3N3-OS-28a-1in2

潜在トピックに基づく複雑ネットワークの生成モデルに関する基礎検討

A Basic Study on Generation Model of Complex Network based on Latent Topic

赤山 郁人　倉持 俊也　土方 嘉徳
Ikuto Akayama　Toshiya Kuramochi　Yoshinori Hijikata

大阪大学大学院 基礎工学研究科
Graduate School of Engineering Science, Osaka University

Complex networks describe a wide range of systems in nature and society such as the WWW, social network between humans and citation network of research papers. While network generation models proposed so far are designed to reproduce the network structure of the existing complex network, some real networks are created according to topics which each node is interested in. We propose a network generation model that creates links according to the topic distributions of each node. In this work, we implemented the proposed model and conducted an initial evaluation.

1. はじめに

多くの研究者によって World Wide Web や人の社会ネットワーク，論文の被引用関係のネットワークなどの複雑ネットワークに対する解析が行われ，様々な共通の性質が観測されている [Albert 02]．代表的な性質にスモールワールド性，スケールフリー性，クラスタ性がある．これらの性質を作り出すネットワークの振る舞いに対する知見を得るために，今までネットワーク生成モデルが幾つも考えられてきた [Albert 02]．

しかし，従来のモデルはネットワークにおける構造上の特徴を作るためにエッジが生成される過程をモデル化したものであり，ネットワークにおけるエッジの持つ意味を扱ってこなかった．実際に人と人とが関わりを持つ時，すなわち 2 つのノードが互いの間にエッジを生成するときには，所属するコミュニティのつながりや趣味のつながり，興味のつながり等を考慮している．特に Twitter のフォローネットワークは社会的な知人関係だけでなく，興味や関心の関係も表していると報告されており [Kwak 10]，意味を考慮したエッジ生成の機構がネットワーク生成モデルにも必要であるといえる．

そこで我々はエッジが生成されている 2 つのノード間には共通の潜在的なトピックが存在している可能性に着目し，トピックに基づいてエッジを生成するネットワーク生成モデルを提案する．すなわち，各ノードはトピックを選択する確率分布をもっており，それによってトピックを選択する．さらに，トピックはノードを選択する確率分布を持っており，それによってノードを選択する．これは自然言語処理の分野において有名なトピックモデルである LDA(Latent Dirichlet Allocation)[Blei 03] が前提とする文書が生成される過程と同じである．そのため，提案モデルには LDA に使われる変数を使用する．

本稿ではまず我々が提案するモデルを示し，計算機によるネットワーク生成のシミュレーションを行う．そして提案するモデルが多様な複雑ネットワークを生成できるかを考察する．

連絡先: 赤山 郁人，大阪大学大学院 基礎工学研究科，
大阪府 豊中市 待兼山町 1-3，06-6850-6383，
akayama@hlab.sys.es.osaka-u.ac.jp

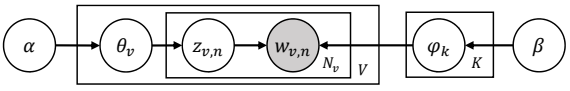

図 1: トピックに基づくエッジ生成過程のグラフィカルモデル

2. 関連研究

2.1 ネットワーク生成モデル

スモールワールド性を作り出す要因を探るため，ネットワークの成長過程に着目した研究が多くなされ，スケールフリーなネットワークを生成する BA モデル [Barabasi 99] が提案されている．また，BA モデルの派生として Holme-Kim モデル [Holme 02] や適応度モデル [Bianconi 01] が提案されている．これらはネットワーク全体を統一的にとらえたモデルである．本研究ではネットワークは個々のノードが持つ潜在トピックに基づいて生成されると考える．

2.2 LDA のネットワークへの応用

LDA は自然言語処理の分野でよく用いられる機械学習の手法である．近年 LDA をネットワークデータに応用したアプローチの研究が進められており，LDA を使ったコミュニティ抽出が盛んに行われている．倉持らは，LDA の考えに基づくトピックの概念を持つネットワーク生成モデルを仮定し，Twitter のフォローネットワークを使い，トピック推定が正しく行われているかどうかを検証するためにリンク予測を行った [Kuramochi 13]．Henderson らの提案している LDA-G は，ノード v の隣接ノード集合を文書 d の単語集合とみなした LDA であり，LDA で使われる機械学習のアルゴリズムを使うことで，コミュニティの発見を行った [Henderson 09]．

3. 提案手法

本研究では，図 1 に示したエッジを生成する過程を既存のネットワーク生成モデルに組み込む．それを新たな複雑ネットワークの生成モデルとして定義する．図 1 はネットワークにおいてノード d がトピック k を選択し，トピック k がノード w を選択する過程を表したものである．また，α, β はハイパーパラメータ，v, w はノード，z はトピック，K はトピック数，θ はトピックを選択する確率分布，ϕ はノードを選択する確率

アルゴリズム 1 ： 提案モデル

Start with K_m complete graph
repeat
  if Probability p **then**
    New node v_{new} is added to V, v_{new} has no edges.
    $n \leftarrow n + 1$
    for $i = 0$ to m **do**
      Topic z is selected distributed under $\theta_{v_{new}}$.
      Node v_{old} is selected distributed under ϕ_z.
      New edge $e_{new} = (v_{new}, v_{old})$ is added to E.
    end for
  else
    Node x is selected distributed under outdegree.
    Topic z is selected distributed under θ_x.
    Node y is selected distributed under ϕ_z.
    New edge $e_{new} = (x, y)$ is added to E.
  end if
until $n = \#$ of node.

分布をそれぞれ表す．図 1 の一連のプロセスではエッジに向きが存在する．従って，まず無向グラフで考えられてきた既存のモデルを有向グラフに対応させる．その後，トピックに基づいたエッジ生成過程を有向グラフに対応させた既存モデルに導入する．以上の手順で提案モデルを考えた (アルゴリズム 1)．具体的には，我々は既存のネットワーク生成モデルとして，スケールフリー性を持つネットワークを生成する BA モデル [Barabasi 99] を選択し，有向グラフに対応させた．そして，図 1 のアルゴリズムを導入した．

4. シミュレーション実験

提案モデルのネットワーク生成を計算機上でシミュレーションした．また，生成したネットワークに対して任意の 2 つのノード間の平均経路長，クラスタ係数を計算した．ただし，ノード数が 2000 に到達した時，提案モデルのネットワークを生成するアルゴリズムを停止した．

その結果，ハイパーパラメータ β の値を増加させると平均経路長，クラスタ係数の値が減少することが確認された (図 2)．β はノードを選択する確率分布の平滑化の度合いを変更するパラメータであり，値が大きくなると平滑化の影響を強く受ける．そのため，各トピックの選択するノードにばらつきが出てしまい，トピック内でクラスタが生成されにくくなる，従ってクラスタが減少した分だけ最短経路が減少し平均経路が伸びる結果となったと考えられる．

また，トピック数 K を増やしていくと平均経路が長くなり，クラスタ係数が下がることも確認された (図 3)．トピック数が増加すると，目標ノードが一定であるため 1 つのトピックに属するノードの数が減少する．同一トピックに属するノード間ではクラスタが生成されやすいが，異なるトピックに属するノード間ではクラスタが生成されにくい，したがってトピック数 K が増えるとクラスタ係数が減少したと考えられる．平均経路が長くなったことは，β を大きくした時と同様と考えられる．

5. おわりに

本稿では，提案モデルを計算機上でシミュレーションを行い，提案モデルの挙動と指標の値を計算し，得られた値に関しての考察を行った．提案モデルでは，α, β, K を変更すること

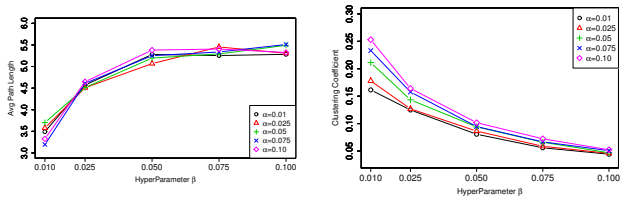

図 2: ハイパーパラメータを変更した時の平均経路長とクラスタ係数の変化

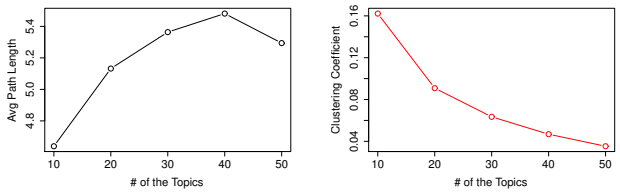

図 3: トピック数を変化させた時の平均経路長とクラスタ係数の変化

で複雑ネットワークの特徴である，スモールワールド性やクラスタ性に関する特徴を制御できた．今後は，エッジの生成がどのトピックを介するかということを中心に，より細かなネットワークの分析を行う．また，他の既存の複雑ネットワーク生成モデルと比較し評価実験を行う．

参考文献

[Albert 02] R, Albert., A.-L, Barabasi. ”Statistical mechanics of complex networks.” Reviews of modern physics 74.1 (2002): 47.

[Kwak 10] H, Kwak, et al. ”What is Twitter, a social network or a news media?.” Proc. WWW’10, 591-600

[Blei 03] D.-M, Blei., et al. ”Latent dirichlet allocation.” the Journal of machine Learning research 3 (2003): 993-1022.

[Barabasi 99] A.-L, Barabasi., R, Albert. ”Emergence of scaling in random networks.” science 286.5439 (1999): 509-512.

[Holme 02] P, Holme., B.-J, Kim. ”Growing scale-free networks with tunable clustering.” Physical review E 65.2 (2002): 026107.

[Bianconi 01] G, Bianconi., A.-L, Barabasi. ”Bose-Einstein condensation in complex networks.” Physical Review Letters 86.24 (2001): 5632.

[Kuramochi 13] 倉持 俊也，土方 嘉徳，西田 正吾：潜在嗜好トピックに基づくネットワーク生成モデルの提案．第 6 回 Web とデータベースに関するフォーラム (WebDB Forum) (2013)

[Henderson 09] K, Henderson., E.-R, Tina. ”Applying latent dirichlet allocation to group discovery in large graphs.” Proceedings of the 2009 ACM symposium on Applied Computing. ACM, 2009.