

Organization of Internet Standards

Mehmet Gençer¹, Beyza Oba², Bülent Özel¹, and V. Sinan Tunaloğlu¹

¹ İstanbul Bilgi University, Department of Computer Science
Kurtuluş Deresi Cad. No:47 34440 Dolapdere, İstanbul, Turkey
{mgencer,bulento,vst}@cs.bilgi.edu.tr
WWW home page <http://cs.bilgi.edu.tr/~mgencer>

² İstanbul Bilgi University, Institute of Social Sciences
İnönü Cad. No:28 34387 Şişli, İstanbul, Turkey boba@bilgi.edu.tr

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. Our results demonstrate major groups centered around key standard tracks, and application of network metrics reflect different levels of cohesion for these groups. As application of these techniques in such domains is unusual, possible uses in open source projects for strategizing are discussed.

1 Introduction

Open Source Software(OSS) 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 social network analysis methodology[15]. Most such externalities are formalized elsewhere by authoritative bodies of standardization, with close cooperation with the open source 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.

Please use the following format when citing this chapter:

Gençer, M., Oba, B., Özel, B., and Tunaloğlu, V.S., 2006, in IFIP International Federation for Information Processing, Volume 203, Open Source Systems, eds. Damiani, E., Fitzgerald, B., Scacchi, W., Scotto, M., Succi, G., (Boston: Springer), pp. 267-272

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 importance of Internet standards, and inter-dependencies among them using techniques from social network analysis practice, (2)to 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)to 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, and similar strategizing tasks.

An overview of data and the network analysis approach is summarized in section 2. Results for influence and its historical development are presented in section 3, and findings for specialization in section 4. An overview of results and possible other applications of social network analysis methods in OSS processes are discussed in section 5.

2 Standards data and network analysis methods used

Software development processes are studied for the mechanisms of their evolution as a coherent system, and as a community practice of actors[13, 6]. Other research on software call our attention to importance of discursive practices and alignment of software development efforts[10]. 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[1, 12, 11, 7, 8, 9]. 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[5]. 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[4, 14, 2].

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 referral relations between 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. Selection of structurally important standards based on prestige measures. These standards have more influence than others.
2. Identification of subgroups formed around influential standards, key technological questions addressed by them, and their cohesion levels. Subgroups, analysis of their cohesion, and connections between subgroups are important in understanding specialization in growing networks.
3. Sampling of historical patterns of centrality metrics for some key standards and demonstration of stabilization patterns in structural development of Internet standards.

3 Structural importance and influence

Degree prestige(number of references) and relative in-degree prestige[14] are used for assessment of structural importance of a node in a network. Table 1 shows top 15 RFCs according to these centrality measures. Fig.1 is a graphical representation of top 55 nodes, where labels reflect the ranking of RFCs. Density(ratio of existing relations to possible number of relations between nodes) of the RFC network is found to be 0.003716. As best demonstrated by top 7 nodes which have many relations to each other, success of an Internet standard is closely related to its positioning with other standards and success of its siblings.

3.1 Historical development of influence

There are not many established methods available for longitudinal analysis of network formation. One would expect that standards that appear earlier would have higher centrality measures as recent standards are built by referencing the older ones. However results shown in Table 1 only partially confirms this insight.

Table 2a shows changes in density of the network through years. Unlike earlier years of Internet standards, the density decreases as standards becomes more specialized on certain issues, but the rate of decrease is becoming lower.

Table 2b shows changes in relative degree prestige of some key standards through years. This sample is insufficient to suggest a unifying pattern. However it is worth noting that in all of the first three cases, centrality measure first rises to a climax, followed by a decrease as the standard ages and possibly replaced by newer versions at a later stage.

4 Subgroups and specialization

Many standards are related to some others in terms of the technical issues they address. Fig.1 shows how relations concentrated around standards that are influential. Three major groups are identifiable in the network. One group

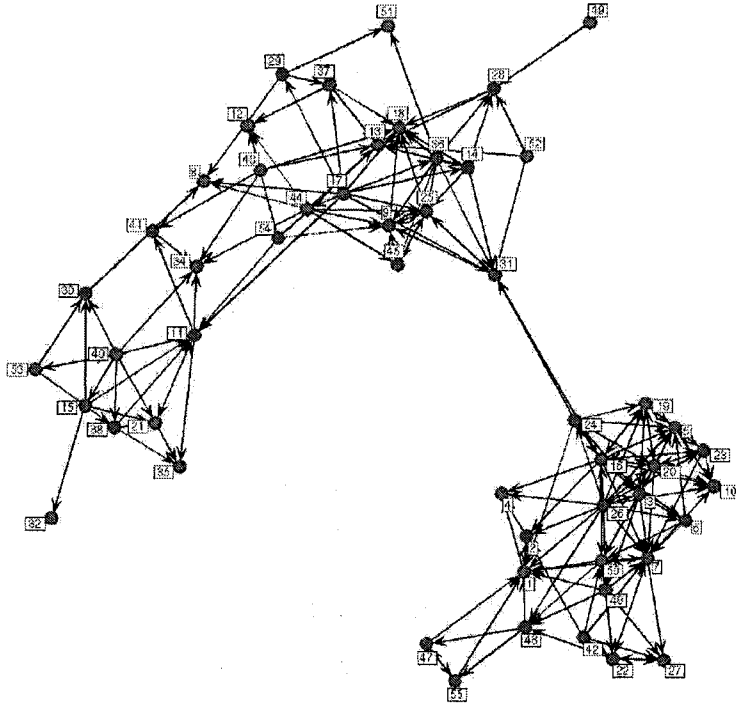


Fig. 1. Groups of standard class RFCs. Top 55 nodes according to degree prestige. Three subgroups are identifiable here: (1) the group on the bottom-right is "network management" related standards, (2) bottom-left group is mostly related to Internet protocol and its security extensions, and (3) top-middle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

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, however there are many links between the second and third groups.

Further assessment 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[14]. 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(network management protocols) which exhibit this level of cohesion. Its only link with other major standards is indirectly

Table 1. Top ranking RFCs according to in-degree and relative in-degree prestige measures.

	Deg.(rel)	RFC: Year, Short title
1	141(0.0966)	1213 : 1991, MIB-II for Network Man. of TCP/IP internets
2	129(0.0884)	1212 : 1991, Concise MIB definitions
3	127(0.0870)	2578 : 1999, Structure of Management Information(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 SNMPv2
8	108(0.0740)	2234 : 1997, Augmented BNF for Syntax Specifications
9	89(0.0610)	2045 : 1996, MIME Part One: Format of Internet Message Bodies
10	89(0.0610)	1906 : 1996, Transport Mappings for SNMPv2
11	79(0.0541)	2401 : 1998, Security Architecture for the IP
12	76(0.0521)	1035 : 1987, Domain names - implementation and spec.
13	72(0.0493)	1034 : 1987, Domain names - concepts and facilities
14	69(0.0473)	2396 : 1998, URI: Generic Syntax
15	64(0.0438)	2460 : 1998, IPv6 Specification

Table 2. Historical changes in structure: (a)changes in the network density through years, and (b)changes in relative in-degree prestige of some RFCs.

(a)		(b)					
Year	Density		1992	1995	1998	2001	2004
1992	0.033560						
1995	0.017480	RFC-1035(1987)	0.0242	0.0495	0.0608	0.0553	0.05584
1998	0.011691	RFC-1213(1991)	0.3273	0.2473	0.1597	0.1369	0.1269
2001	0.008563	RFC-1738(1994)	-	0.0177	0.0486	0.0415	0.0393
2004	0.007312	RFC-2045(1996)	-	-	0.0608	0.0636	0.06980

through node 31(UTF-8 standard), which has an important role in this sense not captured by degree prestige measure.

5 Conclusion

Our results for structural features of the interrelated system of IETF standards demonstrate that methods from social network analysis can be applied to standards or software processes, and to our best knowledge such cross applications of these methods are rare. Structural measures are valuable in determining which artifacts in a system are more influential, can deteriorate the overall quality of a system when they malfunction, or whether introduction of new relations may compromise integrity. As our results suggest, higher levels of subgroup cohesion(i.e. refined specialization) brings success.

Our research was limited as there are many more centrality measures such as betweenness centrality[14, 3]. These were not preferred in this study as their interpretation may be problematic in a first probation, compared to more direct measures we have used. However, note that role of some standards such as UTF-8 which is not captured by prestige centrality can be successfully assessed by incorporation of other measures, such as betweenness centrality. There has been criticisms in the past regarding the meaning of several network analysis instruments[4]. 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.

Approaches for historical analysis of such networks are limited in the literature[14]. Our results are very limited but nevertheless hints on existence of common patterns. Further research is needed, for example to understand whether subgroup cohesion levels show any such patterns over time, or whether any of these instruments can be consolidated into models for forecasting structural features.

Most parts of our analysis can be applied to similar systems. For example releases of Debian distributions are known to have timing problems. Identification of structural bottlenecks and subgroups in software processes, can improve release schedules and further help in successful management of workforce allocation in such development efforts.

References

1. Borgatti SP (2005) *Social Networks* 27:
2. Bonacich P, Lloyd P (2001) *Social Networks* 23/3:191–201
3. Burt RS (1995) *Structural holes : the social structure of competition*. Harvard University Press, Cambridge, Mass
4. Cook KS, Whitmeyer JM (1992) *Annual Review of Sociology* 18/1:109–127
5. Kleinberg J (1999) *Journal of the ACM* 46/5:604–632
6. Lehmann F (2004) *First Monday* 9/11
7. Oliver AL, Ebers M (1998) *Organization Studies* 19/4:549–583
8. Oliver AL (2001) *Organization Studies* 22/3:467–489
9. Powell WW, Koput KW, Smith-Doerr L (1996) *Administrative Science Quarterly* 41:116–145
10. Raymond ES 2004 *The art of Unix programming*. Addison-Wesley, Boston
11. Ring PS, van de Ven AH (1994) *Academy of Management Review* 19/1:90–118
12. Stephenson K, Zelen M (1989) *Social Networks* 11:1–37
13. Tuomi I (2000) *First Monday* 6/1
14. Wasserman S, Faust K (1999) *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge New York
15. West J (2002) *Journal of Research Policy* 32/7:1259–1285