

GRIPS Discussion Paper 24-7

**アカデミアにおける将来のスター人材の早期検知システムの
試行的開発**
**Trial development of a system for early detection of future stars
in academia**

By

**池内 健太
佐々木 達郎
隅藏 康一
登坂 万結
野村 ともこ
上島 早織**

August 2024



GRIPS

NATIONAL GRADUATE INSTITUTE
FOR POLICY STUDIES

National Graduate Institute for Policy Studies
7-22-1 Roppongi, Minato-ku,
Tokyo, Japan 106-8677

アカデミアにおける将来のスター人材の早期検知システムの試行的開発

Trial development of a system for early detection of future stars in academia.

池内 健太

独立行政法人経済産業研究所 上席研究員

政策研究大学院大学 科学技術イノベーション政策研究センター プログラムコンサルタント

佐々木 達郎

文部科学省 科学技術・学術政策研究所 第2研究グループ 主任研究官

隅藏 康一

政策研究大学院大学 教授

登坂 万結

三菱電機株式会社 開発戦略部 連携推進グループ 主席研究員

野村 ともこ

三菱電機株式会社 開発戦略部 連携推進グループ 主席研究員

上島 早織

三菱電機株式会社 開発戦略部 連携推進グループ研究員

目次

Abstract	3
エグゼクティブサマリー	4
1 調査研究の背景・目的と調査手法	5
1.1 調査研究の背景・目的	5
1.2 調査研究の方法	6
2 将来のスター研究者の早期検知モデルの分析結果	13
3 まとめ	27
謝辞	28
参考文献	28
付録	29

Abstract

This paper reports on our efforts to develop a system for Dynamic Early-warning by Learning to Predict High Impact-Researchers (DELPHI-R) on a trial basis. We attempted to analyze a model that connects information on KAKENHI (Grants-in-Aid for Scientific Research) projects with patent and scientific article information and predicts future research performance, based on current and past information. For researchers who received the KAKENHI project between 2000 and 2009, we utilize a machine learning method to predict the top 10% of KAKENHI budget from 2010 to 2019 based on the status of research projects up to 2009, the number of collaborative researchers, and patent/paper information. Finally, we estimate a model, and thus, we obtained high prediction accuracy (ROC-AUC of validation data was 93%) to detect the researchers with top 10% budget and the presence or absence of highly cited papers (HCP).

エグゼクティブサマリー

本稿では将来のスター研究者を早期検出するシステム(DELPHI-R)の試行的な開発に関する取り組みについて報告する。科研費の研究プロジェクトの情報に特許情報と論文情報を接続し、現在および過去の情報に基づいて将来の研究パフォーマンスを予測するモデルの分析を試みた。2000～2009年度に科研費を獲得した研究者について、機械学習の手法により2009年度までの科研費の獲得状況や共同研究者数、特許・論文情報から2010～2019年度の科研費獲得額トップ10%に入るか否かと高頻度引用論文の有無を予測するモデルを分析したところ、高い予測精度(検証用データのROC-AUCが93%)が得られた。

1 調査研究の背景・目的と調査手法

1.1 調査研究の背景・目的

本調査研究では、先行研究 Weis-Jacobson (2021)で提案された重要論文の早期検出システム (DELPHI: Dynamic Early-warning by Learning to Predict High Impact)を応用して、将来のスター研究者を早期検出するシステム (DELPHI-Researcher)を開発する。

図 1 は本研究で参考とした Weis-Jacobson (2021)の分析モデルの概要を示している。Weis-Jacobson (2021)はある論文が出版された時点あるいは直後のデータから出版 5 年後におけるその論文の影響度(被引用度)を予測するモデルを提案している。具体的には、42 のバイオテクノロジー関連の研究雑誌に 1980 年~2019 年に収録された 780 万の論文を分析対象としている。予測の対象とする論文の影響度の指標としては、論文の引用ネットワークにおける中心性指標のトップ 5 に入る論文をハインパクト論文と定義し、出版直後から 4 年後までの各時点のデータによって出版から 5 年後の時点でハインパクト論文となるかどうかを予測するモデルを推定している。

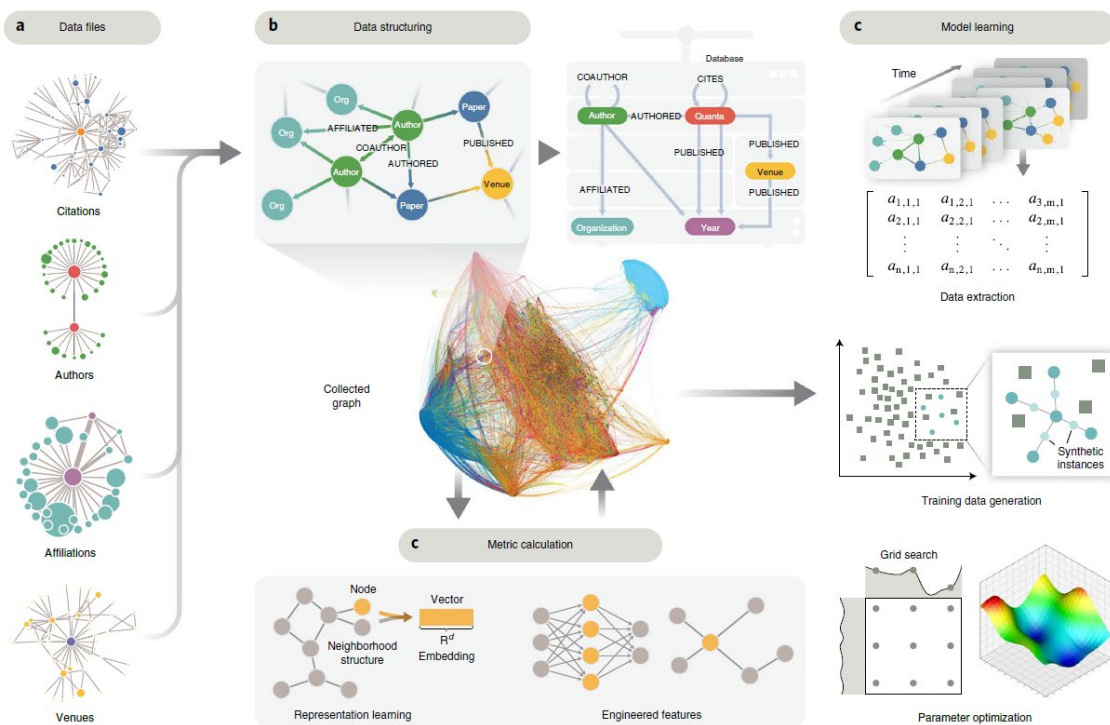


図 1 Weis-Jacobson (2021)の分析モデルの概要

Weis-Jacobson (2021)において予測に用いている説明変数は、当該論文の著者の過去の論文の被引用件数やその増加傾向、著者の過去の論文の量と質のバランスを示す h-index の水準とその変化、当該論文が掲載された雑誌の被引用件数とその増加傾向や h-index といった論文・著者・雑誌の特性であり、また引用ネットワークと共著ネットワークの指標を加えている(表 1)。Weis-Jacobson (2021)では、ネットワーク指標を説明変数として加えることで、予測精度の顕著な向上が見られたと強調している。

表 1 Weis-Jacobson (2021)が予測に用いている変数

Table 1 Features used in machine learning-based early-warning system		
Category	Variable	Description
Paper	Citations per paper	Mean number of citations per paper for papers the author has published
	Δ Citations per paper	Change in the mean number of citations per paper for the author over the preceding 2 years
	Citations per year	Average number of citations per year for papers the author has published
	Maximum citations	Maximum number of citations a paper has received out of all the papers the author has currently published
	Rank citations per year	Rank of the author among all other authors in terms of mean citations per year
	Total citations	Number of citations author has received
	Δ Total citations	Change in the total number of citations for the author over the preceding 2 years
	Total papers	Total number of papers published by the author
	Δ Total papers	Change in the total number of papers over the preceding 2 years
	Citations	Citations collected in the current year
Author	Adopters	Number of unique citing authors in the current year
	Author age	Number of years since the year of publication of the author's first paper
	h-index	Author's h-index
	Δ h-index	Change in the author's h-index over the past 2 years
	Recent coauthors	Number of coauthors the author has had in the current and immediately preceding year
Journal	Δ Mean journal citations per paper	Two-year change in the mean number of citations per paper of the journals the author has published in
	Mean journal citations per paper	Mean number of citations per paper for the journals the author has published in
	Δ Mean journal h-index	Two-year change in the mean h-index for the journals the author has published in
	Mean journal h-index	Mean h-index for the journals the author has published in
	Mean journal maximum citations	Mean of the maximum number of citations any paper published in a journal has received for each journal the author has published in
	Mean journal rank citations per paper	Rank of journal in which the author has published, as determined by the mean number of citations per paper
	Mean Δ journal total papers	Change in the mean of the total number of papers published in journals the author has published in
	Total journals	Total number of journals published in by the author
	Mean journal total papers	Mean of the total number of papers published in journals the author has published in
	Network	Learned network embedding
Time-rescaled node centrality		Time-balanced network centrality calculated using the full citation network
Unweighted PageRank		PageRank score of author, calculated on the unweighted coauthorship network
Weighted PageRank		PageRank score of author, calculated on the weighted coauthorship network

1.2 調査研究の方法

本研究では Weis-Jacobson (2021)が対象としている論文ではなく、将来ハイインパクトな研究成果を生み出す可能性の高い研究者の早期検知を目指したシステムの開発を試行する。ハイインパクトな研究者 (Researcher) の早期検知の意味で「DELPHI-R」と呼ぶ。分析対象は「科学研究費助成事業データベース」(JSPS/NII) から抽出した 2000~2009 年度に研究代表者もしくは研究分担者として科研費プロジェクトに採択された研究者とする。それらの研究者に論文情報及び特許情報を接続して分析に用いる。論文情報は OurResearch 社が無償で提供している「OpenAlex (OA)」¹を用い、論文の著者と研究者を氏名と所属機関を用いて接続する。特許情報は知的財産研究所が無償提供している「IIP パテントデータベース」²と特許庁の「特許庁の一括ダウンロードサービスによる提供データ」³を用い、特許の発明者と研究者を氏名及び研究者の所属機関と特許の出願人の情報をキーに接続する。なお、論文著者と特許発明者の接続は科研費 PJ「大学等の公的研究

¹ OpenAlex についての詳細は以下の URL を参照されたい。 <https://openalex.org/>

² IIP パテントデータベースについての詳細は以下の URL を参照されたい。

<https://www.iip.or.jp/patentdb/>

³ 特許庁の一括ダウンロードサービスについての詳細は以下の URL を参照されたい。

<https://www.jpo.go.jp/system/laws/sesaku/data/download.html>

機関に関する制度改革と経済社会インパクトに関する研究」(23H00843、研究代表者:元橋一之東京大学教授)の研究成果の一部を活用した。

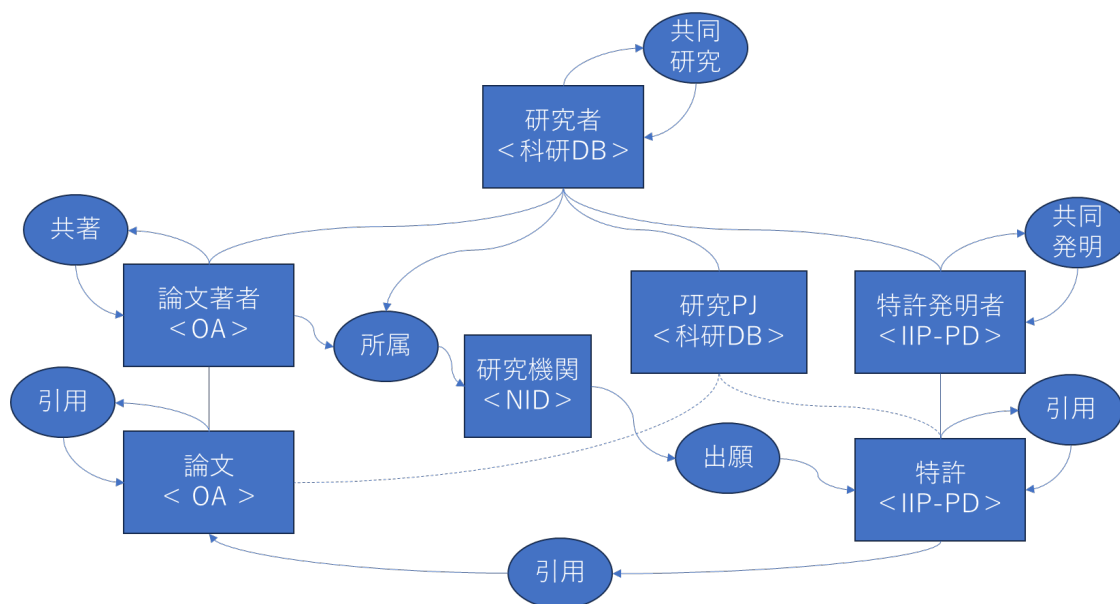


図 2 分析に用いるデータの構造

表 2 は本研究で分析に用いる変数の一覧を示している。まず、予測の対象とする被説明変数は 2 つあり、1 つ目は 2010~2019 年度に研究代表者として参画した科研費 PJ の総配分額トップ 10%に入っているか否かを示すダミー変数(t10p_fund_pi_1019)、2 つ目は 2010~2019 年度のトップ 1%の Highly Cited Paper(HCP)の有無を示すダミー変数(hcr1p_09)である。HCP は Clarivate 社の定義にしたがって、各分野・各年度で被引用数がトップ x%の論文を「トップ x%の HCP」と定義して分析に用いる。

HCP を定義する研究分野には Open Alex における「Concept」の Level-0 として設定されている以下の 19 分類を用いた。

1. Mathematics
2. Physics
3. Chemistry
4. Biology
5. Medicine
6. Materials science
7. Engineering
8. Computer science
9. Geology
10. Geography
11. Environmental science
12. Psychology
13. Business
14. Economics
15. Political science
16. Sociology
17. History
18. Philosophy

19. Art

予測に用いる説明変数は 2009 年度以前に入手可能なデータを用いる。産学連携の取り組みが研究者の将来のパフォーマンスにどのような影響を与えるかを分析するために、特に注目する変数として、2000～2009 年度の産学連携（企業との共同発明特許・共著論文）の有無を示すダミー変数（firm_0009）を加えている。その他の変数は Weis-Jacobson (2021)を参考としている。特に、Weis-Jacobson (2021)において予測精度の向上への寄与の大きかったネットワーク中心性指標及びネットワーク上の位置をあらわす分散表現（node2vec）も含めている。また、ネットワーク中心性の指標として、3つの代表的な指標である固有ベクトル中心性（eigenvector centrality）、媒介中心性（betweenness centrality）、ページランク（pagerank）をそれぞれ用いることとする。

表 2 分析に用いる変数の一覧

変数名	定義
t1p_fund_pi_1019	2010～2019 年度に研究代表者として参画した科研費 PJ の総配分額 トップ 1%（被説明変数）
t10p_fund_pi_1019	2010～2019 年度に研究代表者として参画した科研費 PJ の総配分額 トップ 10%（被説明変数）
hcr1p_09	2010～2019 年度のトップ 1%の Highly Cited Paper（HCP）の有 無（被説明変数）
hcr10p_09	2010～2019 年度のトップ 10%の Highly Cited Paper（HCP）の有 無（被説明変数）
firm_0009	2000～2009 年度の産学連携（企業との共同発明特許・共著論文） の有無
age_10	2010 年度時点での研究経験年数（初めて科研費 PJ に参画してから の経過年数）
t10p_fund_pi_0009	2000～2009 年度に研究代表者として参画した科研費 PJ の総配分額 トップ 10%
nproj_pi_0009	2000～2009 年度に研究代表者として参画した科研費 PJ 数
nproj_co_0009	2000～2009 年度に研究分担者として参画した科研費 PJ 数
lnfund_pi_0009	2000～2009 年度に研究代表者として参画した科研費 PJ の総配分額 の対数
lnfund_co_0009	2000～2009 年度に研究分担者として参画した科研費 PJ の総配分額 の対数
npat_0009	2000～2009 年度に出願した特許数
npap_0009	2000～2009 年度に出版した論文数

hcr1p_09	2009 年度時点でのトップ 1%HCP の有無
hcr10p_09	2009 年度時点でのトップ 10%HCP の有無
npap_top1p_09	2009 年度時点でのトップ 1%HCP の数
npap_top10p_09	2009 年度時点でのトップ 10%HCP の数
hindex_09	2009 年度時点での研究者の h-index
nco_0009	2000～2009 年度に参画した科研費 PJ の共同研究者数
nco_pi_0009	2000～2009 年度に研究代表者として参画した科研費 PJ の共同研究者数
eigenvector_centrality	2000～2009 年度の科研費 PJ 共同研究ネットワークの固有ベクトル中心性
betweenness_centrality	2000～2009 年度の科研費 PJ 共同研究ネットワークの媒介中心性
pagerank	2000～2009 年度の科研費 PJ 共同研究ネットワークのページランク
node2vec	2000～2009 年度の node2vec によって定義される科研費 PJ 共同研究ネットワーク上の位置の特徴をあらわす分散表現 (64 次元)

分析に用いるモデルはロジスティック回帰モデルとランダム・フォレストの 2 種類を用いる。ランダム・フォレストは変数の非線形効果を柔軟に考慮できるため予測精度の高められるモデルである反面、結果の解釈が困難である。一方、ロジスティック回帰モデルはランダム・フォレストに予測精度は劣るが、結果の解釈が容易であるというメリットがある。

また、分析対象は 2000～2009 年度に研究代表者あるいは研究分担者として科研費に採択された 129,142 名の研究者であり、このうち無作為に抽出した 75%をモデルの学習に用いる訓練用データとして用い、残りの 25%を予測精度の検証用のデータとして用いる。

表 3 は分析対象の研究者のうち、2009 年度までのトップ 1%HCP の有無別の研究者数とその後の 10 年間でトップ 1%HCP を創出した研究者数とその割合を示している。2009 年度までにトップ 1%HCP を創出している研究者数約 3,111 人であり、その後の 10 年間でトップ 1%HCP を創出した研究者は 3,466 人であった。2009 年度までにトップ 1%HCP を創出した研究者の約 30%がその後の 10 年間でも継続してトップ 1%HCP を創出している。また、2009 年度時点で初めて科研費を獲得してから 10 年以内の若手の研究者 70,621 名に限定すると、そのうち 1,112 人が 2009 年度までにトップ 1%HCP を創出しており、2010～2019 年度にトップ 1%HCP を創出したのは 308 名である。

表 3 トップ 1%HCP の頻度

	2009 年度までの トップ 1% HCP	合計	2010～2019 年度に トップ 1%HCP あり	
			人数	%
全体	なし	126,016	2,572	2.0%

	あり	3,111	894	28.7%
	合計	129,127	3,466	2.7%
経験年数 10年以下	なし	69,509	1,422	2.0%
	あり	1,112	308	27.7%
	合計	70,621	1,730	2.4%

表 4 は 2000～2009 年度及び 2010～2019 年度の科研費獲得額トップ 10%の研究者数とその割合を示している。サンプル全体では 2000～2009 年度に科研費獲得額トップ 10%に入っている研究者は 12,917 名であり、そのうち 45%にあたる 5,815 名がその後の 10 年間でも科研費獲得額トップ 10%に継続して入っている。一方、2000～2009 年度に科研費獲得額トップ 10%に入っていない 7,136 名ではその後の 10 年間で科研費獲得額トップ 10%に入った研究者は 6.1%となっている。これらの割合は初めて科研費を獲得してから 10 年以内の若手研究者でも概ね同様である。

表 4 科研費獲得額トップ 10%の頻度

	2000～2009 年 度の科研費獲得 額トップ 10%	合計	2010～2019 年度の 科研費獲得額トッ プ 10%	
			人数	%
全体	なし	116,210	7,136	6.1%
	あり	12,917	5,815	45.0%
	合計	129,127	12,951	10.0%
経験年数 10年以下	なし	68,494	4,441	6.5%
	あり	2,127	1,091	51.3%
	合計	70,621	5,532	7.8%

表 5 にサンプル全体の記述統計、表 6 にサンプル全体の変数間の相関係数を示している。なお、付録の表 16、表 17、表 18 に分析に用いたサブサンプルの記述統計を示している。

表 5 記述統計(サンプル全体)

変数	N	平均	標準偏差	最小値	最大値
t1p_fund_pi_1019	129,127	0.0100	0.0995	0.0000	1.0000
t10p_fund_pi_0009	129,127	0.1003	0.3004	0.0000	1.0000
hcr1p_1019	129,127	0.0268	0.1616	0.0000	1.0000
hcr10p_1019	129,127	0.1393	0.3463	0.0000	1.0000
age_10	129,127	11.0501	7.3553	1.0000	25.0000
t10p_fund_pi_0009	129,127	0.1000	0.3000	0.0000	1.0000
t1p_fund_pi_0009	129,127	0.0100	0.0995	0.0000	1.0000
nproj_pi_0009	129,127	1.4878	1.6190	0.0000	25.0000
nproj_co_0009	129,127	1.7992	2.2958	0.0000	56.0000
lnfund_pi_0009	129,127	4.7975	3.0675	0.0000	13.1959
lnfund_co_0009	129,127	5.3713	3.5116	0.0000	13.5363

npat_0009	129,127	0.7855	5.0167	0.0000	316.0000
npap_0009	129,127	5.2809	15.3150	0.0000	768.0000
hcr1p_09	129,127	0.0241	0.1533	0.0000	1.0000
hcr10p_09	129,127	0.1229	0.3283	0.0000	1.0000
npap_top1p_09	129,127	0.0433	0.4232	0.0000	54.0000
npap_top10p_09	129,127	0.4574	2.2211	0.0000	152.0000
hindex_09	129,127	1.4122	3.3826	0.0000	65.0000
nco_0009	129,127	6.2920	7.6278	0.0000	167.0000
nco_pi_0009	129,127	1.3901	2.6717	0.0000	54.0000
firm_0009	129,127	0.1181	0.3227	0.0000	1.0000
eigenvector_centrality	129,127	0.0000	0.0028	0.0000	0.5140
betweenness_centrality	129,127	0.0000	0.0002	0.0000	0.0157
pagerank	129,127	0.0000	0.0000	0.0000	0.0001
node2vec_v1	129,127	0.0812	0.5852	-2.4519	2.6459
node2vec_v2	129,127	-0.6035	0.5800	-3.4423	1.3844
node2vec_v3	129,127	-0.4766	0.5898	-3.2972	2.0679
node2vec_v4	129,127	0.1049	0.5164	-2.2290	2.3335
node2vec_v5	129,127	0.1685	0.5144	-2.2282	2.7564
node2vec_v6	129,127	-0.1740	0.5220	-2.6525	2.2374
node2vec_v7	129,127	-0.0658	0.5751	-2.5508	2.3970
node2vec_v8	129,127	-0.4464	0.5545	-3.0698	2.0794
node2vec_v9	129,127	-0.4162	0.5857	-2.9005	1.8006
node2vec_v10	129,127	-0.0507	0.5128	-2.3590	2.3642
node2vec_v11	129,127	0.6260	0.5898	-1.3171	3.5888
node2vec_v12	129,127	-0.0001	0.4966	-2.4715	2.5712
node2vec_v13	129,127	0.1986	0.5243	-2.1284	2.5832
node2vec_v14	129,127	-0.4296	0.5575	-3.0263	1.7134
node2vec_v15	129,127	-0.1785	0.5049	-2.5956	2.0505
node2vec_v16	129,127	-0.0898	0.5294	-2.4564	1.9939
node2vec_v17	129,127	-0.1487	0.5403	-2.9284	2.3611
node2vec_v18	129,127	-0.2114	0.5291	-3.1128	1.9287
node2vec_v19	129,127	0.0956	0.5291	-2.2786	2.5105
node2vec_v20	129,127	0.6731	0.5805	-1.3164	3.8462
node2vec_v21	129,127	0.0529	0.5359	-2.5554	2.2730
node2vec_v22	129,127	0.5365	0.5762	-1.6413	3.2702
node2vec_v23	129,127	-0.1645	0.5240	-2.4926	1.9815
node2vec_v24	129,127	-0.5290	0.5640	-3.2105	1.4337
node2vec_v25	129,127	0.1327	0.5302	-2.2442	2.6653
node2vec_v26	129,127	-0.0019	0.5143	-2.1722	2.5573
node2vec_v27	129,127	-0.7725	0.6306	-3.7800	1.1278
node2vec_v28	129,127	0.4267	0.5540	-1.8807	3.0860
node2vec_v29	129,127	-0.1419	0.5013	-2.7095	2.0950
node2vec_v30	129,127	0.1327	0.5099	-2.2448	2.8839
node2vec_v31	129,127	0.4141	0.5692	-1.6267	3.3762
node2vec_v32	129,127	-0.3209	0.5452	-3.0547	1.8496
node2vec_v33	129,127	-0.1277	0.5053	-2.6092	2.2263

node2vec_v34	129,127	0.2775	0.6094	-2.0444	3.2219
node2vec_v35	129,127	-0.2337	0.5173	-2.9094	1.9662
node2vec_v36	129,127	-0.2154	0.5363	-3.0963	1.9941
node2vec_v37	129,127	-0.1363	0.5214	-2.5035	2.0356
node2vec_v38	129,127	-0.3711	0.5598	-3.2397	1.7670
node2vec_v39	129,127	-0.0825	0.5128	-2.5738	2.0709
node2vec_v40	129,127	0.3615	0.5489	-2.1121	2.8202
node2vec_v41	129,127	0.6749	0.5959	-1.2734	3.1549
node2vec_v42	129,127	0.2352	0.5537	-1.9540	2.9153
node2vec_v43	129,127	0.0512	0.5202	-2.1753	2.5402
node2vec_v44	129,127	-0.2638	0.5285	-3.0546	1.8711
node2vec_v45	129,127	-0.3994	0.5406	-3.1413	1.7425
node2vec_v46	129,127	0.1182	0.5179	-2.4605	2.3470
node2vec_v47	129,127	-0.0478	0.5510	-2.4696	2.3778
node2vec_v48	129,127	-0.1537	0.5566	-2.8159	2.1909
node2vec_v49	129,127	0.1508	0.5191	-1.9581	2.6690
node2vec_v50	129,127	-0.0972	0.5081	-2.8055	2.1979
node2vec_v51	129,127	0.1842	0.5226	-1.9124	2.8612
node2vec_v52	129,127	-0.2054	0.5223	-2.5596	2.1814
node2vec_v53	129,127	-0.0928	0.5058	-2.4320	2.1287
node2vec_v54	129,127	0.2287	0.5180	-1.9188	2.5807
node2vec_v55	129,127	0.3007	0.5820	-2.1828	3.2118
node2vec_v56	129,127	0.2100	0.5101	-1.8574	2.6346
node2vec_v57	129,127	-0.2154	0.5138	-2.9370	1.8099
node2vec_v58	129,127	0.1661	0.5123	-1.9534	2.7259
node2vec_v59	129,127	-0.2605	0.5515	-2.7394	1.8157
node2vec_v60	129,127	0.1231	0.5234	-2.3513	3.0047
node2vec_v61	129,127	-0.6397	0.5941	-3.5291	1.3676
node2vec_v62	129,127	-0.2067	0.5323	-2.7178	2.2157
node2vec_v63	129,127	-0.0230	0.5196	-2.4710	2.8211
node2vec_v64	129,127	0.3249	0.5676	-2.0796	3.0440

表 6 変数間の相関係数(サンプル全体)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
[1] t1p_fund_pi_1019	1.000												
[2] t10p_fund_pi_0009	0.301	1.000											
[3] hcr1p_1019	0.146	0.201	1.000										
[4] hcr10p_1019	0.164	0.338	0.413	1.000									
[5] age_10	0.081	0.072	0.013	-0.003	1.000								
[6] t10p_fund_pi_0009	0.248	0.388	0.143	0.230	0.301	1.000							
[7] t1p_fund_pi_0009	0.283	0.172	0.118	0.129	0.134	0.302	1.000						
[8] nproj_pi_0009	0.219	0.372	0.172	0.302	0.309	0.581	0.285	1.000					
[9] nproj_co_0009	0.122	0.231	0.101	0.187	0.269	0.290	0.108	0.332	1.000				
[10] Infund_pi_0009	0.134	0.275	0.115	0.226	0.223	0.421	0.179	0.716	0.168	1.000			
[11] Infund_co_0009	0.075	0.156	0.062	0.112	0.208	0.183	0.070	0.091	0.618	-0.145	1.000		
[12] npat_0009	0.113	0.144	0.113	0.189	0.036	0.145	0.100	0.133	0.079	0.107	0.052	1.000	

[13] npap_0009	0.210	0.303	0.356	0.529	0.145	0.333	0.247	0.394	0.288	0.264	0.151	0.298	1.000
[14] hcr1p_09	0.157	0.186	0.253	0.292	0.065	0.201	0.184	0.226	0.128	0.141	0.079	0.155	0.426
[15] hcr10p_09	0.159	0.306	0.300	0.556	0.094	0.289	0.164	0.366	0.230	0.258	0.141	0.218	0.626
[16] npap_top1p_09	0.165	0.142	0.227	0.207	0.059	0.164	0.201	0.172	0.105	0.107	0.058	0.141	0.421
[17] npap_top10p_09	0.223	0.241	0.325	0.381	0.118	0.279	0.290	0.318	0.204	0.193	0.108	0.235	0.730
[18] hindex_09	0.213	0.336	0.348	0.592	0.158	0.360	0.259	0.457	0.288	0.303	0.163	0.268	0.861
[19] nco_0009	0.123	0.213	0.058	0.084	0.294	0.325	0.134	0.271	0.722	0.168	0.568	0.042	0.166
[20] nco_pi_0009	0.149	0.227	0.079	0.118	0.369	0.512	0.231	0.475	0.381	0.466	0.173	0.065	0.222
[21] firm_0009	0.138	0.279	0.222	0.429	0.083	0.258	0.130	0.302	0.197	0.230	0.121	0.394	0.536
[22] eigenvector_centrality	0.001	0.007	0.000	-0.001	0.007	0.017	0.000	0.023	0.099	0.011	0.014	0.001	0.007
[23] betweenness_centrality	0.122	0.165	0.068	0.100	0.203	0.265	0.194	0.232	0.418	0.169	0.219	0.070	0.203
[24] pagerank	0.155	0.242	0.099	0.155	0.440	0.418	0.193	0.407	0.780	0.233	0.594	0.092	0.297

	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]
[14] hcr1p_09	1.000										
[15] hcr10p_09	0.420	1.000									
[16] npap_t~1p_09	0.651	0.273	1.000								
[17] npap_t~0p_09	0.568	0.550	0.735	1.000							
[18] hindex_09	0.499	0.759	0.461	0.769	1.000						
[19] nco_0009	0.080	0.113	0.071	0.130	0.149	1.000					
[20] nco_pi_0009	0.107	0.159	0.092	0.170	0.215	0.564	1.000				
[21] firm_0009	0.252	0.481	0.186	0.348	0.564	0.087	0.139	1.000			
[22] eigenvecto~y	0.002	0.011	0.001	0.006	0.012	0.032	0.022	0.007	1.000		
[23] betweennes~y	0.104	0.122	0.109	0.175	0.177	0.541	0.347	0.113	0.011	1.000	
[24] pagerank	0.138	0.213	0.117	0.218	0.288	0.803	0.614	0.194	0.042	0.496	1.000

2 将来のスター研究者の早期検知モデルの分析結果

表 7 は 2010～2019 年度のトップ 1%HCP の有無を被説明変数とするロジスティック回帰モデルの係数の推定結果を示している。t10p_fund_pi_0009 の係数が正であり、有意水準 10%未満であることから、2000～2009 年度に科研費獲得額トップ 10%に入っている研究者はそうでない研究者と比べて、将来の 10 年以内に HCP を生み出す可能性が高いことを示している。同様に、nproj_pi_0009、Infund_pi_0009、Infund_co_0009 の係数が正で有意水準 1%であり、2000～2009 年度に研究代表者として採択された科研費のプロジェクト数、研究代表者及び研究分担者として獲得した科研費プロジェクトの予算額が多いほど、将来の HCP 創出確率が高いことを示している。一方で nproj_co_0009 の係数は負で有意水準 1%未満であるため、研究分担者として多くの科研費プロジェクトに関与していると将来 HCP を創出する確率が低下する傾向にあることを示している。

次に、npat_0009 と npap_0009 の係数はいずれも統計的に有意に正(有意水準はそれぞれ 10%と 1%)であることから、多くの特許及び論文を出している研究者は将来 HCP を生み出す可能性が高い傾向が示されている。hcr1p_09 と hcr10p_09 の係数も正で有意であることから、2009 年度時点でトップ 1%HCP 及びトップ 10%HCP を生み出している研究者は将来トップ 1%HCP を生み出す可能性が高い傾向がある。2009 年時点でのトップ 1%HCP の数(npap_top1p_09)の係数は正で有意だが、トップ 10%HCP の数(npap_top10p_09)は反対に負で有意となっており、トップ 1%HCP

を多く出している研究者は将来トップ 1%HCP を創出する可能性は高いが、トップ 1%に入らないトップ 10%HCP を多く出している研究者はトップ 1%HCP を生み出す可能性は低くなる傾向を示している。

研究者の論文の量と質のバランスを示す h-index については正で有意となっており、現在の h-index は将来の HCP を創出する研究者のシグナルとして有効である可能性を示している。一方、2000～2009 年度における共同研究者の総数(nco_0009)及び研究代表者として関与した研究プロジェクトにおける共同研究者数(nco_pi_0009)の係数は有意ではなく、3つの共同研究ネットワークのデータに基づいて測定したネットワーク中心性指標(eigenvector_centrality、betweenness_centrality、pagerank)の係数はいずれも統計的に有意ではなく、科研費の共同研究ネットワークの情報から将来の HCP を創出する研究者の早期検知における有効性を示す結果は得られなかった。

また、尤度比検定の結果によれば、共同研究ネットワーク上の位置の特徴をあらゆる 64 次元の分散表現(node2vec)は有意水準 1%未満で将来のトップ 1%HCP の有無に対して有意な効果が見られる。

表 7 トップ 1%HCP の有無に関するロジスティック回帰モデルの分析結果(1)

	[1]	[2]	[3]	[4]	[5]	[6]
t10p_fund_pi_0009	0.132*	0.133*	0.132*	0.132*	0.133*	0.138*
	[0.0712]	[0.0712]	[0.0712]	[0.0712]	[0.0712]	[0.0731]
nproj_pi_0009	0.0341***	0.0341***	0.0343***	0.0337**	0.0338**	0.0256*
	[0.0132]	[0.0132]	[0.0132]	[0.0133]	[0.0133]	[0.0138]
nproj_co_0009	-0.0365***	-0.0356***	-0.0364***	-0.0377***	-0.0368***	-0.0455***
	[0.0102]	[0.0103]	[0.0102]	[0.0118]	[0.0119]	[0.0128]
lnfund_pi_0009	0.120***	0.120***	0.120***	0.120***	0.120***	0.109***
	[0.0131]	[0.0131]	[0.0131]	[0.0132]	[0.0132]	[0.0134]
lnfund_co_0009	0.0566***	0.0564***	0.0561***	0.0564***	0.0556***	0.0301***
	[0.00843]	[0.00844]	[0.00851]	[0.00850]	[0.00860]	[0.0102]
npat_0009	0.00378*	0.00377*	0.00381*	0.00376*	0.00379*	0.00364*
	[0.00202]	[0.00202]	[0.00202]	[0.00202]	[0.00202]	[0.00208]
npap_0009	0.0170***	0.0170***	0.0170***	0.0170***	0.0170***	0.0173***
	[0.00122]	[0.00122]	[0.00122]	[0.00122]	[0.00122]	[0.00127]
hcr1p_09	0.410***	0.411***	0.411***	0.410***	0.411***	0.399***
	[0.0797]	[0.0797]	[0.0797]	[0.0797]	[0.0797]	[0.0808]
hcr10p_09	1.592***	1.592***	1.591***	1.592***	1.592***	1.459***
	[0.0674]	[0.0674]	[0.0674]	[0.0675]	[0.0675]	[0.0679]
npap_top1p_09	0.113***	0.112***	0.113***	0.113***	0.112***	0.106***
	[0.0353]	[0.0353]	[0.0353]	[0.0353]	[0.0353]	[0.0357]
npap_top10p_09	-0.0638***	-0.0638***	-0.0638***	-0.0637***	-0.0636***	-0.0593***
	[0.00937]	[0.00937]	[0.00936]	[0.00938]	[0.00937]	[0.00960]
hindex_09	0.103***	0.103***	0.103***	0.103***	0.103***	0.0931***
	[0.00953]	[0.00953]	[0.00953]	[0.00953]	[0.00953]	[0.00999]
nco_0009	0.0058	0.00575	0.00659	0.00535	0.00615	0.0110*
	[0.00455]	[0.00455]	[0.00502]	[0.00508]	[0.00543]	[0.00614]
nco_pi_0009	-0.0089	-0.00886	-0.00928	-0.0096	-0.01	-0.0171

	[0.00996]	[0.00996]	[0.0100]	[0.0106]	[0.0107]	[0.0110]
eigenvector_centrality		-19.86			-19.79	-13.01
		[37.33]			[37.27]	[31.62]
betweenness_centrality			-33.22		-36.67	-114.4
			[89.45]		[89.84]	[91.05]
pagerank				1381	1464.5	8854
				[6981.1]	[7022.4]	[7843.8]
_cons	-4.792***	-4.791***	-4.793***	-4.793***	-4.794***	-4.647***
	[0.105]	[0.105]	[0.105]	[0.106]	[0.106]	[0.111]
*.age_10	Yes	Yes	Yes	Yes	Yes	Yes
node2vec_*	No	No	No	No	No	Yes
N	96,880	96,880	96,880	96,880	96,880	96,880
r2_p	0.283	0.283	0.283	0.283	0.283	0.298
AUC (train)	0.905	0.905	0.904	0.904	0.904	0.910
AUC (test)	0.902	0.902	0.902	0.902	0.902	0.908

注)カッコ内は標準誤差、* p < 0.1, ** p < 0.05, *** p < 0.01。

表 8 は産学連携への取り組みをあらわすダミー変数(firm_0009)を説明変数に加えた、2010～2019 年度のトップ 1%HCP の有無を被説明変数とするロジスティック回帰モデルの係数の推定結果を示している。注目している産学連携への取り組みをあらわすダミー変数(firm_0009)の係数は正で統計的に有意であり、産学連携に取り組んでいる研究者はそうでない研究者と比べて、将来 HCP を生み出す可能性が高いことを示唆する結果である。また、その係数の大きさは約 0.55 であり、これは産学連携に取り組んでいる研究者はそうでない研究者と比べて、他の条件が一定であれば将来 HCP を生み出す確率の期待値が 1.7 倍(=exp(0.55))大きいことを示している。一方、共同研究ネットワーク上の中心性の係数は統計的に有意ではなく、ネットワーク中心性が高い研究者は将来 HCP を生み出す確率が高いとは言えないという結果である。

なお、産学連携への取り組みを説明変数に加えたことで、特許の数(npat_0009)の係数が統計的に有意ではなくなった。また、尤度比検定の結果によれば、共同研究ネットワーク上の位置の特徴をあらわす 64 次元の分散表現(node2vec)は有意水準 1%未満で将来のトップ 1%HCP の有無に対して有意な効果がみられる。

表 8 トップ 1%HCP の有無に関するロジスティック回帰モデルの分析結果(2)

	[1]	[2]	[3]	[4]	[5]	[6]
t10p_fund_pi_0009	0.106	0.107	0.107	0.106	0.108	0.124*
	[0.0713]	[0.0713]	[0.0713]	[0.0713]	[0.0713]	[0.0731]
nproj_pi_0009	0.0374***	0.0373***	0.0376***	0.0374***	0.0375***	0.0281**
	[0.0132]	[0.0132]	[0.0132]	[0.0134]	[0.0134]	[0.0138]
nproj_co_0009	-0.0404***	-0.0396***	-0.0404***	-0.0403***	-0.0395***	-0.0475***
	[0.0102]	[0.0103]	[0.0102]	[0.0119]	[0.0120]	[0.0128]
lnfund_pi_0009	0.115***	0.115***	0.115***	0.115***	0.115***	0.104***
	[0.0131]	[0.0131]	[0.0131]	[0.0132]	[0.0132]	[0.0134]
lnfund_co_0009	0.0528***	0.0526***	0.0522***	0.0528***	0.0519***	0.0280***
	[0.00845]	[0.00845]	[0.00853]	[0.00851]	[0.00861]	[0.0102]
npat_0009	-0.00111	-0.00112	-0.00108	-0.00111	-0.00109	-0.000562

	[0.00220]	[0.00220]	[0.00220]	[0.00220]	[0.00220]	[0.00221]
npap_0009	0.0152***	0.0151***	0.0152***	0.0152***	0.0152***	0.0157***
	[0.00122]	[0.00122]	[0.00123]	[0.00123]	[0.00123]	[0.00128]
hcr1p_09	0.408***	0.408***	0.409***	0.408***	0.409***	0.393***
	[0.0798]	[0.0798]	[0.0798]	[0.0798]	[0.0798]	[0.0809]
hcr10p_09	1.487***	1.487***	1.486***	1.487***	1.486***	1.371***
	[0.0678]	[0.0678]	[0.0678]	[0.0679]	[0.0679]	[0.0681]
npap_top1p_09	0.101***	0.101***	0.101***	0.101***	0.101***	0.0939***
	[0.0352]	[0.0352]	[0.0352]	[0.0352]	[0.0352]	[0.0356]
npap_top10p_09	-0.0517***	-0.0516***	-0.0516***	-0.0517***	-0.0516***	-0.0466***
	[0.00944]	[0.00945]	[0.00944]	[0.00945]	[0.00945]	[0.00969]
hindex_09	0.0904***	0.0904***	0.0904***	0.0904***	0.0904***	0.0792***
	[0.00960]	[0.00960]	[0.00960]	[0.00961]	[0.00961]	[0.0101]
nco_0009	0.00826*	0.00823*	0.00937*	0.0083	0.00938*	0.0131**
	[0.00456]	[0.00456]	[0.00503]	[0.00509]	[0.00543]	[0.00614]
nco_pi_0009	-0.0106	-0.0106	-0.0111	-0.0105	-0.0111	-0.018
	[0.00997]	[0.00997]	[0.0100]	[0.0106]	[0.0107]	[0.0110]
firm_0009	0.545***	0.545***	0.545***	0.545***	0.546***	0.525***
	[0.0549]	[0.0549]	[0.0550]	[0.0550]	[0.0550]	[0.0555]
eigenvector_centrality		-23.06			-23.31	-14.67
		[37.90]			[38.10]	[32.56]
betweenness_centrality			-45.87		-47.36	-122
			[88.46]		[88.88]	[91.35]
pagerank				-123.7	11.35	7640.9
				[7004.6]	[7043.3]	[7860.7]
_cons	-4.798***	-4.798***	-4.800***	-4.798***	-4.799***	-4.658***
	[0.105]	[0.105]	[0.106]	[0.106]	[0.106]	[0.111]
*.age_10	Yes	Yes	Yes	Yes	Yes	Yes
node2vec_*	No	No	No	No	No	Yes
N	96,880	96,880	96,880	96,880	96,880	96,880
r2_p	0.287	0.287	0.287	0.287	0.287	0.302
AUC (train)	0.907	0.907	0.907	0.907	0.907	0.913
AUC (test)	0.904	0.904	0.904	0.904	0.904	0.909

注)カッコ内は標準誤差。* p < 0.1, ** p < 0.05, *** p < 0.01。

表 9 は 2010 年度時点において初めて研究代表者あるいは研究分担者として科研費に採択されてから 10 年以内の比較的若手と考えられる研究者にサンプルを限定した 2010~2019 年度のトップ 1%HCP の有無を被説明変数とするロジスティック回帰モデルの係数の推定結果を示している。注目している産学連携への取り組みをあらわすダミー変数(firm_0009)の統計的に有意な正の係数を含めて、サンプル全体の推定結果と概ね整合的な結果である。ただし、研究代表者あるいは研究分担者として関与する科研費プロジェクトの数(nproj_pi_0009 と nproj_co_0009)の係数はサンプル全体では統計的に有意(係数はそれぞれ正と負)であったが、若手研究者にサンプルを限定すると統計的に有意ではなくなった。

なお、有意水準 10%であるものの媒介中心性(betweenness_centrality)の係数が正で統計的にも有意となっており、若手研究者に関しては共同ネットワーク上の媒介中心性が高いほど、将来の

トップ 1%HCP を創出する可能性が高いことを示唆する結果であった。媒介中心性は複数の共同研究グループの接続するような位置にある場合に高くなる指標であり、若手研究者にとっては決まった共同研究グループではなく、複数の研究グループに所属していることが、将来の成長のために重要である可能性を示唆している可能性が考えられる。他方、固有ベクトル中心性 (eigenvector centrality) の係数は統計的に有意ではなく、ページランクは有意水準 10%の負の係数であった。

また、尤度比検定の結果によれば、共同研究ネットワーク上の位置の特徴をあらわす 64 次元の分散表現 (node2vec) は有意水準 1%未満で将来のトップ 1%HCP の有無に対して有意な効果が見られる。

表 9 トップ 1%HCP の有無に関するロジスティック回帰モデルの分析結果(若手サンプル)

	[1]	[2]	[3]	[4]	[5]	[6]
t10p_fund_pi_0009	0.029 [0.119]	0.0286 [0.119]	0.02 [0.120]	0.0327 [0.119]	0.0232 [0.120]	0.00186 [0.122]
nproj_pi_0009	0.0534 [0.0340]	0.0536 [0.0340]	0.0517 [0.0340]	0.0551 [0.0339]	0.0538 [0.0340]	0.055 [0.0345]
nproj_co_0009	-0.0128 [0.0216]	-0.0119 [0.0216]	-0.0146 [0.0217]	0.00836 [0.0241]	0.00785 [0.0243]	-0.0508* [0.0276]
lnfund_pi_0009	0.120*** [0.0177]	0.120*** [0.0177]	0.120*** [0.0177]	0.117*** [0.0178]	0.116*** [0.0178]	0.0956*** [0.0182]
lnfund_co_0009	0.0669*** [0.0120]	0.0666*** [0.0120]	0.0715*** [0.0123]	0.0735*** [0.0124]	0.0776*** [0.0127]	0.0594*** [0.0172]
npat_0009	-0.00478 [0.00345]	-0.0048 [0.00345]	-0.00487 [0.00346]	-0.00457 [0.00343]	-0.00468 [0.00344]	-0.00394 [0.00341]
npap_0009	0.0217*** [0.00257]	0.0217*** [0.00257]	0.0216*** [0.00257]	0.0221*** [0.00259]	0.0220*** [0.00259]	0.0230*** [0.00266]
hcr1p_09	0.471*** [0.135]	0.470*** [0.135]	0.468*** [0.135]	0.472*** [0.135]	0.468*** [0.135]	0.442*** [0.138]
hcr10p_09	1.248*** [0.0903]	1.248*** [0.0903]	1.247*** [0.0902]	1.244*** [0.0903]	1.242*** [0.0902]	1.122*** [0.0903]
npap_top1p_09	0.150** [0.0661]	0.150** [0.0661]	0.151** [0.0662]	0.151** [0.0662]	0.152** [0.0663]	0.149** [0.0681]
npap_top10p_09	-0.112*** [0.0191]	-0.112*** [0.0191]	-0.112*** [0.0191]	-0.113*** [0.0191]	-0.113*** [0.0191]	-0.0950*** [0.0194]
hindex_09	0.144*** [0.0164]	0.144*** [0.0164]	0.144*** [0.0163]	0.143*** [0.0164]	0.144*** [0.0164]	0.124*** [0.0169]
nco_0009	0.00027 [0.00823]	0.000552 [0.00825]	-0.0081 [0.00975]	0.00887 [0.00932]	0.00103 [0.0107]	-0.00118 [0.0127]
nco_pi_0009	-0.0355* [0.0213]	-0.0355* [0.0213]	-0.0324 [0.0215]	-0.0175 [0.0230]	-0.0142 [0.0232]	-0.0407 [0.0248]
firm_0009	0.548*** [0.0761]	0.550*** [0.0761]	0.543*** [0.0762]	0.552*** [0.0761]	0.549*** [0.0762]	0.512*** [0.0768]
eigenvector_centrality		-42.78 [81.73]			-60.14 [109.5]	-40.16 [92.93]
betweenness_centrality			699.1* [409.7]		696.1* [406.3]	443.4 [428.2]

pagerank				-28102.3*	-28327.9*	24018.1
				[14611.3]	[14645.6]	[18302.6]
_cons	-4.918***	-4.917***	-4.911***	-4.890***	-4.883***	-4.646***
	[0.117]	[0.117]	[0.118]	[0.118]	[0.118]	[0.124]
*.age_10	Yes	Yes	Yes	Yes	Yes	Yes
node2vec_*	No	No	No	No	No	Yes
N	53,141	53,141	53,141	53,141	53,141	53,141
r2_p	0.249	0.249	0.249	0.249	0.250	0.268
AUC (train)	0.899	0.899	0.899	0.899	0.899	0.903
AUC (test)	0.891	0.891	0.891	0.892	0.892	0.895

表 10 は若手サンプルからさらに 2009 年度時点で既にトップ 1%HCP のある研究者を除いたサンプルの 2010~2019 年度のトップ 1%HCP の有無を被説明変数とするロジスティック回帰モデルの係数の推定結果を示している。注目している産学連携への取り組みをあらわすダミー変数 (firm_0009) の統計的に有意な正の係数を含めて、サンプル全体の推定結果と概ね整合的な結果であるが、媒介中心性 (betweenness centrality) の係数は統計的に有意ではなくなった。他方、ページランクは有意水準 10% の負の係数であった。また、尤度比検定の結果によれば、共同研究ネットワーク上の位置の特徴をあらわす 64 次元の分散表現 (node2vec) は有意水準 1% 未満で将来のトップ 1%HCP の有無に対して有意な効果がみられる。

表 10 トップ 1%HCP の有無に関するロジスティック回帰モデルの分析結果 (若手&トップ 1%HCP なしのサンプル)

	[1]	[2]	[3]	[4]	[5]	[6]
t10p_fund_pi_0009	0.0253	0.0249	0.0225	0.0309	0.0273	0.0116
	[0.138]	[0.138]	[0.138]	[0.138]	[0.138]	[0.141]
nproj_pi_0009	0.0710*	0.0713*	0.0701*	0.0738*	0.0734*	0.0743*
	[0.0383]	[0.0383]	[0.0383]	[0.0383]	[0.0383]	[0.0388]
nproj_co_0009	-0.01	-0.00873	-0.0106	0.0282	0.0296	-0.0292
	[0.0248]	[0.0249]	[0.0249]	[0.0274]	[0.0276]	[0.0311]
lnfund_pi_0009	0.124***	0.124***	0.124***	0.118***	0.118***	0.0960***
	[0.0192]	[0.0192]	[0.0192]	[0.0193]	[0.0193]	[0.0197]
lnfund_co_0009	0.0761***	0.0756***	0.0777***	0.0883***	0.0895***	0.0706***
	[0.0132]	[0.0132]	[0.0135]	[0.0137]	[0.0140]	[0.0190]
npat_0009	-0.000701	-0.000723	-0.000752	-0.000389	-0.000462	0.000119
	[0.00390]	[0.00391]	[0.00392]	[0.00387]	[0.00388]	[0.00392]
npap_0009	0.0230***	0.0229***	0.0229***	0.0238***	0.0237***	0.0240***
	[0.00308]	[0.00308]	[0.00308]	[0.00310]	[0.00310]	[0.00319]
hcr10p_09	0.999***	0.999***	0.999***	0.990***	0.989***	0.881***
	[0.0991]	[0.0991]	[0.0991]	[0.0992]	[0.0992]	[0.0996]
npap_top10p_09	-0.103***	-0.103***	-0.103***	-0.104***	-0.104***	-0.0882***
	[0.0267]	[0.0267]	[0.0267]	[0.0267]	[0.0267]	[0.0273]
hindex_09	0.179***	0.179***	0.179***	0.178***	0.179***	0.159***
	[0.0190]	[0.0190]	[0.0190]	[0.0190]	[0.0190]	[0.0197]
nco_0009	-0.00774	-0.00739	-0.0108	0.00808	0.00558	0.00997
	[0.00935]	[0.00937]	[0.0108]	[0.0106]	[0.0119]	[0.0140]

nco_pi_0009	-0.0192 [0.0233]	-0.0193 [0.0232]	-0.0184 [0.0234]	0.0114 [0.0247]	0.0125 [0.0248]	-0.0145 [0.0266]
firm_0009	0.466*** [0.0838]	0.468*** [0.0838]	0.464*** [0.0839]	0.469*** [0.0838]	0.469*** [0.0839]	0.434*** [0.0845]
eigenvector_centrality		-45.94 [87.36]			-61.45 [108.4]	-45.11 [95.45]
betweenness_centrality			284.3 [488.2]		293.3 [483.1]	-97.77 [514.3]
pagerank				-50940.7*** [16362.5]	-51541.6*** [16391.1]	-3262.7 [20334.0]
_cons	-4.985*** [0.125]	-4.984*** [0.125]	-4.983*** [0.125]	-4.935*** [0.126]	-4.932*** [0.126]	-4.700*** [0.132]
*.age_10	Yes	Yes	Yes	Yes	Yes	Yes
node2vec_*	No	No	No	No	No	Yes
N	52,311	52,311	52,311	52,311	52,311	52,311
r2_p	0.214	0.214	0.214	0.215	0.215	0.236
AUC (train)	0.889	0.889	0.889	0.889	0.889	0.894
AUC (test)	0.883	0.883	0.883	0.883	0.883	0.887

表 11 は 2010～2019 年度のトップ 10%HCP の有無を被説明変数とするロジスティック回帰モデルの係数の推定結果を示している。注目している産学連携への取り組みをあらわすダミー変数 (firm_0009) の統計的に有意な正の係数を含めて、上記のトップ 1%HCP の有無を被説明変数とする推定結果と概ね整合的な結果である。ただし、媒介中心性 (betweenness_centrality) の係数は統計的に有意で正となり、固有ベクトル中心性とページランクの係数は統計的に有意で負となった。また、尤度比検定の結果によれば、共同研究ネットワーク上の位置の特徴をあらわす 64 次元の分散表現 (node2vec) は有意水準 1%未満で将来のトップ 1%HCP の有無に対して有意な効果がみられる。

表 11 トップ 10%HCP の有無に関するロジスティック回帰モデルの分析結果(サンプル全体)

	[1]	[2]	[3]	[4]	[5]	[6]
t10p_fund_pi_0009	0.0337 [0.0475]	0.0359 [0.0475]	0.028 [0.0476]	0.039 [0.0475]	0.035 [0.0476]	0.04 [0.0485]
nproj_pi_0009	0.0706*** [0.0113]	0.0708*** [0.0113]	0.0699*** [0.0114]	0.0764*** [0.0114]	0.0764*** [0.0114]	0.0637*** [0.0116]
nproj_co_0009	0.0172** [0.00805]	0.0218*** [0.00814]	0.0165** [0.00807]	0.0364*** [0.00885]	0.0418*** [0.00896]	0.00373 [0.00963]
lnfund_pi_0009	0.128*** [0.00682]	0.127*** [0.00683]	0.128*** [0.00682]	0.125*** [0.00686]	0.124*** [0.00686]	0.119*** [0.00704]
lnfund_co_0009	0.0622*** [0.00488]	0.0611*** [0.00489]	0.0640*** [0.00492]	0.0675*** [0.00497]	0.0688*** [0.00503]	0.0456*** [0.00613]
npat_0009	0.00211 [0.00208]	0.0021 [0.00208]	0.00206 [0.00208]	0.00237 [0.00208]	0.00233 [0.00209]	0.00357* [0.00210]
npap_0009	0.0287*** [0.00157]	0.0286*** [0.00157]	0.0286*** [0.00157]	0.0292*** [0.00157]	0.0290*** [0.00157]	0.0293*** [0.00159]
hcr1p_09	-0.0033	-0.00179	-0.00106	0.000974	0.00534	-0.0195

	[0.101]	[0.101]	[0.101]	[0.101]	[0.101]	[0.102]
hcr10p_09	1.240***	1.240***	1.240***	1.234***	1.233***	1.171***
	[0.0379]	[0.0379]	[0.0379]	[0.0379]	[0.0379]	[0.0382]
npap_top1p_09	0.276***	0.274***	0.273***	0.274***	0.268***	0.265***
	[0.0656]	[0.0655]	[0.0654]	[0.0656]	[0.0654]	[0.0657]
npap_top10p_09	-0.240***	-0.240***	-0.240***	-0.240***	-0.240***	-0.228***
	[0.0100]	[0.0100]	[0.0100]	[0.0100]	[0.0101]	[0.0102]
hindex_09	0.281***	0.281***	0.281***	0.281***	0.281***	0.260***
	[0.00759]	[0.00759]	[0.00759]	[0.00759]	[0.00759]	[0.00777]
nco_0009	-0.00106	-0.00153	-0.00428	0.00539*	0.00168	0.0158***
	[0.00283]	[0.00283]	[0.00308]	[0.00307]	[0.00326]	[0.00370]
nco_pi_0009	-0.0147**	-0.0146**	-0.0143**	-0.00174	-0.000225	-0.0220***
	[0.00669]	[0.00669]	[0.00670]	[0.00710]	[0.00710]	[0.00750]
firm_0009	0.631***	0.630***	0.629***	0.636***	0.633***	0.612***
	[0.0335]	[0.0336]	[0.0336]	[0.0336]	[0.0336]	[0.0341]
eigenvector_centrality		-14.13**			-14.64**	-13.45**
		[5.889]			[6.044]	[6.228]
betweenness_centrality			219.7***		256.1***	83.92
			[80.19]		[79.34]	[82.10]
pagerank				-23283.3***	-25061.9***	-9469.6*
				[4505.0]	[4534.7]	[5247.2]
_cons	-2.928***	-2.924***	-2.925***	-2.904***	-2.895***	-2.824***
	[0.0507]	[0.0508]	[0.0508]	[0.0510]	[0.0510]	[0.0538]
*.age_10	Yes	Yes	Yes	Yes	Yes	Yes
node2vec_*	No	No	No	No	No	Yes
N	96,880	96,880	96,880	96,880	96,880	96,880
r2_p	0.406	0.406	0.406	0.406	0.406	0.420
AUC (train)	0.924	0.924	0.924	0.924	0.925	0.928
AUC (test)	0.922	0.922	0.922	0.922	0.923	0.925

次に、図 3 はトップ 1%HCP の有無に関して、ロジスティック回帰モデルに加えて、ランダム・フォレストを用いた場合の予測精度を示している。予測精度の指標としては、ROC-AUC を用いる。ROC-AUC は分類モデルの予測精度を示す代表的な指標であり、誤差なく完全に予測できる場合は 1、でたらめな予測をしている場合に 0.5 をとり、1 に近いほど予測精度が高いことを示す。図 4 ではモデルの学習に用いた訓練用のデータに対する予測精度と学習には用いていない検証用データに対する予測精度をそれぞれ示している。ここで、「Base」はネットワーク指標以外の変数を全て説明変数に入れた場合の予測精度を示しており、それ以外はこの「Base」のモデルに各ネットワーク指標を説明変数として加えた場合の予測精度をそれぞれ示している。訓練用データに対する予測精度は過学習することがあるため、検証用のデータへの予測精度がより重要な指標である。ロジスティック回帰モデルを用いた場合は訓練用データ・検証用データの精度がほとんど変わらず、ROC-AUC が 0.91 前後となっている。ランダム・フォレストを用いた場合、検証用データの ROC-AUC は 0.93 となり、予測精度がやや改善することがわかる。ただし、ネットワーク指標を加えても予測精度はほとんど向上していない。

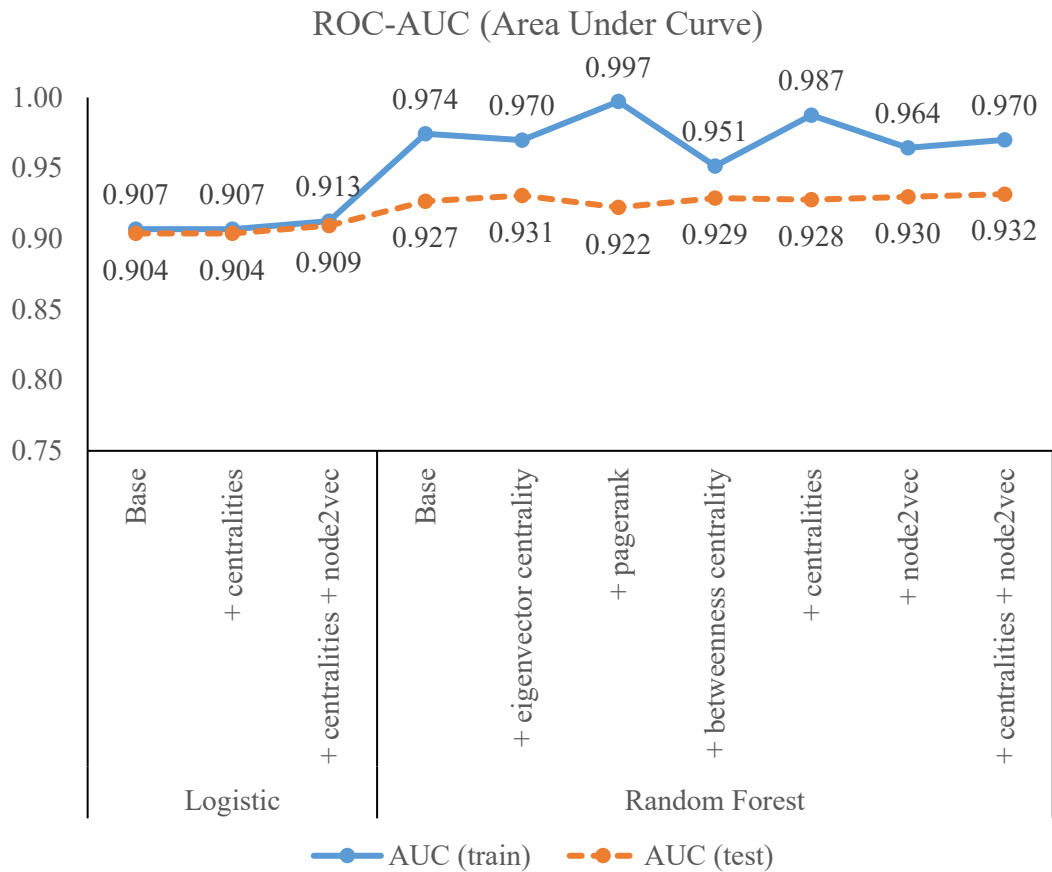


図 3 トップ 1%HCP の有無の予測モデルの精度(ROC-AUC)

次に、図 4 はランダム・フォレストを用いた予測における各変数の重要度を示している。共同研究ネットワーク上の位置の特徴をあらわす分散表現(node2vec)を説明変数に加えると重要度が高くなり、その他の変数の重要度が低くなる傾向がみられる。これは共同研究ネットワーク上の位置は将来のスターサイエンティストの予測において重要なシグナルとなっている可能性を示している。その他の変数としては、過去 1~3 年の論文の被引用件数や論文の質と量のバランスを示す h-index、科研費の獲得額の重要度が比較的高い。

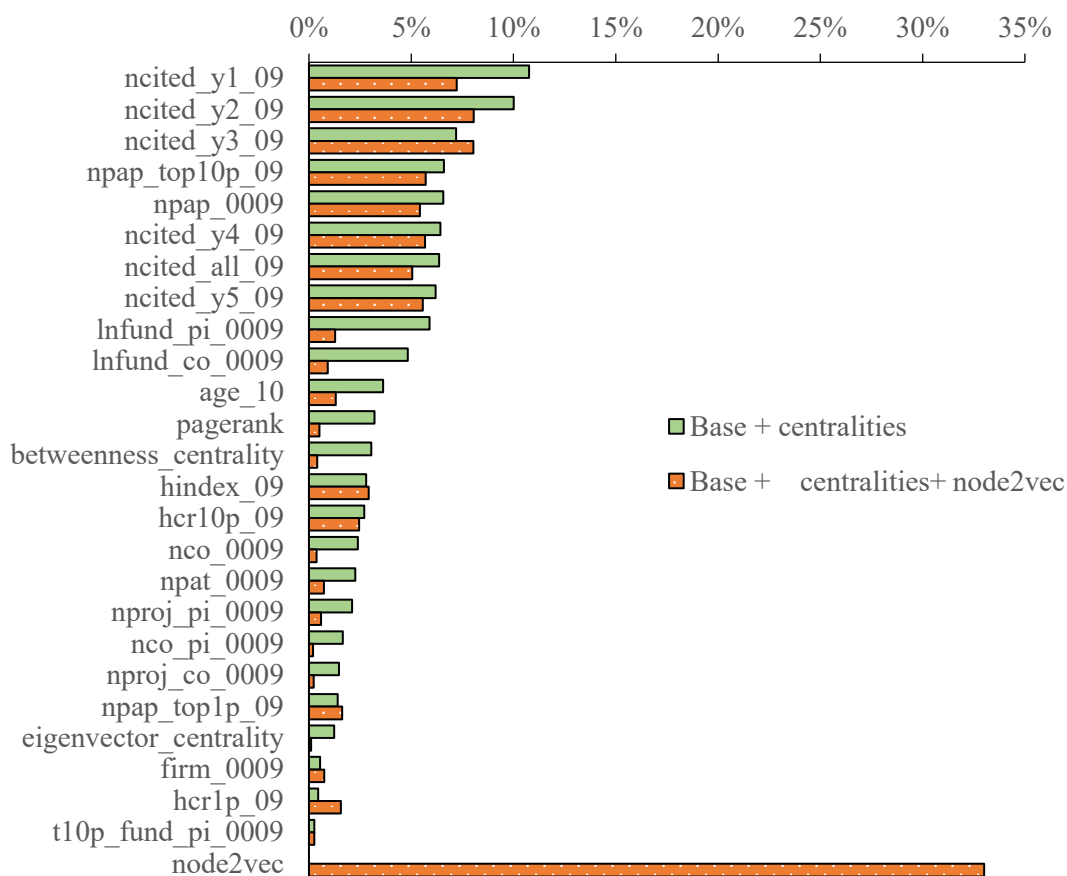


図 4 トップ 1%HCP の有無のランダム・フォレストにおける説明変数の重要度

表 12 は 2010～2019 年度の科研費獲得額トップ 10%に入っているか否かを被説明変数とするロジスティック回帰モデルの係数の推定結果を示している。t10p_fund_pi_0009 の係数が正であり、有意水準 10%未満であることから、2000～2009 年度に科研費獲得額トップ 10%に入っている研究者はそうでない研究者と比べて、将来の 10 年間で科研費獲得額トップ 10%に入る可能性も高いことを示している。同様に、nproj_pi_0009、nproj_co_0009、lnfund_pi_0009、lnfund_co_0009 の係数が全て正で有意水準 1%であり、2000～2009 年度に研究代表者または研究分担者として採択された科研費のプロジェクト数、研究代表者及び研究分担者として獲得した科研費プロジェクトの予算額が多いほど、将来の科研費獲得額トップ 10%に入る可能性が高いことを示している。

次に、2000～2009 年度における共同研究者の総数(nco_0009)及び研究代表者として関与した研究プロジェクトにおける共同研究者数(nco_pi_0009)の係数は正で有意であり、共同研究者が多いほど、将来の科研費獲得額が多い傾向を示している。一方、3つの共同研究ネットワークのデータに基づいて測定した固有ベクトル中心性の係数は負で有意であり、共同ネットワーク上の中心性が高いほど将来の科研費獲得額は少ない傾向を示している。ただし、残り2つのネットワーク中心性指標(betweenness centrality、pagerank)については統計的に有意ではなく、科研費の共同研究ネットワークの情報は将来の科研費獲得額が多くなる研究者の早期検知における有効性を示す結果は得られなかった。

また、尤度比検定の結果によれば、共同研究ネットワーク上の位置の特徴をあらわす 64 次元の分散表現(node2vec)は有意水準 1%未満で将来のトップ 1%HCP の有無に対して有意な効果が見られる。

表 12 科研費獲得額トップ 10%に関するロジスティック回帰モデルの分析結果(サンプル全体)

	[1]	[2]	[3]	[4]	[5]	[6]
t10p_fund_pi_0009	1.225*** [0.0385]	1.225*** [0.0385]	1.223*** [0.0386]	1.228*** [0.0386]	1.227*** [0.0386]	1.094*** [0.0394]
nproj_pi_0009	0.210*** [0.00918]	0.210*** [0.00918]	0.210*** [0.00919]	0.212*** [0.00927]	0.212*** [0.00927]	0.234*** [0.00966]
nproj_co_0009	0.0247*** [0.00620]	0.0268*** [0.00627]	0.0243*** [0.00621]	0.0301*** [0.00715]	0.0330*** [0.00723]	0.0401*** [0.00794]
lnfund_pi_0009	0.337*** [0.00949]	0.336*** [0.00949]	0.336*** [0.00948]	0.336*** [0.00950]	0.335*** [0.00950]	0.305*** [0.00945]
lnfund_co_0009	0.125*** [0.00487]	0.124*** [0.00488]	0.126*** [0.00489]	0.126*** [0.00493]	0.127*** [0.00497]	0.112*** [0.00589]
nco_0009	0.0178*** [0.00209]	0.0175*** [0.00209]	0.0165*** [0.00227]	0.0193*** [0.00231]	0.0179*** [0.00243]	0.0150*** [0.00288]
nco_pi_0009	-0.0458*** [0.00504]	-0.0457*** [0.00504]	-0.0455*** [0.00505]	-0.0433*** [0.00531]	-0.0422*** [0.00535]	-0.0301*** [0.00551]
eigenvector_centrality		-4.812** [2.307]			-4.873** [2.315]	-5.701** [2.320]
betweenness_centrality			80.68 [57.22]		95.78* [58.14]	-5.414 [59.71]
pagerank				-5743.1 [3797.0]	-7061.8* [3857.1]	4559.6 [4365.0]
_cons	-4.647*** [0.0725]	-4.645*** [0.0725]	-4.645*** [0.0725]	-4.642*** [0.0727]	-4.636*** [0.0727]	-4.417*** [0.0742]
*.age_10	Yes	Yes	Yes	Yes	Yes	Yes
node2vec_*	No	No	No	No	No	Yes
N (train)	96,891	96,891	96,891	96,891	96,891	96,891
r2_p (train)	0.254	0.254	0.254	0.254	0.254	0.268
AUC (train)	0.843	0.843	0.843	0.843	0.843	0.852
AUC (test)	0.838	0.838	0.838	0.838	0.838	0.845

注)カッコ内は標準誤差、* p < 0.1, ** p < 0.05, *** p < 0.01。

表 13 は 2010 年度時点において初めて研究代表者あるいは研究分担者として科研費に採択されてから 10 年以内の比較的若手と考えられる研究者にサンプルを限定した 2010~2019 年度の科研費獲得額トップ 10%に入っているか否かを被説明変数とするロジスティック回帰モデルの係数の推定結果を示している。媒介中心性 (betweenness_centrality) の係数が正で統計的にも有意となっており、若手研究者に関しては共同ネットワーク上の媒介中心性が高いほど、将来の科研費の獲得額が多い可能性が高いことを示唆する結果であった。そのような傾向は他の 2 つの中心性指標 (固有ベクトル中心性及びページランク) には見られなかった。媒介中心性は複数の共同研究グループの接続するような位置にある場合に高くなる指標であり、若手研究者にとっては決まった共同研究グループではなく、複数の研究グループに所属していることが、将来の成長のために重要である可能性を示唆している可能性が考えられる。

また、尤度比検定の結果によれば、共同研究ネットワーク上の位置の特徴をあらわす 64 次元の分散表現 (node2vec) は有意水準 1%未満で将来のトップ 1%HCP の有無に対して有意な効果がみられる。

表 13 科研費獲得額トップ 10%に関するロジスティック回帰モデルの分析結果(若手サンプル)

	[1]	[2]	[3]	[4]	[5]	[6]
t10p_fund_pi_0009	1.327*** [0.0676]	1.328*** [0.0676]	1.313*** [0.0678]	1.341*** [0.0678]	1.327*** [0.0680]	1.139*** [0.0701]
nproj_pi_0009	0.301*** [0.0226]	0.301*** [0.0226]	0.300*** [0.0226]	0.307*** [0.0226]	0.306*** [0.0226]	0.348*** [0.0235]
nproj_co_0009	0.110*** [0.0132]	0.109*** [0.0133]	0.106*** [0.0133]	0.148*** [0.0150]	0.143*** [0.0151]	0.0939*** [0.0173]
lnfund_pi_0009	0.306*** [0.0123]	0.306*** [0.0123]	0.305*** [0.0123]	0.300*** [0.0124]	0.299*** [0.0124]	0.261*** [0.0124]
lnfund_co_0009	0.144*** [0.00709]	0.145*** [0.00709]	0.150*** [0.00718]	0.155*** [0.00733]	0.160*** [0.00743]	0.172*** [0.0105]
nco_0009	-0.000697 [0.00395]	-0.000629 [0.00395]	-0.0104** [0.00451]	0.0125*** [0.00464]	0.00279 [0.00512]	-0.0104 [0.00638]
nco_pi_0009	-0.0767*** [0.0106]	-0.0770*** [0.0106]	-0.0758*** [0.0107]	-0.0534*** [0.0113]	-0.0523*** [0.0114]	-0.0342*** [0.0120]
eigenvector_centrality		25.85 [24.69]			22.78 [23.87]	30.26 [25.58]
betweenness_centrality			1130.5*** [236.7]		1145.9*** [236.6]	748.4*** [247.0]
pagerank				-44511.2*** [8318.2]	-44758.0*** [8341.2]	12521.5 [10476.8]
_cons	-4.601*** [0.0788]	-4.601*** [0.0788]	-4.591*** [0.0788]	-4.552*** [0.0793]	-4.543*** [0.0793]	-4.267*** [0.0809]
*.age_10	Yes	Yes	Yes	Yes	Yes	Yes
node2vec_*	No	No	No	No	No	Yes
N (train)	53,052	53,052	53,052	53,052	53,052	53,052
r2_p (train)	0.187	0.187	0.187	0.188	0.188	0.213
AUC (train)	0.810	0.810	0.810	0.813	0.813	0.823
AUC (test)	0.809	0.809	0.810	0.812	0.812	0.818

注)カッコ内は標準誤差、* p < 0.1, ** p < 0.05, *** p < 0.01。

表 14 は若手研究者のサンプルから、2000～2009 年度において既に科研費獲得額トップ 10%に入っている研究者を除いたサンプルでの 2010～2019 年度の科研費獲得額トップ 10%に入っているか否かを被説明変数とするロジスティック回帰モデルの係数の推定結果を示している。媒介中心性 (betweenness_centrality) の係数が正で統計的にも有意となっており、上記の若手研究者に関する結果の頑健性を示している。

また、尤度比検定の結果によれば、共同研究ネットワーク上の位置の特徴をあらわす 64 次元の分散表現 (node2vec) は有意水準 1%未満で将来のトップ 1%HCP の有無に対して有意な効果がみられる。

表 14 科研費獲得額トップ 10%に関するロジスティック回帰モデルの分析結果(若手&科研費獲得額
トップ 10%圏外サンプル)

	[1]	[2]	[3]	[4]	[5]	[6]
t10p_fund_pi_0009						
nproj_pi_0009	0.363*** [0.0295]	0.364*** [0.0295]	0.362*** [0.0296]	0.363*** [0.0296]	0.363*** [0.0296]	0.402*** [0.0305]
nproj_co_0009	0.124*** [0.0146]	0.122*** [0.0147]	0.119*** [0.0148]	0.171*** [0.0166]	0.165*** [0.0167]	0.0974*** [0.0194]
lnfund_pi_0009	0.288*** [0.0136]	0.288*** [0.0136]	0.286*** [0.0136]	0.282*** [0.0136]	0.281*** [0.0136]	0.242*** [0.0137]
lnfund_co_0009	0.154*** [0.00767]	0.154*** [0.00767]	0.160*** [0.00777]	0.167*** [0.00799]	0.175*** [0.00809]	0.195*** [0.0121]
nco_0009	-0.00634 [0.00430]	-0.00626 [0.00430]	-0.0188*** [0.00492]	0.0104** [0.00511]	-0.00198 [0.00563]	-0.0186*** [0.00710]
nco_pi_0009	-0.0882*** [0.0132]	-0.0885*** [0.0132]	-0.0874*** [0.0133]	-0.0558*** [0.0140]	-0.0549*** [0.0142]	-0.0243 [0.0148]
eigenvector_centrality		31.81 [27.51]			27.55 [26.15]	36.23 [28.36]
betweenness_centrality			1552.0*** [266.1]		1562.0*** [265.1]	1069.3*** [279.7]
pagerank				-55932.2*** [9391.3]	-56198.2*** [9409.7]	15790.2 [11745.2]
_cons	-4.578*** [0.0800]	-4.578*** [0.0800]	-4.565*** [0.0800]	-4.523*** [0.0805]	-4.511*** [0.0805]	-4.222*** [0.0822]
*.age_10	Yes	Yes	Yes	Yes	Yes	Yes
node2vec_*	No	No	No	No	No	Yes
N (train)	51,448	51,448	51,448	51,448	51,448	51,448
r2_p (train)	0.125	0.125	0.126	0.127	0.128	0.159
AUC (train)	0.797	0.797	0.797	0.800	0.800	0.813
AUC (test)	0.794	0.794	0.795	0.797	0.798	0.808

注)カッコ内は標準誤差、* p < 0.1, ** p < 0.05, *** p < 0.01。

表 15 は 2010～2019 年度の科研費獲得額トップ 1%に入っているか否かを被説明変数とするロジスティック回帰モデルの係数の推定結果を示している。上記の科研費獲得額トップ 10%に入っているか否かを被説明変数とする推定結果と概ね整合的な結果であった。ただし、研究分担者として参画した科研費プロジェクトの数(nproj_co_0009)の係数が統計的に有意ではなくなった。

表 15 科研費獲得額トップ 1%に関するロジスティック回帰モデルの分析結果(サンプル全体)

	[1]	[2]	[3]	[4]	[5]	[6]
t10p_fund_pi_0009	1.105*** [0.132]	1.105*** [0.132]	1.099*** [0.132]	1.119*** [0.132]	1.110*** [0.132]	0.979*** [0.134]
nproj_pi_0009	0.0703*** [0.0131]	0.0700*** [0.0131]	0.0707*** [0.0131]	0.0630*** [0.0134]	0.0629*** [0.0134]	0.0906*** [0.0146]
nproj_co_0009	0.00242	0.00479	0.00277	-0.0205	-0.0199	-0.00957

	[0.0116]	[0.0119]	[0.0116]	[0.0142]	[0.0145]	[0.0157]
Infund_pi_0009	1.029***	1.029***	1.034***	1.018***	1.025***	0.977***
	[0.0457]	[0.0457]	[0.0460]	[0.0459]	[0.0462]	[0.0492]
Infund_co_0009	0.0665***	0.0658***	0.0659***	0.0642***	0.0622***	0.0435***
	[0.0140]	[0.0140]	[0.0140]	[0.0140]	[0.0141]	[0.0148]
nco_0009	0.0200***	0.0197***	0.0214***	0.0139***	0.0161***	0.0143**
	[0.00475]	[0.00477]	[0.00514]	[0.00525]	[0.00548]	[0.00643]
nco_pi_0009	-0.0529***	-0.0527***	-0.0539***	-0.0626***	-0.0654***	-0.0582***
	[0.0108]	[0.0109]	[0.0109]	[0.0115]	[0.0117]	[0.0122]
eigenvector_centrality		-38.68			-23.59	-15.43
		[85.81]			[66.35]	[43.60]
betweenness_centrality			-56.34		-112.4	-148.0*
			[76.96]		[80.66]	[84.77]
pagerank				23401.8***	25435.2***	31109.7***
				[8155.7]	[8349.6]	[9023.2]
_cons	-12.99***	-12.99***	-13.02***	-12.94***	-12.98***	-12.58***
	[0.528]	[0.528]	[0.529]	[0.529]	[0.530]	[0.542]
*.age_10	Yes	Yes	Yes	Yes	Yes	Yes
node2vec_*	No	No	No	No	No	Yes
N (train)	96,880	96,880	96,880	96,880	96,880	96,880
r2_p (train)	0.361	0.361	0.361	0.362	0.362	0.380
AUC (train)	0.940	0.940	0.940	0.940	0.940	0.946
AUC (test)	0.939	0.939	0.939	0.938	0.938	0.940

次に、図 5 は科研費獲得額トップ 10%に入っているか否かに関して、ロジスティック回帰モデルに加えて、ランダム・フォレストを用いた場合の予測精度を示している。予測精度の指標としては、上記同様 ROC-AUC を用いる。図 5 ではモデルの学習に用いた訓練用のデータに対する予測精度と学習には用いていない検証用データに対する予測精度をそれぞれ示している。ここで、「Base」はネットワーク指標以外の変数を全て説明変数に入れた場合の予測精度を示しており、それ以外はこの「Base」のモデルに各ネットワーク指標を説明変数として加えた場合の予測精度をそれぞれ示している。訓練用データに対する予測精度は過学習することがあるため、検証用のデータへの予測精度がより重要な指標である。ロジスティック回帰モデルを用いた場合は訓練用データ・検証用データの精度がほとんど変わらず、ROC-AUC が 0.85 前後となっている。ランダム・フォレストを用いた場合、検証用データの ROC-AUC は 0.86 となり、予測精度がやや改善することがわかる。ただし、ネットワーク指標を加えても予測精度はほとんど向上していない。

一方、「Base+centralities+node2vec+npat+npap」に示すように、特許・論文データを説明変数に加えると AUC は 0.88 と予測精度が改善している。

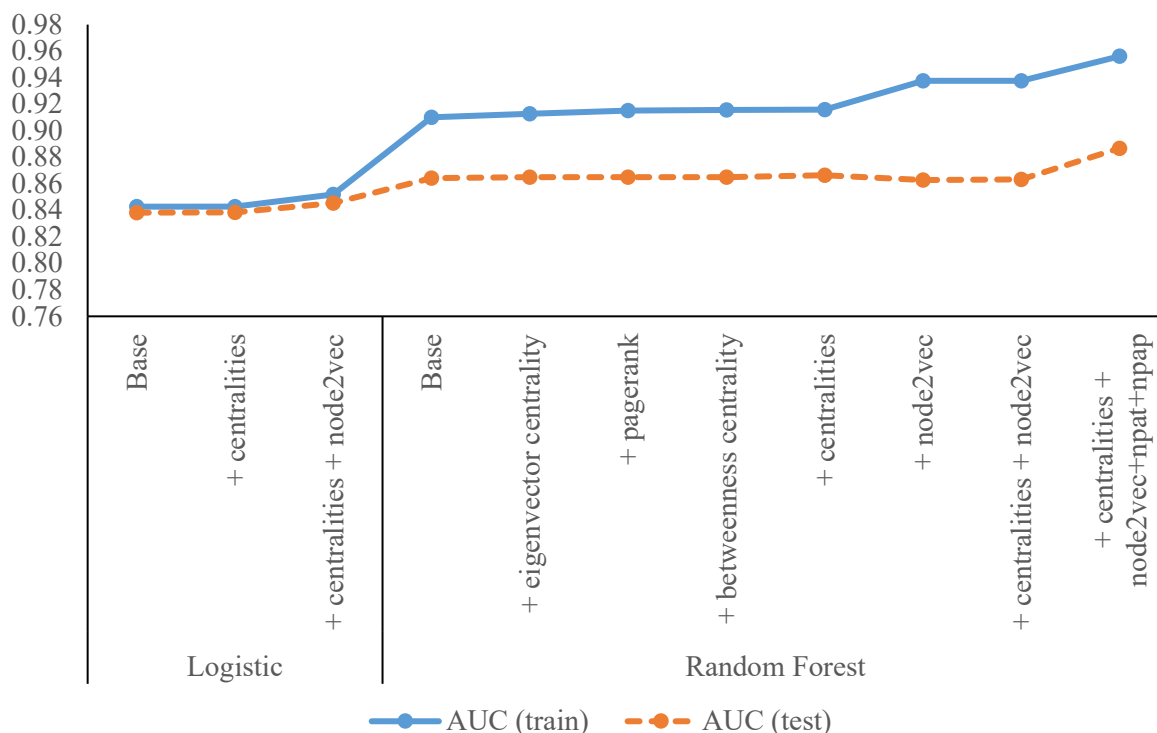


図 5 科研費獲得額トップ 10%の予測モデルの精度 (ROC-AUC)

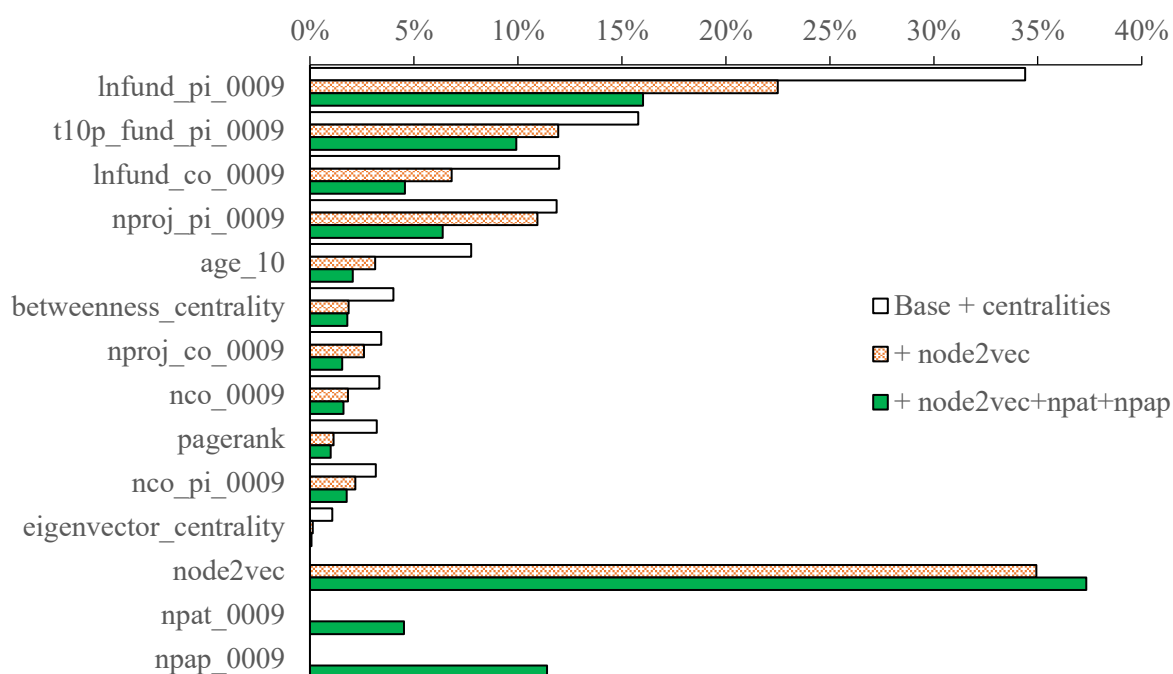


図 6 科研費獲得額トップ 10%を予測するランダム・フォレストにおける説明変数の重要度

3 まとめ

本調査研究では、スター研究者の早期検知システム DELPHI-R の試行的な開発を行った。科研費の研究プロジェクト由来のデータと特許・論文データを接続したデータベースを構築し、分析を行

ったところ、将来の 10 年間の間に分野別のトップ 1% の高頻度引用論文 (HCP) を生み出す研究者を十分に高い精度 (ROC-AUC が 93%) で予測可能なモデルを開発することができた。さらに、産学連携への取り組みは将来のスターサイエンティストを予測するためのシグナルとして利用できる可能性を示唆する結果が得られた。また、将来科研費獲得額が上位となるかどうかを予測するモデルも推定したところ、こちらも十分に高い予測精度 (ROC-AUC が 88%) を示す結果が得られた。

今後の課題としては、さらなる予測精度の向上を目指し、分野の違い (理系・文系) や国際連携の指標 (海外の研究者との共同発明特許・共著論文)、特許の共同発明ネットワーク、論文の共著ネットワーク、特許及び論文の引用ネットワークを用いたネットワーク指標、所属機関や地域の特性、所属機関の変化の有無など、予測に用いる説明変数をさらに充実させることである。また、「日本の研究.com」などの他のデータソースを接続することで、将来国プロを獲得する研究者を早期検知するようなシステムの開発も有効であると考えられる。

謝辞

本研究は国立大学法人政策研究大学院大学・三菱電機株式会社共同研究「エビデンスに基づく研究開発戦略支援システムの構築」の元で実施されたものである。本論文の執筆にあたり、三菱電機株式会社堀田厚氏による様々なサポートを受けた。記して感謝する。

参考文献

Weis, J. W., & Jacobson, J. M.

2021. “Learning on knowledge graph dynamics provides an early warning of impactful research.” *Nature Biotechnology*. <https://doi.org/10.1038/s41587-021-00907-6>

科研費 PJ「大学等の公的研究機関に関する制度改革と経済社会インパクトに関する研究」
(23H00843、研究代表者:元橋一之東京大学教授)

<https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-23K25540/>

特許庁の一括ダウンロードサービスについて

<https://www.jpo.go.jp/system/laws/sesaku/data/download.html>

IIP パテントデータベースについて <https://www.iip.or.jp/patentdb/>

OpenAlex について <https://openalex.org/>

付録

表 16 記述統計(初の科研費採択後 10 年以内の若手サンプル)

変数	N	平均	標準偏差	最小値	最大値
t1p_fund_pi_1019	70,621	0.0036	0.0599	0.0000	1.0000
t10p_fund_pi_0009	70,621	0.0783	0.2687	0.0000	1.0000
hcr1p_1019	70,621	0.0245	0.1546	0.0000	1.0000
hcr10p_1019	70,621	0.1346	0.3413	0.0000	1.0000
age_10	70,621	5.2708	2.7911	1.0000	10.0000
t10p_fund_pi_0009	70,621	0.0301	0.1709	0.0000	1.0000
t1p_fund_pi_0009	70,621	0.0008	0.0286	0.0000	1.0000
nproj_pi_0009	70,621	1.0602	1.0657	0.0000	17.0000
nproj_co_0009	70,621	1.2733	1.5530	0.0000	27.0000
lnfund_pi_0009	70,621	4.2111	2.9483	0.0000	11.2427
lnfund_co_0009	70,621	4.8009	3.5406	0.0000	12.8749
npat_0009	70,621	0.6384	4.7589	0.0000	316.0000
npap_0009	70,621	3.3696	9.7695	0.0000	323.0000
hcr1p_09	70,621	0.0157	0.1245	0.0000	1.0000
hcr10p_09	70,621	0.0940	0.2918	0.0000	1.0000
npap_top1p_09	70,621	0.0241	0.2611	0.0000	17.0000
npap_top10p_09	70,621	0.2493	1.2284	0.0000	64.0000
hindex_09	70,621	0.9393	2.3129	0.0000	39.0000
nco_0009	70,621	4.4738	5.4315	0.0000	88.0000
nco_pi_0009	70,621	0.6039	1.5888	0.0000	38.0000
firm_0009	70,621	0.0929	0.2903	0.0000	1.0000
eigenvector_centrality	70,621	0.0000	0.0015	0.0000	0.2431
betweenness_centrality	70,621	0.0000	0.0001	0.0000	0.0028
pagerank	70,621	0.0000	0.0000	0.0000	0.0001
node2vec_v1	70,621	0.0629	0.5576	-2.4519	2.5741
node2vec_v2	70,621	-0.5446	0.5866	-3.3758	1.3844
node2vec_v3	70,621	-0.4253	0.5842	-3.2972	2.0679
node2vec_v4	70,621	0.0903	0.4988	-2.2290	2.3191
node2vec_v5	70,621	0.1477	0.4974	-2.1624	2.7564
node2vec_v6	70,621	-0.1594	0.5020	-2.6525	2.2374
node2vec_v7	70,621	-0.0857	0.5482	-2.5508	2.3970
node2vec_v8	70,621	-0.4012	0.5491	-3.0698	2.0794
node2vec_v9	70,621	-0.3669	0.5713	-2.9005	1.8006
node2vec_v10	70,621	-0.0513	0.4929	-2.2942	2.1722
node2vec_v11	70,621	0.5769	0.6002	-1.3171	3.5712
node2vec_v12	70,621	-0.0094	0.4767	-2.3408	2.5712
node2vec_v13	70,621	0.1998	0.5046	-2.0180	2.5207
node2vec_v14	70,621	-0.3946	0.5525	-3.0263	1.5861
node2vec_v15	70,621	-0.1561	0.4894	-2.5956	2.0505
node2vec_v16	70,621	-0.0752	0.5000	-2.1987	1.9630
node2vec_v17	70,621	-0.1387	0.5179	-2.8713	1.9880
node2vec_v18	70,621	-0.1834	0.5075	-2.8765	1.9287

node2vec_v19	70,621	0.0852	0.5061	-2.2285	2.3345
node2vec_v20	70,621	0.6191	0.5950	-1.1552	3.5770
node2vec_v21	70,621	0.0660	0.5111	-2.5554	2.2730
node2vec_v22	70,621	0.4961	0.5680	-1.6413	3.1067
node2vec_v23	70,621	-0.1656	0.5132	-2.4926	1.8613
node2vec_v24	70,621	-0.4893	0.5637	-3.2105	1.4337
node2vec_v25	70,621	0.1297	0.5092	-2.2155	2.4550
node2vec_v26	70,621	-0.0144	0.4857	-2.1722	2.5573
node2vec_v27	70,621	-0.7197	0.6510	-3.5361	1.0243
node2vec_v28	70,621	0.3919	0.5484	-1.6927	2.9977
node2vec_v29	70,621	-0.1409	0.4805	-2.6732	2.0950
node2vec_v30	70,621	0.1071	0.4865	-2.1103	2.8839
node2vec_v31	70,621	0.3626	0.5539	-1.6267	3.3237
node2vec_v32	70,621	-0.2931	0.5298	-3.0292	1.7873
node2vec_v33	70,621	-0.1172	0.4907	-2.5568	2.2263
node2vec_v34	70,621	0.2754	0.6001	-1.7925	3.2219
node2vec_v35	70,621	-0.2169	0.5016	-2.9094	1.9662
node2vec_v36	70,621	-0.2051	0.5190	-3.0030	1.7998
node2vec_v37	70,621	-0.1484	0.5045	-2.5035	1.9912
node2vec_v38	70,621	-0.3055	0.5393	-3.0365	1.7670
node2vec_v39	70,621	-0.0804	0.4845	-2.5621	2.0709
node2vec_v40	70,621	0.3318	0.5386	-2.0558	2.7178
node2vec_v41	70,621	0.6377	0.6143	-1.1091	3.1009
node2vec_v42	70,621	0.1964	0.5289	-1.9540	2.9153
node2vec_v43	70,621	0.0368	0.4970	-2.1753	2.5402
node2vec_v44	70,621	-0.2406	0.5133	-3.0546	1.7454
node2vec_v45	70,621	-0.3618	0.5332	-3.1413	1.6125
node2vec_v46	70,621	0.1098	0.4975	-2.4605	2.3470
node2vec_v47	70,621	-0.0557	0.5249	-2.4696	2.3778
node2vec_v48	70,621	-0.1484	0.5346	-2.8147	2.1909
node2vec_v49	70,621	0.1363	0.4990	-1.9581	2.6690
node2vec_v50	70,621	-0.0899	0.4906	-2.6860	2.1332
node2vec_v51	70,621	0.1843	0.5072	-1.7618	2.4284
node2vec_v52	70,621	-0.1738	0.4960	-2.5596	2.1814
node2vec_v53	70,621	-0.0816	0.4832	-2.3974	2.1287
node2vec_v54	70,621	0.2111	0.5037	-1.9103	2.5807
node2vec_v55	70,621	0.2660	0.5560	-1.8284	3.2118
node2vec_v56	70,621	0.1953	0.4906	-1.6990	2.6346
node2vec_v57	70,621	-0.1993	0.4963	-2.8545	1.7148
node2vec_v58	70,621	0.1425	0.4930	-1.9534	2.7259
node2vec_v59	70,621	-0.2312	0.5264	-2.7394	1.8153
node2vec_v60	70,621	0.1237	0.5013	-2.3513	2.9668
node2vec_v61	70,621	-0.5833	0.6022	-3.5291	1.2954
node2vec_v62	70,621	-0.1944	0.5120	-2.6201	2.2157
node2vec_v63	70,621	-0.0287	0.4966	-2.4710	2.8211
node2vec_v64	70,621	0.2941	0.5488	-2.0796	3.0429

表 17 記述統計(若手&トップ1%HCPなしのサンプル)

変数	N	平均	標準偏差	最小値	最大値
t1p_fund_pi_1019	69,509	0.0030	0.0545	0.0000	1.0000
t10p_fund_pi_0009	69,509	0.0736	0.2611	0.0000	1.0000
hcr1p_1019	69,509	0.0205	0.1416	0.0000	1.0000
hcr10p_1019	69,509	0.1243	0.3299	0.0000	1.0000
age_10	69,509	5.2620	2.7908	1.0000	10.0000
t10p_fund_pi_0009	69,509	0.0273	0.1631	0.0000	1.0000
t1p_fund_pi_0009	69,509	0.0006	0.0254	0.0000	1.0000
nproj_pi_0009	69,509	1.0428	1.0427	0.0000	17.0000
nproj_co_0009	69,509	1.2580	1.5241	0.0000	27.0000
lnfund_pi_0009	69,509	4.1777	2.9453	0.0000	11.2048
lnfund_co_0009	69,509	4.7812	3.5306	0.0000	12.8749
npat_0009	69,509	0.5606	4.1904	0.0000	282.0000
npap_0009	69,509	2.9396	8.3724	0.0000	184.0000
hcr1p_09	69,509	0.0000	0.0000	0.0000	0.0000
hcr10p_09	69,509	0.0795	0.2705	0.0000	1.0000
npap_top1p_09	69,509	0.0000	0.0000	0.0000	0.0000
npap_top10p_09	69,509	0.1663	0.7588	0.0000	20.0000
hindex_09	69,509	0.8168	2.0208	0.0000	24.0000
nco_0009	69,509	4.4589	5.4141	0.0000	88.0000
nco_pi_0009	69,509	0.5966	1.5814	0.0000	38.0000
firm_0009	69,509	0.0855	0.2796	0.0000	1.0000
eigenvector_centrality	69,509	0.0000	0.0015	0.0000	0.2431
betweenness_centrality	69,509	0.0000	0.0001	0.0000	0.0028
pagerank	69,509	0.0000	0.0000	0.0000	0.0001
node2vec_v1	69,509	0.0595	0.5576	-2.4519	2.5741
node2vec_v2	69,509	-0.5439	0.5862	-3.3758	1.3844
node2vec_v3	69,509	-0.4252	0.5847	-3.2972	2.0679
node2vec_v4	69,509	0.0875	0.4985	-2.2290	2.3191
node2vec_v5	69,509	0.1482	0.4974	-2.1624	2.7564
node2vec_v6	69,509	-0.1581	0.5024	-2.6525	2.2374
node2vec_v7	69,509	-0.0857	0.5470	-2.5508	2.2763
node2vec_v8	69,509	-0.3986	0.5485	-3.0698	2.0794
node2vec_v9	69,509	-0.3641	0.5713	-2.9005	1.8006
node2vec_v10	69,509	-0.0510	0.4926	-2.2942	2.1722
node2vec_v11	69,509	0.5760	0.6003	-1.3171	3.5712
node2vec_v12	69,509	-0.0092	0.4764	-2.3408	2.5712
node2vec_v13	69,509	0.2017	0.5046	-1.9650	2.5207
node2vec_v14	69,509	-0.3946	0.5524	-3.0263	1.5861
node2vec_v15	69,509	-0.1547	0.4892	-2.5956	2.0505
node2vec_v16	69,509	-0.0762	0.4995	-2.1987	1.9630
node2vec_v17	69,509	-0.1372	0.5180	-2.8713	1.9880
node2vec_v18	69,509	-0.1813	0.5072	-2.8765	1.9287
node2vec_v19	69,509	0.0873	0.5059	-2.2285	2.3345
node2vec_v20	69,509	0.6207	0.5958	-1.1552	3.5770

node2vec_v21	69,509	0.0668	0.5093	-2.5554	2.2730
node2vec_v22	69,509	0.4952	0.5681	-1.6413	3.1067
node2vec_v23	69,509	-0.1668	0.5138	-2.4926	1.8613
node2vec_v24	69,509	-0.4889	0.5638	-3.2105	1.4337
node2vec_v25	69,509	0.1290	0.5093	-2.2155	2.4550
node2vec_v26	69,509	-0.0150	0.4852	-2.0837	2.5573
node2vec_v27	69,509	-0.7205	0.6514	-3.5361	1.0243
node2vec_v28	69,509	0.3919	0.5486	-1.6927	2.9977
node2vec_v29	69,509	-0.1413	0.4806	-2.6732	2.0950
node2vec_v30	69,509	0.1075	0.4865	-2.1103	2.8839
node2vec_v31	69,509	0.3619	0.5538	-1.6267	3.3237
node2vec_v32	69,509	-0.2946	0.5301	-3.0292	1.7873
node2vec_v33	69,509	-0.1167	0.4908	-2.5568	2.2263
node2vec_v34	69,509	0.2796	0.6009	-1.7925	3.2219
node2vec_v35	69,509	-0.2180	0.5015	-2.9094	1.9662
node2vec_v36	69,509	-0.2048	0.5189	-3.0030	1.7998
node2vec_v37	69,509	-0.1498	0.5044	-2.5035	1.9912
node2vec_v38	69,509	-0.3047	0.5386	-3.0365	1.7670
node2vec_v39	69,509	-0.0813	0.4844	-2.5621	2.0709
node2vec_v40	69,509	0.3306	0.5384	-2.0558	2.7178
node2vec_v41	69,509	0.6398	0.6153	-1.1091	3.1009
node2vec_v42	69,509	0.1935	0.5289	-1.9540	2.9153
node2vec_v43	69,509	0.0365	0.4968	-2.1753	2.5402
node2vec_v44	69,509	-0.2404	0.5135	-3.0546	1.7454
node2vec_v45	69,509	-0.3622	0.5338	-3.1413	1.6125
node2vec_v46	69,509	0.1105	0.4972	-2.4605	2.3470
node2vec_v47	69,509	-0.0527	0.5237	-2.4696	2.3778
node2vec_v48	69,509	-0.1467	0.5351	-2.8147	2.1909
node2vec_v49	69,509	0.1358	0.4992	-1.9581	2.6690
node2vec_v50	69,509	-0.0903	0.4906	-2.6860	2.1332
node2vec_v51	69,509	0.1848	0.5074	-1.7618	2.4284
node2vec_v52	69,509	-0.1734	0.4959	-2.5596	2.1814
node2vec_v53	69,509	-0.0798	0.4834	-2.3716	2.1287
node2vec_v54	69,509	0.2095	0.5039	-1.9103	2.5807
node2vec_v55	69,509	0.2633	0.5551	-1.8284	3.2118
node2vec_v56	69,509	0.1943	0.4904	-1.6990	2.6346
node2vec_v57	69,509	-0.1996	0.4962	-2.8545	1.7148
node2vec_v58	69,509	0.1406	0.4925	-1.9534	2.7259
node2vec_v59	69,509	-0.2310	0.5264	-2.7394	1.8153
node2vec_v60	69,509	0.1245	0.5014	-2.3157	2.9668
node2vec_v61	69,509	-0.5826	0.6024	-3.5291	1.2954
node2vec_v62	69,509	-0.1923	0.5120	-2.6201	2.2157
node2vec_v63	69,509	-0.0292	0.4961	-2.4710	2.8211
node2vec_v64	69,509	0.2948	0.5479	-2.0796	3.0429

表 18 記述統計(若手&科研費獲得額トップ 10%圏外のサンプル)

変数	N	平均	標準偏差	最小値	最大値
t1p_fund_pi_1019	68,494	0.0016	0.0397	0.0000	1.0000
t10p_fund_pi_0009	68,494	0.0648	0.2462	0.0000	1.0000
hcr1p_1019	68,494	0.0220	0.1467	0.0000	1.0000
hcr10p_1019	68,494	0.1258	0.3316	0.0000	1.0000
age_10	68,494	5.2086	2.7825	1.0000	10.0000
t10p_fund_pi_0009	68,494	0.0000	0.0000	0.0000	0.0000
t1p_fund_pi_0009	68,494	0.0000	0.0000	0.0000	0.0000
nproj_pi_0009	68,494	0.9870	0.9273	0.0000	8.0000
nproj_co_0009	68,494	1.2284	1.4674	0.0000	26.0000
lnfund_pi_0009	68,494	4.0812	2.8974	0.0000	7.8713
lnfund_co_0009	68,494	4.7447	3.5179	0.0000	12.8749
npat_0009	68,494	0.5656	4.5712	0.0000	316.0000
npap_0009	68,494	3.0227	9.0389	0.0000	323.0000
hcr1p_09	68,494	0.0129	0.1130	0.0000	1.0000
hcr10p_09	68,494	0.0846	0.2783	0.0000	1.0000
npap_top1p_09	68,494	0.0188	0.2256	0.0000	17.0000
npap_top10p_09	68,494	0.2099	1.0955	0.0000	64.0000
hindex_09	68,494	0.8502	2.1542	0.0000	39.0000
nco_0009	68,494	4.3382	5.2414	0.0000	70.0000
nco_pi_0009	68,494	0.5253	1.4022	0.0000	26.0000
firm_0009	68,494	0.0848	0.2785	0.0000	1.0000
eigenvector_centrality	68,494	0.0000	0.0015	0.0000	0.2431
betweenness_centrality	68,494	0.0000	0.0001	0.0000	0.0025
pagerank	68,494	0.0000	0.0000	0.0000	0.0000
node2vec_v1	68,494	0.0606	0.5566	-2.4519	2.5741
node2vec_v2	68,494	-0.5417	0.5873	-3.3758	1.3844
node2vec_v3	68,494	-0.4220	0.5839	-3.2972	2.0679
node2vec_v4	68,494	0.0873	0.4975	-2.2290	2.3191
node2vec_v5	68,494	0.1448	0.4963	-2.1624	2.7564
node2vec_v6	68,494	-0.1577	0.5016	-2.6525	2.2374
node2vec_v7	68,494	-0.0889	0.5447	-2.5508	2.2763
node2vec_v8	68,494	-0.3965	0.5485	-3.0698	2.0794
node2vec_v9	68,494	-0.3630	0.5706	-2.9005	1.8006
node2vec_v10	68,494	-0.0512	0.4918	-2.2942	2.1722
node2vec_v11	68,494	0.5742	0.6007	-1.3171	3.5712
node2vec_v12	68,494	-0.0092	0.4763	-2.3408	2.5712
node2vec_v13	68,494	0.2029	0.5031	-2.0180	2.5207
node2vec_v14	68,494	-0.3931	0.5521	-3.0263	1.5861
node2vec_v15	68,494	-0.1537	0.4891	-2.5956	2.0505
node2vec_v16	68,494	-0.0765	0.4975	-2.1987	1.9630
node2vec_v17	68,494	-0.1357	0.5166	-2.8713	1.9880
node2vec_v18	68,494	-0.1793	0.5064	-2.8765	1.9287
node2vec_v19	68,494	0.0881	0.5056	-2.2285	2.3345
node2vec_v20	68,494	0.6185	0.5969	-1.1552	3.5770

node2vec_v21	68,494	0.0687	0.5086	-2.5554	2.2730
node2vec_v22	68,494	0.4928	0.5681	-1.6413	3.1067
node2vec_v23	68,494	-0.1676	0.5138	-2.4926	1.8613
node2vec_v24	68,494	-0.4877	0.5643	-3.2105	1.4337
node2vec_v25	68,494	0.1289	0.5076	-2.2155	2.4550
node2vec_v26	68,494	-0.0147	0.4840	-2.1722	2.5573
node2vec_v27	68,494	-0.7186	0.6531	-3.5361	1.0243
node2vec_v28	68,494	0.3889	0.5478	-1.6927	2.9977
node2vec_v29	68,494	-0.1400	0.4799	-2.6732	2.0950
node2vec_v30	68,494	0.1045	0.4849	-2.1103	2.8839
node2vec_v31	68,494	0.3581	0.5523	-1.6267	3.3237
node2vec_v32	68,494	-0.2941	0.5296	-3.0292	1.7873
node2vec_v33	68,494	-0.1163	0.4901	-2.5568	2.2263
node2vec_v34	68,494	0.2791	0.6007	-1.7925	3.2219
node2vec_v35	68,494	-0.2181	0.5016	-2.9094	1.9662
node2vec_v36	68,494	-0.2057	0.5184	-3.0030	1.7998
node2vec_v37	68,494	-0.1515	0.5034	-2.5035	1.9912
node2vec_v38	68,494	-0.3024	0.5384	-3.0365	1.7670
node2vec_v39	68,494	-0.0809	0.4830	-2.5621	2.0709
node2vec_v40	68,494	0.3301	0.5387	-2.0558	2.7178
node2vec_v41	68,494	0.6393	0.6162	-1.1091	3.1009
node2vec_v42	68,494	0.1928	0.5281	-1.9540	2.9153
node2vec_v43	68,494	0.0353	0.4953	-2.1753	2.5402
node2vec_v44	68,494	-0.2394	0.5135	-3.0546	1.7454
node2vec_v45	68,494	-0.3604	0.5336	-3.1413	1.6125
node2vec_v46	68,494	0.1112	0.4967	-2.4605	2.3470
node2vec_v47	68,494	-0.0529	0.5226	-2.4696	2.3778
node2vec_v48	68,494	-0.1477	0.5338	-2.8147	2.1909
node2vec_v49	68,494	0.1359	0.4985	-1.9581	2.6690
node2vec_v50	68,494	-0.0905	0.4903	-2.6860	2.1332
node2vec_v51	68,494	0.1858	0.5065	-1.7618	2.4284
node2vec_v52	68,494	-0.1726	0.4945	-2.5596	2.1814
node2vec_v53	68,494	-0.0793	0.4818	-2.3716	2.1287
node2vec_v54	68,494	0.2098	0.5035	-1.9103	2.5807
node2vec_v55	68,494	0.2626	0.5548	-1.8284	3.2118
node2vec_v56	68,494	0.1941	0.4895	-1.6990	2.6290
node2vec_v57	68,494	-0.2002	0.4956	-2.8545	1.7148
node2vec_v58	68,494	0.1393	0.4913	-1.9534	2.7259
node2vec_v59	68,494	-0.2296	0.5255	-2.7394	1.8153
node2vec_v60	68,494	0.1244	0.5006	-2.3513	2.9668
node2vec_v61	68,494	-0.5812	0.6036	-3.5291	1.2954
node2vec_v62	68,494	-0.1927	0.5115	-2.6201	2.2157
node2vec_v63	68,494	-0.0305	0.4952	-2.4710	2.8211
node2vec_v64	68,494	0.2937	0.5465	-2.0796	2.9886