

令和2年度 日本学生支援機構 学生支援の推進に資する調査研究事業（JASSO リサーチ）

研究成果報告書

官民協働海外留学創出プロジェクト（トビタテ）における
コミュニティ形成力学と潜在的な学生間類似性の可視化研究

2021年2月28日

研究代表者：後藤 弘光

金沢学院大学経済情報学部講師

目次

1	はじめに	2
2	ネットワークによるつながりデータのモデル化	3
3	オンラインコミュニティを活用した学生支援の可能性	5
4	留学計画から探る潜在的な学生間のつながり可視化	13
5	おわりに	23

1 はじめに

官民協働海外留学支援制度「トビタテ！留学 JAPAN 日本代表プログラム」(以下、「トビタテ」という。)[1]は、「日本再興戦略」(平成 25 年 6 月 14 日閣議決定)に基づき、2020 年までに日本学生の海外留学倍増という政府の目標の下、官民が協力して海外留学を支援するために創設された「グローバル人材育成コミュニティ」が運営するものである。トビタテは、企業インターンシップや学生自らが立案したプロジェクト等、「実践活動」に焦点を当てた留学を推奨し、多様な経験と自ら考え行動できるような体験の機会を提供、多様な経験を積んだ個性溢れる留学生のネットワークを形成することで、産業界を中心に社会で求められる人材、世界で、又は世界を視野に入れて活躍できる将来のグローバルリーダーの育成を目指している。トビタテの主な支援内容は以下の3点である、

- 海外留学費用(奨学金、留学準備金、授業料)の給付
- 留学事前・事後に行う研修の提供
- 継続的な学習や交流の場としての留学生ネットワークの提供

学生自らが立案した留学計画に対する金銭的支援に加え、留学事前事後の合同研修やコミュニティを活用した学習支援を含む点は、国費による単位取得型や協定派遣型の海外留学支援制度とは大きく異なる。2014 年よりスタートしたこのトビタテは、2020 年 1 月までに約 6 千人の高校・大学生の海外挑戦を支援し、留学という共通点を通じて、所属や留学国、地域や専門性の垣根を越えた交流の場、多様な人脈形成の機会を提供している。2020 年 1 月以降、新型コロナウイルス感染症のパンデミックが世界中に広がり、派遣留学生の一時帰国や留学の中止など大きな影響を及ぼしている。一方、オンライン上でのコミュニケーションツールの活用により、留学事前・事後研修のオンライン化やオンラインでの勉強会など、実際の留学だけではない、「継続的な学習や交流の場としての留学生ネットワークの提供」として、学生の学びに対する支援が継続されている。

トビタテに参加する学生は全国・全世界で活動しており、一般にその交流は地理的な条件によって制限されていた。一方で、ソーシャル・ネットワーキング・サービス(SNS)などのオンライン上でのコミュニケーションツールは、これらの地理的な制限を克服するだけでなく、新型コロナ禍における学生支援のためのプラットフォームとなっている。Facebook は世界最大の SNS であり、ユーザーが実名登録された個人アカウントを使用する特徴を持つ。そして、ユーザーは友人関係を持つ友人と Facebook 上でも友達として関係を形成することができ、Facebook 上で友達になることで、その友人の投稿がホーム画面に表示されるほか、Facebook 上でメッセージのやり取りが可能となる。また投稿に対する「いいね」や「コメント」によってユーザー同士の交流を深めたり、投稿のシェアによって情報の拡散をしたりすることができ、グループ機能を活用することで、グループ内に限定した情報共有やイベントの開催などコミュニティでの活動を推進することができる。トビタテは 2014 年よりスタートし、同時期にトビタテの参加学生のオンライン上での交流や情報伝達の場として、承認制の Facebook グループが発足した。2020 年までに数千規模のトビタテ参加学生が加入するオンラインコミュニティとして成長している。さらに、2020 年 1 月には日本学生支援機構グローバル人材育成部(以下、「トビタテ事務局」という。)の主導で、留学後の継続的な成長促進と支援企業等との接点、コラボレーションの創出を狙いとした「トビタテ学習プラットフォーム」が発足し、有識者やスポンサー企業、トビタテに参加した学生が登壇するセミナー等のイベントを定期的開催、承認制の Facebook グループ内において、企画の案内やライブ配信動画の配信・公開が行われている。これらコミュニティにおける力学を調査・分析し、学生支援との相関を明らかにすることは、新型コロナ禍における困難を克服し、データ駆動型支援を推進する上で非常に重要であろう。

ネットワーク科学^{*1}は、インターネットや企業間の取引、食物連鎖や道路網などの複雑な「つながり」からなる現象を、ネットワーク（グラフ）を用いてモデル化することで研究する学術分野であり、1998年の現実世界で広く観測される「世間は狭い」性質を持つネットワークのモデル化 [2] を契機に、広い分野で活用されるようになった。近年、インターネットの普及や計算機能力の向上に伴い、主に社会学で扱われていた人間関係に関する研究にも広く用いられている。特に本調査と関連する Facebook における友人関係ネットワークの研究として、例えば、アメリカの大学 100 校の学生間の Facebook 上での友人関係形成における属性（性別、学年、高校や住居など）の役割及びコミュニティ調査 [3, 4] などが挙げられる。また教育機関における同窓会の実態調査のため、ウェブクローラーによってデータを収集、テクノロジー業界のリーダーに着目した同窓会ネットワークを可視化・分析した研究 [5] もある。日本における実践的かつ網羅的な初の研究として、筆者は令和元年度 JASSO リサーチ「官民協働留学創出プロジェクト（トビタテ）における友人関係ネットワークの成長とコミュニティ構造の可視化研究」[6] に取り組み、トビタテに参加した大学生の 7 割以上がオンラインコミュニティに参加していること、トビタテの学習支援の特徴でもある留学事前・事後研修によって、地理的近接性に依存しない派遣期毎の群れ構造が形成されていることを明らかにした。一方、申請コースなど専門性や興味関心に基づくつながり、派遣期を跨ぐつながりの乏しさが可視化されている。本調査研究はこれらの課題解決に向けて、コミュニティ内の交流機会であるイベントに着目し、その参加傾向から興味関心に基づくつながりを可視化する。さらに、学生の興味関心、専門性の豊富な情報を含むテキストデータである留学計画情報に対して、機械学習を用いた自然言語処理技術を応用することで、これまで数値化が困難であった潜在的な学生間のつながりを可視化することで、データ駆動型のコラボレーション創出に貢献する。

本報告書の構成は以下である。2 節において分析手法について説明する。3 節においてオンライン上の学生間の人脈構造の特性、イベントへの参加傾向分析を可視化し、オンラインコミュニティを活用した新たな学生支援の可能性を議論する。4 節において留学計画のテキストデータを分析することで、潜在的な学生間のつながりを可視化した結果を紹介する。最後に 5 節において本報告書をまとめる。

2 ネットワークによるつながりデータのモデル化

はじめに本調査研究における主な手法について説明する。ネットワークは、点（ノード）と線（リンク）で定義される、いくつかの点同士が繋がった構造体を指す。グラフ理論においては、ノードの集合 $V = \{v_1, \dots, v_N\}$ とノード $v_i - v_j$ 間のリンク集合 $E = \{(v_i, v_j)\}$ で表現される。例えば友人関係ネットワークは、ノードを人、2 人の間に友人関係がある場合にリンクを定義する。具体的には 4 人の友人関係がノードとリンクの集合、

$$V = \{v_1, v_2, v_3, v_4\}, \quad E = \{(v_1, v_2), (v_1, v_3), (v_1, v_4), (v_2, v_3)\}, \quad (1)$$

として与えられたとき、その友人関係ネットワークは図 1（左）として表現できる。一般に任意の 2 つのノード間の距離や関係の強さは、図 1（右）に示すようなリンクの重みとして表現することができる。

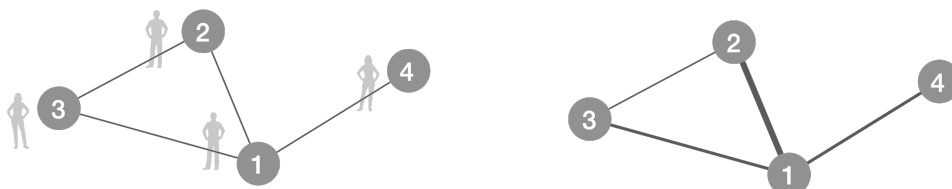


図 1: ネットワークの例。(左) 無向ネットワーク, (右) 重み付き無向ネットワーク。

^{*1} 複雑ネットワーク、ネットワークサイエンスなどとも呼ばれる。

本調査ではノードが持つ属性を用いた同類性によって、ネットワークにおけるリンクレベルの属性間の結合傾向を評価する。一般にノードは複数の属性を持ち、似たもの同士が結合しているか否かによってネットワークを特徴づけることができる。属性 a の要素 α, β 間の結合確率 $e_{\alpha\beta}$ を、全リンク数 $|E|$ における属性 a の要素 α, β 間に張られるリンク数の割合として次のように定義する、

$$e_{\alpha\beta}(a) = \frac{|E(a(v_i) = \alpha, a(v_j) = \beta)|}{|E|} . \quad (2)$$

また各属性に対する結合確率を用いて、ネットワーク全体としてリンクで結ばれた2つのノードが似る度合いを測る特徴量、同類性 [7] を計算できる、

$$r_a = \frac{\text{tr}(e(a)) - \|e^2(a)\|}{1 - \|e^2(a)\|} , \quad (3)$$

同じ属性が結び付いたネットワークであれば正の相関（親和的）を示し、そうでない場合には負の相関（排他的）を示す。例えば学生間のつながりを考える場合、派遣期や申請コースなどの属性において、同じ派遣期や申請コースでの友人関係が比較的多い場合には強い正の相関が期待される。

リンクによって結ばれた2つのノードの属性が一致する度合いを用いてネットワークの結合傾向である同類性を評価できる一方、複数の属性を持つ2つのノード間における属性の要素の一致傾向を用いて、ノード間の類似度を評価することもできる。2つノード v_1, v_2 が持つ複数の属性を一般にベクトル変数 $\mathbf{x}_1, \mathbf{x}_2$ と表すとき、コサイン類似度はこれらベクトル変数同士のなす角を用いて次のように定義できる、

$$\text{similarity}(\mathbf{x}_1, \mathbf{x}_2) = \frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{\|\mathbf{x}_1\| \times \|\mathbf{x}_2\|} . \quad (4)$$

コサイン類似度を用いることで任意の2つのノードのベクトル変数から重み付きネットワークを形成することができる。

本調査研究の概要

本調査研究の全体像を図2に示す。本調査研究は人と人のつながりである友人関係ネットワークに加え、イベント参加履歴や留学計画からトビタテ参加学生、ノードに対するベクトル変数を定義することで、コサイン類似度の基づくクラスタリング、重み付きネットワークにおける申請コースなどの興味や関心、専門性に関する同類性やコミュニティ構造を調査する。学生とイベント、学生間の潜在的なつながりを可視化することで、イベントのレコメンドや専門性や興味関心が類似した学生間のマッチングの実現可能性を示す。

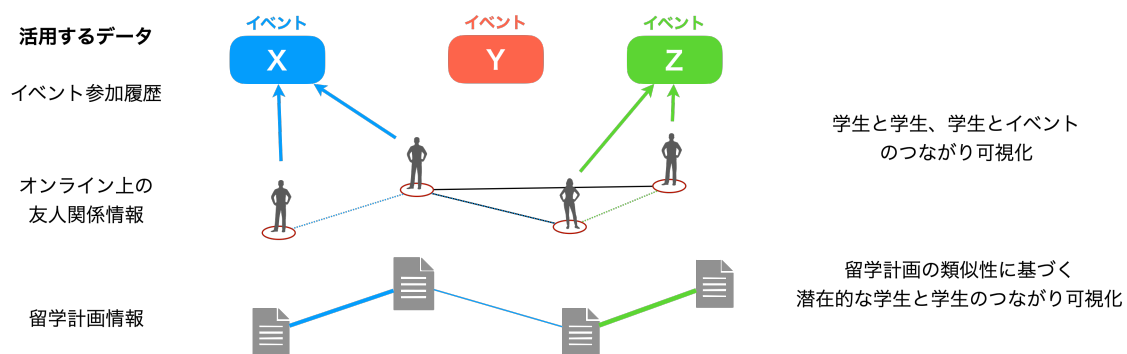


図2: 本調査研究の全体像.

3 オンラインコミュニティを活用した学生支援の可能性

本節ではオンライン上のデータを活用することで可視化される学生及びイベント間のつながりを分析した結果を報告する。具体的には、Facebook グループ内の友人関係情報に基づく学生と学生のつながりを可視化した令和元年度の調査研究の結果を概観する。トビタテ学習プラットフォームイベントの参加履歴を用いて、学生と学生、学生とイベント、イベントとイベントのつながりを可視化した結果を紹介し、オンラインコミュニティを活用した新たな学生支援の可能性を議論する。

オンライン上のトビタテ友人関係ネットワーク

ここではトビタテに参加する学生の主な交流の場である Facebook の友人情報を用いて人脈構造を調査した令和元年度調査研究の結果 [6] を振り返り、本調査研究で解決を試みる課題について言及したい。トビタテに参加した学生の同窓会組織とまりぎのホームページ [8] に記載のある Facebook グループのうち、メンバー数が最大のグループ（グループ ID: 1585059971706689）を主な調査対象とした。対象の Facebook グループには、トビタテ大学生等コース及び高校生コースに参加した学生、日本学生支援機構グローバル人材育成部（トビタテ事務局）の実務者、支援企業及び大学関係者が参加している。本調査ではトビタテ事務局より提供を受けた参加学生及び、トビタテ事務局員一覧情報を用いて、Facebook グループのメンバーを同定、人脈構造における特性を調査した。ただし、同定が困難であった参加学生、支援企業及び大学関係者は調査対象から除いている。本調査では以下の手順で Facebook グループにおける友人関係情報を取得した。

- Facebook グループ（ID: 1585059971706689）の参加者一覧を取得する。
対象の Facebook グループは非公開グループであるため、参加者一覧はグループ参加者のみ閲覧が可能である。したがって本調査では、対象グループ参加者の協力の下、グループの参加者情報を取得した。
- 取得した各メンバーの個人ページで公開されている友達一覧を取得する。
Facebook の個人ページで公開されている友達情報は、個人のプライバシー設定によって、友達のみ閲覧可能である場合があるため、本調査では、対象グループの友達数が 0 人である調査用アカウントを使用し、各グループ参加者の個人ページで一般公開されている友達一覧のみを取得する。このとき、グループの参加者数は数千規模であるため、参加者情報の取得と同日に友人関係情報を取得することは困難である。したがって本調査では、参加者一覧の取得日から友人関係の取得完了日までの友人関係の増加を許容することとした。
- Facebook グループ内の友人関係情報を抽出する。
Facebook グループの参加者一覧に含まれる者同士の友人関係情報のみ抽出することで、トビタテコミュニティにおける友人関係ネットワークを抽出する。このとき、参加者一覧の取得日から友人関係の取得完了日まで期間にグループに参加した者は、抽出したネットワークに含まれないことに注意。

上述の手順で 2019 年 4 月及び 11 月、2020 年 3 月及び 7 月の計 4 回、上述の手順にて友人関係情報を取得した。ただし、2019 年 4 月及び 11 月は令和元年度調査において取得したものである。

派遣期及び申請コース毎のトビタテ採択学生数と 2020 年 3 月及び、7 月時点での Facebook グループ参加人数を図 3、及び表 1、表 2 に示す。2020 年 7 月の時点で 4,394 人の採択学生が対象の Facebook グループに参加しており、大学生コースにおいて派遣期や申請コース毎で大きな偏りはない。一方、大学生コースのコミュニティ加入率は 75.3%、高校生コースについては 15.1% であり、この結果には大学生と高校生の Facebook 利用率の差が反映していると考えられる。また 2020 年 7 月時点において 67 人のトビタテ事務局員（研修講師を含む）が参加しており、参加学生に対する情報発信及びコミュニティ形成の援助を担っている。ただしトビタ

テ事務局員には、大学や企業からの出向者及び、非常勤職員も含まれるため、データ取得時において既にトビタテ事務局に在籍していない者も含む。

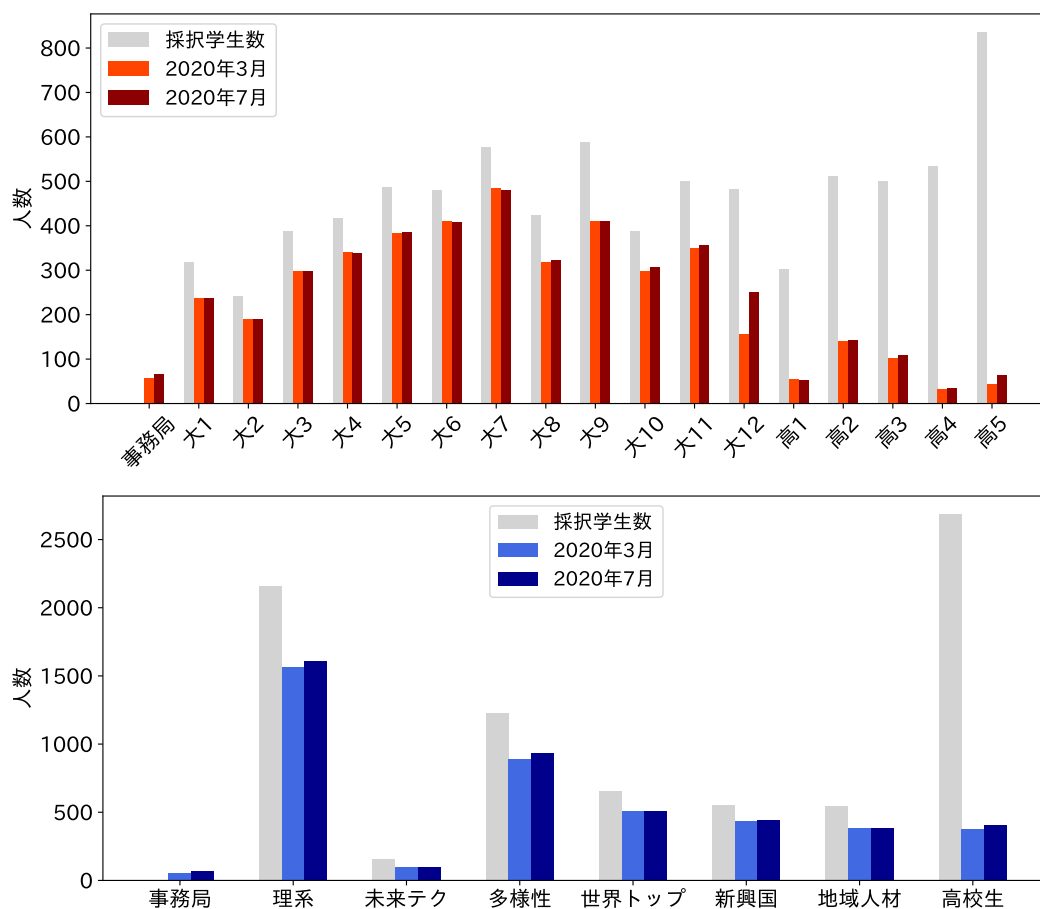


図 3: 派遣期及び申請コース毎のトビタテ採用学生数と 2020 年 3 月及び、7 月時点での Facebook グループ参加人数。派遣期 (上)、申請コース (下)。詳細はそれぞれ表 1、表 2 に示す。

Facebook 上の友人関係情報をもとに構成されるトビタテ友人関係ネットワークの群れ構造及び成長過程を可視化した結果を図 4 に示す。ただし、ネットワークの最大連結成分のみ可視化し、各ノードの位置座標はばね・電気モデル [9] を用いて決定した。ばね・電気モデルはネットワークをばね (リンク) によって結合した同符号電荷 (ノード) の力学系とみなし、系の定常状態によってノードの位置座標が決定される手法であり、グループ内の友人数が多い参加者ほど中心に配置される可視化手法である。図 4 (上) は 2020 年 7 月時点のトビタテ友人関係ネットワークを派遣期毎にノードの色を変えて可視化した結果である。非常に密な人脈構造が形成され、派遣期ごとにコミュニティが形成されていることが確認できる。また図 4 (下) に、2019 年 4 月から 2020 年 7 月までのトビタテ友人関係ネットワーク成長過程を可視化した。ただし比較のためノードの座標は図 4 (上) に固定している。2019 年 4 月から 11 月の間で新たな参加者群 (青色) が、2020 年 3 月及び 7 月の時点も同様に新たな参加者群 (緑色、紫色) が既存のコミュニティ群に接続する形で全体の最大連結成分が大きく成長していることがわかる。2019 年 11 月時点の新たな参加者群 (青色) の多くが、大学 10 期生及び 11 期生に対応しており、2020 年 3 月及び 7 月時点の新たな参加者群 (緑色、紫色) の多くが大学 12 期生に対応している。2020 年 1 月以降、新型コロナウイルスの影響により対面式の事前・事後研修等がオンライン化されたが、派遣期毎の交流機会が提供された結果、他の派遣期同様に群れ構造が形成されたと考えられる。

派遣期	採択学生数	2020年3月	2020年7月
事務局	–	57	67
大学1期	318	238	238
大学2期	241	190	191
大学3期	388	299	298
大学4期	417	340	339
大学5期	488	383	385
大学6期	480	410	409
大学7期	578	485	481
大学8期	424	319	322
大学9期	588	410	410
大学10期	389	299	307
大学11期	500	351	357
大学12期	482	157	252
高校1期	303	56	54
高校2期	511	140	142
高校3期	501	103	110
高校4期	535	32	34
高校5期	835	45	65

表1: 派遣期毎のオンラインコミュニティ参加者数状況.

申請コース	採択学生数	2020年3月	2020年7月
理系、複合・融合系人材コース	2160	1566	1612
未来テクノロジー人材枠	155	95	101
多様性人材コース	1225	891	936
世界トップレベル大学等コース	655	510	512
新興国コース	550	437	443
地域人材コース	548	382	385
高校生コース	2685	376	405

表2: 申請コース毎のオンラインコミュニティ参加者数状況.

2020年7月時点におけるトビタテ友人関係ネットワークにおけるリンクレベルの属性間の結合傾向を可視化するため、派遣期と申請コースに対する結合確率 $e_{\alpha\beta}$ を図5に示した。図5(左)の派遣期間の結合確率においては、同じ派遣期同士での結合確率がどの派遣期においても高く、参加率が少ない高校生コースの学生も同様の結果を示していることがわかる。一方、図5(右)の申請コース間の結合確率は、各コースの人数に比例するような振り舞いをしており、高校生コース以外には同じコース内での顕著な結合が無いことがわかる。図5の派遣期間及び申請コース間の結合確率を用いて同類性を計算した結果が以下である、

$$r_a = \begin{cases} 0.411 & a : \text{派遣期} \\ 0.139 & a : \text{申請コース} \end{cases} \quad (5)$$

トビタテの主な学生支援の一つである「留学事前・事後に行う研修の提供」によって、地理的近接性に依存

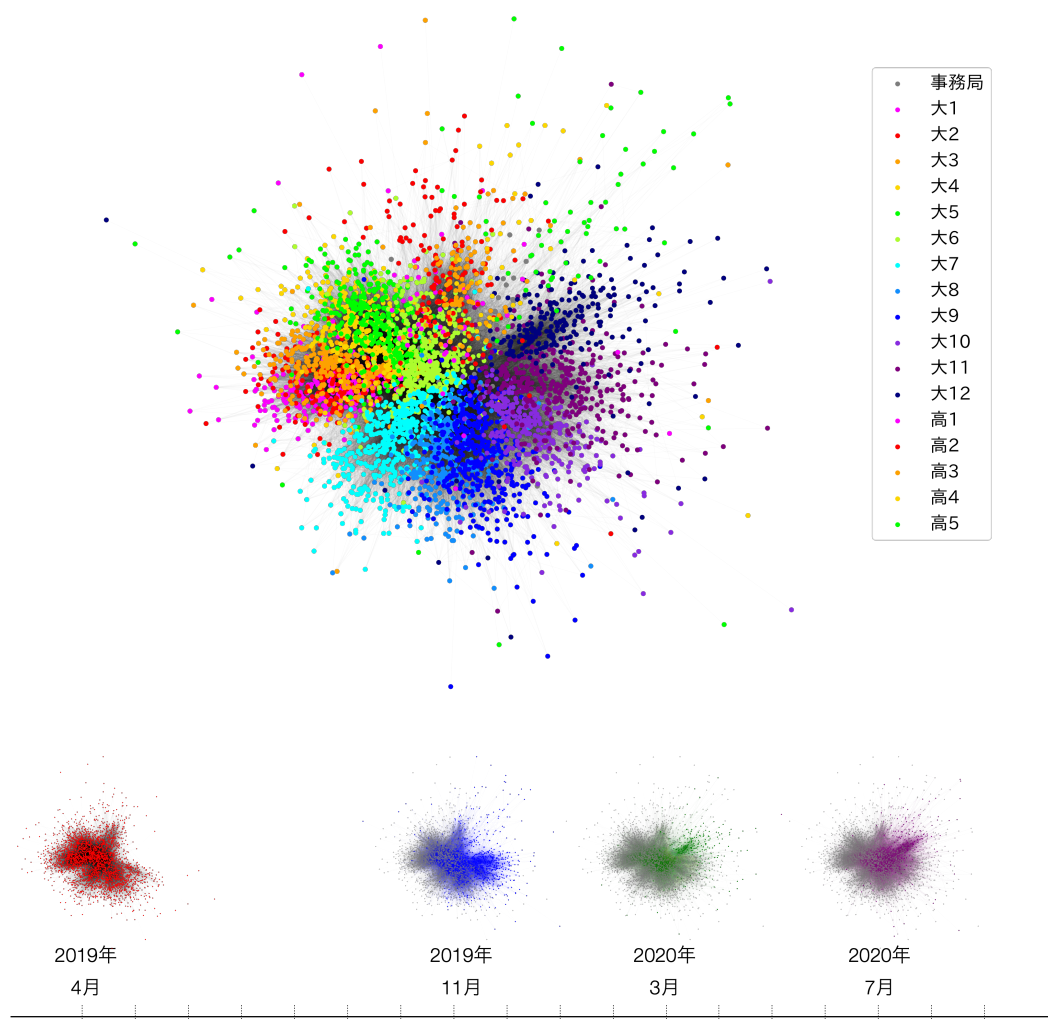


図4: トビタテ友人関係ネットワークの群れ構造及び成長過程の可視化。(上)参加学生の派遣期に基づいて色付けした2020年7月時点のトビタテ友人関係ネットワーク。(下)2019年4月から2020年7月までの成長過程の可視化。

しない多様な人脈構造が形成されたと期待される一方、派遣期を跨ぐ専門性や興味関心に基づくつながりは未だ乏しいことが明らかになった。本調査研究では派遣期を跨ぐ専門性や興味関心に基づくつながりを、多様なデータ活用することで抽出する手法を提案する。

イベント参加履歴から可視化される学生の興味関心

トビタテの主な学生支援の一つである「継続的な学習や交流の場としての留学生ネットワークの提供」は、国費による単位取得型や協定派遣型の海外留学支援制度とは大きく異なる点であった。実際、2020年1月にトビタテ事務局主導で発足した「トビタテ学習プラットフォーム」によって、留学後の継続的な成長促進と支援企業等との接点、コラボレーションの創出を目的として、有識者やスポンサー企業、トビタテに参加した学生が登壇するセミナー等のイベントがオンラインで定期的に行われていた。新型コロナ禍にも関わらず、トビタテ友人関係ネットワークにおいて大学11期生や12期の学生間につながりが形成されていた。これらは定

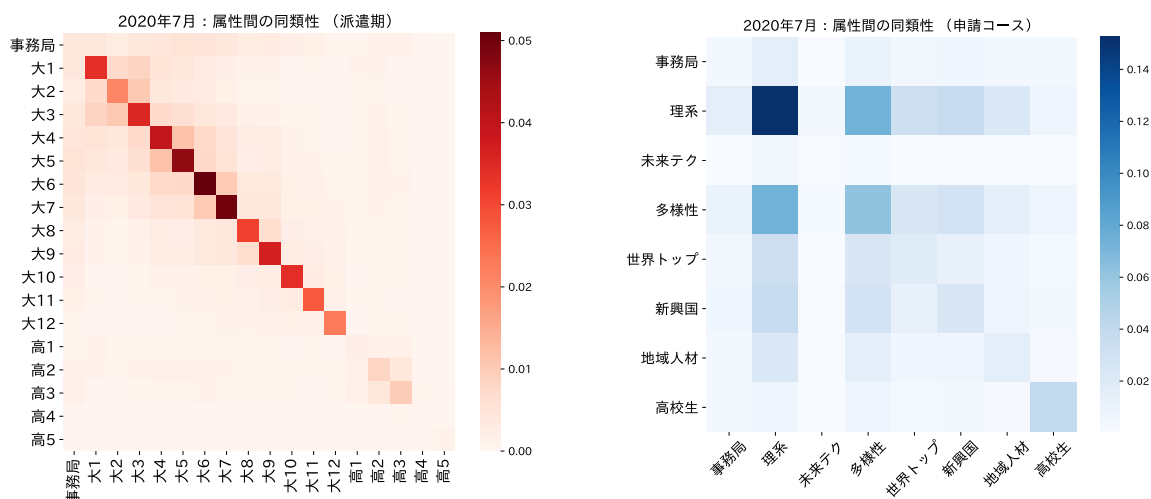


図 5: 2020 年 7 月時点におけるトビタテ友人関係ネットワークの各属性に対する要素間の結合確率。(左) 派遣期及び (右) 申請コース。

期的なイベントが開催されることで交流機会が創出された結果であると期待できる。

トビタテ学習プラットフォームによるセミナー等のイベント開催は、学生の学びの場の提供という具体的な学生支援であったと同時に、学生の興味関心をリアルタイムに把握する情報収集機会であったと考えられる。ここではどのような学生がどのようなイベントに参加したのか、学生とイベントのつながりのデータであるイベント参加履歴を用いて学生の興味関心を可視化した結果を報告する。

ジャンル	イベント数	参加人数	うち紐付け済
勉強会	20	590	492
起業家	8	735	629
基礎教養	4	495	433
リーダーシップ	6	411	338
グローバル	3	259	219
オープンイノベーション	3	223	202
SDGs	3	94	71

表 3: ジャンル別のイベント数及び累積参加人数の集計。

本調査ではトビタテ学習プラットフォームによって 2020 年 2 月から 2020 年 9 月までに開催された、47 回のイベントに対する学生の参加履歴を分析する。表 3 に本調査の対象としたジャンル別イベントのイベント数及び参加人数の集計を示す。イベントは内容によって分類されており、セミナー等の登壇者がトビタテ参加学生である勉強会が最も開催回数が多い。その他のジャンルのイベントは有識者やトビタテのスポンサー企業からセミナー講師が登壇する形式となっており、開催回数は少ない一方で、累積の参加者数は勉強会に比べて多くなっている。イベントは全てウェブ会議システムを用いてオンラインで開催されており、学生のイベント参加履歴情報はウェブ会議システムへのアクセス履歴をもとに、トビタテ事務局に所属するトビタテ学習プラットフォーム運営者によって作成されたものである。ただし、学生の個人照合はウェブ会議システムに用いた氏名及びメールアドレス、イベント開催時にアンケートによって収集した派遣期情報によって行われたため、ト

ビタテ事務局が保有する学生情報と紐付けられていないものが含まれる。本調査では、派遣期や申請コースなどの学生情報と紐付けられた学生のイベント参加履歴を用いて、イベントと学生のつながりを分析し、学生の興味関心の傾向を可視化する。ただし、本調査では大学生コースのみを対象とし、学生情報と紐付けられた学生数は756名である。

図6に派遣期及び申請コースに着目したイベントの累積参加者数を示す。申請コース毎の累積参加者数は、採択学生数と同じ傾向を持つ一方、派遣期毎には偏りが生じていることがわかる。特に大学7期と9期、11期と12期の参加者数が多いことが見て取れる。これらは大学7期以降の事前・事後研修においてトビタテ学習プラットフォームに関する周知活動を実施していた効果と考えられる。特に大学11期及び12期の累積参加数は顕著であり、これらは新型コロナの影響によって実際の留学の中止や延期が余儀なくされた結果、オンラインでの学びの機会に積極的に参加しようという傾向が現れたのではと考えられる。

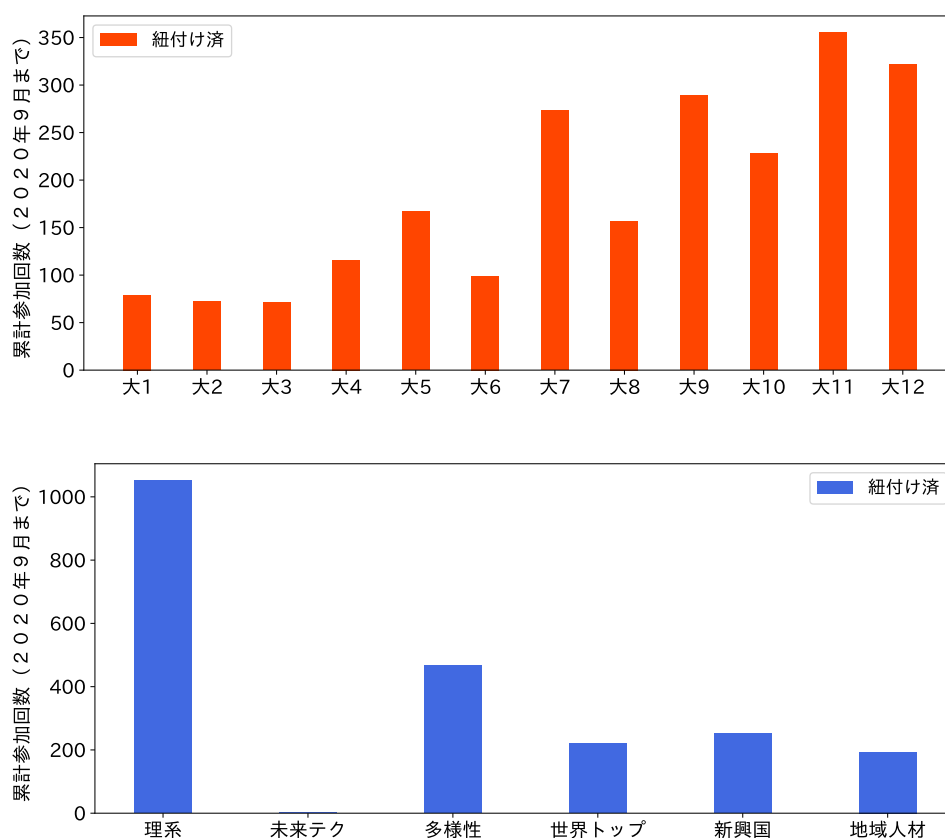


図6: イベントの累積参加者数。派遣期(上)及び申請コース(下)。

図7にイベント参加回数とトビタテ友人関係ネットワークにおける友人数の関係を示す。ただし、イベント参加履歴と学生情報と紐付けられた756名、友人数は2020年7月時点の結果である。イベントに1回以上参加した学生のイベント参加回数は、平均2.3回であり中央値は2回であり、2020年7月時点でのトビタテ友人関係ネットワーク友人数は、平均95.9人であり中央値は70.5人であった。またイベント参加回数と友人数は右裾の長い分布となっていることがわかる。イベント参加回数と友人数の相関係数は0.168であり、正ではあるがその相関は非常に弱いことがわかる。コミュニティに対する学生のニーズは多様であるため、これらの相関は必ずしも強くなく、学生間のつながりを求めている学生は多い友人数、学びの機会を求めている学生は多数のイベント参加回数として、それらのニーズが可視化されていると解釈できる。

学生のイベント参加履歴を用いて、学生とイベントのつながりに関する分析を行う。イベント参加履歴の類

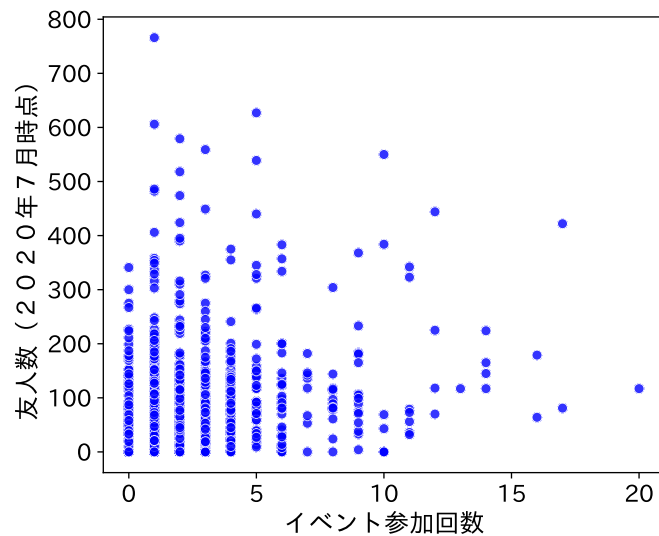


図 7: イベント参加回数とトビタテ友人関係ネットワークにおける友人数の関係。ただし、イベント参加履歴と学生情報と紐付けられた 756 名、友人数は 2020 年 7 月時点の結果。

似性は、学生間の興味関心の類似性を表すと共に、あるイベントに参加している学生はある特定のイベントにも参加しているという共起性は、イベント間の類似性を表す。ここではオンライン上の友人関係の分析と同様に、トビタテに参加する学生の派遣期及び申請コースに毎のジャンル別のイベント参加傾向を可視化する。

イベントの参加履歴を派遣期及び申請コース毎に集計し、派遣期とイベントのジャンルとの結合性を可視化した結果を図 8 と図 9 に示す。ヒートマップの各要素の数字は、ある派遣期または申請コースの学生があるジャンルのイベントに参加した回数を示す。例えば図 8 から大学 12 期の学生が勉強会イベントに累計で 116 回参加していることがわかる。これらある属性を持つ学生の参加回数を成分を持つベクトル変量を用いて、イベントのジャンル間及び属性間のコサイン類似度を計算し、イベントやある属性を持つ学生間の類似性に基づく階層的なクラスタリングの結果を樹形図を用いて可視化している。

図 8 より、トビタテ学習プラットフォームによって開催されているイベントに対する派遣期毎の興味関心の傾向は、およそ大学 1 期から 6 期とそれ以降で傾向が異なることがわかる。大学 6 期以降の学生は、勉強会や起業家、基礎教養及びリーダーシップなど、幅広いジャンルのイベントに参加する傾向が見て取れる。一方、大学 1 期から 6 期の学生は、基礎教養と起業家のジャンルに分類されるイベントへの参加傾向が高いことがわかる。大学 1 期から 6 期に分類されるイベント参加者は既に社会人になっている場合が多く、有識者やトビタテのスポンサー企業からセミナー講師が登壇する形式のイベントに興味や関心を示す傾向が結果として現れていることが考えられる。一方、大学 11 期と 12 期はイベント開催回数が多い勉強会への参加者数が特に多い傾向を持つ。勉強会はトビタテ参加学生がセミナー登壇者となるため、小規模であるが学生間の交流機会となるイベントが多い。大学 11 期と 12 期の学生は新型コロナウイルスの影響によって実際に留学が開始できていない場合が多く、オンラインでの学習や交流機会へのニーズが高かった結果であると解釈できる。一方、図 9 で可視化された申請コース毎のイベント参加傾向においては、特徴的なクラスタ構造は確認できなかった。

図 8 と図 9 で可視化されたイベントのジャンル間のクラスタ構造に着目すると、勉強会イベントに参加している学生は起業家イベントにも参加している傾向があることがわかる。これらイベント間の共起性は、勉強会と起業家だけでなく、基礎教養とリーダーシップなどにも現れており、これら学生とイベント間のつながりデータを活用することで、学生に対するイベントのレコメンドなどへの応用が期待できる。

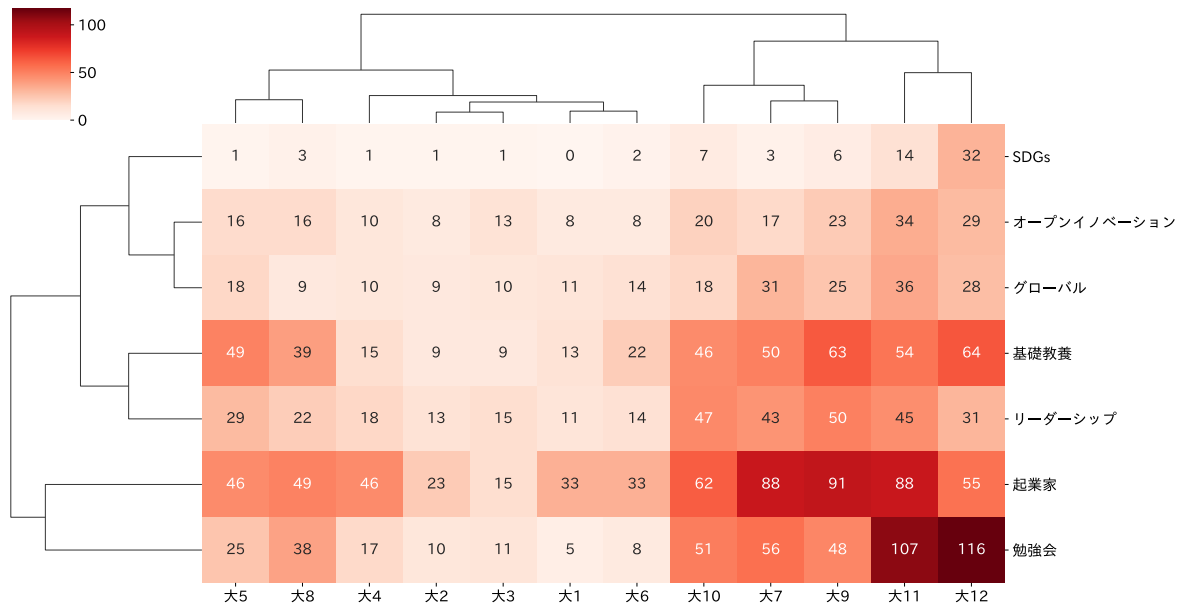


図 8: 派遣期毎のイベント参加傾向の可視化. 派遣期毎に参加回数を成分に持つベクトル変量を用いて, イベントのジャンル間及び派遣期間のコサイン類似度を計算し, イベントや各派遣期に属する学生間の類似性に基づく階層的なクラスタリングの結果を樹形図を用いて可視化している.

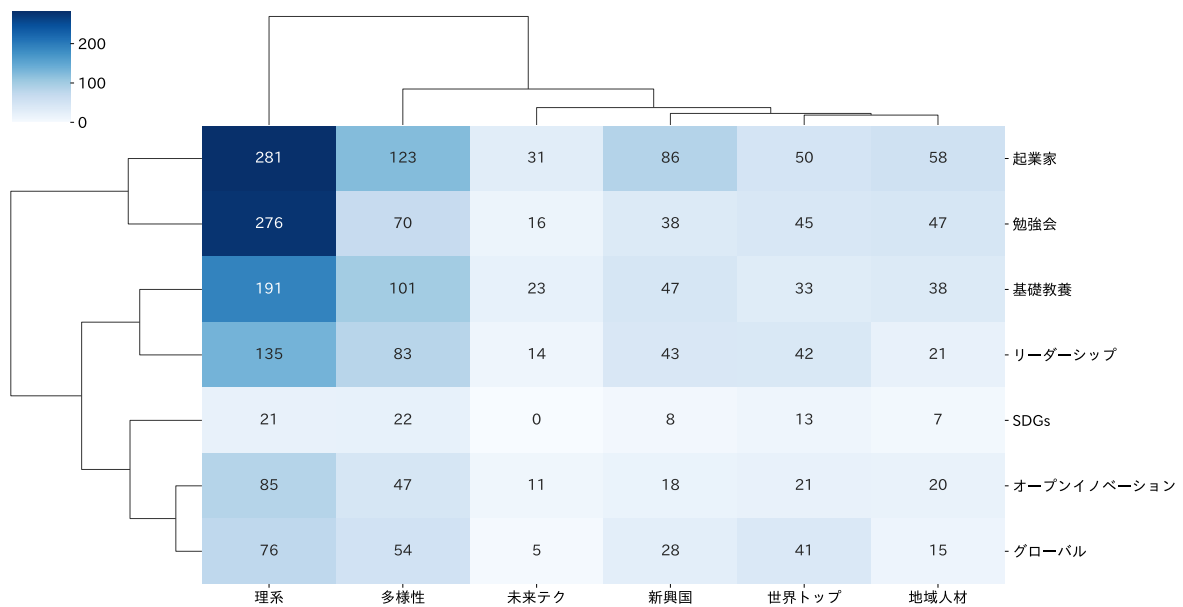


図 9: 申請コース毎のイベント参加傾向の可視化. 申請コース毎に参加回数を成分に持つベクトル変量を用いて, イベントのジャンル間及び申請コース間のコサイン類似度を計算し, イベントや各申請コースの学生間の類似性に基づく階層的なクラスタリングの結果を樹形図を用いて可視化している.

4 留学計画から探る潜在的な学生間のつながり可視化

前節ではオンライン上のデータを活用することで可視化される学生及びイベント間のつながりを分析した結果を報告した。オンラインコミュニティ内の友人関係情報に基づく学生と学生のつながりを分析することで、トビタテの主な学生支援の一つである「留学事前・事後に行う研修の提供」によって、地理的近接性に依存しない多様な人脈構造が形成されている一方、派遣期を跨ぐ専門性や興味関心に基づくつながりの乏しさが明らかになった。またイベントの参加履歴を分析することで、学生間の興味関心の類似性、あるイベントに参加している学生はある特定のイベントにも参加しているという共起性を可視化した。これらは学生の興味関心をリアルタイムに把握、学生に対するイベントのレコメンドなど、データ駆動型の「継続的な学習や交流の場としての留学生ネットワークの提供」の新たな可能性である。本節では専門性や興味関心に基づく潜在的な学生間のつながりを、学生が作成した留学計画から可視化する。具体的には、留学計画のテキストデータに対して機械学習を用いた自然言語処理を行うことで、留学計画の類似度の高い学生間を抽出する手法を紹介する。

留学計画から見るトビタテ参加学生の特性

学生が作成する留学計画は、学生の興味関心、専門性の豊富な情報を含む自然言語によるテキストデータである。トビタテのホームページ [1] には「留学計画書の作り方」として、留学計画作成時におけるポイント等が掲載されており、自身の留学計画に沿った応募コースを選択し申請を行う。専門性や興味関心、留学における実践活動が類似している学生間の潜在的なつながりは、留学計画の類似性として可視化されると期待できる。ここでは留学計画における「タイトル」と「目的と概要（留学計画を説明する主要部分）」のテキストデータに対して自然言語処理を行い、申請コース毎で使用されている単語の出現頻度から学生の特性を概観する。本調査では、自然言語処理が可能な形にデータ化された大学2期から12期に採用された学生3,879人の留学計画を分析する。図10に留学計画におけるタイトルと目的と概要に含まれる単語数の分布を示す。留学計画のタイトルにおける平均の単語数は19であり、目的と概要における平均の単語数は438であった。

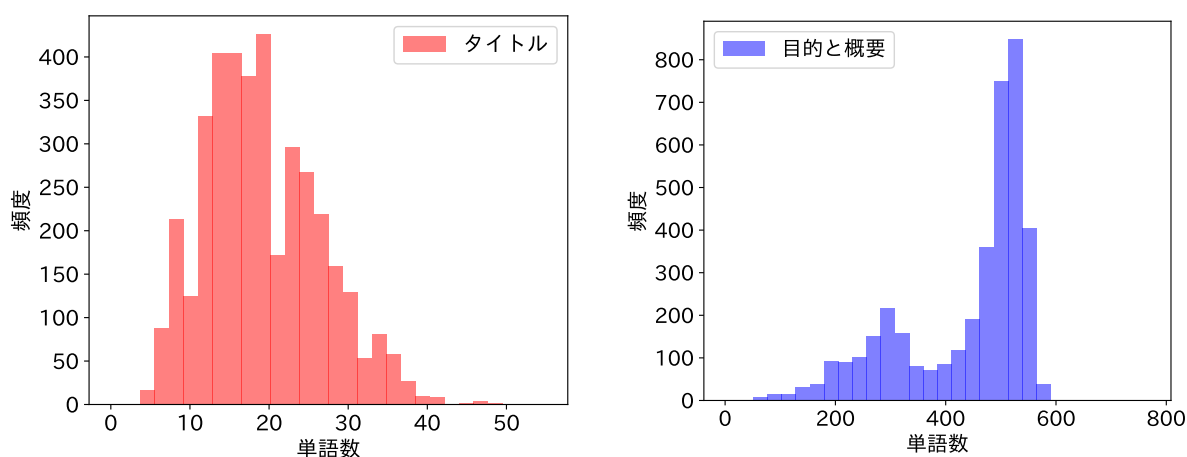


図 10: 留学計画におけるタイトルと目的と概要に含まれる単語数の分布。

留学計画のタイトルは少ない単語数であるが、留学計画の内容が端的に表現されていると考えられる。また留学計画の目的と概要には、学生の専門性や興味関心、留学における実践活動などが具体的に記載されていると期待できる。図11と図12に申請コース毎で留学計画のタイトル及び目的と概要における文章中で使用さ

れている単語の違いを、出現頻度が高い単語をその頻度に応じた大きさで図示する手法であるワードクラウドを用いて可視化した。ただし、単語は名詞のみ可視化しており、文字の色には意味はない。留学計画のタイトルにおける単語の出現頻度に着目すると、全ての申請コースにおいて「日本」や「世界」という単語が多く利用されていることがわかる。また、理系、複合・融合系人材コースや未来テクノロジー人材枠のタイトルには「研究」や「開発」や「技術」、その他の3つのコースのタイトルには「社会」や「教育」など、申請コースの特色が留学計画のタイトルにも反映していることがわかる。また、留学計画の目的と概要にはタイトルと異なり、「私」などの文章特有の名詞や、「活動」や「経験」といった具体的な留学内容を記載するための名詞が可視化されていることがわかる。

機械学習モデルによる類似した留学計画の見える化

個々の留学計画における各単語の出現頻度は、前節のイベント参加履歴の分析におけるジャンル別のイベント参加数と同様に、各学生のベクトル変数としての特徴量として定義することができる。このように文章中の単語の出現頻度をベクトル化する手法として、Bag-of-Words (BOW) が知られている。一方、BOW は単語の語順に関する情報が失われることや、単語の意味を表現することが困難であるという欠点を持つ。このような自然言語処理の課題を克服する機械学習モデルとして、文章データにおける単語の分散表現を学習し、単語や文章をベクトル化する Word2Vec (w2v) [10] 及び Doc2Vec (d2v) [11] がある。

本調査研究では大学2期から12期に採用された学生3,879人の留学計画の文章データから単語の分散表現を学習した2つの機械学習モデル(w2v, d2v)を用いて、各学生の留学計画の「タイトル」と「目的と概要」に対応するベクトル変数を求め、全ての学生間のコサイン類似度を計算することで留学計画の類似性を評価した。これら2つのモデルを用いて計算された学生の留学計画間のコサイン類似度の分布を図13に示す。これら学生の留学計画間のコサイン類似度によって定義された学生間の類似性が、潜在的な学生の専門性や興味関心に基づく類似性であることが期待される。

実際にどのような留学計画間のコサイン類似度が高い傾向があるかを評価するため、各学生の留学計画に対して、コサイン類似度が高い上位5人をリンクで結合することでネットワークを定義する。この定義に基づいて大学2期から12期に採用された学生3,879人の留学計画の類似性に基づくネットワークを可視化した結果を図14(モデル:w2v)と図15(モデル:d2v)に示す。ただし、各ノードは学生の申請コースに基づいて色付けし、その位置座標はばね・電気モデル[9]を用いて決定した。留学計画のタイトルの類似性に基づくネットワークには目立った群れ構造が確認できない一方、留学計画の目的と概要の類似性に基づくネットワークには申請コースに関する群れ構造が確認できる。ワードクラウドによって可視化された留学計画のタイトル及び目的と概要における文章中で使用されている単語の違い、研究などを目的とした留学計画が多い理系とその他のコースの頻出単語の違いが、留学計画の目的と概要の類似性に基づくネットワークの群れ構造として現れていることが見て取れる。これらの結果は、留学計画の文章データから単語の分散表現を学習した機械学習モデルによって計算したコサイン類似度を用いて、申請コース、学生の専門性や興味関心に即した学生間のマッチングが可能であると期待できる。

留学計画の類似性に基づくマッチングによるコミュニティ形成の可能性

本節ではこれまでに、留学計画のテキストデータに対して機械学習を用いた自然言語処理を行い、留学計画の類似度の高い学生間を抽出することで、専門性や興味関心に基づく潜在的な学生間のつながりを可視化できる可能性を示した。ここでは実際にこの手法を用いて留学計画の類似性に基づくマッチングができた場合に、どのような同類性やコミュニティ構造を持つ学生間のネットワークが形成されるかを評価する。学生の専門性をより詳細に特徴付けるため、申請コースに加え留学計画の分野を用いて留学計画の類似性に基づくネット

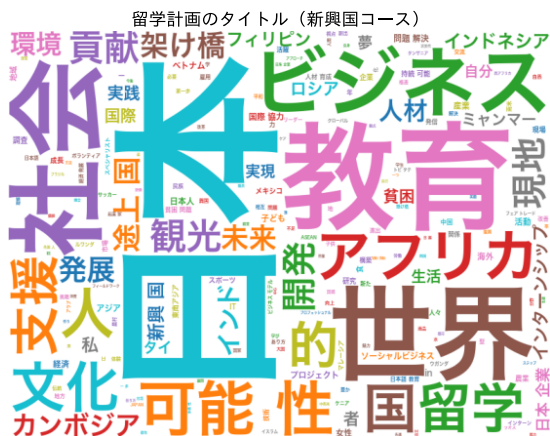
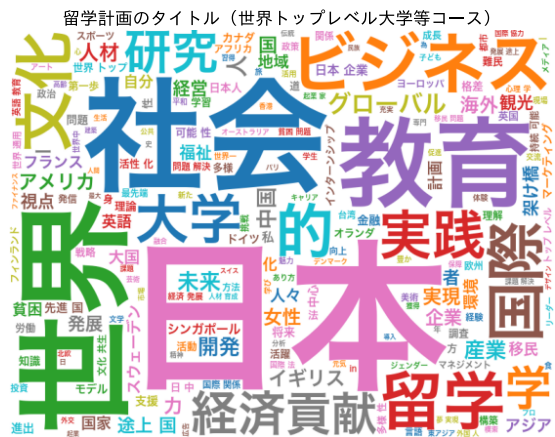


図 11: 申請コース毎留学計画のタイトルに対するワードクラウド。出現頻度が高い単語をその頻度に応じて大きく可視化している。ただし、単語は名詞のみ可視化しており、文字の色には意味はない。

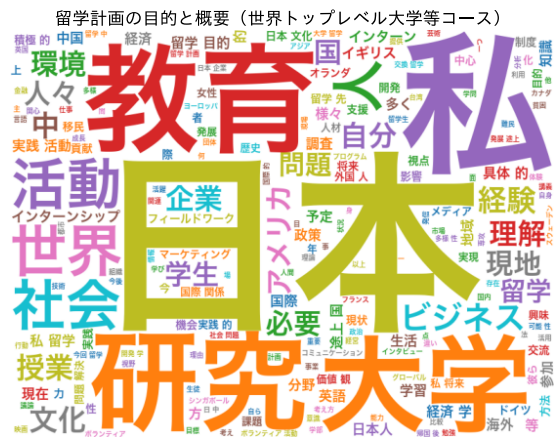
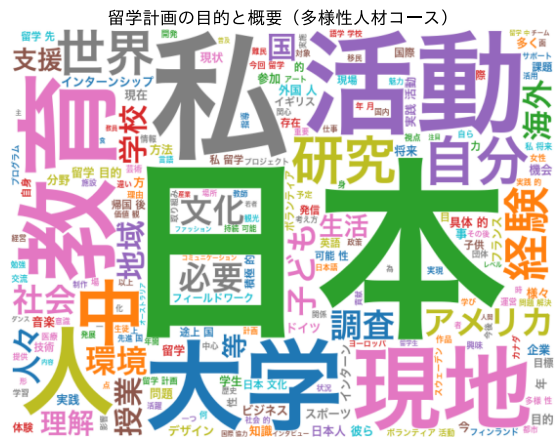
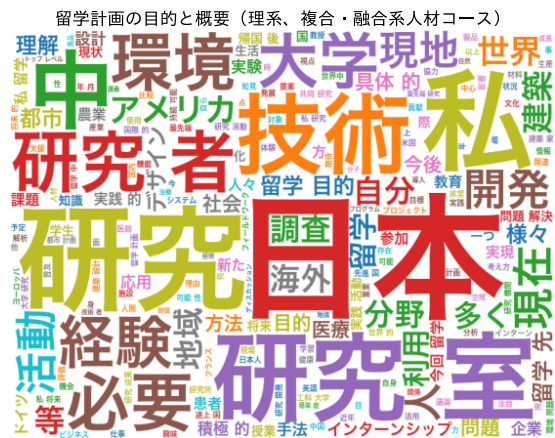


図 12: 申請コース毎留学計画の留学計画の目的と概要に対するワードクラウド。出現頻度が高い単語をその頻度に応じて大きく可視化している。ただし、単語は名詞のみ可視化しており、文字の色には意味はない。

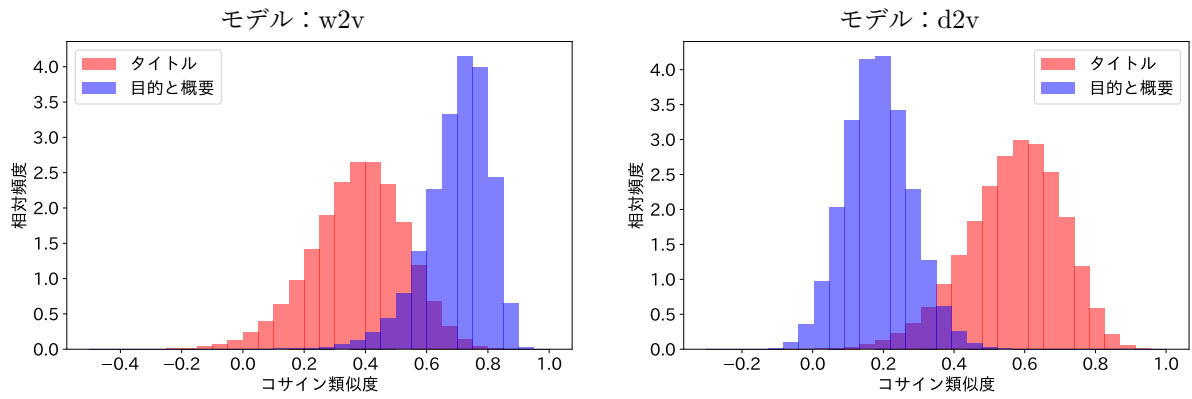


図 13: 大学 2 期から 12 期に採用された学生 3,879 人の留学計画における「タイトル (赤色)」と「目的と概要 (青色)」のコサイン類似度の分布. (左) モデル：w2v, (右) モデル：d2v.

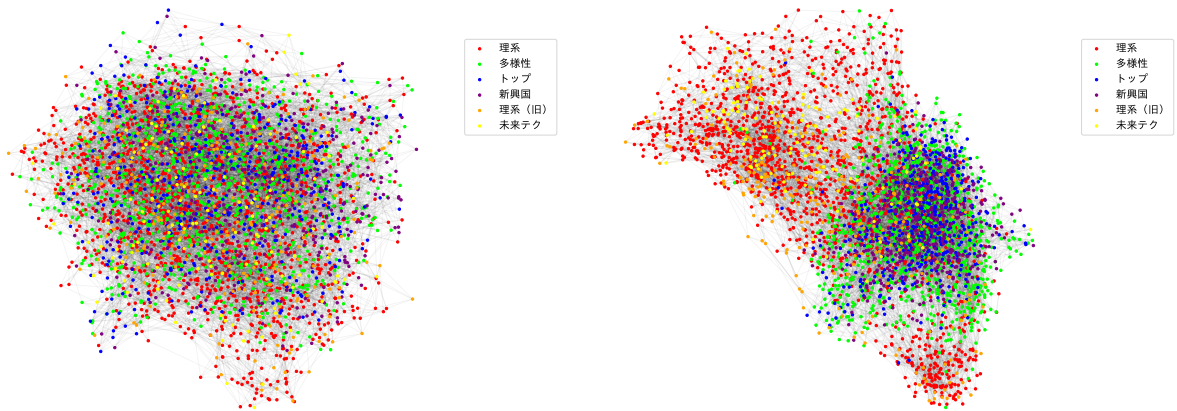


図 14: 大学 2 期から 12 期に採用された学生 3,879 人の留学計画の類似性 (モデル：w2v) に基づくネットワークの可視化. (左) タイトル, (右) 目的と概要. ただし, 学生の申請コースに基づいて色付けしている.

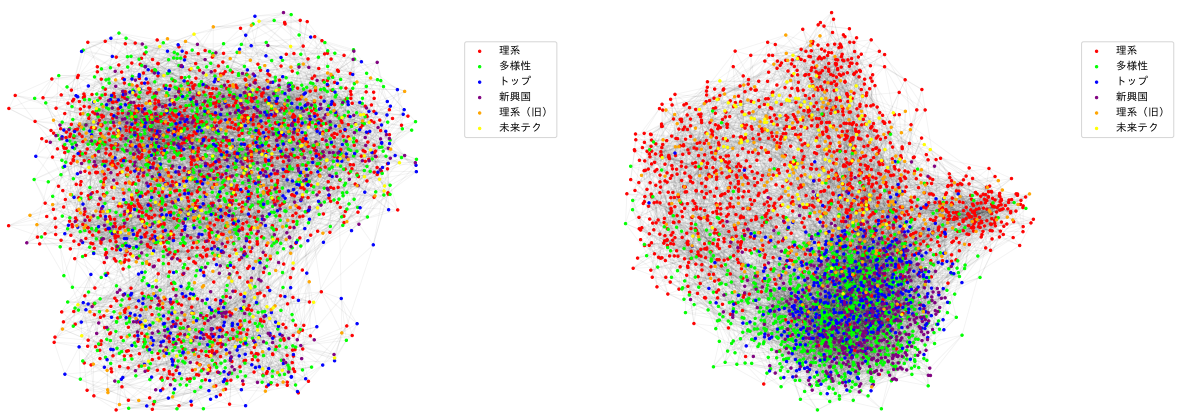


図 15: 大学 2 期から 12 期に採用された学生 3,879 人の留学計画の類似性 (モデル：d2v) に基づくネットワークの可視化. (左) タイトル, (右) 目的と概要. ただし, 学生の申請コースに基づいて色付けしている.

ワークの同類性等を評価する。ただし留学計画の分野の項目は、大学5期の募集から導入された留学計画情報の項目であり、大学12期から分類項目が変更・細分化されている。そのためここでは、留学計画の分野（18項目）が特定できる大学5期から大学11期の学生2,509人*2の留学計画の類似性に基づくネットワークの同類性及びコミュニティ構造を評価した結果を報告する。

図14と図15で可視化した留学計画の類似性に基づくネットワーク同様、留学計画の分野（18項目）が特定できる学生2,509人の留学計画に対して、コサイン類似度が高い上位5人をリンクで結合することでネットワークを定義する。留学計画の類似性に基づくネットワークにおけるリンクレベルの属性間の結合傾向を可視化するため、申請コースと留学計画の分野に対する結合確率 $e_{\alpha\beta}$ を図16と図17に示した。ただし、図16は留学計画のタイトル、図17は留学計画の目的と概要の類似性に基づくネットワークの結合確率を示す。また、これら申請コースと留学計画の分野の結合確率を用いて同類性を計算した結果が表4にまとめた。2つのモデルによって計算された留学計画の類似性に基づくネットワーク同類性を比較すると、留学計画のタイトルに比べ、留学計画の目的と概要を用いた方がより同じ申請コースや留学計画の分野同士の学生をマッチングできていることがわかる。また、d2vを用いて計算された留学計画のタイトルの類似性に基づくネットワークの同類性は、w2vのものに比べて著しく低いが、これはd2vが文章全体から単語の分散表現を学習し、文章をベクトル化するモデルであるためであると考えられる。

同類性 r_a	w2v	d2v	同類性 r_a	w2v	d2v
申請コース	0.250	0.017	申請コース	0.446	0.456
留学計画の分野	0.188	0.024	留学計画の分野	0.359	0.354
利用データ	タイトル		利用データ	目的と概要	

表4: 留学計画のタイトル（左）及び目的と概要（右）の類似性に基づくネットワークにおける隣接する2つのノードの属性の同類性。

前節及び令和元年度の調査研究では、オンラインコミュニティ内の友人関係情報に基づく学生と学生のつながりを分析することで、トビタテの主な学生支援の一つである「留学事前・事後に行う研修の提供」によって、派遣期毎の結合確率が高く申請コース毎の結合確率は低い友人関係が形成されていることが明らかになった。これら式(5)の同類性とマッチングによって形成され得る学生間ネットワークの同類性(表4)を比較すると、これまでに形成されたオンライン上での友人関係ネットワークにおける派遣期毎の同類性と同程度の専門性に関する同類性を持つネットワークを形成することができると期待できる。

最後に留学計画の類似性に基づくネットワークの階層的クラスタリング、コミュニティ抽出の手法適用することで、留学計画情報を活用したマッチングによって形成され得るコミュニティ構造を定量的に評価する。ネットワーク上のランダムウォークのエンコードを最短化することでコミュニティ分割を求める Infomap 法 [12, 13] は、同種のコミュニティ抽出手法と比較して高精度にコミュニティ分割ができることが示されている [14]。本調査研究ではこの手法を用いて、特に同類性が高い結果が得られた留学計画の目的と概要の類似性に基づくネットワークにおけるコミュニティを抽出し、各コミュニティの特性をノードの属性を用いて評価する*3。留学計画の目的と概要の類似性に基づくネットワークにおいて抽出されたコミュニティの集計結果を表5に示す。ここで階層レベルは、階層クラスタリングにおける階層度であり、階層レベル2のコミュニティ数は、階層レベル1のコミュニティに内包されるサブコミュニティ数を表す。抽出されたコミュニティの内、サ

*2 ただし、留学計画変更時に留学計画の分野（18項目）を選択した大学5期以前の学生も数名含まれる。

*3 Infomap 法 [12, 13] のコードは、MapEquation ホームページ (<https://www.mapequation.org>) にて入手可能である。

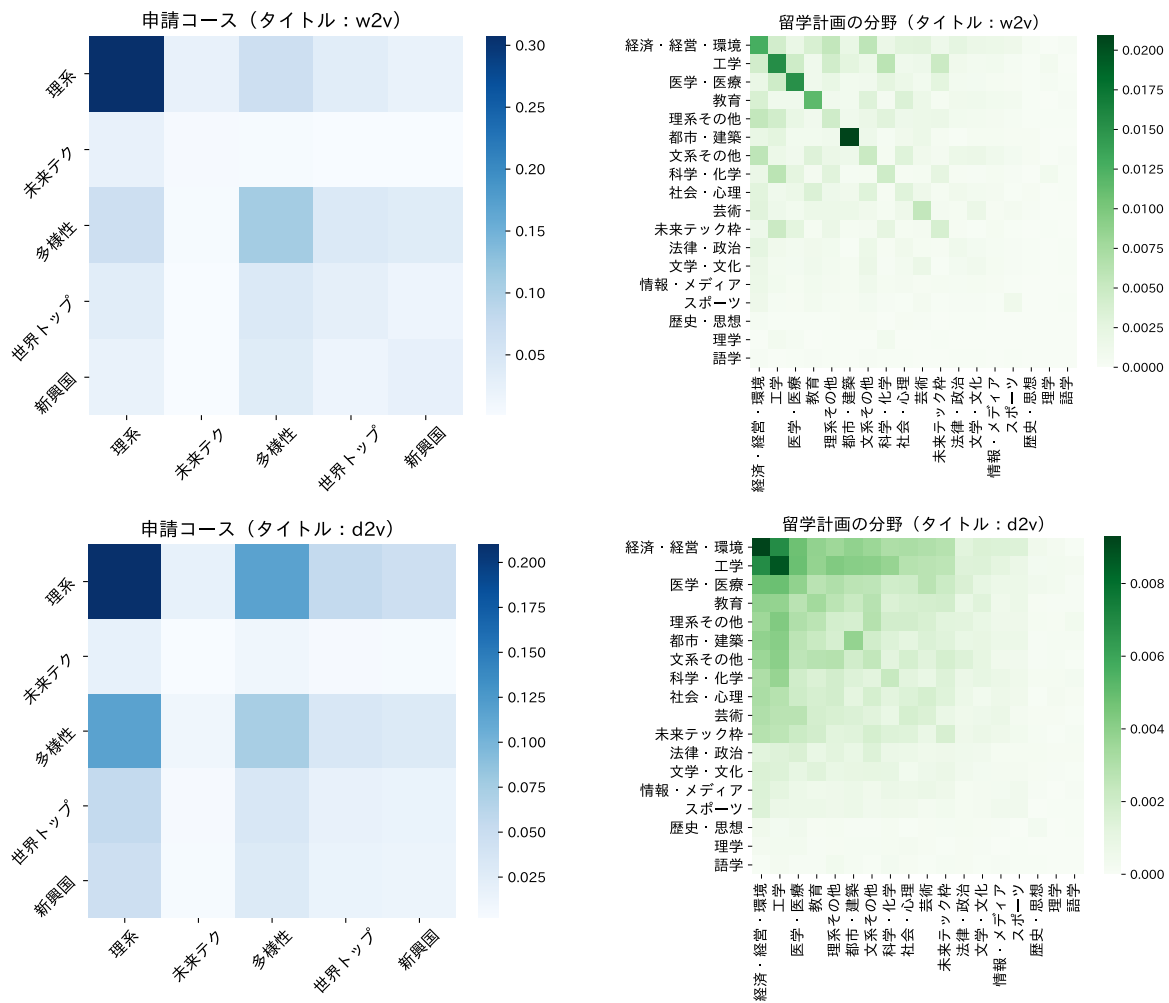


図 16: 留学計画のタイトルの類似性に基づくネットワークにおける属性 (左) 申請コース及び (右) 留学計画の分野に対する要素間の結合確率. (上) モデル: w2v, (下) モデル: d2v.

ブコミュニティを含まないコミュニティを既約と表現し、各階層レベルにおける既約なコミュニティに属する人数を記載している。階層レベル 1 において 50 人以上が含まれるコミュニティに着目し、そのコミュニティ特性をコミュニティに属する学生の属性 (申請コースと留学計画の分野) の割合を用いて評価する。コミュニティ抽出を行った留学計画の分野が特定されている学生 2,509 人の留学計画の目的と概要の類似性に基づくネットワークの可視化を可視化した結果を図 18 (モデル: w2v) と図 19 (モデル: d2v) に示す。ただし、各ノードは学生の申請コース及び抽出されたコミュニティに基づいて色付けし、その位置座標はばね・電気モデル [9] を用いて決定した。またコミュニティ特性を評価した結果を、表 6 (モデル: w2v) と表 7 (モデル: d2v) にそれぞれ示した。ここでコミュニティの番号は、内包するサブコミュニティの数が多い順に番号を付け、コミュニティに占める各属性の割合は 10% 以上であるもののみ記載している。2つのモデルによる留学計画の目的と概要の類似性に基づくネットワークは、階層的なコミュニティ構造の違いはあるものの、留学計画の分野が「都市・建築」や「医学・医療」の理系コミュニティだけでなく、「教育」や「芸術」といった留学計画の分野が申請コースを跨ぐような場合においても専門分野に基づくコミュニティを内包することがわかる。

本調査研究では、代表的な 2つの機械学習モデルを用いて留学計画におけるテキストデータの類似性を計算することで、同じ留学計画の分野、専門性を持つ学生間のマッチングできる可能性を示した。イベントの参加

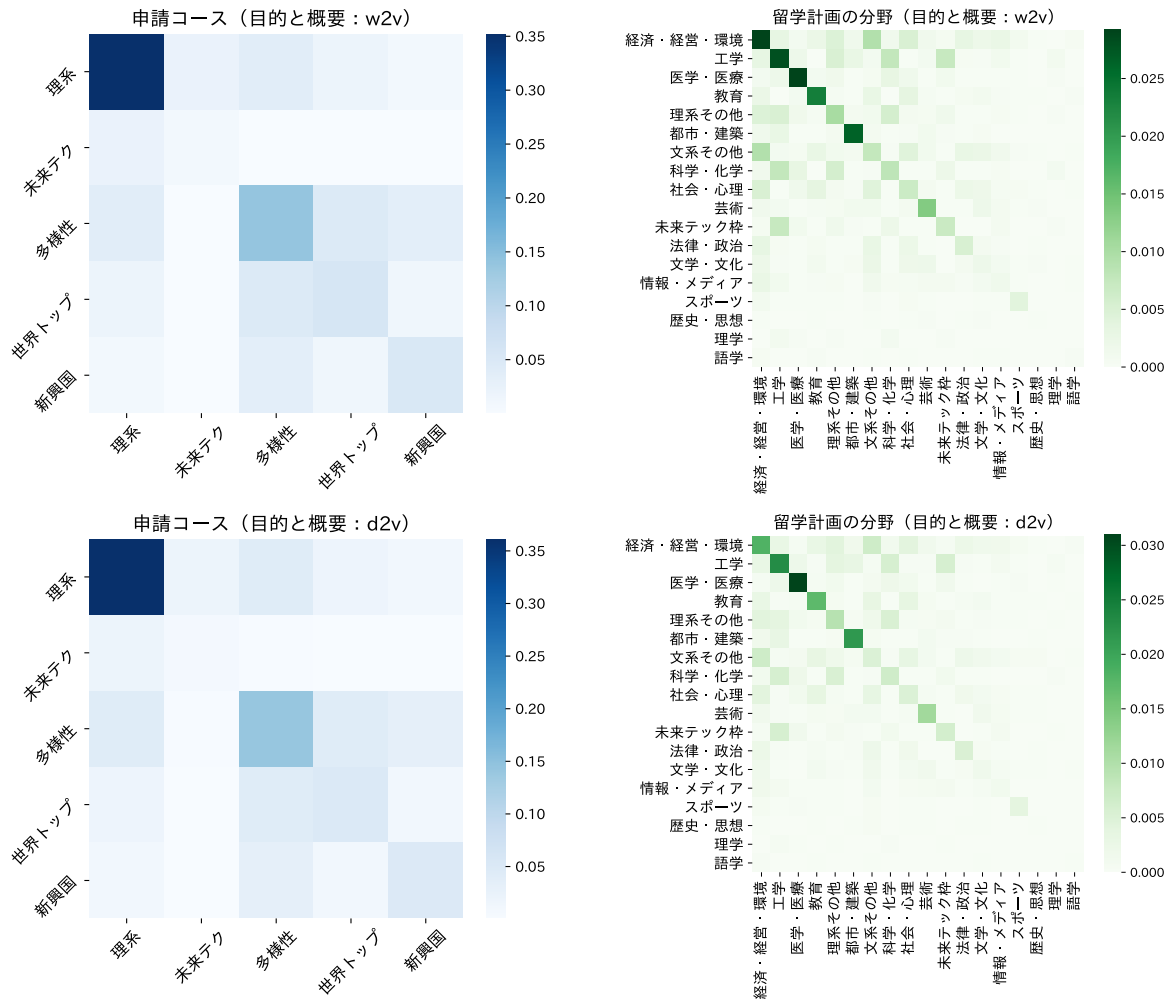


図 17: 留学計画の目的と概要の類似性に基づくネットワークにおける属性（左）申請コース及び（右）留学計画の分野に対する要素間の結合確率。（上）モデル：w2v,（下）モデル：d2v.

履歴や留学の成果報告書など、様々なデータと統計手法を組み合わせることで、より学生のニーズにあったマッチング、データ駆動型の学生支援の実現が期待できる。

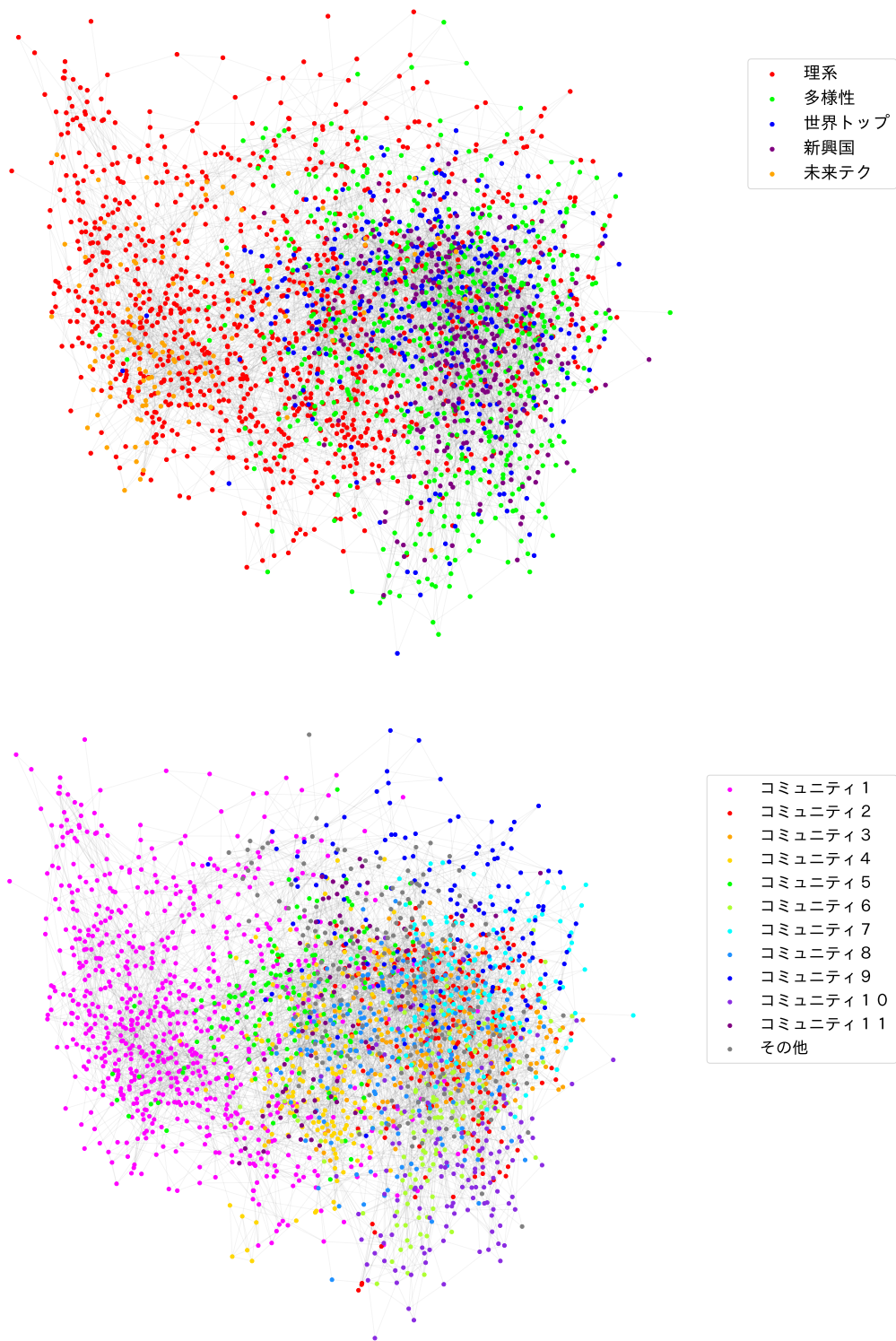


図 18: 留学計画の分野が特定されている学生 2,509 人の留学計画の目的と概要の類似性 (モデル: w2v) に基づくネットワークの可視化。ただし、各ノードは学生の申請コース及び抽出されたコミュニティに基づいて色付けしている。抽出されたコミュニティ特性の詳細は表 6 に記載。

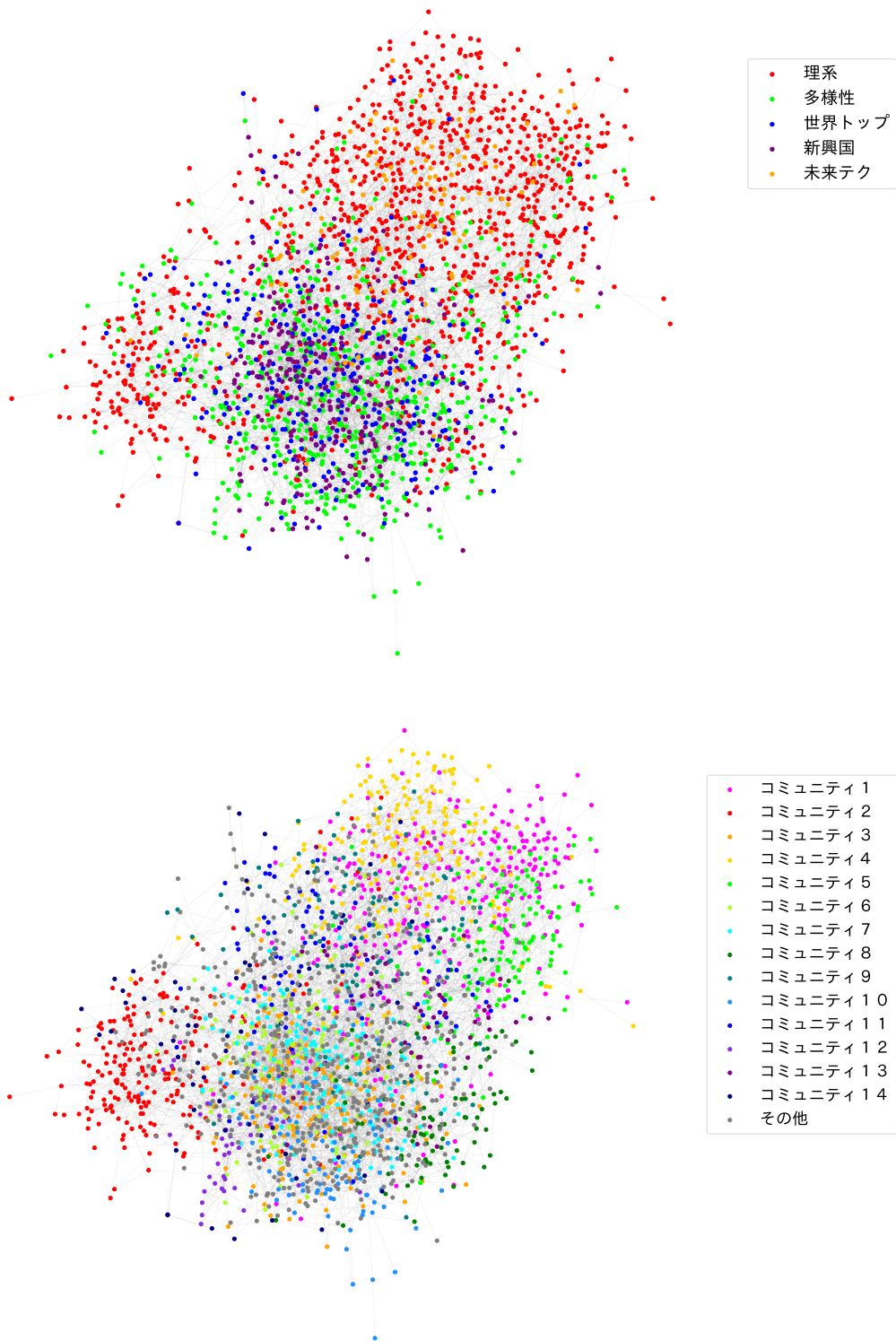


図 19: 留学計画の分野が特定されている学生 2,509 人の留学計画の目的と概要の類似性 (モデル: d2v) に基づくネットワークの可視化。ただし、各ノードは学生の申請コース及び抽出されたコミュニティに基づいて色付けしている。抽出されたコミュニティ特性の詳細は表 7 に記載。

モデル：w2v

階層レベル	コミュニティ数	(うち既約)	人数(既約)
1	21	(4)	19
2	303	(264)	1,519
3	212	(198)	876
4	30	(28)	86
5	4	(4)	9

モデル：d2v

階層レベル	コミュニティ数	(うち既約)	人数(既約)
1	50	(14)	24
2	528	(468)	2,081
3	154	(153)	398
4	2	(2)	6

表 5: 留学計画の目的と概要の類似性に基づくネットワークにおいて Infomap 法 [12] によって抽出されたコミュニティの集計結果。(上) モデル：w2v, (下) モデル：d2v. ここで階層レベルは、階層クラスタリングにおける階層度であり、階層レベル $n+1$ のコミュニティ数は、階層レベル n のコミュニティに内包されるサブコミュニティ数を表す。サブコミュニティを含まないコミュニティを既約と表現し、各階層レベルにおける既約なコミュニティに属する人数を記載した。

5 おわりに

日本初の官民協働のグローバル人材育成事業であるトビタテは、留学に対する金銭的な支援に加え、「留学事前・事後研修」や「継続的な学習や交流の場」を設計することで、所属や留学国、地域や専門性の垣根を越えた交流の場、多様な人脈形成の機会を提供している。令和元年度調査研究では、Facebook の個人ページにおいて公開されている友人情報を活用して、オンラインコミュニティに属する数千規模の人脈構造、トビタテ友人関係ネットワークを可視化し、ネットワーク科学の観点からコミュニティの特性、それらの成長過程を可視化することで、学習支援との関連性を明らかにした。トビタテの学習支援の特徴でもある留学事前・事後研修は、トビタテに採用された留学派遣期の近い学生を全国から一同に会する機会を提供し、グループワーク形式の研修によって、参加学生間の相互理解の場を提供する。これらオフラインでの交流機会によって結ばれたと考えられる友人関係は、オンラインの友人関係として顕著に現れていた。一方、2020年1月以降、新型コロナウイルス感染症のパンデミックが世界中に広がり、派遣留学生の一時帰国や留学の中止など大きな影響を及ぼしている。2020年2月末には大学12期の選考結果が発表されるも対面での壮行会の中止が発表され、3月上旬に予定されていた大学12期に対する事前研修も任意のオンライン開催となった。オンライン上でのコミュニケーションツールの活用により、実際の留学だけではない、「継続的な学習や交流の場としての留学生ネットワークの提供」として、学生の学びに対する支援が今もなお継続されている。

「継続的な学習や交流の場としての留学生ネットワークの提供」の新たなカタチとして、2020年1月にはトビタテ事務局の主導で、留学後の継続的な成長促進と支援企業等との接点、コラボレーションの創出を狙いとしたり「トビタテ学習プラットフォーム」が発足した。トビタテ学習プラットフォームによって、有識者やスポ

	人数	申請コース	留学計画の分野
1	759	理系 (83.1) 未来テク (13.2)	工学 (33.5) 科学・化学 (17.1) 未来テック枠 (13.20) 医学・医療 (12.5) 理系その他 (12.3)
2	221	多様性 (41.2) 新興国 (32.1) 世界トップ (15.8)	経済・経営・環境 (19.1) 教育 (15.5) 社会・心理 (13.6) 文系その他 (10.9)
3	186	理系 (34.9) 新興国 (24.2) 多様性 (22.6) 世界トップ (18.3)	経済・経営・環境 (40.3) 理系その他 (18.8) 文系その他 (11.3)
4	183	理系 (88.0)	都市・建築 (75.4)
5	159	世界トップ (29.6) 多様性 (27.7) 理系 (27.0) 新興国 (10.7)	経済・経営・環境 (32.1) 工学 (11.3) 文系その他 (11.3) 芸術 (10.1)
6	146	理系 (39.7) 多様性 (32.9) 新興国 (13.7) 世界トップ (13.0)	経済・経営・環境 (24.7) 文系その他 (21.9) 理系その他 (19.2)
7	154	多様性 (37.0) 世界トップ (31.8) 新興国 (25.3)	社会・心理 (20.1) 経済・経営・環境 (19.5) 文系その他 (18.8) 法律・政治 (14.3) 教育 (12.3)
8	169	多様性 (47.9) 世界トップ (27.8) 新興国 (20.1)	社会・心理 (18.9) 文系その他 (16.0) 経済・経営・環境 (15.4) 法律・政治 (13.6) 教育 (10.1)
9	134	理系 (74.6) 多様性 (20.1)	医学・医療 (83.6)
10	142	多様性 (57.0) 世界トップ (23.9) 新興国 (15.5)	教育 (70.4)
11	73	多様性 (78.1) 世界トップ (19.2)	芸術 (71.2)

表 6: 留学計画の目的と概要の類似性 (モデル: w2v) に基づくネットワークにおけるコミュニティ特性. ただしコミュニティ抽出には, Infomap 法 [12] を用いた. それぞれのコミュニティにおけるノード数及び, 属性ごとの割合. ただし, 50 人以上のコミュニティを内包するサブコミュニティの数が多い順に列挙している. 括弧内にはコミュニティに占める各属性の割合を示し, 割合は 10% 以上であるもののみ記載した.

ンサー企業, トピタテに参加した学生が登壇するセミナー等のイベントを定期的で開催されている. これらコミュニティにおける力学を調査・分析し, 学生支援との相関を明らかにすることは, 新型コロナ禍における困難を克服し, データ駆動型支援を推進する上で非常に重要である.

本調査研究では, オンラインコミュニティ内の友人関係情報, イベントの参加履歴を分析することで, 新型コロナ禍におけるオンライン上での継続的なコミュニティ形成の実態を可視化し, オンラインイベントの参加履歴から学生間の興味関心の類似性, あるイベントに参加している学生はある特定のイベントにも参加しているという共起性を可視化した. これらは学生の興味関心をリアルタイムに把握, 学生のニーズに沿ったイベントのレコメンドなど, データ駆動型の「継続的な学習や交流の場としての留学生ネットワークの提供」の新たな可能性である. さらに本調査研究では, 機械学習を用いて留学計画が類似する学生同士を抽出することで, 同じ留学計画の分野, 専門性を持つ学生間のマッチングできる可能性を示した. これらは従来の「留学事前・事後に行う研修の提供」のみでは形成されていなかった, 派遣期を跨ぐ専門性や興味関心に基づくつながり創出の糸口となるだろう. イベントの参加履歴や留学の成果報告書など, 様々なデータと統計手法を組み合わせ

	人数	申請コース	留学計画の分野
1	311	理系 (72.0) 未来テク (22.5)	医学・医療 (27.0) 工学 (24.4) 未来テック枠 (22.5) 科学・化学 (10.9)
2	211	理系 (79.6)	都市・建築 (67.3)
3	186	新興国 (43.0) 多様性 (31.2) 世界トップ (18.8)	経済・経営・環境 (24.3) 文系その他 (16.8) 教育 (15.1)
4	231	理系 (85.7)	工学 (41.6) 科学・化学 (25.1)
5	149	理系 (67.8) 多様性 (24.2)	医学・医療 (79.9)
6	139	新興国 (53.2) 多様性 (28.1) 世界トップ (11.5)	経済・経営・環境 (30.9) 教育 (21.6) 文系その他 (20.1) 社会・心理 (10.8)
7	156	多様性 (50.0) 世界トップ (37.8)	社会・心理 (23.1) 法律・政治 (23.1) 文系その他 (16.7) 経済・経営・環境 (12.8)
8	84	理系 (39.3) 多様性 (31.0) 世界トップ (16.7) 新興国 (13.1)	理系その他 (29.8) 経済・経営・環境 (21.4) 文系その他 (13.1)
9	117	理系 (81.2)	理系その他 (43.6) 経済・経営・環境 (20.5) 科学・化学 (14.5)
10	109	多様性 (56.0) 世界トップ (23.9) 新興国 (13.8)	教育 (60.6)
11	90	理系 (62.2) 多様性 (12.2)	工学 (37.8) 経済・経営・環境 (21.1)
12	50	多様性 (70.0) 理系 (14.0) 世界トップ (12.0)	芸術 (56.0)
13	64	理系 (75.0) 多様性 (15.6)	理系その他 (43.8) 経済・経営・環境 (17.2) 科学・化学 (17.2)
14	62	多様性 (41.7) 世界トップ (20.0) 新興国 (20.0) 理系 (16.7)	芸術 (33.9) 工学 (12.9) 文学・文化 (11.3)

表 7: 留学計画の目的と概要の類似性 (モデル: d2v) に基づくネットワークにおけるコミュニティ特性. ただしコミュニティ抽出には, Infomap 法 [12] を用いた. それぞれのコミュニティにおけるノード数及び, 属性ごとの割合. ただし, 50 人以上のコミュニティを内包するサブコミュニティの数が多い順に列挙している. 括弧内にはコミュニティに占める各属性の割合を示し, 割合は 10% 以上であるもののみ記載した.

ることで, より学生のニーズにあったマッチングが可能になるかもしれない. データ駆動型の学生支援の継続によって, これらが実現されることを期待したい.

参考文献

- [1] トビタテ！留学japan ホームページ (<https://tobitate.mext.go.jp/>) .
- [2] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, Vol. 393, No. 6684, p. 440, 1998.
- [3] Amanda L Traud, Peter J Mucha, and Mason A Porter. Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications*, Vol. 391, No. 16, pp. 4165–4180, 2012.
- [4] Leto Peel, Jean-Charles Delvenne, and Renaud Lambiotte. Multiscale mixing patterns in networks. *Proceedings of the National Academy of Sciences*, Vol. 115, No. 16, pp. 4057–4062, 2018.
- [5] Neil Rubens, Martha Russell, Rafael Perez, Jukka Huhtamäki, Kaisa Still, Dain Kaplan, and Toshio Okamoto. Alumni network analysis. In *2011 IEEE Global Engineering Education Conference (EDUCON)*, pp. 606–611. IEEE, 2011.
- [6] 令和元年度日本学生支援機構学生支援の推進に資する調査研究事業「官民協働留学創出プロジェクト（トビタテ）における友人関係ネットワークの成長とコミュニティ構造の可視化研究」(<https://www.jasso.go.jp/about/statistics/jasso-research/2019.html>) .
- [7] Mark EJ Newman. Mixing patterns in networks. *Physical Review E*, Vol. 67, No. 2, p. 026126, 2003.
- [8] 同窓会組織とまりぎのホームページ (<http://tobitate-net.com/>) .
- [9] Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, Vol. 21, No. 11, pp. 1129–1164, 1991.
- [10] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.
- [11] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents, 2014.
- [12] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, Vol. 105, No. 4, pp. 1118–1123, 2008.
- [13] Martin Rosvall and Carl T Bergstrom. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PloS one*, Vol. 6, No. 4, p. e18209, 2011.
- [14] Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. *Physical review E*, Vol. 80, No. 5, p. 056117, 2009.