

APSA Preprints

Version 1

April 10th, 2025

This content is an early research output and has not been peer-reviewed at the time of posting

Sara Morrell and Christoph Riedl

Northeastern University, NULab for Digital Humanities and Computational Social Science

Human Rights in the Internet Engineering Task Force: Using Large Language Models and Network Analysis to Explore Norm Entrepreneurship

Abstract

Introduction

The extent to which human rights considerations are integrated into Internet technical standards shapes how people exercise their rights online. Understanding the policy implications of engineering decisions helps support the Internet as a right-enabling environment. Prior literature has provided valuable qualitative evidence on how values and rights are (or are not) considered in technical processes and organizations. However, few examine these processes at a large scale.

Methods

We analyze all emails sent by Internet Engineering Task Force (IETF) and Internet Research Task Force (IRTF) participants recorded in the public IETF Mail Archive containing the term “human rights” sent from 1992 to early 2025 ($n = 3,377$). Emails are annotated by a large language model to indicate the author’s position on integrating human rights. They are then analyzed using time series, regression, and network analysis.

Results

This analysis reveals how processes of norm entrepreneurship regarding human rights considerations in Internet standards unfolded over the course of 33 years. We find evidence of a new cohort of norm entrepreneurs emerging around 2013. Based on our network analysis, it appears that these entrepreneurs had limited success in diffusing norms to the broader IETF community.

Conclusion

This study provides insight into how processes of norm entrepreneurship and diffusion evolved in the IETF and IRTF community. It builds on prior literature using email data to map normative evolution with a novel dataset from the IETF Mail Archives. Understanding these informal Internet-mediated interactions can help explain policy-making outcomes in Internet governance.

Keywords

Norms, Large Language Model, Network Analysis, Human Rights, Internet Engineering Task Force

Introduction

The integration of human rights into Internet infrastructure is controversial. Some in the technical community responsible for building and maintaining the Internet argue that technology is neutral, while others argue that engineering decisions impact human rights (Krishnamurthy, 2019). This study analyzes the online processes surrounding the integration of human rights considerations into technical standards using theories from Science and Technology Studies. It applies time series, regression, and network analysis to a novel dataset of Internet Engineering Task Force and Internet Research Task Force emails pertaining to integrating human rights considerations into technical standards. This analysis reveals how processes of norm entrepreneurship and diffusion evolved in the Internet Engineering Task Force and Internet Research Task Force community.

The Internet Engineering Task Force (IETF) is a standards development organization (SDO) for the Internet. SDOs create and refine the standards that allow different entities on the Internet to connect; developing and maintaining the Internet as a communications network. According to the IETF's mission statement: "the overall goal of the IETF is to make the Internet work better" (Internet Engineering Task Force a, n.d.). Cath and Floridi (2017, p. 2) write that the IETF is "one of the most influential SDOs, capable of making the most significant difference in the architectural implementation of values or human rights." While the ability of SDOs in general to influence human rights is questioned by some scholars (Mueller and Badiei, 2019), the IETF is a key player in this space. Therefore, this study will focus on normative development in the IETF.

Individuals voluntarily participate in the IETF in their personal capacity. According to its website, "The IETF welcomes all interested individuals and participants come from all over the world and from many different parts of the Internet industry" (Internet Engineering Task Force a, n.d.). Participants develop technical standards through an open process of offline (Andrade et al, 2018) and online collaboration. According to Cath and Floridi (2017), at the time of their study, IETF participants were mostly white, male, and Western and worked for large companies. Khare

et al (2022) explore the relationship between participant influence and draft success through social network analysis of emails and document co-authorship. They find that participants who are influential in mailing lists also tend to find more success in authoring document drafts. Furthermore, they find that participants who are affiliated with more prominent organizations have a higher likelihood of their draft being adopted.

The Internet Research Task Force (IRTF) works in parallel with the IETF on “longer term research issues related to the Internet” (Internet Research Task Force, n.d.). It is composed of 16 research groups focused on Internet technology, applications, architecture, and protocol-related topics. The Human Rights Protocol Considerations Research Group is particularly relevant to this study. It researches “whether standards and protocols can enable, strengthen or threaten human rights, as defined in the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR), specifically, but not limited to the right to freedom of expression and the right to freedom of assembly” (Internet Research Task Force b, n.d.) While IETF participants tend to come from industry, many IRTF participants are members of the academic research community or non-governmental organizations.

Both of these organizations are committed to making their work processes open. The IETF Mail Archive contains emails documenting the processes through which both IETF and IRTF participants collaborate on and debate research and technical issues (Internet Engineering Task Force b, n.d.). The resulting documents are also available on the IETF Datatracker (Internet Engineering Task Force c, n.d.). Most of these documents are Internet standards and technical guidelines referred to as RFCs (Requests for Comments). This data will be used in this study to examine the normative development surrounding human rights in the IETF and IRTF.

Literature Review

As the importance of the Internet for daily life increases, the nexus between human rights and technical standards is coming under increasing scrutiny. Cath and Floridi (2017) identify several major instances of international bodies affirming the link between the Internet and human rights. These bodies include the United Nations, the World Summit on the Information Society, and the Netmundial Initiative. Mueller and Badiei (2019) summarize academic arguments pertaining to human rights and standards development into three categories: “Code is Law” (Mueller and Badiei, 2019, p. 4), “Architecture as Mediator of Struggles Over Rights” (Mueller and Badiei, 2019, p. 6), and “Values in Design” (Mueller and Badiei, 2019, p. 7). Each of these perspectives describes a proposed relationship between standards and rights. The “Code is Law” (Mueller and Badiei, 2019, p. 4) perspective emphasizes the importance of Internet architecture and infrastructure in shaping behavior and society. The architecture as a mediator (Mueller and Badiei, 2019, p. 6) perspective also takes into account how the implementation of a technology impacts human rights. It provides for the agency of policy makers and other political actors to influence how technology is used to regulate behavior. The “Values in Design” perspective (Mueller and Badiei, 2019, p. 7-8) holds that values should be integrated into design processes as

part of a system's constraints and functionalities. While Mueller and Badiei (2019) express skepticism over whether these perspectives are realistic, other scholars have illustrated their relevance to the case of the IETF. Cath and Floridi (2016) take the position that Internet architecture mediates rights and practices. They directly address "the debate on SDOs' ethical responsibility to bring their work in line with human rights" (Cath and Floridi, 2016, p. 1). They advance three arguments: the work of SDOs is political, the work of the IETF should support human rights, and it should do so by implementing a "responsibility-by-design approach" (Cath and Floridi, 2016, p. 1). Cath and Floridi clarify that this approach is distinct from a "human-rights-by-design" approach which could increase the risk of Internet fragmentation.

Integrating human rights considerations into the work of the IETF, however, has faced resistance. This resistance primarily takes the form of conflict over norms around the role and characteristics of the Internet and the role of engineers responsible for developing and maintaining it. Following Finnemore and Sikkink (1998, p. 891), we consider norms "a standard of appropriate behavior for actors with a given identity...". According to Radu et al. (2021, p. 3), "Entrepreneurs - be it state or non-state actors - define and establish standards of behaviour that become influential within the community mostly by adhering to relevant networks". These norms then diffuse further, in what Finnemore and Sikkink (1998, p. 895) have termed a 'norm cascade'. Cath and Floridi (2016, p. 3) identify the four base architectural principles of the Internet as "openness, interoperability, redundancy, and the end-to-end principle." They also found that engineers tend to argue for the neutrality of technology. The importance of these norms to the broader Internet technical community is reaffirmed in more recent work by ten Oever (2023). ten Oever (2023) develops the concept of 'infrastructural norms' on the basis of a case study of norm conflict in Internet routing. They situate this work in the broader "transnational Internet governance regime complex", or "polycentric Internet governance network" (ten Oever, 2023, p. 184-185). They find that established infrastructural norms were used to resist new norms related to human rights and data protection. The primary infrastructural norm they found was 'voluntary interconnection'. This norm has economic, institutional, technical, and supranational roots that keep it from being marginalized by other norms. Human rights advocates in technical communities have faced entrenched resistance, not necessarily due to disputes over the importance of human rights in general, but due to concerns that integrating human rights considerations into technical work would jeopardize other deeply held values prescribing what the Internet should be and the proper role of engineers.

Academic literature has documented these normative debates using mixed methods research designs. Cath and Floridi (2016) use discourse analysis, participant observation, and qualitative interviews to test their arguments. The data for ten Oever's (2023) case study was pulled from the technical documents and mailing lists of RIPE, the Regional Internet Registry for the former Soviet Union, West Asia, and Europe. This data was analyzed using statistical, network, and discourse analysis. ten Oever (2023) then introduced a proposal on human rights and data protection and qualitatively documented participants' responses through semi-structured

interviews, participant observation, and mailing lists. This work reveals normative conflicts over human rights in particular time frames and around specific issues.

Much work in the IETF, and the technical community more broadly, is carried out over email lists. Becker et al (2022) use 3rd Generation Partnership Project (3GPP) published standards and mailing lists to examine collaboration on communications network standards development. They describe mailing lists as “unique decision-making cloud chamber[s]” (Becker et al, 2022, p. 7). According to ten Oever, Milan, and Beraldo (2020, p. 214) email lists hold “precious insights on process but also on feelings, values, relationships, and backstage dynamics”. They encourage mixed-methods research designs for exploring this data and argue that network analysis can be used to reveal “the structural basis of discourse, power relations among actors, and intergroup dynamics” (ten Oever, Milan, and Beraldo, 2020, p. 221). Doty (2015) analyzes the integration of privacy concerns into Internet standards at the IETF in part through comparing the activity of relevant email lists over time. Milan and ten Oever (2017, p. 8) perform an analysis of ICANN email lists. They examine when correspondents sent their first email on an ICANN email list and use that data to group correspondents into three cohorts. They then mapped changes in ICANN policy to who was joining the email list and changes in activity across the cohorts. Nanni (2022) compares how networks of individuals on IETF email lists changed over time in terms of which individuals were the most influential in the network, based on centrality measures. This work underlines the usefulness of email data in understanding online collaboration and contestation in the technical community.

Founded in 1986, the IETF is likely the first organization to utilize online collaboration for distributed creative work. Subsequent literature has described collaborative online work processes and the impact of controversy on online social networks. The type of work structure used by the IETF can be considered a form of crowdsourcing. Riedl et al. (2024) show that competitive or collaborative behaviors impact participants' long term engagement with the community. As described, human rights are a controversial issue in portions of the technical community. In some online settings, controversy is evident in the development of social networks and use of sentiment in communications. On some social media platforms, this can lead to the development of ‘echo chambers’ (Cinelli et al., 2021), while on others users prefer to engage with those holding opposing views (Morales et al., 2021).

Despite controversy and opposition, there is headway being made in the consideration of human rights in Internet standards design. Requests for Comments (RFC) 8280, “Research into Human Rights Protocol Considerations” has been heralded as one of the first times an Internet technical community has formally acknowledged that their engineering decisions have human rights consequences (Krishnamurthy, 2019). RFC 8280 summarizes research on human rights and technical architecture, relates the principles supporting Internet protocols to human rights, and lastly presents a questionnaire engineers can use to assess the impact of new protocols on human rights. This RFC was followed by RFC 9620 “Guidelines for Human Rights Protocol and Architecture Considerations” (Grover and ten Oever, 2024). According to RFC 9620’s abstract, it “sets guidelines for human rights considerations for developers working on network protocols

and architectures, similar to the work done on the guidelines for privacy considerations (RFC 6973).” It outlines threats to human rights and proposes various methods for conducting human rights reviews of protocols or standards. These methods are “Analyzing Internet-Drafts Based on Guidelines for Human Rights Considerations Model”, “Analyzing Internet-Drafts Based on Their Perceived or Speculated Impact”, “Expert Interviews”, “Interviews with Impacted Persons and Communities”, and “Tracing Impacts of Implementations”. The majority of RFC 9620 is dedicated to detailing the “Guidelines for Human Rights Considerations” method by providing:

[Q]uestions that engineers should ask themselves when developing or improving protocols if they want to understand how their decisions can potentially influence the exercise of human rights on the Internet.

These guidelines enable engineers to conduct a human rights review of their standard or protocol based on a list of questions. Each set of questions is accompanied by an explanation of its relevance to human rights and a list of impacted human rights. RFC 9620 draws its conception of human rights from the International Bill of Human Rights and the Charter of Human Rights and Principles for the Internet, and focuses on rights considered most pertinent to information and communications technologies. The purpose of RFC 9620 is to summarize the relevance of Internet standards and protocols to human rights and provide engineers with a structured method to evaluate the extent to which their standards and protocols can be used to enable or undermine human rights (Grover and ten Oever, 2024).

This paper will use emails to examine the processes of collaboration and contestation surrounding the integration of human rights considerations into Internet protocols in the IETF. In doing so, it builds on prior literature examining debates over human rights in the technical community and on prior literature pertaining to online collaboration and contestation more broadly. It contributes a comprehensive overview of this debate and explores its evolution at scale.

Methods

This study integrates time series, regression, and network analysis to explore norm entrepreneurship and diffusion in IETF/IRTF email networks. Emails are annotated by OpenAI’s GPT-4o, a large language model, to designate whether they support, oppose, or are neutral towards integrating human rights considerations in technical standards (Open AI Platform, n.d.). This data is then graphed in a time series to reveal overall changes in activity level and average support. Sending email accounts are also categorized by top level domain (eg .com or .edu), providing insight into the affiliation of the sender. Top level domain and how many emails an account sent (network out-degree) were then regressed on support. These results were used to group email accounts into cohorts based on when they joined the network. Cohorts are then

compared on the most frequent top level domain and the average level of support over time. Lastly, cohort networks are compared.

Data

This study uses a dataset of all publicly available IETF and IRTF emails containing the phrase “human rights” from 1992 to early 2025 not sent by accounts with the domain ietf.org. It is composed of 3,377 emails. Email metadata, such as sender address and data sent are extracted from the email headers. This metadata is then supplemented with annotations by GPT indicating the opinion of the author on whether human rights considerations should be integrated into technical standards, based on email content.

Data Collection

To collect email data from the IETF Mail Archive we navigated to <https://mailarchive.ietf.org/arch/> used to search bar to identify emails with the text “human rights”. This search was performed on February 3rd, 2025. It yielded 3664 messages sent between November 3rd, 1992 and January 13th, 2025. To export the emails, we created and logged into a Mail Archive account. Account creation required no particular credentials and could be completed by any member of the public with an email address. Once logged in, we selected Export and the URLs export type. This generated a list of unique URLs pointing to each email. We downloaded this list as a text file. Using Python code co-authored with Gemini in Google Colab, we scraped the email header, message, and links to the two prior emails in the email chain and their content. We then extracted relevant metadata from the header. [Python code](#) is available in the supplementary materials. We then exported the email data as a CSV and uploaded it to Google Sheets, where we filtered out emails where the sender account had the domain ietf.org, as these are typically institutional or administrative emails and do not reflect the opinions of individuals. This is similar to the method used by Welzl et al (2021) to filter out email addresses not associated with individuals.

Annotation using GPT

Emails were annotated by GPT 4o-2024-08-06 (Open AI Platform, n.d.) to indicate their support for or opposition to integrating human rights in Internet technical standards. Annotations represent the opinion of the primary email author as identified by the sending email address. The two prior emails from the email chain (if present) are used to provide context to the primary author’s email. For example, if the primary author’s email contained only “That’s a bad idea” and an email signature the prior email from the chain would be used to provide context for what idea they are referring to and whether it is relevant to human rights. Emails were assigned a score of 1, 0, or -1. A score 1 indicates that the email is relevant to human rights and the author is

supportive of integrating human rights considerations into technical standards. A score of -1 indicates that the email is relevant to human rights and the author opposes integrating human rights considerations into technical standards. A score of 0 indicates that the email is not relevant to human rights or the author expressed a neutral opinion.

Annotation was done using the OpenAI API to query GPT. [Python code](#) for submitting the queries and processing the results was co-authored with ChatGPT 4o and Gemini in Google Colab and is available in the supplementary materials. GPT 4o-2024-08-06 was queried with the below prompt:

"You are an expert research assistant in human rights, specializing in assessing support for human rights, particularly in relation to Internet technical standards. Your task is to analyze and assess text-based conversations to detect emails that support, oppose, or are neutral/not relevant to integrating human rights considerations into Internet technical standards. Human rights are those rights defined in the International Covenant on Civil and Political Rights, the International Covenant on Economic, Social and Cultural Rights, and the Universal Declaration of Human Rights. An Internet technical standard "is a normative specification of a technology or methodology applicable to the Internet" (https://en.wikipedia.org/wiki/Internet_Standard#:~:text=In%20computer%20network%20engineering%2C%20an,methodology%20applicable%20to%20the%20Internet). What to identify: You are looking for emails where the primary author supports integrating human rights considerations into technical standards, opposes integrating human rights considerations into technical standards, or is neutral on whether human rights considerations should be integrated into technical standards; or where the author's email is not relevant to integrating human rights considerations into technical standards or technology development more broadly. You should score the email based on the email text. You should use the context text to help determine whether the email is relevant to human rights. If the email text contains multiple potential authors you should use the from email address information to determine the primary author. Scoring: End your answer with the final score 0 = Email or context are relevant to integrating human rights considerations into Internet standards and the author does not express support or opposition to integrating human rights considerations into technical standards; or email and context are not relevant to integrating human rights considerations into technical standards or technology development more broadly. If an email discusses a particular country's human rights record or policies, but does not discuss integrating human rights considerations into Internet standards, it should be scored as not relevant since it is not about integrating human rights into Internet standards. If an email makes a general reference to International Human Rights Law, but does not discuss integrating human rights considerations into Internet standards, it should be scored as not relevant since it is not about integrating human rights into Internet standards. 1 = Email or context are relevant to integrating human rights considerations into Internet standards and the author

expresses support for integrating human rights considerations into technical standards. Support can be in the implicit form of providing constructive feedback on drafts or documents related to integrating human rights considerations, or promoting meetings, events, or research on human rights or related topics. Support for integrating human rights considerations can also take the form of opposing the use of a technology or standard because it might negatively impact human rights. Support can also take the form of a general philosophical stance that considering human rights in internet standards development is important, it does not have to present specific methods for integration. -1 = Email or context are relevant to integrating human rights considerations into Internet standards and the author expresses opposition to integrating human rights considerations into technical standards. Opposition can be in the implicit form of providing sarcastic or otherwise unhelpful feedback on drafts. Opposition can also be in the form of arguing that human rights are not relevant to technical or engineering work. Examples: Neutral or not relevant (Score = 0): I have done a review of the HRPC [Human Rights Protocol Considerations] guidelines doc as well. Please see the attached. This email is relevant to human rights but the author does not express support or opposition to integrating human rights into technical standards, as it does not give information about the referenced review, therefore it is neutral; it should be rated as Score = 0. We have now received all necessary approvals and consider AUTH48 complete Thank you for your attention and guidance during the AUTH48 process. We will move this document forward in the publication process at this time. This email is not relevant to integrating human rights considerations into technical standards or technology development more broadly, it merely describes the document publication process; it should be rated Score = 0. Supportive emails (Score = 1): For example, protection of whistleblowers is widely recognized as an important element necessary to underpin the right of freedom of expression (and access to information), yet it often results in loss of personal privacy and/or security for the individuals involved – most would agree that using communication technology to disclose an official's embezzlement of public funds is reasonable and yet disclosing their spouse's serious medical condition is not appropriate. The communication protocols are agnostic on content, and so we need to be intellectually honest and document if a protocol's attributes preclude support for legal remedy. It doesn't mean that such a protocol is bad but simply means that there's a conscious tradeoff that's been made on prioritization of functionality in favor of certain human rights (in that case freedom of expression over legal remedy.). In this email, the author discusses trade-offs between different human rights in what technical protocols prioritize, their promotion of documentation and intellectual honesty around the relation between human rights and Internet protocols means that this email is supportive of integrating human rights into Internet protocols; this message should be rated Score = 1 I think we don't want new considerations, but it would make sense to perhaps deepen, or address, the existing considerations on association, otherwise there is not much connection

between them. One suggestion is to perhaps bring forward the relevant 8280 sections by expanding a paragraph starting with this sentence: '[RFC8280] established the relationship between human rights and Internet protocols, therefore in this document we seek to uncover the relation between two specific human rights and the Internet infrastructure and architecture.'. This email describes further developing the consideration of the relationship between human rights and Internet standards, therefore its author is supportive of integrating human rights into technical standards; this message should be rated Score = 1 Opposing emails (Score = -1): The difference between protocol and policy is the difference between engineering and law. IETF is engineering. For example, engineers build a tunnel that can work for both cars and buses. Policy may preference one or the other but the engineers have a duty to avoid tilting that choice unless it has direct economic impact. The GNAP example I have brought to this group is asking the engineers to stay out of the decision of whether the resource server or the resource owner has a stronger privacy interest. The GNAP protocol can and therefore should treat both as equally important and let constraints be a matter of policy and law. The author of this email argues that Internet protocols and engineering decisions should not support a preference for one use case over another, and that those decisions should instead be a matter of policy, therefore, this email opposes integrating human rights into technical standards; this message should be rated Score = -1 Be precise and consistent in scoring. Before you provide an answer, explain why the email was given a particular score. Score Explanation: Score:"

"Context:"

"Email:"

"Author:"

To evaluate the annotations, we reviewed 20 randomly selected emails and their annotations to determine whether the scores were appropriate based on the primary email and context emails. Out of these 20, 19 were annotated correctly. This initial random sample did not include any scored emails scored -1, likely due to the low prevalence of opposing emails in the dataset. Therefore, we also randomly selected five emails scored -1 for review. Four out of the five randomly selected -1 scores were appropriate. Overall, 23 out of the 25 emails were scored correctly (92%).

Descriptive Statistics

The email dataset contains 3,377 emails sent between November 3rd, 1992 and January 13th, 2025. All descriptive statistics and further analysis were performed using the [Python code](#) co-authored with Gemini in Google Colab and available in the supplementary materials. Figure 1 displays the number of emails sent each year.

Figure 1

Number of Emails per Year

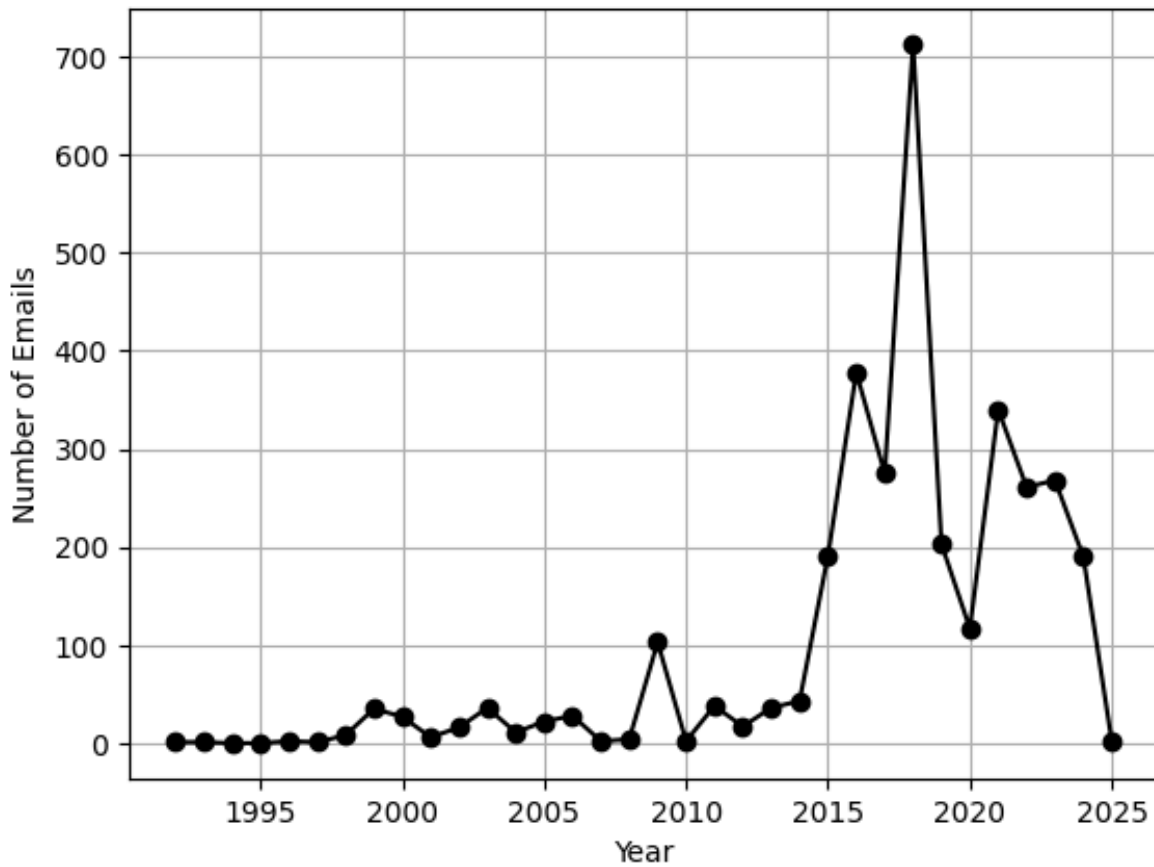
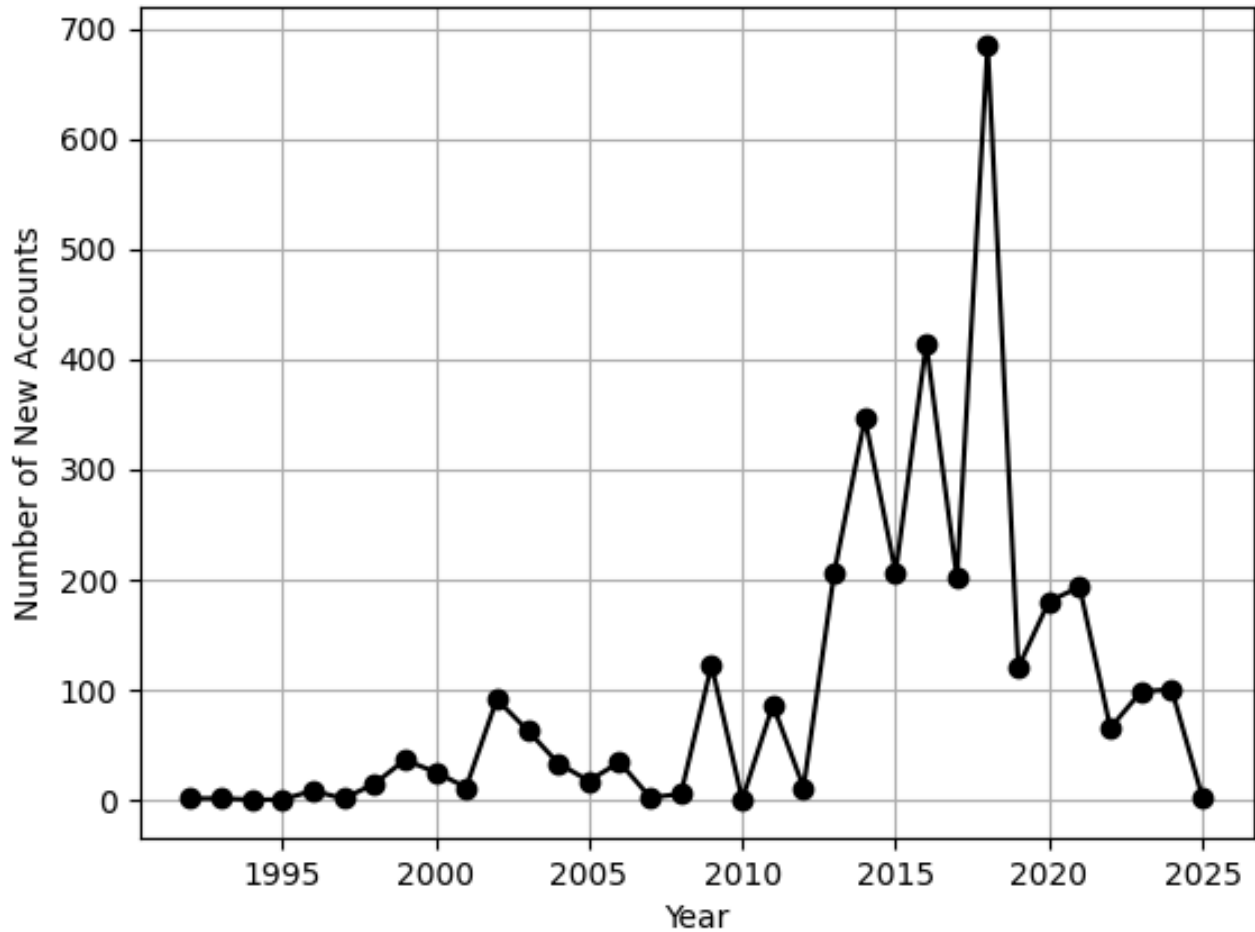


Figure 1 shows spikes in email frequency for the years 2016, 2018, and 2021. Figure 2 displays the number of new email accounts joining the network each year.

Figure 2

Number of New Accounts Joining per Year

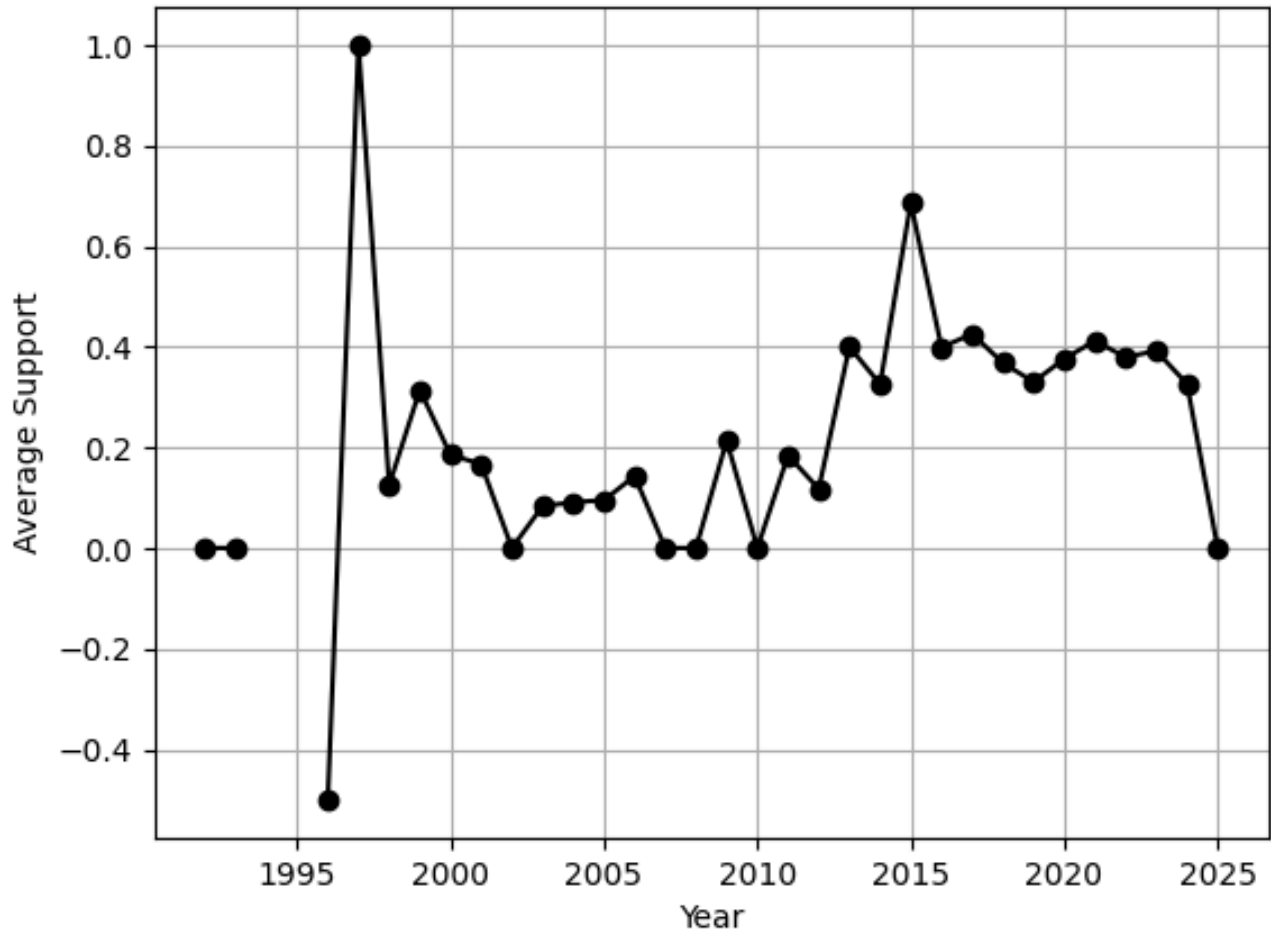


This figure shows that around 2013 many new email accounts began joining the network.

Out of the 3,377 emails, 190 were scored with -1; 1,728 were scored with 0; 1,454 were scored with 1; and five were scored with NA (due to missing message text). Figure 3 displays the average support score for each year.

Figure 3

Average Support Score of Emails per Year



The years with the highest average support are 1997 and 2015.

Figure 4 displays the number of emails per year for the top five most frequent top level domains. The top level domain provides insight into the potential affiliation of the sender. For example, an email ending in edu indicates that the sender is likely affiliated with an academic institution, while an email ending in com is likely a personal or company email.

Figure 4

Number of Emails per Year for the Top 5 TLDs

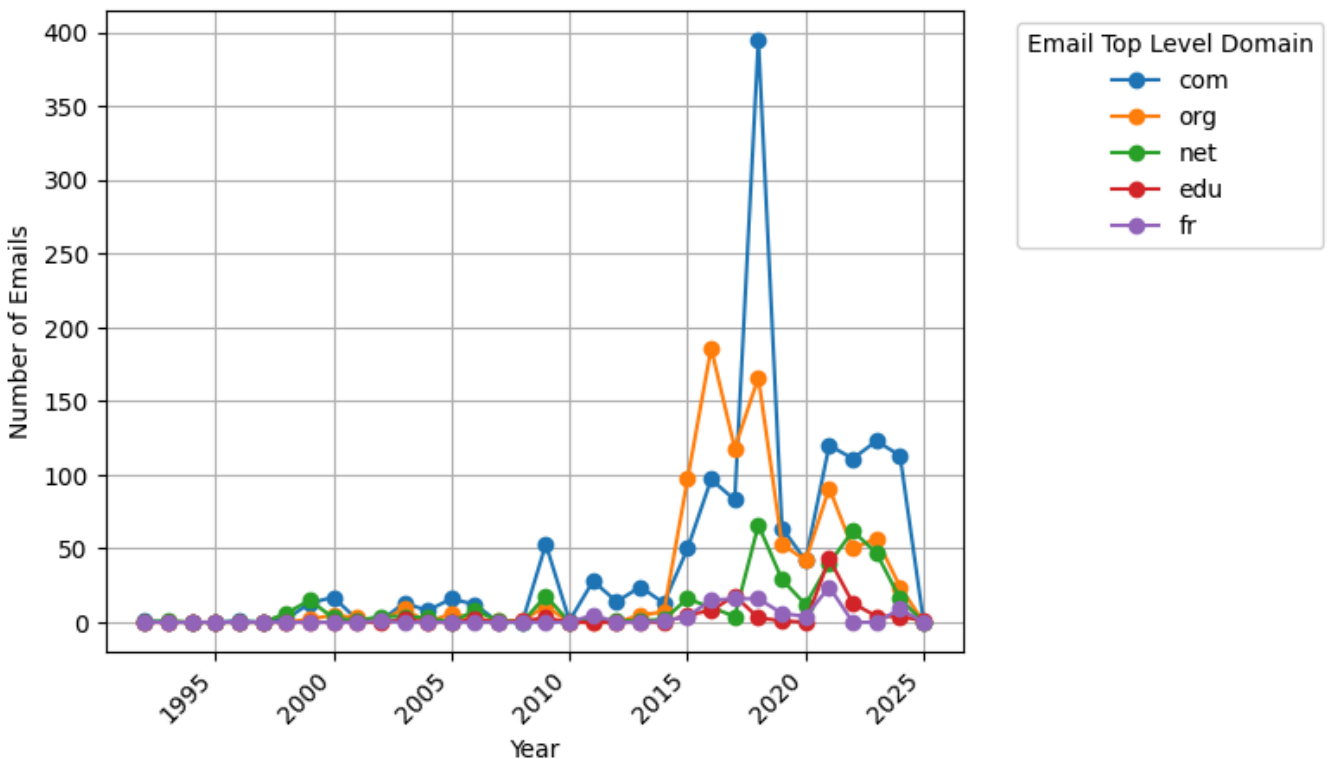


Figure 4 shows that for most years com is the most prevalent top level domain, with the exception of the years 2015, 2016, 2020, and 2025. In 2025, edu was the most prevalent TLD (though this is likely due to the limited data for 2025). In the other years, org was either the most prevalent TLD or tied with com (as in the year 2020). This could indicate that participants associated with organizations, potentially members of civil society, were the most active participants on the network in 2015 and 2016.

Levels of support for integrating human rights considerations varied by TLD. Table 1 displays the total average support score for the top five TLDs for all years as well as the number of total emails sent from that domain.

Table 1*Support and Frequency by Top Level Domain*

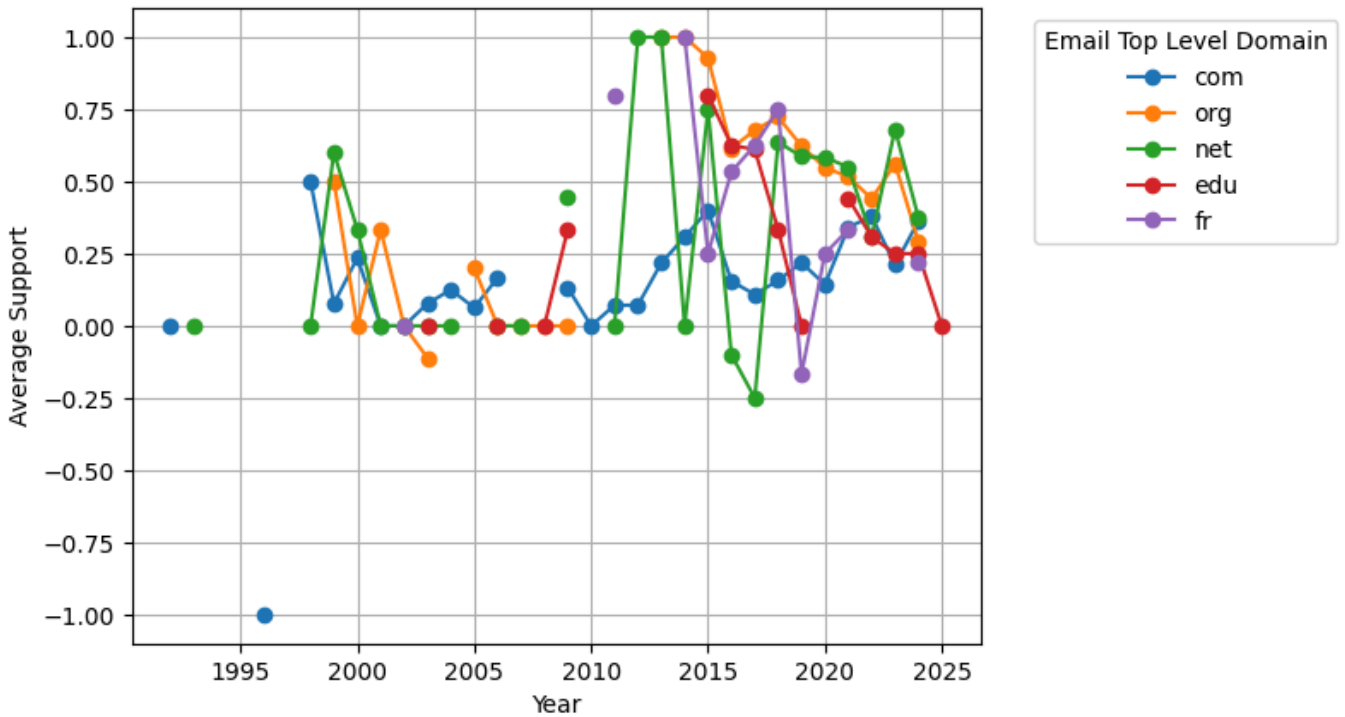
Email Top Level Domain	Average Support Score	Total Number of Emails Sent
edu	0.44	108
com	0.22	1417
net	0.47	372
org	0.62	937
fr	0.46	101

Note. This table includes data only for the top five most frequent top level domains.

These average scores indicate that email authors writing from addresses ending in com, likely company or personal emails, displayed the lowest average support for integrating human rights considerations. Authors writing from addresses ending org, likely organizations, displayed the highest average support. There were 30 emails in the dataset where the top level domain could not be automatically extracted. Figure 5 shows the average support score per year for emails with each TLD.

Figure 5

Average Support Score of Emails per Year by TLD



For most years, emails ending in net, org, or edu showed the highest level of support, while emails ending in com showed the lowest level of support. When comparing average scores for emails ending in org across years, there is a marked increase in average scores in 2013 and after compared to the time period 1999 to 2009.

Figure 6 displays a histogram of average support by email domain.

Figure 6

Average Support by Domain

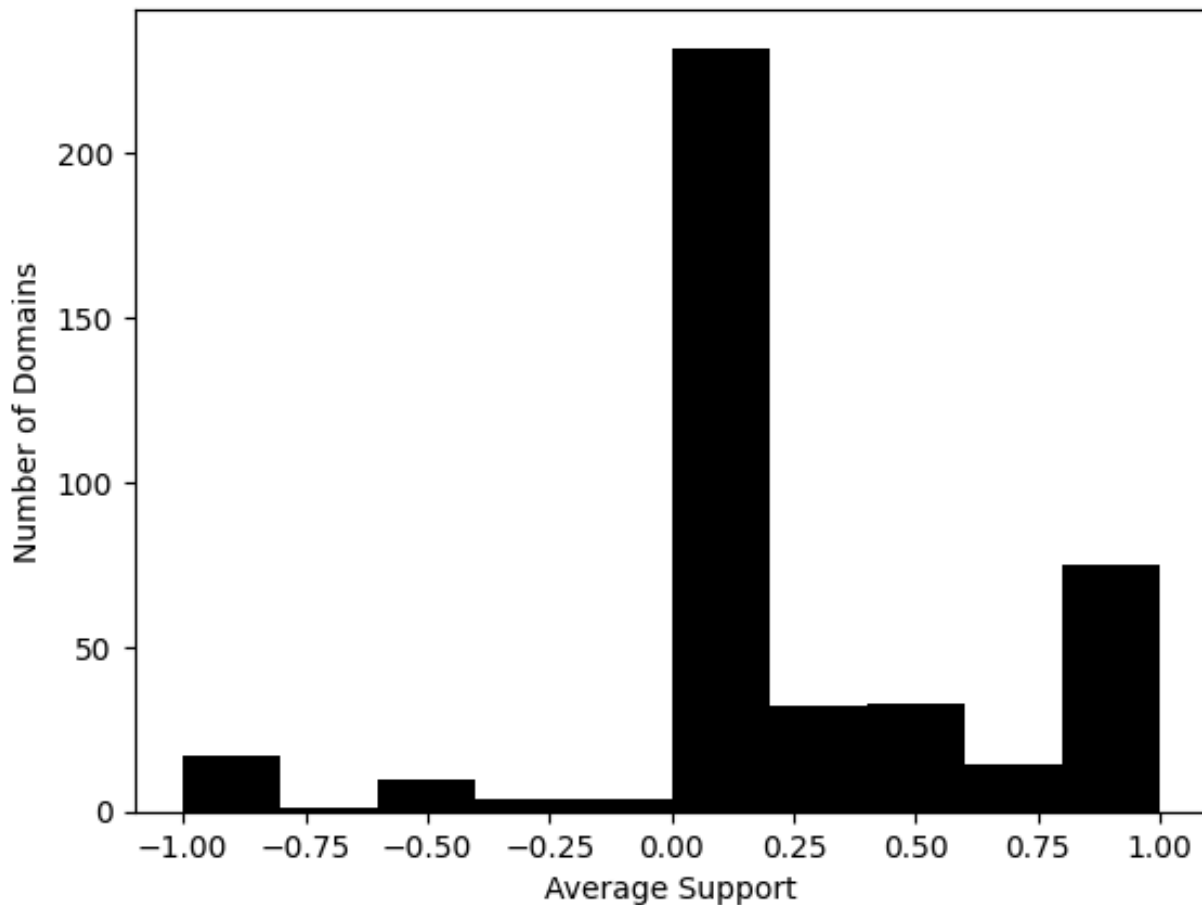


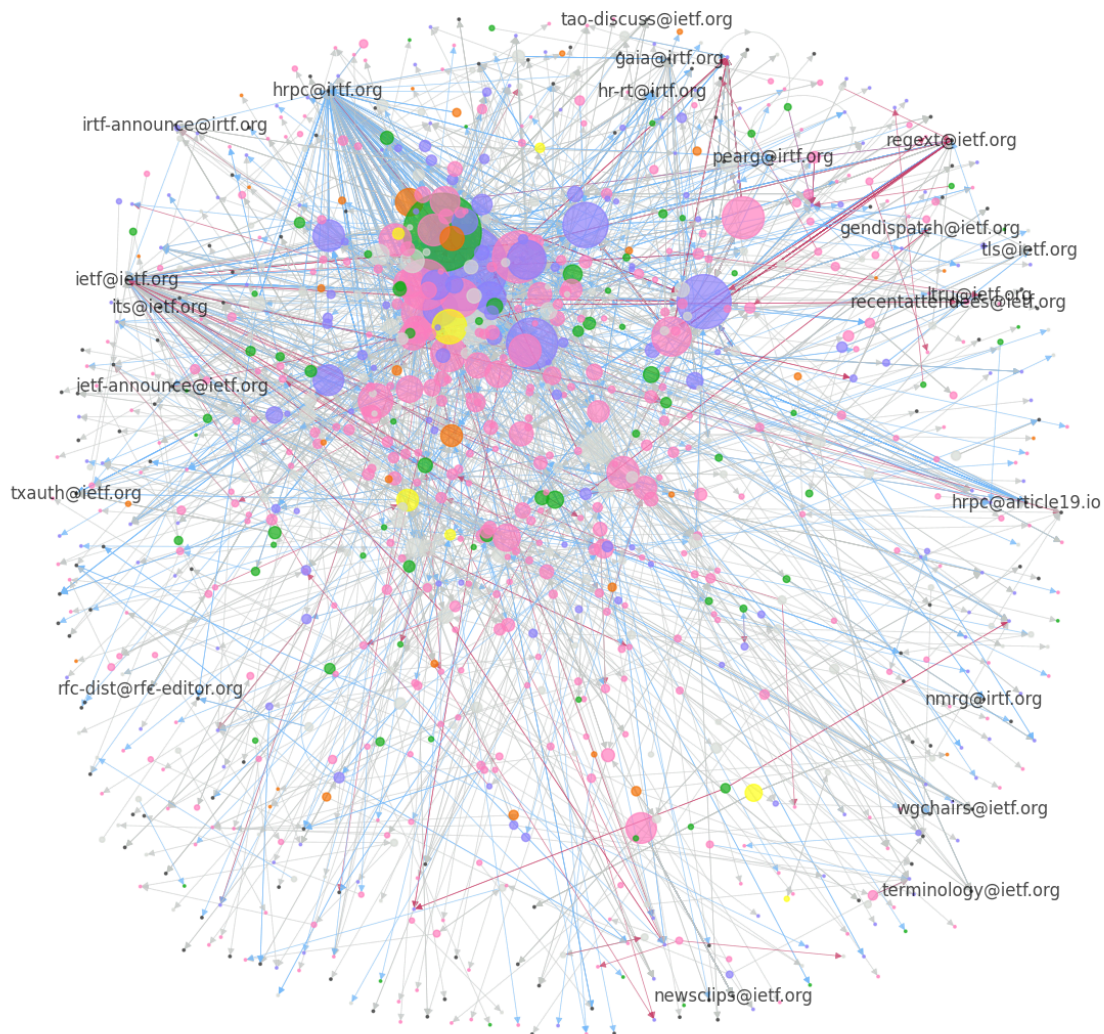
Figure 6 shows that most domains have an average support of between 0 and .25, indicating a neutral to slightly positive stance on integrating human rights considerations into technical standards. Overall, a higher number of domains have a positive average support. This indicates that the promotion of human rights considerations is more distributed over a larger number of email domains, while opposition is more concentrated in fewer domains.

Figure 7 displays the network graph of all emails in the dataset. Nodes are either sender or receiver email accounts. They are sized based on the number of emails sent from that account. They are colored based on the account top level domain: addresses ending in com are pink, addresses ending edu are orange, net is green, fr is yellow, and org is purple. Each edge represents an email sent between two nodes. Since one email can go to multiple recipients, each edge represents one sender-receiver dyad. For example, an email from one account to three others would be represented in three separate edges. Edges are colored based on the support score of the email: -1 is red, 0 is grey, and 1 is blue. Due to privacy considerations, email accounts associated with individuals are not labeled with their addresses. Mailing list email

accounts that received 10 or more emails are labeled. For 177 emails, the mailing list address could not be automatically extracted. In total, this network has 4,998 edges and 971 nodes.

Figure 7

Email Network



Note. Nodes are colored by top level domain and if they are a mailing list: addresses ending in com are pink, addresses ending edu are orange, net is green, fr is yellow, org is purple, and mailing lists are black. Nodes are sized based on the number of outgoing emails. Edges are colored based on the support score of the email: -1 is red, 0 is grey, and 1 is blue. Mailing lists that received at least 10 emails are labeled.

The most prominent email lists in this network are the [Human Rights Protocol Considerations Research Group](mailto:hrpc@irtf.org) mailing list (hrpc@irtf.org), used for the work of the HRPC research group, and the [IETF-Discussion](mailto:ietf@ietf.org) mailing list (ietf@ietf.org), a general discussion list. The average support score for emails sent to the hrpc@irtf.org mailing list is 0.51, while the average support score for emails sent to the ietf@ietf.org mailing list is 0.14. These differences are also visually reflected in the network, with noticeably more blue lines connected to the hrpc@irtf.org node than the ietf@ietf.org node.

Results

Our results suggest the presence of two distinct phases of norm entrepreneurship in the network. We use these phases to separate participants into two cohorts based on when they joined the network. This is similar to the method used by Milan and ten Oever (2017) to analyse cohorts in ICANN. We then evaluate the network and participant attributes and level of support for each cohort to explore norm entrepreneurship and diffusion.

Regression

Top level domain and out-degree (the number of emails sent from an account) were regressed on support using OLS regression from the statsmodels package in Python. No top level domain was a significant predictor of support. Regressing out-degree on support was done first for all years then for individual years to determine change over time in the relationship between out-degree and support. For all years, support was averaged for each email address across all of the emails they sent. Out-degree was calculated for each email address based on how many emails it had sent for all years. Out-degree was then regressed on support. The results show that out-degree is a significant predictor of average support ($p < .01$) though the effect is very small: the coefficient is 0.002.

This calculation was then repeated for each year in the dataset. For each year, an email network was created including only emails sent in that year. If the network included at least 25 addresses, the out-degree of each address was then regressed on average support for emails sent in that year by each address. There were no individual years in which the relationship was significant.

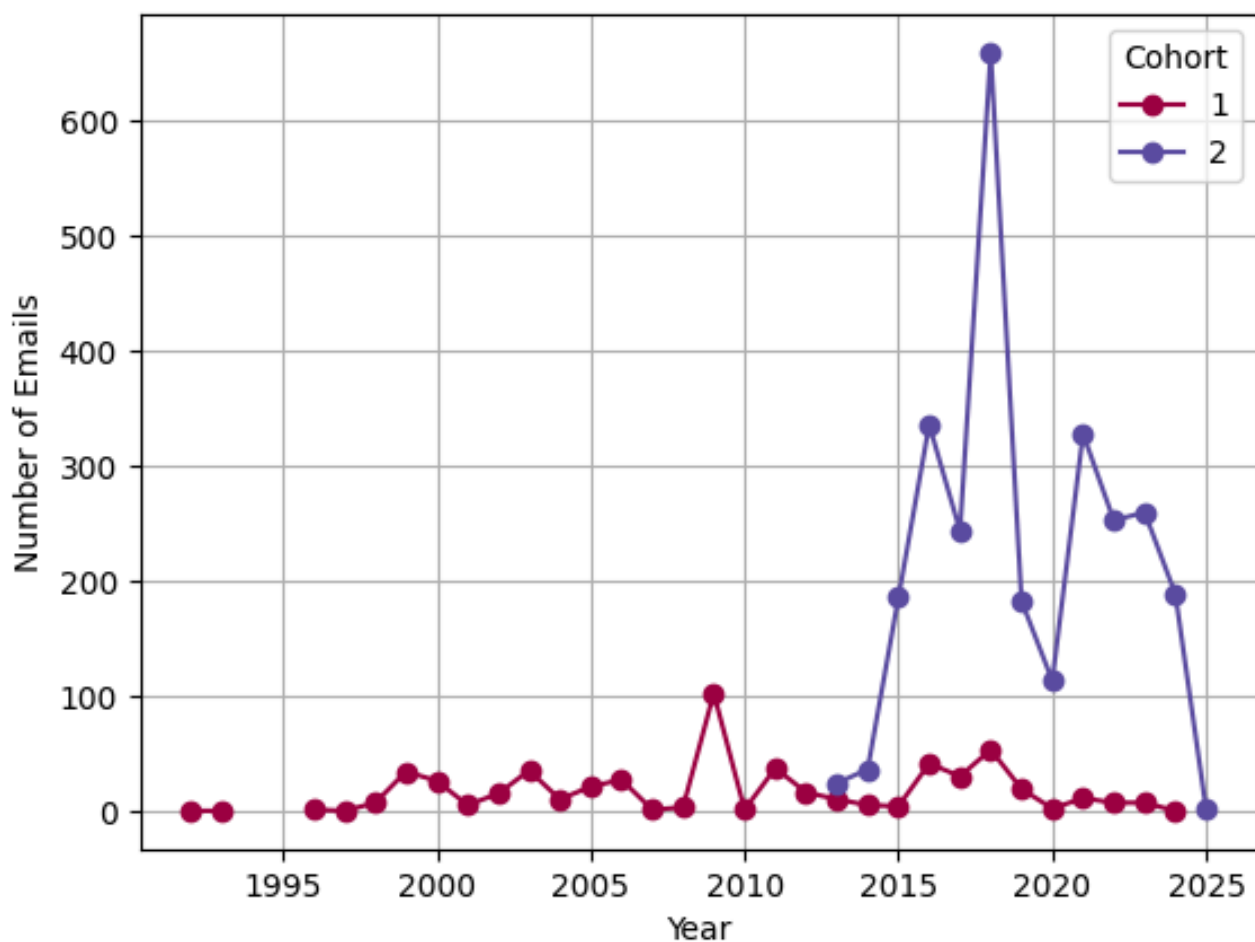
The relationship between out-degree and average support was then assessed for the time period before 2013 and the time period during and after 2013. The year 2013 was selected due to the marked increase in activity in the years following and the increase in average support for integrating human rights. In the network of emails sent between 1992 and 2012, out-degree was not a significant predictor of average score. However, in the network of emails sent between 2013 and 2025, out-degree was a significant predictor of average score ($p < .01$), though the impact is very small (coefficient is 0.0032).

Cohort Comparison

Based on the regression results and descriptive statistics presented, a new cohort of norm entrepreneurs sending many emails supporting the integration of human rights considerations may have joined the network around 2013. Email accounts are grouped into cohorts based on whether they joined the network before 2013 or during or after 2013. Figure 8 shows the yearly number of emails sent by each cohort.

Figure 8

Yearly Email Frequency by Cohort



Cohorts are then compared by the most frequent top level domain, the average level of support over time, and their network characteristics.

Top Level Domain

The total number of emails sent for each of the 10 most frequent domains and their average level of support for cohorts 1 and 2 are displayed in tables 2 and 3.

Table 2

Most Frequent Domains in Cohort 1

Email Top Level Domain	Total Number of Emails Sent	Average Support
com	275	0.05
net	73	0.26
fr	63	0.49
org	46	0.07
es	14	0.07
tv	8	0.13
edu	8	0.13
jp	7	-0.43
im	6	0.50
uk	4	0.25

Note. For 30 emails in this cohort, the top level domain could not be automatically extracted.

Table 3

Most Frequent Domains in Cohort 2

Email Top Level Domain	Total Number of Emails Sent	Average Support
com	1142	0.26
org	891	0.65
net	299	0.52

edu	100	0.46
ie	73	0.18
fr	38	0.39
it	28	0.29
uk	27	0.22
ch	25	0.48
ca	21	0.52

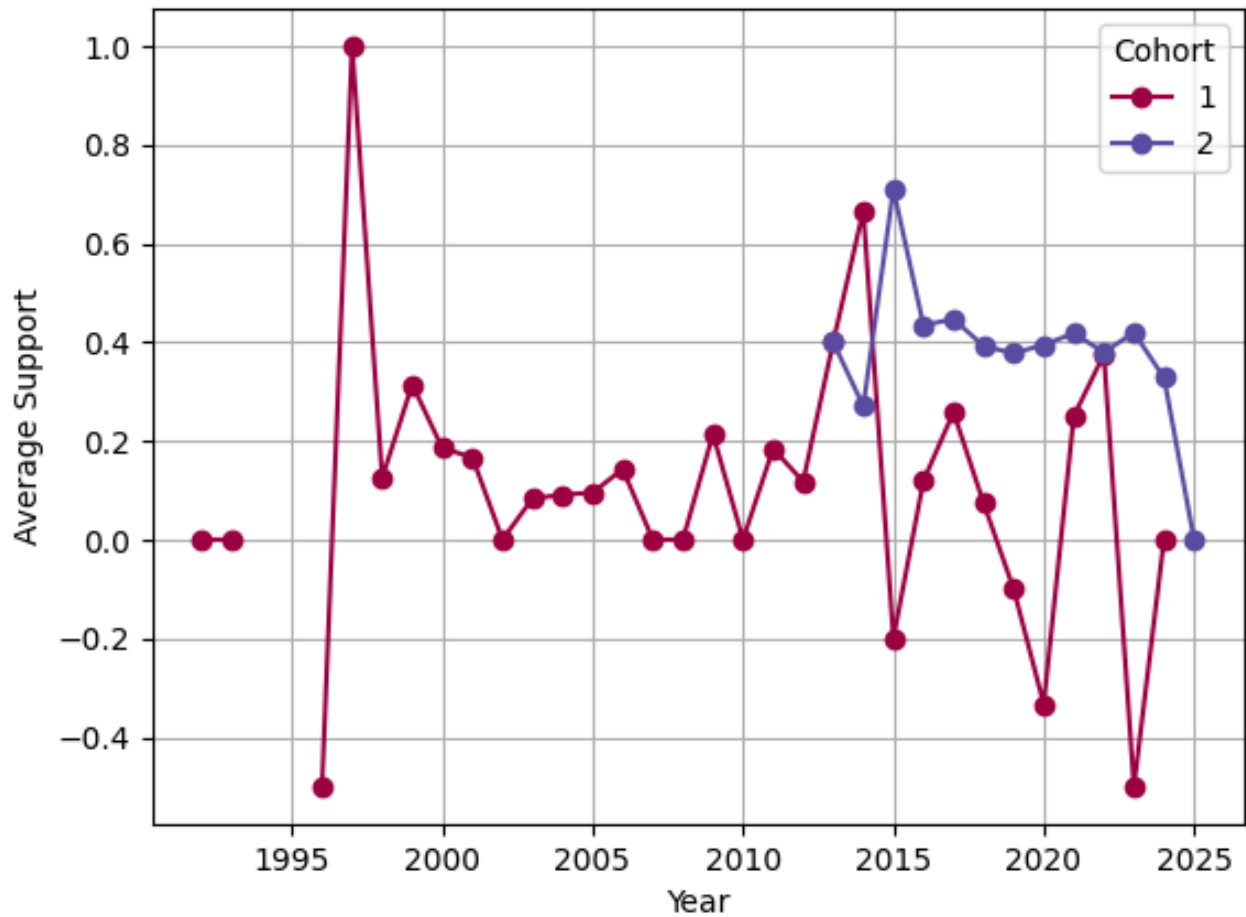
For cohort 2, the prevalence of emails ending in org and edu is much higher than for cohort 1. These domains also have higher average support scores in cohort 2 than in cohort 1. Emails ending in com also saw an increase in average support score from cohort 1 to cohort 2, though the change was not as large as for org and edu.

Average Level of Support

Figure 9 shows the yearly average support for each cohort.

Figure 9

Yearly Average Support by Cohort



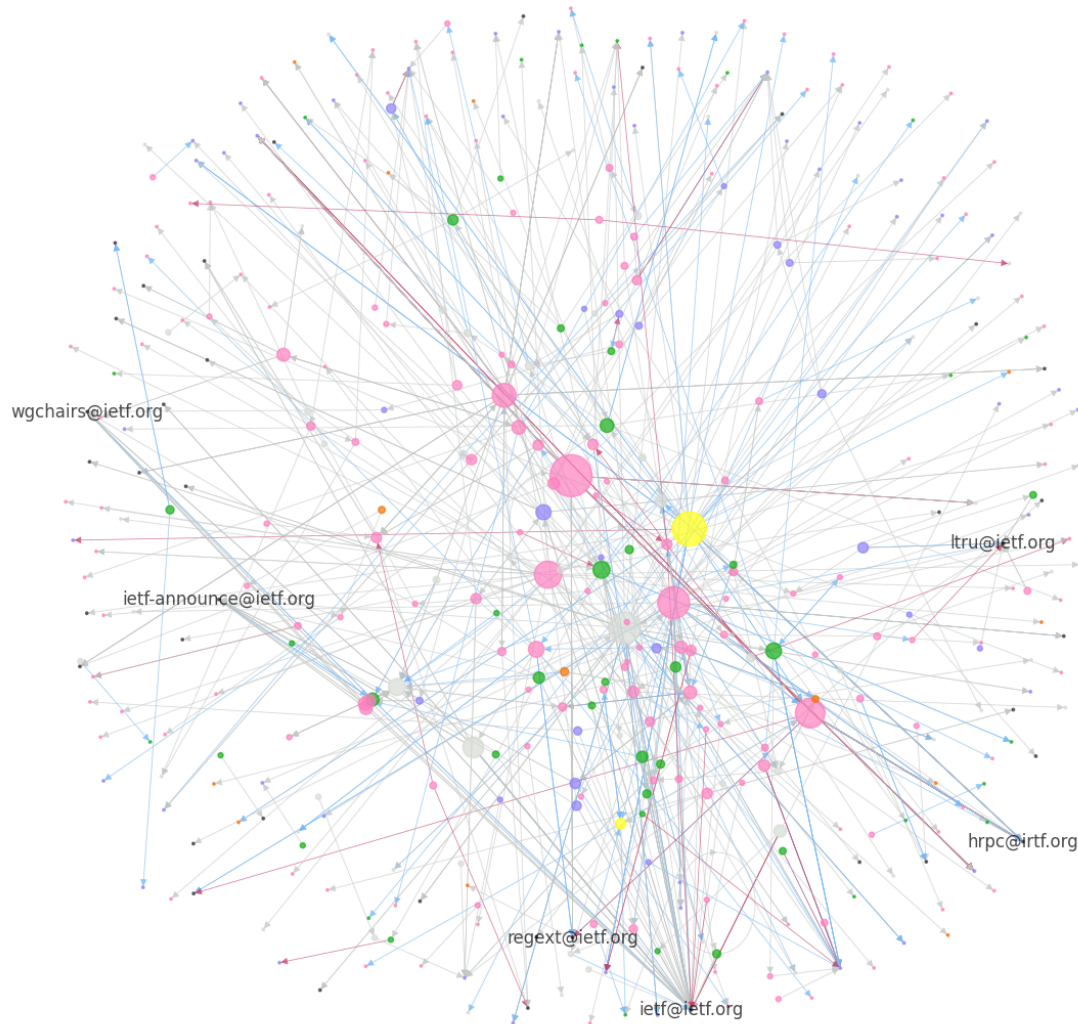
With the exception of 2014, the average support of cohort 2 is equal to or greater than the average support of cohort 1 for each year after 2013.

Network Characteristics

Figures 10 and 11 display the email networks for cohorts 1 and 2.

Figure 10

Cohort 1 Email Network



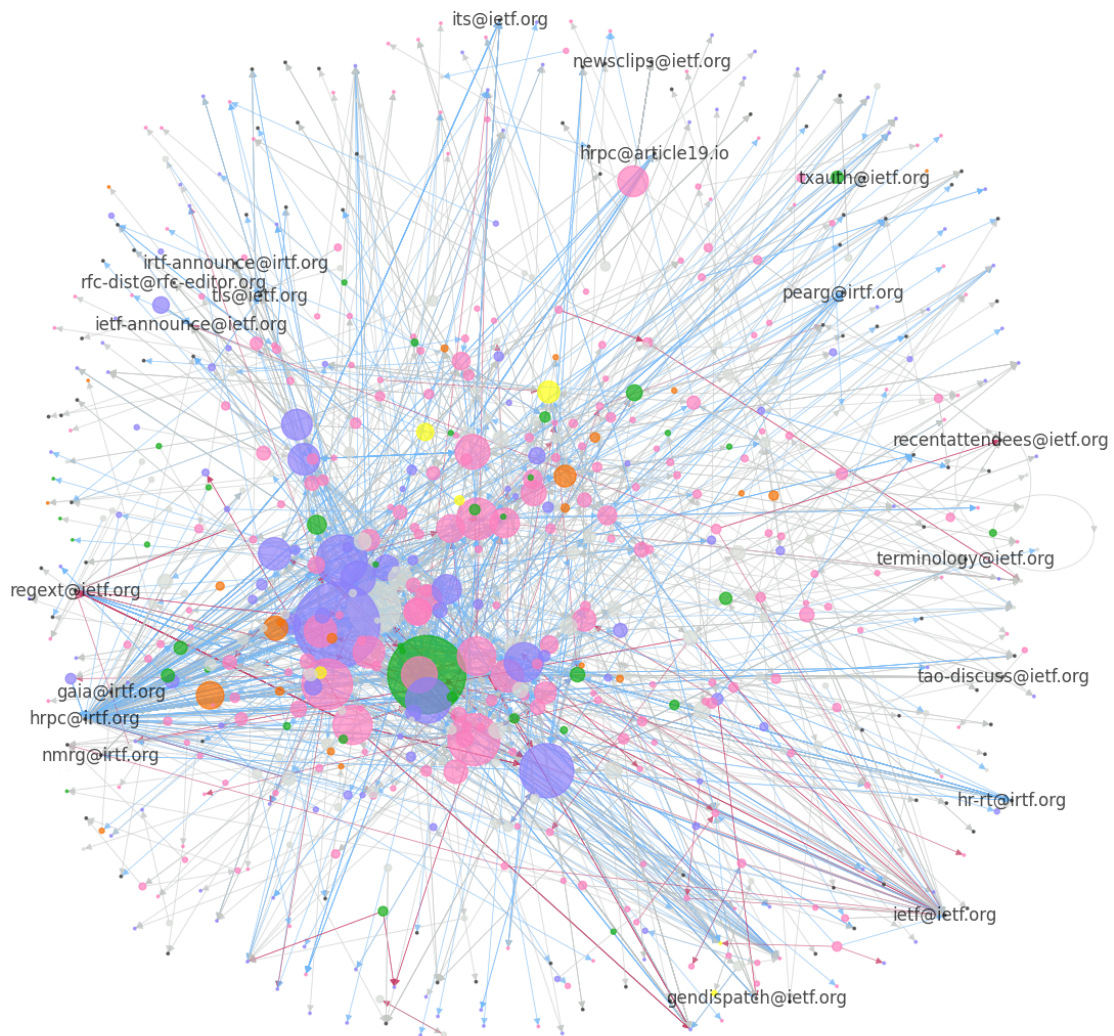
Note. Nodes are colored by top level domain and if they are a mailing list: addresses ending in com are pink, addresses ending edu are orange, net is green, fr is yellow, org is purple, and mailing lists are black. Nodes are sized based on the number of outgoing emails. Edges are colored based on the support score of the email: -1 is red, 0 is grey, and 1 is blue. Mailing lists that received at least 10 emails are labeled.

The network for cohort 1 displays emails sent by email accounts that joined the network prior to 2013. Given that some cohort 1 emails were still active after 2013 this network includes some emails sent after 2013. Furthermore, while all accounts sending emails in this network are from

cohort 1, some cohort 2 accounts may be included in this network as nodes if they received emails from accounts in cohort 1.

Figure 11

Cohort 2 Email Network



Note. Nodes are colored by top level domain and if they are a mailing list: addresses ending in com are pink, addresses ending edu are orange, net is green, fr is yellow, org is purple, and mailing lists are black. Nodes are sized based on the number of outgoing emails. Edges are colored based on the support score of the email: -1 is red, 0 is grey, and 1 is blue. Mailing lists that received at least 10 emails are labeled.

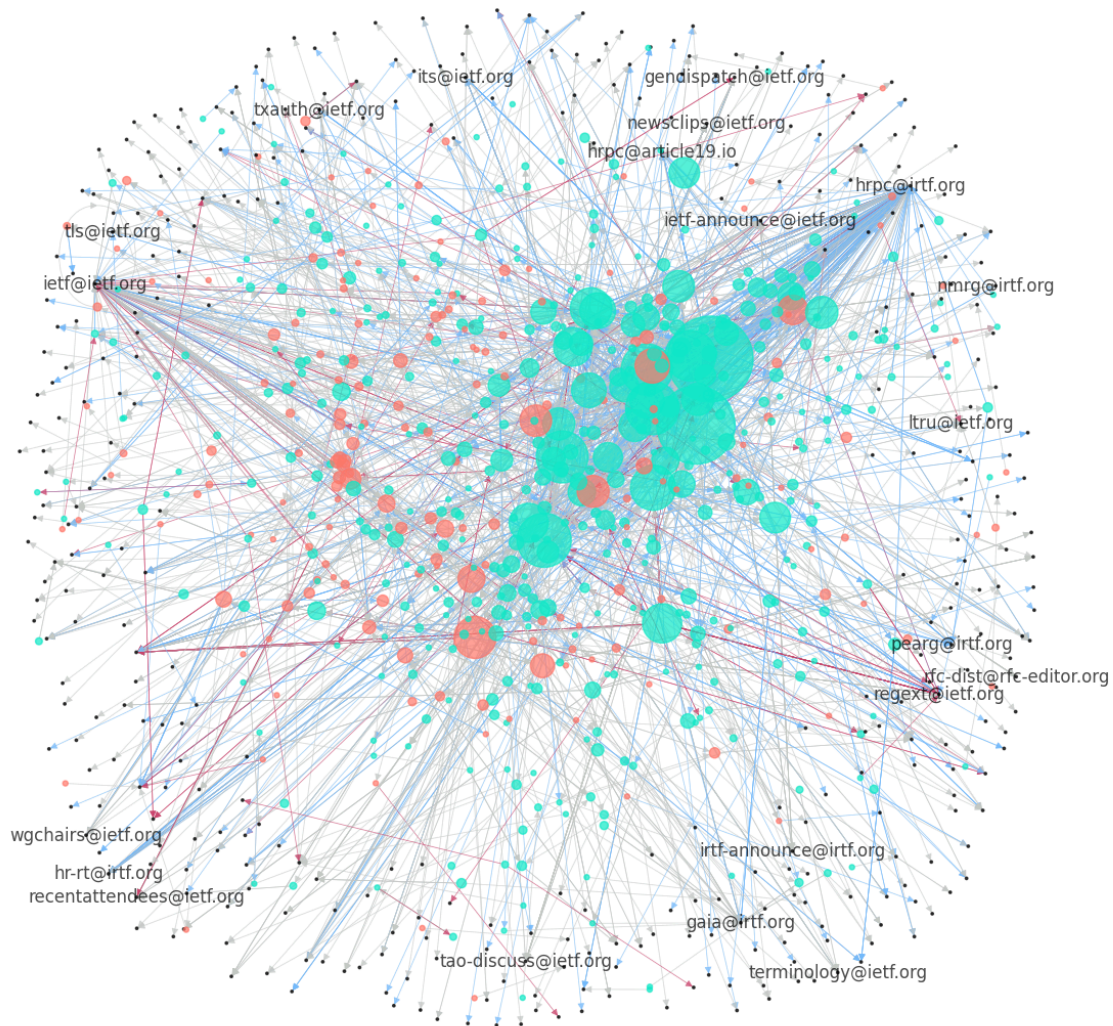
The network for cohort 2 displays emails sent by email accounts that joined the network during or after 2013. If emails were sent from cohort 2 accounts to cohort 1 accounts, these emails would be included in the network.

There are several notable differences between the cohort 1 and 2 networks. Firstly, the presence of accounts with the top level domain org (represented in purple) sending many emails is much more pronounced in the cohort 2 network. The email list ietf@ietf.org is prominent in both networks and receives both supportive and opposing emails from both cohorts. The email list hrpc@irtf.org is much more frequently emailed by cohort 2 accounts than by cohort 1 accounts, and these emails appear to be mostly supportive. Across the two cohorts, these email lists receive emails with differing levels of support. Emails sent by cohort 1 accounts to ietf@ietf.org have an average support of 0.18, while emails sent by cohort 2 accounts have an average support of 0.13. Emails sent by cohort 1 accounts to hrpc@irtf.org have an average support of 0.16, while emails sent by cohort 2 accounts have an average support of 0.54. For 177 emails, the mailing list address could not be automatically extracted.

Figure 12 shows the full network with nodes colored by cohort.

Figure 12

Email Network: Cohort Comparison



Note. Nodes belonging to cohort 1 are colored orange and nodes belonging to cohort 2 are colored aqua. Nodes are sized based on the number of outgoing emails. Edges are colored based on the support score of the email: -1 is red, 0 is grey, and 1 is blue. Mailing lists that received at least 10 emails are labeled.

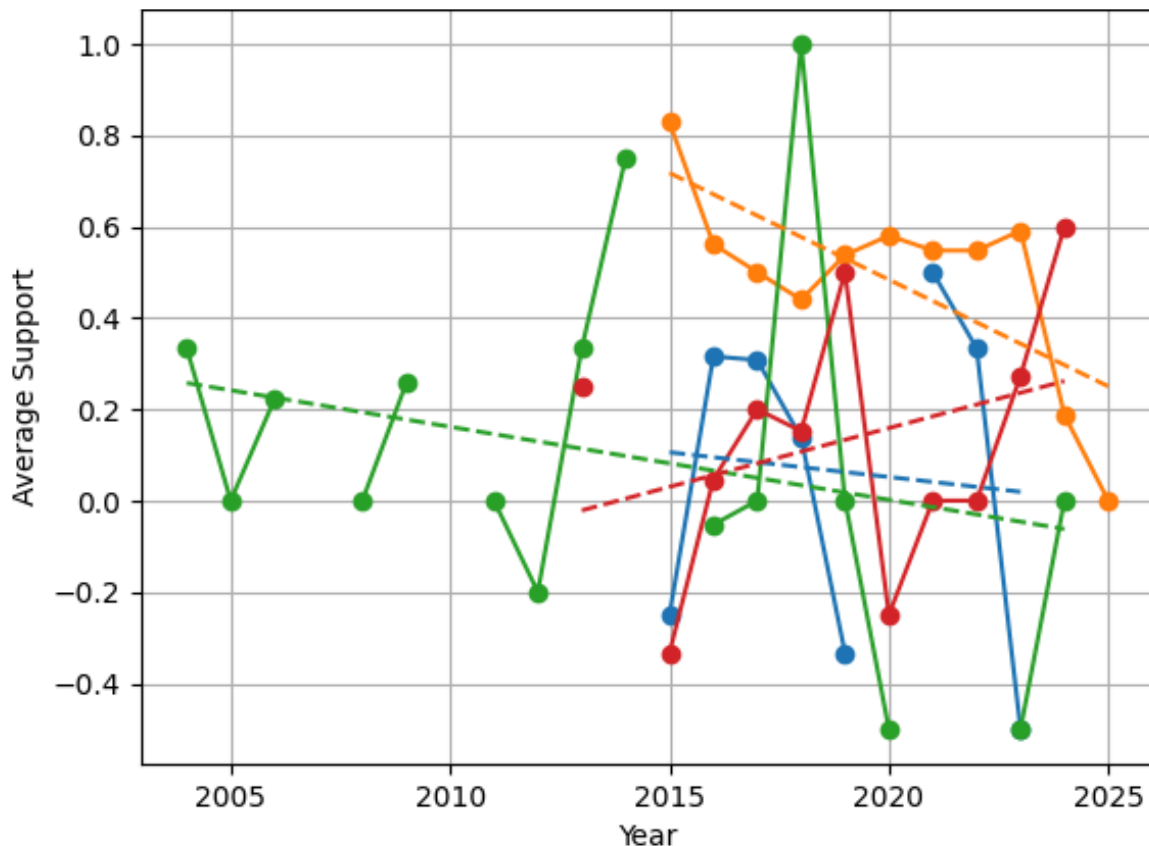
Nodes belonging to cohort 1 are colored orange, nodes belonging to cohort 2 are colored aqua. This network shows that many of the high out-degree cohort 2 nodes (email accounts that send many emails) are displayed near the hrpc@irtf.org mailing list, which is receiving many supportive emails. These nodes can be considered norm entrepreneurs. The nodes displayed near

ietf@ietf.org are more mixed between the two cohorts and generally have a lower degree. Emails sent to ietf@ietf.org are also more distributed in terms of supportive and opposing emails.

Figure 13 shows the average level of support each mailing list received over time by members of cohort 1 and cohort 2.

Figure 13

Yearly Average Support by Cohort and Mailing List



Note. Yearly average support by cohort 1 emails sent to hrpc@irft.org is shown in blue. Yearly average support by cohort 2 emails sent to hrpc@irft.org is shown in orange. Yearly average support by cohort 1 emails sent to ietf@ietf.org is shown in green. Yearly average support by cohort 2 emails sent to ietf@ietf.org is shown in red. Trendlines for each series are shown as a dashed line of the same color as that series. The only significant trendline is that for cohort 2 emails to hrpc@irft.org.

From 2015 to 2023, cohort 2 emails to hrpc@irft.org had the highest average support for each year except for 2018. It is also the only series that never had a yearly average below 0. However, it is also the only series with a significant trendline ($p < 0.05$), which slopes downward. This

could be due to the increasing presence of emails discussing how to implement integrating human rights considerations, rather than advocating for human rights to be integrated.

Discussion and Conclusions

As the Internet becomes more central to daily life, the impact of online harms and the importance of protecting the Internet as a rights-enabling space have become more pronounced. This has been recognized by international organizations and is increasingly being acknowledged by the technical community. In 2011, a United Nations report highlighted the role of the Internet in enabling human rights (UN Human Rights Council, 2011). One year later, the UN Human Rights Council published the resolution on “The promotion, protection and enjoyment of human rights on the Internet” (UN Human Rights Council, 2012). Also in 2012, the Internet Society published the report “Human Rights and Internet Protocols: Comparing Processes and Principles” (Liddicoat and Doria, 2012). In 2013, the Snowden leaks were published, revealing U.S. government surveillance (British Broadcasting Corporation, 2013).

These events were followed by more formal recognition in Internet governance technical organizations, including the IETF and Internet Corporation for Assigned Names and Numbers (ICANN), of the relevance of human rights to their work. In 2015, the IRTF and IETF began to formally consider and study human rights by [chartering](#) the Human Rights Protocol Considerations Research Group (Internet Research Task Force b, n.d.). This group was chartered one year after the idea to “research how Internet standards and protocols impacted human rights values and develop guidelines that IETF engineers could use to mitigate the negative impact of their work” was presented to the IETF (Cath, 2021, p. 2-3). A 2016 update to ICANN bylaws also included human rights for the first time (Internet Corporation for Assigned Names and Numbers, n.d.). The work of the HRPC resulted in the publication of RFC 8280 in 2017 (Cath and ten Oever, 2017) and RFC 9620 in 2024 (Grover and ten Oever, 2024).

These policy developments are reflected in the analysis conducted in this paper. The entry of the second cohort of participants, many of which are part of organizations, into the email network around 2013 aligns with the increasing interest of international organizations in human rights online. This increasing interest is also reflected in the increase in yearly average support scores around 2013. Both of these developments predate, and potentially help explain, the chartering of the HRPC in 2015.

This study further contributes to our understanding of human rights in Internet governance by exploring the IETF/IRTF email network at scale. Much of the prior literature analyzing email data from the Internet governance community focuses on particular thematically-relevant email lists (Becker, ten Oever, and Nanni, 2022; Doty, 2015), with the notable exception of Welzl et al (2021). Our study shows how support for integrating human rights varies across email lists. If only emails on the list hrpc@irtf.org had been included in the analysis, many of those who are critical of integrating human rights considerations would have been left out of the analysis. This would have created a false picture of homogeneity and

consensus in the email network, while examining all email lists where the term “human rights” is used reveals the contention and variety of viewpoints in the debate.

This study also provides further nuance to our understanding of norm entrepreneurship and diffusion in Internet governance. It shows that a second cohort of network participants became active on the network around 2013. Compared to the first cohort, who joined the network between 1992 and 2012, the second cohort had greater representation of participants affiliated with organizations. The second cohort potentially included human rights norm entrepreneurs, distinguished by their relatively high number of supportive emails. Despite their level of activity, these norms may not have diffused to the rest of the network. Time series analysis of average levels of support among those who joined before 2013 reveals that their average level of support was consistently lower than those in the newer cohort. Network analysis reveals that the activity of potential norm entrepreneurs was primarily concentrated around the email list hrpc@irtf.org. The average level of support of emails sent to the list ietf@ietf.org, a more general purpose discussion list, declined from the first cohort to the second cohort. These results indicate that while norm entrepreneurs may have been active on the network, their levels of support for integrating human rights considerations in technical standards may not have diffused to the prior cohort of participants or to more general discussion in the IETF.

This study shows that the dynamics of norm development can be quantitatively measured by integrating annotation by a large language model with time series, regression, and network analysis; allowing for a more comprehensive overview than would be possible with exclusively qualitative tools. It provides insight into the development and diffusion of norms pertaining to integrating human rights considerations into technical standards on IETF/IRTF email networks.

Limitations

This study has several important limitations. Firstly, it only considers emails that use the term “human rights”. This may exclude discussions that are relevant to specific human rights. For example, if an email included “freedom of speech” but not “human rights” it would not be included in our dataset. According to Milan and ten Oever (2017), advocates in ICANN began frequently invoking “human rights” explicitly around 2014. If this discursive shift is also present in IETF/IRTF discussions, it could contribute to the increasing frequency of emails starting around 2015 seen in our dataset. Secondly, while a sample of GPT annotations were reviewed, not all annotations were reviewed and could contain unknown errors or biases.

References

Andrade, L., Braga, J., Pereira, S., Roque, R., & Santos, M. (2018). In-Person and Remote Participation Review at IETF: Collaborating Without Borders. *arXiv*. <https://arxiv.org/abs/1805.09140>

- Becker, C., ten Oever, N., & Nanni, R. (2022). *The Standardisation of Lawful Interception Technologies in the 3GPP: Interrogating 5G and Surveillance Amid US-China Competition*. SSRN. <http://dx.doi.org/10.2139/ssrn.4167105>
- British Broadcasting Corporation. (2013, August 20). *Edward Snowden: Timeline*. <https://www.bbc.com/news/world-us-canada-23768248>
- Benthall, Sebastian. 2015. “Testing Generative Models of Online Collaboration with BigBang.” In , 182–89. https://conference.scipy.org/proceedings/scipy2015/sebastian_benthall.html
- Cath, C., & Floridi, L. (2017). The Design of the Internet's Architecture by the Internet Engineering Task Force (IETF) and Human Rights. *Sci Eng Ethics*, 23(2), 449-468. <https://doi.org/10.1007/s11948-016-9793-y>
- Cath, C., & ten Oever, N. (2017). *Research into Human Rights Protocol Considerations* (Request for Comments: 8280). Internet Research Task Force. <https://datatracker.ietf.org/doc/rfc8280/>
- Cath, C. (2021). The technology we choose to create: Human rights advocacy in the Internet Engineering Task Force. *Telecommunications Policy*, 45(6), 102144.
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrocioni, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118.
- De Francisci Morales, G., Monti, C., & Starnini, M. (2021). No echo in the chambers of political interactions on Reddit. *Sci Rep*, 11(1), 2818. <https://doi.org/10.1038/s41598-021-81531-x>
- Doty, Nick. 2015. “Reviewing for Privacy in Internet and Web Standard-Setting.” In *Security and Privacy Workshops (SPW)*, 2015 IEEE, 185–192. IEEE. <https://ieeexplore.ieee.org/document/7163224/>
- Finnemore, M., & Sikkink, K. (1998). International Norm Dynamics and Political Change. *International Organization*, 52(4), 887–917.
- Garimella, K., Morales, G. D. F., Gionis, A., & Mathioudakis, M. (2018). Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1), 1-27.
- Grover, G. & ten Oever, N. (2024). *Guidelines for Human Rights Protocol and Architecture Considerations* (Request for Comments: 9620). Internet Research Task Force. <https://datatracker.ietf.org/doc/rfc9620/>
- Internet Engineering Task Force a. (n.d.). *Introduction to the IETF*. Retrieved April 7, 2025, from <https://www.ietf.org/about/introduction/>
- Internet Engineering Task Force b. (n.d.). *Mail Archive*. Retrieved April 7, 2025, from <https://mailarchive.ietf.org/arch/>
- Internet Engineering Task Force c. (n.d.). *IETF Datatracker*. Retrieved April 7, 2025, from <https://datatracker.ietf.org/>
- Internet Research Task Force a. (n.d.). *Overview*. Retrieved April 7, 2025, from <https://www.irtf.org/>
- Internet Research Task Force b. (n.d.). *Human Rights Protocol Considerations Research Group HRPC*. Retrieved April 7, 2025, from <https://www.irtf.org/hrpc.html>

- Internet Corporation for Assigned Names and Numbers. (n.d.). *Bylaws FOR INTERNET CORPORATION FOR ASSIGNED NAMES AND NUMBERS | A California Nonprofit Public-Benefit Corporation - ICANN*. Retrieved April 7, 2025, from <https://www.icann.org/en/governance/documents/bylaws-for-internet-corporation-for-assigned-names-and-numbers-a-california-nonprofit-public-benefit-corporation-icann-01-10-2016-en>
- Khare, P., Karan, M., McQuistin, S., Perkins, C., Tyson, G., Purver, M., Healey, P., & Castro, I. (2022). The Web We Weave: Untangling the Social Graph of the IETF. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1). 500-511. <https://doi.org/10.1609/icwsm.v16i1.19310>
- Krishnamurthy, V. (2019). Are Internet Protocols the New Human Rights Protocols? Understanding ‘RFC 8280 – Research into Human Rights Protocol Considerations.’ *Business and Human Rights Journal*, 4(1), 163–169. doi:10.1017/bhj.2018.30
- Liddicoat, J., & Doria, A. (2012). Human Rights and Internet Protocols: Comparing Processes and Principles. *Internet Society*. <https://www.internetsociety.org/resources/doc/2012/human-rights-and-internet-protocols-comparing-processes-and-principles/>
- Marino, M. C. (2020). *Critical code studies*. MIT Press.
- Milan, Stefania, and Niels ten Oever. 2017. “Coding and Encoding Rights in Internet Infrastructure.” *Internet Policy Review* 6 (1)
- Mueller, M. L., & Badiei, F. (2019). Requiem for a Dream: On Advancing Human Rights via Internet Architecture. *Policy & Internet*, 11(1), 61-83. <https://doi.org/10.1002/poi3.190>
- Nanni, Riccardo. “Digital Sovereignty and Internet Standards: Normative Implications of Public-Private Relations among Chinese Stakeholders in the Internet Engineering Task Force.” *Information, Communication & Society* 0, no. 0 (October 1, 2022): 1–21. <https://doi.org/10.1080/1369118X.2022.2129270>.
- Open AI Platform. (n.d.). GPT-4o. Retrieved April 7, 2025, from <https://platform.openai.com/docs/models/gpt-4o>
- Radu, R., Kettemann, M. C., Meyer, T., & Shahin, J. (2021). Normfare: Norm entrepreneurship in internet governance. *Telecommunications Policy*, 45(6). <https://doi.org/10.1016/j.telpol.2021.102148>
- Riedl, C., Grad, T., & Lettl, C. (2024). Competition and Collaboration in Crowdsourcing Communities: What Happens When Peers Evaluate Each Other? *Organization Science*, 35, 1957–1976. <https://doi.org/10.1287/orsc.2021.15163>
- ten Oever, N., Milan, S., & Beraldo, D. (2020). Studying Discourse in Internet Governance through Mailing-list Analysis. In D. L. Cogburn, L. DeNardis, N. S. Levinson, & F. Musiani (Eds.), *Research Methods in Internet Governance*. Cambridge, MA: MIT Press. <https://direct.mit.edu/books/oa-monograph/4936/chapter/625914/Studying-Discourse-in-Internet-Governance-through>

- ten Oever, N. (2023). Norm conflict in the governance of transnational and distributed infrastructures: the case of Internet routing. *Globalizations*, 20(1), 184-200.
- UN Human Rights Council. (2011). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue* (A/HRC/17/27). United Nations General Assembly. https://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf
- UN Human Rights Council. (2012). *The promotion, protection and enjoyment of human rights on the Internet : resolution / adopted by the Human Rights Council* (A/HRC/RES/20/8). United Nations General Assembly. <https://digitallibrary.un.org/record/731540?ln=en&v=pdf>
- Welzl, M., Oepen, S., Jaskula, C., Griwodz, C., & Islam, S. (2021). Collaboration in the IETF: An Initial Analysis of Two Decades in Email Discussions. *ACM SIGCOMM Computer Communication Review*, 51(3), 29–32. <https://dl.acm.org/doi/pdf/10.1145/3477482.3477488>