Submission to Show me the numbers: Workshop on Analyzing IETF Data (AID), 2021

Using Complex Systems Analysis to Identify Organizational Interventions

Sebastian Benthall - Information Law Institute - NYU School of Law

The IETF has historically emphasized the individuality of its participants. Consider this summary statement from "The Tao of IETF": "Participation in the IETF or of its WGs is not fee-based or organizationally defined, but is based upon self-identification and active participation by individuals."[1] Nevertheless, it is well known that many participants in IETF are operating on behalf of their employer.

This potentially dual nature of IETF participation raises interesting normative questions. Are IETF participants, as individuals, expected to make arguments and deploy reasons in a kind of Habermasian discourse for the public's interest? (Froomkin, 2002) Or is the IETF rather the site of contesting, powerful commercial interests? (Kuerbis & Mueller, 2011) Does the latter tendency erode the IETF's legitimacy as a standards organization?

This proposal does not aim to answer these questions. Rather, it addresses a descriptive question of whether it is possible to detect, empirically, when an organization is acting within the IETF through individuals that it employs, as opposed to letting those employees act in a more personal capacity. I aim to address this question using BigBang (Benthall, 2015), an open source toolkit for studying the dynamics of collaboration that has been used to instrument and analyze Internet standards setting organizations. (ten Oever, N., Milan, S., & Beraldo, D., 2020)

To explore this hypothesis, I will draw on structural aspects of public IETF data. In particular, the IETF is organized into a number of different working groups, each of which has a set of (often overlapping) individual participants. These participants are identified individually, but may also have an affiliation that is either explicit (in a RFC authorship listing) or implicit (in an email address domain). Hence, IETF data can be viewed at multiple different *scales* of analysis: individual, email domain, and working group.

I draw on the general complex systems theory result of Siegenfeld and Bar-Yam (2020) that systems can vary by their *complexity profile*, the plot of their complexity at different scales, depending on whether they are "random", "coordinated", or "coherent" systems. The actions of individuals that are acting on behalf of an organization will be more coherent -- appearing with complexity that is similar across different scales of analysis -- than the actions of individuals who are interacting more randomly. In other words, those acting as individuals should be identifiable by virtue of how they do not contribute to complexity at higher scales of analysis.

Preliminary results have been promising. An analysis of email domains used in IETF working group mailing lists has shown that representativeemail domains that are typed as either *generic hosts* (gmail.com, gmx.de), *companies* (apple.com, cisco.com), or *personal* (mnot.net,

---

[1] https://www.ietf.org/about/participate/tao/

csperkins.org) have noticeably different signatures across a few summary statistics of their participation at different scales.

Statistics of Email Domain Distributions
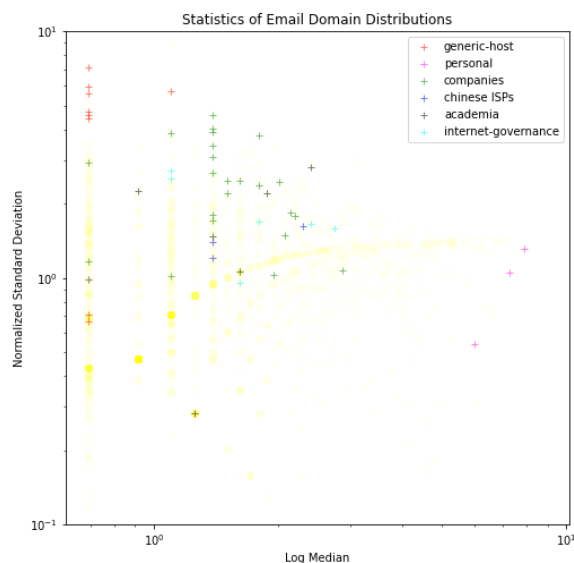
The plot shows these preliminary results. Each point is an email domain that is active in the IETF. Each domain has a corresponding *email distribution* -- the distribution of counts of the number of messages sent by each individual email address within that domain, to any working group. The X-axis shows the log of the median of this email distribution. The Y-axis shows the standard deviation of that distribution, normalized by the distribution's mean. The log median score is particularly useful in differentiating generic hosts, personal domains, and company domains.

The results of this work may be of interest to IETF leadership as a way of identifying which organizations are most active in influencing the IETF via their individual employees. This work is aimed for publication in a general science journal such as *Royal Society Interface*. This may also contribute to more social scientific and normative studies of Internet governance.

Ethical Considerations

This research will use exclusively information that is (a) public, in the sense that those who provided it had no reasonable expectation that it would not be recorded and made public, and (b) not obtained through interaction or intervention with those persons. It is therefore not human subject research and is not subject to the ethical considerations raised in the Bellmont Report.

While there can be privacy concerns when "public" data is recontextualized and analyzed (Nissenbaum, 2020), in this case the main focus of the research is on institutions and not natural persons. Those natural persons that are identified in any resulting materials will be discussed only insofar as they have acted in a public capacity as a participant in IETF.

References

Benthall, S. (2015). Testing generative models of online collaboration with BigBang. In *Proceedings of the 14th Python in Science Conference.*

Froomkin, A. M. (2002). Habermas@ discourse. net: Toward a critical theory of cyberspace. *Harv. L. Rev.*, *116*, 749.

Kuerbis, B., & Mueller, M. (2011). Negotiating a new governance hierarchy: An analysis of the conflicting incentives to secure internet routing. *Communications and Strategies*, (81), 125-142.

Nissenbaum, H. (2020). *Privacy in context*. Stanford University Press.

Siegenfeld, A. F., & Bar-Yam, Y. (2020). An introduction to complex systems science and its applications. *Complexity*, *2020*.

ten Oever, N., Milan, S., & Beraldo, D. (2020). 10 Studying Discourse in Internet Governance through Mailing-List Analysis. *Researching Internet Governance: Methods, Frameworks, Futures*, 231.