

引用ダイナミクスを用いた萌芽的論文の早期特定に関する研究

(Early Detection of Emerging Research Fronts
Using Citation Dynamics)

岩見紫乃

論文要旨

本論文は、論文の引用ネットワークに関する時系列変化の情報を用いて、早期に萌芽的論文を特定する基礎的な手法を提供するための研究である。

本論文は7章から成り、第1章に、本研究が必要である背景と研究目的を述べ、第2章に提案手法を説明し、第3章と第4章に、時系列分析から得られた結果と考察を記述し、第5章に、時系列分析を観察した上で、研究目的を達成するための設計を検証し、第6章に、提案手法を実装したシステムを紹介し、第7章に結論としてまとめている。

まず、第1章では、研究の背景、研究目的等について述べている。今日、情報量が急速に増加して、専門家でも把握しきれないほど科学技術の知識が増えるといった知識の爆発現象が起こっている。先端技術企業は萌芽的技術シーズの探索に力を入れていることが多いが、この現象の中で、探索が次第に難しくなっている。一方で、技術シーズが製品化・産業化され、市場において利益をあげ得る期間(ライフサイクル)は短くなっており、競合者に対し有利な立場を得るには、早期に有望な技術シーズを見つけることがますます重要になっている。こうしたことを背景として、先行研究では、論文の引用その他の書誌情報を利用して、萌芽的技術を探索する手法を開拓することが行われている。具体的には、例えば、論文の引用先は、論文の著者が当該被引用論文を自著と内容的な関連が深いとして選定したものであるという性質を利用して、引用ネットワークを構築し、クラスタリング手法を用いて論文をグループ化することで、学術知識や産業技術知識をボトムアップに構造化する手法が実用的なものとして完成されている。このような引用ネットワークまたは各クラスターのネットワークの中心に位置する論文は、当該分野において、多くの専門家から深い関心を持たれ、または、支持されているものと評価することができ、それら論文を精査することで、萌芽的技術シーズを特定することも可能であるとの議論がなされている。ネットワーク中心性を測定するにあたって適切な指標としては、複数のものが候補として挙げられている。論文は学術分野の成果物であるが、近年、サイエンス・リンケージから産業応用に繋げることに注目が集まっており、特許やインターネット上のデータもある中から、本研究では論文のみ注力している。また、論文のデータは、特許などのデータに比べるとデータの質が高く、手法の確立にあたってはより正確な結果が導出されることが期待される。一方、こうした先行研究に基づく手法は、十分に論文が被引用数を得た後でなければ、重要性を推測できないものであることや、参考とする被引用数が直近の数年のみである等といった制約があることから、被引用数は少ないが将来大きな分野を開拓することにつながる萌芽的論文を早い段階で見つけ出すこ

とが難しいという問題点がある。そこで、本論文では、各年の論文の引用ネットワークからネットワーク構造に基づく指標群とそれらに関する特徴量を計算し、分野の登場初期からの指標の中長期的な変化を可視化、定量化することで、萌芽的論文を早期に特定する基礎的な手法を提案する。萌芽的論文とは、複数の重要論文の定義があるうち、特に、分野を開拓した重要論文を指している。

第2章では、本論文の提案手法について述べている。本論文で着目した時系列分析は、検知すべきパターンが定まっていないものを特定するにあたって、時間の前後のデータを利用することにより、質の安定した恒常的に供給されるデータを利用することができるというメリットがある。提案手法は、次のようなステップで行うものである。まず、論文の書誌情報を論文データベースサービスである **Web of Science** から収集し、論文の引用関係から引用ネットワークを形成する。引用ネットワークのうち、最大連結成分に含まれる各論文に対して、**in-Degree Centrality** を始めとする複数のネットワーク中心性指標を計算する。指標は各年まで論文を集めたデータセットごとに計算する。1つの論文に対して各年の指標が算出されると、横軸を時間軸、縦軸を指標の値とし、各論文を1本の線として指標の値を繋げ、時系列で指標の変化を見ることを可能とする。これら時系列分析から萌芽的論文の特徴を抽出し、未知の分野でも萌芽的論文を特定できる特徴量として数値化する。

第3章では、この提案手法に対して、ノーベル賞受賞分野や **ACM チューリング賞** から7分野を選定し、受賞対象となった論文(非常に重要な萌芽的論文)に関して本提案手法を適用することで、萌芽的論文の特定の観点から、指標の有効性を具体的に検証している。ノーベル賞は、分野でもっとも重要な発見、発明、改良をした人物に贈る意図で創始されており、分野の始まりに貢献した研究者に贈られるため、本研究における萌芽的論文の特徴を抽出するのに適した素材であると考えた。いくつかのノーベル賞受賞分野では、分野が開拓された当初、その研究成果が顧みられなかった逸話が残っている分野もあり、そのような分野での萌芽的論文の特徴を抽出することで、未知の分野における萌芽的論文を特定することに繋がると考えられる。12の指標群で時系列分析を行い、比較検討を行った結果、**in-Degree Centrality**、**Communicability Centrality**、**PageRank** の3指標が萌芽的論文の特定に有効であることが明らかとなった。これらの指標においては、萌芽的論文が分野の登場初期に高い値を持つ傾向が明確に見られた。

第4章では、7分野を横断した考察を行っている。重要論文の特定に有効な指標どうしを比較すると、精度と早期特定には、**in-Degree Centrality** が最も適切だが、データセットによっては

PageRankの方が有用な場合があることが示された。また、手法を活用するに当たっての制限事項として、学術分野の登場初期において引用があまりに疎であるために引用ネットワークが形成できない場合があること、同じく初期に重要論文が2番目以降の連結成分に含まれてしまい、重要論文の特定が遅れる場合があることも明らかとなった。更に、Web of Scienceに収録されていない論文や書籍が萌芽となる重要な文献に該当した場合には、データセットの性質上、特定ができないことも示された。

以上の第3章、第4章では、萌芽的論文の早期特定に適した指標を見つけるために探索的研究であり、提案手法による時系列の可視化から有効な指標群を見出すことができた。しかしながら、可視化した時系列を観察して発行年の異なる論文を比較することは容易ではない。そこで、第5章においては、分野の成長初期において高い精度で特定可能となるように、萌芽的論文を客観的な数値で特定するための特徴量群の開発を試みている。最初に、第3章、第4章の分析結果から時系列中における各論文の指標の最大値(Height)、各論文の指標の増加率の最大値(Slope)、指標の増加率が最大値に達するまでの増加率の平均値(SlopeAve)等の候補特徴量を設定した。これら候補特徴量群について、第3章と同様の7つのデータセットを用いて実験を行った結果、in-Degree Centrality および PageRank の時系列変化のデータについて、各論文の指標の最大値(Height)と毎年の変化量のうちの最大値(Slope)が、萌芽的論文を特定する特徴量として有効であることが明らかになった。また、それら特徴量間の相関が高いことも示された。一部のデータセットでは、対象と設定した分野ではなく、不可分に関連する分野の萌芽的論文を特定するという現象が見られたが、関連領域も含めて萌芽的論文を特定する特徴量として有用であることが判明した。さらに、専門家に確認いただけた分野では、未受賞の萌芽的論文が、特徴量で挙げられる萌芽的論文候補に特定されているという評価が得られた。これらの特徴量の成果は、将来、萌芽的論文を自動的に抽出するにあたって、利用可能であると期待される。

第6章では、開発した指標、特徴量のシステム化について述べている。本システムは、論文データベースと引用ネットワークの形成については既存のシステムを利用しているが、それ以外の部分の分析に関しては新規に作成したものである。

第7章では、結論を述べている。先述したように、政策や経営の場において投資すべき技術分野を決定せねばならない意思決定者や研究テーマを模索する研究者にとって、萌芽的な技術シーズの早期特定が重要な技術経営上の課題となっている。本論文では、過去の計量書誌学による技術経営研究の成果も踏まえつつ、重要な萌芽的論文の早期特定に役立つ独自の手法を提案し、ケー

ス研究により、その有効性を確認した。本研究の新規性は、ネットワーク指標を時系列で表した時に萌芽的論文ならば初期に高い値を持つことから、学術分野の中で萌芽的論文の特定する特徴量を設定したことである。この特徴量は、異なる出版年で複数の萌芽的論文があっても、時代を横断して一元的に評価できる点で優れている。また、先行研究では、**Betweenness Centrality**で特定できる萌芽的論文に限られていたのを、それ以外の萌芽的論文にも適用できるようになった。本研究の研究成果は、先端技術を巡る市場化競争が激化する今日のイノベーション環境において、意思決定者を効果的に支援する技術経営上の基礎的な手法を提供するものであると確信する。

目次

論文要旨	2
1. 序論	23
1.1. 背景：情報爆発時代	23
1.1.1. 選択と集中が求められる研究開発	25
1.1.2. 客観的根拠に基づく意思決定の重要性	27
1.1.3. イノベーションに関する情報	27
1.1.4. 萌芽的技術シーズの発見の必要性	28
1.1.5. 萌芽領域の特定	29
1.2. 研究の目的	31
1.3. 先行研究	32
1.3.1. 萌芽領域の発見に関する先行研究	32
1.3.2. 時系列分析に関する先行研究	33
1.3.3. 科学計量学・書誌計量学に関する先行研究	34
1.4. 本研究の貢献	37
1.5. 本論文の構成	38
2. 提案手法	39
2.1. 書誌情報の収集	40
2.2. 引用ネットワークの形成とクラスター分析	40
2.3. 指標の算出方法	43

2.4.	時系列分析の可視化	44
2.5.	使用するデータセット	45
3.	引用ネットワークの時系列分析の結果	48
3.1.	暗号(2002年 ACM チューリング賞 [crypto])	48
3.1.1.	分析対象分野の概要	48
3.1.2.	各指標の時系列分析	52
3.2.	蛍光プロテイン(2008年ノーベル化学賞 [gfp])	58
3.2.1.	分析対象分野の概要	58
3.2.2.	各指標の時系列分析	62
3.3.	G タンパク質共益受容体(2012年ノーベル化学賞 [2012chem])	66
3.3.1.	分析対象分野の概要	66
3.3.2.	各指標の時系列分析	69
3.4.	iPS 細胞(2012年ノーベル生理学・医学賞 [2012med])	73
3.4.1.	分析対象分野の概要	73
3.4.2.	各指標の時系列分析	75
3.5.	準結晶(2011年ノーベル化学賞 [2011chem])	81
3.5.1.	分析対象分野の概要	81
3.5.2.	各指標の時系列分析	84
3.6.	パラジウム触媒クロスカップリング反応(2010年ノーベル化学賞 [2010chem])	91
3.6.1.	分析対象分野の概要	91
3.6.2.	各指標の時系列分析	94

3.7.	体外受精技術(2010年ノーベル生理学・医学賞 [2010med]).....	100
3.7.1.	分析対象分野の概要.....	100
3.7.2.	各指標の時系列分析.....	102
3.8.	まとめ.....	107
4.	引用ネットワークの時系列分析の考察.....	108
4.1.	指標の一般性の検証.....	108
4.1.1.	適合率と再現率.....	108
4.1.2.	平均逆順位(MRR; Mean Reciprocal Rank).....	112
4.2.	検知の時期.....	114
4.3.	社会的認知度による違い.....	116
4.4.	まとめ.....	118
5.	萌芽的論文特定のための時系列変化特徴量の考察.....	120
5.1.	特徴量の設計.....	120
5.2.	特徴量の上位論文、および、受賞論文のリスト.....	122
5.3.	特徴量の妥当性の検討.....	138
5.4.	まとめ.....	145
6.	提案手法のシステム化.....	146
6.1.	提案手法の適用例.....	149
7.	結論.....	155
	参考文献.....	159
	謝辞.....	170

研究業績	171
査読付論文(英文)	171
査読付論文(和文)	171
国際会議.....	171
国内学会.....	172
付録	173
付録 A: ネットワーク指標の計算方法	173
A.1. Degree Centrality (次数中心性) と in-Degree Centrality	173
A.2. Closeness Centrality (近接中心性)	174
A.3. Betweenness Centrality (媒介中心性).....	175
A.4. Average Neighbor Degree.....	175
A.5. Closeness Vitality.....	175
A.6. Clustering Coefficient	175
A.7. Communicability Centrality.....	176
A.8. Core Number	176
A.9. Eigenvector Centrality.....	176
A.10. Load Centrality	177
A.11. Node Redundancy	177
A.12. PageRank.....	178
付録 B: 指標の時系列変化の比較.....	179
B.1. クラスタごと引用ネットワークから計算した場合	179

B.2. データセット内の引用ネットワークから計算した場合	181
付録 C: 学術分野の選定	185
C.1. クラスタ分析によるトピックの抽出	185
C.2. 頻出キーワードによるトピックの抽出	186

目次

Figure 1 : Web of Science 上の各年の論文数 (1950-2012)	24
Figure 2 : Business enterprise R-D expenditure (OECD, 2013)	26
Figure 3 : 発見・発明からイノベーションへのフェーズ (Auerswald, 2003)	28
Figure 4 : 重要論文と萌芽的論文の範囲.....	30
Figure 5 : 計量学問の関係 (Björneborn et al., 2004)	35
Figure 6 : 学問の萌芽領域・重要論文を検出する研究の分類	36
Figure 7 : 本研究の要素	37
Figure 8 : 指標の時系列から各年の引用ネットワークを形成する概念図	39
Figure 9 : Flow of Cluster Analysis	41
Figure 10 : 引用関係 (直接引用、共引用、書誌結合).....	42
Figure 11 : 引用ネットワークの連結成分.....	42
Figure 12 : How to read figures of time transition	45
Figure 13 : ライフサイクルと E/N 比における、各データセットのポジション	47
Figure 14 : ライフサイクルと上位論文の特徴における、各データセットのポジション	47
Figure 15 : 出版年ごとの論文数の分布 [crypto].....	50
Figure 16 : Academic Landscape Map of Cryptology [crypto].....	51
Figure 17 : Heatmap of Cosine Similarity between Top 7 Clusters [crypto]	51
Figure 18 : キーワードに基づくクラスター間の類似度 [crypto]	53
Figure 19 : 論文の遷移量に基づくクラスター間の類似度 [crypto].....	53

Figure 20 : Time Transition about Number of within-Dataset Citation of 419 Papers having more than 100 citations (最大値=1288) [crypto]	54
Figure 21 : Time Transition about in-Degree Centralities of 419 Papers having more than 100 citations (最大値=0.4136) [crypto]	54
Figure 22 : Time Transition about PageRank of 419 Papers having more than 100 citations (最大値=0.1691) [crypto]	54
Figure 23 : Time Transition about in-Degree Centralities of 1978 Papers having 1 citation (最大値=0.09091) [crypto]	55
Figure 24 : Time Transition about PageRank of 1978 Papers having 1 citation (最大値=0.08200) [crypto]	55
Figure 25 : Time Transition about Communicability Centralities of 419 Papers having more than 100 citations (最大値= 1.160e+29) [crypto]	56
Figure 26 : Time Transition about Communicability Centralities of Papers having more than 100 citations (最大値= 175) [crypto]	56
Figure 27 : Correlation Coefficient between Current Number of Citations and Indicators [crypto]	57
Figure 28 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [crypto]	57
Figure 29 : 出版年ごとの論文数の分布 [gfp]	60
Figure 30 : Academic Landscape Map of Fluorescent Protein [gfp]	61
Figure 31 : Heatmap of Cosine Similarity between Top 7 Clusters [gfp]	61
Figure 32 : キーワードに基づくクラスター間の類似度 [gfp]	63
Figure 33 : 論文の遷移量に基づくクラスター間の類似度 [gfp]	63

Figure 34 : Time Transition about Number of within-Dataset Citation of 1,967 Papers having more than 100 citations (最大値=2371) [gfp]	64
Figure 35 : Time Transition about in-Degree Centrality of 1,967 Papers having more than 100 citations (最大値=0.7333) [gfp]	64
Figure 36 : Time Transition about PageRank of 1,967 Papers having more than 100 citations (最大値=0.2276) [gfp]	64
Figure 37 : Correlation Coefficient between Current Number of Citations and Indicators [gfp]	65
Figure 38 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [gfp]	65
Figure 39 : 出版年ごとの論文数の分布 [2012chem]	67
Figure 40 : Academic Landscape Map of G-protein-coupled receptors [2012chem]	68
Figure 41 : Heatmap of Cosine Similarity between Top 7 Clusters [2012chem]	68
Figure 42 : キーワードに基づくクラスター間の類似度 [2012chem]	70
Figure 43 : 論文の遷移量に基づくクラスター間の類似度 [2012chem]	70
Figure 44 : Time Transition about Number of within-Dataset Citation of 3,557 Papers having more than 100 citations (最大値=1792) [2012chem]	71
Figure 45 : Time Transition about in-Degree Centrality of 3,557 Papers having more than 100 citations (最大値=1) [2012chem]	71
Figure 46 : Time Transition about PageRank of 3,557 Papers having more than 100 citations (最大値=1) [2012chem]	71
Figure 47 : Correlation Coefficient between Current Number of Citations and Indicators [2012chem]	72

Figure 48 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [2012chem]	72
Figure 49 : 出版年ごとの論文数の分布 [2012med]	74
Figure 50 : Academic Landscape Map of iPS cells [2012med]	75
Figure 51 : Heatmap of Cosine Similarity between Top 7 Clusters [2012med]	75
Figure 52 : キーワードに基づくクラスター間の類似度 [2012med]	77
Figure 53 : 論文の遷移量に基づくクラスター間の類似度 [2012med]	77
Figure 54 : Time Transition about Number of within-Dataset Citation of 1,064 Papers having more than 100 citations (最大値=2649, 最小値=0) [2012med]	78
Figure 55 : Time Transition about in-Degree Centralities of 1,064 Papers having more than 100 citations (最大値=0.25) [2012med]	78
Figure 56 : Time Transition about PageRank of 1,064 Papers having more than 100 citations (最大値=0.3412) [2012med]	78
Figure 57 : Time Transition about Communicability Centralities of 1,064 Papers having more than 100 citations [2012med]	79
Figure 58 : Correlation Coefficient between Current Number of Citations and Indicators [2012med]	80
Figure 59 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [2012med]	80
Figure 60 : 出版年ごとの論文数の分布 [2011chem]	82
Figure 61 : Academic Landscape Map of Quasicrystal [2011chem]	83
Figure 62 : Heatmap of Cosine Similarity between Top 7 Clusters [2011chem]	83
Figure 63 : キーワードに基づくクラスター間の類似度 [2011chem]	85

Figure 64 : 論文の遷移量に基づくクラスター間の類似度 [2011chem]	85
Figure 65 : Time Transition about Number of within-Dataset Citation of 155 Papers having more than 100 citations (最大値=1745) [2011chem]	86
Figure 66 : Time Transition about in-Degree Centralities of 155 Papers having more than 100 citations (最大値=1) [2011chem]	86
Figure 67 : Time Transition about PageRank of 155 Papers having more than 100 citations (最大値=0.3383) [2011chem]	87
Figure 68 : Time Transition about in-Degree Centralities of 689 Papers having 1 citation (最大値=0.004345) [2011chem]	88
Figure 69 : Time Transition about PageRank of 689 Papers having 1 citation (最大値=0.002622) [2011chem]	88
Figure 70 : Time Transition about Communicability Centralities of 155 Papers having more than 100 citations (最大値= 1.529e+23) [2011chem]	88
Figure 71 : Time Transition about Communicability Centralities of 312 Papers published until 1990 (最大値= 2.131e+7) [2011chem]	89
Figure 72 : Time Transition about Communicability Centralities of 312 Papers published until 1990 (最大値=2754) [2011chem]	89
Figure 73 : Correlation Coefficient between Current Number of Citations and Indicators [2011chem]	90
Figure 74 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [2011chem]	90
Figure 75 : 出版年ごとの論文数の分布 [2010chem]	93
Figure 76 : Academic Landscape Map of palladium-catalyzed cross couplings in organic synthesis [2010chem]	94

Figure 77 : Heatmap of Cosine Similarity between Top 7 Clusters [2010chem]	94
Figure 78 : キーワードに基づくクラスター間の類似度 [2010chem].....	96
Figure 79 : 論文の遷移量に基づくクラスター間の類似度 [2010chem]	97
Figure 80 : Time Transition about Number of within-Dataset Citation of 580 Papers having more than 200 citations (最大値=1928) [2010chem]	97
Figure 81 : Time Transition about in-Degree Centralities of 580 Papers having more than 200 citations (最大値=0.05898) [2010chem]	98
Figure 82 : Time Transition about PageRank of 580 Papers having more than 200 citations (最大値=0.02162) [2010chem].....	98
Figure 83 : Correlation Coefficient between Current Number of Citations and Indicators [2010chem]	98
Figure 84 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [2010chem]	99
Figure 85 : 出版年ごとの論文数の分布 [2010med].....	101
Figure 86 : Academic Landscape Map of in vitro fertilization [2010med]	102
Figure 87 : Heatmap of Cosine Similarity between Top 7 Clusters [2010med].....	102
Figure 88 : キーワードに基づくクラスター間の類似度 [2010med]	104
Figure 89 : 論文の遷移量に基づくクラスター間の類似度 [2010med].....	104
Figure 90 : Time Transition about Number of within-Dataset Citation of 2,546 Papers having more than 100 citations (最大値=488) [2010med]	105
Figure 91 : Time Transition about in-Degree Centralities of 2,546 Papers having more than 100 citations (最大値=0.5714) [2010med]	105

Figure 92 : Time Transition about PageRank of 2,546 Papers having more than 100 citations (最大値=0.1720) [2010med].....	105
Figure 93 : Correlation Coefficient between Current Number of Citations and Indicators [2010med]	106
Figure 94 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [2010med]	106
Figure 95 : 適合率と再現率の算出方法	109
Figure 96 : MRR の概念	113
Figure 97 : 社会的認知度の違いによる、in-Degree Centrality および PageRank の形状の違い	116
Figure 98 : in-Degree Centrality や PageRank の一般的な形状	121
Figure 99 : Height, Area, Slope の定義	121
Figure 100 : PageRank が in-Degree Centrality より萌芽的論文を特定するのに有効な理由	141
Figure 101 : インパルス応答の模式図	141
Figure 102 : 本研究における手法において、ツールとしての貢献の範囲	146
Figure 103 : 時系列分析における Web インターフェース	148
Figure 104 : 出版年ごとの論文数の分布	150
Figure 105 : Time Transition about in-Degree Centralities of 2,664 Papers having more than 100 citations (最大値=0.75) [fMRI].....	151
Figure 106 : Time Transition about PageRank of 2,664 Papers having more than 100 citations (最大値=0.4683) [fMRI].....	151
Figure 107 : 本研究と選定された書誌情報との関係	158

Figure 108 : 次数中心性 (Degree Centrality) の定義と、ネットワークの成長に伴う変化..	174
Figure 109 : in-Degree Centrality の定義と、ネットワークの成長に伴う変化	174
Figure 110 : 伝達性中心性 (Communicability Centrality) の、ネットワークの成長に伴う変化	176
Figure 111 : load (左) と betweenness (右) の、始点ノードと終点ノード間での重みの違い	177
Figure 112 : Node Redundancy の例	178
Figure 113 : PageRank の概念	178
Figure 114 : Time Transition about Cluster Number of 155 Papers having more than 100 citations (最大値=33)	179
Figure 115 : Time Transition about Number of within-Cluster Citation of 155 Papers having more than 100 citations (最大値=803)	179
Figure 116 : Time Transition about within-Cluster Degree Centralities of 155 Papers having more than 100 citations (最大値=1)	180
Figure 117 : Time Transition about within-Cluster in-Degree Centralities of 155 Papers having more than 100 citations (最大値=1)	180
Figure 118 : Time Transition about within-Cluster out-Degree Centralities of 155 Papers having more than 100 citations (最大値=1)	180
Figure 119 : Time Transition about within-Cluster Closeness Centralities of 155 Papers having more than 100 citations (最大値=1)	180
Figure 120 : Time Transition about within-Cluster Betweenness Centralities of 155 Papers having more than 100 citations (最大値=0.05556)	180
Figure 121 : Time Transition about Number of within-Dataset Citation of 155 Papers having more than 100 citations (最大値=1745)	181

Figure 122 : Time Transition about within-Dataset Degree Centralities of 155 Papers having more than 100 citations (最大値=1)	181
Figure 123 : Time Transition about within-Dataset in-Degree Centralities of 155 Papers having more than 100 citations (最大値=1)	181
Figure 124 : Time Transition about within-Dataset out-Degree Centralities of 155 Papers having more than 100 citations (最大値=0.2917)	181
Figure 125 : Time Transition about within-Dataset Closeness Centralities of 155 Papers having more than 100 citations (最大値=0.2963)	182
Figure 126 : Time Transition about within-Dataset Betweenness Centralities of 155 Papers having more than 100 citations (最大値=0.008890)	182
Figure 127 : Time Transition about within-Dataset Average Neighbor Degree of 155 Papers having more than 100 citations (最大値=24)	182
Figure 128 : Time Transition about within-Dataset Closeness Vitality of 155 Papers having more than 100 citations (最大値= 2.166e+5, 2001 年以降は最小値がマイナス)	182
Figure 129 : Time Transition about within-Dataset Clustering Coefficient of 155 Papers having more than 100 citations (最大値=1)	182
Figure 130 : Time Transition about within-Dataset Communicability Centralities of 155 Papers having more than 100 citations (最大値= 1.529e+23)	183
Figure 131 : Time Transition about within-Dataset Core Number of 155 Papers having more than 100 citations (最大値= 17)	183
Figure 132 : Time Transition about within-Dataset Eigenvector of 155 Papers having more than 100 citations (最大値= 1)	183
Figure 133 : Time Transition about within-Dataset Load Centralities of 155 Papers having more than 100 citations (最大値= 0.008890)	183

Figure 134 : Time Transition about within-Dataset Node Redundancy of 155 Papers having more than 100 citations (最大値= 1)	183
---	-----

Figure 135 : Time Transition about within-Dataset PageRank of 155 Papers having more than 100 citations (最大値= 0.3383).....	184
--	-----

表目次

Table 1 : 科学計量学・書誌計量学関連分野の普及状況	35
Table 2 : 各指標の説明.....	43
Table 3 : データセットの特徴.....	46
Table 4 : 最新被引用数ごとの論文数の分布 [crypto]	50
Table 5 : Information of TOP 7 Clusters [crypto]	51
Table 6 : TOP 10 Papers on 2012-11-16 [crypto]	52
Table 7 : 最新被引用数ごとの論文数の分布 [gfp].....	60
Table 8 : Information of TOP 7 Clusters [gfp].....	61
Table 9 : TOP 10 and the Nobel Laureate's Papers on 2012-11-16 [gfp]	62
Table 10 : 最新被引用数ごとの論文数の分布 [2012chem]	67
Table 11 : Information of TOP 7 Clusters [2012chem]	68
Table 12 : TOP 10 and the Nobel Laureate's Papers on 2013-03-27 [2012chem].....	69
Table 13 : 最新被引用数ごとの論文数の分布 [2012med].....	74
Table 14 : Information of TOP 7 Clusters [2012med].....	75
Table 15 : TOP 10 and the Nobel Laureate's Papers on 2013-03-27 [2012med].....	76

Table 16 : 最新被引用数ごとの論文数の分布 [2011chem]	82
Table 17 : Information of TOP 7 Clusters [2011chem]	83
Table 18 : TOP 10 Papers on 2013-03-29 [2011chem]	84
Table 19 : 最新被引用数ごとの論文数の分布 [2010chem]	93
Table 20 : Information of TOP 7 Clusters [2010chem]	94
Table 21 : TOP 10 and the Nobel Laureate's Papers on 2013-05-08 [2010chem]	96
Table 22 : 最新被引用数ごとの論文数の分布 [2010med]	101
Table 23 : Information of TOP 7 Clusters [2010med]	102
Table 24 : TOP 10 and the Nobel Laureate's Papers on 2013-03-29 [2010med]	103
Table 25 : 指標の適合率一覧	110
Table 26 : 指標の再現率一覧	111
Table 27 : MRR が持つ値(理論値)	113
Table 28 : TOP 10 論文の平均逆順位	114
Table 29 : 萌芽的論文の検知できる時期	115
Table 30 : 指標が最大値に達するまでの期間	117
Table 31 : 各リストにおける上位 10 位論文、および、受賞論文	122
Table 32 : 各データセットにおける、特徴量の適合度	139
Table 33 : 蛍光プロテイン(gfp)の特徴量間の相関係数	143
Table 34 : 体外受精技術(2010med)の特徴量間の相関係数	143
Table 35 : iPS 細胞(2010med)の特徴量間の相関係数	144
Table 36 : 暗号(crypto)の特徴量間の相関係数	144

Table 37: 準結晶(2011chem)の特徴量間の相関係数	145
Table 38: 代表的な脳機能の計測装置	150
Table 39: 最新被引用数ごとの論文数の分布	150
Table 40: fMRI の時系列分析の図より特徴的な論文	152
Table 41: fMRI の各特徴量の上位 10 論文	153
Table 42: Clusters of TOP 0.1% Papers	186
Table 43: キーワードの出現頻度	188
Table 44: キーワードの共起頻度	188
Table 45: 共起するキーワードの Dice 係数	189

1. 序論

1.1. 背景：情報爆発時代

古くは、科学技術を含めた複数の分野で歴史的な業績を残す者がいた。15～16世紀のイタリア、ルネサンス期を代表する芸術家であるレオナルド・ダ・ビンチ(Leonardo da Vinci)は、絵画、建築、科学、技術ほか複数の分野にわたって後世に残る業績を残した(森, 1976)。17～18世紀のドイツでは、ゴットフリート・ヴィルヘルム・ライプニッツ(Gottfried Wilhelm Leibniz)も、2進法や微積分学と数学のほか、哲学の分野でも思想家として有名である(Voulker, 1983; 阿部, 2006)。また、学問分野の分化が進んだ20世紀においても、ハーバート・アレクサンダー・サイモン(Herbert Alexander Simon)は、政治学、経済学、認知心理学、工学と多岐に渡り才能を発揮し、1975年に人工知能への貢献から ACM チューリング賞を、1978年にはノーベル経済学賞を受賞している(Augier, 2002; サイモン, 1999)。彼らのように、一人で多くの分野の知識を掌握した博学者が存在する一方で、社会的には、社会に流通する情報量は急激に増加し、専門家が専門家であるために知識の分化が進んだ。農業革命と産業革命に続く「第三の波」(Toffler, 1980)として、先進国を中心に人類は情報革命を迎え、情報通信技術の発達に伴い情報量は急激に増加したのである。

情報通信技術の発達にあたっては、1946年に、現在のコンピュータの基礎的概念を実装した von Neumann 型の ENIAC (Electronic Numerical Integrator and Computer) が完成した。ENIAC は、当時、利用できる素子として信頼性の低い真空管を用いても、複雑な計算を行うことができるハードウェア構成を持っており、ソフトウェアの部分においてはプログラムをいつでも変更できる汎用性を備えた(Grier, 1996; Winegrad, 1996) ことで、コンピュータが多くの用途で利用できる土台ができあがったのである。さらに、1970年代には、アメリカ合衆国の ARPA (現在の DARPA、すなわち、Defense Advanced Research Projects Agency) において、ネットワークの一部が軍事攻撃を受けて機能しなくなっても、残りのネットワーク自律的に協調してネットワークとしての機能を維持する、拡張性と冗長性の高いインターネットが発明され、1993年には、情報を簡便に伝えるプロトコルである World Wide Web が発明されて Mosaic ブラウザが登場(Mowery, 2002) してからは、情報が増殖し拡散する技術的な要素が整った。日本においては、2001年に、総務省が「ブロードバンド元年」と位置付けて以降、帯域の広い回線を定額で利用できるサービスが普及し、経済的な障壁もはるかに軽減されることとなったのである。このよう

な情報通信技術をふんだんに利用することが可能となった環境下において、2001年の 3.83×10^{21} ビットに対し、10年後の2011年には 7.61×10^{21} ビットと流通情報量は約2倍となった(総務省, 2011)。

Figure 1 を見ていただきたい。これは、Thomson Reuters 社が提供する論文データベースサービス Web of Science 上で各年に発表された学術論文の数の経年変化であるが、学術領域においても情報爆発時代にあることが伺える。世界人口が研究者の数をそのまま表すものでもないが、世界人口が1950年に約25億人であったのが2012年には70億を超え3倍弱であること(United Nations, 2013) を考えると、2012年の1年あたりに発表される論文数は1950年の18倍強になっており、人類の知識のアウトプットは一人当たりの量としても増加している。さらに、知識の総量は過去からの蓄積であるので、インプットとして取り扱う知識は、アウトプットの増加量より加速している状況にある。

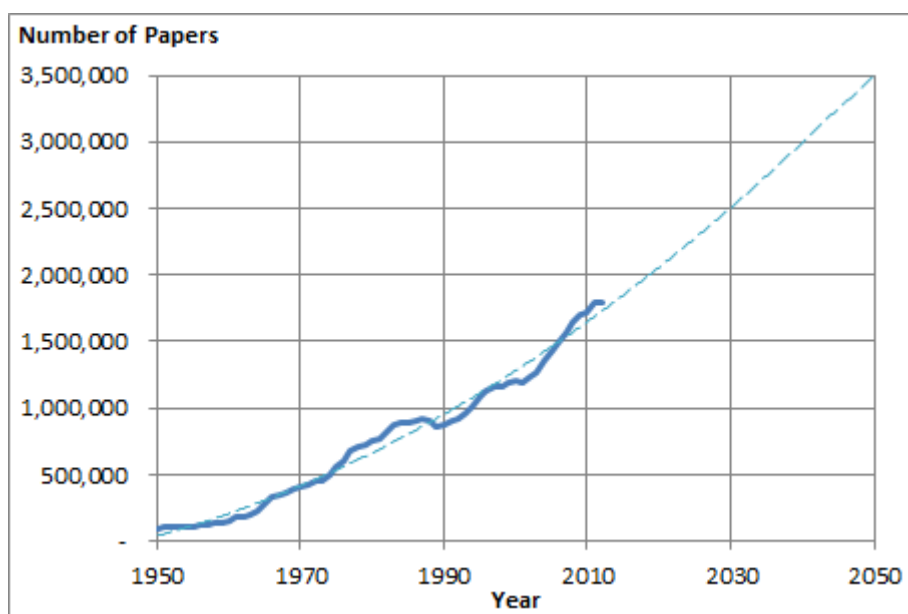


Figure 1 : Web of Science 上の各年の論文数 (1950-2012)

学術分野における情報量の増加とその把握の困難を証明するものとして、学術論文のピアレビューに関し、その信頼性を確認した研究がある。著名なジャーナルに掲載された12本の論文を、18-32か月後に著者とタイトルを変更してから再度、投稿して採択されるかを試したところ、38

名による査読を受けて3名のみが再投稿に気付いたというものである(Douglas et al., 1982)。このように、その分野の専門家でも専門分野を全て把握することは不可能である。

科学・産業技術に関する情報量の爆発的増加と同時に、知識の細分化や構造変化も起こっており、科学・産業技術における情報活用はハードルが高まっている。従来、政策の場に学術的な知見を導入するような政府の委員会活動においては、代表的な専門家を招いてワークショップ方式(T-Plan 法やデルファイ法)で検討を行うことがあるが、増加した情報量が細分化し構造変化を起こした最新の知見を全て取り入れることは難しく、また、ワークショップ方式では時間やコストもかかるため、アウトプットまでに陳腐化する虞もある。ワークショップ方式自体は、専門家どうしの対面のインターアクションによりその場で新たな知が構成されるという点で意義はあるもの、偏りなく全てを俯瞰することが難しい。これらの状況の変化に伴い、知識の未利用が発生しうる環境となっており、結果として、国際競争力の低下、科学・産業技術に対する投資効果の低下、卓越した科学技術と市場との溝、社会課題と優れた技術との谷間を招く虞がある。望むらくは、ワークショップ方式と、全てを俯瞰する方式を併用したいところである。

そこで、全てを俯瞰する方式を実現し、科学・産業技術に係る意思決定者が効率的に情報を取捨選択するサポートとして、情報・ウェブ工学を利用した「知の埋没」(国家戦略会議 フロンティア分科会, 2012)を乗り越える手段の開発に今が重要なタイミングであると考え。論文データベースの機能を見ると、一般的には、対象分野の被引用数の多い方から優先順位をつけたり、Impact Factor の大きなジャーナルを優先するといった方法で読むべきに学术论文を絞り込むことが行われている。

1.1.1.1. 選択と集中が求められる研究開発

一方で、2007年以降の世界金融危機により、研究開発費を削減する企業も現れた。Figure 2 (OECD, 2013)は、企業の研究開発費について OECD の統計から 2012 年の論文数上位 10 国に関して時系列で表したものであるが、アメリカ、日本、カナダに関しては、企業の研究開発費の減少が明確に見られる。日本の財政状況に限ると、一般会計歳出が一般会計税収より多状況で差が徐々に拡大しており、2009 年以降は歳出が税収の 2 倍以上となっている(財務省, 2012)。これらのように、景気の影響により研究開発費が削減されたり、財政状況により研究開発に潤沢に資金が提供される状況ではなくなっている。研究開発に投資すべき人・物・金といったリソースが効

率的に運用されるために、また、研究開発が社会に受け入れられ説明が可能となるためには、合理的な根拠に基づく判断が求められる。

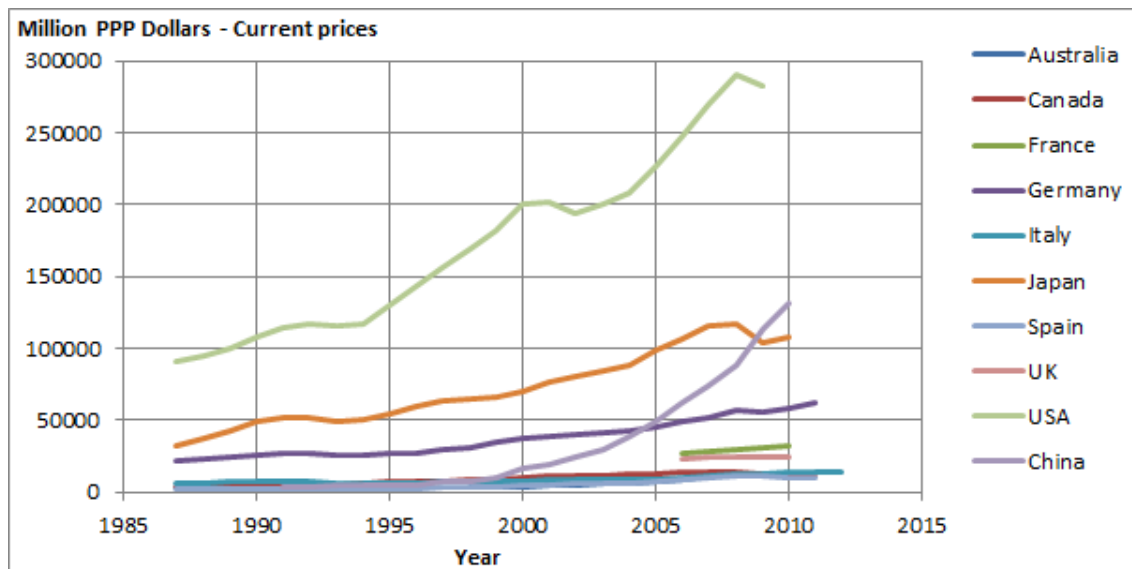


Figure 2 : Business enterprise R-D expenditure (OECD, 2013)

もちろん、研究開発に投資して即座に企業の業績向上に繋がり、景況が回復するというだけでもない。Knott(2012)は、Booz & Company 社が毎年、発表している「The Global Innovation 1000」から、「研究開発への投資額は株式時価総額や成長性と相関はなく、業績向上に最も大きな要因では、研究開発投資と戦略との整合性、および、イノベーションを支える企業文化である」ということを受けて、RQ(research quotient)という指標を、自社の研究開発への投資効果を推定して他社と比較するだけでなく、研究開発支出の増減が損益に与える影響および自社の株式時価総額への影響を把握できるものとして提案した。この RQ は、資本と労働力と研究開発投資から算出されるものである。このように科学技術が事業として実を結ぶには、萌芽領域の発見のみではなく、資金面や労働力も考慮する必要があるが、本研究は、萌芽領域に注目して分析を行うことに注力する。

1.1.2. 客観的根拠に基づく意思決定の重要性

研究開発に投資するリソースの効率的な活用が重要になっていることに加えて、日本においては2011年3月11日以降の東日本大震災に伴う福島第一原子力発電所事故において、科学的根拠が政策に十分に活かされていなかったこと、ならびに、非常時において政府が所有する観測データの情報開示が不十分であったことが指摘されており、国民および世界に不信感を抱かせる結果となった(Baba, 2013; Figueroa, 2013)。このような事態を引き起こした反省から、政策形成の場において、客観的根拠(エビデンス)に基づく合理的で透明性の高いプロセスが求められており、第4期科学技術基本計画(2011-2015年度)に基づき、日本政府は「科学技術イノベーション政策のための科学」(文部科学省, 2013)を推進している。

一方、アメリカ合衆国では、2006年段階で、客観的根拠に基づく(evidence-based)政策は表面的なパスワードに留まっており、MBA取得者であっても学んだことを実践に活かしていないことが指摘されている。客観的根拠に基づくための研究・教育と実践との差異を埋めるために、(1)客観的根拠に基づく実践モデルを提供すること、(2)客観的根拠の利用を促進すること、(3)実践者と研究者・教育者との協力を構築すること、(4)当該分野を学ぶ学生の期待を刺激することが挙げられる(Rousseau, 2006)が、本研究では、4点中の前者2点に関係して、客観的根拠を提供すること、および、その客観的根拠を導出する手法を提供するものとなる。

1.1.3. イノベーションに関する情報

企業が長期にわたって事業を維持するために、イノベーションを行うことが望ましい。研究開発により導き出された科学技術はイノベーションのシーズではあるが、科学技術がすなわちイノベーションということではなく、イノベーションは科学技術のみで構成されるものでもない。一時期、「イノベーション」の日本語訳に「技術革新」が当てられていたこともあるが、現在では、技術革新や経営革新ではないと認識されており、「イノベーション25」(内閣府, 2008)といった政策も取りまとめられている。政府や企業が事業を行う際に製品の機能のみに着目すると、顧客のニーズを見落とすことが指摘されている(Levitt, 1960)が、これは科学技術のみにイノベーションを求めるとこのようなケースも発生するであろう。イノベーションとは、顧客にとっての価値を新規創造することであり、value innovation や business model innovation と呼ばれるよ

うに、プロセスやサービスによっても提供できるものである(Johnson et al., 2008; Kim et al., 2004)。研究開発における発見・発明からイノベーションが発生するまでのフェーズにおいて、プロセスやサービスのイノベーションは、研究開発が製品化・事業化する Figure 3 の流れの下流に当たる。



Figure 3: 発見・発明からイノベーションへのフェーズ (Auerswald, 2003)

しかしながら、本研究では、Figure 3 の研究段階にあたる論文に注目して分析を行う。論文については、情報が整理された良質のデータベースが存在している。他のフェーズにおいても利用できるデータベースとして、開発フェーズにおいては特許データが存在するが、下流ほど世界的に整備された良質なデータを得ることは難しくなる。データから有用なエビデンスを引き出そうという時に品質の悪いデータであると、実態とずれを生じる可能性が高くなり、分析結果が重大な瑕疵を引き起こす虞がある。例えば、2007年に始まった米サブプライム問題では、データの不正確性や欠如が根源となり国際的に重大な問題を引き起こした事例であり、住宅ローンの申込時に不正確な情報が記入されなかったために審査の正確性を欠く結果となって、信用収縮時に資産は回収不能となった(レドマン, 2010)。データの取得・維持・加工・破棄のプロセスを考えた場合、情報爆発時代のデータの取得・維持・破棄にも重要な課題が潜むが、それらは本研究とは別に十分な検討をされるべきであり、本研究では、意思決定に寄与するためにデータの加工に注力し、既存の良質な論文データベースを活用するものである。

1.1.4. 萌芽的技術シーズの発見の必要性

企業が新商品・新規事業を立ち上げるとき、顧客ニーズを探る市場調査を行うことが多い。企業の経営者が、企業の投資判断を行うに先立って、自分たちの顧客がそれを求めているのか、市場はどのくらい大きいのか、投資は利益を生むのかを考える。しかしながら、顧客が自らの要求しているものを全て把握しているとは限らない。事前調査で顧客ニーズに合致しない新商品・新規事業が、新しい市場を開拓し顧客に受け入れられることがある。

例えば、ゼロックスのコア顧客であるコピー・サービス会社は、低スピードの小型卓上コピー機を必要としていなかったが、キヤノンはオフィス内で印刷できる小型卓上コピー機を提供してコピーのために外出する手間を減らした。ソニーのウォークマンは従来、室内のステレオで聞いていた音楽を、音響機器の小型化に成功して室外でも音楽を聴けるようにし利用者の生活様式を変えた。キヤノンやソニーの事例は、製品を企画するにあたって顧客ニーズの高まりがあって開発されたものではなく、メーカーの技術シーズが先にあって実現したものである。その他の事例では、ハード・ディスク・ドライブ(HDD)の小型化は、コンピュータの主流がメインフレームであった登場当初は小型化の代償に記憶容量が小さく、大手コンピュータメーカーに興味を持たれなかったが、小型化技術の進展につれ記憶容量が増加し、また、電源の内部化、ステッパー・モーターの低コスト化、耐久性の向上、軽量化、省電力化といった効果も伴って、1970年代後半から1980年代半ばのミニコン、デスクトップPC、ポータブルPCといったコンピュータの新しい形状の市場開拓に寄与したのである(Christensen, 2001)。

また、先のHDDの例では、ライフサイクルが2年といわれており(Bower, 1995)、投資を回収するためには先行者利益を確保するために、研究開発を行った技術は早く製品に実現する必要がある。一方で、例えば、基礎研究から上市するまでに数十年単位の時間がかかり医薬品のような長いライフサイクルが必要となる分野もある。研究、開発、製品化、事業化の間にはそれぞれ、「死の谷」(あるいは、Figure 3のように、3つの境界を別々に呼称して、それぞれを「魔の川」「死の谷」「ダーウィンの海」と呼ばれる障壁があると言われており、前段階が早く達成したからといって後段階が順調に進捗するとは限らない。しかしながら、早く技術シーズを把握することで、競合に先んじる可能性が高まることは間違いない。いずれにしても、早いことが技術シーズの権利を得る大きなファクターである。

これらの社会的背景から、「技術シーズ」を発見すること、それが「萌芽的」なものであることは、研究者、および、政策・経営の意思決定者には、競合により有利な立場に立つために重要である。実際に、企業からは「萌芽的技術シーズを早く特定したい」という要望が聞かれる。

1.1.5. 萌芽領域の特定

重要論文とは、その分野の源流となる論文や、源流ではないが被引用数の多い論文など、複数の考え方があり。後者を利用して、従来の指標や、ノーベル賞有力候補者と挙げているトムソン・

ロイター引用栄誉賞、そして、分野ごとの被引用数上位 1%論文を採用したサイエンスマップなどは重要な論文を決定しているが、分野を開拓した論文が被引用数の多い論文とは限らない。例えば、2012 年のノーベル生理学・医学賞において、John Bertrand Gurdon 氏と山中伸弥氏が受賞しているが、山中氏の論文(2006)は被引用数が 2012 年末時点で約 5,000 件と後者による重要論文にも当てはまるが、山中氏にも源流の論文と認められている Gurdon 氏の論文(1962)の被引用数は 300 件程度であり、被引用数で判断すると分野の源流となる論文が重要論文から除外されかねない。

そこで、本研究では、分野を開拓した源流となる論文を萌芽的論文と定義し、これを特定の対象とする。萌芽的論文は、重要論文の一形態である。一方、萌芽領域とは、将来の成長の大きな源泉となる萌芽的技術シーズ群を含み、研究者や社会の関心を集め急速に立ち上がりつつある技術領域と定義される。萌芽的技術シーズを記述したものの一つが萌芽的論文となる。本研究で主に扱うノーベル賞受賞論文は、そのコンセプトから、受賞の動機となった論文の中で一番早い時期に発表された論文に対して、萌芽的論文に相当すると考えることができる。

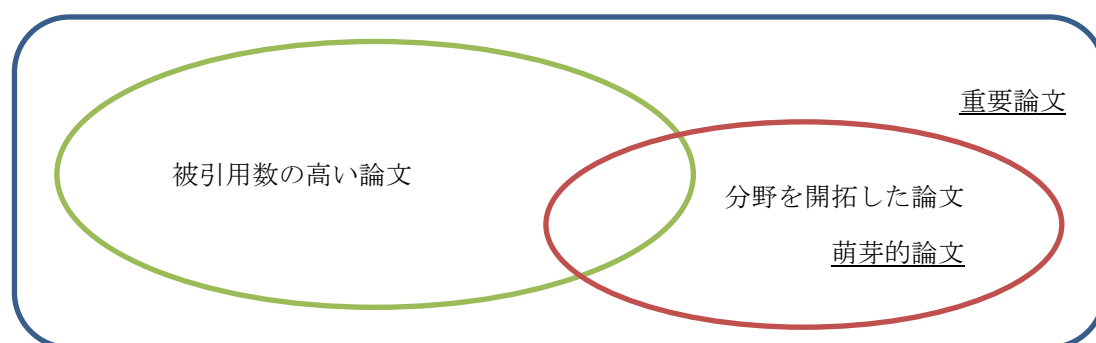


Figure 4：重要論文と萌芽的論文の範囲

ノーベル賞の 3 分野(物理学、化学、医学生理学)は、現在、評価された自然科学の証明として世界で認められている最高峰であることは間違いない。ノーベル賞は、スウェーデンのアルフレッド・ベルンハルド・ノーベル(Alfred Bernhard Nobel)の遺言により、物理学、化学、医学生理学、文学、平和の 5 分野について 1901 年から授与が始まった。経済学についても、ノーベルの名を冠した重要な賞が存在するが、ノーベルの遺言にあった賞ではなく、1968 年、スウェーデン国立銀行で創設された賞である。ノーベル賞の自然科学 3 分野について、ノーベルの遺言では、(a)物理学の分野でもっとも重要な発見か発明をした人物、(b)化学の分野でもっとも重要

な発見や改良をした人物、(c)生理学または医学の領域でもっとも重要な発見をした人物に与えられるように規定されていた。また、ノーベルの遺言では「前年に」功績を上げた人物に対して賞を与えるとあったが、前年の発見では十分に評価が定まっておらず、実際は、5年から20年前、時にはもっと前の発見に対して賞が授与されている。物理学賞、および、化学賞はスウェーデン王立科学アカデミー、生理学・医学賞はカロリンスカ研究所が各ノーベル賞の選考機関であり、選考機関から指名された人物のみ推薦する権利がある。推薦をおこなう権利があるのは、スカンディナヴィア諸国の各分野の教授や、過去の受賞者などであり、過去数十年の間、自然科学3分野について委員会に寄せられる推薦状の数は毎年200~500通であった(ノルビ, 2011)。このような賞創設の意図を鑑みるに、自然科学3分野のノーベル賞受賞論文は、萌芽的論文に相応しいと考える。

ここで、萌芽的論文の論文に絞ることで、一見、産業応用からは距離があるようにも考えられる。しかし、近年では、企業のヒアリングより、サイエンス・リンケージを産業応用に結びようという動きがあり、論文という学術分野に焦点を当てることも、十分に、科学技術の産業応用への貢献となる。技術のライフサイクルが短くなっている現在、より早く開発・製品化・事業化に繋げるには、その前段階の研究フェーズで技術シーズを見つけることは早さを稼ぐことにも繋がる。また、論文以外に、特許やインターネット上のデータを利用することも考えられるが、データの質を鑑みたときに、特許等のデータと比べると論文データの質は高く、分析結果の正確な検証が可能であると期待できる。

1.2. 研究の目的

本研究の目的は、科学技術に関する情報量の爆発的増加という背景から、大量の情報の中から萌芽的技術シーズを早期に特定し、科学技術に係る意思決定をサポートする方法論を提供することにある。具体的には、膨大な論文情報を利用し、情報工学的な技法を導入して、特徴量の時系列変化に着目し、萌芽的論文を早期に特定する。ここでは、萌芽的論文とは、分野の源流にあたる論文を指す。論文情報の中でも、特に引用関係を利用する。論文の引用は、論文の著者が自著に関連があるとして選定したものであることから、著者がどのレベルで引用に用いたかの差はあるものの、引用ネットワークは専門家の知識をアド・ホックに集大成して、知のネットワークを構成しているものである。このような引用ネットワークの中心にある論文は、多くの専門家から支

持されていると期待できる。一方、引用するという行為は直近数年に重点が置かれることが多く、かつ、年々論文数が増えているがために、数十年前に重要論文があっても把握することが難しい。本研究では、政策や経営の場において投資すべき技術分野を決定せねばならない意思決定者や研究テーマを模索する研究者を支援するために、論文の引用ネットワークからネットワーク構造に基づく指標を年ごとに計算し、学術分野の登場初期からの指標の変化を可視化することにより特徴を抽出して、萌芽的論文を特定する基礎的な手法を提供する。

また、本研究の提案手法を実装して、科学技術に関わる意思決定をサポートする情報システムを設計することも本研究の目的である。

1.3. 先行研究

本研究の先行研究は、科学計量学・書誌計量学に多く見られ、情報通信技術の発達し身近なものになってから利用できるリソースが格段に増えており、定量的な分析に利用できるデータが多く残っている。そのなかでも、論文の引用ネットワークを用いた研究は古くからなされており、ツール化して実用化も行われているが、萌芽領域を発見する先行研究は「現在」「被引用数の多さ」が重視されており、近年の情報量急増によって近年の発表された論文が多く被引用数を集めるといったような偏りが考慮されていない。本研究は、論文の引用ネットワークを用いて、萌芽領域の発見を目指すものでありながら、論文の時系列の変化に注目する。

1.3.1. 萌芽領域の発見に関する先行研究

萌芽領域の発見に関する研究を紹介するに先立ち、科学計量学・書誌計量学の分野では、論文間の引用関係に基づいて重要分野を発見する引用分析が活用されている(Boyack,2005; Klavans, 2009)。その中には、クラスタリングや可視化に取り組んだ研究(Chen, 1999; Chen, 2003; Small, 1999)や、学術論文における大規模調査を行った研究(Leydesdorff, 2004; Leydesdorff, 2009)がある。柴田らは、統計的な処理のみでなく、引用ネットワークのトポロジカルおよびセマンティックな分析を萌芽領域の発見に活用する研究を行った(Shibata, 2008)。

萌芽領域を発見する手法として、論文間の関係が緊密なクラスターを発見するという方法がある。Price(1965)は、新興分野は発展途上で論文間の引用関係が緊密になるという考えを導出した。研究者は、新しい論文を引用する傾向にあり、萌芽領域にあるネットワークは緊密になる。Smallら(1974)は、活性化している分野の研究者は、共引用による引用ネットワークでクラスターを形成することを示した。Small(2006)は、被引用数がTOP1%の上位に属する論文について共引用で引用ネットワークを形成し、成長分野の追跡および予測も行っている。

一方、引用パターンが萌芽領域の発見に影響を与えることも分かっている。Klavansら(2006)は、直接引用と共引用によるジャーナル引用ネットワークのクラスター性能を比較し、直接引用の方がセマンティックなコヒーレンス測定に優れていることを明らかにした。柴田ら(2009)は、直接引用、共引用、書誌結合の異なる3つの引用パターンで比較を行い、萌芽領域の発見性能について調べた。さらに、藤田ら(2012a; 2012b)は、引用に重み付けを行った場合や引用パターンを組み合わせた場合について研究を行っている。

萌芽領域の発見のために時間の経過に伴う変化を導入した研究においては、引用ネットワークをクラスター分析した後、クラスター内の論文の平均年を各年の段階で計算して変化を見ることで、新興分野を検知する手法が提案されている(Shibata, 2011; Chen, 2012)。Tuら(2012)は、新興トピックを検出するために、トピックに対する新規性の指標を設定した。Liuら(2012)は、引用ネットワークとテキスト・マイニングから新興トレンドを発見しようという研究を行った。

ノーベル賞受賞候補の予測という点においては、かつては被引用数の高さと同関があった(Garfield, 1992)が、近年では、論文数の増加と分野の細分化により、必ずしも被引用数とは連動しないことが報告されている(Gingras, 2010)。

現在では、Thomson Reutersなどの大手情報企業やElsevierなどの大手の出版社が、引用分析を用いたツールを自社のビジネスとして実装し事業化を行っている。

1.3.2. 時系列分析に関する先行研究

時系列分析の有効性は、一般に、検知すべき構造が判明していない場合に同じ質の情報と比較可能ということ、および、時間の経過とともに情報が常に生成されデータが豊富に入手できるとい

うことにある。このことは、分析を始めやすくする一方、膨大な情報量のために分析に時間がかかるというデメリットも存在する。

時系列分析は、科学計量学・書誌計量学の分野の中でも行われている。2000年以前では、コンピュータが潤沢に利用できず、少ない論文数を短い期間分しか分析を行うことができなかった。しかし、近年では、多くのデータを対象に、時系列を利用した分析も増えてきている。論文が時間とともに変化する属性からフェーズの変化を明らかにしたり(Shibata, 2008; Joung, 2010)、複数の論文の引用関係から情報の伝播を辿ったり(Garfield, 2009; Lu, 2013)、学術分野のグループとしての指標や構造について時系列変化を追う(Shibata, 2008; Leydesdorff, 2012; Manh Cuong, 2012)といった分析がされている。Hughes(2010)は、分野の異なる機関どうしの共著関係からの、ネットワーク構造の時系列変化を使って、異なる分野間の協力関係を明らかにしている。Leydesdorff(2011)は、特定分野のジャーナル間の引用ネットワークにおいて学際的な構造変化を明らかにする中で、**Betweenness Centrality**の時系列を利用している。

1.3.3. 科学計量学・書誌計量学に関する先行研究

本研究で用いる手法は、図書館情報学(Library and Information Science)と呼ばれる分野のうち、細区分としては、科学計量学(scientometrics)、あるいは、書誌計量学(bibliometrics)に属する。これら、科学計量学、および、書誌計量学は、情報を計量する学術分野(informetrics)の一部を構成しながら、データをウェブサービスから取得することから、Figure 5における cybermetrics、webometrics の要素を含んだ分析となる。Table 1の数値は、2013-06-19時点で Thomson Reuters社の論文データベースサービス Web of Science、および、Microsoft社の検索エンジン bingにおける検索結果であるが、1990年代に論文が初めて登場した cybermetrics と webometrics は論文数もインターネット上の情報も低い状況になっているが、古くからある bibliometrics、scientometrics、informetrics に関しては、最古の論文が最も古い bibliometrics が Web of Science も bing 上もヒット数の総量が一番が多かったが、bibliometrics より scientometricsの方が学術に寄った語であることが分かった。

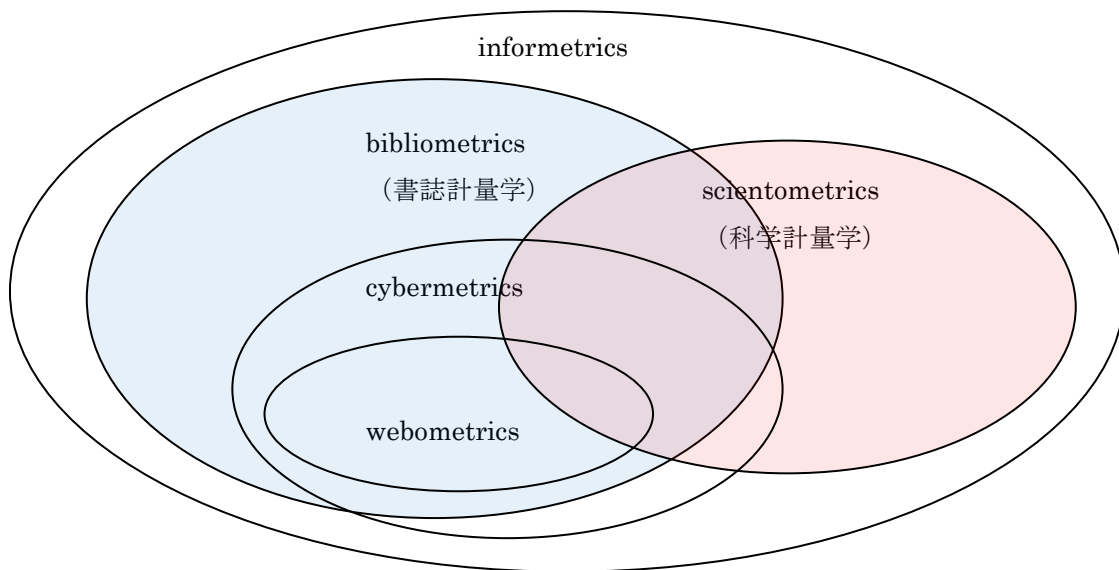


Figure 5 : 計量学間の関係 (Björneborn et al., 2004)

Table 1 : 科学計量学・書誌計量学関連分野の普及状況

Query	WoS#の最古出版年	WoS 上の論文数	Bing でのヒット数	WoS / Bing#
Informetric*	1979	330	4,870	0.0678
scientometric*	1976	1,186	13,400	0.0885
bibliometric*	1969	3,888	77,600	0.0501
cybermetric*	1999	34	4,750	0.0072
webometric*	1997	180	11,500	0.0157

WoS : Thomson Reuters 社の論文データベースサービス Web of Science
 WoS / Bing : Web of Science での論文数を、Bing での検索結果で除したもの

書誌計量学において、萌芽領域の発見や研究者を評価する研究は多く行われてきた。Figure 6 は、本研究に関わる研究を整理するための切り口であるが、科学技術を評価するにあたって Figure 6 で示した軸の他に、特徴抽出、特徴量の設定、検出、予測といったプロセスのうち、どこまで対象にするのかという点も論文により異なる。

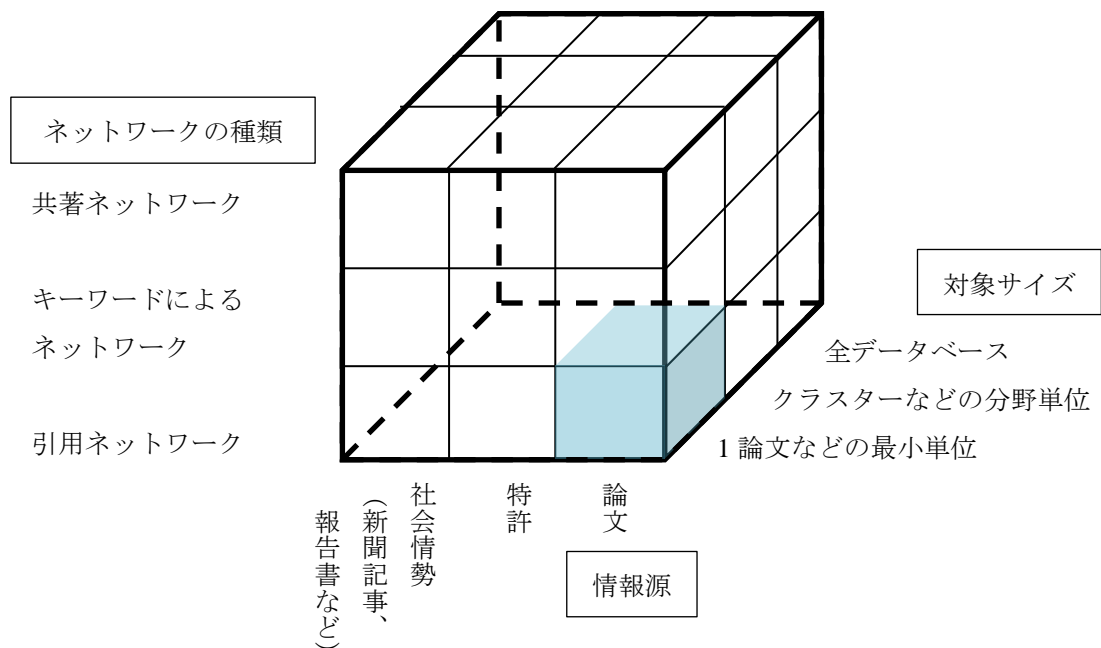


Figure 6: 学問の萌芽領域・重要論文を検出する研究の分類

論文においては単体を評価する指標として、個々の論文ごとの被引用数が使用されることが多く、学術ジャーナルの重要度を評価する Impact Factor (Garfield, 1955) も、研究者の科学技術への貢献度を測る h-index (Hirsch, 2005) も、論文への被引用数を活用したものである。主要な指標で使われる被引用数は、専門知識を持った研究者によって引用した研究内容が評価された結果であるため、引用を多く受けた研究ほど優れた研究であると判断することは合理的である。被引用数は有効な指標ではあるものの、しかしながら、引用関係がないグループ同士では、内容上の類似度があっても別のテーマとして扱われることになる。さらに、例えば、医学の論文は数が多く被引用数も多く集める傾向にあるように、学術分野が異なれば比較ができないという問題がある。また、被引用数は、投稿したジャーナルの Impact Factor や所属組織の著名度、共著者の国籍の広がりなどによって顕著な傾向がある (Länzel, 2001; Slyder et al., 2011) ことから、論文の内容だけでなく、属性によって評価が左右されるという問題はある。

キーワードによるネットワークを形成する方法もある (Roche et al., 2013)。この場合には、著者の引用文献の見落としや、同時期に発行されたことにより引用関係が切れたグループ間を繋げてネットワークを形成できるメリットがある。しかし、本当に関係が存在するのか内容を確認するレビューのようなプロセスがないため、引用関係を使用する場合より精度が落ちる繋がりとなる。

例えば、情報工学のエントロピーの議論で、エントロピーの語源が熱力学に由来することを記述すれば、情報工学と熱力学の分野がつながりかねない。

共著によるネットワークでは、著者同士の繋がりだけでなく、組織間、国・地域間の協力関係を表し(Sakata et al., 2011; Sasaki et al., 2010)、どの人物、どの場所に知識が集まっているのか、把握することができる。

情報源には複数考えられ、論文、特許、新聞等のその他のデータがあるが、引用ネットワークが構成できるのは、論文と特許となる。イノベーションが、研究、開発、製品化、事業化というプロセスを経るにあたって、論文と特許は研究と開発のプロセスに相当すると考えられる。

ノーベル賞の分析に書誌計量学を取り入れたケースには、ノーベル経済学賞受賞者の論文に対する被引用数を用いて、受賞の前後でライフサイクルのフェーズにあるかを分析したものがある(Bjork, 2013)。

1.4. 本研究の貢献

従来の萌芽領域発見に関する研究は、現在の被引用数から重要論文を決定する手法が大勢を占めた。本研究においても、被引用数に影響を受けることには違いないが、現在の被引用数を見るだけでは論文数の増加に飲み込まれて埋もれてしまった萌芽的論文を、時系列分析を導入することで抽出することに意義がある。これまでは、ノーベル賞受賞者選定過程のように、萌芽的論文は権威のある専門家の精査によって探し出されるものであった。本研究にて提案する手法では、機械的に萌芽的論文を抽出することができ、専門家による精査の手間を削減することに役立つだろう。



Figure 7: 本研究の要素

1.5. 本論文の構成

本節では本論文の構成を記す。第 1 章では、序論として、背景、研究の目的を定義し、先行研究、本研究の貢献について述べる。

続いて、第 2 章では、研究の目的を達成する手法を提案し、提案手法を検証するために使用したデータセットについて、概要を示す。

次に、第 3 章では、各データセットについて提案手法での分析結果を示し、第 4 章では、手法の一般性について検証し、結果について考察を行う。

そして、第 5 章では、第 4 章までの分析結果とその可視化による観察の上で、萌芽的論文を特定するための特徴量の候補について提案を行い、その検証を通じて、有効な特徴量群を特定する。

第 6 章では、提案手法を、ウェブインターフェースを使って可視化したシステムを紹介する。

最後に、第 7 章では、結論と今後の展望を述べる。

本論ののち、巻末には、参考文献、謝辞、付録が付してある。

2. 提案手法

本研究における分析は、以下の手順で行う。

- (1) 書誌情報の収集
- (2) 引用ネットワークの形成とクラスター分析
- (3) 指標の算出
- (4) 時系列分析の可視化
- (5) 萌芽的論文の特徴抽出 (第3章、第4章の結果・考察より判断する)
- (6) 特徴量の設定 (第3章、第4章を受け、第5章において詳細を検討する)

対象となる学術分野について収集した学术论文の書誌情報を、各年までに出版された範囲で引用ネットワークを形成してクラスター分析を行う (Figure 8 上段)。各年で形成された引用ネットワークに対して指標の計算を行い、値を時系列で表示する (Figure 8 下段)。時系列で表現された指標の特徴により、萌芽的論文を検出する。本研究で扱うデータは、(データセット数 : 7 セット) × (年数 : 37 ~ 52 年) × (指標の数 : 12) となる。

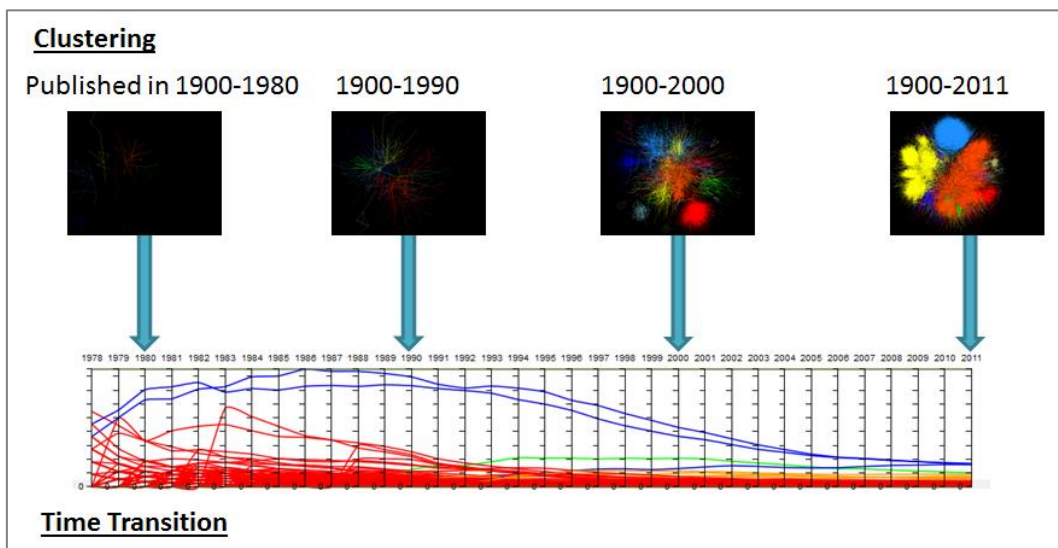


Figure 8 : 指標の時系列から各年の引用ネットワークを形成する概念図

2.1. 書誌情報の収集

論文の書誌情報収集に先立って、まず、分析対象とする分野のノーベル賞受賞論文を特定する。ノーベル賞受賞論文は、ノーベル賞の公式サイト (<http://www.nobelprize.org/>) で発表される、(a) Press Release の Key publications 欄、(b) Popular Information の Original articles 欄、あるいは、(c) Advanced Information の参考文献などうち、明示的に受賞論文を示されている場合はその論文を、明示的でない場合には参考文献と解説文書の内容を照らし合わせてノーベル賞受賞論文を特定した。

また、書誌情報を収集するためにキーワードは、ノーベル賞の公式ホームページには受賞理由が記載されているので、この受賞理由より当該分野を特徴づけるキーワードを選定し、書誌情報収集ためのクエリーとして使用した。ただし、受賞理由を用いたキーワードでは、先に特定したノーベル賞受賞論文を含まれない場合がある。その場合は、含まれないノーベル賞受賞論文のタイトルから特徴的な語を選定して、論理和(OR)でクエリーに追加し、論文の書誌情報を収集する。つまり、ノーベル賞受賞論文の特定も、キーワードの選定も、客観的に特定するようにした。例外として、ACM チューリング賞を受賞した RSA 暗号に関して、専門知により、受賞論文を特定しキーワードを選定している。

これらの手順を通して決定したクエリーを Thomson Reuters 社の論文データベース Web of Science で検索し、得られた書誌情報を集めたひとまとまりをデータセットという。

2.2. 引用ネットワークの形成とクラスター分析

引用ネットワーク分析は、1955年に Garfield により提案され、その後、人気のあるトピックや重要な学術論文を効果的に抽出するツールが開発されている(Börner et al., 2003)。引用ネットワークでは、学術論文が引用ネットワークのノードとして、学術論文間の引用関係がエッジとして定義される。引用ネットワークは、各学術論文の著者が自分の論文と関連する論文を引用するため、その分野の専門家の判断が含まれたものになると考えられる。

引用ネットワークを形成する際に同時に行われるクラスター分析は、Figure 9のように、以下の手順で行う。

- (2-1) 分析の対象となる論文の書誌情報を収集する。
- (2-2) 引用ネットワークを形成する。
- (2-3) 引用ネットワークの最大連結成分を取り出す。
- (2-4) クラスタに分割する。
- (2-5) クラスタ分析の結果により、可視化や各クラスタの内容特定、さらなる分析を加える。

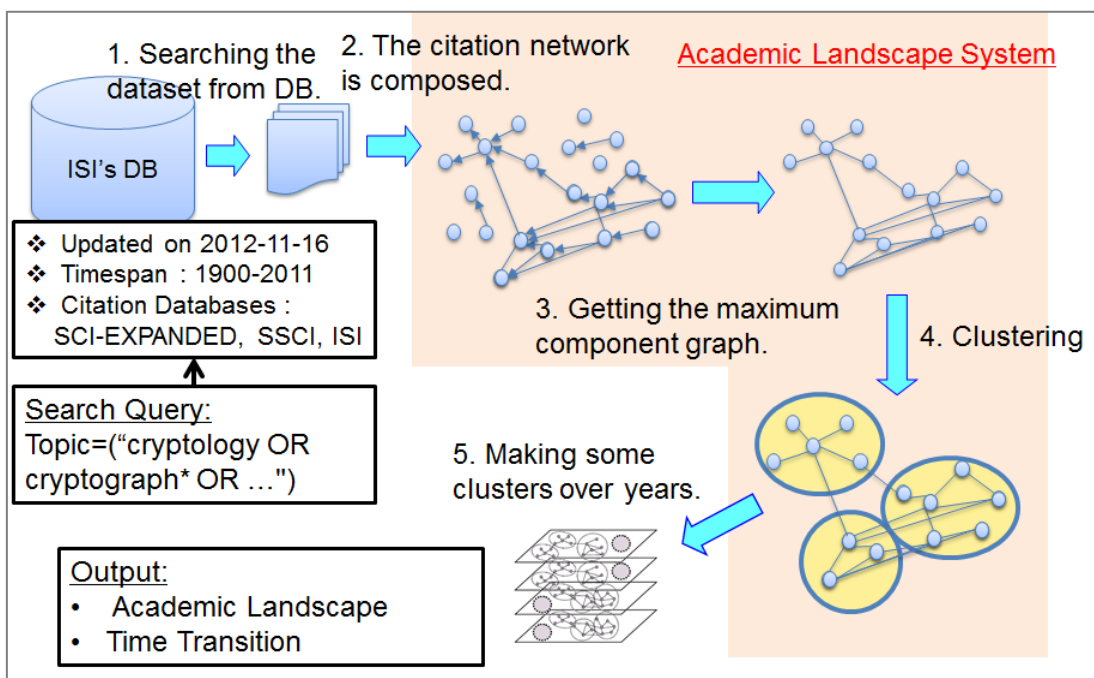


Figure 9 : Flow of Cluster Analysis

ステップ(2-1)は、前節で説明したように、分析対象の分野についてデータセットを得る。データ取得時ではその年に出版される論文が確定していないため、前年までに出版されたデータを用いて分析する。

ステップ(2-2)においては、論文の引用関係より引用ネットワークを形成する。柴田ら(2009)の大きく若い新興クラスタを早い時期に検出するには直接引用が適切であった旨の報告があり、また、論文数の多いデータセットを分析するにあたって、他の引用関係に比べて計算が速く終わる為、本研究における引用ネットワークを形成するための引用関係には、直接引用(direct citation)

を用いる。直接引用とは、学术论文の引用・被引用の関係である。直接引用の他に、引用・被引用の関係がない2つの論文が第三の論文に共に引用されることで関係性が発生する共引用(co-citation)、引用・被引用の関係がない2つの論文が第三の同じ論文を引用することで関係性が発生する書誌結合(bibliographic coupling)がある。Figure 10に、直接引用、共引用、書誌結合の3つの引用関係を図示する。

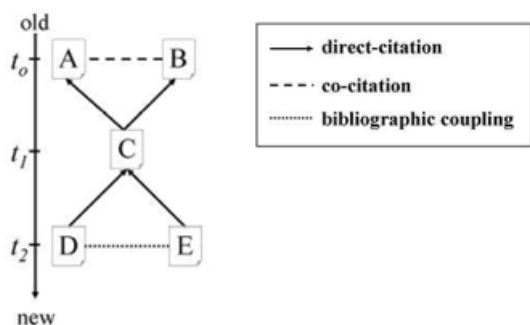


Figure 10: 引用関係 (直接引用、共引用、書誌結合)

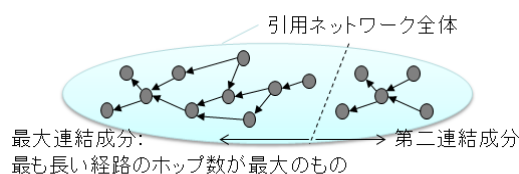


Figure 11: 引用ネットワークの連結成分

ステップ(2-3)において、引用ネットワークの最大連結成分を取り出す。最大連結成分とは引用ネットワークのすべてのノード間の経路で最も長い経路が最大のホップ数を持つネットワークである。引用ネットワークでは、Figure 11のように、相互に繋がりを持たない複数のネットワークが形成されることがある。ステップ(2-2)で複数の独立したネットワークを形成した場合、最大連結成分とは繋がりのなかったネットワークは、ステップ(2-4)のクラスターに分割する際に対象とならない。また、時系列分析で年ごとにクラスターに分割すると、二番目以降だったネットワークが最大連結成分に成長したり、複数のネットワークが合流して一つになることがある。

ステップ(2-4)において、Newman法(2004)を用いて、クラスターに分割する。クラスタリング手法は複数存在しているが、Newman法は、1)大規模ネットワーク(数万ノード規模)に適用して、現実可能な計算時間でクラスタリングが実行可能である、2)ネットワーク構造を局所的ではなく、俯瞰的に見てクラスター間のエッジを切断できる、という点で採用している。特に、より多くの情報量をより短い時間で分析できることは、学術分野の論文数が数万~数十万に達することがあることから合理的な時間内に計算を完了するために重要な要因となる。

なお、ステップ(2-2)から(2-4)の分析については、東京大学イノベーション政策研究センターより学術俯瞰システム (Innovation Policy Research Center, 2013) といった分析ツールが提供さ

れており、ステップ(2-1)で得たデータセットを本システムに投入して、ステップ(2-2)から(2-4)の分析を経て得られた結果を用いて、ステップ(2-5)として以降の分析を行っている。既存のシステムを利用することで、論文の同一性認定において、これまでに蓄積された知見を利用することができる。

2.3. 指標の算出方法

次に、各データセットにおいて、各年の引用ネットワークの各論文のネットワーク中心性指標を、Table 2 のように 12 種類、計算した。近年の研究において、ネットワーク中心性指標と萌芽領域との関係が注目されている(Leydesdorff, 2006; Shibata et al., 2008; Chen et al., 2009)が、これらの研究では数が少ないので、より多くの指標を採用し網羅的に扱っている。これらの指標は、python のモジュールである NetworkX (NetworkX developer Team, 2013) を用いて算出を行っている。なお、本節で述べるネットワーク指標の定義は、付録 B において説明する。

Table 2: 各指標の説明

指標	説明
<i>次数および固有値</i>	
In-Degree Centrality (入次数中心性)	ネットワーク内での関係が多いほど大きくなる指標である。ネットワーク内での人気の高さを表す。
Average Neighbor Degree	あるノードに隣接するノードが持つ次数を平均したものである。
Eigenvector Centrality (固有ベクトル中心性)	本当の有力者は他の有力者と多くつながっていることを示す指標である。
Communicability Centrality (伝達性中心性)	ノード n を始まりと終わりとする閉鎖経路の長さの合計である。
PageRank	ランダムアクセスで辿り着く可能性の高さを示す。影響力の流れを測ることができる。
<i>距離</i>	
Closeness Centrality (近接中心性)	他の論文との関係に近いほど高くなる指標である。情報を伝達力の高さを表す。
Closeness Vitality	そのノードを除き、すべてのノードのペア間の距離の和の変化である。

媒介性	
Betweenness Centrality (媒介中心性)	ボトルネックとなるような最短経路上にいるほど高くなる指標である。
Load Centrality	ボトルネックとなるような最短経路上にいるほど高くなる指標である。Betweenness と計算方法が異なる。
その他	
Clustering Coefficient (クラスター係数)	グラフ内のノードが集中する傾向を示す。
Core Number	最大のサブグラフの次数を示す。
Node Redundancy	エッジの密度から、他の論文との関係が近いほど高くなる指標である。

なお、他にも利用可能なネットワーク指標は存在するが、計算に時間がかかり過ぎる、あるいは、値の欠損が多い為に利用を断念したネットワーク指標もある。実際の分析に置いては、合理的な時間内に分析が完了することも重要な要素であるため、前者 3 指標は必須とし、後者 9 指標はデータセットの状況により計算しない場合がある。In-Degree Centrality については、重要論文は時間が経過しても被引用数が増加する(Adams, 2005; Wang, 2012)ということから、萌芽的論文の検出への利用が期待される。

2.4. 時系列分析の可視化

各年の引用関係をもって指標を算出すると、各論文で出版年(正確には、引用関係が発生した年)から最新年まで、毎年の指標の値の変動が見られることになる。時系列分析では、1つの論文に対して毎年の指標の値を1本の線としてつなげて、Figure 12のように図示する。可視化にあたっては、Data-Driven Documents(Bostock, 2013)で提供されているライブラリを利用して、Webインターフェースで描画している。各線の色は、最新被引用数の最大値と最小値を5等分して、値の高い方から、青、アクア、グリーン、オレンジ、赤で示す。論文の最新被引用数の分布は、最新被引用数が多いほど論文数が少なく、最新被引用数が少ないほど論文数が多くなるので、当然に青の本数が少なく、赤の本数が多くなる。上下のスケールは表示対象のデータが持つ最大値と最小値に規定される。

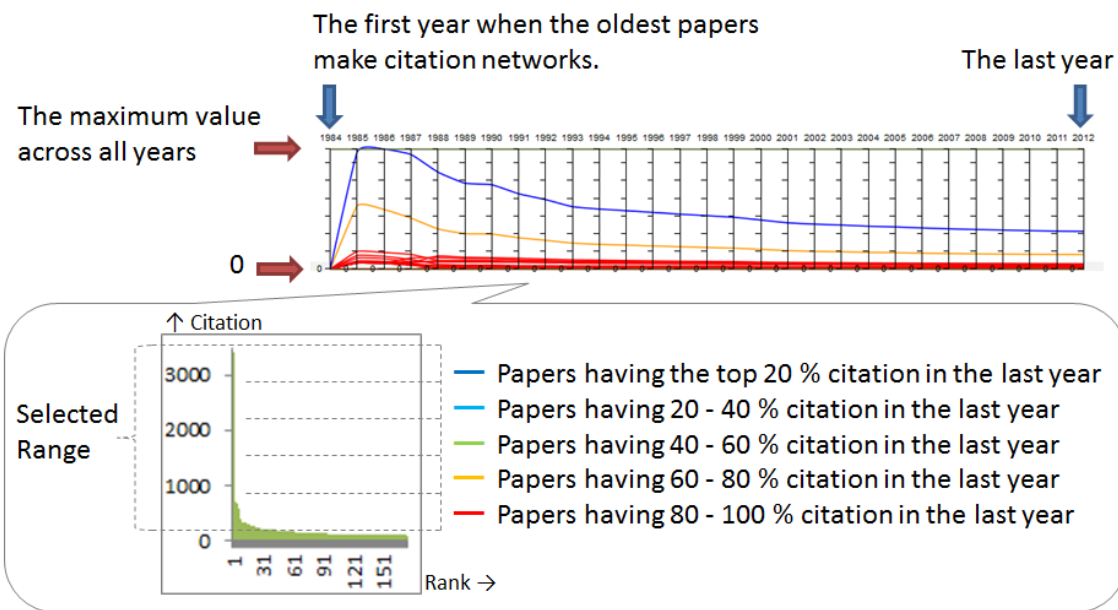


Figure 12 : How to read figures of time transition

2.5. 使用するデータセット

本研究では、ノーベル化学賞、ノーベル生理学・医学賞、ACM チューリング賞から、以下のデータセットを使用する。暗号(crypto)と蛍光プロテイン(gfp)以外は、ノーベル化学賞およびノーベル生理学・医学賞の分析開始時点での最新の受賞分野を選定している（ただし、分析に時間がかかったため、2011年ノーベル生理学・医学賞は除く）。Table 3にデータセットの特徴をまとめる。一般に、ライフサイクルは、導入期・成長期・成熟期・衰退期の4期間であるが、各年に出版される論文の増減で判断するにあたって論文数が増加中である場合、導入期から成熟期の前半まで区別することは難しい。ゆえに、以下の3期間で判断した。

- 成長期：論文数が、最大数の年から増加傾向にある場合
- 成熟期：論文数が、最大数の年から減少を示し、最新年では最大数の3分の2以上を維持している場合
- 衰退期：論文数が、最大数の年から減少傾向にあり、最新年では最大数の3分の2未満となった場合

暗号(crypto)とパラジウム触媒クロスカップリング反応(2010chem)のデータセットにおいて、一時期、基調から外れた突発的な論文数の増減が見られるが、その前後の時期から推測して基調の傾向からライフサイクルを判断する。

E/N 比とは、エッジ数をノード数で割ったものであり、値が大きいほど密接に論文どうしが繋がっていることを表す。上位論文の特徴は、最新被引用数 TOP 論文が 4,000 を越える論文から引用されている場合を高位、4,000 以下の場合を低位とし、最新被引用数 TOP 論文と 10 位論文との最新被引用数の差が大きい場合をチャンピオン、差が大きい場合を大競争とする。

Table 3: データセットの特徴

	使用論文数 取得論文数	平均出版年 ライフサイクル	E/N 比 上位論文の特徴
暗号 [crypto]	15,273 19,448	2004.5 成長期	5.73 低位・大競争
蛍光プロテイン [gfp]	28,926 32,459	2005.2 成熟期	5.89 低位・大競争
G タンパク質共益受容体 [2012chem]	46,421 54,774	2000.8 成熟期	10.90 低位・大競争
iPS 細胞 [2012med]	12,513 18,538	2007.5 成長期	9.58 高位・チャンピオン
準結晶 [2011chem]	6,906 9,335	2001.1 衰退期	7.68 低位・チャンピオン
パラジウム触媒 クロスカップリング反応 [2010chem]	26,546 66,656	1999.4 成長期	5.65 高位・チャンピオン
体外受精技術 [2010med]	50,609 66,561	2003.4 成長期	7.06 低位・大競争

これらのデータセットを、ライフサイクルと E/N 比、および、ライフサイクルと上位論文の特徴でポジショニングすると、Figure 13、および、Figure 14 のようになる。

(E/N 比) 9 - 10	iPS 細胞[2012med]	G タンパク質共益受容体[2012chem]	
7 - 8	体外受精技術 [2010med]		準結晶[2011chem]
5 - 6	暗号[<i>crypto</i>] パラジウム触媒クロス カップリング反応 [2010chem]	蛍光プロテイン[<i>gfp</i>]	
	成長期	成熟期	衰退期

Figure 13: ライフサイクルと E/N 比における、各データセットのポジション

(上位論文の特徴) チャンピオン	パラジウム触媒クロス カップリング反応 [2010chem] iPS 細胞[2012med]		準結晶[2011chem]
大競争	暗号[<i>crypto</i>] 体外受精技術 [2010med]	蛍光プロテイン[<i>gfp</i>] G タンパク質共益受容 体[2012chem]	
	成長期	成熟期	衰退期

Figure 14: ライフサイクルと上位論文の特徴における、各データセットのポジション

3. 引用ネットワークの時系列分析の結果

本章では、引用ネットワーク上の指標を時系列で分析して、それぞれのデータセットについて、複数の指標を比較することで目的に適った指標の選定を行う。本章のデータセットは、書誌情報の収集の際に、ノーベル賞受賞論文、および、ACM チューリング賞受賞論文を含むように選定しており、本研究の分析に際しては、正解としての萌芽的論文が既にそれぞれのデータセットに含まれている。正解とする論文から萌芽的論文となるべき特徴を抽出することができる。

3.1. 暗号(2002年 ACM チューリング賞 [CRYPTO])

本節では、2002年の ACM チューリング賞より、暗号分野について、指標を時系列で図示することにより、有効な指標を選定する。本データセットのみ、ノーベル賞受賞論文ではないが、本研究の端緒となったデータセットであり、ノーベル賞受賞論文との違いから示唆が得られることから、重要な位置づけにあるデータセットである。

3.1.1. 分析対象分野の概要

暗号は、すでに古代ギリシャに存在しており、例えば、シーザー暗号は、暗号化の基本として有名なシフト暗号である。もちろん、このシーザーは紀元前1世紀にローマ共和国の一つで政治家である。現代の学界では、暗号のいくつかの成果は、DES や RSA のように実用化されているが、一方、解決されるのを待っている現在進行中のトピックでもある。暗号理論は、数学的な基礎研究に基づいているが、コンピュータ化が浸透した日常生活で適用されている応用研究も進んでいる。このように、暗号は、古くて新しいトピックであり、基礎研究にも応用研究にも成果を上げている分野である。また、現在の暗号学は、暗号学全般として **cryptology** がある中、暗号方式やその性能など生成側に関する分野を **cryptography**、暗号解読法に関する分野を **cryptoanalysis**、データ隠蔽技術に関する分野を **steganography** といった複数の分野から成る (Mollin, 2002)。

このように、長い期間をかけて漸進的に進展してきた暗号分野であるが、Rivest 氏、Shamir 氏、Adleman 氏の3名は、RSA 公開鍵暗号(1978)について、2002年に ACM チューリング賞

を受賞した。このチューリング賞は、コンピュータ科学のノーベル賞と呼ばれているほど、計算機分野では重要な賞である。RSA 公開鍵暗号は、大きな素数を分解することの難しさに基づいたアルゴリズムの一つである。現在では、RSA 公開鍵暗号は、暗号化やデジタル署名などの日常の通信の中で使用されている。

暗号分野の分析をするにあたって、受賞の際に発表される資料から、以下の論文をに貢献した論文として抽出した。

- Rivest, RL, A Shamir and L Adleman (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2), pp.120–126. [WOS:000285713900009]

Thomson Reuters 社の論文サービス Web of Science より、「cryptology OR cryptograph* OR cryptanalysis* OR cryptanalyz* OR steganograph* OR cryptosystem* OR cryptogram OR encrypt* OR decrypt* OR cipher* OR encipher OR decrypt*」というクエリーで、タイトル、アブストラクト、キーワード、キーワード・プラスを検索し、19,448 論文の書誌情報を得た。そのうち、2011 年末時点で引用ネットワークを形成可能な 15,273 論文の書誌情報を対象に、複数の指標を各年について算出した。このクエリーを決定するにあたっては、専門家の知見により一般的に暗号分野のみを示す語を選定した。Figure 15、および、Table 4 に、対象となる論文の出版年の分布と、書誌情報を収集した 2012-11-16 時点での被引用数(以降、最新被引用数とする)の分布を示す。Figure 15 の論文数の経年増加をみると、2002 年から 2006 年に、その期間以外とは異質な急増が見られるが、暗号分野は最新年まで増加を続けている分野である。RSA 暗号の論文(1978)は、Web of Science 上の集計で 3,150 論文から引用されている。

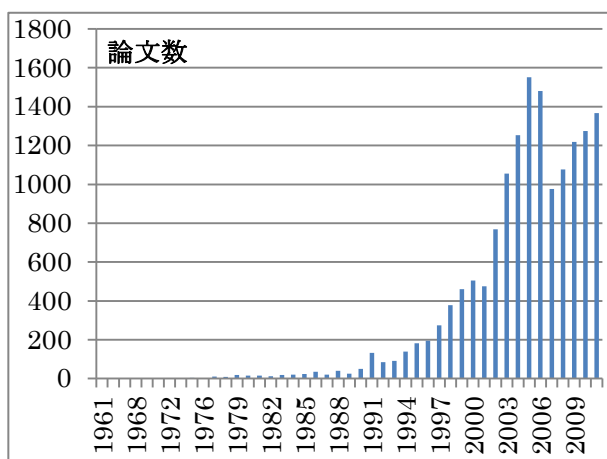


Figure 15: 出版年ごとの論文数の分布 [crypto]

Table 4: 最新被引用数ごとの論文数の分布 [crypto]

Times Cited	Count
0	3,119
1 - 100	11,735
101 - 200	275
201 - 300	73
301 - 400	35
401 - 500	6
501 - 600	6
601 - 700	7
701 - 800	3
801 - 900	2
901 - 1000	1
1001 - 2000	5
2001 - 3000	3
3001 - 4000	3

これらの論文の 2011 年末時点での引用ネットワークは Figure 16 のように形成される。クラスター分析を行った結果の各クラスターは各色で示され、TOP 7 クラスターの内容を Table 5 に、クラスター間のキーワード類似度を Figure 17 に示す。

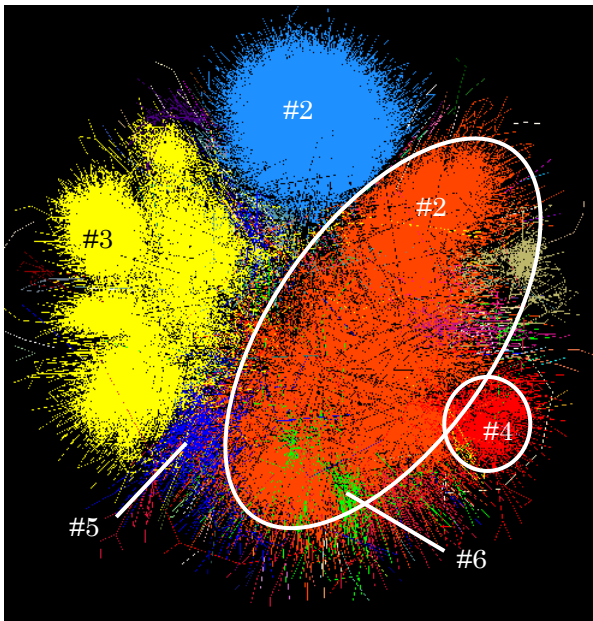


Figure 16 : Academic Landscape Map of Cryptology [crypto]

	1	2	3	4	5	6	7
1	1.00	0.03	0.09	0.18	0.04	0.19	0.10
2	0.03	1.00	0.03	0.06	0.01	0.01	0.02
3	0.09	0.03	1.00	0.01	0.00	0.06	0.04
4	0.18	0.06	0.01	1.00	0.04	0.09	0.13
5	0.04	0.01	0.00	0.04	1.00	0.05	0.01
6	0.19	0.01	0.06	0.09	0.05	1.00	0.08
7	0.10	0.02	0.04	0.13	0.01	0.08	1.00

Figure 17 : Heatmap of Cosine Similarity between Top 7 Clusters [crypto]

Table 5 : Information of TOP 7 Clusters [crypto]

	Ave.PY*	Nodes	Edges	E/N*	Topics
All	2004.5	15,273	87,507	5.73	
#1	2003.6	6,693	30,999	4.63	cryptosystem
#2	2005.6	2,853	29,369	10.29	quantum cryptography
#3	2006.2	2,850	17,132	6.01	image encryption, steganography
#4	2002.2	540	1,276	2.36	protocol
#5	2006.2	509	1,043	2.05	implementation, hardware
#6	2003.6	463	1,480	3.20	access control
#7	2006.7	314	427	1.36	sensor network

*Ave.PY: Average year that papers are published

E/N: Density that number of edges is divided by number of papers

3.1.2. 各指標の時系列分析

指標については、in-Degree Centrality、Closeness Centrality、Betweenness Centrality ほか、複数の指標について各年の引用ネットワークの中で算出しているが、本項では、重要論文を検出するのに有効な指標について、詳細を解説する。

Table 6 は、この引用ネットワークで最新被引用数が多い順から TOP 10 の論文リストである。Figure 18 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。Figure 19 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。ノードのサイズがそのクラスターの所属する論文数に比例し、クラスターの 8 以降は煩雑になるため省略している。オレンジの破線は、RSA 暗号に関する Rivest 氏、Shamir 氏、Adleman 氏 3 名の共著論文が辿った軌跡である。この論文は、2 位までの上位のクラスターにあって、多くの論文が経由する軌跡も辿っている。

Table 6 : TOP 10 Papers on 2012-11-16 [crypto]

R	1 st Author	PY	TC	Cluster Number							
				1980	1985	1990	1995	2000	2005	2010	2011
1	Diffie, W	1976	3,339	1	3	3	2	1	1	1	1
# 2	Rivest, RL	1978	3,150	1	2	2	1	1	1	1	1
3	Ekert, AK	1991	3,125	-	-	-	5	4	2	2	2
4	Gisin, N	2002	2,269	-	-	-	-	-	2	2	2
5	Bennett, CH	1996	2,166	-	-	-	-	4	2	2	2
6	Cox, IJ	1997	2,064	-	-	-	-	6	4	3	3
7	Elgamal, T	1985	1,552	-	4	5	2	1	1	1	1
8	Shor, PW	1997	1,280	-	-	-	-	4	2	2	2
9	Koblitz, N	1987	1,098	-	-	5	2	1	1	1	1
10	Bennett, CH	1992	1,082	-	-	-	5	4	2	2	2

* R : Rank based on TC, PY : Published Year(出版年), TC : Times Cited(最新被引用数), #: 受賞論文

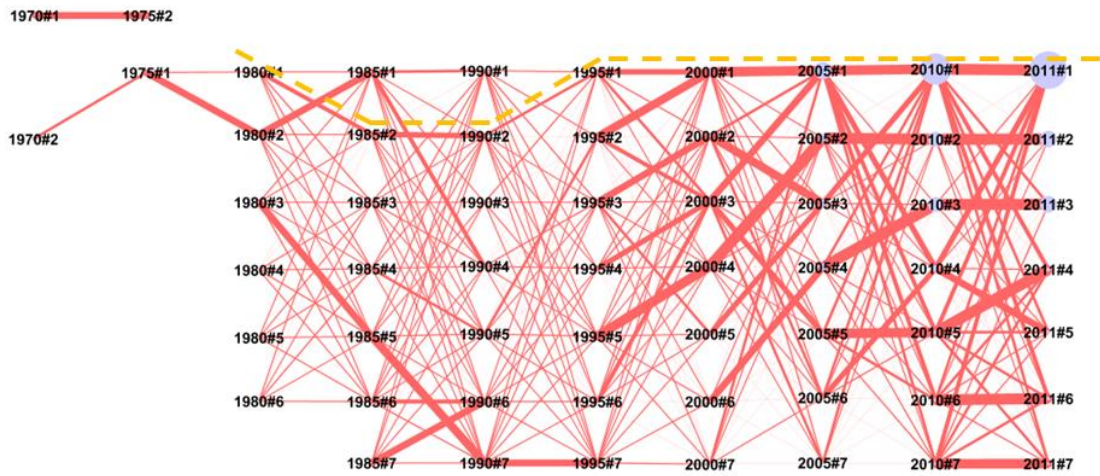


Figure 18: キーワードに基づくクラスター間の類似度 [crypto]

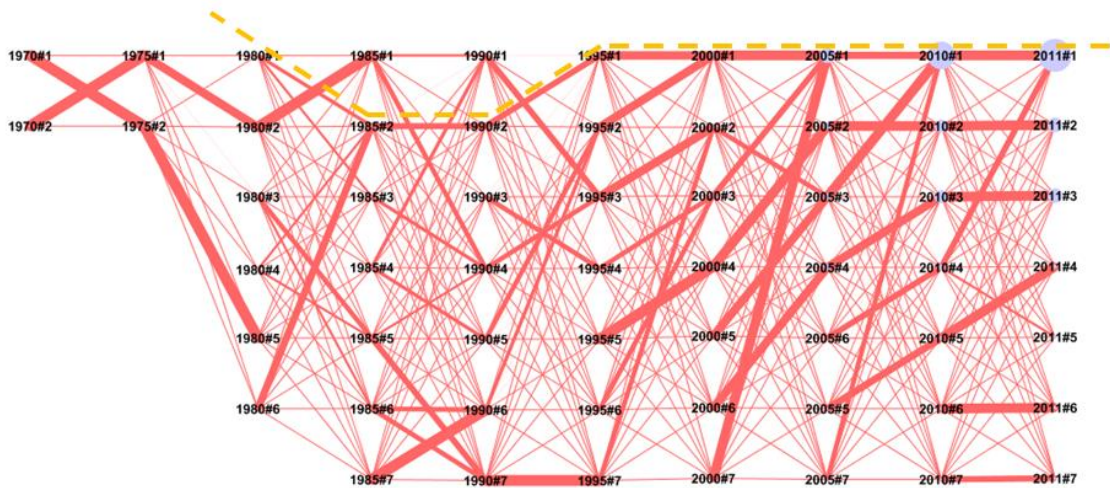


Figure 19: 論文の遷移量に基づくクラスター間の類似度 [crypto]

Figure 20 は、データセット内被引用数の時系列変化であり、青線は最新被引用数が 3,000 を越える論文であり、Table 6 における上位 3 論文を示す。なお、Figure 20、Figure 21、Figure 22 では、最新被引用数が 100 を越える 419 論文について表示している。

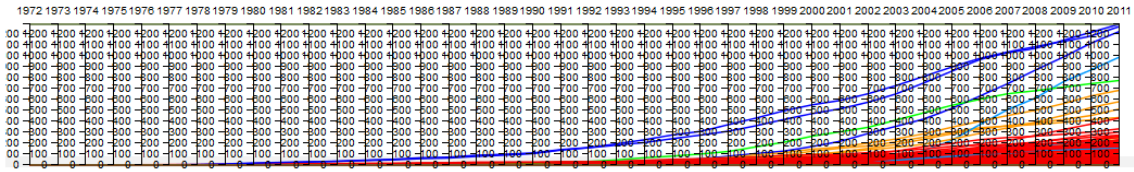


Figure 20 : Time Transition about Number of within-Dataset Citation of 419 Papers having more than 100 citations (最大値=1288) [crypto]

Figure 21 は、in-Degree Centrality の時系列変化であり、青線は最新被引用数が 3,000 を越える論文であり、Table 6 における上位 3 論文を示す。これら 3 論文のうち、1970 年代に出版された 2 論文は、他の論文も登場する 1980 年代から 2000 年代前半にかけて、他と乖離する大きな in-Degree Centrality を持つ。2 論文の一つは、本分析で取り上げる対象とした RSA 暗号であるが、もう一方は、Diffie-Hellman 鍵交換であり、ともに、公開鍵暗号システムの基本的な理論として、現在、広く実用化されている技術となっている。

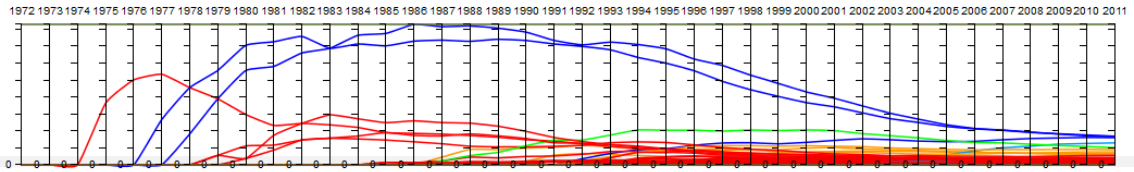


Figure 21 : Time Transition about in-Degree Centralities of 419 Papers having more than 100 citations (最大値=0.4136) [crypto]

Figure 22 は、PageRank の時系列変化であり、Figure 21 の in-Degree Centrality と同様に、1984 年以降は青線 2 本（Diffie-Hellman 鍵交換、および、RSA 暗号）が顕著に他の論文と乖離しており、PageRank を利用した重要論文の特定も可能であると考えられる。

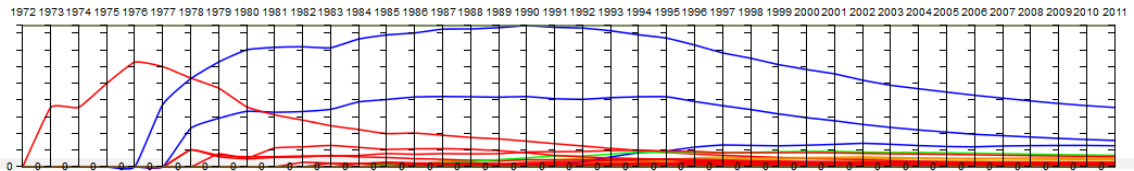


Figure 22 : Time Transition about PageRank of 419 Papers having more than 100 citations (最大値=0.1691) [crypto]

以上から、in-Degree Centrality と PageRank において重要論文は発行年から数年経つと高い数値を持つことから、他と比べて値の高い論文を重要論文として検知できると考えられる。

しかしながら、Figure 21、および、Figure 22 において、最新被引用数が 100 を超える論文のみに限定して描画したため、最新被引用数が低い論文が高い論文よりも高い in-Degree Centrality と PageRank を持つ懸念が残る。そこで、Figure 23 と Figure 24 は、重要論文を検出できる可能性の高い in-Degree Centrality と PageRank において、最新被引用数の低い論文が高い値を持つ可能性を検証するために、最新被引用数が 1 の論文群について、時系列で値を描画したものである。Figure 21、および、Figure 22 と比較すると、最新被引用数の低い論文群における in-Degree Centrality と PageRank の最大値は、最新被引用数の高い論文群の 4 分の 1、および、2 分の 1 以下であるので、in-Degree Centrality と PageRank により重要論文は検知できることが分かる。

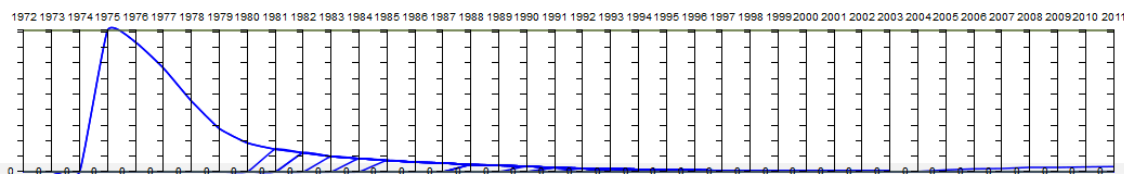


Figure 23 : Time Transition about in-Degree Centralities of 1978 Papers having 1 citation (最大値=0.09091) [crypto]

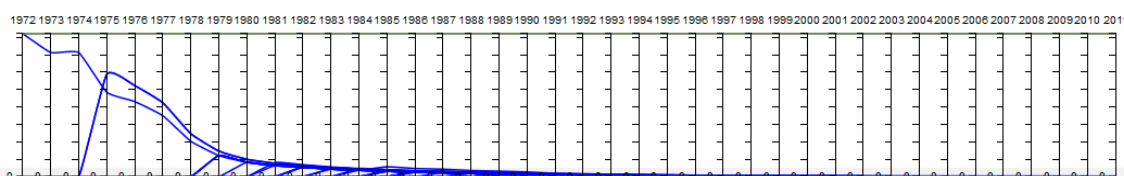


Figure 24 : Time Transition about PageRank of 1978 Papers having 1 citation (最大値=0.08200) [crypto]

Communicability Centrality については、Figure 21 と Figure 22 の in-Degree Centrality と PageRank とは違った形で重要論文を検出できる可能性がある。Figure 25 は、in-Degree Centrality と PageRank と同様に、最新被引用数が 100 を越える論文について Communicability Centrality の時系列変化を描画したものであるが、最新年まで描画した状態ですでに重要論文と分かっている論文の最新被引用数が大き過ぎて、新興の重要論文を検出するのは難しい。しかし

ながら、Figure 26 のように、出版年直後の数年の論文に限定して比較すると重要論文の Communicability Centrality が他の論文と乖離して重要論文を検出できるであろう。

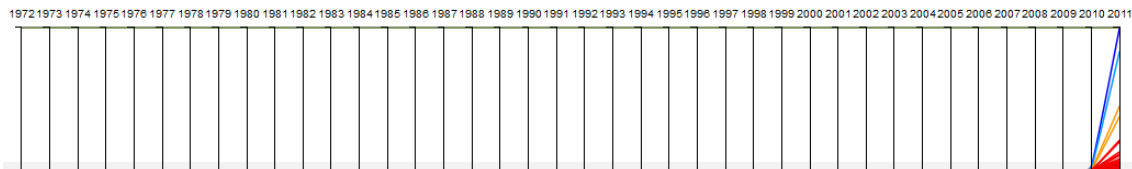


Figure 25 : Time Transition about Communicability Centralities of 419 Papers having more than 100 citations (最大値= 1.160e+29) [crypto]

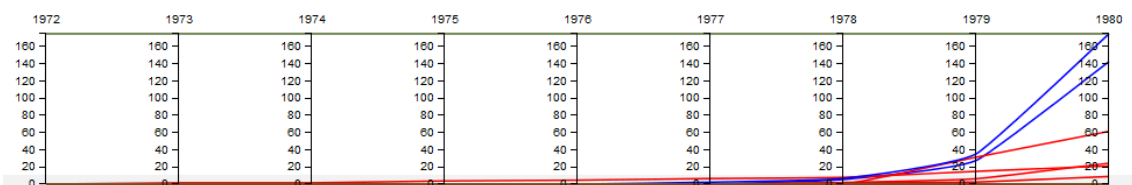


Figure 26 : Time Transition about Communicability Centralities of Papers having more than 100 citations (最大値= 175) [crypto]

どの指標が有効であるかを明らかにするために、Figure 27 に最新被引用数と指標の相関係数を、Figure 28 に各年のデータセット内の被引用数と指標の相関係数を、年ごとに計算した。Figure 27 では、相関係数の高い in-Degree Centrality は、このデータセットの初期から現在まで 0.7 以上と高い相関係数を保っており、また、Communicability Centrality、PageRank も相関係数が 0.6 以上、0.5 以上と、最新被引用数との相関があり、重要論文として特別に突出した論文だけに有効な指標ではなく、重要論文に次ぐレベルの論文群も検出可能である。Figure 28 では、被引用数の正規化した値である in-Degree Centrality は相関係数が常にほぼ 1 であり、次いで、Communicability Centrality、PageRank が弱い相関を示している。

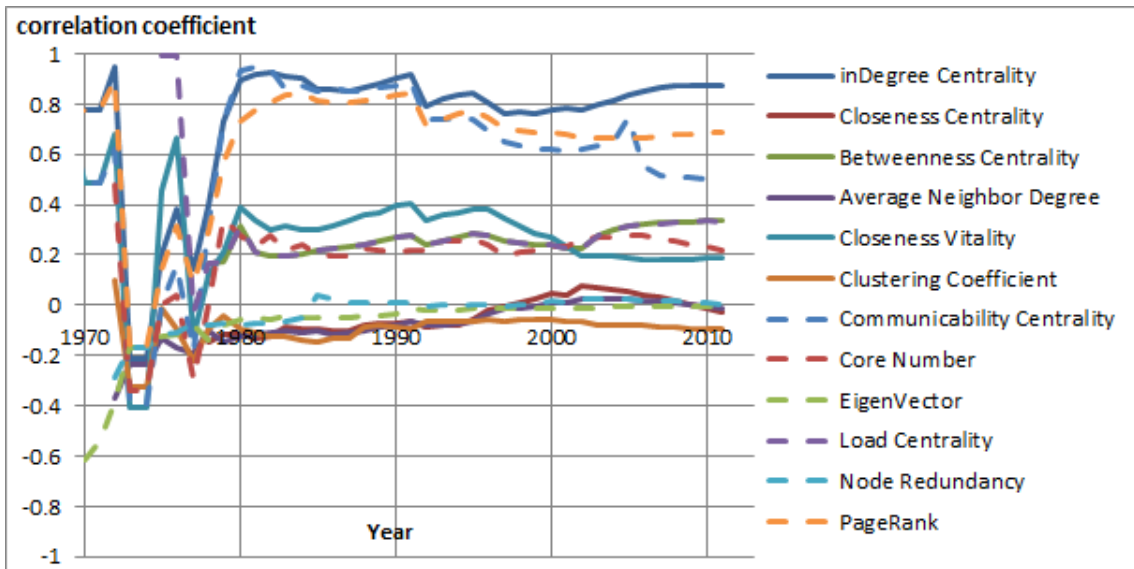


Figure 27 : Correlation Coefficient between Current Number of Citations and Indicators [crypto]

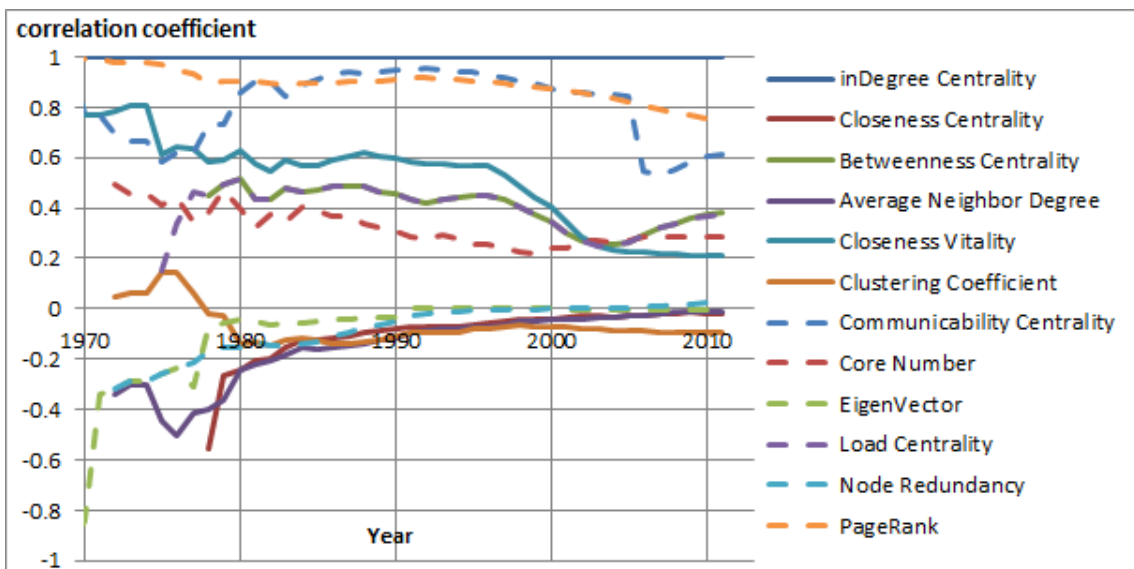


Figure 28 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [crypto]

3.2. 蛍光プロテイン(2008年ノーベル化学賞 [GFP])

本節では、2008年のノーベル化学賞より、蛍光プロテインについて、指標を時系列で図示することにより、有効な指標を選定する。

3.2.1. 分析対象分野の概要

2008年ノーベル賞受賞者の1人である下村脩氏は、1962年、海の中で発光クラゲであるオワンクラゲから緑色蛍光タンパク質(GFP)を発見した。さらに、彼は自ら発光するタンパク質を"イクオリン"と名付け、また、pH値およびカルシウムイオン(Ca^{2+})濃度でGFPの輝度を制御することが可能であるという事象を発見した。GFPはイクオリンから青色光を受け取って緑色の光を発する。下村氏の発見の後、蛍光タンパク質より発光タンパク質の方が、主流の研究テーマであった。しかし、下村氏の発見から30年後、Chalfie氏は、1993年に他の生き物へのGFP導入を実現した。Tsien氏は、1998年で、様々な色の蛍光タンパク質を作ることに成功した。

蛍光タンパク質は、生命科学分野で生命の詳細を知るのに役に立っている。これまでの生命活動を究明する際には解剖を行い死んだ細胞を観察していたが、蛍光タンパク質のおかげで、生きた細胞内の個々のタンパク質をトレースすることができるようになった。異なる色の蛍光タンパク質を使うことで、複数のタンパク質の関係を調べるために活用することができる。例えば、蛍光タンパク質は、癌の広がりやニューロンの構造を明らかにするために使用される(石浦, 2009)。

蛍光タンパク質の分析をするにあたって、受賞の際に発表される資料から、以下の論文をに貢献した論文として抽出した。

- Shimomura, O., Johnson, F.H. & Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J. Cell. Comp. Physiol.* 59: 223–239. [WOS:A19624094B00012]
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., & Prasher, D. C. (1994). Green fluorescent protein as a marker for gene-expression. *SCIENCE*, 263(5148), 802–805. [WOS:A1994MV86600033]

- Tsien, R.Y. (1998). The green fluorescent protein. *Annual Review of Biochemistry*, 67, pp.509-544. [WOS:000075721700018]

Thomson Reuters 社の論文サービス Web of Science より、「Topic=("fluorescent protein*" OR "bioluminescent protein*" OR "luminescent protein*" OR "photoprotein*") AND Year Published=(1900-2011)」というクエリーで、タイトル、アブストラクト、キーワード、キーワード・プラスを検索し、2011 年末時点で 32,439 論文の書誌情報を得た。そのうち、引用ネットワークを形成可能な 28,926 論文の書誌情報を対象に、複数の指標を各年について算出した。このクエリーを決定するにあたっては、文献調査により、下村氏の蛍光プロテイン発行当初の状況を考慮して、蛍光プロテインと発光プロテインを示す一般語を選定した。Figure 29、および、Table 7 に、対象となる論文の出版年の分布と、書誌情報を収集した 2012-11-16 時点での被引用数(以降、書誌情報収集時点での被引用数を最新被引用数とする)の分布を示す。この分野の論文は、2008 年にその年の論文数がピークの 2,500 本に達して、その後は微減傾向にある。2008 年はノーベル賞受賞年でもあり、その影響も多少あることは考えられるが、ノーベル賞の発表が 10 月であり、前年がそれまでの中で論文数がピークであることから、ノーベル賞受賞の影響はわずかだと考えられる。

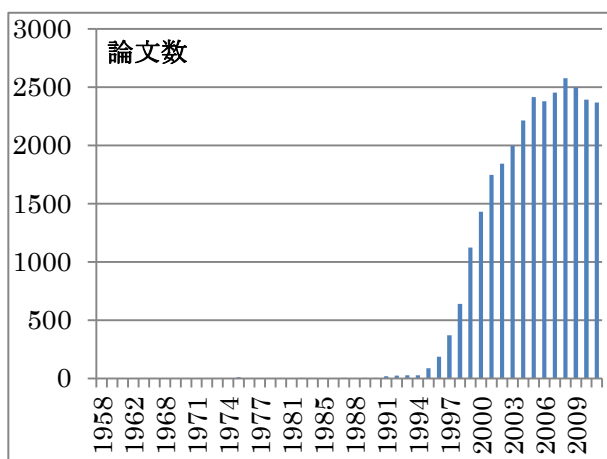


Figure 29: 出版年ごとの論文数の分布 [gfp]

Table 7: 最新被引用数ごとの論文数の分布 [gfp]

Times Cited	Count
0	1,203
1 - 100	25,756
101 - 200	1,368
201 - 300	349
301 - 400	112
401 - 500	54
501 - 600	21
601 - 700	15
701 - 800	5
801 - 900	11
901 - 1000	12
1001 - 2000	15
2001 - 3000	4
3001 - 4000	1

これらの論文の 2011 年末時点での引用ネットワークは Figure 30 のように形成される。クラスター分析を行った結果の各クラスターは各色で示され、TOP 7 クラスターの内容を Table 8 に、クラスター間のキーワード類似度を Figure 31 に示す。

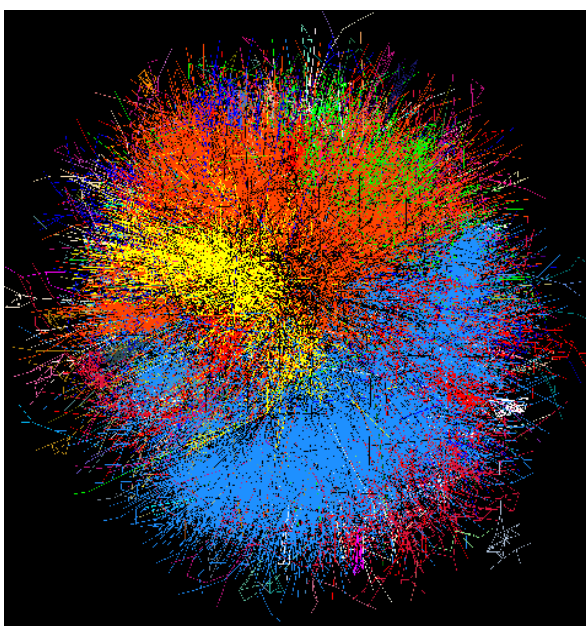


Figure 30 : Academic Landscape Map of Fluorescent Protein [gfp]

	1	2	3	4	5	6	7
1	1.00	0.00	0.04	0.20	0.06	0.06	0.02
2	0.00	1.00	0.03	0.12	0.09	0.04	0.07
3	0.04	0.03	1.00	0.00	0.00	0.00	0.03
4	0.20	0.12	0.00	1.00	0.02	0.11	0.01
5	0.06	0.09	0.00	0.02	1.00	0.00	0.00
6	0.06	0.04	0.00	0.11	0.00	1.00	0.01
7	0.02	0.07	0.03	0.01	0.00	0.01	1.00

Figure 31 : Heatmap of Cosine Similarity between Top 7 Clusters [gfp]

Table 8 : Information of TOP 7 Clusters [gfp]

	Ave.PY*	Nodes	Edges	E/N*	Top Keywords
All	2005.2	28,926	170,387	5.89	
#1	2004.0	8,354	41,497	4.97	plant, arabidopsis, actin, microtubule, root
#2	2005.6	7,481	25,740	3.44	neuron, stem cell, receptor, mouse
#3	2006.2	5,562	55,516	9.98	chromophore, fret, imaging, energy transfer
#4	2005.5	2,366	9,330	3.94	virus, viral, plant, infection, movement
#5	2004.9	772	1,314	1.70	elegans, autophagy, peroxisomal, neuron, mitochondrial
#6	2005.3	742	2,148	2.89	parasite, plasmodium, mosquito, malaria, leishmania
#7	2006.0	533	869	1.63	ultrasound, transfection, pei, gene delivery

*Ave.PY: Average year that papers are published(平均出版年)

E/N: Density that number of edges is divided by number of papers(ノードに対するエッジの密度)

3.2.2. 各指標の時系列分析

本項では、各指標の時系列分析を使って、萌芽的論文を検出するのに有効な指標について、詳細を解説する。

Table 9 は、この引用ネットワークで最新被引用数が多い順から TOP 10 の論文リストである。Figure 32 は、5年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が5年後のクラスターに遷移するかを、エッジの太さで表した図である。Figure 33 は、5年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が5年後のクラスターに遷移するかを、エッジの太さで表した図である。ノードのサイズがそのクラスターの所属する論文数に比例し、クラスターの8以降は煩雑になるため省略している。2008年にノーベル化学賞を受賞した下村氏(1984)、Chalfie氏(1994)、Tsien氏(1998)の論文がクラスター上を辿った軌跡はそれぞれ、オレンジ、グリーン、アクアの破線で示す。

Table 9 : TOP 10 and the Nobel Laureate's Papers on 2012-11-16 [gfp]

	R*	1st Author	PY*	TC*	Cluster Number								
					1975	1980	1985	1990	1995	2000	2005	2010	2011
#	1	Chalfie, M	1994	3,627	-	-	-	-	2	1	1	2	1
	2	Orlic, D	2001	2,824	-	-	-	-	-	-	2	1	2
#	3	Tsien, RY	1998	2,761	-	-	-	-	-	3	1	2	3
	4	He, TC	1998	2,469	-	-	-	-	-	11	6	1	2
	5	Longtine, MS	1998	2,267	-	-	-	-	-	2	3	3	1
	6	Cormack, BP	1996	1,748	-	-	-	-	-	1	1	2	1
	7	Huh, WK	2003	1,705	-	-	-	-	-	-	3	3	1
	8	Shaner, NC	2004	1,551	-	-	-	-	-	-	1	2	3
	9	Okabe, M	1997	1,515	-	-	-	-	-	1	2	1	3
	10	Miyawaki, A	1997	1,475	-	-	-	-	-	3	1	2	3
#	-	Shimomura, O	1962	841	2	3	1	1	1	3	2	2	3

* R : Rank based on TC, PY : Published Year(出版年), TC : Times Cited(最新被引用数), #: 受賞論文

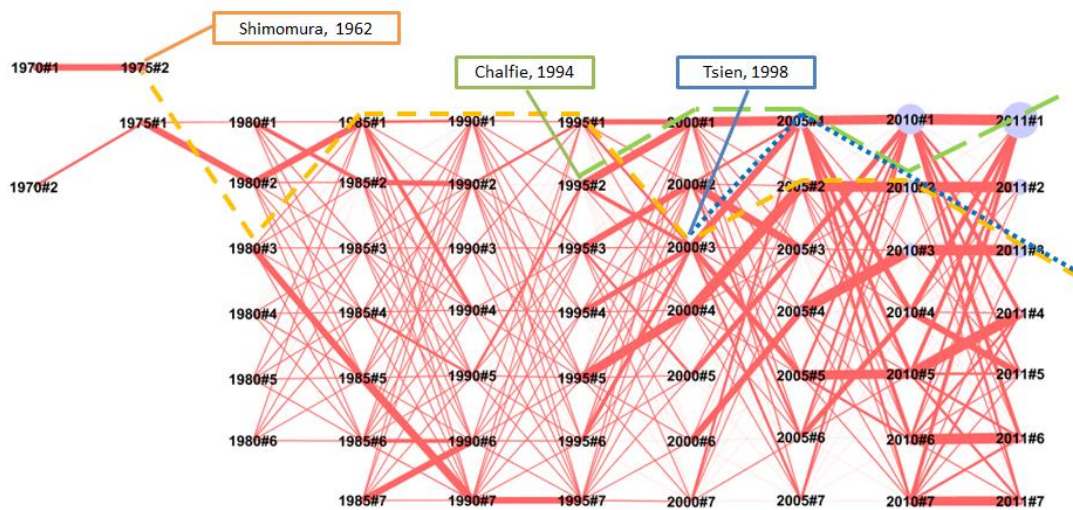


Figure 32: キーワードに基づくクラスター間の類似度 [gfp]

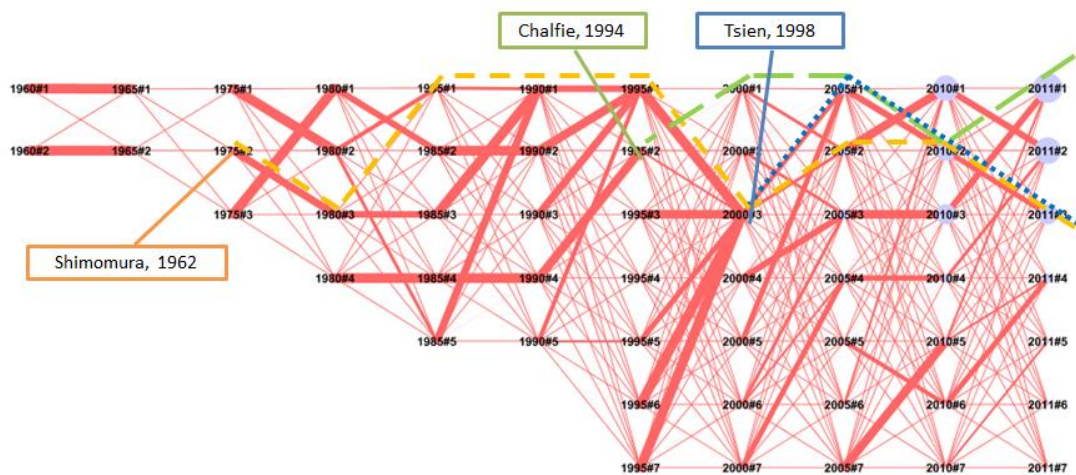


Figure 33: 論文の遷移量に基づくクラスター間の類似度 [gfp]

Figure 34 は各年の被引用数、Figure 35 は in-Degree Centrality、Figure 36 は PageRank の時系列変化であり、データセット内で引用ネットワークで計算を行っている。Figure 34、Figure 35、Figure 36 ではいずれも、最新被引用数が 100 を越える 1,967 論文について表示しており、各線は最新被引用数で 5 等分して多い順に青、アクア、グリーン、オレンジ、赤の着色を行っている。青線は最新被引用数が最も多い Chalfie 氏の論文(1994)であり、アクアの線は Table 9

の2位から4位までの論文であるが、その中で青線に続いて集団から飛び出しているアクアの線は Tsien 氏の論文(1998)である。



Figure 34 : Time Transition about Number of within-Dataset Citation of 1,967 Papers having more than 100 citations (最大値=2371) [gfp]

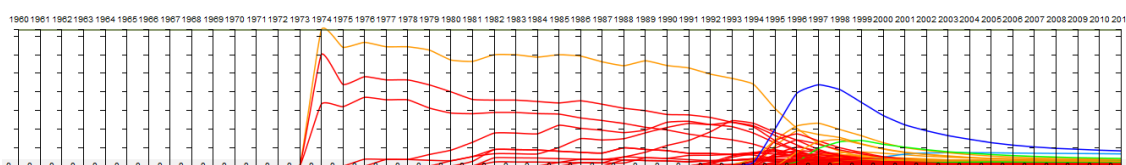


Figure 35 : Time Transition about in-Degree Centrality of 1,967 Papers having more than 100 citations (最大値=0.7333) [gfp]

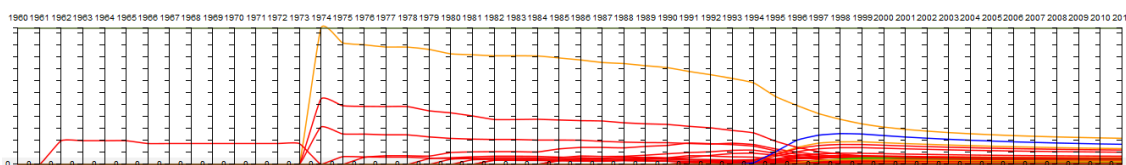


Figure 36 : Time Transition about PageRank of 1,967 Papers having more than 100 citations (最大値=0.2276) [gfp]

以上から、in-Degree Centrality と PageRank において萌芽的論文は発行年から数年経つと高い数値を持つ。

どの指標が有効であるかを明らかにするために、Figure 37 に最新被引用数と指標の相関係数を、Figure 38 に各年のデータセット内の被引用数と指標の相関係数を、年ごとに計算した。Figure 38 では、被引用数の正規化した値である in-Degree Centrality は相関係数が常にほぼ1であり、次いで、Betweenness Centrality、Core Number、Load Centrality、PageRank が年々相関が低くなっていくものの、データセット内の被引用数に相関を示している。

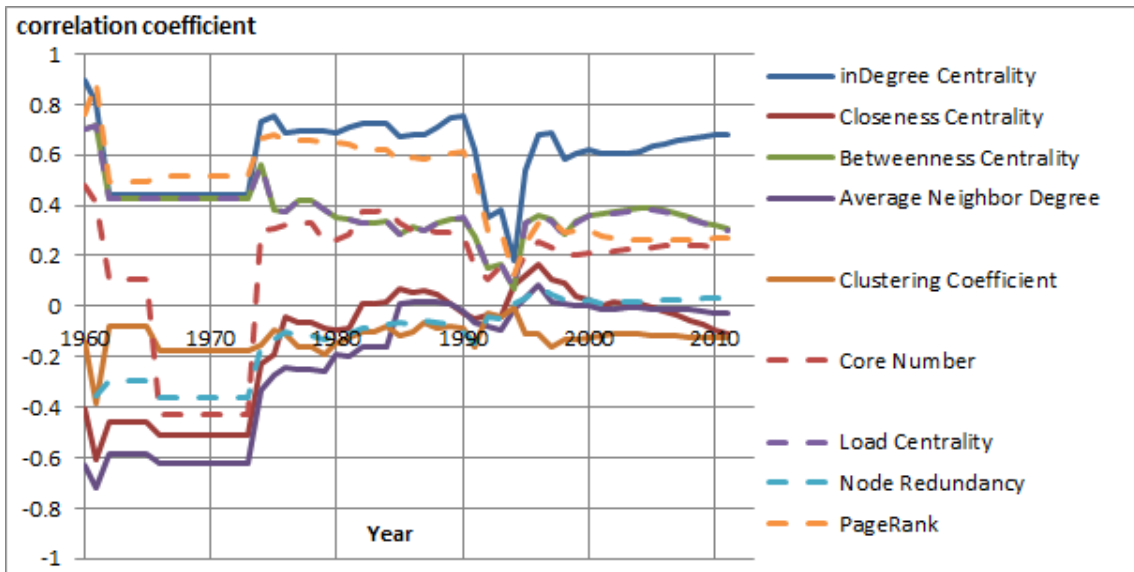


Figure 37 : Correlation Coefficient between Current Number of Citations and Indicators [gfp]

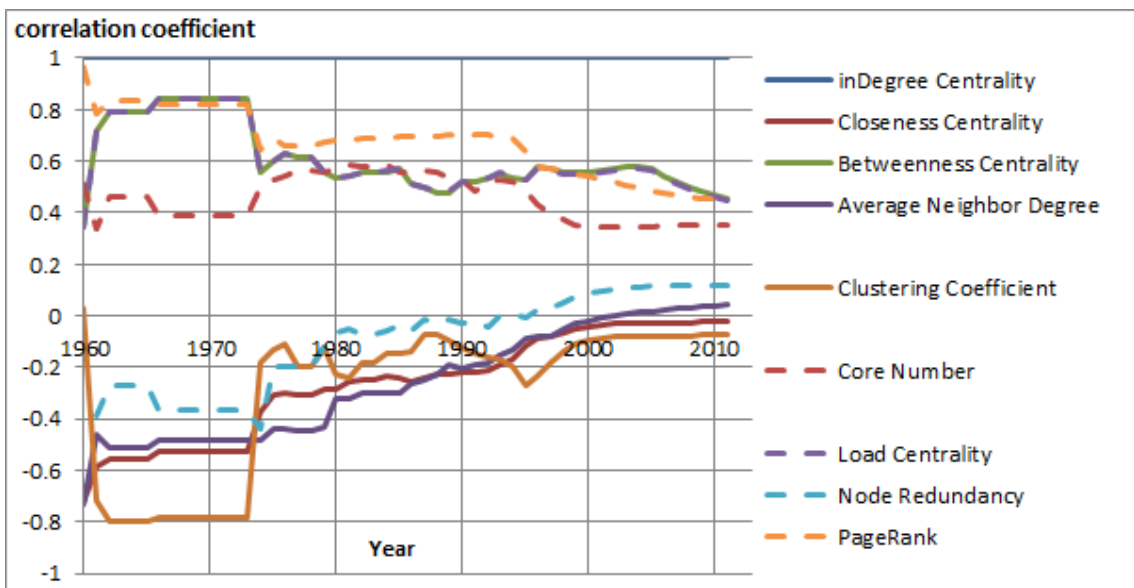


Figure 38 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [gfp]

3.3. G タンパク質共益受容体 (2012 年ノーベル化学賞 [2012CHEM])

本節では、2012 年のノーベル化学賞より、G タンパク質共益受容体(G-protein-coupled receptors、GPCR)について、指標を時系列で図示することにより、有効な指標を選定する。

3.3.1. 分析対象分野の概要

細胞内部に外部から情報を伝達するのに細胞膜表面に存在するタンパク質が情報仲介の役割を担っており、このようや役割を行うタンパク質を受容体と呼ぶが、複数の細胞膜受容体があるうち、G タンパク質共益受容体とは、もっとも多くを占める細胞膜受容体である。G タンパク質共益受容体の研究に対して、Robert J. Lefkowitz 氏と Brian K. Kobilka 氏に 2012 年のノーベル化学賞が授与された。

G タンパク質共益受容体の分析をするにあたって、受賞の際に発表される資料から、以下の論文に貢献した論文として抽出した。

- Lefkowitz RJ, Roth J, Pricer W, Pastan I (1970a) ACTH receptors in the adrenal: specific binding of ACTH-125I and its relation to adenylyl cyclase. Proc Natl Acad Sci USA65, 745-752. [WOS:A1970F821700040]
- Lefkowitz RJ, Roth J, Pastan I (1970b) Radioreceptor assay for adrenocorticotropic hormone: new approach to assay of polypeptide hormones in plasma. Science 170, 633-635. [WOS:A1970H672100027]
- De Lean A, Stadel JM, Lefkowitz RL (1980) A ternary complex model explains the agonistspecific binding properties and the adenylate cyclase-coupled beta-adrenergic receptor. J Biol Chem 255, 7108-7117. [WOS:A1980KD39000017]
- Dixon RA, Kobilka BK, Strader DJ, Benovic JL, Dohlman HG, Frielle T, Bolanowski MA, Bennet CD, Rands E, Diehl RE, Mumford RA, Slater EE, Sigal IS, Caron MG, Lefkowitz RJ, Strader CD (1986) Cloning of the gene and cDNA for mammalian beta-adrenergic receptor: primary structure and membrane topology. Nature 321, 75-79. [WOS:A1986C169000052]

- Kobilka KB, Frielle T, Dohlman HG, Bolanowski MA, Dixon RA, Keller P, Caron MG, Lefkowitz RJ (1987) Delineation of the intronless nature of the genes for the human and hamster beta 2-adrenergic receptor and their putative promoter regions. J Biol Chem 262, 7321-7327. [WOS:A1987H414700060]

Thomson Reuters 社の論文サービス Web of Science より、「Topic=("G-protein-coupled receptor*" OR "ACTH receptor*" OR radioreceptor* OR "adrenergic receptor*") AND Year Published=(1900-2012)」というクエリーで、タイトル、アブストラクト、キーワード、キーワード・プラスを検索し、54,774 論文の書誌情報を得た。そのうち、2012 年末時点で引用ネットワークを形成可能な 46,421 論文の書誌情報を対象に、複数の指標を各年について算出した。Figure 39、および、Table 10 に、対象となる論文の出版年の分布と、書誌情報を収集した 2013-03-27 時点での被引用数(以降、最新被引用数とする)の分布を示す。

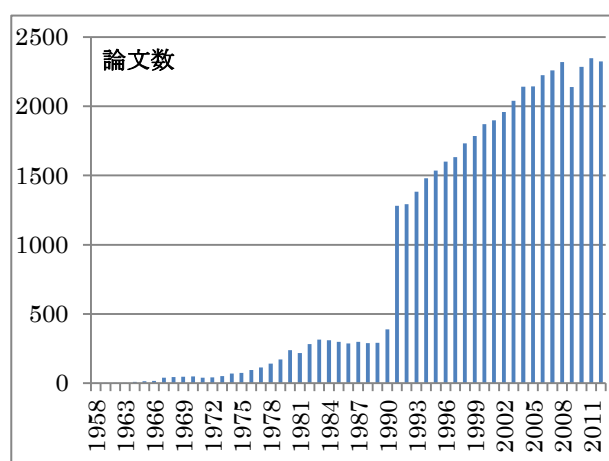


Figure 39: 出版年ごとの論文数の分布 [2012chem]

Table 10: 最新被引用数ごとの論文数の分布 [2012chem]

最新被引用数	論文数
0	2,807
1 - 100	39,557
101 - 200	2,338
201 - 300	593
301 - 400	271
401 - 500	138
501 - 600	68
601 - 700	41
701 - 800	38
801 - 900	23
901 - 1000	9
1001 - 2000	30
2001 - 3000	6
3001 - 4000	2

これらの論文の 2012 年末時点での引用ネットワークは Figure 40 のように形成される。クラスター分析を行った結果の各クラスターは各色で示され、TOP 7 クラスターの内容を Table 11 に、クラスター間のキーワード類似度を Figure 41 に示す。クラスターのキーワード類似度は、Figure 41 のように、クラスター 2 と 3、および、1 と 3 の類似度が高い。

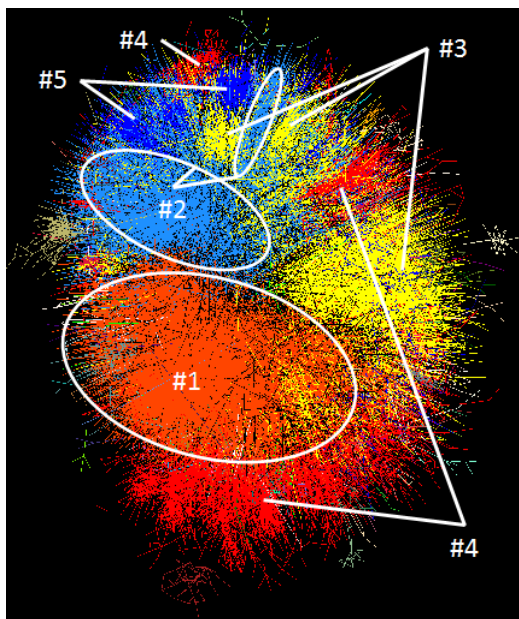


Figure 40 : Academic Landscape Map of G-protein-coupled receptors [2012chem]

	1	2	3	4	5	6	7
1	1.00	0.12	0.29	0.04	0.04	0.10	0.04
2	0.12	1.00	0.34	0.06	0.23	0.02	0.05
3	0.29	0.34	1.00	0.04	0.15	0.00	0.10
4	0.04	0.06	0.04	1.00	0.03	0.02	0.01
5	0.04	0.23	0.15	0.03	1.00	0.00	0.02
6	0.10	0.02	0.00	0.02	0.00	1.00	0.00
7	0.04	0.05	0.10	0.01	0.02	0.00	1.00

Figure 41 : Heatmap of Cosine Similarity between Top 7 Clusters [2012chem]

Table 11 : Information of TOP 7 Clusters [2012chem]

	Ave.PY*	Nodes	Edges	E/N*	Top Keywords
All	2000.8	46,421	505,929	10.90	
#1	1998.6	15,675	155,956	9.95	arrestin, desensitization, polymorphism, heart, kinase
#2	2003.6	13,328	166,012	12.46	protein coupled receptor
#3	2001.9	11,700	82,895	7.09	alpha, lpa, slp, cell, kinase
#4	1990.9	2,050	6,613	3.23	acth receptor, melanocortin, adrenal, igf
#5	2004.9	1,450	9,365	6.46	melatonin, apelin, kisspeptin, gpr54, pineal
#6	2004.1	267	1,578	5.91	autoantibody, antibody, dilated cardiomyopathy, dcm
#7	1999.0	224	484	2.16	rgr, ca2+, retinal, inositol, trisphosphate

*Ave.PY: Average year that papers are published(平均出版年)

E/N: Density that number of edges is divided by number of papers(ノードに対するエッジの密度)

3.3.2. 各指標の時系列分析

本項では、各指標の時系列分析を使って、萌芽的論文を検出するのに有効な指標について、詳細を解説する。

Table 12 は、この引用ネットワークで最新被引用数が多い順から TOP 10 の論文リストである。Figure 42 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。Figure 43 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。ノードのサイズがそのクラスターの所属する論文数に比例し、クラスターの 8 以降は煩雑になるため省略している。2012 年にノーベル化学賞を受賞した Lefkowitz 氏、Kobilka 氏の論文がクラスター上を辿った軌跡はそれぞれ、オレンジ、黄、グリーン、アクア、青の線で示す。

Table 12 : TOP 10 and the Nobel Laureate's Papers on 2013-03-27 [2012chem]

R*	1 st Author	PY*	TC*	1975	1980	1985	1990	1995	2000	2005	2010	2012
1	Kojima, M	1999	3,850	-	-	-	-	-	3	2	3	2
2	Palczewski, K	2000	3,470	-	-	-	-	-	2	2	3	2
3	Feng, Y	1996	2,946	-	-	-	-	-	3	2	2	2
4	Vu, TKH	1991	2,390	-	-	-	-	2	3	4	7	3
5	Munro, S	1993	2,294	-	-	-	-	2	6	2	3	2
6	Sakurai, T	1998	2,178	-	-	-	-	-	3	2	3	2
7	Buck, L	1991	2,156	-	-	-	-	2	2	2	3	2
8	Alkhatib, G	1996	2,039	-	-	-	-	-	3	2	2	2
9	Baggiolini, M	1994	1,994	-	-	-	-	2	3	2	2	3
10	Conn, PJ	1997	1,938	-	-	-	-	-	2	2	3	2
#	Lefkowitz, RJ	1970a	326	4	2	2	3	1	1	2	3	4
#	Lefkowitz, RJ	1970b	127	4	2	2	3	5	1	2	5	4
#	De Lean, A	1980	1,287	-	2	2	3	1	2	1	2	2
#	Dixon, RA	1986	974	-	-	-	5	2	2	1	1	2
#	Kobilka, KB	1987	237	-	-	-	2	2	4	1	1	1

* R : Rank based on TC, PY : Published Year(出版年), TC : Times Cited(最新被引用数), #: 受賞論文

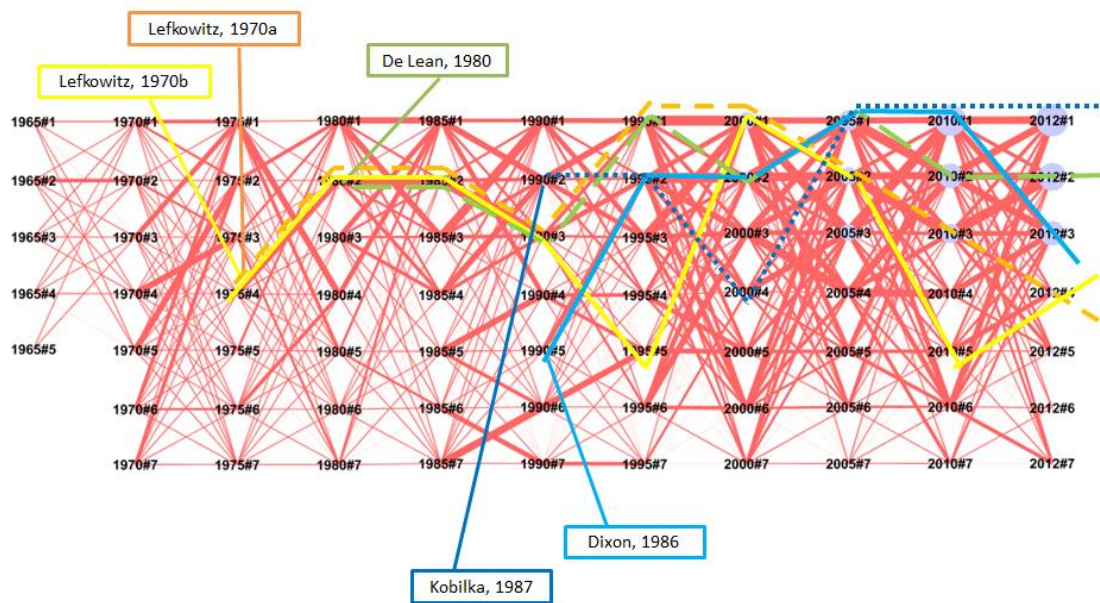


Figure 42: キーワードに基づくクラスター間の類似度 [2012chem]

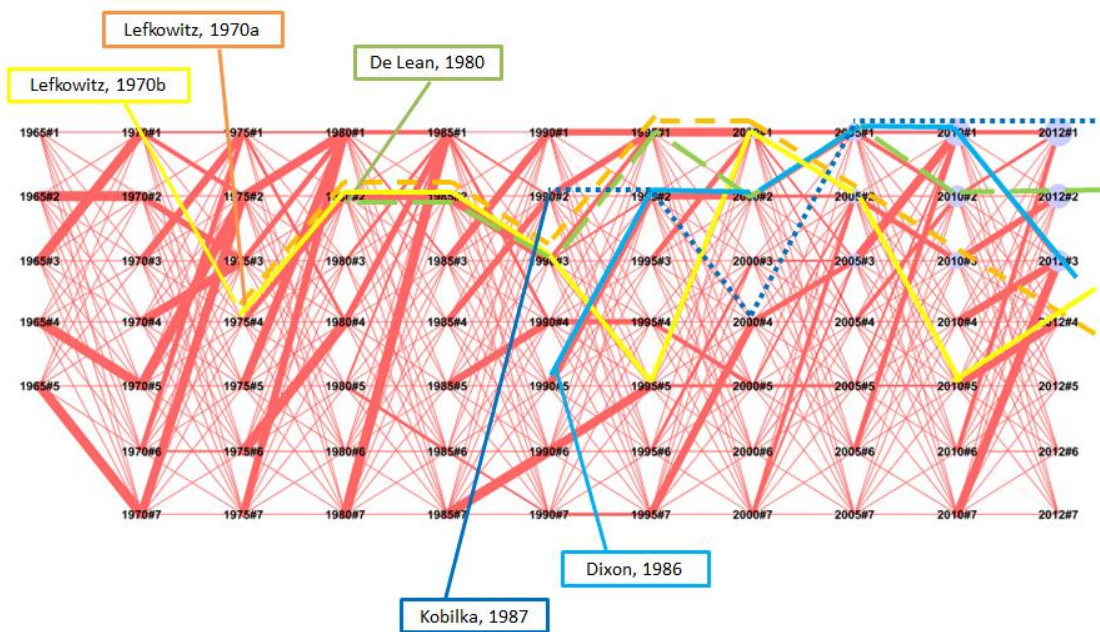


Figure 43: 論文の遷移量に基づくクラスター間の類似度 [2012chem]

Figure 44 は各年の被引用数、Figure 45 は in-Degree Centrality、Figure 46 は PageRank の時系列変化であり、データセット内で引用ネットワークで計算を行っている。Figure 44、Figure

45、Figure 46 ではいずれも、最新被引用数が 100 を越える 3,557 論文について表示しており、各線は最新被引用数で 5 等分して多い順に青、アクア、グリーン、オレンジ、赤の着色を行っている。

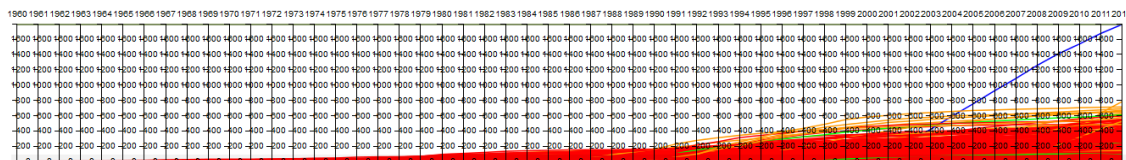


Figure 44 : Time Transition about Number of within-Dataset Citation of 3,557 Papers having more than 100 citations (最大値=1792) [2012chem]



Figure 45 : Time Transition about in-Degree Centrality of 3,557 Papers having more than 100 citations (最大値=1) [2012chem]

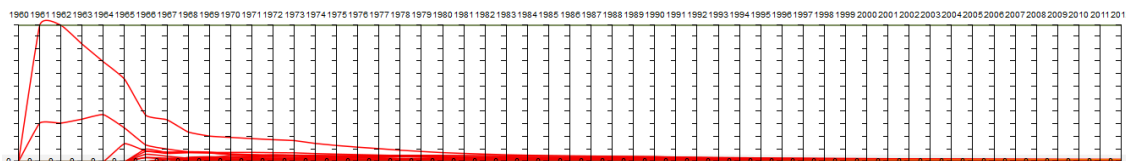


Figure 46 : Time Transition about PageRank of 3,557 Papers having more than 100 citations (最大値=1) [2012chem]

どの指標が有効であるかを明らかにするために、Figure 47 に最新被引用数と指標の相関係数を、Figure 48 に各年のデータセット内の被引用数と指標の相関係数を、年ごとに計算した。Figure 48 では、被引用数の正規化した値である in-Degree Centrality は相関係数が常にほぼ 1 であり、次いで、Betweenness Centrality、Core Number、Load Centrality、PageRank もデータセット内の被引用数に弱い相関を示している。

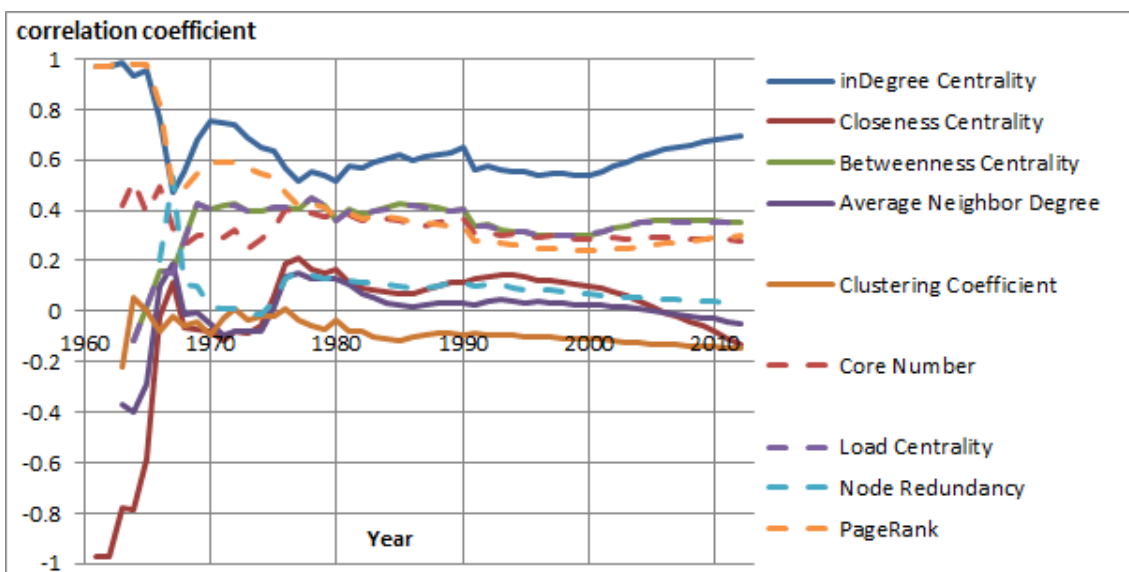


Figure 47 : Correlation Coefficient between Current Number of Citations and Indicators [2012chem]

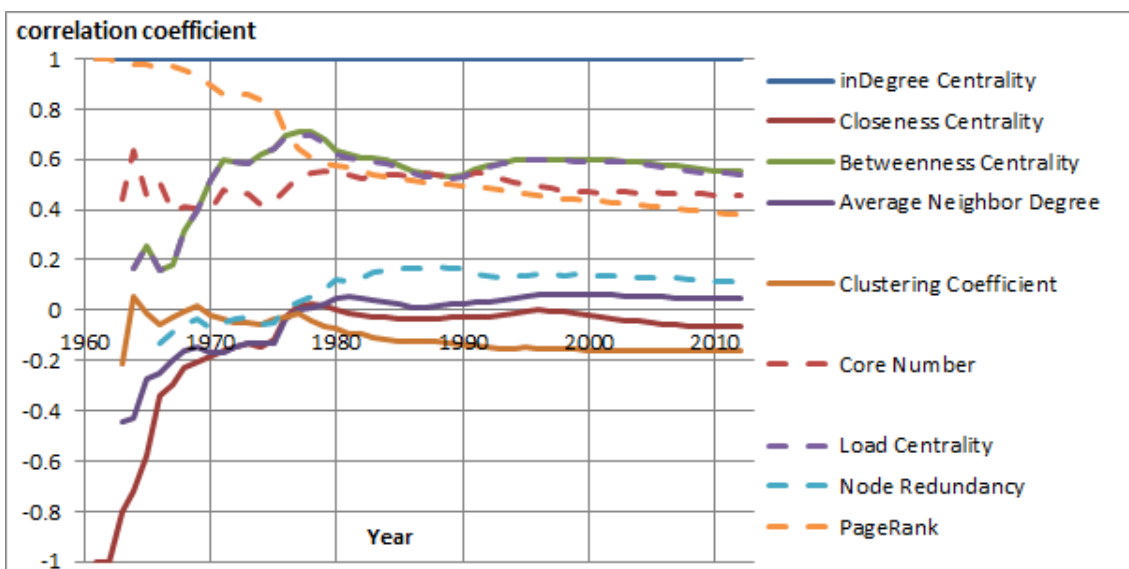


Figure 48 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [2012chem]

3.4. IPS 細胞(2012 年ノーベル生理学・医学賞 [2012MED])

本節では、2012 年のノーベル生理学・医学賞より、成熟細胞が再プログラムされ多能性を獲得する発見について、指標を時系列で図示することにより、有効な指標を選定する。

3.4.1. 分析対象分野の概要

1962 年、John B. Gurdon 氏がカエルの体細胞核移植に成功し、この技術が、2007 年に生理学・医学賞を序章した ES 細胞の研究や、2012 年に Gurdon 氏と同時受賞した山中伸弥氏の研究である iPS 細胞に繋がった。細胞は、発生初期においてどの機能にも分化する多能性を持つが、山中氏は、成熟細胞が多能性を持つ状態に戻す因子を特定した。これらの成果から、John Bertrand Gurdon 氏と山中伸弥氏は 2012 年のノーベル生理学・医学賞を授与された(Newton Press, 2012)。

成熟細胞が再プログラムされ多能性を獲得する発見の分析をするにあたって、受賞の際に発表される資料から、以下の論文をに貢献した論文として抽出した。

- Gurdon JB (1962). Developmental Capacity of Nuclei Taken from Intestinal Epithelium Cells of Feeding Tadpoles. *J Embryol Exp Morph* 10: 622-640. [WOS:A19624723B00005]
- Takahashi K, Yamanaka S (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126: 663-676. [WOS:000240276700016]

Thomson Reuters 社の論文サービス Web of Science より、「Topic=("mature cell*" OR reprogram* OR pluripotent OR "developmental capacity of nuclei" OR "developmental capacity of nucleus") AND Year Published=(1900-2012)」というクエリーで、タイトル、アブストラクト、キーワード、キーワード・プラスを検索し、18,538 論文の書誌情報を得た。そのうち、2012 年末時点で引用ネットワークを形成可能な 12,513 論文の書誌情報を対象に、複数の指標を各年について算出した。Figure 49、および、Table 13 に、対象となる論文の出版年の

分布と、書誌情報を収集した 2013-03-27 時点での被引用数(以降、最新被引用数とする)の分布を示す。

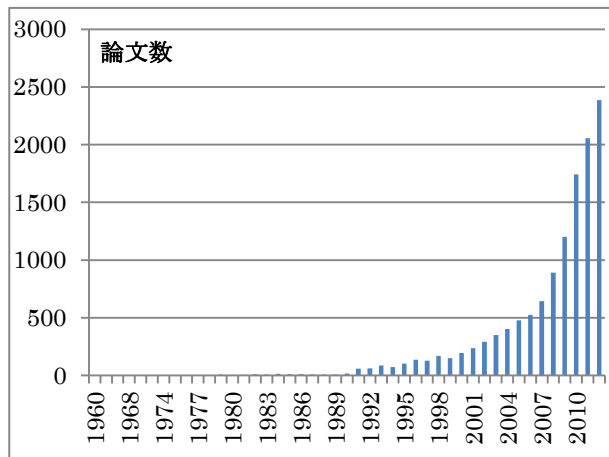


Figure 49: 出版年ごとの論文数の分布 [2012med]

Table 13: 最新被引用数ごとの論文数の分布 [2012med]

最新被引用数	論文数
0	1,649
1 - 100	9,800
101 - 200	674
201 - 300	176
301 - 400	80
401 - 500	53
501 - 600	28
601 - 700	12
701 - 800	8
801 - 900	5
901 - 1000	3
1001 - 2000	14
2001 - 3000	8
3001 - 4000	1
4001 - 5000	1
5001 - 6000	1

これらの論文の 2012 年末時点での引用ネットワークは Figure 50 のように形成される。クラスター分析を行った結果の各クラスターは各色で示され、TOP 7 クラスターの内容を Table 14 に、クラスター間のキーワード類似度を Figure 51 に示す。

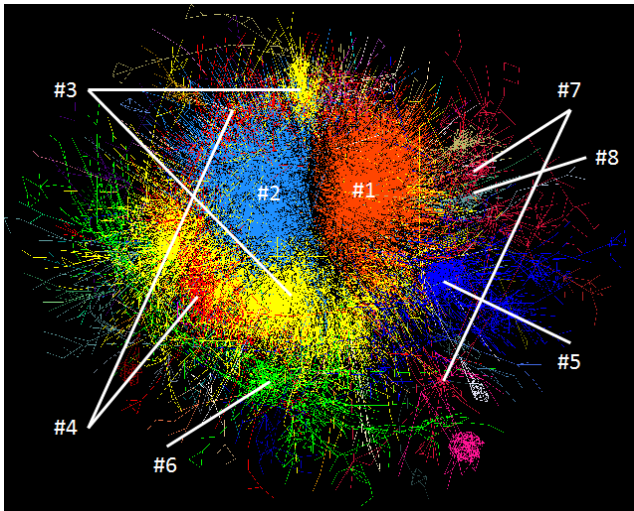


Figure 50 : Academic Landscape Map of iPS cells [2012med]

	1	2	3	4	5	6	7
1	1.00	0.26	0.14	0.06	0.02	0.00	0.00
2	0.26	1.00	0.12	0.06	0.00	0.00	0.00
3	0.14	0.12	1.00	0.01	0.03	0.00	0.00
4	0.06	0.06	0.01	1.00	0.00	0.00	0.00
5	0.02	0.00	0.03	0.00	1.00	0.03	0.05
6	0.00	0.00	0.00	0.00	0.03	1.00	0.03
7	0.00	0.00	0.00	0.00	0.05	0.03	1.00

Figure 51 : Heatmap of Cosine Similarity between Top 7 Clusters [2012med]

Table 14 : Information of TOP 7 Clusters [2012med]

	Ave.PY*	Nodes	Edges	E/N*	Top Keywords
All	2007.5	12,513	119,858	9.58	
#1	2010.6	3,622	49,658	13.71	ipscs, induced pluripotent stem, somatic cell
#2	2007.1	3,560	27,529	7.73	germ cell, embryonic stem cell
#3	2004.9	2,625	12,880	4.91	oocyte, embryo, nuclear transfer, methylation, somatic
#4	2006.6	663	1,179	1.78	osteoblast, bmp, mesenchymal, msc, bone
#5	2009.0	451	800	1.77	hif, cancer, metabolism, metabolic, hypoxia
#6	2007.6	337	602	1.79	plant, meristem, arabidopsis, leaf, shoot
#7	2007.8	233	372	1.60	trans splicing, yeast, cerevisiae

*Ave.PY: Average year that papers are published(平均出版年)

E/N: Density that number of edges is divided by number of papers(ノードに対するエッジの密度)

3.4.2. 各指標の時系列分析

本項では、各指標の時系列分析を使って、萌芽的論文を検出するのに有効な指標について、詳細を解説する。

Table 15 は、この引用ネットワークで最新被引用数が多い順から TOP 10 の論文リストである。Figure 52 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。Figure 53 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。ノードのサイズがそのクラスターの所属する論文数に比例し、クラスターの 8 以降は煩雑になるため省略している。2012 年にノーベル生理学・医学賞を受賞した Gurdon 氏 (1962)、山中氏 (Takahashi, 2006) の論文がクラスター上を辿った軌跡はそれぞれ、オレンジおよびグリーンの破線で示す。

Table 15 : TOP 10 and the Nobel Laureate's Papers on 2013-03-27 [2012med]

	R*	1st Author	PY*	TC*	Cluster Number								
					1975	1980	1985	1990	1995	2000	2005	2010	2012
	1	Thomson, JA	1998	5,392	-	-	-	-	-	2	2	1	2
#	2	Takahashi, K	2006	4,706	-	-	-	-	-	-	-	3	1
	3	Takahashi, K	2007	3,786	-	-	-	-	-	-	-	3	1
	4	DeRisi, JL	1997	2,926	-	-	-	-	-	-	-	10	7
	5	Yu, JY	2007	2,873	-	-	-	-	-	-	-	3	1
	6	Karin, M	2000	2,598	-	-	-	-	-	-	-	6	5
	7	Lehmann, JM	1995	2,546	-	-	-	-	-	11	4	5	4
	8	Inaba, K	1992	2,379	-	-	-	-	-	15	12	6	8
	9	Martin, GR	1981	2,337	-	-	-	-	2	2	2	1	2
	10	Engler, AJ	2006	2,221	-	-	-	-	-	-	-	5	2
#	-	Gurdon, JB	1962	284	1	-	-	-	4	3	3	2	3

*R : Rank based on TC, PY : Published Year(出版年), TC : Times Cited(最新被引用数), #: 受賞論文

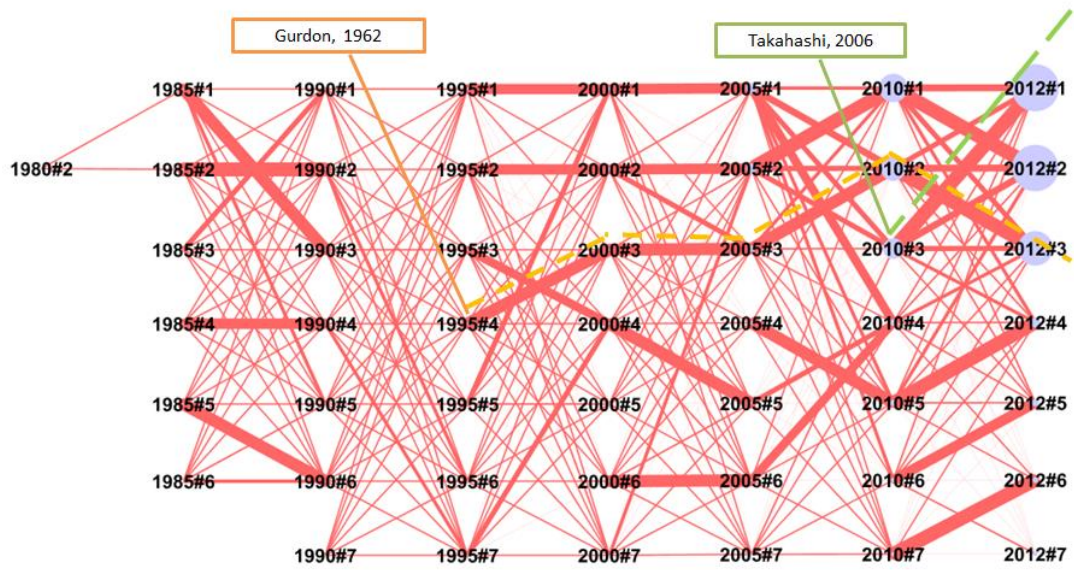


Figure 52: キーワードに基づくクラスター間の類似度 [2012med]

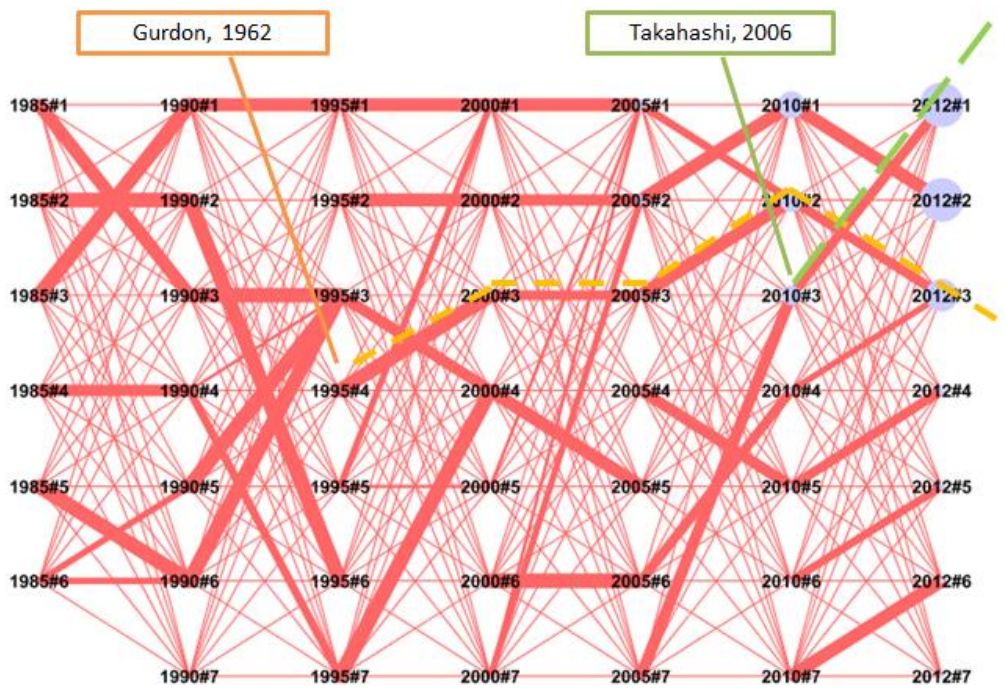


Figure 53: 論文の遷移量に基づくクラスター間の類似度 [2012med]

Figure 54 は各年の被引用数、Figure 55 は in-Degree Centrality、Figure 56 は PageRank の時系列変化であり、データセット内で引用ネットワークで計算を行っている。Figure 54、Figure 55、Figure 56 ではいずれも、最新被引用数が 100 を越える 1,064 論文について表示しており、各線は最新被引用数で 5 等分して多い順に青、アクア、グリーン、オレンジ、赤の着色を行っている。

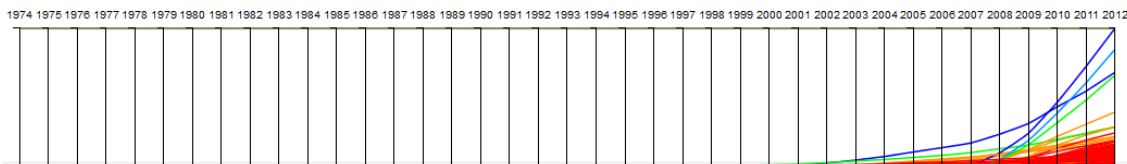


Figure 54 : Time Transition about Number of within-Dataset Citation of 1,064 Papers having more than 100 citations (最大値=2649, 最小値=0) [2012med]

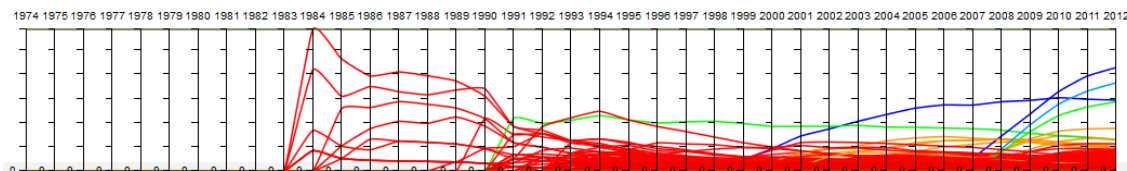


Figure 55 : Time Transition about in-Degree Centralities of 1,064 Papers having more than 100 citations (最大値=0.25) [2012med]

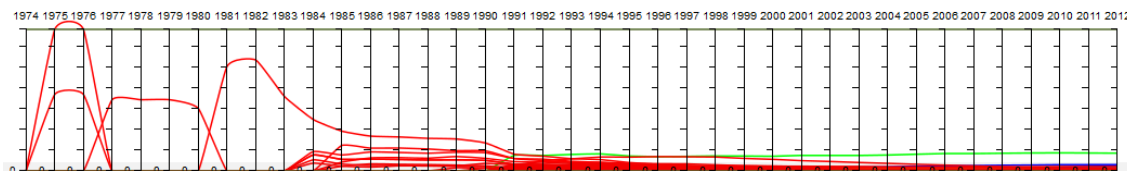
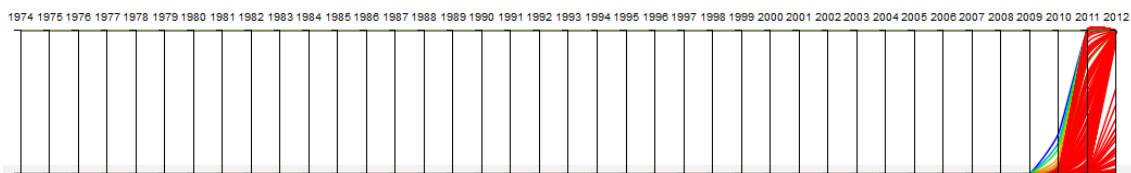


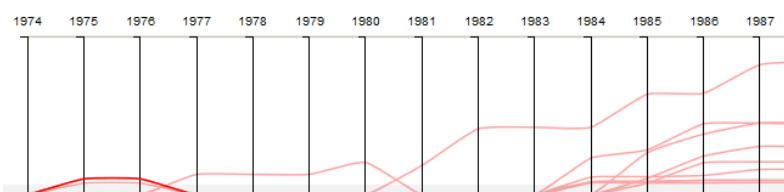
Figure 56 : Time Transition about PageRank of 1,064 Papers having more than 100 citations (最大値=0.3412) [2012med]

Figure 57 は、最新被引用数が 100 を越える論文について Communicability Centrality の時系列変化を描画したものである。この分野は、被引用数が相対的に少ない論文であっても Communicability Centrality の値が高く萌芽的論文が分離しないため、Communicability Centrality による萌芽的論文の検知は難しい。

(a) (最大値=3.403e+38、最小値=0)



(b) (最大値=21.51、最小値=0、フォーカスされた線はノーベル賞受賞者 Gurdon 氏の論文(1962))



(c) (最大値=3.464e+20)

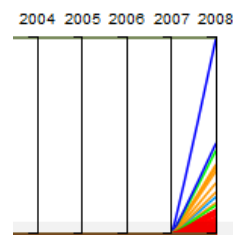


Figure 57 : Time Transition about Communicability Centralities of 1,064 Papers having more than 100 citations [2012med]

どの指標が有効であるかを明らかにするために、Figure 58 に最新被引用数と指標の相関係数を、Figure 59 に各年のデータセット内の被引用数と指標の相関係数を、年ごとに計算した。Figure 58 では、in-Degree Centrality、Communicability Centrality、PageRank は 1980 年以降は相関を高めているが、比較上、このデータセットではあまりよい相関ではない。Figure 59 では、被引用数の正規化した値である in-Degree Centrality は相関係数が常にほぼ 1 であり、Communicability Centrality が 2010 年まで良い相関を示していたが、その後、急速に悪化した。

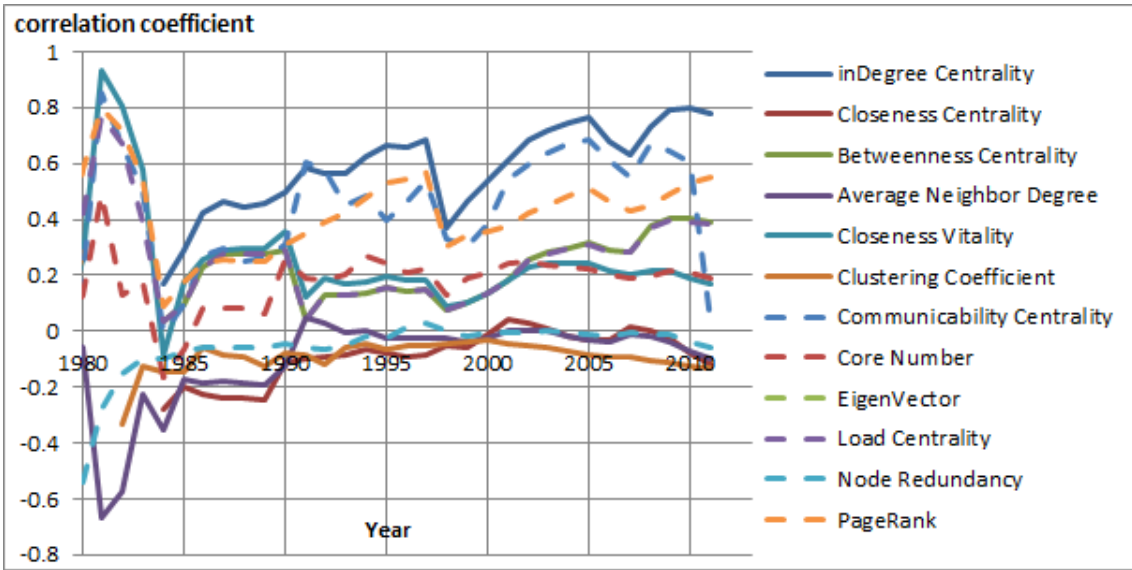


Figure 58 : Correlation Coefficient between Current Number of Citations and Indicators [2012med]

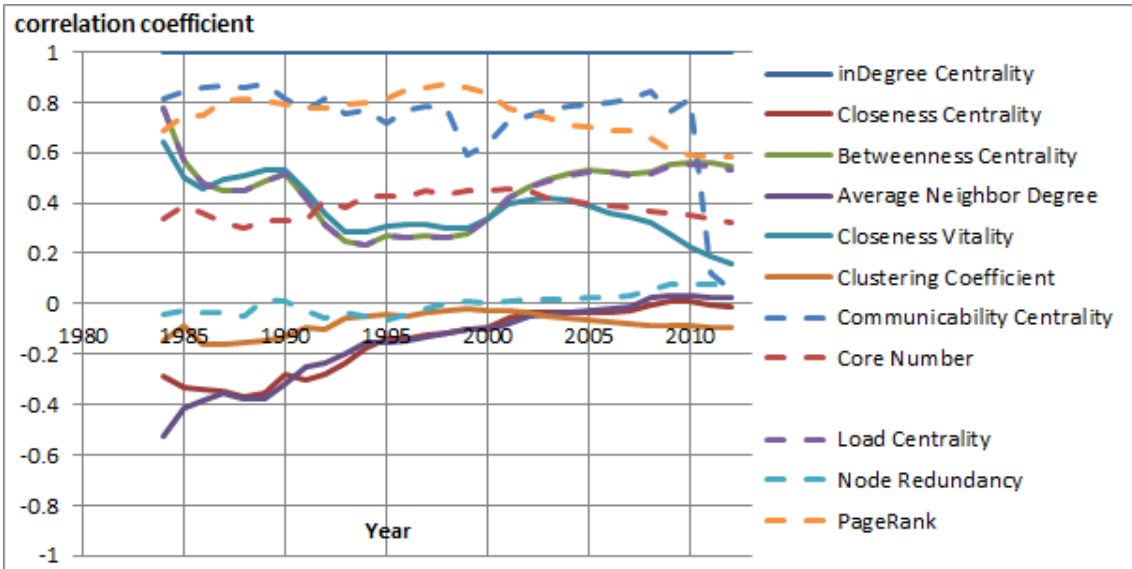


Figure 59 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [2012med]

3.5. 準結晶(2011年ノーベル化学賞 [2011CHEM])

本節では、2011年のノーベル化学賞より、準結晶(quasicrystal)について、指標を時系列で図示することにより、有効な指標を選定する。

3.5.1. 分析対象分野の概要

準結晶とは、従来、周期性を持つ結晶と非結晶(アモルファス構造)の固体に分けられると考えられていたところ、それらの中間状態の固体をいう(甲元, 1987)。準結晶は、Dan Shechtman氏により1982年に発見されたが、あまりにも革新的な発見であったため、当初は理解を示す研究者が少なく、論文も信憑性を疑われて不採択になったこともあり、実際に論文として発表されたのは1984年となった。このような経緯から、分野の起源の論文が明確な分野である(The Royal Swedish Academy of Sciences, 2011a; The Royal Swedish Academy of Sciences, 2011b)。

準結晶の分析をするにあたって、受賞の際に発表される資料から、以下の論文をに貢献した論文として抽出した。

- Shechtman, D., Blech, I., Gratias, D., and Cahn, J.W. (1984) Metallic phase with long-range orientational order and no translational symmetry, *Phys. Rev. Lett.* 53(20):1951-1954. [WOS:A1984TR73100024]

Thomson Reuters社の論文サービスWeb of Scienceより、「Topic=(quasicrystal* OR "metallic phase*") AND Year Published=(1900-2012)」というクエリーで、タイトル、アブストラクト、キーワード、キーワード・プラスを検索し、2012年末時点で9,335論文の書誌情報を得た。そのうち、引用ネットワークを形成可能な6,906論文の書誌情報を対象に、複数の指標を各年について算出した。Figure 60、および、Table 16に、対象となる論文の出版年の分布と、書誌情報を収集した2013-03-29時点での被引用数(以降、書誌情報収集時点での被引用数を最新被引用数とする)の分布を示す。この分野の論文は、2000年にその年の論文数がピークの約500本に達したが、その後は減少傾向にあるが、現在でも安定して200-300本もの論文が発行されている。Dan Shechtman氏の論文(1984)は、Web of Science上の集計で3,402論文から引用されており、準結晶の引用ネットワークの中では最大の最新被引用数である。

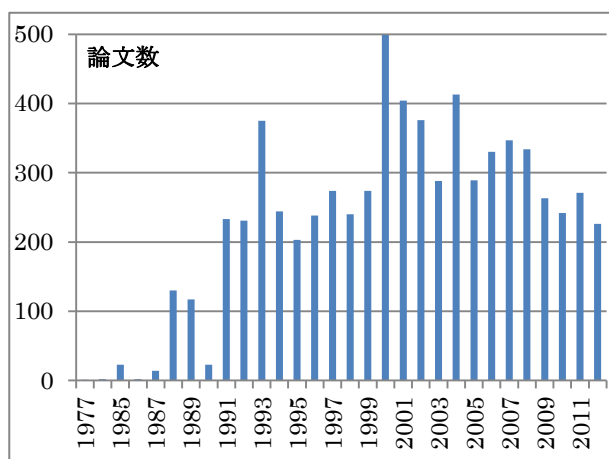


Figure 60: 出版年ごとの論文数の分布 [2011chem]

Table 16: 最新被引用数ごとの論文数の分布 [2011chem]

最新被引用数	論文数
0	879
1 - 100	5,872
101 - 200	118
201 - 300	19
301 - 400	11
401 - 500	0
501 - 600	1
601 - 700	1
701 - 800	2
801 - 900	0
901 - 1000	0
1001 - 2000	1
2001 - 3000	1
3001 - 4000	1

これらの論文の 2012 年末時点での引用ネットワークは Figure 61 のように形成される。クラスター分析を行った結果の各クラスターは各色で示され、TOP 7 クラスターの内容を Table 17 に、クラスター間のキーワード類似度を Figure 62 に示す。Shechtman 氏の論文(1984)が所属するクラスター1は、全体よりも平均出版年が古く、クラスター2とキーワード類似度が強い。クラスター2は、5と7とキーワード類似度が高い。Figure 62 から、TOP 7 クラスターは、{1, 2, 5, 7}、{3}、{4, 6}のグループに分かれる。

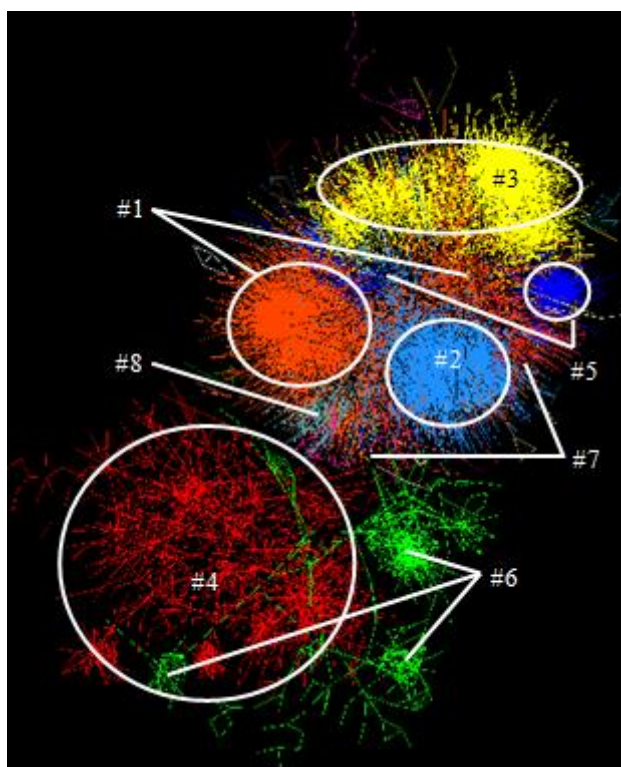


Figure 61 : Academic Landscape Map of Quasicrystal [2011chem]

	1	2	3	4	5	6	7
1	1.00	0.48	0.23	0.03	0.35	0.08	0.38
2	0.48	1.00	0.39	0.11	0.46	0.14	0.52
3	0.23	0.39	1.00	0.09	0.38	0.08	0.36
4	0.03	0.11	0.09	1.00	0.09	0.69	0.14
5	0.35	0.46	0.38	0.09	1.00	0.06	0.35
6	0.08	0.14	0.08	0.69	0.06	1.00	0.13
7	0.38	0.52	0.36	0.14	0.35	0.13	1.00

Figure 62 : Heatmap of Cosine Similarity between Top 7 Clusters [2011chem]

Table 17 : Information of TOP 7 Clusters [2011chem]

	Ave.PY*	Nodes	Edges	E/N*	Top Keywords
All	2001.1	6,906	53,030	7.68	
#1	1999.7	2,266	13,536	5.97	decagonal, tiling, photonic, quasicrystal, phase
#2	2000.6	1,687	14,480	8.58	icosahedral, approximant, quasicrystal, alloy, temperature
#3	2003.0	1,345	8,821	6.56	alloy, icosahedral, metallic glass, dislocation
#4	2003.4	604	1,104	1.83	insulator, charge, manganite, metallic, vo2
#5	2004.3	324	1,993	6.15	alloy, surface, icosahedral, quasicrystal, film
#6	2004.2	221	503	2.28	insulator transition, mott, metallic, hubbard
#7	1994.9	188	410	2.18	magnetic, icosahedral, mossbauer, alloy, channeling

*Ave.PY: Average year that papers are published(平均出版年)

E/N: Density that number of edges is divided by number of papers(ノードに対するエッジの密度)

3.5.2. 各指標の時系列分析

本項では、各指標の時系列分析を使って、萌芽的論文を検出するのに有効な指標について、詳細を解説する。

Table 18 は、この引用ネットワークで最新被引用数が多い順から TOP 10 の論文リストである。Figure 63 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターを特徴づけるキーワードと 5 年後のクラスターを特徴づけるキーワードとを、エッジの太さで表した図である。Figure 64 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。ノードのサイズがそのクラスターの所属する論文数に比例し、クラスターの 8 以降は煩雑になるため省略している。オレンジの破線は、2011 年にノーベル科学賞を受賞した Dan Shechtman 氏の論文(1984)が辿った軌跡である。Figure 64 において、1995 年のクラスター1 および 2005 年のクラスター1 を見ると、5 年前のクラスターの 2 以降を吸収して成長していることが見て取れる。そのような他の論文の大勢に関わらず、オレンジの破線、1985#3 と 1990#1 の間、2005#2 と 2010#1 の間のように、他の論文が経由しない軌跡も辿っている。

Table 18 : TOP 10 Papers on 2013-03-29 [2011chem]

#	R*	1st Author	PY*	TC*	Cluster Number						
					1985	1990	1995	2000	2005	2010	2012
1	1	Shechtman, D	1984	3,402	3	1	1	3	2	1	1
2	2	Suryanarayana, C	2001	2,342	-	-	-	-	1	3	1
3	3	Levine, D	1984	1,149	1	1	3	2	1	1	1
4	4	Uchida, S	1991	712	-	-	-	-	4	4	4
5	5	Elser, V	1985	702	2	2	4	1	1	2	2
6	6	Elser, V	1985	680	3	5	1	2	1	1	1
7	7	Bendersky, L	1985	579	1	3	2	1	3	1	1
8	8	Hanaguri, T	2004	390	-	-	-	-	-	4	4
9	9	Tokura, Y	1993	378	-	-	-	-	4	4	4
10	10	Phillips, JC	1996	343	-	-	-	2	1	2	1

* R : Rank based on TC, PY : Published Year(出版年), TC : Times Cited(最新被引用数), #: 受賞論文

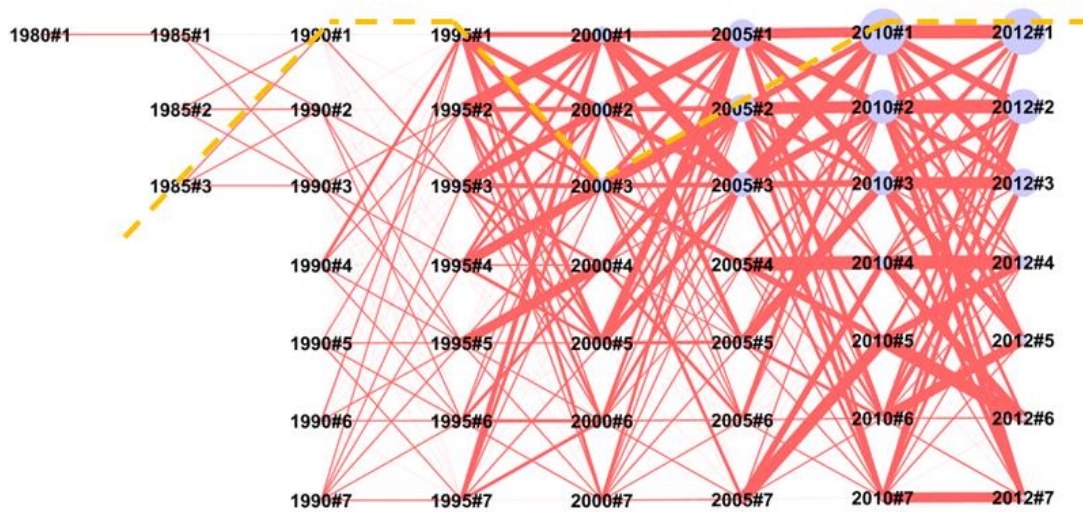


Figure 63: キーワードに基づくクラスター間の類似度 [2011chem]

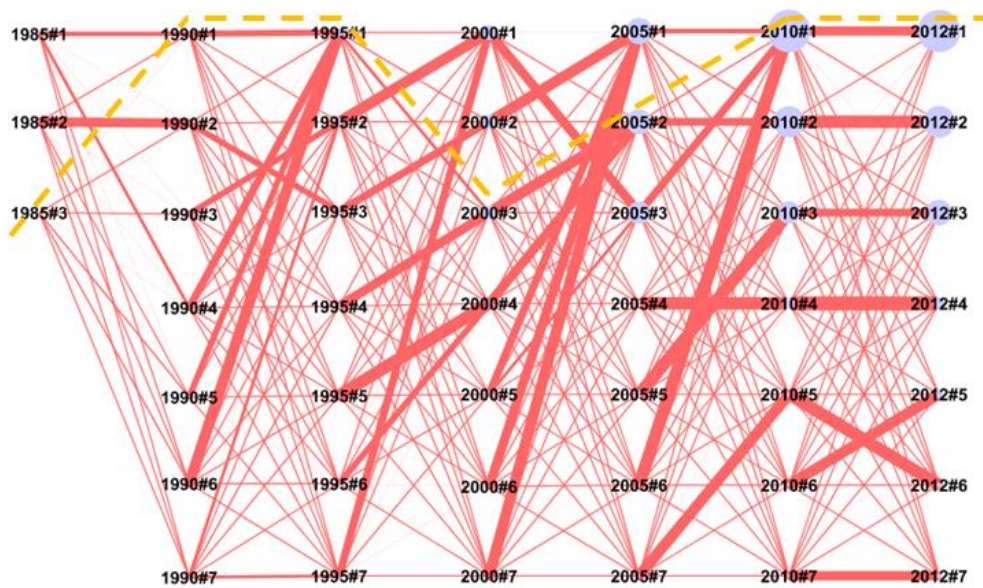


Figure 64: 論文の遷移量に基づくクラスター間の類似度 [2011chem]

Figure 65 は、データセット内被引用数の時系列変化であり、青線はノーベル賞受賞対象であり最新被引用数が最も多い Dan Shechtman 氏の論文(1984)である。最新被引用数が 2 位および 4 位の論文は本データセット内においては出版年から常に 20 未満と被引用数が少なく、最新被

引用数が3位の論文は1998-2008において5位の論文とデータセット内被引用数の順位が前後している。

なお、Figure 65、Figure 66、Figure 67、Figure 70では、Dan Shechtman氏の論文(1984)を含み、最新被引用数が100を越える155論文について表示している。

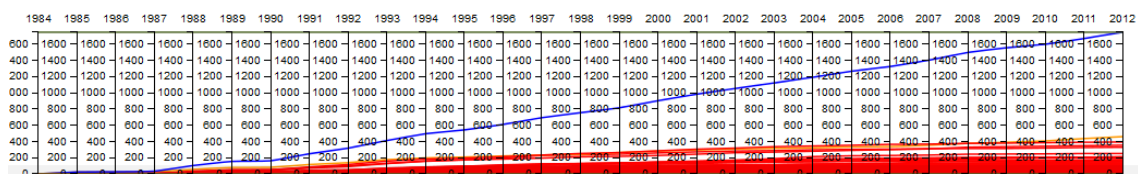


Figure 65 : Time Transition about Number of within-Dataset Citation of 155 Papers having more than 100 citations (最大値=1745) [2011chem]

Figure 66は、in-Degree Centralityの時系列変化であり、1位の青線は常に、3位のオレンジ線は1992年まで他の論文とは乖離している。最新被引用数が5位から7位の3論文は1985年に発表されているが、1988年にin-Degree Centralityの数値においても3位から5位に浮上している。なお、最新被引用数が2位の論文は2001年に出版されており、最新被引用数が4位の論文は1991年に出版されたものの2001年に始めて引用ネットワークに接続しており、in-Degree Centralityの時系列変化において明確な特徴は見られない。

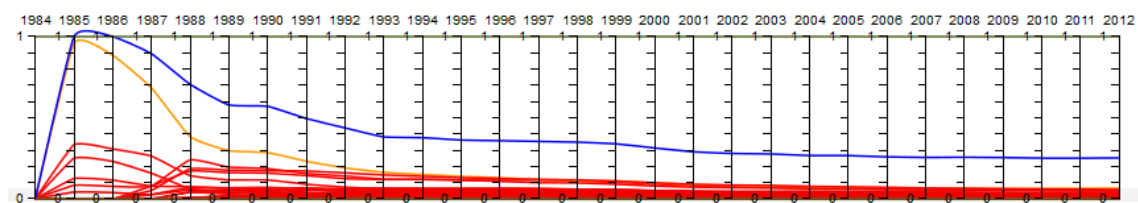


Figure 66 : Time Transition about in-Degree Centralities of 155 Papers having more than 100 citations (最大値=1) [2011chem]

Figure 67は、PageRankの時系列変化であり、Figure 66のin-Degree Centralityと同様に、青線（最新被引用数1位、ノーベル賞受賞論文）、および、オレンジ線（最新被引用数3位）が顕著に他論文と乖離しており、PageRankを利用した萌芽的論文の検出も可能であると考えられる。

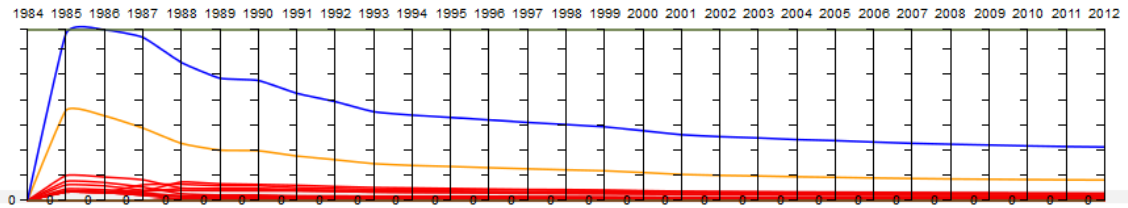


Figure 67 : Time Transition about PageRank of 155 Papers having more than 100 citations (最大値=0.3383) [2011chem]

以上から、in-Degree Centrality と PageRank において萌芽的論文は高い数値を持つことから、萌芽的論文は早い段階で検知できると考えられる。Figure 65 の引用数を単純に数えただけでも、2011 年時点においては十分に Dan Shechtman 氏の論文(1984)が特殊なものであると判断できるが、出版年直後の 1985-1987 年時点では他の論文に紛れて区別することは難しい。Figure 66、および、Figure 67 のように、in-Degree Centrality と PageRank であれば、1985 年時点で他の論文と区別することが可能であり、これらの特徴は、萌芽的論文の早期発見に利用できると考えられる。

ところで、Figure 66、および、Figure 67 において、最新被引用数が 100 を超える論文のみに限定して描画したため、最新被引用数が低い論文が高い論文よりも高い in-Degree Centrality と PageRank を持つ懸念が残る。そこで、Figure 68 と Figure 69 は、萌芽的論文を検出できる可能性の高い in-Degree Centrality と PageRank において、最新被引用数の低い論文が高い値を持つ可能性を検証するために、最新被引用数が 1 の論文群について、時系列で値を描画したものである。Figure 66、および、Figure 67 と比較すると、最新被引用数の低い論文群における in-Degree Centrality と PageRank の最大値は、最新被引用数の高い論文群の 100 分の 1 以下であるので、Figure 66、および、Figure 67 における「in-Degree Centrality と PageRank において萌芽的論文は高い数値を持つことから、萌芽的論文は早い段階で検知できる」と結果は変わらないと考えられる。

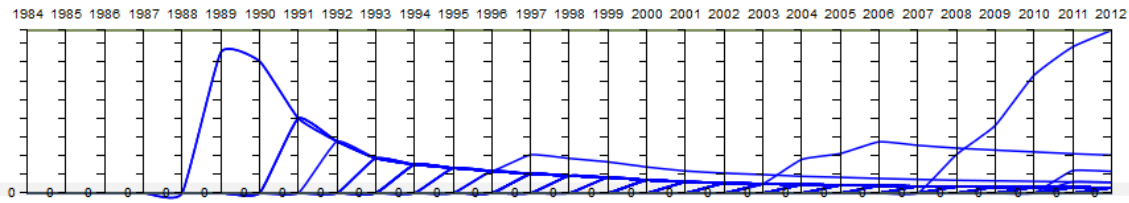


Figure 68 : Time Transition about in-Degree Centralities of 689 Papers having 1 citation (最大値=0.004345) [2011chem]

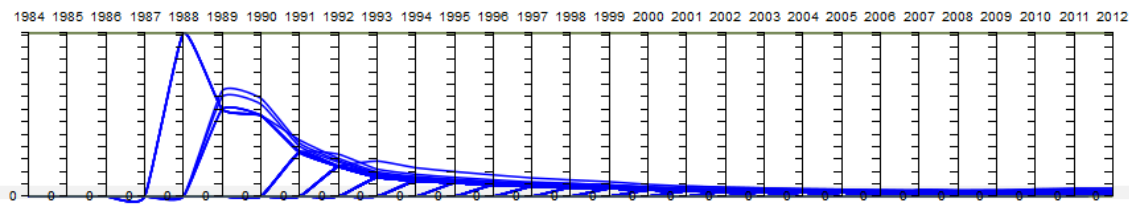


Figure 69 : Time Transition about PageRank of 689 Papers having 1 citation (最大値=0.002622) [2011chem]

Communicability Centrality については、in-Degree Centrality と PageRank と違った形で萌芽的論文を検出できる可能性がある。Figure 70 は、Figure 66 と Figure 67 の in-Degree Centrality と PageRank と同様に、最新被引用数が 100 を超える論文について Communicability Centrality の時系列変化を描画したものであるが、1 本の萌芽的論文（青線）の数値が突出して大きいため、他の論文の兆候が把握できず、最新年まで描画した場合においては他の萌芽的論文を検出するのは難しい。

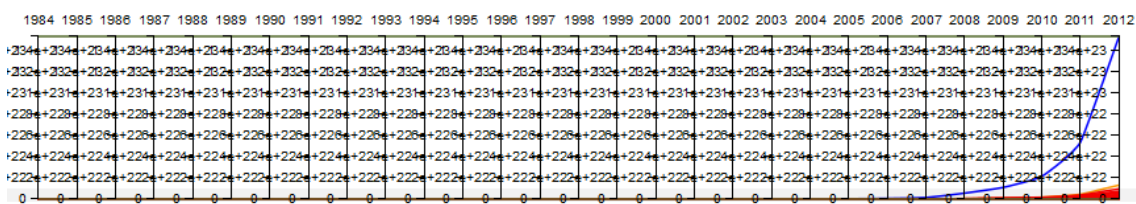


Figure 70 : Time Transition about Communicability Centralities of 155 Papers having more than 100 citations (最大値= 1.529e+23) [2011chem]

そこで、Figure 71 と Figure 72 のように、1990 年までに出版された論文について、1990 年まで、および、1989 年までに限定して描画すると、最新被引用数の上位の論文が顕著に乖離した大きな値を持つことが分かる。Figure 71 からは、1990 年時点で最新被引用数の上位 4 論文が

最新被引用数の順に大きな値を持って他の論文群と乖離しており、Figure 72 においては、青線（最新被引用数 1 位、ノーベル賞受賞論文）、および、オレンジ線（最新被引用数 3 位）が 1985 年時点で検出可能である。

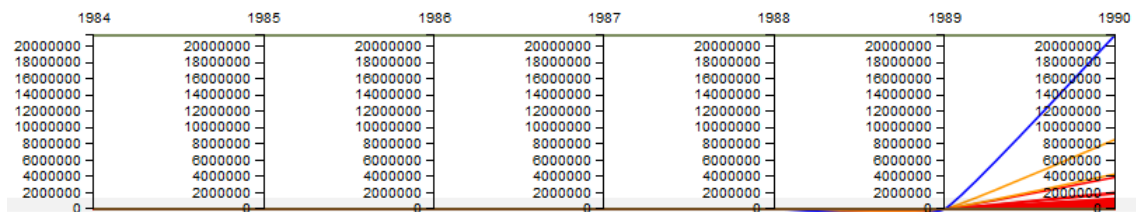


Figure 71 : Time Transition about Communicability Centralities of 312 Papers published until 1990 (最大値=2.131e+7) [2011chem]

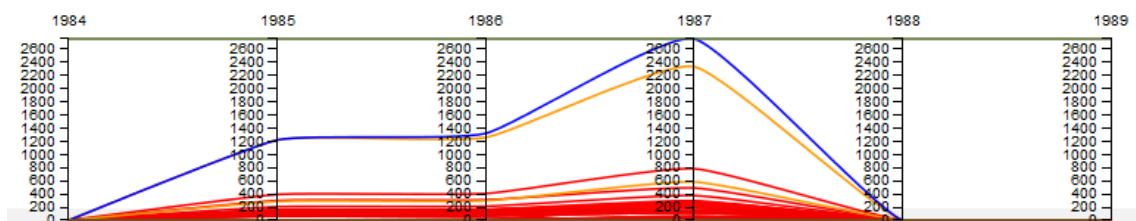


Figure 72 : Time Transition about Communicability Centralities of 312 Papers published until 1990 (最大値=2754) [2011chem]

どの指標が有効であるかを明らかにするために、Figure 73 に最新被引用数と指標の相関係数を、Figure 74 に各年のデータセット内の被引用数と指標の相関係数を、年ごとに計算した。Figure 73 では、相関係数の高い in-Degree Centrality、Communicability Centrality、PageRank の 3 指標は、このデータセットの初期から現在まで、0.7 以上と高い相関係数を保っており、萌芽的論文として特別に突出した論文だけに有効な指標ではなく、萌芽的論文に次ぐレベルの論文群も検出可能である。Figure 74 では、被引用数の正規化した値である in-Degree Centrality は相関係数が常にほぼ 1 であり、次いで、Communicability Centrality、PageRank も 0.8 以上の高い相関を示している。

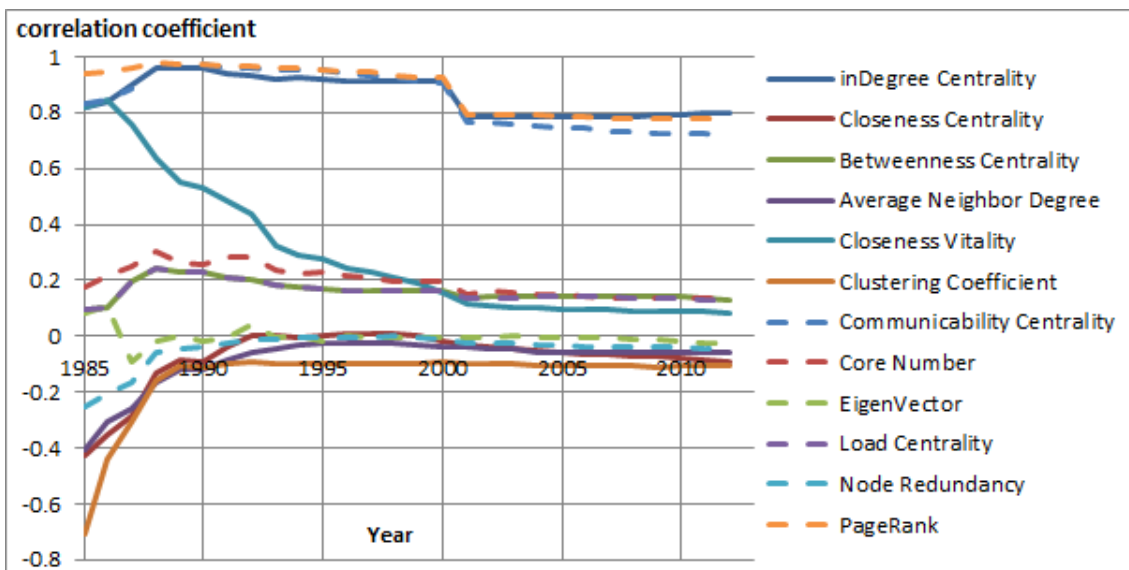


Figure 73 : Correlation Coefficient between Current Number of Citations and Indicators [2011chem]

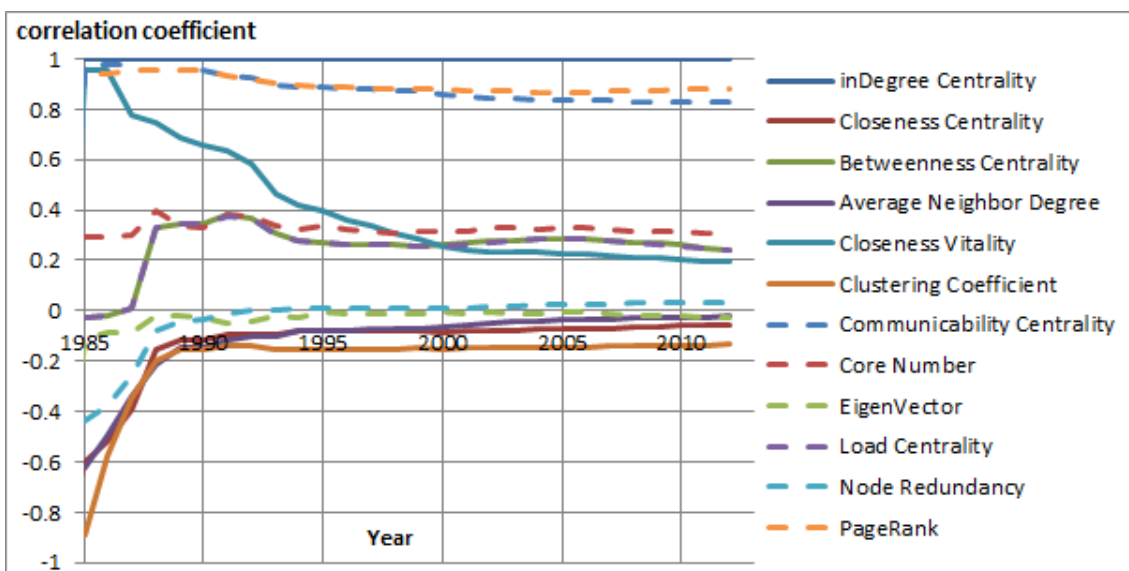


Figure 74 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [2011chem]

3.6. パラジウム触媒クロスカップリング反応(2010年ノーベル化学賞 [2010CHEM])

本節では、2010年のノーベル化学賞より、パラジウム触媒クロスカップリング反応について、指標を時系列で図示することにより、有効な指標を選定する。

3.6.1. 分析対象分野の概要

自然界において炭素原子を含む結合は安定的であるため、医薬品や材料のような炭素原子を多く含む複雑な分子を作り出し、かつ、希望と異なる副産物の分子を抑制するために、有機合成においては、様々な触媒が試行錯誤されてきた。また、商業的に成功するかどうかは、より安価により効率よく化学反応を行う触媒の選択が大きな影響を与える。

1971年に溝呂木勉氏、1972年に Richard Fred Heck 氏が、有機ハロゲンとアルケンという2種類の炭素化合物を結合させる際に、パラジウム化合物を併用すると反応がうまくいくことを発見し、「溝呂木・ヘック反応」と名付けられている。Heck氏はこの発見により、2010年のノーベル化学賞を受賞したが、溝呂木氏はすでに故人であったため、存命の科学者に与えられる賞であるノーベル賞は受賞できなかった。その後、1977年に、根岸栄一氏がパラジウムに亜鉛を加えて安定的なカップリング反応を起こすことに成功した(根岸カップリング反応)。さらに、1979年に鈴木章氏がパラジウムにホウ素を加えることで、効率的に合成を行うことができ、触媒を再利用することのできる、商業的により有利な合成方法を実現したのである(鈴木カップリング反応、あるいは、鈴木・宮浦カップリング反応。宮浦氏は、鈴木氏の論文の共著者)(中西, 2012)。

パラジウム触媒クロスカップリング反応の分析をするにあたって、受賞の際に発表される資料から、以下の論文をに貢献した論文として抽出した。

- Heck, R. F. and Nolley, J. P. (1972) J. Org. Chem. 37, p. 2320.
[WOS:A1972M962500024]
 - PALLADIUM-CATALYZED VINYLIC HYDROGEN SUBSTITUTION REACTIONS WITH ARYL, BENZYL, AND STYRYL HALIDES

- Negishi, E.-I., King, A. O. and Okukado, N. (1977) J. Org. Chem. 42, p. 1821.
[WOS:A1977DF31700041]
 - SELECTIVE CARBON-CARBON BOND FORMATION VIA
TRANSITION-METAL CATALYSIS .3. HIGHLY SELECTIVE SYNTHESIS
OF UNSYMMETRICAL BIARYLS AND DIARYLMETHANES BY
NICKEL-CATALYZED OR PALLADIUM-CATALYZED REACTION OF
ARYL DERIVATIVES AND BENZYLZINC DERIVATIVES WITH ARYL
HALIDES
- Miyaura, N., Yamada, K., and Suzuki, A. (1979a), Tetrahedron Lett. 20, 3437.
[WOS:A1979HK10000026]
 - NEW STEREOSPECIFIC CROSS-COUPPLING BY THE
PALLADIUM-CATALYZED REACTION OF 1-ALKENYLBORANES WITH
1-ALKENYL OR 1-ALKYNYL HALIDES
- Miyaura, N. and Suzuki, A. (1979b), J. Chem. Soc. Chem. Commun., p 866.
[WOS:A1979HQ98200031]
 - STEREOSELECTIVE SYNTHESIS OF ARYLATED (E)-ALKENES BY
THE REACTION OF ALK-1-ENYLBORANES WITH ARYL HALIDES IN
THE PRESENCE OF PALLADIUM CATALYST

Thomson Reuters 社の論文サービス Web of Science より、「Topic=("palladium-catalyz*" OR "palladium-catalys*" OR "cross coupling*" OR "organic synthesis" OR "organic syntheses") AND Year Published=(1900-2012)」というクエリーで、タイトル、アブストラクト、キーワード、キーワード・プラスを検索し、66,656 論文の書誌情報を得た。そのうち、2012 年末時点で引用ネットワークを形成可能な 26,546 論文の書誌情報を対象に、複数の指標を各年について算出した。Figure 75、および、Table 13 に、対象となる論文の出版年の分布と、書誌情報を収集した 2013-05-08 時点での被引用数(以降、最新被引用数とする)の分布を示す。

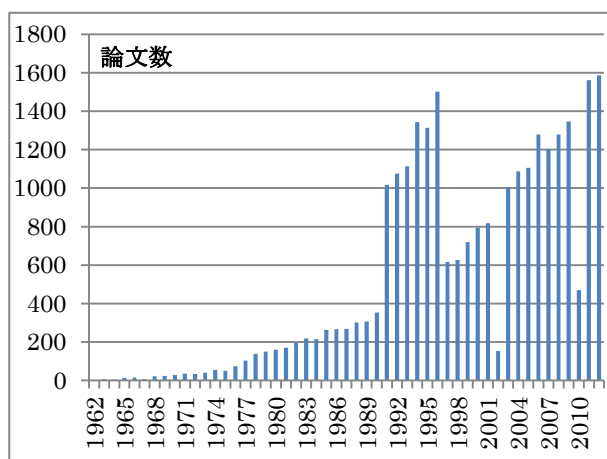


Figure 75: 出版年ごとの論文数の分布 [2010chem]

Table 19: 最新被引用数ごとの論文数の分布 [2010chem]

最新被引用数	論文数
0	1,325
1 - 100	23,743
101 - 200	981
201 - 300	239
301 - 400	97
401 - 500	45
501 - 600	38
601 - 700	13
701 - 800	10
801 - 900	14
901 - 1000	10
1001 - 2000	25
2001 - 3000	5
3001 - 4000	0
4001 - 5000	0
5001 - 6000	0
6001 - 7000	1

これらの論文の 2012 年末時点での引用ネットワークは Figure 76 のように形成される。クラスター分析を行った結果の各クラスターは各色で示され、TOP 7 クラスターの内容を Table 20 に、クラスター間のキーワード類似度を

	1	2	3	4	5	6	7
1	1.00	0.30	0.05	0.06	0.12	0.16	0.41
2	0.30	1.00	0.29	0.20	0.13	0.34	0.14
3	0.05	0.29	1.00	0.09	0.05	0.24	0.02
4	0.06	0.20	0.09	1.00	0.01	0.16	0.06
5	0.12	0.13	0.05	0.01	1.00	0.14	0.03
6	0.16	0.34	0.24	0.16	0.14	1.00	0.07
7	0.41	0.14	0.02	0.06	0.03	0.07	1.00

Figure 77 に示す。

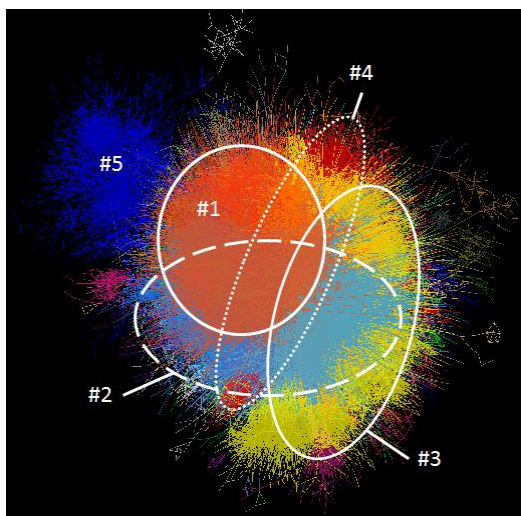


Figure 76 : Academic Landscape Map of palladium-catalyzed cross couplings in organic synthesis [2010chem]

	1	2	3	4	5	6	7
1	1.00	0.30	0.05	0.06	0.12	0.16	0.41
2	0.30	1.00	0.29	0.20	0.13	0.34	0.14
3	0.05	0.29	1.00	0.09	0.05	0.24	0.02
4	0.06	0.20	0.09	1.00	0.01	0.16	0.06
5	0.12	0.13	0.05	0.01	1.00	0.14	0.03
6	0.16	0.34	0.24	0.16	0.14	1.00	0.07
7	0.41	0.14	0.02	0.06	0.03	0.07	1.00

Figure 77 : Heatmap of Cosine Similarity between Top 7 Clusters [2010chem]

Table 20 : Information of TOP 7 Clusters [2010chem]

	Ave.PY*	Nodes	Edges	E/N*	Top Keywords
All	1999.4	26,546	150,073	5.65	
#1	2000.7	7,459	52,528	7.04	aryl, cross coupling, palladium
#2	1999.3	6,241	27,718	4.44	allylic, asymmetric, complex, chiral, ligand
#3	1997.6	5,726	24,127	4.21	radical, organic synthesis, samarium, enzyme, alpha
#4	2001.2	2,741	8,568	3.13	microwave, metathesis, carbene complex
#5	1998.6	1,098	2,905	2.65	hydrogenation, supported, palladium catalyst, surface
#6	2002.3	432	715	1.66	bismuth, iron, complex, catalyst, organic synthesis
#7	2003.0	410	962	2.35	porphyrin, palladium catalyzed cyanation, meso

*Ave.PY: Average year that papers are published(平均出版年)

E/N: Density that number of edges is divided by number of papers(ノードに対するエッジの密度)

3.6.2. 各指標の時系列分析

本項では、各指標の時系列分析を使って、萌芽的論文を検出するのに有効な指標について、詳細を解説する。

Table 21 は、この引用ネットワークで最新被引用数が多い順から TOP 10 の論文リストである。Figure 78 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。Figure 79 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。ノードのサイズがそのクラスターの所属する論文数に比例し、クラスターの 8 以降は煩雑になるため省略している。オレンジの破線は、2010 年にノーベル賞を受賞した Richard Fred Heck 氏の論文(1972) が辿った軌跡である。グリーン破線は、同賞を受賞した根岸氏の論文(1977)と鈴木氏の論文(Miyaura, 1979) が辿った軌跡である。

受賞論文として選定し、かつ、初期に発表された論文は、最新被引用数の上位論文に比べて、被引用数は少ないものの、1985 年以降、クラスター1 の上位論文に入っている。一方で、最新被引用数の 1 位と 2 位の論文は、2000 年に一度、クラスター4 に遷移している。この 2000 年のクラスター4 について Figure 78 を見ると、前後の年のクラスター1 とキーワード類似度が高くなっており、また、Figure 79 を見ると、前年のクラスター1 から多くの論文が遷移して来て、後年のクラスター1 に多くの論文が遷移して行っていることから、2000 年に内容的に非常に近いクラスター1 と 4 が分離している状況が分かる。2000 年のヒートマップを確認したところ、0.5699 と非常に高いコサイン類似度が得られた。この変動が発生した 1995 年から 2005 年を見ると、最新被引用数の 1 位の論文がデータセット内での当時の被引用数において順位を上げている現象が見られる。

Table 21 : TOP 10 and the Nobel Laureate's Papers on 2013-05-08 [2010chem]

R*	1 st Author	PY*	TC*	Cluster Number							
				1980	1985	1990	1995	2000	2005	2010	2012
1	Miyaura, N	1995	6,365	-	-	-	1	4	1	1	1
2	Stille, JK	1986	2,865	-	-	1	1	4	1	1	1
3	Desiraju, GR	1995	2,618	-	-	-	11	12	10	4	10
4	Kolb, HC	1994	2,548	-	-	-	3	2	2	2	2
5	Tmka, TM	2001	2,498	-	-	-	-	-	4	6	4
6	Tietze, LF	1996	2,180	-	-	-	-	1	2	2	2
7	Trost, BM	1991	1,817	-	-	-	1	1	1	2	2
8	Dess, DB	1991	1,738	-	-	-	4	2	2	3	3
9	Corey, EJ	1965	1,597	6	7	11	3	2	2	2	3
10	Heck, RF	1982	1,448	-	1	1	1	1	1	1	1
#	-	Heck, RF	1972	928	3	1	1	1	1	1	1
#	-	Negishi, E	1977	579	2	1	1	1	1	1	1
#	-	Miyaura, N	1979a	556	2	1	1	1	1	1	1
#	-	Miyaura, N	1979b	239	2	1	1	1	1	1	1

*R : Rank based on TC, PY : Published Year(出版年), TC : Times Cited(最新被引用数), #: 受賞論文

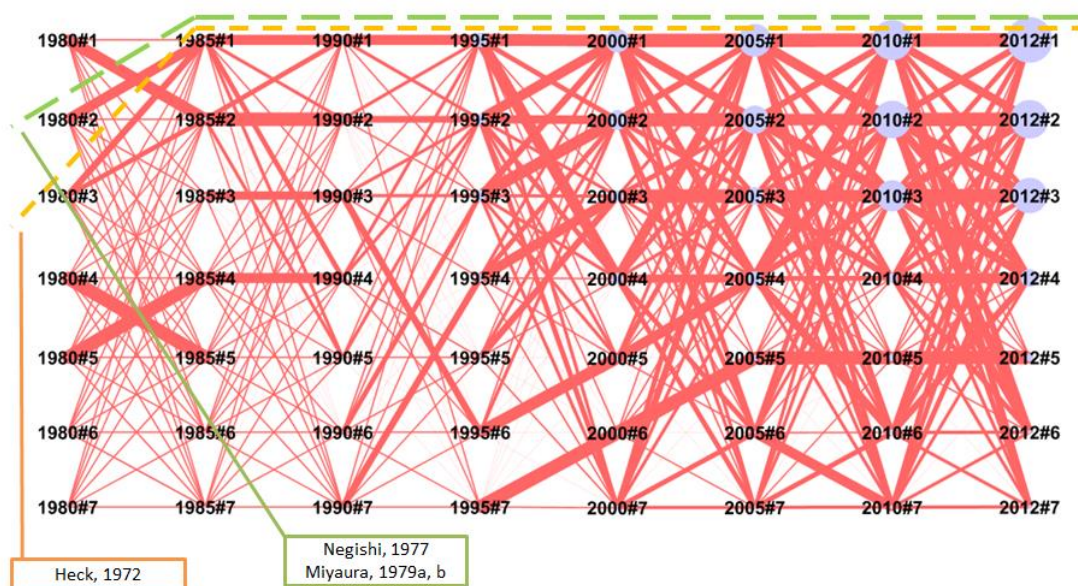


Figure 78 : キーワードに基づくクラスター間の類似度 [2010chem]

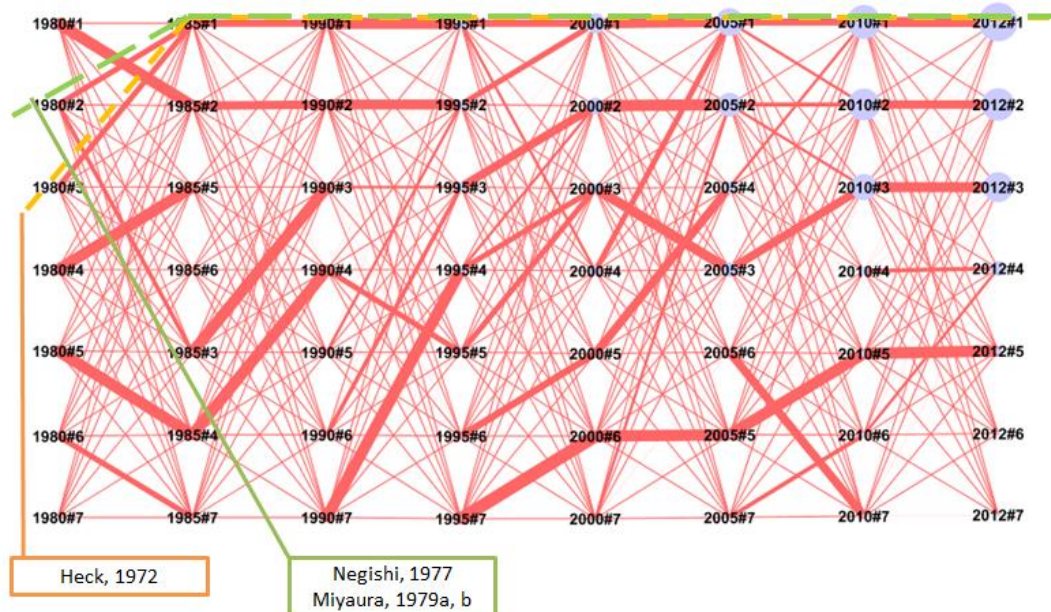


Figure 79: 論文の遷移量に基づくクラスター間の類似度 [2010chem]

Figure 80 は各年の被引用数、Figure 81 は in-Degree Centrality、Figure 82 は PageRank の時系列変化であり、データセット内で引用ネットワークで計算を行っている。Figure 80、Figure 81、Figure 82 ではいずれも、ノーベル賞受賞論文を含むよう、最新被引用数が 200 を越える 580 論文について表示しており、各線は最新被引用数で 5 等分して多い順に青、アクア、グリーン、オレンジ、赤の着色を行っている。青線は Table 21 で最も最新被引用数が多い宮浦氏・鈴木氏の共著論文(1995)である。グリーン線は、最新被引用数が 2 位の論文(1986)であり、スティルカップリング(または、右田・小杉・スティルカップリング)と呼ばれるパラジウム触媒カップリング反応に関する論文である。

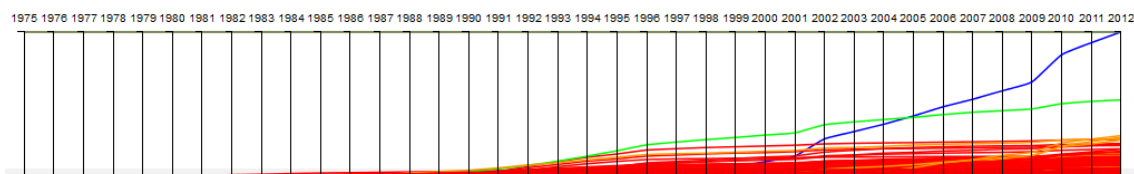


Figure 80: Time Transition about Number of within-Dataset Citation of 580 Papers having more than 200 citations (最大値=1928) [2010chem]

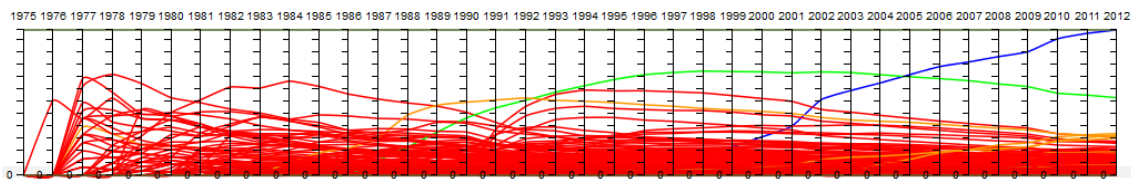


Figure 81 : Time Transition about in-Degree Centralities of 580 Papers having more than 200 citations (最大値=0.05898) [2010chem]

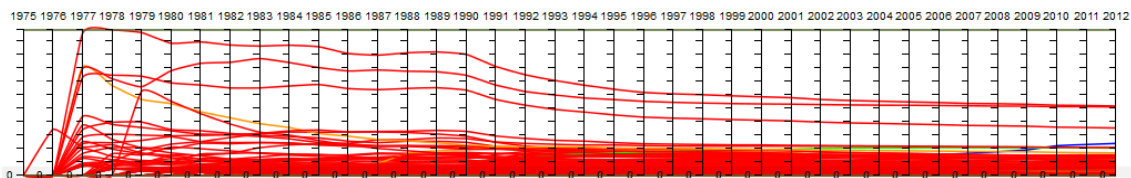


Figure 82 : Time Transition about PageRank of 580 Papers having more than 200 citations (最大値=0.02162) [2010chem]

どの指標が有効であるかを明らかにするために、Figure 83 に最新被引用数と指標の相関係数を、Figure 84 に各年のデータセット内の被引用数と指標の相関係数を、年ごとに計算した。Figure 84 では、被引用数の正規化した値である in-Degree Centrality は相関係数が常にほぼ 1 であり、PageRank、Core Number は年々、データセット内での被引用数との相関が低くなっていくものの、まだ、弱い相関を保っている。一方で、Betweenness Centrality、Load Centrality は年々、データセット内での被引用数との相関が強くなっている。

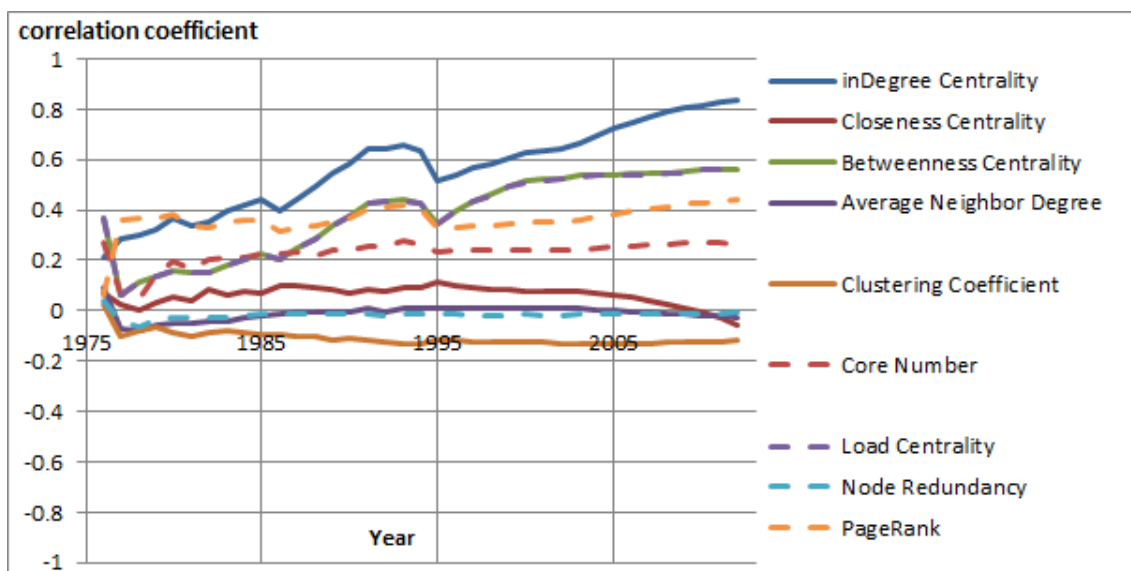


Figure 83 : Correlation Coefficient between Current Number of Citations and Indicators [2010chem]

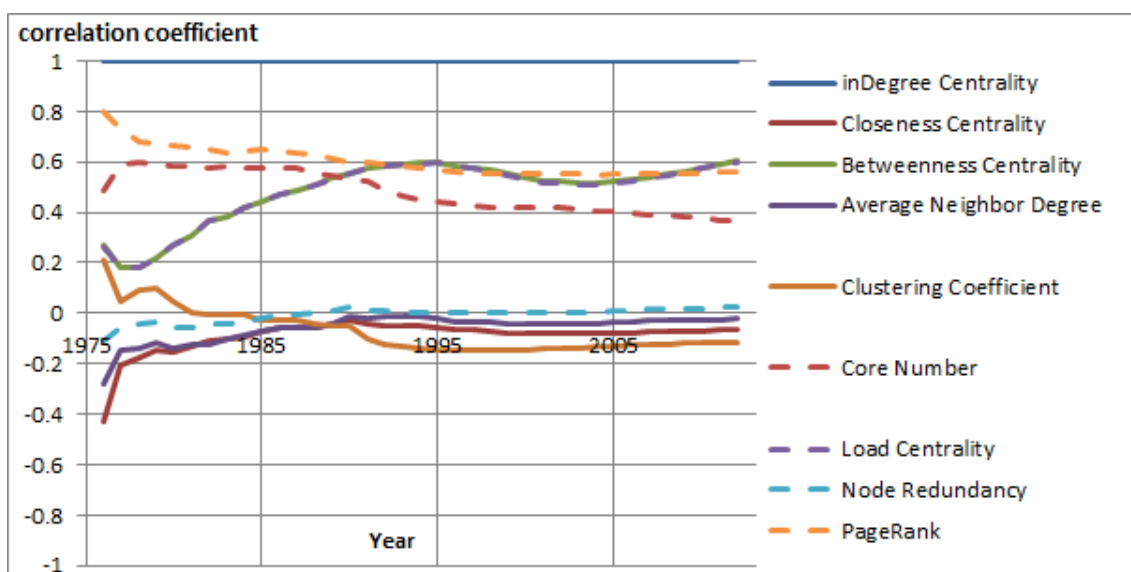


Figure 84 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [2010chem]

3.7. 体外受精技術(2010年ノーベル生理学・医学賞 [2010MED])

本節では、2010年のノーベル生理学・医学賞より、体外受精技術について、指標を時系列で図示することにより、有効な指標を選定する。

3.7.1. 分析対象分野の概要

1978年、世界初の体外受精児、**Louise Brown** が誕生した。2010年のノーベル生理学・医学賞受賞時点では、すでに約4万人の体外受精児が誕生し親の年齢に成長した者も登場している。この技術の成果により、体外受精は不妊症治療恩恵を与えており、胚性幹細胞(ES細胞)研究の道を開くことにもつながった(Abbott, 2010)。

体外受精技術の分析をするにあたって、受賞の際に発表される資料から、以下の論文をに貢献した論文として抽出した。

- Edwards RG. Maturation in vitro of human ovarian oocytes. *Lancet* 1965; 2:926-929. [WOS:A19656954700005]
- Edwards RG, Bavister BD, Steptoe PC. Early stages of fertilization in vitro of human oocytes matured in vitro. *Nature* 1969; 221:632-635. [WOS:A1969C573400017]
- Edwards RG, Steptoe PC, Purdy JM. Fertilization and cleavage in vitro of human oocytes matured in vivo. *Nature* 1970; 227:1307-1309. [WOS:A1970H331200031]
- Steptoe PC, Edwards RG. Birth after the reimplantation of a human embryo. *Lancet* 1978; 2:366. [WOS:A1978FL08800018]

Thomson Reuters社の論文サービス Web of Science より、「Topic=((("in vitro" AND ("fertilization*" OR "maturation*" OR "cleavage*")) OR (reimplantation* OR "human embryo")) AND Year Published=(1900-2012)」というクエリーで、タイトル、アブストラクト、キーワード、キーワード・プラスを検索し、66,561論文の書誌情報を得た。そのうち、2012年末時点で引用ネットワークを形成可能な50,609論文の書誌情報を対象に、複数の指標を各年に

ついて算出した。Figure 85、および、Table 22 に、対象となる論文の出版年の分布と、書誌情報を収集した 2013-04-05 時点での被引用数(以降、最新被引用数とする)の分布を示す。

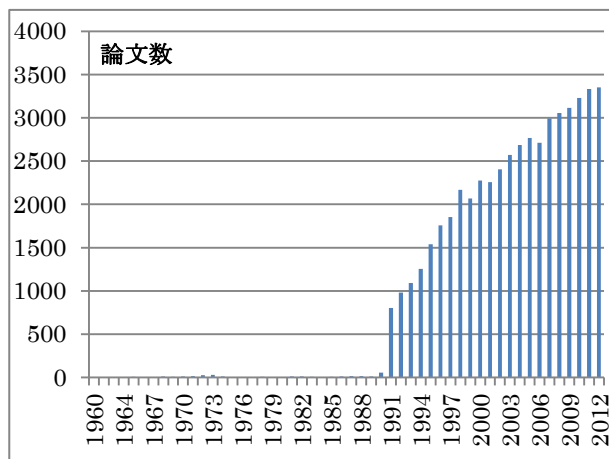


Figure 85: 出版年ごとの論文数の分布 [2010med]

Table 22: 最新被引用数ごとの論文数の分布 [2010med]

最新被引用数	論文数
0	4,527
1 - 100	43,536
101 - 200	1,883
201 - 300	385
301 - 400	138
401 - 500	41
501 - 600	38
601 - 700	10
701 - 800	15
801 - 900	6
901 - 1000	6
1001 - 2000	20
2001 - 3000	1
3001 - 4000	3

これらの論文の 2012 年末時点での引用ネットワークは Figure 86 のように形成される。クラスター分析を行った結果の各クラスターは各色で示され、TOP 7 クラスターの内容を Table 19 に、クラスター間のキーワード類似度を Figure 87 に示す。Edwards 氏の 4 論文(Table 24 の太字)が所属するクラスター1と2と4は、Figure 87 からキーワード類似度が高いことが分かるが、他のクラスター相互間はキーワード類似度が低い。クラスター1と2と4に関しては、Table 19 の E/N の値が高いように、クラスター内での相互のリンクが多く密なクラスターを形成している。

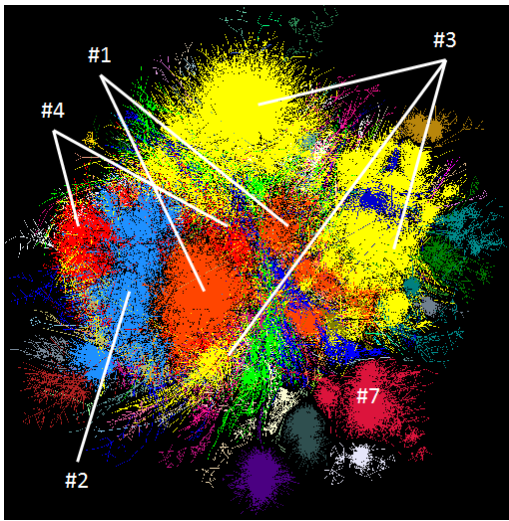


Figure 86 : Academic Landscape Map of in vitro fertilization [2010med]

	1	2	3	4	5	6	7
1	1.00	0.38	0.02	0.38	0.03	0.00	0.00
2	0.38	1.00	0.00	0.75	0.00	0.00	0.03
3	0.02	0.00	1.00	0.00	0.04	0.02	0.00
4	0.38	0.75	0.00	1.00	0.00	0.00	0.03
5	0.03	0.00	0.04	0.00	1.00	0.02	0.00
6	0.00	0.00	0.02	0.00	0.02	1.00	0.00
7	0.00	0.03	0.00	0.03	0.00	0.00	1.00

Figure 87 : Heatmap of Cosine Similarity between Top 7 Clusters [2010med]

Table 23 : Information of TOP 7 Clusters [2010med]

	Ave.PY*	Nodes	Edges	E/N*	Top Keywords
All	2003.4	50,609	357,530	7.06	
#1	2003.6	12,811	122,250	9.54	oocyte, embryo, blastocyst, follicle, sperm
#2	2003.6	11,960	117,452	9.82	pregnancy, sperm, ivf, fertilization, embryo
#3	2003.4	10,334	35,282	3.41	dendritic cell, ribozyme, virus, rna
#4	2002.5	4,274	23,194	5.43	pregnancy, woman, ivf, endometriosis, fertilization
#5	2003.7	1,513	2,600	1.72	neuron, oligodendrocyte, gaba, neuronal, brain
#6	2004.5	1,408	2,743	1.95	cartilage, chondrocytes, osteoblast, bone, collagen
#7	2001.9	821	2,982	3.63	ureteral, reimplantation, vesicoureteral reflux

*Ave.PY: Average year that papers are published(平均出版年)

E/N: Density that number of edges is divided by number of papers(ノードに対するエッジの密度)

3.7.2. 各指標の時系列分析

本項では、各指標の時系列分析を使って、萌芽的論文を検出するのに有効な指標について、詳細を解説する。

Table 24 は、この引用ネットワークで最新被引用数が多い順から TOP 10 の論文リストである。Figure 88 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。Figure 89 は、5 年ごとに引用ネットワークをクラスターに分割した後、ある年のクラスターに含まれる論文のうち何%の論文が 5 年後のクラスターに遷移するかを、エッジの太さで表した図である。ノードのサイズがそのクラスターの所属する論文数に比例し、クラスターの 8 以降は煩雑になるため省略している。オレンジ、グリーン、アクアの破線は、2010 年にノーベル生理学・医学賞を受賞した Edwards 氏の 4 論文が辿った軌跡である。

Table 24 : TOP 10 and the Nobel Laureate's Papers on 2013-03-29 [2010med]

R*	1 st Author	PY*	TC*	Cluster Number								
				1980	1985	1990	1995	2000	2005	2010	2012	
1	Kluck, RM	1997	3,360	-	-	-	-	4	1	2	1	
2	Liu, XS	1996	3,235	-	-	-	-	4	1	2	1	
3	Nicholson, DW	1995	3,056	-	-	-	14	4	1	2	1	
4	Cohen, GM	1997	2,852	-	-	-	-	4	1	2	1	
5	Lazebnik, YA	1994	1,836	-	-	-	14	4	1	2	1	
6	Elbashir, SM	2001	1,783	-	-	-	-	-	4	2	1	
7	Lee, Y	2003	1,541	-	-	-	-	-	4	2	1	
8	Heil, F	2004	1,502	-	-	-	-	-	5	4	3	
9	Rossi, D	2000	1,410	-	-	-	-	6	5	4	3	
10	Zamore, PD	2000	1,386	-	-	-	-	-	4	2	1	
#	-	Edwards, RG	1965	178	1	1	1	1	2	1	2	1
#	-	Edwards, RG	1969	253	1	1	1	1	2	1	2	1
#	-	Edwards, RG	1970	116	1	1	3	1	5	1	10	4
#	-	Steptoe, PC	1978	697	-	-	4	2	1	2	1	2

* R : Rank based on TC, PY : Published Year(出版年), TC : Times Cited(最新被引用数), #: 受賞論文

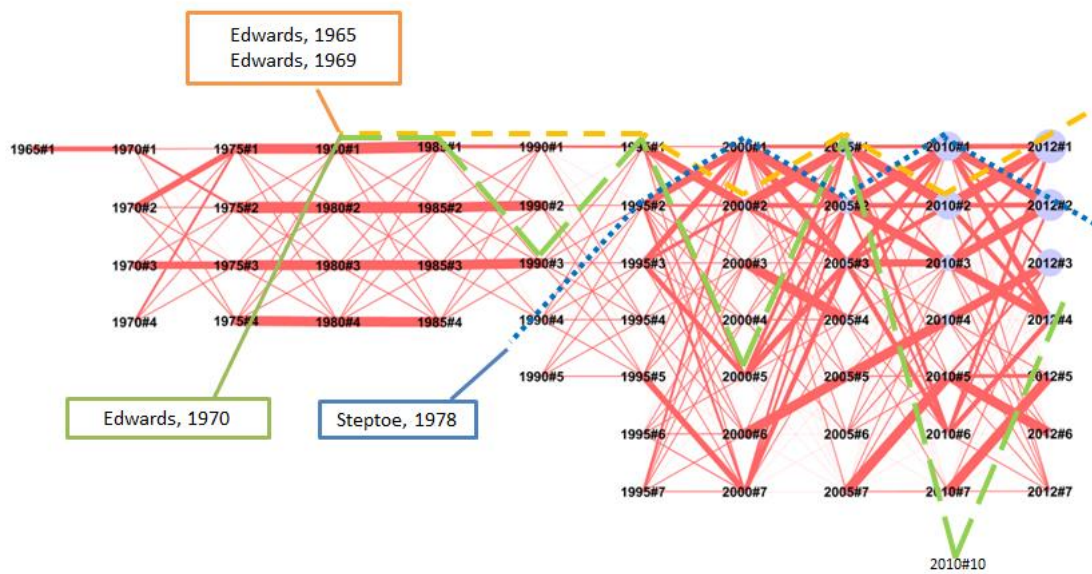


Figure 88 : キーワードに基づくクラスター間の類似度 [2010med]

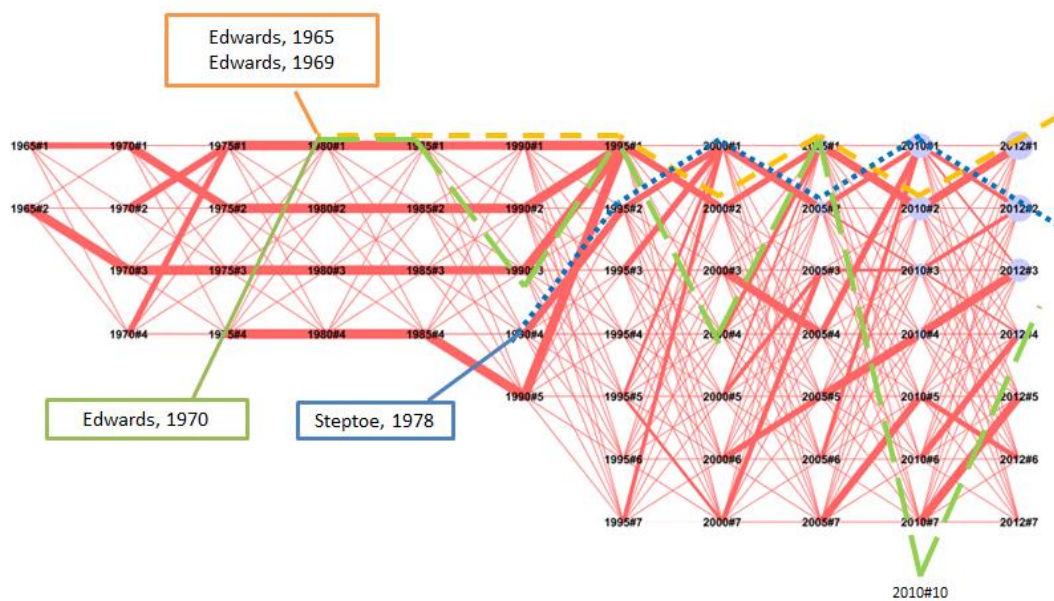


Figure 89 : 論文の遷移量に基づくクラスター間の類似度 [2010med]

Figure 90 は各年の被引用数、Figure 91 は in-Degree Centrality、Figure 92 は PageRank の時系列変化であり、データセット内で引用ネットワークで計算を行っている。Figure 90、Figure

91、Figure 92 ではいずれも、最新被引用数が 100 を越える 2,546 論文について表示しており、各線は最新被引用数で 5 等分して多い順に青、アクア、グリーン、オレンジ、赤の着色を行っている。

Figure 87 の 1 と 2 と 4 以外のクラスター間のキーワード類似度が低いこと、および、Figure 90 において、Web of Science 内での最新被引用数上位である青線の論文が、データセット内での被引用数において上位にならないことから、ノーベル賞受賞分野から離れたトピックの論文が多く紛れ込んでいると考えられる。

Figure 85 にあるように 1991 年以降の論文数が急増しており、この急増の前では、in-Degree Centrality(Figure 91)、および、PageRank(Figure 92)において、1966 年に Edwards 氏の論文が 2 番手、1974-1990 年の論文数がほとんど増加しない時期では 3 番手にあり、他のデータセットほどではないが、早期に萌芽的論文として特定することが可能である。

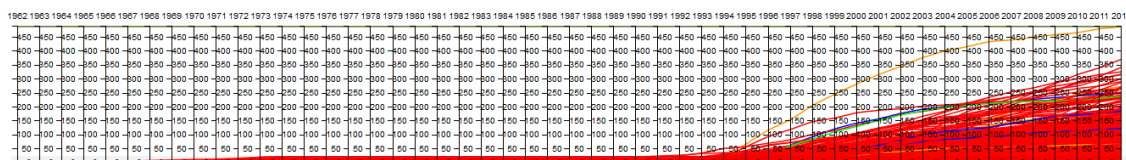


Figure 90 : Time Transition about Number of within-Dataset Citation of 2,546 Papers having more than 100 citations (最大値=488) [2010med]

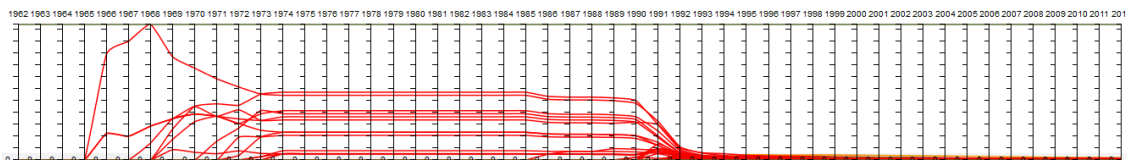


Figure 91 : Time Transition about in-Degree Centralities of 2,546 Papers having more than 100 citations (最大値=0.5714) [2010med]

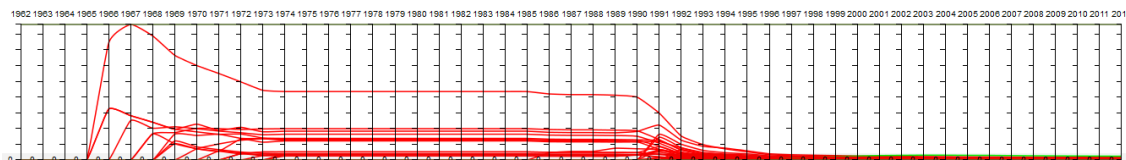


Figure 92 : Time Transition about PageRank of 2,546 Papers having more than 100 citations (最大値=0.1720) [2010med]

どの指標が有効であるかを明らかにするために、Figure 93 に最新被引用数と指標の相関係数を、Figure 94 に各年のデータセット内の被引用数と指標の相関係数を、年ごとに計算した。Figure 94 では、被引用数の正規化した値である in-Degree Centrality は相関係数が常にほぼ 1 であり、論文数が急増した 1990 年までは PageRank が良い相関を示している。PageRank に次いで、Betweenness Centrality、Core Number、Load Centrality が弱い相関を示している。

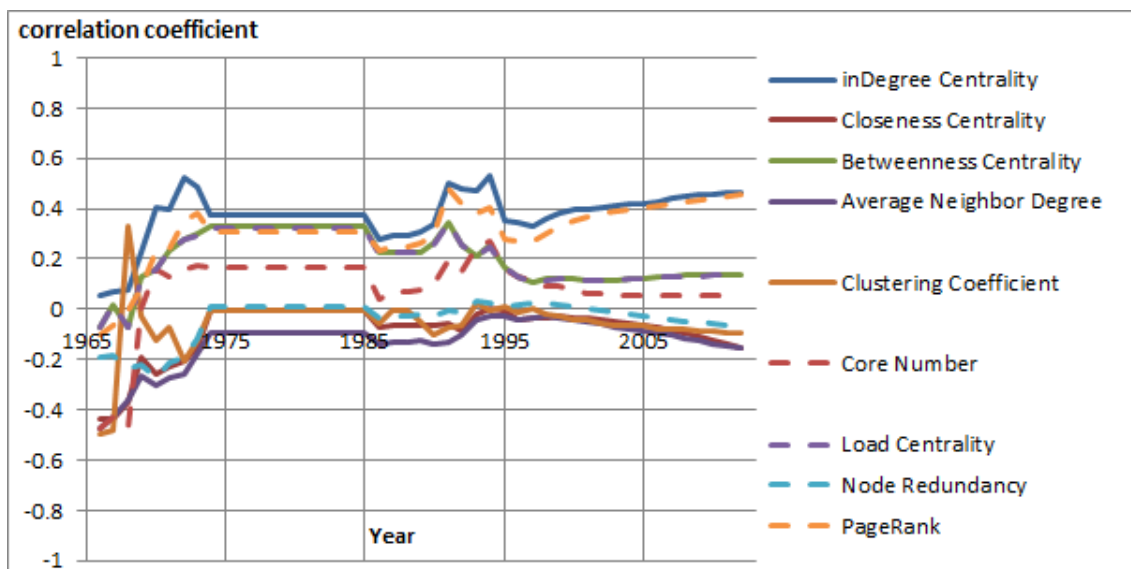


Figure 93 : Correlation Coefficient between Current Number of Citations and Indicators [2010med]

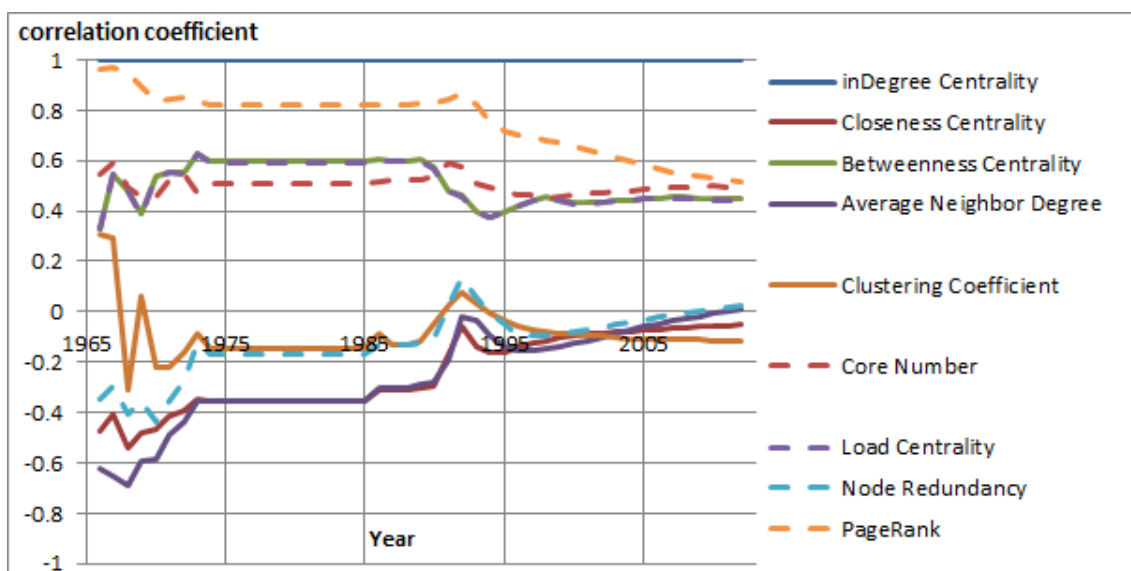


Figure 94 : Correlation Coefficient between Citations and Indicators at Each Year within Dataset [2010med]

3.8. まとめ

以上の結果より、in-Degree Centrality、PageRank、Communicability Centrality が萌芽的論文の検知に有効であるものと考えられる。これら 3 指標のうち、in-Degree Centrality は被引用数を正規化したものであり、それに対し重要論文が他の多くの論文から引用されることから、in-Degree Centrality にも反映されることから一定の有効性を持つことは予想された結果であるが、正規化した値を時系列で表示したことにより、(1)出版直後に高い in-Degree Centrality を持った論文は、初期での検知が容易になり、(2)出版後、数年を経て in-Degree Centrality が増加する論文は時系列で追いつけることで、重要性を増したことが検知しやすくなる。

PageRank についても、in-Degree Centrality と同様の傾向を持ち、時系列により萌芽的論文の検知に有効性を持つことが見出された。

Communicability Centrality は、萌芽的論文の検知に有効ではあるが、データセットのサイズが大きくなるほど計算量が増え合理的な時間内に計算が完了しないため、実際の利用には難しいと考えられる。前節までのデータセットにおいても、論文数が 2 万件を切る場合のみ計算している。

先の 3 指標の他、Betweenness Centrality、および、Load Centrality は、最新年で引用ネットワークに 2 万件以上の論文が含まれる場合に、各年のデータセット内の被引用数と指標の相関係数において弱い相関を示すようになる。また、Core Number は、最新年で引用ネットワークに 4 万件以上の論文が含まれる場合に、各年のデータセット内の被引用数と指標の相関係数において弱い相関を示すようになる。

4. 引用ネットワークの時系列分析の考察

本章では、複数データセットを横断して比較することで、個別のデータセットで抽出した萌芽的論文の特徴が重要論文としての一般性を持つかを確認し、さらに、分野による相違を特定する。また、萌芽的論文の特定に有用な指標を用いた場合に、萌芽的論文の検知に時期について考察する。本章の最後に、萌芽的論文の特徴について総括し、次章における検知に有効な特徴量の導出につなげる。

4.1. 指標の一般性の検証

前章において、個別の学術分野について時系列分析の結果を示し、**in-Degree Centrality**、**PageRank**、および、**Communicability Centrality** が萌芽的論文の検出に利用できる可能性について共通性を見出したが、前章では、ノーベル賞受賞論文や ACM チューリング賞受賞論文などは、重要論文の中でも世界的な賞を受賞した論文(以降、特別な重要論文)であり、データセットに数万本の論文のうちの数本でしかない。しかしながら、特別な重要論文の他にも、その学術分野を補完し、あるいは、その知見が広く認知されるのに貢献した論文など、特別な重要論文に準ずる一般の重要論文も存在するはずである。

本節では、これら指標のどのくらいの重要論文まで活用できるか、被引用数が多い論文を重要論文として、適合率、再現率、平均逆順位を用いて指標の一般性を検証する。

4.1.1. 適合率と再現率

本項では、重要論文の判断時を 2 年前、5 年前、10 年前の時点に置き、その時点の指標から、結果である現在の最新被引用数がどの程度、特定可能であるか、適合率(precision)と再現率(recall)を算出する(Manning, 2008)。各指標の適合率、および、再現率は、Figure 95 のように、2 年前、5 年前、10 年前のそれぞれの時点における指標の値と、Web of Science 内の最新被引用数を用いて算出する。

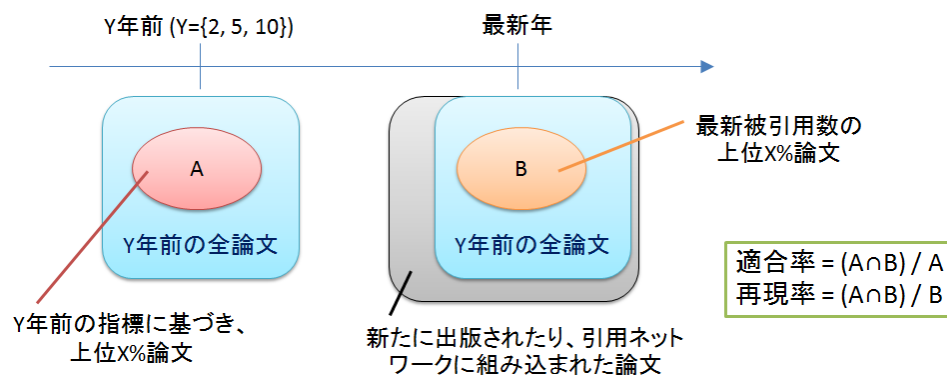


Figure 95： 適合率と再現率の算出方法

各指標の適合率は、Y年前の3指標(in-Degree Centrality、PageRank、および、Communicability Centrality)それぞれの指標の値を大きい方から X%の論文と、最新被引用数の値を大きい方から X%の論文と、どれだけ合致したか、Xの値を変動させて計算した結果の図一覧が、Table 25 である。各指標の再現率は、Y年前の3指標(in-Degree Centrality、PageRank、および、Communicability Centrality)それぞれの指標の値を大きい方から X%の論文と、最新被引用数の値を大きい方から X%の論文と、どれだけ合致したか、Xの値を変動させて計算した結果の図一覧が、Table 26 である。適合率と再現率の両図一覧とも、in-Degree Centrality を青線、PageRank を赤線、および、Communicability Centrality をグリーン破線で示す。なお、これら適合率と再現率の算出の際、指標の年に値が算出されていない論文は指標の年の時点で重要論文の候補に含まれていないため、適合率と再現率の算出にも使用しない。また、G タンパク質共益受容体(2012chem)と体外受精技術(2010med)は、前章のクラスター分析の内容を見ると、1990年代前半の論文数急増がありデータセットが年を経る中で変質したと予想されるが、内容に関して専門家の視点を入れることが難しい為、Table 25 と Table 26 では除外した。

年差による違いを比較すると、分野の成果が安定してきている蛍光プロテイン(gfp)や準結晶(2011chem)は、適合率が指標と最新被引用数の年差の違いに影響されにくい。現在でも盛んに論文が発表されている iPS 細胞(2012med)では、指標と最新被引用数の年差が開くと適合率が悪化する傾向がある。

3指標を比較すると、一般的には、in-Degree Centrality と Communicability Centrality が適合率が良く、再現率は in-Degree Centrality がよい。しかしながら、PageRank はデータセッ

トによって適合率の良否に振幅があるものの、暗号(crypto)と準結晶(2011chem)と成果の安定している分野においては適合率が良い。ただし、Communicability Centrality は、合理的な時間内に計算が完了しないため、引用ネットワークの論文数が2万未満のデータセットのみについて、適合率・再現率を計算している。

Table 25: 指標の適合率一覧

	2年差	5年差	10年差
crypto			
sfjp			
2012med			
2011chem			

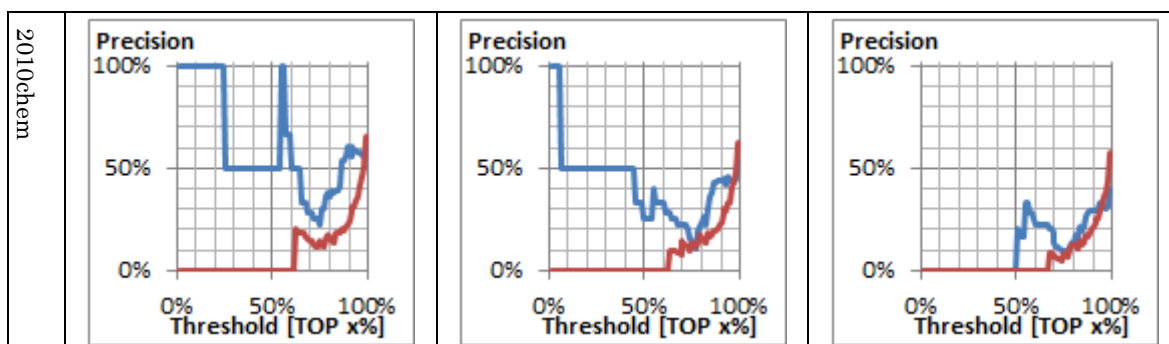
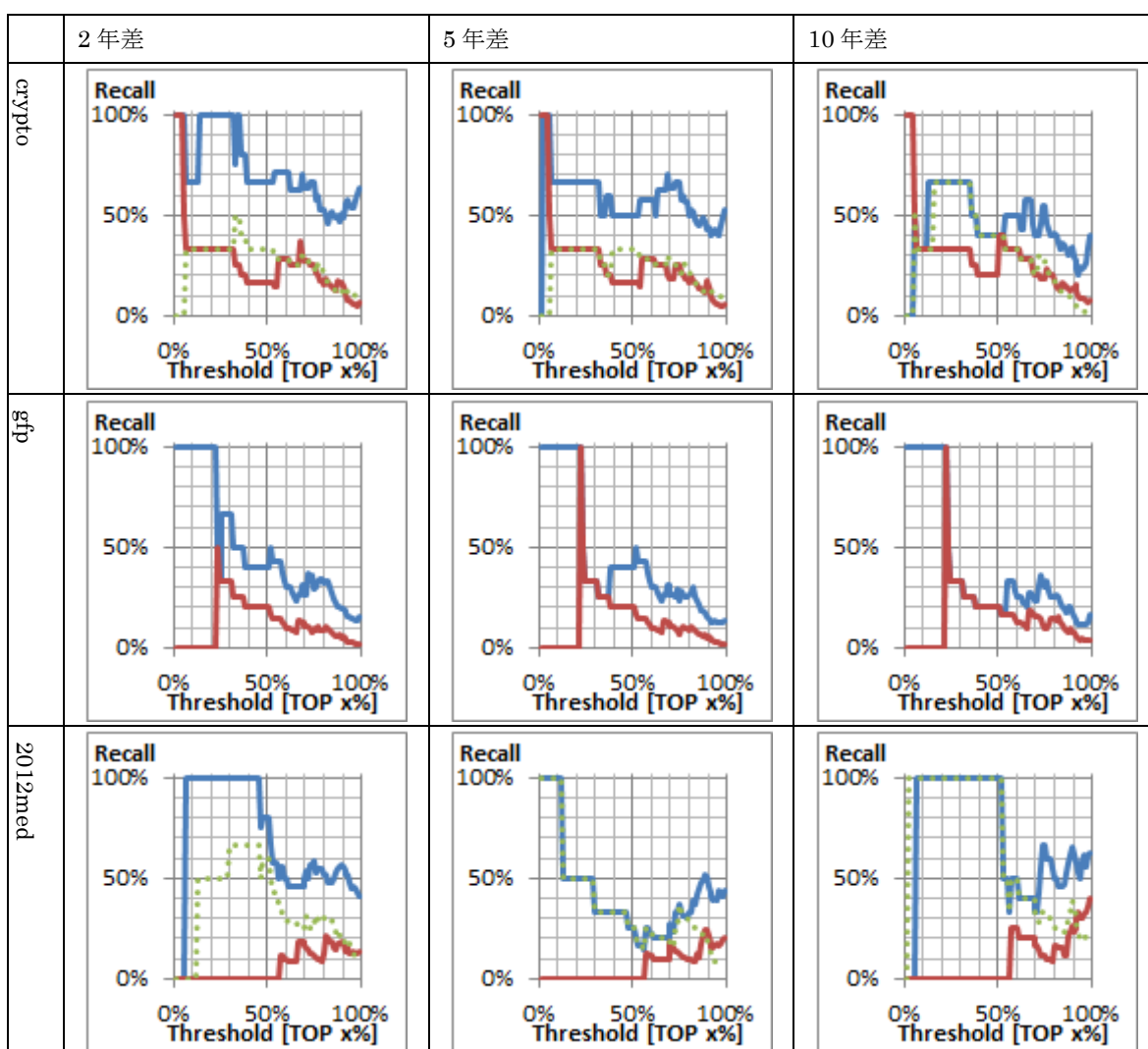
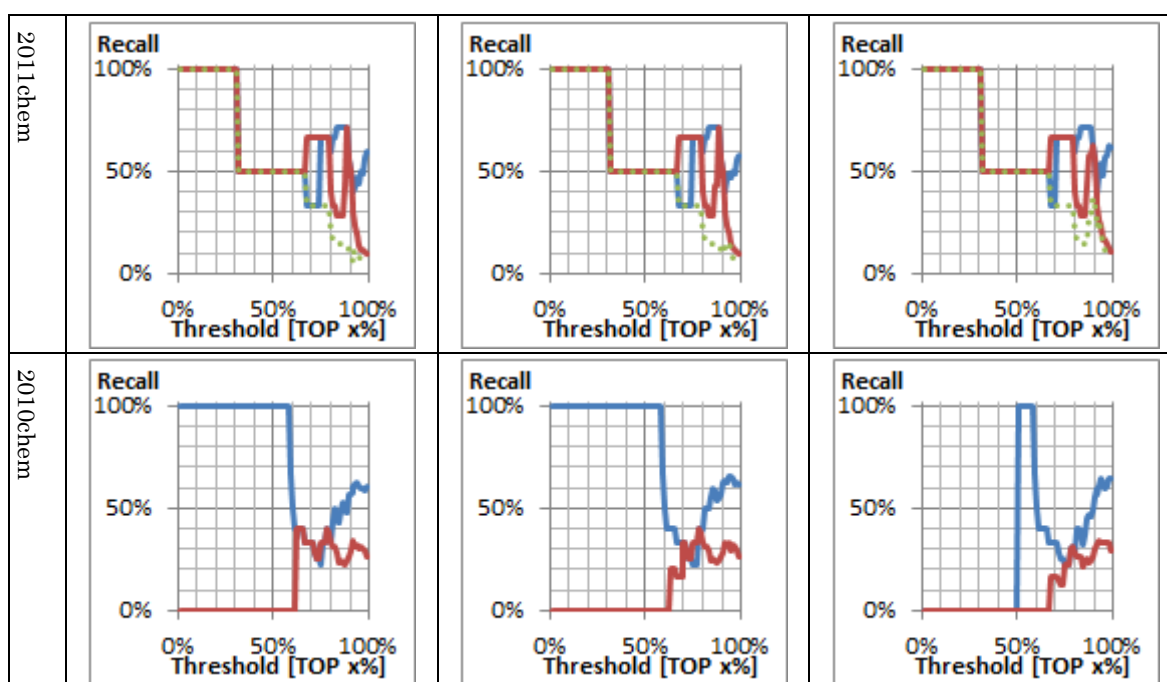


Table 26: 指標の再現率一覧





4.1.2. 平均逆順位 (MRR; MEAN RECIPROCAL RANK)

前項の適合率と再現率は、順位ではなく値で X% という閾値を設けたため、X の値が小さいほど少ない上位論文が配分されることになり極端な値を示すことになる。そこで、本項では、順位に関して評価に平均逆順位を用い、指標がどの程度の有効性を持つかを一覧する。逆順位 (Reciprocal Rank) という情報検索尺度は、文書の順位の逆数を計算する。文書が 1 位の場合は 1 で、2 位の場合は 0.5 という値を持つ。平均逆順位は、逆順位の平均で求められ、次の式で計算される (Springer, 2013)。

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

本項では、Figure 96 のように、最新年における指標の値と最新年におけるデータセット内の被引用数との、それぞれ TOP 10 論文の順位が合致した場合には逆順位を算出し、合致しない場合には逆順位を 0 として、平均逆順位を算出している。ただし、前項の適合率と再現率と、本項での平均逆順位が異なるのは、被引用数が Web of Science 内ではなくデータセット内の被引用数であり、年差の途中で出版された論文を除外してはいない。

Table 27 に、平均逆順位が具体的にどのような値も持つか、サンプルとしての理論値を示し、Table 28 に実際の計算値を示す。順位の合致を求める場合、PageRank や Communicability Centality に in-Degree Centrality ほどの精度は求めえないが、PageRank において、前節において in-Degree Centrality より良い値を示した暗号(crypto)と準結晶(2011chem)については値を持っている。

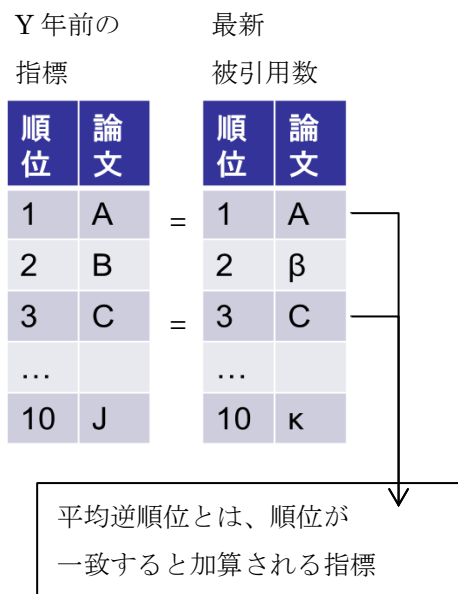


Figure 96 : MRR の概念

Table 27 : MRR が持つ値(理論値)

順位	MRR (a)	MRR (b)
1		1.0000
2		0.2500
3		0.1111
4		0.0625
5		0.0400
6		0.0278
7		0.0204
8		0.0156
9		0.0123
10	0.2929	0.0100

* (a)は、当該順位までの全てが合致した場合の値、(b)は、当該順位のみが合致した場合の値

Table 28 : TOP 10 論文の平均逆順位

in-Degree Centrality				
	0年差	2年差	5年差	10年差
crypto		0.2929	0.0458	0
gfp		0.2561	0.2550	0.1000
2012chem		0.1492	0.1617	0
2012med		0.1450	0	0
2011chem		0.2550	0.1167	0.1000
2010chem		0.2226	0.2326	0.0111
2010med		0.1500	0.1200	0.1000
PageRank				
	0年差	2年差	5年差	10年差
crypto	0.2287	0.2287	0.2033	0.2287
gfp	0	0	0	0
2012chem	0	0	0	0
2012med	0	0	0.0100	0
2011chem	0.1611	0.1611	0.1611	0.1611
2010chem	0	0	0	0
2010med	0	0	0	0
Communicability Centrality				
	0年差	2年差	5年差	10年差
crypto	0	0	0	0
2012med	0	0.1500	0	0
2011chem	0.1833	0.1833	0.1000	0.1300

* Communicability Centrality は、データセットの論文が多いほど算出に時間がかかり、合理的な時間内に終了しないため、2万本以上の論文を含むデータセットでは算出していない。

4.2. 検知の時期

Table 29 に、in-Degree Centrality と PageRank の時系列分析の図から萌芽的論文の検知できる時期を示す。検知年は、他の論文と比べて高い値を持ち始めた年とした。数値で明確に示すならば、上位 10 位に入った年としたいところではあるが、分野の登場初期において指標の値が

低い論文数が少ない為に無為に上位に入るため、単純に上位で判断することは難しいためである。

In-Degree Centrality と PageRank を比較すると、in-Degree Centrality の方が早く検知できる傾向にある。PageRank は検知時期としては遅れるものの、in-Degree Centrality より他の論文と区別できるために検知できることがある。検知の時期が判断でき出版直後から指標の値を持つ論文の in-Degree Centrality を見ると、G タンパク質共益受容体(2012chem)は5~7年と遅めではあるが、1~3年と早期に検知できている。

Table 29: 萌芽的論文の検知できる時期

データセット	筆頭著者	出版年	In-Degree 検知年	In-Degree 経過年数	PageRank 検知年	PageRank 経過年数
crypto	Rivest	1978	1980	2	1983	5
gfp	Shimomura (※1)	1962	1974	12	1974	12
gfp	Chalfie	1994	1995	1	1996	2
gfp	Tsien	1998	2001	3	2002	4
2012chem	Lefkowitz	1970a	検知不可	-	1977 (※3)	7
2012chem	Lefkowitz	1970b	検知不可	-	1977 (※3)	7
2012chem	De Lean	1980	1993	7	検知不可	-
2012chem	Dixon	1986	1991	5	1996 (※3)	10
2012chem	Kobilka	1987	1993	7	検知不可	-
2012med	Gurdon (※1)	1962	検知不可	-	1975	13
2012med	Takahashi	2006	2008	2	2008	2
2011chem	Shechtman	1984	1985	1	1985	1
2010chem	Heck (※1)	1972	検知不可(※2)	-	1977	5
2010chem	Negishi	1977	検知不可(※2)	-	2006 (※3)	29
2010chem	Miyaura	1979a	検知不可(※2)	-	検知不可	-
2010chem	Miyaura	1979b	検知不可(※2)	-	検知不可	-
2010med	Edwards	1965	検知不可(※4)	-	検知不可(※4)	-
2010med	Edwards	1969	検知不可(※4)	-	検知不可(※4)	-
2010med	Edwards	1970	検知不可(※4)	-	検知不可(※4)	-
2010med	Steptoe	1978	検知不可(※4)	-	検知不可(※4)	-

(※1) 引用ネットワークの最大連結成分に含まれ始めた年。出版年からしばらく引用ネットワークが形成できない、あるいは、引用ネットワークの最大連結成分に含まれず、指数が算出されていない場合がある。

(※2) 前章であらかじめ設定した論文では検知は難しいが、受賞者の後年に発表された論文では検知可能な論文があった。

(※3) 二番手の集団にあり、見落とす可能性も高い。

(※4) 初期に論文数が少なく区別が難しい。

4.3. 社会的認知度による違い

準結晶(2011chem)と暗号分野(crypto)の in-Degree Centrality および PageRank を比較すると、重要論文(受賞論文、および、最新被引用数上位の論文)の形状に大きな違いがある。

準結晶は、論文発表当初はその内容について査読者から信憑性が疑われたほど、その研究内容が社会的に認知されていなかった。この場合、出版年の初年に in-Degree Centrality および PageRank が上昇し、その後、他の論文に比べれば高い値を持つものの関連論文の登場により急激に減衰している(Figure 66、Figure 67)。一方、暗号分野は、古代ギリシャ・ローマに起源があり、良く知られた分野である。この場合、長い時間かけて値の頂点に達し、その後、なだらかに下っていく。暗号分野では、値の最大値が RSA 暗号の論文が、発行年の 11 年後に値の最大値を持った(Figure 21、Figure 22)。これは、認知度の低い分野では、萌芽的論文が唯一の引用元となるが、時間が経つにつれて萌芽的論文を引用しない論文(以降、雑音とよぶ)が増加するためと考えられる。一方、認知度の高い分野では、世界的な賞を受賞する重要論文が登場した当初にも、引用ネットワーク内に雑音が多く存在しているため、分野の中で論文自体が認知される時間を必要とし、後年に最高値に到達するものと考えられる。Figure 97 は、社会的認知度の違いによる形状の違いについて模式的に示す。

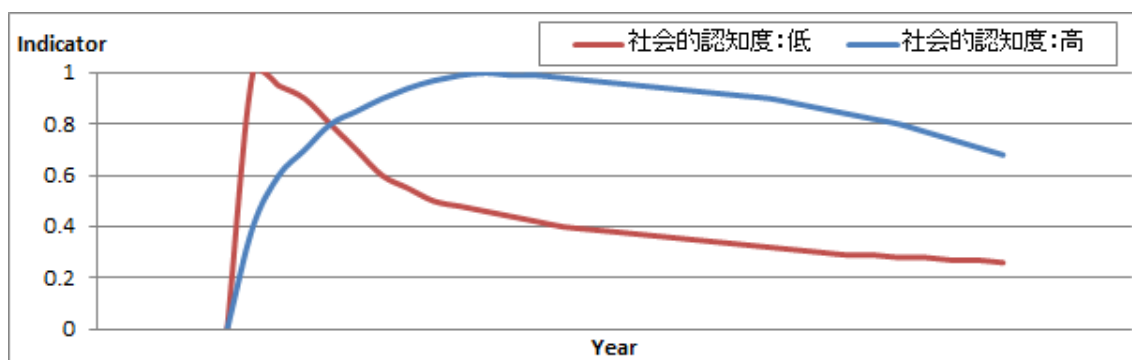


Figure 97: 社会的認知度の違いによる、in-Degree Centrality および PageRank の形状の違い

理論的には社会的認知の違いにより形状が異なることが説明できるものの、各データセットについて受賞論文が出版年からどのくらいで **in-Degree Centrality** および **PageRank** の最大値を持つ年を調べると、Table 30 のように、一般性を論じることは難しい。実際の分析においては、蛍光プロテイン(gfp)のように、分野が誕生した当初は認知度が低いがゆえに、分析対象となる引用ネットワークの最大連結成分に入れられないこともある。本研究の提案する手法で萌芽的論文を特定する際には、第2以降の連結成分も分析する必要があるものとする。なお、Table 30 においては、各論文の学術分野について、既に知られた分野か否かは、その研究が当時の主流派から外れた観点から研究されたか否かのエピソードに基づいて判断している。

Table 30: 指標が最大値に達するまでの期間

データセット	筆頭著者	出版年	登場年 (※1)	In-Degree 最高年	In-Degree 到達期間	PageRank 最高年	PageRank 到達期間	認知 度低 ／高
crypto	Rivest	1978	1978	1986	8	1987	9	高
gfp	Shimomura	1962	1974	1974	12	1974	12	低
gfp	Chalfie	1994	1994	1997	3	1998	4	高
gfp	Tsien	1998	1998	2004	6	2011	13+	高
2012chem	Lefkowitz	1970a	1971	1975	5	1978	8	高
2012chem	Lefkowitz	1970b	1971	1975	5	1977	7	高
2012chem	De Lean	1980	1980	1989	9	1999	19	高
2012chem	Dixon	1986	1986	1991	5	2006	20	高
2012chem	Kobilka	1987	1987	1992	5	2002	15	高
2012med	Gurdon (※2)	1962	1995	2006	44	1975,1976	13	-
2012med	Takahashi	2006	2006	2012	6+	2012	6+	高
2011chem	Shechtman	1984	1985	1985	1	1985	1	低
2010chem	Heck	1972	1977	1977	5	1978	6	高
2010chem	Negishi	1977	1977	1981	4	1989	12	高
2010chem	Miyaura	1979a	1979	1983	4	1987 (※3)	8	高
2010chem	Miyaura	1979b	1979	1982	3	2012 (※4)	33	高
2010med	Edwards	1965	1966	1970	5	1966	1	-
2010med	Edwards	1969	1969	1975-1985	6	1969	0	高
2010med	Edwards	1970	1970	1973	3	1973	3	高

2010med	Steptoe	1978	1986	1989	11	1989	11	高
---------	---------	------	------	------	----	------	----	---

(※1) 引用ネットワークの最大連結成分に含まれ始めた年。出版年からしばらく引用ネットワークが形成できない、あるいは、引用ネットワークの最大連結成分に含まれず、指数が算出されていない場合がある。

(※2) 1975年～1976年に引用ネットワークの最大連結成分に含まれた後、一旦、最大連結成分に含まれず、1995年以降に再度、含まれるようになった。

(※3) 1990年代に再び増加し始め、2012年以降に最大値を迎える兆しが見られる。

(※4) 1981年に一度、ピークを迎えるが、1990年代に再び増加し始め、2012年も増加傾向の中にある。

4.4. まとめ

以上の考察から、重要論文は、in-Degree Centrality、PageRank、Communicability Centrality の3指標のいずれかを用いて、時系列分析にて高い値を持つ論文として特定することができる。In-Degree Centrality が最も精度が高く、Communicability Centrality、PageRank の順で精度が低くなるが、いずれも重要論文を検知できる。Communicability Centrality は、論文数が大きなデータセットの場合、合理的な時間内に計算が終了しないことがあるので、本研究の結果からは、in-Degree Centrality、または、PageRank の利用が実用的であると考えられる。PageRank は、データセットに依存して in-Degree Centrality より精度が高い場合や検出しやすい場合がある。検知の時期については、in-Degree Centrality と PageRank とで比較すると、in-Degree Centrality の方が早く検知できる傾向にある。

この手法では、本研究で萌芽的論文の正解として選択した論文は、重要論文の一つである最新被引用数が高い論文と必ずしも一致するものではない。しかしながら、本研究で萌芽的論文の正解として選択した論文は、少なくとも100を越える最新被引用数を持っていた。

また、学術分野の社会的認知度が高いと、10年以上が経過しても in-Degree Centrality、および、PageRank は上昇し続ける場合がある。Vanclay(2011)が、今日、一般的に使用されている Impact Factor が直近2年または5年の被引用数で算出されていることに疑念を提起してことを裏付けられよう。特別な重要論文を特定するには、厳密には、2年や5年では不十分である。

なお、制限事項としては、学術分野の登場初期に引用ネットワークが形成できなかつたり、同じく初期に萌芽的論文が2番以降の連結成分に含まれてしまい、萌芽的論文の特定が遅れる場合がある。また、Web of Science に収録されていない論文や書籍が重要な文献であっても特定できない。例えば、社会科学・人文科学分野の現代正義論では、ジョン・ロールズが橋頭堡的な理論を発表し、アマルティア・センが最新の理論を発表しているが、書籍が特別に重要な場合でも Web of Science に収録されていないので、重要な理論として特定することは難しい。

次の第5章では、ここまでの議論で有効と認められた in-Degree Centrality と PageRank について、さらに追究する。

5. 萌芽的論文特定のための時系列変化特徴量の考察

前章までは、in-Degree Centrality や PageRank などの指標を時系列で可視化した図を観察し、考察を行ってきた。本章では、前章までの考察の結果に基づき、前章において有効と示された指標群について、データセット内における萌芽的論文の客観的な特定に寄与すると考えられる候補の特徴量群を提案し、前章と同じ7つのデータセットを用いて、それらの有効性について実際に検証を行い、萌芽的論文の特定に有効な特徴量群を議論する。

5.1. 特徴量の設計

ここまでの議論においては、受賞論文が他の論文より高い指標の値を持つことから、時系列分析の図を見て萌芽的論文を判断していた。しかしながら、発行年の異なる論文を比較する場合に、各年の状況は論文数も時代背景も異なり、そのまま比較するのは難しい。ゆえに、その分野に知見が十分でない人や自動的に萌芽的論文を特定するには、一元化して判断できる数値が要求される。

本章で特徴量を設定するにあたっては、多くの論文では、Figure 98 のように最大値を迎えるまでに、in-Degree Centrality で 5 年前後、PageRank で 10 年前後である。本研究では、重要論文の中でも萌芽的である需要論文に注目していることから、in-Degree Centrality や PageRank が最大値を迎えるまでの値で特徴量を設定する。設定する特徴量は、In-Degree Centrality、PageRank、および、比較のため、データセット内での被引用数(Number of Citation)について、Height、Area(参考値)、Slope、SlopeAve を以下のように定義した。

- **Height**：指標の値が時系列変化する中で各論文の最も高い値。前章までの議論で重要論文を判断する材料として使われてきた値となる。
- **Area(参考値)**：各論文の時系列変化の線と時間軸と間の面積。最新年まで計算しており、最新年における指標によるインパクトを示すこととなる。In-Degree Centrality、PageRank、では登場初期を、データセット内での被引用数では最新年を重み付けた値となる。原理的には、in-Degree Centrality、PageRank において、萌芽的論文で高い値となる。設定意図から考えると、指標の値が最大値に到達するまでで算出すべきであるが、そうすると、短期で in-Degree Centrality、PageRank の最大値に到達する萌芽

的論文では小さくなることが予想され、萌芽的でない重要論文のための特徴量となってしまうため、最新年までの計算となっている。他の特徴量と合わせて、より精度を高めるために使う事が想定される補助的な特徴量である。

- Slope：各論文の指標の増加率の最大値。認知度が低い分野において、論文の発表直後に急峻な in-Degree Centrality、PageRank の上昇が見られるため、萌芽的論文で高い値になることが期待される。
- SlopeAve：各論文の指標の増加率を指標が最大値に到達するまでを平均したもの。Slope と同様に、論文の発表直後に急峻に上昇する in-Degree Centrality、PageRank が反映するものと期待される。

なお、これら特徴量の算出にあたっては、計算量を可能な範囲に抑えるために、現在の被引用数が 100 を超える論文のみを対象にしている。時系列変化の観察から、設定した特徴量が低くなることが予想され、本節の議論には影響しないと考えられる。

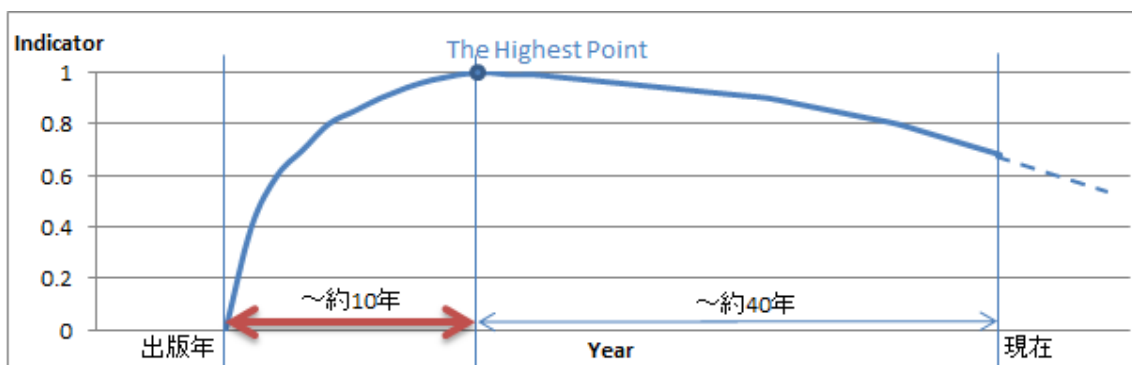


Figure 98 : in-Degree Centrality や PageRank の一般的な形状

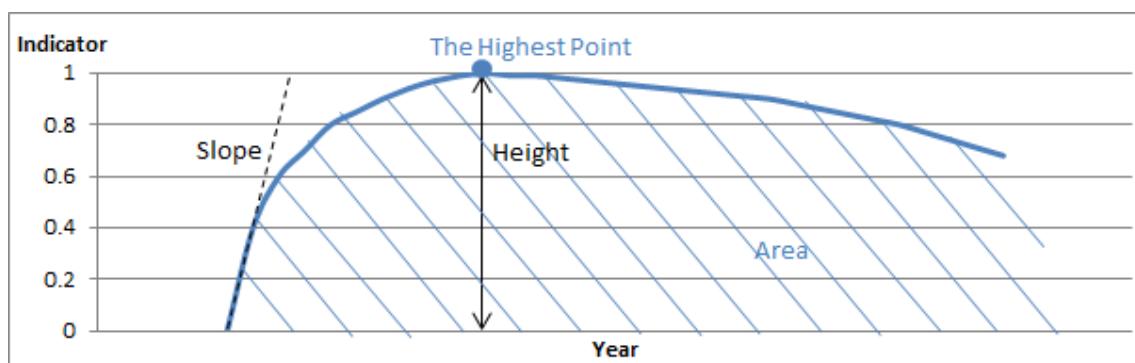


Figure 99 : Height, Area, Slope の定義

5.2. 特徴量の上位論文、および、受賞論文のリスト

全ての論文に対し特徴量を計算した結果、上位 10 論文、および、受賞論文を Table 31 に示す。各表において、Height、Area、Slope、SlopeAve の順に横に並べて示す。Rank の背景に着色のあるものは受賞論文であり、各表内において同じ論文を表している。縦には、in-Degree Centrality、PageRank、データセット内での被引用数(Number of Citation)を分けて表にまとめ、7 データセットについて示す。左端の着色されたバーは、データセットごとのまとまりが判別しやすいように設置したものである。

Table 31: 各リストにおける上位 10 位論文、および、受賞論文

gfp (1,967 papers)

In-Degree Centrality

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	SHIMOMURA, O	1962	0.7333	1	SHIMOMURA, O	1962	13.2249	1	SHIMOMURA, O	1962	0.7333	1	HEIM, R	1994	49.6667
2	SHIMOMURA, O	1963	0.6000	2	SHIMOMURA, O	1963	7.9612	2	SHIMOMURA, O	1963	0.6000	2	Patterson, GH	1997	23.5833
3	CHALFIE, M	1994	0.4380	3	SHIMOMUR.O	1969	5.7043	3	SHIMOMUR.O	1969	0.3333	3	SHEEN, J	1995	12.5714
4	SHIMOMUR.O	1969	0.3704	4	CHALFIE, M	1994	3.4929	4	CHALFIE, M	1994	0.1911	4	DELAGRAVE, S	1995	12.0667
5	INOUYE, S	1985	0.2441	5	BLINKS, JR	1976	3.2939	5	WANG, SX	1994	0.1300	5	SHIMOMURA, O	1962	11.3056
6	BLINKS, JR	1976	0.2414	6	SHIMOMURA, O	1975	2.7530	6	HEIM, R	1994	0.1084	6	Gerdes, HH	1996	9.6923
7	SHIMOMURA, O	1975	0.2362	7	INOUYE, S	1985	1.7498	7	PRASHER, DC	1992	0.1011	7	OGAWA, H	1995	9.5714
8	PRASHER, DC	1992	0.2309	8	PRASHER, DC	1992	1.7384	8	CUBITT, AB	1995	0.0736	8	Siegel, MS	1997	7.9167
9	HEIM, R	1994	0.1900	9	WARD, WW	1980	1.4662	9	Cormack, BP	1996	0.0712	9	Cheng, LZ	1996	6.8462
10	WANG, SX	1994	0.1734	10	SHIMOMURA, O	1979	1.4048	10	CODY, CW	1993	0.0701	10	AMSTERDAM, A	1995	6.7143
21	Tsien, RY	1998	0.0724	15	Tsien, RY	1998	0.8389	37	Tsien, RY	1998	0.0259	307	CHALFIE, M	1994	0.1861

347	Tsien, RY	1998	0.0138
-----	-----------	------	--------

PageRank

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	SHIMOMURA, O	1962	0.2276	1	SHIMOMURA, O	1962	4.7741	1	SHIMOMURA, O	1962	0.2276	1	SHIMOMURA, O	1962	11.3056
2	SHIMOMURA, O	1963	0.1092	2	SHIMOMURA, O	1963	1.8754	2	SHIMOMURA, O	1963	0.1092	2	Siegel, MS	1997	7.9167
3	SHIMOMUR.O	1969	0.0625	3	SHIMOMUR.O	1969	1.0662	3	SHIMOMUR.O	1969	0.0625	3	Stauber, RH	1998	5.6364
4	CHALFIE, M	1994	0.0515	4	CHALFIE, M	1994	0.6874	4	STEINER, RF	1962	0.0396	4	Oparka, KJ	1997	5.5000
5	STEINER, RF	1962	0.0396	5	SHIMOMURA, O	1979	0.6417	5	CHALFIE, M	1994	0.0197	5	Burke, NV	1997	5.2727
6	PRASHER, DC	1992	0.0377	6	BLINKS, JR	1976	0.6144	6	SHIMOMURA, O	1975	0.0128	6	WARD, WW	1982	4.6786
7	BLINKS, JR	1976	0.0356	7	PRASHER, DC	1992	0.5304	7	PRASHER, DC	1992	0.0127	7	OLSON, KR	1995	4.5000
8	SHIMOMURA, O	1979	0.0333	8	WARD, WW	1980	0.5258	8	BLINKS, JR	1976	0.0118	8	BRINI, M	1995	4.4000
9	WARD, WW	1980	0.0279	9	STEINER, RF	1962	0.4371	9	SHIMOMURA, O	1979	0.0097	9	MONTERO, M	1995	4.1429
10	SHIMOMURA, O	1975	0.0237	10	SHIMOMURA, O	1975	0.3620	10	WARD, WW	1980	0.0087	10	Klein, RL	1998	4.0909
27	Tsien, RY	1998	0.0049	29	Tsien, RY	1998	0.0483	66	Tsien, RY	1998	0.0009	279	CHALFIE, M	1994	0.0153
												399	Tsien, RY	1998	0.0004

Number of Citation

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	CHALFIE, M	1994	2371	1	CHALFIE, M	1994	22918	1	CHALFIE, M	1994	261	1	Tsien, RY	1998	147
2	Tsien, RY	1998	1899	2	Tsien, RY	1998	12740	2	Tsien, RY	1998	165	2	CHALFIE, M	1994	141
3	Cormack, BP	1996	1196	3	Cormack, BP	1996	10894	3	Shaner, NC	2004	155	3	Shaner, NC	2004	123
4	PRASHER, DC	1992	1028	4	PRASHER, DC	1992	10116	4	Cormack, BP	1996	151	4	Campbell, RE	2002	84
5	Ormo, M	1996	949	5	HEIM, R	1994	7819	5	PRASHER, DC	1992	124	5	Cormack, BP	1996	81

6	HEIM, R	1994	820	6	Ormo, M	1996	7590	6	Campbell, RE	2002	115	6	Shaner, NC	2005	70
7	Shaner, NC	2004	772	7	CUBITT, AB	1995	7256	7	Shaner, NC	2005	94	7	Ormo, M	1996	64
8	Campbell, RE	2002	737	8	Miyawaki, A	1997	5391	8	Nagai, T	2002	90	8	Matz, MV	1999	61
9	Miyawaki, A	1997	718	9	Yang, F	1996	5369	9	CUBITT, AB	1995	89	9	PRASHER, DC	1992	58
10	Matz, MV	1999	715	10	Crameri, A	1996	4932	10	Ormo, M	1996	85	10	Nagai, T	2002	58
17	SHIMOMURA, O	1962	469	12	SHIMOMURA, O	1962	4347	16	SHIMOMURA, O	1962	58	139	SHIMOMURA, O	1962	11

in-vitro fertilization [2010med] (2,546papers)

In-Degree Centrality

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	YANAGIMACHI, R	1964	0.5714	1	YANAGIMACHI, R	1964	8.1125	1	YANAGIMACHI, R	1964	0.4444	1	CHA, KY	1991	11.8947
2	WHITTING.DG	1968	0.2857	2	WHITTING.DG	1968	6.1044	2	WHITTING.DG	1968	0.1304	2	TAKAHASHI, Y	1992	11.7222
3	EDWARDS, RG	1965	0.2258	3	EDWARDS, RG	1965	4.5501	3	EDWARDS, RG	1965	0.1111	3	MUNNE, S	1993	8.0000
4	EDWARDS, RG	1970	0.2083	4	EDWARDS, RG	1969	4.4634	4	EDWARDS, RG	1965	0.1111	4	LASSALLE, B	1985	5.9500
5	EDWARDS, RG	1969	0.2078	5	BARROS, C	1967	4.1459	5	BARROS, C	1967	0.1025	5	BRACKETT, BG	1993	5.1667
6	BARROS, C	1967	0.1935	6	EDWARDS, RG	1970	3.8335	6	DONAHUE, RP	1968	0.0962	6	YOSHIDA, M	1992	5.1111
7	EDWARDS, RG	1965	0.1935	7	EDWARDS, RG	1965	3.2757	7	EDWARDS, RG	1969	0.0870	7	LANE, M	1992	5.0000
8	IWAMATSU, T	1971	0.1169	8	IWAMATSU, T	1971	2.1707	8	EDWARDS, RG	1970	0.0789	8	FUNAHASHI, H	1993	4.9444
9	DONAHUE, RP	1968	0.1039	9	DONAHUE, RP	1968	2.0452	9	IWAMATSU, T	1971	0.0780	9	CHO, WK	1974	4.6190
10	LASSALLE, B	1985	0.0676	10	YANAGIMA, R	1972	0.7116	10	LASSALLE, B	1985	0.0676	10	LEIBO, SP	1993	4.6111
11	STEPTOE, PC	1978	0.0476	15	STEPTOE, PC	1978	0.3992	19	STEPTOE, PC	1978	0.0244	201	STEPTOE, PC	1978	11.0000
												1719	EDWARDS, RG	1965	5.0000

1774	EDWARDS, RG	1969	5.0000
2103	EDWARDS, RG	1970	3.0000

PageRank

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	YANAGIMACHI, R	1964	0.1720	1	YANAGIMACHI, R	1964	2.6731	1	YANAGIMACHI, R	1964	0.1486	1	CHA, KY	1991	11.8947
2	EDWARDS, RG	1965	0.0645	2	EDWARDS, RG	1965	1.1740	2	EDWARDS, RG	1965	0.0645	2	TAKAHASHI, Y	1992	11.7222
3	EDWARDS, RG	1965	0.0645	3	WHITTING.DG	1968	0.9678	3	EDWARDS, RG	1965	0.0645	3	LASSALLE, B	1985	5.9500
4	BARROS, C	1967	0.0501	4	BARROS, C	1967	0.8851	4	BARROS, C	1967	0.0501	4	BRACKETT, BG	1993	5.1667
5	WHITTING.DG	1968	0.0402	5	EDWARDS, RG	1965	0.8823	5	DONAHUE, RP	1968	0.0333	5	YOSHIDA, M	1992	5.1111
6	DONAHUE, RP	1968	0.0333	6	EDWARDS, RG	1969	0.6433	6	WHITTING.DG	1968	0.0333	6	LANE, M	1992	5.0000
7	EDWARDS, RG	1969	0.0328	7	DONAHUE, RP	1968	0.6229	7	EDWARDS, RG	1969	0.0328	7	FUNAHASHI, H	1993	4.9444
8	BOWMAN, P	1970	0.0326	8	EDWARDS, RG	1970	0.5346	8	BOWMAN, P	1970	0.0326	8	LEIBO, SP	1993	4.6111
9	LASSALLE, B	1985	0.0288	9	FOOT, WD	1969	0.2914	9	LASSALLE, B	1985	0.0288	9	KRANZ, E	1991	4.6000
10	EDWARDS, RG	1970	0.0270	10	DONAHUE, RP	1968	0.2474	10	DONAHUE, RP	1968	0.0241	10	TESTART, J	1986	4.4500
13	STEPTOE, PC	1978	0.0149	15	STEPTOE, PC	1978	0.1400	12	EDWARDS, RG	1970	0.0142	343	EDWARDS, RG	1969	0.0269
								17	STEPTOE, PC	1978	0.0091	345	EDWARDS, RG	1965	0.0209
												347	EDWARDS, RG	1970	0.0054
												365	STEPTOE, PC	1978	0.0009

Number of Citation

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	VANSTEIRTEGH	1993	488	1	VANSTEIRTEGHE	1993	6166	1	INSALL, JN	1983	115	1	Hansen, M	2002	37

	EM, AC				M, AC										
2	Hansen, M	2002	371	2	EDWARDS, RG	1965	3735	2	Hansen, M	2002	56	2	Broekmans, FJ	2006	36
3	STEPTOE, PC	1978	342	3	VANSTEIRTEGHE M, AC	1993	3353	3	Jackson, RA	2004	52	3	DeBaun, MR	2003	33
4	MUNNE, S	1995	326	4	MUNNE, S	1995	3099	4	CABROL, C	1981	51	4	Helmerhorst, FM	2004	33
5	Bergh, T	1999	318	5	NICHOLSON, DW	1995	3010	5	CABROL, C	1986	51	5	Jackson, RA	2004	30
6	Gardner, DK	1998	316	6	LAZEBNIK, YA	1994	2898	6	VANSTEIRTEGH EM, AC	1993	51	6	Mastenbroek, S	2007	29
7	DeBaun, MR	2003	301	7	STEPTOE, PC	1978	2888	7	Schieve, LA	2002	50	7	Schieve, LA	2002	28
8	EDWARDS, RG	1965	300	8	TONER, JP	1991	2829	8	SARSAM, MAI	1993	47	8	Thurin, A	2004	27
9	STEER, CV	1992	291	9	Gardner, DK	1998	2706	9	KOUCHOUKOS, NT	1991	46	9	VANSTEIRTEGH EM, AC	1993	27
10	Schieve, LA	2002	288	10	STEER, CV	1992	2683	10	NICHOLSON, DW	1995	43	10	Bergh, T	1999	26
307	EDWARDS, RG	1965	83	78	EDWARDS, RG	1969	1287	29	STEPTOE, PC	1978	32	84	STEPTOE, PC	1978	12
274	EDWARDS, RG	1969	89	129	EDWARDS, RG	1965	1052	605	EDWARDS, RG	1970	8	951	EDWARDS, RG	1969	2
718	EDWARDS, RG	1970	36	245	EDWARDS, RG	1970	763	661	EDWARDS, RG	1965	8	1026	EDWARDS, RG	1965	2
								685	EDWARDS, RG	1969	8	1533	EDWARDS, RG	1970	1

iPS cell [2012med] (1,064 papres)

In-Degree Centrality

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	MESSNER, HA	1980	0.2500	1	MARTIN, GR	1981	2.0209	1	MESSNER, HA	1980	0.2500	1	MARTIN, GR	1981	33.5000
2	Takahashi, K	2006	0.2117	2	MESSNER, HA	1980	1.7463	2	WOLF, NS	1968	0.1282	2	Meyer, N	2008	4.5000

3	Takahashi, K	2007	0.1803	3	Thomson, JA	1998	1.5737	3	GOLDSCHNEIDE R, I	1980	0.1250	3	Hwang, WS	2005	4.1667
4	Thomson, JA	1998	0.1507	4	GOLDSCHNEIDE R, I	1980	1.2705	4	MARTIN, GR	1981	0.1083	4	Youngson, NA	2008	3.5000
5	WOLF, NS	1968	0.1429	5	WOLF, NS	1968	1.0604	5	JONES, RJ	1990	0.0659	5	Chang, TC	2009	3.5000
6	Yu, JY	2007	0.1416	6	JONES, RJ	1990	1.0239	6	WOBUS, AM	1984	0.0583	6	ANDREWS, PW	1984	3.4667
7	GOLDSCHNEIDE R, I	1980	0.1385	7	ABRAMSON, S	1977	0.7937	7	Takahashi, K	2007	0.0503	7	MALTSEV, VA	1994	3.1875
8	JONES, RJ	1990	0.1228	8	Takahashi, K	2006	0.7760	8	Takahashi, K	2006	0.0472	8	WOBUS, AM	1984	3.0500
9	MARTIN, GR	1981	0.1140	9	MATSUI, Y	1992	0.7204	9	PLATZER, E	1985	0.0435	9	Wada, H	2008	3.0000
10	WELTE, K	1985	0.0926	10	WELTE, K	1985	0.7043	10	Yu, JY	2007	0.0419	10	Gang, EJ	2007	3.0000
268	GURDON, JB	1962	0.0098	102	GURDON, JB	1962	0.1295	170	GURDON, JB	1962	0.0066	144	Takahashi, K	2006	0.0382
												774	GURDON, JB	1962	0.0002

PageRank

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	GURDON, JB	1962	0.3412	1	MESSNER, HA	1980	1.5432	1	GURDON, JB	1962	0.3412	1	Meyer, N	2008	4.5000
2	MESSNER, HA	1980	0.2674	2	MARTIN, GR	1981	0.8573	2	MESSNER, HA	1980	0.2517	2	Youngson, NA	2008	3.5000
3	GURDON, JB	1975	0.1844	3	GURDON, JB	1962	0.7488	3	GURDON, JB	1975	0.1844	3	Chang, TC	2009	3.5000
4	MILLER, RA	1976	0.1716	4	MILLER, RA	1976	0.7296	4	MILLER, RA	1976	0.1716	4	ANDREWS, PW	1984	3.4667
5	WOLF, NS	1968	0.0614	5	GOLDSCHNEIDE R, I	1980	0.5103	5	WOLF, NS	1968	0.0614	5	MALTSEV, VA	1994	3.1875
6	GOLDSCHNEIDE	1980	0.0465	6	WOLF, NS	1968	0.4906	6	GOLDSCHNEIDE	1980	0.0465	6	WOBUS, AM	1984	3.0500

	R, I								R, I						
7	MARTIN, GR	1981	0.0436	7	JONES, RJ	1990	0.4532	7	GOLDSCHNEIDE R, I	1978	0.0385	7	Wada, H	2008	3.0000
8	GOLDSCHNEIDE R, I	1978	0.0385	8	GURDON, JB	1975	0.4147	8	CUDKOWICZ, G	1964	0.0385	8	Hong, YK	2004	2.6667
9	CUDKOWICZ, G	1964	0.0385	9	ABRAMSON, S	1977	0.3412	9	MARTIN, GR	1981	0.0374	9	Petrova, TV	2002	2.6667
10	WELTE, K	1985	0.0347	10	WELTE, K	1985	0.3316	10	ABRAMSON, S	1977	0.0266	10	STICE, SL	1988	2.6250
22	Takahashi, K	2006	0.0118	53	Takahashi, K	2006	0.0421	49	Takahashi, K	2006	0.0026	137	Takahashi, K	2006	0.0021
												313	GURDON, JB	1962	0.0002

Number of Citation

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	Takahashi, K	2006	2649	1	Thomson, JA	1998	7976	1	Takahashi, K	2006	711	1	Takahashi, K	2007	407
2	Takahashi, K	2007	2256	2	Takahashi, K	2006	6953	2	Takahashi, K	2007	627	2	Takahashi, K	2006	386
3	Thomson, JA	1998	1825	3	Takahashi, K	2007	5704	3	Yu, JY	2007	464	3	Yu, JY	2007	327
4	Yu, JY	2007	1772	4	MARTIN, GR	1981	4779	4	Thomson, JA	1998	351	4	Okita, K	2007	207
5	Okita, K	2007	1083	5	Yu, JY	2007	4612	5	Okita, K	2007	245	5	Park, IH	2008	187
6	Wernig, M	2007	807	6	Okita, K	2007	3163	6	Park, IH	2008	210	6	Wernig, M	2007	165
7	Park, IH	2008	802	7	Nichols, J	1998	3076	7	Wernig, M	2007	194	7	Kim, K	2010	163
8	MARTIN, GR	1981	799	8	Niwa, H	2000	3042	8	Nakagawa, M	2008	172	8	Nakagawa, M	2008	156
9	Nakagawa, M	2008	694	9	Reubinoff, BE	2000	2891	9	Kim, K	2010	163	9	Yu, JY	2009	146
10	Niwa, H	2000	624	10	Wernig, M	2007	2618	10	Okita, K	2008	155	10	Okita, K	2008	138
224	GURDON, JB	1962	95	128	GURDON, JB	1962	490	276	GURDON, JB	1962	20	458	GURDON, JB	1962	5

cryptology (419 papers)

In-Degree Centrality

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	RIVEST, RL	1978	0.4136	1	RIVEST, RL	1978	8.7548	1	FEISTEL, H	1973	0.1818	1	AnnovazziLodi, V	1996	3.2000
2	DIFFIE, W	1976	0.3780	2	DIFFIE, W	1976	8.6631	2	DIFFIE, W	1976	0.1333	2	NANDI, S	1994	2.6429
3	FEISTEL, H	1973	0.2667	3	FEISTEL, H	1973	2.8215	3	RIVEST, RL	1978	0.1035	3	Sackett, CA	2000	2.0000
4	DIFFIE, W	1979	0.1471	4	ELGAMAL, T	1985	1.8596	4	DIFFIE, W	1979	0.0684	4	LAMPSON, B	1992	1.7333
5	ELGAMAL, T	1985	0.1028	5	DIFFIE, W	1979	1.7732	5	NEEDHAM, RM	1978	0.0297	5	Ohbuchi, R	1998	1.2727
6	POHLIG, SC	1978	0.0949	6	POHLIG, SC	1978	1.5573	6	POHLIG, SC	1978	0.0278	6	Bennett, CH	1998	1.2222
7	EKERT, AK	1991	0.0807	7	NEEDHAM, RM	1978	1.4348	7	GOLDWASSER, S	1984	0.0229	7	Yang, T	1997	1.1250
8	NEEDHAM, RM	1978	0.0784	8	EKERT, AK	1991	1.2791	8	ELGAMAL, T	1985	0.0198	8	Chen, HF	2000	0.9000
9	Gisin, N	2002	0.0646	9	GOLDWASSER, S	1984	1.0625	9	EKERT, AK	1992	0.0183	9	CHUANG, IL	1995	0.7273
10	GOLDWASSER, S	1984	0.0574	10	BENNETT, CH	1992	0.9076	10	KOBLITZ, N	1987	0.0180	10	Lloyd, S	1999	0.6250
												17	RIVEST, RL	1978	0.0423

PageRank

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	DIFFIE, W	1976	0.1691	1	DIFFIE, W	1976	4.4363	1	DIFFIE, W	1976	0.0746	1	AnnovazziLodi, V	1996	3.2000
2	FEISTEL, H	1973	0.1254	2	RIVEST, RL	1978	2.1544	2	FEISTEL, H	1973	0.0712	2	LAMPSON, B	1992	1.7333
3	RIVEST, RL	1978	0.0843	3	FEISTEL, H	1973	1.4329	3	RIVEST, RL	1978	0.0462	3	Ohbuchi, R	1998	1.2727
4	EKERT, AK	1991	0.0282	4	POHLIG, SC	1978	0.5337	4	NEEDHAM, RM	1978	0.0203	4	Bennett, CH	1998	1.2222
5	DIFFIE, W	1979	0.0257	5	EKERT, AK	1991	0.4634	5	POHLIG, SC	1978	0.0203	5	Yang, T	1997	1.1250
6	NEEDHAM, RM	1978	0.0203	6	ELGAMAL, T	1985	0.3784	6	DIFFIE, W	1979	0.0155	6	Chen, HF	2000	0.9000

7	POHLIG, SC	1978	0.0203	7	DIFFIE, W	1979	0.3500	7	EKERT, AK	1991	0.0062	7	Lloyd, S	1999	0.6250
8	ELGAMAL, T	1985	0.0186	8	NEEDHAM, RM	1978	0.2727	8	QUISQUATER, JJ	1982	0.0053	8	WEI, VK	1991	0.5000
9	BENNETT, CH	1992	0.0118	9	GOLDWASSER, S	1984	0.2323	9	GOLDWASSER, S	1984	0.0044	9	ABADI, M	1993	0.4375
10	BENNETT, CH	1992	0.0112	10	BENNETT, CH	1992	0.1966	10	AKL, SG	1983	0.0043	10	Tancevski, L	1996	0.2000
												16	RIVEST, RL	1978	0.0047

Number of Citation

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	DIFFIE, W	1976	1288	1	RIVEST, RL	1978	14145	1	Gisin, N	2002	148	1	Gisin, N	2002	107
2	RIVEST, RL	1978	1267	2	DIFFIE, W	1976	13424	2	EKERT, AK	1991	128	2	EKERT, AK	1991	62
3	EKERT, AK	1991	1222	3	EKERT, AK	1991	8002	3	DIFFIE, W	1976	107	3	RIVEST, RL	1978	37
4	Gisin, N	2002	987	4	ELGAMAL, T	1985	7404	4	RIVEST, RL	1978	86	4	DIFFIE, W	1976	37
5	ELGAMAL, T	1985	774	5	BENNETT, CH	1992	5018	5	BENNETT, CH	1992	72	5	Deng, FG	2003	37
6	BENNETT, CH	1992	684	6	Gisin, N	2002	4548	6	BENNETT, CH	1992	71	6	Deng, FG	2003	37
7	BENNETT, CH	1992	580	7	GOLDWASSER, S	1984	3973	7	ELGAMAL, T	1985	69	7	BENNETT, CH	1992	35
8	KOBLITZ, N	1987	512	8	BENNETT, CH	1992	3853	8	KOBLITZ, N	1987	58	8	ELGAMAL, T	1985	32
9	GOLDWASSER, S	1984	435	9	KOBLITZ, N	1987	3804	9	REFREGIER, P	1995	54	9	Hwang, WY	2003	31
10	REFREGIER, P	1995	430	10	MILLER, VS	1986	2538	10	Deng, FG	2003	53	10	REFREGIER, P	1995	30

quasicrystal [2011chem] (155 papers)

In-Degree Centrality

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	SHECHTMAN, D	1984	1.0000	1	SHECHTMAN, D	1984	11.6782	1	SHECHTMAN, D	1984	1.0000	1	SHECHTMAN, D	1984	63.1923
2	LEVINE, D	1984	0.9583	2	LEVINE, D	1984	5.7812	2	LEVINE, D	1984	0.9583	2	ELSER, V	1985	16.1667

3	LEVINE, D	1985	0.3333	3	ELSER, V	1985	2.7526	3	LEVINE, D	1985	0.3333	3	LEVINE, D	1984	15.8462
4	MERMIN, ND	1985	0.2500	4	ELSER, V	1985	2.7052	4	MERMIN, ND	1985	0.2500	4	BENDERSKY, L	1985	13.4167
5	ELSER, V	1985	0.2397	5	BENDERSKY, L	1985	2.4477	5	ELSER, V	1985	0.1608	5	ELSER, V	1985	13.2692
6	ELSER, V	1985	0.1849	6	LEVINE, D	1985	2.2969	6	ELSER, V	1985	0.1323	6	LEVINE, D	1985	7.5000
7	BENDERSKY, L	1985	0.1712	7	ELSER, V	1985	1.2651	7	ELSER, V	1985	0.1250	7	BANCEL, PA	1989	4.8571
8	ELSER, V	1985	0.1250	8	HE, LX	1988	1.0126	8	BENDERSKY, L	1985	0.0923	8	ELSER, V	1985	4.6923
9	LUBENSKY, TC	1985	0.0704	9	MERMIN, ND	1985	0.9621	9	BANCEL, PA	1989	0.0493	9	LILIENFELD, DA	1985	2.4286
10	HE, LX	1988	0.0669	10	TSAI, AP	1989	0.9375	10	LUBENSKY, TC	1985	0.0422	10	URBAN, K	1985	1.7143

PageRank

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	SHECHTMAN, D	1984	0.3383	1	SHECHTMAN, D	1984	4.8086	1	SHECHTMAN, D	1984	0.3306	1	SHECHTMAN, D	1984	63.1923
2	LEVINE, D	1984	0.1787	2	LEVINE, D	1984	2.0263	2	LEVINE, D	1984	0.1787	2	LEVINE, D	1984	15.8462
3	LEVINE, D	1985	0.0497	3	ELSER, V	1985	0.6010	3	LEVINE, D	1985	0.0497	3	ELSER, V	1985	13.2692
4	MERMIN, ND	1985	0.0381	4	BENDERSKY, L	1985	0.5562	4	MERMIN, ND	1985	0.0381	4	LEVINE, D	1985	7.5000
5	ELSER, V	1985	0.0368	5	LEVINE, D	1985	0.5010	5	ELSER, V	1985	0.0311	5	ELSER, V	1985	4.6923
6	BENDERSKY, L	1985	0.0323	6	ELSER, V	1985	0.4783	6	ELSER, V	1985	0.0230	6	LILIENFELD, DA	1985	2.4286
7	ELSER, V	1985	0.0311	7	ELSER, V	1985	0.4248	7	HENLEY, CL	1985	0.0193	7	URBAN, K	1985	1.7143
8	ELSER, V	1985	0.0241	8	MERMIN, ND	1985	0.2573	8	BENDERSKY, L	1985	0.0172	8	TOKURA, Y	1993	1.6667
9	HENLEY, CL	1985	0.0193	9	HENLEY, CL	1985	0.2315	9	LILIENFELD, DA	1985	0.0172	9	MERMIN, ND	1985	0.8846
10	LILIENFELD, DA	1985	0.0172	10	LUBENSKY, TC	1985	0.1968	10	GUYOT, P	1985	0.0172	10	Cox, DE	1998	0.5000

Number of Citation

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	SHECHTMAN, D	1984	1745	1	SHECHTMAN, D	1984	22678	1	SHECHTMAN, D	1984	95	1	SHECHTMAN, D	1984	63
2	LEVINE, D	1984	459	2	LEVINE, D	1984	6694	2	KLEIN, T	1991	45	2	ELSER, V	1985	16
3	ELSER, V	1985	393	3	ELSER, V	1985	6413	3	ELSER, V	1985	42	3	LEVINE, D	1984	16
4	ELSER, V	1985	354	4	BENDERSKY, L	1985	5551	4	Inoue, A	1999	38	4	Tsai, AP	2000	15
5	BENDERSKY, L	1985	329	5	ELSER, V	1985	5479	5	BENDERSKY, L	1985	38	5	BOUDARD, M	1992	13
6	BOUDARD, M	1992	252	6	BOUDARD, M	1992	3161	6	LEVINE, D	1984	38	6	BENDERSKY, L	1985	13
7	LEVINE, D	1985	212	7	LEVINE, D	1985	3107	7	Chen, MW	1999	35	7	ELSER, V	1985	13
8	POON, SJ	1992	208	8	KLEIN, T	1991	3080	8	BOUDARD, M	1992	34	8	Koster, U	1996	12
9	TSAI, AP	1989	207	9	TSAI, AP	1989	3069	9	ELSER, V	1985	32	9	POON, SJ	1992	11
10	KLEIN, T	1991	199	10	POON, SJ	1992	3043	10	Eckert, J	1998	32	10	Chen, MW	1999	10

G-protein-coupled receptors [2012chem] (3,557 papers)

In-Degree Centrality

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	POWELL, CE	1958	1.0000	1	POWELL, CE	1958	6.9451	1	POWELL, CE	1958	1.0000	1	TOLKOVSKY, AM	1978	3.5000
2	FURCHGOTT, RF	1959	0.4286	2	FURCHGOTT, RF	1959	2.1897	2	FURCHGOTT, RF	1959	0.4000	2	EVA, C	1990	1.7500
3	KAISER, GA	1964	0.1538	3	FURCHGOT.RF	1967	1.6317	3	KAISER, GA	1964	0.1538	3	LIMBIRD, LE	1975	1.6857
4	FURCHGOT.RF	1967	0.1235	4	MAGUIRE, ME	1976	1.0540	4	LISH, PM	1965	0.0606	4	WOODCOCK, EA	1979	1.5294
5	LISH, PM	1965	0.1042	5	ROBISON, GA	1967	1.0208	5	FURCHGOT.RF	1967	0.0527	5	RICE, WR	1995	1.3636
6	ROBISON, GA	1967	0.0909	6	MUKHERJEE, C	1975	0.9909	6	AURBACH, GD	1974	0.0525	6	VOIGT, MM	1991	1.2667

7	MAGUIRE, ME	1976	0.0802	7	AURBACH, GD	1974	0.9354	7	LEFKOWIT.RJ	1974	0.0460	7	POWELL, CE	1958	1.2041
8	AURBACH, GD	1974	0.0797	8	MUKHERJEE, C	1975	0.9317	8	LEVITZKI, A	1974	0.0418	8	EPSTEIN, SE	1966	1.1667
9	LEFKOWIT.RJ	1974	0.0750	9	KAISER, GA	1964	0.9294	9	BELLEAU, B	1967	0.0417	9	SHEWRING, G	1982	1.1579
10	MUKHERJEE, C	1975	0.0733	10	LEFKOWIT.RJ	1974	0.9167	10	MUKHERJEE, C	1975	0.0406	10	WILLIAMS, JA	1993	1.0000
29	DIXON, RAF	1986	0.0431	13	DELEAN, A	1980	0.8032	35	LEFKOWIT.RJ	1970b	0.0162	118	DIXON, RAF	1986	0.0093
36	DELEAN, A	1980	0.0387	23	DIXON, RAF	1986	0.6297	45	DIXON, RAF	1986	0.0142	133	LEFKOWIT.RJ	1970b	0.0075
47	LEFKOWIT.RJ	1970b	0.0323	73	LEFKOWIT.RJ	1970b	0.3295	72	DELEAN, A	1980	0.0098	166	DELEAN, A	1980	0.0047
81	LEFKOWIT.RJ	1970a	0.0231	98	LEFKOWIT.RJ	1970a	0.2788	78	LEFKOWIT.RJ	1970a	0.0096	181	LEFKOWIT.RJ	1970a	0.0041
262	KOBILKA, BK	1987	0.0098	204	KOBILKA, BK	1987	0.1704	256	KOBILKA, BK	1987	0.0035	295	KOBILKA, BK	1987	0.0020

PageRank

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	POWELL, CE	1958	0.5420	1	POWELL, CE	1958	4.0142	1	POWELL, CE	1958	0.5420	1	WOODCOCK, EA	1979	1.5294
2	FURCHGOTT, RF	1959	0.1864	2	FURCHGOTT, RF	1959	1.6358	2	FURCHGOTT, RF	1959	0.1527	2	RICE, WR	1995	1.3636
3	KAISER, GA	1964	0.0706	3	ROBISON, GA	1967	0.7520	3	KAISER, GA	1964	0.0706	3	POWELL, CE	1958	1.2041
4	STANTON, HC	1965	0.0492	4	KAISER, GA	1964	0.5642	4	STANTON, HC	1965	0.0492	4	EPSTEIN, SE	1966	1.1667
5	LISH, PM	1965	0.0396	5	FURCHGOT.RF	1967	0.5576	5	LISH, PM	1965	0.0396	5	SHEWRING, G	1982	1.1579
6	TURNER, P	1965	0.0286	6	AURBACH, GD	1974	0.4566	6	TURNER, P	1965	0.0286	6	STRITTMATTER , WJ	1979	1.0000
7	FURCHGOT.RF	1967	0.0269	7	LEFKOWIT.RJ	1970a	0.4396	7	FOSTER, RW	1966	0.0163	7	DEGUCHI, T	1972	0.9355
8	ROBISON, GA	1967	0.0247	8	STANTON, HC	1965	0.4351	8	EPSTEIN, SE	1966	0.0155	8	HARDEN, TK	1982	0.7200
9	LEFKOWIT.RJ	1970a	0.0219	9	LISH, PM	1965	0.3979	9	AHLQUIST, RP	1966	0.0155	9	Fathi, Z	1997	0.6667
10	LEFKOWIT.RJ	1970b	0.0193	10	LEFKOWIT.RJ	1974	0.3886	10	BELLEAU, B	1967	0.0147	10	GOVIER, WC	1966	0.6129

28	DIXON, RAF	1986	0.0071	13	LEFKOWIT.RJ	1970b	0.3192	18	LEFKOWIT.RJ	1970a	0.0055	87	LEFKOWIT.RJ	1970b	0.0029
91	DELEAN, A	1980	0.0025	23	DIXON, RAF	1986	0.1451	19	LEFKOWIT.RJ	1970b	0.0054	88	LEFKOWIT.RJ	1970a	0.0028
148	KOBILKA, BK	1987	0.0014	61	DELEAN, A	1980	0.0645	83	DIXON, RAF	1986	0.0011	152	DIXON, RAF	1986	0.0004
				138	KOBILKA, BK	1987	0.0289	228	DELEAN, A	1980	0.0003	262	DELEAN, A	1980	0.0001
								321	KOBILKA, BK	1987	0.0002	294	KOBILKA, BK	1987	0.0001

Number of Citation

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	Palczewski, K	2000	1792	1	Palczewski, K	2000	11822	1	Cherezov, V	2007	189	1	Cherezov, V	2007	158
2	Cherezov, V	2007	782	2	HAUSDORFF, WP	1990	11373	2	Palczewski, K	2000	171	2	Palczewski, K	2000	152
3	HAUSDORFF, WP	1990	713	3	DELEAN, A	1980	10907	3	Jaakola, VP	2008	148	3	Jaakola, VP	2008	133
4	EMORINE, LJ	1989	690	4	DOHLMAN, HG	1991	10225	4	Warne, T	2008	126	4	Warne, T	2008	113
5	SAMAMA, P	1993	670	5	EMORINE, LJ	1989	10078	5	Rasmussen, SGF	2011	119	5	Rasmussen, SGF	2007	107
6	DELEAN, A	1980	660	6	DIXON, RAF	1986	10042	6	Rasmussen, SGF	2007	116	6	Scheerer, P	2008	83
7	DOHLMAN, HG	1991	643	7	KOBILKA, BK	1987	9599	7	Rasmussen, SGF	2011	104	7	Rosenbaum, DM	2009	78
8	Ferguson, SSG	2001	626	8	BRISTOW, MR	1982	9302	8	Wu, BL	2010	99	8	Rosenbaum, DM	2007	77
9	Pierce, KL	2002	614	9	COTECCHIA, S	1988	8405	9	Rosenbaum, DM	2007	98	9	Park, JH	2008	73
10	BRISTOW, MR	1982	609	10	SAMAMA, P	1993	8066	10	Rosenbaum, DM	2009	93	10	Pierce, KL	2002	62
12	DIXON, RAF	1986	575	132	KOBILKA, BK	1987	3101	25	DIXON, RAF	1986	69	89	LEFKOWIT.RJ	1970a	1
233	KOBILKA, BK	1987	179	779	LEFKOWIT.RJ	1970b	1024	70	DELEAN, A	1980	45	107	DIXON, RAF	1986	22
1795	LEFKOWIT.RJ	1970a	43	833	LEFKOWIT.RJ	1970a	968	604	KOBILKA, BK	1987	18	738	KOBILKA, BK	1987	7
1916	LEFKOWIT.RJ	1970b	39					2500	LEFKOWIT.RJ	1970b	5	2878	LEFKOWIT.RJ	1970b	1
								2755	LEFKOWIT.RJ	1970a	4	2913	DELEAN, A	1980	21

palladium catalyst [2010chem] (1728 papers)

In-Degree Centrality

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	TAYLOR, EC	1968	0.106	1	STILLE, JK	1986	0.8880	1	TAYLOR, EC	1968	0.1061	1	HECK, RF	1972	6.8182
2	TAYLOR, EC	1971	0.106	2	HECK, RF	1982	0.6537	2	TAYLOR, EC	1971	0.1061	2	TAMAO, K	1972	5.3636
3	MCKILLOP, A	1971	0.106	3	TROST, BM	1977	0.5788	3	MCKILLOP, A	1971	0.1061	3	DIECK, HA	1975	4.5758
4	MCKILLOP, A	1973	0.076	4	CURRAN, DP	1988	0.5780	4	MCKILLOP, A	1973	0.0758	4	SCHOENBEA	1974	3.7576
5	MCKILLOP, A	1973	0.076	5	MIYAUURA, N	1995	0.5700	5	MCKILLOP, A	1973	0.0758	5	MELPOLDER, JB	1976	2.6667
6	MCKILLOP, A	1971	0.076	6	HECK, RF	1979	0.5218	6	MCKILLOP, A	1971	0.0758	6	BALLESTRI, M	1991	2.4375
7	MCKILLOP, A	1973	0.076	7	MELPOLDER, JB	1976	0.4826	7	MCKILLOP, A	1973	0.0758	7	CROWLEY, JI	1976	2.0909
8	LEZNOFF, CC	1973	0.076	8	CURRAN, DP	1988	0.4722	8	LEZNOFF, CC	1973	0.0758	8	HART, DW	1974	2.0909
9	TAYLOR, EC	1970	0.061	9	TAMAO, K	1972	0.4396	9	TAYLOR, EC	1970	0.0606	9	COREY, EJ	1965	2.0882
10	MIYAUURA, N	1995	0.059	10	CHALK, AJ	1976	0.4127	10	TSUJI, J	1964	0.0490	10	EVANS, DA	1974	2.0625
20	HECK, RF	1972	0.029	18	HECK, RF	1972	0.3818	15	HECK, RF	1972	0.0294	115	MIYAUURA, N	1979b	0.0056
42	NEGISHI, E	1977	0.020	13	NEGISHI, E	1977	0.3994	81	NEGISHI, E	1977	0.0058	116	MIYAUURA, N	1979a	0.0052
77	MIYAUURA, N	1979a	0.013	73	MIYAUURA, N	1979a	0.1912	60	MIYAUURA, N	1979	0.0083	118	NEGISHI, E	1977	0.0051
122	MIYAUURA, N	1979b	0.009	130	MIYAUURA, N	1979b	0.1188	82	MIYAUURA, N	1979	0.0056				

PageRank

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	HECK, RF	1972	0.0216	1	HECK, RF	1972	0.53261	1	HECK, RF	1972	0.0211	1	HECK, RF	1972	6.8182
2	TAMAO, K	1972	0.0174	2	TAMAO, K	1972	0.45894	2	COREY, EJ	1965	0.0161	2	SCHOENBEA	1974	3.7576

3	COREY, EJ	1965	0.0161	3	SCHOENBE.A	1974	0.37112	3	TAYLOR, EC	1968	0.0161	3	BALLESTRI, M	1991	2.4375
4	TAYLOR, EC	1968	0.0161	4	SCHOENBE.A	1974	0.18972	4	TSUJI, J	1964	0.0160	4	CROWLEY, JI	1976	2.0909
5	TSUJI, J	1964	0.016	5	COREY, EJ	1965	0.18838	5	TAMAO, K	1972	0.0159	5	COREY, EJ	1965	2.0882
6	MCKILLOP, A	1971	0.015	6	BABA, S	1976	0.1673	6	MCKILLOP, A	1971	0.0150	6	EVANS, DA	1974	2.0625
7	SCHOENBE.A	1974	0.0149	7	MORRELL, DG	1975	0.16418	7	SCHOENBE.A	1974	0.0145	7	VOLLHARDT, KPC	1977	1.8125
8	MCKILLOP, A	1971	0.0137	8	TAMAO, K	1972	0.16385	8	MCKILLOP, A	1971	0.0137	8	CHALK, AJ	1976	1.7879
9	COREY, EJ	1969	0.0125	9	MELPOLDER, JB	1976	0.1503	9	COREY, EJ	1969	0.0125	9	POUTSMA, ML	1978	1.6500
10	TAYLOR, EC	1971	0.0125	10	EVANS, DA	1974	0.14969	10	TAYLOR, EC	1971	0.0125	10	OIKAWA, Y	1978	1.5556
31	NEGISHI, E	1977	0.0056	11	NEGISHI, E	1977	0.14943	73	NEGISHI, E	1977	0.0014	123	NEGISHI, E	1977	0.0004
100	MIYAURA, N	1979a	0.0016	61	MIYAURA, N	1979a	0.04523	132	MIYAURA, N	1979a	0.0007	148	MIYAURA, N	1979a	0.0002
151	MIYAURA, N	1979b	0.0011	96	MIYAURA, N	1979b	0.03112	137	MIYAURA, N	1979b	0.0007	700	MIYAURA, N	1979b	0.0000

Number of Citation

Rank	1st Author	PY	Height	Rank	1st Author	PY	Area	Rank	1st Author	PY	Slope	Rank	1st Author	PY	SlopeAve
1	MIYAURA, N	1995	1786	1	STILLE, JK	1986	12853	1	MIYAURA, N	1995	360	1	MIYAURA, N	1995	116
2	STILLE, JK	1986	1009	2	MIYAURA, N	1995	11294	2	Nicolaou, KC	2005	141	2	Littke, AF	2002	62
3	HECK, RF	1982	497	3	CURRAN, DP	1988	7635	3	Hassan, J	2002	132	3	Chinchilla, R	2007	59
4	CURRAN, DP	1988	496	4	HECK, RF	1982	7211	4	Littke, AF	2002	124	4	Corbet, JP	2006	52
5	Littke, AF	2002	457	5	CURRAN, DP	1988	6222	5	Chinchilla, R	2007	113	5	Hassan, J	2002	51
6	Tietze, LF	1996	443	6	MOLANDER, GA	1992	5713	6	Tietze, LF	1996	112	6	Nicolaou, KC	2005	48
7	MOLANDER, GA	1992	441	7	GIRARD, P	1980	5090	7	STILLE, JK	1986	108	7	Kantchev, EAB	2007	42
8	Hassan, J	2002	416	8	MIYAURA, N	1981	4944	8	Corbet, JP	2006	100	8	STILLE, JK	1986	42

9	CURRAN, DP	1988	411	9	RAMAIAH, M	1987	4691	9	Ley, SV	2003	94	9	Kotha, S	2002	40
10	MIYAURA, N	1981	403	10	HECK, RF	1979	4122	10	Kotha, S	2002	87	10	McGlacken, GP	2009	38
28	HECK, RF	1972	234	22	NEGISHI, E	1977	2967	21	HECK, RF	1972	60	250	HECK, RF	1972	7
36	NEGISHI, E	1977	214	38	HECK, RF	1972	2240	36	MIYAURA, N	1979a	46	261	NEGISHI, E	1977	7
67	MIYAURA, N	1979a	156	100	MIYAURA, N	1979a	1355	163	NEGISHI, E	1977	22	375	MIYAURA, N	1979a	5
328	MIYAURA, N	1979b	62	288	MIYAURA, N	1979b	732	401	MIYAURA, N	1979b	13	940	MIYAURA, N	1979b	2

5.3. 特徴量の妥当性の検討

前節の計算結果は、全般に、in-Degree Centrality および PageRank の特徴量 Height、Area、Slope は、被引用数の特徴量 Height、Area、Slope より、萌芽的論文を特定するのに有効か同程度である。また、in-Degree Centrality と PageRank において特徴量 Height、Area、Slope を比較すると、PageRank の特徴量の方が有効という傾向にある。SlopeAve は、Height、Area、Slope と比較すると萌芽的論文の特定において劣る傾向にあるので、以降では Height、Area、Slope に関して述べる。

データセットを、in-Degree Centrality および PageRank の特徴量 Height、Area、Slope が萌芽的論文を特定できた度合で分類すると、Table 32 のように、大きく 3 つの類型に分けられる。類型 I は、被引用数の特徴量 Height、Area、Slope を使用するより、in-Degree Centrality および PageRank の特徴量 Height、Area、Slope を使用する方が萌芽的論文の特定に優れている場合、類型 II は、in-Degree Centrality および PageRank の特徴量 Height、Area、Slope が被引用数の特徴量 Height、Area、Slope と明確に優劣が判断できない場合、類型 III は、in-Degree Centrality および PageRank の特徴量 Height、Area、Slope が部分的にしか当てはまらない場合である。これらは、受賞論文が上位 10 位以内、あるいは、準ずるレベルの上位にある場合(1000 論文あるうちの 15 位など)を、萌芽的論文の適切に特定したとみている。受賞論文以外にも研究対象としたデータセットの領域において、萌芽的論文が存在することも事実である。そこで、in-Degree Centrality 及び Pagerank の特徴量 Height、Area、Slope の上位 10 位以内に抽出された論文群について、当該分野の学術専門家に、萌芽的論文に該当する論文であるかどうか評価をしてもらった。その結果、コンタクトのとれた領域の専門家より、上位 10 位内の論文群の多くは、萌芽的論文またはそれに準じる位置付けの論文群であるとの評価を受けた。本研究では、ノーベル賞等の受賞論文を萌芽的論文の代表としてそれに関する検証を行ってきたが、先の評価結果から、ノーベル賞等に厳密に限定されない萌芽的論文の抽出にも提案手法は有効であると考えられる。

Table 32: 各データセットにおける、特徴量の適合度

類型 I 特徴量が被引用数よりも優れたデータセット	類型 II 特徴量が被引用数の場合と優劣が判断できないデータセット	類型 III 特徴量が部分的にしか当てはまらないデータセット
<ul style="list-style-type: none"> • 蛍光プロテイン(gfp) ■ • 体外受精技術(2010med) ■ • iPS 細胞(2012med) ■ ※PageRank のみ	<ul style="list-style-type: none"> • 暗号(crypto) ■ • 準結晶(2011chem) ■ 	<ul style="list-style-type: none"> • パラジウム触媒(2010chem) ■ • G タンパク質共益受容体(2012chem) ■

類型 I について、1 つのデータセット内に複数の受賞論文がある場合につき、in-Degree Centrality と PageRank の 2 指標における特徴量 Height、Area、Slope と、被引用数の特徴量 Height、Area、Slope を比較すると、in-Degree Centrality、PageRank では、先に発表された論文が後年に発表された論文より上位になる傾向が強い。他方、被引用数(Number of Citation) の特徴量 Height、Area、Slope は、後年に発表された論文の方が上位となる。このことから、萌芽的であることに注目するならば、in-Degree Centrality、および、PageRank を使用すべきと言えよう。

類型 II について、データセット内に受賞論文が一編しかなく、この一編が有力であることから、In-Degree Centrality と PageRank と共に、データセット内の被引用数によっても重要論文として特定可能である。

類型 III について、in-Degree Centrality と PageRank の特徴量 Height、Area、Slope が萌芽的論文の特定に、部分的に機能しているように見える。

パラジウム触媒(2010chem)の上位にノーベル賞受賞論文以外が多く選出されているが、この分野は、溝呂木・Heck 反応、熊田・玉尾・Corriu 反応、菌頭反応、根岸反応、Stille 反応、鈴木・宮浦反応と多様な反応名あるように、ノーベル賞は多くの研究者からわずか 3 名を選らるのであり、Table 31 には Stille 氏、Tamao 氏などの受賞はしていない萌芽的論文も含まれる。また、しばしば登場している Taylor 氏、Mckillop 氏は、タリウムを使った有機合成の研究者であり、関連領域の研究も抽出されている。

G タンパク質共益受容体(2012chem)も、受賞論文が TOP 10 に入っていない異質なデータセットであるが、これは、G タンパク質共益受容体の関連分野もデータセット内に多く含まれてしまっているためである。In-Degree Centrality、および、PageRank の上位にある、Furchgott 氏(あるいは、Furchgot 氏。表記ゆれがある)は、1998 年のノーベル生理学・医学賞の受賞者の一人であり、循環器系における情報伝達物質としての一酸化窒素の発見について受賞されている。なお、in-Degree Centrality、および、PageRank のすべてに一位の Powell 氏は、アドレナリン受容体に対する受容体遮断薬の研究者である。これらの分野は、データセット作成段階のキーワードを再検討しても分けがたく、クラスター分析においても同じサブクラスターとなり分離が難しい。しかし、萌芽的論文を特定するという点においては、特徴量は目的を果たしていると十分考え得る。

次に、in-Degree Centrality と PageRank でパフォーマンスを比較する時、PageRank の方が in-Degree Centrality よりも萌芽的論文を特定するのに、有利な傾向がある。受賞論文が 10 位以内に入らない G タンパク質共益受容体(2012chem)においても、関連分野を考慮しないで 2012 年ノーベル化学賞受賞論文どうしで比較すると、PageRank の方が in-Degree Centrality より先に発表された論文が上位にある傾向は他のデータセットと同様に見られる。

PageRank の方が in-Degree Centrality より、先に発表された論文が上位になる理由について、PageRank と in-Degree Centrality の原理より説明すると、PageRank では、多くの論文の引用を集めた論文が引用する論文に大きな評価を与えられる。一方、in-Degree Centrality は、引用する論文が受けた評価が高くても、引用される論文に重み付けされない。つまり、PageRank では、in-Degree Centrality より Figure 100 の赤矢印の分が重み付けとなって評価され、先に出版された論文の方が PageRank による重み付けの効果がより強調されるため、PageRank が萌芽的論文を特定するのに適していると理論的に考えられる。ゆえに、PageRank の重み付けが十分に機能していると、iPS 細胞(2012med)のように、in-Degree Centrality の Height、Area、Slope で出版年の新しい論文が上位になるようなことがあっても、PageRank では、出版年の古い論文が上位になり萌芽的論文を特定する役割を果たすものと考えられる。

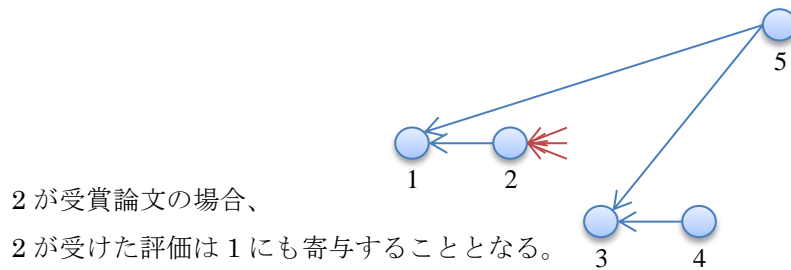


Figure 100 : PageRank が in-Degree Centrality より萌芽的論文を特定するのに有効な理由

Height、Area、Slope の間で比較すると、この 3 特徴量ではほとんど差は見られない。変化率や時間の累積を考慮した Slope、Area の導入にも関わらず、萌芽的論文を特定するには Height で十分という帰結が得られるが、このことは、Figure 101 のような、システム工学のインパルス応答で説明することができる。すなわち、システムへの入力値が大きいと、出力の波形の高さとなって表れる。また、システム内部の回路において抵抗が大きいと、入力減衰し最大値が現れる遅れ時間は増加する。本来、電気的な現象を説明するものであるが、Koop(1996)らにより米国 GDP と失業率といった経済分野に適用されたこともある。萌芽的論文を特定するというシステムにおいては、本研究では出力は存在するが入力値は不明であり、システムの入力値を求められれば萌芽的論文のポテンシャルが判明すると考えられる。ゆえに、Height で萌芽的論文を特定する役割を果たすと考えられる。

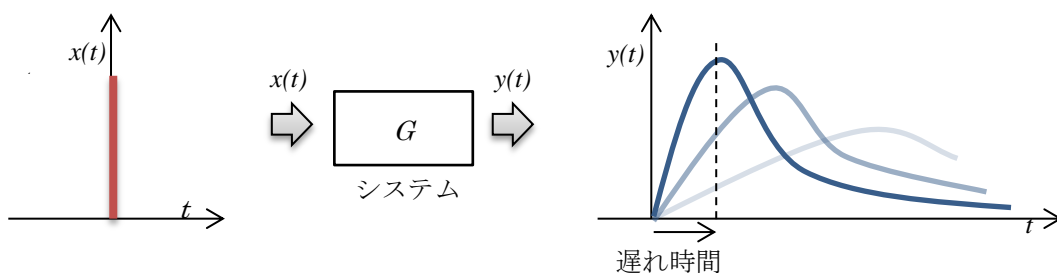


Figure 101 : インパルス応答の模式図

ここまで複数の特徴量を用いて分析を行ってきた。相関の低い複数の特徴量で萌芽的論文を特定出来ていれば、複数の特徴量を組み合わせることにより、より高い確度で萌芽的論文を検出することが可能であると考えられる。そこで、類型Ⅰ、または、類型Ⅱのデータセットについて、特徴量間の相関係数を計算してみたところ、Table 33・Table 37 のようになった。緑は、相関係数 0.7 以上または -0.7 以下を示し特徴量間の相関が高く、赤は、相関係数 -0.2~0.2 を表し特徴量間の相関がないことを示している。この分析結果から、in-Degree Centrality と PageRank の特徴量 Height、Area、Slope の 6 つの中では、iPS 細胞(2012med)のデータセットを除き、いずれも相関がある。一方で、これら 6 つの特徴量の間相関は高いことから、いずれの特徴量を利用しても大差なく萌芽的論文を特定できるものと評価される。

Table 33 : 蛍光プロテイン(gfp)の特徴量間の相関係数

	in-Degree: Height	in-Degree: Area	in-Degree: Slope	in-Degree: SlopeAve	PageRank: Height	PageRank: Area	PageRank: Slope	PageRank: SlopeAve	Citation: Height	Citation: Area	Citation: Slope	Citation: SlopeAve
in-Degree: Height	1.000	0.949	0.935	0.497	0.899	0.852	0.820	0.519	0.468	0.551	0.432	0.323
in-Degree: Area	0.949	1.000	0.968	0.267	0.971	0.957	0.940	0.291	0.316	0.373	0.288	0.195
in-Degree: Slope	0.935	0.968	1.000	0.296	0.943	0.924	0.931	0.291	0.301	0.365	0.280	0.183
in-Degree: SlopeAve	0.497	0.267	0.296	1.000	0.242	0.166	0.096	0.884	0.641	0.748	0.598	0.487
PageRank: Height	0.899	0.971	0.943	0.242	1.000	0.990	0.975	0.278	0.265	0.322	0.243	0.144
PageRank: Area	0.852	0.957	0.924	0.166	0.990	1.000	0.982	0.197	0.212	0.260	0.196	0.102
PageRank: Slope	0.820	0.940	0.931	0.096	0.975	0.982	1.000	0.107	0.151	0.190	0.141	0.059
PageRank: SlopeAve	0.519	0.291	0.291	0.884	0.278	0.197	0.107	1.000	0.651	0.768	0.598	0.476
Citation: Height	0.468	0.316	0.301	0.641	0.265	0.212	0.151	0.651	1.000	0.967	0.953	0.928
Citation: Area	0.551	0.373	0.365	0.748	0.322	0.260	0.190	0.768	0.967	1.000	0.892	0.820
Citation: Slope	0.432	0.288	0.280	0.598	0.243	0.196	0.141	0.598	0.953	0.892	1.000	0.965
Citation: SlopeAve	0.323	0.195	0.183	0.487	0.144	0.102	0.059	0.476	0.928	0.820	0.965	1.000

Table 34 : 体外受精技術(2010med)の特徴量間の相関係数

	in-Degree: Height	in-Degree: Area	in-Degree: Slope	in-Degree: SlopeAve	PageRank: Height	PageRank: Area	PageRank: Slope	PageRank: SlopeAve	Citation: Height	Citation: Area	Citation: Slope	Citation: SlopeAve
in-Degree: Height	1.000	0.977	0.967	0.741	0.948	0.976	0.933	0.336	0.142	0.244	0.087	0.050
in-Degree: Area	0.977	1.000	0.902	0.781	0.884	0.945	0.875	0.364	0.112	0.212	0.057	0.022
in-Degree: Slope	0.967	0.902	1.000	0.657	0.974	0.971	0.959	0.204	0.101	0.204	0.060	0.016
in-Degree: SlopeAve	0.741	0.781	0.657	1.000	0.619	0.672	0.612	0.400	0.094	0.177	0.061	0.037
PageRank: Height	0.948	0.884	0.974	0.619	1.000	0.976	0.995	0.172	0.099	0.202	0.049	0.007
PageRank: Area	0.976	0.945	0.971	0.672	0.976	1.000	0.967	0.217	0.087	0.186	0.038	0.001
PageRank: Slope	0.933	0.875	0.959	0.612	0.995	0.967	1.000	0.112	0.089	0.191	0.040	-0.003
PageRank: SlopeAve	0.336	0.364	0.204	0.400	0.172	0.217	0.112	1.000	0.111	0.171	0.092	0.059
Citation: Height	0.142	0.112	0.101	0.094	0.099	0.087	0.089	0.111	1.000	0.923	0.821	0.889
Citation: Area	0.244	0.212	0.204	0.177	0.202	0.186	0.191	0.171	0.923	1.000	0.697	0.690
Citation: Slope	0.087	0.057	0.060	0.061	0.049	0.038	0.040	0.092	0.821	0.697	1.000	0.808
Citation: SlopeAve	0.050	0.022	0.016	0.037	0.007	0.001	-0.003	0.059	0.889	0.690	0.808	1.000

Table 35 : iPS 細胞(2010med)の特徴量間の相関係数

	in-Degree: Height	in-Degree: Area	in-Degree: Slope	in-Degree: SlopeAve	PageRank: Height	PageRank: Area	PageRank: Slope	PageRank: SlopeAve	Citation: Height	Citation: Area	Citation: Slope	Citation: SlopeAve
in-Degree: Height	1.000	0.881	0.849	0.693	0.379	0.611	0.324	0.568	0.732	0.775	0.665	0.601
in-Degree: Area	0.881	1.000	0.832	0.437	0.402	0.726	0.347	0.488	0.490	0.685	0.355	0.267
in-Degree: Slope	0.849	0.832	1.000	0.353	0.533	0.830	0.496	0.333	0.333	0.395	0.302	0.279
in-Degree: SlopeAve	0.693	0.437	0.353	1.000	0.032	0.085	-0.015	0.675	0.793	0.725	0.791	0.833
PageRank: Height	0.379	0.402	0.533	0.032	1.000	0.836	0.997	0.088	0.041	0.094	0.018	-0.004
PageRank: Area	0.611	0.726	0.830	0.085	0.836	1.000	0.811	0.227	0.075	0.198	0.020	-0.020
PageRank: Slope	0.324	0.347	0.496	-0.015	0.997	0.811	1.000	0.024	0.002	0.044	-0.013	-0.029
PageRank: SlopeAve	0.568	0.488	0.333	0.675	0.088	0.227	0.024	1.000	0.430	0.482	0.379	0.353
Citation: Height	0.732	0.490	0.333	0.793	0.041	0.075	0.002	0.430	1.000	0.927	0.957	0.890
Citation: Area	0.775	0.685	0.395	0.725	0.094	0.198	0.044	0.482	0.927	1.000	0.799	0.707
Citation: Slope	0.665	0.355	0.302	0.791	0.018	0.020	-0.013	0.379	0.957	0.799	1.000	0.940
Citation: SlopeAve	0.601	0.267	0.279	0.833	-0.004	-0.020	-0.029	0.353	0.890	0.707	0.940	1.000

Table 36 : 暗号(crypto)の特徴量間の相関係数

	in-Degree: Height	in-Degree: Area	in-Degree: Slope	in-Degree: SlopeAve	PageRank: Height	PageRank: Area	PageRank: Slope	PageRank: SlopeAve	Citation: Height	Citation: Area	Citation: Slope	Citation: SlopeAve
in-Degree: Height	1.000	0.968	0.918	0.918	0.923	0.900	0.917	0.806	0.724	0.856	0.526	0.374
in-Degree: Area	0.968	1.000	0.822	0.827	0.893	0.938	0.868	0.687	0.718	0.878	0.488	0.314
in-Degree: Slope	0.918	0.822	1.000	0.979	0.943	0.824	0.963	0.955	0.521	0.648	0.358	0.226
in-Degree: SlopeAve	0.918	0.827	0.979	1.000	0.924	0.820	0.949	0.910	0.533	0.658	0.380	0.258
PageRank: Height	0.923	0.893	0.943	0.924	1.000	0.953	0.980	0.900	0.588	0.723	0.400	0.237
PageRank: Area	0.900	0.938	0.824	0.820	0.953	1.000	0.906	0.736	0.630	0.780	0.425	0.250
PageRank: Slope	0.917	0.868	0.963	0.949	0.980	0.906	1.000	0.907	0.505	0.654	0.322	0.172
PageRank: SlopeAve	0.806	0.687	0.955	0.910	0.900	0.736	0.907	1.000	0.422	0.518	0.288	0.179
Citation: Height	0.724	0.718	0.521	0.533	0.588	0.630	0.505	0.422	1.000	0.946	0.920	0.806
Citation: Area	0.856	0.878	0.648	0.658	0.723	0.780	0.654	0.518	0.946	1.000	0.768	0.600
Citation: Slope	0.526	0.488	0.358	0.380	0.400	0.425	0.322	0.288	0.920	0.768	1.000	0.935
Citation: SlopeAve	0.374	0.314	0.226	0.258	0.237	0.250	0.172	0.179	0.806	0.600	0.935	1.000

Table 37: 準結晶(2011chem)の特徴量間の相関係数

	in-Degree: Height	in-Degree: Area	in-Degree: Slope	in-Degree: SlopeAve	PageRank: Height	PageRank: Area	PageRank: Slope	PageRank: SlopeAve	Citation: Height	Citation: Area	Citation: Slope	Citation: SlopeAve
in-Degree: Height	1.000	0.943	0.994	0.016	0.946	0.918	0.941	0.114	0.826	0.841	0.638	0.732
in-Degree: Area	0.943	1.000	0.924	0.083	0.965	0.974	0.952	0.197	0.952	0.966	0.746	0.857
in-Degree: Slope	0.994	0.924	1.000	-0.037	0.952	0.919	0.953	0.051	0.796	0.806	0.584	0.695
in-Degree: SlopeAve	0.016	0.083	-0.037	1.000	-0.043	-0.029	-0.055	0.126	0.164	0.200	0.373	0.198
PageRank: Height	0.946	0.965	0.952	-0.043	1.000	0.992	0.998	0.059	0.895	0.890	0.627	0.790
PageRank: Area	0.918	0.974	0.919	-0.029	0.992	1.000	0.987	0.093	0.927	0.922	0.658	0.824
PageRank: Slope	0.941	0.952	0.953	-0.055	0.998	0.987	1.000	0.017	0.878	0.870	0.602	0.772
PageRank: SlopeAve	0.114	0.197	0.051	0.126	0.059	0.093	0.017	1.000	0.195	0.257	0.320	0.197
Citation: Height	0.826	0.952	0.796	0.164	0.895	0.927	0.878	0.195	1.000	0.989	0.840	0.958
Citation: Area	0.841	0.966	0.806	0.200	0.890	0.922	0.870	0.257	0.989	1.000	0.836	0.924
Citation: Slope	0.638	0.746	0.584	0.373	0.627	0.658	0.602	0.320	0.840	0.836	1.000	0.893
Citation: SlopeAve	0.732	0.857	0.695	0.198	0.790	0.824	0.772	0.197	0.958	0.924	0.893	1.000

5.4. まとめ

本章では、指標の時系列変化を観察した上で、発行年の異なる論文群を横断して萌芽的論文を特定する特徴量 Height、Area、Slope、SlopeAve を、in-Degree Centrality、PageRank、データセット内の被引用数に対して設定した。これらの特徴量の上位論文と受賞論文を調べたところ、in-Degree Centrality、PageRank の特徴量 Height、Area、Slope が萌芽的論文を特定することに役立つことが分かった。In-Degree Centrality と PageRank との比較では、PageRank の Height、Area、Slope が最も萌芽的論文を特定するのに効果的であることがわかった。ただし、データセットによっては、分離できないほどの関連分野の論文を含んでいると、関連分野の方の萌芽的論文が特定されるケースが見られた。データセットの範囲をどのように設定するか、本研究と別のテーマとして留意が必要ではあるが、萌芽的論文の特定に、in-Degree Centrality と PageRank の Height、Area、Slop といった特徴量を、ここに提案するものである。

6. 提案手法のシステム化

本研究は、手法の結果を分かりやすく表示するというツールとしての貢献がある。利用する論文データベース、および、クラスター分析の部分については、Thomson Reuters社のWeb of Science、および、東京大学イノベーション政策研究センターの学術俯瞰システムを利用して。本研究における手法を実現するにあたって、ツールとしての貢献があった部分は、Figure 102のように、(a)データセットごとに数十年分の分析を学術俯瞰システムと連携して、各年の指標計算を行ったバックエンドでの作業、(b)操作性に優れたWebインターフェースの提供、および、(c)特徴量の計算にある。バックエンドでの作業(a)について、論文数と分析対象年数と同時分析数にもよるが、学術俯瞰システム(Innovation Policy Research Center, 2013)との連携と自動化により、1データセットに対し、1週間～2か月程度でWebインターフェースが利用できる状況となる。

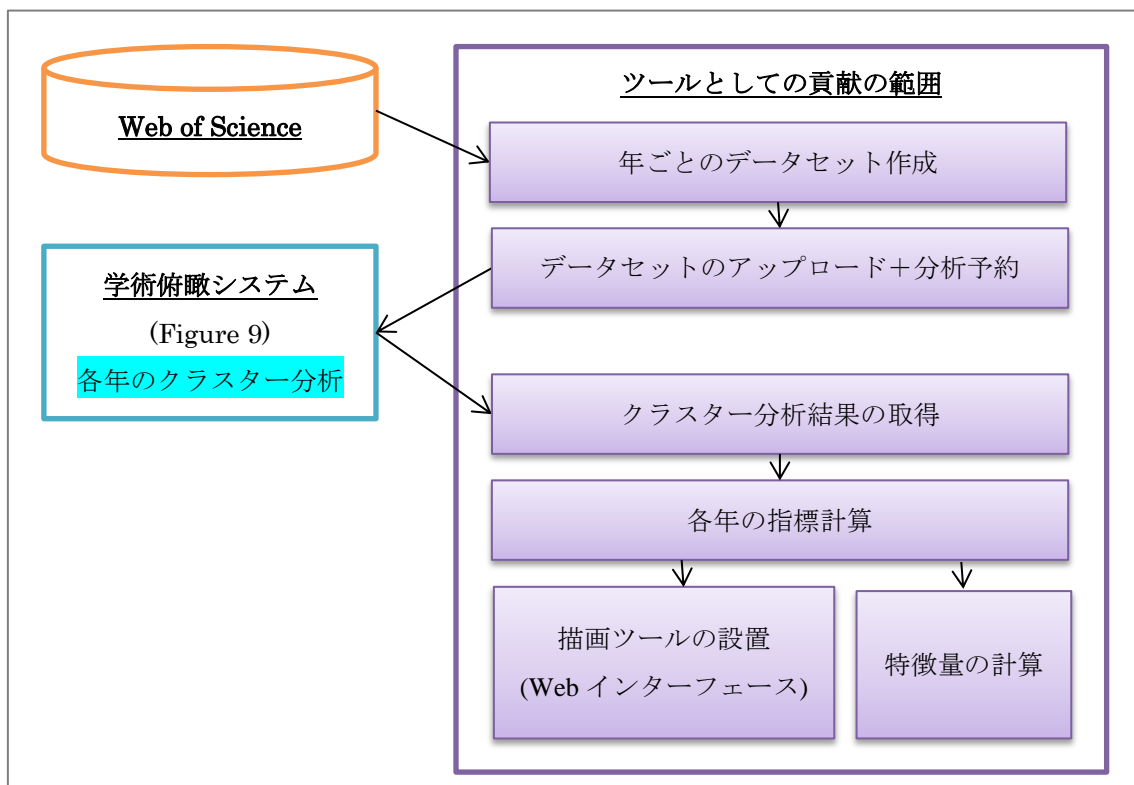


Figure 102: 本研究における手法において、ツールとしての貢献の範囲

Figure 103 は、Web インターフェース(b)上において、被引用数を 100 以上の論文に対して in-Degree Centrality を表示した後、1988 年の TOP5 論文にフォーカスした例である。この Web インターフェースでは、(b-1)バックエンドに問い合わせでデータを取得する機能と、(b-2)Web ブラウザ上で操作する機能がある。前者(b-1)については、以下の機能がある。

- 指標選択 (Showing Indicator)
- 表示年選択 (Year)
- 表示件数絞込み (Rank)
- 論文発行年絞込み (Published Year)
- 著者絞込み (Author)
- 所属クラスター絞込み (Path)
- 閾値絞込み (Threshold)

後者(b-2)には、以下の機能がある。

- 時系列上での絞込み機能 (1988 年灰色部分)
- 論文リスト (Figure 103 の 1/3 より下)での選択が時系列上でフォーカスされる機能
- Export 機能

後者(b-2)の表示は、クライアントサイドで処理を行っているため、表示するコンピュータの性能に依存するがレスポンスが早く、前者(b-1)は、サーバサイドで処理を行っているため、数万件のデータから選択するため数分の時間がかかるが安定した動作が期待できる。

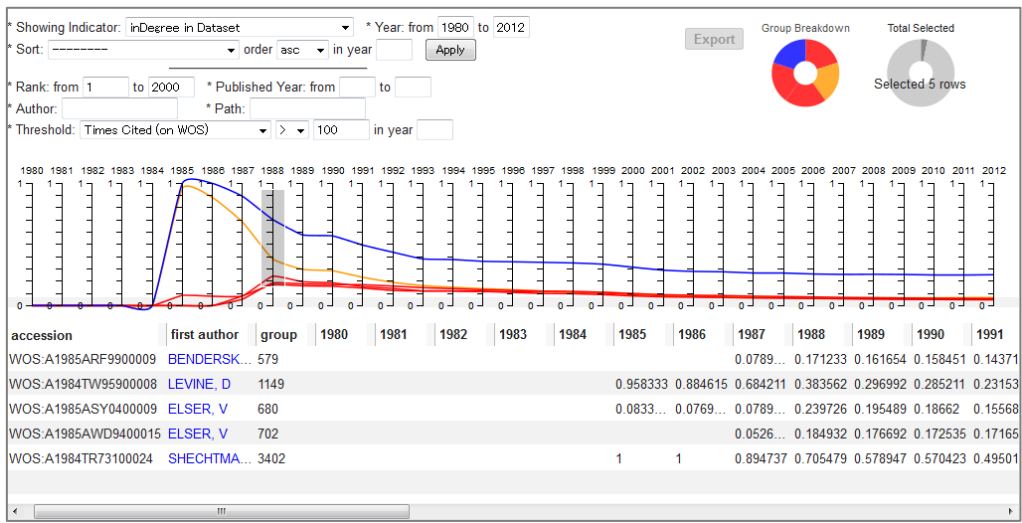


Figure 103: 時系列分析における Web インターフェース

特徴量の計算(c)は、データセットごとに全体を見渡して計算を行っており、順位、特徴量とともに論文のリストとして提供される。

6.1. 提案手法の適用例

本節では、システム化の実用として、未だ、ノーベル賞を受賞していない分野に適応し、萌芽的論文を予測する。予測を行う学術分野は、Table 42 のクラスター分析、および、Table 45 の頻出キーワードから、既にノーベル賞を受賞している分野を除き、これまでの Thomson Reuters 社のノーベル賞候補の予測(Thomson Reuters, 2012)等を加味した結果、機能的核磁気共鳴画像法(fMRI; functional magnetic resonance imaging)を取り上げる。

核磁気共鳴画像法(MRI)やその元の原理である核磁気共鳴(NMR; nuclear magnetic resonance)はすでに、2003 年のノーベル生理学・医学賞、および、1944 年のノーベル物理学賞を受賞している。また、医工学分野のノーベル受賞はノーベル賞全体に対する割合は少ないが、Wilhelm Conrad Röntgen が 1901 年に X 線の発見が第 1 回ノーベル物理学賞を受賞しており、その医学への応用であるコンピュータ断層撮影が 1979 年にノーベル生理学・医学賞を受賞している。

すでにノーベル賞を受賞している MRI ではあるが、MRI は、どこに何があるのか、静止面を取るものであるのに対し、fMRI はどの部位が働くのか、生理現象を動的に捉えることができるという点で、医学的には新たな役割を果たす医療測定機器となっている。近年では、Table 42 の最大サブクラスターである C1-1、および、Table 45 のキーワードに現れているように、脳・神経分野の研究が盛んになっている。脳内の機能をマッピングするために活性部位を計測する装置が必要であり、それら装置は大きく、手術等で装置を埋め込む侵襲型と、頭の表面から計測する非侵襲型がある。さらに、非侵襲型の計測装置には、脳活動に伴う微弱な電磁界変化を計測する方法と、局所的な血液量変化を計測する方法に大別される。本節で分析を行う fMRI は非侵襲型であり、非侵襲型の後者である局所的な血液量変化を計測する方法を使う。その他にも Table 38 に示すような非侵襲型の計測装置があり、非侵襲型の前者には脳波計と脳磁計、非侵襲型の後者には fMRI、陽電子放射断層撮影法(PET)、機能的近赤外分光法(fNIRS)がある。これら脳機能計測装置の中で、fMRI は、ノーベル賞の予想候補としても提供されているトムソン・ロイター引用栄誉賞を、小川誠二氏が 2009 年に受賞している分野でもある(Thomson Reuters, 2012b)。

今回、fMRI の分析をするにあたって、Thomson Reuters 社の論文サービス Web of Science より、「functional and ("magnetic resonance imaging" OR MRI)」というクエリーで、タイトル、アブストラクト、キーワード、キーワード・プラスを検索し、35,985 論文の書誌情報を得た。そのうち、2012 年末時点で引用ネットワークを形成可能な 30,490 論文の書誌情報を対象に、

複数の指標を各年について算出した。Figure 104、および、Table 39 に、対象となる論文の出版年の分布と、書誌情報を収集した 2013-08-22 時点での被引用数(以降、最新被引用数とする)の分布を示す。

Table 38：代表的な脳機能の計測装置

	MEG 脳磁図	EEG 脳波計	PET 陽電子放射断層 撮影法	fMRI 機能的磁気共鳴 イメージング	fNIRS 機能的近赤外分 光法
測定対象	磁気 神経電流	電位 神経電流	ガンマ線 脳血流量・脳血 液量代謝物質変 化	電磁波 脳血流変化	近赤外光 脳血流変化
時間分解能	ミリ秒	ミリ秒	分	秒	百ミリ秒

※島津製作所のカタログより

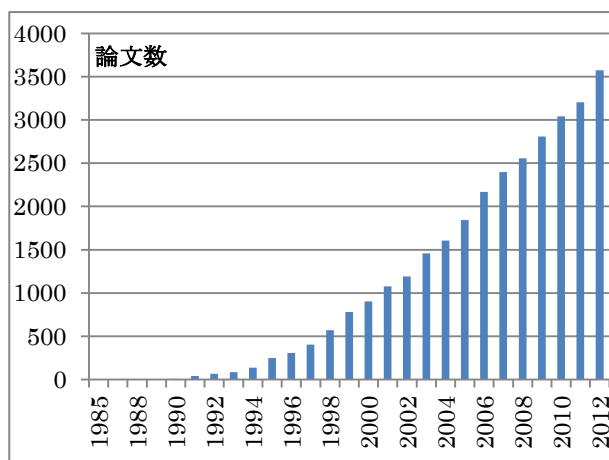


Figure 104：出版年ごとの論文数の分布

Table 39：最新被引用数ごとの論文数の分布

最新被引用数	論文数
0	2,183
1 - 100	25,644
101 - 200	1,743
201 - 300	466
301 - 400	197
401 - 500	104
501 - 600	47
601 - 700	30
701 - 800	14
801 - 900	20
901 - 1000	7
1001 - 2000	27
2001 - 3000	8

このように、機能的核磁気共鳴画像法は現状も成長期にある分野であり、最新被引用数の上位論文が複数、競い合っている分野であるが、この分野において、Figure 105 および Figure 106 のように、in-Degree Centrality と PageRank の時系列分析を行い、時系列分析を観察して特徴的な論文は Table 40 のようになる。また、in-Degree Centrality、および、PageRank の特徴量である Height と Slope を適用し、上位 10 論文について Table 41 に示す。

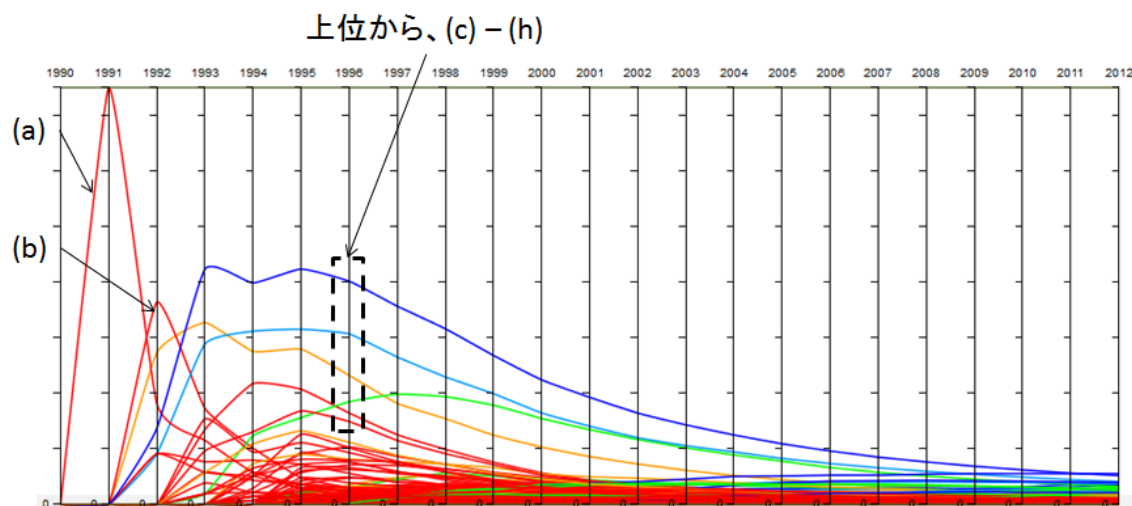


Figure 105 : Time Transition about in-Degree Centralities of 2,664 Papers having more than 100 citations (最大値=0.75) [EMRI]

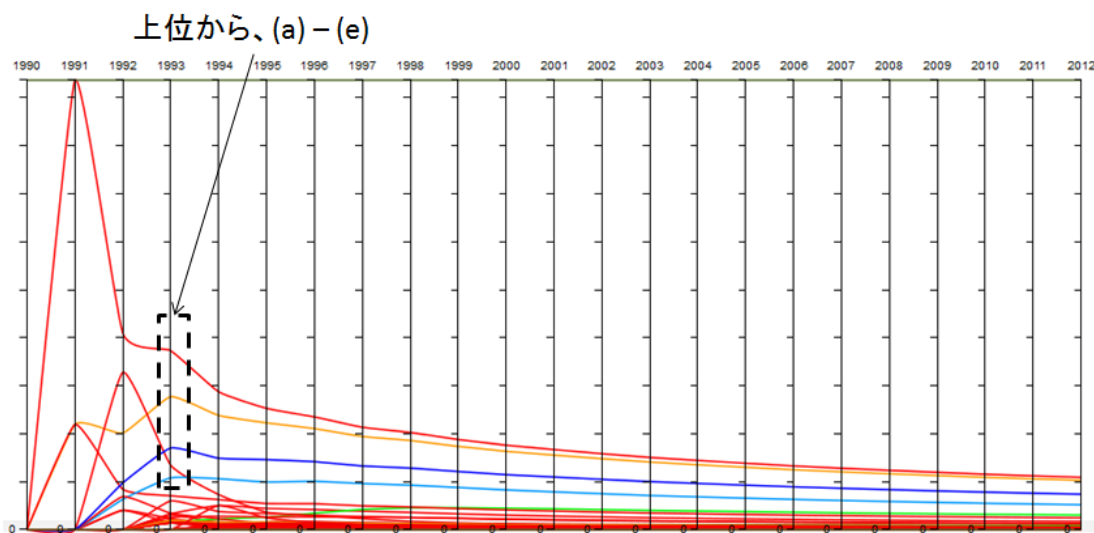


Figure 106 : Time Transition about PageRank of 2,664 Papers having more than 100 citations (最大値=0.4683) [EMRI]

Table 40 : fMRI の時系列分析の図より特徴的な論文

In-Degree Centrality				Page Rank			
	筆頭著者	出版年	最新被引用数		筆頭著者	出版年	最新被引用数
(a)	Moonen, CTW	1990	139	(a)	Moonen, CTW	1990	139
(b)	Evans, AC	1988	181	(b)	Belliveau, JW	1991	1,058
(c)	Kwong, KK	1992	2,463	(c)	Kwong, KK	1992	2,463
(d)	Ogawa, S	1992	1,950	(d)	Evans, AC	1988	181
(e)	Belliveau, JW	1991	1,058	(e)	Ogawa, S	1992	1,950
(f)	Bandettini, PA	1993	1,240				
(g)	Frahm, J	1993	294				
(h)	Turner, R	1993	345				

Table 41 : fMRI の各特徴量の上位 10 論文

In-Degree Centrality

Rank	1st Author	PY	Height	Rank	1st Author	PY	Slope
1	MOONEN, CTW	1990	0.7500	1	MOONEN, CTW	1990	0.7500
2	KWONG, KK	1992	0.4234	2	EVANS, AC	1988	0.3636
3	EVANS, AC	1988	0.3636	3	KWONG, KK	1992	0.2867
4	BELLIVEAU, JW	1991	0.3269	4	BELLIVEAU, JW	1991	0.2727
5	OGAWA, S	1992	0.3153	5	OGAWA, S	1992	0.1976
6	FRAHM, J	1993	0.2174	6	KIM, SG	1993	0.1538
7	BANDETTINI, PA	1993	0.1984	7	FRAHM, J	1993	0.1346
8	TURNER, R	1993	0.1682	8	BANDETTINI, PA	1993	0.1040
9	KIM, SG	1993	0.1538	9	TURNER, R	1993	0.0962
10	KIM, SG	1993	0.1321	10	STEHLING, MK	1991	0.0909

PageRank

Rank	1st Author	PY	Height	Rank	1st Author	PY	Slope
1	MOONEN, CTW	1990	0.4683	1	MOONEN, CTW	1990	0.4683
2	EVANS, AC	1988	0.1639	2	EVANS, AC	1988	0.1639
3	BELLIVEAU, JW	1991	0.1385	3	BELLIVEAU, JW	1991	0.1096
4	STEHLING, MK	1991	0.1096	4	STEHLING, MK	1991	0.1096
5	KWONG, KK	1992	0.0851	5	KWONG, KK	1992	0.0487
6	OGAWA, S	1992	0.0541	6	GREITZ, T	1991	0.0344
7	GREITZ, T	1991	0.0344	7	OGAWA, S	1992	0.0316
8	KIM, SG	1993	0.0300	8	KIM, SG	1993	0.0300
9	STEFAN, H	1987	0.0252	9	STEFAN, H	1987	0.0252
10	FRAHM, J	1993	0.0251	10	SERGENT, J	1992	0.0201

日経サイエンスの記事(青木, 2011)において、小川誠二氏がノーベル賞に近い人たちの一人として挙げられているが、本研究で提示した手法においても、小川氏の論文(1992)が **in-Degree Centrality**、および、**PageRank** において、特徴量の上位に登場しており、萌芽的論文の一つと認められる。その他、先行研究としてではなく、fMRI を含む医用イメージングについて記述されたサーベイ論文(遠藤, 2004)では、提案手法で特定した萌芽的論文リストに含まれる **Belliveau** 氏の論文(1991)、小川氏の論文(1992)に言及されており、提案手法で特定した萌芽的論文の中に

専門家が萌芽的と考える論文が含まれる。しかしながら、1992年の小川氏の論文も重要であるものの、fMRIの第一番目とされる萌芽期の1990年の小川氏の論文を本システムでは特定できていない。これは、fMRIやBOLD法といった本学術分野を特徴づけるキーワードが確立しておらず、データセットの収集段階で含まれていないためである。本研究とは別にデータセットを適切に切り出すクエリーエクспанションが研究として取り組まれているが、現在の手法でデータセットの適切な切り出しが可能な範囲においては、対象分野に専門的な知見が十分ない状態でも、本システムを利用することにより萌芽的論文を自動的に抽出できることが示された。

7. 結論

本研究は、論文の引用ネットワークの時系列変化から、早期に萌芽的論文を特定するための研究である。

昨今、情報量が急速に増加して、専門家でも把握しきれないほど科学技術の知識が増えている。一方で、技術シーズが製品化・産業化に到達するライフサイクルは短くなっており、競合に対し有利な立場に立つには、早く有望な技術シーズを見つけることが重要である。従来手法では、論文の引用は、論文の著者が自著に関連があるとして選定したものであることから、著者がどんな目的で引用に用いたかの差はあるものの、引用ネットワークを利用することで、専門家の知識をアド・ホックに接続し集大成した知のネットワークを利用することができる。このような引用ネットワークの中心にある論文は、多くの専門家から支持されていると期待できる。一方、論文が多く被引用数を得ることで判断する手法や、直近の数年の情報で判断する手法では、被引用数は少ないが分野を開拓する萌芽的論文を見つけ出すことは難しいという問題点がある。そこで、本論文では、各年の論文の引用ネットワークからネットワーク構造に基づく指標群とそれらに関する特徴量を計算し、分野の登場初期からの指標の中長期的な変化を可視化、定量化することで、萌芽的論文を早期に特定する基礎的な手法を提案した。萌芽的論文とは、複数の重要論文の定義があるうち、特に、分野を開拓した重要論文を指す。

本研究での分析の手法は、まず、論文の書誌情報を Web of Science から収集し、論文の引用情報から引用ネットワークを形成する。引用ネットワークのうち、最大連結成分に関して in-Degree Centrality を始めとする複数の指標を計算する。指標は各年まで論文を集めたデータセットごとに計算し、横軸を時間軸、縦軸を指標の値として、各論文を 1 本の線で指標の値を繋げる。

手法を提案するにあたっては、ノーベル賞受賞分野や ACM チューリング賞から 7 分野を選定し提案手法を適用して、in-Degree Centrality、Communicability Centrality、PageRank が有効であることを明らかにした。これらの 3 指標のうち、Communicability Centrality に関しては、計算量が莫大になるため実用には難しことが実際の分析において判明した。残りの 2 指標では、in-Degree Centrality が多少優れているものの、データセットによっては PageRank の方が有用な場合があり、in-Degree Centrality と PageRank を萌芽的論文の特定に利用できる指標として見出した。制限事項としては、学術分野の登場初期に引用ネットワークが形成できなかった

り、萌芽的論文が2番以降の連結成分に含まれてしまい、萌芽的論文の特定が遅れる場合がある。また、Web of Science に収録されていない論文や書籍は特定範囲から除外される。

本研究で提案手法の有用性を検証することができたので、各データセットで時間を横断して萌芽的論文を判断するために、時系列分析の形状を観察して、in-Degree Centrality と PageRank に対し4つの特徴量を設定した。これらの特徴量のうち、Height、Area、Slope は、萌芽的論文を特定するのに有効である。さらに、先に出版された論文が高い評価を得る性質を PageRank が持つことから、in-Degree Centrality より PageRank の Height、Area、Slope の方が萌芽的論文を特定するのに有効である。これらの特徴量により、基本的には、受賞論文を萌芽的論文として特定可能ではあるが、一部のデータセットにおいて、受賞論文が萌芽的論文の優先的な候補として上がらないことがある。これは、関連領域と分析対象の領域とが一つのデータセットのまま分離できない場合があり、分析対象の領域としたノーベル賞受賞論文の代わりに、他の年のノーベル賞受賞論文や受賞論文と同様に萌芽的論文が検出されている。一部のデータセットにおいては、特徴量で特定した萌芽的論文候補のリストを専門家に見せたところ、受賞論文以外の萌芽的論文が抽出されていることを確認いただいた。

本研究の最後に、本研究で提案する手法を実装したシステムを紹介し、ノーベル賞や ACM チューリング賞を受賞していない分野へ適用した。これからも脳科学分野への貢献が予想される機能的核磁気共鳴画像法(fMRI)に提案手法を適用し、萌芽的論文を特定できることを確認した。本研究の研究成果は、政策や経営の場において投資すべき技術分野を決定せねばならない意思決定者や研究テーマを模索する研究者が、学術分野の重点を探し出す指針を与えるものと確信する。

なお、本研究を実施する中で課題として挙がり、追究しきれなかった事項に、論文の書誌情報を収集するクエリーが適切であったかという点に課題が残る。書誌情報を収集する際に、ノーベル賞の授賞理由や専門家の経験からクエリーを選定する手順では、新しく作られた語が一般に認知されるまで、その新しい語に関する将来有望な領域や学術論文が検知されないのではないかという懸念がある。本研究で扱ったデータセットの中では、準結晶(2011chem)は全く新規の概念であったし、iPS細胞(2012med)においても「stem cell」というのは現在では定着した用語ではあるが、Gurdon氏が論文を発表した1962年当時には語が定着しておらず、「stem cell」の語だけではGurdon氏が論文を抽出できない可能性もあった。本研究の目的における「早期に」萌芽

的論文を特定することを考えると、(A)設定したクエリーでは不足する情報を適切に拡張する手法、ならびに、(B)新規の専門用語をデータベースから自動的に抽出する手法が求められるものとする。前者(A)の手法については、コーパス拡張やクエリーエクспанションとして、少し周辺領域を編入するだけで膨大にデータが拡大するところ、適切なサイズに切り出す研究が行われている。後者(B)の手法については、定期的に作動するシステムを構築すれば実装可能である。例えば、以下のようなルールで、常に Web of Science 等の論文データベースを監視すれば、最近人気のトピックや新規に登場したトピックが抽出できよう。

- 最新年に発行された論文のキーワードを抽出し、出現頻度の高いキーワードを特定する。
- 最新年に発行された論文のキーワードを抽出し、前年と比較して伸びが大きいキーワードを特定する。

出現頻度の高いキーワードに関しては、論文データベースに既に設定されたキーワードが存在しており、付録 C.2 のように、既存のキーワードを集計することで簡単に実現できる。学術分野に特化した専門用語(Hjørland, 2002)の抽出においては、前年と比較して伸びが大きいキーワードを抽出することが有効であると考えられる。ただし、手法の(A)(B)ともに、必要とするコンピュータ能力や時間が作業途中で膨張したり初めから膨大であったり、研究周辺のリソースにおいてもチャレンジングな課題である。

本研究においては、キーワードに基づき書誌情報を収集したため、キーワードに関係する課題も抽出されたが、書誌情報を収集するにあたっては、論文誌や国・地域など、別の切り口でデータセットを準備することも考えられる。ただし、論文誌や国・地域では、他の論文誌や国・地域を介して接続する論文が引用ネットワークに含まれないことが予想されるが、これも本研究とは別の課題である。本研究は、Figure 107 のように、選定された書誌情報の先に適用する手法について研究したものである。本研究、ならびに、大量の書誌情報を扱う分析は時間がかかるため、将来的には、適切なクエリーを生成するシステムと共に、提案手法で分析するシステムが定期的に作動し、常に分析の完了した情報を備えておくといった利用方法が想定される。

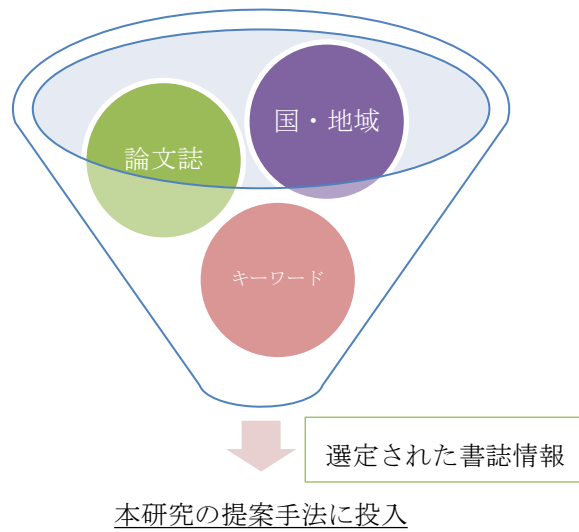


Figure 107: 本研究と選定された書誌情報との関係

いくつかの課題や制限事項はあるものの、本研究で議論した、論文の引用ネットワークの分析に時間軸を取り入れて萌芽的論文を特定する手法は、科学技術に携わる意思決定者や研究者の支援に役立つであろう。例えば、実務における利用場面を考えると、政策の決定プロセスにおいて科学的証拠を提供できるものと考え。また、企業や機関における研究開発において、有望なテーマを探すため、あるいは、研究から開発の段階に進める際に既存の技術シーズから選定する判断材料として利用できるものと考え。

参考文献

- Abbott, A., 2010. 体外受精の先駆者にノーベル医学生理学賞. *Nature Digest*, 7(12), pp.3-4.
Available at: <http://www.nature.com/doi/10.1038/ndigest.2010.101203> [Accessed August 27, 2013].
- Adams, J., 2005. Early citation counts correlate with accumulated impact. *Scientometrics*, 63(3), pp.567–581.
- Aschoff, V., 1983. The early history of the binary code. *IEEE Communications Magazine*, 21(1), pp.4–10.
- Auerswald, P. & Branscomb, L., 2003. Valleys of Death and Darwinian Seas: Financing the Invention to Innovation Transition in the United States. *The Journal of Technology Transfer*, 28(3-4), pp.227–239. Available at: <http://dx.doi.org/10.1023/A:1024980525678>.
- Augier, M. & March, J.G., 2001. Remembering Herbert A. Simon (1916-2001). *Public Administration Review*, 61(4), pp.396–402. Available at: <http://doi.wiley.com/10.1111/0033-3352.00043> [Accessed February 20, 2014].
- Baba, M., 2013. Fukushima accident: What happened? *Radiation Measurements*, 55, pp.17–21. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1350448713000267> [Accessed February 20, 2014].
- Bjork, S., Offer, A. & Söderberg, G., 2013. Time series citation data: the Nobel Prize in economics. *Scientometrics*, pp.1–12. Available at: <http://dx.doi.org/10.1007/s11192-013-0989-5>.
- Björneborn, L. & Ingwersen, P., 2004. Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), pp.1216–1227. Available at: <http://doi.wiley.com/10.1002/asi.20077> [Accessed July 29, 2012].
- Börner, K., Chen, C. & Boyack, K.W., 2003. Visualizing Knowledge Domains. *Annual Review of Information Science and Technology*, pp.179–255.
- Bostock, M., 2013. Data Driven Documents. Available at: <http://d3js.org/>.

- Bower, J.L. & Christensen, C.M., 1995. Disruptive Technologies: Catching the Wave. *Harvard Business Review*, 73(1), pp.43–53. Available at: <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=9501303527&lang=ja&site=ehost-live>.
- Boyack, K.W., Klavans, R. & Börner, K., 2005. Mapping the backbone of science. *Scientometrics*, 64(3), pp.351–374. Available at: <http://link.springer.com/10.1007/s11192-005-0255-6> [Accessed September 8, 2013].
- Brandes, U., 2008. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2), pp.136–145. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0378873307000731> [Accessed August 19, 2013].
- CHALFIE, M. et al., 1994. GREEN FLUORESCENT PROTEIN AS A MARKER FOR GENE-EXPRESSION. *SCIENCE*, 263(5148), pp.802–805.
- Chen, C. et al., 2009. Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, 3(3), pp.191–209. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1751157709000236> [Accessed January 12, 2014].
- Chen, C., 1999. Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing & Management*, 35(3), pp.401–420. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0306457398000685> [Accessed September 9, 2013].
- Chen, C. et al., 2002. Visualizing and tracking the growth of competing paradigms: Two case studies. *Journal of the American Society for Information Science and Technology*, 53(8), pp.678–689. Available at: <http://doi.wiley.com/10.1002/asi.10075> [Accessed August 8, 2013].
- Chen, S.-H., Huang, M.-H. & Chen, D.-Z., 2012. Identifying and visualizing technology evolution: A case study of smart grid technology. *Technological Forecasting and Social Change*, 79(6), pp.1099–1110. Available at: <http://www.sciencedirect.com/science/article/pii/S0040162511002897>.

- Figueroa, P.M., 2013. Risk communication surrounding the Fukushima nuclear disaster: an anthropological approach. *Asia Europe Journal*, 11(1), pp.53–64. Available at: <http://link.springer.com/10.1007/s10308-013-0343-9> [Accessed February 20, 2014].
- Fujita, K. et al., 2012a. Detecting Research Fronts Using Different Types of Combinational Citation. In 17th International Conference on Science and Technology Indicators (STI 2012). Montreal.
- Fujita, K et al., 2012b. Detecting research fronts using different types of weighted citation networks. In *Technology Management for Emerging Technologies (PICMET), 2012 Proceedings of PICMET '12*: pp. 267–275.
- Garfield, E., 2006. Citation indexes for science. A new dimension in documentation through association of ideas. 1955. *International journal of epidemiology*, 35(5), pp.1123–7; discussion 1127–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16987841> [Accessed July 16, 2012].
- Garfield, E., 2009. From the science of science to Scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics*, 3(3), pp.173–179. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1751157709000297> [Accessed November 22, 2013].
- Garfield, E. & Welljams-Dorof, A., 1992. Of Nobel class: A citation perspective on high impact research authors. *Theoretical Medicine*, 13(2), pp.117–135. Available at: <http://link.springer.com/10.1007/BF02163625> [Accessed November 25, 2013].
- Gingras, Y. & Wallace, M.L., 2009. Why it has become more difficult to predict Nobel Prize winners: a bibliometric analysis of nominees and winners of the chemistry and physics prizes (1901–2007). *Scientometrics*, 82(2), pp.401–412. Available at: <http://link.springer.com/10.1007/s11192-009-0035-9> [Accessed November 25, 2013].
- Grier, D.A., 1996. The ENIAC, the verb “to program” and the emergence of digital computers. *IEEE Annals of the History of Computing*, 18(1), pp.51–55. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=476561> [Accessed June 20, 2013].
- Hirsch, J.E., 2005. An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), pp.16569–72. Available at:

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1283832&tool=pmcentrez&rendertype=abstract> [Accessed May 21, 2013].

Hughes, M.E., Peeler, J. & Hogenesch, J.B., 2010. Network dynamics to evaluate performance of an academic institution. *Science translational medicine*, 2(53), p.53ps49. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20944088> [Accessed December 21, 2013].

Innovation Policy Research Center, 2013. Academic Landscape System. Available at: <http://academic-landscape.com/> [Accessed August 31, 2013].

Johnson, M.W., Christensen, C.M. & Kagermann, H., 2008. Reinventing Your Business Model. (cover story). *Harvard Business Review*, 86(12), pp.50–59. Available at: <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=35386627&lang=ja&site=ehost-live>.

Kim, W.C. & Mauborgne, R., 2004. BLUE OCEAN STRATEGY. *Harvard Business Review*, 82(10), pp.76–84. Available at: <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=14599913&lang=ja&site=ehost-live>.

Klavans, R. & Boyack, K.W., 2006. Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology*, 57(2), pp.251–263. Available at: <http://doi.wiley.com/10.1002/asi.20274> [Accessed September 9, 2013].

Klavans, R. & Boyack, K.W., 2009. Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, 60(3), pp.455–476. Available at: <http://doi.wiley.com/10.1002/asi.20991> [Accessed September 9, 2013].

Knott, A.M., 2012. The Trillion-Dollar R&D Fix. *Harvard Business Review*, 90(5), pp.76–82. Available at: <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=74440193&lang=ja&site=ehost-live>.

Koop, G., Pesaran, M.H. & Potter, S.M., 1996. Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1), pp.119–147. Available at: <http://linkinghub.elsevier.com/retrieve/pii/0304407695017534> [Accessed November 26, 2013].

- Länzel, W.O.G., 2001. National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), pp.69–115.
- Latapy, M., Magnien, C. & Vecchio, N. Del, 2008. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1), pp.31–48. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0378873307000494> [Accessed September 11, 2013].
- Levitt, T., 1960. *MARKETING MYOPIA*. *Harvard Business Review*, 38(4), pp.45–56. Available at: <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=6774995&lang=ja&site=ehost-live>.
- Leydesdorff, L., 2004. Clusters and maps of science journals based on bi-connected graphs in. *Journal of Documentation*, 60(4), pp.371–427. Available at: <http://www.emeraldinsight.com/10.1108/00220410410548144> [Accessed August 7, 2013].
- Leydesdorff, L., 2012. Statistics for the dynamic analysis of scientometric data: the evolution of the sciences in terms of trajectories and regimes. *Scientometrics*, 96(3), pp.731–741. Available at: <http://link.springer.com/10.1007/s11192-012-0917-0> [Accessed December 21, 2013].
- Leydesdorff, L., 2011. “Structuration” by intellectual organization: the configuration of knowledge in relations among structural components in networks of science. *Scientometrics*, 88(2), pp.499–520. Available at: <http://dx.doi.org/10.1007/s11192-011-0397-7>.
- Leydesdorff, L. & Rafols, I., 2009. A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, 60(2), pp.348–362. Available at: <http://doi.wiley.com/10.1002/asi.20967> [Accessed September 9, 2013].
- Liu, X., Jiang, T. & Ma, F., 2013. Collective dynamics in knowledge networks: Emerging trends analysis. *Journal of Informetrics*, 7(2), pp.425–438. Available at: <http://www.sciencedirect.com/science/article/pii/S1751157713000059>.

- Lu, L.Y. & Liu, J., 2013. An innovative approach to identify the knowledge diffusion path: the case of resource-based theory. *Scientometrics*, 94(1), pp.225–246. Available at: <http://dx.doi.org/10.1007/s11192-012-0744-3>.
- Manh Cuong, P., Derntl, M. & Klamma, R., 2012. Development Patterns of Scientific Communities in Technology Enhanced Learning. *Journal of Educational Technology & Society*, 15(3), pp.323–335. Available at: <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=79816986&lang=ja&site=ehost-live>.
- Manning, C.D.P.R. & Schütze, H., 2008. *Introduction to Information Retrieval*, Cambridge University Press. Available at: <http://www-nlp.stanford.edu/IR-book/>.
- Miller, G.A., 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review*, 63, pp.81–97. Available at: <http://cogprints.org/730/1/miller.html>.
- Mollin, R.A., 2002. *Rsa and Public-Key Cryptography*, Chapman and Hall/CRC.
- Mowery, D.C. & Simcoe, T., 2002. Is the Internet a US invention?—an economic and technological history of computer networking. *Research Policy*, 31(8-9), pp.1369–1387. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0048733302000690> [Accessed June 19, 2013].
- NetworkX developer Team, 2013. *NetworkX*. Available at: <http://networkx.github.com/>.
- Newman, M.E.J., 2004. Fast Algorithm for detecting community structure in networks. *Physical Review E*, 69(066133).
- Newton Press, 2012. *iPS 細胞—夢の再生医療を実現する 第2版*, ニュートンプレス.
- No, H.J. & Park, Y., 2010. Trajectory patterns of technology fusion: Trend analysis and taxonomical grouping in nanobiotechnology. *Technological Forecasting and Social Change*, 77(1), pp.63–75. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0040162509000778> [Accessed November 22, 2013].
- OECD, 2013. *OECD Science, Technology and R&D Statistics*. Available at: <http://www.oecd.org/statistics/> [Accessed June 6, 2013].

- Page, L. et al., 1999. The PageRank Citation Ranking: Bringing Order to the Web., Stanford InfoLab. Available at: <http://ilpubs.stanford.edu:8090/422/>.
- Peters, D.P. & Ceci, S.J., 2010. Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, 5(02), p.187. Available at: http://www.journals.cambridge.org/abstract_S0140525X00011183 [Accessed May 23, 2013].
- ROCHE, I. et al., 2013. ARE RESEARCH LANDSCAPES FROM SUBMITTED PROJECT PROPOSALS OR FROM THE S&T LITERATURE SIMILAR? A COMPARISON USING TEXT MINING AND CLUSTERING. In IAMOT 2013. Porto Alegre, Brazil.
- Rousseau, D.M., 2006. IS THERE SUCH A THING AS “EVIDENCE-BASED MANAGEMENT”? *Academy of Management Review*, 31(2), pp.256–269. Available at: <http://amr.aom.org/cgi/doi/10.5465/AMR.2006.20208679> [Accessed June 4, 2013].
- Sakata, I., Sasaki, H. & Inoue, T., 2011. Structure of International Research Collaboration in Wind and Solar Energy. *Industrial Engineering and Engineering Management (IEEM)*, 2011 IEEE International Conference on.
- Sasaki, H. et al., 2010. Detecting the Valley of International Academic Collaboration in Renewable Energy. *Fuel Cell*, pp.99–103.
- Shibata, N. et al., 2009. Comparative study on methods of detecting research fronts using different types of citation. *J. Am. Soc. Inf. Sci. Technol.*, 60(3), pp.571–580. Available at: <http://portal.acm.org/citation.cfm?id=1527090.1527098>.
- Shibata, N. et al., 2008. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, 28(11), pp.758–775. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0166497208000436>.
- Shibata, N. et al., 2011. Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications. *Technological Forecasting and Social Change*, 78(2), pp.274–282. Available at: <http://www.sciencedirect.com/science/article/B6V71-50NY2N0-2/2/1f6adc9f81d4fe5ff5c4740f88d9d3a4>.

- Slyder, J.B. et al., 2011. Citation pattern and lifespan: a comparison of discipline, institution, and individual. *Scientometrics*, 89(3), pp.955–966. Available at: <http://link.springer.com/10.1007/s11192-011-0467-x> [Accessed June 21, 2013].
- Small, H., 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), pp.265–269. Available at: <http://doi.wiley.com/10.1002/asi.4630240406> [Accessed August 27, 2013].
- Small, H., 2013. Tracking and predicting growth areas in science. *Scientometrics*, 68(3), pp.595–610. Available at: <http://link.springer.com/10.1007/s11192-006-0132-y> [Accessed August 15, 2013].
- Small, H., 2006. Tracking and predicting growth areas in science. *Scientometrics*, 68(3), pp.595–610. Available at: <http://www.akademai.com/index/711P28W6178K22X0.pdf> [Accessed July 19, 2012].
- Small, H., 1999. Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50(9), pp.799–813. Available at: [http://doi.wiley.com/10.1002/\(SICI\)1097-4571\(1999\)50:9<799::AID-ASI9>3.3.CO;2-7](http://doi.wiley.com/10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.3.CO;2-7) [Accessed September 9, 2013].
- De Solla Price, D.J., 1965. Networks of Scientific Papers. *Science*, 149(3683), pp.510–515. Available at: <http://www.sciencemag.org/cgi/doi/10.1126/science.149.3683.510> [Accessed August 15, 2013].
- Springer, 2013. Mean Reciprocal Rank - SpringerReference. Springer. Available at: <http://www.springerreference.com/docs/html/chapterdbid/63599.html> [Accessed September 6, 2013].
- The Royal Swedish Academy of Sciences, 2011a. Scientific Background on the Nobel Prize in Chemistry 2011 - The Discovery of Quasicrystals. Available at: http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2011/advanced-chemistry-prize2011.pdf [Accessed July 4, 2013].
- The Royal Swedish Academy of Sciences, 2011b. THE NOBEL PRIZE IN CHEMISTRY 2011 (INFORMATION FOR THE PUBLIC) - Crystals of golden proportions. Available at: http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2011/popular-chemistry-prize2011.pdf [Accessed July 4, 2013].

- Thomson Reuters, 2012a. 「トムソン・ロイター引用栄誉賞」 2002-2012 の日本人受賞者一覧.
Thomson Reuters. Available at:
<http://ip-science.thomsonreuters.jp/press/release/2012/2012-Citation-Laureates/laureates-jp/> [Accessed September 11, 2013].
- Thomson Reuters, 2012b. 2012 Citation Laureates. Available at:
<http://sciencewatch.com/nobel> [Accessed July 28, 2013].
- Toffler, A., 1980. 第三の波, 日本放送出版協会.
- Tsvetovat, M., Kouznetsov, A. & 長尾高弘 (訳), 2012. オープンソースで学ぶ社会ネットワーク分析——ソーシャル Web の「つながり」を見つけ出す, オライリー・ジャパン.
- Tu, Y.-N. & Seng, J.-L., 2012. Indices of novelty for emerging topic detection. *Information Processing & Management*, 48(2), pp.303–325. Available at:
<http://www.sciencedirect.com/science/article/pii/S0306457311000768>.
- United Nations, World Population Prospects, the 2012 Revision. Available at:
<http://esa.un.org/unpd/wpp/index.htm> [Accessed June 30, 2013].
- Vanclay, J.K., 2011. Impact factor: outdated artefact or stepping-stone to journal certification? *Scientometrics*, 92(2), pp.211–238. Available at:
<http://www.springerlink.com/index/10.1007/s11192-011-0561-0> [Accessed November 9, 2012].
- Wang, J., 2012. Citation time window choice for research impact evaluation. *Scientometrics*, 94(3), pp.851–872. Available at: <http://link.springer.com/10.1007/s11192-012-0775-9> [Accessed June 20, 2013].
- Winegrad, D., 1996. Celebrating the birth of modern computing: the fiftieth anniversary of a discovery at the Moore School of Engineering of the University of Pennsylvania. *IEEE Annals of the History of Computing*, 18(1), pp.5–9. Available at:
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=476556> [Accessed February 20, 2014].
- アーリング・ノルビ & 千葉喜久枝(訳), 2011. ノーベル賞はこうして決まる: 選考者が語る自然科学三賞の真実, 創元社.

- クレイトン・クリステンセン, 玉田俊平太 (監修) & 伊豆原弓 (訳), 2001. イノベーションのジレンマ—技術革新が巨大企業を滅ぼすとき (Harvard business school press), 翔泳社.
- トーマス・C・レドマン & 栗原潔(訳), 2010. 戦略的データマネジメント 企業利益は真のデータ価値にあり, 翔泳社.
- ハーバート・A. サイモン, 稲葉元吉(訳) & 吉原英樹(訳), 1999. システムの科学, パーソナルメディア.
- 阿部剛久, 2006. 科学の近代化からライプニッツに至る話題の思想史的断片. 数理解析研究所講究録. Available at:
<http://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/pdf/1513-2.pdf>.
- 遠藤真広, 2004. 医用イメージング技術の最近の発展. 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理, 87(1), pp.3-18. Available at:
<http://ci.nii.ac.jp/naid/110003170977/>.
- 甲元真人, 1987. 一次元, 二次元準結晶の電子状態. 日本物理學會誌, 42(5), pp.433-442.
Available at: <http://ci.nii.ac.jp/naid/110002068435/>.
- 国家戦略会議 フロンティア分科会, 2012. 叡智のフロンティア部会報告書, Available at:
<http://www.cas.go.jp/jp/seisaku/npu/policy09/pdf/20120706/hokoku4.pdf>.
- 財務省, 2012. 日本の財政関係資料 平成 24 年度予算案 補足資料. Available at:
http://www.mof.go.jp/budget/fiscal_condition/related_data/sy014_24_02.pdf [Accessed June 6, 2013].
- 阪彩香, 伊神正貫 & 桑原輝隆, 2010. サイエンスマップ 2008—論文データベース分析 (2003年から 2008 年) による注目される研究領域の動向調査—報告書, Available at:
<http://hdl.handle.net/11035/686>.
- 森政弘, 1976. レオナルド・ダ・ビンチから学ぶべきもの : その合一性と発想. 日本機械學會誌, 79(697), pp.1115-1118.
- 青木慎一 & 古田彩, 2011. ブレークスルーの実現者. 日経サイエンス 11 月号, pp.46-49.
- 石浦章一(監修) & 生化学若い研究者の会(編), 2009. 光るクラゲがノーベル賞をとった理由, 日本評論社.

総務省, 2011. 情報通信白書平成 23 年版. Available at:

<http://www.soumu.go.jp/johotsusintokei/whitepaper/h23.html> [Accessed June 20, 2013].

中西貴之, 2012. 実はおもしろい化学反応, 株式会社技術評論社.

朝日新聞大阪本社科学医療グループ, 2011. iPS 細胞とはなにかー万能細胞研究の現在, 講談社.

島津製作所, 近赤外光脳機能イメージングによる脳機能研究の展開, Available at:

<http://www.an.shimadzu.co.jp/products/general/life.htm>.

内閣府, 2008. イノベーション 25. Available at: <http://www.cao.go.jp/innovation/index.html>.

文部科学省, 2013. SciREX 科学技術イノベーション政策における「政策のための科学」.

Available at: <http://scirex.mext.go.jp/>.

謝辞

本研究を行うにあたり、東京大学大学院工学系研究科技術経営戦略学専攻 坂田一郎教授、同専攻 森純一郎特任講師には、研究活動の方向性を導き、また、研究活動の支援をいただき、心より感謝いたします。先生方のご指導は、研究活動を通して人生の転機にもなる多くを経験する機会をいただきました。

東京工業大学大学院イノベーションマネジメント研究科 梶川裕矢准教授には、研究のアプローチ方法について背中を押していただきました。先生に研究のアプローチ方法を見出していたかなければ、本研究の手法は試行錯誤したうちのひとつとして漫然と押し流されて、私の研究活動そのものが未だ迷子になっていたかもしれません。また、東京大学大学院工学系研究科総合研究機構イノベーション政策研究センター 学術俯瞰システムの開発チームには、研究活動の技術面においてサポートいただきました。ここに深く感謝の気持ちを表したいと思います。

また、東京大学大学院工学系研究科技術経営戦略学専攻 影山和郎教授、同専攻 阿部力也特任教授、同専攻 松尾豊准教授、東京大学政策ビジョン研究センター 谷口武俊教授には研究上のアドバイスをいただき、より充実した論文にすることができました。立命館大学 上原哲太郎教授、京都大学 仙石慎太郎准教授、東邦大学 金岡晃講師には、深い専門知識に基づき、本研究の分析に関してご教示いただきました。本当にありがとうございます。

東京大学大学院工学系研究科総合研究機構イノベーション政策研究センター 石原絢さん、粥川敬子さんには、研究活動から精神的なフォローまでサポートをいただきました。深く感謝しております。

最後に、私が博士を取得したいと考え関係者の説得を始めた頃から、四半世紀以上が過ぎました。14年前に、それまでの長年の積み重ねむなしく学業を続けることが不能となりましたが、その時に、命を繋ぐ助けをご紹介くださったのが、当時の北海道大学工学研究科所属の松永太氏でした。彼の親切がなければ、本研究が存在することはなかったでしょう。ここに多大なる感謝の意を表明します。

研究業績

以下に示す研究業績のうち、本研究に特に関係が深いものには(*)を記す。

査読付論文(英文)

(*)Shino Iwami, Junichiro Mori, Yuya Kajikawa and Ichiro Sakata, "Detection Method of Emerging Leading Papers using Time Transition", *Scientometrics*, (submitted).

査読付論文(和文)

内藤理, 岩見紫乃, 森純一郎, 梶川裕矢, 佐藤啓宏, 工藤俊亮, 池内克史, 2013. 実社会との関係形成で変化するロボット研究の現状-ロボットに関する学術論文の引用関係からの考察, *日本ロボット学会誌*, 31(8).

国際会議

(*)Shino Iwami, Junichiro Mori, Yuya Kajikawa and Ichiro Sakata, "Comparison of Indicators to Detect Emerging Researches using Time Transition in Quasicrystals", *The IEEE International Conference on Industrial Engineering and Engineering Management (IEEM2013)*, Thailand, December 2013.

(*)Shino Iwami, Junichiro Mori, Yuya Kajikawa and Ichiro Sakata, "Detection of Next Researches using Time Transition in Fluorescent Proteins", *The 14th International Society for Informetrics and Scientometrics (ISSI2013)*, Austria, July 2013.

(*)Shino Iwami, Junichiro Mori, Yuya Kajikawa, Tetsutaro Uehara and Ichiro Sakata, "Detection of Promising Fields using Time Transitions in Cryptology", *The 22th International Conference for Management of Technology (IAMOT2013)*, Brazil, April 2013.

Asato Tashiro, Hisato Tashiro, Shino Iwami and Ichiro Sakata, "Bibliometric Analysis of Net Zero Energy Buildings", *Portland International Conference on Management Engineering and Technology 2013 (PICMET'13)*, USA, July-August, 2013.

Shino Iwami and Ichiro Sakata, "Classification of Renewable Policies with Indicators of Science and Market", the 21th International Conference for Management of Technology (IAMOT2012), Taiwan, March 2012.

国内学会

内藤理, 岩見紫乃, 森純一郎, 梶川裕矢, 工藤俊亮, 池内克文. 実社会との関係を深めるロボットに関する研究の現状と今後. 日本ロボット学会 第30回記念 学術講演会, September, 2012.

付録 A： ネットワーク指標の計算方法

本節では、2.3 節で述べたネットワーク指標(NetworkX developer Team, 2013)について、その算出方法を説明する。

A.1. DEGREE CENTRALITY (度数中心性) と IN-DEGREE CENTRALITY

Degree Centrality (度数中心性) は、ネットワーク内での関係が多いほど大きくなる指標である。

度数(degree)とは、引用ネットワークの場合、引用関係の数を表す。引用関係は関係の方向性を考えると二種類あり、ある論文が他の論文を引用する引用数を **out-Degree**、ある論文が他の論文から引用される被引用数を **in-Degree** とする。本研究では、**out-Degree** は論文筆者の意思により自由に選択でき、既に発表された論文の引用数は変化しないことから、**out-Degree** は使用せず、**in-Degree** のみを使用する。

論文の度数中心性は、引用ネットワーク内のすべての論文数を m 、論文 X が実際に持っている引用関係の数を n とすると、論文 X がすべての論文と関係を持つ可能性 $m-1$ で割った値となる。

In-Degree の場合は、 n が被引用数のみとなり、 m は論文数のままである。

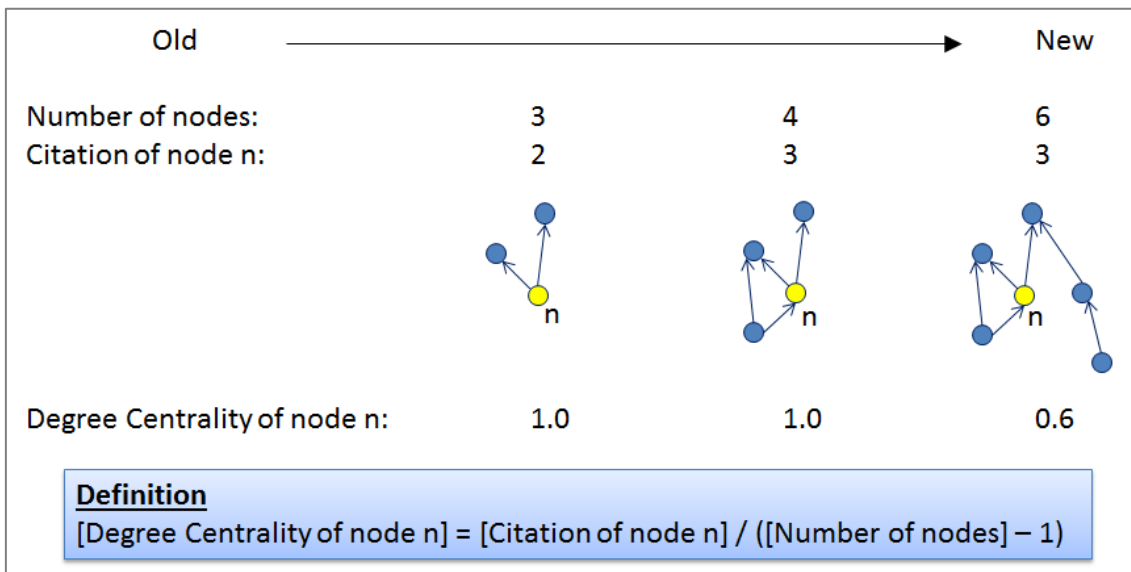


Figure 108: 次数中心性 (Degree Centrality) の定義と、ネットワークの成長に伴う変化

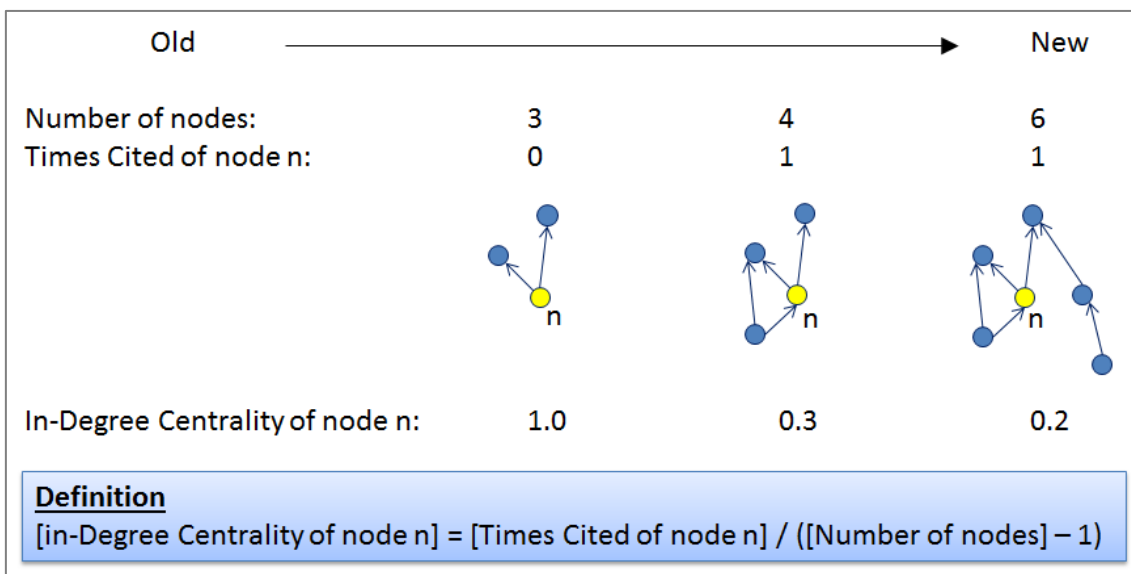


Figure 109: in-Degree Centrality の定義と、ネットワークの成長に伴う変化

A.2. CLOSENESS CENTRALITY (近接中心性)

Closeness Centrality (近接中心性) は、他の論文との関係が近いほど高くなる指標である。あるノードにおける Closeness Centrality は、他のノードへの平均距離の逆数となる。

A.3. BETWEENNESS CENTRALITY (媒介中心性)

Betweenness Centrality (媒介中心性) は、ボトルネックとなるような最短経路上にいるほど高くなる指標である。あるノード v における Betweenness Centrality は、 v を通る最短距離を全ての最短距離でわったものの合計である。

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

上の式において、 V はすべてのノードの集合からなり、始点ノードを s 、終点ノードを t と表す。 $\sigma(s,t)$ は最短距離、 $\sigma(s,t|v)$ は v を通る場合の最短距離である。

A.4. AVERAGE NEIGHBOR DEGREE

Average Neighbor Degree とは、あるノードに隣接するノードが持つ次数を平均したものである。

A.5. CLOSENESS VITALITY

Closeness Vitality とは、そのノードを除き、すべてのノードのペア間の距離の和の変化である。

A.6. CLUSTERING COEFFICIENT

Clustering Coefficient (クラスタリング係数) は、グラフ内のノードが一緒に集中する傾向の度合いである。

A.7. COMMUNICABILITY CENTRALITY

Communicability Centrality (伝達性中心性) は、Subgraph Centrality と呼ばれ、ノード n における Communicability Centrality は、ノード n に始まりノード n に終わる閉路 (closed walk) の長さの合計である。

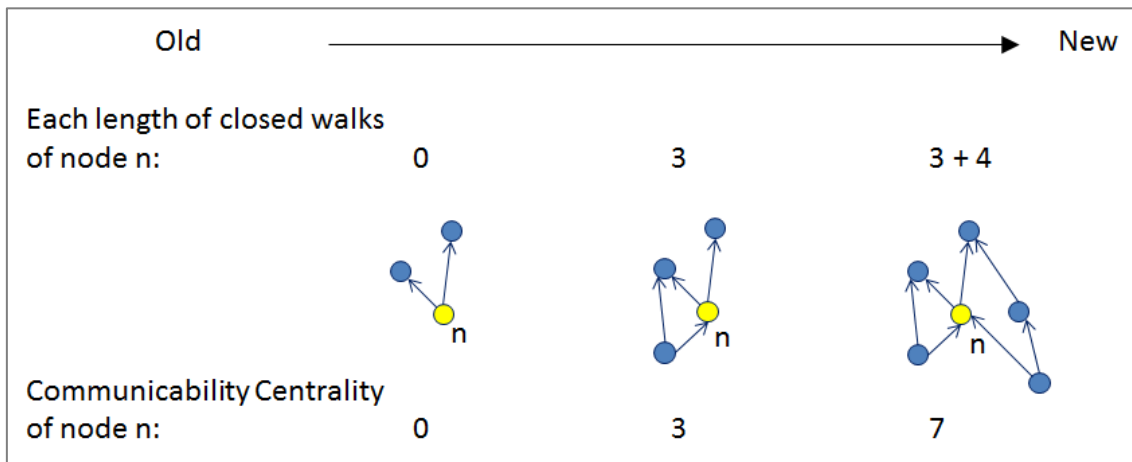


Figure 110: 伝達性中心性 (Communicability Centrality) の、ネットワークの成長に伴う変化

A.8. CORE NUMBER

K -core は、次数が k 以上のノードを含む最大のサブグラフである。Core Number とは、そのノードを含む K -core の値 k である。ここで次数は入次数と出次数の和を指す。

A.9. EIGENVECTOR CENTRALITY

Eigenvector Centrality (固有ベクトル中心性) は、中心的な存在になっている他のノードにどれだけつながっているかによって中心的な存在として扱う。本当の有力者は他の有力者と多くつながっているという考え方で示される。

理論上は有望な指標ではあるが、実際に NextworkX での Eigenvector Centrality の計算では非常に負担が大きく計算の欠損も発生したため、本研究では、論文数の少ないデータセットのみの計算となった。

A.10. LOAD CENTRALITY

Load Centrality (負荷中心性) は、そのノードを通過するすべての最短経路の割合である。Load Centrality は Betweenness Centrality とは若干異なり、Figure 111 のように、始点ノードから終点ノードを辿る間の重みが異なる (Brandes, 2008)。



Figure 111 : load (左) と betweenness (右) の、始点ノードと終点ノード間での重みの違い

A.11. NODE REDUNDANCY

ノード v の Node Redundancy (あるいは、Redundancy Coefficient) は、ノード v の隣接しているノードのうち、 v の隣接しているノード同士で共引用しているペアの割合である。ノード v に隣接しているノードの集合を $N(v)$ とすると、Node Redundancy は以下の式で求められる (Latapy, 2008)。

$$rc(v) = \frac{|\{\{u, w\} \subseteq N(v), \exists v' \neq v, (v', u) \in E \text{ and } (v', w) \in E\}|}{\frac{|N(v)|(|N(v)|-1)}{2}}$$

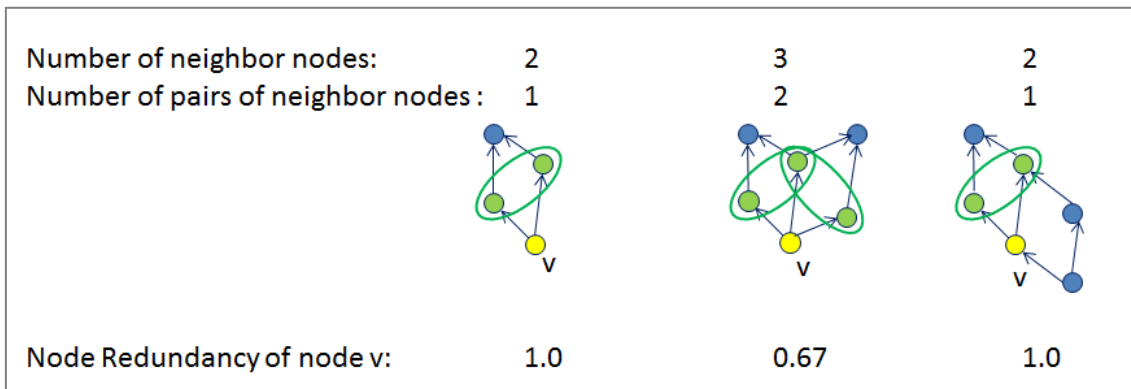


Figure 112 : Node Redundancy の例

A.12. PAGERANK

PageRank はもともと、ウェブのグラフに基づいてすべてのウェブページのランキングを計算する方法であり、Google の検索結果を順位付けするために採用されている。PageRank は論文における被引用数の考えに近く、Figure 113 のように、ウェブページのリンクされた数を人気投票のスコアと考え、人気のあるウェブページからのリンクをリンク元のスコアで重み付けして再帰的に計算する方法である (Page, 1999)。

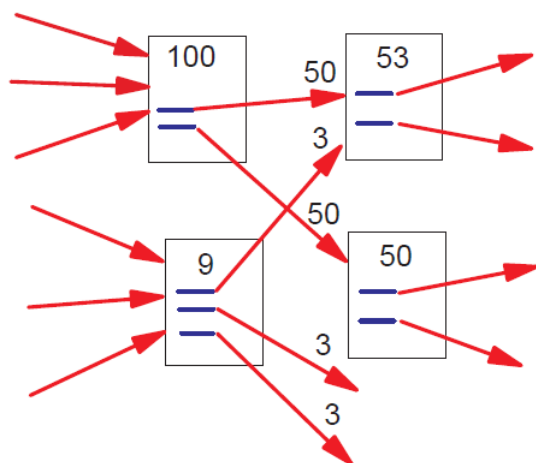


Figure 113 : PageRank の概念

付録 B: 指標の時系列変化の比較

本論にて、各論文のデータセットについて、パフォーマンスのよい3指標(in-Degree Centrality, PageRank, Communicability Centrality)とそのデータセット特有の状況を論ずる場合に絞って、時系列変化の図を提示しているが、実際には、引用ネットワークをクラスターごとにした場合や、他の指標も計算し、図を提供することが可能である。ここでは、本論で除外された理由を明らかにするために、一部のデータセットを代表として紹介する。

B.1. クラスターごとの引用ネットワークから計算した場合

ここでは、3.5 で用いた準結晶のデータセットを用いて引用ネットワークをクラスターごとにした場合の指標の時系列を示す。クラスターごとで計算すると、小さなクラスターの中で大きな指標を持つ論文が時系列変化の中でも大きく表現されるため、同じ指標であってもデータセット内で計算した場合より重要論文の明確性が失われる。一方で、Figure 116 と Figure 117 のように、本論で論じた重要論文が突出した値を持つ特徴はノーベル賞受賞論文(青線)を見ると引き継がれるため、計算の母集団となる論文数に関わらず Degree Centrality および in-Degree Centrality の特徴が出現する。

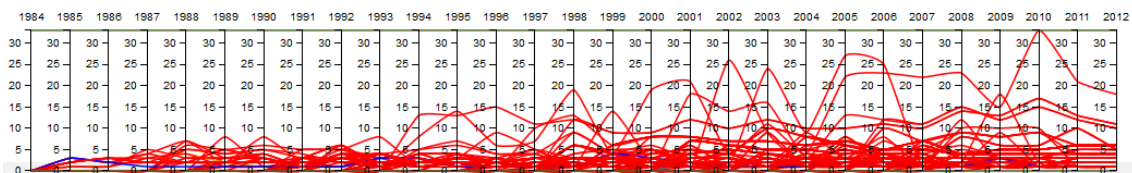


Figure 114 : Time Transition about Cluster Number of 155 Papers having more than 100 citations (最大値=33)

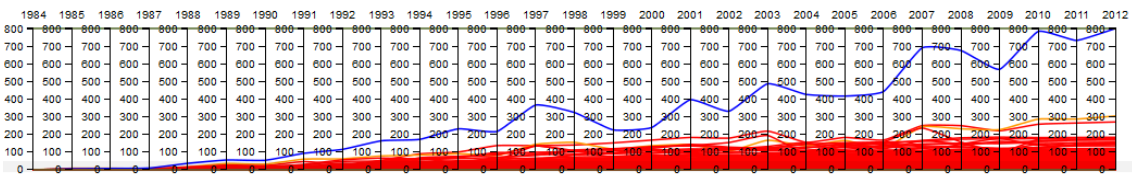


Figure 115 : Time Transition about Number of within-Cluster Citation of 155 Papers having more than 100 citations (最大値=803)

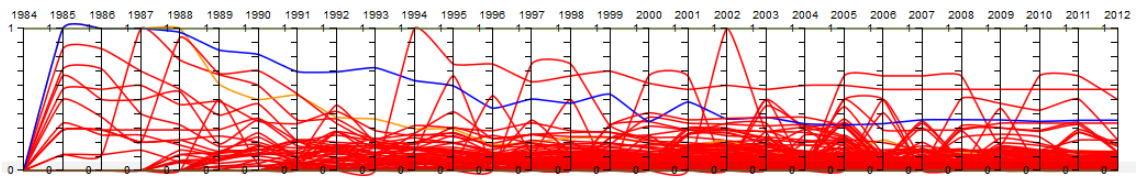


Figure 116 : Time Transition about within-Cluster Degree Centralities of 155 Papers having more than 100 citations (最大值=1)

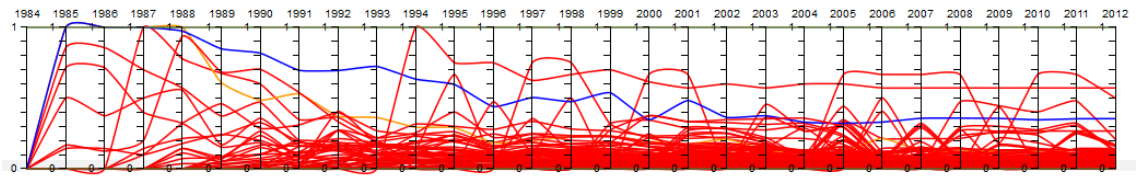


Figure 117 : Time Transition about within-Cluster in-Degree Centralities of 155 Papers having more than 100 citations (最大值=1)

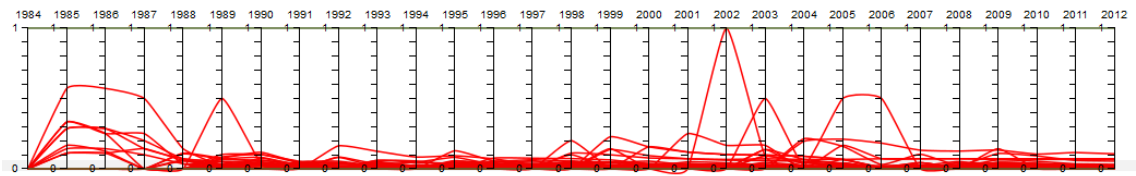


Figure 118 : Time Transition about within-Cluster out-Degree Centralities of 155 Papers having more than 100 citations (最大值=1)

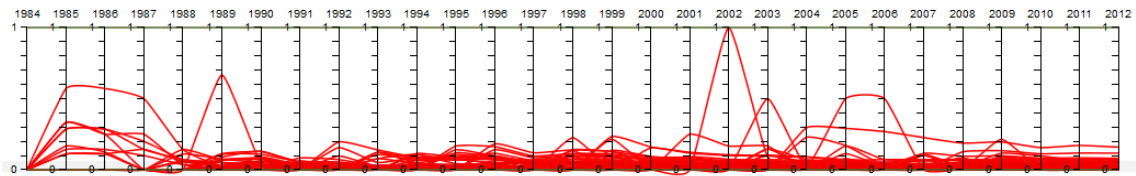


Figure 119 : Time Transition about within-Cluster Closeness Centralities of 155 Papers having more than 100 citations (最大值=1)

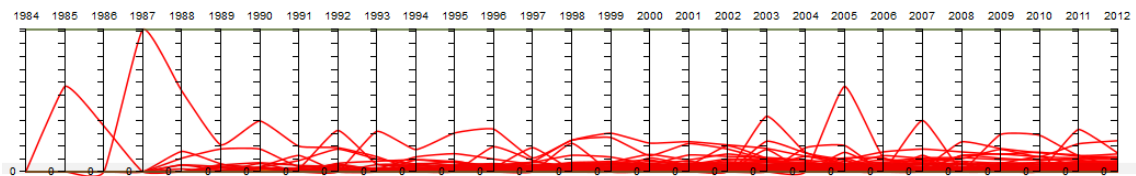


Figure 120 : Time Transition about within-Cluster Betweenness Centralities of 155 Papers having more than 100 citations (最大值=0.05556)

B.2. データセット内の引用ネットワークから計算した場合

ここでは、3.5 で用いた準結晶のデータセットを用いてデータセットの最大連結成分を引用ネットワークとした場合の指標の時系列を示す。なお、ここでの3指標(in-Degree Centrality, PageRank, Communicability Centrality)は、3.5 で提示したものと同一である。

本論では、Degree Centrality の代わりに in-Degree Centrality を使用したが、被引用数が引用数よりはるかに大きいので、Degree Centrality をそのまま使用することができよう。

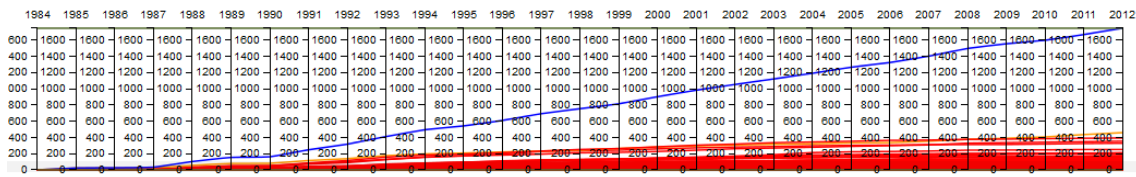


Figure 121 : Time Transition about Number of within-Dataset Citation of 155 Papers having more than 100 citations (最大値=1745)

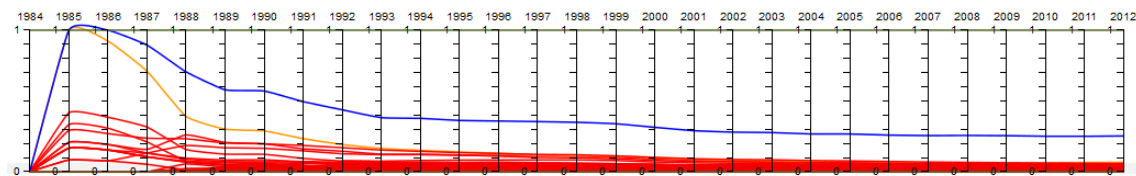


Figure 122 : Time Transition about within-Dataset Degree Centralities of 155 Papers having more than 100 citations (最大値=1)

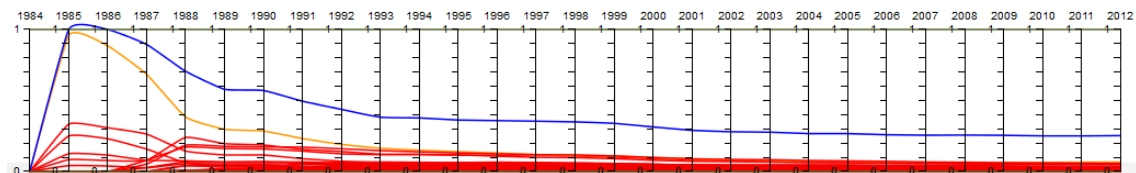


Figure 123 : Time Transition about within-Dataset in-Degree Centralities of 155 Papers having more than 100 citations (最大値=1)

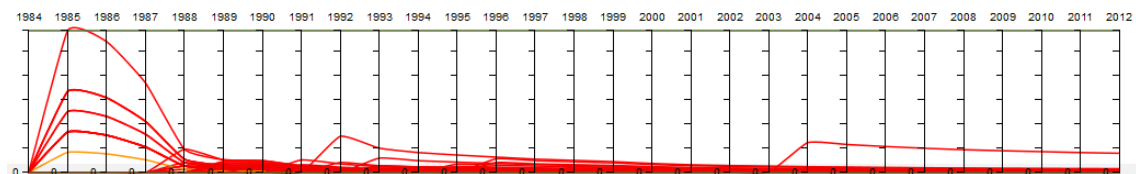


Figure 124 : Time Transition about within-Dataset out-Degree Centralities of 155 Papers having more than 100 citations (最大値=0.2917)

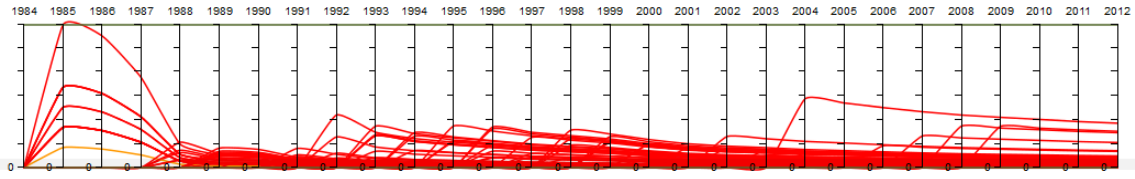


Figure 125 : Time Transition about within-Dataset Closeness Centralities of 155 Papers having more than 100 citations (最大値=0.2963)

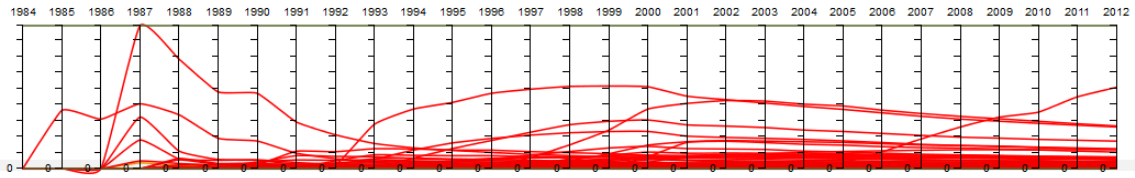


Figure 126 : Time Transition about within-Dataset Betweenness Centralities of 155 Papers having more than 100 citations (最大値=0.008890)

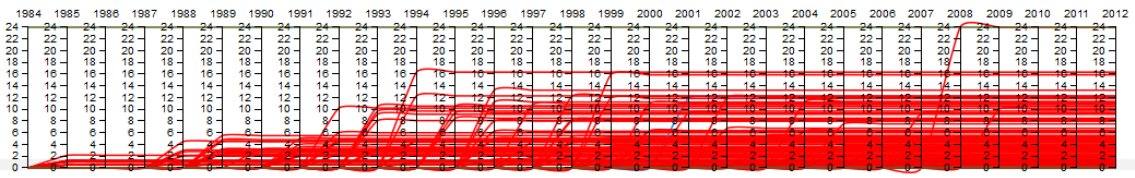


Figure 127 : Time Transition about within-Dataset Average Neighbor Degree of 155 Papers having more than 100 citations (最大値=24)

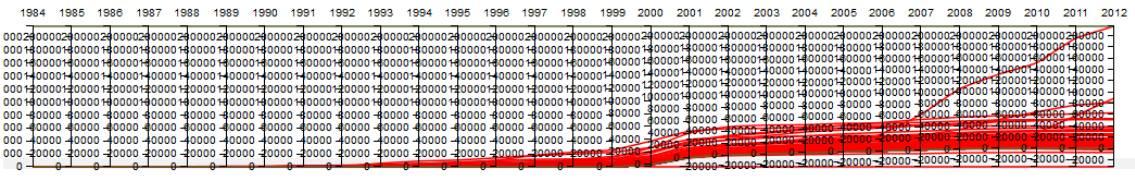


Figure 128 : Time Transition about within-Dataset Closeness Vitality of 155 Papers having more than 100 citations (最大値= 2.166e+5, 2001 年以降は最小値がマイナス)

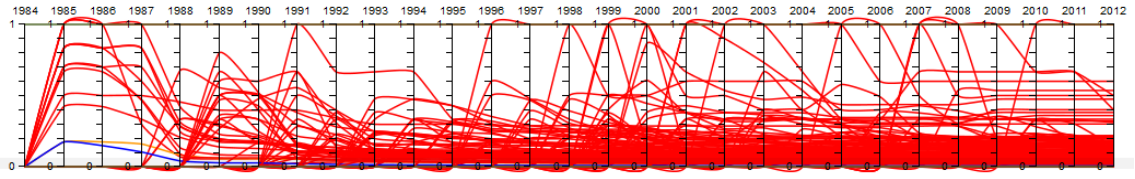


Figure 129 : Time Transition about within-Dataset Clustering Coefficient of 155 Papers having more than 100 citations (最大値=1)

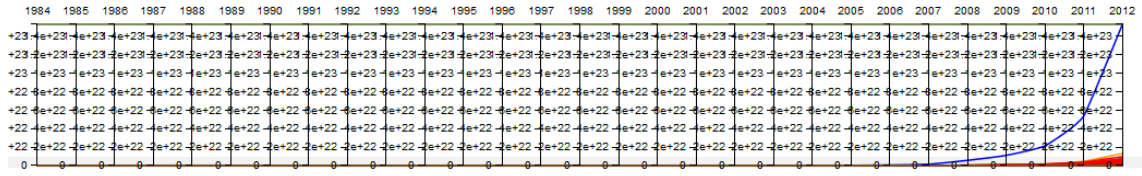


Figure 130 : Time Transition about within-Dataset Communicability Centralities of 155 Papers having more than 100 citations (最大値= 1.529e+23)

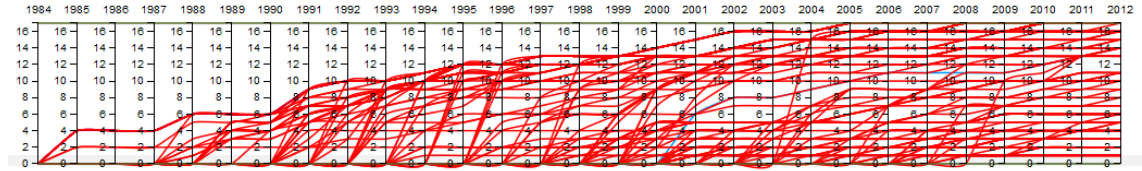


Figure 131 : Time Transition about within-Dataset Core Number of 155 Papers having more than 100 citations (最大値= 17)

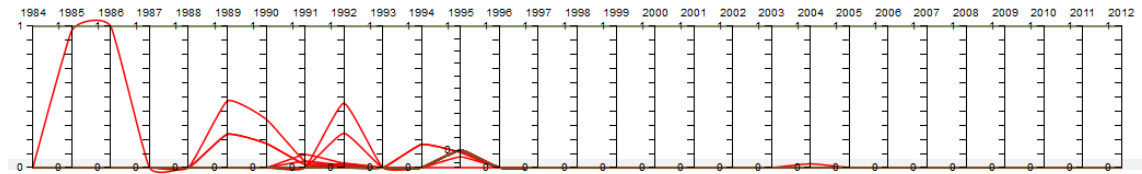


Figure 132 : Time Transition about within-Dataset Eigenvector of 155 Papers having more than 100 citations (最大値= 1)

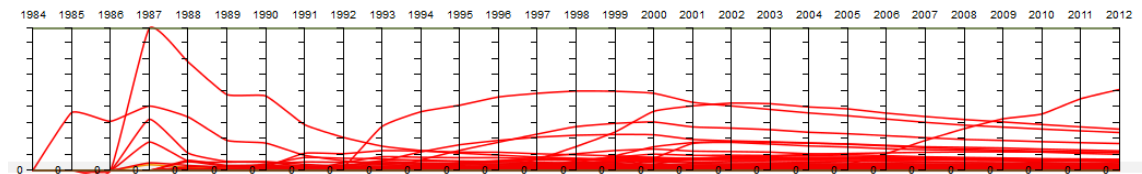


Figure 133 : Time Transition about within-Dataset Load Centralities of 155 Papers having more than 100 citations (最大値= 0.008890)

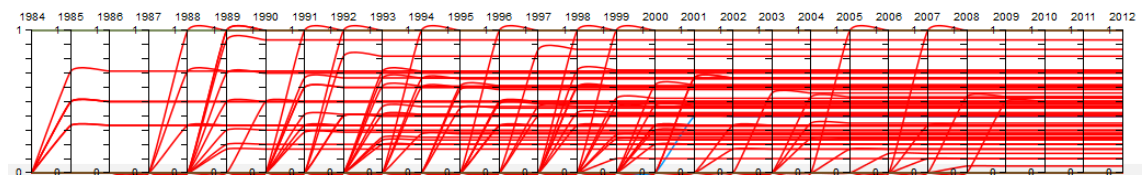


Figure 134 : Time Transition about within-Dataset Node Redundancy of 155 Papers having more than 100 citations (最大値= 1)

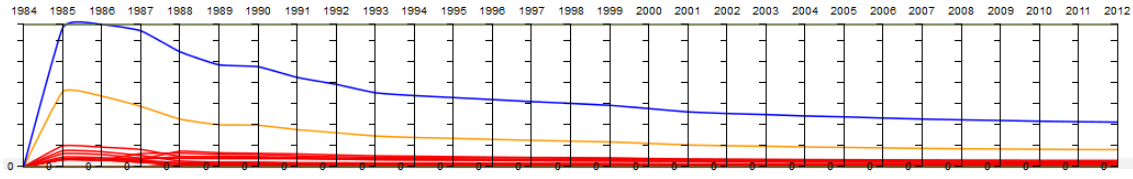


Figure 135 : Time Transition about within-Dataset PageRank of 155 Papers having more than 100 citations (最大值 = 0.3383)

付録 C: 学術分野の選定

萌芽領域を特定するために利用できるデータ全てを使用すると、現在では、情報量がすでに膨大になっているので、余程潤沢なリソースがないと合理的な時間内で計算が完了しない。すなわち、情報量を絞ったデータセットで計算する必要があるが、一般的に、データセットを決める方法は、以下の収集方法が見られる。

- a) 分野を代表するキーワードにより収集する
- b) 特定のジャーナルに掲載されている論文を対象にする
- c) 全領域のうち被引用数が上位 TOP 数% (1%、または、0.1%) のみの論文を分析対象にする
- d) 論文データベースからランダムに選択する

以上の他に、Web of Science 上の「Web of Science の分野」や「研究分野」で分析対象を絞り込む方法も考えられるが、この方法も分野によっては 1,000,000 を越える論文が所属するために、分析を行うには論文数が大きくなり過ぎる傾向にあり、現実的ではない。(c) は、文部科学省科学技術政策研究所のサイエンスマップ (坂ほか, 2008) の他にも複数で行われている方法である。

ここでは、重要論文の抽出に先立って、全学術分野から人気トピックを俯瞰的に抽出した。Web of Science には、2013-06-26 時点で、約 5,000 万件の論文が登録されており、これらの論文の最新被引用数トップ 0.1% では 50,459 件の論文書誌情報を得ることができる。このデータセットから、クラスター分析によるトピック抽出、および、キーワードによる人気トピック抽出を試みた。

C.1. クラスター分析によるトピックの抽出

最新被引用数トップ 0.1% の論文から引用ネットワークを形成したところ、49,026 件の論文で最大連結成分の引用ネットワークが得られた。最大連結成分の引用ネットワークをクラスター分析したところ、134 のクラスターに分かれた。最大連結成分に含まれなかった 1,433 論文で再度、引用ネットワークを形成したところ、第 2 に大きい連結成分の引用ネットワークは 4 論文と少

ないので、第 2 に大きい連結成分は考慮しない。最大連結成分では、上位 3 クラスターが 1 万を超える大きさであったため、再度、クラスター分析を行ってサブクラスターに分解した。

Table 42 : Clusters of TOP 0.1% Papers

#	Ave.PY*	Nodes	Edges	E/N	WCC*	Keywords	Topic
All	1987.4	49,026	395,431	8.066			
C1-1	1984.9	4,588	31,759	6.922	95	cortex, brain, receptor, neuron, memory	脳・神経
C1-2	1987.2	4,181	29,431	7.039	207	coronary, risk, patient, disease, myocardial	成人病
C1-3	1994.6	3,364	20,069	5.966	139	gene, protein, sequence, expression, rna	RNA
C1-4	1983.4	1,990	7,638	3.838	56	channel, firm, organizational, receptor, social	社会行動
C2-1	1983.4	3,770	24,278	6.440	276	galaxy, star, quantum, graphene, mass	宇宙・素粒子
C2-2	1982.2	3,358	20,931	6.233	149	protein, structure, energy, molecular, calculation	タンパク質
C2-3	1991.0	3,244	21,835	6.731	107	nanotube, polymer, carbon nanotube, nanoparticles, surface	ナノ・有機材料
C3-1	1991.8	4,722	44,851	9.498	213	cell, receptor, tumor, antigen, factor	免疫
C3-2	1992.9	3,284	42,170	12.841	133	kinase, cell, protein, apoptosis, receptor	生化学
C3-3	1990.3	3,255	27,248	8.371	227	gene, cell, p53, dna, protein	遺伝子
C4	1983.0	3,566	11,284	3.164	47	patient, cancer, percent, survival, hepatitis	細胞増殖
C5	1984.4	2,248	7,828	3.482	57	climate, ecosystem, ocean, specie, carbon	地球環境

*ヘッダーの省略について

Ave.PY: Average year that papers are published (平均出版年)

WCC : 1st of Within-cluster Citations

C.2. 頻出キーワードによるトピックの抽出

トップ 0.1%論文を利用して、著者キーワードとキーワード・プラスを集計し、キーワード単体の出現頻度、キーワード 2 語の共起頻度、前者 2 者を用いた Dice 係数を算出し、Table 43、Table 44、そして、Table 45 に上位を示す。キーワードの出現頻度と共起頻度とに関して単純に集計すると、医学関係の論文が多いという一般的な状況が反映される。そこで、キーワードの出現頻度と共起頻度を使って分野による偏向を可能な限り取り除こうというのが Dice 係数となる。また、キーワード 2 語の共起頻度を用いているのは、異なる分野でも同じ語が使用される場合が

ある（例えば、エントロピーは、熱力学と情報理論の両分野で使われる。熱力学で使用されていた語が、情報理論分野でも使われるようになったものである）が、2語以上の共起を利用すると、同じ語で異なる分野の場合に別のトピックとして取り出せる可能性が高くなるためである。また、ノーベル賞に選考される分野は評価が定まった分野であるため、評価が安定すると使用されるキーワードも安定してくるため、分野と特徴づけるキーワードの共起頻度も高まると考えられるためである。Dice 係数は、以下の式で求められる。

$$\text{Dice 係数} = \frac{2 \times \text{キーワード}A, B \text{の共起頻度}}{\text{キーワード}A \text{の出現頻度} + \text{キーワード}B \text{の出現頻度}}$$

Table 43: キーワードの出現頻度

順位	出現頻度	スコア
1	expression	1,775
2	gene	925
3	cells	825
4	identification	820
5	protein	810
6	activation	785
7	in-vivo	513
8	gene-expression	484
9	mice	466
10	saccharomyces-cerevisiae	457
11	tumor-necrosis-factor	445
12	disease	438
13	receptor	431
14	binding	418
15	dna	414
16	messenger-rna	402
17	growth	397
18	model	394
19	escherichia-coli	393
20	sequence	361
21	apoptosis	355
22	mortality	354
23	molecular-cloning	348
24	differentiation	347
25	cancer	345
25	therapy	345
27	phosphorylation	341
28	proteins	322
29	induction	319
30	nf-kappa-b	318

Table 44: キーワードの共起頻度

順位	共起頻度	スコア
1	expression; gene	287
2	cells; expression	200
3	expression; protein	172
4	expression; identification	163
5	activation; expression	147
6	expression; mice	121
7	cloning; expression	119
8	gene; identification	114
9	expression; receptor	107
10	differentiation; expression	105
11	gene; protein	101
12	expression; molecular-cloning	92
13	cells; gene	91
14	expression; messenger-rna	90
15	expression; mouse	84
16	factor-alpha; tumor-necrosis-factor	82
16	identification; protein	82
18	activation; protein	77
19	activation; phosphorylation	76
20	gene; mice	75
21	c-elegans; caenorhabditis-elegans	73
22	saccharomyces-cerevisiae; yeast	72
22	binding; expression	72
22	expression; sequence	72
25	expression; induction	71
26	expression; in-vivo	70
27	activation; gene	69
28	cells; protein	68
29	cells; identification	65
30	apoptosis; expression	64

Table 45: 共起するキーワードの Dice 係数

順位	共起するキーワード	Dice
1	non-a; non-b-hepatitis	0.6429
2	sv40-transformed cells; t-antigen	0.5714
3	galaxies : evolution; galaxies : stellar content	0.5581
4	deep-vein thrombosis; low-dose heparin	0.5405
5	galaxies : evolution; galaxies : formation	0.5185
6	iv collagenase; tissue inhibitor	0.5106
7	facial expressions; human amygdala	0.4898
8	sv40-transformed cells; transformed-cells	0.4889
9	mitogen; permeability factor	0.4590
10	enhanced raman-scattering; silver nanoparticles	0.4561
11	immunological self-tolerance; lymphocyte-associated antigen-4	0.4400
12	dark matter; galaxy formation	0.4348
13	chronic myelogenous leukemia; chronic myeloid-leukemia	0.4324
14	digital sky survey; galaxies : evolution	0.4138
15	bcr-abl; chronic myeloid-leukemia	0.4074
16	chronic myelogenous leukemia; philadelphia-chromosome	0.4063
17	c-elegans; caenorhabditis-elegans	0.4056
18	cosmology : observations; supernovae : general	0.4000
18	cosmological parameters; cosmology : observations	0.4000
18	catalogs; surveys	0.4000
21	cosmic microwave background; cosmology : observations	0.3929
22	mammalian target; p70 s6 kinase	0.3860
23	rat nucleus-accumbens; ventral tegmental area	0.3774
23	bax; bcl-x(l)	0.3774
25	enhanced raman-scattering; surface-plasmon resonance	0.3768
26	cytochrome-c release; permeability transition pore	0.3750
27	l-arginine; relaxing factor	0.3723
28	non-b-hepatitis; posttransfusion hepatitis	0.3448
28	dopamine; nucleus accumbens	0.3448
30	neurofibrillary tangles; paired helical filaments	0.3333
30	ischemia; reperfusion	0.3333
30	ataxia-telangiectasia; ionizing-radiation	0.3333

順位	共起するキーワード	Dice
33	reliability; validity	0.3188
34	experimental allergic encephalomyelitis; myelin basic-protein	0.3171
35	bcr-abl; chronic myelogenous leukemia	0.3143
35	bax; cytochrome-c release	0.3143
37	gene-related peptide; rat spinal-cord	0.3125
38	double-blind; placebo-controlled trial	0.3071
39	chemical-vapor-deposition; molecular-beam epitaxy	0.3059
40	field-effect transistors; thin-film transistors	0.3014
40	efficacy; safety	0.3014
42	neurofibrillary tangles; senile plaques	0.2899
42	neurofibrillary tangles; senile plaques	0.2899
44	estrogen-receptor; thyroid-hormone receptor	0.2892
45	human fibroblasts; senescence	0.2857
46	c-elegans; double-stranded-rna	0.2823
47	factor-alpha; tumor-necrosis-factor	0.2803
48	macaque monkey; rhesus-monkey	0.2791
49	macaque monkey; superior temporal sulcus	0.2785
50	body-mass index; overweight	0.2759
51	human blastocysts; lines	0.2740
52	glycogen-synthase kinase-3; p70 s6 kinase	0.2703
53	drosophila toll; lipopolysaccharide	0.2619
54	gene amplification; wild-type p53	0.2569
55	molecular recognition; supramolecular chemistry	0.2564
55	dopamine; ventral tegmental area	0.2564
57	relaxing factor; smooth-muscle	0.2556
58	acute myeloid-leukemia; acute promyelocytic leukemia	0.2553
59	conjugated polymers; light-emitting-diodes	0.2549
60	dna-damage; p53-dependent apoptosis	0.2532