

A Dissertation for the Degree of Ph.D. in Engineering

**Fast Index-based Random Walks
on Dynamic Graphs
for Personalized Analysis**

February 2025

Graduate School of Science and Technology,
Keio University

Tsuyoshi Yamashita

報告番号	㊦ 乙 第	号	氏 名	山下 剛志
主 論 文 題 名 : Fast Index-based Random Walks on Dynamic Graphs for Personalized Analysis (動的グラフにおけるパーソナライズ解析のためのインデックスを用いた高速なランダムウォーク)				
(内容の要旨) 近年、多種多様なデジタルデータが指数的に増加している中、ユーザの興味や嗜好を柔軟に反映させたパーソナライズ解析としてグラフにおけるランダムウォーク (RW) が注目されている。 第1章ではランダムウォークを用いたパーソナライズ解析手法を述べる。RW はユーザの興味を表現する起点ノードと探索の範囲を制御する終了確率を入力とするグラフ演算であり、両パラメータを効果的に設定する必要がある。また、高速な演算のためには事前演算した RW の経路群を保持するインデックスの活用が重要である。 第2章では本博士論文における前提を整理した。具体的には、RW に基づくパーソナライズ解析手法である Personalized PageRank (PPR) を導入し、最先端の PPR の演算手法である FORA フレームワークを示す。そして、インデックスを用いた高速化手法である FORA+ を説明する。 第3章では本博士論文が取り組む3種類の課題について述べる。第4,5,6章においてそれぞれの課題に対する本博士論文の貢献を示す。 第4章では、PPR におけるパラメータの設定指針を確立する。PPR のパラメータである起点ノードの設定手法は数多く検討されてきた一方で、終了確率は盲目的に固定値に設定されてきた。そこで本博士論文では PPR 演算結果における全体重要性和起点近接性の影響を RW の平均経路長を制御する終了確率によって単調にバランスすることを検討する。映画評価データセットを用いたケーススタディでは、RW が短くなるように終了確率を設定することで起点ノードと直接関連が強いノードの評価値が単調に増加することが示された。また、統計的評価では、RW の平均経路長を 1.05 から 100 まで変化させることによって、PPR と全体重要性的ベクトルのコサイン類似度が最大で 0.003 から 0.76 まで単調に変化することが明らかになった。 第5章では、任意のパラメータに対する高速な RW 経路生成手法である α FlexWalk を提案する。高速な演算のためには事前実行された経路を保存するインデックスを参照することが重要であるが、インデックスを用いて出力された RW 経路の終了確率は事前実行時の値に制約される。そこで本博士論文ではインデックス内の経路を加工すること				

により任意の終了確率に対する RW 経路を高速に生成するアルゴリズムを提案する。具体的には、終了確率が変化すると RW の経路長が確率的に変化することに注目しながら、インデックス内の経路の連結・切断によって数学的に精度が保証された経路を生成する。評価では、任意の終了確率に対してインデックスを用いる提案手法はインデックスを用いない既存手法より最大で 10 倍高速であることが明らかになった。

第 6 章では、動的グラフにおける高精度な演算のための高速で軽量なインデックス管理手法を提案する。インデックス生成後にグラフが更新される場合、演算精度の保証のためには一部のインデックスを再生成することが必要であるが、その時間・空間計算コストが課題となっている。そこで本博士論文では、PPR 演算におけるインデックス参照は再生成後にあまりインデックスの内容が変化しないノードに集中することに注目して、グラフ更新時のインデックスの再生成を排除するインデックス管理手法を提案する。評価では、インデックス生成時からエッジ数が倍または半分になったとしても、演算精度の低下は 0.4% 以下であることが明らかになった。

最後に第 7 章では、本博士論文の結論を述べる。

Thesis Abstract

No. _____

Registration Number	<input checked="" type="checkbox"/> “KOU” <input type="checkbox"/> “OTSU” No. *Office use only	Name	Tsuyoshi Yamashita
Thesis Title Fast Index-based Random Walks on Dynamic Graphs for Personalized Analysis			
Thesis Summary <p>With the exponential growth of a wide variety of digital data in recent years, random walks on graphs have attracted attention as a personalized analysis that flexibly reflects user interests and preferences.</p> <p>Chapter 1 describes a personalized analysis method using random walks. Random walk is a graph computation whose input is a source node that represents the user’s interest and a termination probability that controls the exploration range. Both parameters need to be set effectively. In addition, it is important to utilize an index that maintains a set of pre-performed random walk paths for fast computation.</p> <p>Chapter 2 organizes the preliminaries of this dissertation. In particular, it introduces Personalized PageRank (PPR), a personalized analysis method based on random walks, and presents the FORA framework, a state-of-the-art PPR computation method. Then, it explains FORA+, an index-based acceleration method.</p> <p>Chapter 3 describes the three challenges addressed by this dissertation. Chapter 4, Chapter 5, and Chapter 6 present the dissertation’s contribution to each of these challenges.</p> <p>Chapter 4 establishes parameter-setting guidelines for PPR. Although many methods for setting the source node have been discussed, the termination probability has been blindly set to a fixed value. This chapter investigates how to monotonically balance the influence of global importance and source proximity on PPR results with termination probability, which controls the average path length of the random walk. A case study using a movie rating dataset shows that setting termination probability to shorten random walk monotonically increases the rating of nodes that are directly related to the source node. Statistical evaluation also revealed that varying the average path length of the random walk from 1.05 to 100 resulted in a monotonic change in the cosine similarity between PPR and the global importance vector from 0.003 to 0.76 at the maximum.</p> <p>Chapter 5 proposes a fast random walk path generation method for arbitrary parameters. For fast computation, it is important to reference an index, but the accepted termination probability is limited to the value at the time of index generation. As a result, to generate paths for arbitrary termination probabilities, it was necessary to perform random walks without the index. This dissertation proposes an algorithm, αFlexWalk, to quickly generate random walk paths for arbitrary termination probabilities by manipulating the paths in the index. In particular, αFlexWalk focuses on the fact that the random walk path length varies stochastically as the termination probability changes. The algorithm generates mathematically guaranteed paths by connecting and cutting the indexed paths. Evaluations showed</p>			

that α FlexWalk was up to 11.2 times faster than existing index-free methods.

Chapter 6 proposes a fast and lightweight index management method for high-accuracy computation on dynamic graphs. When a graph is updated after index generation, it is necessary to re-generate some indexes to guarantee accuracy, but the time and space computation cost is significant. This dissertation proposes an index management method that eliminates index re-generation for graph updates, focusing on the fact that index references in PPR computation concentrate on nodes whose index does not change much after re-generation. The evaluation revealed that even if the number of edges is doubled or halved from the time of index generation, the loss of accuracy is less than 0.3%.

Finally, Chapter 7 concludes this dissertation.