敵対的攻撃下におけるランダムウォークの初回到着時間に関する一検討

A Study on the Hitting Time of a Random Walk under Adversarial Attacks

椎名 智 Satoshi Shiina　河村 宇記 Hiroki Kawamura　ハンネーアウン Han Nay Aung　大崎 博之 Hiroyuki Ohsaki

関西学院大学 大学院理工学研究科 情報科学専攻
Department of Informatics, Graduate School of Science and Technology, Kwansei Gakuin University

1 はじめに

未知のグラフ内で対象ノードを発見する手法や効率的な探索手法として、ランダムウォークに基づくアルゴリズムが広く使用されている。従来の研究では、グラフ上のランダムウォークアルゴリズムが敵対的攻撃 (リンクの張り替え) に対してどの程度堅牢であるかをシミュレーションによって調べているが [1]、これらの攻撃が実際に探索効率に与える影響の数理的解析は行われていない。

我々がこれまでに分析した敵対的リンクの張り替え攻撃手法は、ランダムウォークによるグラフ上のノードの探索や探査に大きな影響を与えることが明らかになっている。しかし、この攻撃手法が具体的にランダムウォークの探索や探査をどのように効果的に妨害するかについての特性はまだ十分に解明されていない。特に、リンクの張り替え攻撃が、探索効率、具体的にはランダムウォークの平均初回到着時間に与える影響についての詳細な理解が必要である。

本研究の目的は、敵対的攻撃の環境下において、ランダムウォークに基づくノード探索アルゴリズムの探索効率、特に平均初回到着時間がどのように低下するかを数理的に解析することである。具体的には、ランダムウォークとして単純ランダムウォーク (SRW: Simple Random Walk) を用い、攻撃手法として始点法を対象とし、これらの条件下での平均初回到着時間を解析的に導出することで、その特性を明らかにすることを目指す。

本研究では、敵対的攻撃者が、重みなし無向グラフ G 上でのエージェントのランダムウォークを妨害するシナリオを考える。攻撃者はエージェントの過去の訪問ノードを知ることができ、各ステップごとにレート λ でグラフ中のリンクを 1 本張り替えることができる。攻撃者の戦略として始点法を対象とし、エージェントが訪問しているノードに接続されているリンクの一つを、エージェントがランダムウォークを開始した始点ノードに接続するとする。この状況下におけるエージェントの初回到着時間を導出する。

2 解析

ランダムウォークの始点ノードを s、目的ノードを t とし、攻撃者がエージェントの現在位置に基づいてリンクを張り替えることで探索を妨害する。この攻撃下でのランダムウォークを再スタート付きランダムウォーク (RWR) として解析し、遷移確率行列の変形を通じて平均初回到着時間を導出する。

重みなし無向グラフ $G = (V, E)$, $V = \{1, 2, ..., N\}$ 上での単純ランダムウォークを考える。ランダムウォークの始点ノードを s、ランダムウォークの目的ノードを t と表記する。エージェントは、時刻 $k = 0$ に開始したランダムウォークによって目的ノード t を探索する。

攻撃者は各ステップごとにレート λ でエージェントに対して始点法による攻撃を行う。つまり、エージェントが現在訪問しているノード u に接続されているリンクのうち、ランダムに選択した 1 本のリンク (u, v) を、一時的にリンク (u, s) へと張り替える。

本解析では、始点法による攻撃下にあるランダムウォークの挙動を、再スタートのあるランダムウォーク RWR (Random Walk with Restart) によって近似する。

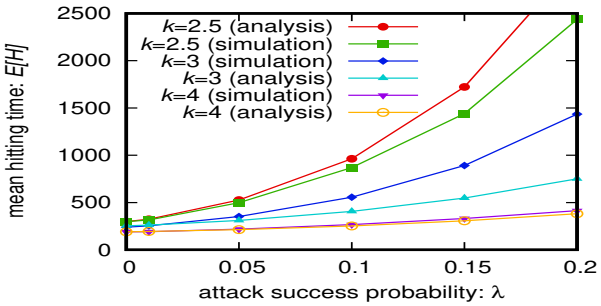

図 1　攻撃頻度 λ と平均初回到着時間 $E[H_{s,t}]$ の関係

グラフ G におけるエージェントの遷移確率行列を P とすると、始点法による攻撃下にあるランダムウォークの遷移確率行列 P_λ は

$$P_\lambda = (1 - \lambda)P + \lambda R$$

で与えられる。ここで R は $N \times N$ の行列であり、s 列の要素のみが 1 であり、それ以外の要素はすべて 0 である。

ノード s からノード t への平均初回到着時間は $\mathbf{h}_t$

$$\mathbf{h}_t = (I - Q_\lambda(t))^{-1}\mathbf{1}$$

の s 番目の要素によって与えられる。ここで、I は単位行列、$Q_\lambda(t)$ は P_λ からノード t を除外した部分グラフに対応する遷移確率行列、$\mathbf{1}$ は全要素が 1 のベクトルである。

3 数値例

以下では、いくつかの数値例により、攻撃レート λ がエージェントの平均初回到着時間 $E[H_{s,t}]$ に与える影響を分析する。

ER モデルによって生成したノード数 100 のランダムグラフを用いる。グラフの平均次数を 2.5、3、4 と変化させ、それぞれのグラフにおけるエージェントの平均初回到着時間を求める。解析の妥当性検証のため、同一のグラフにおけるランダムウォークのシミュレーションを実行し、その時の平均初回到着時間を計測した。

図 1 に、攻撃頻度 λ と平均初回到着時間 $E[H_{s,t}]$ の関係を示す。この結果から、攻撃レート λ が大きくなるにつれ、平均初回到着時間が急速に増大することがわかる。また、特に攻撃レート λ が小さい場合には、解析結果とシミュレーション結果が十分に一致していることもわかる。

謝辞

本研究の一部は JSPS 科研費 24K02936 の助成を受けたものである。

参考文献

[1] H. Kawamura, S. Shiina, H. N. Aung, and H. Ohsaki, "Robustness of random walk on a graph against adversary attacks," in *Proceedings of the IEEE Signature Conference on Computers, Software, and Applications (COMPSAC 2024)*, July 2024.