

事が出来ていない。統計的に正確な分析を行うためには、一部のネットワーク構造のみネットワーク構造ではなく、図 2 のようなあらゆるネットワーク構造を網羅するようなデータセットを用いる必要がある。本研究では、あらゆる構造を含むネットワークデータセットを構築し、現象とネットワーク構造の関係を明確にする事を目指す。

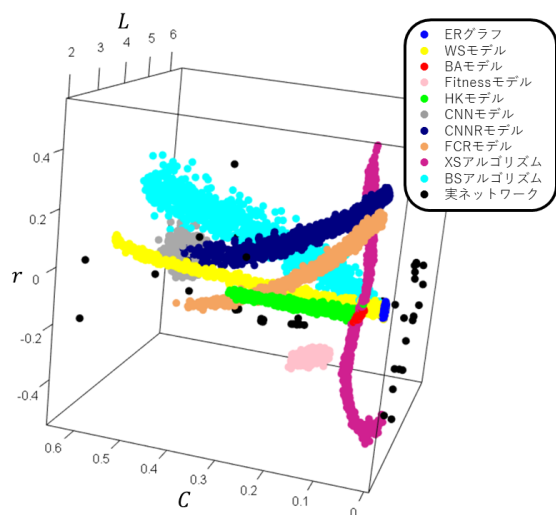


図 1: 既存ネットワーク生成モデルで生成したネットワークの構造

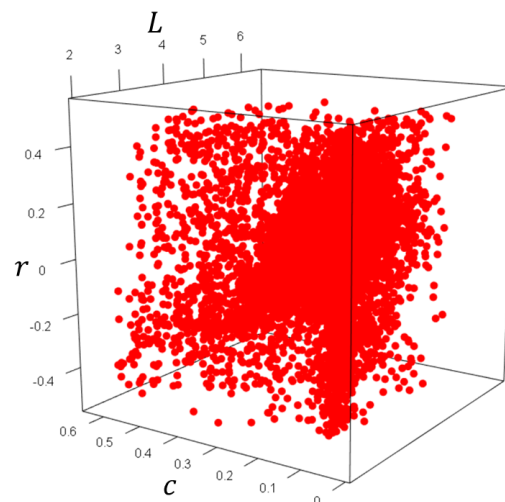


図 2: 提案手法による網羅的ネットワークデータセット

現在提案されている多くのネットワーク生成モデルは、生成したネットワークが図 1 に示すように一定の構造傾向を持つように提案されている。従って、ネットワーク構造を網羅的に生成する事はできない。そこで、この構造特徴空間上の任意の特徴点を取り、その特徴点のネットワーク構造特徴から逆にネットワークを生成する手法を考案する。

本研究では Arbitrary Network Generation Algorithm (ANGA) を提案した。ANGA は、構造特徴空間上の任意の点を設定し、貪欲法とネットワーク成長モデルを用いてネットワークをその特徴点に近いネットワークに成長させるアルゴリズムである。各ステップでは、ノードやリンクを追加する際に複数の候補ネットワークを生成する。候補の生成には、既存モデルで用いられる、優先的選択や Connecting nearest neighbor (CNN) を用いる。次に生成された候補ネットワークの中から、特徴点に最も近い候補ネットワークを選択し、次のステップに進む。これを繰り返し行う事によって、ネットワークは特徴点に近づきながら成長する。

ANGA が様々な特徴点を生成できるかどうかを評価する。本研究で利用するネットワーク構造特徴として、平均経路長、クラスタ係数、次数相関に加え、次数分布にベータ分布をフィッティングしたパラメータ $\log \alpha$ 及び $\log \beta$ を用いた。評価には、実ネットワークから抽出した特徴点と、既存ネットワーク生成モデルによって生成したネットワークから抽出

した特徴点を用いる。これらの特徴点を対象として ANGA によるネットワーク生成実験を行い、ANGA が設定された様々な特徴点に近い点にネットワークを生成できるかを確認した。実験の結果、与えられた特徴点の多くに高精度でネットワークを生成できる事を示した。また、ANGA はノード数とリンク数がある程度大きいネットワークでなければ生成できない事がわかった。また、候補の生成の際には優先的選択や CNN を用いる事が重要な要素となる事も示した。

この ANGA を用いて、様々なネットワークを含むデータセットを構築する。構築には以下の手順を用いて行う。

- ネットワーク構造特徴空間上から任意の特徴点を抽出する。
- 抽出された特徴点に対して ANGA によりネットワークの生成をする。

この手順を繰り返し行う事によって、様々なネットワークを構築し、これをネットワークデータセットとした。本研究で用意したデータセットは 7028 個のネットワークを含む。この 7028 個のネットワークを平均経路長、クラスタ係数、次数相関の 3 次元特徴空間にプロットすると図 2 のようになる。様々な特徴を持つネットワークを含んでいる事がわかる。

提案データセットが様々なネットワークを含んでいる事を定量的に示すために、図 1 に示す既存モデルによって生成した比較対象データセットと比較しながら評価した。その結果、以下のような結果が得られた。

- 提案手法を用いて構築したデータセットは比較対象データセットよりも各特徴間に相関のないデータセットである事を示した。
- 提案手法を用いて構築したデータセットは、比較対象データセットに含まれる 90.1% のネットワークと類似したネットワークを含んでいる事を示した。
- 提案手法によるデータセットは、68 個の実ネットワーク全てに対して、比較対象データセットよりも近い特徴を持つネットワークを含んでいる事を示した。
- 提案手法を用いて構築したデータセットの充填率は 3.40×10^{-2} であり、既存モデルによって構築したデータセットの充填率 3.38×10^{-3} の 10 倍以上であり、提案手法を用いて構築したデータセットは、より様々な構造を含み、空間を満たしている事が示した。

これらの結果から、提案手法によるデータセットは様々なネットワークを含んだネットワークデータセットであり、統計的分析を行うのに有用である事が示された。

次に本データセットを用いて、現象とネットワーク構造の関係について分析を行う。ここでは、現象の例として、構造と関係が深く、古くから分析が多く行われている情報拡散と協調行動を挙げる。

まず情報拡散について、情報の拡散率とネットワーク構造の関係について分析した。本研究では、情報拡散モデルとして、独立カスケード (Independent cascade : IC) モデルを用いる。まず、既存研究で得られた知見との比較を行った。Badham らは次数相関と拡散率

の相関関係を示しているが、本データセットでは、相関関係は得られなかった。この分析によって、Badham らの示した、次数相関と拡散率の相関関係は平均経路長と拡散率の相関関係の偽相関である事を示した。

次に、各特徴と拡散率の間の相関を分析した。拡散率と平均経路長との相関-0.502、クラスタ係数との相関-0.511、次数分布のパラメータ $\log \alpha$ との相関0.398が示された。さらに、本研究で用いた5つの特徴から拡散率の推定を行った。推定にはSVRを用い、シミュレーション結果と推定値の決定係数は0.906であり、この5つの特徴によって、拡散率が決定される事を示した。また、各特徴を除いた4特徴からの推定結果を分析する事によって、各特徴の拡散率の推定に対する寄与度を示した。推定結果から、次数分布のパラメータ $\log \alpha$ が最も寄与度が高く、次数相関 r はほとんど寄与していない事が示された。また、最後に別の情報拡散モデルである線形閾値 (Linear threshold : LT) モデルの拡散率とICモデルの拡散率を比較した。スケールフリーネットワークはICモデルの拡散率に対してLTモデルの拡散率が低い事が示された。このようにして、本章ではネットワーク構造が情報拡散に与える影響を明らかにする事に成功した。

協調行動について、協調戦略エージェントの割合とネットワーク構造の関係について分析を行った。まず、既存研究で得られた知見との比較を行った。Assenza らはクラスタ係数と協調戦略エージェントの割合の相関関係を示しているが、本データセットでは、相関関係は得られなかった。この分析によって、Assenza らの示した、クラスタ係数と協調戦略エージェントの割合の相関関係は、平均経路長と協調戦略エージェントの割合の相関関係の偽相関である事を示した。

次に、各特徴と協調戦略エージェントの割合の間の相関を分析した。協調戦略エージェントの割合と平均経路長との相関-0.198、次数分布のパラメータ $\log \alpha$ との相関0.274、次数分布のパラメータ $\log \alpha$ との相関\$0.372\$が得られた。

さらに、本研究で用いた5つの特徴から協調戦略エージェントの割合の推定を行った。シミュレーション結果と推定値の決定係数は0.651であり、この5つの特徴によって協調戦略エージェントの割合は推定しきれない事がわかった。従って、協調行動に影響を与えるネットワーク構造特徴がこの5つのネットワーク構造特徴以外に存在する事が示された。また、各特徴を除いた4特徴からの推定結果を分析する事によって、各特徴に関して協調戦略エージェントの割合の推定に対する寄与度を示した。推定結果から、次数分布のパラメータ $\log \alpha$ 及び $\log \beta$ が最も寄与度が高く、クラスタ係数 C はほとんど寄与していない事が示された。

この2つの現象の例の分析を通して、本研究では、本データセットを用いて、ネットワーク構造と現象の関係を明らかにできる事を示した。本データセットが有用なのはこの2つの現象に限らない。ネットワーク上で起こる様々な現象とネットワーク構造の関係を明らかにするために、本データセットは有用である。本データセットは、様々な現象の解析に貢献できる事が期待される。