

グラフ生成モデルの発展と今後の展望

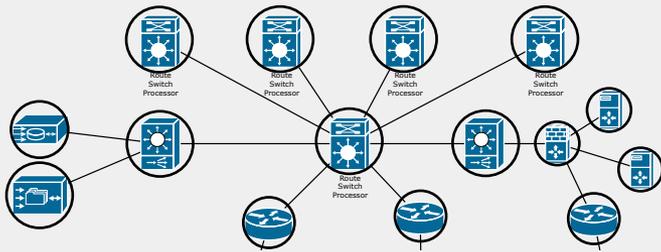
～統計的生成モデルからDeep Graph Generatorsまで～

長岡技術科学大学 大学院工学研究科
渡部康平

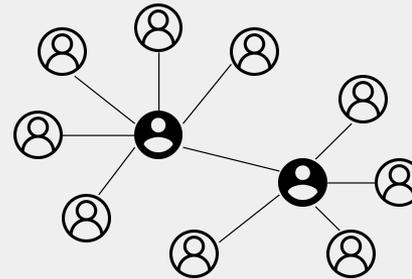
本講演の概要

- ◆ ノードとエッジからなるグラフは極めて汎用性の高いデータ構造。

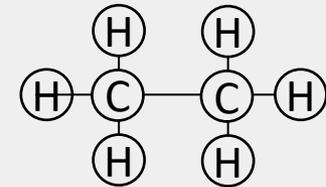
通信ネットワーク



ソーシャルネットワーク



分子構造

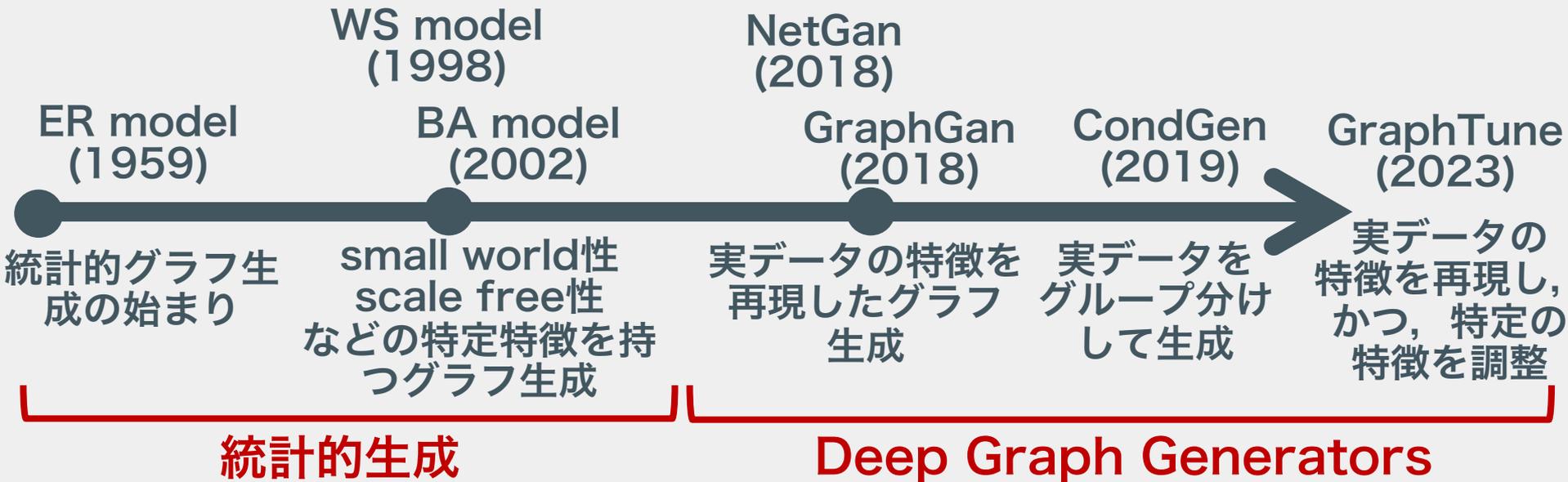


○ ノード (頂点, vertex) — エッジ (辺, リンク)

- ◆ 本講演の主テーマは、グラフ構造を人工的に作るグラフ生成技術。
 - ◆ 少数のパラメータやグラフデータから**存在しない構造や未来の構造を作り出す。**
- ◆ グラフ生成のアプリケーション
 - ◆ 通信プロトコルのシミュレーション
 - ◆ ソーシャルネットワークの情報伝播やコミュニティ予測
 - ◆ プログラムコードのサジェスト
 - ◆ 新しい分子構造の生成による新薬開発

グラフ生成技術の歴史

- ◆ グラフ生成技術は1959年のERモデルに始まり、現代に至る。
- ◆ 広大なグラフの空間から所望の特徴を有したグラフを如何にサンプリングするか。



- ◆ 同時に様々な特徴を定量化した特徴量が開発される。

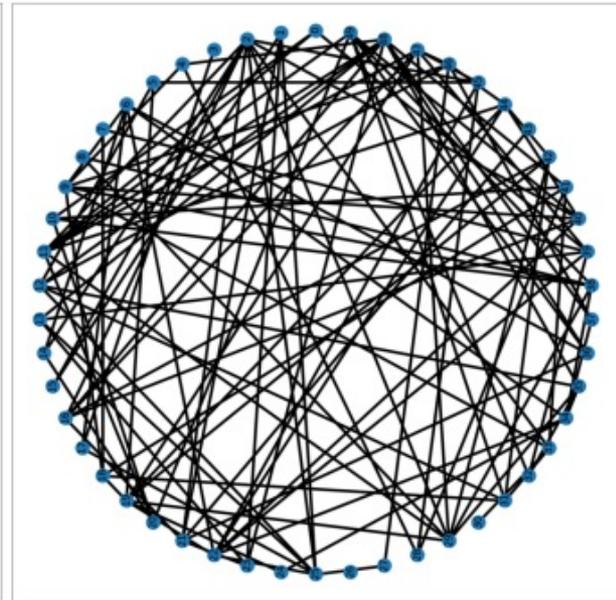
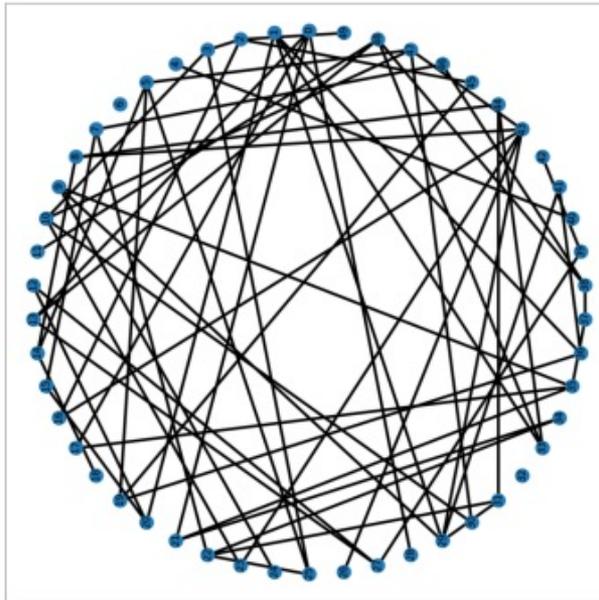
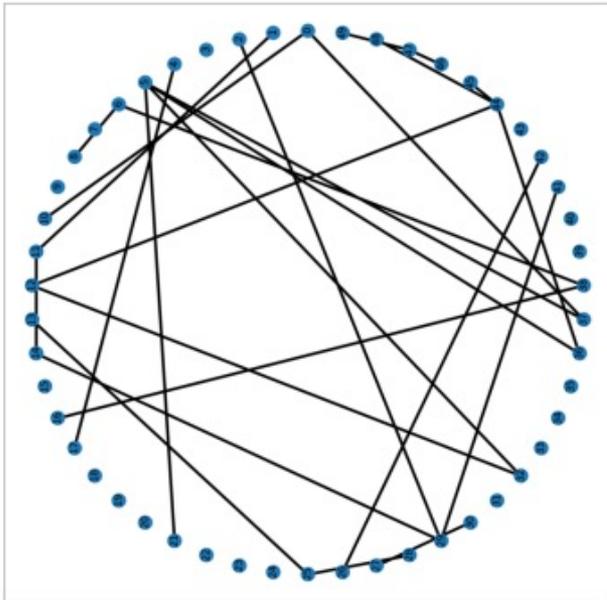
- ◆ 平均最短経路長, 直径
- ◆ クラスタリング係数, モジュラリティ
- ◆ 平均次数, 次数分布のべき指数, エッジ密度
- ◆ 固有値, 次数中心性, 媒介中心性, PageRank

統計的な生成モデル – ERモデル

◆ Erdős–Rényiモデルは最も原始的なグラフ生成モデル。

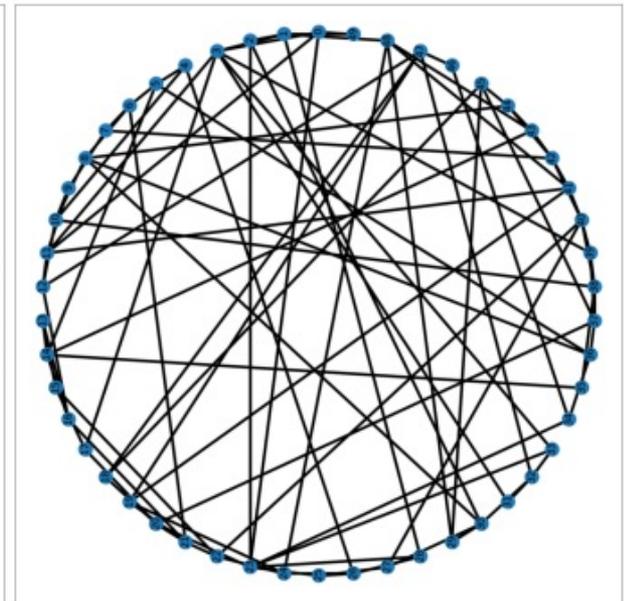
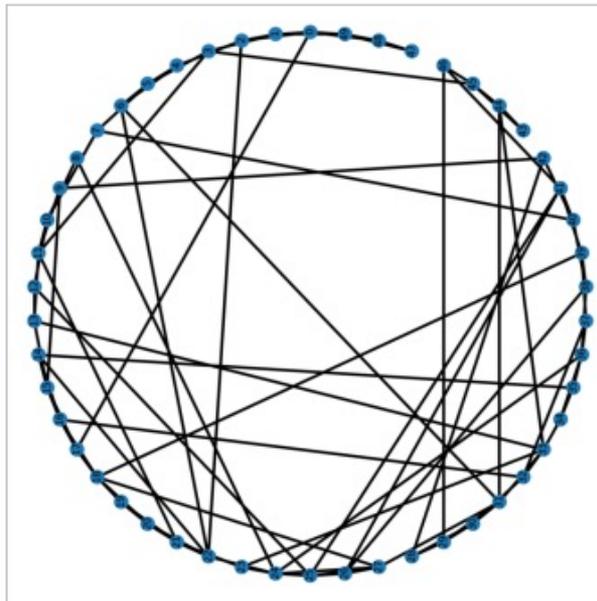
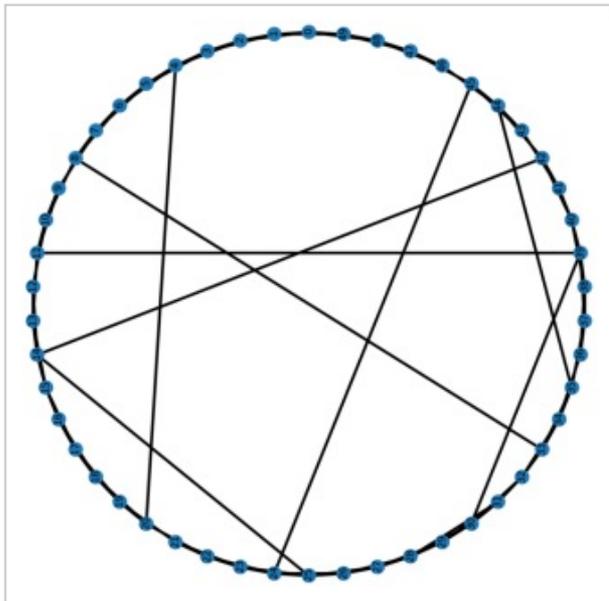
◆ 入力パラメータ: ノード数 n , エッジ密度 p

1. n 個のノードを作成する。
2. すべての2ノード間について確率 p でエッジを張る。



統計的な生成モデル – WSモデル

- ◆ Watts-Strogatzモデルは**スモールワールド性を再現する**グラフ生成モデル。
 - ◆ スモールワールド性: ノード数 n に対して平均経路長 L が $\log n$ でしか増えない。
 - ◆ 入力パラメータ: ノード数 n , 平均次数 $2K$, エッジ張替え確率 p
 1. n 個のノードを作成する。
 2. 平均次数が $2K$ であるリング格子を構成する。
 3. すべてのエッジについて確率 p でエッジを張り替える。



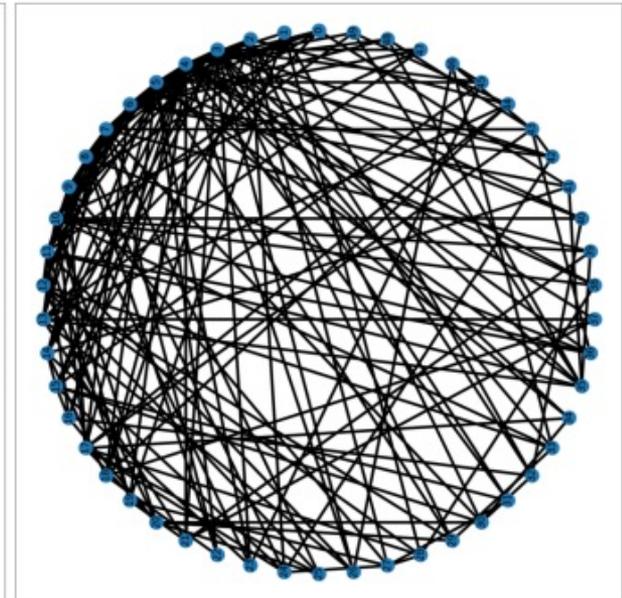
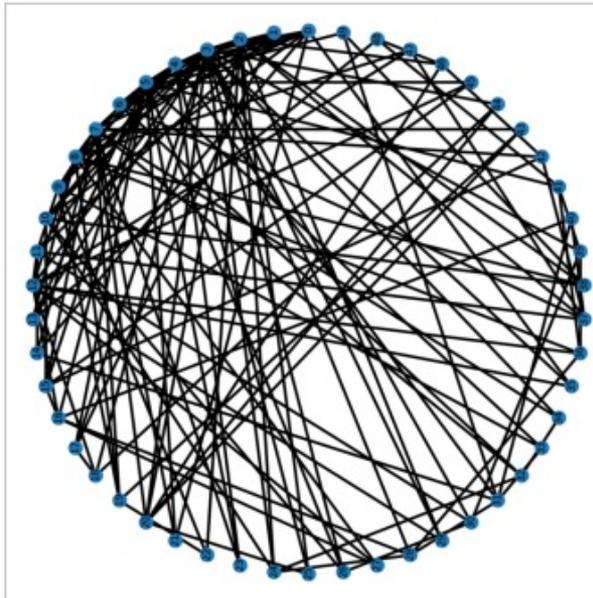
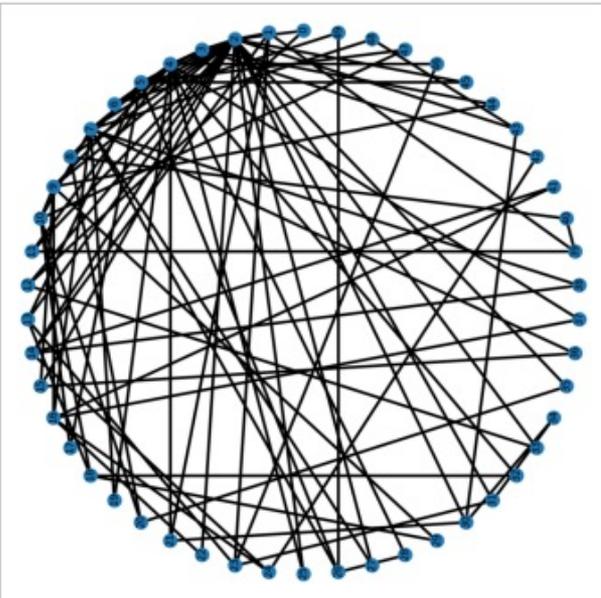
統計的な生成モデル – BAモデル

◆ Barabási–Albertモデルは**スケールフリー性を再現する**グラフ生成モデル。

◆ スケールフリー性: 次数分布 $f(k)$ が $f(k) \propto k^{-\gamma}$ ($2 \leq \gamma \leq 3$) となる。

◆ 入力パラメータ: 追加エッジ数 m , ノード数 n

1. m 個のノードから成る完全グラフを作る。
2. m 本のエッジを持つノードを追加する。ただし、新規ノードからのエッジが張られる確率は、つながる相手先のノードの次数に比例させる。
3. ノード数が n になるまでステップ2を繰り返す。



Deep Graph Generators – GraphGen

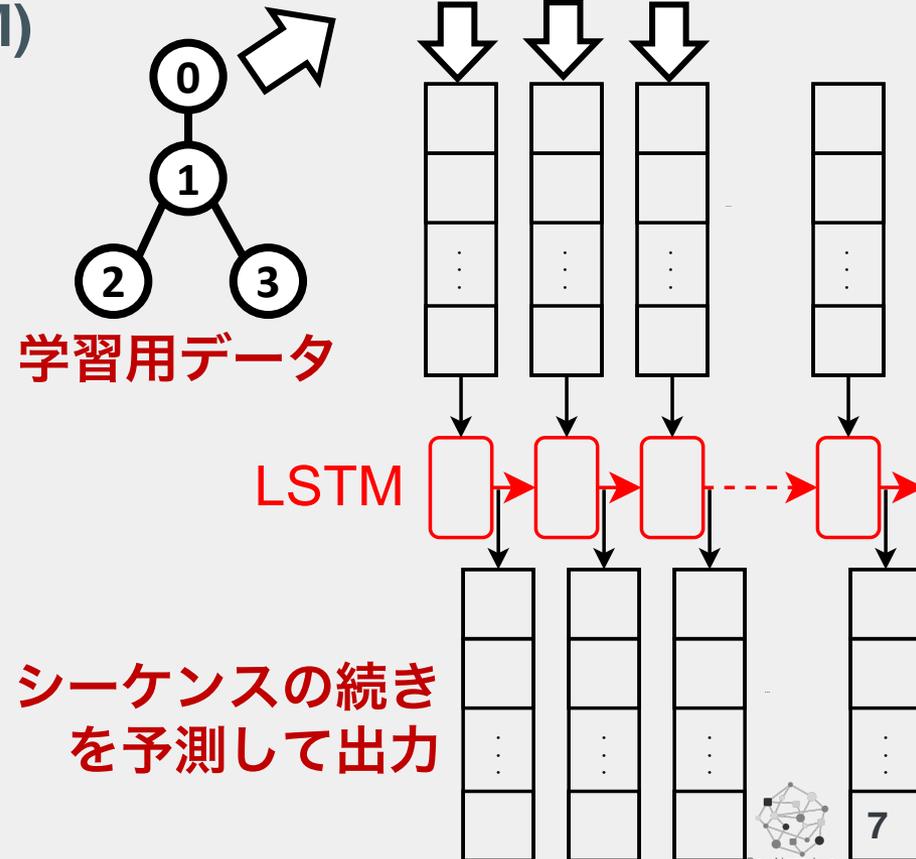
◆統計的なグラフ生成とは対比的に、**Deep Learning** を使って**実グラフを模倣したグラフを生成**。

◆シーケンス予測が可能な深層学習の**Long Short Term Memory (LSTM)** を使って生成。

1. 学習用グラフをエッジのシーケンスに変換する。
2. シーケンスデータを入力して、次のエッジを出力(予測)するよう学習を進める。
3. 学習済みモデルを使って再帰的に出力させて、シーケンスを生成。
4. シーケンスをグラフに逆変換することで、生成グラフを得る。

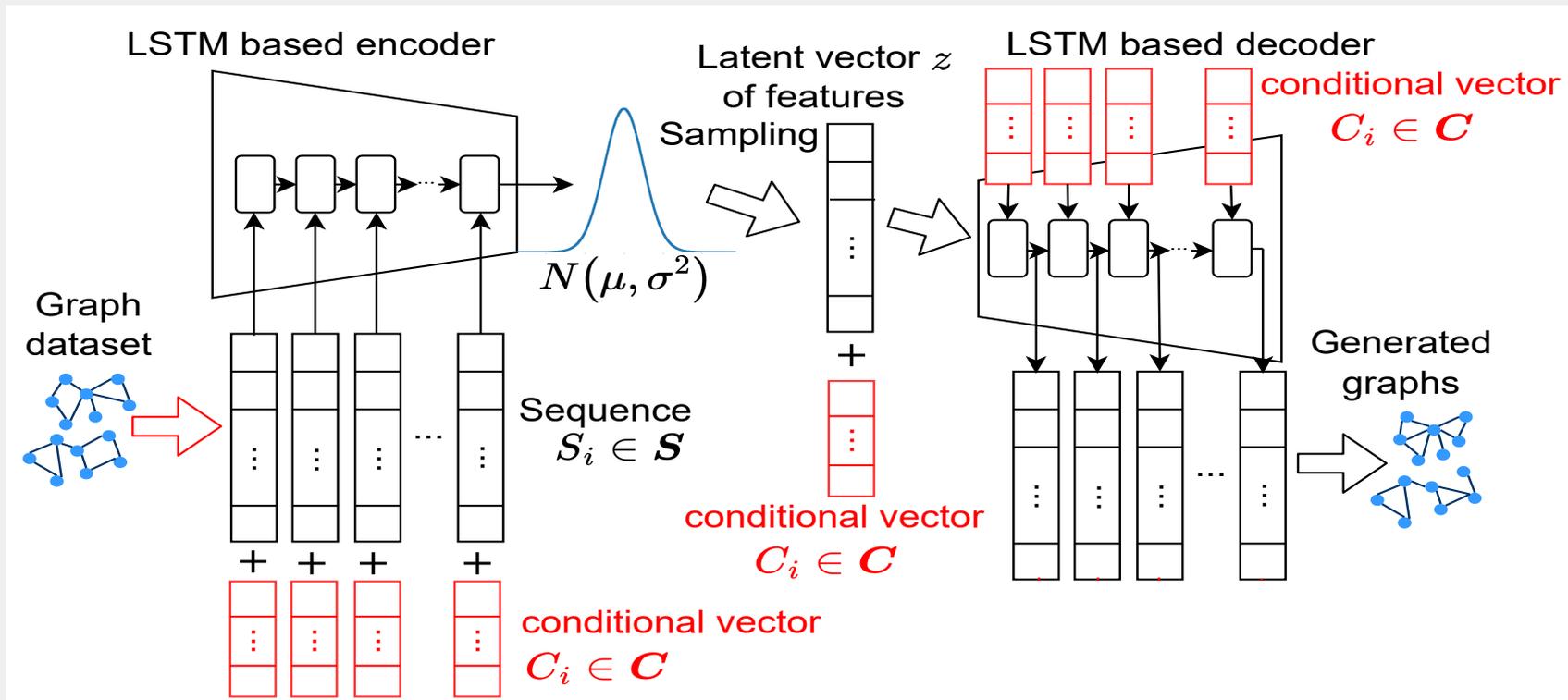
深さ優先探索でエッジのシーケンスデータに変換

n	0	1	2
元ノードID	0	1	1
先ノードID	1	2	3



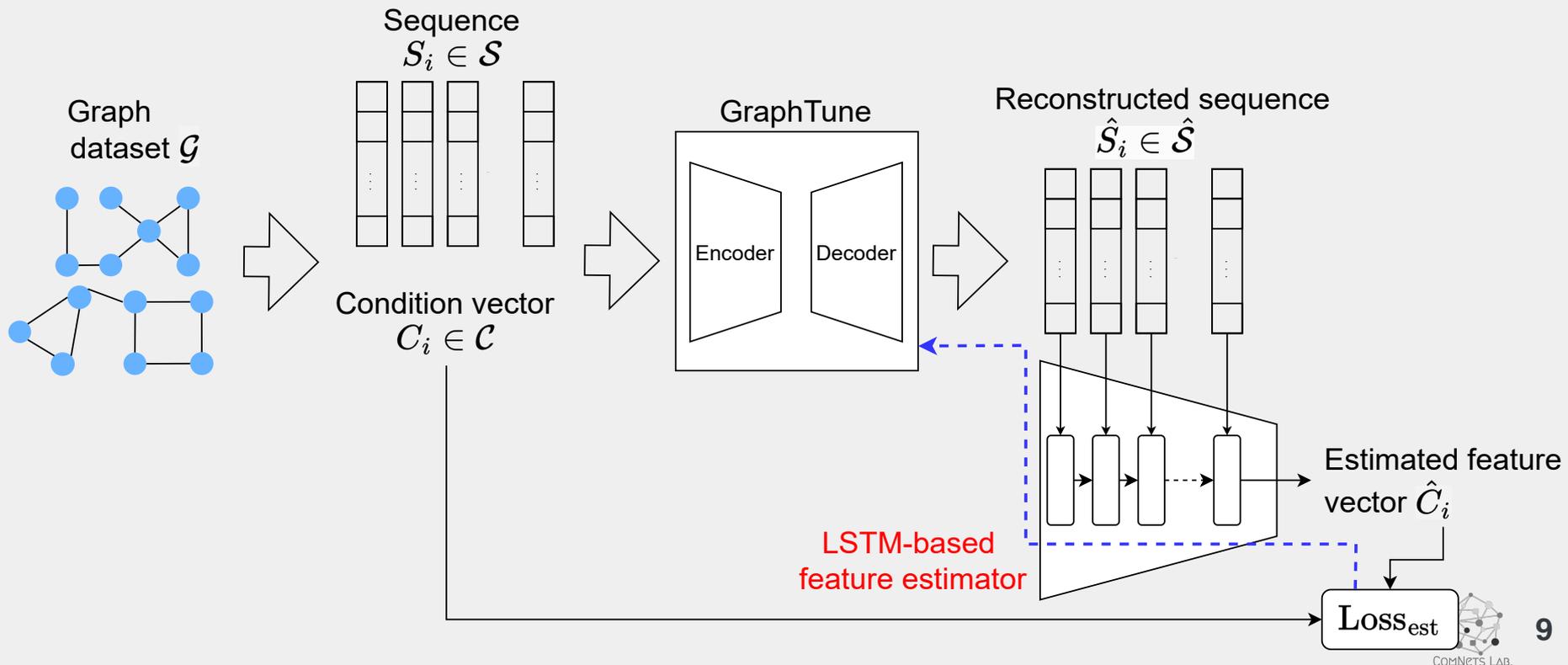
Deep Graph Generators – GraphTune

- ◆ Conditional Variational AutoEncoder (CVAE)と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[1][2].
- ◆ グラフを一度ベクトル空間にマッピングして特徴量の情報を付加することで連続的な調整を実現.
 1. シーケンス化したグラフをベクトルに圧縮して、再びもとのシーケンスに解凍するよう学習.
 2. 学習済みモデルの解凍パートに乱数ベクトルを与えて生成.



Deep Graph Generators – GraphTune-FE

- ◆ GraphTuneをさらに高精度化するために、Feature estimator 特徴量のフィードバック機構を加えた生成モデル。
 - ◆ 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
 - ◆ GraphTuneとFeature estimatorを交互学習



グラフ生成モデルによる生成グラフの検証

- ◆ **グラフ生成問題の目標は、広大なグラフの空間から所望の特徴を有したグラフを如何にサンプリングするか。**
 - ◆ WSモデル: スモールワールド性(平均最短経路長)を再現したい。
 - ◆ BAモデル: スケールフリー性(次数分布のべき指数)を再現したい。
 - ◆ GraphGen: 入力グラフと同じ特徴を再現したい。
 - ◆ GraphTune: 入力グラフの特徴を保持しつつ特定の特徴量だけ調整したい。
- ◆ **グラフを生成して、各種特徴量の分布がどの様になるか。**
 - ◆ 再現したいグラフがある場合、ある特徴量を調整しても別の特徴量が連動して再現対象から離れたりする。
 - ◆ つまり、**多面的な視点で見て再現対象に近いことが重要。**

WSモデルが生成するグラフの空間的特徴

- ◆ WSモデルでエッジ張替え確率 p を変更した場合の特徴量の分布
- ◆ 平均最短経路長は調整されている一方で、次数分布のべき指数は調整不能。

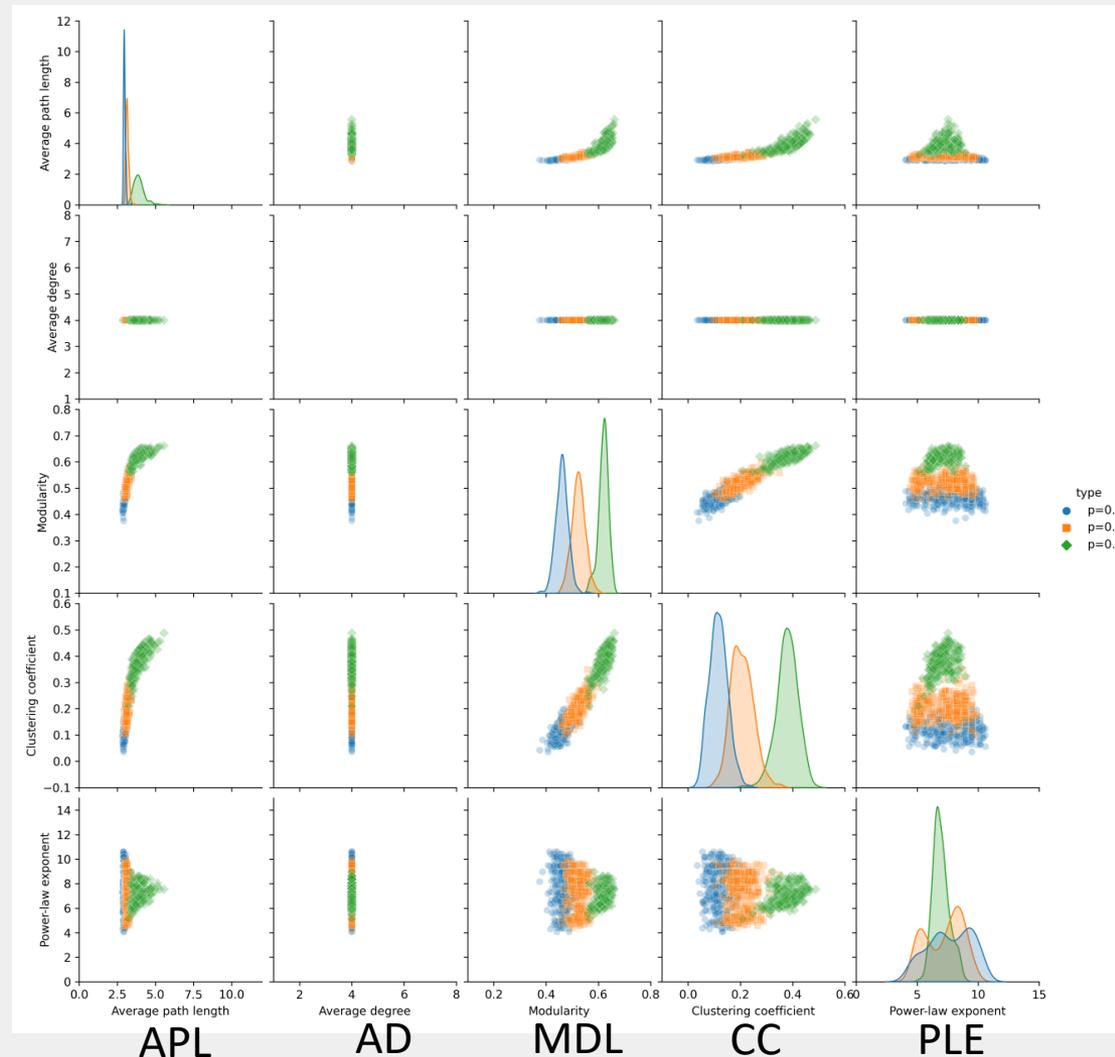
Average Path Length (APL)

Average degree (AD)

Modularity (MDL)

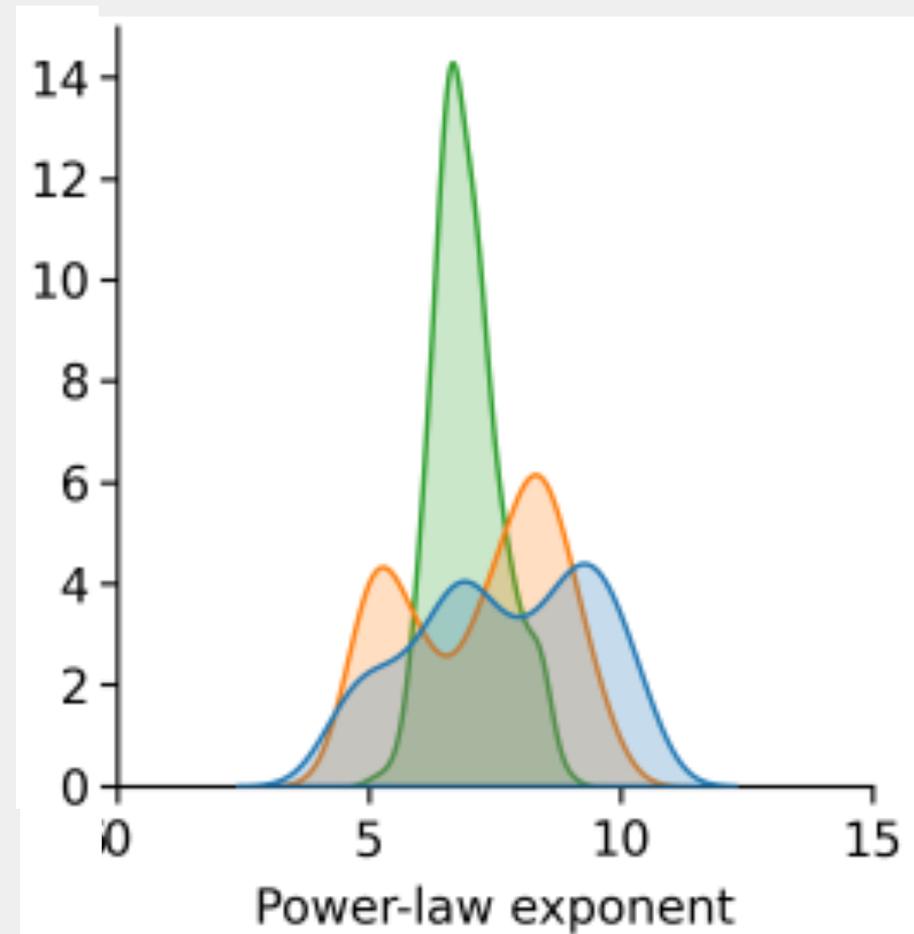
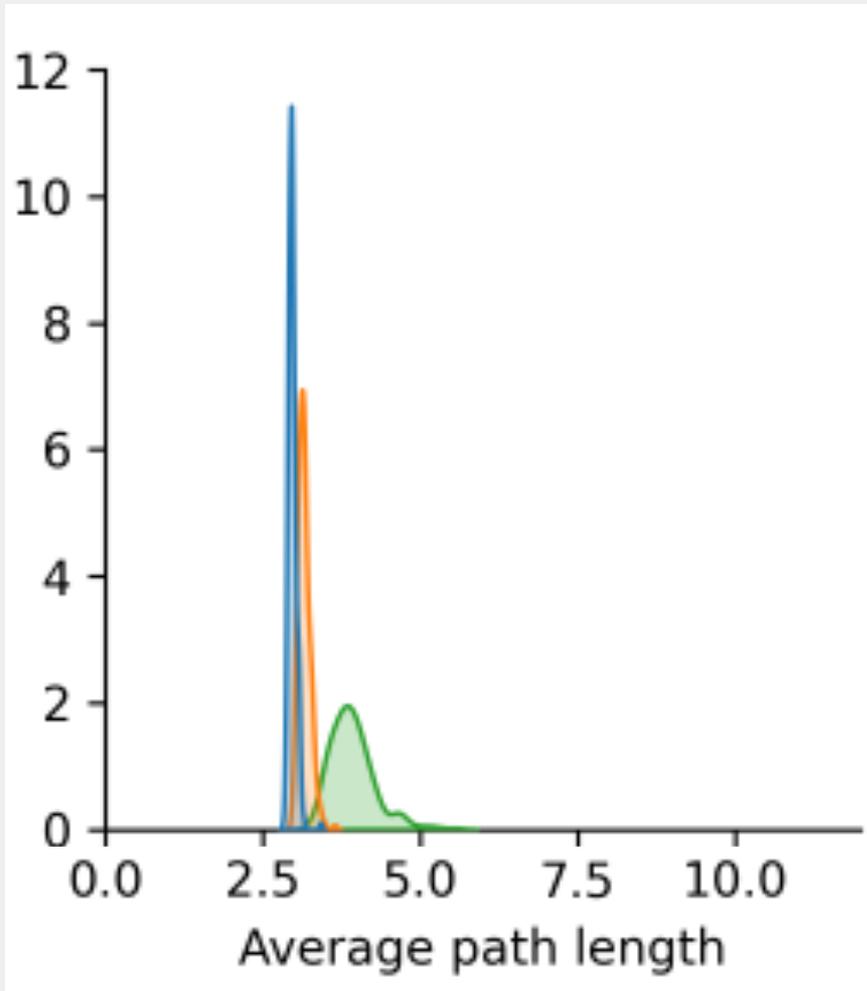
Clustering coefficient (CC)

Power-law exponent (PLE)



WSモデルが生成するグラフの空間的特徴

◆平均経路長と次数分布のべき指数の分布の拡大図.



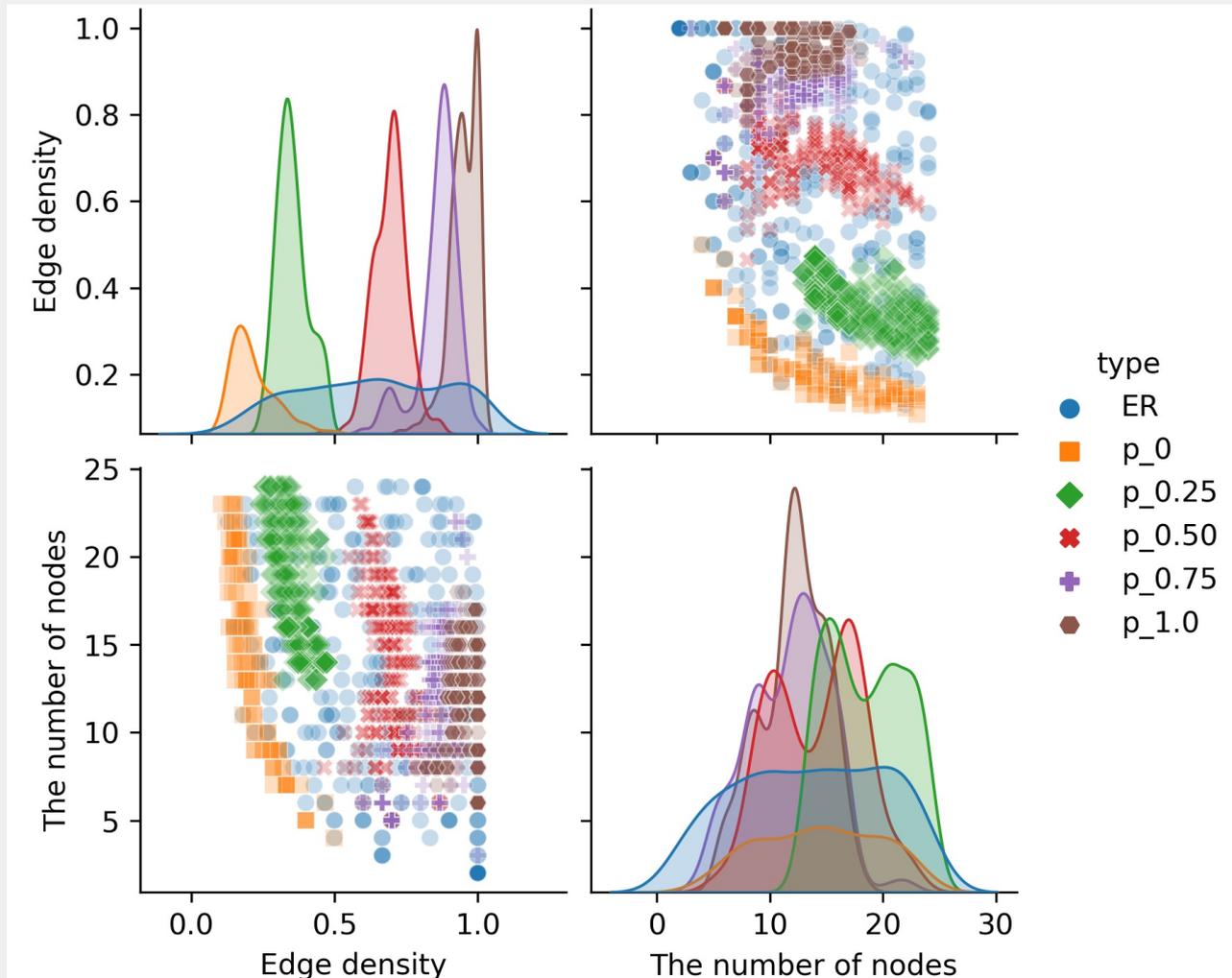
GraphTune-FEでの実験条件

- ◆以下の2つのデータセットを使用して GraphTune-FE を訓練.
 - ◆ERモデルで生成したグラフ (ER dataset)
 - ◆Twitterの部分グラフ (Twitter dataset)
- ◆ER dataset
 - ◆条件ベクトルとして**エッジ密度**を与えて、ノード数と独立に調整可能か検証.
 - ◆エッジ密度を 0.0, 0.25, 0.5, 0.75, 1.0 の5パターンで指定.
- ◆Twitter dataset
 - ◆条件ベクトルとして**平均最短経路長**を与えて、学習グラフの基本的な特徴を保持しつつ、平均最短経路長のみ調整することを試みる.
 - ◆平均最短経路長を 3.0, 4.0, 5.0 と3パターンで指定する

GraphTune-FEが生成するグラフの空間的特徴

◆ ER datasetで学習した場合の生成グラフの分布.

◆ 独立に調整可能な特徴量であるノード数 n とエッジ密度 p が独立に調整可能であることが確認できる.



GraphTune-FEが生成するグラフの空間的特徴

- ◆ Twitter dataset で学習した場合の生成グラフの分布.
- ◆ 全体的な分布が学習データ(Real data)と一致している.

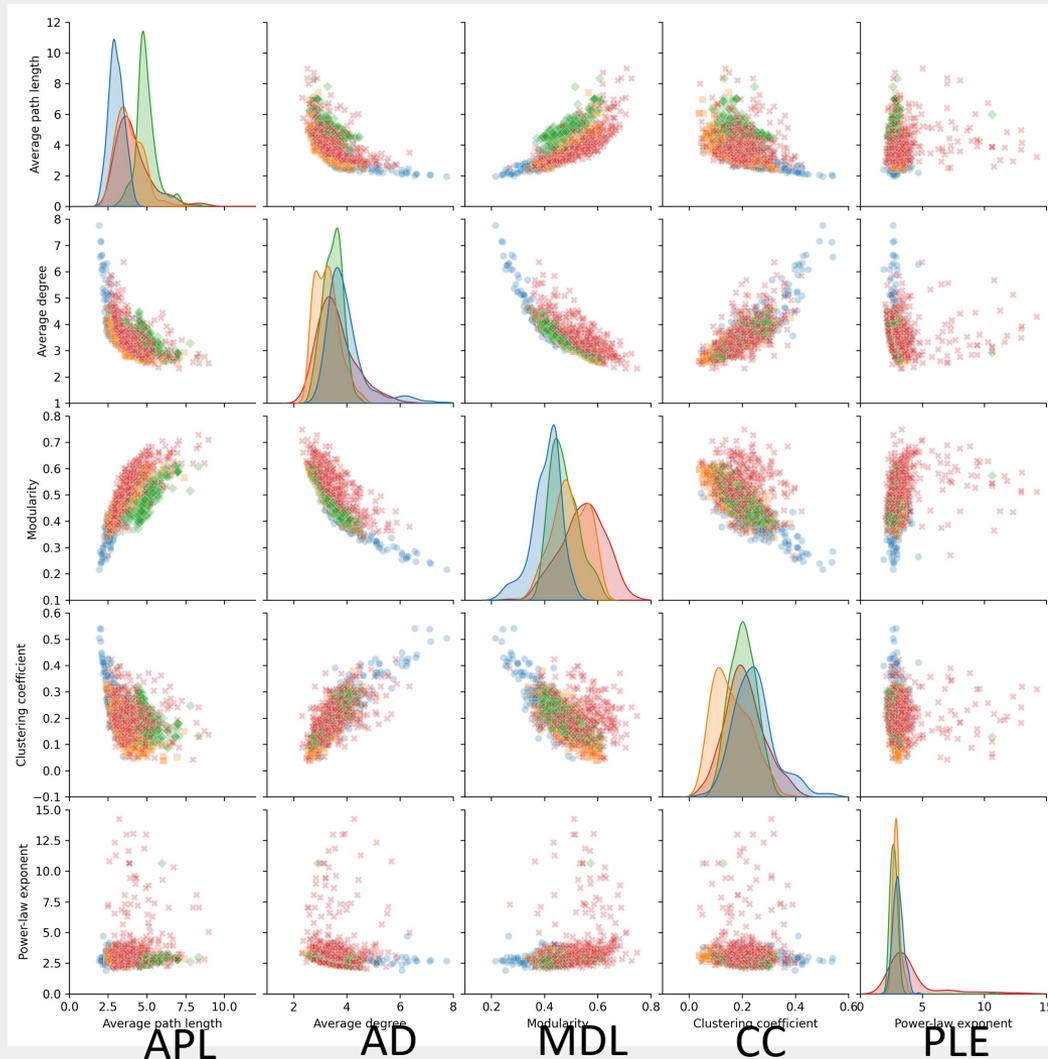
Average Path Length (APL)

Average degree(AD)

Modularity (MDL)

Clustering coefficient (CC)

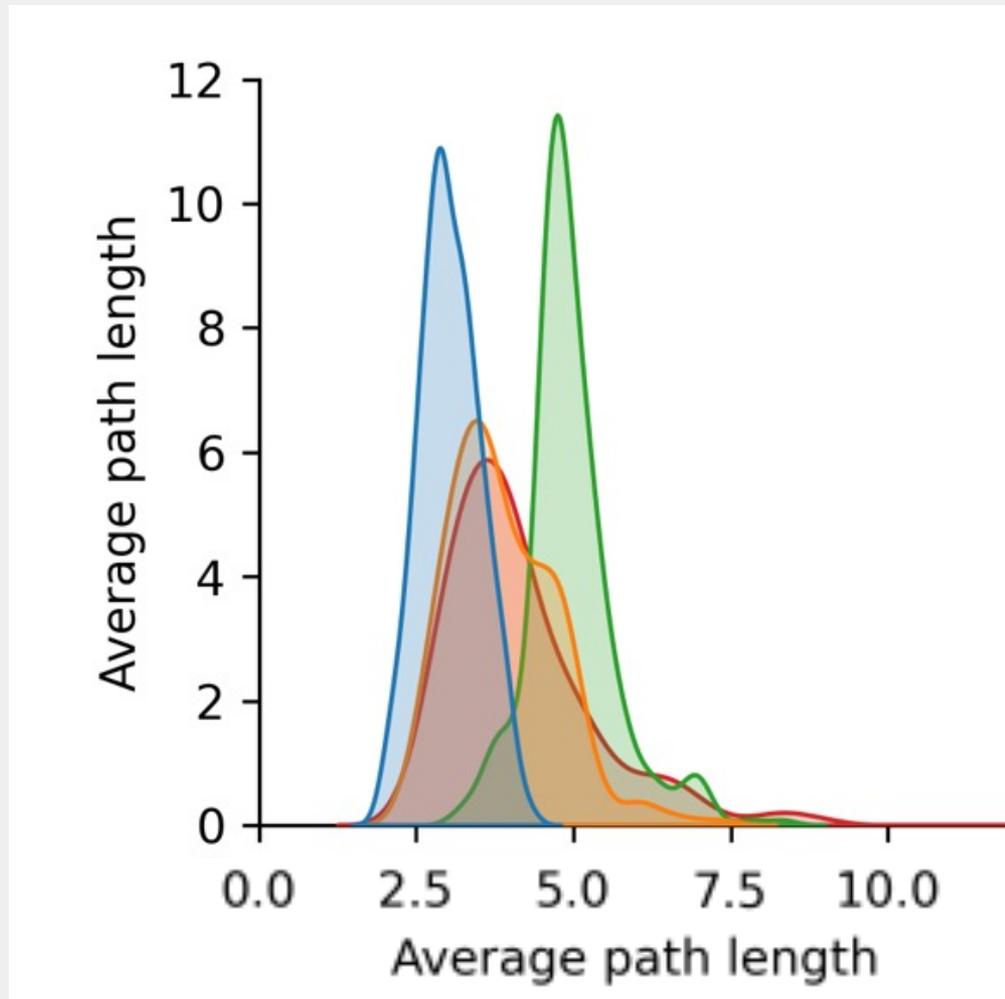
Power-law exponent (PLE)



- Length = 3.0
- Length = 4.0
- ◆ Length = 5.0
- * Real data

GraphTune-FEが生成するグラフの空間的特徴

◆平均最短経路長の分布の拡大図.



- Length = 3.0
- Length = 4.0
- ◆ Length = 5.0
- * Real data

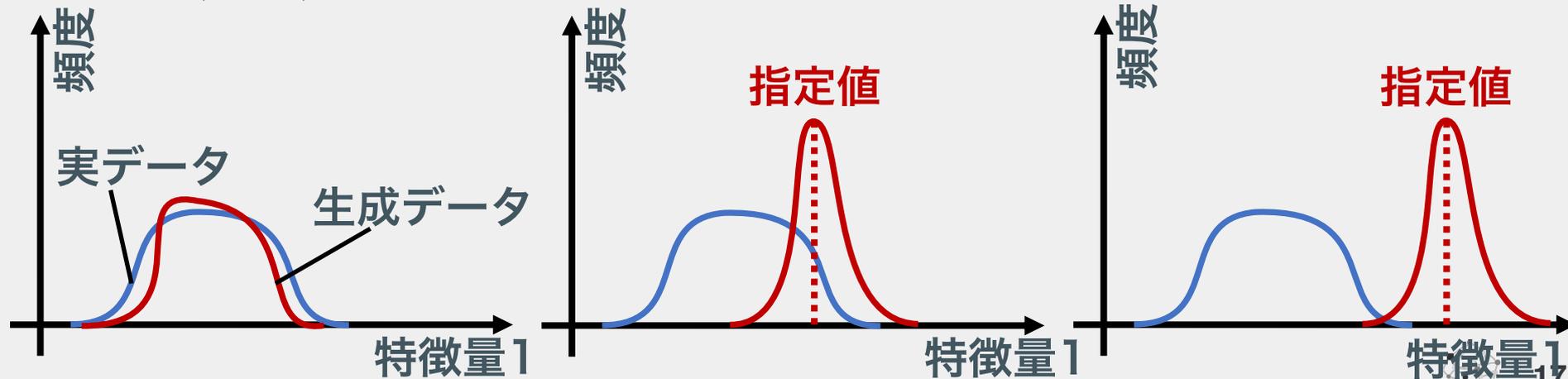
今後の展望 – 条件付き生成技術

- ◆ グラフ生成技術は、学習データを模倣した生成は成熟してきている。
- ◆ 今後は**条件付き生成技術の開発**が中心になると考えている。
 - ◆ 調整精度の高精度化
 - ◆ 調整対象の多角化
 - ◆ 調整領域の拡張
- ◆ アプリケーションの多様さを考えても条件付き生成は有望。

学習データに類似したグラフを生成
(従来)

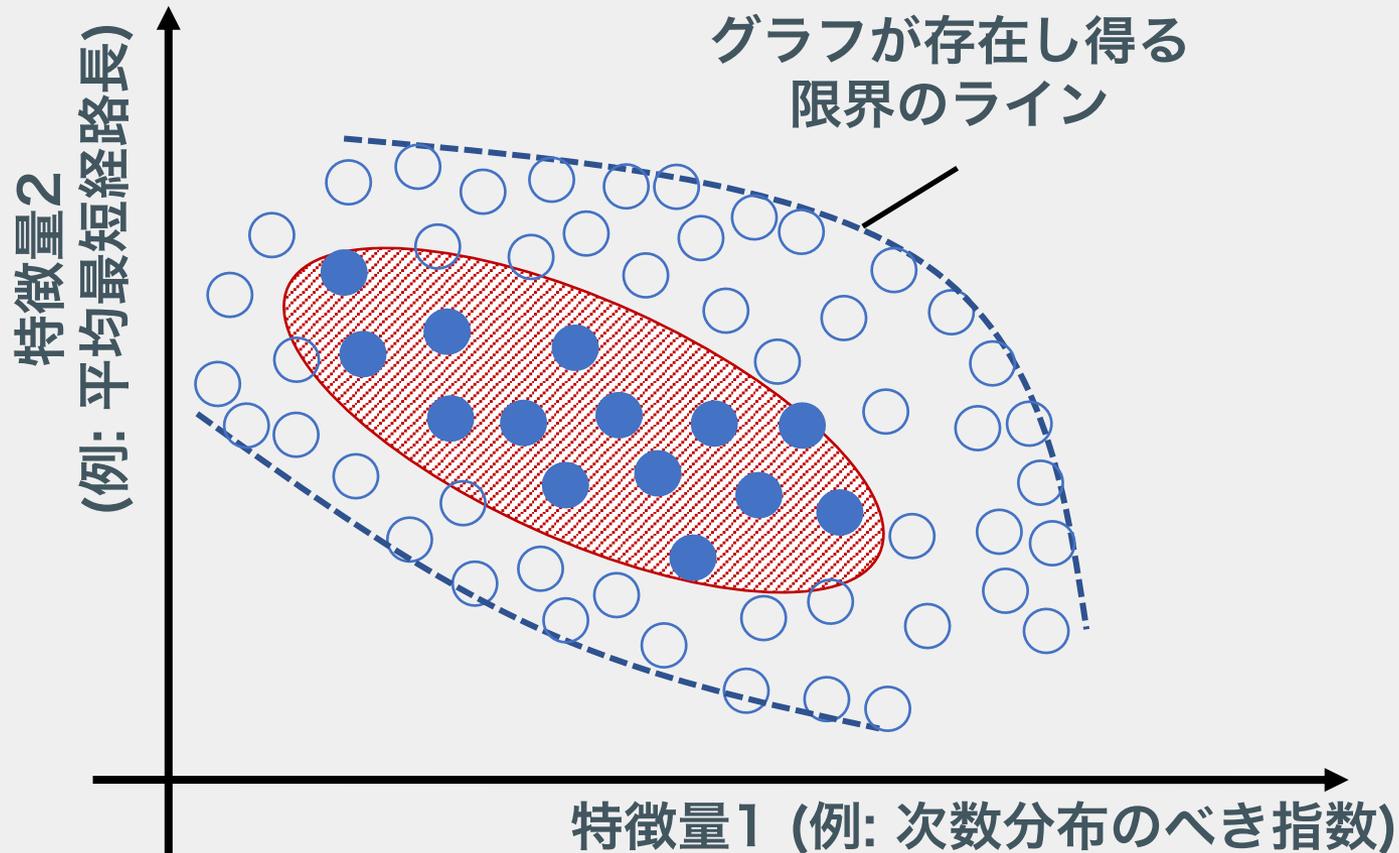
学習データの一部に類似したグラフを指定して生成 (現状)

学習データに存在しない領域のグラフを生成 (今後)



今後の展望 – Data driven なグラフ理論

- ◆条件付きグラフ生成技術の成熟すると、任意の特徴量を持つグラフを得ることが可能になる。
 - ◆グラフが存在し得る空間の全体像を把握可能。
 - ◆グラフの特徴量間の関係性を解析・検証することが可能。
 - ◆グラフの空間を均質にサンプリングしたデータセットを作れる。



まとめ・今後の展望

◆まとめ

- ◆グラフ生成問題は、広大なグラフの空間から所望の特徴を有したグラフをサンプリングする問題。
- ◆生成技術は統計的生成から入力グラフの特徴を学習的に再現する研究が活発になってきた。
 - ◆ER, WS, BAモデル ⇒ GraphGen, GraphTune

◆今後の展望

- ◆特に**条件付き生成技術がまだまだ未成熟**であるがゆえに有望。
- ◆条件付きグラフ生成の技術が成熟すると、**古典的なグラフ理論の研究にも恩恵があるかも。**

関連文献

- [1] 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.
- [2] 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, 2023.

ご清聴ありがとうございました