The Evolution of IETF Standards and Their Production

Mehmet Gencer
Department of Computer Science, Istanbul Bilgi University, Istanbul, Turkey

Abstract

This paper reports the results of an exploratory study on the Internet Engineering Task Force (IETF) specifications corpus in relation to the changes in volume, structure, and production of Internet standards. Using data spanning three decades we examine changes in the production volume and type composition of IETF documents, their interdependency, and the level of collaboration involved in their production. Longitudinal changes in the standards corpus exhibit an increasing trend in interdependency, number of refinement steps, and number of authors, and additionally reveal that standards production is of an episodic nature with regular peaks in output volume. Complementary analysis on the network structure of dependencies highlights a trend toward compartmentalization of the system over the years involving the emergence of relatively isolated subsystems of related standards. We suggest that a perspective which considers a system like IETF as an organization itself, rather than a constellation of extra-organizational activities, is needed to understand and manage standardization processes like this one.

Keywords: standards evolution, standards collaboration

Introduction

Information and Communication Technology (ICT) industry is strongly dependent on standards and this dependency is increasing. Despite the wider debate on whether standardization is an obstacle to, or catalyst of, innovation, most actors in the Software and Internet technology industry seem to embrace standardization, and, in particular, open standards (Capek et al.., 2005). Increasing numbers of industry consortia in the ICT industry are being formed in order to deliver the standards needed and these play a major role in the standardization process (Blind et al.., 2010). While dominant firms' attempts to impose their own technologies occasionally create conflicts in the standardization process, there is a general tendency in the ICT industry to embrace collaborative processes of standards creation in the public domain (Garcia, 1992; Bonaccorsi et al.., 2006; Simcoe, 2006).

Driven by either industry consortia or formal bodies, participation in the standardization process seems to be becoming more widespread in the ICT and other similar industries. This is particularly true for Internet technologies, where mobility of digital information over the globe across a variety of hardware and software platforms demands a high level of compatibility and durable standards. Certain levels of this technology stack favor formal or semi-formal standardization bodies like the Internet Engineering Task Force (IETF), Institute of Electrical and Electronics Engineers (IEEE), or ISO. Despite extensive collaboration within these bodies, or perhaps because of it, the creation of standards is often delayed or even fails completely (Besen & Farrell, 1991); both outcomes are costly for the industries which rely on these standards.

Recently research on standardization is concerned with a variety of issues including motivations for standardization, its impact on, and diffusion into industry, legal issues and business strategies related to standardization. On the other hand, empirical studies concerning how standards are produced are rather rare. An understanding of how production takes place in standardization bodies, how systems of standards are organized, and how all these change over time, can contribute to our understanding of failures and delays in standardization.

This paper is an attempt to contribute in this direction. We present results of an empirical study building upon our previous work (Gençer et al.., 2006) concerning Internet standards published by the IETF and the collaborative processes involved in their production. We follow these processes through several decades of IETF's existence. In addition to conventional exploratory statistics,

we borrow concepts, methods, and tools from social network analysis to examine production related features and referential relations of standards in relation with the ways in which a system of standards change over time. Our study (1) highlights long term trends in the volume, composition, and collaborative features of IETF output, (2) explores changes in the subject focus of the standardization community, (3) examines changes in the interdependency structures of IETF driven standards using social network analysis methods, and discusses the consequences of certain structural changes for the future of standardization work.

The next section summarizes the background literature and describes the research methodology adopted in the study. Then we present the IETF case and the data set used in this research, along with the analysis methods employed. In the following sections we first present findings regarding changes in the features of standards and the extent of collaboration in their production, then we summarize findings about changes in interdependency structures through the decades. Finally we discuss consequences of the findings for the standardization process, followed by a summary of our conclusions.

Background and Methodology

Existing research on standardization tends to focus on subjects such as the way in which standards diffuse into industries and affect innovation and their relation to the competitive strategy and performance of firms (Choi et al.., 2011). While there are several studies which concern systems of standards as a whole and the processes within them (e.g. Egyedi 2003), the majority of research take the firm/organization level as the unit of analysis. Similarly, a good portion of MIS research concerns standards to some degree, but its focus is on the business organization or intra-organization level and role attributed to standards is not explicit or clear (West, 2003). At the opposite end, research on technological systems (Allen & Sriram, 2000) focusses on the relation between standards and whole societies or industries. As a consequence our understanding of how the systems of related standards are produced at the intersection of organizations, and how such systems change over time, is both limited and fragmented.

Among the few exceptions is a study by Nickerson and Muehlen (2006) which uses institutional theory to study evolution of the bodies that form Internet standards. They note that “economic self-interest alone cannot explain all aspects of the Internet standard-making process... An approach that describes an ecology (a set of

relations between different standards institutions, ideas, and participants) provides needed explanations"(Nickerson & Muehlen, 2006, pp17). Using relational examination , Gamber et al. (2008) analyze the referential relations between clusters of standards in an industry, and identify spillover effects in standards clusters similar to the case of patents. Although yet limited in terms of scope and practical implications, these lines of research indicate the value of relational analysis for understanding internal dynamics of systems of standards, separate from the market dynamics surrounding these systems.

Concepts and methods developed in the field of social networks research provide a powerful basis for such relational analysis. Empirical methods developed in this field are based on the mathematical concept of graph which is used to represent relations between people in a variety of social phenomena from relations between school children, to corporate board interlocks (Scott, 2000). A rich variety of graph based metrics have been proposed to assess features of individuals, groups or whole networks in relation to knowledge processes, power, and trust in social systems (see Wasserman & Faust 1994 for a review). One finds numerous different social network metrics for assessing different aspects of same structural feature (Scott, 2000), such as the 'information centrality' which considers collaborative relations, versus 'structural hole' which considers competition and opportunity. Often these structural measures are subjected to different frames of interpretation depending on the social phenomena of interest.

With the recent growth in the popularity of the term "social networks" in reference to Internet based virtual communities, the field has experienced a boom in the availability of network visualization and analysis tools. In addition these new fields of application have stimulated research interest in study of social groups and cliques. Beyond the well known concept of 'small worlds' (Watts, 1999) which refers to the highly clustered nature of relations in social and natural phenomena, social networks researchers have started to further develop techniques and concepts to understand clustering and its dynamics in social systems.

Social Networks Analysis Methods Used in the Study

We suggest that the generality of structural assessment methods and tools, and the rich variety of frames of interpretation makes social networks analysis a useful resource for understanding the endogenous dynamics of systems of standards and the processes within these systems. Social network structural metrics, together

with a repertoire of their interpretations in various applications provide a suitable toolset for investigation of systems of artifacts such as groups of standards. Parts of the empirical study presented here borrow centrality measures and visualization tools from social network analysis to explore essentially non-social referential relations within a corpus of Internet standards. While social network analysis is rather weak in terms of longitudinal analysis methods, we use the relevant metrics and visualizations to compare non-overlapping time frames in order to understand how the corpus evolves over the several decades of its history.

Although this is mainly an empirical study of the referential relations between standards, we have applied certain network metrics to co-authorship relations as well, in order to assess the possible similarities between the relational structure of artifacts (standards) and of their producers (standards authors).

The most essential network metric used in this study is *in-degree centrality*, which is simply a count of references made to a standard from other standards documents. This is a direct measure of 'internal' importance of a standard within the system. We also refer to measures such as *betweenness centrality* (Wasserman & Faust, 1994) which assesses the role of an element in integrating a system.

Measures such as in-degree or betweenness centrality are used to assess the features of individual elements in a system of relations. However, one is often interested in the emergent sub-systems within the larger system. Such sub-systems are groups of elements which are relatively highly interconnected with each other. If the system at hand represents friendship relations, for example, a group of friends can be identified by looking at the fact that they come together more often with people within the group, in comparison with people outside the group. Since such sub-systems/communities are emergent features of a system, several methods have been proposed for their identification. Some of these methods allow one to assess 'how much' a system is compartmentalized into sub-systems, by considering the number of sub-systems and how isolated these systems are. For this purpose we have used commonly implemented metrics of *clustering coefficient* (Watts & Strogatz, 1998) and *modularity* (Blondel et al.., 2008) as a measure of how much a system is compartmentalized, specifically to measure how general structure becomes fragmented as the system grows over time.

The IETF Standards

Internet standards date back to the beginning of the conception and engineering of the Internet within ARPA (Advanced Research Projects Agency) in 1969. As internet technology emerged from its defense project cocoon, IETF was established in 1986 as a civil organization, which remains intimately connected to Internet and one of the important platforms that drives the processes for production of Internet standards (Alvestrand, 2004). Most, if not all, standards produced before 1986 were later cataloged by IETF.

IETF organizes three open meetings in a year. As Alvestrand notes in his insider review of IETF processes, decision making is not based on voting, but on "rough consensus and running code" (2004, p1372). The formal output of IETF processes is published as part of the "Request for Comments" (RFC) document series, which contains "each distinct version of an Internet standards-related specification" (Bradner, 1996, p5).

Although all RFCs are standards related specifications, only some are Internet standards. Among others are those called BCP (Best Current Practice) which "standardize the results of community deliberations about statements of principle or conclusions about what is the best way to perform some operations or IETF process function" (Bradner, 1996, p6). Only some RFCs contain specifications which are intended to become standards (so called standards track). Other than these, IETF publishes RFCs which are categorized by labels such as 'experimental' or 'informational', whose circulation is seen beneficial to the Internet standards community. Table 1 presents a list of standards and non-standards track RFC types, along with their common abbreviations. The order of these listings indicates an increasing 'level' of RFC types. For example a specification enters the standards track as a proposed standard, and over time, and depending on community feedback, may escalate to draft standard, then to standard level.

Table 1: Types of RFC documents published by IETF

STANDARDS TRACK (in order of maturity)
PS : Proposed Standard
DS : Draft Standard
S : Standard
NON-STANDARDS TRACK
H : Historic
I : Informational
E : Experimental

IETF has established rules and procedures for the format and content of different types of documents, how they circulate and progress in their relevant tracks, etc. Not all standards-track specifications progress towards becoming a standard and their status depends on community feedback from their application. Since its establishment, IETF has maintained good support from, and representation of, the industry. The authorship of RFC documents demonstrates this support, as most authors are company representatives, although many academics are also involved in the process.

The number of RFCs published by IETF each year indicates a growing trend as shown in Figure 1. Alvestrand's (2004) report covering numbers of participation to meetings up to 2003 indicates a boom which subsided in the year 2000. However, there seems to have been a new wave of activity in recent years. Figure 1 exhibits an ongoing increasing trend in IETF's RFC publication counts, overlaid with peaks, as shown.

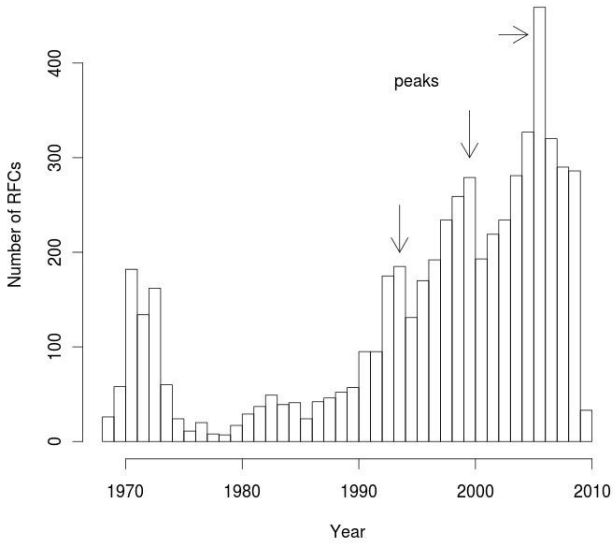


Figure 1: Number of RFCs published by IETF each year, with peaks of activity marked.

Data-set and Analysis Methods

IETF makes available both meta-data and complete texts for each of its RFC documents, some of which date prior to IETF's establishment in 1986. We have used custom software to retrieve both meta-data and texts and to produce the necessary data sets. Among the meta-data about each RFC are the year of publication, list of authors, and RFC type. In cases where specification in an RFC obsoletes an existing one, this information is also included in its meta-data, thus providing information about both the life-times of and relations between RFCs. Although some of these meta-data were not collected in earlier RFCs and were retrospectively attached by IETF, the meta-data is generally reliable. We have further analyzed RFC texts with custom software to extract referential relations between documents. The data used for this study included 5781 RFC documents.

These pieces of information are used to generate the following set of data for each RFC in the IETF corpus:

1. *Year* of publication.
2. List of document authors, whose length we denote as *number of authors*.
3. Current *type* of document, where possible types are listed in Table 1.
4. Whether the RFC is replaced by publication of another one, and if so the identification of replacement. For such RFCs we have also added the *lifetime* information of RFC by checking the difference between the year of publication of the RFC and the one which replaces it. The lifetime data is produced only for 882 such documents.
5. The referential relations between RFCs are generally not available other than the references implied by one making another obsolete (although the meta-data scheme appears to reflect intentions to include such referential information). For this reason we have resorted to scanning document texts using a custom computer program. As a result of this process, we have found which other RFCs are mentioned in an RFC body, and how often (i.e. the strength of reference). The *number of dependencies* in each RFC is appended to a per RFC data set which denotes the number of those RFCs which an RFC has references to. The reference strengths are used for visualization and analysis of the interdependency network.

The referential relations between RFCs form the basis of a directed network of dependency relations. Figure 2 demonstrates how such a network is formed. If an RFC document mentions another one, a directed edge from the former to the latter is added to the network, and edge weight is set to the number of times the latter RFC is mentioned in the text.

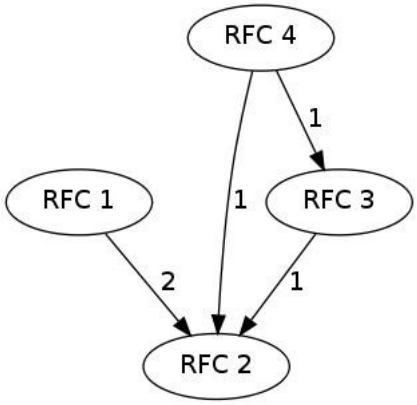


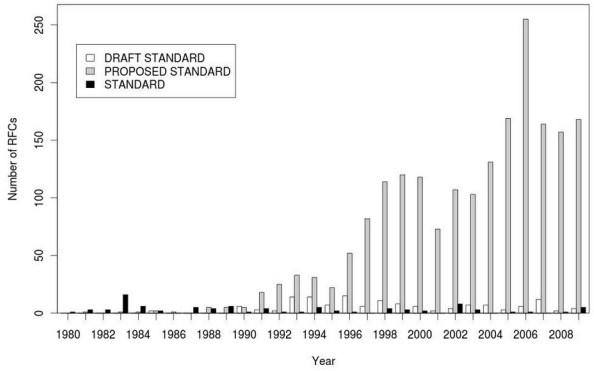
Figure 2: An example of referencing relations between RFC documents. In this example RFC-2 is mentioned twice in the text of RFC-1, thus the relation strength is two. In the other relations the relation strength is one.

The data set is analyzed using descriptive tools from the R statistics package (R Development Core Team, 2009). In addition to this data set, the strengths of references are used to lay out the interdependency network of the RFC corpus, which is, in turn, analyzed and visualized using the ORA software toolkit for social network analysis (Carley et al.., 2009) and Gephi (Bastian et al.., 2009). The interdependency network is represented as a directed and weighted graph (Scott, 2000) in which the vertices are RFC documents, edge direction corresponds to referencing direction, and the weight of an edge is found by counting the number of occurrences of the referenced RFC in the referencing RFC document's text.

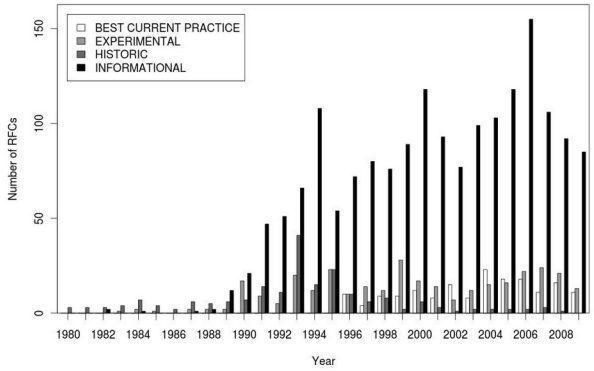
Changes in IETF standardization output and process

The volume and composition of the IETF documents published each year reveal certain aspects of the evolution of IETF standards and their production. The total number of IETF RFCs published each year is shown in Figure 1. This output volume is composed of different types of documents. The composition of publication volume across different types of documents is shown in figures 3a

and 3b, separately for standards track, and non-standards track RFCs . When interpreting this data one must remember that the IETF was established in 1986, although some RFCs were released prior to that and included in the IETF data set. Therefore fluctuations prior to this year, during what we may call the infancy period of Internet related institutions, are disregarded here. The composition has shown that a large portion of documents in this period are labeled as 'unknown', which justifies the decision to disregard this early period in the context of this study.



(a) Standards track RFCs over the years.



(b) Non-standards track RFCs over the years.

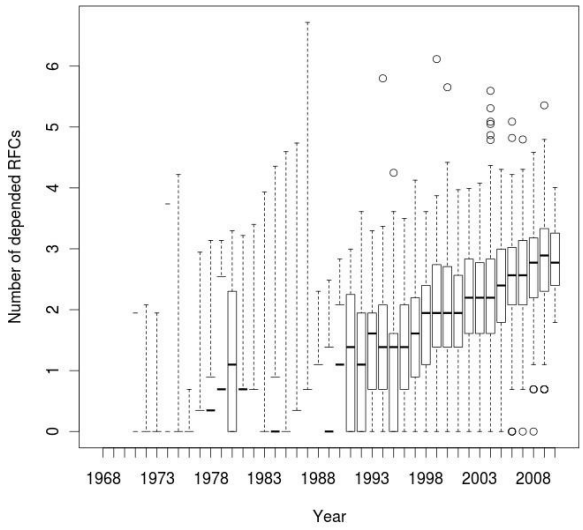
Figure 3: Changes in the volume and composition of standards and non-standards track RFCs over the years.

These data for RFC publication volume exhibit two distinct features: (i) there is a consistent increase trend in volume continuing to the present day, and (ii) there are peaks of output volume overlaid with this increase trend, approximately every six

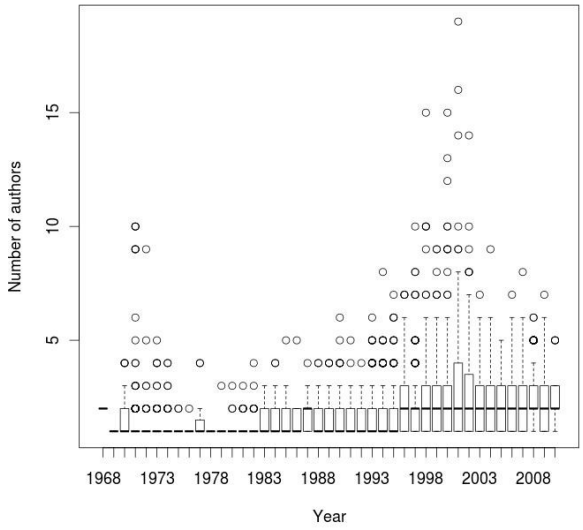
years. These peaks are marked in Figure 1, and seems to be common in standards and non-standards track volumes in Figures 3a and 3b.

The first of these features is rather expected, since IETF has retained its role uninterruptedly since its formation, and the spread of Internet use increased demand for the production of an increasing volume of standards. The plot of document production volume composition with respect to the type attributes of documents, shown in figures 3a and 3b, also reveals some changes over time. The majority of recent output is composed of PS (proposed standard) and I (informational) type documents, which are at the less mature levels of their respective tracks. Although BCP and E type documents display a steady volume for the same period, volume is low. On the other hand, more mature specifications, i.e. those of DS (draft standard), S (standard), or BCP (Best Current Practice) type, have a low and more or less stable volume. This data therefore suggests that over time it takes increasingly more refinement stages and more extended collaboration to produce mature standards and specifications. Recent updates by IETF to its processes seems to acknowledge a slowdown and reflect the need to make these processes both more inclusive (of independent contributors) and faster (by adopting parallel approval processes)[i]. This seems to be similar to the task of modifying complex artifacts such as software (Ghezzi et al.., 2002) which becomes more complex and time consuming as the product grows. More interestingly, it also shows that IETF specifications production increasingly relies on two document types (I and PS types). This, in turn, prompts a further question concerning whether use of too many types makes coordination for standards production more difficult, and whether this is the reason for the community converging on a small subset of document types.

Changes in RFC document volume composition hint at the changes in standards production progresses. Dependency relations between standards and the extent of their authorship can provide further explanations about such changes. Figures 4a and 4b show the changes in two related features through years in box-plot form: number of dependencies (references) and number of authors. Both measures exhibit a clearly increasing trend after 1986. This data shows that it takes more people and wider consideration of, and integration with, existing standards to produce a new one. This resonates with the above finding that it takes more stages and requires more corrections for proposed standards to mature into standard status.



(a) Number of dependencies of RFCs published each year.



(b) Number of authors of RFCs each year.

Figure 4: Distributions of number of dependencies and number of authors of RFCs published each year. The solid boxes mark the first and third quartiles of the distributions with the mean value shown as a bold line, whereas the bubbles beyond the dotted whiskers show individual extreme values in data.

The second feature of RFC production volume, that it has an episodic nature, worths more attention. Systems science can

provide a probable interpretation for these episodes. As Simon (Simon, 1962) describes, while a system grows, relatively distinct subsystems emerge within the larger system as the need arises. Such an evolutionary pattern over time may well be expected from the system consisting of the IETF specifications corpus. If that is the case, the IETF community's attention should be shifting between these subsystems depending on the immediate and common expectations within the community. It seems reasonable to expect that a contemporary challenge will invite efforts to create a set of interrelated specifications, with these efforts becoming a major attractor of the attention of IETF community. Indeed, the special attention of the community to security and virtual private network related technologies, which was noted by Alvestrand (2004) in reference to the state of IETF affairs in year 2003, shows itself in the emergence of a relatively independent subsystem of specifications which we shall discuss in the following sections. However, creation and stabilization of such a subsystem of standards or non-standards track specifications will probably start with a low volume of output, and only after reaching a consensus on key issues will it then exhibit a higher output volume. Definitive ascertainment of the causes of this episodic volume requires further qualitative and quantitative work on both the RFC corpus and the IETF community.

Exploration of document lifetimes in this study was limited to a small portion of the RFC document corpus and we did not observe any apparent temporal pattern; hence no results are reported here.

Changes in the structure of the RFC corpus

It is only in rare cases that one has extensive records of a community's work and the interactions involved in conducting a collective work. On the other hand the outcomes of such work, often reveal essential features of their production. This is the case for IETF community, since it is possible to determine interdependencies between RFC documents produced by the community, which to some extent reflects the collaborative structures underlying their production.

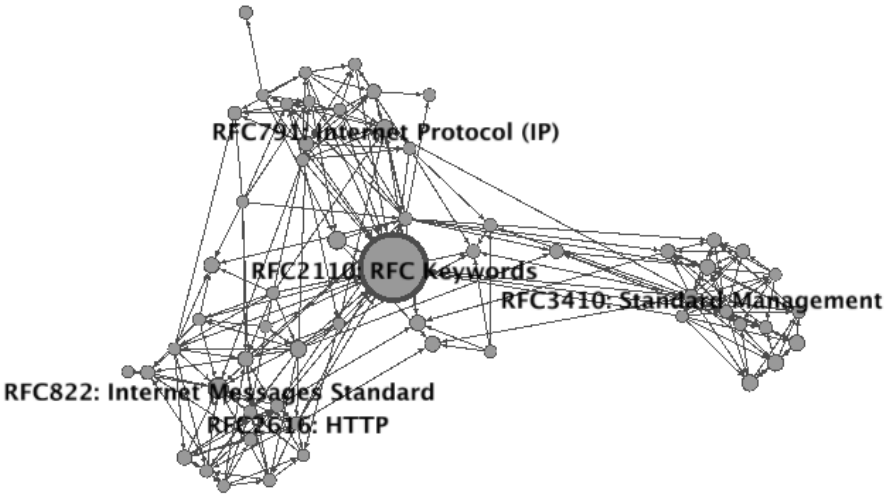


Figure 5: The dependency network structure of the whole IETF RFC corpus, with only the most central nodes (90 out of about 6.000) included in the visualization. The sizes of vertices representing RFCs are proportional to their in-degrees, while the lengths of the arrowed edges representing referential relations are proportional to reference strength/count.

The features of the RFC interdependency network structure explored in this section are presented below, first for the whole RFC corpus, then for the three decades between 1980 and 2010. A visualization of the network structure for the whole RFC corpus is shown in Figure 5 which was created with the network analysis tool Gephi (Bastian et al.., 2009). In this visualization, each vertex corresponds to an RFC and each edge corresponds to a referencing relation. The graph contains only a subset of RFCs, however, since full visualizations of such large networks usually convey little information (Bender-Demoll & Mcfarland, 2006). Thus the visualization in this figure shows, not all, but only most important nodes, selected on the basis of the *in-degree* measure that is often used in social network analysis as an essential measure of positional importance (Wasserman & Faust, 1994; Scott, 2000; Nooteboom, 2008). Graph vertices (RFCs) are shown in sizes proportional to their in-degrees.

The whole RFC corpus:

Figure 5 demonstrates the historical build up of referential relations between IETF RFCs. A variety of RFCs from different years depend on one another, thus the corpus is not a collection of isolated islands of documents but an interconnected group whose connections do not follow any initial plan. Nevertheless, as is

common in most real networks the RFC corpus network exhibits a "small worlds" structure (Watts, 1999). One can observe several subgroups which are tightly connected to one another, more so than they are connected to documents in other subgroups. The visualization only shows labels of the most central RFCs in their respective subgroups, for readability reasons.

The distinctive subgroups in Figure 5 are: (i) the subgroup of user level application protocols such as HTTP (the standard protocol for the transfer of web pages) and MIME (standard for file attachments in e-mail and web pages), at the bottom left of the figure, (ii) the subgroup of network management related protocols, at the right of the figure, and (iii) the subgroup of basic data transfer protocols such as the Internet Protocol (IP). Since the latter subgroup is essential to the operation of the Internet and depended on by others, it has a relatively higher proportion of links to other subgroups compared to within-group links. At the center of Fig. 5 is an RFC which defines numerous keywords that are frequently referenced from other RFCs. It is an example of similar RFCs which bind the system of RFCs, and which typically define standards for the specification of standards.

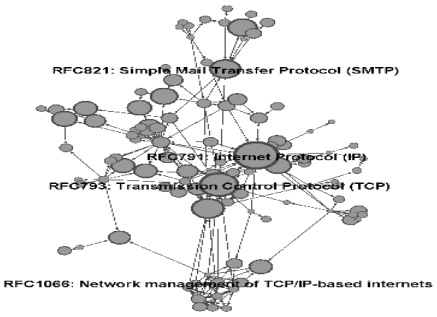
The measure of betweenness centrality (which assesses the integrative role of an element in a graph) points to different elements in the system from those indicated by the direct measure of in-degree centrality. The highest betweenness centrality in this graph belongs to an RFC (RFC2279) which specifies a standard for universal character encoding to support multiple content languages in a variety of situations. The second rank in terms of betweenness centrality is an RFC (RFC2049) which specifies improvements to support a variety of content types on the Internet. This finding highlights the pressures on Internet standardization efforts stemming from both globalization and increasing content variety. It is worth noting here the utility of ostensibly irrelevant social network analysis methods and metrics in exposing systemic features of the structure of a network which is itself not social.

We next look at the structures for the three decades of the 1980s, 1990s, and 2000s separately.

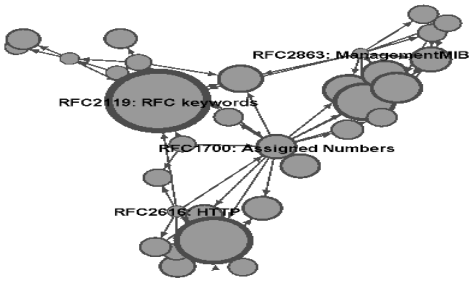
1980s

The network for RFCs published between 1980-1990 is shown in Figure 6a. The structure seems to reveal the beginnings of the emergence of specialized sub-groups within the system. At the center of the graph are standards labeled TCP and IP, which are abbreviations of protocol names that are the most essential

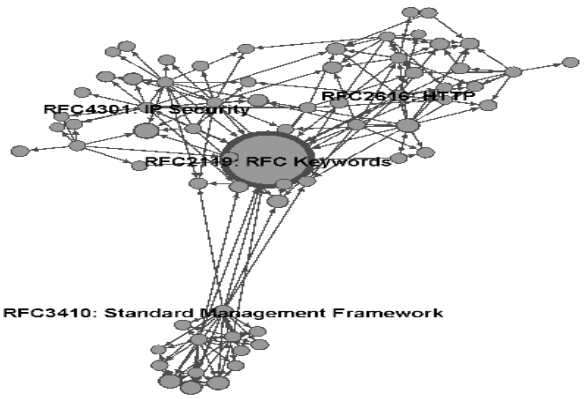
standards for transfer of digital data packets over the Internet between computers. Two peripheral subgroups seem to be emerging in the 80s: (i) the application level standards which specify the content structure of data carried in the Internet for specialized uses such as e-mail (top of the figure), and (ii) network management related standards (bottom of the figure). This structure indicates that the focus of work in the 1980s decade is making the Internet communication infrastructure work, with the newly emerging interest in network scaling and use of Internet communication for services like e-mail.



(a) RFCs produced in 1980s.



(b) RFCs produced in 1990s.



(c) RFCs produced in 2000s.

Figure 6: Dependency network structure of RFCs produced in different decades, with limited numbers of most central nodes included in each. The sizes of vertices representing RFCs are proportional to their in-degrees, while the lengths of arrowed edges representing referential relations are proportional to reference strength/count.

1990s

The network for the period 1990-2000 is shown in Figure 6b. Here the network management group continues to be distinguishable (top right of the figure). The application level specifications subgroup also continues to exist in this decade (bottom of the figure), but the activity is now concentrated on the HTTP protocol which is at the center of web technologies that started to become popular after 1991. At the center of the network are groups of specifications which attempt to clarify both the standards for RFC authoring (e.g. RFC keywords) and matters such as assigned numbers which become an important issue for the growing Internet. These are numbers which globally standardise the computer port addresses web page services must use, etc. The focus of work in the 90s decade seems to be two fold: (i) to sustain compatibility between the increasing variety of technologies, and (ii) to impose structure and discipline on the community's work under an increasing production load.

2000s

The network for the 2000s, shown in Figure 6c, exhibits a different structure with some subsystems of standards becoming relatively more isolated from the network core. Specifically, the network management subgroup (bottom of figure 6c) is relatively more isolated than it was previously. A newly emerging subgroup at the top left contains RFCs related to Internet protocol security. The emergence of this subgroup reflects the increasing concern with Internet security and its use in electronic commerce in the new millennium with the spreading of web usage. The third subgroup, at the top right, is once again related to application level specifications, and HTTP remains the most important standard in this subgroup. The keyword definitions RFC remain the most central RFC.

Changes in the structure of standards and collaboration

Changes in the network structure, seem to indicate an ongoing compartmentalization of the RFC corpus. This prompts us to question whether the same may be true for the structures of collaborative work behind standards production. Findings in social studies (Kogut, 2000; Langlois & Robertson, 1992) suggest that such an evolutionary path towards progressive specialization may well be the case for systems of standards and the work behind their production.

While the tools of social network studies do not provide well developed longitudinal methods to assess such changes in the structure, a comparison over the decades is possible. For such a comparison we have used some measures to assess the level of compartmentalization separately for the network of each decade. As described in the social network analysis methods section above, there are metrics proposed in the literature which assess compartmentalization of a network by looking at how it is partitioned into sub-systems and how isolated these sub-systems are from one another. For this purpose we have applied two commonly used metrics of *clustering coefficient* and *modularity* available in the two corresponding software tools of Gephi (Blondel et al.., 2008; Bastian et al.., 2009) and ORA (Watts & Strogatz, 1998; Carley et al.., 2009), respectively. These metrics were applied to the RFC networks of the three decades separately, to make sure that relative changes in the two metrics are in check with one another. One of the metrics is also applied to the co-authorship networks of the corresponding decades to check whether the compartmentalization of the system of standards is, as we expected, in parallel with that of the collaboration behind its production. The results are shown in Table 2.

Table 2: Changes in compartmentalization of RFC corpus and co-authorship network for three decades, using two measures: (#1) *modularity* (Blondel et al.., 2008), and (#2) *clustering coefficient* (Watts & Strogatz, 1998).

Network/Measure	1980s	1990s	2000s
RFC, #1	0.453	0.581	0.563
RFC, #2	0.133	0.164	0.171
Co-authorship, #1	0.785	0.851	0.803

These results show an increase in compartmentalization from the 80s to the 90s, followed by stability from the 90s to the 2000s. This is in general agreement with the system science insight that a growing system gets more compartmentalized in time, and the stability achieved during the last decade is in parallel to that of activity. In addition, the results provide preliminary support for our expectation that compartmentalization of the organization of production (co-authorship network) parallels that of the products, while determining the particular characteristics of the parallelism between the two requires further study.

Discussion of Findings

The findings reported here reveal several features of the IETF's RFC specifications corpus and the way it has changed over time. An obvious feature of the IETF corpus is that it is a growing system. The findings here demonstrate some important properties of this system. First of all, the production output level suggests that this system acts or reacts as a whole. The output volume of RFCs published by the IETF exhibit fluctuations, or episodes, marked with seemingly regular peaks. The changes in the system's output suggest that IETF is an organization in the traditional sense and that it reacts to environmental demands as a whole. However, empirical investigations focusing on the motivations and behavior of standardization process participants (e.g. commercial firms) are mostly blind to such system level features. In other words, our results underline the need to consider a system of standardization such as IETF as an organization with its own identity, organizational culture, dynamics, goals, etc.

Few reports on the IETF in the literature seem to support this interpretation. Bradner (Bradner, 1999) describes how IETF work is structured to improve efficiency and work focus, and how certain practices are employed to increase output quality. Alvestrand's account of the IETF (Alvestrand, 2004) speaks of a strong work culture, how participants associate certain elements of organizational identity with the IETF, and in addition indicates that this organization as a whole directs its attention to certain contemporary challenges. A contrary view that a system such as the IETF is simply a constellation of the extra-organizational activities of contributors from different organizations would neither concur with these accounts, nor explain the non-random, episodic output volume of the IETF standardization process reported here. Even when one considers the possibly conflicting agendas of IETF

contributors, such aspects are well suited to be investigated within organizational research frameworks.

Presence of episodes in organizational output is not a widely known phenomenon and our research methods are not suitable to offer an explanation. However, looking at Alvestrand's account in conjunction with our findings, we suggest that since organizational attention is possibly focused on a certain few contemporary challenges, the episodes reflect how these challenges are tackled, then a basic consensus on their solution is reached, and eventually reflected in IETF specifications. If one subscribes to the view that IETF is an organization with shared goals and relatively unified focus, it makes sense to assume that RFC production within the IETF proceeds through such phases of organizational problem solving. A full explanation of this episodic output, however, will require a qualitative and retrospective study.

The second property of the IETF system, which is somewhat contrary to the first, is that, as the system grows, it becomes divided into specialized subsystems. This is evident in the structural changes in the RFC interdependency networks. While such compartmentalization is at first contrary to the view of the IETF as a unified system, it is a common evolutionary path in organizational growth (Aldrich, 1999). Our findings also reveal that the level of compartmentalization has stopped increasing during the last decade, and in parallel the production volume of RFCs has started to decline. Since the demand for Internet related standards appear to be growing, this indicates that IETF's position in Internet standardization is shrinking, possibly in parallel with other standards organizations taking over certain sub-fields. In addition, we have found that it takes more authors, deeper consideration of existing specifications, and more revisions to develop specifications into mature standards. In considering these findings together, it should be asked whether it is inevitable that the work of standards production in a growing system of standards become more complex or inefficient. More precisely, it is worth considering whether such inefficiency could be addressed by introducing changes in administrative or organizational practices. A failure to address growth, and the structural changes it brings about, may eventually lead to further inefficiency and shrinking of IETF's role in Internet standardization.

Conclusions

This study explored features of production and relational structure of Internet standards related specifications produced by the IETF, as

it has grown through several decades. By combining social network analysis approaches, we have investigated changes in standards production in relation to the network structure. The findings reported here indicate the increasing complexity of the work on standards specifications, and development of specialization as the system of Internet standards grows. Certain longitudinal features of IETF production call for a view of this 'task force' as an organization with relatively unified goals and focus, rather than merely a collection of contributions from different organizations.

References

Aldrich, H. (1999). *Organizations evolving*. Sage.

Allen, R. H., & Sriram, R. D. (2000). The role of standards in innovation. *Technological forecasting and social change*, *64*, p171-181.

Alvestrand, H. (2004). The role of the standards process in shaping the internet. *Proceedings of the IEEE*, *92* (9), p1371 - 1374.

Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: An open source software for exploring and manipulating networks.* International AAAI Conference on Weblogs and Social Media.

Bender-Demoll, S., & Mcfarland, D. A. (2006). The art and science of dynamic network visualization. *Journal of Social Structure*, *7*(2), p1-38.

Besen, S. M., & Farrell, J. (1991). The role of the ITU in standardization : Pre-eminence, impotence or rubber stamp? *Telecommunications Policy*, *15*(4), p311 - 321.

Blind, K., Gauch, S., & Hawkins, R. (2010). How stakeholders view the impacts of international ICT standards. *Telecommunications Policy*, *34*(3), p162 - 174.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008, 10October). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10).

Bonaccorsi, A., Giannangeli, S., & Rossi, C. (2006). Entry strategies under competing standards: Hybris business models in the open source software industry. *Management Science*, *52*(7), p1085–1098.

Bradner, S. (1996). *The Internet Standards Process – Revision 3*. RFC 2026, IETF.

Bradner, S. (1999). The Internet Engineering task Force. In *Open sources: Voices from the open source revolution*, O'Reilly.

Capek, P. G., Frank, S. P., Gerdt, S., & Shields, D. (2005). A history of IBM's open-source involvement and strategy. *IBM Systems Journal*, *44*(2), p249–257.

Carley, K., Reminga, J., Storrick, J., & DeReno, M. (2009). *ORA user's guide* (Tech. Rep.). Carnegie Mellon University, School of Computer Science, Institute for Software Research.

Choi, D. G., Lee, H., & Sung, T. Kyung. (2011). Research profiling for standardization and innovation. *Scientometrics*, *88*(1), p259–278.

Egyedi, T. M. (2003). Consortium problem redefined: Negotiating democracy in the actor network on standardization. *International Journal of IT Standards and Standardization Research (IJITSR)*, *1*(2), p22–38.

Gamber, T., Friedrich-Nishio, M., & Grupp, H. (2008). Science and technology in standardization: A statistical analysis of merging knowledge structures. *Scientometrics*, *74*(1), p89–108.

Garcia, D. L. (1992). Standard setting in the united states: Public and private sector roles. *Journal of the American Sociaety for Information Science*, *43*(8), p531–537.

Gençer, M., Oba, B., Özel, B., & Tunalıoğlu, V. S. (2006). Organization of Internet Standards. In *Open source systems*, E. Damiani, B. Fitzgerald, W. Scacchi, & M. Scotto (Eds.). Springer.

Ghezzi, C., Jazayeri, M., & Mandrioli, D. (2002). *Fundamentals of software engineering*. Upper Saddle River, NJ, USA: Prentice Hall PTR.

Kogut, B. (2000). The network as knowledge: Generative rules and the emergence of structure. *Strategic Management Journal*, *21*, p405–425.

Langlois, R. N., & Robertson, P. L. (1992). Networks and innovation in a modular system: Lessons from the microcomputer and stereo component industries. *Research Policy*, *21*, p297–313.

Nickerson, J. V., & Muehlen, M. Z. (2006). Standard making, legitimacy, organizational ecology, institutionalism, Internet standards, web services choreography. *MIS Quarterly*, *30*(SI), p467–488.

Nooteboom, B. (2008). Learning and innovation in inter-organizational relationships. In *The oxford handbook of inter-organizational relations*, S. Cropper, M. Ebers, C. Huxham, & P. S. Ring (Eds.). Oxford University Press.

R Development Core Team. (2009). *R: A language and environment for statistical computing.* Vienna, Austria. (ISBN 3-900051-07-0)

Scott, J. (2000). *Social network analysis*. Sage.

Simcoe, T. S. (2006). Open standards and intellectual property rights. In *Open innovation: Reaching a new paradigm*, H. Chesbrough, W. Vanhverbeke, & J. West (Eds.). Oxford University Press.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*(6), p467–482.

Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge.

Watts, D. J. (1999). Networks, dynamics, and the Small-World phenomenon. *The American journal of sociology.*, *105*(2), p493-527.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, p440–442.

West, J. (2003). The role of standards in the creation and use of information systems. In *Proceedings of the workshop on standard making: A critical research frontier for information systems*, MISQ Special Issue Workshop, J. L. King & K. Lyytinen (Eds.), p314–326.

.Notes

1

An IETF memo summarizes policy changes to make standardization process more inclusive of independent contributors (`http://tools.ietf.org/id/draft-klensin-rfc-independent-02.txt`, retrieved Nov. 10, 2011). However, the statistics of RFC editor still indicate a relatively longer processing time for independent contributions (`http://www.rfc-editor.org/overview.html`, retrieved Nov. 10, 2011).