修 士 論 文 概 要 書

Master's Thesis Summary

Date of submission: 01/25/2021 (MM/DD/YYYY)

専攻名（専門分野） Department	情報理工・ 情報通信専攻	氏 名 Name	浅見 莉絵子	指 導 教 員 Advisor	渡辺裕 印 Seal
研究指導名 Research guidance	オーディオ・ビジュアル情報処理研究	学籍番号 Student ID number	CD 5119F001-5		
研究題目 Title	グラフ類似性測定のためのグラフの局所構造を考慮した Gromov-Wasserstein 距離 Gromov-Wasserstein Distance Considering Graph Local Structure for Graph Similarity Measurement				

1. はじめに

近年，グラフやネットワーク，ツリー等の``構造"を有するデータを用いた様々なアプリケーションが検討されている．例えば，ソーシャルネットワークやバイオインフォマティクス，交通トラヒックなどが挙げられ，広範囲に亘る．しかしこれらの構造は一般的に取り扱いが難しいことが知られており，その問題に対して，例えば Graph/network Embedding では，データ構造を低次元のベクトル情報で表現することで処理の簡略化を実現している．一方で，グラフ構造やネットワーク構造そのものを処理する手法も数多く提案されており，その中の一つとして Gromov-Wasserstein（GW）距離を用いたグラフ処理が提案されている[1]．しかし GW 距離では，ノード間あるいはグラフ間の距離を最短距離等から計算することから，グラフ構造，特にエッジ構造を正確に表現できているとは言い難い．また，GW を基にしたラベルを使用したグラフ処理手法は，ラベルの種類が異なるグラフの比較ができない．そこで本研究では，Gromov-Wasserstein 距離の距離尺度としてグラフの局所構造に着目した．局所的なエッジ構造を GW 距離に組み込み，グラフ固有の特性を表現する手法を提案した．また，ラベル構造に着目した新たな距離尺度を組み込み，ラベルの異なるグラフの比較が可能な手法を提案した．

2. 関連研究

2.1. 最適輸送 [2,3]

最適輸送は，同じ空間上または事前に登録された複数の空間上で定義される．最適輸送問題は質量移動問題として解釈され，ある分布の質量を，総質量を変えずに別の分布に移動する時，最小のコストで移動するための最適な計画を求める．最適輸送は，辞書学習からドメイン適応，クラスタリング，半教師あり学習まで，さまざまな機械学習問題で注目されている．しかし，オブジェクトの特徴表現にのみ依存しているため，構造情報を活用できない．

2.2. Gromov-Wasserstein discrepancy [1,4]

確率測度が異なる確率空間にある場合，最適輸送ベースの距離は無効になる．そこで Gromov-Wasserstein（GW）discrepancy では，空間が異なる場合に最適輸送を拡張する．二つの類似度行列$(\boldsymbol{C_1},\boldsymbol{p}) \in \mathbb{R}_+^{n_1\times n_1}\times\Delta_{n_1}$と$(\boldsymbol{C_2},\boldsymbol{q}) \in \mathbb{R}_+^{n_2\times n_2}\times\Delta_{n_2}$の GW discrepancy $d_{gw}(\boldsymbol{C_1},\boldsymbol{C_2},\boldsymbol{p},\boldsymbol{q})$は，以下のように定義される．

$$d_{gw}(\boldsymbol{C_1},\boldsymbol{C_2},\boldsymbol{p},\boldsymbol{q}) = \min_{T\in\mathcal{U}_{n_1n_2}} E_{C_1,C_2}(\boldsymbol{T}) \quad (1)$$

ここで，行列$\boldsymbol{T}$は二つの空間の結合であり，$E_{C_1,C_2}(\boldsymbol{T})$は以下のように定義される．

$$E_{C_1,C_2}(\boldsymbol{T}) = \sum_{i,j,k,l} L\big(C_1(i,k),C_2(j,l)\big)\boldsymbol{T}_{i,j}\boldsymbol{T}_{k,l} \quad (2)$$

ここで，$L(a,b)$はaとbの discrepancy を測定するための損失関数を表す．GW discrepancy が小さいことは類似度が高いことを意味する．

3. Edge-structure-enhanced Gromov Wasserstein（EGW）距離

3.1. EGW 距離

本節では，グラフのエッジ構造を用いて構造的距離を表した EGW 距離を提案する．本手法は GW discrepancy に基づく．$(\boldsymbol{D_1},\boldsymbol{p}) \in \mathbb{R}_+^{n_1\times n_1}\times\Delta_{n_1}$と$(\boldsymbol{D_2},\boldsymbol{q}) \in \mathbb{R}_+^{n_2\times n_2}\times\Delta_{n_2}$がそれぞれ，エッジ構造が強化された二つの類似度行列を表す．二つの類似度行列を用いた距離を EGW 距離$d_{egw}(\boldsymbol{D_1},\boldsymbol{D_2},\boldsymbol{p},\boldsymbol{q})$とし，以下のように定義する．

$$d_{gw}(\boldsymbol{D_1},\boldsymbol{D_2},\boldsymbol{p},\boldsymbol{q}) = \min_{T\in\mathcal{U}_{n_1n_2}} \sum_{i,j,k,l} |\boldsymbol{D_1}(i,k)-\boldsymbol{D_2}(j,l)|\boldsymbol{T}_{i,j}\boldsymbol{T}_{k,l} \quad (3)$$

上式における$\boldsymbol{D_1}(i,k)$，$\boldsymbol{D_2}(j,l)$はそれぞれエッジ構造を強化した距離を表し，本研究では二つのアプローチを検討した．一つ目は，着目するノードが持つエッジの数と，そのエッジが構成する三角形の数を考慮した手法である．本手法を EGW-Degree Triangle（EGW-DT）と呼び，以下のように定義する．

$$\boldsymbol{D_D}(i,k) = |p_{i,k}^*| + \frac{|\sigma_i-\sigma_k|}{1+\sigma_{max}} + \frac{1}{1+\tau_{max}}\left|\sum_{s=1}^{m_i}\tau(e_i^s) - \sum_{t=1}^{m_k}\tau(e_k^t)\right| \quad (4)$$

上式で，$p_{i,k}^*$は i 番目のノードから k 番目のノードへの shortest path を表す．σ_iはi番目のノードが持つエッジの数を表す．σ_{max}はグラフ内の最大エッジ数を表す．$\tau(e_i^s)$はi番目ノードに接続する s 番目のエッジが構成する三角形の数を示す．τ_{max}はグラフ内の三角形の最大数を表す．

次に，二つのノード間のパスに沿ったエッジの構造について考慮した手法を提案した．本手法を EGW-Triangle Weight（EGW-TW）と呼び，以下のように定義する．

$$\boldsymbol{D_{TW}}(i,k) = \min \sum_{(u,v)\in p} \frac{1}{1+\tau(e_{u,v})} \quad (5)$$
$$\textit{Subject to}\ \ p \in P_{i,k}$$

上式において，(u,v)はパス p に沿った一つのエッジを表す．パスpはi番目のノードとk番目のノード間の複数パス$P_{i,k}$の一つである．

3.2. 評価実験

提案手法の有効性を評価するために，Support Vector Machine（SVM）分類器によるグラフ分類を行った．比較として，GW，Shortest Path Kernel（SPK），Random Walk Kernel（RWK），Pyramid Matched Kernel（PMK）を用いた．グラフに含まれるノードの数に対してエッジの数が多く，構造が複雑なデータセットを使用し，2 対 1 で分割してトレーニングデータとテストデータを作成した．データセットの分割，トレーニング，テストを 10 回繰り返し，分類精度の平均値と標準偏差を算出した．結果を表 1 に示す．

平均値±標準偏差として数値を示し，最も精度が高いものは太字，二番目に高いものは下線で示した．γは各手法をカーネルとして用いるときの重み係数である．

表 1 グラフの平均分類精度

メソッド		データセット	
	γ	IMDB-B	IMDB-M
SPK		50.36±2.06	34.12±2.76
RWK		48.15±3.38	30.12±1.45
PMK		64.27±1.99	43.37±1.92
GW	0.1	57.36±1.73	38.28±0.95
	1.0	56.03±2.09	41.54±1.96
	10	63.45±1.55	44.81±1.49
	100	59.39±3.67	<u>46.81±1.84</u>
EGW-DT	0.1	57.64±1.89	39.90±0.90
	1.0	61.70±2.20	45.68±1.64
	10	59.48±2.46	46.65±1.84
	100	58.24±3.48	**46.93±1.82**
EGW-TW	0.1	61.85±3.18	33.92±3.54
	1.0	62.39±1.53	40.79±1.21
	10	<u>64.24±0.94</u>	41.81±2.00
	100	**66.94±1.65**	45.82±1.63

表 1 から，EGW-TW と EGW-DT は，ノードに対してエッジの数が多い複雑なグラフのデータセットにおいて他の手法よりも優れていた．

4. Label-invariant Gromov-Wasserstein（LGW）距離

4.1. LGW 距離

本節では，GW discrepancy に基づいてグラフのノードラベルを考慮した LGW 距離を提案する．$(\boldsymbol{F_1}, \boldsymbol{p}) \in \mathbb{R}_+^{n_1 \times n_1} \times \Delta_{n_1}$と$(\boldsymbol{F_2}, \boldsymbol{q}) \in \mathbb{R}_+^{n_2 \times n_2} \times \Delta_{n_2}$を，それぞれノードラベルを用いた構造が考慮された，比較する二つの類似度行列を表すとする．二つの類似度行列を用いた距離を LGW 距離$d_{egw}(\boldsymbol{F_1}, \boldsymbol{F_2}, \boldsymbol{p}, \boldsymbol{q})$とし，以下のように定義する．

$$d_{gw}(\boldsymbol{F_1}, \boldsymbol{F_2}, \boldsymbol{p}, \boldsymbol{q}) = \min_{T \in \mathcal{U}_{n_1 n_2}} \sum\nolimits_{i,j,k,l} |\boldsymbol{F_1}(i,k) - \boldsymbol{F_2}(j,l)| \boldsymbol{T}_{i,j} \boldsymbol{T}_{k,l} \qquad (6)$$

上式における$\boldsymbol{F_1}(i,k)$，$\boldsymbol{F_2}(j,l)$はそれぞれノードラベルの局所的な配置を用いた距離を表す．本研究では二つのアプローチを検討した．一つ目は，ラベルの発生頻度に着目した手法であり，LGW-Label Frequency（LGW-LF）として，以下のように定義する．

$$\boldsymbol{F_{LF}}(i,k) = \min \sum_{(u,v) \in p} \overline{\psi_{uv}} \qquad (7)$$
$$\textit{Subject to } \; p \in P_{i,k}$$

上式で，$\overline{\psi_{uv}}$はエッジ(u,v)の両端のノードにおいて，隣接するノードのグループ内でのラベルの発生頻度を表す．

二つ目は，ラベルの一致度に着目した LGW-Label Matching（LGW-LM）であり，以下のように定義する．

$$\boldsymbol{F_{LM}}(i,k) = \min \sum_{(u,v) \in p} \frac{1}{\chi(u,v)} \qquad (8)$$
$$\textit{Subject to } \; p \in P_{i,k}$$

上式で，$\chi(u,v)$はエッジ(u,v)の両端のノードにおいて，それぞれのノードと隣接するノードをグループとしたとき，隣接するグループ同士で一致しているラベルの数をカウントした値（一致度）である．

4.2. 評価実験

3.2 節と同様に，SVM 分類器による評価を行った．比較手法は，Fused Gromov-Wasserstein（FGW），Weisfeiler-Lehmen Kernel（WLK）Hadamard Code Kernel（HCK），Neighborhood Hash Kernel（NHK），Vertex Hash Kernel（VHK）を用いた．ノードラベルを持つグラフデータセットを使用し，3.2 節と同様の処理を行った．また，テストデータのみラベルの値を一つ巡回シフトし，トレーニングデータと異なるラベルの値となるようにした．データセットの分割，リラベル，トレーニング，テストを 10 回繰り返し，分類精度の平均と標準偏差を算出した．結果を表 2 に示す．

表 2 グラフのノードラベルを用いた平均分類精度

メソッド	データセット		
	MUTAG	ENZYMES	ER_MD
FGW	74.28±4.86	23.28±2.60	60.00±2.15
WLK	65.87±6.32	17.37±3.68	61.69±2.79
HCK	73.49±14.09	<u>27.67±4.31</u>	<u>62.70±2.21</u>
NHK	71.43±13.26	22.83±3.63	61.15±2.23
VHK	77.62±6.08	18.94±2.62	62.36±2.64
LGW-LF	**81.42±5.81**	**30.61±2.63**	**66.89±3.76**
LGW-LM	<u>80.16±4.67</u>	22.93±2.35	60.00±2.15

表 2 より，LGW-LF は分類精度が高く，ノードラベルを用いた手法として有効であることが分かった．一方，LGW-LM は MUTAG を除いて既存手法よりも低かった．これは，ラベルの一致度を用いたパラメータが入っており，純粋な構造分析ではなかったためであると考えられる．

5. むすび

本研究では，グラフ類似性測定精度を向上させることを目的に，GW 距離の距離尺度としてグラフの局所構造を使用した二つの手法を提案した．第一の手法として，グラフのエッジ構造を用いて構造的距離を表す EGW-DT と EGW-TW を提案し，グラフ分類による評価では，構造が複雑なデータの分類に適していることが分かった．第二の手法として，ラベルを用いたグラフの構造的類似性を表現する， LGW-LF と LGW-LM を提案した．グラフ分類による評価では，LGW-LF は従来の手法より精度が高く，ラベルの構造情報のみで分類できることが分かった．つまり，ラベルが異なる種類のグラフデータの分類にも適用でき，実用性が高いことを示している．一方，LGW-LM は既存手法の分類精度を下回っていた．今後の課題として，より大きなデータセットで実験する必要がある．また，実際にアプリケーションに適用し，その実用性を確かめる必要がある．

参考文献

[1] G. Peyr´e, M. Cuturi, and J. Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In Thirty-third International Conference on Machine Learning (ICML), 2016.

[2] C. Villani. Optimal transport: Old and new. Springer, New York, 2008.

[3] G. Peyr´e and M. Cuturi. Computational optimal transport. Foundations and Trends in Machine Learning, Vol. 11, No. 5-6, pp. 355–607, 2019.

[4] F. M´emoli. Gromov–Wasserstein distances and the metric approach to object matching. Found. Comput. Math., Vol. 11, pp. 417–487, 2011.

2020 年度

早稲田大学大学院基幹理工学研究科 情報理工・情報通信専攻　修士論文

グラフ類似性測定のための
グラフ局所構造を考慮したGromov-Wasserstein距離

Gromov-Wasserstein Distance Considering Graph Local Structure for Graph Similarity Measurement

浅見　莉絵子

(5119F001-5)

提出日：2021 年 1 月 25 日

指導教員：渡辺　裕

研究指導名：オーディオ・ビジュアル情報処理研究

目 次

図 目 次

表 目 次

第1章　序論

1.1　背景

機械学習で使用するデータの多くは，ツリー，グループ，クラスター，シーケンスなどの構造化データである．そして多くの構造化データはグラフ構造として表現することができる [1]．図 1.1 にグラフ構造の例を示す．グラフ構造はノードと，ノード間のつながりや経路を表すエッジから構成される．近年グラフ構造は，ソーシャルネットワークから医療工学 [2]，バイオインフォマティクス [3]，化学 [4]，および天文学まで，多くの分野で注目されている．例えばソーシャルネットワークでは，ユーザー同士の繋がりをグラフ構造で表すことができる．これは現実世界における学校の交友関係や企業の同僚との関係を表す．ソーシャルネットワークのグラフを分析することにより，ユーザーの分類，コミュニティの検出，友達の推薦などを行うことができる [5]．また化学における分子化合物では，各原子をノードとし，原子同士の結合をエッジとするグラフ構造表現によって記述できる．さらに，タンパク質の相互作用をグラフ構造で表現すると，ノードがタンパク質に対応し，エッジが相互作用するタンパク質間のリンクを表す．このように，グラフ構造表現は多くのアプリケーションで構造化データを記述する手段として有効であり，グラフ表現を用いることで数学的処理が可能となる．本論文では，グラフ構造を持ったデータをグラフデータと呼ぶ．

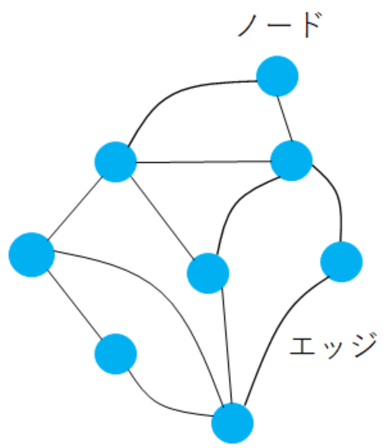


図 1.1: グラフ構造の例

近年，機械学習を用いたデータ解析が注目されているが，ベクトル値を持つデータに特化した既存の機械学習手法では，グラフデータに対するパフォーマンスが大幅に低下する場合がある．機械学習をグラフデータに対して効果的に適用するためには，機械学習アルゴリズムに対

してノードとエッジに関連付けられたグラフ構造を活用することが重要である．

グラフデータに対する機械学習，統計，コンピュータビジョンの基本的な問題を表す一例として，Graph Matching（GM）がある [6]．GM では，グラフデータのペア間でノードやエッジの不一致を最小限に抑えるような，最適なマッチングパラメータを探して調整する．これは，理論的にも実際的にも難しい問題であり，NP 完全問題 [7] としてよく知られている 2 次割り当て問題になる．Lawler QAP [8] と Koopmans Beckmann QAP [9] の二つはよく使われている定式化だが，計算が非常に複雑である．しかし GM 問題は，コンピュータビジョン [10]，ネットワーク分析 [11]，バイオインフォマティクス，生物学 [12]，データマイニングなどの多くのアプリケーションで頻繁に発生する．例として，マルウェアサンプルの検出が挙げられる [11]．グラフ構造表現でサンプルをコーディングすると，未知のサンプルに対する既知のマルウェアサンプルやクリーンサンプルとの比較を検出の難易度として，GM 問題と同様に再定式化できる．また，データベースの中の既知のタンパク質に対するグラフ表現されたタンパク質の分類では，実験で得られた新しいタンパク質を識別することができる．

上記のような GM 問題は，一般的に二つのカテゴリに分類される．一つ目は，1 対 1 マッピングとも呼ばれる Exact Graph Matching であり，ノードの数が同じ二つのグラフ間の厳密な対応が必要である．二つ目は Inexact Graph Matching で，多対多マッピングまたはベストグラフマッチングとも呼ばれる．Inexact Graph Matching では Exact Graph Matching が緩和され，複数のノードをマージすることで大きさの違うグラフをマッチングできるようになる．したがって，Inexact Graph Matching はノードの数が異なるグラフ間のマッチングに適用でき，より現実世界に近いデータに対応することができる．GM，グラフ分類，グラフクラスタリングなどのグラフベースの問題で扱うグラフ類似性を検討すると，ノード間のペアワイズ距離の類似性行列または共分散行列が，固有の構造情報を用いるために利用できる最も効果的な手法である．これらの手法は，着目するノードを絶対的な方法で表現するのではなく，残りのノードに対する相対的な方法で表現する．既存研究では，Random Walk [13, 14] および Shortest Path [15]，Subtrees [4, 14]，Cycles [16]，Graphlets [17]，Pyramid Matched Graphs [18] に基づくカーネルがあり，二つのグラフ構造化データ間の類似性を測定するグラフカーネルは，効果的なアプローチである [19]．さらに，Weisfeiler–Lehman 手法 [20] を使用してパフォーマンスを改善することができる．このアプローチの顕著な利点は，取得したカーネルを Support Vector Machine（SVM）などの既存の機械学習手法にプラグインできることである．他にも，グラフ畳み込みネットワークは，エンドツーエンドでグラフ上の機能の最適な組み合わせを学習する手法であり既存手法よりも優れている [1, 21–24]．また，最適輸送 [25, 26] に基づく手法では，グラフデータに確率測度を与える．例えば，最適輸送を用いた Gromov– Wasserstein (GW) 距離 [27–29] では，二つの異なる距離空間において，空間内のペアワイズ距離を用いて，二つの確率測度間の距離を計算する．しかし，いずれの手法でも完全な GM，グラフ分類，グラフクラスタリングを行うには至っていない．そこで本研究では GW に基づき，さらに構造の類似性を表す新たなパラメータを検討した．

1.2 目的

本研究の目的は，最適輸送に基づいたグラフ比較手法を拡張し，グラフ類似性測定精度を向上させることである．そのために，グラフのエッジ構造とノードラベルに着目した．

1.2.1 グラフのエッジ構造を用いた類似性測定

最適輸送に基づく手法の代表例である GW 距離は，さまざまなタイプのオブジェクトに適用できるため，多様な GW 距離の変形手法が提案され，多くの GM およびグラフ分類問題で既存手法を上回っている [30–34]．しかしこれらの GW 距離の変形手法は，グラフを含むオブジェクトの類似性を測定するのに有効であるが，実際のグラフ構造データセットの幾何学的特性を完全には反映していない．具体的には，グラフ内のノードの任意のペア間の最短経路距離や調和距離 [35] などの構造情報を組み込んでいるが，グラフの二つのノード間の構造的類似性と構造的同等性を完全には取り入れてない．そこで，本研究ではエッジ構造に着目した．エッジ構造をグラフ構造化データに組み入れるために，GW 距離に対する新しい手法を提案した．本提案手法を Edge-structure-enhanced Gromov-Wasserstein (EGW) 距離と呼ぶ．

1.2.2 グラフのノードラベルを用いた類似性測定

本研究では，グラフ類似性測定精度を向上させるもう一つの方法として，グラフのノードラベルに着目した．既存のグラフのラベルを用いた類似性測定手法には，Weisfeiler–Lehman Kernel [20]，Fused Gromov-Wasserstein [30]，Hadamard Code Kernel [36]，Neighborhood Hash Kernel [37, 38]，Vertex Histogram Kernel [39] などがある．これらの既存研究では比較するグラフ同士のラベルの種類が一致していることを前提としている．したがって，ラベル付きグラフを比較する際，比較するグラフのラベルのタイプが異なっている場合に分類精度が下がってしまうため，使用できるデータが限られる．そこで，本研究ではグラフの構造的類似性表現として，ラベルのタイプが異なるグラフに対しても適用することができる手法を提案した．本提案手法を Label-invariant Gromov-Wasserstein (LGW) 距離と呼ぶ．

1.3 論文の構成

本論文の構成は以下のとおりである．
第 1 章は，本章であり，本研究の背景，目的について述べる．本研究の目的は，最適輸送に基づいたグラフ比較手法を拡張し，グラフ類似性測定精度を向上させることである．そのために，グラフのエッジ構造とノードラベルに着目した．
第 2 章は，関連研究について述べる．最初に本研究で使用するグラフの用語を定義する．その後，本研究のベースとなる最適輸送と GW 距離を紹介する．さらに，最適輸送ベースの discrepancy に関する最先端の研究を紹介する．

第 3 章は，グラフのエッジ構造を用いた提案手法 (EGW) について述べる．初めに，GW discrepancy についての説明を示す．続いて，提案手法である EGW 距離について示し，提案手法によるグラフ分類の実験と結果，考察を示す．
第 4 章は，グラフのラベルを用いた提案手法 (LGW) について述べる．初めに，既存のラベルを用いたグラフの類似性測定手法と問題点について述べる．次に，提案手法である LGW 距離について示し，提案手法によるグラフ分類の実験と結果，考察を示す．
第 5 章は，本研究の結論について述べる．初めに本研究のまとめを示し，次に今後の課題について述べる．

第2章　関連研究

2.1　まえがき

本章では，最初に本研究で使用するグラフの用語を定義する．その後，本研究のベースとなる最適輸送と GW 距離を紹介する．さらに，最適輸送ベースの discrepancy に関する最先端の研究を紹介する．以下では，スカラーを小文字 $(a, b, \ldots)$ で表し，ベクトルを太字の小文字 $(\boldsymbol{a}, \boldsymbol{b}, \ldots)$ で表し，行列を太字の大文字 $(\mathbf{A}, \mathbf{B}, \ldots)$ で表す．

2.2　用語の定義

本節では，本研究で使用するグラフと関連する用語を定義する．

定義 1 (グラフ)**.** グラフは，n 個のノード $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ と m 個のエッジ $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ で構成される構造 $G = (\mathcal{V}, \mathcal{E})$ である．また，グラフ G が方向を示さないエッジのみで構成されえている場合，G は無向グラフである．

図 2.1 に無向グラフの例を示す．ノードの数は $|\mathcal{V}|$，エッジの数は $|\mathcal{E}|$ である． G に含まれる二つのノード $v_i, v_j \in \mathcal{V}$ がエッジ e で接続されるとき，エッジ e は e_{ij} と表記する．ここで，e_{ij} で接続される二つのノード v_i, v_j は「隣接している」と言う．また本研究では自己ループのない無向グラフのみを考慮する． i 番目のノードに接続するエッジの数を m_i として表すと，$|\mathcal{E}| = \sum_{i=1}^{n} m_i/2$ となり，これらのエッジのセットは $e_i = \{e_i^1, e_i^2, \ldots, e_i^{m_i}\}$ と表す．

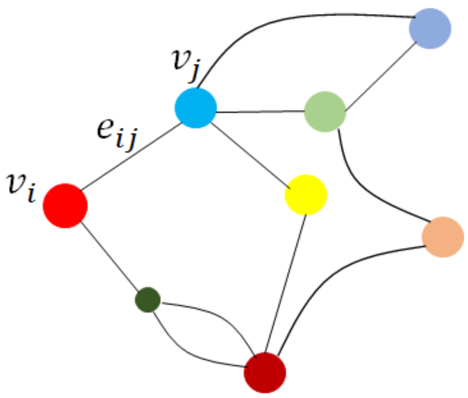


図 2.1: 無向グラフの例

定義 2 (Degree). 無向グラフ $G = (\mathcal{V}, \mathcal{E})$ とノード $v_i \in \mathcal{V}$ があるとき，v_i に接続するエッジ e_i の数 (degree) は，σ_i で表される．$\mathcal{N}(v_i)$ を v_i と隣接するノードのセットとすると，定義は次の通りである．

$$\sigma_i \quad = \quad |\{v_i : e_i \in \mathcal{E}\}| = |\mathcal{N}(v_i)|. \tag{2.1}$$

定義 3 (パス). グラフ $G = (\mathcal{V}, \mathcal{E})$ のノード $v_i \in \mathcal{V}$ から $v_j \in \mathcal{V}$ への経路 (walk) はノード $v_1, v_2, ..., v_{k+1}$ のシーケンスである．ここで，$1 \leq i \leq k+1$ であり，$v_i, v_{i+1} \in \mathcal{V}$ の場合，$1 \leq i \leq k$ である．walk の長さは，シーケンス内のエッジの数に等しくなる．つまり，上記の場合は k である． $v_i \neq v_j \Leftrightarrow i \neq j$ となる walk はパスと呼ばれる．隣接する二つのノード間のパスは，エッジに相当する．隣接していない二つのノード間のパスのセットは $\mathcal{P}_{i,j}$ と示される．$\mathcal{P}_{i,j}$ 内の一つのパスは $p_{i,j}$ と示される．また，$p_{i,j}$ の長さを $|p_{i,j}|$ と表記する．

定義 4 (Shortest Path). グラフ G のノード v_i から v_j までの Shortest Path(SP) $p^{\star}_{i,j}$ は， v_i から v_j へのパスにおいて，最も長さが短いパスである．

定義 5 (ラベル). グラフ G のノード v_i の属性をラベルで分類し，l_i で表す．

2.3 最適輸送

最適輸送 [25, 26] は，確率分布として有限次元のヒストグラムを使用する．最適輸送は，同じ空間上 (ground space と呼ぶ) または事前登録された複数の ground space 上で定義される．最適輸送は質量移動問題として解釈され，ある分布の質量を，総質量を変えずに別の分布に移動する時，最小のコストで移動するための最適な計画を求める．図 2.2 に最適輸送の概念図を示す．

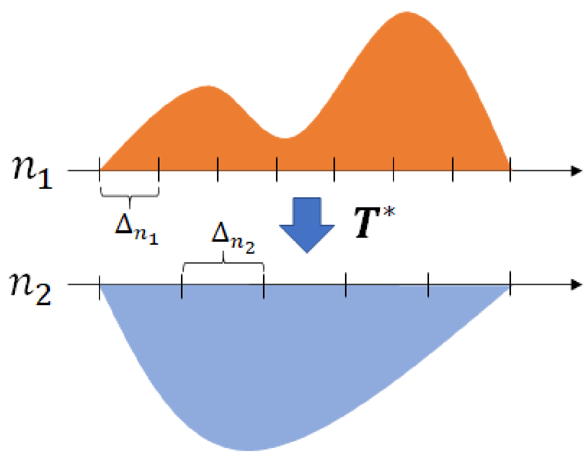


図 2.2: 最適輸送の概念図

同じ二つの距離空間 n_1，n_2 上のヒストグラムのシンプレックスを，それぞれ $\Delta_{n_1} = \{\boldsymbol{p} \in \mathbb{R}^{n_1}_{+}; \sum_i p_i = 1\}$，$\Delta_{n_2} = \{\boldsymbol{q} \in \mathbb{R}^{n_2}_{+}; \sum_j q_j = 1\}$ と定義する．次に，二つの確率測度を以下のよ

うに定義する．

$$\mu = \sum_{i=1}^{n_1} p_i \delta_{x_i}, \nu = \sum_{j=1}^{n_2} q_j \delta_{y_j}. \tag{2.2}$$

ここで，δ_{x_i} は x_i に対する Kronecker のデルタを表し，p_i は x_i に対するヒストグラムである．また，$x_i \neq x_j$ (ただし $i \neq j$) は，一般性を失うことなく想定される．また，輸送コスト行列 (ground cost matrix) $\mathbf{C} \in \mathbb{R}_+^{n_1 \times n_2}$ についても検討する．$\mathbf{C}(i,j)$ は，i 番目の要素と j 番目の要素の間の輸送コストを示す．次に，これら二つのヒストグラム間の最適輸送問題を以下のように定義する．

$$\mathbf{T}^*(\mathbf{C}, \boldsymbol{p}, \boldsymbol{q}) = \underset{\mathbf{T} \in \mathcal{U}_{n_1 n_2}}{\arg\min} \langle \mathbf{T}, \mathbf{C} \rangle. \tag{2.3}$$

ここで，$\mathcal{U}_{n_1 n_2}$ は

$$\mathcal{U}_{n_1 n_2} := \left\{ \mathbf{T} \in \mathbb{R}_+^{n_1 \times n_2} : \mathbf{T}\mathbf{1}_{n_2} = \boldsymbol{p},\ \mathbf{T}^T \mathbf{1}_{n_1} = \boldsymbol{q} \right\} \tag{2.4}$$

と表される．$\mathcal{U}_{n_1 n_2}$ は，$n_1 \times n_2$ 非負行列の超多面体 (polytope) を表し，行と列の周辺分布がそれぞれ p_i と q_j に等しくなる．この最小化問題は，線形計画問題と凸最適化問題である．したがって，線形計画問題の generic solver を使用して解くことができる．次に，$\mathcal{W}(\mu, \nu)$ と表される二つの確率測度間の Wasserstein 距離は，最適輸送計画 $\mathbf{T}^*$ の下で質量が移動した合計距離に等しくなる．さらに，エントロピー正則化された最適輸送問題の解であるエントロピー正則化項 $H(\mathbf{T}) = -\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \mathbf{T}(i,j)(\log(\mathbf{T}(i,j)) - 1)$ を追加する．これは，Sinkhorn の固定少数点反復 [40–42] を使用して効率的に解くことができる．空間に関する事前情報が分からない場合は，$\boldsymbol{p}$ と $\boldsymbol{q}$ をそれぞれ $\boldsymbol{p} = \frac{1}{n_1}\mathbf{1}_{n_1}$，$\boldsymbol{q} = \frac{1}{n_2}\mathbf{1}_{n_2}$ として一様分布に設定できる．

最適輸送は、低ランクの近似 [43] から辞書学習 [44]，ドメイン適応 [45]，クラスタリング [46]，半教師あり学習 [47] まで，さまざまな機械学習問題で注目されている．しかし，オブジェクトの特徴表現を比較するコスト関数にのみ依存しているため，オブジェクトの構造情報を活用できない．

2.4 Gromov-Wasserstein 距離

確率測度が異なる確率空間にある場合，2.3 節の最適輸送での距離は無効になる．そこで，Gromov-Wasserstein 距離 (GW 距離) [27,28,31] は，最適輸送を ground space が事前に整列および登録されていない場合に拡張する．その結果，GW 距離問題で得られる輸送行列は，二つのドメイン間の緩い結合と解釈できる．GW 距離の定義は，以下のように与えられる．

定義 6 (Gromov–Wasserstein distance [27,28,48])**.** (X, d_X, μ_X) と (Y, d_Y, μ_Y) を二つの *metric measure spaces* とする．ここで，(X, d_X) は *compact metric space* であり，μ_X は X の *Borel probability measure* である [1]．(Y, d_Y, μ_Y) も同様に定義される．そして，GW 距離は以下のように定義される．

$$\inf_{\pi \in \prod(\mu_X, \mu_Y)} \int_{X \times Y} \int_{X \times Y} L(x_i, y_j, x_k, y_l) \mathrm{d}\pi(x_i, y_j) \mathrm{d}\pi(x_k, y_l). \tag{2.5}$$

ここで， $L(x_i, y_j, x_k, y_l) = |d_X(x_i, x_k) - d_Y(y_j, y_l)|$ は損失関数であり，$\prod(\mu_X, \mu_Y)$ は μ_X と μ_Y を周辺分布として持つ $X \times Y$ 上の全ての確率測度のセットを表す．

GW 距離は，距離空間に対して最適輸送ベースの距離を与える．注目すべき点は，各ドメイン内のサンプルのペア間の距離が計算されることである．これらの距離を他の領域の距離と比較する．したがって，GW では異なる空間でのサンプルの距離を直接比較する必要はない．その結果，ターゲット空間は様々な次元を持つことができる．

2.5 最適輸送ベースのグラフ類似性測定手法の関連研究

本節では，グラフベースの問題と関連ドメインの最適輸送ベースの不一致 (discrepancy) を使った方法に関するいくつかの最先端の研究を紹介する．グラフカーネルについては [19]，グラフマッチングについては [6] を参照している．

定義 6 の GW 距離は，空間内のペアワイズ距離を計算することにより，二つの異なる距離空間における二つの確率測度間の距離を計算する．GW 距離の特徴は，異なる次元の空間にあるオブジェクトを比較できることである．GW 距離のフレームワークに基づいて構築された一部の手法では，幾何学的特性が最適化問題に追加されている．例えば Peyré らは，さまざまな実世界のネットワークデータセットを比較する GW discrepancy として，確率測度を備えた距離測度空間をモデル化した [29]．GW discrepancy は，エントロピー正則化とシンクホーンの予測を使用したアルゴリズムであり，ペアワイズ非類似度行列の discrepancy と重心を計算する [40, 41]． Vayer らは，構造化された情報だけでなく，学習フレームワークの特徴情報にも対応できるように，Fused Gromov-Wasserstein（FGW）と呼ばれる GW 距離の拡張を提案した [30]．ただし，異なるドメイン内のエンティティ間の距離が既知であるべきであるという前提があるため，アプリケーションが限定される． Xu らは，グラフマッチングとノード埋め込みのための Gromov–Wasserstein Learning（GWL）フレームワークを提案した [31]．これにより，各グラフのトポロジー構造だけが学習され，各グラフに対応させる．Xu らはさらに，グラフ分割とマッチングが考慮されるグラフ分割技術に基づく大規模なグラフのためのスケーラブルな GWL フレームワークを提案した [32]． また Maretic らは，グラフのラプラシアン行列に関する滑らかなグラフ信号の確率分布を利用することにより，同じサイズのグラフを比較するための最適輸送に基づくフレームワークを示した [49]．Maretic らはさらに，異なるサイズの非整列グラフを比較するための新しい方法を提案した [50]．Bunne らは GW 距離を使用して，次元の異なる空間や異なるデータタイプの空間など，比較できない空間全体で生成モデルを学習するアプローチを提案した [33]． 提案されたアプローチは，そのロバスト性を改善し公平な学習を保証することにより，さまざまな学習設定で GW 距離を使用できるようにした．Yan らは，エントロピー GW discrepancy にラベルデータを組み込むことにより，異なるドメイン間の意味の一貫性を維持することを提案した [34]．提案された半教師付きスキームは，同じラベルを持つターゲットサンプルと転送されたソースサンプルを強制的に類似した分布にする．

2.6　むすび

本章では，本研究で使用するグラフの用語について定義した．その後，本研究のベースとなる最適輸送と GW 距離を紹介した．また，最適輸送ベースの discrepancy を使った最先端の研究を紹介した．

第3章　Edge-structure-enhanced Gromov-Wasserstein（EGW）距離

3.1　まえがき

本章では，グラフ類似性測定精度を向上させるために考案した，Edge-structure-enhanced GW 距離 (EGW 距離) について記述する．EGW 距離では，GW discrepancy にエッジの幾何学的特性を適用するため，ノードの degree とエッジから構成される三角形の数を組み込んだ．本章では初めに，GW discrepancy についての説明を示す [29]．続いて，提案手法である EGW 距離について示し，提案手法によるグラフ分類の実験と結果，考察を示す．

3.2　Gromov-Wasserstein discrepancy

GW discrepancy [29] は，GW 距離 [28] を拡張したグラフ類似性測定手法である．二つの測定された類似度行列間の GW discrepancy は，以下のように定義される．

定義 7 (Gromov–Wasserstein discrepancy)**.** $(\mathbf{C}_1, \boldsymbol{p}) \in \mathbb{R}_+^{n_1 \times n_1} \times \Delta_{n_1}$ と $(\mathbf{C}_2, \boldsymbol{q}) \in \mathbb{R}_+^{n_2 \times n_2} \times \Delta_{n_2}$ がそれぞれ二つの測定された類似度行列を表すとする．次に，二つの測定された類似度行列間の GW discrepancy を $d_{\mathcal{GW}}(\mathbf{C}_1, \mathbf{C}_2, \boldsymbol{p}, \boldsymbol{q})$ とすると，GW discrepancy は次のように定義される．

$$d_{\mathcal{GW}}(\mathbf{C}_1, \mathbf{C}_2, \boldsymbol{p}, \boldsymbol{q}) = \min_{\mathbf{T} \in \mathcal{U}_{n_1 n_2}} E_{\mathbf{C}_1, \mathbf{C}_2}(\mathbf{T}). \tag{3.1}$$

ここで，行列 $\mathbf{T}$ は二つの空間の結合であり [26]，$E_{\mathbf{C}_1, \mathbf{C}_2}(\mathbf{T})$ は以下のように定義される．

$$E_{\mathbf{C}_1, \mathbf{C}_2}(\mathbf{T}) = \sum_{i,j,k,l} L(\mathbf{C}_1(i,k), \mathbf{C}_2(j,l)) \mathbf{T}_{i,j} \mathbf{T}_{k,l}. \tag{3.2}$$

ここで，$L(a,b)$ は a と b の discrepancy を測定するための損失関数を表す．損失関数には例えば，$L(a,b) = \frac{1}{2}|a-b|^2$ や，Kullback–Leibler divergence $L(a,b) = K(a|b) = a\log(a/b) - a + b$ がある．図 3.1 に GW discrepancy の概念図を示す．GW discrepancy が小さいことは類似度が高いことを意味する．

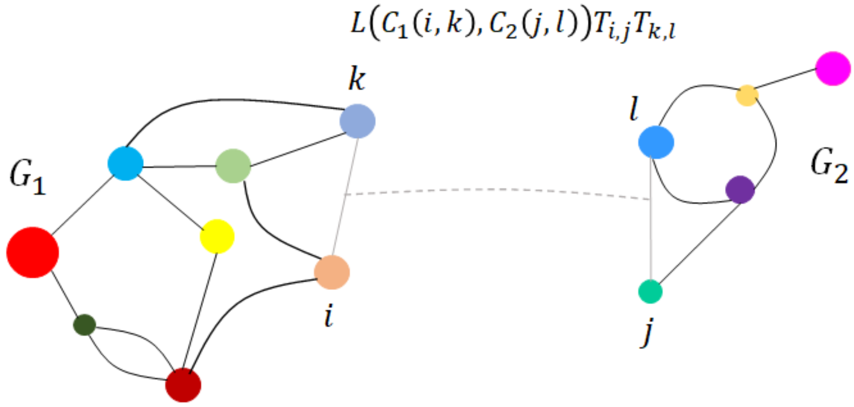


図 3.1: GW discrepancy の概念図

3.3 Edge-structure-enhanced Gromov–Wasserstein (EGW) 距離

本研究では，グラフのエッジ構造を用いて構造的距離を表した EGW 距離を提案する．本手法は 3.2 節の GW discrepancy に基づく．まず 3.3.1 項で提案手法の定式化をし，次にエッジ構造を用いた 2 種類の距離計算手法を記述する．

3.3.1 EGW 距離の定式化

各グラフ内のすべてのペアワイズ距離間の類似度を用いて構造間の類似性を表す 4 次元テンソル $L(\mathbf{C}_1(i,k), \mathbf{C}_2(j,l))$ [30] を利用し，以下のように EGW を定義する．

定義 8 (Edge-structure-enhanced Gromov-Wasserstein 距離)**.** 二つのグラフ G_1 ， G_2 を比較することを考える．$(\mathbf{D}_1, \boldsymbol{p}) \in \mathbb{R}_+^{n_1 \times n_1} \times \Delta_{n_1}$ と $(\mathbf{D}_2, \boldsymbol{q}) \in \mathbb{R}_+^{n_2 \times n_2} \times \Delta_{n_2}$ がそれぞれ，シンプレックスを含むエッジ構造が強化された二つの類似度行列を表すとする．ここで，これら二つの測定された類似度行列間の EGW である $d_{\mathcal{EGW}}(\mathbf{D}_1, \mathbf{D}_2, \boldsymbol{p}, \boldsymbol{q})$ は，次のように定義される．

$$d_{\mathcal{EGW}}(\mathbf{D}_1, \mathbf{D}_2, \boldsymbol{p}, \boldsymbol{q}) = \min_{\mathbf{T} \in \mathcal{U}_{n_1 n_2}} \sum_{i,j,k,l} |\mathbf{D}_1(i,k) - \mathbf{D}_2(j,l))| \mathbf{T}_{i,j} \mathbf{T}_{k,l}. \tag{3.3}$$

上式で， $\mathbf{D}_1(i,k)$ と $\mathbf{D}_2(j,l)$ はそれぞれ，G_1 と G_2 内のエッジ構造を考慮した距離を表し，次節で詳しく述べる．

EGW の定式化は GW の不一致の定式化に類似しているが，EGW はエッジ構造を用いた距離に依存している．また，以下の定理と証明を示す．

定理 1. $d_{\mathcal{EGW}}(\mathbf{D}_1, \mathbf{D}_2, \boldsymbol{p}, \boldsymbol{q})$ は距離である．

Proof. $d_{\mathcal{EGW}}(\mathbf{D}_1, \mathbf{D}_2, \boldsymbol{p}, \boldsymbol{q})$ は対称的である．つまり，$d_{\mathcal{EGW}}(\mathbf{D}_1, \mathbf{D}_2, \boldsymbol{p}, \boldsymbol{q}) = d_{\mathcal{EGW}}(\mathbf{D}_2, \mathbf{D}_1, \boldsymbol{q}, \boldsymbol{p})$ である．また，$d_{\mathcal{EGW}}(\mathbf{D}_1, \mathbf{D}_1, \boldsymbol{p}, \boldsymbol{p}) = 0$ である．そして，3.4 節で示されるように， $\mathbf{D}_1$ と $\mathbf{D}_2$ は距離の公理を満たすため， $d_{\mathcal{EGW}}(\mathbf{D}_1, \mathbf{D}_2, p, q)$ は距離の公理を満たす．証明は [29, 30] と類似しているため，本論文では完全な証明は省略されている． □

以降，これらの議論が明確である場合，簡単にするために，$d_{\mathcal{EGW}}(\mathbf{D}_1, \mathbf{D}_2, \boldsymbol{p}, \boldsymbol{q})$ を $d_{\mathcal{EGW}}(\mathbf{D}_1, \mathbf{D}_2)$ または $d_{\mathcal{EGW}}$ と表記する．

次に，式 (3.3) で提案されたエッジ構造の各長距離 D について詳しく述べる．本研究では，構造的な距離をより具体的に表現するために，二つのノード間のパスに沿ったエッジが含まれる三角形の数について考える．二つのノード間のパスに沿ったエッジを含む三角形の数が増えると，構造上の観点から構造距離が短くなることが予想される．ここで，次の定義を明示的に示す．

定義 9 (三角形とエッジ番号)**.** 無向グラフ $G = (\mathcal{V}, \mathcal{E})$ が与えられた場合，三角形は G で相互に隣接する三つのノードのセットである．次に，エッジ $e \in \mathcal{E}$ に着目するとき，エッジ e が成す三角形のエッジ番号は，e に対応する三角形の数と定義される．m 番目のエッジが成す三角形の数は $\tau(m)$ で表される．

図 3.2 に三角形とエッジ番号の例を示す．図 3.2 において，$\tau(m) = 2$ である．

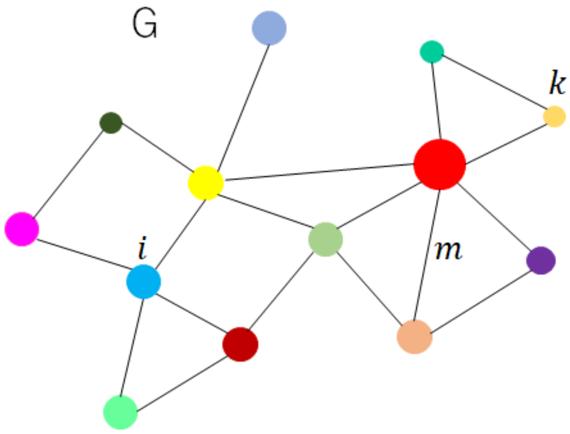


図 3.2: 三角形とエッジ番号の例

次に，i 番目と k 番目のノード間の距離 $\mathbf{D}(i, k)$ に，**定義 9** で定義された三角形とエッジ番号と， **定義 2** で定義された degree を組み込む．本研究では，二つのアプローチを検討した．一つ目は，グラフ内の二つのノードの局所的な構造を組み込んだ．二つ目は，二つのノード間のパスに沿ったエッジの構造を組み込んだ．

3.3.2 ノードベースの EGW 距離

本節では，二つのグラフの discrepancy を測定する，二つのノードのエッジに関する局所的な構造の違いを評価する提案手法について説明する．具体的には，ノードが持つエッジの数と，

そのエッジが持つ三角形の数を考慮する．エッジが持つ三角形の数に関しては，次式のように，グラフの i 番目のノードと k 番目のノードの間の正規化された差を考慮する．

$$\frac{1}{1+\tau_{\max}}\left|\sum_{s=1}^{m_i}\tau(e_i^s)-\sum_{t=1}^{m_k}\tau(e_k^t)\right|. \tag{3.4}$$

ここで，$\tau(e_i^s)$ は i 番目のノードに接続する s 番目のエッジの三角形の数を表す．したがって，第 1 の項 $\sum_{s=1}^{m_i}\tau(e_i^s)$ は，i 番目のノードに接続するすべてのエッジの三角形の総数を表す．2 番目の項も同様である．さらに， $\tau_{\max}$ は，グラフ内のエッジの三角形の最大数，つまり $\tau_{\max}=\max_{(i,j)\in\mathcal{E}}\tau(e_{i,j})$ である．またここでは，オーバーフローとゼロ徐算を避けるため，$(1+\tau_{\max})$ による除算が行われる．

次に，エッジの三角形の数と同様に，次式のような正規化された degree の差を考慮する．

$$\frac{|\sigma_i-\sigma_k|}{1+\sigma_{\max}}. \tag{3.5}$$

式 (3.5) で， $\sigma_{\max}$ はグラフの最大 Degree を表す．つまり，$\sigma_{\max}=\max_{i\in n=|\mathcal{V}|}\sigma_i$．式 (3.4) と式 (3.5) は，二つのエッジの局所的な構造特性が完全に同じである場合にゼロになる．したがって，discrepancy を測定するにはグローバルな違いが必要になる．そこで，**定義 4** の最短距離 $|p^\star|$ が考慮される．

提案手法では，式 (3.4) と 式 (3.5)， $|p^\star|$ を加算して一つの式にまとめ， $\mathbf{D}_{DT}$ として i 番目のノードと k 番目のノードの距離を次のように表す.

$$\mathbf{D}_{DT}(i,k) \quad = \quad |p_{i,k}^\star|+\frac{|\sigma_i-\sigma_k|}{1+\sigma_{\max}}+\ \frac{1}{1+\tau_{\max}}\left|\sum_{s=1}^{m_i}\tau(e_i^s)-\sum_{t=1}^{m_k}\tau(e_k^t)\right|. \tag{3.6}$$

ここで，degree と三角形 (triangle) を使用した EGW 距離を EGW-Degree Triangle (EGW-DT) とする．

さらに， 式 (3.4) と式 (3.5) の二つの距離のうち一つだけを考慮すると，それぞれ EGW-Triangle (EGW-T) および EGW-Degree (EGW-D) という EGW-DT の特別な変形が得られる．この距離を，以下のように定義する．

$$\mathbf{D}_D(i,k)=|p_{i,k}^\star|+\frac{|\sigma_i-\sigma_k|}{1+\sigma_{\max}}. \tag{3.7}$$

$$\mathbf{D}_T(i,k)=|p_{i,k}^\star|+\frac{1}{1+\tau_{\max}}\left|\sum_{s=1}^{m_i}\tau(e_i^s)-\sum_{t=1}^{m_k}\tau(e_k^t)\right|. \tag{3.8}$$

3.3.3 パスベースの EGW 距離

3.3.2 項の EGW-DT では，二つのノード間の局所的な構造の違いについて考慮したが，本節では，二つのノード間のパスに沿ったエッジの構造について考察する．

定義 3 の i 番目のノードと k 番目のノード間の全てのパス $\mathcal{P}_{i,k}$ を考慮して，各パス $p\in\mathcal{P}_{i,k}$ のすべてのエッジの三角形を列挙する．これは，エッジの冗長性またはパスの冗長性を表すとみ

なすことができる．したがって，この総数は，これら 2 つのノード間の相互の構造的近さを適切に表しているとみなすことができる．したがって，EGW-DT と同様にゼロ除算を考慮し，1 を足して逆数をとる．この数を用いて次のような最小化問題を解くことにより，新しい $\mathbf{D}_{TW}(i,k)$ を導出する．

$$\begin{aligned} \mathbf{D}_{TW}(i,k) &= \min \sum_{(u,v)\in p} \frac{1}{1+\tau(e_{u,v})} \\ \text{subject to} &\quad p \in \mathcal{P}_{i,k}. \end{aligned} \tag{3.9}$$

ここで，(u,v) は，パス p に沿った一つのエッジを表す．これは，i 番目のノードと k 番目のノード間の複数パス $\mathcal{P}_{i,k}$ の一つである． 式 (3.9) で定義された最小化問題は，一般的な重み付き最短経路問題である．したがってこの点は，ダイクストラのアルゴリズムなどの一般的なアルゴリズムを使用して解決できる．ここでは，三角形の重みを使用した EGW 距離を EGW-Triangle Weight (EGW-TW) とする．

3.4 $\mathbf{D}_{DT}$ 及び $\mathbf{D}_{TW}$ の距離の証明

本節では，提案手法が距離の公理を満たすことの証明をする．$\mathbf{D}(i,k)$ が最短経路距離である場合，EGW に距離空間を与えることができる．ここでは，$\mathbf{D}_D(i,k)$ と $\mathbf{D}_T(i,k)$ の証明が $\mathbf{D}_{DT}(i,k)$ に類似しているため，$\mathbf{D}_{DT}(i,k)$ の証明について説明する．ここに，以下の定理を与える．

定理 2. 式 *(3.6)* で定義された $\mathbf{D}_{DT}(i,k)$ は距離である．

Proof. $\mathbf{D}_{DT}(i,j) \geq 0$ と $\mathbf{D}_{DT}(i,i) = 0$ は定義から明らかである．$\mathbf{D}_{DT}(i,k)$ は定義から $\mathbf{D}_{DT}(k,i)$ に等しいため， $\mathbf{D}_{DT}(i,k)$ は対称的である．したがって，$\mathbf{D}_{DT}(i,k)$ が三角不等式を満たすことを示すことにより，距離の公理を満たすことの証明とする． 式 (3.6) より，

$$
\begin{aligned}
&\mathbf{D}_{DT}(i,k) \\
= \;& |p^{\star}_{i,k}| + \frac{|\sigma_i - \sigma_k|}{1+\sigma_{\max}} + \frac{1}{1+\tau_{\max}} \left| \sum_{s=1}^{m_i} \tau(e_i^s) - \sum_{t=1}^{m_k} \tau(e_k^t) \right| \\
= \;& |p^{\star}_{i,k}| + \frac{|\sigma_i - \sigma_h - (\sigma_k - \sigma_h)|}{1+\sigma_{\max}} \\
& + \frac{1}{1+\tau_{\max}} \left| \sum_{s=1}^{m_i} \tau(e_i^s) - \sum_{u=1}^{m_h} \tau(e_h^u) \right. \\
& \left. - \left(\sum_{t=1}^{m_k} \tau(e_k^t) - \sum_{u=1}^{m_h} \tau(e_h^u) \right) \right| \\
\leq \;& |p^{\star}_{i,k}| + \frac{|\sigma_i - \sigma_h| + |(\sigma_k - \sigma_h)|}{1+\sigma_{\max}} \\
& + \frac{1}{1+\tau_{\max}} \left[\left| \sum_{s=1}^{m_i} \tau(e_i^s) - \sum_{u=1}^{m_h} \tau(e_h^u) \right| \right. \\
& \left. + \left| \sum_{t=1}^{m_k} \tau(e_k^t) - \sum_{u=1}^{m_h} \tau(e_h^u) \right| \right] \\
\leq \;& |p^{\star}_{i,h}| + |p^{\star}_{k,h}| + \frac{|\sigma_i - \sigma_h| + |\sigma_k - \sigma_h|}{1+\sigma_{\max}} \\
& + \frac{1}{1+\tau_{\max}} \left[\left| \sum_{s=1}^{m_i} \tau(e_i^s) - \sum_{u=1}^{m_h} \tau(e_h^u) \right| \right. \\
& \left. + \left| \sum_{t=1}^{m_k} \tau(e_k^t) - \sum_{u=1}^{m_h} \tau(e_h^u) \right| \right] \\
= \;& |p^{\star}_{i,h}| + \frac{|\sigma_i - \sigma_h|}{1+\sigma_{\max}} + \frac{1}{1+\tau_{\max}} \left| \sum_{s=1}^{m_i} \tau(e_i^s) - \sum_{u=1}^{m_h} \tau(e_h^u) \right| \\
& + |p^{\star}_{k,h}| + \frac{|\sigma_k - \sigma_h|}{1+\sigma_{\max}} + \frac{1}{1+\tau_{\max}} \left| \sum_{t=1}^{m_k} \tau(e_k^t) - \sum_{u=1}^{m_h} \tau(e_h^u) \right| \\
= \;& \mathbf{D}_{DT}(i,h) + \mathbf{D}_{DT}(k,h).
\end{aligned}
$$

最初の不等式は，Cauchy–Schwarz 不等式を使用した．二番目の不等式は，最短経路の定義に由来する．したがって，式 (3.6) で定義された $\mathbf{D}_{DT}(i,k)$ は，三角不等式を満たす．これにより，距離の公理を満たす． □

$\mathbf{D}_{TW}(i,k)$ について，ここに以下の定理を与える．

定理 3. 式 *(3.9)* で定義された $\mathbf{D}_{TW}(i,k)$ は距離である．

定理 3 は， $\mathbf{D}_{TW}(i,k)$ の定義から明らかであるため，証明は省略する．

3.5 評価実験

3.5.1 実験準備

比較手法：提案手法の有効性を評価するために，k-NN 分類器と SVM 分類器によるグラフ分類問題を行った．比較手法には，最適輸送に基づく最先端のフラームワークである GW [29] と FGW [30] を用いた．さらに，最先端のグラフカーネルである Shortest Path Kernel（SPK）[15]，および Random Walk Kernel（RWK） [13,14]，Pyramid Matched Kernel（PMK）[18] と比較した．また，グラフニューラルネットワークフレームワークである PSCN [24] も用いる．

データセット：グラフ分類実験にグラフのベンチマークデータセット[1]を使用する．使用するデータセットの詳細については，表 3.1 に示す．

表 3.1: EGW の評価実験で用いたデータセットの詳細

データセット	プロパティ				ラベル		カテゴリ
	グラフ数	クラス数	平均ノード数	平均エッジ数	ノード	エッジ	
AIDS	2000	2	15.69	16.20	+	+	小分子化合物
MUTAG	188	2	17.93	19.79	+	+	小分子化合物
PTC_MR	344	2	14.29	14.69	+	+	小分子化合物
BZR	405	2	35.75	38.36	+	-	バイオインフォマティクス
COX2	467	2	41.22	43.45	+	-	バイオインフォマティクス
ENZYMES	600	6	32.63	62.14	+	-	バイオインフォマティクス
IMDB-B	1000	2	19.77	96.53	-	-	ソーシャルネットワーク
IMDB-M	1500	3	13.00	65.94	-	-	ソーシャルネットワーク
HIGHSCHOOL	180	2	52.32	544.81	+	-	ソーシャルネットワーク
INFECTIOUS	200	2	50	459.72	+	-	ソーシャルネットワーク
MIT	97	2	20	1469.15	+	-	ソーシャルネットワーク
OHSU	79	2	82.01	199.66	+	-	ブレインネットワーク
PEKING	85	2	39.31	77.35	+	-	ブレインネットワーク
MSRC-9	221	8	40.58	97.94	+	-	コンピュータビジョン
MSRC-21c	209	20	40.28	96.60	+	-	コンピュータビジョン

AIDS データセットは分子化合物を表すグラフで構成されており，グラフは活性化化合物のエイズ抗ウイルススクリーニングデータベースから作成されている．原子をノードとし，共有結合をエッジとして表す [51]．MUTAG データセットは，分子化合物を表すグラフで構成されており，細菌に対する変異原性の影響に応じてクラス分類されている [52,53]．PTC_MR データセットは雄のマウスの発がん性に依って識別された化合物をグラフで表したデータセットである．

[1] `https://chrsmrrs.github.io/datasets/`

ノードは原子を表し，エッジは化学結合を表す [53]．BZR データセットはベンゾジアゼピン受容体をグラフ表現したデータセットである [54]．COX2 データセットはシクロオキシゲナーゼ抑制剤をグラフ表現したデータセットである [54]．ENZYMES データセットは，タンパク質を表したデータセットであり，タンパク質の二次構造要素がノードを表し，アミノ酸配列に沿ってエッジが設定されている [15]．IMDB-B データセットは、IMDB(Internet Movie Database) の映画に登場する 1000 人の俳優または女優のエゴネットワーク (個人を中心に集められたデータ) で構成される映画コラボレーションデータセットである．各グラフのノードは俳優または女優を表し，二つのノード間のエッジは，それらが同じ映画に登場することを示す．IMDB-B データセットのグラフは，アクションとロマンスのジャンルから成る [55]．IMDB-M データセットは，Sci-Fi，ロマンス，コメディの三つのジャンルから生成されており，IMDB-B データセットに類似している [55]．HIGHSCHOOL データセットは連絡先のネットワークであり，7 日間の 20 秒間隔での学生間の相互作用を表す．INFECTIOUS データセットは，展示会の訪問者間の対面の接触を表す [56, 57]．MIT データセットは，マサチューセッツ工科大学の学生間の相互作用である [56, 58]．OHSU データセットと PEKING データセットは，機能的なブレインネットワークデータであり，各ノードは関心領域（ROI）に対応し，各エッジは脳の fMRI アトラスからの二つの ROI 間の相関を示す． OHSU データセットは，ハイパーアクティブインパルス分類タスク用，PEKING データセットは性別分類タスク用に構築されている [59]．MSRC-9 および MSRC-21c データセットは，セマンティック画像処理用である．各画像は，条件付きマルコフ確率グラフで表される．頂点は画像をセグメント化することによって導出され，各頂点は一つのスーパーピクセルを表す [60]．

手順：実験では，各データセットを 2 対 1 でランダムに分割し，トレーニングデータセットとテストデータセットを作成した．データセットの分割は乱数を使用して行い，使用した乱数はすべての手法で同じである．図 3.3 に実験手順を示し，表 3.2 に使用した乱数の値を示す．

表 3.2: EGW の評価実験で用いたパラメータ

乱数	42	44	51	49	89	45	80	65	5	54

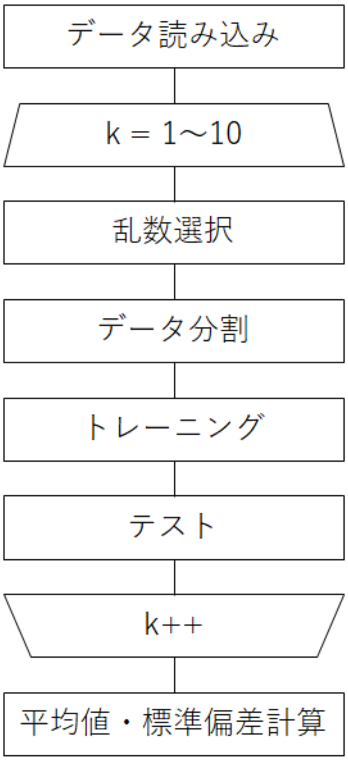


図 3.3: 実験手順

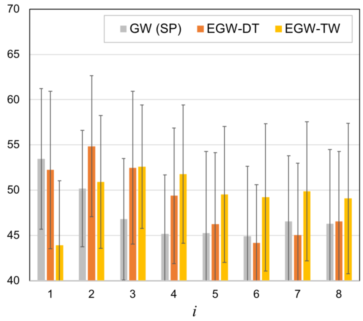
実装：提案手法の EGW は Python で実装された．SVM 分類器には scikit-learn toolbox[2]を使用する [61]．GW と FGW のコードは著者の web サイト[3]から取得できる．比較手法のグラフカーネルのソースコードは，有名なグラフカーネルを提供するライブラリである Grakel[4] [62] を使用する．PSCN のコードは web サイト[5]から取得できる．

3.5.2 k-NN 分類器を用いたグラフ分類

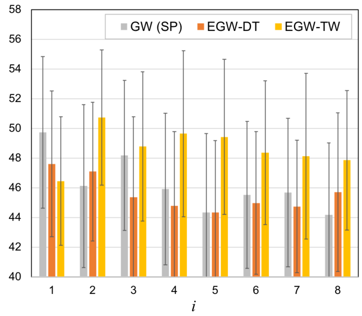
初めに，式 (3.3) で提案された $d_{\mathcal{EGW}}$ 距離の有効性を直接評価するため，k 近傍 (k-NN) 分類器を使用して，比較的小さいサイズのデータセットでグラフ分類の予備実験を行った．この比較では，ベースライン方法を Shortest Path（SP）を用いた式 (3.1) の GW 距離 $d_{\mathcal{GW}}$ とする．k-NN 分類器では，$k = \{1, 3, 5, 7, 9, 11, 13, 15\}$ を使用した．使用するデータセットは，トレーニングデータとテストデータにランダムに 50 回分割し，標準偏差を使用して平均分類精度を評価した．結果を図 3.4 に示す．ここで，i は $k = 2(i-1)+1$ であり k に対応する．図 3.4 から，提案された EGW-TW および EGW-DT は，MIT，HIGHSCHOOL，および OHSU データセットの全ての k 値にわたって，GW よりも安定して分類精度が高かった．EGW-DT は PEKING データセットで他の手法より高い精度であり，EGW-TW は INFECTIOUS データセットで安

[2] `https://scikit-learn.org/`
[3] `https://github.com/tvayer/FGW`
[4] `https://github.com/ysig/GraKeL`
[5] `https://github.com/tvayer/PSCN`

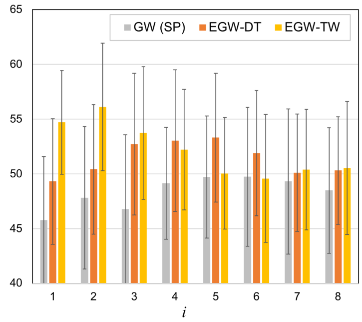
定して GW より良い精度であった．本実験では，EGW 距離はグラフ分類問題に有効であると言える．



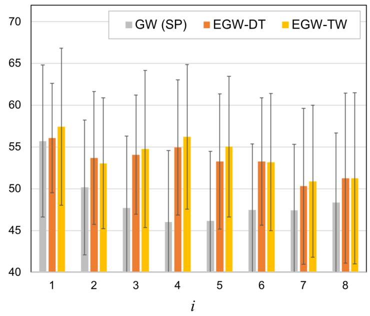
(a) MIT



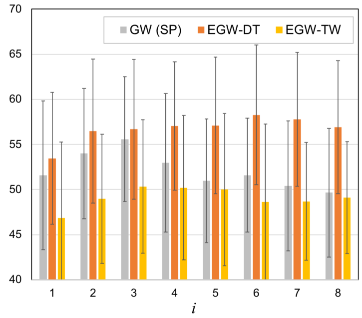
(b) INFECTIOUS



(c) HIGHSCHOOL



(d) OHSU



(e) PEKING

図 3.4: k-NN 分類器を用いた分類結果

3.5.3　SVM 分類器を用いたグラフ分類実験

本実験の目的は，比較的大きなサイズのデータセットを用いて，提案された EGW 距離の実際的有効性を明らかにすることである．本実験では，グラフ分類問題で多く使用されているベンチマークデータセットを用い，カテゴリ別に結果を示す．小分子化合物カテゴリのデータセットは AIDS，MUTAG，PTC_MR を用いた．バイオインフォマティクスカテゴリのデータセットは BZR，COX2，ENZYMES を用いた．最後に，ソーシャルネットワークカテゴリのデータセットは IMDB-B，IMDB-M を用いた．そして，SVM 分類器によって N 個のグラフのグラフ分類を行った．実験では，$N \times N$ カーネル行列を作成し，(p,q) 要素は $\exp(-\gamma \cdot d_{\mathcal{EGW}}(\mathbf{D}_p, \mathbf{D}_q))$ によって計算される．ここで、$\gamma \in \mathbb{R}$ はハイパーパラメータであり，$d_{\mathcal{EGW}}$ は二つのグラフ G_p，G_q 間の式 (3.3) の EGW 距離である．γ の値は，$\{0.1, 1, 10, 100\}$ に設定した．比較手法は，SPK [15]，および RWK [13,14]，PMK [18]，GW である．SVM 分類器の設定は，比較するすべての手法で同じであり，手法ごとに図 3.3 の実験を行い，分類精度の平均値と標準偏差を算出した．データセットの分割に用いた乱数は，表 3.2 と同様である．

表 3.3 に小分子化合物カテゴリのデータセットを分類した結果を示す．表 3.4 にバイオインフォマティクスカテゴリのデータセットを分類した結果を示す．表 3.5 にソーシャルネットワークカテゴリのデータセットを分類した結果を示す．結果は，平均値 $\pm$ 標準偏差として数値を示した．最も良い精度を太線，二番目に良い精度を下線で示す．表 3.3，表 3.4 より，小分子化合物とバイオインフォマティクスのデータセットでは BZR データセットを除き，提案手法は既存手法よりも分類精度が低かった．表 3.5 より，IMDB-B データセットでは，EGW-TW の分類精度が，特に γ が高い場合で，他の手法よりも高かった．IMDB-M データセットでは，EGW-DT の分類精度が最も高かった．

分類結果に違いが出た理由として，グラフの複雑さが関係していると考えられる．表 3.1 より，小分子化合物やバイオインフォマティクスのデータセットは平均ノード数と平均エッジ数が同等のため，一つのノードに対して一つまたは二つ程度のエッジ数のグラフが考えられ，提案手法で使用した三角形があまり含まれていなかったことが考えられる．一方ソーシャルネットワークのデータセットはノードの数に対してエッジの数が多く，複雑なグラフになっている．つまり三角形が多く含まれているため，提案手法が有効であったと考えられる．したがって，提案手法は構造が複雑なグラフに対して有効であることが分かる．

表 3.3: 小分子化合物データセットの分類精度 (± : 標準偏差)

メソッド		データセット		
	γ	AIDS	MUTAG	PTC_MR
SPK	–	99.74±0.15	86.51±4.09	52.37±3.33
RWK	–	99.68±0.13	**87.94±4.44**	52.28±4.61
PMK	–	**99.76±0.15**	83.17±5.73	49.30±4.85
GW (SP)	0.1	93.38±0.85	78.41±4.27	56.58±4.43
	1.0	95.86±0.41	81.11±3.79	53.94±5.03
	10	84.91±1.75	77.62±3.36	**57.28±3.40**
	100	84.17±1.75	76.19±3.76	57.19±3.37
EGW-DT	0.1	92.36±0.42	78.57±4.83	54.56±3.37
	1.0	95.15±0.41	81.26±4.36	51.14±3.84
	10	85.24±1.69	76.51±3.31	57.19±2.74
	100	84.91±1.66	75.71±3.41	57.19±2.74
EGW-TW	0.1	93.73±0.84	78.41±4.27	56.93±5.12
	1.0	95.86±0.63	81.11±3.79	54.12±4.77
	10	84.89±1.74	77.62±3.36	56.84±3.57
	100	84.33±1.66	76.19±3.76	56.93±3.47

表 3.4: バイオインフォマティクスデータセットの分類精度 (± : 標準偏差)(* : 実験ができなかったデータ)

メソッド		データセット		
	γ	BZR	COX2	ENZYMES
SPK	–	<u>80.90±2.34</u>	**76.58±3.84**	27.53±3.88
RWK	–	78.88±2.71	*	17.98±1.88
PMK	–	75.87±2.85	71.61±3.55	28.54±2.73
GW (SP)	0.1	78.13±2.69	74.39±3.46	22.83±2.15
	1.0	80.07±1.83	73.10±2.72	**31.11±2.12**
	10	78.66±2.06	75.81±4.11	20.96±4.28
	100	78.51±2.18	75.81±4.00	14.75±1.76
EGW-DT	0.1	74.93±2.39	71.48±3.45	22.47±2.26
	1.0	**81.12±2.43**	74.06±1.77	<u>31.01±1.58</u>
	10	78.66±1.86	<u>75.81±3.92</u>	19.90±3.98
	100	78.21±2.08	75.23±3.32	14.29±1.69
EGW-TW	0.1	77.91±2.55	72.65±2.47	20.56±2.06
	1.0	80.00±1.73	73.74±3.39	29.44±2.25
	10	78.66±2.06	75.81±4.11	30.96±2.75
	100	78.51±2.18	75.81±4.00	15.40±1.84

表 3.5: ソーシャルネットワークデータセットの分類精度 (±：標準偏差)

メソッド		データセット	
	γ	IMDB-B	IMDB-M
SPK	–	50.36±2.06	34.12±2.76
RWK	–	48.15±3.38	30.12±1.45
PMK	–	64.27±1.99	43.37±1.92
GW (SP)	0.1	57.36±1.73	38.28±0.95
	1	56.03±2.09	41.54±1.96
	10	63.45±1.55	44.81±1.49
	100	59.39±3.67	<u>46.81±1.84</u>
EGW-DT	0.1	57.64±1.89	39.90±0.90
	1.0	61.70±2.20	45.68±1.64
	10	64.24±0.94	41.81±2.00
	100	58.24±3.48	**46.93**±**1.82**
EGW-TW	0.1	61.85±3.18	33.92±3.54
	1.0	62.39±1.53	40.79±1.21
	10	<u>66.82±2.68</u>	43.94±2.66
	100	**66.94**±**1.65**	45.82±1.63

3.5.4　グラフの属性情報を考慮した分類実験

本節では，提案された距離が各ノードの属性データと一緒に適用されたときの有効性を評価する．本実験で使用するデータセットは，OHSU，PEKING，MSRC-21c，MSRC-9 であり，グラフに含まれるノード数に対してエッジ数が多く，ラベルデータを持つ．本実験では，EGW 距離は構造情報だけでなく属性情報も処理する FGW [30] フレームワークに基づいて構築される．属性情報と構造情報の重みを同等にするため，α に 0.5 を設定した．さらに，γ は 1 に設定した．比較手法には，3.5.3 項の実験での SPK，RWK，PMK に加えて，Weisfeiler Lehman Kernel（WLK）[20]，PSCN [24] が使用した．WLK は，height parameter h は $\{1, 2\}$ で実験する．PSCN の receptive field size parameter k は $\{5, 10\}$ に設定し，width parameter w は 18，batch size は 32 である．さらに，FGW と提案手法では，Weisfeiler Lehman labeling scheme [20] を用いない場合と，$h = \{2\}$ として適用した場合で実験した．3.5.3 項の実験と同様に，各手法ごとに図 3.3 の実験を行い，分類精度の平均値と標準偏差を算出した．データセットの分割に用いた乱数は，表 3.2 と同様である．表 3.6 に結果を示す． 表 3.6 において，平均値 $\pm$ 標準偏差として数値を示した．最も良い精度を太字，二番目に良い精度を下線で示す．

表 3.6 より，EGW は MSRC-21c データセットでは他の手法よりも劣っているが，他のデータセットでは安定して優れた精度を出している．したがって，提案された EGW は，属性データに基づく距離と組み合わせた場合でも有効であることが分かる．

表 3.6: 属性情報を持つグラフの分類精度 (±：標準偏差)(*：実験ができなかったデータ)

メソッド (parameter)	データセット OHSU	PEKING	MSRC-21c	MSRC-9
SPK	42.59±5.30	52.07±5.66	<u>80.43±3.56</u>	<u>90.55±3.38</u>
RWK	*	*	6.96±2.23	12.19±4.60
PMK	54.07±6.87	54.48±6.13	**81.74±2.17**	89.86±2.68
WLK(1)	51.85±6.42	55.86±4.83	4.83±5.10	86.71±4.59
WLK(2)	52.96±8.77	55.17±4.88	73.48±5.02	86.58±4.54
PSCN(5)	45.93±11.01	52.41±12.6	13.33±3.93	12.47±3.80
PSCN(10)	48.52±7.30	47.24±5.98	13.33±3.93	12.47±3.80
FGW	49.63±10.50	53.10±3.84	68.99±4.01	87.81±4.22
FGW(2)	54.07±7.07	52.76±5.78	60.58±6.64	87.95±4.28
EGW-DT	55.19±10.27	<u>58.97±5.66</u>	51.16±5.19	86.85±5.49
EGW-DT(2)	<u>55.56±8.28</u>	**59.66±6.91**	44.20±7.65	85.89±3.73
EGW-TW	**57.78±12.19**	52.76±5.98	73.62±3.65	89.32±3.91
EGW-TW(2)	54.44±9.08	53.45±4.94	64.06±5.05	**91.10±3.93**

3.6 むすび

本章では，グラフのエッジ構造を考慮した距離である EGW を提案した．EGW では，ノードの周辺構造に着目した手法である EGW-DT と，二つのノード間のパスの構造に着目した EGW-TW を提案した．EGW を使用して SVM によるグラフ分類を行った結果，構造が複雑なグラフに対して分類精度が高いことが分かった．

第4章　Label-invariant Gromov-Wasserstein (LGW) 距離

4.1　まえがき

本章では，グラフのノードラベルを用いたグラフ類似性測定精度を向上させるために考案した，Label-invariant GW 距離 (LGW 距離) について記述する．LGW 距離では，GW discrepancy にノードラベル情報を適用するため，ノードラベルの発生頻度と一致度を組み込んだ．本章では初めに，既存のラベルを用いたグラフの類似性測定手法と問題点について述べる．次に，提案手法である LGW 距離について示し，提案手法によるグラフ分類の実験と結果，考察を示す．

4.2　グラフのノードラベルの一致性を用いたグラフ分類方法と問題点

4.2.1　従来のノードラベルを用いた手法と課題

グラフのノードラベルを用いたグラフの類似性測定手法の代表的な手法として，Weisfeiler-Lehman Kernel（WLK）[20] がある．WLK では，ノードラベルを決められた法則で再度ラベリング (リラベル) していき，それを一定回数繰り返したとき，二つのグラフでラベルが一致していれば二つのグラフは類似している，と判定する手法である．また最先端の手法である FGW は，GW にノードの属性情報を組み込むことで，ラベルを持つグラフデータの類似性測定を行う．FGW [30] では，比較する二つのグラフの類似性を見るためにノードラベルのハミング距離を使用している．他にも，ノードラベルを用いたグラフカーネルとして，Hadamard Code Kernel (HCK) [36]，Neighborhood Hash Kernel (NHK) [37, 38]，Vertex Histogram Kernel (VHK) [39] がある．

これらの手法は，ノードラベルの値そのものを比較するため，比較するグラフデータがもつラベルのタイプが一致している必要があり，利用できるアプリケーションが限られてしまう．例えば，トレーニングデータのラベルとして $\{0, 1, 2\}$ を使い，テストデータのラベルとして $\{a, b, c\}$ を使っている場合，パフォーマンスが劣化することが考えられる．次の項でパフォーマンスの劣化について検証する．

4.2.2 従来手法の検証実験

本節では，先行研究の検証のために行った SVM 分類器によるグラフ分類実験について述べる．トレーニングデータとテストデータのラベルが同じ場合と，テストデータのみラベルを入れ替えた場合で，分類精度に違いが出るかどうかを検証する．使用するグラフカーネルは，FGW を用いたカーネル，および WLK，HCK，NHK，VHK である．FGW は Weisfeiler-Lehman ラベル [20] を用いない場合と，$h = \{2\}$ と適用した場合で実験した．データセットは MUTAG [52,53]，ENZYMES [15]，ER-MD [53, 54] を用いた．データセットの詳細を，表 4.1 に示す．

表 4.1: 先行研究の検証実験で用いたデータセットの詳細

データセット	プロパティ					カテゴリ
	グラフ数	クラス数	平均 ノード数	平均 エッジ数	ラベルの種類数	
MUTAG	188	2	17.93	19.79	7	小分子化合物
ENZYMES	600	6	32.63	62.14	3	バイオインフォマティクス
ER-MD	446	2	21.33	234.85	10	小分子化合物

MUTAG データセットは，分子化合物で構成されており，細菌に対する変異原性の影響に応じてクラス分類されている．ENZYMES データセットは，タンパク質を表したデータセットであり，タンパク質の二次構造要素がノードを表し，アミノ酸配列に沿ってエッジが設定されている．ER-MD データセットは，3D 座標が付属した分子化合物データセットである．まず，手法ごとに図 3.3 の実験を行い，平均分類精度と標準偏差を算出した．次に，図 4.1 に示すように，トレーニングデータに対してテストデータのみラベルの値を変え，トレーニングデータと異なるラベルの値となるようにして実験を行った．ここで，ラベルは整数値を使用し，テストデータのリラベルはラベルの値を 1 だけ巡回シフトした．また，データセットの分割に用いた乱数は，表 3.2 と同様である．

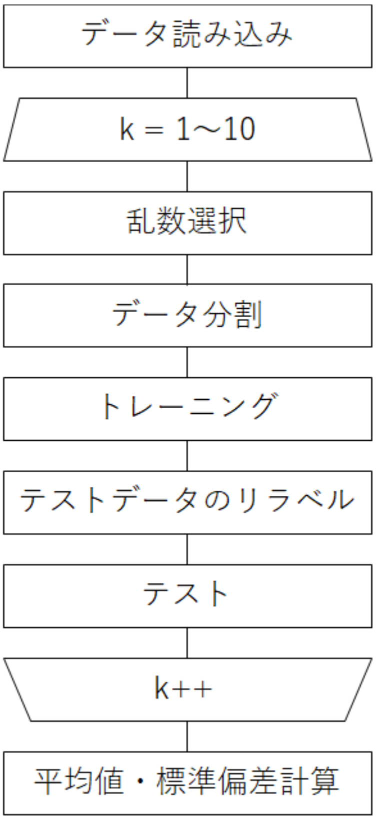


図 4.1: 実験手順

各データセットについてリラベル前とリラベル後の分類結果を表 4.2 に示す．表 4.2 において，平均値 ± 標準偏差として数値を示した．

表 4.2 より，先行研究ではテストデータのラベルを入れ替えると 10 ポイント前後で精度が落ちることが分かる．

表 4.2: ラベルに依る分類精度 (± : 標準偏差)

メソッド	データセット					
	MUTAG		ENZYMES		ER-MD	
	original	relabeled	original	relabeled	original	relabeled
FGW	84.44±5.21	74.28±4.86	39.34±2.40	23.38±2.60	68.65±5.23	60.00±2.15
FGW(2)	85.87±4.73	78.57±4.61	48.08±3.75	25.91±3.10	67.97±2.72	60.00±2.15
WLK	73.97±7.45	65.87±6.32	27.63±3.00	17.37±3.68	67.70±2.78	61.69±2.79
HCK	85.71±4.54	73.49±14.09	46.66±3.00	27.67±4.31	67.50±2.46	62.70±2.21
NHK	83.92±5.28	71.43±13.26	35.61±2.01	22.83±3.63	65.54±1.93	61.15±2.23
VHK	85.08±5.08	77.62±6.08	21.41±2.64	18.94±2.62	68.24±2.98	62.36±2.64

4.3 Label-invariant Gromov-Wasserstein (LGW) 距離

4.2.2 項より，既存手法ではトレーニングデータに対してテストデータのラベルを入れ替えたときに分類精度が悪くなることが分かった．そこで本研究では，ノードラベルの種類の変化にロバストな手法を提案する．本手法は 3.2 節の GW discrepancy に基づく．まず 4.3.1 項で提案手法の定式化をし，次にラベルを用いた 2 種類の計算方法を記述する．

4.3.1 LGW 距離の定式化

各グラフ内のすべてのペアワイズ距離間の類似度を用いて構造間の類似性を表す 4 次元テンソル $L(\mathbf{C}_1(i,k), \mathbf{C}_2(j,l))$ [30] を利用し，以下のように LGW を定義する．

定義 10 (Label based Gromov-Wasserstein distance)**.** 二つのグラフ G_1，G_2 を比較することを考える．$(\mathbf{F}_1, \boldsymbol{p}) \in \mathbb{R}_+^{n_1 \times n_1} \times \Delta_{n_1}$ と $(\mathbf{F}_2, \boldsymbol{q}) \in \mathbb{R}_+^{n_2 \times n_2} \times \Delta_{n_2}$ がそれぞれ，シンプレックスを含むノードラベル情報を付加した二つの類似性行列を表すとする．ここで，これら二つの測定された類似性行列間の LGW 距離 $d_{\mathcal{LGW}}(\mathbf{F}_1, \mathbf{F}_2, \boldsymbol{p}, \boldsymbol{q})$ は，次のように定義される．

$$d_{\mathcal{LGW}}(\mathbf{F}_1, \mathbf{F}_2, \boldsymbol{p}, \boldsymbol{q}) = \min_{\mathbf{T} \in \mathcal{U}_{n_1 n_2}} \sum_{i,j,k,l} |\mathbf{F}_1(i,k) - \mathbf{F}_2(j,l))| \mathbf{T}_{i,j} \mathbf{T}_{k,l}. \tag{4.1}$$

上式で， $\mathbf{F}_1(i,k)$ と $\mathbf{F}_2(j,l)$ はそれぞれ，G_1 と G_2 内のラベルを用いた距離を表し，次節で詳しく述べる．

LGW の定義の定式化は GW descrepancy の定式化に類似しているが，LGW はラベルを用いた距離に依存している．次に，距離に関する定理を示す．

定理 4. $d_{\mathcal{LGW}}(\mathbf{F}_1, \mathbf{F}_2, \boldsymbol{p}, \boldsymbol{q})$ は距離である．

定理 4 の証明は，**定理 1** の証明と同様である．以降，$d_{\mathcal{LGW}}(\mathbf{F}_1, \mathbf{F}_2, \boldsymbol{p}, \boldsymbol{q})$ を $d_{\mathcal{LGW}}(\mathbf{F}_1, \mathbf{F}_2)$ または $d_{\mathcal{LGW}}$ と表記する．

次に，式 (4.1) で提案されたラベルを用いた GW の拡張距離 $\mathbf{F}$ について詳しく述べる．本研究では，i 番目と k 番目のノード間のパスに沿ったラベル情報を考慮した．本研究では，ノードラベル情報を用いた二つのアプローチを検討した．一つ目は，ノードラベルの局所的な配置に着目した手法である．二つ目の提案手法は，比較するノードとその周辺ノードのラベルの一致度を考慮する手法である．

4.3.2 局所的なラベルの配置を距離とした LGW

本節では，二つのノード間のパスに沿ったラベルの配置について考慮した距離 $\mathbf{F}$ ついて説明する．**定義 3** より，i 番目のノードと k 番目のノード間の全てのパス $\mathcal{P}_{i,k}$ を考慮し，各パス $p \in \mathcal{P}_{i,k}$ 上にあるすべてのノードラベルを用いた．ここで，以下の定義を示す．

定義 11 (ラベルの発生頻度)**.** ラベルの発生頻度について定義

i 番目のノードに着目するとき，i と隣接するノードがある．i と，それに隣接するノードのラベルを一つの集合 g_i とする．その集合の中で，それぞれのラベルの発生頻度 ψ_{ij} を調べる．集合の中のあるラベル l_j の数 w_j を g_i 内のノードの数 z_i で割った値が，ラベル l_i の発生頻度 ψ_{ij} となる．つまり，

$$\psi_{ij} \quad = \quad \frac{w_j}{z_i}. \tag{4.2}$$

そしての発生頻度 ψ_{ij} を各エッジの重みとする．図 4.2 にラベルの発生頻度の概念図を示す．

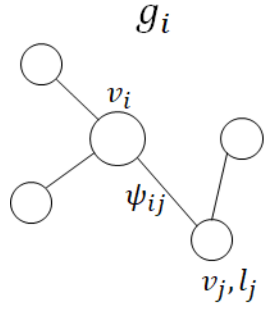


図 4.2: ラベルの発生頻度の概念図

ここで，全てのノードにこの処理を行うと選択するノードによってエッジの重みが変わる，つまり $\psi_{ij} \neq \psi_{ji}$ となる．そこで，以下のように平均化処理を行い重みとする．

$$\bar{\psi}_{ij} = \bar{\psi}_{ji} = \frac{\psi_{ij} + \psi_{ji}}{2}. \tag{4.3}$$

定義 11 はラベル集合から見たノード同士のつながりの強さを表しており，中心のノードと集合の中に多いラベルを持っているノードとのつながりは強いとみなすことができる．これを用いて次のような最小化問題を解くことにより，新しい $\mathbf{F}_{LF}(i,k)$ を導出する．

$$\begin{aligned} \mathbf{F}_{LF}(i,k) &= \min \sum_{(u,v)\in p} \bar{\psi}_{uv} \\ \text{subject to} &\quad p \in \mathcal{P}_{i,k}. \end{aligned} \tag{4.4}$$

ここで，(u,v) は，パス p に沿った一つのエッジを表す．これは，i 番目のノードと k 番目のノード間の複数パス $\mathcal{P}_{i,k}$ の一つである．式 (4.4) で定義された最小化問題は，一般的な重み付き最短経路問題である．したがって，ダイクストラのアルゴリズムなどの一般的なアルゴリズムを使用して解決できる．本手法を，LGW-Label Frequency (LGW-LF) と呼ぶ．

4.3.3 ラベルの一致度を距離とした LGW

次に，二つのノード間のパスに沿ったラベルの一致度について考慮した提案手法について説明する．具体的には，あるエッジの両端のノードに対して，その周辺ノードのラベルの一致度を使用する手法である．

定義 12 (ラベルの一致度)．エッジ e_{ij} の両端のノード v_i, v_j に対して，v_i と v_i に隣接にするノードの集合を g_i とし，v_i と v_i に隣接にするノードの集合を g_j とする．g_i のノードと g_j のノードを総当たりで比較し，一致しているラベルの数をラベルの一致度という．

図 4.3 にラベルの一致度の概念図を示す．

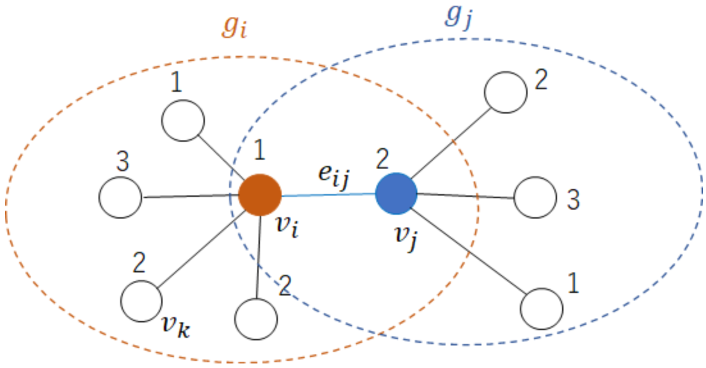


図 4.3: ラベルの一致度の概念図

図 4.3 より，例えば，v_k のラベルは 2 である．それに対して，g_j 内の全てのノードと比較すると，一致するラベルの数は 2 個である．これを各ノードに対して行い，一致するラベルの数を計数する．エッジの両端で一致するラベルが多ければ，ノードの繋がりが強いと言える．これは，エッジの両端のノードの類似性に着目し，二つのノードの距離に反映している．そして，ノード v_i とノード v_j に対するラベルの一致度を $\chi_{ij}(=\chi_{ji})$ として，その逆数を用いた次のような最小化問題を解くことにより，新しい $\mathbf{F}_{LW2}(i,k)$ を導出する．

$$\begin{aligned} \mathbf{F}_{LM}(i,k) &= \min \sum_{(u,v)\in p} \frac{1}{\chi(u,v)} \\ \text{subject to} & \quad p \in \mathcal{P}_{i,k}. \end{aligned} \tag{4.5}$$

ここで (u,v) は，パス p に沿った一つのエッジを表す．これは，i 番目のノードと k 番目のノード間の複数パス $\mathcal{P}_{i,k}$ の一つである．発散を防ぐために一致度の逆数を重みとする．またラベルが一致しないときはカウントせず，0 の時は一致度を使用しないため，ゼロ割りは考慮しない．式 (3.9) で定義された最小化問題は，一般的な重み付き最短経路問題である．したがってこの点は，ダイクストラのアルゴリズムなどの一般的なアルゴリズムを使用して解決できる．ここでは，ラベルの一致度を使用した LGW 距離を LGW-Label Matching (LGW-LM) とする．

4.4　評価実験

本節では，SVM 分類器を用いたグラフ分類実験によって提案手法の評価を行う．本実験の目的は，提案された LGW 距離がラベルの種類の違いにロバストな手法であり，実際に有効であることを明らかにすることである．そのため，4.2.2 項の実験と同様にテストデータのみラベルを入れ替え，SVM 分類器による N 個のグラフのグラフ分類実験を行った．実験では，$N \times N$ カーネル行列を作成し，(p,q) 要素は $\exp(-\gamma \cdot d_{\mathcal{LGW}}(\mathbf{F}_p, \mathbf{F}_q))$ によって計算される．ここで、$\gamma \in \mathbb{R}$ はハイパーパラメータであり，$d_{\mathcal{LGW}}$ は 2 つのグラフ G_p，G_q 間の式 (4.1) の LGW 距離である．γ は 1 に設定した．比較手法は，4.2.1 項の実験で用いた，FGW，WLK，HCK，NHK，VHK である．WLK は，height parameter h は 2 に設定した．SVM 分類器の設定は，比較するすべての手法で同じであり，手法ごとに図 4.1 の実験を行い，分類精度の平均値と標準偏差を算出した．データセットの分割に用いた乱数は，表 3.2 と同様である．表 4.3 に結果を示す．ここでは，平均値 ± 標準偏差として数値を示した．最も良い精度を太字，二番目に良い精度を下線で示す．

表 4.2 より，LGW-LF は三つのデータセットで既存手法を上回っており，ラベルの構造を用いたグラフ分類手法の有効性が確認できた．一方，LGW-LM は MUTAG データセットを除いて，既存手法の分類精度を下回っている．これは，ラベルの構造の分析方法の中に一致度のパラメータが入っており，純粋な構造分析ではなかったためであると考えられる．

表 4.3: グラフのノードラベルを用いた分類精度

メソッド	データセット		
	MUTAG	ENZYMES	ER-MD
FGW	74.28±4.86	23.38±2.60	60.00±2.15
FGW(2)	78.57±4.61	25.91±3.10	60.00±2.15
WLK	65.87±6.32	17.37±3.68	61.69±2.79
HCK	73.49±14.09	27.67±4.31	62.70±2.21
NHK	71.43±13.26	22.83±3.63	61.15±2.23
VHK	77.62±6.08	18.94±2.62	62.36±2.64
LGW-LF	**81.42± 5.81**	**30.61±2.63**	**66.89±3.76**
LGW-LM	80.16±4.67	22.93±2.35	60.00±2.15

4.5 むすび

本章では，グラフのノードラベルの構造に着目した距離である，LGW 距離を提案した．LGW では，ラベルの絶対値ではなく，ラベルの構造をパスの重みとして使用している．トレーニングデータに対して，テストデータのラベルの絶対値を変えてグラフ分類を行った結果，LGW-LF は従来の手法よりも精度が高く，ラベルの構造情報のみで分類できることが分かった．したがって，ラベルが異なる種類のグラフデータの分類にも適用でき，実用性が高いことが分かった．

第5章　結論

5.1　まとめ

本研究では，グラフ類似性測定精度を向上させることを目的に，GW に基づいた新たな手法を提案した．類似性を表す指標となる距離について，二つの手法を検討した．

第一の手法として，グラフのエッジ構造を用いて構造的距離を表す EGW 距離を提案した．そして EGW 距離のパラメータとしてノードの周辺構造に着目した手法である EGW-DT を提案した．これは，ノードが持つエッジの数とエッジが成す三角形の数を構造的距離を表す手法として組み込んでいる．次に，二つのノード間のパスの構造に着目した EGW-TW を提案した．これは，二つのノード間のパス上にある三角形の数をエッジの重みとした手法である．二つの手法のグラフ分類による評価では，構造が複雑なデータの分類に適していることが分かった．

第二の手法として，グラフのノードラベルを用いて構造的距離を表し，ラベルのタイプが異なるグラフに対しても適用することができる LGW 距離を提案した．従来手法ではラベルの絶対値を比較しており，比較するグラフ同士のラベルの種類が一致している必要があった．本研究では異なる種類のラベルを持つグラフを比較できるようにするため，グラフノードラベルの構造に着目した．まず，ノードラベルの発生頻度をエッジの重みとした LGW-LF を提案した．次に，エッジの両端のノードラベルの一致度をエッジの重みとした LGW-LM を提案した．グラフ分類による評価では，LGW-LF は従来の手法より精度が高く，ラベルの構造情報のみで分類できることが分かった．つまり，ラベルが異なる種類のグラフデータの分類にも適用でき，実用性が高いことを示している．一方，LGW-LM は既存手法の分類精度を下回っていた．これは，ラベル構造の分析方法の中にラベルの一致度を用いたパラメータが入っており，純粋な構造分析ではなかったためであると考えられる．

以上のことから，ラベルが同一で構造が複雑なデータには EGW が有効であると考えられる．一方，ラベルの種類が異なるデータ間での比較では LGW-LF が有効である．

5.2　今後の課題

本研究で提案した EGW 距離は，グラフ構造が複雑なデータセットに有効であることが分かった．本研究で行ったソーシャルネットワークに対する実験ではその有効性が確認されたが，その他の構造が複雑なデータセットでも実験する必要がある．また，グラフ構造に三角形を含まないグラフに対しては精度が低くなったため，全てのグラフに対して有効である手法として，全てのグラフに共通な構造特性を利用した手法が必要であると考えられる．LGW 距離は，ノードラ

ベルの種類が異なるグラフを比較する手法として有効であることが分かった．本論文で行った実験では比較的サイズの小さなデータセットを使用したため，より大きなデータセットで実験する必要がある．また，グラフに含まれるノードラベルの種類が少ないデータで実験を行ったため，ラベルの種類の数によってどのように変化するかを確かめる必要がある．さらに，EGW 距離，LGW 距離共に，実際のアプリケーションに適用し，その実用性を測る必要がある．

謝辞

本研究にあたり，環境を整えてくださり，研究についてアドバイスをくださった，早稲田大学基幹理工学部情報通信学科の渡辺教授に感謝申し上げます．また，研究の進め方についてご指導くださり，お手伝いいただいた早稲田大学基幹理工学部情報通信学科の笠井裕之教授に心から感謝いたします．さらに，研究についてアドバイスをくださった早稲田大学国際情報通信センターの石川孝明氏に感謝申し上げます．

また，学業に専念するための環境をくださり，生活を支えてくださった家族に感謝申し上げます．

参考文献

[1] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[2] T. M. Lehmann, M. O. Güld, C. Thies, B. Fischer, K. Spitzer, D. Keysers, H. Ney, M. Kohnen, H. Schubert, and B. B. Wein. Content-based image retrieval in medical applications. *Methods of Information in Medicine*, Vol. 43, No. 4, pp. 354–361, 2004.

[3] D. J. Jacobs, L. A. Rader, A. J. Kuhn, and M. F. Thorpe. Protein flexibility predictions using graph theory. *Proteins*, Vol. 44, No. 2, pp. 150–165, 2001.

[4] P. Mahé and J.-P. Vert. Graph kernels based on tree patterns for molecules. *Machine learning*, Vol. 75, No. 1, pp. 3–35, 2009.

[5] H. Cai and K. C. Zheng, V. W.and Chang. A comprehensive survey of graph embedding: Problems, techniques and applications. *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[6] F. Emmert-Streib, M. Dehmer, and Y. Shi. Fifty years of graph matching, network alignment and network comparison. *Information Sciences*, Vol. 346–347, pp. 180–197, 2016.

[7] E. M. Loiola, N. M. M. de Abreu, P. O. Boaventura-Netto, H. Peter, and T. Querido. A survey for the quadratic assignment problem. *European Journal of Operational Research*, Vol. 176, No. 2, pp. 657–690, 2007.

[8] Eugene L. Lawler. The quadratic assignment problem. *Management Science*, Vol. 9, pp. 586–599, 1963.

[9] T. C. Koopmans and M. J. Beckmann. Assignment problems and the location of economic activities. *Econometrica*, Vol. 25, pp. 53–76, 1957.

[10] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 24, No. 4, 2002.

[11] S. Wang, Z. Chen, X. Yu, L. Ding, J. Ni, L.-A. Tang, J. Gui, Z. Li, H. Chen, and P. S. Yu. Heterogeneous graph matching networks for unknown malware detection. In *Proceedings*

of the Twenty-Eight International Joint Conference on Artificial Intelligence (IJCAI), 2019.

[12] B. Schölkopf, K. Tsuda, and J.-P. Vert. *Kernel Methods in Computational Biology.* The MIT Press, 2004.

[13] H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized kernels between labeled graphs. In *20th Internatioal Conference in Machine Learning (ICML)*, 2003.

[14] T. Gärtner, P. Flach, and S. Wrobel. On graph ker- nels: Hardness results and efficient alternatives. In *16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel*, 2003.

[15] K. M. Borgwardt and H.-P. Kriegel. Shortest-path kernels on graphs. In *5th International Conference on Data Mining*, Vol. 5, pp. 74–81, 2005.

[16] T. Horvath, T. Gartner, and S. Wrobel. Cyclic pattern kernels for predictive graph mining. In *10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2004.

[17] N. Shervashidze, S.V.N. Vishwanathan, T. H. Petri, K. Mehlhorn, and K. M. Borgwardt. Efficient graphlet kernels for large graph comparison. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2009.

[18] K. Grauman and T. Darrell. The pyramid match kernel: Efficient learning with sets of feature. *The Journal of Machine Learning Research*, Vol. 8, pp. 725–760, 2007.

[19] N. M. Kriege, F. D. Johansson, and C. Morris. A survey on graph kernels. *Applied Network Science*, Vol. 5, No. 6, 2020.

[20] N. Shervashidze, P. Schweitzer, E. J. v Leeuwen, K. Mehlhorn, and K. M. Borgwardt. Weisfeiler-lehman graph kernels. *The Journal of Machine Learning Research*, Vol. 12, pp. 2539–2561, 2011.

[21] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, Vol. 34, No. 4, pp. 18–42, 2017.

[22] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint: arXiv:1609.02907*, 2017.

[23] M. Defferrard, X. Bresson, and P. Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

[24] M. Niepert, M. Ahmed, and K. Kutzkov. Learning convolutional neural networks for graphs. In *Thirty-third International Conference on Machine Learning (ICML)*, 2016.

[25] Cedric. Villani. *Optimal transport: Old and new.* Springer, New York, 2008.

[26] G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, Vol. 11, No. 5-6, pp. 355–607, 2019.

[27] Facundo Mémoli. On the use of gromov–Hausdorff distances for shape compasrison. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007.

[28] Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Found. Comput. Math.*, Vol. 11, pp. 417–487, 2011.

[29] G. Peyré, M. Cuturi, and J. Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *Thirty-third International Conference on Machine Learning (ICML)*, 2016.

[30] T. Vayer, L. Chapel, R. Flamary, R. Tavenard, and N. Courty. Optimal transport for structured data with application on graphs. In *Thirty-sixth International Conference on Machine Learning (ICML)*, 2019.

[31] H. Xu, D. Luo, H. Zha, and L. Carin. Gromov-wasserstein learning for graph matching and node embedding. In *Thirty-sixth International Conference on Machine Learning (ICML)*, 2019.

[32] H. Xu, D. Luo, and L. Carin. Scalable Gromov-Wasserstein learning for graph partitioning and matching. In *33rd Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

[33] C. Bunne, D. Alvarez-Melis, A. Krause, and S. Jegelka. Learning generative models across incomparable spaces. In *Thirty-sixth International Conference on Machine Learning (ICML)*, 2019.

[34] Y. Yan, W. Li, H. Wu, H. Min, M. Tan, and Q. Wu. Semi-supervised optimal transport for heterogeneous domain adaptation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, 2018.

[35] Y. Lipman, R. M. Rustamov, and T. A. Funkhouser. Biharmonic distance. *ACM Transactions on Graphics (TOG)*, Vol. 29, No. 3, p. 27, 2010.

[36] T. Kataoka and A. Inokuchi. Hadamard code graph kernels for classifying graphs. *5th International Conference on Pattern Recognition Applications and Methods*, pp. 24–32, 2016.

[37] S. Hido and H. Kashima. A linear-time graph kernel. *9th International Conference on Data Mining*, pp. 179–188, 2009.

[38] John C. Gower. A general coefficient of similarity and some of its properties. *Biometrics*, pp. 857–871, 1971.

[39] M. Sugiyama and K. M. Borgwardt. Halting in random walk kernels. *In Advances in Neural Information Processing Systems*, pp. 1639–1647, 2015.

[40] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, Vol. 21, No. 2, pp. 343–348, 1967.

[41] Marco. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Annual Conference on Neural Information Processing Systems (NIPS)*, 2013.

[42] J. D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing,*, Vol. 37, No. 2, pp. 1111–A1138, 2015.

[43] V. Seguy and M. Cuturi. Principal geodesic analysis for probability measures under the optimal transport metric. In *Advances in Neural Information Processing Systems (NIPS)*, 2015.

[44] A. Rolet, M. Cuturi, and G. Peyré. Fast dictionary learning with a smoothed wasserstein loss. In *Nineteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2016.

[45] N. Courty, R. Flamary, and D. Tuia. Domain adaptation with regularized optimal transport. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML)*, 2014.

[46] M. Cuturi and A. Doucet. Fast computation of wasserstein barycenters. In *Thirty-first International Conference on Machine Learning (ICML)*, 2014.

[47] J. Solomon, G. Peyré, V. Kim, and S. Sra. Entropic metric alignment for correspondence problems. *ACM Transactions on Graphics (TOG)*, Vol. 35, No. 4, p. 72, 2016.

[48] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*, Vol. 33 of *AMS Graduate Sudies in Math.* American Mathematical Society, 2001.

[49] H. Petric Maretic, M. El Gheche, G. Chierchia, and P. Frossard. GOT: An optimal transport framework for graph comparison. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[50] H. Petric Maretic, M. El Gheche, G. Chierchia, and P. Frossard. Wasserstein-based graph alignment. *arXiv preprint: arXiv:2003.06048*, 2020.

[51] K. Riesen and H. Bunke. Iam graph database repository for graph based pattern recognition and machine learning. *SSPRSPR*, Vol. 5342, pp. 287–297, 2008.

[52] A. K. Debnath, R. L. Lopez de Compadre, G. Debnath, A. J. Shusterman, and C. Hansch. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. Vol. 2, pp. 786–797, 1991.

[53] N. Kriegel and P. Mutzel. Subgraph matching kernels for attributed graphs. *International Conference on Machine Learning*, 2012.

[54] J. Sutherland, L. A.O ' Brien, D. F. Weaver. Spline-fitting with a genetic algorithm: a method for developing classification structure-activity relationships. *J. Chem. Inf. Comput. Sci.*, Vol. 43, pp. 1906–1915, 2003.

[55] P. Yanardag and S. V. N. Vishwanathan. Deep Graph Kernels. In *21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1365–1374. ACM Press, 2015.

[56] L. Oettershagen, N. Kriege, C. Morris, and P. Mutzel. Temporal graph kernels for classifying dissemination processes. In *SIAM International Conference on Data Mining (SDM)*, 2020.

[57] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, Vol. 271, No. 1, pp. 166–180, 2011.

[58] N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. *Personal Ubiquitous Computing*, Vol. 10, No. 4, pp. 255–268, 2006.

[59] S. Pan, J. Wu, X. Zhu, G. Long, and C. Zhang. Task sensitive feature exploration and learning for multi-task graph classification. *IEEE Transactions on Cybernetics*, Vol. 47, No. 3, pp. 744–758, 2017.

[60] M. Neumann, N. Patricia, R. Garnett, and K. Kersting. Efficient graph kernels by randomization. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML)*, 2012.

[61] Uirike. von. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, Vol. 17, pp. 395–416, 2007.

[62] G. Siglidis, G. Nikolentzos, S. Limnios, C. Giatsidis, K. Skianis, and M. Vazirgianis. GraKeL: A Graph Kernel Library in Python. *Journal of Machine Learning Research*, Vol. 21, pp. 1–5, 2020.

研究業績

[1] 浅見莉絵子，石川孝明，渡辺裕，” 機械学習による日本語話者の自動読唇の基礎検討 ”，第 33 回画像符号化シンポジウム・第 23 回映像メディア処理シンポジウム（PCSJ/IMPS2018），P-3-08，Nov. 2018.

[2] 浅見莉絵子，石川孝明，渡辺裕，” 機械学習による日本語話者の自動読唇 ”，電子情報通信学会総合大会，Mar. 2018.

[3] 横井真也，浅見莉絵子，石川孝明，渡辺裕，” ボロノイ図に基づく 3 次元優勢領域によるパスコース評価について ”，映像情報メディア学会年次大会，13D-3, Sep. 2017.