Trends in Networking via Topic Modeling

Matthew Penza '19
Department of Computer Science
Princeton University
mpenza@cs.princeton.edu

Manik Dhar GS
Department of Computer Science
Princeton University
manikd@princeton.edu

15 January 2019

Abstract

Since the advent of the internet, networking has developed substantially. As the field has matured, the areas of inquiry have expanded and changed rapidly. How has the distribution of interest in various topics within the research community changed over the years? By leveraging tools from natural language processing, we attempt to formulate a data-informed answer to this question. In particular, we use latent Dirichlet allocation (LDA) on a corpus of titles, keywords, and abstracts extracted from the IEEE International Conference on Computer Communications (INFOCOM) conference from 1988 to 2018.

1 Introduction

Computer networking is a field with a long history spanning more than half a century. Over this time period, a large body of exciting research has been done to address the pressing needs of the times, develop new technologies, and lay the foundation for the future of the internet. Indeed, this body of work is now so large that the interaction of an individual researcher with it is limited to restricted parts thereof.

Systematically engaging with a large section of the networking literature might lead us to new insights into interesting questions, such as:

- How have computer networking conferences evolved over the past three decades?
- During what periods have the topics in a given paper been prominent in the past?
- How have classic papers affected the direction the field took around the time they were published?
- How similar or dissimilar are trends observed in various conferences?

While an expert deeply familiar with the field might attempt to answer these questions with some confidence, working towards a more systematic approach may lead to answers of superior reliability. To that end, we leverage topic models, [1] a tool from Machine Learning and natural language processing, to attempt to answer some of these questions. On a conceptual level, topic models are distributions where each topic represents a collection of related words. More precisely, it is a distribution over those words. The words in a document are thought to have arisen independently from a subset of these topics where each topic is weighed differently. In the reverse direction, given a new document, the model tries to assign weights to each candidate topic.

We will be using these models to investigate the proceedings IEEE International Conference on Computer Communications (INFOCOM) conference from 1988 to 2018.

2 Background

2.1 Related Work

Topic models have seen much use in analyzing scientific literature. These models have been used to gain insights into articles from the journal *Science*, the *Yale Law Review*, grant proposals to the American National Institute of Health, and other archives of documents in a variety of fields including history, literature, and physics. [1] [2] [3]

Of greater relevance to computer scientists is that topic models have been used in analyzing the trends in programming languages over the years. [4] Our work and methodology is indeed derived from such analyses.

2.2 Topic Models

We will be working with the latent Dirichlet allocation (LDA) model. [5] In this model, a topic is represented as a distribution over a fixed collection of words. The probability, or weight, of a given word represents how strongly the given word is related to the topic. For example, for a topic "Software Defined Networking", such words as "Openflow", "controller", and "data-plane" will probably be associated with it. A word like "TCP" may also be associated, but we would not expect it to have as high a weight. If there were a topic "Congestion Control", we would expect "TCP" to have a high weight in it, but we would not expect a high weight for "Openflow". Each topic is itself given a weight via a distribution which denotes how common a given topic is relative to the others in the corpus.

LDA treats each given document as a bag of words, i.e. the order is ignored and the document is treated as a collection of words where we keep track of duplicates (a multiset). A given document in the LDA model is produced by a multistage generative process.

Recall that each topic has a weight relative to the other topics in the corpus, and that each topic is a distribution over words. First, using the relative topic weights we randomly generate a mixture of topics. That is, we generate a list of numbers which denote how strongly each topic contributes to a document relative to all the other topics. Next, for each word we independently pick a topic and sample the word from that topic distribution. Therefore, for each document we produce a new random vector representing the relative proportions of each topic. Based on this the words are generated, implying different papers will likely have topics mixed in different proportions. The set of possible topics in a given LDA model is fixed.

In our problem, we begin with a corpus of documents and we want to infer what the "right set" of topics will be to produce the given corpus. This is not immediate from the given set. Concretely, we want to infer parameters for a set of distributions and proportions between them. An important problem in machine learning is to solve this problem in a principled fashion. Several methods exist to infer such a model, including expectation maximization and variational methods. [5]

3 Implementation

3.1 Dataset

We are working with data from the IEEE International Conference on Computer Communications (INFOCOM). We collected titles, keywords, and abstracts from the proceedings of INFOCOM from 1988 to 2018, giving us a dataset of 8,257 documents. This data is stored in the BibTex format, making it easily readable. In recent years the conference has had over 300 papers per proceeding, although 30 years ago the number was approximately half of that.

3.2 Methods

The data was stored in the form of BibTex citations. It includes the text of each paper's title, keywords, and abstract. We used the open-source PyPi module bibtexparser[1] to read and process the BibTex citations. Some papers are missing their abstracts, keywords, or both; we included these partial entries regardless. To train the model, for each paper we append these three fields together and treat the result as a single document.

Before running LDA, we have to be careful to disregard words such as "the", "or", and "because", as they are be very common but do not provide any meaningful information about the topics represented in the documents. Words like these are collected in a list of "stopwords" to be eliminated. We used the list of English stopwords provided by the Natural Language Toolkit (NLTK) library [6]. We augmented NLTK's stoplist with additional entries such as "exist", "because", etc., which we felt did not provide much information. We also removed dashes and underscores, such that "data-plane", for example, becomes "dataplane". During preprocessing, we transform each abstract to a bag of words. We next eliminate punctuation, and then we remove all words appearing in the stoplist. The bag of words is represented as a set of tuples. Each tuple consists of a word and the number of times it appears. To ensure that different forms of the same root word, such

[1] https://github.com/sciunto-org/python-bibtexparser

as "router" and "routers", are treated as the same word, we use the NLTK library's stemming functionality. [6]

To train the LDA model we used the PyPi module gensim[7]. It is required that the number of topics that the LDA model is to represent be specified at the beginning of training. The model does not provide human-readable names for the topics, instead specifying a word distribution for each topic learned. Readable topic names must be assigned manually if they are desired. We decided to run our topic model with 30 topics for simplicity and to satisfy time constraints.

3.3 Code Repository

We have made our implementation publicly available as a GitHub repository, which can be found at `https://github.com/mpenza19/LatentDirichletAlloc`

4 Results

4.1 Topics Generated

After training the model over our 30-year corpus, we manually assigned names to the topic distributions generated. To achieve this, for each document we generated a distribution over the topics representing the most likely mixture of topics the document could be generated from. The probability of a topic can be interpreted as the weight given to that topic. The higher the weight, the greater the contribution that topic has to the given document. Then for each topic we took the top 10 documents which gave that topic the highest weight. We inspected these papers and gave them names to the best of our ability. We list them below with some example titles. In some cases there was some ambiguity, which we specify where applicable.

1) Congestion control and Connection Management: "JetMax: Scalable Max-Min Congestion Control for High-Speed Heterogeneous Networks" [8] and "Efficient Distributed Admission Control for Core-Stateless Networks" [9].

2) Flow measurements: "A Quasi-Likelihood Approach for Accurate Traffic Matrix Estimation in a High Speed Network" [10] and "Decentralizing network inference problems with Multiple-Description Fusion Estimation (MDFE)" [11].

3) Information retrieval: *In this topic (like a few others later) the connection between the papers is weaker. Most of the papers dealt with issues related to information retrieval like caching and error correction. There was a paper or two which exactly did not fit this paradigm.* "From Uncertain Photos to Certain Coverage: a Novel Photo Selection Approach to Mobile Crowdsensing" [12] and "Asymptotic Miss Ratio of LRU Caching with Consistent Hashing" [13].

4) Optical Networks: "Integrated Intermediate Waveband and Wavelength Switching for Optical WDM Mesh Networks" [14] and "A generalized framework for analyzing time-space switched optical networks" [15].

5) Routing (Packet Classification, table lookup): "A fast IP routing lookup scheme for gigabit switching routers" [16] and "CutSplit: A Decision-Tree Combining Cutting and Splitting for Scalable Packet Classification" [17].

6) Unclear 1: *A common theme for this topic is unclear. A number of these papers are related to transport protocols and TCP in particular. The remaining papers are related to different topics like network topology and email protocols.* "TCP Vegas revisited" [18] and "An analysis of Internet inter-domain topology and route stability" [19].

7) Data Centers (Multicasting): "A combined group/tree approach for scalable many-to-many reliable multicast" [20] and "How bad is reliable multicast without local recovery?" [21].

8) RFID systems: "Season: Shelving interference and joint identification in large-scale RFID systems" [22] and "Tag size profiling in multiple reader RFID systems" [23].

9) Cognitive radio networks: "Utility-based cooperative spectrum sensing scheduling in cognitive radio networks" [24] and "Improved rendezvous algorithms for heterogeneous cognitive radio networks" [25].

10) Security (Authentication and Cryptography): "Cross-domain password-based authenticated key exchange revisited" [26] and "Checks and balances: A tripartite public key infrastructure for secure web-based connections" [27].

11) Graph Data Processing (Privacy, Machine Learning): "AppDNA: App Behavior Profiling via

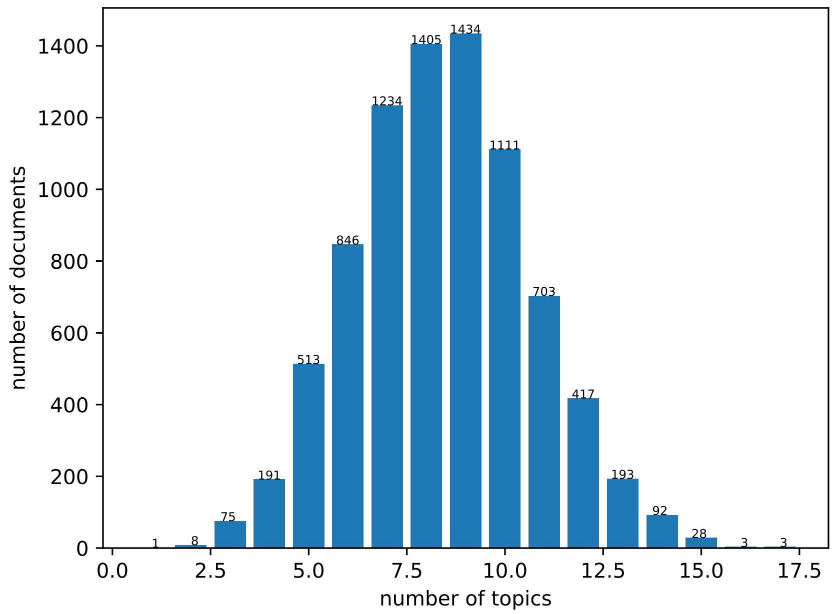


Figure 1: A histogram representing documents containing a given number of topics.

Graph-based Deep Learning" [28] and "Quantifying Graph Anonymity, Utility, and De-anonymity" [29].

12) Game Theory (online auctions and markets): "Truthful online double auctions for dynamic mobile crowdsourcing" [30] and "Flexauc: Serving dynamic demands in spectrum trading markets with flexible auction" [31].

13) Load Balancing / Network reliability: *There were two themes among high-weight documents for this topic distribution: load balancing and network reliability.* "An algorithm for capacity expansion of local access networks" [32] and "Optimizing Network Reliability via Best-First Search over Decision Diagrams" [33].

14) Network Security (Malware and Botnets): "Can we beat legitimate cyber behavior mimicking attacks from botnets?" [34] and "PeerClean: Unveiling peer-to-peer botnets through dynamic group behavior analysis" [35].

15) Queuing Analysis: "On stochastic recursive equations and infinite server queues" [36] and "A simple approximation for modeling nonstationary queues" [37].

16) Smart Grids: "Robust Multi-stage Power Grid Operations with Energy Storage" [38] and "Restricting Involuntary Extension of Failures in Smart Grids using Social Network Metrics" [39].

17) Task Scheduling: "Cluster fair queueing: Speeding up data-parallel jobs with delay guarantees" [40] and "Coupling task progress for MapReduce resource-aware scheduling" [41].

18) Algorithims (Graph / Optimization): "The k-Constrained Bipartite Matching Problem: Approximation Algorithms and Applications to Wireless Networks" [42] and "The online disjoint set cover problem and its applications" [43].

19) Video Streaming: "Statistical characteristics and multiplexing of MPEG streams" [44] and "Tracker-assisted rate adaptation for MPEG DASH live streaming" [45].

20) Web Hosting: "Design and performance of a Web server accelerator" [46] and "DotSlash: handling Web hotspots at dynamic content Web sites" [47].

21) Virtualization: "The Impact of Virtualization on Network Performance of Amazon EC2 Data Center" [48] and "Consolidating complementary VMs with spatial/temporal-awareness in cloud datacenters" [49].

22) Sensor Networks: "On full-view coverage in camera sensor networks" and "Extending Network

Lifetime for Precision-Constrained Data Aggregation in Wireless Sensor Networks".

23) Unclear 2: *Here a decent chunk of papers were on crowdsourcing but the rest were on unrelated topics.* "Promela++: a language for constructing correct and efficient protocols" [50] and "Efficient and flexible crowdsourcing of specialized tasks with precedence constraints" [51].

24) Wireless Scheduling: "Delay Guarantees for Throughput-Optimal Wireless Link Scheduling" [52] and "Optimal delay bound for maximum weight scheduling policy in wireless networks" [53].

25) Mobile networking: "EnLoc: Energy-Efficient Localization for Mobile Phones" [54] and "Efficient location management based on moving location areas" [55].

26) Unclear 3: *Multiple themes seem to be represented, such as distributed computing, ad hoc networking, and graph algorithms.* "Localized low-weight graph and its applications in wireless ad hoc networks" [56] and "Optimal Construction of Redundant Multicast Trees in Directed Graphs" [57].

27) Wireless transmission: "CARA: Collision-Aware Rate Adaptation for IEEE 802.11 WLANs" [58] and "Decoding interfering signals with fewer receiving antennas" [59].

28) Machine Learning (Sensor data): "MV-Sports: A Motion and Vision Sensor Integration-Based Sports Analysis System"

29) Privacy (Sensor Data): "Data perturbation with state-dependent noise for participatory sensing" [60] and "Towards Privacy-Preserving Speech Data Publishing" [61].

30) ATM networks / Packet switching: "Queueing analysis for ATM switching of mixed continuous-bit-rate and bursty traffic" [62] and "Performance of a crosspoint buffered ATM switch fabric" [63].

We see that although it is fairly easy to assign a name to most topic distributions, some seem to be trying to capture multiple ideas at once. Also of note is that some broader ideas seem to be appear in multiple topics. For example, we have wireless scheduling and wireless transmission both under the umbrella of wireless networking. Interestingly, very specific topics like smart grids and RFID sensors have their own topics while a topic like software-defined networking (SDN) doesn't appear. This could be happening because SDN can be represented well by a mixture of a few topic distributions but topics like smart grids are not amenable to that.

4.2 Trends

For each document we produced a vector representing a mixture of topics which generated the document. For most documents will not use all possible topics. In figure 1 we see that that no document has more than 17 topics included in it. Most documents in fact have no more than ten topics included. We have also checked that each document gives some non trivial probability (more than 0.05) to at least one topic.

Figure 2 gives a bird's eye view of how each topic has varied in importance over the years. For each year, we look at the weights assigned to each topic among papers published that year. By grouping and normalizing them we get a measure of percentage contribution of a topic in a given year.

A number of interesting trends can be observed. Such topics as sensor networks, mobile networking, and virtualization have grown in prominence in recent years, which aligns well with our intuition. Topics like congestion control and ATM networks/packet switching account for fewer papers over the years. This is reasonable since, during the early years of the conference, problems like congestion control and connection management would have been immensely popular. As the complexity of the field has grown, more problems came into the limelight, leading these rudimentary topics to account for smaller and smaller proportion as the years went on.

We also plotted the data as individual line graphs for each topic in Figure 3 and 4 at the end. This helps us look at trends not as readily apparent in Figure 2. One of the things we observe is how cryptography and authentication (plot 10) had a dip in popularity but regained the ground it lost in recent years. Conversely, congestion control (plot 1) experienced a surge of popularity from c. 1995-2005, but has dropped off drastically ever since.

We also notice that there are no topics that could be described as stable. The vast majority show either a dip or peak as discussed immediately above, or else a clear decline or increase in interest over time, perhaps with a few sudden spikes upwards or downwards along the way. Optical networks (plot 4) and virtualization (plot 21) are prime examples of such

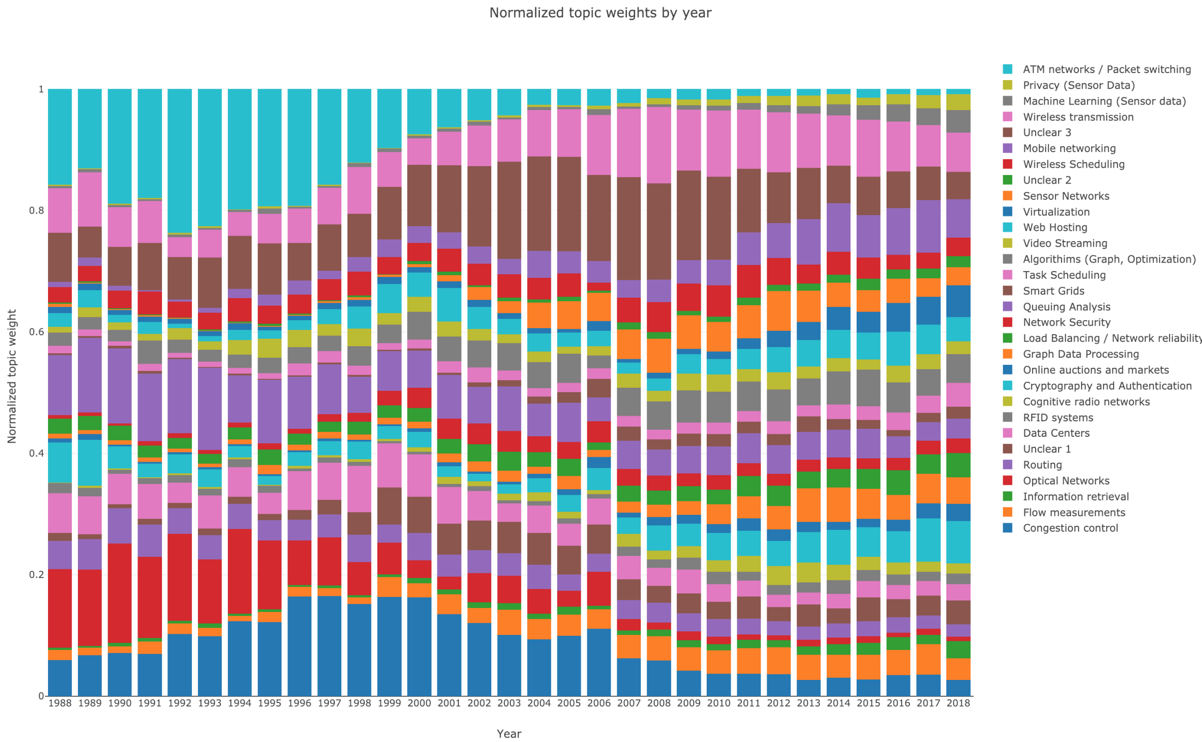


Figure 2: Normalized weights of each topic in the corpus over time.

decline and growth, respectively. The rest are mostly erratic, as exemplified by video streaming (plot 19) and wireless transmission (plot 27). The only topics that approach consistency over time are wireless scheduling (plot 24) and *unclear 1*, but only since c. 2010 in the latter case.

4.3 Finding similar papers

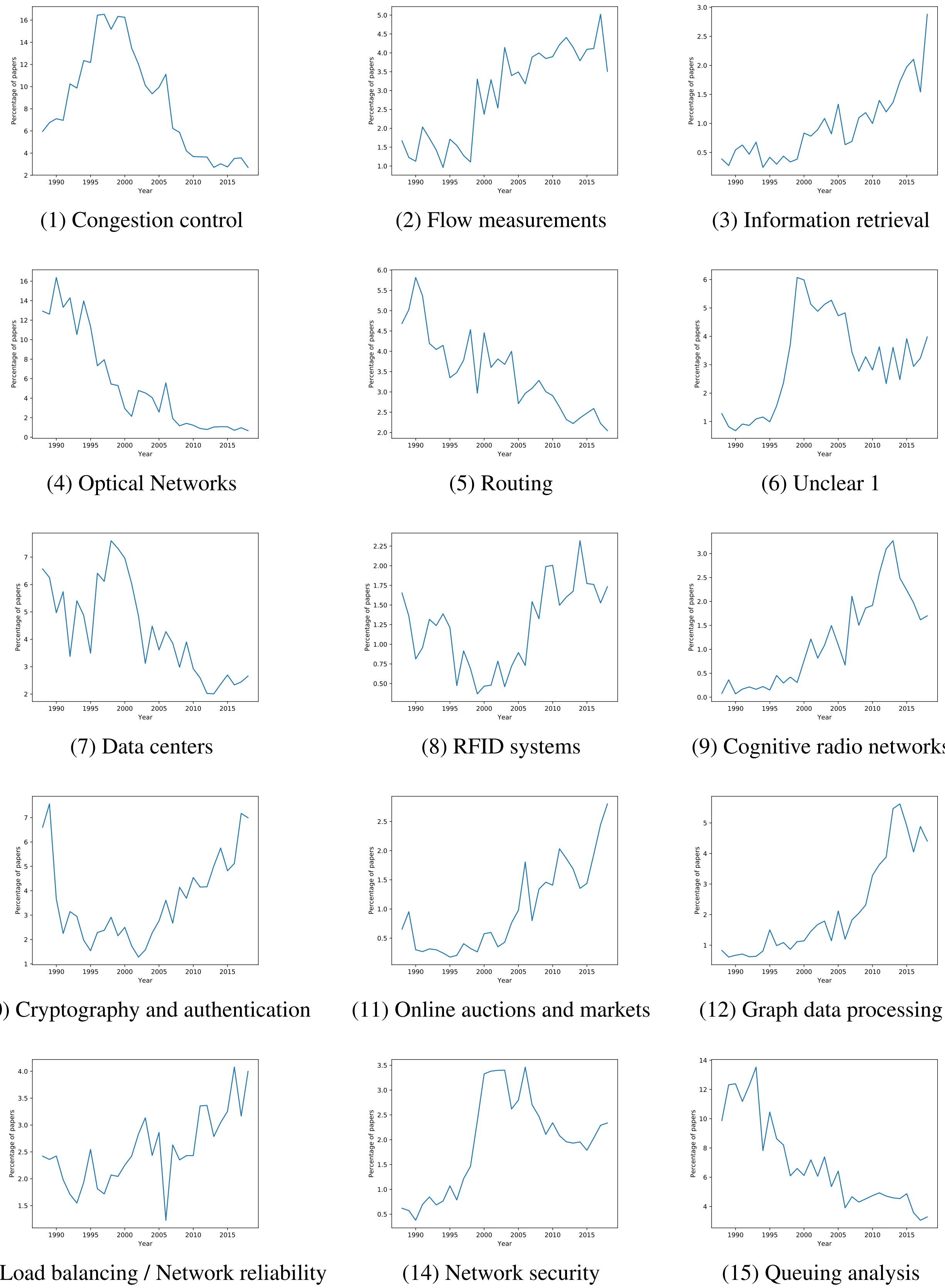
Now that we have a trained topic model, we can use it to compare documents. Given a document, the model can give a vector which contains each topic's weight. Now this weight vector can be used to compare it to the weight vectors of other documents. Given a paper from a different source, we can use this to find the most similar papers to the input paper.

We examined some papers from the 2016 SIGCOMM conference using our model. First we turned to the paper "Virtualized Congestion Control". [64] Upon examining the topic weight vector generated, we see the largest weight is assigned to the topic *congestion control* (0.37), followed by *virtualization* (0.13) and *task scheduling* (0.1), which appears to be reasonable for this particular paper. The three closest papers to it in the INFOCOM dataset are:

1. "Deadline-aware bandwidth sharing by allocating switch buffer in data center networks" [65],
2. "The cost of QoS support in edge devices an experimental study" [66], and
3. "New bandwidth sharing and pricing policies to achieve a win-win situation for cloud provider and tenants" [67].

We note that the first and the third paper are similar, as they too deal with the problem of bandwidth allocation.

Next, we examined the paper "SNAP: Stateful Network-Wide Abstractions for Packet Processing" [68]. The key topic here is related to software-defined networking (SDN), which our model doesn't explicitly represent in a single distribution. However, we did observe some SDN-related papers among the high-weight papers while naming the topics. The topic weight vector has multiple topics with similar weight. The three closest papers to it in the INFOCOM dataset are:



(1) Congestion control (2) Flow measurements (3) Information retrieval

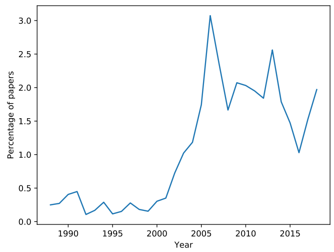
(4) Optical Networks (5) Routing (6) Unclear 1

(7) Data centers (8) RFID systems (9) Cognitive radio networks

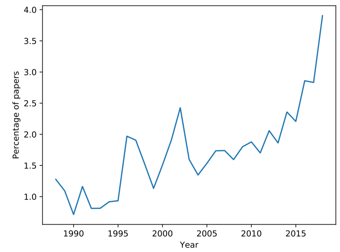
(10) Cryptography and authentication (11) Online auctions and markets (12) Graph data processing

(13) Load balancing / Network reliability (14) Network security (15) Queuing analysis

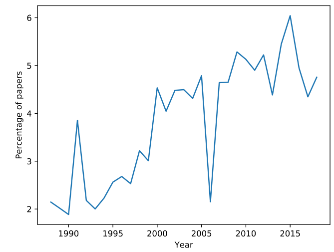
Figure 3: Individual trend lines plotting how the percentage of a topics contribution changes over the years



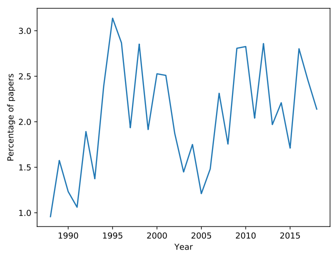
(16) Smart Grids



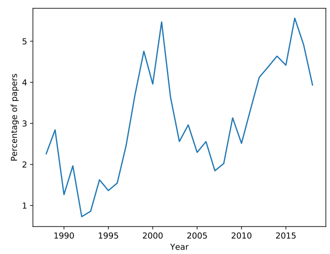
(17) Task Scheduling



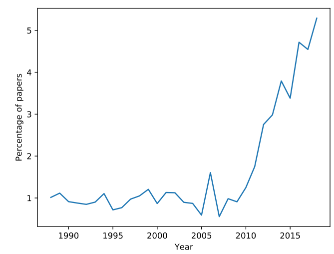
(18) Algorithms (Graph, Optimization)



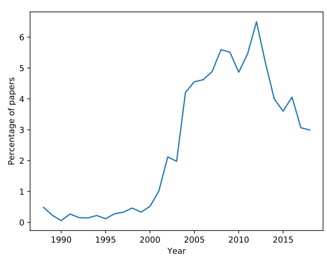
(19) Video streaming



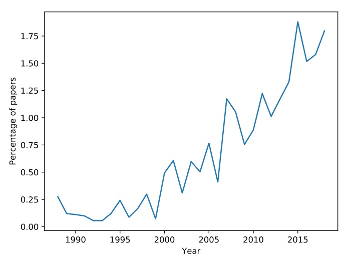
(20) Web hosting



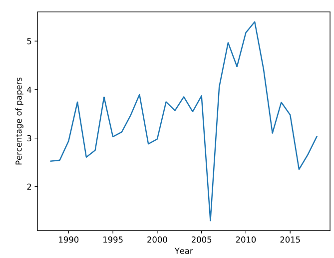
(21) Virtualization



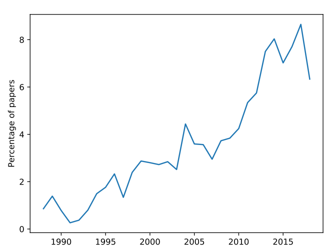
(22) Sensor networks



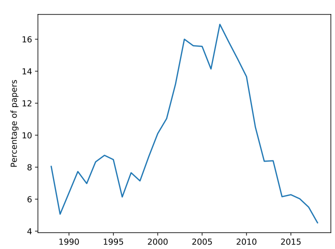
(23) Unclear 2



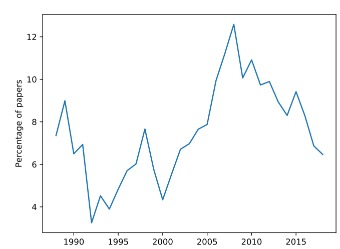
(24) Wireless scheduling



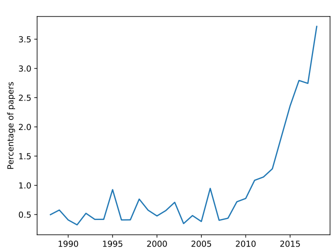
(25) Mobile networking



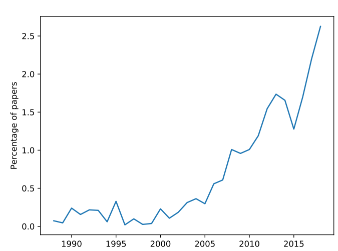
(26) Unclear 3



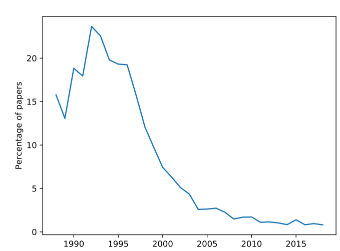
(27) Wireless transmission



(28) Machine learning (sensor data)



(29) Privacy (sensor data)



(30) ATM networks / Packet switching

Figure 4: Individual trend lines plotting how the percentage of a topics contribution changes over the years (continued from Figure 3)

1. "Network anti-spoofing with SDN data plane" [69],
2. "GMPLS Control Plane, Policy-Based Management, and Information Modeling" [70], and
3. "A distributed and robust SDN control plane for transactional network updates". [71]

Two of these are SDN-related papers. The second is also related, albeit less directly, as it deals with network management.

Finally, we examined the paper "The Deforestation of L2" [72]. This is another topic that did not appear in our topic list. The topic weight vector for this paper gave the highest weights to *unclear 1* and *unclear 3*. The closest 3 papers to it in the INFOCOMM dataset were:

1. "Flow labelled IP: a connectionless approach to ATM" [73],
2. "Switch-aided flooding operations in ATM networks" [74], and
3. "Analysis of point-to-point packet delay in an operational network" [75].

The second mentions link flooding and network spanning tree protocols, which are concepts related to the SIGCOMM paper in question.

5 Limitations and Future Work

From the list of topics generated we see that a lot of topics are mixed in a single one and there is ambiguity in some others. One way to deal with that is possibly to train the model with more topics allowed. Another is to use more sophisticated models. Latent Dirchlet Allocation (LDA) doesn't take into fact that some topics can be correlated to each other. To address that we can use Correlated Topic Models [76] and Pachinko Allocation Model [77].

It would also be interesting to compare results from different conferences or try to analyse them together. We did very simple experiments on papers from SIGCOMM which gave qualitative results. Training models on conferences like SIGCOMM and NSDI can allow us to compare trends between conferences. We could run the similarity experiment in both directions using multiple models. INFOCOM is a huge conference with a very broad scope. Working with a conference with a smaller scope may lead to more interesting results with simpler models. Instead of working on papers published in a conference, one can look at the papers published by an individual researcher and see how that has evolved over time.

Another possibility is to use dyanmic LDA [78] which allows us to model changes in topic contents. If we are working with multiple conferences it can help us observe if topics appearing in one conference influence other conferences.

6 Conclusion

We see that using LDA does give some interesting results and clearly shows broad trends in networking researchers' interests. We are able to see the attention paid to certain topics diminish to near zero despite having once been major focuses of the field, and others emerge from obscurity to take their place. This analysis has its place as a blunt instrument. For more refined results, however, the limitations addressed above ought to be dealt with, the more sensitive suggested models employed, and data from other conferences and journals analyzed.

References

[1] David M Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

[2] Jordan Boyd-Graber, Yuening Hu, David Mimno, et al. Applications of topic models. *Foundations and Trends® in Information Retrieval*, 11(2-3):143–296, 2017.

[3] Xiaolin Shi, Ramesh Nallapati, Jure Leskovec, Dan McFarland, and Dan Jurafsky. Who leads whom: Topical lead-lag analysis across corpora.

[4] Michael Greenberg, Kathleen Fisher, and David Walker. Tracking the flow of ideas through the programming languages literature. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 32. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.

[5] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[6] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text*

with the natural language toolkit. " O'Reilly Media, Inc.", 2009.

[7] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. http://is.muni.cz/publication/884893/en.

[8] Yueping Zhang, Derek Leonard, and Dmitri Loguinov. Jetmax: scalable max–min congestion control for high-speed heterogeneous networks. *Computer Networks*, 52(6):1193–1219, 2008.

[9] Avadora Dumitrescu. Efficient distributed admission control for core-stateless networks. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 4, pages 2864–2869. IEEE, 2005.

[10] Mehdi Malboubi, Cuong Vu, Chen-Nee Chuah, and Puneet Sharma. Decentralizing network inference problems with multiple-description fusion estimation (mdfe). *IEEE/ACM Transactions on Networking*, 24(4):2539–2552, 2016.

[11] J. Cao, A. Chen, and T. Bu. A quasi-likelihood approach for accurate traffic matrix estimation in a high speed network. In *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications*, pages 21–25, April 2008.

[12] Tongqing Zhou, Bin Xiao, Zhiping Cai, Ming Xu, and Xuan Liu. From uncertain photos to certain coverage: a novel photo selection approach to mobile crowdsensing. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pages 1979–1987. IEEE, 2018.

[13] Kaiyi Ji, Guocong Quan, and Jian Tan. Asymptotic miss ratio of lru caching with consistent hashing. *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pages 450–458, 2018.

[14] Mengke Li and Byrav Ramamurthy. Integrated intermediate waveband and wavelength switching for optical wdm mesh networks. 04 2006.

[15] R Srinivasan and Arun K Somani. A generalized framework for analyzing time-space switched optical networks. *IEEE Journal on Selected Areas in Communications*, 20(1):202–215, 2002.

[16] Nen-Fu Huang, Shi-Ming Zhao, Jen-Yi Pan, and Chi-An Su. A fast ip routing lookup scheme for gigabit switching routers. In *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1429–1436. IEEE, 1999.

[17] Wenjun Li, Xianfeng Li, Hui Li, and Gaogang Xie. Cutsplit: A decision-tree combining cutting and splitting for scalable packet classification. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pages 2645–2653, 04 2018.

[18] Urs Hengartner, Jürg Bolliger, and Thomas Gross. Tcp vegas revisited. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1546–1555. IEEE, 2000.

[19] Ramesh Govindan and Anoop Reddy. An analysis of internet inter-domain topology and route stability. In *INFOCOM'97. Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Driving the Information Revolution., Proceedings IEEE*, volume 2, pages 850–857. IEEE, 1997.

[20] Wonyong Yoon, Dongman Lee, Hee Yong Youn, Seungik Lee, and Seok Joo Koh. A combined group/tree approach for scalable many-to-many reliable multicast. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1336–1345. IEEE, 2002.

[21] Jörg Nonnenmacher, Martin Lacher, Matthias Jung, Ernst W Biersack, and Georg Carle. How bad is reliable multicast without local recovery? In *INFOCOM'98. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 972–979. IEEE.

[22] Lei Yang, Jinsong Han, Yong Qi, Cheng Wang, Tao Gu, and Yunhao Liu. Season: Shelving interference and joint identification in large-scale rfid systems. In *INFOCOM, 2011 Proceedings IEEE*, pages 3092–3100. IEEE, 2011.

[23] Shigeng Zhang, Xuan Liu, Jianxin Wang, and Jiannong Cao. Tag size profiling in multiple reader rfid systems. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pages 1–9. IEEE, 2017.

[24] H. Li, X. Xing, J. Zhu, X. Cheng, K. Li, R. Bie, and T. Jing. Utility-based cooperative spectrum sensing scheduling in cognitive radio networks. *IEEE Transactions on Vehicular Technology*, 66(1):645–655, Jan 2017.

[25] Zhaoquan Gu, Haosen Pu, Qiang-Sheng Hua, and Francis C M Lau. Improved rendezvous algorithms for heterogeneous cognitive radio networks. In *2015 IEEE conference on computer communications (INFOCOM)*, pages 154–162. IEEE, 2015.

[26] L. Chen, H. W. Lim, and G. Yang. Cross-domain password-based authenticated key exchange revisited. In *2013 Proceedings IEEE INFOCOM*, pages 1052–1060, April 2013.

[27] Jing Chen, Shixiong Yao, Quan Yuan, Ruiying Du, and Guoliang Xue. Checks and balances: A tripartite public key infrastructure for secure web-based connections. In *INFOCOM 2017-IEEE Conference on Computer Communications, IEEE*, pages 1–9. IEEE, 2017.

[28] Shuangshuang Xu, Lan Zhang, Anran Li, Xiang-Yang Li, Chuyan Liu, and Wenchao Huang. Appdna: App behavior profiling via graph-based deep learning. *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pages 1475–1483, 2018.

[29] Shouling Ji, Tianyu Du, Zhen Hong, Ting Wang, and Raheem A. Beyah. Quantifying graph anonymity, utility, and de-anonymity. *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pages 1736–1744, 2018.

[30] Yueming Wei, Yanmin Zhu, Hongzi Zhu, Qian Zhang, and Guangtao Xue. Truthful online double auctions for dynamic mobile crowdsourcing. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 2074–2082. IEEE, 2015.

[31] Peng Lin, Xiaojun Feng, and Qian Zhang. Flexauc: Serving dynamic demands in spectrum trading markets with flexible auction. In *INFOCOM, 2014 Proceedings IEEE*, pages 2265–2273. IEEE, 2014.

[32] Alexander Shulman and Rita Vachani. An algorithm for capacity expansion of local access networks. In *INFOCOM'90, Ninth Annual Joint Conference of the IEEE Computer and Communication Societies. The Multiple Facets of Integration. Proceedings, IEEE*, pages 221–229. IEEE, 1990.

[33] Masaaki Nishino, Takeru Inoue, Norihito Yasuda, Shin-ichi Minato, and Masaaki Nagata. Optimizing network reliability via best-first search over decision diagrams. *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pages 1817–1825, 04 2018.

[34] Shui Yu, Song Guo, and Ivan Stojmenovic. Can we beat legitimate cyber behavior mimicking attacks from botnets? In *INFOCOM, 2012 Proceedings IEEE*, pages 2851–2855. IEEE, 2012.

[35] Qiben Yan, Yao Zheng, Tingting Jiang, Wenjing Lou, and Y Thomas Hou. Peerclean: Unveiling peer-to-peer botnets through dynamic group behavior analysis. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 316–324. IEEE, 2015.

[36] Eitan Altman. On stochastic recursive equations and infinite server queues. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 2, pages 1295–1302. IEEE, 2005.

[37] Wei-Ping Wang, David Tipper, and Sujata Banerjee. A simple approximation for modeling nonstationary queues. In *INFOCOM'96. Fifteenth Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation. Proceedings IEEE*, volume 1, pages 255–262. IEEE, 1996.

[38] Yihan Zou, Xiaojun Lin, Dionysios Aliprantis, and Minghua Chen. Robust multi-stage power grid operations with energy storage. *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pages 2483–2491, 04 2018.

[39] Jose Cordova-Garcia, Dong-Liang Xie, and Xin Wang. Restricting involuntary extension of failures in smart grids using social network metrics. *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pages 2510–2518, 04 2018.

[40] C. Chen, W. Wang, S. Zhang, and B. Li. Cluster fair queueing: Speeding up data-parallel jobs with delay guarantees. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, May 2017.

[41] J. Tan, X. Meng, and L. Zhang. Coupling task progress for mapreduce resource-aware scheduling. In *2013 Proceedings IEEE INFOCOM*, pages 1618–1626, April 2013.

[42] A. Berger, J. Gross, and T. Harks. The k-constrained bipartite matching problem: Approximation algorithms and applications to wireless networks. In *2010 Proceedings IEEE INFOCOM*, pages 1–9, March 2010.

[43] A. Pananjady, V. K. Bagaria, and R. Vaze. The online disjoint set cover problem and its applications. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 1221–1229, April 2015.

[44] M. Krunz, R. Sass, and H. Hughes. Statistical characteristics and multiplexing of mpeg streams. In *Proceedings of INFOCOM'95*, volume 2, pages 455–462 vol.2, April 1995.

[45] A. Detti, B. Ricci, and N. Blefari-Melazzi. Tracker-assisted rate adaptation for mpeg dash live streaming. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, April 2016.

[46] E. Levy-Abegnoli, A. Iyengar, Junehwa Song, and D. Dias. Design and performance of a web server accelerator. In *IEEE INFOCOM '99. Conference on Computer Communications. Proceedings. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. The Future is Now (Cat. No.99CH36320)*, volume 1, pages 135–143 vol.1, March 1999.

[47] W. Zhao and H. Schulzrinne. Dotslash: handling web hotspots at dynamic content web sites. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 4, pages 2836–2840 vol. 4, March 2005.

[48] G. Wang and T. S. E. Ng. The impact of virtualization on network performance of amazon ec2 data center. In *2010 Proceedings IEEE INFOCOM*, pages 1–9, March 2010.

[49] L. Chen and H. Shen. Consolidating complementary vms with spatial/temporal-awareness in cloud datacenters. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pages 1033–1041, April 2014.

[50] A. Basu, G. Morrisett, and T. Von Eicken. Promela++: a language for constructing correct and efficient protocols. In *Proceedings. IEEE INFOCOM '98, the Conference on Computer Communications. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Gateway to the 21st Century (Cat. No.98*, volume 2, pages 455–462 vol.2, March 1998.

[51] A. Chatterjee, M. Borokhovich, L. R. Varshney, and S. Vishwanath. Efficient and flexible crowdsourcing of specialized tasks with precedence constraints. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, April 2016.

[52] K. Kar, X. Luo, and S. Sarkar. Delay guarantees for throughput-optimal wireless link scheduling. In *IEEE INFOCOM 2009*, pages 2331–2339, April 2009.

[53] C. BoyacÄś and Y. Xia. Optimal delay bound for maximum weight scheduling policy in wireless networks. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pages 565–573, April 2014.

[54] I. Constandache, S. Gaonkar, M. Sayler, R. R. Choudhury, and L. Cox. Enloc: Energy-efficient localization for mobile phones. In *IEEE INFOCOM 2009*, pages 2716–2720, April 2009.

[55] Y. Bejerano and I. Cidon. Efficient location management based on moving location areas. In *Proceedings IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No.01CH37213)*, volume 1, pages 3–12 vol.1, April 2001.

[56] Xiang-Yang Li, Yu Weng, Peng-Jun Wan, Wen-Zhan Song, and O. Frieder. Localized low-weight graph and its applications in wireless ad hoc networks. In *IEEE INFOCOM 2004*, volume 1, pages 431–442, March 2004.

[57] Y. Bejerano and P. V. Koppol. Optimal construction of redundant multicast trees in directed graphs. In *IEEE INFOCOM 2009*, pages 2696–2700, April 2009.

[58] J. Kim, S. Kim, S. Choi, and D. Qiao. Cara: Collision-aware rate adaptation for ieee 802.11 wlans. In *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, pages 1–11, April 2006.

[59] Z. Li, X. Dai, and K. G. Shin. Decoding interfering signals with fewer receiving antennas. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, April 2016.

[60] Fan Zhang, Li He, Wenbo He, and Xue Liu. Data perturbation with state-dependent noise for participatory sensing. In *2012 Proceedings IEEE INFOCOM*, pages 2246–2254, March 2012.

[61] Jianwei Qian, âĞd' Feng, Jiahui Hou, Chunhong Zhang, Yu Wang, and Xiang-Yang Li. Towards privacy-preserving speech data publishing. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 04 2018.

[62] T. . Hou and A. K. Wong. Queueing analysis for atm switching of mixed continuous-bit-rate and bursty traffic. In *Proceedings. IEEE INFOCOM '90: Ninth Annual Joint Conference of the IEEE Computer and Communications*, pages 660–667 vol.2, June 1990.

[63] P. Goli and V. Kumar. Performance of a crosspoint buffered atm switch fabric. In *[Proceedings] IEEE INFOCOM '92: The Conference on Computer Communications*, pages 426–435 vol.1, May 1992.

[64] Bryce Cronkite-Ratcliff, Aran Bergman, Shay Vargaftik, Madhusudhan Ravi, Nick McKeown, Ittai Abraham, and Isaac Keslassy. Virtualized congestion control. In *Proceedings of the 2016 ACM SIGCOMM Conference*, pages 230–243. ACM, 2016.

[65] J. Zhang. Deadline-aware bandwidth sharing by allocating switch buffer in data center networks. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, April 2016.

[66] R. Guerin, L. Li, S. Nadas, P. Pan, and V. Peris. The cost of qos support in edge devices an experimental study. In *IEEE INFOCOM '99. Conference on Computer Communications. Proceedings. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. The Future is Now (Cat. No.99CH36320)*, volume 2, pages 873–882 vol.2, March 1999.

[67] H. Shen and Z. Li. New bandwidth sharing and pricing policies to achieve a win-win situation for cloud provider and tenants. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pages 835–843, April 2014.

[68] Mina Tahmasbi Arashloo, Yaron Koral, Michael Greenberg, Jennifer Rexford, and David Walker. Snap: Stateful network-wide abstractions for packet processing. In *Proceedings of the 2016 ACM SIGCOMM Conference*, pages 29–43. ACM, 2016.

[69] Y. Afek, A. Bremler-Barr, and L. Shafir. Network anti-spoofing with sdn data plane. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, May 2017.

[70] H. Lonsethagen, A. Karasen, A. Welin, B. Berde, and A. Hajjaoui. Gmpls control plane, policy-based management, and information modeling. In *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, pages 1–12, April 2006.

[71] M. Canini, P. Kuznetsov, D. Levin, and S. Schmid. A distributed and robust sdn control plane for transactional network updates. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 190–198, April 2015.

[72] James McCauley, Mingjie Zhao, Ethan J Jackson, Barath Raghavan, Sylvia Ratnasamy, and Scott Shenker. The deforestation of l2. In *Proceedings of the 2016 ACM SIGCOMM Conference*, pages 497–510. ACM, 2016.

[73] P. Newman, T. Lyon, and G. Minshall. Flow labelled ip: a connectionless approach to atm. In *Proceedings of IEEE INFOCOM '96. Conference on Computer Communications*, volume 3, pages 1251–1260 vol.3, March 1996.

[74] Yih Huang and P. K. McKinley. Switch-aided flooding operations in atm networks. In *Proceedings of INFOCOM '97*, volume 3, pages 1080–1087 vol.3, April 1997.

[75] B. . Choi, S. Moon, Zhi-Li Zhang, K. Papagiannaki, and C. Diot. Analysis of point-to-point packet delay in an operational network. In *IEEE INFOCOM 2004*, volume 3, pages 1797–1807 vol.3, March 2004.

[76] David M. Blei and John D. Lafferty. Correlated topic models. In *Proceedings of the 18th International Conference on Neural Information Processing Systems*, NIPS'05, pages 147–154, Cambridge, MA, USA, 2005. MIT Press.

[77] Wei Li and Andrew McCallum. Pachinko allocation: Dag-structured mixture models of topic correlations. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 577–584, New York, NY, USA, 2006. ACM.

[78] David M. Blei and John D. Lafferty. Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 113–120, New York, NY, USA, 2006. ACM.