

SNS 型の特徴量分布を持つグラフ生成モデルに向けて Towards graph generation model with SNS-type network metrics

○石田咲起人¹, 平石秀史²
*Sakito Ishida¹, Hiraishi Hidefumi²

Abstract: Graph generation models are a well-researched and interdisciplinary area covering fields such as physics, computer science, biology, chemistry, social science and others. In this research, we incorporate a procedure to simulate 'going viral' into the classical Watts-Strogatz model, aiming to achieve SNS-type network metrics. We experimentally evaluate network metrics of our model.

1. はじめに

現実に現れる様々なデータには、繋がりを表すグラフ構造で表現できるものが多い。SNS のフォロー関係のような人の繋がりのデータや、路線図のような場所の繋がりのデータは代表例である。一方、同じグラフ構造でもデータの属するドメインにより特徴は大きく異なる。各々のドメインに現れるグラフ構造の不変量・特徴量を人工的に再現するグラフ生成モデルの研究は、ネットワーク科学と呼ばれる分野で盛んに行われており、Watts-Strogatz モデル [1]・Barabási-Albert モデル [2] をはじめ多くのモデルが提案されている。

本稿では Watts-Strogatz モデルをもとに、SNS に特異的なネットワーク特徴量の分布を再現するネットワーク生成モデルを構築することを目指して、予備的な実験を行う。また、実験結果をもとに、SNS が持つと想定されるネットワーク特徴量の傾向について検討を行う。

2. ネットワークの特徴量と特性

本稿では次の3つの特徴量に着目し、実験を行うこととする。以下では実験の前提となる特徴量やネットワーク生成モデルについて説明をする。各特徴量や生成モデルの詳細については [3] に詳しい。今回は無向グラフを考える。また頂点集合 $V = \{1, 2, \dots, N\}$ とし、 E を V 上の辺集合として、グラフを $G = (V, E)$ と表す(この場合、頂点数は N である)。

2.1 クラスタ係数

クラスタ係数とは、局所的な辺の密度を表す指標である。まず、頂点 i と隣接している頂点集合を頂点 i の近傍 N_i とし、 N_i で誘導される部分グラフの辺数を m_i と表す。このとき頂点 i のクラスタ係数 C_i は、次数 k_i と辺数 m_i を用いて次のように定義される。

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

クラスタ係数の平均値が高いとき、クラスタ性をもつという。クラスタ性をもつようなネットワークでは、「友達の友達は友達」という状況が頻繁に起きている状況を意味する。

2.2 次数中心性

中心性とは、ある頂点や辺がネットワーク全体にとってどの程度重要かを表す指標のことである。様々な種類の中心性の指標が存在するが、今回は最も基礎的な中心性である次数中心性に着目し実験を行う。ただし、頂点 i の次数とは、その頂点に接続されている辺の数のことをいう。

次数中心性とは、ある頂点がネットワーク全体の大きさに比して、どの程度の割合の頂点数と隣接しているかを表す指標である。また次数中心性 K_i は、頂点数 N と頂点 i の次数 k_i を用いて次のように定義される。

$$K_i = \frac{k_i}{N - 1}$$

現実に存在する大規模なネットワークでは、次数の分布がべき乗則に従う場合が多いことが経験的に知られている。べき乗則に従うとは、ある頂点が次数 k を持つ確率密度関数を $P(k)$ とすると以下の関係をもつことをいう。

$$P(k) \propto k^{-\gamma}$$

このような次数分布に関する性質をスケールフリー性という。この性質の特徴として、極端に大きい次数も、小さい確率ながらある程度存在しうることが挙げられる。

2.3 平均最短経路長

ネットワークに含まれる頂点全てのペアの最短経路長を平均した値のことである。

人と人の繋がりを表すネットワークではスモールワールド性と呼ばれる特徴を持つことが知られている。スモールワールド性とは、平均最短経路長が $O(\log N)$ 程度になる性質と定義される場合が多い。

2.3.1 Watts-Strogatz (WS) モデル

WSモデルによるネットワークは、以下のようなアルゴリズムによって形成される。

- (1) N 個の頂点をサイクルグラフ状に並べる。
- (2) 各頂点とその z -近傍までの頂点との間に辺を結ぶ。このとき、総辺数は zN であり、各頂点の次数は $2z$ となる。
- (3) (2) で張られた zN 本の辺のそれぞれに対し、確率 p で辺の一方の端(ランダムに選ぶ)の結合を切り離し、 N 個の頂点からランダムに選ばれた頂点に繋ぎ替える。この際、自己ループや多重辺ができないようにする。

3. SNS 型ネットワークモデル構築に向けた基礎実験

WSモデルでは、多くの頂点で同じような次数を持つネットワークが現れる。本研究では、先のような構成から次数が特別多い頂点をどのように作ることができるのかを考える。WSモデルに、SNSで頻繁に起こる“バズる”という現象を組み込んだモデルを作成し、そのネットワーク特徴量を調べるといふ実験を行う。

“バズる”という現象を表現するために、WSモデルに以下の操作を組み込む。

- (4) 頂点を以下の確率に従って選ぶ。

$$\frac{k_i}{\sum_{i \in V} k_i}$$

最初に選んだ頂点以外に、複数の頂点をランダムに選択し、最初に選んだ頂点との間に辺を張る。選択する頂点の個数は、指数分布に従って決める。

3.1 実験詳細

今回提案したアルゴリズムでは事前に定めるパラメータが複数ある。今回の実験では、頂点数は50000、WSモデルにおける z の値を300、 p の値を0.01、ステップ(4)の繰り返し回数を10000回とする。

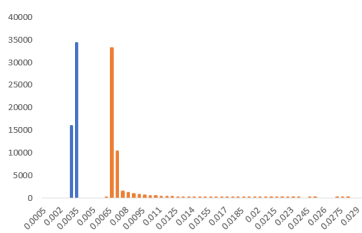


図1: 次数中心性(横軸)のヒストグラム
青色: Watts-Strogatz モデル, オレンジ色: 本実験のモデル

3.2 実験結果

元のWSモデルの辺の数3750000に対し、本モデルの辺の数は4240000であるので、WSモデルと同程度に疎なグラフと考えられる。また、図1、図2に本モデルでの次数中心性やクラスター係数の実験値を載せる。これらを見る限り、「スケールフリー性」「クラスター性」「スモールワールド性」を満たすネットワークを実験的に実現できていることが予想できる。

4. 考察と今後の展望

図1で示した二つの実験結果を比較すると、WSモデルにおいては次数中心性の取りうる値が一箇所に集中している。一方で、本実験モデルの結果には少数の裾の尾が見られることが分かる。この結果はWSモデルの性質を満たしつつ、次数中心性の大きな頂点を生み出すことができたと考えられる。一方、SNSでは、次数が大きい頂点では、その頂点の友達同士が友達である可能性は低いことが自然と想定できる。図2は、本モデルがこの傾向を持つグラフを生成していることを示唆している。

今後の展望として、SNS型ネットワークに近づけるために、図1,2ともに本実験モデルの結果を階段状に下がる形にしたいと考えている。これは人気者となりかけている人とそうでない人の間に、次数中心性の値に明確な差が生じると予想しているためである。この予想を基に新たなアルゴリズムを追加し、他の特徴量に関する実験を行うことで、SNS上の特徴を捉えたネットワークモデルを完成させたいと考える。

5. 参考文献

- [1] D.J.Watts and S.H.Strogatz, *Collective dynamics of 'small-world' networks*, Nature, 1998, 509, 440-442.
- [2] A.-L.Barabási-Albert and R.Albert, *Emergence of scaling in random networks*, Science, 1999, 286, 509-512.
- [3] 矢久保孝介, 複雑ネットワークとその構造, 共立出版, 2013.

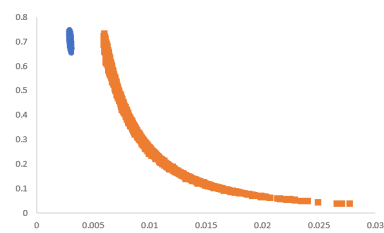


図2: 次数中心性(横軸)とクラスター係数(縦軸)の相関関係
青色: Watts-Strogatz モデル, オレンジ色: 本実験のモデル