

ノードの特徴量を考慮したネットワーク間類似度

Similarity Between Networks Considering Node Feature Quantities

滝見 優太†
Yuta Takimi鈴木 育男‡
Ikuo Suzuki山本 雅人†
Masahito Yamamoto古川 正志†
Masashi Furukawa

1. 序論

世界に存在する情報は点(ノード)と線(リンク)によるネットワークとして見ることで見られる人と人との繋がり, WWW のネットワーク上でのホームページとそれらを繋ぐリンクの関係, 都市と都市を結ぶ交通網, 生物が他の生物を栄養源とする様子を表わした食物連鎖など様々な事象がネットワークとして表わされる.

一般に二つのネットワークが類似しているかの比較には特徴量を用いることが多い. しかし単一の特徴値による比較ではある一つの特徴しか考慮することができない. リンクのトポロジを基に比較する GBI 法^{[1][2]}も存在するが, トポロジ比較は膨大なパターンが考えられ, 大規模なネットワークに対して行うことは難しい.

本研究ではネットワークを構成するノードのネットワーク中での役割に注目する. ネットワークの特徴値を複合的に用いることでそのノードの役割を特徴付け, 似た役割を持つノードがどの程度双方のネットワークに存在・分布し, どの程度似ているのかを測定して, ネットワークの類似度を測定する.

2. ネットワークにおける特徴量

本研究では重みなし有向ネットワークを対象とする. 有向グラフ上では 1 本のリンクに対して入と出があり, ノードの特徴となる要素が増え判定が容易となる. 実際に存在するネットワークでも twitter のフォロー・フォロワー関係のように双方で関係が必ずしも一致しないものは多い. また, 簡単のためにリンクの種類は連結が有るか無いかのみを表現する重み無しネットワークを取り扱う.

以下に本論文で取り扱う特徴値について述べる.

2.1 次数 k

あるノード i に接続しているリンクの本数を次数 k_i と表わす. 有向グラフの場合当該ノード i へと入ってくるリンクの本数を入次数 k_i^{in} , ノードから出ていくリンクの本数を出次数 k_i^{out} と呼び区別する.

2.2 最短経路長 d

あるノード i から任意のノード j へ到達するために経由するリンクの本数の最小値を最短経路長 d_{ij} と表わす. 有向グラフの場合, ノード i からノード j への最短経路長と, ノード j からノード i への最短経路長は必ずしも等しくならない.

最短経路長 d_{ij} はノード i から見るとノード j へと到達する出の最短経路長であり, ノード j から見るとノード i から入ってくる入の最短経路長である.

2.3 クラスタリング係数 c

ノードに接続されたノード同士で三角形を作りその対象

ノード周辺において密なクラスタを形成しているかを測る指標である.

あるノード i と, i に接続しているノード j, m が存在するとき, あるノード i のクラスタリング係数は

$$c_i = \frac{1}{k_i(k_i - 1)} \sum_j \sum_m \sigma(i, j, m) \quad (1)$$

の式で表わす. $\sigma(i, j, m)$ はノード i, j, m が全て繋がっている時 1, それ以外の場合 0 の値となる. クラスタリング係数 c_i は $0 \leq c_i \leq 1$ の値をとる.

2.4 媒介中心性 b

媒介中心性 b_i は, あるノード i がネットワーク中でどれだけ重要かを, 全ての最短経路長の中でノード i がどれだけ通過されたかを考慮し計算される. 有向グラフにおける任意のノード i の媒介中心性は

$$b_i = \frac{\mu(i)}{(n-1)(n-2)} \quad (2)$$

の式で求められる. ここで, n はネットワークに存在するノードの総数, $\mu(i)$ はノード i を端としない全ての最短経路長のうち, ノード i が通過された回数を示す. 式の分母は i を端としない全ての最短経路長の数である.

2.5 隣接特徴値分布

あるノード i に接続しているノード集合についての各特徴値の分布を隣接特徴値分布 p_i と呼ぶ. 例として, 対象とする特徴値が次数であれば隣接次数分布となる.

有向ネットワークでは接続しているノード集合についても入ノード集合と出ノード集合に分けて考える. よって, 「ノード i の隣接入ノード集合の隣接出次数分布」を $p_i^{in(k^{out})}(x)$ と表わす.

3. 提案手法

2つのネットワーク X, Y のネットワーク間類似度を測定する手順は以下に沿って行う.

- (1) 両ネットワーク中の各ノードの特徴値を算出する
- (2) 両ネットワーク中のノード同士の特徴値を比較し, ノード間非類似度を算出する
- (3) ノード間非類似度の小さいノード同士をマッチングする
- (4) マッチングしたノード同士のノード間非類似度からネットワーク間非類似度を算出する

各手順について詳しく説明する.

- (1) 両ネットワーク中の各ノードの特徴値を算出する
まずは両方のネットワークに存在する各ノードの特徴値を算出する.
- (2) 両ネットワーク中のノード同士の特徴値を比較し, ノード間非類似度を算出する

算出した各ノードの特徴値を他方のネットワークに存在する任意のノードと比較し, その差からノード間非類似度

† 北海道大学大学院 情報科学研究科

‡ 北見工業大学

を算出する. ネットワーク X に存在するあるノードを i , ネットワーク Y に存在するあるノードを j , ノード i と j のノード間非類似度 $D(i, j)$ と表現する. 各要素の差の算出方法は用いる特徴値によって異なる. ノード間非類似度 $D(i, j)$ を要素 D_{ij} としたノード間非類似度行列 D を作成する.

(3) ノード間非類似度の小さいノード同士をマッチングする

ノード間非類似度の小さいノード同士が対になるよう, ルールに従いマッチングする. ノード i にマッチングしたノードが j のとき, $M(i)=j$ となるようなマッチング対応関数 $M(x)$ をネットワーク X に存在する全てのノードに対して作る.

(4) マッチングしたノード同士のノード間非類似度からネットワーク間非類似度を算出する

算出したノード間非類似度行列 D とマッチング対応関数 $M(x)$ からネットワーク間非類似度を算出する. ネットワーク X とネットワーク Y のネットワーク間非類似度 $|D_{AB}|$ は,

$$|D_{XY}| = \sum_{i=0}^n D(i, M(i)) \quad (3)$$

の計算式で算出する.

この値は非類似度であり, $0 \leq |D_{XY}| \leq 1$ であるため,

$$|S_{XY}| = 1 - |D_{XY}| \quad (4)$$

と計算することでネットワーク間類似度 $|S_{XY}|$ として表わすことができる.

4. 数値演算実験

提案手法と指標の妥当性を示すため, 実験を行った.

4.1 実験条件

ネットワークは以下のパラメータの有向グラフの WS モデル^[3]ネットワークを用いた.

ノード数: 100

リンク数: 600

初期次数: 12

リンク張替確率 α : 0.5

このネットワーク中に存在する全てのリンクに対して確率 β でリンクを繋ぎ替える. 確率 β を徐々に上げていくことで元々のネットワークの形状から変化していくため, β の上昇に比例して各特徴値差が増加していけば, その特徴量はこの手法においてネットワークの類似度測定に用いることに適していると考えられる.

使用した特徴量は, 次数(入次数, 出次数, 隣接ノード入次数分布, 隣接ノード出次数分布, 隣接ノード入次数分布, 隣接ノード出次数分布), クラスタリング係数(クラスタリング係数, 隣接ノードクラスタリング係数分布, 隣接ノードクラスタリング係数分布), 経路長(入平均経路長, 出平均経路長, 入平均経路長分布, 出平均経路長分布), 媒介中心性(平均媒介中心性, 隣接ノード媒介中心性分布, 隣接ノード媒介中心性分布)の, 4 つの特徴値を基とした 16 項目の差のデータを確認した.

試行は β を 0.05 刻みで増加させ, 50 回平均でデータを取った.

4.2 実験結果

実験結果のグラフを図 1, 2 に示す.

図 1 のグラフは β に比例し増加する傾向がみられた特徴量差のグラフである. 各特徴量差は β に比例して増加する傾向が見られるが, β がある程度の値になると増加が緩や

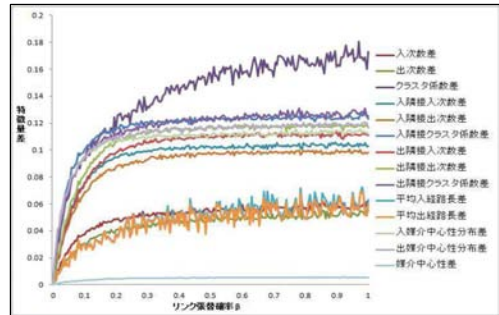


Figure1. Result of Experiment with parameter chart 1

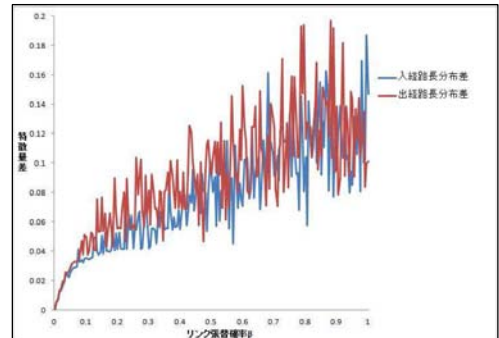


Figure2. Result of Experiment with parameter chart 2

かになる. これは一定以上リンクを張り替えてしまうとどちらも元の構造を失ってしまい, 両者同様に全く似ていないネットワークだと判断されるためだと考えられる.

一方, 図 2 は β の増加に無関係に大きく変動した特徴量(入平均経路長, 出平均経路長)のグラフである. β の値に関わらず増減が激しくなっており, この特徴量に関してはネットワークの類似度を捉えるための特徴量としては適していないと考えられる.

5 結論

ネットワークの特徴量を用いてノードの役割という観点からネットワークの類似度を比較する手法を提案した. その中で, 用いる特徴量はどのようなものを用いるのが妥当かを実験により検証した.

今後はこれの特徴量がどのように構造上の特徴を捉えるのか, 特徴づけたノードがどのように分布しているのか, など更に詳細な検証を行い, より実用に即した手法・指標に近づけていく.

参考文献

- [1] NGUYEN PHU CHIEN, OHARA KOUZOU MOTODA HIROSHI WASHIO TAKASHI "CI-GBI: A Novel Strategy to Extract Typical Patterns from Graph Data", 情報処理学会研究報告. ICS, [知能と複雑系] 2004(125), 105-110, 2004-12-04
- [2] 和田 貴久, 大野 博之, 稲積 宏誠 「部分構造に基づく構造類似性を用いた特徴抽出システムとその応用」, 日本データベース学会論文誌 Vol.7, No.1 (2004)
- [3] Duncan J. Watts and Steven H. Strogatz: Collective dynamics of 'small-world' networks