

# 潜在トピックに基づく複雑ネットワーク生成における クラスタ性に関する評価

赤山 郁人<sup>a</sup> 倉持 俊也 土方 嘉徳<sup>b</sup>

大阪大学大学院基礎工学研究科

a) akayama@hlab.sys.es.osaka-u.ac.jp b) hijikata@sys.es.osaka-u.ac.jp

**概要** ソーシャルネットワーキングサービスにおいて、人はその人の興味のあるトピックに基づいてフォロー先を決定していると思われる。我々は、ノードが潜在的なトピックに基づき、有向リンクのリンク先を決定する考えに基づく複雑ネットワークの生成モデルを提案する。本稿では、まず計算機によるシミュレーションを行い、その結果が複雑ネットワークの様々な性質を満たしているかどうかを議論する。また、ネットワークを評価する指標である平均経路長、クラスタ係数、べき指数を計算し、得られた値について分析を行った。さらに、提案モデルが生成するネットワークを可視化し、考察を行った。

**キーワード** 複雑ネットワーク, トピックモデル, シミュレーション, ネットワーク生成モデル

## 1 はじめに

近年多くの研究者によって World Wide Web や人の社会ネットワーク、論文の被引用関係のネットワークなどのグラフ構造に対する解析が行われている。それらは複雑ネットワークと呼ばれ、多数の因子が影響しあうことで全体の性質が決定されるネットワークである。多くの複雑ネットワークには、いくつか共通する特徴がみられ、中でもスモールワールドとスケールフリーはよく知られた特徴である [1]。スモールワールドは、ノード間の平均経路長が短く、クラスタが多数生成されているネットワークとして定義される [2]。またスケールフリーネットワークは、数学的にはノードの次数が  $k$  である確率分布がロングテイル、すなわちべき分布  $P(k) \sim k^{-\gamma}$  (ただし  $\gamma$  は定数) に従うグラフである [3]。

これまで、これらの性質を作り出すネットワークの振る舞いに対する知見を得るために、ネットワークを人工的に生成するネットワーク生成モデルが幾つも考えられてきた [5, 6, 7, 8, 9]。BA モデルでは、ノードの次数に基づいたエッジの優先的選択によりネットワークを成長させることで、スケールフリーなネットワークが生成されることが明らかになっている [1, 10]。他にも、スモールワールド・ネットワークの生成を説明するモデルが提案されている [4]。従来のモデルは、ネットワークにおける構造上の特徴を生成するためにエッジが生成される過程をモデル化したものであり、ネットワーク全体を統一的な始点からとらえたモデルである。これに対し我々は、ネットワークにおける個々のエッジが持つ意味や役割に着目した。例えば人と人が関わりを持つ時、2人は所属しているコミュニティが同じである場合や、共通する趣味・興味を持っていることがある。特にオンラインで

の人と人との繋がりに焦点を当てると、Twitter のフォローネットワークは社会的な知人関係だけではなく、趣味や関心の関係も表していると報告されている [11]。また、論文の引用関係においてもトピックに依存した分析が行われている [12]。ネットワークの成長に2つのノードの間に意味持たせたエッジ生成の過程を導入することは、実ネットワークの成長を模倣するという点で必要であると考えられる。

そこで我々は、エッジが生成されている2つのノードの間には、共通の潜在的なトピックが存在している可能性に着目し、トピックに基づいてエッジを生成するネットワーク生成モデルを提案する。すなわち、ノードはどのトピックを選びやすいかという確率分布を持っており、それにしたがってトピックを選択する。さらに、トピックはどのノードを選びやすいかという確率分布を持っており、それにしたがってノードを選択する。このようにしてノードからノードへのエッジが生成されるモデルを提案する。これは、自然言語処理の分野において有名なトピックモデルである LDA (Latent Dirichlet Allocation) [13] が前提とする文書が生成される過程と同じである。そのため、我々の提案モデルには LDA で行われる計算処理を用いる。

本研究では、我々が提案するモデルを示し、計算機によるネットワーク生成のシミュレーションを行う。そして、提案するモデルが多様な複雑ネットワークを生成できるかを議論する。また、提案モデルのパラメータを変更することにより、生成されるネットワークの特徴の程度を変更することができるかを検証する。さらに、我々の提案モデルが生成するネットワークと他のネットワーク生成モデルが生成するネットワークを可視化して比較し、提案モデルが生成するネットワークのノードのまと

まり具合を中心に議論する。

2章では関連研究を述べ、3章で提案モデルのパラメータを変更したときの定量的な評価を行う。また4章では既存モデルと提案モデルの比較を行い、5章で結論を述べる。

## 2 関連研究

### 2.1 ネットワーク生成モデル

ネットワークが成長していく過程で、何が複雑ネットワークの共通する特徴を創り出すのかということに注目した研究が行われ、スモールワールドなネットワークを生成する WS モデルが提案された [4]。これを皮切りに、共通する特徴を作り出すという目的で、ネットワークの成長過程に着目した研究が行われ、スケールフリーなネットワークを生成する BA モデル [3] が提案された。これらのモデルは複雑ネットワークの研究において先駆的なものであるが、不自然な点も存在する。WS モデルでは格子状のネットワークのノード間のエッジを張り替えるというアルゴリズムであるが、一般的に現実のネットワークでエッジの張り替えは起こりえない。また BA モデルにより生成されるネットワークはクラスタ性を満たさない。クラスタ性を満たさないという問題を解決するために、Holme-Kim モデル [6] や適応度モデル [7] が提案された。これらは接続先をネットワーク全体のノードから選択することを前提としたモデルである。

一方で、各ノードを意思決定を行う主体とみなし、新しいエッジの生成はノードの投票によって行われるモデル [14] や、「友達の友達にエッジを張る」というルールに基づく CNN モデル [15] も提案されている。これらのモデルは、あるノードからみてネットワークの近傍から、接続先のノードを選択することを前提としたモデルである。

この研究ではネットワークはノードの次数やノード間の距離に基づいて生成されると考えられているが、本研究ではネットワークは個々のノードが持つ潜在トピックに基づいて生成されると考える。

### 2.2 LDA のネットワークへの応用

LDA は自然言語処理の分野でよく用いられる機械学習の手法である。近年 LDA をネットワークデータに応用したアプローチの研究が進められており、LDA を使ったコミュニティ抽出が盛んに行われている。Henderson らの提案している LDA-G は、ノード  $v$  の隣接ノード集合を文書  $d$  の単語集合とみなした LDA であり、LDA で使われる機械学習のアルゴリズムを使うことで、コミュニティの発見を行った [16]。また、Cha らは同様のアイデアに基づいて LDA を Twitter のフォローネットワークに適用し、有向エッジに対してユーザの興味を

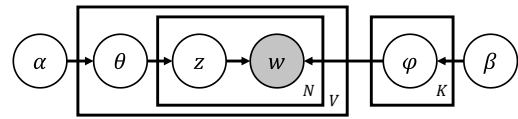


図1: 提案モデルを表すグラフィカルモデル

表すラベル付けを行っている [17]。このように LDA がネットワーク上のトピックを推定することが可能であり、LDA の前提とする文書を生成する過程を満たしていると言える。

本研究では、ノードがトピックを介してノードを選択することで生成されるネットワークがどのようなものを調査する。

## 3 提案モデル

図1に潜在的トピックに基づいたエッジを生成する過程を示す。本研究では、図1の過程を提案し、これを導入した新たな複雑ネットワークの生成モデルを提案モデルとして定義する。図1中の  $\alpha, \beta$  はハイパーパラメータ、 $w$  はノード、 $z$  はトピックを表し、 $\theta$  はノードが選択するトピックの確率分布、 $\phi$  はトピックが選択するノードの確率分布をそれぞれ表す。

ノード  $v$  がどのトピックを選択しやすいかを表す確率分布は  $\theta_v$  と表され、ノード  $v$  がトピック  $k$  を選択する確率は (1) 式で表される。

$$\theta_{v,k} = \frac{n_{v,k} + \alpha}{\sum_j (n_{v,j} + \alpha)} \quad (1)$$

(1) 式の  $n_{v,k}, n_{v,j}$  はノード  $v$  がトピック  $k, j$  を選択した回数、 $\alpha$  はハイパーパラメータを表す。またトピック  $z$  がどのノードを選択しやすいかを表す確率分布は  $\phi_z$  と表され、トピック  $z$  がノード  $w$  を選択する確率は (2) 式で表される。

$$\phi_{z,w} = \frac{n_{z,w} + \beta}{\sum_i (n_{z,i} + \beta)} \quad (2)$$

(2) 式の  $n_{z,w}, n_{z,i}$  はトピック  $z$  がノード  $w, i$  を選択した回数、 $\beta$  はハイパーパラメータを表す。

我々の提案モデルでは、全てのノードはどのトピックを選択しやすいかを表す確率分布  $\theta$  と各トピックはどのノードを選択しやすいかを表す確率分布  $\phi$  を持っている。ノード  $v$  から  $w$  へのエッジを生成する時、ノード  $v$  は  $\theta_v$  に基づいた優先的選択でトピック  $z$  を選択する。トピック  $z$  が定まった時、トピック  $z$  は  $\phi_z$  に基づいた優先的選択でノード  $w$  を選択する。このプロセスに従うことでノード  $v$  からノード  $w$  へのエッジがトピック  $z$  という属性を持って生成される。

---

アルゴリズム 1 : 提案モデル

---

Start with  $K_m$  complete graph  $G(V, E)$

**repeat**

    Generate  $P \in [0, 1]$  randomly.

**if** Probability  $p \leq P$  **then**

        Add node  $v_{new}$  to  $V$ ,

$v_{new}$  has no edges.

$n \leftarrow n + 1$

**for**  $i = 0$  to  $m$  **do**

            Select topic  $z$  distributed under  $\theta_{v_{new}}$ .

            Select node  $v_{old}$  distributed under  $\phi_z$ .

            Add edge  $e_{new} = (v_{new}, v_{old})$  to  $E$ .

**end for**

**else**

        Select node  $x$  distributed under out-degree.

        Select topic  $z$  distributed under  $\theta_x$ .

        Select node  $y$  distributed under  $\phi_z$ .

        Add edge  $e_{new} = (x, y)$  to  $E$ .

**end if**

**until**  $n = \#$  of node.

---

また、図1の過程で得られるエッジには向きが存在する。始点となるノード  $v$  から端点となるノード  $v$  へのエッジが提案モデルでは生成されるため、得られるネットワークは有向グラフとなる。

提案モデルの詳細をアルゴリズム1に示す。このモデルでは初期状態のネットワークは  $m$  個のノードからなる完全グラフ  $K_m$  としている。またアルゴリズム中の  $n$  はノードの番号を示し、 $N$  はアルゴリズムを停止する時のノード数である。ネットワークに新たにノードを生成した時、ノードは  $m$  本のエッジを同時に生成する。これは、BAモデル [3] を模倣したものとなっている。また、ノード生成時のエッジ生成とは別に、エッジのみを生成するプロセスも存在する。これは、現実のネットワークによく見られる行動をモデル化したものである。例えば、Webサイトを新たに作ってしばらく時間が経った時、Webサイトの管理者が自分のWebサイトと関連があるサイト、あるいは興味のあるサイトに対するリンクをWebサイトの中に掲載することがある。またSNSにおいても、SNSに新規登録をしてしばらく時間が経った時、ユーザが自分の興味のあるトピックの中で気になるユーザを見つけると、そのユーザをフォローすることがある。このような状況が考えられるため、ネットワーク中のノードが逐次他のノードへのエッジを生成する過程をモデルに組み込んだ。このエッジ生成の過程でも、ノードの持つ潜在的なトピックに基づいてエッジを生成する。

## 4 シミュレーション実験

提案モデルによって生成されるネットワークが複雑ネットワークの性質をみたすのか否か、提案モデルのパラメータを変更することで生成されるネットワークの特徴の程度を制御できるかを検証するためにシミュレーション実験を行った。

### 4.1 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の特徴量を用いて、ネットワークを定量的に表す。本研究で用いた特徴量は、平均経路長、クラスタ係数、べき指数、モジュラリティである。以下にそれぞれの計算方法を示す。

#### 4.1.1 平均経路長 [10]

任意のノード間の最短経路長の平均値である。この値が小さいほどネットワークは強いスモールワールド性を持つといえる。接続されているノードのグラフ上の距離は1であり、次の式で与えられる。

$$L = \frac{1}{N} \sum_{i \neq j} \min(L(i, j)) \quad (3)$$

$L(i, j)$  はノード  $i$  からノード  $j$  への経路長で、 $N$  は経路が存在したノードの組み合わせの個数である。ただし、本研究で扱うグラフは有向グラフであるため、到達不可能なノードの組み合わせが存在することに注意する。本研究では、到達可能なノードの組のみを用いて、上記の値を算出する。

#### 4.1.2 クラスタ係数 [18]

異なる3つのノードが互いに繋がっている状態のものをクラスタとみなし、ノードがクラスタの一部となる確率の平均を表す。この値が大きいほど、ネットワークは強いクラスタ性(ノード間の結びつきが強い状態のこと。例えば、人の社会ネットワークでは、任意の2人を取り出したときに、共通の友人がいる可能性が高い状態のこと)を持つと言える。ただし、クラスタ係数は無向グラフにおいて定義されているため、有向グラフでも計算可能なよう拡張された計算方法を用いる [19]。ネットワークのノードの総数を  $N$ 、ノードとノードの接続状態を表す隣接行列を  $A$ 、その  $(i, j)$  成分を  $a_{ij}$  とすると、次の式で与えられる。

$$d_i^{\leftrightarrow} = \sum_j a_{ij} a_{ji} \quad (4)$$

$$d_i^{total} = \sum_j (a_{ij} + a_{ji}) \quad (5)$$

$$C = \frac{1}{N} \sum_i \frac{\frac{1}{2} \sum_j \sum_h (a_{ij} + a_{ji})(a_{ih} + a_{hi})(a_{jh} + a_{hj})}{[d_i^{total}(d_i^{total} - 1) - 2d_i^{\leftrightarrow}]} \quad (6)$$

表 1: 各パラメータの基準となる値

変更可能なパラメータ	値
トピック数 $K$	3
初期シードの数 $m_0$	3
ノード生成過程のエッジ生成本数 $m$	1
ノード生成過程とエッジ生成過程の分岐確率 $p$	0.5
ハイパーパラメータ $\alpha$	0.05
ハイパーパラメータ $\beta$	0.05

#### 4.1.3 べき指数 $\gamma_{in}, \gamma_{out}$ [3]

次数分布がべき分布  $P(k) \sim k^{-\gamma}$  ( $\gamma$  は定数) に従うと仮定した時の  $\gamma$  を表す。有向グラフでは、ネットワークの入次数、出次数ごとに、べき指数を計算する。この値が大きいくほど、より少数のノードが大きな次数を持つことを示す。従来研究によるとべき指数を計算する方法は、1) 次数分布の両対数を取ったものへ直接線形回帰を適用したもの [1, 10], 2) 両対数を取ったヒストグラムの最初の 5 点を用いて線形回帰を適用したもの [20], 3) 両対数を取った時に正の実数になるものだけを利用して線形回帰を適用したもの [3, 21] の 3 つがある。1) と 3) の方法では、次数分布における次数の高い部分の分布によって推定する直線が大きいくずれの可能性があり、それを改善するための手法も提案されている [22]。しかし、ここでは簡単のため 2) の方法を使ってべき指数の推定を行う。なぜなら、ネットワーク中の大半のノードは次数が低く、分布の中で次数が低い部分は多くのノードの情報を含んでいるため、次数分布の大半をカバーできるからである。

#### 4.2 シミュレーション実験の環境

提案モデルにおけるパラメータは、トピックの数を表す  $K$ 、ハイパーパラメータ  $\alpha, \beta$ 、初期シードの数  $m_0$ 、ノード生成過程におけるエッジ生成本数  $m$ 、ノード生成過程とエッジ生成過程の分岐確率  $p$ 、目標ノード数  $N$  がある。この実験では、目標ノード数  $N$  を 2000 とする。またこの実験では、各パラメータによりどの程度、複雑ネットワークを評価するための特徴量を変化させることができるかを調べる。そこで、他のパラメータは固定にしておき、注目するパラメータのみを変動させて、特徴量の値を算出する。パラメータを固定する際の基準となる値を表 1 のように設定した。

これらのうち、ネットワークの構造に大きく影響を及ぼすのはトピック数  $K$ 、ハイパーパラメータ  $\alpha, \beta$  であると考えられる。 $K, \alpha, \beta$  は本研究の核心部であるトピックに関するパラメータであり、後に与える影響が大きいくと考えるからである、従ってそれらのパラメータの値を変動させ、生成されたネットワークのsmallワールド

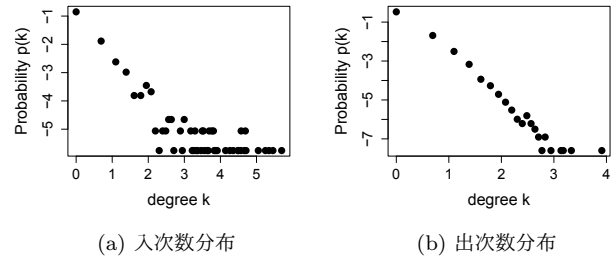


図 2: 次数分布の例 (パラメータは全ての基準値に設定)

表 2: トピック数  $K$  を変更したときの結果

$K$	$\gamma_{in}$	$\gamma_{out}$	$L$	$C$
1	1.52	2.00	3.01	0.22
3	1.53	2.00	3.65	0.088
10	1.48	1.99	4.64	0.025
50	1.26	1.99	5.92	0.0050

な性質やスケールフリーな性質の程度を変更できるかを確認するための実験を行った。また 1 つのパラメータの組み合わせにつき 100 回行い、その状態におけるネットワーク構造の特徴量の平均を計算した。

#### 4.3 実験の結果と考察

図 2 では全てのパラメータを基準となる値に設定した時の入次数と出次数の分布を表している。図から入・出次数の分布がべき分布に従っていることがわかる。べき分布に従っている一方で、次数が高い部分で分布が横に広がりを見せている。これは高い次数のノードが偶然 1 個ないし 2 個生成されていたためである。これはネットワークを構成するノードの数を増やすことで横への広がりが更に右に、つまり高次数へとシフトするものと考えられる。いずれのパラメータの組み合わせにおいても、傾きは異なるがこのようなべき分布が確認された。このことから我々が提案するネットワーク生成モデルはスケールフリーなネットワークを生成できると言える。

トピック数  $K$  を変動させたときのネットワーク構造の特徴量を表 2 に示す。表からトピック数を増やすと平均経路長  $L$  の値が大きくなりクラスタ係数  $C$  が小さくなることがわかる。一般に複雑ネットワークはノードの数に比べて小さい値になり、比較的大きなクラスタ係数になることが知られている [24, 4]。従って我々の提案モデルではトピック数  $K$  が小さい時にsmallワールドなネットワークを生成しうることが分かる。また、トピック数を変動させることで平均経路長とクラスタ係

表 3: ハイパーパラメータ  $\alpha, \beta$  を変更したときの結果

$\alpha$	$\gamma_{in}$	$\gamma_{out}$	$L$	$C$
0.01	1.55	2.00	3.70	0.084
0.05	1.53	2.00	3.65	0.088
0.1	1.53	2.00	3.64	0.091
1	1.54	2.00	3.59	0.11
$\beta$	$\gamma_{in}$	$\gamma_{out}$	$L$	$C$
0.01	1.37	2.00	3.15	0.13
0.05	1.53	2.00	3.65	0.088
0.1	1.53	1.98	4.20	0.055
1	1.24	2.00	6.01	0.0047

数, ベキ指数  $\gamma_{in}$  を調整できることが分かる. これは, トピックの数が増えたとき, ネットワーク全体のエッジの数は変化しないが, トピック数が増加することでトピックごとのエッジが少なくなるために得られる結果であると考え.

ハイパーパラメータ  $\alpha, \beta$  を変動させたときのネットワーク構造の特徴量を表 3 に示す. ハイパーパラメータ  $\alpha$  はノードが潜在トピックを選択する時の確率に影響し, ハイパーパラメータ  $\beta$  は潜在トピックがノードを選択する時の確率に影響する.  $\alpha$  を大きくすると平均経路長  $L$  の値が減少し, クラスタ係数はわずかに上昇することがわかった. ハイパーパラメータ  $\alpha$  の値を大きくすると, 各ノードの持つ潜在トピックを選ぶ確率の分布が平滑化され, ノードは潜在トピックをランダムに選ぶことに近くなる. この結果はネットワーク生成において, ノードとノードを繋ぐための仲介をするトピックはランダムに選択する方が, クラスタ係数がわずかに大きくすることができることを示している. また, 平均経路長  $L$  を僅かに短くすることも示している. すなわち, わずかにスモールワールド性に影響を与えることができることを示している. しかしトピック数  $K$  を変更した時の特徴量の変化よりも少ないことから, ネットワークの特徴を大幅に変更するためのパラメータであるとはいえない.

またハイパーパラメータ  $\beta$  の値を大きくすると平均経路長  $L$  の値が大きくなりクラスタ係数  $C$  が小さくなることがわかる. ハイパーパラメータ  $\beta$  の値を大きくすると, 各潜在トピックの持つノードを選ぶ確率の分布が平滑化され, 潜在トピックはノードをランダムに選ぶことに近くなる. 今,  $\beta$  が小さい時に平均経路長  $L$  が短くクラスタ係数  $C$  が大きいから, トピックがそのトピックで選ばれやすいノードをより優先的に選択することで, スモールワールドなネットワークが生成できるというこ

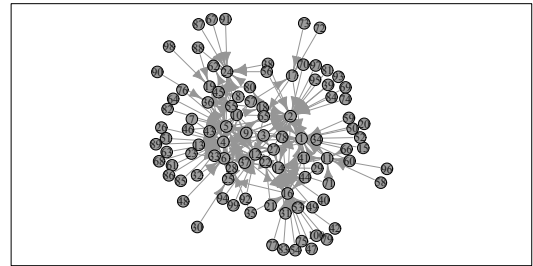


図 3: 提案モデルが生成したネットワーク

とを表している. すなわち, 我々の提案モデルではハイパーパラメータ  $\beta$  を小さくすれば, スモールワールドなネットワークを生成できることが分かる.

ここまでの結果をまとめると, いずれのパラメータを変更しても, 提案モデルが生成するネットワークはスケールフリーな性質を持つことがわかった. またトピック数  $K$  とハイパーパラメータ  $\beta$  を変更することで, ネットワーク構造の特徴量を変更することができることがわかった. そしてトピック数  $K$ , ハイパーパラメータ  $\beta$  の値がそれぞれ小さい時, スモールワールドな性質を持つネットワークを生成できることがわかった. 一方で,  $\alpha$  によりネットワークのスモールワールドな性質を大きく変更することはできないこともわかった.

ここまでは主に提案モデルが生成するネットワークの特徴と, その特徴の程度を変更することができるかについて分析を行った. しかし, これらはいくまでも統計的な値であるため, 実際のネットワークがどのような状態になっているかわからない. そこで, 我々の提案モデルによって生成されたネットワークを可視化し, どのようなネットワークが生成されているか確認する.

可視化するネットワークは表 1 のように設定された提案モデルによって生成されたもので, ノードの数が 100 である.

図 3 に提案モデルが生成したネットワークを示す. 図 3 から, ごく限られたノードが多数のノードからのエッジを獲得している事がわかる. また, 任意のノードが次数の高いノードまでの距離が非常に短い事がわかる. これは, トピックの中で次数の高いノードを選んでいるためであると考え.

## 5 結論

本稿では, 提案モデルを計算機上でシミュレーションを行い, 提案モデルの挙動と指標の値を計算し, 得られた値に関する考察を行った. 提案モデルでは, いずれ

のパラメータを変更してもスケールフリーなネットワークになることが確認された。また、主に  $\beta, K$  を変更することで複雑ネットワークの特徴である、スモールワールド性やクラスタ性に関する特徴を制御できた。パラメータの中でも特にトピック数  $K$  とハイパーパラメータ  $\beta$  を変更することで、ネットワーク構造の特徴量を大きく変更することができることがわかった。いずれも値が小さい時、スモールワールドな性質を持つネットワークを生成できることが明らかになった。さらに、ネットワークを可視化することにより、任意のノードから次数の高いノードまでの距離が短いことも明らかになった。

今後は実ネットワークを用いて提案モデルの各パラメータの値を推定し、推定されたパラメータを基にネットワークをさらに成長させることでどのようなネットワークが生成されるか実験を行う。

#### 参考文献

- [1] Albert, Réka, and Albert-László Barabási. "Statistical mechanics of complex networks." *Reviews of modern physics* 74.1 (2002): 47.
- [2] Watts, Duncan J. *Small worlds: the dynamics of networks between order and randomness*. Princeton university press, 1999.
- [3] Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *science* 286.5439 (1999): 509-512.
- [4] Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." *nature* 393.6684 (1998): 440-442.
- [5] Krapivsky, P. L., and S. Redner. "Network growth by copying." *Physical Review E* 71.3 (2005): 036118.
- [6] Holme, Petter, and Beom Jun Kim. "Growing scale-free networks with tunable clustering." *Physical review E* 65.2 (2002): 026107.
- [7] Bianconi, Ginestra, and A-L. Barabasi. "Competition and multiscaling in evolving networks." *EPL (Europhysics Letters)* 54.4 (2001): 436.
- [8] Klemm, Konstantin, and Victor M. Eguiluz. "Highly clustered scale-free networks." *Physical Review E* 65.3 (2002): 036123.
- [9] Barabasi, Albert-László, Erzsebet Ravasz, and Tamas Vicsek. "Deterministic scale-free networks." *Physica A: Statistical Mechanics and its Applications* 299.3 (2001): 559-564.
- [10] Newman, Mark EJ. "The structure and function of complex networks." *SIAM review* 45.2 (2003): 167-256.
- [11] Kwak, Haewoon, et al. "What is Twitter, a social network or a news media?." *Proceedings of the 19th international conference on World wide web*. ACM, 2010.
- [12] 岡本洋, 坪下幸寛, 園田隆志: トピックに依存した論文ランキング: 学術論文引用ネットワーク分析-, 情報プロフェッショナルシンポジウム予稿集 2009(0), 95-99, 2009
- [13] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3.Jan (2003): 993-1022.
- [14] Shinoda, Kosuke, Yutaka Matsuo, and Hideyuki Nakashima. "Emergence of global network property based on multi-agent voting model." *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*. ACM, 2007.
- [15] Vázquez, Alexei. "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations." *Physical Review E* 67.5 (2003): 056104.
- [16] Henderson, Keith, and Tina Eliassi-Rad. "Applying latent dirichlet allocation to group discovery in large graphs." *Proceedings of the 2009 ACM symposium on Applied Computing*. ACM, 2009.
- [17] Cha, Youngchul, and Junghoo Cho. "Social-network analysis using topic models." *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2012.
- [18] Newman, Mark EJ, and Michelle Girvan. "Finding and evaluating community structure in networks." *Physical review E* 69.2 (2004): 026113.
- [19] Fagiolo, Giorgio. "Clustering in complex directed networks." *Physical Review E* 76.2 (2007): 026107.
- [20] Jones, James Holland, and Mark S. Handcock. "An assessment of preferential attachment as a mechanism for human sexual network formation." *Proceedings of the Royal Society of London B: Biological Sciences* 270.1520 (2003): 1123-1128.
- [21] Newman, Mark EJ, and Juyong Park. "Why social networks are different from other types of networks." *Physical Review E* 68.3 (2003): 036122.
- [22] Goldstein, Michel L., Steven A. Morris, and Gary G. Yen. "Problems with fitting to the power-law distribution." *The European Physical Journal B-Condensed Matter and Complex Systems* 41.2 (2004): 255-258.
- [23] Newman, Mark EJ. "Modularity and community structure in networks." *Proceedings of the national academy of sciences* 103.23 (2006): 8577-8582.
- [24] Milgram, Stanley. "The small world problem." *Psychology today* 2.1 (1967): 60-67.