

【緒言】データのつながりや関係を表現する方法として、グラフが用いられる。しかし、グラフへの機械学習手法の直接的な適用は困難である。その中で、グラフをベクトル空間に埋め込む手法においては、ネットワーク上ランダムウォーク[1]が用いられている。また、素過程に変化を加えることにより、精度が向上することも知られている。以上のことから、ネットワークの変化に対するランダムウォーク変化の知見を得ることの重要性が示唆される。そこで本研究では、シンプルランダムウォークとノードの次数情報を考慮したランダムウォークの様々な素過程を考え、ランダムウォークがネットワークの変化に対してどの様に振る舞うかを調べた。

【方法】ネットワークデータとして、3種のモデル（Barabasi-Albertモデル、ランダムグラフ、2モデル混合グラフ）による生成データと現実系データを用いた。これらのモデル上で、3種類の素過程（シンプルランダムウォーク、次数比例、次数の逆数比例）に基づいたランダムウォークシミュレーションを行い、モデル変化（成長）に伴うランダムウォークの特徴量（探索数指数（図1）、遍歴ノード平均次数）を算出した。また、ランダムウォークの結果とネットワーク構造の関係を調べるために、ネットワークの特徴量（次数分布、次数相関）を算出し、比較を行った。さらに、コンフィグモデルを用いて各特徴量の変化の有無を調べた。

【結果】Barabasi-Albertモデル、ランダムグラフの両ネットワークにおいて、平均次数の上昇（成長）に対してランダムウォークの結果は変化しなかった。しかし、2モデル混合グラフでは特定の条件において、成長に伴いランダムウォークの結果が変化した。シンプルランダムウォークでは、2モデル混合グラフとそのコンフィグモデルで探索数指数は変化しなかった。しかし、ノード情報を考慮した素過程（次数比例（図2）、次数の逆数比例）では、探索数指数に変化が見られた。このことは、素過程の変化により、ネットワークのつながりを反映した探索が行われることを示唆している。また、次数相関関数による当てはめ指数の結果と探索数に関するべき乗則による当てはめ指数の結果は、同様の振る舞いを示した（図3）。これらの結果は、次数相関を反映した探索（ノードサンプリング）の実現を示している。

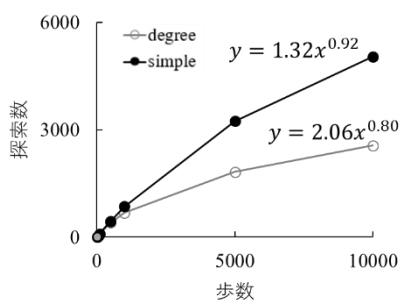


図1：歩数と探索数の推移

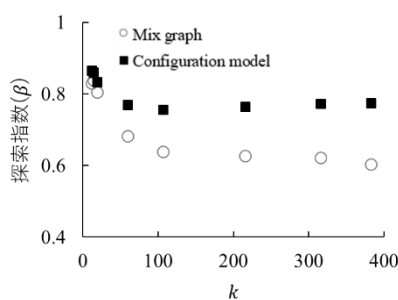


図2：探索数指数(次数比例)

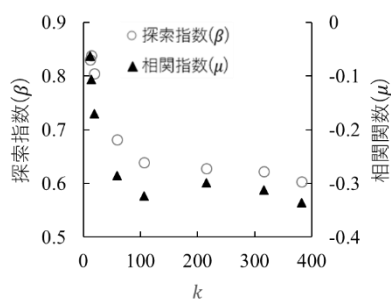


図3：探索数指数(次数比例)と相関指数

【参考文献】

[1] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 701-710).