

# ネットワーク上でのコンテンツ分布のための 距離減衰関数の検討

伏見 卓恭<sup>2</sup> 佐藤 哲司<sup>2</sup> 齊藤 和巳<sup>3</sup> 風間 一洋<sup>4</sup>

**概要:** 本研究では、ネットワークにおけるノードの活動履歴や性質などから得られるコンテンツベクトルを用いて、コンテンツ中心性という中心性指標を提案する。ネットワークでは、類似のコンテンツを有するノード同士は偏って分布しており、分布の中心に存在するノードからの距離が離れるほど、コンテンツ密度は徐々にあるいは急激に減少すると想像できる。そこで各ノードに対して、自身からの距離に従って減衰する重みを付しながらコンテンツベクトルを合成し、自身のコンテンツベクトルとのコサイン類似度により各ノード周りのコンテンツの集中度を定量化し、ランキングする。特に本稿では、逆シグモイド関数による重み付けをすることで、著者らの従来研究より自由度の高いモデルを提案する。実ネットワークを用いた評価実験では、中心性ランキングの妥当性を確認するとともに、減衰レベルを調整するパラメータの推定結果とノードの性質、想定される分布の性質を比較しながら考察する。

## Distance-Based Decay Function for Contents Distributions over a Social Network

TAKAYASU FUSHIMI<sup>2</sup> TETSUJI SATOH<sup>2</sup> KAZUMI SAITO<sup>3</sup> KAZUHIRO KAZAMA<sup>4</sup>

### 1. はじめに

Twitter や Facebook などの SNS や、レビューサイト、ブログサイトなどのソーシャルメディアでは、ユーザ間に多くのインタラクションが存在し、ネットワークとして分析することにより、様々な知見が得られている [1], [2], [3], [4]. このようなネットワークにおいて隣接関係にあるユーザ間には、共通の特徴があると考えられる [5]. 例えば、化粧品に関するレビューサイトを利用するユーザを、商品に対するレビュー評点を要素とする商品次元のベクトルで表現する。この時、フォロー関係にあるユーザ間のベクトルは比較的類似する傾向にある。しかし、ネットワーク上のすべてのユーザ同士のベクトルが類似していることは、通常

考えられない。すなわち、ネットワークのどこかで嗜好の変化点が存在すると考えられる。さらに、各ノードは隣接関係にあるノードと類似特徴を有する傾向にはあるが、その傾向はノードによって様々である。さらに、ソーシャルサーチなどの観点でのユーザ推薦では、類似ノードとの隣接度が高いノードや、類似ノードへのアクセシビリティが高いノードが望まれる。

一方、ネットワークを構成する各ノードに対して、構造特性からノードをランキングする指標として、社会ネットワーク分析で用いられる中心性指標があげられる [6], [7], [8], [9], [10], [11], [12]. 代表的なものとして、他ノードとの隣接度に着目した次数中心性、他ノードへのアクセシビリティに着目した近接中心性、任意のノードペア間を媒介する度合いに着目した媒介中心性、隣接ノードの中心性を加味して自身の中心性を再帰的に求める固有ベクトル中心性、所属するコミュニティへの帰属度に着目したコミュニティ中心性 [13] などが広く知られている。さらには、Web ページのランキング手法である、PageRank [14] や HITS [15] など中心性としての役割を果たす。これら

<sup>1</sup> 筑波大学図書館情報メディア系・日本学術振興会特別研究員  
Faculty of Library, Information and Media Science

<sup>2</sup> 筑波大学図書館情報メディア系  
Faculty of Library, Information and Media Science

<sup>3</sup> 静岡県立大学経営情報学部  
School of Management and Information

<sup>4</sup> 和歌山大学システム工学部  
Faculty of Systems Engineering, Wakayama University

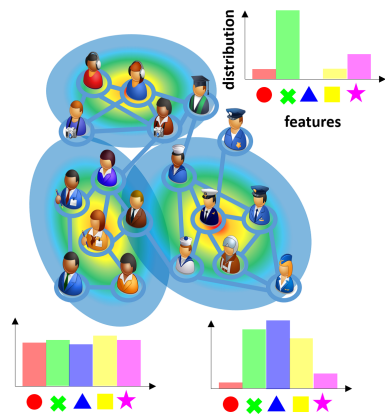


図 1 ネットワーク上でのコンテンツ分布

の中心性指標は、構造特性のみに着目しているため、活動履歴などから得られる各ノードの特徴量を十分に反映出来ていない。すなわち、ノード A とノード B が隣接関係にあっても、これらのノードがどの程度類似しているかは示していない。

本研究では、各ノードの活動履歴から得られる特徴ベクトル（以下、コンテンツベクトル）に基づき、ノード間の類似度を定義する。そして、近傍に類似ノードが多く集まっているノードを抽出する。すなわち、ネットワーク構造上のコンテンツ分布を想定し、コンテンツ分布のモード（最頻値）を抽出することを試みる。ソーシャルサーチ [16], [17] などでは、ノード（ユーザ）がどのような内容の発言を投稿しているかや、投稿内容とクエリとの類似性が重要な役割を果たす。コンテンツ分布の中心にいるようなノードを抽出することで、ソーシャルサーチなどにおいても有用と考えられる。図 1 は、隣接ノードが同一のコンテンツを有するため、コンテンツ分布の密度が高くなるが、ノード間距離が離れるに連れて、密度が低くなることを表している。

著者らは、コンテンツ分布のモード（最頻値）を発見するために、各ノードごとにその近辺にコンテンツが偏在する度合いを表すコンテンツ中心性という中心性指標を提案した [18]。具体的には、各ノードの特徴量をコンテンツベクトルで表現し、近隣ノードのコンテンツベクトルを距離減衰重みを乗じながら合成したベクトルとのコサイン類似度により類似度を定義する。一般に、離れたノードの影響も少なからず受けるが、その影響は遠いほど小さくなることを考えるのが自然であるため、ノード間の距離に応じて減衰する重みを乗じながら合成ベクトルを構築する。

著者らは、これまでに、減衰関数として指数的減衰関数とべき乗的減衰関数を評価・検討してきたが、本論文では、表現の自由度が高い逆シグモイド関数の適用を検討する。実験的に生成したネットワーク構造、および、人々の社会活動によって形成された、いくつかの実際のネットワークに、提案する距離減衰関数を適用し、その有効性と限界を検討したので報告する。

## 2. 関連研究

中心性指標に関する多くの先行研究が存在する。各ノードのトピックに着目し、トピックに応じた重要ノード抽出手法として文献 [19] がある。これらは、PageRank アルゴリズムにより、リンク構造から計算する通常の PageRank スコアベクトルだけでなく、トピック確率ベクトル（非一様確率ベクトル）から計算する PageRank スコアベクトルも加味することで、トピックを考慮した PageRank スコア計算を実現している。このアルゴリズムでは、ランダムジャンプ確率としてトピック確率を用いており、非隣接ノード間の関係を想定しているが、本稿では隣接するノード間の関係に着目する点で異なる。また文書をノードとし、文書間の類似度により構築したグラフに対して、固有ベクトル中心性を求めることで、重要文書を抽出する手法 [20] も存在するが、ノード間の物理的なつながりをネットワークとしていない点で異なる。

コミュニティ抽出に関する先行研究では、構造特性と意味特性の両方に基づく手法がいくつか提案されている。そこで、ノードのコンテンツを用いるコミュニティ抽出の関連研究について説明する。

Kuramochi ら [21] は、与えられたグラフ構造から、極大クリークなどの密なノード集合をノード、クリーク間のリンクをリンクとした交グラフを構築する。交グラフにおけるノード間のリンクには、特徴量より算出する重みを付与する。この際に、交グラフのノード（密なノード集合に相当）内のノードの特徴量を併合し、TF・IDF をかけている。本研究の提案手法でも、周辺ノードの特徴ベクトルを合成するが、距離に従って減衰させながら合成する点、および、減衰の強弱をノードごとに推定する点で異なる。

Wu らは、与えられたネットワークに対し、ノード間の類似度などを重みとした Conceptual ネットワークにおける重みの和が最大で、Physical ネットワーク（実際の接続関係）において連結となる Densest Connected Subgraph を抽出する手法を提案している [22]。この手法では、低次数ノードを枝刈りすることで構造的に密な部分を抽出し、効率的なアルゴリズムを実現している。本稿の提案指標では、構造的な密度ではなく意味的な密度に着目しており、全ノードの中心性スコアを計算する点で異なる。

これらの関連研究と異なり、我々の手法では、自身のコンテンツベクトルと近隣ノードのコンテンツベクトルを合成したベクトル間のコサイン類似度により各ノードの中心性スコアを計算する。

## 3. 提案手法

この節では、本稿の提案指標であるコンテンツ中心性について説明する。コンテンツ中心性は、類似コンテンツを持つノードがそのノード近隣にどの程度偏在しているか、

すなわち、各ノード周辺のコンテンツ集中度を各ノードの中心性スコアとして定量化し、ノードをランキングする。まず、コンテンツ中心性の概念および仮定について説明し、中心性スコアの計算法について述べる。

### 3.1 概念と仮定

現実のネットワークにおいて、ソーシャルメディアでの投稿内容などノードの活動履歴から得られる特徴量は、隣接関係にあるノード同士で類似する傾向が観測されている [5]。本研究では、こうした類似の特徴量を有するノード群は、ネットワーク上に偏在していると仮定する。すなわち、コンテンツに関して隣接ノード間の類似性が高い（同類選択性が高い）ネットワークを対象とする。さらに、特徴量分布の密度は分布の中心に位置するノードとの距離が離れるにつれて、徐々にあるいは急激に減衰すると仮定する（図 2）。そこで各ノードに対して、自身とその近隣に類似コンテンツがどの程度集中しているかを定量化することを試みる。具体的には、投稿内容の TF・IDF などコンテンツベクトルを定義し、距離に従って減衰する重みを付しながら、近隣ノードのコンテンツベクトルを合成する。自身のコンテンツベクトルと近隣ノードの合成ベクトル間のコサイン類似度により集中度を定量化する。前述のように、隣接関係にあるノード同士は類似のコンテンツを有する、すなわち、互いに強い影響を与えていると言える。逆に、遠く離れたノードのコンテンツはほとんど影響しないと自然に想像できる。この影響を反映するために、図 2 のようにノード間の距離に基づく減衰関数を導入する。

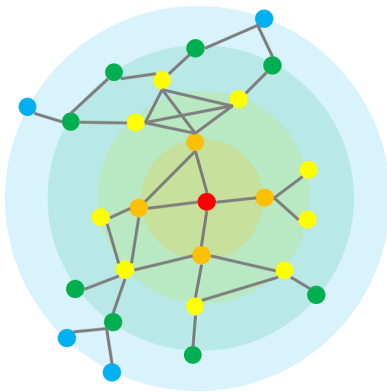


図 2 コンテンツ分布と影響度（中心：赤ノード）

本研究では、コンテンツの集中度をコンテンツ中心性として定義する。中心性の高いノードの近隣には、そのノードのコンテンツと類似のコンテンツを有するノードが偏在していることを意味する。

また、ネットワーク全体を俯瞰すると、類似のコンテンツをもつノード群が、1箇所だけではなく全く離れた場所にも存在する場合も想定できる。このように距離的に離れて存在している場合は、別の分布（多峰性）として扱うこ

とができる。

### 3.2 中心性スコア計算

ノード集合  $V$  とリンク集合  $E$  からなる単純無向ネットワーク  $G = (V, E)$  の各ノード  $u \in V$  は、 $J$  次元コンテンツベクトル  $\mathbf{x}_u$  を有する。各ノード  $u$  に対して、他ノード  $v$  へのグラフ距離（最短パス長）を  $d(u, v)$  とする。ただし、 $d(u, v) = d(v, u)$  であり、 $d(u, u) = 0$  である。ノード  $u$  の距離  $d$  である近隣ノード集合を  $\Gamma_d(u) = \{v : d(u, v) = d\} \subset V$  とする。コンテンツ分布の仮定より、離れたノードはほとんど影響しないようにするために、逆シグモイド関数による減衰を導入する。

$$\rho(d; \lambda, \delta) = \frac{1}{1 + \exp(\lambda \cdot d - \delta)} \quad (1)$$

ここで、 $d$  はノード間距離を表す変数、 $\lambda$  と  $\delta$  は減衰の強さと、減衰開始の距離オフセットを制御するパラメータであり、これらのパラメータを最適化することで、現実のコンテンツ分布を表現することができると考える。

各ノードに対して、距離に基づく減衰重みを付しながら、隣接するノードのコンテンツベクトルを以下のように合成する：

$$\begin{aligned} \mathbf{y}_u &= \sum_{d=1}^{D_u} \rho(d; \lambda, \delta) \sum_{v \in \Gamma_d(u)} \mathbf{x}_v \\ &= \sum_{v \in V \setminus \{u\}} \rho(d(u, v); \lambda, \delta) \mathbf{x}_v. \end{aligned}$$

ここで、 $D_u = \max_{v \in V} d(u, v)$  であり、この合成ベクトルを RVwD (Resultant Vector with Decay) と表記する。各ノードの RVwD は、直接隣接するノードを含め、近隣ノードのコンテンツベクトルを大きい重みで、遠方ノードのコンテンツベクトルを小さい重みで合成したものである。したがって、RVwD は幾分か均ざされている。図 3 は、パラメータを変化させた際の減衰度合いを表している。パラメータ  $\lambda$  は、大きくするほど急激に減衰する。パラメータ  $\delta$  は、大きくするほど減衰開始の距離が遠くなる。すなわち、直近だけでなく幾分か離れたノードも大きな重みで足し込むことになる。

次に、各ノードに対して、元々のコンテンツベクトルと RVwD 間のコサイン類似度によりコンテンツ中心性のスコアを計算する：

$$\begin{aligned} \text{CDC}(u) &= \left\langle \frac{\mathbf{x}_u}{\|\mathbf{x}_u\|}, \frac{\mathbf{y}_u}{\|\mathbf{y}_u\|} \right\rangle \\ &= \left\langle \frac{\mathbf{x}_u}{\|\mathbf{x}_u\|}, \frac{\sum_{d=1}^{D_u} \rho(d; \lambda, \delta) \sum_{v \in \Gamma_d(u)} \mathbf{x}_v}{\|\sum_{d=1}^{D_u} \rho(d; \lambda, \delta) \sum_{v \in \Gamma_d(u)} \mathbf{x}_v\|} \right\rangle. \end{aligned}$$

ノード  $u$  のスコア  $\text{CDC}(u)$  が他のノードより大きければ、 $u$  の界隈に類似コンテンツを有するノードが偏在しており、ノード  $u$  はコンテンツ中心性ランキングで上位となる。

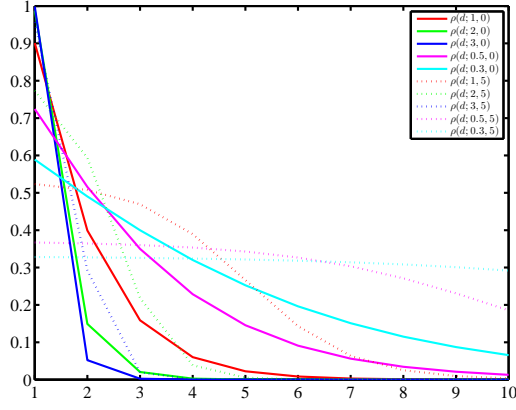


図 3 パラメータと減衰度合い

#### 4. 比較手法

逆シグモイド関数による減衰と比較するために、著者の既存研究で提案した2つの減衰関数を用いる。1つ目は、以下に定義する指数的減衰関数であり、

$$\rho(d; \lambda) = \exp(-\lambda d),$$

$\lambda$  は減衰の程度を制御するパラメータである。2つ目は、以下に定義するべき乗的減衰関数である：

$$\rho(d; \lambda) = \exp(-\lambda \log d).$$

ネットワーク構造的に近いノードほどコンテンツベクトルが類似する傾向にあるという前提に従って、ノードごとに減衰度合いを調整する。本稿では、各ノードの RVwD がコンテンツベクトルとのコサイン類似度の意味で最も類似するように各ノードのパラメータ  $\lambda_u$  を設定する。L2 ノルムを 1 に正規化したコンテンツベクトルを  $\mathbf{x}_u$  とし、以下の目的関数を定義する：

$$F_u(\lambda_u) = \mathbf{x}_u^T \frac{\sum_{v \in V \setminus \{u\}} \rho(d(u, v); \lambda_u) \mathbf{x}_v}{\|\sum_{v \in V \setminus \{u\}} \rho(d(u, v); \lambda_u) \mathbf{x}_v\|}. \quad (2)$$

目的関数 (2) を最大化するようなパラメータ  $\lambda_u$  を求める手順を説明する。ここで、 $d$  を  $\log(d)$  と置き換えることで、以下で説明する導出はべき乗的減衰でも成り立つため、以下では指数減衰重みを用いて説明を進める。ノード  $u$  に対して、距離  $d$  にあるノードのコンテンツベクトルの合成ベクトルを

$$\mathbf{f}_{u,d} = \sum_{v \in \Gamma_d(u)} \mathbf{x}_v$$

とし、ノード  $u$  のコンテンツベクトルとの内積を  $g_{u,d} = \mathbf{x}_u^T \mathbf{f}_{u,d}$  とする。そして、ノード  $u$  からの距離の和が  $d$  になる合成ベクトル  $\mathbf{f}_{u,d_1}$  と  $\mathbf{f}_{u,d_2}$  ペア間の内積を足し合わせ

$$h_{u,d} = \sum_{d_1+d_2=d} \mathbf{f}_{u,d_1}^T \mathbf{f}_{u,d_2}$$

とすると、式 (2) は以下のように書き換えられる：

$$F_u(\lambda_u) = \frac{\sum_{d=1}^{D_u} \exp(-\lambda_u d) g_{u,d}}{\sqrt{\sum_{d=2}^{2D_u} \exp(-\lambda_u d) h_{u,d}}}.$$

計算の便宜上、対数をとった以下の目的関数を最大にするようなパラメータ  $\lambda_u$  を求める：

$$\begin{aligned} \log F_u(\lambda_u) &= \log \sum_{d=1}^{D_u} \exp(-\lambda_u d) g_{u,d} \\ &\quad - \frac{1}{2} \log \sum_{d=2}^{2D_u} \exp(-\lambda_u d) h_{u,d}. \end{aligned} \quad (3)$$

ここで、事後確率関数を

$$r_{u,d} = \frac{\exp(-\lambda_u d) g_{u,d}}{\sum_{d'=1}^{D_u} \exp(-\lambda_u d') g_{u,d'}}$$

とすると、式 (3) は以下のように書き換えられる：

$$\begin{aligned} \log F_u(\lambda_u) &= \sum_{d=1}^{D_u} \bar{r}_{u,d} \{(-\lambda_u d) + \log g_{u,d}\} - \sum_{d=1}^{D_u} \bar{r}_{u,d} \log r_{u,d} \\ &\quad - \frac{1}{2} \log \sum_{d=2}^{2D_u} \exp(-\lambda_u d) h_{u,d}. \end{aligned}$$

パラメータ  $\lambda_u$  に関係のない項などを除くと

$$Q_u(\lambda_u) = -\lambda_u \sum_{d=1}^{D_u} \bar{r}_{u,d} \cdot d - \frac{1}{2} \log \sum_{d=2}^{2D_u} \exp(-\lambda_u d) h_{u,d}$$

となり、1階微分は、

$$\frac{dQ_u(\lambda_u)}{d\lambda_u} = -\sum_{d=1}^{D_u} \bar{r}_{u,d} \cdot d + \frac{\sum_{d=2}^{2D_u} \exp(-\lambda_u d) \cdot d \cdot h_{u,d}}{2 \sum_{d=2}^{2D_u} \exp(-\lambda_u d) h_{u,d}}$$

となる。ここで、

$$s_{u,d} = \frac{\exp(-\lambda_u d) h_{u,d}}{\sum_{d'=2}^{2D_u} \exp(-\lambda_u d') h_{u,d}}$$

とすると2階微分は、

$$\frac{d^2 Q_u(\lambda_u)}{d\lambda_u^2} = -\frac{1}{2} \left\{ \sum_{d=2}^{2D_u} s_{u,d} \cdot d^2 - \left( \sum_{d=2}^{2D_u} s_{u,d} \cdot d \right)^2 \right\}$$

となり、ブレースの中は2次のモーメント同様に非負であるため、2階微分自体は常に0以下となる。1階微分が  $\lambda_u$  に関して閉じた形で書けないため、本研究ではニュートン法によりパラメータを求める。この推定されたパラメータは値が大きいほど強い減衰を実現し、近隣ノードの値のみを大きな重みで、遠くのノードの値はほとんど無視する。逆に値が0に近いほど弱い減衰を実現し、近隣も遠方も同程度の重みで合成する。すなわち、近隣に類似のコンテンツベクトルを有するノードが存在するか否かにより値が異なり、局所的なノード集合の中に順応しているノードは値が大きく、近隣に類似するノードが存在しない異端児ノードの場合は、多くのノードのコンテンツベクトルを均等に合成しなければコサイン類似度を高くできないため、値が

0に近くなる。

コンテンツベクトルの次元を  $J$ 、平均ノード間距離を  $\bar{D}$  とすると、全ノードのパラメータ推定に要する時間計算量は、 $h_{u,d}$  計算に要する  $O(|V| \times 2\bar{D} \times J)$  である。本提案指標では、様々な次元圧縮技術を用いて圧縮したベクトルを用いることも可能である。

## 5. 評価実験

### 5.1 ネットワークデータ

本研究では、コンテンツに関して隣接ノード間の類似性(同類性)が高いネットワークを対象として評価する。1つ目のネットワークは、ある大学のウェブサイトにおけるハイパーリンク構造である\*1。ウェブページをノード、ハイパーリンクを無向化しリンクとした。各ノードのコンテンツベクトルは、ウェブページの内容を形態素解析して得られる名詞群の Bag of Words とした。ノード数は 600、リンク数は 1,299、コンテンツベクトルの次元数は 4,412 である。本稿では Web ネットワークと呼ぶ。

2つ目のネットワークは、日本語ウィキペディア\*2の人名の共起ネットワークである。人物記事をノード、5つ以上の記事において共起関係のある人物間にリンクを張った無向ネットワークである。各ノードのコンテンツベクトルは、記事内に出現する名詞群の Bag of Words とした。ノード数は 9,481、リンク数は 122,522、コンテンツベクトルの次元数は 20,411 である。本稿では Wiki ネットワークと呼ぶ。

### 5.2 推定パラメータに関する考察

提案手法における距離減衰の強さおよび開始オフセットを制御するパラメータに関して、推定した結果  $\hat{\lambda}, \hat{\delta}$  について考察する。図 5 に、横軸を  $\hat{\delta}$ 、縦軸を  $\hat{\lambda}$  とした散布図を示す。各点が各ノードを表し、各点は指数的減衰(従来法)におけるパラメータ  $\lambda$  により赤から青のグラデーションで着色した。

図 5(a)(b) いずれの結果においても、指数的減衰で同一視されていた青色のノード群が、オフセットパラメータ  $\delta$  により分割され、ノードごとにより詳細な表現が可能となった。Web ネットワークにおいて  $\hat{\lambda} > 5$  となるようなノードは、「教員紹介ページ」が多くを占めていた。これらのページには、教員の経歴や研究内容などが書かれており、隣接するページとして、詳細な研究内容が書かれたページなどが存在する。すなわち、ごく近隣のみ類似ノードが存在するため、強い距離減衰重みを実現させたと考えられる。Wiki ネットワークにおいて  $\hat{\lambda} > 5$  となるようなノードは、「タレント芸能人ページ」が多くを占めていた。これらはバンドやコンビ、チームなどのメンバーになっている

割合が高く、共起関係にある(同じチームに所属)ノード同士は、類似の単語を使用する傾向にあるため、このような結果になったと考えられる。すなわち、タレントの共起関係は、類似のコンテンツベクトルをもったノード同士が結びついているという直感に合致した結果となった。またいずれのネットワークでも、 $\hat{\lambda} > 5 \wedge \hat{\delta} > 1$  となるノードはほとんど存在しなかった。

$1 \leq \hat{\lambda} \leq 4$  となるようなノードは、上述のノードより、近傍に類似ノードがある程度存在するノードが占めていた。Wiki ネットワークでは、「俳優や女優などの芸能人ページ」ノードがこの範囲に含まれていた。これらの芸能人ページ内には、来歴・人物やエピソードなどが書かれているが、共起関係にある俳優同士(同じドラマに出演など)だからといって、これらに用いられる単語が類似するとは限らないため、それを反映した結果と考えられる。すなわち、俳優・女優の共起関係は、様々なコンテンツベクトルをもったノード同士が結びついているという直感に合致した結果となった。そのため、上述の「タレント芸能人」ノードと比べると緩やかな減衰重みを実現したと考えられる。さらに、この範囲のノードは、オフセットパラメータ  $\delta$  の値により詳細に分類できる。 $\hat{\delta} < 2$  なノードとして、次数が高く、1,2 ステップで多くのノードに辿り着けるノードが多く含まれていた。そのため、減衰のオフセットが比較的小さく推定されたと考えられる。対して、 $\hat{\delta} > 5$  なノードとして、次数が低くネットワークの端にいるようなノードが多く含まれていた。

$\hat{\lambda} \simeq 0$  となるようなノードは、Web ネットワークでは、「ニュースページ」や「お知らせページ」などが見られた。これらのページには、所属学生や教員の受賞ニュースなどが書かれており、近隣はおろかネットワーク内に類似するページが存在しないため、パラメータの値が小さくなり、弱い距離減衰重みを実現させたと考えられる。Wiki ネットワークでは、「特異な固有名詞を使用するページ」や「内容が少ないページ」などが見られた。これらのページは、ネットワーク内に類似するページが存在しないため、距離に関係なく様々なページのコンテンツベクトルを合成することで、自身のコンテンツベクトルとのコサイン類似度を高くしようとする。そのためパラメータの値が小さくなり、弱い距離減衰重みを実現させたと考えられる。

### 5.3 中心性ランキングに関する考察

この節では、提案指標であるコンテンツ中心性ランキングの結果について、上位ノードの性質や他の中心性指標と比較しながら考察する。図 6 に、コンテンツ中心性スコアにより赤から青のグラデーションで着色した可視化結果を示す。Web ネットワークでは、上位 10 ノードの全てがシラバスページであり、実際にこれらのページ群には類似の単語(名詞)が多く含まれており、シラバスページ群が大きい

\*1 法政大学情報科学部(2010年8月時点) <http://cis.k.hosei.ac.jp/>

\*2 <https://ja.wikipedia.org/>

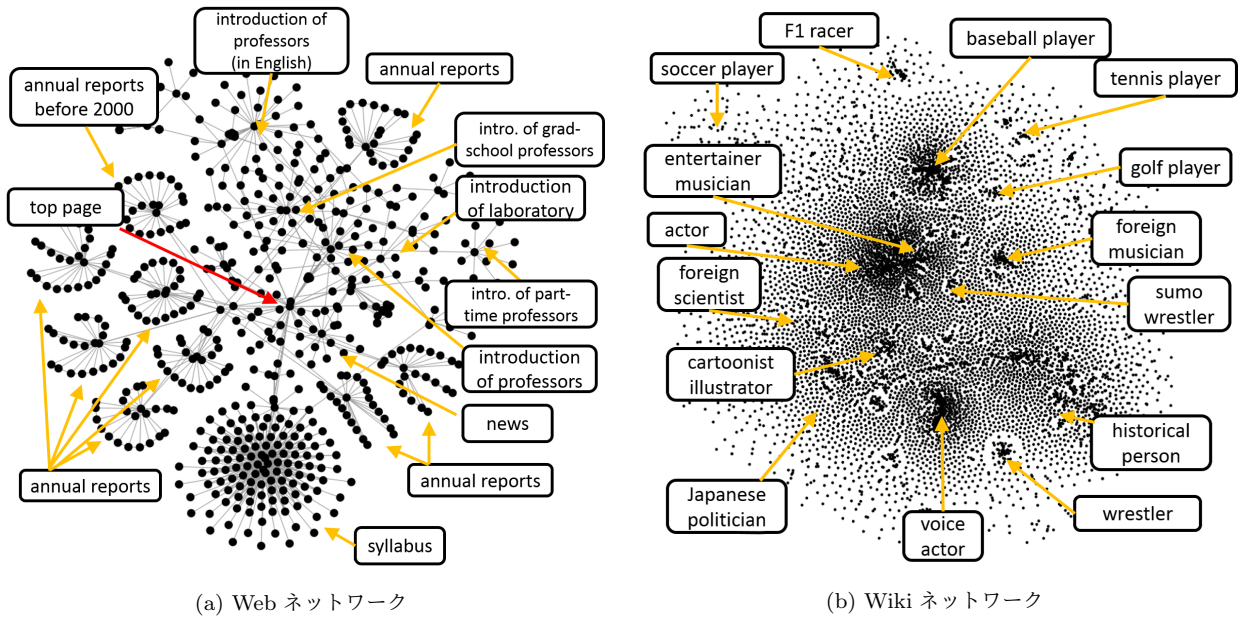


図 4 ラベル付き可視化結果

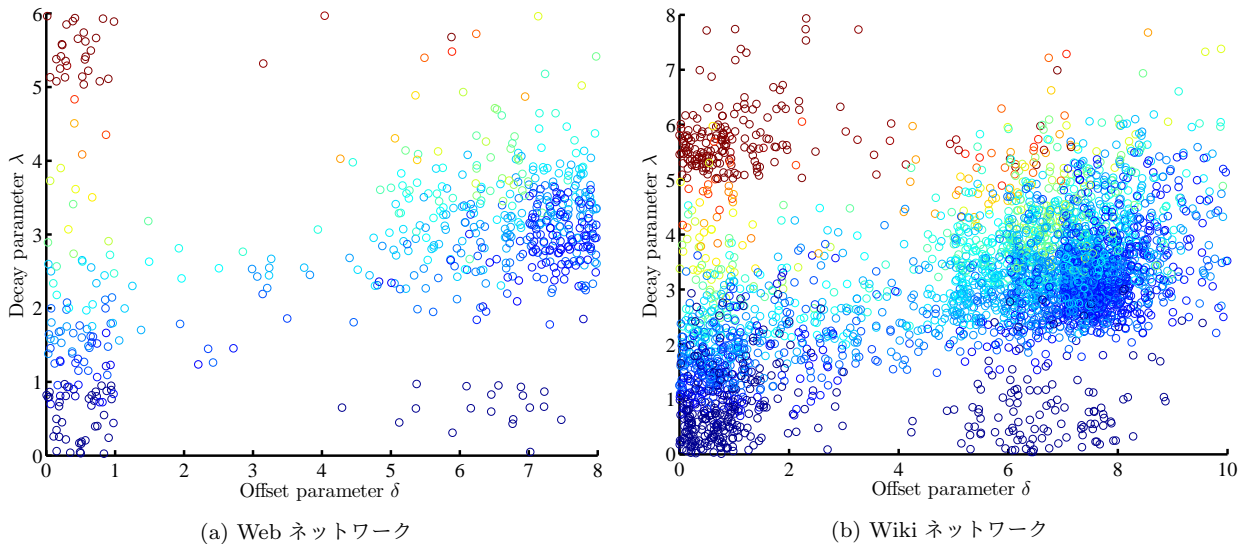


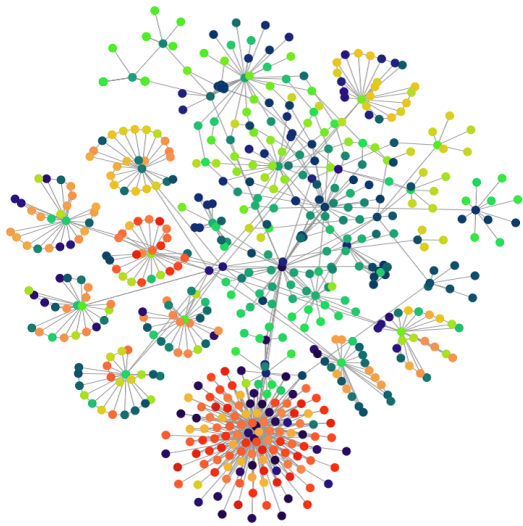
図 5 パラメータの推定結果

なコンテンツ分布を構成していることが伺える。Wiki ネットワークでは、上位ノードとして、ジャニーズ事務所所属のアイドルに関するページが抽出された。これらは、同一グループに所属するアイドルのページは互いにつながっており、かつ、ページ内に含まれる名詞も類似のものが多い。次いで、野球選手、サッカー選手、関取のページが抽出された。これらについても、近傍に類似のコンテンツ（勝敗表など）を有するノードが多く存在することから、上位として抽出されたと考えられる。コンテンツ中心性は、類似のコンテンツベクトルを有するノードが多く集まる部分を抽出できたといえる。

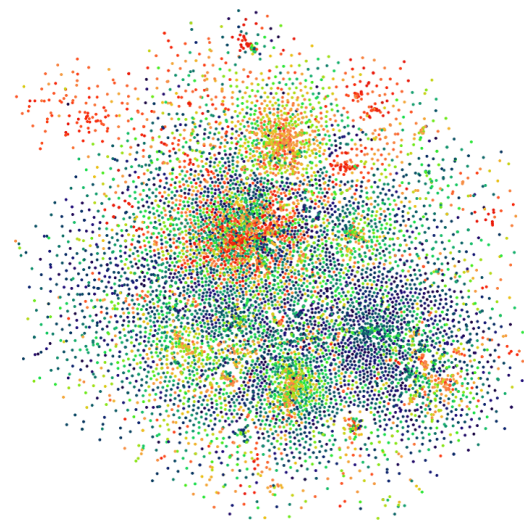
次に、構造に基づく従来の中心性指標との関係性を評価する。図 7 に、横軸に順位、縦軸にコンテンツ中心性その他の中心性指標ランキングのノード一致率（F 値）をプロットした。具体的には、コンテンツ中心性と中心性 A の上位

$r$  件のノード集合をそれぞれ  $CDC(r), A(r)$  とするとき、 $|CDC(r) \cap A(r)|/r$  をプロットした。評価実験に用いた 2 つのネットワークにおいて、以下の関係が見られた。

- 近接中心性とコミュニティ中心性はコンテンツ中心性とある程度の相関がある。これは、コンテンツ分布のモードはコミュニティなどの中心に位置していることから、近接中心性、コミュニティ中心性が高いノードであるという直感に合致した結果である。特にコミュニティ中心性は、上位の一致率が顕著に高い。
- 次数中心性、PageRank はやや相関がある。次数中心性と PageRank の間には強い相関関係があるが、次数が高いノードは、相対的に多くのコンテンツベクトルを強い重みで合成するため、コサイン類似度も相対的に高くなるためと考えられる。
- 媒介中心性、固有ベクトル中心性、コンテンツベクトル

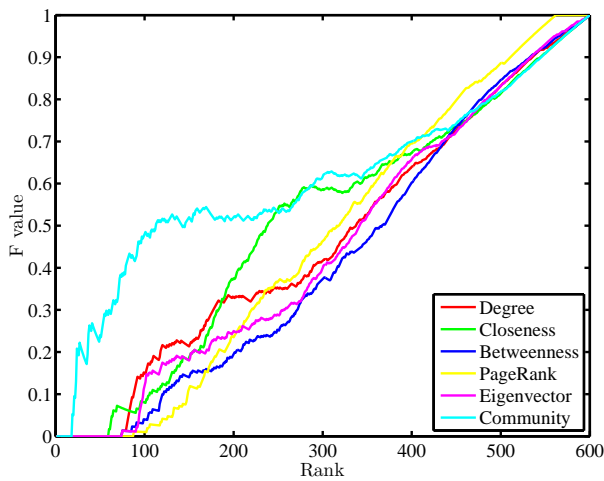


(a) Web ネットワーク

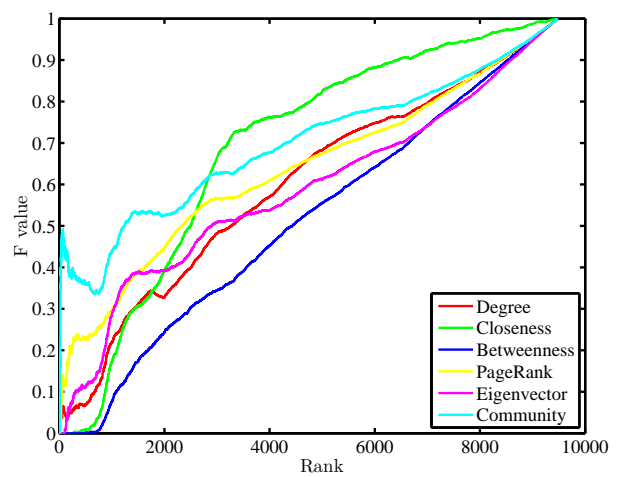


(b) Wiki ネットワーク

図 6 コンテンツ中心性スコアによる色付け



(a) Web ネットワーク



(b) Wiki ネットワーク

図 7 中心性指標とのランキンガー致率

ルの次元は、相対的に相関係数が低い傾向にある。

次いで、目的関数値という側面から提案手法（逆シグモイド的減衰）と従来手法（指数的減衰重み、べき乗的減衰重み）を比較する。図 8 に、提案手法の目的関数値順にノードを並べ、各ノードに対する各種法の目的関数値をプロットした。図 8 から、いずれのネットワークにおいても、上位のノードでは従来手法より提案手法が逆シグモイド関数による表現力ゆえに高い目的関数値を実現している。逆に下位のノードに対しては、指数的減衰による適切でない減衰が起因し、RVwD が平均的なベクトルになり、本来中心性スコアが低いノードでも高く推定されてしまったと考えられる。

## 6. おわりに

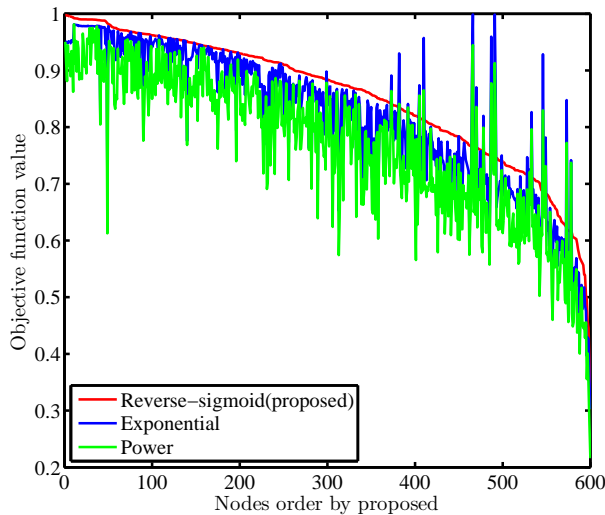
本研究では、ネットワーク上でのコンテンツ分布を仮定

して、分布の中心に存在するようなノードを抽出するコンテンツ中心性という新たな指標を提案した。特に、著者らの従来研究における指数的減衰関数やべき乗的減衰関数より表現の自由度の高い逆シグモイド関数を用いた。2つの実ネットワークを用いた評価実験により、ある程度妥当なノードを抽出できることを確認した。今後は、さらに多様なネットワークを用いて評価する。

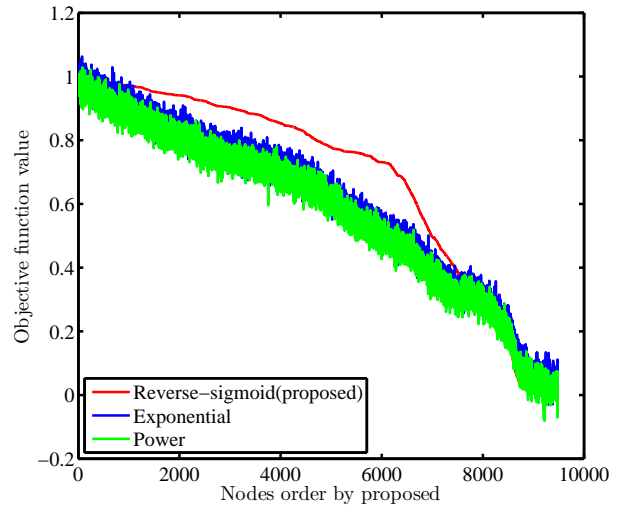
謝辞 本研究は、JSPS 特別研究員奨励費 15J00735 の助成を受けたものである。

## 参考文献

- [1] Newman, M. E. J., Forrest, S. and Balthrop, J.: Email networks and the spread of computer viruses, *Physical Review E*, Vol. 66, p. 035101 (2002).
- [2] Newman, M. E. J.: The Structure and Function of Complex Networks, *SIAM Review*, Vol. 45, pp. 167–256



(a) Web ネットワーク



(b) Wiki ネットワーク

図 8 目的関数値の比較

- (2003).
- [3] Gruhl, D., Guha, R., Liben-Nowell, D. and Tomkins, A.: Information Diffusion Through Blogspace, *Proceedings of the 13th International Conference on World Wide Web, WWW '04*, New York, NY, USA, ACM, pp. 491–501 (2004).
- [4] Domingos, P.: Mining Social Networks for Viral Marketing, *IEEE Intelligent Systems*, Vol. 20, No. 1, pp. 80–82 (2005).
- [5] Newman, M. E. J.: Assortative mixing in networks, *Structure*, Vol. 2, No. 4, p. 5 (2002).
- [6] Freeman, L.: Centrality in social networks: Conceptual clarification, *Social Networks*, Vol. 1, No. 3, pp. 215–239 (1979).
- [7] Wasserman, S. and Faust, K.: *Social Network Analysis: Methods and Applications*, Cambridge University Press (1994).
- [8] Brandes, U.: A Faster Algorithm for Betweenness Centrality, *Journal of Mathematical Sociology*, Vol. 25, pp. 163–177 (2001).
- [9] Kang, C., Molinaro, C., Kraus, S., Shavitt, Y. and Subrahmanian, V. S.: Diffusion Centrality in Social Networks, *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, ASONAM '12, Washington, DC, USA, IEEE Computer Society, pp. 558–564 (2012).
- [10] 伏見卓恭, 斉藤和巳, 池田哲夫, 武藤伸明: ノード群の協調的振舞いに着目した集合媒介中心性の提案と応用, *電子情報通信学会和文論文誌 D*, Vol. J96-D, No. 5, pp. 1158–1165 (2013-05).
- [11] 伏見卓恭, 斉藤和巳, 武藤伸明, 池田哲夫, 風間一洋: 実距離を考慮した中心性指標の提案と重要観光スポット抽出への応用, *人工知能学会論文誌*, Vol. 30, No. 6, pp. 703–712 (2015).
- [12] Saito, K., Kimura, M., Ohara, K. and Motoda, H.: Super mediator - A new centrality measure of node importance for information diffusion over social network, *Information Sciences*, Vol. 329, pp. 985 – 1000 (2016).
- [13] Newman, M. E. J.: Finding community structure in networks using the eigenvectors of matrices, *Physical Review E*, Vol. 74, No. 3, pp. 036104+ (2006).
- [14] Langville, A. N. and Meyer, C. D.: Deeper Inside PageRank, *Internet Mathematics*, Vol. 1, No. 3, pp. 335–380 (2004).
- [15] Kleinberg, J. M.: Authoritative sources in a hyperlinked environment, *J. ACM*, Vol. 46, pp. 604–632 (1999).
- [16] Carmel, D., Zwerdling, N., Guy, I., Ofek-Koifman, S., Har'el, N., Ronen, I., Uziel, E., Yogev, S. and Chernov, S.: Personalized Social Search Based on the User's Social Network, *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09*, New York, NY, USA, ACM, pp. 1227–1236 (2009).
- [17] Hsieh, H.-P., Li, C.-T. and Yan, R.: I See You: Person-of-Interest Search in Social Networks, *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15*, New York, NY, USA, ACM, pp. 839–842 (2015).
- [18] 伏見卓恭, 佐藤哲司, 斉藤和巳, 風間一洋: 構造特性と意味特性を考慮した中心性指標の提案, 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016) (2016).
- [19] Haveliwala, T. H.: Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search, *IEEE Trans. on Knowl. and Data Eng.*, Vol. 15, No. 4, pp. 784–796 (2003).
- [20] Erkan, G. and Radev, D. R.: LexRank: Graph-based Lexical Centrality As Saliency in Text Summarization, *J. Artif. Int. Res.*, Vol. 22, No. 1, pp. 457–479 (2004).
- [21] Kuramochi, T., Okada, N., Tanikawa, K., Hijikata, Y. and Nishida, S.: Community Extracting Using Intersection Graph and Content Analysis in Complex Network, *Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology, WI-IAT '12*, Vol. 1, Washington, DC, USA, IEEE Computer Society, pp. 222–229 (2012).
- [22] Wu, Y., Jin, R., Zhu, X. and Zhang, X.: Finding Dense and Connected Subgraphs in Dual Networks, *Proceedings of the IEEE 31st International Conference on Data Engineering (ICDE2015)*, pp. 915–926 (2015).