Internet Complexity Through Five Decades of RFCs

ABSTRACT

The functioning of the Internet requires interoperability between networks, systems, and applications, and cooperation amongst a growing number of stakeholders. The Request for Comments (RFC) series plays a critical role in enabling this cooperation and interoperability. As the series marks its 50th anniversary, we analyse the shifts and trends that have emerged, including the rise and fall of large Internet players, and the shift in relevance between industry and academia. In addition, we show that the Internet's growth and maturity has given rise to a longer standardisation process, and provide initial insight into the growing complexity.

ACM Reference Format:

. 2020. Internet Complexity Through Five Decades of RFCs. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The behaviour of the Internet is defined by a bewildering array of protocols. For more than 50 years, from the earliest experimentation with the ARPANET to the modern Internet, these protocols have been described in the Request for Comments (RFCs) series of documents (https://www.rfc-editor.org/). The creation of RFC document has been led by the Internet Engineering Task Force (IETF), organised into a set of Working Groups (WGs). These bring together diverse stakeholders who debate and formulate final versions of the protocols. RFCs therefore provide a unique historical record of the development of the Internet, and of how the process by which the Internet is developed has evolved.

Despite this, to date this diverse body of data has seen little analaytical attention. We argue that by studying the RFC series and the IETF, we can begin to understand how the complexity of both the network, and of the standards process that defines the network, have grown over time. For example, this can not only reveal the social process that has driven decades of protocol design but also potentially shed light on major bottlenecks that could enable future streamlining of the standardisation process.

In this paper, we perform the first measurement study of RFCs to shed light on 50 years of IETF activities. To this end, we have collected a dataset comprising 8643 RFCs, from 5548 authors, working at 2175 institutions. Using this, we perform a bibliometric analysis that shows how the scope and complexity of the protocols, and the standards development process, have evolved. Our analysis is driven by two perspectives. We start by investigating the generation of the RFC documents, particularly in terms of their spread across categories (Section 3). Then we study the patterns of authorship and collaboration (Section 4). Our analysis reveals how RFC data can be exploited to shed light on network protocols standardisation activities, as well as delineating a set of bibliometric methodologies that can be used to achieve this. We also provide preliminary evidence regarding the complexity of the underlying social process that drives standardisation. To the best of our knowledge this is the first paper studying the social dimension of the development of Internet protocols.

the paper.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00

https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 DATASET & METHODOLOGY

The RFC series started in 1969 as a way to organise work-in-progress notes of the community creating the ARPANET. As the ARPANET grew, developed and became the modern Internet, the RFC series also evolved to become the primary publication series for technical standards developed by the IETF, and for other documents relating to the development of the Internet [8].

RFCs were originally relatively informal working notes, but the series has become more formalised and structured over time. It currently comprises four active *publication streams*, one for each of the organisations that produce RFCs: the IETF stream, the Internet Research Task Force (IRTF) stream, the Internet Architecture Board (IAB) stream, and the Independent Submission stream. The IETF is an open standards forum that develops technical specifications and standards for the Internet. The IRTF is an associated venue that conducts longer-term research. Both are organised as activities under the IAB. Independent submissions provide a non-standards route to RFC publication, with a separate review process. The publication streams coordinate closely to avoid publishing conflicting material. The RFC Editor (https://www.rfc-editor.org/) provides an archival publication services for the four streams. RFCs published prior to the formalisation of publication streams comprise the Legacy stream.

Each RFC also has a *category*. The categories are Standards Track, Best Current Practice, Experimental, Informational, and Historic. Some RFCs prior to the formalisation of categories have Unknown category. Not all RFCs are standards, and the standards track is subdivided into Proposed, Draft, and Full Standard documents depending on maturity. Only the IETF stream can publish RFCs in the Standards Track category. Only the IETF and IAB streams can publish RFCs in the Best Current Practice category. Experimental and Informational RFCs can be published by any stream, and RFCs from any stream can be categorised as Historic [13, 15, 16]. Each stream has its own procedures and review processes. Early RFCs were published with only limited review, but RFCs from all streams now undergo extensive peer review prior to publication [3].

The RFC Editor maintains the master archive of the RFC documents, along with an index providing metadata about the RFCs, their authors, and their inter-relationships. From this index, we extract the following data: title, publication date, author names and affiliations, relationship with other RFCs, category, and publication stream. For RFCs published on the IETF stream, we also extract the name of the Working Group that developed the RFC, and the Area of the IETF in which that working group was chartered. The IETF is organised in working groups working in a particular topic (e.g., the "idr" working group focuses on interdomain routing and belongs to the "routing")

We augment this dataset with information from the IETF Datatracker (https://datatracker.ietf.org). The Datatracker is the main administrative database used by the IETF and IRTF. It records the status of work-in-progress *Internet-Drafts* that eventually become RFCs, as well as information about participants, working groups, meetings, etc. This information can be used to track the development of documents before they're published as RFCs. The Datatracker was introduced in the early 2000s, and contains little historical data prior to that, but holds comprehensive information about more recent IETF activities. We use information from the Datatracker in Section 3.3 to make an initial attempt at quantifying the impact of growing complexity within the standardisation process. The Datatracker provides a REST API, that we access with an open source library we developed.

All in all, we collect information on 5548 authors, belonging to 2175 institutions, from 8643 RFCs covering the period 1969–2019.

3 DOCUMENTS

We begin by looking at the documents published in the RFC series, how they fit into the different publication streams and categories, and their inter-relationships.

ToDo: We should have a plot showing the cumulative number of RFCs in each publication stream over time

3.1 Overview of Document Generation

Figure 1a shows the cumulative number of publications in each category over time. Three phases can be seen in the development of the RFC series. First, in the period 1969-1974, RFCs are published at a rapid rate during the initial development of the ARPANET. Second, growth in the period 1974–1985 slows, as the relatively small community gains experience with using the network and develops the core set of applications and protocols. Finally, with the creation of the IETF and the introduction of the National Science Foundation Network (NSFNET) in 1986, the community and the number of RFCs published starts to expand rapidly. This is further driven by the opening of the network to commercial and public use in the mid-1990s, and continues to this day.

Figure **??** shows the cumulative number of RFCs published in each stream over time.

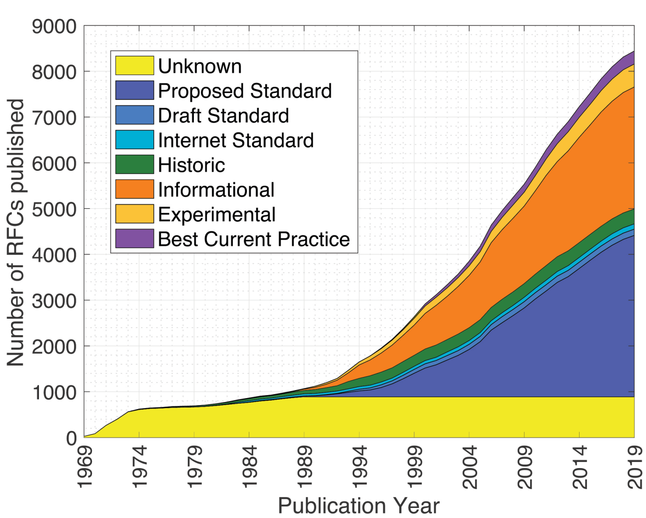
RFCs are initially seen to comprise a homogeneous series, with categories and the three-stage standards track (*i.e.*, Proposed, Draft, and Internet Standard) being introduced shortly after the formation of the IETF in the mid-1980s.

New RFCs since that time have been broadly split between Standards Track and Informational categories, with much smaller numbers of Best Current Practice and Experimental RFCs, and a steady trickle of older documents (about 318) being classified as Historic as they become obsolete.

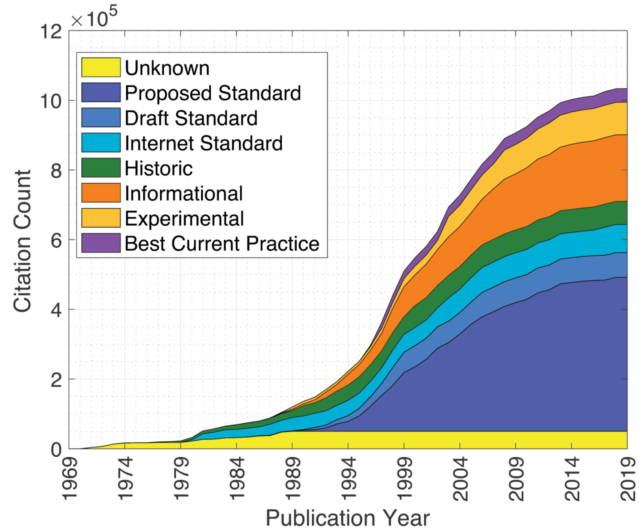
There is a steady growth in the number of Standards Track RFCs published, but it is clear that the three-stage standards track was not successful. Very few RFCs are advanced from Proposed Standard, to Draft Standard, to Full Standard. Recognising this, the IETF abolished the Draft Standard maturity level in 2011, leaving only a two-stage standards process [14]. This change also does not seem to have been a success. The Internet still runs primarily on Proposed Standards and few RFCs advance to Internet Standard.

Informational RFCs have been the second largest group of publications since the early 1990s. This category includes different types of RFCs including those providing design rationales, operational guidance, deployment guidelines and considerations, IETF procedural concerns, and "April 1st" satirical publications.

The highest number of contributions in Internet RFCs are observed in 2005, at the peak of the standardisation of SIP and related standards relating to voice-over-IP and Internet Telephony. The rate of publication has slowed in the last few

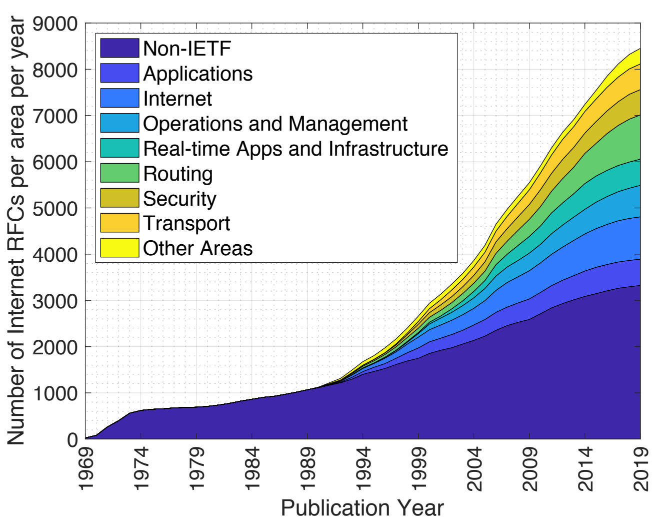


(a) RFC publications

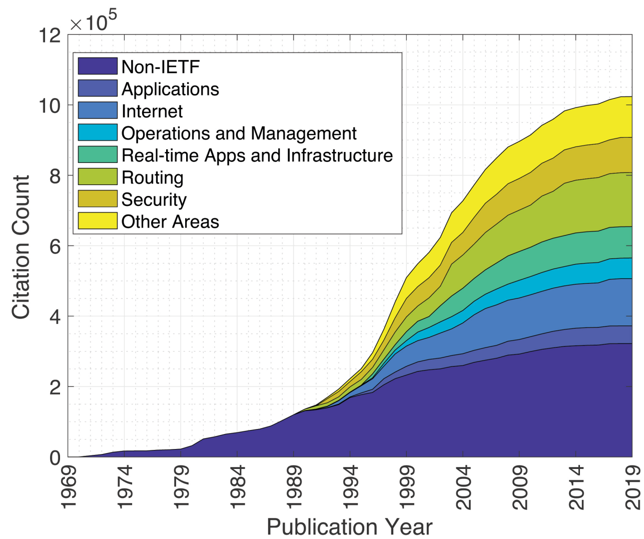


(b) RFC citations

Figure 1: Cumulative RFCs by category during 1969–2019



(a) RFC publications.



(b) RFC citations.

Figure 2: Cumulative RFCs by research area during 1969–2019

years, following the completion of large work programmes relating to HTTP/2 and WebRTC.

Figure 2a shows the cumulative number of RFCs published, subdivided by area, where applicable. The introduction of areas for IETF RFCs is observed around 1990. Areas remain relatively stable over time, with the most notable shift when the Real-time Applications and Infrastructure area became the Applications and Real-Time area around 2014. Another important trend that can be observed is the significant rise in routing RFCs, owing to the development of MPLS, service function chaining, and fat tree routing for data centres.

3.2 Citation Trends

Figure 2b shows the number of citations received by each RFC category in each year. The standards track draws a disproportionate attention in terms of citations, with Proposed Standards being by far the most popular RFCs. This points again to the heavy reliance of the Internet in this particular category and the difficulties of the Standardisation process previously discussed. The most cited RFC[1], however, belongs to the Experimental category, one of the least cited ones. This RFC together with all those in the top most cited ones, have been published by the Networks group,one of the most active ones. In contrast, Informational RFCs receive proportionally less citations. This is partially due to the administrative nature of the content contained by many of these documents. From the point of view of the publication streams, the most cited RFCs are overwhelmingly published through the IETF stream, with RFCs published through the Independent and IRTF stream lagging far beyond.

With respect to the area, Internet, Routing, Real-time Apps and Infrastructure as well as Security are the most popular ones. This reflects that the standardisation efforts mentioned were welcomed in the research community where this topics were rather popular.

Figure 2b shows the cumulative number of citations received by RFCs over time, by status. While Internet standards are few in comparison with other RFC statuses, they attract a disproportionate attention in terms of citations. However, the RFCs attracting most citations are Proposed Standards,

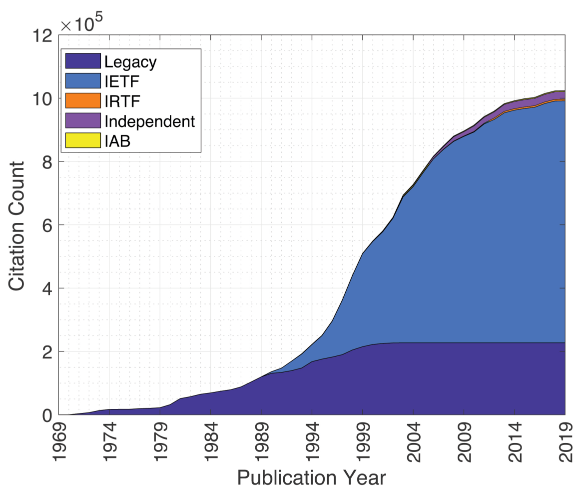


Figure 3: Cumulative RFC citations by streams during 1969–2019.IG: to be removed, just for commenting

[1]https://tools.ietf.org/html/rfc3561

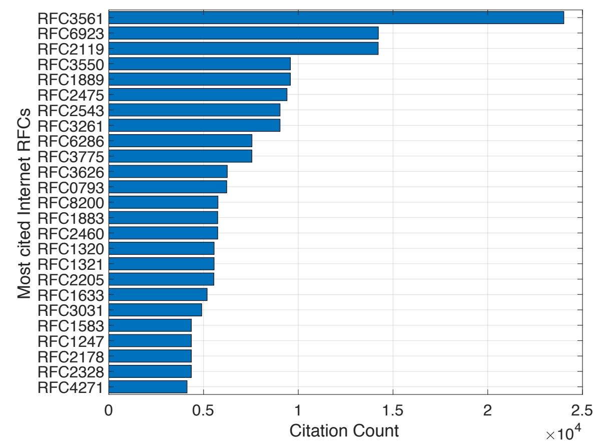


Figure 4: Most Cited Internet RFCs during 1969–2019.IG: to be removed. just for commenting

3.3 Complexity

While, as identified in the previous section, the standardisation has changed over time, leading to more RFCs from a larger set of areas, has it become more complex? In this section, we make an initial attempt to quantify the complexity of the process. To do this, we analyse the time taken for an RFC to be published, beginning from its first draft. We use data from the IETF Datatracker, looking at RFCs published between 2000–2019, inclusive, where a timeline (consisting of at least one prior draft) can be constructed. This gives a dataset of 5747 RFCs.

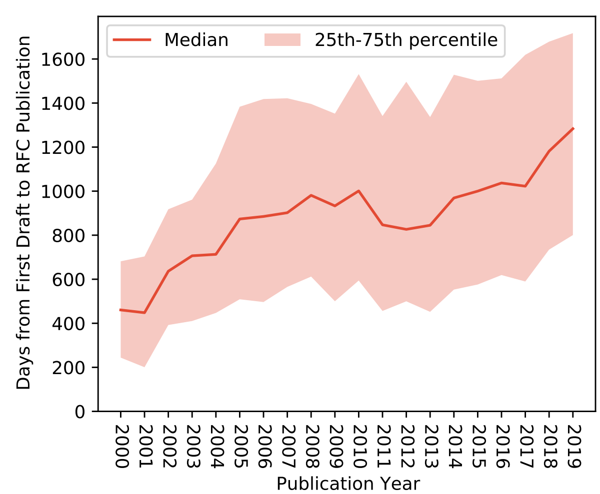


Figure 5: Time to publish for RFCs published 2000–2019.

Figure 5 shows the *time to publish* for the RFCs in the Datatracker dataset. We define time to publish as the number of days from the submission of the first draft recorded for that RFC, to the RFC's date of publication. This gives a broad measure of *complexity*, encompassing the coordination costs between authors, the process of building consensus, and the editing and publication process itself.

We can see that the standardisation process is taking longer over time: the median time to publish was 460 days in 2000, and rose to 1283 days in 2019. While, as noted, this metric encompasses many aspects of the process, we can begin to analyse the factors that may contribute to this increasing complexity.

Stream	Page Count	References	Authors
All	0.22 (5747)	0.18 (5077)	0.18 (5747)
IAB	0.34 (91)	0.22 (70)	0.21 (91)
Independent	0.07 (274)	0.10 (252)	0.10 (274)
Legacy	0.12 (177)	0.32 (15)	0.03 (177)
IRTF	0.19 (76)	0.04 (76)	0.05 (76)
IETF	0.23 (5129)	0.19 (4664)	0.19 (5129)
None	0.23 (878)	0.21 (841)	0.16 (878)
app	0.18 (343)	0.13 (281)	0.01 (343)
art	0.48 (185)	0.37 (185)	0.32 (185)
gen	0.43 (18)	-0.30 (13)	0.14 (18)
int	0.29 (686)	0.22 (613)	0.17 (686)
ops	0.17 (610)	0.06 (527)	0.01 (610)
rai	0.22 (544)	0.34 (511)	0.17 (544)
rtg	0.24 (884)	0.14 (841)	0.15 (884)
sec	0.17 (465)	0.16 (420)	0.16 (465)
subip	0.16 (19)	-0.16 (12)	-0.48 (19)
tsv	0.27 (497)	0.13 (420)	0.10 (497)

Table 1: Correlation coefficients with time to publish for page counts, number of referenced documents, and number of authors (total RFCs in parenthesis).

Table 1 lists the correlations between time to publish and three features of the RFCs in the Datatracker dataset: the number of pages, referenced documents, and authors. The dataset used to measure references is a subset of the Datatracker dataset, looking only at those RFCs where that referenced at least one document.

As shown, there are weak, but statistically significant, correlations between all three factors and time to publish. This follows the intuition: as documents become longer, include more references, and have more authors, it takes longer for them to be standardised.

4 AUTHORSHIP

Given the correlation between increasing complexity and the number of authors that an RFC has, we now further explore trends in authorship.

We begin, in Figure **??**, by analysing top authors, quantified by their total number of authored RFCs. Unsurprisingly, Jon Postel tops the list, due to his substantial role in the early standardisation of the Internet, including as the first Editor of the RFC series. The same is true for other RFC Editors, including Steve Crocker and Joyce Reynolds.

Since one of the reasons pointing to an increasing complexity and time to publish RFCs is the number of authors, we know explore the authorship of RFCs.

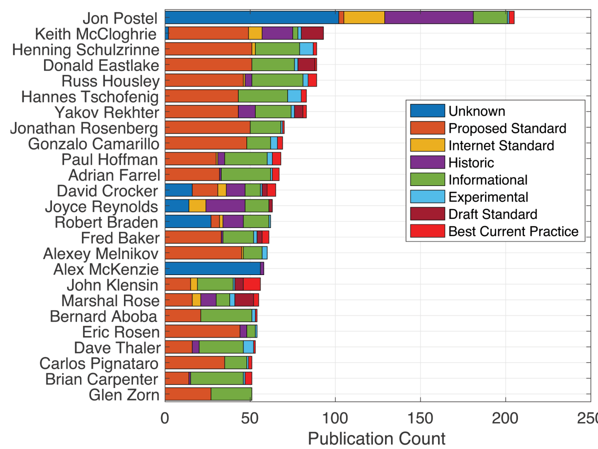


Figure 6: Most publishing authors during 1969–2019 according to their paper count in RFCs.

Moving beyond individuals, Figure 7 shows the cumulative number of RFCs published by authors at different companies, and overall by authors from academic and industrial institutions. This figure depicts the broad journey of the Internet, from an academic project, to a critical infrastructure component whose operation relies upon the cooperation of large industrial organisations.

Figure 7 depicts the top RFC publishers according to the number of authored RFCs. Since RFCs stem from the need of interoperability and coordination, the affiliation of RFC authors reflects to a great extent the stakeholders of the underlying ecosystem. While RFC publication was dominated in its very early years by authors from academic institutions, as the ARPANET evolved into the early Internet, academic and commercial institutions became responsible for similar fractions of RFCs. Through this period, the academic institutions responsible for the largest number of RFCs were frequently also nodes of the ARPANET (*e.g.*, UCLA, MIT).

With the greater involvement of commercial organisations in the early 90s this balance definitely shifted away from academic organisations. Nowadays, the vast majority of RFCs is authored by authors with industrial affiliation.

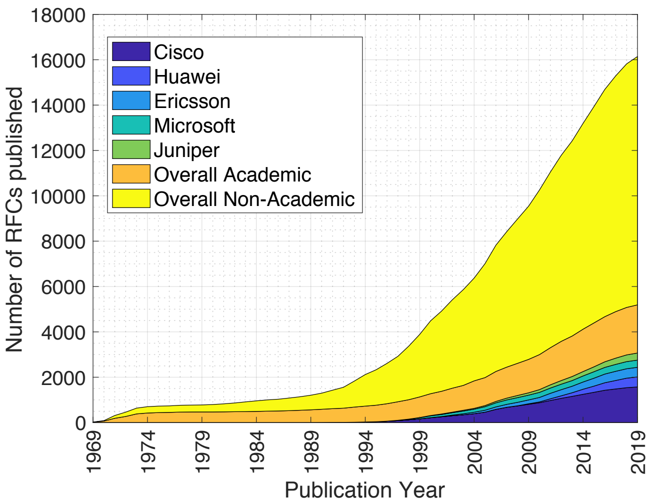


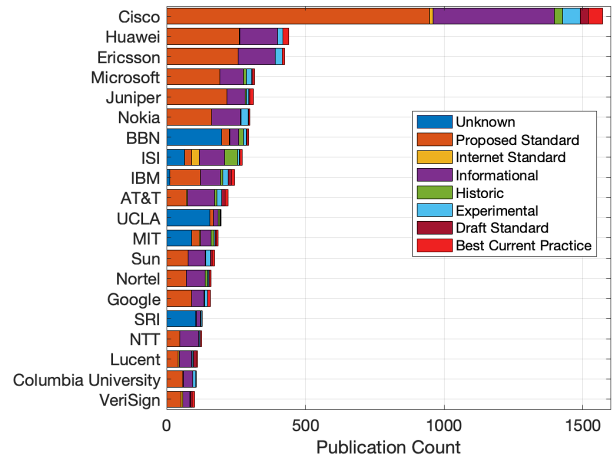
Figure 7: Cumulative Distribution of publications in academic vs non-academic organizations during 1969–2019.

Figure 8 illustrates this further, showing the top institutions by authorship of published RFCs. As shown, the top 5 institutions are all industrial contributors. While it is clear that Cisco is dominant, there are underlying shifts that are shown in Figure 7. Much of Cisco's contributions were made in the early 2000s, with significant work on routing standards, including MPLS. However, there is a marked rise in contributions from Huawei, beginning around 2010.

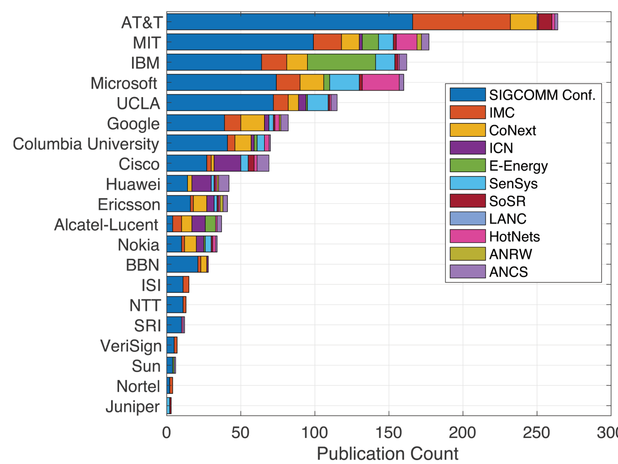
5 RELATED WORK

To the best of our knowledge, we are the first to study the IETF's publication practices. That said, there have been a number of related works performing bibliometric analysis [4, 7, 19, 20, 22, 24, 27]. Some work have also undertaken quantitative analysis on the top ACM conferences. The purpose of these studies is to determine the genre of the article and to understand the publication culture of these conferences [9]. Many previous works have performed an analysis on the content of various research areas using topic modeling [21] and keyword-based analysis [5].

A number of studies have used social networking analysis for social and medical science research [2, 6, 23, 25, 26], using social network analysis on generally social media data and altmetric data [11]. Social media analysis has not been



(a) Internet RFCs



(b) SIGCOMM conferences

Figure 8: Most publishing institutes during 1969–2019 according to their paper count in Internet RFCs and SIGCOMM.

used to determine the communities in computer networking research due to which we do not yet have complete insights into the collaborating patterns that exist in computer networking research.

Different researchers analyzed referencing patterns in research articles to identify incorporation of the latest studies relating to a research article [12] and citation analysis of the productivity of various research entities [1, 10].

Limited work has focused on using bibliometric or scientometric techniques to analyze the publication mores of the field of computer networks. Chiu et al. [4] have performed an analysis of author productivity in computer networking venues in 2010. Iqbal et al. [18] studied computer networking conferences and journals, whereas in [17] they studied 50 years of SIGCOMM sponsored publications.

6 CONCLUSIONS

Internet protocol standards have a critical socioeconomic role, in underpinning the Internet, and ensuring its interoperability. In this paper, we have shown that this process is becoming increasingly diverse, with more authors, from more regions, companies, and institutions. While this is beneficial to the process, we have also demonstrated that it increases its complexity: RFCs are taking longer to publish than ever before.

While we have provided initial insights into the growing complexity of the process, the importance of Internet protocol standardisation means that we must seek to better understand *how* decisions are made, such that the benefits of diversity can be maintained.

REFERENCES

[1] Christoph Bartneck and Jun Hu. 2009. Scientometric analysis of the CHI proceedings. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 699–708.

[2] Stephen P Borgatti, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. 2009. Network analysis in the social sciences. *science* 323, 5916 (2009), 892–895.

[3] B. E. Carpenter and C. Partridge. 2010. Internet Requests for Comments (RFCs) as Scholarly Publications. *Computer Communication Review* 40, 1 (Jan. 2010), 31–33. https://doi.org/10.1145/1672308.1672315

[4] Dah Ming Chiu and Tom ZJ Fu. 2010. Publish or perish in the internet age: a study of publication statistics in computer networking research. *ACM SIGCOMM Computer Communication Review* 40, 1 (2010), 34–43.

[5] Jinho Choi, Sangyoon Yi, and Kun Chang Lee. 2011. Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. *Information & Management* 48, 8 (2011), 371–381.

[6] Fereshteh Didegah and Mike Thelwall. 2018. Co-saved, co-tweeted, and co-cited networks. *Journal of the Association for Information Science and Technology* (2018).

[7] João M Fernandes and Miguel P Monteiro. 2017. Evolution in the number of authors of computer science publications. *Scientometrics* 110, 2 (2017), 529–539.

[8] H. Flanagan. 2019. RFC 8700: Fifty Years of RFCs. *IETF RFC-editor* (2019).

[9] Matthias Flittner, Mohamed Naoufal Mahfoudi, Damien Saucez, Matthias Wählisch, Luigi Iannone, Vaibhav Bajpai, and Alex Afanasyev. 2018. A Survey on Artifacts from CoNEXT, ICN, IMC, and SIGCOMM Conferences in 2017. *ACM SIGCOMM Computer Communication Review* 48, 1 (2018), 75–80.

[10] Brahim Hamadicharef. 2012. Scientometric study of the IEEE transactions on software engineering 1980-2010. In *Proceedings of the 2011 2nd International Congress on Computer Applications and Computational Science*. Springer, 101–106.

[11] Saeed-Ul Hassan, Mubashir Imran, Uzair Gillani, Naif Radi Aljohani, Timothy D Bowman, and Fereshteh Didegah. 2017. Measuring social media activity of scientific literature: an exhaustive comparison of scopus and novel altmetrics big data. *Scientometrics* 113, 2 (2017), 1037–1057.

[12] Leonard Heilig and Stefan Voß. 2014. A scientometric analysis of cloud computing literature. *IEEE Transactions on Cloud Computing* 2, 3 (2014), 266–278.

[13] Russ Housley and Harald T. Alvestrand. 2009. IESG Procedures for Handling of Independent and IRTF Stream Submissions. RFC 5742. (Dec. 2009). https://doi.org/10.17487/RFC5742

[14] Russ Housley, D Crocker, and E Burger. 2011. Reducing the standards track to two maturity levels. *IETF RFC 6410* (2011).

[15] IAB and Leslie Daigle. 2007. Process for Publication of IAB RFCs. RFC 4845. (July 2007). https://doi.org/10.17487/RFC4845

[16] IAB and Leslie Daigle. 2007. The RFC Series and RFC Editor. RFC 4844. (July 2007). https://doi.org/10.17487/RFC4844

[17] Waleed Iqbal, Junaid Qadir, Saeed-Ul Hassan, Rana Tallal Javed, Adnan Noor Mian, Jon Crowcroft, and Gareth Tyson. 2019. Five decades of the ACM special interest group on data communications (SIGCOMM) a bibliometric perspective. *ACM SIGCOMM Computer Communication Review* 49, 5 (2019), 29–37.

[18] Waleed Iqbal, Junaid Qadir, Gareth Tyson, Adnan Noor Mian, Saeed-ul Hassan, and Jon Crowcroft. 2019. A bibliometric analysis of publications in computer networking research. *Scientometrics* 119, 2 (2019), 1121–1155.

[19] S Nattar. 2009. Indian journal of physics: A scientometric analysis. *International Journal of Library and Information Science* 1, 4 (2009), 043–61.

[20] Gustavo Cattelan Nobre and Elaine Tavares. 2017. Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study. *Scientometrics* 111, 1 (2017), 463–492.

[21] Michael Paul and Roxana Girju. 2009. Topic modeling of research fields: An interdisciplinary perspective. In *Proceedings of the International Conference RANLP-2009*. 337–342.

[22] Periyaswamy Rajendran, R Jeyshankar, and B Elango. 2011. Scientometric analysis of contributions to Journal of Scientific and Industrial Research. *International Journal of Digital Library Services* 1, 2 (2011), 79–89.

[23] Miloš Savić, Mirjana Ivanović, and Bojana Dimić Surla. 2017. Analysis of intra-institutional research collaboration: a case of a Serbian faculty of sciences. *Scientometrics* 110, 1 (2017), 195–216.

[24] Alexander Serenko, Nick Bontis, and Joshua Grant. 2009. A scientometric analysis of the proceedings of the McMaster world congress on the management of intellectual capital and innovation for the 1996-2008 period. *Journal of Intellectual Capital* 10, 1 (2009), 8–21.

[25] Caroline S Wagner, Travis A Whetsell, and Loet Leydesdorff. 2017. Growth of international collaboration in science: revisiting six specialties. *Scientometrics* 110, 3 (2017), 1633–1652.

[26] Hajra Waheed, Saeed-Ul Hassan, Naif Radi Aljohani, and Muhammad Wasif. 2018. A bibliometric perspective of learning analytics research landscape. *Behaviour & Information Technology* (2018), 1–17.

[27] Zhifeng Yin and Qiang Zhi. 2017. Dancing with the academic elite: a promotion or hindrance of research production? *Scientometrics* 110, 1 (2017), 17–41.