

ウェブ掲示板を対象としたネットワークコミュニティ分析支援システム：CMINER

井上 智雄[†] 小林 哲郎^{††} 池田 謙一^{††}
重野 寛^{†††} 岡田 謙一^{†††}

情報ネットワークの普及により、ネットワークコミュニケーションが一般化し、ネットワークコミュニティが形成されてきた。これにともない、ネットワークコミュニティの社会科学的研究がさかになってきている。本論文で説明するネットワークコミュニティ分析支援システム CMINER は、ネットワークコミュニティの典型的メディアの 1 つであるウェブ掲示板を対象として、その社会科学的研究を支援するソフトウェアシステムである。本システムにより、ウェブ掲示板のログデータを収集し、ログ分析の基本的指標に加えて、基本的な社会ネットワーク分析を自動的に一貫して行うことができ、その結果はグラフにより視覚化できる。従来ソフトウェアとの比較実験による作業の効率化も確認した。

A System to Support Network Community Analysis for Web BBS: CMINER

TOMO'O INOUE,[†] TETSURO KOBAYASHI,^{††} KEN'ICHI IKEDA,^{††}
HIROSHI SHIGENO^{†††} and KEN-ICHI OKADA^{†††}

The spread of information networks has made computer-mediated communication common, which has brought various network communities. Social scientific research on network communities has been increasingly conducted consequently. A network community analysis system named CMINER presented in this paper is a software system to support social scientific analysis on network communities that deals with Web bulletin boards. CMINER is able to get Web BBS logs and then is continuously able to calculate basic social network analysis measures as well as basic log analysis measures. It is also able to visually present the results of analysis in graphs. Task efficiency compared with the use of current software is also investigated.

1. はじめに

インターネットが一般の人々にまで普及し、PC や携帯電話などの情報機器を用いて情報ネットワークごとにコミュニケーションすることが一般的になってきた現在、ネットワークコミュニティが注目を集めている。コミュニティという用語は様々な定義され^{1),2)}、また、ネットワークコミュニティという用語も様々な定義され使われている^{3),4)}。ここでは、情報ネットワークを利用してコミュニケーションする不特定多数の集団を

ネットワークコミュニティとし、典型的なネットワークコミュニティのメディアの 1 つとしてウェブ掲示板を研究の対象とする。本論文では、ウェブ掲示板を対象としたネットワークコミュニティ分析支援システム CMINER (Community MINER) について述べる。

ネットワーク社会の現出にともない、社会学、社会心理学などの社会科学は従来の社会だけでなくネットワーク社会を研究対象に加えてきた⁵⁾。1 対 1 のコミュニケーションのほかに様々な規模・性質の集団におけるコミュニケーションが対象であり、電子会議室、電子掲示板、フォーラム、ニュースグループといったネットコミュニティを扱う例が多い^{5)~7)}。この理由として、金子は『ネットワークを分析し、その動きをとらえるには、1 人の人、1 つの企業、1 つの組織を対象にするのではなく、ゆるやかに伸縮する関係の集合体——それをここではコミュニティと呼ぶ——をユニット(単位)として考えなくてはならない。ネットワークでは「つながる」ことこそが存在の本質だからである』と

[†] 筑波大学図書館情報学系

Institute of Library and Information Science, University of Tsukuba

^{††} 東京大学大学院人文社会系研究科

Department of Social Psychology, The University of Tokyo

^{†††} 慶應義塾大学理工学部情報工学科

Department of Information and Computer Science, Faculty of Science and Technology, Keio University

述べている⁷⁾。

また、CSCW(コンピュータによる協調作業支援)研究では、当初比較的流動性の低い職場の作業グループなどを対象にすることが多かったが、最近の仕事では、協調作業は組織の構造を反映した固定的なグループでなされるよりもむしろ、各自の個人的人脈を通じてなされる場合が増加していることが示されている⁸⁾。このこともネットワークコミュニティが注目される理由になっていると考えられる。

ネットワーク社会、ネットワークコミュニティに関する研究は、人々の活動の場がネットワーク上に広がるにつれて活発化しており、今後ますます重要かつ活発になるものと予想される。これまでのネットワークコミュニティの社会科学的研究では、研究に必要なデータは通信ログをはじめとして基本的に人手で調査分析してきたが、それらデータを適切に情報処理するソフトウェアシステムにより、ネットワークコミュニティの特徴量抽出・類型化を通じてその性質を明らかにするネットワークコミュニティマイニングのような研究の支援が可能である。

本研究では、ネットワークコミュニティの様相を分析するための道具を開発し、その適用によりネットワークコミュニティの分析を行い、その結果をフィードバックしてさらに道具を改良するという方法を通じて、汎用性の高いネットワークコミュニティマイニングシステムを構築することを目標としている。本論文では、その最初のシステムについて述べる。以下の章では、研究背景と関連研究の面からその意義を述べ、分析方法論とシステムについて順に述べる。

2. ネットワークコミュニティ支援

ネットワークコミュニティ研究は主として情報工学、社会科学を専門とする双方の研究者が、ある程度相互に影響を与え合い、また独立にも進めてきた複合的な研究分野である。

情報工学領域においては、分散情報処理、モバイルコンピューティング、ユビキタスコンピューティング、通信ネットワーク、エージェントなどの技術的研究とともに、その社会的応用としてコミュニティ形成支援、コミュニティ活動支援、コミュニティ管理支援などが研究されている。

コミュニティの活性化を目的としたパブリック・オピニオンチャンネルという放送型コミュニティの研究では、コミュニティによって個人の体験に基づく話を共有し、メンバが相互にそれらの話を編集しあうことによる知識共有支援が提案された^{9),10)}。

梅木らはネットワークコミュニティ形態と形成過程について考察し、ネットワークコミュニティの形成過程における異なるポイントを支援する2つのシステムを提案している。1つはコミュニティの発生期に対するもので、ウェブページのブックマークの協調フィルタリングによるページ推薦システムに、関心を共有するユーザ間のコミュニケーションチャネルを加えることによりコミュニティの発生と成長の促進を狙っている。もう1つはすでに存在するネットワークコミュニティにおける意思決定機能、信用を利用した自治機能を提供する。メッセージに対してユーザが投票行動をとり、その信頼度の評価に従ってメッセージを投稿したユーザにコミュニティ運営の権限を与えることで、協調的信頼モデルによるコミュニティの安定化を図っている¹¹⁾。

Community Organizer は、コミュニティの形成の初期段階を「眺める」「見つける」「近づく」「会話する」の4つのフェーズにモデル化し、人々の興味や関心の近さを2次元空間上のユーザアイコンの位置により表示する、コミュニティ形成支援システムである。関心の近いユーザをリスト表示した場合に比べ、より強いコミュニティ感覚を提供することが確認されている¹²⁾。

ネットワークコミュニティ支援の研究はグループウェアや情報システムの研究分野を中心としてここにあげたほかにも活発に行われつつあるが、ある特定のコミュニティを対象としてその可視化をしたり、その活性化を狙ったりするものが多く^{13),14)}、多数のネットワークコミュニティの分析を支援する試みはほとんど見られない。

3. ネットワークコミュニティ分析

社会科学領域においては、情報ネットワークの出現とともにCMC(コンピュータを媒介としたコミュニケーション)が研究されてきた。2章に述べたネットワークコミュニティ支援では分析という視点はあまり見られないが、社会科学では分析が主体であり、ネットワークコミュニティ分析につながっている。

CMC研究では当初、対面ではない対人コミュニケーションという特徴から、その差異に焦点が当てられてきたが、コンピュータと情報ネットワークがより普及するにつれて、確立された社会の一部としてネットワーク社会をとらえ、その特徴、様相を分析する風潮が起こった¹⁵⁾。インターネット普及以前のパソコン通信ネットワークを対象とした研究においても、電子会議室という新たなコミュニティについて研究されている⁵⁾。インターネットの爆発的普及がウェブの発達と

ともに起こると、ネットワークコミュニティ研究が本格化し、パソコン通信の電子会議室を対象とした社会調査手法による分析⁶⁾、内容分析を主とした質的アプローチに加え、ログの数量的分析、社会ネットワーク分析もなされるようになってきた⁷⁾。また、チャットやウェブ掲示板も対象となってきた^{16),17)}。

ネットワークコミュニティの分析を支援するシステムとしては、Netscan がある¹⁸⁾。ネットニュースを収集し分析することができるシステムだが、Usenet しか扱うことができず、また計算できる指標は従来のログ解析でよく使われてきた基本的なものだけである。一方、近年はウェブ掲示板によるネットワークコミュニティが増加しておりその分析が望まれる。また、ネットワークコミュニティ分析では数理社会学で発達した手法である、社会ネットワーク分析がよく利用されるようになってきた。ウェブ掲示板のデータを取得してネットワーク分析を行うことは考えられても、そのような分析支援システムがなく、手間との兼ね合いでまだ研究が進んでいない。CMINER はまずこの点をターゲットにして開発した。

4. 社会ネットワーク分析

4.1 社会ネットワーク分析の概要

社会ネットワーク分析は、組織内のインフォーマルなコミュニケーションや企業間取引などの様々な関係構造を定量的に評価する手法である¹⁹⁾。一般に、ある事象の関係構造を点と紐帯のグラフ（ソシオグラム）として抽象化して表現する。そして、ソシオグラムを特徴付けるデータ（ネットワークデータ）を行列を用いて記述、計算することにより分析する。

社会ネットワーク分析では様々なコミュニケーション構造指標が使用される。指標には、ネットワークサイズ、次数（入次数、出次数）、ネットワーク密度、直接結合度、中心性、集中化、などがあるが、これらのほかに多くの指標が提案されている²⁰⁾。

4.2 ネットワークコミュニティ分析への社会ネットワーク分析の利用

ネットワークコミュニティ分析では客観性・信頼性のあるデータを大量に時系列的に収集できるという利点があり、社会ネットワーク分析の適用は自然である。ログデータを元に行列計算するのは人手では既存のツールを使っても手間であるため従来はあまり研究されていなかったが、近年限定的なデータに対していくつかの例を見ることができる。

北山は、ニフティサーブの電子会議室であるフォーラムを対象にして密度や排他性、パワーなどの指標を

使ってフォーラム内のコミュニケーションのされ方を明らかにし、フォーラム間で比較検討した²¹⁾。

高橋らは、ネットワーク分析をメーリングリストに適用し、抽出された組織構造的指標を視覚化してネットワークコミュニティの内外にフィードバックすることで、組織アウェアネスの生成を支援した²²⁾。

藤田らは、サイバーコミュニティの発生・形成過程における人々の活動を支援するシステムの有効性を、実験の結果のネットワーク分析により評価した²³⁾。

福原らは、コミュニティの知識共有を目的とした話の共有システムのサブシステムとして、ネットワーク分析に基づき投稿されたメッセージ間の関係を分析するツールを開発した。指定されたキーワードを含むメッセージ集合のリンク構造と、メッセージ間の参照関係を元にしたリンク構造の表示の機能を有する^{10),24)}。

4.3 社会ネットワーク分析ソフトウェア

UCINET は最もよく知られている社会ネットワーク分析ソフトウェアの 1 つである。入力データとしてデータエディタによる直接データ入力や CSV ファイルの取り込みが可能であるが基本的に行列データを使用する。多数のネットワーク指標の算出と関連する統計処理を行うことができる。分析結果はログファイルとして出力する²⁵⁾。

STRUCTURE は MS-DOS 上のフリーソフトウェアで、行列データに対して ASSISTANT というソフトウェアで分析コマンドを指定して実行すると分析結果が別ファイルに出力される²⁶⁾。

これら以外にもグラフを視覚化するソフトウェアや、それらを組み合わせたものがある²⁷⁾、いずれもデータ入力は社会ネットワーク分析に必要な行列データ形式からである。

5. ネットワークコミュニティ分析支援システム：CMINER

前章までに述べた関連研究と関連研究分野のニーズに基づいて、著者らは本章で説明するネットワークコミュニティ分析支援システム CMINER を開発した。

5.1 システム要件

ネットワークコミュニティを支えるメディアは、電子メール、メーリングリスト、インスタントメッセージングなど多様だが、その 1 つであるウェブ掲示板を対象としたシステムを開発した。本システムは、ネットワークコミュニティ分析を行っている、あるいは行いたい社会科学分野の専門家を対象とし、これら分野におけるネットワークコミュニティ研究、ネットワーク社会研究を加速することがねらいの 1 つである。本

システムが提供する機能は、分析作業のフェーズに対応して、データ収集、データ分析、コミュニティ比較の3つに分かれる。

5.1.1 データの収集

社会科学における従来研究では掲示板ログは運用者が提供していたが、ウェブ掲示板の場合はネットワーク上に HTML データとして公開されているため、必要に応じ許可を得ることを別にすればデータを機械的に収集することができる。分析支援という点では、大量のウェブページを容易に収集できる仕組みが必要である。

分析に必要なデータはウェブページ全体ではなく、ウェブ掲示板ページの掲示板の内容である。ウェブ掲示板ページの HTML データは、そのタグ構造が掲示板ごとに異なるため、統一的なログデータを抽出するためには、それら掲示板ごとに対応する必要がある。

5.1.2 データの分析

メッセージ数など従来研究でよく使われてきた指標と社会ネットワーク分析の指標が、本システムのデフォルトで利用できるが、これら以外の指標も考えられる。分析者によって新たな指標が生まれることも予想される。このため、新たな分析指標の追加などができるだけ容易に行える仕組みが望ましい。

先に述べたように、行列データを入力とする社会ネットワーク分析ソフトウェアが存在する。しかし、生のログデータをそのままでは利用できず、行列データへの変換作業が必要である。この作業はデータ量が多いと困難である。収集したデータからのデータ加工を自動化することが支援システムには求められる。

5.1.3 コミュニティ比較

本システムではウェブ掲示板を対象とし、コミュニティとはある1つのトピックについてのスレッドを指すものとする。1つのスレッドには複数のコメントチェーンが含まれる。ここでコメントチェーンとは、あるメッセージを起点として、それに対する返信メッセージの連鎖を指す。

コミュニティの比較を行うためにはそれぞれのコミュニティを特徴づける指標値を比較することが基本となる。そのためには、まず比較対象を様々な条件によって選択するフィルタリング機能が必要である。

5.2 システム構成

本システムはすべて Windows2000 上で実装されている。システムの実装には Java 言語を使用し、システムのメイン部分はプラットフォーム非依存に動作する Java アプリケーションで作成した。また、掲示板からデータを取得するスクリプトプログラムには Perl を利

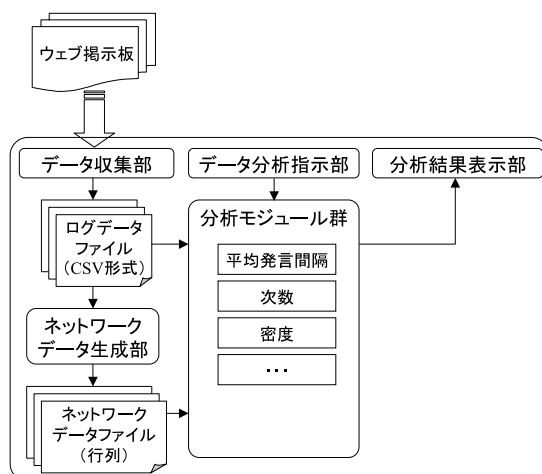


図 1 システム構成

Fig. 1 System architecture.

用した。具体的には、Windows 上で動作する Perl インタープリタに ActivePerl 5.8.0 build 804 を、Java 環境に J2SE 1.4.1 を利用した。

本システムの全体の構成を図 1 に示す。システムは HTTP プロトコルにより、ウェブ上の掲示板からデータを取得する。システムの各部分の役割は以下のとおりである。それぞれの詳細は後述する。

5.2.1 データ収集部

指定したウェブ掲示板ページから掲示板ログデータを収集し、分析に必要な情報を抽出する処理と、個々の掲示板ログデータに整理のための情報を加える処理を行う。

5.2.2 ネットワークデータ生成部

データ収集部分によって収集された掲示板ログデータから、発言内容などを削除し、メンバ同士のコミュニケーションの関係構造のみを記述したネットワークデータを表す行列に変換する。ログデータをネットワークデータに変換することで社会ネットワーク分析が行えるようにする。

5.2.3 データ分析指示部

ユーザからの指示入力インタフェースの主要部で、取得されたデータの分析方法について指示を行う。具体的には分析対象の選択、分析指標の選択、分析結果の表示方法を指示する。実際の分析指標の算出は、次に説明する分析モジュール群によって行われている。

5.2.4 分析モジュール群

取得されたデータから各分析指標値の計算を行うプログラムモジュール群である。本システムでは、5.1.2 項の要件である分析指標の変更への対応として、分析指標ごとにプログラムのモジュール化を行うことによ

り、新たな分析指標の追加・修正などの要求に応えやすくしている。個々の分析モジュールは指標の種類によって、掲示板ログデータかネットワークデータのどちらか一方を利用する。分析対象に指定されたコミュニティのデータを入力とし、指標値を出力する。

5.2.5 分析結果表示部

分析モジュールの結果を、視覚的に表示する。複数の分析モジュールの出力値を用いて、2次元グラフなどを生成する。

5.3 ウェブ掲示板ログデータ収集

本節では、ウェブ掲示板ページからメッセージなどの掲示板ログデータを取得する方法とその保存形式について述べる。

ウェブ掲示板ページにはログと無関係なデータも多く含まれているため、そのままの状態ではログデータとして扱うことができない。ページデータからログデータだけを取り出す必要があるが、ウェブ上には多くの種類の掲示板が存在しており、掲示板ごとにフォーマットが異なる。いずれのページにも名前や題名など、ログデータとして必要な情報が含まれるが、単一の方法ではうまく切り出せない。一方、掲示板の種類を特定すれば、その範囲ではフォーマットは同一なので、HTMLデータからログデータの抽出操作を行うことができる。

この問題に対し、掲示板ログデータ収集部分は、各種掲示板別の複数のスクリプトをプラグインとして用意し、それらを選択使用するという方法が考えられる。本システムではこの方法を実現するために、フリーソフトウェアである INCM 用プラグインの一部を利用している。INCM は掲示板巡回ソフトの 1 つで、ウェブ掲示板との情報交換フォーマットとして CMT フォーマットと呼ぶテキストファイルの形式を提唱して使用している²⁸⁾。そして、ウェブ掲示板からこのフォーマットでログデータを収集するために多数の掲示板に対してプラグインが提供されている^{29),30)}。プラグインは、ウェブ掲示板ページの HTML データを HTTP プロトコルで読み込み、特定名のファイルに CMT フォーマットでその情報を書き出す。本システムではウェブ掲示板とのログ収集インタフェースにこのプラグインを利用している。これにより 5.1.1 項の要件である複数掲示板データに対応でき、必要に応じたプラグインの追加による拡張性を有する。プラグインの出力ファイルから本システムで利用するファイル形式への変換部分は 1 種類である。システムでどのプラグインを使用するかはユーザが判断し選択する。

本システムで以後利用するログデータは、CSV 形

式で 1 つのスレッドにつき 1 つのファイルとして保存している。1 発言 1 行でこの順で次の 6 種類の項目を含む。

1. 発言番号：スレッドの最初の発言から順に通し番号を付ける。個々の発言は一意に識別される。
2. 返信先番号：ある発言が別の発言に対する返信である場合、返信先の発言番号を格納する。発言番号と返信先番号から、発言間の関係構造が分かる。発言が返信でない場合は、“*”を格納する。
3. 発言時間：発言時刻を“西暦年/月/日,時:分”という形式で格納している。時系列的分析に有効な情報である。
4. 題名：発言の題名を表す文字列を格納する。
5. 発言者名：発言者名、いわゆるハンドルネームを表す文字列を格納する。
6. 発言内容：発言本文を格納する。

ログデータファイルのファイル名は、収集された順番で 4 桁の通し番号としている。また、この 4 桁数字のファイル名と対応させた形で、スレッドの説明情報(トピック名)およびフィルタリングのための付加情報(カテゴリ名)を 1 行として全ログデータファイル分をまとめ、説明ファイルとして CSV 形式で別ファイルに書き出している。

5.4 データ分析指標

本節では、本システムでデータ分析に利用できる指標について説明する。指標には従来から CMC 研究で用いられてきたものを含む基本指標と、発言間の関係構造に関わる社会ネットワーク分析指標がある。

5.4.1 基本指標

以下に述べる指標は、ログデータから計算される。平均発言間隔

隣り合う個々の発言の平均時間間隔である。総発言数を N 、発言番号 i の発言時間を T_i とすると、平均発言間隔 $average(t)$ は式 (1) により求められる。

$$average(t) = \frac{\sum_{i=1}^{N-1} T_{i+1} - T_i}{N - 1} \quad (1)$$

この値が大きい場合は、個々の発言の間隔が長く議論が活発でないといえ、この値が小さい場合は、個々の発言の間隔が短く議論が活発であるといえる。この指標からコミュニティの発言の活発さが読み取れると考えられる。

発言間隔標準偏差

発言間隔のばらつき具合を示す。発言間隔標準偏差 $sd(t)$ は式 (2) から求められる。

$$sd(t) = \sqrt{\frac{\sum_{i=1}^{N-1} \{(T_{i+1} - T_i) - average(t)\}^2}{N-1}} \quad (2)$$

平均発言量

発言 1 つあたりの文字数の平均である．各発言 i における文字数を c_i とすると，平均発言量 $average(c)$ は式 (3) で求められる．

$$average(c) = \frac{\sum_{i=1}^N c_i}{N} \quad (3)$$

発言量標準偏差

1 発言あたりの文字数のばらつき具合を示す．発言量標準偏差 $sd(c)$ は式 (4) から求められる．

$$sd(c) = \sqrt{\frac{\sum_{i=1}^N (c_i - average(c))^2}{N}} \quad (4)$$

傾き

x 軸を時間， y 軸を発言番号として x - y 座標平面上に発言をプロットしたグラフにおける，最小 2 乗法による 1 次近似直線の傾きである．この傾き a は，単位時間あたりの発言数を表すので，この値が大きい場合には，単位時間あたりの発言数が多い，つまりそのスレッドの進み具合が速く，小さい場合は，進み具合が遅いといえる．

年齢

あるスレッドの最初の発言から最後の発言までの経過時間をこう呼ぶことにする．コミュニティの時間的な大きさの目安となると考えられる．

総発言数

あるスレッドのすべての発言数である．コミュニティの大きさの目安となると考えられる．

5.4.2 社会ネットワーク分析指標

5.4.2.1 ネットワークデータの準備

社会ネットワーク指標を求めるためには，まず行為者間の関係を表すネットワークデータが必要である．掲示板ログデータの発言番号と返信先番号からは，どの発言がどの発言に対する返信であるかということが分かる．そして，各発言者に一意の発言者番号を割り当て，各発言番号を発言者番号に書き換えることにより，ログデータの発言間の関係から発言者間の関係を得ることができる．

本システムでは，後述する分析指標の計算に適宜用いるために，このような行為者間の関係を表すネットワークデータとして，2 値非対称行列，2 値対称行列，重みつき行列，の 3 種類の行列を掲示板ログデータ収集時に同時に作成している．いずれの行列も，行を発言元，列を返信先として，各行為者に同一の行番号と

列番号を与える．

2 値非対称行列は，行為者間の関係に，発言する，発言されるという方向性があることから，関係の回数は考えない場合である．行為者間に関係がある，つまり返信先のある発言の場合には 1，関係がない場合には 0 の成分をそれぞれ対応する要素に入れる．対角要素は行為者本人への関係を示す数値が入る部分で，今回は発言がある場合に 1 を，発言がない場合には 0 を入れた．以上のような方法で生成された行列は，完全に相互的な行為者間の関係を全行為者が持っていないかぎり非対称行列となる．

2 値対称行列は，行為者間の関係に方向性がないことから，関係の回数は考えない場合である．A さんが B さんに発言した場合と，B さんが A さんに発言した場合を区別せず，A 行 B 列と B 行 A 列の要素にともに 1 を入れる．

重みつき行列は，行為者間の関係についてその回数も考慮する．各要素はそこで交差する行為者間の関係の回数である．ここでは関係の方向性があるものとして，行をその発言元，列を返信先とした非対称行列とした．

実装上は，1 つのログデータからまず発言者リストを作成する．発言者リストはユニークな ID 番号，発言者名，最初の発言日時をタブ区切りの 1 行としたテキストファイルで，行数はユニーク発言者数である．次にログデータと発言者リストから各発言がだれのだれへの発言かをユニーク発言者番号の対で表したファイルを作成する．ファイルの 1 行は 1 発言に対応し，発言者 ID 番号と返信先 ID 番号をタブで区切った 2 つの数値からなる．行数はコミュニティの発言数である．このファイルから 3 種類の行列ファイルを生成している．以上の処理をログデータ収集直後に自動的に行うことにより，5.1.2 項の要件であるデータ加工の自動化に対応している．

5.4.2.2 指標の説明

以下の社会ネットワーク分析指標は，ネットワークデータを利用する．

サイズ

コミュニティを構成する行為者数，すなわち発言者数を表す．行列の行 (列) の数 g となる．

次数 (入次数，出次数)

次数とはネットワークのある点の持つ紐帯数である．これは各行為者の他の行為者との関わり，すなわち掲示板コミュニティにおいては発言する，発言されるという関係を示す．次数は，発言の方向性について考慮する場合には，ある点に入る紐帯数の入次数と，ある

点から出る紐帯数の出次数に分けることができる。これらの指標では、関係の回数は考慮していない。行為者 i の次数 n_i は 2 値対称行列における i に対応する行(列)に含まれる対角成分以外の 1 の数となる。入次数は i に対して発言した人数で、2 値非対称行列における対応列に含まれる対角成分以外の 1 の数、出次数は i が発言を行った他の発言者の人数で、2 値非対称行列における対応行に含まれる対角成分以外の 1 の数となる。

次数中心性集中化

中心性とは、ある行為者がネットワーク内でどの程度中心的存在であるかを示す指標であり、何を基準とするかにより様々な中心性が考案されている。ここで扱うのは、式 (5) に示す Wasserman らによる次数に基づく標準化された中心性 $Cd(n_i)$ で、0 から 1 までの値をとる³¹⁾。

$$centrality_i \quad Cd(n_i) = \frac{n_i}{g-1} \quad (5)$$

ネットワークにおける行為者の中心性の偏り具合を示すために集中化という指標がある。次数に基づく中心性の集中化 Cd は、 $Cd(n^*)$ を最大の中心性として式 (6) で示される。

$$centralization \quad Cd = \frac{\sum_{i=1}^g \{Cd(n^*) - Cd(n_i)\}}{g-2} \quad (6)$$

密度

行為者同士の関係がどのくらい緊密であるかを示し、理論的に存在可能な紐帯数に対する実際の紐帯数の割合である。無向グラフのネットワーク密度 d は式 (7) で求められる。ここで n はネットワークサイズ、 n_i は点 i における次数である。

$$density \quad d = \frac{\sum_{i=1}^n (n_i)}{n(n-1)} \quad (7)$$

直接結合度

行為者間で双方向的な関係がどの程度存在しているかを示す。直接結合 coh は式 (8) から求められる。行為者間の双方向的な関係があるときのみ分子がカウントされる。行為者 i から j への有向紐帯を z_{ij} として、 $z_{ij} \oplus z_{ji}$ は $z_{ij} = 1$ かつ $z_{ji} = 1$ のときのみ 1 をとり、それ以外では 0 をとる。

$$cohesion \quad coh = \frac{\sum_{i=1}^n \sum_{j=1}^n (z_{ij} \oplus z_{ji})}{n(n-1)/2} \quad (8)$$

5.5 システム動作例・分析実施例

本節ではここまで説明してきた CMINER の動作例をシステムの画面を交えて示す。

図 2 は掲示板ログデータ収集画面である。まず対象



図 2 ログデータ収集画面
Fig. 2 Log collection window.



図 3 メイン画面
Fig. 3 Main window.

となる掲示板の種類を選択する。これが掲示板別のログインの選択となる。トピック URL 欄にコミュニティの URL を入力し、必要に応じてトピック名やカテゴリ名を入力し、「Get CmtFile」ボタンを押すことにより、指定 URL のログを収集する。同時にログからネットワークデータを生成している。その進み具合はボタン右のプログレッシバースタックに表示される。処理が終了すると画面下部のリストに取得したコミュニティが追加され、またメイン画面のリストも更新される。なお、現在「Yahoo! 掲示板³²⁾」についてはトピック一覧 URL を指定し「Get Topic」ボタンを押すことにより、そこに含まれるトピックの URL、トピック名データを自動的に収集できる。

図 3 はシステムのメイン画面で、ここから本システムのすべての機能を利用できる。上段左側の「新規取得」ボタンは、上述の掲示板ログデータ収集画面を起動させる。上段中央は、データ分析指標が並んでいる。表示させたい値を選択し、上段右側の「表示」ボタンを押すと、各コミュニティの指標値が表示される。

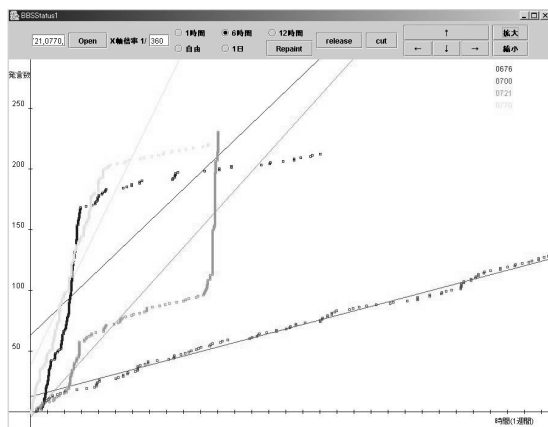


図 4 時間—発言数グラフ表示画面
Fig. 4 Time-posts graph window.

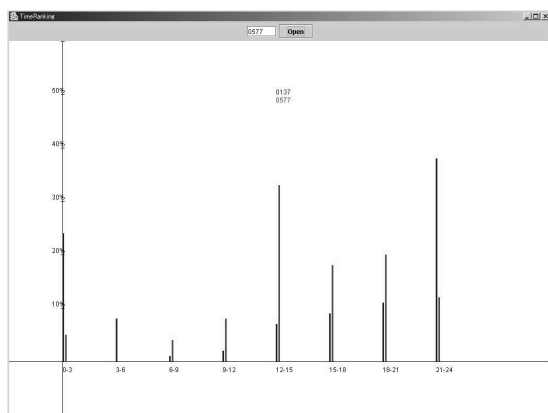


図 5 活動時間グラフ表示画面
Fig. 5 Active hours graph window.

中段は収集されたコミュニティのリストで、ここから分析対象を選択する。下段のボタンは分析結果の視覚化などの他機能を起動する。主なものについて簡単に説明する。

分析対象を選択した上で「時間—発言数グラフ」ボタンを押すと、図 4 のようなグラフが表示される。選択されたデータはその番号が上段左枠内に表示されており、右隣の「Open」ボタンで表示する。X 軸の時間は、最初の発言を 0 とした相対時間で、表示スケールは上部の「X 軸比率」部分で変更できる。Y 軸は発言数で、これもスケール変更可能である。このように発言数を時系列表示することで、コミュニティごとに発言頻度が異なり、1 つのコミュニティ中でも変化していることが分かる。傾きが大きい部分は、短時間の間に多くの発言が行われたことを示し、活動が活発であったといえる。この傾きによって発言活性度の時系列変化が分かる。



図 6 ログデータフィルタリング画面
Fig. 6 Log data filtering window.

「活動時間グラフ」ボタンを押すと、図 5 のようなグラフが表示される。これは、各コミュニティ内の発言を時間帯別に分類したグラフである。昼間に活発なコミュニティや夜間に活発なコミュニティがあることが分かる。

「次数グラフ」ボタンを押すと、コミュニティ内の発言者の次数、入次数、出次数を棒グラフで表示し、またその値でソートすることができる。これにより、次数による中心性の存在が分かりやすくなる。

「時系列変化グラフ」では、コミュニティにおける、各種分析指標値の推移を表示できる。X 軸が時間で、そのスケールは可変である。

「指標比較グラフ」では、選択した分析指標値について選択した複数コミュニティ間の比較を視覚化する。

「類型化ウィンドウ」ボタンにより起動する図 6 の画面では、ログデータの絞り込みが可能である。画面左上ではコミュニティに付加したカテゴリ名による選択絞り込みができる。画面上段では分析指標を選択してその値の範囲による絞り込みができる。分析指標は 3 つまで組み合わせることができる。絞り込まれたコミュニティのリストが中段に表示される。この絞り込まれたデータに対して下段のボタンにより視覚化が行える。これにより 5.1.3 項の要件であるコミュニティ選択機能に対応している。

6. 評価実験

CMINER によるコミュニティ分析自体は、本論文の域を超えるが、システムの有効性を確認するために評価実験を行った。実験の目的は CMINER を用いることにより従来手法と比較してどの程度作業が効率化されるのかを検証することである。

6.1 実験概要

ネットワークコミュニティ分析に関心のある社会心理学専攻の大学生・大学院生 5 名が、CMINER および従来手法により社会ネットワーク分析を含むログ分析の課題を行った。その必要作業時間と質問紙調査により両者を比較した。比較ソフトウェアは 4.3 節に述べた UCINET とした。現在よく知られている社会ネットワーク分析ソフトウェアで GUI ベースである点から適切な比較対象である。

6.2 実験方法

被験者 5 名はいずれもネットワークコミュニティ分析に関心があるが、両ソフトウェアの使用経験はない。エンドユーザとしてのコンピュータの使用には十分慣れている。社会ネットワーク分析についても基礎的知識を有するが、専門家とはいえないことと、被験者個人に起因するばらつきを減らすため、一定のデータ処理手続きとするため、実験に先立ち次の確認、教示をした。

1) 社会ネットワーク分析についての知識(中心性、密度の指標の意味)を確認し、不足していると判断された場合は実験者が教示した。

2) 比較方法におけるログ取得方法について教示し、被験者は中断なく課題を遂行できるようになるまで練習した。内容は、Yahoo! 掲示板からコピー&ペーストにより実験者が指定するスレッドのログを Excel に取得し、行列を作成する方法である。

3) UCINET および CMINER の操作方法を実験に必要な部分のみ教示した。UCINET については行列データの入力方法、指標の算出方法を、CMINER についてはオンラインでのログデータの取得方法、指標の算出方法を教示した。

以上の準備の後、次の実験を実施した。

1) 使用するウェブ掲示板のスレッドを実験者が各被験者に指示した。これは事前練習と異なる Yahoo! 掲示板スレッドである。実験に適当な時間で終了するように小規模なスレッドを指示した。

2) 被験者は各自くじを引き、UCINET と CMINER のどちらを先に使用するかを決めた。

3) 被験者は各自課題を行った。課題は指示されたスレッドについて、総発言数、サイズ、中心性、密度を求めることである。UCINET の場合は、掲示板からコピー&ペーストにより必要データを Excel に取得し、行列を作成のうえ、UCINET に入力するテキスト形式のデータファイルにするまでの作業時間を T1 とした。この作業の途中で総発言数、サイズを得ている。またその後データファイルを UCINET に入力し

表 1 作業時間結果

Table 1 Required time result.

被験者	スレッド 総発言数	スレッド サイズ	UCINET+Excel			CMINER		
			T1(s)	T2(s)	T1+T2	T1(s)	T2(s)	T1+T2
T	13	7	852	91	943	86	25	111
K	14	5	500	67	567	110	29	139
M	12	10	676	61	737	68	16	84
E	12	7	470	62	532	63	19	82
Y	30	23	1878	55	1933	74	19	93
平均	16.2	10.4	875.2	67.2	942.4	80.2	21.6	101.8

表 2 質問紙調査結果

Table 2 Questionnaire result.

被験者	ネットワークコミュニティを分析する際、UCINET は使いやすい	ネットワークコミュニティを分析する際、CMINER は使いやすい	ネットワークコミュニティを分析しようとしている人には、UCINET を勧めたい	ネットワークコミュニティを分析しようとしている人には、CMINER を勧めたい
T	3	2	2	2
K	3	1	1	2
M	4	2	3	2
E	4	1	2	1
Y	3	1	3	2
平均	3.4	1.4	2.2	1.8

て、中心性、密度を算出するまでの作業時間を T2 として測定した。CMINER の場合は、掲示板から「新規取得」ボタンによりログデータを取得し(自動的に行列も作成される)終わるまでの作業時間を T1 とし、その後、総発言数、サイズ、中心性、密度を算出するまでの作業時間を T2 として測定した。

実験終了後、被験者に簡単な質問紙調査を実施した。

6.3 実験結果

作業時間結果を表 1 に示す。各被験者に対して作業を指示したスレッドの総発言数とサイズも示した。UCINET での作業に比べて CMINER での作業は必要時間が T1、T2 のいずれにおいても大幅に短縮されたことが分かる。特に UCINET 条件で Excel を利用する T1 部分はその差が顕著であった。

次に質問紙調査結果を表 2 に示す。「そう思う(1点)」「まあそう思う(2点)」「あまりそうは思わない(3点)」「そうは思わない(4点)」のリッカート尺度による回答である。これ以外に、両者でネットワークコミュニティを分析する際に使いにくかった点、優れていた点についても自由記入形式で回答を求めた。現在の CMINER はプロトタイプで完成度が高いとはいえないが、従来ソフトより高い評価を得た。使いやすさに関する項目では、UCINET では平均 3.4、CMINER は平均 1.4 となった($t = 6.32, p < .01$)。人に勧めたいかに関する項目では、UCINET では平均 2.2、CMINER は平均 1.8 となった($t = 1.00, n.s.$)。自由記入の意見から、UCINET が入力するデータファイルを作成する必要があるのに対して、CMINER は

データ取得から分析まで一貫して支援している点が高く評価されたことが分かった。

7. ま と め

ネットコミュニティの典型的1つであるウェブ掲示板によるコミュニティ, すなわちスレッドおよびその集合を対象として, ログデータを容易に収集し, ログ分析の指標と社会ネットワーク分析の指標により分析することができる, ネットワークコミュニティ分析支援システム CMINER を開発した。従来の, 個別ケースに限定されたデータ取得の工夫や社会ネットワーク分析ソフトウェアの使用に比べ, 本システムはログデータの取得から分析を一貫支援し, ウェブ掲示板という範囲内ではあるが, ある程度の汎用性を持たせている。システムの支援効果についても従来の代表的な分析ソフトウェアとの比較実験により確認した。本システムの社会科学研究への適用を今後進めていきたい。

謝辞 本研究にご協力いただいた根本啓一氏に深謝いたします。また, 本研究の一部は国際コミュニケーション基金調査研究助成および国立情報学研究所研究企画推進室特別プロジェクトの支援によるものです。記して感謝いたします。

参 考 文 献

- MacIver, R.M.: コミュニティ, ミネルヴァ書房 (1975).
- Hillery, G.A. Jr.: Definitions of community: Area of agreement, *Rural Sociology*, 20.
- Preece, J.: *Online communities: designing usability, supporting sociability*, John Wiley & Sons (2000).
- Kim, A.J.: *Community building on the web*, Peachpit Press, CA (2000).
- 川上善郎, 川浦康至, 池田謙一, 古川良治: 電子ネットワーキングの社会心理—コンピュータ・コミュニケーションへのパスポート, 誠信書房, p.216 (1993).
- 池田謙一(編): ネットワーキング・コミュニティ, p.202, 東京大学出版会 (1997).
- NIFTY ネットワークコミュニティ研究会, 金子郁容, 松岡正剛, 中村雄二郎, 岡田智雄: 電縁交響主義, p.375, NTT 出版 (1997).
- Nardi, B.A., Whittaker, S. and Schwarz, H.: NetWORKers and their Activity in Intensional Networks, *Computer Supported Cooperative Work*, Vol.11, No.1-2, pp.205-242, Kluwer Academic Publishers, Netherlands (2002).
- 西田豊明: インタラクションの理解とデザイン, 岩波書店 (2000).
- 福原知宏, 近間正樹, 西田豊明: コミュニティの知識共有を目的とした話の共有システムの提案, 第16回人工知能学会大会, 2C3-05 (2002).
- 梅木秀雄, 下郡信宏, 横田健彦: ネットワークコミュニティ形成支援, 情報処理学会研究報告, 情報メディア 37-5, pp.25-30 (2000).
- 亀井剛次, エバジェットマー, 藤田邦彦, 吉田仙, 桑原和宏: ネットワークコミュニティの形成を支援するシステム “Community Organizer” における情報提示手法の検討, 信学論(D-I), Vol.J84-D-I, No.9, pp.1440-1449 (2001).
- Erickson, T., Halverson, C., Kellogg, W.A., Laff, M. and Wolf, T.: Social Translucence: Designing Social Infrastructures That Make Collective Activity Visible, *Comm. ACM*, Vol.45, No.4, pp.40-44 (2002).
- Donath, J.: A Semantic Approach to Visualizing Online Conversations, *Comm. ACM*, Vol.45, No.4, pp.45-50 (2002).
- 野島久雄, 新垣紀子: CMC 研究の歴史と広がり, 人工知能学会第6回ことば工学会研究 (2000).
- 三浦麻子, 篠原一光: チャット・コミュニケーションに関する心理学的研究—ログ記録の解析に基づく探索的検討, 対人社会心理学研究, No.2, pp.25-34 (2002).
- 篠原一光, 三浦麻子: WWW 掲示板を用いた電子コミュニティ形成過程に関する研究, 社会心理学研究, Vol.14, No.3, pp.144-154 (1999).
- Smith, M.: Netscan: A tool for measuring and mapping social cyberspaces (2001). <http://netscan.research.microsoft.com>
- International Network for Social Network Analysis: What is Network Analysis?. <http://www.sfu.ca/~insna/INSNA/na.inf.html>
- 安田 雪: 実践ネットワーク分析, p.188, 新曜社 (2001).
- 北山 聡: フォーラムの生態—インタビュー + 計量分析の試み, in NIFTY ネットワークコミュニティ研究会(企画), 電縁交響主義, pp.34-67, NTT 出版 (1997).
- 高橋正道, 北山 聡, 金子郁容: ネットワーク・コミュニティにおける組織アウェアネスの計量と可視化, 情報処理学会論文誌, Vol.40, No.11, pp.3988-3999 (1999).
- 藤田邦彦, 亀井剛次, Eva Jettmar, 吉田 仙, 桑原和宏: サイバーコミュニティ形成支援システムのネットワーク分析による評価, 情報処理学会研究報告グループウェア 39-1, pp.1-6 (2001).
- 福原知宏, 近間正樹, 西田豊明: ネットワークコミュニティにおける活動分析支援システムの提案, 情報科学技術フォーラム講演論文集, pp.145-146 (2002).
- Analytic Technologies, Inc.: UCINET 5. http://www.analytictech.com/ucinet_5_description.htm

- 26) Burt, S.R.: STRUCTURE Version4.2 Reference Manual. <http://gsbwww.uchicago.edu/fac/ronald.burt/teaching/STRUCmanual.pdf>
- 27) GRADAP. <http://www.assess.com/Software/GRADAP.htm>
- 28) INCM@HOME – INCM ホームページ .
<http://www.kaede.sakura.ne.jp/~nikeno/incm/>
- 29) air – material – INCM plug-in.
http://air.vis.ne.jp/material/incm/incm_pl.html
- 30) 纏読館 (まとめよみかん).
<http://kt.sakura.ne.jp/~timeflow/M/>
- 31) Wasserman, S. and Faust, K.: *Social Network Analysis: Methods and Applications*, Cambridge University Press, p.825 (1994).
- 32) Yahoo! 掲示板: <http://messages.yahoo.co.jp/index.html>

(平成 15 年 5 月 28 日受付)

(平成 15 年 11 月 4 日採録)



井上 智雄 (正会員)

1998 年慶應義塾大学大学院理工学研究科計測工学専攻博士課程修了。博士 (工学)。国立情報学研究所知能システム研究系助手を経て、現在筑波大学図書館情報学系助教授。社会的インタラクションの理解と支援、CSCW、e-Learning の研究に従事。本会論文賞、同山下記念研究賞、電気通信普及財団テレコムシステム技術学生賞他受賞。本会論文誌応用グループ主査、グループウェアとネットワークサービス研究会幹事、日本 VR 学会サイバースペース研究会運営委員。電子情報通信学会、教育システム情報学会、ACM ほか各会員。



小林 哲郎

2002 年東京大学文学部社会心理学専修課程卒業。東京大学大学院人文社会系研究科社会心理学専攻修士課程。専攻は、コミュニケーション論、社会的ネットワーク論、社会関係資本論。現在の主要研究テーマはインターネット利用の社会的帰結。コミュニケーションメディアとしてのインターネット利用が生み出す社会レベルでの効果を、サーヴェイデータを用いて実証的に検討している。



池田 謙一

1982 年 3 月東京大学大学院社会学研究科博士課程中途退学。後に博士 (社会心理学)。1987 年 4 月明治学院大学法学部専任講師、1992 年東京大学文学部助教授を経て、現在同大学院人文社会系研究科教授。専攻はコミュニケーション研究、社会心理学。



重野 寛 (正会員)

1997 年慶應義塾大学大学院理工学研究科博士課程修了。現在、同大学理工学部情報工学科助教授。工学博士。無線 LAN の構成法と媒体アクセス制御方式、計算機ネットワークにおけるステーション移動サポート、モバイル・コンピューティング、マルチエージェントシステム、アクティブネットワーク、遠隔教育システム等の研究に従事。著書『～ネットワーク・ユーズのための～無線 LAN 技術講座』(ソフト・リサーチ・センター)、『コンピュータネットワーク』(オーム社)等。電子情報通信学会、IEEE、ACM 各会員。



岡田 謙一 (正会員)

慶應義塾大学理工学部情報工学科教授。工学博士。情報処理学会学会誌編集主査、同論文誌編集主査、電子情報通信学会論文誌編集委員、情報処理学会グループウェア研究会主査、同モバイル研究会幹事、電子情報通信学会マルチメディアインフラストラクチャ&サービス研究会幹事、日本 VR 学会サイバースペース研究会副委員長等を歴任。情報処理学会論文賞 (1996, 2001)、情報処理学会 40 周年記念論文賞、日本バーチャルリアリティ学会サイバースペース研究賞を受賞。