

PageRankに基づく動的ネットワークの構造変化抽出

Extracting Structural Changes of Dynamic Networks based on the PageRank

伏見卓恭^{1,2*} 斉藤和巳¹ 風間一洋³

Takayasu Fushimi^{1,2}, Kazumi Saito¹, and Kazuhiro Kazama³

¹ 静岡県立大学 経営情報イノベーション研究科

¹ Graduate School of Management and Information of Innovation, University of Shizuoka

² 日本学術振興会 特別研究員

² Japan Society for the Promotion of Science, Research Fellow

³ 和歌山大学 システム工学部

³ Faculty of Systems Engineering, Wakayama University

Abstract: In this paper, we attempt to detect change points of a dynamic network structure. We focus on the nodes' functions in a network and define the node's function as the convergence curve of the PageRank score. For each node, we calculate the correlation coefficients between the convergence curves in adjacent two snapshots of a time-varying network. Then, we propose the average of correlation coefficients of all nodes as a measure of the change point of a network structure and refer to this measure as average similarity. Especially, when the average similarity shows the lower value, we assume that the network structure changes significantly. In our experiments using synthetic and real networks with artificial changes, we evaluate the effectiveness of our proposed measure.

1 はじめに

近年、Web上でのハイパーリンクネットワークや、SNS内でのユーザ関係ネットワークなど、現実のあらゆる場面で複雑ネットワークが観察されるようになってきている。ネットワークの構造を分析し理解することは、様々な活動や現象を理解することにつながる。無向ネットワーク構造の指標として、平均クラスタ係数や平均ノード間距離 [Watts 98]、次数分布のべき指数 [Albert 02] など、有向ネットワーク構造の指標として、モチーフ [Milo 02] などがあげられる。これらの指標の多くはネットワークの静的な構造に着目したものである。

現実のネットワークでは、時間とともにネットワークのリンク構造が動的に変化するが、どのようなパターンで、どのような周期で変化するかなどを検出する研究が広く行われている。しかし先行研究ではネットワーク上のどのような性質を持つノード間に、どの程度の、どのような変化が発生したかについてはわからない。本研究では、各ノードのネットワークに対する機能に着目し、その機能が変化した度合いを新たな指標として

提案する。ノードの機能的な特徴は PageRank 収束曲線のパターンとして現れ、類似パターンのノードは同じ機能を持つと推測できる [伏見 12]。本稿では、動的ネットワークの構造変化に伴い、各ノードの果たす役割も変化すると仮定して、その変化を定量化する。具体的には、ネットワーク構造に埋め込まれた各ノードの機能を PageRank スコアの収束曲線で表し、ある時刻におけるノードの機能の変化を、その前の時刻の収束曲線との相関係数で計算する。さらにある時刻におけるネットワーク全体の機能変化を表す指標として、各ノードの相関係数の平均である平均類似度を用いる。この平均類似度が小さくなる時点で、ネットワークの多くのノードの機能が変化したと考える。評価実験では、人工ネットワークなどに対し人工的な変化を発生させ平均類似度の挙動について評価する。

本稿の構成は以下の通りである。最初に提案指標について 2 章で詳しく説明する。3 章において、変化の規模、集中具合、対象ノードの違いなどの点から評価し、結果について考察する。動的ネットワークの構造変化に関する先行研究について 4 章で簡単に触れ、最後に 5 章で本稿をまとめる。

*連絡先：静岡県立大学経営情報イノベーション研究科
〒422-8521 静岡県静岡市駿河区谷田 52-1
E-mail: takayasu.fushimi@gmail.com

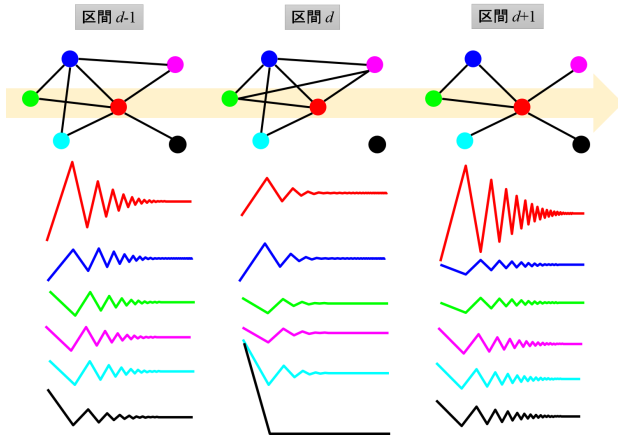


図 1: ネットワーク構造と収束曲線

2 提案指標

本研究では、各ノードのネットワークに対する機能に着目し、動的ネットワークの変化に伴いその機能が変化すると仮定し、各ノードの機能変化の平均により、ネットワーク構造の変化を定量化した指標を提案する。

提案指標では、時系列における任意の区間 d でのネットワークのスナップショットを $G^{(d)}$ とし、ネットワーク系列 $\mathcal{G} = (G^{(1)}, \dots, G^{(D)})$ が与えられ、以下のステップで変化点を検出する（図 1 参照）。

1. 各区間のネットワーク $G^{(d)}$, ($d = 1, \dots, D$) において、各ノード $u \in V$ の収束曲線 $x_u^{(d)}$ を計算;
2. 各ノード $u \in V$ に対して、隣接区間 d と $d+1$ の収束曲線間の相関係数 $r_u^{(d)}$ を計算;
3. 全ノードの相関係数の平均 $\bar{r}^{(d)}$ を計算;
4. $\bar{r}^{(d)}$ の値が小さい区間 d と $d+1$ の間の時刻を大きな変化が起きた時刻として抽出;

次節で、ネットワークにおけるノードの機能の定量化と、ある区間のネットワークにおける各ノードの PageRank スコアの収束曲線について説明する。

2.1 ノードの機能・役割

文献 [伏見 12] において、各ノードの機能・役割はネットワーク構造に埋め込まれていると仮定している。ノードの機能として、ネットワーク内での階層的地位や相対的な位置、次数や周辺ノードの次数、周辺ノードとのつながり方などを意図しているが、これらが類似するノード同士は、PageRank スコアの収束過程も類似すると推測できる。従って、ネットワーク構造上でのランダムウォークのモデルである PageRank [Langville 04]

を用いて、各ノードの機能を表す特徴ベクトルを計算する。

以下に、ネットワーク構造から窺い知ることのできるノードの機能を表す収束曲線の計算法を以下に示す。収束曲線は、大域ジャンプなしの PageRank を用いて計算する。無向ネットワーク $G = (V, E)$ の各ノードに 1 から $|V|$ までの整数値を一意的に割り振る。ここで、 $(u, v) \in E$ のとき $a(u, v) = 1$ 、それ以外の場合 $a(u, v) = 0$ とし隣接行列 $A \in \{0, 1\}^{|V| \times |V|}$ を定義する。各ノード $u \in V$ に対して、 $\Gamma(u)$ をノード u の隣接ノード集合とする。すなわち、 $\Gamma(u) = \{v \in V; (u, v) \in E\}$ となる。ここで、行推移確率行列 P の各要素を $p(u, v) = a(u, v)/|\Gamma(u)|$ とする。各ノードの PageRank スコアを要素としたベクトル y は、 $y(v) \geq 0$ で $\sum_{v \in V} y(v) = 1$ となる。初期ベクトルを $y_0 = (1/|V|, \dots, 1/|V|)^T$ とし、繰り返しステップ数 s を用い、PageRank スコアベクトル y は以下の更新式の極限分布として定義される:

$$y_s^T = y_{s-1}^T P \quad (1)$$

ここで b^T は b ベクトルの転置を表わす。また、ノード u に注目すると、

$$\begin{aligned} y_s(u) &= \sum_{v \in \Gamma(u)} y_{s-1}(v) \cdot p(v, u) \\ &= \sum_{v \in \Gamma(u)} \frac{y_{s-1}(v)}{|\Gamma(v)|} \end{aligned} \quad (2)$$

で計算される。PageRank スコアの反復計算を所定の回数 S まで繰り返し、各反復回数でのノード u の値を要素としたベクトルを $x_u = (y_1(u), y_2(u), \dots, y_S(u))^T$ と定義する。このベクトル x_u をノード u の収束曲線と呼ぶ。各ノードの収束する値は、各ノードの次数のみでは決まらない。周辺ノードの影響や周辺ノードとの相対的な位置関係、ネットワーク構造の影響を受ける。初期ベクトル $y_0 = (1/|V|, \dots, 1/|V|)^T$ で収束曲線を計算する。

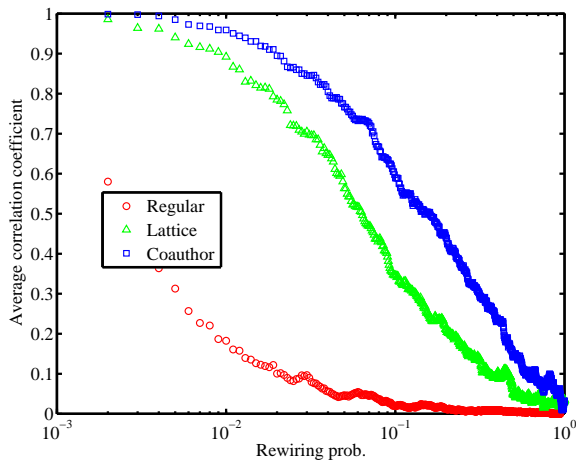
3 評価実験

提案指標の有効性を示すために、人工データを用いて 4 つの観点から評価する。

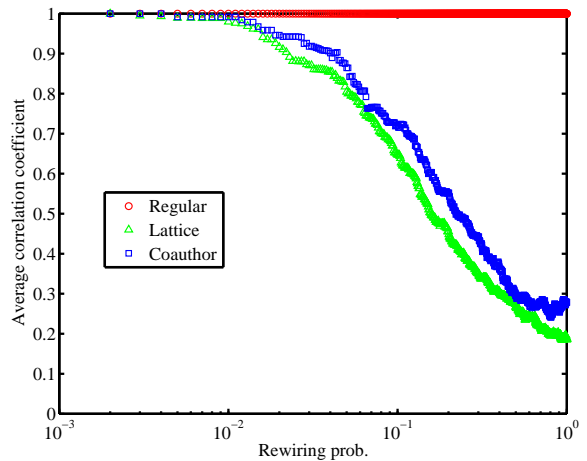
3.1 データセット

1 つ目のデータは、レギュラーグラフである。本稿では、5 近傍ノードとリンクを張り次数 10 のレギュラーグラフを使用する。ノード数は 1,000、無向リンク数は 5,000 である。本稿では Regular ネットワークと呼ぶ。

2 つ目のデータは、2 次元平面上の正方格子である。本稿では、格子点を縦に 30、横に 30 に配置し、ノード



(a) 次数分布可変張替



(b) 次数分布不変張替

図 2: 張替確率と平均類似度

ド数は 900, リンク数は 1,740 のネットワークを作り, Lattice ネットワークと呼ぶ.

3 つ目のデータは, ネットワーク分析のベンチマークとして広く用いられている, 複雑ネットワーク研究分野の共著関係ネットワークである [Newman 06]. ノード数は 379, リンク数は 914 である. 社会ネットワークの特徴であるスケールフリー性とスモールワールド性を有する. 本稿では Coauthor ネットワークと呼ぶ.

4 つ目のデータは, 階層構造を持つネットワークである. 階層性のあるネットワークとは, 企業内の社員のネットワークや Web サイトのハイパーリンクネットワークのようにトップノードと他のすべてのノード間にはリンクが張られているが, その他のノード同士は限られた範囲でのみリンクが張られているような構造を持っている. すなわちトップノード (社長やトップページほか) は高い次数を有しているが, クラスタ係数が非常に小さいことになる. 一方, その他のノード (一般社員や普通のページほか) は低い次数を有しているが, 狭い範囲内で密につながっているためクラスタ係数が大きくなる. 本稿では, Ravasz らによって提案された HN モデル [Ravasz 03] によりノード数は 125, リンク数は 410 のネットワークを作り, Hierarchical ネットワークと呼ぶ.

3.2 変化の規模に関する評価結果

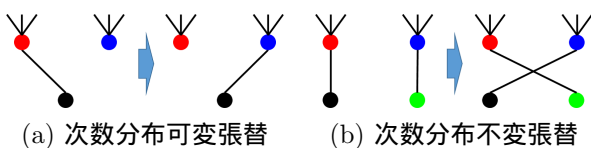


図 3: 張替法

リンク構造の変化の規模が大きいほど, 提案指標である平均類似度は低くなるのが望ましい. この節では, 2 種類のリンク張替による構造変化を人工的に生成し, 平均類似度が張替リンク数に比例することを検証する.

1 つ目の張替手法として, Watts らのスモールワールド実験 [Watts 98] で用いられた張替法をベースとした張替を考える. 各ノードに対して, リンク先のノードを張替確率 p で変更する (3(a)). 但し, 多重リンクや自己リンクを形成しないように張替える. 本稿では, この張替法を次数分布可変張替と呼ぶ.

2 つ目の張替手法として, Newman のコンフィギュレーションモデル [Newman 03] を採用する. 各リンクに対して, 張替確率 p で他のリンクと接続先ノードを交換する (3(b)). 但し, 多重リンクや自己リンクを形成しないように張替える. 本稿では, この張替法を次数分布不変張替と呼ぶ.

図 2 に, Regular, Lattice, Coauthor ネットワークに対して, 2 つの張替法を張替確率を変えて生成したネットワークによる評価結果を示す. 横軸に張替確率, 縦軸に平均類似度をプロットした.

図 2(a) に次数分布可変張替での結果を示す. 3 つのネットワークすべてにおいて, 張替確率を大きくするにつれ平均類似度も減少傾向にあり, 変化の規模に比例した尺度であることが確認できた. しかし, Regular ネットワークと他の 2 つのネットワークでは, 提案指標の減少の仕方が異なっている. Regular ネットワークのような規則正しいネットワークでは, すべてのノードの収束曲線は一定値であるため, 少しでも変化が起きた場合, 変化が発生したノードおよびその周辺ノードの収束曲線が大きく変化してしまう. そのため, 張替確率が低い状態で, 平均類似度が顕著に大きく減少する.

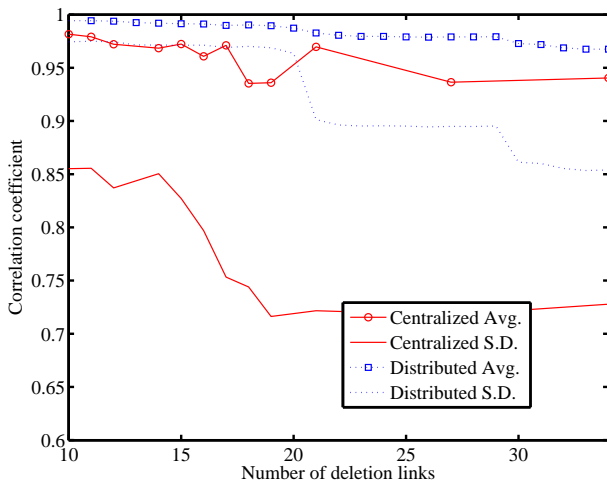


図 4: 集中度と平均類似度

図 2(b) に次数分布不変張替での結果を示す。Regular ネットワークでは、張替確率に関係なく平均類似度が 1.000 となり一定である。Regular ネットワークは、元々すべてのノードの次数が等しく、張替後においてもすべてのノードの次数が等しい。そのため、PageRank の各ステップにおける流入量と流出量のバランスが始終一定であり、各ノードの収束曲線は全く変化しないためこのような結果が得られた。Lattice ネットワークと Coauthor ネットワークでは、次数分布可変張替と同様に、張替確率を大きくするにつれ平均類似度も減少傾向にあり、変化の規模に比例した尺度であることが確認できた。すなわち、各ノードにおいて、少ないステップで辿り着けるノード群の次数が変化することは、各ノードの周辺環境が変化したことを意味し、提案指標はその大きさを定量化できることが示唆された。

3.3 変化の集中度に関する評価結果

リンク構造の変化は、あるノード周辺に集中的に発生する状況や、ネットワーク全体に分散的に発生する状況が考えられる。小さな変化が分散して発生するより、大きな変化が集中して発生することの方が意義深いと考えられる。これらの変化を人工的に生成し、提案指標である平均類似度の挙動について検証する。集中的なリンク構造変化として、ある高次数ノードのリンクをすべて削除することにより人工的に変化を発生させる。分散的なリンク構造変化として、全リンク集合からランダムに選択したリンクを削除することにより人工的に変化を発生させる。

図 4 に、Coauthor ネットワークに対する評価結果を示す。横軸に削除リンク数、縦軸に平均類似度および平均類似度から $(-1) \times$ 標準偏差の値をプロットした。

実線で集中的リンク削除、点線で分散的リンク削除の結果を示す。平均類似度に関しては、同数のリンクを削除しても、分散的リンク削除より集中的リンク削除の方が幾分か低いが、大きな違いはない。一方、収束曲線の類似度の標準偏差に関しては、分散的リンク削除より集中的リンク削除の方が顕著に大きいことがわかる。一か所に集中して起きる変化の場合、変化が起きたノードおよびその周辺ノードは収束曲線が大きく変化し、収束曲線の類似度は低下するが、離れているノードには大きな影響がなく収束曲線の類似度が高い状態となるため、類似度の値に幅ができることを表している。すなわち、集中的リンク削除の方がネットワーク構造に大きくインパクトを与えていると考えられる。

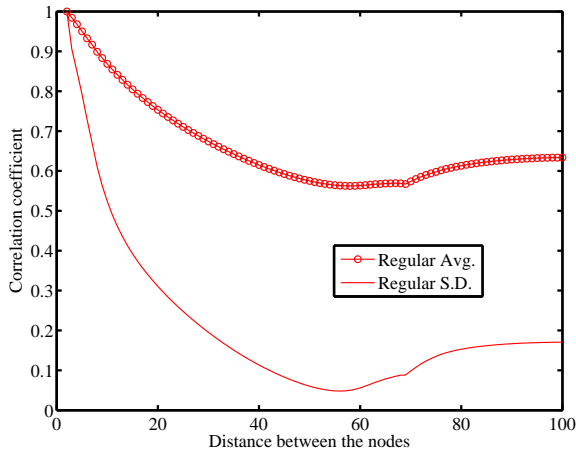
3.4 変化対象の距離に関する評価結果

リンク構造の変化は、近接するノード同士にリンクが追加される状況や、比較的遠くに存在するノードとの間にリンクが追加される状況が考えられる。これらの変化を人工的に生成し、提案指標である平均類似度の挙動について検証する。近くのノードより遠くのノードへのリンクが追加されたことを検出することの方が意義深いと考えられる。なぜならば、遠くに位置するノード同士にリンクが張られることで、ネットワーク全体の平均ノード間距離に大きな影響を及ぼすからである。そのことを反映し、追加リンクの両端ノード間距離が大きいほど、提案指標である平均類似度は低くなることを望ましい。Regular ネットワークと Lattice ネットワークに対して、あるノードからの距離に従って、リンク対象を選択し、それぞれ 1 本のリンクを追加することで変化を発生させる。

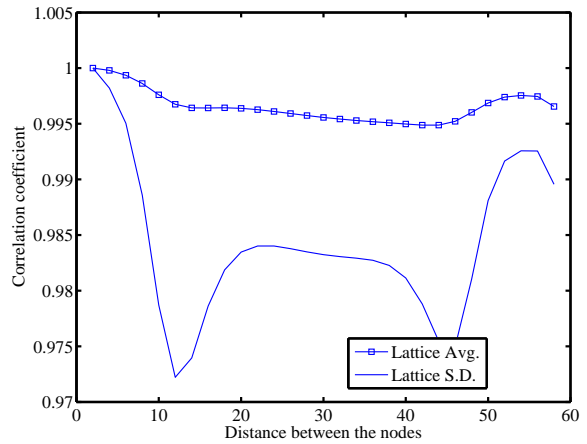
図 5 に、Regular, Lattice ネットワークに対する評価結果を示す。横軸に追加リンクの両端ノード間の（リンク追加前）距離、縦軸に平均類似度および平均類似度から $(-1) \times$ 標準偏差の値をプロットした。

図 5(a) に Regular ネットワークに対する結果を示す。追加リンクの両端ノード間距離が大きいほど、より多くのノードの収束曲線に影響を与えるため、平均類似度の値が減少する傾向にあるのがわかる。反対に、両端ノード間距離が小さいと周辺ノードが重複するため、一部のノードにしか大きな影響を与えられないため、比較的減少幅は小さいこともわかる。

図 5(b) に Lattice ネットワークに対する結果を示す。Regular ネットワークと同様に、追加リンクの両端ノード間距離が大きいほど、平均類似度の値が減少する傾向にある。Lattice ネットワークは Regular ネットワークと異なり、中心部分のノードと外側部分のノードでは性質が異なるので、遠くに存在するノードにリンクを付与しても、近くに存在するノードへのリンクより



(a) Regular ネットワーク



(b) Lattice ネットワーク

図 5: 張替確率と平均類似度

影響が小さいとは、一概には言えない。しかしどちらのネットワークでも、両端ノード間距離が大きい追加リンクほど、収束曲線への影響は大きい傾向にある。

3.5 変化対象の性質に関する評価結果

ネットワークにおいて、すべてのノードは均質ではなく、ハブノードやぶら下がりノードなどが存在する。ネットワークの末端に存在するようなノード同士のリンク追加・削除などの変化より、ハブノードとのリンクが追加・削除されることを検出することの方が意義深いと考えられる。この節では、Hierarchical ネットワークに対して、リンクが追加される対象ノードの性質と平均類似度の挙動について 2 つのケースにより検証する。いずれのケースも 3.4 節の知見を踏まえて、距離が等しいノードペアで比較する。

ケース 1 として、PageRank スコアが最も高い赤ノードと最も低い青ノード間にリンクを追加する(図 6(a))。図 6(b) に、ケース 1 における代表ノードの収束曲線を示す。但し、平均 0、ノルム 1 に正規化してある。点線がリンク追加前、実線がリンク追加後を表す。リンクを追加したハブノードである赤ノードの収束曲線は、リンク追加前後で大きな変化は起きておらず、前後の相関係数は 0.999 である。これは、ハブノードに 1 つ一般ノードへのリンクが追加されたところで、大きな影響がないという直観に対応する。一方、リンクを追加した非ハブノード(一般ノード)である青ノードの収束曲線は、リンクの追加前後で大きな変化が起きており、前後の相関係数は 0.144 である。これは、一般ノードに 1 つハブノードへのリンクが追加されることが、大きな影響を与えるという直観に対応する。リンクを追加した周辺ノードである、桃ノードに関して、大きな影響を受けていないが、水

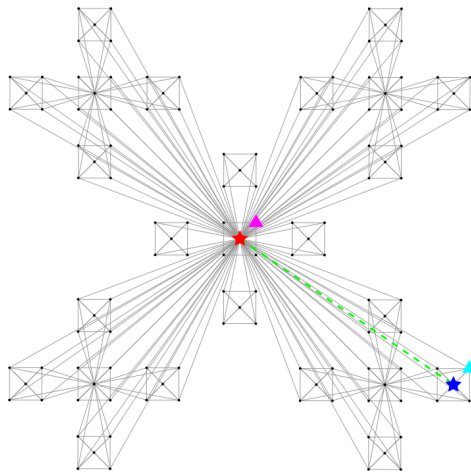
ノードには、青ノードほどではないが大きな影響を受けている。

ケース 2 として、PageRank スコアが低い赤ノード、青ノード間にリンクを追加する(図 7(a))。図 7(b) に、ケース 2 における代表ノードの収束曲線を示す。リンクを追加した赤ノード(青ノードも同様)の収束曲線は、リンク追加前後である程度の変化が起きているが、前後の相関係数は 0.946 であり、ケース 1 の場合に比べると変化は少ない。これは、一般ノード同士にリンクが 1 つ追加されたところで、大きな影響がないという直観に対応する。周辺ノードに関して同様に、桃ノード(水ノードも同様)の収束曲線は、赤ノードより変化が小さく、前後の相関係数は 0.952 である。茶ノード(紺ノードも同様;追加前は赤ノード、青ノードと等しい)の収束曲線は、桃ノードよりさらに変化が小さく、前後の相関係数は 0.991 である。

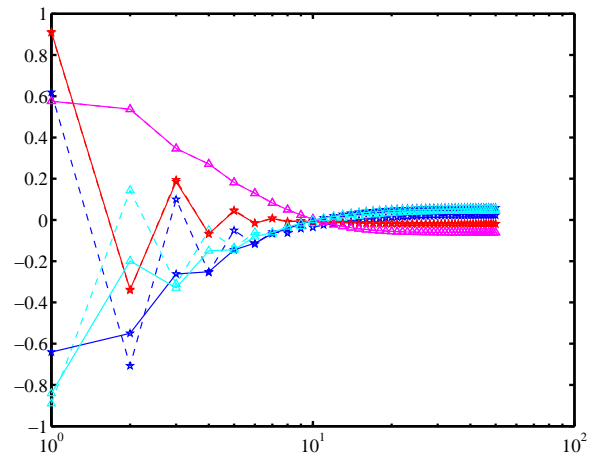
ネットワーク全体の平均類似度は、ケース 1 では 0.890、ケース 2 では 0.997 であり、追加リンク数が少ないため有意な差は出ていないが、変化対象ノードの性質の違いにより、平均類似度の値も異なることが示唆された。

4 関連研究

現実のネットワークでは、時間とともにネットワークのリンク構造が動的に変化するが、どのようなパターンで、どのような周期で変化するかなどを検出する研究が広く行われている。文献 [高田 06] では、モチーフパターンの推移数の Z スコアにより、有意なパターン変化を抽出している。文献 [安田 06] では、人工知能学会の共著関係ネットワークに対して基本統計量や中心

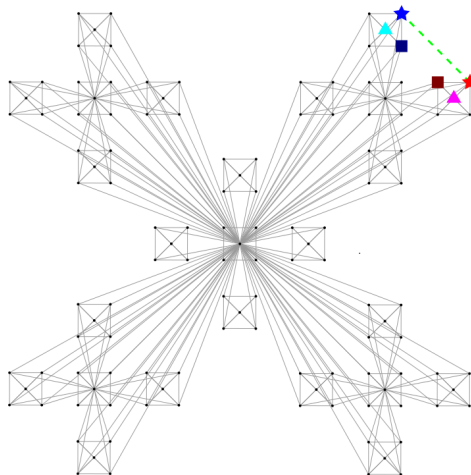


(a) 可視化結果

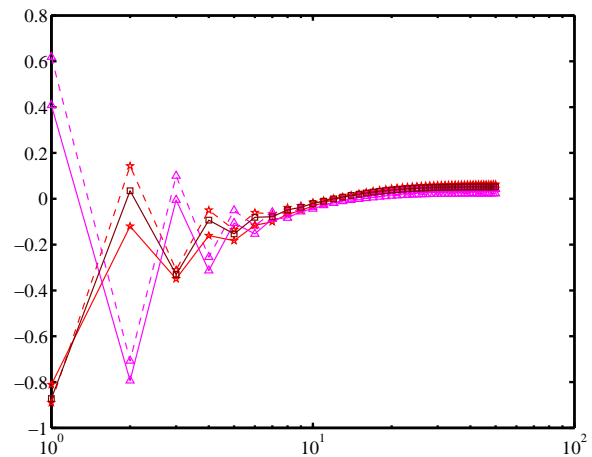


(b) 収束曲線 (鎖線: オリジナル, 実線: リンク追加後)

図 6: ケース 1 の実験結果



(a) 可視化結果



(b) 収束曲線 (鎖線: オリジナル, 実線: リンク追加後)

図 7: ケース 2 の実験結果

性について比較し,年ごとのネットワーク構造の違いについて考察している.文献 [Koujaku 13] では,密につながるノードが分離したり,疎な状態のノードが密につながるなど大きな変化を検出するべく,隣接行列の固有値分解を利用し動的ネットワークに対する異常値検出法を提案している.文献 [渡部 14] では, Twitter ネットワークを対象に,成長モデルの観点からフォロー/フォロー解除のパターンをモデル化し,時間的構造変化の解析手法を提案している.文献 [Peel 14] では,階層ランダムグラフモデル [Clauset 08] を一般化した GHRG モデルを用いて,ネットワークのコミュニティ階層性を学習し,尤度比検定により有意に階層性が変化した時点を検出する手法を提案している.

本研究では,各ノードの機能・役割に着目し,時間とともに変化するネットワーク構造の中でノードの果たす役割も変化するかと仮定し,その変化の平均値によりネットワーク構造の変化を定量化している.これにより,変化が発生した場所やネットワーク全体に与えるインパクトなどを考慮した指標となっている点で,先行研究とは異なる.

5 おわりに

本稿では,動的にリンク構造の変化するネットワークに対して,大きな影響を及ぼす変化を検出するため

の指標を提案した。複数のネットワークに人工的な変化を発生させた評価実験より、提案指標である PageRank 収束曲線の平均類似度は、1) 変化の規模に比例した値を示し、2) 集中的に発生する変化、3) 対象ノード間の距離が大きいような変化、4) 重要ノードを巻き込んだ変化を検出可能であることを示した。ネットワーク構造の指標であるクラスタ係数、ノード間距離、トライアドを観測するだけでは、変化が起きたことは検出できても、変化の重大性を反映するのは困難である。今後は、多様なネットワークの実際の変化に対して、提案指標が適用できるかについて評価を進めていきたい。

謝辞

本研究は科学研究費補助金 (No.25・10411) の補助を受けた。

参考文献

- [Albert 02] Albert, A. and Barabási, A. L.: Statistical mechanics of complex networks, *Rev. Mod. Phys.*, pp. 47–97 (2002)
- [Clauset 08] Clauset, A., Moore, C., and Newman, M. E. J.: Hierarchical structure and the prediction of missing links in networks, *Nature*, Vol. 453, pp. 98–101 (2008)
- [Koujaku 13] Koujaku, S., Kudo, M., Takigawa, I., and Imai, H.: Structural Change Point Detection for Evolutional Networks, in *Proceedings of the World Congress on Engineering*, Vol. 1 (2013)
- [Langville 04] Langville, A. N. and Meyer, C. D.: Deeper inside pagerank, *Internet Mathematics*, Vol. 1, p. 2004 (2004)
- [Milo 02] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U.: Network motifs: simple building blocks of complex networks., *Science (New York, N.Y.)*, Vol. 298, No. 5594, pp. 824–827 (2002)
- [Newman 03] Newman, M. E. J.: The structure and function of complex networks, *SIAM Review*, Vol. 45, pp. 167–256 (2003)
- [Newman 06] Newman, M. E. J.: Finding community structure in networks using the eigenvectors of matrices, *Physical Review E*, Vol. 74, No. 3, pp. 036104+ (2006)
- [Peel 14] Peel, L. and Clauset, A.: Detecting change points in the large-scale structure of evolving networks, *CoRR*, Vol. abs/1403.0989, (2014)
- [Ravasz 03] Ravasz, E. and Barabási, A. L.: Hierarchical organization in complex networks, *Physical Review E*, Vol. 67, No. 2, pp. 026112+ (2003)
- [Watts 98] Watts, D. J. and Strogatz, S. H.: Collective dynamics of ‘small-world’ networks, *Nature*, Vol. 393, No. 6684, pp. 440–442 (1998)
- [安田 06] 安田 雪, 松尾 豊, 武田 英明: 人工知能学会におけるネットワーク構造と変化, 第 20 回人工知能学会全国大会 (JSAI2006) (2006)
- [高田 06] 高田 寛喜, 斉藤 和巳, 上田 修功: 時系列情報を考慮したモチーフパターン変化の分析, 電子情報通信学会第 7 回 Web インテリジェンスとインタラクション研究会 (2006)
- [渡部 14] 渡部 優, 鈴村 豊太郎: 大規模ソーシャルネットワークにおける時間的変化の解析手法の研究, 第 6 回データ工学と情報マネジメントに関するフォーラム (DEIM2014) (2014)
- [伏見 12] 伏見 卓恭, 斉藤 和巳, 風間 一洋: ネットワーク機能コミュニティ抽出法, 日本データベース学会論文誌, Vol. 10, No. 3, pp. 13–18 (2012)