。たとえば、FA記事であれば過去にBクラスやGAに昇格した日付の取得を行った。これらのデータの中から、英語版ウィキペディアの昇格の遷移として、StartクラスからBクラスへ昇格した記事、BクラスからGAへ昇格した記事、GAからFAへと昇格した記事を対象とした。昇格した記事との比較のため、昇格していない記事群としてBクラスにとどまっている記事群とGAにとどまっている記事群も分析対象とした。日本語版のウィキペディアでは、評

価レベルがGAとFAしかないため、GAからFAへ移行したものを対象とした。GAからFAへと昇格した日付を取得できた昇格記事群は24件であった。英語版ウィキペディアと同様に昇格した記事との比較のため、昇格していない記事群としてGAにとどまっている記事群も分析対象とした。

それぞれ、Start, B, GA, FAの順番で記事の質が高くなるため、StartからBに記事の質を向上させるタスクは、BからGAにするタスクよりも難易度は低いといえる。同様に、BからGAはGAからFAよりも難易度は低く、GAからFAが最も難易度の高いタスクとなる。

たとえば、BからGAとGAからFAにおける編集作業を取り上げると、GAに昇格するためには、「記事のカバーする範囲が十分か」が判断基準の1つとなるが、FAに昇格するためには、「重要な事項が記述されており、さらに unnecessary部分を除かれていること」などが基準としてあげられている*7。すなわち、GAは必要なポイントをカバーするという編集であり、含めるべき項目をあげ、それらの項目に関して個々に記事編集を行うことができると考えられる。一方、FAでは重要な項目と unnecessary項目を整理する必要があり、記事に関わる編集者間で、含めるべき項目と削除すべき項目の意思決定の調整作業が多く発生し、それらを解消することが記事の質の向上につながるといえる。すなわち、FAでは質の向上のために、このような編集者間の調整作業が多く発生する難易度の高い編集作業となることが考えられる。

3.4 生存時間分析

3.4.1 Cox 比例ハザードモデル

編集に関わるユーザ間の社会ネットワークと、その記事作成のパフォーマンスとの関係性を推定するため、記事編集に関わるユーザが事前に保持していた社会ネットワークと記事昇格の時間との関係をモデル化した。記事昇格の時間は、有限の時間で観測を行っていることから、観測終了時点で記事昇格というイベントが発生しなかった記事も含まれる。しかし、このようなイベント発生までの時間データは、有限の観測期間でのデータであり、イベントが発生しなかったのではなく、あくまで観測期間内にイベントが発生しなかったと考えられ、このような不完全な観測値を扱うことができるモデルが必要となる。そこで、このような時間ともなうイベント発生に対して、イベントが発生しなかったデータを打ち切りデータとして扱うことができるCox比例ハザードモデル [40] を用いて分析した。Cox比例ハザードモデルでは、時間ともなうイベント発生の確率を、共変量の変化にもなうイベントの発生確率の比を求めることで、各共変量のイベント発生時間への寄与を

*4 <http://en.wikipedia.org/wiki/Wikipedia:Version.1.0.Editorial.Team/Assessment>

*5 http://outreach.wikimedia.org/wiki/Life_of_an_Article

*6 一部のウィキプロジェクトではA, B, Cでの記事の質を評価している (<http://ja.wikipedia.org/wiki/プロジェクト:執筆支援>) が、日本語版ウィキペディア全体での運用とはなっていないため本稿では良質/秀逸のみを対象とした。

*7 <http://en.wikipedia.org/?title=Wikipedia:Version.1.0.Editorial.Team/Assessment>

算出することができる。

目的変数は、3.3.2 項で詳述する記事の昇格というイベントまでの時間であり、独立変数を 3.3.4 項と 3.3.5 項で説明するウィキプロジェクトに関する指標と、既存協業ネットワーク、すなわち編集者間での社会ネットワークを表す指標とした。

Cox 比例ハザードモデルでは、ある時期 t における既存協業ネットワークの指標と記事昇格確率は以下のように定義される。

$$h(t|X) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots)$$

$h(t|X)$ が時刻 t での昇格というイベントが発生する確率であり、 $h_0(t)$ がすべての独立変数が 0 のときの基本確率となり、 x は共変量である。

分析対象には、昇格した記事とともに、観測段階で記事の昇格というイベントがまだ発生していない記事群、すなわち B に昇格したが、データを取得した段階で B にとどまっている記事、もしくは GA へ昇格したが、その後 GA にとどまり FA へ昇格していない記事も取り入れ、イベントが発生していない打ち切りデータとしてモデルに加えた。

3.4.2 目的変数：記事昇格までの時間

記事の質向上というタスクにおけるパフォーマンスを表す目的変数として、1 つ前の記事クラスから次の記事クラスへの昇格までの時間を目的変数とした。この時間は、記事編集に関わるユーザが、記事の質を 1 つ上のレベルに押し上げるタスクを遂行する時間と考えることができ、この時間が短いほど、ユーザのコラボレーションのパフォーマンスが高いと考えることができる。分析対象としたものは、英語版の記事では、FA へ昇格、もしくは GA へ昇格した記事であり、それぞれに対して、1 つ前のレベルに昇格した日付 (FA では GA へ、GA は B へ) が特定できるものとした。日本語版の記事では、FA のうち GA へ昇格した日付が得られた記事を対象とした。

3.4.3 独立変数：ウィキプロジェクト

英語版のウィキペディアでは、各記事はウィキプロジェクトと呼ばれる、編集者間でのプロジェクトマネジメントのような機能も果たすコミュニティに属しているものが多い。ウィキプロジェクトでは、共通のテーマの記事に対して、記事作成のガイドラインの作成、テーマに関連する記事のリスト化や、記事の優先順位付けなどを行う、記事の優先順位としては、テーマに関する記事を Top から Low の 4 段階に設定している。

たとえば、台風に関するウィキプロジェクト^{*8}では、記事の候補となる台風のリストを作成しており、その中でハリケーン・カトリナ^{*9}は主要な記事として選出されてお

*8 http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Tropical_cyclones

*9 http://en.wikipedia.org/wiki/Hurricane_Katrina

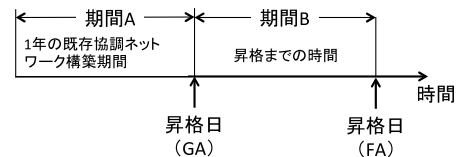


図 4 既存協業ネットワーク構築期間

Fig. 4 Explanation of pre-existing network construction.

り、重要度が Top に指定されている。

一方、ハリケーン・カトリナ記事からプロジェクトを見ると、4 つのテーマ (台風、米国、災害管理、ニューヨーク) のプロジェクトに属している。先の台風のプロジェクトと同様に、各プロジェクトが個々に記事の優先順位やガイドラインに沿って編集に関与していると考えられる。

ウィキプロジェクトはあくまで関心を共有するユーザによるコミュニティであり、記事編集に対して特別な権限があるわけではなく、どの程度記事に対して影響を与えているかは不明である。たとえば、重要度が高く設定されると、ユーザがその記事を積極的に編集するのかなどについては明らかになっていない。しかし、上記のように、各記事は複数のウィキプロジェクトから重要度を設定され、また複数のウィキプロジェクトにリストアップされていることから、このようなウィキプロジェクトの関与が記事編集に与える影響を考慮する必要があると考える。そこで、ウィキプロジェクトが記事の質向上に寄与しているかを調べるため、記事とウィキプロジェクトの関係を指標化し、因子として考慮した。

ウィキプロジェクトに関する指標は、ウィキプロジェクト関与数とウィキプロジェクト重要度の 2 つとした。関与数は記事に関与しているウィキプロジェクト数とした。重要度はウィキプロジェクトのメンバによって、記事に付与されている 4 段階の重要度 (Low から Top) の評価とした。なお、複数のウィキプロジェクト間で記事に付与される重要度が異なる場合は、最大のものを採用することとした。

3.4.4 独立変数：既存協業ネットワーク

記事編集を行ったユーザが、記事編集のタスクにとりかかる前に有していたユーザ間の協業ネットワークを既存協業ネットワークと定義とした。GA から FA に昇格した記事の場合、1 つ前の昇格 (GA) から次の昇格 (FA) までの期間に編集を行ったユーザを取得し (図 4 の期間 B)、1 つ前の昇格から過去 1 年間 (図 4 の期間 A) でのユーザ間で行ったユーザノートページへの書き込みを対象として既存協業ネットワークを構築した。この過去 1 年間の協業ネットワークが、記事の質を向上させようとするユーザ間にあらかじめ存在する既存の社会ネットワークを表す。

既存協業ネットワークを表す指標として、既存協業ネットワーク割合、グループ中心性、クラスタリング係数の 3 つを利用した。

図 4 の期間 B に記事編集に関わったユーザのうち、期

間 A で既存協業ネットワークを形成していたユーザの人数の割合を既存協業ネットワーク割合 R_{pc} とし、以下のよう
に定義した。

$$R_{pc} = N_p / N_{all}$$

N_p は期間 B に記事編集を行ったユーザのうち、期間 A で
すでに期間 B の他のユーザと協業ネットワークを構築して
いたユーザ数を表す。 N_{all} は期間 B に記事編集を行った
ユーザ数を表す。

R_{pc} が 1 のとき、昇格のために記事編集に関わった期間
B での編集ユーザ全員が期間 A の期間で既存協業ネット
ワークを形成していたことを表す。

既存協業ネットワークの構造を表す指標として、3.1 節で
記述した中心性を表す指標である、グループ中心性 (Group
Degree Centrality: GDC) と、凝集性を表すクラスタリン
グ係数 (Clustering Coefficient: CC) を利用した。

4. 結果

記事編集の難易度と協業ネットワークの分析結果と、記
事の質向上の時間へのウィキプロジェクトや既存協業ネッ
トワークの影響度の分析結果を英語版ウィキペディアと日
本語版ウィキペディアに分けて以下に示す。

4.1 英語版ウィキペディア

(1) 記事編集協業ネットワーク分析

表 1 に Start から B, B から GA, GA から FA の 3 つ
記事昇格パスに対して、記事の昇格にかかった時間 (図 4
の期間 B), その間に編集に関わった編集者数, 期間 B での
協業ネットワークの構造を表すグループ中心性 (GDC)
と凝集性を表すクラスタリング係数 (CC) の値を示す。本
文ではグループ中心性 (GDC) を中心性, クラスタリング
係数 (CC) を凝集性と表記する。なお, 昇格までの期間が
1 日以下の記事は, 評価の間違いや荒らし行為の可能性を
考慮してデータから除外した。また, 協業ネットワークを
構成するユーザが 2 人以下の記事も, ネットワーク指標を
計算することができないため, データから除外した。

ネットワークの構造に関わる指標は, そのネットワーク
サイズに影響を受ける可能性があるため, ネットワークサ
イズをそろえたランダムサンプルを抽出し, 中心性と凝集
性の比較を行うこととした。サンプルの抽出では, 各ネッ
トワークサイズにおける協業ネットワークから 10 件ずつ
のサンプルを取得し, 10 件のサンプルが取得できなくな
った時点でサンプルの取得を中止した。その結果, ネット
ワークサイズが 3 から 25 までのものがサンプルとして抽出
された。これらネットワークサイズが 25 以下のものが各
変化パターンの全対象に占める割合は, それぞれ Start-B
は 78% (5598 件), B-GA は 77% (2138 件), GA-FA は
80% (1106 件) であった。

表 1 記事の質変化パターンごとの協業ネットワーク指標の平均値
と (標準偏差)

Table 1 Mean and (standard deviation) of collaboration net-
works for articles moving up in quality levels.

記事の変化パターン	Start-B	B-GA	GA-FA	p (anova)
N	7175	2786	1379	
昇格までの期間(日)	514.5 (327.2)	248.6 (292.8)	189.9 (205.5)	***
編集ユーザ数	42.7 (60.9)	56.6 (119.1)	41.8 (95.0)	***
協業ネットワーク指標				
N(サンプリング)	230	230	230	
グループ中心性(GDC)	0.347 (0.210)	0.400 (0.223)	0.462 (0.215)	***
クラスタリング係数(CC)	0.234 (0.217)	0.256 (0.212)	0.339 (0.207)	***

*** $p < 0.001$

表 2 昇格記事と非昇格記事間での協業ネットワーク指標の比較
Table 2 Collaboration network metrics of promoted and not
promoted articles.

記事の昇格パターン	B-B	B-GA	p (anova)	GA-GA	GA-FA	p (anova)
N	39883	2786		7600	1379	
経過時間(日)	978.4 (401.6)	248.6 (292.8)	***	766.2 (367.5)	189.9 (205.5)	***
編集ユーザ数	65.29 (102.9)	56.56 (119.1)	***	51.43 (106.3)	41.76 (95.00)	**
協業ネットワーク指標						
N(サンプリング)	350	350		230	230	
グループ中心性(GDC)	0.348 (0.185)	0.398 (0.194)	***	0.397 (0.192)	0.459 (0.208)	***
クラスタリング係数(CC)	0.221 (0.156)	0.258 (0.177)	**	0.305 (0.205)	0.351 (0.214)	*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

分散分析の結果, すべてのパターン間で, 昇格時間, 編
集者数, GDC, CC が異なることが分かった。GDC と CC
の平均値はともに, Start から B までのパスで最も小さく,
B から GA, GA から FA と変化するについて, 値が大き
くなることが分かる。多重分析の結果から, 質変化パスで
の GDC の値は有意 ($p < 0.05$) に異なることが分かった。
一方, GA から FA の変化パスでの CC は他の 2 つのパス
と比較して有意に大きい値であったが, Start から B, B から
GA の間には有意な差は見られなかった。

表 2 に, 昇格した記事と昇格しなかった記事における
協業ネットワークの指標を示す。ここでの経過時間とは,
図 4 の期間 B の時間を表す。昇格しなかった記事は, 打
ち切りデータとして, 1 つ前の昇格からデータ取得時まで
(2010 年 12 月 1 日) までを経過時間とした。ネットワーク
指標に関しては, 表 1 と同様にネットワークサイズを揃え,
各ネットワークサイズから 10 件のランダムなサンプルを
作成して比較を行った。B から GA に昇格したパターンと
B にとどまっているパターンでは, ネットワークサイズが
3 から 37 までのものがサンプルとして抽出された。GA から
FA に昇格したパターンと GA にとどまっているパター
ンではネットワークサイズが 3 から 25 までのものがサン
プルとして抽出された。これらの昇格したパターンでは,

表 3 GA から FA への昇格イベント分析における基本統計量

Table 3 Descriptive statistics between variables for event-history model for FA-Class promotion from GA-Class.

変数	Mean	S.D.	Min.	Max
1. 生存時間(日)	665.51	414.4	3.23	1673
2. 既存協業ネットワーク割合 R_{pc}	0.4253	0.2082	0.042	1
3. グループ中心性(GDC)	0.3976	0.2713	0	1
4. クラスタリング係数(CC)	0.24	0.26	0	1
5. ウィキプロジェクト数	1.164	1.2635	0	15

表 4 B から FA への昇格イベント分析における基本統計量

Table 4 Descriptive statistics and correlation between variables for event-history model for GA-Class promotion from B-Class.

変数	Mean	S.D.	Min.	Max
1. 生存時間(日)	918.75	447.06	1.084	1681
2. 既存協業ネットワーク割合 R_{pc}	0.2834	0.1688	0.022	1
3. グループ中心性(GDC)	0.3741	0.2895	0	1
4. クラスタリング係数(CC)	0.1702	0.2304	0	1
5. ウィキプロジェクト数	0.5121	1.1201	0	23

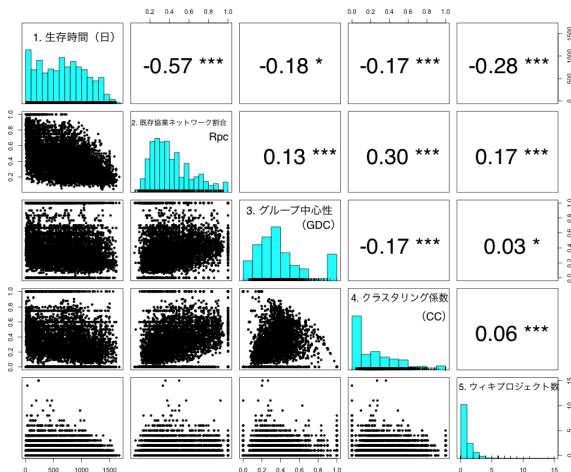


図 5 GA から FA への昇格イベント分析における相関係数と散布図 (* $p < 0.05$, *** $p < 0.001$)

Fig. 5 Correlation and scatter plot between variables for event-history model for FA-Class promotion from GA-Class.

昇格しなかったパターンと比較して、協業ネットワークは中心性が高く、凝集性も高いネットワークとなっていることが分かった。

(2) 既存協業ネットワーク分析

表 3 では、GA から FA への昇格間での時間、表 4 では B から GA への昇格までの時間というイベントに対する生存時間分析に取り入れた指標である、既存協業ネットワークを表す指標の基本統計量(表 3, 表 4)と指標間の相関と散布図(図 5, 図 6)を表す。生存時間とは、1つ前の昇格(GA, B)から次の昇格(FA, GA)までの時間を表している。昇格していない記事に関しては、打ち切りデータとして、表 2 と同様にデータ取得までの時間(2010年12月1日)を生存時間としている。

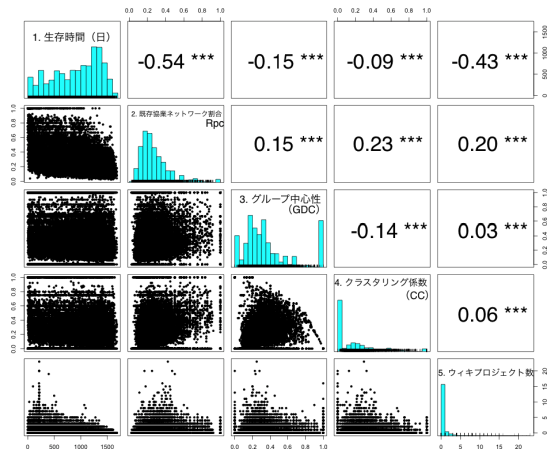


図 6 B から FA への昇格イベント分析における相関係数と散布図(*** $p < 0.001$)

Fig. 6 Correlation and scatter plot between variables for event-history model for GA-Class promotion from B-Class.

図 5 と図 6 では、対角線上に各指標値の分布を表すヒストグラムを表示し、右上部分に変数間の相関係数を、左下部分に変数間の散布図を示している。たとえば3行1列目の散布図は、横軸が生存時間、縦軸がグループ中心性(GDC)の散布図となり、1行3列目が生存時間とGDCの相関係数となる。図 5 と図 6 から、生存時間には大きなはずれ値が存在していないことが分かる。

既存協業ネットワークの特徴を見ると、生存時間と他のすべての指標の間には、既存ネットワーク割合を除いて、弱い負の相関が見られた。負の相関とは、いずれの指標においても、値が高くなることで昇格までの時間が短くなることを示している。

表 5 に Cox 比例ハザードモデルによる生存時間分析結果を示す。それぞれの独立変数の寄与率は標準化するため、独立変数はすべて平均が0、標準偏差が1となるように標準化した。表 5 での $\exp(\text{coef.})$ がハザード比を表しており、その共変量の標準偏差分の増減によって目的変数が変化する割合を表している。すなわち、ハザード比が1よりも大きい場合、標準化した共変量の標準偏差分が増加することにより、ハザード比の割合で記事昇格の確率が増加することを表す。逆にハザード比が1よりも小さい場合、標準化した共変量の標準偏差分が増加することにより、ハザード比の割合で記事昇格の確率が減少することを表す。データ取得期間において、7,900件のGAのうち、1,369記事がFAに昇格した。33,033のBから2,503記事がGAに昇格した。生存時間分析の結果から、既存協業ネットワークの各指標、ウィキプロジェクトの重要度、ウィキプロジェクトの関与数のいずれの指標も、記事昇格というイベントへ寄与していることが示された。

ウィキプロジェクトの重要度では、FAへの昇格、GAへの昇格の両方において、ハザード比がそれぞれ1.273, 1.196と1より大きく、有意にポジティブに寄与していることが

表 5 英語版 Wikipedia を対象とした Cox 比例ハザードモデル
Table 5 Cox proportional hazard model for the English Wikipedia.

目的変数	FA昇格率			GA昇格率		
N	7900			33033		
イベント数	1369			2503		
既存ネットワーク期間	1年			1年		
	exp(coef.)	se(coef.)	p	exp(coef.)	se(coef.)	p
ウィキプロジェクト重要度	1.273	0.0376	***	1.196	0.0292	***
ウィキプロジェクト関与数	0.683	0.0449	***	0.973	0.0295	
既存協調ネットワーク関連指標						
既存ネットワーク割合 R_{pc}	2.532	0.0269	***	2.278	0.0138	***
グループ中心性 (GDC)	1.074	0.0314	*	1.165	0.0201	***
クラスタリング係数 (CC)	1.345	0.0256	***	1.113	0.0173	***

* $p < 0.05$, *** $p < 0.001$

分かった。一方、ウィキプロジェクトの関与数はFAにおいてはハザード比が0.683, 0.973と1より小さく、有意にネガティブに寄与していることが分かった。

既存協業ネットワークに関する指標では、割合を表す R_{pc} がFAとGAへの昇格へ有意に寄与することが分かった。この結果から、記事編集に関わるユーザのうち、すでに関係を保持しているユーザの割合が大きいほど、その記事は早く昇格することを表している。

さらに、GDCとCCもまた、FAとGAにおいて記事の質向上にポジティブに寄与していることが分かった。FAへの昇格を見ると、CCのハザード比が1.345であり、標準化したCCが標準偏差分増加することで、記事昇格の確率が1.345倍大きくなることを表す。GDCは1.074倍となっていることから、CCの方がGDCより記事昇格確率への寄与度が大きいことが分かった。一方、GAへの昇格を見ると、GDCのハザード比(1.165)の方がCC(1.113)よりも大きいことが分かった。すなわち、FAでは中心性よりも凝集性の方が記事昇格への寄与が大きく、GAでは中心性の方が記事昇格への寄与が大きいことを示している。

4.2 日本語版ウィキペディア

(1) 記事編集協業ネットワーク分析

日本語版のウィキペディアでは、記事の質評価がGAとFAのみしか存在していない。また、GAの評価が付与され始めたのが2010年7月からであるため、2010年7月にGAに昇格した記事が多数存在していた。しかし、これらの同時期に昇格した記事は事前にGAの質があるとされていた記事であることが考えられ、この日付がGAの質に到達した日ではない。そこで、これらのデータを除外し、2011年以降にGAに昇格した記事のみを対象データとした。対象とするデータは、着目期間において3回以上編集があるものとした。

GAに昇格後、FAに昇格した記事と(GA-FA)、GAに昇格後、GAから変化していない記事群(GA-GA)を対象

表 6 昇格しなかった群と昇格した群での協業ネットワーク指標
Table 6 Collaboration network metrics of promoted and not promoted articles.

	GA-GA	GA-FA	p
N	79	6	
期間(日)	759.8 (262.9)	195.7 (176.8)	***
編集者数	13.35 (7.65)	11.00 (6.60)	
グループ中心性 (GDC)	0.597 (0.334)	0.480 (0.288)	
クラスタリング係数 (CC)	0.126 (0.259)	0.185 (0.205)	

*** $p < 0.001$

として、図4の期間Bに示す協業ネットワークを抽出を行った。FAに昇格した記事は、GA昇格後FA昇格までの期間、GAから変化していない記事は、観測時点(2014年3月13日)までの期間を期間Bとした。協業ネットワークを構成するユーザが2人以下の記事は、ネットワーク指標を計算することができないため、データから除外した。

表6にGA昇格後、GAにとどまっている79記事とGAからFAに昇格した6記事の指標を示す。期間以外では有意な差は見られなかったが、グループ中心性では昇格しなかった群で、クラスタリング係数では昇格した群で高くなる傾向が見られた。

(2) 既存協業ネットワーク分析

表7にGAからFAへの昇格パスを対象とした既存協業ネットワークを表す各指標の基本統計量と指標間の相関を示す。GAからFAへの昇格期間を生存時間とし、昇格しなかった記事群は、データ取得日(2014年3月13日)で打ち切りとして、データとして含めた。図7より生存時間と既存協業ネットワーク割合、クラスタリング係数、グループ中心性とクラスタリング割合との間には弱い負の相関が見られた。また、生存時間には大きなずれ値は見られなかった。

表 7 GA から FA への昇格イベント分析における基本統計量
Table 7 Descriptive statistics between variables for event-history model for FA-Class promotion from GA-Class.

	Mean	S.D.	Min.	Max
1. 生存時間(日)	534.6	321.5	11.23	1122.4
2. 既存協業ネットワーク割合 R_{pc}	0.403	0.211	0.130	1
3. グループ中心性(GDC)	0.781	0.319	0	1
4. クラスタリング係数(CC)	0.088	0.228	0	1

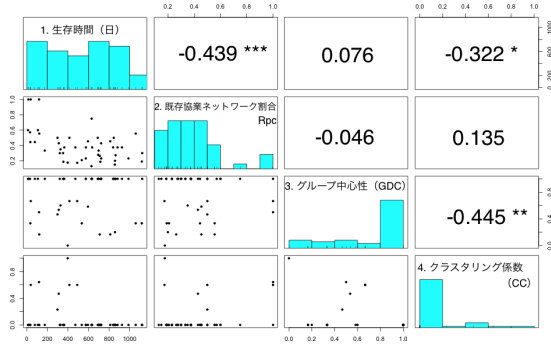


図 7 GA から FA への昇格イベント分析における相関係数と散布図 (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

Fig. 7 Correlation and scatter plot between variables for event-history model for FA-Class promotion from GA-Class.

表 8 日本語版 Wikipedia を対象とした Cox 比例ハザードモデル
Table 8 Cox proportional hazard model for the Japanese Wikipedia.

目的変数	FA昇格率		
N	47		
イベント数	10		
事前協調期間	1年		
	exp(coef.)	se(coef.)	p
既存協業ネットワーク割合 R_{pc}	3.206	0.355	**
グループ中心性(GDC)	0.767	0.360	
クラスタリング係数(CC)	0.839	0.379	

** $p < 0.01$

表 8 に Cox 比例ハザードモデルによる FA 昇格率への既存協業ネットワークの各指標の寄与を示す。これらの指標値も独立変数間の寄与率を標準化するため、平均を 0、標準偏差を 1 となるように変換している。FA 昇格に有意に寄与するのは既存ネットワーク割合のみという結果となった。

5. 考察

5.1 タスクの難易度と協業ネットワーク構造

本稿では、英語版と日本語版のウィキペディアを対象として、1つ前の記事の昇格日から次の昇格日までの期間、すなわち記事の質を1段階向上させるというタスクにおけるコラボレーションを行うユーザー同士の協業ネットワークの抽出を行った。表 1 の結果から、編集作業に関わったユーザー間の協業ネットワークの構造は、記事の質変化のパター

ンによって異なる構造を有していることが分かった。Start から B, B から GA, GA から FA という 3つのパターンの中では、GA から FA のパターンは最も凝集性の高いネットワークを構築しており、中心性も最も高いネットワークを構築していた。B から GA では中心性は中程度の協業ネットワークを構築しているが凝集性は低く Start から B と同程度の凝集性であった。Start から B では中心性が最も低く、凝集性の低い疎なネットワークを構築していることが示唆された。記事評価を行うウィキプロジェクトのガイドからも、FA のような質の高い記事を作成することは、B の記事を作成するのとは比べ、より困難で複雑なタスクであるということが出来る。表 1 から最も難易度が高いタスクでは、編集者間が相互にコミュニケーションを行い、中心性と凝集性の高い協業ネットワークを築いていることが示唆される。中程度の難易度のタスクでは、下位の質の記事と凝集性は同程度であるが、中心性の高い協業ネットワーク、すなわち少数のアクティブなユーザーが他の参加者へコミュニケーションを行っていることが示唆される。

表 2 では、ある記事レベルから次の質へ昇格しているコラボレーションと、昇格していないコラボレーションを比較した。その結果、B から GA へ昇格した記事の協業ネットワーク、より難易度の高い GA から FA へ昇格した記事の協業ネットワークの両方で、昇格していない記事と比較して有意に中心性が高く、凝集性も高い協業ネットワークが得られ、コミュニケーションの中心となるような少数のユーザーの存在とともに、編集に関わるユーザー同士でも密な関係を築いていることが分かった。これは、OSS 開発のコミュニティに見られたオニオンモデル [27] に近い構造であると考えられる。すなわち、オンライン上でのパフォーマンスの高いコラボレーションは、OSS 開発とはタスク内容自体は大きく異なるウィキペディアにおいても、高い中心性と凝集性というユーザー間の関係構造が難易度の高いタスクの遂行と関連することが示唆された。

さらに、これらのタスクの難易度と協業ネットワークの構造の関係は、表 1 で示した記事昇格の 3つのパターンに見られたものと、ある記事レベルから昇格した記事としなかった記事を比較した表 2 の結果の両方において一貫した結果となっている。この結果からも、難易度の高いタスクを成功させるには、中心性と凝集性を持った協業ネットワークの構築が有効であることが示唆された。

一方、表 6 に示した日本語版のウィキペディアでは、FA の数が少なく、有効な協業ネットワークが抽出できた件数が少なかったため、有意な差は見られなかった。これは、英語版と日本語版での協業ネットワークへの活動量の違いとも関係していると考えられる。英語版のウィキペディアでの協業ネットワークは、ユーザーノートページへの編集数の増加率という観点から、2003年から2005年の間で78倍増加していた [36]。一方、日本語版では図 1 に示したとお

り、2011年頃になって記事ノートページよりも、ユーザノートページに対する編集が上回る傾向が見られてきている。このように、日本語版のウィキペディアでは、ユーザ間の関係の構築が英語版と比較して遅れていることも考えられる。しかし、協業ネットワークへの活動が増加している傾向からも、今後日本語版ウィキペディアでユーザ間の社会ネットワークの構築が進んでいくと考えられる。

5.2 質向上時間への既存協業ネットワークの寄与

BやGAレベルに昇格した記事を、さらにGAやFAといった1段階上の質に向上させる編集作業に着目し、その作業効率の指標として、昇格までの時間を目的変数とした生存時間分析を行った。独立変数として、編集に関わったユーザ間の既存協業ネットワークの指標を用いた。加えて、英語版ウィキペディアではウィキプロジェクトによる指標としてウィキプロジェクトによる重要度の評価と、関与しているウィキプロジェクトの数を変数に加えた。

表5に示したとおり、英語版ウィキペディアでは、BからGA、GAからFAに記事の質を向上させる時間は、ウィキプロジェクトの重要度が両方のケースで大きくポジティブに寄与していることが分かった。すなわち、ウィキペディアの編集者は、ウィキプロジェクトが定めた重要度の高い記事により注力していることが示唆される。一方、ウィキプロジェクトの関与数はFAにおいて有意にネガティブに寄与していることが分かった。ウィキプロジェクトの関与数が多いということは、その記事内容が様々な分野にまたがっている、すなわち複雑な記事であることが推測される。そのため、より多くのウィキプロジェクトが関与するほど、ユーザが記事の質を向上させるのが難しくなっていることが考えられる。GAに昇格した記事では関与したウィキプロジェクト数による影響は見られなかった。

記事編集に関わったユーザ間の既存協業ネットワークの指標では、編集開始以前にすでに関係を形成しているユーザの割合が大きいほど、その記事は早く昇格することが示された。つまり、既存の協業ネットワークで示される社会ネットワークは、記事の質を早期に高めるというタスクにおいて、ユーザの生産性の向上に寄与することを示唆している。

既存協業ネットワークの構造では、FAでは中心性よりも凝集性の方が記事昇格への寄与が大きく、GAでは中心性の方が記事昇格への寄与が大きいことが分かった。これらの結果は表1と表2で示した結果と同様な傾向を表している。すなわち記事編集集中の協業ネットワークと、記事編集に関わる以前の既存の協業ネットワークの両方において、中心性の高い組織構造に加えて凝集性の高い関係構造が複雑なタスクにおいてパフォーマンスに関連することを示唆している。GAからFAに昇格させるという複雑なタスクに関与することで、凝集性の高い関係構造が築かれた

だけではなく、そのような関係構造が事前に存在することが、複雑なタスクにおいて有効であることが示唆された。さらに、これは表2の結果と合わせて、タスクの難易度や複雑さに応じて協業ネットワークのパターンが異なる点において一貫した傾向があることを示唆している。

表8では、日本語版のウィキペディアの結果から既存協業ネットワークの記事昇格の期間への寄与を示している。GAからFAに昇格した記事のうち、有効な既存協業ネットワークが形成された10記事を対象とした結果、既存協業ネットワーク割合、すなわち記事編集に関わる以前から関係性のあるユーザが多いほど、記事昇格の期間が短くなることが分かった。一方、既存協業ネットワークの構造に有意な寄与は見られなかった。既存協業ネットワーク割合の寄与は英語版のウィキペディアでも同様な傾向が見られており、日本語ウィキペディアコミュニティにおいても、ユーザ間に存在する関係性、すなわち社会ネットワークはグループコラボレーションにポジティブに寄与することが示唆された。

これらの結果はKitturら[33]が明らかにした「潜在的なコーディネーション」との関連性も示唆される。記事ノートページへ編集方針などを記述せずに、複雑な記事編集をこなす特定のユーザ間の潜在的コーディネーションの背景には、本稿で示したような社会ネットワークが築かれていた可能性が考えられる。事前に社会ネットワークがあることで、記事作成でのコーディネーションコストを低減させ、早期に記事の質向上をはかることができたと考えられる。

6. 結論と今後の課題

本稿では、インターネット上のオープンなコラボレーションによる集合知の事例の1つであるウィキペディアを対象とし、記事編集に関わるユーザ間の社会関係資本の前提となる社会ネットワークがグループコラボレーションへ与える効果を分析した。ユーザ間の関係の定義を、ウィキペディアユーザ同士が直接コミュニケーションできるユーザノートページへの書き込みとした。社会ネットワーク分析の手法を用いてユーザ間の関係構造を抽出し指標化した。グループコラボレーションの対象の記事の質を向上させるタスクとし、タスクの難易度や遂行スピードをコラボレーションの質の測定指標とした。その結果、ユーザ同士の関係構造、すなわちユーザ間の社会ネットワークが記事の質向上というグループコラボレーションタスクのパフォーマンスを向上させることが示唆された。

今後は、ウィキペディアで見られた社会ネットワークが、どのようにユーザ間で構築され発達してきたかといったプロセスを明らかにしていきたい。さらに、記事ごとではなく、複数の記事群を横断して活動するユーザグループの活動から、より広範囲に存在する社会ネットワークについて

も明らかにしていきたい。

今回得られたユーザ間の社会ネットワークとグループパフォーマンスの関係が、ウィキペディア以外のコラボレーション環境において一般化できるかは、検証していく必要がある。今後増加していく多様で複雑な社会的課題に対し、オンライン・分散環境のオープンなコラボレーションによって英知を結集し解決を図っていくためにも、集合知を実現するコラボレーション環境におけるユーザ間の社会ネットワークの役割と効果を明らかにしていき、様々な組織やコミュニティ環境での分散コラボレーションの支援につなげていきたい。

謝辞 論文作成にあたり、有益なコメントをいただいた Stephanie Woerner 氏に、謹んで感謝の意を表す。

参考文献

- [1] Giles, G.: Internet encyclopaedias go head to head, *Nature*, Vol.438, 7070, pp.900-901 (2005).
- [2] Pfeil, U., Zaphiris, P. and Ang, C.A.: Cultural Differences in Collaborative Authoring of Wikipedia, *Journal of Computer-Mediated Communication*, Vol.12, pp.88-113 (2006).
- [3] Bourdieu, P.: The forms of capital, *Handbook of Theory and Research for the Sociology of Education*, Richardson, J. (Ed.), pp.241-258, New York, Greenwood (1986).
- [4] Putnam, R.D.: Bowling Alone: America's Declining Social Capital, *Journal of Democracy*, Vol.6, No.1, pp.65-78 (1995).
- [5] Boden, A., Nett, B. and Wulf, W.: Trust and Social Capital: Revisiting an Offshoring Failure Story of a Small German Software Company, *Proc. ECSCW'09*, pp.123-142 (2009).
- [6] パットナム, R.D. (著), 河田潤一 (訳): 哲学する民主主義—伝統と改革の市民的構造, NTT 出版 (2001).
- [7] パットナム, R.D. (著), 柴内康文 (訳): 孤独なボウリング—米国コミュニティの崩壊と再生, 柏書房 (2006).
- [8] パート R.S. (著), 安田 雪 (訳): 競争の社会的構造—構造的空隙の理論, 新曜社 (2006).
- [9] Borgatti, S.P., Jones, C. and Everett, M.G.: Network Measures of Social Capital, *CONNECTIONS*, Vol.21, No.2, pp.27-36 (1998).
- [10] リン, N. (著), 筒井淳也ほか (訳): ソーシャル・キャピタル—社会構造と行為の理論, ミネルヴァ書房 (2008).
- [11] Worldbank. Measuring Social Capital, available from (<http://go.worldbank.org/A77F30UIX0>) (2014年4月12日取得).
- [12] Van Der Gaag, M., Tom, A.B. and Snijders, T.B.A.: The Resource Generator: social capital quantification with concrete items, *Social Networks*, Vol.27, pp.1-29 (2005).
- [13] Krackhardt, D. and Hanson, J.R.: Informal Networks: The Company Behind the Chart, *Harvard Business Review*, July-August (1993).
- [14] Cummings, J.N. and Cross, R.: Structural properties of work groups and their consequences for performance, *Social Networks*, Vol.25, No.3, pp.197-210 (2003).
- [15] Aral, S., Brynjolfsson, E. and Van Alstyne, M.: Information, Technology, and Information Worker Productivity, *Info. Sys. Research*, Vol.23, 3-Part-2 (2012).
- [16] Anthony, D., Smith, S.W. and Williamson, T.: Explaining Quality in Internet Collective Goods: Zealots and Good Samaritans in the Case of Wikipedia, Working Paper, Department of Computer Science, Dartmouth College, November (2005).
- [17] Benkler, Y.: *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, Yale University Press, New Haven and London (2007).
- [18] Howe, J.: The Rise of Crowdsourcing, *Wired* (14:06) (June 2006).
- [19] Gloor, P.: *Swarm Creativity: Competitive Advantage Through Collaborative Innovation Networks*, Oxford University Press, New York (2005).
- [20] Malone, T. Laubacher, R. and Dellarocas, C.: The Collective Intelligence Genome, *Sloan Management Review*, Vol.51, No.3, Spring (2010).
- [21] Ahuja, M.K. and Carley, K.M.: Network Structure in Virtual Organizations, *Organization Science*, Vol.10, No.6, pp.741-757 (1999).
- [22] Cummings, J.N. and Cross, R.: Structural properties of work groups and their consequences for performance, *Social Networks*, Vol.25, No.3, pp.197-210 (2003).
- [23] Porter, K.A., Bunker Whittington, K.C. and Powell, W.W.: The institutional embeddedness of high-tech regions: Relational foundations of the Boston biotechnology community, *Clusters, networks, and innovation*, Breschi, S. and Malerba, F. (Eds.), pp.261-294, Oxford University Press (2005).
- [24] Uzzi, B. and Spiro, J.: Collaboration and Creativity: The Small World Problem, *American Journal of Sociology*, Vol.111, No.2, pp.447-504 (2005).
- [25] von Hippel, E.: Innovation by User Communities: Learning From Open-Source Software, *Sloan Management Review* (2001).
- [26] Lakhani, K.R. and von Hippel, E.: How open source software works: "Free" user-to-user assistance, *Research Policy*, Vol.32, No.6, pp.923-943 (2003).
- [27] Crowston, K. and Scozzi, B.: Coordination practices for bug fixing within FLOSS development teams, *Proc. 1st International Workshop on Computer Supported Activity Coordination (CSAC)*, Portugal (2004).
- [28] Aberdour, M.: Achieving Quality in Open Source Software, *IEEE Software*, January-February (2007).
- [29] Kidane, Y. and Gloor, P.: Correlating temporal communication patterns of the Eclipse open source community with performance and creativity, *Computational & Mathematical Organization Theory*, Vol.13, No.1, pp.17-27 (2007).
- [30] Anthony, D., Smith, S.W. and Williamson, T.: Explaining Quality in Internet Collective Goods: Zealots and Good Samaritans in the Case of Wikipedia, Working Paper, Department of Computer Science, Dartmouth College, November (2005).
- [31] Adler, B.T. and de Alfaro, L.: A content-driven reputation system for the Wikipedia, *Proc. 16th International Conference on the World Wide Web*, pp.261-270 (2007).
- [32] Priedhorsky, R., Chen, J.K., Lam, S.K., Panciera, K., Terveen, L. and Riedl, J.: Creating, Destroying, and Restoring Value in Wikipedia, *ACM GROUP'07*, November 4-7, 2007, Sanibel Island, Florida, USA. (2007).
- [33] Kittur, A. and Kraut, R.E.: Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination, *Proc. ACM 2008 Conference on Computer Supported Cooperative Work—CSCW '08*, ACM Press, New York (2008).

- [34] Wilkinson, D. and Huberman, B.: Cooperation and Quality in Wikipedia, *WikiSym'07* (2007).
- [35] Liu, J. and Ram, S.: Who Does What: Collaboration Patterns in the Wikipedia and Their Impact on Data Quality, *ACM Trans. Manage. Inf. Syst.* Vol.2, No.2, Article 11 (2011).
- [36] Viégas, F.B., Wattenberg, M., Kriss, J. and Van Ham, J.: Talk Before You Type: Coordination in Wikipedia, *Proc. 40th Hawaii International Conference on System Sciences* (2007).
- [37] Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J. and Suri, S.: Feedback Effects between Similarity and Social Influence in Online Communities, *ACM KDD'08* (2008).
- [38] Wasserman, S. and Faust, K.: *Social Network Analysis: Methods and Applications*, Cambridge, Cambridge University Press (1994).
- [39] Watts, D.J. and Strogatz, S.H.: Collective dynamics of 'small-world' networks, *Nature*, Vol.393, 440-4 (1998).
- [40] Cox, D.R.: Regression Models and Life Tables, *Journal of the Royal Statistical Society Series B*, Vol.34, No.2, pp.187-220 (1972).



根本 啓一 (正会員)

慶應義塾大学大学院。1978年生。2001年慶應義塾大学理工学情報工学科卒業。2003年同大学大学院理工学研究科修士課程修了。同年富士ゼロックス株式会社入社。2009年MIT Center for Collective Intelligence 客員研究員。2011年より富士ゼロックス株式会社にて、集合知を促進するコラボレーションメディア・システムに関する研究に従事。2014年4月より慶應義塾大学大学院理工学研究科後期博士課程在学。ACM 会員。



ピーター グロア

MIT Center for Collective Intelligence 研究員として協調イノベーションネットワークの研究に従事。ソーシャルメディアソフトウェア会社、Garaxyadviro のチーフ・クリエイティブ・オフィサー。ケルン大学名誉教授，アアルト大学講師，Catolica Santiago de Chile 大学客員教授。



ロバート ローバッカー

MIT Center for Collective Intelligence エグゼクティブディレクター・研究員として集合知の構成要素を成す集合知ゲノムプロジェクト，気候変動に対する創造的なアイデア創出をクラウドソースする仕組みである Climate CoLab プロジェクトに従事。ノースウェスタン大学卒業，ハーバード大学修士・博士課程修了。



岡田 謙一 (フェロー)

慶應義塾大学。慶應義塾大学理工学部情報工学科教授，工学博士。専門は，CSCW，グループウェア，HCI。情報処理学会理事，学会誌編集主査，論文誌編集主査，GN研究会主査，IE 領域委員長，日本 VR 学会理事等を歴任。現在，情報処理学会論文誌：デジタルコンテンツ編集長，DCC 研究会運営委員，電子情報通信学会 HB/KB 幹事長。情報処理学会論文賞（1996，2001，2008），情報処理学会 40 周年記念論文賞，IEEE SAINT '04，ICAT '07 最優秀論文賞等を受賞。情報処理学会フェロー，日本 VR 学会フェロー，IEEE，ACM，電子情報通信学会各会員。