Technical Analysis Report
Codd's World: Topics and their Evolution in the Database Community Publication Graph

Rutuja Shivraj Pawar, Sepideh Sobhgol, Gabriel Campero Durand, Marcus Pinnecke, Jacob Krüger, David Broneske, and Gunter Saake

Otto von Guericke University Magdeburg Germany
rutuja.pawar@ovgu.de, sepideh.sobhgol@st.ovgu.de, campero@ovgu.de, pinnecke@ovgu.de, jacob.krueger@ovgu.de, david.broneske@ovgu.de, saake@ovgu.de

Abstract. This is a detailed Technical Analysis Report supplementing the original paper on Codd's World: Topics and their Evolution in the Database Community Publication Graph.

1 Topic Overview

At its core, our analysis is based on the global 30 topics extracted over all the years from the dataset. The 30 topics including their words and giving them a representative meaningful name are,

0. **Numerical Methods:** method proposed method proposed methods based new using method based results estimation
1. **Applications:** research social study design analysis knowledge technology use human online
2. **Networking:** network networks nodes routing wireless traffic sensor node protocol neural
3. **Optimization:** problem problems optimization solution optimal solutions linear set solve function
4. **Data Mining:** data mining data sets data mining sets clustering analysis large database query
5. **HCI:** performance memory parallel applications high architecture design hardware implementation processor
6. **Modeling And Simulation:** model models modeling parameters based process model based simulation proposed model markov
7. **Communication:** channel signal interference frequency noise channels rate performance multiple error
8. **Operating Systems:** system proposed system design based developed paper system performance using monitoring describes
9. **Cognitive Learning:** learning students machine machine learning training learn neural student knowledge supervised
10. **Semantic Web:** web search semantic query web services pages content services queries documents

11. **Algorithms:** algorithm algorithms proposed algorithm proposed search based algorithm based clustering new genetic
12. **Energy:** energy power consumption energy consumption sensor power consumption voltage low efficiency energy efficiency
13. **Logic Programming:** language logic languages object semantics programming knowledge programs program semantic
14. **Image Processing:** image images segmentation color 3d object visual resolution regions objects
15. **Cloud Computing:** service services cloud qos computing management quality business resource resources
16. **Cryptography:** scheme proposed schemes proposed scheme based coding propose signature simulation key
17. **Control Theory:** control robot controller robots motion feedback tracking stability loop nonlinear
18. **Network Analysis:** graph graphs vertices vertex number edge edges set connected tree
19. **Time Series:** time real real time scheduling time series series delay space temporal varying
20. **Software Engineering:** software development engineering process software development design requirements project tools hardware
21. **Machine Learning:** features classification feature recognition speech accuracy classifier training detection based
22. **Video Processing:** video motion quality coding frame videos frames content 3d temporal
23. **Decision Support:** fuzzy decision rules sets rule logic clustering set neural controller
24. **Testing:** test testing fault faults detection tests coverage circuit circuits generation
25. **Security:** security protocol attacks secure key attack authentication protocols privacy encryption
26. **Distributed Systems:** agent agents distributed multi communication complex information systems state based
27. **Block Coding And Decoding:** codes code error decoding coding source binary rate length block
28. **Information Retrieval:** information retrieval information systems knowledge sources context documents text document available
29. **GPS Navigation:** user users mobile devices interface interaction device mobile devices location access

2 Research Questions

The below research questions are answered through analysis on the network data:

- **RQ**$_1$: How did topics evolve in their popularity through time?
- **RQ**$_2$: Which are the most cited papers per topic per year? (*with and without self-citations)*

- **RQ**$_3$: Which is the most influential paper per topic per year? *(with and without self-citations)*
- **RQ**$_4$: How many citations per topic per year? *(with and without self-citations)*
- **RQ**$_5$: Who is the most important author per topic, looking at collaboration only, citation only, and mixed? *(with and without self-citations)*

2.1 Evolution of Topics Through Time

Relevance: Visualizing topic evolution depicts the popular research topics, measures topic change over time, merge or split of a topic, increase or decrease of importance for a topic and other topic evolutionary characteristics, thus helping to better understand the research trend in the database field.
Results and Discussion: Fig. 1 shows the evolution of the identified 30 topics over the years. It is seen from the figure that Topic 1 (named as Numerical methods) has seen a steady evolution over the years, with its highest evolution in the year 2018, whereas Topic 15 (named as Cloud Computing) had the highest evolution in the year 1965 and is not visible in the year 2018. Similarly, we can visualize other topics which have evolved or suddenly disappeared over the years. Additionally, Fig. 2 shows the non-overlapping 30 topic assignment in the form of clusters. This figure merely illustrates that each paper is assigned to a single topic.

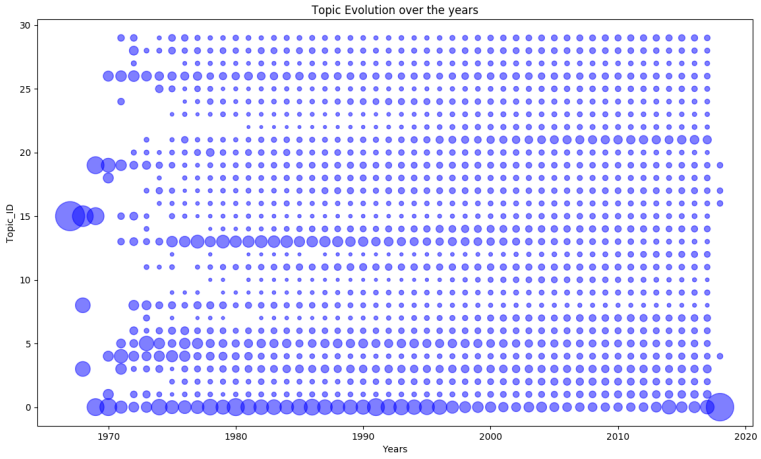

Fig. 1: Evolution of Topics Through Time

2.1.1 Top Citations per Topic per Year

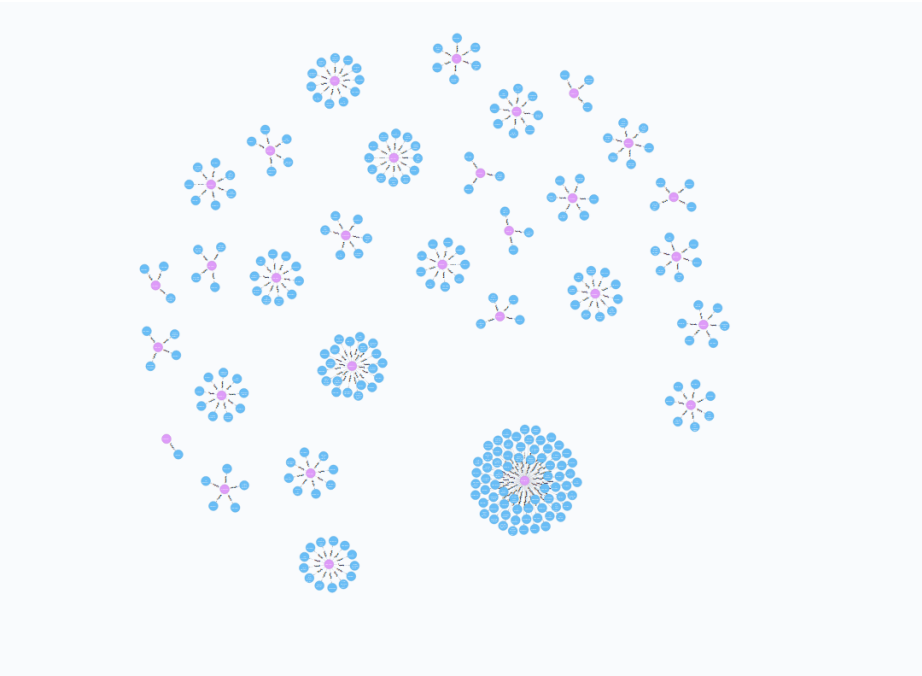

Fig. 2: Non-Overlapping Topic Assignments

Relevance: Understanding the most cited papers helps to measure the overall scientific impact made by the paper. Recognizing the most cited papers per topic per year facilitates deep analysis through measuring the trends in the scientific impact along the years.
Results and Discussion: The below Cypher query returns the most cited papers for a particular year with self-citation,

```
MATCH(p:TopicDescription)-
[:Topicality]->(s)<-[rel:CitationWithSC]-(r Year:"1970")
RETURN (s.Title), (p.TopicName), COUNT(rel)
ORDER BY COUNT(rel) DESC LIMIT 100;
```

Tables 1 and 2 summarize the query output for the years 1970 and 2017.

The below Cypher query returns the most cited papers for a particular year without self-citation,

```
MATCH(p:TopicDescription)-
[:Topicality]->(s)<-[rel:CitationWithoutSC]-(r Year:"1970")
RETURN (s.Title), (p.TopicName), COUNT(rel)
ORDER BY COUNT(rel) DESC LIMIT 100;
```

Tile	**TopicName**	**Count**
A Survey of Analytical Time-Sharing Models	NumericalMethods	3
A relational model of data for large shared data banks	DataMining	3
Optimizing the Performance of a Drum-Like Storage	TimeSeries	2
Principles of Optimal Page Replacement	Optimization	1

Table 1: Most cited papers in 1970 with self-citation

Title	TopicName	Count
ImageNet Classification with Deep Convolutional Neural Networks	Testing	736
Caffe: Convolutional Architecture for Fast Feature Embedding	CognitiveLearning	734
LIBSVM: A library for support vector machines	MachineLearning	585
Distinctive Image Features from Scale-Invariant Keypoints	MachineLearning	573
Very Deep Convolutional Networks for Large -Scale Image Recognition	MachineLearning	562
Random Forests	MachineLearning	540
Distributed Representations of Words and Phrases and their Compositionality	CognitiveLearning	490
Histograms of oriented gradients for human detection	MachineLearning	449
Image quality assessment: from error visibility to structural similarity	ImageProcessing	407
Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift	MachineLearning	406

Table 2: Most Cited Papers in 2017 with Self-Citation

Tables 3 and 4 summarize the query output for the years 1970 and 2017.

Title	TopicName	Count
A Survey of Analytical Time-Sharing Models	NumericalMethods	3
A relational model of data for large shared data banks	DataMining	3
Optimizing the Performance of a Drum-Like Storage	TimeSeries	2

Table 3: Most Cited Papers in 1970 without Self-Citation

Title	TopicName	Count
ImageNet Classification with Deep Convolutional Neural Networks	Testing	736
Caffe: Convolutional Architecture for Fast Feature Embedding	CognitiveLearning	734
LIBSVM: A library for support vector machines	MachineLearning	585
Distinctive Image Features from Scale-Invariant Keypoints	MachineLearning	573
Very Deep Convolutional Networks for Large-Scale Image Recognition	MachineLearning	562
Random Forests	MachineLearning	540
Distributed Representations of Words and Phrases and their Compositionality	CognitiveLearning	490
Histograms of oriented gradients for human detection	MachineLearning	449
Image quality assessment: from error visibility to structural similarity	ImageProcessing	407
Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift	MachineLearning	406

Table 4: Most Cited Papers in 2017 without Self-Citation

Comparing the tables for the years 1970 and 2017 for with and without self-citation, it is observed that majority of the returned papers with their topics are the same in both the queries. This suggests that the top papers returned do not achieve their most cited criteria through self-citation. Additionally Fig. 3 depicts the top 200 papers cited in the year 2017 with their topic names and without self-citation.

2.2 Top Influence per Topic per Year

Relevance: Measuring the most Influential paper based on its ranking in the network is an indicator of high acceptance of the research work by the scientific community. Understanding the top influential paper per topic per year helps to visualize the trend

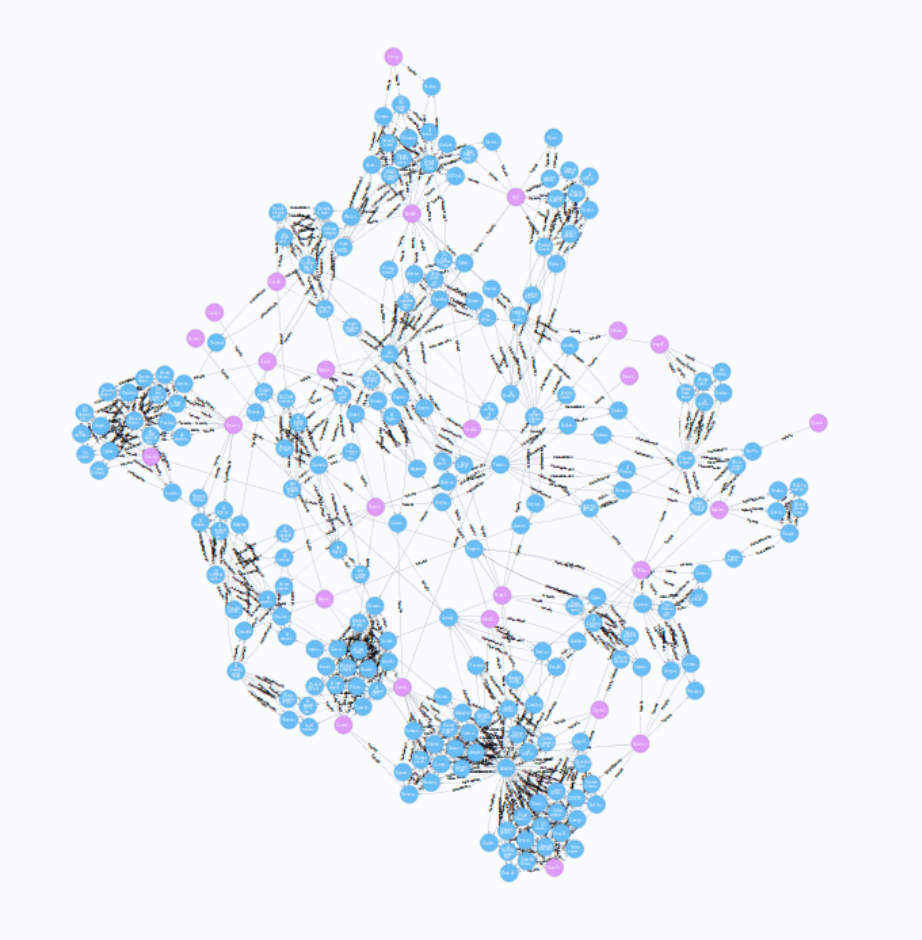

Fig. 3: Top 200 papers without Self-Citations in year 2017

of this acceptance over the years.
Results and Discussion: The below Cypher query returns the 25 most influential papers (based on Page Rank score) with self-citation for all the years,

```
MATCH(p:TopicDescription)-[:Topicality]->(s)
RETURN p.TopicName, s.Title, s.ScoreWithSC AS PR
ORDER BY PR DESC LIMIT 25;
```

Table 5 summarize the query output for all the years. The below Cypher query returns the 10 most influential papers (based on Page Rank score) with self-citation for a particular year,

```
MATCH(p:TopicDescription)-[:Topicality]->(s Year:"1970")
```

Title	TopicName	Score
A relational model of data for large shared data banks	DataMining	814.4239501953125
Induction of Decision Trees	DistributedSystems	722.2256469726562
Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference	DistributedSystems	544.8001098632812
Snakes: Active Contour Models	MachineLearning	530.5615234375
A theory for multiresolution signal decomposition: the wavelet representation	ImageProcessing	484.8881530761719
A training algorithm for optimal margin classifiers	MachineLearning	431.5468444824219
A robust layered control system for a mobile robot	ControlTheory	362.0459899902344
Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers	Networking	357.85980224609375
Support-Vector Networks	MachineLearning	357.1878967285156
A learning algorithm for boltzmann machines	HCI	347.3739929199219
The Concept of a Linguistic Variable and its Application to Approximate Reasoning	DecisionSupport	337.3362121582031
A simple transmit diversity technique for wireless communications	Security	334.5489501953125
MACAW: a media access protocol for wireless LAN's	Security	327.1644592285156
Independent component analysis, a new concept?	DataMining	323.7842712402344
Indexing by Latent Semantic Analysis	SemanticWeb	313.26165771484375
Compliance and Force Control for Computer Controlled Manipulators	ControlTheory	308.8047790527344
Computer architecture: a quantitative approach	HCI	308.58050537109375
Handwritten Digit Recognition with a Back-Propagation Network	Networking	306.81317138671875
A stochastic parts program and noun phrase parser for unrestricted text	MachineLearning	301.9481201171875
Distinctive Image Features from Scale-Invariant Keypoint	MachineLearning	301.7352294921875
LIBSVM: A library for support vector machines	MachineLearning	299.5965270996094
Analysis and simulation of a fair queueing algorithm	Algorithms	296.6663513183594
Supporting real-time applications in an Integrated Services Packet Network: architecture and mechanism	TimeSeries	295.6441345214844
Fast learning in networks of locally-tuned processing units	CognitiveLearning	283.77789306640625
What Size Net Gives Valid Generalization	Testing	281.984619140625

Table 5: 25 most Influential Papers (based on Page Rank score) with Self-Citation for all the years

```
RETURN p.TopicName, s.Title, s.ScoreWithSC AS PR
ORDER BY PR DESC LIMIT 10;
```

Tables 6 and 7 summarize the query output for the years 1970 and 2018.

Title	TopicName	Score
A relational model of data for large shared data banks	DataMining	814.4239501953125
Virtual memory	NumericalMethods	151.17889404296875
Toward an understanding of data structures	NetworkAnalysis	27.3781681060791
A schema for describing a relational data base	NumericalMethods	18.157360076904297
Introduction to storage structure definition	NumericalMethods	3.3184258937835693
Time-sharing for OS	TimeSeries	1.6499865055084229
TICKETRON: a successfully operating system without an operating system	DistributedSystems	0.2359350025653839
Swap-Time Considerations in Time-Shared Systems	TimeSeries	0.18187500536441803
A contiuum of time-sharing scheduling algorithms	Applications	0.15000000596046448

Table 6: 10 most Influential Papers (based on Page Rank score) with Self-Citation 1970

The below Cypher query returns the 25 most influential papers (based on Page Rank score) without self-citation for all the years,

```
MATCH(p:TopicDescription)-[:Topicality]->(s)
RETURN p.TopicName, s.Title, s.ScoreWithoutSC AS PR
ORDER BY PR DESC LIMIT 25;
```

Table 8 summarize the query output for all the years.
The below Cypher query returns the 10 most influential papers (based on Page Rank score) without self-citation for a particular year,

```
MATCH(p:TopicDescription)-[:Topicality]->(s Year:"1970")
RETURN p.TopicName, s.Title, s.ScoreWithoutSC AS PR
ORDER BY PR DESC LIMIT 10;
```

Tables 9 and 10 summarize the query output for the years 1970 and 2017. Observation of the tables, suggest that the highest Page Rank is indeed associated with the old papers but is not necessarily always true. As expected, the foundational paper of Edgar Codd on relational databases remains the most influential over all the years (with and without self-citation). The results of this research question cannot be compared with the results of RQ2 as, self-citation makes a difference on the network dynamics (given that Page Rank scores depend on the complete network structure) but not on the citation count of the most cited papers. Furthermore we observe that removing self-

Title	TopicName	Score
Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks	Networking	3.801413059234619
Random Graphs and Complex Networks	NumericalMethods	1.9076725244522095
Minimizing finite sums with the stochastic average gradient	Optimization	1.75485098361969
A Temporal Logic Approach to Binding-Time Analysis	LogicProgramming	1.6572284698486328
On the Linear Convergence of the Alternating Direction Method of Multipliers	Optimization	1.4891154766082764
Order-Optimal Rate of Caching and Coded Multicasting With Random Demands	Security	0.9819459915161133
SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation	HCI	0.7875764966011047
Inventory rebalancing and vehicle routing in bike sharing systems	Optimization	0.7597730159759521
A messy state of the union: taming the composite state machines of TLS	Security	0.6841909885406494
Salient Object Detection: A Discriminative Regional Feature Integration Approach	MachineLearning	0.6486610174179077

Table 7: 10 most Influential Papers (based on Page Rank score) with Self-Citation 2018

citations leads to a higher range for the scores of the most influential paper, showing that self-citation does indeed make a difference in the scoring. Additionally, Fig. 4 shows the top Influential papers with their topics depicting that most influential papers are cited across topics and have a high Page Rank, leading to the formation of big clusters. The small isolated clustered topics like Communication, Video Processing, Applications, Block Coding And Decoding are disconnected and indicate lower values for Page Rank.

2.3 Citations per Topic Through Time

Relevance: Measuring citation count for a topic helps to understand its research popularity among the scientific community. Analyzing citation count per topic per year helps to measure the relevant trends of research on a topic over the years.
Results and Discussion: The below Cypher query returns the citation count for a topic for all the years,

```
MATCH(p:TopicDescription TopicName:"MachineLearning")-
[:Topicality]->(s)<-[rel:CitationWithSC]-(r)
RETURN p.TopicName, r.Year, COUNT(rel)
AS CitationCount ORDER BY r.Year DESC LIMIT 100;
```

Title	TopicName	Score
A relational model of data for large shared data banks	DataMining	13669.4931640625
Jobshop-Like Queueing Systems	CloudComputing	5750.12548828125
A model and stack implementation of multiple environments	ControlTheory	5621.76611328125
Toward an understanding of data structures	NetworkAnalysis	5092.0205078125
Procedural embedding of knowledge in planner	Optimization	4842.31005859375
Optimizing the Performance of a Drum-Like Storage	TimeSeries	4267.1123046875
Virtual memory	NumericalMethods	3988.57568359375
New Programming Languages for Artificial Intelligence Research	NumericalMethods	3860.931884765625
Queues with State-Dependent Stochastic Service Rates	CloudComputing	3685.233642578125
Correctness-preserving program transformations	LogicProgramming	3680.9638671875
A universal modular ACTOR formalism for artificial intelligence	NumericalMethods	3617.47900390625
Multiple evaluators in an extensible programming system	LogicProgramming	3111.29833984375
Requirements for advanced programming systems for list processing	DistributedSystems	2911.9736328125
Uniqueness of the Gaussian Kernel for Scale-Space Filtering	Communication	2821.81005859375
Scale-space filtering: A new approach to multi-scale description	ImageProcessing	2751.470947265625
Relational Completeness of Data Base Sublanguages	NumericalMethods	2484.283935546875
A Survey of Data Structures for Computer Graphics Systems	DataMining	2473.419677734375
Interference detection among solids and surfaces	Communication	2460.12158203125
Forward Reasoning and Dependency-Directed Backtracking in a System for Computer-Aided Circuit Analysis	OperatingSystems	2418.098388671875
A total standard WIP estimation method for wafer fabrication	Algorithms	2403.2353515625
Higher order approximations for the single server queue with splitting, merging and feedback	DistributedSystems	2403.017578125
Symbolic reasoning among 3-d models and 2-d images	ImageProcessing	2240.699951171875
Abstract data types and software validation	LogicProgramming	2222.979248046875
Induction of Decision Trees	DistributedSystems	2193.14404296875
How to construct random functions	TimeSeries	2152.89111328125

Table 8: 25 most Influential Papers (based on Page Rank score) without Self-Citation for all the years

Title	TopicName	Score
A relational model of data for large shared data banks	DataMining	13669.4931640625
Toward an understanding of data structures	NetworkAnalysis	5092.0205078125
Virtual memory	NumericalMethods	3988.57568359375
A schema for describing a relational data base	NumericalMethods	264.00726318359375
Introduction to storage structure definition	NumericalMethods	18.38025665283203
TICKETRON: a successfully operating system without an operating system	DistributedSystems	12.225720405578613
Time-sharing for OS	TimeSeries	5.367920398712158
Swap-Time Considerations in Time-Shared Systems	TimeSeries	0.21375000476837158
A contiuum of time-sharing scheduling algorithms	Applications	0.15000000596046448

Table 9: 10 most Influential Papers (based on Page Rank score) without Self-Citation 1970

Title	TopicName	Count
Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks	Networking	3.972501039505005
Random Graphs and Complex Networks	NumericalMethods	2.2444255352020264
A Temporal Logic Approach to Binding-Time Analysis	LogicProgramming	2.216188430786133
Minimizing finite sums with the stochastic average gradient	Optimization	1.8653680086135864
On the Linear Convergence of the Alternating Direction Method of Multipliers	Optimization	1.761472463607788
Order-Optimal Rate of Caching and Coded Multicasting With Random Demands	Security	1.1023739576339722
Counting flags in triangle-free digraphs	NetworkAnalysis	0.8560609817504883
SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation	HCI	0.8167909979820251
Inventory rebalancing and vehicle routing in bike sharing systems	Optimization	0.812651515007019
A messy state of the union: taming the composite state machines of TLS	Security	0.6916624903678894

Table 10: 10 most Influential Papers (based on Page Rank score) without Self-Citation 2017

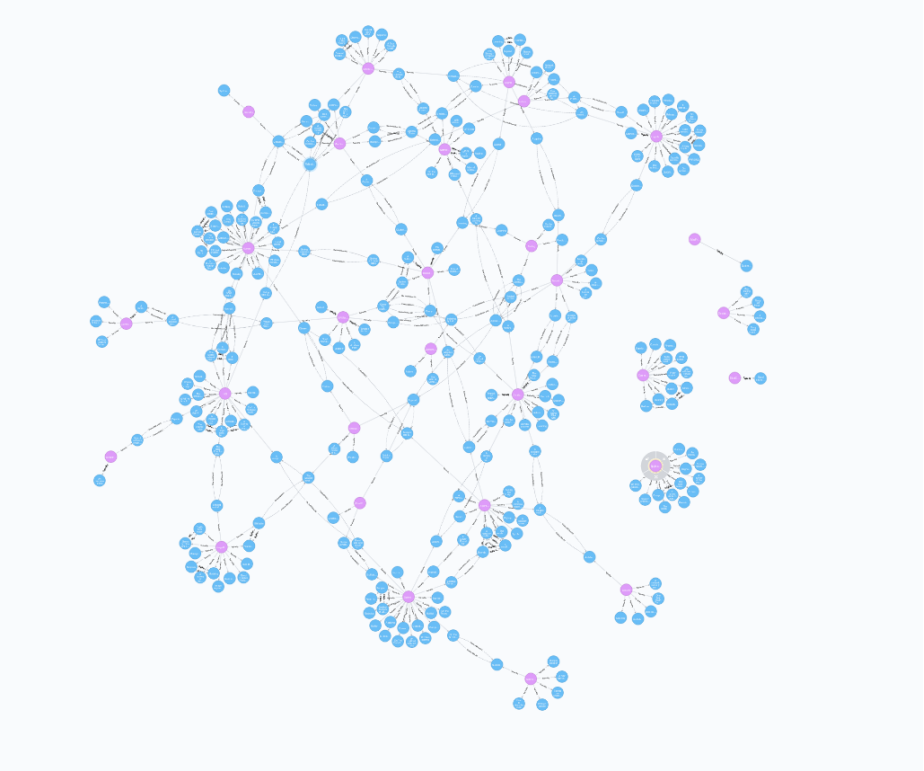

Fig. 4: Top Influence per Topic per Year

Tables 11 and 12 summarize the query output for topic Machine Learning and Data Mining.
Fig. 5 shows a histogram depicting the total citation count per year. Observing the tables indicate an increasing trend for the selected topics Machine Learning and Data Mining over the years. Additionally running the query for other topics identified no significant downtrend for any topic. This could be given the fact that we have not included information regarding distribution of the papers having high citation counts. Power Law analysis [1] can be used to solve this problem by drilling down into papers.

2.4 Top Influential Author per Topic

Relevance: Measuring the top influential author per topic combined and ranked over all the years, helps to understand the popular acceptance of the author's research on a particular topic among the scientific community. It is also an indicator of the valuable contribution made by the author towards the research topic.

TopicName	Year	CitationCount
MachineLearning	2018	110
MachineLearning	2017	100094
MachineLearning	2016	265503
MachineLearning	2015	234204
MachineLearning	2014	212706
MachineLearning	2013	177887
MachineLearning	2012	153578
MachineLearning	2011	130260
MachineLearning	2010	114068
MachineLearning	2009	98024
MachineLearning	2008	80878
MachineLearning	2007	68705
MachineLearning	2006	54832
MachineLearning	2005	42364
MachineLearning	2004	32709
MachineLearning	2003	22761
MachineLearning	2002	17970
MachineLearning	2001	13054
MachineLearning	2000	12189
MachineLearning	1999	8461
MachineLearning	1998	7545
MachineLearning	1997	6271
MachineLearning	1996	5040
MachineLearning	1995	3434
MachineLearning	1994	2508
MachineLearning	1993	1776
MachineLearning	1992	1511
MachineLearning	1991	1084
MachineLearning	1990	696
MachineLearning	1989	589
MachineLearning	1988	373
MachineLearning	1987	245
MachineLearning	1986	164
MachineLearning	1985	128
MachineLearning	1984	100
MachineLearning	1983	66
MachineLearning	1982	56
MachineLearning	1981	37
MachineLearning	1980	29
MachineLearning	1979	23
MachineLearning	1978	26
MachineLearning	1977	17
MachineLearning	1976	11
MachineLearning	1975	2

Table 11: Citation count for Topic Machine Learning over all the years

TopicName	Year	CitationCount
DataMining	2018	230
DataMining	2017	33505
DataMining	2016	91126
DataMining	2015	86158
DataMining	2014	81443
DataMining	2013	71882
DataMining	2012	62897
DataMining	2011	56531
DataMining	2010	49424
DataMining	2009	45940
DataMining	2008	38756
DataMining	2007	34282
DataMining	2006	28626
DataMining	2005	23338
DataMining	2004	17710
DataMining	2003	13269
DataMining	2002	9710
DataMining	2001	7245
DataMining	2000	5942
DataMining	1999	5062
DataMining	1998	3816
DataMining	1997	3139
DataMining	1996	2644
DataMining	1995	2138
DataMining	1994	1830
DataMining	1993	1693
DataMining	1992	1351
DataMining	1991	1199
DataMining	1990	1121
DataMining	1989	945
DataMining	1988	823
DataMining	1987	553
DataMining	1986	478
DataMining	1985	443
DataMining	1984	516
DataMining	1983	443
DataMining	1982	388
DataMining	1981	344
DataMining	1980	272
DataMining	1979	279
DataMining	1978	283
DataMining	1977	199
DataMining	1976	199
DataMining	1975	161
DataMining	1974	44
DataMining	1973	13
DataMining	1972	7
DataMining	1971	17
DataMining	1970	3

Table 12: Citation count for Topic Data Mining over all the years

counts per year.png

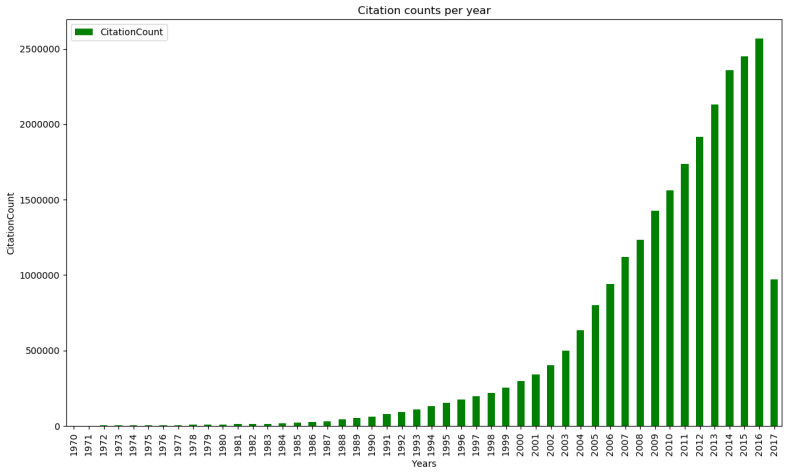

Fig. 5: Total citation count per year

Results and Discussion: The below Cypher query returns the 25 most influential authors through author rank on all topics involving Author Rank only on collaboration/co-authorship network,

```
MATCH(p:TopicDescription)-[:Topicality]->(s)<-[:Authorship]-(a)
RETURN DISTINCT a.AUTHOR_NAME, a.ARScore AS score
ORDER BY score DESC LIMIT 25;
```

Table 13 summarizes the query output for Author Rank on all topics indicating how spread is their collaboration with other authors.

The below Cypher query returns the 25 most influential authors through Page Rank on all topics on collaboration/co-authorship network with self-citation,

```
MATCH(p:TopicDescription)-[:Topicality]->(s)<-[:Authorship]-(a)
RETURN DISTINCT a.AUTHOR_NAME, a.PRwithSC AS score
ORDER BY score DESC LIMIT 25;
```

Table 14 summarizes the query output for Page Rank on all topics with self-citation.

The below Cypher query returns the 25 most influential authors through Page Rank on all topics on collaboration/co-authorship network without self-citation,

```
MATCH(p:TopicDescription)-[:Topicality]->(s)<-[:Authorship]-(a)
RETURN DISTINCT a.AUTHOR_NAME, a.PRwithoutSC AS score
ORDER BY score DESC LIMIT 25;
```

AuthorName	Score
Wei Wang	215.6251220703125
Wei Zhang	146.0390625
Wei Liu	143.38377380371094
Lei Wang	142.0850372314453
Yang Liu	137.28176879882812
Lei Zhang	128.57749938964844
Wei Chen	121.05441284179688
Jun Wang	117.1329345703125
Wei Liu	115.75605010986328
Xin Liu	108.65744018554688
Yan Zhang	106.84562683105469
Li Zhang	104.4130859375
Jun Zhang	104.29021453857422
Yang Yang	99.6176986694336
Jing Wang	98.92251586914062
Yu Zhang	97.22993469238281
Xin Wang	95.66221618652344
Li Li	94.99837493896484
Jing Li	92.20679473876953
Jie Zhang	90.3498764038086
Jun Li	89.5878677368164
Yu Wang	89.09492492675781
Hui Li	88.64311218261719
Yan Li	85.41546630859375
Yang Li	83.77407836914062

Table 13: Author Rank on all Topics (25 most Influential Authors)

AuthorName	Score
Scott Shenker	2323.373779296875
Demetri Terzopoulos	1693.306396484375
Robert L. Mercer	1608.9576416015625
Geoffrey E. Hinton	1563.3994140625
Hari Balakrishnan	1534.552978515625
Rakesh Agrawal	1505.221923828125
Vladimir Vapnik	1460.100830078125
Andrew P. Witkin	1459.3623046875
Deborah Estrin	1458.5408935546875
Lixia Zhang	1445.181884765625
Alex Pentland	1413.5933837890625
E. F. Codd	1409.331787109375
David E. Culler	1386.3238525390625
Anil K. Jain	1343.75537109375
David Haussler	1284.4566650390625
Robert E. Schapire	1277.0406494140625
Frederick Jelinek	1244.083740234375
Ian T. Foster	1220.2354736328125
Judea Pearl	1195.59033203125
Rodney A. Brooks	1189.1766357421875
Takeo Kanade	1180.5196533203125
Bernhard Schölkopf	1172.0855712890625
Sally Floyd	1149.4163818359375
Michael I. Jordan	1136.1651611328125
Michael Stonebraker	1126.252685546875

Table 14: Page Rank on all Topics with Self-Citation (25 most Influential Authors)

AuthorName	Score
E. F. Codd	18399.134765625
Daniel G. Bobrow	14275.8740234375
Carl Hewitt	12347.6787109375
Ben Wegbreit	9271.6328125
Andrew P. Witkin	7430.35205078125
Rakesh Agrawal	1505.221923828125
Vladimir Vapnik	1460.100830078125
Andrew P. Witkin	1459.3623046875
Peter J. Denning	7198.57421875
Robert Endre Tarjan	6088.55419921875
Peter Boehler Bishop	5915.26806640625
Richard Steiger	5915.26806640625
James R. Jackson	5750.12548828125
Jay Earley	5562.37841796875
H. T. Kung	4987.71875
Rodney A. Brooks	4939.78271484375
Geoffrey E. Hinton	4618.77734375
Joseph Abate	4354.34765625
Richard P. Brent	4338.82666015625
David R. Musser	4316.86474609375
Harvey Dubner	4279.337890625
Ellis Horowitz	4257.8330078125
Robert L. Mercer	4257.7353515625
Larry S. Davis	3981.80810546875
Bertram Raphael	3860.931884765625
Susan L. Gerhart	3806.7587890625
Oded Goldreich	3787.566650390625
David Haussler	3755.4345703125

Table 15: Page Rank on all topics without Self-Citation (25 most Influential Authors)

AuthorName	Score
Scott Shenker	2323.373779296875
Demetri Terzopoulos	1693.306396484375
Geoffrey E. Hinton	1563.3994140625
Hari Balakrishnan	1534.552978515625
Rakesh Agrawal	1505.221923828125
Vladimir Vapnik	1460.100830078125
Deborah Estrin	1458.5408935546875
Lixia Zhang	1445.181884765625
Alex Pentland	1413.5933837890625
E. F. Codd	1409.331787109375
David E. Culler	1386.3238525390625
Anil K. Jain	1343.75537109375
David Haussler	1284.4566650390625
Robert E. Schapire	1277.0406494140625
Ian T. Foster	1220.2354736328125
Judea Pearl	1195.59033203125
Takeo Kanade	1180.5196533203125
Bernhard Schölkopf	1172.0855712890625
Michael I. Jordan	1136.1651611328125
Jitendra Malik	1117.4991455078125
Alan J. Demers	1100.455322265625
Christos Faloutsos	1052.6434326171875
David R. Karger	1042.8453369140625
Robert Morris	1023.3004150390625

Table 16: Combination of Author Rank and Page Rank with self-citation for Topic Data Mining (25 most Influential Authors)

Table 15 summarizes the query output for Page Rank on all topics without self-citation.

The below Cypher query returns the 25 most influential authors through combination of Author Rank and Page Rank on collaboration/co-authorship network with self-citation for a particular topic,

```
MATCH(p:TopicDescription TopicName:"DataMining")- [:Topicality]->(s)<-[:Authorship]-(a)
RETURN DISTINCT a.AUTHOR_NAME, a.ARPRScorewithSC AS score
ORDER BY score DESC LIMIT 25;
```

Tables 16 summarize the query output for topic Data Mining with self-citation.

The below Cypher query returns the 25 most influential authors through combination of Author Rank and Page Rank on collaboration/co-authorship network with self-citation for a particular topic without self-citation,

```
MATCH(p:TopicDescription TopicName:"DataMining")-
[:Topicality]->(s)<-[:Authorship]-(a)
RETURN DISTINCT a.AUTHOR_NAME, a.ARPRScorewithoutSC AS score
ORDER BY score DESC LIMIT 25;
```

Tables 17 and 18 summarize the query output for topics Data Mining and Machine Learning without self-citation.

The below Cypher query returns the top most influential authors through combination of Author Rank and Page Rank on collaboration/co-authorship network without self-citation for all topics,

```
MATCH(p:TopicDescription)-[:Topicality]->(s)<-[:Authorship]-(a)
RETURN DISTINCT a.AUTHOR_NAME, a.ARPRScorewithoutSC AS score
ORDER BY score DESC LIMIT 100;
```

Table 19 summarizes the query output for 100 most influential authors without self-citation. Observing the tables, it is found that combining Author Rank and Page Rank did not lead to different results in the ranking of authors in comparison to Page Rank alone mostly because the collaboration network had very small weights. Further, it is found that most collaborative authors publish papers in all the topics. The limitation of the dataset should be also noted that some authors may have same names which may result in vagueness of results for ranking authors. Additionally, Fig. 6 shows the collaboration/co-authorship network of Prof. Gunter Saake.

We observe that the ranking of authors on the citation network is more informative than the ranking on the collaboration network.

AuthorName	Score
E. F. Codd	18399.134765625
Daniel G. Bobrow	14275.8740234375
Carl Hewitt	12347.6787109375
Ben Wegbreit	9271.6328125
Peter J. Denning	7198.57421875
Robert Endre Tarjan	6088.55419921875
Peter Boehler Bishop	5915.26806640625
Richard Steiger	5915.26806640625
H. T. Kung	4987.71875
Geoffrey E. Hinton	4618.77734375
David R. Musser	4316.86474609375
Ellis Horowitz	4257.8330078125
Larry S. Davis	3981.80810546875
David Haussler	3755.4345703125
Michael Stonebraker	3747.854248046875
Don Chamberlin	3745.54296875
Linda G. Shapiro	3681.767333984375
Silvio Micali	3540.778076171875
Jim Gray	3488.064453125
Raymond A. Lorie	3453.755615234375
Zohar Manna	3450.918701171875
Terrence J. Sejnowski	3410.999755859375
Demetri Terzopoulos	3395.916748046875
Scott Shenker	3378.406494140625
Azriel Rosenfeld	3370.689697265625

Table 17: Combination of Author Rank and Page Rank without self-citation for Topic Data Mining (25 most Influential Authors)

AuthorName	Score
Andrew P. Witkin	7430.35205078125
H. T. Kung	4987.71875
Geoffrey E. Hinton	4618.77734375
Richard P. Brent	4338.82666015625
Ellis Horowitz	4257.8330078125
Robert L. Mercer	4257.7353515625
Larry S. Davis	3981.80810546875
David Haussler	3755.4345703125
Don Chamberlin	3745.54296875
Frederick Jelinek	3727.75146484375
Linda G. Shapiro	3681.767333984375
Tomas Lozano-Perez	3597.21337890625
Lalit R. Bahl	3484.365478515625
Raymond A. Lorie	3453.755615234375
Terrence J. Sejnowski	3410.999755859375
Demetri Terzopoulos	3395.916748046875
Azriel Rosenfeld	3370.689697265625
Butler W. Lampson	3314.319580078125
Robert M. Haralick	3271.701416015625
Alan L. Yuille	3184.633056640625
Alex Pentland	3104.677978515625
Richard O. Duda	2840.168212890625
Anil K. Jain	2833.53271484375
Jon Louis Bentley	2791.63330078125
John V. Guttag	2688.9921875

Table 18: Combination of Author Rank and Page Rank with self-citation for Topic Machine Learning (25 most Influential Authors)

AuthorName	Score	AuthorName	Score
E. F. Codd	18399.134765625	Eugene C. Freuder	2852.529052734375
Daniel G. Bobrow	14275.8740234375	Richard O. Duda	2840.168212890625
Carl Hewitt	12347.6787109375	Gerald Jay Sussman	2838.2021484375
Ben Wegbreit	9271.6328125	Anil K. Jain	2833.53271484375
Andrew P. Witkin	7430.35205078125	Jean Babaud	2821.81005859375
Peter J. Denning	7198.57421875	M. Baudin	2821.81005859375
Robert Endre Tarjan	6088.55419921875	Alan J. Demers	2801.77978515625
Peter Boehler Bishop	5915.26806640625	Jon Louis Bentley	2791.63330078125
Richard Steiger	5915.26806640625	Muckai K. Girish	2750.896728515625
James R. Jackson	5750.12548828125	K. Mani Chandy	2711.01806640625
Jay Earley	5562.37841796875	Irene Greif	2699.468994140625
H. T. Kung	4987.71875	Irving L. Traiger	2691.023193359375
Rodney A. Brooks	4939.78271484375	John V. Guttag	2688.9921875
Geoffrey E. Hinton	4618.77734375	Richard J. Waldinger	2671.83984375
Joseph Abate	4354.34765625	Forest Baskett	2665.68359375
Richard P. Brent	4338.82666015625	Judea Pearl	2665.228759765625
David R. Musser	4316.86474609375	John Ross Quinlan	2613.190673828125
Harvey Dubner	4279.337890625	Jian-Qiang Hu	2604.32421875
Ellis Horowitz	4257.8330078125	Stéphane Mallat	2603.43603515625
Robert L. Mercer	4257.7353515625	Richard R. Muntz	2544.54443359375
Larry S. Davis	3981.80810546875	Robin Williams	2531.242919921875
Bertram Raphael	3860.931884765625	Brian Cantwell Smith	2495.075439453125
Susan L. Gerhart	3806.7587890625	Thomas O. Binford	2464.228759765625
Oded Goldreich	3787.566650390625	John W. Boyse	2460.12158203125
David Haussler	3755.4345703125	Franco P. Preparata	2448.564453125
Carl M. Harris	3749.491943359375	Tomaso Poggis	2435.02587890625
Michael Stonebraker	3747.854248046875	Richard M. Stallman	2418.098388671875
Don Chamberlin	3745.54296875	Yu-Hsin Lin	2407.6494140625
Frederick Jelinek	3727.75146484375	Andrew Birrell	2405.43798828125
Linda G. Shapiro	3681.767333984375	Ching-En Lee	2403.2353515625
Tomas Lozano-Perez	3597.21337890625	Leslie Lamport	2396.341064453125
Silvio Micali	3540.778076171875	David Harel	2330.34619140625
Jim Gray	3488.064453125	Steven W. Zucker	2311.266845703125
Lalit R. Bahl	3484.365478515625	Todd Matson	2297.789306640625
Raymond A. Lorie	3453.755615234375	Terry Winograd	2246.6162109375
Zohar Manna	3450.918701171875	King-Sun Fu	2234.163330078125
Terrence J. Sejnowski	3410.999755859375	Eugene Wong	2226.120361328125
Demetri Terzopoulos	3395.916748046875	Ward Whitt	2212.03857421875
Scott Shenker	3378.406494140625	David A. Patterson	2191.583984375
Azriel Rosenfeld	3370.689697265625	David R. Cheriton	2181.14208984375
Michael A. Wesley	3366.160888671875	Laveen N. Kanal	2154.9892578125
Butler W. Lampson	3314.319580078125	Alexandra Duel-Hallen	2146.438720703125
Robert M. Haralick	3271.701416015625	Raymond F. Boyce	2097.845458984375
Alan L. Yuille	3184.633056640625	Rakesh Agrawal	2096.28125
Ronald L. Rivest	3184.265625	Jean Vuillemin	2095.0771484375
Leslie G. Valiant	3164.701171875	Vladimir Vapnik	2087.853271484375
Alex Pentland	3104.677978515625	John M. Cioffi	2087.64697265625
Shafi Goldwasser	3093.48486328125	Michael Brady	2062.149658203125
Amir Pnueli	2984.33642578125	Rod M. Burstall	2057.70556640625
Kapali P. Eswaran	2877.698486328125	Adi Shamir	2054.127197265625

Table 19: 100 most Influential Authors without Self-Citation (Combination of Author Rank and Page Rank)

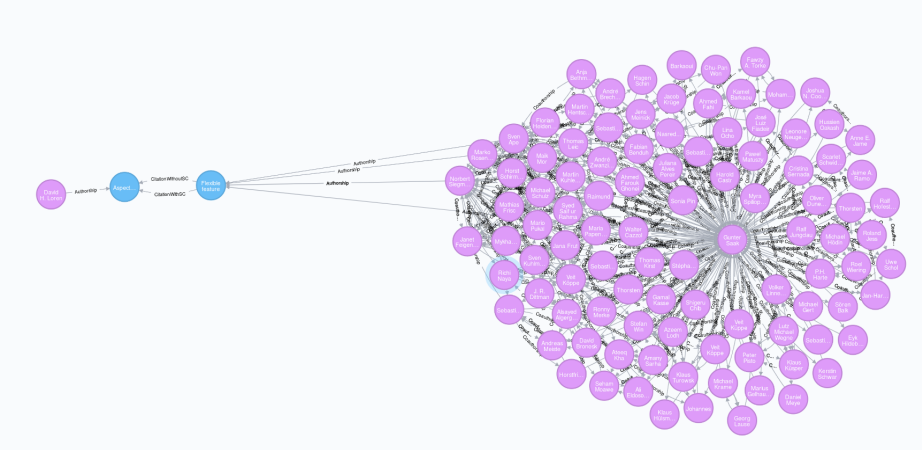

Fig. 6: Collaboration/Co-Authorship Network of Prof. Gunter Saake

References

1. Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.