Organization of Internet Standards: A Case Study of Interdependence and its Historical Development

Mehmet Gençer*, Bülent Özel*, V. Sinan Tunalıoğlu*, Beyza Oba**.

* İstanbul Bilgi University, Department of Computer Science

** İstanbul Bilgi University, Institute of Social Sciences

THIS IS A DRAFT COPY. PLEASE DO NOT CITE.

Abstract

In this study we look at a body of standards documents in RFCs(Request For Comments) of IETF(Internet Engineering Task Force). The cross references between these documents form a network. Approaches from social network analysis are deployed to assess centrality of artifacts in this network and identify cohesive subgroups and levels of cohesion. Availability of historical data makes it possible to identify structural stabilization patterns among the RFCs. Our results demonstrate three major groups centered around some key standards. Application of network metrics reflect different levels of cohesion for these groups. Some metrics adopted from social network analysis indicate key roles of some standards in connecting these different groups, which may be overlooked otherwise. Techniques for assessment of structural interdependency and insights about evolution of the structure can be equally applicable to similar domains such as dependency relations of software conglomerates like Debian GNU/Linux.

Keywords

Internet Standards, Interdependence, Structural Features, Patterns of success and failure, Diffusion and adoption of innovations, Dynamics of project teams.

1 Introduction

Open Source Software has a good reputation for its compliance with standards. Capability of open source processes for handling such externalities is a major reason of interest on and adoption of this methodology[West 2002]. Most such externalities are formalized elsewhere by authoritative bodies of standardization, with close cooperation with the software development community.

In this study we analyze several aspects of the body of standards documents in RFCs(Request For Comments) of IETF(Internet Engineering Task Force). IETF is an organization with major influence in development of Internet standards. Formation of IETF standards resembles very much the processes in open source development: influential members first issue RFCs reporting current practices and propose solutions to interoperability problems of Internet technologies, later these proposals are converged into standards.

The process is similar to the development and release cycles in software development.

There are some major motivations which makes the organization of IETF standards interesting for us: (1)full history of its development is recorded in RFCs themselves and available for longitudinal analysis, and (2)techniques for assessment of structural interdependency and insights about its evolution which may be gained from such analysis, can be equally applicable to other domains such as structure of software conglomerates, like Debian GNU/Linux packages. Also as a practical result, such analysis provide hints on importance level of some contemporary standardization efforts.

Our aim in this research is (1)to assess relative structural centrality of Internet standards, in order to understand inter-dependencies among them using techniques from social network analysis practice, (2)identify groups of standards that are related to each other more so than they are to the rest, and levels of cohesion in these groups, and (3)find stabilization patterns of structural centrality through longitudinal analysis. Development of such approaches can be valuable, for example, in identifying critical segments of similar conglomerates(e.g. software conglomerates like Debian GNU/Linux), in management of processes within them(e.g. release scheduling and team splitting), in partitioning of training programs, etc.

An overview of data and the network analysis approach is summarized in next section. Next three sections present results of analysis regarding structural centrality, subgroups and historical development of centrality, respectively. Followed by discussion and conclusions of the analysis.

2 Standards data and network analysis methods used

Software development processes are studied for the mechanisms of their evolution as a coherent system [Lehman, Ramil 2001], and as a community practice of actors [Tuomi, Lehmann 2004]. Other research on software call our attention to importance of discursive practices and alignment of software development efforts[Raymond 2004, Magnus, Szczepanska 2002]. Clusters and their formation in similar collaboration systems have been a subject of interest. There exist in social sciences research, valuable frameworks and methodologies for assessment of structural features of networks and their evolution [Borgatti 2005, Stephenson, Zelen 1989, White, Smyth 2003, Ring, van de Ven 1994, Oliver, Ebers 1998, Oliver 2001]
[Powell, Koput, Smith-Doerr 1996]. There is also a group of methods in computer and informatics developed for analyzing different structures(such as for web page rankings) within surprisingly similar terms[Kleinberg 1999]. However, not only that, to our interest, their application to domains of software processes and standards formation is limited, but also there is much way to go for developing frameworks for sensibly combining these different lenses for a better identification and understanding of structural features common in different contexts [Cook, Whitmeyer 1992, Wasserman, Faust 1994, Bonacich, Lloyd 2001].

There are over four thousand RFC documents published by the IETF. Most standards start as *informational* class documents. *Best Practices* documents are more influential than *informational* ones. But *standard* class RFC are by far the most important within this collection. In this study we have used only the 1.460 standard class RFCs for analysis. The relations between the nodes in this graph formed by the RFCs is a directed relation. Although there may be several references from one document to another, a dichotomous relation is assumed in the analysis, as the number of references varies greatly.

Our method for analyzing this data consists of several steps:

1. A quantitative selection of structurally important standards. Many standards are built upon these key elements, attributing more relative importance to them in the making.

2. Identification of subgroups, key technological questions addressed by them, and their cohesion levels. Subgroups and analysis of their cohesion is important in understanding specialization in growing networks.

3. Sampling of historical patterns of centrality metrics for some key standards and demonstration of stabilization patterns. Based on this analysis, a commentary of structural development of Internet standards is also presented.

3 Structural importance

Degree prestige(number of references) and relative in-degree prestige[Wasserman, Faust 1994] are used for assessment of structural importance of a node in the network. Table 1 shows top 15 RFCs according to this centrality measure. Density of the network(ratio of existing relations to possible number of relations between nodes) is found to be 0.003716.

Group of nodes 1-7 and 10 in the network are related to network management protocols. 11 and 15 are related to Internet protocol and its security. The rest is related to web and mail, except number 8 which seems to be distinct from the rest.

One possible extension of this centrality measure is number of descendants of each node, i.e. accounting for indirect references to an RFC. Top results are shown in table 2. Top five items according to degree prestige also appear in this list. However inconsistency of the two measures indicate that it is necessary to use some sort of attenuation factor when accounting for indirect references. Which is not included in this study.

4 Subgroups and specialization

Many standards are related to some others in terms of the technical issues they address. Figure 1 shows the relations between several standards that are found to be important according to degree prestige. Two major groups are identifiable in the network. One group which seems quite isolated is related to network management protocols. Another group includes Internet protocol and its security extensions. There is also a third group in fig. 1(a), however as more nodes are included in the graph, as in fig 1(b), we see that this rather eclectic group does not exhibit much coherence. In fact a look at the content of the RFCs in the group indicates that they are not conceptually centered around a single central issue. A reference list of standards corresponding to labels in this figure is given in Appendix A.

Rank	Degree	Relative deg.	RFC-No : Year, Title
1	141	0.0966	1213 : 1991, Management Information Base for Network Management of TCP/IP-based internets:MIB-II
2	129	0.0884	1212 : 1991, Concise MIB definitions
3	127	0.0870	2578 : 1999, Structure of Management Information Version 2 (SMIv2)
4	126	0.0863	1155 : 1990, Structure and identification of management information for TCP/IP-based internets
5	125	0.0856	2579 : 1999, Textual Conventions for SMIv2
6	118	0.0808	2580 : 1999, Conformance Statements for SMIv2
7	111	0.0760	1905 : 1996, Protocol Operations for Version 2 of the Simple Network Management Protocol (SNMPv2)
8	108	0.0740	2234 : 1997, Augmented BNF for Syntax Specifications: ABNF
9	89	0.0610	2045 : 1996, Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies
10	89	0.0610	1906 : 1996, Transport Mappings for Version 2 of the Simple Network Management Protocol (SNMPv2)
11	79	0.0541	2401 : 1998, Security Architecture for the Internet Protocol
12	76	0.0521	1035 : 1987, Domain names - implementation and specification
13	72	0.0493	1034 : 1987, Domain names - concepts and facilities
14	69	0.0473	2396 : 1998, Uniform Resource Identifiers (URI): Generic Syntax
15	64	0.0438	2460 : 1998, Internet Protocol, Version 6 (IPv6) Specification

Table 1: In-degree and relative in-degree prestiges of most central RFCs.

Rank(in-degree)	Descendants	RFC-No : Year, Title
1(25)	554	2046 : 1996, Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types
2(41)	501	1191 : 1990, Path MTU discovery
3(387)	278	2262 : 1998, Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)
4(328)	230	1158 : 1990, Management Information Base for network management of TCP/IP-based internets: MIB-II
5(18)	200	1123 : 1989, Requirements for Internet Hosts - Application and Support
6(229)	145	1284 : 1991, Definitions of Managed Objects for the Ethernet-like Interface Types
7(1)	141	1213 : 1991, Management Information Base for Network Management of TCP/IP-based internets:MIB-II
8(60)	136	1521 : 1993, MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies
9(4)	131	1155 : 1990, Structure and identification of management information for TCP/IP-based internets
10(388)	130	1065 : 1988, Structure and identification of management information for TCP/IP-based internets
11(188)	129	1212 : 1991, Concise MIB definitions
12(179)	129	2440 : 1998, OpenPGP Message Format
13(2)	127	2578 : 1999, Structure of Management Information Version 2 (SMIv2)
14(3)	125	2579 : 1999, Textual Conventions for SMIv2
15(5)	118	2580 : 1999, Conformance Statements for SMIv2

Table 2: Number of descendants(indirect references to) of RFCs. For comparison, their ranks according to in-degree prestige are also shown.

Some further analysis is helpful in understanding the cohesion of these groups. Relative cohesion of a group is defined as the ratio of the number of ties between group members to the number of ties to outside nodes[Wasserman, Faust 1994]. That ratio can be regarded as relative strength of "centripetal" and "centrifugal" properties of the group. This measure for the first group in fig. 1 is found to be 2.25, whereas it is 0.47 and 0.65 for the second and third group, respectively. A value larger than one should be regarded as an indicator of stronger in-group ties(centripetal). Thus it is only the first group which exhibits such level of cohesion.

Note also that some nodes such as number 31(UTF-8) and 8(BNF for Syntax Specifications) seem structurally important in terms of how they connect different groups to each other, independent of their degree prestige.

5 Historical development of standards

There are not many methods available for longitudinal analysis of network formation. Table 1 indicates, in general, higher degree prestige for standards that appear earlier, as one would expect as recent standards are built by referencing the older ones.

Table 3 shows changes in density of the network through years. Unlike earlier years of Internet standards, the density seems to decrease as standards becomes more specialized on certain issues. But it also appears to be stabilizing.

Table 4 shows changes in relative degree prestige of some key standards through years. There is no unique pattern to be found in this centrality

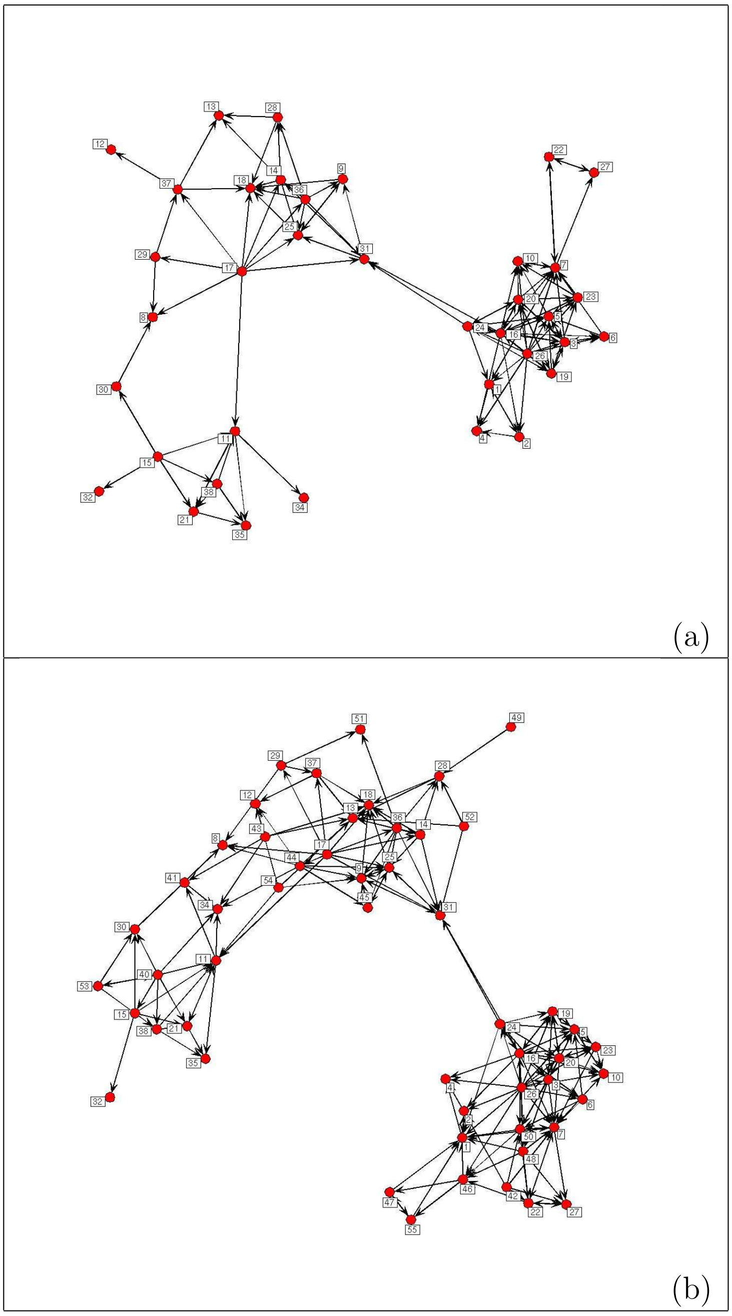

Figure 1: Groups of standard class RFCs. Top 38 and 55 nodes(according to degree prestige) are shown in (a) and (b), respectively. Three subgroups are identifiable : (1) the group on the right is "network management" related standards, (2) bottom-left group is mostly related to Internet protocol and its security extensions, and (3)top-left is a mixed group including standards such as domain names, e-mail content.

Year	Density
1992	0.033560
1995	0.017480
1998	0.011691
2001	0.008563
2004	0.007312

Table 3: Changes in the network density through years. Isolated nodes are excluded from the network before densities are calculated.

RFC	1992	1995	1998	2001	2004
RFC-1035	0.0242	0.0495	0.0608	0.0553	0.05584
RFC-1213	0.3273	0.2473	0.1597	0.1369	0.1269
RFC-1738	-	0.0177	0.0486	0.0415	0.0393
RFC-2045	-	-	0.0608	0.0636	0.06980

Table 4: Changes in relative degree prestige of some key standards through time.

measures. The ever top-ranking RFC-1213 shows a consistently decreasing prestige measure which stabilizes later. Where RFC-2*45 is quite stable through its lifetime, the other two exhibit a rather bumpy ride. Despite these differences, fluctuations in the degree prestige of all nodes seems to get smaller at later stages of their existence.

We have not included and *unsuccessful* examples in our study. Our evaluation of the longitudinal patterns of structuration is nowhere near to be comprehensive.

6 Conclusion

Our results for structural importance show that some nodes in the directed graph formed by IETF standards are far important than others, and network analysis methods, that are used more frequently in social sciences research, can be deployed for quantitative assessment of this difference. Centrality measures are valuable in determining which artifacts in a network are more influential, have effects when that are changed, or can deteriorate the overall quality of a system when they malfunction. In the case of IETF standards, more central standards are much harder to change as such a change would be reciprocated through a much greater number of artifacts than changes in less central nodes.

The measures we have applied for groups within the system is more crude than centrality measures. But nevertheless this group of methods are valuable in determining level of cohesion within the group, relative to outside of the group. More comprehensive application of such methods are used, for example in industrial cluster analysis. Further information can be obtained with this group of methods, for example in understanding density of relations between the groups.

One outcome of the first two group of analysis is the fact that standards related to network management are both more central individually and more cohesive as a group. This is a likely indication that specialization brings success in this system, and prompts for further comparison with similar systems.

Methods for historical analysis of such networks are limited [Wasserman, Faust 1994], but nevertheless our results showing the changes in network density and relative centrality measures confirm that the Internet standards are becoming more specialized over the time. However they do not indicate any possible ways of predicting future success of individual standards, as the fluctuation of centrality seem to exhibit many different patterns. We can not suggest purely quantitative methods for such prediction.

There are many more network analysis metrics available in addition to what was used in this study. These include methods for estimating relative advantage of some artifacts in networks in introducing inventions or changes(for example [Burt, 1995]). There has been criticisms in the past regarding the meaning of several network analysis instruments[Cook, Whitmeyer 1992]. Despite their value in quantitative assessment of structural features of interlinked artifacts, most network metrics has to be combined with due attention to the discourse of application.

Most parts of our analysis can be applied to similar groups of entities in different contexts. For example releases of Debian distributions are known to have timing problems. Identification of structural bottlenecks and subgroups among the package dependencies with these methods, can improve release schedules and further help in successful management of workforce allocation in such development efforts.

References

[Borgatti 2005] Borgatti, S.P. 2005. Centrality and network flow. *Social Networks*, 27.

[Bonacich, Lloyd 2001] Bonacich, P., Lloyd, P. 2001. Eigenvector-like measures of centrality for asymmetric relations.*Social Networks*, 23

[Burt, 1995] Ronald S. Burt, 1995, Structural holes : the social structure of competition. Harvard University Press.

[Cook, Whitmeyer 1992] Cook, K.S., Whitmeyer, J.M. 1992. Two Approaches to Social Structure: Exchange Theory and Network Analysis. *Annual Review of Sociology*, 18

[Kleinberg 1999] Kleinberg, J. 1999. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46/5

[Lehman, Ramil 2001] Lehman, M.M., Ramil, J.F. 2001. An Approach to a Theory of Software Evolution. *International Workshop on Principles of Software Evolution*, Vienna

[Lehmann 2004] Lehmann, F. (2004) FLOSS developers as a social formation. *First Monday.*

[Magnus, Szczepanska 2002] Bergquist, M., Szczepanska, A.M. 2002. Creating a Common Ground: Developing Discursive Practices as Means for Aligning. Systems Development Projects in Virtual Communities. *Proceedings of IRIS(Information Systems Research in Scandinavia) 25.* Copenhagen, Denmark.

[Oliver, Ebers 1998] Oliver, A.L., Ebers, M. 1998.Networking Network Studies: An Analysis of Conceptual Configurations in the Study of Interorganizational Relationships. *Organization Studies*, 19/4

[Oliver 2001] Oliver, A.L. 2001. Strategic Alliances and the Learning Life-cycle of Biotechnology Firms. *Organization Studies*, 22/3

[Powell, Koput, Smith-Doerr 1996] Powell, W.W., Koput, K.W., Smith-Doerr, L. 1996. Interorganizational Collaboration and the Locus of Innovation: Networks of learning in Biotechnology. *Administrative Science Quarterly*, 41.

[Raymond 2004] Raymond, E.S. 2004. The art of Unix programming. Addison-Wesley, Boston.

[Ring, van de Ven 1994] Ring, P.S., van de Ven, A.H. Developmental Processes of Cooperative Inter-organizational Relationships. *Academy of Management Review*, Vol. 19, No. 1.

[Stephenson, Zelen 1989] Stephenson, K., Zelen, M. 1989. Rethinking Centrality: Methods and Examples. *Social Networks*, 11.

[Tuomi] Tuomi, I. 2000. Internet, Innovation, and Open Source: Actors in the Network. *Association of Internet Researchers conference*, Lawrence, Kansas.

[Wasserman, Faust 1994] Wasserman, S., Faust, K. 1994. Social Network Analysis: Methods and Applications. Cambridge University Press.

[West 2002] West, J. 2002. How Open is Open Enough?: Melding Proprietary and Open Source Strategies. *Journal of Research Policy* Special Issue on "Open Source Software Development".

[White, Smyth 2003] White, S., Smyth, P. 2003. Algorithms for Estimating Relative Importance in Networks. *SIGKDD '03*.

A List of top ranking standards

Label	RFC
1	1213 : Mar-01-1991 Management Information Base for Network Management of TCP/IP-based internets:MIB-II
2	1212 : Mar-01-1991 Concise MIB definitions
3	2578 : April 1999 Structure of Management Information Version 2 (SMIv2)
4	1155 : May-01-1990 Structure and identification of management information for TCP/IP-based internets
5	2579 : April 1999 Textual Conventions for SMIv2
6	2580 : April 1999 Conformance Statements for SMIv2
7	1905 : January 1996 Protocol Operations for Version 2 of the Simple Network Management Protocol (SNMPv2)
8	2234 : November 1997 Augmented BNF for Syntax Specifications: ABNF
9	2045 : November 1996 Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies
10	1906 : January 1996 Transport Mappings for Version 2 of the Simple Network Management Protocol (SNMPv2)
11	2401 : November 1998 Security Architecture for the Internet Protocol
12	1035 : Nov-01-1987 Domain names - implementation and specification
13	1034 : Nov-01-1987 Domain names - concepts and facilities
14	2396 : August 1998 Uniform Resource Identifiers (URI): Generic Syntax
15	2460 : December 1998 Internet Protocol, Version 6 (IPv6) Specification
16	2571 : April 1999 An Architecture for Describing SNMP Management Frameworks
17	3261 : June 2002 SIP: Session Initiation Protocol
18	1123 : October 1989 Requirements for Internet Hosts - Application and Support
19	2575 : April 1999 View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)
20	2572 : April 1999 Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)
21	2406 : November 1998 IP Encapsulating Security Payload (ESP)
22	1902 : January 1996 Structure of Management Information for Version 2 of the Simple Network Management Protocol (SNMPv2)
23	2574 : April 1999 User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)
24	2573 : April 1999 SNMP Applications
25	2046 : November 1996 Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types

Label	RFC
26	2863 : June 2000 The Interfaces Group MIB
27	1903 : January 1996 Textual Conventions for Version 2 of the Simple Network Management Protocol (SNMPv2)
28	1738 : December 1994 Uniform Resource Locators (URL)
29	2327 : April 1998 SDP: Session Description Protocol
30	2373 : July 1998 IP Version 6 Addressing Architecture
31	2279 : January 1998 UTF-8, a transformation format of ISO 10646
32	1661 : July 1994 The Point-to-Point Protocol (PPP)
33	2205 : September 1997 Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification
34	1122 : October 1989 Requirements for Internet Hosts - Communication Layers
35	2409 : November 1998 The Internet Key Exchange (IKE)
36	2616 : June 1999 Hypertext Transfer Protocol – HTTP/1
37	1889 : January 1996 RTP: A Transport Protocol for Real-Time Applications
38	2402 : November 1998 IP Authentication Header
39	2246 : January 1999 The TLS Protocol Version 1
40	2461 : December 1998 Neighbor Discovery for IP Version 6 (IPv6)
41	1191 : Nov-01-1990 Path MTU discovery
42	1904 : January 1996 Conformance Statements for Version 2 of the Simple Network Management Protocol (SNMPv2)
43	2131 : March 1997 Dynamic Host Configuration Protocol
44	2822 : April 2001 Internet Message Format
45	2047 : November 1996 MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text
46	1573 : January 1994 Evolution of the Interfaces Group of MIB-II
47	1448 : April 1993 Protocol Operations for version 2 of the Simple Network Management Protocol (SNMPv2)
48	2233 : November 1997 The Interfaces Group MIB using SMIv2
49	2251 : December 1997 Lightweight Directory Access Protocol (v3)
50	1907 : January 1996 Management Information Base for Version 2 of the Simple Network Management Protocol (SNMPv2)
51	1766 : March 1995 Tags for the Identification of Languages
52	3280 : April 2002 Internet X
53	2462 : December 1998 IPv6 Stateless Address Autoconfiguration
54	2535 : March 1999 Domain Name System Security Extensions
55	1442 : April 1993 Structure of Management Information for version 2 of the Simple Network Management Protocol (SNMPv2)