

# グラフ埋め込み手法に基づく従業員の ビジネスコミュニケーション分析に関する一考察

野中 賢也<sup>1,a)</sup> 山下 遥<sup>2,b)</sup> 三浦 豊史<sup>3,c)</sup> 後藤 正幸<sup>1,d)</sup>

受付日 2022年4月5日, 採録日 2022年12月2日

**概要:** 近年の多くの企業では、生産性の高いビジネスコミュニケーションを目的としてチャットアプリケーションが活用されるようになった。その主たる導入目的は社員間のコミュニケーションであるが、ビジネスアナリティクスの観点からは、ビジネスチャットシステムに蓄積されたデータを活用して社員間コミュニケーションの分析が可能となっており、社内のコミュニティの実態を把握することで、適切な人材マネジメントへの活用が期待されている。その際、チャットアプリケーションではチャンネルと呼ばれるグループ単位で社員間のコミュニケーションが行われることから、社員間のコミュニケーションの特徴をチャンネルごとに分析することが望まれる。たとえば、各チャンネルのコミュニケーション状態を表現したグラフデータを構造化データへと変換できれば、チャンネルのクラスタリングや可視化分析が可能になると考えられる。本研究では、グラフデータを構造化データに変換する Deep Divergence Graph Kernel (DDGK) を改良し、チャンネルグラフの分析に適した埋め込み表現を得るための手法を提案する。本手法は、チャンネルグラフの重要な構造を表現する埋め込み空間を構築し、各チャンネルグラフをこの空間上の点として埋め込むことができる。さらに、某企業のビジネス現場で導入されている Slack 上に保存された実際の会話履歴データに適用することで、提案モデルの有効性を示す。提案モデルを用いたデータ分析により、特徴ベクトルとして表されたグラフ構造を活用してチャンネルグラフが分類でき、各チャンネルにおけるコミュニケーションの特徴を分析することが可能となる。

**キーワード:** ビジネスチャットデータ, Deep Divergence Graph Kernel, Embedding, グラフマイニング, グラフ埋め込み

## Graph Embedding for Analysing Business Communication between Employees

KENYA NONAKA<sup>1,a)</sup> HARUKA YAMASHITA<sup>2,b)</sup> TOYOFUMI MIURA<sup>3,c)</sup> MASAYUKI GOTO<sup>1,d)</sup>

Received: April 5, 2022, Accepted: December 2, 2022

**Abstract:** On business chat applications that are widely used in many offices, communications between the employees are made in each group called channel. From the viewpoint of business analytics, the activation of human resource can be realized by grasping the real situation of the office community based on the analysis of the communication represented as the graph structure by utilizing the data stored on the business chat system. Especially, it is desirable to analyze the characteristics of business communication between employees for each channel, which is a group created for specific communication. For the purpose of analyzing employees' communication, it is desirable to apply clustering or visualizing analyses of channels and it seems to be possible if many channel graphs are converted to structured data. In this study, we improve the Deep Divergence Graph Kernel (DDGK) which converts the graph data to the structured data for chatting data analysis, and propose a method for obtaining embedding representation suitable to analyze the channel graphs. The proposed method enables to construct the embedding space which represents the important structure of channel graphs. Moreover, we show the effectiveness of the proposed model by applying to an actual communication data stored on Slack at a Japanese company. In this data analysis, feature vectors are activated by the model, and channel graphs can be classified by graph structures.

**Keywords:** business chat data, Deep Divergence Graph Kernel, embedding, graph mining, graph embedding

## 1. はじめに

企業経営において、人的リソースの有効活用とパフォーマンスを最大化する職場コミュニティの醸成と、そのための自社内の様々なチームのコミュニケーション状態の客観的な分析は重要な課題である。近年、企業の職場コミュニティにおいても様々な ICT (Information and Communication Technology) ツールが用いられるようになり、様々なコミュニケーションデータが蓄積されるようになったことから、これらのデータを活用した職場コミュニティのコミュニケーション状態を分析することへの期待が高まっている [1]。このような課題に対して、E-mail の送受信履歴データを解析することによって、社内の会話状況を把握する研究などがなされている [2]。これに対し最近では、E-mail の代替ツールとして、Slack [3] や Teams [4] などのビジネスチャットアプリが職場内のコミュニケーションのために使用され始めており、その存在感は年々大きくなっている。これらのチャットアプリは、これまで研究対象とされてきた E-mail より気軽な意思疎通にも適しているといったメリットがあり、これらのアプリに蓄積された会話履歴データの利活用により、より実態に近いコミュニケーション状態の分析が期待される。

チャットアプリでは、社員間の会話はチャンネルと呼ばれるプロジェクトごとに作成された社員グループ上で行われる。このチャットアプリ上での会話量により、各社員同士のコミュニケーション強度を定量化することができるため、すべての社員ペアに対して関係性の強度が与えられたグラフ構造のデータと見なすことができる。すなわち、チャンネル上での会話履歴をもとに、社員をノード、社員間の業務上の関係性の有無をエッジとして表現したグラフを構築することが可能である (以下、チャンネルグラフと呼ぶ)。このチャンネルグラフは参加者の会話傾向によって形状が異なり、各チャンネルが対象としている社員グループや業務内容によってその特徴が異なると考えられる。たとえば、1 人に会話が集中するチャンネルや参加者各々が対等に会話を行うチャンネルなど、グラフ構造によって各チャンネルのコミュニケーション特性をとらえられる可能性が高い。これらの部署やタスク別のチャンネルグラフの生起パターンを分析することによって、社員間の会話傾向の特徴を明らかにし、職場チーム構成の評価や適切な人材マネジメントにつながる

ると期待される。しかし、チャンネルグラフの集合は複雑な非構造化データであり、多数のチャンネルをグラフ表現のまま分析することは困難である。そのため、各チャンネルグラフの特徴を抽出してテーブル (行がグラフ、列がその特徴を表すデータ) の形式で表現した構造化データへと変換する (構造化する) ことで、従来研究されてきたクラスタリングや回帰・分類モデルといった各手法の適用が期待される。そこで本研究では、抽出された特徴ベクトルに対してクラスタリング手法 [5] を適用し、各チャンネルグラフの類型化を行う。

一方、非構造化データの特徴抽出の方法として、ニューラルネットワークの高い表現能力を利用した埋め込み表現モデルの研究が進められている。特に、チャンネルグラフの特徴抽出では、構造が類似したグラフを特徴空間上の近接領域に埋め込むグラフ埋め込み表現モデル [6] が有効であると考えられる。本研究では、Al-Rfou らによって提案された Deep Divergence Graph Kernel (以下、DDGK と呼ぶ) [7] に注目する。この手法は、グラフ埋め込み表現の学習にノードやエッジのラベルを必要としない汎用性の高い手法である。DDGK では、埋め込み空間の軸に対応するグラフ (以下、ソースグラフ) を埋め込み対象のグラフ (以下、ターゲットグラフ) からのランダムサンプリングにより選定している。しかし、このような選定方法では、構成される埋め込み空間の各軸が相関を持ってしまい、類似した特徴を表す軸が作られてしまう恐れがある。また、チャンネルグラフの重要な特徴を表す軸が選ばれていないケースも考えられ、その場合は分析に必要な観点が漏れてしまっていることになるため、これらの問題を解決するような方法が必要となる。

そこで、本研究では、Barabási-Albert (BA) モデル [8] に基づき適切なソースグラフを生成することで、チャンネルグラフの埋め込み表現を得るための特徴空間軸構成法を提案する。これにより、チャンネルグラフの重要な構造を多面的に表現できる埋め込み空間の構成が可能となる。加えて、チャットアプリの 1 つである Slack をビジネスコミュニケーションに導入している実企業の会話履歴データに提案モデルを適用する。提案モデルによって得られる特徴ベクトルを活用して各チャンネルのグラフ構造を類型化し、その分析結果から得られる知見について考察を与え、提案モデルの有効性を示す。

## 2. 関連研究

ネットワーク分析 (グラフマイニング) は、各ノード (頂点) がエッジ (辺) で連結されたグラフデータから有用な

<sup>1</sup> 早稲田大学  
Graduate School of Creative Science and Engineering,  
Waseda University, Shinjuku, Tokyo 169-8555, Japan  
<sup>2</sup> 上智大学  
Sophia University, Chiyoda, Tokyo 102-8554, Japan  
<sup>3</sup> Laboratik 株式会社  
Laboratik Inc., Shinagawa, Tokyo 142-0041, Japan  
a) nonakakenya@akane.waseda.jp  
b) h-yamashita-lg8@sophia.ac.jp  
c) toyofumi@laboratik.com  
d) masagoto@waseda.jp

本研究は 2019 年の日本計算機統計学会シンポジウムで発表した社員間コミュニケーションの類型化を可能とするグラフ構造の分散表現モデル [13] に対して研究を深め、その成果をまとめたものである。

知見を抽出することを目的として研究されてきた。ネットワーク構造の中でも特に、人間の社会的な関係性をグラフとして表現したソーシャルグラフの分析は、社会集団における個人の性質や社会集団全体の振舞いを理解するうえで重要なアプローチとなっている。代表的なものは、Twitter や Facebook などの友達関係から構成される Social Network Service (SNS) コミュニティに関する分析である。たとえば、Twitter 上の影響力のあるアカウントを特定するために PageRank などの中心性を測る手法 [10], [11] や、組織において優れた功績を与えている社員を発見するための手法 [9] が提案されている。また、Facebook における友達関係のデータを用いたコミュニティの状態を表現するためのネットワーク分析に関する研究 [16], [17] やリンクに着目したコミュニティ検出の研究 [18] も行われている。さらに、Instagram を対象とした分析も行われている [19]。これらの研究では、1つの巨大なネットワークに対して、様々な角度から有用な分析手法が提案されている。

本研究では、社員をノードとして、また業務上のつながりをエッジとして位置付けている。構成される「組織内ネットワーク」の一種である、コミュニケーションのネットワークを研究対象とする。組織内の社員は、日頃の業務のために物理的に直接会って会話を行うか、E-mail や後述のチャットアプリを使用してコミュニケーションをとっている。前者の口頭会話は、同一空間で仕事が行われる場合には、最も頻度高く使用されるコミュニケーション形態である。Araらは、社員にウェアラブルセンサをつけることで物理的な会話ログから社内コミュニティの分析を行っている [12]。

しかしながら、社員に対してセンサ装着を依頼しなければならず、コミュニケーションデータを収集する際のコストが大きいため、多くの企業に対して同様の実験を展開することは難しいといえる。後者の ICT ツール上のコミュニケーションログデータは、データベースに自動的に蓄積されていくため、データ収集コストが比較的小さく、より多くの企業を対象とした分析が期待できる。代表的なコミュニケーションログデータとしては E-mail データがあげられる。E-mail は宛先 (To) と送り主 (From) を記述するプロトコルであるため、これらの情報を使ってユーザー間の関係性をグラフデータで表現しやすい。E-mail データを活用した研究としては、企業内の中心人物を発見する中心性指標の提案 [2], [14] や企業内のコミュニティ構造の発見 [15] などの分析がすでに行われている。また、山口らは、E-mail データの中でも特に、メーリングリスト機能に着目し、各種ネットワーク指標による評価を行っている [20]。

一方、Slack や Teams などのビジネスチャットアプリは、E-mail に代わるコミュニケーションツールとして、近年、急速に普及している。ビジネスチャットアプリは、E-mail と比較してリアルタイム性が高く、短い返答でも許容され

る雰囲気があるなど、気軽にコミュニケーションがとれることが大きな特徴となっている。これらのビジネスチャットアプリ上で交わされる会話から得られるデータを可視化することにより、各社員のチャットコミュニケーションの状況を可視化することも可能となっている。

一方で、これらのコミュニケーションツールでは、E-mail とは異なり、チャンネルと呼ばれるコミュニケーショングループの機能があり、個別に構築されるコミュニケーションチャンネル上にユーザが投稿する形式で会話が進められる。すなわち、職場や仕事単位でこのチャンネル（グループ）を使い分けることができることが特徴の1つとなっている。実際、同じ企業内であっても、職場単位や仕事単位で多くのチャンネルが作られており、各ユーザは関係する複数のグループに所属し、それぞれにおいて並行してコミュニケーションが行われている。すなわち、職場や仕事単位別にメンバ間のコミュニケーションの状態を分析するためには、このチャンネルごとに構築される多数のコミュニケーションネットワークを可視化する手法が必要となる。これに対し、ビジネスチャット上での発言までの時間に着目したネットワークの構築に関する研究はなされている [21]。また、グループごとの会話の内容に着目した企業におけるコミュニケーションの違いを分析するためのネットワークモデルの構築方法も提案されている [22]。しかしながら、これらの研究では、本研究において着目する各チャンネルごとのコミュニケーション形態の把握という観点でのモデル構築は行っていない。

以上の点を考慮して、本研究では、チャンネルごとに会話状況を表現するグラフデータを作成し、これらを多次元特徴空間上に埋め込むことで、コミュニケーションスタイルの類型化を行う方法を提案する。

### 3. 準備

#### 3.1 対象とするグラフデータ

本研究で用いるデータは、Slack や Teams といったチャットアプリ上に蓄積された会話履歴データである。アプリ上には、特定業務の連絡を目的としたチャンネルが多数存在し、チャンネル内で社員同士の会話が記録される。チャンネルには、2名の参加者が直接やりとりする DM (Direct Message) チャンネルと複数のユーザが参加するグループチャンネルが存在する。グループチャンネルでは、1つのメッセージに対して、その発信者と受信者（当該チャンネルの参加者）が1対多で結び付いている。発信者が受信者を特定する機能（メンション機能）を使用した場合に限り、発信者と受信者は1対1で結び付く。

ある企業のチャットアプリ上に、 $N$  個のチャンネルが存在し、アプリ上に登録されたユーザの集合を  $V$  とする。 $n$  番目 ( $n = 1, \dots, N$ ) のチャンネルで投稿されている  $K_n$  件の会話履歴データを  $D_n = \{(s_k^n, t_k^n)\}_{k=1}^{K_n}$  と記述する。

$(s_k^n, t_k^n)$  は、 $n$  番目のチャンネルにおける  $k$  番目の発言を表す。  $s_k^n \in V$  は、 $k$  番目の発言者であり、また、  $t_k^n \in \mathbb{R}_+$  は  $k$  番目の発言の時刻を表す。ただし、  $\mathbb{R}_+$  は非負の実数からなる集合とする。  $n$  番目のチャンネルの会話履歴データ  $D_n$  から、ノードをチャンネル参加者、エッジを参加者同士のつながりとした無向重みなしグラフ (チャンネルグラフ)  $G_n$  を構築することができる。構築の手順としては、事前に設定した閾値  $\zeta > 0$  を用いて、会話の間隔が閾値 (本研究では 6 時間と設定) 以下の  $n$  番目のチャンネルにおける連続した発言のペア ( $k$  番目と  $k+1$  番目の発言者  $s_k^n$ , および  $s_{k+1}^n$  のペア) について「連絡をとった」と見なし、それぞれの発言者の連絡回数をカウントしていく。そして連絡回数の多いペアに対してエッジを構築する以下の手順を用いる。

**STEP1**  $n$  番目のチャンネルのユーザ  $s_i^n$  とユーザ  $s_j^n$  の連絡回数を  $W_{s_i^n, s_j^n}$  とし、すべてのユーザの組について、  $W_{s_i^n, s_j^n} = 0$  と初期化する。また、発言番号のインデックス  $k$  を  $k=1$  として初期化する。

**STEP2**  $t_{k+1}^n > t_k^n + \zeta$  のとき、  $s_k^n$ , および  $s_{k+1}^n$  に該当する  $s_i^n, s_j^n$  に対して、  $W_{s_i^n, s_j^n} = W_{s_i^n, s_j^n} + 1$  とする。

**STEP3**  $k = k+1$  とする。  $k < K_n$  であれば、STEP2 に戻る。それ以外であれば、STEP4 を実行する。

**STEP4** 連絡回数  $W_{s_i^n, s_j^n}$  の平均値を  $\bar{W}$  とする。  $W_{s_i^n, s_j^n} < \bar{W}$  となるような  $s_i^n, s_j^n$  の組に対してエッジを構築したグラフを  $G_n$  として出力する。

このように構築されたチャンネルグラフは、当該チャンネル内で平均以上のコミュニケーションを行っているペアに対してのみエッジが構築されており、ネットワークデータと位置づけることができる。本研究では、  $N$  個のチャンネルグラフ  $G_n$  を  $M$  次元の多次元特徴空間へと埋め込んだ表現  $\Psi(G_n) \in \mathbb{R}^M$  ( $n = 1, \dots, N$ ) を得る手法を提案する。ただし、  $\mathbb{R}^M$  は  $M$  次元のユークリッド空間を表すものとする。

### 3.2 従来手法

Deep Divergence Graph Kernel (DDGK) [7] は、埋め込みの対象となるグラフの集合 (ターゲットグラフ) を埋め込み空間の軸に相当するグラフ (ソースグラフ) との非類似度によって、多次元特徴空間上へと埋め込む手法である。ターゲットグラフ集合を  $\{G_1, \dots, G_N\}$ 、ソースグラフ集合を  $\{S_1, \dots, S_M\}$  とおく。ここで、ソースグラフ集合は、ターゲットグラフからのランダムサンプリングによって事前に設定されるものとする。

ターゲットグラフ  $G_n$  の埋め込み表現  $\Psi(G_n) \in \mathbb{R}^M$  ( $n = 1, \dots, N$ ) の第  $m$  要素は、ターゲットグラフと第  $m$  番目のソースグラフとの非類似度  $D(G_n|S_m)$  となる。

$$\Psi(G_n) = [D(G_n|S_1), \dots, D(G_n|S_M)] \quad (1)$$

非類似度  $D(G_n|S_m)$  は、ソースグラフの構造を学習したニューラルネットワーク (エンコーダ) のパラメータ (重み)  $\theta_m$  を用いて、ターゲットグラフの構造を予測した際の損失として計算される。

#### 3.2.1 ソースグラフの構造学習

ソースグラフ  $S_m$  ( $m = 1, \dots, M$ ) の構造を学習したニューラルネットワークをソースグラフ数  $M$  個分、事前に学習する。ソースグラフ構造の学習は、ソースグラフ上の隣接ノードの予測タスクとして与えられる。隣接ノードの予測タスクとは、One-Hot 表現されたノード  $\mathbf{v}_i$  ( $i = 1, \dots, I$ ) をエンコーダの入力とし、その隣接ノード  $\mathbf{v}_j \in N(\mathbf{v}_i)$  を予測するタスクのことである。ここで  $N(\mathbf{v}_i)$  は、ノード  $\mathbf{v}_i$  の隣接ノード集合である。学習は以下の対数尤度  $L(\theta_m)$  を最大化させるように誤差逆伝搬法によりパラメータ  $\theta_m$  を更新する。

$$L(\theta_m) = \sum_{i=1}^I \sum_{\mathbf{v}_j \in N(\mathbf{v}_i)} \log P(\mathbf{v}_j | \mathbf{v}_i, \theta_m) \quad (2)$$

あるノードの隣接ノードを予測するように最適化されたパラメータ  $\theta_m$  は、ソースグラフ  $S_m$  の構造を表現するように学習がなされているといえる。

#### 3.2.2 ターゲット・ソース間のノード対応の学習

ターゲットグラフとソースグラフの非類似度を計算するために、学習されたエンコーダを用いて、エンコーダ学習時と同様の隣接ノード予測タスクを、ターゲットグラフに対して行う。つまり、ターゲットグラフのあるノード  $\text{id}$  を入力として、その隣接ノードの  $\text{id}$  を予測するタスクを重みパラメータ  $\theta_m$  を持つエンコーダを用いて行う。ここで、  $\theta_m$  でパラメタライズされたエンコーダの入力部および出力部には、ソースグラフ  $S_m$  のノード数分のユニットが存在するが、ターゲットグラフとソースグラフでは、一般的にノード数が異なる。したがって、入力部と出力部にターゲットグラフ  $G_n$  のノード数分のユニットを加え、エンコーダの入力部と出力部に全結合層で結び付ける。

この全結合層は、アテンション層 [23] の役割を果たし、ソースグラフとターゲットグラフのノード間の対応を学習する (入力部, 出力部の全結合層それぞれの重みを  $\theta_{in}, \theta_{out}$  とする)。エンコーダに入力部と出力部のアテンション層が加えられたモデルにおいて、式 (3) の対数尤度  $L(\theta_{in}, \theta_{out})$  を最大化させるように誤差逆伝搬法により  $\theta_{in}, \theta_{out}$  を更新する。ここで  $\hat{\theta}_m$  は、前ステップで学習済みのエンコーダの重みであり、ターゲットグラフの構造学習時には更新されないパラメータである。また、  $\mathbf{u}_i$  はターゲットグラフの One-Hot 表現されたノードであり、  $N(\mathbf{u}_i)$  はその隣接ノード集合を示す。

$$L(\theta_{in}, \theta_{out}) = \sum_{i=1}^I \sum_{u_j \in N(u_i)} \log P(u_j | u_i, \hat{\theta}_m, \theta_{in}, \theta_{out}) \quad (3)$$

### 3.2.3 非類似度の定義

非類似度  $D(G_n | S_m)$  は、学習後の重み  $\hat{\theta}_m, \hat{\theta}_{in}, \hat{\theta}_{out}$  を用いて、ターゲットグラフのすべてのノード  $u_i$  ( $i = 1, \dots, I$ ) について、その隣接ノードを予測した際の損失として式 (4) のように定義される。

$$D(G_n | S_m) = - \sum_{i=1}^I \sum_{u_j \in N(u_i)} \log P(u_j | u_i, \hat{\theta}_m, \hat{\theta}_{in}, \hat{\theta}_{out}) \quad (4)$$

仮に、エンコーダに学習されているソースグラフがターゲットグラフと同型であれば、同じ構造を予測するためのパラメータ  $\theta_m$  を用いて、ターゲットグラフの構造を予測していることになるため、損失は小さくなることが期待される。反対にエンコーダに学習されているソースグラフとターゲットグラフの形状が異なると、異なる構造を予測するためのパラメータ  $\theta_m$  を用いて、ターゲットグラフの構造を予測しているため、損失は大きくなることが期待される。したがって、この損失をターゲットグラフとソースグラフの非類似度と見なすことができる。

このように算出される式 (1) および式 (4) から、ターゲットグラフ  $G_n$  の埋め込み表現  $\Psi(G_n)$  の埋め込み表現を得ることができる。

### 3.3 Barabási-Albert (BA) モデル

本研究では、BA モデル [8] により生成されたグラフを DDGK のソースグラフとして用いることを考える。BA モデルは、次数分布がべき乗分布に従うネットワークを生成する代表的なモデルであり、グラフの成長と優先的選択を生成アルゴリズムに取り入れている。具体的には、ノード数  $a$  とノード追加時のエッジ数 (初期ノード数)  $b$  を引数として以下の方法でグラフを生成する。

- (1)  $b$  個のノードを持つグラフ  $G$  を構築する。
- (2)  $G$  にノードを 1 個追加する (グラフの成長)。
- (3) 既存のノードから  $b$  個のノードをそれぞれの次数に比例した確率で、ランダムに選択し、追加されたノードをエッジにより結び付ける (優先的選択)。
- (4)  $G$  のノード数が  $a$  に等しければ  $G$  を出力、そうでなければ (2) に戻る。

## 4. 提案手法

### 4.1 埋め込み空間の軸が満たすべき要件

本研究では、ノードの数 (社員数) が比較的少ないネットワークに焦点を当て、社員間コミュニケーションの分析の

ためのチャンネルグラフ構造の類型化を行う。具体的には、チャンネルグラフの埋め込み表現にクラスタリング手法を適用することを考える。このとき、クラスタリング手法の基礎となるデータ間の類似度には、チャンネルグラフにおける重要な構造が多面的に反映されている必要がある。データ間の類似度は、埋め込み空間の各軸に大きく依存するため、チャンネルグラフの重要な構造を多面的に評価する埋め込み空間軸が必要となる。

よって、本研究目的における埋め込み空間の軸は、以下の 2 つの要件を満たす必要がある。1 点目は、チャンネルグラフにおける重要な特徴を漏れなくとらえていることである。2 点目は、チャンネルグラフの多面的な形状的特徴を表現するため、埋め込み表現の各軸は互いに異なる形状を評価する、すなわち、軸どうしの相関が低いことである。しかし、Al-Rfou ら [7] は、埋め込み空間の軸に対応するソースグラフをターゲットグラフからのランダムサンプリングにより設定している。すなわち、上述の要件を満たすようなソースグラフの選定は行っていない。

### 4.2 要件を満たすソースグラフの生成

本研究では、上述の 2 つの要件を満たすような軸となるソースグラフをネットワーク生成モデルを用いて生成する。ネットワークの生成モデルは、生成されるネットワークの種類に応じて多数存在するが、チャンネルグラフの埋め込みには、Barabási-Albert (BA) モデル [8] によって生成されたグラフをソースグラフに用いることが適切である。その理由は、第 1 の要件であるチャンネルグラフにおける重要なネットワーク特徴の表現という観点から説明することができる。

チャンネルグラフを表現するうえで、重要なネットワーク特徴を特定するため、チャンネルグラフ集合に対して、従来使用されてきたネットワーク基本特徴量 [24] のうち 13 の代表的な特徴量 (平均次数、エッジ数、最大次数、次数分散、ノード数、孤立ノード数、連結成分数、密度、次数相関、グラフ直径、平均クラスタ係数、最小次数、の 13 の特徴量。以下、「基本特徴量セット」) を計算した。これらの特徴量の共分散行列を求めて可視化したものが、図 1 である。図 1 では、各ノードが基本特徴量、エッジが特徴量間の共分散を示している。エッジの太さが各特徴量間の共分散の絶対値の大きさを示す。ただし、共分散の閾値を 0.7 と設定し、共分散の絶対値が低いエッジは剪定した。これは、ノード間の関係性をよく表しているようなグラフを表示するために、より関係の強いノード間にエッジが張られ、人間にとって解釈がしやすいような閾値を探索的に調べ、当該値に設定したものである。図 1 より、チャンネルグラフを表現する重要な特徴量は、サイズに関連する特徴量および密度に関連する特徴量の大きく 2 グループに分けられることが分かる。

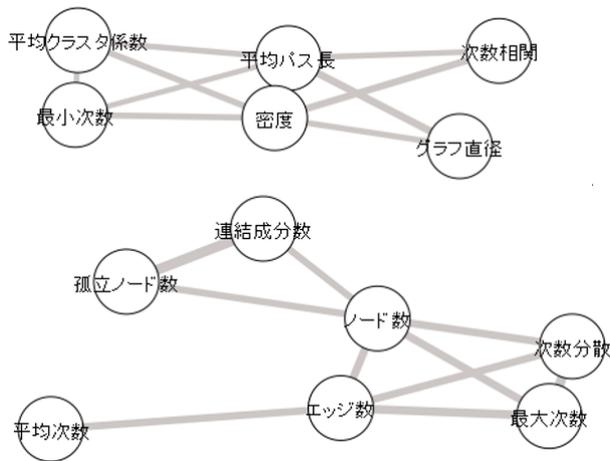


図 1 チャネルグラフの基本特徴量間の共分散可視化

Fig. 1 Visualization of covariance between basic features of a channel graph.

したがって、チャネルグラフを分析するための軸となるソースグラフ集合は、サイズおよび密度が異なる様々なグラフの集合であることが望ましい。BA モデルでは、このサイズおよび密度を生成アルゴリズムの引数として扱うことができる。具体的には、3.3 節で論じた BA モデルの生成アルゴリズムにおいて、ノード数とノード追加時のエッジ数という 2 つの引数に対してその値の候補集合をそれぞれ  $A, B$  とし、そのときの候補の数をそれぞれ  $|A|, |B|$  とおく。このとき、 $|A| \times |B|$  個のすべての組合せから、ソースグラフとなるグラフを生成する。このように生成されたグラフの集合は、チャネルグラフの重要なネットワーク特徴量をとらえており、第 1 の要件を満たす。また、それぞれサイズおよび密度が異なっており、第 2 の要件を満たすことが期待できる。

#### 4.3 提案アルゴリズム

BA モデルの引数の候補集合  $A, B$  をあらかじめ適切に設定した場合に、ターゲットグラフの埋め込み表現  $\Psi(G_n)$  ( $n = 1, \dots, N$ ) を学習する提案モデルのアルゴリズムを以下に示す。

- STEP1**  $A, B$  の要素すべての組合せに対し、BA モデルに基づき、 $M = |A| \times |B|$  個のソースグラフ集合  $\{S_1, \dots, S_M\}$  を得る。また、ターゲットグラフのインデックスを  $n = 1$  として初期化する。
- STEP2** 式 (1) および式 (4) から、ターゲットグラフ  $G_n$  の埋め込み表現  $\Psi(G_n) \in \mathbb{R}^M$  を得る。
- STEP3**  $n = n + 1$  とする。  $n \leq N$  であれば STEP2 に戻る。  $n \geq N + 1$  であれば終了する。

## 5. 実データ分析

ここでは、提案手法の有用性を検証するため、実データ

を用いた分析を行い、その結果を示す。具体的には、本研究の提案手法によるグラフの埋め込みが、従来手法である DDGK よりも各グラフの特徴をとらえたものになっているかを確認することに加え、実際の企業で収集された Slack の会話履歴データから実際にどのような情報が抽出できるかを明らかにする。

### 5.1 分析対象データとパラメータの設定

分析に用いた会話履歴データは、社員数 343 名、総チャンネル数 14,786 件の某企業内 Slack である。このデータに対して、4.1 節で示したチャネルグラフの構築手法を適用し、チャネルグラフ集合  $\{G_1, \dots, G_N\}$  を構築した。構築されたグラフ集合は、総グラフ数 111、平均ノード数 7.52、平均エッジ数 7.62 となった。また、ノード数の標準偏差値は 4.01、エッジ数の標準偏差値は 6.86 となった。

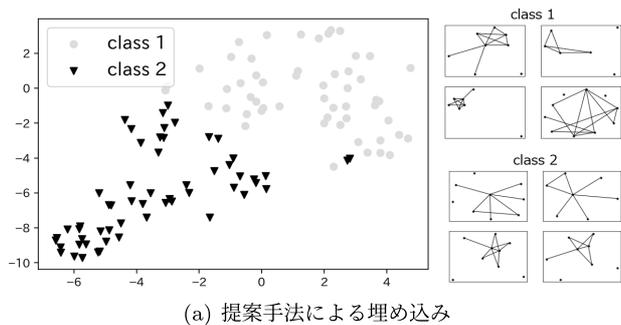
また、提案手法と従来手法の詳細な実験設定は以下のとおりである。提案手法では、BA モデルの引数にあたるノード数および追加エッジ数の候補を  $A = \{6, 9, 15\}$ ,  $B = \{1, 2, 3\}$  として設定する。  $A, B$  のすべての組合せに対し、グラフを生成し、9 個のソースグラフを得る。従来手法 (DDGK) では、提案手法と同数の 9 個のソースグラフを、ターゲットグラフからのサンプリングによって得る。提案手法、従来手法とも学習に使用したエンコーダの層数は  $L = 4$ 、誤差逆伝搬法の学習率は  $\alpha = 0.01$ 、エポック数は  $e = 1,000$  とした。

### 5.2 相関係数行列の比較

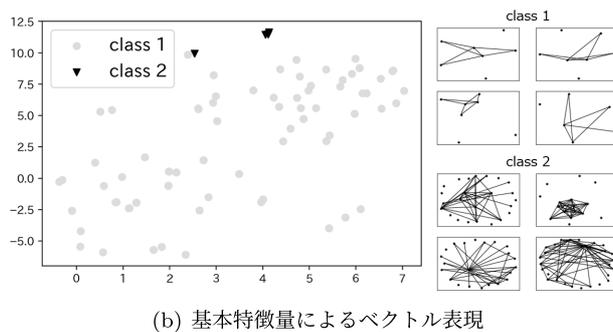
提案手法および従来手法で得られたチャネルグラフの埋め込み表現に対して、その相関係数行列を計算した結果を図 2、図 3 に示す。各セルの濃度が各特徴量の相関の強さを示している。図 3 (提案手法) では、濃い色のセルが少なく特徴量間に強い相関が出ていないのに対して、図 2 (従来手法) では、特徴量 2, 4, 7, 8 に対応するセルが比較的濃い色となっている。これは、ソースグラフをランダムサンプリングし、形状の似たグラフがサンプリングされた結果、形状の似たソースグラフに対応する軸どうしに相関が発生したためである。一方で、図 3 に示されたように、提案手法では異なる形状を持ったソースグラフによってターゲットグラフを埋め込んだ結果、各軸の相関が低くなったと考えられる。以上より、提案手法は従来手法と比較して、相関の低い埋め込み空間の軸を構成していることが分かる。

### 5.3 提案手法によるチャネルグラフの分析

チャネルグラフの特徴をベクトル表現するうえで最も単純な方法は、4.2 節で論じたネットワーク基本特徴量を計算することである。チャネルグラフの分析において、この単純な方法と提案手法を比較する。図 4 に示した t-SNE [25]



(a) 提案手法による埋め込み



(b) 基本特徴量によるベクトル表現

図 4 t-SNE による埋め込み空間の低次元可視化

Fig. 4 Low-dimensional visualization of embedding space by t-SNE.

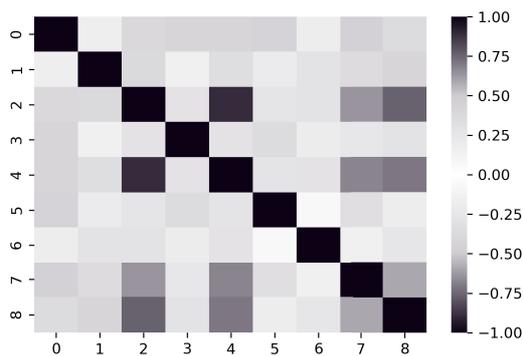


図 2 埋め込み表現の相関係数行列のヒートマップ (従来手法)

Fig. 2 Heatmap of the correlation coefficient matrix of the embedded representation (Conventional method).

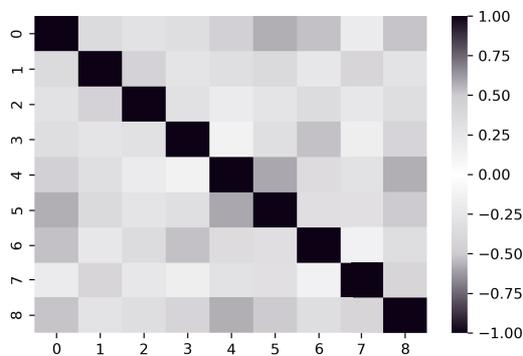


図 3 埋め込み表現の相関係数行列のヒートマップ (提案手法)

Fig. 3 Heatmap of the correlation coefficient matrix of the embedded representation (Proposed method).

による埋め込み空間の低次元可視化は、提案手法および基本特徴量セットによって計算されたチャネルグラフの特徴量ベクトルに対し、k-means 法 [26] によるクラスタリングを行い、t-SNE を用いて、2次元に写像・可視化したものである。k-means 法のクラス数は、 $k$  を 2 から 10 まで変化させて分析を行い、シルエット係数 [27] が最大となる  $k = 2$  を選択した。

また、各図の右部に可視化されているグラフは、各クラスを中心ベクトルから最も近い 4 つのデータ点に対応するチャネルグラフである。図 4(b) から、ネットワーク基本特徴量に対してクラスタリングを行うことで、少数のデー

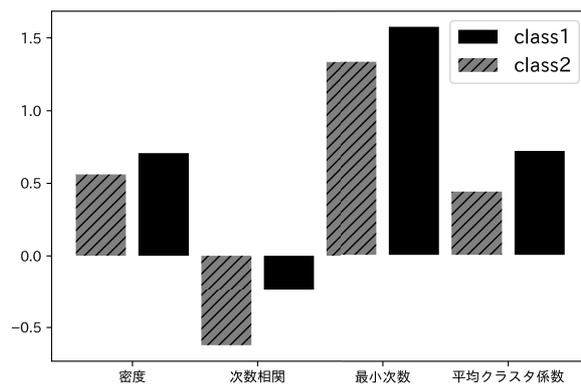


図 5 クラスごとのネットワーク基本特徴量の平均値比較

Fig. 5 Comparison of average values of network basic features by class.

タが所属するクラスと多数のデータが所属するクラスに分かれることが分かる。これは、ノード数やエッジ数が大きな少数のグラフ、すなわち、外れ値的なグラフが存在するためである。したがって、グラフ基本特徴量を単純に用いてベクトル表現するだけでは、意味のあるクラスを抽出することが難しいといえる。提案手法による埋め込み表現の図 4(a) では、図 4(b) と異なり、各クラスに所属するデータ数に偏りは存在しない。この抽出されたクラスが、ネットワーク基本特徴量によって特徴づけられ、解釈可能であることを以下に示す。

図 5 では、提案法の埋め込み表現から抽出されたクラス 1 およびクラス 2 に所属するチャネルグラフについて、ネットワーク基本特徴量 [24] を計算し、クラスごとの差が大きい基本特徴量の平均を図示した。図 5 および以下の各特徴量の性質から各クラスに所属するグラフは表 1 のように特徴づけられる。

(1) 密度：グラフ上のエッジ数を  $E$ 、ノード数を  $V$  とする。グラフ上に構築することができるエッジ数  $S$  は、 $V$  個のノードから任意の 2 つを選ぶすべての組合せ数  ${}_V C_2$  に等しい。密度は、 $S$  と  $E$  の比として計算される。密度が高ければ、グラフのサイズと比べて、多くのエッジが構築された密なネットワークとなる。

(2) 次数相関：隣接するノードどうしの次数の相関によ

表 1 各クラスの基本特徴量による特徴づけ  
Table 1 Characterization of each class by basic features.

	クラス 1 に所属するグラフ	クラス 2 に所属するグラフ
(1) 密度	密なネットワーク	疎なネットワーク
(2) 次数相関	隣接するノード間の次数は等しい傾向	中心的なノードに次数の小さなノードが結合する傾向
(3) 最小次数	次数の小さな疎外されたノードが少ない	次数の小さな疎外されたノードが多い
(4) 平均クラスタ係数	三角形的な関係が発生しやすい	三角形的な関係が発生しにくい

り計算される。次数の高いノードに次数の低いノードが結び付くグラフであれば次数相関は低くなり、隣接するノードどうしの次数が近ければ次数相関は高くなる。

(3) 最小次数：グラフ中で最も次数の低いノードの次数である。4.2 節の方法により構築したグラフは、隣接行列の平均値でエッジを剪定しており、次数が 0 のノードが発生しやすく、ほとんどのグラフで最小次数の値が 0 となる。よって、元のグラフから次数 0 のノードを除くため、最もノード数が多くなるように連結した部分グラフを抽出し、この部分グラフの最小次数を特徴量として計算した。最小次数が大きければ、次数の小さな疎外されたノードが少ないといえる。

(4) 平均クラスタ係数：ノード  $v_i$  のクラスタ係数は、ノード  $v_i$  に隣接するノード  $N(v_i)$  について、 $N(v_i)$  間に実際に構築されたエッジ数と  $N(v_i)$  から任意の 2 つを選んで構築できるエッジ数との比で計算される。平均クラスタ係数は、グラフ上のすべてのノードに対して、上記のクラスタ係数を算出し、平均したものである。グラフ上に 3 つのノードが相互に接続された三角形的な関係が発生しやすければ、平均クラスタ係数が高くなる。

会話に積極的な社員が多いチャンネルの場合、その会話履歴から構築されるグラフには多くのエッジが構築され密度が高くなり、いずれの参加者でも、ある程度の参加者と結び付いて最小次数が高くなる。また、少数の参加者に会話が集中することが少ないため、隣接するノードどうしの次数すなわち次数相関が高くなる。加えて、多くのエッジが構築された結果、三角形的な関係性が発生し平均クラスタ係数が高くなる傾向にあるといえる。このような密でフラットな会話が行われているチャンネルのグラフがクラス 1 として抽出されている。逆に、疎で 1 人集中型の会話が行われているチャンネルのグラフはクラス 2 で抽出されているといえる。

## 6. 考察

### 6.1 グラフ埋め込み手法の検討

本研究では、DDGK におけるソースグラフを適当に設定することで、チャットアプリ上のコミュニケーション分析に適したモデルを提案した。グラフ埋め込み表現手法は DDGK 以外にも様々なモデルや手法が提案されている。

DDGK 以外の多くの手法は、グラフの同型性を判定する

ために、Weisfeiler-Lehman カーネル [28] を用いている。たとえば、代表的なグラフ埋め込み手法である Graph2vec [29] は、Weisfeiler-Lehman 再ラベル操作によって得られた根付き部分グラフに対して、Document2vec のモデル構造 [30] を用いたものである。

しかし、Weisfeiler-Lehman カーネルは、グラフの同型性を判定する有力な手段であるものの、ノードラベルがない場合に適応が困難であるという問題点がある。本研究の対象問題である、チャンネル上のコミュニケーションの分析においては、ノードラベルは、各社員の個人情報に該当する 경우가多く、データとしてノードに対する何らかの属性情報が得られていることを前提とするのは実務上現実的ではない。したがって、本研究では、グラフ埋め込み表現の学習にノードラベルの存在を前提としない DDGK をベースにチャンネルグラフの分析に有用な埋め込み表現空間の構成を提案している。

### 6.2 提案手法のランダム性に関する考察

本研究では、ノードの数（社員数）が比較的少ないネットワークに焦点を当て、社員間コミュニケーションの分析のためのチャンネルグラフ構造の類型化を行っている。その際、BA アルゴリズムに基づきソースグラフをランダムに生成している。一般的には、BA モデルは確率的にグラフを生成するため、実行のたびにソースグラフの形状が変化すると考えられ、このランダム性を考慮する必要がある。

これに対して、本研究ではノードの数がある程度少ない場合という前提条件を設定しているため、このソースグラフのランダム性の影響は小さく、問題となっていない。すなわち、ノードの数が多くなれば生成されるグラフの種類数も多いため、生成するグラフの数をある程度増やすことで、これらの組合せは網羅されると考えられる。そのため、今回の前提条件においては、BA モデルの生成グラフのランダム性、およびその影響は大きくないと考えられる。

一方で、ノードの数が大きくなると、膨大な種類のグラフが BA アルゴリズムによって生成されるため、乱数のシード値によって生成されるグラフの集合に大きな違いが出てくることが考えられる。すなわち、本研究の提案法を、よりサイズの大きいチャンネルを有する大企業に適用した場合の挙動については検証の余地がある。この問題について

の検討は、今後の課題とする。

### 6.3 分析するデータに関する考察

本研究を実施した時期に、Slack においてスレッド機能がリリースされている。そのため、収集されたデータは Slack にスレッド機能がなかった時期のものであり、発言の時系列のみによってしか会話の連続性についてとらえることができなかった。しかしながら、スレッド機能に関するデータを活用することで、現実に即した分析が可能になることが考えられる。具体的には、同じスレッド内の会話については、より重視するようなグラフ作成の手法へと拡張していくことが考えられ、この点も今後の課題とする。

### 6.4 実験に対する考察

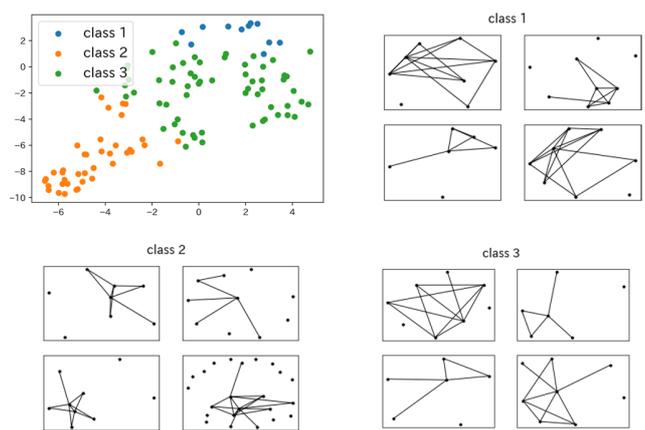
#### 6.4.1 ハイパーパラメータの設定について

BA モデルにおける引数候補の集合は、埋め込み対象のグラフのノード数やエッジ数の統計量をとることで、分析者が発見的に決定するパラメータである。本実験では、引数の候補を  $A = \{6, 9, 15\}$ ,  $B = \{1, 2, 3\}$  と設定した。これは、構築されたチャネルグラフの平均ノード数 7.52, 平均エッジ数 7.62 から発見的に決定している。このように設定することで、図 6(a) に示されるとおり、相関の低い埋め込み空間の軸の構成が可能となっていることから、このパラメータの設定は妥当であるといえる。

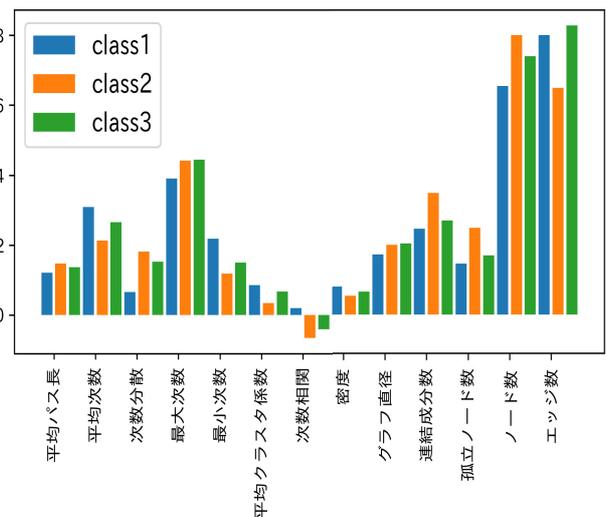
#### 6.4.2 クラス数を増やした場合の埋め込み空間

5.3 節では、k-means 法を適用する際のクラス数をシルエット係数が最大となるように  $k = 2$  として決定した。本項では、このクラス数を変化させた場合にも、提案手法を用いて構築された埋め込み空間であれば、同様の結論を引き出せることを示す。図 6(b) は、実験と同様のデータに対して、k-means 法のクラス数を 3 として適用し、抽出されたクラスごとにネットワーク基本特微量の平均値をプロットしたものである。図 6(b) から、クラス 1 に所属するネットワークは、5.3 節におけるクラス 1 のネットワーク、すなわち、密度、次数相関、最小次数、平均クラスタ係数が高く、密でフラットなコミュニケーションをとるタイプのチャネルに対応することが分かる。逆に、クラス 2 に所属するネットワークは、5.3 節におけるクラス 2 のネットワーク、すなわち密度、次数相関、最小次数、平均クラスタ係数が低く、疎で 1 人集中型のコミュニケーションをとるタイプのチャネルに対応する。新しく抽出されたクラス 3 のネットワークは、この 2 つのクラスの間位置するクラスタであり、両方クラスの性質を持っていると考えられる。

以上より、クラス数を増加させた場合でも、提案手法により得られた埋め込み空間から得られる結論は、クラス数が少ない場合から得られる結論と同種のものであり、クラス数を増加させた分、より詳細な類型化がなされるといえる。



(a) 埋め込み空間の低次元可視化と各クラスのサンプル



(b) 基本特微量によるベクトル表現

図 6 クラス数を 3 とした場合のチャネルグラフの分析

Fig. 6 Analysis of the channel graphs when the number of classes is set as 3.

このようにクラス数の増加にともない、より詳細なチャネルの類型化がなされる傾向にあるが、どの程度詳細に類型化を行えばよいかを定量的に決定することが望ましい。これは、k-means 法のクラス数を得られた埋め込み表現に合わせて、定量的に決定することに相当する。5.3 節では、この決定に必要な指標として、得られたクラス内のデータがどの程度密にクラスタリングされているかを図る指標であるシルエット係数を最大化させるように決定している。

### 6.5 本手法の組織マネジメント上の有用性

5.3 節における分析結果は、各チャネルのチームマネージャーに対する現状の可視化や意思決定支援に有用であると考えられる。たとえば、自身がマネジメントするチームのチャネルグラフのクラスが 1 人集中型である場合、よりフラットな関係性を目指し、メンバー間が気軽に意見の出せる雰囲気作りのための施策を打つ動機となりうる。また、同一チャネルの施策実施以降の会話履歴データから構築さ

れたチャネルグラフを同一空間内に埋め込むことで、施策の結果、実際にフラットな関係性が構築できたかを確認できる。

このような組織の構造に着目した分析が可能になることにより、ビジネスチャット上のデータに基づくマネジメントへの活用が期待される。

## 7. まとめと今後の課題

本研究では、ビジネスチャットアプリ上のコミュニケーショングループで形成される会話状態をグラフによって表現し、その埋め込み表現を分析するための手法を提案した。具体的には、グラフ埋め込み表現を学習するうえで汎用性の高い手法である Deep Divergence Graph Kernel に着目し、そのソースグラフの構成方法として、BA モデルによって生成されたグラフにより構成することを提案した。さらに、実際の Slack 会話履歴データを用いて、このような構成方法をとることによって、埋め込み空間の各軸の相関が小さくなり、より望ましい埋め込み空間が構築できることを確認した。また、各チャネルグラフの埋め込み表現に対してクラスタリング手法を適用することで、コミュニケーション状況の類型化が可能であることを示した。

今後の課題としては、部署・プロジェクト名などのチャネル属性情報を活用することがあげられる。それらの属性情報を活用し、埋め込み表現のクラスタリングから得られるグラフ形状に関する情報とあわせて統合的に分析することで、さらに有用な知見が得られる可能性がある。

**謝辞** 本研究を行うにあたり、貴重なデータをご提供いただいた Laboratik 株式会社に深く感謝の意を表します。また、本研究の一部は JSPS 科研費 21H04600 の助成を受けたものです。

## 参考文献

- [1] 牛丸 元, 穴澤 務, 山田仁一郎, 神吉直人, 山下 勝: 組織ネットワーク分析の可能性を探る: 理論的展開と経営行動への応用, 経営行動科学学会年次大会発表論文集, pp.339–344 (2008).
- [2] Yang, H., Luo, J., Liu, Y. Yin, M. and Cao, D.: Sensible Organizations: Changing Our Businesses and Work Styles through Sensor Data, *2010 3rd International Conference on Biomedical Engineering and Informatics*, pp.3041–3045 (2007).
- [3] Laboratik Inc., available from (<https://neworg.laboratik.com/>)(accessed 2022-03-22).
- [4] Ilag, B.N.: *Introducing Microsoft Teams*, Springer (2018).
- [5] Saima, B. and Naem, K.: A Survey of Data Clustering Methods, *International Journal of Advanced Science and Technology*, Vol.113, pp.113–142 (1993).
- [6] Hong, Y.C., Zheng, V. and Chang, K.: A Comprehensive Survey of Graph Embedding: Problems, Techniques, and Applications, *IEEE Trans. Knowledge and Data Engineering*, Vol.30, pp.1616–1637 (2018).
- [7] Al-Rfou, R., Zelle, D. and Perozzi, B.: DDGK: Learning Graph Representations for Deep Divergence Graph Kernels, *Proc. 2019 World Wide Web Conference (WWW '19)*, pp.37–48 (2019).
- [8] Barabási, A.L. and Albert, R.: Emergence of Scaling in Random Networks, *Science*, Vol.286, pp.509–512 (2019).
- [9] Naraine, M.L., Wear, H.T. and Whitburn, D.J.: User Engagement from Within the Twitter Community of Professional Sport Organizations, *Managing Sport and Leisure*, Vol.24, No.5, pp.275–293 (2019).
- [10] Kwak, H., Lee, C., Park, H. and Moon, S.: What is Twitter, a Social Network or a News Media?, *Proc. 19th International Conference on World Wide Web (WWW '10)*, pp.591–600 (2010).
- [11] Jianshu, W., Ee-Peng, L., Jing, J. and He, Q.: Twitter-Rank: Finding Topic-Sensitive Influential Twitterers, *Proc. 3rd ACM International Conference on Web Search and Data Mining (WSDM '10)*, pp.591–600 (2010).
- [12] Ara, K., Kanehira, N., Olguín, D., Waber, B., Kim, T., Mohan, A., Gloor, P., Laubacher, R., Oster, D., Pentland, A. and Yano, K.: Sensible Organizations: Changing Our Businesses and Work Styles through Sensor Data, *Journal of Information Processing*, Vol.16, pp.1–12 (2010).
- [13] 野中賢也, 山下 遥, 堀田 創, 後藤正幸: 社員間コミュニケーションの類型化を可能とするグラフ構造の分散表現モデル, 日本計算機統計学会シンポジウム論文集, Vol.34, pp.13–16 (2020).
- [14] Jitesh, S. and Jafar, A.: Discovering Important Nodes through Graph Entropy the Case of Enron Email Database, *Proc. 3rd International Workshop on Link Discovery*, pp.74–81 (2005).
- [15] Qian, R., Zhang, W. and Yang, B.: Detect Community Structure from the Enron Email Corpus Based on Link Mining, *Proc. 6th International Conference on Intelligent Systems Design and Applications*, pp.850–855 (2006).
- [16] Traud, A.L., Mucha, P.J. and Porter, M.A.: Social Structure of Facebook Networks, *Physica A: Statistical Mechanics and its Applications*, Vol.391, No.16, pp.4165–4180 (2012).
- [17] Xin, Z., Lai, C.M., Chapman, J.W., Barnett, G. and Wu, S.F.: Multi-view Community Detection in Facebook Public Pages, *International Computer Symposium 2018*, pp.131–138 (2018).
- [18] Mellah, M.A., Amine, A., Hamou, R.M. and Kumar, A.V.: Link Analysis for Communities Detection on Facebook, *International Journal of Data Mining and Emerging Technologies*, Vol.4, No.1, pp.16–24 (2014).
- [19] Massey, P.M., Kearney, M.D., Hauer, M.K., Selvan, P., Koku, E. and Leader, A.E.: Dimensions of Misinformation about the HPV Vaccine on Instagram: Content and Network Analysis of Social Media Characteristics, *Journal of Medical Internet Research*, Vol.22, No.12, e21451 (2020).
- [20] 山口 哲, 武田英明, 市瀬龍太郎, 大向一輝, 原誠一郎, 千葉大作: 複数の業務メーリングリストからの企業内ソーシャルネットワーク分析, 2006 年度人工知能学会全国大会論文集, p.242 (2006).
- [21] 野中賢也, 山下 遥, 堀田 創, 後藤正幸: ビジネスチャットアプリ上のコミュニケーションデータに基づくネットワーク構築手法, 人工知能学会論文誌, Vol.37, No.2, E-L63, pp.1–11 (2006).
- [22] Kawakami, T., Yamashita, H., Hotta, H. and Goto, M.: Visualization Analysis of Relationships between Employees Focusing on Content of Communication on Business

- Chat, 第 44 回情報理論とその応用シンポジウム予稿集, 1.4.1 (2021)
- [23] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I.: Attention Is All You Need, arXiv, 1706.03762 (2017).
- [24] Costa, L.F., Rodrigues, F.A., Traverso, G. and Villas Boas, P.R.: Characterization of Complex Networks: A Survey of Measurements, *Advances in Physics*, Vol.56, pp.1167–242 (2007).
- [25] Maaten, L. and Hinton, G.: Visualizing Data using t-SNE, *Journal of Machine Learning Research*, Vol.9, pp.2579–2605 (2008).
- [26] Xin, J. and Jiawei, H.: K-Means Clustering, *Encyclopedia of Machine Learning*, pp.563–564 (2010).
- [27] Rousseeuw, J.R.: Graphical Display, Cluster Analysis, Clustering Validity, Classification, *Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis*, Vol.20, pp.53–65 (1987).
- [28] Shervashidze, N., Pascal, S., Leeuwen, E., Mehlhorn, K., Borgwardt, M.K.: Weisfeiler-Lehman Graph Kernels, *Journal of Machine Learning Research*, Vol.12, pp.2539–2561 (2011).
- [29] Narayanan A., Chandramohan, M., Venkatesan, R., Chen, L., Liu, Y. and Jaiswal, S.: Graph2vec: Learning Distributed Representations of Graphs, ArXiv, 1707.05005 (2017).
- [30] Le, Q. and Mikolov, T.: Distributed Representations of Sentences and Documents, *Proc. 31st International Conference on Machine Learning*, pp.1188–1196 (2014).



野中 賢也

2019 年一橋大学卒業。2021 年早稲田大学大学院創造理工学専攻修士課程修了。在学時には機械学習手法を活用したビジネスデータ分析に関する研究に従事。



山下 遥

2010 年東京理科大学卒業。2012 年慶應義塾大学大学院博士前期課程修了。2015 年同大学院博士後期課程修了。博士 (工学) 取得。2015 年早稲田大学助手。2017 年上智大学助教。2022 年同大学准教授。機械学習のアプローチ

に基づくビジネスアナリティクスモデルの構築, 統計的手法に基づく農業データの分析, 深層学習に基づくスポーツアナリティクス等に興味を持つ。経営工学会, 品質管理学会, 応用統計学会各会員。



三浦 豊史

ラボラテック株式会社代表取締役社長。2004 年にニューヨーク市立大学芸術学部卒業後, 現地のクリエイティブエージェンシー R/GA New York でデザイナーとして勤務。2007 年に帰国後は, Google でインダストリーマネージャーとして AdWords や YouTube の広告営業・コンサルに携わる。2015 年に同社退社後 Laboratik Inc. 創業。早稲田大学大学院商学研究科修了。



後藤 正幸 (正会員)

1994 年武蔵工業大学大学院修士課程修了。2000 年早稲田大学大学院博士課程修了。博士 (工学)。1997 年早稲田大学理工学部助手。2000 年東京大学大学院工学系研究科助手。2002 年武蔵工業大学環境情報学部情報メディア学科助教授。2008 年早稲田大学創造理工学部経営システム工学科准教授。2011 年同大教授。情報数理応用とデータサイエンス, ならびにビジネスアナリティクスの研究に従事。著書に、『入門パターン認識と機械学習』, コロナ社 (2014), 『データサイエンス入門』, 有斐閣 (2022), 『ビジネス統計のための基礎理論』, オデッセイコミュニケーションズ (2023) 等。IEEE, INFORMS, 電子情報通信学会, 人工知能学会, 日本経営工学会, 経営情報学会等各会員。本会シニア会員。