

# LSTMと条件付きVAEを用いた 構造的特徴を高精度に指定可能な グラフ生成モデル

†長岡技術科学大学 大学院工学研究科

††筑波大学 システム情報系

横山昂大† 佐藤良紀† 津川翔†† 渡部康平†

# 目次

1. 目的・背景
2. 先行研究
3. 提案法
4. 実験
5. まとめ・今後の展望

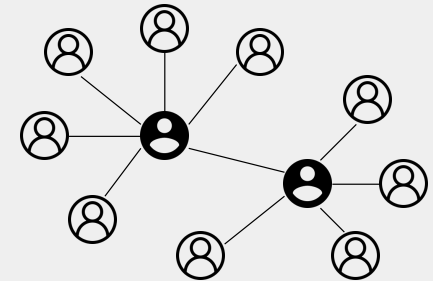
# 目的・背景

# 目的・背景

◆近年、グラフを用いるアプリケーションの需要は増加している

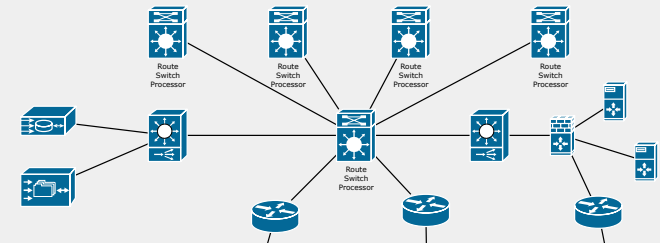
◆グラフとして表現できる例

- ◆ソーシャルネットワーク：アカウントの繋がり
- ◆通信ネットワーク：通信機器の接続関係



◆グラフを用いたアプリケーションの例

- ◆通信プロトコルのシミュレーション
- ◆コミュニティ予測



◆グラフ上でのシミュレーションでは、**現実の環境の特徴を備えたグラフ**が必要となる

- ◆プライバシーなどの観点から、これを用意するのは難しい

# 目的・背景

◆ **グラフ生成モデル**は実世界のグラフデータを補完するアプローチ

◆ 既存のグラフ生成モデルにおける問題

◆ 統計的手法を用いたグラフ生成モデル

➢ 特定の実在するグラフの特徴すべてを捉えた生成は難しい

◆ 機械学習を用いたグラフ生成モデル

➢ 特徴の指定に関する生成手法の研究は非常に少ない

➢ 特徴を指定した生成の精度には課題がある

◆ 目的

実在するデータを学習し、**任意の特徴量をより高精度に指定することが可能なグラフ生成モデルの構築**

# 目的・背景

## ◆アプローチ

- ◆先行研究であるGraphTuneモデルを拡張し、グラフからその特徴値を推定するFeature estimatorを追加して、交互学習アルゴリズムを適用することによって特徴の指定を高精度に行う

## ◆Feature estimator

- ◆グラフから特定の特徴値を推定する
- ◆特徴値の再現に関するフィードバックをGraphTuneに与えられる

## ◆交互学習アルゴリズム

- ◆GraphTuneとFeature estimatorを一方の重みを凍結させて他方を学習する作業を交互に繰り返す
- ◆特徴値を含む条件ベクトルの情報がリークすることを防ぐ

# ネットワークの 代表的な特徴

# ネットワークの代表的な特徴

◆現実のネットワークが持つ特徴は多く存在し、ネットワーク毎にさまざまな特徴を有している

## ◆スケールフリー性

- ◆ごく少数のノードが持つリンクの数(次数)が大きく、大多数のノードの次数は小さいという性質
- ◆特徴量：次数分布のべき指数

## ◆クラスタ性

- ◆ある人の知人同士が知り合いというように、ネットワークが三角形を取る割合が高いという性質
- ◆特徴量：平均クラスタ係数

## ◆スモールワールド性

- ◆ネットワークの平均ノード間距離が小さいという性質
- ◆特徴量：平均最短経路長



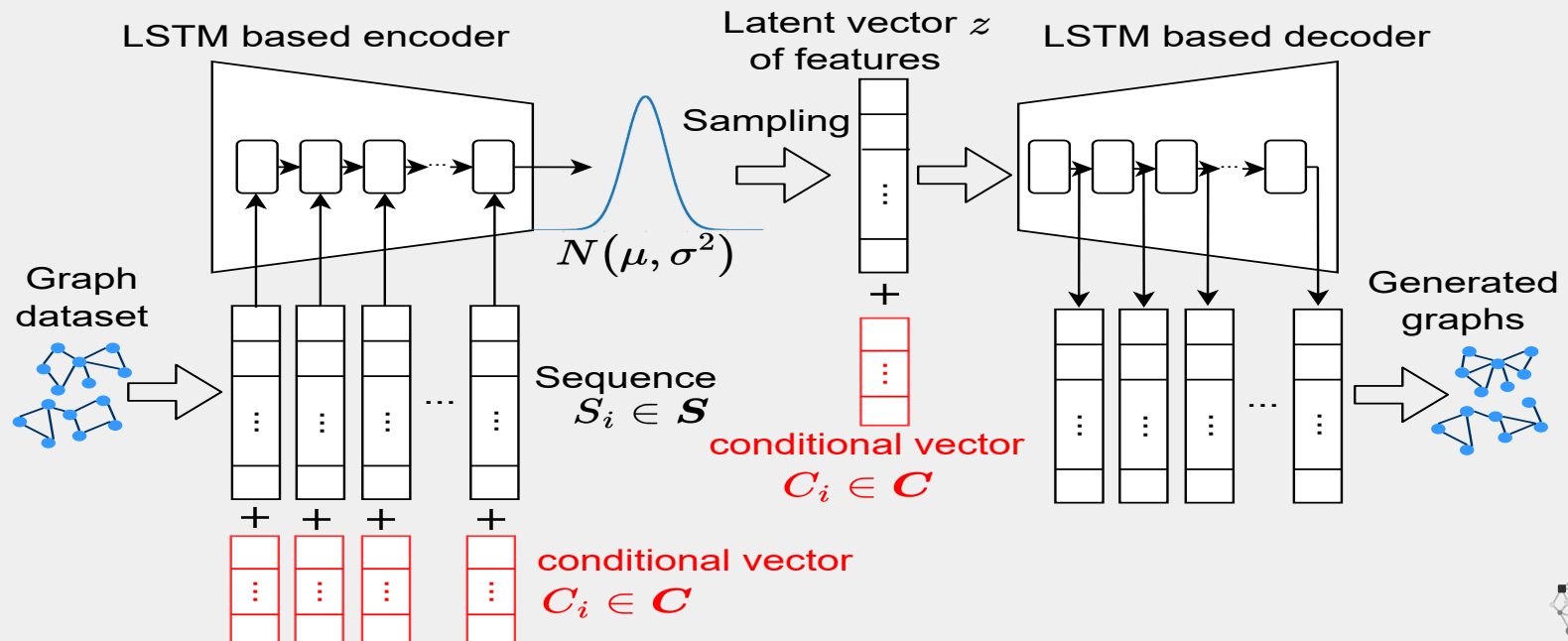
# 先行研究

# GraphTune [3][4]

## ◆ Conditional VAE(CVAE)とLong Short Term Memory(LSTM)を組み合わせた任意の特徴量を指定可能な生成モデル

### ◆ 前処理

- ◆ グラフをDFSコードによってシーケンスへ変換し、グラフから計算される特徴量を条件ベクトルとして連結させる
- ◆ この条件ベクトルを与えることにより、特徴の学習が可能となる

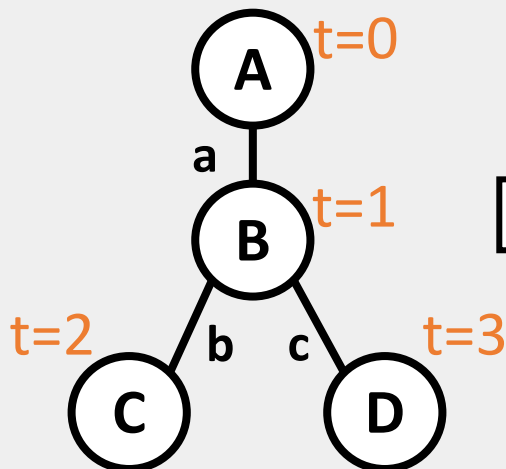


# DFSコード (Depth First Search Code)

## ◆ グラフをシーケンスにする手法

1. グラフに**深さ優先探索**を行い探索順に各ノードにタイムスタンプを付与する
2. タイムスタンプ情報とノードラベル、ノード間のエッジラベルを基に**5要素のベクトル**に変換する
3. 各ノードを探索順毎にベクトルに変換することでグラフをシーケンスなデータに変換する

### グラフを深さ優先探索

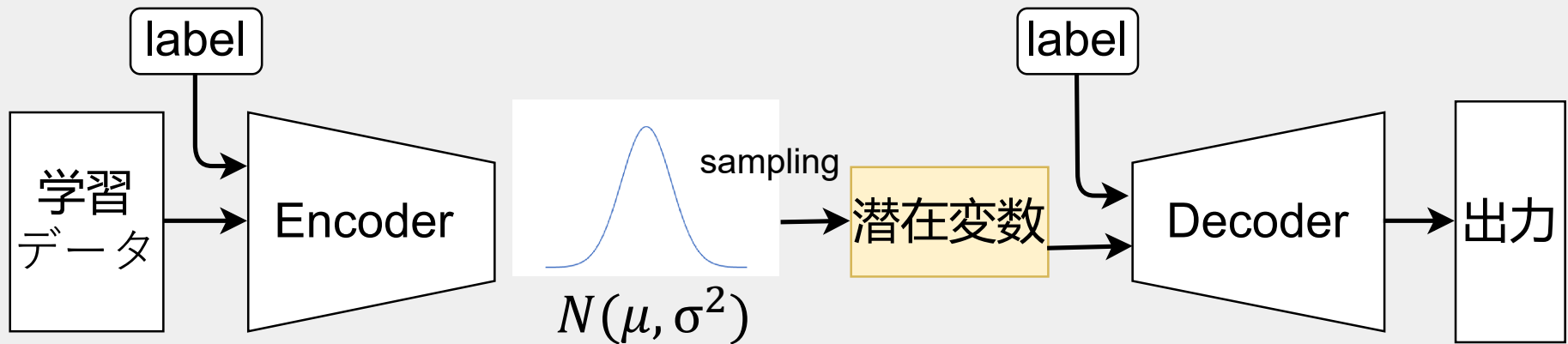


### シーケンスデータへ

Time	Time stamp u	Time stamp v	Node label u	Node label v	Edge label
0	0	1	A	B	a
1	1	2	B	C	b
2	1	3	B	D	c

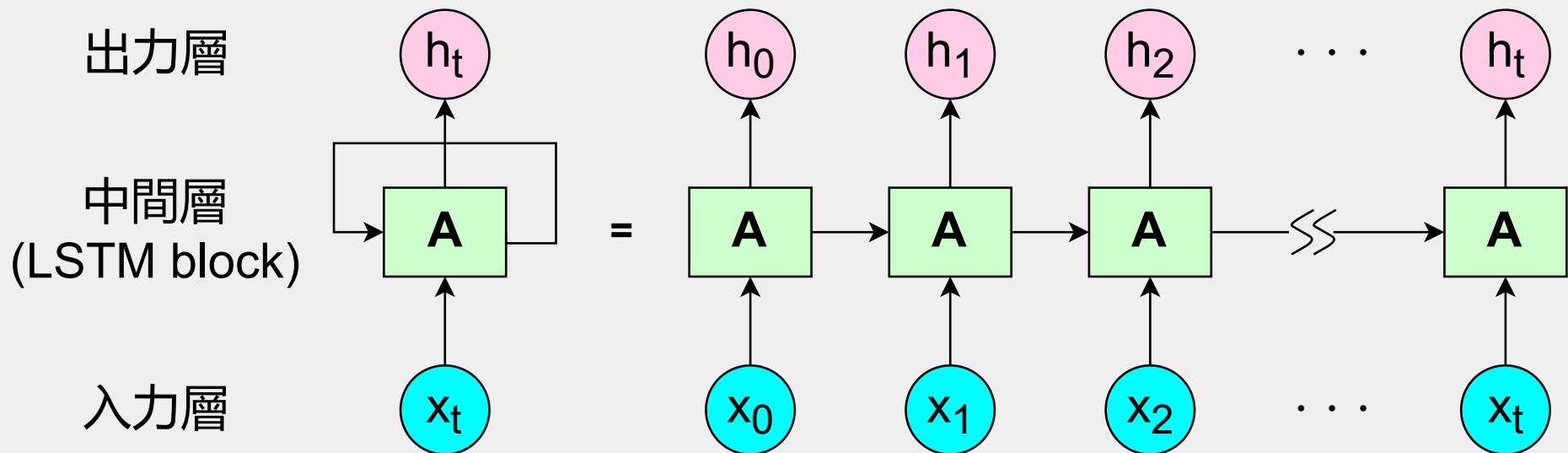
# Conditional VAE[1]

- ◆ EncoderとDecoderの2つのニューラルネットワークからなる条件付き生成モデル
- ◆ Encoder
  - ◆ 学習データとlabelから潜在変数が正規分布に従うように学習
- ◆ Decoder
  - ◆ 潜在変数とlabelの情報を基に学習データを復元するように学習



# Long Short Term Memory [2]

- ◆ 長期的な依存関係を学習することが可能なモデルで、シーケンスデータの学習に用いられる
- ◆ シーケンスデータを固定サイズのベクトルにすることができる
- ◆ 前の時系列の出力を再帰的に入力することで生成に利用できる



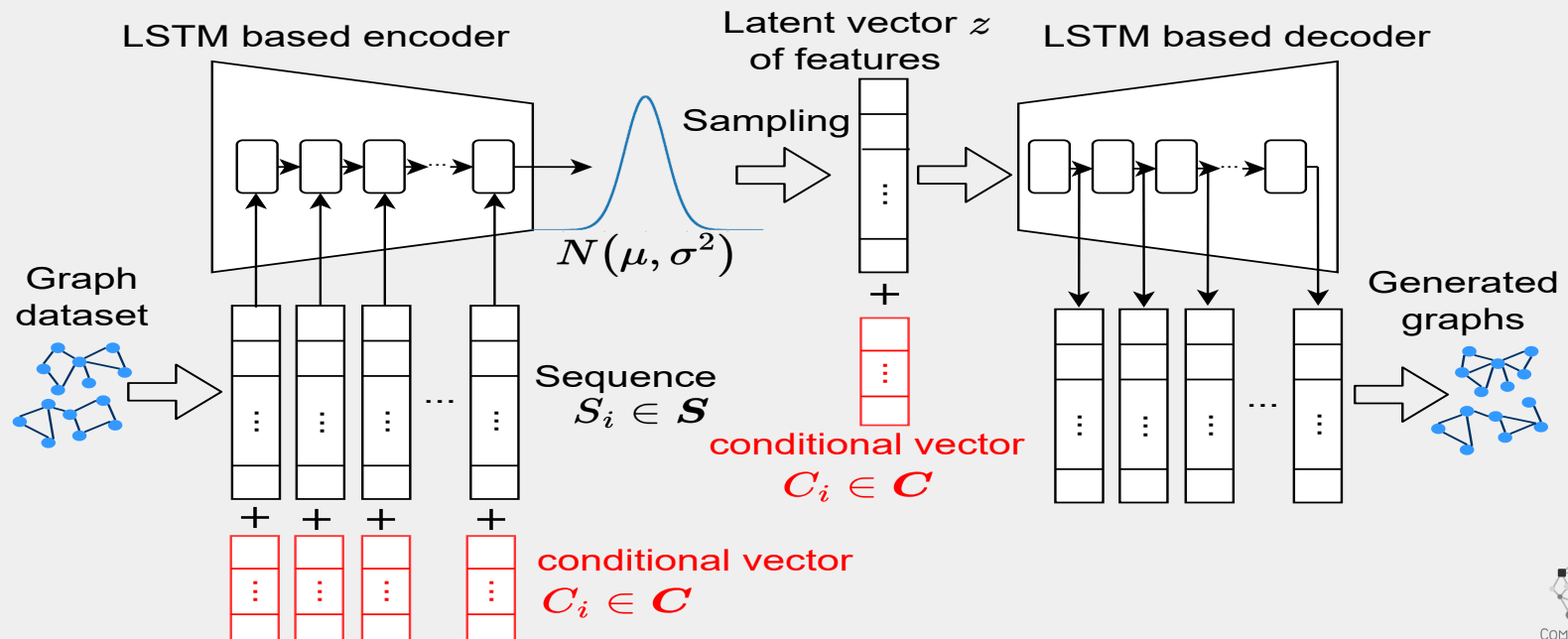
# GraphTune [3][4]

## ◆学習

- ◆Encoderはシーケンスから潜在変数を正規分布となるように学習する
- ◆Decoderは条件ベクトルとサンプリングした潜在変数から入力シーケンスを復元するように学習する

## ◆生成

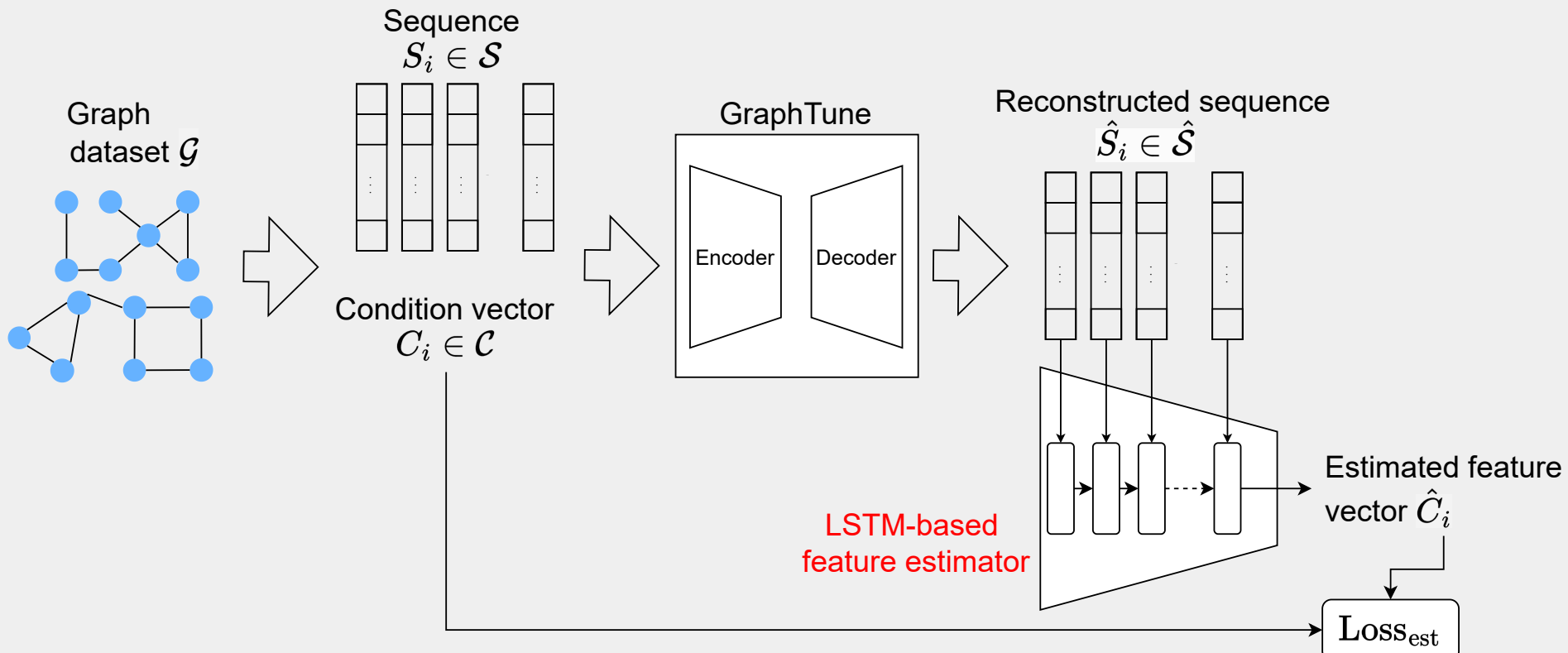
- ◆指定したい特徴値の条件ベクトルと正規分布からサンプリングされた潜在変数をDecoderに入力することで生成される



# 提案法

# 提案モデルの構造

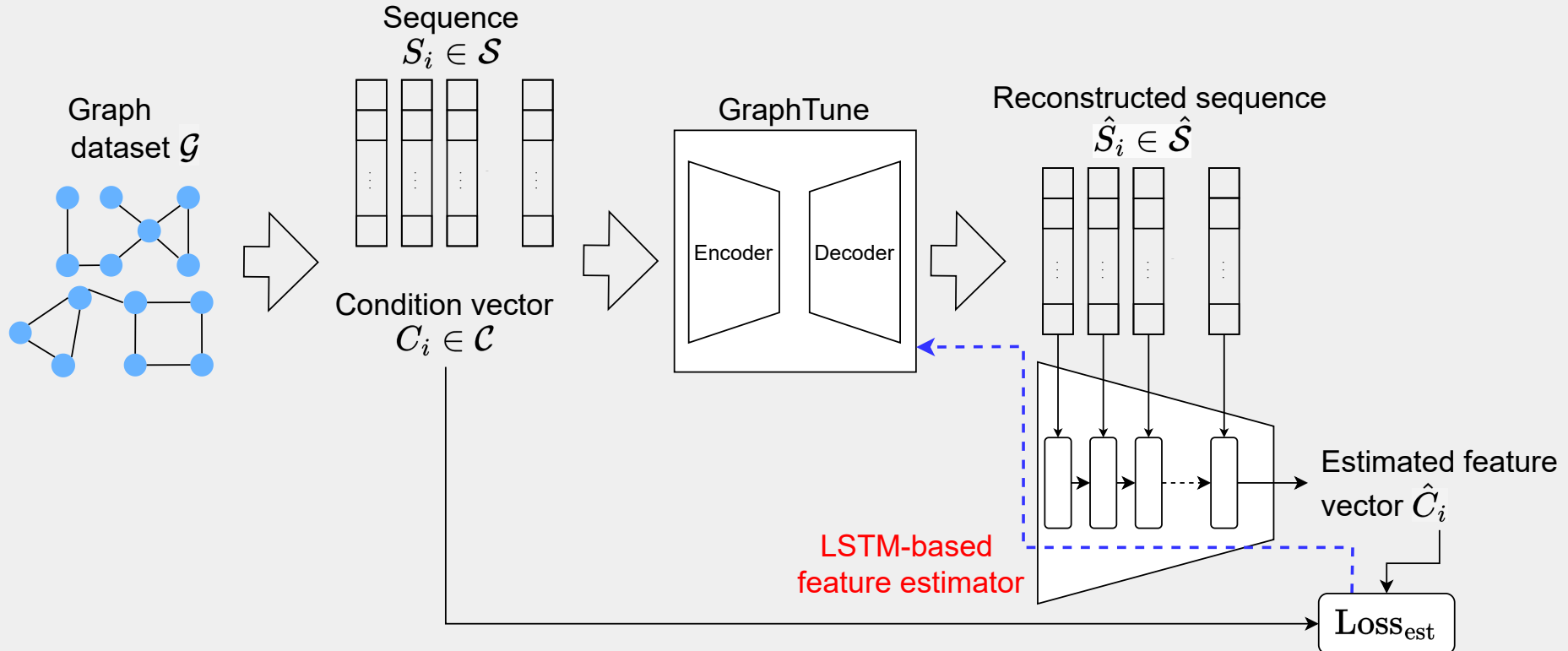
- ◆ GraphTuneにFeature estimatorモデルを追加した生成モデル
  - ◆ Feature estimatorは再構築されたシーケンスからグラフの特徴値を表す条件ベクトルを推定する





# 提案モデルの構造

- ◆ 提案モデルの学習を通して、GraphTuneがFeature estimatorによる特徴値の推定誤差のフィードバックを受けることが可能となる



# 提案モデルの学習

◆提案モデルは、交互学習アルゴリズムによって学習される

## ◆交互学習アルゴリズム

- ◆GraphTuneとFeature estimatorの一方の重みを凍結させ、他方の重みを更新することを繰り返す
- ◆同時に学習した場合、GraphTuneに入力される条件ベクトルの情報がFeature estimatorへリークする可能性がある

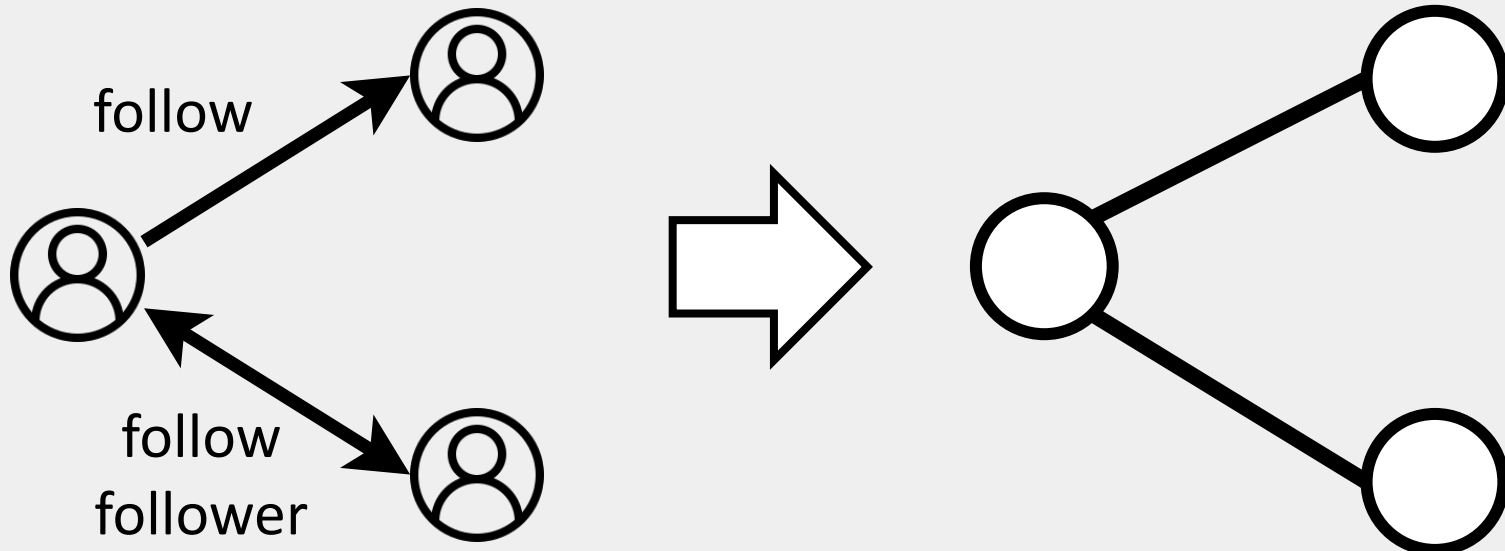
## ◆具体的な手順

- (1) GraphTuneのみを学習させる
- (2) GraphTuneによって生成されたグラフを学習データセットに追加する
- (3) Feature estimatorのみを学習させる
- (4) (1) ~ (3) を繰り返す

# 実験

# データセット

- ◆ Twitter上のアカウントのフォロー・フォロワー関係を表すグラフ
- ◆ Twitterの実グラフからノード数が50となるように、ランダムに探索することで得られた2,000個の部分グラフ
- ◆ 簡易化のため、無向グラフとして扱う



# 実験内容

- ◆ Twitterの部分グラフのデータセットを使用してGraphTuneと提案モデルを学習する
- ◆ 条件ベクトルとして、**平均最短経路長**を用いる
- ◆ データセットを訓練用と検証用に分割
  - ◆ 訓練データ数：1,800
  - ◆ 検証データ数：200
- ◆ 生成
  - ◆ 平均最短経路長を3, 4, 5 と3パターン指定する
  - ◆ 生成するグラフ数：300

# ハイパーパラメータ

## ◆提案モデル

- ◆最初の全結合層のサイズ：256
- ◆LSTMの隠れ層のサイズ：512
- ◆最後の全結合層のサイズ：1
- ◆GraphTune：提案論文で評価されたものを使用

## ◆学習

- ◆最適化関数：Adam
- ◆学習率：0.001
- ◆Weight decay：0.0
- ◆勾配クリップの閾値：1.0
- ◆交互学習の反復回数：2
- ◆バッチサイズ：37
- ◆エポック数：10,000

# 生成結果の評価

## ◆定性的な評価

- ◆生成されたグラフの特徴量の**分布**を確認する

## ◆定量的な評価

- ◆生成されたグラフの特徴量の**平均値**

=> 指定値との差異を評価

- ◆生成されたグラフの特徴量の**RMSE**

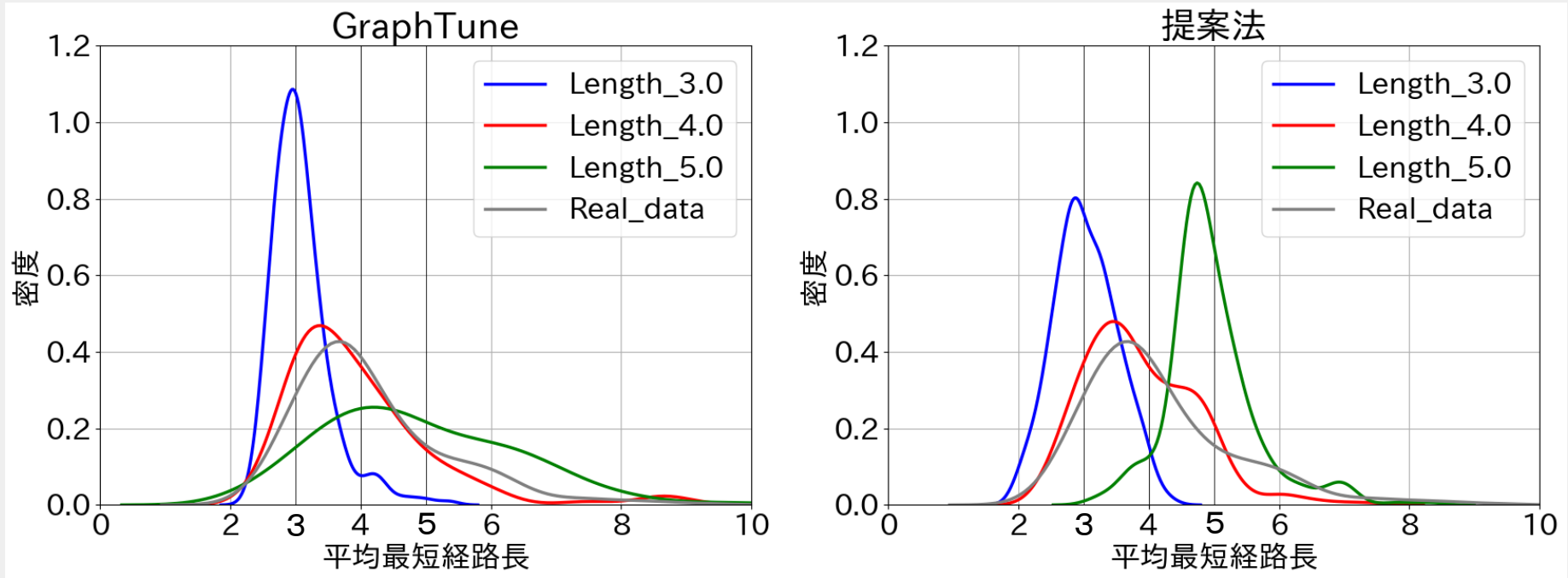
=> 平均値と指定値との差異および分散を同時に評価

$$\text{RMSE} = \left( (E[c_{\hat{S}} - c])^2 + \text{Var}[c_{\hat{S}}] \right)^{\frac{1}{2}} \quad (13)$$

$c_{\hat{S}}$  : 生成されたグラフの特徴量の値

$c$  : 指定した値

# 生成結果：分布



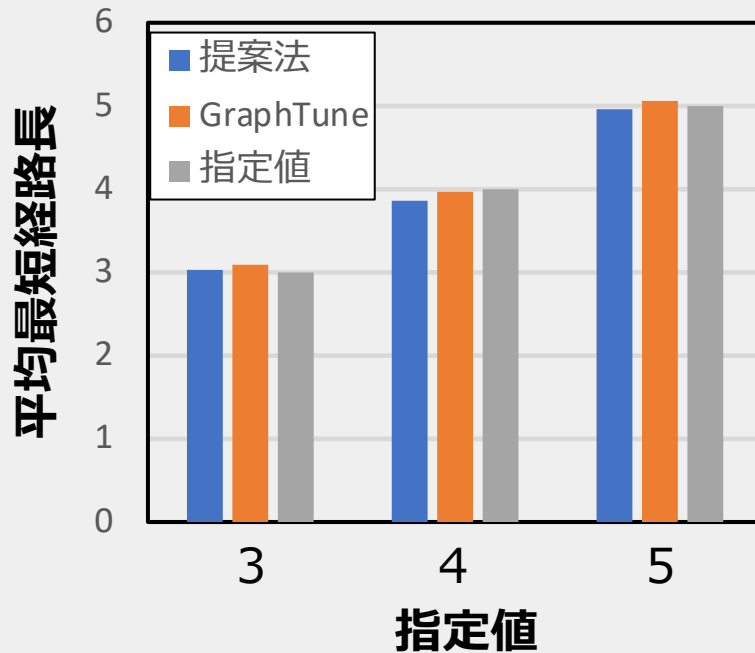
- ◆各手法による生成結果のカーネル密度推定プロットで分布を表示
- ◆平均最短経路長を5と指定した場合の分布は、GraphTuneよりも提案法の方がより指定値の近くに集中して分布している



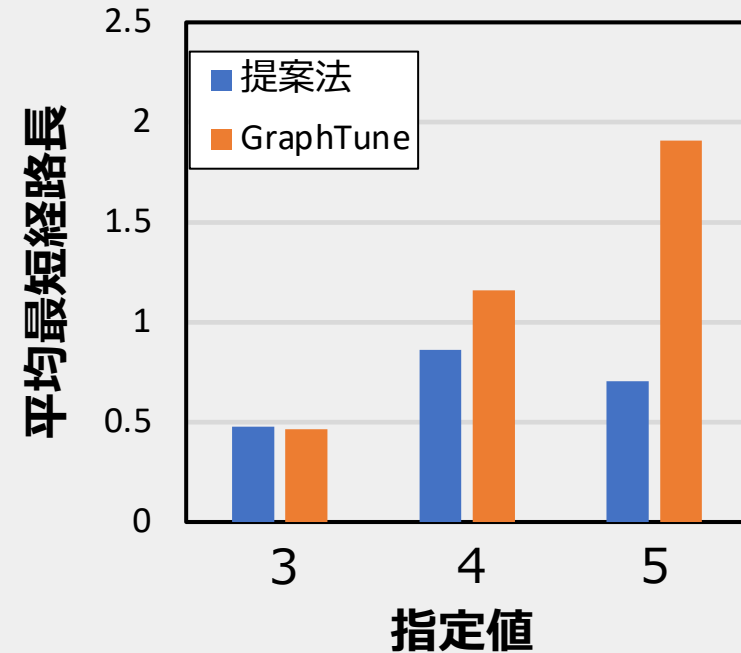
# 生成結果：評価指標

## ◆評価指標による比較

### 平均値による比較



### RMSEによる比較



◆平均値では2つの手法は指定値通りに平均値が推移

◆RMSEでは提案法がGraphTuneと同等かそれ以上の性能を示す

➤提案法は生成精度を高めることができたといえる

# まとめ・今後の展望

# まとめ・今後の展望

## ◆まとめ

- ◆先行研究のGraphTuneをFeature estimatorを追加することで拡張し、交互学習アルゴリズムを適用する手法を提案した
- ◆生成結果の分布の確認による定性的な評価、平均値とRMSEによる定量的な評価を通して、**提案法がGraphTuneよりも高精度な条件付き生成が可能であることを実験的に示した**

## ◆今後の展望

- ◆他の特徴を指定した場合の評価
- ◆交互学習アルゴリズムの反復回数の増加

# 参考文献

- ◆ [1] D.P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” Proc. of the 2nd International Conference on Learning Representations (ICLR 2014), 2014.
- ◆ [2] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol.9, no.8, pp.1735–1780, 1997.
- ◆ [3] S. Nakazawa, Y. Sato, K. Nakagawa, S. Tsugawa, and K. Watabe, “A Tunable Model for Graph Generation Using LSTM and Conditional VAE,” Proc. of the 41st IEEE International Conference on Distributed Computing Systems (ICDCS 2021) Poster Track, 2021.

# 参考文献

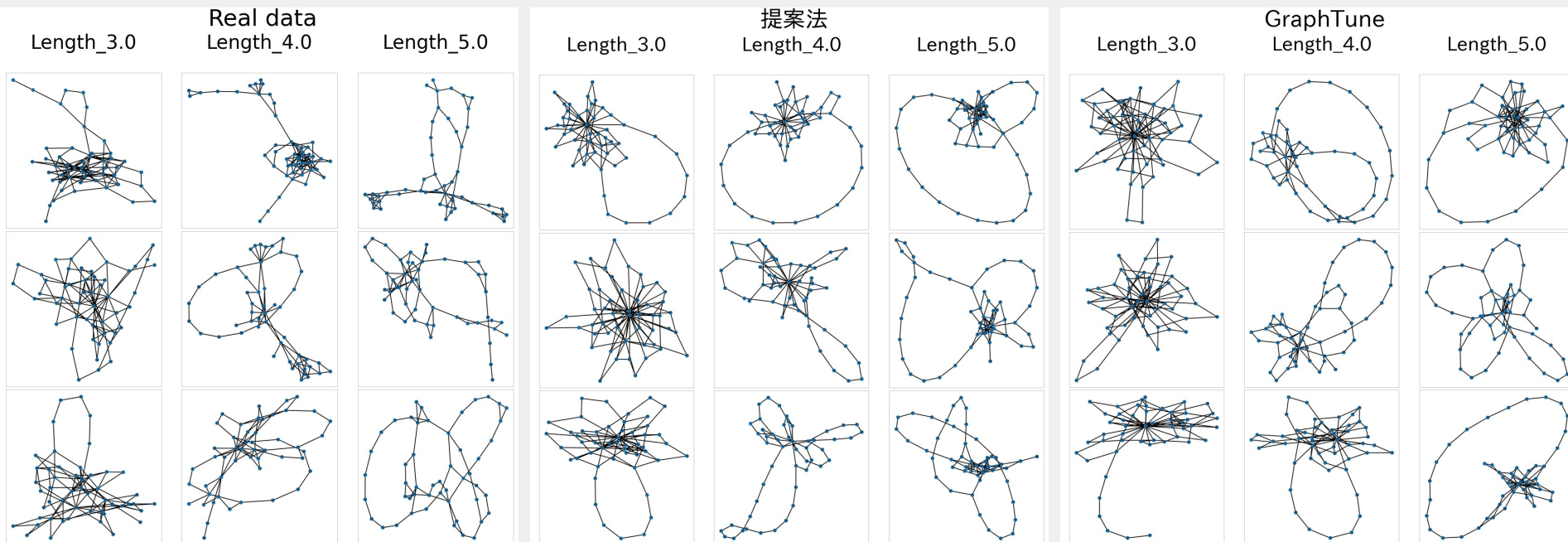
- ◆[4] S. Nakazawa, Y. Sato, S. Tsugawa, K. Nakagawa, and K. Watabe, “Graptune: A learning-based graph generative model with tunable structural features,” CoRR, vol.abs/2201.11494, , 2022. <https://arxiv.org/abs/2201.11494>



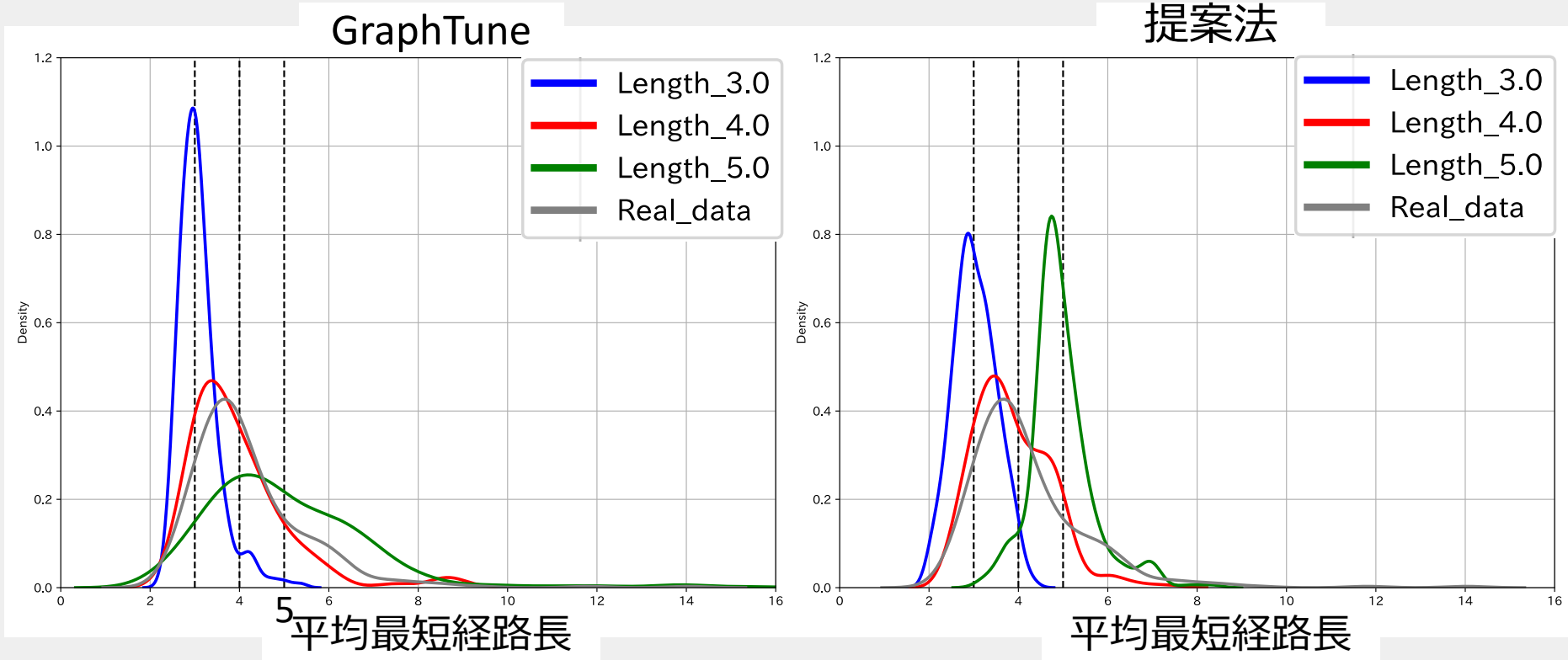
# 補足：生成されたグラフ

◆ 指定値ごとに生成されたグラフの一部を抜粋

◆ Real dataはデータセットから指定値に最も近い3つのグラフを抜粋



# 補足：生成結果の分布の全体



- ◆各手法による生成結果のカーネル密度推定プロットで分布を表示
- ◆平均最短経路長を5と指定した場合の分布は、GraphTuneよりも提案法の方がより指定値の近くに集中して分布している



# 補足：生成結果の定量的評価

## ◆評価指標による比較

指標	Condition	提案法	GraphTune	Real data
平均値	3.0	<b>3.03</b>	3.09	4.23
	4.0	3.86	<b>3.97</b>	
	5.0	<b>4.96</b>	5.06	
RMSE	3.0	0.476	<b>0.463</b>	-
	4.0	<b>0.861</b>	1.16	
	5.0	<b>0.705</b>	1.91	

- ◆多くの場合で提案法はGraphTuneよりも優れたスコアである
- ◆GraphTuneが優れている場合でも、提案法との差は小さい
  - 提案法は生成精度を高めることができたといえる