1409 複雑ネットワーク分析による利用者間の情報伝播モデルの構築

Development of Information Spreading Model among Users by Complex Network Analysis

学生員 ○内田 誠 (東京大学大学院)

正員 白山 晋 (東京大学人工物)

UCHIDA Makoto (the University of Tokyo)
uchida@race.u-tokyo.ac.jp

SHIRAYAMA Susumu (RACE, the University of Tokyo)
sirayama@race.u-tokyo.ac.jp

Key Words: *Complex Network Analysis, Social Network Analysis, Information Diffusion/Spreading, Digital Value Engineering.*

Varying user's sense of value, it is getting more important to deal with users' individual needs and value standard, when designing a product or a service. Such needs or value standard would be not only formed by their own attributes, but also affected by an interaction with other users. To analyze such interaction among users and emergence of needs and value standard, and to build a model of user which could imply such interaction and emergence, we observed a complex network structure which is composed of users. This paper reports an outline and a structure of real network instances, Weblog network and social networking service, applying complex network analysis. Then we propose an information diffusion/spreading model on the complex network, and experiment it by computer simulation on the real users' network.

1. はじめに

近年の製品やサービスの設計において、平均的な利用者モデルの限界が指摘されている。これは、利用者の価値観の多様化に起因するものである[1]。多様な価値観や価値の定量化に対しては、個の属性に加え、周囲の環境による影響や、群の中で他の主体との相互作用を考える必要がある。それらを統合的に扱う方法としては、アンケート分析などによる個の属性の抽出と、主体のエージェント化によるエージェントシミュレーションがあるが、現実的な意味での個体同士の情報伝達が考慮されていない、また、群としての行動の表現が難しいといった課題がある。一方、情報伝達や群の構造を分析するための手法として、複雑ネットワーク分析という新たな手法が提案され、様々な適用事例ともいくつかの成果が報告されている[2,3]。

本研究では、個と全体の間に位置する群としての構造をモデル化し、群の内部と外部の相互作用までを考慮した利用者モデルの構築を目標としている。また、個々の要求や価値尺度を適切に観測し、それを製品やサービスの設計に活かす仕組みの提案を目指している。

本稿では、個と群のモデル化を考える際に重要となる主体間の情報伝達や合意形成に着目し、利用者のネットワーク構造の抽出を試みる。いくつかの実世界のネットワークについて、複雑ネットワーク分析により、その構造や性質を明らかにする。次に、分析したネットワーク上での情報伝播モデルについて考察する。

2. 利用者のネットワーク

2-1 社会的ネットワークの特徴

近年、情報通信手段の発達により個人間のコミュニケーションに要する時間が短縮され、コストも劇的に低下している。このため、ある時点や場所からの情報の伝播に要する時間が極めて短いものになっている。また、インターネット上での掲示板やブログ (Weblog)、SNS (ソーシャルネットワーキングサービス) などの普及により、地理的・組織的制約条件を越えた個人間のコミュニケーションが容易なものになっている。その結果、属地的・属組織的な人的ネットワークとは異なる新たな形態の社会的ネットワークが形成されており、その上での情報交換も活発に行なわれている。

このように、個人がネットワークを通じて周囲から影響を受ける、あるいは周囲に影響を与える機会は増加しているが、

それは均質なものではなく、現実の社会的ネットワークでは、不均一性や不完全性が顕在化する。ネットワークに属する個人を主体と考えたときに、主体は一様に分布し、他の主体と均一に結びついているわけではなく、リンクの強さの強弱や、非常に多くのリンクを持つ「ハブ」の存在や、主体同士が相互に密に結びついている「コミュニティ」の存在が認められることが多い (図1)。このような現実の社会的ネットワークに生じる特徴は、以下のようにまとめられる。

- 個々が周囲から得られる情報量や影響の強さは均一ではない。
- 個々がコミュニケーションをとれる他者の数は平等ではない。
- 個々が発信できる情報量や周囲に与えることができる影響の強さは平等ではない。

群としての構造を持つ利用者のモデル化を考える際には、現実の社会的ネットワークに存在するこれらの性質を考慮することが不可欠となる。

図1. 現実のネットワークの例

2-2 複雑ネットワーク分析

前節で述べたネットワークの特徴を分析するための手法として、複雑ネットワーク分析がある[2]。複雑ネットワーク分析では、計量社会学での社会ネットワーク・コミュニティ分析や、グラフ理論、また統計物理学的手法などによってノードとエッジに抽象化されたネットワークの構造や性質を扱う。その適用範囲は広く、人間同士の社会的ネットワークから、インターネットのトポロジ、WEB ページのハイパーリンク、タンパク質の反応連鎖ネットワークなどに至るまで、互いに異なる分野で扱われるネットワークの事例がいくつかの共通のモデルで説明可能であることが明らかになりつつある[3]。

2-3 ネットワーク構造に対する指標

本節では、複雑ネットワーク分析で用いられる代表的な

ネットワーク指標について説明する。

(a) 次数分布

ノードが持つエッジの数の数 (次数) の分布である。次数 k を持つノードの存在確率分布 $p(k)$ 、もしくはその累積確率である $P(k \leq X)$ が指標となる。現実のネットワークでは $p(k)$ が k の冪乗に比例する冪分布になる例が多く報告されている。

(b) 平均経路長

ネットワーク中の任意の2ノード間の最短経路上に存在するエッジの数を経路長 (距離) と定義する。ネットワーク中の任意の2ノード間の経路長の平均として平均経路長 L を、

$$L = \frac{1}{n(n-1)/2} \sum_{i,j} d_{ij} \quad (1)$$

で定義する。ここで、 n はネットワーク中のノード数、 d_{ij} はノード i と j との間の経路長である。一般に、 L は n に対して比較的小さいことが知られている。

(c) クラスタ率

クラスタ率 C はネットワークの中でのノードの凝集度を示す指標であり、「あるノードの隣接ノード間にエッジが存在する確率」として定義される。 k_i をノード i の次数、 E_i をノード i の隣接ノード間に存在するエッジの数とし、 C を、

$$C = \frac{1}{n} \sum_i C_i, \quad C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (2)$$

と定義する。現実のネットワークでは、 n が非常に大きいネットワークでも C の値は比較的大きく、ネットワークの凝集性が保たれることが報告されている[3]。

3. 提案手法

3-1 ネットワーク構造の分析と利用者のクラスタリング

本研究で行なう利用者間の情報伝播モデルの構築は、図2に示すような二段階のステップで行なう。

はじめに、ネットワークに属する個人間の情報が到達し、相互作用が発生する範囲や経路は、そのネットワークの構造によって規定されると考える。したがって、ネットワークの静的な構造を分析し、情報伝播モデルの構築に必要な要素を抽出する必要がある。また、この段階でクラスタリング手法によって、利用者をグループ化する。これをグループユーザと名付ける。

図2 ネットワーク分析による2段階のアプローチ

3-2 情報伝播モデルの構築

第二に、構造が明らかになったネットワーク上での、グループユーザに対しての情報流通や相互作用のダイナミクスを考察する。手法として、マルチエージェントシミュレーション等を利用する。構造を所与としての伝播や合意形成の過程に関する研究は多く行なわれており、そのモデルや手法、知見を援用する。

3-2.1 局所情報伝播モデル

本稿では、情報伝播のモデルとしてニューラルネットワークで用いられるものを発展させる。ノード v は $+1$ か -1 のいずれかの状態を持ち、周囲の状況によって次ステップの状態が決まるものとする。この決め方に対して、本稿では、二つのモデルを用いる。一つは、多数決モデルである。 $\sigma_i(t)$ をノード v_i の t での状態を示すものとする、このモデルは式(3)で示される。

$$\sigma_i(t+1) = \text{sgn}\left\{\sum_j c_{ij} \sigma_j(t)\right\} \quad (3)$$

$$c_{ij} = c_{ji} = \begin{cases} 1 & v_i \text{ と } v_j \text{ の間にリンク有} \\ 0 & v_i \text{ と } v_j \text{ の間にリンク無} \end{cases}$$

もう一つは、(4)式で示す重み付き多数決モデルである。

$$\sigma_i(t+1) = \text{sgn}\left\{\sum_j k_j c_{ij} \sigma_j(t)\right\} \quad (4)$$

ここで、 k_j はノード v_j のエッジの次数である。これらのモデルを用いて伝播の様子を分析する。

3-2.2 実ネットワーク上の情報伝播シミュレーション

ネットワーク上の情報伝播や合意形成のシミュレーションの多くは、ネットワーク構造を2次元格子やランダムなネットワークなどの理想化されたもので行われることがほとんどである。このため、現実のネットワーク構造上で得られる結果とは異なることが考えられる (特に局在性が問題になるものと予想できる)。そこで、従来のものとの類似点、相違点を明らかにするために、現実のネットワークに基づく情報伝播シミュレーションを行なう。いくつかのネットワークモデルとの比較を行うことで、伝播過程の違いを分析する。

3-2.3 大域的信息伝播モデル

図2で示した二つのプロセスは、階層化されているが独立ではない。ネットワークを構成する個人間の相互作用によって、リンクの有無や強さといったネットワークそのものの構造もダイナミックに変化する。必要に応じてフィードバックプロセスを加えることにする。

利用者間の相互作用ネットワークの構造、その上での情報伝播のモデル、実ネットワークを対象とした情報伝播シミュレーション、再モデリングというプロセスを繰り返しながら、個、群、全体の位相的構造の動的変化と相互作用を考慮し、ネットワーク上の大域的な情報伝播のモデル化を試みる。

3-3 利用者モデルの構築

以上の手順で明らかにされる利用者間の相互作用ネットワークの構造と、その上でなされる情報伝播によって、利用者個々の固有属性に加え、群の中での周囲との相互作用の結果として創発的に生じる性質までを対象とすることができるようになる。結果として、従来の手法では取り扱うことができなかった群による創発的性質を利用者モデルに含めることが可能になる。このモデルができれば、従来のモデルでは十分な記述や説明が難しかった個々人の多様な要求や価値尺度を実世界に即した形でモデル上に記述できるだろう。このフレームワークについては、講演時に示す。

4. ネットワーク分析の事例

現実の利用者ネットワークの例として、Weblogの記事ネットワークと、ソーシャルネットワーキングサービス (SNS) について、静的な構造を分析する。

4-1 Weblog ネットワーク

Weblog は、WWW上で個人が簡単に情報発信を行なうことが可能なシステムの総称である。記事毎に独立したコンテンツとして管理され、「トラックバック」と呼ばれる仕組みによって、記事同士の繋がりが明示される。トラックバックは自分もしくは他人の記事との引用や内容の関連の繋がりを示すものであり、記事をノード、トラックバックを有向のエッジとすれば、相互に情報がやり取りされた記事の巨大なネットワークとして捉えることができる。

はじめに、実際のWWW空間のある期間に作成された7183記事を起点として、そこからトラックバックを遡るアルゴリズムによってネットワークを抽出する。このためネットワークは非連結である。

収集したサンプルの記事数は 60,892、トラックバック数 187,446 であり、これらがネットワークのノード数とエッジ数に相当する。クラスタ率は 0.13 であった。なお、ネットワークが有向であり非連結であるため、平均経路長を調べることは本質的な意味はない。

次に、度数分布を調べる。各記事が持つトラックバックの度数の平均は、入次数・出次数とも約 3.08 となるが、その分布は図 3 に示すように、平均付近にピークを持たない分布である。図 3 に度数の確率分布を、エッジの入次数・出次数に分けて示す。横軸はエッジの度数、縦軸は確率を表す。

入次数は両対数スケールで直線的に分布し、冪乗分布的になっている様子がわかる。ただし、低次領域と高次領域で傾きが異なり、度数の上位 10% である累積確率 $P < 0.1$ の高次領域 ($k \geq 8$ に相当) では冪指数は 1.99、 $P \geq 0.1$ の低次領域では 0.98 となり、低次領域では傾きが緩やかになっている。一方、出次数も両対数スケールで直線的に分布するが、全ての領域で傾きがほぼ一定で冪指数は 2.58 となり、入次数の場合よりも傾きが急である。

この結果の違いは、トラックバックの送信は記事の書き手が他の記事に対して制限なく行えるが、受信は他の書き手がその記事に関連する記事を書いたときのみ行えるというトラックバックの非対称性によるものと考えられる。

図 3. Weblog ネットワークの度数の確率分布 (両対数プロット)
(上段: 入次数、下段: 出次数)

4-2 SNS (ソーシャルネットワーキングサービス)

Weblog が、ある個人が書いた記事と記事との繋がりを示すネットワークであるのに対し、SNS はインターネットを介した人と人との繋がりのネットワークであるといえる。SNS の運営企業から提供された実データを対象として分析する。

このデータは、ユーザーとその知人関係を表わすリンクの情報からなり、これらがネットワークのノードとエッジに相当する。また、SIG (Special Interest Group) と呼ばれるユーザーの興味に応じたグループとユーザーを紐づけるデータもある。すなわち、サンプル全体では図 4 のようにユーザーと SIG の階層型のネットワークをなしている。なお、データからは個人や SIG を特定できる情報は一切消去されており、純粋にトポロジカルな情報のみとなっている。データの規模は、ユーザー数 363,819、リンク数 (有向) 3,813,702 である。ネットワークは非連結であるが、全体の 99% 以上のユーザーが最大連結成分に含まれるため、最大連結成分 (ユーザー数 360,802、リンク数 3,809,228) のみを分析対象とする。クラ

スタ率は $C=0.330$ 、平均経路長は $L=5.48$ であった。

次に、ユーザー同士のネットワークの度数分布を調べる。エッジに向きのあるネットワークであるが、必ず逆向きのエッジも持つので、実質的に無向のネットワークとみなせる。度数の確率分布を図 5 に示す。リンクの平均次数は 10.55、標準偏差は 19.52 である。Weblog の入次数分布に似て、低次領域と高次領域で傾きの異なる冪乗分布の様子がみられる。度数の上位 15% である累積確率 $P < 0.15$ の高次領域 ($k \geq 19$ に相当) では冪指数は 2.64、それより低次の領域では冪指数は 1.08 と顕著な差が表れ、低次領域では傾きが緩やかになっている。

さらに、ユーザーと SIG との関係について調べる。SIG は全体で 90,795 存在し、ユーザーは平均して 16.18 の SIG に参加し、また 1 つの SIG には平均して 64.29 のユーザーが所属している。また、標準偏差はそれぞれ 45.88、210.71 であり、平均に比べて大きい。それらの度数の確率分布を図 6、図 7 に示す。これらの度数分布も冪乗分布的の性質を示すが、やはり高次領域と低次領域では冪指数が異なり、これも低次領域では傾きが緩やかになっている。

図 4. SNS のサンプルデータの概要

図 5. SNS ユーザーのリンク度数の確率分布 (両対数プロット)

図 6. ユーザーの参加 SIG 数の確率分布 (左) と、SIG の参加ユーザー数の確率分布 (右) (両対数プロット)

4-3 考察

Weblog のネットワークと SNS のネットワークともにクラスタ率は、同規模のノードをランダムに配置するネットワークの場合よりも 3 桁以上大きく、高い凝集性を示している。また、SNS のネットワークでは小さな平均経路長も見られ、30 万人以上のユーザーにわずか 4、5 人の知人を介することで到達するという「小さな世界」[3]の性質が見られた。また、エッジの度数分布は冪乗分布的の性質を示していた。これらのことより、実世界のネットワークではノードとエッジが均一

ではなく、ハブや密なコミュニティの存在といった不均質性を持つことが確認された。

これらの性質は、Watts や Barabasi らによって指摘され、このような性質を持つネットワークが生成するメカニズムも提案されている[3]。しかし、それらのモデルだけでは、分布の冪指数の大きさや、今回複数の事例で共通して観察されるような、低次数領域で冪指数が小さくなる次数分布のカットオフ現象、高次数領域での確率分布の不均一性などを十分に説明することはできない。これらの現象は、人間が介在するいくつかのネットワークで観察されている[3]が、定説はない。本研究では、いくつかの考察からネットワーク生成の過程において影響を与える何らかのヒューマンファクターの存在を予測している。

5. 実ネットワーク上での情報伝播現象シミュレーション

次に、現実のネットワークの静的構造に基づき、その上での情報伝播現象について、3-2 節で述べたシミュレーションによって分析を行なう。実ネットワーク上での伝播の様子を観察し、初期条件を変化させたときの内部構造の違いによる伝播過程を調べる。ネットワークサンプルとしては、4-2 節で扱った SNS のネットワーク事例を用いることにする。

5-1 シミュレーションの概要

異なる初期条件に従い状態+1と-1となるノードを決定し、ステップごとに+1の状態のノードの割合を算出する。初期状態として+1を与えるノードをランダムに決定する試行のほか、ネットワークの内部構造によって1)エッジ次数順(正順(少ない順)/逆順(多い順))、2)他ノードへの平均距離順(正順/逆順)、3)クラスター率順(正順/逆順)、の順に従って決定する試行を、それぞれ初期に状態+1を与えるノードの割合 r を0.1~0.9まで、0.1刻みで変えながら行なう。観測するステップは50ステップとする。

5-2 結果と考察

多数決モデルに従った場合の結果を図7、図8に示す。縦軸は+1の状態を持つノードの割合の推移である。図7はランダムに初期状態を与えた時の結果の例である。複数の試行によって同様の結果が得られることを確認している。図8はネットワークの内部構造によって初期状態+1を与える基準を変えた場合の結果であり、例として次数順に与えた場合の結果を示した。その他の基準によって初期状態+1を与えたものも、収束時の状態+1の割合に若干の違いはあるものの、収束までの挙動はほとんど同じであった。

多数決モデルでは、初期条件の与え方にかかわらず、その後の挙動はほとんど同じであり、どちらかが圧倒的に多数の状態に収束している。また、初期に与えるノードの状態の割合が大きく偏っていても、片方が完全に淘汰されることはなく、一定割合が残って共存の状態に収束している。

次に、重み付き多数決モデルに従った場合の結果を図9、図10に示す。図9は初期状態をランダムに与えた場合の2回の試行の結果であるが、初期状態のランダムな選択によって、 $r=0.5$ で初期状態+1を与えた場合の挙動が大きく変動している。その他の条件で初期状態+1を与えた場合でも、 $r=0.5$ の初期条件での振る舞いが大きく異なった。代表的な結果として、次数順に初期状態+1を与えたものの結果を、図10に示す。次数で重みをつけているにもかかわらず、次数の高いノードに初期状態+1を与えた時の方が収束時の割合が低いなど、いくつか興味深い結果も得られている。しかし、初期状態の割合がどちらかにわずかも偏っている場合は、どのような条件で初期状態を与えてもその後の伝播の挙動に影響を与えることはなかった。

図7. 多数決モデル・初期状態ランダム

図8. 多数決モデル・初期状態を次数順に与えた場合

図9. 重み付き多数決モデル・初期状態ランダム

図10. 重み付き多数決モデル・初期状態を次数順に与えた場合

6. おわりに

利用者ネットワークの分析により、「群」としての構造までも含めた利用者モデルを作るために、実際の Weblog の記事ネットワークと SNS のネットワークを事例として静的なネットワークの構造の分析を行なった。その結果、ネットワークとしての不均質性ともいえるべき性質が両者に共通して存在することを明らかにした。そして、その性質の背後にネットワーク生成の際に介入するヒューマンファクターの存在があるものと思われる。

また、実ネットワークをもとに、情報伝播のシミュレーションを行なった。その結果、実ネットワークにおいては、伝播の起点となるノードのネットワーク中における統計的性質の差によって、ネットワーク全体の情報伝播の挙動に差が生じる可能性があることを明らかにした。

謝辞

本発表に先立ち、議論相手として多くの有意義なアドバイスをいただいたほか、データ分析の一部を行なっていただいた東京大学大学院の柴田尚樹さん、早稲田大学国際情報通信研究センターの森祐治さん他の方々に感謝の意を表する。

参考文献

- [1] 内田誠, 白山晋: 不完全ネットワーク環境下におけるデジタル価値工学の方法論, 日本機械学会第14回設計工学・システム部門講演会講演論文集, No. 04-38, 2004, pp190-193
- [2] 増田直紀, 今野紀雄: 複雑ネットワークの科学, 産業図書, 2005
- [3] M. E. J. Newman: The structure and function of complex networks, SIAM Review, 45, 2003, pp. 167-256.