

Title	情報科学の国際学会セッションの時系列ネットワーク分析
Author(s)	森, 薫; 有野, 和真; 古川, 貴雄
Citation	年次学術大会講演要旨集, 27: 21-24
Issue Date	2012-10-27
Type	Conference Paper
Text version	publisher
URL	<a href="http://hdl.handle.net/10119/10966">http://hdl.handle.net/10119/10966</a>
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

## 情報科学の国際学会セッションの時系列ネットワーク分析

○森 薫(慶應義塾大学 / 文部科学省 科学技術政策研究所)  
有野 和真 古川 貴雄(文部科学省 科学技術政策研究所)

## 1. はじめに

革新的技術 (Emerging Technology)は、既存産業を変革し、新たな産業を創出する技術と定義され<sup>1)</sup>、既存技術の改善とは根本的に異なる。これまでに、革新的技術に関する事例分析<sup>2)</sup>は行われているが、定量的な分析手法は未開拓である。

類似した論文が発表される国際学会のセッションは、研究者コミュニティが創出した知識の一つと考えられ、最新の研究動向を示す概念がセッション名に反映されている。本稿では、研究動向の変化が大きと思われる情報科学分野について、国際会議のセッションの変遷と主要な研究組織について分析し、革新的技術の発展過程とその要因を検討する。

## 2. セッションネットワークの時系列分析

## 2.1 分析対象

WWW やデータベースを基盤とする技術は、様々な流通サービスだけでなく、個人のコミュニケーションまで変化させている。本稿では、WWWとデータベースの研究において主要な国際学会とされる WWW、VLDB、SIGMOD、ICDE を選び、2002～2011年のセッションと論文を発表した組織の推移を分析した。なお、Research sessionとIndustrial sessionを対象とし、Poster sessionやDemo sessionは含めていない。

- (1) WWW: International World Wide Web Conference
- (2) ICDE: IEEE International Conference on Data Engineering
- (3) VLDB: International Conference on Very Large Data Base
- (4) SIGMOD: ACM SIGMOD (Special Interest Group on Management of Data) Conference

これら4つの国際学会で2002～2011年に発表された論文について、組織別に著者数を求めた結果を表1に示す。ここでは、論文別にカウントとしたのべ著者数を著者数とした。

## 2.2 セッションネットワークの生成方法

国際学会のセッションネットワークは以下の手順<sup>3)</sup>で生成する。

- (1) 各論文のアブストラクトから単語を抽出し、出現頻度に基づく tfidf を求める。抽出された単語の tfidf の値を要素とするベクトルを用いて各論文を数値的に表す。
- (2) 論文の類似度はベクトルのなす角を用いて計算

する。セッションに含まれる論文間の類似度の平均値をセッション間の類似度とする。

- (3) 基点となるセッションを選択し、設定した閾値よりもセッション間の類似度が大きくなるセッションを順次接続してネットワークを形成する。

表1 組織別の著者数(論文別にカウントしたのべ著者数を示す)

(a) WWW		(b) VLDB	
組織	著者数	組織	著者数
1 Microsoft	305	1 IBM	301
2 Yahoo!	287	2 Microsoft	143
3 IBM	268	3 Oracle	135
4 Google	84	4 Univ. of Wisconsin-Madison	117
5 Stanford Univ.	73	5 AT&T	96
6 Univ. Illinois at Urbana-Champaign	55	6 National Univ. of Singapore	84
7 Carnegie Mellon Univ.	52	7 Univ. Illinois at Urbana-Champaign	84
8 Univ. of Washington	51	8 Stanford Univ.	74
9 AT&T	48	9 Hong Kong Univ. of Sci. and Tech	64
10 Hewlett-Packard	48	10 Univ. of Toronto	57
11 Cornell Univ.	46	11 ETH Zurich	56
12 Univ. of California, Santa Barbara	43	12 Purdue Univ.	53
13 Shanghai Jiao Tung University	38	13 Chinese Univ. of Hong Kong	51
14 Georgia Inst. of Tech.	36	14 Univ. of Michigan	51
15 NEC	34	15 Carnegie Mellon Univ.	50

(c) SIGMOD		(d) ICDE	
組織	著者数	組織	著者数
1 IBM	287	1 IBM	257
2 Microsoft	195	2 Microsoft	148
3 National Univ. of Singapore	114	3 National Univ. of Singapore	131
4 Oracle	86	4 Purdue University	78
5 Univ. of Wisconsin-Madison	73	5 Hong Kong Univ. of Sci. and Tech	76
6 Hong Kong Univ. of Sci. and Tech	72	6 Univ. of Wisconsin-Madison	70
7 AT&T	62	7 Univ. Illinois at Urbana-Champaign	69
8 Univ. Illinois at Urbana-Champaign	56	8 Oracle	63
9 Stanford Univ.	51	9 AT&T	59
10 Cornell University	50	10 Univ. of New South Wales	52
11 Univ. of California, Berkeley	40	11 Univ. of Toronto	52
12 Univ. of Washington	40	12 Chinese Univ. of Hong Kong	49
13 Univ. of Toronto	39	13 Carnegie Mellon Univ.	44
14 Univ. of California, Santa Barbara	37	14 Univ. of California, Berkeley	41
15 Purdue Univ.	33	15 Univ. of California, Santa Barbara	40

## 3. 結果

2002～2011年に開催された4つの国際学会から生成したセッションネットワークの例を図1～4に示す。矢印の太さはセッション間の類似度を反映している。

## 3.1 WWWにみるソーシャルネットワーク研究の発展

国際学会 WWW 2011 の Social Network Analysis というセッションに至るネットワーク(図1)には多くのセッションと接続される特徴的なセッションが存在する。過去のセッションとの接続が多い場合、従来の研究を集約したセッションとみなすことができる。2008年の Social Networks: Discovery and Evolution of Communities が典型例で、2007年の Web Graphs、E-Communities などの研究を集約したセッションと考えられる。反対に、多くのセッションに分岐し、将来の研究に影響を与えたと考えられるセッションも存在する。2003年の Using the Semantic Web や 2005年の Link-based Similarity がこれに相当する。

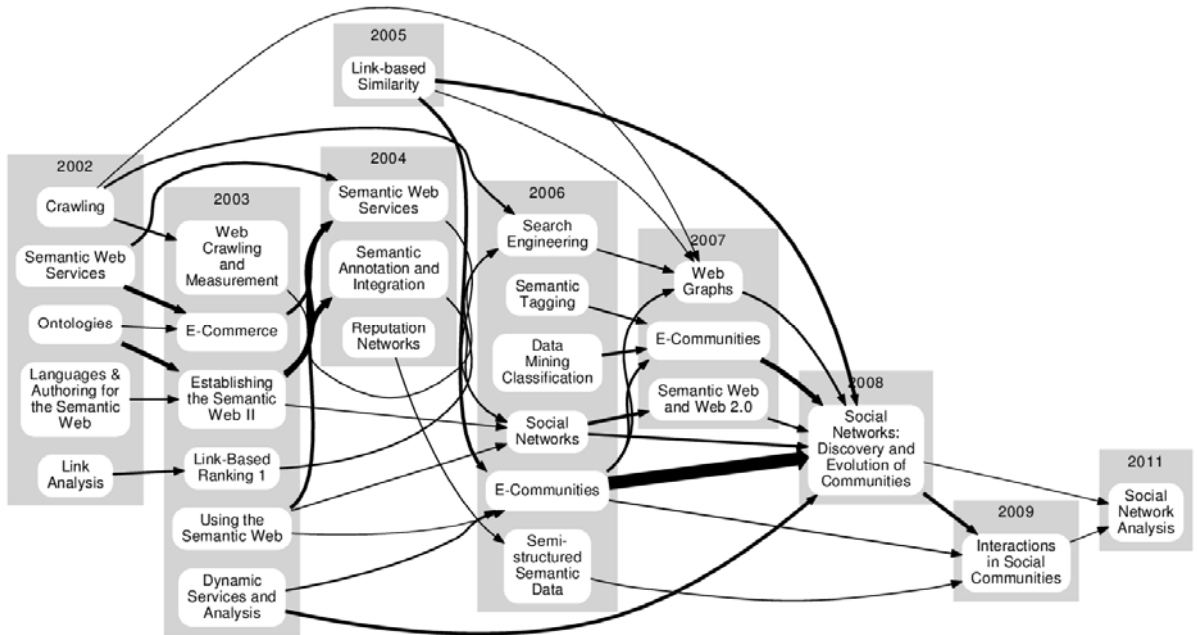


図 1 WWW 2011 の Social Network Analysis に至る時系列セッションネットワーク (矢印の太さにセッション間の類似度を反映している)

2006 年の E-Communities と 2008 年の Social Networks: Discovery and Evolution of Communities を接続する矢印が最も太く、これらのセッションが類似していることを示す。これは、2006 年に E-Communities とされていた研究が、他の研究を合わせて 2008 年に Social Networks の研究に変化したことを意味する。この変化は、その後の Social Networks に関する研究を推進する要因となったと考えられる。

### 3.2 ICDE にみるクラウド研究の発展

国際学会 ICDE 2011 の Cloud Computing というセッションに至るネットワーク(図 2)は、クラウドコンピューティング研究が、2002~2004 年の Heterogeneous, Distributed Database, Distributed, Parallel, Mobile Database から 2010 年の Distributed Data を経由して発展したことを示す。さらに、2007 年の Information Management に接続される 2002~2004 年のセッションには、Semi-structured Data が含まれていることから、これらの研究もクラウドコンピューティング研究の発展に寄与したものと考えられる。また、2006 年と 2008~2009 年のセッションが含まれていないことから、2000 年代後半の変化は連続的ではなかったと推察される。

### 3.3 VLDB にみるクラウド基盤技術の発展

MapReduce は大規模分散環境で実行される新しい並列分散処理方式であり、Hadoop は Google のデータ処理の基盤環境をオープンソースで実装したものである。

Hadoop 上で MapReduce 処理が実行されることから、クラウドコンピューティングの基盤技術として注目を集めている。

国際学会 VLDB のセッションネットワーク(図 3)は、MapReduce and Hadoop が、2009 年の MapReduce、2010 年の Cloud Computing と Experimental Analysis and Performance から派生したことを示している。さらに、2005 年の Data Mining や 2007 年の Indexing and Search から派生した研究とみなすこともできる。2002~2005 年のネットワークは、XML を中心としたデータ処理研究が密接に関連した構造となっている。この期間に、XML に関する類似研究が多くセッションで発表されたことを示している。

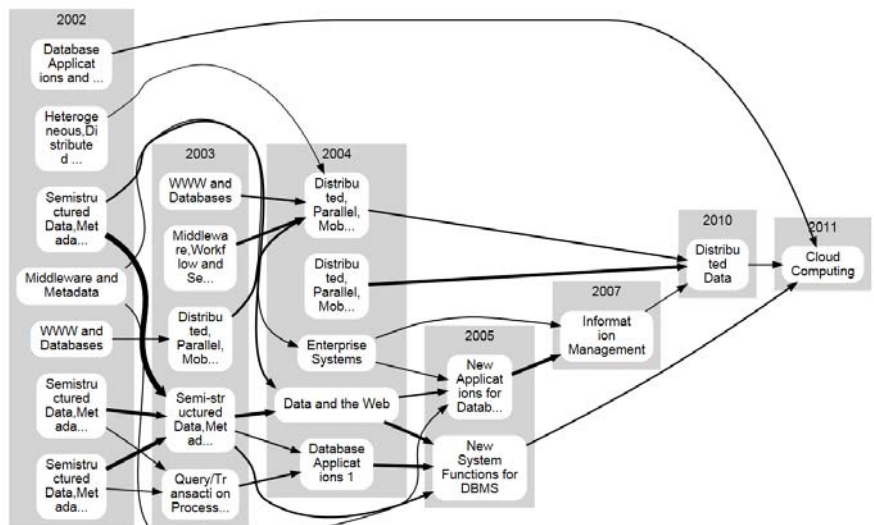


図 2 ICDE 2011 の Cloud Computing に至る時系列セッションネットワーク (矢印の太さにセッション間の類似度を反映している)

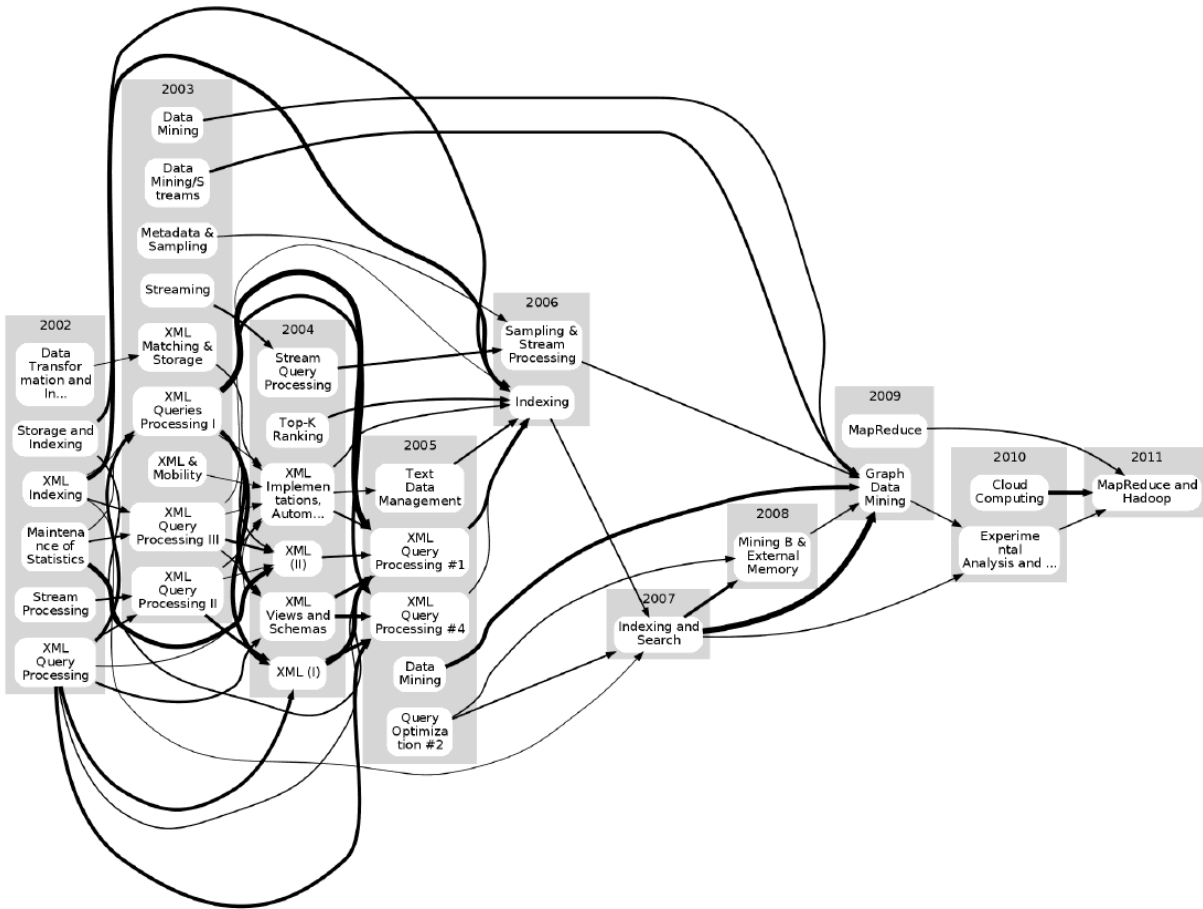


図3 VLDB 2011 の MapReduce and Hadoop に至る時系列セッションネットワーク (矢印の太さはセッション間の類似度の高さを反映している)

### 3.4 SIGMOD にみるサービス指向研究の発展

国際学会 SIGMOD 2011 の Service Oriented Computing, Data Management in the Cloud というセッションに至るネットワークを図4に示す。この研究は2010年の Cloud Computing & Internet Scale Computing から派生している。セッション名に Cloud が含まれることから、クラウドコンピューティングとサービスに関する研究が密接に関係することがわかる。また2004年の Security and Privacy、2009年の Transactions, Security, and Caching から派生していることから、セキュリティがサービスを支える技術要素として重視されている。2010年の Cloud computing & Internet Scale Computing は、2002年の Commercial Implementation Techniques と2004年の Index and Tuning から派生している。2002年頃に商業システムの技術としてインデックスやチューニングが議論され、その後、クラウドコンピューティングやセキュリティを含めて、サービス志向技術に発展したと考えられる。

### 3.5 セッションネットワークに見る研究の変化

図1~3をみると、2005年から2008年の間にセッション間を接続するネットワークの構造が変化している。これは、インターネットやデータベースを基盤とする

技術が大きく変化したことを示している。2000年代前半は、Web コンテンツに意味情報を付加するセマンティックウェブ技術に関する研究や、XML データベースに関する研究が多く発表されていたが、その一方で研究の新規性が薄れ、次第に研究が停滞していった期間とみなすこともできる。2005年頃に、MapReduce や Hadoop といったクラウドコンピューティングを支える要素技術が登場し、その後、研究者コミ

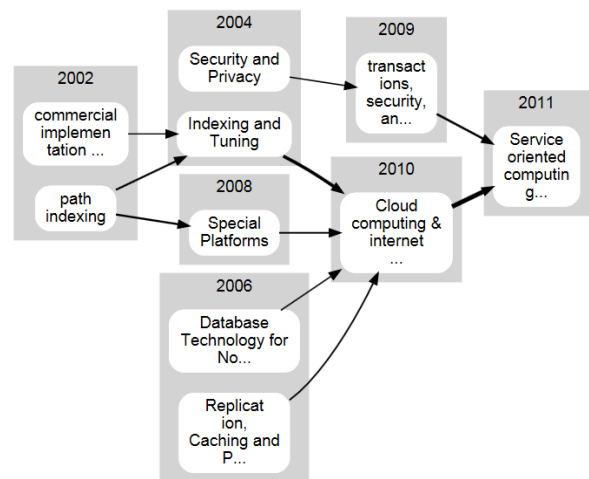


図4 SIGMOD 2011 の Service Oriented Computing, Data Management in the Cloud に至る時系列セッションネットワーク (矢印の太さにセッション間の類似度を反映している)

ユニティで普及し、WWW やデータベースなどの研究の方向性を大きく変える革新的技術となったと考えられる。

#### 4. 革新的技術を牽引する組織

2000 年代前半は停滞していた WWW とデータベース研究に、クラウドコンピューティングは 2005 年以降の研究に大きな変化をもたらした。このように研究が大きく変化した過程で、主要研究組織に起きた変化を著者数の推移から検討する。組織別の著者数(表 1)をみると、大学よりも企業が主要な研究組織となっているため、IBM、Microsoft、Yahoo!、Google、Facebook を選び、各国際学会の著者数の変化を求めた。ここでは、これら企業の特徴が明確な WWW と VLDB の著者数変化を図 5 に示す。

##### 4.1 著者数の減少した IBM

IBM の場合、WWW の著者数は大きく減少したことから、WWW に関する研究に注力しなくなったと言える。一方、VLDB の著者はやや減少傾向を示しているが、WWW の場合ほど顕著ではない。データベース研究に関しては、クラウドコンピューティングによって 2005 年以降に大きく変化したが、IBM はその変化に適応してきたと考えられる。

##### 4.2 WWW の著者数が増加した Microsoft

Microsoft の場合、WWW の著者数が 2005 年以降に大きく増加している。このことから、Microsoft は WWW に関する研究開発に注力するようになったと考えられる。一方、VLDB の著者数には大きな変化が見られない。

##### 4.3 WWW の著者数が急激に増加した Yahoo!

Yahoo! の場合、2005 年以降に WWW の著者数が急激に増え、2008 年には IBM、Microsoft よりも著者数が多くなっている。VLDB の著者数も 2005 年以降に増えているが、2008 年をピークにその後は変動している。Yahoo! はクラウドコンピューティングの基盤技術である Hadoop の開発を牽引したことが、WWW と

VLDB の著者数増加につながったと考えられる。

#### 4.4 主要研究組織が交代した WWW と均衡する VLDB

WWW の著者数変化から、主要研究組織が IBM から Microsoft、Yahoo! に移行したことがわかる。Google も著者数が増える傾向にあり、今後は Facebook の著者数が増えることが予想される。WWW はクラウドコンピューティングの影響が大きな研究であるため、主要研究組織が交代が起きたと考えられる。VLDB の著者数変化を見ると、Yahoo!、Google が台頭し、IBM、Microsoft を含めて均衡している状態になっている。この中でも、著者数の増えている Yahoo! と Google はデータベース研究に注力していることが読み取れる。

#### 5. まとめ

本稿では、情報科学の国際学会におけるセッションの変遷を分析することによって、2005 年頃にクラウドコンピューティングを支える技術が登場し、その後の研究を大きく変化させる要因となったことを示した。さらに、論文著者の所属組織に関する分析から、研究の変遷と主要研究組織との関係について述べた。

#### 謝辞

本研究は JSPS 科研費 22500875 の助成を受けたものである。

#### 文献

- 1) Day, G. S., Shoemaker P. J. H., (2000). A different game, in Day, G. S. and Shoemaker P. J. H. (Eds.), Wharton on Managing Emerging Technologies, John Wiley and Sons Inc., Hoboken, NJ, pp. 1-23.
- 2) Hesse-Biber, S. N. (Eds.) (2011), The Hand book of Emergent Technologies in Social Research, Oxford University Press, New York.
- 3) Arino, K., Furukawa, T., Shirakawa, N., Okuwada, K., Temporal Network Analysis of Emerging Technologies: Topic Transition in World Wide Web (WWW) Conferences, In Proc. IEEE International Conference on Industrial Engineering and Engineering Management (IEEM 2012). (to appear)

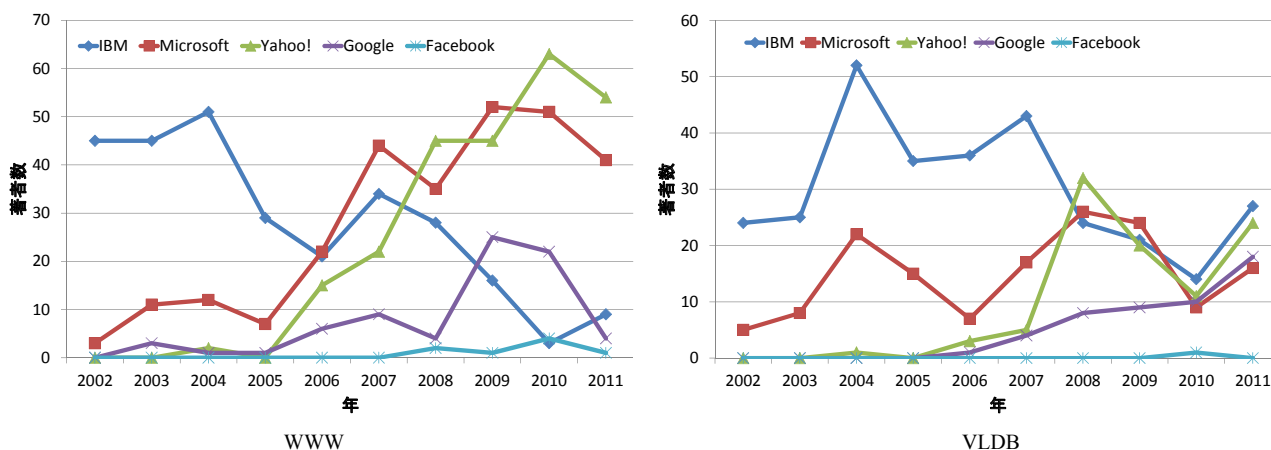


図 5 主要企業別の著者数変化