

グラフ生成モデルにおける 入力情報が構造的特徴量の 制御精度に与える影響の検証

佐藤 天樹¹ 津川 翔² 眞田 亜紀子¹ 渡部 康平³

¹長岡技術科学大学 大学院工学研究科

²筑波大学 システム情報系

³埼玉大学 大学院理工学研究科

Contents

- 1.研究背景・目的
- 2.関連研究
- 3.提案手法
- 4.実験
- 5.結果
- 6.まとめ・今後の展望

- 1.研究背景・目的
- 2.関連研究
- 3.提案手法
- 4.実験
- 5.結果
- 6.まとめ・今後の展望

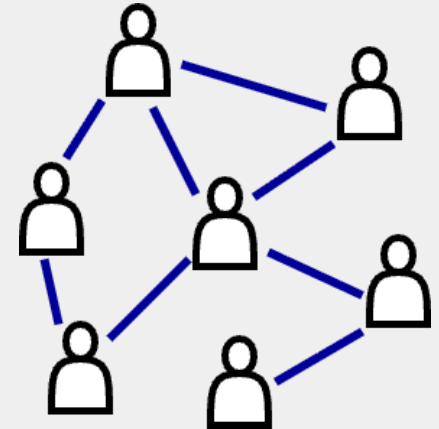
研究背景 - 1

◆グラフ構造はノード(接続点)とエッジ(繋がり)で表現される

◆多くの**関係性を表現可能**である

(例：人間関係, 交通網, 通信ネットワーク)

◆応用分野でのグラフ構造の利用は増加している

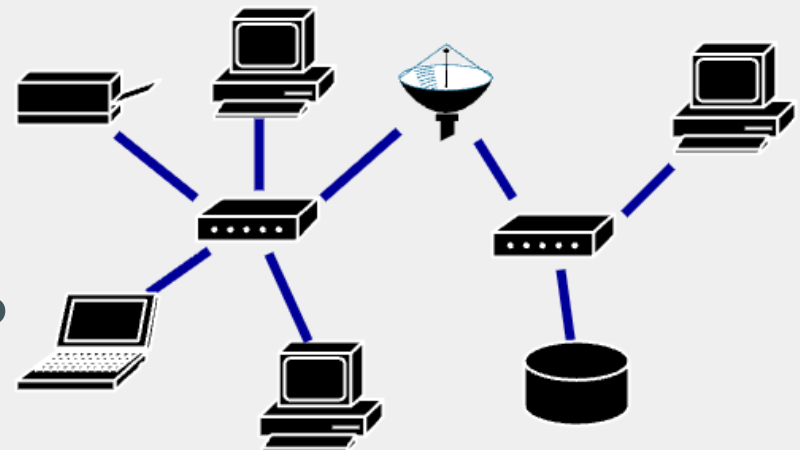


◆グラフ構造によるシミュレーションでは、

現実世界のグラフ構造の特徴を捉えた

グラフを用意する必要がある

◆セキュリティやプライバシー、
計測データの不足などにより
現実世界の環境から直接再現される
グラフ(実グラフ)の数は限られている



研究背景 - 2

◆ グラフ生成モデル

◆ 統計的手法を用いたモデル

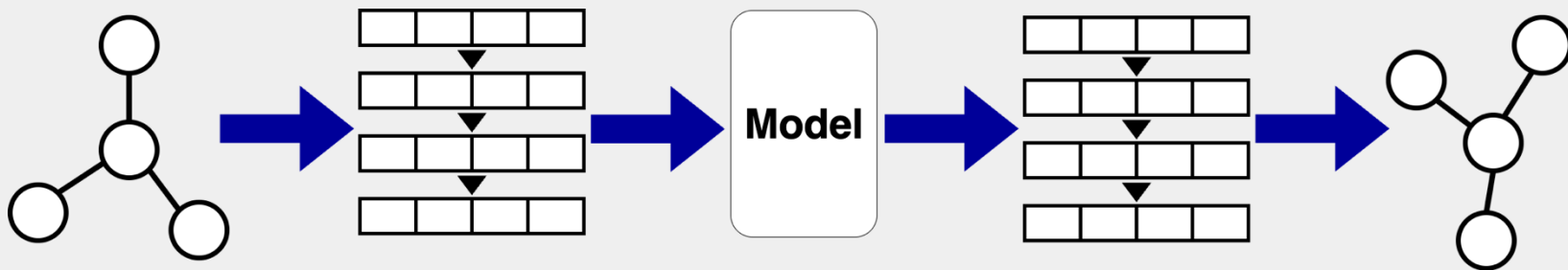
- ◆ 実グラフが持つ特徴の一部を再現可能
- ◆ 複数の特徴を同時に再現することは困難

◆ 機械学習を用いたモデル

- ◆ 実グラフが持つ特徴の多くを同時に再現可能
- ◆ 特徴の正確な制御が困難

◆ 先行研究：GraphTune

- ◆ グラフをシーケンスとして扱う
- ◆ 任意の構造的特徴量を制御することが可能

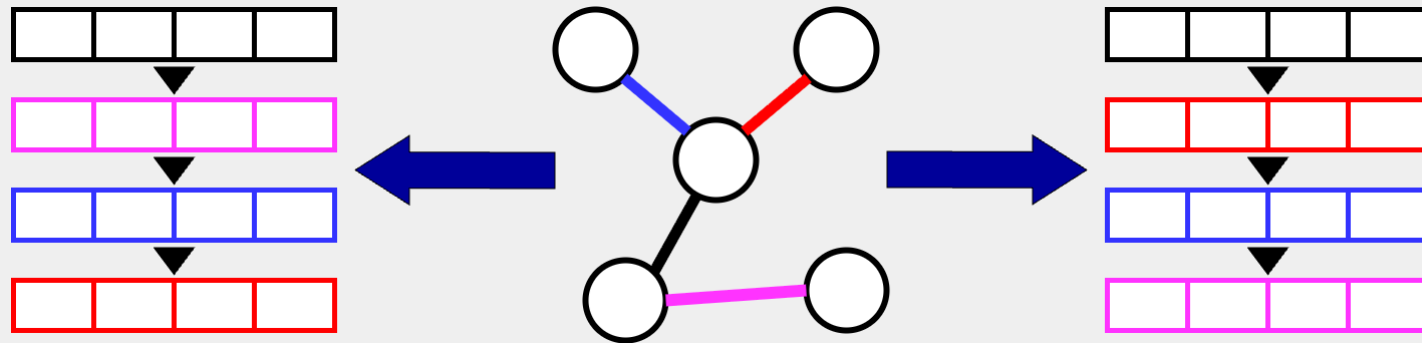


研究目的

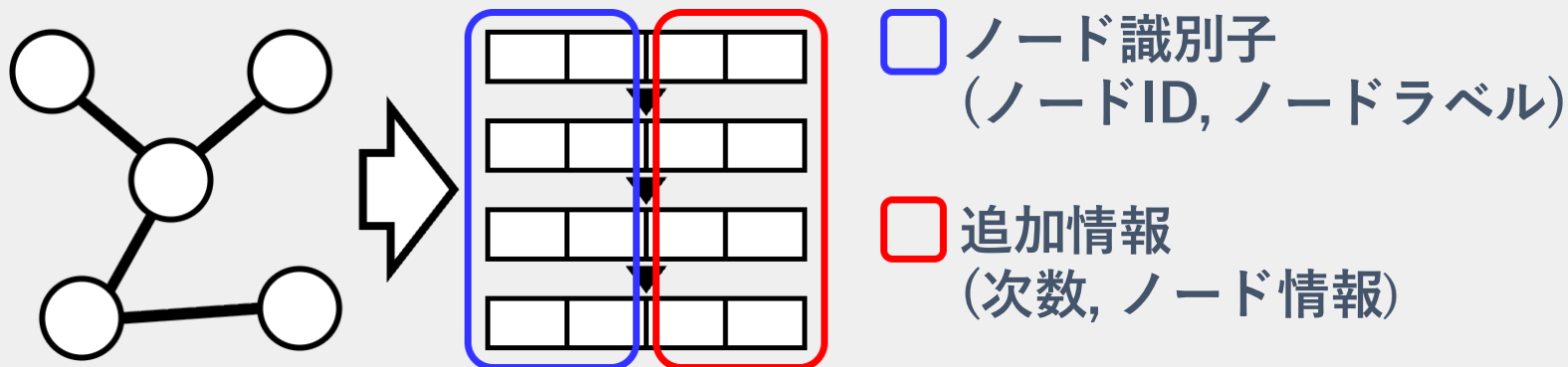
◆シーケンスが生成されるグラフ構造に影響を与える

◆最適な順序を求める研究は存在

ランダムウォーク, 深さ優先探索(DFS), 幅優先探索(BFS) など



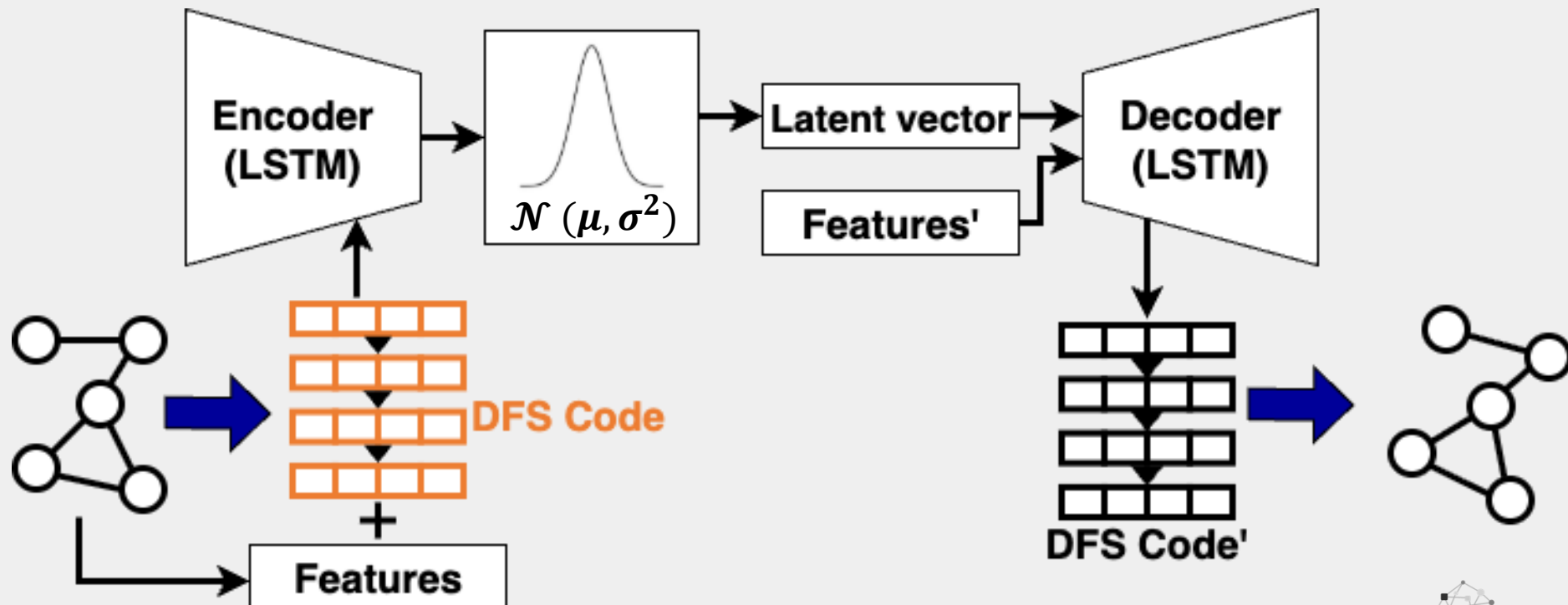
◆順序以外の情報(ノード情報, 時系列的信息)の影響を検証する



- 1.研究背景・目的
- 2.関連研究**
- 3.提案手法
- 4.実験
- 5.結果
- 6.まとめ・今後の展望

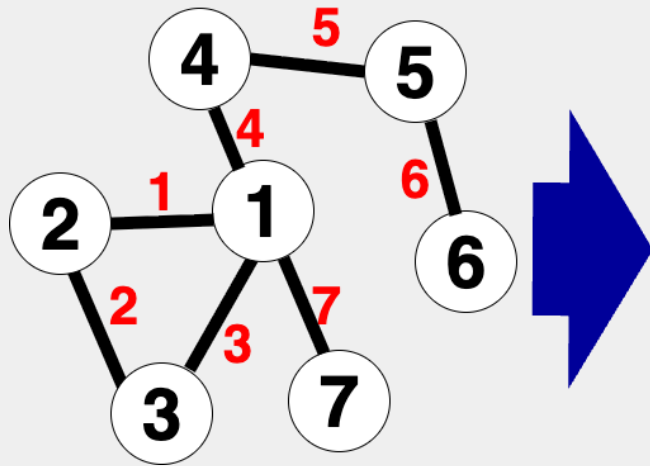
関連研究：GraphTune [1]

- ◆ グラフを **DFS Code** と呼ばれるシーケンスに変換し、LSTMによってシーケンスを学習する
- ◆ グラフの構造的特徴量を入力することで、生成されるグラフの特徴量を制御可能



DFS Code

- ◆ グラフからシーケンスへの変換手法の一つ
- ◆ グラフをDFSで探索した履歴を表現
- ◆ 高次数のノードを優先して全エッジを探索



ノードには発見順を、
エッジには探索順序を
表示している

探索 順序	現在の ノードの タイム スタンプ	次の ノードの タイム スタンプ	現在の ノード ラベル (次数)	次の ノード ラベル (次数)	エッジ の ラベル
1	1	2	4	3	0
2	2	3	3	2	0
3	3	4	2	1	0
4	2	5	3	1	0
5	1	6	4	2	0
6	6	7	2	2	0
7	7	1	2	4	0

- 1.研究背景・目的
- 2.関連研究
- 3.提案手法**
- 4.実験
- 5.結果
- 6.まとめ・今後の展望

シーケンスへの要素の追加 - 1

- ◆ GraphTune で使用している **DFS Code** および、
順序を幅優先探索(BFS) に変更した **BFS Code**
に対して要素の追加を行う
- ◆ グラフの構造的特徴に基づく情報
 - ◆ **固有ベクトル中心性指標**
 - ◆ グラフ内におけるノードの重要性を示す
 - ◆ シーケンスが持つノード情報の追加
- ◆ 探索に基づく時系列的な情報
 - ◆ **カウントダウン**
 - ◆ 局所的な残りのステップ数を表す
 - ◆ モデルが捉えにくい値の急激な変化のタイミングを示唆

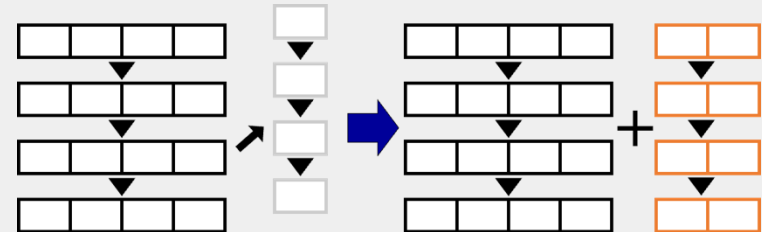
シーケンスへの要素の追加 - 2

◆DFS Code は、以下の5要素を基本としている

- ◆始点・終点のノードのタイムスタンプ
- ◆始点・終点のノードの次数
- ◆エッジのラベル（GraphTune では 0 で統一）

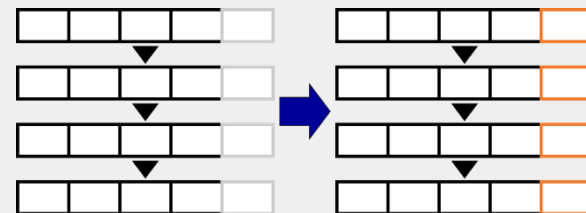
◆固有ベクトル中心性

- ◆エッジのラベルを除去し、始点・終点のノードの固有ベクトル中心性のランク（**EVC Rank**）を導入
- ◆5 要素 $\Rightarrow$ 6 要素 に



◆カウントダウン

- ◆エッジのラベルとして**カウントダウンラベル**を導入
- ◆エッジに関する情報を追加



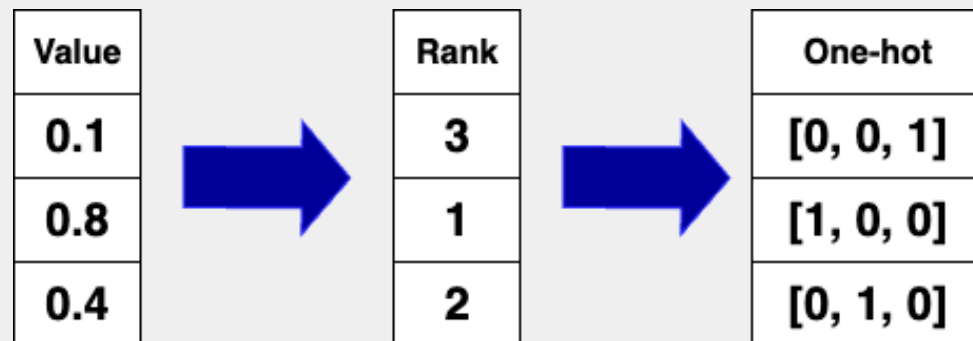
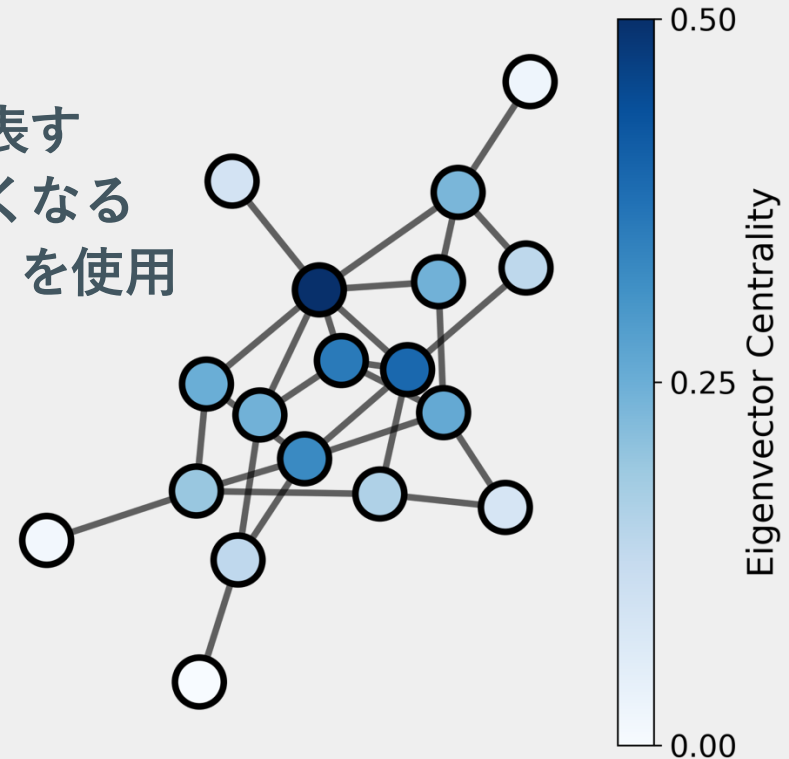
固有ベクトル中心性

◆固有ベクトル中心性

- ◆グラフ内におけるノードの重要性を表す
- ◆重要なノードと繋がるほど値は大きくなる
- ◆ランキング化したもの (**EVC Rank**) を使用

◆ランキング化

- ◆GraphTune では
One-hot エンコーディングを採用
 - ◆数値を 0, 1 で表現する手法の一つ
 - ◆連続値は表現不可
- ◆値の大小関係のみを保持



DFS におけるカウントダウン

◆DFSでは、可能な限り長い経路を辿ろうとする

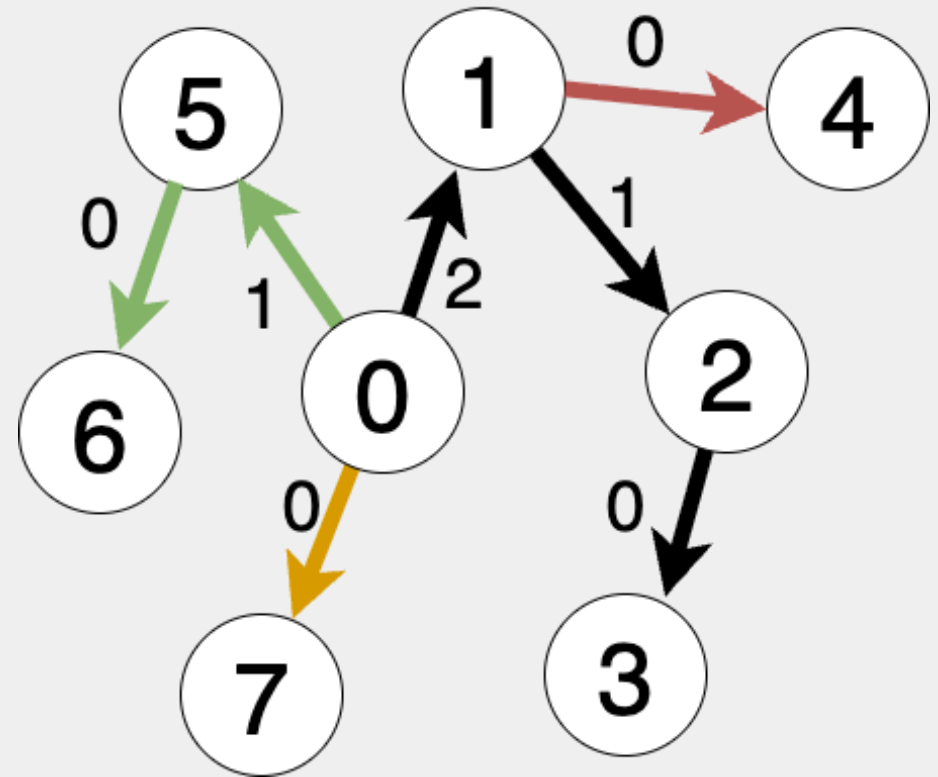
◆現在の終点が次の始点になる

⇒ 1本の経路を進んでいる

◆現在の終点と次の始点が変わる

⇒ 経路の終点に到達

◆終点までの残りのステップ数を明示する



BFS におけるカウントダウン

◆BFSでは、近い場所から順に幅広く探索を行う

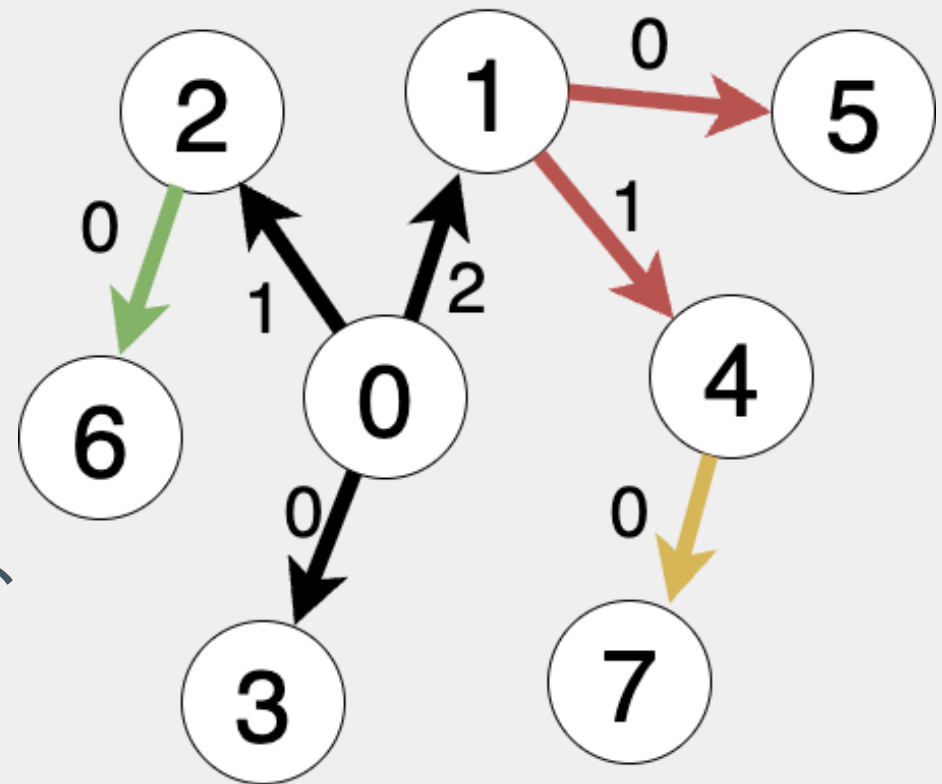
◆現在の始点と次の始点が同一

⇒ 近隣に未探索のエッジが存在

◆現在の始点と次の始点異なる

⇒ 当該ノードの探索が終了

◆ノードに隣接するエッジのうち、
未発見のエッジの本数を明示



- 1.研究背景・目的
- 2.関連研究
- 3.提案手法
- 4.実験**
- 5.結果
- 6.まとめ・今後の展望

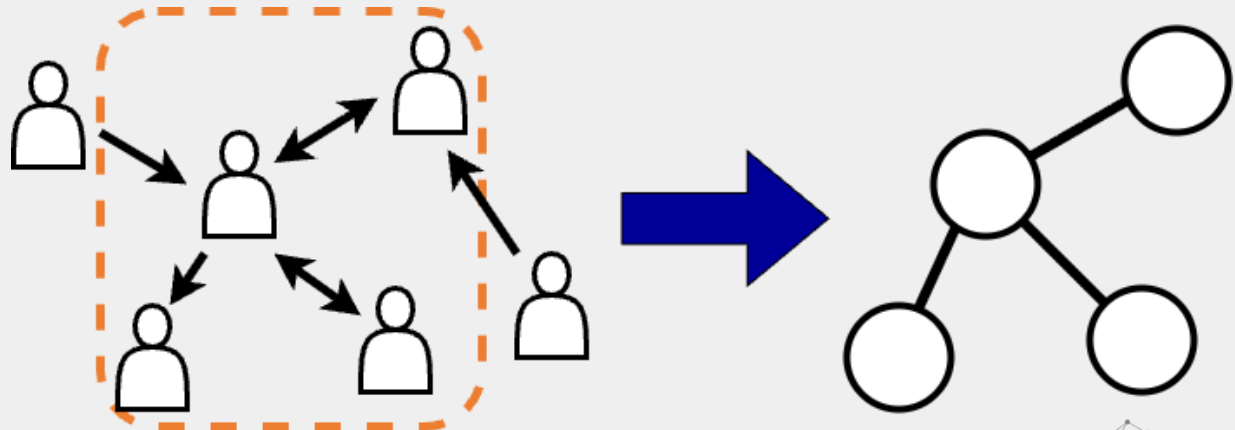
データセット

◆元データ：Higgs Twitter Dataset [2]

- ◆ノード：ユーザー
- ◆エッジ：フォロワー・フォロワーの関係
- ◆ノード数：456,626
- ◆エッジ数：14,855,842

◆GraphTune に従い、元データからランダムウォークでサンプリングした部分グラフを使用

- ◆グラフ数：2,000
- ◆ノード数：50
- ◆無向
- ◆属性・重みなし



実験設定 - 1

- ◆ Twitterのデータセットを使用してモデルの学習を行い、
複数の条件により特徴量の制御精度を比較する
- ◆ 比較を行う条件
 - ◆ 探索手法
 - ◆ DFS, BFS
 - ◆ シーケンスに付加する情報
 - ◆ なし(ベースライン), **EVC Rank**, **カウントダウン**
 - ◆ 制御する特徴量と指定値
 - ◆ 平均最短経路長(APL) : [3.00, 4.00, 5.00]
 - ◆ クラスタ係数(CC) : [0.10, 0.20, 0.30]
 - ◆ モジュール性(MQ) : [0.40, 0.55, 0.70]
 - ◆ 平均次数(AD) : [3.00, 3.50, 4.00]
 - ◆ べき乗指数(PLE) : [2.60, 3.00, 3.40]

実験設定 - 2

◆データセットの配分

◆学習用：1800, 検証用：200

◆モデルパラメータ

◆GraphTune と同様

◆評価指標

◆生成したグラフが持つ特徴量の値と、
実際の指定値の間の二乗平均平方根誤差(RMSE)を使用

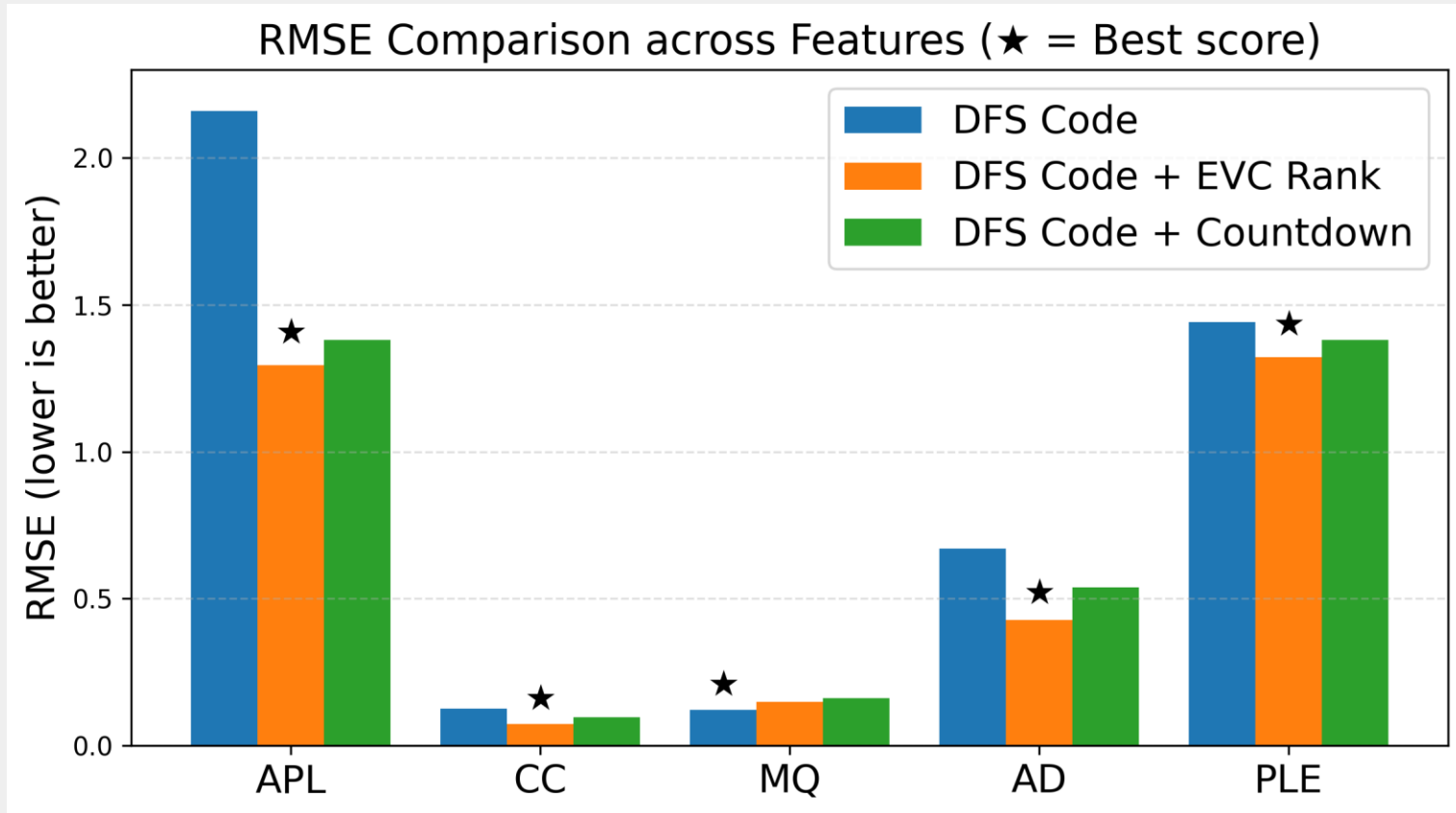
◆結果表示では、**各特徴量の3条件を平均して可視化**

◆各条件で、ランダム要素となるSeed 値を変更し、10回ずつ実施

◆RMSEの平均と分散を記録

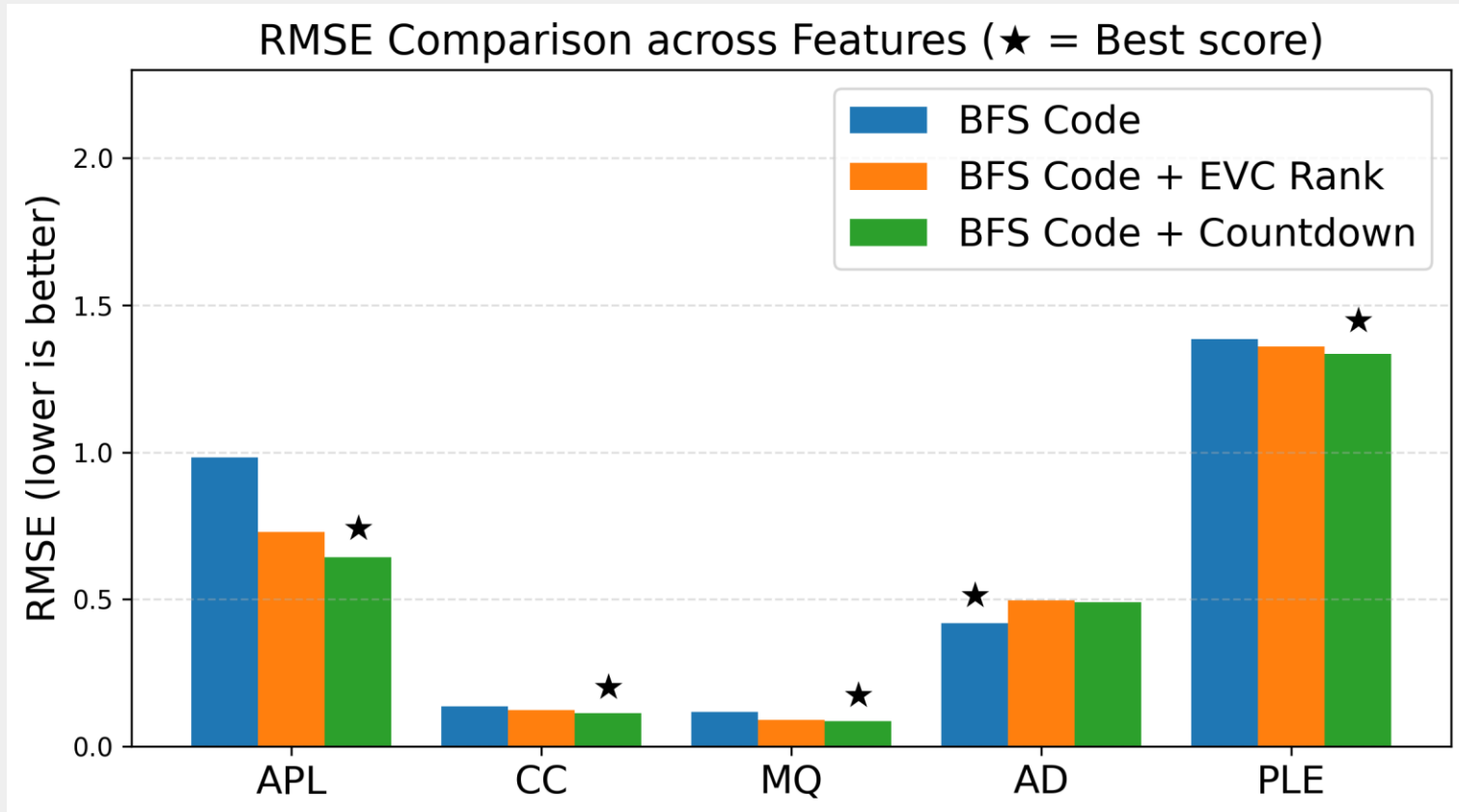
- 1.研究背景・目的
- 2.関連研究
- 3.提案手法
- 4.実験
- 5.結果**
- 6.まとめ・今後の展望

実験結果：DFS



- ◆EVC Rank を導入した手法が、4種類の特徴量で最良の結果を記録
- ◆カウントダウンでも、従来手法より小さなRMSE を記録

実験結果：BFS



◆ カウントダウンを導入した手法が、4条件で最良の結果を記録

◆ EVC Rank を導入した手法も、従来手法より小さなRMSEを記録

- 1.研究背景・目的
- 2.関連研究
- 3.提案手法
- 4.実験
- 5.結果
- 6.まとめ・今後の展望**

まとめ・今後の展望

◆まとめ

- ◆GraphTune におけるグラフからシーケンスへの変換手法である DFS Code、および BFS Code に対し、EVC Rankや カウントダウンラベルといった、追加情報を導入したシーケンスで 特徴量の制御精度の比較を行った
- ◆RMSEによる定量的評価から、提案手法を使用することで 従来手法を上回る精度で特徴量を制御可能であることを示した

◆今後の展望

- ◆次数情報と今回導入した情報との有効性比較
- ◆グラフが持つ他の情報を使用したより広範な実験
- ◆GraphTune 以外のシーケンスベースのモデルでの検証

参考文献

- ◆[1] K. Watabe, S. Nakazawa, Y. Sato, S. Tsugawa, and K. Nakagawa, “Graphtune: A learning-based graph generative model with tunable structural features,” *IEEE Transactions on Network Science and Engineering*, vol.10, no.4, pp.2226–2238, 2023.
- ◆[2] M. D. Domenico, A. Lima, P. Mougél, and M. Musolesi, “The Anatomy of a Scientific Rumor,” *Scientific Reports*, vol. 3, no. 2980, 2013.

補足：グラフの構造的特徴量

- ◆ グラフの構造的特徴を表す特徴量は複数存在する
- ◆ 本研究では、以下の特徴量を使用する
 - ◆ 平均最短経路長 (APL)
 - ◆ すべての2ノード間の最短経路長の平均値
 - ◆ クラスタ係数 (CC)
 - ◆ 3つのノードが三角形を形成する割合のグラフ全体における平均
 - ◆ モジュール性 (MQ)
 - ◆ グラフを複数のコミュニティに分割したとき、コミュニティ内でのまとまりの強さ
 - ◆ 平均次数 (AD)
 - ◆ あるノードが何本のエッジを持つかの平均値
 - ◆ べき乗指数 (PLE)
 - ◆ 次数分布において、両軸を対数とした場合の直線の傾き

補足：RMSEの算出方法

◆特徴量の指定値と、生成されたグラフの特徴量とのRMSE

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (c^* - c_i)^2}$$

n : 生成されたグラフの個数

c^* : 特徴量の指定値

c_i : 生成グラフの特徴量

補足：実験結果(DFS)

◆最良の結果を太字、次点の結果を下線で示す

Features	Specified Value	Normal DFS Code	DFS Code + EVC Rank	DFS Code + Countdown
APL	3.0	1.343 ± 0.902	0.520 ± 0.122	<u>0.674 ± 0.179</u>
	4.0	2.102 ± 1.054	1.332 ± 0.179	<u>1.355 ± 0.157</u>
	5.0	3.034 ± 1.220	2.034 ± 0.193	<u>2.114 ± 0.225</u>
CC	0.10	0.108 ± 0.050	0.063 ± 0.023	<u>0.081 ± 0.017</u>
	0.20	0.135 ± 0.034	0.072 ± 0.012	<u>0.102 ± 0.027</u>
	0.30	0.133 ± 0.030	0.087 ± 0.013	<u>0.105 ± 0.014</u>
MQ	0.40	0.153 ± 0.045	<u>0.099 ± 0.031</u>	0.083 ± 0.017
	0.55	0.119 ± 0.012	<u>0.135 ± 0.038</u>	0.149 ± 0.031
	0.70	0.092 ± 0.048	<u>0.211 ± 0.059</u>	0.255 ± 0.044
AD	3.0	<u>0.513 ± 0.094</u>	0.337 ± 0.085	0.514 ± 0.171
	3.5	0.725 ± 0.182	0.441 ± 0.099	<u>0.601 ± 0.131</u>
	4.0	0.775 ± 0.121	<u>0.504 ± 0.159</u>	0.499 ± 0.243
PLE	2.6	1.036 ± 0.108	0.965 ± 0.060	<u>1.033 ± 0.055</u>
	3.0	1.456 ± 0.123	1.316 ± 0.065	<u>1.364 ± 0.046</u>
	3.4	1.828 ± 0.153	1.681 ± 0.068	<u>1.744 ± 0.052</u>

補足：実験結果(BFS)

◆最良の結果を太字、次点の結果を下線で示す

Features	Specified Value	Normal BFS Code	BFS Code + EVC Rank	BFS Code + Countdown
APL	3.0	0.152 ± 0.043	0.568 ± 0.188	<u>0.321 ± 0.179</u>
	4.0	0.973 ± 0.065	0.512 ± 0.151	<u>0.594 ± 0.166</u>
	5.0	1.824 ± 0.220	<u>1.107 ± 0.373</u>	1.011 ± 0.373
CC	0.10	0.071 ± 0.006	<u>0.070 ± 0.014</u>	0.058 ± 0.013
	0.20	0.139 ± 0.016	<u>0.122 ± 0.022</u>	0.119 ± 0.018
	0.30	0.198 ± 0.048	<u>0.178 ± 0.022</u>	0.160 ± 0.051
MQ	0.40	0.057 ± 0.020	0.044 ± 0.014	<u>0.053 ± 0.023</u>
	0.55	0.108 ± 0.050	0.059 ± 0.018	<u>0.074 ± 0.038</u>
	0.70	0.186 ± 0.034	<u>0.164 ± 0.033</u>	0.132 ± 0.043
AD	3.0	0.371 ± 0.218	0.336 ± 0.075	<u>0.361 ± 0.127</u>
	3.5	0.383 ± 0.126	0.456 ± 0.102	<u>0.401 ± 0.087</u>
	4.0	0.499 ± 0.153	<u>0.697 ± 0.211</u>	0.708 ± 0.135
PLE	2.6	0.994 ± 0.056	<u>0.970 ± 0.033</u>	0.935 ± 0.040
	3.0	1.382 ± 0.056	<u>1.356 ± 0.035</u>	1.346 ± 0.036
	3.4	1.780 ± 0.052	<u>1.750 ± 0.034</u>	1.721 ± 0.029