

MULTI-DATA MINING FOR UNDERSTANDING LEADERSHIP BEHAVIOR

Naohiro Matsumura^{1} and Yoshihiro Sasaki²*

^{1}Graduate School of Economics, Osaka University, 1-7 Machikaneyana, Toyonaka, Osaka, 560-0043 Japan*

E-mail: matumura@econ.osaka-u.ac.jp

²School of Economics, Osaka University, 1-7 Machikaneyana, Toyonaka, Osaka, 560-0043 Japan

E-mail: cdg36390@par.odn.ne.jp

ABSTRACT

We propose an approach for understanding leadership behavior in dot-jp, a non-profit organization, by analyzing heterogeneous multi-data composed of questionnaires and mailing list archives. Attitudes toward leaders were obtained from the questionnaires, and human networks were extracted from the mailing list archives. By integrating the results, we discovered that leaders must receive messages from other people as well as send messages to construct reliable relationships.

Keywords: Leadership behavior, IDM, Non-profit organization (NPO)

1 INTRODUCTION

We studied dot-jp (<http://www.dot-jp.or.jp/>), a non-profit organization (NPO) in Japan that organizes seminars and internship programs, which give university students the opportunity to participate in political activities with diet members. Through the internship program, students learn how diet members engage in political activities.

The headquarters of dot-jp is in Osaka, and seven branch offices are distributed all over Japan (Branches A, B, C, D, E, F, and G¹). Each branch office has nine to twenty one staff members, and one of the staff members is appointed to be the area manager responsible for managing all staff members.

Most staff members in dot-jp are university students, and having face-to-face meetings with each other is difficult because of the distances involved. For these reasons, staff members primarily use e-mail to exchange information, plan and arrange events, and discuss matters. Communication via e-mail creates a

¹ We used fake names to maintain secrecy.

virtual office and complements real office communication. An overview of the e-mail archives in the seven branch offices of the 14th period (from October 2004 to March 2005) that we analyzed is shown in Table 1. Thousands of e-mails were exchanged during the period.

Table 1. Number of staff members and e-mails exchanged during the 14th period (October 2004 – March 2005)

	Branch A	Branch B	Branch C	Branch D	Branch E	Branch F	Branch G
# of staffs	21	9	16	14	16	16	12
# of e-mails	2297	1198	2465	2076	3258	1309	1717

Much research on NPOs has showed that reliable relationships among staff members are crucial to make the most of human and knowledge capital (Krackhardt & Hanson, 1993; Drucker, 1993; Wenger, 1999). Leadership behaviors also play an important role in determining the atmosphere or culture of an organization (Perkins & Wilson, 1999) and the ability to create knowledge from the experiences of staff members (Krogh, Ichijo, & Nonaka, 2002). In an ongoing project about NPOs (OSSIP04, 2004), the aspects of capital, scale of operation, human resources, and partnerships with governments and organizations are being studied. However, the relationships among staff members are not being investigated although they are quite relevant for sharing missions and motivating voluntarism.

We propose an approach for understanding leadership behavior in organizations by using questionnaires and human influence networks. We first determined the results of the questionnaire for 97 staff members working in dot-jp to understand the degrees of satisfaction of staff members with leaders. Then, we extracted human influence networks from the e-mail archives used at dot-jp to understand the relationships between leaders and other staff members. Finally, we integrated the results of the questionnaire with human influence networks and determined reliable leadership behaviors.

2 QUESTIONNAIRES

Understanding the degree of satisfaction of staff members in an organization can be a key to discovering specific problems related to faults in human resource management (Cohen & Prusak, 2001). We sent questionnaires to all staff members (104) working for dot-jp to understand their degree of satisfaction, specifically with their managers and their branch offices. Then, by comparing the results of questionnaires received from 97 staff members (correction rate: 93%) with the achievement rate of activity, explained later, we determined factors that lead to successful activities at dot-jp.

2.1 Questionnaires

To investigate the degree of satisfaction of staff members with their branch offices, area, seminar, and internship managers, we sent questionnaires to 104 staff members working in the seven branch offices in March 2005, the last month of the 14th period. The questions in the questionnaire were as follows:

- **Q1.** Please rate your degree of satisfaction with your branch office.
- **Q2.** Please rate your degree of satisfaction with the area manager in your branch office.
- **Q3.** Please list up to three substantive leaders in your branch office.

2.2 Achievement Rate

The activity for each branch office was numerically evaluated by comparing the total number of students coming to seminars and diet members accepting student interns with the desired number for both groups set at an early stage of each period. The achievement rate was calculated using the following equation:

$$\text{Achievement rate (\%)} = x / y * 100$$

where

x= actual number of students and diet members

y= desired number of students and diet members

2.3 Assumptions of Leadership Behavior

The results of the questionnaire (Q1, Q2, Q3, and Q4) and the achievement rate for each branch office are shown in Table 2. To understand the relationships between the degrees of satisfaction and the achievement rates, we classified the branch offices into three groups based on their achievement rates.

- **Group-1:** The high-achievement-rate group, including Branches A, B, and C. Staff members in Branches A and B had high degrees of satisfaction with their branch offices and area managers. Staff members in Branch C had a high degree of satisfaction with their branch office but a low degree of satisfaction with their area manager.
- **Group-2:** The middle-achievement-rate group, including Branches D and E. The degree of satisfaction with the area manager in Branch D was quite high compared to that in Branch E.
- **Group-3:** The low-achievement-rate group, including Branches F and G. The degree of satisfaction with the area manager in Branch G was much higher than that in Branch F.

Table 2. Achievement rate and averaged degree of satisfaction for each branch office and area manager. Note that the answers were selected from five alternatives (5: Very satisfied, 4: Satisfied, 3: Neutral, 2: Dissatisfied, 1: Very dissatisfied).

Branch	Achievement rate (%)	Averaged degree of satisfaction	
		Branch	Area mgr.
A	135	4.78	5.00
B	109	4.37	4.37
C	106	2.69	3.00
D	97	3.58	4.50
E	99	2.69	2.75
F	79	2.87	3.69
G	69	2.82	4.50

We did not consider that these groups reflect the classification of management ability of area managers because the achievement rates depend on uncontrollable external factors. It means, however, that the degrees of satisfaction within the same group reflected the management ability of area managers since they excluded effects related to the achievement rate. We made an assumption regarding the management ability of area managers by comparing branch offices in the same groups with different degrees of satisfaction with area managers.

Assumption: The high degrees of satisfaction with area managers come from a high level of management ability.

We interviewed 10 staff members individually, face-to-face, and made sure that the assumption was empirically getting to the point, (i.e., area managers of Branches C, E, and F were not trusted regarding their management ability, whereas area managers of other branches had the esteem of their staff members). In the following sections, we further investigate leadership behavior by comparing human influence networks obtained from the archives of e-mail with this assumption.

3 HUMAN INFLUENCE NETWORKS

3.1 IDM

The IDM was originally an algorithm for measuring values of influence of messages, senders, and terms from threaded messages (Matsumura, 2003) (Matsumura, 2005a (Matsumura, 2005b). In the IDM, the influence between a pair of senders is defined as the number of terms propagating human-to-human influence via messages. Details are listed in the above references.

By mapping the propagating terms between senders, a human influence network can be obtained, where the number of outgoing terms corresponds to the values of influence ($I_p^{<out>}$) and the number of incoming terms corresponds to the influenced values ($I_p^{<in>}$). From the figure, the relationships of influence among the staff members can be visually understood.

3.2 Verification of Assumption

We applied the IDM to the archives of e-mail in each branch office and obtained $I_p^{<out>}$ and $I_p^{<in>}$ as influence and influenced values of each staff member. The top six staff members by $I_p^{<out>}$ for each branch office are shown with $I_p^{<in>}$ in Table 3, where the same names across branch offices represent distinct staff members². The staff names with “*” (asterisks) represent area managers. From Table 3, the following tendencies can be seen:

- Three out of seven area managers have the highest $I_p^{<out>}$ value in their branch offices.
- Those who have high $I_p^{<out>}$ values do not always have high $I_p^{<in>}$ values, and vice versa.

Table 3. For each branch office, the top six staffs by $I_p^{<out>}$ are shown with $I_p^{<in>}$. Note that the same staff names in different branches represent distinct staffs, and the staff names with “*” (asterisk mark) represent area managers.

Staff in Branch A	$I_p^{<out>}$	$I_p^{<in>}$	Staff in Branch B	$I_p^{<out>}$	$I_p^{<in>}$	Staff in Branch C	$I_p^{<out>}$	$I_p^{<in>}$
p2*	541	827	p1*	841	385	p6	1113	403
p39	518	0	p9	772	1051	p1*	836	142
p18	506	177	p5	542	317	p8	515	1257
p4	470	244	p2	484	666	p18	490	95
p41	454	910	p11	304	198	p2	470	163
p5	417	235	p17	251	537	p17	405	261

Staff in Branch D	$I_p^{<out>}$	$I_p^{<in>}$	Staff in Branch E	$I_p^{<out>}$	$I_p^{<in>}$	Staff in Branch F	$I_p^{<out>}$	$I_p^{<in>}$	Staff in Branch G	$I_p^{<out>}$	$I_p^{<in>}$
p1*	1495	717	p4	1022	8	p11	578	524	p3	779	1270
p4	923	68	p3	1019	1032	p20	506	871	p4*	559	623
p13	893	1880	p11	754	563	p18	432	52	p6	537	36
p3	794	927	p7	737	664	p1*	404	263	p8	521	466
p5	790	386	p1*	686	849	p16	355	98	p1	513	480
p14	655	205	p2	560	491	p28	325	0	p28	445	0

² For example, p1 can be seen in every branch office, but it represents distinct staff members.

To investigate further based on the assumption in Section 2, we compared $I_p^{<out>}$ and $I_p^{<in>}$ with the results of Q5 of the questionnaire where staff members were asked to pick up to three staff members to be substantive leaders. The number of votes represents how well the management ability of staff members was evaluated regardless of their position. The rankings of area managers by votes, $I_p^{<out>}$, and $I_p^{<in>}$, shown in Table 4, support the assumption, that is, area managers with high $I_p^{<out>}$ and $I_p^{<in>}$ values tend to have a large number of votes in Branches A, B, D, and G.

Table 4. Rankings of area managers for each branch office

Branch	Ranking of area managers		
	Vote	$I_p^{<out>}$	$I_p^{<in>}$
A	1st	1st	3rd
B	2nd	1st	4th
C	N/A	2nd	14th
D	2nd	1st	6th
E	4th	5th	4th
F	N/A	4th	8th
G	1st	2nd	3rd

By contrast, area managers in Branches C and F did not get any votes. In these cases, we identified the medium rankings of $I_p^{<out>}$ (second in Branch C and fourth in Branch F) and the low rankings of $I_p^{<in>}$ (fourteenth in Branch C and eighth in Branch F). Considering the results with the assumption that the lack of management abilities of area managers in Branches C and F arises from their low level of $I_p^{<in>}$ (i.e., they rarely replied to e-mails sent by other staff members).

For Assumption 3, the $I_p^{<out>}$ value of the area manager in Branch E ranked much lower than the one in Branch D, whereas the rankings by $I_p^{<in>}$ were almost uniform. Thinking of the assumption, the low degree of satisfaction with the area manager in Branch E comes from the low level of $I_p^{<out>}$ (i.e., he rarely sent useful e-mails to other staff members.).

In Branch G, the area manager ranked first by votes and ranked high by $I_p^{<out>}$ and $I_p^{<in>}$ as well although the achievement rate was the worst (only 69%). From this result and the assumption, the degree of satisfaction with the area manager in Branch G comes from the high level of $I_p^{<out>}$ and $I_p^{<in>}$ (i.e., he communicated often with other staff members by e-mail.).

In summary, having high $I_p^{<out>}$ and $I_p^{<in>}$ values should be required of area managers. They should send useful e-mails to get high $I_p^{<out>}$ values and reply back to e-mails sent by other staff members with as much information as possible to getting high $I_p^{<in>}$ values.

3.3 Interpretation of Human Influence Network

The graphs of human influence networks help us understand the behavior of area managers from a structural point of view. The graphs of the IDM obtained from the archives of e-mails of Branches A, C, F, and G are shown in Figures 1, 2, 3, and 4 respectively. Here, the figures are composed of the top six most influential staff members and influential links between them with values of influence of more than 10. In the figures, area managers are depicted as gray nodes, and the relationships among staff members are shown as directed links with values of influence. These figures illustrate the behavior of area managers toward their staff members. We summarized the leadership behaviors based on human influence networks and the analysis of the questionnaires.

- **Interactive behavior:** In Branches A, B, and D (Upper left figure in Figure 1), for example, the behaviors of the area managers are “interactive” enough to give and receive influence from other staff members. The questionnaires showed that these area managers are highly trusted.
- **Partially interactive behavior:** In Branch E and G (Lower right figure in Figure 1), for example, the behaviors of the area managers are “partially interactive”, that is the area managers receive influence from five staff members while influencing three staff members. From the results of the questionnaires, the reason that the area manager in Branch G succeeded in managing the Branch came from the high influential and influenced values. By contrast, the area manager in Branch E failed because of the low influential and influenced values as compared to other managers.
- **Preferential behavior:** In Branch C (Upper right figure in Figure 1), the behavior of the area manager is “preferential”, stricter than partially interactive behavior, that is the area manager influences four staff members and is influenced by only two staff members. The lack of interaction, specifically from his low values of influence, resulted in the failure of managing the Branch.
- **Passive behavior:** In Branch F (Lower left figure in Figure 1), the behavior of the area manager was “passive”, that is the area manager was influenced by four staff members and influenced only one staff member. The apparent lack of communication caused staff members to distrust the area manager in this Branch.

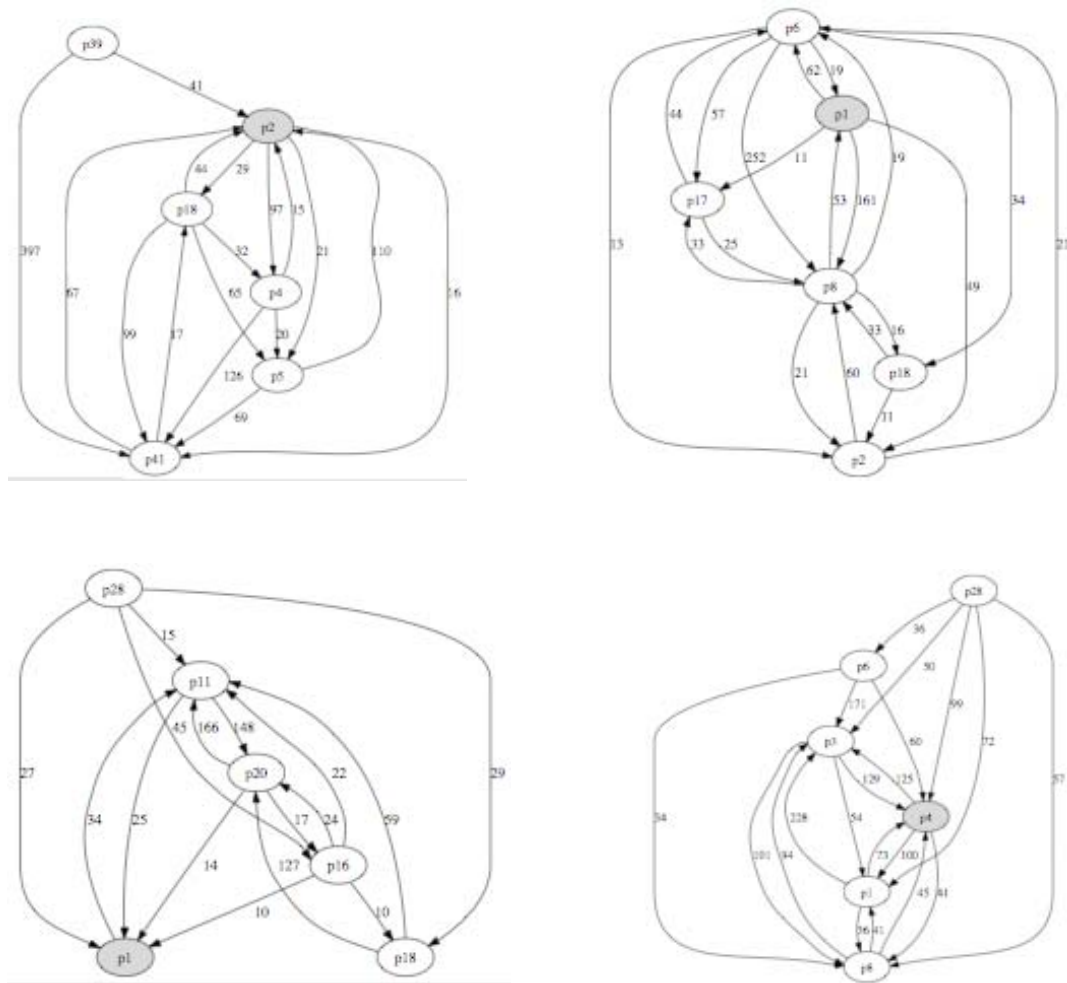


Figure 1. Human Influence Networks (Upper left: Branch A, Upper right: Branch C, Lower left: Branch F, Lower right: Branch G)

From these results, we concluded that interactive behavior is ideal for leadership. By relating the interpretation of human influence networks to $I_p^{<out>}$ and $I_p^{<in>}$, we identified the types of leadership behavior as follows:

- **Interactive:** High $I_p^{<out>}$ and high $I_p^{<in>}$
- **Partly interactive:** Middle $I_p^{<out>}$ and high $I_p^{<in>}$
- **Preferential:** High $I_p^{<out>}$ and low $I_p^{<in>}$
- **Passive:** Low $I_p^{<out>}$ and high $I_p^{<in>}$

This classification provides the possibility of other behavioral types, such as *high $I_p^{<out>}$ and low $I_p^{<in>}$* ;

however, further research is needed.

4 CONCLUSION

The activities of NPOs are based on the voluntarism of staff members. Unfortunately, staff members often lose this motivation and eventually leave organizations. Part of the reason for this comes from staff members' dissatisfaction with the lack of their leaders' management ability. We proposed an approach for integrating questionnaires and IDM analysis and determined ideal leadership behavior.

For future work, we will apply the proposed approach to other popular communication tools, such as weblogs or BBSs, to accumulate further knowledge on leadership behavior. In addition, we are planning to establish guidelines regarding leadership behaviors relating to communication to put the results of our research to use.

5 REFERENCES

- Cohen, D. & Prusak, L. (2001) *In Good Company: How Social Capital Makes Organizations Work*, Harvard Business School Press, Cambridge, MA
- Drucker, P. (1993) *Post-Capitalist Society*, *Harper Business*
- Krackhardt, D. & Hanson, J. (1993) Informal networks: The company behind the chart, *Harvard Business Review*, pp. 104-111
- Krogh, G., Ichijo, K., & Nonaka, I. (2002) *Enabling Knowledge Creation: How to Unlock the Mystery of Tacit Knowledge and Release the Power of Innovation*, Oxford University Press
- Matsumura, N. (2003) Topic Diffusion in a Community, In Ohsawa, Y. & McBurney, P. (Eds.) *Chance Discovery*, Springer Verlag
- Matsumura, N. (2005a) Collaborative communication strategies in online community, Proc. of the *Fourth International Workshop on Social Intelligence Design (SID2005)*
- Matsumura, N. Llorca, X., & Goldberg, D. (2005b) Mining directed social network from message board, Proc. of the *Fourteenth International World Wide Web Conference (WWW2005)*, pp. 1092-1093
- OSSIP04 (2004) Ossip center for nonprofit research & information: npo 2004 white paper-- the Japan nonprofit almanac --, Technical report, OSSIP, Osaka University
- Perkins, D. & Wilson, D. (1999) Bridging the idea-action gap, Knowledge Directions, *The Journal of the Institute for Knowledge Management* (1), pp. 65-77
- Wenger, E. (1999) *Communities of Practice: Learning, Meaning, and Identity*, Cambridge University Press