The Web We Weave: Untangling the Social Graph of the IETF

Prashant Khare,[1] **Mladen Karan,**[1] **Stephen McQuistin,**[2] **Colin Perkins,**[2]
Gareth Tyson,[1,3] **Matthew Purver,**[1,4] **Patrick Healey,**[1] **Ignacio Castro**[1]
[1] Queen Mary University of London
[2] University of Glasgow
[3] Hong Kong University of Science & Technology
[4] Jožef Stefan Institute
p.khare@qmul.ac.uk, m.karan@qmul.ac.uk, sm@smcquistin.uk, csp@csperkins.org, g.tyson@qmul.ac.uk,
m.purver@qmul.ac.uk, p.healey@qmul.ac.uk, i.castro@qmul.ac.uk

Abstract

The Internet Engineering Task Force (IETF) has developed many of the technical standards that underpin the Internet. The standards development process followed by the IETF is open and consensus-driven, but is inherently both a social and political activity, and latent influential structures might exist within the community. Exploring and understanding these structures is essential to ensuring the IETF's resilience and openness. We use network analysis to explore the social graph of IETF participants, based on public email discussions and co-author relationships, and the influence of key contributors. We show that a small core of participants dominates: the top 10% contribute almost half (43.75%) of the emails and come from a relatively small group of organisations. On the other hand, we also find that influence has become relatively *more* decentralised with time. IETF participants also propose and work on drafts that are either adopted by a working group for further refinement or get rejected at an early stage. Using the social graph features combined with email text features, we perform regression analysis to understand the effect of user influence on the success of new work being adopted by the IETF. Our findings shed useful insights into the behavior of participants across time, correlation between influence and success in draft adoption, and the significance of affiliated organisations in the authorship of drafts.

1 Introduction

The global success of the Internet owes much to its open development process, a focus on permissionless innovation, and the ready interoperability enabled by its underpinning technical protocol standards. These standards support interworking between a diverse range of systems implemented by different vendors, and encourage the development of a vibrant, open, ecosystem. Given how crucial the Internet has become, it is, however, vital to understand *who* develops and maintains these standards, as they, and the companies they are affiliated with, have the power to fundamentally shape the Internet.

The technical standards that define the Internet are largely developed and maintained by the Internet Engineering Task Force (IETF). The IETF develops and maintains Internet protocols, including those for internetworking and transport (TCP/IP and QUIC), routing (BGP, MPLS), security (TLS), and application protocols such as HTTP and WebRTC. The IETF follows an open, consensus-driven process and does not have a formal membership, thereby posing few barriers to entry. The standards it develops are publicly available at no cost and, more importantly for our purposes, the IETF also makes available its email archives, working documents, meeting minutes, etc., providing transparent access to rich datasets that document decades of activities. This allows us to study the process by which the Internet protocols where developed in unprecedented detail. The IETF data provides a representative use case of large-scale, long-lived, distributed online collaboration, and since the dataset long pre-dates the COVID-19 pandemic by several decades, lets us to generate longitudinal insights and patterns pertaining to our research questions.

Protocol standardisation is an inherently social process. Most day-to-day work happens on public mailing lists, aided by meetings, video conference calls, and open document and code repositories. We are specifically interested in better understanding how influence is distributed across stakeholders and how it might affect the standardisation process. This is of critical societal importance: the IETF has a major impact on global Internet technologies, and understanding the social processes involved would give us insight into not only the driving forces behind standardisation, but also its resilience to the loss of major participants. Thus, we ask the following *research questions*:

(i) How centralised is the active IETF community, and to what extent is it reliant on a small core of participants?

(ii) How do the most influential participants behave?

(iii) How does influence (determined by mailing list participation) relate to wider impacts throughout the IETF?

(iv) Does the organisational affiliation of participants also influence the innovation (adoption of new work) within IETF?

To answer these questions, we collect public mailing list archives (2000–2019) containing more than 2.1M messages from almost 45K senders. We then generate a social interaction graph from these public mailing lists (§2). We find that, akin to many prior social graph studies (Kourtellis et al. 2013; Weitzel, Quaresma, and de Oliveira 2012), influence

Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

resides primarily with a small cohort of influential participants (§3.1). In fact, removing just 5% of IETF participants disintegrates the social graph, reducing the Largest Connected Component (LCC) by around 41%. We show that members of this group contribute a disproportionate fraction of the email communications, participate in a wider range of IETF areas, remain active in the community for longer, and discuss a broader range of topics (§3.2).

We then demonstrate that this mailing list influence also translates into draft authorship and leadership roles in the IETF, finding significant overlaps in these different communities (§3.3). Overall, up to 42% of the working drafts in the IETF come from the top 10% of mailing list participants. We examine participants' affiliations to determine which organisations have influence in the IETF, finding that a large fraction of influential participants are affiliated with a small number of prominent organisations.

Finally, we conjecture that the success of new work proposals may be predictable, based on prior social interactions by their authors. To test this, we build models to quantify the most determinant features (§4). We discover that, indeed, the social graph does play a significant role in predicting the adoption of work by the IETF.

To the best of our knowledge, this is the first study to characterise the social graph of the IETF standards development community. Our key findings include: (*i*) influence evident in the mailing list social graph is also reflected in document authorship and in IETF leadership roles, i.e., in defining the technical standards themselves; (*ii*) participants are affiliated with a relatively small set of organisations, but show an increasingly willingness to collaborate with authors from other organisations: 772 (2000-2004) vs. 3083 (2015-2019) jointly authored documents; and (*iii*) the social graphs of participants demonstrate influence, and, combined with email text features, they are a predictor of the success of their documents, allowing us to build a model that predicts if new work proposals will be adopted by the IETF.

2 Background & Datasets

Standardisation — The technical standards that define the Internet are primarily developed and maintained by the IETF. The IETF was established in 1986 to coordinate the development of Internet technologies, as a follow-on to the early research projects that developed the ARPAnet and other Internet precursor networks. Through the evolution of the Internet (Böttger et al. 2018), the IETF has been instrumental in developing many core protocols, such as TCP/IP, HTTP, or more recent ones such as QUIC and DoH (Böttger et al. 2019). The IETF has also worked with the W3C to jointly develop many recent critical web standards, including WebRTC (Jennings, Hardie, and Westerlund 2013) and WebSockets. For instance, many of the video conferencing applications, which have been so vital in supporting remote work during the COVID-19 pandemic, are built on WebRTC (Arkko et al. 2020).

The IETF is an open standards development organisation with no formal membership, and rather comprises a large, and long-lived, volunteer community. This openness, and the resulting social graph, exposes dynamics of the standards development process, and makes it interesting to study.

Lifecycle of IETF standards — The IETF develops technical standards, published in the RFC series of documents,[1] via a collaborative work process that reflects the inherent social and political nature of standardisation. This is managed through public mailing lists, supported by regular plenary and working group (WG) meetings. RFCs begin as *Internet-Drafts* submitted to the IETF for community discussion. Before an Internet-Draft can be *published* as a standards-track RFC, it must first be *adopted* by a WG that will conduct a technical review of the material. Drafts are authored by individual participants, and that model of explicit authorship is maintained even after WG adoption. Around 20% of drafts are adopted by a WG, and around 75% of these WG-adopted drafts are published as an RFC. Both of these stages, prior to and after WG adoption, involve multiple rounds of review and revision.

Datasets — We use of three data sources: the IETF Datatracker, the corpus of Internet-Drafts, and the public IETF mail archives. The Datatracker is an administrative database used to manage the IETF's work. It contains metadata about Internet-Draft submissions and their authors. This, when combined with the Internet-Draft corpus, gives us data about 32,872 drafts and 8216 authors, spanning from 2000-2019. Note, this period excludes the COVID-19 pandemic and removes the impact of this highly disruptive event. Studying this constitutes future work.

The IETF provides a rich email archive, including lists discussing WG activities, meetings, and administration. We gather 2,106,804 emails from 56,733 email addresses. To match email senders to IETF participants listed in the Datatracker, we apply the entity resolution approach of McQuistin et al. (2021), finding 44,741 unique participants.

Ethical considerations — Participation in the IETF is bound by agreements and policies explicitly stating that mailing list discussions and Datatracker metadata will be made publicly available.[2] We use only this publicly available data in our analysis. We have discussed our work with the IETF leadership to confirm that it fits their acceptable use policies. We have also made provisions to manage the data securely, and retain it only as necessary for our work. Our work is reproducible.[3] We open source our tooling, which accesses the public datasets.

3 The IETF as a Social Graph

Much of the interaction between IETF participants occurs on public mailing lists. For each year, we build a social graph based on the *active community* of participants (nodes) who have interactions (edges) with any other participant in the previous 5 years. McQuistin et al. (2021) demonstrated that

[1] RFC used to stand for *Request For Comments*, but the RFC series of documents has evolved over the past fifty years to become the publication venue for Internet standards (Flanagan 2019).

[2] See both https://www.ietf.org/about/note-well/ and the IETF privacy policy available at https://www.ietf.org/privacy-statement/.

[3] GitHub repository for codebase- https://github.com/sodestream/icwsm22

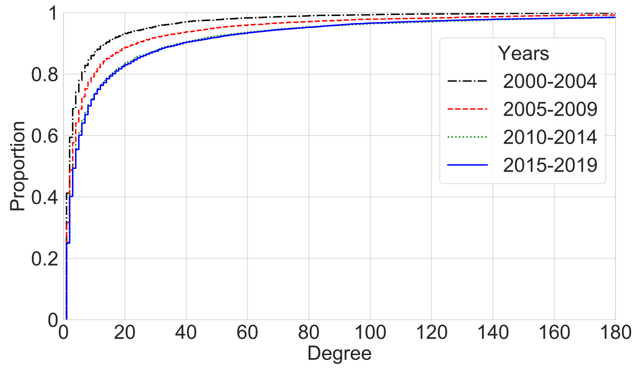

Figure 1: Cumulative degree distribution of the email graph for different year-periods.

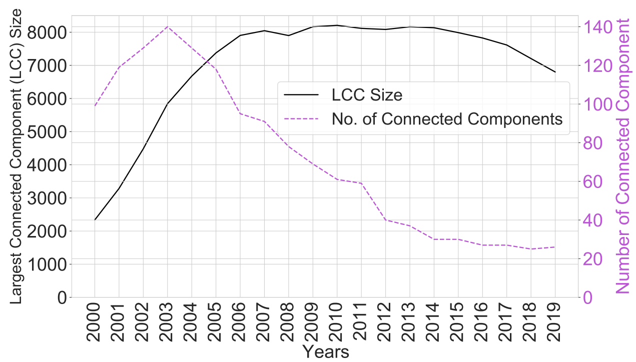

Figure 2: Size of Largest Connected Component (LCC) and Number of Connected Components (NCC)

there are three categories of participants in IETF - (a) *young contributors* – who leave within one year of their first year mailing list contribution; (b) *mid-age contributors* – participants who stay active for up-to 5 years; and (c) *senior contributors* – who remain active for more than 5 years. Following this, a 5 year period window is chosen to observe interactions. There are no direct participant-to-participant email exchanges in our dataset: the emails and responses we capture are sent to the public mailing lists, and we observe an interaction between two participants when one replies to an email sent by another on any mailing list. This yields a social graph based on 1,049,793 emails (out of 2.1M) from 22,138 unique participants across 840 mailing lists. Through the lens of these discussions, we next explore the dynamics of the standardisation process and how stakeholders influence it.

3.1 Measuring Influence

Participation — To examine the variation in participation between IETF participants, in Figure 1 we plot the cumulative degree distribution of the social graph – i.e., the cumulative number of people a participant interacts with. We observe a core group that always interacts with substantially more people than the rest of the community, with around 80% of the participants having degree <15, but 5%-10%

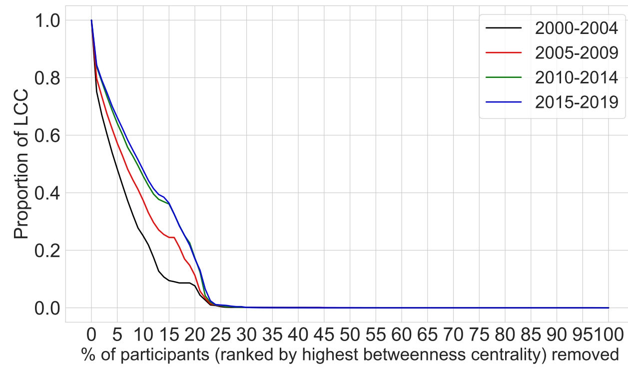

Figure 3: Impact of removing participants by their influence on the size of the LCC.

having degree >40. However, this difference decreases over time: in 2005-2009 around 6% of participants had degree >40, but by the period 2015-2019 this increases to $\sim 10\%$. This shows that while there is always a core of more active participants, the degree of participation has spread out over time.

To understand the structure of the community, and its reliance on specific groups, we analyse the connected components of this graph. Each connected component reflects a maximal set of nodes such that each pair of nodes is connected by a path. We compute the size of the Largest Connected Component (LCC) and the Number of Connected Components (NCC) for each year in Figure 2. The NCC peaks in 2003 before declining, and broadly aligns with the variation in the number of meeting attendees,[4] suggesting a lag between participating for the first time and integrating into the wider community. In contrast, the size of the LCC increases until 2006 before stabilising. Overall, we observe that the IETF community has become less fragmented.

Influence — Relying on a small group to interconnect the community could undermine the resilience of the IETF. To study the influence of participants and their role in interconnecting the wider community, we compute the betweenness centrality of each participant in a given time period (Kourtellis et al. 2013; Weitzel, Quaresma, and de Oliveira 2012; Solé-Ribalta et al. 2014). We also considered other graph based influence metrics such as eigenvector centrality: eigenvector centrality reflects the importance of a node as per its neighbours, while betweenness centrality is based on shortest paths which is independent of the influence of neighbours. However, we found a very strong correlation between the two measures, in line with similar experiments (Valente et al. 2008; He, Meghanathan et al. 2016): the Spearman's rank correlation of participants ranked by betweenness centrality and eigenvector centrality ranged between 0.51-0.72, with a strong statistical significance ($p < 0.01$) for the period 2000-2019. We therefore use just betweenness centrality in our analysis; this is often studied and acknowledged as a measure of influence in social and complex networks, particularly built on on-

[4]https://datatracker.ietf.org/stats/meeting/overview/

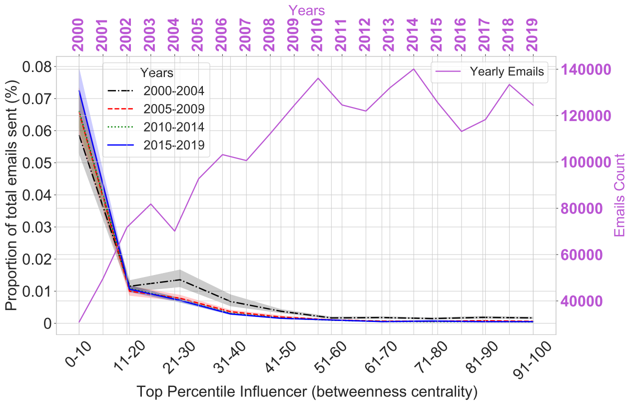

Figure 4: Proportion of emails sent (%) by participants in each period (with 95% confidence interval) according to their betweenness-centrality percentile (x-axis). Y2-axis (purple) shows yearly count of emails.

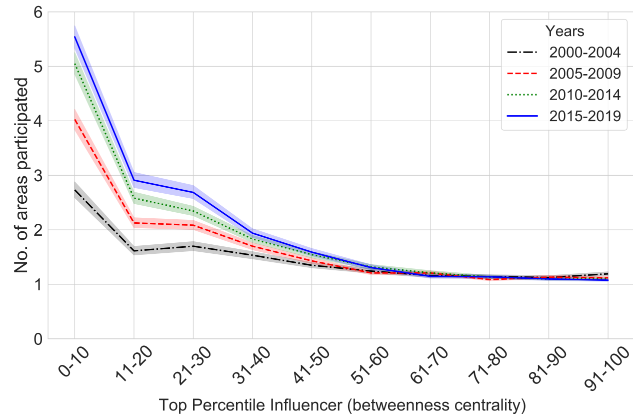

Figure 5: Mean number of areas participated in (with 95% confidence interval) according to their betweenness-centrality percentile in the x-axis, ranked from top (0-10) to bottom percentile (91-100).

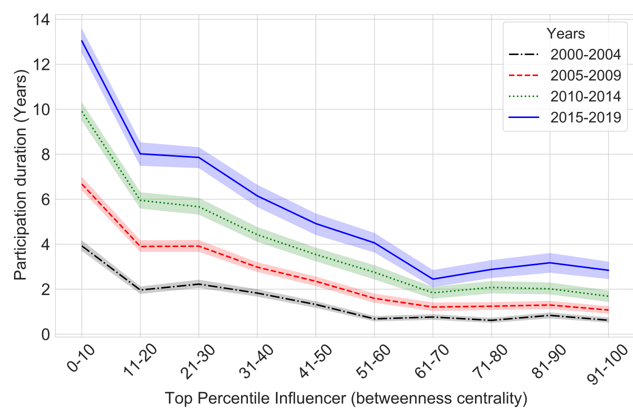

Figure 6: Mean participation duration (with 95% confidence interval) according to their betweenness-centrality percentile in the x-axis, ranked from top (0-10) to bottom percentile (91-100).

line communication (Hagen et al. 2018; Chen et al. 2012; Ghalmane et al. 2019). Figure 3 then shows the effect of removing the most influential participants (in 1% increments from most to least, moving left-to-right on the x-axis) on the size of the LCC. Worryingly, we find that removal of just 20-25% of the most influential participants causes the LCC to shrink by 90%. However, we also find that this impact has decreased over time: for instance, in 2000-2004, removing the top ∼5% influential participants, reduces the size of LCC by more than half, whereas in 2015-2019, it takes the removal of the top ∼15% of participants to have the same effect. This shows that the community has become *more* cohesive and resilient over time, and the IETF can now sustain a larger amount of churn while maintaining a well-connected social graph.

3.2 Behaviour of Influential Participants

We now characterise the behaviour of the most influential participants, in terms of the volume of emails sent, length of time active within the community, and topics discussed.

Email volume — Figure 4 shows that each participant in the top 10% most influential participants sends on average around 0.05%-0.08% of total emails in a given period. Collectively, the top 10% most influential participants account for 43.75%, on average, of the total emails, a substantially larger proportion than the others. This dynamic seems stable over time, making differences even more acute. At the same time, the overall number of emails sent increases up to 2010 and remains roughly stable from then onward. Along with the results from §3.1, this shows a worrying, if slowly improving, dependence of the IETF on a small number of influential participants.

Cross-area review — In addition to sending more emails, we test if influential participants engage with different parts of the community more. The work of the IETF is divided into several areas (e.g., Applications & Real-time, Security, and Routing). This eases administration, but might also act as a barrier to broader discussion and review. For example, WebRTC standards developed in the Applications and Real-time area might contain elements that could benefit from expertise from the Security area, but participants in one area might not review work in another.

Figure 5 shows the mean number of areas where IETF participants are active (derived from the mailing lists they use). We see that influential participants engage in more areas of the IETF, on average, and that cross-area engagement has improved over time, indicating that the community itself has matured. This indicates that influential participants benefit the IETF in enabling cross-area discussion and review, and can bridge administrative divisions.

Participation duration — We next consider for how long participants remain associated with the IETF (measured as the delta between first and last emails sent). Figure 6 shows the mean participation duration distribution for participants, ranked by influence. Participation duration increases over time, with the most influential participants typically being

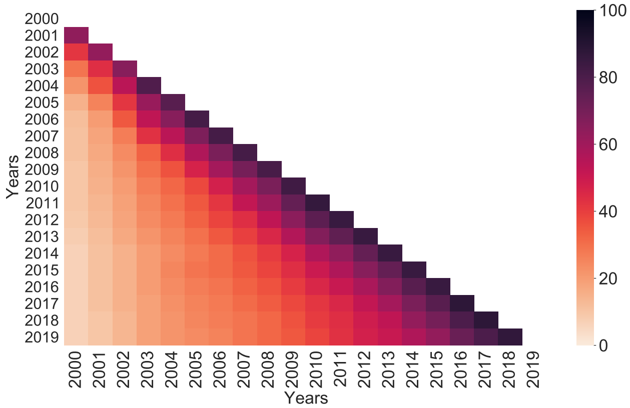

Figure 7: Yearly overlap (%) of top 10% influencers.

those who have been active for longer. While this might be expected, it might be a good sign for the IETF community in that it implies that the most influential participants are also the most experienced.

We extend this observation by considering what happens to participants once they become influential. Figure 7 shows what proportion of the top 10% influential participants in any given year (*y-axis*) was also among the top 10% influential participants in any other year (*x-axis*). This shows that a significant majority of participants that become influential, continue to be influential for a number of years: the top 10% are influential for at least 6-7 years on average. Further, Figure 8 shows that breaking into the top 10% of influencers requires an increasing number of years of participation. This may be beneficial, showing that the IETF is maturing and is capable of retaining influential and experienced participants, but it may also point to an increasingly ossified structure that is not welcoming to newcomers.

Topics of Discussion — The topics discussed on WG mailing lists are a good indicator of the focus of technical work, and it might be expected that influential participants will set the direction of that work. To explore this, we use the Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003; Hoffman, Bach, and Blei 2010) model from `gensim` (Řehůřek and Sojka 2010) to induce 100 topics on the entire set of email texts. Each topic is a distribution over words: e.g., a *security* topic might have high probability for *cypher*, *rsa*, *auth*, and related words. We can also use the model to obtain a vector for any input text as a sparse distribution over all topics, e.g., finding that a text is comprised of 30% security and 70% video streaming topics. With this, we generate a vector for each participant in a given time period by concatenating all messages sent by that participant in the period and feeding it into the LDA model. The result can be loosely interpreted as a distribution of topics on which each participant works.

Manual inspection of the induced topics reveals some interesting trends. We find that topics related to *routing* and *email* protocols are seeing a steady decline in popularity, while topics like *streaming* and *cryptography* are becoming more prominent. This reflects wider trends in standardisation, as these efforts have adjusted to the public's increased awareness of privacy, especially in light of the Edward Snowden leaks (Farrell and Tschofenig 2014). We also find that as videoconferencing became increasingly popular, so did the WebRTC protocol that frequently underpins them.

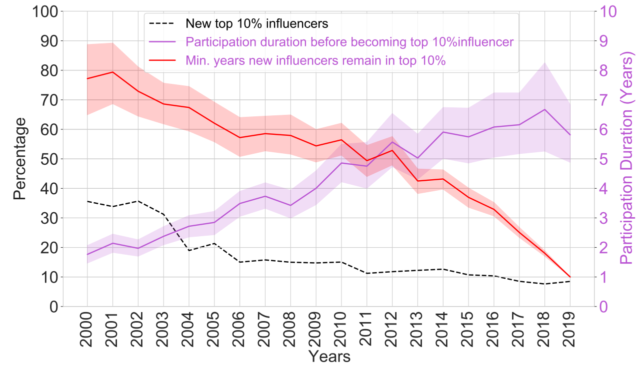

Figure 8: Percentage of new influential participants breaking into top 10% influencers (black); the average years of participation before entering top 10% (purple); and minimum # years new influencers remain in top 10% (with 95% confidence intervals).

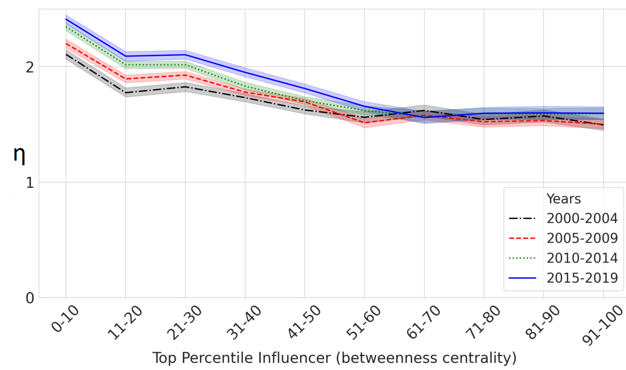

Figure 9: Mean topic entropy (with 95% confidence interval) according to their betweenness-centrality percentile in the x-axis, ranked from top (0-10) to bottom percentile (91-100).

We next measure the topical diversity of a participant by observing the entropy of their topic vector, their *topic entropy*, defined as follows:

$$\eta = -\sum_{i=1}^{|T|} p_i \log p_i \quad (1)$$

where $T = [p_1, ..., p_N]$ is a topic vector, which defines a probability distribution over N topics, each p_i is a probability of the i-th topic and $\sum_i p_i = 1$. Figure 9 shows topic entropy distributions of participants in different time periods and influence percentiles. While we initially experimented with measuring diversity by simply counting the number of topics that account for the majority ($\geq 95\%$) of a participant's topic distribution probability mass, we found that most participants engage in a relatively small number of top-

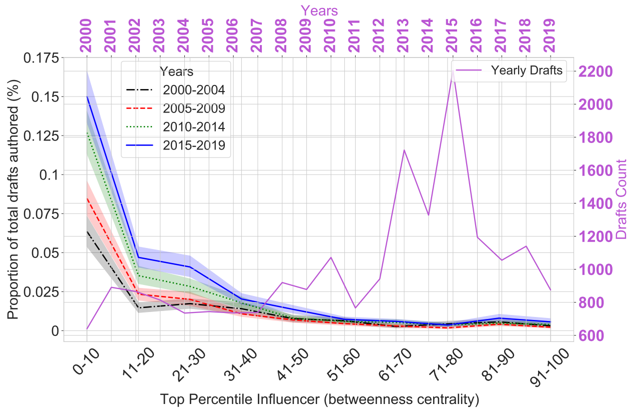

Figure 10: Average drafts authored (with 95% confidence interval) according to their betweenness-centrality percentile in the x-axis, ranked from top (0-10) to bottom percentile (91-100). Right-y axis (purple) shows yearly drafts per year.

ics, regardless of influence. However, after measuring diversity using entropy, we identify that the activity of the more influential participants tends to be more evenly spread across the topics they participate in[5]. This difference becomes more pronounced over time, implying that increased topic entropy (i.e., more evenly participating in different topics) is an increasingly salient property of influencers. This is aligned with our earlier findings on cross-area review, showing the growing number of areas in which influencers participate.

3.3 Impact of Influential Participants

While these top participants are influential within the *mailing lists*, it is unclear how this influence translates into document authorship and into gaining leadership roles within the organisation.

Draft authorship — The output of the IETF is technical standards and other documents published in the RFC series. As described in §2, RFCs are developed by WGs from a sequence of Internet-Drafts, with individuals acting as named authors. Having identified the most influential participants in the mailing list community, we can determine whether or not these same participants are also the most active authors.

Figure 10 shows the distribution of the proportion of documents authored by mailing list participants, sorted by their influence rank. We observe a growing proportion of the community involved in draft authoring, skewed towards influential participants who tend to write more drafts. During 2010-2019, each participant in the top 10 percentile, authored 0.125% to 0.175% of the total number of drafts in that period. We find that 32%-42% of the total drafts, in different years during this period, were authored by the top 10% most influential mailing list participants.

Co-authorship and email graph correlation — Figure 10 shows that influential mailing list participants tend to author more drafts than others. We next ask if they are also influencers in the draft co-authorship graph. Thus, we create a draft co-authorship graph where each author is a node, and draft co-authorship is an edge. We then measure influence with betweenness centrality of the authors in each time period. Table 1 compares the top 20% of influencers of the mailing lists with those of the co-authorship graph. We find a significant overlap between both groups, ranging from 48.2% to 67.2%. There is a significant ($p < 0.05$) positive correlation between the rankings of the overlapping members of each community: participants that are influential in the mailing lists are also likely to be influential in draft authorship.

Top 20% draft authors & all email participants					
Years	Sub-Network Size		Overlap	Spearman's r_s	
	Co-author	Email		r_s	p-value
2000-04	398	1390	48.24%	.323	4.75e-06
2005-09	427	1662	67.21%	.332	7.39e-09
2010-14	728	1639	63.05%	.299	5.73e-11
2015-19	915	1370	55.85%	.337	4.13e-15

Table 1: Overlap in the co-author and email graph.

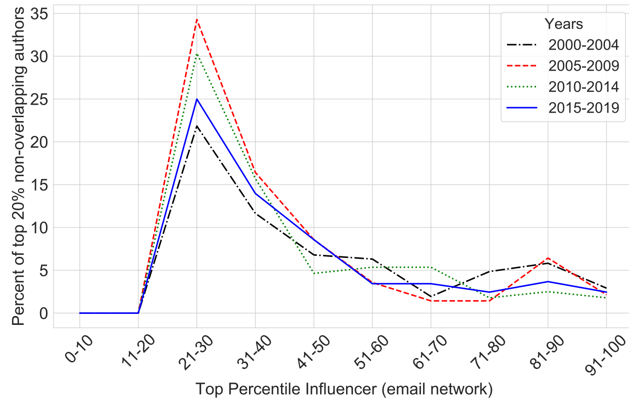

Figure 11: Percentage of Top 20% authors that are not 20% influencers according to their betweenness-centrality (email graph) percentile in the x-axis, ranked from top (0-10) to bottom percentile (91-100).

We also look at the top 20% participants from the co-authorship network who are *not* part of the top 20% influencers in the email network in Figure 11. We observe that not all the prolific authors are that engaged in the email discussion: 40%-50% of non-overlapping authors are ranked between 20th to 40th percentile of influence in the email network. These non-overlapping authors are typically more junior with respect to their participation duration.

Leadership roles — WG chairs are selected from the community, and we might expect those selected to be influential in the community. Figure 12 shows that this is indeed the case: 87.5% of WG chairs are in the top 20% of mailing list influencers and 67.4% are in the top 20% of document authors in the year before they became chairs. This also shows the impact of taking up a leadership role: influence in both

[5] An example with 3 topics: a distribution of $[0.8, 0.1, 0.1]$ is less evenly spread than $[0.3, 0.4, 0.3]$.

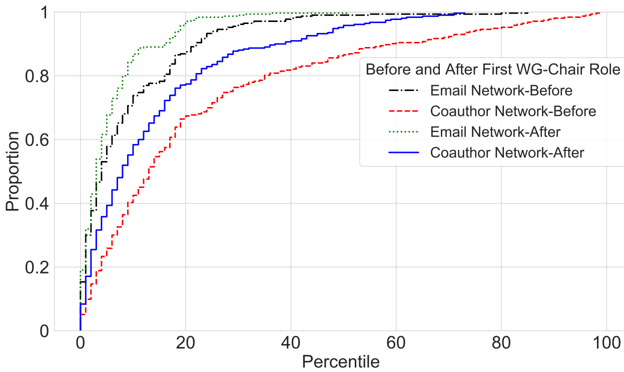

Figure 12: CDF of the percentile of betweenness-centrality of WG chairs in the email and co-authorship graph, one year before and after becoming chairs.

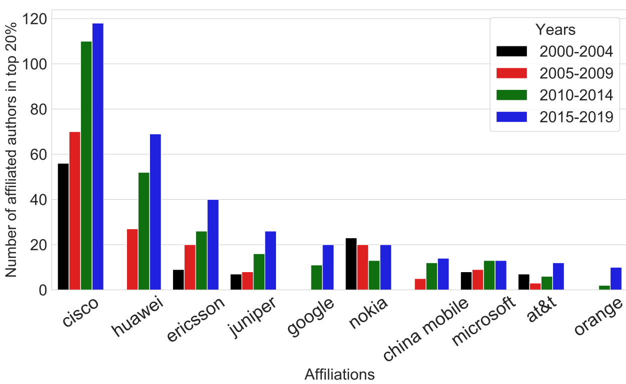

Figure 13: Authors in the most common affiliations of top 20% authors (co-authorship graph) in 2015-2019.

the mailing list and authorship communities grows in the year after participants become a WG chair for the first time.

Organisations — While participants in the IETF contribute as individuals, they are usually affiliated with an organisation.To study the potential influence of organisations, for each draft we obtain the authors' affiliations from the Datatracker. If this is not available, we use the domain name of the authors' email addresses, which we map to the relevant organisation (e.g., `@cisco.com` maps to Cisco). For generic email addresses, such as `@gmail.com`, we use the participant name if no affiliation is available.

Figure 13 shows the ten most frequent affiliated organisations of the top 20% authors in the period 2015-2019, and the number of authors affiliated with these organisations in each period. While the dominance of Cisco is clear, other organisations, such as Huawei, have gained a larger presence over time. In general, a small number of organisations employ a significant fraction of the influential participants in the standards process. If we look at the period 2015-2019, for example, out of the 915 authors ranked in the top 20% of most influential authors, 342 belong to the ten most frequently affiliated organisations, and nearly 253 are from Cisco, Huawei, Ericsson, or Juniper alone.

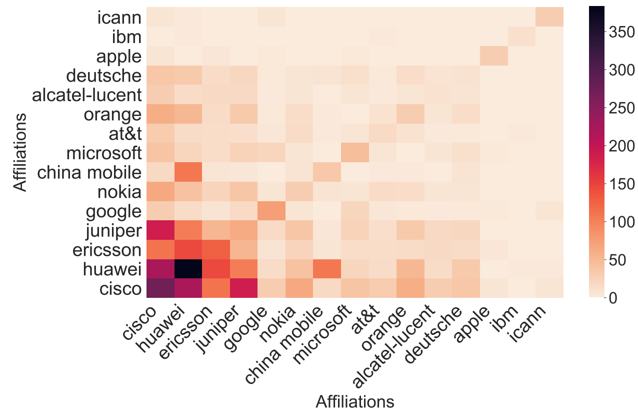

Figure 14: Number of drafts co-authored by top 15 most frequently affiliated organisations. We only include the top 20% authors (co-authorship graph) of 2015-2019.

While authors from the same organisation often co-author drafts together, collaborations between authors from different organisations are also common. Figure 14 shows the most common collaborations between authors from the top 15 most frequently affiliated organisations. Collaborations between authors at competing organisations, such as between authors from Cisco, Huawei, Ericsson, and Juniper, occur frequently. We also computed that over the years, joint collaboration (authors from multiple affiliations) has increased: 772 co-authored drafts in the period 2000-2004 vs. 3083 co-authored drafts in the period 2015-2019.

Further inspection of the collaboration trends reveals that the volume of collaborations between authors from Huawei and other organisations appears to be unaffected by the recent addition of Huawei to the U.S. Entity List (of Industry and Security 2019) as the trends in Figure 14 also hold for 2020-2021.[6]

4 Predicting Author Success

We have identified that influential authors write more drafts (§3), but are these drafts more successful? Before a standards track RFC is *published*, the corresponding draft must be *adopted* by a WG for review and development.[7] Many documents, though, fail to be adopted or published. We employ logistic regression to analyse predictability of this outcome. Using the Datatracker we compile a dataset for the *adoption* of a draft by a WG, comprising 11632 drafts. The positive and negative labels correspond to *adopted* and *not adopted*, respectively.

4.1 Methodology

Features — We compute a set of features that may impact adoption or publication for each draft, most of them motivated by the characteristics of influential participants, including the following *feature groups*: (1) *Text*, we derive a

[6] https://mailarchive.ietf.org/arch/msg/ietf-announce/0ywjgSS4LlO0DaWDoLJLRHxJdUk/

[7] There are occasional exceptions, known as Area Director sponsored drafts, but these too rare (∼3%) to affect our analysis.

term frequency – inverse document frequency (Salton and McGill 1983) weighted bag of words vector representation of all email text of messages that either explicitly mention a draft or are part of a thread where the subject line explicitly mentions a draft; (2) Communication patterns (i.e., volumes of incoming/outgoing communication between draft authors and IETF participants of varying experience levels), following the scheme proposed by McQuistin et al. (2021); (3) Betweenness *centrality* for authors; (4) *Number of emails* sent by authors; (5) *Number of drafts* submitted by authors; (6) *Length of participation* duration of authors; (7) *Number of authors*; (8) *Number of areas* in which the author is active; (9) *Proportion of top authors* in the top 10 and top 20 percent of influential participants; (10) *Topic entropy* (η) scores for authors; (11) *Author affiliations* for each of the 20 most influential organizations (20 boolean features); (12) *Number of mailing lists* that each author participates in, (13) *Number of years active*.

For numerical features that relate to each author individually instead of the draft as a whole (e.g., *Number of emails sent*, *Centrality*, or *Number of years active* is defined for each author separately) in three variants: (*i*) for the least influential author, (*ii*) for the most influential author, and (*iii*) as the average value across all authors of the draft. We compute individual author features using information from the 5 year period prior to draft submission.

Regarding the text based features, we found that the models tend to assign high weights to words such as surnames of active contributors, prolific WG names, and common technical terms. While this does make the models perform slightly better, such terms should not be relevant for solving the task as they do not model the relevant part of the conversation. To explore how models would behave in a scenario wherein such words were not available, we construct a domain-specific stop word list, consisting of: (*i*) all last names from the Datatracker, (*ii*) all WG names from the Datatracker, and (*iii*) technological jargon terms obtained from the web[8]. We remove from this list any terms appearing in top 5K English terms (e.g., sometimes people's names can have identical surface forms as some of the terms relevant to organisational activities). We will refer to this variant as *Text (S)*.

Experimental setup — To gain empirical insights into which features provide better prediction results, models are run using a *feature group alone* variant and a variant combining each feature group with the *Text* features. This approach was motivated by the finding that *Text*, even though conceptually quite different from the rest of the feature set and not our main focus, does provide surprisingly strong results on its own. Thus, we wanted to empirically investigate in more detail how well it complements the rest of the graph-based features. Moreover, we compare all models to a baseline which simply assigns all examples to the majority class.

We split the data into *training* (70%), *development* (10%), and *test* (20%) subsets. As a scoring function, we use the F1 score macro averaged across the positive and negative classes (but we also report area under the curve (AUC), precision, and recall). We train the models on the training set, optimise hyper-parameters on the development set, and report final scores on the test set. The final scores are those obtained by the model variant that fared best on the development set. We use logistic regression implemented in `scikit-learn` (Pedregosa et al. 2011), given that it is widely used and well interpretable. While we do not perform explicit feature selection, we do have implicit feature selection through L1 regularisaton. We consider the regularisation strength a hyper-parameter and consider values from $[2^{-7}, 2^{-6}, ..., 2^{6}]$ on the development set. As our data set is quite imbalanced (in a roughly 4:1 ratio, 17% is *adopted*), we use different class weights during optimization to counteract this. We employ a non-parametric random shuffling test (Yeh 2000) to check statistical significance of score differences against the baseline. Table 2 summarises prediction results, which will be discussed in the next section.

Next, we wanted to confirm these results at the level of individual features. To this end, we inspect the statistically significant coefficients learned by the model corresponding to each individual feature. For this experiment, we use a slightly different setup. We do not consider the *Text* features[9]. We begin by first applying the Variance Inflation Factor (VIF) to exclude all features with $VIF > 5$. This, to an extent, mitigates the collinearity that we know exists as some features are by construction highly correlated. We then standardise each remaining feature and then fit a Logistic Regression model from the `statsmodels` package (Seabold and Perktold 2010) on the entire data set. Table 3 presents the results of this experiment.

4.2 Results

Predicting adoption — Scores from Table 2 reveal several interesting trends. Most importantly, the best performing feature groups are *Centrality* and *Proportion in top percentiles*. They consistently stand out (AUC 0.655 and AUC 0.663, respectively) among all feature groups in terms of performance. This observation holds in terms of both AUC and F1 scores regardless of whether or not the text features are included. This suggests that features related to participants' influence in the network affect prediction. Some other well performing features are *Email count*, and *N. mailing lists*, thereby highlighting the importance of the extent of participants' engagement within the community. This engagement is also related to influence, as we have shown that influential participants are more engaged (§3.2). Statistical analysis of features better reflect these empirical implications.

Combining the the text features with most individual feature groups yields performance gains in both AUC and F1. This persists when combining the text with the combination of all other feature groups. Finally, the (S) variants of text features, that use the domain-specific stopword list, provide slightly worse standalone performance and benefits when

[8]https://www.computerhope.com/jargon.htm, we used a union of the *Internet terms*, *Network terms*, and *Security terms* categories.

[9]There are thousands of text features (terms), making it prohibitively complex to include them. Moreover, they are not central to our research questions about the IETF community.

	$\neg TXT$				TXT			
	AUC	F1	P	R	AUC	F1	P	R
Baseline	.500	.445	.401	.500	.500	.445	.401	.500
Comm. patterns	.644	.583	.566	.601	.723	.626	.609	.644
Centrality	.655	.577	.573	.581	.717	.639	.625	.653
Email count	.632	.566	.555	.577	.718	.650	.639	.662
N. years active	.608	.569	.555	.584	.702	.610	.594	.626
Number of authors	.578	.536	.529	.544	.697	.631	.621	.641
Proportion in top	.663	.594	.580	.609	.728	.634	.614	.655
Draft count	.500	.445	.401	.500	.696	.630	.624	.636
N. Areas	.634	.572	.557	.588	.721	.639	.622	.657
N. Mailing lists	.634	.570	.557	.584	.718	.641	.630	.653
Affiliations	.605	.584	.576	.592	.708	.626	.609	.643
Topic entropy (η)	.622	.563	.550	.577	.715	.642	.628	.657
Text	-	-	-	-	.681	.624	.622	.627
Text (S)	-	-	-	-	.667	.598	.593	.614
All feats	.692	.624	.600	.651	.744	.645	.625	.666
All feats (S)	.692	.624	.600	.651	.726	.636	.627	.664

Table 2: Results for predicting adoption. Each row presents the scores of a model using the corresponding feature group either alone ($\neg TXT$) or combined with the Text features (TXT). *All feats* denotes all features except the text. Rows labeled with (S) use the *Text (S)* variant of the text features.

Feature	Weight	P-value
Centrality of most influential	0.1691	0.000
Centrality of least influential	-0.0698	0.048
Proportion of authors in top 10	0.1520	0.000
Has a Cisco author	0.1301	0.000
Has an AT&T author	0.1003	0.000
Has a Juniper author	0.0738	0.000
Has a China Mobile author	-0.0697	0.000
Has an Alcatel-Lucent author	0.0474	0.022
Topic entropy max	0.1728	0.000
Topic entropy min	-0.1135	0.002

Table 3: Weights of the model that are statistically significant at $p \leq 0.05$ grouped by similarity of features – Author influence measures vs. Affiliation features vs. Topic entropy (η).

combined with the rest of the features. As detailed analysis of text features is out of scope for this paper, we leave a more in depth investigation of these phenomena for future work.

While all model variants are statistically significantly better than the baseline, the absolute scores are still not very high. This indicates there is still considerable variance in the data that the current model and feature set cannot capture. As a future work, we intend to refine the contextual text features from the email conversations, possibly using some of the recently developed discourse-aware neural language models (Gu, Yoo, and Ha 2021).

Statistical analysis — To confirm empirical implications from the previous section and provide a more in-depth look at the features, we perform statistical analysis summarised in Table 3. The most interesting insight is that the coefficients corresponding to the *Centrality of the most influential author* and *Proportion of authors in the top 10 percentile in influence*, are, indeed, positive and among the largest in absolute value. This highlights that among the people who are influential in the email networks, there is a substantial number of individuals who are proficient in contributing to successful drafts.

Another interesting observation is that author affiliations (which are also indirectly connected to influential participants, as demonstrated in §3.3, Figure 13) are an important feature group, particularly affiliations such as *Cisco* and *AT&T*. Moreover, topic entropy is also an important feature group, which was also shown to be related to influential participants in our analysis from §3.3, Figure 9.

As regards the initial question in this chapter about whether influential participants write *successful* drafts, we have confirmed that the answer is positive, i.e., we have shown that having authors that are high in influence scores or are affiliated with high influence organisations does indeed increase the chances of a draft being adopted.

4.3 Discussion

We now answer the research questions (RQ) in §1 with the analysis conducted so far.

RQ (i) — We conclude that the IETF is still reasonably centralised, but that this has improved over the years. However, removing around 20%-25% most influential participants still fragments the entire network (§3.1, Figure 3). The most influential people are the ones with highest degree of engagements within the community who also tend to be more involved in draft authoring activities (§3.2, Figure 10). However, we also observe that over the years more people (even the ones less influential in the email network) are getting involved in draft authoring activities (Figure 10). This shows that these activities, while centralised to a certain degree, are still open to contributions from the wider IETF community.

RQ (ii) — We observe that the influential participants show a much higher level of engagement with the community (§3.2). A considerable proportion of total emails is sent by the top 10% of influential participants (Figure 4). Compared to the rest of participants, they are active in more areas (Figure 5), are active within the IETF for a longer period of time (Figure 6), and participate in a more diverse set of topics (Figure 9). Finally, their influence extends from the mailing lists to other activities such as draft authorship.

RQ (iii) — A significant overlap and correlation is observed between influential authors from co-authorship network and email networks (§3.3, Table 1). This shows that a large set of authors exhibit an ability to co-author drafts as well as engage with participants on the email networks, thereby translating their influence from email networks to co-authorship and vice versa. The statistical analysis shows that influence of draft authors in the email networks does impact the possibility of a draft getting adopted (§4.2, Table 2 & 3). This might hint at the ability of participants who hold a domain expertise to be able to engage better with the community. Several WG chairs are already in the top percentile influential category in both the email and co-authorship networks before taking up these leadership roles, which further elevates after taking up such leadership roles (Figure 12).

RQ (iv) — A considerable portion of influential participants ($\sim$30%) are affiliated to one of the more prominent organisations (e.g., *Cisco* or *Ericsson*) (§3.3, Figure 13). Participants from different organisations do considerably collaborate (Figure 14). Moreover, the level of such collaboration is found to have increased with time. Most importantly the statistical analysis shows that being affiliated with a prominent organisation positively impacts the chances of a draft in getting adopted by a WG, thereby directly driving the innovation process (Table 2 & 3).

5 Related Work

Influence in social graphs has been studied in many domains including Twitter (Cha et al. 2010; Weitzel, Quaresma, and de Oliveira 2012; Anger and Kittl 2011), Instagram (Zarei et al. 2020) and decentralised social networks (Bin Zia et al. 2022; Hassan et al. 2021). For example, (Ye and Wu 2010) measured the impact of follower influence on message propagation. Others have focused on devising metrics to capture influence: similarly to (Kourtellis et al. 2013; Weitzel, Quaresma, and de Oliveira 2012; Solé-Ribalta et al. 2014), we use betweenness centrality as a metric of influence.

Online collaborative communities such as Wikimedia (Bosu and Carver 2014a) and Open Stack have also been studied for understanding communities, collaboration, strategy making, and organisational structures, through mailing lists and platform interactions (Dobusch and Kapeller 2018; Bosu and Carver 2014a; Zhang et al. 2020). Bosu and Carver (2014b) highlight that contributors with much higher reputation are successful in seeking reviews from the community in a much shorter time span and are more likely to get their suggestions accepted. This is well aligned with most of our findings related to the impact of influential people on the Internet-Draft adoption process. Zhang et al. (2020) observe that in the open source software (OSS) ecosystem, several participating organisations (firms or companies) may engage in intentional or passive collaborations, or they may contribute in an isolated way. They find that an organisation's influence in the collaboration network is positively correlated with its scale of contribution within the ecosystem. In some ways, this is similar to our observation related to prominent affiliated organisations in the co-authorship network and their ability to collaborate and produce drafts.

Other studies focusing on mailing list data include OSS mailing lists (Bird et al. 2006; Rigby and Hassan 2007), and the popular Enron dataset (Klimt and Yang 2004). While the Enron data set is often used for classification of emails into categories (Madjarov et al. 2012), there are also several studies of organisation interaction patterns on it. Namely, (Tang et al. 2008; Kossinets, Kleinberg, and Watts 2008) are concerned with developing methodologies for studying how communication evolves over time. In contrast, Diesner, Frantz, and Carley (2005) focus on providing insights more than on developing a methodology, revealing that organisational structure is reflected in email communication patterns.

To the best of our knowledge, this is the first social graph study of the Internet standards community, although we build on a number of prior studies and tools (Benthall 2015). Niedermayer et al. (2017) investigate participant behaviour in IETF mailing lists, and examine RFC authorship, classifying the Twitter popularity of recent RFCs (Niedermayer et al. 2016). Our modelling work mainly builds on two prior studies. McQuistin et al. (2021) characterise the IETF's authorship community, analysing the factors that lead to the successful deployment of protocol standards. Similarly, Nikkhah et al. (2017) analyse RFC adoption, but through the characteristics of the RFCs themselves. In contrast, we provide a more extensive analysis of the social graphs, augmented with IETF Datatracker data, providing context for how influence translates to other IETF activities.

6 Conclusions

This paper characterised the social graph of the Internet standards community. We observed a core group of participants (§3.1), revealing that the most influential participants send more messages, to more IETF areas, participate for longer, and discuss a wider range of topics (§3.2). We further demonstrated through correlation analysis in §3.3 that influence in the mailing list community translates into influence in draft authorship network and WG leadership. Finally, we showed that the social graph properties of IETF participants positively impact the success of their drafts, leading to greater chances of WG adoption.

We have revealed a community that is a product of its open, consensus-driven approach to protocol standardisation. Despite significant growth over time, and influential participants remaining an important driving force, the community has grown better connected and less fragmented, growing in resilience to departure of influential participants.

This study could support further research in the areas of online collaborative systems, social science, and the role of various stakeholders in the development of Internet standards. For instance, an interesting research avenue would be tracking the emergence of new technologies, and exploring whether they are a consequence of new communities forming within the IETF social graph, curating innovative ideas. Another could be to explore interventions that can ensure that the correct participants engage early with relevant documents and WGs. This would help participants to more easily navigate the growing IETF social graph, and ultimately improve the process and resulting Internet standards.

Acknowledgements

This work is supported by the EPSRC under grants, SODESTRAM: Streamlining Social Decision Making for Improved Internet Standards (EP/S033564/1, EP/S036075/1), and AP4L: Adaptive PETs to Protect emPower People during Life Transitions (EP/W032473/1). More information about this project and related papers are available from https://sodestream.github.io/.

References

Anger, I.; and Kittl, C. 2011. Measuring influence on Twitter. In *Proceedings of the 11th international conference on knowledge management and knowledge technologies*, 1–4. Graz, Austria: ACM.

Arkko, J.; Cooper, A.; Pauly, T.; and Perkins, C. S. 2020. Evolving the Internet Through COVID-19 and Beyond. CircleID.

Benthall, S. 2015. Testing Generative Models of Online Collaboration with BigBang. In *Proceedings of the 14th Python in Science Conference*, 182–189. Austin, Texas: SciPy Organizers.

Bin Zia, H.; Raman, A.; Castro, I.; Anaobi, I.; Cristofaro, E. D.; Sastry, N.; and Tyson, G. 2022. Toxicity in the Decentralized Web and the Potential for Model Sharing. In *Proceedings of ACM SIGMETRICS*.

Bird, C.; Gourley, A.; Devanbu, P.; Gertz, M.; and Swaminathan, A. 2006. Mining email social networks. In *Proceedings of the 2006 international workshop on Mining software repositories*, 137–143.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3: 993–1022.

Bosu, A.; and Carver, J. C. 2014a. How do social interaction networks influence peer impressions formation? a case study. In *IFIP International Conference on Open Source Systems*, 31–40. Springer.

Bosu, A.; and Carver, J. C. 2014b. Impact of developer reputation on code review outcomes in oss projects: An empirical investigation. In *Proceedings of the 8th ACM/IEEE international symposium on empirical software engineering and measurement*, 1–10.

Böttger, T.; Antichi, G.; Fernandes, E. L.; di Lallo, R.; Bruyere, M.; Uhlig, S.; Tyson, G.; and Castro, I. 2018. Shaping the internet: 10 years of IXP growth. *arXiv preprint arXiv:1810.10963*.

Böttger, T.; Cuadrado, F.; Antichi, G.; Fernandes, E. L.; Tyson, G.; Castro, I.; and Uhlig, S. 2019. An Empirical Study of the Cost of DNS-over-HTTPS. In *Proceedings of IMC*, 15–21.

Cha, M.; Haddadi, H.; Benevenuto, F.; and Gummadi, K. 2010. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the international conference on web and social media*, 8. Washington, DC, USA: Association for the Advancement of Artificial Intelligence.

Chen, D.; Lü, L.; Shang, M.-S.; Zhang, Y.-C.; and Zhou, T. 2012. Identifying influential nodes in complex networks. *Physica a: Statistical mechanics and its applications*, 391(4): 1777–1787.

Diesner, J.; Frantz, T. L.; and Carley, K. M. 2005. Communication networks from the Enron email corpus "It's always about the people. Enron is no different". *Computational & Mathematical Organization Theory*, 11(3): 201–228.

Dobusch, L.; and Kapeller, J. 2018. Open strategy-making with crowds and communities: Comparing Wikimedia and Creative Commons. *Long Range Planning*, 51(4): 561–579.

Farrell, S.; and Tschofenig, H. 2014. Pervasive Monitoring Is an Attack. Internet Engineering Task Force. RFC 7258.

Flanagan, H. 2019. Fifty Years of RFCs. RFC Editor. RFC 8700.

Ghalmane, Z.; El Hassouni, M.; Cherifi, C.; and Cherifi, H. 2019. Centrality in modular networks. *EPJ Data Science*, 8(1): 15.

Gu, X.; Yoo, K. M.; and Ha, J.-W. 2021. DialogBERT: Discourse-Aware Response Generation via Learning to Recover and Rank Utterances. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 12911–12919.

Hagen, L.; Keller, T.; Neely, S.; DePaula, N.; and Robert-Cooperman, C. 2018. Crisis communications in the age of social media: A network analysis of Zika-related tweets. *Social Science Computer Review*, 36(5): 523–541.

Hassan, A. I.; Raman, A.; Castro, I.; Zia, H. B.; De Cristofaro, E.; Sastry, N.; and Tyson, G. 2021. Exploring Content Moderation in the Decentralised Web: The Pleroma Case. In *Proceedings of ACM CoNEXT*.

He, X.; Meghanathan, N.; et al. 2016. Correlation of eigenvector centrality to other centrality measures: random, small-world and real-world networks. *NeCoM, CSITEC*, 09–18.

Hoffman, M.; Bach, F.; and Blei, D. 2010. Online learning for latent dirichlet allocation. *advances in neural information processing systems*, 23: 856–864.

Jennings, C.; Hardie, T.; and Westerlund, M. 2013. Real-Time Communications for the Web. *IEEE Communications Magazine*, 51(4): 20–26.

Klimt, B.; and Yang, Y. 2004. The enron corpus: A new dataset for email classification research. In *European Conference on Machine Learning*, 217–226. Springer.

Kossinets, G.; Kleinberg, J.; and Watts, D. 2008. The structure of information pathways in a social communication network. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 435–443.

Kourtellis, N.; Alahakoon, T.; Simha, R.; Iamnitchi, A.; and Tripathi, R. 2013. Identifying high betweenness centrality nodes in large social networks. *Social Network Analysis and Mining*, 3(4): 899–914.

Madjarov, G.; Kocev, D.; Gjorgjevikj, D.; and Džeroski, S. 2012. An extensive experimental comparison of methods for multi-label learning. *Pattern recognition*, 45(9): 3084–3104.

McQuistin, S.; Karan, M.; Khare, P.; Perkins, C.; Tyson, G.; Purver, M.; Healey, P.; Iqbal, W.; Qadir, J.; and Castro, I. 2021. Characterising the IETF Through the Lens of RFC Deployment. In *Proceedings of IMC 2021*. Online: ACM.

Niedermayer, H.; Raumer, D.; Schwellnus, N.; Cordeiro, E.; and Carle, G. 2016. An analysis of IETF activities using mailing lists and social media. In *International Conference on Internet Science*, 218–230. Florence, Italy: Springer.

Niedermayer, H.; Schwellnus, N.; Raumer, D.; Cordeiro, E.; and Carle, G. 2017. Information Mining from Public Mailing Lists: A Case Study on IETF Mailing Lists. In *International Conference on Internet Science*, 301–309. Thessaloniki, Greece: Springer.

Nikkhah, M.; Mangal, A.; Dovrolis, C.; and Guérin, R. 2017. A statistical exploration of protocol adoption. *IEEE/ACM Transactions on Networking*, 25(5): 2858–2871.

of Industry, U. B.; and Security, C. 2019. Addition of Entities to the Entity List. Rule 84 FR 22961.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830.

Řehůřek, R.; and Sojka, P. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA.

Rigby, P. C.; and Hassan, A. E. 2007. What can oss mailing lists tell us? a preliminary psychometric text analysis of the apache developer mailing list. In *Fourth International Workshop on Mining Software Repositories (MSR'07: ICSE Workshops 2007)*, 23–23. IEEE.

Salton, G.; and McGill, M. J. 1983. *Introduction to modern information retrieval*. New York: McGraw-Hill.

Seabold, S.; and Perktold, J. 2010. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

Solé-Ribalta, A.; De Domenico, M.; Gómez, S.; and Arenas, A. 2014. Centrality rankings in multiplex networks. In *Proceedings of the conference on Web science*, 149–155. Bloomington, IA, USA: ACM.

Tang, L.; Liu, H.; Zhang, J.; and Nazeri, Z. 2008. Community evolution in dynamic multi-mode networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 677–685.

Valente, T. W.; Coronges, K.; Lakon, C.; and Costenbader, E. 2008. How correlated are network centrality measures? *Connections (Toronto, Ont.)*, 28(1): 16.

Weitzel, L.; Quaresma, P.; and de Oliveira, J. P. M. 2012. Measuring node importance on twitter microblogging. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, 1–7. Craiova, Romania: ACM.

Ye, S.; and Wu, S. F. 2010. Measuring message propagation and social influence on Twitter. com. In *International conference on social informatics*, 216–231. Laxenburg, Austria: Springer.

Yeh, A. 2000. More accurate tests for the statistical significance of result differences. In *Proceedings of the 18th International Conference on Computational Linguistics*. Saarbrücken, Germany: Association for Computational Linguistics.

Zarei, K.; Ibosiola, D.; Farahbakhsh, R.; Gilani, Z.; Garimella, K.; Crespi, N.; and Tyson, G. 2020. Characterising and detecting sponsored influencer posts on Instagram. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 327–331. Online: IEEE/ACM.

Zhang, Y.; Zhou, M.; Stol, K.-J.; Wu, J.; and Jin, Z. 2020. How do companies collaborate in open source ecosystems? an empirical study of openstack. In *2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE)*, 1196–1208. IEEE.