

次数分布を考慮した情報拡散の分析

○臼井翔平 鳥海不二夫 (東京大学)

Analysis of the information diffusion focusing on degree distribution

*S. Usui and F. Toriumi (University of Tokyo)

概要— ソーシャルメディアの情報拡散能力は、災害のような非常事態時での活躍が期待されている。この情報拡散能力は情報が拡散されるネットワークの構造から影響を受ける。現在、平均経路長と次数相関が情報拡散に影響を与える事がわかっている。しかし、既存研究ではネットワークの構造特徴を個別に扱い、複合的な影響は考慮していない。本研究では、複数の構造特徴が複合的に情報拡散に与える影響を明らかにするため、決定木分析を行う。決定木を構築するためには、様々な構造を持つネットワークが大量に必要である。そこで本研究では、Greedy Growth Model(GGM)を用いて様々なネットワークを生成する。その際、ネットワークの構造特徴として、4つの特徴を用いる。既存研究においてGGMに用いられている次数分布に関する特徴は、べき分布を仮定した特徴であり、次数分布がべき分布になっていなければ、当該ネットワークの特徴を十分に表す事ができない。そこで本研究では、次数分布をベータ分布の形状に近づける事によって様々な次数分布を実現する。決定木分析では平均経路長による分岐が大きく、平均経路長が最も影響を与えている事を確認した。次に、平均経路長が大きい際には、次数相関が低い値である時には高い情報拡散能力を持っている事がわかった。また、平均経路長が小さい際には、次数分布が情報拡散能力に影響を与え、次数分布が正規分布に近い時に情報拡散能力が高い事が分かった。

キーワード: 情報拡散, 複雑ネットワーク, マルチエージェントシミュレーション

1 はじめに

近年、ソーシャルメディア上での情報拡散が注目を集めており、多くのユーザがソーシャルメディアによって情報を収集・発信している。この情報収集・発信は、個人的な趣味趣向に関する情報だけに留まらず、政治や経済といった情報にまで至っている。これに伴って日本では、2013年4月19日に改正公職選挙法によって、これまで禁止されていたインターネットを用いた選挙活動の解禁を定めた。同年に行われた第23回参議院議員通常選挙においては、インターネットを利用した選挙運動を展開していた候補者は一部に留まったが、今後このような動きが拡大される事は明らかである。

ソーシャルメディアの情報拡散能力は、災害のような非常事態時での活躍が期待されている。一般に、非常事態時には、電話等が使用不可能となり、被災地との連絡が取れない事態が起こると想定されている。その際、ソーシャルメディアを介する事でインターネット回線に大きな負担をかけることなく情報拡散が可能であると考えられている。実際に2011年3月11日に日本で発生した東日本大震災時にはTwitterが情報拡散に貢献したと報告されている¹。災害のような状況下では、情報は広く拡散される事が求められる。情報の拡散は、拡散されるネットワーク(情報拡散ネットワーク)の構造に大きく影響を受けると考えられる。篠田らは、東日本大震災の際に多く利用されたTwitterに着目し、震災前と震災後のユーザ間のコミュニケーションによるネットワークを作成し、構造の変化について指摘した¹。このようなネットワークの構造変化は存在し、この構造変化は情報の拡散に影響を与えている²。現在、情報拡散に影響を与える構造特徴としては、平均経路長及び次数相関が挙げられている^{3,4}。しかし、この研究では構造特徴を個別で扱い、特徴の複合的な作用については考慮していない。また、次数分布をべき分布であると仮定しているため、べき分布でな

いネットワークに対する分析が不十分である。そこで本研究では、様々な次数分布を考慮した上で、複数特徴の情報拡散への複合的な影響を明らかにする事を目的とする。

本研究では、決定木を構築する事によって、複数の構造特徴が複合的に情報拡散に与える影響を分析する。決定木を構築するためには、様々な構造を持つネットワークが大量に必要である。そこで本研究では、Greedy Growth Modelを用いて様々なネットワークを生成する。第3章でデータの構築について述べ、第4章で決定木を構築して分析を行う。

2 関連研究

ネットワーク上での情報拡散の分析は多く行われている。Kwakらは、Twitter上で3つの影響力(retweet, follower, page-rank)を測る尺度を比較し、それによって影響力の高いユーザを発見した⁵。またChaらは、follower数の大きいユーザについては、これら以外のスコアが高い事が影響力に関係ないことを発見した⁶。Wengらは、follower数とpage-rankを比較し、page-rankをトピックに合わせて修正を行い、影響力の順位を算出した⁷。Bakshyらは、情報の拡散に貢献しているユーザ同士の関係や、拡散される範囲との関係を分析した⁸。Saitouらは、非同期時間遅れを用いて2つのモデル、非同期時間遅れ付き独立カスケードモデル(ASICモデル)と、非同期時間遅れ付き線形閾値モデル(ASLTモデル)を提案し、いくつかのトピックの情報拡散を再現した⁹。本研究もまたネットワーク上での情報拡散の分析を行う。これらの研究の着目点はネットワーク上のユーザに置かれている。一方本研究では、情報拡散が行われるネットワークの構造に着目し、ネットワークの構造特徴と情報拡散の関係について分析を行う。

3 データセットの構築

本研究では、決定木を構築する事によって複数の構造特徴が複合的に情報拡散へ与える影響を分析する。そのために、様々な構造を持つネットワークが大量に必

¹総務省：平成23年度情報通信白書、
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h23/pdf/index.html>

要である。更に、情報拡散はノード数やリンク数によって大きく変化してしまうため、ノード数やリンク数が同一のネットワークが好ましい。現実のネットワークでは、大量のデータを揃える事は困難であり、またノード数やリンク数が大きく異なってしまう。一方、Barabási Albert モデル¹⁰⁾や CNN モデル¹¹⁾のようなネットワーク生成モデルでは、同一ノード数とリンク数を持つネットワークを大量に生成する事が可能であるが、生成されるネットワーク構造に一定の傾向が表れてしまうため、様々な構造のネットワークを揃える事は困難である。そこで、本研究では任意の特徴を持つネットワークが生成可能な Greedy Growth Model(GGM)^{3,4)}を用いる。既存研究において GGM に用いられている次数分布に関する特徴は、べき分布を仮定した特徴であり、次数分布がべき分布になっていなければ、そのネットワークの特徴を十分に表す事ができない。本研究では、別の次数分布に関する特徴を用いる。

以下では、本研究で用いるネットワークの特徴量を説明し、その後、本研究で使用する GGM について解説する。

3.1 ネットワーク構造の特徴

本研究では、ネットワークの構造特徴を以下に示す特徴によりネットワークを定量的に表現する。

- (1) 平均経路長 L ¹²⁾
- (2) クラスタリング係数 C ¹²⁾
- (3) 次数相関 r ¹³⁾
- (4) 次数分布 $p(k)$

それぞれ以下で説明する。

3.1.1 平均経路長： L

平均経路長 L はネットワーク上における任意のノード間の最短経路長を平均した値を表す。ネットワーク上のノード数を N 、ノード a_i とノード a_j の間の最短経路長を l_{ij} とすると、平均経路長 L は以下の式で定義される。

$$L = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i \geq j} l_{ij} \quad (1)$$

3.1.2 クラスタリング係数： C

クラスタリング係数 C は、ネットワークの局所的な凝縮性を示す指標であり、「あるノードの隣接ノード間にリンクが存在する確率」として定義される。クラスタリング係数 C は以下の式で定義される。

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (2)$$

$$C = \frac{1}{N} \sum_i C_i \quad (3)$$

ここで、 k_i はノード a_i の次数を表し、 E_i はノード a_i に関するクラスタ数を表す。クラスタ数 E_i とはノード a_i の隣接ノード間に存在するリンクの数である。

3.1.3 次数相関 (Assortativity)： r

次数相関 r は、各リンクの終点であるノード間の次数相関を示す指標で、 $-1 \leq r \leq 1$ の範囲の値をとる。次数相関 r が正の値をとるとき、次数が高いノード同士が接続されている傾向があることを示し、 r が負の値をとるときは、次数が高いノードと低いノードが接続されている傾向があることを示す。次数相関 r は以下の式で定義される。

$$r = \frac{M^{-1} \sum_m k_m l_m - [M^{-1} \sum_m \frac{1}{2}(k_m + l_m)]^2}{\frac{1}{2}M^{-1} \sum_m (k_m^2 + l_m^2) - [M^{-1} \sum_m \frac{1}{2}(k_m + l_m)]^2} \quad (4)$$

ここで、 M はネットワーク上の全リンク数を表し、 k_m 、 l_m はリンク m によって結合している 2 つのノードの次数を表す。

3.1.4 次数分布： $p(k)$

次数分布 $p(k)$ は、ネットワーク中の次数 k であるノードの存在確率である。本研究では、 β 分布 $q(k)$ を用いる。

$$q(k) = \frac{x^\alpha (1-x)^\beta}{B(\alpha, \beta)} \quad (5)$$

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt \quad (6)$$

ここで、 α 及び β は β 分布のパラメータである。また、 x は $0 \leq x \leq 1$ であるため、 x を以下の式で与える。

$$x = \frac{k}{k_{max}} \quad (7)$$

β 分布 $q(k)$ は確率密度関数であるため、以下の処理を行う。

$$q'(k) = \frac{q(k)}{\sum_{k'} q(k')} \quad (8)$$

k_{max} は以下の式を満たす最大の値とする。

$$\sum k N q'(k) \leq 2M \quad (9)$$

なお、 N はノード数であり、 M はリンク数である。

β 分布は α 及び β を様々な値にする事で様々な形状の分布を実現する。この β 分布の形状に次数分布 $p(k)$ を近づける事で、様々な次数分布 $p(k)$ を持つネットワークを生成する。本研究ではパラメータ α 及び β を特徴として与える。

3.2 Greedy Growth Model

GGM では、ネットワーク成長モデルのプロセスを用いて、ノードの追加とリンクの追加を繰り返し行い、ネットワークを成長させていく。毎ステップのノードの追加、またはリンクの追加の際に複数のネットワーク候補を生成する。その候補の中から評価関数によって最良のものを選択し、これを次のステップのネットワークとして採用する。これを繰り返し行うことによって要求した特徴に近づきながらネットワークを成長させる。

3.2.1 ネットワーク成長モデル

一般的に、ネットワーク成長モデルでは、毎ステップ以下のいずれかの操作を行い、ネットワークを徐々に成長させていく。

- ノードの追加
- リンクの追加

本モデルではノードの追加を最初に行い、ノード数が必要数に達した時点で、リンクの追加を行う。リンク数が必要数まで達した時点で終了する。

3.2.2 候補ネットワークの生成

ノードの追加とリンクの追加のそれぞれにおいて、候補ネットワークを生成する手法について解説する。

ノードの追加 ノードの追加では、ネットワークに新規ノードを1つ追加する。さらに既存のノードの中から1つのノードを選択し、追加した新規ノードと選択した既存ノードの間にリンクを1つ生成する。この際既存のネットワークからランダムにノードを選択する。選択されたノードと新規ノードの間にリンクを生成したネットワークを候補ネットワークとする。この操作を m 回繰り返す事によって、 m 個の候補ネットワークを生成する。

リンクの追加 リンクの追加では既存のノードの中から2つのノードを選択し、その2つのノードの間にリンクを1つ生成する。まず、既存のノードから1つのノードをランダムに選択する。次に、2つ目のノードはランダム選択か、一つ目のノードから2ステップ先のノードの中からランダムで選択する。2つの選択方法は同じ確率で行われるようにする。この操作によって選択された2つのノード間にリンクを結ぶ事によって候補ネットワークを生成する。この操作を m 回繰り返す事によって、 m 個の候補ネットワークを生成する。

3.2.3 評価関数

生成した候補ネットワークの中から要求指標値に近いネットワークを選択するためには、ネットワーク間の指標値の差を定量的に評価する必要がある。そこでまず、使用する指標に関する距離を以下のようにそれぞれ定義する¹⁴⁾。

$$D_L = \left(\frac{L_i - L_a}{\sigma_L}\right)^2, D_C = \left(\frac{C_i - C_a}{\sigma_C}\right)^2 \quad (10)$$

$$D_r = \left(\frac{r_i - r_a}{\sigma_r}\right)^2 \quad (11)$$

また、生成中のネットワークの次数分布 $p(k)$ と β 分布 $p'(k)$ との距離をユークリッド距離によって定義する。

$$D_{p(k)} = \frac{\sum (p(k) - q'(k))^2}{\sigma_{p(k)}^2} \quad (12)$$

ただし、 L_i , C_i , r_i , $p(k)$ はそれぞれ生成中のネットワークの指標値であり、 L_a , C_a , r_a , $q'(k)$ はそれぞれ要求指標値である。各項の分母 $\sigma_L, \sigma_C, \sigma_r, \sigma_{p(k)}$ は各指標の標準偏差を表す。標準偏差は本章の検証に用い

Table 1: 各特徴の値域

特徴	値域
L	$1 \leq L \leq 10$
C	$0 \leq C \leq 1$
r	$-1 \leq r \leq 1$
α	$0 < \alpha \leq 10$
β	$0 < \beta \leq 10$

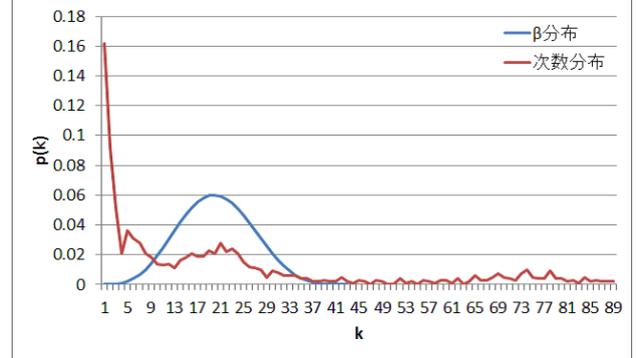


Fig. 1: $D_{p(k)} = 0.239$ の次数分布と β 分布

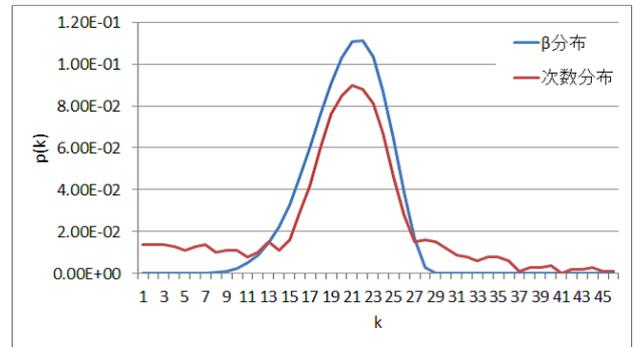


Fig. 2: $D_{p(k)} = 0.0970$ の次数分布と β 分布

るデータセットから算出する。さらに、それぞれの指標の距離の和をネットワーク間相異度 D とする。

$$D = \sum_j D_j \quad (13)$$

D の値が小さい程、類似した特徴を持つネットワークであると言える。

3.3 データセットの作成

ここでは GGM を用いてネットワークのデータセットを作成する。各特徴の値域の中でランダムに値を設定し、その特徴値を持つネットワークを GGM によって生成する。それぞれの特徴の値域を **Table 1** に示す。なお、ノード数は 1000、リンク数は 10000 とする。

生成されたネットワークの次数分布の特徴として、 α 及び β を用いるためには、生成されたネットワークの次数分布と β 分布 $q'(k)$ がある程度近い必要がある。**Fig. 1** に $D_{p(k)} = 0.239$ の次数分布と β 分布を示す。この時、 β 分布は次数分布の特徴を捉えているとは言えず、 α と β はネットワークの特徴とは言えない。そこで、 $D_{p(k)} \leq 0.1$ であるネットワークを用いる。**Fig. 2** に、 $D_{p(k)} = 0.0970$ であるネットワークの次数分布とその時の β 分布をそれぞれ示す。**Fig. 2** より、生成されたネットワークの次数分布は β 分布の特徴をとらえている事がわかる。実際に用いるデータはこの例よりも類似した次数分布を持つネットワークである。

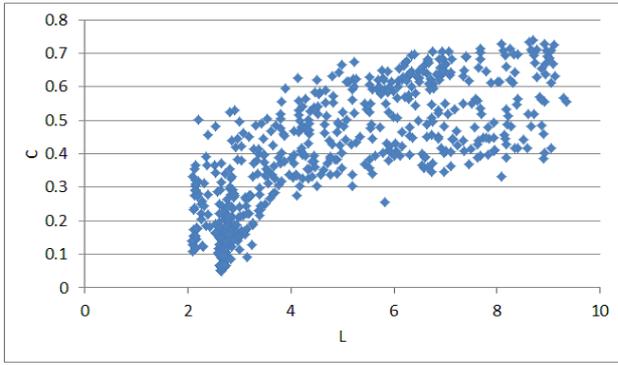


Fig. 3: 平均経路長とクラスタ係数の関係

Table 2: 指標間の相関係数

	C	r	α	β
L	0.770	-0.167	0.0478	0.112
C	-	-0.0919	-0.0695	0.155
r	-	-	-0.0136	-0.0290
α	-	-	-	0.179

3.4 特徴間の相関関係

ここでは、各特徴間に関係がないことを確認する。Fig. 2に、各指標間の相関係数を示す。これより、平均経路長 L とクラスタ係数 C の間に大きな相関がある事がわかる。Fig. 3に平均経路長 L とクラスタ係数 C の関係を示す。データセットの生成の際には平均経路長とクラスタ係数はランダムな値を設定し、その値に近い特徴を持つようにネットワークを生成している。そのため、このような偏りは発生しづらい。しかし、 $D_p(k) \leq 0.1$ にデータを絞った時に偏りが発生したものと考えられる。つまり、次数分布を決定した状態では平均経路長とクラスタ係数に相関が生まれてしまう事がわかる。これが、ネットワーク生成モデルに起因するものなのか、ネットワークの特性なのかは定かではないため、今後の課題とする。

本研究では、平均経路長とクラスタ係数に強い相関関係がある事から、2つの内平均経路長のみを用いる。

4 情報拡散に影響する特徴の分析

ここでは、前章で得られたデータを用いて決定木を構築し、複数の構造特徴が複合的に情報拡散へ与える影響の分析を行う。決定木は、ある事項に対する観測値からその事項の目的変数に関する結論を導くものである。情報拡散の能力を目的変数として、決定木を構築する事によって、複数のネットワーク特徴がどのような値を取っていれば情報拡散能力が高いのかを分析する事が可能である。本章ではまず、使用するデータについて説明を行い、その後作成した決定木を用いて分析を行う。

4.1 決定木の構築

4.1.1 データセット

ネットワークは第3章で生成したネットワーク796個を用いる。ネットワークの持つ特徴として、以下の5つを用いる。

- 平均経路長 L ²⁾
- 同類選択性 r ¹³⁾

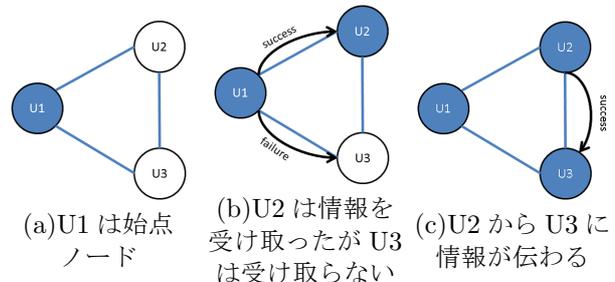


Fig. 4: ICモデルの操作

- β 分布のパラメータ α
- β 分布のパラメータ β

さらにこのネットワークに対して情報拡散シミュレーションを行い、ネットワークの情報拡散能力を測定する。以下では情報拡散シミュレーションに用いる情報拡散モデルを解説し、その後ネットワークの情報拡散能力を定義する。

4.1.2 情報拡散モデル

情報拡散シミュレーションは情報拡散モデルとして、ICモデルを用いる¹⁵⁾。ICモデルは、感染症の広がり方を示す基本的な確率モデルである。Fig. 4にICモデルの基本的な流れを示す。U1が始点ノードのとき、隣接ノードであるU2とU3に対し1度だけ情報を発信する。この際U1はノード間のリンクに設定された発信確率に従って情報を発信するかどうかを決定する。ユーザに情報が伝わると、受信ユーザが発信源となり、再び1度だけ情報発信を行う。この操作を繰り返すことにより情報拡散を表現する。

4.1.3 情報拡散能力を表す指標

ここではネットワークにおける情報拡散のしやすさを定量的に表す指標として、Ability of Information Diffusion(AID)⁴⁾を用いる。ネットワーク上のあるノード v_i から情報拡散を開始し、情報が伝わったノードの割合を $\delta(v)$ とする。さらに、全てのノードから情報拡散を開始し、 $\delta(v_i)$ の平均をとった値をネットワークの情報拡散能力(AID) δ として以下の式で定義する。

$$\delta = \frac{\sum \delta(v_i)}{N} \quad (14)$$

AIDが高い程、ネットワーク上での情報拡散が行われやすい。なお、 $\delta(v_i)$ を求める際には情報拡散シミュレーションを100回を行い、平均を算出する。

4.1.4 決定木

AIDを目的変数として、クラスタ係数を除いた4つの特徴を説明変数として決定木を構築する。構築された決定木をFig. 5に示す。なお、枝は左がtrue, 右がfalseとなっている。また葉の先端には集合内のAIDの平均を示し、 n はその集合に含まれるデータ数を表す。

4.2 決定木分析

まず最も根の部分で平均経路長 L によって分岐している。したがって、平均経路長 L が最も作用している事がわかる。これはDoerrらの知見と一致する³⁾。 $\alpha < 4.813$ の集合のAIDの平均は0.521であるのに対し、 $4.813 \leq L$ の集合のAID平均は0.114である。

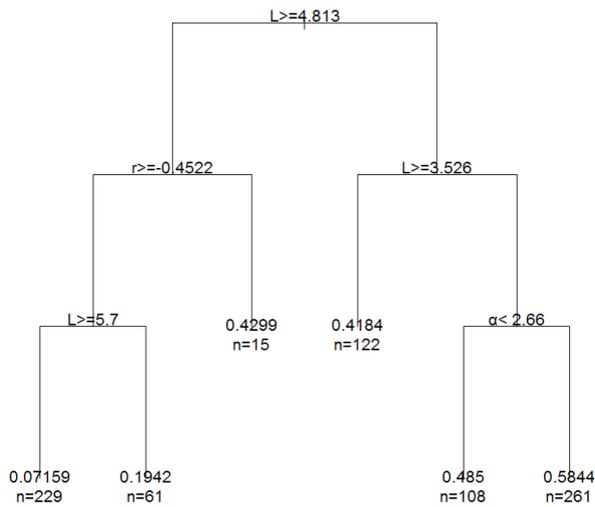
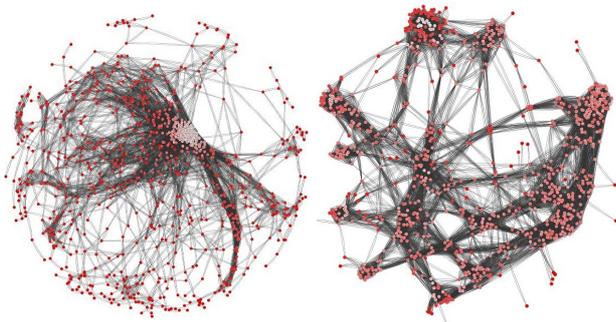


Fig. 5: 決定木



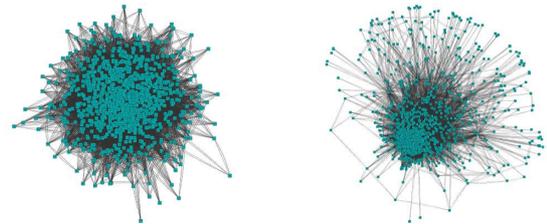
(a) $L = 5.19, r = 0.810, \delta = 0.199$

(b) $L = 4.99, r = -0.536, \delta = 0.589$

Fig. 6: ネットワーク例

$\alpha < 4.813$ であるネットワークの方が非常に高い AID を持っている事がわかる。

しかし、 $4.813 \leq L$ なネットワークであっても $r \leq -0.4522$, つまり disassortative なネットワークであれば高い AID を示している。 $4.813 \leq L$ における $r \leq -0.4522$ の集合の AID 平均は 0.430 であるのに対し、 $-0.4522 \leq r$ であるネットワークの AID 平均は 0.0974 と非常に低い。 したがって、平均経路長 L が高い値である時には、次数相関 r は小さい値である方が AID が高い事がわかる。 この知見は、以前に得られている知見と異なる⁴⁾。 この原因は、既存研究が平均経路長を小さい値に固定して行っているためであると考えられる。 ここで、Fig. 6(a) に $4.813 \leq L$ かつ $-0.4522 \leq r$ の集合に属するネットワークを、Fig. 6(a) に $4.813 \leq L$ かつ $r \leq -0.4522$ の集合に属するネットワークをそれぞれ示す。 なお Fig. 6 では、次数の大きいノードを白で、次数の小さいノードを赤であらわしている。 Fig. 6(a) のネットワークでは次数の大きいノードが集中して存在し、それ以外の箇所では次数の小さいノードが多く存在している事がわかる。 このネットワークでは情報は次数の高いノードまで達する事がなければ拡散されにくいと考えられる。 これに対し、Fig. 6(b) のネットワークでは、次数の大きいノードを中心としたコミュニティ構造が見られる。 また、そのコミュニティ間を次数の小さいノードがつないでいる。 このような次数の高いノードと次数の低いノードの接続によって、



(c) $\alpha = 4.83, \beta = 9.44, \delta = 0.592$ (d) $\alpha = 1.10, \beta = 3.11, \delta = 0.429$

Fig. 7: ネットワーク例

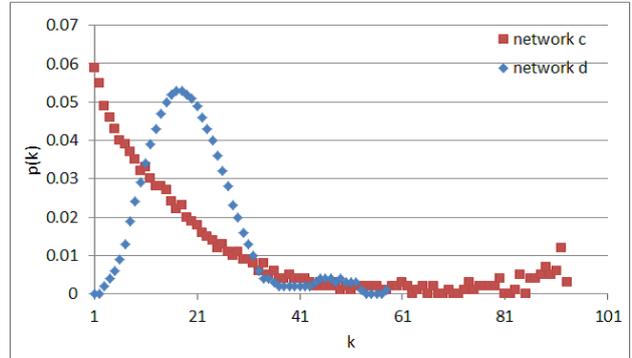


Fig. 8: ネットワーク例の次数分布

次数相関は小さくになると考えられる。 このネットワークでは次数の大きいハブノードへ到達するのが容易であるため、情報拡散能力が高いと考えられる。 したがって、平均経路長が大きくても、コミュニティ構造を有している事で、情報拡散能力が高くなる事がわかった。

さらに、Fig. 5 から最左の葉である $4.813 \leq L, r \leq -0.4522$ における $6.464 \leq L$ の集合が AID 平均 0.0716 で最小をとる集合となっている事がわかる。 一方、 $L \leq 3.53$ の集合は AID 平均 0.555 と非常に高い AID をもつ事から、平均経路長 L は短い程高い AID の値をとる事がわかる。

次に、 $L \leq 3.53$ の集合を見る。 ここでは α による分岐が存在し、 $\alpha < 2.66$ の集合の AID 平均は 0.485 であり、 $2.66 \leq \alpha$ の集合の AID 平均は 0.584 となっている。 また、全体の中の $\alpha < 2.66$ の AID 平均は 0.337 であり、 $2.66 \leq \alpha$ の集合の AID 平均は 0.300 であった。 したがって、 α が大きい方が AID が高い事を表している。 α が大きいと、ネットワークの次数分布は正規分布に近い分布をとる。 一方、 α が小さいとネットワークの次数分布はべき分布に近い分布をとる。 Fig. 7 にそれぞれの集合に属するネットワークの例を、そして Fig. 8 にその次数分布を示す。 したがって、 α の値が大きい時、つまり次数分布が正規分布に近い方が、AID が高くなる傾向にある事がわかる。

最後に、今回の分析では、次数相関が高い程 AID が高いという現象が見られなかった⁴⁾。 既存研究との大きな違いは、次数分布の特徴である。 本研究では、分布間距離を用いているのに対し、既存研究ではべき分布決定係数とべき指数を用いている。 べき分布決定係数とべき指数を用いた場合、べき分布決定係数が非常に高い場合は次数分布は一意に決定するが、べき分布決定係数が低い時には、分布は一意に決定しない。 つまり、べき分布決定係数とべき指数を用いている場合、次数分布の動きはある程度許容している。 それに対し、分布間距離を用いた本手法では、次数分布は設定した

β 分布に類似した分布となる。したがって、次数関連の変化による次数分布の変化が影響を与えたのではないかと考える。

5 まとめ

本研究では、複数の構造特徴が複合的に情報拡散に与える影響を明らかにするため、決定木分析を行なった。決定木を構築するためには、様々な構造を持つネットワークが大量に必要である。そこで本研究では、Greedy Growth Model(GGM)を用いて様々なネットワークを生成した。その際、ネットワークの構造特徴として、4つの特徴を用いる。既存研究においてGGMに用いられている次数分布に関する特徴は、べき分布を仮定した特徴であり、次数分布がべき分布になっていなければ、そのネットワークの特徴を十分に表す事ができない。そこで本研究では、次数分布をベータ分布の形状に近づける事によって様々な次数分布を実現した。まず、それぞれの特徴間の関係について分析をした。すると、平均経路長とクラスタ係数の間に強い相関が出た。そこで、クラスタ係数を除く3つの特徴を用いて決定木を構築した。決定木分析では平均経路長による分岐が大きく、平均経路長が最も影響を与えている事を確認した。次に、平均経路長が大きい際には、次数相関が低い値である時には高い情報拡散能力を持っている事がわかった。また、平均経路長が小さい際には、次数分布が情報拡散能力に影響を与え、次数分布が正規分布に近い時に情報拡散能力が高い事が分かった。

今後の課題としては、本研究で用いた4つの特徴以外の特徴も用いた分析を行う事を挙げる。具体的には、コミュニティに関する特徴を加える事を挙げる。次に、本研究では平均経路長とクラスタ係数の間に相関が存在した。これが、ネットワーク生成モデルに起因するものなのか、ネットワークの特性なのかは定かではないため、今後の課題とする。

参考文献

- 1) 篠田孝祐, 榊剛史, 鳥海不二夫, 風間一洋, 栗原聡, 野田五十樹, 松尾豊: 東日本大震災時におけるtwitterの活用状況とコミュニケーション構造の分析, 知能と情報, **25**-1, 598/608 (2013)
- 2) 白井翔平, 鳥海不二夫, 平山高嗣, 榎堀優, 間瀬健二: なぜ震災後デマが拡散したのか: ネットワーク構造の影響分析一, 電気学会論文誌C (電子・情報・システム部門誌), **133**-9, 1796/1805 (2013).
- 3) Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich: Why rumors spread so quickly in social networks, *Commun. ACM*, **55**-6, 70/75 (2012)
- 4) 白井翔平, 鳥海不二夫: 情報拡散に影響するネットワーク構造特徴. 人工知能学会論文誌, **30**-1, 195/203 (2015)
- 5) Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon: What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, WWW '10, pp. 591–600, New York, NY, USA (2010) ACM.
- 6) Meeyoung Cha, Hamed Haddadi, Fabricio Benvenuto, and Krishna P. Gummadi: Measuring user influence in twitter: The million follower fallacy, In *in ICWSM10: Proceedings of international AAAI Conference on Weblogs and Social* (2010).
- 7) Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He: Twiterrank: finding topic-sensitive influential twitterers, In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, 261/270, New York, NY, USA (2010). ACM.
- 8) Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts: Everyone's an influencer: quantifying influence on twitter, In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, 65/74, New York, NY, USA (2011). ACM.
- 9) Kazumi Saito, Masahiro Kimura, Kouzou Ohara, and Hiroshi Motoda: Learning asynchronous-time information diffusion models and its application to behavioral data analysis over social networks, *CoRR*, Vol. abs/1204.4528 (2012)
- 10) Albert-Laszlo Barabasi and Reka Albert: Emergence of scaling in random networks, *Science*, **286**-5439, 509/512 (1999)
- 11) Alexei Vazquez: Growing networks with local rules: preferential attachment, clustering hi-erarchy and degree correlations, *arXiv:cond-mat/0211528 v2* (2003)
- 12) D.J. Watts and S.H. Strogatz: Collective dynamics of 'small-world' networks, *Nature*, No. 393, pp. 440–442, (1998).
- 13) M. E. J. Newman: Mixing patterns in networks, *Phys. Rev. E*, **67**-2, 026126, (2003)
- 14) 鳥海不二夫, 石田健, 石井健一郎: Snsにおけるネットワーク成長モデルの提案(コミュニティ形成支援, インターネット技術とその応用論文), 電子情報通信学会論文誌. D, 情報・システム, **93**-7, 1135/1143 (2010)
- 15) Jacob Goldenberg, Barak Libai, and Eitan Muller: Talk of the network: A complex systems look at the underlying process of word-of-mouth, *Marketing Letters* (2001)