APSA Preprints
Version 2
April 28th, 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This content is an early research output and has not been peer-reviewed at the time of posting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: Changing Norm Entrepreneurship

Abstract

Introduction

The extent human rights considerations are integrated into Internet standards shapes how people can exercise their rights online. For example, an Internet standard that leaks information about online communications can compromise the right to privacy. Prior literature provides evidence on how values and rights are (or are not) considered in technical processes and organizations. However, few examine these processes at scale. We focus on the role of norm entrepreneurship, which refers to the process by which individuals or groups actively promote new social norms, in advancing the consideration of human rights.

Methods

We analyze all emails in the public Internet Engineering Task Force 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

We show how norm entrepreneurship around human rights considerations in Internet standards unfolded over the course of 33 years. We find evidence of a cohort of norm entrepreneurs emerging around 2013. Based on our network analysis, it appears that they had limited success in diffusing norms to the broader Internet Engineering Task Force community.

Conclusion

This study provides insight into normative processes. It builds on prior literature using email data to map normative evolution with a novel dataset from the IETF Mail Archives. Understanding

these Internet-mediated interactions can help explain policy-making outcomes in Internet governance.

Keywords

Norms, Large Language Models, Network Analysis, Human Rights, Internet Governance

Introduction

Integrating human rights into Internet infrastructure is controversial. Some in the technical community argue that technology is neutral; others argue that engineering decisions impact human rights (Krishnamurthy, 2019; Zalnieriute and Milan, 2019). With the rise of artificial intelligence (AI), and associated risks to human rights (Teo, 2024; Knodel, 2025), these debates are increasingly consequential. They also shape what digital regulation (Moss, 2025) is technically feasible. This study analyzes the online processes surrounding the integration of human rights considerations into technical standards. It applies time series, regression, and network analysis to a novel dataset of emails sent to public Internet Engineering Task Force and Internet Research Task Force mailing lists. We analyze the support for or against integrating human rights considerations into technical standards expressed in these emails. This analysis reveals how processes of norm entrepreneurship and diffusion evolved in the Internet Engineering Task Force and Internet Research Task Force community. Norm entrepreneurship refers to the process by which individuals or groups actively promote new social norms. One example of this is the policy advocacy that led to the inclusion of human rights in the Internet Corporation for Assigned Names and Numbers' (ICANN) bylaws (ten Oever, 2018).

The Internet Engineering Task Force (IETF) is a standards development organization (SDO). Internet SDOs create and refine the standards that allow different entities on the Internet to connect and communicate; 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.). The mission statement goes on to declare that:

> The mission of the IETF is to produce high quality, relevant technical and engineering documents that influence the way people design, use, and manage the Internet in such a way as to make the Internet work better. These documents include protocol standards, best current practices, and informational documents of various kinds… The IETF community wants the Internet to succeed because we believe that the existence of the Internet, and its influence on economics, communication, and education, will help us to build a better human society…The Internet isn't value-neutral, and neither is the IETF. We want the Internet to be useful for communities that share our commitment to openness and fairness. We embrace technical concepts such as decentralized control,

> edge-user empowerment and sharing of resources, because those concepts resonate with the core values of the IETF community. These concepts have little to do with the technology that's possible, and much to do with the technology that we choose to create (Alvestrand, 2004).

In this mission statement, the term "better" refers to effective technical operation and societal attributes. The statement also acknowledges the role of values in deciding what technology to create. 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; Mueller and Kiernan, 2020), the IETF is a key player in this space.

Individuals participate in the IETF and IRTF in their personal capacity as volunteers. 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 standards through an open process of collaboration (Andrade et al, 2018). 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. This gender bias is also reflected in the broader field of Internet governance: scholars bearing male names are cited and referred to as experts more often than scholars with female names, despite gender parity in general Internet governance conference attendance (Rosa et al, 2024). 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.). While IETF participants tend to come from industry, many IRTF participants are members of the academic research community or non-governmental organizations.

The IETF and IRTF are committed to making their work processes open. The IETF Mail Archive contains emails documenting debates and collaboration among IETF and IRTF participants on 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, contributing a comprehensive overview of these debates and exploring their evolution at scale.

Literature Review

The importance of protecting the Internet as a rights-enabling space is recognized by international organizations and 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) and prompting further discussion of rights online.

Surveillance is one area, among others (Broeders and Sukumar, 2024; Bradshaw and DeNardis, 2019), where technical standards impact human rights. Standards enable people to use devices, like smart-phones and computers, to communicate online. They can leak more or less information about the communication depending on decisions made by the developers (Varon, 2016). If the standard shares more information, it is easier to surveil communications (such as planning a protest). If standards share less information, surveillance is harder. Information gathered from network traffic could lead a state to label a person as threatening, potentially leading to that person's "torture, extra-judicial killing, or detention" (Grover and ten Oever, 2024). This example illustrates how technical decisions on standards impact peoples' ability to exercise their human rights.

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 researches:

> [W]hether 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.)

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). RFC 8280, "Research into Human Rights Protocol Considerations" was one of the first times an Internet technical body formally acknowledged that their engineering decisions have human rights consequences (Krishnamurthy, 2019). This RFC was followed by RFC 9620 "Guidelines for Human Rights Protocol and Architecture Considerations" (Grover and ten Oever, 2024). RFC 9620 outlines threats to human rights and proposes methods for conducting human rights reviews of protocols or standards. 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. 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).

Despite these developments, many in the IETF believe that engineers should prioritize maintaining the neutrality of technology and the ability of any device to join the network and connect with other devices (Cath and Floridi, 2017; ten Oever, 2023) to the exclusion of other

objectives. Resistance to integrating human rights into standards primarily takes the form of conflict over norms around the 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'. In this study, we use the term "diffusion" to refer to influencing discussion and individual support for integrating human rights considerations into internet standards. We are not referring to resultant changes in internet standards. Human rights norm entrepreneurs 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 values. For example, one IETF engineer interviewed by Cath (2021, p. 11) stated that "All that [standards] needs to be consistent and predictable, you cannot create a suite of human rights considerations for protocols, then a completely different suite for whatever because they need to be interoperable. Everything on the Internet needs to be interoperable in order to make sense." Here, integrating human rights is seen as potentially jeopardizing interoperability (Cath, 2021), leading to resistance.

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 (e.g. Khare et al, 2022) 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 map changes in ICANN policy to changes in cohorts. Nanni (2022) compares how communication networks 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. Our study builds on this prior work by using email data to examine normative evolution.

Methods

We integrate 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 time series to reveal changes in activity level and average support. Sender email accounts are also categorized by top level domain (e.g., .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 their most frequent top level domain, their average level of support over time, and their network characteristics.

Data

This study uses a dataset of all publicly available IETF and IRTF emails containing the phrase "human rights" in the subject or body 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. A detailed explanation of our data collection process is available in the appendix and the Python code (co-authored with ChatGPT 4o and Gemini in Google Colab) used to scrape the emails is available in the supplementary materials.

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 focal email contained only "That's a bad idea" we use prior emails from the chain for context: what idea are they referring to and whether it is relevant to human rights. Emails were assigned the following scores:

> 1 = the email is relevant to human rights and the author is supportive of integrating human rights considerations into technical standards;
> -1 = the email is relevant to human rights and the author opposes integrating human rights considerations into technical standards;

0 = 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. The full text of the prompt is available in the appendix and supplementary materials. Below is a portion of the prompt that has been edited for readability:

Role and Task
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.

Definitions
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" [Wikipedia] [...]

Annotation Criteria
[...] 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.[...]

Neutral
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 [...]

Supportive
1 = Email or context are relevant [...] and the author expresses support for integrating human rights considerations into technical standards. [...]

Opposing
-1 = Email or context are relevant to [...] and the author expresses opposition to integrating human rights considerations into technical standards. [...]

Annotation Examples

Neutral
[...] I have done a review of the HRPC [Human Rights Protocol Considerations] guidelines doc as well. Please see the attached.
[...] 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.

Supportive
[...] 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 involve – most would agree that using communication technology to disclosure an officials 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.).
[...] 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.'.

Opposing
[...] 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. [...]

Further Guidance
Be precise and consistent in scoring. Before you provide an answer, explain why the email was given a particular score. [...]

This prompt leverages in-context learning by providing examples of how to annotate emails (Lightman et al, 2024).

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%).

Our dataset is available in the supplementary materials. We removed fields such as the sender's full email address, the message text, and the written GPT score explanation to avoid identifying individuals. Some individuals appear to have used their own emails as addresses for mailing lists. We assume that these individuals were aware that they were sharing their email publicly and therefore do not remove the mailing list address from our dataset. The dataset contains URLs pointing to each email and each context email, allowing the full dataset to be recreated by scraping the emails, assuming the emails are still publicly available on the IETF Mail Archive at the time of recreation.

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

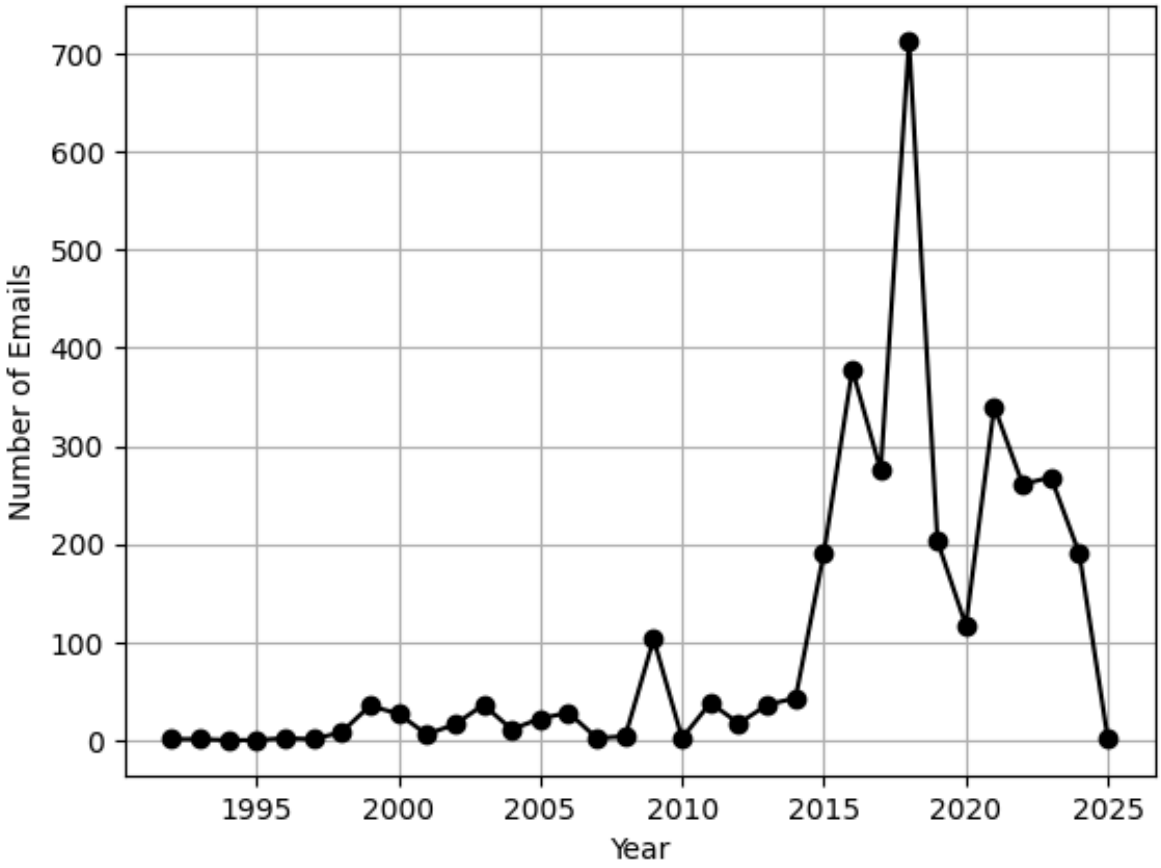
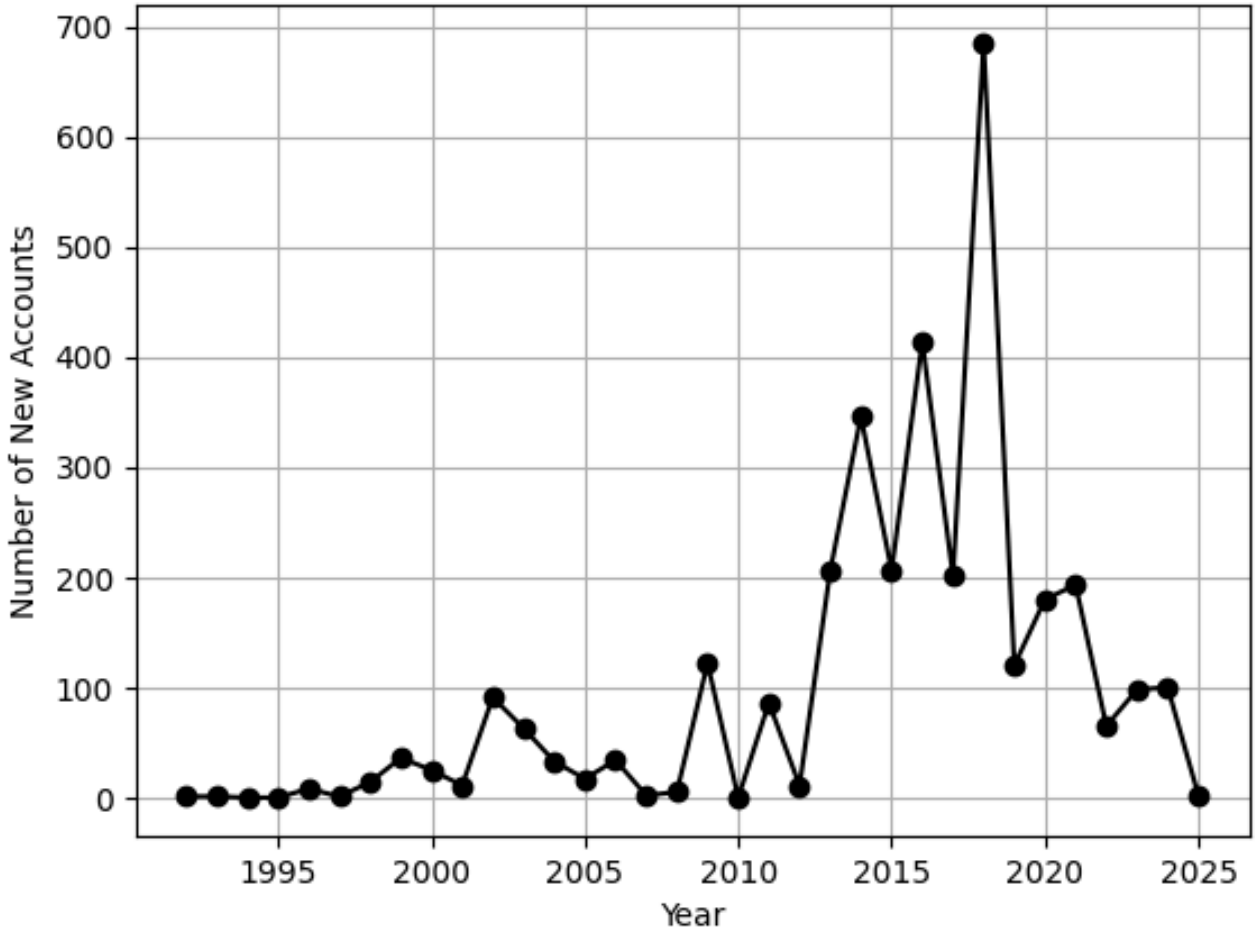


Figure 1 shows spikes in email frequency for the years 2016, 2018, and 2021. Figure 2 displays the number of new email accounts joining the network each year.

Figure 2

Number of New Accounts Joining per Year

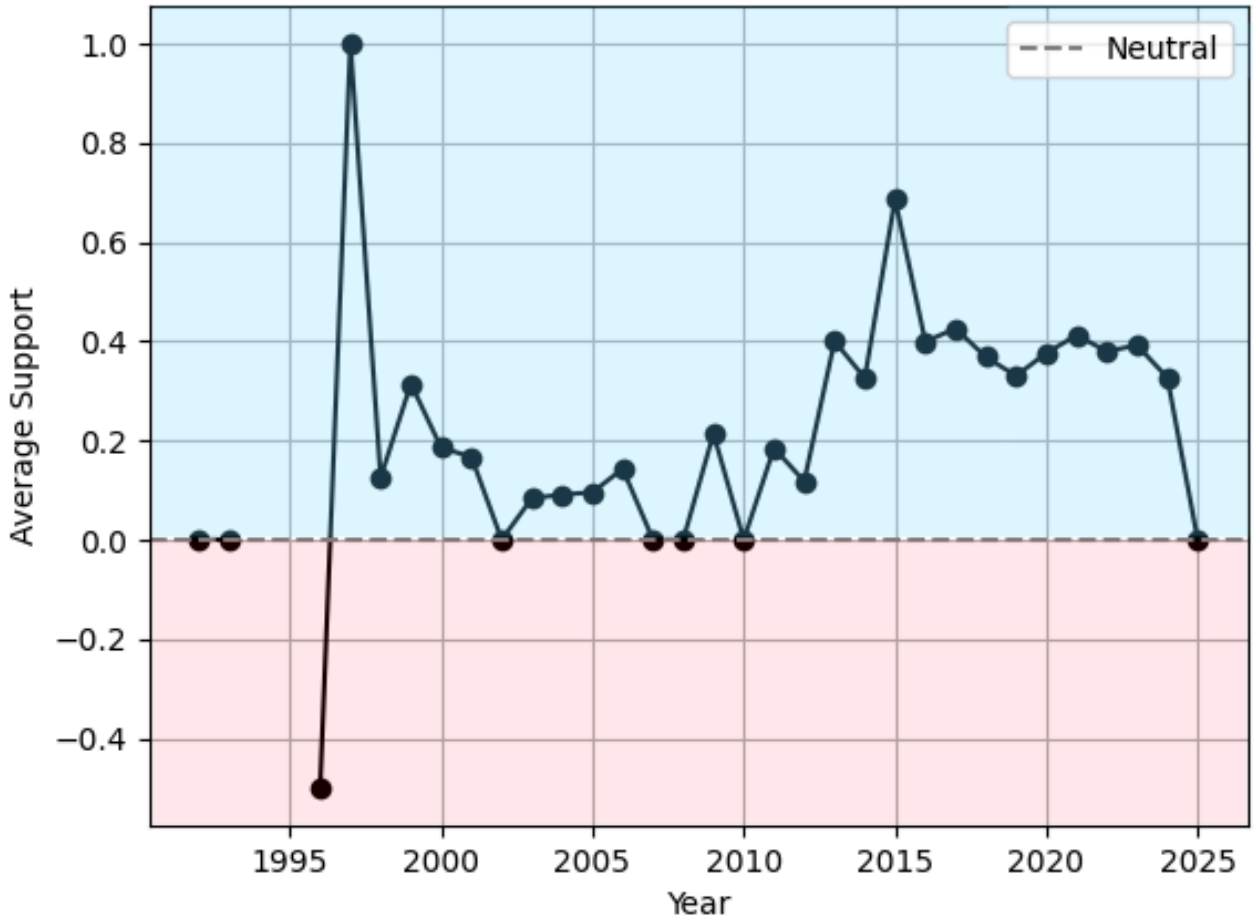


This figure shows that around 2013 many new email accounts began joining the network.

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.

Figure 4

Number of Emails per Year for the Top 5 TLDs

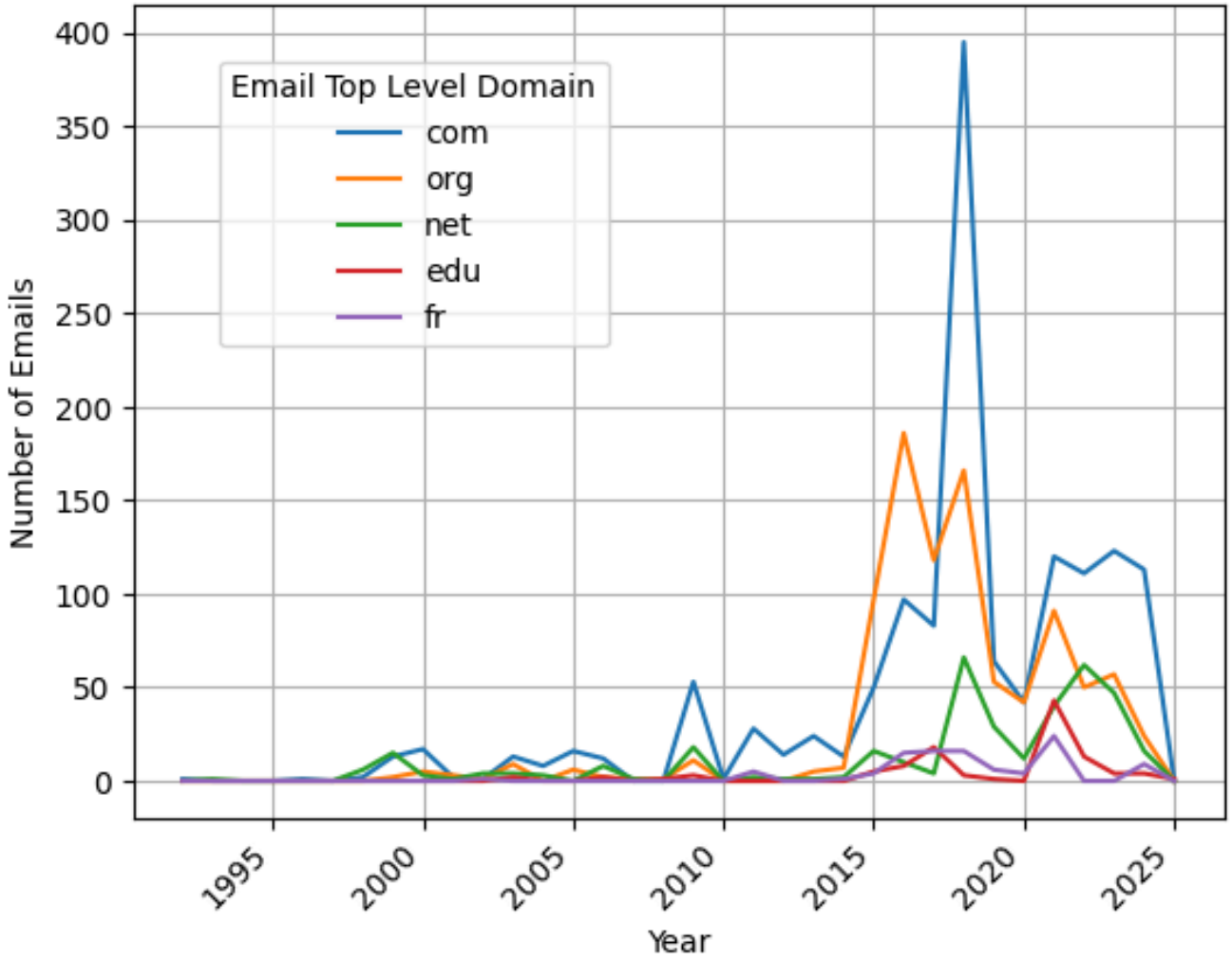


Figure 4 shows that for most years com is the most prevalent top level domain, with the exception of the years 2015, 2016, 2017, 2020, and 2025. In 2025, edu was the most prevalent TLD (though this is likely due to the limited data for 2025). In the other years, org was either the most prevalent TLD 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, 2016, and 2017. Trithara (2024, p. 2), citing Abbott and Snidal (2009, p. 64-5), summarizes civil society as including "a diverse range of value- and interest-driven actors that seek to shape the rules of social life."

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, and the number of total emails sent from that domain.

Table 1

Support and Frequency by Top Level Domain

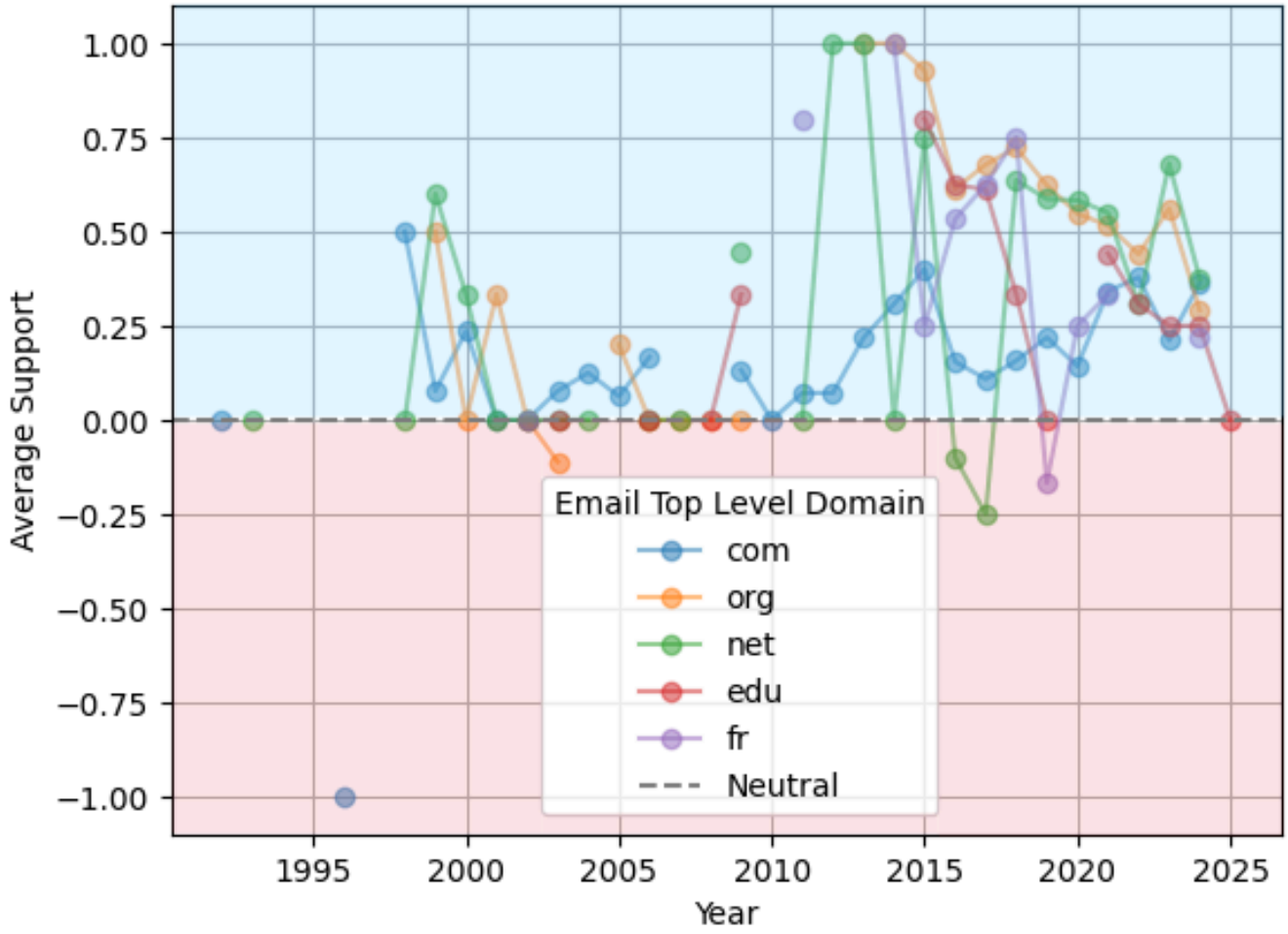
Email Top Level Domain	Average Support Score	Total Number of Emails Sent
edu	0.44	108
com	0.22	1417
net	0.47	372
org	0.62	937
fr	0.46	101

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 displayed the lowest average support for integrating human rights considerations. Authors writing from addresses ending org 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

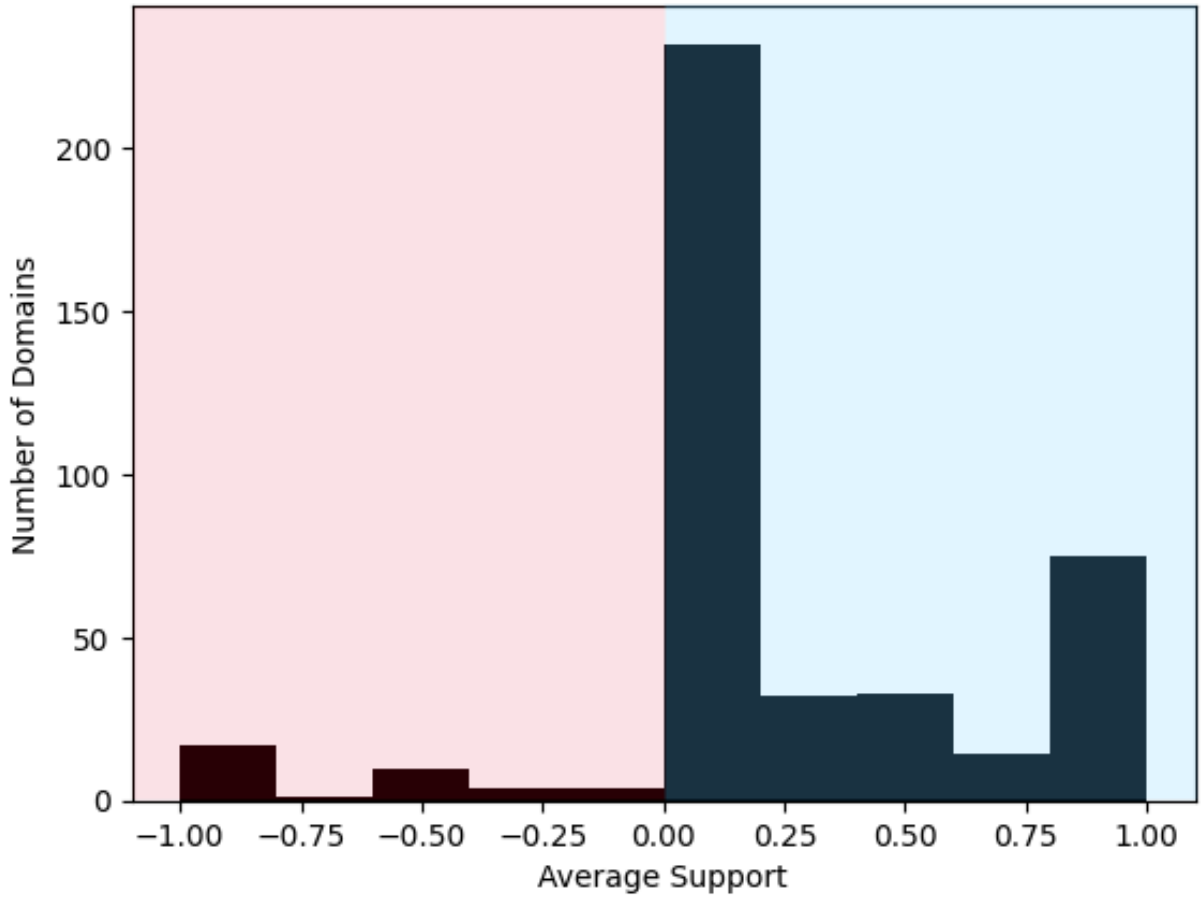


For most years, emails ending in net, org, or edu showed the highest level of support, while emails ending in com showed the lowest level of support. When comparing average scores for emails ending in org across years, there is a marked increase in average scores in 2013 and after compared to the time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

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

Figure 6

Average Support by Domain



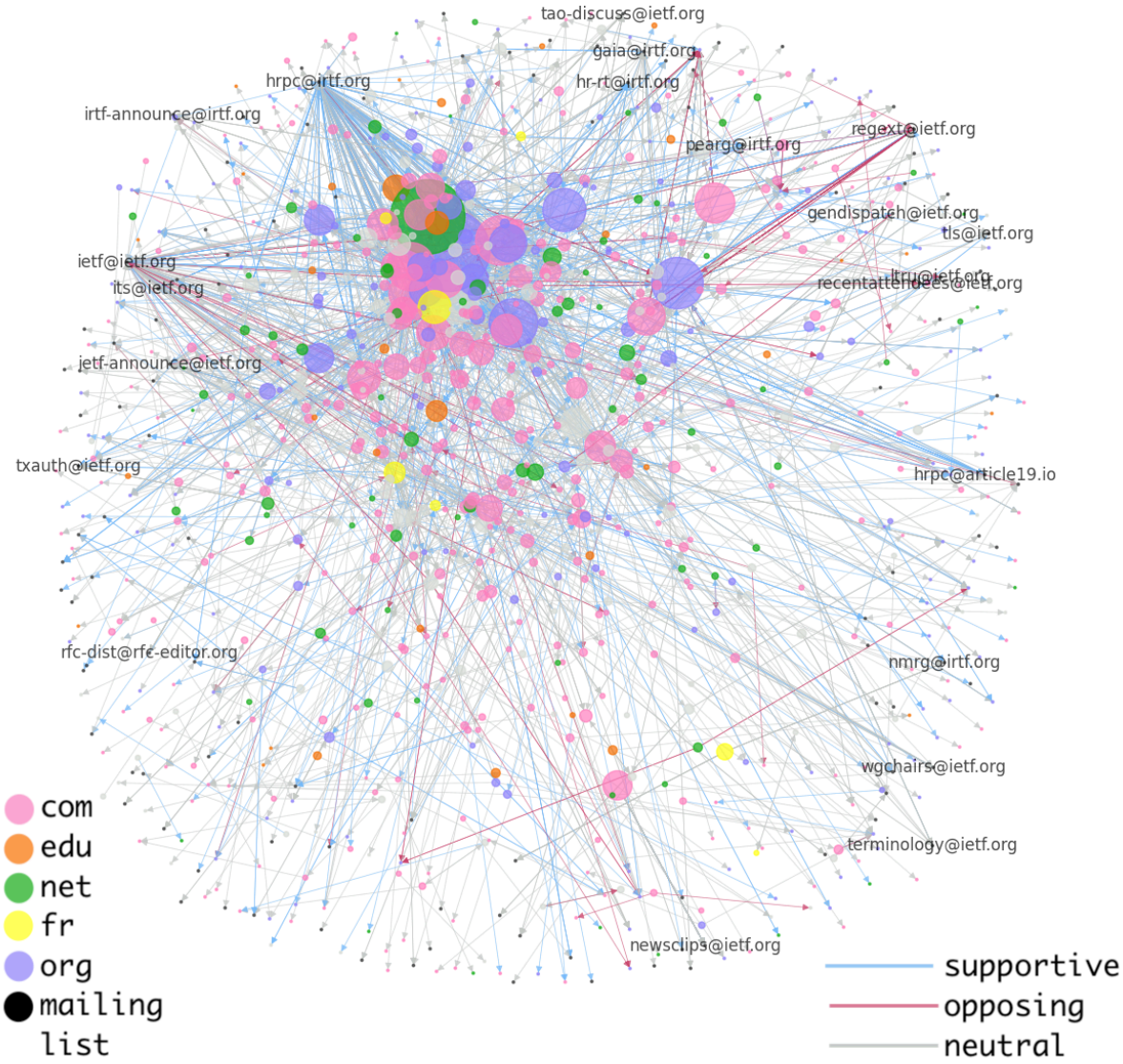
Most domains have an average support of between 0 and .25, indicating a neutral to slightly positive stance on integrating human rights considerations into technical standards. Overall, a higher number of domains have a positive average support. This indicates the promotion of human rights considerations 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. 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. Edges are colored based on the support score of the email: -1 is red, 0 is grey, and 1 is blue. Email accounts associated with individuals are not labeled with their addresses for privacy. Mailing list addresses 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 mailing list (hrpc@irtf.org), used for the work of the HRPC research group, and the IETF-Discussion 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 two phases of norm entrepreneurship in the network. Can we quantify these phases and separate participants into two cohorts based on when they joined the network? We explore such a cohort analysis, 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.

Regression

This section explores the extent to which affiliation and email activity are associated with support. It addresses the questions: are email authors associated with organizations or universities more supportive of human rights than authors with other affiliations? And are authors who send more emails more supportive of human rights than authors who send fewer emails? Top level domain (TLD), a proxy for affiliation, and out-degree (the number of emails sent from an account) were regressed on support using OLS regression from the statsmodels module in Python. In this data structure for regressing domain on support the rows are individual emails and the columns are attributes of each email, including its support score and the TLD of the sender email address. No top level domain was a significant predictor of support, indicating that affiliation does not significantly impact 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. Out-degree was calculated for each email address based on how many emails it had sent for all years to how many other email addresses. For example, a sender address sending two emails, each with two recipients, would have an out-degree of four. In the data structure for this regression, the rows were unique email addresses and the columns were their out-degree in the network and the average support across emails they sent. For example, if a sender address sent one email that was supportive (1) and one email that was neutral (0), they would have an average support of 0.5. Out-degree was then regressed on average 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. There were no individual years in which the relationship was significant.

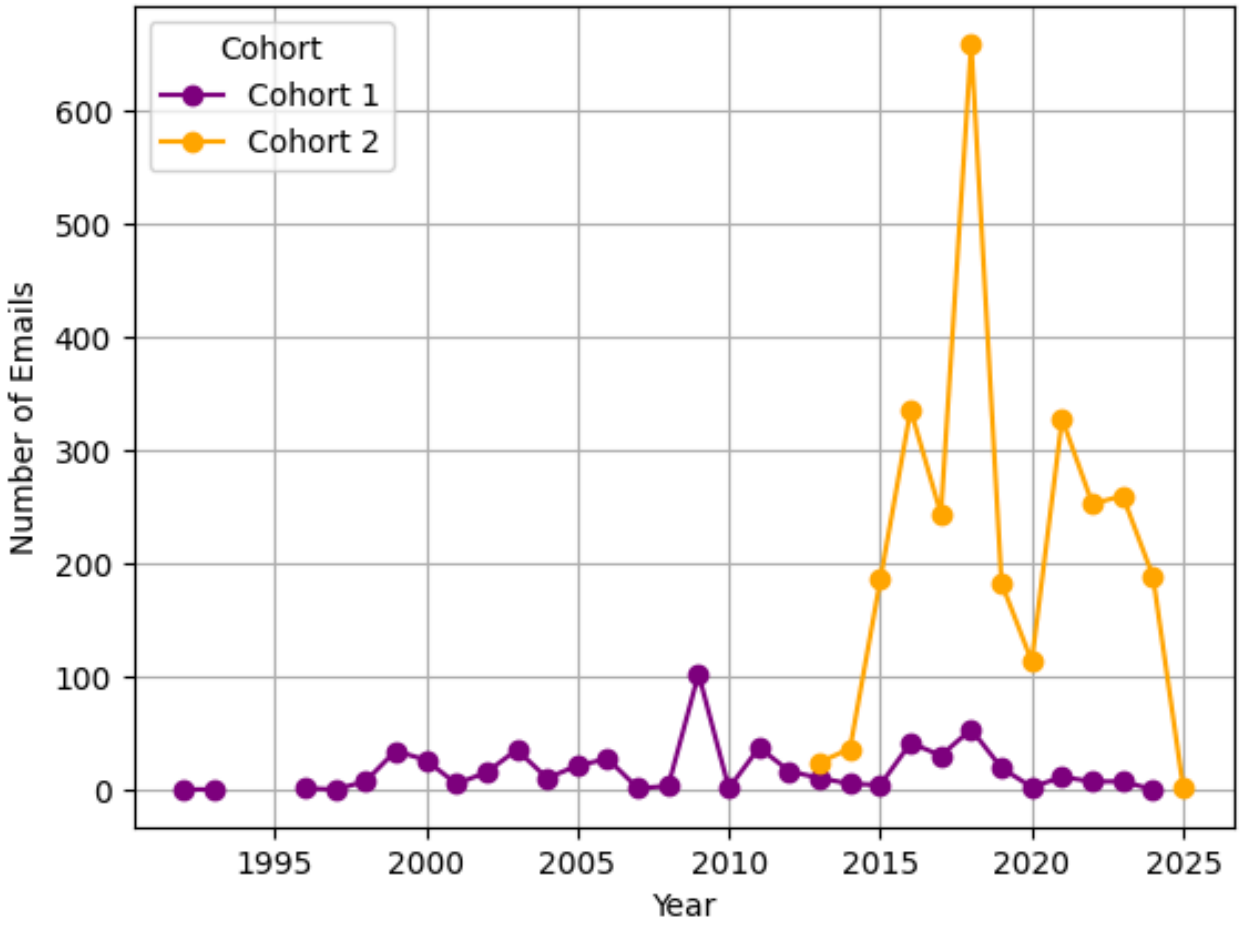
The relationship between out-degree and average support was then assessed for the time period before 2013 and the time period during and after 2013. The year 2013 was selected due to the marked increase in activity in the years following and the increase in average support for integrating human rights. 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 still small (coefficient is 0.0032). This indicates that there was a shift in communications patterns among participants who joined before 2013 and those who joined during and after 2013. For those who joined before, the number of emails they sent is not related to their level of support. For those who joined during and after 2013, the number of emails they sent is related to their level of support. This could indicate a deliberate effort by norm entrepreneurs to spread support for human rights by sending many emails.

Cohort Comparison

This section examines the question: how have network participation and communication patterns evolved over time? Based on the regression results and email activity, 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 compared by the most frequent top level domain, the average level of support over time, and their network characteristics.

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

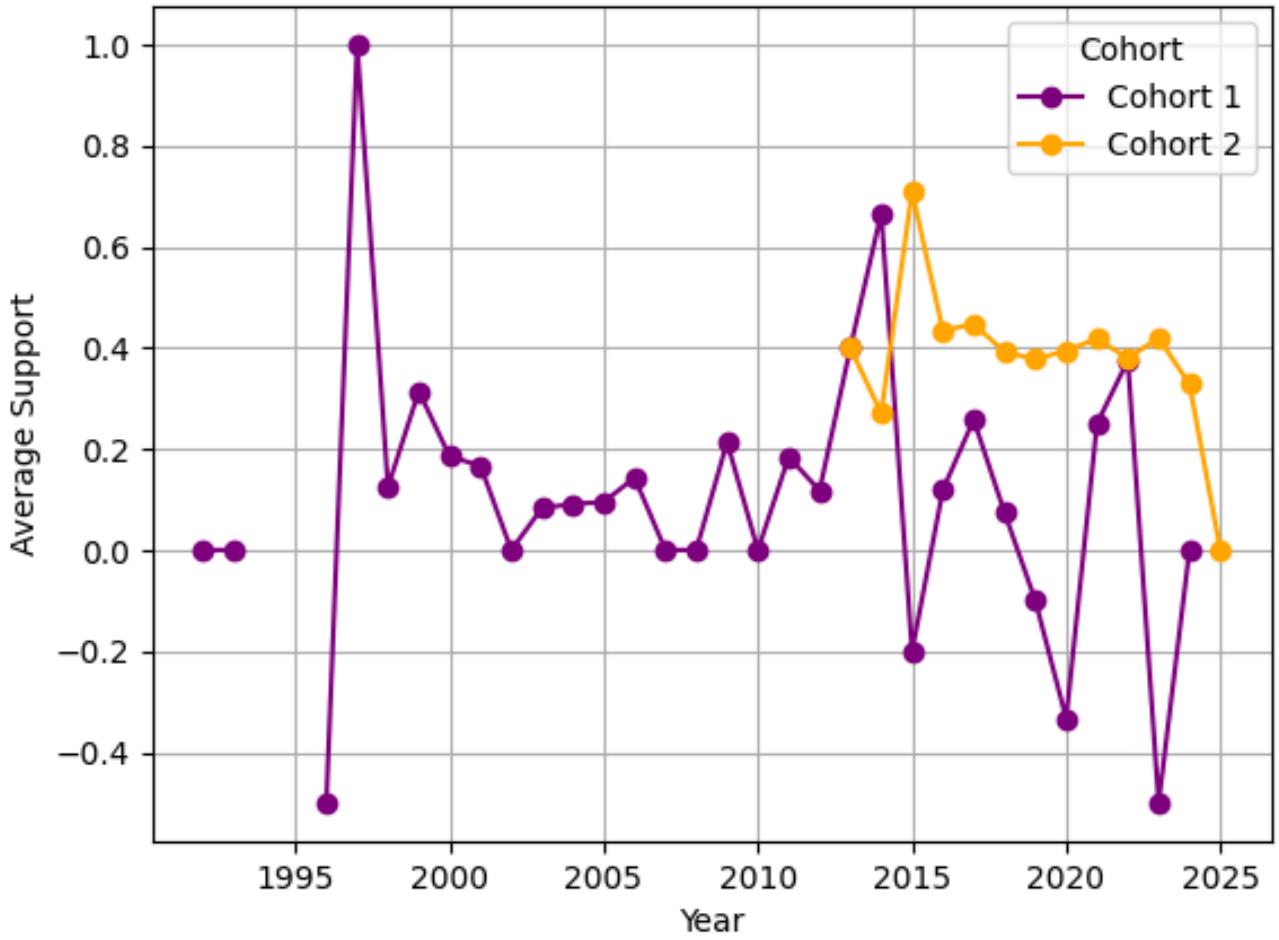
it	28	0.29
uk	27	0.22
ch	25	0.48
ca	21	0.52

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.

Figure 9 shows the yearly average support for each cohort.

Figure 9

Yearly Average Support by Cohort

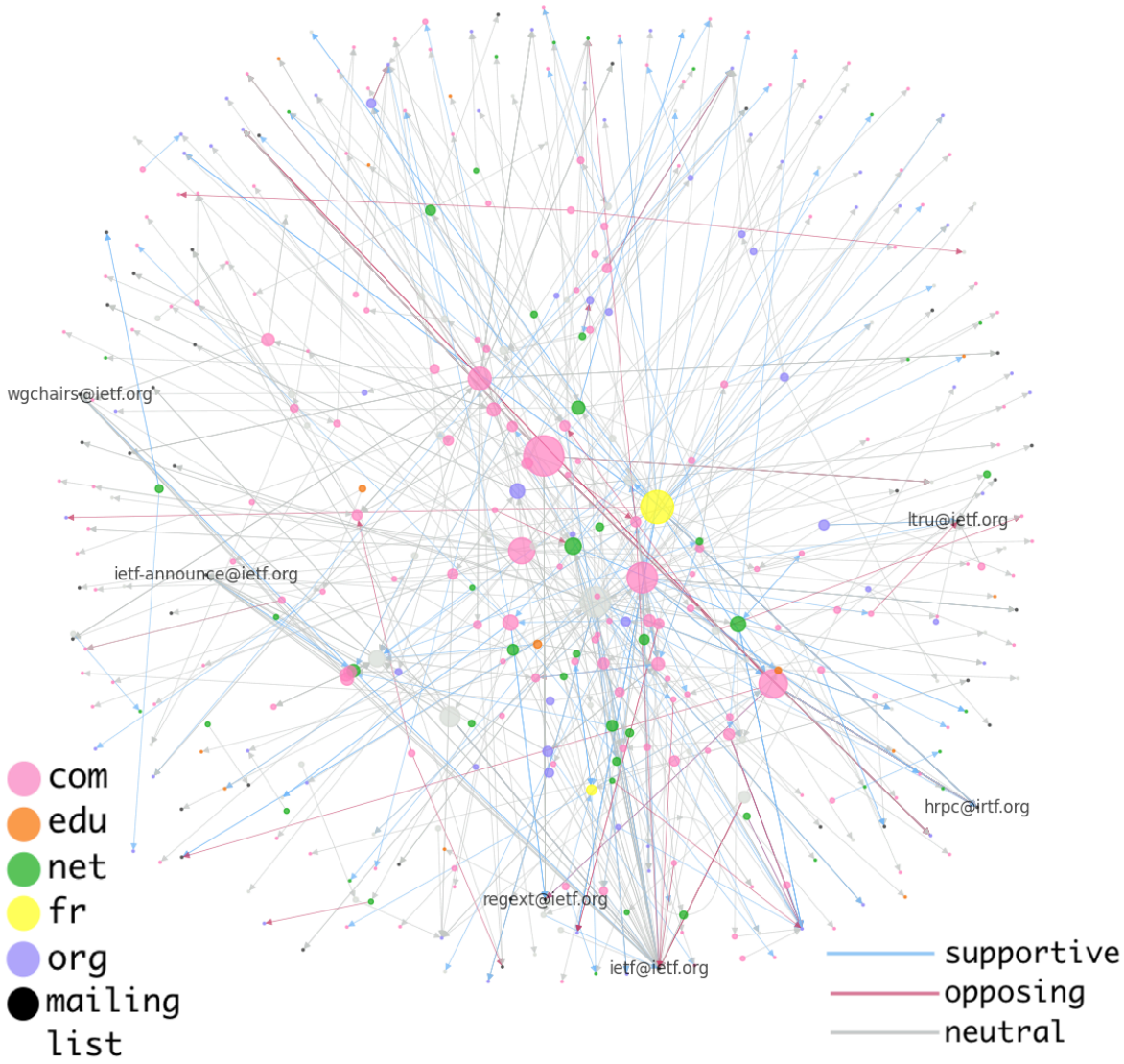


With the exception of 2014, the average support of cohort 2 is equal to or greater than the average support of cohort 1 for each year after 2013.

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

Figure 10

Cohort 1 Email Network



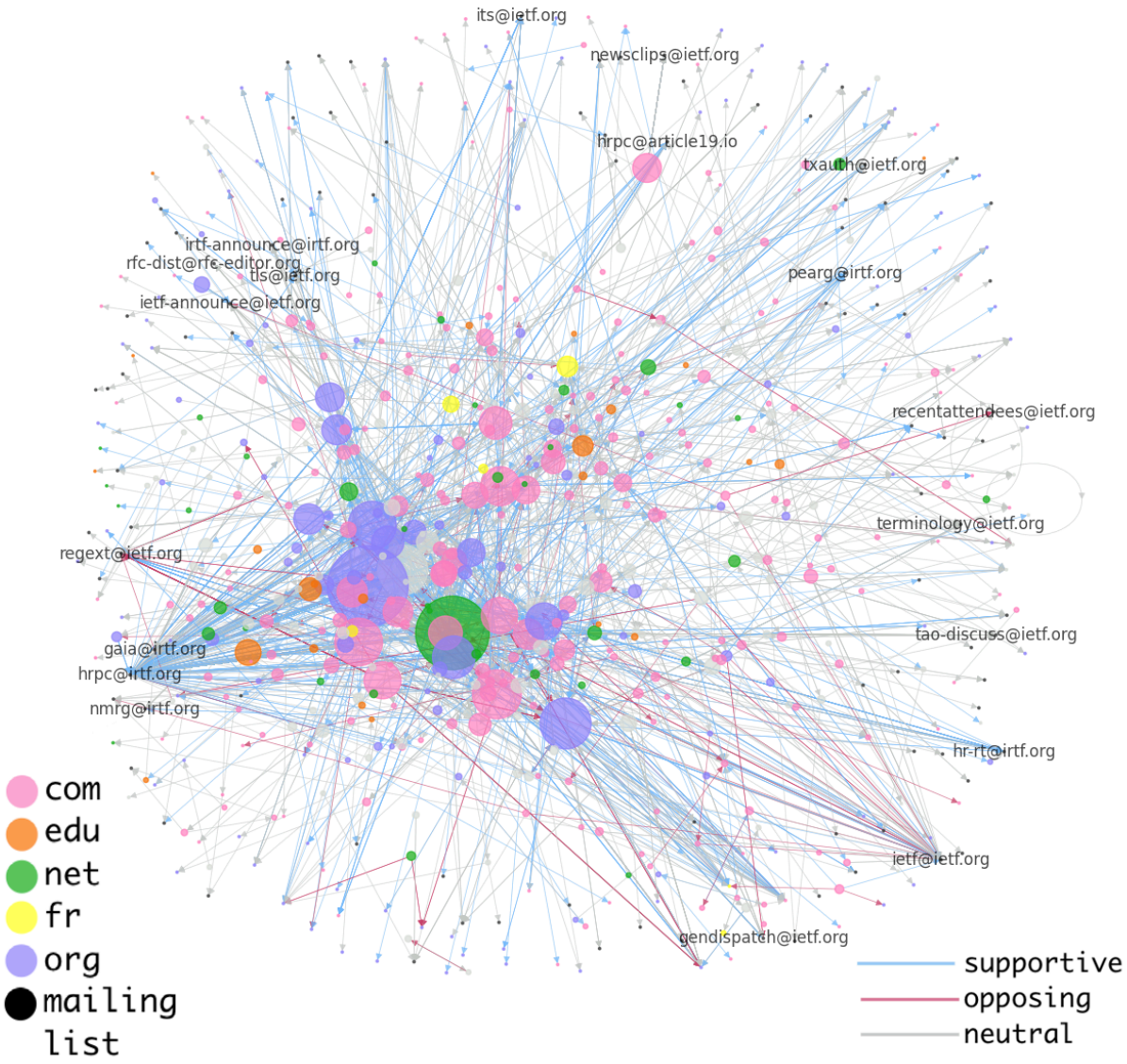
Note. Nodes are colored by top level domain and if they are a mailing list: addresses ending in com are pink, addresses ending edu are orange, net is green, fr is yellow, org is purple, and mailing lists are black. Nodes are sized based on the number of outgoing emails. Edges are colored based on the 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.

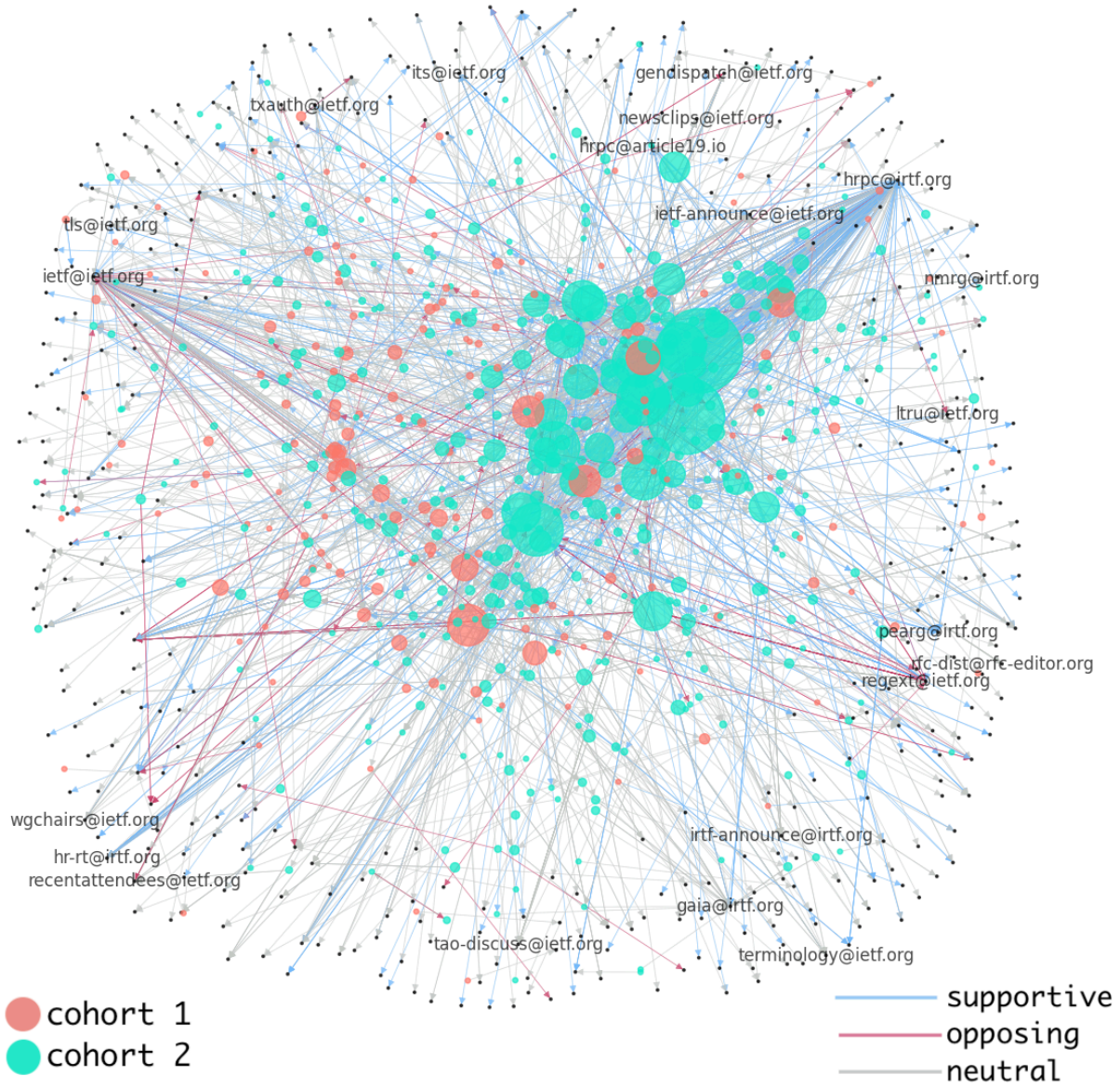
The network for cohort 2 displays emails sent by email accounts that joined the network during or after 2013. If emails were sent from cohort 2 accounts to cohort 1 accounts, these emails would be included in the network.

There are several notable differences between the cohort 1 and 2 networks. Firstly, the presence of accounts with the top level domain org 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.

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.

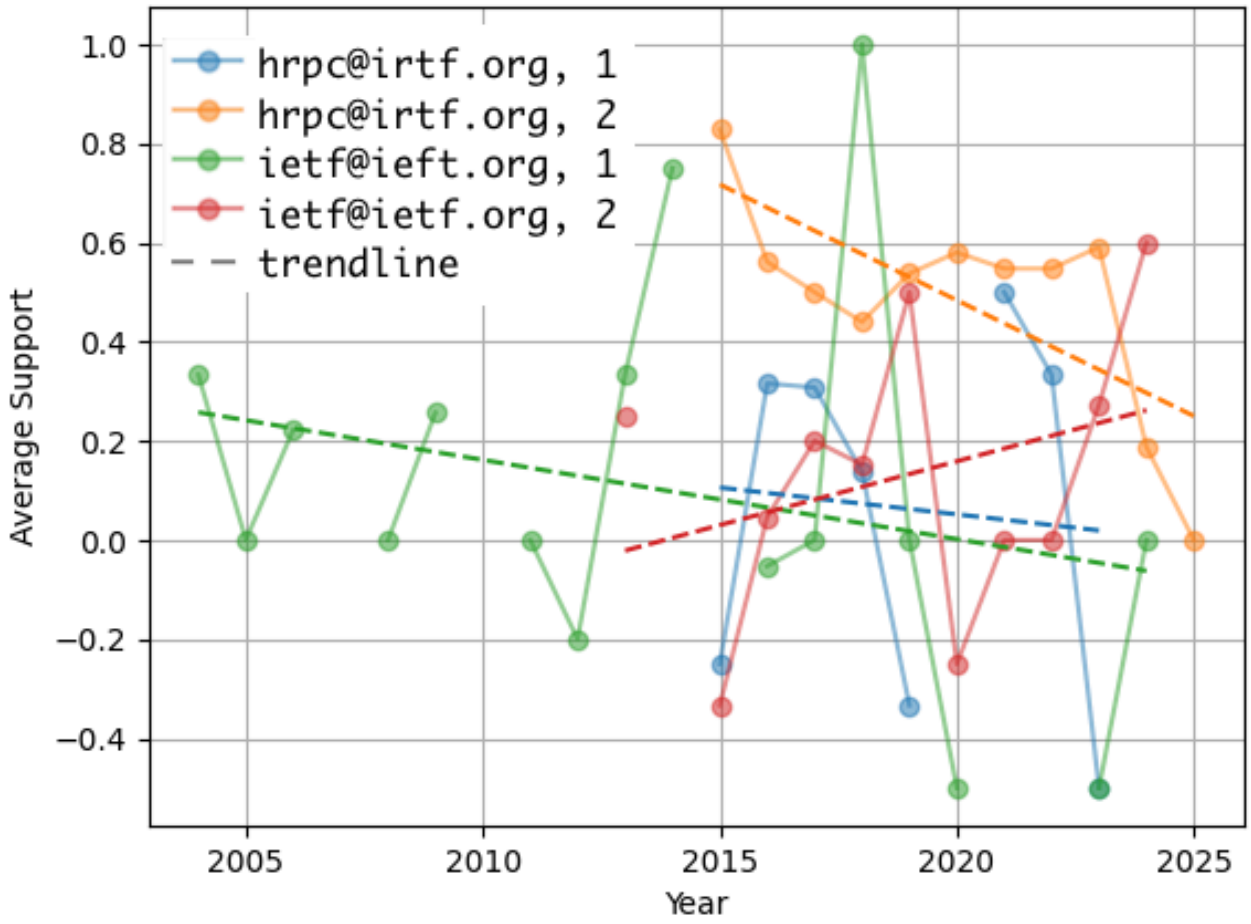
Nodes belonging to cohort 1 are colored orange, nodes belonging to cohort 2 are colored aqua. This network shows that many of the high out-degree cohort 2 nodes (email accounts that send

many emails) are displayed near the hrpc@irft.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

This study furthers 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. The entry of this second cohort 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.

Compared to the first cohort, who joined the network between 1992 and 2012, the second cohort had greater representation of participants affiliated with organizations. It also potentially included human rights norm entrepreneurs, distinguished by their relatively high number of supportive emails. Despite this 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 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. 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.

Limitations

This study has several limitations. First, 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. Second, while a sample of GPT annotations were reviewed, not all annotations were reviewed and could contain unknown errors or biases.

References

Abbott, K. W., & Snidal, D. (2021). The governance triangle: Regulatory standards institutions and the shadow of the state. In *The spectrum of international institutions* (pp. 52-91). Routledge.

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

Alvestrand, H. (2004). *A Mission Statement for the IETF* (Request for Comments: 3935). Internet Engineering Task Force. https://www.rfc-editor.org/rfc/pdfrfc/rfc3935.txt.pdf

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

Bradshaw, S., & DeNardis, L. (2019). Privacy by Infrastructure: The Unresolved Case of the Domain Name System. *Policy & Internet*, *11*(1), 16-36. https://doi.org/10.1002/poi3.195

British Broadcasting Corporation. (2013, August 20). *Edward Snowden: Timeline*. https://www.bbc.com/news/world-us-canada-23768248

Broeders, D., & Sukumar, A. (2024). Core concerns: The need for a governance framework to protect global Internet infrastructure. *Policy & Internet*. https://doi.org/10.1002/poi3.382

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.

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.

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

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

Knodel, M., Fábrega, A., Ferrari, D., Leiken, J., Hou, B. L., Yen, D., de Alfaro, S., Cho, K. & Park, S. (2025). How To Think About End-To-End Encryption and AI: Training, Processing, Disclosure, and Consent. *arXiv preprint arXiv:2412.20231*. https://arxiv.org/pdf/2412.20231

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/

Lightman, H., Kosaraju, V., Burda, Y., Edwards, H., Baker, B., Lee, T., Leike, J., Schulman, J., Sutskever, I., & Cobbe, K. (2024). Let's verify step by step. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/pdf?id=v8L0pN6EOi

Milan, Stefania, and Niels ten Oever. 2017. "Coding and Encoding Rights in Internet Infrastructure." *Internet Policy Review 6* (1)

Moss, G. (2025). Digital Regulation and Questions of Legitimacy. *Policy & Internet*. https://doi.org/10.1002/poi3.433

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

Mueller, M. L., & Kiernan, C. J. (2020). *Standardizing Security: Surveillance, Human Rights, and TLS 1.3*. Georgia Institute of Technology. https://www.internetgovernance.org/wp-content/uploads/StandardizingSecurity-TLS.pdf

Nanni, R. (2022). Digital sovereignty and Internet standards: normative implications of public-private relations among Chinese stakeholders in the Internet Engineering Task Force. *Information, Communication & Society*, *25*(16), 2342–2362. 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

Rosa, F. R., Anastácio, K., Pereira de Jesus, M. V., & Veras, H. J. A. (2024). Citation politics: The gender gap in internet governance. *Telecommunications Policy*, *48*(5). https://doi.org/10.1016/j.telpol.2024.102734

ten Oever, N. (2018). Productive Contestation, Civil Society, and Global Governance: Human Rights as a Boundary Object in ICANN. *Policy & Internet*, *11*(1), 37-60. https://doi.org/10.1002/poi3.172

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.

Teo, S. A. (2024). Artificial intelligence and its 'slow violence' to human rights. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00547-x

Trithara, D. (2024). Agents of platform governance: Analyzing U.S. civil society's role in contesting online content moderation. *Telecommunications Policy*, *48*(4). https://doi.org/10.1016/j.telpol.2024.102735

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

Varon, Joana. (Direction). (2016). *Net of Rights* [Film]. Article19; Coding Rights. https://hrpc.io/

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

Zalnieriute, M., & Milan, S. (2019). Internet Architecture and Human Rights: Beyond the Human Rights Gap. *Policy & Internet*, *11*(1), 6-15. https://doi.org/10.1002/poi3.200