Standards as Interdependent Artifacts: Application to Prediction of Promotion for Internet Standards

Mo-Han Hsieh Engineering Systems Division, Massachusetts Institute of Technology Email: mohan76@mit.edu Christopher L. Magee Engineering Systems Division & Department of Mechanical Engineering, Massachusetts Institute of Technology Email: cmagee@mit.edu

Abstract

We report on a study of a set of standards as a system and examine by network analysis techniques the interdependency among these standards. We provide empirical evidence that a standard's interdependency with other standards is correlated with the standard's level of acceptance. By using the standards developed by the Internet Engineering Task Force (IETF), we construct a citation network of the standards and a coauthor network of the standards' developers. We demonstrate that standards in a higher maturity level have higher prestige and higher implicit social influence. By studying the IETF standards over time, we also demonstrate that the interdependency among a related set of standards has predictive power for whether a standard increases its level of acceptance at a later date.

1. Introduction

Technical standards^{*} play important roles in business activities and industrial development. Companies now spend large amount of resources to follow and participate in the development of standards and strategically adjust their position as being the leaders or followers of setting specific standards (Funk and Methe 2001). Though the choices of standardization range from governmental processes through standards setting organizations to *de facto* processes of market competition, companies have to make decisions based on the trade-off between the predictable market acceptance and the speed of setting technical specifications (Cargill 1998). Being aware of some infamous cases that the "standards war" (Shapiro and Varian 1999) caused excessive loss for both the participants and the market, more and more companies are willing to hedge their standards setting activity by participating in the cooperative work within organizations.

Usually, setting standards through the operation of formal or informal organizations can reduce transaction costs, streamline the information exchange, and facilitate negotiation among key players (Farrell and Saloner 1988). Moreover, voluntary standards organizations can workout compromises on various issues, preserving more social welfare in the process. With the benefits that the formal standards setting organization can potentially provide, these organizations are, nevertheless, composed of participants that have their respective and usually very different goals. Practitioners participate in the standards setting organizations for different reasons. They may have their own agenda to influence the adoption of certain standards, or they may just want to be more informed about the latest development of standards (Schmidt and Werle 1998). Though companies prefer setting standards through organizations for its more predictable market acceptance, with the combination of different intentions and behaviors among the participants, plus the typical prolonged process of reaching consensus, the level of market acceptance when a standard has finally gone through the process can be hard to predict.

Standards depend on each other. Very often, a later developed standard incorporates some previously developed standards as part of its specification. With the notion of the interdependency of standards, the uncertainty of a standard's acceptance can possibly be reduced if viewed in the context of a relevant set of standards. In other words, instead of viewing a standard as a standalone technical specification, viewing the standard in terms of its interdependency with other standards has the potential of helping us better evaluate its market acceptance.

The term, "standard", is broadly used in different industrial and technical contexts as the allied terms such as protocol, agreement, convention, etc. Because these terms are usually used in similar fashion in different engineering fields, it is not possible to consistently differentiate among them. In this paper, we simply use "standard" to cover all of them when used in a technical context.

In this paper, we empirically study whether a standard's interdependency with other standards is correlated with the standard's level of acceptance. We examine whether the adoption of one standard in another standard represents an indication that the developers have in some measure affirmed the authority of the cited standard. Moreover, we empirically examine whether the status of standards is correlated with the social influences of the standards' developers. We also explore whether the interdependency among a relevant set of standards has predictive power in the dynamic process of new standards continuously being proposed and accepted or not.

This paper is organized as follows. In Sec. 2, we describe the rationale and methodology of our analysis. In Sec. 3, we introduce the empirical data used in our investigation. In Sec. 4, we discuss the relationship between the prestige of standards and their level of acceptance. In Sec. 5, we discuss the relationship between the social influences of the standards' authors and the related standards' level of acceptance. In Sec. 6, we test a logistic regression model for its effectiveness in predicting a standard's future acceptance and compare this to the actual data. In Sec. 7, we provide discussion and conclusions.

2. Standards as Interdependent Artifacts

While specifying the elements of a technology or an engineering system, standards depend upon each other. Very often, a later developed standard incorporates some previously developed standards as part of its specification. In fact, it is a common practice that a new standard explicitly "cites" some previously developed standards as part of its implementation requirements. For example, among the Internet standards, the Simple Mail Transfer Protocol (SMTP) uses the Transmission Control Protocol (TCP) as one of its transport services, while the SMTP is itself adopted by another Internet standard, the SMTP Service Extension for Message Size Declaration (SMTP-SIZE), as part of its specification.

With the incorporation of other standards, a new standard does not have to specify every technical detail from scratch; moreover, the chance of conflicts among standards is significantly decreased. As more and more new standards are developed based upon the previously developed standards, the interdependency among standards gradually increases, and an interdependency network of standards thus emerges.

One representation of the network structure of a relevant set of technical standards, with the standards as nodes and the citation as directed links, is analyzed in this paper. We intend to explore whether the citation network of standards gives us a way to identify the interrelationships among various components of these technologies or engineering systems. It can be hypothesized that citations among standards encode a considerable amount of latent human judgment, and this type of judgment can be an important element underlying the development of the technology or engineering system.

As with all technical artifacts, non-trivial social influences upon the standard setting process are expected. We examine one social aspect of a relevant set of standards by looking into the set's coauthor network. In this network, the nodes are the developers of the standards, and the undirected links are the co-authorship of the standards' developers. The coauthor network can be related to the citation network of standards by assigning the nodal properties of the coauthor network to the corresponding standards. In this case, the hypothesis that the interdependency among standards is gradually developed under the influence of not only the technical considerations but also social considerations can be tested.

3. The Case: Internet Standards

In this research, we took the local approach of studying the interdependency among a relevant set of standards to avoid the problems of representing and filtering large volumes of information. To make observation and analysis about the differences between standards with high acceptance and that without, the standards of Internet, developed by the Internet Engineering Task Force (IETF), were used.

In the IETF practice, standards were published as part of the Requests for Comments (RFC) document series. With the first of this series, RFC1, been published in April 1969, the series currently has more that 4,400 RFCs. Cumulatively, there have been about 2,800 authors who have participated in writing these RFCs. One special (and useful for this research) feature of the RFC document series is that, once a RFC has been published, the document contents never changes. Any attempt to revise the previously published RFC is in the form of a new RFC. Therefore, the new RFCs can update or obsolete the previous RFCs, but the old RFCs would never disappear from the public database. Since the IETF periodically publishes a RFC that lists the latest "Official Internet Protocol Standards", historical snapshots

of the IETF official standards in a number of different times can be captured.

In the IETF practice, the official standards were categorized into three maturity levels – the "Proposed Standard", the "Draft Standard", and the "Internet Standard". With the Proposed Standards being the entry level specifications and the Internet Standards the most recognized standards, the differences among the three maturity levels generally suggest different level of endorsements from IETF on the standards. Vendors or practitioners of the Internet take these recommendations when implementing the standards in various environments (Bradner 1996).

With these differences, the three-level categorization can be viewed as a natural reference to the standards' level of acceptance. Therefore, the relationship between the standards' interdependency and the standards' actual level of acceptance can be examined. This also allows us to test the methods that predict the acceptance level of standards based upon the interdependency among standards.

To take the approach of studying the interdependency of a relevant set of standards, the official Internet standards announced in 2004 was used as the basis for further analysis. The "Official Internet Protocol Standards" published in 2004 (i.e. RFC3700) has 998 official standards. Among the 998 official standards, 848 were Proposed Standards, 75 were Draft Standards, and 75 were Internet Standards.

Since it is a common practice for a later developed IETF standard to explicitly incorporate previously developed IETF standards as part of its specification, in most of these standards, references to other standards were typically provided in a dedicated section. With these standards and their references, a network of the standards' technological interdependency can be generated. In this network, the nodes are the 998 standards, and directed links are the 4,418 references pointing among them.

For the IETF standards, each standard was developed by either one or several developers. Since the first RFC published in 1969, a great number of people have participated in the IETF community to develop standards for the Internet. As of 2006, there have been about 2,800 developers whose names had appeared in at least one of the 4,400 RFCs. A social network of these standards developers is likely to exist by various means. In this research, we focused on the co-authorship among the standards' developers. By identifying whether or not the names of two developers had ever appeared as the authors of the same RFC, we can construct the coauthor network of the RFC developers. In this social network, the nodes are the developers, and the undirected links are the co-authorship between them.

Since we are interested in the social influences that any specific developer has accumulated over the years in the IETF community, when constructing the social network, we included all of the developers whose names had ever appeared in the RFCs that were published before the publication of the 2004 IETF official standards (i.e. RFC3700). In this case, there are 2,550 developers included in our social network.

The citation network of the standards and the coauthor network of the standards' developers are the starting points of our further analysis.

4. The Prestige of Standards

By examining the citation network of the 2004 IETF official standards, we observed that relationships exists between the authoritative standards that receive large amounts of citations and standards that cite many authoritative standards. With this observation, the link-based algorithm developed by Kleinberg (1999) for identifying the *authority* and *hub* of a directed network appears quite relevant.

Kleinberg proposed an algorithm that quantifies the rank that a particular node has within a set of nodes. He refers to a node that is pointed to by highly ranked nodes as an authority, while a node pointing to highly ranked nodes is referred to as a hub. In this paper, we call the two kinds of weights that he assigns to each node the *prestige* and the *acquaintance*. The prestige represents the weighted proportion of links pointing to a node, and the acquaintance represents the weighted proportion of links pointing from a node.

In Kleinberg's algorithm, a node's prestige depends simultaneously on the acquaintance of nodes pointing to it, while a node's acquaintance depends simultaneously on the prestige of nodes it points to. A simple way of presenting the solution to this iterative referencing situation is to define x_i as the prestige and y_i as the acquaintance of node *i*. With a citation network having *n* standards, by defining the adjacency matrix *A* to be the *n* by *n* matrix whose (*i*, *j*) entry equals to 1 if standard *i* has standard *j* as its reference, we have the iterative relationship between a node's prestige and its acquaintance as:

 $Ax = \lambda y, \quad A^{\mathsf{T}}y = \mu x \qquad (1),$

where λ and μ are the scaling factors. In the situation where there is no multiplicity of the principal

eigenvalue, Kleinberg (1999) proposed to use the principal eigenvector of $A^T A$ and AA^T as the prestige and acquaintance for the nodes respectively.

After applying the algorithm to the citation network of the 2004 IETF official standards, we ranked the 998 standards according to their value of prestige. If two standards had the same value of prestige, we gave them the same rank.

According to our results, among all of the 998 official standards, the RFC2578 has the highest prestige. We will not argue in this paper that RFC2578 is really the most important standard. We are interested in only whether the more important standards tend to have higher prestige as calculated from Kleinberg's algorithm. In this regard, some obviously important standards, for example, the Internet Protocol (RFC791) and the User Datagram Protocol (RFC768), are among the standards having prestige rankings in the top five percent.

It is worth noticing that the top five percent ranked standards are composed of 22.7 percent of the Internet Standards, 5.3 percent of the Draft Standards, and 1.2 percent of the Proposed Standards. Since we are more interested in whether the prestige of standards has any relationship with the standards' level of acceptance, we conducted statistical tests that compared the prestige of standards in different maturity levels. Because our observations were not independent, instead of using parametric tests, we used randomization tests (Edgington 1995) to obtain the significance of the mean level differences of the prestige.

Table 1 is the result of pairwise comparison of the mean level differences of the prestige for standards having different maturity levels.

Basic statistics on each group							
			Maturity Level ¹				
		1	2	3			
Me	ean	0.0015	0.0059	0.0376			
Std	Dev	0.0045	0.0204	0.1066			
Num. Of Observation		848	75	75			
Significan	Significance Tests						
Maturity Level		Difference in Means	One-Tailed Tests				
Group 1	Group 2	(Group 1 - Group 2)	Group 1 > Group 2	Group 2 > Group 1			
2	1	0.004	0.0003 *	0.9997			
3	1	0.036	0.0000 *	1.0000			
3	2	0.032	0.0036 *	0.9964			

Table 1.	Pairwise	comparison	of the mean	prestige	versus	maturity	levels
----------	----------	------------	-------------	----------	--------	----------	--------

¹ Maturity Level: 1 = Proposed Standard, 2 = Draft Standard, and 3 = Internet Standard

* Statistically significant at, at least, the .05 level.

In Table 1, the average prestige of the Draft Standards is significantly higher than that of the Proposed Standards, and that of the Internet Standards in turn is significantly higher than that of the Draft Standards. With this result, we conclude that standards at higher acceptance levels have higher prestige. For the IETF 2004 official standards, a standard having higher prestige in general implies that it has a higher level of acceptance. We hypothesize this to be true for standards outside the realm of the IETF, though most of these standards provide no natural reference like the maturity level to be tested against.

5. The Social Influence of Standards

Based upon the coauthor network of the RFC writers, we want to infer the social influences of the standards' developers in the IETF community. These inferred social influences can later be related to the standards by assigning the nodal properties of the coauthor network to the corresponding standards.

Among a great number of social network metrics designed to differentiate between important and non-important actors, the concept of "Betweenness Centrality" developed by Freeman (1977) has been most widely accepted and used by sociologist (Wasserman and Faust 1994) and was used in this research.

For a network, Freeman defined a node's betweenness centrality as the total number of times that a node is in the geodesics of a pair of nodes other than itself. For the pair of nodes with more than one geodesic, we count the number of times that the node is in the geodesics of the pair and divide it by the total number of geodesics of the pair. Therefore, for an undirected network with n nodes, the betweenness centrality for node n_i is simply (Wasserman and Faust 1994),

 $C_B(n_i) = \sum_{j \neq k} g_{jk}(n_i) / g_{jk}$, (2) where *i* is distinct/from *j* and *k*, g_{jk} is the number of geodesics linking *j* and *k*, and $g_{jk}(n_i)$ is the number of such geodesics that contain actor *i*.

The betweenness centrality can be normalized as:

 $C'_{R}(n_{i}) = C_{R}(n_{i})/[(n-1)(n-2)/2],$ (3)

where the denominator, (n-1)(n-2)/2, is simply the maximum possible number of geodesics among all pairs of nodes other than the ego.

After calculating the betweenness centrality of the RFC authors using the coauthor network in 2004, Jon B. Postel was the most influential RFC author. Though Postel's massive contribution as the sole editor of the RFCs for 28 years is well recognized by the IETF community, we will not argue in this paper whether he was truly the most influential author in the IETF community.

With the betweenness centrality of the developers of the 2004 official standards, we then summed these values according to their appearances in different RFCs as the cumulative social influences of the authors of these RFCs. For example, if a 2004 IETF official standard was written by three developers, we sum the betweenness centrality of these three developers and assign the result to the standard. In this case, if a standard's developers have higher betweenness centrality, the standard would have a higher In this paper, we call a standard's summation of its authors' normalized score assigned to it. betweenness centrality the "implicit social influence (ISI)" of the standard.

After assigning every 2004 IETF standard its ISI, we then ranked the 998 standards according to these assigned values. If two standards had the same ISI, we gave them the same rank

According to our results, among all of the 998 official standards, the RFC3035 has the highest ISI. Again we will not argue in this paper that RFC3035 is the most important Internet standard but whether this ranking does correlate with the standards' level of acceptance. In this regard, the top five percent ranked standards are composed of 18.7 percent of the Internet Standards, 5.3 percent of the Draft Standards, and 2.1 percent of the Proposed Standards. Since we are more interested in whether a standard's ISI has any relationship with the standard's level of acceptance, we conducted statistical tests that compare the ISI of standards having different maturity levels. Because our observations were not independent, we again used randomization tests to obtain the significance of the mean level differences of ISI.

Table 2 is the result of pairwise comparison of the mean level differences of the ISI for standards having different maturity levels.

Basic statistics on each group						
		Maturity Level ¹				
	1	2	3			
Mean	0.0099	0.0128	0.0249			
Std Dev	0.0138	0.0164	0.0214			
Num. Of Observation	848	75	75			

Table 2. Pairwise com	parison of the	mean ISI versus	maturity levels
-----------------------	----------------	-----------------	-----------------

Maturity Level		Difference in Means	One-Tailed Tests		
Group 1	Group 2	(Group 1 - Group 2)	Group 1 > Group 2	Group 2 > Group 1	
2	1	0.0029	0.0468 *	0.9532	
3	1	0.0150	0.0000 *	1.0000	
3	2	0.0121	0.0000 *	1.0000	

¹ Maturity Level: 1 = Proposed Standard, 2 = Draft Standard, and 3 = Internet Standard

* Statistically significant at, at least, the .05 level.

Significance Tests

In Table 2, the average ISI of the Draft Standards is significantly higher than that of the Proposed Standards, and that of the Internet Standards is significantly higher than that of the Draft Standards. With this result, for the IETF 2004 official standards, a standard having higher ISI in general implies that it has a higher level of acceptance. We hypothesize this to be true for standards outside the realm of the IETF.

Regarding the association between a standard's prestige and its ISI, with the Pearson correlation coefficient at the level of 0.205, we can only infer a weak association between the two. More investigation is needed in order to understand the association between the two.

6. Predicting a Standard's Future Acceptance

From the previous results, a standard with higher prestige and implicit social influence is more likely to be in higher maturity levels. With the result being statistically significant, we further explore how useful that higher prestige and implicit social influence is in predicting the promotion of a Proposed or Draft Standard (i.e. increase its maturity level in the future.)

To explore this possibility, we fitted a logistics regression model with the data of 1994 IETF official standards (i.e. RFC1610). In the logistics regression model, the binary dependent variable is whether or not the Proposed or Draft Standard would increase its maturity level in the next five years. The variable is coded one if a Proposed or Draft Standard had increased its maturity level between year 1995 and 1999; otherwise the variable is coded zero. The independent variable is the prestige and the implicit social influence of the standards in 1994. The model was later verified with the data of the 1999 IETF official standards (i.e. official standards listed in RFC2500).

The regression models fitted with the 1994 IETF official standards are shown in Table 3. Though there are different criteria of choosing a cut value for categorizing model outputs, we first used 0.5 as the cut value in these models. Influences of different cut values are examined later in this paper.

	Model 1		Model 2		Model 3	
Variable Coefficient Wald		Coefficient	Wald	Coefficient	Wald	
Constant	-1.482 *	7.886	-1.499 *	31.947	-1.497 *	31.605
Prestige	31.685 *	40.862			30.921 *	5.689
Implicit Social Influence			0.227 *	4.585	0.016	0.014
Model Chi-square [df]	12.759 [1]		4.474 [1]		12.773 [2]	
% Correct Predictions	79.9		75.5		79.9	
Hosmer & Lemeshow Test	0.403		0.114		0.364	
Note: The Wald statistics are distributed chi-square with 1 degree of freedom.						
* Indicates that the coefficient is statistically significant at, at least, the .05 level.						

 Table 3. Regression models fitted with the 1994 IETF official standards

Model 1 was fitted with the prestige of standards as the sole independent variable. The Wald statistics indicates that the variable is significant at the 0.05 level. Model 2 was fitted with the implicit social influence of standards as the sole independent variable. The Wald statistics also indicates that the variable is significant at the 0.05 level. Model 3 was fitted with both the prestige and the implicit social influence as the independent variables. The Wald statistics indicates that the prestige of standards is significant at the 0.05 level but the implicit social influence is not. Though the Hosmer and Lemeshow goodness-of-fit test indicates that all of the three models are well fitted (i.e. with statistics greater than 0.05), Model 3 is discarded due to the non-significance of one of its independent variable. With both Model 1 and Model 2 being feasible models, we chose Model 1 (with higher Chi-square statistics) for further demonstration. The classification table of Model 1 is shown in Table 4.

Table 4. Classification table of Model 1 (with 0.5 as the cut value)

	Predicted		
Observed	0	1	Percentage Correct
0	105	1	99.1
1	28	5	15.2
Overall Percentage			79.1

According to Table 4, with the cut value of 0.5, if a Proposed or Draft Standard in 1994 increased its maturity level in the next five years, Model 1 has 15.2 percent correct rate of making the prediction at 1994. Similarly, if a Proposed or Draft Standard in 1994 did not increase its maturity level in the next five years, Model 1 has 99.1 percent correct rate of making the prediction at 1994. The overall percentage of correct rate is about 79.1 percent. In this case, the false positive rate is 1/(1+5)=0.17, and the false negative rate is 28/(28+105)=0.21.

It should be noticed that, by changing the cut value for categorizing model outputs, the correct rate of prediction can be changed. However, there is a trade-off between the correct rate of prediction, the false positive rate, and the false negative rate, depending on the costs of making wrong categorizations.

To better evaluate the model, the cumulative gains chart is drawn in Fig. 1 and the lift chart is drawn in Fig. 2. Lift is a measure of the effectiveness of a predictive model. It is the ratio between the positive prediction obtained with and without the model. The cumulative gains and lift charts can help us assess model performance; the greater the area between the lift curve and the baseline, the better the model.



Figure 1 and Fig. 2 both show that our model is effective in making predictions. As mentioned previously, we used 0.5 as the cut value for categorizing whether a Proposed or Draft Standard would increase its maturity level in the next five years. In this case, we made only six positive predictions, which is about four percent of all cases. According to Fig. 1, if we are willing to predict 20 percent of all cases (i.e. about 28 cases) as positive, we can capture about 40 percent of all positive cases (i.e. about 13 cases). In Fig. 2, 20 percent of all cases correspond to the lift of about 1.8, which means our model is 1.8 times more effective than making prediction by random guess.

Finally, the model constructed by using the data of 1994 IETF official standards was verified by using the data of 1999 IETF official standards. With 0.5 as the cut value, the categorization table is shown in Table 5. The Cumulative Gains Chart of the verification data is shown in Figure 3.



According to Table 5, with the cut value of 0.5, if a Proposed or Draft Standard in 1999 increased its maturity level in the next five years, Model 1 had 11.4 percent correct rate of making the prediction. In this case, we made only five cases of positive prediction, which is about 1.1 percent of all cases. The cumulative gains chart in Fig. 3 shows that the model is effective in making prediction. To increase the

correct rate of capturing positive cases, we may increase the number of positive prediction. However, there is a trade-off between the correct rate and false positive rate as mentioned previously.

After verifying the model with the data of 1999 IETF official standards, we conclude that the logistic regression model with the prestige of standards as the sole independent variable has its merit in predicting whether a Proposed or Draft Standard would increase its maturity level in the next five years. Though not demonstrated in this paper, Model 2, which used the implicit social influence of standards as the sole independent variable, has the same effectiveness in making the prediction.

7. Conclusion and Future Research

Standards are important for modern economic activity and industrial development. This paper takes a systems approach to look at a relevant set of standards as a "standards system". We argue that the interdependency among a relevant set of standards can provide insights into the development of the technology or the engineering system.

In this paper, we provided empirical evidence that a standard's interdependency with others is correlated with the standard's level of acceptance. By using the IETF standards as our data, we demonstrated that a standard in the higher maturity level has, on average, higher prestige. To some extent, it implies that the adoption of one standard in another standard represents an indication that the developers have in some measure affirmed the authority of the cited standard. Moreover, we demonstrated that a standard in the higher maturity level has, on average, higher implicit social influence. This finding implies that the acceptance of a standard has a positive correlation with the social influences of the standards' developers.

To test the hypothesis that a Proposed Standard or a Draft Standard would be more likely to increase its maturity level in the future if it has higher prestige or implicit social influence, we fitted a logistic regression model with the data of 1994 IETF official standards. The regression model, verified with the 1999 IETF official standards, is effective in predicting whether a Proposed or Draft Standard would increase its maturity level in the next five years. The model has either the prestige or the implicit social influence of standards as the sole independent variable. It has successfully demonstrated the usefulness of studying the interdependency among standards.

References

Bradner, S. (1996). The Internet Standards Process -- Revision 3. <u>RFC2026</u>, Internet Engineering Task Force.

Cargill, C. F. (1998). "Standardization: art or discipline?" <u>IEEE Micro</u> 18(3): 18-24.

Edgington, E. S. (1995). Randomization tests. New York, M. Dekker.

Farrell, J. and G. Saloner (1988). "Coordination through Committees and Markets." <u>Rand Journal of</u> <u>Economics</u> **19**(2): 235-252.

Freeman, L. C. (1977). "A set of measures of centrality based on betweenness." Sociometry 40: 35-41.

- Funk, J. L. and D. T. Methe (2001). "Market- and committee-based mechanisms in the creation and diffusion of global industry standards: the case of mobile communication." <u>Research Policy</u> **30**(4): 589-610.
- Kleinberg, J. M. (1999). "Authoritative sources in a hyperlinked environment." Journal of the Acm **46**(5): 604-632.

Schmidt, S. K. and R. Werle (1998). <u>Coordinating technology: studies in the international standardization</u> of telecommunications. Cambridge, Mass., MIT Press.

Shapiro, C. and H. R. Varian (1999). "The art of standards wars." <u>California Management Review</u> **41**(2): 8-+.

Wasserman, S. and K. Faust (1994). <u>Social network analysis: methods and applications</u>. Cambridge; New York, Cambridge University Press.